Process Book

Team Members:

Aashay Sanghvi

Patrick Connors

Will Smiles

This process book will cover our project, Uber Everywhere (a site containing web-based visualizations), and the process we took to build it. We will present the Project Plan, the Project Proposal, an overview of the design decisions, and a summary.

Project Plan

Here is our initial project proposal we submitted:

Call Me An Uber

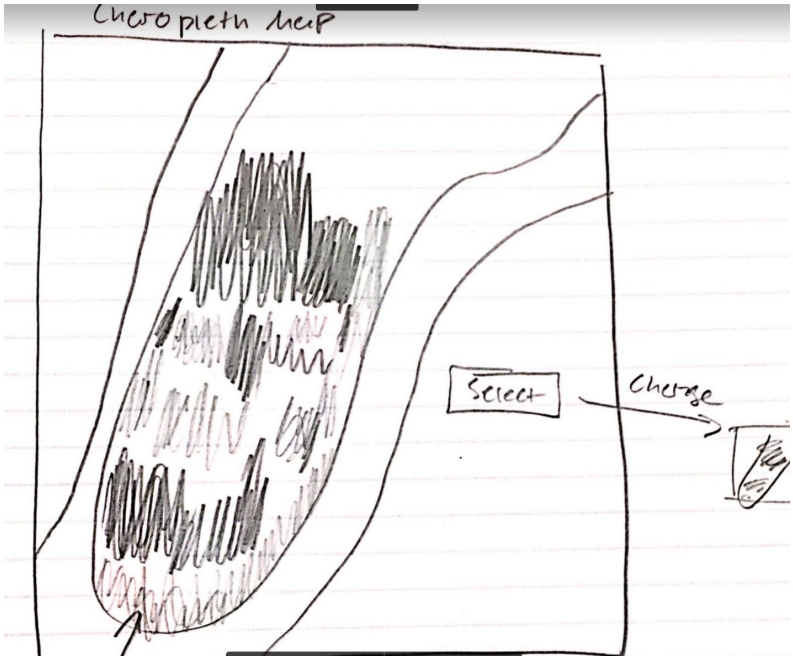
Our group is especially in interested in the future of transportation and how companies like Uber and Lyft are affecting the future of cities. We want to weave a story around how Uber's presence is growing in New York City and in what parts of the city in particular. We also want to compare that with taxi traffic and visualize how the two are affecting each other.

We plan to use a dataset provided open-source by 538 which contains information on over 4.5M Uber pickups in NYC in 2014 and 2015. The data is provided here, on GitHub: https://github.com/fivethirtyeight/uber-tlc-foil-response

We really based a lot of what we wanted to work on based off the data we found. The dataset provided by FiveThirtyEight included pickup data from Uber for a set of individual csvs containing Uber data by months, a csv file containing daily aggregate data for a few months, and a data file with information on weekly averages per hour for pickups. There was also data on other transportation and for-hire vehicles in New York.

At that point, we really started to think through how we could best represent this data and tell a story. We looked at some of the writing samples FiveThirtyEight put together

for inspiration. We also took a look at some of the past projects in the class to get an idea of what direction we wanted to take our project. Primary themes and issues we were concerned with from the start were trying to figure out how Uber was growing, what a typical day for the service looked like in the city, how it stacked up to other transportation services, and if there were any additional factors that influenced pickups. We knew we definitely needed a map of some kind, and this might be a primary feature. This is a very rough sketch of a choropleth map we could have potentially used.



We went through debating if we wanted to do a static choropleth map or a map that included a transition and animation of some kind. This was our primary focus then. Soon, we submitted the project proposal.

Project Proposal

Here is the project proposal we submitted:

Goals and Tasks

Uber launched with in 2010 with a simple, yet revolutionary premise. Tap a button and within minutes, a driver would arrive to take you to your destination. Over the past couple years, Uber has grown at an extraordinary

pace, serving millions of rides per day across the globe and defining new methods of transportation across cities.

The purpose behind our project is to understand what impact Uber has had on transportation in New York City, as well as how specifically the company has spread its reach through the city. Our goals include educating our audience about Uber's growth in New York City through 2014 and 2015, visualizing what impact Uber has had on the taxi cab system in the city, as well as understanding what neighborhoods of New York City are most affected by the growth of this new transportation juggernaut.

Tasks for our team include wrangling, cleaning, and organizing the Uber pickup data and other for-hire vehicle data. We need to make sure all the data are consistent and in the same structures. Other tasks include deciding storytelling methods that will be effective, as well as designing mockups and storyboards of our visualizations. We also need to work on implementing the visualizations and make sure everything looks clean and polished at the end.

Data Description

Our primary data source is this GitHub repository from FiveThirtyEight: https://github.com/fivethirtyeight/uber-tlc-foil-response. Uber is a private company, so normally they wouldn't release their data, but they gave this up in a freedom of information request FiveThirtyEight submitted to NYC's Taxi and Limousine Commission. According to FiveThirtyEight, the data includes information on over 4.5 million Uber pickups in New York City from April 2014 to September 2014, as well as 14.3 million pickups from January 2015 to June 2015. There is also aggregated data on taxis and other for-hire vehicles operating in NYC. The data is split into CSV files, which we know how to load properly into d3. We can create object structures around data, time, latitude, and longitude of the pickups.

Feature List

- Map that shows Uber's growth over time (Must Have)
- Choropleth map that show's Uber's pickups versus taxi pickups (Must Have)
- Area chart showing total rides (Must Have)
- Force chart showing how the different For Hire Vehicles interact with each other (Good-to-have)
- Line and Bar charts (Optional)

Team Roles

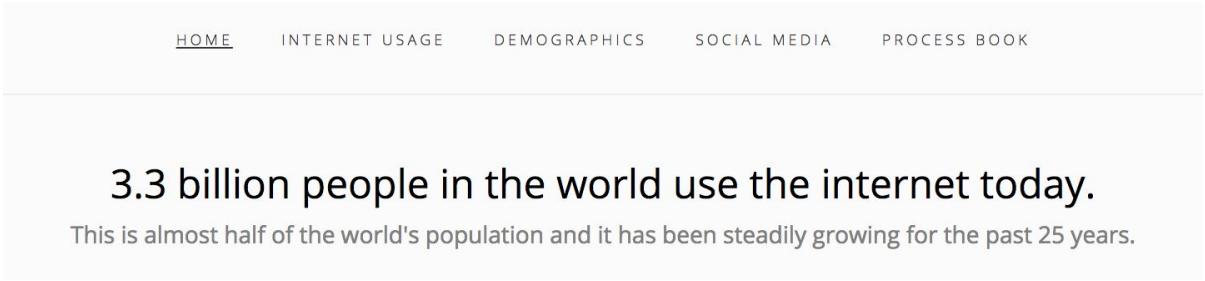
- Aashay will focus on project planning, orchestration, and targeting. He will also work on design and implementation.
- Patrick will pour his effort around data wrangling and the trickier implementations.
- Will will work on the implementation with Patrick and the website storytelling.

