HKUST
VISLAB

HKUST
HCI Initiative

COMP 4462
Data Visualization Tutorial

PAN Ziqi
CHEN Chang

Friday 20 September, 2024

Logistics

- **We make data visible! And beautiful!**
- Course homepage: https://canvas.ust.hk/courses/58234
- About assessment:
  - In-class exercises (10%)
  - Lab sessions (10%)
  - Top-vis competition and essay (10%)
  - Final project (30%)
  - Two In-Class Midterms (40%)
- Tutorial session
  - **Date & Time:** Friday 03:00PM - 03:50PM
  - **Venue:** Multi-function Room, LG4, LIB
  - **Tutors:** PAN Ziqi(zpanar@connect.ust.hk) and CHEN Chang(cchenda@connect.ust.hk)

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 complete 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

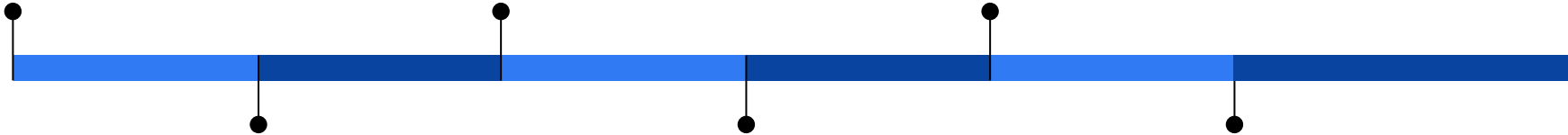
MS Excel
Tableau
MS PowerBI
Google Data Studio

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/DelPanz7/DataVisualization2024
  - Download the .xlsx and .csv in the directory “tutorial01”
- We will go through the followings with a simple dataset (lab1.xlsx):
  - VLOOKUP function
  - Pivot table
  - Filtering
  - Plotting
  - Customizing charts, reverse axis and labels
- Then, tasks on a bigger datasets (university_rankings.csv and university_countries.csv)
- Remember to submit your work to Canvas

VLOOKUP

- Use VLOOKUP when you need to find things in a table
- We will use it to lookup the country of an university and joint two tables
- See documentation

Subject	Ranking
CS	14
EE	23
CHEM	23
ACCT	16

Subject	School
CS	Engineering
EE	Engineering
CHEM	Science
ACCT	Business

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

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

Subject Ranking

Subject to School

VLOOKUP

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

Table joined!

=VLOOKUP(What you want to look up,

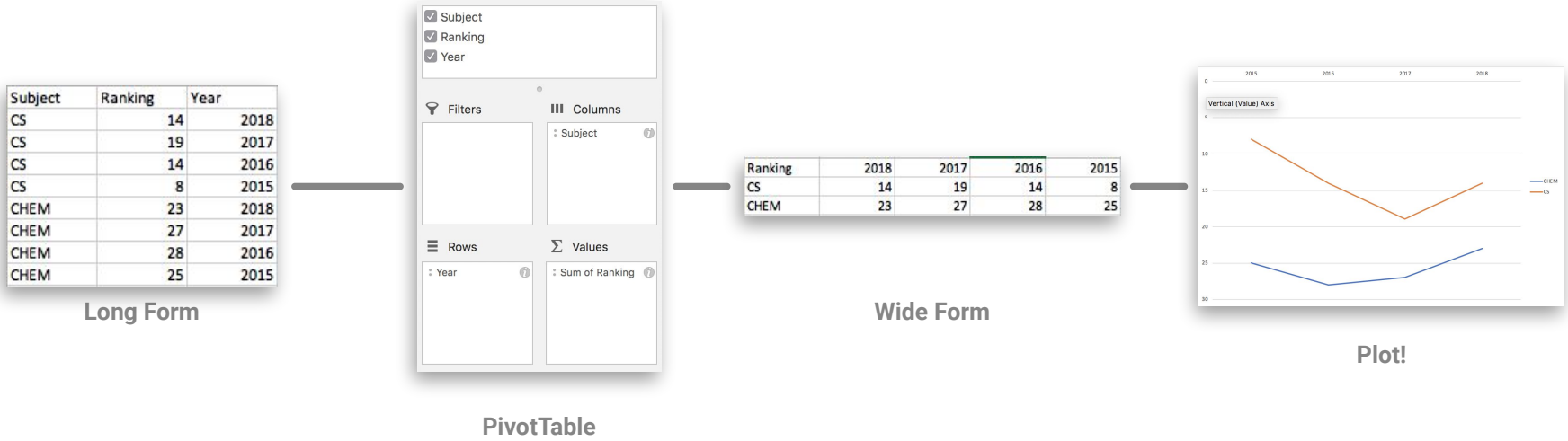
where you want to look for it,

the column number in the range containing the value to return,

return an Approximate or Exact match – indicated as 1/TRUE, or 0/FALSE)

