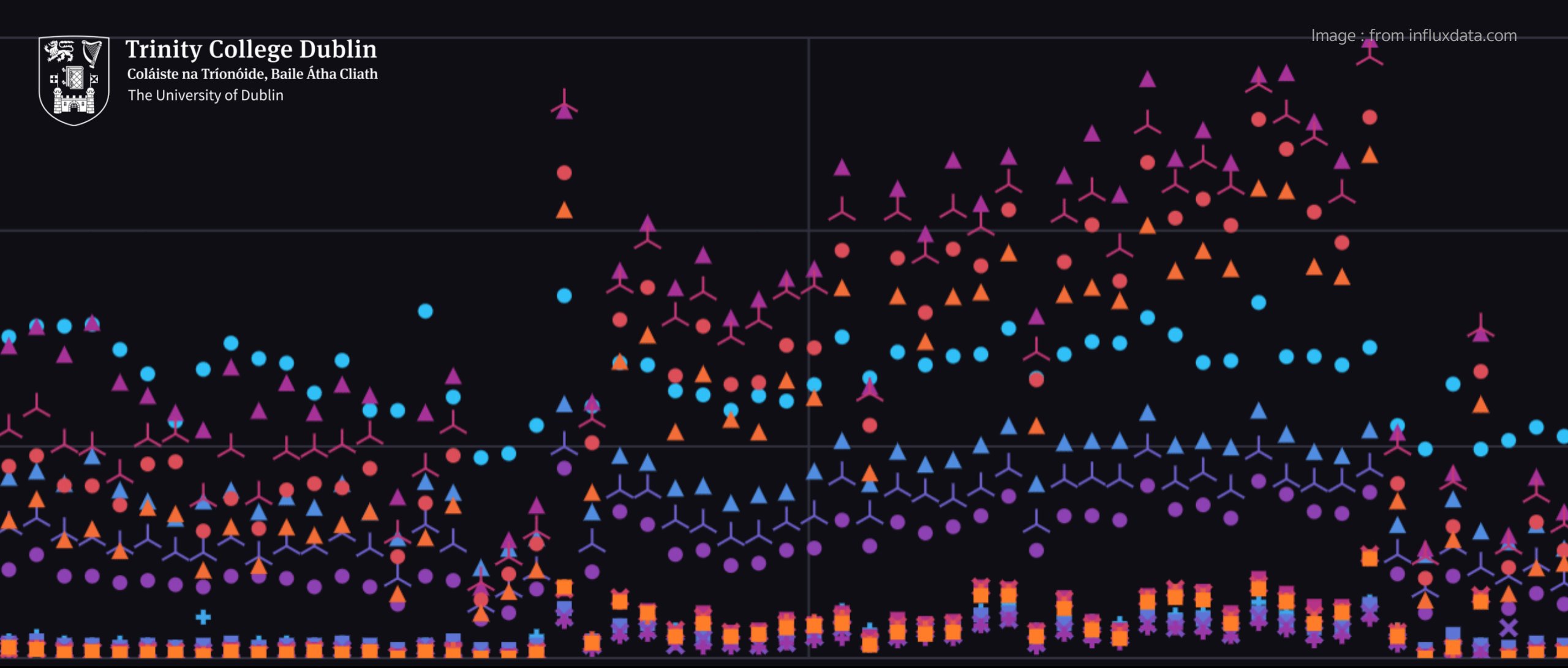


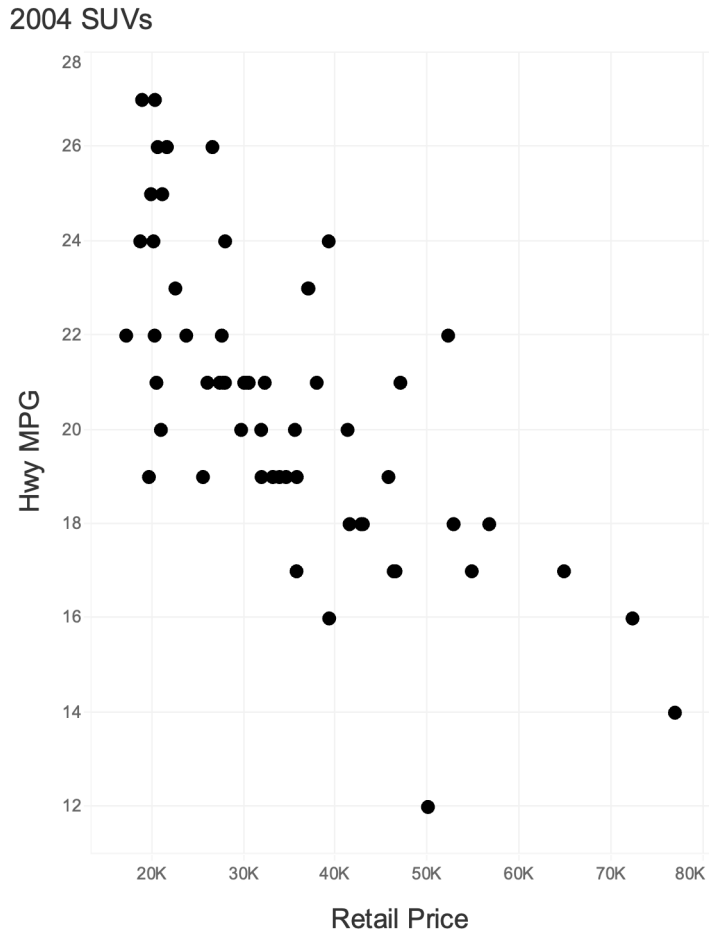
Assignment 1.2: Visual Encoding

Worth 14% of the module

