Homework 3

Shreya Rao sr3843

5/27/2021

```r
#Loading libraries
library(plyr)

#Loding the file
debt <- read.csv("debt.csv", as.is = TRUE)
dim(debt)
```

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

```r
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
```

```r
#ratio: debt to GDP ratio
```

1. Calculate the average GDP growth rate for each country (averaging over years). This is a classic split/apply/combine problem, and you will use daply() to solve it.

```r
#a)
mean.growth <- function(df) {
  return(signif(mean(df$growth, na.rm = TRUE), 3))
}

#b)
avg_growth <- daply(debt, .(Country), mean.growth)
avg_growth["Australia"]
```

```
## Australia
##      3.72
```

```r
avg_growth["Netherlands"]
```

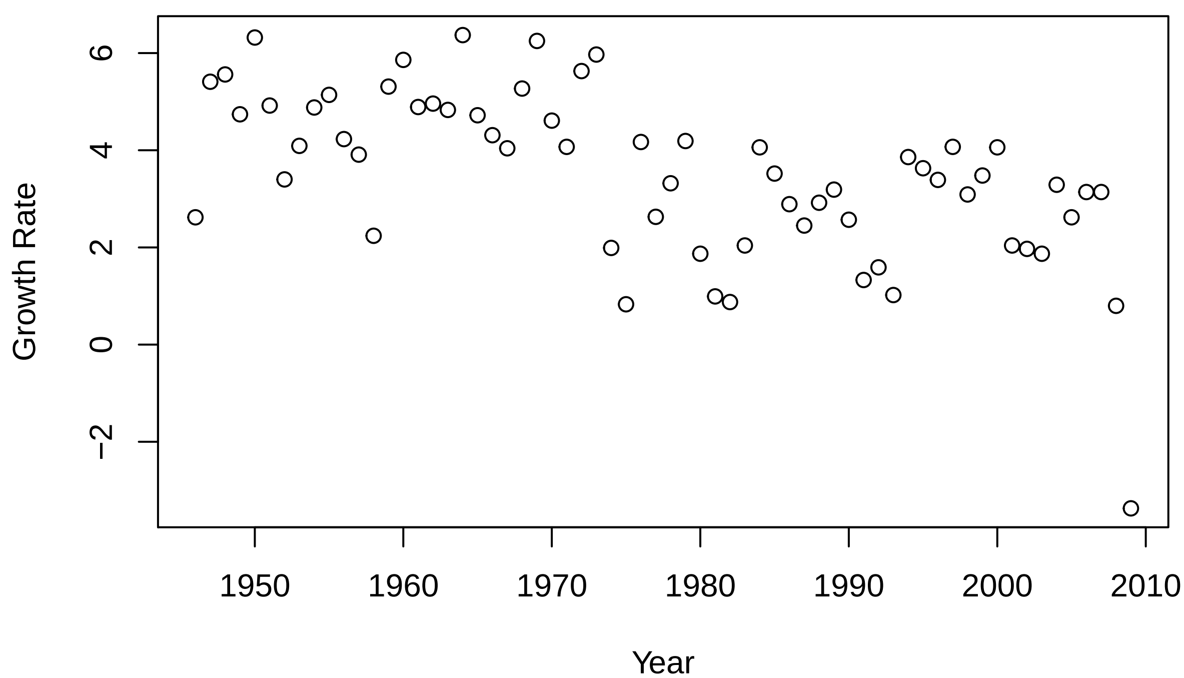
```
## Netherlands
##        3.03
```

2. Using the same instructions as problem 1, calculate the average GDP growth rate for each year (now averaging over countries). (The average growth rates for 1972 and 1989 should be 5.63 and 3.19, respectively. Print these values in your output.) Make a plot of the growth rates (y-axis) versus the year (x-axis). Make sure the axes are labeled appropriately.

```
year_means <- daply(debt, .(Year), mean.growth)
year_means[c("1972", "1989")]
```

```
## 1972 1989
## 5.63 3.19
```

```
plot(x=names(year_means), y=year_means, xlab = "Year", ylab="Growth Rate")
```



