Katherine Loboda & Samuel Stone
"The CCRB"
CS171 Final Project Process Book

Week 1: Project Proposal

Project Title: The Disparate Impact of the NYPD's Stop and Frisk Program

Abstract

The New York Police Department ("NYPD") is tasked with patrolling a city with over 8 million residents and 50 million annual tourists.[1] As part of the department's crime-fighting initiatives, officers rely on the Stop, Question, and Frisk ("SQF") program to detect and remove weapons and contraband from city streets. Though the NYPD has conducted SQF operations since the 1990s, the practice became far more widespread under the mayorship of Michael Bloomberg that began in 2002. By 2011, there were over 650,000 NYPD stop-and-frisks across the city. Since these programs are mainly implemented in New York's most dangerous neighborhoods, Hispanic and Black civilians are stopped at a disparate rate (over 52% of SQF stops involved black suspects, and 31% involved Hispanic suspects).[2] Though the program has been curtailed under Mayor Bill de Blasio, SQF remains part of NYPD policing protocols.

Though the racial biases of SQF are well known, we believe effective visualizations of SQF data will help quantify the scale of the program and illuminate how stark the racial disparities in NYPD stops are for minority communities. The NYPD has a meticulous database of thousands of SQF records for each year (including geographic information, time of day, characteristics of suspects, follow-up arrest information, etc.), which will serve as the data underlying our visualizations. These datasets are publicly available at the following link, part of the New York City government site: http://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page.

---

[1] The numbers add up to one fact: cops are a blessing to NYC," New York Post, http://nypost.com/2014/12/31/the-numbers-add-up-to-one-fact-cops-are-a-blessing-to-nyc/.

[2] Here's what you need to know about stop and frisk - and why the courts shut it down," Washington Post, https://www.washingtonpost.com/news/wonk/wp/2013/08/13/heres-what-you-need- to-know-about-stop-and-frisk-and-why-the-courts-shut-it-down/.

Week 2: Project Plan

Names of Contributors to Project Plan: Katherine Loboda & Samuel Stone

Goals and Tasks for Final Project:

Goals:

1. Present users with a compelling and interesting website about the NYPD's stop-and-frisk program that convinces them of the importance of studying the program and hooks them into dedicating the time to navigating through our site and learning about the program and its implications.
2. Educate users on the history of the NYPD's stop-and-frisk program, and contextualize the program in the broader landscape of law enforcement in New York City and around the country.
3. Give users the ability to see how stop-and-frisk was actually conducted in New York City and allow them to independently elucidate the harmful racial biases that arose as a result of its implementation.
4. Illuminate other less well-known disparities involved in the program, e.g. geographic or socioeconomic ones, so users can get a broader sense of the program than is typically involved in mainstream media narratives.
5. Engage users enough for them to conduct further research of their own into the program, and inspire viewers of our site to think deeply about the program.

Tasks for Each Goal [Based on cs171.org, which asks that we list in detail which data and visual manipulations we will implement to support our outlined goals]:

- Goals 1 & 2: We hook users into our website by providing them with a history of the Stop & Frisk program by showing them a timeline of how the number of stops by police officers in New York City have changed over time with related news articles for key years of the program implementation. To do this we will wrangle the data to only find the number of stops in each year, and then we will allow the user to click through a timeline that shows a summary with "annotated" related articles for each period of time (as per martini-glass storytelling), and the "guiding" visual taxonomy. By clicking through this timeline the user will hopefully get interested about how much has been written on the topic and how many opinions there are about this program. Additionally, the summaries that we provide and key articles chosen will hopefully elucidate the key biases and opinions that have been dispelled by the media.
- Goal 3: We will provide users with one main visualization that clearly highlights the racial disparities of the program and understand better how the program was

actually implemented on the streets of New York. This visualization will be a tree map which allows users to filter by certain categories of data (e.g. race of person stopped, type of contraband found, if any) and will allow users to quickly see how many stops were of each type. This visualization will hopefully eventually also allow users to select certain data points, namely individual stops, and learn more information about how the stop occurred (e.g. what was found on the suspect's person). There will also be an element of guiding users through the data, showing them the some of the filtered categories that we deem most notable.

- Goal 4: To illuminate other less well-known disparities involved in Stop & Frisk we will provide the user with a choropleth map of New York broken down by precinct. The choropleth will be able to show variables such as number of stops and percentage of stops that led to weapon or drug discovery, and thus we will "filter" the data based on the user's selection and also calculate the percentages. The user will also be able to select the range of time (can be just one year) they are most interested in exploring, which will be another level of data filtering that the visualization will accomplish. Finally, the visualization will allow the user to "select" their precinct of interest and by hovering over it they will learn more about the demographics and policing done in that precinct. We hope that this interactive and user-driven visualization, that incorporates data selection and filtering, will allow them to make their own opinions about the program and validate or dispute the biases in the media that we previously showed them.
- Goal 5: To actively engage users we will create a section at the end that displays qualitative information about recent political events relating to stop and frisk, and encourages active participation in the political process that leads to overhauls of police policies like this program. We hope that this qualitative information, when taken alongside the plethora of quantitative information we will visualize, may help inspire individuals to take future action relating to programs like this one.

Description of Data:

We have NYPD provided data on stop-and-frisk stops available from 2003 to 2016. We have not decided whether we will ultimately constrain the number of years (e.g. to one year or a limited set of a few years), and we expect to make this decision largely on an as-needed basis. For example, if we see that adding multiple years creates so much computational strain that our visualizations or website renders unacceptably slowly, we will likely just constrain our dataset to include a slightly smaller number of years.

Each year's dataset contains thousands of rows (corresponding to individual stops) and roughly 100 columns, though it varies slightly from year to year as the NYPD added and

subtracted certain variables. The dataset also has a convenient data dictionary that provides information for otherwise obscure variables, as shown in Exhibit 1.

List of Variables

variable	description	values
year	year of stop	yyyy
pct	precinct of stop	1 through 123
ser_num	UF-250 serial number	nnn
datestop	date of stop	mmddyyyy
timestop	time of stop	hhmm
city	location of stop city	1 - Manhattan 2 - Brooklyn 3 - Bronx 4 - Queens 5 - Staten Island
sex	suspect's sex	0 - female 1 - male
race	suspect's race	1 - black 2 - black Hispanic 3 - white Hispanic 4 - white 5 - Asian/Pacific Islander 6 - Am. Indian/Native

Exhibit 1. Excerpt from NYPD data dictionary

The variables in the datasets generally fall into three main categories: demographic, stop-related, and geographic. We believe this mix of different types of variables will lead to a more interesting set of visualizations and allow us to tell a more compelling story. Obviously, the most famous “story” associated with the stop-and-frisk program is the racial disparity associated with the programs (a quick glance indicates that over 75% of stops in New York were of Black or Hispanic individuals, a clear overrepresentation). But we have data available to tell many other interesting, and less commonly reported, stories. For example, we have geographic coordinates, evidently to the nearest square foot, of every stop, which could be used to show geographic disparities in the program. We also have data on whether different types of physical force were used, or duration of the stop, so we could see whether certain types of individuals are more likely to end up being involved in a “violent” or otherwise intense stop.

There will doubtless be plenty of other questions that users hope to answer using this dataset, so we will continue to fine-tune our approach to analyzing the data over the coming weeks.

Sketches of Visualization Ideas:

Visualization #1: This visualization illustrates the percentage of stops in which the police found different items on their defendants (nothing, drugs, weapons, or both), with each category broken down by race. This way, individuals can quickly compare across races whether there was a different success rate in finding illegal items. Our hypothesis is that there will be very little difference, suggesting that stopping many black civilians was not only biased, but also an ineffective strategy.

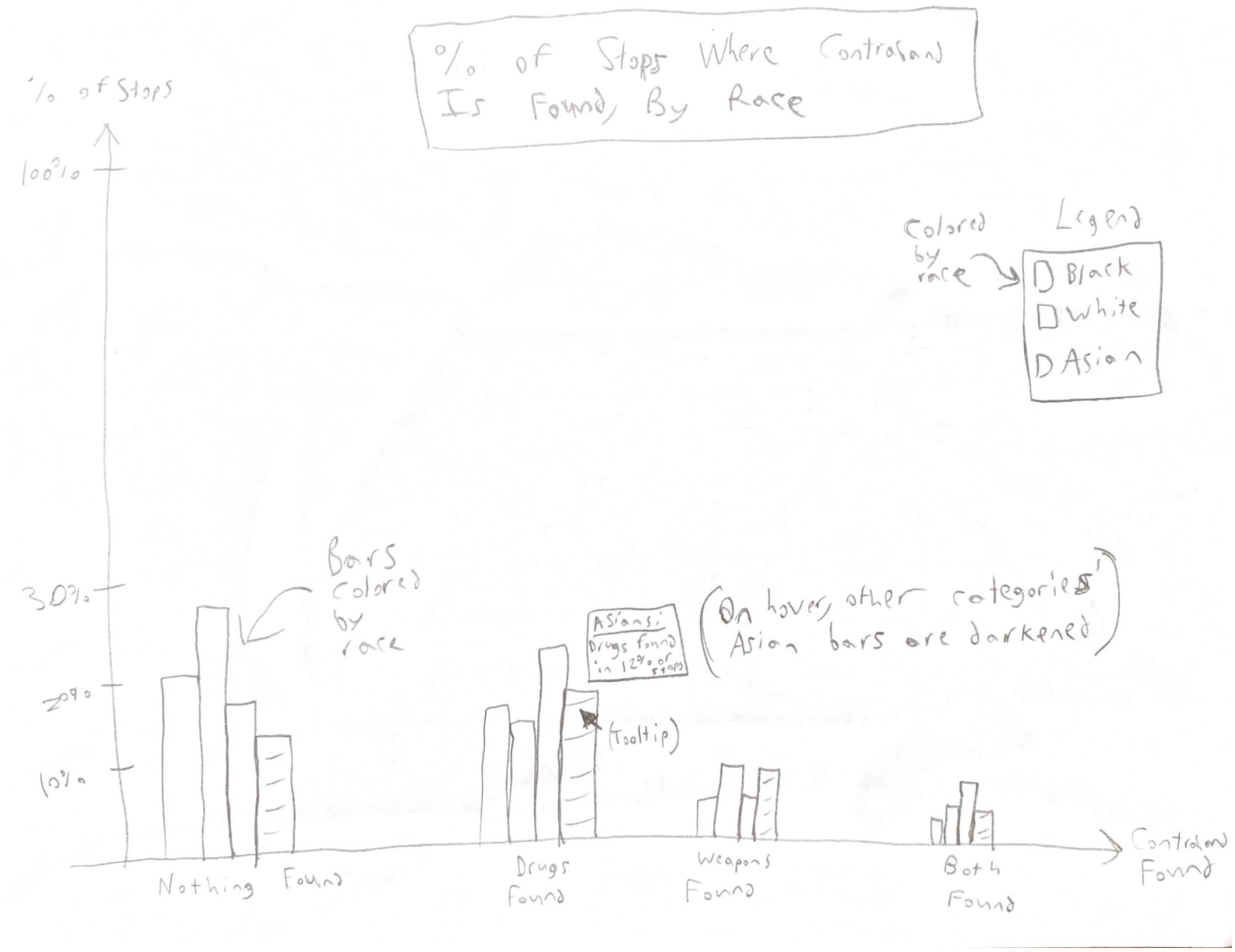


Exhibit 2. Percentage of Stops where Contraband Is Found, by Race

Visualization #2: This visualization will provide users the chance to quickly and intuitively see “nested” breakdowns of all stops, allowing them to drill down on which characteristics they care about. For example, they can begin by specifying that they only want to see stops in which drugs were found. Then, they can specify they only want to see stops where drugs were found *and* the stopped individual was black. Of course, can keep adding all the criteria this way until they have exhausted our provided categories. We could also incorporate tooltips into this view that show summary information. Note: We were inspired for this visualization by FiveThirtyEight.