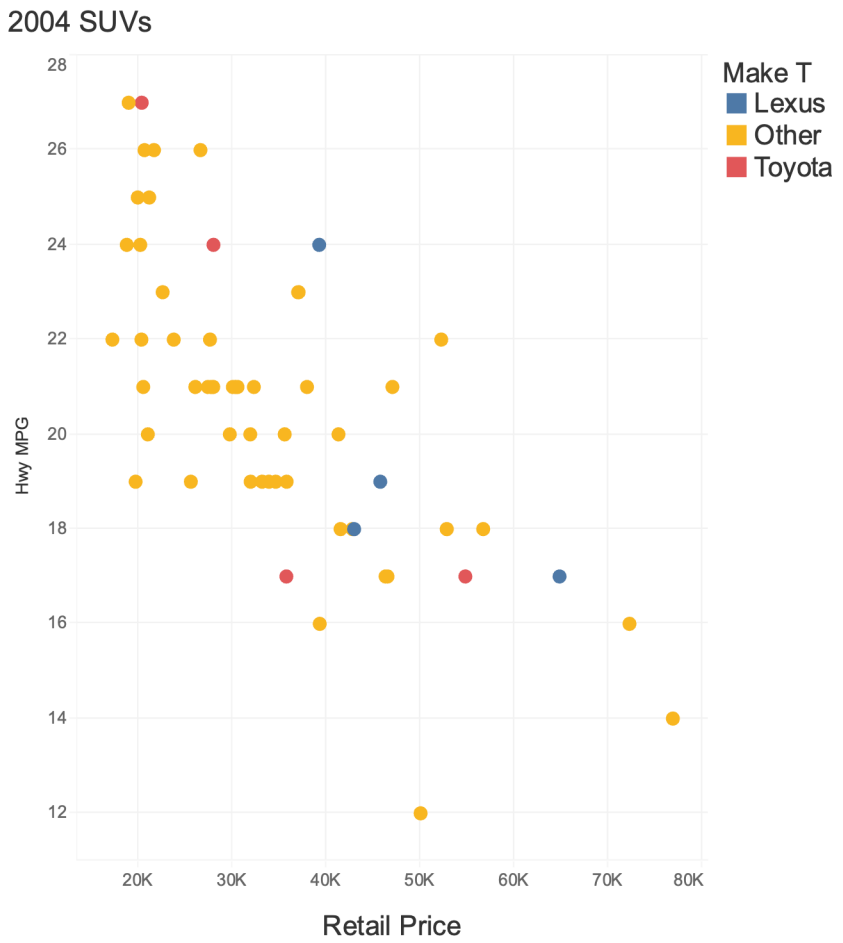
Due 21/10/2024

@ 23:59hrs

Recap: **Visual Encoding Channels**



Retail Price vs. Hwy MPG.

