

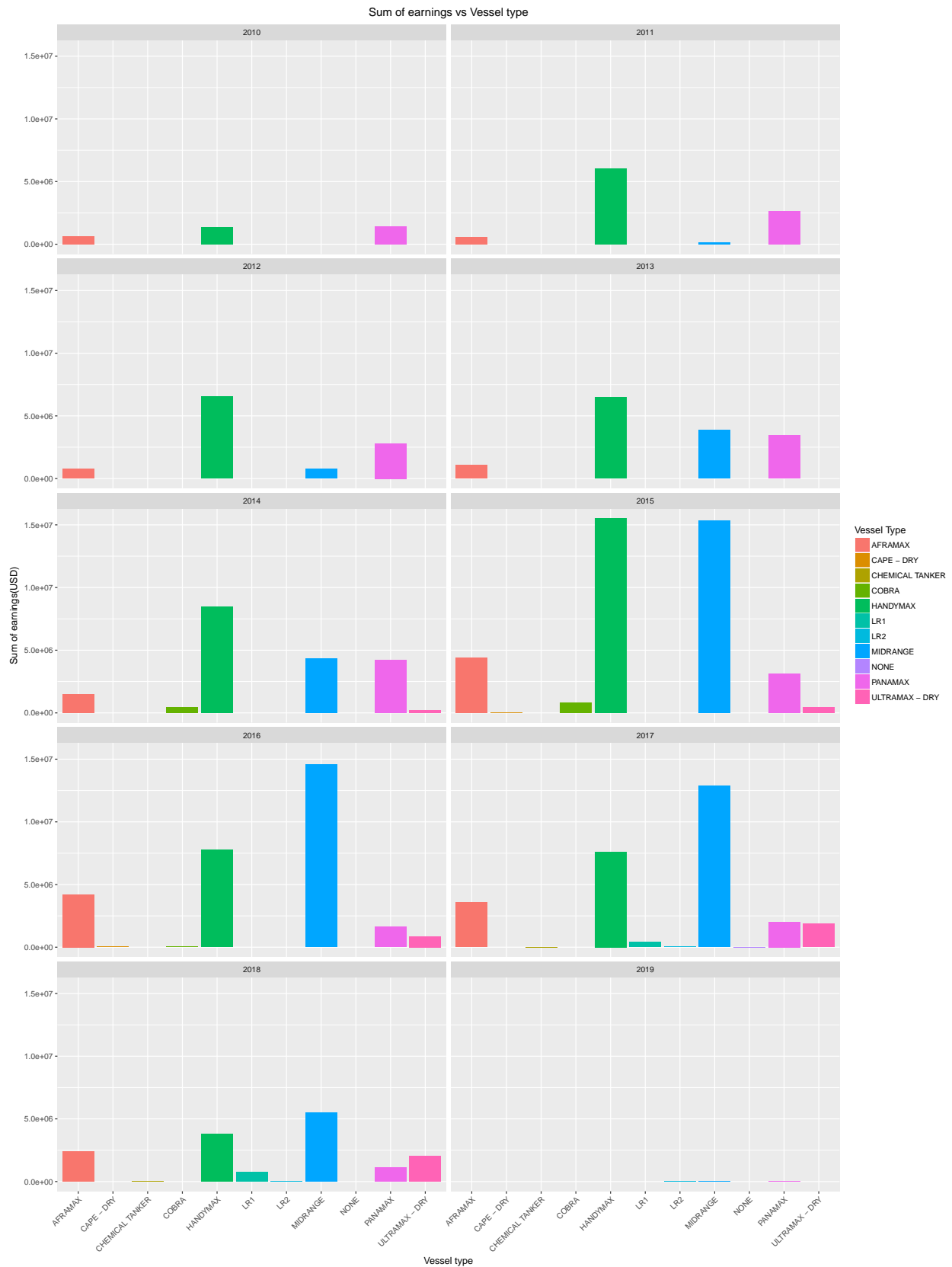
## Statistical analysis for Springboard Capstone Project

After understanding the data set, doing data cleaning and data wrangling now it is time to do some more analysis.

