Lab 2: Tableau Basic & Critique

> **Submission by 11:59 AM —3/18 (Saturday)**
> **Point**: 4
>
> 中英文作答皆可
>
> N.B. You will still get some scores if you fail to come up with a perfect answer but explain your thought process/concerns.

Section 0 -

Your name: 廖泓傑

Your student ID number: b07703001

**Section 1 – Getting Started with Tableau (2 points)**

First things first:

1) Download and install Tableau Desktop
2) Download the datasets @ COOL

For this section, spend some time explore the features in Tableau. You could access a cheat sheet @ COOL. These steps may help you to get familiar Tableau:

Task 1: A Bivariate Chart (0.5 points)

Research goal: Compare the number of Master's/ Bachelor's degrees awarded by US regions

1. Download the dataset "American Universities" @COOL
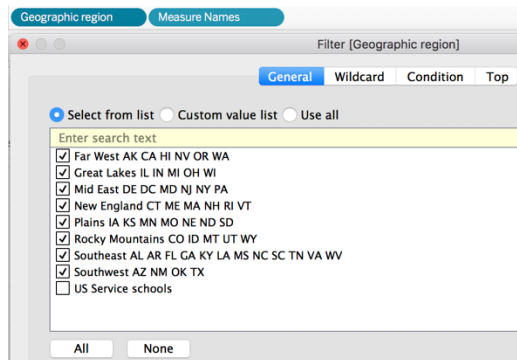2. Import the data. Select Data.
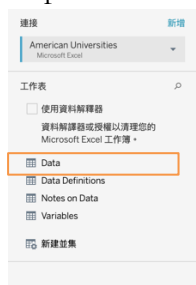
3. Feel free to enable the "Show Caption" and let Tableau remind us what we are visualizing right now. Path: Worksheet > Show Caption
4. Drag "Geographic region" from the Dimensions pane into Columns
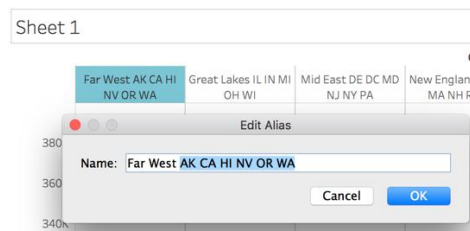5. Drag "Master's degrees awarded" from the Measures pane into Rows

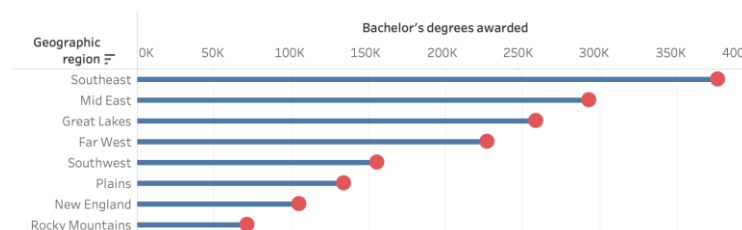6. Drag "Bachelor's degrees awarded" from the Measures pane into Rows. Now you will have two measures in Rows.

7. Now, select "side-by-side bar" on your Show Me pane.

8. Click the side arrow of "Geographic region" on the Columns, select Filters, uncheck "US Service School" to exclude unwanted data.

Geographic region	Measure Names

Filter [Geographic region]

General Wildcard Condition Top

◉ Select from list ○ Custom value list ○ Use all

Enter search text
- ☑ Far West AK CA HI NV OR WA
- ☑ Great Lakes IL IN MI OH WI
- ☑ Mid East DE DC MD NJ NY PA
- ☑ New England CT ME MA NH RI VT
- ☑ Plains IA KS MN MO NE ND SD
- ☑ Rocky Mountains CO ID MT UT WY
- ☑ Southeast AL AR FL GA KY LA MS NC SC TN VA WV
- ☑ Southwest AZ NM OK TX
- ☐ US Service schools

All   None

9. Let's finalize the graph. Right-click the caption, select "Edit Alias." Delete all the abbreviations of states for each row.

