# Assignment 9

### 1

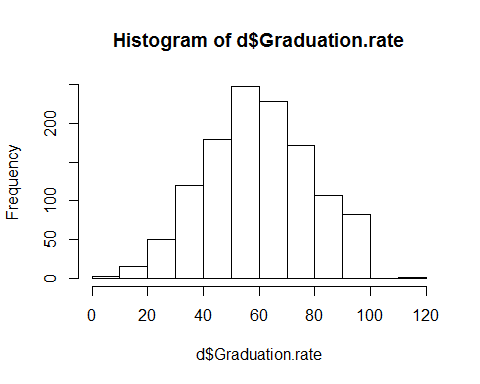
#1a.  
d <- read.csv('usnews\_clean.csv')  
head(d)

## ID.Number College.Name State Public.private  
## 1 1061 Alaska Pacific University AK 2  
## 2 1063 University of Alaska at Fairbanks AK 1  
## 3 1065 University of Alaska Southeast AK 1  
## 4 11462 University of Alaska at Anchorage AK 1  
## 5 1002 Alabama Agri. & Mech. Univ. AL 1  
## 6 1003 Faulkner University AL 2  
## Avg.Math.SAT Avg.Verbal.SAT Avg.combined.SAT Avg.ACT  
## 1 490 482 972 20  
## 2 499 462 961 22  
## 3 NA NA NA NA  
## 4 459 422 881 20  
## 5 NA NA NA 17  
## 6 NA NA NA 20  
## First.quartile.Math.SAT Third.quartile.Math.SAT  
## 1 440 530  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA  
## First.quartile.Verbal.SAT Third.quartile.Verbal.SAT First.quartile.ACT  
## 1 430 550 18  
## 2 NA NA NA  
## 3 NA NA NA  
## 4 NA NA NA  
## 5 NA NA 14  
## 6 NA NA NA  
## Third.quartile.ACT Num.applications.received Num.applicants.accepted  
## 1 22 193 146  
## 2 NA 1852 1427  
## 3 NA 146 117  
## 4 NA 2065 1598  
## 5 17 2817 1920  
## 6 NA 345 320  
## Num.students.enrolled Pct.new.students.from.top.10pct.of.HS.class  
## 1 55 16  
## 2 928 NA  
## 3 89 4  
## 4 1162 NA  
## 5 984 NA  
## 6 179 NA  
## Pct.new.students.from.top.25pct.of.HS.class Num.full.time.undergraduates  
## 1 44 249  
## 2 NA 3885  
## 3 24 492  
## 4 NA 6209  
## 5 NA 3958  
## 6 27 1367  
## Num.part.time.undergraduates In.state.tuition Out.of.state.tuition  
## 1 869 7560 7560  
## 2 4519 1742 5226  
## 3 1849 1742 5226  
## 4 10537 1742 5226  
## 5 305 1700 3400  
## 6 578 5600 5600  
## Room.and.Board.costs Room.costs Board.costs Additional.fees  
## 1 4120 1620 2500 130  
## 2 3590 1800 1790 155  
## 3 4764 2514 2250 34  
## 4 5120 2600 2520 114  
## 5 2550 1108 1442 155  
## 6 3250 1550 1700 300  
## Estimated.book.costs Estimated.personal.spending  
## 1 800 1500  
## 2 650 2304  
## 3 500 1162  
## 4 580 1260  
## 5 500 850  
## 6 350 NA  
## Pct.of.faculty.with.PhDs Pct.of.faculty.with.terminal.degree  
## 1 76 72  
## 2 67 NA  
## 3 39 51  
## 4 48 NA  
## 5 53 53  
## 6 52 56  
## Student.faculty.ratio Pct.alumni.who.donate  
## 1 11.9 2  
## 2 10.0 8  
## 3 9.5 NA  
## 4 13.7 6  
## 5 14.3 NA  
## 6 32.8 NA  
## Instructional.expenditure.per.student Graduation.rate Public.private.cat  
## 1 10922 15 Private  
## 2 11935 NA Public  
## 3 9584 39 Public  
## 4 8046 NA Public  
## 5 7043 40 Public  
## 6 3971 55 Private  
## IQR.math IQR.verbal  
## 1 90 120  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA

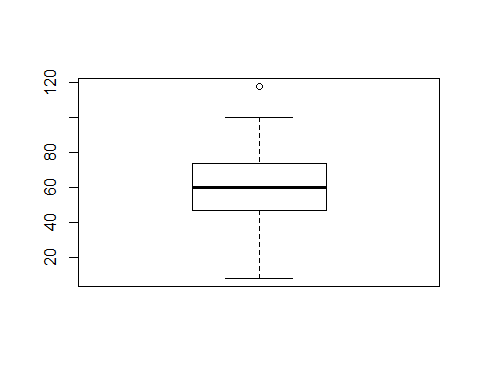
#1b  
summary(d$Graduation.rate)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 8.00 47.00 60.00 60.41 74.00 118.00 98

hist(d$Graduation.rate)



boxplot(d$Graduation.rate)



d$College.Name[which(d$Graduation.rate > 100)]

## [1] Cazenovia College  
## 1274 Levels: Abilene Christian University ... Youngstown State University

d$Graduation.rate[which(d$Graduation.rate > 100)] = NA  
  
#1c.  
m <- aggregate(d$Pct.alumni.who.donate, by=list(d$Public.private.cat), FUN=mean, na.rm=TRUE)  
names(m) = c('Type', 'Mean')  
m

## Type Mean  
## 1 Private 24.58287  
## 2 Public 13.44944

#1d.  
t.test(d$Pct.alumni.who.donate ~ d$Public.private.cat)

