

Handout 4

Put your name here!

In this handout, we will learn how to calculate correlations, plot points of different colors and types, and add a line to a scatterplot.

Topics and Concepts Covered

- Calculating a correlation coefficient
- Fitting a linear model
- Creating a scatterplot
- Adding a regression line
- Setting the color of points
- Setting the character plotted
- *Auto-correlation*. The correlation between a variable and its lagged value
- *Panel data*. A data structure where we observe each of the same set of units at the same time, over multiple time periods

R Commands Covered

- Calculating correlation coefficients with `cor`
- Estimating a linear model with `lm`
- Creating a scatterplot through `plot`
- Adjusting the color of points by setting the value of `col`
- Adjusting the character plotted by setting the value `pch`

Before beginning this handout, do not forget to set your working directory!

Introduction to Data

To study sovereign debt and economic growth, the economists Carmen Reinhart and Ken Rogoff (Reinhart and Rogoff 2010) have constructed an impressive data set that encompasses over 70 countries and 200 years worth of observations including variables such as inflation, exchange rates, debt-to-GDP ratios and categorical variables indicating whether a country is suffering from a debt, inflation or banking crisis. While the research is multi-faceted, the aspect that has received the most public attention is the claim that once a country's debt-to-GDP ratio crosses a threshold of approximately 90%, growth rates plummet.

This argument is not strictly academic; the current US debt-to-GDP ratio has almost doubled since the financial crisis and currently stands at about 105%. In Japan, the debt-to-GDP ratio is over 200%. If Reinhart and Rogoff's conclusions are correct then these countries could increase their GDP growth rates by cutting spending and/or raising taxes. This policy is referred to as "austerity". Policy makers argue over the potential benefits of austerity versus the possible harm that could come from cutting spending for government programs. US House Speaker Paul Ryan has even cited Reinhart and Rogoff when making arguments for cutting spending.

The following articles help explain the data and the debate involved. You don't have to focus on details, but these articles do give some useful background for really understanding this analysis:

- The original paper positing the 90% threshold: Reinhart and Rogoff (2010)
- Herndon et al.'s reply: Herndon, Ash, and Pollin (2014); you should focus on the first 10 pages or so
- Stephen Colbert weighs in on the issue

A few more resources, if you find the topic interesting, are:

- Reinhart and Rogoff respond in the New York Times to the Herndon et al. paper

- Herndon's co-authors respond to their response, also in the New York Times
- An early critique by Irons and Bivens: (Irons and Bivens 2010) (link). See especially the section on causality
- A blog post on 'Rortybomb' laying out the issues involved

Correlations and Linear Models

You will find the file `RR.handout.csv` in the data folder. The file contains a subset of the variables considered by Herndon et al.:

- `Country`. Country name
- `Year`. Year of observation
- `dRGDP` Percent difference in real GDP (i.e., inflation-adjusted) from last year to the *current* year
- `debtgdp` The debt-to-GDP ratio for the *current* year
- `lag.dRGDP` The percentage difference in real GDP (i.e., inflation-adjusted) from last year to the *previous* year
- `lag.debtgdp` The debt-to-GDP ratio for the *previous* year

These data are an example of **panel data**, a type of data where the same observations are made over multiple time periods. Identifying causal effects in panel data, such as estimating whether GDP shifts affect debt, vice versa, neither, or both, is among the trickiest of problems in social science. In this handout, we're going to conduct an analysis that (begins) to help address this causal question.

