HOW THE NEW YORK TIMES USES INFORMATION GRAPHICS AND DATA VISUALIZATIONS FOR HARD NEWS AND SOFT NEWS AND TO FOSTER AUDIENCE ENGAGEMENT

A Professional Project

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In partial fulfillment

of the Requirements for the Degree of

MASTER OF JOURNALISM

by

NG, YEE MAN (MARGARET)

May 2014

Committee:

David Herzog, Chair
Scott Swafford, Member
Barbara Cochran, Member

ACKNOWLEDGMENTS

As I reflect on this journey I must first acknowledge my faith, which has been my strength and my shield. As I moved along this path, there have been many people who prayed for me and believed in me and in my ability to finish the race. I thank each and every one who has helped me reach my goal.

To the professors who served on my committee, Prof. David Herzog, Prof. Scott Swafford and Prof. Barbara Cochran: I express my appreciation for your guidance, hours of assistance, inspiration, and input. This project could not be done so successfully without all of you. Thank you.

To Prof. Yong Volz: Thanks for all your advice and suggestion on my academic career. I appreciate all your words. They help me to become a better scholar.

To my adviser at CPI, David Donald: Thanks for being such a great mentor and boss. Your patient and your sharing mean a lot to me. Thanks for your time to listen to my inner thought and hesitation on my future career. Your encouragement always touched me.

To Ruisha and Qing Tian: I am so blessed to meet you both here in Mizzou. I enjoy every time we gathered together to share our thought. I cherish every moment we thrived together. Thanks for being my true friends.

To Jenny Chun: You are my long time friend and confidant. You are the one who always willing to listen to me even we are separated by the Pacific Ocean. I appreciate every little cards and gifts you gave me. They gave me encouragement during my time lonely staying abroad.

To my parents: Thanks for being so confident in me and let me go out and try. When I was alone these years, pieces of childhood memory appeared. I value every hardship we faced before as a family. Those experience shapes me to be a much stronger person. I am sorry that I still cannot give support the family financially and you both do not have a chance to come to my graduate ceremony, but I hope you would feel proud of me.

To my younger brother and sister: Sorry that I don't have time to do many things for you both. I wish I set a great model for you both to you dare to achieve a much higher goal in life.

To Liu Kun: Thanks for being with me and promise you would not leave me alone and walk away from my life. I love your childish giggling. They are my source of energy. I love you.

To myself: Thank you so much for your hard work these years. I understand you more along the way. Thanks for thriving through and you finally survive among all the tears and hardships. It is time to start a new chapter of your life. I wish you all the best. Add oil!

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Chapter One: Introduction

Visual communication has always been my strength. Since my childhood, I have learned best from visual displays such as diagrams, illustrated textbooks, maps, PowerPoint slideshows, photos and videos. I drew charts, graphs and tables in all of my high school notes to help me memorize all those complicated biological terms or chemical reactions. I had a sketchbook to draw mind-maps to help me develop ideas to write a story.

My graduate study at the Missouri School of Journalism extended my definition of visual communication. I found that storytelling is not restricted to words, photographs or videos. Data and graphics are other promising devices for journalists to tell stories. The combination of data analysis and graphic visualizations can be very powerful. Therefore, I planned my graduate coursework in this growing field, starting with my first semester at the J-school. I have taken courses such as Information Graphics, Computer-Assisted Reporting, Introduction to Geographical Information Systems and Convergence Reporting. I dusted off the coding skills I had learned in high school and started to play with MySQL, Python, Ruby and JavaScript. I did an independent study on interactive data visualizations with JavaScript/d3 library. I built interactive maps with the Google API. I explained government spending and politicians and business actions in infographics and visualizations. I attended the Society of News Design conference, where I met design professionals from around the world. My working experience at the Columbia Missourian, National Geographic Magazine and Investigative Reporters and Editors as a graphics reporter or a web designer have provided me valuable opportunities

to explore new forms of storytelling at the confluence of design, data and coding. All these have prepared me for my master's professional project, working in a professional newsroom and studying how The New York Times uses infographics and data visualization.

This professional project will be another valuable experience for me to gain insight and experience working at a professional newsroom. Every inspiration I gain and storytelling skill I learn will help me achieve my creative and career goals, as a data reporter or a visual communication researcher.

For my professional project, I chose to participate in the J-school's Washington Program in Washington, D.C. I worked at the Center for Public Integrity as a data and interactive design intern. I worked from Jan. 16 to April 23, 2014, 30 hours per week.

Chapter Two: Field Notes

Weekly Report 1

I started my internship at the Center for Public Integrity (CPI) on 16th January 2014. In the coming 14 weeks, my data reporting and visualization skills will be put to the test.

My advisor, David Donald, is a very nice and patient mentor. In the very first day, he gave me some of his training materials for IRE computer-assisted reporting class as warm-up exercises. I recalled and picked up my SQL skills very soon and the exercises were not tough. However, when I was first introduced to the Federal Election Commission (FEC) database, I got nervous. Many questions came to my mind: what does the FEC do? What are federal campaign finance laws about? How often are FEC databases updated? What is a political action committee? What is the donation process? I soon realized that I should start from the basics and get to know how the election commission system works in detail. This reminded me once again that good data reporting is not about the how advance the tools we use, but about veteran journalistic research and thinking process. Even if I have the best data analysis skill, I need to know how the society and political system works as this is what journalist is meant to be.

My first investigative project is to examine the growing trend of medical advisory groups' donation and the increasing expenditure on lobbying for health care issues by the health insurance companies in the past 10 years. I would have several weeks to work on this data reporting.

Besides this project, I was also involved in a project called, “Secrecy for Sale: Inside the Global Offshore Money Maze” with ICIJ. I helped proofread and fact checks the Chinese edition and the interactive components. During the past week, Offshore Leaks Asia Release gained more than a million pageviews. We were very excited.

After the first week at CPI, I was so impressed with “slow journalism”. Investigative journalism takes more time and more experienced journalists to produce, and it often involves legal battles. It is generally the most expensive work the news media undertakes. However, the Internet is killing mainstream media, sending the Fourth Estate into record-breaking revenue declines. Online ads garner only a fraction of the dropping print revenue. When faced with cuts, investigative reporting is often the first target. CPI is set up to be a non-profit journalism center to focus on “slow journalism”. My advisor, David Donald, told me he turned down a job offer from the Wall Street Journal and chose to work for CPI because he wanted to do his whole investigative reporting instead of chasing deadlines. I think I would love getting to work in this newsroom too.

We had two guest lectures in this week’s seminar. One was held at the American Political Science Association. Jeff Biggs, the director of the Congressional Fellowship Program, shared with us his insights in the Congressional Research Service. It was a new idea for me to have teams of political scientists and journalists to work exclusively for the United States Congress, providing policy and legal analysis to committees and Members of both the House and Senate, regardless of party affiliation. Mr Biggs generously offered the tips of making the most of our Washington experience and a list of mailservs to receive daily update of foreign and domestic policy. I have subscribed to several of them. Another guest lecture was held at NPR. We had lunch with Jessica Pupovac,

producer/editor at NPR's Morning Edition. She gave us useful advice to live in DC and excel in an internship. She reminded us not to afraid to make mistakes and keep learning. After her sharing, we had a tour at NPR's new digital facility. I really like the open space working setting at NPR.

For the professional analysis part, I started to gather the data visualizations I need for the analysis and reach out to contact my potential interviewees at the New York Times. Unfortunately, I had a skin allergy for a week that prevented me from doing much research, however, I would start to work right after I get better.

Weekly Report 2

I spent most of my time working on FEC data analysis with SQL and Excel last week. I am very happy to see I have become more familiar with the format of the database. This helps speed up my thinking process significantly.

I reviewed the history of the FEC database: Ten years ago, the Federal Election Commission introduced electronic filing for political committees that raise and spend money to influence elections to the House and the White House. The filings contain aggregate information about a committee's work (what it has spent, what it owes) and more detailed listings of its interactions with the public (who has donated to it, who it has paid for services). Journalists who work with these filings extracted their data from complex text files and turn them into usable data using the FEC's data dictionaries to match all the fields to their positions in the data.

When dealing with the tables, I found that available fields have changed over time, and subsequent versions did not always match up, such as the time structure. Before 2008, filings were done biannually, but, starting from 2009, filings were done quarterly.

To ensure a fair comparison, I need to make sure I analyze data in the correct time frame. What's more, I need to pay extra attention to avoid data duplication. Sometimes, both contributors and receivers filed the same record twice, and for some of the cases, the donation was refunded.

A more concrete story budget is: We want to see how Medicare Advantage has morphed into a powerful force by analyzing industry campaign contributions, growing lobbying by Medicare Advantage groups and contributions from individuals employed by Medicare Advantage. The key questions to me: what has been the pattern in giving and how has it changed from 2004 through the 2012? I would focus on the top 10 health insurance companies with highest enrollment and some of the top advocates for Medicare Advantage, including Senate Orrin Hatch, Mike Enzi, Mitch McConnell, Marco Rubio, etc.

We had two guest lectures in this week's seminar. In the morning, we met with the communications team for Sen. Claire McCaskill, D-MO. Anamarie Rebori, the New Media director, told us that the senator actively uses social media. Senator McCaskill has a YouTube channel and a Tumblr blog that she uses to share information with her constituents and to respond to questions. She has been using Twitter regularly and is able to utilize social media to communicate. For politicians, authenticity is an important part of PR. I believe the public responds best to an authentic touch over Twitter, and that is something Senator McCaskill has been able to convey in her use of social media.

In the afternoon, we had a free tour in Newseum. This was my third time to visit Newseum. Attracted by all the compelling pictures and detailed records of news history, I have to say it is my favorite museum in the world. We met with producers of Al Jazeera

America's America Tonight. While mainstream media focus doing stories on sports, entertainment, politics, Al Jazeera aims to tell stories which are often conspicuous by their absence, such as topics about immigration or prisoner rights. It is nice to see Al Jazeera America focus on the other end of the news spectrum.

For the professional analysis part, I started to review the results returned from the New York Times' search engine and eliminated the results not related to my topic on data visualizations. I am also arranging a trip to join the NICAR conference in late February to meet some of my potential interviewees.

Weekly Report 3

In data journalism, we always talk about how beautiful data could be. Data reveals the public truth and brings powerful impact to political change. However, without systemic analysis and an understandable presentation, data could be very hard to understand.

This week, I worked mainly with ill-defined, chaotic and incomplete data within the FEC database. There were mistakes in the federal filings and the tables failed to capture some data in the filings. I struggled whether I should call and clarify the data I have. But if I plan to do that, how many calls would I end up doing? I turned my question to my advisor. David told me bad data is a fact of life. Coping with bad data is a valuable, learned skill. It is great that I fight for accuracy for the information we need. But some time, for some data, I should think about whether or not it is worthy to spend time to make it clean. Data is just a reduction of reality. Besides learning how to deal with data, we also need to learn to have correct judgment on data.

This week, I found that it was generally easier to do some basic analysis of data to

look for problems: gaps, inconsistencies, unusual distributions. Doing so will give me insight into what I am dealing with. Going through the actual data file, rather than trusting the metadata and documentation, is the only way to really know what sort of issues are lying in wait.

What's more, there is lots of interesting data that's structured for human consumption rather than machine-driven analysis. Restructuring it to be in a format that is more amenable for machine analysis can be tedious, but it is also automatable. I learnt that I should not shy away from writing code to transform data into something useful and always expect that to be an iterative process.

This week our guest speaker is Mike Dorning, the White House correspondent for Bloomberg. I was particular surprised by the story that a former Bloomberg reporter wrote about Goldman Sachs partner's employment status after noting the person had not logged on in the Bloomberg Terminal for some time. Bloomberg, whose computer terminals are widely used on Wall Street, had allowed journalists to see some information about terminal usage, including when customers had last logged in, and how often they used messaging or looked up data on broad categories, such as equities or bonds. It might be a valuable way for journalists to break in a story, but it raised privacy concerns.

I have been spending time on picking up my coding skills in D3/JavaScript, as I would work on data visualization next week. Therefore, I did not spend much time on my professional analysis this week.

Weekly Report 4

We started our interactive news project about the controversy of the spectrum auctions between wireless carriers this week.

The Federal Communications Commission is preparing to sell off more of the airwaves that are crucial for wireless companies to continue to deliver the increasing amount of information to smartphones and computer tablets. To remain competitive, smaller wireless carriers such as T-Mobile will need to win a significant chunk of the so-called spectrum, or likely never be able to compete with AT&T and Verizon.
The big stakes have set off a lobbying frenzy in Washington, where T-Mobile, Sprint and smaller companies are fighting to ensure they get a fair shot by calling on the FCC to limit how much bandwidth their giant competitors can buy. Verizon and AT&T, in turn, are pushing the commissioners not to limit how much they can buy.

We came up with lots of graphic ideas, such as spectrum propagation graphics, lobbying spending graphics, average cell phone bill, etc. I started with illustrating a table to show the professors or scholars hired by each firms during the lobbying.

At last week's seminar we had USA Today's Donna Leinw and Leger to share her experience with disaster reporting. She has been to Baghdad to cover the war, Indonesia and Thailand after the tsunami, London during the subway bombings and Haiti after the massive earthquake that struck Port-au-Prince. She reminded us that a war and disaster reporter needs to be self-sufficient to survive, be technical-sufficient to fix your own computers and file reports and be active to get involved in others' cultures. Her experience was so unique, yet challenging and appalling. I really admired her job and her personality. Everyone can imagine her trips were full of hardship, but during her sharing, she described all her encounters in a really fun and relaxing way. I can see she is a very easygoing person and she is always prepared her next assignment.

For the professional analysis part, I successfully contacted with Aron Pilhofer at

the New York Times. Unfortunately, he mentioned he was not the right person for data visualization. But, he was so nice and suggested that I reach out to other graphic editors like Matt Ericson and Steve Duenes. I would keep on sending emails.

Weekly Report 5

I continued my week with FEC data analysis and the interactive news design for the spectrum competition.

It was the first time I got to code with some professional coder with a computer sciences background. I used to finish a coding project all on my own. Everyone just saw the final product instead of my "messy" codes behind the scenes. I finally learnt some coding manner that I did not learn since I am self-taught. It was really great there were professionals to clear my thought and teach me the basics that I thought I knew.

My first HTML table showing the professors or scholars hired by each wireless communication companies during the lobbying went very well. My team liked it and I was satisfied with my work too. I am looking forward to working on more interactive graphic works.

"Coding" is an endless hole of knowledge, trial, error and repetiton. New languages and platforms are introduced. It takes time to become familiar with a new language, but there are elements that are similar across languages. If you understand the syntax of one, it isn't that hard to learn the syntax of another. The real difficulty is in your approach to problem solving, as mentioned above. How exactly do you break down the challenge to make something that works the way you envisioned? It's moving something from idea to execution that is thrilling. Learning to code should be part of a lifelong learning strategy.

For the professional analysis part, I have successfully set interviews with Aron Pilhofer and Alastair Dant of the New York Times during the NICAR conference the coming week.

We did not have seminar this week.

Weekly Report 6

I was at the NICAR conference last week. It was a privilege to visit and be exposed to ideas, stories and techniques from the cutting edge of investigative reporting, data visualization and the development of news applications.

I used to spend a lot of time learning about data publication and visualization tools, but this is just one small slice of the needs of computer-assisted reporting. A story might take months of investigative work- gathering data, cleaning data, interviewing people, assembling scraps of paper — and a presentation of that data is only prepared in the final run-up to publication. That presentation isn't always a wiz-bang interactive graphic, either. Many times a data-intensive story might be presented entirely as narrative, if the medium fits better. There is so much for me to learn.

For my professional analysis part, I have finished the interviews with Aron Pilhofer, the associate managing editor of digital strategy at the New York Times, and Alastair Dant, the interactive developer at the New York Times. During the interview, I got to know personalized interactive elements, such as quizzes, are very popular in the newsrooms. However, I found that the New York Times had not done a lot of user analytics and trackings on their online news (according to Aron, there was only one person working on the analysis in the past year). So, some of my questions, such as what kinds of visual elements generate more clicks and views, still are not supported by

statistical data. There is still much room for research on interactive news strategy.

Weekly Report 7

This week I continued to look at the contributions to top advocates for Medicare Advantage, but this time, I focused on the donation by the Board of Directors and the Executive Officers from the health insurance companies. I built a table to include every name of the senior management team listed on the companies' web and wrote a really long SQL code aiming to extract all the donation records with those names.

Then problems emerged. I needed to find a way to ensure that, for example, the names appeared as "David Jones" were 100% the "David Jones" from the Board of Director of Humana, instead of other donators with the same names. The FEC database did not have a distinct ID to distinguish every person. So I spent a long time checking their associated companies, zip codes, middle names etc to make sure they are the right person. This was a time to test my judgment with "bad data" again.

On Sunday, we had an invaluable opportunity to watch the Sunday show "Meet the Press" live at NBC studio. Several officials and politicians discussed the latest in Ukraine as well as the Malaysian airliner, one-year anniversary of Pope Francis and the future of the GOP. In class, we discussed the purpose of the Sunday show. One major impact of it is to set the political agenda for the week. A major struggle for any politician with powerful ambitions is getting enough airtime. The whole goal of becoming a political household name is largely dependent on how many appearances you can clock on all the major television networks. The best place for this political opportunity is to hit the Sunday talk shows.

For my professional analysis part, I am working on the transcripts for my

interviews.

Weekly Report 8

This week, I finally got a chance to read the draft of the Medicare story written by Fred Schulte, the senior reporter. It is a long story. There are some placeholders reserved for the data David and I are currently working on. This is a really great opportunity for me to learn how to humanize all the big data in a story.

Meanwhile, I need to do more data analysis for the story. I need to look for the finances of The Coalition for Medicare Choices and the other group called The Association of Mature American Citizens. Another thing we are paying extra attention to are the dates of contributions. Do they suggest to influence a proposed Health plan rate cut? David reminded me, data reporters are not just looking for rows and columns, we also need to read through plenty of pdf reports in order to get a grasp of the financial situation of a particular organization.

On Friday, we went to meet with two lobbyists, Terry Bracy and Jim Brown, who represent lots of municipalities, including St. Louis. Things I learnt were money and information play the biggest role in Washington politics. But information, in some sense, is more powerful than money. A big story told in a right time and through a right platform could “kill anything”. Brown summarized his work as “pick up a phone call, offer solution to fix the problems, go through the bureaucratic process, gather all the senates, congressmen to sit down and get them all tuned.” This sounds very straightforward. During class, we discussed a lot about bad reputation of being a lobbyist. I think I need to do more critical researches on the news about lobbyists.

For my professional analysis, by the end of this week, I would finish the content analysis of 600 datavizs I collected from the New York Times. During the spring break, I aim to finish two in-depth case analysis and the analysis of the two interviews I have. First draft of the professional project would be ready after the spring break.

Weekly Report 9

This week, the two interactive news elements I worked on finally got published. The story was called “Wireless companies fight for their futures”. My interactive elements show the lobbying spending and the professionals hired by the top four wireless carriers. It has been a long wait. But I realize that even for some really simple graphics, we still need to considerate many things, such as the graphic size, the color, placement, in order to make the entire story package coherent.

I continued with more work for the Medicare story. There were more than 160 members of Congress raised concerns to the Medicare agency about the negative impact a proposed new cut would have on seniors. The Coalition for Medicare Choices openly thanked them in a letter. Therefore, for this week, I researched on how much donation members of Congress received from the health services industry. I relied on the OpenSecret database to do the scrapping and I built an Excel file to do all the calculation.

We had Mike McCurry, the former White House press secretary for our seminar this week. He helped steer President Clinton through three and a half years of hostile media fire during the Whitewater and Monica Lewinsky years, but he joked that his daughter always mentioned, “You might be once famous, but it was last century”.

Of course, we started asking questions about the Monica Lewinsky case. McCurry talked about how he handled the press and how to balance his answers with lawyers’

advice. He reminded us to rely on basic values of decency and respect for others. One thing he said was thought-provoking. He mentioned that some investigative newsrooms, like CPI, always looking at data to assume the government is going wrong and the politicians are corrupted. He said fundraising and contribution is inevitable in political world. Many politicians actually are doing good things and the government passes good laws. But positive things done by them could not make to the headlines of newspapers. This made me have a second thought on my attitude when I am doing my investigative reporting. I think a reporter should open to the possibility that corruption is there but not being biased against the possibility that good work is being done as well.

Weekly Report 10

I finished looking at the donation received by the 160 members of Congress mentioned in the Coalition for Medicare Choices' letter from the health service industry. At first, I just counted the money directed to the campaign committees of the members. But later, I found that my analysis did not cover the entire donation exhaustively. I found that the Leadership PAC of these members was also a major recipient. So, I redid the whole table again. That reminded me to think comprehensively before starting to do any analysis.

This week, we were very happy that Chris Zubak-Skees, my colleague who is my "programming mentor", won Melofiej International Infographics Awards bronze medal in Online Graphics Portfolios for his "Breathless and Burdened". I was very excited, as Melofiej is such a big award to honor the best infographics. That reassured me of the potential that the investigative newsroom has for creative graphics and interactive elements to do storytelling.

We did not have seminar this week.

Weekly Report 11

This week, I worked on the FEC dataset to see if the dates of contributions reveal the influence of congressmen to vote on the Health plan rate cut. The vote was held on 1 April 2013. Therefore, we set up two time frames to make a comparison. One is from 1 March to 1 April, right before the vote, and the other is from 1 April to 1 May, right after the vote. We could investigate if there is any obvious increase of donation because of the voting period.

