**Basic R commands**

|  |  |  |
| --- | --- | --- |
| **SNo.** | **Command** | **Comments** |
| 1 | cardata=read.csv(file.choose()) | #importing data |
| 2 | head(cardata) | # specify first 6 rows of data |
| 3 | head(cardata,20) | # specify first 20 rows of data |
| 4 | tail(cardata) | # specify last 6 rows of data |
| 5 | tail(cardata,20) | # specify last 20 rows of data |
| 6 | str(cardata) | # display the internal structure of an r object (Here dataframe: cardata) |
| 7 | head(cardata$Price) | # display price variable data under cardata dataframe |
| 8 | class(cardata) | # outcome to this command will be: [1] "data.frame" |
| 9 | ncol(cardata) | # display the total number of columns |
| 10 | summary(cardata) | # summarizes the Min,Q1, Median, Mean, Q3 and Max data information of the cardata dataframe |
| 11 | summary(cardata$Mileage) | # summarizes the Min,Q1, Median, Mean, Q3 and Max data information for only Mileage variable in the cardata dataframe |
| 12 | summary(cardata$Price) | # summarizes similar information as above for price varaible |
| 13 | sd(cardata$Price) | # return the standard deviation for all price variable values in cardata dataframe |
| 14 | hist(cardata$Price) | # return a histogram plot for price variable |
| 15 | hist(cardata$Mileage) | # return a histogram plot for mileage variable |
| 16 | boxplot(cardata$Price,horizontal = T) | # return a horizontal boxplot for price variable |
| 17 | boxplot(cardata$Mileage, horizontal = T) | # return a horizontal boxplot for mileage variable |
| 18 | plot(density(cardata$Price)) | # return a density curve for price variable |
| 19 | plot(density(cardata$Mileage)) | # return a density curve for Mileage variable |
| 20 | plot(density(log(cardata$Price))) | # return a logrithmic transformed density curve for price variable  # this plot will be nearly unskewed with outliers removed to max possibility  # compare this with the regular density plot of price variable |
| 21 | cardata1=cardata[-c(3:6)] | # return a dataframe cardata1 by removing column 3,4,5,6 from original dataframe: cardata  # this is done to remove categorical data availble in c3 to c6 here  # this is done to execute next two commands of fing correlation and covariance(these can only be done for numerical data) |
| 22 | cor(cardata1) | #find correlation for all the variables in cardata1 |
| 23 | cov(cardata1) | #find covariance for all the variables in cardata1 |
| 24 | plot(cardata$Price,cardata$Mileage) | # return a scatter plot for price vs mileage variables |
| 25 | pairs(cardata1) | # return an array of scatter plots considering all combinations for variables |
| 26 | table(cardata$Make) | # return frequency count for all "Make" categories/variable |
| 27 | prop.table(table(cardata$Make))\*100 | # return percentages share for all "Make" categories/variable |
| 28 | table(cardata$Cylinder) | # return frequency count for all "cylinder" categories/variable |
| 29 | prop.table(table(cardata$Cylinder))\*100 | # return percentages share for all "cylinder" categories/variable |
| 30 | table(cardata$Leather) | # return frequency count for all "leather" categories/variable |
| 31 | table(cardata$Sound) | # return frequency count for all "sound" categories/variable |
| 32 | barplot(table(cardata$Make)) | # return a bar chart for all Make variable data |
| 33 | pie(table(cardata$Make)) | # return a bar chart for all Make variable data |
| 34 | table(cardata$Make,cardata$Sound) | # return the information on total number of cars among different Make(model or company) with and without Sound variable (Sound system) |
| 35 | barplot(table(cardata$Sound,cardata$Make)) | # return a stack chart on above information |
| 36 | barplot(table(cardata$Sound,cardata$Make),legend.text = T) | # return the same stack chart as above with legends defined |
| 37 | barplot(table(cardata$Sound,cardata$Make),legend.text = T,beside = T) | # return the same chart with two individual bars for sound variable for each make type |
| 38 | buickcadillac=subset(cardata,cardata$Make=="Buick" | cardata$Make=="Cadillac") | # sub setting |
| 39 | aggregate(cardata$Price~cardata$Sound,FUN = mean) | # return the mean prices for cars with and without sound system |
| 40 | aggregate(cardata$Price~cardata$Sound,FUN = median) | # return the median prices for cars with and without sound system |
| 41 | aggregate(cardata$Price~cardata$Make,FUN = mean) | # return the mean prices for all cars for different Make(model/company) |
| 42 | makeavg=aggregate(cardata$Price~cardata$Make,FUN = mean) | #assigning a user-defined function (here makeavg) for the above mean price command  #now "makeavg" can be used further in the code to return same output as of the above command |
| 43 | aggregate(cardata$Price~cardata$Make+cardata$Leather,FUN = mean) | #return mean prices for Individual Car Make categories, with and without Leather(leather finishing) |
| 44 | aggregate(cardata$Price~cardata$Make+cardata$Leather+cardata$Cruise,FUN = mean) | #return mean prices for Individual Car Make categories, with and without Leather(leather finishing) and cruise control |
| 45 | aggregate(cardata$Price~cardata$Type+cardata$Mileage,FUN = mean) | #return mean prices for Individual Car Type categories, with different Mileage values |
| 46 | aggregate(cbind(cardata$Price,cardata$Mileage)~cardata$Type,FUN = mean) | # return mean price and mean mileage for individual car Type  # the mean values will be displayed under column v1 and v2 |
| 47 | aggregate(cbind(Price=cardata$Price,Mile=cardata$Mileage)~cardata$Type, FUN=mean) | # same as above just the v1 and v2 columns will be displayed as Price and Mile |