WEEK 04 -

4.1 Learning Computational Tools Excel 1

Excel, Google Sheets, libreoffice etc.
(https://edu.gcfglobal.org/en/excel/)

4.2 Absolute and Relative Referencing

Objective

- Understand absolute referencing and relative referencing
- Plotting, visualization and functional fitting
- Mathematical formula and functions
- Loading and cleaning data

Absolute and Relative Referencing

- Implement logistics population model and Excel
- Through the process, understand absolute and relative referencing

- Predict the population for 200 years starting from 1900 until 2100
- Assume maximum population that earth can hold, K = 15 billion
- Population during 1900, P_0 = 1.654 billon
- Assume population growth rate r = 1.5%

- $\text{Compute } A = \frac{K - P_0}{P_0}$
- Excel stores values in cells. Each cell has a unique address (represented with row and column indices)
- Row - numbers, columns - alphabets

$$\text{Compute } A = \frac{K - P_0}{P_0}$$

Referred with the excel formula: =(C5-C6)/C6

Model solution

- years from 1900 to 2100 are represented from 0 to 200 years for the variable t.
- This is because the initial population P0 is given for 1900.
- In the model, initial population P0 is for the year t = 0.

- let us type the expression or formula for the logistics population model

$$P(t) = K/(1 + Ae^{-rt})$$

The expression should be=$C$5/(1+$C$9∗EXP(−$C$7∗B17))

- Note we use absolute referencing for accessing values for K, A and r in cells C5,C9 and C7 with $ symbols
- We use relative referencing for accessing value for variable t in cell B17 (no $ symbol)

- Important to know how to use absolute and relative reference appropriately
- Practice with assignment questions and reading materials.

4.3 Plotting and Visualization in Excel

- Select the values containing time t, hold the control button and select the values containing population P(t)
- Insert -> Scatter Chat -> Scatter with smooth lines and markers
- Also, I have changed the growth rate from 1.5% to 1.75% (by editing the cell C7)

- We can do all kind of formatting on the chart
- Explore yourself all possibilities, later we will do the same in python with matplotlib

- Explore the influence of model parameters on the population
- Changing the parameters, automatically changed the chart

- Explore, tinker, experiment
- Data science and ML practitioners should play around the data and explore the trends

4.4 learning Computational Tools Excel 4

Objective

- Understand absolute referencing and relative referencing
- plotting, Visualization and functional fitting
- Mathematical formula and functions
- Loading and cleaning data

- visit (https://datahub.io/core/gdp)
- Download the GDP csv (comma separated values) file
- The file contains the GDP (Gross domestic product) of all the countries of the world.

1. Copy the contents of CSV file.
2. paste into Excel sheets; All data is single column A
3. Select Column A
4. Go to: Data tab
5. Click: Test to columns
6. Use: Delimited; You may preview data
7. Click: Next
8. Checkmark: Comma
9. Click: Finish
10. Play around with data, make it readable
11. Save Excel file

1. Select Row 1
2. Go to Data > Filter
3. We want only 2016 data; choose 2016 by filtering column C contents
4. Select the data for countries only, we do not want any region wise data: Copy relevant contents and paste into a new sheet

Later we will be doing all these activities in Python/Colab. Students new to programming can mentally visualize the steps needed to the task, by doing the steps in Excel in a very organized way.

Mathematical functions/formula in Excel

- Find the total sum of the GDP of all countries

  ```
  =SUM(cells)
  ```

- Find the min GDP of the countries

  ```
  =MIN(cells)
  ```

- Find the max of the countries

  ```
  =MAX(cells)
  ```

- Find the average or mean GDP of all the countries

  ```
  =AVERAGE(cells)
  ```

- Find the standard deviation GDP of all countries

  ```
  =STDEV(cells)
  ```

- Find the median GDP of the countries

  ```
  =MEDIAN(cells)
  ```

- Count the number of countries with less than GDP of 2 billion USD

  ```
  = Count(range, criteria)
  criteria: "< 2"
  ```

- Count the number of countries with less than average GDP of all countries criteria: "<"&cell_address

- Find the country with highest GDP

  ```
  =VLOOKUP(I3,E:G,2,FALSE)
  * Range E:G has three columns E,F and G
  * GDP in Column E, which is 1st column in range E:G
  * F has country name, which is 2nd column in range E:G
  * G has country code, which is 3rd column in range E:G
  * Since country is in 2nd row, we specify 2 as third argument of
  VLOOKUP
  ```

- Find the country with lowest GDP

  ```
  * VLOOPUP(I1,E:G,2,FALSE)
  ```

Look up value should be in the first column of the table that we specify

Table is comprised of E, F,G
Therefore, lookup value that we specified should be available in the first column (column E)