```
if (!require('tidyverse'))
{
  install.packages('tidyverse');
  library(tidyverse);
}
```

```
## Loading required package: tidyverse
```

```
## Warning: package 'tidyverse' was built under R version 4.1.1
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
## Warning: package 'readr' was built under R version 4.1.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
if(!require('readxl'))
{
install.packages("readxl")
library(readxl)
}
```

```
## Loading required package: readxl
```

```
## Warning: package 'readxl' was built under R version 4.1.1
```

```
if(!require('readr'))
{
  install.packages('readr')
library(readr)
}
```

```
if(!require('caret'))
{
  install.packages('caret')
library(caret)
}
```

```
## Loading required package: caret
```

```
## Warning: package 'caret' was built under R version 4.1.1
```

```
## Loading required package: lattice
```

```
## Warning: package 'lattice' was built under R version 4.1.1
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

Practical descriptive analytics assignment (Lecture 1, Lecture3)

```
    Automotive market data
```

This is the dataset for Russian auto market sales (total), starting from the year 2005.

- Data are monthly, from January 2005 until June 2018
- The following are the dataset variables:

– Date - month and year – Total Sales - the number of new cars sold in this month – CPI - Consumer price index (inflation). Given in percentage points, so value 12.5 means 12.5%. – Oil - Prices for Brent oil, per barrel, in USD. – CCI - Consumer confidence index. Given in percentage points, so value 99.67382 means 99.67382% – BCI - Business confidence index. Given in percentage points, same as CPI. – Prime rate - interbank exchange rate. Given in percentage points, same as previous variables.

```
#read dataset:
auto_data <- read_excel("Auto market.xlsx")
```

```
# see summary of data:
summary(auto_data)
```

```
##       Date                        Total Sales          CPI
##  Min.   :2005-01-01 00:00:00   Min.   : 67670   Min.   : 2.200
##  1st Qu.:2008-05-08 18:00:00   1st Qu.:118181   1st Qu.: 6.450
##  Median :2011-09-16 00:00:00   Median :150011   Median : 8.110
##  Mean   :2011-09-16 03:42:13   Mean   :161963   Mean   : 8.859
##  3rd Qu.:2015-01-24 06:00:00   3rd Qu.:213279   3rd Qu.:11.822
##  Max.   :2018-06-01 00:00:00   Max.   :270914   Max.   :16.930
##       Oil              USD             CCI              BCI
##  Min.   : 30.70   Min.   :23.35   Min.   : 93.26   Min.   : 96.13
##  1st Qu.: 55.95   1st Qu.:25.36   1st Qu.: 98.84   1st Qu.: 99.90
##  Median : 72.31   Median :31.02   Median :101.02   Median :100.30
##  Mean   : 77.48   Mean   :37.74   Mean   :100.10   Mean   :100.07
##  3rd Qu.:106.01   3rd Qu.:54.25   3rd Qu.:101.75   3rd Qu.:100.55
##  Max.   :132.72   Max.   :77.29   Max.   :103.64   Max.   :101.41
##    Prime rate
##  Min.   : 5.500
##  1st Qu.: 8.062
##  Median :10.000
```

```
##  Mean   : 9.812
##  3rd Qu.:11.000
##  Max.   :17.000
```

```
head(auto_data)
```

```
## # A tibble: 6 x 8
##   Date                ‘Total Sales‘   CPI   Oil   USD   CCI   BCI ‘Prime rate‘
##   <dttm>                      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>        <dbl>
## 1 2005-01-01 00:00:00         74918  12.7  44.5  24.8  99.7  100.           13
## 2 2005-02-01 00:00:00         88947  13.0  45.5  24.6  99.8  100.           13
## 3 2005-03-01 00:00:00        108563  13.6  53.1  25.0 100.   100.           13
## 4 2005-04-01 00:00:00        113716  13.8  51.9  25.4 100.   100.           13
## 5 2005-05-01 00:00:00        100923  13.8  48.6  24.9 101.   100.           13
## 6 2005-06-01 00:00:00        107553  13.7  54.4  25.2 101.   100.           13
```

```
# to see the column names in the dataset:
colnames(auto_data)
```

```
## [1] "Date"        "Total Sales" "CPI"         "Oil"         "USD"
## [6] "CCI"         "BCI"         "Prime rate"
```