3. The function cor(x,y) calculates the correlation coefficient between two vectors x and y.

a. Calculate the correlation coefficient between GDP growth and the debt ratio over the whole data set (all countries, all years). Your answer should be -0.1995.

```
signif(cor(debt$growth, debt$ratio), 4)
```

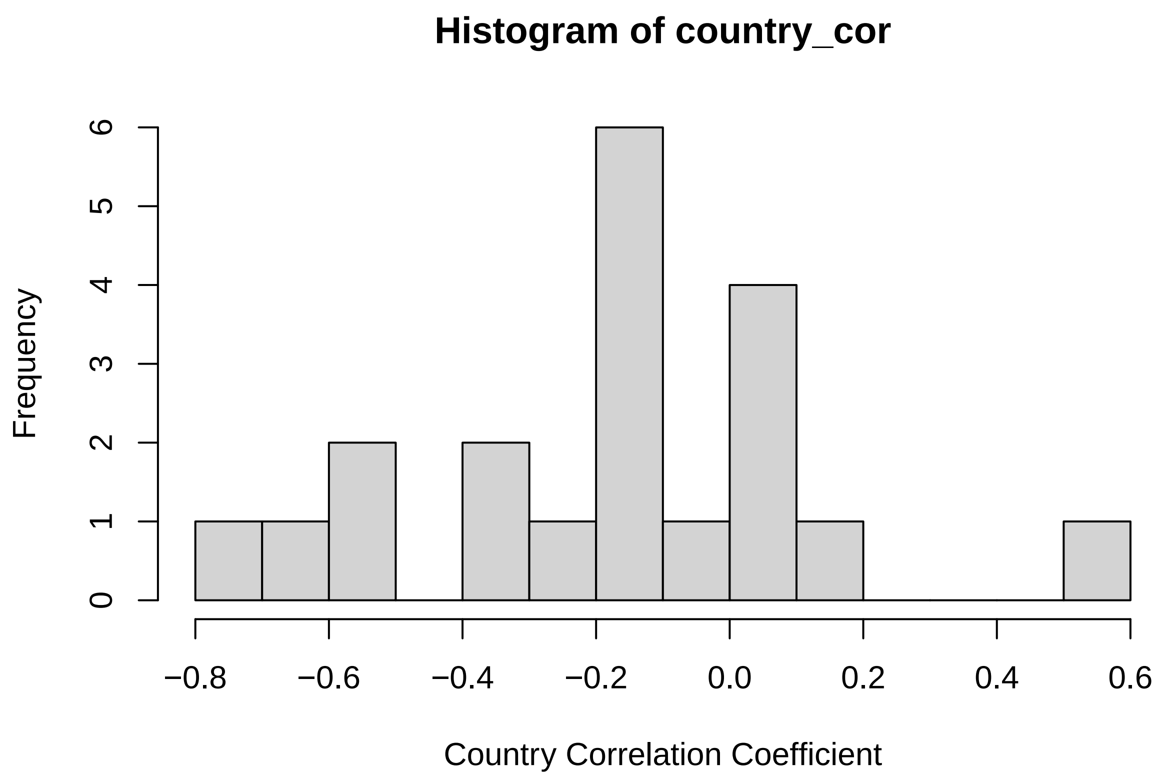
```
## [1] -0.1995
```

b. Compute the correlation coefficient separately for each country, and plot a histogram of these coefficients (with 10 breaks). The mean of these correlations should be -0.1778. Do not use a loop. (Hint: consider writing a function and then making it an argument to daply()).

```r
cor_func <- function(df) {
  return(signif(cor(df$growth, df$ratio), 3))
}

country_cor <- daply(debt, .(Country), cor_func)
#country_cor
signif(mean(country_cor), 4)
```

```
## [1] -0.1778
```

```r
hist(country_cor, breaks = 10, xlab="Country Correlation Coefficient")
```