First, we read in the data, and make sure we loaded it in properly.

```
RR <- read.csv("data/RR.handout.csv")
head(RR)
```

	Country	Year	dRGDP	debtgdp	lag.dRGDP	lag.debtgdp
1	Australia	1947	2.4594746	177.32137	-3.557951	190.41908
2	Australia	1948	6.4375341	148.92981	2.459475	177.32137
3	Australia	1949	6.6119938	125.82870	6.437534	148.92981
4	Australia	1950	6.9202012	109.80940	6.611994	125.82870
5	Australia	1951	4.2726115	87.09448	6.920201	109.80940
6	Australia	1952	0.9046516	86.06644	4.272612	87.09448

```
summary(RR)
```

	Country	Year	dRGDP	debtgdp
Australia	: 63	Min. :1947	Min. : -9.922	Min. : 3.279
Canada	: 63	1st Qu.:1965	1st Qu.: 1.919	1st Qu.: 22.148
Finland	: 63	Median :1981	Median : 3.258	Median : 40.401
New Zealand	: 63	Mean :1980	Mean : 3.389	Mean : 45.916
Norway	: 63	3rd Qu.:1995	3rd Qu.: 5.065	3rd Qu.: 61.385
Sweden	: 63	Max. :2009	Max. :18.902	Max. :237.941
(Other)	:769			
lag.dRGDP		lag.debtgdp		
Min.	: -10.942	Min.	: 3.279	
1st Qu.:	1.981	1st Qu.:	22.083	
Median :	3.319	Median :	40.389	
Mean :	3.538	Mean :	46.172	
3rd Qu.:	5.109	3rd Qu.:	61.474	
Max.	: 27.329	Max.	:247.482	

Correlations and the Line of Best Fit

Calculating correlations in R is done with the `cor` function. It takes two arguments `x` and `y`, as

```
cor(x, y)
```

For example, if we wanted to know the correlation between current change in GDP and the current debt ratio, we would use

```
cor(RR$debtgdp, RR$dRGDP)
```

```
[1] -0.1914904
```

which gives a correlation of approximately -0.19. So, current debt-to-GDP ratios and current growth show a negative relationship. We could also look at the correlation between past and current debt-to-GDP ratios:

```
cor(RR$debtgdp, RR$lag.debtgdp)
```

```
[1] 0.9892828
```

and growth:

```
cor(RR$dRGDP, RR$lag.dRGDP)
```

```
[1] 0.4088232
```

It appears that the debt-to-GDP ratio is highly *autocorrelated*. In other words, its past and current values correlate at a really high level. The problem is less acute (though still present) for growth rates.

Autocorrelation creates a problem in inference, as it becomes difficult to disentangle whether the causal effect is driven by the *current* levels or *past* levels of the variable. When the correlation is so high, it becomes easy to confuse one effect with the other. One common way of reducing autocorrelation is to take a “first difference”:

```
cor(RR$debtgdp - RR$lag.debtgdp, RR$lag.debtgdp)
```

```
[1] -0.2403264
```

The first difference of debt ratio (`RR$debtgdp - RR$lag.debtgdp`) correlates with previous debt ratio at about -0.24 , which is much less (in magnitude) than the original correlation of approximately 0.99.

Calculating linear models in R is done using the `lm` function. It takes one “formula” argument that connects one variable `y` to one (or more) variables `x`, as

```
lm(y ~ x)
```

Here, `y` is the outcome, or dependent, variable and `x` is a single independent variable. As mentioned above, the argument passed to the `lm` function is of the form `y ~ x`, where `~` shows that it is a formula. In other words, we are assuming, for observation y_i , a model of the form

$$y_i = \alpha + \beta x_i + \epsilon_i$$

The `lm` function allows us to estimate the coefficients $\hat{\alpha}$ and $\hat{\beta}$ such that the line is as close to the data as possible.

For example, let’s say we wanted to find the line of best fit from regressing GDP growth on debt ratios.

```
lm1 <- lm(RR$dRGDP ~ RR$debtgdp)
lm1
```

Call:

```
lm(formula = RR$dRGDP ~ RR$debtgdp)
```

Coefficients:

```
(Intercept)    RR$debtgdp
      4.17512      -0.01713
```

where `lm1` is an object that contains all the information we need about the linear regression. According to our model, the estimated marginal effect of a one percentage point increase in debt ratio on GDP growth is captured by the estimated coefficient $\hat{\beta} \approx -0.017$.

