# Homework 4 Solutions

Gross domestic product (GDP) is a measure of the total market value of all goods and services produced in a given country in a given year. The percentage growth rate of GDP in year $t$ is

$$100 \times \left(\frac{GDP_{t+1} - GDP_t}{GDP_t}\right) - 100$$

An important claim in economics is that the rate of GDP growth is closely related to the level of government debt, specifically with the ratio of the government's debt to the GDP. The file `debt.csv` on the class website contains measurements of GDP growth and of the debt-to-GDP ratio for twenty countries around the world, from the 1940s to 2010. Note that not every country has data for the same years, and some years in the middle of the period are missing data for some countries but not others. **Throughout, use 3 significant digits for numerical answers!!** (That is, `signif(mydat,3)` is your friend).

```
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. 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.

    a. Begin by writing a function, `mean.growth()`, that takes a data frame as its argument and returns the mean of the `growth` column of that data frame.

```
library(plyr)
mean.growth <- function(country.df) {
  return(signif(mean(country.df$growth), 3))
}
```

```
b. Use `daply()` to apply `mean.growth()` to each country in `debt`. You should not need to use a loop t
```

```
country.avgs <- daply(debt, .(Country), mean.growth)
country.avgs["Australia"]
```

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

```
country.avgs["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.) Make a plot of the growth rates (y-axis) versus the year (x-axis). Make sure the axes are labeled appropriately.

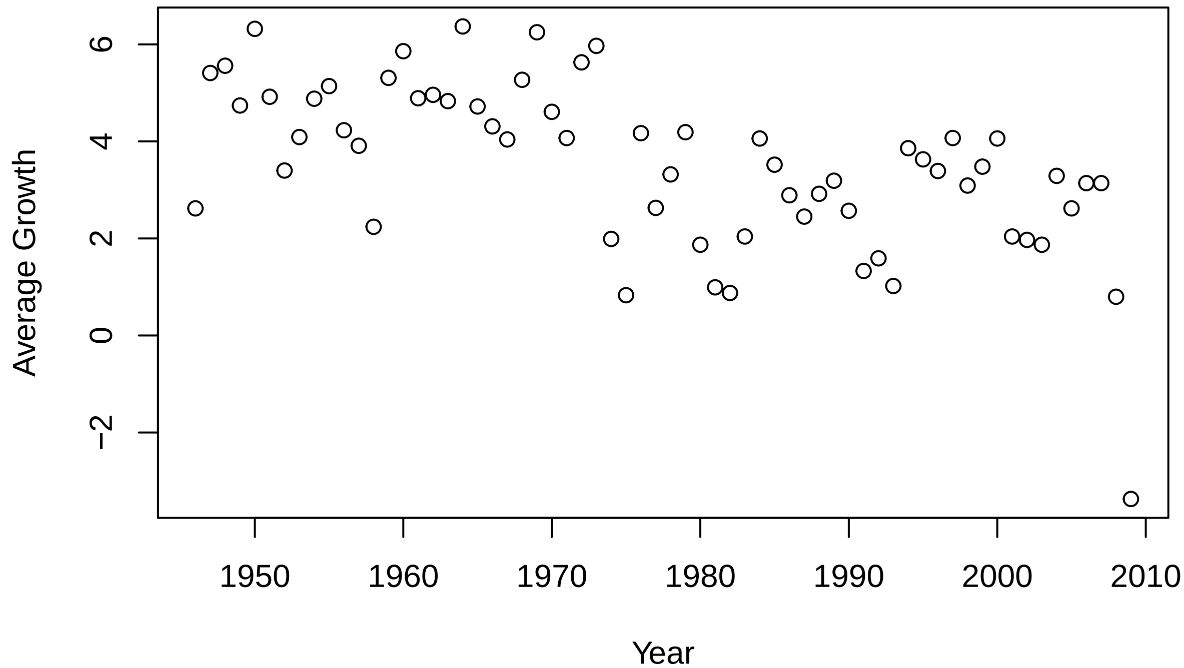
```r
year.avgs <- daply(debt, .(Year), mean.growth)
year.avgs["1972"]
```

```
## 1972
## 5.63
```

```r
year.avgs["1989"]
```

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

```r
plot(names(year.avgs), year.avgs, xlab = "Year", ylab = "Average Growth")
```




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 $-0.1995$.

