

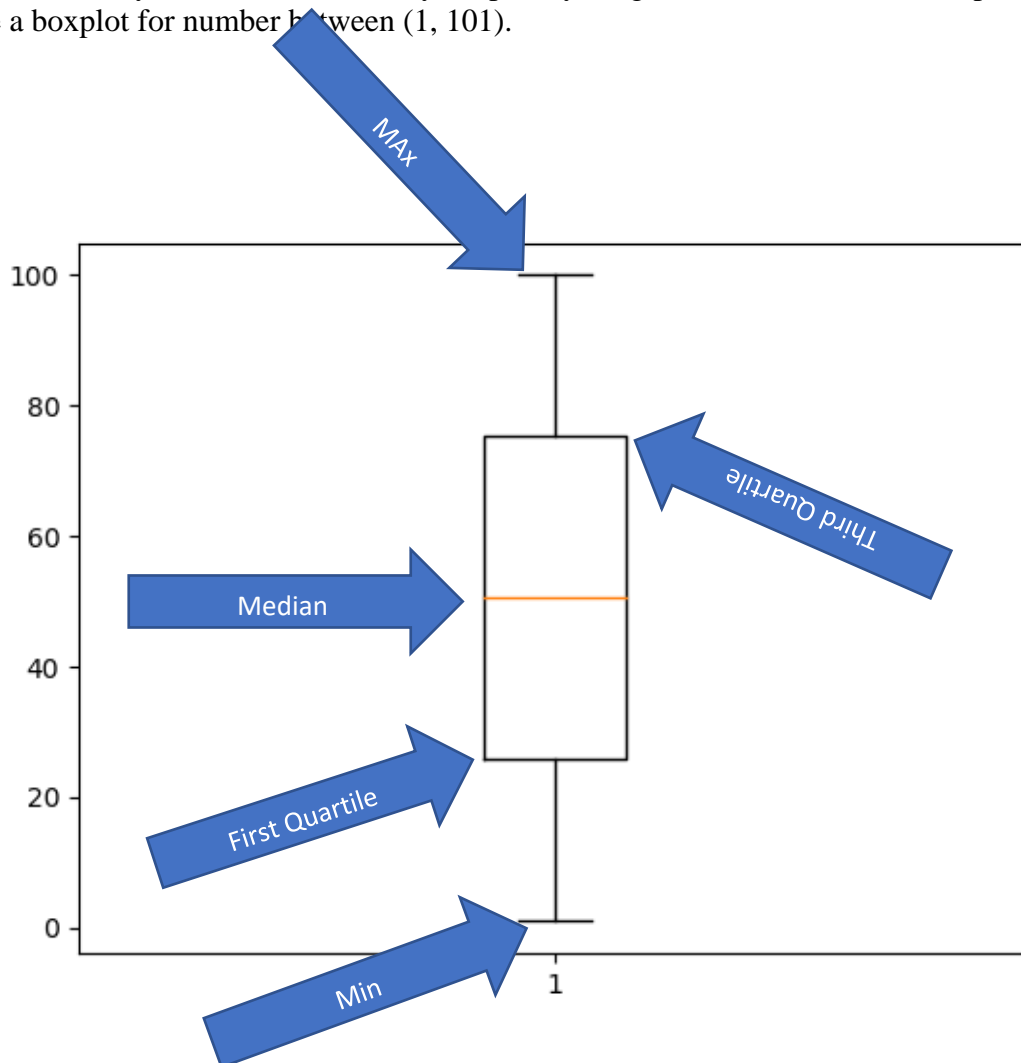
Part 1

1. Create an array with 100 elements from 1 to 100 in order: Create a box plot to visualization your data.

The short accurate description of the most any set of data. In this Question. The Orange color in pictures blow show the median.

The purpose of this question to use box plot. I used python and I used different library in python such as numpy and matplotlib. Numpy according to defection on numpy web site “adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.” And matplotlib lib is “matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy.”

I created an array of 100 element, `array = np.array(range(1, 101))` and I used `boxplot` function to create a boxplot for number between (1, 101).

