HW4_cq2203

Chang Qu (cq2203)

2018/10/18

1. Calculate the average GDP growth rate for each country (averaging over years).

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.

```r
library(plyr)

debt <- read.csv("/Users/courtneyqu/Desktop/GR5206/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
# write function to calculate the avarage growth rate for a given data frame
mean.growth <- function(df){
  mean_growth <- signif(mean(df$growth), 3)
  return(mean_growth)
}
```

b. Use daply() to apply mean.growth() to each country in debt. Report the average GDP growth rates of Australia and Netherland.

```r
# use daply, split by country
country.mean.growth <- daply(debt, .(Country), mean.growth)
country.mean.growth
```

```
##   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
```

```r
country.mean.growth["Australia"]
```

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

```r
country.mean.growth["Netherlands"]
```

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

The above result gives the average GDP growth rate for each country

2. Calculate the average GDP growth rate for each year (now averaging over countries). Make a plot of the growth rates (y-axis) versus the year (x-axis). Make sure the axes are labeled appropriately.

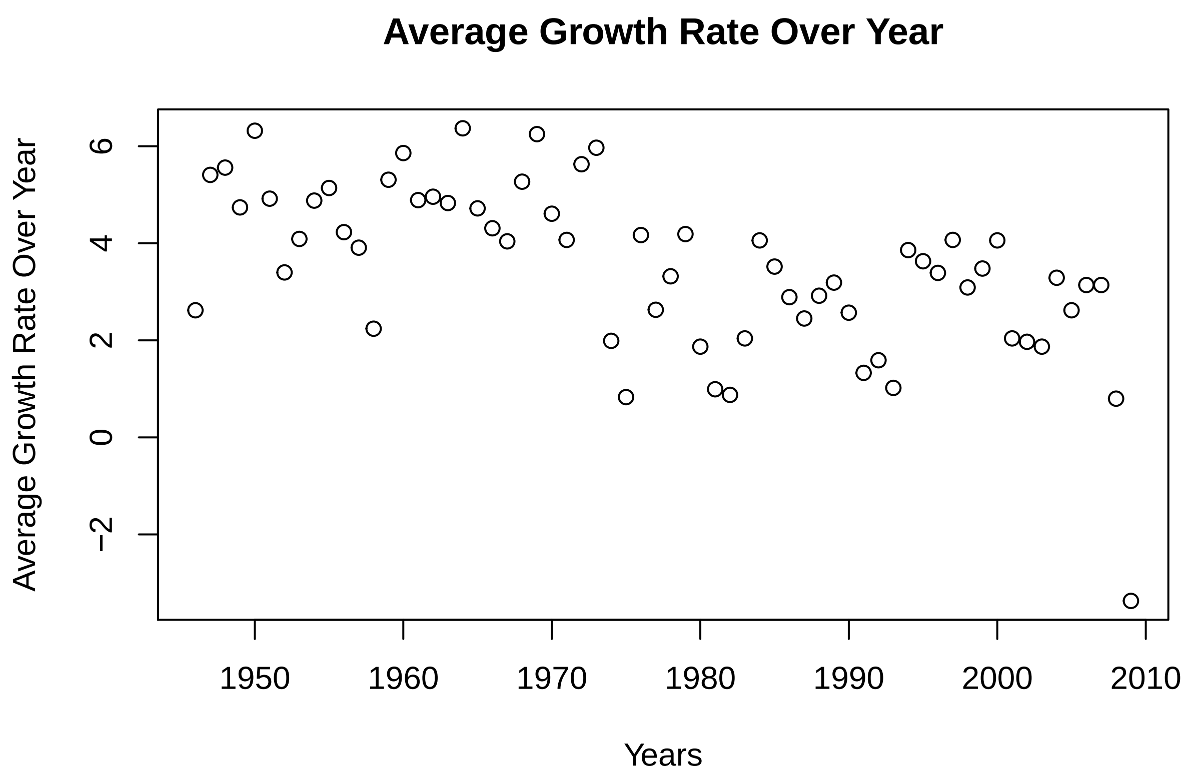
```r
#use daply, split by year
mean_rate_yr <- daply(debt, .(Year), mean.growth)

mean_rate_yr["1972"]
```