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

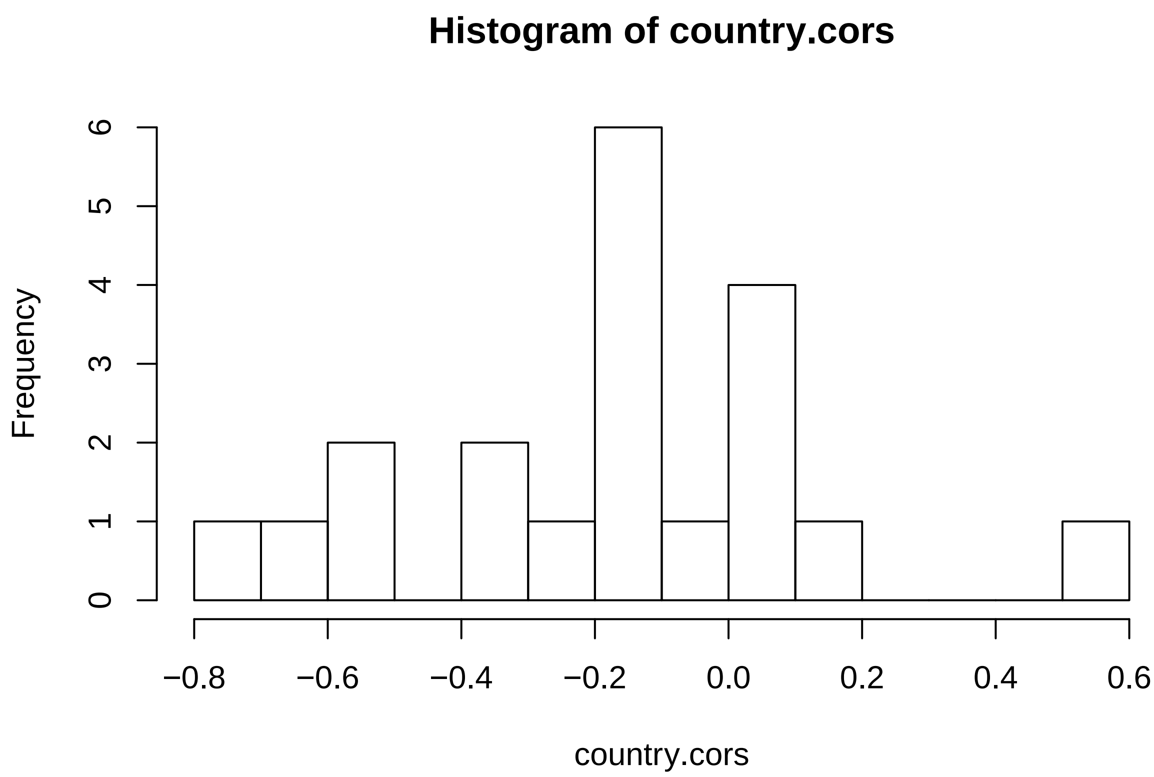
```
## [1] -0.199
```

```
b. Compute the correlation coefficient separately for each country, and plot a histogram of these coeffi
```

```r
cor.calc <- function(country.df) {
  return(signif(cor(country.df$growth, country.df$ratio), 3))
}
country.cors <- daply(debt, .(Country), cor.calc)
mean(country.cors)
```

```
## [1] -0.177766
```

```r
hist(country.cors,breaks=10)
```