Sheet 1

Far West AK CA HI NV OR WA	Great Lakes IL IN MI OH WI	Mid East DE DC MD NJ NY PA	New England MA NH RI

Edit Alias

380

Name: Far West AK CA HI NV OR WA

360   Cancel   OK

340K

10. On your menu bar, click "File" > "Save"

11. [bonus, up to +0.25] Challenge yourself to make a lollipop chart. N.B. The part of the candy and its stick should be in two different colors. Please select variables other than "Geographic Region" and "Bachelor's degrees awarded," so we know you can create the lollipop chart on your own :)

Bachelor's degrees awarded

Geographic region 0K 50K 100K 150K 200K 250K 300K 350K 400K

- Southeast
- Mid East
- Great Lakes
- Far West
- Southwest
- Plains
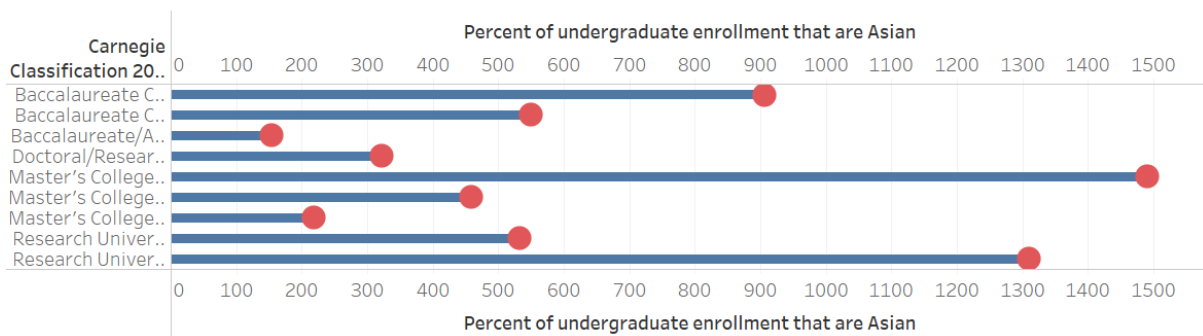- New England
- Rocky Mountains

For completing the Task 1, you are free to delete all the instructions above and simply **paste a screenshot** of your final output of Task 1 here. (and your screenshot for the bonus, if applied)

Task 1 output:
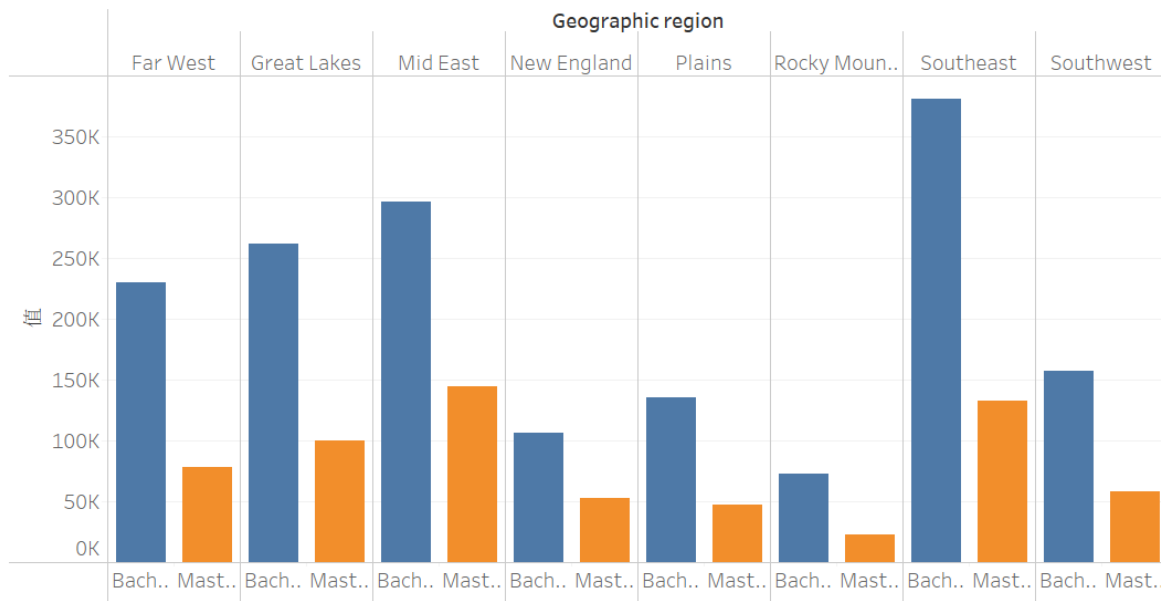
Task 1 bonus:

Lollipop Chart


Task 2: A Multivariate Chart (0.5 points)

Continuing with the same dataset, our practice goal is to find out US universities with the larger enrollment (>20k) and know how we can display data points on the map.

