Time Series Lab 3

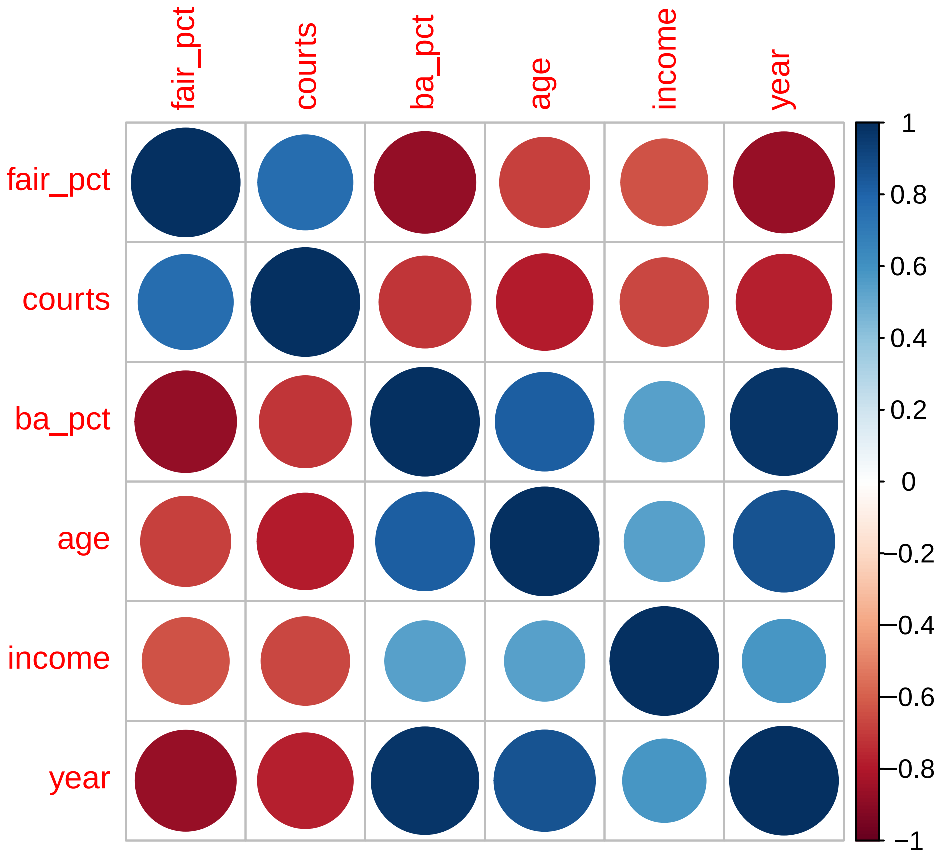
Nathan Colbert

12/7/2017

1. Create a multivariate time series; perform any interpolations.

```
trends_gss <- read_csv("/Users/nathancolbert/Downloads/trends-gss 3.csv")

mark <- colSums(is.na(trends_gss) == 0)

mark <- sapply(trends_gss, FUN = is.numeric)

trends_gss <- trends_gss[, mark]

vars <- c("year", "courts", "fair", "sex",
          "age", "partyid", "degree", "realinc")
sub <- trends_gss[, vars]

sub <- mutate(sub,
              baplus = ifelse(degree >= 3, 1, 0),
              courts = ifelse(courts == 2, 1, 0),
              optimism = ifelse(fair == 2, 1, 0),
              income = realinc)

by_year <- aggregate(subset(sub, sel = -year),
                     list(year = sub$year), mean, na.rm = T)

by_year[30:40, "year"] <- c(1979, 1981, 1992,
                            1995, seq(1997, 2009, 2))

by_year <- arrange(by_year, year)

by_year_ts <- ts(by_year)

by_year_ts <- na.approx(by_year_ts)

by_year <- aggregate(subset(sub, sel = -year),
                     list(year = sub$year), mean, na.rm = T)

by_year_ts <- as.data.frame(by_year_ts)

by_year_ts <- mutate(by_year_ts,
                     courts = courts*100,
                     fair_pct = optimism*100,
                     ba_pct = baplus*100)

cor_vars <- c("fair_pct", "courts", "ba_pct",
              "age", "income", "year")

cor_data <- by_year_ts[, cor_vars]
```