Timeline

Date	Work
11/14	Wrangle the data, work on mockups, and tackle 1 visualization
11/21	Nail down the two primary visualizations
11/28	Solidify all the implementations
12/5	Beautify and create storytelling

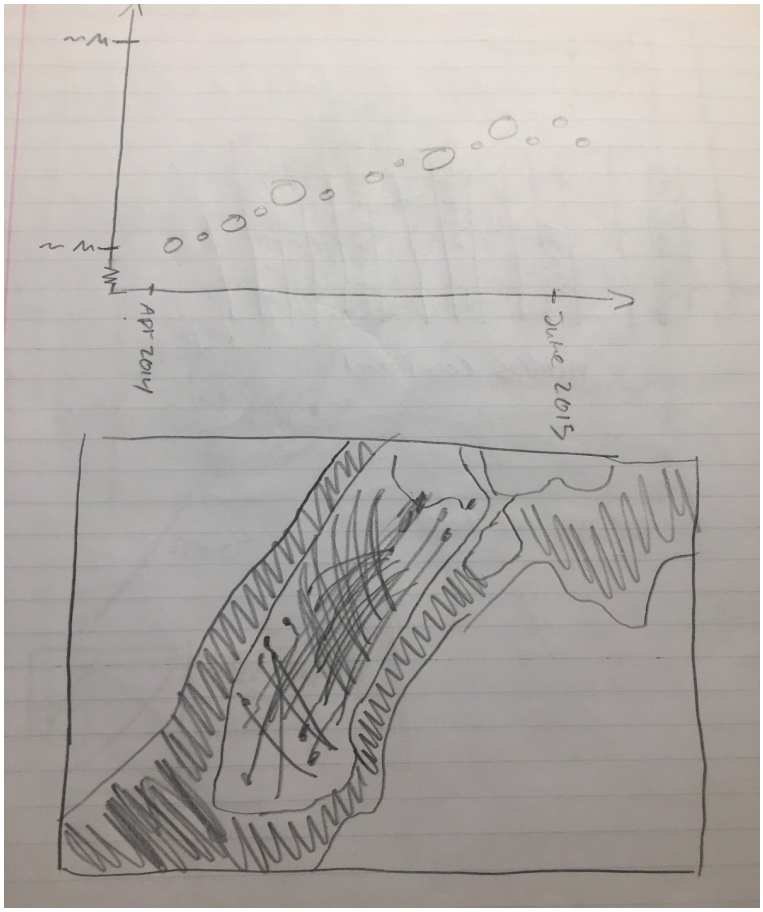
Our goals were slightly lofty and as we dove deeper into wrangling the data, we saw that some of our efforts would be challenged. Because there are so many Uber rides each month, the CSV files we were working with were huge and barely loading in the browser or crashing it completely. We also were not thinking completely through the story we wanted to tell and how all the different visualization components were going to fit together.

As we started to dive in, we thought that a horizontal navigation structure was best, with a top-level navigation that has each visualization on a different page. We were particularly inspired by a project we found from last year's Hall of Fame, which explored Internet usage across the globe. The project was done by Jacob Scherba, Maria Zlatkova, Aksel Reiten, and Magnus Moan. Here is the photo of their top-level navigation.



Using that as inspiration, here was our initial design for the top-level navigation.

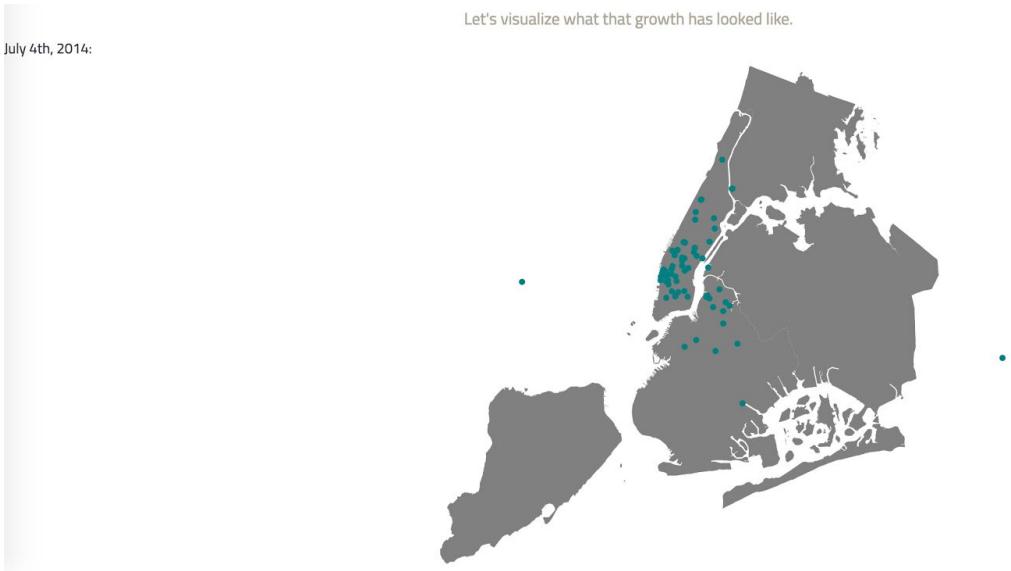
Another part we really struggled with early on was how to handle the data. Our initial idea for the main visualization was a map of New York City with a transition of dots across a month representing how many rides happened for that duration. Here is an initial sketch of what that might have looked like. The sketch is located at the bottom of the page.