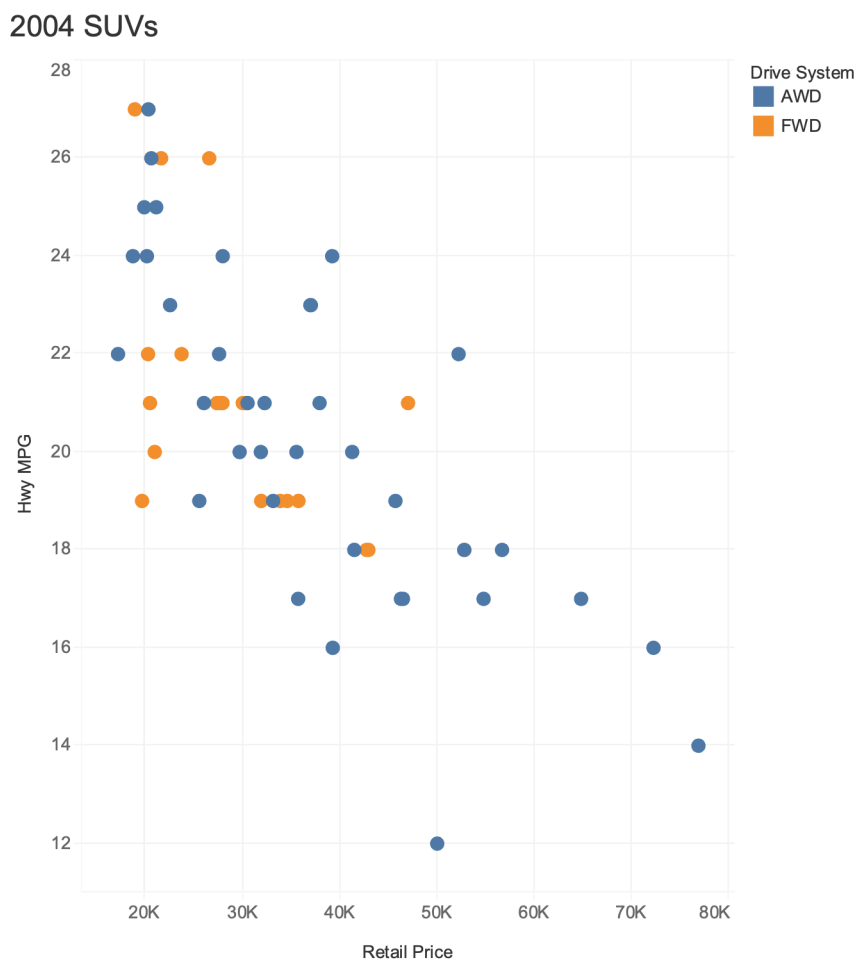


Retail Price vs. Hwy MPG. Color shows details about Make T.

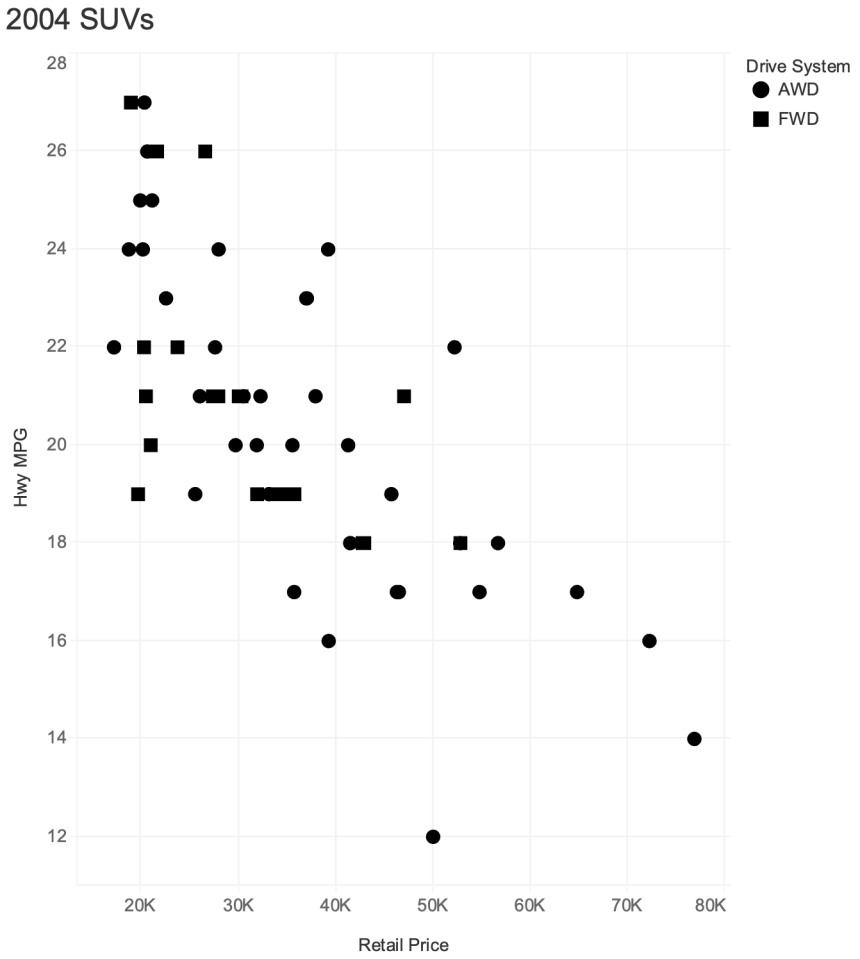
Example: adding colour helps distinguish a data point from others