PivotTable

- Sometimes, data are in “Long Form”, but Excel plots charts with “Wide Form”
- We **transform** data with PivotTable (A pivot table is a table of grouped values that aggregates the individual items of a more extensive table)
- 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

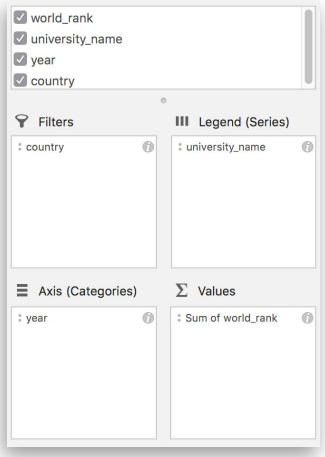
Ranking	2018	2017	2016	2015
CS	14	19	14	8
CHEM	23	27	28	25

Wide Form

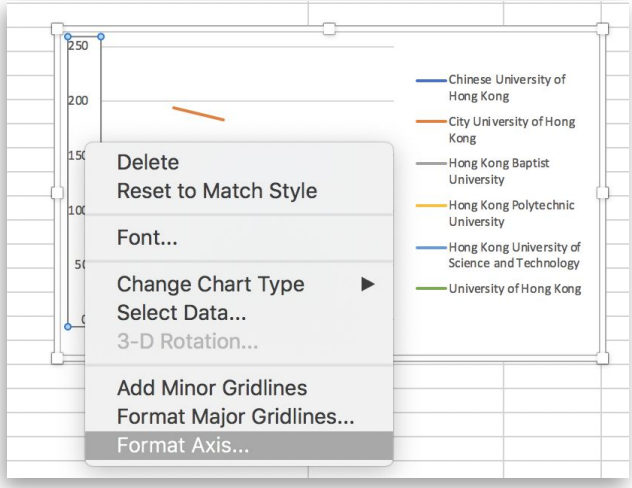
Plot!

Filtering and reverse index

- Use the “Filters” field in PivotTable



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



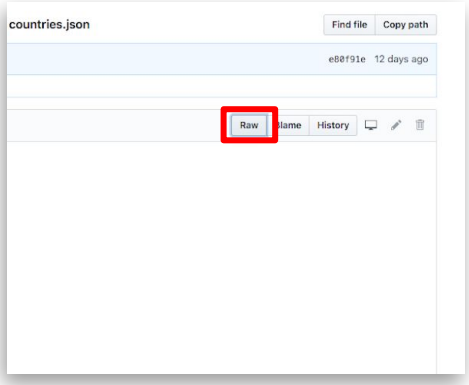
Lab exercise

- 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
  - Repeat for all the universities from Canada, China, Australia, UK and USA (Done by yourself)
    - To reduce your burden, you only need to choose any two of them
    - One sheet for one country
- Remember to upload your .xlsx file to Canvas (Due: Sep 24 11:59 PM)
  - Your .xlsx file should have the following sheets: Countries, Ranking, Pivot Table_all, Pivot Table_HK, and two other Pivot tables.
- Credit:
  - Data source from University Rankings.ch