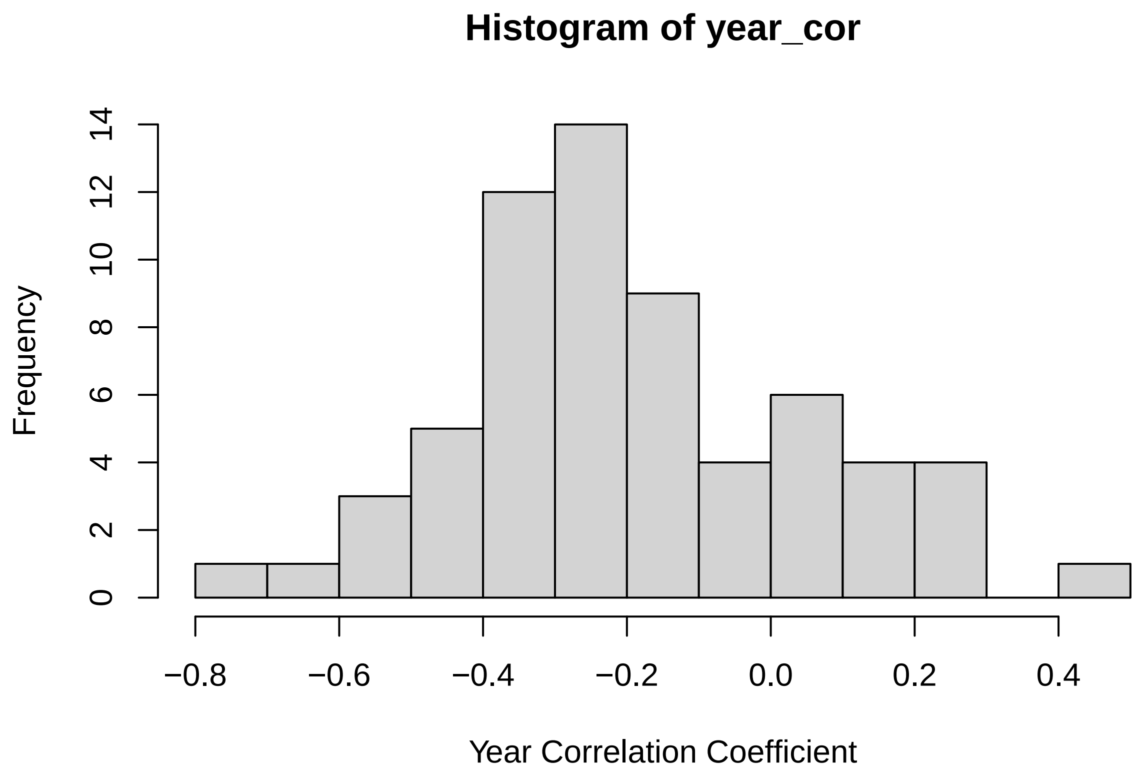


c. Calculate the correlation coefficient separately for each year, and plot a histogram of these coefficients. The mean of these correlations should be -0.1906.

```r
year_cor <- daply(debt, .(Year), cor_func)
signif(mean(year_cor), 3)
```

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

```r
hist(year_cor, breaks = 10, xlab="Year Correlation Coefficient")
```



d. Are there any countries or years where the correlation goes against the general trend?

```
country_cor
```

```
##   Australia     Austria     Belgium      Canada     Denmark     Finland
##    0.025200   -0.253000   -0.192000    0.075000   -0.168000    0.000581
##      France     Germany      Greece     Ireland       Italy       Japan
##   -0.502000   -0.576000   -0.093500   -0.140000   -0.645000   -0.702000
## Netherlands New Zealand      Norway    Portugal       Spain      Sweden
##   -0.199000    0.161000    0.563000   -0.352000    0.081400   -0.161000
##          UK          US
##   -0.137000   -0.341000
```

```
country_cor[country_cor > 0.2 | country_cor < -0.6]
```

```
##  Italy  Japan Norway
## -0.645 -0.702  0.563
```

Norway has an unsually high positive correlation of 0.563 while the rest have correlation close to 0 or negative correlation. Japan and Italy have high negative correlations (below -0.65).

