Quantitative Methods - Computer Lab 1

Ralf Becker

1 February 2019

For this computer lab we are using exactly the same data as for the first lecture. So please make sure that you have the datafile `RRdata.xlsx` in the folder you want to work from.

Preparing your script file

Via the RStudio menu (FILE - NEW FILE - R SCRIPT) open a new script file and safe it in the folder from which you want to work (best to create a dedicated folder for your R work).

Start by creating a comment line at the top of that file which may say something like

```
# File for first QM Computer Lab
# February 2019
```

If not mentioned otherwise all the following code should be added to that script file and executed from there.

Next you should ensure that you set the working directory to the directory where your scriptfile is in.

```
setwd("P:/Rcode/QM")   # replace the specific path with your path
```

Note that R only understands forward slashes `/`.

Load the libraries which we want to use.

```
library(tidyverse)    # for almost all data handling tasks
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readxl)       # to import Excel data
library(ggplot2)      # to produce nice graphiscs
library(stargazer)    # to produce nice results tables
```

```
##
## Please cite as:
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

If RStudio tells you that one or more of these libraries are not installed then install these (not any others) on the machine you are working from. For instance, if `stargazer` was not installed you would receive an error message like "Error in library(stargazer) : there is no package called 'stargazer' ", and in that case you should run:

```
install.packages("stargazer")
```

Once you have done that you can call `library(stargazer)` again without running into problems.

Data Upload

Make sure you have set the working directory to the directory in which you saved your script file (see above). Make sure that your data upload is successful.

```
RRData <- read_excel("RRdata.xlsx")
RRData <- as.data.frame(RRData)
str(RRData)  # prints some basic info on variables
```

