**ISA 520 Final Project**

**Civilian Complaints Against NYC Police**

Liam Smith



Index

[1. Introduction / Dashboard Story and Objectives 2](#_Toc1270829833)

[2. Dataset Description 4](#_Toc1625789422)

[3. Data Visualizations and Final Dashboard Elements 8](#_Toc1089700318)

[Filters and Interactivity 9](#_Toc794601808)

[F1. Board Disposition 9](#_Toc840745105)

[F2. Precinct Location 10](#_Toc1704619381)

[I3. Data Source Button 12](#_Toc1866012069)

[Dashboard Visualizations 14](#_Toc1418533073)

[D1. Number of Allegation: Board Disposition – Score Cards 14](#_Toc517789735)

[D2. Allegations Received by Offense Type – Area Chart 15](#_Toc311134587)

[D3. Allegations Towards NYC Police – Lollipop Bar Chart 17](#_Toc1820327953)

[D4. Complainant Demographics – Drilled Bar Chart 19](#_Toc1199026016)

[D5. Officers Change in Rank – Sanky Chart 21](#_Toc562257035)

[D6. Substantiated Complaint Board Disposition – Waffle Chart 23](#_Toc380823468)

[D7. Action Time by Offense Type – Box and Whisker Plot 26](#_Toc367243938)

[D8. Complaints by Precinct Locations – Map 28](#_Toc371309104)

[Additional Visualizations 30](#_Toc2044728534)

[V1. Complaint Category – Tree Map 30](#_Toc1727413110)

[V2. Reason for Contact – Pie Chart 31](#_Toc513407826)

[V3. Most Complained About Officers – Bar Chart 32](#_Toc1889109223)

[V4. Complaints Over The Years – Histogram 33](#_Toc1538859156)

[V5. Top 5 Allegations Towards NYC Police – Stacked Bar Chart 34](#_Toc1936279042)

[V6. Action Time by Officers Demographics – Violin Plot 35](#_Toc1519217806)

[V7. Officers Change in Rank – Stream Graph 36](#_Toc1438977886)

[V8. Substantiated Complaint Board Disposition – Donut Chart 37](#_Toc833977884)

[V9. Action Time by Offense Type – Ridgeline Plot 38](#_Toc1508411346)

[V10. Complaints by Precinct Location – Bubble Chart 39](#_Toc1811786776)

[V11. Complaint Category – Sunburst Chart 40](#_Toc1263090426)

[V12. Civilians’ Outcome to PD Suspecting Violation/Crime - Lollipop Chart 41](#_Toc1958805085)

[V13. Top Officers Board Dispositions – 100% Stacked Bar Chart 42](#_Toc1953848683)

[Final Dashboard 44](#_Toc1680502081)

[4. Reflection, Future Improvements and Conclusion 45](#_Toc1336561703)

# **1. Introduction / Dashboard Story and Objectives**

1. Names of each group members

Our group members are Liam Smith, Sabrina Rao, Victoria Olivieri and Karen Abdallah.

1. Index of what items are included in the report

Index is displayed on the second page of this report.

1. Briefly describe the overall objective of your dashboard project.
   1. What is the main story or insight you aim to convey through your dashboard?

In our dashboard, we aim to show the trend of civilian complaints against the NYPD from 1958 to 2019. In doing so, we not only show the tendency of complaints received by the Civilian Complaint Review Board, but also, we uncover insights related to these complaints.

* 1. What are the specific objectives you want to achieve with your dashboard?

The specific objectives we want to achieve with our dashboard are to dig deeper into the complaints against the NYPD and understand the characteristics of the complainants, the reason for these complaints, the nature of them as well as their outcomes. In addition, using the complaints’ nature, we aim to uncover patterns related to distribution of complaints across precincts, the amount of time it takes to close the complaint and characteristics related directly to the NYPD personnel.

* 1. Why is a dashboard appropriate for this story / goal verses just one or two independent visuals?

The use of a dashboard to tell the story about the growing number of civilian complaints against the NYPD from 1858 to 2019 is necessary because it allows us to dig into all the factors related to these complaints at once. Using a dashboard to present the nature of the complaints, we are able show the viewers the trend of complaints against the NYPD, information about the types of allegations, the outcomes of the complaints as well as findings about the complainants. The viewers can see the individual and group impact in one area and use the instructiveness of the visuals on the dashboard to uncover more insights or patterns about these complaints. This could not be accomplished with separate visuals.

1. Provide an overview of the dataset you selected and its relevance to business analytics and data science.

The dataset chosen is from ProPublica collected from New York City’s Civilian Complaint Review Board. It contains a total of 33,358 allegations with 31 total features about the type of allegation, characteristics about the police personnel is question, characteristics of the complainants, as well as the outcome of the complaint. These allegations have been filed and closed from 1958 to 2020. This dataset is relevant to both business analytics as well as data science. For example, if the police department is looking to decrease the amount of complaints they get. This dataset will help them discover a pattern regarding the allegation type and then can train the police personnel depending on these indicators.

1. Describe the software selected for this project
   1. Detail how Tableau, PowerBI, and Python were used in the project, if at all.

We used two different softwares in this project. First, we read in the data retrieved from ProPublica into Python where we did different types of data manipulations and data cleaning. The dataset contained precinct numbers but not the locations of these precincts. We found a different dataset on “nyc.gov” that indicated the precinct number and the location associated with it. We then merged the precinct location dataset with precinct number on the civilian complaints dataset to get the address and the location of the precinct. We also found the date was separated into day, month, and year fields. We concatenated these fields to create one datetime field. We remove precinct numbers 0 and 1000 as these precincts do not exist and we replaced null values in all columns by unknown. We featured engineering on one of the columns by grouping similar attributes in these fields together using a function. We then exported that dataset into csv form and uploaded it into Power BI. In Power BI, we did some more data manipulation and feature engineering. We created an “Action Time” column by subtracting the date the complaint was received from the date that the complaint was closed. Furthermore, we split a column by a delimiter, and we did feature engineering on one field where we grouped similar values together. Power BI is the software we chose to create our dashboard. In addition, we created a variation of every visualization on the dashboard in Power BI. We used Power BI to create 14 data visualizations and Python to create 6 visuals.

* 1. Explain the specific roles of each software in generating insights from the data. This can be pretty general; I just want to get a sense of what you used and what for.

Python was used in this project to do the bulk of the data cleaning and the data manipulation as well as feature engineering. We also created 6 data visualizations using Python. Power BI was used to do minimal data manipulation and feature engineering and we used Power BI to create 14 data visualizations. 7 of these visualizations are a part of the dashboard and the other 7 are a variation of those visuals.

# **2. Dataset Description**

1. Provide a detailed description of the dataset, including its source, size, and key variables.

This dataset was collected by the NYPD between the years 1985 and January 2020. It records contain each of the allegations made against NYPD officers and has been filtered to only allegations made against those who were still on the force as of June 2020. This dataset contains 33,358 allegations with 31 total features. Some of the key variables that will be shown in our visualizations are date the allegation received, the date the allegation was closed, the shield number of the police personnel, the precinct location, the rank of the police at the incident and now, complainant gender, complainant ethnicity, allegation, contact reason, board disposition etc.

1. Explain why this dataset was chosen and how it is relevant to the field of business analytics.

This dataset was chosen because it was a topic that is both relevant and interesting to look into. Also, the police department, especially in New York City, is well known in the United States. Therefore, we thought this was something that we could analyze in more detail. This is relevant to the field of business analytics because it has a large amount of data ranging from 1985-2020, so this is something that can be made sense of if visualized in the right way. Essentially, individuals may want to know the officer who had the most complaints for example, and just giving an individual a spreadsheet doesn’t allow them to see this information easily. Therefore, choosing this dataset to create both a dashboard and several other visualizations allows information like this to be expressed in a clear and concise way.

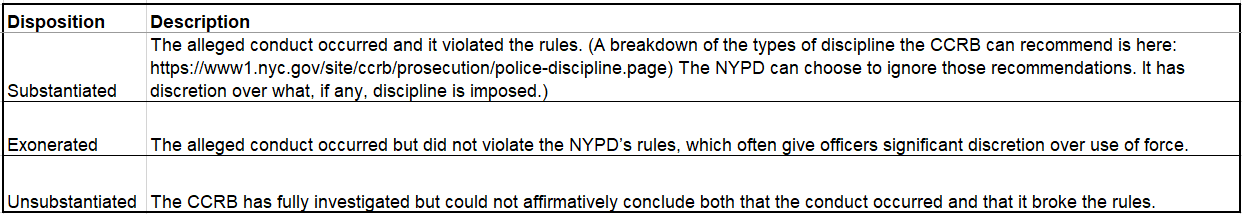
1. What is the primary focus or theme of the dataset you have chosen for your data visualization project?

The primary focus or theme of the dataset we chose is to explore the civilian complaints against New York City police officers. This project explores the claims about officers that were previously kept hidden from the public for decades. Through the use of several visualizations, we discovered patterns within the New York City police department.

1. Provide any required background information for a reader to be able to understand the data/data visualization project (i.e. are there some domain words or variables that someone might not know?) If so it may be advantageous to include a glossary of some sort.

Some required background information for the reader to understand this project and the dataset used is that each record in the dataset is a complaint against the NYPD. In addition, the dataset includes only closed cases and only complaints against officers who as of 2020 still work for the NYPD. This means that if an officer was let go from the force because of a complaint against them, this complaint would not be listed as a record in the dataset. In addition, the field name “fado\_type” was used as an essential part of our dashboard, FADO type means “complaints” so this field identifies the category of misconduct.

Some additional glossary is the board disposition which represents Finding by the CCRB (Civilian Complaint Review Board): Unsubstantiated, substantiated and exonerated. The meaning of these terms is below:



1. What is the source of this dataset? If it is publicly available, please provide the link, if it is not, please specify how you accessed the data and who gave you approval to use.

The source of this dataset is the NYPD. It was provided to ProPublica after a request to the “New York City’s Civilian Complaint Review Board”

https://www.propublica.org/datastore/dataset/civilian-complaints-against-new-york- city-police-officers

To further investigate the precinct field, this data was combined with precinct details publicly available on NYC.gov. This allows us to add a more commonly understood location field. We used the precinct number on the above dataset and merged it with precinct number on the below dataset to get the address and the location of the precinct.

Precincts - NYPD (nyc.gov)

1. In addition, to the best of your knowledge, explain how the data was collected or the nature of the original data.

This data was collected by the NYPD and provided from “New York City’s Civilian Complaint Review Board” to ProPublica to us. ProPublica did some filtering on the data before releasing it. The provided data was only those that involved an officer who was still active in June 2020 and ranged from 1985 to Jan 2020.

1. What interested you in using this dataset? For example, is it the subject matter, the data itself? Why do you think this dataset makes for an interesting story for a dashboard?

Something that interested us about this dataset was the topic matter of reported complaints from NY civilians about police officers. We felt that it could provide interesting and possibly new perspectives on what has become an increasingly hot topic over the past few years. This data was also remarkably clean and provided all of the many types of data that we were looking to get for this project in order to create visualizations.

