AS404 Assignment 01

S/16/342 D.G.M.D.Dimbulkumbura

3/24/2022

```
#Call to the tidyverse package
library(tidyverse)
```

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1

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

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':
##
##     set_names

## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(dplyr)
```

#.....Q(a).....

```
#Read the dataset
crimeData=read_csv(file="E:/400 Level Sem 1/AS 404/Assignment/AS404_S16_342/crimeData.csv")
```

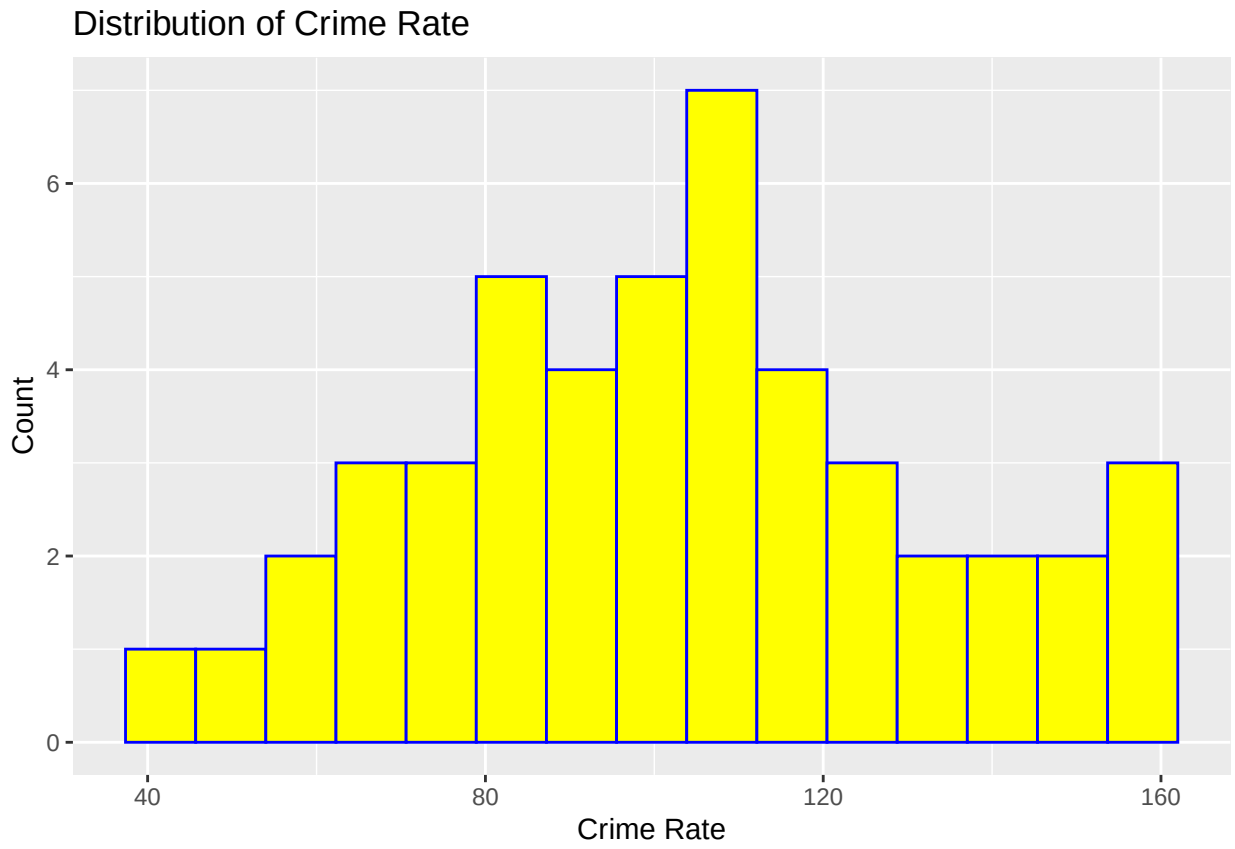
```
## Rows: 47 Columns: 14

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (14): CrimeRate, Youth, Southern, Education, ExpenditureYear0, LabourFor...
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
#Defining the column types

crimeData=read_csv(file="E:/400 Level Sem 1/AS 404/Assignment/AS404_S16_342/crimeData.csv",
                   col_types = c("dildiiiliiiiii"))


  #rename the column names in 'snake case' format

crimeData=crimeData%>%
rename(crime_rate=CrimeRate,
       youth=Youth,
       southern=Southern,
       education=Education,
       expenditure_year_0=ExpenditureYear0,
       labour_force=LabourForce,
       males=Males,
       more_males=MoreMales,
       state_size=StateSize,
       youth_unemployment=YouthUnemployment,
       mature_unemployment=MatureUnemployment,
       high_youth_unemploy=HighYouthUnemploy,
       wage=Wage,
       below_wage=BelowWage)
```

#.....Q(b).....