If we wanted to access the coefficients from the regression, we can use them directly within `lm1` object

```
lm1$coef # Both coefficients
```

```
(Intercept)    RR$debtgdp
      4.1751238    -0.0171267
```

```
lm1$coef[1] # The intercept
```

```
(Intercept)
      4.175124
```

```
lm1$coef[2] # The slope
```

```
RR$debtgdp
-0.0171267
```

or we could use the `coef` function to take a copy of the model coefficients

```
mycoefs <- coef(lm1) # Both coefficients
mycoefs
```

```
(Intercept)    RR$debtgdp
      4.1751238    -0.0171267
```

```
mycoefs[1] # The intercept
```

```
(Intercept)
      4.175124
```

```
mycoefs[2] # The slope
```

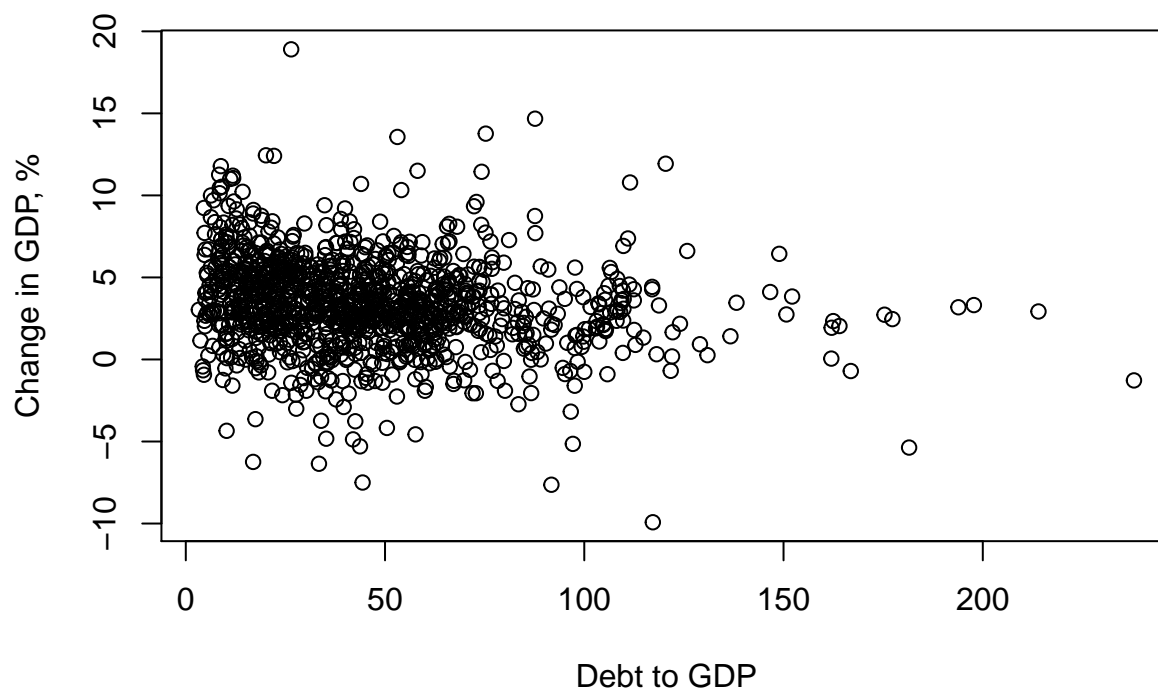
```
RR$debtgdp
-0.0171267
```

Adding a Regression Line and Coloring Points in a Figure

In this section, we are going to create scatterplots, which are a plot of a variable x against a variable y . We will then change the color and plotting character of subsets of the points, and then add a regression line to the figure.

```
plot(RR$debtgdp, RR$dRGDP,
     xlab = "Debt to GDP", ylab = "Change in GDP, %",
     main = "Debt-to-GDP Ratio versus Growth")
```