```
## 'data.frame':    1171 obs. of  4 variables:
##  $ Year   : num  1946 1947 1948 1949 1950 ...
##  $ Country: chr  "Australia" "Australia" "Australia" "Australia" ...
##  $ debtgdp: num  190 177 149 126 110 ...
##  $ dRGDP  : num  -3.56 2.46 6.44 6.61 6.92 ...
```

You should now see the `RRData` object in your Environment (right hand pane).

Change the `Country` variable into a factor variable and confirm that the change of data type worked.

Check whether Poland is one of the countries in the dataset.

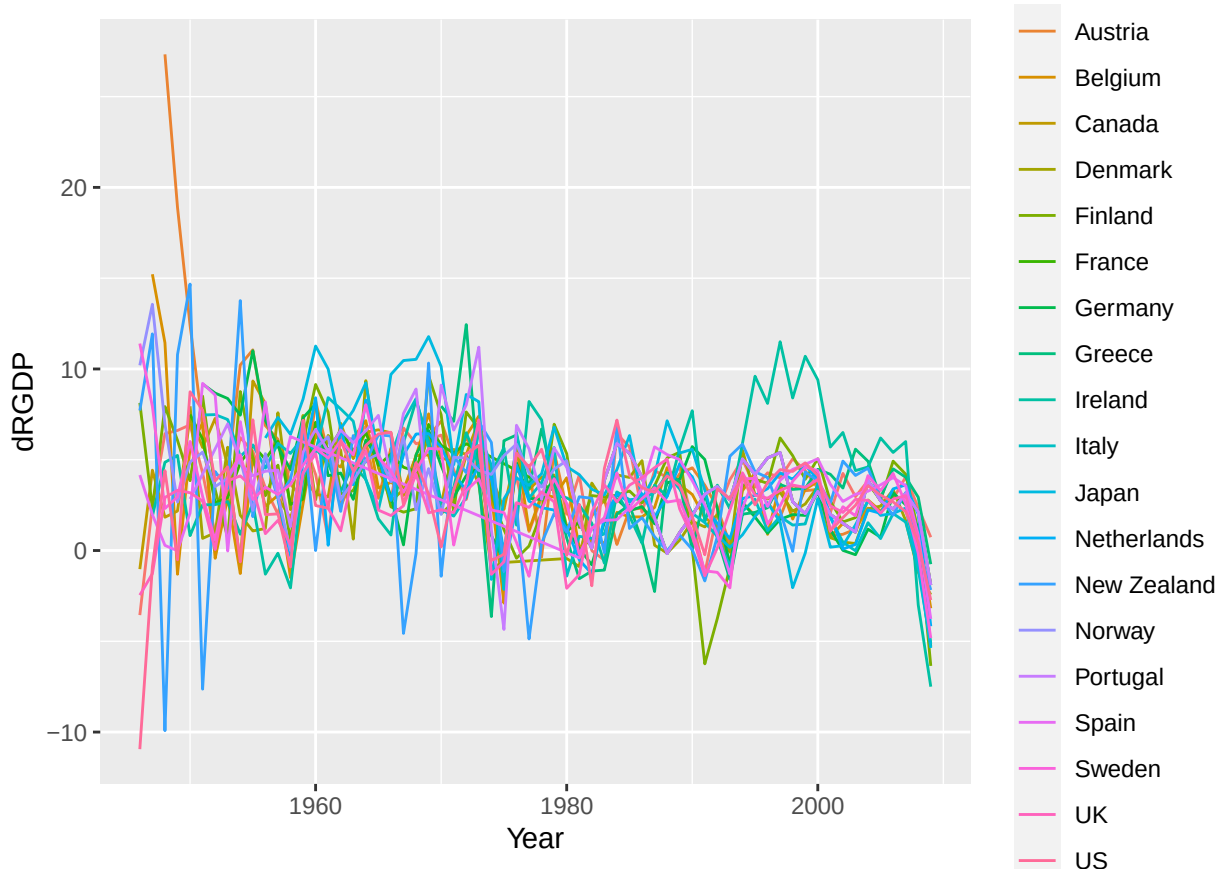
Analyse early post-war years

The analysis undertaken by Reinhard and Rogoff was for the years 1946 to 2009. Here in this lab we shall repeat some of the analysis but for a subsample of data. The argument is that some of the early postwar data were so atypical that they shouldn't be included in the analysis.

To determine which of the early years we should exclude we will first investigate how the variables behaved in these years. We will do this mainly graphically.

Complete the code below to replicate the following image, by finding the right variables to replace the `XXXX` terms.

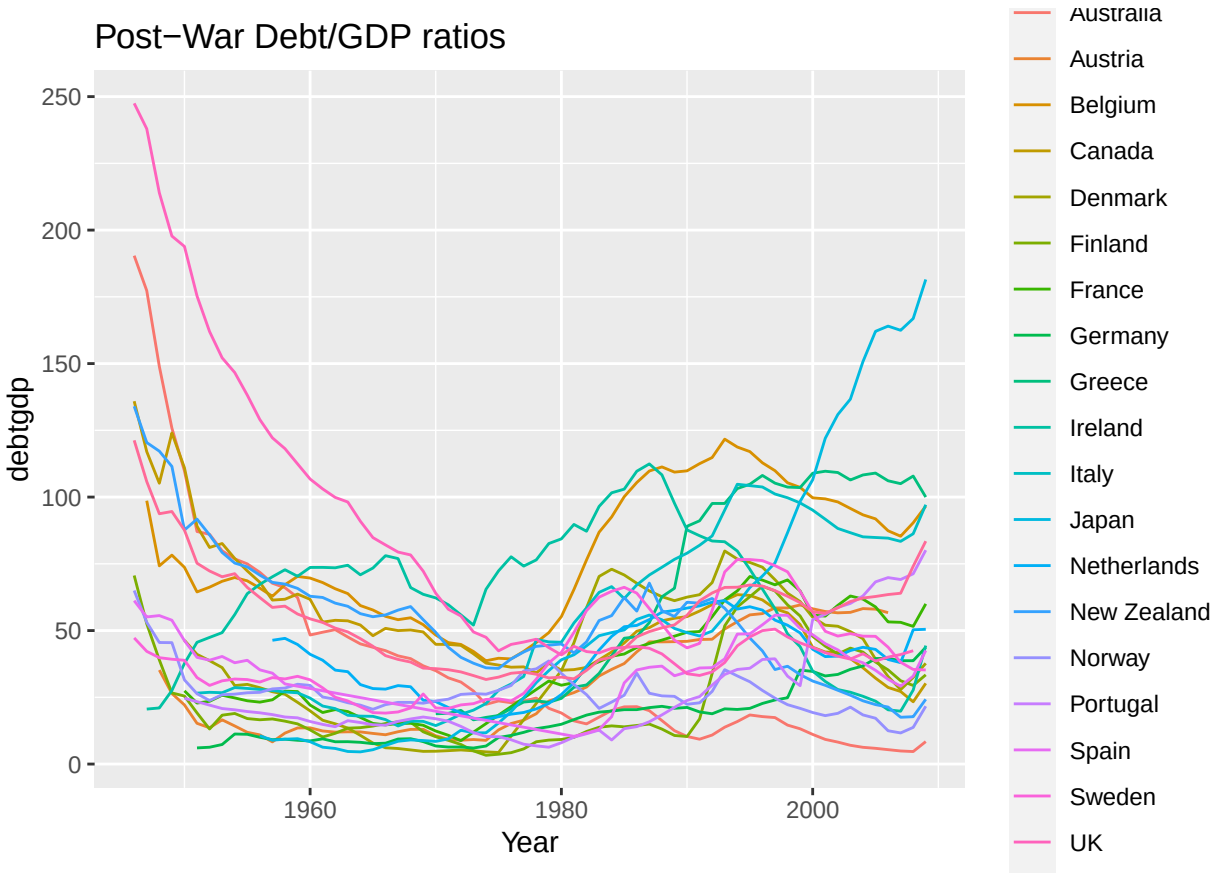
```
ggplot(RRData,aes(x = XXXX,y=XXXX,color = Country)) +
  geom_line()
```



As you can see there are some very erratic growth rates up until approximately 1952. Try what happens if you run the above figure without the `color = Country` option. Also google to figure out how you can add a Title to the figure.

Let's repeat the same for the Debt to GDP ratios.

