

## PROBABILITY THEORY AND INTRODUCTORY STATISTICS



ALY6010, FALL 2019

WEEK 1 PROJECT ASSIGNMENT (US POPULATION AND MIGRATION RATES)

SUBMITTED BY: SHIVANI ADSAR, ANUPREETA MISHRANUID: 001399374, 001050752CRN: 71447SUBMITTED TO: DR. DEE CHILUIZA REYESDATE: 09/22/2019

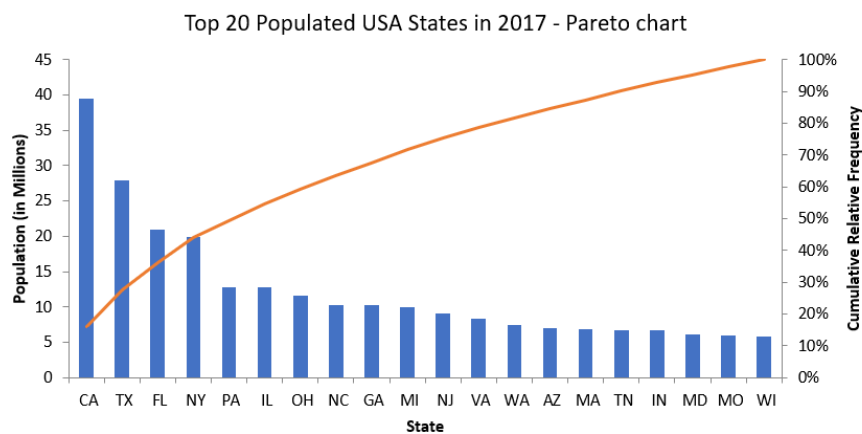
## Introduction

The assignment aims at executing descriptive statistical methods on the data set. The project displays the population and migration rates of United States. Using descriptive statistics, we can represent the population of data using graphical methods. The data set is displayed in the form of Pareto Chart, Histogram, Scatter Plot, Line Chart, Box plot and Pie Chart using numerical and categorical data.

## Analysis

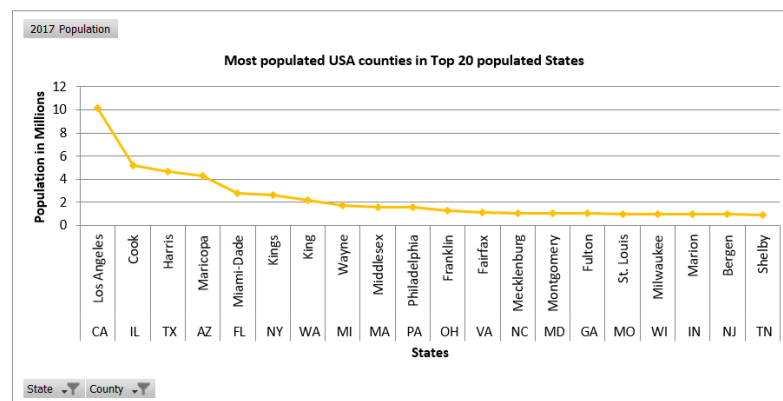
### Q1. Pareto Chart for populated states in United States

Pareto charts displays in order of importance of inputs that matter is simple in a presentable visual format. It is based on the 80-20 principle that states 80% of the effects come from 20% of the causes. According to the graph, it is observed that, California has the highest population and Wisconsin has the minimum population.



Graph 1: Pareto Chart for top 20 populated states in United States, 2017

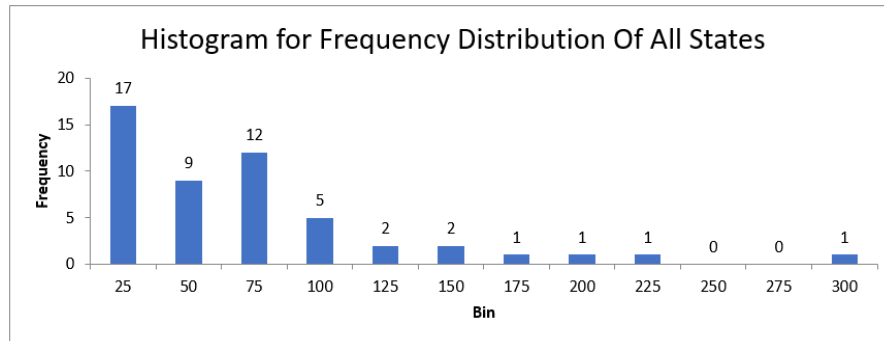
Line graphs are easy to interpret the data easily to find the drastic differences in data. It is observed from the graph, that Los Angeles is the most populated and Shelby is the least populated county. Example of a line graph is ECG for monitoring the heart rate.



