Homework 1

Austin Lee

January 23, 2019

This is Homework 1, Problem 1. This first block is to load all of the necessary libraries I used to finish the homework.I used the describe function to gather more information about the dataset. options(scipen =999) allows us to look at the numeric values without scientific notation.

```
#install.packages('leaps',repos ="Http://cran.us.r-project.org")
require('leaps')
```

```
## Loading required package: leaps
```

```
## Warning: package 'leaps' was built under R version 3.5.2
```

```
library('olsrr')
```

```
## Warning: package 'olsrr' was built under R version 3.5.2
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
library('psych')
```

```
## Warning: package 'psych' was built under R version 3.5.2
```

```
library('DAAG')
```

```
## Warning: package 'DAAG' was built under R version 3.5.2
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'DAAG'
```

```
## The following object is masked from 'package:psych':
##
##     cities
```

```
library('corrplot')
```

```
## Warning: package 'corrplot' was built under R version 3.5.2
```

```
## corrplot 0.84 loaded
```

```
library('MASS')
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:DAAG':
##
##     hills
```

```
## The following object is masked from 'package:olsrr':
##
```

```
##     cement
```

```r
library('dynlm')
```

```
## Warning: package 'dynlm' was built under R version 3.5.2
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 3.5.2
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library('car')
```

```
## Warning: package 'car' was built under R version 3.5.2
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 3.5.2
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:DAAG':
##
##     vif
```

```
## The following object is masked from 'package:psych':
##
##     logit
```

```r
library('stats')

describe(nsw74psid1)
```

```
##       vars    n     mean       sd   median  trimmed      mad min      max
## trt      1 2675     0.07     0.25     0.00     0.00     0.00   0      1.0
## age      2 2675    34.23    10.50    32.00    33.69    11.86  17     55.0
## educ     3 2675    11.99     3.05    12.00    12.13     2.97   0     17.0
## black    4 2675     0.29     0.45     0.00     0.24     0.00   0      1.0
## hisp     5 2675     0.03     0.18     0.00     0.00     0.00   0      1.0
## marr     6 2675     0.82     0.38     1.00     0.90     0.00   0      1.0
## nodeg    7 2675     0.33     0.47     0.00     0.29     0.00   0      1.0
## re74     8 2675 18230.00 13722.25 17437.47 17103.86 12490.68   0 137148.7
## re75     9 2675 17850.89 13877.78 17008.06 16624.38 13271.66   0 156653.2
## re78    10 2675 20502.38 15632.52 19432.10 19155.53 14569.33   0 121173.6
##          range  skew kurtosis     se
## trt        1.0  3.39     9.52   0.00
## age       38.0  0.38    -1.12   0.20
## educ      17.0 -0.43     0.34   0.06
## black      1.0  0.92    -1.16   0.01
## hisp       1.0  5.11    24.09   0.00
## marr       1.0 -1.66     0.76   0.01
## nodeg      1.0  0.71    -1.50   0.01
## re74  137148.7  1.24     4.56 265.32
## re75  156653.2  1.35     5.84 268.32
```

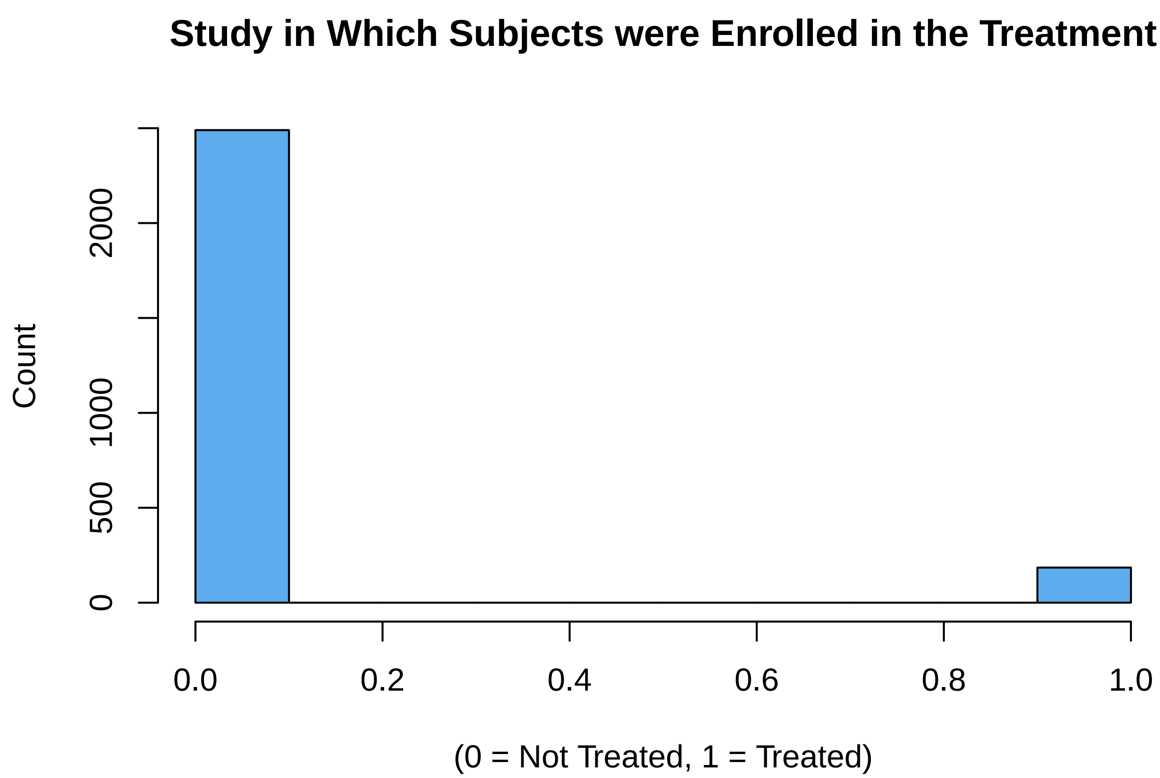
```
## re78  121173.6  1.25     3.73 302.25
```

```r
options(scipen=999)
attach(nsw74psid1)
```

Part A

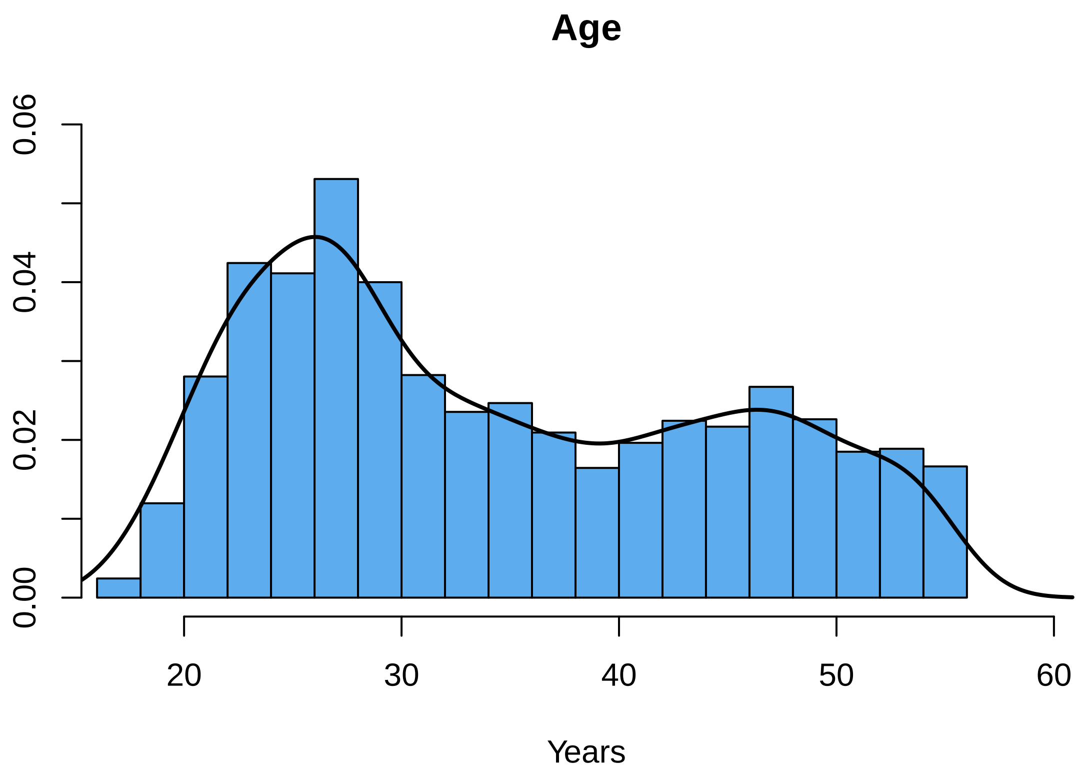
Trt is an indicator variable that represents whether the subjects were enrolled in the treatment program or not. If they apart of the treatment group, they were cateogorized with the bar at one, else they are categorized as 0. The data shows that more participants were not within the treatment group.

```r
hist(trt, col = 'steelblue2', main = 'Study in Which Subjects were Enrolled in the Treatment', xlab = '
```



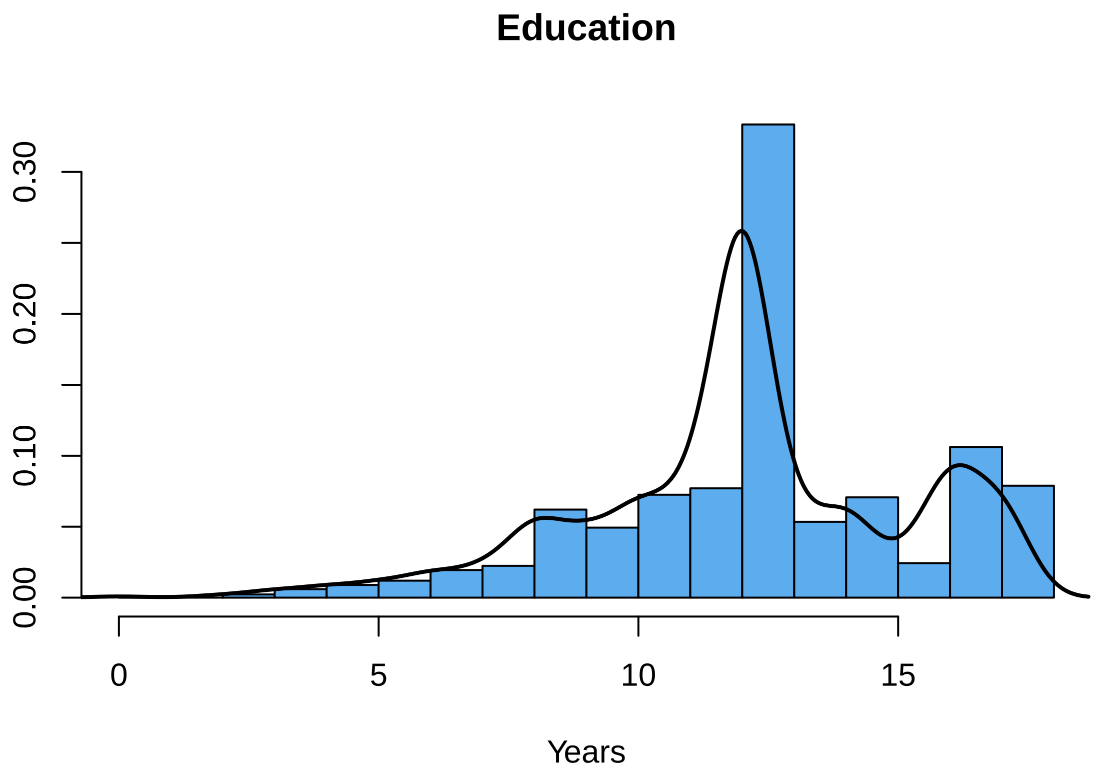
Age represents the age of each participant in the study. It seems the histogram would indicate that there is a bias in age. The histogram would indicate a bimodal distribution that is skewed slightly right.

```r
truehist(age,col = 'steelblue2', main = 'Age', xlab = 'Years', xlim = c(17,60), ylim = c(0,.06))
lines(density(age),lwd=2)
```



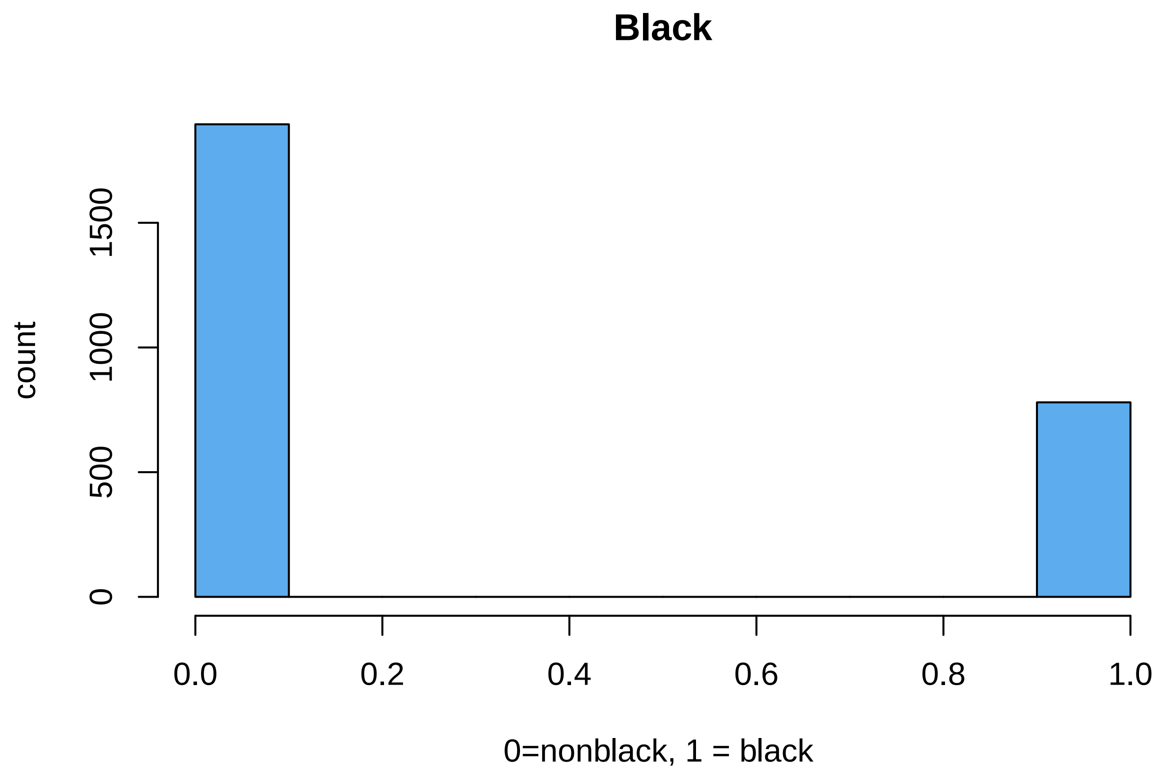
educ represents the amount of education each participant had in the study. It seems most of the subjects education peaked below 12 years.

```
truehist(educ,col = 'steelblue2', main ="Education", xlab = "Years")
lines(density(educ),lwd=2)
```