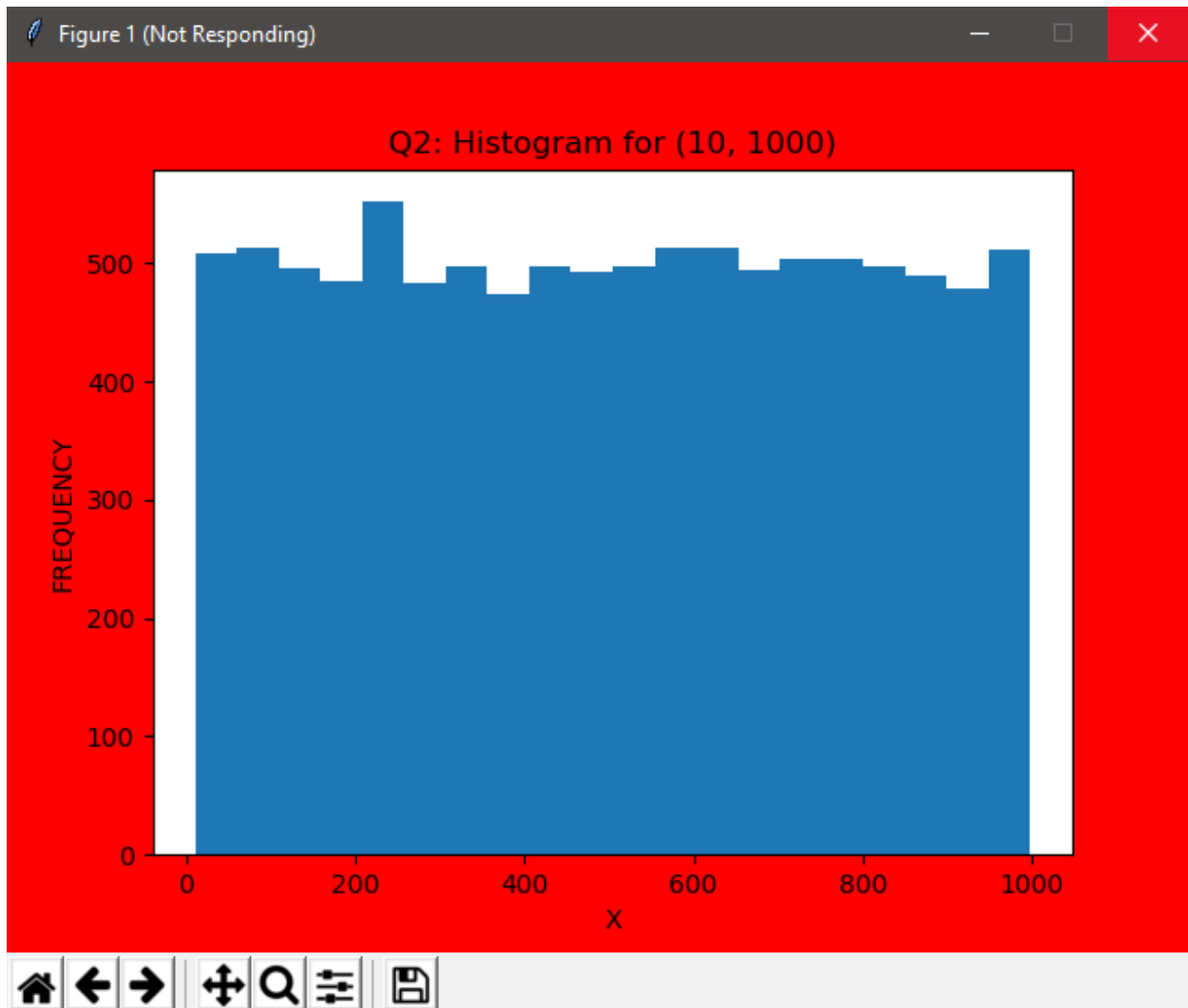


2. Create an array with 10,000 random numbers. Create a histogram of the data using 20 bins.

The purpose of a histogram is to graphically summarize the distribution of a univariate data set.

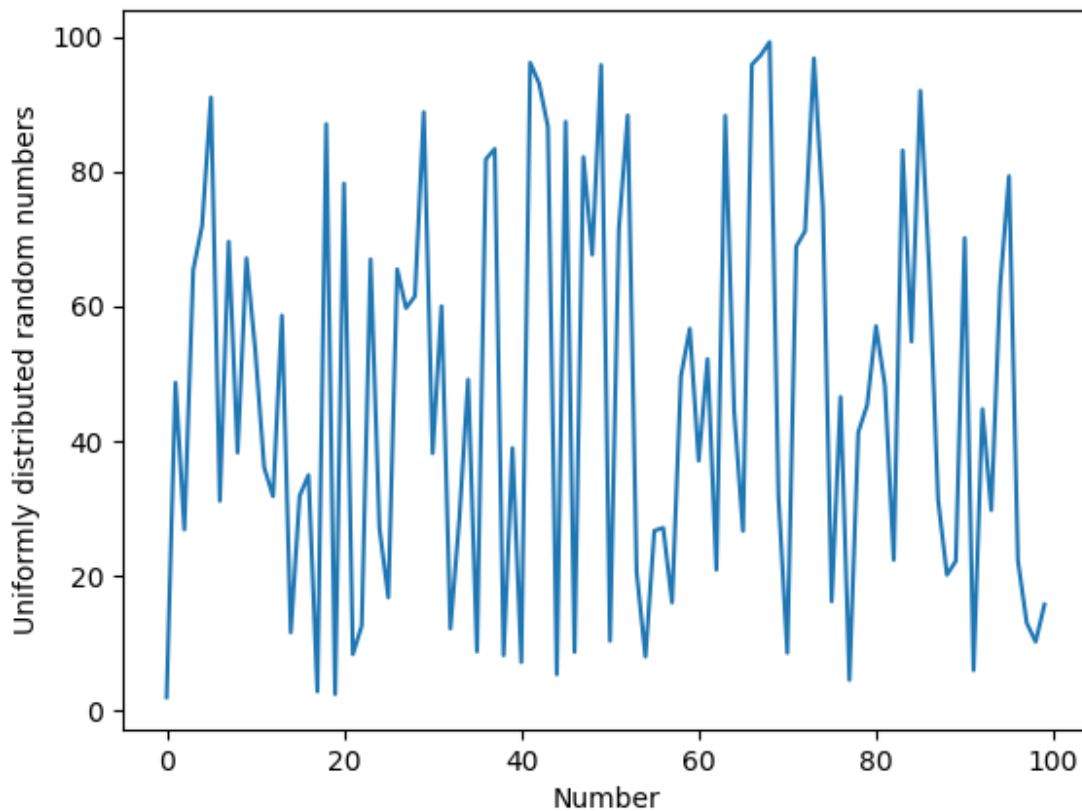
In general histogram shows three things:

- 1- Center
- 2- Spread
- 3- Skewness of the data
- 4- presence of outliers
- 5- presence of multiple mods in the data

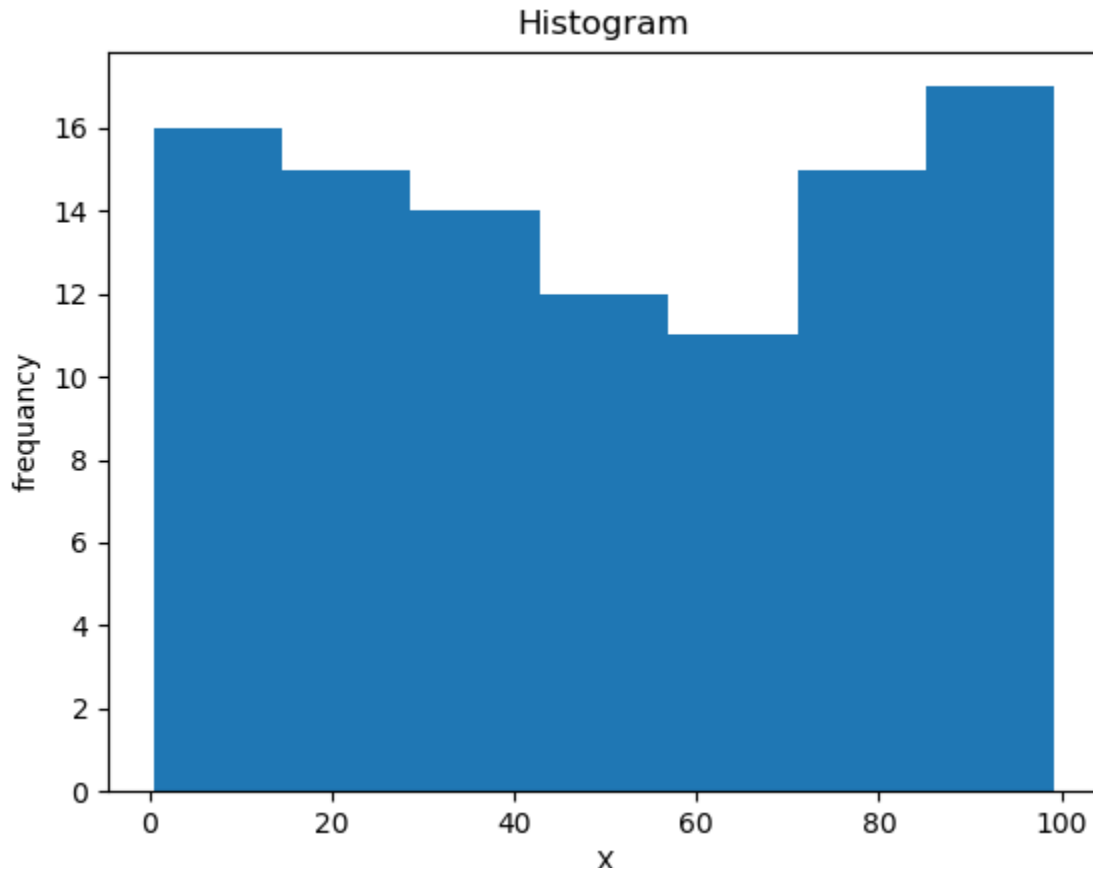


3. Write a program to generate 100 random number uniformly distributed between 1 and 100. Write the numbers out to a binary file and use a line graph to draw the 100 Numbers

When we generate random number uniformly distributed means interval have equal probability of being selected or happening. I used randi to generate random number. In my graph ith Number are in x axis and Uniformly distributed random number in y axis. For example when number is 19 we generate close to 90 uniformly distributed random numbers.



4. Write a program to read the binary file back, divide the range between 1 and 100 into 7 intervals, and calculate the frequency for each interval: display a histogram of your result.