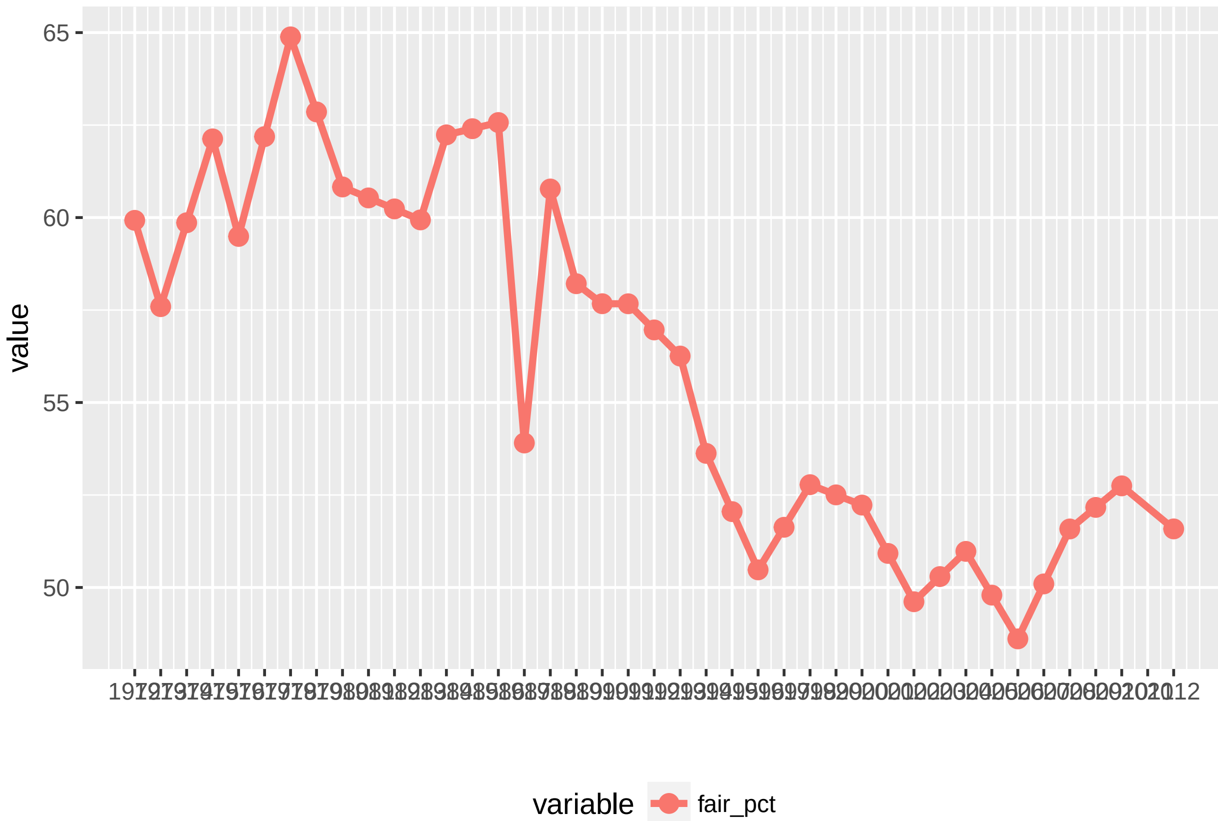
```r
corrplot(cor(cor_data))
```



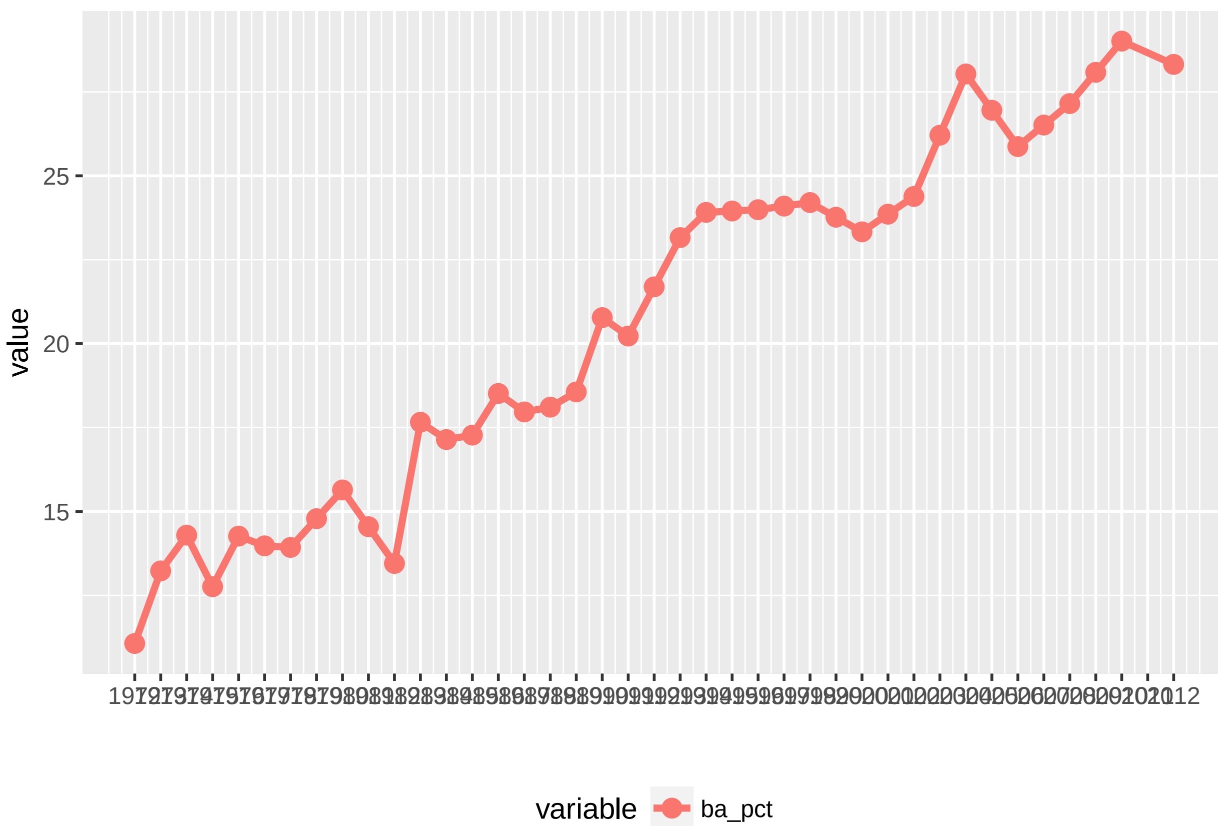
2. Graph the relationships between X and Y. Explain how you think Y should relate to your key Xs.

I would assume those who believe people are generally fair would er on the side of the courts either punishing appropriately, or even too much. However, those who believe that it is the nature of others to take advantage of one another would likely find the court system too lax. I would assume both of these relationships would move in harmony across time and that an increase in belief that the courts should be more stringent would beget an increase in people who believe that others govern themselves selfishly.

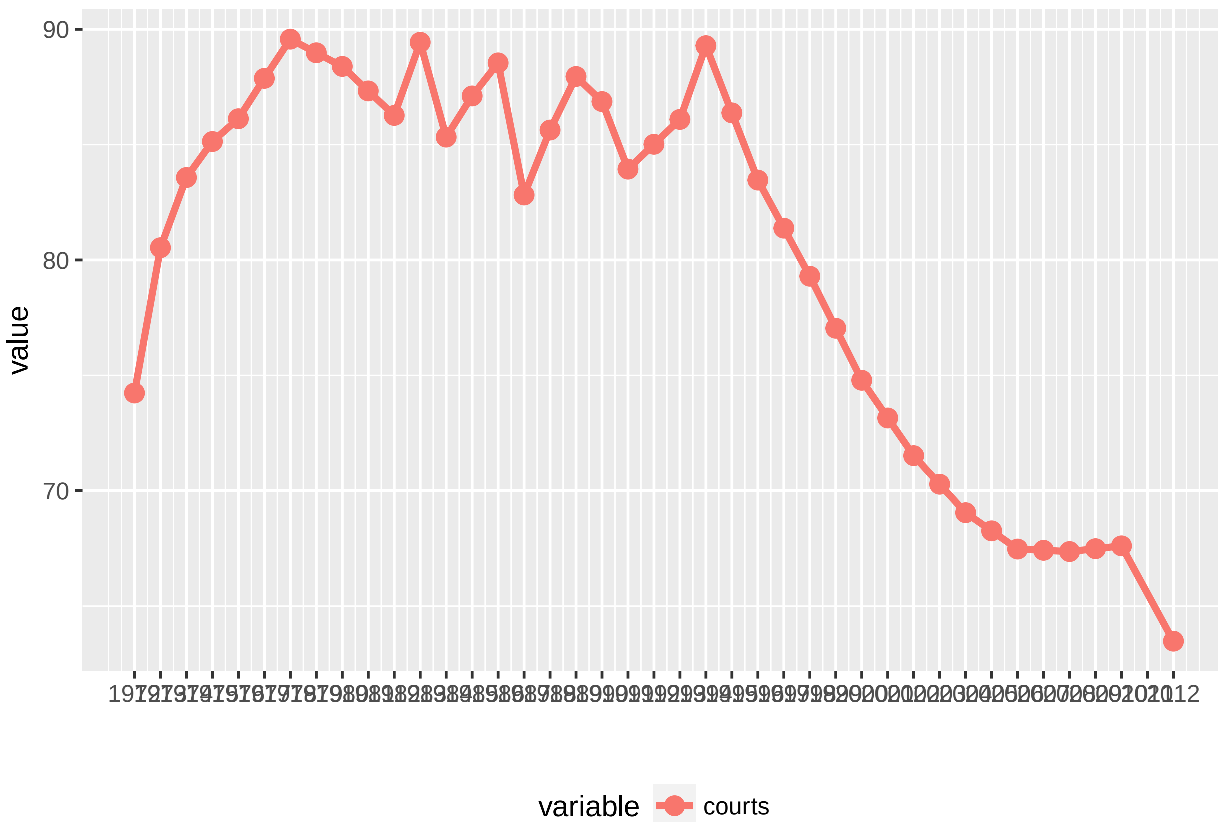
```r
keep_vars <- c("year", "fair_pct", "courts",
               "age", "ba_pct", "income")

plot_dat <- meltMyTS(mv.ts.object = by_year_ts,
                     time.var = "year", keep.vars = keep_vars)

(g_trust <- ggMyTS(df = plot_dat, varlist = c("fair_pct")))
```