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

```r
mean_rate_yr["1989"]
```

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

```r
plot(levels(as.factor(debt$Year)), mean_rate_yr, xlab = "Years", ylab = "Average Growth Rate Over Year"
```



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).

```
cor_whole <- signif(cor(debt$growth, debt$ratio), 4)
cor_whole
```

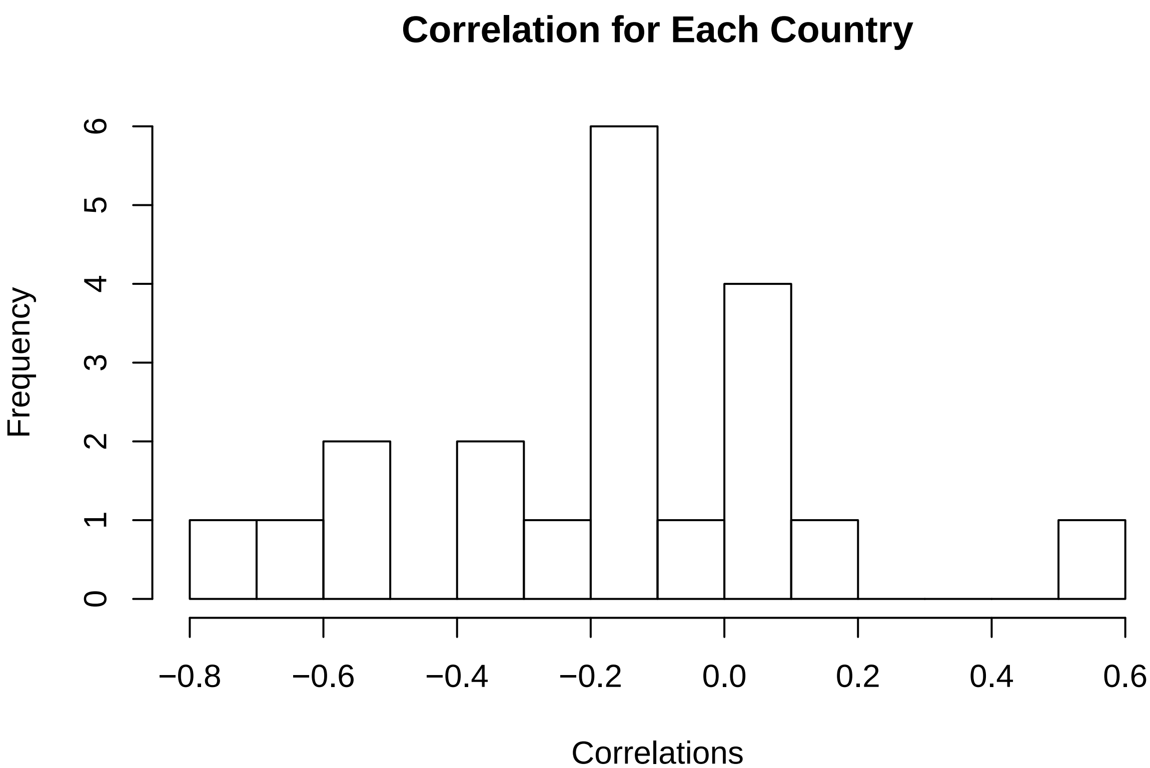
```
## [1] -0.1995
```

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

```
# write function to calculate correlation between growth rate and debt ratio
# of a given dataframe
correlation <- function(df){
  corr <- signif(cor(df$growth, df$ratio), 3)
  return(corr)
}

# use daply to split over country
corr_country <- daply(debt, .(Country), correlation)
signif(mean(corr_country),4)
```

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

```
hist(corr_country, xlab = "Correlations",  main = "Correlation for Each Country", breaks = 10)
```