This time frame was easy to understand and to compute. The main challenge I encountered was to understand all the relationships among the tables I got in hand from the OpenSearch. I spent a lot of time reading the documentation to figure out which tables I should join together to get the result. For example, there was a table for individual donations to candidate, a separate table for PAC contributions to campaign committees, and one for PAC donations to leadership PAC… etc. But sometimes, there might be duplication of records, which make things complicated.

This week, we went to the Covington and Burling law firm. Steven Weiswasser, a renowned media attorney and former vice president of ABC News, and several of his colleagues gave us an informative lecture and notes on current issues in media law. We spent some time to discuss about FOIA request. They were very honest to say bringing a lawyer to the negotiation would lengthen the time for the process. Sometimes, developing a personal relationship with the agency, rephrasing the request to a more specific way help us get the data faster. They reminded us “We have the right to request records, but we don't have the right to ask the agency to create a record for us.”

Weekly Report 12

It has been a very busy week for me. I started a new project to review the most recent three years of personal disclosures for all 258 sitting federal appellate judges. Our investigation found more than two dozen instances in which judges have broken the rules and ruled on cases in which they had a financial interest, even though the judges have been warned in the past about such conflicts and were clear rules and systems to prevent them. Thirty more cases involving potential conflicts of interest in courtroom were found. I would make an interactive map to show the divisions of circuit of court of appeal in US. Also, there would be box plots to show the amount of stocks the judges held. I found that the coding part is not an easy task this time as there is not much reference out there in the Internet, but I am happy that I can do something original.

This week, we had two seminars, one with Clarence Page, a Pulitzer Prize-winning liberal columnist, for the Chicago Tribune and the other one with Fred Barnes, a conservative columnist and executive editor of the Weekly Standard. I learned that columnists are free to express their personal opinions as that is the primary part of their job. Page said he got far more emails from his readers when he started his career as a columnist than he was a reporter. A lot of his loyal readers were actually people who disagree with what he said. “There are lots of people out there love to argue and criticize.” he said. Barnes commented “future politics is something we write so much about, but you know nothing about.” Politic is not a science. There is no path of projection. Barnes further elaborated his views, “Journalism on politics is based on the current assumption, that is who is popular now, would keep on being popular. But that is not the case. Politics could change dramatically within a short time.”

Weekly Report 13

Last week was quite a week in the history of The Center for Public Integrity. The Center received its first Pulitzer Prize, for Chris Hamby's groundbreaking investigation revealing how the coal industry rigged a system to deny benefits to coal miners stricken with black-lung disease. We were all astonished. Although it was the last day for Chris on Wednesday, the Center thanked him for all the great work he has done for the Center and wished him the best to continue his stellar journalism career at Buzz Feed.

For this week's seminar, we went to The Washington Post, which won two Pulitzer Prizes this year, included the prestigious public service medal for a series of stories that exposed the National Security Agency's massive global surveillance programs. Jeff Leen, investigations editor of the Washington Post, gave us a tour around the building and we got a chance to have a close look at the Pulitzer Prize' medals Washington Post won. Leen shared that investigative reporting was expensive and risky. "There are many ways to fail and could cost the newspaper a lot of money." Leen said. Therefore, resources of the newsroom were always allocated to stories that impact wide scope and do a lot of harm. Leen reckoned the best Investigative reporting was those finding patterns and breaks. Human sources and documents were the main tools the Post used to find the official secrecy.

This week, I kept working on with my final data analysis on Medicare care stories and the interactive map and chart for the appellate project.

Weekly Report 14

This is the last week of the internship, and it feels bitter sweet. I am happy to go back to Mizzou because I can finally claim a completion of my research project. On the

other hand, I feel like I am more experienced in database and coding now, and oddly feel like I am leaving a job that I enjoyed and learned so much.

This week I have been busy trying to finish all my work. I wrote the documentation and went through every database I built for the Medicare story with David. We found some mistakes and worked on the correction. I tidied up the code for the interactive project and passed it to my colleague for publication next week.

David had a final talk with me and told me he was very satisfied with my work attitude. I am one of the few interns he would consider to recruit if there is position open at the Center. What he said was a great recognition to me. I was so blessed to have him as my advisor and mentor, who I felt very comfortable to share my thought to. I cannot put into words how valuable my time there has been. I am wiser and have a greater understanding in the working process of investigative journalism. I am so thankful for the opportunities, guidance and help I received from everyone I met with the company.

Chapter Three: Evaluation of work experience

Self-evaluation

I am so grateful to be part of the Center for Public Integrity. This is my very first time to work at an investigative newsroom and do data-driven reporting. I am so honored to get to work with the investigative reporters as a newbie data intern and learned a ton from building the database, creating interactive graphics and doing data analysis with all my awesome mentors. David Donald was the first to treat me like a real reporter and taught me how to deal with real-world data. Alex Cohen showed me how to find untold stories in a massive database, and Chris Zubak-Skees gave me short and intense lessons on programming. My mentors remind me what makes journalists great is that they have an instinct to self-teach and explore.

My days at CPI often involved writing SQL queries, analyzing data, brainstorming graphics and designing interactives. I focused on the FEC database and used Excel and T-SQL for data analysis and Javascript/D3 and chartjs for web graphics. NodeXL crept into my workflow for exploring the network relationships and experimenting with visualizations. I paid attention to all my assignments, and I think I did some good work.

The most valuable thing I learned was that a data journalist needs news judgment and attention to detail in order to identify the newsworthiness and limitations of datasets. In every situation you face, there will be unique considerations about whether and how to publish a set of data. Statistics can help explain a dataset's strengths and weaknesses, so I wish I had paid more attention during my stats classes in school. In addition to finding the stories, data journalists also need to be able to explain why data is significant to their

audience, so visual interactive elements are there for the stories — and, of course, reporting and writing.

Another lesson I learned is not to assume data is inherently accurate, fair and objective. It is inevitable that there will be inconsistencies and errors in the FEC's data, given the size and complexity of the FEC's reporting mandate and the tendency for human error in reporting. It's important to have clean, structured, easy-to-find data — because journalism is about getting things right while beating the clock. But that does not mean one must have exact numbers. The way to present the data is all that matters. Paying attention to the margin of error and rounding-up or down is always preferred. A detailed note on methodology can help an audience understand the calculation of the numbers.

Data is at the heart of what journalism is — and the more substantive it is, the more organized it is, the more easily accessible it is, the better we all can understand the events that affect our world, our nation, our communities and ourselves. I am so happy to have spent a semester working with that dataset as an intern at the Center for Public Integrity.

Supervisor Evaluation

Evaluation of Margaret Ng

Margaret has been an outstanding intern at the Center for Public Integrity during spring semester 2014. As we would expect, she is hard working and dedicated. She took her responsibilities seriously and acted professionally at all times. We have seen such consistency in interns from the Missouri School of Journalism for many years now.

But a few traits and actions set Margaret apart. She is conscientious in everything she approaches, giving her editors confidence in the exacting work demanded of her. She accepts every challenge, even those moving her outside her comfort zone. She's not afraid to ask questions when needed, yet takes the answers and builds upon them.

And finally, there is her passion for learning and growing. We saw this early in her time with us in that she wanted us to keep pushing her so that she could develop new skills, especially in the technical side of data journalism. That was accomplished. What was refreshing was her openness and willingness to rethink working in journalism when presented new ideas and practices.

Here are a few specifics of her duties and achievements.

For an in-depth investigation of the political connections between congressmen and the private insurance industry, Margaret built a database using Federal Elections Commission campaign finance data and lobbying data gathered from the Senate Office of Public Records. From those databases, Margaret analyzed private insurers in the Medicare Advantage program donated substantial amounts of money to the key senators and congressmen behind Medicare Advantage. The analysis covered many elections cycles, showing trends over time.

Margaret then used lobbying data to show how much pressure such companies produced on Congress, especially at times of important votes on bills of interest to all stakeholders in Medicare Advantage. All of Margaret's analysis was completed accurately and on time. She asked questions when needed and sought help to make sure she understood the data and how to use them. Her work added greatly to the political reporting for this Medicare project. While the overall project was not completed before her internship ended, Margaret completed her part promptly and will receive byline credit when the project runs in the near future.

Margaret has shown good skill with her data analysis capabilities. For the next step, which usually comes only after some experience working with data, Margaret should push herself to be able to see patterns and trends in data that go beyond what's asked and find results that even she did not expect.

Margaret also worked on several of our investigative projects to contribute interactive news graphics during her Center internship.

We found her to be quite effective at getting to the heart of a story and translating it into visual journalism. It's immediately clear what her work is trying to convey: she understands the distinction between a graphic and a piece of visual journalism.

Her work adheres to industry best practices in being responsive and mobile-first, and her graphics are executed in a clean and elegant visual style.

She took the internship as an opportunity to use and create data-driven interactives using the D3 library: a tool and a skillset that will serve her well in today's job market.

When Margaret first started working on coding for us, she said that her client-side coding skill was limited to copying what other people had done and modifying it a little. But this is how everyone starts — the question is whether they limit themselves to this and give up, or persevere and go further. Margaret does not yet have a full understanding of D3 or JavaScript, which I can tell is frustrating to her, but in her time at the Center she has persevered and done a good deal more than merely copycat work.

The interactive elements she has produced show originality and inventiveness as well as quick comprehension of complex systems. She has created interactive maps, charts and displays of information that rank among the best of breed, and which, with a little bit of clean up, add immensely to the story.

If this is what she wants pursue, we encourage her to work to more fully understand JS and D3, by the hard road of continuing to work at it, and to challenge herself to become a more disciplined coder. She's made excellent progress already.

And her attitude could not be better. Margaret is willing to learn, takes feedback with a smile, is incredibly dedicated, and eager to be of use. Early on in her internship she responded to a request for help by coming into the office on a day when she wasn't scheduled to be with us. Her work that day (translation, copy editing, and some writing) was outside her internship scope, yet invaluable to the launch of a major project.

David Donald
Data Editor

Kimberley Porteous
Chief Digital Officer

Chris Zubak-Skees
News Developer

Chapter Four: Work with The Center for Public Integrity

Interactive graphics

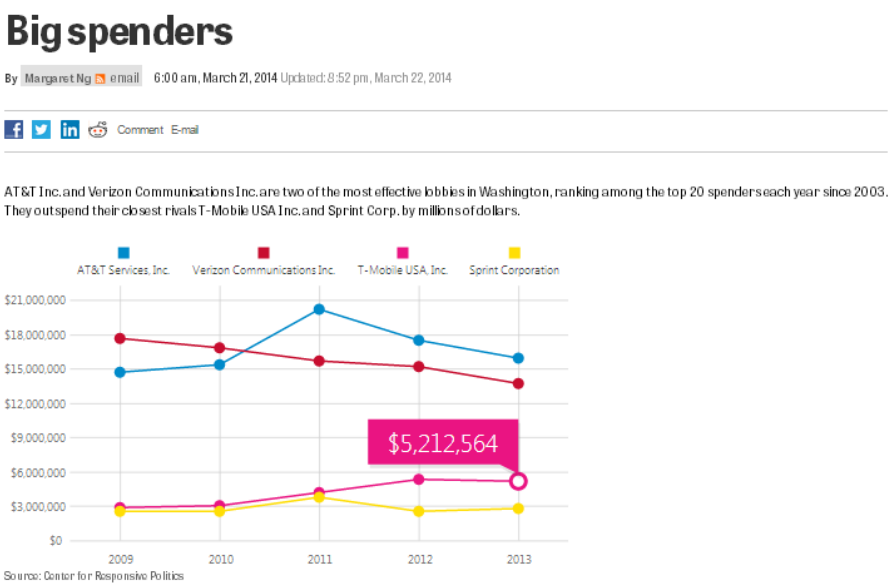


Figure 1: Big Spenders - How much wireless carriers spent lobbying.

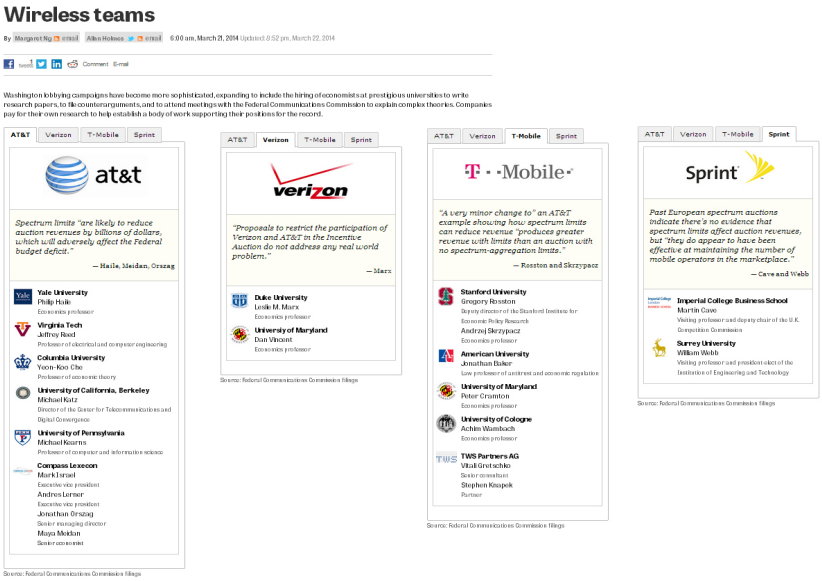


Figure 2: Wireless Team - Lobbying campaigns have become more sophisticated, expanding to include the hiring of economists at prestigious universities.

(Two more graphics would be published in late April. Here are the drafts.)

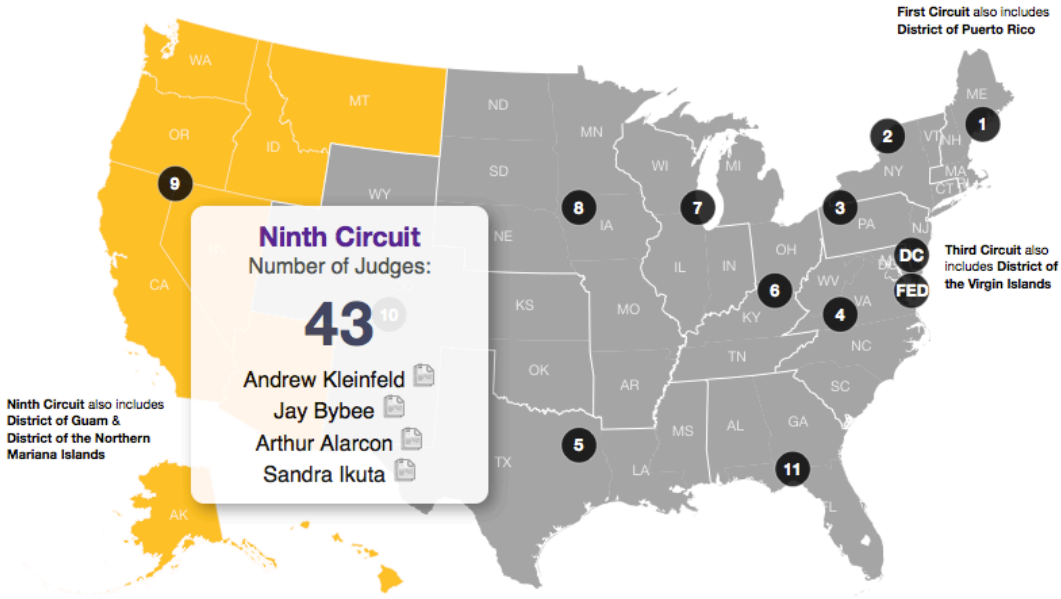


Figure 3: Federal Appellate Judges - Geographic Boundaries of United States Courts of Appeals

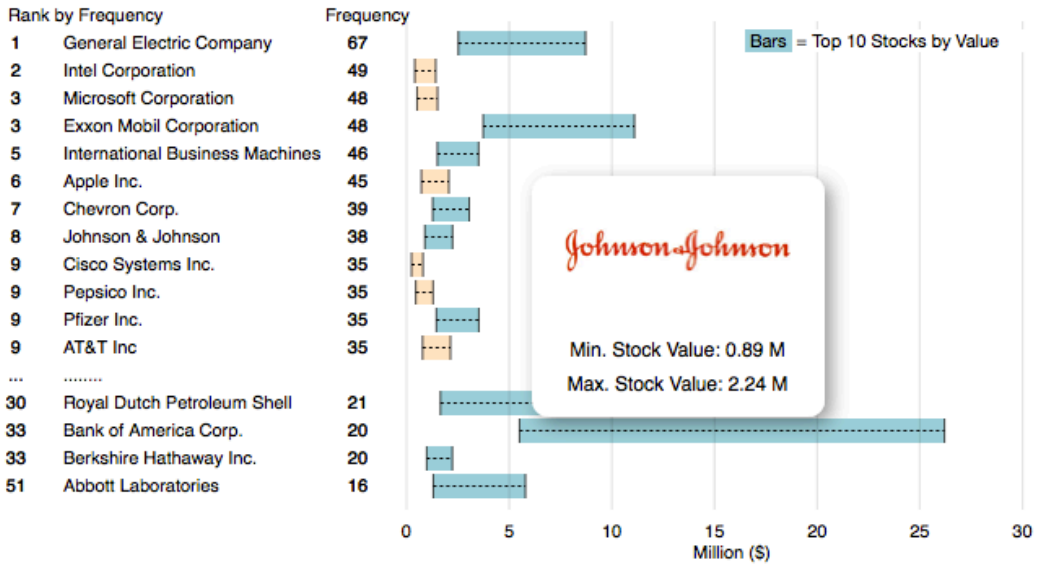


Figure 4: Stocks Judges Held - Top 10 stocks judges held by frequency & by value

Data analysis | Notes and SQL queries

(The stories would be only published in May 2014)

LOBBYING QUERIES

```
/* Basic lobbying query */
SELECT t1.*
FROM SOPR.vLD2Matcherv3 t1
WHERE t1.ClientName LIKE '%AETNA%'
AND t1.FilingYear > 2012
ORDER BY t1.PeriodShort,t1.RegistrantId

/* List of lobbyists for Aetna, eliminates duplication of lobbyists listed by alternate
names */
SELECT t1.RegistrantName,t1.ClientName,MAX(t2.LobbyistName) AS
LobbyistName,t2.LobbyistLname
FROM SOPR.vLD2Matcherv3 t1
JOIN SOPR.LD2 t2
ON t1.FilingId = t2.FilingId
WHERE t1.ClientName LIKE '%AETNA%'
AND t1.FilingYear > 2012
AND t2.LobbyistName IS NOT NULL
GROUP BY t1.RegistrantName,t1.ClientName,t2.LobbyistLname
ORDER BY t1.RegistrantName,
LobbyistName

/* Basic list of lobbyists for Aetna with former government positions*/
SELECT MAX(t2.LobbyistName) AS
LobbyistName,t2.LobbyistLname,t3.LobbyistName AS LobbyistPositionName,
t3.LobbyistPosition
FROM SOPR.vLD2Matcherv3 t1
JOIN SOPR.LD2 t2
ON t1.FilingId = t2.FilingId
JOIN SOPR.LD2 t3
ON t2.LobbyistLname = t3.LobbyistLname
AND t2.RegistrantId = t3.RegistrantId
WHERE t1.ClientName LIKE '%AETNA%'
AND t1.FilingYear > 2012
AND t2.LobbyistName IS NOT NULL
AND t3.LobbyistPosition IS NOT NULL
AND t3.LobbyistPosition <> 'N/A'
GROUP BY t2.LobbyistLname,t3.LobbyistName,t3.LobbyistPosition
ORDER BY LobbyistName
```

```
/* List of issues for Aetna */
SELECT t1.RegistrantName,t1.ClientName,
t2.Issues,t2.IssueDescription
FROM SOPR.vLD2Matcherv3 t1
JOIN SOPR.LD2 t2
ON t1.FilingId = t2.FilingId
WHERE t1.ClientName LIKE '%AETNA%'
AND t1.FilingYear > 2012
AND t2.Issues IS NOT NULL
GROUP BY  t1.FilingID,t1.RegistrantName,t1.ClientName,t2.Issues,t2.IssueDescription
ORDER BY t1.RegistrantName

/* Very basic query on Aetna employee contributions */
SELECT t2.committee_name,
t1.*
FROM Weekly.Contribs t1
JOIN Weekly.Cmtes t2
ON t1.cmte_id = t2.committee_id
AND t1.ftpcycle = t2.ftpcycle

WHERE t1.ftpcycle = 2014
AND employer LIKE '%Aetna%'
AND cmte_id <> 'C00181826' -- Remove employee contributions to Aetna PAC
AND employer NOT IN (
'AETNA BRIDGE CO.',
'AETNA FIRE ALARM SERVICE CO., INC.',
'AETNA LIFE & CASUALTY')
AND tran_tp IN ( '15', '22Y' ) --Contributions and refunds

Change to <> and examine records for transfers

/* Very basic query of Aetna PAC contributions, report from Aetna's PAC */
SELECT t2.committee_name,
t1.*
FROM Weekly.Contribs t1
JOIN Weekly.Cmtes t2
ON t1.cmte_id = t2.committee_id
AND t1.ftpcycle = t2.ftpcycle

WHERE t1.ftpcycle = 2014
AND cmte_id = 'C00181826' -- Aetna PAC
AND tran_tp = '24K' --Change to <> and examine records for transfers

/* Very basic query of Aetna PAC contributions, report from group/candidate receiving
contribution */
```

```
SELECT t2.committee_name, t1.*
FROM Weekly.Contribs t1
JOIN Weekly.Cmtes t2
ON t1.cmte_id = t2.committee_id
AND t1.ftpcycle = t2.ftpcycle

WHERE t1.ftpcycle = 2014
AND other_id = 'C00181826' -- Aetna PAC
AND tran_tp IN ( '18K' ,'22Z' ) -- Checks for contributions and refunds
```