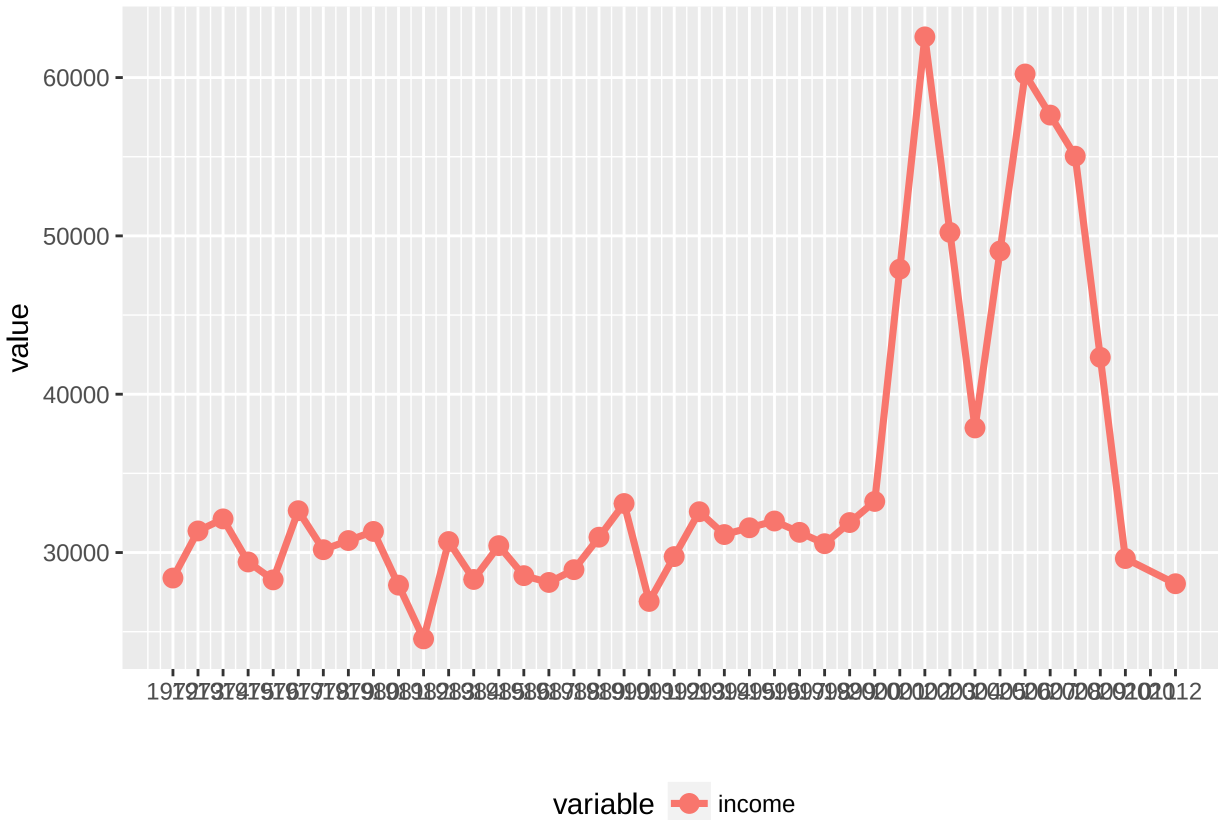

```
(g_degreelt50_pct <- ggMyTS(df = plot_dat, varlist = c("ba_pct")))
```



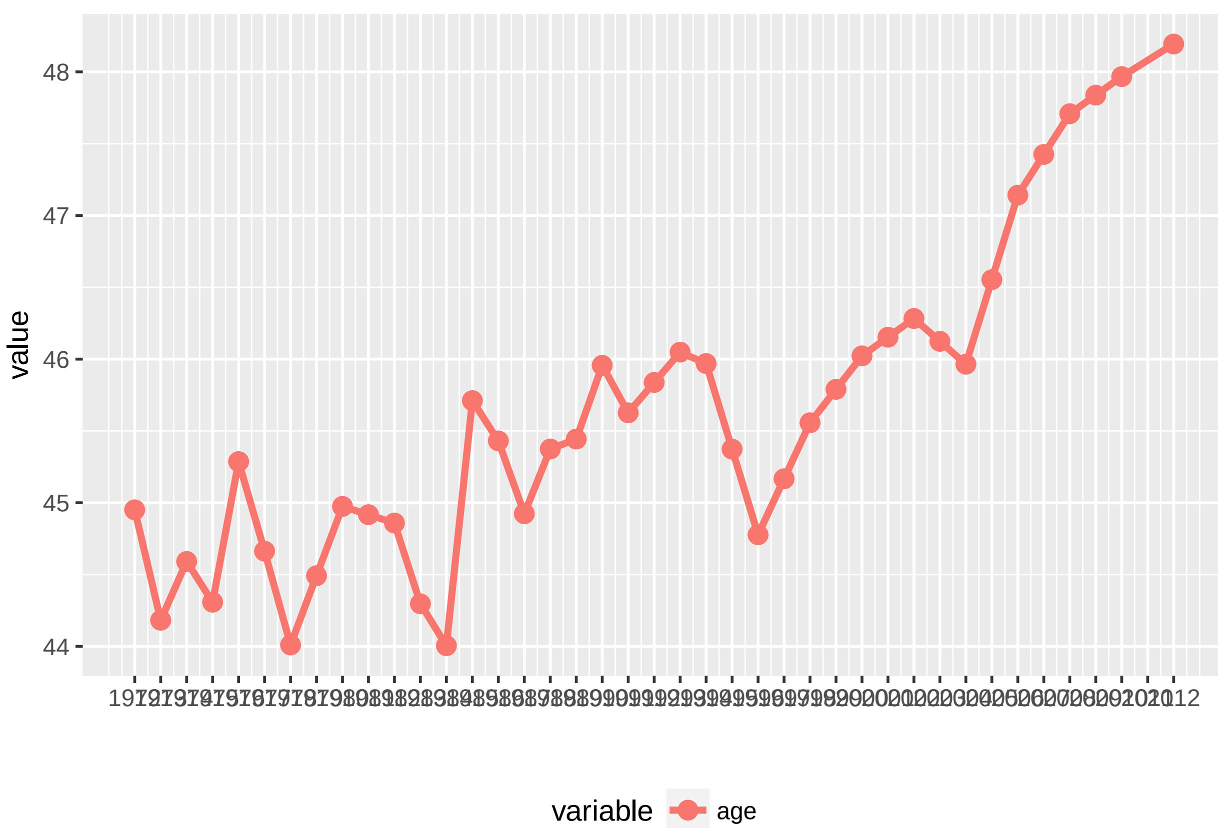
```r
(g_degreelt50_pct <- ggMyTS(df = plot_dat, varlist = c("courts")))
```




```r
(g_degreelt50_pct <- ggMyTS(df = plot_dat, varlist = c("income")))
```




```r
(g_age <- ggMyTS(df = plot_dat, varlist = c("age")))
```