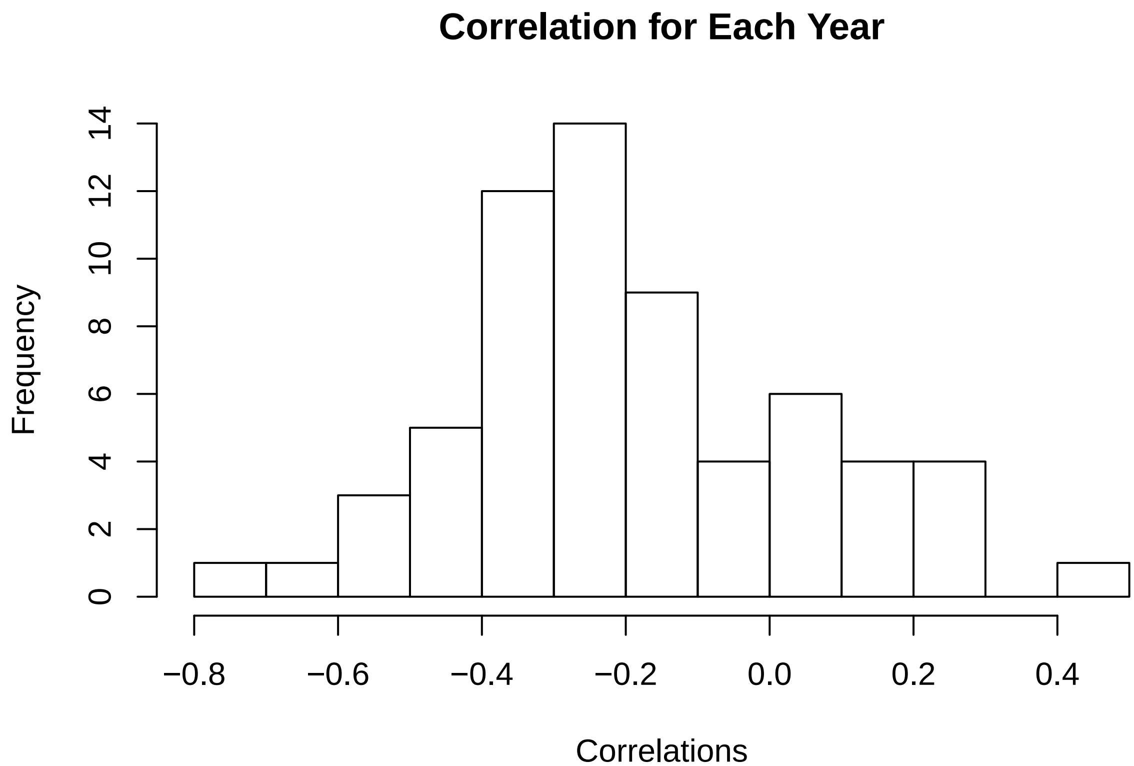
c. Calculate the correlation coefficient separately for each year, and plot a histogram of these coefficients.

```
# use daply to split over year
corr_year <- daply(debt, .(Year), correlation)
```

```r
signif(mean(corr_year),4)
```

```
## [1] -0.1905
```

```r
hist(corr_year, xlab = "Correlations",  main = "Correlation for Each Year", breaks = 10)
```



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

From the histgram, we can see that Norway with correlation 0.563, and year 1978 with correlation 0.431 seem to go against the general trend.

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

```r
lm_whole <- lm(debt$growth~debt$ratio)
summary(lm_whole)
```

```
##
## Call:
## lm(formula = debt$growth ~ debt$ratio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.9958  -1.5200  -0.0774   1.5707  23.6960
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.279290   0.148970   28.73   < 2e-16 ***
```

```
## debt$ratio  -0.018355   0.002637   -6.96 5.67e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.922 on 1169 degrees of freedom
## Multiple R-squared:  0.03979,    Adjusted R-squared:  0.03897
## F-statistic: 48.44 on 1 and 1169 DF,  p-value: 5.666e-12
```