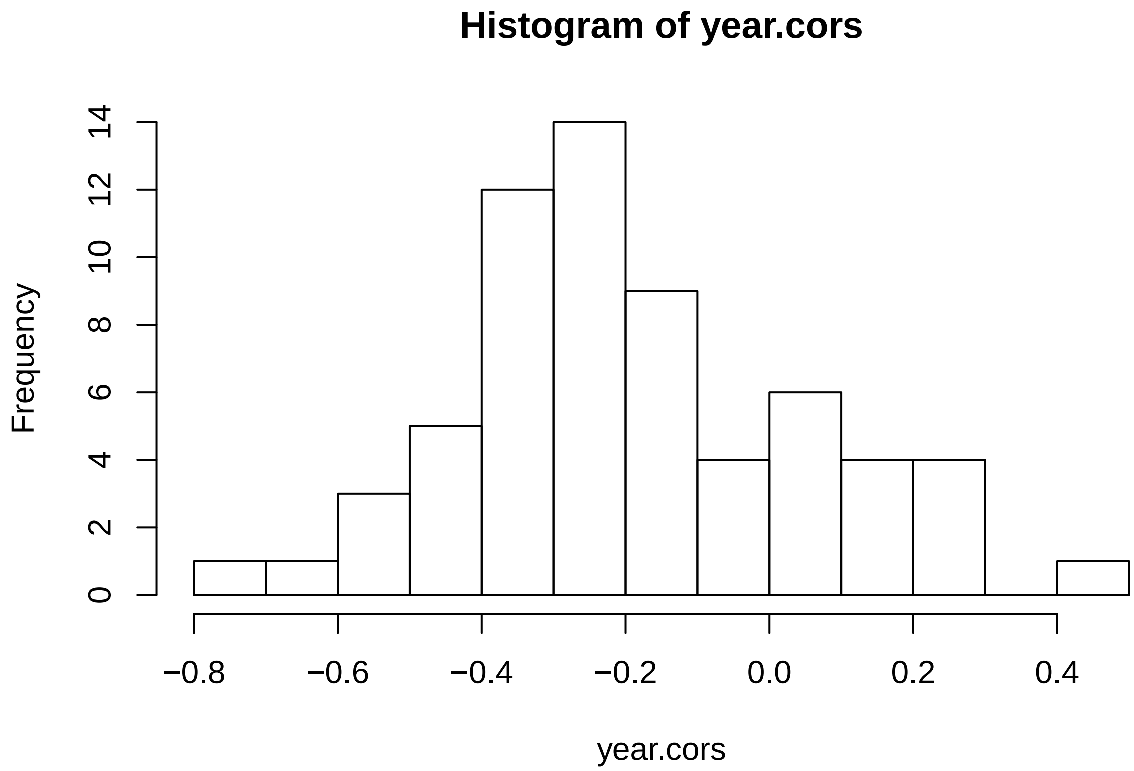



```
c. Calculate the correlation coefficient separately for each year, and plot a histogram of these coeffi
```

```r
year.cors <- daply(debt, .(Year), cor.calc)
mean(year.cors)
```

```
## [1] -0.190525
```

```r
hist(year.cors,breaks=10)
```




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

```r
sort(country.cors)
```

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

```r
sort(year.cors)
```

```
##     1957     1946     1961     1970     1960     1956     1958     1985
## -0.75500 -0.62000 -0.53900 -0.51200 -0.50400 -0.45800 -0.45400 -0.44900
##     1979     1951     1962     1993     1983     1964     1986     1996
## -0.42900 -0.41600 -0.38300 -0.37200 -0.36200 -0.36100 -0.35800 -0.35700
##     2002     2007     1948     2005     1965     1966     1959     1967
## -0.34900 -0.34400 -0.34000 -0.31400 -0.31100 -0.31100 -0.28500 -0.27800
##     1952     1954     1947     1998     1999     1969     2001     1955
## -0.27700 -0.27500 -0.27400 -0.26500 -0.25800 -0.25000 -0.23800 -0.22700
##     1994     1953     2009     1949     1972     2006     1968     1976
## -0.22400 -0.20500 -0.20500 -0.20000 -0.19600 -0.19600 -0.18100 -0.17100
##     2004     1984     2000     1980     1997     2008     1987     2003
## -0.17100 -0.15600 -0.13400 -0.12700 -0.11100 -0.09450 -0.06890 -0.06790
##     1992     1971     1981     1950     1995     1989     1988     1973
## -0.00222  0.00872  0.03040  0.03980  0.05190  0.06640  0.07970  0.11400
##     1963     1990     1977     1991     1982     1974     1975     1978
##  0.12800  0.15600  0.16400  0.20200  0.23900  0.26000  0.27100  0.43100
```

```
*Solution* Norway stands out for having a particularly large, positive correlation.  1978 seems to be pa
```

