hw\_01

Abby Bergman

1/8/2019

#set wd -> didn't actually have to do this because I worked in a markdown  
setwd("/Users/AbigailBergman/Desktop/Grad School/Winter Quarter 2019/Data Science/datascience\_repo/week\_01/hw\_01")  
  
#import data  
bicycle <- read.csv("Bicycle.csv")

#Risk taking scores  
helmet <- bicycle %>%  
 filter(Condition ==1)  
  
hat <- bicycle %>%  
 filter(Condition == 2)  
  
#overall descriptive stats  
mean(bicycle$BART)

## [1] 35.6165

mean(bicycle$SSS\_total)

## [1] 20.95

#BART  
mean(helmet$BART)

## [1] 40.40308

sd(helmet$BART)

## [1] 18.17778

mean(hat$BART)

## [1] 31.06341

sd(hat$BART)

## [1] 13.29115

#independent t test (not Welch's)  
t.test(helmet$BART, hat$BART, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: helmet$BART and hat$BART  
## t = 2.6326, df = 78, p-value = 0.01021  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.276655 16.402669  
## sample estimates:  
## mean of x mean of y   
## 40.40308 31.06341

#SSS  
mean(helmet$SSS\_total)

## [1] 23.23077

sd(helmet$SSS\_total)

## [1] 6.997975

mean(hat$SSS\_total)

## [1] 18.78049

sd(hat$SSS\_total)

## [1] 5.086807

#independent t test  
t.test(helmet$SSS\_total, hat$SSS\_total)

##   
## Welch Two Sample t-test  
##   
## data: helmet$SSS\_total and hat$SSS\_total  
## t = 3.2399, df = 69.192, p-value = 0.001839  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.710146 7.190416  
## sample estimates:  
## mean of x mean of y   
## 23.23077 18.78049

#gender  
male <- bicycle %>%  
 filter(Sex ==1)  
  
female <- bicycle %>%  
 filter(Sex == 2)  
  
mean(male$BART)

## [1] 34.65882

sd(male$BART)

## [1] 15.0565

mean(female$BART)

## [1] 36.32435

sd(female$BART)

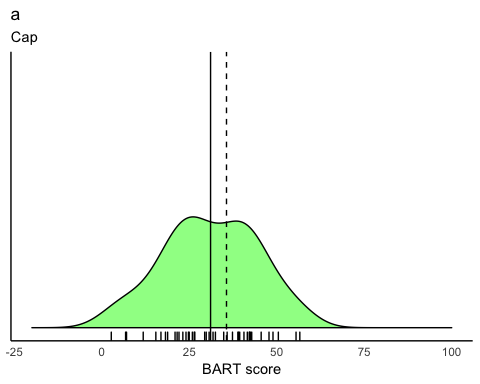
## [1] 17.53145

#independent t test (not Welch's)  
t.test(male$BART, female$BART, var.equal = TRUE)

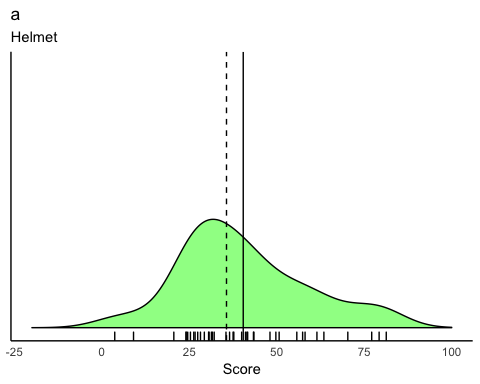
##   
## Two Sample t-test  
##   
## data: male$BART and female$BART  
## t = -0.44551, df = 78, p-value = 0.6572  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -9.108180 5.777132  
## sample estimates:  
## mean of x mean of y   
## 34.65882 36.32435

#rename Condition  
bicycle <- bicycle%>%  
 mutate(Condition = factor(Condition, levels=(c(1,2)), labels=(c("Helmet", "Cap"))))

