Data Visualisation

COS30045

[PROJECT TOPIC]

Fabian Sigmund 105015135

Vasudevan Karthikeyan 14085344

Tutor: Afzal Azeem Chowdhary

Tutorial Time: Monday 12:30

2024 Semester 1

Word Count:

https://mercury.swin.edu.au/cos30045/s105015135/data-vis-project/

Table of Contents

1. Introduction

1.1 Background and Motivation

The Concept of Life expectancy is a well-regarded subject by health authorities, governments, social services, and other interested parties. The notion of a person's expected life span being determined by specific environmental variables is of fascination. The subject is of relevance in the assessment of measuring the growth and development of countries. The gap between life expectancies of developed nations in Europe, to under-developed countries in Africa, have been instrumental in such countries earing the status of a "Third-world nation."

1.2 Visualisation Purpose

The lower life expectancies of such countries are determined by their lack of health care, violence, poverty and multiple other factors. However, some factors play a significant role more than others. The purpose of showcasing of validating or disproving any pre-determined notions, regarding life expectancy, is the motivation behind our visualisation.

2. Data

2.1 Data Source

The primary data source for the visualization is the life expectancy table from the OECD Health Statistics, accessible via the following link: [OECD Data Explorer • Health status].

Time period	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
Reference area																		
Sex: Total																		
Austria	68.7	69.7	69.4	69.6	70	69.8	70.2	70	70.2	69.9	70.1	70.3	70.7	71.2	71.2	71.4	71.8	72.2
Canada																		
Chile	E 57.3	E 57.7	E 58.1	E 58.6	E 59	E 59.5	E 60	E 60.6	E 61.1	E 61.7	E 62.3	E 63	E 63.6	E 64.3	E 65	E 65.7	E 66.4	E 67.1
Colombia	55.2	55.6	56.1	56.5	56.9	57.3	57.8	58.2	58.7	59.1	59.5	59.9	60.3	60.7	61.1	61.5	62	62.5
Costa Rica	60.4	61	61.6	62.2	62.8	63.3	63.9	64.4	64.8	65.3	65.8	66.4	66.9	67.5	68.2	68.8	69.5	70.2
Czechia	70.7	70.7	70	70.5	70.7	70.4	70.6	70.5	70	69.6	69.6	69.8	70.3	70.1	70.2	70.5	70.7	70.7
Denmark	72.4							72.6	73	73.2	73.3	73.3	73.4	73.6	73.9	74.2	73.8	74.7
France	70.3	71	70.5	70.5	71.4	71.3	71.5	71.5	71.5	71.3	72.2	72.1	72.4	72.5	72.8	73	73.2	73.8
Germany	69.1	69.7	69.9	70	70.6	70.5	70.6	70.9	70.5	70.3	70.6	70.8	71	71.3	71.5	71.4	71.8	72.5
Greece		72.1	71.7	72	71.8	72.4	72.6	72.4	72.4	73	73.8	74.2	73.8	74.1	74.6	74.4	74.5	74.7
Hungary	68.1	69	67.9	69	69.5	69.2	69.9	69.5	69.3	69.4	69.2	69.1	69.8	69.6	69.3	69.4	69.7	69.9
Iceland		73.5	73.7	73	73.5	73.8	73.3	73.8	74	73.7	73.8	73.5	74.4	74.4	74.4	75.5	76.9	76.3
Italy		69.8										72			72.6	72.7	72.8	73.4

Figure 1: Life Expectancy Table of the OECD Health Statistics

Figure 1 displays a table that outlines the life expectancy in various countries from 1960 to 2022. It measures life expectancy in years and provides separate sections for women, men, and the overall population.