Recap: **Visual Encoding Channels**

Associative?



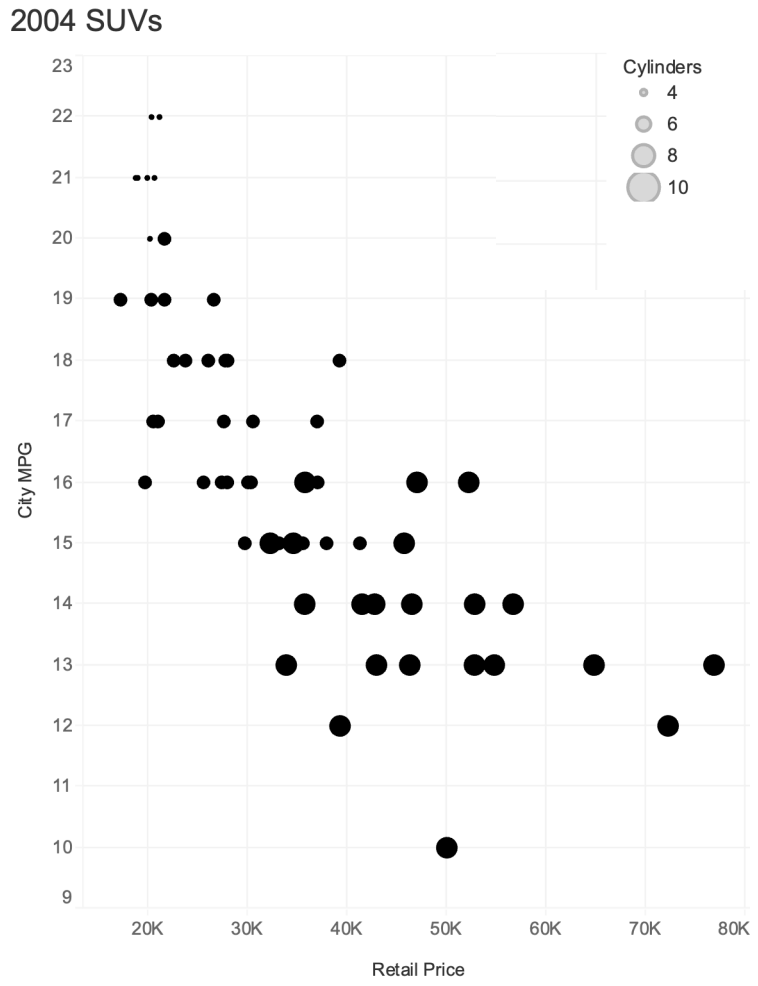
Retail Price vs. Hwy MPG. Color shows details about Drive System.



