hw4_yw3204

wyh

10/13/2018

1.

```
library(plyr)
debt <- read.csv("debt.csv", as.is = TRUE)
dim(debt)
```

```
## [1] 1171    4
```

```
head(debt)
```

```
##     Country Year    growth     ratio
## 1 Australia 1946 -3.557951 190.41908
## 2 Australia 1947  2.459475 177.32137
## 3 Australia 1948  6.437534 148.92981
## 4 Australia 1949  6.611994 125.82870
## 5 Australia 1950  6.920201 109.80940
## 6 Australia 1951  4.272612  87.09448
```

```
# a
# calculate the mean of the growth column of a data frame
mean.growth <- function(df) {
  return (mean(df$growth))
}

# b
avg_g <- daply(debt, 'Country', mean.growth)
signif(avg_g, 3) # print
```

```
##   Australia     Austria     Belgium      Canada     Denmark     Finland
##        3.72        4.44        3.18        3.65        2.66        3.57
##      France     Germany      Greece     Ireland       Italy       Japan
##        3.78        3.31        2.93        3.93        3.25        4.45
## Netherlands New Zealand      Norway    Portugal       Spain      Sweden
##        3.03        3.07        3.83        4.00        3.20        3.07
##          UK          US
##        2.41        3.00
```

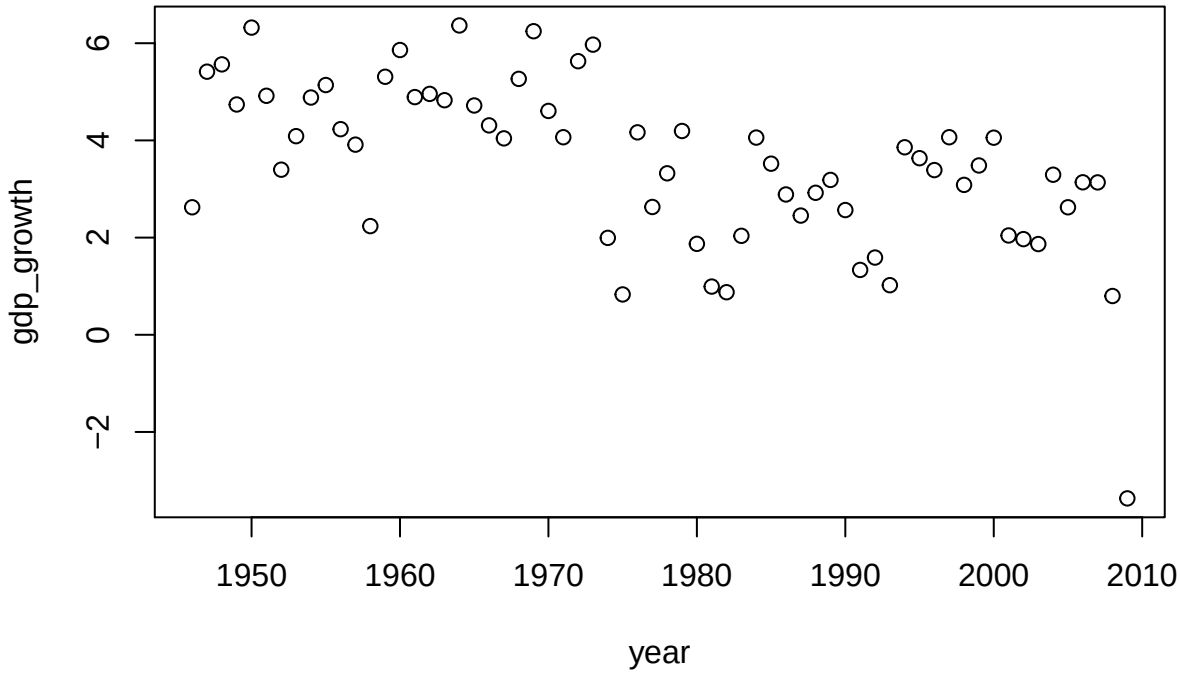
2.

```
avg_g1 <- daply(debt, 'Year', mean.growth)
signif(avg_g1, 3) # print
```

```
##   1946   1947   1948   1949   1950   1951   1952   1953   1954   1955
##  2.620  5.410  5.560  4.740  6.320  4.920  3.400  4.090  4.880  5.140
##   1956   1957   1958   1959   1960   1961   1962   1963   1964   1965
##  4.230  3.910  2.240  5.310  5.860  4.890  4.960  4.830  6.370  4.720
##   1966   1967   1968   1969   1970   1971   1972   1973   1974   1975
```

```
##  4.310  4.040  5.270  6.250  4.610  4.070  5.630  5.970  1.990  0.830
##   1976   1977   1978   1979   1980   1981   1982   1983   1984   1985
##  4.170  2.630  3.320  4.190  1.870  0.992  0.876  2.040  4.060  3.520
##   1986   1987   1988   1989   1990   1991   1992   1993   1994   1995
##  2.890  2.450  2.920  3.190  2.570  1.330  1.590  1.020  3.860  3.630
##   1996   1997   1998   1999   2000   2001   2002   2003   2004   2005
##  3.390  4.070  3.090  3.480  4.060  2.040  1.970  1.870  3.290  2.620
##   2006   2007   2008   2009
##  3.140  3.140  0.798 -3.370
```

```
# plot
plot(c(1946:2009), avg_g1, xlab = 'year', ylab = 'gdp_growth')
```

