Assignment 01

Vasanthakumar Kalaikkovan

07/09/2021

1. Import, Plot, Summarize, and Save Data

Set the working directory to the root of your DSC 520 directory

```
setwd("C:/Users/vasan/Predictive-Analytics/Week 01")
```

```
# importing the data
df<-read.csv("Dataset.csv")
##checking the head
head(df)
```

```
##   ï..Year   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov
## 1    2011 184.5 186.7 189.3 191.6 192.7 191.6 192.4 191.8 192.8 191.9 191.9
## 2    2012 192.1 193.0 194.5 195.0 193.8 192.9 193.3 195.5 196.9 196.4 194.5
## 3    2013 194.9 196.4 196.7 196.0 196.9 197.3 197.3 197.9 197.3 196.9 196.1
## 4    2014 198.1 198.9 200.3 202.1 201.8 202.9 203.0 202.5 201.7 200.4 198.2
## 5    2015 192.2 192.6 193.6 193.1 196.0 197.7 197.4 196.3 193.4 192.4 191.6
## 6    2016 189.9 188.8 189.2 190.3 191.7 193.8 193.5 192.6 193.2 193.7 192.4
##     Dec
## 1 191.2
## 2 193.8
## 3 196.5
## 4 195.4
## 5 190.1
## 6 193.7
```

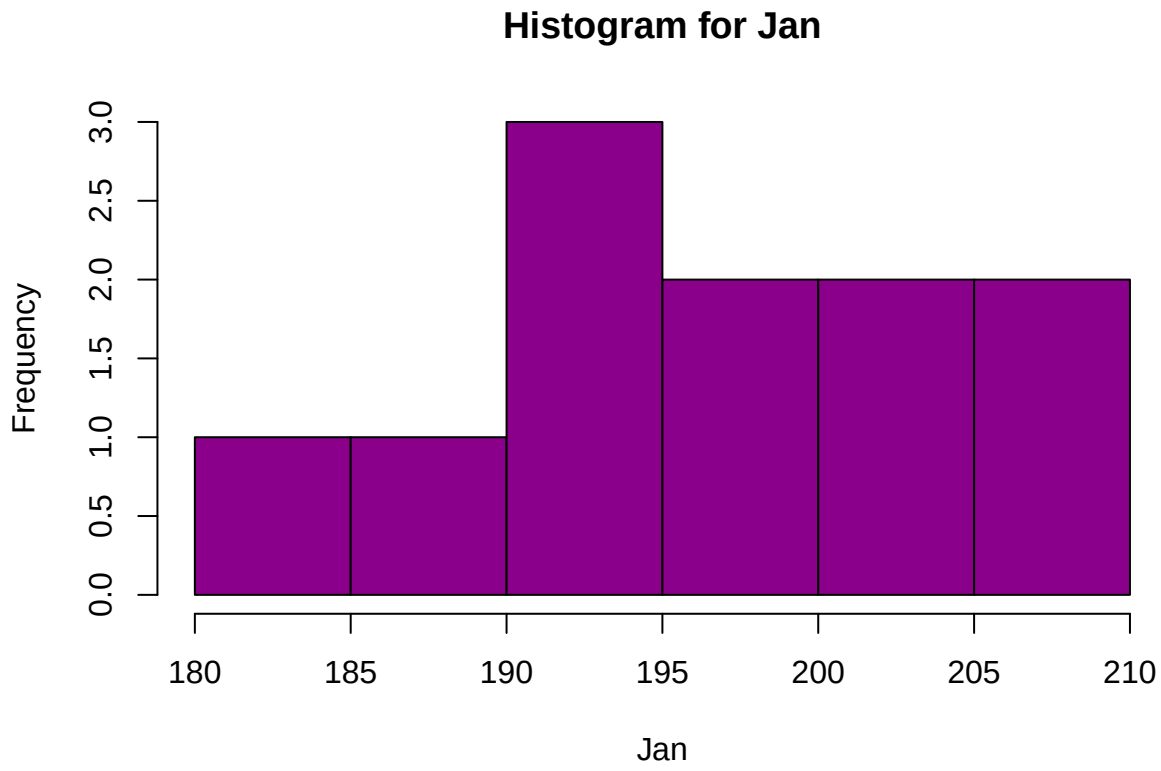
```
# summary for  Jan Variable
summary(df['Jan'])
```

```
##       Jan
##  Min.   :184.5
##  1st Qu.:192.2
##  Median :195.4
##  Mean   :196.8
##  3rd Qu.:201.4
##  Max.   :207.6
```

```
# summary for Year variable
summary(df['Feb'])
```