Health Insurance Companies' spending on lobbying

Steps:

1. SQL – copy the result to Excel 0219_lobbying.xlxs
2. [Raw] Add column Internal/External
3. [Depleted_Duplicate] Remove Duplicate – with same year and period,same client,same register and same amount – 2 rows removed
4. [InitialCal]Add up External using Pivot Tables
5. [Checking]Check for missing for Total spending that have not been self reported in total
6. [InitialCal]Add up Internal using Pivot Table + missing values after checking
7. [Result]Spreadsheet manually calucation

Remarks:

A. Time period focus after checking: **2004 – 2013**
B. Database used: FROM **[LobbyingRaw].[SOPR].[vLD2Matcherv3]** AS t1 JOIN **[LobbyingRaw].[SOPR].[LD2]** AS t2

1. Aetna, BCBS, Coventry and Humana are more simple to calcuate their total spendings as they have less "branches."
2. Health Net: Health Net + Health Net Federal services. I treated self-reported spending for Health Net Federal services as part of the internal spendings for health net as a whole.
3. Kaiser foundation Health plan includes lobbying expenses reported by subsidiary permanente medical group and permanente foundation.I treated the subsidiary spending as external spending as they are not self-reported spending.
4. Wellpoint: Wellpoint + Amerigroup Group + BCBS of Georgia
5. Healthspring: includes Bravo Health
6. Highmark: includes Gateway Health plan
7. UnitedHealth" includes Subsidiary AmeriChoice Corp, Subsidiary Ingenix Inc, Subsidiary Ovations, Subsidiary Oxford Health Plans and Lobbying Expenses Reported by Subcontractor UnitedHealth Group

SQL used:

```
SELECT t1.FilingId,t2.RegistrantName, t2.ClientName, t2.Period,t2.Amount,
t2.FilingYear,
CASE WHEN  ((t2.ClientName LIKE '%UnitedHealth%' OR t2.ClientName LIKE
'United Health%' OR t2.ClientName LIKE '%GOLDEN RULE INSURANCE%'
OR t2.ClientName LIKE '%AMERICHOICE%' OR t2.ClientName LIKE '%ARIZONA
PHYSICIANS%' OR t2.ClientName LIKE '%AZ PHYSICIANS%' OR t2.ClientName
LIKE '%EVERCARE OF TEXAS%' OR t2.ClientName LIKE '%EVERCARE OF TX%'
OR t2.ClientName LIKE '%GREAT LAKES HEALTH%' OR t2.ClientName LIKE
'%HEALTH PLAN OF NEVADA%' OR t2.ClientName LIKE '%HEALTH PLAN OF
NV%' OR t2.ClientName LIKE '%OXFORD HEALTH PLANS%' OR t2.ClientName
LIKE '%PACIFICARE%' OR t2.ClientName LIKE '%SIERRA HEALTH%' OR
t2.ClientName LIKE '%UNISON HEALTH%' OR t2.ClientName LIKE '%UNITED
HEALTH%CARE%' OR t2.ClientName LIKE '%Lewin%'  OR t2.ClientName LIKE
'Ovations%') AND (t2.ClientName NOT LIKE '1199SEIU United Healthcare Workers
East' )) THEN 'UnitedHealth'
     WHEN  (t2.ClientName LIKE '%Humana%' OR t2.ClientName LIKE
'%Compbenefits%' OR t2.ClientName LIKE '%CAREPLUS%' OR t2.ClientName LIKE
'%CARITEN HEALTH PLAN%' OR t2.ClientName LIKE '%CARITEN INSURANCE
COMPANY%' OR t2.ClientName LIKE '%OSF HEALTHPLANS%' )  THEN 'Humana'
     WHEN  (t2.ClientName LIKE '%Kaiser%' AND t2.ClientName LIKE '%Health%' )
OR ( t2.ClientName LIKE '%KAISER FOUNDATION%') OR (t2.ClientName LIKE
'%Permanente%')  THEN 'Kaiser'
     WHEN  (t2.ClientName LIKE '%Wellpoint%' OR  t2.ClientName LIKE
'%Amerigroup%' OR t2.ClientName LIKE '%ANTHEM%' OR t2.ClientName LIKE
'%BC LIFE%' OR t2.ClientName LIKE '%BLUE CROSS BLUE SHIELD OF G%' OR
t2.ClientName LIKE '%BLUE CROSS OF CA%'OR t2.ClientName LIKE
'COMMUNITY INSURANCE%' OR t2.ClientName LIKE '%COMPCARE HEALTH%'
OR t2.ClientName LIKE '%EMPIRE HEALTHCHOICE%' OR t2.ClientName LIKE
'%HEALTHY ALLIANCE LIFE%' OR t2.ClientName LIKE '%HMO CO%' OR
t2.ClientName LIKE '%HMO MISSOURI%' OR t2.ClientName LIKE '%HMO MO%'
OR t2.ClientName LIKE '%ROCKY MOUNTAIN HOSPITAL%' OR t2.ClientName
LIKE '%UNICARE LIFE%')  THEN 'Wellpoint'
     WHEN  (t2.ClientName LIKE '%Aetna%')  THEN 'Aetna'
     WHEN  (t2.ClientName LIKE '%HealthSpring%' OR t2.ClientName LIKE
'%BRAVO HEALTH%' OR t2.ClientName LIKE '%HEALTH SPRING%')  THEN
'HealthSpring'
     WHEN  (t2.ClientName LIKE '%Highmark%' OR t2.ClientName LIKE
'%GATEWAY HEALTH%' OR t2.ClientName LIKE '%HM HEALTH%' OR
t2.ClientName LIKE '%KEYSTONE HEALTH PLAN WEST%')  THEN 'Highmark'
     WHEN  (t2.ClientName LIKE '%BLUE CROSS BLUE SHIELD OF MICHIGAN%'
OR t2.ClientName LIKE '%BLUE CARE NETWORK OF MICHIGAN%')  THEN
'BLUE CROSS BLUE SHIELD OF MICHIGAN'
     WHEN  (t2.ClientName LIKE '%COVENTRY HEALTH%' OR t2.ClientName LIKE
'%ALTIUS%' OR t2.ClientName LIKE '%CAMBRIDGE LIFE%' OR t2.ClientName
LIKE '%FIRST HEALTH LIFE%' OR t2.ClientName LIKE '%GROUP HEALTH
PLAN%' OR t2.ClientName LIKE '%HEALTHAMERICA%' OR t2.ClientName LIKE
```

```
'%HEALTHASSURANCE%' OR t2.ClientName LIKE '%MERCY HEALTH PLANS%'
OR t2.ClientName LIKE '%PERSONAL CARE INSURANCE OF ILLINOIS%' OR
t2.ClientName LIKE '%PERSONAL CARE INSURANCE OF IL%' OR t2.ClientName
LIKE '%SUMMIT HEALTH PLAN%' OR t2.ClientName LIKE '%VISTA
HEALTHPLAN%')  THEN 'COVENTRY'
     WHEN  (t2.ClientName LIKE '%Health Net%Inc%' OR t2.ClientName LIKE
'%HEALTH%NET%Federal%' OR  t2.ClientName LIKE '%HEALTH NET' OR
t2.ClientName LIKE '%HEALTH NET OF%') AND (t2.ClientName NOT LIKE
'%NETWORK%' )  THEN 'HEALTH NET'
     ELSE 'ERROR'
  END AS Clean_ClientNamename

FROM [LobbyingRaw].[SOPR].[vLD2Matcherv3] AS t1
JOIN [LobbyingRaw].[SOPR].[LD2] AS t2
ON t1.FilingId = t2.FilingId
WHERE
(t2.ClientName LIKE '%UnitedHealth%' OR t2.ClientName LIKE 'United Health%' OR
t2.ClientName LIKE '%GOLDEN RULE INSURANCE%' OR t2.ClientName LIKE
'%AMERICHOICE%' OR t2.ClientName LIKE '%ARIZONA PHYSICIANS%' OR
t2.ClientName LIKE '%AZ PHYSICIANS%' OR t2.ClientName LIKE '%EVERCARE
OF TEXAS%' OR t2.ClientName LIKE '%EVERCARE OF TX%' OR t2.ClientName
LIKE '%GREAT LAKES HEALTH%' OR t2.ClientName LIKE '%HEALTH PLAN OF
NEVADA%' OR t2.ClientName LIKE '%HEALTH PLAN OF NV%' OR t2.ClientName
LIKE '%OXFORD HEALTH PLANS%' OR t2.ClientName LIKE '%PACIFICARE%'
OR t2.ClientName LIKE '%SIERRA HEALTH%' OR t2.ClientName LIKE '%UNISON
HEALTH%' OR t2.ClientName LIKE '%UNITED HEALTH%CARE%' OR
t2.ClientName LIKE '%Lewin%' OR t2.ClientName LIKE 'Ovations%') AND
(t2.ClientName NOT LIKE '1199SEIU United Healthcare Workers East' ) OR
(t2.ClientName LIKE '%Humana%' OR t2.ClientName LIKE '%Compbenefits%' OR
t2.ClientName LIKE '%CAREPLUS%' OR t2.ClientName LIKE '%CARITEN HEALTH
PLAN%' OR t2.ClientName LIKE '%CARITEN INSURANCE COMPANY%' OR
t2.ClientName LIKE '%OSF HEALTHPLANS%') OR
(t2.ClientName LIKE '%Kaiser%' AND t2.ClientName LIKE '%Health%' ) OR (
t2.ClientName LIKE '%KAISER FOUNDATION%') OR (t2.ClientName LIKE
'%Permanente%') OR
(t2.ClientName LIKE '%Wellpoint%' OR  t2.ClientName LIKE '%Amerigroup%' OR
t2.ClientName LIKE '%ANTHEM%' OR t2.ClientName LIKE '%BC LIFE%' OR
t2.ClientName LIKE '%BLUE CROSS BLUE SHIELD OF G%' OR t2.ClientName
LIKE '%BLUE CROSS OF CA%'OR t2.ClientName LIKE 'COMMUNITY
INSURANCE%' OR t2.ClientName LIKE '%COMPCARE HEALTH%' OR
t2.ClientName LIKE '%EMPIRE HEALTHCHOICE%' OR t2.ClientName LIKE
'%HEALTHY ALLIANCE LIFE%' OR t2.ClientName LIKE '%HMO CO%' OR
t2.ClientName LIKE '%HMO MISSOURI%' OR t2.ClientName LIKE '%HMO MO%'
OR t2.ClientName LIKE '%ROCKY MOUNTAIN HOSPITAL%' OR t2.ClientName
LIKE '%UNICARE LIFE%') OR
(t2.ClientName LIKE '%Aetna%' ) OR
```

```
(t2.ClientName LIKE '%HealthSpring%' OR t2.ClientName LIKE '%BRAVO
HEALTH%' OR t2.ClientName LIKE '%HEALTH SPRING%') OR
(t2.ClientName LIKE '%Highmark%' OR t2.ClientName LIKE '%GATEWAY
HEALTH%' OR t2.ClientName LIKE '%HM HEALTH%' OR t2.ClientName LIKE
'%KEYSTONE HEALTH PLAN WEST%') OR
(t2.ClientName LIKE '%BLUE CROSS BLUE SHIELD OF MICHIGAN%' OR
t2.ClientName LIKE '%BLUE CARE NETWORK OF MICHIGAN%') OR
(t2.ClientName LIKE '%COVENTRY HEALTH%' OR t2.ClientName LIKE
'%ALTIUS%' OR t2.ClientName LIKE '%CAMBRIDGE LIFE%' OR t2.ClientName
LIKE '%FIRST HEALTH LIFE%' OR t2.ClientName LIKE '%GROUP HEALTH
PLAN%' OR t2.ClientName LIKE '%HEALTHAMERICA%' OR t2.ClientName LIKE
'%HEALTHASSURANCE%' OR t2.ClientName LIKE '%MERCY HEALTH PLANS%'
OR t2.ClientName LIKE '%PERSONAL CARE INSURANCE OF ILLINOIS%' OR
t2.ClientName LIKE '%PERSONAL CARE INSURANCE OF IL%' OR t2.ClientName
LIKE '%SUMMIT HEALTH PLAN%' OR t2.ClientName LIKE '%VISTA
HEALTHPLAN%') OR
(t2.ClientName LIKE '%Health Net%Inc%' OR t2.ClientName LIKE
'%HEALTH%NET%Federal%' OR  t2.ClientName LIKE '%HEALTH NET' OR
t2.ClientName LIKE '%HEALTH NET OF%') AND (t2.ClientName NOT LIKE
'%NETWORK%' )

GROUP BY t1.FilingId,t2.RegistrantName,t2.ClientName, t2.Period,t2.Amount,
t2.FilingYear
ORDER BY t2.FilingYear, t2.RegistrantName, t2.Period
```

Finding number of lobbyists involved

Steps:

1. SQL – Copy the result to Excel [0220-lobbyist]
2. [Remove Duplicate] – lobbyists involved in lobbying for a company per year
   Remove Duplicate (select Year, Clean_Clientname, Lobbyist_ID) 1629 duplications are deleted, 4 different quarters still only count one
3. [result1]Pivot Table – show how many lobbyist hired each year for each company
4. [result2]Pivot Table- which lobbying firms involved in multiple insurance companies
   - How many lobbyist are appointed by that lobbying firms to help the insurance company each year.
5. [result3] Pivot Table – which lobbyist involved the most

Remarks:
Limitation: Year 2013 is not complete.

```
SELECT DISTINCT

t1.year, t1.client, t1.registrant, t2.lobbyist_id, t2.lobbyist,t1.amount,
```

```
    CASE WHEN  ((t1.client LIKE '%UnitedHealth%' OR t1.client LIKE 'United
Health%' OR t1.client LIKE '%GOLDEN RULE INSURANCE%'
OR t1.client LIKE '%AMERICHOICE%' OR t1.client LIKE '%ARIZONA
PHYSICIANS%' OR t1.client LIKE '%AZ PHYSICIANS%' OR t1.client LIKE
'%EVERCARE OF TEXAS%' OR t1.client LIKE '%EVERCARE OF TX%' OR t1.client
LIKE '%GREAT LAKES HEALTH%' OR t1.client LIKE '%HEALTH PLAN OF
NEVADA%' OR t1.client LIKE '%HEALTH PLAN OF NV%' OR t1.client LIKE
'%OXFORD HEALTH PLANS%' OR t1.client LIKE '%PACIFICARE%' OR t1.client
LIKE '%SIERRA HEALTH%' OR t1.client LIKE '%UNISON HEALTH%' OR t1.client
LIKE '%UNITED HEALTH%CARE%' OR t1.client LIKE '%Lewin%'  OR t1.client
LIKE 'Ovations%') AND (t1.client NOT LIKE '1199SEIU United Healthcare Workers
East' )) THEN 'UnitedHealth'
    WHEN  (t1.client LIKE '%Humana%' OR t1.client LIKE '%Compbenefits%' OR
t1.client LIKE '%CAREPLUS%' OR t1.client LIKE '%CARITEN HEALTH PLAN%'
OR t1.client LIKE '%CARITEN INSURANCE COMPANY%' OR t1.client LIKE
'%OSF HEALTHPLANS%' )  THEN 'Humana'
    WHEN  (t1.client LIKE '%Kaiser%' AND t1.client LIKE '%Health%' ) OR ( t1.client
LIKE '%KAISER FOUNDATION%') OR (t1.client LIKE '%Permanente%')  THEN
'Kaiser'
    WHEN  (t1.client LIKE '%Wellpoint%' OR  t1.client LIKE '%Amerigroup%' OR
t1.client LIKE '%ANTHEM%' OR t1.client LIKE '%BC LIFE%' OR t1.client LIKE
'%BLUE CROSS BLUE SHIELD OF G%' OR t1.client LIKE '%BLUE CROSS OF
CA%'OR t1.client LIKE 'COMMUNITY INSURANCE%' OR t1.client LIKE
'%COMPCARE HEALTH%' OR t1.client LIKE '%EMPIRE HEALTHCHOICE%' OR
t1.client LIKE '%HEALTHY ALLIANCE LIFE%' OR t1.client LIKE '%HMO CO%' OR
t1.client LIKE '%HMO MISSOURI%' OR t1.client LIKE '%HMO MO%' OR t1.client
LIKE '%ROCKY MOUNTAIN HOSPITAL%' OR t1.client LIKE '%UNICARE LIFE%')
THEN 'Wellpoint'
    WHEN  (t1.client LIKE '%Aetna%')  THEN 'Aetna'
    WHEN  (t1.client LIKE '%HealthSpring%' OR t1.client LIKE '%BRAVO
HEALTH%' OR t1.client LIKE '%HEALTH SPRING%')  THEN 'HealthSpring'
    WHEN  (t1.client LIKE '%Highmark%' OR t1.client LIKE '%GATEWAY
HEALTH%' OR t1.client LIKE '%HM HEALTH%' OR t1.client LIKE '%KEYSTONE
HEALTH PLAN WEST%')  THEN 'Highmark'
    WHEN  (t1.client LIKE '%BLUE CROSS BLUE SHIELD OF MICHIGAN%' OR
t1.client LIKE '%BLUE CARE NETWORK OF MICHIGAN%')  THEN 'BLUE CROSS
BLUE SHIELD OF MICHIGAN'
    WHEN  (t1.client LIKE '%COVENTRY HEALTH%' OR t1.client LIKE
'%ALTIUS%' OR t1.client LIKE '%CAMBRIDGE LIFE%' OR t1.client LIKE '%FIRST
HEALTH LIFE%' OR t1.client LIKE '%GROUP HEALTH PLAN%' OR t1.client LIKE
'%HEALTHAMERICA%' OR t1.client LIKE '%HEALTHASSURANCE%' OR t1.client
LIKE '%MERCY HEALTH PLANS%' OR t1.client LIKE '%PERSONAL CARE
INSURANCE OF ILLINOIS%' OR t1.client LIKE '%PERSONAL CARE INSURANCE
OF IL%' OR t1.client LIKE '%SUMMIT HEALTH PLAN%' OR t1.client LIKE
'%VISTA HEALTHPLAN%')  THEN 'COVENTRY'
```

```
    WHEN  (t1.client LIKE '%Health Net%Inc%' OR t1.client LIKE
'%HEALTH%NET%Federal%' OR  t1.client LIKE '%HEALTH NET' OR t1.client LIKE
'%HEALTH NET OF%') AND (t1.client NOT LIKE '%NETWORK%' )  THEN
'HEALTH NET'

    ELSE 'ERROR'
  END AS Clean_Clientname


FROM CRP.Lobbying t1
JOIN CRP.Lobbyists t2
ON t1.uniqid = t2.uniqID
WHERE

((t1.client LIKE '%UnitedHealth%' OR t1.client LIKE 'United Health%' OR t1.client
LIKE '%GOLDEN RULE INSURANCE%' OR t1.client LIKE '%AMERICHOICE%'
OR t1.client LIKE '%ARIZONA PHYSICIANS%' OR t1.client LIKE '%AZ
PHYSICIANS%' OR t1.client LIKE '%EVERCARE OF TEXAS%' OR t1.client LIKE
'%EVERCARE OF TX%' OR t1.client LIKE '%GREAT LAKES HEALTH%' OR
t1.client LIKE '%HEALTH PLAN OF NEVADA%' OR t1.client LIKE '%HEALTH
PLAN OF NV%' OR t1.client LIKE '%OXFORD HEALTH PLANS%' OR t1.client
LIKE '%PACIFICARE%' OR t1.client LIKE '%SIERRA HEALTH%' OR t1.client LIKE
'%UNISON HEALTH%' OR t1.client LIKE '%UNITED HEALTH%CARE%' OR
t1.client LIKE '%Lewin%' OR t1.client LIKE 'Ovations%') AND (t1.client NOT LIKE
'1199SEIU United Healthcare Workers East' ) OR
(t1.client LIKE '%Humana%' OR t1.client LIKE '%Compbenefits%' OR t1.client LIKE
'%CAREPLUS%' OR t1.client LIKE '%CARITEN HEALTH PLAN%' OR t1.client
LIKE '%CARITEN INSURANCE COMPANY%' OR t1.client LIKE '%OSF
HEALTHPLANS%') OR
(t1.client LIKE '%Kaiser%' AND t1.client LIKE '%Health%' ) OR ( t1.client LIKE
'%KAISER FOUNDATION%') OR (t1.client LIKE '%Permanente%') OR
(t1.client LIKE '%Wellpoint%' OR  t1.client LIKE '%Amerigroup%' OR t1.client LIKE
'%ANTHEM%' OR t1.client LIKE '%BC LIFE%' OR t1.client LIKE '%BLUE CROSS
BLUE SHIELD OF G%' OR t1.client LIKE '%BLUE CROSS OF CA%'OR t1.client
LIKE 'COMMUNITY INSURANCE%' OR t1.client LIKE '%COMPCARE HEALTH%'
OR t1.client LIKE '%EMPIRE HEALTHCHOICE%' OR t1.client LIKE '%HEALTHY
ALLIANCE LIFE%' OR t1.client LIKE '%HMO CO%' OR t1.client LIKE '%HMO
MISSOURI%' OR t1.client LIKE '%HMO MO%' OR t1.client LIKE '%ROCKY
MOUNTAIN HOSPITAL%' OR t1.client LIKE '%UNICARE LIFE%') OR
(t1.client LIKE '%Aetna%' ) OR
(t1.client LIKE '%HealthSpring%' OR t1.client LIKE '%BRAVO HEALTH%' OR
t1.client LIKE '%HEALTH SPRING%') OR
(t1.client LIKE '%Highmark%' OR t1.client LIKE '%GATEWAY HEALTH%' OR
t1.client LIKE '%HM HEALTH%' OR t1.client LIKE '%KEYSTONE HEALTH PLAN
WEST%') OR
```

```
(t1.client LIKE '%BLUE CROSS BLUE SHIELD OF MICHIGAN%' OR t1.client LIKE
'%BLUE CARE NETWORK OF MICHIGAN%') OR
(t1.client LIKE '%COVENTRY HEALTH%' OR t1.client LIKE '%ALTIUS%' OR
t1.client LIKE '%CAMBRIDGE LIFE%' OR t1.client LIKE '%FIRST HEALTH LIFE%'
OR t1.client LIKE '%GROUP HEALTH PLAN%' OR t1.client LIKE
'%HEALTHAMERICA%' OR t1.client LIKE '%HEALTHASSURANCE%' OR t1.client
LIKE '%MERCY HEALTH PLANS%' OR t1.client LIKE '%PERSONAL CARE
INSURANCE OF ILLINOIS%' OR t1.client LIKE '%PERSONAL CARE INSURANCE
OF IL%' OR t1.client LIKE '%SUMMIT HEALTH PLAN%' OR t1.client LIKE
'%VISTA HEALTHPLAN%') OR
(t1.client LIKE '%Health Net%Inc%' OR t1.client LIKE '%HEALTH%NET%Federal%'
OR  t1.client LIKE '%HEALTH NET' OR t1.client LIKE '%HEALTH NET OF%') AND
(t1.client NOT LIKE '%NETWORK%' ))


AND [use] = 'y' AND t1.amount <> '0.0'

GROUP BY t1.year, t1.client, t1.registrant, t2.lobbyist_id, t2.lobbyist,t1.amount
ORDER BY Clean_Clientname DESC, t1.year, t1.registrant, t2.lobbyist_id

Indivs to Cands/PACs:
SELECT *
FROM CRP.Indivs
WHERE Cycle = 2014
AND RealCode = 'H3700'


PACs to Cands:
SELECT t2.FirstLastP,
t1.*
FROM CRP.PACs t1
JOIN CRP.Cands t2
ON t1.FECCandID = t2.FECCandID
AND t1.Cycle = t2.Cycle
WHERE RealCode = 'H3700' AND t1.Cycle = 2014

PACs to PACs:
SELECT t2.PACShort,
t1.*
FROM CRP.PacOther t1
JOIN CRP.Cmtes t2
ON t1.Cycle = t2.Cycle
AND t1.RecipID=t2.RecipID
WHERE RealCode = 'H3700' AND t1.Cycle = 2014 AND RecipPrimcode NOT LIKE
'J2%'
```