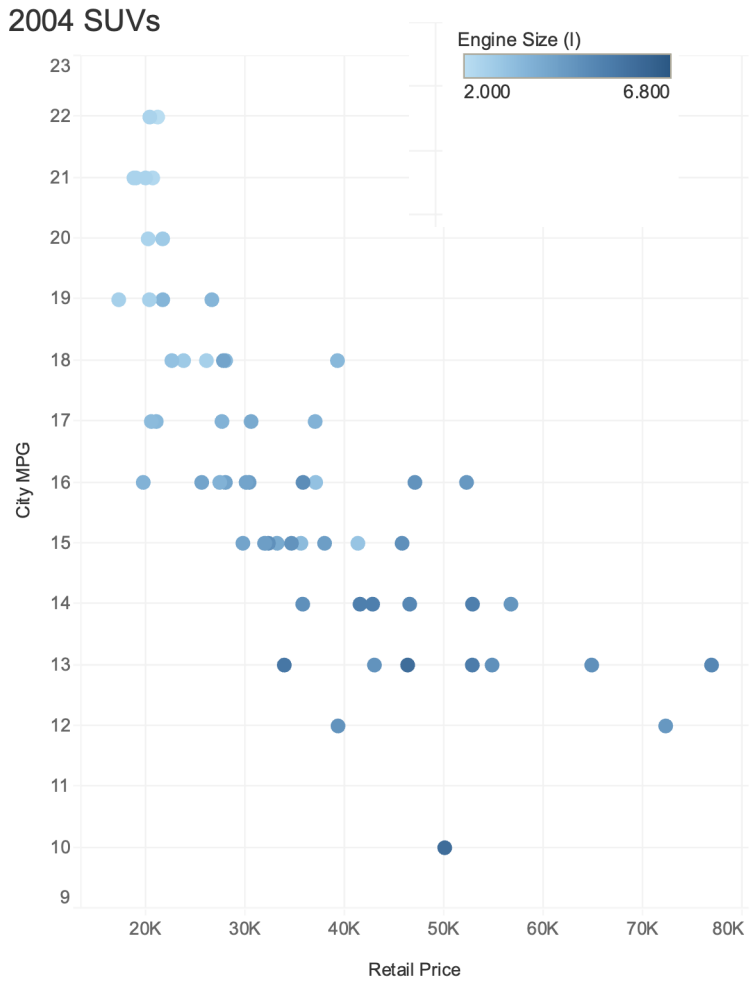
Retail Price vs. Hwy MPG. Shape shows details about Drive System.

Example: here colour (left) is better than shape (right) at associating/grouping data points of the same category

Recap: **Visual Encoding Channels**



Retail Price vs. City MPG. Size shows details about Cylinders.



Retail Price vs. City MPG. Color shows sum of Engine Size (l).

Left: size is used to encode an ordinal attribute; Right: brightness is used to encode a quantitative attribute

Assignment 1.2: Visual Encoding

[14% of the module Mark | Due on 21st October, 2024]

Motivations: Compare visual encoding channels in visualization; get accustomed to tools for implementing visualizations of any given data set

Data set: *gapminder.csv*, *available on on mymodule.tcd.ie* comprises the following attributes

- country, a categorical attribute
- continent, categorical attribute
- year, ordinal attribute, 5 year intervals from 1952-2007
- lifeExp, quantitative attribute representing health in terms of average life expectancy in years
- pop, a quantitative attribute representing population of the country in a particular year
- gdpPercap income, a quantitative attribute representing the wealth of the country in terms of GDP per capita

Original source (including documentation if needed) is available at:

https://cran.r-project.org/web/packages/gapminder/index.html

Note that the original Gapminder visualizations are famous for their use of animation, you are NOT required nor expected to use animation for this assignment

gapminder.csv

(just an excerpt shown here for space reasons)