```
str(auto_data)
```

```
## tibble [162 x 8] (S3: tbl_df/tbl/data.frame)
##  $ Date       : POSIXct[1:162], format: "2005-01-01" "2005-02-01" ...
##  $ Total Sales: num [1:162] 74918 88947 108563 113716 100923 ...
##  $ CPI        : num [1:162] 12.7 13 13.6 13.8 13.8 ...
##  $ Oil        : num [1:162] 44.5 45.5 53.1 51.9 48.6 ...
##  $ USD        : num [1:162] 24.8 24.6 25 25.4 24.9 ...
##  $ CCI        : num [1:162] 99.7 99.8 100 100.3 100.7 ...
##  $ BCI        : num [1:162] 100 100 100 100 100 ...
##  $ Prime rate : num [1:162] 13 13 13 13 13 13 13 13 13 13 ...
```

```
# confirm date is stored quantitative:
date <- unclass(auto_data$Date)
# see dataset of date:-
head(date)
```

```
## [1] 1104537600 1107216000 1109635200 1112313600 1114905600 1117584000
```

#In the dataset two variable names having space in the name so we have to correct them first by removing space in between the name of the variable removing space between the variable name

```
auto_data <- auto_data %>%
  rename(TotalSales = "Total Sales", Primerate = "Prime rate")
```

1. Create a table that will contain the following: Variable Name, Quantitative or Qualitative, Measurement level, Appropriate charts:

2. Fill the table with all of the variables in the dataset, classifying them appropriately. You will need to indicate whether variable is quantitative or qualitative, its measurement level (nominal, ordinal,interval, ratio, counts), and which charts (pie, bar, line, boxplot, etc.) are appropriate for this variable.

```
VariableName <- names(auto_data)

QuanQual <- c("quantitative")

measureLevel <- c("interval","ratio","ratio","ratio","ratio","ratio","ratio","ratio")

appChart <- c("lineplot, barplot","boxplot/ Histogram/ Density plots/ QQ plot", "boxplot/ Histogram/ De
              "boxplot/ Histogram/ Density plots/ QQ plot", "boxplot/ Histogram/ Density plots/ QQ plot
              "boxplot/ Histogram/ Density plots/ QQ plot","boxplot/ Histogram/ Density plots/ QQ plot"
```

```
DataTable <- data.frame(VariableName,QuanQual,measureLevel,appChart) %>%
  rename("Variable Name" = VariableName, "Quantitative or Qualitative" =QuanQual," Measurement level" =
```

```
# installing packages knitr
if(!require('knitr'))
{
install.packages('knitr')
library(knitr)
}
```

```
## Loading required package: knitr
```

```
## Warning: package 'knitr' was built under R version 4.1.1
```

```
if(!require('kableExtra'))
{
install.packages("kableExtra")
library(kableExtra)
}
```

```
## Loading required package: kableExtra
```

```
## Warning: package 'kableExtra' was built under R version 4.1.1
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
kable(DataTable, booktabs = TRUE) %>%
  kable_styling(font_size = 10)
```

#3. For any three variables in the dataset, create an appropriate chart. Feel free to be creative with colors,titles, etc.

Variable Name	Quantitative or Qualitative	Measurement level	Appropriate charts
Date	quantitative	interval	lineplot, barplot
TotalSales	quantitative	ratio	boxplot/ Histogram/ Density plots/ QQ plot
CPI	quantitative	ratio	boxplot/ Histogram/ Density plots/ QQ plot
Oil	quantitative	ratio	boxplot/ Histogram/ Density plots/ QQ plot
USD	quantitative	ratio	boxplot/ Histogram/ Density plots/ QQ plot
CCI	quantitative	ratio	boxplot/ Histogram/ Density plots/ QQ plot
BCI	quantitative	ratio	boxplot/ Histogram/ Density plots/ QQ plot
Primerate	quantitative	ratio	boxplot/ Histogram/ Density plots/ QQ plot

```
if(!require('ggplot2'))
{
install.packages("ggplot2")
   library(ggplot2)
}
```

```
if(!require('lattice'))
{
install.packages('lattice')
  library(lattice)
}
```

```
# plotting boxplot for total sales-
   boxplot(auto_data[,2],main="Boxplot for Total Sale (from jan 2005 to jun 2018)",col="red",sub= "data
```