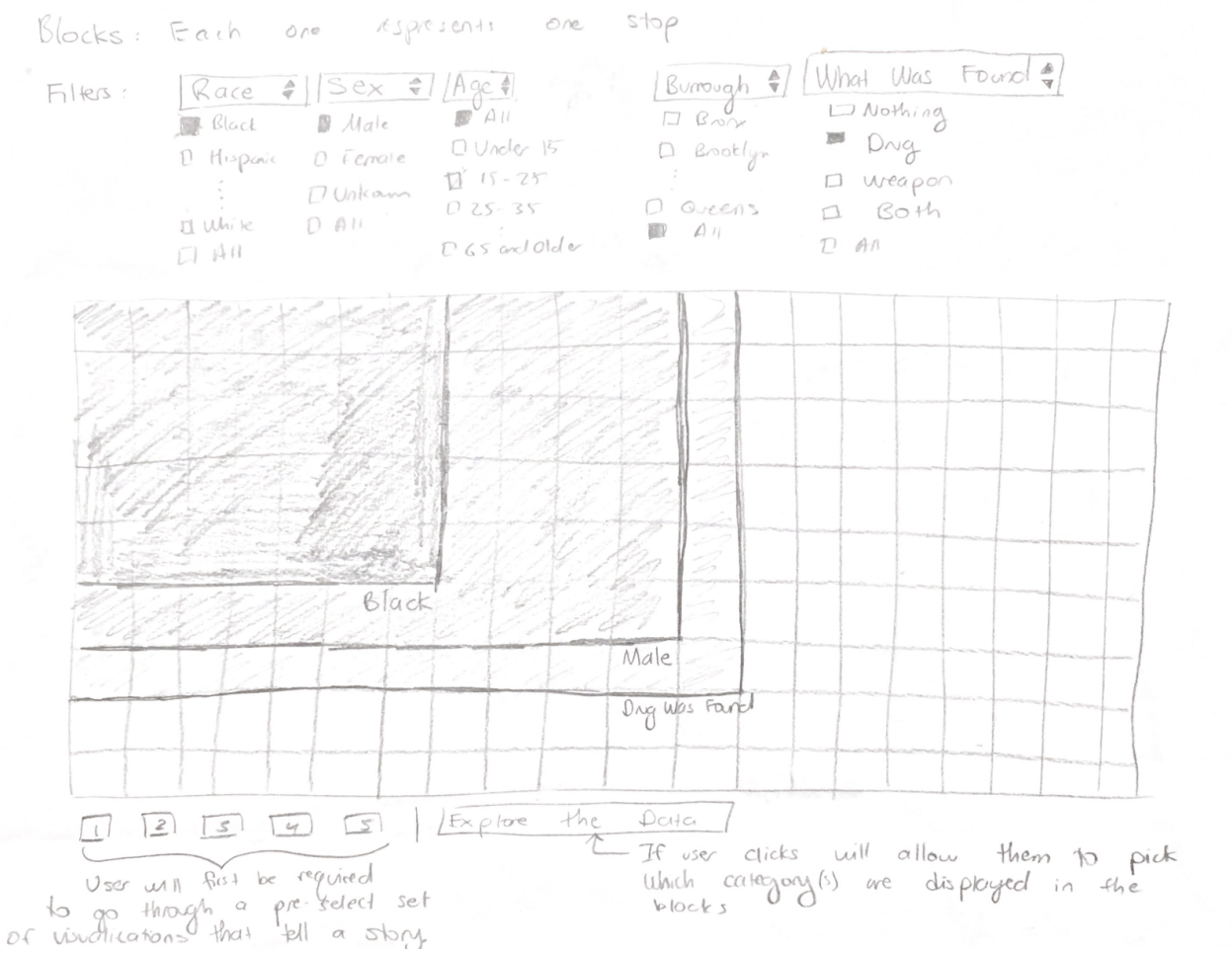


Exhibit 3. Treemap of Stops by Race, Sex, Age, etc.

Visualization #3: This map visualization, which will likely ultimately serve as the centerpiece of our website, depicts a choropleth in which each police precinct in the five boroughs of New York City is colored according to a given type of data the user specifies. We will likely link this visualization to several other summary data visualizations like the line graphs in this sketch, though the exact linkage we will establish remains TBD for now. We would allow users to select which type of data to visualize (e.g. total number of stops vs. % of stops that yield weapons) and specify a date range across which to display information.

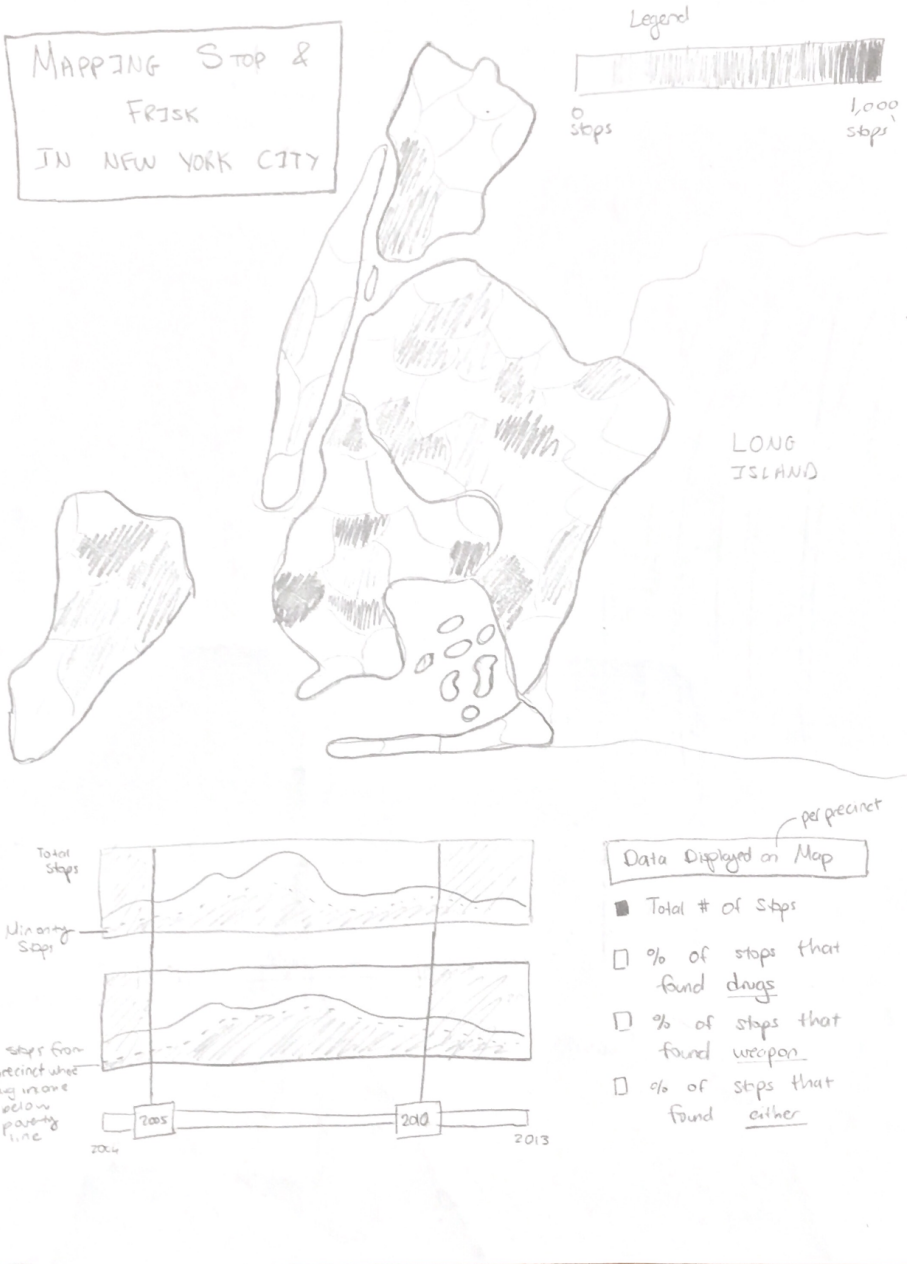


Exhibit 4. Mapping Stop and Frisk in New York City

Visualization #4: This visualization, which we hope to incorporate as a way to emphasize the narrative we want to emphasize, will allow users to follow a timeline of key historical events relating to the stop-and-frisk program over time and overlay this information with data on the number of stops that were actually conducted. In other words, users can see on which date the court system appointed a monitor to oversee the NYPD, and also see whether the program saw a decrease in stops accordingly. This visualization would likely serve as our most site-directed visualization, i.e. not allowing a great deal of user customization, because it seems more effective to us to convey these historical events in a very prescribed manner.

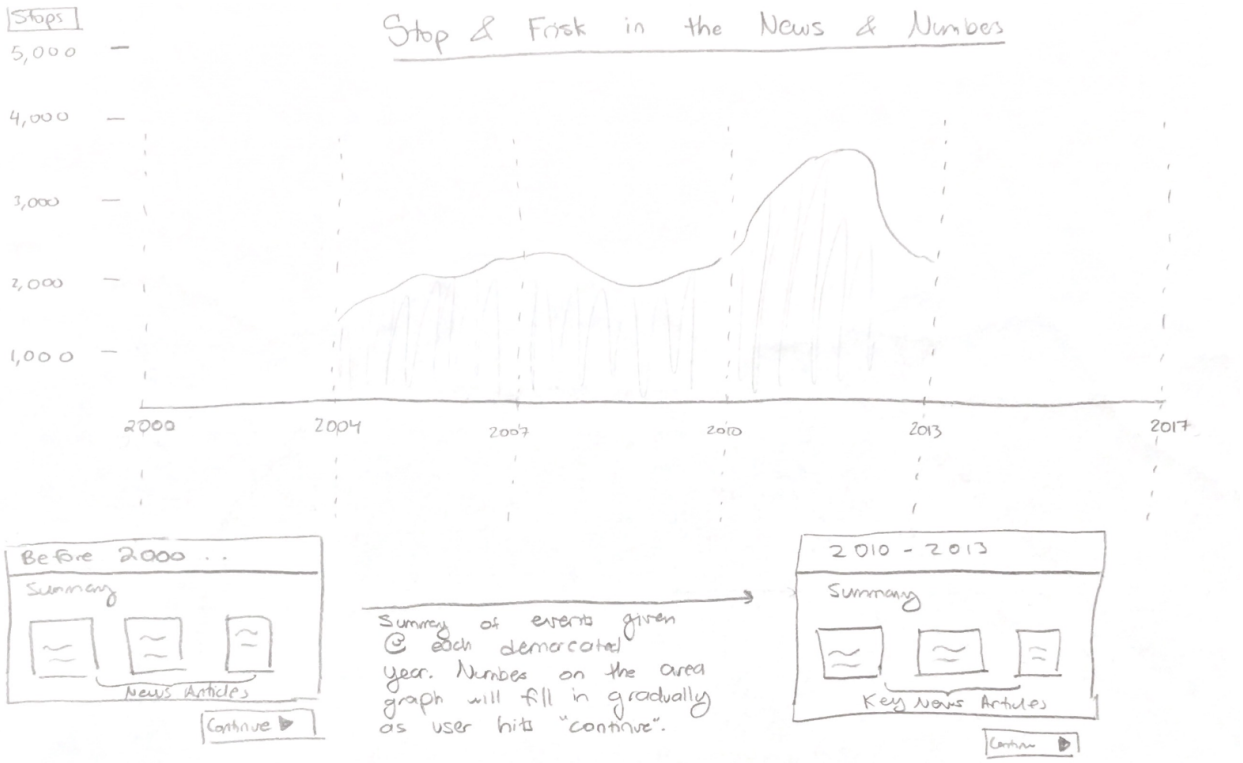


Exhibit 5. Stop and Frisk in the News and Numbers

Sketch of Interaction Storyboard: Below, see our interaction storyboard. The primary interactions shown are:

1. (Top-left frame): By hovering over a precinct, the user can view information about it, e.g. the racial composition of people living in that precinct and information on the variables that are not currently selected to be visualized in the choropleth.
2. (Top-right frame): By selecting different data to be viewed, our choropleth map updates to show whatever the user wants to see, and updates its scale as needed (from divergent to convergent, etc.)
3. (Bottom frame): By dragging the date range filter, the map will add or remove data points that are from earlier or later than the range that a user specifies