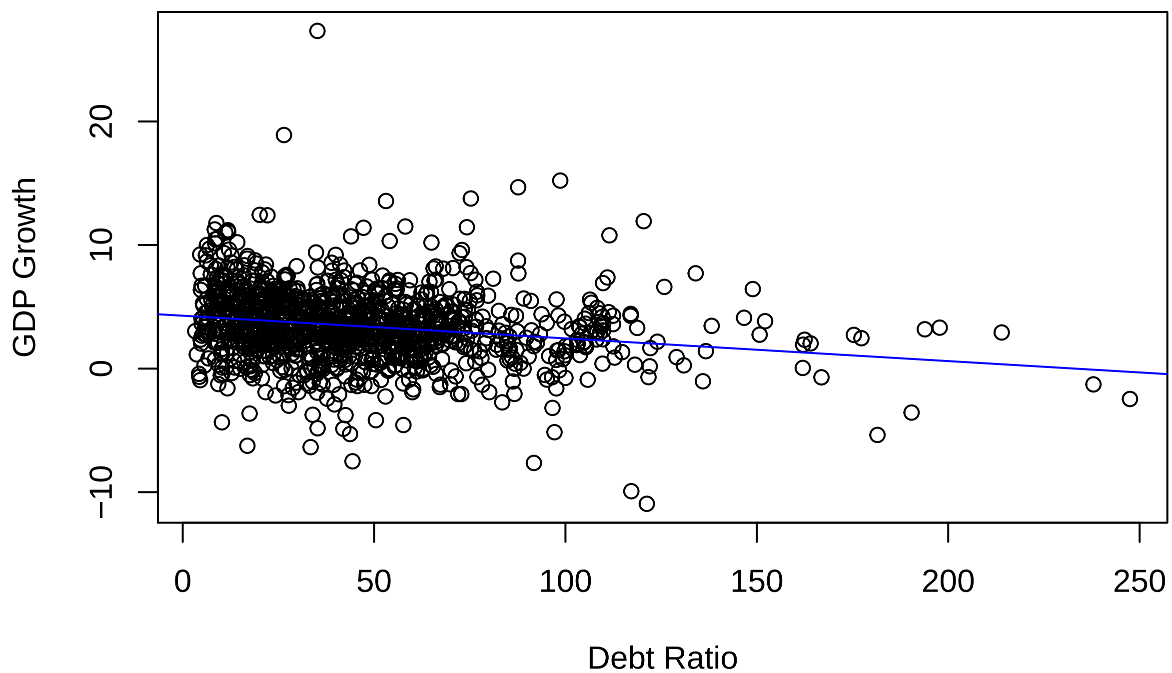
```
#plot(names(year_cor), year_cor)
year_cor[year_cor > quantile(year_cor, 0.975) | year_cor < quantile(year_cor, 0.025)]
```

```
##    1946    1957    1975    1978
## -0.620 -0.755  0.271  0.431
```

The years 1946 and 1975 have high correlations that are negatively correlated (below -0.6). 1978 has an evidently high postive correlation of 0.431.

4. Fit a linear model of overall growth on the debt ratio, using lm(). Report the intercept and slope. Make a scatter-plot of overall GDP growth (vertical) against the overall debt ratio (horizontal). Add a line to your scatterplot showing the fitted regression line

