Assignment 4

Tushar Ponkshe, tvp2110

October 25, 2018

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
setwd("C:/Users/tusha/OneDrive/Documents/Columbia Fall 2018 Courses/STAT COMP & INTRO TO DATA SC
IENCE/Datasets")
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
```

1.

```
#signif(debt, digits = 3)
#a
mean.growth = function(x) {
  mean(x[,"growth"])
}
signif(mean.growth(debt),3)
```

```
## [1] 3.43
```

```
#b
require(plyr)
```

```
## Loading required package: plyr
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
signif(daply(debt, .(Country), mean.growth), 3)
```

```
##   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
```

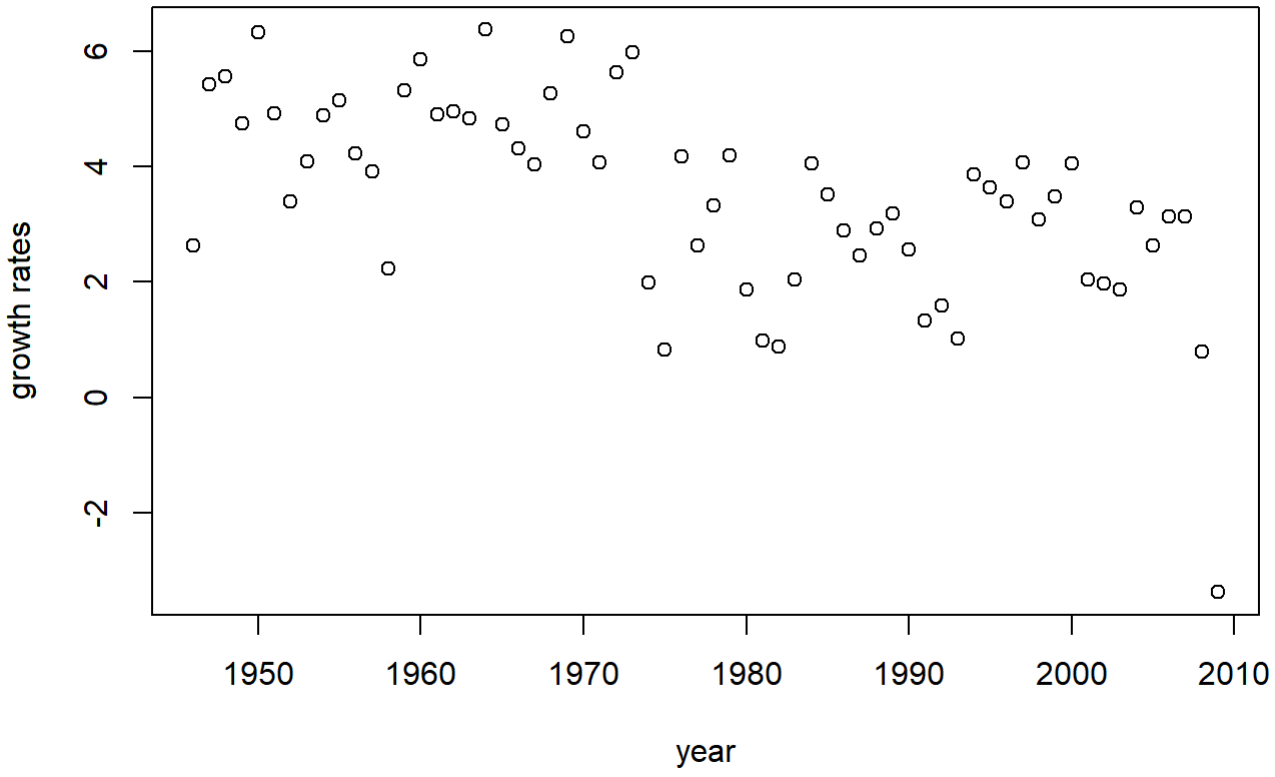
```
#Checking results
#mean(debt$growth[debt$Country=="Australia"])
```

2.

```
library(ggplot2)
yearly.growth = daply(debt, .(Year), mean.growth)
signif(yearly.growth, 3)
```

```
##   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
```

```
Years = unique(debt$Year)
#ggplot(yearly.growth, aes(x = Years, y = mean.growth)) + geom_line()
plot(Years, yearly.growth, xlab = "year", ylab = "growth rates")
```



3.