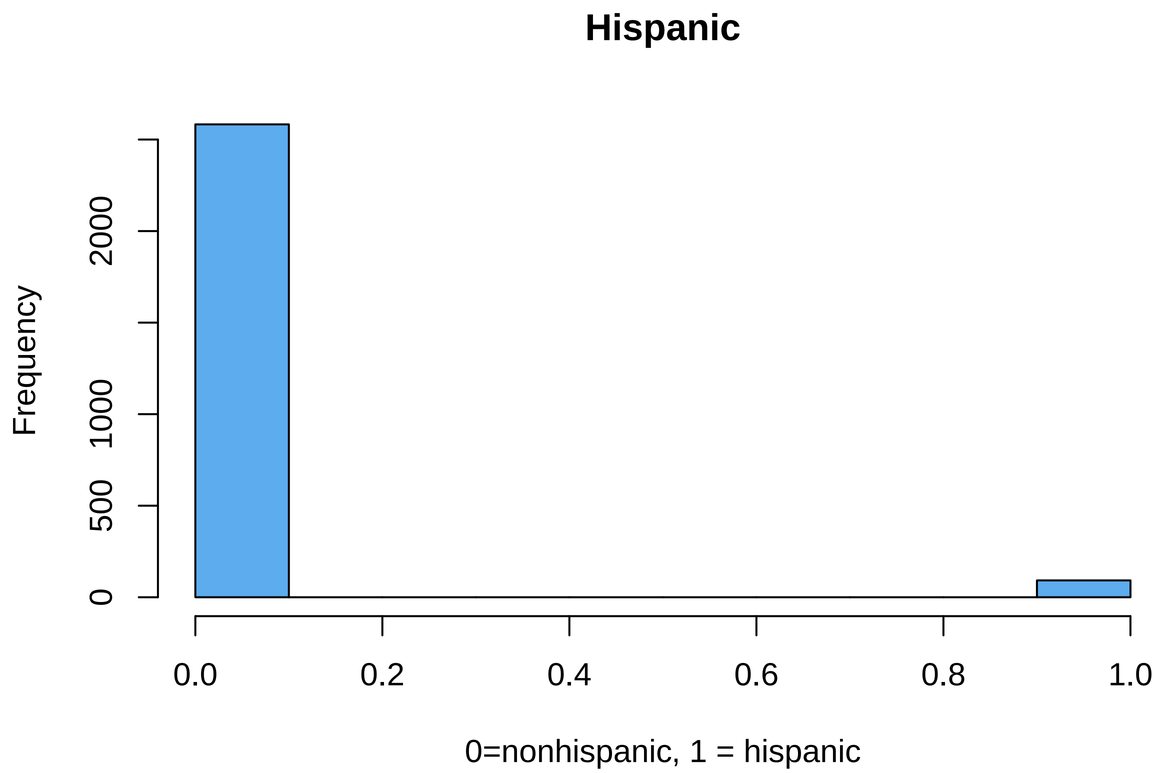
black is an indicator variable that represents whether the subjects are black or not. Most participants were not black.If they were, its represented by the bar at 1, or else they are categorized on the 0.

```
hist(black,col = 'steelblue2',main ='Black', xlab = '0=nonblack, 1 = black ',ylab = 'count')
```



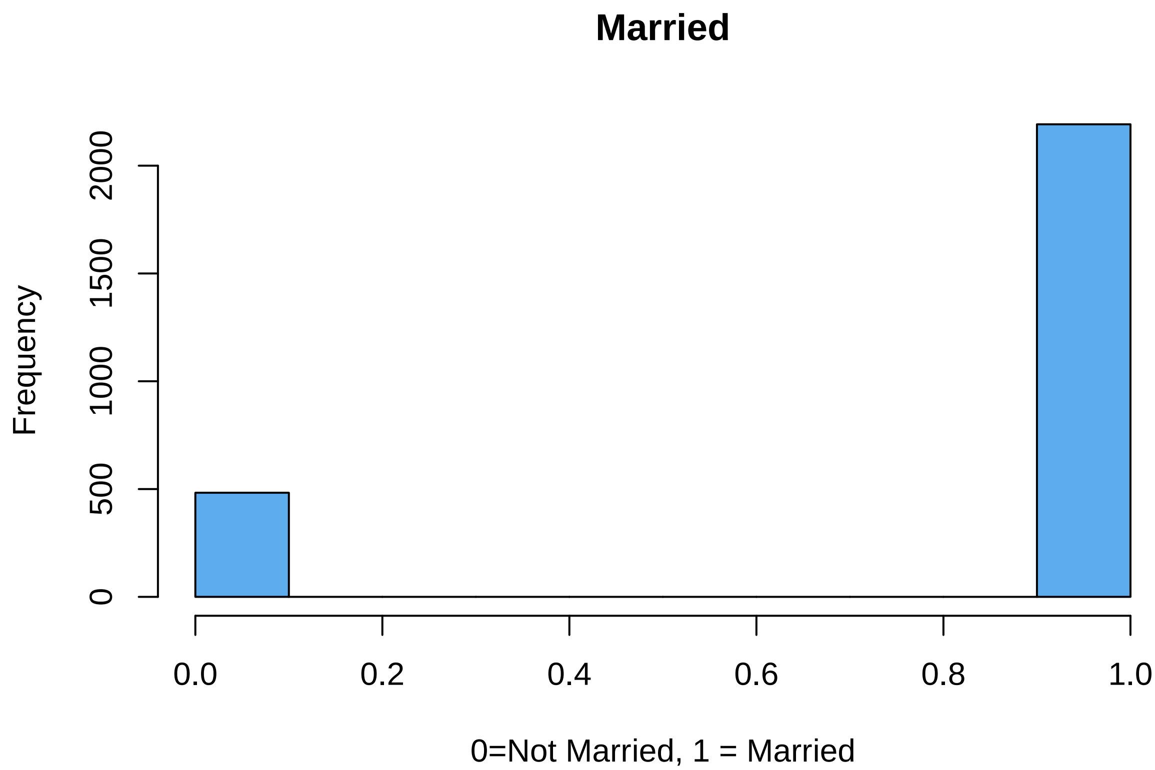
hisp is an indicator variable that represents whether the subjects are Hispanic or not. Most of the participants were not Hispanic. If they were, its represented by the bar at 1, or else they are categorized on the 0.

```
hist(hisp,col = 'steelblue2', main ='Hispanic', xlab= '0=nonhispanic, 1 = hispanic')
```



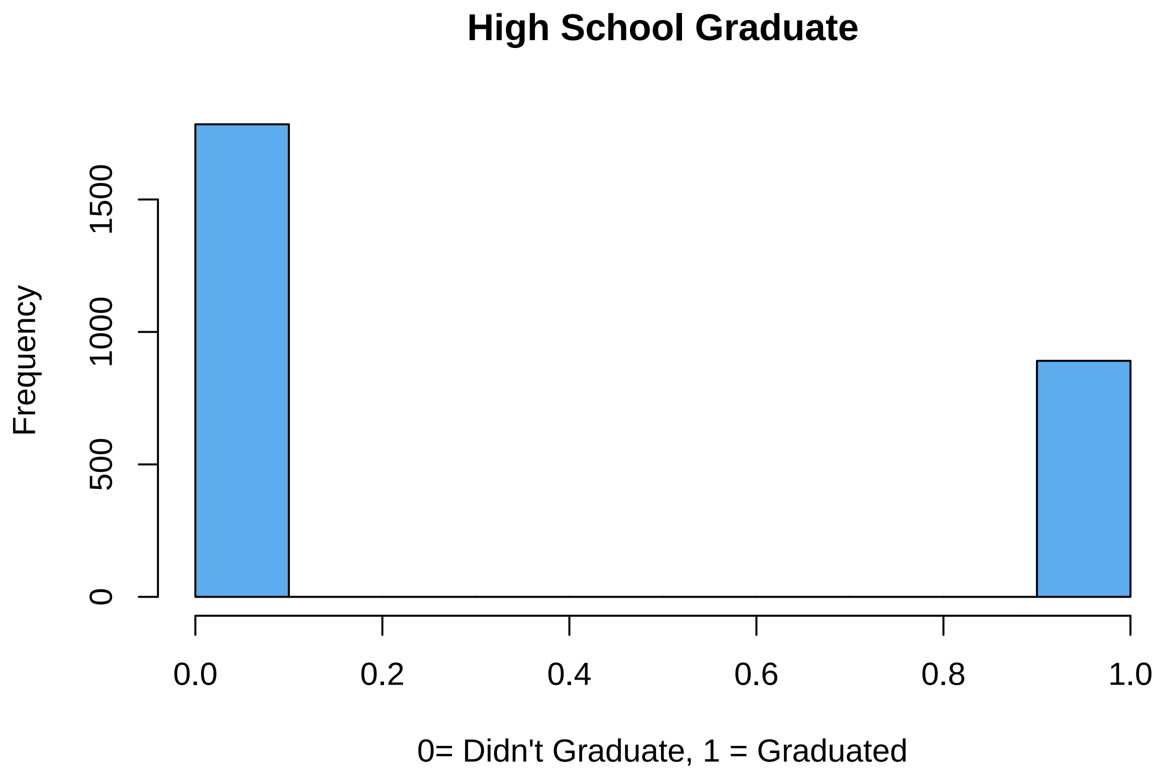
marr is an indicator variable that represents whether the subjects are married or not. Approximately 82%of those who participated in the study were married. If they were, its represented by the bar at 1, or else they are categorized on the 0.

```
hist(marr,col = 'steelblue2',main = 'Married', xlab = '0=Not Married, 1 = Married')
```



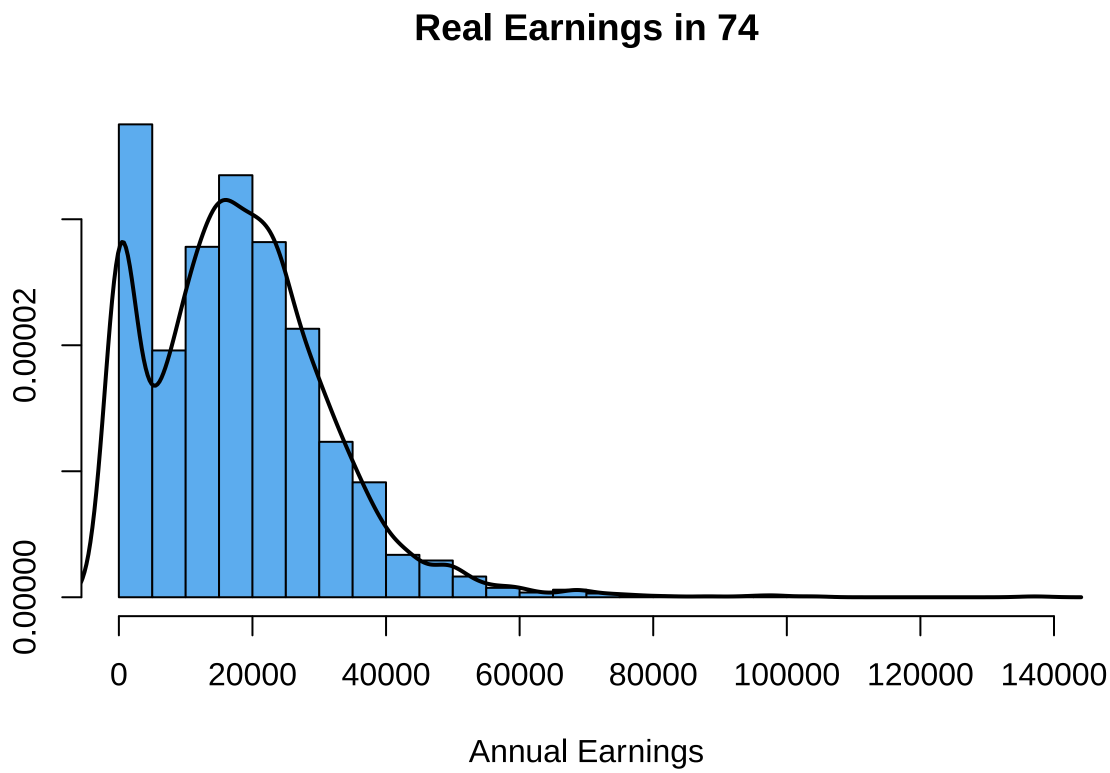
nodeg is an indicator variable that represents whether the subjects graduated high school or not Approxmiately 33% of the subjects did not graduate. If they did not graduate, its represented by the bar at 1, or else they are categorized on the 0.

```
hist(nodeg,col = 'steelblue2',main = "High School Graduate", xlab = "0= Didn't Graduate, 1 = Graduated")
```



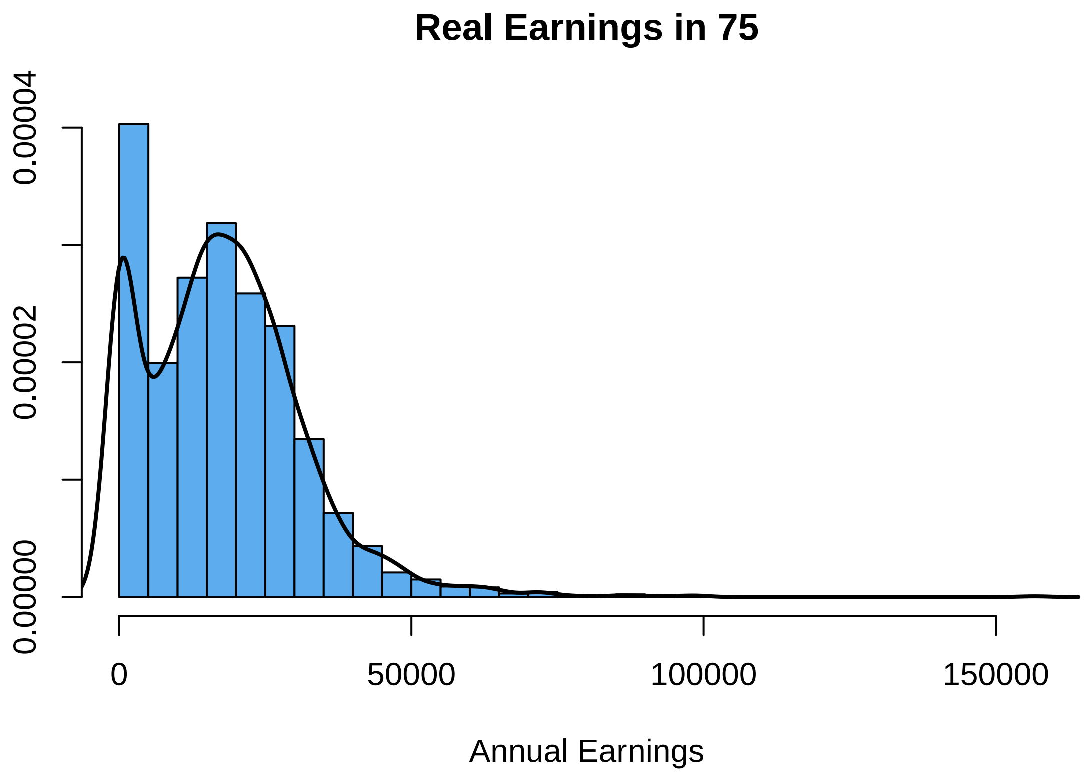
The histogram indicates a right skewed graph with bimodal modes at 0 and 20,000

```
truehist(re74,col = 'steelblue2', main = 'Real Earnings in 74', xlab = 'Annual Earnings')
lines(density(re74),lwd=2)
```



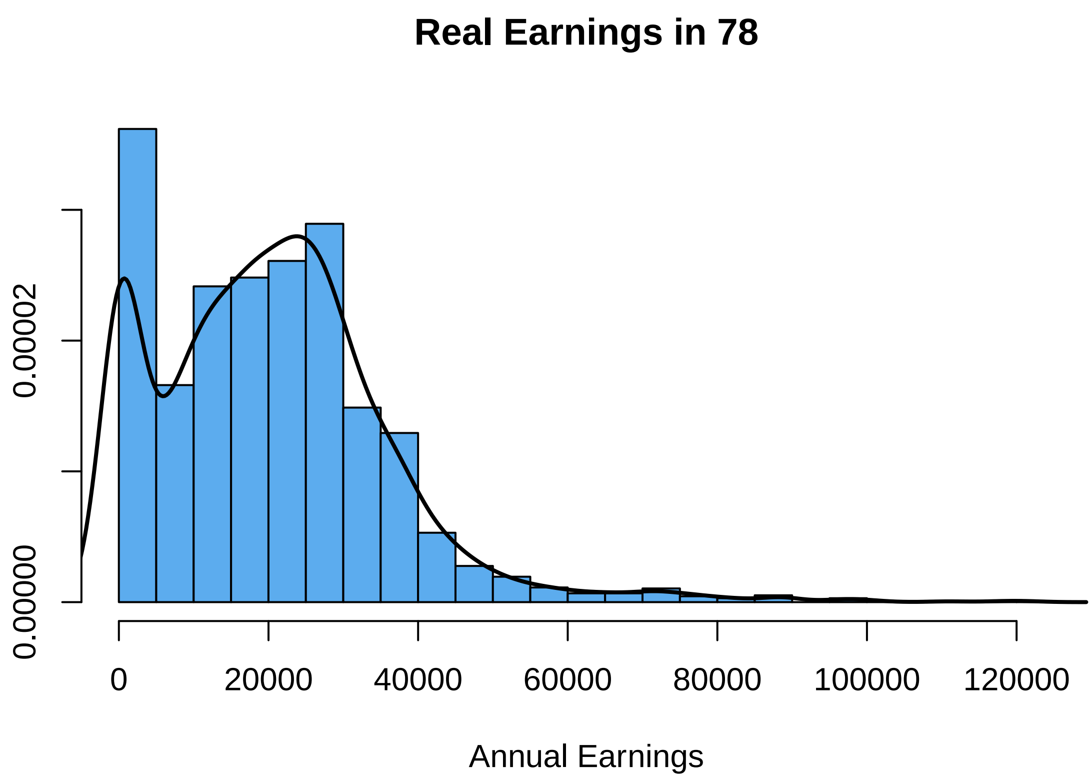
The histogram indicates a right skewed graph with bimodal means at 0 and around 20000

```
truehist(re75,col = 'steelblue2',main = 'Real Earnings in 75', xlab = 'Annual Earnings')
lines(density(re75),lwd=2)
```