##   
## Welch Two Sample t-test  
##   
## data: d$Pct.alumni.who.donate by d$Public.private.cat  
## t = 17.317, df = 1018.9, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 9.871826 12.395044  
## sample estimates:  
## mean in group Private mean in group Public   
## 24.58287 13.44944

#1e.  
write.csv(d, file='usnews\_r.csv', quote=FALSE, row.names=FALSE)

### 1b.

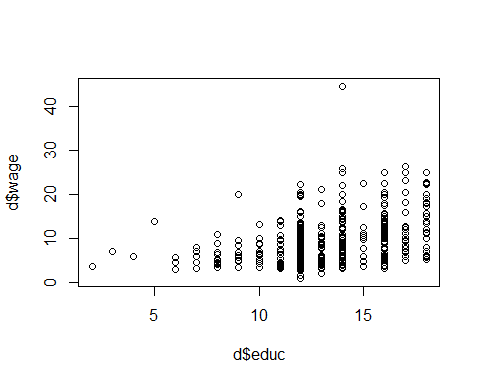
The graduation rate for Cazenovia College was over 100%. This entry was converted to NA.

### 1d.

We have enough evidence to reject the null hypothesis. In other words, it appears the means are statistically different. As such, it appears that whether one graduated from a public vs. private university is associated with whether one donates or not.

# 2

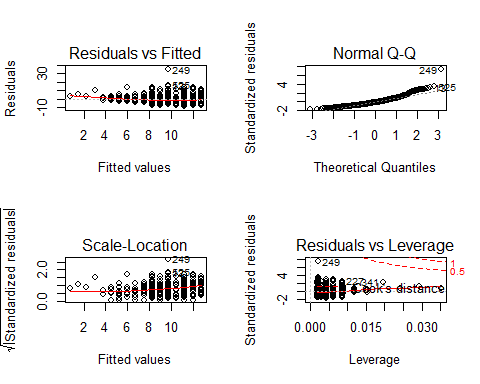
#2a.  
d = read.csv('cps.csv')  
plot(d$wage ~ d$educ)



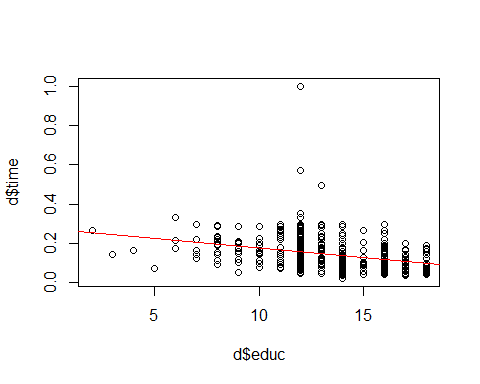
#2b.  
model = lm(d$wage ~ d$educ)  
model

##   
## Call:  
## lm(formula = d$wage ~ d$educ)  
##   
## Coefficients:  
## (Intercept) d$educ   
## -0.7460 0.7505

par(mfrow=c(2,2))  
plot(model)



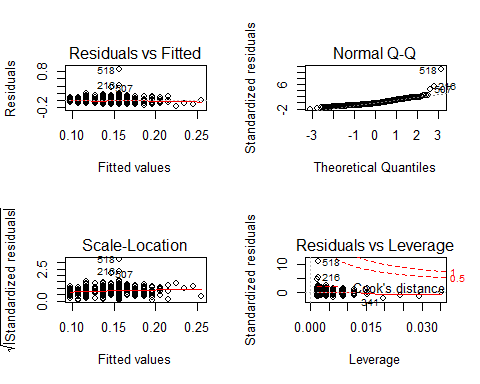
#2c.  
d$time <- 1 / d$wage  
  
#2d.  
par(mfrow=c(1,1))  
plot(d$time ~ d$educ)  
  
#2e.  
model = lm(d$time ~ d$educ)  
abline(model, col='red')



summary(model)

##   
## Call:  
## lm(formula = d$time ~ d$educ)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.15393 -0.05180 -0.02021 0.04361 0.84371   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.274700 0.017251 15.924 < 2e-16 \*\*\*  
## d$educ -0.009867 0.001299 -7.595 1.39e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.07845 on 532 degrees of freedom  
## Multiple R-squared: 0.09782, Adjusted R-squared: 0.09613   
## F-statistic: 57.68 on 1 and 532 DF, p-value: 1.393e-13

#2f.  
par(mfrow=c(2,2))  
plot(model)



d$time[c(216,507, 518)] <- NA  
model = lm(d$time ~ d$educ, na.action=na.exclude)  
summary(model)

##   
## Call:  
## lm(formula = d$time ~ d$educ, na.action = na.exclude)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.14805 -0.04856 -0.01807 0.04557 0.19798   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.267039 0.014387 18.561 <2e-16 \*\*\*  
## d$educ -0.009512 0.001083 -8.783 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.06538 on 529 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.1273, Adjusted R-squared: 0.1256   
## F-statistic: 77.13 on 1 and 529 DF, p-value: < 2.2e-16

### 2a.

It apppears there is a positive relationship between wages and education but is sightly hard to tell to do how "zoomed in" the plot is.

### 2b.

Without tranforming, we are missing some of the upper values of wages.

### 2d.

The relationship between time and education is negative.

### 2e.

Yes, I am happy with my decision to pursue a master's degree. Besides the enjoyment of the subject matter, it appears as one furthers education the time it takes to make $1 goes down.

### 2f.

The outliers appear to be rows 216, 507, and 518. After I remove them, the relationship is more significant. We now how stronger evidence that time is negatively associated with education.