Data Visualization Homework 3

Brandon Hosley

UIN: 676399238

Yanhui Guo, Ph.D

Data Visualization Homework 3

Problem 1

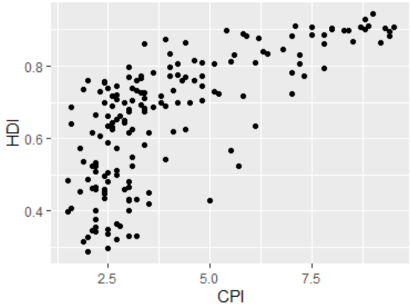
The attached CSV data file consists of Human Development Index and Corruption Perception Index scores for several countries. Load the attached CSV file to finish the following steps and put the drawing result in the file one by one: (10 points)

```
file_path <- #######/EconimistData.csv
EconData <- read.csv(file_path)
library(ggplot2)
```

(a)

Create a scatter plot with CPI on the x axis and HDI on the y axis.

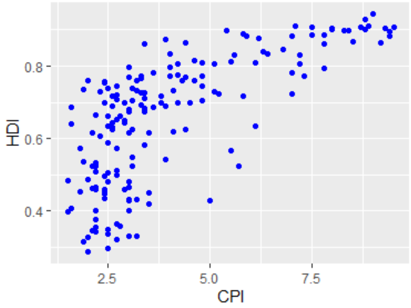
```
p <- ggplot(EconData, aes(x= CPI, y= HDI))
p + geom_point()
```



(b)

Color the points in the previous plot blue.

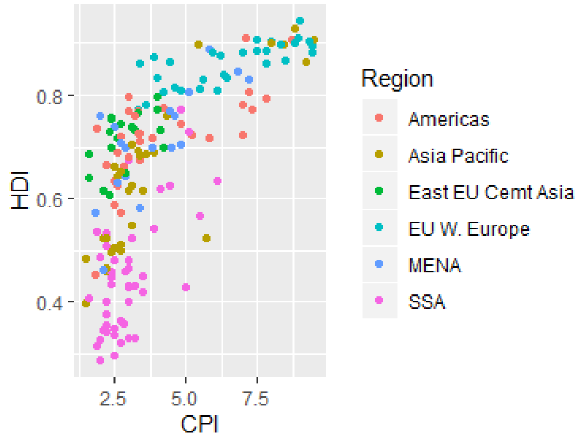
```
p + geom_point(color= "blue")
```



(c)

Color the points in the previous plot according to Region.

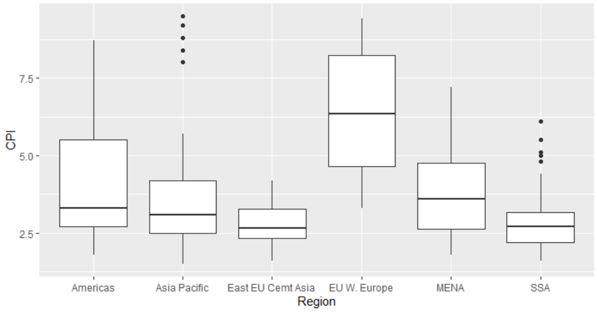
```
p + geom_point(aes(color= Region))
```



(d)

Create boxplots of CPI by Region

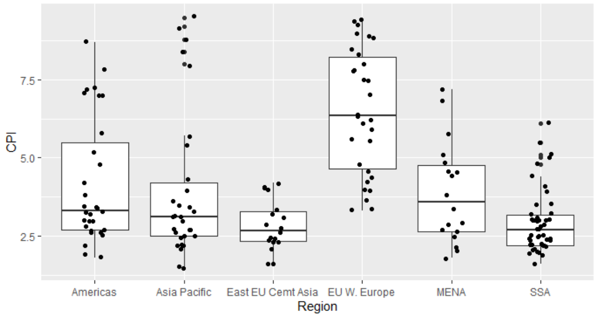
```
p + geom_boxplot()
```



(d)

Overlay points on top of the box plots

```
p + geom_boxplot() + geom_point(position= position_jitter(width= 0.1))
```



Problem 2

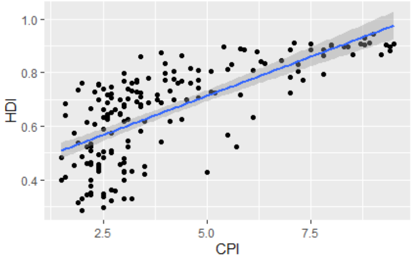
Re-create a scatter plot with CPI on the x axis and HDI on the y axis as previous question. Then finish the following steps and put the drawing result in the file one by one: (15 points)

```
p <- ggplot(EconData, aes(x= CPI, y= HDI))
```

(a)

Overlay a smoothing line on top of the scatter plot using the lm method.