Yet our big issue was that the data files were too large. We had to figure out how alter our story and use the data in a different way. We couldn't figure out how to show the whole month of rides or several months without crashing our browser. Thus, we decided to explore what a day looked in like New York City in terms of Uber pickups. We had a file that contained all the pickup data for a whole month. We used a python script our TF (Emily) helped us with to strip down the CSV to show data for a single day. That script can be found here:

https://github.com/aashaysanghvi/cs171-finalproject/blob/master/chopper.py

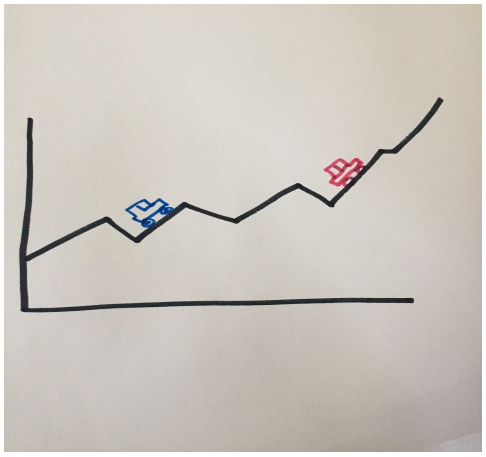
This way, we decided to alter the story of the first visualization to walk through what a typical day looked for Uber in New York City. At first, we wanted to do a special day, so we used July 4th, 2014. Here is what the draft looked like.

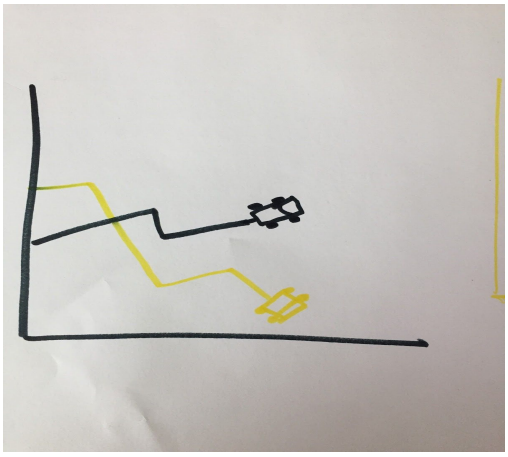


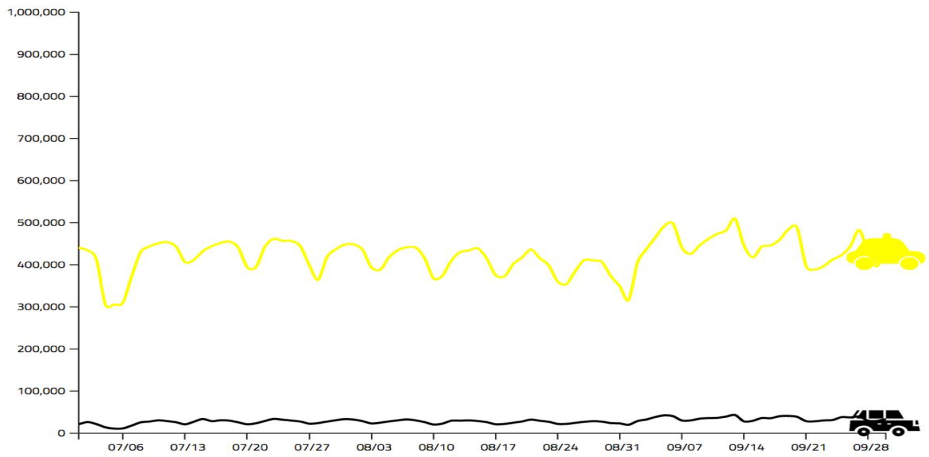
This is a screenshot of an animation that would play out for a whole day. We especially liked the way this visualization showed the fast rate of Uber Pick-Ups in Manhattan. We

later iterated on the visual design, but ended up keeping in line with this general concept (as you'll see in the redesign).

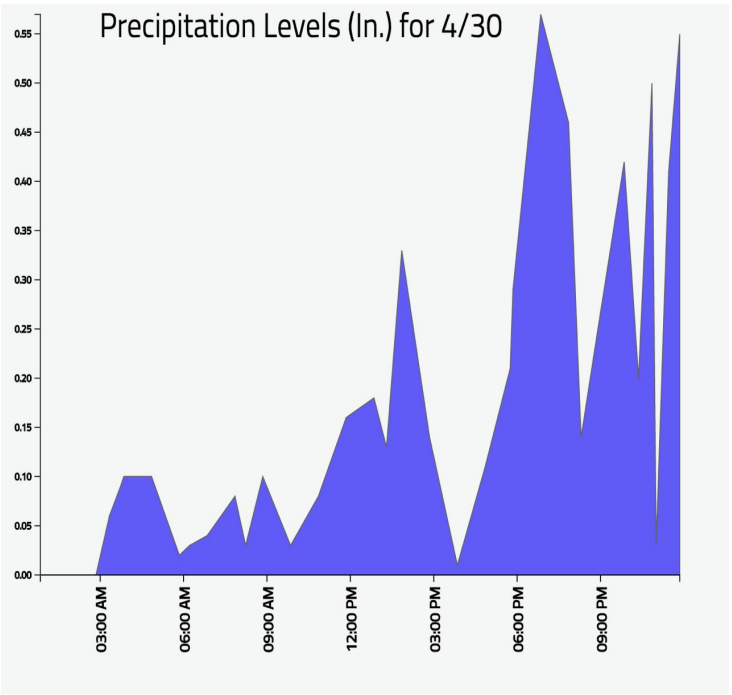
In order to illustrate the growth of Uber, we first went with a line chart of Uber's growth vs. the taxi's growth. While the animations of this visualization were interesting (each car moves across the screen as the lines are drawn), we realized that it did not adequately illustrate a significant growth difference between Uber and taxis. Thus, we knew had to dive further in the data to help solve this issue (using other vehicle data).