An indicator of life expectancy is the gross domestic product (GDP). The following table was retrieved from the OECD Economy Statistics, accessible via the following link: [OECD Data Explorer • NAAG Chapter 1: GDP]

Time period	1970	1971	1972	1973	1974	1975	1976	1977
Reference area								
Combined unit of measure: US dollars, PPP converted, Billions, Current prices								
Australia	58.91	64.19	69.87	78.91	86.39	96.93	104.37	110.36
Austria	E 28.57	E 31.56	E 34.97	E 38.68	E 43.83	E 47.71	E 52.64	E 58.76
Belgium	E 37.28	E 40.64	E 44.63	E 49.96	E 56.74	E 61.17	E 68.19	E 72.88
Canada	99.26	108.49	119.42	134.44	151.46	168.05	187.74	206.45
Chile								
Colombia						E 33.21	E 36.66	E 40.57
Costa Rica								
Czechia								
Denmark	20.96	22.69	24.60	27.01	29.11	31.34	35.03	37.90
Estonia								
Finland	E 15.57	E 16.75	E 18.82	E 21.24	E 23.90	E 26.59	E 28.15	E 29.97
France	191.73	212.16	231.32	259.47	294.98	319.21	351.45	386.22
Germany	E 314.13	E 340.40	E 370.38	E 409.34	E 450.15	E 487.58	E 539.88	E 592.61
Greece	E 27.51	E 31.17	E 35.83	E 40.85	E 41.66	E 48.41	E 54.58	E 59.67
Hungary								
Iceland	E 0.79	E 0.93	E 1.04	E 1.17	E 1.34	E 1.48	E 1.65	E 1.91
Ireland	E 7.25	E 7.88	E 8.75	E 9.67	E 10.99	E 12.69	E 13.57	E 15.60
Israel								E 23.25
Italy	E 195.36	E 208.99	E 226.07	E 255.45	E 293.76	E 314.26	E 355.17	E 386.90
Japan	E 348.11	E 382.95	E 433.11	E 493.54	E 531.37	E 598.53	E 656.57	E 727.99
Korea	19.36	22.49	25.15	30.48	36.39	42.87	51.21	61.11
Latvia								

Figure 2: Gross Domestic Product of the OECD Economy Statistics

GDP per capita [GDP per capita (ourworldindata.org)]

	A	B	C	D	E
1	Entity	Code	Year	GDP per capita, PPP (constant 2017 international $)	
2	Afghanistan	AFG	2002	1280.4631	
3	Afghanistan	AFG	2003	1292.3335	
4	Afghanistan	AFG	2004	1260.0605	
5	Afghanistan	AFG	2005	1352.3207	
6	Afghanistan	AFG	2006	1366.9932	
7	Afghanistan	AFG	2007	1528.3446	
8	Afghanistan	AFG	2008	1556.8445	
9	Afghanistan	AFG	2009	1823.7426	
10	Afghanistan	AFG	2010	2026.1638	
11	Afghanistan	AFG	2011	1961.0963	
12	Afghanistan	AFG	2012	2122.8308	
13	Afghanistan	AFG	2013	2165.3408	
14	Afghanistan	AFG	2014	2144.4497	
15	Afghanistan	AFG	2015	2108.714	
16	Afghanistan	AFG	2016	2101.422	
17	Afghanistan	AFG	2017	2096.093	
18	Afghanistan	AFG	2018	2060.699	
19	Afghanistan	AFG	2019	2079.9219	
20	Afghanistan	AFG	2020	1968.341	
21	Afghanistan	AFG	2021	1516.3057	
22	Albania	ALB	1990	4827.9053	
23	Albania	ALB	1991	3497.0054	
24	Albania	ALB	1992	3265.4143	
25	Albania	ALB	1993	3599.4646	

Figure 3: GDP per Capita

2.2 Data Processing

Many of the tables lack data points for certain years, which pose a challenge when utilizing this data for visualization purposes. We decided to only visualise the data of the countries which are not missing any data points from 1980 to 2021.

2.2.1 Life Expectancy

1. Start with the excel file retrieved from the OECD Health statistics and highlight all blank fields to get an overview.