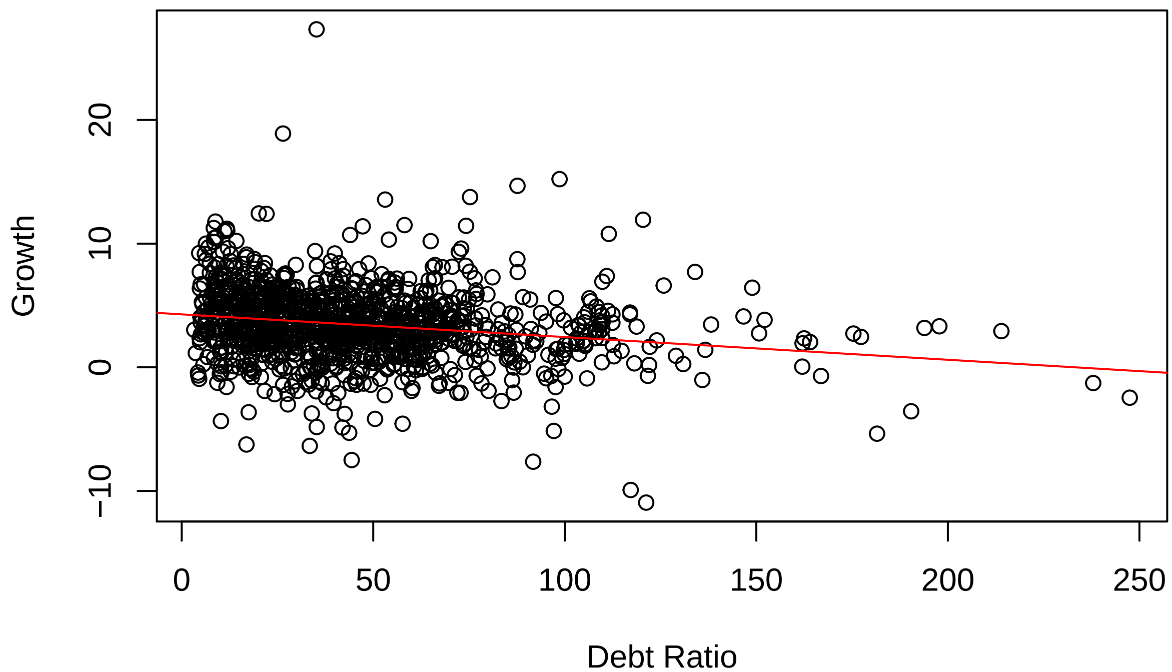
4. Fit a linear model of overall growth on the debt ratio, using `lm()`. Report the intercept and slope. Make a scatter-plot of overall GDP growth (vertical) against the overall debt ratio (horizontal). Add a line to your scatterplot from question 5 showing the fitted regression line.

```r
plot(debt$ratio, debt$growth, xlab = "Debt Ratio", ylab = "Growth")

lm0 <- lm(debt$growth ~ debt$ratio)
lm0$coef
```