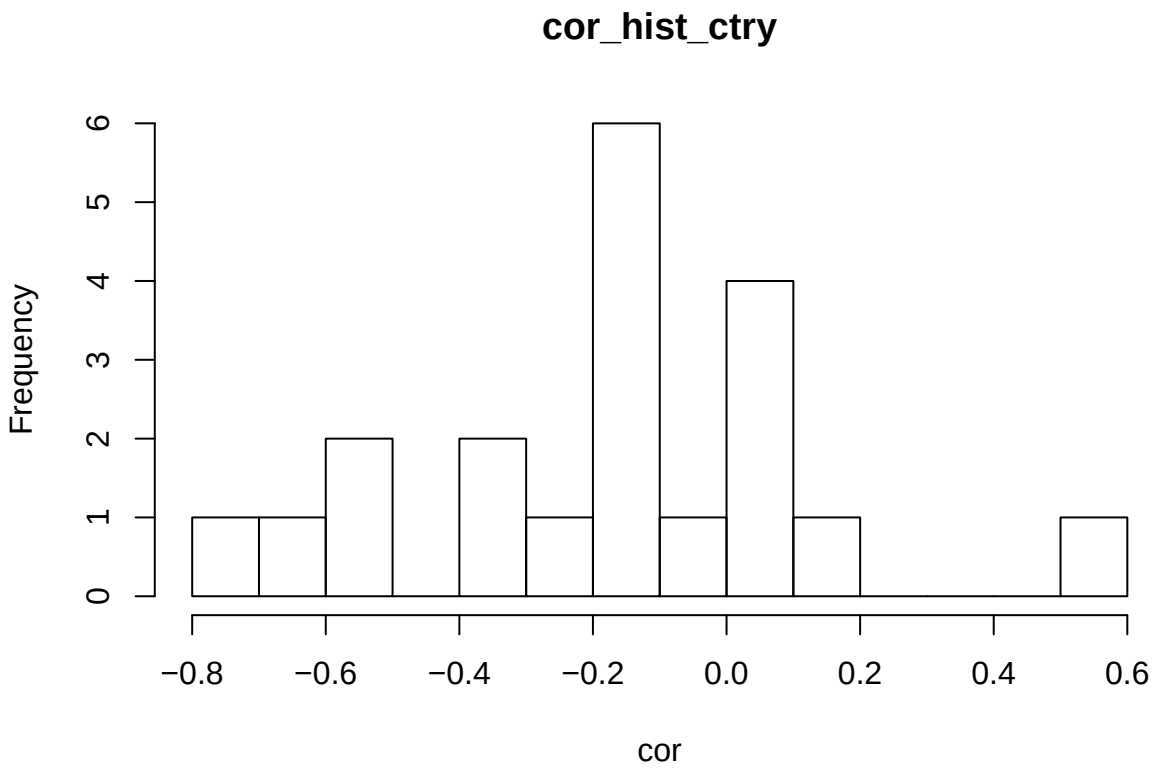


3.