- Using conditional formatting -> New Rule… -> Format only cells that contain -> Blanks

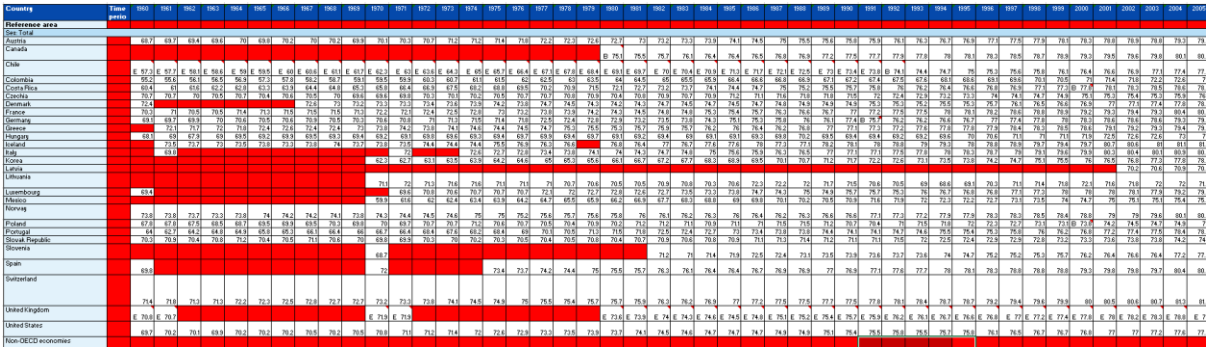
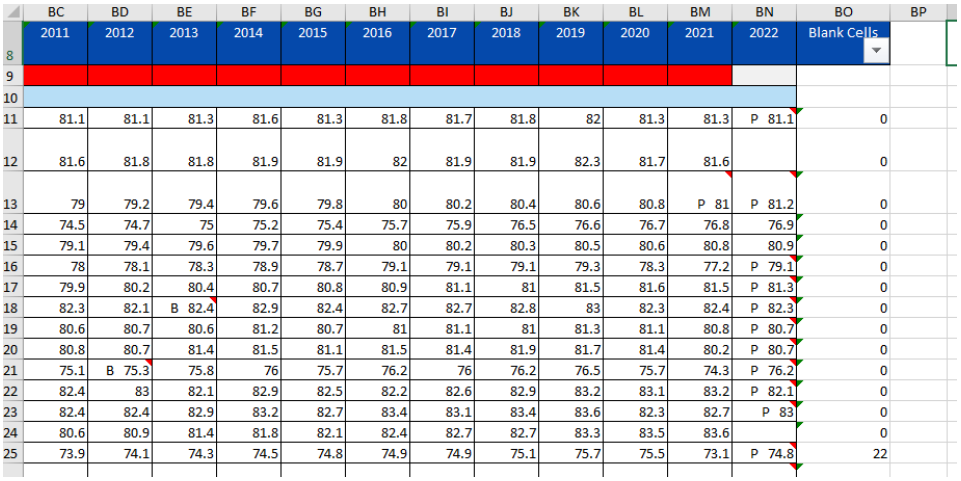


Figure 4*:* Life expectancy table with highlighted blank cells.

2. Create a helper column that counts the blank cells in each row

- Using the Formula '=COUNTBLANK(X11:BM11)' and using the fill handle we can count the blank cells from 1980 to 2021 for each row