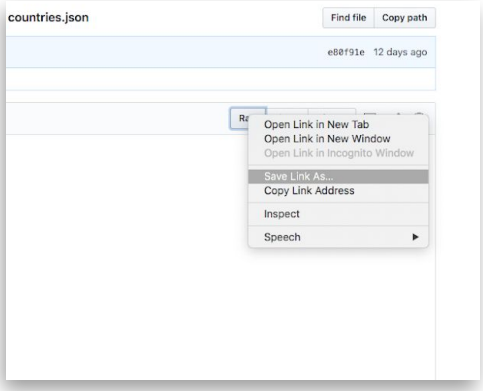
Download dataset from GitHub

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

3. Right click “Raw”

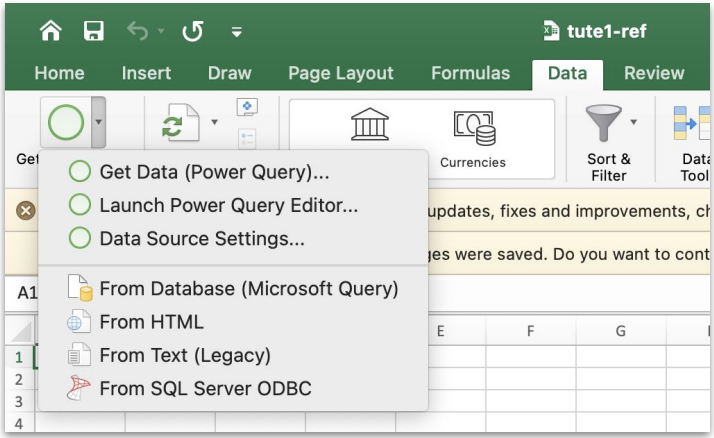


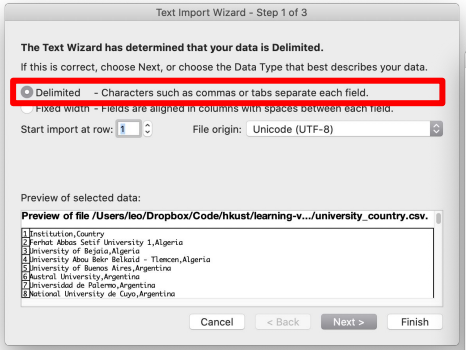
4. Save as file

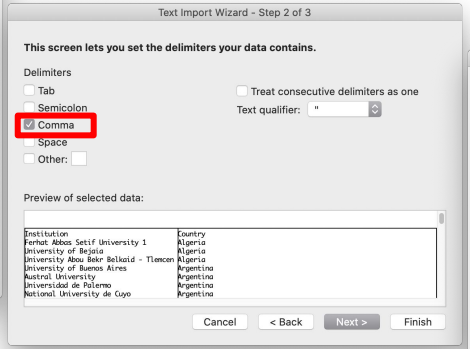


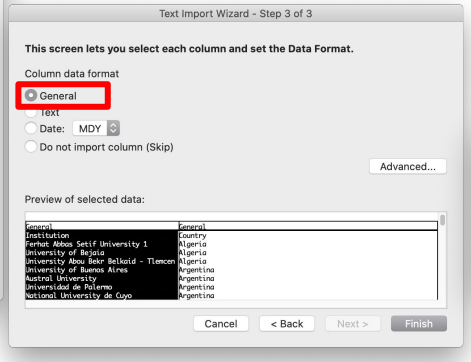
Import dataset into Microsoft Excel

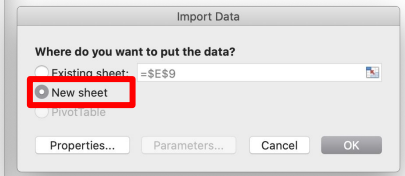
1. Data -> From Text
2. Choose file (“university_countries.csv”)
3. In import wizard:
    a. Step 1: Select “Delimited”
    b. Step 2: Select “Comma”
    c. Step 3: Select “General”
    d. Last: Put data in “New Sheet”





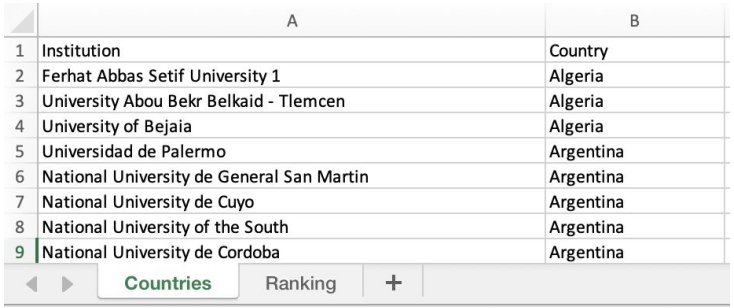






Lookup the countries of all the universities

1. After importing the data, you should have two sheets. Name them as *Countries* and *Ranking*
2. Type =VLOOKUP(B2, Countries!A$2:B$1521, 2, FALSE) in Cell D2
   a. A$2:B$1302 is the absolute positions in Excel. Remember to type $
3. Move your cursor to the bottom right of Cell D2 and you will see your cursor becomes a black cross. Then Double Click the black cross to apply the formula to the whole column.