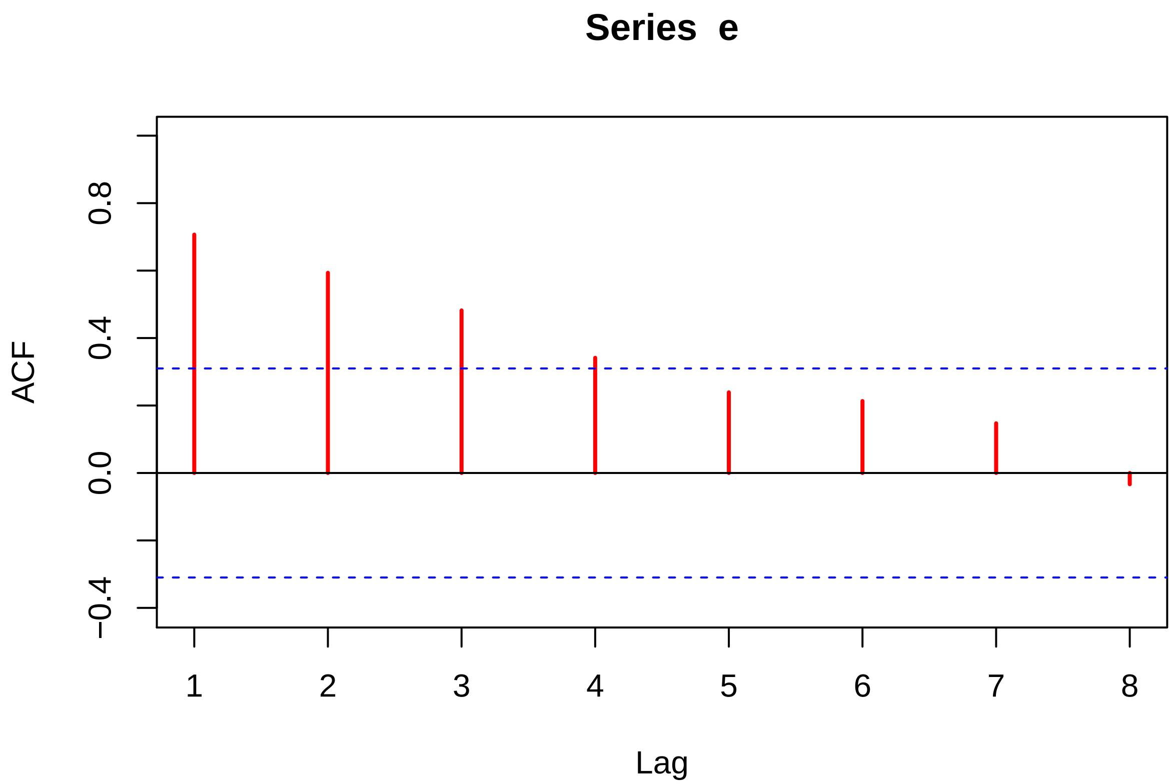


3. Run a simple time series regression, with one X and no trend. Interpret it.

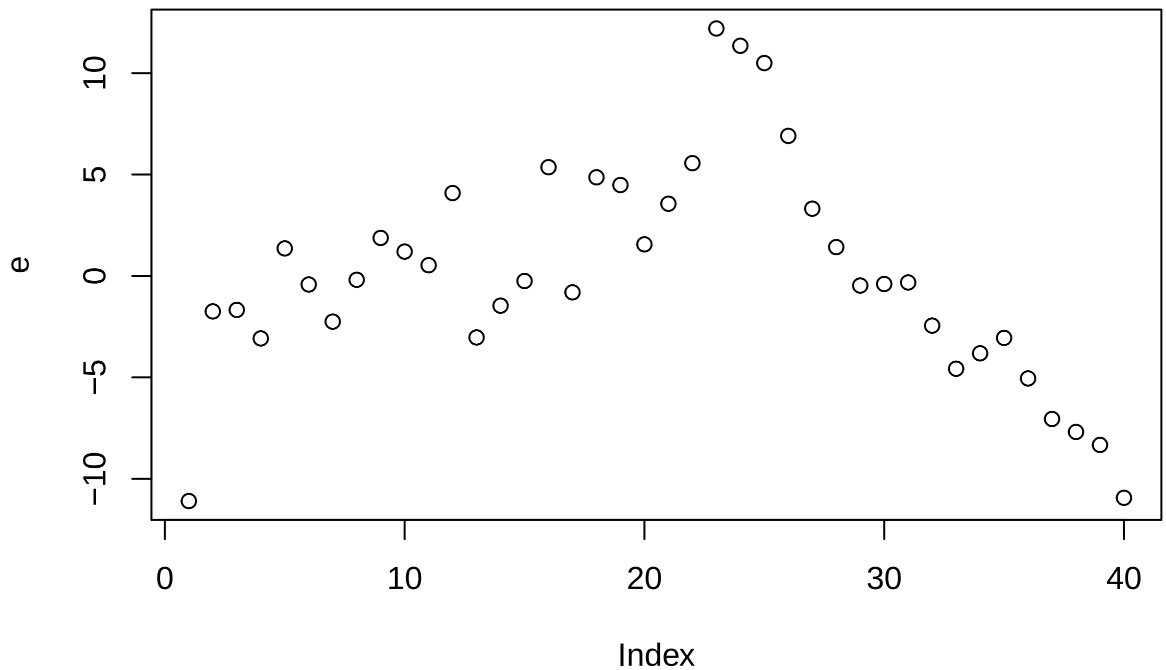
```r
lm_courts <- lm(courts ~ fair_pct, data = by_year_ts)

summary(lm_courts)
```

```
##
## Call:
## lm(formula = courts ~ fair_pct, data = by_year_ts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0973  -3.0360  -0.3587   3.3775  12.2012
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.9038    10.1347   0.681      0.5
## fair_pct      1.3088     0.1801   7.267 1.07e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.432 on 38 degrees of freedom
## Multiple R-squared:  0.5815, Adjusted R-squared:  0.5705
## F-statistic: 52.81 on 1 and 38 DF,  p-value: 1.074e-08
```

```r
e <- lm_courts$resid

acf(e, xlim = c(1,8), col = "red", lwd = 2)
```




```r
plot(e)
```




```r
dwtest(lm_courts)
```

```
##
##  Durbin-Watson test
##
## data:  lm_courts
```

```
## DW = 0.37011, p-value = 3.062e-12
## alternative hypothesis: true autocorrelation is greater than 0
```

```r
bgtest(lm_courts)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  lm_courts
## LM test = 22.431, df = 1, p-value = 2.178e-06
```

```r
durbinWatsonTest(lm_courts, max.lag=3)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.7067052     0.3701087       0
##    2       0.5935024     0.5319192       0
##    3       0.4820335     0.6996272       0
##  Alternative hypothesis: rho[lag] != 0
```

