Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Programme	:	**B.Tech – ECE and ECM**	Semester	:	**Win 2022**
Course	:	**Essentials of Data Analytics Lab**	Code	:	**CSE3506**
Faculty	:	**Gobinath N**	Slot	:	**L51 + L52**

Consider Temperature data of any Indian city for Jan 2019, Jan 2020, Jan 2021. Compare their mean for any significant difference.

- ⇨ Here, I have collected Data for Mumbai Temperature for January month for three consecutive years i.e. 2017,18,19, as the mentioned year data was not available.

Setting the working directories:

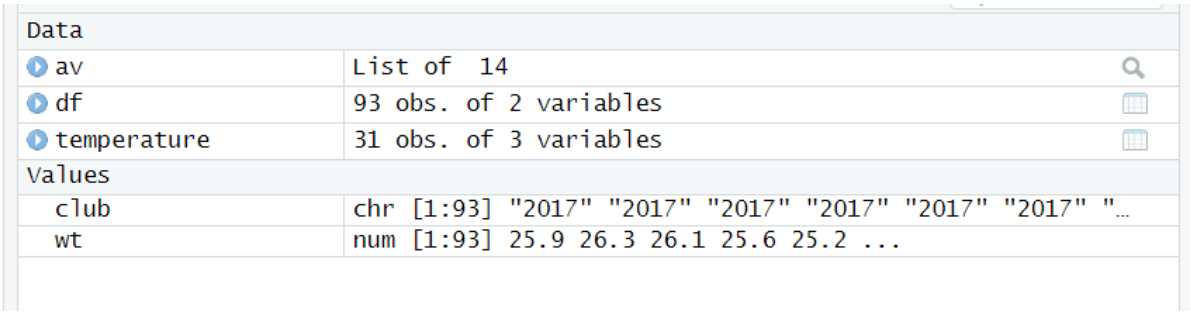
```
#Lab3_AssignedWork

rm(list=ls())

setwd("C:\\Users\\Rituraj Anand\\Desktop\\Sem6\\CSE3506\\LAB\\Lab 3")
```

Reading the csv file:

```
temperature=read.csv("mumbaiTemp.csv")
```



Clubbing the data

```
club=c(rep('2017',31),rep('2018',31),rep('2019',31))

club
```

```
> club=c(rep('2017',31),rep('2018',31),rep('2019',31))
> club
 [1] "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017"
[14] "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017" "2017"
[27] "2017" "2017" "2017" "2017" "2017" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018"
[40] "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018"
[53] "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2018" "2019" "2019" "2019"
[66] "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019"
[79] "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019" "2019"
[92] "2019" "2019"
```

```
wt=c(temperature$Jan.17,temperature$Jan.18,temperature$Jan.19)

wt
```