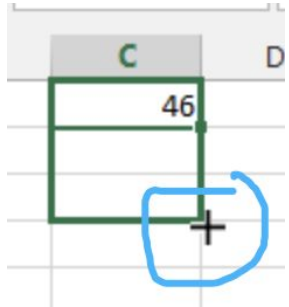


	A	B
1	Institution	Country
2	Ferhat Abbas Setif University 1	Algeria
3	University Abou Bekr Belkaid - Tlemcen	Algeria
4	University of Bejaia	Algeria
5	Universidad de Palermo	Argentina
6	National University de General San Martin	Argentina
7	National University de Cuyo	Argentina
8	National University of the South	Argentina
9	National University de Cordoba	Argentina

Countries | Ranking | +

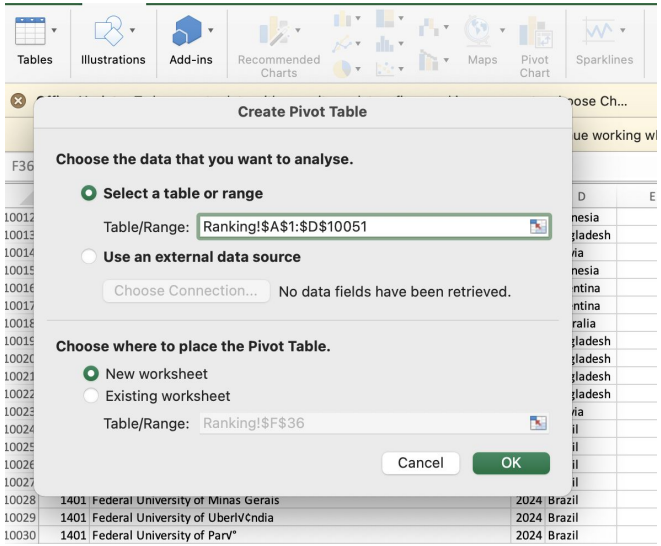
=VLOOKUP(B2, Countries!A$2:B$1521, 2, FALSE)

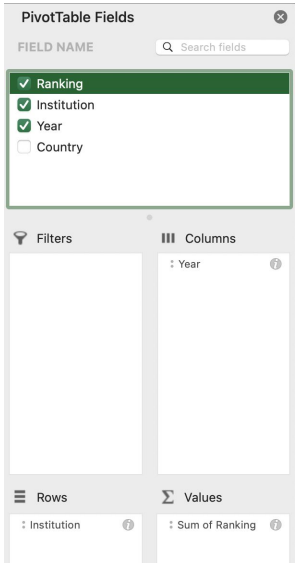
B	C	D
	Year	Country
tute of Technology, MIT	2015	USA
ıdon	2015	UK
idge	2015	UK
	2015	USA
ondon	2015	UK

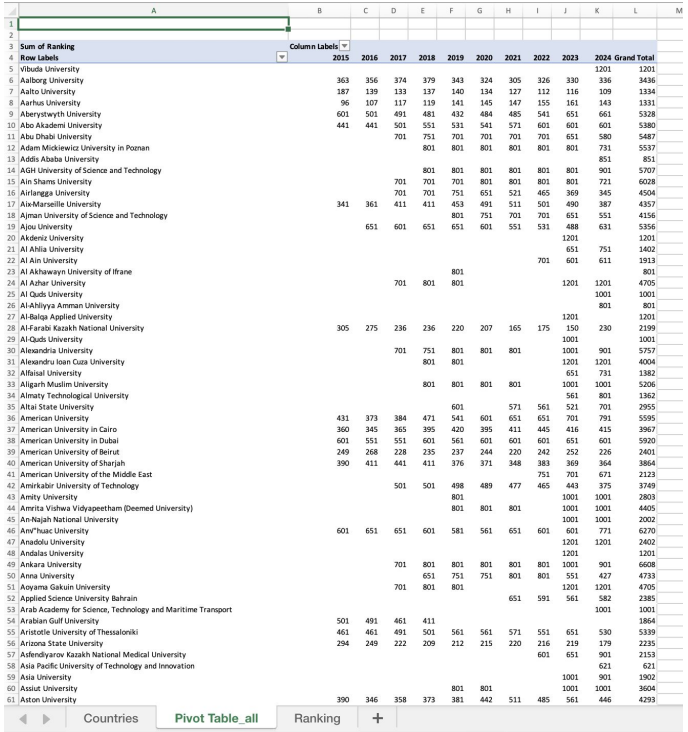


Apply PivotTable to transform “long form” to “wide form”