Industry Codes:

```
SELECT [Catcode]
      ,[Catname]
      ,[Catorder]
      ,[Industry]
      ,[Sector]
      ,[Sector Long]
  FROM [FEC].[LookupTable].[CRPCategories]
WHERE Catcode LIKE 'J%'
```

Chapter Five: Research Component

Introduction

A new visual-based communication era is upon us. Thanks to the open government data movement and the advancement in computing for processing and presenting data, visualization is becoming the "next mass communication medium" and is emerging as a critical tool for helping readers navigate the abundance of information (Viegas & Wattenberg, 2011).

News stories are increasingly accompanied by informational graphics and data visualizations, and these are becoming more sophisticated than ever (Giardina & Medina, 2013; Utts & Pasternak, 2000). Illustrations, charts, databases, maps and other interactive elements are built to encourage interaction and provide readers multiple layers of news detail. Visualization has the potential to reveal unnoticed information, especially in large datasets; to give answers faster; to help journalists investigate cause-effect relationships; and to help audiences with limited education and short attention spans understand data more easily. News sites such as usatoday.com, elmundo.es, or bbc.co.uk use information graphics and data visualizations to explain complex information clearly and intelligibly.

However, even though information visualization has a long history in other disciplines, such as engineering and statistical modeling (Chen, 2004), there is little agreement on the best way to integrate visualizations into the news production process (Weber & Rall, 2012). Newsrooms are experimenting and finding the best processes for producing infographics and visualizations.

This study will explore and compare the use of infographics and visualizations in hard and soft news, examining the digital portal of The New York Times. My research

will examine what built-in features editors and reporters tend to use to foster conversations and audience engagement. Content analysis, case studies and interviews would be conducted to address the following research questions:

RQ1: Between hard and soft news at the New York Times' websites, what, if any, are the differences in terms of number and nature of infographics and visualizations used?

RQ2: What, if anything, motivates editors and reporters to come up with different features in infographics and visualizations in hard versus soft news?

RQ3: What kind of mechanisms do editors and reporters tend to use in infographics and visualizations to foster conversations and audience engagement?

Theoretical Framework

To understand the use of infographics and visualizations on the New York Times' website, visual rhetoric theory and narrative visualizations categories framework derived by Segel and Heer (2010) are adopted.

Rhetoric theory.

Rhetoric is an interpretive theory that frames a message as an interested party's attempt to influence an audience. The sender's intention manifests in the argument, the evidence, the order of argumentation, and the style of delivery (Corbett, 1965). The sender crafts the message in anticipation of the audience's probable response, using shared knowledge of various vocabularies and conventions, as well as common experiences. Receivers use the same body of cultural knowledge to read the message and the sender's argument, and to formulate a response.

Visual rhetoric in a broader sense.

Visual rhetoric is one of the new disciplines taken into account by semantic multimedia and visual communication researchers. Rhetoric was first largely employed in the domain of speech and writing programs. However, in 1965, Gui Bonsiepe argued that all of the decisions made in the layout and presentation of information are rhetorical and that theories of rhetoric should include the interplay of words and images (Veltsos, 2009). Twenty years later Robin Kinross (1985) analyzed the typeface and color used in a railroad timetable and asserted that designers make choices about when and how to use conventions based upon audience and context. This reaffirmed Bonsiepe's argument that all design is rhetorical.

The term visual rhetoric refers to the intended meanings that are represented in the visualization via a designer's choices and then shaped by individual end-user characteristics, contextual factors involving societal or cultural codes, and the end-user's interaction. With a targeted audience in mind, the designer can select elements that tap into or coincide with the audience's beliefs and cultural expectations and make visual arguments more persuasive. Cultural knowledge thus provides the basis for normative interaction and persuasion (Scott, 1990).

Foss (2004) summarized visual rhetoric as "both a visual object or artifact and a perspective on the study of visual data." It is a product individuals create as they use visual symbols for the purpose of communication and a perspective that scholars apply to symbolic processes by which visual artifacts perform communication. As an area of focus, visual rhetoric has three areas that scholars typically study. They are:

(1) Nature. It focuses on the components (space, color, etc.) of the image.

(2) Function. It concerns the effect of an image on the audience, not necessarily the creator's purpose.

(3) Evaluation. It tries to assess the effectiveness of the image. However, because function isn't singular in visual rhetoric, assessing the effectiveness of an image to meet its purpose is difficult.

Rhetoric associated with persuasion.

Rhetoric has come to be associated with persuasion as a result of the implicit motivation of the sender to gain other adherents to a preconceived view or conclusion (Bogost, 2007). McGuire (2000), the pioneer of persuasion theory, suggested that a better understanding of visual rhetorical figures is most likely to provide new insights into persuasion processes. Ann C. Tyler (1992) summarized the three purposes of design: to persuade the audience to act, to educate the audience and to provide the audience with an experience. The goal of visual communication is to persuade an audience to adopt a new belief. The visual rhetoric pertains to the visual elements of the document and affects the reader's initial impression of the document (Brumberger, 2001). It can also affect the tone, author's voice and credibility of a document (Kostelnick and Roberts, 1998). Strachan and Kendall's (2004) analysis of political candidates' convention films was an example of the evaluation of visual rhetoric. The study pointed out that politicians' emotional appeals via visual cues encourage "unquestioned acceptance" of politicians and policies.

Studies of visual rhetoric in infographics and visualizations.

Although the outlines and depths of visual rhetoric of data visualization are still being explored, a few studies recognize that all information graphics are visually inscribed rhetoric. Kress and Van Leeuwen (2006) reminded that the study of visual rhetoric is different from that of visual or graphic design. Rather than purely aesthetic consideration, it emphasizes images as sensory expressions of cultural meaning.

Allen (1996) summarized that visual rhetoric, which helps us understand how visuals communicate, focuses on seven purposes: teaching visual information, heightening awareness of visually informative features, evaluating the artistry of visual features, processing visuals and text, obeying graphic grammar, integrating visual and verbal languages aesthetically and using efficiency and clarity to inform visually.

Kimball (2006) examined Charles Booth's maps of London poverty (1889-1902) and analyzed the cultural basis of ideas of transparency and clarity in information graphics. He argued that Booth's maps derive their rhetorical power from contemporary visual culture as much as from their scientific authority. The visual rhetoric of the maps depended upon an ironic inversion of visual culture to make poverty seem a problem that could be addressed, rather than an insurmountable crisis. Information graphics are inherently rhetorical and have the power to influence social policy.

Visualizations can be appealing, enjoyable and understandable. Data visualizations grant huge rewards for society regarding enhancement of perception, persuasion and interpretation. Visual rhetoric is used as an analytical framework in this study to help understand how design techniques prioritize particular interpretations in

visualization storytelling and their influence on end-user interpretation.

Segel and Heer (2010) narrative visualizations categories.

In response to the growing number of online visualizations designed to convey a story, Segel and Heer (2010) identified clear categories to distinguish different forms of data visualization. Although the samples they used might not be exhaustive, their study establishes a very useful framework to access the visual narrative, structuring, and storytelling aspects of data visualization. The framework is very useful to my study. They presented three categories to distinguish narrative visualizations: (1) Balance of narrative, (2) visual narrative devices and (3) narrative structure devices.

(1) **Balance of narrative** refers to the spectrum of author-driven, reader-driven and hybrid approaches that balance a narrative together with interaction and messaging.

- An author-driven approach refers to a linear path through the visualization, which relies heavily on messaging and has no interactivity. Examples include watching a film, educational videos or training materials.
- A reader-driven approach, on the other hand, does not prescribe an order of viewing and usually involves a lot of interactivity. Visual analysis tools commonly have this function for tasks such as data diagnostics, patterns discovery and hypothesis formation.
- A hybrid-mix, as the name implies, falls in between the balance of the above approaches. Segel and Heer (2010) say this category gains increasing popularity with visualizations.

(2) **Visual narrative devices** are the visual mechanisms that assist and facilitate the narrative. The authors divide these devices into visual structuring mechanisms,

progress bars, consistent visual platforms and highlighting.

- Visual structuring mechanisms communicate the overall structure of the narrative to the viewer and allow him to identify his position within the larger organization of the visualization. Some use visual structuring mechanisms to orient the viewer early on with an overall view or consistent visual platform and to allow the viewer to track his progress through the visualization.
- Progress bars or timeline bars indicate the length of visualization. They provide users with a mechanism to navigate and allow them to skip around the visualization to parts deemed more interesting.
- Consistent visual platform refers to when a visualization or a slideshow progresses and how only the content within each panel changes while leaving the general layout of the visual elements the same. Each new slide changes the text, while animated transitions propel the story forward.
- Highlighting refers to visual mechanisms that draw users' attention to specific areas on the screen by augmenting it with distinctive features such as color, motion, framing, size, and audio.

(3) **Narrative structure devices** are the non-visual mechanisms that assist and facilitate the narrative. The authors divide these tactics into three sections: (i) ordering, (ii) interactivity, and (iii) messaging.

- Ordering refers to the different ways of arranging the path viewers take through a visualization, where sometimes the path is prescribed by the authors (linear), sometimes there is no suggested path at all (random access), and other times the user selects among multiple alternatives (user-directed).

- Interactivity categorizes the different ways a user can manipulate the visualization. For example, by using navigation buttons, filtering, selecting, searching or hover highlighting.
- Messaging denotes the ways in which a visualization communicates with observations and commentary to the viewer (achieved by labels, captions, headlines and annotations). It helps clearly communicate through the interaction of text on one side with annotations and graphic elements on the other side by providing related but different information.

Segel and Heer's contribution of abstract structures and genres provides a general framework that opens the discussion of narrative visualization to a wider range of examples. Adapting the framework in this research would allow comparisons between visualizations produced in newsrooms.

Literature Review

Definitions of hard news and soft news.

Media scholars have regularly made distinctions about the production processes of various types of news content, their effects on the resulting products, and the subsequent social and political consequences. The distinction between hard and soft news is one of the foremost examples of this analytical strategy and has been widely employed by communication scholars (e.g., Patterson 2000; Scott and Gobetz 1992).

Hard news is "the coverage of breaking events involving top leaders, major issues, or significant disruptions in the routines of daily life, such as an earthquake or airline disaster" (Patterson, 2000, p.3). On the contrary, soft news aims more to

entertain. It is usually less political in content, but more human interest stories and special news. It is typically "more sensational, more personality-centered, less time-bound…and more incident-based than other news" (Patterson, 2000, p.4). Shoemaker and Cohen (2006) defined hard and soft news according to topicality or timeliness. Hard news items are urgent occurrences that have to be reported right away because they become obsolete very quickly. In contrast, soft news items are usually based on nonscheduled events.

Lehman-Wilzog and Seletzky (2008) determined that hard news is often defined as news content that covers political, social or economic topics, demands immediate reporting due to its importance in order to stay relevant and has actual ramifications over a wide spectrum of society. Soft news, on the other hand, is defined as having little or no intrinsic social or personal importance, so that it can be reported on at any time, as well as news that rather than being relevant to the lives of those receiving the news instead affects only a tiny fraction of the viewing audience.

Conceptual definitions of hard news and soft news in this study.

For the purposes of my study, stories that fall under the umbrella of hard news often deal with topics like business, politics and international news. Hard news stories are accounts of events that have just happened or are about to happen. For example, crimes, fires, meetings, court testimony, speeches, protest rallies, acts of war, traffic accidents and elections are all typical topics of hard news stories. They emphasize facts but not opinion. Soft news would be more likely to be reports about celebrities, human interest, sports and other entertainment-centered stories.

However, for news about crimes, fires, accidents, I would also keep an eye on the stories' topicality or timeliness. Hard news items consists of urgent occurrences that have

to be reported right away because they become obsolete very quickly. Breaking news stories that develop overnight or on the same day are hard news. Episodic reporting of crimes and fires, for example, would fall under the definition of soft news.

Definitions of infographics and data visualizations.

Card et al. (1999) defined infographics as "The use of computer-supported, interactive visual representations of data to amplify cognition;" Newsom & Haynes (2004) defined infographics as graphic visual representations of information, data or knowledge intended to clarify and integrate difficult information quickly and clearly. They are usually used to summarize data. Mike Scaife (1996) defined visualization as a mechanism by which humans perceive, interpret, use and communicate visual information. It focuses more on design to allow users to explore datasets for their own purposes. That is, where infographics tell stories designed by communicators, information visualization helps readers discover stories by themselves.

Card et al. (1999) illustrate that the field of data visualization covers the properties of visual perception to resolve logical problems. It investigates how a visual display of information — by automatically assembling thousands of data objects into pictures, revealing hidden patterns — can serve as a new method for amplifying cognition and generate new knowledge and insight about the world. Card et al. conclude that diagrams can help in six ways: increasing the memory and processing resources available to users; reducing the search for information; using visual representations to enhance the detection of patterns; enabling perceptual inference operations; using perceptual attention mechanisms for monitoring; and encoding information in a medium that can be manipulated. Steve Pasternack and Sandra Utts (1990) reported that readers

use information graphics strategically to seek out information. J. Votika Ramaprasad (1991) found that the reader's understanding from information graphics was limited and that the information presented in the graphic sometimes misled the reader. However, in a systematic program of research, Jeffrey Griffin and Robert Stevenson (1994) demonstrated how various information graphic tools such as locator maps, explanatory graphics and graphs facilitate learning.

Several studies have stressed that infographics allow newspapers to optimize the understanding processes thanks to compressed quantity of information and a greater precision, anchored in image and text. Recent infographics researchers have emphasized the problem of recognizing infographics (Huang & Tan, 2007), arguing that understanding infographics is a discourse-level problem while others have explored the intersection between infographics and games (Diakopoulos et al., 2011). Meyer (2004) found that interactive infographics could help newspapers add value and enhance the quality of their informational product and therefore have a pivotal function on the influence and credibility of a newspaper company (Meyer, 2004).

Previous research on visualizations' storytelling power.

Infographics were largely used to support the work of journalism back in the 1980s. They used to accommodate newsletters, newspapers, magazines, and reports. Now storytelling has become a new focus in visualization research and practice. More recently, Hullman and Diakopoulos (2011) presented a rhetorical framework for narrative visualizations that includes design choices about the dataset, visualization and interactivity as well as "extra-representational" factors on how a visualization may be interpreted. They identified a number of approaches to communicate authority,

completeness of data, etc., and showed how these cues can be used to prioritize particular interpretations.

Segel and Heer (2010) studied how specific data visualizations are produced and integrated into online news by identifying narrative design differences and by making recommendations on best practices. One of the most interesting structures is what they called the Martini glass, which starts with a broad introduction, then narrows to make a particular point, and then opens up interaction and exploration to the viewer.

Tactinsky and Meyer (1999) found that the presenters created different displays when using data they viewed favorably as opposed to those they disliked. Their findings showed that people tend to create more complicated graphics when they want to persuade or impress the audience.

Little work has been done on understanding and comparing how infographics and visualizations are used differently based on the nature of the news being reported. Scientific research examining the effect of infographics and visualizations on improving audiences' awareness about social issues is rare. George-Palilonis (2006) gave more practical guidelines and described the professional skills for designers and journalists. Cairo (2005) provided valuable insights into the journalistic process and the common roots of visualization research, perceptual background, and journalistic mission. Dörk et al (2010) has explored the politics of visualization and took engaging visualizations as a starting point and outlined a critical approach that promotes disclosure, plurality, contingency, and empowerment.

In this research, I am trying to learn about the differences in the use and production of infographics and data visualizations in hard and soft news. I focus on three

aims: (1) gaining insight into the use of infographics and visualizations in different news stories; (2) exploring and investigating the distribution and differences in infographics and visualizations in different news stories; (3) understanding how infographics and visualizations could help audience engagement.

Methodology

This study employs a mixed-research approach designed to collect both qualitative and quantitative data. The approach includes the following three parts:

- A content analysis of all the infographics and visualizations published on The New York Times website during 2012. The method is chosen to help understand how infographics and visualizations are presented in hard and soft news and, specifically, to address research question 1.
- An in-depth study of two cases of infographics or visualizations, one on hard news and one on soft news.
- Interviews with three editors at The New York Times who regularly display infographics and visualizations on the newspaper's website. This method is employed to provide deeper insight into what, if any, differences are present in the infographics and visualizations used between hard and soft news and what strategies are adopted by editors and reporters to foster citizens' conversations and audience engagement. These interviews are intended to address research questions 2 and 3 and serve as an extension to my content analysis and case study. The Institutional Review Board has approved the interviewing protocol.

Rationale for choosing The New York Times.

The New York Times has invested heavily in experimenting with design practices related to infographics production and dissemination. As of 2012, The New York Times infographics department employed 25 highly specialized journalists to research and create diagrams, maps and charts for the newspaper and the website. The New York Times' infographics have been researched in many scholarly studies (Segel and Heer, 2010). Notable examples of their work include a 3-D video explaining how New York Yankees pitcher Mariano Rivera dominates hitters, before-and-after-satellite maps of the earthquake and tsunami in Japan in March 2011, an interactive budget puzzle and a customizable electoral map. Therefore, The New York Times is a logical choice to study in this research.

Rationale for choosing year 2012.

2012 was a year with several big national and international events. It was a presidential election year in the United States. A lot of graphics were designed to tell stories about the elections. A data visualization called "512 Paths to the White House" by The New York Times is an example. It calculates the likelihood of Romney and Obama winning, based on which direction each swing states ends up going. 2012 was also the year of the London Olympic games. The New York Times produced many graphics to illustrate the record-breaking moments and the intense competition among athletes. For example, an infographic video called "Men's 100-Meter Dash" creates imaginary events in which all the medalists from every Olympic games since 1896 compete together. A lot of graphics and visualizations were produced in this year.

Content Analysis.

The use of content analysis as a quantitative approach in studying newspapers is very popular. In an examination of articles published in *Journalism & Mass Communication Quarterly* from 1971-95, Riffe and Freitag (1997) found that the primary focus of articles using content analysis was on news/editorial content. Krippendorff (1980) defined content analysis as ‘‘a research technique for making replicable and valid inferences from data to their context.’’ According to Kolbe and Burnett (1991), content analysis possesses the following benefits:

- It is unobtrusive, which is particularly valuable in situations in which other methods yield biased results.
- It is helpful in summarizing large bodies of communication messages. If one wants to know, for example, how frequently an issue was discussed in the newspaper in the past year, content analysis would be an appropriate method.
- It enables people to systematically study historical moments and trends over time. For example, it is not possible to interview George Washington, but one could conduct a content analysis of his writings.