Part 2

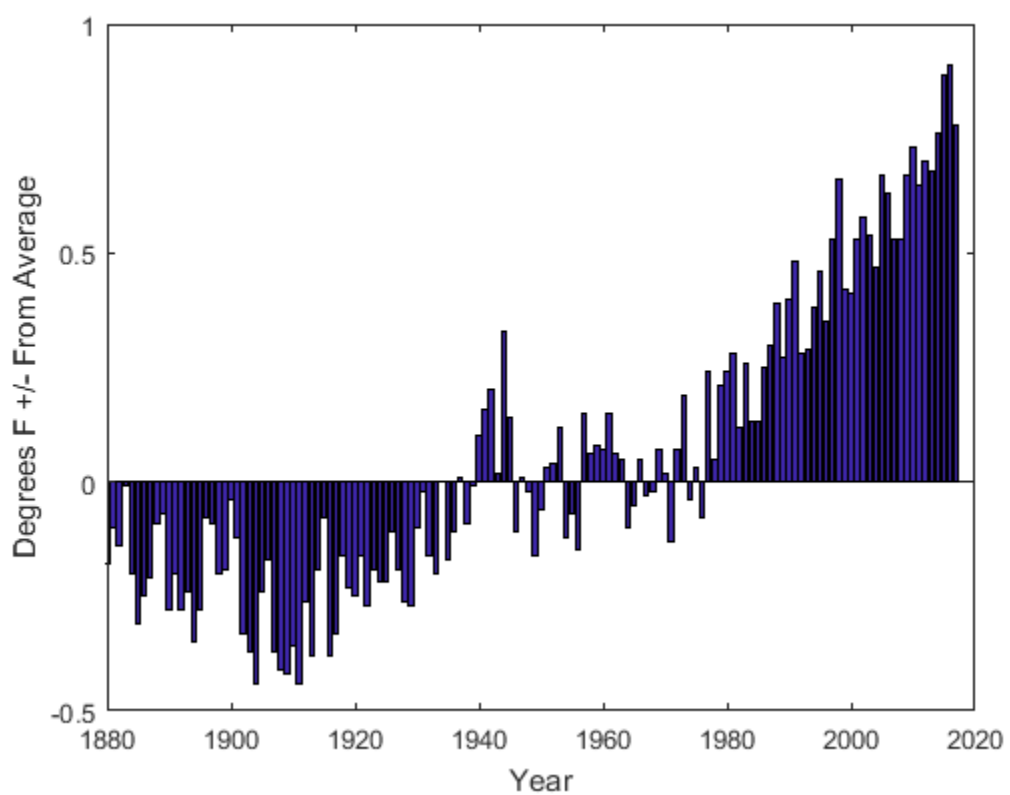
1. Download the NOAA Land Ocean Temperature Anomalies Data Set:

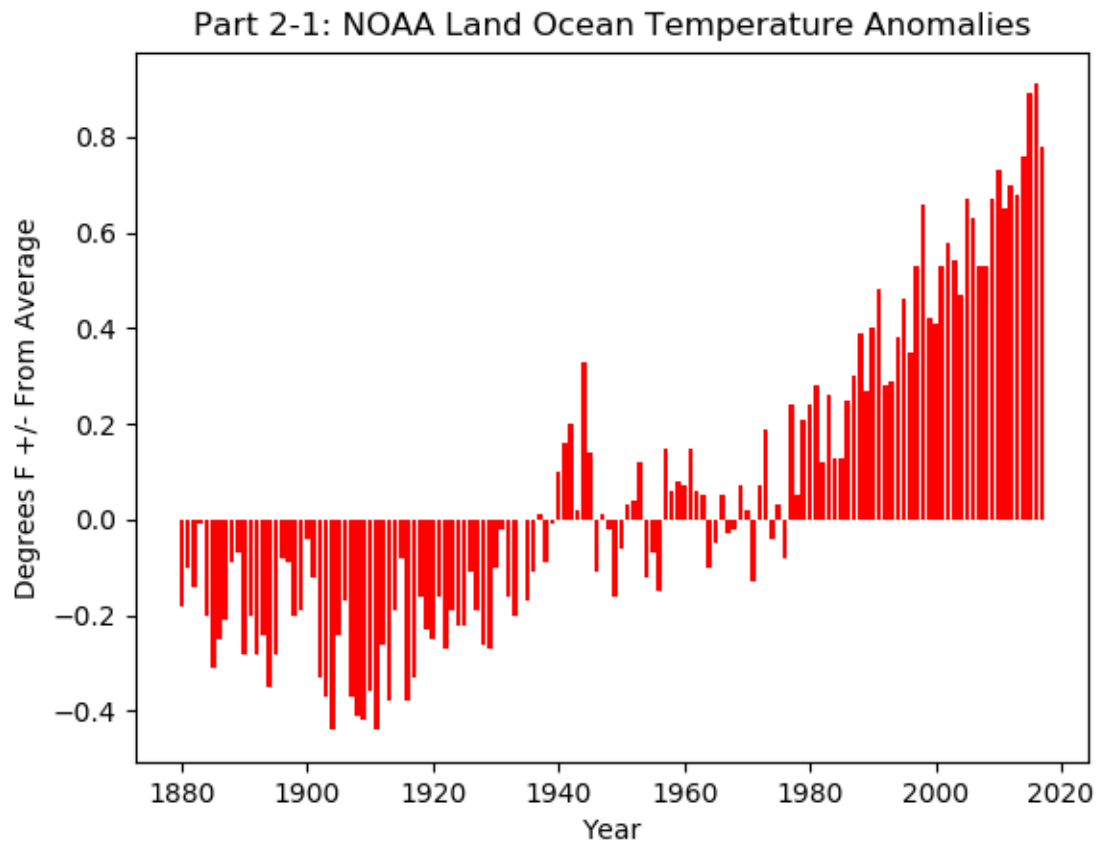
https://www.ncdc.noaa.gov/cag/time-eries/global/globe/land_ocean/1/6/1880-2017.csv.

Create a bar chart of the data. Include a label called "Year" along the x-axis and a label called Degrees F +/- From Average along the y-axis. Describe trends in the data.

I finished most of the question in part 2 with matlab to and some of them with both include this part.

The purpose of using the bar chart I think to display and compare the number. In this question I create bar chart to show the anomalies of temperature. It can be seen that up to around 1940, the temperature was below average, then it stayed around average with some fluctuations, but around 1980 the temperature started to be consistently above average and is continuing to increase.





