Lab8

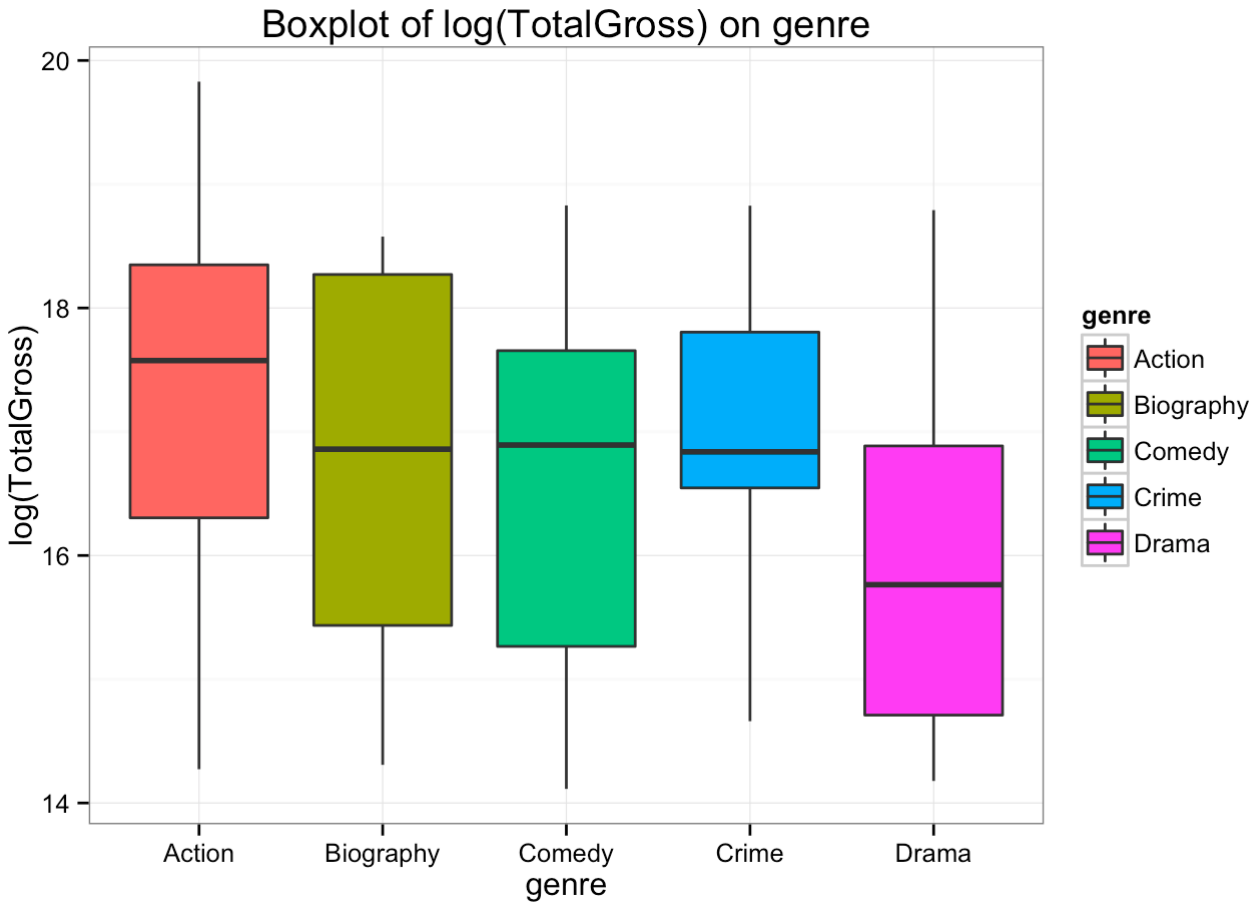
Vergil

November 25, 2015

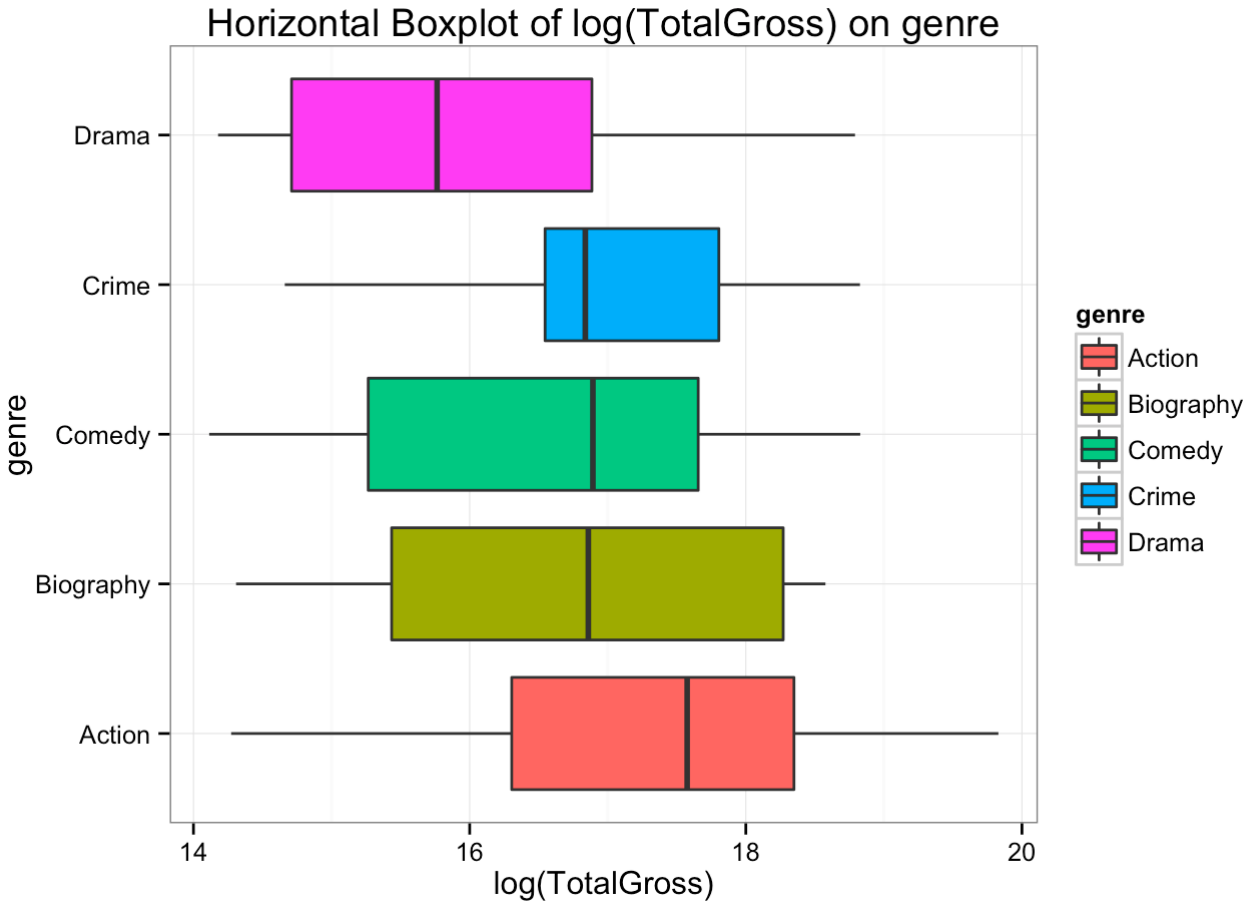
```
library(ggplot2)
library(tidyr)
library(reshape2)
```

2.

```
bom_imdb <- read.csv("~/Desktop/bom_imdb.csv", stringsAsFactors=FALSE)
bom_imdb2 <- bom_imdb[bom_imdb$genre %in% c("Action","Comedy","Drama","Biography","Crim
e"),c(4,20)]