1. Provide a summary of the number of data features (variables) in the data, such as data type (categorical, numeric, etc.) and a description of what each represents and give a few examples of possible values/format for each feature. Note you can choose to provide this information in any form you desire, I demonstrated with a table earlier in this document.

|  |  |  |  |
| --- | --- | --- | --- |
| **field name** | **Data type** | **description** | **Example** |
| unique\_mos\_id | Nominal Categorical | unique ID of the officer ("member of service") | 10004, 10007 |
| first\_name | Nominal Categorical | Officer's first name | Jonathan, Paula |
| last\_name | Nominal Categorical | Officer's last name | Ruiz, Smith |
| command\_now | Nominal Categorical | Officer's command assignment as of July 2020 | 078 PCT, 079 PCT |
| complaint\_id | Ordinal Categorical | Unique ID of the complaint | 42835, 24601 |
| month\_received | Ordinal Categorical | Month the complaint was received by CCRB | 7, 11 |
| year\_received | Ordinal Categorical | Year the complaint was received by CCRB | 2019, 2015 |
| month\_closed | Ordinal Categorical | Month the complaint investigation was closed by CCRB | 5, 3 |
| year\_closed | Ordinal Categorical | Year the complaint investigation was closed by CCRB | 2020, 2017 |
| command\_at\_incident | Nominal Categorical | Officer's command assignment at the time of the incident | 078 PCT, PBBS |
| rank\_abbrev\_incident | Nominal Categorical | Officer's rank at the time of the incident, abbreviation | POM, SGT |
| rank\_abbrev\_now | Nominal Categorical | Officer's rank as of July 2020, abbreviation | POM, POF |
| rank\_now | Nominal Categorical | Officer's rank as of July 2020 | Police Officer, Sergeant |
| rank\_incident | Nominal Categorical | Officer's rank at the time of the incident | Police Officer, Detective |
| mos\_ethnicity | Nominal Categorical | Officer's ethnicity | Hispanic, Black |
| mos\_gender | Nominal Categorical | Officer's gender | M, F |
| mos\_age\_incident | Continuous Quantitative | Officer's age at time of incident | 32, 39 |
| complainant\_ethnicity | Nominal Categorical | Complainant's ethnicity | Black, White |
| complainant\_gender | Nominal Categorical | Complainant's gender | Female, Male |
| complainant\_age\_incident | Continuous Quantitative | Complainant's age at time of incident | 38, 34 |
| fado\_type | Nominal Categorical | Top-level category of complaint | Abuse of Authority, Discourtesy |
| allegation | Nominal Categorical | Specific category of complaint | Frisk, Physical force |
| precinct | Nominal Categorical | Precinct associated with the complaint | 66, 78 |
| contact\_reason | Nominal Categorical | Reason officer made contact with complainant | Moving Violation, Report-dispute |
| outcome\_description | Nominal Categorical | Outcome of the contact between office and complainant | Moving violation summons issued, Arrest – other violation/crime |
| board\_disposition | Nominal Categorical | Finding by the CCRB | Substantiated (Formalized Training), Unsubstantiated |
| Precinct | Nominal Categorical | Precinct associated with the complaint | 66th Precinct, 65th Precinct |
| Phone | Nominal Categorical | Phone number of the precinct | 718-636-6411 |
| Address | Nominal Categorical | Address of the precinct | 65 6th Avenue, 2820 Snyder Avenue |
| district | Nominal Categorical | District of the precinct | Brooklyn, Manhattan |

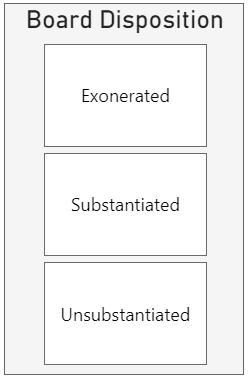
1. Did you perform any preprocessing on the dataset before using it for your data visualization / in order to make a visual in your project? (e.g., cleaning, feature engineering, etc.) If so, could you briefly describe the preprocessing steps and rationale behind them?

Yes, we did preprocessing on the dataset in both Python and Power BI before using it for our data visualizations. We used Python for data manipulation and data cleaning. The dataset contained precinct numbers but not the locations of these precincts. We found a different dataset on “nyc.gov” that indicated the precinct number and the location associated with it. We then merged the precinct location dataset with precinct number on the civilian complaints dataset to get the address and the location of the precinct. We also found the date was separated into day, month, and year fields. We concatenated these fields to create one datetime field. We remove precinct numbers 0 and 1000 as these precincts do not exist and we replaced null values in all columns by unknown. We featured engineering on one of the columns by grouping similar attributes in these fields together using a function. We then exported that dataset into csv form and uploaded it into Power BI. In Power BI, we did some more data manipulation and feature engineering. We created an “Action Time” column by subtracting the date the complaint was received from the date that the complaint was closed. Furthermore, we split a column by a delimiter, and we did feature engineering on one field where we grouped similar values together. In addition, there are only 4 data points that represent January 2020 so we decided to filter out 2020 data from our dataset. Lastly, we created a field which is the total number of all substantiated complaints. This was created because in order to get the waffle chart percentages to add up, we had to create a maximum field.

# **3. Data Visualizations and Final Dashboard Elements**

## Filters and Interactivity

### F1. Board Disposition



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

The filter displayed reflects the board disposition which represents finding by the CCRB (Civilian Complaint Review Board).

* + Why did you display the data this way?

We decided to display the board disposition as a filter because we thought it would be interesting to view how each one of the three findings to the complaints change the visuals on the dashboard. The viewers will be able to gain additional insights regarding how variable of the complaints change depending on the outcome decided on by the CCRB.

* + Was any data manipulation required to make this data visualization?

Yes, we did data manipulation on the column that represents the board disposition. The column stated the finding and a description of the finding in parenthesis. We split this column into two by the space delimiter. One is the actual finding which is the one we see in the filter and the other is the board disposition description.

* + What other ways could you have visualized this data, why did you choose the method you did?

We could have visualized the data behind this filter in a pie chart or a donut chart and then used the interactive feature in Power BI by selecting any of the slices and highlighting the rest of the dashboard. We chose to go with a filter instead because it is cleaner than using a pie/donut chart.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

Yes, the feature “substantiated” played an important role in the dashboard because it represents the complaints that actually occurred and deemed by the CCRB a misconduct that violated the rules of police discipline. This feature as well as the other options helps us, and the viewer finds patterns in the types of allegations depending on the finding. For example, we found the top 1 allegation was stopping civilians which was deemed as an abuse of authority. In addition, even though the complaints are trending upwards, the receiving substantiated complaints reached its peak in 2015. Furthermore, in the same way that substantiated findings played a vital role in telling the data story, so did unsubstantiated and exonerated findings.

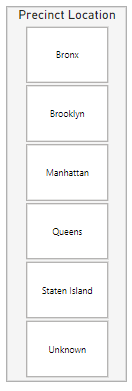
* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

We used fined tuned colors for this tile filter. This helps with decreasing the clutter that a filter might add to the dashboard and allows the filter visualization look clean and clear. Both filters are on the left-hand side of the dashboard. The reason we decided to place them on the left (upper and lower) because we know that upper is where the viewer will immediately draw attention to and then the lower because, although it is in the neutral zone, since the filter will catch the eye of the viewer because of both of their location. This gives the option to the viewer to filter on any of the variables if they would like to deeper into the data.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

We utilized Gestalt principle of enclosure in different elements in the filter. Creating a border around each board disposition helps the viewer group each one separately and it provides a clear identification to distinguish the different variables from each other especially when selecting one of them. In addition, once a tile is selected the color of the tile changes which also brings up Gestalt principle of similarity because the user/viewer can differentiate between which filters are selected and the ones that are not selected.

### F2. Precinct Location



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

The filter displayed reflects the Precinct Location which represents the various locations of different precincts of the police personnel in question.

* + Why did you display the data this way?

We decided to display the precinct location as a filter because we thought it would be interesting to view how each one of the locations that the police personnel is stationed at change the other visuals on the dashboard. The viewers will gain additional insights regarding how variables of the complaints change depending on the precinct's location.

* + Was any data manipulation required to make this data visualization?

Originally, the dataset only displayed the precinct number but not the location of the precinct. We found a new dataset on nyc.gov that displays the different precinct numbers associated with the NYPD as well as their locations and address. In python, we did a left join on the precinct number in our original dataset to the precinct number in the precinct location dataset which allowed us to retrieve the precinct location and make up a part of the original dataset. This also allowed us to use precinct location in our analysis rather the precinct number.

* + What other ways could you have visualized this data, why did you choose the method you did?

We could have visualized the data behind this filter in a pie chart or a donut chart and then used the interactive feature in Power BI by selecting any of the slices and highlighting the rest of the dashboard. We chose to go with a filter instead because it is cleaner than using a pie/donut chart. We also created a heat map that shows the number of complaints by location. Technically, we can use the interactive highlight feature to filter on a specific location using the heat map, but we decided to go with a filter. The reason is because when we click on the location in the heat map, only the associated color with the location will be highlighted. However, when we select on one of the tiles in the filter, not only does it filter on all the visuals, but also the heat map zooms into the location selected. This way, the viewer can easily visualize the distribution of complaints in any specific location.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

The location of NYPD personnel in which complaints were filed played a vital role in the dashboard and storytelling. In our analysis, we found that precincts located in Brooklyn received the most complaints over the years. This is crucial to show especially when presenting from NYPD perspective. This helps the NYPD understand which locations have been receiving the most complaints as well as what are the complaints characteristic specific to each precinct.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

We used fined tuned colors for this tile filter. This helps with decreasing the clutter that a filter might add to the dashboard and allows the filter visualization look clean and clear. Both filters are on the left-hand side of the dashboard. The reason we decided to place them on the left (upper and lower) because we know that upper is where the viewer will immediately draw attention to and then the lower because, although it is in the neutral zone, since the filter will catch the eye of the viewer because of both of their location. This gives the option to the viewer to filter on any of the variables if they would like to deeper into the data.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

We utilized Gestalt principle of enclosure in different elements in the filter. Creating a border around each location helps the viewer group each one separately and it provides a clear identification to distinguish the different variables from each other especially when selecting one of them. In addition, once a tile is selected the color of the tile changes which also brings up Gestalt principle of similarity because the user/viewer can differentiate between which filters are selected and the ones that are not selected.

### F3. Data Source Button



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

The interactivity displayed is a button that directs users to the website where the data utilized in this dashboard was sourced from. The button is an image of the NYPD shield.

* + Why did you display the data this way?

We decided to display this interactivity in the way we did because we thought users should have a quick and easy way to allow viewers to reach not only the data shown in the dashboard but also lengthy information about the data’s collection and originally state. We believed it would be impractical to put this information in plain text on the dashboard as it would take away from the storytelling and visuals.

* + Was any data manipulation required to make this data visualization?

There was no manipulation done to the data in order to create this visualization.

* + What other ways could you have visualized this data, why did you choose the method you did?

In order to show the information that viewers can access when using the hyperlink, another option would have been to write a brief summary of the information and place it somewhere on the dashboard. We decided against this option because as stated above this would take up a large amount of space on the dashboard and take away from the overall storytelling and visuals.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

The location of NYPD personnel in which complaints were filed played a vital role in the dashboard and storytelling. In our analysis, we found that precincts located in Brooklyn received the most complaints over the years. This is crucial to show especially when presenting from NYPD perspective. This helps the NYPD understand which locations have been receiving the most complaints as well as what are the complaints characteristic specific to each precinct.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

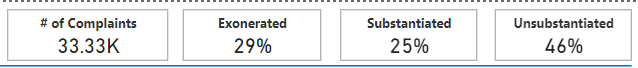
We used fined tuned colors for this tile filter. This helps with decreasing the clutter that a filter might add to the dashboard and allows the filter visualization look clean and clear. Both filters are on the left-hand side of the dashboard. The reason we decided to place them on the left (upper and lower) because we know that upper is where the viewer will immediately draw attention to and then the lower because, although it is in the neutral zone, since the filter will catch the eye of the viewer because of both of their location. This gives the option to the viewer to filter on any of the variables if they would like to deeper into the data.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

We utilized Gestalt principle of enclosure in different elements in the filter. Creating a border around each location helps the viewer group each one separately and it provides a clear identification to distinguish the different variables from each other especially when selecting one of them. In addition, once a tile is selected the color of the tile changes which also brings up Gestalt principle of similarity because the user/viewer can differentiate between which filters are selected and the ones that are not selected.

## Dashboard Visualizations

### D1. Number of Allegation: Board Disposition – Score Cards



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

The data that is being shown in these 4 score cards are the number of complaints filed by civilians against the NYPD and the percentage of the number of complaints that were deemed exonerated, substantiated and unsubstantiated by the CCRB. The number of complaints and percentages reference to the time range of 1958 to 2019. The three score cards showing exonerated, substantiated and unsubstantiated percentages are calculated by finding the number of complaints in each category and dividing it by the total number of complaints which is shown in the score card.

* + Why did you display the data this way?

We chose to display the data in score cards that are placed at the top of the dashboard because we aim to highlight these numbers. In addition, we would like for the viewer to keep the number of complaints and the substantiated, unsubstantiated, and exonerated rates in mind when viewing the dashboard.

* + Was any data manipulation required to make this data visualization?

Yes, we did data manipulation on the column that represents the board disposition. The column stated the finding and a description of the finding in parenthesis. We split this column into two by the space delimiter. One is the actual finding which is the one we see in the last 3 score cards and the other is the board disposition description. In addition, we used 3 different measures to calculate the percentage of each board disposition by the total number of complaints.