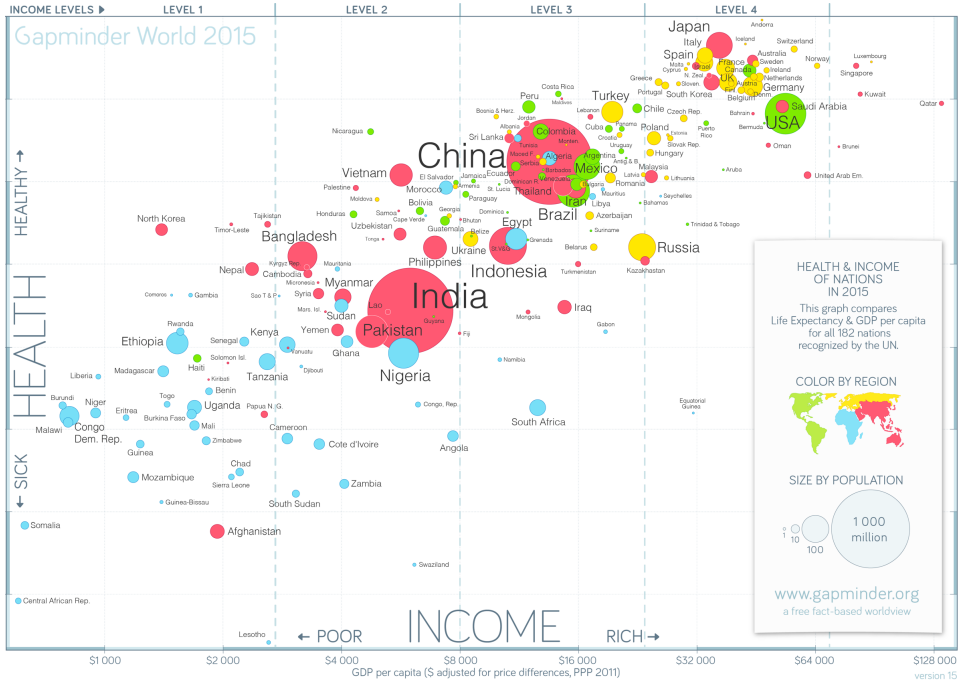
country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1952	28.801	8425333	779.4453145
Afghanistan	Asia	1957	30.332	9240934	820.8530296
Afghanistan	Asia	1962	31.997	10267083	853.10071
Afghanistan	Asia	1967	34.02	11537966	836.1971382
Afghanistan	Asia	1997	41.763	22227415	635.341351
Afghanistan	Asia	2002	42.129	25268405	726.7340548
Afghanistan	Asia	2007	43.828	31889923	974.5803384
Albania	Europe	1952	55.23	1282697	1601.056136
Albania	Europe	1957	59.28	1476505	1942.284244
Albania	Europe	1962	64.82	1728137	2312.888958
Albania	Europe	1967	66.22	1984060	2760.196931
Zimbabwe	Africa	1997	46.809	11404948	792.4499603
Zimbabwe	Africa	2002	39.989	11926563	672.0386227
Zimbabwe	Africa	2007	43.487	12311143	469.7092981

Objectives – Part A

Before you start, you should review the lecture notes on Visual Encoding channels in visualization.

Chart #1 - #3 : Visualize the correlation between wealth and health in the year 2002:

- Choose a visual encoding channel to express GDP (gdpPercap) and life expectancy (lifeExp) respectively (e.g. x-position-position). Use the same channel for these attributes in all charts of Part A
- Then create 3 variants of the charts that encode, the quantitative attribute population (pop) using 3 other encoding channels.
- Ensure that Ireland is distinguishable from other countries (BUT it is not required that every single country is identifiable i.e., other data points can be anonymous)
- Each chart must created for viewing as a figure no larger than half an A4 page
- It is optional to visualize the remaining attributes in this part.
- You are only required to visualize 2002 data for this part.
- In max half a page, add some brief text stating the tool(s) used for creating the visualizations. Discuss briefly your opinion on the efficacy of the visualizations in Charts #1 – 3.



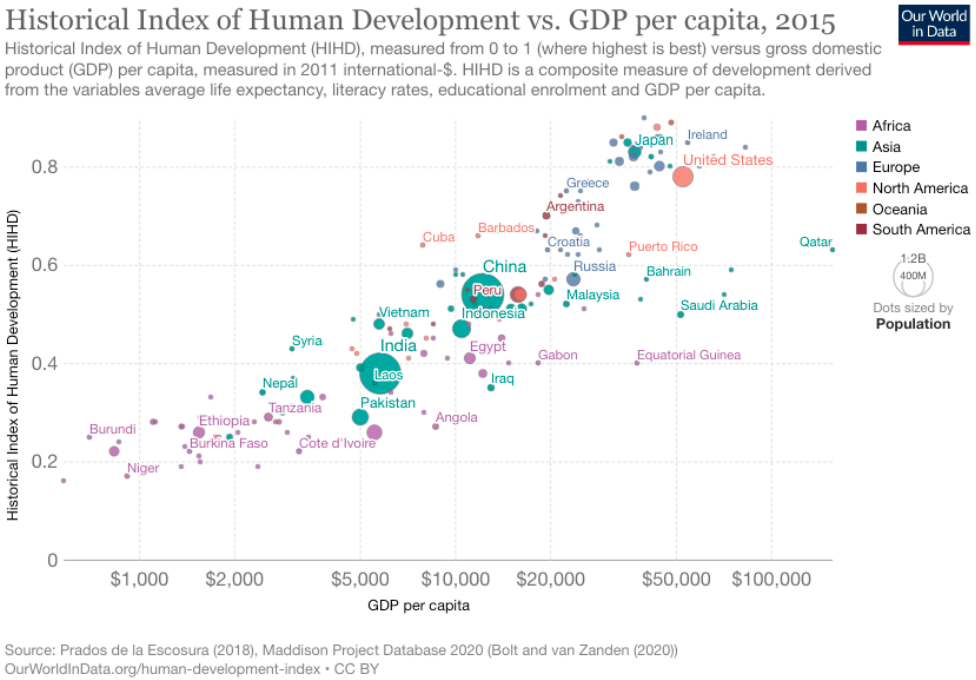
Example from gapminder.org You do NOT have to specifically replicate this chart, which is more complex than required for this assignment.