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

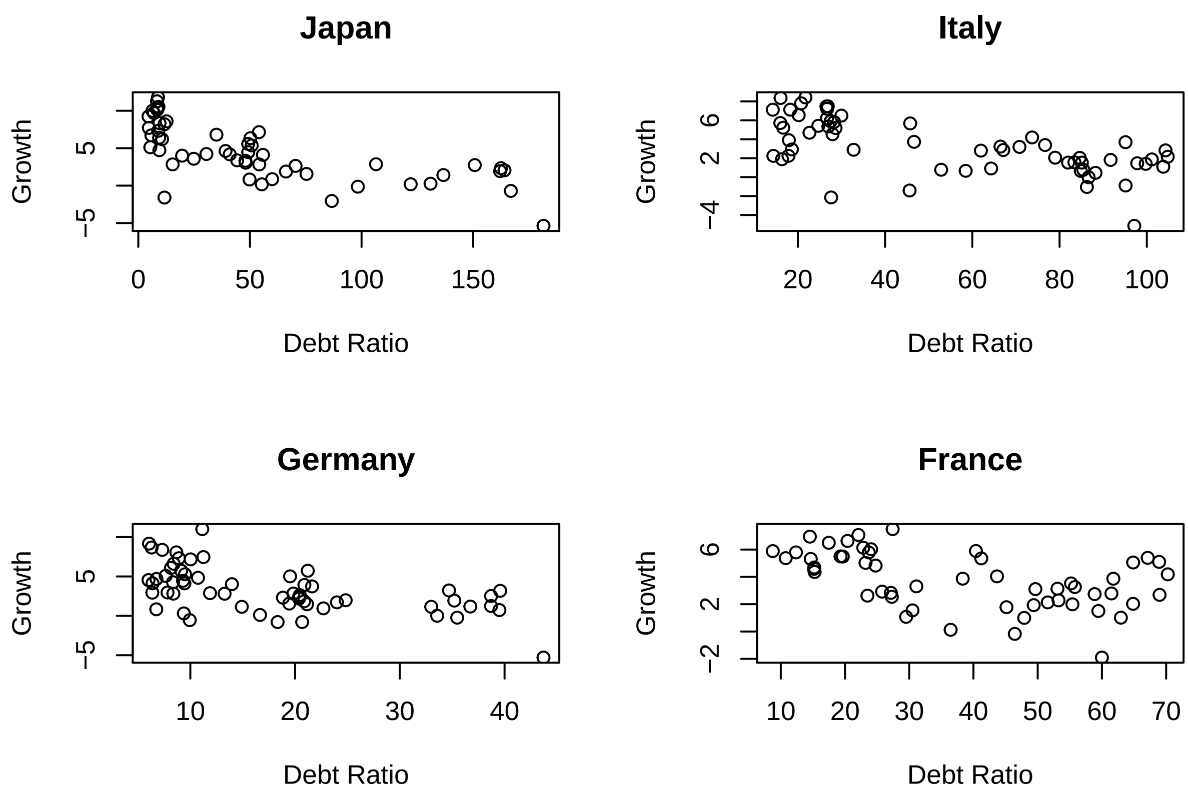
```r
abline(lm0, col = "red")
```




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. Call `par(mfrow=c(2,2))` before plotting so all four plots will appear in the same figure.

(Think about what this shows: individual relationships at the country level are sometimes concealed or "smudged out" when data is aggregated over *all* groups (countries). This conveys the importance of careful analysis at a more granular group level, when such groupings are available!)

```r
par(mfrow=c(2,2))
four.countries <- names(sort(country.cors))[1:4]
for (i in 1:4) {
  plot(debt$ratio[debt$Country == four.countries[i]], debt$growth[debt$Country == four.countries[i]], xl
}
```




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 dataframe which just contains the rows of `debt` for France, but contains all those rows. It should have 54 rows and 4 columns. Note that some years are missing from the middle of this data set.