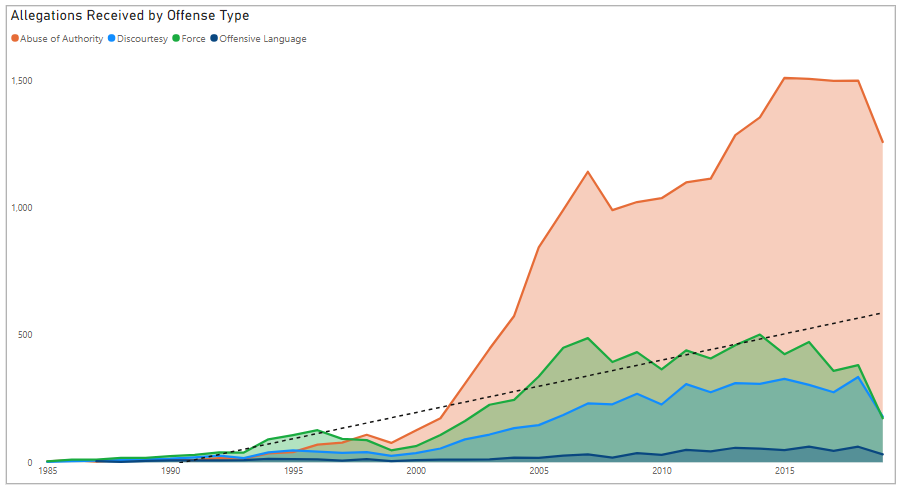
* + What other ways could you have visualized this data, why did you choose the method you did?

There are multiple ways we could have visualized this data, for example we could have used a pie chart, donut chart or even a bar graph. We chose to go with score cards instead is because not only do we aim to highlight these percentages when showing other factors of the complaints, but also we would like for the viewer to keep these numbers in mind when digging deeper into other elements of the dashboard.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

This visualization was chosen to be on our final dashboard because it is important for the viewer to keep in mind that out of the 33.33K complaints received from 1958 to 2019, only 25% of them were substantiated meaning that the allegation brought by the civilian actually occurred and deemed by the CCRB a misconduct that violated the rules. Knowing that the substantiated disposition is the least of the total allegations’ board disposition, the majority of these cases were unsubstantiated. This means 46% of the allegations received from 1958 to 2019 did occur but the investigation done by the CCRB had reached no result regarding whether the complaint against the police did break a rule. In our dashboard, we aim to not only dive deeper into the increasing number of complaints against the police but also to show the outcome of these complaints in our analysis, which these score cards clearly do.

### D2. Allegations Received by Offense Type – Area Chart



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

This area chart is located on the upper left side of the dashboard. Since it is an area chart, the area beneath each line is shaded to show the density of the values. This area chart shows the allegations received by offense type. The number of allegations for each of the offense types is shown on the y axis. The offense types are displayed on the legend located on the top of the visual. The 4 offense types shown are: Abuse of Authority, Discourtesy, Force, and Offensive Language. The x axis shows the year distribution from 1985 to 2019. There is also a trendline present to highlight the overall trend of the data.

* + Why did you display the data this way?

This data was displayed this way because we wanted to incorporate a time series element in the dashboard and we thought that it would be interesting to explore how the allegations and offense types are related and how they fluctuate throughout the years. Therefore, a line chart captures this data in a way that is easy to clearly see for each year.

* + Was any data manipulation required to make this data visualization?

For this visualization, a filter was applied to eliminate the year 2020 from date\_received category. This was filtered because there didn’t seem to be enough information for this year since some of the cases are still open because they are from recently.

* + What other ways could you have visualized this data, why did you choose the method you did?

Other ways we could have visualized this data is through a basic line chart or even a stacked bar chart that showed the allegation and offense type for each of the years. We decided to display the data in this area chart because this seemed to be the most logical way to show the data. This is especially because the years displayed on the x axis make it intuitive to see the trend and it is also easy to see the allegations for each of the offense types and how they continue for each of the years.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

The offense types seem to play a crucial role in the storytelling for this particular visual. This is because each offense type of the 4 offense types have its own line and area shaded within it. Based on these lines, it is clear that the line displaying the allegation count for abuse of authority is highest. The highest allegations received for abuse of authority peaked in 2015 with a total of 1511. Therefore, this indicates that there was a high number of allegations received based on abuse of authority indicating that many individuals felt that there was an improper use of a position of power.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

In this area chart, the color scheme used is a bright color scheme that separates the colors in ways that are distinct from one another. Each line for the offense type categories are represented in a dark color and shaded in a lighter color to indicate its density. The colors include orange, green, blue and dark blue. These colors were used in this way because they are easily distinguishable from one another and they do not overlap in ways that make any category unreadable. It’s position is located in the upper right hand corner because it catches the reader’s attention and it displays important information. Also, size was used by ensuring that the text, such as the x & y axis were large enough to read but still small enough where the graph is the largest object for the visual.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

In this visual, continuity was used because the lines are a continuous flow and is easy for viewers to see and follow the trend. In addition, the trendline displayed on the chart indicates continuity because although it is dotted, it allows the reader to see the overall trend of the line and see that it is traveling in an upward direction. We also utilized Gestalt principle of similarity where we used a different color for each offense category which helps the viewer group these different categories separately.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

This visual is crucial for the story we are trying to tell in our dashboard. This visual shows us that the allegations received toward police personnel have been on an increasing trend. This is important because from the NYPD standpoint, we would aim to decrease the allegations received. The basis of our dashboard is to dig into the increasing allegations received by civilians in NYC. In addition, this visual digs into the trend of each offense type which is also vital to our data story. This tells us that the driving factor behind the increasing amount allegations received from 1958 to 2019 is the abuse of authority.

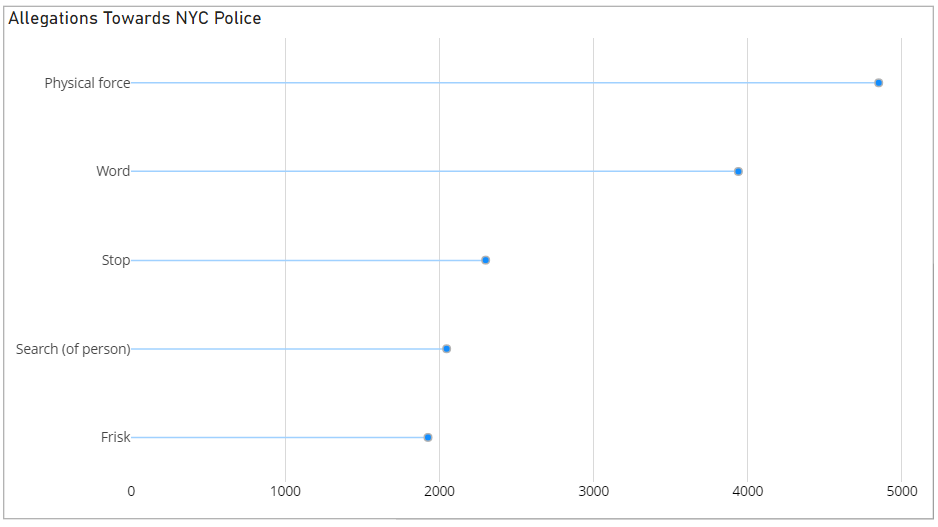
* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?

We put this visual in the upper left corner, in the 1st quadrant of the dashboard. The trend of the allegations over time as well as the offence type is the primary focus of our story. We wanted to focus on the growing number of allegations, specifically abuse of authority, which is the driving factor behind the increase of allegations. This was strategically placed because it is known that the upper left corner is the location in which the viewer’s eyes are drawn to.

* + What color or other aesthetics did you choose for this dashboard and why?
    - How do these style choices relate to the story you aim to tell and why?

This visualization kept the same colors as used in the other dashboard elements. Our two primary colors were blue and orange. The blue is a common color associated with the police and orange is the complementary color. These two colors work well together and allow for clear differentiation among categories. All additional colors are different enough from one another to prevent any confusion between offense types.

### D3. Allegations Towards NYC Police – Lollipop Bar Chart



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

The data shown is the count of allegations made against NYPD, categorized by the allegation type. This chart shows only the top 5 allegation types. Allegation type is on the y axis and count is the x axis.

* + Why did you display the data this way?

This data was chosen to be displayed in this manner because we wanted to visualize the most frequent allegation types and compare them easily. We felt a lollipop bar chart would be best for this data as it allows for easy comparison of a small number of categories when dealing with a single data point such as count.

* + Was any data manipulation required to make this data visualization?

The only data manipulation needed to create this visualization is the filtering of allegation types to the top 5 based on their frequency in the dataset. These were ‘Physical force’, ‘Word’, ‘Stop’, ‘Search (of person)’, and ‘Frisk’.

* + What other ways could you have visualized this data, why did you choose the method you did?

One way that we could have visualized this data besides a lollipop bar chart would be a standard bar chart. We believe this would be an ill-advised method of visualization for this specific data because when compared to a lollipop bar chart, a standard bar chart is slightly more difficult to discern small differences between categories which is one of the key points of this visual.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

The most important feature to this story of this visual is the type of allegation. Of the top 5 allegation types ‘Physical force’ has a solid lead in frequency with nearly 5000 occurrences. The second greatest in count is ‘Word’ at about 4000 allegations with the next 3 closest only having around 2000 allegations. This is important to our storytelling because it shows viewers that physical allegation types (force, search of person, and frisk) have a strong presence in the top 5.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

The primary color in this visualization was blue, with a light being used for the line and a darker utilized for the marker. Our purpose for the use of blue in this visual is that blue is our dashboard’s primary color and also a color typically associated with the police, including the NYPD. The chart is on a white background to highlight the chart and text. The chart’s title is black to further highlight it against the background, and to set it apart from the axis labels which are a light gray. This visual is located in the middle-left region of the dashboard. This is because it is directly related to the visual above it which deals with the count of allegations received over time categorized into offense types. The size of the visual is the smallest on our dashboard. This is because unlike many of our other visuals it does require a large amount of space to see and understand its data.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

We used Gestalt principle of connection in this visual. Connection was utilized in this visual by the connection made between the allegation using a line to the marker at the end. This helped group the allegations together as connected objects are seen as members of the same group.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

When showing the increasing number of allegations against the NYPD, it is important to dig into the top 5 allegations that the civilians filed complaints about. By looking at this chart, we can easily identify the top 1 allegation that was filed a complaint about, the allegation is the use of physical force by NYPD personnel.

* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?

We designed the dashboard so that this visual would be located in the middle-left region of the dashboard. This is because it is directly related to the visual above it (D1) which deals with the count of allegations received over time categorized into offense types

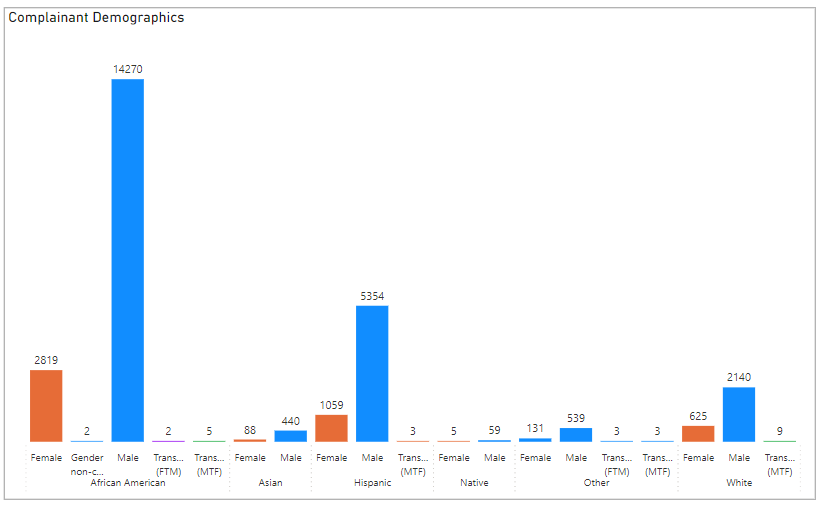
* + What color or other aesthetics did you choose for this dashboard and why?

The primary color in this visualization was blue, with a light being used for the line and a darker utilized for the marker. Our purpose for the use of blue in this visual is that blue is our dashboard’s primary color and also a color typically associated with the police, including the NYPD. The chart is on a white background to highlight the chart and text. The chart’s title is black to further highlight it against the background, and to set it apart from the axis labels which are a light gray.

* + - How do these style choices relate to the story you aim to tell and why?

This visualization kept the same colors as used in the other dashboard elements. Blue is one of our primary colors because it is commonly associated with the police.

