Hello everyone. Thank you for the time in listening my presentation. Today, my topic is about the analysis of avocados in the United States, I would conduct both explorative and predictive data analysis from 2015 to 2020. Specifically, the whole work was finished by R. This is the overview of my whole presentation. It composed of three parts: data preprocessing, explorative data analysis and predictive data analysis. I would mainly focus on the EDA part, including the comparison between two types avocados, the association analysis of price and sold volumes, seasonal patterns analysis of avocado’s price and volumes, the geographical analysis to return the cheapest region in the united states. Meanwhile, the future price trend would be predicted to conduct the analysis.

Now, let’s move to the first part, data preprocessing part. I returned the head of dataset to have an overview of inside components. In this case, I would use six features in the dataset, average price, total volume, year, type and geography, to conduct the following analysis. Furthermore, it is significant to explore the number of missing values, the NAs. From this figure, we could see that the dataset I used do not contain any missing value. As a result, I could move on to the explorative data analysis, which is the most important part of my research. To figure out the difference of the avocado types, I firstly used the density plot to compare the average price distribution of the conventional or organic avocados. As can be seen from the plot, the green stands for the conventional type while orange stands for the organic type. The conventional type suggests the right skewed shape and the organic type suggests the normal distributed shape. In other words, the conventional avocado tends to center around the relative low price and almost none of the avocados’ average price higher than 2$. Contrarily, the organic type distributed more evenly, the center could be around 1.5$. Meanwhile, for better illustration of the price trend in these two types, I plotted the avocado price from 2015 to 2020. It is clear that the organic type always has a higher price compared to the conventional type. Based on previous research, I made a table to summarize the results. I also incorporated the average sold volume feature into consideration. As can be seen from the table, conventional avocado could account for approximately 97% in the markets and the average price would be 28% lower than organic type.