### **Preliminary Exploration:**

The goal of preliminary exploration was to find independent variables that appear to have some predictive power.

I. First we tried to find the top earning vessels by plotting sum of the earnings per year for each vessel type.



It shows that every year Handymax Pool, Midrange Pool are doing good business. Chemical tanker is in business in only 2017. Cobra was in business just for 3 years from 2014, 2015, 2016.

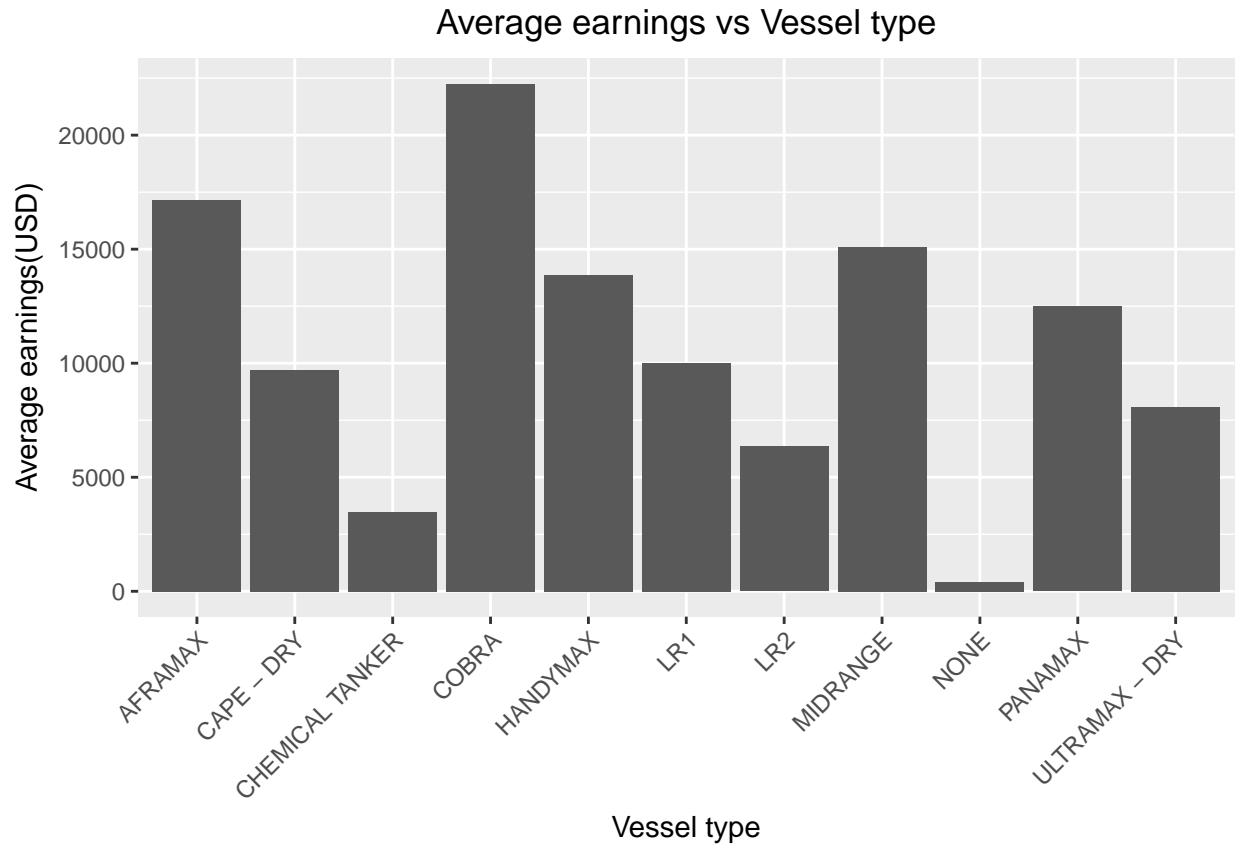
From the plots, out of the total earnings of all the vessel types for all the years, about 80% of the earnings are from Midrange Pool, Handymax Pool, Panamax Pool and Aframax LR2 Pool. Therefore, we can say that majority of business is done by these pool vessel type.