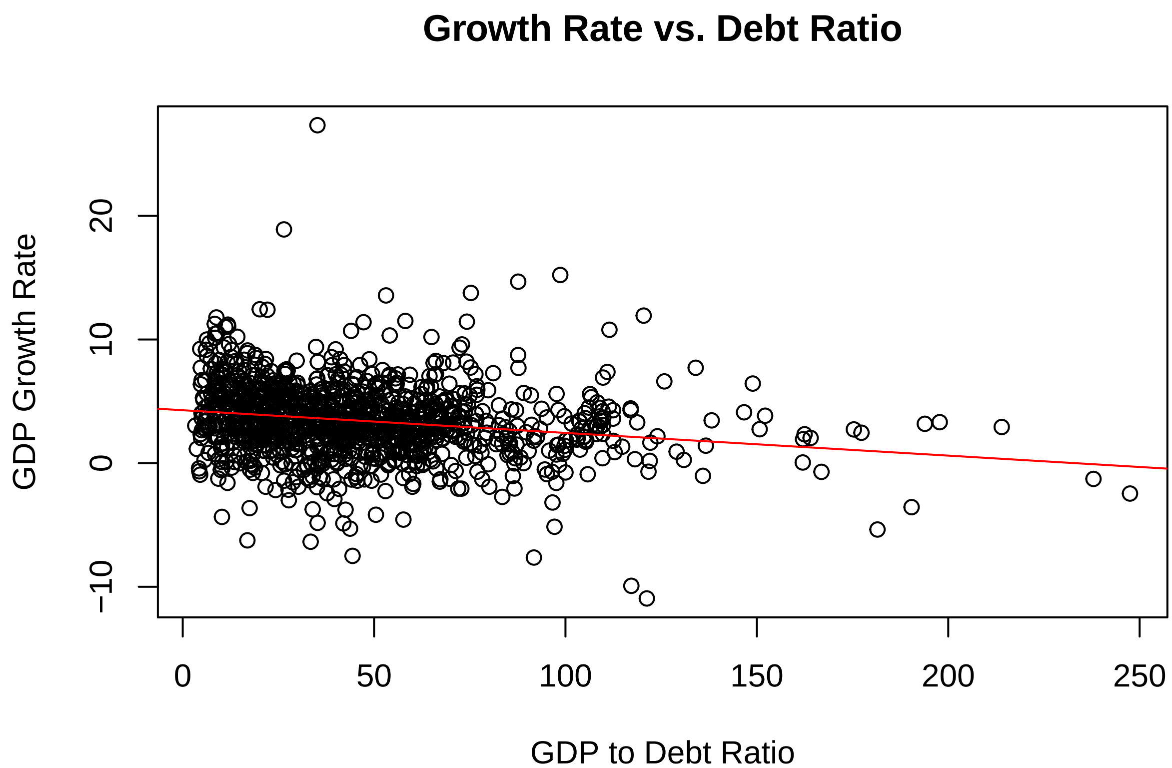
```r
b0 <- signif(lm_whole$coefficients[1],3)
b0
```

```
## (Intercept)
##        4.28
```

```r
b1 <- signif(lm_whole$coefficients[2],3)
b1
```

```
## debt$ratio
##    -0.0184
```