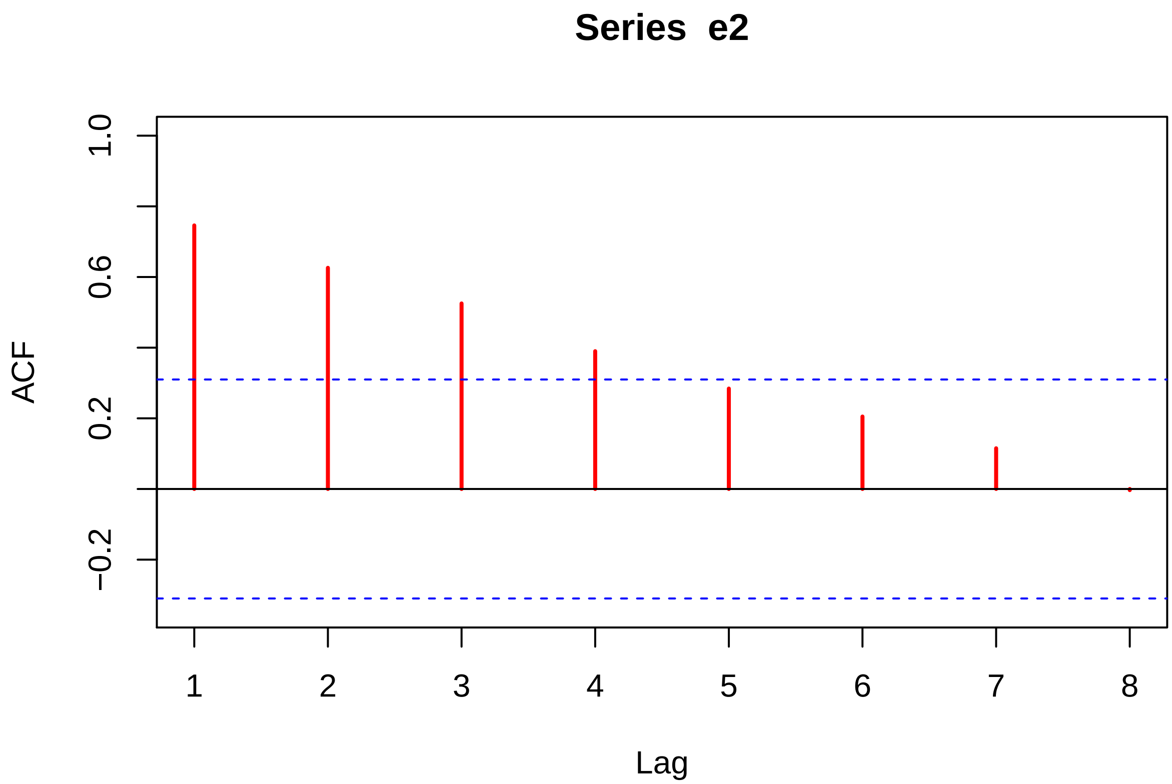
Each percent more people increase on their opinions towards fairness is a 1.3088%*** increase in their belief that courts are not strict enough. This is opposite of what I thought would happen in every way. Looking at the Durbin Watson Test and the Breusch-Godfrey Test, we definitely have autocorrelation in the errors. Including the year trend will potentially correct this.

4. Run a time series regression with one X and trend. Interpret it. Perform autocorrelation diagnostics. Explain what you found.

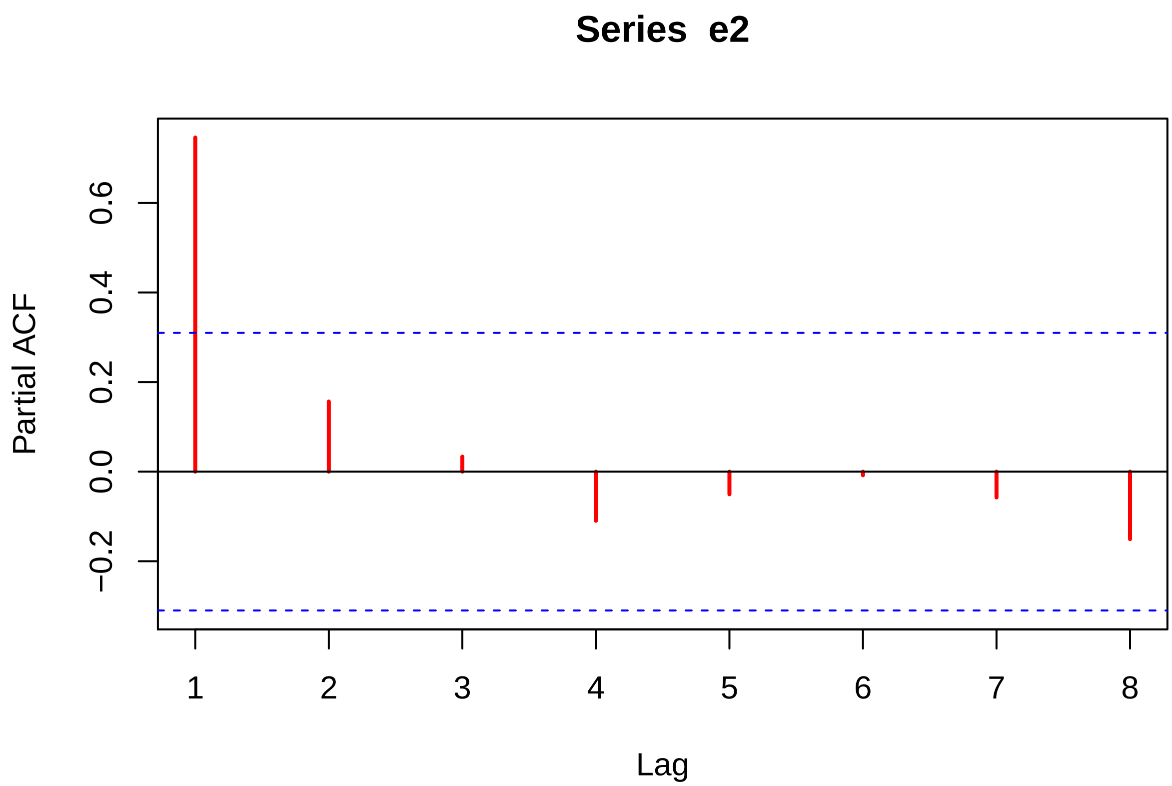
```r
lm_courts2 <- update(lm_courts, ~ . + year)

summary(lm_courts2)
```

```
##
## Call:
## lm(formula = courts ~ fair_pct + year, data = by_year_ts)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.7229  -3.7819  -0.0308   2.6676  11.3641
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 674.6365   292.8817   2.303   0.0270 *
## fair_pct      0.6457     0.3372   1.915   0.0633 .
## year         -0.3166     0.1388  -2.281   0.0284 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.155 on 37 degrees of freedom
## Multiple R-squared:  0.6331, Adjusted R-squared:  0.6133
## F-statistic: 31.93 on 2 and 37 DF,  p-value: 8.782e-09
```

