```
### Wallace O'Rear Homework 2
### Chapter 3: 1, 3, 4, 5, 8
library(readr)
library(ggplot2)
library(car)
library(MASS)
library(alr3)
setwd("C:/code/Math550")

##1
airfares <- read.csv("Data/airfares.txt", sep = "")
airfarePlot <- ggplot(airfares,aes(Distance,Fare, label=City)) + geom_point() + geom_text(aes(label=City),hjust=0,
vjust=0)
airfarePlot
plot(airfares$Distance,airfares$Fare)
abline(lsfit(airfares$Distance,airfares$Fare))


airfares.lm <- lm(Fare ~ Distance, data=airfares)
summary(airfares.lm)
par(mfrow=c(2,2))
plot(airfares.lm)

tail(airfares)


airfares.lm.sq <- lm(Fare ~ sqrt(Distance), data=airfares)




##3
### part A
par(mfrow=c(1,1))
adrevenue <- read.csv("Data/AdRevenue.csv")
plot(adrevenue$Circulation,adrevenue$AdRevenue)
## build a non-transformed lm
ad.lm <- lm(AdRevenue ~ Circulation, data=adrevenue)
summary(ad.lm)
par(mfrow=c(2,2))
plot(ad.lm)

### let's try a log transformation
par(mfrow=c(1,1))
plot(log(adrevenue$Circulation),log(adrevenue$AdRevenue))
ad.loglm <- lm(log(AdRevenue) ~ log(Circulation), data=adrevenue)
summary(ad.loglm)
par(mfrow=c(2,2))
plot(ad.loglm)


#that looks better
exp(predict(ad.loglm, data.frame(Circulation = c(0.5,20)), interval = "prediction"))

### part B
par(mfrow=c(1,1))
ad.poly <- lm(AdRevenue ~ poly(Circulation,3), data=adrevenue)
summary(ad.poly) #.9333 for Rsquared, thats more than the log model
par(mfrow=c(2,2))
plot(ad.poly)
predict(ad.poly, data.frame(Circulation = c(0.5,20)), interval = "prediction")


### part C
#I would probably choose the polynomial model, the rsquared and the prediction intervals look more right

#8
par(mfrow=c(1,1))
diamonds <- read.csv("Data/diamonds.txt",sep="")
plot(diamonds)
abline(lsfit(x=diamonds$Size,y=diamonds$Price))

diamond.lm <- lm(Price ~ Size, data=diamonds)
summary(diamond.lm)
par(mfrow=c(2,2))
plot(diamond.lm)

par(mfrow=c(1,1))
inverseResponsePlot(diamond.lm)
boxcox(diamond.lm,lambda=seq(-2,2,length=20))

summary(tranxy <- powerTransform(diamonds$Price ~ diamonds$Size, diamonds))

#both the boxcox and power transform say the lambda should be close to 1, so I'm confused

#I'll try a log just for fun
```

```
diamond.log.lm <- lm(log(Price) ~ log(Size), data=diamonds)
summary(diamond.log.lm)
par(mfrow=c(2,2))
plot(diamond.log.lm)
```

```
diamond.log.lm <- lm(log(Price) ~ log(Size), data=diamonds)
summary(diamond.log.lm)
par(mfrow=c(2,2))
plot(diamond.log.lm)
```