The histogram indicates a right skewed graph with bimodal means at 0 and around 25000

```
truehist(re78,col = 'steelblue2',main = 'Real Earnings in 78', xlab = 'Annual Earnings')
lines(density(re78),lwd=2)
```



Part B

After regressing the model, we have a R^2 .2102 which indicates statistical measure of statistical measure of how close the dependent variable is explained by the explanatory variables collectively. It appears the trt, black, and nodeg have low t-values, rendering their predictive powers statistically insignificant

```
mod1 <- lm(re78 ~ trt + age + educ+ black + hisp + marr + nodeg+re74+re75, data = nsw74psid1)
summary(mod1)
```

```
##
## Call:
## lm(formula = re78 ~ trt + age + educ + black + hisp + marr +
##     nodeg + re74 + re75, data = nsw74psid1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -64870  -4302   -435   3786 110412
##
## Coefficients:
##                Estimate Std. Error t value             Pr(>|t|)
## (Intercept) -129.74276 1688.51706  -0.077               0.9388
## trt          751.94643  915.25723   0.822               0.4114
## age          -83.56559   20.81380  -4.015         0.0000611093 ***
## educ         592.61020  103.30278   5.737         0.0000000107 ***
## black       -570.92797  495.17772  -1.153               0.2490
## hisp        2163.28118 1092.29036   1.981               0.0478 *
```

```
## marr        1240.51952  586.25391   2.116              0.0344 *
## nodeg        590.46695  646.78417   0.913              0.3614
## re74           0.27812    0.02792   9.960 < 0.0000000000000002 ***
## re75           0.56809    0.02756  20.613 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10070 on 2665 degrees of freedom
## Multiple R-squared:  0.5864, Adjusted R-squared:  0.585
## F-statistic: 419.8 on 9 and 2665 DF,  p-value: < 0.00000000000000022
```

Part C

The Mallows CP statistic estimates the size of bias introduced into the predicted response. In a regression model variance and bias are at play which impede a model's predicting power. To pick the model with the least amount of bias, so that we may also pick the lowest variation. Models with CP Mallows values a lot larger than its predictors indicate there is substantial bias. The best CP Mallows values are the ones that are only slightly above their predictors. Therefore, the best model we should use is the model with predictors: age, educ, hisp, marr, marr, r74, r75 because its cp is 6.59499, the model closest to the number of predictors.

```
leaps(x = nsw74psid1[,1:9], y = nsw74psid1[,10], names = names(nsw74psid1)[1:9],method = 'Cp')
```

```
## $which
##     trt   age  educ black  hisp  marr nodeg  re74  re75
## 1 FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## 1 FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
## 1 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
## 1 FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
## 1 FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 1  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 1 FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
## 1 FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 1 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE
## 2 FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 2 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
## 2 FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE
## 2 FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## 2 FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE
## 2 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE
## 2 FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE
## 2  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## 2 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE
## 2 FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
## 3 FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 3 FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
## 3 FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 3 FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE
## 3 FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## 3 FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 3  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 3 FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE
## 3 FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE  TRUE
## 3 FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE
## 4 FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
```

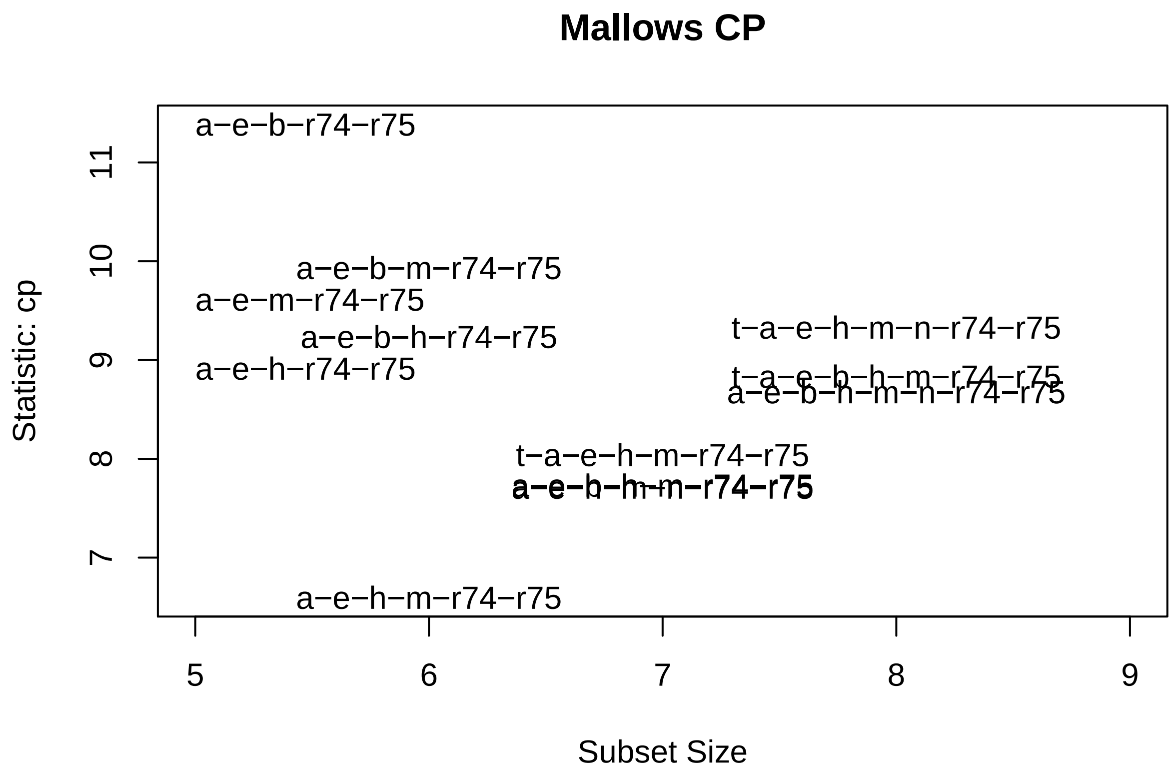
```
## 4 FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## 4 FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 4 FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE
## 4  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 4 FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE
## 4 FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
## 4 FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE
## 4 FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE  TRUE  TRUE
## 4 FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
## 5 FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## 5 FALSE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 5 FALSE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE
## 5 FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE
## 5  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE
## 5 FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
## 5 FALSE FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
## 5  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## 5 FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
## 5  TRUE FALSE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 6 FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
## 6 FALSE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
## 6 FALSE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE
## 6 FALSE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
## 6 FALSE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
## 6  TRUE  TRUE  TRUE FALSE  TRUE FALSE FALSE  TRUE  TRUE
## 6  TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE  TRUE
## 6 FALSE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE
## 6  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE
## 6  TRUE  TRUE  TRUE FALSE FALSE FALSE  TRUE  TRUE  TRUE
## 7 FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
## 7 FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 7  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE  TRUE
## 7 FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
## 7  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE
## 7 FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## 7  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
## 7  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE  TRUE
## 7  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE
## 7  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE
## 8 FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## 8  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE
## 8  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
## 8  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## 8  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
## 8  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## 8  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
## 8  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE
## 8  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
## 9  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE
##
## $label
##  [1] "(Intercept)" "trt"         "age"         "educ"        "black"
##  [6] "hisp"        "marr"        "nodeg"       "re74"        "re75"
##
```

```
## $size
##  [1]  2  2  2  2  2  2  2  2  2  3  3  3  3  3  3  3  3  3  3  4  4  4  4
## [24]  4  4  4  4  4  4  5  5  5  5  5  5  5  5  5  5  6  6  6  6  6  6  6
## [47]  6  6  6  7  7  7  7  7  7  7  7  7  7  8  8  8  8  8  8  8  8  8  8
## [70]  9  9  9  9  9  9  9  9  9 10
##
## $Cp
##  [1]  210.370496  580.785912 2884.988555 3202.136626 3302.179736
##  [6] 3379.606731 3456.308825 3699.442465 3771.056298  102.800135
## [11]  117.558157  164.491345  187.945278  194.021200  208.551290
## [16]  208.875510  211.854762  444.085270  510.399514   23.464124
## [21]   65.295330   66.417758   92.313692  102.123410  104.151836
## [26]  104.727389  112.232088  113.998847  114.679052   12.169519
## [31]   19.920420   23.705490   23.990497   25.127798   25.258108
## [36]   45.210911   54.442114   63.306807   63.427462    8.914197
## [41]    9.609933   11.384562   13.674672   14.168450   20.289980
## [46]   21.229743   21.643800   21.696798   24.274736    6.594990
## [51]    9.234664    9.932122   10.396771   10.742827   10.907879
## [56]   10.942812   12.601860   13.239594   15.650559    7.709776
## [61]    7.730800    8.039315   10.492335   10.764152   10.830300
## [66]   11.172161   12.266104   12.354467   14.539275    8.674975
## [71]    8.833436    9.329353   11.922381   12.477501   24.119526
## [76]   40.908967  107.196967  432.894236   10.000000
```

```
ss = regsubsets(re78 ~ trt + age + educ + black + hisp + marr + nodeg+re74+re75, method = c('exhaustive
subsets(ss,statistic = "cp", legend = F, main = 'Mallows CP', min.size =5 )
```

```
##       Abbreviation
## trt              t
## age              a
## educ             e
## black            b
## hisp             h
## marr             m
## nodeg            n
## re74           r74
## re75           r75
```

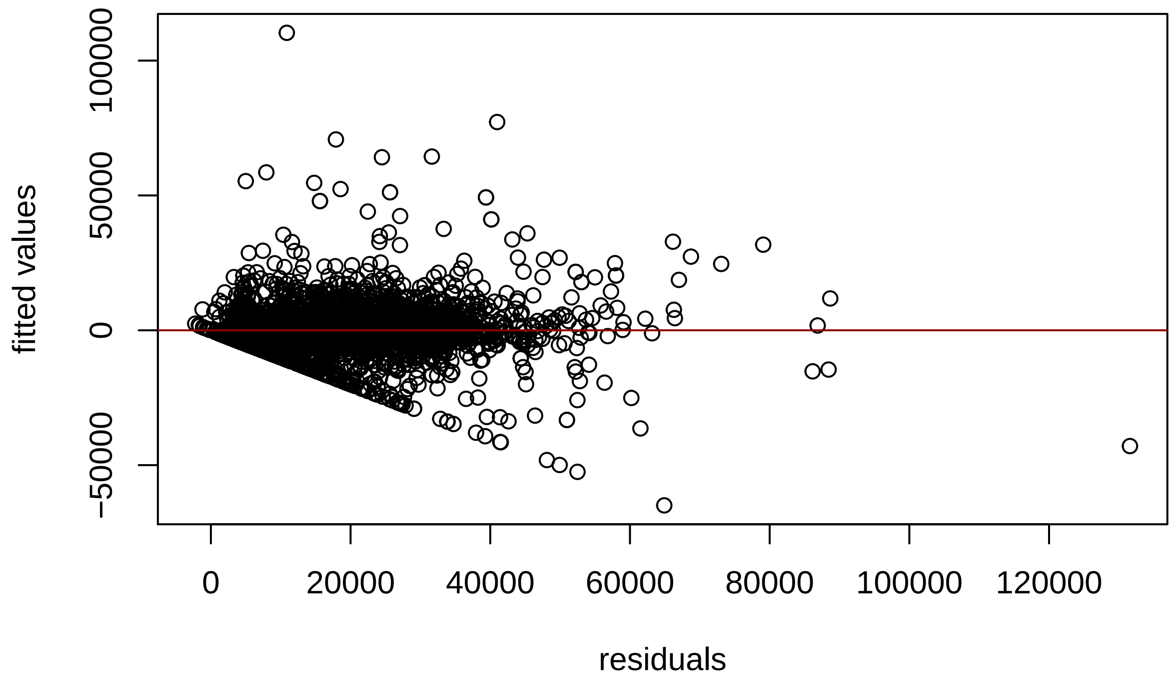
```
legend(6,700,legend=c('t = treatment', 'a = age', 'e = education', 'b = black', 'h = hispanic',' m = ma
```



Part D

Plotting fitted vs residuals values

```r
mod2 <-lm(re78 ~ age + educ + hisp + marr+ re74 +re75, data = nsw74psid1)
plot(mod2$fitted.values, mod2$residuals, xlab = "residuals", ylab = "fitted values")
abline(0,0, col = "dark red")
```



Part E

Our vif factors are close to 1. Our model is stronger because there is less multicolinearity.

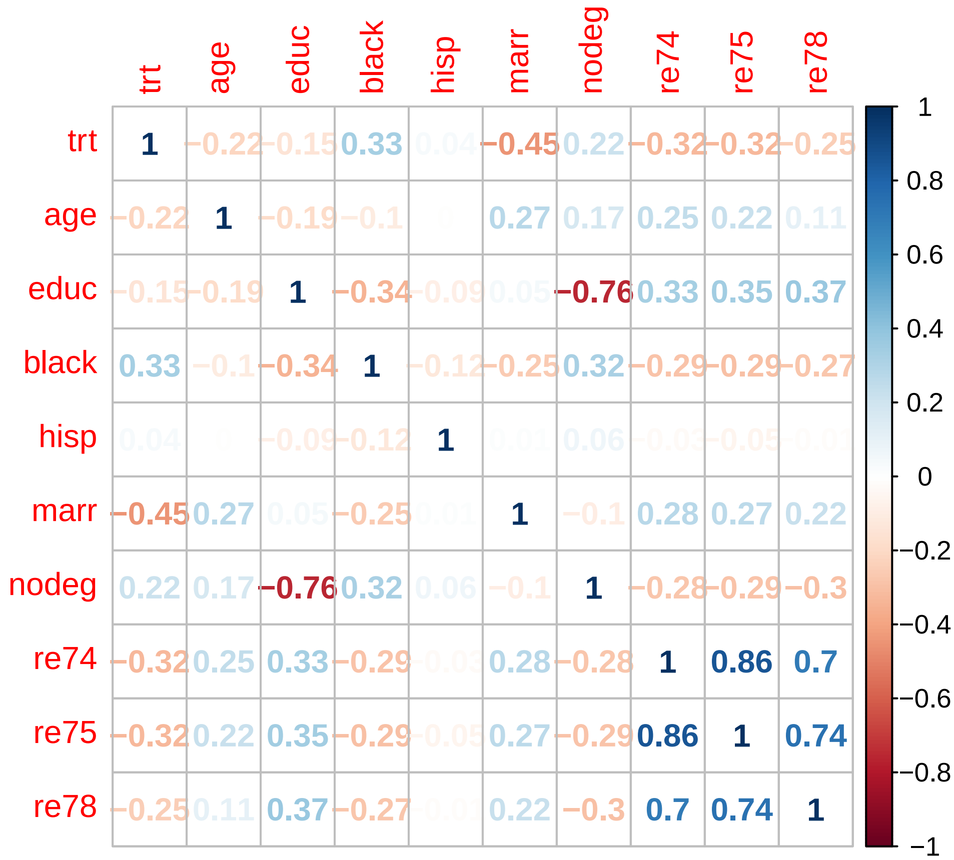
```
vif(mod2)
```

```
##      age     educ     hisp     marr     re74     re75
## 1.227765 1.264142 1.010674 1.142693 3.863304 3.833144
```

Part F

There is very high correlation between education and nodeg. This is a good reason why the Mallows CP suggested to take nodeg out and limit bias within the model. Although there is high correlation between re74 and re75, it is heavily correlated to re78, which may amplify the model's predicting power, justifying their relevance.