```
#a
#newdebt = na.omit(debt)
func = function(x) {
  cor(x$growth, x$ratio)
}

signif(func(debt), 3)
```

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

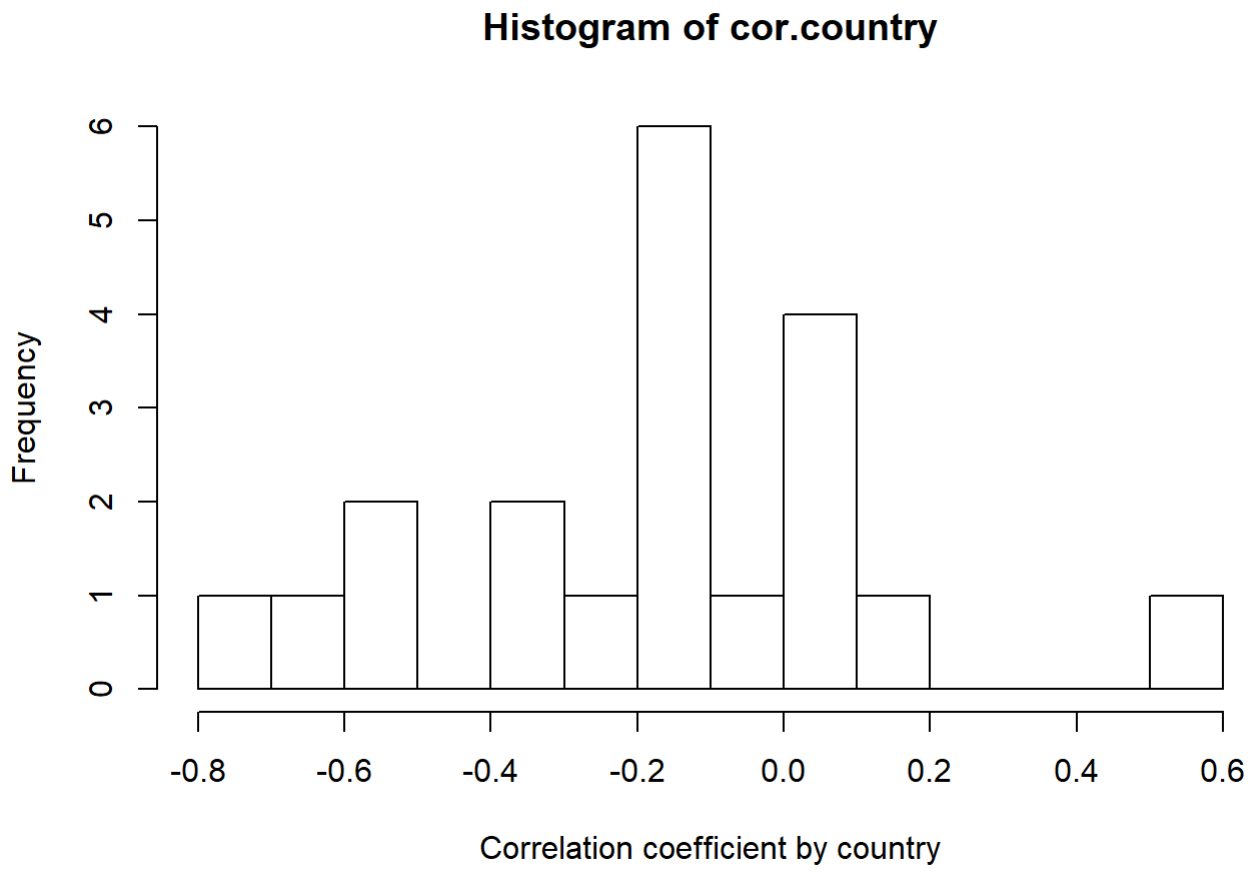
```
#b
cor.country = daply(debt, .(Country), func)
signif(head(cor.country), 3)
```

```
## Australia   Austria   Belgium    Canada   Denmark   Finland
##  0.025200 -0.253000 -0.192000  0.075000 -0.168000  0.000581
```