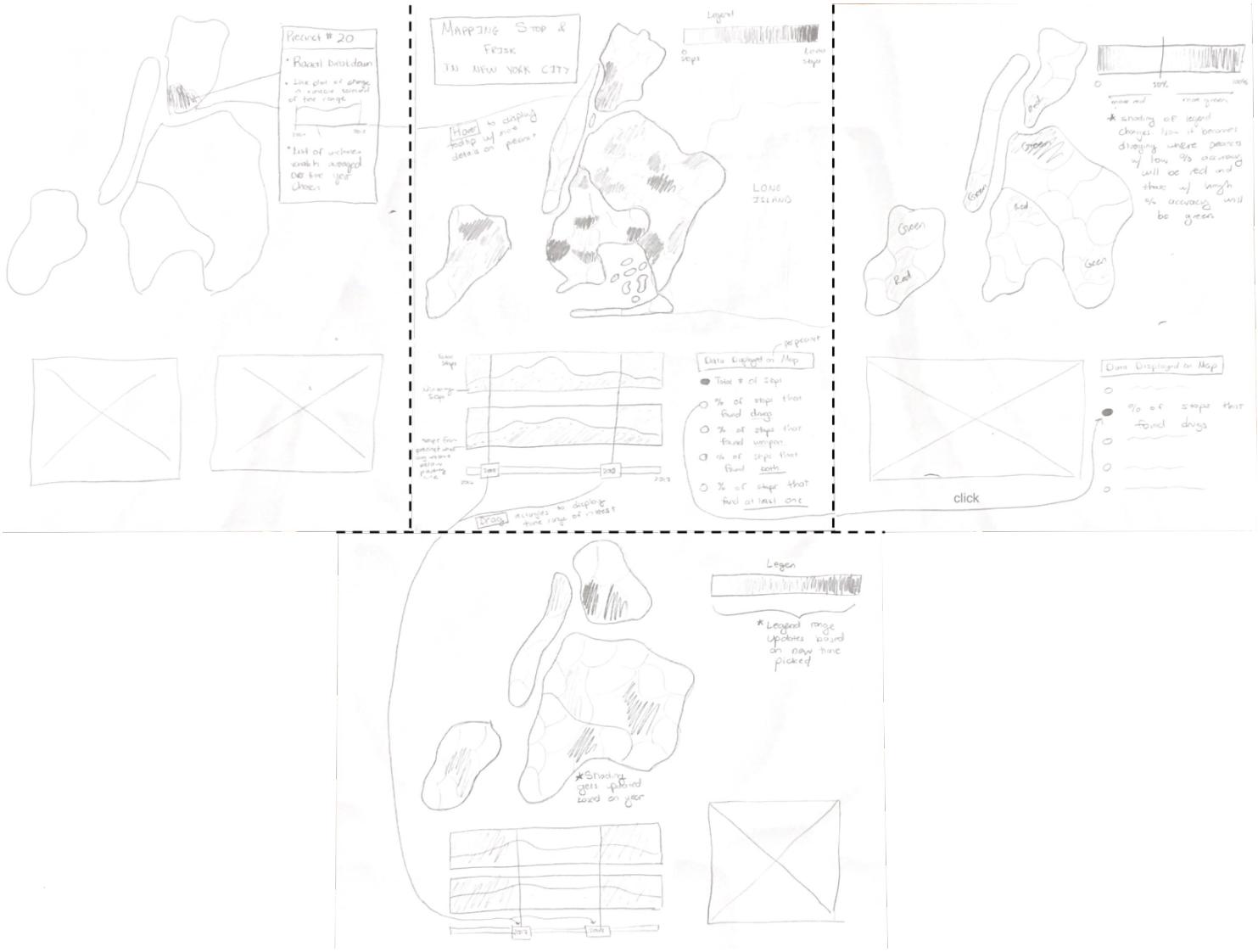


Exhibit 6. Proposed Interaction Storyboard for Choropleth Visualization

Sketch of Webpage Layout / Storytelling Approach:

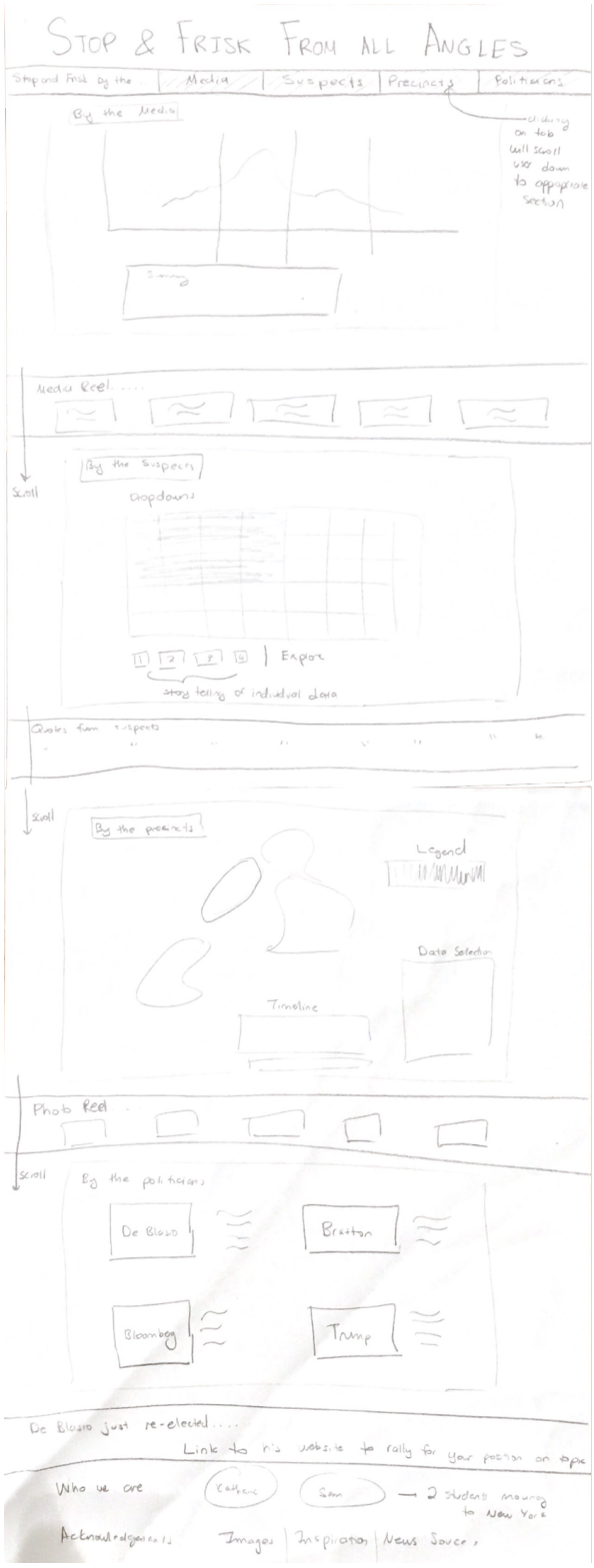


Exhibit 7. Proposed Website Layout

Project Timeline:

November 13: Prototype v1 due (External class deadline)

- Finalize our selection of data from the NYPD's stop and frisk datasets, and determine which years we want to include and how much cleaning we want to implement (versus potentially dropping certain columns to minimize workload)
- Complete a partial implementation of our news clipping / total stops over time visualization, and gather the most relevant and popular historical news articles to build a corpus of information that we will provide to users of our site
- We will additionally aim to complete, or at least mostly implement, the static bar graph visualization that shows the "success rate" of stops by race
- Finally, we will have a skeleton of our website prepared by this point that will follow our storytelling structure (likely "martini glass") and begin crafting the key messages we want to highlight through repetition and other approaches for emphasizing certain points

November 20: Prototype v1.5 due (Internal group deadline)

- Continue refining / complete the visualizations we implemented for prototype v1, and begin gathering feedback from friends, roommates, etc. about our visualizations through observational "think aloud" studies
- Implement our map visualization, with a primary focus on the "must-have" features of that visualization that we describe below

November 27: Prototype v2 due (External class deadline)

- Tinker with our website structure, i.e. fonts, images, and background colors to ensure that the overall feel of the website is matching the type of story we want to tell and impressions we want to convey to users
- Implement our nested treemap visualization, with a focus on the "must-have" features of that visualization that we describe below

December 4: Prototype v2.5 due (Internal group deadline)

- Conduct another round of observational "think aloud" studies and tweak (or entirely redesign, as necessary) visualizations to improve user comprehension
- Fine-tune the design/CSS of our website based on user feedback from the observational "think aloud" studies.
- Smooth out any remaining issues with previously implemented visualizations and implement the remaining good-to-have and/or optional features of the treemap / map visualizations, and ensure that linked visualizations are robust to various changes and unexpected user behaviors

December 11: Final Project due (External class deadline)

- Based on Hofstadter's Law ("It always takes longer than you expect, even when you take into account Hofstadter's Law"), we will leave the week from 12/4-12/11 available for any unexpected delays in completing prior goals rather than specifically leaving new tasks for this week
- Assuming we stay on track, this week will be a chance to relax and do final debugging; if not, we will use this week to correct previous errors in timing and complete any last-minute implementations or updates that are urgently required

Feature List:

Must-Have:

- Overview of stop-and-frisk with necessary background to get users settled into the topic smoothly when they open our site
- Map visualization that shows which New York City precincts most often conduct stops via stop-and-frisk, with a linked view to a visualization like the line chart shown in our Visualization #3
- A visualization like our Visualization #4 that shows articles describing key events in stop-and-frisk history along with quantitative data information corresponding to time periods in which these events occurred
- A visualization like our Visualization #2 that shows a treemap of stop-and-frisk incidents broken down into several nested categories and subcategories to allow users to see a variety of ways in which these stops can be broken down and analyzed accordingly

Good-to-Have:

- A visualization like our Visualization #1 in which users can see, at a glance, the percentages of stops that are successful (according to several metrics) and how these success rates vary by race of the stopped individual
- For our map visualization (Vis #3), provide users the ability to show choropleths for several datatypes, like % of stops that were successful in each precinct, in addition to total number of stops by precinct
- Add tooltips to our map visualization (Vis #3) so users can find more relevant information about each precinct on the fly by hovering over these precincts with their cursors
- For our treemap visualization (Vis #2), provide both a site-directed approach (where certain trends are highlighted) as well as allowing users to manually select the data groupings they want to see visualized

Optional:

- Integrate news stories elsewhere in our visualizations to provide a more comprehensive overview of the program that blends mainstream news media coverage with our interactive user-directed visualizations
- For our map visualization (Vis #3), scrape additional data beyond what the NYPD provides for each precinct that could be interesting, like socioeconomic breakdowns for each precinct
- For our treemap visualization (Vis #2), implement tooltips that allow users to get information on every single datapoint depicted with a box (e.g. race and age of user, borough of stop, etc.)
- Implement some sort of "poll" whereby we can measure how user's attitudes toward the stop-and-frisk program were changed by our website; then after users have voted, display information on how all prior users voted

Description of Team Roles:

As described in our team expectation agreement, we will aim to split all tasks as evenly as possible. That being said, we each have slightly different individual strengths, so we will leverage those accordingly. As we mention in our agreement, Sam will be the primary point-person for the "target" and "implementation" aspects of our final project, while Katherine will lead our "data wrangling" and "design" portion, with both members splitting work on the "evaluation" phase.

While Katherine's relative penchant for sketching and Sam's relative penchant for code debugging will certainly lead to some specialization, we both plan to regularly contribute to all areas of our project.

Week 3: Prototype v1

Names of Contributors to Prototype v1: Katherine Loboda & Samuel Stone

Data Scraping and Cleaning:

As we mentioned earlier in our process book, the stop and frisk datasets that we had available to us contained data from ~15 years of the program, with each year's dataset containing tens or hundreds of thousands of entries (earlier years had closer to a million entries, since the program has become less widespread in the past few years). Further, each dataset had over 100 features available for each individual who was stopped by the NYPD.

It quickly became apparent to us that we needed to constrain the number of datapoints we were using or our website would not be able to load or update itself in any reasonable amount of time. Therefore, after consultation with Charlene, we decided to take the following overarching approach:

1. Choose a manageable subset of years on which to focus. We decided to use the years 2012-2016 (inclusive) for several reasons:
    a. First, Bill de Blasio became the Mayor of New York City in early 2014, and almost immediately began reducing the prevalence of stop and frisk. Thus, this selection of dates allowed us to have data both before and after de Blasio's tenure began, which we hoped would make it simpler to incorporate interesting content into our timeline-based visualization.
    b. Second, because stop and frisk has been on the decline since 2014, more recent years have the pragmatic benefit of containing fewer stops (under ~100,000 per year), which makes it more feasible to include a full 5 year span without having to worry about the same excess data concerns.
2. Choose a subset of features that have practical significance / accessibility, and combine features where possible to allow a smaller set of features that will permit faster data loading and processing. Specifically, we took two primary actions with respect to our datasets:
    a. Dropped a number of variables that we believed were unlikely to be of interest or useful anywhere in our project. For example, one variable was the geographic state where the stop took place -- not only would this information be duplicated in the X/Y coordinates we stored, but presumably every stop was in New York state.
    b. Combined other variables that we felt were useful but too granular. For example, the dataset contained binary variables for whether several types

of weapons were found (machine gun, assault rifle, pistol, knife, etc.) but we felt that given the technical knowledge of most potential readers of our site, we could simplify our dataset without sacrificing crucial information by redefining a variable to encode whether any kind of weapon was found.