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

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

```
b. Create a new column in your dataframe for France, `next.growth`, which gives next year's growth _if_
```

```
next.growth <- function(year, country.df) {

  if(any(country.df$Year == (year + 1))) {
    return(country.df$growth[country.df$Year == (year + 1)])
  } else {
    return(NA)
  }
}

debt.fr$next.growth <- sapply(debt.fr$Year, next.growth, debt.fr)
debt.fr$next.growth[debt.fr$Year  == 1971]
```

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

```
debt.fr$next.growth[debt.fr$Year  == 1972]
```

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

You can also use a loop

```
next.growth.loop <- function(country.df) {

  my.years <- country.df$Year
  next.growth <- rep(NA,length(my.years))
  counter <- 1

  for (year in my.years) {

    if (any(my.years==year+1)) {next.growth[counter] <- country.df$growth[counter+1]}
    else (next.growth[counter] <- NA)
    counter <- counter+1
  }

  country.df$next.growth <- next.growth
  return(country.df)

}

# Test the loop on France
debt.fr <- debt[debt$Country == "France", ]

debt.fr <- next.growth.loop(debt.fr)
debt.fr$next.growth[debt.fr$Year  == 1971]
```

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

```
debt.fr$next.growth[debt.fr$Year  == 1972]
```

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

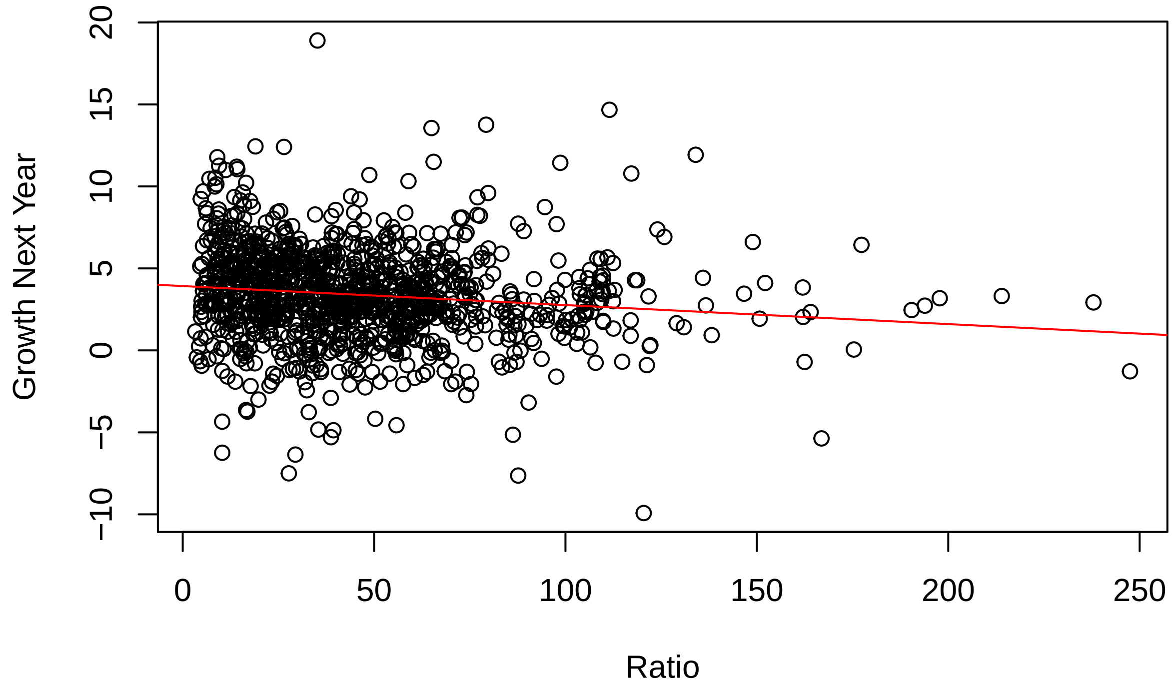
7. Add a `next.growth` column, as in question 4, to the *whole* of the `debt` data frame. Make sure that you do not accidentally put the first growth value for one country as the `next.growth` value for another. (The `next.growth` for France in 2009 should be `NA`, not 9.167.) *Hints*: Write a function to encapsulate what you did in question 4, and apply it using `ddply()`.

```
next.growth.all <- function(country.df) {
  country.df$next.growth <- sapply(country.df$Year, next.growth, country.df)
  return(country.df)
}
debt <- ddply(debt, .(Country), next.growth.all)
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?