```
signif(mean(cor.country),3)
```

```
## [1] -0.178
```

```
hist(cor.country, breaks = 10, xlab = "Correlation coefficient by country")
```



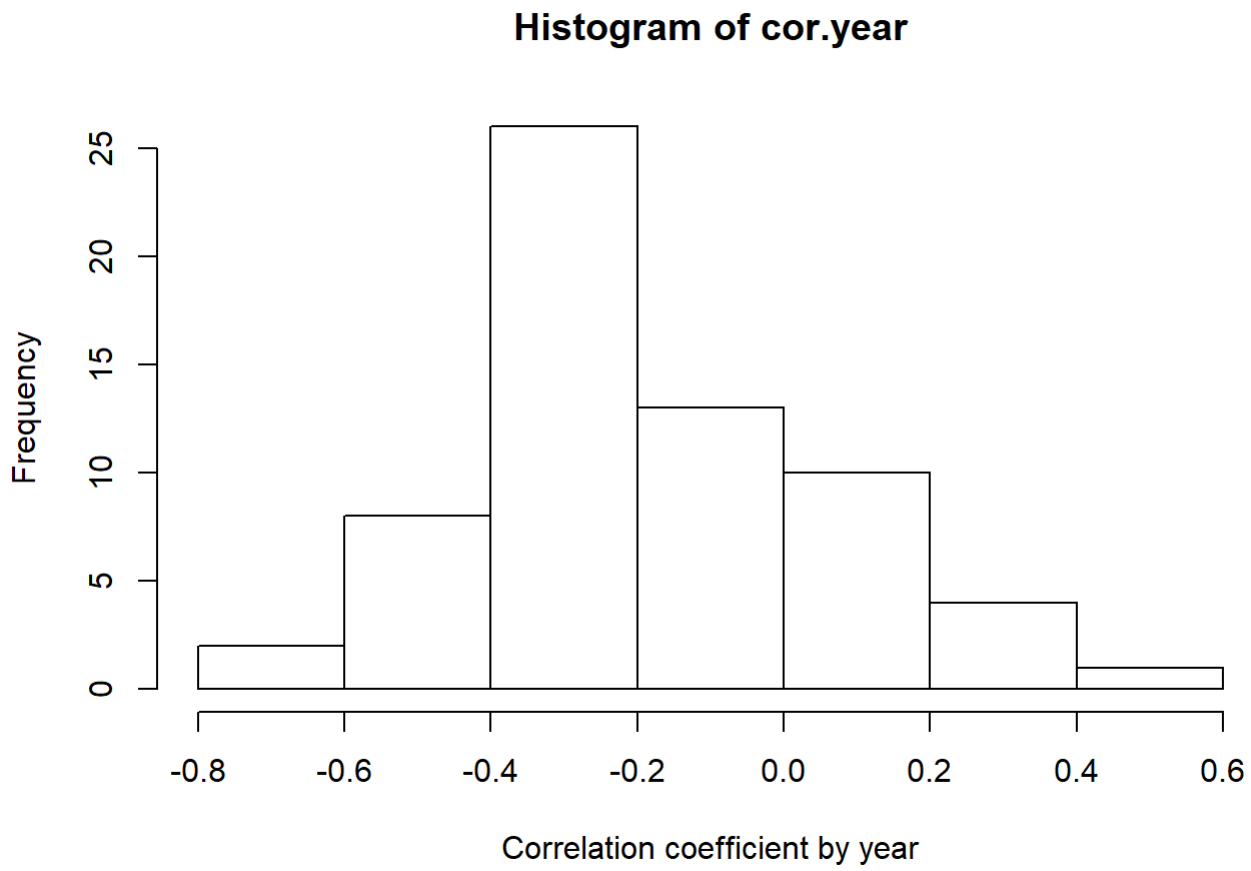

```
#c
cor.year = daply(debt, .(Year), func)
signif(head(cor.year), 3)
```

```
##    1946    1947    1948    1949    1950    1951
## -0.6200 -0.2740 -0.3400 -0.2000  0.0398 -0.4160
```

