INFO 5100 Project 1
Pujaa Rajan and Tyler Alicea

We retrieved our data from public sources, including:
- Complaint/crime data: https://catalog.data.gov/dataset/nypd-complaint-data-current-ytd
- Shape for map (Census Tracts 2010 (Clipped to Shoreline)):
  http://www1.nyc.gov/site/planning/data-maps/open-data/districts-download-metadata.page

The complaint data provided us with all of the reported crimes documented by the New York City Police Department in 2016. We modified NYPD_Complaint.json to only focus on murders, which we did by editing a CSV file in excel to remove extraneous columns and using an online tool to convert that to a JSON file. To make our file more compatible with the map.json, we extracted the crime data from the complaint file to create crimes.json.

Our variables include murder report date, longitude and latitude (used for mapping the incidents to our map of New York City), the borough, and time of day. All of this information came from the initial complaint JSON file provided on data.gov. The initial dataset had a handful of columns/variables related to internal police use that we found unuseful for our purposes, which is why we removed them from the data file. We then integrated the latitude and longitude information with the map, plotting each incident with a single dot.

In looking at the first two visualizations, we used linear scales to create bar graphs of murder reports in each borough and then by time of day. We wrote functions involving for each loops and conditional statements to parse the data, and then input that data into our linear scale. We also created a scatterplot to show the when and where most crimes were occurring by plotting time of day against day of the year. An interesting, but understanding, finding is most crimes occurred early in the morning or late at night. This is probably because it's dark outside. There aren't many murders executed at noon.

On the map, we used the a SHP file converted to a GEOJSON and then a TOPOJSON file to draw the paths representing New York City. Each of the subareas in each borough represent the census tracts as they existed in 2010, which are used in conducting the census every decade. We appended circles onto the map, with each representing a reported murder from the crimes.json file that we created from the other data source.

For the colors, we used the schemeCategory10 from the Scale Chromatic library using an ordinal scale, giving each borough of New York City a different color. Unfortunately, the color palette offerings were somewhat limited and too bright for what we were looking for. To resolve this, we overlaid a semi-transparent black rectangle over the map to darken the colors. Then we pulled the hex codes from a screenshot of the map to apply them to each of the boroughs in the other visualizations.

The visualization shows us a variety of information about murders in New York City by borough. The visualizations begin by showing how the 300 reported murders were spread out across the boroughs. Then by showing different characteristics of those murders, such as whether they occurred inside or outdoors, as well as when they occurred and at what time of day. Finally, by looking at the map, you can see exactly where each of the murders were reported, allowing for a user to see clusters and possible location patterns. Personally, we found the concentrated clusters of murders to be interesting, and if we had more time, we would want to compare our data set with additional information such as socioeconomic data, which could be correlated with the clusters.