```
ggplot(RRData,aes(x = XXXX,y=XXXX,color = Country)) +
  geom_line() +  # adds the line graph
  XXXX           # adds the title
```



We can again see that until the early 1950s there were dramatic debt reductions almost across the bord. Most countries had very significant post-war debt which they were able to rapidly reduce.

On the basis of this analysis we decide to start our analysis from 1955

Restrict the data

We want to restrict the dataset to start at 1955 and save this restricted dataset in `RR1955p`. What does `XXXX` below need to be replaced with?

```
## Remove early post-war years
RR1955p <- RRData %>% filter(XXXX)
```

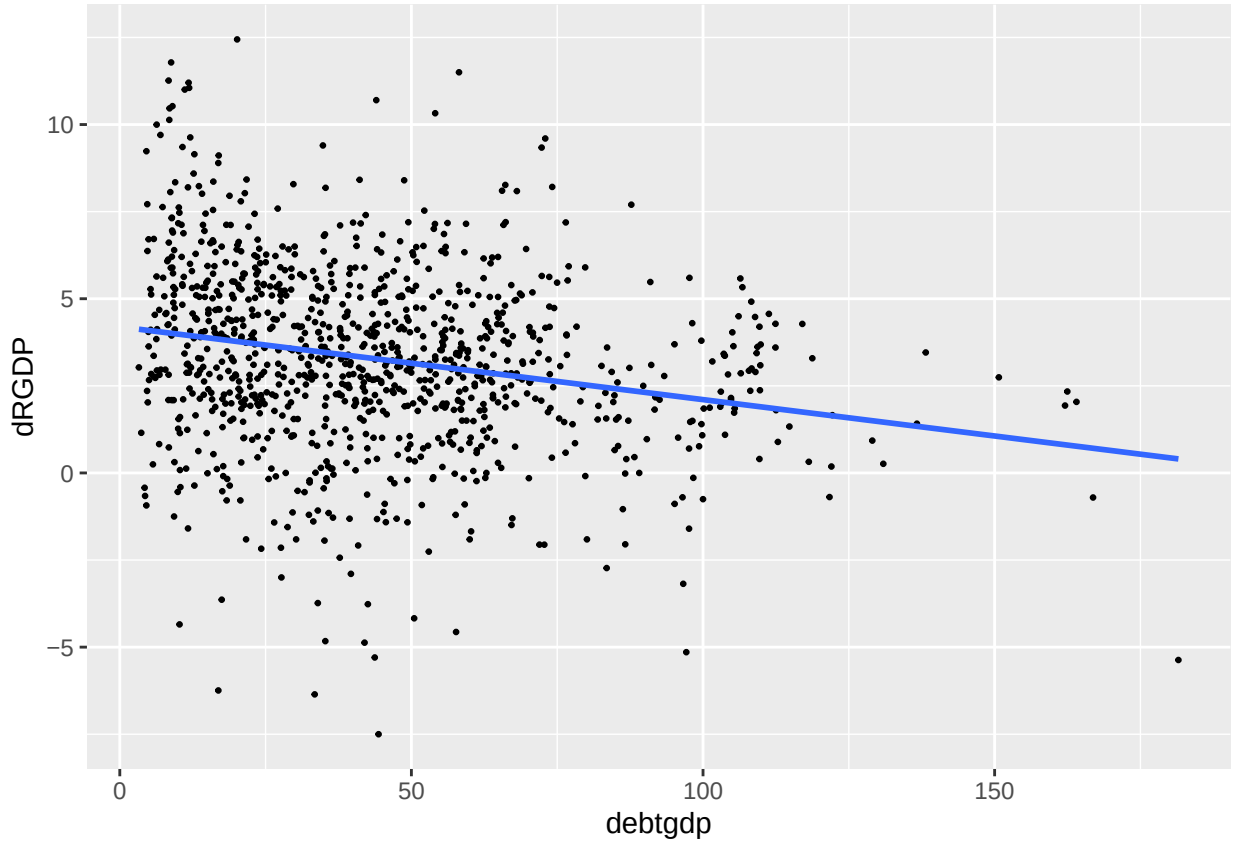
Use the`stargazer` function to confirm that the average `debtgdp` value is 43.803 and the upper quartile of `dRGDP` is 4.974.

Analysis - New dataset

Create a scatterplot (x-axis: `debtgdp`, y-axis: `dRGDP`) and a line of best fit for `RR1955p`. Replace all `XXXX` terms below.