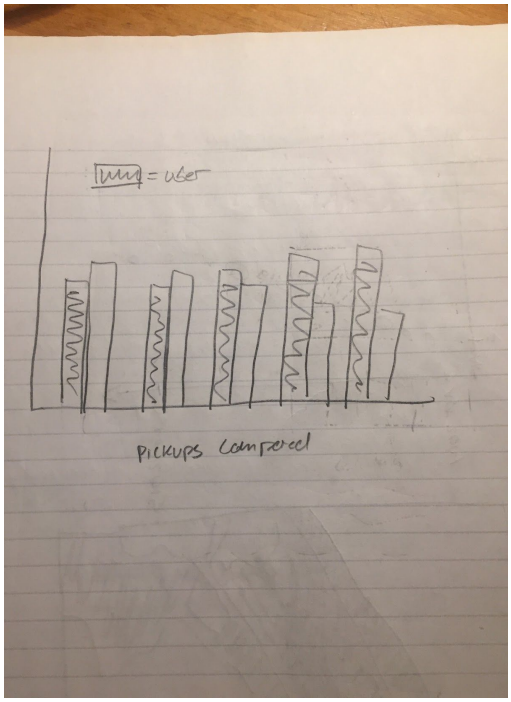






Our weather chart was still in its initial stages of being developed: at this point in v1 we only had an area chart of a single day of precipitation completed, using data from https://www.wunderground.com/history/airport/KNYC. We wanted to pair each area chart with a bar chart showing how precipitation levels show changes in total number of rides, but we would later realize that doing so would require careful thinking (as you'll see in our redesign).





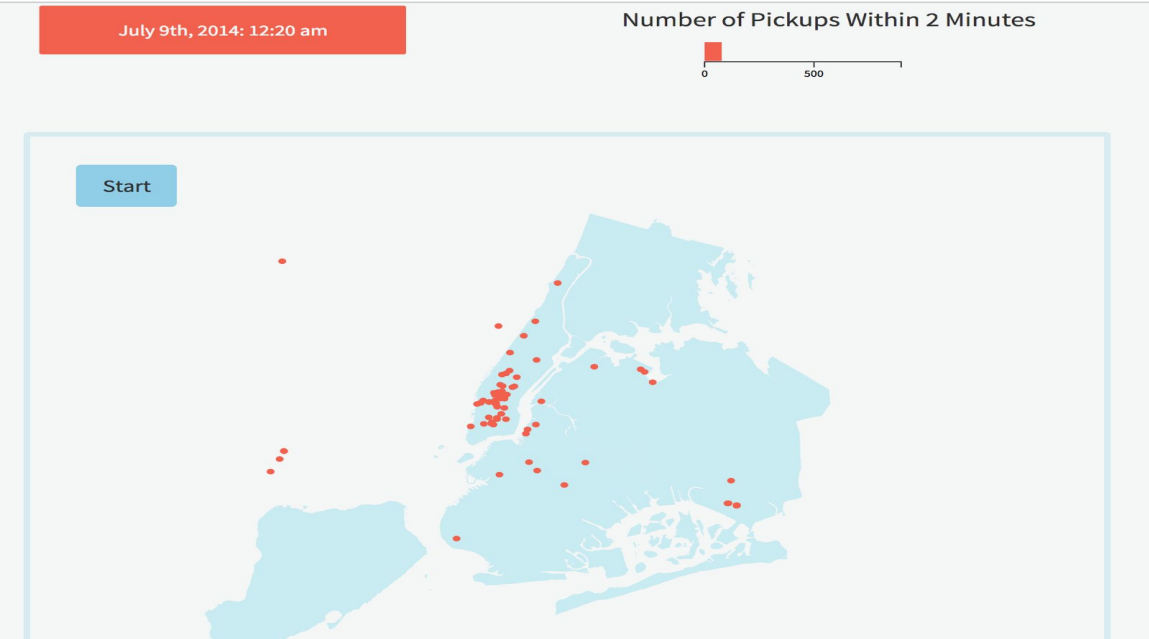
Another problem we found was that we had rough ideas of what we wanted to accomplish with our visualizations, but still no story to tie it together. Thus, we realized we needed to begin thinking more about what kind of text we could use to illustrate how these visualizations could be connected. Telling the story of this company would be no

easy task-- so we had to rethink which features we thought were most important to our message.

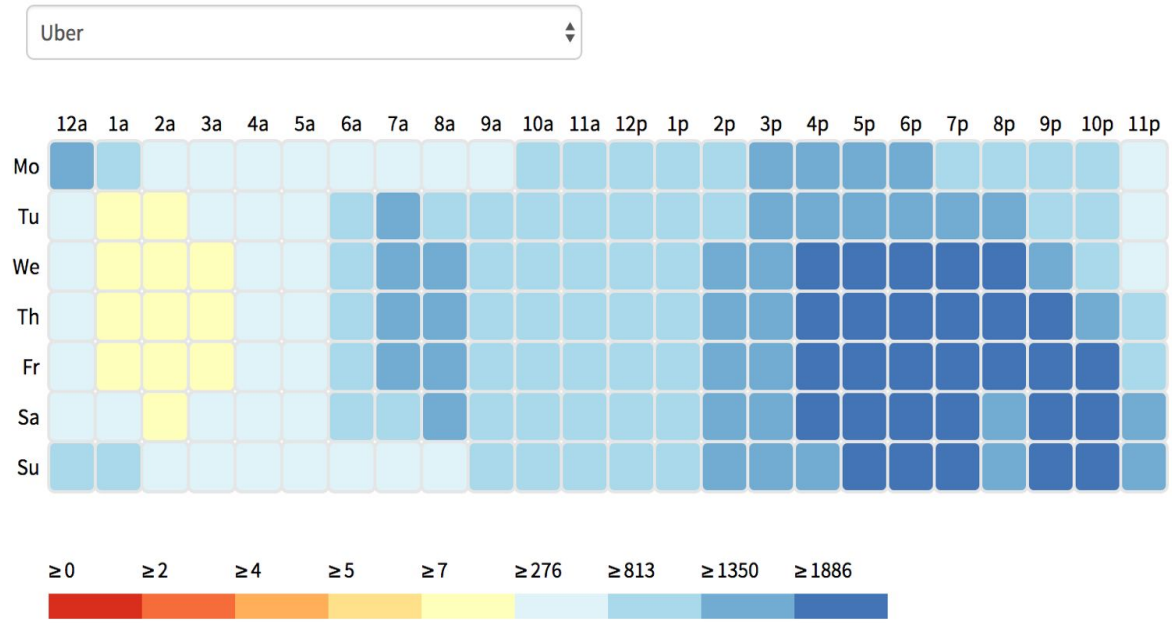
Project Re-Design:

After receiving feedback from our TF's and peers, we received helpful information on how to improve the clarity of our project's message. First and foremost, we decided that the entirety of the project should be placed in one html page, rather than breaking them apart into separate pages. In doing so, it became easier for our viewers to follow the sequence of our project, rather than having to skip around from page to page.