ggplot(bom_imdb2,aes(x=genre,y=log(TotalGross),fill=genre))+geom_boxplot()+ggtitle("Boxp
lot of log(TotalGross) on genre")+theme_bw()
```

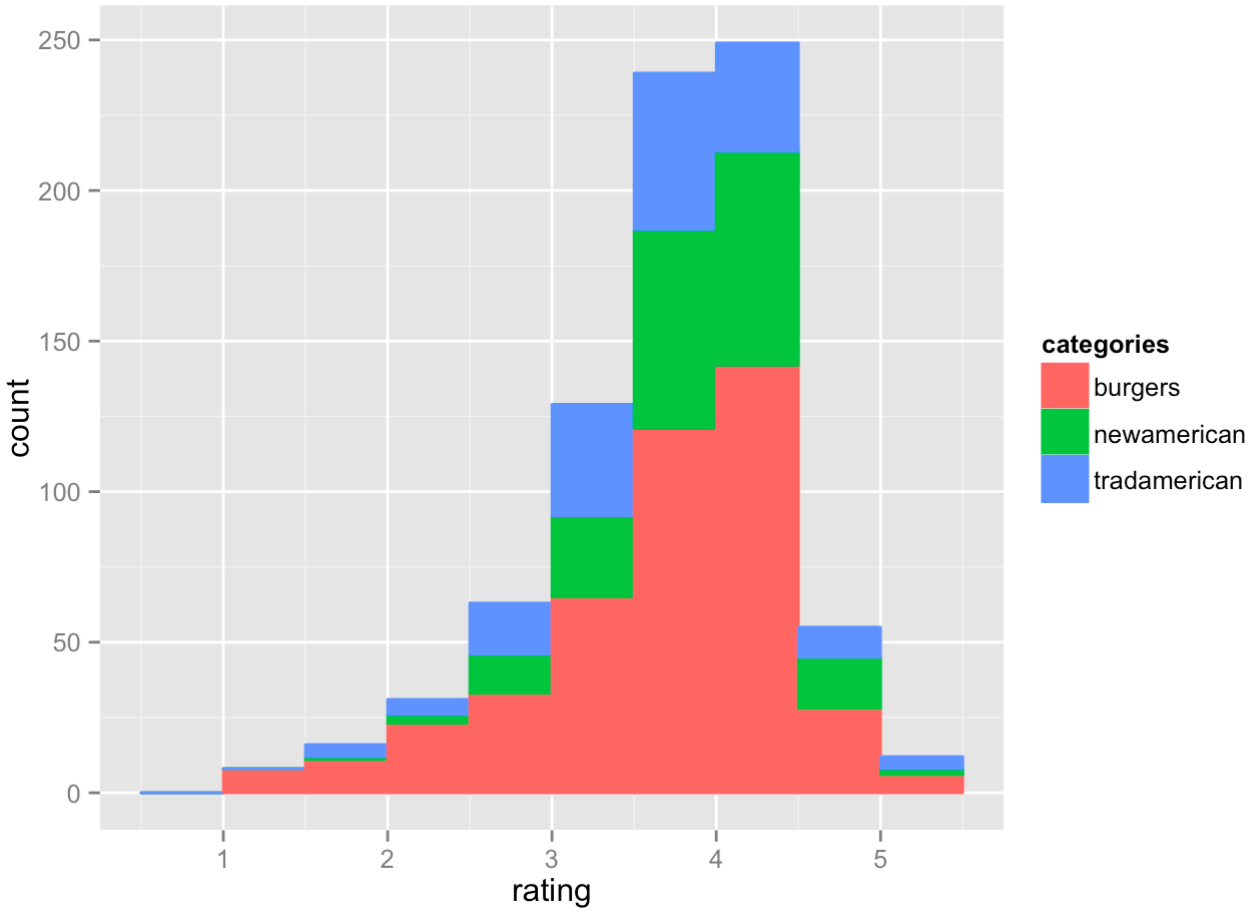



```
ggplot(bom_imdb2,aes(x=genre,y=log(TotalGross),fill=genre))+geom_boxplot()+ggtitle("Hori
zontal Boxplot of log(TotalGross) on genre")+coord_flip()+theme_bw()
```