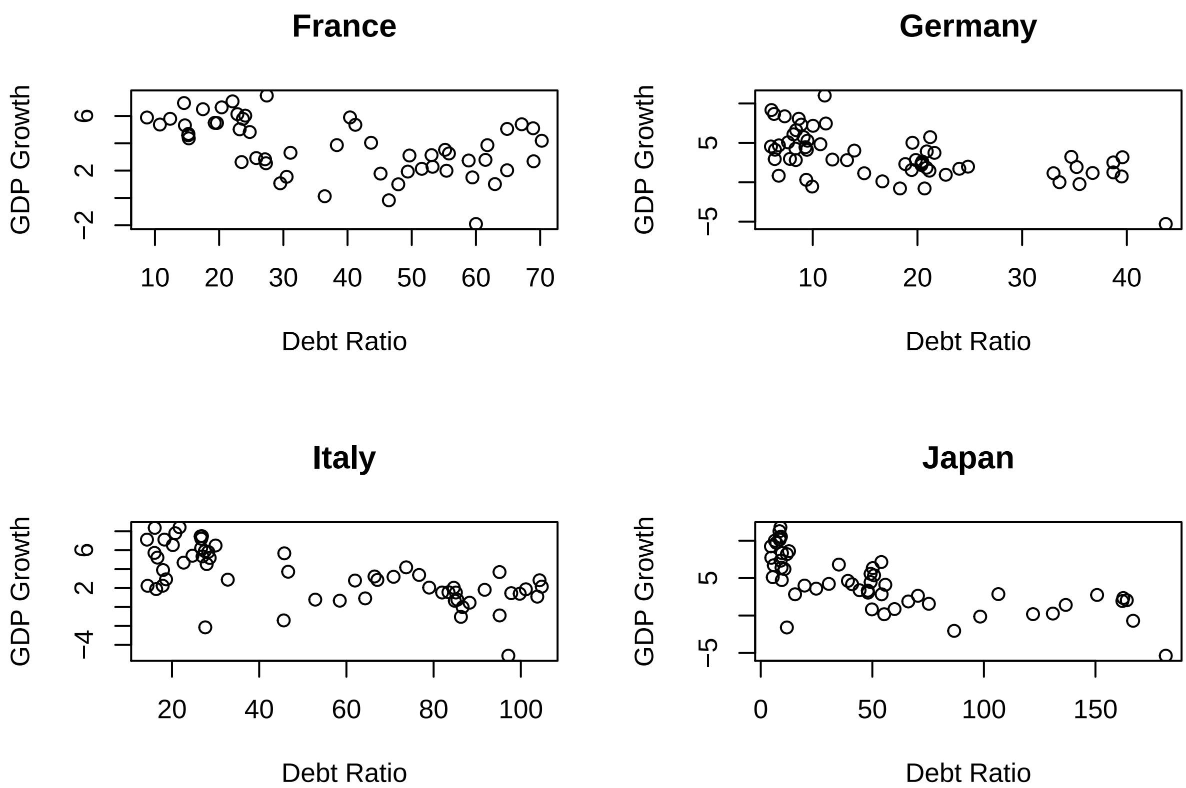
```r
lm_growth_debt <- lm(growth~ratio, data=debt)
plot(x=debt$ratio, y=debt$growth, xlab="Debt Ratio", ylab="GDP Growth")
abline(lm_growth_debt, col="blue")
```



5. There should be four countries with a correlation smaller than -0.5. Separately, plot GDP growth versus debt ratio from each of these four countries and put the country names in the titles. This should be four plots. Call par(mfrow=c(2,2)) before plotting so all four plots will appear in the same figure.

```r
countries_4 <- country_cor[country_cor < -0.5]
names_4 <- names(countries_4)
par(mfrow=c(2,2))

for(i in names_4) {
```

```
  plot(y=debt$growth[debt$Country == i], x=debt$ratio[debt$Country == i], xlab="Debt Ratio", ylab="GDP
  title(main = i)
}
```



6. Some economists claim that high levels of government debt cause slower growth. Other economists claim that low economic growth leads to higher levels of government debt. The data file, as given, lets us relate this year's debt to this year's growth rate; to check these claims, we need to relate current debt to future growth.

a. Create a new data frame which just contains the rows of debt for France, but contains all those rows. It should have 54 rows and 4 columns (print the dimensions of your data frame). Note that some years are missing from the middle of this data set.

```
france <- debt[debt$Country == "France",]
dim(france)
```

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

b. Create a new column in your data frame for France, next.growth, which gives next year's growth if the next year is in the data frame, or NA if the next year is missing. (next.growth for 1971 should be (rounded) 5.886, but for 1972 it should be NA. Print these two values.)

```r
for(i in 1:(nrow(france))) {
  if ((france$Year[i]+1 == france$Year[i+1]) & (i != nrow(france))) {
    france$next.growth[i] <- signif(france$growth[i+1], 4)
  }
  else {
    france$next.growth[i] <- NA
  }
}

france[france$Year == 1971 | france$Year == 1972,]
```

```
##     Country Year   growth     ratio next.growth
## 392  France 1971 5.372329 10.770552       5.886
## 393  France 1972 5.885827  8.757901          NA
```