### D4. Complainant Demographics – Drilled Bar Chart



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

This visual is a clustered column chart. The ethnicity of each complainant is displayed on the bottom of the x axis and is split into 6 ethnic groups: African American, Asian, Hispanic, Native American, Other, and White. These ethnic groups are further broken down into categories of gender which varies slightly for each ethnic group. These gender categories include female, gender non-conforming, transman, and transwoman. The count of allegation is what is portrayed on the y axis for each bar indicated by values for each ethnic group category and the subgroup of gender.

* + Why did you display the data this way?

This data was displayed in a clustered column chart because this seemed to be the most effective way to show the allegation count for the category of ethnicity and the subgroup of gender for the complainants. By breaking down the data into hierarchical subgroups, this made it easier to display a wide range of data without it being cluttered or confusing to the audience.

* + Was any data manipulation required to make this data visualization?

Due to the limited amount of space allotted for any dashboard, some data points were chosen not to be displayed on the visualization. For example, for ethnicity, the categories of refused and unknown were filtered out of the visual. This is mainly because these were not significant categories that would propel our story. Similarly, for gender, the categories of not described and unknown were filtered out for this same reason.

* + What other ways could you have visualized this data, why did you choose the method you did?

We maybe could have visualized this data in a radial chart or could have created small multiples. However, we chose to visualize the data in a clustered column chart because it seemed to be the best way to effectively display all the data categories in ways that were easy to understand. In addition, due to the fact that most individuals are familiar with reading column charts, we thought this would allow the audience to understand the chart immediately without taking time to decipher a new chart, such as the radial chart.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

The complainant demographic allegations displayed for ethnicity and gender played a crucial role in the dashboard storytelling. This is because the complainants play a key role in storytelling because without information on them, there likely wouldn’t be a case. Also, it is important to understand their demographics because this can help draw conclusions about any trends or patterns occur for any particular ethnicity or gender. For example, there seems to be the most allegations for African American males as opposed to Hispanic transwomen.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

The color blue was used to match the NYPD colors and therefore the overall dashboard theme. In addition, this color was chosen as well as orange because both colors are bright and will likely draw the audience in. In addition, size was used by ensuring the bars were the largest element on the visual. Also, the font sizes were matching in size while still being legible but not overpowering.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

Proximity was applied for this visualization due to the bars being equally spaced from one another. This was used in this way because it made it easier to read the data and distinguish between each bar representing each category. If the bars were not equidistant from one another, it would be much more difficult to understand each category and what it represented.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

This data visualization was chosen for the final dashboard because it gives more insight into the demographics of the complainants. By displaying the complainant gender and ethnicity, it can help to narrow down which group may have filed the most complaints. Essentially, this clustered column chart makes it is easy to identify the gender for the specific ethnicities.

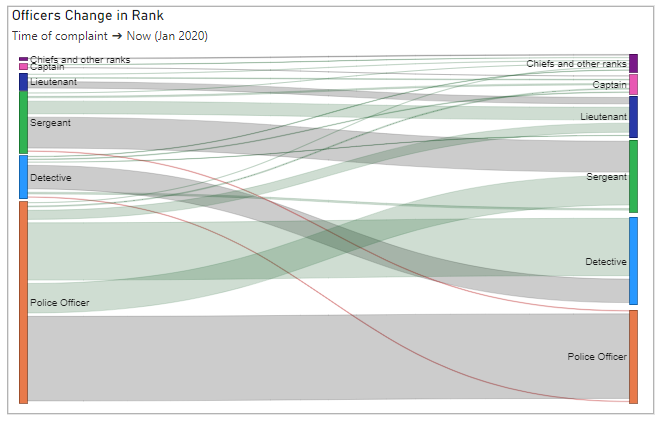
* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?

This visual was placed in the lower left corner of the dashboard because it is something that is important for the viewer to see, but is not necessarily something that is the main purpose of our dashboard. It is essentially a supplement to the overall story of analyzing civilian complaints.

* + What color or other aesthetics did you choose for this dashboard and why?
    - How do these style choices relate to the story you aim to tell and why?

The dashboard color theme is blue to match the NYC police department colors. Therefore, that is why this visual uses both blue. Orange is also used because it is bright and appealing while simultaneously capturing the audiences attention. This relates to the story we aim to tell because the complaints reflect on the NYC police department so therefore it made sense to match their color theme.

### D5. Officers Change in Rank – Sanky Chart



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

This visualization shows the officers rank at the time of the incident and then at the time the data was collected in January 2020. The width of the bars represents the count of how many officers fall into that category. This visualization is responsive to the filters on our dashboard. This allows us to filter down on a subset of our data and see how the distribution changes. This does create some complication as the order will shift, but the colors do remain the same.

* + Why did you display the data this way?

The data was shown in this way to be able to show the relationship between two categorical variables. The original thought was to present this data in an arc chart however the directional aspect was not made as clear. The use of a sanky chart allowed us to also demonstrate a hierarchy. The higher ranks were placed at the top while the lower ranks were at the bottom.

* + Was any data manipulation required to make this data visualization?

This visualization filtered out inspectors and deputy inspectors. This was a very small category and did not classify as easily into the rank structure. The inclusion of the inspector categories created additional clutter on the diagram and did not provide any additional insights. Inspectors also include their own sub ranking which made it difficult to place on the vertical spectrum. For example, within Inspectors, there are Deputy Inspectors and Chief Inspectors. Since it was not able to clearly identify increase of decrease, we chose to remove these two categories from the visualization.

* + What other ways could you have visualized this data, why did you choose the method you did?

This could have also been produced using a sunburst chart. The issue with presenting the data in this way made it difficult to read within each category as the slices quickly became very small. Another option for the visual was a heat map with the start rank on one axis and the current ranking on the other. This would require a lot of eye movement and trying to match categories both up and down. The final option thought of for this visual was the Stream graph which can be seen later in this report. Ultimately, we chose this visual because we thought it was the clearest way to show the relationship between categories. We were able to show a lot of information by showing the ranks in order, color coding to show increases and decreases, as well as the width of the bars representing the count.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

This visualization allowed us to show that the most complaints were about police officers and not frequently about upper-level positions. The height of the bars on the left (at the time of the incident) were highly skewed whereas the height of the bars of the right (Jan 2020) were more normally distributed. This also allowed us to add to the story by demonstrating that complaints did not prevent officers from moving up in the ranking system. There were temporary consequences but in the grand scheme, it appears that the NYPD is rather forgiving. Showing the two decreased lines in red allowed us to clearly show this idea.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

Blue and orange were used for the two largest categories on this visualization as they have been used throughout the entire dashboard. The blue matches to the police departments colors as well as is easy on the eyes. The orange complements the blue very well and allows for the pop to show difference in categories. All categories were represented in their own unique color to show difference. The colors on the left match the same color on the right associated with the same category. Having the same category represented with the same color on both sides allowed for the ranks to be grouped as belonging to one another. Color was also used in the connecting lines. The grey lines represented the rank stayed the same, the green represented the officers ranked increase, and the red lines represented the officers rank decreased. In addition, size was used to represent the count. This allows the reader's attention to be directed to those categories. We are attracted to larger sections which have more meaning as more of our data is associated with that particular category.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

Similarity was used to group the increasing, decreasing, and constant lines together. This was made possible through the use of color. Continuity is used within each ribbon to allow the connections to easily flow from one side of the graph to the other. Finally, enclosure was used in this visual as well as all others on our dashboard to clearly differentiate them from one another.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

This visualization was chosen for the final dashboard because it encompasses a large amount of information in a relatively small amount of space. It also is a different type of visualization which may help intrigue the reader to want to learn more about it. This visual allows the reader to see that the most complaints were received about police officers. This makes sense as they are more frequently interacting with civilians. It also allows the reader to see that not many officers decreased in rank, but a large number saw an increase.

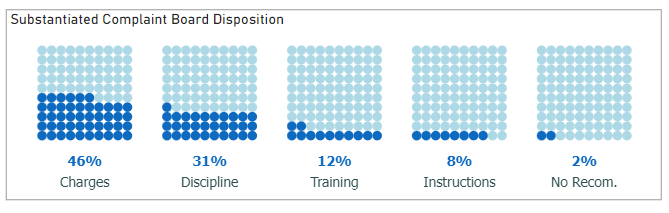
* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?

This visualization was placed in the top middle of the screen. This visualization did require a large amount of space to be able to clearly see the connections from one rank to another. Making this visual any smaller would prevent the user from being able to comprehend useful information. It is important enough to be in the middle because it shows what rank of police the complaints were about.

* + What color or other aesthetics did you choose for this dashboard and why?
    - How do these style choices relate to the story you aim to tell and why?

This visualization kept the same colors as used in the other dashboard elements. Our two primary colors were blue and orange. The blue is a common color associated with the police and orange is the complementary color. These two colors work well together and allow for clear differentiation among categories. All additional colors are different enough from one another to prevent any confusion etween categories. The increases were shown in green as green is commonly associated with a positive attribute. Red similarly was used to show a bad thing, i.e. a demotion. The remaining lines showing a constant rank stayed grey to prevent adding color overload to the dashboard.

### D6. Substantiated Complaint Board Disposition – Waffle Chart



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

The waffle chart above shows the recommendation given by the Civilian Complaint Review Board to the NYPD’s substantiated cases. This means the if the CCRB found that the alleged conduct occurred, this visualization shows their suggestion about how the NYPD should proceed with their police’s misconduct. Note that the NYPD can choose to ignore these recommendations, or they could impose no or a different discipline.

* + Why did you display the data this way?

We decided to display the data in a waffle chart because we want to visualize the recommendations by the CCRB as part of the whole which represent the police personnel whom the CCRB deemed that the allegation against them did occur and found their contact with the civilian as a misconduct. A waffle chart, in this case, was a great option to show the percentage of substantiated allegations in each recommendation category.

* + Was any data manipulation required to make this data visualization?

Yes, we did data manipulation on the column that represents the board disposition. The column stated the finding and a description/recommendation of the finding in parenthesis. We split this column into two by the space delimiter. For this visual we used the column that represents the board disposition recommendation. In addition, we did feature engineering on this column by grouping similar recommendations together. Further, the filters on this visual is only substantiated board disposition finding. We also filtered out the recommendations “Officer Unidentified” because not only does it represent less than 1% of our data, but also it is not a valid recommendation that the NYPD can derive action from. Lastly, we added a total column to the dataset that represents the total number of substantiated cases and added it to the max field in building the waffle chart.

* + What other ways could you have visualized this data, why did you choose the method you did?

The data on the waffle chart could have been visualizated in many different ways as long as the visuals show the parts of a whole. A treemap, donut, and pie charts could all have been used to represent the data in this visualization. There are many reasons why we chose to visualize that data in a waffle chart. First, we often refrain from using a donut chart or a pie chart in a dashboard but specifically in our case, these types of charts would have been hard to read with the limited amount of space on the dashboard. Additionally, we could have used a treemap but a treemap would have caused clutter to our dashboard which is being avoided by using a waffle chart that is clean, clear and concise. Furthermore, a treemap would not have shown all the different recommendation labels, especially for the categories that have a small percentage in relation to the whole.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

This visualization is vital in understanding the recommendations given to the NYPD for these cases. The visual shows that CCRB recommended upon conducting an investigation that 46% of the allegations towards police personnel who showed misconduct to be charged. In addition, discipline was the top 2 recommendation by the CCRB with 31% of these allegations having the same recommendation. Interestingly enough, even though the CCRB had deemed the allegations as substantiated, 2% of them received no recommendation.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

Our purpose for the use of blue in this visual is that blue is our dashboard’s primary color and also a color typically associated with the police, including the NYPD. We used the police color which is the darker shade of blue to represent the part of the whole in the waffle chart. The color that we used for the parts that were outside of the whole is lighter blue. This allowed us to stay consistent with the theme of the dashboard. In addition, we used fine tune colors for the label on this visual which helps with decreasing the clutter and delivering a clean, clear and concise visual.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

Gestalt principle of continuity, closure and similarity were utilized in this visual. We can observe Gestalt principle of similarity being used in the color in each of the waffle charts. Light blue was used to represent the whole and dark blue was used to represent the parts of the whole. This helps the viewer differentiate between the whole and the parts and groups them separately. In addition, continuity and closure are observed when looking at the parts of the whole. The viewer can look at the parts as a continuous flow as well as group the shaded square together even though there is white space separating each square.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

Showing the board disposition recommendations for substantiated cases played an important role in telling the story about the civilian complaints and the outcome of these complaints. In showing not only the outcome but also the recommendation reached for these allegation helps the viewer understand the severity of these allegations as well as how substantiated board recommendation is correlated with the allegations and in doing so, we can uncover patters and insight between elements of the allegations and the recommendation. Furthermore, speaking from a NYPD perspective, they will be able to understand what the recommendations are given by CCRB to their personnel.

