For the Principal component analysis of this Dataset, the data of the 20 columns have first been visualized to know which of them may be of special interest.

In the same way that from a database of results of Olympic athletes we can work on the score of each one, for this database the Horsepower variable has been chosen since it is a determining factor in the performance of a car in terms of at speed and consumption.

Once the dataset is imported, the classes of each column are analyzed to first know what type of data each one contains. In the R language, each column can only contain one data type. This also removes less relevant data columns using the names of the columns themselves to clear up data that may seem unnecessary.

Something very important is to eliminate the outliers, to be able to generate the graphs without problems in scaling or visualization, these being the data that differ excessively from the rest of the set and do not add value to the analysis.

Thus, after "cleaning" the dataset and saving it with the name "cars\_short", tables are made where a header of the first cars in the database is obtained along with the data that we have maintained and then another table that shows the 20 cars with higher horsepower than the interesting fact to analyze.

Next, the Data Exploratory Plots are carried out together with the PCA, obtaining in turn a summary of the data of the 15 cars with the highest power.

Due to complications with the programming of the command line 234, "(cars\_pca <- princomp(cars\_short2, cor=TRUE,) ", where various errors were obtained, to solve it the number of cars has been limited to 15 and the column has been eliminated. “Name” of the “Character” class.

The rest of the graphs show the relative correlations of the data, always based on horsepower.