```
plot(debt$ratio, debt$next.growth, xlab = "Ratio", ylab = "Growth Next Year")
lm1 <- lm(debt$next.growth ~ debt$ratio)
abline(lm1, col = "red")
```




```
lm0$coef
```

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

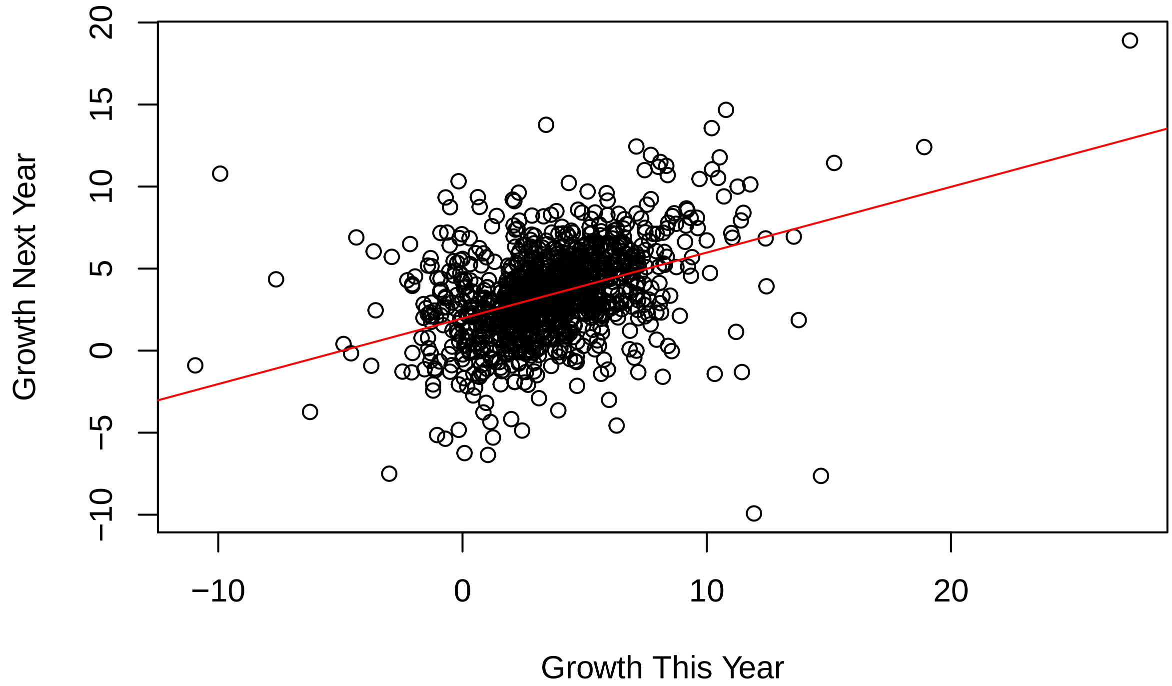
```
lm1$coef
```

```
## (Intercept)  debt$ratio
##  3.92472155 -0.01160842
```

The two regressions are similar. Both have a slightly negative slope with an intercept somewhere around 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 question 9), whether current growth or current debt is a better predictor of future growth?

```
plot(debt$growth, debt$next.growth, xlab = "Growth This Year", ylab = "Growth Next Year")
lm2 <- lm(debt$next.growth ~ debt$growth)
abline(lm2, col = "red")
```




```
lm2$coef
```

```
## (Intercept) debt$growth
##   1.9710648   0.4006517
```

I can't tell! Though this regression has a slightly larger value of $R^2$ and slightly more significant coefficients.