1. In Ranking sheet (with countries combined), insert->Pivot Table, choose *New worksheet*, then click OK.
2. Set PivotTable Fields
3. Name the sheet as *Pivot Table_all*

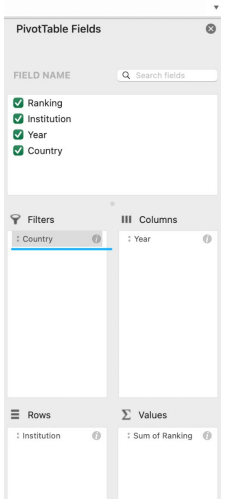


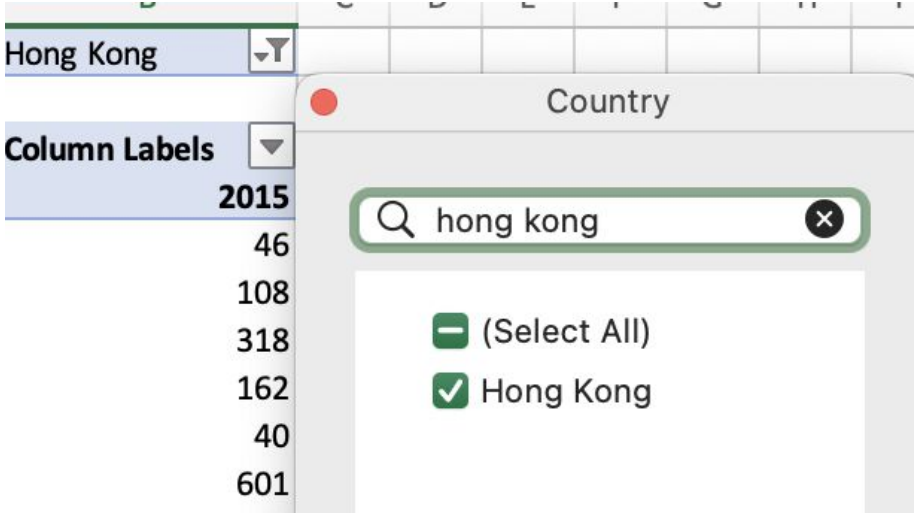




Plot the rankings of all the universities from Hong Kong

- Make a copy of *Pivot Table_all* and rename it as *Pivot Table_HK*
- Utilize the filter field in PivotTable
- Only select Hong Kong in the filter

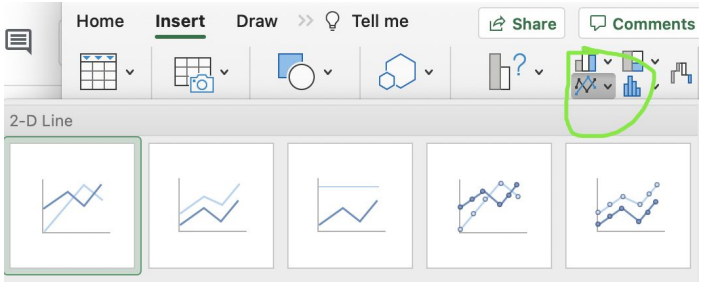


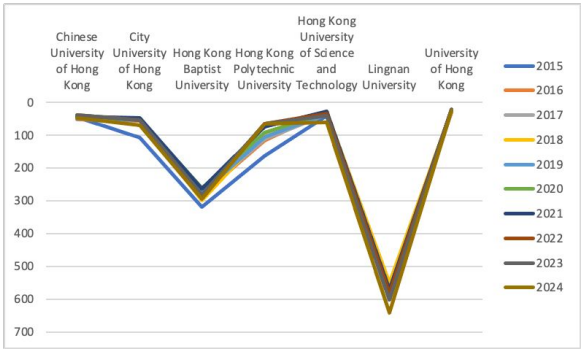


Plot the rankings of all the universities from Hong Kong

- Get line chart in sheet *Pivot Table_HK*
  - Select the data
  - Insert->2D-Line