```
signif(mean(cor.year), 3)
```

```
## [1] -0.191
```

```
hist(cor.year, xlab = "Correlation coefficient by year")
```



d

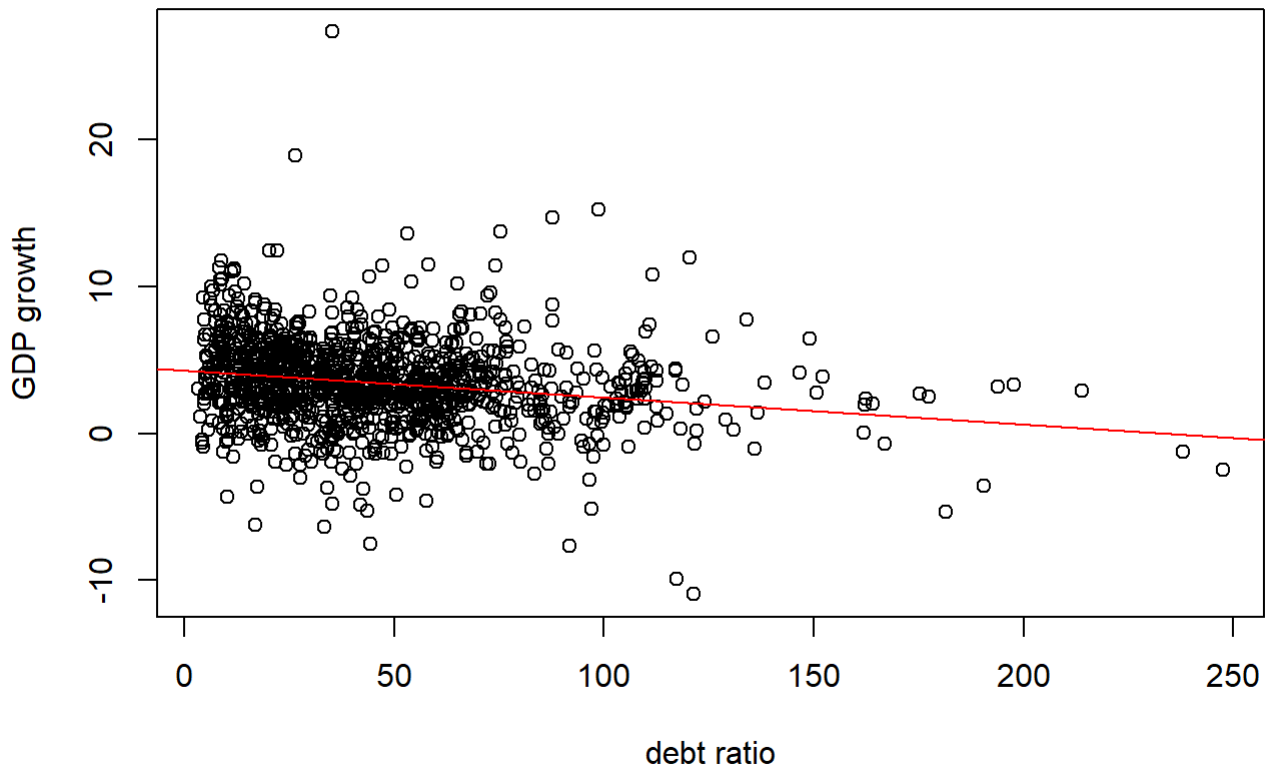
The correlation coefficient of correlation for Norway seems to be very high compared to the rest

4.

```
fit = lm(debt$growth~debt$ratio)
signif(fit$coefficients, 3)
```

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