1. Start a new worksheet.

2. Drag "zip code" from Dimensions to the Visualization field. Please note that Tableau typically uses longitude and latitude coordinates to generate geographical layouts. However, for this particular case, we will be using zip codes to link the geographical information.

3. If you do not see any data after completing Step 2, please locate the "UNKNOWN" option at the bottom right corner of your screen, click on it, and select "Edit Location" followed by "United States". This step ensures that Tableau recognizes that the zip codes are in the US.

4. Drag "Name" from Dimensions "onto" the Label.
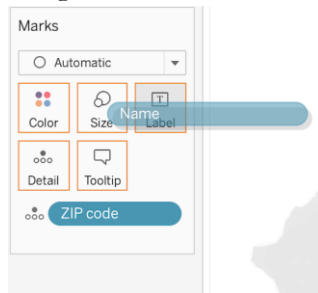
5. Drag "Full-time enrollment" from Measures to into the Size. We should see the size of the bubbles vary.
6. Click Color and set the Opacity (i.e., Transparency controller) to 60%
7. Let's target at universities with larger enrollment. Click the side of "Sum of Full-time enrollment," then click filter, select 20000- 50012.
8. Drag "Sector of institution" from Dimensions to the Visualization Field.
9. Feel free to explore more functions, try to change colors, size of circles, or shapes.
10. Paste a screenshot of your final output of Task 2.
11. Save your project.

For completing the Section, you are free to delete all the instructions above and simply **paste a screenshot** of your final output of Task 2 here.

Task 2
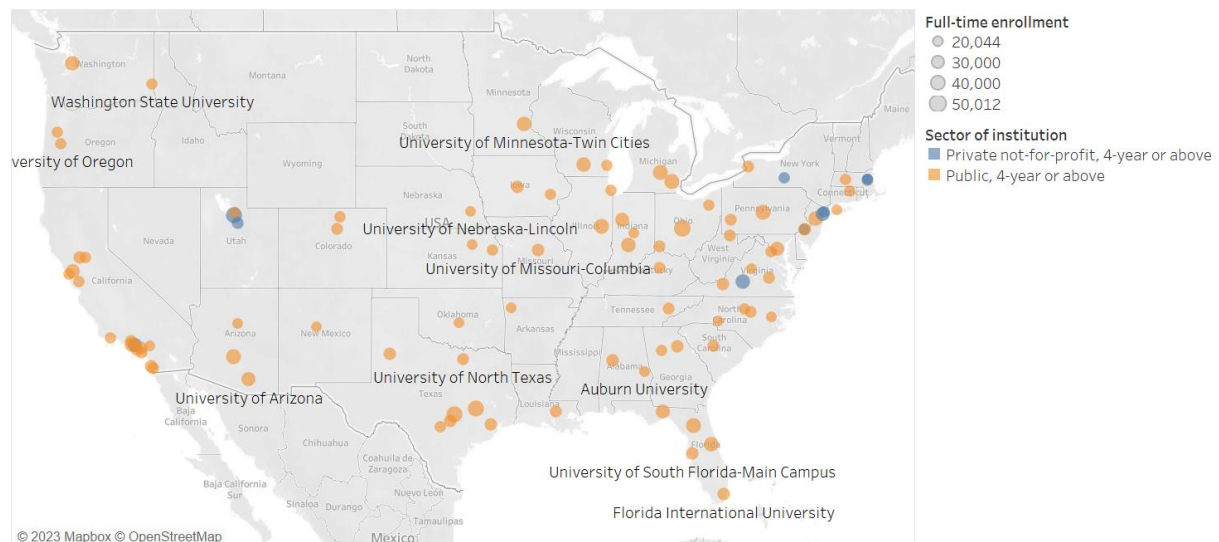
Data source: National Center for Education Statistics
https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx

Task 3 — Top Baby Names in the US (1 points)

Download the dataset "Top Baby Names" @COOL

Find a research goal based on the dataset, visualize it with Tableau Desktop, and complete the following questions:

1. Think of your own visualization goal: What your research goal or research question? (e.g., what are the name adoption trends of NameA, NameB, and NameC …, over time? Wondering which name(s)'s adoption dropped most dramatically over time?)
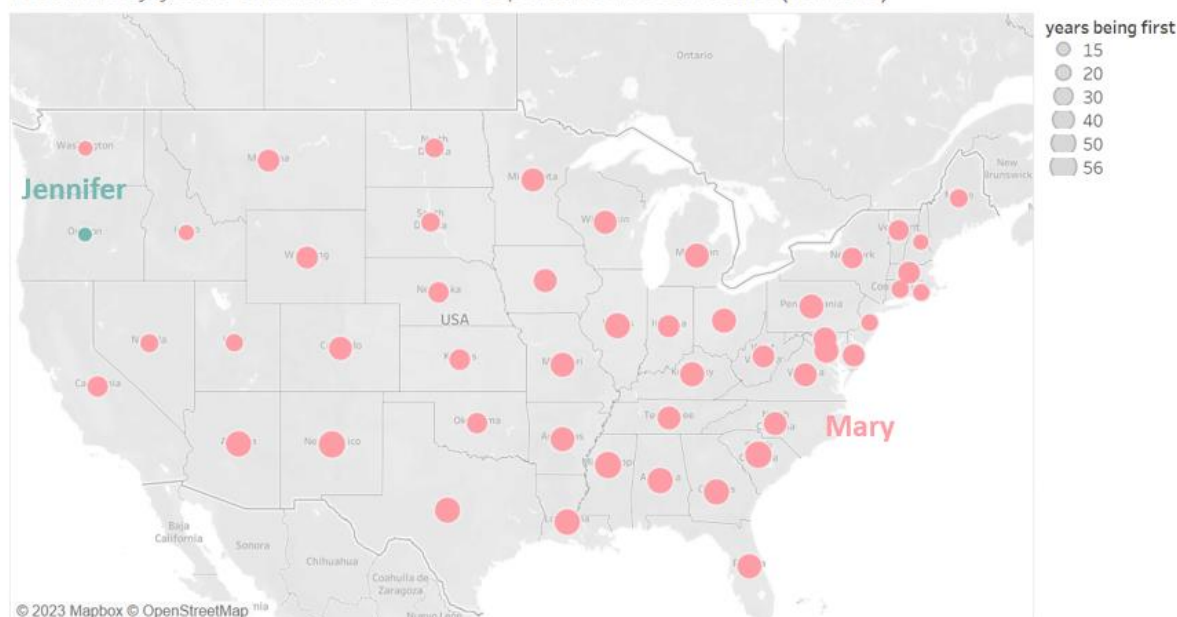   Research question: 美國在 1910~2012 年間最常取的嬰兒名稱是否有地理上的關聯性?

2. Paste a screenshot of your final product here. Briefly summarize your finding. Do you apply any special design or tweaking? (e.g., markers, color, …?