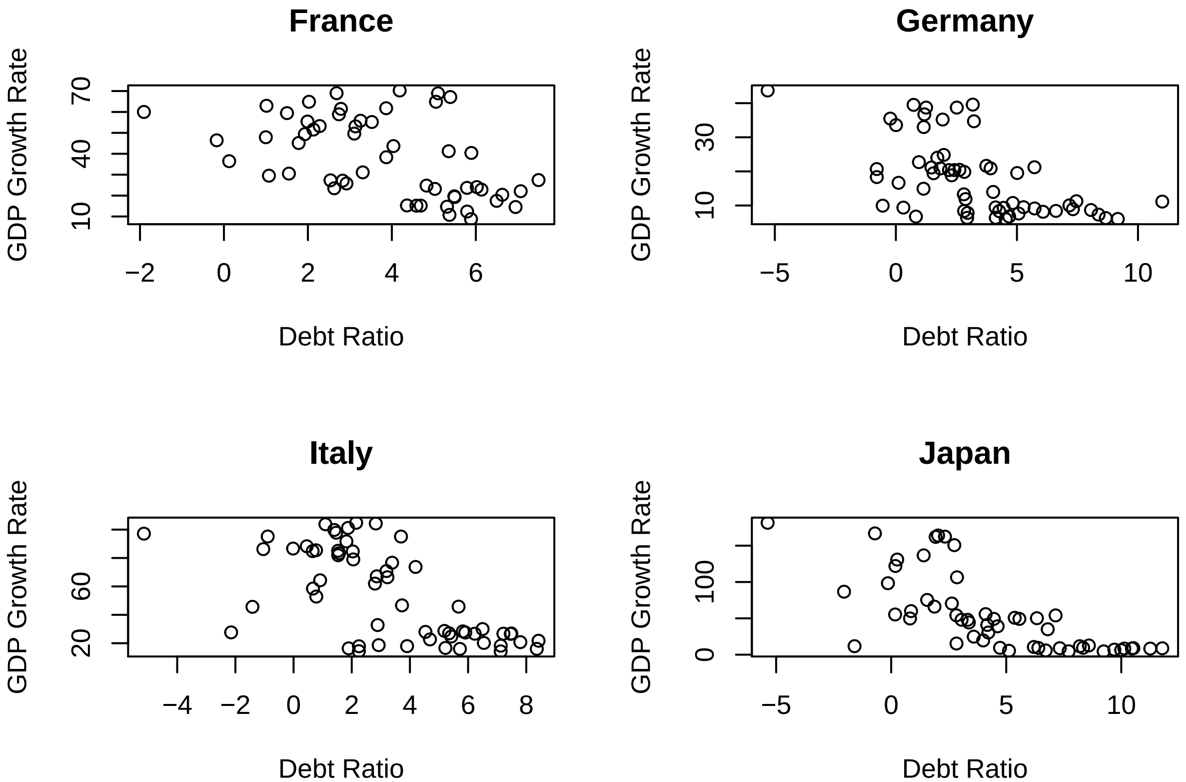
```r
plot(debt$ratio, debt$growth, xlab = "GDP to Debt Ratio", ylab = "GDP Growth Rate",
     main = "Growth Rate vs. Debt Ratio")
abline(lm_whole$coefficients[1], lm_whole$coefficients[2], col="red")
```



The intercept for the linear regression model is 4.28, the slope is -0.0184.

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.

```
four_country <- levels(factor(debt$Country))[which(corr_country < -0.5)]
debt_four <- rbind(debt[debt$Country=="France",],
                   debt[debt$Country=="Germany",],
                   debt[debt$Country=="Italy",],
                   debt[debt$Country=="Japan",])
plot.function <- function(df){
  plot(df$growth, df$ratio, xlab = "Debt Ratio", ylab = "GDP Growth Rate", main = levels(factor(df$Coun
}
par(mfrow=c(2,2))
d_ply(debt_four, .(Country), plot.function)
```



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.

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

```
## [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.

```
france_debt$Next.growth <- rep(NA, length(france_debt$Country))

for(i in 1:length(france_debt$growth)){
  france_debt$Next.growth[i] <- ifelse(france_debt$Year[i+1]==france_debt$Year[i]+1,
                                      france_debt$growth[i+1], NA)
}
france_debt[france_debt$Year==1971,"Next.growth"]
```

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

```
france_debt[france_debt$Year==1972,"Next.growth"]
```

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

7. Add a next.growth column, as in the previous question, to the whole of the debt data frame.

```
next.growth.func <- function(df){
  df$Next.growth <- rep(NA, length(df$growth))
  for(i in 1:length(df$growth)){
    df$Next.growth[i] <- ifelse(df$Year[i+1]==df$Year[i]+1, df$growth[i+1], NA)
  }
  return(df)
}

new.debt <- ddply(debt, .(Country), next.growth.func)
head(new.debt)
```

```
##     Country Year    growth     ratio Next.growth
## 1 Australia 1946 -3.557951 190.41908   2.4594746
## 2 Australia 1947  2.459475 177.32137   6.4375341
## 3 Australia 1948  6.437534 148.92981   6.6119938
## 4 Australia 1949  6.611994 125.82870   6.9202012
## 5 Australia 1950  6.920201 109.80940   4.2726115
## 6 Australia 1951  4.272612  87.09448   0.9046516
```

