HKUST
VISLAB

COMP 4462
Data Visualization Tutorial

Leo Yu Ho, Lo
Qian Zhu

Monday 13 September, 2021
https://bit.ly/vis-t01

Logistics

- **We make data visible! And beautiful!**
- Course homepage: https://canvas.ust.hk/courses/38279
- About assessment:
  - In-class exercises (10%)
  - Lab sessions (10%)
  - Top-vis competition and essay (10%)
  - Final project (35%)
  - Final exam (35%)
  - Reference materials can be found on course homepage
- Tutorial session
  - **Date & Time:** Monday 9:30 am to 10:20 am
  - **Venue:** Via Zoom
  - **Tutors:** Leo and Qian

Data Visualization

- Week 1: Introduction
- Principles:
  - Week 2: Color and Perception
  - Week 3: Design Principles
  - Week 4: Tasks and Rules
  - **Week 5: Top Vis Competition**
- Specific type of data
  - Week 6: Trajectories
  - Week 7: Multi-Dimensional Data
  - Week 8: Text
  - **Week 9: Project Proposal**
  - Week 10: Graph
- Miscellaneous:
  - Week 11: Interaction and Evaluations
  - Week 12: Narrative Visualization
  - **Week 13: Final Project Presentation**

About this tutorial

- Focus on tools, more hands-on, more coding
  - Get your hands dirty, learn by doing
- Cover most of common tools in data scientist toolbox
  - Visualization oriented, obviously
- Time allocation:
  - 20 mins go through slides, 30 mins hands-on
  - Use your own computer
  - Submit your work to Canvas
- Some programming experience will help, but not necessary (we will help)
  - To help you completing the course project
  - First two weeks will be no programming (Excel and Tableau)
  - Then, more and more coding (Python and Javascript)
- One session for “where to find data” and “where to find visualizations”
  - To help you on top-vis competition and project topics

Visualization tools

GUI base vis tools

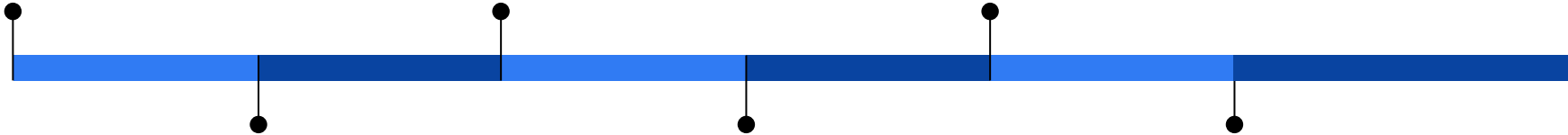
MS Excel
Tableau
MS PowerBI
Google Data Studio

Python vis tools

Matplotlib
Seaborn
Bokeh
Altair

More Expressive JS vis tools

D3.js
Three.js (WebGL based)
p5.js (HTML5 Canvas based)
Leaflet (for maps)



R language vis tools

ggplot2
ggvis

Specification base JS vis tools

Vega-lite
Plotly.js
Highcharts
ECharts

Frontend Framework, Backend Server & DB

Frontend (React, Vue.js)
Node.js (express, koa)
Python (Flask, Django)
MongoDB, PostgreSQL

And many more upon discovery!