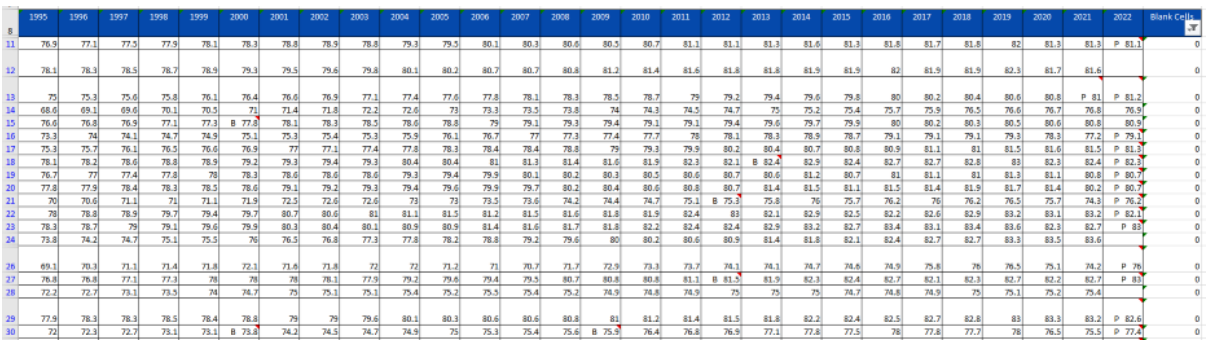


	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP
8	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Blank Cells	
9														
10														
11	81.1	81.1	81.3	81.6	81.3	81.8	81.7	81.8	82	81.3	81.3	P 81.1	0	
12	81.6	81.8	81.8	81.9	81.9	82	81.9	81.9	82.3	81.7	81.6		0	
13	79	79.2	79.4	79.6	79.8	80	80.2	80.4	80.6	80.8	P 81	P 81.2	0	
14	74.5	74.7	75	75.2	75.4	75.7	75.9	76.5	76.6	76.7	76.8	76.9	0	
15	79.1	79.4	79.6	79.7	79.9	80	80.2	80.3	80.5	80.6	80.8	80.9	0	
16	78	78.1	78.3	78.9	78.7	79.1	79.1	79.1	79.3	78.3	77.2	P 79.1	0	
17	79.9	80.2	80.4	80.7	80.8	80.9	81.1	81	81.5	81.6	81.5	P 81.3	0	
18	82.3	82.1	B 82.4	82.9	82.4	82.7	82.7	82.8	83	82.3	82.4	P 82.3	0	
19	80.6	80.7	80.6	81.2	80.7	81	81.1	81	81.3	81.1	80.8	P 80.7	0	
20	80.8	80.7	81.4	81.5	81.1	81.5	81.4	81.9	81.7	81.4	80.2	P 80.7	0	
21	75.1	B 75.3	75.8	76	75.7	76.2	76	76.2	76.5	75.7	74.3	P 76.2	0	
22	82.4	83	82.1	82.9	82.5	82.2	82.6	82.9	83.2	83.1	83.2	P 82.1	0	
23	82.4	82.4	82.9	83.2	82.7	83.4	83.1	83.4	83.6	82.3	82.7	P 83	0	
24	80.6	80.9	81.4	81.8	82.1	82.4	82.7	82.7	83.3	83.5	83.6		0	
25	73.9	74.1	74.3	74.5	74.8	74.9	74.9	75.1	75.7	75.5	73.1	P 74.8	22	

Figure 5: Life expectancy table with blank cells count.

3. Filter the years which have a blank cell count unequal to zero

- Using the excels built in filtering function we can remove any rows that have missing data points by using the helper column "blank cells" that we created in step 2.