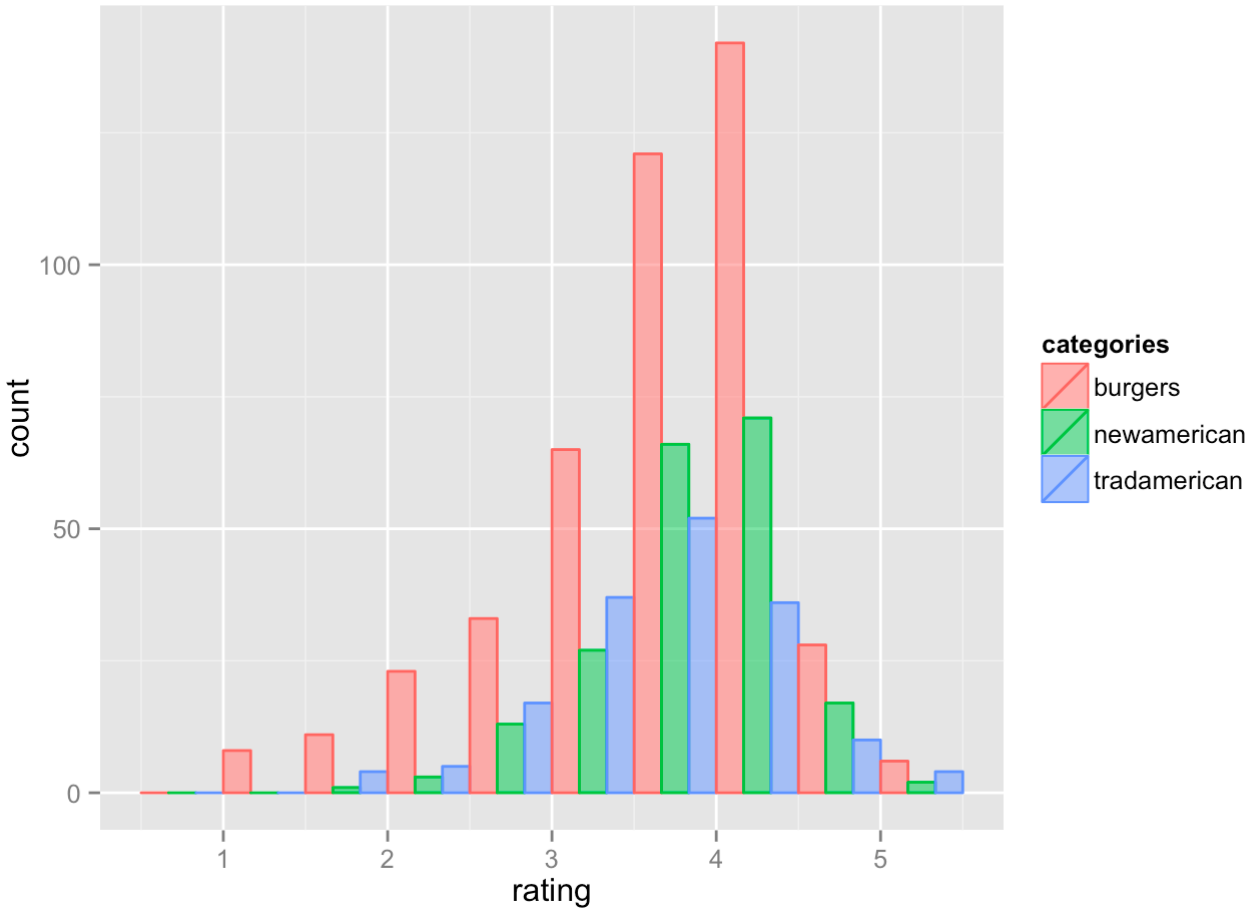


3.

```
burgers <- read.csv("~/Desktop/burgers.csv", stringsAsFactors=FALSE)
ggplot(burgers,aes(x=rating,color=categories,fill=categories))+geom_histogram(binwidth=0.5
+theme_gray()
```

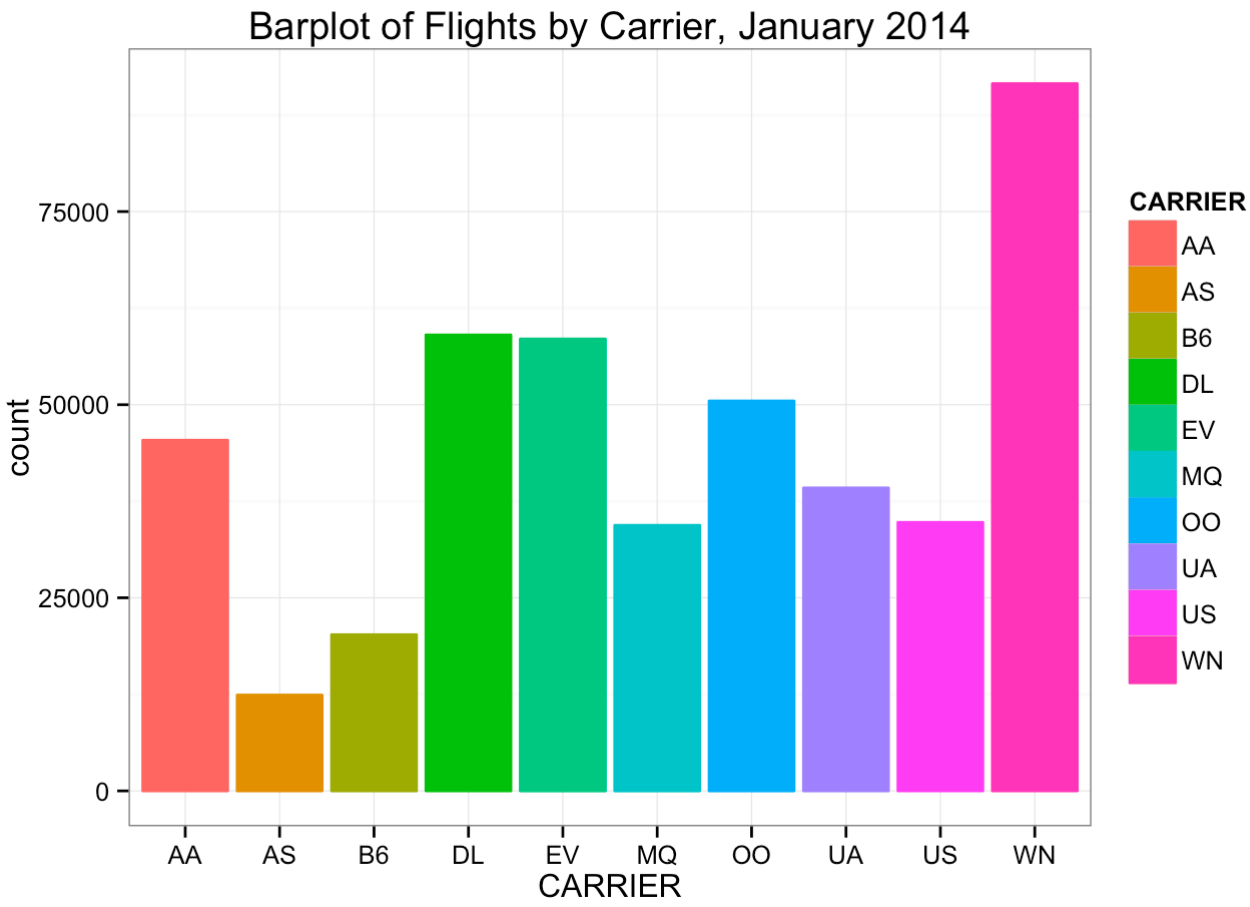



```
ggplot(burgers,aes(x=rating,color=categories,fill=categories))+geom_histogram(binwidth=0.5
position = "dodge",alpha=0.5)+theme_gray()
```