Debt-to-GDP Ratio versus Growth

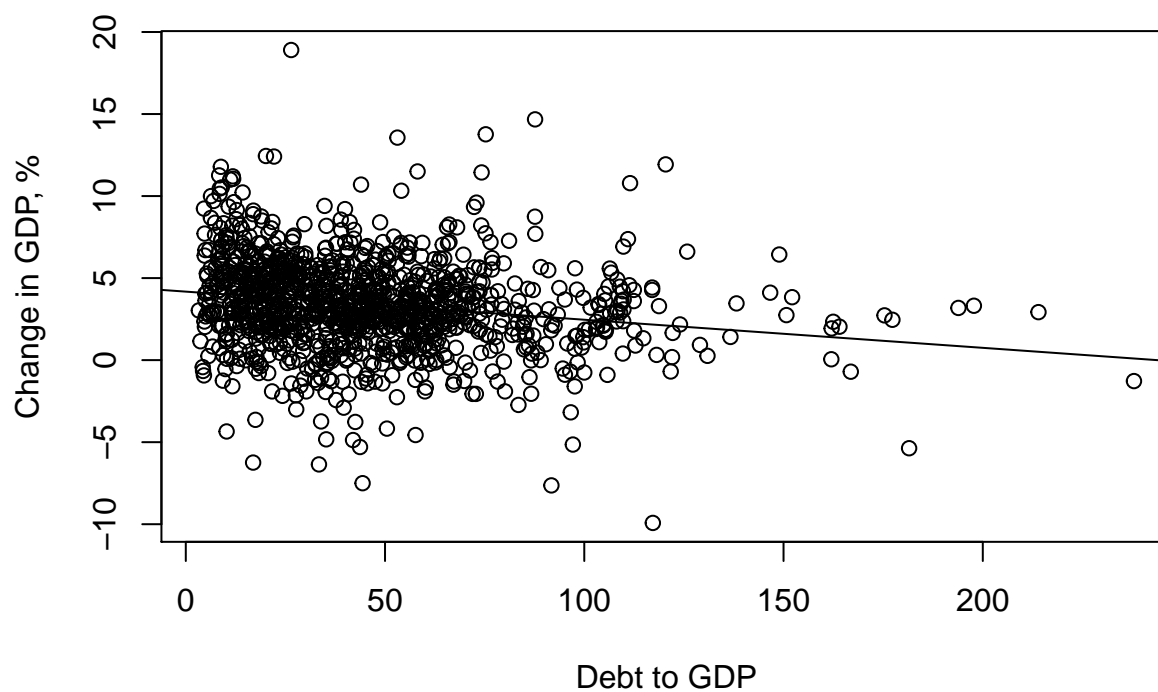


Recall

that we have the object `lm1` defined above. We can add this line to the plot using the `abline` function

```
plot(RR$debtgdp, RR$dRGDP,  
     xlab = "Debt to GDP", ylab = "Change in GDP, %",  
     main = "Debt-to-GDP Ratio versus Growth")  
abline(lm1)
```