	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Blank Cells
8																													
11	76.9	77.1	77.5	77.9	78.1	78.3	78.8	78.9	78.8	79.3	79.5	80.1	80.3	80.6	80.5	80.7	81.1	81.1	81.3	81.6	81.3	81.8	81.7	81.8	82	81.3	81.3	P 81.1	0
12	78.1	78.3	78.5	78.7	78.9	79.3	79.5	79.6	79.8	80.1	80.2	80.7	80.7	80.8	81.2	81.4	81.6	81.8	81.8	81.9	81.9	82	81.9	81.9	82.3	81.7	81.6		0
13	75	75.3	75.6	75.8	76.1	76.4	76.6	76.9	77.1	77.4	77.6	77.8	78.1	78.3	78.5	78.7	79	79.2	79.4	79.6	79.8	80	80.2	80.4	80.6	80.8	P 81	P 81.2	0
14	68.6	69.1	69.6	70.1	70.5	71	71.4	71.8	72.2	72.6	73	73.3	73.5	73.8	74	74.3	74.5	74.7	75	75.2	75.4	75.7	75.9	76.5	76.6	76.7	76.8	76.9	0
15	76.6	76.8	76.9	77.1	77.3	B 77.8	78.1	78.3	78.5	78.6	78.8	79	79.1	79.3	79.4	79.1	79.3	79.4	79.6	79.7	79.9	80	80.2	80.3	80.5	80.6	80.8	80.9	0
16	73.3	74	74.1	74.7	74.9	75.1	75.3	75.4	75.3	75.9	76.1	76.7	77	77.3	77.4	77.7	78	78.1	78.3	78.9	78.7	79.1	79.1	79.3	78.3	77.2	P 79.1	0	
17	75.3	75.7	76.1	76.5	76.6	76.9	77	77.1	77.4	77.8	78.3	78.4	78.4	78.8	79	79.3	79.5	80.2	80.4	80.7	80.8	80.9	81.1	81	81.5	81.6	81.5	P 81.3	0
18	78.1	78.2	78.6	78.8	78.9	79.2	79.3	79.4	79.3	80.4	80.4	81	81.3	81.4	81.6	81.9	82.3	82.1	B 82.4	82.9	82.4	82.7	82.7	82.8	83	82.3	82.4	P 82.3	0
19	76.7	77	77.4	77.8	78	78.3	78.6	78.6	78.6	79.3	79.4	79.9	80.1	80.2	80.3	80.5	80.6	80.7	80.6	81.2	80.7	81	81.1	81	81.3	81.1	80.8	P 80.7	0
20	77.8	77.9	78.4	78.3	78.5	78.6	79.1	79.2	79.3	79.4	79.6	79.9	79.7	80.2	80.4	80.6	80.8	80.7	81.4	81.5	81.1	81.5	81.4	81.9	81.7	81.4	80.2	P 80.7	0
21	70	70.6	71.1	71	71.1	71.9	72.5	72.6	72.6	73	73	73.5	73.6	74.2	74.4	74.7	75.1	B 75.3	75.8	76	75.7	76.2	76	76.2	76.5	75.7	74.3	P 76.2	0
22	78	78.8	78.9	79.7	79.4	79.7	80.7	80.6	81	81.1	81.5	81.2	81.5	81.6	81.8	81.9	82.4	83	82.1	82.9	82.5	82.2	82.6	82.9	83.2	83.1	83.2	P 82.1	0
23	78.3	78.7	79	79.1	79.6	79.9	80.3	80.4	80.1	80.9	80.9	81.4	81.6	81.7	81.8	82.2	82.4	82.4	82.9	83.2	82.7	83.4	83.1	83.4	83.6	82.3	82.7	P 83	0
24	73.8	74.2	74.7	75.1	75.5	76	76.5	76.8	77.3	77.8	78.2	78.8	79.2	79.6	80	80.2	80.6	80.9	81.4	81.8	82.1	82.4	82.7	82.7	83.3	83.5	83.6		0
26	69.1	70.3	71.1	71.4	71.8	72.1	71.6	71.8	72	72	71.2	71	70.7	71.7	72.9	73.3	73.7	74.1	74.1	74.7	74.6	74.9	75.8	76	76.5	75.1	74.2	P 76	0
27	76.8	76.8	77.1	77.3	78	78	78	78.1	77.9	79.2	79.6	79.4	79.5	80.7	80.8	80.8	81.1	B 81.5	81.9	82.3	82.4	82.7	82.1	82.3	82.7	82.2	82.7	P 83	0
28	72.2	72.7	73.1	73.5	74	74.7	75	75.1	75.1	75.4	75.2	75.5	75.4	75.2	74.9	74.8	74.9	75	75	75	74.7	74.8	74.9	75	75.1	75.2	75.4		0
29	77.9	78.3	78.3	78.5	78.4	78.8	79	79	79.6	80.1	80.3	80.6	80.6	80.8	81	81.2	81.4	81.5	81.8	82.2	82.4	82.5	82.7	82.8	83	83.3	83.2	P 82.6	0
30	72	72.3	72.7	73.1	73.1	B 73.8	74.2	74.5	74.7	74.9	75	75.3	75.4	75.6	B 75.9	76.4	76.8	76.9	77.1	77.8	77.5	78	77.8	77.7	78	76.5	75.5	P 77.4	0

Figure 6: Life expectancy table with removed rows that have missing values.

3. Visualisation Design

The choice of visualisation suitable for expressing as much information as possible was a challenge. Life Expectancy of a country corresponds to multiple facets of a nation's infrastructure ranging from GDP to quality health services. Additionally, it was our intention to highlight the growth or decline of certain OECD countries over a period. Therefore, a visualisation capable of displaying time-series data was required.