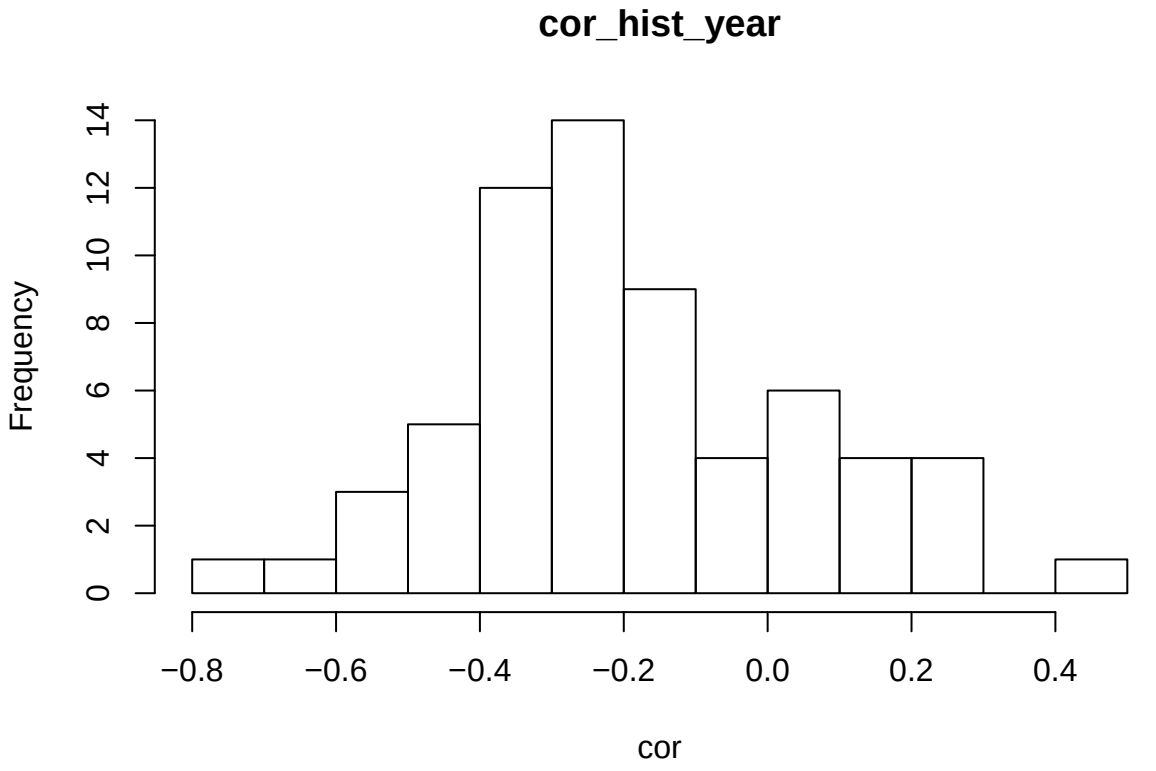
```
# a
c1 <- cor(debt$growth, debt$ratio)
signif(c1, 3)
```

```
## [1] -0.199
```

```
# b
# calculate the correlation between growth and ratio column in a data frame
cor_fun <- function(df) {
  return (cor(df$growth, df$ratio))
}

c2 <- daply(debt, 'Country', cor_fun)
#mean(c2) #check, if you may
hist(c2, breaks = 10, main = "cor_hist_ctry", xlab = "cor")
```




```r
# c
c3 <- daply(debt, 'Year', cor_fun)
#mean(c3) #check
hist(c3, breaks = 10, main = "cor_hist_year", xlab = "cor")
```



```
# d
```