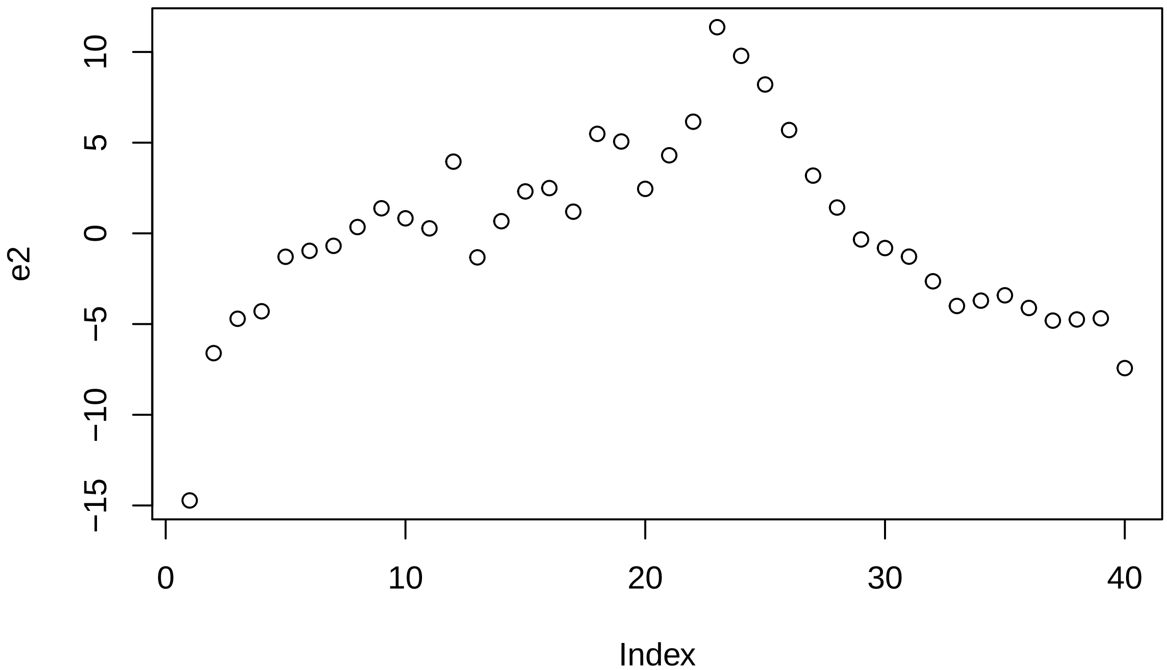
```r
e2 <- lm_courts2$resid

acf(e2, xlim = c(1,8), col = "red", lwd = 2)
```




```
pacf(e2, xlim = c(1,8), col = "red", lwd = 2)
```




```
plot(e2)
```




```r
dwtest(lm_courts2)
```

```
##
##  Durbin-Watson test
##
## data:  lm_courts2
## DW = 0.23136, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```r
bgtest(lm_courts2)
```

```
##
##  Breusch-Godfrey test for serial correlation of order up to 1
##
## data:  lm_courts2
## LM test = 23.741, df = 1, p-value = 1.102e-06
```

```r
durbinWatsonTest(lm_courts2, max.lag=3)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.7460180     0.2313558       0
##    2       0.6259914     0.4047513       0
##    3       0.5251718     0.5609273       0
##  Alternative hypothesis: rho[lag] != 0
```

Through including trend, this relationship all but disappears. Now, for every 1% increase in belief in fairness we see a .64% increase in belief that the courts should be more stringent. However, this is barely statistically significant. Furthermore, net of percenntage of change in belief in the fairness of others, with each passing year we are seeing a decrease by .3%* of peoples belief that courts should prosecute more harshly. However, we still have extremely correlated errors over multiple lags.

5. Consider running a time series regression with many Xs and trend. Interpret that.

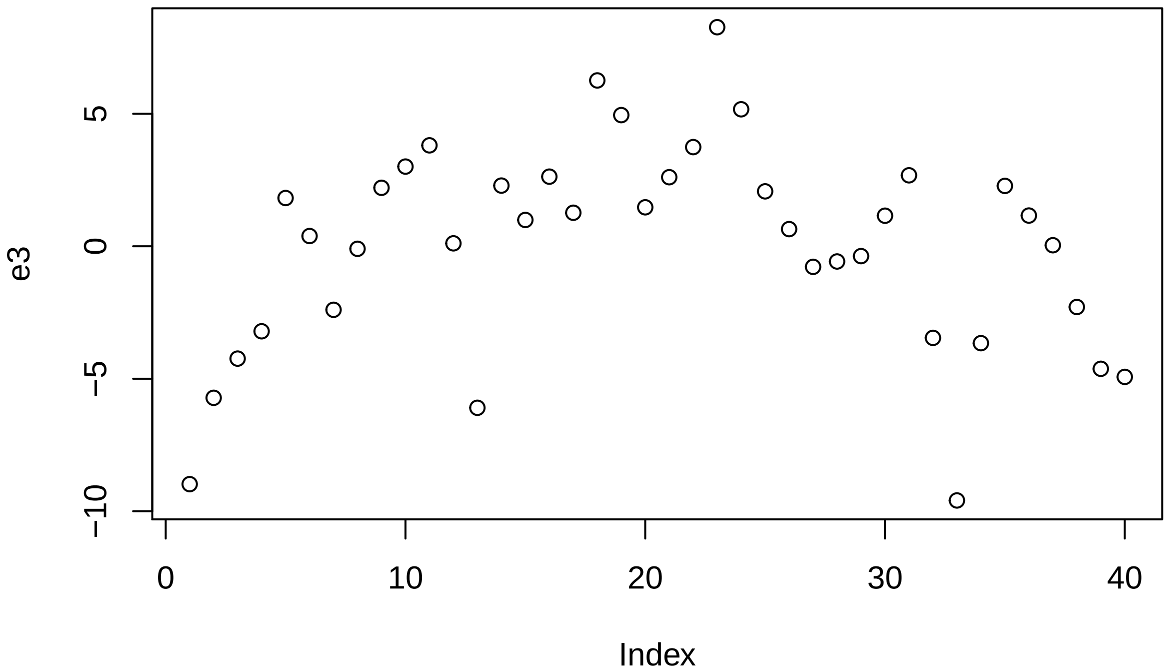
```r
options(scipen = 999)

lm_courts3 <- update(lm_courts2, ~ . + age + ba_pct + income)

summary(lm_courts3)
```

```
##
## Call:
## lm(formula = courts ~ fair_pct + year + age + ba_pct + income,
##     data = by_year_ts)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.5937 -2.6008  0.8204  2.3696  8.2684
##
## Coefficients:
##                  Estimate    Std. Error t value Pr(>|t|)
## (Intercept) 1252.21389447  639.94722751   1.957  0.05863 .
## fair_pct       0.90731072    0.31930767   2.841  0.00754 **
## year          -0.54497401    0.34153475  -1.596  0.11982
## age           -3.60613319    1.31271531  -2.747  0.00955 **
## ba_pct         1.51968070    0.67556691   2.249  0.03107 *
## income        -0.00011998    0.00009332  -1.286  0.20726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.208 on 34 degrees of freedom
## Multiple R-squared:  0.7753, Adjusted R-squared:  0.7423
## F-statistic: 23.47 on 5 and 34 DF,  p-value: 0.0000000003884
```

```r
durbinWatsonTest(lm_courts3, max.lag=2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.5485593     0.7286292   0.000
##    2       0.1631980     1.4094823   0.048
##  Alternative hypothesis: rho[lag] != 0
```

```r
e3 <- lm_courts3$residuals

plot(e3)
```



Including more variables we see that, net of all other variables, the previously perceived relationship between belief in fairness and harsher courts continues its statistical significance has returned with a 1% increase in fairness leading to a .91** increase in desiring harsher court sentences. Another interesting outcome is that having a bachelors degree or more, net of all other variables, leads to a 1.5% increase in your belief in a more strict justice system. I suppose in one light this makes sense, although in general more education tends to lead to greater overall tolerance in other areas, and thus would suggest the opposite relationship than is seen here. Per usual, with age comes a 3.6% decrease in desiring stricter court systems which is statistcally significant. This is in line with most other ideologies that with aging one may begin to digress in stringent veiws. Yet again it appears we have autocorellation in the errors. The graph still shows a high corellation.