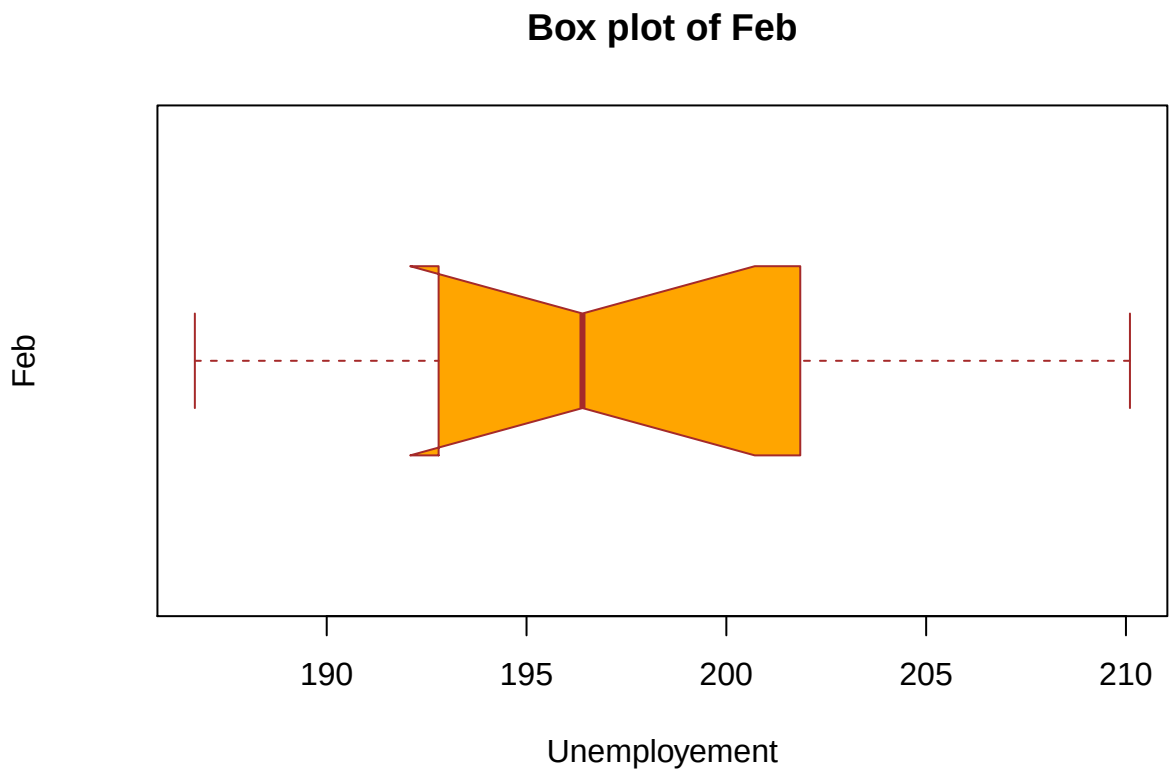
```
##       Feb
##  Min.   :186.7
##  1st Qu.:192.8
##  Median :196.4
##  Mean   :197.4
##  3rd Qu.:201.8
##  Max.   :210.1
```

```
# Histogram
hist(df$Jan,main="Histogram for Jan",
     xlab="Jan",
     col="darkmagenta")
```