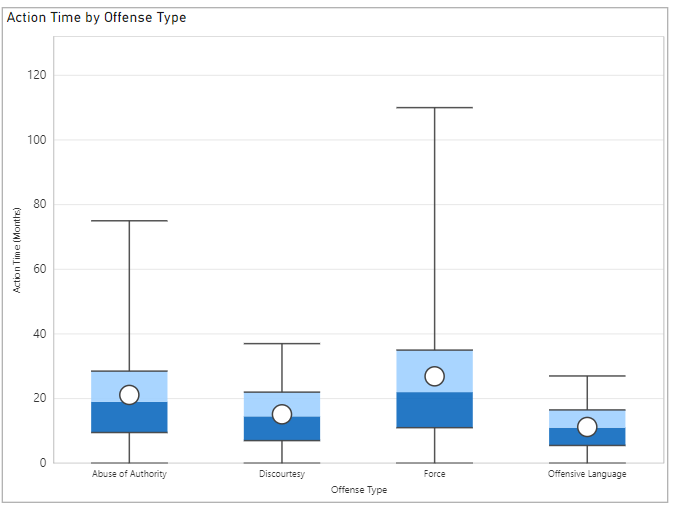
* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?

We decided to put this visual in the lower half of the dashboard between the 3rd and the 4th quadrant. This is a neutral zone because the story of this visual is not the most important of all visuals but is important enough. It is also strategically placed right under the Officer Change in Rank visual. The reason why it was placed right under is because first the viewer can observe that most of the police personnel with allegations filed against them either stayed in the same rank or increased in rank. This would be shocking for most because the thought process would be a decrease in rank, but after showing the board disposition recommendation underneath, we notice that the board recommendations have no effect on the rank change. These two visuals are specific to the police personnel which is why they are placed near each other.

* + What color or other aesthetics did you choose for this dashboard and why?
    - How do these style choices relate to the story you aim to tell and why?

This visual is consistent with other elements of the dashboard. We used dark blue to indicate the parts of the whole of the waffle chart which aligns with the NYPD color. We also used a lighter shade of blue to stay consistent with the theme of the dashboard when showing the whole in the waffle chart. By doing this, the viewer can clearly distinguish between the parts and the whole and also the integrity of the theme of the dashboard is still intact.

### D7. Action Time by Offense Type – Box and Whisker Plot



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

This is a box a whisker plot is displaying the action time by offense type. Essentially, the action time in months is what the y axis is measuring which corresponds to the distribution of each of the 4 box and whisker plots displayed. The x axis is displaying the offense type which are: abuse of authority, discourtesy, force and offensive language.

* + Why did you display the data this way?

The data was displayed in this way because we were interested in comparing the distribution of action time (in months) for offense types. Therefore, we chose a box and whisker plot because is a good way to show a distribution while still comparing categories. Also, we chose to display the data in the form of a box and whisker plot because this plot condenses a lot of data into a small amount of space which showing the overall distribution.

* + Was any data manipulation required to make this data visualization?

Other than the original preprocessing done on the original dataset, there was no data manipulation required to make this visualization.

* + What other ways could you have visualized this data, why did you choose the method you did?

We could have maybe visualized the data as a violin plot because that shows the distribution of categories as well. However, we chose this method because a box and whisker plot clearly defines the mean, median, outliers, etc., to effectively see the distribution. However, a violin plot does not make it as clear to see all of these measures.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

A specific feature that played a particularly crucial role in our dashboard storytelling are the categories of offense type as well as the action time. The offense type gives more insight into what offenses were most commonly attributed to allegations. After looking at these categories of offense type, we can then delve into the action time to understand which types of offenses resulted in the highest action time on average.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

The color blue was used in this visualization to match both the theme of the dashboard and the overall color scheme used throughout the project. This color is also appealing to the audience due to its brightness. In addition, size was used to ensure that each box and whisker plot had the same dimensions so one would not appear more important than another. These box and whisker plots were also positioned vertically instead of horizontally because this appears to be easier to read.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

Proximity was used because each of these box and whisker plots are all equidistant from one another. There is no disparity in the amount of space between the box and whisker plots because that may lead to confusion if there is a wide gap between one plot and a small gap between another plot. In addition, continuity was used because a viewer can see that quartile 3, for example, continues/is connected to the upper whisker.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

This visualization was chosen for the final dashboard because it is simple to understand while simultaneously providing a summary of the top offense types and what the action type time was for each one. This helps to give more information about the civilian complaints and the outcome of these complaints. A viewer can essentially see how long the action time of the offense lasted to see if there are any trends or changes in the future.

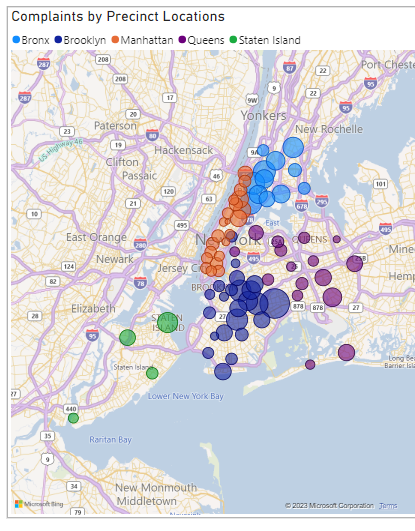
* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?

This visual was placed in the upper right corner of the dashboard because it is something that is important for the viewer to see, but is not the main point to be highlighted. Essentially, this visual provides more information to the viewer about the complaints but should not be the sole main take away from the analysis displayed on the dashboard.

* + What color or other aesthetics did you choose for this dashboard and why?
    - How do these style choices relate to the story you aim to tell and why?

The color of this visual is blue to match the overall theme of the dashboard. We chose this color because it matches the NYC police department colors. By keeping the colors consistent, a viewer can easily understand that the visuals displayed in the dashboard are interrelated to one another.

### D8. Complaints by Precinct Locations – Map



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

This visualization shows the count of complaints for each precinct. This map was made possible through joining the data with their corresponding longitude and latitude coordinates. Color is also used to allow this visualization to show the data by borough. A slicer is created for these boroughs on the left side of our dashboard. Depending on who is viewing our dashboard, they may be interested in only looking at a particular borough.

* + Why did you display the data this way?

We displayed the data in this way because we wanted to show where the complaints came from. The original data had the precincts stored as their appropriate number. Adding the coordinates allowed us to make more sense of the precinct numbers. As an average individual, precinct 78 means absolutely nothing but seeing it on a map is a lot easier to comprehend. This also allowed us to see the cutoffs of the different boroughs which are more commonly known.

* + Was any data manipulation required to make this data visualization?

This visualization required us to join the original dataset with a precinct dataset that included longitudinal and latitudinal coordinated as well as what borough it belonged to. There were some (24) complaints in our dataset that did not have a precinct listed. We labeled these boroughs as unknown. For this visualization we filtered out the unknown category because you cannot plot unknown on a map.

* + What other ways could you have visualized this data, why did you choose the method you did?

This visualization could have been shown in a tree map which still allowed us to separate the precincts by borough. This allowed us to more easily see the top locations, but it was not clear as to where in New York it was associated with. We could have also used a gridded, graphical, or contiguous cartogram for this visualization however all those options increased the chance of distorting the commonly known and perceived map. Ultimately, we chose the basic map to allow the viewer to get a clear understanding of where the complaints came from.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

This visualization allows us to see that the greatest number of complaints are associated with Brooklyn; in particular, borough number 75. The borough number is not clear on the map visualization, but you can see where in Brooklyn it is associated with. You can also see that Manhattan is separated into more precincts which is likely due to its high density. Ther are not many complaints coming from Staten Island.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

Color was used to associate each precinct with its appropriate borough. Position was also used in a similar matter but happened naturally due to the way the boroughs and map are structured. You can see that all Manhatten precincts are next to one another. This same effect would not have happened if we did not plot the values on a map. Size was used to show the count of complaints in each precinct. This allows the reader's attention to be directed towards the highest precinct. The smaller ones do not capture as much attention which makes sense as they are less important in our story.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

Proximity are similarity were used to group the boroughs together. By using color and the precincts being mapped close to one another we avoided having to use enclosure to differentiate from one borough to the next. Enclosure is used around the entire map to differentiate the end of this visualization from the start of next.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

This visualization allows the viewer to see where the complaints are coming from. If the viewer is familiar with the New York area, they will be able to quickly draw association. It also allows us to keep an eye out if a particular attribute causes a change in the map distribution. If one precinct or borough dominates in a particular category, the NYPD may want to investigate further what is happening internally.

* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?

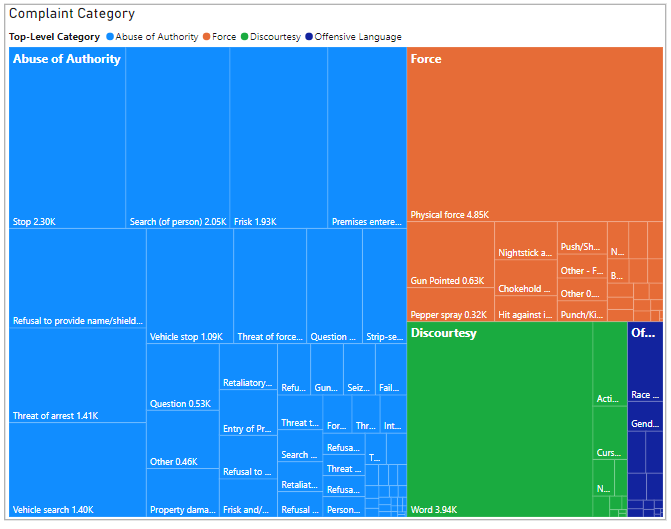
This visualization was placed on the bottom right corner of the dashboard. This prioritizes it last out all the dashboard elements. It is interesting to see where the complaints came from but does not necessarily add to the story about the type of complaints the NYPD are receiving.

* + What color or other aesthetics did you choose for this dashboard and why?
    - How do these style choices relate to the story you aim to tell and why?

This visualization maintained the same categorical colors used through the rest of the dashboard. These colors are easy to differentiate from one another. The darker blue was used as the highest color on this visualization because it stood out more than the light blue used previously. Since the map shows water, the light blue color captures less attention than the darker blue does.

## Additional Visualizations

### V1. Complaint Category – Tree Map



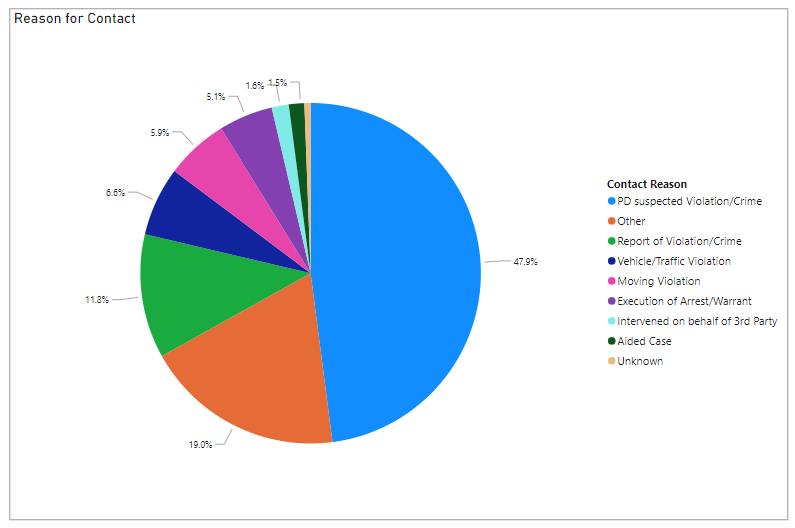
* What data is shown?