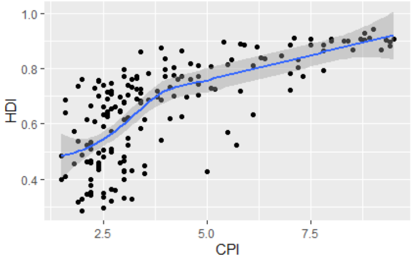
```
p + geom_point() + stat_smooth(method= lm)
```



(b)

Overlay a smoothing line on top of the scatter plot using the default method.

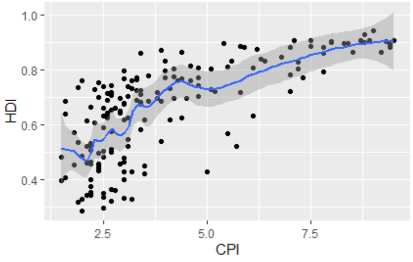
```
p + geom_point() + stat_smooth()
```



(c)

Overlay a smoothing line on top of the scatter plot using the default loess method, but make it less smooth.

```
p + geom_point() + stat_smooth(span= 0.25)
```



Problem 3

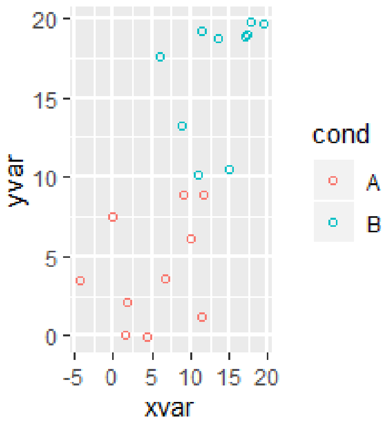
Create a data frame using the following functions:

```
set.seed(955)
dat <- data.frame(cond = rep(c("A", "B"), each=10),
        xvar = 1:20 + rnorm(20,sd=3),
        yvar = 1:20 + rnorm(20,sd=3))
```

Then use and write functions to draw four graphs, respectively. (25 points)

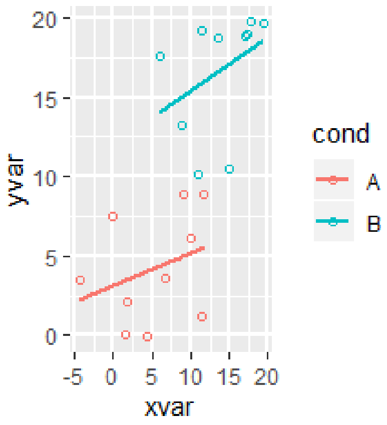
(**a**)

```
p <- ggplot(dat, aes(x= xvar, y= yvar))
p + geom_point(aes(color= cond), shape= 21)
  + theme(panel.grid.major = element_line(color= "white", size= 1),
                  panel.grid.minor = element_line(colour= "white", size=0.5))
```



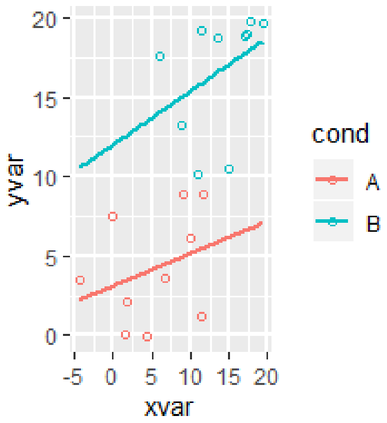
(**b**)

```
p + geom_point(aes(color= cond), shape= 21)
  + stat_smooth(aes(color= cond), method= lm, se= FALSE)
  + theme(panel.grid.major = element_line(color= "white", size= 1),
                  panel.grid.minor = element_line(colour= "white", size=0.5))
```