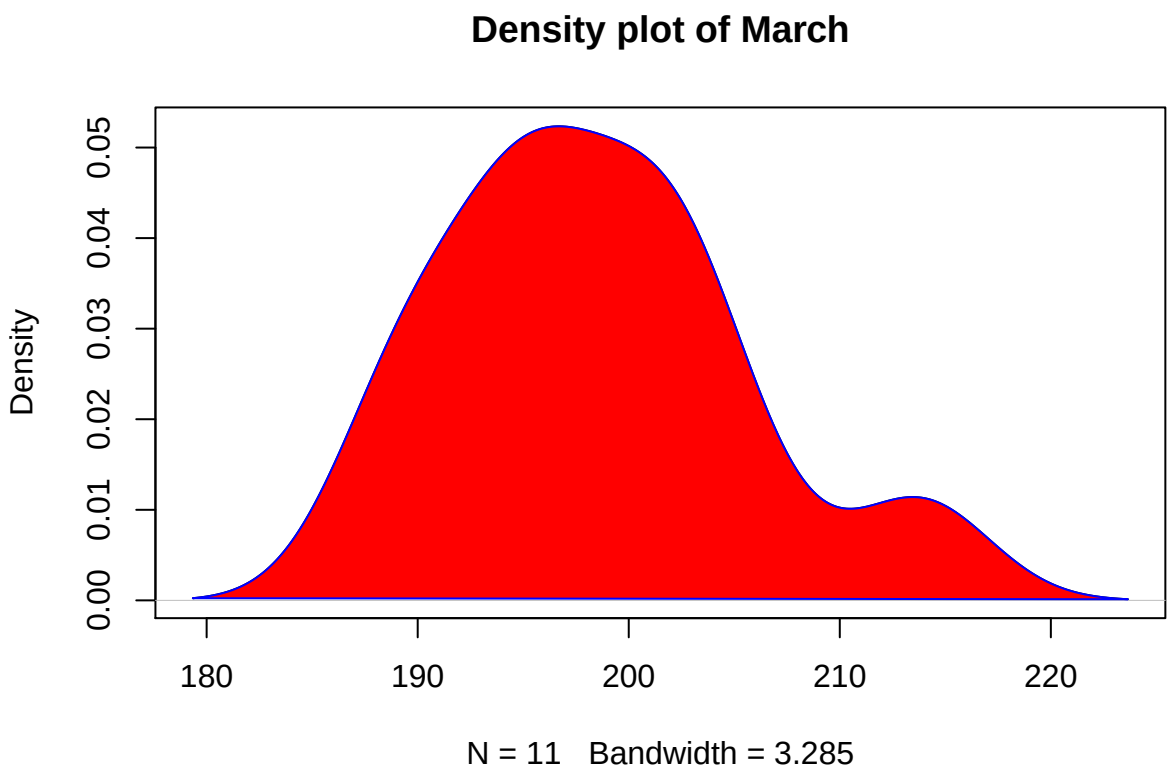



```
#Box Plot
boxplot(df$Feb,main = "Box plot of Feb",
        xlab = "Unemployement",
        ylab = "Feb",
        col = "orange",
        border = "brown",
        horizontal = TRUE,
        notch = TRUE)
```

```
## Warning in bxp(list(stats = structure(c(186.7, 192.8, 196.4, 201.85, 210.1: some
## notches went outside hinges ('box'): maybe set notch=FALSE
```




```
#Density plot
d<-density(df$Mar)
plot(d,main="Density plot of March")
polygon(d,col="red",border="blue")
```




```
#Save the data locally as csv
write.csv(df,"Export.csv", row.names = FALSE)
```