The goal of this study is to analyze the common practice of employing infographics and visualization in newspapers' digital (web) forms and to highlight the distribution of these visual elements in each news section. Through a content analysis of published placement on the webpage, categories, sources, interactive elements and genres of visualization narratives, the use of infographics and visualizations in different news sections can be evaluated.

I would collect infographics and visualizations produced online by The New York

Times in 2012. Unfortunately, The New York Times only includes text in its current archives. Photos, charts, illustrations and other graphics are not included. However, The New York Times has a multimedia search engine inside its website. I would use that search engine to do a comprehensive search, including the following keywords: "interactive," "infographic," "chart," "graphic," "visualization,", "diagram," and "timeline." To make sure I get all the infographics and visualizations published on the web, I also would go to The New York Times' Twitter accounts to follow its graphics posting history. This could help to double check if the search engine misses some of the graphics. The initial estimation of the number of graphics on the website would be more than 600. Graphics that were produced by the newsrooms and for the purpose of storytelling would be collected and coded through a coding scheme.

Table 1: Coding sheet

The coding sheet would record the basic information of the graphics, their placements with either hard or soft news and Segel and Heer's (2010) narrative visualizations categories.

Month		
Day		
Title		
News type	1. Politics 2. Public Affairs 3. Education 4. Economy/Business 5. Crime/Accident/War	Hard News
	6. Health/Science/Technology 7. Sports 8. Celebrities 9. Arts/Fashion & Style 10. Episodic reporting of crimes/ accident/war	Soft News
Interactivity	1. Static 2. Motional 3. Interactive	
Forms (can be multiple inputs)	1. Table(s) 2. Map(s) (Geographical presentation of data) 3. Illustration(s) 4. Photo(s) 5. Time series 6. Flow chart(s) (Showing concepts or processes) 7. Relational/ Organizational diagram(s) 8. Graph(s) (Showing data trends or the proportion of data) 9. Video(s)/Audio(s) 10. Other(s) (e.g. Document Reader)	
Balance of Narrative	1. Author-driven 2. Reader-driven 3. Hybrid approaches	
Visual narrative devices (can be multiple inputs)	Visual Structuring	1. Establishing Shot/Splash Screen 2. Consistent Visual Platform 3. Progressive Bar/Time bar
	Highlighting	1. Close-Ups 2. Feature Distinction 3. Character Distinction 4. Motion 5. Audio 6. Zooming

Narrative structure devices (can be multiple inputs)	Ordering	1. Linear 2. Random access 3. User-directed
	Interactivity categorizes (only for interactive graphics)	1. Hover Highlighting/Details 2. Filtering/Selection/Search 3. Navigation Buttons 4. Explicit Instruction
	Messaging	1. Captions/Headlines 2. Annotations 3. Accompanying Article

A pilot-coding test would be carried out to test the reliability of the coding sheet. I would invite a peer researcher to try coding for 10 percent of The New York Times infographics and visualizations. I would compare the coding result from my results and my peer's. Intercoder reliability would be calculated with SPSS Crosstab function.

Intercoder reliability is the widely used term for the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion. It is a critical component of content analysis. It is acknowledged that "given that a goal of content analysis is to identify and record relatively objective characteristics of messages, reliability is paramount. Without the establishment of reliability, content analysis measures are useless" (Neuendorf, 2002, p. 141).

Data would be recorded in an Excel spreadsheet. This can conveniently convert to a .csv file to do more quantitative analysis in SPSS. The percentage of the distribution of infographics between hard news and soft news, the features they are using and other characteristics would be summarized.

Case Study.

Case study research excels at bringing us to an understanding of a complex issue or object and can extend experience or add strength to what is already known through previous research. Researcher Robert K. Yin (1984) defined the case study research method as an empirical inquiry that investigates a contemporary phenomenon within its real-life context. It is used when the boundaries between phenomenon and context are not clearly evident and when multiple sources of evidence are used. Case studies emphasize detailed contextual analysis of a limited number of events or conditions and their relationships.

Two case studies would be carried out as a narrative description and a more detailed analysis of infographics and visualizations between hard news and soft news in The New York Times. Selection of the case would be based on representative examples encountered during the content analysis. At this point, graphics on the U.S. presidential election and London Olympic games would be applicable choices for the case studies.

Interview.

As previously stated, qualitative study also would be conducted. In order to do so, at least three open-ended, semi-structured, in-depth interviews would be carried out. According to Newton (2010), interviews have the following benefits for scholarly research:

- They provide the opportunity to generate rich data;
- The language used by participants is considered essential in gaining insight into their perceptions and values;
- Contextual and relational aspects are seen as significant to understanding others'

perceptions;

- Data generated can be analyzed in different ways.

Interviews will help me to delve into research questions 2 and 3. It helps to provide insight into how news organizations use infographics and visualizations to tell stories and how these practices could help foster conversations and audience engagement. It serves as an extension of my content analysis and case studies.

Open-ended interviews allow participants to discuss their opinions, views and experiences fully in detail, whereas a set interview with closed-ended questions might inhibit them from expressing their full opinions and feelings. The interviews will consist of more than 10 open-ended questions, uniquely developed for the sole purpose of this study (see Table 2 for a sample interview guide). The interviews will last an estimated 30 to 45 minutes. However, these questions and times are merely a guide; the participants' responses will determine the direction and length of the interviews. The interviews will be audiotaped with permission from the participants to ascertain an accurate account of the interview, which can be replayed for analytic purposes. The interviews will be carried out over a period of two months, which allows the researcher to reflect and make adjustments as necessary.

Interviewees will be information graphic designers, graphic editors and art directors who work in editorial departments for print and online journalism or TV and regularly plan and design interactive graphics. My potential interviewees include:

- Kevin Quealy, graphics editor at The New York Times; Mizzou alumnus;
- Amanda Cox, graphics editor at The New York Times;
- Matthew Bloch, graphics editor at The New York Times;

- Aaron Pilhofer, associate managing editor for digital strategy/editor of the Interactive News desk at The New York Times.

Table 2: *Interview Guide*

These questions were developed and framed according to the subjects' experiences of terminology and process. A mixture of open and closed questions would be used. The weighting of responses is broadly proportionate to the range of participants but reflects individual concerns.

RQ2. What, if anything, motivates editors and reporters to come up with different features in infographics and visualizations in hard versus soft news?
• How do you decide if it is necessary to include infographics and visualizations in an article to tell a story? • How would you categorize graphics? What is the most common type you've seen or worked on? • What are the advantages and disadvantages of each type of information graphic? • What is the proportion of time your team works on hard news graphics vs. soft news graphics? • Which type of news needs infographics and visualizations the most? Why? • Are there any different features on infographics and visualizations between hard and soft news? • If there are, what motivates the graphic team to come up with different features? • Do you think there are different characteristics for hard news' infographics and soft news' infographics? What are they? • Do audiences prefer infographics on hard news or soft news? • Within the team, is there work division between doing infographics for hard news and soft news?
RQ3. What kinds of mechanisms do editors and reporters tend to use on infographics and visualizations to foster conversations and audience engagement?
• What is the ultimate goal of producing infographics and visualizations to accompany stories? • Do you think infographics and visualizations can help foster conversations and audience engagement? • Do you think infographics and visualizations are better than text to accomplish this mission? • Can you describe the key elements infographics and visualizations need to have in order to foster conversations and audience engagement?

- Do articles with infographics and visualizations generate more views or comments?
- What do readers want in infographics and visualizations?
- Can you give examples of your work that you believe greatly helped to foster conversations or audience engagement? How do you know it had an impact?
- What is the future of infographics and visualizations in helping audience empowerment?

Strategies to improve validity.

Intercoder reliability would be calculated and the coding sheet would be modified to keep researcher adjustment objective.

Triangulation of approaches helps secure different perspectives on the material. Methodological triangulation involves the use of multiple qualitative and/or quantitative methods to study research questions. In this study, content analysis and then case study is accompanied by interviews. It is essentially a strategy that will aid in the elimination of bias and allow the dismissal of plausible rival explanations such that a truthful proposition about some social phenomenon can be made (Campbell and Fiske, 1959).

Also, in this paper, when discussing the results, the researcher would use low inference descriptors to promote the validity of qualitative research. Low inference descriptors are the descriptions phrased very closely to the participants' accounts and researchers' field notes (Johnson, 1997). Verbatim quotes are commonly used as a type of low inference descriptor, and therefore this paper will use direct quotes from the subjects to improve validity. Such examples of data not only validate the conclusions but also provide rich illustrations of the topic.

Analysis

Content Analysis.

This chapter reports the major findings of this study. A content analysis was conducted to answer RQ1: Between hard and soft news at the New York Times' websites, what, if any, are the differences in terms of the number and nature of infographics and visualizations used? This study includes 643 infographics and data visualizations at the New York Times' website in 2012. Intercoder reliability was computed as the percentage of agreement with another journalism master's student who was assigned 50 randomly selected graphics for coding. The Holsti-coefficient of inter-coder reliability was .92.

According to Table 1, there were an average of 50 infographics and data visualizations published per month in 2012. The proportion of news type for infographics and visualizations for each month was approximately 11 for politics, six for public affairs, two for education, 19 for business and economy, five for crime and accident, five for health/science, five for sports, and two for episodic reporting of crimes/accidents/war. But the proportion changed depending on the major events happening during a certain period of time. For example, the number of graphics surged to 81 and the number of graphics on sports was threefold the month-average in August. The sharp increase can be explained by the occurrence of the London Olympics during that time.

Of 643 infographics and data visualizations studied, there were a total of 508 graphics built for hard news (79%), which is significantly more than those for soft news (135, 21%). Graphics for business and economy stories made up the largest percentage

(34.5%) among all the news types. News involving government and politics was second with 21.2 percent; public affairs graphics were third with 11 percent.

Static infographics made up the largest category of graphics. There were a total of 425 static infographics (85.1%), while 13 percent were interactive graphics and the remaining two percent were motional graphics, which are graphics that mainly involve videos and animation (refer to Table 2A & 2B). For business stories, 95 percent of the graphics were static, which can be explained by the quest for fast and simple graphics to comply with the frequent update of the stock market and the market tax rate. Approximately one-fourth of sports news graphics were motional or interactive (26%).

Graphs, such as line charts and bar charts are the most popular elements (refer to Table 3). Among all the graphics studied, 57.9 percent of them had graphs. Graphs are built to present statistic data in a more comprehensible style, such as in business news. This is one of the major reasons why graphs are largely used. The categories with the next highest number of elements are maps and photos. Approximately 20 percent of graphics studied incorporated maps or photos in their storytelling.

Table 1: Number of infographics and data visualizations produced by month & by news content

Hard news						Soft news					
Month	Politics	Public Affairs	Education	Economy Business	Crime Tragedy War	Health Science	Sports	Celebrities	Arts Fashion Style	Episodic reporting of crimes/ accident/war	Total
January	13	3	1	21	3	3					44
February	5	7		17	3	6	2	1		1	42
March	16	9	1	15	3	6			2	1	53
April	10	4		17	2	6	3		1	1	44
May	7	10	5	20	2	5	4		1	1	55
June	11	14	2	13	5	7	6		2	1	61
July	6	4	3	16	3	3	9		1	1	46
August	15	6	1	21	12	7	16		1	2	81
September	10	4	2	18	10	7	1		2	1	55
October	15	2	1	23	9	5	2		1		58
November	18	1		15	3	5			2		44
December	7	6	2	26	4	5	2			1	53
Across-month project	3	1			2		1				7
Grand Total	136 ***(21.2%)***	71 (11.0%)	18 (2.8%)	***222 (34.5%)***	61 (9.5%)	***65 (10.1%)***	46 (7.2%)	1 (0.2%)	13 (2.0%)	10 (1.6%)	643
508 (79%)						135 (21.0%)					(100%)

Table 2A: Number of infographics and data visualizations by mode and news content

	Politics	Public Affairs	Education	Economy Business	Crime Accident War	Health Science	Sports	Celebrities	Arts Fashion Style	Episodic reporting of crimes/ accident/ war	Total
Static	107 (78.7%)	63 (88.7%)	16 (88.9%)	***211*** ***(95.0%)***	46 (75.4%)	59 (90.8%)	26 (56.5%)		12	7	547 (85.1%)
Motional	5 (3.7%)						8 (17.4%)				13 (2.0%)
Interactive	24 (17.6%)	8 (11.3%)	2 (11.1%)	11 (5.0%)	15 (24.6%)	6 (9.2%)	12 ***(26.1%)***	1 (100%)	1 (7.7%)	3 (30%)	83 (12.9%)
Grand Total	136 (100%)	71 (100%)	18 (100%)	222 (100%)	61 (100%)	65 (100%)	46 (100%)	1 (100%)	13 (100%)	10 (100%)	643 (100%)

Table 2B: Number of infographics and data visualizations by mode and news type

Mode	Hard news	Soft news
Static	***87.2%***	77.0%
Motional	1.0%	5.9%
Interactive	11.8%	***17.0%***
Total	(100%)	(100%)

Table 3: Number of infographics and data visualizations by form of element

	Hard news	% rel. to total (508)	Soft news	% rel. to total (135)	Total
Table(s)	87	***17.1%***	9	6.7%	96
Map(s)	107	21.1%	18	13.3%	125
Illustration(s)	31	6.1%	64	***47.4%***	95
Photo(s)	72	14.1%	48	***35.5%***	120
Time series	25	4.9%	18	13.3%	43
Flow Chart(s)	17	3.3%	13	9.6%	30
Relation Diagram(s)	11	2.1%	2	1.5%	13
Graph(s)	340	***66.9%***	33	24.4%	373
Video(s)/Audio(s)	9	1.8%	22	16.3%	31
Other(s)	3	0.6%	2	1.5%	5

Table 4: Number of infographics and data visualizations by balance of narratives								
		Author-driven	% rel. to total	Reader-driven	% rel. to total	Hybrid approaches	% rel. to total	
Hard News Total: 508	Politics	116	455	11	25	9	28	136
	Public Affairs	63	***(89.6%)***	6	(4.9%)	2	(5.5%)	71
	Education	16		2				18
	Economy/Business	215		4		3		222
	Crime/Accident/War	45		2		14		61
Soft News Total: 135	Health/Science	58	113	3	11	4	11	65
	Sports	35		4		7		46
	Celebrities		(83.7%)	1	***(8.1%)***		***(8.1%)***	1
	Arts/Fashion/Style	13						13
	Episodic reporting of crimes/Accident/War	7		3				10
	Grand Total	**568**		**36**		**39**		**643**

Table 5: Number of infographics and data visualizations by visual structuring							
		Establishing shot/ Splash Screen	% rel. to total	Consistent Visual	% rel. to total	Progressive Bar/Timeline	% rel. to total
Hard News Total: 508	Politics	11	30 (5.9%)	132	494 ***(97.2%)***	12	30 (5.9%)
	Public affairs	4		70		4	
	Education	2		18		1	
	Economy/business	4		219		3	
	Crime/accidents/war	9		55		10	
Soft News Total: 135	Health/science	3	15 ***(11.1%)***	61	126 (93.3%)	7	19 ***(14.1%)***
	Sports	9		45		8	
	Celebrities			1		1	
	Arts/fashion/style	3		10			
	Episodic reporting of crimes/accidents/war			9		3	
	Grand total	45		620		49	

Table 6: Number of infographics and data visualizations by highlighting													
		Close -ups		Feature Distinction		Character Distinction		Motion		Audio		Zooming	
Hard News Total: 508	Politics	4	14	105	395	20	66	8	23	2	7	1	10
	Public affairs	1		59		10		3				2	
	Education	2		13		1		1				1	2.0
	Economy/business	3	2.8%	174	**77.8%**	21	13.0%	3	4.5%	1	1.4%	1	%
	Crime/accidents/war	4		44		14		8		4		5	
Soft News Total: 135	Health/science	4	10	46	**87**	5	22	3	17	2	16	1	7
	Sports	5		31		7		13		13		5	
	Celebrities					1							**5.2**
	Arts/fashion/style	1	**7.4%**	5	**64.4%**	4	**16.3%**	1	**12.6%**	1	**11.9 %**	1	**%**
	Episodic reporting of crimes/accidents/war			5		5							
	Grand total	**24**		**482**		88		**40**		**23**		**17**	

Table 7: Number of infographics and data visualizations by ordering								
		Linear	% rel. to total	Random access	% rel. to total	User-directed	% rel. to total	Grand Total
Hard News Total: 508	Politics	11	20 (3.9%)	114	450 **(88.6%)**	11	37 **(7.3%)**	136
	Public affairs			62		8		70
	Education			16		2		18
	Economy/business	6		210		6		222
	Crime/accidents/war	3		48		10		61
Soft News Total: 135	Health/science	9	23 **(17.0%)**	53	104 (77.0%)	3	8 (5.9%)	65
	Sports	8		33		5		46
	Celebrities	1						1
	Arts/fashion/style	1		12				13
	Episodic reporting of crimes/accidents/war	4		6				10
	Grand total	**43**		**554**		**45**		**642**

Table 8: Number of Infographics and data visualizations by interactivity				
	Hover Highlighting	Filtering/Selection/Search	Navigation Button	Explicit Button
Interactive	60 (44.4%)	61 (45.2%)	54(40.0%)	49(36.3%)

Table 9: Number of infographics and data visualizations by messaging					
	Hard News		Soft News		
		% rel. to all hard news		% rel.to all hard news	
Captions/headline	507	(99.8%)	135	**(100%)**	642
Annotations	68	(13.4%)	40	**(29.6%)**	108
Accompanying article	354	***(69.7%)***	80	(59.2)	434

Author-driven narrative is the most common approach used for infographics and visualization. Ninety percent of hard news and 83.7 percent of soft news are author-driven, which relies heavily on messaging and has no interactivity. Soft news has a relatively higher percentage of approaches that are reader-driven (8.1%) and hybrid (8.1%). These approaches do not prescribe an order of viewing and usually involve a lot of interactivity (refer to Table 4).

Almost all graphics, 97.2 percent for hard news and 93.3 percent for soft news, employ a consistent visual platform for their visual storytelling. This is the basic technique to keep their viewers undistracted. A slightly higher proportion of soft news makes use of establishing shot and progression bar/timeline (11.1% & 14.1% respectively) (refer to Table 5).

Soft news generally uses more highlighting techniques than hard news. These highlighting techniques include close-ups, character distinction, motion, audio and zooming) than hard news. Feature highlighting is the only exception, which is more prevalent for hard news graphics (77.8%) (refer to Table 6). Feature highlighting is a common highlighting technique used in graphs. The discrete colors used in charts are effective to distinguish and compare items or trends.

Random access is the most popular way of ordering for both hard news (88.6%) and soft news (77.0%) (refer to Table 7). This suggests no suggested path at all for an infographics or visualizations. Viewers can land their eyes on items that are more attractive to them first. Linear paths are usually constructed when using flow charts, which are relatively commonly used in soft news.

All four interactivity features, which include hover highlighting, filtering/selection/search, navigation buttons, and explicit instruction, are evenly used in interactive graphics (refer to Table 8). Sixty percent of interactive graphic use at least one of these features.

Almost all graphics come with a caption/title (99.7%). Soft news has a higher percentage of annotation (29.6%), which helps to reveal interesting patterns and trends viewers might not know about or, worse, may get unnoticed. Seventy percent of hard news is accompanied by articles and 59.2 percent of soft news is paired with articles (refer to Table 9). This number reveals that there are more independent graphic projects for soft news.

Case Studies.

In this section, I present two selected case studies of narrative visualizations. My goal is to give a general idea of the selection process and thinking approaches I used during the content analysis when I coded the 643 infographics and visualization. Throughout, design strategies are marked in **bold face**. These two examples are chosen to highlight a diverse sample of design techniques in political and sports stories' visualizations.

A. 512 Paths to the White House

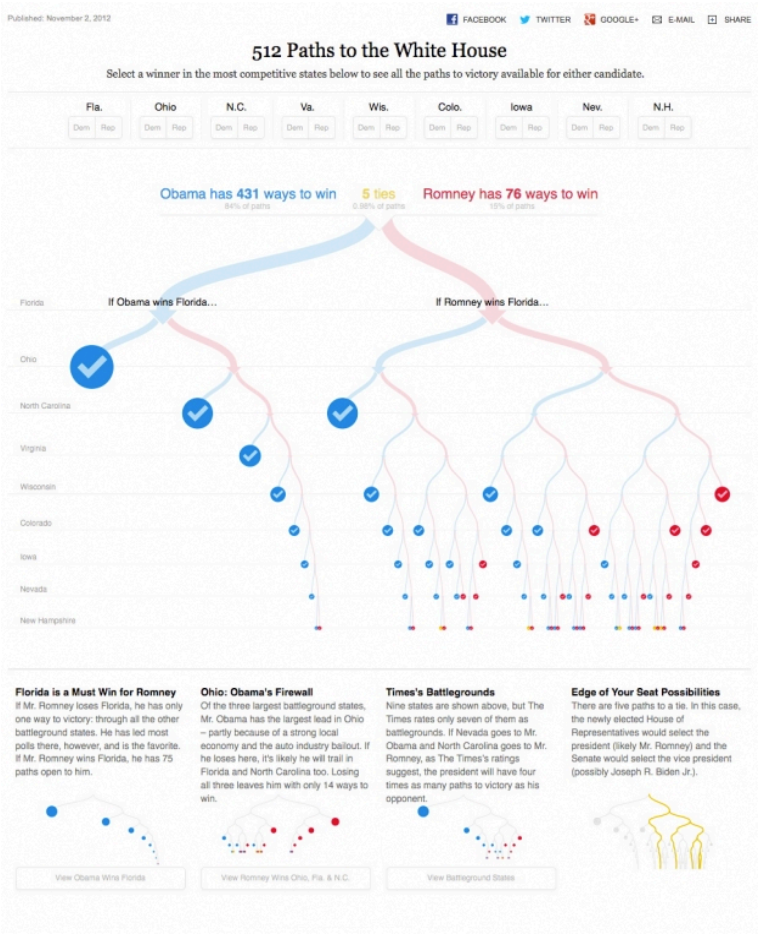


Figure 5: A screenshot from "512 Paths to the White House" a NYT interactive binary tree that shows every possible outcome for the 2012 presidential election

Date		November 2, 2013
Title		512 Paths to the White House
News Content		Politics – Hard news
Interactivity		Interactive
Forms		Flow Chart/Photo
Balance of Narrative		Reader-driven
Visual narrative devices	Visual Structuring	Establishing Shot Consistent Visual
	Highlighting	Feature Distinction Character Distinction Motion
Narrative structure devices	Ordering	User-directed
	Interactivity	Hover highlighting Selection Explicit Instruction
	Messaging	Captions Annotations

With 48 hours left until polls for the 2012 presidential election closed, the New York Times explored all the various ways state votes could add up to a victory for either candidate. Based on a mountain of political analysis and number crunching, this graphic simplifies all the complicated statistics behind it and presents the information in a format that anyone can understand. It demonstrates that of the 512 paths the election can go, 431 of them lead to a Barack Obama victory while only 76 lead to a Mitt Romney win.