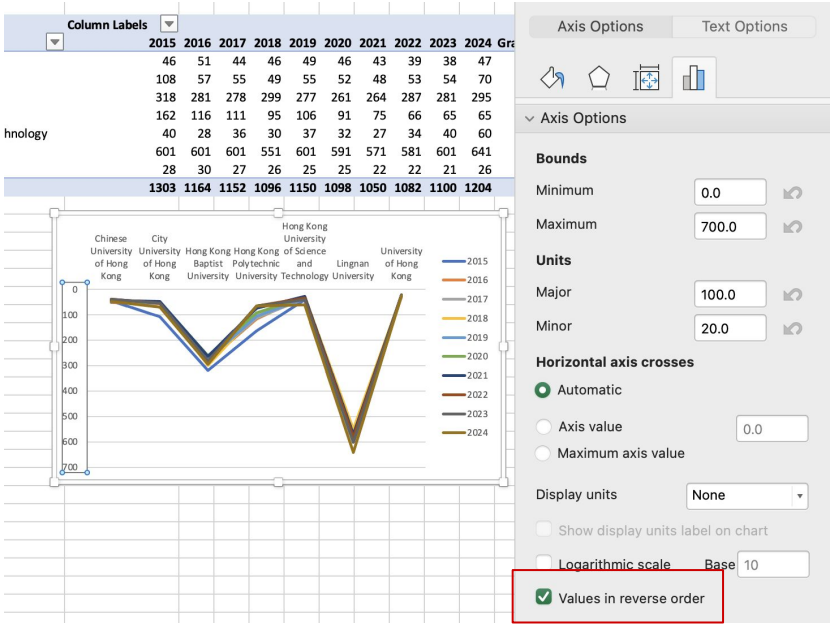
Country	Hong Kong										
Sum of Ranking	**Column Labels**										
Row Labels		**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**Grand Total**
Chinese University of Hong Kong		46	51	44	46	49	46	43	39	38	47	449
City University of Hong Kong		108	57	55	49	55	52	48	53	54	70	601
Hong Kong Baptist University		318	281	278	299	277	261	264	287	281	295	2841
Hong Kong Polytechnic University		162	116	111	95	106	91	75	66	65	65	952
Hong Kong University of Science and Technology		40	28	36	30	37	32	27	34	40	60	364
Lingnan University		601	601	601	551	601	591	571	581	601	641	5940
University of Hong Kong		28	30	27	26	25	25	22	22	21	26	252
Grand Total		**1303**	**1164**	**1152**	**1096**	**1150**	**1098**	**1050**	**1082**	**1100**	**1204**	**11399**





Plot the rankings of all the universities from Hong Kong

- flip the y-axis, zero at the top-left



Checklist for your submission

- Countries sheet

Institution	Country
National University de La Plata	Argentina
Pontifical Catholic University Argentina	Argentina
Universidad de Palermo	Argentina
University of San Andres	Argentina
National Technological University	Argentina
National University de TucumV°n	Argentina
National University de Rosario	Argentina
Buenos Aires Institute of Technology	Argentina
National University de Mar del Plata	Argentina
National University of the South	Argentina
National University de RV≠o Cuarto	Argentina
National University de Cuyo	Argentina
University Torcuato di Tella	Argentina
University of Salvador	Argentina
National University de Quilmes	Argentina
National University de General San MartV≠n	Argentina
Catholic University of CV≥rdoba	Argentina
National University of the Littoral	Argentina
University of Belgrano	Argentina
University of Buenos Aires	Argentina
National University de San Luis	Argentina
National University de CV≥rdoba	Argentina
National University of Central Buenos Aires	Argentina
National University del Comahue	Argentina
Austral University	Argentina
Yerevan State University	Armenia
Russian-Armenian (Slavonic) State University	Armenia
Flinders University	Australia