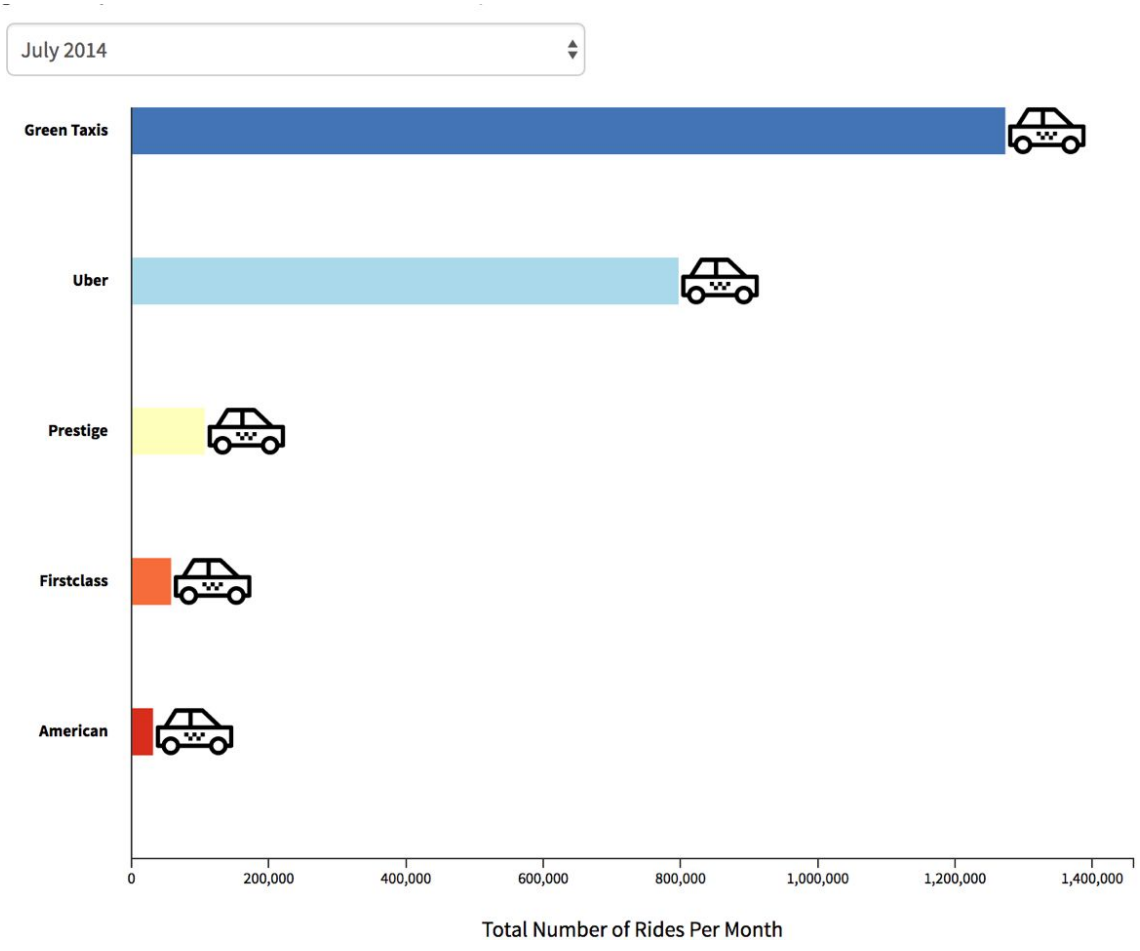
Secondly, we touched up the color scheme of our time lapse map and sped up the rate at which the time elapsed-- allowing users to view an entire day of Uber's all in the span of 72 seconds (previously this had taken multiple minutes, making it difficult for our viewers to understand the purpose of this visualization). Additionally, we added a small bar chart feature in the upper right-hand corner that demonstrated how many uber-picks per 2-minutes there were at that time. This visualization helped viewers get a better sense of peak hours, allowing them to better understand the business during rush hour.



Although this bar-chart did shed some light on the peak hours of the day, we still felt that we could do much more in highlighting the hourly differences. To do so, we included a heatmap of Uber pick-ups by the hour, along with similar heatmaps of Lyft and other car-services. We felt that the inclusion of these heat maps really helped add an extra temporal element to the geographical feature of the time lapse map.