3. In combination, these data cleaning procedures resulted in a new dataset with 805,516 entries from January 1, 2012 to December 31, 2016. Each individual stop contains 20 features, some demographic and some stop-specific, as displayed below:


```
addrpct: 40
age: 20
arstmade: "N"
crimsusp: "ROBBERY"
▶ datestop: Sun Jan 01 2012 00:00:00 GMT-0500 (EST) {}
frisked: "Y"
ht_feet: 5
ht_inch: 9
itemBinary: "None Found"
itemfound: "None"
physforce: "N"
race: "B"
searched: "N"
sex: "M"
sumissue: "N"
timestop: "115"
weight: 155
xcoord: "1008031"
ycoord: "233036"
year: "2012"
```

Three D3 Visualizations (1 implemented, 2 drafts):

Note: We decided to begin implementing our choropleth rather than our timeline first, which is the only part of this prototype that differs meaningfully from our proposal.

1. Our first visualization, which we have essentially implemented fully is an aggregated “success” bar chart, shown below. This visualization displays data on the outcomes of each stop (whether illegal items were actually found, constituting “success”) and groups these results by race. Though we did have granular enough data to distinguish between stops where drugs, weapons, or both were found, we found the chart more easily interpretable with only two categories. By switching between percentage and number of stops with two buttons, users can discover that similarly equally low success rates across races, NYPD officers stop Black and Hispanic New Yorkers at disparately high rates.

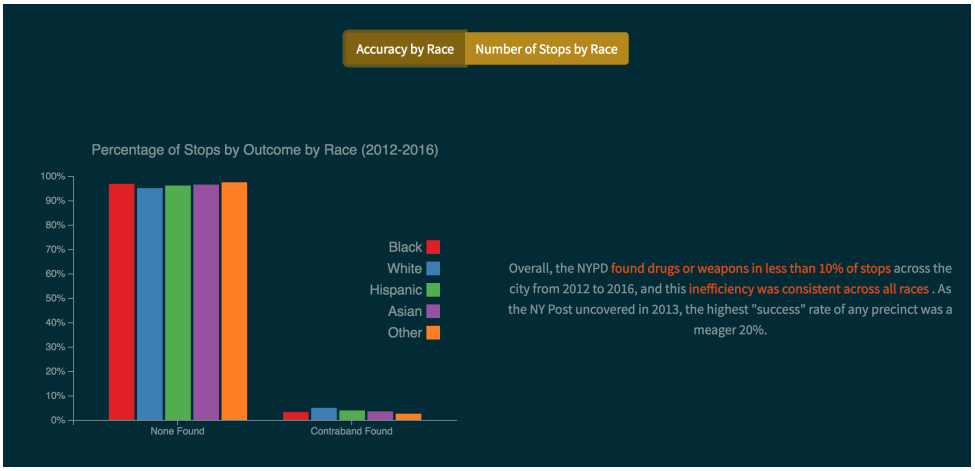


Exhibit 8a. Draft Visualization 1(i): Displaying Success of Stops by Race (Percentage)

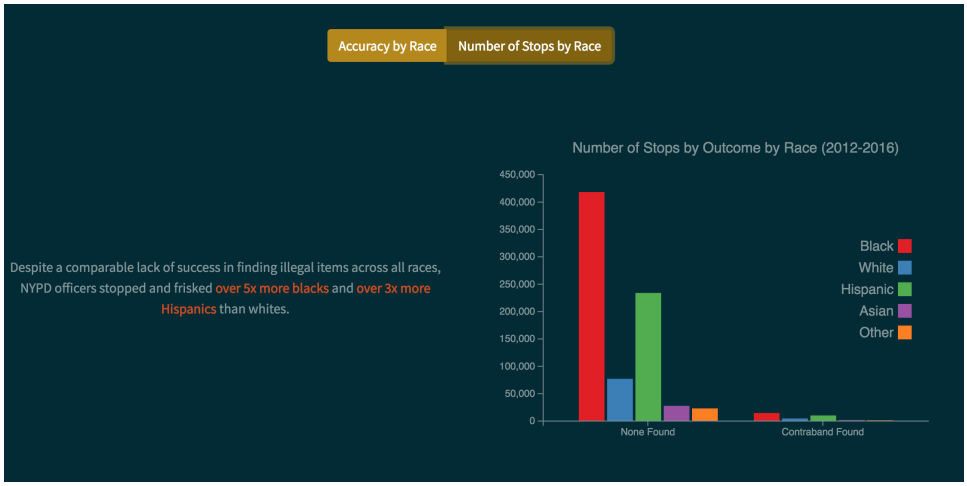


Exhibit 8b. Draft visualization 1(ii): Displaying Success of Stops by Race (Total Number)

2. Our second visualization, which we have made significant strides toward implementing in D3 but have yet to complete, is our choropleth map that will show the prevalence of stop-and-frisk incidents across the varying geographic areas of New York City. A screenshot of our current visualization, shown below in *Exhibit 9*, shows what we have achieved thus far. *Exhibit 4*, from the first process project plan, shows what we expect to achieve eventually. We have successfully mapped New York City using a D3 projection, and have created individual paths corresponding to NYPD police precincts based on a GeoJSON file that we converted to a TopoJSON file that allowed us to graph these individual precincts.

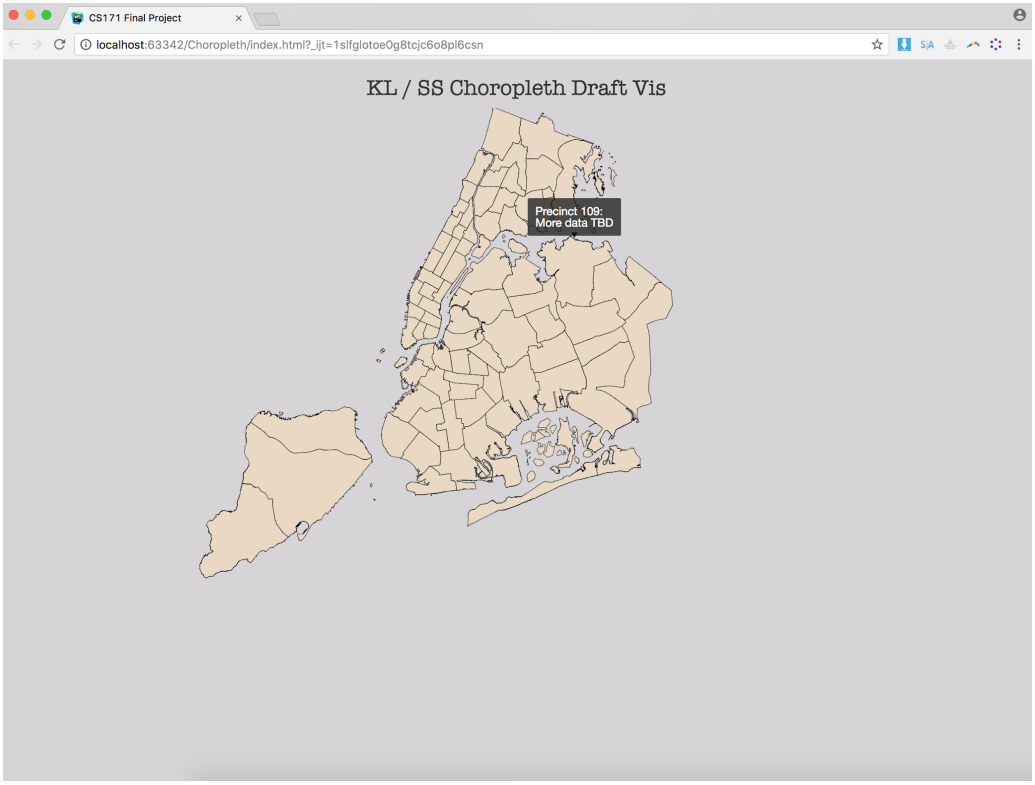


Exhibit 9. Partial Implementation of Choropleth Vis

Now, our challenge is merging this geographic data with our underlying stop-and-frisk datasets, i.e. displaying key statistics about the program in individual precincts. Though we believe this will be very doable, one wrench in our plan is that the NYPD unfortunately encoded their stop-and-frisk X/Y coordinates using the relatively niche “New York Long Island State Plane Coordinate System.” Though we believe there are several reasonable methods for converting these coordinates to standard Mercator ones, we did not want to rush through this important conversion, and will instead tackle mapping these coordinates correctly a little further into our project implementation.

3. Finally, we designed a new visualization, shown below in *Exhibit 10*, which we are considering implementing in lieu of the treemap visualization that we had previously proposed *(Exhibit 3).* We found similar “sunburst” visualizations online and believed creating one might offer a more intuitive way to get the same information (a breakdown of all stops into several subcategories) across to users. Also, we plan to use two smaller circles alongside our main sunburst diagram, as shown below, which will allow us to provide overall demographic information about New York that we believe is significant (i.e. if 40% of people stopped are black and only 20% of the population is black, the story is very different from one in which 60% of the population is black, so we need citywide demographics)

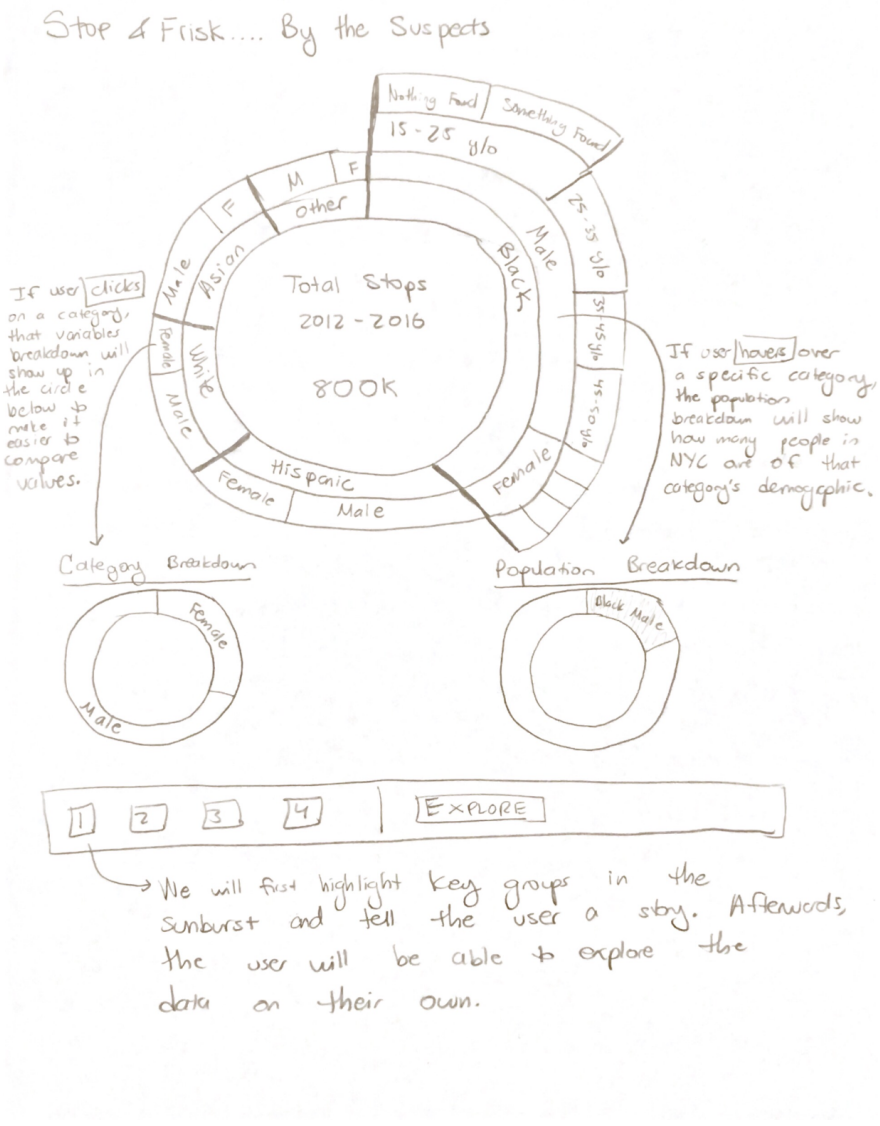


Exhibit 10. Proposed Sunburst Demographics Visualization

Rough Webpage Design:

Note: Pase open our *KL-SS_Website_V1* in WebStorm *to* see our draft site.

We chose a theme and style for our website. At the top of the site, we have a navigation bar that will allow the user to always know where they as they scroll through our site. we We also implemented a “slide scrolling” method with the *fullPage* bootstrap plugin which ensures that a user can see a full visualization each time that they scroll. Having this functionality is important to since we hope to have interactive visualizations that require the user to see a full visualization to understand all the information being displayed.

Thus far, we have not included text within our website since we plan on incorporating it into our visualizations. As users interact with certain elements of visualizations, we will have pop-up text boxes that clarify and summarize what they are seeing. For more detailed information, please see the sketches of our visualizations.