How many years the name was the top name in each state(male)
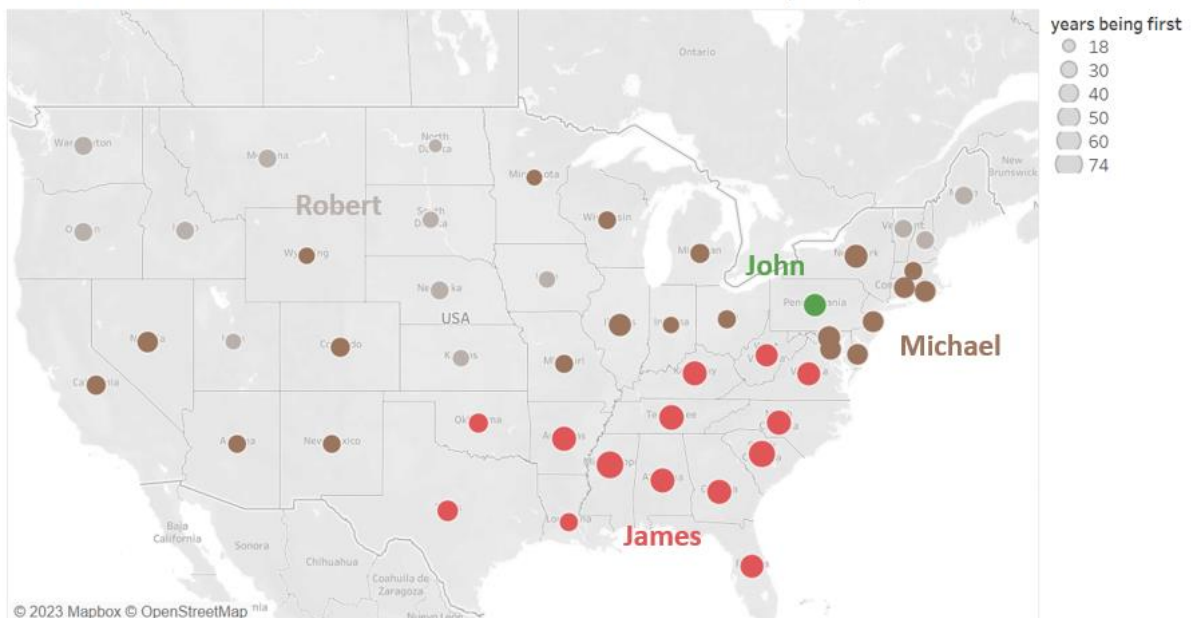
How many years the name was the top name in each state(female)

在這兩張圖中，統計的變數為各州自 1910 年至 2012 年間，各個名字成為男/女該年在各州最常取的名字的次數，並找出各州次數最多的名字，其中顏色代表不同的名字，圓圈的大小則代表其為最常取名字的年數。目標是找出名字和次數是否和美國各州地理上的分布有關聯性。

在第一張男性的分布圖中，可以看到名字的分布和各州的分布有密切的關係，如 Robert 大多在北方、James 集中在東南方等等，並不是隨機散佈在地圖的各處。

在第二張女性的分布圖中，除了 Oregon 以外，其他的州常年最常取的名子皆是 Mary，然而觀察其圓圈大小，亦即 Mary 在多少年中為最常見女嬰名，仍能發現圓圈較大的大多集中於東南部，西北部則較小，仍然可以看出在 Mary 流行時間上，仍和各州於地理上的分布有一定關係。

Hint: Feel free to draw an external dataset to enrich your findings--- Remember what we learned in Week 2 about 'numeracy' or 'landmark numbers'? This would be an excellent opportunity to put those skills to practice!

Section 2: Critique with tool/framework

Task 4 — Critique Framework 1: Trifecta checkup (1 points)

Read the Trifecta checkup carefully at
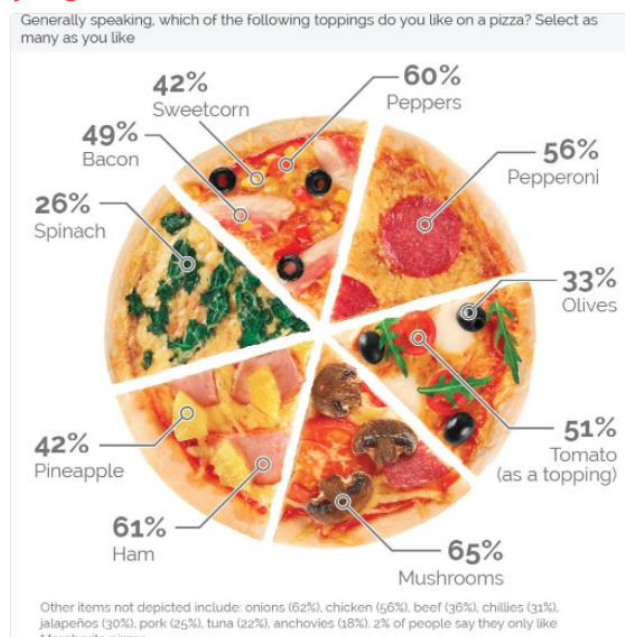http://junkcharts.typepad.com/junk_charts/junk-charts-trifecta-checkup-the-definitive-guide.html

and answer following questions:

1) Find one Type DV example on the Internet and tell us why you think it is.