```
crimeData%>%
  ggplot(mapping = aes(x=crime_rate))+
  geom_histogram(bins=15,color="blue",fill="yellow")+
labs(title="Distribution of Crime Rate",x="Crime Rate",y="Count")
```



.....Q(c).....

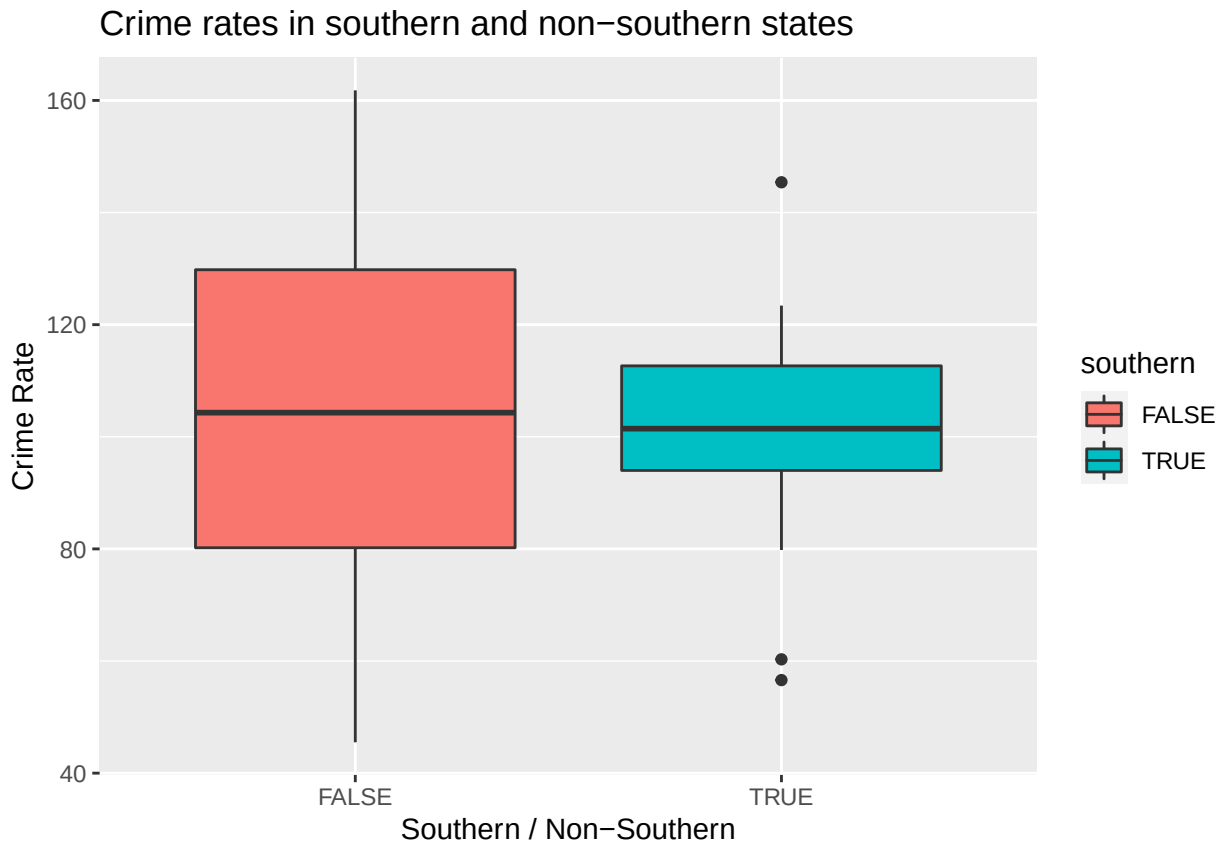
```
crimeData%>%
  group_by(southern)%>%
  summarise(avg_crime_rate=mean(crime_rate))
```