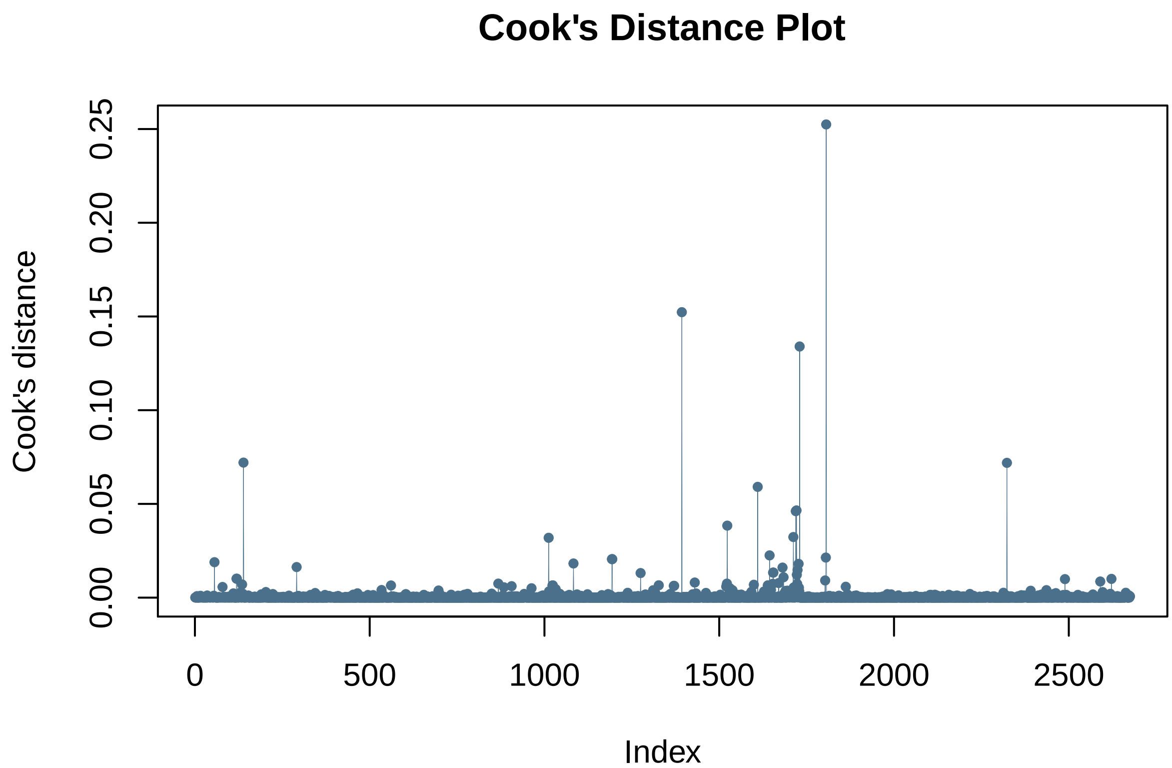
```
corrplot(cor(nsw74psid1),method= "number")
```



Part G

It appears that there are a few outliers, but their Cook's Distance values are less than 1, so it would be inappropriate to drop outliers as they may hold legitimacy. It is important to investigate into those few outlier datapoints to determine why they deviate from the mean.

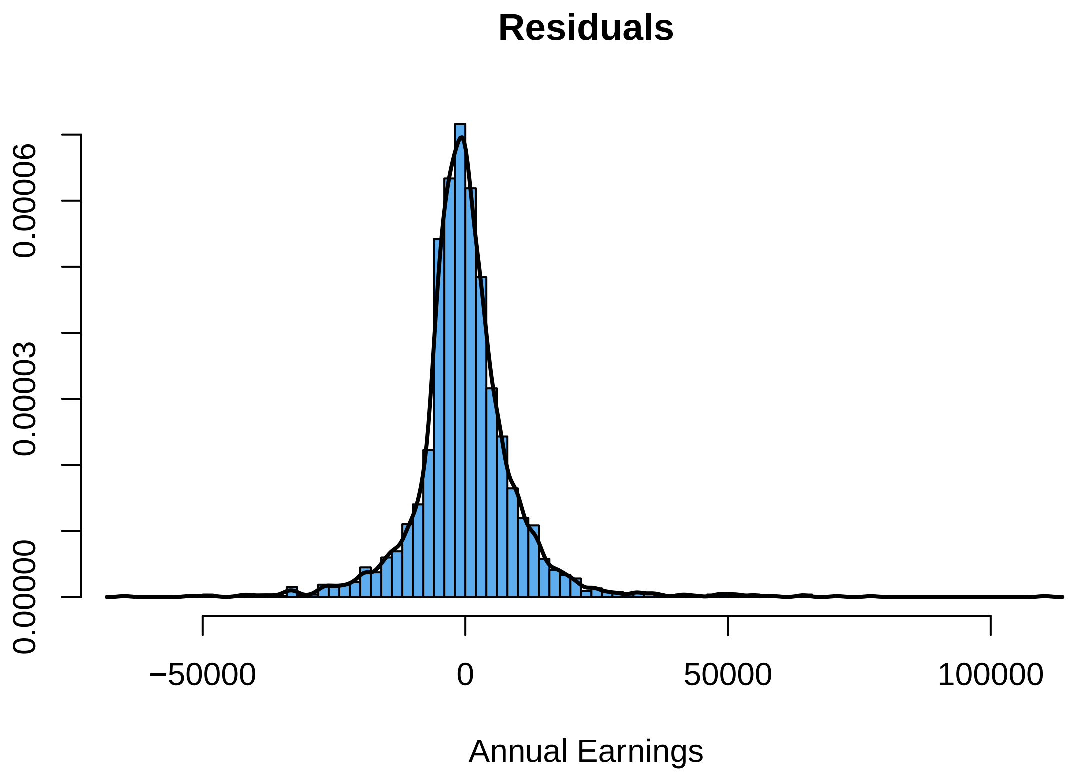
```
mod2_cooks = cooks.distance(mod2)
plot(mod2_cooks, ylab="Cook's distance",type='o',main="Cook's Distance Plot",col="skyblue4", pch=20,lwd=
```



Part H

Residuals are mostly centered at 0, which is a good indication that the model is close to meeting the requirement for the Gauss Markov assumption. However the histogram plot seems to skew right, indicating that the distribution of residuals arent exactly perfectly distributed at 0.

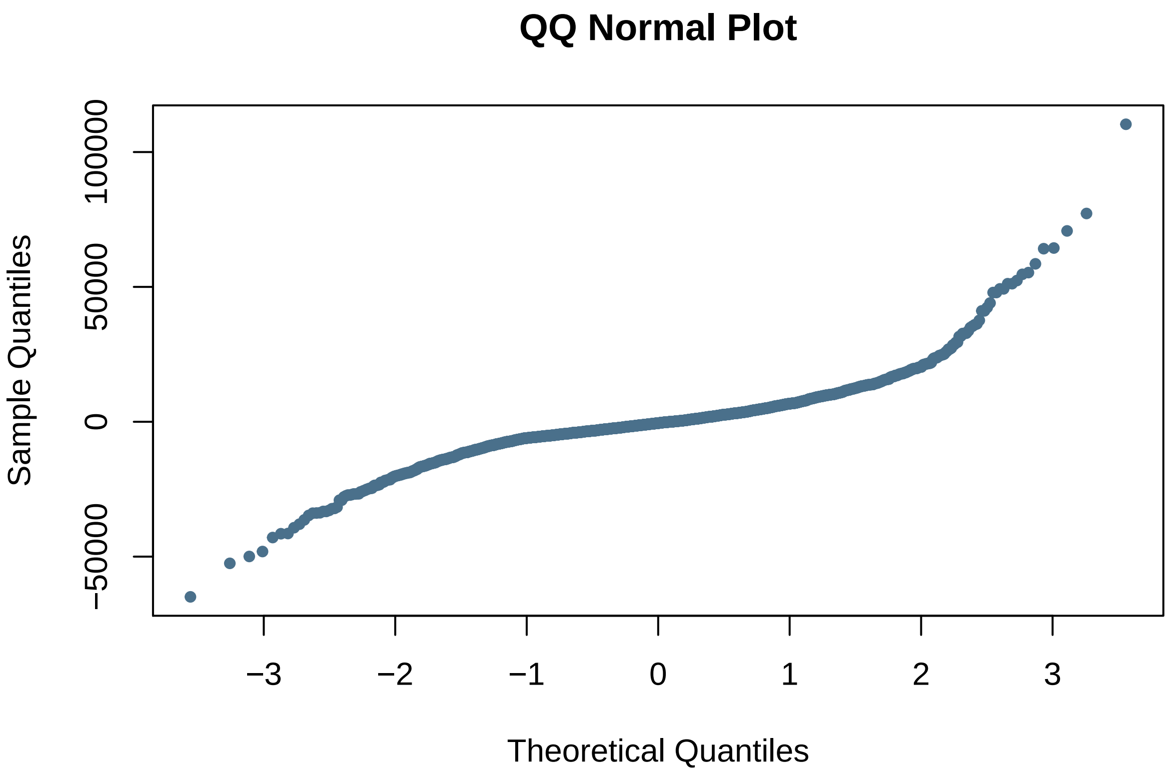
```
truehist(mod2$res,col = 'steelblue2', main = 'Residuals', xlab = 'Annual Earnings')
lines(density(mod2$res),lwd=2)
```



Part I

The dots in the qqnorm plot do not fit a straight line, the residual distribution do not follow a normal distribution as indicated by the histogram above. the qqnormal plot also indicates that the residual distribution is skewed right, as mentioned previously.

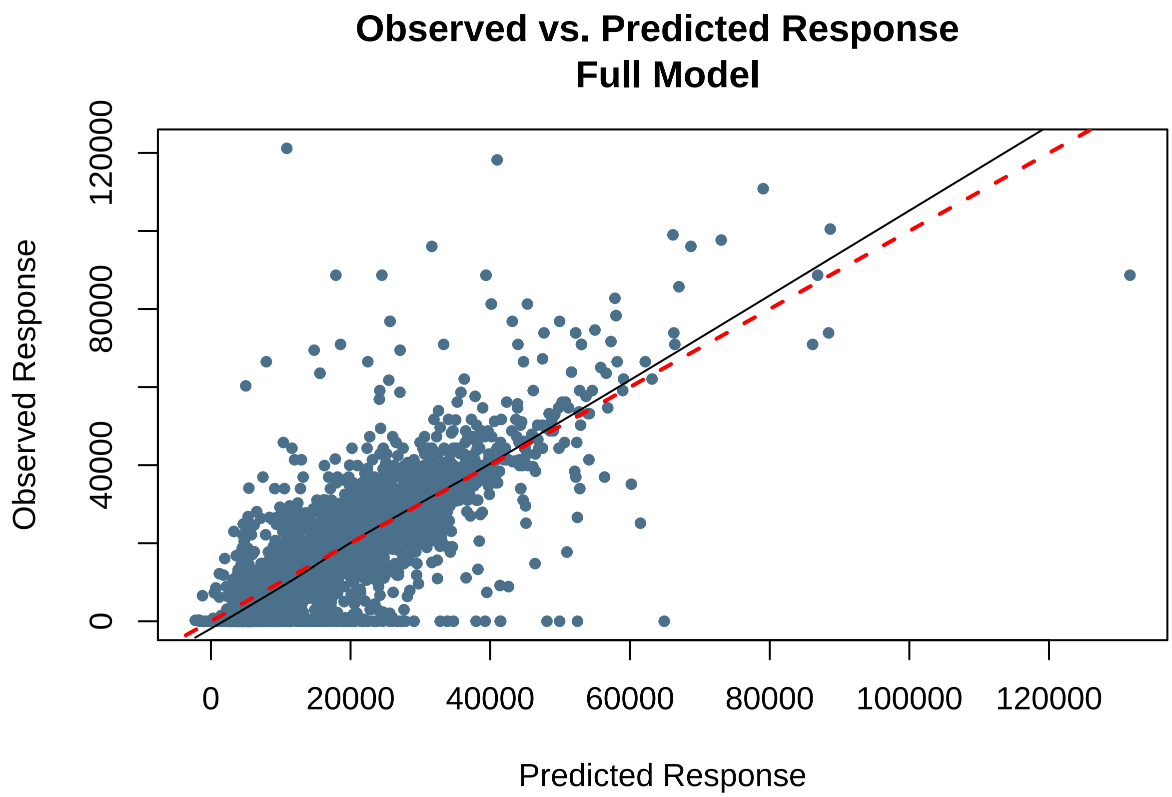
```
qqnorm(mod2$res,col="skyblue4", pch=20,lwd=1,main="QQ Normal Plot")
```



Part J

Locally Weighted Smoothing helps with graphically visualization of a regression line, where lots of noisy data is present.There also seems to be a lot of observed responses where real earnings = 0, our predicted model does not have the same.To enhance our model, we may need to somehow account for the observed values at 0. A valuable question to survey is whether these observed data points are unemployed or not, so further investigation may be realized.

```
plot(mod2$fitted.values,re78,pch=20,col="skyblue4",cex=1,xlab="Predicted Response",ylab="Observed Respo
lines(lowess(mod2$fitted.values,re78))
abline(0,1,col="red",lwd=2,lty=2)
```



Problem Two

```
data2 <- read.table(file ='c:\\Users\\Austin\\Documents\\R\\Copy_of_Chapter2_exercises_data.csv', heade
attach(data2)

describe(data2$GRGDP)
```

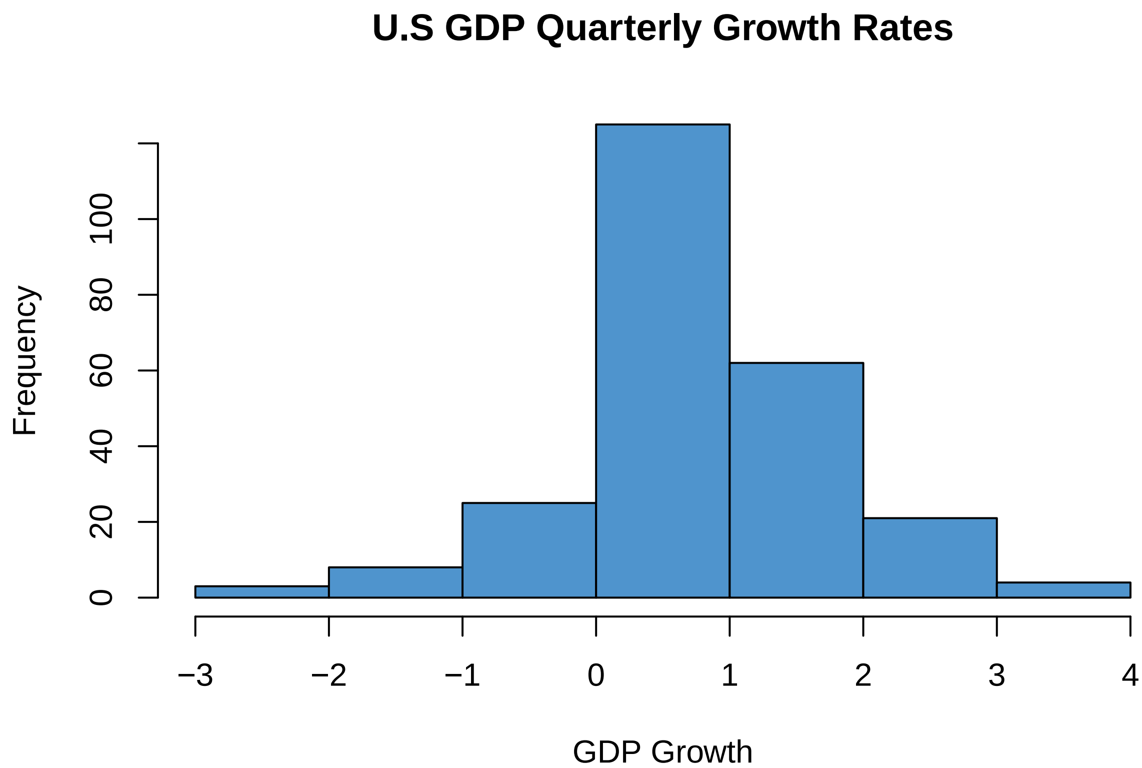
```
##    vars   n mean   sd median trimmed  mad   min  max range  skew kurtosis
## X1    1 248  0.8 0.98   0.77    0.82 0.73 -2.71 3.93  6.64 -0.19     1.38
##      se
## X1 0.06
```

```
describe(data2$RETURN)
```