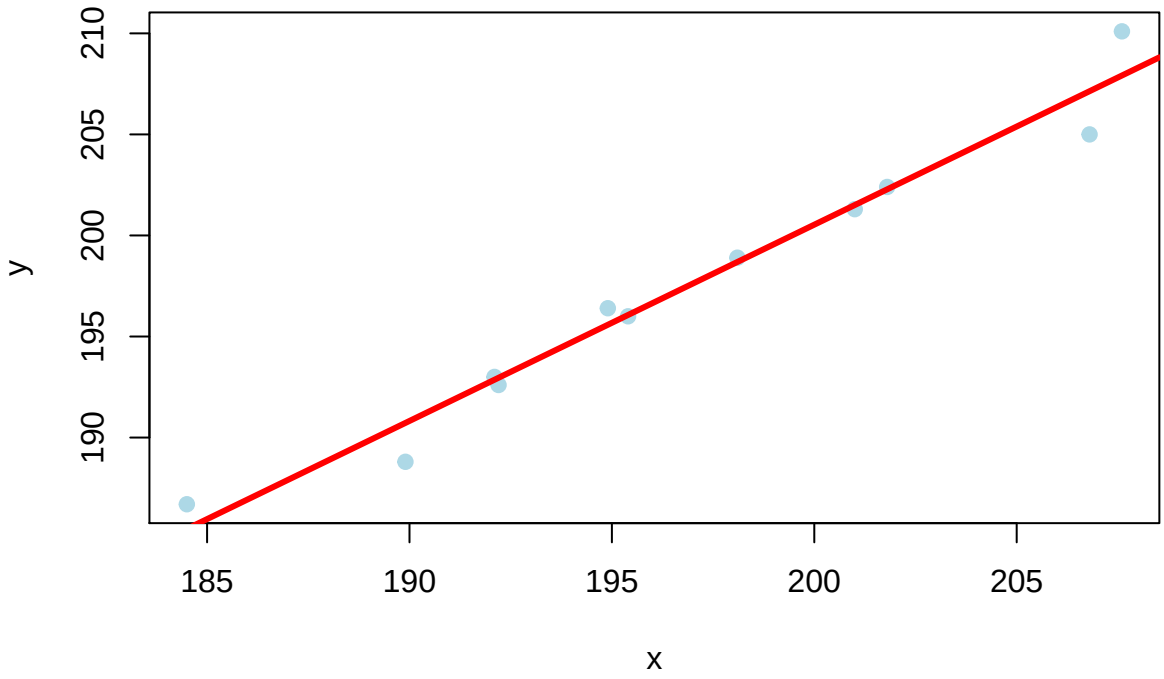
2. Explore Some Bivariate Relations

```
# Correlation plot
x<-df$Jan
y<-df$Feb

# Creating the plot
plot(x, y, pch = 19, col = "lightblue")

# Regression line
abline(lm(y ~ x), col = "red", lwd = 3)

# Pearson correlation
text(paste("Correlation:", round(cor(x, y), 2)), x = 25, y = 95)
```




```
#Cross Tabulation
library(gmodels)
```

```
## Warning: package 'gmodels' was built under R version 4.0.5
```

```
CrossTable(df$Jan, df$Feb)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## | Chi-square contribution |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  11
##
##
##              | df$Feb
##       df$Jan |     186.7 |     188.8 |     192.6 |       193 |       196 |     196.4 |     198.9 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
```

```
##        184.5 |         1 |         0 |         0 |         0 |         0 |         0 |         0 |
##              |     9.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |
##              |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.091 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        189.9 |         0 |         1 |         0 |         0 |         0 |         0 |         0 |
##              |     0.091 |     9.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |
##              |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.091 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        192.1 |         0 |         0 |         0 |         1 |         0 |         0 |         0 |
##              |     0.091 |     0.091 |     0.091 |     9.091 |     0.091 |     0.091 |     0.091 |
##              |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.091 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        192.2 |         0 |         0 |         1 |         0 |         0 |         0 |         0 |
##              |     0.091 |     0.091 |     9.091 |     0.091 |     0.091 |     0.091 |     0.091 |
##              |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.091 |     0.000 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        194.9 |         0 |         0 |         0 |         0 |         0 |         1 |         0 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     9.091 |     0.091 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.091 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        195.4 |         0 |         0 |         0 |         0 |         1 |         0 |         0 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     9.091 |     0.091 |     0.091 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.091 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        198.1 |         0 |         0 |         0 |         0 |         0 |         0 |         1 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     9.091 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     1.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.091 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##          201 |         0 |         0 |         0 |         0 |         0 |         0 |         0 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        201.8 |         0 |         0 |         0 |         0 |         0 |         0 |         0 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
```

```
##        206.8 |         0 |         0 |         0 |         0 |         0 |         0 |         0 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##        207.6 |         0 |         0 |         0 |         0 |         0 |         0 |         0 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
##              |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
## Column Total |         1 |         1 |         1 |         1 |         1 |         1 |         1 |
##              |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |     0.091 |
## -------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|---
##
##
```