Design 1 – Line graph

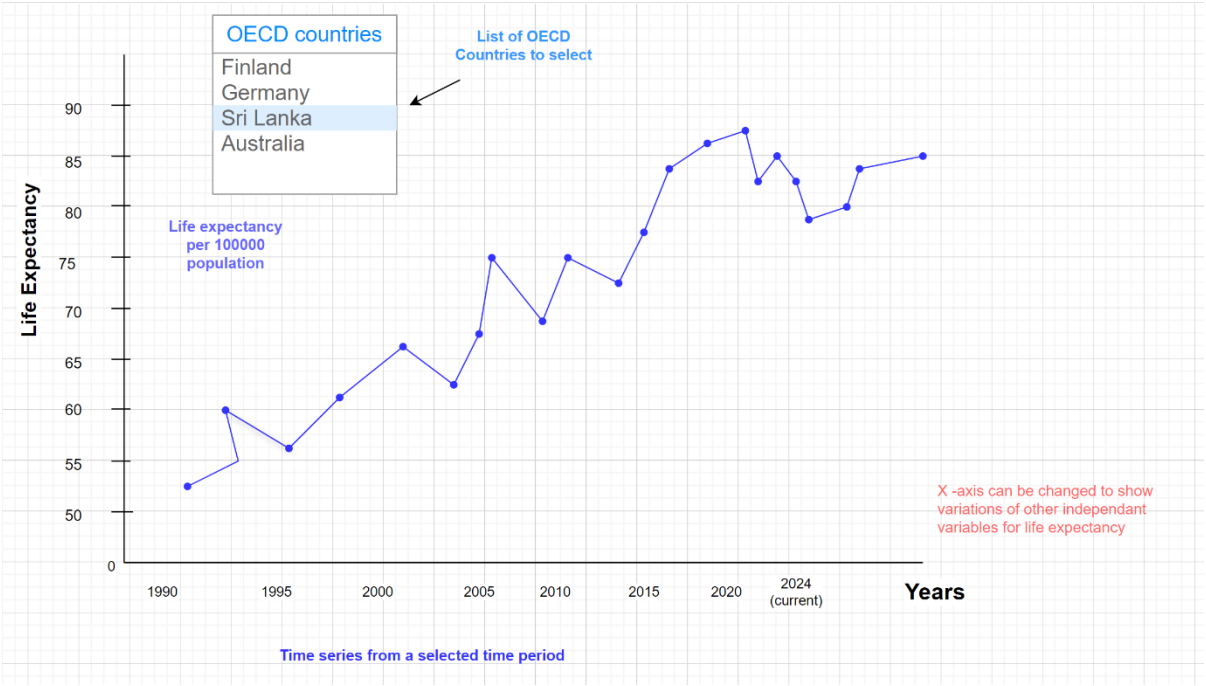


Figure 7: Sketch 1 – Line graph

A line chart was the initial chart though of when considering a time series. The above sketch illustrates a chart plotting Life Expectancy vs time. Line charts employ the positioning on the common scale as a channel to encode the trend of its variables. The sketch shows that chart can be altered through a drop-down list, to highlight the trend of life expectancy in different countries.

However, it was evident that the visualisation and design was too simplistic and un-interesting to proceed with. Despite its expressiveness with time, line charts are limited to the amount of information it can hold. Our intention was to create visualisation which can highlight different variations of life expectancy with different varieties of datasets. Therefore, a different approach was needed.