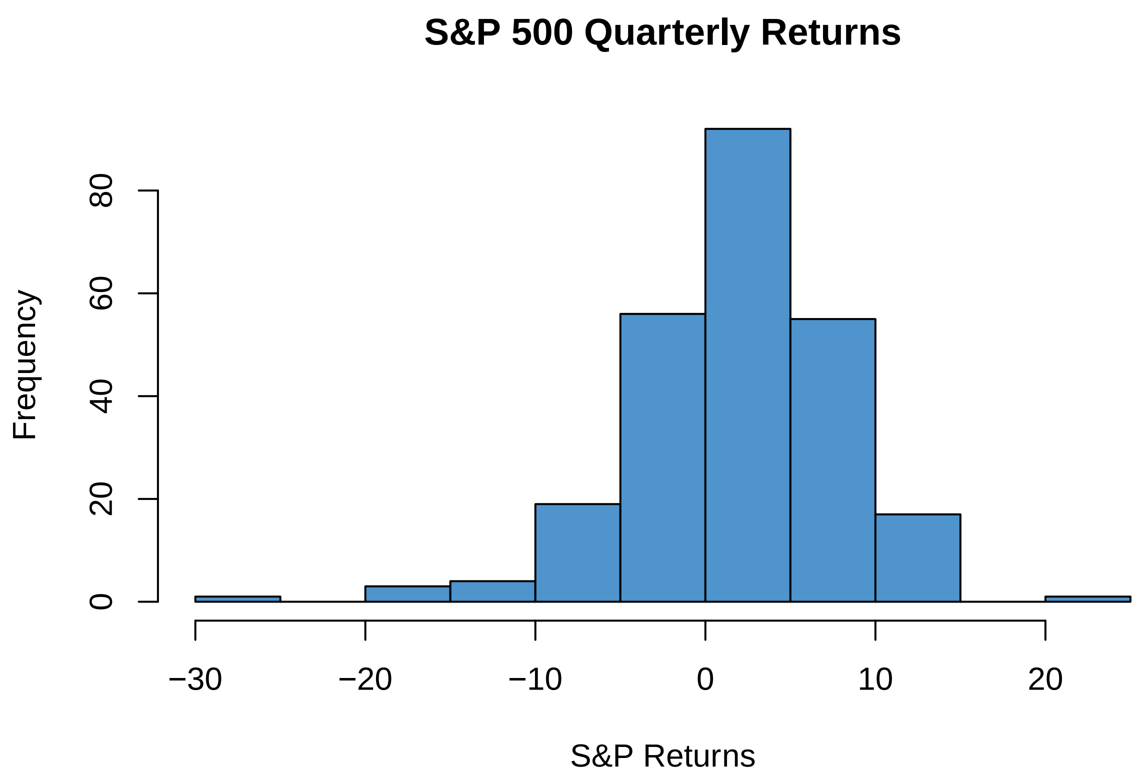
```
##    vars   n mean   sd median trimmed  mad    min   max range  skew
## X1    1 248 1.96 6.08   2.02    2.21 5.09 -26.94 20.12 47.05 -0.68
##    kurtosis   se
## X1     2.23 0.39
```

The histograms seem to both be left skewed.

```
hist(data2$GRGDP,col = 'steelblue3', main = "U.S GDP Quarterly Growth Rates", xlab = "GDP Growth")
```




```r
hist(data2$RETURN, col = 'steelblue3', main= "S&P 500 Quarterly Returns", xlab = "S&P Returns")
```



The correlation is .2702427. GDP growth is positively correlated, but is not a perfect indication of the S&P 500.

```
cor(data2$GRGDP, data2$RETURN)
```

```
## [1] 0.2702427
```

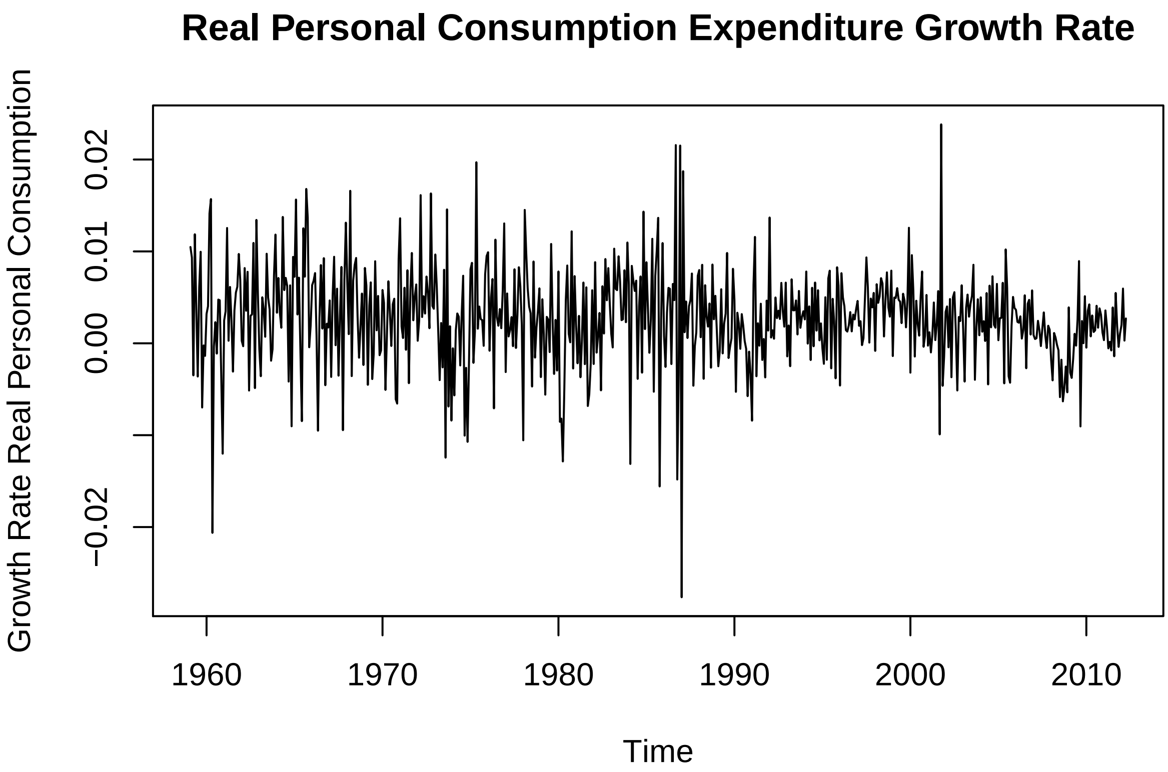
Problem Three

Here, I load in the data and create the time series for real personal consumption expenditure, real disposable personal income

```
data <- read.table(file ='C:\\Users\\Austin\\Documents\\R\\rpe rdi.csv', header =T, sep = ",")
ts_rpce <- ts(data$rpce, start = 1959, freq = 12)
ts_rdpi <- ts(data = data$rdpi, start= 1959, freq = 12)
```

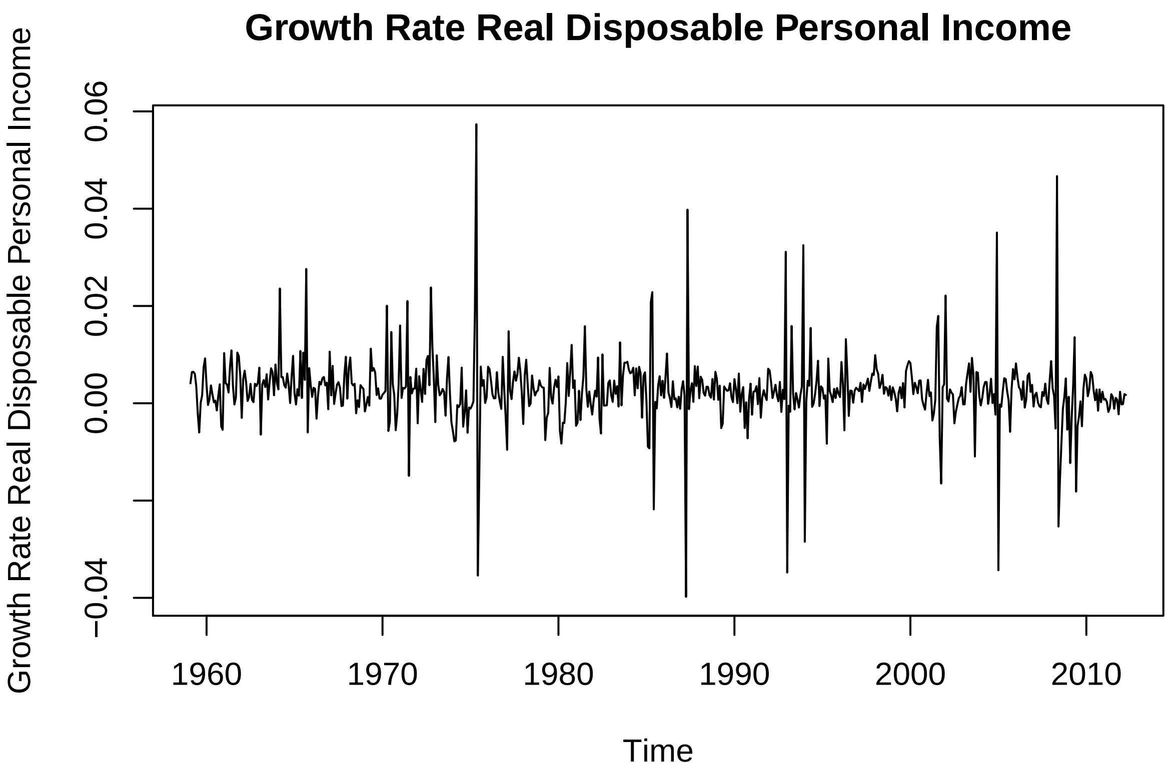
Now, I calculate the growth rate for real personal consumption expenditure and plot its time series.

```
log_ts_rpce <- log(ts_rpce)
log_ts_rpce_diff <- diff(log_ts_rpce,lag =1)
plot(log_ts_rpce_diff,ylab = "Growth Rate Real Personal Consumption", main = "Real Personal Consumption
```



Afterwards, I plot the growth for real disposable personal income.

```
log_ts_rdpi <-log(ts_rdpi)
log_ts_rdpi_diff <- diff(log_ts_rdpi, lag =1)
plot(log_ts_rdpi_diff, ylab = "Growth Rate Real Disposable Personal Income", main = "Growth Rate Real Di
```



Using Macroeconomic theory, we may come to understand why real personal disposable income is more volatile than real consumption expenditure. In theory, a person's consumption is not determined by their current income, but by their expected income in future years, which would lead to consumption smoothing.The data suggests that real disposable income peaks and 6% and -4% maximum and minimum, whereas personal consumption expenditure never rises above or below 3% absolute. As a rational actor, a person will consume based on what he or she expects in the future. An increase in temporary disposable income does not mean the person will use it all; their consumption will be smooth over time.

Part B

Our coefficients are statistically significant, however, the R^2d is low, which may indicate that real disposable personal income does not precisely explain the growth rate of consumption expenditure entirely. ***Interpretation***: if real disposable income increases by 1%, we expect real personal expenditure consumption to increase by .17%

```r
mod1 <- lm(log_ts_rpce_diff ~ log_ts_rdpi_diff)
summary(mod1)
```

```
##
## Call:
## lm(formula = log_ts_rpce_diff ~ log_ts_rdpi_diff)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0303967 -0.0029727  0.0001525  0.0030417  0.0244545
##
```

```
## Coefficients:
##                   Estimate Std. Error t value            Pr(>|t|)
## (Intercept)      0.0022542  0.0002242  10.055 < 0.0000000000000002 ***
## log_ts_rdpi_diff 0.1745671  0.0292029   5.978        0.00000000377 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.005318 on 637 degrees of freedom
## Multiple R-squared:  0.05312,    Adjusted R-squared:  0.05163
## F-statistic: 35.73 on 1 and 637 DF,  p-value: 0.000000003768
```

Part C

This model has a slightly higher R^2, and both coefficients are statistically significant. Adding a lag would theoretically make sense sense because our model acquires more information pertaining to the next year and what future real disposable personal income may be.

```
mod2 <-dynlm(log_ts_rpce_diff ~ log_ts_rdpi_diff+ L(log_ts_rdpi_diff,1))
summary(mod2)
```

```
##
## Time series regression with "ts" data:
## Start = 1959(3), End = 2012(4)
##
## Call:
## dynlm(formula = log_ts_rpce_diff ~ log_ts_rdpi_diff + L(log_ts_rdpi_diff,
##     1))
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0300734 -0.0028702 -0.0000006  0.0029797  0.0255131
##
## Coefficients:
##                         Estimate Std. Error t value             Pr(>|t|)
## (Intercept)            0.0019887  0.0002405   8.268 0.000000000000000798
## log_ts_rdpi_diff       0.1872692  0.0293883   6.372 0.000000000358039089
## L(log_ts_rdpi_diff, 1) 0.0828596  0.0293877   2.820              0.00496
##
## (Intercept)            ***
## log_ts_rdpi_diff       ***
## L(log_ts_rdpi_diff, 1) **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.005285 on 635 degrees of freedom
## Multiple R-squared:  0.06479,    Adjusted R-squared:  0.06185
## F-statistic:    22 on 2 and 635 DF,  p-value: 0.00000000058
```