##3. Organize a Data Report

```
#Summary for each variable
summary(df$ï..Year)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2011    2014    2016    2016    2018    2021
```

```
summary(df$Jan)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   184.5   192.2   195.4   196.8   201.4   207.6
```

```
summary(df$Feb)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   186.7   192.8   196.4   197.4   201.8   210.1
```

```
summary(df$Mar)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   189.2   194.1   196.7   198.4   201.9   213.8
```

```
summary(df$Apr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   190.3   194.1   196.1   198.8   202.4   214.7
```

```
summary(df$May)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   191.7   194.9   197.0   200.2   203.4   218.4
```

```
summary(df$Jun)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   191.6   195.6   197.8   200.9   204.4   221.3
```

```
summary(df$Jul)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   192.4   195.4   197.6   201.3   204.7   222.9
```

```
summary(df$Aug)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   194.1   197.9   180.9   202.8   206.4
```

```
summary(df$Sep)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   193.3   197.3   180.9   202.5   205.9
```

```
summary(df$Oct)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   193.1   196.9   180.7   202.1   206.7
```

```
summary(df$Nov)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   192.2   196.1   180.0   202.1   206.4
```

```
summary(df$Dec)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   192.4   195.4   179.4   201.2   206.0
```

```
#Structure
structure(df)
```

