

# Homework: Learning the Tools

Due Earlier Than Usual

## 1 Introduction

Among the essential skills for a quant are the *technological* skills. Any quant is expected to be able to find, extract, and preprocess the data necessary for analysis with little or no help from business partners. This assignment is aimed at familiarizing you with (or refreshing you in) the means of connecting standard toolsets to data sources, of which Quandl is the primary one in this class, along with simple plotting in R and Python.

## 2 Preparation

Choose one US stock  $S_1$  whose ticker symbol starts with the same letter as your first name, and another US stock  $S_2$  whose ticker symbol starts with the same letter as your last name. Take the first Monday on or after your birthday in 2018 as a starting date, and the fourth Friday after that as an ending date. (This period will encompass 19 or 20 trading days). We'll call this your "birthday month".

## 3 Using the Tools

1. Write a Jupyter notebook with a Python kernel. Download adjusted closing prices  $S_1$  and  $S_2$  in your birthday month from Quandl. Convert prices to returns. Fit regressions of  $S_1$  on  $S_2$  on (i) daily closing prices and on (ii) returns. Create
  - (a) a plot of daily prices with  $S_2$  on the  $x$  axis and  $S_1$  on the  $y$  axis, with your price regression line superimposed
  - (b) a plot of daily returns with  $S_2$  on the  $x$  axis and  $S_1$  on the  $y$  axis, with your return regression line superimposed

Your notebook *must* automatically download the data from Quandl using the API.

2. Repeat Exercise 1 in a Jupyter notebook with an R kernel.
3. Take the econometrics data archive from the class website, consisting of 5 tab-delimited data files. The file `HEET.tab` contains *Growth* at various values of the *Timestamp*. The other files contain *DiscountRate* at various values of the *Timestamp* that do not exactly match those in `HEET.tab`. Make a Python or R Jupyter notebook (your choice) that does the following
  - (a) Read the data files from the current working directory
  - (b) Interpret values as coming from
    - i. `C00_0.tab` Belgium
    - ii. `C00_2700.tab` Canada
    - iii. `C00_2762.tab` France
    - iv. `C00_3026.tab` Sweden
  - (c) Linearly interpolate the *Growth* as a function of *Timestamp* to find values at the same timestamps from which *DiscountRate* was available.
  - (d) Make a plot (see Figure 1) with the *Growth* as a function of *DiscountRate*, with different colors for points from different countries.

Your score will depend on correctness, axis labels, plot title and legend. More points for repeating the exercise with *GDP* instead of *Timestamp* as the variable of coincidence<sup>1</sup> and making both plots appear as subplots of a single figure.

4. Create a directory `{StudentID}_{FirstName}_{LastName}_HW1` substituting your first and last names with no periods, spaces or special characters. Put your notebooks into this directory<sup>2</sup>. Zip the directory up into a file named `{FirstName}_{LastName}_HW1.zip`, and submit it to the class website.

---

<sup>1</sup>Ignore Belgium for GDP.

<sup>2</sup>You do not need to include the tab-delimited files. We already have those!

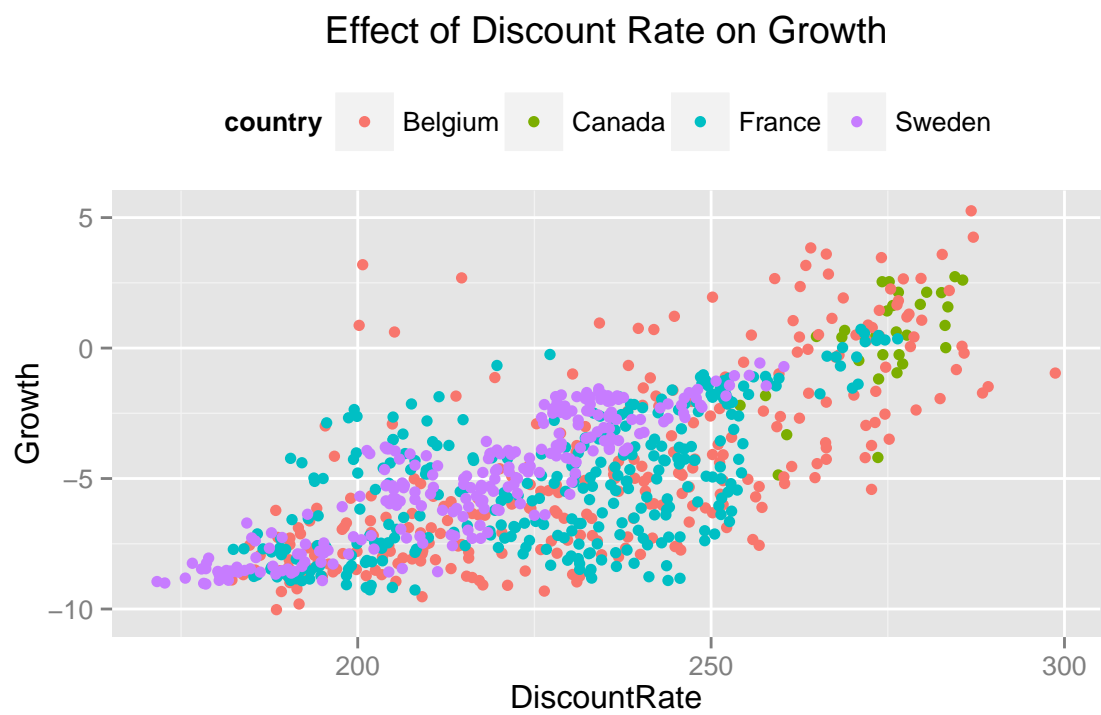


Figure 1: Sample plot.

## 4 Example Data period

The first Monday after my birthday was 18 Jun 2018 . So my data period runs from 18 Jun through 13 July, i.e. over 19 trading days (since July 4 is a US holiday).

## 5 Further Exploration

I encourage you to explore further, for example by experimenting with `xts` versus `data.frame` or `data.table` objects in R, by experimenting with `pandas.DataFrame` versus `numpy.ndarray` or `list` objects in Python, by working with `scatter3` and `ggplot` for plots, and by learning to download alternative data series such as risk-free rates, option prices, financial ratios, interest rate curves and futures information.