Forget pepperoni - mushroom is Britain's most liked pizza topping (65%), followed by onion (62%) and then ham (61%)
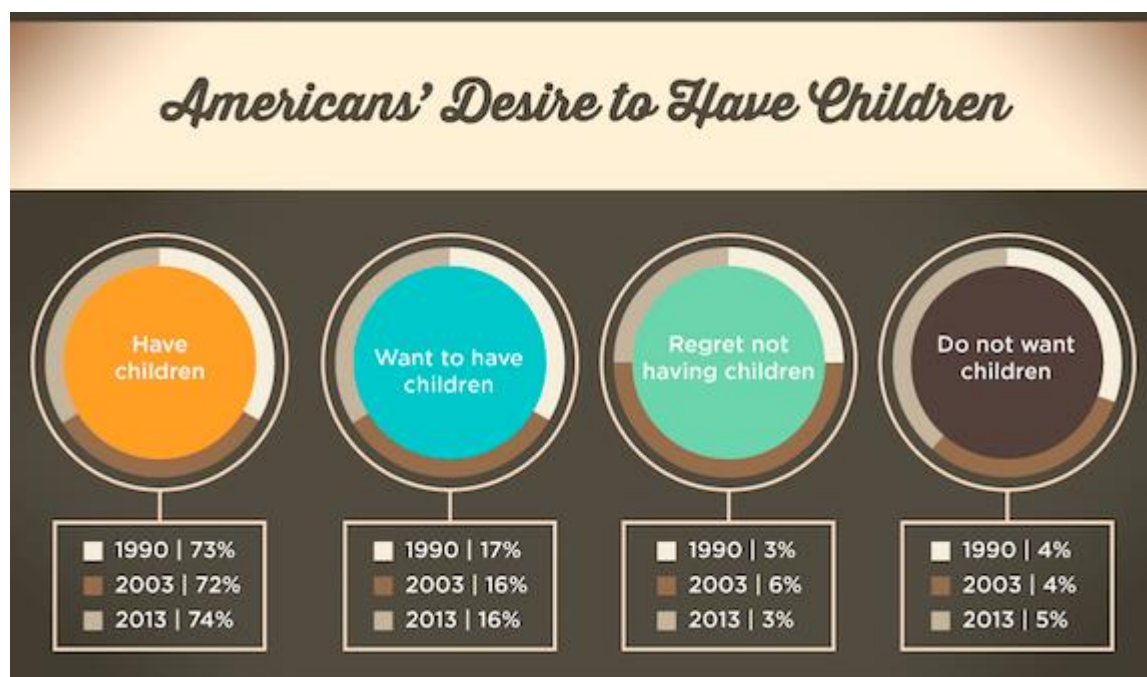yougov.co.uk/news/2017/03/0 ...

Generally speaking, which of the following toppings do you like on a pizza? Select as many as you like

42% Sweetcorn
49% Bacon
26% Spinach
60% Peppers
56% Pepperoni
33% Olives
51% Tomato (as a topping)
42% Pineapple
61% Ham
65% Mushrooms

Other items not depicted include: onions (62%), chicken (56%), beef (36%), chillies (31%), jalapeños (30%), pork (25%), tuna (22%), anchovies (18%). 2% of people say they only like Margherita pizzas.

Retrieved from: https://getdolphins.com/blog/the-worst-graphs-of-2017/

I consider is as a DV chart. First of all, the question is interesting and it briefly conclude the result of the study. However, some data is missing in the graph and do not state why they are only putted in the footnote. Last, the visualization is awful because the slices of pizza do not represent the proportion of percentage. There is also no particular order to the toppings. It is not desirable to be presented as a pie chart. Consequently, it is a DV chart.

2) Find one Type QDV example on the Internet and explain why you think it is.

Americans' Desire to Have Children

Have children
1990 | 73%
2003 | 72%
2013 | 74%

Want to have children
1990 | 17%
2003 | 16%
2013 | 16%

Regret not having children
1990 | 3%
2003 | 6%
2013 | 3%

Do not want children
1990 | 4%
2003 | 4%
2013 | 5%

Retrieved from: https://analythical.com/blog/examples-of-awful-data-visualization

I consider the graph as a QDV example. First, it addresses a stale question that is not intriguing for the readers, and it fails to represent what was found during the study with the given result. Moreover, the data is not sufficient or contains error that all choices within the same year do not have a sum of 100%. Lastly, the visualization is both inaccurate and misleading, the pie chart does not have a sum of 100%, and does not compare each choice within a given year. It fails to answer the question. Consequently, it is a QDV graph.