Problem 4

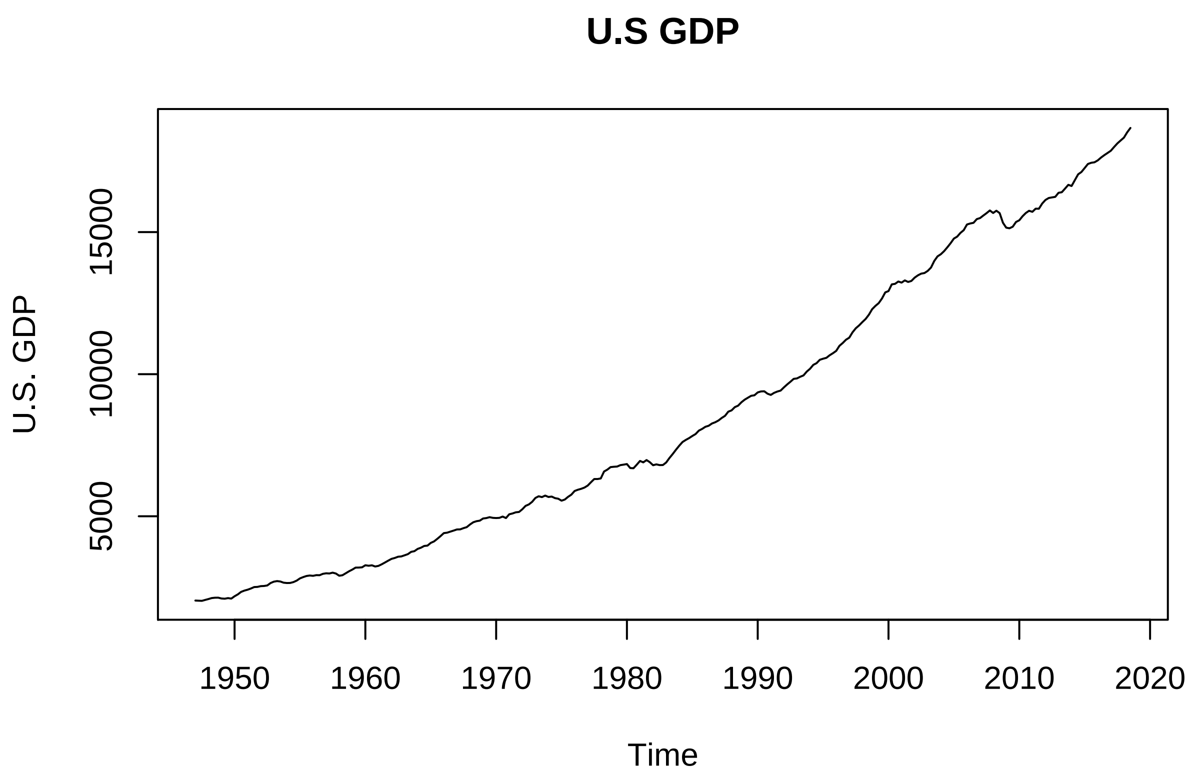
Here, I download GDP, Yen to Dollar Exchange, Ten yield rates, and unemployment rate

```
GDP <- read.table(file ='C:\\Users\\Austin\\Documents\\R\\GDPC1.csv', header = T, sep = "," )
yendol <- read.table(file = 'C:\\Users\\Austin\\Documents\\R\\EXJPUS.csv', header = T, sep = ",")
```

```r
tenyield <- read.table(file = 'C:\\Users\\Austin\\Documents\\R\\GS10.csv', header =T, sep = ",")
unemploy <- read.table(file = 'C:\\Users\\Austin\\Documents\\R\\UNRATE.csv', header = T, sep = ",")
```

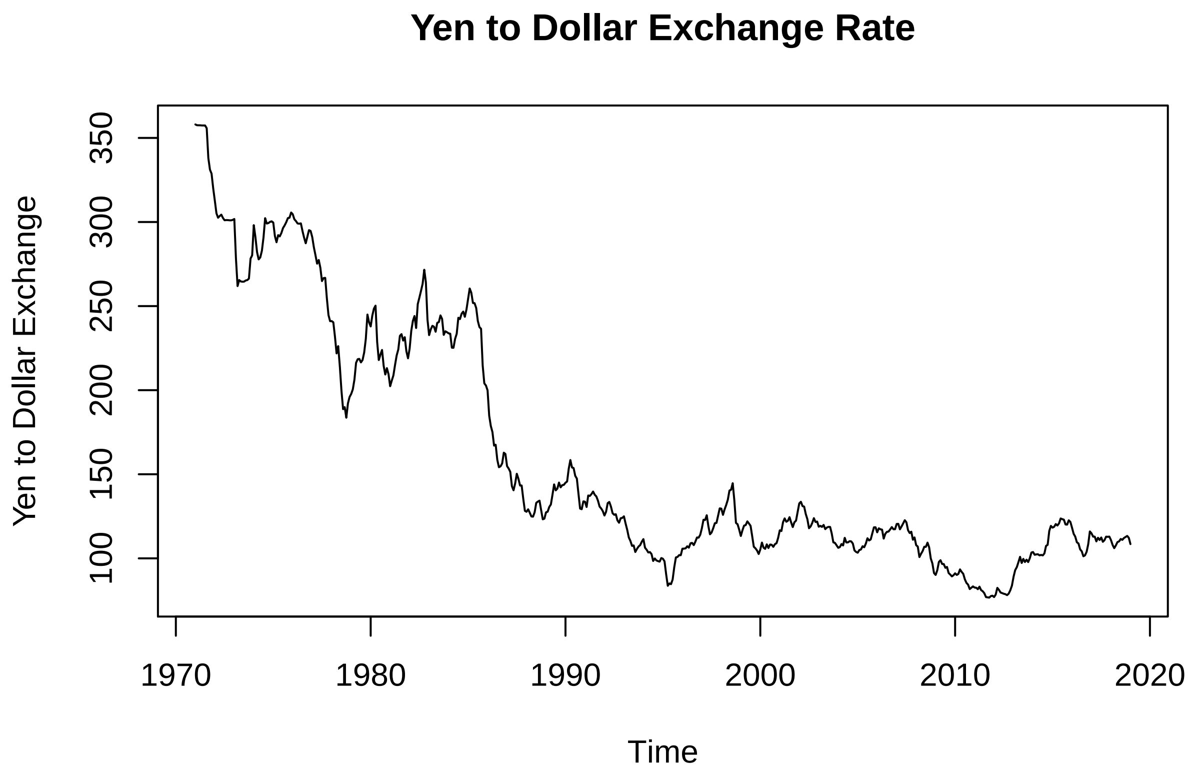
This is not first order or second order weakly stationary because it does not revert to any mean on a consistent basis.

```r
ts_GDP <-ts(data = GDP$GDPC1, start = 1947, freq = 4)
seq_GDP <- seq(1947,2018, length = length(ts_GDP))
plot(ts_GDP, main = "U.S GDP", ylab = "U.S. GDP")
```



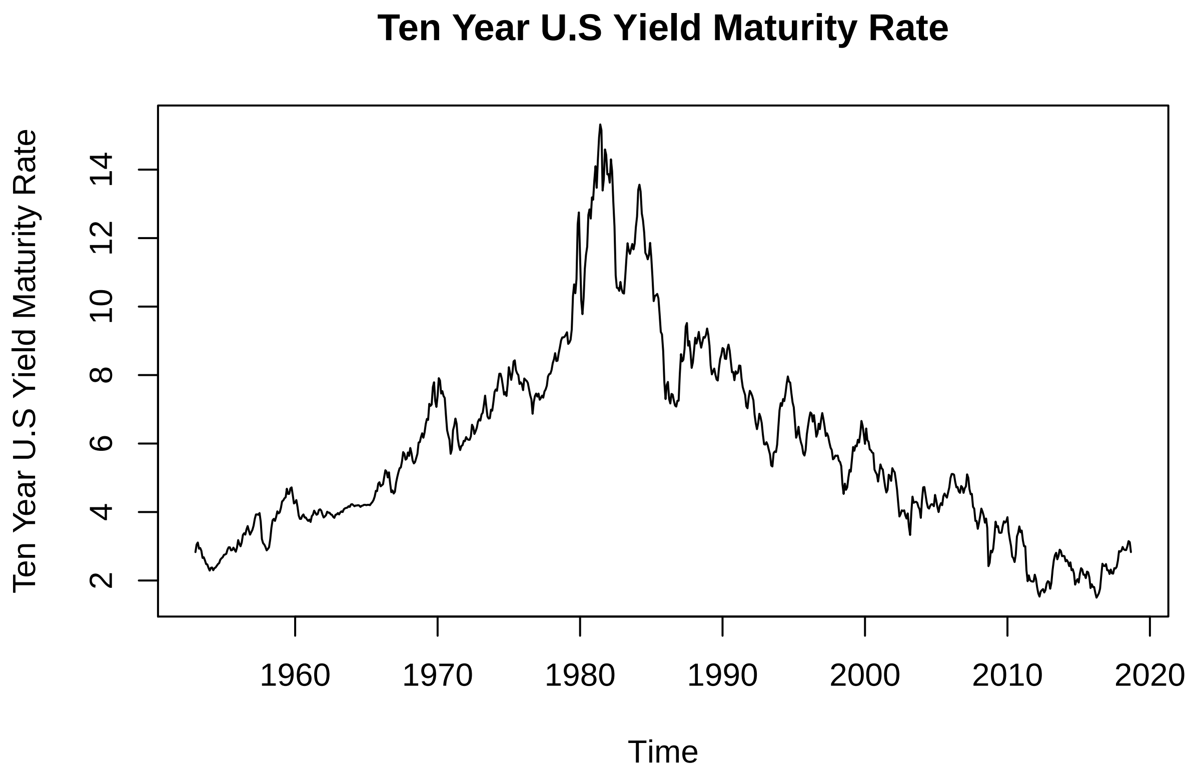
This is not first order or second order weakly stationary because it does not revert to any mean on a consistent basis.

```r
ts_yendol <- ts(data = yendol$EXJPUS, start = 1971, freq = 12)
plot(ts_yendol, main = "Yen to Dollar Exchange Rate", ylab ="Yen to Dollar Exchange")
```



This is not first order or second order weakly stationary, it does not revert to any mean on a consistent basis

```
ts_tenyield <- ts(data = tenyield$GS10, start = 1953, freq = 12)
plot(ts_tenyield, main = "Ten Year U.S Yield Maturity Rate", ylab = "Ten Year U.S Yield Maturity Rate")
```



More data would suggest a first order weakly stationary, however, there is insignificant oscilation at any particular mean. Therefore, it does not revert to any mean on a consistent basis

```
ts_unemploy <- ts(data = unemploy$UNRATE, start = 1948, freq = 12)
summary(unemploy$UNRATE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.500   4.600   5.600   5.763   6.800  10.800
```

```
plot(ts_unemploy, main =  "Unemployment Rate in the U.S", ylab = "Unemployment Rate in the U.S")
```

Problem 5

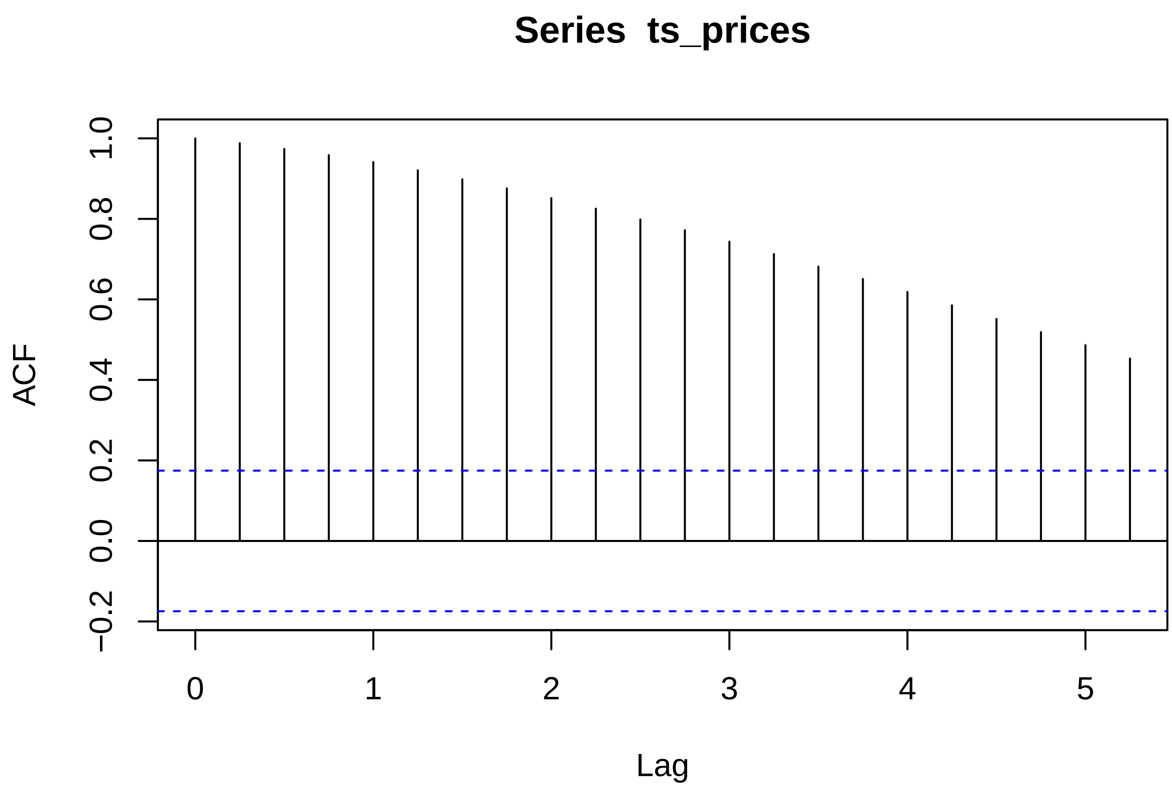
Here, I load the data and create the time series for prices, interest rates, price growth, and interest rate growth

```
data1 <- read.table(file = 'C:\\Users\\Austin\\Documents\\R\\HW143.csv', header =T, sep = ",")
ts_prices <- ts(data1$P,start = 1980, freq = 4)
ts_int_rate <- ts(data1$R..in..., start = 1980, freq = 4)
ts_prices_growth <- diff(log(ts_prices),1)
ts_int_rate_growth <- diff(log(ts_int_rate),1)
```

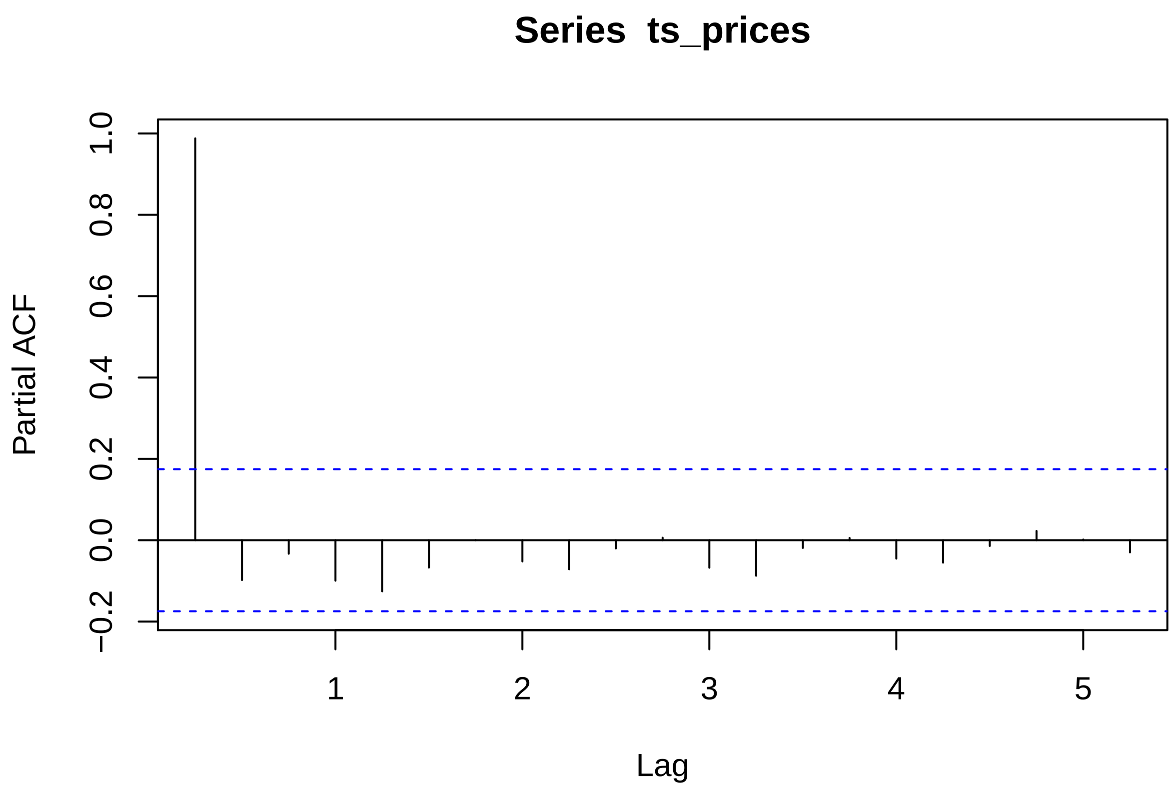
Autocorrelation gives us information about temporal advantages and patterns in specific time series. For extended periods for a time series, the autocorrelation function explains whether they are correlated at seperated points. The partial autocorrelated function removes separated observations when comparing between distant lags. As points become more separated, we may expect ACF and PACF to be 0 if the time series is not time dependent.