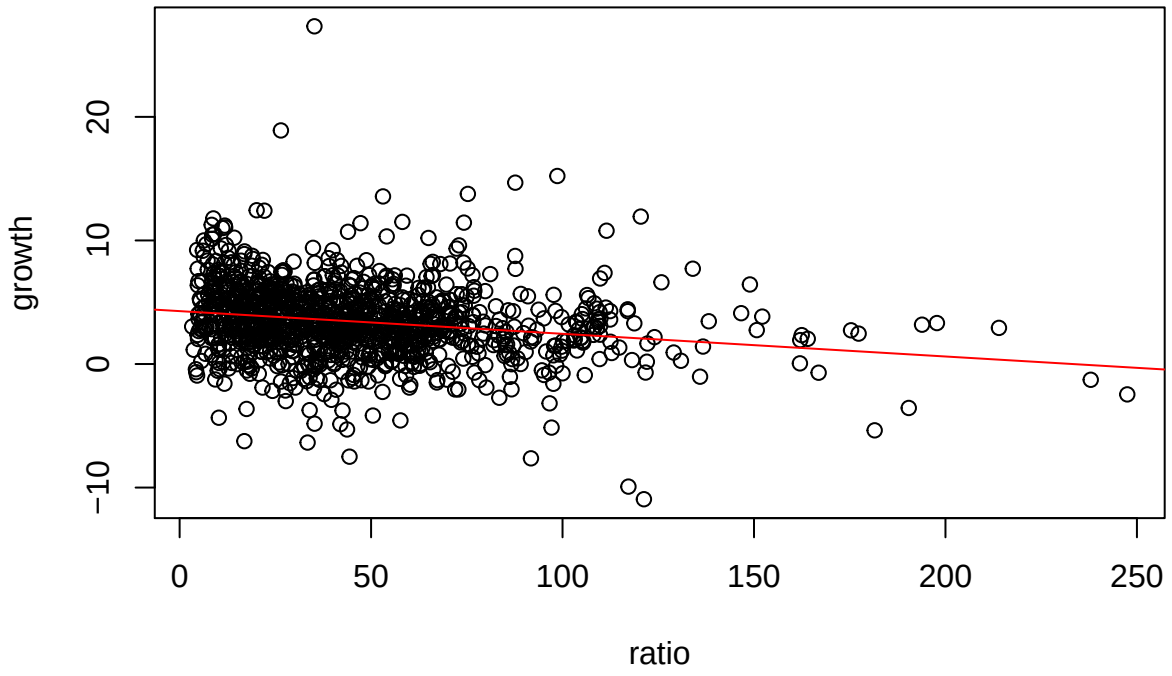
When splitting by country, we find that most country have negative correlation except Australia, Canada, New Zealand, Norway and Spain. When splitting by year, we also find most year have negative correlation but for year 1978, it has the highest positive corrrelation.

4.

```
l1 <- lm(growth ~ ratio, debt)
signif(l1$coefficients, 3)
```

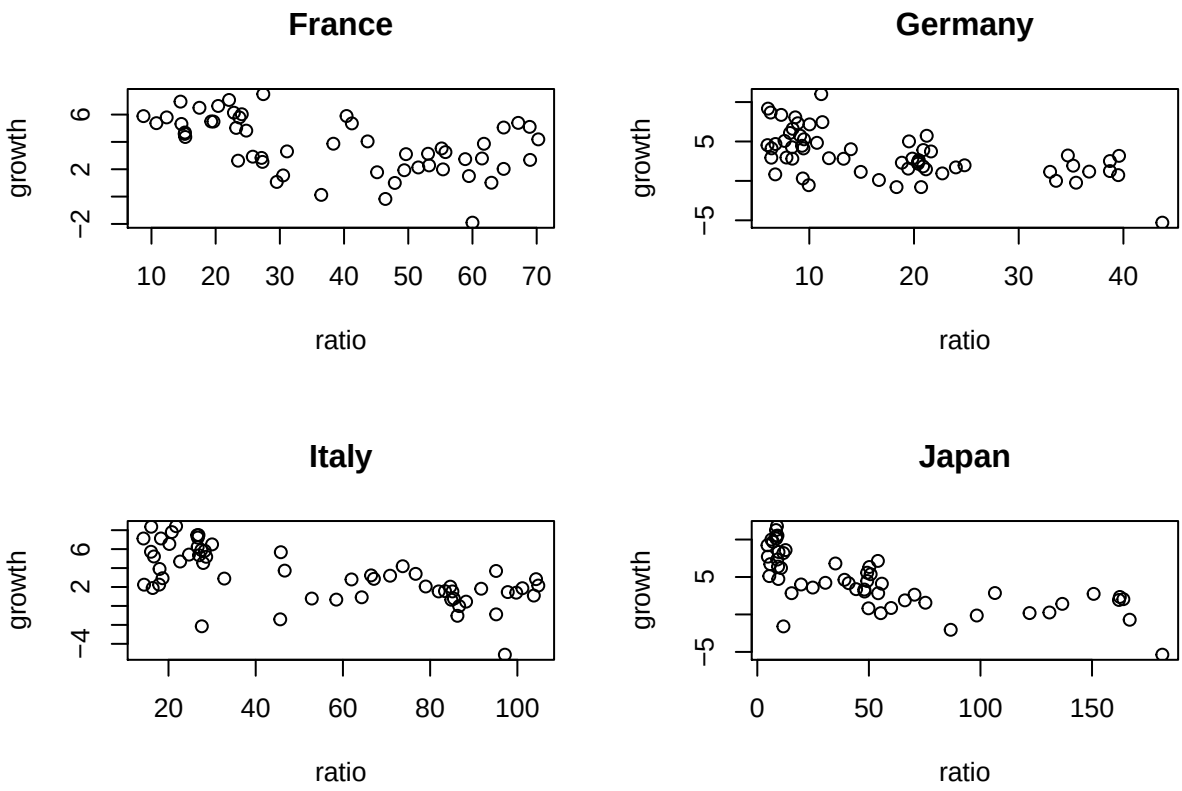
```
## (Intercept)       ratio
##      4.2800     -0.0184
```

```
plot(debt$ratio, debt$growth, xlab = "ratio", ylab = "growth")
abline(l1, col = "red")
```



