5206_hw4_sw3449

Shuyuan Wang, sw3449

11/7/2019

```
debt <- read.csv('debt.csv',as.is=T)
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. Calculate the average GDP growth rate for each country (averaging over years).

```
library(plyr)
# write a function: take a data frame as argument and
# return the mean of the "growth" column of that data frame
mean.growth <- function(my_df){
  return(signif(mean(my_df[,'growth']),3))
}

# Use daply() to apply mean.growth() to each country in debt
growth_for_country <- daply(debt,.(Country),mean.growth)

# check
growth_for_country['Australia']
```

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

```
growth_for_country['Netherlands']
```

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

2. Calculate the average GDP growth rate for each year (averaging over countries). Make a plot of the growth rate versus the year.

```
avg_for_year <- daply(debt,.(Year),mean.growth)

# check
avg_for_year['1972']
```

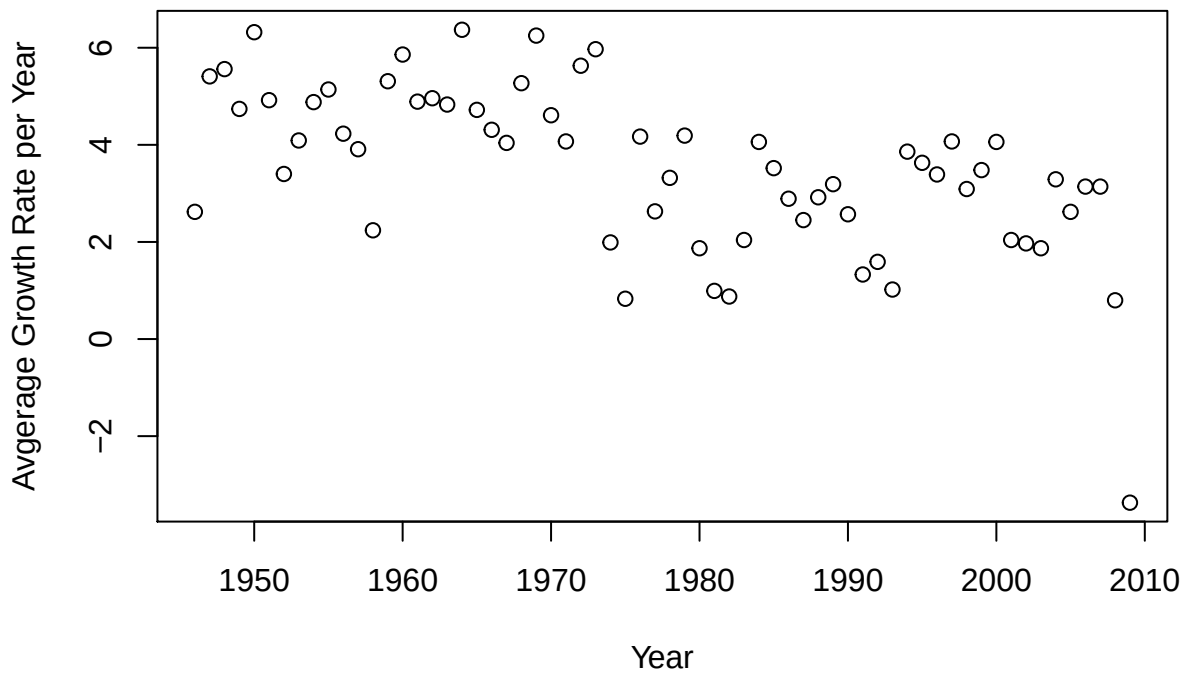
```
## 1972
```

```
## 5.63
```

```
avg_for_year['1989']
```

```
## 1989
## 3.19
```

```
# plot of the growth rates (y-axis) versus the year (x-axis)
plot(x=min(debt$Year):max(debt$Year),y=avg_for_year,
     xlab='Year',ylab='Avgerage Growth Rate per Year')
```



3.a Calculate the correlation coefficient between GDP growth rate and the debt ratio over the whole data set.