In 2017 out of \$28 million business \$23 million business is done by Midrange Pool, Handymax Pool, Aframax LR2 pool.

II. The next plot is average of the earnings per year for all vessel types.

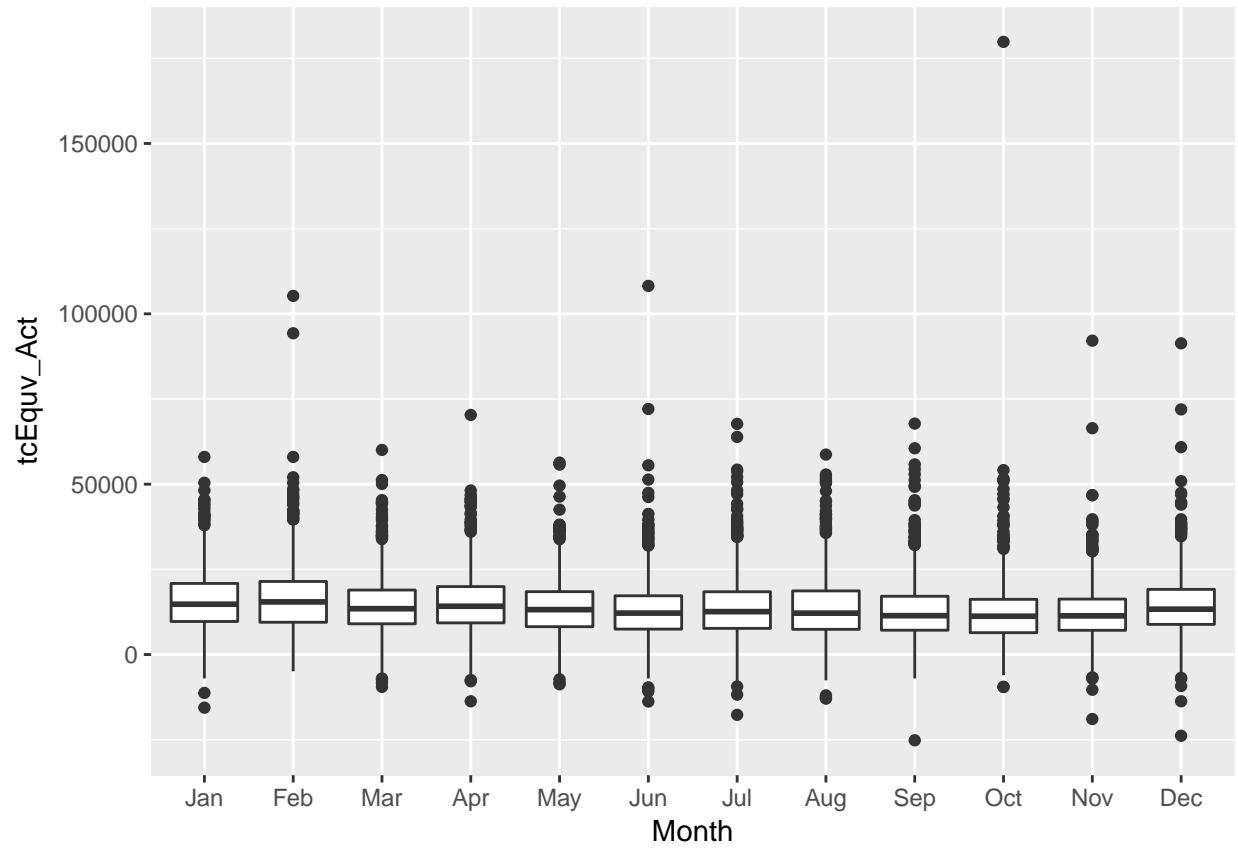


III. Next plots give average earnings for each vessel type for all the years.