```
## # A tibble: 2 x 2
##   southern avg_crime_rate
##   <lgl>              <dbl>
## 1 FALSE               104.
## 2 TRUE                101.
```

Southern states have high average crime rate than the non- southern states

.....Q(d).....

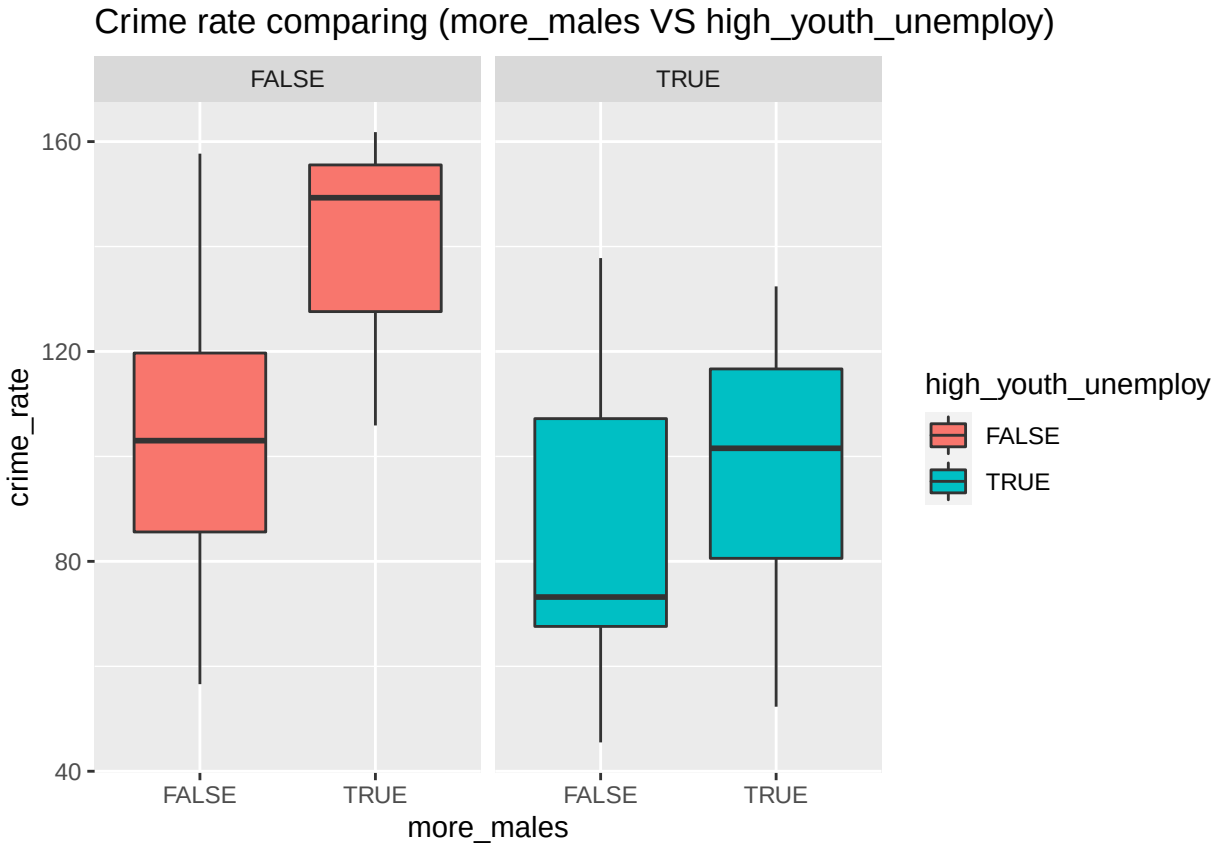
```
crimeData%>%
  ggplot()+
  geom_boxplot(aes(x=southern,y=crime_rate,fill=southern))+
labs(title="Crime rates in southern and non-southern states",x="Southern / Non-Southern",y="Crime Rate"
```



The first quartile of Non-southerns' lower than the Southerns. But median & 3rd quartile of Non-southerns' higher than the Southerns. Also Non-southerns have higher range compared to the Southerns.

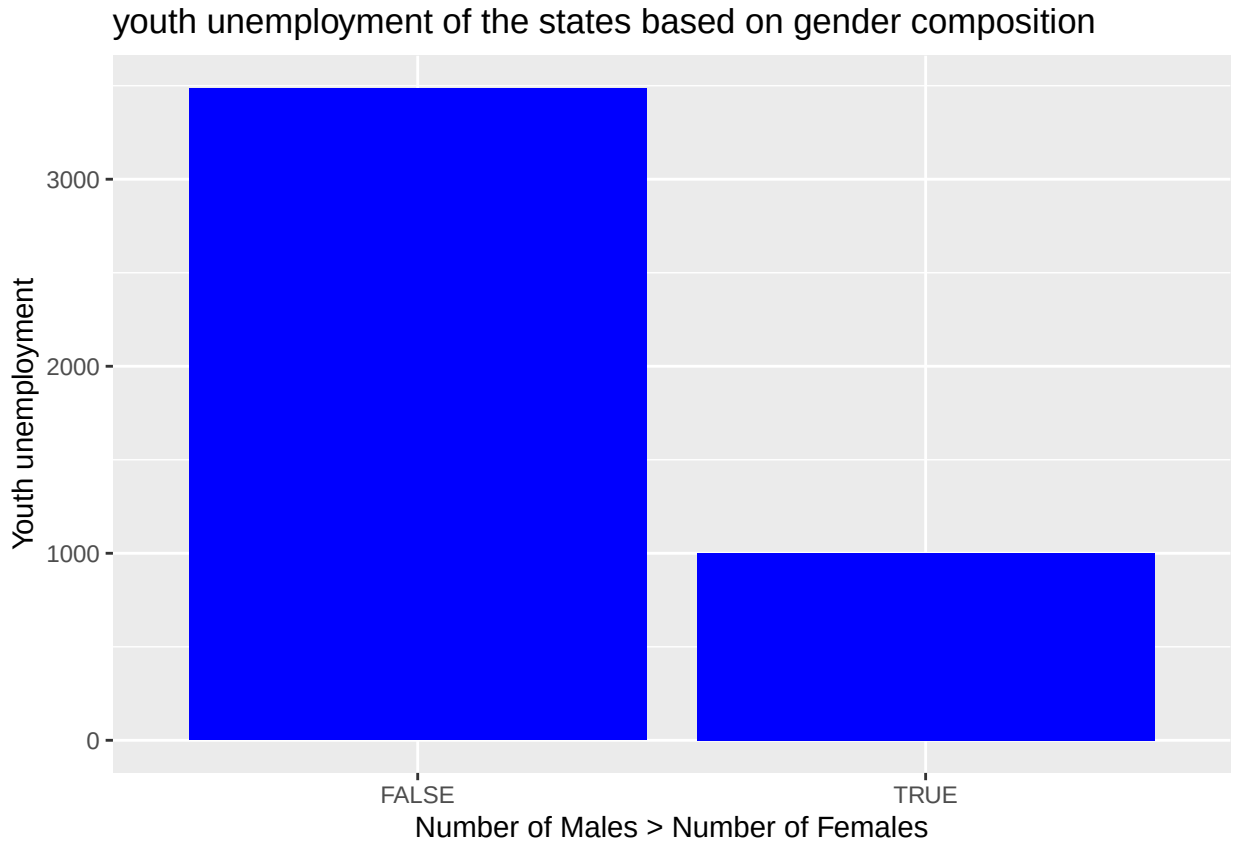
.....Q(e).....