```
p1 <- ggplot(XXXX,aes(x = XXXX,y = XXXX)) +
  geom_XXXX(size=0.5) +    # this produces the scatter plot
  XXXX(method = "lm", se = FALSE)   # adds the line
p1
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Use a simple regression model

$$dRGDP = \beta_0 + \beta_1\ debtgdp + u$$

to investigate whether the negative relationship displayed in your graph is stronger for the full dataset (`RRData`) or the restricted dataset (`RR1955p`). Complete the following code and also find the typos in the code (easiest to be guided by any error messages)

```
model_full <- lm(dGDP ~ XXXX, data = RRData)
model_1955p <- XXXX
stargazer(model_full,model_1955, type = "text")
```

YOu got it right if the Adjusted R2 for the models are 0.039 and 0.050. If at all it is slightly stronger in the `RR1955p` data.

Create categorical debt to GDP ratios - Reinhard-Rogoff analysis I

In order to replicate some of the analysis in Reinhard and Rogoff we create categorical variables (`dgcat`) for the Debt to GDP ratio. For this we use piping and the mutate command. First we do that for the full dataset, `RRData`:

```
RRData <- RRData %>% mutate(dgcat = cut(RRData$debtgdp, breaks=c(0,30,60,90,Inf)))

RRData %>% group_by(dgcat) %>%
            summarise_at("dRGDP", funs(mean, median)) %>%
            print()
```

```
## # A tibble: 4 x 3
##   dgcat     mean median
##   <fct>    <dbl>  <dbl>
## 1 (0,30]    4.17   4.15
## 2 (30,60]   3.12   3.11
## 3 (60,90]   3.22   2.9
## 4 (90,Inf]  2.17   2.34
```

Repeat this analysis for `RR1955p`.

You have worked correctly if your median growth rate in the highest debt category is 2.152.

Create categorical debt to GDP ratios - Reinhard-Rogoff analysis II

Now we implement the Reinhard-Rogoff weighting scheme to the complete and the restricted dataset. First for the complete dataset.

```
RRData_w <- RRData %>%
        group_by(dgcat,Country) %>%          # create category and country groups
        summarize( m1 = mean(dRGDP, na.rm = TRUE)) %>%  # calculate average growth in each cat
        group_by(dgcat) %>%                                  # group again by category
        summarize(  n = n(), mean = mean(m1, na.rm = TRUE), median = median(m1, na.rm = TRUE)) %>%  # c
        print()
```

```
## `summarise()` has grouped output by 'dgcat'. You can override using the `.groups` argument.
```

```
## # A tibble: 4 x 4
##   dgcat        n  mean median
##   <fct>    <int> <dbl>  <dbl>
## 1 (0,30]      17  3.98   3.87
## 2 (30,60]     20  3.04   3.15
## 3 (60,90]     18  3.42   3.05
## 4 (90,Inf]    10  1.93   2.50
```

Repeat this analysis for the restricted dataset (`RR1955p`).

```
## `summarise()` has grouped output by 'dgcat'. You can override using the `.groups` argument.
```

Is there any evidence for a large drop in growth rate for Debt/GDP ratios larger than 90%?