



HKUST
VISLAB

COMP 4462

Data Visualization Tutorial

Leo Yu Ho, Lo
Ming Yao

Tuesday 12 February, 2019

Logistics

- **We make data visible! And beautiful!**
- Course homepage: <https://canvas.ust.hk/courses/23932>
- About assessment:
 - In-class exercises and labs (10%)
 - Top-vis competition and essay (10%)
 - TBD
 - Final project (50%)
 - Phase 1: Proposal presentation (27 Mar & 29 Mar)
 - Phase 2: Project presentation (3 May & 8 May)
 - Final exam (30%)
 - Reference materials can be found on course homepage
- Tutorial session
 - **Date & Time:** Tuesday 5:00 pm to 5:50 pm
 - **Venue:** Room 4210, Lift 19
 - **Tutors:** [Leo](#) and [Ming Yao](#)

Data Visualization

- Week 1: Introduction
- Principles:
 - Week 2: Color and Perception
 - Week 3: Design Principles
 - Week 4: Tasks and Rules
- Specific type of data
 - Week 5: Trajectories
 - Week 6: Multi-Dimensional Data
 - Week 7: Text
 - Week 8: Graph
- Miscellaneous:
 - Week 9: How to know you've made the right visualization?
 - Week 10: Storytelling with visualization
 - Week 11: Scientific visualization
 - Week 12: Extra topics e.g. Explainable AI, Financial Data Analysis

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
 - Bring your own laptop or use the lab 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)

R language vis tools

[ggplot2](#)
[ggvis](#)

Specification base JS vis tools

[Vega-lite](#)
[Plotly.js](#)
[Highcharts](#)
[ECharts](#)
[BizCharts](#)
[Leaflet](#) (for maps)

Backend & DB

[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, [matplotlib](#) 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 folder “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 Ranking

Subject	School
CS	Engineering
EE	Engineering
CHEM	Science
ACCT	Business

Subject to School

fx `=VLOOKUP(A2,Schools!A$2:B$5, 2, FALSE)`

VLOOKUP

VLOOKUP(lookup_value,
table_array, col_index_num,
[range_lookup])

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

Table joined!

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

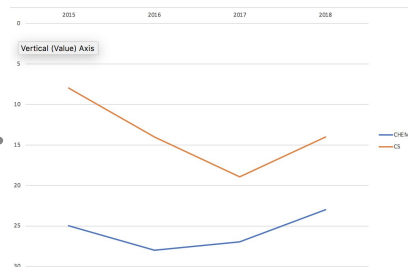
Long Form

<input checked="" type="checkbox"/> Subject
<input checked="" type="checkbox"/> Ranking
<input checked="" type="checkbox"/> Year
Filters
Columns
Subject
Rows
Year
Values
Sum of Ranking

PivotTable

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

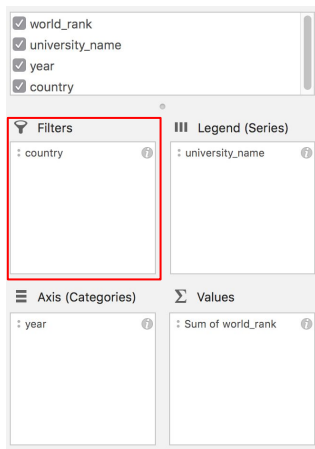
Wide Form



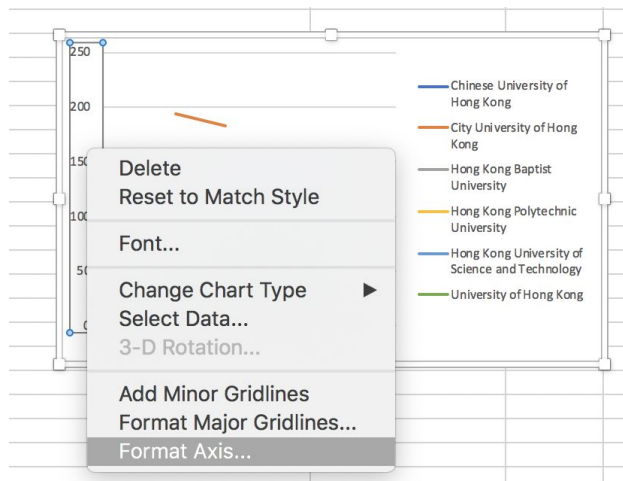
Plot!

Filtering and reverse index

- Use the “Filters” field in PivotTable



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



Lab exercise

- Tasks
 - Download the two csv files 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, USA, UK and Australia
- Remember to upload your .xlsx file to Canvas
- Credit:
 - Data source from [Kaggle Dataset by Myles O'Neill](#)

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 Public: <https://public.tableau.com>
 - Or Tableau student:
<https://www.tableau.com/academic/students>
 - Either will be fine