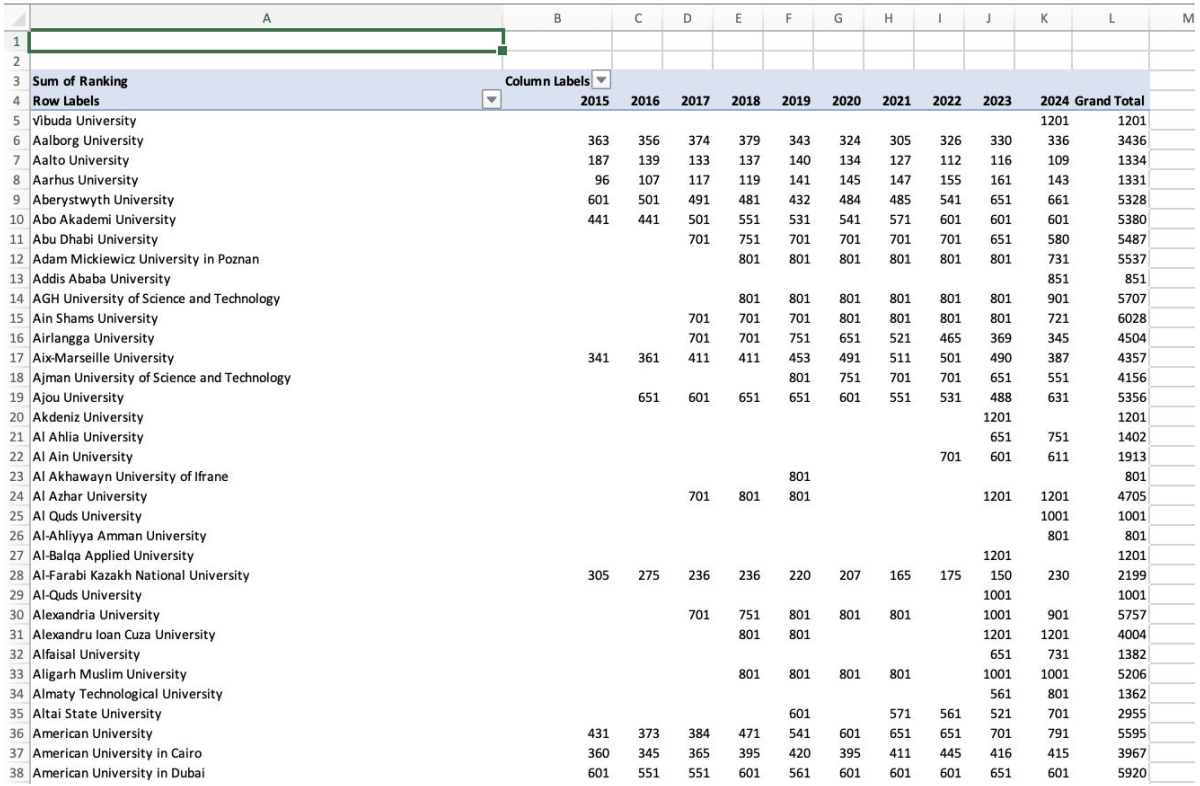
Checklist for your submission

- Ranking sheet with **Country Column**

A	B	C	D
Ranking	Institution	Year	Country
1	Massachusetts Institute of Technology, MIT	2015	USA
2	Imperial College London	2015	UK
2	University of Cambridge	2015	UK
4	Harvard University	2015	USA
5	University College London	2015	UK
5	University of Oxford	2015	UK
7	Stanford University	2015	USA
8	California Institute of Technology, Caltech	2015	USA
9	Princeton University	2015	USA
10	Yale University	2015	USA
11	University of Chicago	2015	USA
12	Swiss Federal Institute of Technology Zurich, ETHZ	2015	Switzerland
13	University of Pennsylvania	2015	USA
14	Johns Hopkins University	2015	USA
14	Columbia University	2015	USA
16	King,Âôs College London	2015	UK
17	Swiss Federal Institute of Technology Lausanne, EPFL	2015	Switzerland
17	University of Edinburgh	2015	UK
19	Cornell University	2015	USA
20	University of Toronto	2015	Canada
21	McGill University	2015	Canada
22	National University of Singapore	2015	Singapore
23	University of Michigan - Ann Arbor	2015	USA
24	Ecole Normale SupV©rieure de Paris	2015	France
25	Australian National University	2015	Australia
25	Duke University	2015	USA
27	University of California, Berkeley	2015	USA
28	University of Hong Kong	2015	Hong Kong
29	University of Bristol	2015	UK
30	University of Manchester	2015	UK
31	The University of Tokyo	2015	Japan
31	Seoul National University	2015	South Korea
33	University of Melbourne	2015	Australia
34	Northwestern University	2015	USA
35	Ecole Polytechnique	2015	France
36	Kyoto University	2015	Japan
37	University of Sydney	2015	Australia
37	University of California, Los Angeles	2015	USA
39	Nanyang Technological University	2015	Singapore
40	Hong Kong University of Science and Technology	2015	Hong Kong