According to acf, there is a great autocorrelation dependence within the time series. However, when we break down into the partial acf, we see a lack in time dependence.

```
acf(ts_prices)
```

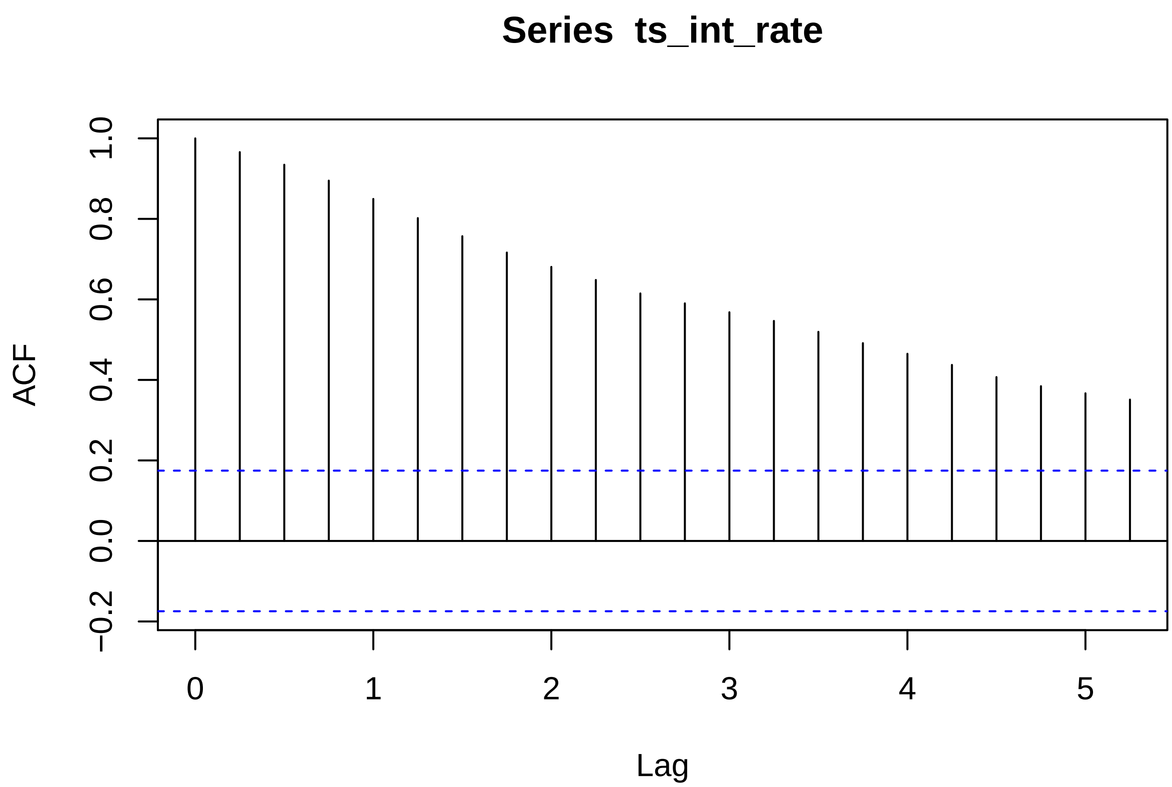



```r
pacf(ts_prices)
```

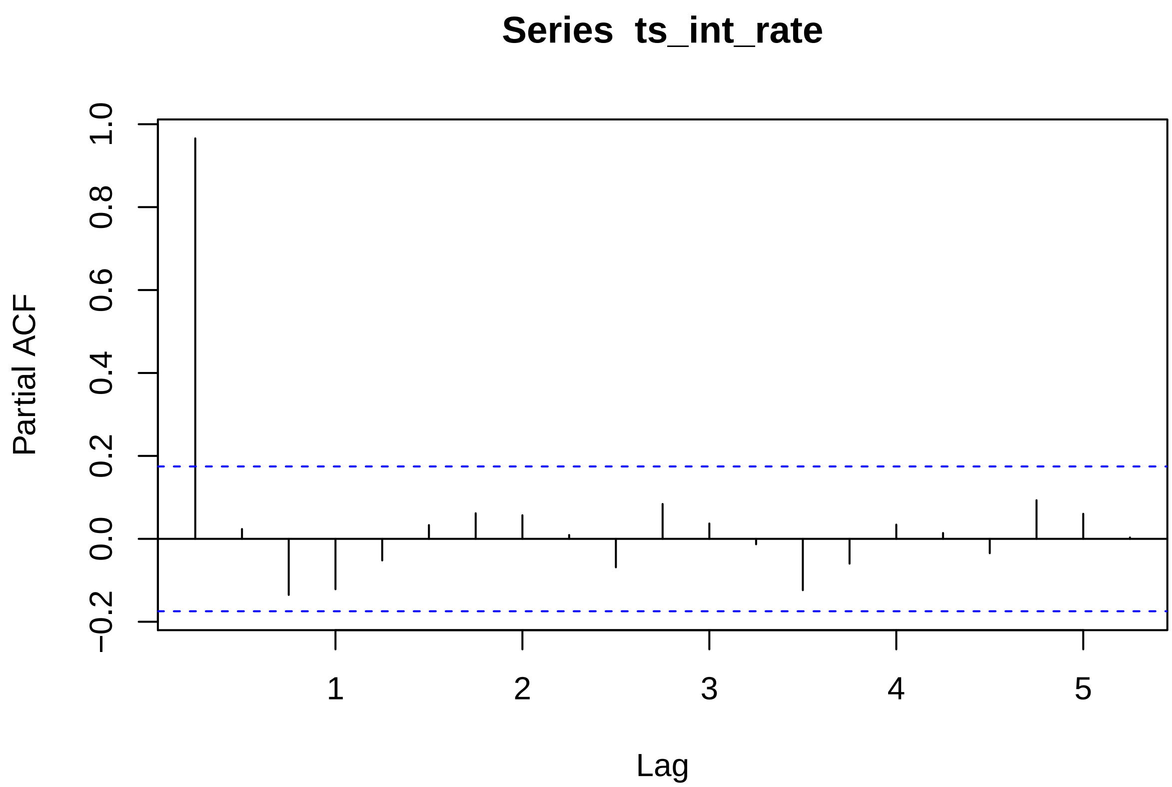


similarly, there is a great autocorrelation dependence within the time series, according to acf function. However, like the pacf for ts_prices, ts_int_rate, lack in time dependence when using pacf.

```
acf(ts_int_rate)
```

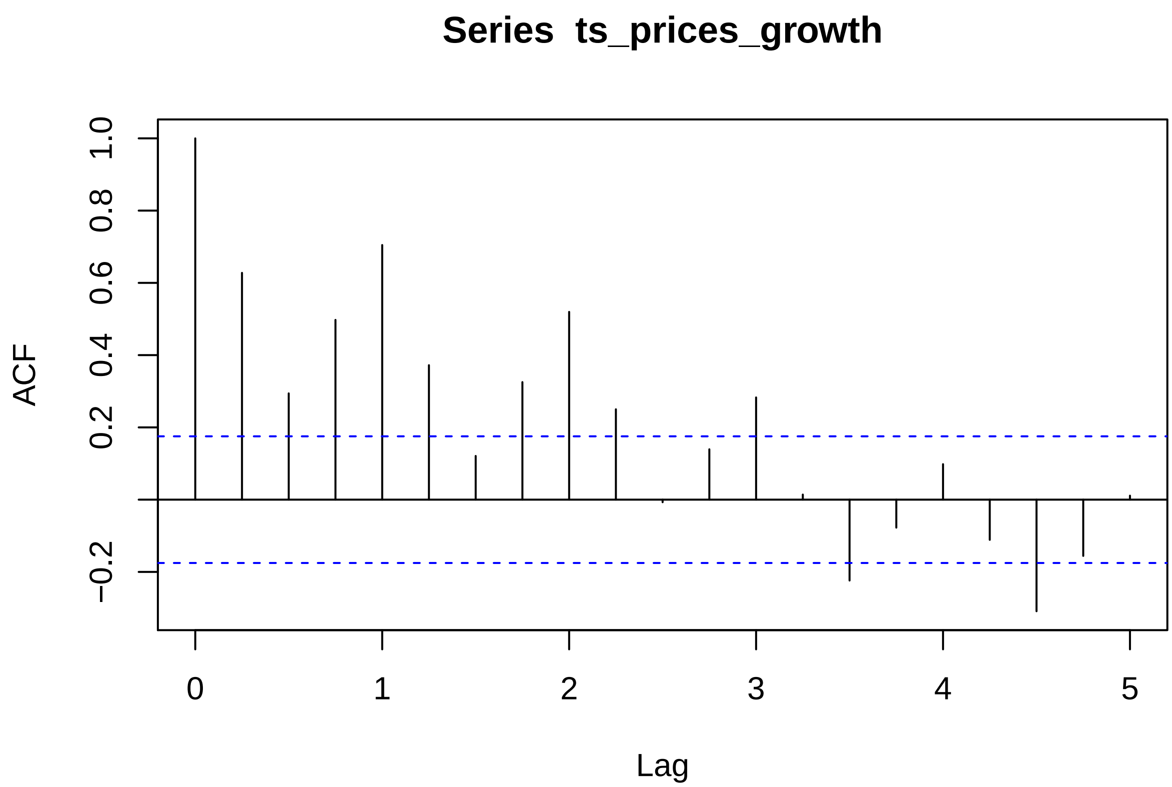



```
pacf(ts_int_rate)
```

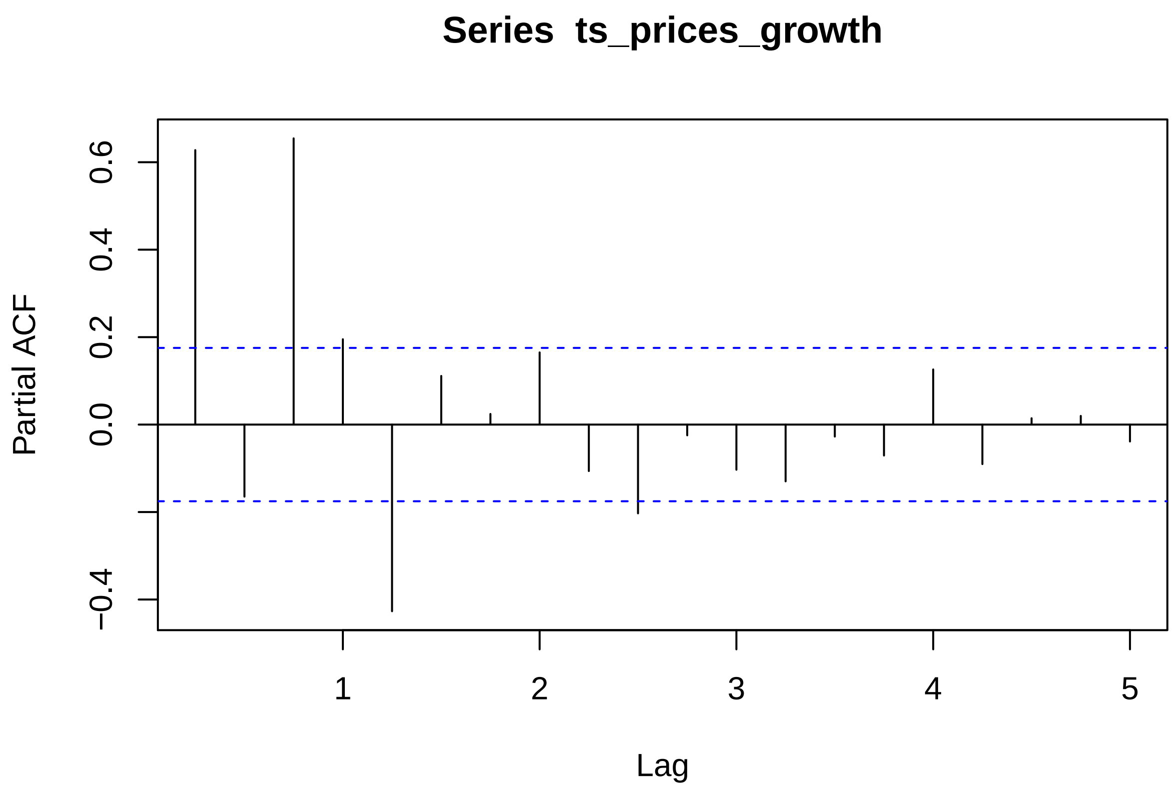


After the intial 3 lags, we begin to see a lack in time dependence for price growth according to the acf function. Likewise in the previous examples, we see that pacf also indicates no significant time dependence.

```
acf(ts_prices_growth)
```