Storytelling:

As we move along with the design and implementation of our visualizations and our webpage in general, we have continued to focus on storytelling about the stop-and-frisk program, and generally continue to believe that our opinionated approach to our storytelling is appropriate in the context of this program. Delivering an “unbiased” assessment of the program would likely be more difficult and less rewarding to us than trying to make several targeted points against the program, so we are continuing to implement a relatively editorializing approach for presenting data about the program.

In terms of our visualizations thus far, we believe the “Accuracy by Race” vs. “Number of Stops by Race” visualizations effectively help further this story, since we explicitly contend to users that despite similar success rates between stops of minorities and stops of white individuals, the NYPD stops 3x and 5x more Hispanic and Black New Yorkers, respectively, than White New Yorkers. Furthermore, we will use our choropleth to show the disparate occurrence of stop-and-frisks in predominantly minority neighborhoods. As we implement more fully-baked visualizations and expand our website’s focus further, we aim to continue with our approach of opinionated storytelling that makes the key points about stop and frisk that we want to convey to readers within the framework of the available underlying data about the program.

Innovative View Design:

For our innovative view, we wanted to implement an interactive view that allows users to specify several of their key demographic characteristics and that subsequently calculates how likely (or unlikely) someone of their background is to be subjected to a stop-and-frisk procedure than someone from another demographic category. In other words, if a user is a 20 year old black male, it is likely that their probability of being stopped by an NYPD officer is orders of magnitude higher than the likelihood of being stopped as a 30 year old white female. We know that of course our likelihood estimates will be based on only the observed data that we have access to from the NYPD, but we believe that prevalence within our dataset is a reasonable proxy for overall likelihood.

We are also excited to implement this visualization because we were both intrigued by the glyph-based Chernoff Face visualizations that we saw in a previous lecture, and wanted to implement our own similar glyphs. While they will not be as complicated as the Trump / Hillary faces we saw, we hope to at least have a few different characteristics or accessories to display depending on the demographics that a user selects.

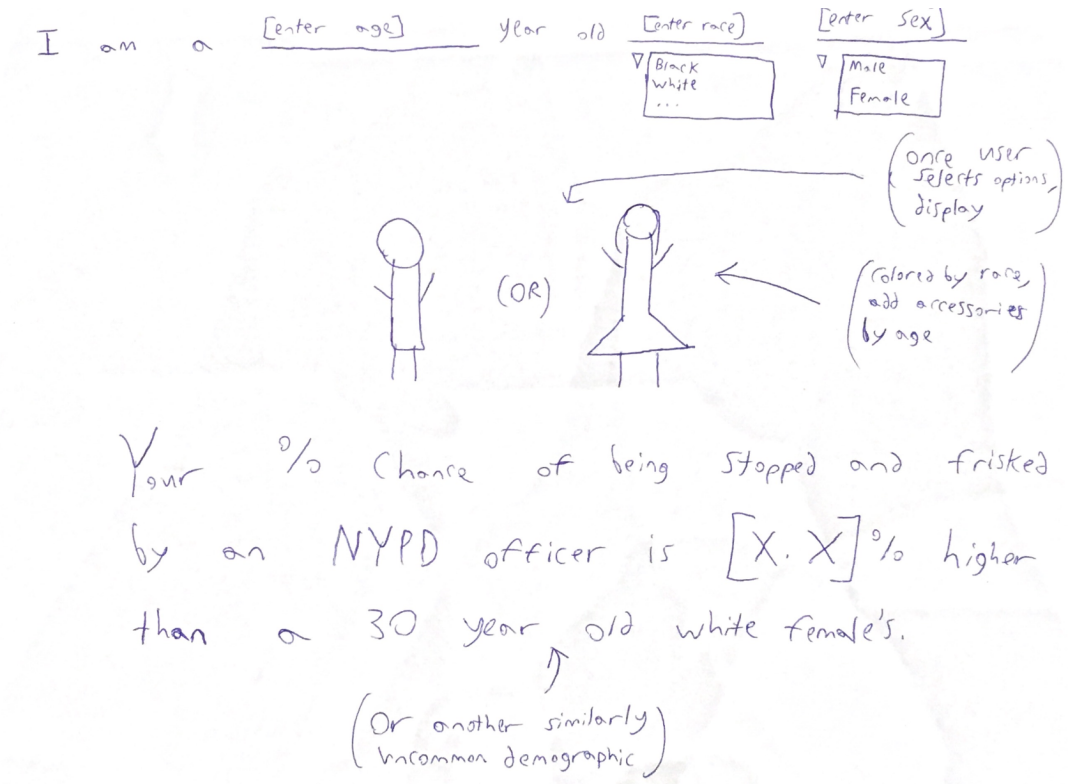


Exhibit 11. Proposed Innovative Visualization: "Your Chances of Being Stopped and Frisked"

Interaction Storyboard:

After reviewing our primary interactive visualization, our choropleth and corresponding timeline visualizations *(Exhibit 6)*, we decided that based on our data cleaning and website implementation, there were no major changes that came to mind -- in other words, we remain comfortable with our previously proposed interaction design for this visualization. Therefore, see *Exhibit 12* below for a copy of the interaction storyboard that we initially proposed to implement, as well as *Exhibit 9* above (and "The Precincts" section of our webpage itself), which represents the initial progress that we have made with respect to this geographical visualization.

That said, we felt that now was an appropriate time to take stock of all the other uses of interactivity, both big and small, that we plan to incorporate into our visualizations, To summarize our interactions, we will go through each visualization and describe how the user will engage with the data presented:

Stop & Frisk by the...

1. Numbers
    a. The user can click between two radio buttons to show two different cuts of stop and frisk data by race, and we may ultimately add tooltips for more granular numbers when the user hovers over a given bar.
2. Media
    a. The user will be guided through a story of how the media has dominated discourse of Stop & Frisk. The user will interact with this visualization by clicking a "next" button every time they are satisfied with learning about a particular time point, and will be able to click hyperlinks to visit the news sources that we provide.
3. Suspects
    a. Again, the user will be guided through a story on which demographics (race, age, gender) are most affected by Stop & Frisk and will be able to click "next" to move forward through sections.
    b. If we wind up using the sunburst then the user will be able to hover over each demographic breakdown to see how that demographic's frisk proportion relates to their proportion in NYC as a whole. Additionally, the user will be able to click on categories of interest (age, gender, race), to more clearly see their frisk breakdown since a lot of different variables will be displayed in the sunburst.
    c. Alternatively, if we use the grid layout visualization that we designed last week, the user will be able to interact with the visualization by choosing

which variables they want to see displayed (i.e. they can chose “hispanic” and “female” and the proportion of hispanic females in the total number of people stopped from 2012-2016 will be highlighted in our grid).

4. Precincts
    a. See *Exhibit 12* interaction storyboard below -- users can visualize different data on the choropleth, filter data by time period, and select individual areas by hovering to view tooltips with extra detail about each precinct.
5. Politicians
    a. This interaction will not be interactive. Instead we will provide information on the stances that politicians have taken on Stop & Frisk over the years.
6. YOU
    a. This visualization is entirely reliant on user interaction. To begin the visualization the user will be asked to input their age, race, and gender.

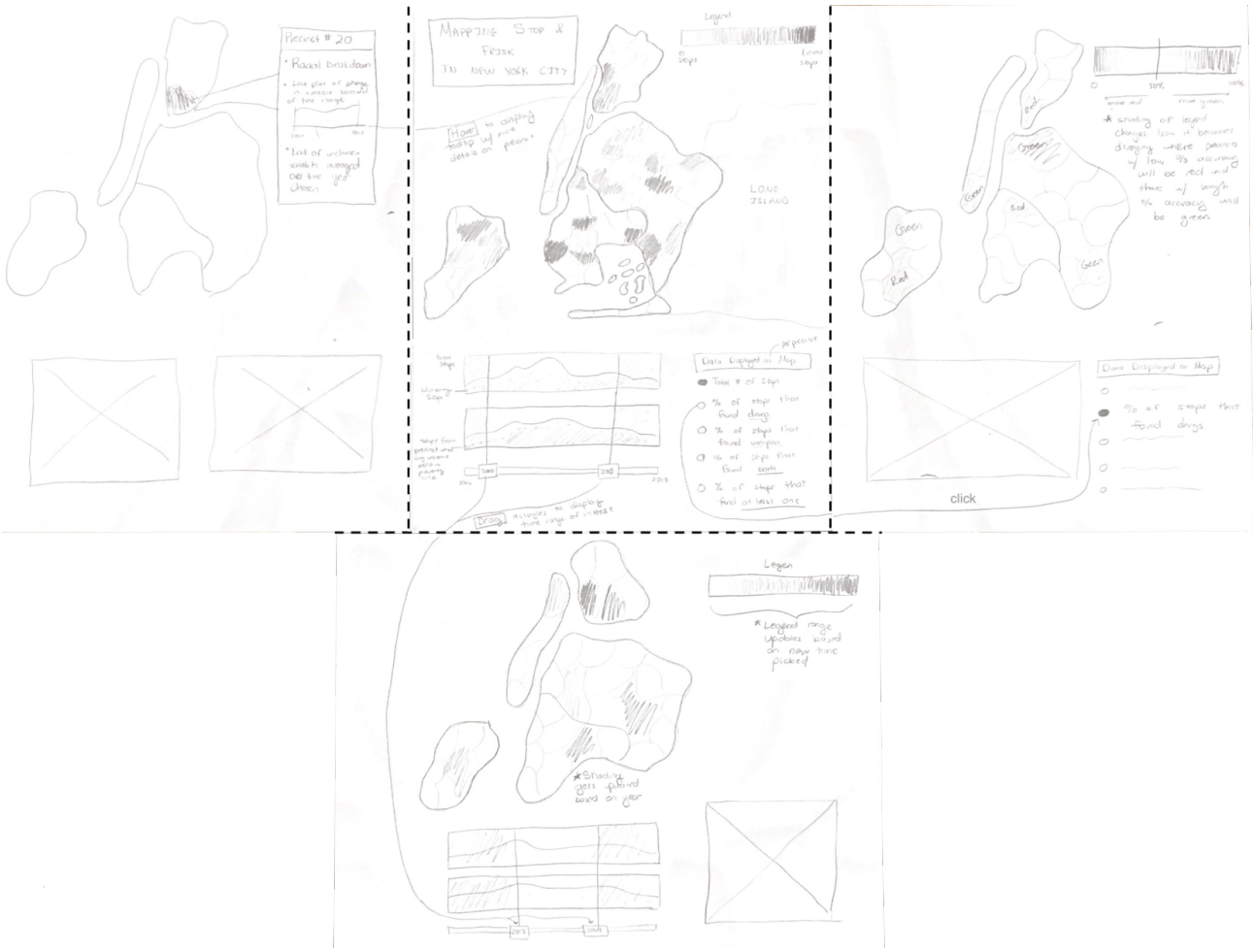


Exhibit 12. Copy of Proposed Interaction Storyboard for Choropleth Visualization

Week 4: Prototype v1.5 (Internal Deadline)

Summary of Progress:

We accomplished our two primary goals for this week with respect to our final project, and despite a weekend shortened by the Harvard-Yale game:

1. Fully implementing our choropleth visualization and its “must-have” features, such as a correctly linked timeline visualization with brushing capabilities.
2. Soliciting feedback from our friends on our overall website structure and the visualizations we had previously implemented.

Choropleth Visualization:

Our first hurdle was finding a way to map individual stop and frisk incidents to geographic data-points. In our previous process update, we were nervous about the potential requirement to convert coordinates from the New York State plane coordinate system into traditional latitude / longitude coordinates and the best approach for converting these coordinates to precincts.