This is a standalone graphic, which does not accompany any articles. However, the graphic does a good job of preserving minute details (512 paths possibilities) while providing an effective visual summary of the presidential race. The narratives are rooted in a clear starting point with an **establishing shot** of the binary tree preset to introduce the scene at default. This interactive graphic is classified with **reader-driven approach** and **user-directed narrative structure** as the user can create and test their own paths.

A large **headline** is followed by **explicit text** encouraging the user to explore the routes through the electoral battleground and plot victory for either candidate. Below the text is a series of **buttons** labeled with the swing states (Florida, Ohio, North Carolina, Virginia, Wisconsin, Colorado, Iowa, Nevada and New Hampshire). Users can click on the buttons for each state, giving it to the Democrat or Republican. With each pick, a **binary tree**, which is computed recursively by summing the electoral votes for each path, appears to illustrate what has to happen for each candidate to win. When you flip states, there would be an **animated transition** to show the change of the nodes. A reset button is on the left, which facilitates the exploration of viewers. Each interactive component is clearly adorned with markers of interactivity, explicitly pointing out the potential for interaction.

As the user steps through the visualization, the graphic maintains a **consistent visual platform**, remapping the nodes of the binary tree, but leaving the layout of the graphic the same. **Feature distinction** is adopted: blue paths for Obama and red paths for Romney ensure that the viewer does not lose his place in the narrative during the exploratory stage. When **mousing over** the charts, the path closest to the mouse is highlighted and a human-readable sentence is constructed to explain how the graphic works. Users can create and test their own paths, but there are **annotations** for highlighted paths at the bottom of the page to help them get context. The presentation ends with four-selected scenarios to guide the user. They help apply background knowledge to make the visualization that much richer, such as Obama has the largest lead in Ohio because of a strong local economy and the auto industry bailout. Annotations serve to both highlight interesting views and to reinforce the user's understanding of the context of a graphic and interpretation of its key messages.

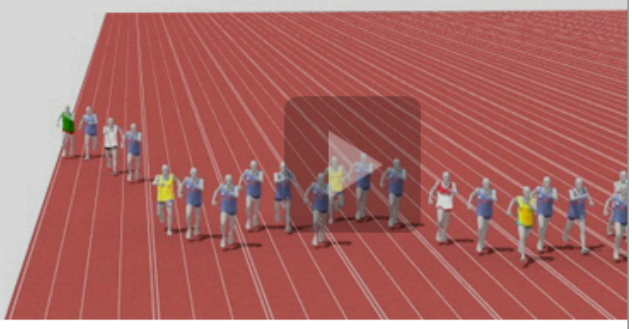
B. One Race, Every Medalist Ever

All the Medalists: **Men's 100-Meter Sprint**

One Race, Every Medalist Ever

Usain Bolt's 9.63 set an Olympic record in the 100. So how far ahead of every Olympic medalist is he?

By KEVIN QUEALY and GRAHAM ROBERTS



Sources: "The Complete Book of the Olympics" by David Wallechinsky and Jaime Loucky, International Olympic Committee; Amateur Athletic Association; Photographs: Chang W. Lee/The New York Times, Getty Images, International Olympic Committee

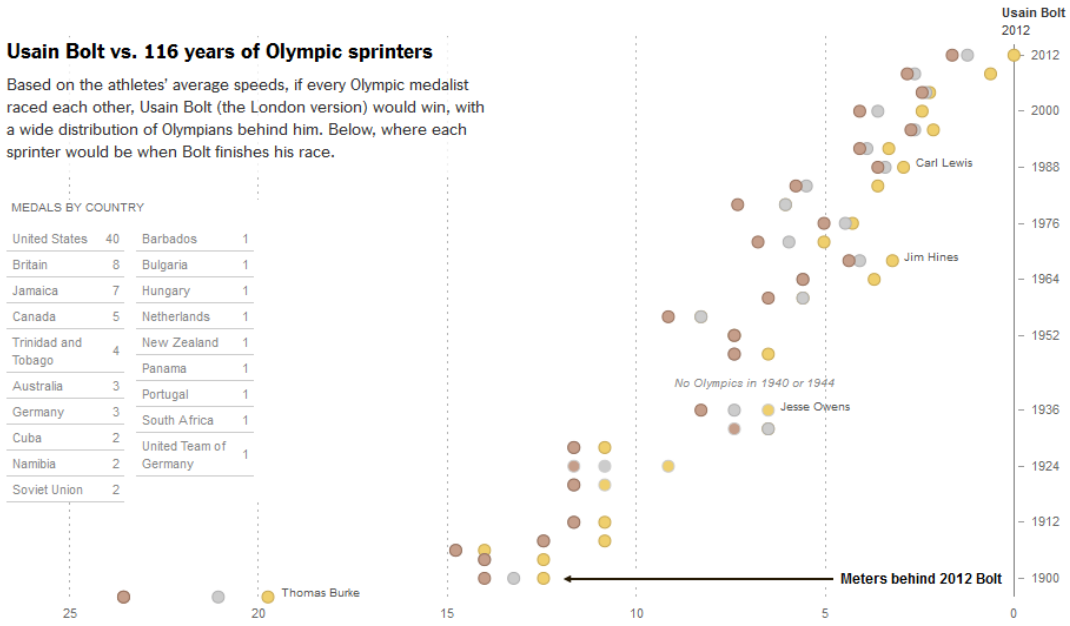
FACEBOOK TWITTER GOOGLE+ EMAIL SHARE

Usain Bolt vs. 116 years of Olympic sprinters

Based on the athletes' average speeds, if every Olympic medalist raced each other, Usain Bolt (the London version) would win, with a wide distribution of Olympians behind him. Below, where each sprinter would be when Bolt finishes his race.

MEDALS BY COUNTRY

United States	40	Barbados	1
Britain	8	Bulgaria	1
Jamaica	7	Hungary	1
Canada	5	Netherlands	1
Trinidad and Tobago	4	New Zealand	1
Australia	3	Panama	1
Germany	3	Portugal	1
Cuba	2	South Africa	1
Namibia	2	United Team of Germany	1
Soviet Union	2		



This chart includes medals for the United States and Australia in the "Intermediary" Games of 1906, which the I.O.C. does not formally recognize.

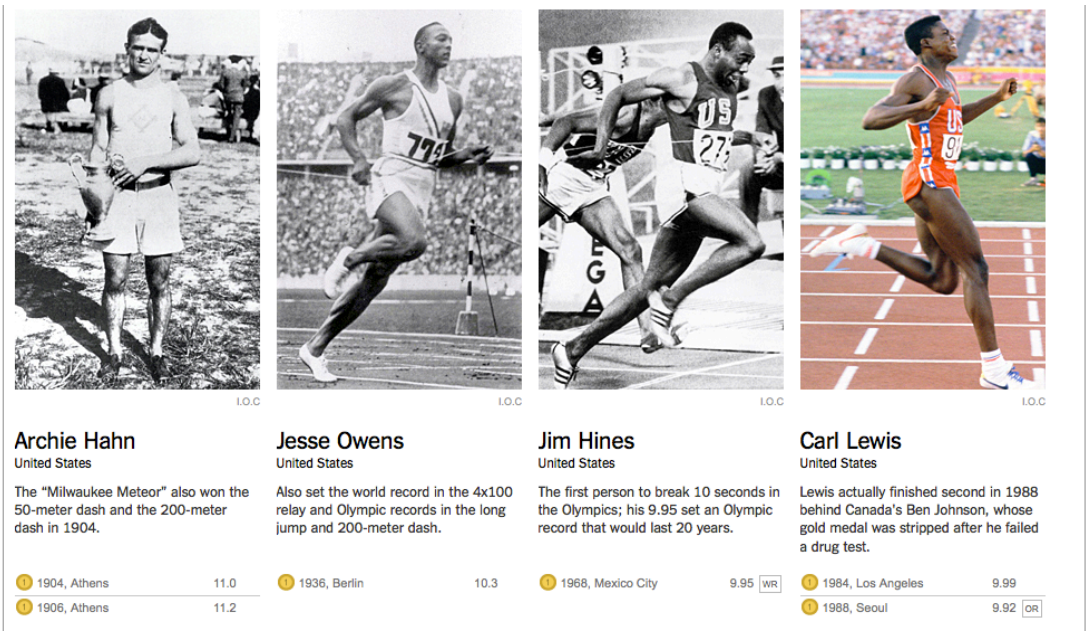


Figure 6: A screenshot from "One Race, Every Medalist Ever," a New York Times animated visualization showing how record-holder Usain Bolt's performance compares to past Olympic medalists.

Date		August 5, 2013
Title		All the Medalists: Men's 100-Meter Sprint
News Content		Sports – Soft news
Interactivity		Motional
Forms		Table, Photo, Graph, Video
Balance of Narrative		Random Access
Visual narrative devices	Visual Structuring	Establishing Shot Consistent Visual
	Highlighting	Feature Distinction Character Distinction Motion Audio Zooming
Narrative structure devices	Ordering	Random access
	Interactivity	Hover highlighting
	Messaging	Captions

The Olympic Games provide engaging content and visuals, potential for multimedia, and a plethora of data for presentation and interaction. This second example shows how data integrates with multimedia and animation in sports news. During the

2012 London Olympics, NYT produced a series of "Racing Against History" graphics. They are visuals that display the results of every Olympic gold medalist in the long jump, the 100-meter sprint and the 100-meter freestyle since 1896.

The graphics adopted an **author-driven approach**, which works best when the goal is storytelling or efficient communication. Viewers can choose their own order to go through the package. They can watch the movie first or head to the dot chart in the middle, therefore the ordering is regarded as **random access**.

Take the data visualization package of "All the Medalists: Men's 100-Meter Sprint" as an example to examine. Like any other visualizations, the graphic starts with its own title. Right below the title is a **short movie** to compare Usain Bolt against 116 years of Olympic sprinters in the 100-meter dash. According to the finding from previous content analysis, audio and movie are very popular elements that sports news uses to present the story. Therefore, when the difference between the gold and silver medal for sprint is less than half a second, using a traditional charting form is not so effective; instead a movie with human figures on a 3D track more easily shows the difference of how far back a split second put the silver medalist.

In this movie, the image they use is basically a dot plot (but using runners instead of dots). They tell their story by **zooming** and **panning** to highlight certain runners, distances, specific years and technological achievements. It also very nicely rotates the axis labels as the image moves. At the end of the movie, "**audiolization**" is used to show the duration. All the runners are aligned like keys on a piano, and the resulting sound is like playing a scale across the keyboard. It offers readers a unique audio way to understand the speed for every sprinter.

The second part of the story package is an interactive **dot plot**. On the left panel, there is a subtitle to differentiate the second part of the package. An introductory text is there, too, conveying a sense of "what the dot plot is about." Introductions may engage or disengage potential views, depending on whether the theme of the visualization is relevant to individual interests. Right below the text is a **table** listing the total number of medals each country won from 1896 to 2012. On the right panel, it comes to the dot plot. Instead of showing the actual winning times, they converted the differences in winning times into distances. It is another visual way to convey the meaning of the tenths and hundredths of a second that separate the top performers. Users can discover additional statistics by **mousing over** the chart, revealing further details in which they are interested. Gold, sliver and bronze medals are colored in the chart to give a sense of **feature distinction**. **Different font sizes** prioritize the importance of the information.

The final part of the visualization is showing the notable winners of the 100-meter sprint throughout history. Large portrait **photos** and illustrations are used to enrich each narrative.

Interviews.

Three NYT reporters, who are known for their high-quality work in interactive storytelling and graphics, were invited to participate in the study.

- Amanda Cox, Graphics Editor
- Alastair Dant, Interactive Developer
- Aron Pilhofer, Associate Managing Editor, Digital Strategy

My goal in this section is to enrich my research with industry perspective. I seek evidence and proof of the results I found in my content analysis. Each of them talked about their insight into infographics and data visualization for hard news and soft news. They offered examples to illustrate how interactivity helps audience engagement.

I did face-to-face interviews with Dant and Pilhofer during the NICAR conference (an industry conference focused on discussing topics such as data journalism and data visualization) in late February 2014. Both Dant and Pilhofer were speakers. They shared their knowledge and the work they did at NYT during their sessions. Their sessions offered some insight to this research and some follow-up questions were asked. After two face-to-face interviews, an e-mail interview was done with Amanda Cox.

When asking questions about **what motivates editors and reporters to come up with different features in infographics and visualizations in hard versus soft news**, all of the interviewees agreed that the features to use mainly depend on the story. There is no one-size-fits-all approach to how interactive graphics are created. There is no decision made based on whether it is a hard news or soft news. Pilhofer said features used mainly depend on the story. "There is not a preference. It is something we want to contribute that matches the story content."

Cox pointed out "things involving scale, context, or patterns" need visualizations the most. She also raised a point that "Hard news is not a good place to experiment with new forms." The same for Dant, who also explained that breaking news is much harder to visualize due to tighter deadlines.

> "There is a process that cannot be easily brought to breaking news. Doing a visualization package is like having a studio. Designers, people who work on the CMS, people who create social media, the graphic team…all the staff are combining together, and you need time to coordinate."

However, he pointed out that maps could be a better option for hard news.

> "I noticed that, for graphics in hard news, maps are pretty simple but [they are some] very useful description of geography. Maps offer clear and concise visualizations of data for readers. For example, [during the Boston Bombing] people [could] have a sense of where the bombs went off in relation to the geography. There is always a race against [time in] newsrooms during the happening of breaking news. The best graphic is any graphic that the newsroom offers at the first time."

Concerning **what kinds of mechanisms do editors and reporters tend to use on infographics and visualizations to foster conversations and audience engagement,** all of them agreed that interactivity may aid in learning and provide an entertaining use of media. Dant noted that graphics' essential capacity is interactivity. Interactivity can bring interpersonal communication and shifts the controlling power to the readers. This offers far more engagement to the readers than traditional mass media forms, such as newspapers, television, radio stations and magazines, in which readers tend to receive information passively. Cox mentioned that the process of interaction helps people to discover stories they may not have found otherwise.

Pilhofer said people love interactive:

> "Do readers use the searchers and filters [in the visualization]? Do they actually stay and engage more deeply with the interactive than those [story] just with pictures? The answer is yes."

Pilhofer brought up The Red Carpet Project as an example to illustrate the popularity of interactive features. The Red Carpet Project explores an archive of Oscar red carpet looks that spans 17 years. Pilhofer elaborated that the project represents a new take on the slide show, enabling filtering, searching and sharing of certain slices of slides. Built into the project is an implicit logic of engagement; while the images are on display, readers are encouraged to explore the dresses and suits curated by editors and filter the images by time, style and color. If all goes well, the readers share their own selection of outfits on social media. NYT found that more than half of the users viewed only slides and 5% clicked every slide without engaging other features. Pilhofer mentioned:

> "We discovered that 5% of readers were very obsessive to click through all the 500 more slides and did nothing else, [but] just to explore the dresses. [Imagine] Click… click… click… People love slide shows, but we did not know people would click 500 times."

Yet, there is still little understanding of how effectively interactive features help to foster conversation and audience engagement. The New York Times just started the analytical team in April 2013, and there is only one person involved in the work. Pilhofer explained: "We don't really do in-depth [research] on reader behavior. We don't track interaction. We don't really have a deep understanding of how readers interacted with stuff."

However, during the interview, both Dant and Pilhofer bought up the "How Y'all, Youse and You Guys Talk" project published in mid-December 2014 and showed their optimism about interactivity. "How Y'all, Youse and You Guys Talk," an interactive quiz about dialects, made No. 1 on the list of the publication's most-visited stories after being published about two weeks before the end of 2013.

Pilhofer described the piece as a graphic project with a quiz and a map at the end of it. The creator, Josh Katz, once described dialect as all about people's sense of identity — "this is who I am, this is where I come from". However, beyond sentimentality or being able to identify your roots, it's an entertaining feature.

Dant agreed that audiences are leaning toward more-interactive consumption of news, and quizzes offer publications another way to provide that for readers, along with a little fun.

> "Having an interactive means enabling [viewers] to find themselves in the data. It is actually a powerful thing. Rather than reading about it, you are actually participating. We don't have data [to prove this], but we are sure that is the case."

Cox shared the same opinion on how to gain audience engagement, but she also was concerned about the challenges. "The challenge of telling a story with a dataset is identifying individuals in the visualization in such a way that their stories form part of the broader epic," she said.

Pilhofer speculated about the reasons behind the popularity of personality quizzes and offered curiosity as an explanation. "That would be my guess. We saw a lot of indications that there is lots of traffic through the social media," Pilhofer said. Pilhofer pointed out that offering an interactive package which is entertaining and engaging so

that people wanted to share with their friends — that helped people figure out where their speech patterns fit in the bigger picture — helps keep people coming back.

Discussion

This study found several notable differences, in terms of number and nature, between infographics and data visualizations used in hard news and soft news at the New York Times in 2012.

First, in terms of number, there are significantly more graphics built for hard news. Graphics about business and the economy made up the largest portion. Data visualizations can make unnoticed numbers and significant numerical data in a business reporting easily recognizable. However, graphics for soft news demonstrate more highlighting features and interactivity. Especially for sports news, more than a quarter of them involve animation and interactivity in their storytelling. Reporters found that with hard news it is much harder to explore unique features due to a tight working schedule. In contrast, visualizations for soft news are usually planned ahead. This allows sufficient time for reporters, designers and coders to form a team to experiment with unique interactive features. This explains the reason why soft news comes with animation, audio and timelines more frequently.

In the content analysis, it was also found that the most common element used in graphics are graphs/charts. Graphs allow us to explore statistical data and observe patterns that no other approach can achieve. Graphs/charts usually adopt the feature distinction highlighting technique. Using discrete colors in graphs/charts can help distinguish and compare items or trends.

Author-driven and random access are the main approaches of narrative for both hard news and soft news. This reveals that infographics and data visualizations are still relying heavily on messaging. The use of captions, introductory text and annotation has the advantage of providing an additional layer of context over the data, which helps the audience come to a valid interpretation of what it really means.

Revising the visual rhetoric theory discussed in the literature review, designers' choices of design principles largely shape how a viewer interprets a story. With a target audience in mind, a designer can select elements that tap into or coincide with the audience's beliefs and cultural expectations. Cultural knowledge provides the basis for interaction (Scott, 1990). Based on the case studies and the interviews, it is discovered that there are several design principles NYT stringently follows.

Simple Layout: Simple interface and easily accessible content are the keys. One reality is that, if a user is confused by an interactive, or if they feel it takes too much time to get to the content, they are going to leave. Therefore, the layout just needs to include fairly simple items, with a limited amount of tabs and navigation buttons. Short and explicit instruction alongside is encouraged.

Clear Content: In most of the journalism visualization, context is added to data visualization through the use of labels, captions and other annotations — texts — of various kinds. Indeed, from the examples studied at the content analysis and case studies, visualizations not only have integrated textual annotations, but an entire one- or two-paragraph introductory article associated with them. In addition to adding an angle and story to the piece, such contextual journalism helps flesh out what the data means and

guides the reader's interpretation towards valid inferences from the data. Textual annotations help enrich the story content and promise some insightful knowledge to viewers.

Unique Presentation: Infographics and data visualization with subtle and interesting presentation can also have a positive impact on the project and encourage users to spend more time with the content.

Engaging Exploration: From these examples studied, a pattern emerges: A "successful" visualization is one that presents a reader with a story and offers them an opportunity to dig deeper. Users love to find that they are in control. This ability to explore makes it possible for users to highlight the pertinent information they want to share.

However, it is found that the New York Times has not done much to understand users' interest when it come to data visualization. Pilhofer once wrote a blog post that reflected on the impoverished status of newsroom analytics, claiming:

> "…the benchmarks we use now are so ill suited. They are the simplistic, one-dimensional metrics we all know: page views, time on site, unique [visitors]. We use them largely because they are there and because they are easy."