```
# next growth for France in 2009
new.debt[new.debt$Country=="France"&new.debt$Year=="2009", "Next.growth"]
```

```
## [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 compareto the regression of the current year's growth on the current year's debt ratio?

```
lm.next <- lm(Next.growth~ratio, data = new.debt)
summary(lm.next)
```

```
##
## Call:
## lm(formula = Next.growth ~ ratio, data = new.debt)
##
## 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 ***
## 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
```

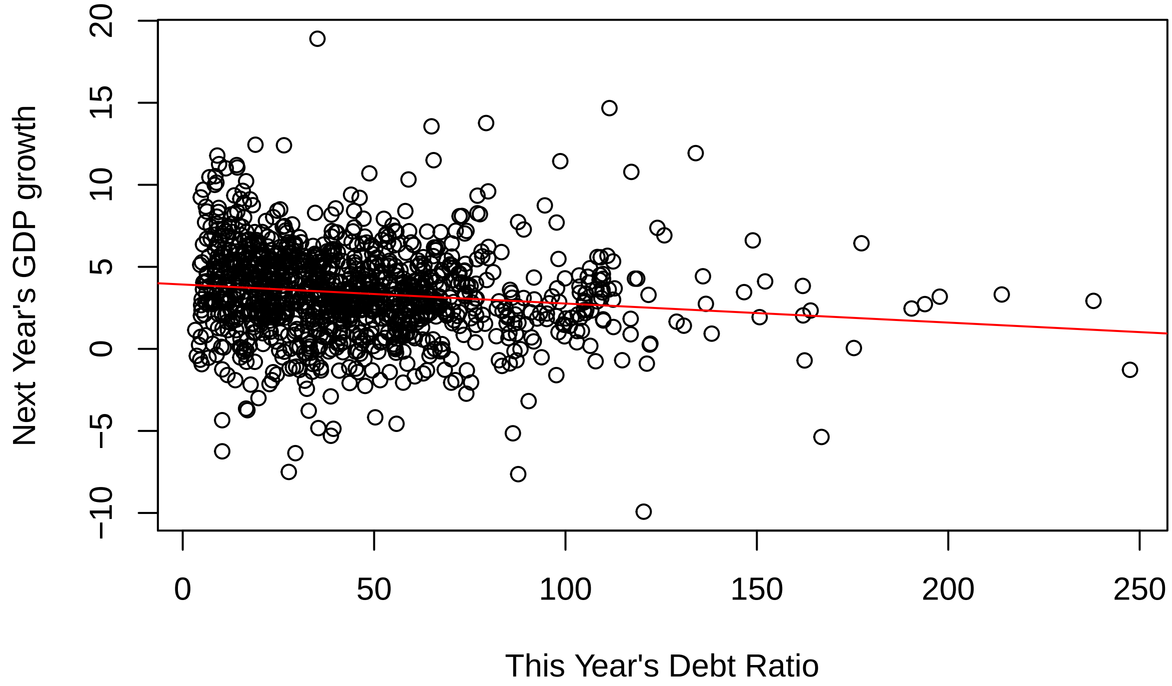
```r
b0.next <- signif(lm.next$coefficients[1],3)
b0.next
```

```
## (Intercept)
##        3.92
```

```r
b1.next <- signif(lm.next$coefficients[2],3)
b1.next
```

```
##   ratio
## -0.0116
```

```r
plot(new.debt$ratio, new.debt$Next.growth, ylab = "Next Year's GDP growth", xlab = "This Year's Debt Ra
abline(a=lm.next$coefficients[1], b=lm.next$coefficients[2], col = "red")
```



The intercept is 3.92, the slope is -0.0116.

Comparing to the regression of current year's growth and current year's debt ratio, the coefficient of determination– R-square is slightly lower than the one of previous model, which means the effect of next year's GDP growth in reducing the variance of current year's debt ratio is slightly lower than that of current year's GDP growth.