Luckily, Katherine saved the day for our team by noticing that each stop also had an additional column *addrpct*, which we had briefly neglected to consider, which corresponds to the police precinct in which a stop occurred, which enabled us to quickly wrangle precinct-level data without needing to do geographic calculations. At this point, we were able to implement a choropleth like the one shown below in *Exhibit 13.*

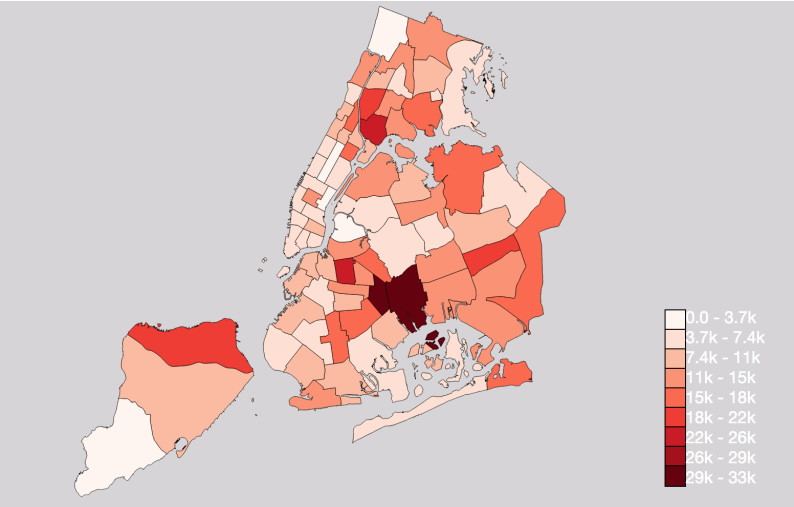


Exhibit 13. First draft of police precincts choropleth

Next, we wanted to follow through in allowing individuals to select their own visualization category, which involved just wrangling a few percentages for each precinct (percentage of stops that yielded contraband, percentage of stops that led to arrest, and percentage of stops that were of minorities). We were able to accomplish this relatively quickly, leading to a second draft of the choropleth depicted in *Exhibit 14* below.

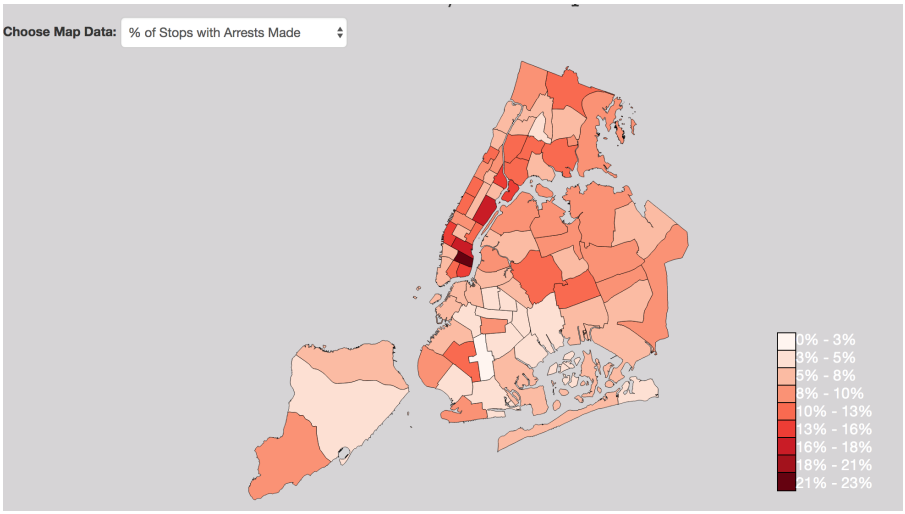


Exhibit 14. Second draft of police precincts choropleth

Our third choropleth-related task was to link our visualization to an area chart that showed stops per day from the beginning of 2012 to 2016, as we planned in our initial interaction storyboards. Though we have implemented linked views in the past for homework and lab projects, creating a linked view with real-world data proved more difficult because of errors in our data and some inconsistencies that we had to address. A couple that posed particular challenges were:

- Due to data entry errors by NYPD officers, a few stops were indicated to have occurred in “precinct 0”, a non-existent precinct, and since our legend and scale are dynamically generated, our min and max percentages were heavily skewed, making our scale meaningless for many categories. We addressed this issue by removing the 3 points in question, since a dataset of 700k points is unlikely to suffer major data integrity issues from such a small change.
- If users filtered to a narrow enough time range, certain precincts may have data while others do not, so we had to modify our initial code to allow the choropleth to display data for some precincts (and their corresponding tooltips) while appropriately indicating that others had no data available.

Now, our choropleth appeared as depicted in *Exhibit 15* below.

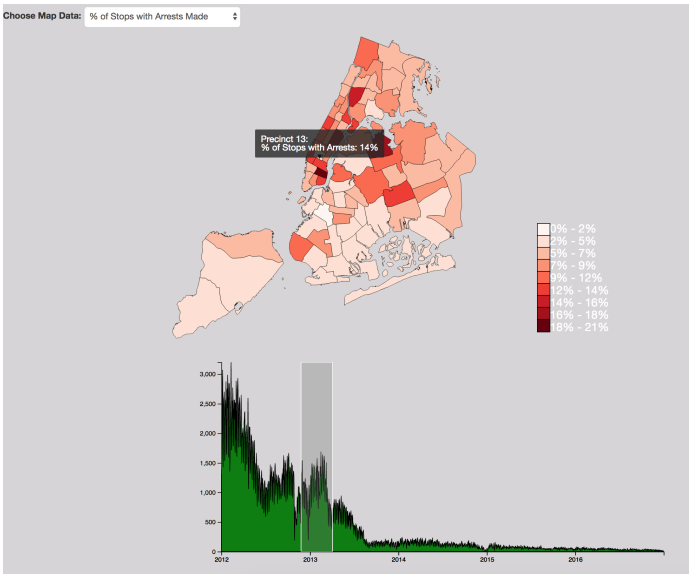


Exhibit 15. Third draft of police precincts choropleth

At this point, we were ready to incorporate the functioning linked visualization into our full final project website (as shown in *Exhibit 16)*. This integration process included many aesthetic tweaks, but also included one major functional one: we created a standalone tooltip on the side of the visualization and integrated external racial composition data for each precinct (based on peer feedback as described below), so users can see at a glance how prevalent minorities are in each precinct that they hover their mouses over, which we hoped would further contextualize the stop and frisk data.

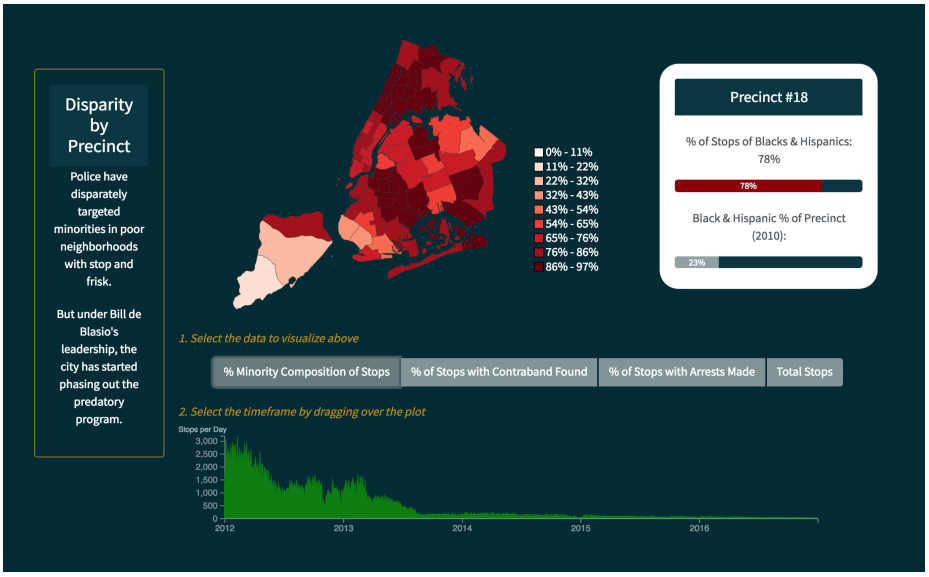


Exhibit 16. Final draft of police precincts choropleth as incorporated into the site

Peer Feedback (Round 1):

We both discussed our current website design with our roommates, and several main pieces of feedback came up consistently that we tried to address.

1. As part of our *fullpage.js* approach to creating a flowing webpage, we had been using the default margins / padding settings, and several of our friends noted that our frames were too short and too narrow, leading to visualizations taking up odd dimensions that were not well-suited to the messages we were trying to show. Accordingly, we learned how to change the height and width of these frames and adjusted moving forward.
2. The tooltip for our choropleth visualization appeared directly above our cursor, obfuscating other precincts that became hidden “behind” the tooltip. Some of our friends complained that it was distracting to have these tooltips, so we decided to instead use one tooltip box on the side of our visualization. We also decided to include precinct-level statistics, and (inspired by a previous final project) also added moving “progress bar” visualizations to show all percentages visually in addition to textually.
3. Finally, we got several different (and unfortunately, sometimes contradictory) ideas about how to improve our site color scheme, which helped us realize that an important goal of ours once we have our visualizations properly implemented will be sitting down as a team and perhaps discussing with Charlene several ideas for appropriate color choices as we conduct the final stages of refining our project.

Week 5: Prototype v2

Summary of Progress:

For this week, we had initially aimed to complete three main goals: tinkering to improve our website layout / CSS and implementing a nested treemap. But since parts of our site unfolded somewhat differently from our initial intentions, and we became more attached to some of our alternative visualization designs over time, our plans for the week evolved to be slightly different. Thus, we had three main goals for the week:

1. Implement a sunburst visualization that intuitively shows users what a “typical” stop and frisk victim looks like (i.e. the likelihood of belonging to a certain race and gender for each stop and the likelihood of the stop uncovering contraband). We decided on this visualization in lieu of our previously intended treemap, since we thought it was more intuitive and more aesthetically pleasing.
2. Implement our non-D3 visualizations (e.g. the newsreel and list of key political stances on stop and frisk issues) and properly incorporate them into our broader website design while keeping in mind design principles and storytelling.
3. As we initially intended, continue working on refining our website / CSS.

Sunburst Visualization:

Our initial intent for this visualization was to implement a treemap that would break down stops based on demographic factors. As we proceeded, however, we decided a “sunburst” might be nicer aesthetically and more innovative and captivating for users of our site. Therefore, we began implementing a sunburst based on aggregated data of stops by different demographic characteristics. The initial draft of our visualization is shown below in *Exhibit 17*.

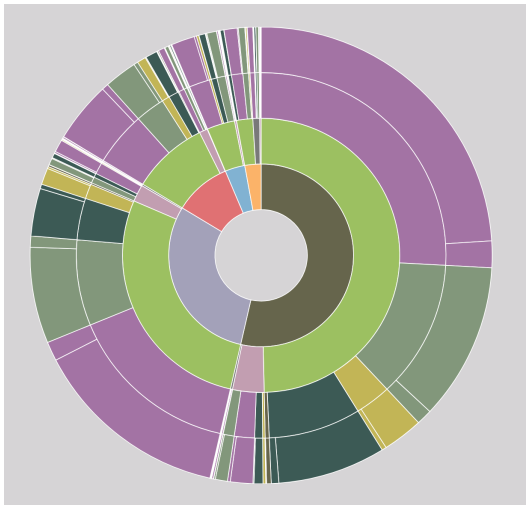


Exhibit 17. First draft of sunburst visualization

Once we saw this basic sunburst skeleton, we both agreed that it was a unique and interesting way to visualize the data we wanted to show users, and proceeded to improve and develop our design. The first hurdle was adding text to our diagram, which proved far more complicated than we anticipated because we needed to rotate certain text boxes but not others (e.g. for very small slivers). Luckily, we were able to modify an algorithm a previous D3 user had advocated, allowing us to implement the visualization shown in *Exhibit 18* below.

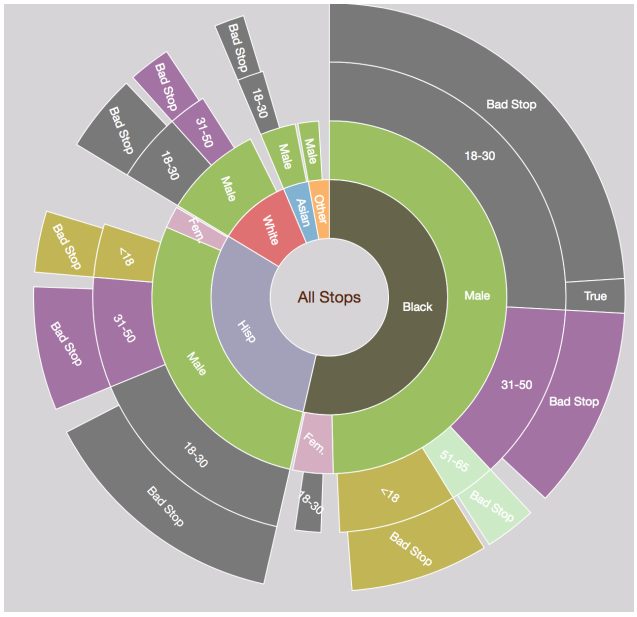


Exhibit 18. Second draft of sunburst visualization

Finally, we wanted to incorporate interesting tooltip-esque interactivity into our visualization, since we knew that small labels without explicitly stated percentages would lead to a relative degree of uncertainty for our users about how to actually interpret these sunburst. Therefore, we decided to incorporate a “breadcrumb” sequence as a tooltip, much like what appears on the top of a computer’s file browser (e.g. “Desktop/CS171/Final Project…”). In other words, as the user hovered over each segment of our sunburst, sequential arrows would appear at the top of their screen to show them where they are in the grand scheme of the breakdown of stops and what percentage of stops fell within that category. Luckily, we were again able to modify some sample code from online to create a final working draft of our sunburst that we think is not only a particularly compelling and innovative visualization, but also an interpretable one, as shown in *Exhibit 19* below.

