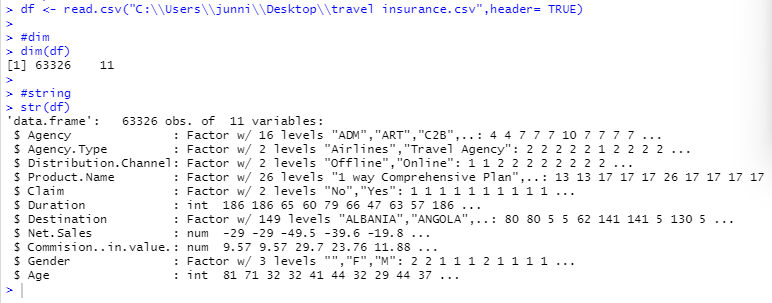
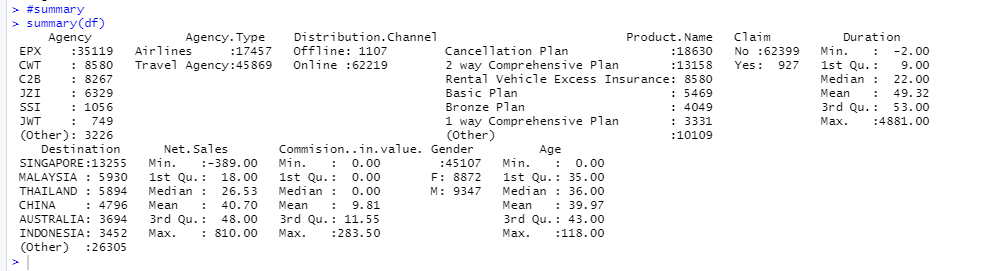
**Yijun Wang**

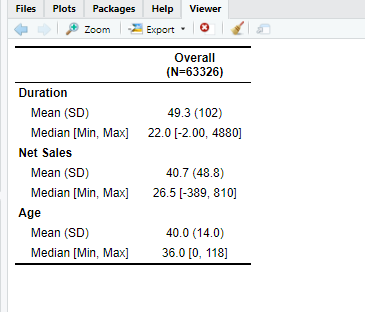
**Aly 6010**



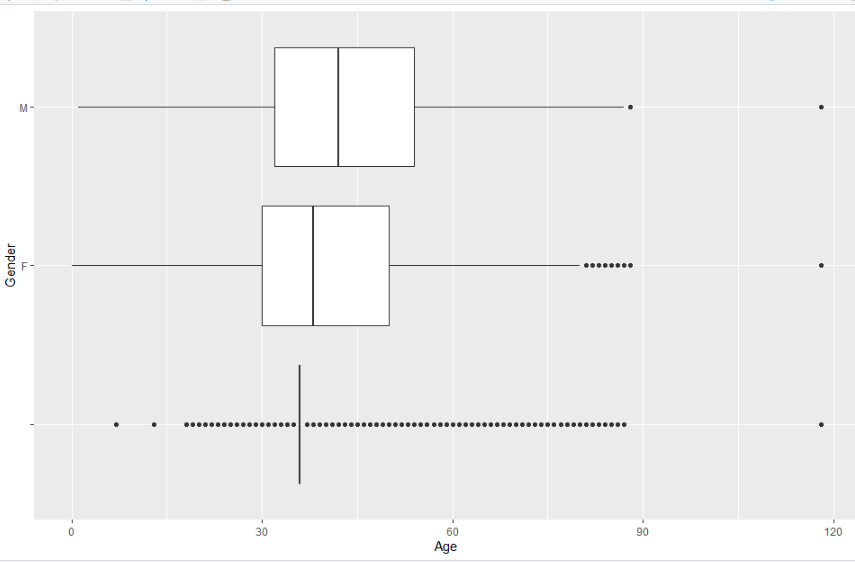
This dataset has 63326 observations and 11 columns. This dataset is made up of factor, int and numbers.