```
plot(debt$ratio, debt$growth, ylab = "GDP growth", xlab = "debt ratio")
abline(fit, col = 2)
```



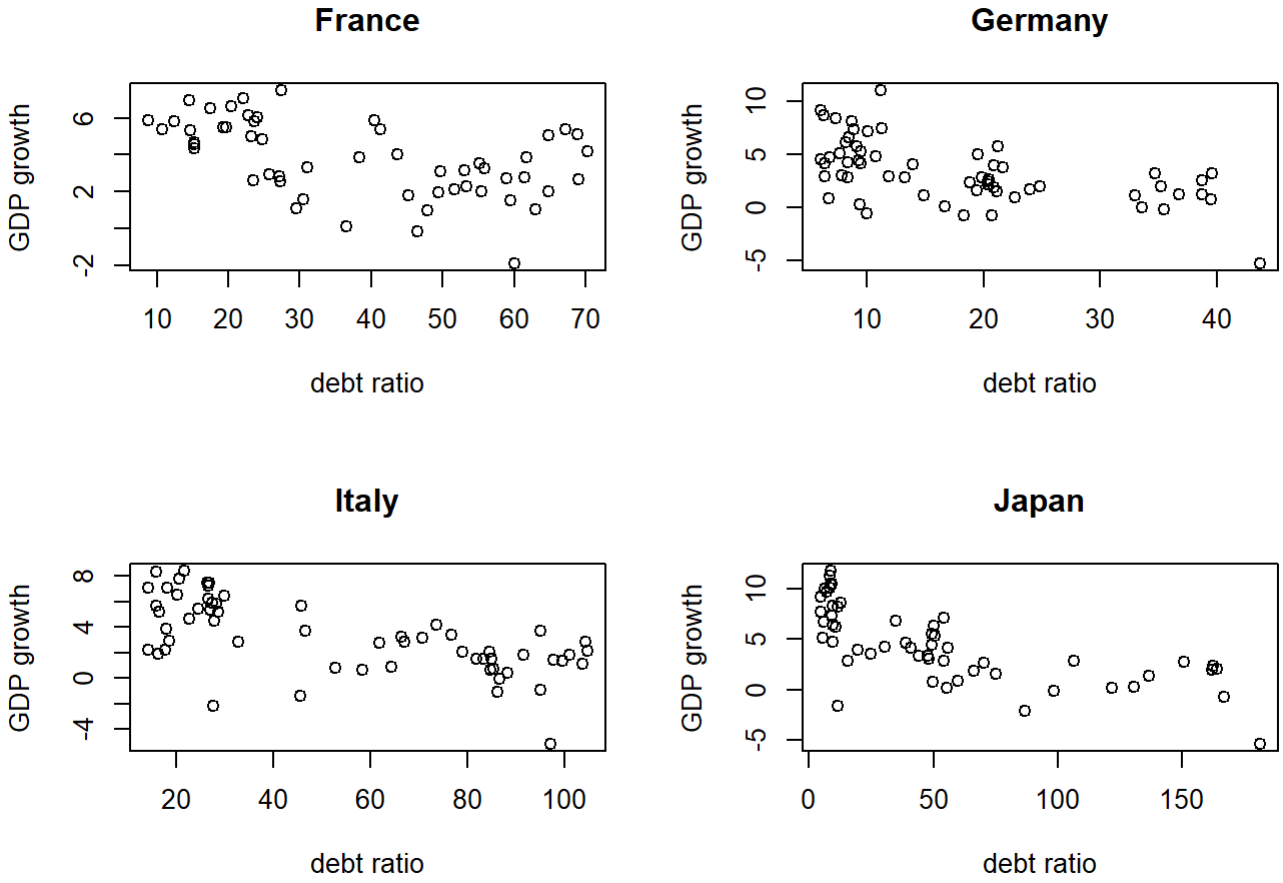
5.

```
four.countries = cor.country[which(cor.country < -0.5)]
signif(four.countries, 3)
```

```
##  France Germany   Italy   Japan
##  -0.502  -0.576  -0.645  -0.702
```

```
by.country = split(debt, debt$Country)

par(mfrow=c(2,2))
plot(by.country$France$ratio, by.country$France$growth, main = "France", xlab = "debt ratio", yl
ab = "GDP growth")
plot(by.country$Germany$ratio, by.country$Germany$growth, main = "Germany", xlab = "debt ratio",
 ylab = "GDP growth")
plot(by.country$Italy$ratio, by.country$Italy$growth, main = "Italy", xlab = "debt ratio", ylab
 = "GDP growth")
plot(by.country$Japan$ratio, by.country$Japan$growth, main = "Japan", xlab = "debt ratio", ylab
 = "GDP growth")
```



6.

```
library(dplyr)
#a
debt.france = by.country$France
dim(debt.france)
```

```
## [1] 54  4
```

```
#b
debt.france$next.growth = mutate(debt.france, next.growth = ifelse((lead(Year) - Year) == 1, lea
d(growth), NA))
debt.france[debt.france$Year=="1971","next.growth"]
```

```
##     Country Year   growth    ratio next.growth
## 392  France 1971 5.372329 10.77055    5.885827
```

7

```
func1 = function(df) {
  df$next.growth = mutate(df, next.growth = ifelse((lead(Year) - Year) == 1, lead(growth), NA))
}
debt.next = ddply(debt, .(Country), func1)
debt.next[(debt.next$Country=="France" & debt.next$Year=="2009"),]
```