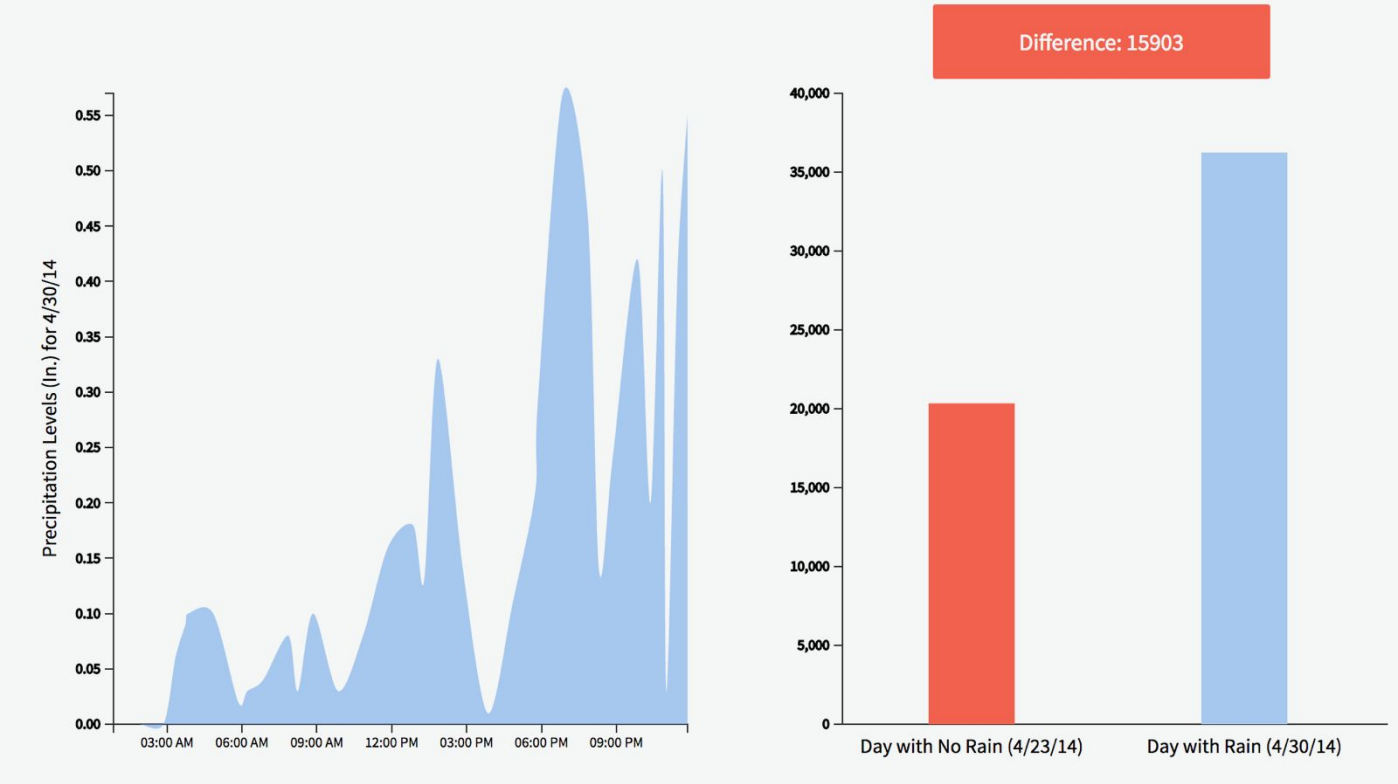


We also received helpful feedback on our line chart of Uber growth vs taxi growth. The relative difference between these two lines had seemed quite insignificant, so we overhauled our design and started from scratch. Our new visualization involved multiple bar chart figures, allowing us to add extra data while still keeping our original intent of including moving car icons in the visualization The end-result expanded on the growth of uber the past few years, while also demonstrating that other similar vehicular companies have not experienced such similar growth.



Finally, we also realized that we needed to change our way of thinking when it came to measuring the effect of weather on Uber. At first, we wanted to utilize area

charts in order demonstrate average precipitation levels for multiple days/years, but we realized that doing so would not accurately account for peak/quiet hours (for example: if it heavily rained at 3:30 AM, this data would not necessarily have a large effect on the number of rides because practically no one is using Uber at 3:30 AM). Thus, we realized we needed to take a snapshot of the effect of rain over the course of one day-- and so we decided to compare the number of rides on a rainy day (high precipitation levels) vs. the number of rides on a dry day (no precipitation). The end result was an area chart that was interactive with two bar charts-- demonstrating a large difference in rides during high precipitation hours.



Wrapping Up

Based off feedback we received at the project demos and from out TF Ronell, we had a couple of minor adjustments we still needed to make. We needed one concise, clear story for our four visualizations. The storyline would walk through a typical day and typical week for Uber in New York, answering questions like when is the service busiest. It would then transition to exploring if Uber seems busier than other services, comparing their average pickups and pickups over time. Is Uber growing and having an effect on them? The last piece will explore the impact of weather, an additional factor on Uber pickups. After adding some colorful commentary and analysis to each of these visualizations, we finally felt that we had accomplished our project's goal.

General Summary

Overview and Motivation

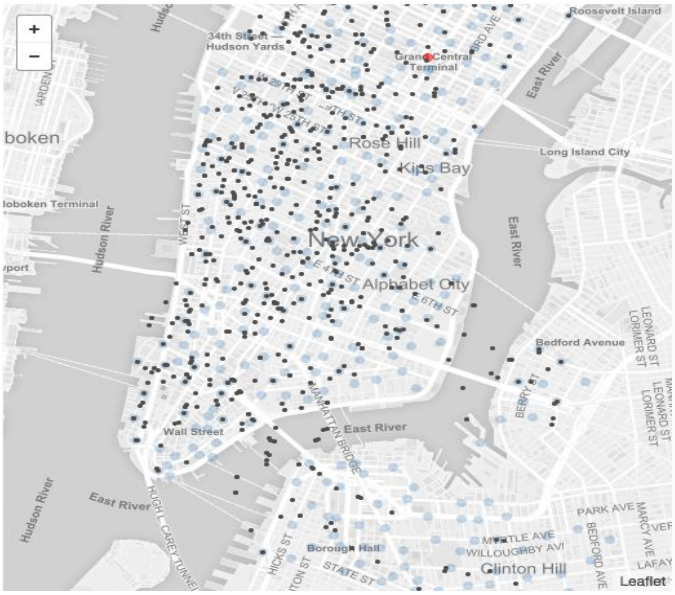
Uber launched with in 2010 with a simple, yet revolutionary premise. Tap a button and within minutes, a driver would arrive to take you to your destination. Over the past couple years, Uber has grown at an extraordinary pace, serving millions of rides per day across the globe and defining new methods of transportation across cities. The purpose behind our project is to understand what impact Uber has had on transportation in New York City, as well as how specifically the company has spread its reach through the city.
Our goals include educating our audience about Uber's growth in New York City through 2014 and 2015, visualizing what impact Uber has had on the taxi cab system in the city, as well as understanding what neighborhoods of New York City are most affected by the growth of this new transportation juggernaut. Tasks for our team include wrangling,

