

I. The Worldwide Bureaucracy Indicators

The Worldwide Bureaucracy Indicators (WWBI) are a cross-national dataset on public sector employment and wages developed by the World Bank's 'Bureaucracy Lab'. The core data are included in the attached file `wwbi_data.csv`. The data are in a long format with the following variables:

<code>country_code</code>	Three letter country code
<code>indicator_code</code>	Indicator
<code>year</code>	Year
<code>value</code>	Value

For each country/year, there are typically several measurements, each given with a code in the `indicator_code`. A codebook for the series is given in the attached file `wwbi_series.csv`.

- Read the data from `wwbi_data.csv` into a tibble `wwbi`. Extract the data on "Public sector employment, as a share of formal employment" (indicator `BI.EMP.FRML.PB.ZS`) for Norway (country_code `NOR`). Which years are available and what was public sector employment as a share of formal employment in each of these years?
- Construct a boxplot indicating the values of "Public sector employment, as a share of formal employment" by year for the whole dataset. Remember to label the axes properly. Comment briefly on what we can read from the graph.
- We want to see the relationship between public sector employment and GDP levels. The attached file `gdp.csv` gives GDP per capita (in constant 2015 US\$) together with country_code and year. Combine these data with the data you used in question b) and provide a scatter plot of public sector employment versus the log of GDP per capita. Comment on what we can read in the graph.
- For most countries, the data set contain multiple values. For each country, we want to only keep the most recent observation for which we have a value. Construct a dataset with the "Public sector wage premium, by industry: Health (compared to formal wage employees)" (indicator `BI.WAG.PREM.HE`), keeping only the last observation for each country. Show the data in a scatter plot against the log GDP numbers used in question c). Each data point should be labelled with the country code. Comment on what you can read out of the graph.
- The trend in a variable can be found by regressing it on the variable year (e.g. using `lm`). Find the trend in the wage premium studied in question d) for each country separately, restricting attention to countries with at least 3 observations. Which countries have a negative trend?

II. Brian's bakery

Brian runs a bakery where he produces and sells bread using a subscription scheme. Customers subscribe to a given number of breads delivered daily over a period.

- a) The file `subscribers.xlsx` gives all the subscribers Brian has at the moment together with their required number of breads and the period where they subscribe. Compute the number of breads he has to produce on April 15th 2024.
- b) Brian wants to make a graph where he plots the number of required breads for every day from January 1st to December 31st. Help him make this graph.
- c) Customers subscribe to Brian's services by filling in an Excel sheet. All the forms can be found in the zip file `subscribers.zip`. Put the content of the zip file in a directory. Write a script that reads all the files in the directory with file structure `s#.xlsx` where # is a number (with potentially more than the three files supplied here), read the customer's name, required number of breads and subscription period, and places everything nicely in a file like the one you used in question a).

III. Floating point numbers

Explain what it means that numbers in a computer are stored as floating point numbers. What are the advantages of this way of storing numbers compared to other ways? Where can this cause problems? Illustrate with examples.