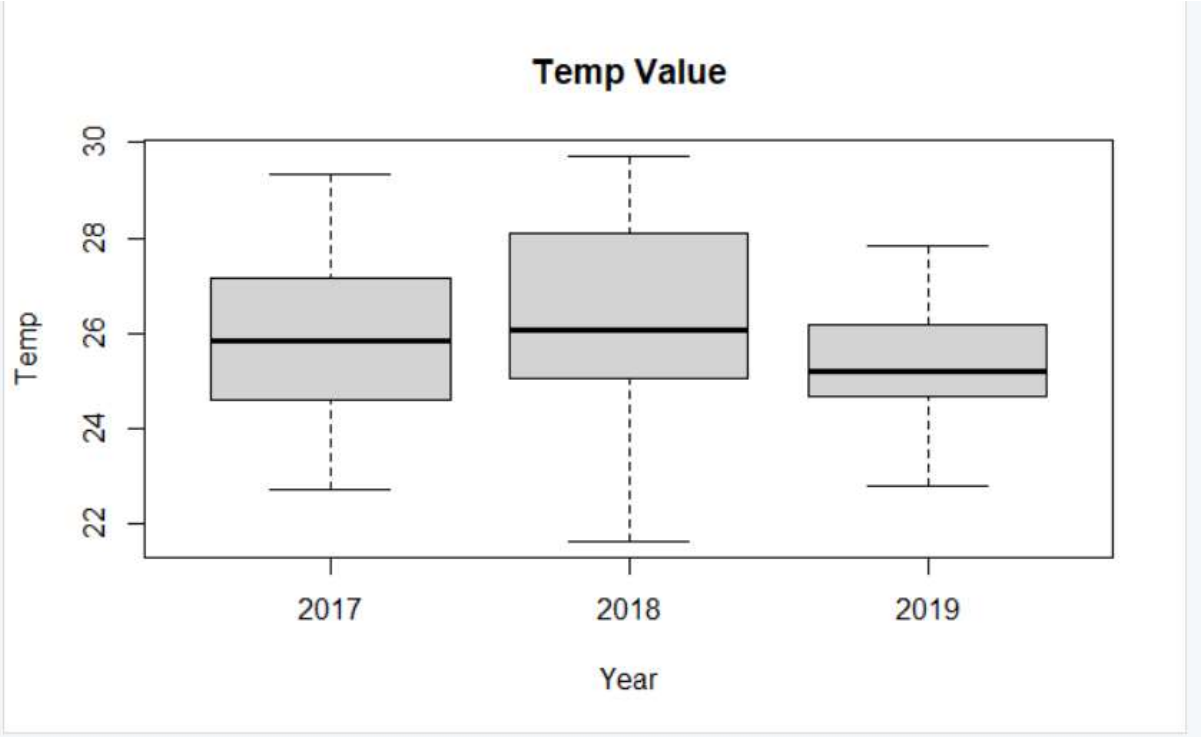
```
> wt=c(temperature$Jan.17,temperature$Jan.18,temperature$Jan.19)
> wt
 [1] 25.88889 26.33333 26.05556 25.61111 25.22222 24.11111 24.22222 24.61111 24.50000 23.16667
[11] 22.72222 23.27778 24.44444 25.50000 25.83333 25.50000 24.55556 24.83333 26.33333 26.33333
[21] 27.05556 28.16667 29.33333 29.05556 28.50000 27.05556 25.05556 28.50000 28.61111 28.11111
[31] 27.27778 25.11111 24.00000 24.94444 21.61111 25.22222 24.55556 23.33333 23.72222 25.44444
[41] 27.16667 27.94444 28.27778 28.16667 28.33333 29.00000 29.72222 28.66667 28.77778 27.27778
[51] 27.00000 27.27778 26.05556 24.33333 23.94444 25.22222 25.83333 25.22222 25.11111 26.66667
[61] 28.00000 28.33333 24.83333 25.50000 25.44444 26.27778 24.83333 24.72222 25.66667 25.22222
[71] 24.22222 25.50000 24.55556 24.00000 24.16667 26.16667 27.83333 26.94444 26.88889 27.11111
[81] 27.27778 27.22222 25.22222 25.11111 23.05556 22.77778 23.61111 24.66667 24.83333 25.05556
[91] 25.88889 25.16667       NA
```

```
> df=data.frame(club,wt)
> df
   club       wt
1  2017 25.88889
2  2017 26.33333
3  2017 26.05556
4  2017 25.61111
5  2017 25.22222
6  2017 24.11111
7  2017 24.22222
8  2017 24.61111
9  2017 24.50000
10 2017 23.16667
11 2017 22.72222
12 2017 23.27778
13 2017 24.44444
14 2017 25.50000
84 2019 25.11111
85 2019 23.05556
86 2019 22.77778
87 2019 23.61111
88 2019 24.66667
89 2019 24.83333
90 2019 25.05556
91 2019 25.88889
92 2019 25.16667
93 2019       NA
> library('dplyr')
```

Plotting the Boxplot:

```
library('dplyr')

boxplot(wt~club,data=df,xlab="Year",ylab = "Temp",main="Temp Value")
```



Applying Anova

```
av=aov(wt~club,data=df)

av
```

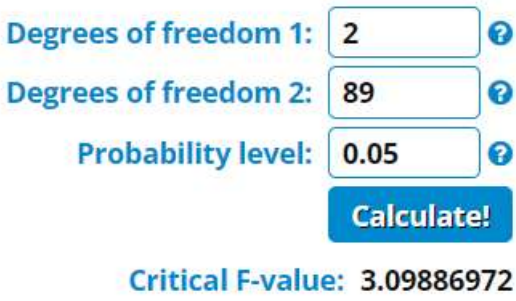
```
> av=aov(wt~club,data=df)
> av
Call:
   aov(formula = wt ~ club, data = df)

Terms:
                     club Residuals
Sum of Squares   14.23425 263.11046
Deg. of Freedom         2        89

Residual standard error: 1.719389
Estimated effects may be unbalanced
1 observation deleted due to missingness
```

Summary:

```
1 observation deleted due to missingness
> summary(av)
             Df Sum Sq Mean Sq F value Pr(>F)
club          2  14.23   7.117   2.407 0.0959 .
Residuals    89 263.11   2.956
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
1 observation deleted due to missingness
>
```



Inference:

We have performed ANOVA of the given Dataset, and we have found its Fstatistic value to be 0.885, while its Fcritical value to be 3.0988. Thus since Fstatistic < Fcritical We can accept null hypothesis and state that there exists no significant difference amongst their means.