This treemap shows the top 4 categories of offense type (fado\_type) as the main category. Then, the allegation for each offense type is displayed within the main category of offense type. The size of the treemap is determined by the count of offense type for each allegation within the offense type main category. Essentially, each offense type is counted, and each count represents the size of the allegation square for the overarching category of offense type. Also the 4 categories of offense types are different sizes based on both the count of offense types and the number of allegations for that offense type. The 4 offense type categories are distinctly different colors: (blue, orange, green and dark blue). The blue matches the overall NYPD theme and the other colors allow readers to group the data into different categories and understand what is being displayed for each.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

This treemap allows us to show the top 4 categories of offense type (fado\_type) as the main category and a subcategory of allegation with the count of offense type as well as the distribution of these categories. In addition, we thought the treemap was a visualization method that was relatively easy to read and was intriguing to the audience. A reader can examine the allegations and offense type which can help to inform them about how the allegations make up the offense type. When analyzing complaints this is useful because it is helpful to understand what allegations were made and how many offense types make up these complaints.

### V2. Reason for Contact – Pie Chart



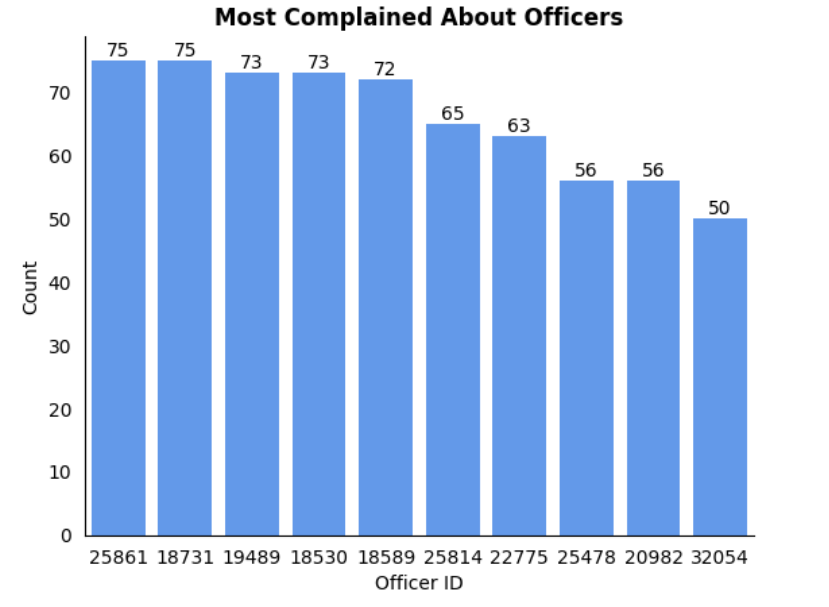
* What data is shown?

This pie chart demonstrates the distribution of various reasons for contact. Each section is labeled with the calculated percentage and the legend on the right shows what reason the color represents. As this is a pie chart, it shows parts of a whole meaning it is accounting for all reasons and complaints in our dataset.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take  
  away from the plot.

Each contact reason has its own slice in the pie chart. The size of the slice is based on the count of complaints that fall into that particular category. The percentage is automatically calculated in PowerBI and is displayed attached to the appropriate section through the use of connection. The colors were carefully selected to ensure that it was considerably easy to differentiate among each category. This plot may be limited based on accessibility limitations. When the reader views this visualization, they can see the largest section is PD suspected of Violation/Crime. This category takes up nearly half of our data at 47.9%. Due to this large amount, we dive further into this category in [visual 12](#_V12._Civilians’_Outcome). It is important to note that this visualization required cleaning based on the 80-20 rule. For this reason, it is not significant that the ‘*other’* category is shown to represent 19% of our data. This category is a combination of a whole bunch of other smaller categories.

### V3. Most Complained About Officers – Bar Chart



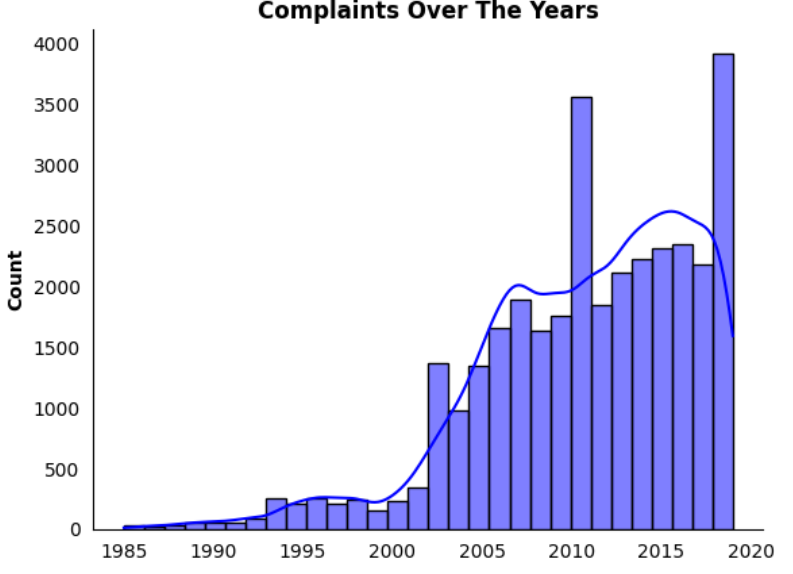
* What data is shown?

This bar plot/column chart shows the top 10 NYC police officers who were the most complained about officers. This chart displays the ID for the top 10 officers by the count of how many complaints each officer received. In addition, data labels are included for the audience to easily see the most complained about officer immediately. The color blue was used to match the theme of the NYC police department as well as the Civilian Complaint Review Board, and also matches the color scheme used in all other visuals. This was done to keep consistency and make the visual appealing.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

The x axis displays the ID for the top 10 officers, while the y axis shows the count of how many complaints each officer received. Based on the counts being displayed as a series of columns in descending order with data labels, allows it allows the reader to see the officer with the most complaints vs the least complaints easily. In addition, this column chart makes it easy to see the officer IDs and understand the trend while still comparing each officer to one another. A reader can also examine the most complained about officers to infer which officers were receiving the most complaints. Seeing the officer with the most complaints may signify that they may need to be monitored more closely to investigate their behavior. Seeing that there are at least 10 officers with more than 50 complaints may signify that a lot of officers are not obeying protocols as they should.

### V4. Complaints Over The Years – Histogram



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

The data that is being shown in this histogram is the number of complaints received by civilians against the NYPD from 1958 to 2019. The visual above shows us that the number of complaints over these years has been increasing. We also see the first peak in 2011 and then the second peak in 2019. The trend of the number of complaints is on an incline.

* + Why did you display the data this way?

The data is being shown in a histogram because it allows us to view the frequency of the distribution of frequency of complaints over the years. In addition, kernel density estimate line has been added on top of the histogram bars to show the shape of the distribution. Furthermore, this histogram is a variation of the area graph of complaints over the year by offense type. Using this variation, we aim to highlight the two peaks in 2011 and 2015.

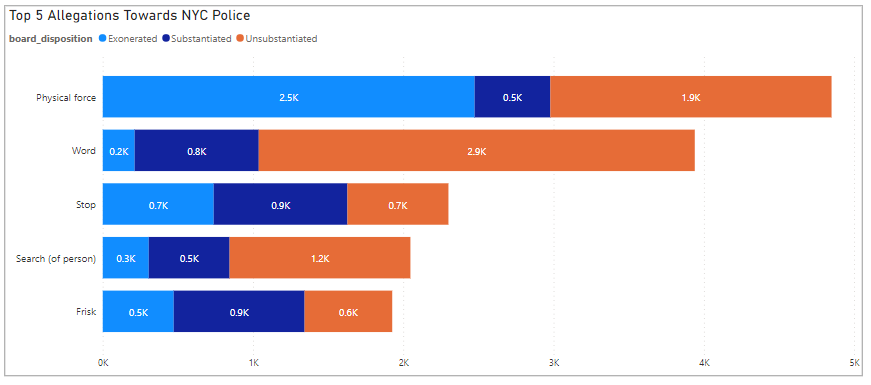
* + Was any data manipulation required to make this data visualization?

The dataset included day complaint receiver, month complaint received, and year complaint received. We concatenated these fields to create one datetime field. We then used this concatenated field to create this histogram.

* + What other ways could you have visualized this data, why did you choose the method you did?

We could have visualized the same data using a line graph, area graph or a bar graph. This histogram is a variation of the area graph of complaints over the year by offense type. Using this variation, we aim to highlight the two peaks in 2011 and 2015 and show the frequency distribution of complaints clearly over these years

### V5. Top 5 Allegations Towards NYC Police – Stacked Bar Chart



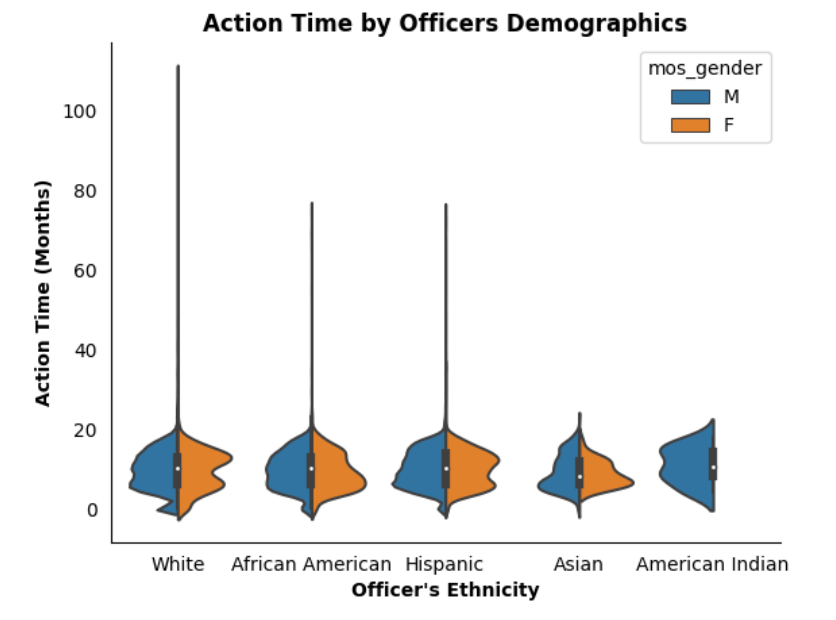
* What data is shown?

This visualization shows the count of how many cases fall into the specified allegation category. It breaks downs each category by the board disposition. Only the top 5 categories are displayed on this bar chart, and they are ordered to be displayed in descending order based on the cumulative count.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take  
  away from the plot.

The x axis shows the count of how many complaints are recorded. They are in one thousand increments and labeled accordingly with a ‘k’. There is no x axis label. The y axis shows the top 5 allegations. Since this is a horizontal bar chart, the width of the bar is used to show the count. To show the 3 different board dispositions, color is used. Each section is clearly labeled with units inside the respective section. This prevents the reader from having to do any estimating. The reader can quickly see that the top allegation towards NYPD is Physical Force. This category also has the highest number of exonerated cases. This means that although the police did use physical force, it did not go against their rules. The reader may be more interested in only looking at substantiated cases in which case they can focus on the dark blue sections. The leading category becomes difficult to single out as stop and frisk both have 0.9k complaints and word is not far behind with 0.8k.

### V6. Action Time by Officers Demographics – Violin Plot



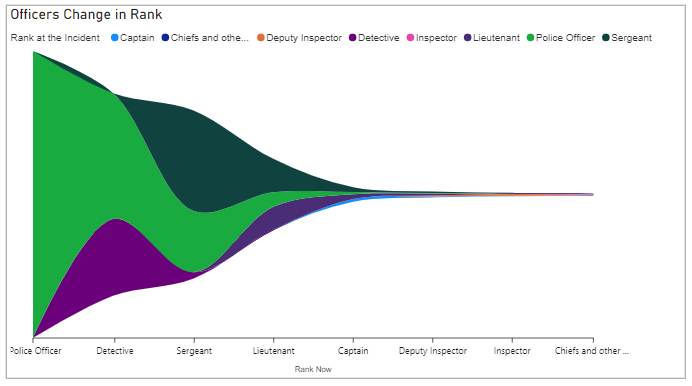
* What data is shown?

This is a violin plot displaying the action time (in months) by police officer ethnicity. These plots are also split by gender (male or female). The hues of the violin plots are colored blue and orange to match the overall theme of the project, since the NYPD color is blue and orange is a bright color that matches the color scheme.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

For this violin plot, the y axis shows the action time (in months) for each officer. In other words, it is displaying how long the officer was part of the police force. The x axis shows the distribution of the officer’s ethnicity (White, African American, Hispanic, Asian, and American Indian). On each side of the violin plot line, it is split into gender (male or female). Male is displayed in blue on the left side of the line, while Female is displayed in orange on the right side of the line. A reader can infer that the action time in months for officer ethnicity of White, African American, Hispanic and Asian all seem to have relatively similar distributions in which the action time ranges from roughly 0-20 months and seems to dip somewhere around 10 months for females. A reader could maybe infer that this is due to there being less females in the police force, so they may leave after about a year.