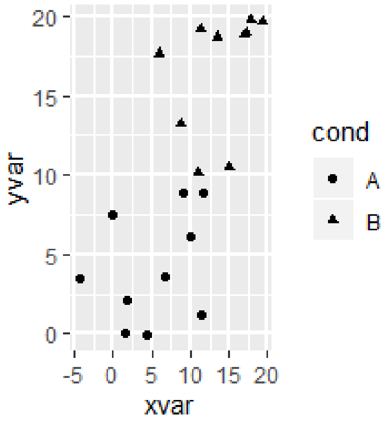
(**c**)

```
 p + geom_point(aes(color= cond), shape= 21)
   + stat_smooth(aes(color= cond), method= lm, se= FALSE, fullrange= TRUE)
   + theme(panel.grid.major = element_line(color= "white", size= 1),
                          panel.grid.minor = element_line(colour= "white", size=0.5))
```



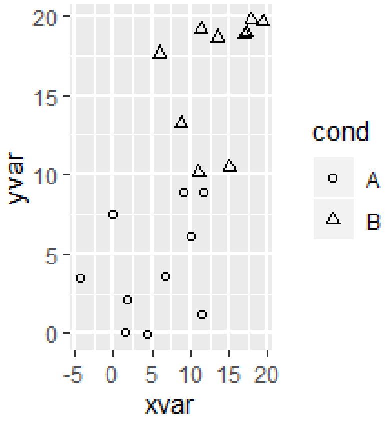
(**d**)

```
p + geom_point(aes(shape= cond))
  + theme(panel.grid.major = element_line(color= "white", size= 1),
                         panel.grid.minor = element_line(colour= "white", size=0.5))
```



(e)

```
p + geom_point(aes(shape= cond))
  + scale_shape_discrete(solid=F)
  + theme(panel.grid.major = element_line(color= "white", size= 1),
                   panel.grid.minor = element_line(colour= "white", size=0.5))
```



Problem 4

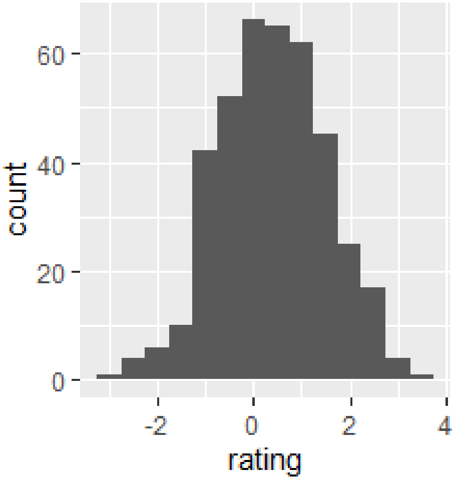
Create a data frame using the following functions:

```
set.seed(1234)
dat <- data.frame(cond = factor(rep(c("A","B"), each=200)),
                  rating = c(rnorm(200),rnorm(200, mean=.8)))
```

Then use and write functions to draw graphs, respectively. (20 points)

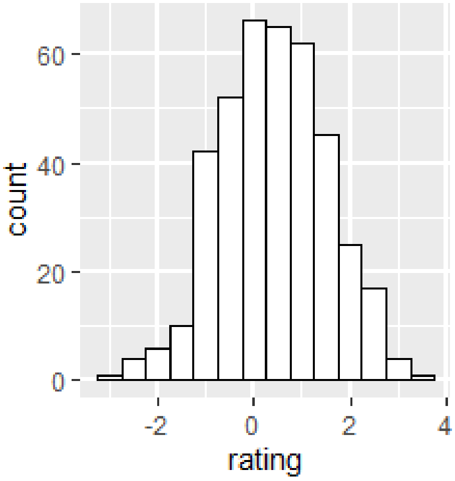
(a)

```
p <- ggplot(dat, aes(x= rating))
p + geom_histogram(binwidth= 0.5)
```



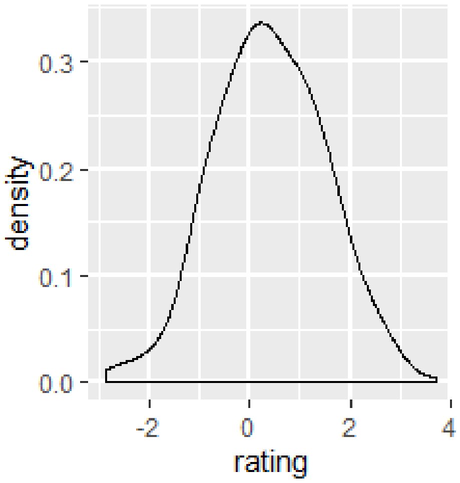
(b)

```
p + geom_histogram(binwidth= 0.5, color= "black", fill= "white")
  + theme(panel.grid.major = element_line(color= "white", size= 1),
                       panel.grid.minor = element_line(colour= "white", size=0.5))
```