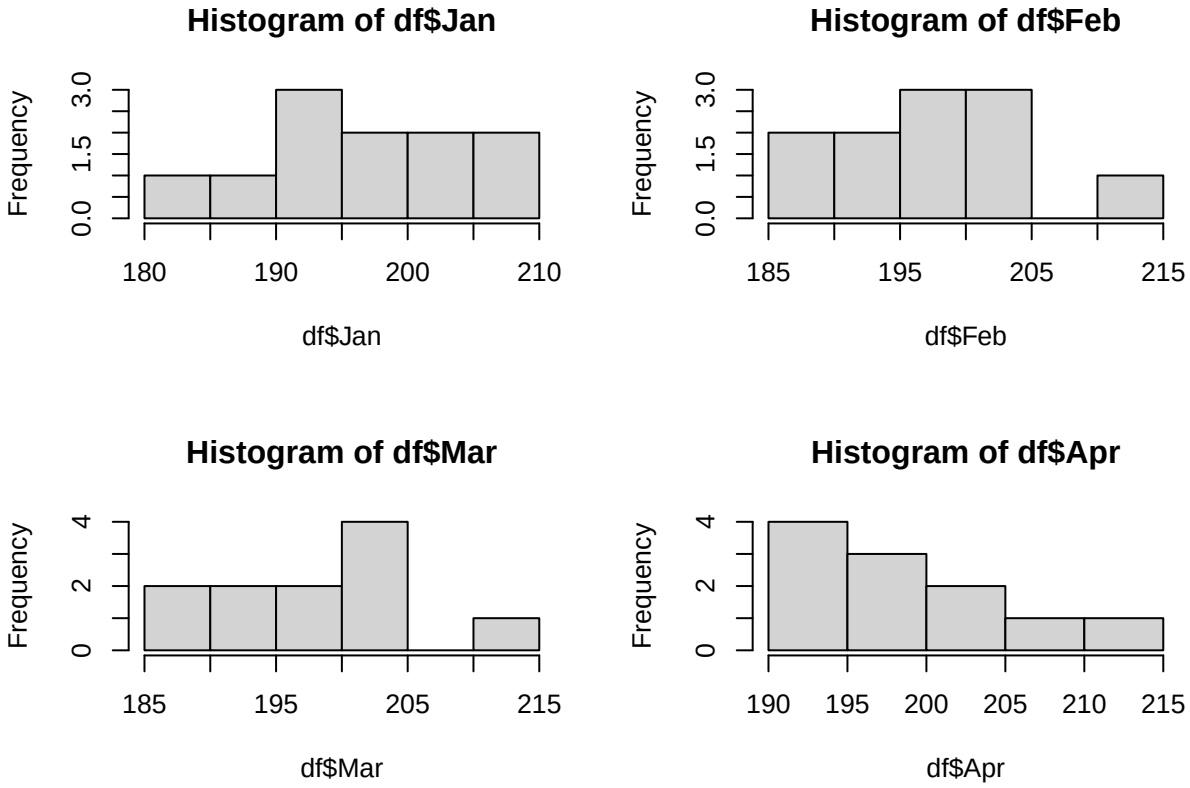
```
##    ï..Year   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov
## 1     2011 184.5 186.7 189.3 191.6 192.7 191.6 192.4 191.8 192.8 191.9 191.9
## 2     2012 192.1 193.0 194.5 195.0 193.8 192.9 193.3 195.5 196.9 196.4 194.5
## 3     2013 194.9 196.4 196.7 196.0 196.9 197.3 197.3 197.9 197.3 196.9 196.1
## 4     2014 198.1 198.9 200.3 202.1 201.8 202.9 203.0 202.5 201.7 200.4 198.2
## 5     2015 192.2 192.6 193.6 193.1 196.0 197.7 197.4 196.3 193.4 192.4 191.6
## 6     2016 189.9 188.8 189.2 190.3 191.7 193.8 193.5 192.6 193.2 193.7 192.4
## 7     2017 195.4 196.0 196.3 198.0 197.0 197.8 197.6 198.4 199.6 199.4 200.4
## 8     2018 201.0 201.3 202.1 202.8 205.0 205.9 206.0 205.8 205.9 206.7 204.4
## 9     2019 201.8 202.4 204.9 207.1 207.6 206.9 207.4 206.4 205.7 206.3 206.4
```

```
## 10    2020 206.8 205.0 201.8 196.1 200.9 202.2 203.3 203.2 203.3 203.8 203.8
## 11    2021 207.6 210.1 213.8 214.7 218.4 221.3 222.9   0.0   0.0   0.0   0.0
##      Dec
## 1  191.2
## 2  193.8
## 3  196.5
## 4  195.4
## 5  190.1
## 6  193.7
## 7  199.9
## 8  202.5
## 9  206.0
## 10 204.4
## 11   0.0
```

```
#Type of data elements
str(df)
```

```
## 'data.frame':    11 obs. of  13 variables:
##  $ ï..Year: int  2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 ...
##  $ Jan    : num  184 192 195 198 192 ...
##  $ Feb    : num  187 193 196 199 193 ...
##  $ Mar    : num  189 194 197 200 194 ...
##  $ Apr    : num  192 195 196 202 193 ...
##  $ May    : num  193 194 197 202 196 ...
##  $ Jun    : num  192 193 197 203 198 ...
##  $ Jul    : num  192 193 197 203 197 ...
##  $ Aug    : num  192 196 198 202 196 ...
##  $ Sep    : num  193 197 197 202 193 ...
##  $ Oct    : num  192 196 197 200 192 ...
##  $ Nov    : num  192 194 196 198 192 ...
##  $ Dec    : num  191 194 196 195 190 ...
```

```
#Results
par(mfrow=c(2,2))
hist(df$Jan)
hist(df$Feb)
hist(df$Mar)
hist(df$Apr)
```



Results: 1. For the last three years the unemployment in January was almost same and its continuous. 2. In Feb Month there is a missing in the continuity. 3. Similarly to the Jan there is almost same amount of unemployment in the initial stage and also at the end there is no continuity. 4. For April month there is a continuity through out all years but the frequency is not same.