```
signif(cor(debt$growth,debt$ratio),3)
```

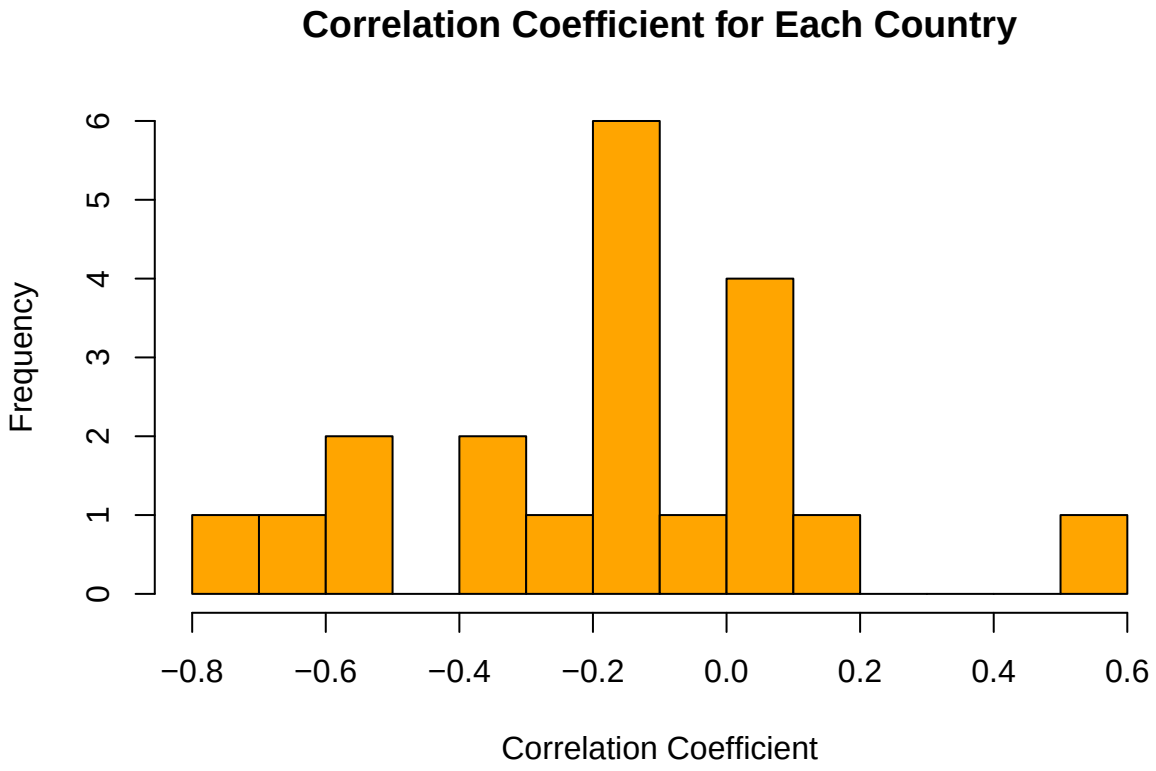
```
## [1] -0.199
```

3.b Compute the correlation coefficient separately for each country, and plot a histogram of these coefficients (10 breaks)

```
# write a function to calculate the correlation
cor.cal <- function(my_df){
  return(signif(cor(my_df$growth,my_df$ratio),3))
}

cor_per_country <- daply(debt,.(Country),cor.cal)