Most of his frustration lies in the inability of these methodologies to capture deeper ideas like attitudinal, behavioral, or legislative change — in a word: impact. A future experimental study could examine this problem further and explore the amount of time users spend on news stories that contain data visualizations and interactive graphics compared to the amount of time they spend on news stories that do not have any. Another limitation of the present study was the scope of infographics and visualization examined.

It is known that technologies used to create visualizations have rapidly improved. A future examination could look into graphics of certain types of news across several years and several newsrooms.

During the study, it is noted that new stories now tend to blend a written narrative with interactive elements and videos clips and package it all in an eye-catchingly beautiful layout. The New York Times' "Snow Fall" is an example for this new direction of interactive feature. Rather than just a usual news article, the story provides an experience to viewers. It would be another interesting topic to look at how effective interactives could provide experiences to viewers.

Conclusion

News stories are increasingly accompanied by infographics and visualizations. It is generally believed that these new forms of storytelling reveal patterns of distribution, clusters, anomalies and correlations. In an era of big data, it will inevitably become an increasingly important cognitive tool. The findings of this study are a start to understanding the use of infographics and visualizations in hard and soft news and what built-in features could help foster conversations and audience engagement. As data visualizations and interactive graphics are gaining more significance in journalistic storytelling, more research is needed to study their impact.

APPENDIX

A: Original Project Proposal

Graduate Project Proposal

NG, Yee Man (Margaret)
Journalism School
University of Missouri

How The New York Times uses information graphics and data visualizations for hard news and soft news and to foster audience engagement

Committee Chair: David Herzog
Committee members: Scott Swafford, Barbara Cochran

Introduction

Visual communication has always been my strength. Since my childhood, I have learned best from visual displays such as diagrams, illustrated textbooks, maps, PowerPoint slideshows, photos and videos. I drew charts, graphs and tables in all of my high school notes to help me memorize all those complicated biological terms or chemical reactions. I had a sketchbook to draw mind-maps to help me develop ideas to write a story.

My graduate study at the Missouri School of Journalism extended my definition of visual communication. I found that storytelling is not restricted to words, photographs or videos. Data and graphics are other promising devices for journalists to tell stories. The combination of data analysis and graphic visualizations can be very powerful. Therefore, I planned my graduate coursework in this growing field, starting with my first semester at the J-school. I have taken courses such as Information Graphics, Computer-assisted reporting, Introduction to Geographical Information Systems and Convergence Reporting. I dusted off the coding skills I had learned in high school and started to play with MySQL, Python, Ruby and JavaScript. I did an independent study on interactive data visualizations with JavaScript/d3 library. I built interactive maps with the Google API. I explained government spending and politicians and business actions in infographics and visualizations. I attended the Society of News Design industry conference, where I met design professionals from around the world. My working experience at the Columbia Missourian, National Geographic Magazine and Investigative Reporters and Editors as a graphics reporter or a web designer have provided me valuable opportunities to explore new forms of storytelling at the confluence of design, data and

coding. All these have prepared me for my master's professional project, working in a professional newsroom and studying how *The New York Times* uses infographics and data visualization.

This professional project will be another valuable experience for me to gain insight and experience working at a professional newsroom. Every inspiration I gain and storytelling skill I learn will help me achieve my creative and career goals, as a data reporter or a visual communication researcher.

For my professional project, I chose to participate in the J-school's Washington Program in Washington, D.C. I am seeking the best environment in which to present news stories with data visualization. I am looking for job opportunities at the graphic desks of USA Today, ABC News and NPR. I am also looking at a design/data editor jobs at the Center for Public Integrity, the Pew Research Center and the Sunlight Foundation. I would work for 14 weeks, starting from mid-January to mid-April, full time from Monday to Thursday.

Research Questions

A new visual-based communication era is upon us. Thanks to the open government data movement and the advancement in computing for processing and presenting data, visualization is becoming the "next mass communication medium" and is emerging as a critical tool for helping readers navigate the abundance of information (Viegas & Wattenberg, 2011).

News stories are increasingly accompanied by informational graphics and data visualizations, and these are becoming more sophisticated than ever (Giardina & Medina, 2013; Utt & Pasternak, 2000). Illustrations, charts, databases, maps and other interactive

elements are built to encourage interaction and provide readers multiple layers of news detail. Visualization has the potential to reveal unnoticed information, especially in large data sets; to give answers faster; to help journalists investigate cause-effect relationships; and to help audiences with limited education and short attention spans understand data more easily. News sites such as usatoday.com, elmundo.es, or bbc.co.uk use information graphics and data visualizations to explain complex information clearly and intelligibly.

However, even though information visualization has a long history in other disciplines, such as engineering and statistical modeling (Chen, 2004), there is little agreement on the best way to integrate visualizations into the news production process (Weber & Rall, 2012). Newsrooms are experimenting and finding the best processes for producing infographics and visualizations.

This study will explore and compare the use of infographics and visualizations in hard and soft news, examining the digital portal of The New York Times. My research will examine what built-in features editors and reporters tend to use to foster conversations and audience engagement. Content analysis, case studies and interviews would be conducted to address the following research questions:

RQ1: Between hard and soft news at the New York Times' websites, what, if any, are the differences in terms of number and nature of infographics and visualizations used?

RQ2: What, if anything, motivates editors and reporters to come up with different features in infographics and visualizations in hard versus soft news?

RQ3: What kind of mechanisms do editors and reporters tend to use in infographics and visualizations to foster conversations and audience engagement?

Theoretical Framework

To understand the use of infographics and visualizations on the New York Times' website, visual rhetoric theory and narrative visualizations categories framework derived by Segel and Heer (2010) are adopted.

Rhetoric theory.

Rhetoric is an interpretive theory that frames a message as an interested party's attempt to influence an audience. The sender's intention manifests in the argument, the evidence, the order of argumentation, and the style of delivery (Corbett, 1965). The sender crafts the message in anticipation of the audience's probable response, using shared knowledge of various vocabularies and conventions, as well as common experiences. Receivers use the same body of cultural knowledge to read the message and the sender's argument, and to formulate a response.

Visual rhetoric in a broader sense.

Visual rhetoric is one of the new disciplines taken into account by semantic multimedia and visual communication researchers. Rhetoric was first largely employed in the domain of speech and writing programs. However, in 1965, Gui Bonsiepe argued that all of the decisions made in the layout and presentation of information are rhetorical and that theories of rhetoric should include the interplay of words and images (Veltsos, 2009). Twenty years later Robin Kinross (1985) analyzed the typeface and color used in a railroad timetable and asserted that designers make choices about when and how to use conventions based upon audience and context. This reaffirmed Bonsiepe's argument that all design is rhetorical.

The term visual rhetoric refers to the intended meanings that are represented in the visualization via a designer's choices and then shaped by individual end-user

characteristics, contextual factors involving societal or cultural codes, and the end-user's interaction. With a targeted audience in mind, the designer can select elements that tap into or coincide with the audience's beliefs and cultural expectations and make visual arguments more persuasive. Cultural knowledge thus provides the basis for normative interaction and persuasion (Scott, 1990).

Foss (2004) summarized visual rhetoric as "both a visual object or artifact and a perspective on the study of visual data." It is a product individuals create as they use visual symbols for the purpose of communication and a perspective that scholars apply to symbolic processes by which visual artifacts perform communication. As an area of focus, visual rhetoric has three areas that scholars typically study. They are:

(1) Nature. It focuses on the components (space, color, etc.) of the image.

(2) Function. It concerns the effect of an image on the audience, not necessarily the creator's purpose.

(3) Evaluation. It tries to assess the effectiveness of the image. However, because function isn't singular in visual rhetoric, assessing the effectiveness of an image to meet its purpose is difficult.

Rhetoric associated with persuasion.

Rhetoric has come to be associated with persuasion as a result of the implicit motivation of the sender to gain other adherents to a preconceived view or conclusion (Bogost, 2007). McGuire (2000), the pioneer of persuasion theory, suggested that a better understanding of visual rhetorical figures is most likely to provide new insights into persuasion processes. Ann C. Tyler (1992) summarized the three purposes of design: to persuade the audience to act, to educate the audience and to provide the audience with an

experience. The goal of visual communication is to persuade an audience to adopt a new belief. The visual rhetoric pertains to the visual elements of the document and affects the reader's initial impression of the document (Brumberger, 2001). It can also affect the tone, author's voice and credibility of a document (Kostelnick and Roberts, 1998). Strachan and Kendall's (2004) analysis of political candidates' convention films was an example of the evaluation of visual rhetoric. The study pointed out that politicians' emotional appeals via visual cues encourage "unquestioned acceptance" of politicians and policies.

Studies of visual rhetoric in infographics and visualizations.

Although the outlines and depths of visual rhetoric of data visualization are still being explored, a few studies recognize that all information graphics are visually inscribed rhetoric. Kress and Van Leeuwen (2006) reminded that the study of visual rhetoric is different from that of visual or graphic design. Rather than purely aesthetic consideration, it emphasizes images as sensory expressions of cultural meaning.

Allen (1996) summarized that visual rhetoric, which helps us understand how visuals communicate, focuses on seven purposes: teaching visual information, heightening awareness of visually informative features, evaluating the artistry of visual features, processing visuals and text, obeying graphic grammar, integrating visual and verbal languages aesthetically and using efficiency and clarity to inform visually.

Kimball (2006) examined Charles Booth's maps of London poverty (1889-1902) and analyzed the cultural basis of ideas of transparency and clarity in information graphics. He argued that Booth's maps derive their rhetorical power from contemporary

visual culture as much as from their scientific authority. The visual rhetoric of the maps depended upon an ironic inversion of visual culture to make poverty seem a problem that could be addressed, rather than an insurmountable crisis. Information graphics are inherently rhetorical and have the power to influence social policy.

Visualizations can be appealing, enjoyable and understandable. Data visualizations grant huge rewards for society regarding enhancement of perception, persuasion and interpretation. Visual rhetoric is used as an analytical framework in this study to help understand how design techniques prioritize particular interpretations in visualization storytelling and their influence on end-user interpretation.

Segel and Heer (2010) narrative visualizations categories.

In response to the growing number of online visualizations designed to convey a story, Segel and Heer (2010) identified clear categories to distinguish different forms of data visualization. Although the samples they used might not be exhaustive, their study establishes a very useful framework to access the visual narrative, structuring, and storytelling aspects of data visualization. The framework is very useful to my study. They presented three categories to distinguish narrative visualizations: (1) Balance of narrative, (2) visual narrative devices, and (3) narrative structure devices.

(1) **Balance of narrative** refers to the spectrum of author-driven, reader-driven and hybrid approaches that balance a narrative together with interaction and messaging.

- An author-driven approach refers to a linear path through the visualization, which relies heavily on messaging and has no interactivity. Examples include watching a film, educational videos or training materials.
- A reader-driven approach, on the other hand, does not prescribe an order of

viewing and usually involves a lot of interactivity. Visual analysis tools commonly have this function for tasks such as data diagnostics, patterns discovery and hypothesis formation.

- A hybrid-mix, as the name implies, falls in between the balance of the above approaches. Segel and Heer (2010) say this category gains increasing popularity with visualizations.

(2) **Visual narrative devices** are the visual mechanisms that assist and facilitate the narrative. The authors divide these devices into visual structuring mechanisms, progress bars, consistent visual platforms and highlighting.

- Visual structuring mechanisms communicate the overall structure of the narrative to the viewer and allow him to identify his position within the larger organization of the visualization. Some use visual structuring mechanisms to orient the viewer early on with an overall view or consistent visual platform and to allow the viewer to track his progress through the visualization.
- Progress bars or timeline bars indicate the length of visualization. They provide users with a mechanism to navigate and allow them to skip around the visualization to parts deemed more interesting.
- Consistent visual platform refers to when a visualization or a slideshow progresses and how only the content within each panel changes while leaving the general layout of the visual elements the same. Each new slide changes the text, while animated transitions propel the story forward.
- Highlighting refers to visual mechanisms that draw users' attention to specific areas on the screen by augmenting it with distinctive features such as color,

motion, framing, size, and audio.

(3) **Narrative structure devices** are the non-visual mechanisms that assist and facilitate the narrative. The authors divide these tactics into three sections: (i) ordering, (ii) interactivity, and (iii) messaging.

- Ordering refers to the different ways of arranging the path viewers take through a visualization, where sometimes the path is prescribed by the authors (linear), sometimes there is no suggested path at all (random access), and other times the user selects among multiple alternatives (user-directed).
- Interactivity categorizes the different ways a user can manipulate the visualization. For example, by using navigation buttons, filtering, selecting, searching or hover highlighting.
- Messaging denotes the ways in which a visualization communicates with observations and commentary to the viewer (achieved by labels, captions, headlines and annotations). It helps clearly communicate through the interaction of text on one side with annotations and graphic elements on the other side by providing related but different information.

Segel and Heer's contribution of abstract structures and genres provides a general framework that opens the discussion of narrative visualization to a wider range of examples. Adapting the framework in this research would allow comparisons between visualizations produced in newsrooms.

Literature Review

Definitions of Hard news and Soft news.

Media scholars have regularly made distinctions about the production processes of various types of news content, their effects on the resulting products, and the subsequent social and political consequences. The distinction between hard and soft news is one of the foremost examples of this analytical strategy and has been widely employed by communication scholars (e.g., Patterson 2000; Scott and Gobetz 1992).

Hard news is "the coverage of breaking events involving top leaders, major issues, or significant disruptions in the routines of daily life, such as an earthquake or airline disaster" (Patterson, 2000, p.3). On the contrary, soft news aims more to entertain. It is usually less political in content, but more human interest stories and special news. It is typically "more sensational, more personality-centered, less time-bound…and more incident-based than other news" (Patterson, 2000, p.4). Shoemaker and Cohen (2006) defined hard and soft news according to topicality or timeliness. Hard news items are urgent occurrences that have to be reported right away because they become obsolete very quickly. In contrast, soft news items are usually based on nonscheduled events.

Lehman-Wilzog and Seletzky (2008) determined that hard news is often defined as news content that covers political, social or economic topics, demands immediate reporting due to its importance in order to stay relevant and has actual ramifications over a wide spectrum of society. Soft news, on the other hand, is defined as having little or no intrinsic social or personal importance, so that it can be reported on at any time, as well as news that rather than being relevant to the lives of those receiving the news instead affects only a tiny fraction of the viewing audience.

Conceptual definitions of hard news and soft news in this study.

For the purposes of my study, hard news stories are accounts of events that have

just happened or are about to happen. For example, crimes, fires, meetings, court testimony, speeches, protest rallies, acts of war, traffic accidents and elections are all typical topics of hard news stories. They emphasize facts but not opinion. Soft news would be more likely to be reports about celebrities, human interest, sports and other entertainment-centered stories. However, for news about crimes, fires, accidents, I would also keep an eye on the stories' topicality or timeliness. Hard news items are urgent occurrences that have to be reported right away because they become obsolete very quickly. Breaking news stories that develop overnight or on the same day are hard news. Episodic reporting of crimes and fires, for example, would fall under the definition of soft news.

Definitions of infographics and data visualizations.

Card et al. (1999) defined infographics as "The use of computer-supported, interactive visual representations of data to amplify cognition"; Newsom & Haynes (2004) defined infographics as graphic visual representations of information, data or knowledge intended to clarify and integrate difficult information quickly and clearly. They are usually used to summarize data. Mike Scaife (1996) defined visualization as a mechanism by which humans perceive, interpret, use and communicate visual information. It focuses more on design to allow users to explore datasets for their own purposes. That is, where infographics tell stories designed by communicators, information visualization helps readers discover stories by themselves.

Card et al. (1999) illustrate that the field of data visualization covers the properties of visual perception to resolve logical problems. It investigates how a visual display of information — by automatically assembling thousands of data objects into

pictures, revealing hidden patterns — can serve as a new method for amplifying cognition and generate new knowledge and insight about the world. Card et al. conclude that diagrams can help in six ways: increasing the memory and processing resources available to users; reducing the search for information; using visual representations to enhance the detection of patterns; enabling perceptual inference operations; using perceptual attention mechanisms for monitoring; and encoding information in a medium that can be manipulated. Steve Pasternack and Sandra Utt (1990) reported that readers use information graphics strategically to seek out information. J. Votika Ramaprasad (1991) found that the reader's understanding from information graphics was limited and that the information presented in the graphic sometimes misled the reader. However, in a systematic program of research, Jeffrey Griffin and Robert Stevenson (1994) demonstrated how various information graphic tools such as locator maps, explanatory graphics and graphs facilitate learning.

Several studies have stressed that infographics allow newspapers to optimize the understanding processes thanks to compressed quantity of information and a greater precision, anchored in image and text. Recent infographics researches have emphasized the problem of recognizing infographics (Huang & Tan, 2007), arguing that understanding infographics is a discourse-level problem while others have explored the intersection between infographics and games (Diakopoulos et al., 2011). Meyer (2004) found that interactive infographics can help newspapers add value and enhance the quality of their informational product and therefore have a pivotal function on the influence and credibility of a newspaper company (Meyer, 2004).

Previous research on visualizations' storytelling power.

Infographics were largely used to support the work of journalism back in the 1980s. It used to accommodate newsletters, newspapers, magazines, and reports. Now storytelling has become a new focus in visualization research and practice. More recently, Hullman and Diakopoulos (2011) presented a rhetorical framework for narrative visualizations that includes design choices about the dataset, visualization, and interactivity as well as "extra-representational" factors on how a visualization may be interpreted. They identified a number of approaches to communicate authority, completeness of data, etc., and showed how these cues can be used to prioritize particular interpretations.

Segel and Heer (2010) studied how specific data visualizations are produced and integrated into online news by identifying narrative design differences and by making recommendations on best practices. One of the most interesting structures is what they called the Martini glass, which starts with a broad introduction, then narrows to make a particular point, and then opens up interaction and exploration to the viewer.

Tactinsky and Meyer (1999) found that the presenters created different displays when using data they viewed favorably as opposed to those they disliked. Their findings showed that people tend to create more complicated graphics when they want to persuade or impress the audience.

Little work has been done on understanding and comparing how infographics and visualizations are used differently based on the nature of the news being reported. Scientific research examining the effect of infographics and visualizations on improving audiences' awareness about social issues is rare. George-Palilonis (2006) gave more practical guidelines and described the professional skills for designers and journalists.

Cairo (2005) provided valuable insights into the journalistic process and the common roots of visualization research, perceptual background, and journalistic mission. Dörk et al (2010) has explored the politics of visualization and took engaging visualizations as a starting point and outlined a critical approach that promotes disclosure, plurality, contingency, and empowerment.

In this research, I am trying to learn about the differences in the use and production of infographics and data visualizations in hard and soft news. I focus on three aims: (1) gaining insight into the use of infographics and visualizations in different news stories; (2) exploring and investigating the distribution and differences in infographics and visualizations in different news stories; (3) understanding how infographics and visualizations could help audience engagement.

Methodology

This study employs a mixed-research approach designed to collect both qualitative and quantitative data. The approach includes the following three parts:

- A content analysis of all the infographics and visualizations published on The New York Times website during 2012. The method is chosen to help understand how infographics and visualizations are presented in hard and soft news and, specifically, to address research question 1.
- Two cases of infographics or visualizations, one on hard news and one on soft news would be studied in depth.
- Interviews with three editors at The New York Times who regularly display infographics and visualizations on the newspaper's website. This method is employed to provide deeper insight into what, if any, differences are present in

the infographics and visualizations used between hard and soft news and what strategies are adopted by editors and reporters to foster citizens' conversations and audience engagement. These interviews are intended to address research questions 2 and 3 and serve as an extension to my content analysis and case study. The Institutional Review Board has approved the interviewing protocol.

Rationale of choosing The New York Times.

The New York Times has invested heavily in experimenting with design practices related to infographics production and dissemination. As of 2012, The New York Times infographics department employed 25 highly specialized journalists to research and create diagrams, maps and charts for the newspaper and the website. The New York Times' infographics have been researched in many scholarly studies (Segel and Heer, 2010). Notable examples of their work include a 3-D video explaining how New York Yankees pitcher Mariano Rivera dominates hitters, before-and-after-satellite maps of the earthquake and tsunami in Japan in March 2011, an interactive budget puzzle and a customizable electoral map. Therefore, The New York Times is a logical choice to study in this research.

Rationale of choosing year 2012 (1 January – 31 December, 2012).

2012 was a year with several big national and international events. It was a presidential election year in the United States. A lot of graphics were designed to tell stories about the elections. A data visualization called "512 Paths to the White House" by The New York Times is an example. It calculates the likelihood of Romney and Obama winning, based on which direction each swing states ends up going. 2012 was also the year of the London Olympic games. The New York Times produced many graphics to

illustrate the record-breaking moments and the intense competition among athletes. For example, an infographic video called “Men’s 100-Meter Dash” creates imaginary events in which all the medalists from every Olympic games since 1896 compete together. A lot of graphics and visualizations were produced in this year.

Content Analysis.

The use of content analysis as a quantitative approach in studying newspapers is very popular. In an examination of articles published in *Journalism & Mass Communication Quarterly* from 1971-95, Riffe and Freitag (1997) found that the primary focus of articles using content analysis was on news/editorial content. Krippendorff (1980) defined content analysis as ‘‘a research technique for making replicable and valid inferences from data to their context.’’ According to Kolbe and Burnett (1991), content analysis possesses the following benefits:

- It is unobtrusive, which is particularly valuable in situations in which other methods yield biased results.
- It is helpful in summarizing large bodies of communication messages. If one wants to know, for example, how frequently an issue was discussed in the newspaper in the past year, content analysis would be an appropriate method.
- It enables people to systematically study historical moments and trends over time. For example, it is not possible to interview George Washington, but one could conduct a content analysis of his writings.