Schedule

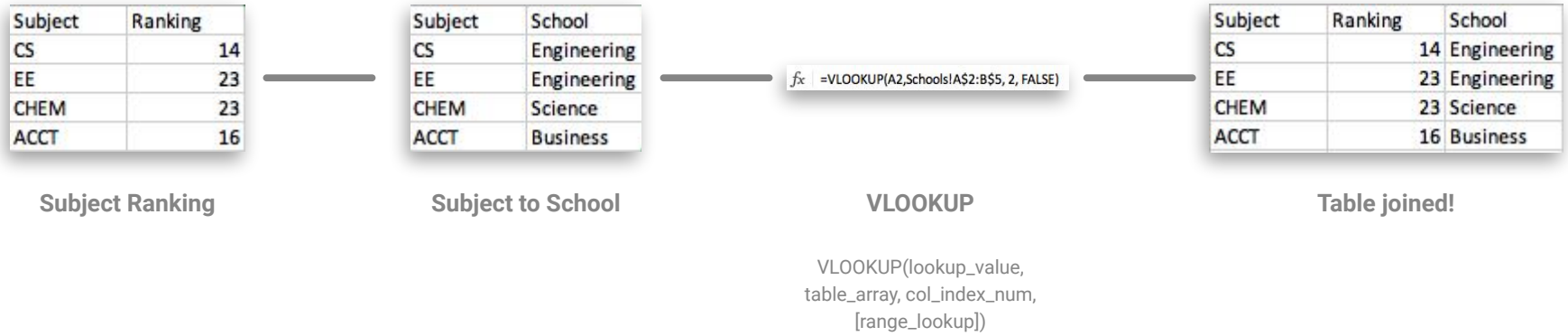
- We will go through a subset of the tools
  - Excel, Tableau, Python (Jupyter, pandas, altair), Javascript (Vega-lite, d3.js)
- Schedule
  - No coding
    - Tutorial 1: Excel
    - Tutorial 2: Tableau
  - Tutorial 3: Where to find data and visualizations
  - Python
    - Tutorial 4: Python, Jupyter and pandas basics
    - Tutorial 5: More on pandas and altair
  - Javascript
    - Tutorial 6: Javascript basics and lodash
    - Tutorial 7: Vega-lite and Observable
    - Tutorial 8: D3.js basics
    - Tutorial 9: D3.js interaction

Warm-up with Microsoft Excel

- Materials are hosted on https://github.com/leoyuholo/learning-vis-tools
  - Download the .xlsx and .csv in the directory “tutorial01”
- We will go through the followings with a simple dataset:
  - VLOOKUP function
  - Pivot table
  - Filtering
  - Plotting
  - Customizing charts, reverse axis and labels
- Then, 3 tasks on a bigger dataset
- Remember to submit your work to Canvas

VLOOKUP

- It's like table join in SQL
- We will use it to lookup the country of an university
- See documentation



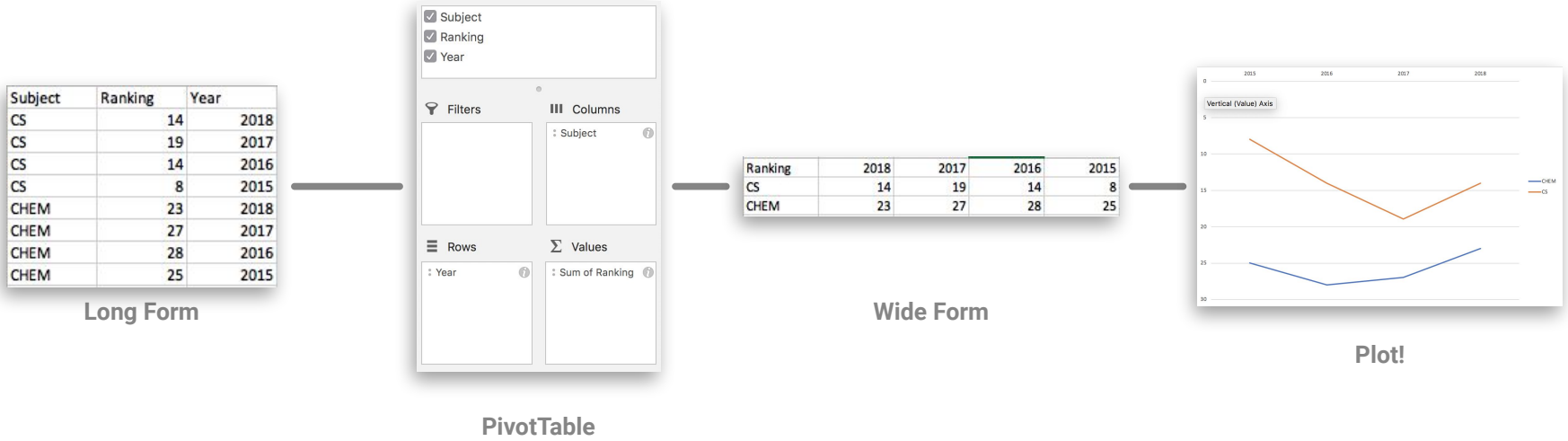
Subject	Ranking
CS	14
EE	23
CHEM	23
ACCT	16

Subject	School
CS	Engineering
EE	Engineering
CHEM	Science
ACCT	Business

Subject	Ranking	School
CS	14	Engineering
EE	23	Engineering
CHEM	23	Science
ACCT	16	Business

PivotTable

- Sometimes, data are in “Long Form”, but Excel plots charts with “Wide Form”
- We **transform** data with PivotTable
- See [documentation]