# check
signif(mean(cor_per_country),3)
```

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

```r
# plot a histogram of these coefficients
hist(cor_per_country,breaks=10,
     main='Correlation Coefficient for Each Country',
     xlab='Correlation Coefficient',col='orange')
```

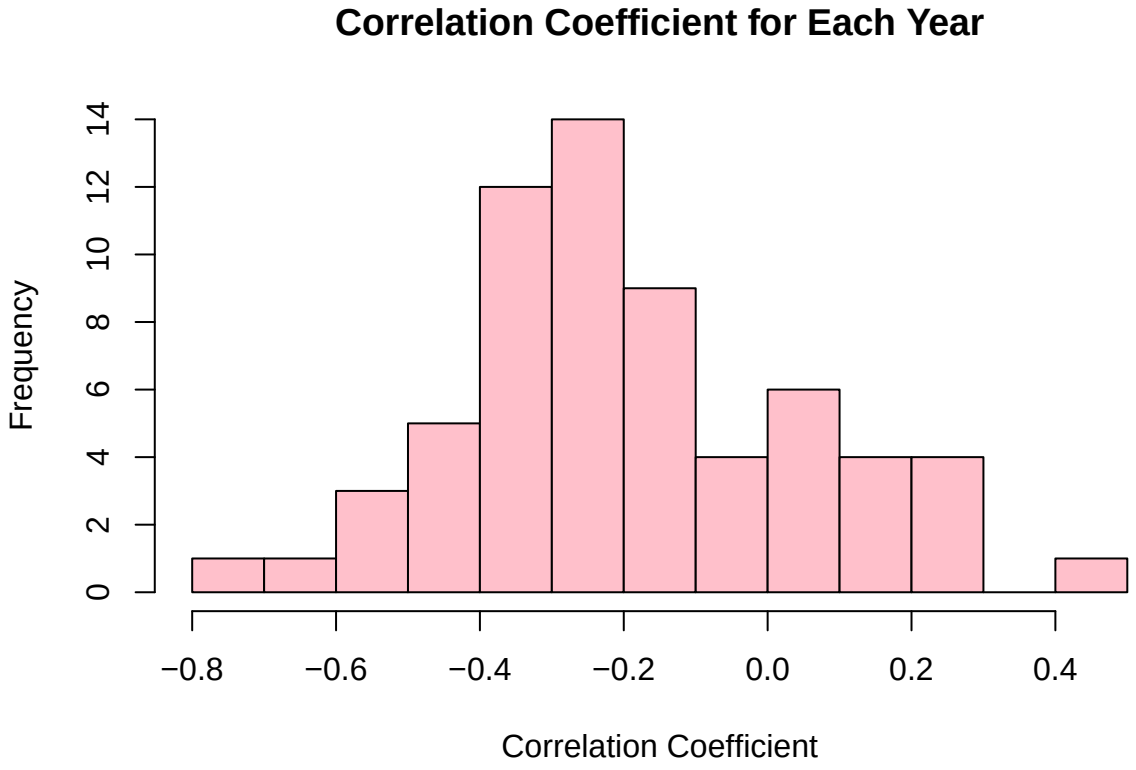


3.c Calculate the correlation coefficient separately for each year, and plot a histogram of these coefficients.

```r
cor_per_year <- daply(debt,.(Year),cor.cal)

# check
signif(mean(cor_per_year),3)
```

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

```r
# plot a histogram of these coefficients
hist(cor_per_year,breaks=10,main='Correlation Coefficient for Each Year',
     xlab='Correlation Coefficient',col='pink')
```



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

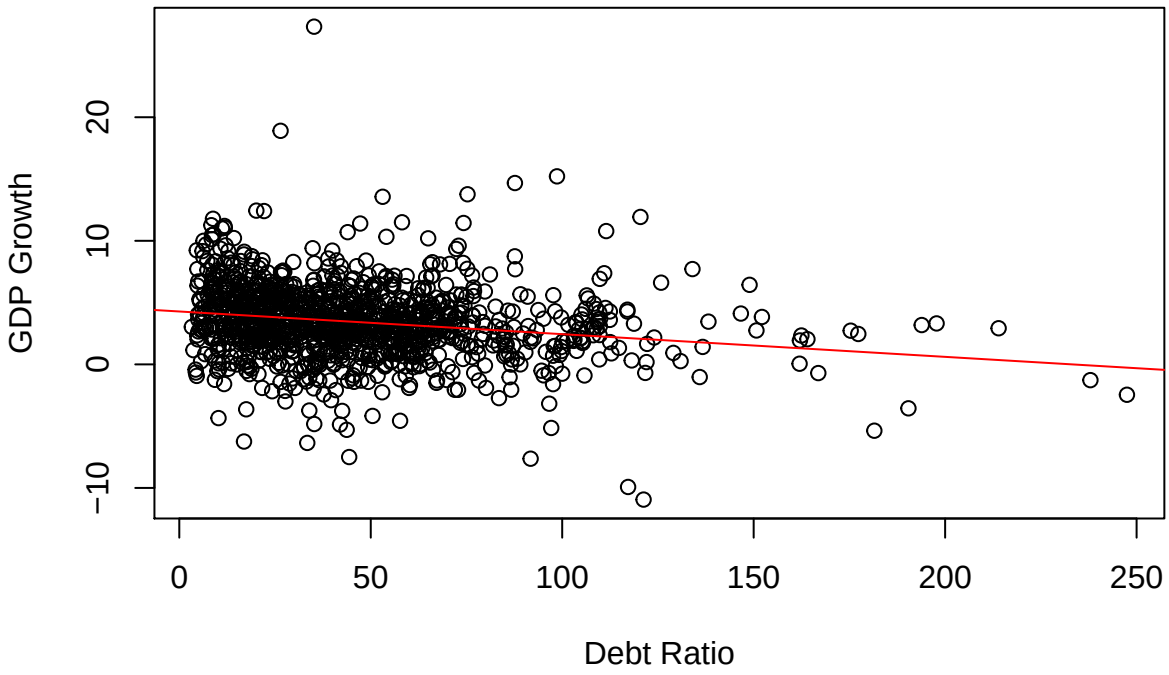
Yes. From the plot in 3.b we can see that there are 4 countries with correlation smaller than -0.5 and one country with correlation larger than 0.5. From 3.c we can see that there is a year with coefficient larger than 0.4.

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.

