Rebecca Ly

February 26th, 2016

INFSCI 2725

Assignment 4: R – Titanic Dataset

Whisker Plot

I created a whisker plot which simply depicts the age groups split into female and male groupings of the passengers among the Titanic. According to the whisker plot, Most of the passengers were between the ages of 20-40 years and averaged evenly among both males and females with few males being above the age of 60 onboard. The median age for the females onboard is estimated to be between 25-30 and around 30 for males.

Histogram

For the Histogram, I visualized the fare distribution of the passengers onboard the Titanic. According to the Histogram, the majority of the passengers paid around $20-$30 on average for their tickets and the rest of the passengers paid $50-$100 and then there are some outliers in the data who paid above $100 which may or may not be correct. Something to improve this Histogram is accounting for the Pclass of the passengers.

Facet Grid

I created a facet grid using Histograms to split up the data into six different charts based on the Pclass, age, and sex of the passengers. The three different rows depict the Pclass of the passengers categorized into 1, 2, and 3. Passengers in the 1st class are considered to be upper level, 2nd class are middle-income and 3rd class passengers were lower income. The two columns depict the sex of the passengers split between male and female. And the x-axis depicts the age of the passengers. These charts show that there is a higher number of male passengers between the ages of 20 and 30 in the 3rd class of the Titanic.

Violin Plot

For the Violin Plot, I decided to focus in on the Pclass of the passengers and Violin Plots are really helpful for focusing on one part of the dataset that shows audiences where the data is heavily concentrated. This violin plot shows that the dataset is heavily concentrated in the area where the age group of the passengers is heavily concentrated between the ages of 20 and 30 in the 2nd class.