2. Download the member of Congress by Age data set:

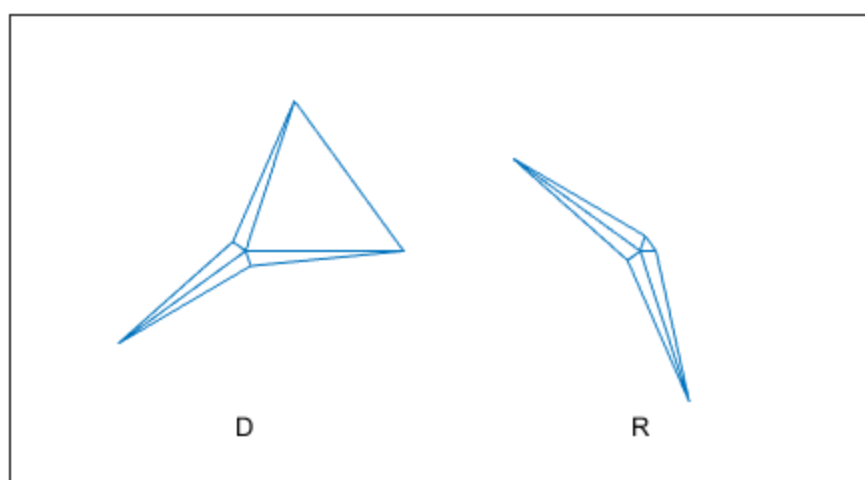
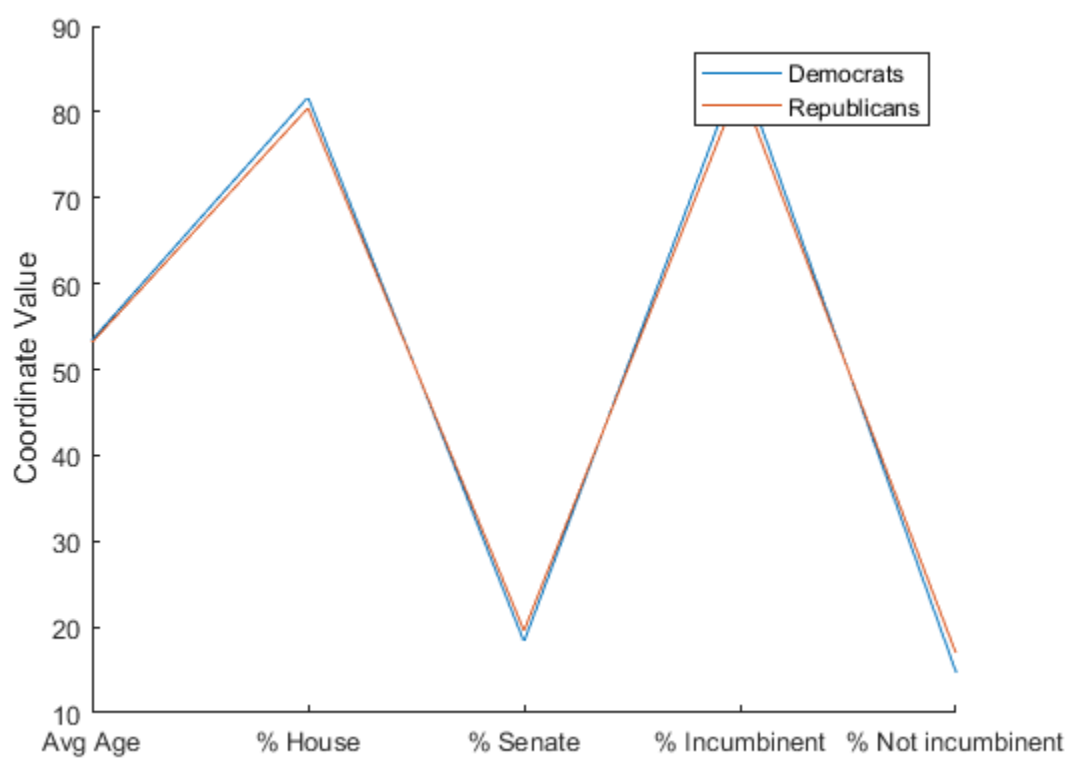
<https://raw.githubusercontent.com/fivethirtyeight/data/master/congress-age/congressterms>. csv.

Create a Star Plot of the data and create a Parallel Coordinates Plot of the data. Describe the trends in the data.

I had a hard time to understand star plot.

The blue line in Coordinate plot shows the Democrats and red color shows the republican. In X axis I created 5 different options such as avg age, percentage of house, percentage of senate and percentage of incumbent and percentage of non incumbent. Also in y axis it just a coordinate value.