```
lm0 <- lm(growth~ratio,data=debt)
lm0$coefficient
```

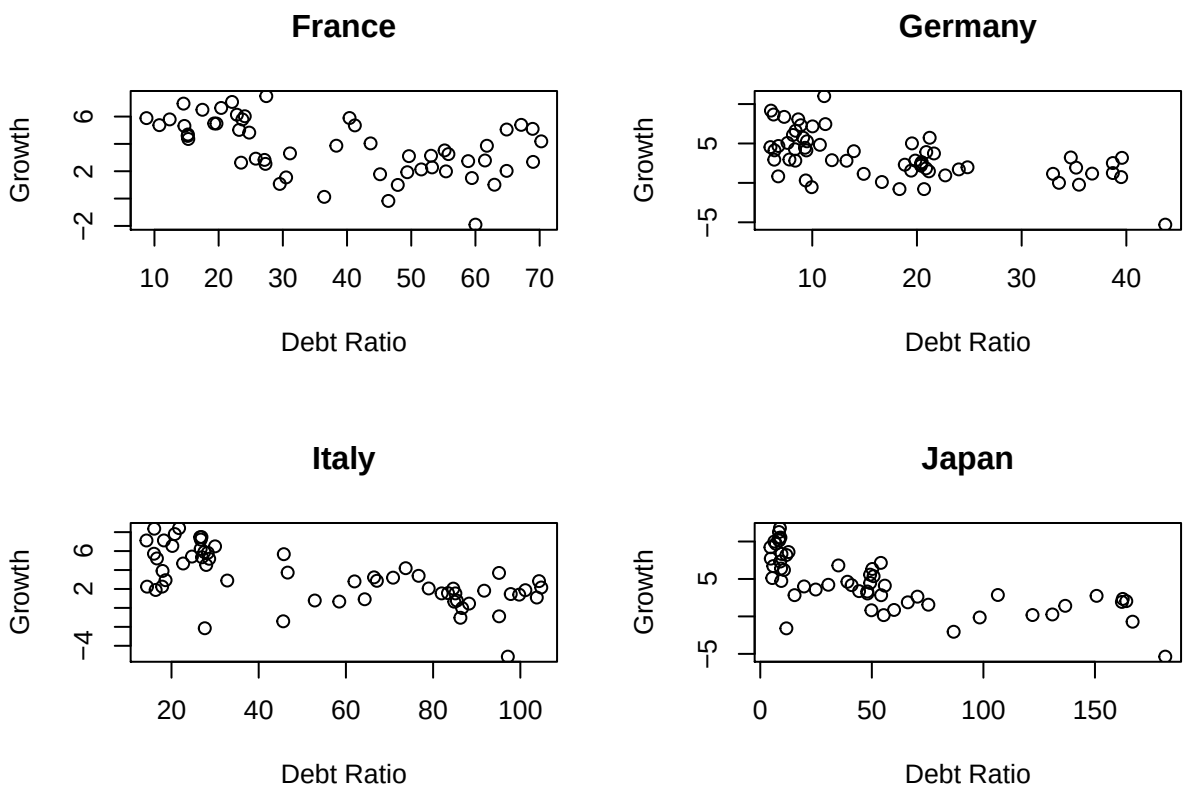
```
## (Intercept)       ratio
##  4.27929049 -0.01835518
```

```
plot(x=debt$ratio,y=debt$growth,xlab='Debt Ratio',ylab='GDP Growth')
abline(lm0,col='red')
```



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.

