Assignment-Chapter 5&6

Student Name: JAYADEEP REDDY KARRI ID:664646469

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:

dat <- read.csv("EconomistData.csv")

library(ggplot2)

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

ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point()

* 1. Color the points in the previous plot blue.

ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point(color = "blue")

* 1. Color the points in the previous plot according to Region.

ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point(aes(color = Region))

* 1. Create boxplots of CPI by Region

ggplot(dat, aes(x = Region, y = CPI)) + geom\_boxplot()

* 1. Overlay points on top of the box plots

ggplot(dat, aes(x = Region, y = CPI)) + geom\_boxplot() + geom\_point()

1. 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:

library(ggplot2)

ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point()

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

ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point() + geom\_smooth(method = "lm")

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

ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point() + geom\_smooth()

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

ggplot(dat, aes(x = CPI, y = HDI)) + geom\_point() + geom\_smooth(span = .4)

1. Create a data frame using the following functions:

set.seed(955)

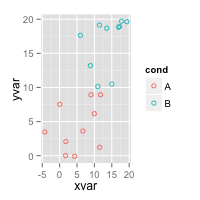
dat <- data.frame(cond = rep(c("A", "B"), each=10),

xvar = 1:20 + rnorm(20,sd=3),

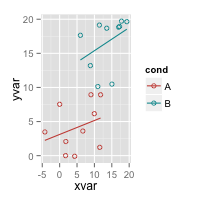
yvar = 1:20 + rnorm(20,sd=3))

Then use and write functions to draw four graphs, respectively.

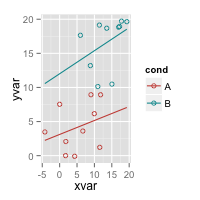
(a)



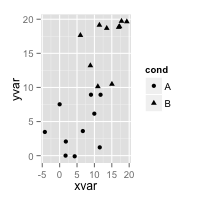
(b)



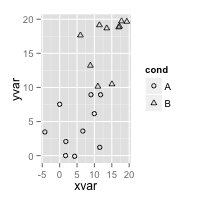
(c)



(d)



(e)



(a) # Set color by cond

ggplot(dat, aes(x=xvar, y=yvar, color=cond)) + geom\_point(shape=1)

(b) # Same, but with different colors and add regression lines

ggplot(dat, aes(x=xvar, y=yvar, color=cond)) +

geom\_point(shape=1) +

scale\_colour\_hue(l=50) + # Use a slightly darker palette than normal geom\_smooth(method=lm, # Add linear regression lines

se=FALSE) # Don't add shaded confidence region

(c) # Extend the regression lines beyond the domain of the data

ggplot(dat, aes(x=xvar, y=yvar, color=cond)) + geom\_point(shape=1) + scale\_colour\_hue(l=50) + # Use a slightly darker palette than normal geom\_smooth(method=lm, # Add linear regression lines

se=FALSE, # Don't add shaded confidence region

fullrange=TRUE) # Extend regression lines

(d) # Set shape by cond

ggplot(dat, aes(x=xvar, y=yvar, shape=cond)) + geom\_point()

(e) # Same, but with different shapes

ggplot(dat, aes(x=xvar, y=yvar, shape=cond)) + geom\_point() + scale\_shape\_manual(values=c(1,2)) # Use a hollow circle and triangle

1. Create a data frame using the following functions:

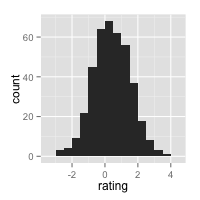
set.seed(1234)

dat <- data.frame(cond = factor(rep(c("A","B"), each=200)),

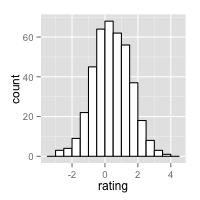
rating = c(rnorm(200),rnorm(200, mean=.8)))

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

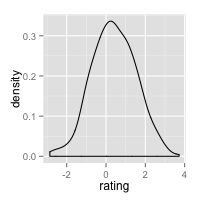
(a)



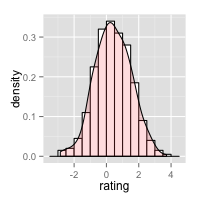
(b)



(c)



(d) For this graph, you can use any color under density curve.



1. ## Basic histogram from the vector "rating". Each bin is .5 wide.

## These both result in the same output:

ggplot(dat, aes(x=rating)) + geom\_histogram(binwidth=.5)

# qplot(dat$rating, binwidth=.5)

(b) # Draw with black outline, white fill

ggplot(dat, aes(x=rating)) +

geom\_histogram(binwidth=.5, colour="black", fill="white")

(c)# Density curve

ggplot(dat, aes(x=rating)) + geom\_density()

(d) #Histogram overlaid with kernel density curve

ggplot(dat, aes(x=rating)) +

geom\_histogram(aes(y=..density..),

#Histogram with density instead of count on y-axis

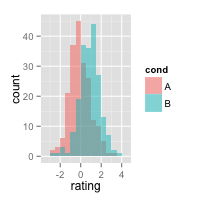
binwidth=.5,

colour="black", fill="white") +

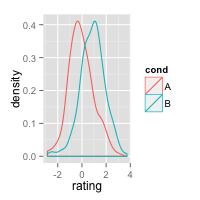
geom\_density(alpha=.2, fill="#FF6666") # Overlay with transparent density plot

1. Following the same data frame in question 4, and use and write functions to draw the following graphs, respectively.

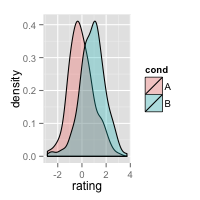
(a)



(b)



(c)



(a) #Overlaid histograms

ggplot(dat, aes(x=rating, fill=cond)) +

geom\_histogram(binwidth=.5, alpha=.5, position="identity")

(b)# Density plots

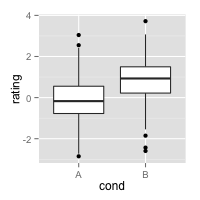
ggplot(dat, aes(x=rating, colour=cond)) + geom\_density()

(c) #Density plots with semi-transparent fill

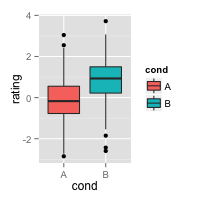
ggplot(dat, aes(x=rating, fill=cond)) + geom\_density(alpha=.3)

1. Following the same data frame in question 4, and use and write functions to draw the following graphs, respectively.

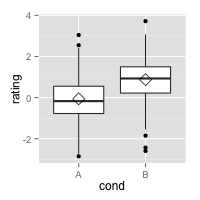
(a)



(b)



(c)



(a) #A basic box plot

ggplot(dat, aes(x=cond, y=rating)) + geom\_boxplot()

(b) #A basic box with the conditions colored

ggplot(dat, aes(x=cond, y=rating, fill=cond)) + geom\_boxplot()

(c) #Add a diamond at the mean, and make it larger

ggplot(dat, aes(x=cond, y=rating)) + geom\_boxplot() +

stat\_summary(fun.y=mean, geom="point", shape=5, size=4)