5.

```
ctrs <- names(c2[c2 < -0.5])
par(mfrow=c(2,2))
for(i in ctrs) {
  plot(debt[debt$Country == i, ]$ratio, debt[debt$Country == i, ]$growth,
       xlab = "ratio", ylab = "growth", main = i)
}
```



6.

```
# a
fra <- debt[debt$Country=="France", ]
# check
# dim(fra)

# b
n <- nrow(fra)
ng <- c()
for(i in c(1:n-1)) {
  ng[i] <- ifelse(fra$Year[i+1] == fra$Year[i]+1, fra$growth[i+1], NA)
}
ng[n] <- NA
fra$next.growth <- ng

signif(fra[fra$Year==1971, ]$next.growth, 3)
```

```
## [1] 5.89
```

```
signif(fra[fra$Year==1972, ]$next.growth, 3)
```

```
## [1] NA
```

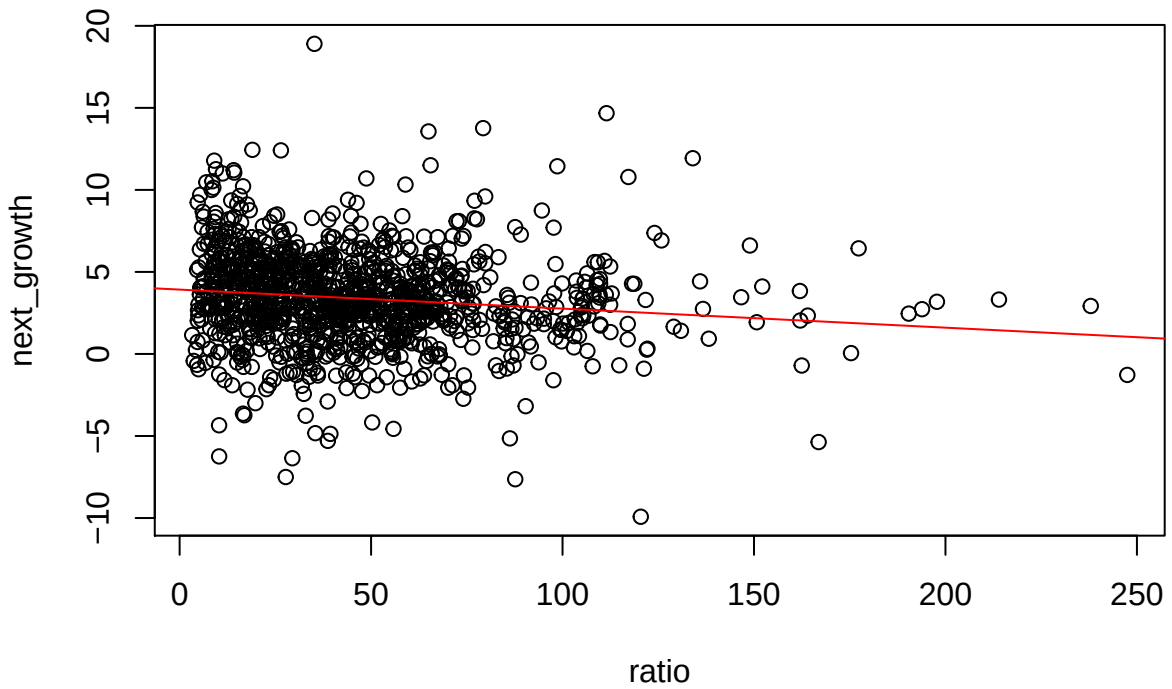
7.

```
add_ng <- function(df) {
  n <- nrow(df)
  ng <- c()
  for(i in c(1:n-1)) {
    ng[i] <- ifelse(df$Year[i+1] == df$Year[i]+1, df$growth[i+1], NA)
  }
  ng[n] <- NA
  df$next.growth <- ng
  return(df)
}

debt <- ddply(debt, "Country", add_ng)

ind <- debt$Country=="France" & debt$Year == 2009
signif(debt[ind, ]$next.growth, 3)
```

```
## [1] NA
```