```
##     Country Year    growth    ratio next.growth
## 424  France 2009 -1.906676 60.00151          NA
```

8

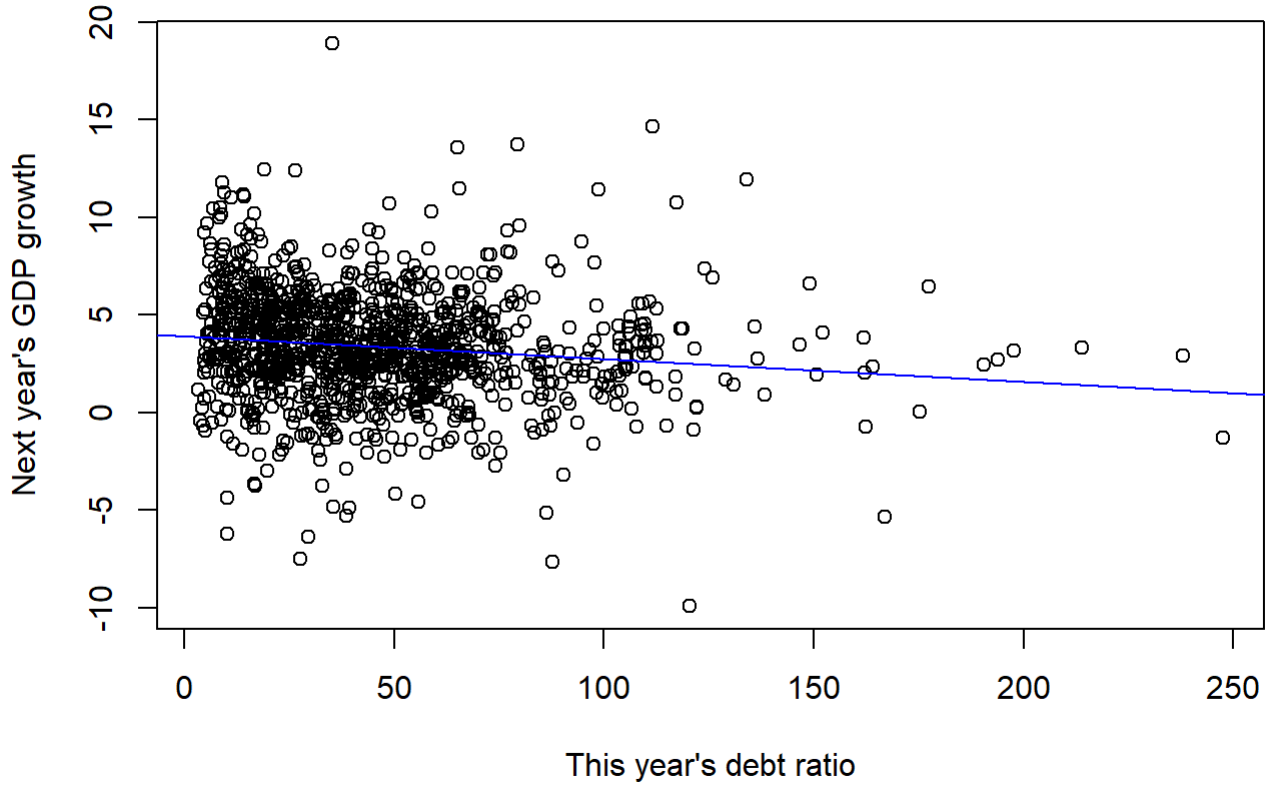
```
fit2 = lm(debt.next$next.growth~debt$ratio)
signif(fit2$coefficients,3)
```

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

```
summary(fit2)
```

```
##
## Call:
## lm(formula = debt.next$next.growth ~ debt$ratio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.4488  -1.4567  -0.0374   1.6331  15.3864
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.924722   0.143852  27.283  < 2e-16 ***
## debt$ratio  -0.011608   0.002555  -4.544 6.11e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.789 on 1145 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.01771,    Adjusted R-squared:  0.01686
## F-statistic: 20.65 on 1 and 1145 DF,  p-value: 6.105e-06
```