Subject	Ranking	Year
CS	14	2018
CS	19	2017
CS	14	2016
CS	8	2015
CHEM	23	2018
CHEM	27	2017
CHEM	28	2016
CHEM	25	2015

Long Form

PivotTable

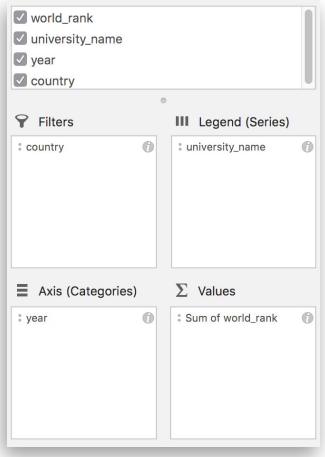
Ranking	2018	2017	2016	2015
CS	14	19	14	8
CHEM	23	27	28	25

Wide Form

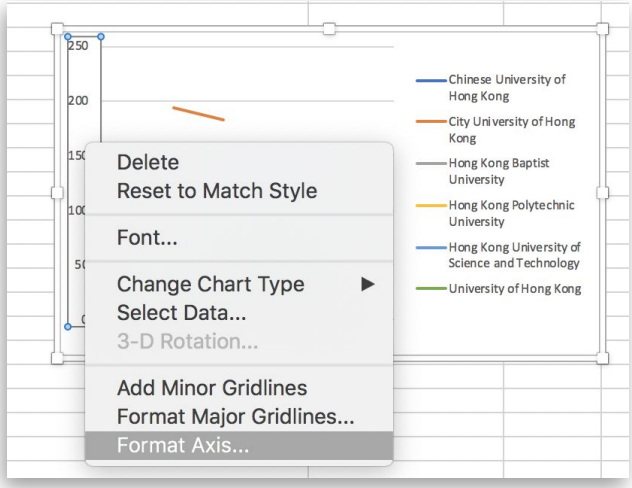
Plot!

Filtering and reverse index

- Use the “Filters” field in PivotTable



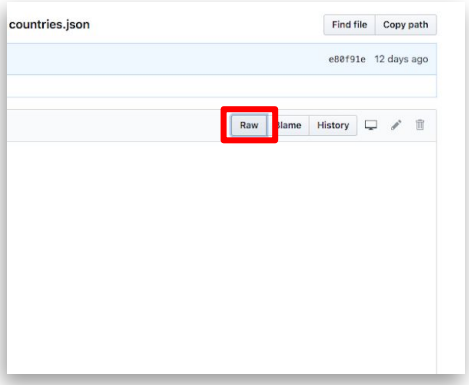
- Use format axis to reverse y-axis
  - Check the option “Values in reverse order”