### V7. Officers Change in Rank – Stream Graph



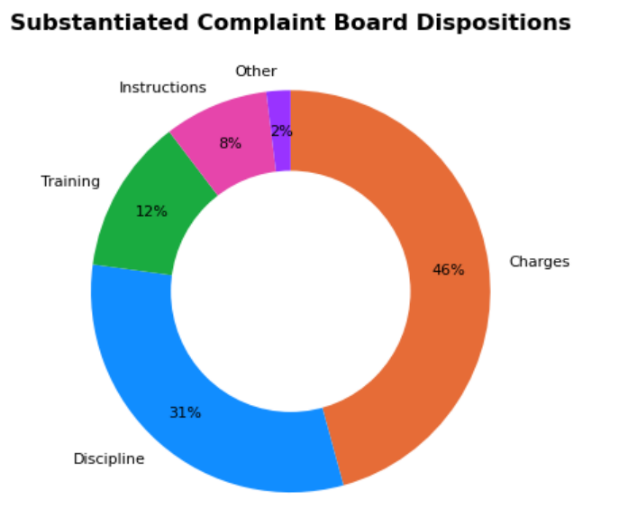
* What data is shown?

The data that is being shown on the stream graph above is the NYPD personnels’ change in rank. The aim in this visual is to show how the NYPD personnels that have a complaint against them changed in rank from the time of the incident when the complaint was filed to when the data was collected which is in January of 2020. This is a variation of the Sankey chart displayed in the dashboard. This chart allows us to not only dig deeper into whether the complaints were against lower rank or higher rank but also, it shows us that regardless of the complaints, did their rank increase or decrease after the incident had occurred.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

In this steam graph, the y-axis represents the ranks of the NYPD personnel after the complaint (January 2020). The x-axis shows the total number of allegations against personnel that had complaints filed against them. The different colors on the chart show the rank of the NYPD personnel in January 2020 post-complaint. We can observe that the majority of the allegations were directed at police officers using green. If we follow the green color on the x-axis, we notice that most of these police officers remained in their rank, while some were promoted to detectives and sergeants. We also see a little of the police officers reach rank of a captain. Even though most of the allegations were directed at police officers, we also observe that sergeants had complaints against them coming in second after police officers by following the dark green color. Most of these sergeants remained in the same rank after the incident, while some of them were demoted to detectives and very few to police officers. Some of these sergeants also got promoted to lieutenant and captain. By following the light purple, we also observe that detectives also had complaints against them. Most of the detectives stayed at the same rank while some were demoted to police officers, and some were promoted to sergeants.

### V8. Substantiated Complaint Board Disposition – Donut Chart



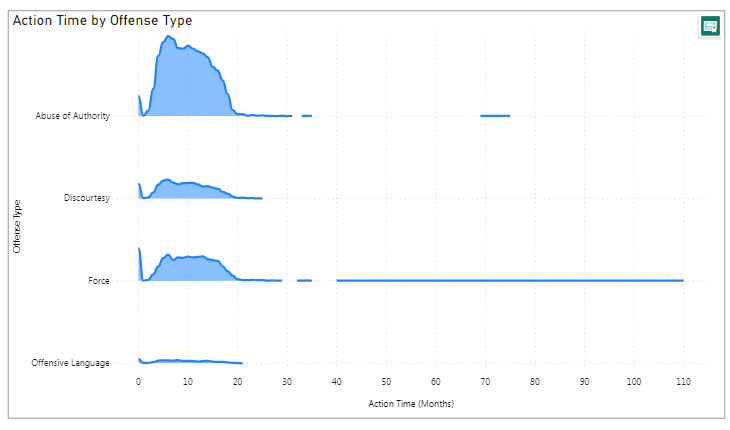
* What data is shown?

This donut chart shows the board dispositions for the complaints that were substantiated, meaning that there was reason to believe exploitation occurred. The board disposition categories are: Charges, Discipline, Training, Instructions, & Other. The percentages within these categories are based on the allegation count for each category that was substantiated. This data includes blue to again match the NYPD colors and the other colors: orange, purple, pink, and green are bright and appealing and also pair well with the blue color scheme.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

The board disposition categories displayed in this donut chart are: Charges, Discipline, Training, Instructions, & Other. The percentages within these categories are based on the allegation count for each category that was substantiated. A reader can infer that the board disposition with the highest number of substantiated allegations was for charges. Therefore, a reader can likely infer that charges are something that likely has some sort of physical evidence, thus making it easier to prove that some form of exploitation occurred.

### V9. Action Time by Offense Type – Ridgeline Plot



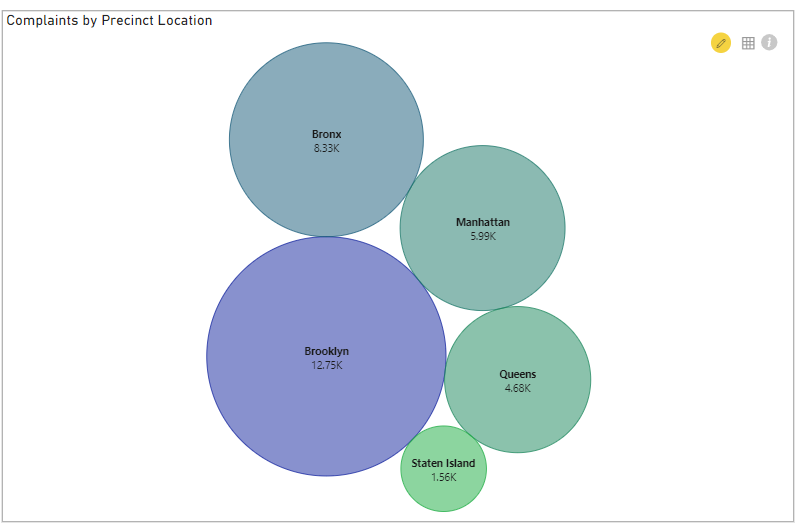
* What data is shown?

This visualization shows the average action time broken down by the various offense types. The action time was calculated using the date the complaint was received and the date it was closed. This ridgeline plot shows a separate distribution for each of the four offense types. The height of the distribution is representative of count similar to a histogram.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

The action time represented in months is shown on the x axis and the offense types are shown on the y axis in alphabetical order. Both axes are labeled, and the plot is titled Action Time by Offense Type. The reader can use the height of each subplot to understand the counts within each category. Abuse of authority is significantly the largest, signally most of our complaints fall into this category. Offensive language barely has any height to the plot suggesting it is our smallest category. The centers of each category are all around 10 months. This initially would suggest that the offense type has no impact on how quickly the complaint is closed. However, when you look further at the spread of each distribution, you can see that force has the greatest range of time going all of the way up to 110 months (over 9 years). Abuse of authority also has a pop up at 70 months (about 6 years). Both of these distributions would require further investigation to see what may have been unique about these complaints.

### V10. Complaints by Precinct Location – Bubble Chart



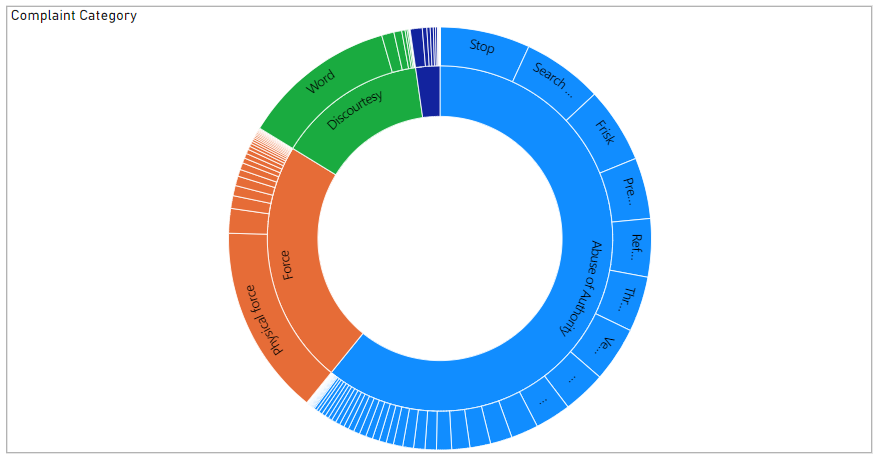
* What data is shown?

This bubble chart shows the count of allegations for the top 5 precinct boroughs in for the NYC police department. The colors range from purple to green which directly corresponds to the size of the bubble. The largest sized bubble is purple and represents the precinct borough with the highest number of allegations, while the smallest bubble is green and represents the precinct borough with the lowest number of allegations.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

This bubble plot displays 5 distinct bubbles with gradient colors ranging from purple to green. This shows the precinct boroughs more clearly. Also, the colors match the values displayed, so Manhattan and Queens for example are relatively close to one another in terms of allegation count, so those are very similar color greens. Also, the color palate has blue to match the NYC police department and the other visuals. In addition, it can also be inferred that each bubble ranges in size based on the count where the largest count has the largest bubble. For a reader, bubble charts are very intuitive to read. From this bubble chart, the reader can see the precinct borough with varying values for each borough. Also, the bubbles are very colorful and simple, so they will likely draw the reader in. In addition, a reader can look at this plot and understand which precinct borough had the most allegations to see where most allegations are concentrated. Understanding which precinct borough has the most allegations can help someone pinpoint which areas have the officers with the most allegations/complaints filed against them. They can then further assess if this is due to the population within the specified city, in this case it is Brooklyn, or if there is a problem internally with the officers.

### V11. Complaint Category – Sunburst Chart



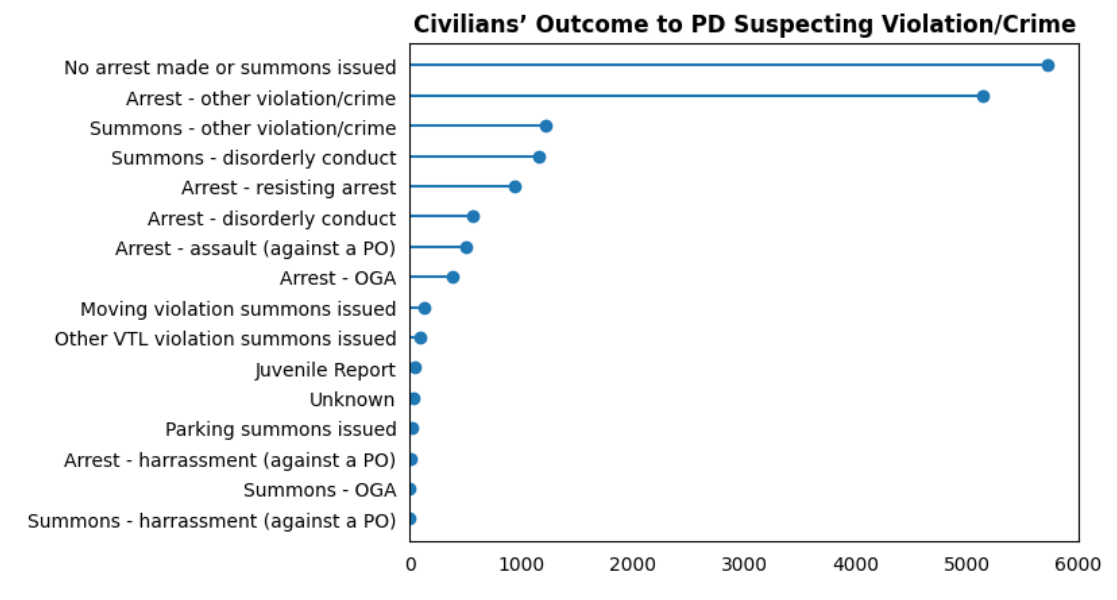
* What data is shown?

This visualization shows counts for fado type and allegation. The fado type represents a top-level category of complaint while the allegation is a more specific category. For this reason, fado type is shown first in the inner circle and the allegation is shown as a secondary level on the outer ring. Both layers show a percentage of the total number of complaints in our dataset.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

This sunburst chart does not have any axis or legends, but does have a clear title, ‘Complaint Category.’ This includes a large number of categories which makes it difficult to understand all the available information in a static image. You are able to quickly see the largest categories which is likely the most interesting concept to the reader. At the top level you can see that nearly two thirds of our data is described as Abuse of Authority. Force and discourtesy take up the majority of the remaining portion of our data. Within Abuse of Authority stop, search, and frisk are the top three specific categories. The top category in force is physical and the top category in discourtesy is word. This is already a large amount of information but there is still a lot more that can be understood if the reader is not looking at the static image. When using this visualization in PowerBI the ability to hover will allow the reader to see all top-level and specific categories within our dataset.