```
# extract the country names
four_countries <- names(cor_per_country)[which(cor_per_country< -0.5)]

# plot the graph
par(mfrow=c(2,2))

for (i in 1:4) {
  plot(x=debt$ratio[debt$Country == four_countries[i]],
       y=debt$growth[debt$Country == four_countries[i]],
       xlab = "Debt Ratio", ylab = "Growth", main = four_countries[i])
}
```



6.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.

```
debt_france <- debt[debt$Country=='France',]
dim(debt_france)
```

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

6.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.

```
# write a function to check if the year of a country has next year growth,
#and fill in the next.growth column
next.growth <- function(year,my_df){
  if(any(my_df$Year==(year+1)))
    return(my_df$growth[my_df$Year==(year+1)])
  else
    return(NA)
}

# apply the function next.growth() to every element in debt_france$Year
debt_france$next.growth <- sapply(debt_france$Year,next.growth,debt_france)

# check
debt_france$next.growth[debt_france$Year==1971]
```

```
## [1] 5.885827
```

```
debt_france$next.growth[debt_france$Year==1972]
```

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

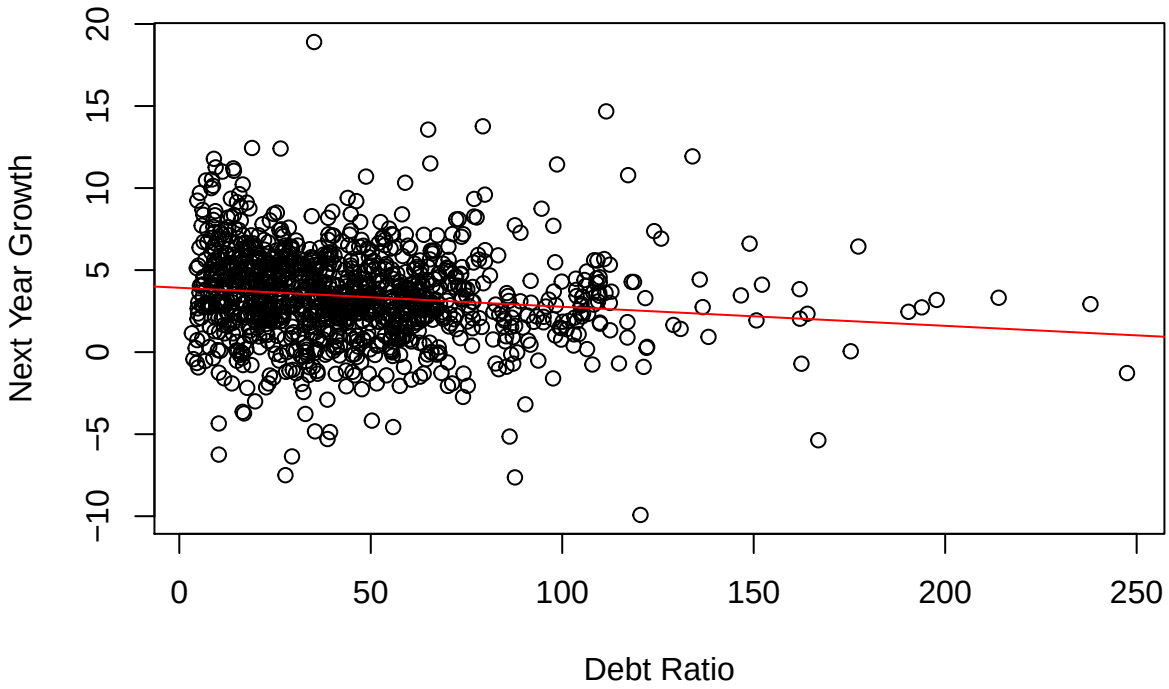
7. Add a `next.growth` column, as in the previous question, to the whole of the debt data frame. Make sure that you do not accidentally put the first growth value for one country as the next.growth value for another. (The next.growth for France in 2009 should be `NA`, not 9.167. Print this value.) Hints: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

