Exploratory Data Analysis Assignment

I choose data from NYC open data resources. This website includes a brunch of data of new york city. The changes in population from different places can reveal many things, like the economy, policy, environment, etc.

This data resource provided unadjusted decennial census data from 1950-2000 and projected figures from 2010-2040: summary table of New York City population numbers and percentage share by Borough, including school-age (5 to 17), 65 and Over, and total population. The data was collected from Census Bureaus' Decennial data dissemination (SF1) for the years 1950, 1960, 1970, 1980, 1990, 2000 and 2010.
Projected estimates for 2020-2040 were produced based on the methodology developed by the Population Division of NYC Department of City Planning. Additional data sets were used such as natural increase, migration, housing stock.

Also, the data dictionary is pretty clear, like Name New York City Population by Borough, 1950 - 2040; Agency Name NYC Department of City Planning
Dataset Keywords population, projected population, 1950,1960,1970,1980,1990, 2000, 2010, 2020, 2030, 2040 growth, dcp, 1950-2040, age; borough, boro, NYC, percent growth; Dataset Category City Government

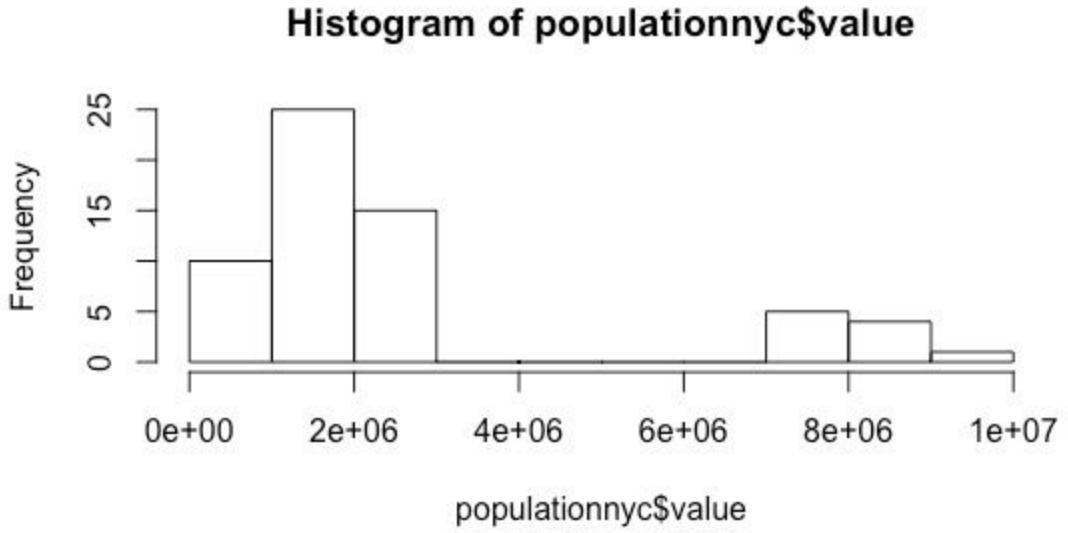
The limitation of the dataset is the numbers are too vague. If the data can provide some details of people to moving place, the visualization can show more meaningful information than data itself. For now, the data can only show the trend of population changes.

By viewing this data, I can obviously find New York City is such an attractive city that its population has an increasing trend in the past few decades and it will continually grow in the future. Every year, the increased rate of population in total is 100%. Those the charts show the population of each location, Brooklyn; Manhsttsn; Queens; Bronx; Staten Island. Brooklyn is the largest group in new york from 1950-2010 and that advantage won't be changed in the future 2040. However, the population trend of Brooklyn had a decreasing trend from 1950-2000 which means people moved out. From 2000-2040, the population starts to increase again. There must be some reason that influenced it. The most reasonable reason is the safety of Brooklyn has increased in 2010. People get more security than before.

People in Manhattan was keeping moving out since 1950. In 1960, Queens became the second large group in new york city. That trend didn't change since 1960. The projected population in 2040 still maintains that trend that more people moving out from manhattan and more people living in Queens.

People living in the Bronx was a kind of stable. It has some waves around 1980, but most of the time, the Bronx's population shows a slowly increasing trend. The same as State Island.

When doing a comparison of each year, I can't show all the data at the same time. I can only view them year by year to see the different place holds a different population. The data is missing in a year row when I want to see the change of specific years. Also, it's interesting to observe the population changes and guess the stories behind those data.



Code

```
#installed.packages("lubridate")

#installed.packages("dplyr")


library(lubridate)

library(dplyr)

library(tidyverse)


population <- read.csv("Desktop/cleannyc.csv", header = TRUE)


class(population)

dim(population)

names(population)

str(population)

glimpse(population)

summary(population)


head(population)

tail(population)


populationnyc <- gather(population, years, value, -Borough)

head(populationnyc)

tail(populationnyc)

glimpse(populationnyc)


populationnyc$years <- gsub("\\X","", populationnyc$years)
```

```
head(populationnyc)


glimpse(populationnyc)

summary(populationnyc)

str(populationnyc)


populationnyc$value <- gsub("\\,","", populationnyc$value)

glimpse(populationnyc)


hist(populationnyc$value)
```