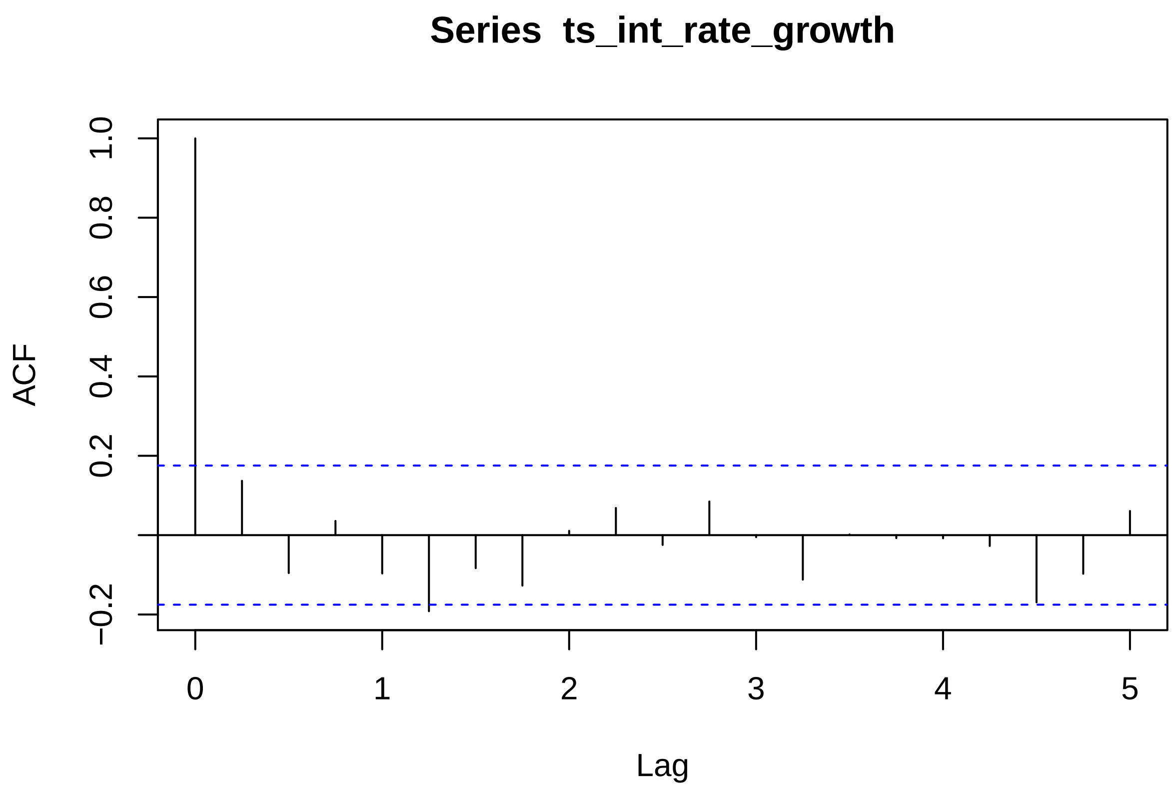

```
pacf(ts_prices_growth)
```



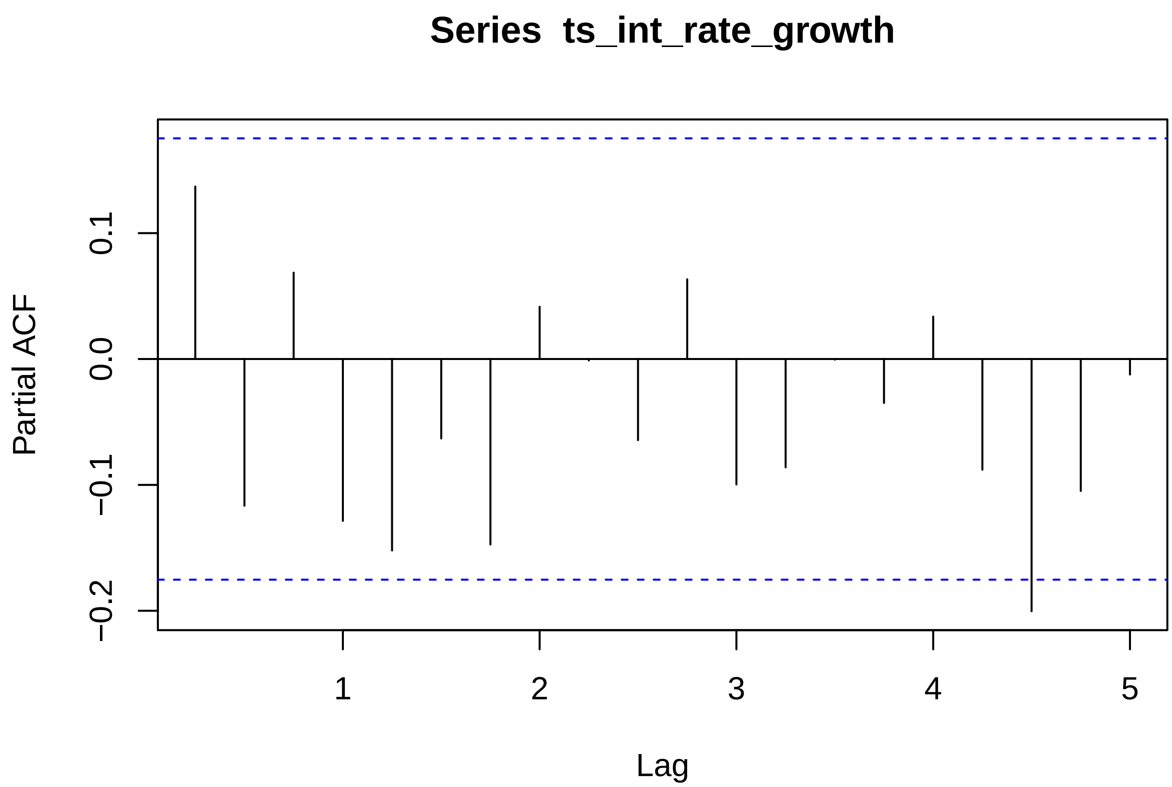
Interest rate growth shows a lack in time dependence according to acf and pacf.

According to the acf and pacf functions, it appears as though there is a greater time dependence with prices and interest rates as opposed to their growth rate counterparts. I would not intuitively expect this answer.

```
acf(ts_int_rate_growth)
```




```
pacf(ts_int_rate_growth)
```

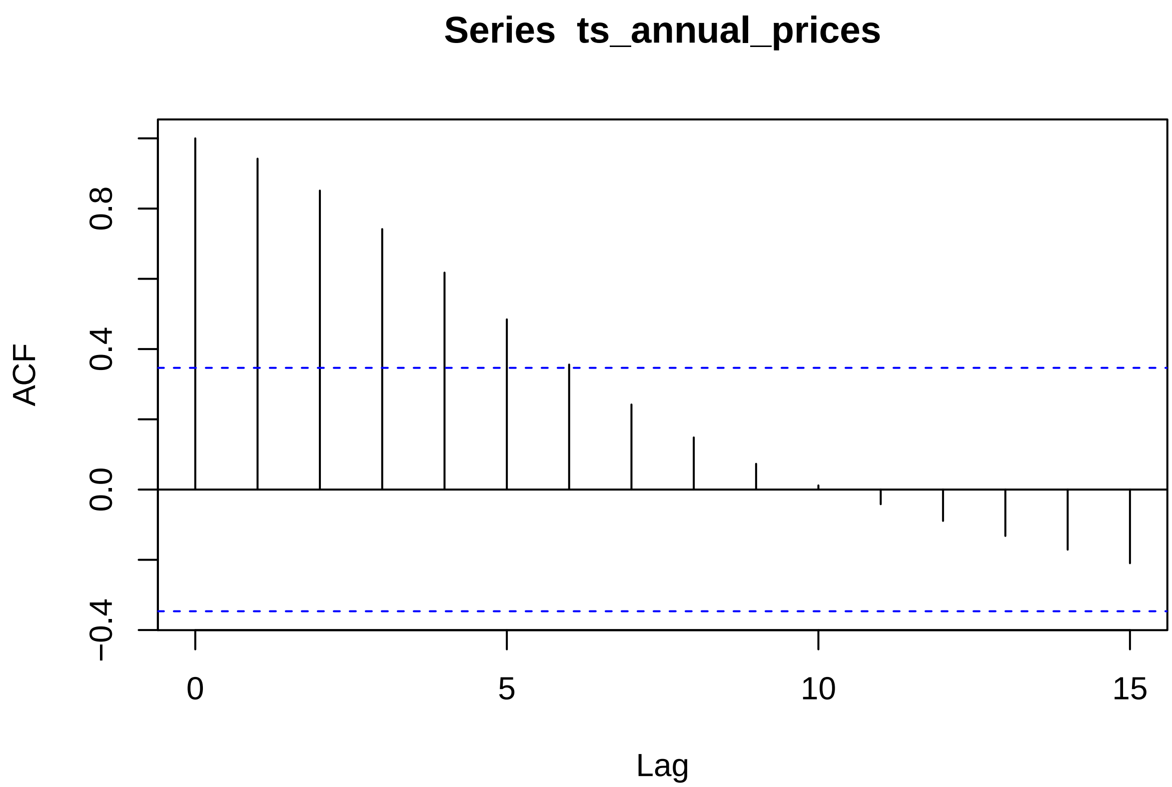


Annual Data Used

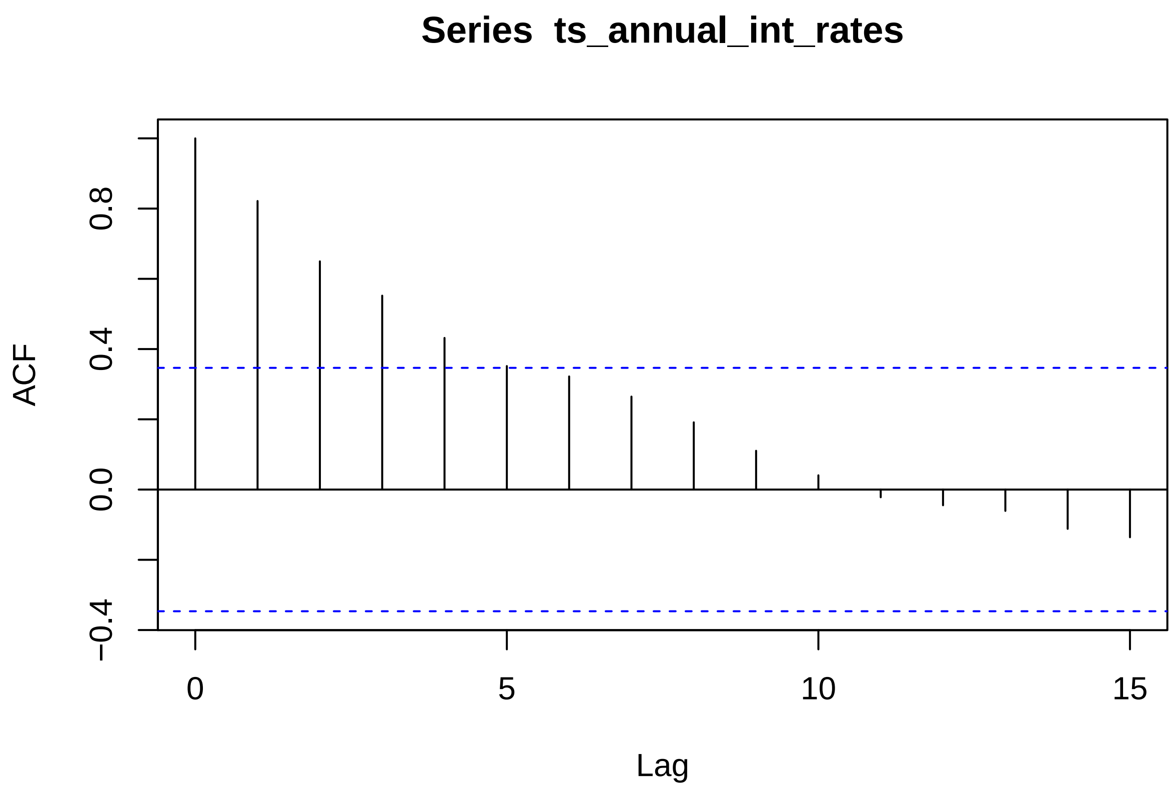
Now, we look at annual data instead of monthly and load in the time series.

```
data3 <- read.table(file = 'C:\\Users\\Austin\\Documents\\R\\HW143pt2.csv', header = T, sep = ",")
ts_annual_prices <- ts(data3$P[7:38], start = 1980, freq = 1)
ts_annual_int_rates <- ts(data3$R..in...[7:38], 1980, freq =1)
ts_annual_prices_growth <- diff(log(ts_annual_prices), lag =1)
ts_annual_int_rates_growth <- diff(log(ts_annual_int_rates), lag =1)
```

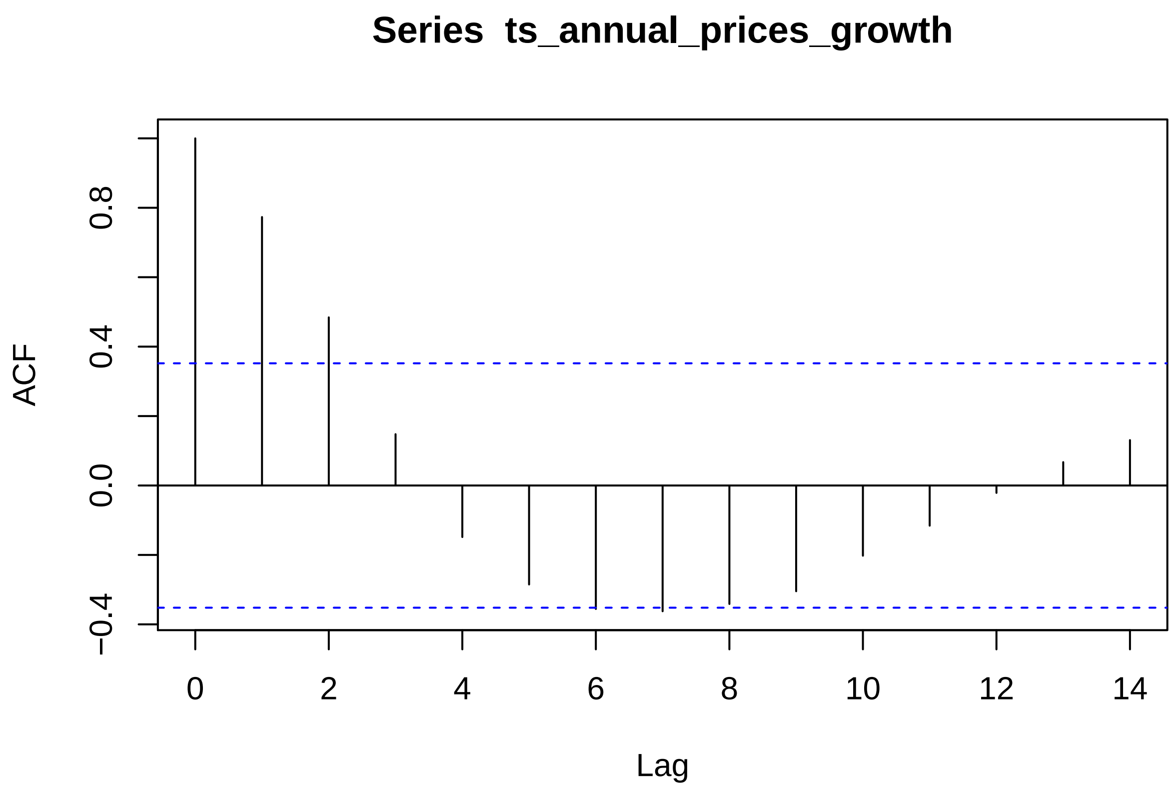
```
acf(ts_annual_prices,na.action = na.pass)
```



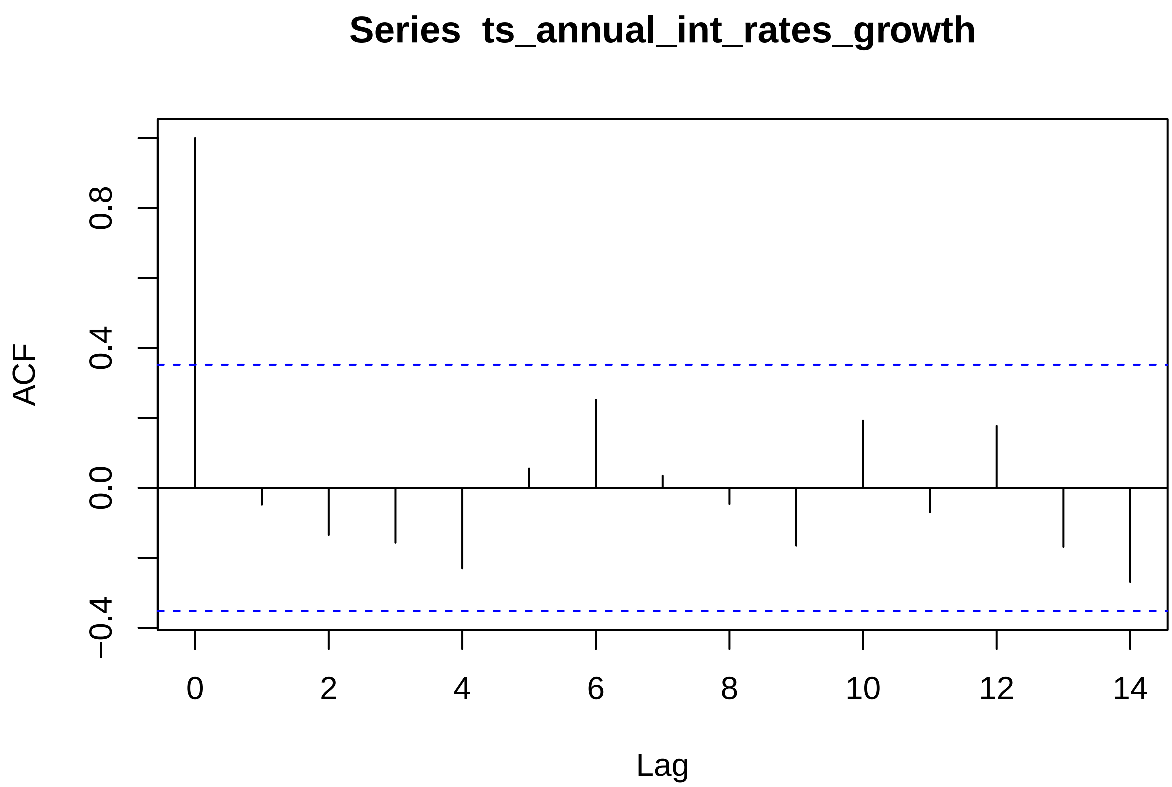

```
acf(ts_annual_int_rates,na.action = na.pass)
```




```r
acf(ts_annual_prices_growth,na.action = na.pass)
```




```r
acf(ts_annual_int_rates_growth,na.action = na.pass)
```



The time dependence is less pronounced because there is a greater time interval between monthly and annually. The lags in the annual years represent a larger time gap, so after the first 5 lags, there is less time dependence. For the quarterly lags, there is more time dependence because of the shorter time span, so the acf shows time dependence for all lags for prices and interest rates after 5 lags.