```
ggplot(crimeData, aes(x=more_males, y=crime_rate, fill=high_youth_unemploy)) +
    geom_boxplot() +
    facet_wrap(~high_youth_unemploy)+
  ggtitle("Crime rate comparing (more_males VS high_youth_unemploy)")
```



.....Q(f).....

```
ggplot(crimeData, aes(x=more_males, y=youth_unemployment))+geom_bar(stat="identity",fill="blue")+
labs(title="youth unemployment of the states based on gender composition",x="Number of Males > Number o
```



.....Q(g).....

```
ggplot(crimeData, aes(x=wage, y=crime_rate)) +
    geom_point(color="red")+
labs(title="Relationship between the Crime rate and Median weekly wage",x="Median weekly wage",y="Crime
```

.....Q(h).....

```
linear_mod=lm(crimeData$crime_rate~crimeData$wage)
```

.....Q(i).....

```
summary(linear_mod)
```

```
##
## Call:
## lm(formula = crimeData$crime_rate ~ crimeData$wage)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.221 -19.666  -0.546  20.849  49.383
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    35.97031   21.57780   1.667  0.10246
## crimeData$wage  0.12722    0.04041   3.148  0.00291 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 26.45 on 45 degrees of freedom
## Multiple R-squared:  0.1805, Adjusted R-squared:  0.1623
## F-statistic: 9.912 on 1 and 45 DF,  p-value: 0.002914
```

Y=(0.12722)*X+(35.97031) ;Where Y=Crime Rate & X=Median weekly wage