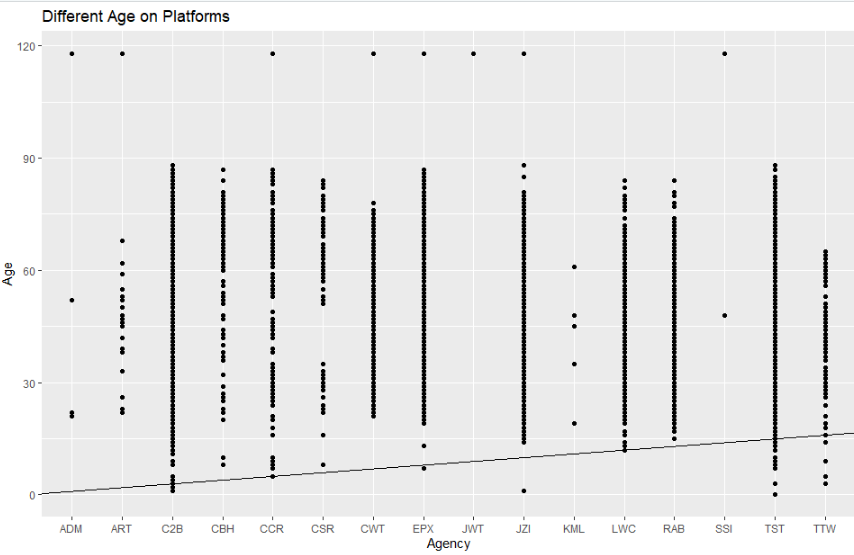
From the summary, we can figure out the number’s min, median, mean, max. To be more specific, I use Three Line Table to explore the information.



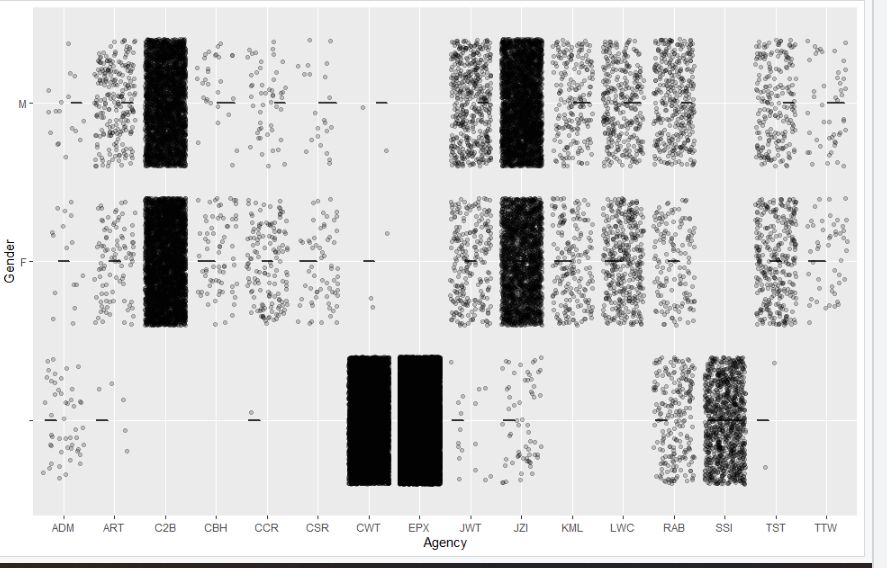
We can see from the graph above that there are 63326 records in this data set. Duration has a mean of 49.3 and a median of 22.0. The mean is bigger than the median, indicating that it is positively skewed. Net Sales have a mean of 40.7 and a median of 26.5. Because the mean is bigger than the median, this is called Positively Skewed. The average age is 40.0, while the median age is 36.0. Because the mean is bigger than the median, this is called Positively Skewed.



The first visualization is a Gender and Age boxplot. We can see the average, quartile, minimum, and maximum values of each group of data from the boxplot. Out of range point, indicating that an outlier exists. In terms of male age, there are two clearly out of range points, suggesting the presence of an outlier. Female age has several out-of-range points as well. When females' ages are compared to males' ages, the female average is younger.



The scatter plot is the second type of visualization. We discover the link and correlation between Agency and Age by monitoring the data analysis of the two. The X-axis represents Agency's name, while the Y-axis represents age. The more authority you have, the more popular it becomes. There are also some out-of-range spots in a few agencies. At the same time, these two variables are positively connected, as the abline of Agency and Age rises.



Jitter plot is the third visualization. This graphic clearly shows the positions of points with darker colors and more data, as well as the placements of points with lighter colors and less data. We can count the number of data points that appear to be clustered. An agency denotes that certain agency occur more frequently than others, and it also denotes that these agencies are more popular with the general population. However, as long as it is a popular agency, the number of men and women is about equal.

Before we can use Jitter Plot, we must first examine several data-related difficulties. What is the distribution of the data points? What are the representative values, and what are the lowest and maximum values? What is the distribution of these points? Are they uniformly spread or concentrated in certain areas? Is the data distribution a long tail distribution? In other words, are there a lot of data points that are far out from the core point set? Or are the majority of the points clustered in one area?