The goal of this study is to analyze the common practice of employing infographics and visualization in newspapers’ digital (web) forms and to highlight the distribution of these visual elements in each news section. Through a content analysis of

published placement on the webpage, categories, sources, interactive elements and genres of visualization narratives, the use of infographics and visualizations in different news sections can be evaluated.

I would collect infographics and visualizations produced online by The New York Times in 2012. Unfortunately, The New York Times only includes text in its current archives. Photos, charts, illustrations and other graphics are not included. However, The New York Times has a multimedia search engine inside its website. I would use that search engine to do a comprehensive search, including the following keywords: "interactive," "infographic," "chart,", "graphic," "visualization,", "diagram," and "timeline." To make sure I get all the infographics and visualizations published on the web, I also would go to The New York Times' Twitter accounts to follow its graphics posting history. This could help to double check if the search engine misses some of the graphics. The initial estimation number of graphics on the website would be more than 600.

Graphics that were produced by the newsrooms and for the purpose of storytelling would be collected and coded through a coding scheme.

Table 1: Coding sheet

The coding sheet would record the basic information of the graphics, their placements with either hard or soft news and Segel and Heer's (2010) narrative visualizations categories.

Month	
Title	
News nature	1. Hard news 2. Soft news

News content	5. Politics 6. Public Affairs 3. Economy/Business 4. Science/Technology 5. Sports 6. Celebrities 7. Health 8. Arts and style 9. Crime 10. Accident 11. Others
Level	1. Standalone graphic 2. Supplement to a story
Type of graphic	1. Infographic 2. Data visualization
Complementary elements with the graphic (Can be multiple inputs)	1. Photos 2. Videos 3. Photo Slide Shows 4. Quizzes 5. Databases 6. Others
Dimensions	1. 2D 2. 3D
Interactivity	4. Static 5. Motional 6. Interactive
Forms (can be multiple inputs)	11. Tables 12. Maps (Geographical presentation of data) 13. Illustrations 14. Time-series 15. Flow charts (Showing concepts or processes) 16. Relational/ Organizational diagrams 17. Graphs (Showing data trends or the proportion of data) 18. Others

Forms of Chart (can be multiple inputs)	1. Bar/Horizontal Bar 2. Line 3. Scatter 4. Area/Mountain 5. Radar 6. Pie 7. Surface 8. Doughnut 9. Stream graph	
Balance of Narrative	4. Author-driven 5. Reader-driven 6. Hybrid approaches	
Visual narrative devices (can be multiple inputs) Interaction strategies	1. Visual structuring mechanisms 2. Progress bar or a timeline bar 3. Consistent visual platform 4. Highlighting	
Narrative structure devices	Ordering	4. Linear 5. Random access 6. User-directed 7. No obvious ordering
	Interactivity categorizes	5. Using navigation buttons 6. Filtering 7. Selecting/Searching 8. Hover highlighting 9. None
	Messaging	4. Use of labels, captions, headlines and annotations 5. No obvious messaging

A pilot-coding test would be carried out to test the reliability of the coding sheet. I would invite a peer researcher to try coding for 10 percent of The New York Times infographics and visualizations. I would compare the coding result from my results and my peer's. Intercoder reliability would be calculated with SPSS Crosstab function.

Intercoder reliability is the widely used term for the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion. It

is a critical component of content analysis. It is acknowledged that "given that a goal of content analysis is to identify and record relatively objective characteristics of messages, reliability is paramount. Without the establishment of reliability, content analysis measures are useless" (Neuendorf, 2002, p. 141).

Data would be recorded in an Excel spreadsheet. This can conveniently convert to a .csv file to do more quantitative analysis in SPSS. The percentage of the distribution of infographics between hard news and soft news, the features they are using and other characteristics would be summarized.

Case Study.

Case study research excels at bringing us to an understanding of a complex issue or object and can extend experience or add strength to what is already known through previous research. Researcher Robert K. Yin (1984) defined the case study research method as an empirical inquiry that investigates a contemporary phenomenon within its real-life context. It is used when the boundaries between phenomenon and context are not clearly evident and when multiple sources of evidence are used. Case studies emphasize detailed contextual analysis of a limited number of events or conditions and their relationships.

Two case studies would be carried out as a narrative description and a more detailed analysis of infographics and visualizations between hard news and soft news in The New York Times. Selection of the case would be based on representative examples encountered during the content analysis. At this point, graphics on the U.S. presidential election and London Olympic games would be applicable choices for the case studies.

Interview.

As previously stated, qualitative study also would be conducted. In order to do so, at least three open ended, semi-structured, in-depth interviews would be carried out. According to Newton (2010), interviews have the following benefits for scholarly research:

- They provide the opportunity to generate rich data;
- The language used by participants is considered essential in gaining insight into their perceptions and values;
- Contextual and relational aspects are seen as significant to understanding others' perceptions;
- Data generated can be analyzed in different ways.

Interviews will help me to delve into research questions 2 and 3. It helps to provide insight into how news organizations use infographics and visualizations to tell stories and how these practices could help foster conversations and audience engagement. It serves as an extension of my content analysis and case studies.

Open-ended interviews allow participants to discuss their opinions, views and experiences fully in detail, whereas a set interview with closed-ended questions might inhibit them from expressing their full opinions and feelings. The interviews will consist of more than 10 open-ended questions, uniquely developed for the sole purpose of this study (see Table 2 for a sample interview guide). The interviews will last an estimated 30 to 45 minutes. However, these questions and times are merely a guide; the participants' responses will determine the direction and length of the interviews. The interviews will be audiotaped with permission from the participants to ascertain an accurate account of the interview, which can be replayed for analytic purposes. The interviews will be carried

out over a period of two months, which allows the researcher to reflect and make adjustments as necessary.

Interviewees will be information graphic designers, graphic editors and art directors who work in editorial departments for print and online journalism or TV and regularly plan, design interactive graphics. My potential interviewees include:

- Kevin Quealy, graphics editor at The New York Times*;* Mizzou alumnus;
- Amanda Cox, graphics editor at The New York Times*;*
- Matthew Bloch, graphics editor at The New York Times*;*
- Aaron Pilhofer, associate managing editor for digital strategy/ editor of the Interactive News desk at The New York Times*;*

Table 2: *Interview Guide*

These questions were developed and framed according to the subjects' experiences of terminology and process. A mixture of open and closed questions would be used. The weighting of responses is broadly proportionate to the range of participants but reflects individual concerns.

RQ2. What, if anything, motivates editors and reporters to come up with different features in infographics and visualizations **in hard versus soft news?**
• How do you decide if it is necessary to include infographics and visualizations in an article to tell a story? • How would you categorize graphics? What is the most common type you've seen or worked on? • What are the advantages and disadvantages of each type of information graphic? • What is the proportion of time your team works on hard news graphics vs. soft news graphics? • Which type of news needs infographics and visualizations the most? Why? • Are there any different features on infographics and visualizations between hard and soft news? • If there are, what motivates the graphic team to come up with different features?

• Do you think there are different characteristics for hard news' infographics and soft news' infographics? What are they? • Do audiences prefer infographics on hard news or soft news? • Within the team, is there work division between doing infographics for hard news and soft news?
RQ3. What kinds of mechanisms do editors and reporters tend to use on infographics and visualizations **to foster conversations and audience engagement?**
• What is the ultimate goal of producing infographics and visualizations to accompany stories? • Do you think infographics and visualizations can help foster conversations and audience engagement? • Do you think infographics and visualizations are better than text to accomplish this mission? • Can you describe the key elements infographics and visualizations need to have in order to foster conversations and audience engagement? • Do articles with infographics and visualizations generate more views or comments? • What do readers want in infographics and visualizations? • Can you give examples of your work that you believe greatly helped to foster conversations or audience engagement? How do you know it had an impact? • What is the future of infographics and visualization in helping audience empowerment?

Strategies to improve validity.

Intercoder reliability would be calculated and the coding sheet would be modified to keep researcher adjustment objective.

Triangulation of approaches helps secure different perspectives on the material. Methodological triangulation involves the use of multiple qualitative and/or quantitative methods to study research questions. In this study, content analysis and then case study is accompanied by interviews. It is essentially a strategy that will aid in the elimination of bias and allow the dismissal of plausible rival explanations such that a truthful proposition about some social phenomenon can be made (Campbell and Fiske, 1959).

Also, in this paper, when discussing the results, the researcher would use low

inference descriptors to promote the validity of qualitative research. Low inference descriptors are the description phrased very closely to the participants' accounts and researchers' field notes (Johnson, 1997). Verbatim quotes are commonly used as a type of low inference descriptors, and therefore this paper will use direct quotes from the subjects to improve validity. Such examples of data not only validate the conclusions but also provide rich illustrations of the topic.

Publication Opportunities

I am looking for an opportunity to publish my study in the Online Journal of Communication and Media Technologies, Communication Arts and the Newspaper Research Journal.

References

Bateman, S., Mandryk, R.L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010, April). Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2573-2582). ACM.

Baum, M.A., & Jamison, A.S. (2006). The Oprah effect: How soft news helps inattentive citizens vote consistently. *Journal of Politics*, *68*(4), 946-959.

Bogost, I., (2007). Persuasive *Games: The Expressive Power of Videogames,* The MIT Press.

Brumberger, E.R. (2001). *The rhetoric of typography: Five experimental studies of typeface personality and its effects on readers and reading*. Unpublished dissertation, New Mexico State University, Las Cruces, NM.

Cairo, A. (2005). The future is now. *Design Journal*, *97*(16).

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81.

Card, S.K., Mackinlay, J.D., & Schneiderman, B. (Eds.). (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann (city?).

Chen, C. (2004). *Information visualization: Beyond the horizon*. Springer (city?).

Corbett, E., (1965). *Classical Rhetoric for the Modern Student*, London: Oxford.

Diakopoulos, N., Kivran-Swaine, F., & Naaman, M. (2011, May). Playable data: characterizing the design space of game-y infographics. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1717-1726). ACM.

Dörk, M., Feng, P., Collins, C., & Carpendale, S. (2013, April). Critical InfoVis: exploring the politics of visualization. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2189-2198). ACM.

Foss, S.K. (2004). Framing the study of visual rhetoric: Toward a transformation of rhetorical theory. *Defining Visual Rhetorics*, 303-313.

Gershon, N., & Page, W. (2001). What storytelling can do for information visualization. *Communications of the ACM, 44*(8), 31-37

George-Palilonis, J. (2006). A Practical Guide to Graphics Reporting. Burlington, MA. Focal Press.

Giardina, M., & Medina, P. (2013). Information graphics design challenges and workflow management. *Online Journal of Communication and Media Technologies 3*(1), 108-124.

Griffin, J.L., & Stevenson, R.L. (1994). The effectiveness of locator maps in increasing reader understanding of the geography of foreign news. *Journalism & Mass Communication Quarterly,* 71(4), 937-946.

Hartman, J. K. (1987). USA Today and young-adult readers: Can a new-style newspaper win them back?. Newspaper Research Journal, 8(2), 1-14.

Huang, W., & Tan, C.L. (2007, August). A system for understanding imaged infographics and its applications. *In Proceedings of the 2007 ACM symposium on Document Engineering* (pp. 9-18).

Hullman, J., & Diakopoulos, N. (2011). Visualization rhetoric: Framing effects in narrative visualization. *Visualization and Computer Graphics, IEEE transactions on visualization and computer graphics*, *17*(12), 2231-2240.

Johnson, R.B. (1997). Examining the validity structure of qualitative research. *Education*, *118*(2), 282-292.

Kinross, Robin. (1985). The rhetoric of neutrality. *Design Issues 2* (2):18-30.

Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design*. Routledge.

Kolbe, R.H., & Burnett, M.S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research*, 243-250.

Kostelnick, C., & Roberts, D.D. (1998). Designing visual language: Strategies for professional communicators. Boston: Allyn and Bacon.

Liu, Z., & Stasko, J.T. (2010). Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *Visualization and Computer Graphics, IEEE Transactions on,* 16(6), 999-1008.

Logan, R.A. (1986). USA Today's innovations and their impact on journalism ethics. *Journal of Mass Media Ethics*, 1(2), 74-87.

Meyer, P. (2004). The influence model and newspaper business. *Newspaper Research Journal*, 25(1), 66-83.

Mol, L. (2011). *The Potential Role for Infographics in Science Communication* (Doctoral dissertation, Master Thesis Communication specialization (9 ECTS), Biomedical

Sciences, Vrije Universiteit Amsterdam).

Newton, N. (2010). The use of semi-structured interviews in qualitative research: strengths and weaknesses. *Exploring qualitative methods*, 1 (1), 1 -11.

Neuendorf, K.A. (2002). The content analysis guidebook. Sage.

Patterson, T.E. (2000). Doing well and doing good: How soft news and critical journalism are shrinking the news audience and weakening democracy — and what news outlets can do about it. Joan Shorenstein Center on the Press, Politics and Public Policy, John F. Kennedy School of Government, Harvard University.

Pasternack, S., & Utts, S.H. (1990). Reader use and understanding of newspaper infographics. *Newspaper Research Journal,* 11(2), 28-41.

Ramaprasad, J. (1991). Informational graphics in newspapers. *Newspaper Research Journal*, 12, 92-101.

Riffe, D., & Freitag, A. (1997). A content analysis of content analyses: Twenty-five years of Journalism Quarterly. *Journalism & Mass Communication Quarterly*, *74*(3), 515-524.

Rojas, I. & Ju, W., 2009. Visualization and empowerment. *Proceeding of the seventh ACM conference on creativity and cognition,* p.401-402.

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work?. *International Journal of Human-Computer Studies*, *45*(2), 185-213.

Schramm, W. (1949). The nature of news. *Journalism Quarterly*, *26*(3), 259-269.

Schwalbe, C.B., Silcock, B.W., & Keith, S. (2008). Visual framing of the early weeks of the US-led invasion of Iraq: Applying the master war narrative to electronic and print images. *Journal of Broadcasting & Electronic Media*, *52*(3), 448-465.

Scott, D.K., & Gobetz, R.H. (1992). Hard news/soft news content of the national broadcast networks, 1972–1987. *Journalism & Mass Communication Quarterly*, *69*(2), 406-412.

Scott, L.M. (1990), "Understanding jingles and needledrop: A rhetorical approach to music in advertising," *Journal of Consumer Research*, 17 (September), 223-237.

Segel, E. & Heer, J., 2010. Narrative visualization: telling stories with data. *IEEE Transactions on Visualization and Computer Graphics,* 16(6), 1,139-48.

Shoemaker, P.J. (2006). News and newsworthiness: A commentary. *Communications*, *31*(1), 105-111.

Strachan, J.C., & Kendall, K.E. (2004). Political candidates' convention films: Finding the perfect image — An overview of political image making. *Defining Visual Rhetorics*, 135.

Tractinsky, N., & Meyer, J. (1999). Chartjunk or goldgraph? Effects of presentation objectives and content desirability on information presentation. *MIS Quarterly*, 397-420.

Tyler, A.C. (1992). Shaping belief: The role of audience in visual communication. *Design Issues*, *9*(1), 21-29.

Ura, J.D. (2009). The supreme court and issue attention: The case of homosexuality. *Political Communication*, *26*(4), 430-446.

Utts, S.H., & Pasternak, S. (2000). Update on infographics in American newspapers. *Newspaper Research Journal, 21*(5), 55-66.

Veltsos, J.R. (2009). More than decoration: An investigation into the role of visual rhetoric and ethos in corporate visual identity. Doctorate in Philosophy thesis. Iowa State University.

Viegas, F., & Wattenberg, M. (2011, April 19). How to make data look sexy — CNN.com. CNN.com - Breaking News, U.S., World, Weather, Entertainment & Video News. Retrieved October 18, 2013, from http://www.cnn.com/2011/OPINION/04/19/sexy.data/

Weber, W. & Rall, H. (2012). Data visualization in online journalism and its implications for the production process. *Paper for the 16th International Conference on Information Visualization*. Retrieved via http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6295837.

Yin, R.K. (2009). Case study research: Design and methods (Vol. 5). Sage.

B: IRB Approval Form

Campus Institutional Review Board
University of Missouri-Columbia

485 McReynolds Hall
Columbia, MO 65211-1150
PHONE: (573) 882-9585
FAX: (573) 884-0663

November 4, 2013

Principal Investigator: Ng, Yee Man
Department: University of Missouri

Your Application to project entitled *The use of infographics and visualizations between hard and soft news and their potential to foster civic engagement* was reviewed and approved by the MU Campus Institutional Review Board according to terms and conditions described below:

IRB Project Number	1209880
Initial Application Approval Date	November 4, 2013
IRB Expiration Date	November 4, 2014
Level of Review	Exempt
Project Status	Active - Open to Enrollment
Regulation	45 CFR 46.101b(2)
Risk Level	Minimal Risk

The principal investigator (PI) is responsible for all aspects and conduct of this study. The PI must comply with the following conditions of the approval:

1. No subjects may be involved in any study procedure prior to the IRB approval date or after the expiration date.
2. All unanticipated problems, serious adverse events, and deviations must be reported to the IRB within 5 days.
3. All modifications must be IRB approved by submitting the Exempt Amendment prior to implementation unless they are intended to reduce risk.
4. All recruitment materials and methods must be approved by the IRB prior to being used.
5. The Annual Exempt Form must be submitted to the IRB for review and approval at least 30 days prior to the project expiration date.
6. Maintain all research records for a period of seven years from the project completion date.
7. Utilize the IRB stamped document informing subjects of the research and other approved research documents located within the document storage section of eIRB.

If you have any questions, please contact the Campus IRB at 573-882-9585 or umcresearchcirb@missouri.edu.

Thank you,

Charles Borduin, PhD
Campus IRB Chair

C: Completion Confirmation Letter IRB

Resent-From: yeeman.ng@mail.missouri.edu
From: **Mitchel, Heidi S <MitchelH@missouri.edu>**
Subject: Completion Confirmation Letter IRB #1209880
Date: April 17, 2014 10:21:59 AM EDT
To: David Herzog , Yee Man Ng

Hide

Inbox - mail.missouri.edu

Hello Investigators,

We received your Completion Report to project entitled "The use of infographics and visualizations between hard and soft news and their potential to foster civic engagement". As requested, your project has been closed. If there are any changes to the status of this project in the future, please let us know. Thank you.

Campus Institutional Review Board

D: Recruitment Email

Dear[subject's name]:

Hi, my name is NG, Yee Man (Margaret). I am a journalism graduate student from University of Missouri, Columbia. I am writing to ask if you would agree to be interviewed in person/through Skype for a research project entitled "The use of infographics and visualizations between hard and soft news and their potential to foster civic engagement".

This research aims to gain insight into the usage of infographics and visualizations in different news stories, explore and investigating the distribution and difference of infographics and visualizations in different news stories and understand how infographics and visualizations could help civic empowerment from journalists' perception. Interviewees are information graphic designers, graphic editors and art directors who work in editorial departments for print and online journalism or TV and regularly plan, design interactive graphics. Your name and design works is very popular in the field and you are one of my potential interviewees. I hope you will be willing help us with my study.

Your participation would involve an interview, which would last 30-60 minutes. Interview would take place in your office/through Skype, and I would meet you at the agreed upon time. The interview would consist generally of questions about the use of infographics and visualizations between hard and soft news and their potential to foster civic engagement. The interview would be audiotaped simply to keep accurate track of information. I would identify you by name and acknowledge you with the obtained information from our interview. Participation is voluntary. Nevertheless, I very much hope you can participate.

I will call you shortly to invite your participation and answer any questions you might have. Please know how greatly I appreciate your time and help with this request.

Yours sincerely,
Margaret Yee Man Ng

Graduate student
University of Missouri, Columbia
832-540-1883

E: Written Consent Form

Project name: The use of infographics and visualizations between hard and soft news and their potential to foster civic engagement

INTERVIEW CONSENT FORM

I volunteer to participate in a research project conducted by NG, Yee Man (Margaret) from University of Missouri, Columbia. I understand that the project is designed to gather information about the use of infographics and visualizations between hard and soft news and their potential to foster civic engagement.

1. My participation in this project is voluntary. I understand that I will not be paid for my participation. I may withdraw and discontinue participation at any time without penalty. If I decline to participate or withdraw from the study, no one on my campus will be told.

2. I understand that most interviewees will find the discussion interesting and thought-provoking. If, however, I feel uncomfortable in any way during the interview session, I have the right to decline to answer any question or to end the interview.

3. Participation involves being interviewed by researcher from University of Missouri, Columbia. The interview will last approximately 30-60 minutes. Notes will be written during the interview. An audiotape of the interview and subsequent dialogue will be making. If I don't want to be taped, I will not be able to participate in the study.

4. I understand that the researcher will identify me by name in the report using information obtained from this interview. Subsequent uses of records and data will be subject to standard data use policies and would only be used only for this study.

5. The knowledge contained in this study will not be given to any non-project staff except in cases where it is useful for protection and preservation purposes. When this material becomes available, it may be read, quoted, or cited from and disseminated for educational and scholarly purposes only.

6. If you have questions about your rights as a research participant please contact the MU Campus Institutional Board (IRB) by phone at 573-882-9585 or e-mail umcresearchcirb@missouri.edu.

7. I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.

8. I have been given a copy of this consent form.

_____ My Signature

_____ My Printed Name

_____ Date

For further information, please contact:
NG, Yee Man (Margaret) Phone: 832-540-1883 Email: ync37@mail.missouri.edu

_____ Signature of the Investigator