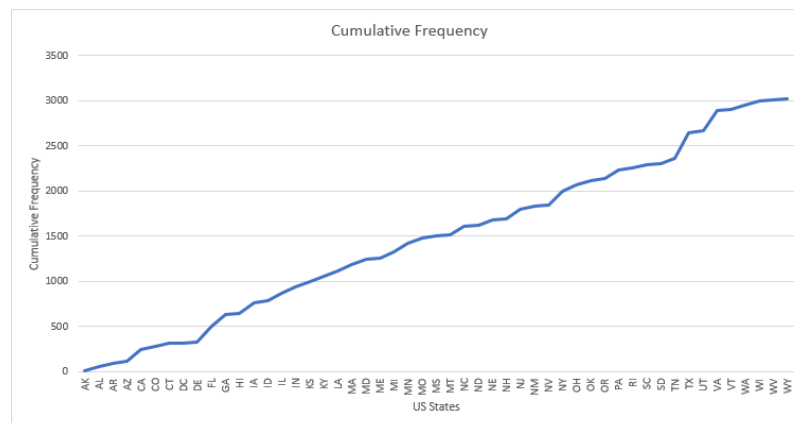
Graph 2: Line Chart for Most populated counties in top 20 populated states

## Q2. Histogram to represent frequency of all states

The frequency and cumulative frequency histogram makes it is easier to identify the data, the frequency of data occurring in the data sets and categorizes the data, making it easy to interpret. The graph indicates, the frequency and cumulative frequency based on the International migration rate of the people in the United States in different States with Texas showing the highest frequency and Washington DC being the lowest.



Graph 3: Histogram to represent frequency of all states



Graph 4: Cumulative frequency for the States in United States

The cumulative frequency is calculated by adding the frequencies for each state.

## Q3. Numerical Descriptive Statistics using Table

The data for East coast and West coast was collected and then segregated. It is observed from the tables that, mean, skewness and median of west coast is more than the mean of east coast. Firstly, we searched the states on the west and east coast, then used the V LOOKUP formula to differentiate between the states and then used the data analysis tools.

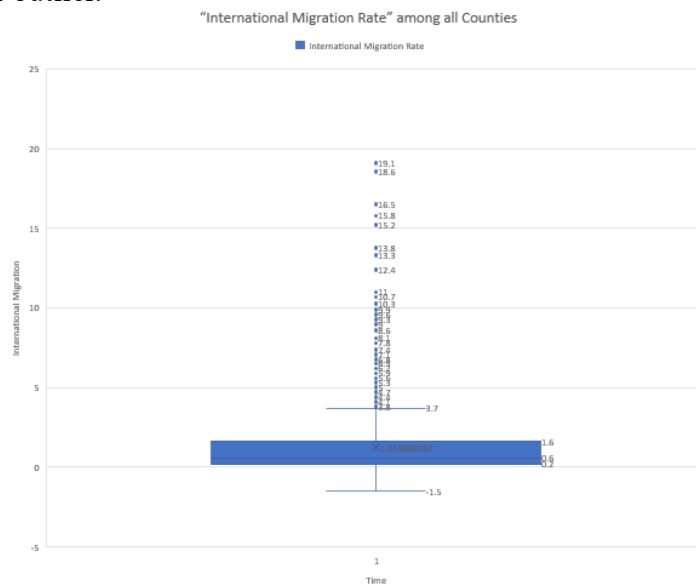
## Probability Theory and Introductory Statistics