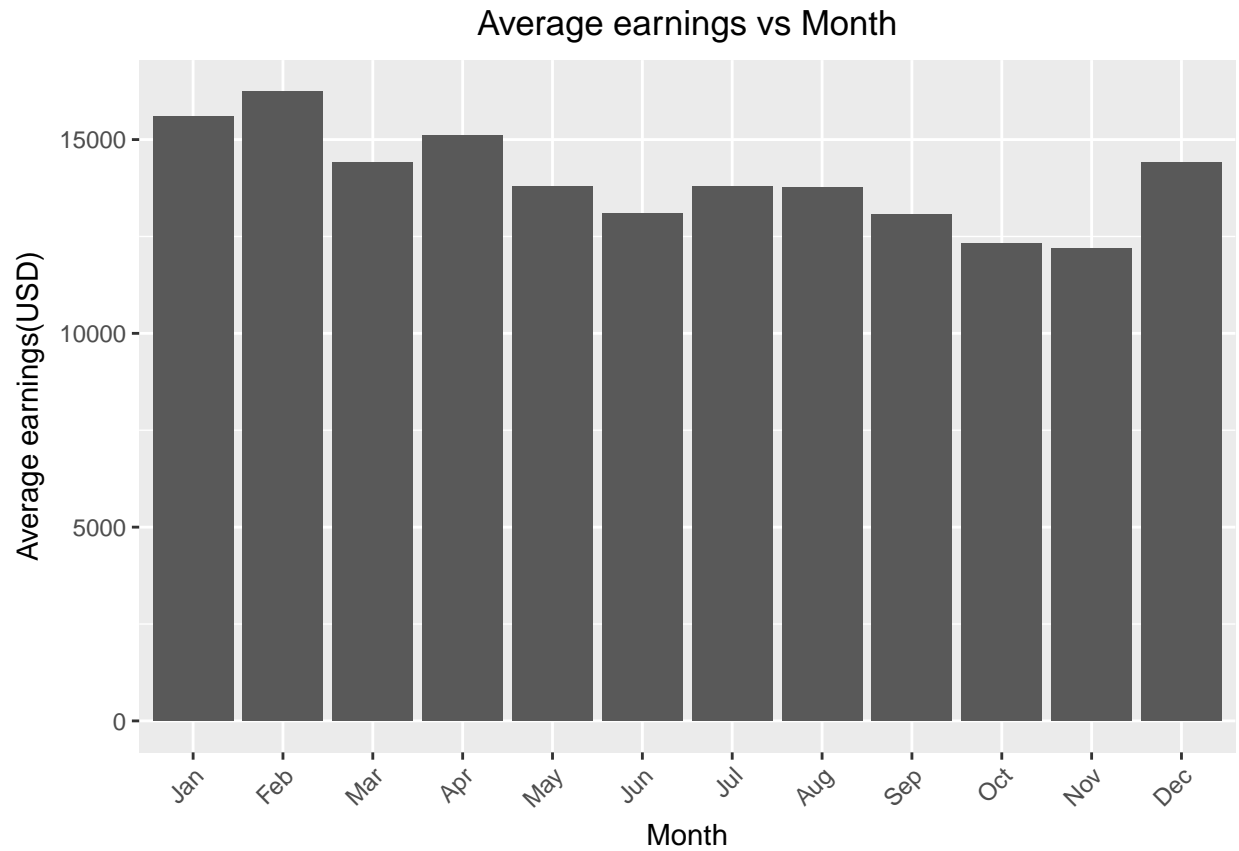
Here it is showing Cobra has highest earnings but Cobra was used just for 3 years (2014,2015,2016). This plot is not so useful as some vessel types has data for 10 years and some are used for couple of years whereas some are recent in the business. It will not be fare to compare these results to find the best one.