Design 2 – Heat Map

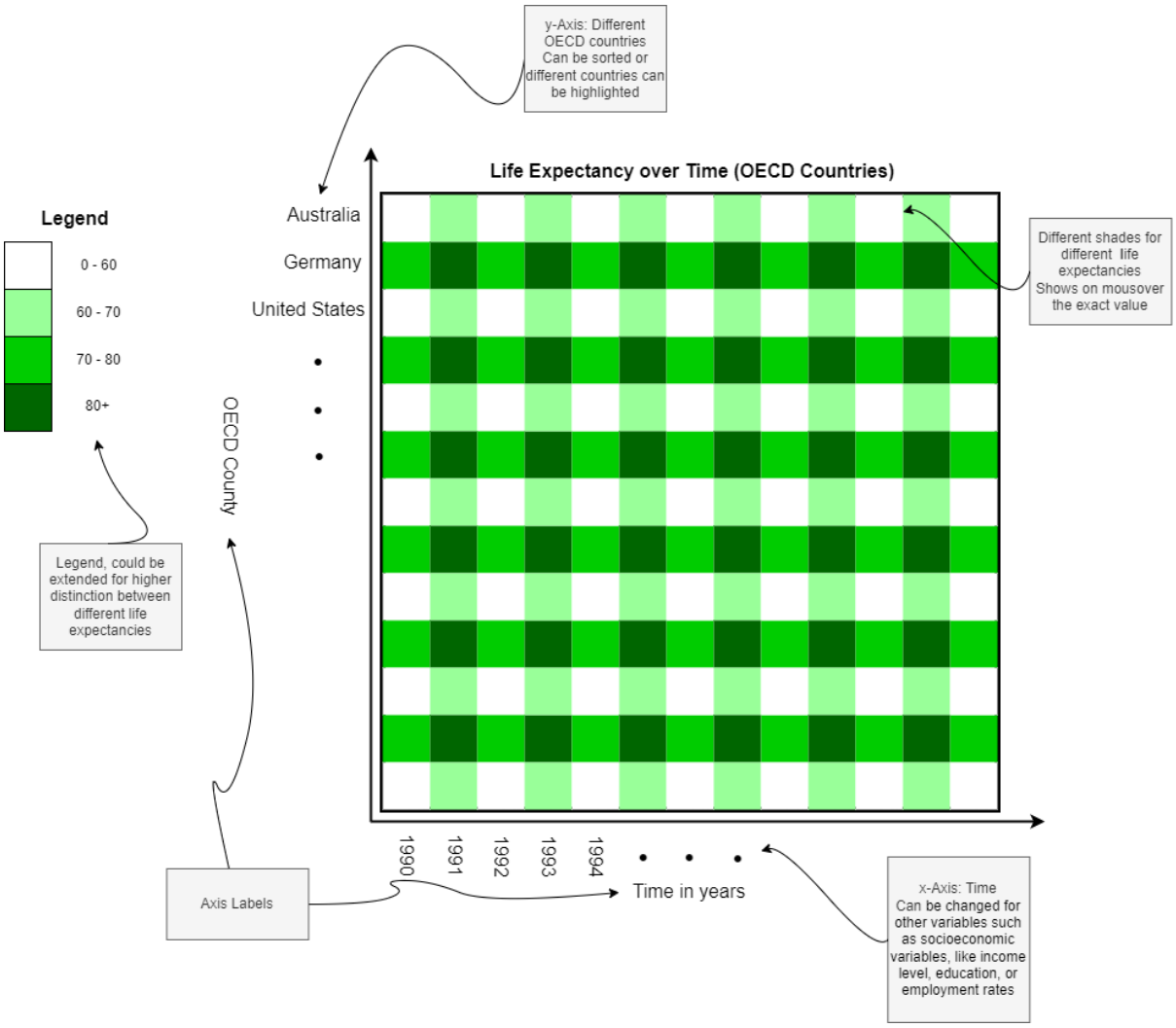


Figure 8: Sketch 2 – Heat Map

The second approach was a heat map. This design could encode the data of multiple OECD countries, in the same chart, without the need for transitions or updates. The time series aspect is encoded as well, like the line graph, but the channel used to encode life expectancy has been changed to colour saturation. A legend is used as reference to the ranges of life expectancy which have been sorted to 4 bins and colour coded.

This design is expected to highlight selected countries and expand the legend though to display further insights. Hence this chart design is very interactive, compared to the previous one. This design was successful in its ability to compare the life expectancies of multiple OECD countries over a period, but it is still short of the ideal visualisation, which can encode different facets of information.