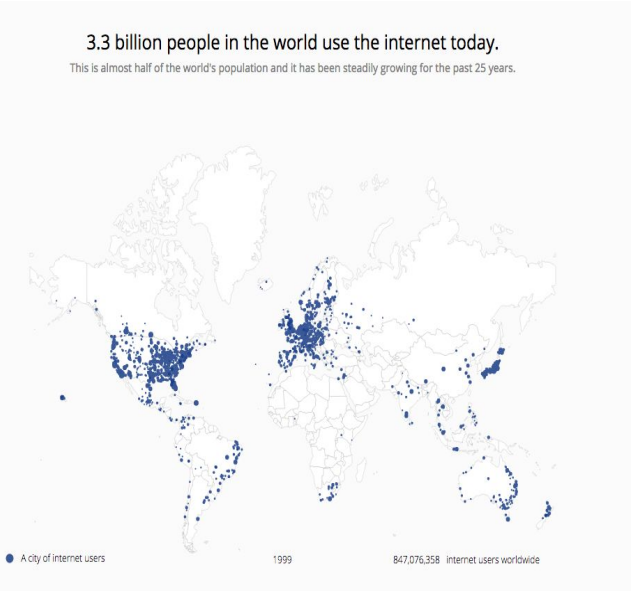
cleaning, and organizing the Uber pickup data and other for-hire vehicle data. We need to make sure all the data are consistent and in the same structures.

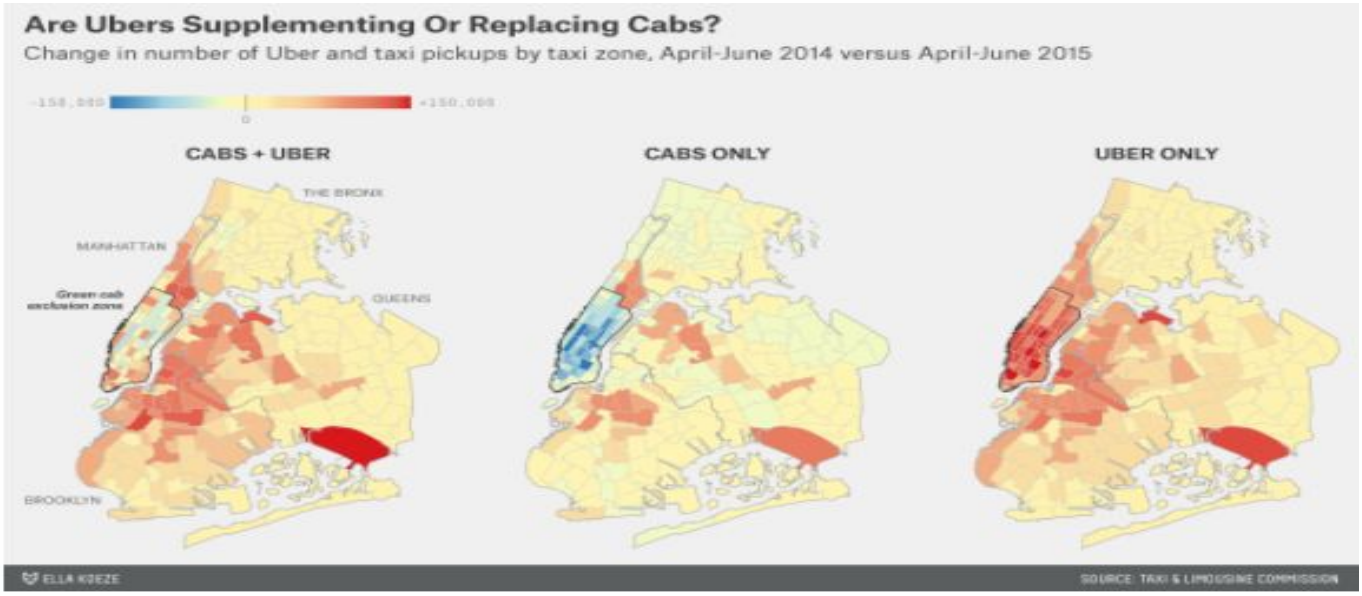
Other tasks include deciding storytelling methods that will be effective, as well as designing mockups and storyboards of our visualizations. We also need to touch up on our implementations of the visualizations and make sure that everything looks clean and polished at the end.

Related Work

One of our major influences for our project was the CitiCycle project from 2015. We really enjoyed the way they were able to depict the traffic of the cyclists moving through the city-- while also identifying the prominence of each type of cyclist through the use of barcharts. Unfortunately, our data was limited to pick-up and drop-off location, so tracking the actual path of each Uber was not within our capabilities. However, we still were able to show peak hours-- similar to the way they demonstrated peak hours for cyclists. We also enjoyed the Global Network project from 2016-- specifically how the internet users populate the map at the beginning of the home screen. We also took pages from various fivethirtyeight articles that dealt with the growth of Uber.







Questions

Initially, we wanted our project to analyze the impact Uber has had on the transportation industry. We wanted to answer large scale questions, illustrating how the growth of Uber has shaped the way people travel in New York City. However, upon further inspection of the data, we decided to move away from the macro, and zoom in on what a typical day of Uber in New York looks like. By moving to a much smaller scale of information, we were able to illustrate an even larger picture of different trends and influences of Uber in New York City.

Data

Our primary data source is this GitHub repository from FiveThirtyEight: https://github.com/fivethirtyeight/uber-tlc-foil-response. Uber is a private company, so normally they wouldn't release their data, but they gave this up in a freedom of information request FiveThirtyEight submitted to NYC's Taxi and Limousine Commission. According to FiveThirtyEight, the data includes information on over 4.5 million Uber pickups in New York City from April 2014 to September 2014, as well as 14.3 million pickups from January 2015 to June 2015. There is also aggregated data on taxis and other for-hire vehicles operating in NYC. We created a script known as "chopper.py" in order to cut the data and save the information for Uber rides into individual csv files for a single day. Once we had the CSV files, we load them properly into d3 and clean up the data. We can create object structures around data, time, latitude, and longitude of the pickups. We used https://www.wunderground.com/history/airport/KNYC, to receive our weather data. We were able to download the data for each day as csv files, and then we scraped the data and transformed strings into numbers.