IV. I used box plot to plot monthly earning distribution to find out seasonality.



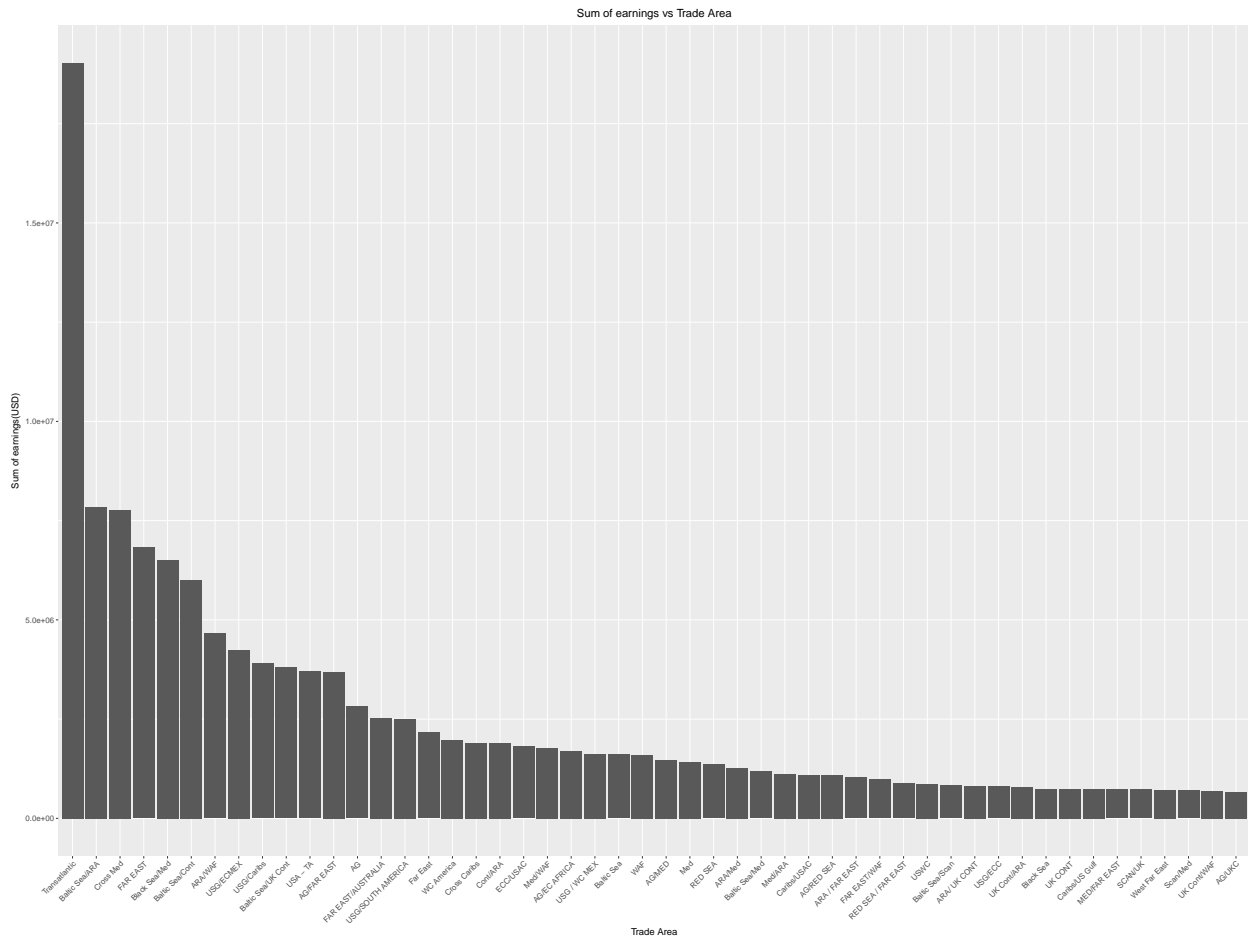
Plot does not show much of seasonality.

Another plot is Average earnings vs Month



This shows the whole year business is going on.