Design 3 – Bubble chart

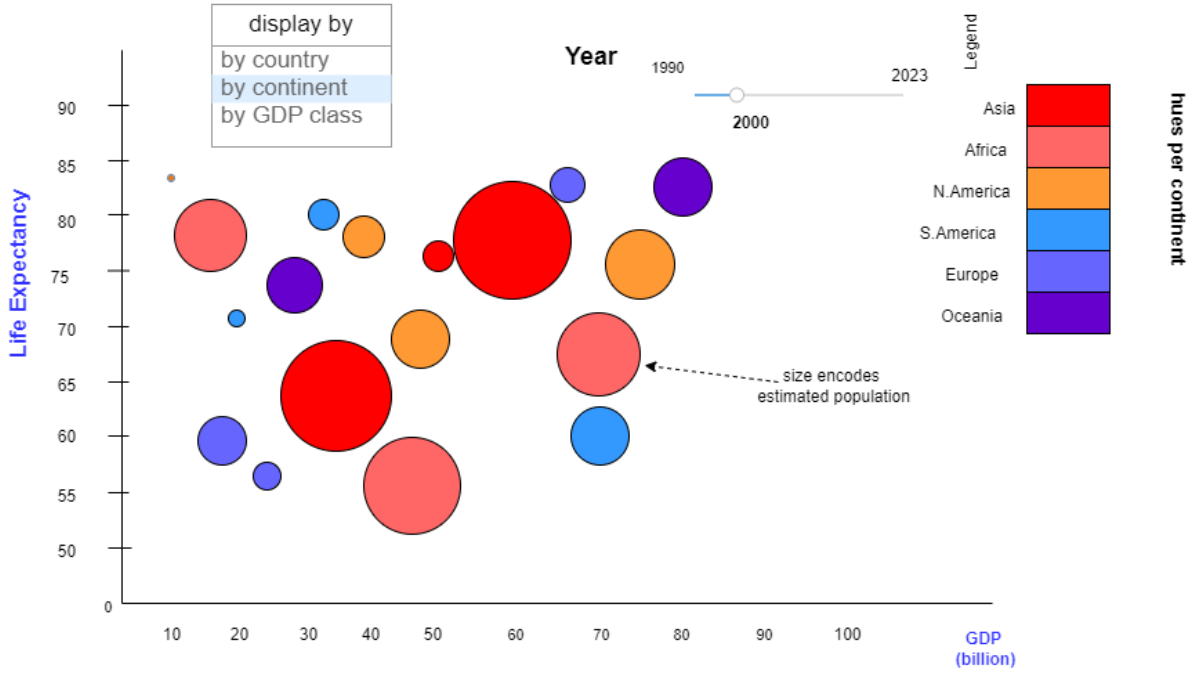


Figure 9: Sketch 3 – Bubble chart

This design encapsulates the successes of both previous designs. The OECD countries are represented by bubbles (mark) and visual encoding has been changes the distinct colour hues to distinguish the qualitative data and the spatial positioning to denote the quantitative data (life expectancy). The different hues are labelled in a legend.

Instead of a time series, this visualisation shows correlation of life expectancy with GDP or any other selected data in the X axis, the change of data across a selected time is viewable by a slider. The viewer can use the slider to update the data in the graph to a continuous sequence of years.

The chart uses the transitioning and updating features of design 1 but is not simplistic or limited. The above design shows the variations of life expectancies in the major regions of the world, but both the chart and legend can be changed to show individual countries or countries categorised based on GDP classes and many more.

The size of bubbles is an additional magnitude channel to be used to encode data, the above sketch shows it encoded for human population in the region, but it can be used to encode other factors such as magnitude of health spending, mortality rate or average GDP.

4. Validation

5. Conclusion

6. References

Notes:

- Week07:
  - Initial idea is to find correlations between mental health and various indicators such as GDP, access to doctors etc.
- Week08:
  - Shift the project topic, because there are no primary databases on mental health status
  - Inspiration: Ted Talk https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?language=en
  - We shifted the topic to “Health status” to be more flexible.

    Life expectancy: https://data-explorer.oecd.org/vis?fs[0]=Topic%2C0%7CHealth%23HEA%23&fs[1]=Topic%2C1%7CHealth%23HEA%23%7CHealth%20status%23HEA_STA%23&pg=0&fc=Topic&snb=16&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_HEALTH_STAT%40DF_HEALTH_STATUS&df[ag]=OECD.ELS.HD&df[vs]=1.0&pd=%2C&dq=.A.LFEXP.Y.Y0.....&ly[rw]=REF_AREA&ly[cl]=TIME_PERIOD&ly[rs]=SEX&to[TIME_PERIOD]=false&vw=tb
  -