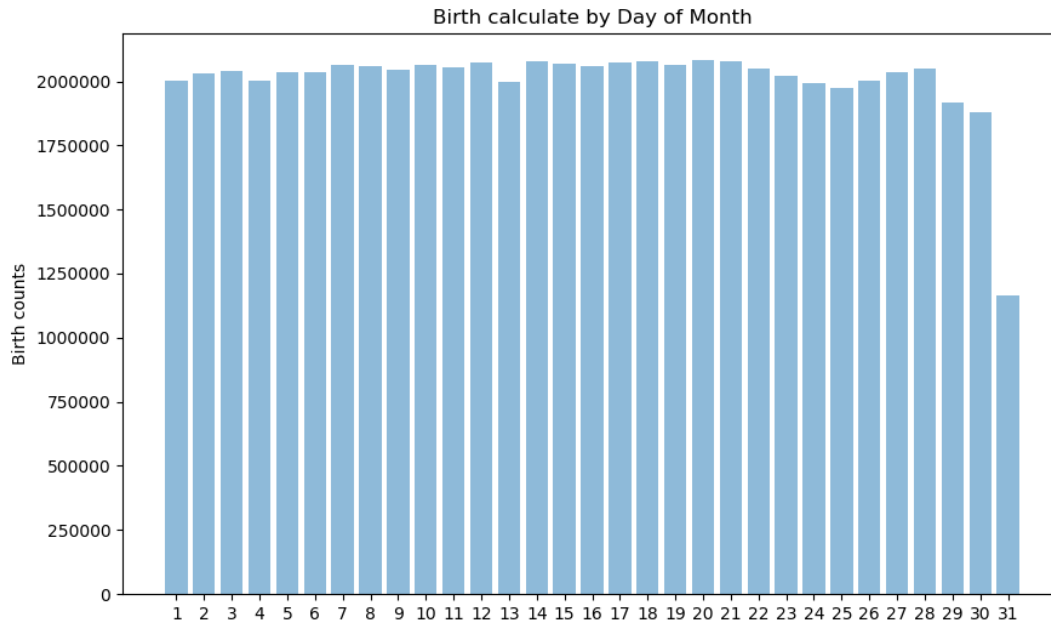
The interesting trend in this data is the democrat and reapplication are really close and similar. So, the blue color and red color on the graph that really close to each other.



3. Download the U.S. Birth data set:

https://raw.githubusercontent.com/fivethirtyeight/data/master/births/US_births_2000-2014_SSA.csv. What day of the month had the highest number of births? What

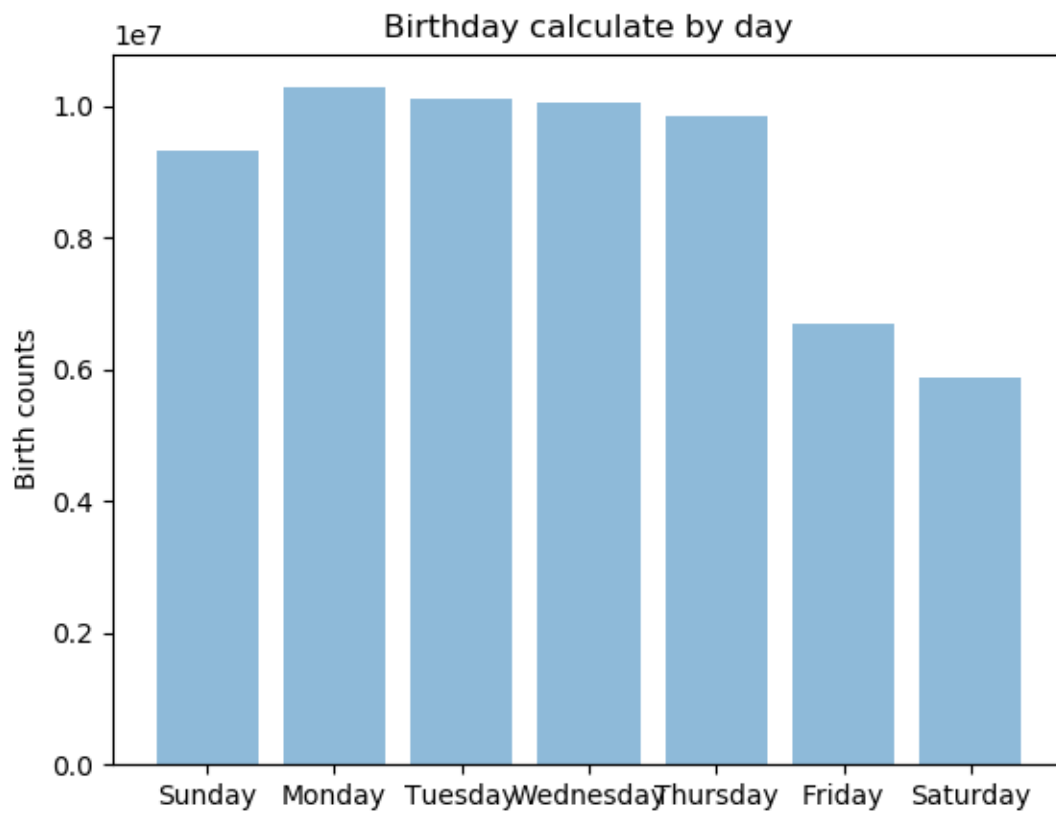
day of the month had the lowest number of births? Are there any interesting trends in the data, i.e. more births in Summer or Winter? What about births on Friday the 13th?