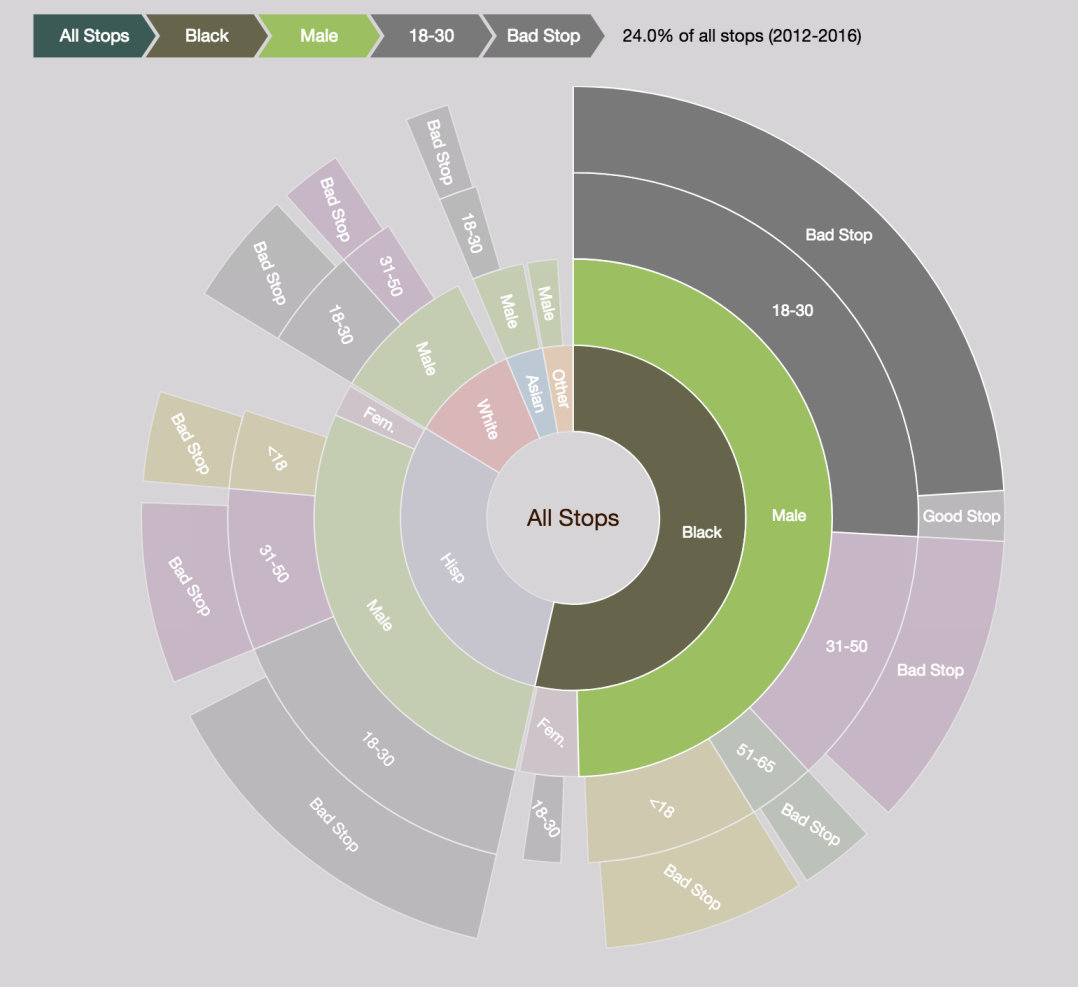


Exhibit 19. Final working draft of sunburst visualization

Non-D3 Content / Pseudo-Visualizations

In addition to the Sunburst, we spent a substantial amount of time this week developing our non-D3 content and pseudo-visualizations, including a timeline of key moments in the history of stop and frisk and information on several key politicians associated with the historical development and policy debates surrounding stop and frisk.

Week 6: Finalizing Our Project

Summary of Progress:

Luckily, we were able to progress smoothly on our project throughout the semester and essentially kept to our initially-planned project timeline throughout, which meant that the last two weeks we had to work on the project, following Prototype v2, were largely focused on final tweaks and improvements based on Charlene's feedback and our peer feedback rather than scrambling to implement a large amount of additional content.

Overall, the majority of the time we spent on the project during these final weeks can be roughly segmented into three main categories:

1. Front-End Website Design Changes: Innovative View and "Odds and Ends"
2. Back-End Code Changes: Adding comments and removing outdated code
3. "About" Section: About Us, Acknowledgements, Screencast, References

Front-End Website Design Changes:

Innovative View:

The major website design addition that we made was with respect to the project's innovative view requirement. After discussion with Charlene, we realized that what we had intended for our innovative view, the section of the "Future" page that told users the amount by which their odds of being stopped-and-frisked declined from 2012 to 2016, did not satisfy the course project's requirements. Upon her suggestion, we decided to instead implement a visualization that conveyed this same data and information to users in a visual manner rather than in words alone, by showing them glyphs of people, which would decline in number to represent the decline in stops from 2012 to 2016.

In addition to the pragmatic advantages of satisfying the course's requirements for an innovative visualization, we also thought this view helped humanize a statistic in a way that might convey it more powerfully to users. Since the decline in stops under Bill de Blasio for essentially any demographic group was stark (over 15x in almost all cases, approaching 75x for certain demographic groups), we thought seeing a rapid decline in the number of stick figure "people" shown on the screen would help remind users that these large numbers are not just statistics, but also representative of human beings.

In terms of the technical approach we took, we used Dave Gandy's Font Awesome for our stick figure images, which allowed us to create a grid of 100 individuals that

represented 100 “people like you” (based on demographic characteristics that the user inputted). Then, when they clicked a radio button to select their race / gender, and/or moved a slider to change their age, a huge number of stick figures in the grid see their opacity reduced, making them appear faded. A picture of how the grid changes upon a user’s interaction with the demographic selectors is provided below in *Exhibit 20*.

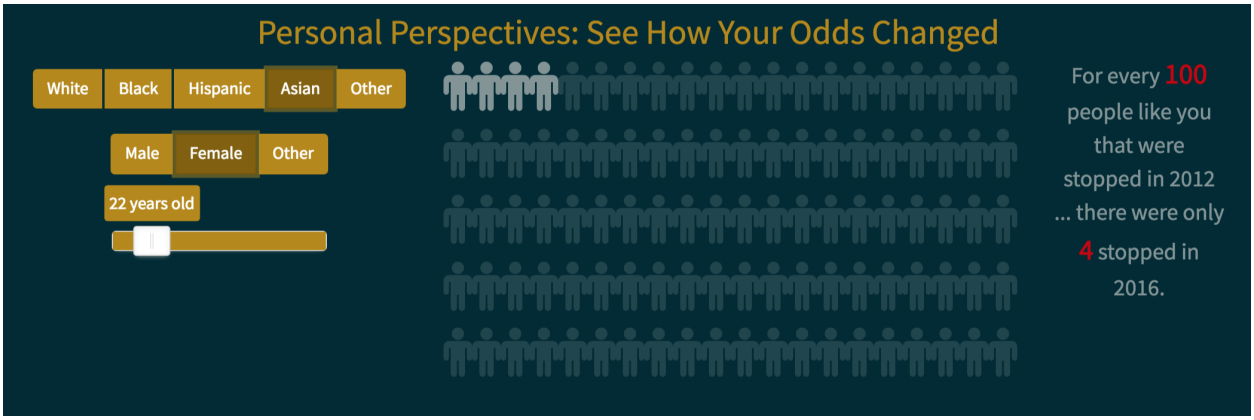


Exhibit 20. Innovative View Visualization

We were very pleased with this final result, and believe the innovative view does a better job of quickly conveying our intended message, that the lives of New Yorkers on the ground have truly improved to a significant extent since de Blasio’s election, than text alone did in our prior iterations of the future visualization. A huge thank you to Charlene for providing us advice and support as we implemented this visualization, i.e. by suggesting that we draw the people initially when the website loads to avoid lag later.

<u>Odds & Ends:</u>

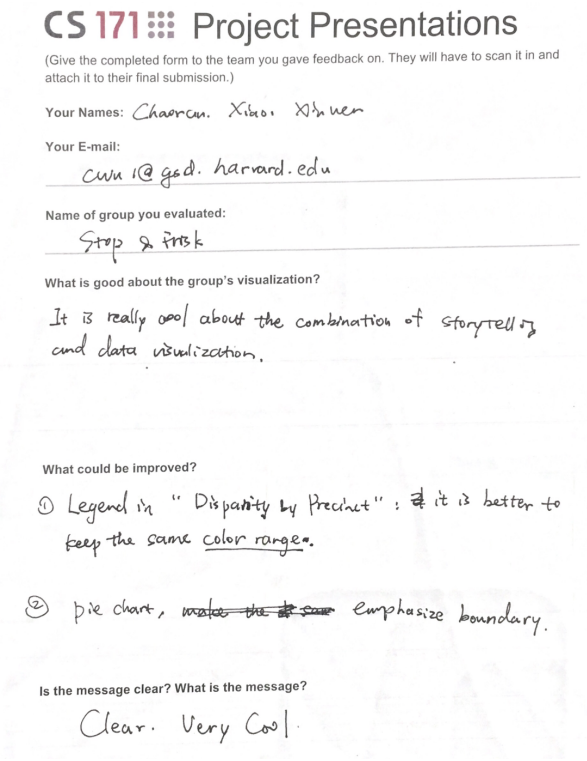
In addition to the innovative view, there were numerous smaller suggestions that Charlene and our classmates made to us in terms of improvements, and we tried to implement as many of them as we feasibly could given our time and site constraints. A few of the suggestions were necessarily overlooked either due to an excessive time burden associated or because our site / data itself would be unlikely conducive to the suggestion. In recognition of the hard work that everyone made in providing us input, however, we decided to include every suggestion we received in our process book below in *Exhibit 21*, as well as a brief explanation of what course of action we decided to take with respect to the suggestion. Please also see *Exhibit 22* and *Exhibit 23* for scanned versions of the official feedback forms from the two teams we worked with during the project workshopping lab meeting.

Exhibit 21. Various Final Project Suggestions (Post-Prototype v2)

Suggester	Suggestion	Response
Charlene Hwang	Fix spacing on cards for each politician	**Implemented fully:** We realized that we had been forcing our photos to take on a constant aspect ratio, which badly distorted some photos.
Charlene Hwang	Implement an innovative view for the “future” page	**Implemented fully:** See “Innovative View” section above for more details.
Charlene Hwang	Potentially reduce lag on choropleth	**Unable to implement:** Even reducing number of columns in our data did not help -- the lag arose from our large number of rows, and removing rows would lead to statistical validity concerns, so unfortunately users will have to deal with slight lag.
Yanni Cho	Fix a typo on sunburst description	**Implemented fully:** Changed description from “left” to “right” to match the actual viz position
Yanni Cho	Remove the “submit” button on the future visualization	**Implemented fully:** As per Yanni’s suggestion, changed our innovative visualization demographic selectors to automatically trigger a change in the visualization when moved, rather than requiring users to click a submit button
Fritz Lekschas	Make colors consistent across visualizations	**Implemented fully:** Fritz correctly noted that the colors for each race were inconsistent between our “Numbers” and “Suspect” visualizations, even though the underlying races were the same, so we changed our “Numbers” visualization to match the color scheme present in our sunburst -- a great catch from Fritz!
Fritz Lekschas	Change dropdown to buttons in choropleth	**Implemented fully:** Fritz suggested that since we only had four choices in our drop-down menu, it would likely look better / simpler as radio buttons, so we followed this advice
Fritz Lekschas	Change color of lower progress bar on choropleth tooltip	**Implemented fully:** We made the precinct racial percentage progress bar color darker so the white font would appear better on top of it
Fritz Lekschas	Allow side-by-side view of two bar charts in “Numbers”	**Implemented fully:** Previously, users could only select a button to view one bar chart at a time in our “numbers” visualization, with text

		displayed alongside the selected bar; now, users can click a third radio button that allows them to see both side-by-side for easier comparability
Frances Shapiro	Allow scaleable windows	**Unable to implement:** Unfortunately, because of the way the *fullpage.js* framework functions, it would be extremely difficult for us to allow for dynamically scalable content (at least given our few weeks worth of experience with *fullpage* methods); if we had significantly more time to work on the project, this would certainly be a change we would consider implementing
Frances Shapiro	Allow clickable photos our timeline	**Implemented fully:** In addition to linking to news articles in headline text itself, we now embed hyperlinks in timeline article photos
Frances Shapiro	Change wording in “Numbers” titles	**Implemented fully:** For our left “numbers” bar graph, we changed our titles to “stop outcomes by race”, hopefully making it clearer to users
Molly Cinnamon, et al.	Reorganize choropleth slide	**Implemented fully:** In addition to changing our drop-down menu to buttons, we removed a horizontal line and instructions that this group thought was a distraction from our visualization
Molly Cinnamon, et al.	Make demographic context clearer	**Declined to implement:** After discussing with the group that made this comment (and showing them our choropleth) we were satisfied that no major changes were needed
Chaoran Wu, et al.	Make choropleth color ranges static	**Declined to implement:** This group suggested we eliminate the dynamically updated color ranges for our choropleth, but since some time ranges have drastically different values than others, this change would lead to unuseful maps with almost entirely dark red / light pink colors
Chaoran Wu, et al.	Thicken sunburst borders slightly	**Implemented fully:** This group pointed out that thicker borders between arcs would allow users to segment our sunburst more quickly and process the data provided more efficiently

Exhibit 22. Feedback Form #1 from Project Workshopping



CS 171 Project Presentations

(Give the completed form to the team you gave feedback on. They will have to scan it in and attach it to their final submission.)