### V12. Civilians’ Outcome to PD Suspecting Violation/Crime - Lollipop Chart



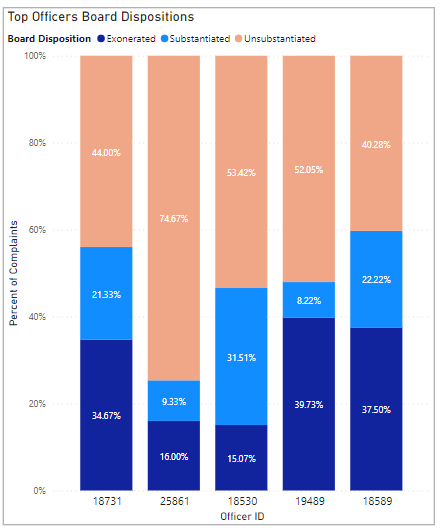
* What data is shown?

The data that is being shown in the lollipop chart above is the civilians’ outcome to police personnel suspecting violation or crime. We noticed from the pie chart showing the reason for the contact of the police toward the civilian that approximately 48% show that PD (police department) suspected a violation or a crime. This prompted us to dive deeper into this contact reason to find out upon the PD suspecting that the civilian is committing a crime or a violation, what measure did the PD take towards the civilian after this suspicion. This is exactly what this visual shows us.

* Describe in sufficient detail what is on the x-axis, y-axis and what the reader can take away from the plot

On the lollipop chart above, the x-axis represents the number of allegations, and the y-axis represents the outcome of the civilian upon the PD suspecting a violation or a crime being committed. The chart shows that out of all the complaints towards police, the top outcome for civilians who were suspected to be committing the violation or a crime is that no arrest or summons was issued. The total number of these outcomes is 5,720. Coming in as the second outcome is that arrest was made by the police due to a violation or a crime. This outcome has a total of 5,146 allegations towards the police. These are the two leading outcomes when the police suspect that the civilian is committing a violation or a crime. The findings in this visual are interesting because the top 2 outcomes are opposite of each other. The first one is that no arrest was made yet the civilian filed a complaint. Moreover, the second leading outcome is that an arrest because the civilian did commit a crime or violation, yet a complaint was filed. Furthermore, it is important to note that only two allegations out of this category of contact reason resulted in a summons due to harassment against the police or obstruction of governmental administration.

### V13. Top Officers Board Dispositions – 100% Stacked Bar Chart



* Describe each element included on the dashboard (e.g., charts, filters, annotations, etc.).
  + What data is shown?

This visualization shows how many complaints the top 5 officers received separated out by the boards disposition. This allows us to get more information than was previously made available in the bar chart without categories. The officer's unique ID is on the x axis and the percentage of complaints is on the y axis. The actual count is displayed within the bar to allow us to easily compare across categories.

* + Why did you display the data this way?

We displayed the data in this way to allow us to see the subcategories distributions for each officer separately. By viewing the data on 100% scale we can see the disposition distribution as part of a whole. Since the bars were all very similar, ranging from 72 – 75, there was not an important difference in seeming them with respective heights.

* + Was any data manipulation required to make this data visualization?

The data was filtered to only show the top 5 officers based on the number of complaints received. The data labels also required us to add a custom field for count of complaints. By default, it was showing the percentages, but the numbers are more direct to understand.

* + What other ways could you have visualized this data, why did you choose the method you did?

This visualization could have also been made possible with a paired bar chart or a bar chart with small multiples. Either of these options would allow all bars to start on the bottom x-axis, which may make comparing officers easier. However, they would have been less compact, and we addressed the issue of comparing across categories by adding in the count label. This method of the 100% stacked bar chart was the ending decision because it shows a lot of information in a limited amount of space, and it is easy to understand through the use of labels. As mentioned previously there were no major differences in the totals from one officer to the next, so the total height was not crucial to the story.

* Are there any specific features /data points that played a particularly crucial role in your dashboard / story telling?
  + Can you explain why?

The light blue sections are the most important because they represent the substantiated cases. It was originally alarming to see that officers were complained about 75 times yet still remain on the force. This visualization allows you to digest that information better by drawing focus mainly to the substantiated cases. The highest officers did not actually have the greatest number of substantiated cases. The officer with the most substantiated cases had 23. Substantiated cases are most important to look at in our story telling because they are the complaints that were proved and violated the rules.

* Identify and explain where preattentive attributes (e.g., color, size, position) were used in your visualizations.
  + Why were they used in the way they were?

Color was used to group the 3 disposition categories together. The primary color was bright blue which was the focus of this visualization. Unsubstantiated cases were in orange to match our theme but were made much lighter to prevent taking away focus.

* Discuss how Gestalt principles (e.g., proximity, similarity, continuity), if any, were applied for better visual organization.
  + Why were they used in the way they were?

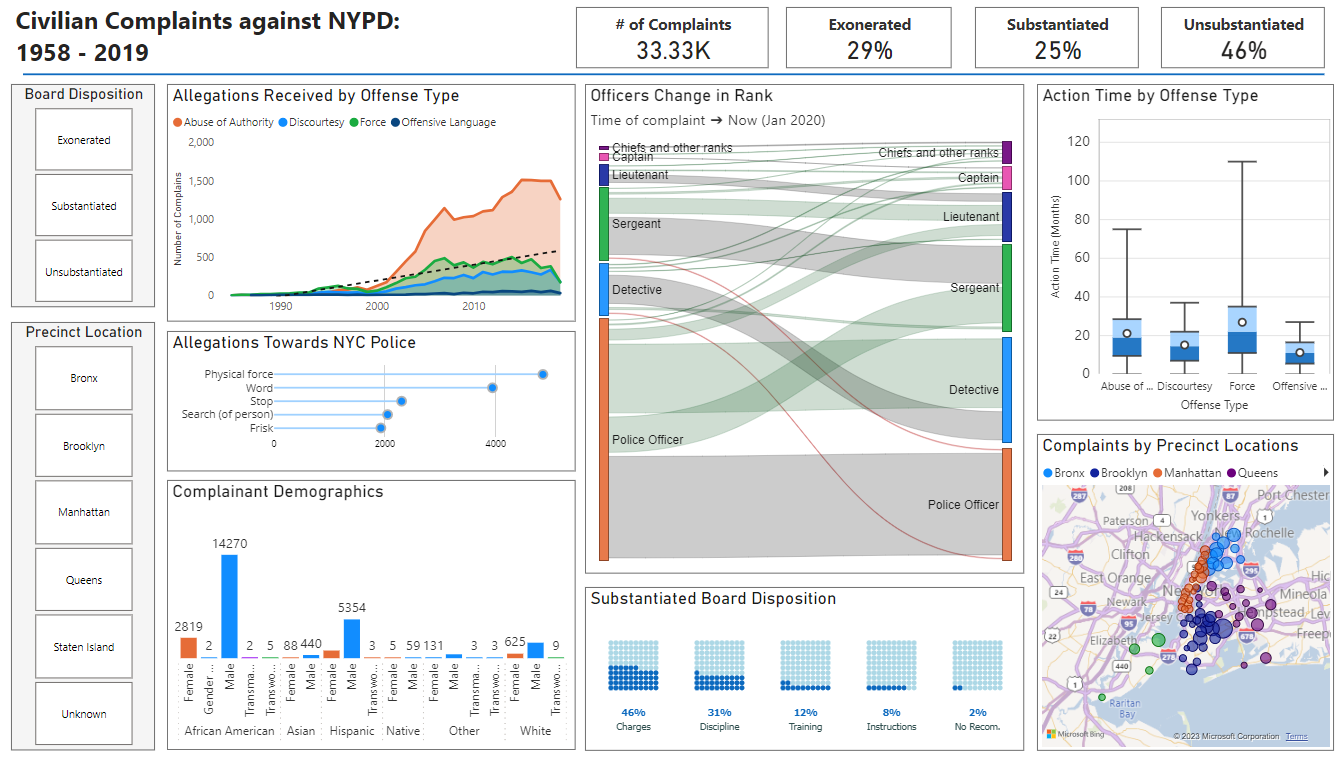
Similarity was used through the use of color to associate the board dispositions categories together. Connection is visible by having the categories stacked on top of one another. This would not have been the case if we used a paired bar chart. Proximity is seen through the labeling of the bars. There is more distance/white space between each ID than there is within an ID. In addition, tick marks are not used but you are able to clearly tell which bar is associated with each ID based on the closeness in location.

* Explain why each of the (at minimum 6) data visualization elements were chosen for the final dashboard and why/ how it contributes to the overall story.

This visualization was not used on the final dashboard because we did not feel as though it was appropriate to be singling out individual officers. Although it is interesting, it does not contribute to the story of the type of complaints being received.

* Describe logic for the layout and overall style choice of the dashboard.
  + Did you put certain visuals in certain locations for a reason?
  + What color or other aesthetics did you choose for this dashboard and why?
    - How do these style choices relate to the story you aim to tell and why?

## Final Dashboard



# 4. Reflection, Future Improvements and Conclusion

* Summarize the key takeaways and insights derived from your dashboard project.

The key takeaways and insights derived from our dashboard project:

1. Complaints against NYPD from 1958 to 2019 have been on an increasing trend. We see the dramatic increase started in 2000 and has been increasing since.
2. The offense type that is the driving factor behind this dramatic increase in complaints over the years is abuse of authority.
3. Only 25% of these complaints were found, by the CCRB, as a misconduct that broke the rules; while the majority, which is about 46% of these allegations were deemed inconclusive.
4. Out of the complaints that were found to be misconduct and broke the rules, the CCRB recommended the NYPD to file charges against these police personnels.
5. Most of the NY civilian complaints were filed against precincts located in Brooklyn.
6. The use of physical force is the top allegation by civilians, and it is the allegation that received the fastest action by the CCRB
7. The top demographic of the complainants are African American males.

* Reflect on the strengths and weaknesses of your dashboard.

The strength of our dashboard is the number of visualizations, their interactivity, and their consistent color theme/design. In addition, the utilization of Gestalt principles of continuity, connection, enclosure and similarity show the strength of our dashboard. In addition, we used trend lines and fine tune colors which helped us use preattentive attributes and reduce clutter while creating a clear, clean and concise dashboard. Some of the weaknesses of the dashboard could be that many of the utilized features had a great deal of options which means we either had to engineer features that simplified the data or only use the highest-ranking data for some of the visualizations. In addition, we were not able to use max line or text to draw the viewers' attention to specific data points because the visuals don’t contain the max line option. Furthermore, if we were to add a constant line or a text, these elements would stay still when using the interactive tools on the dashboard which would cause confusion for the viewer.

* Propose any potential improvements or enhancements you would consider for future iterations.

Although we are satisfied with the outcome of our project, there are still some possible improvements or enhancements that could have made our project even more advanced. In future iterations, maybe we could spend a little more time developing the scope of the project. Although we discussed what we were going to work on for the project and delegated tasks effectively, in the future, we could maybe develop a clearly written scope of the project as well. Developing the scope of the project is very important because it helps the team understand exactly what resources are needed to make the project even more successful. Another possible improvement for future iterations could include changing the font size to be smaller or using understandable keywords for the graphs that have text categories that are too long. This would make the graph look a bit less cluttered/confusing because the user would not have to use the tooltips immediately in order to see the rest of the word that is not readily visible.

* What, if any, parts did you enjoy about this project?

We really enjoyed all the parts of this project. From data exploration, data preprocessing and feature engineering specifically to getting to know and understand new and interesting data points about complaints against the NYPD. Additionally, we also enjoyed creating visuals, building the dashboard and watching the data story come together.

* What were the most challenging aspects of this project?

The most challenging aspect of this project is deciding on the direction we want to go with the data story of the dashboard. There are many different interesting factors in civilian complaints against the NYPD, so it was tough to choose which story we wanted to specifically tell.

* What would you recommend changing about this project for future classes?

No change in this project is needed for future classes. This project came full circle with all the elements that we were taught in class, and it was very exciting to put all the things we learned together in this project.