Exploratory Data Analysis

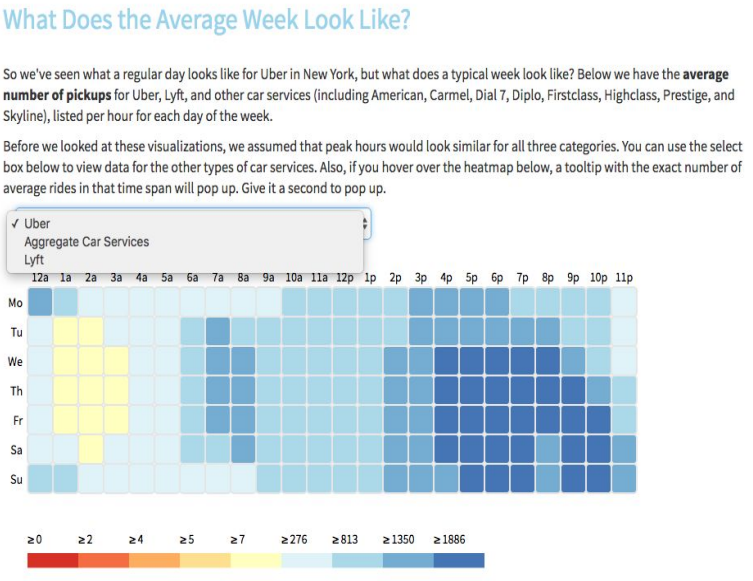
In order to gain a basic understanding of our data, we used bar charts to try and discover common trends in the data. This eventually led to our discovery of the weather factor as we looked at different bar charts of Uber rides on rainy days. We also used Bar Charts to split the data up by hour, which helped us realize that certain peak hours did exist for Uber.

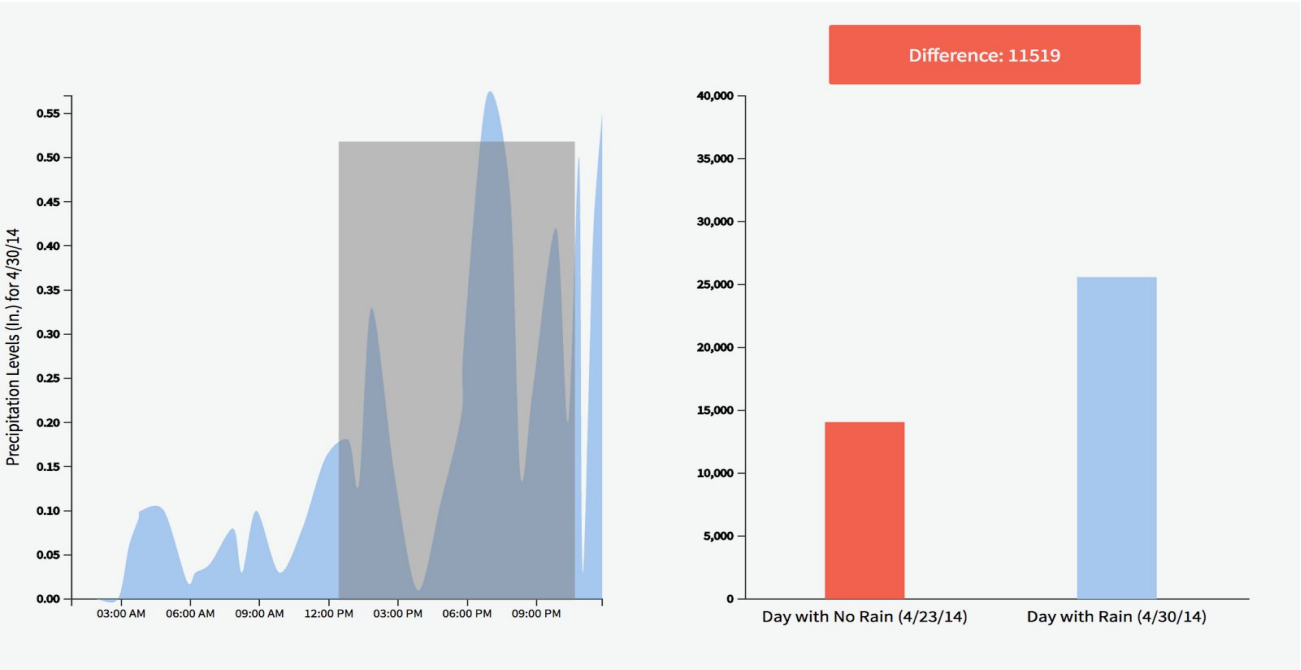
Design Evolution

The evolution of our designs are clearly illustrated by our Re-Design chapter (see above). However, we would like to include a little bit more how we considered design principles when finishing our final designs. Our project centered around a theme of accessibility and simplicity-- as such we didn't want to inundate our viewers with a wide array of interactions and tools that they wouldn't know how to use. To do this, we made sure to keep our visualizations neat and organized, utilizing proper spacing and very simple channels of color (heatmap), length (bar-charts), and position (time lapse map). Furthermore, we made sure to follow Tufte's principles of design, keeping a small data-ink ratio to keep in line with our neat image.

Implementation

The intent of our visualizations if to take our users on an easily navigated story-line. We felt that past projects inundated viewers with information, leaving them hopeless to draw any sort of conclusions from the visualizations. As such, we make sure that our users are aware of the many features of our visualizations, by pointing out necessary steps and conclusions in our commentaries. Examples of our user-interactions include a pause/play button, selection boxes, tooltips, and a brushing tool.





Evaluation

From our visualizations we obtained a snapshot of not only the macro-trends of the transportation industry (the growth of Uber compared to other car services, the differing hours that each company dominates) but also the smaller influences on a day-to-day basis (weather, location). While we did obtain a clear picture of how Uber operates--there is still so much more room for research on this Company; future visualizations could potentially include how income affects the use of Uber/ride-sharing and whether or not this is geographically related. Nevertheless, we were quite proud to create a specific set of well-working individual visualizations that come together to illustrate a larger message about the Uber community.