N.B. Remember to add the source/citation for your DV/QDV examples ☺

Task 5 — Critique Framework 2: Cairo's Visualization Wheel (1 points)

Step 1: Read Cairo's Ch3 - The Beauty Paradox: Art and Communication Efficacy or at least this online article about Cairo's Visualization Wheel at

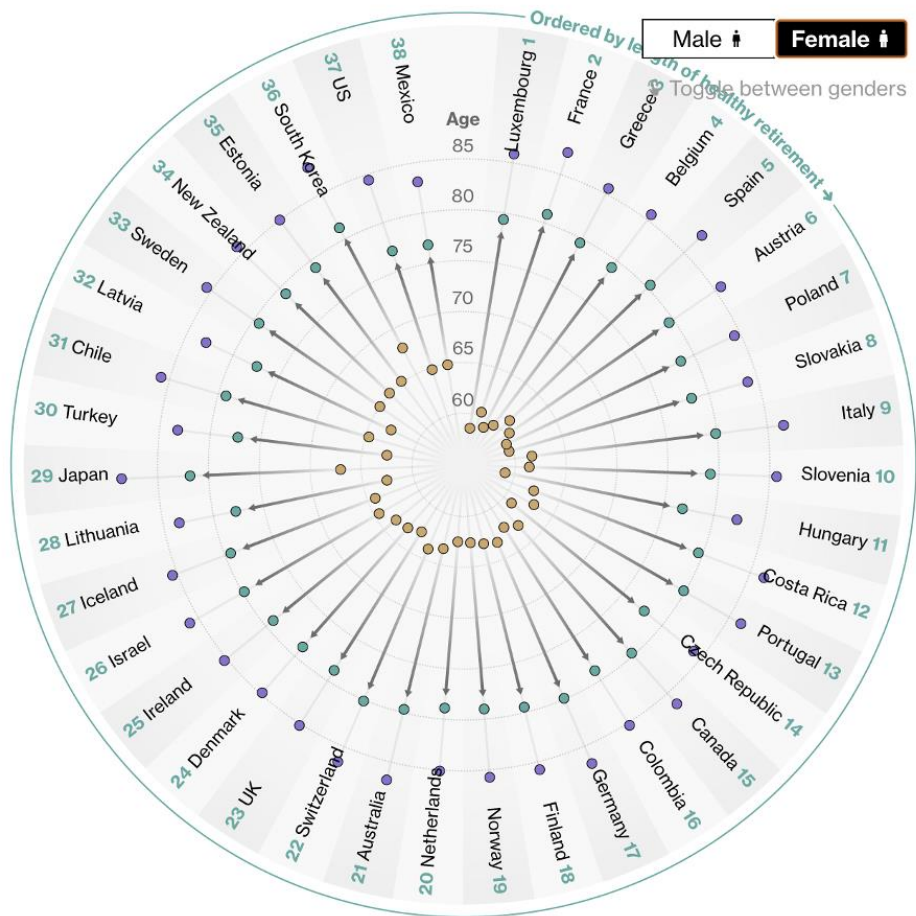https://ryanwingate.com/visualization/guidelines/visualization-wheel/

Step 2: Please use the framework of Cairo's Visualization Wheel to **write a critique** for the following set of visualization(s). Discuss the visualization below with different axes such as functionality-decoration, and so on.