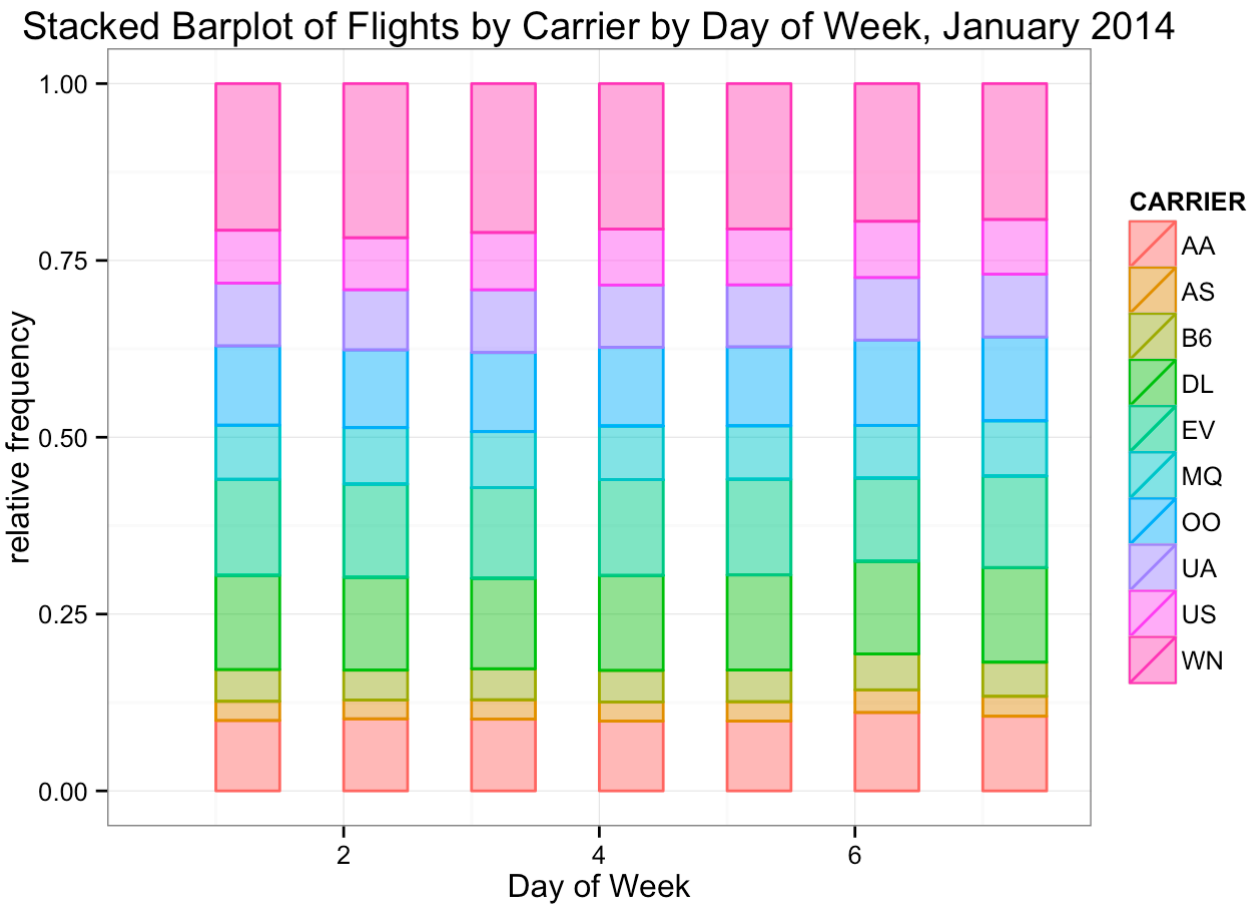


4.

```
JANFLIGHTS <- read.csv("~/Desktop/JANFLIGHTS.csv", stringsAsFactors=FALSE)
ggplot(JANFLIGHTS, aes(x=CARRIER,color=CARRIER,fill=CARRIER)) + geom_bar()+theme_bw()+gg
title("Barplot of Flights by Carrier, January 2014")
```