NOTE: The attributes highlighted in yellow must be graphically represented with some "visual encoding channel" (as discussed in Week 3's lectures). It is NOT enough to merely express them with text or a mouse-over tool-tip.

Objectives - Part B

Chart #4 - 6: Visualize the evolution of life expectancy over the years (1957-2007):

- Choose a visual encoding channels to express year and life expectancy (lifeExp) respectively (position in x and y is recommended). Use the same encoding for these attributes in all charts of Part B
- Then create 3 variants of the chart, encoding the categorical attribute continent using 3 other encoding channels of your choice; one variant of the chart for each channel.
- Ensure that Ireland is distinguishable from other countries (BUT it is not required that every single country is identifiable i.e., other data points can be anonymous)
- Each chart must created for viewing as a figure no larger than half an A4 page
- It is optional to visualize the remaining attributes in this part.
- In max half a page, add some brief text stating the tool(s) used for creating the visualizations. Discuss briefly your opinion on the efficacy of the visualizations in Charts #4 – 6.



Source: Prados de la Escosura (2018), Maddison Project Database 2020 (Bolt and van Zanden (2020))
OurWorldInData.org/human-development-index • CC BY

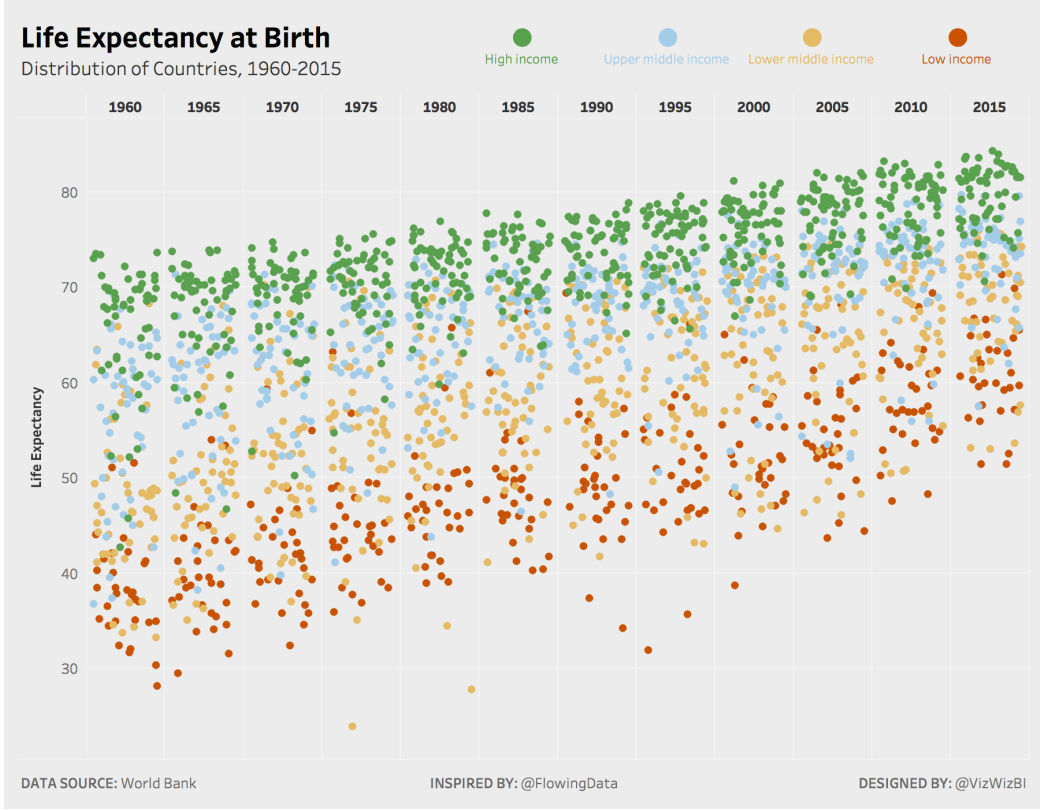
Example from OurWorldInData.org You do NOT have to specifically replicate this chart

NOTE: The attributes highlighted in yellow must be graphically represented with some "visual encoding channel" (as discussed in Week 3's lectures). It is NOT enough to merely express them with text or a mouse-over tool-tip.