Debt-to-GDP Ratio versus Growth



Coding tip

Be careful when adding a line to a figure! The plot command takes the form

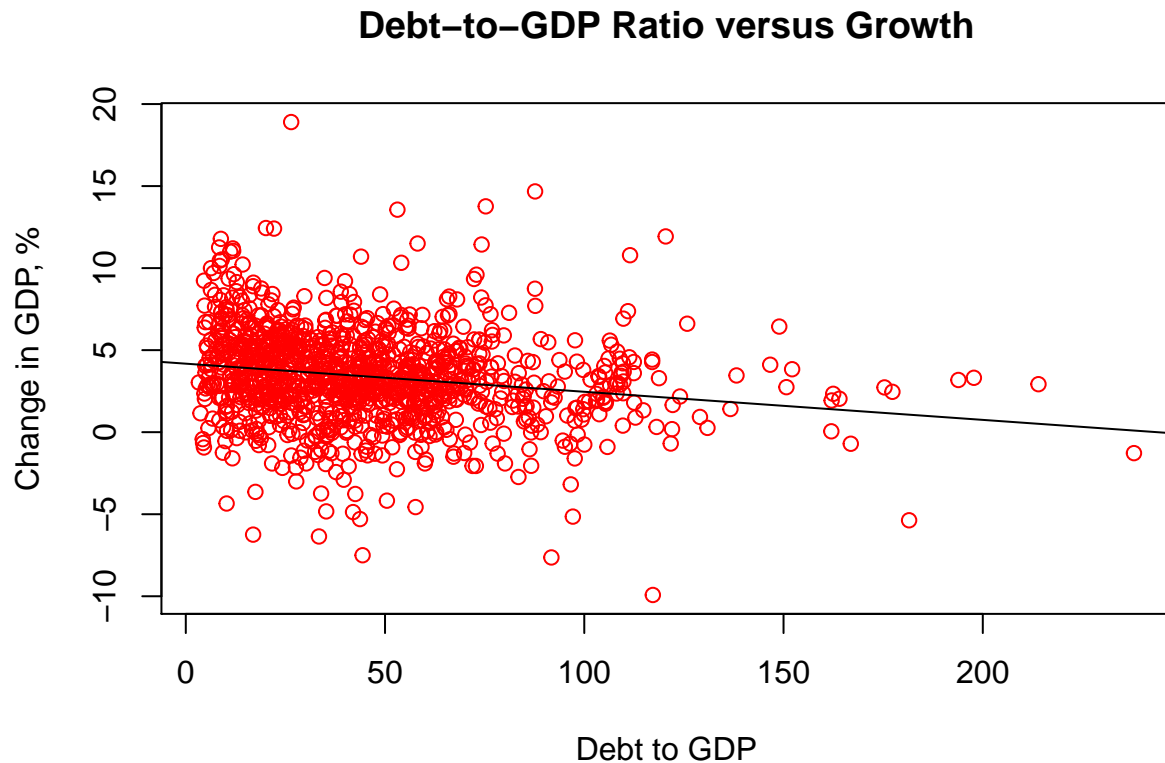
```
plot(x, y)
```

while the two variables in the formula you give to `lm` are in the opposite order:

```
lm(y ~ x)
```

Next, we are going to add some color to the figure. We can do this by specifying a color with `col`:

```
plot(RR$debtgdp, RR$dRGDP,
     xlab = "Debt to GDP", ylab = "Change in GDP, %",
     main = "Debt-to-GDP Ratio versus Growth", col = "red")
abline(lm1)
```