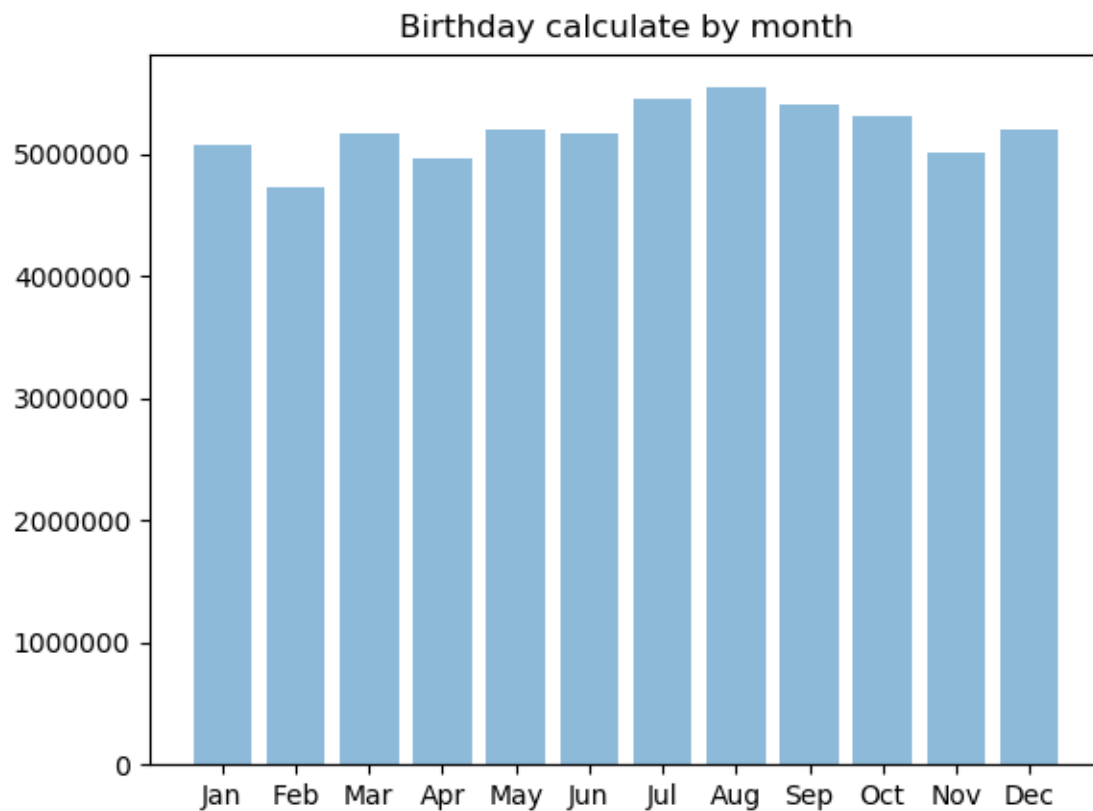


I create 3 graph for Birth calcite by day of month, birthday calculate by the date and birthday, and calculate by the month.

The chance to born in 31 in month is really low it is make sense because we don't have 31 days in each month.

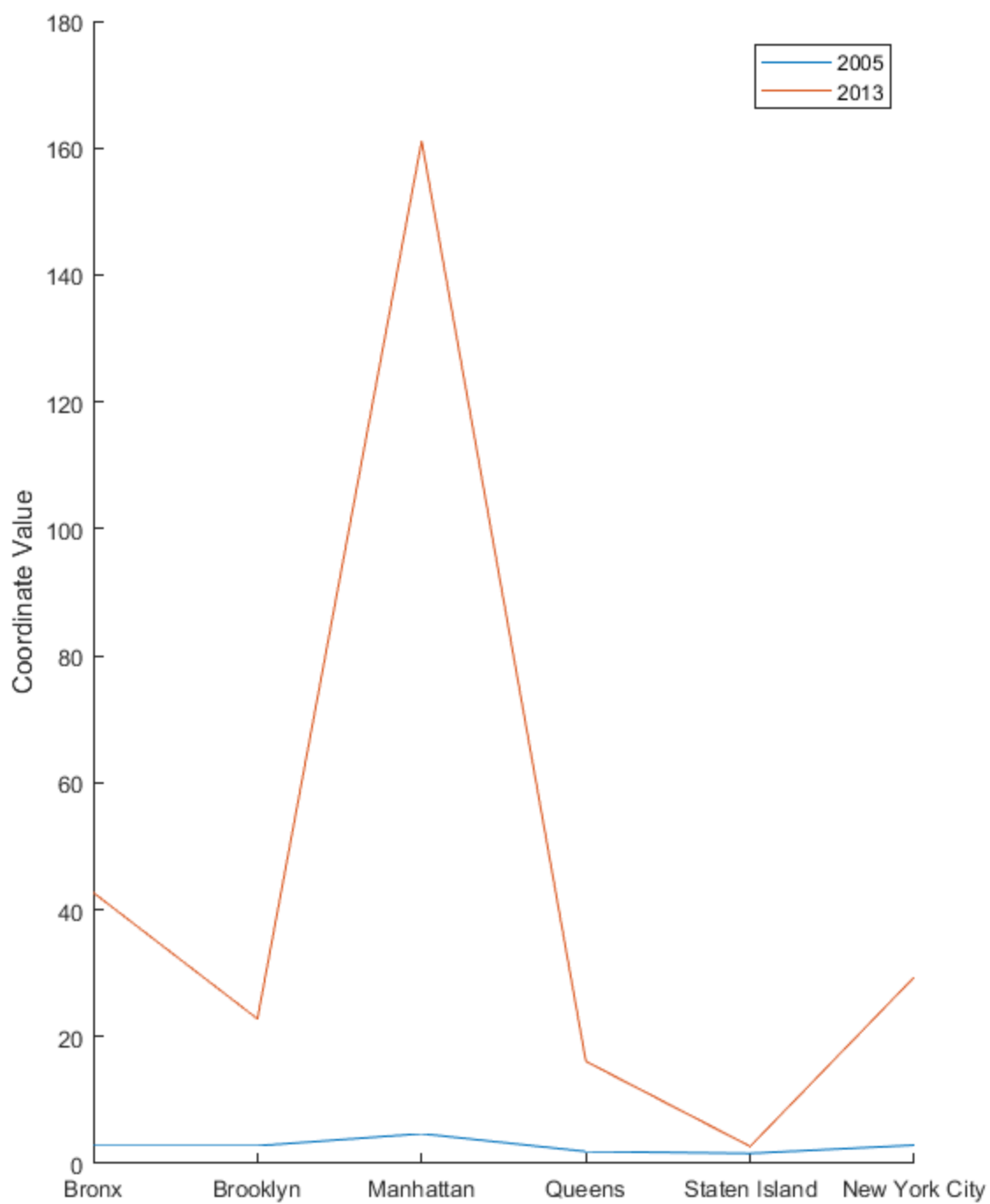
The day of the month with the highest number of birth was on 20th with 2083247 births.