```
# write a function to fill in the next.growth column for a given data frame
fill_next.growth <- function(my_df) {
  my_df$next.growth <- sapply(my_df$Year, next.growth, my_df)
  return(my_df)
}

debt <- ddply(debt, .(Country), fill_next.growth)

# check
debt$next.growth[debt$Country == "France" & debt$Year == 2009]
```

```
## [1] 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?

```
par(mfrow=c(1,1))
lm1 <- lm(next.growth~ratio,data=debt)

plot(x=debt$ratio,y=debt$next.growth,xlab='Debt Ratio',ylab='Next Year Growth')
abline(lm1,col='red')
```




```
# regression of next year's growth against this year's ratio
signif(lm1$coefficient,3)
```

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

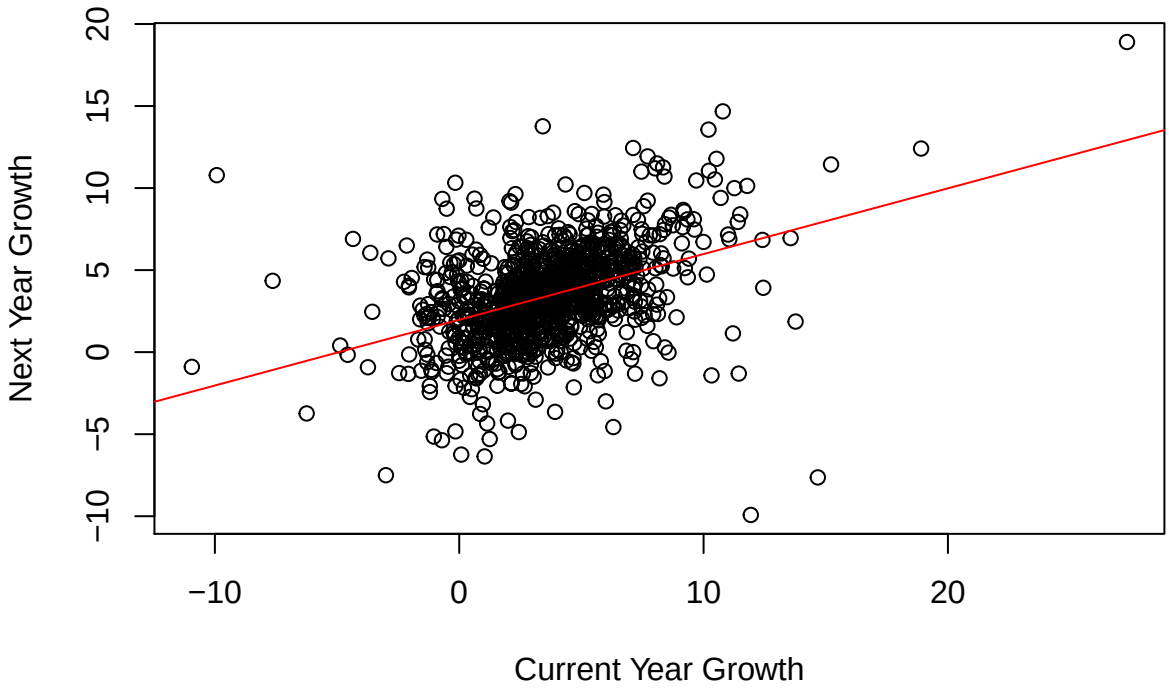
```
# regression of current year's growth against current year's ratio
signif(lm0$coefficients,3)
```

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

They have similar coefficients, both with negative slope about -0.01 and an intercept about 4.

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?

```
lm2 <- lm(next.growth~growth,data=debt)
plot(x=debt$growth,y=debt$next.growth,xlab='Current Year Growth',
     ylab='Next Year Growth')
abline(lm2,col='red')
```




```
signif(lm2$coefficients,3)
```

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

```
# previous regression
summary(lm1)[['adj.r.squared']]
```

```
## [1] 0.01685574
```

```
# current regression
summary(lm2)[['adj.r.squared']]
```

```
## [1] 0.166409
```

Current regression has a much larger adjusted R-squared. So I believe current growth is a better prediction for future growth.