8.

```
l2 <- lm(next.growth~ratio, debt)
plot(debt$ratio, debt$next.growth, xlab = "ratio", ylab = "next_growth")
abline(l2, col = 'red')
```



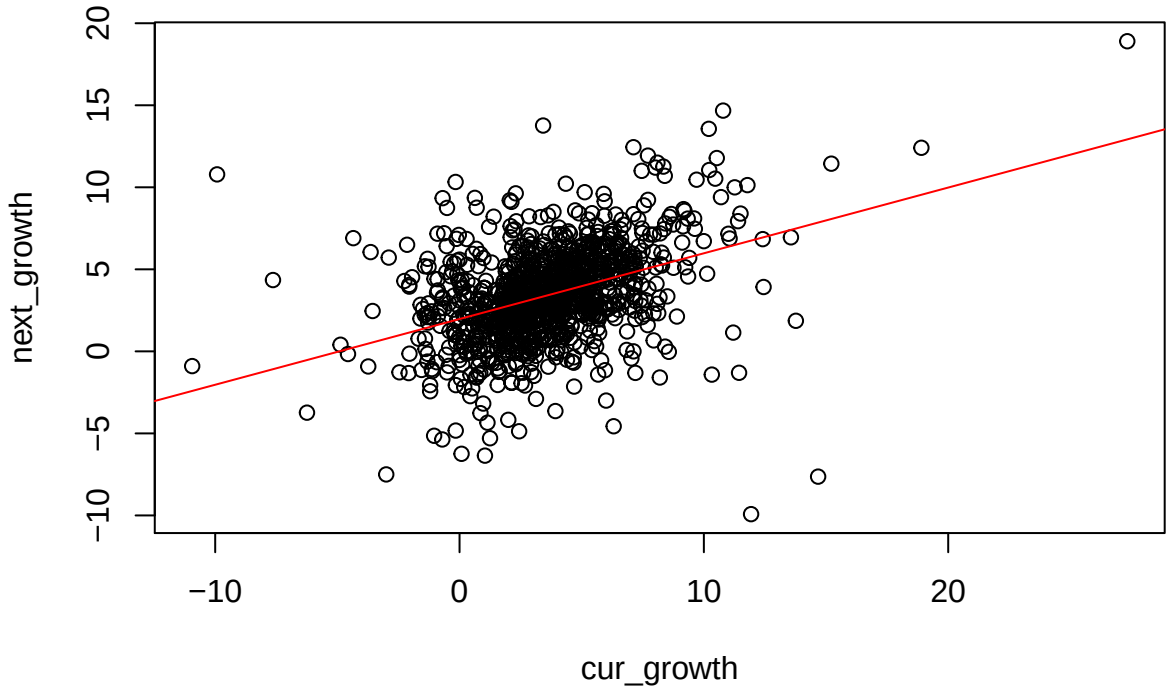

```
signif(l2$coefficients, 3)
```

```
## (Intercept)       ratio
##      3.9200     -0.0116
```

Comparatively speaking, these two perform similarly, both of which have a slightly small negative slope coefficient.

9.

```
l3 <- lm(next.growth~growth, debt)
plot(debt$growth, debt$next.growth, xlab = "cur_growth", ylab = "next_growth")
abline(l3, col = 'red')
```




```
signif(l3$coefficients, 3)
```

```
## (Intercept)      growth
##       1.970       0.401
```

Comparatively speaking, current growth seems to be a better predictor since it has a larger coefficients.