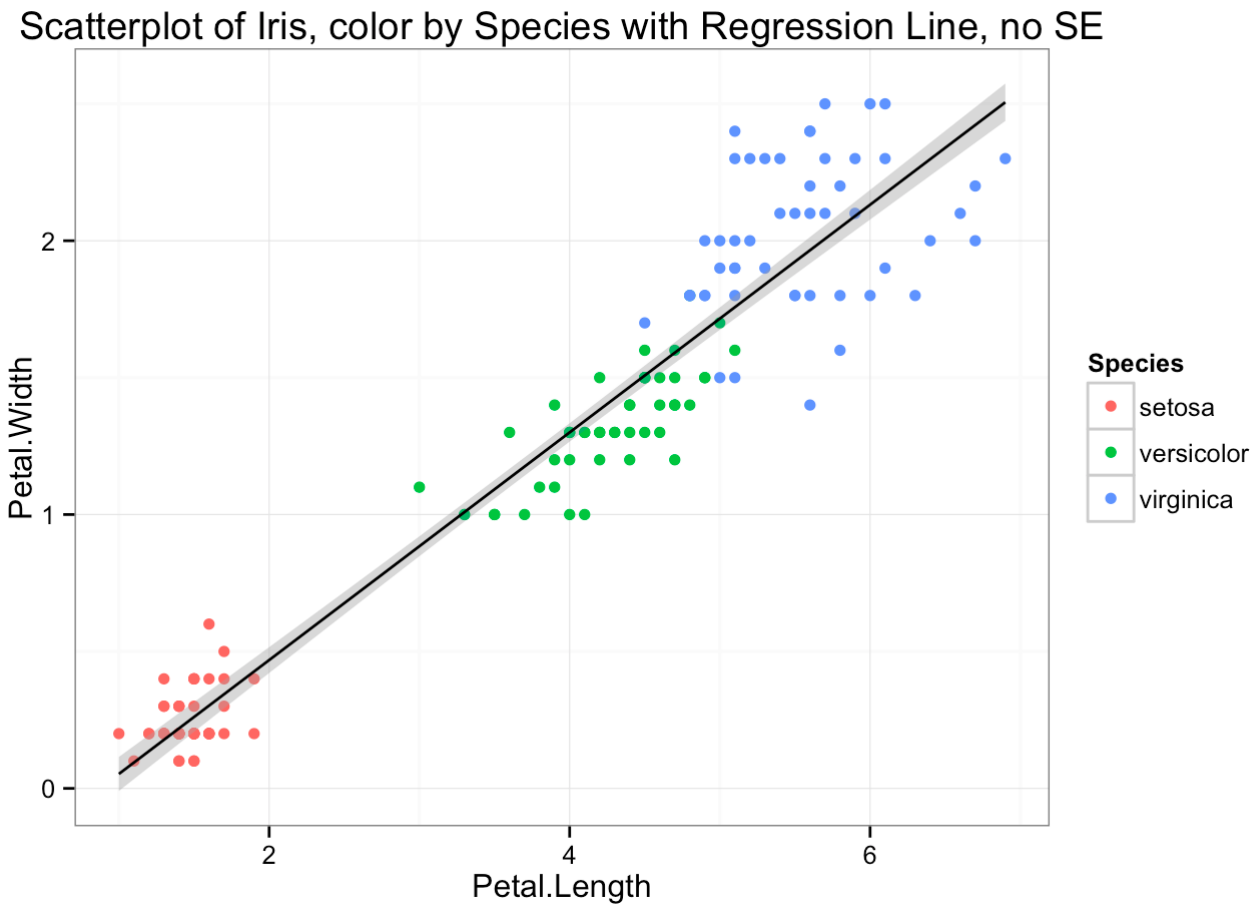



```
ggplot(JANFLIGHTS,aes(x=DAY_OF_WEEK,color=CARRIER,fill=CARRIER))+geom_histogram(position="
ill",alpha=0.5,binwidth=0.5)+theme_bw()+labs(y="relative frequency")+ggtitle("Stacked Ba
rplot of Flights by Carrier by Day of Week, January 2014")+xlab("Day of Week")+ylab("rel
ative frequency")
```