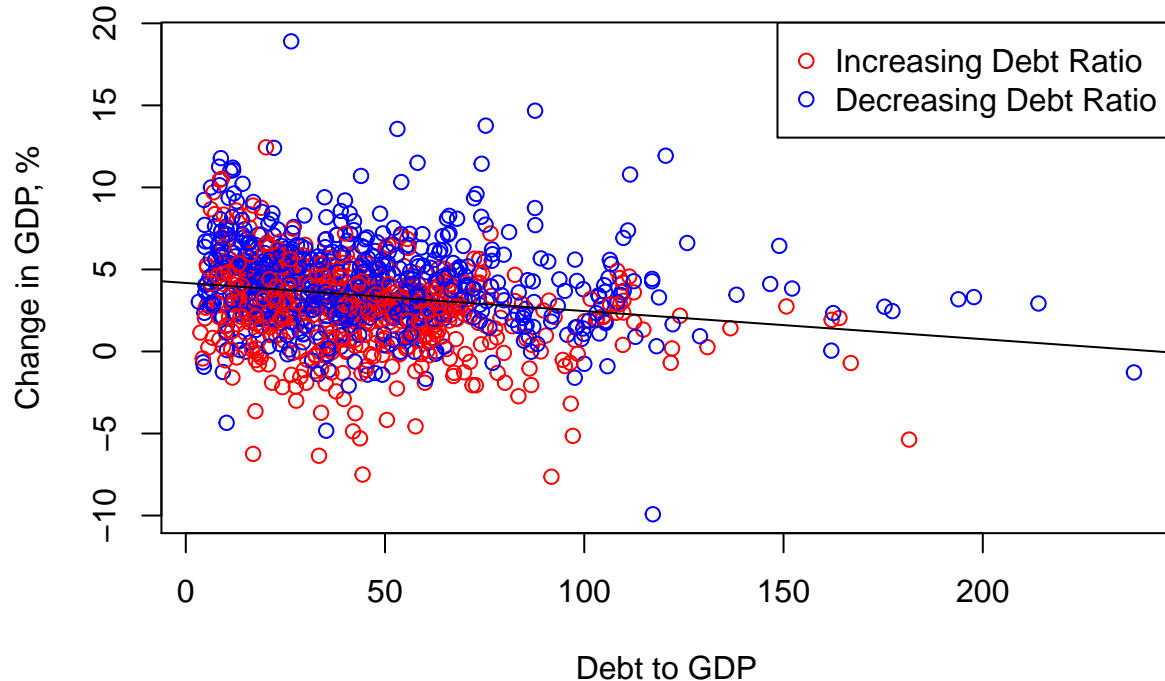


We can make the figure more interesting by giving different subsets of the data their own color.

```
col.plot <- ifelse(RR$debtgdp > RR$lag.debtgdp, "red", "blue")

plot(RR$debtgdp, RR$dRGDP,
     xlab = "Debt to GDP", ylab = "Change in GDP, %",
     main = "Debt-to-GDP Ratio versus Growth", col = col.plot)
abline(lm1)
legend(c("topright"), c("Increasing Debt Ratio", "Decreasing Debt Ratio"),
      col = c("red", "blue"), pch = 1)
```

Debt-to-GDP Ratio versus Growth



Notice that the legend took the option `pch`. This command is short for `plotting character`, and controls the type of character plotted on the figure. We are going to use two different numbers for `pch`

- 1. The default, an open circle
- 19. A solid circle

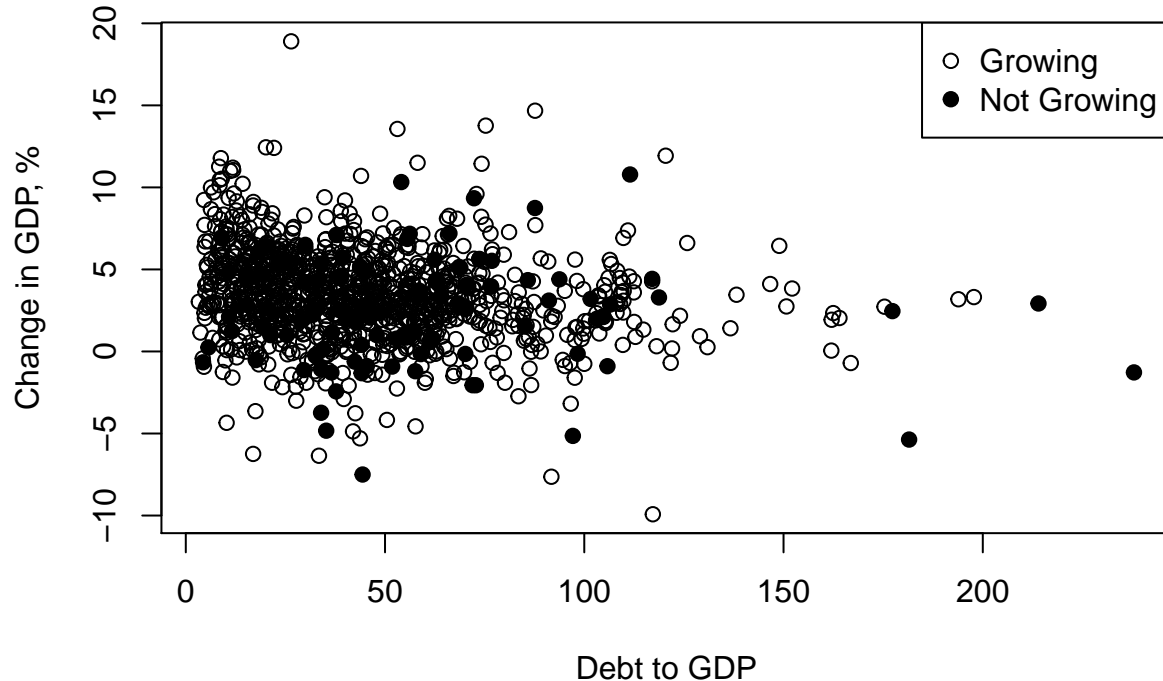
We can also set `pch` to a letter: "A", "f", etc. This will use the letter A or f as the point character.