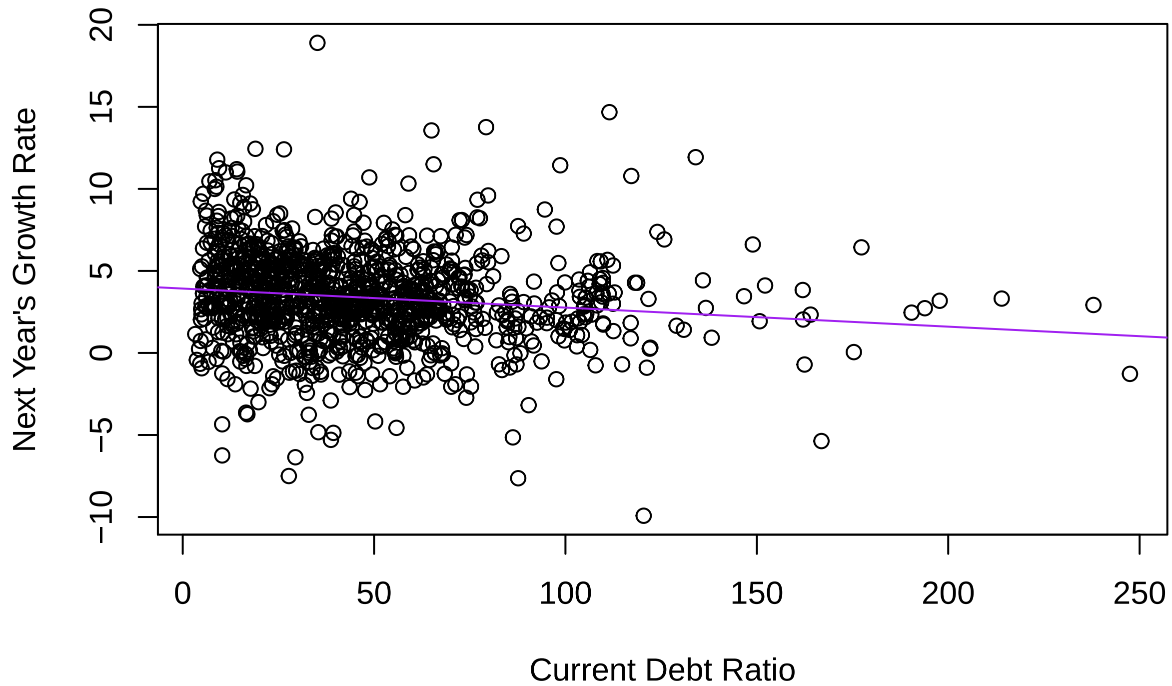
7. Add a next.growth column, as in the previous question, to the whole of the debt data frame. Make sure that you do not accidentally put the first growth value for one country as the next.growth value for another. (The next.growth for France in 2009 should be NA, not 9.167. Print this value.) Hints: Write a function to encapsulate what you did in the previous question, and apply it using ddply().

```r
next_growth <- function(df) {

  temp_growth = NULL
  for(i in 1:(nrow(df))) {
    if ((df$Year[i]+1) == df$Year[i+1] & (i != nrow(df))) {
      df$next.growth[i] <- round(df$growth[i+1], 3)
    }
    else {
      df$next.growth[i] <- NA
    }
  }
  return(df)
}

# t_lists <- split(debt, debt$Country)
# sapply(t_lists, next_growth)

debt <- ddply(debt, .(Country), next_growth)

debt[debt$Country == "France" & debt$Year == 2009,]
```

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

8. Make a scatter-plot of next year's GDP growth against this year's debt ratio. Linearly regress next year's growth rate on the current year's debt ratio, and add the line to the plot. Report the intercept and slope to reasonable precision. How do they compare to the regression of the current year's growth on the current year's debt ratio?

```r
plot(x=debt$ratio, y=debt$next.growth, xlab="Current Debt Ratio", ylab="Next Year's Growth Rate")

reg_line <- lm(next.growth~ratio, data = debt)
abline(reg_line, col="purple")
```




```r
#Next year's coefficients
signif(coefficients(reg_line), 3)
```