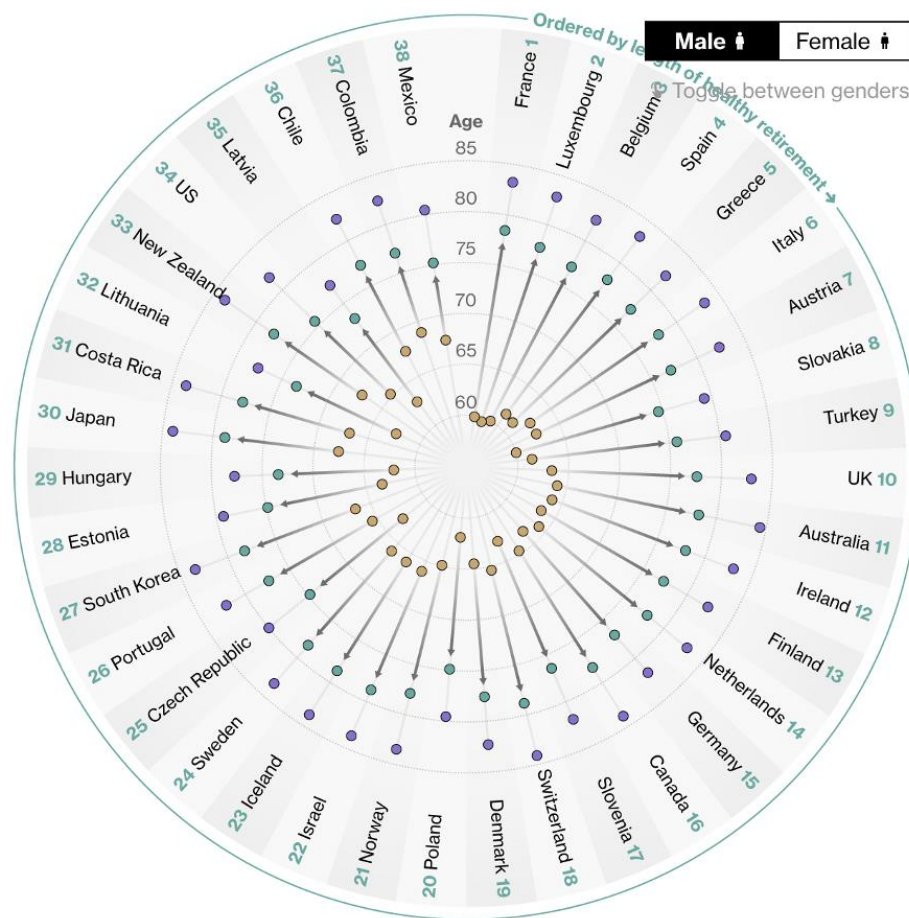The visualizations were from: Bloomberg News. (2023). The Countries with the longest — and shortest — retirements. https://www.bloomberg.com/graphics/2023-global-retirement-age/

Where Healthy Retirement Is the Longest

- Life expectancy
- Healthy life expectancy
- Effective retirement age

Ordered by length of healthy retirement

Male 👨 **Female 👩**

Toggle between genders

Age
85
80
75
70
65
60

1 Luxembourg
2 France
3 Greece
4 Belgium
5 Spain
6 Austria
7 Poland
8 Slovakia
9 Italy
10 Slovenia
11 Hungary
12 Costa Rica
13 Portugal
14 Czech Republic
15 Canada
16 Colombia
17 Germany
18 Finland
19 Norway
20 Netherlands
21 Australia
22 Switzerland
23 UK
24 Denmark
25 Ireland
26 Israel
27 Iceland
28 Lithuania
29 Japan
30 Turkey
31 Chile
32 Latvia
33 Sweden
34 New Zealand
35 Estonia
36 South Korea
37 US
38 Mexico

DDM _SP23 _Lab2_P9

Where Healthy Retirement Is the Longest

● Life expectancy ● Healthy life expectancy ● Effective retirement age

By using the Cairo's Visualization Wheel, we want to analyze the characteristics of this visualization graph. First, I consider the graph more figurative rather than abstract, since the concept of the study was well-shown by the chart; Next, I consider the graph more functional rather than decorative, there are almost no decoration to the chart and it is focused on the content of the data; The graph is also described as light rather than dense, since there is only three variables for each country, and the what the chart wants to show can easily be recognized in first glance. The graph is multidimensional because there are three variable including life expectancy, healthy life expectancy and effective retirement age; I consider the graph more original rather than familiar, since the representation is not commonly used especially how the author sort the countries. Lastly, I consider the graph more redundant rather than novel, since it uses both index and the length of the arrows to indicate the rankings of the length of healthy retirement.

In conclusion, the graph is figurative, functional, light, multidimensional, original, and redundant according to Cairo's Visualization Wheel.

Hello~ Have you recently discovered any interesting datasets that would be suitable for a data visualization exercise, regardless of whether they are excellent or … unusual? Feel free to share some examples and briefly explain why you find them compelling. We could consider incorporating them into future DDM classes! ☺