Next, we are going to use solid points for observations that were growing in the previous period (`RR$lag.dRGDP > 0`), and open circles for observations that were not growing in the previous period. To do so, we are going to make a variable named `pch.plot` that takes on a value of 1 if previous growth is positive, and a 19 otherwise. We then pass that variable, `pch.plot` as the value of `pch`:

```
pch.plot <- ifelse(RR$lag.dRGDP > 0, 1, 19)

plot(RR$debtgdp, RR$dRGDP,
     xlab = "Debt to GDP", ylab = "Change in GDP, %",
     main = "Debt-to-GDP Ratio versus Growth", pch = pch.plot)
legend(c("topright"), c("Growing", "Not Growing"),
     pch = c(1, 19))
```

Debt-to-GDP Ratio versus Growth



Debt and GDP: Exploring the Relationship

In this section, we are going to make a figure with one row and two columns. The two figures will plot

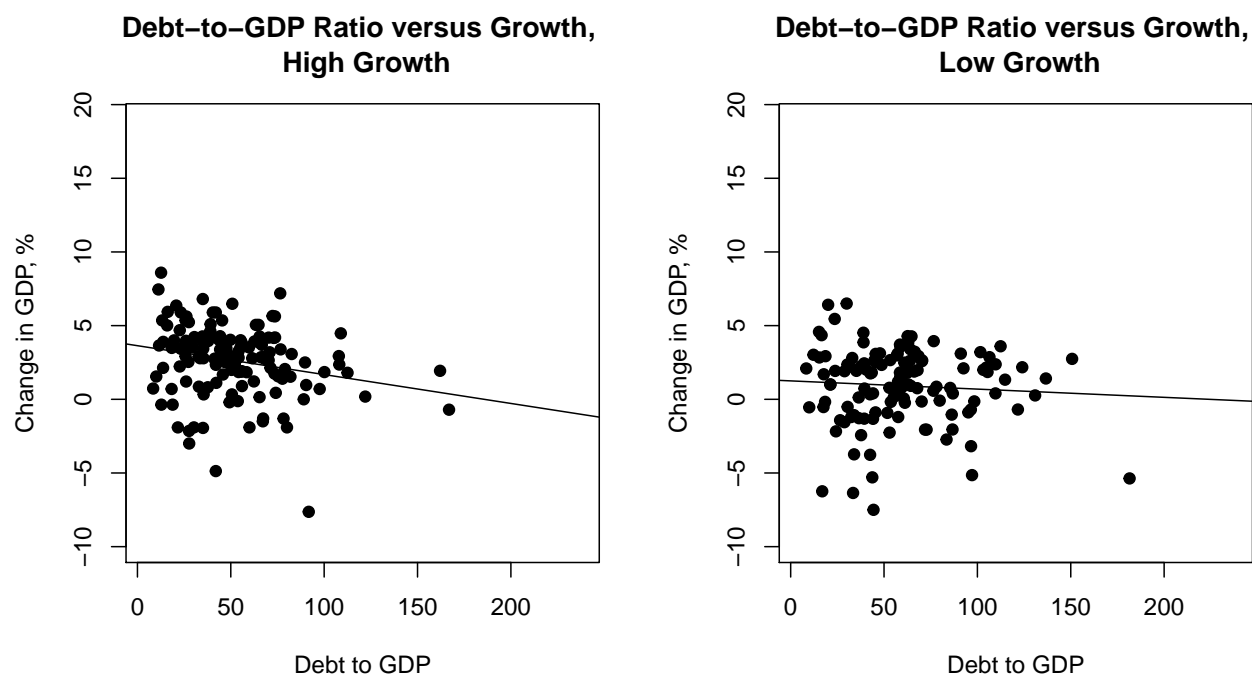
- *Current debt ratio versus current growth* for countries with increasing debt ratios and *high* previous growth
- *Current debt ratio versus current growth* for countries with increasing debt ratios and *low* previous growth