Checklist for your submission

- Pivot Table_all

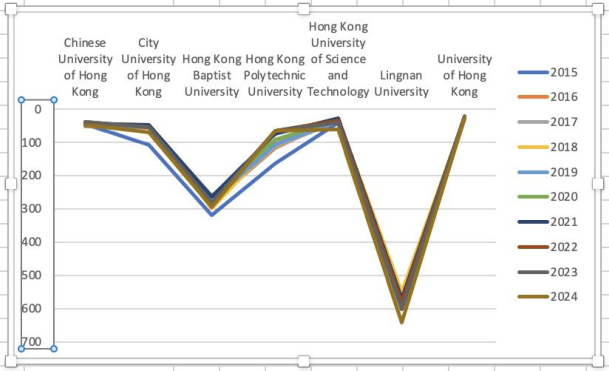


Sum of Ranking	Column Labels										
Row Labels	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Grand Total
Vĩbuda University										1201	1201
Aalborg University	363	356	374	379	343	324	305	326	330	336	3436
Aalto University	187	139	133	137	140	134	127	112	116	109	1334
Aarhus University	96	107	117	119	141	145	147	155	161	143	1331
Aberystwyth University	601	501	491	481	432	484	485	541	651	661	5328
Abo Akademi University	441	441	501	551	531	541	571	601	601	601	5380
Abu Dhabi University			701	751	701	701	701	701	651	580	5487
Adam Mickiewicz University in Poznan				801	801	801	801	801	801	731	5537
Addis Ababa University										851	851
AGH University of Science and Technology				801	801	801	801	801	801	901	5707
Ain Shams University			701	701	701	801	801	801	801	721	6028
Airlangga University			701	701	751	651	521	465	369	345	4504
Aix-Marseille University	341	361	411	411	453	491	511	501	490	387	4357
Ajman University of Science and Technology					801	751	701	701	651	551	4156
Ajou University		651	601	651	651	601	551	531	488	631	5356
Akdeniz University									1201		1201
Al Ahlia University									651	751	1402
Al Ain University								701	601	611	1913
Al Akhawayn University of Ifrane					801						801
Al Azhar University			701	801	801				1201	1201	4705
Al Quds University										1001	1001
Al-Ahliyya Amman University										801	801
Al-Balqa Applied University									1201		1201
Al-Farabi Kazakh National University	305	275	236	236	220	207	165	175	150	230	2199
Al-Quds University									1001		1001
Alexandria University			701	751	801	801	801		1001	901	5757
Alexandru Ioan Cuza University				801	801				1201	1201	4004
Alfaisal University									651	731	1382
Aligarh Muslim University				801	801	801	801		1001	1001	5206
Almaty Technological University									561	801	1362
Altai State University					601		571	561	521	701	2955
American University	431	373	384	471	541	601	651	651	701	791	5595
American University in Cairo	360	345	365	395	420	395	411	445	416	415	3967
American University in Dubai	601	551	551	601	561	601	601	601	651	601	5920

Checklist for your submission

- Pivot Table_HK and two other Pivot table with any two countries
- Line chart with y-axis reversed (Your line chart can have different styles)

Country	Hong Kong										
Sum of Ranking	**Column Labels**										
Row Labels	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**	**Grand Total**
Chinese University of Hong Kong	46	51	44	46	49	46	43	39	38	47	449
City University of Hong Kong	108	57	55	49	55	52	48	53	54	70	601
Hong Kong Baptist University	318	281	278	299	277	261	264	287	281	295	2841
Hong Kong Polytechnic University	162	116	111	95	106	91	75	66	65	65	952
Hong Kong University of Science and Technology	40	28	36	30	37	32	27	34	40	60	364
Lingnan University	601	601	601	551	601	591	571	581	601	641	5940
University of Hong Kong	28	30	27	26	25	25	22	22	21	26	252
Grand Total	**1303**	**1164**	**1152**	**1096**	**1150**	**1098**	**1050**	**1082**	**1100**	**1204**	**11399**



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