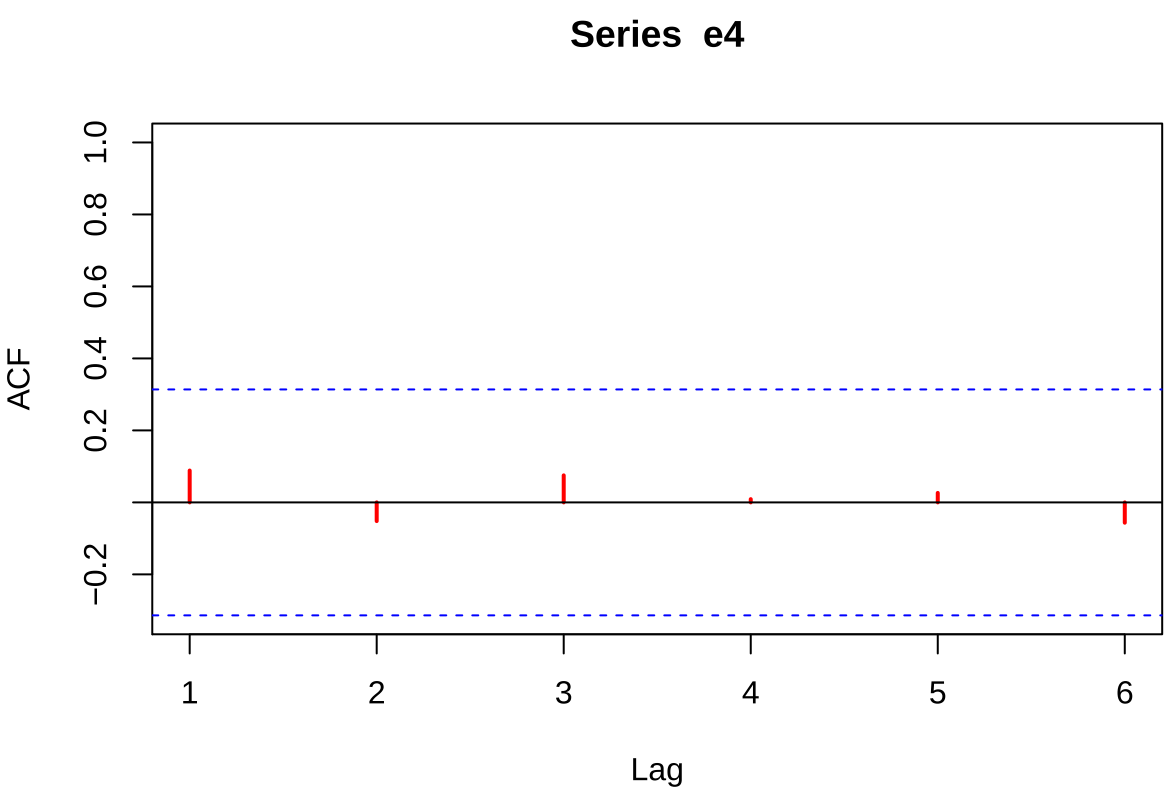
6. Run a first differenced time series regression. Interpret that.

```
by_yearFD <- summarise(data.frame(by_year_ts),
                       courts = firstD(courts),
                       age = firstD(age),
                       fair_pct = firstD(fair_pct),
                       ba_pct = firstD(ba_pct),
                       income = firstD(income),
                       year = year)

lm_courts4 <- update(lm_courts3, data = by_yearFD)

summary(lm_courts4)
```

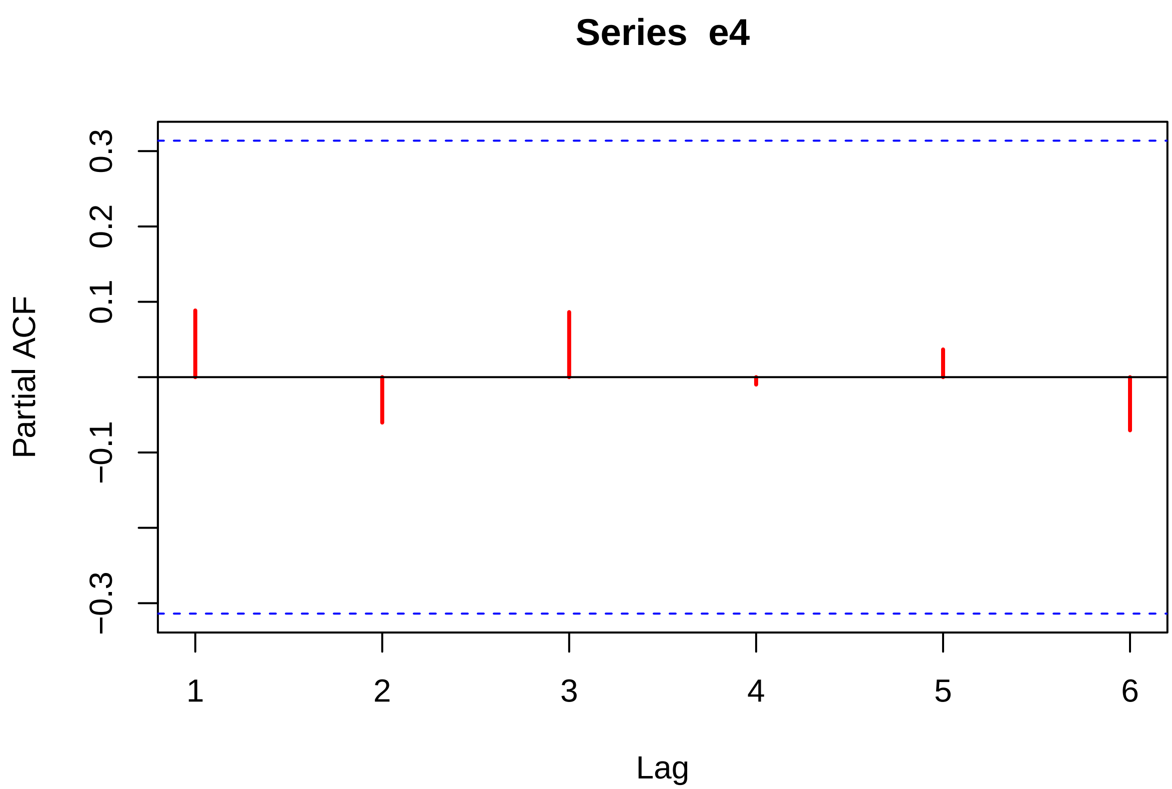
```
##
## Call:
## lm(formula = courts ~ fair_pct + year + age + ba_pct + income,
##     data = by_yearFD)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.1848 -1.1519  0.0182  0.8163  4.2890
##
## Coefficients:
##                Estimate   Std. Error t value Pr(>|t|)
## (Intercept) 167.08671822  53.64413702   3.115 0.003794 **
## fair_pct      0.33923798   0.13214265   2.567 0.014972 *
## year         -0.08420870   0.02693012  -3.127 0.003675 **
## age           0.35530842   0.63000042   0.564 0.576580
## ba_pct        0.96510996   0.26404339   3.655 0.000885 ***
## income        0.00005045   0.00005002   1.009 0.320467
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.846 on 33 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4918, Adjusted R-squared:  0.4148
## F-statistic: 6.387 on 5 and 33 DF,  p-value: 0.0002985
```

```r
e4 <- lm_courts4$resid

acf(e4, xlim = c(1,6), col = "red", lwd = 2)
```



