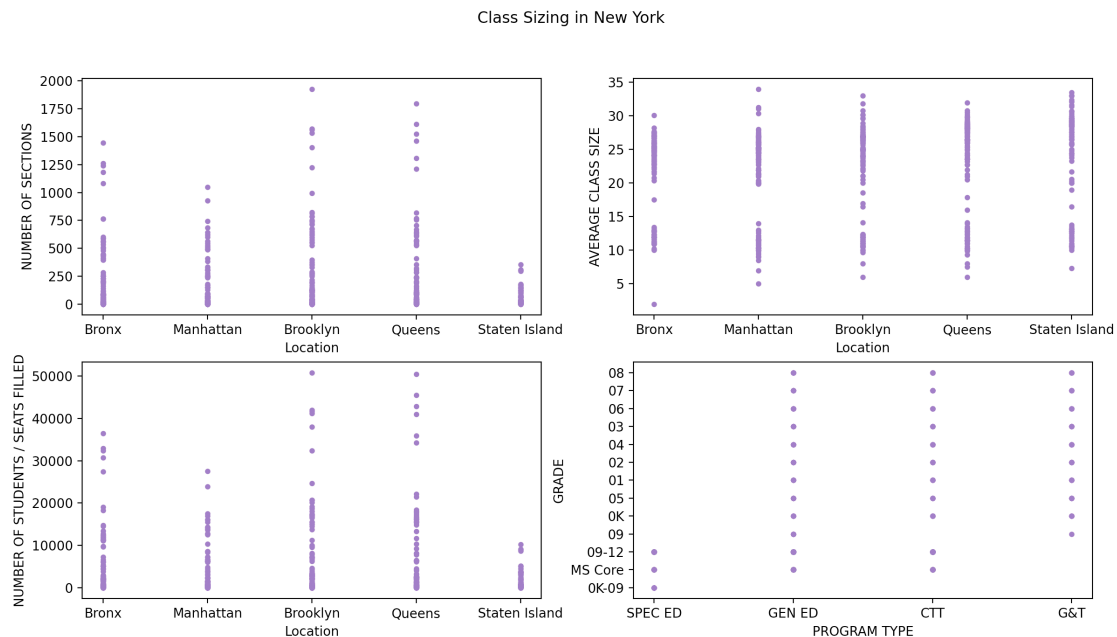


I chose to analyze the “2010-2011 Class Size – Borough Summary” dataset from Data.gov, for class sizing in major locations in New York.

<https://catalog.data.gov/dataset/2010-2011-class-size-borough-summary>



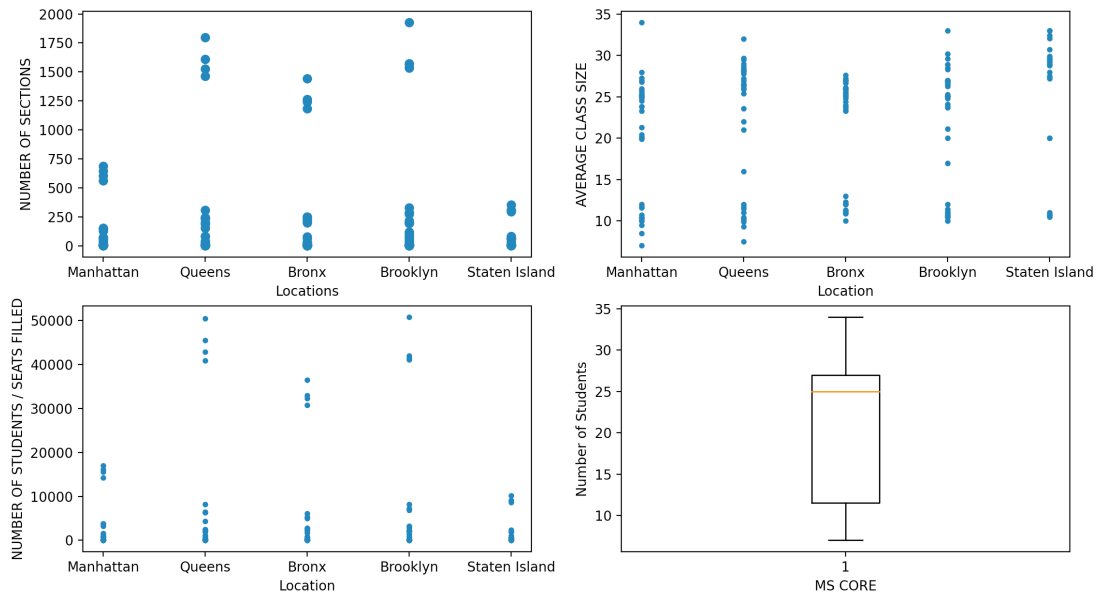
The first question that came to mind for me was initially how many class sections were available in the various locations reported on in New York. In the top left plot, I came to the conclusion that there are a lot more class sections offered in the Brooklyn, and Queens locations relative to other areas.