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

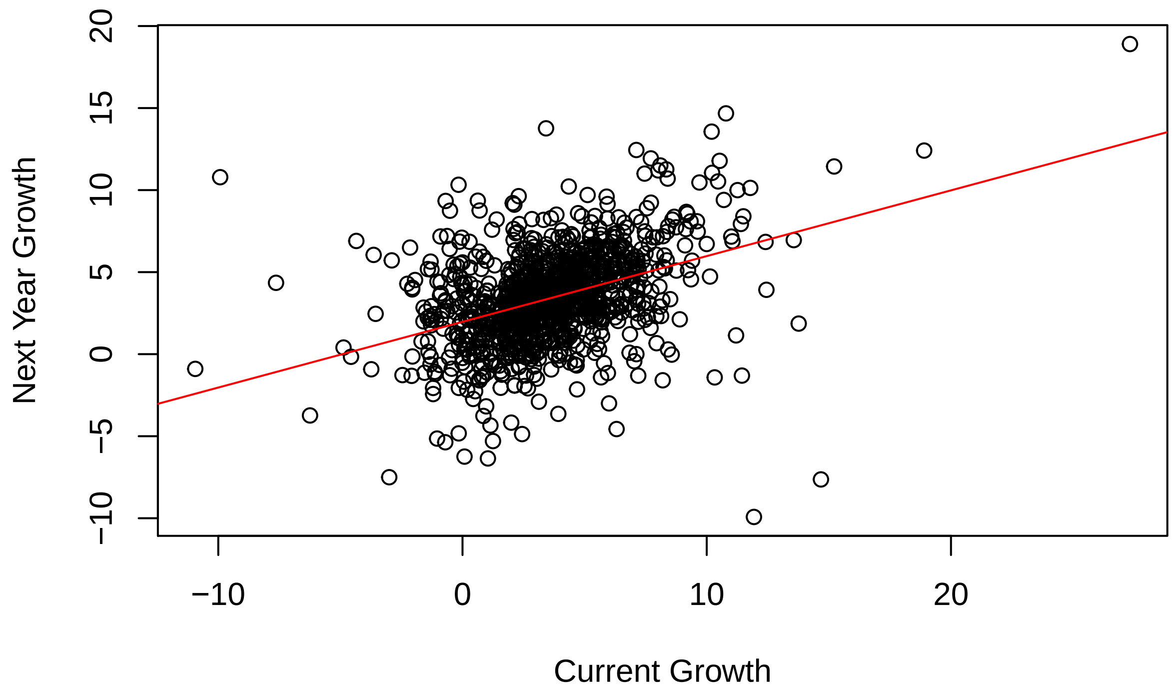
```r
#summary(reg_line)

#This year's coefficients
signif(coefficients(lm_growth_debt), 3)
```

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

Both of them have similar negative slopes, with the current year's regression slope very slightly less that next year's. They also have similar intercepts around 4. Overall, they have similar regression lines.

9. Make a scatter-plot of next year's GDP growth against the current year's GDP growth. Linearly regress next year's growth on this year's growth, and add the line to the plot. Report the coefficients. Can you tell, from comparing these two simple regressions (from the current question, and the previous), whether current growth or current debt is a better predictor of future growth?

```r
plot(x=debt$growth, y=debt$next.growth, xlab="Current Growth", ylab="Next Year Growth")
growth_reg <- lm(next.growth~growth, data = debt)
abline(growth_reg, col="red")
```




```r
signif(coefficients(growth_reg), 3)
```

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

```r
summary(growth_reg)$r.squared
```

```
## [1] 0.1671384
```

```r
#Next year vs debt summary
summary(reg_line)$r.squared
```

```
## [1] 0.01771379
```

There seems to be a more linear relationship between next year's growth and current year's growth. This is also indicated by the slopes. This regression line also seems to be a better fit (as indicated by a higher R squared value) than that of next year's growth against current debt. Thus, current growth seems to be a better predictor of future growth.