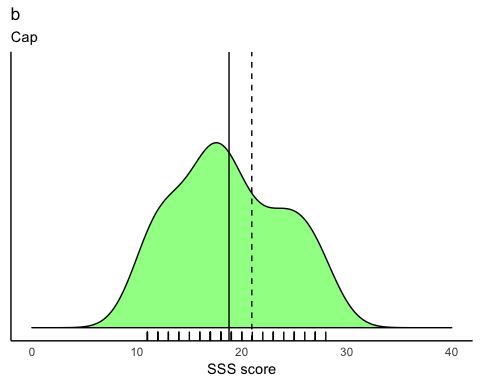
#Plot a  
#ggplot(bicycle, aes(BART))+  
 #facet\_grid(Condition~.)+  
 #geom\_density(fill="green", alpha = .5)+  
 #geom\_rug() +  
 #labs(title = "BART scores for Helmet vs Cap", x = "BART score", y = "")+  
 #geom\_vline(aes(xintercept=mean(BART)), linetype = "dashed")+  
 #ylim(c(0,.03))+xlim(c(-10,90))+  
 #theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),  
#panel.background = element\_blank(), axis.line = element\_line(colour = "black"),axis.text.y=element\_blank(),  
 #axis.ticks=element\_blank(),  
 #axis.title.y=element\_blank())  
  
ggplot(hat, aes(BART)) +  
 geom\_density(fill="green", alpha = .5)+  
 geom\_rug() +  
 labs(title = "a", subtitle = "Cap", x = "BART score", y = "")+  
 geom\_vline(aes(xintercept=mean(hat$BART)))+  
 geom\_vline(aes(xintercept=35.6165), linetype ="dashed")+  
 ylim(c(0,.06))+xlim(c(-20,100))+  
 theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),  
panel.background = element\_blank(), axis.line = element\_line(colour = "black"),axis.text.y=element\_blank(),  
 axis.ticks=element\_blank(),  
 axis.title.y=element\_blank())



ggplot(helmet, aes(BART)) +  
 geom\_density(fill="green", alpha=.5)+  
 geom\_rug()+  
 geom\_vline(aes(xintercept=mean(helmet$BART)))+  
 labs(title = "a", subtitle = "Helmet", x = "Score", y = "")+  
 geom\_vline(aes(xintercept=35.6165), linetype="dashed")+  
 ylim(c(0,.06))+xlim(c(-20,100))+  
 theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),  
panel.background = element\_blank(), axis.line = element\_line(colour = "black"),axis.text.y=element\_blank(),  
 axis.ticks=element\_blank(),  
 axis.title.y=element\_blank())



#plot b  
#ggplot(bicycle, aes(SSS\_total))+  
 # facet\_grid(Condition~.)+  
 # geom\_density(fill="green", alpha = .5)+  
 # geom\_rug() +  
 # labs(title = "SSS scores for Helmet vs Cap", x = "SSS score", y = "")+  
 # ylim(c(0,.08))+xlim(c(0,40))+  
 # geom\_vline(aes(xintercept=mean(SSS\_total)), linetype = "dashed")+  
 # theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),  
#panel.background = element\_blank(), axis.line = element\_line(colour = "black"),axis.text.y=element\_blank(),  
 # axis.ticks=element\_blank(),  
 # axis.title.y=element\_blank())  
  
ggplot(hat, aes(SSS\_total)) +  
 geom\_density(fill="green", alpha = .5)+  
 geom\_rug() +  
 labs(title ="b", subtitle= "Cap", x = "SSS score", y = "")+  
 ylim(c(0,.1))+xlim(c(0,40))+  
 geom\_vline(aes(xintercept=mean(hat$SSS\_total)))+  
 geom\_vline(aes(xintercept=20.95), linetype = "dashed")+  
 theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),  
panel.background = element\_blank(), axis.line = element\_line(colour = "black"),axis.text.y=element\_blank(),  
 axis.ticks=element\_blank(),  
 axis.title.y=element\_blank())



ggplot(helmet, aes(SSS\_total)) +  
 geom\_density(fill="green", alpha=.5)+  
 geom\_rug()+  
 labs(title ="b", subtitle = "Helmet", x = "SSS score", y = "") +  
 ylim(c(0,.1))+xlim(c(0,50))+  
 geom\_vline(aes(xintercept=mean(helmet$SSS\_total)))+  
 geom\_vline(aes(xintercept=20.95), linetype="dashed")+  
 theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),  
panel.background = element\_blank(), axis.line = element\_line(colour = "black"),axis.text.y=element\_blank(),  
 axis.ticks=element\_blank(),  
 axis.title.y=element\_blank())