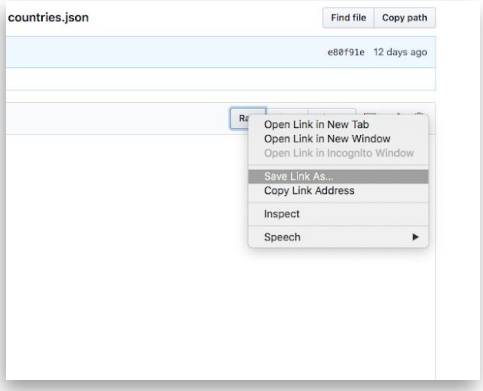
Download dataset from GitHub

1. Go to the tutorial repository
2. Go to the dataset file you want download, e.g. university_rankings.csv

3. Right click “Raw”

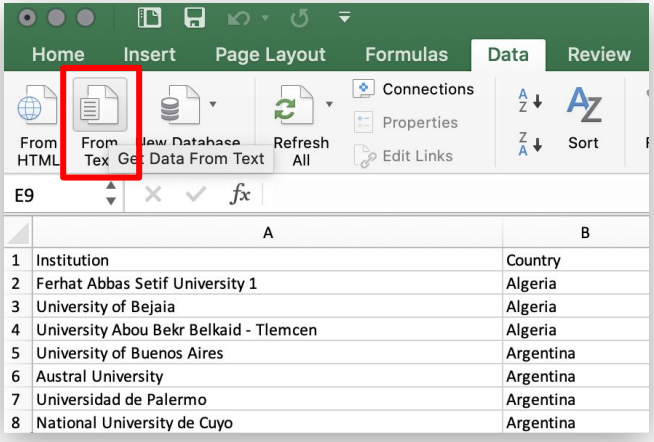


4. Save as file

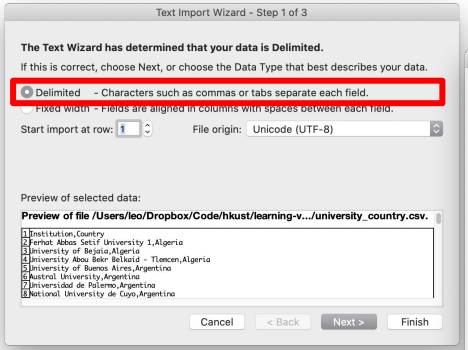


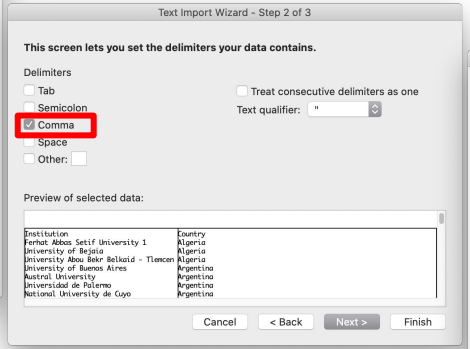
Import dataset into Microsoft Excel

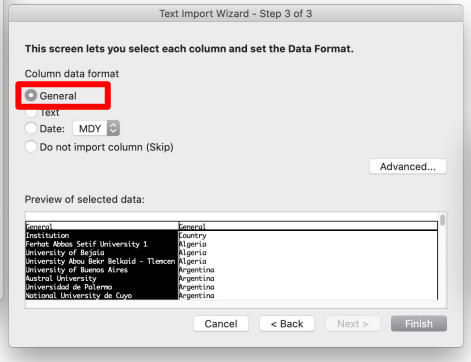
1. Data -> From Text
2. Choose file (“university_countries.csv”)
3. In import wizard:
    a. Step 1: Select “Delimited”
    b. Step 2: Select “Comma”
    c. Step 3: Select “General”
    d. Last: Put data in “New Sheet”

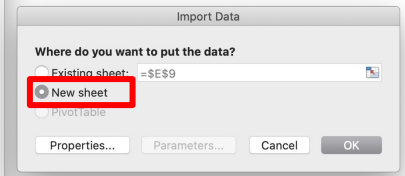


	A	B
1	Institution	Country
2	Ferhat Abbas Setif University 1	Algeria
3	University of Bejaia	Algeria
4	University Abou Bekr Belkaid - Tlemcen	Algeria
5	University of Buenos Aires	Argentina
6	Austral University	Argentina
7	Universidad de Palermo	Argentina
8	National University de Cuyo	Argentina









Lab exercise

- Tasks
  - Download the two csv files (university_rankings.csv and university_countries.csv) from GitHub
  - Import the data into Excel
  - Lookup the countries of all the universities
  - Apply PivotTable to transform “long form” to “wide form”
  - Plot the rankings of all the universities from Hong Kong
    - Utilize the filter field in PivotTable
    - Remember to flip the y-axis, zero at the top-left
    - Also add axis labels and title
  - Repeat for all the universities from Canada, Australia, UK and USA
- Remember to upload your .xlsx file to Canvas
- Credit:
  - Data source from University Rankings.ch

More topics on MS Excel Visualization

- Coursera courses
  - Problem Solving with Excel
  - Data Visualization with Advanced Excel
- Other notable features of MS Excel
  - Power Pivot, PivotCharts, Solver, Goal Seek, Data Tables, Scenario Manager, Simulation Features, ToolPak, Macros, Dashboard, Power View, Conditional Formatting, Form Control, VBA
- A detailed Excel visualization guide
- A list of data visualization with Excel websites

Next tutorial

Data processing and Tableau

- Install Tableau beforehand
  - Tableau student (Full version, preferred): https://www.tableau.com/academic/students
  - Or Tableau Public: https://public.tableau.com

Tableau

- Tableau Public
  - Free
  - All saved works are public
    - Publicly viewable, downloadable
  - Must connect to the internet in order to save
  - Less data connectors
- Tableau Desktop
  - Free for students, need verification
  - Can save locally, use without connecting to the internet
  - More data connectors
- Tableau Server
  - Standalone, dedicated server
  - Enterprise level, expensive