The code is below.

```
par(mfrow = c(1, 2))

growdebt1 <- (RR$debtgdp - RR$lag.debtgdp > 2) & (RR$lag.dRGDP > 2)
plot(RR$debtgdp[growdebt1], RR$dRGDP[growdebt1],
     xlab = "Debt to GDP", ylab = "Change in GDP, %",
     main = "Debt-to-GDP Ratio versus Growth, \n High Growth", pch = 19,
     ylim = range(RR$dRGDP), xlim = range(RR$debtgdp))
abline(lm(RR$dRGDP[growdebt1] ~ RR$debtgdp[growdebt1]))

growdebt2 <- (RR$debtgdp - RR$lag.debtgdp > 2) & (RR$lag.dRGDP <= 2)
plot(RR$debtgdp[growdebt2], RR$dRGDP[growdebt2],
     xlab = "Debt to GDP", ylab = "Change in GDP, %",
     main = "Debt-to-GDP Ratio versus Growth, \n Low Growth", pch = 19,
     ylim = range(RR$dRGDP), xlim = range(RR$debtgdp))
abline(lm(RR$dRGDP[growdebt2] ~ RR$debtgdp[growdebt2]))
```

Precept Questions

Question 1

Create a figure similar to that in the previous section exploring the relationship between debt and GDP. Operationalize 'increasing growth' as an increase in the GDP growth rate from the previous to current year of 2 or more percentage points and *high* and *low* previous debt as previous debt-to-GDP ratios above or below 70.

Now You are going to make a figure with one row and two columns. The two figures will plot:

- *Current* debt ratio versus *current growth* for countries with increasing growth and *high* previous debt
- *Current* debt ratio versus *current growth* for countries with increasing growth and *low* previous debt

Question 2

What is the difference in slopes between the lines in the left and right figures in the example (in the section above)? In the left and right hand figures from Question 1?

Question 3

Is the relationship between debt ratios and GDP more sensitive to changes in previous growth rates (in the section above) of previous shifts in the debt ratio (Question 1)?

Does this suggest

- debt ratios have a causal effect on growth rates
- growth rates have a causal effect on debt ratios
- neither
- both

Please keep in mind, there is no right answer to this question, just well-reasoned and less well-reasoned responses. Also, this is a tricky question to reason through. Give it a shot.

Question 4

The cutoffs we selected, 70% for debt and 2% for growth, were arbitrary. Are the results robust to different values for the debt cutoff and growth cutoff?

Make another figure with different cutoffs.

References

Herndon, T., M. Ash, and R. Pollin. 2014. "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff." *Cambridge Journal of Economics* 38 (2): 257–79. <https://doi.org/10.1093/cje/bet075>.

Irons, John, and Josh Bivens. 2010. "Government Debt and Economic Growth: Overreaching Claims of Debt "Threshold" Suffer from Theoretical and Empirical Flaws." EPI Briefing Paper 271. Economic Policy Institute.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review* 100 (2): 573–78. <https://doi.org/10.1257/aer.100.2.573>.