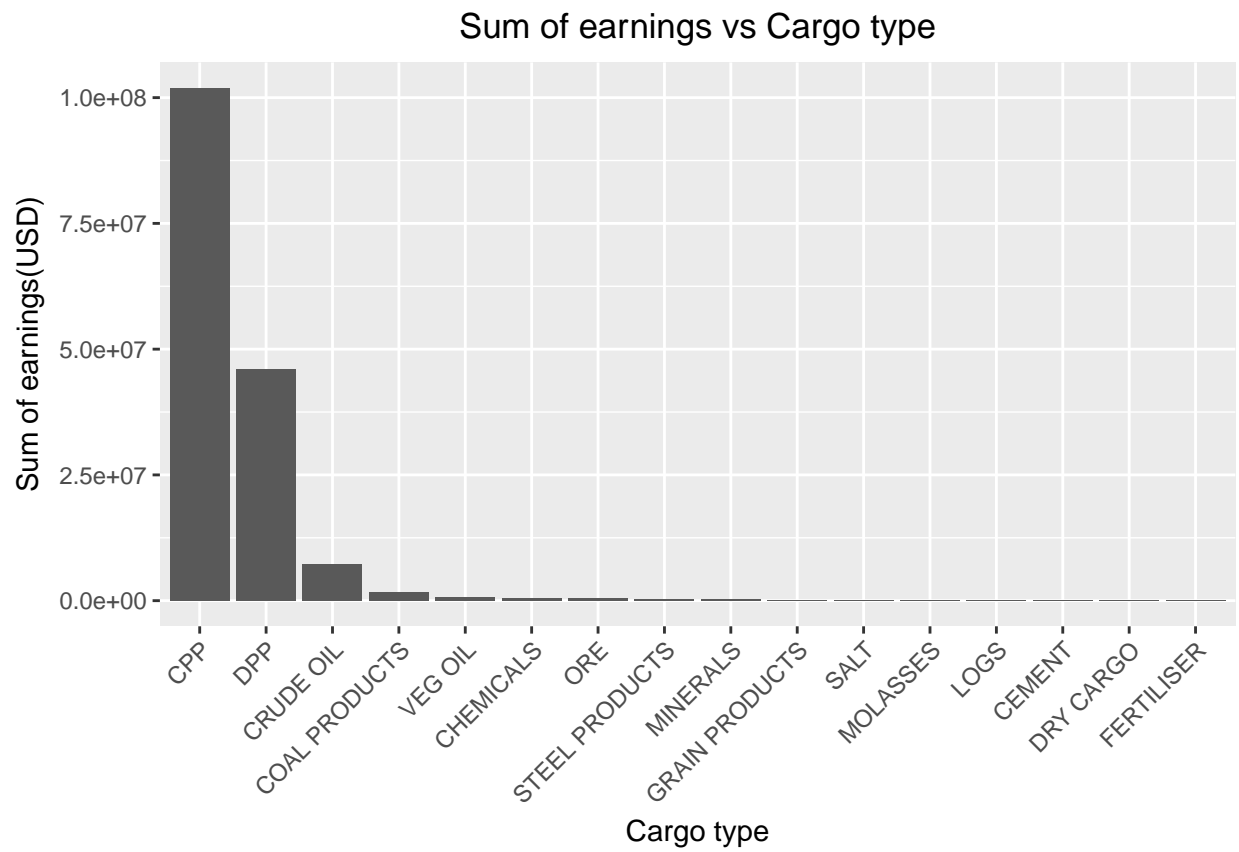
V. This plot has been plotted to find top trade areas from sum of earnings.



This plot shows that major business is done in Transatlantic area.



VI. This plot has been plotted to find top cargo types from sum of earnings.



First 3 cargo types are generating 90% of the business.

Based on this analysis, earnings of these voyages are dependent on mainly below mentioned factors:

- . Vessel type - there are different types of vessels; most profitable are MIDRANGE, HANDYMAX, AFRAMAX, PANAMAX
- . Cargo type- there are mainly two categories of cargo - dry cargo (coal, iron etc.) and liquid cargo (crude oil, gasoline, vegetable oil etc.). Most profitable are FO, CPP, ULSD, Gasoline, Naphtha and Gas Oil etc.
- . Dwt - this is the capacity of the vessel.
- . Actual Cargo lift - How much cargo the particular vessel is carrying affects the earnings.
- . Month - The period of the year in which this voyage is happening also affects the total earnings.
- . Region - the region in which the voyage is starting, loading and discharging also affect its earnings. Most profitable region is Transatlantic area.
- . Daily expense - It includes their daily fuel usage, port fees, jetty charges, food for crew etc. Their 60% of expense is on fuel.

We can say from all of the above analysis and plots that voyages with vessel types like MIDRANGE, HANDYMAX carrying FO, CPP, ULSD, Gasoline, Naphtha and Gas Oil etc. in the transatlantic area have more earnings making the business profitable.