Your Names: Chaorcu. Xiao. Xinuer

Your E-mail:

cwu1@gsd.harvard.edu

Name of group you evaluated:

Stop & frisk

What is good about the group's visualization?

It is really cool about the combination of storytelling and data visualization.

What could be improved?

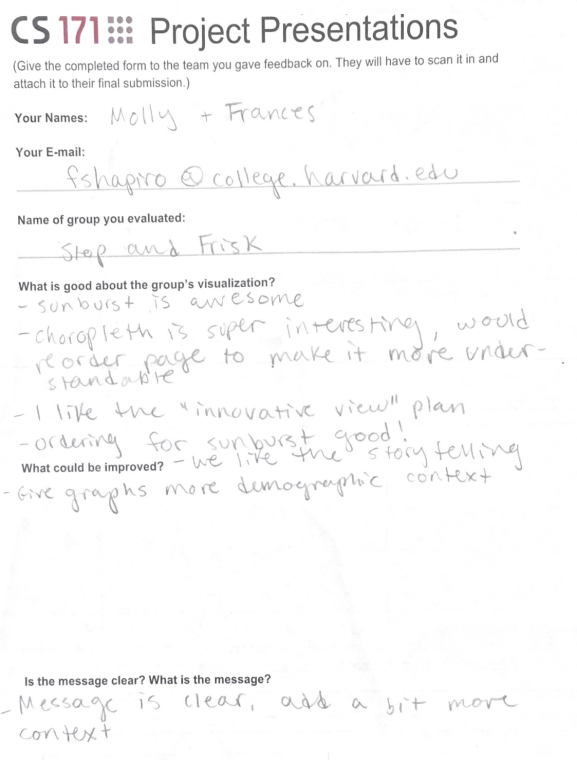
① Legend in "Disparity by Precinct": it is better to keep the same color range.

② pie chart, ~~make the it same~~ emphasize boundary.

Is the message clear? What is the message?

Clear. Very Cool.

Exhibit 23. Feedback Form #2 from Project Workshopping



CS 171 Project Presentations

(Give the completed form to the team you gave feedback on. They will have to scan it in and attach it to their final submission.)

Your Names: Molly + Frances

Your E-mail:

fshapiro@college.harvard.edu

Name of group you evaluated:

Stop and Frisk

What is good about the group's visualization?

- sunburst is awesome
- choropleth is super interesting, would reorder page to make it more understandable
- I like the "innovative view" plan
- ordering for sunburst good!
- we like the story telling

What could be improved?

- Give graphs more demographic context

Is the message clear? What is the message?

- Message is clear, add a bit more context

Back-End Code Changes

We also made a number of small changes to our codebase. It is probably unhelpful to write about these changes in great detail. Instead, we will highlight the few largest changes:

- Removed the old version of the *futurepage.js* file that creates our site's last slide; this version did not have our innovative view, and only had the static text content
- Deleted commented out HTML elements and most of our commented out JS code to improve loading time (though we kept a few old important code remnants in case something breaks over the last few days and we need to revert). In some *JS* files, this old code comprised almost 50% of the file's lines, so it made a large difference in terms of readability for these files
- Add comments to JS where they were too sparse (unsurprisingly, mostly in code segments that required a lot of work right before Prototype v1 and v2 deadlines!)
- Implemented a README.md file in our project directory as suggested on the CS171 website, containing a description of every file in our project and instructions for use

"About" Section

Finally, we added an "About" section complete with acknowledgements as well as links to our data and process book, as well as our embedded screencast. Additionally, the section includes a reference list, which is complete with roughly 25 references, from Javascript frameworks that we used in our code to news outlets that provided us with website photos to individuals who provided a ton of D3 examples and code online for free consumption and modification.

With that, we completed the final tweaks to our website we intended to make following Prototype v2, and called it a semester -- our project was done!

Week 7: Summarizing our Efforts

As we wrapped up our final project, we wanted to aggregate answers to the key questions that Dr. Beyer noted were essential to the data visualization process. Specifically, see below for an answer to these key questions (some sections are replicated from earlier writing in our process book, so feel free to skim these!):

Overview and Motivation

The New York Police Department ("NYPD") is tasked with patrolling a city with over 8 million residents and 50 million annual tourists.[3] As part of the department's crime-fighting initiatives, officers rely on the Stop, Question, and Frisk ("SQF") program to detect and remove weapons and contraband from city streets. Though the NYPD has conducted SQF operations since the 1990s, the practice became far more widespread under the mayorship of Michael Bloomberg that began in 2002. By 2011, there were over 650,000 NYPD stop-and-frisks across the city. Since these programs are mainly implemented in New York's most dangerous neighborhoods, Hispanic and Black civilians are stopped at a disparate rate (over 52% of SQF stops involved black suspects, and 31% involved Hispanic suspects).[4] Though the program has been curtailed under Mayor Bill de Blasio, SQF remains part of NYPD policing protocols.

Though the racial biases of SQF are well known, we believe effective visualizations of SQF data will help quantify the scale of the program and illuminate how stark the racial disparities in NYPD stops are for minority communities. The NYPD has a meticulous database of thousands of SQF records for each year (including geographic information, time of day, characteristics of suspects, follow-up arrest information, etc.), which will serve as the data underlying our visualizations.

Related Work

As we thought about how to best visualize this complicated data, we were inspired by several main visualizations. First, the New York Times's piece on "Mapping the Decline of Stop and Frisk" showed us the power of using a map with this dataset. Second, John Keefe's map of guns found in New York City inspired us to consider a choropleth. Finally, FiveThirtyEight's famous Gun Deaths visualization got our group thinking properly about how to implement our novel visualization (on the "Future" page).

---

[3] The numbers add up to one fact: cops are a blessing to NYC," New York Post, http://nypost.com/2014/12/31/the-numbers-add-up-to-one-fact-cops-are-a-blessing-to-nyc/.

[4] Here's what you need to know about stop and frisk - and why the courts shut it down," Washington Post, https://www.washingtonpost.com/news/wonk/wp/2013/08/13/heres-what-you-need- to-know-about-stop-and-frisk-and-why-the-courts-shut-it-down/.

Questions

Initial Questions:

- What individuals are most affected by Stop & Frisk?
- Are there racial biases to Stop & Frisk? We the ability to see how stop-and-frisk was actually conducted in New York City and allow them to independently elucidate the harmful racial biases that arose as a result of its implementation
- What are the geographic biases to Stop & Frisk?

Evolved Questions:

- As we put together our website we found that it would also be good to highlight Stop & Frisk's role in New York City and answer the question "What is the history of Stop & Frisk in NYC and how did it affect civilians?". We thought that by educating users on the history of the NYPD's stop-and-frisk program, and contextualize the program in the broader landscape of law enforcement in New York City and around the country.
- Additionally, as we continued to explore the data, as explained in a section below, we also found that there were gender and age biases to the program. Thus we wanted to broaden our previous question looking at just race to "What are the demographic biases to Stop & Frisk?"
- Finally, we wanted users to also feel a personal connection to the topic. Thus we wanted to answer the question "How would someone like you be affected by Stop & Frisk today rather as compared to 2012?" to highlight how much the program has improved and decreased in size.

Data

We had official NYPD data on stop-and-frisks from 2003 to 2016, but decided to focus on data from 2012 to 2016 to minimize computational requirements and allow our site to load more quickly. Each year's dataset contains thousands of entries corresponding to individual stops and roughly 20 columns that we chose to use for our analysis.

The variables in the datasets are generally either demographic, stop-related,or geographic. We believe this mix of different types of variables allows the most interesting set of visualizations and lends itself to several compelling stories. While the most famous "story" associated with the stop-and-frisk program is its racial disparity, our data also allows us to tell stories about the geographic distribution and success rate of stops. Ultimately, we feel extremely lucky to have had access to a freely available and comprehensive dataset about such an important topic.
As we mentioned throughout the process book, the data was largely cleaned and required very little additional effort, though as we mentioned earlier, we did need to drop a few entries due to clearly false data that was breaking our code. Ultimately, we felt

confident that dropping only a few out of hundreds of thousands of data points would have essentially no serious effects on our data integrity.

Exploratory Data Analysis

Given that both of us are Statistics concentrators we first ran some regressions assessing the statistical significance of racial and demographic factors in the program to see if our advanced exploration and visualization was merited. We saw that being black and hispanic significantly raised a person's chances of being frisked. Surprisingly, we found that other demographic characteristics were also significant beyond just race, such as gender and age. This initial analysis informed us about what data and patterns in the data we wanted to show in our final website. The output of one of these regressions is provided below in *Exhibit 24*.

Exhibit 24. Regression Assessing Predictors of Frisks

```
> summary(frisks_racial)

Call:
glm(formula = frisked ~ is_blackhisp + age + real_height + weight +
    male, family = binomial(link = "logit"), data = SQF2004)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7531  -1.1021  -0.8006   1.2003   2.3926

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.3556184  0.0905390 -26.018  < 2e-16 ***
is_blackhisp  0.5488477  0.0127255  43.130  < 2e-16 ***
age          -0.0245050  0.0003972 -61.701  < 2e-16 ***
real_height   0.0170971  0.0014629  11.687  < 2e-16 ***
weight        0.0012372  0.0001670   7.407 1.29e-13 ***
male          0.9071062  0.0193398  46.904  < 2e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 347597  on 255194  degrees of freedom
Residual deviance: 337326  on 255189  degrees of freedom
AIC: 337338
```

Design Evolution and Implementation

As you scroll through the process book from *Prototype V1 → Prototype V1.5 → Prototype V2 → Final* you will see screenshots of our design sketches and how our visualizations developed. Additionally, we discuss which design principles we took from the course and when we deviated from them, such as in the final "personal" view with human figures rather than regular barcharts, in order to personalize the website and get the viewers hooked on the topic.

Of course, a picture is worth 1,000 words, and a video is worth 1,000 pictures, so feel free to also see our screencast, which is a narrated walkthrough of our project site and the various interactive (and uninteractive) visualizations it contains.

Evaluation

From the visualizations we were able to discover even more demographic and geographic biases that we did not even think of when initially looking at the data just through a statistical analysis in R. By seeing the data and having live updates through our interactive choropleth visualization, for example, we that there were not only disproportionately more minority stops in the wealthy Upper East Side, but that those minority stops were also much more likely to be unfruitful, thus proving the ineffectiveness of the Stop & Frisk program.

Our visualizations work well in both allowing the user to explore the data for themselves while also giving them enough context to make informed conclusions about the data they see, rather than just considering it in a vacuum. That said, we did identify a couple areas for improvement to make our visualizations even stronger. First, we wish our data could load and update more quickly, but because of our sheer quantity of data, it tends to lag slightly both when initially opening the page and when changing our choropleth. Second, we would loved to have developed an even more creative innovative visualization that startles and excites users when they see it. Finally, we would have experimented with more uncommon yet compelling CSS techniques to develop a more unique website background that differs from standard data visualization sites.

Acknowledgements

As we mention on our site itself, we owe a tremendous amount of gratitude to Dr. Beyer and the CS171 teaching staff, but most of all, to Charlene Hwang, who made everything we accomplished on this final project possible, and whose section helped make CS171 the great class that it was for both of us. We can't wait to work with her again soon.