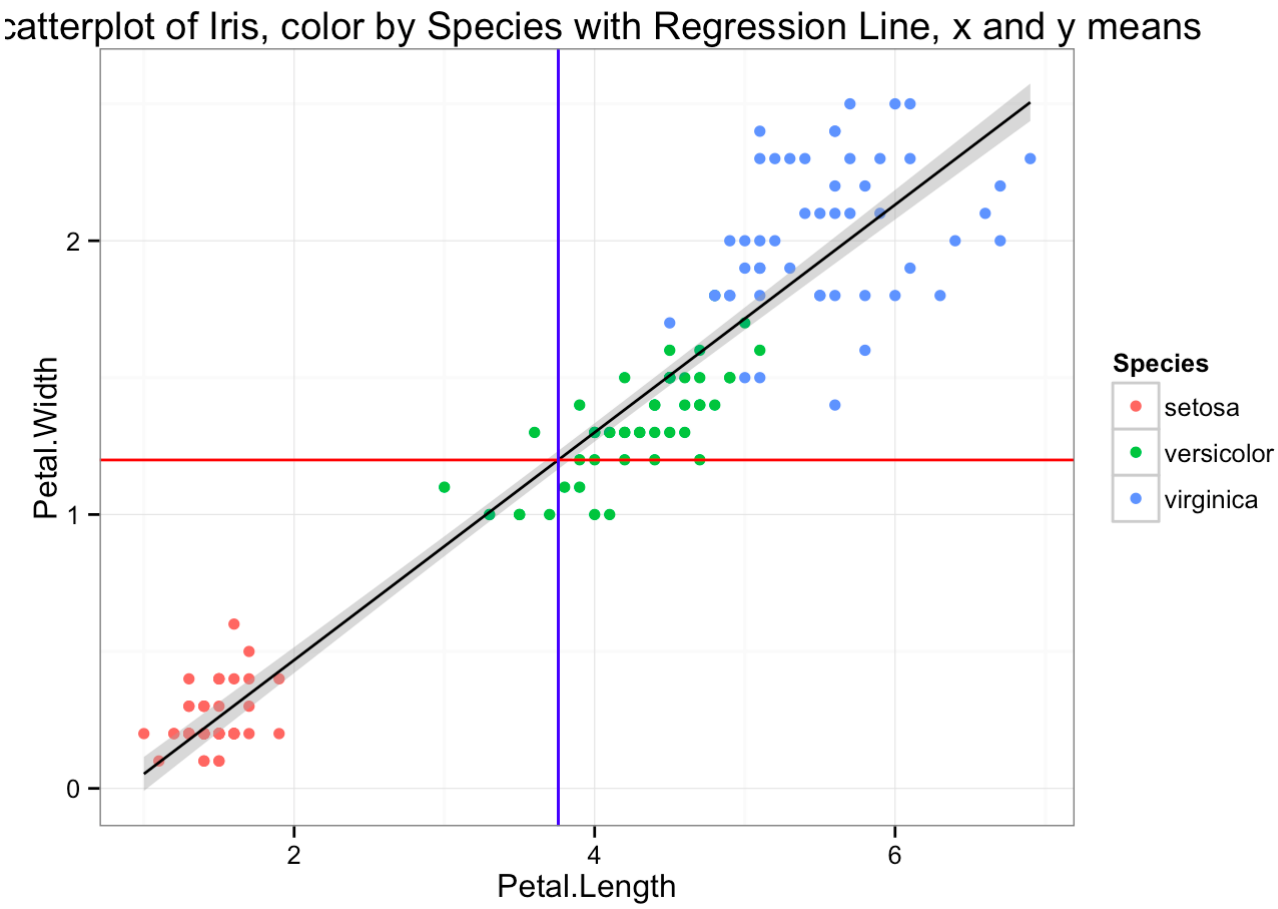


5.

```
ggplot(iris,aes(x=Petal.Length,y=Petal.Width))+geom_point(aes(color=Species))+labs(title="
catterplot of Iris, color by Species with Regression Line, no SE")+geom_smooth(method="l
m",colour="black",formula=y~x)+theme_bw()
```




```
ggplot(iris,aes(x=Petal.Length,y=Petal.Width))+geom_point(aes(color=Species))+labs(title="
catterplot of Iris, color by Species with Regression Line, x and y means")+geom_smooth(m
ethod="lm",colour="black",formula=y~x)+theme_bw()+geom_hline(yintercept=mean(iris$Petal.
Width),color="red")+geom_vline(xintercept=mean(iris$Petal.Length),color="blue")
```



6.

A.

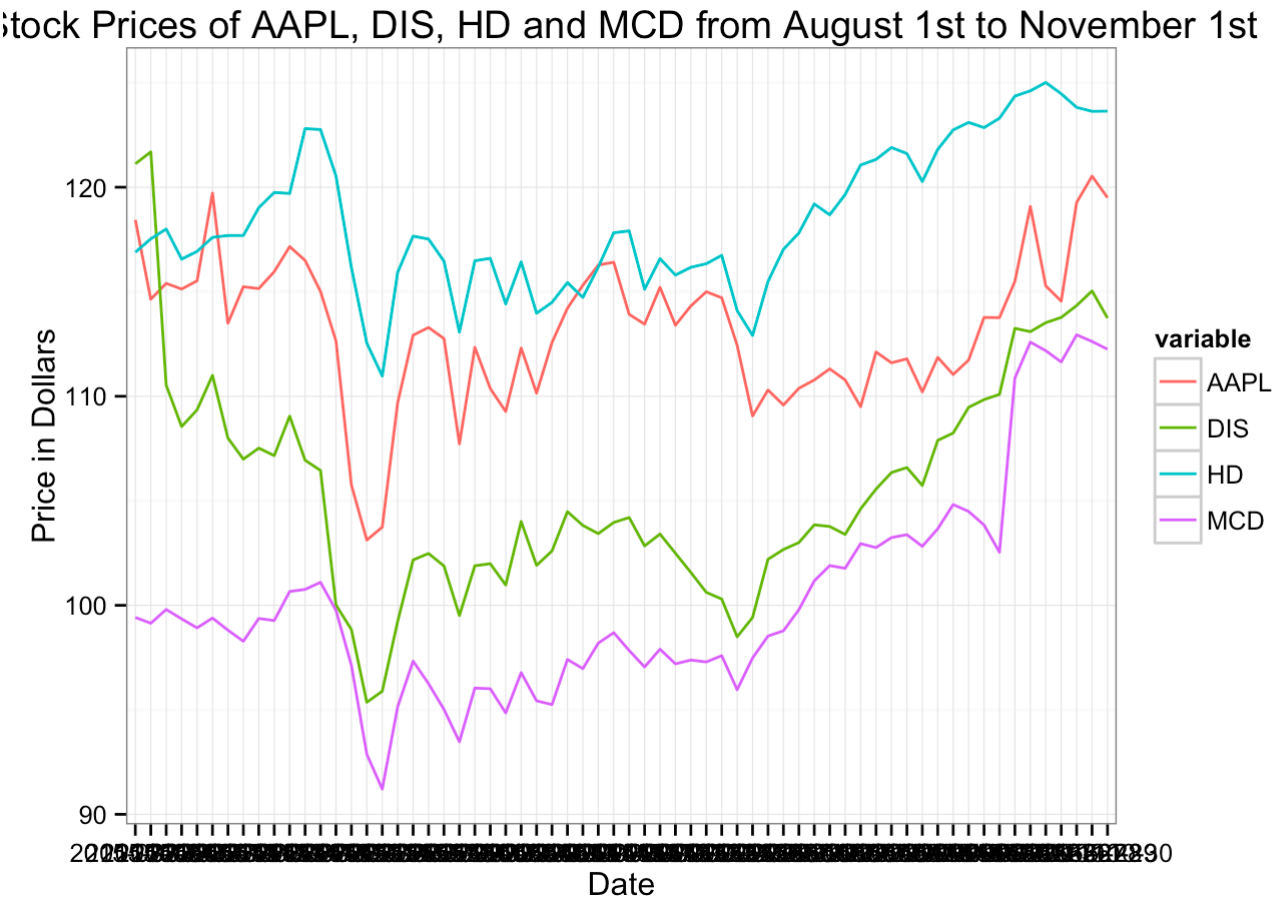
```
STOCKS4 <- read.csv("~/Desktop/STOCKS4.csv", stringsAsFactors=FALSE)
head(STOCKS4)
```

```
##     AAPL    DIS     HD    MCD      DATE1
## 1 119.50 113.74 123.64 112.25 2015-10-30
## 2 120.53 115.04 123.63 112.62 2015-10-29
## 3 119.27 114.34 123.82 112.94 2015-10-28
## 4 114.55 113.77 124.47 111.64 2015-10-27
## 5 115.28 113.52 125.01 112.18 2015-10-26
## 6 119.08 113.09 124.61 112.59 2015-10-23
```

```
STOCKS4_new <- gather(STOCKS4,key=variable,value=value,AAPL,DIS,HD,MCD)
head(STOCKS4_new)
```

```
##        DATE1 variable  value
## 1 2015-10-30     AAPL 119.50
## 2 2015-10-29     AAPL 120.53
## 3 2015-10-28     AAPL 119.27
## 4 2015-10-27     AAPL 114.55
## 5 2015-10-26     AAPL 115.28
## 6 2015-10-23     AAPL 119.08
```

```
ggplot(STOCKS4_new,aes(x=DATE1,y=value,group=variable))+geom_line(aes(col=variable))+the
me_bw()+xlab("Date")+ylab("Price in Dollars")+ggtitle("Stock Prices of AAPL, DIS, HD and
 MCD from August 1st to November 1st")
```