East Coast Natural Change		West Coast Natural Change	
Descriptive Statistics	Value	Descriptive Statistics	Value
Mean	371.954466	Mean	2065.91603
Standard Error	55.7266647	Standard Error	514.272334
Median	16	Median	232
Mode	-15	Mode	-44
Standard Deviation	1331.62244	Standard Deviation	5886.1159
Sample Variance	1773218.32	Sample Variance	34646360.4
Kurtosis	19.6147846	Kurtosis	54.0224104
Skewness	3.93017104	Skewness	6.5181293
Range	14261	Range	56000
Minimum	-3822	Minimum	-328
Maximum	10439	Maximum	55672
Sum	212386	Sum	270635
Count	571	Count	131
Largest(1)	10439	Largest(1)	55672
Smallest(1)	-3822	Smallest(1)	-328
Confidence Level(95.0%)	109.454668	Confidence Level(95.0%)	1017.42631

Graph 5: Numerical Descriptive Statistics for Natural Change in East coast and West Coast

### Q5. Box Plot for International Migration Rate

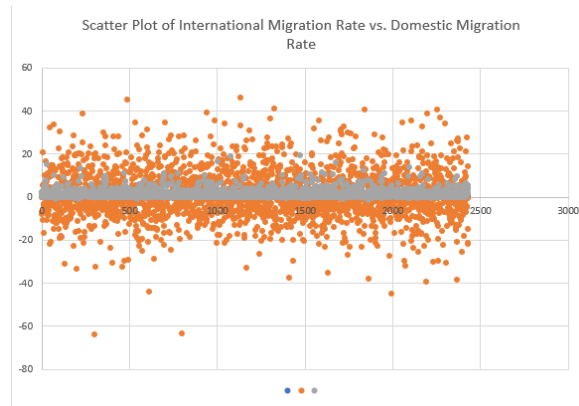
Box plot provides details of data's symmetry, skewness and diversity of data. It is observed from the graph that, 0.2 is the 1<sup>st</sup> quartile, 1.6 is the 3<sup>rd</sup> quartile, 1.4 is the Interquartile range and 3.7 is the upper bound. There are some outliers in the data, 19.1 being the highest outlier.



Graph 6: Box Plot to plot International Migration Rate for all counties

### Q6. Scatter Plot to compare migration rates

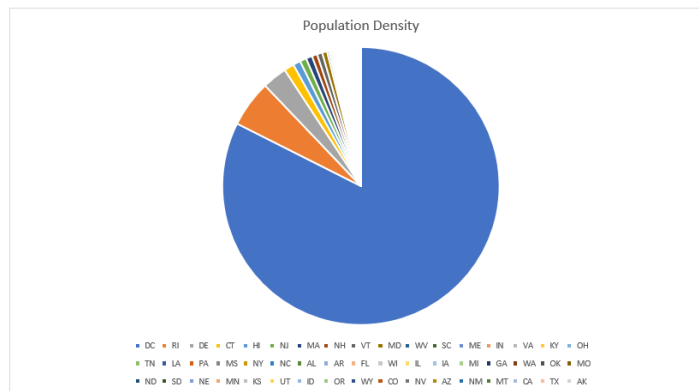
The scatter plot indicates that there exists a correlation between the International Migration Rate and the Domestic Migration Rate. The domestic migration rate is more than the international migration rate.



Graph 7: Scatter plot for International Migration Rate vs. Domestic Migration Rate

### Q7. Pie Chart to display Population Density

Pie chart presents the data in a very understandable format. It displays the regions and shows comparisons at a glance. Hence, pie chart will be an appropriate to display the population density. After observation from the chart, it is seen that Washington DC has the highest population density and Alaska has the least population density.



Graph 8: Pie chart displaying the population density

## Conclusion

It is inferred from the observations of the dataset that using visualization methods like graphs and charts are easy to interpret the data. Also, using statistics methods like mean, median, standard deviation and variance, help to understand the varying patterns and comparisons in the population. So, we can focus our efforts towards the solution according to the problem statement.

## Reference

1. Size of States. (n.d.). Retrieved from <https://statesymbolsusa.org/symbol-official-item/national-us/uncategorized/states-size>
2. Bluman, A.G. (1992). *Elementary Statistics*. Retrieved from <https://archive.org/details/ElementaryStatistics/page/n5>