## STA 141A Homework4 HaozheGu 999200555

**The codes and results derived by using these codes constitute my own work. I have consulted the following resources regarding this assignment: NONE**

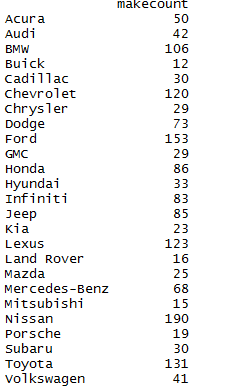
* Q1

The data.frame and .csv file Q1 have columns:

1. Filename
2. Make/Model
3. Model Year
4. VIN Number
5. Price to be sell
6. Mileage
7. Color(Both Exterior and Interior)
8. Transmission
9. Engine Displacement
10. Seller
11. Address of the Company
12. Phone
13. Website

After getting the dataset, I use unique on the model to see if there exist misspelled model name. I correct some of them using code, some of them by hand.

Then I did a maker count for all the cars.



The result show most of the maker in the data frame (excluded some rare maker).

From the result, BMW, Chevrolet, Ford, Lexus. Nissan and Toyota seems to be the most popular maker for the used cars.

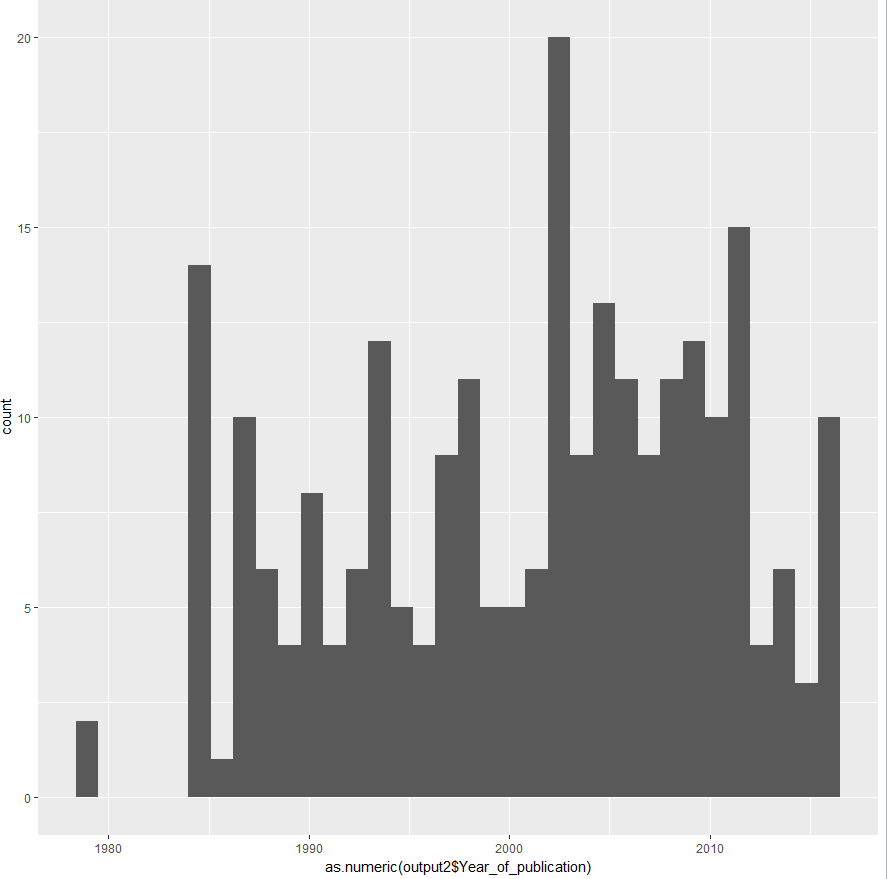
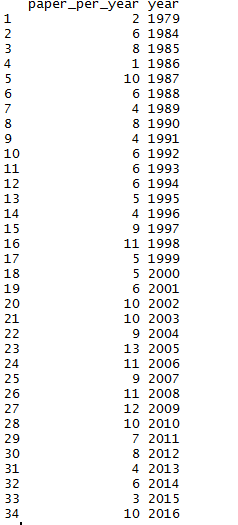
* Q2

The data.frame and .csv file Q2 have columns:

1. Year of Publication
2. Authors
3. Title of Publication
4. Journal of Publication
5. Volume
6. URL

Statistics:

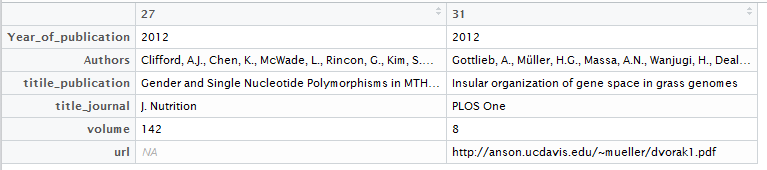
1. Number of publication per year



From the histogram and statistics, year 2009 has the most number of publication. Also, the period from 2002 to 2010 produce the most of the publication.

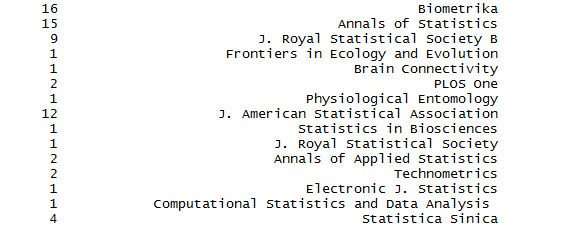
1. Number of Co-Author

The publication with most coauthor has 13 Co-Author, they are at 27, 31



1. Number of Publication in different Journal

There are 90 different Journals, the whole statistic might be too much, so I select some of them in following graph



And Biometrika is also the Journal that appears most frequently

Appendix