B.

```
load("~/Desktop/repPolls.rdata")
head(repPolls)
```

```
##         Poll Donald Trump Ben Carson Marco Rubio Ted Cruz Jeb Bush
## 1 2015-11-04          23%        24%         12%       8%       8%
## 2 2015-11-03          26%        23%         11%      11%       4%
## 3 2015-11-02          24%        23%         14%      13%       4%
## 4 2015-10-29          23%        29%         11%      10%       8%
## 5 2015-10-30          28%        23%         11%       6%       6%
## 6 2015-10-25          22%        26%          8%       4%       7%
##   John Kasich Rand Paul Carly Fiorina Mike Huckabee Chris Christie
## 1          4%        5%            3%            3%             2%
## 2          4%        4%            3%            4%             2%
## 3          3%        2%            3%            1%             3%
## 4          3%        2%            3%            3%             3%
## 5          1%        2%            3%            1%             1%
## 6          4%        4%            7%            4%             1%
##   Unsure or Other
## 1              5%
## 2              7%
## 3              9%
## 4              5%
## 5             15%
## 6              9%
```

```
names(repPolls) <- c("Poll","Donald","Ben","Marco","Ted","Jeb","John","Rand","Carly","Mi
ke","Chris","Unsure")
repPolls_new <- gather(repPolls,key=variable,value=value,Donald,Ben,Marco,Ted,Jeb,John,R
and,Carly,Mike,Chris,Unsure)
repPolls_new$value <- sapply(repPolls_new$value,function(x) gsub("%","",x))
repPolls_new$value <- as.numeric(repPolls_new$value)
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Donald","Donald
 Trump",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Ben","Ben Carso
n",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Marco","Marco Ru
bio",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Ted","Ted Cruz",x)
)
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Jeb","Jeb Bush",x)
)
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("John ","John Kas
ich",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Rand ","Rand Pau
l",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Carly ","Carly F
iorina",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Mike ","Mike Huc
kabee",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Chris ","Chris C
hristie",x))
repPolls_new$variable <- sapply(repPolls_new$variable,function(x) gsub("Unsure","Unsure
 or Other",x))
head(repPolls_new)
```

```
##         Poll     variable value
## 1 2015-11-04 Donald Trump    23
## 2 2015-11-03 Donald Trump    26
## 3 2015-11-02 Donald Trump    24
## 4 2015-10-29 Donald Trump    23
## 5 2015-10-30 Donald Trump    28
## 6 2015-10-25 Donald Trump    22
```

```
ggplot(repPolls_new,aes(x=Poll,y=value,group=variable))+geom_line(aes(col=variable))+theme_bw()+xlab("Poll")+ylab("Popularity")+ggtitle("Republican Poll from October to November")
```