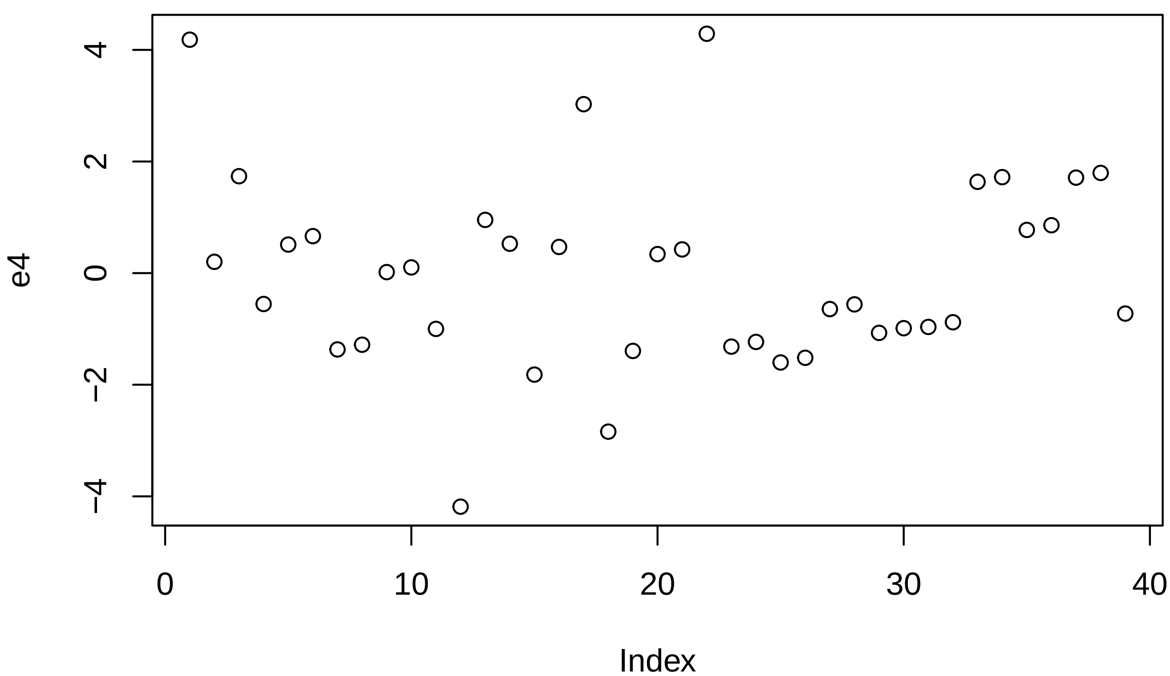

```r
pacf(e4, xlim = c(1,6), col = "red", lwd = 2)
```




```r
durbinWatsonTest(lm_courts4, max.lag = 2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.08849867      1.662819   0.262
##    2     -0.05197356      1.914748   0.936
##  Alternative hypothesis: rho[lag] != 0
```

```r
plot(e4)
```



Through examining the first difference model we see that a 1% positive change in percent greater than a bachelors, net of all other factors, leads to a .96%*** increase in belief that the courts should be more stringent that is highly statistically significant. Year and fair_pct remain almost the same although diminished, and age and income both switch signs and are not significant. This suggests the previous hypothesis about age is likely invalid in this case. Although we still have a bit of autocorellation, it has been reduced significantly, and the errors are begininning to disperse a bit.

7. Check your variables for unit roots. Do some tests. Interpret them.

```r
adfTest(by_year_ts[,"courts"], lags = 0, type="ct")
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 0
##   STATISTIC:
##     Dickey-Fuller: -3.6125
##   P VALUE:
##     0.04468
##
## Description:
##  Mon Dec 11 21:28:51 2017 by user:
```

```r
adfTest(by_year_ts[,"courts"], lags = 4, type="ct")
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 4
##   STATISTIC:
##     Dickey-Fuller: -1.9111
##   P VALUE:
##     0.6084
##
## Description:
##  Mon Dec 11 21:28:51 2017 by user:
```

```r
PP.test(by_year_ts[,"courts"],lshort=TRUE)
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  by_year_ts[, "courts"]
## Dickey-Fuller = -3.7135, Truncation lag parameter = 3, p-value =
## 0.03684
```

According to the augmented Dickey Fuller Test with 0 lags, we can accept the null of not having unit roots, but just barely. Examining the same test with 4 lags, we jump the barrier and due to a .6 p value, we may have a unit root. However, the Phillips-Perron test suggests again that we may not have unit roots.

8. Perform an Automatic ARIMA on the residuals from one of your earlier models. Tell me what it says.

```
auto.arima(e4, trace=TRUE)
```

```
##
##  ARIMA(2,0,2) with non-zero mean : 165.6379
##  ARIMA(0,0,0) with non-zero mean : 156.3273
##  ARIMA(1,0,0) with non-zero mean : 158.324
##  ARIMA(0,0,1) with non-zero mean : 158.2573
##  ARIMA(0,0,0) with zero mean     : 154.1021
##  ARIMA(1,0,1) with non-zero mean : 160.1903
##
##  Best model: ARIMA(0,0,0) with zero mean
```

```
## Series: e4
## ARIMA(0,0,0) with zero mean
##
## sigma^2 estimated as 2.885:  log likelihood=-76
## AIC=153.99   AICc=154.1   BIC=155.66
```

According to auto ARIMA the best model is ARIMA (0, 0, 0) with zero mean. In other words, OLS is more than sufficient.

9. Run an ARIMA that follows from Step 8. Interpret that, too.

```
xvars_fat <- by_year_ts[,c("courts", "year")]

arima_000 <- arima(by_year_ts[,c("fair_pct")],
                   order = c(0,0,0), xreg = xvars_fat)
summary(arima_000)
```

```
##
## Call:
## arima(x = by_year_ts[, c("fair_pct")], order = c(0, 0, 0), xreg = xvars_fat)
##
## Coefficients:
##       intercept  courts     year
##        599.7787  0.1396  -0.2786
## s.e.   103.0853  0.0701   0.0495
##
## sigma^2 estimated as 5.315:  log likelihood = -90.17,  aic = 188.34
##
## Training set error measures:
##                                 ME     RMSE      MAE        MPE     MAPE
## Training set -0.0000000000006412648 2.305534 1.853457 -0.1645617 3.337243
##                  MASE      ACF1
## Training set 1.151077 0.510369
```

```r
Box.test(resid(arima_000), lag = 20,
         type = c("Ljung-Box"), fitdf = 0)
```

```
##
##  Box-Ljung test
##
## data:  resid(arima_000)
## X-squared = 71.034, df = 20, p-value = 0.0000001234
```

Net of time trend, a 1% increase in belief in inherently fair people leads to a .13% increase in belief in more severe courts. However, the Ljung-Box test suggests we do still have autocorellation in the errors. More to be done on this!