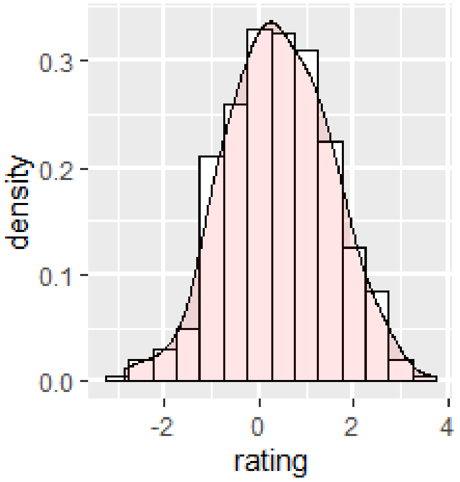
(**c**)

```
p + geom_density()
  + theme(panel.grid.major = element_line(color= "white", size= 1),
                   panel.grid.minor = element_line(colour= "white", size=0.5))
```



(**d**)

```
p + geom_histogram(aes(y=..density..),binwidth= 0.5,color="black",fill="white")
  + geom_density(alpha= 0.1, fill= "red")
  + theme(panel.grid.major = element_line(color= "white", size= 1),
                   panel.grid.minor = element_line(colour= "white", size=0.5))
```

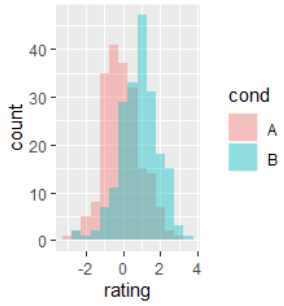


Problem 5

Following the same data frame in question 4, and use and write functions to draw the following graphs, respectively. (15 points)

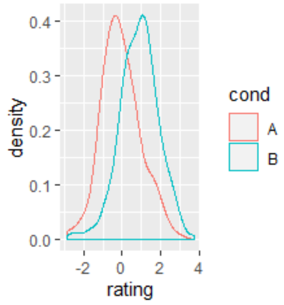
(**a**)

```
p + geom_histogram(aes(fill= cond),binwidth= 0.5,position="identity",alpha=0.4)
```



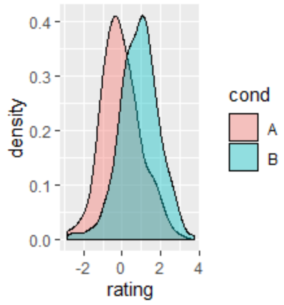
(**b**)

```
p + geom_density(aes(color= cond))
```



(**c**)

```
p + geom_density(aes(fill= cond), color= "black", alpha= 0.4)
```

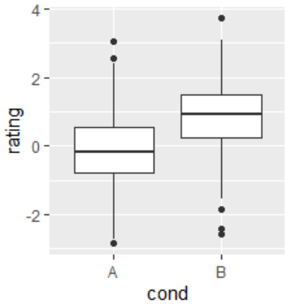


Problem 6

Following the same data frame in question 4, and use and write functions to draw the following graphs, respectively. (15 points)

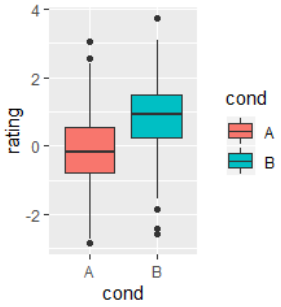
(a)

```
p <- ggplot(dat, aes(x= cond, y= rating))
p + geom_boxplot()
```



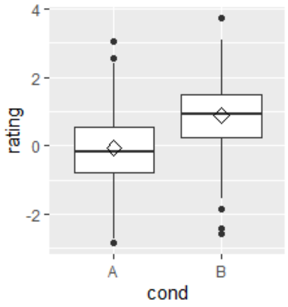
(b)

```
p + geom_boxplot(aes(fill= cond))
```



(c)

```
p + geom_boxplot()
  + stat_summary(fun.y="mean", geom="point", shape=23,size=3)
  + guides(fill=FALSE)
```


rating
A
B
cond