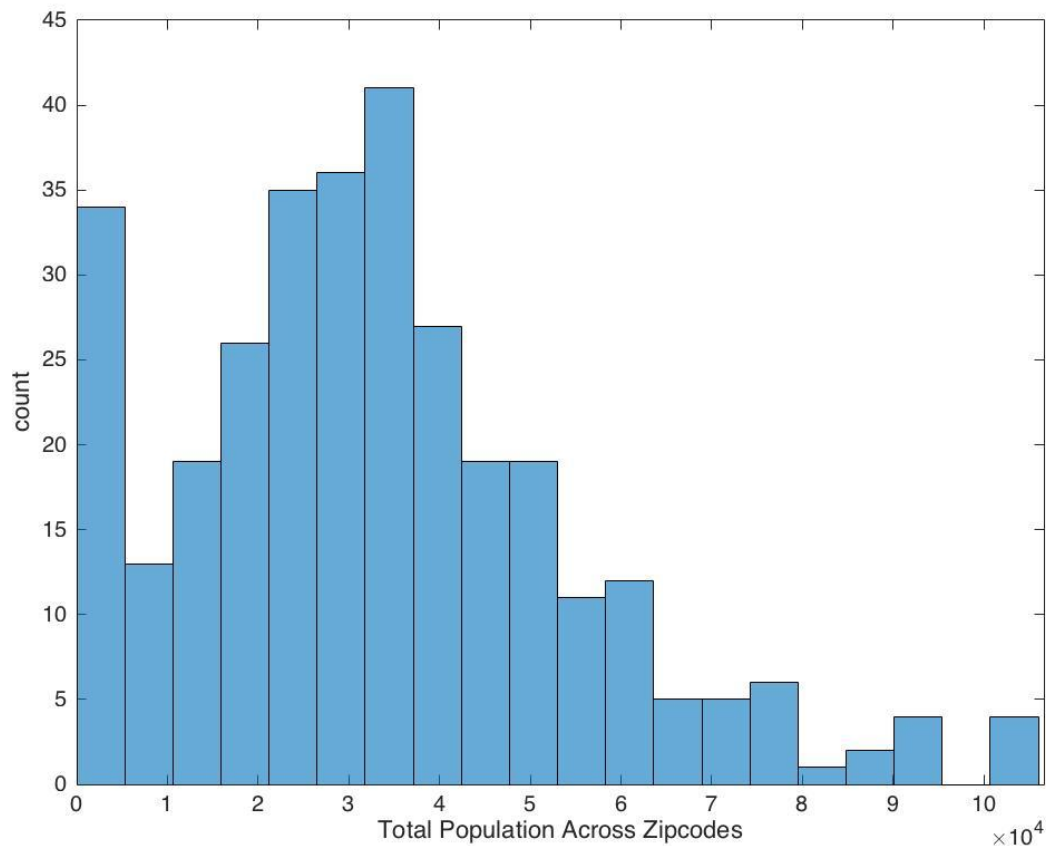


4. The U.S. Government maintains a server with many interesting datasets called Data.Gov: <https://www.data.gov/>. Choose 3 different data sets to visualize. Visualize the data sets in at least 3 different ways. Describe the trends in the data.

I graphed the parallel coordinates plot of the air quality in 6 regions for two different years (2005 and 2013). (Note that the measurement of air quality was different in the 2 years: in 2005, it was measured by the Average Formaldehyde Concentration, and in 2013 it was measured by the Boiler Emission; this is a limitation I had, and it might be better to normalize the y-axis next time). It can be seen that the toxins in Manhattan were much higher than other locations in both years, while the relative air quality between cities was the same in both years (i.e. similar trend).



I plotted a histogram showing how the population is distributed across the zip codes given in the dataset. It can be seen that quite a lot of areas/zip codes have a very low population (less than 5,300 people), while most zip codes have a population of approximately 15,000 to 42,000 (i.e. this is the most common population range in each zip code). On the other hand, there are much fewer zip codes that have a population of more than 53,000, but the population is skewed so that there are as many as 106,000 people in a few of the zip codes.



I plotted a scatterplot showing how the median age of the people in an area (i.e. a zip code) is related to that area's population. The result is interesting: areas with very low population have either a very low median age or a very high median age. This would mean that these areas are either just starting to increase in population (and only have a few young families, for instance) or they are starting to decrease and eventually die out until replaced by new people. On the other hand, as the population increases, the median age of the residents becomes closer and closer to around 30. This likely would mean that these populated areas have a good balance of older and younger people (with the older ones probably those who remained in the area for most of their lives, and the young ones being the children of those who are starting a family; both these groups lead to an increase in the population).