However, when doing the inferene test regarding the slope, both models show that there is a linear relationship between the predictor and response variable at 5% significant level.

Moreover, from the graph, we cannot see much of a difference of these two model.

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 debtis a better predictor of future growth?

```
lm.gdp <- lm(Next.growth~growth, data = new.debt)
summary(lm.gdp)
```

```
##
## Call:
## lm(formula = Next.growth ~ growth, data = new.debt)
##
## 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 ***
## 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
```

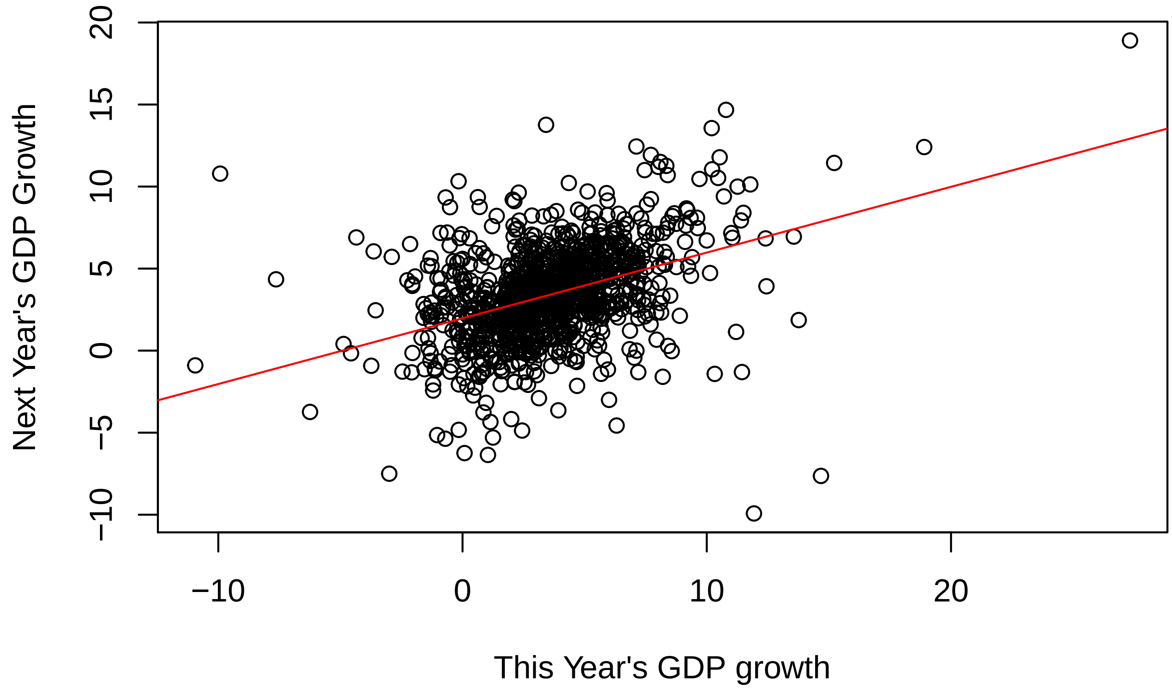
```
b0.gdp <- signif(lm.gdp$coefficients[1],3)
b0.gdp
```

```
## (Intercept)
##        1.97
```

```
b1.gdp <- signif(lm.gdp$coefficients[2],3)
b1.gdp
```

```
## growth
##  0.401
```

```
plot(new.debt$growth, new.debt$Next.growth, xlab = "This Year's GDP growth", ylab = "Next Year's GDP Gr
abline(a=lm.gdp$coefficients[1], b=lm.gdp$coefficients[2], col = "red")
```



The intercept is 1.97, the slope is 0.401.

Comparing this regression model with the previous one, This model shows more obvious relationship from the graph.

Morover, current model has a much more higher R-square than the previous model. The inference test on the current slope also shows that there is a linear relationship between the predictor and the response variable under 5% significant level. Therefore, current growth is a better predictor of next year's growth than current debt ratio.