Objectives – Part C

Chart #7: Exploratory Visualization of the full data set

- Create a single visualization that somehow, encodes the 5 attributes continent, year, lifeExp, pop, gdpPerCap, in a single chart
- Ensure that Ireland is distinguishable from other countries (BUT it is not required that every single country is identifiable i.e., other data points can be anonymous)
- Each chart must created for viewing as a figure no larger than half an A4 page
- In the remaining half page, you should add some brief text stating the tool(s) used for creating the visualizations. Discuss briefly your opinion on the efficacy of the visualization you created for Chart #7



Example from FlowingData.org You do NOT have to specifically replicate this chart

NOTE: The attributes highlighted in yellow must be graphically represented with some "visual encoding channel" (as discussed in Week 3's lectures). It is NOT enough to merely express them with text or a mouse-over tool-tip.

Submission Requirements

- You may use **any tool/library/language** of your choice for this assignment.
- You can use more than one tool and you do not have to use the same tool for any part of the assignment
- You may use **any chart type**, but the simplest and most obvious for this density of data points may be some form of a scatter plot.

Submit your visualization in a PDF file

- Each chart **must not be larger than half an A4 page in size**, including a brief caption noting which encoding channels were used. You should consider the size that the visualizations will be in the submitted PDF and chose an appropriate scale of text and other chart elements for this size
- IF you happen to create an animated or interactive visualization [this is completely optional], you may provide a video or link, but a static screen-shot of this must be included in the PDF (within the page limit) and the chart should be optimized for display on in a size equivalent ½ A4 page

Submit your code/project file and any supplementary data/materials in a Zip file, however, you should not include the main PDF in a Zip file.
Please note that, while mandatory, the zip is merely for the purposes of ensuring fairness and authenticity. Marking will be wholly based on the PDF.

Required Format / Information

This assignment must be submitted along with your submission for A1.2.

You should submit a single PDF of MAX 7 pages in length combining A1.1 (max 2 pages) and A1.2 (max 5 pages) (This assumes the text of the report is roughly in standard 10-12pt font. An optional template is provided on Blackboard)

Your submission MUST include the following information

- CS7DS4 / CSU44065 Data Visualization 2024-25
- Assignment 1
- Student Name : *<your name>*
- Student No: *<your student number>*
- Declaration:

  "I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

  I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write."

Please include your name in the title e.g., A1_SurnameFirstname.pdf

Assessment Policy

If contacting me by Email: John.Dingliana@tcd.ie; Please include the module code 'cs7ds4' or 'csu44056' in the subject header to ensure it is seen

Do NOT submit Assignments by Email

Late submission:

- In this module, for all assignments, unless stated otherwise in writing, a **10% penalty** applies for each day late up to a maximum of 6 days. After this, the mark will be capped at 40%.
- No further submissions will be accepted after 31/12/2024 – the grade will be returned as 0 after this date.
- If you submit after the deadline, you should Email me to let me know that a submission has been made.
- There will be a grace period of 2 hours after the deadline for potential accidental/technical issues. Thus, there is no need to email me if you are just a few minutes late

Extensions (for exceptional circumstances): Submit a request for extension by Email **before the deadline** providing reasons and any relevant documentary evidence of exceptional circumstances

Querying grades/feedback: Email me so that I can refer to your submission and respond in due course. It's a large class, thus I won't necessarily be able to discuss specifics in the classroom.

A1.2 Marking Scheme

The assignment is worth 14% of the module and each part is worth 4%; the writeup is worth a final

Half the marks will go for completeness

- the required data has been visually encoded in some way while meeting the requirements of the spec)

Half the marks will go for quality.

- charts must be sufficiently clear (as clear as can be within the constraints of the assignment), well-labelled, and readable by a third party without over-reliance on textual explanation beyond what is in the caption
- for this assignment, as you are being asked to compare different encoding channels, you won't be marked down for any lack of effectiveness of the choice of encodings in a specific chart