Next was the average class size per specific location as well. In the top right plot, I noticed that the average class size seemed to be similar throughout locations, Staten Island slightly had a heavier average class size than the rest.

Next is the bottom left plot, in which the number of students/seats filled was identified, going more in depth on class turnout per location. Brooklyn and Queens led this area again.

Lastly is the bottom right plot, in which the specific types of programs were identified for each grade. I concluded that the program type is offered in most grades, other than 9-12 and MS CORE classes showing differences.

MS CORE Analysis

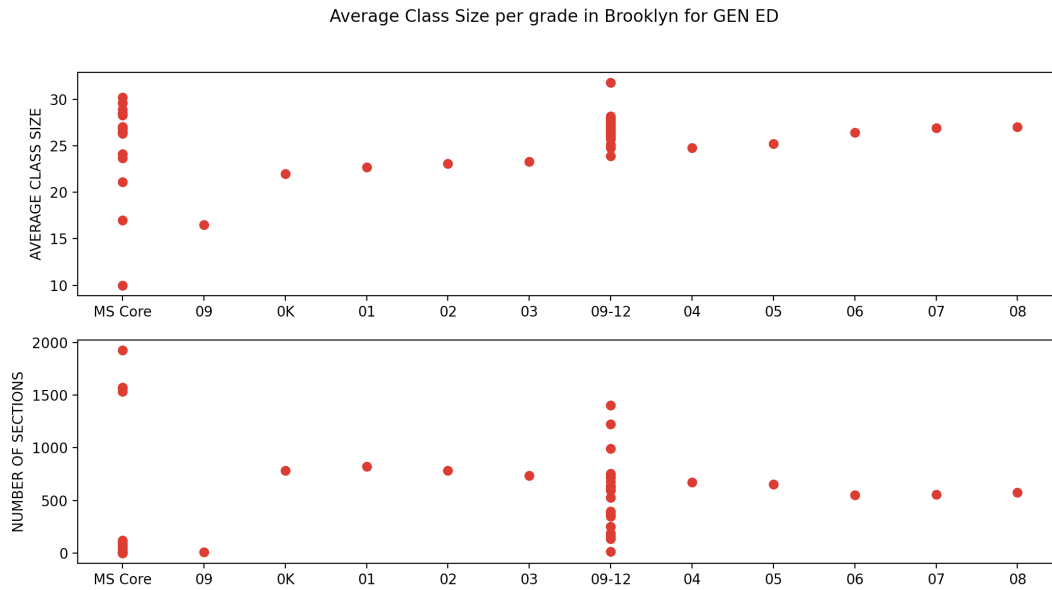


For the next four figures, I was thinking about a specific grade, MS CORE. In the top left plot, I concluded that the number of sections for MS CORE classes was mostly shown in Queens, Bronx, and Brooklyn.

In the top right plot, I found out the average class size per location, once again focused on the MS CORE grade. I noticed that they all had similar averages other than Queens and Brooklyn again.

In the bottom left plot, the number of filled seats was identified for the MS CORE grade, once again with Queens and Brooklyn leading.

Lastly in the bottom right plot, I used a boxplot to look at the average number of students in the MS CORE specific grade. I noticed the mean number of students was around 25, and the max and min students in these classes were about 35 and 10.



Lastly, I dug a little deeper and did analysis on the Average Class Size and Number of Sections specific to the location Brooklyn, in which the program type was GEN ED.

For GEN ED classes in Brooklyn specifically, I noticed the average class size per grade was usually between 15 and 30.

Additionally, the number of sections available for students in Brooklyn in GEN ED showed that grades 9-12 and MS Core classes offered the most sections. I assume this is because students in higher grades have more options in what they are taking relative to elementary classes and middle school students.