```
plot(debt$ratio, debt.next$next.growth, ylab = "Next year's GDP growth", xlab = "This year's deb
t ratio")
abline(fit2, col=4)
```



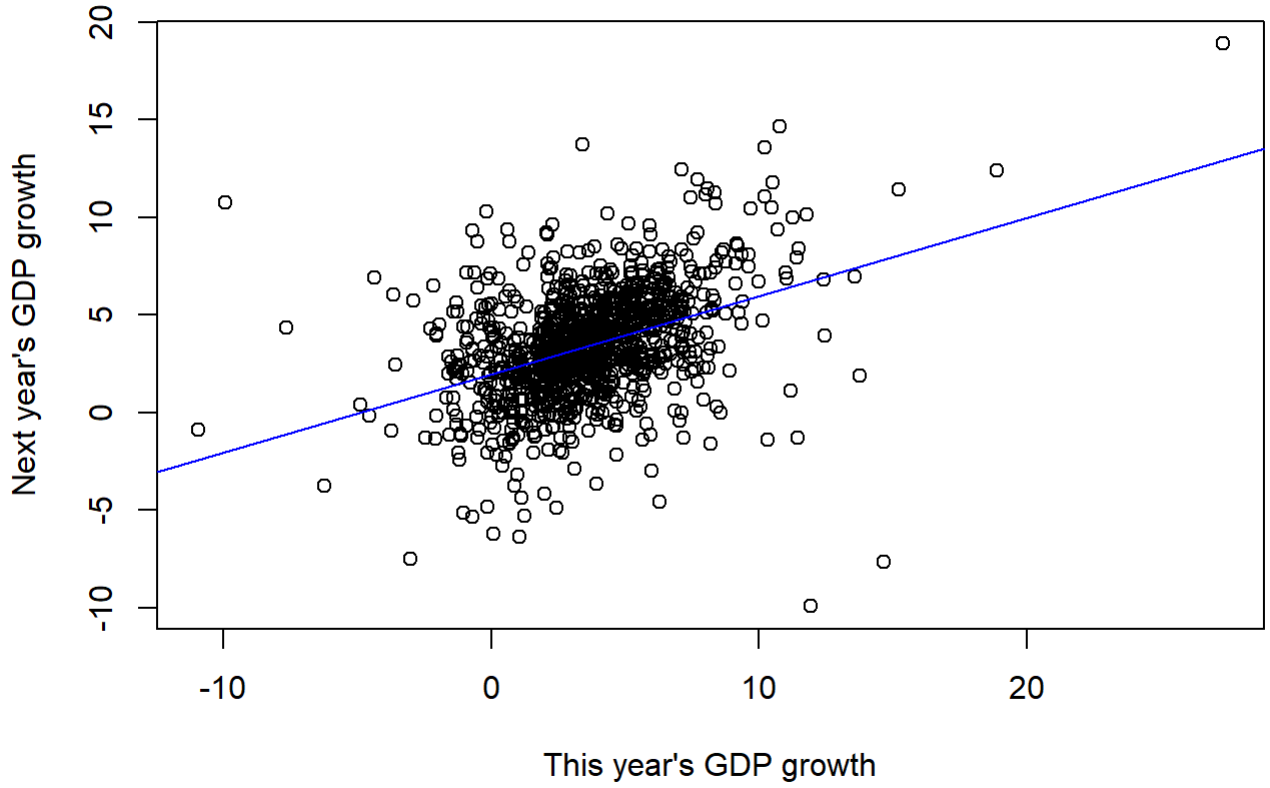
How do they compare to the regression of the current year's growth on the current year's debt ratio? - The coefficients for the intercept and the slope for the two linear models are very close in values.

9

```
fit3 = lm(debt.next$next.growth~debt$growth)
signif(fit3$coefficients, 3)
```

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

```
plot(debt$growth, debt.next$next.growth, ylab = "Next year's GDP growth", xlab = "This year's GD
P growth")
abline(fit3, col = 4)
```




```
summary(fit3)
```

```
##
## Call:
## lm(formula = debt.next$next.growth ~ debt$growth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.6738  -1.3570   0.0401   1.3994  12.7917
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.97106    0.12040   16.37   <2e-16 ***
## debt$growth  0.40065    0.02643   15.16   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.568 on 1145 degrees of freedom
##   (24 observations deleted due to missingness)
## Multiple R-squared:  0.1671, Adjusted R-squared:  0.1664
## F-statistic: 229.8 on 1 and 1145 DF,  p-value: < 2.2e-16
```

Current Year's growth is a better predictor for future growth as seen from R squared in summary of the two fits