Listed below is my code for the project.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import style

#Figure 1- Graphs 1-4
classes = pd.read_csv("2010-2011_Class_Size_-_Borough_Summary.csv")
headers = classes.columns

myFigure, myAxes = plt.subplots(2,2, squeeze=False)
classes.sort_values(by=["AVERAGE CLASS SIZE"], inplace = True, ascending=
True)

myAxes[0,0].scatter(x="BOROUGH", y="NUMBER OF SECTIONS", data=classes,
color='tab:purple', s = 10)
myAxes[0,0].tick_params(labelbottom=True)
myAxes[0,0].set_ylabel("NUMBER OF SECTIONS")
myAxes[0,0].set_xlabel("Location")

myAxes[0,1].scatter(x="BOROUGH", y="AVERAGE CLASS SIZE", data=classes,
color='tab:purple', s = 10)
myAxes[0,1].tick_params(labelbottom=True)
myAxes[0,1].set_ylabel("AVERAGE CLASS SIZE")
myAxes[0,1].set_xlabel("Location")

myAxes[1,0].scatter(x="BOROUGH", y="NUMBER OF STUDENTS / SEATS FILLED",
data=classes, color='tab:purple', s = 10)
myAxes[1,0].tick_params(labelbottom=True)
myAxes[1,0].set_ylabel("NUMBER OF STUDENTS / SEATS FILLED")
myAxes[1,0].set_xlabel("Location")

myAxes[1,1].scatter(classes["PROGRAM TYPE"], classes["GRADE"],
color='tab:purple', marker = "o", s=10)
# myAxes[1,1].tick_params(rotation=90)
myAxes[1,1].set_ylabel("GRADE")
myAxes[1,1].set_xlabel("PROGRAM TYPE")

plt.suptitle("Class Sizing in New York")
plt.show()

# MS CORE FOCUS ONLY- Figure 2- Graphs 5-8
MSCORE = classes[classes['GRADE']=="MS Core"]
print(MSCORE.head())

myFigure, myAxes = plt.subplots(2,2, squeeze=False)
classes.sort_values(by=["AVERAGE CLASS SIZE"], inplace = True, ascending=
True)
```

```

myAxes[0,0].scatter(x="BOROUGH", y= "NUMBER OF SECTIONS", data=MSCORE,
color='tab:blue')
myAxes[0,0].tick_params(labelbottom=True)
myAxes[0,0].set_ylabel("NUMBER OF SECTIONS")
myAxes[0,0].set_xlabel("Locations")

myAxes[0,1].scatter(x="BOROUGH", y="AVERAGE CLASS SIZE", data=MSCORE,
color='tab:blue', s = 10)
myAxes[0,1].tick_params(labelbottom=True)
myAxes[0,1].set_ylabel("AVERAGE CLASS SIZE")
myAxes[0,1].set_xlabel("Location")

myAxes[1,0].scatter(x="BOROUGH", y="NUMBER OF STUDENTS / SEATS FILLED",
data=MSCORE, color='tab:blue', s = 10)
myAxes[1,0].tick_params(labelbottom=True)
myAxes[1,0].set_ylabel("NUMBER OF STUDENTS / SEATS FILLED")
myAxes[1,0].set_xlabel("Location")

myAxes[1,1].boxplot(MSCORE["AVERAGE CLASS SIZE"])
myAxes[1,1].tick_params(labelbottom=True)
myAxes[1,1].set_xlabel("MS CORE")
myAxes[1,1].set_ylabel("Number of Students")

plt.suptitle("MS CORE Analysis")
plt.show()

#Brooklyn GEN ED CLASSES - Figure 3 - Graphs 9-10

Brooklyn = classes[classes['BOROUGH']=="Brooklyn"]
print(Brooklyn.head())

newData = Brooklyn[Brooklyn['PROGRAM TYPE']=="GEN ED"]
print(newData.head())

myFigure, myAxes = plt.subplots(2,1, squeeze=False)
classes.sort_values(by=["AVERAGE CLASS SIZE"], inplace = True, ascending=
True)

myAxes[0,0].scatter("GRADE", "AVERAGE CLASS SIZE", data= newData,
color='tab:red',marker = "o")
myAxes[0,0].set_ylabel("AVERAGE CLASS SIZE")

myAxes[1,0].scatter("GRADE", "NUMBER OF SECTIONS", data= newData,
color='tab:red',marker = "o")
myAxes[1,0].set_ylabel("NUMBER OF SECTIONS")

plt.suptitle("Average Class Size per grade in Brooklyn for GEN ED")
plt.show()

```