Assignment 2 Part B

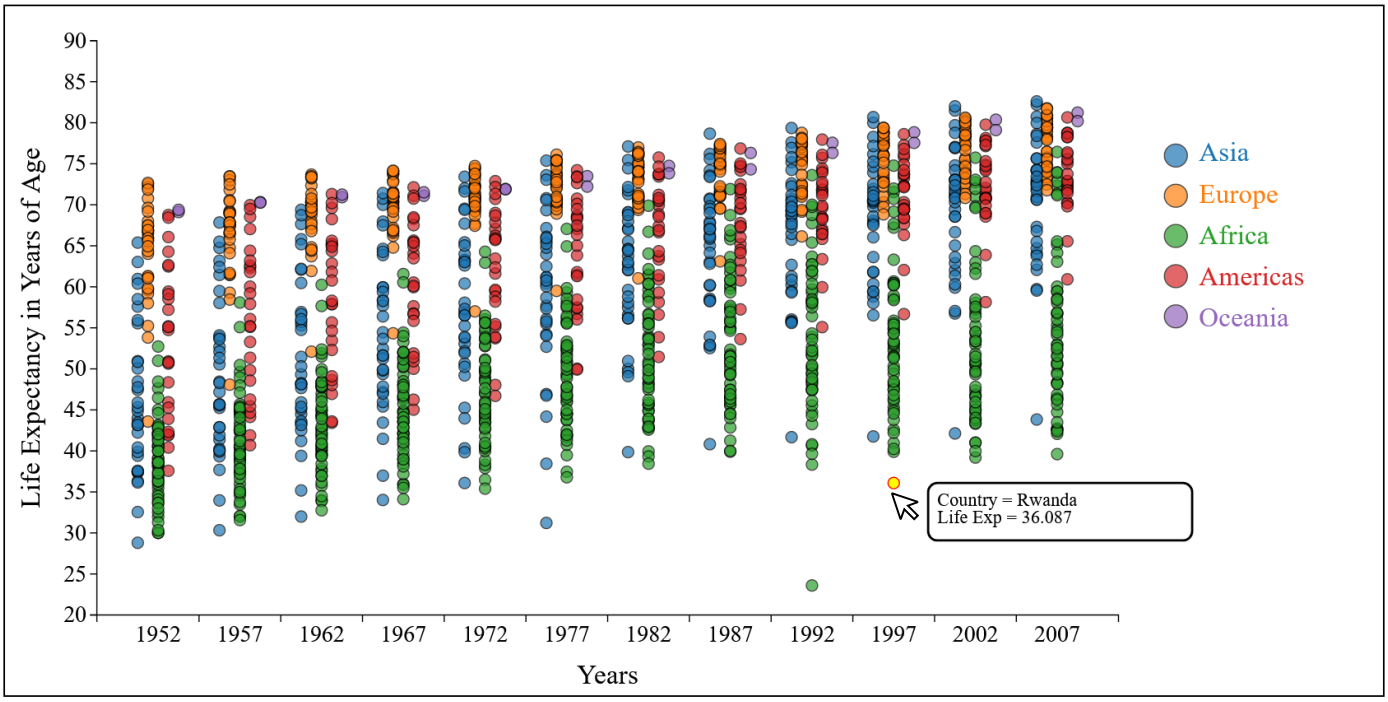
Declaration: I have read, and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcdie.libguides.com/plagiarism/ready-steady-write.

Overall Description: All visualizations were built using ***D3.js*** (version 7) along with ***HTML***, ***CSS***, and ***JS*** in the Firefox browser. A library for in-browser D3 v7 (https://cdnjs.cloudflare.com/ajax/libs/d3-legend/2.25.6/d3-legend.min.js) was used for simplify legend creation as D3 does not natively come with legend functions. The images were obtained by taking screenshots. Provided dataset was not manually modified in any manner. All processing done through code. The `d3.csv()` function was used to load *gapminder.csv*. Code files (html, js, css) for each question is located within a folder called 'question1', 'question2' or 'question3' within the root 'code' folder which also contains a copy of the dataset. Some of the visualizations have interactive tooltips that depict same information as axes but in more detail upon hover while also highlighting hovered point and displaying corresponding country name. The tooltip is designed in such a way that most recently hovered points are sent to back thereby exposing new points for hover. This allows one to explore overlapping points effectively). **All charts are implemented as SVG meaning that they scale up/down in size well.**

QUESTION 1

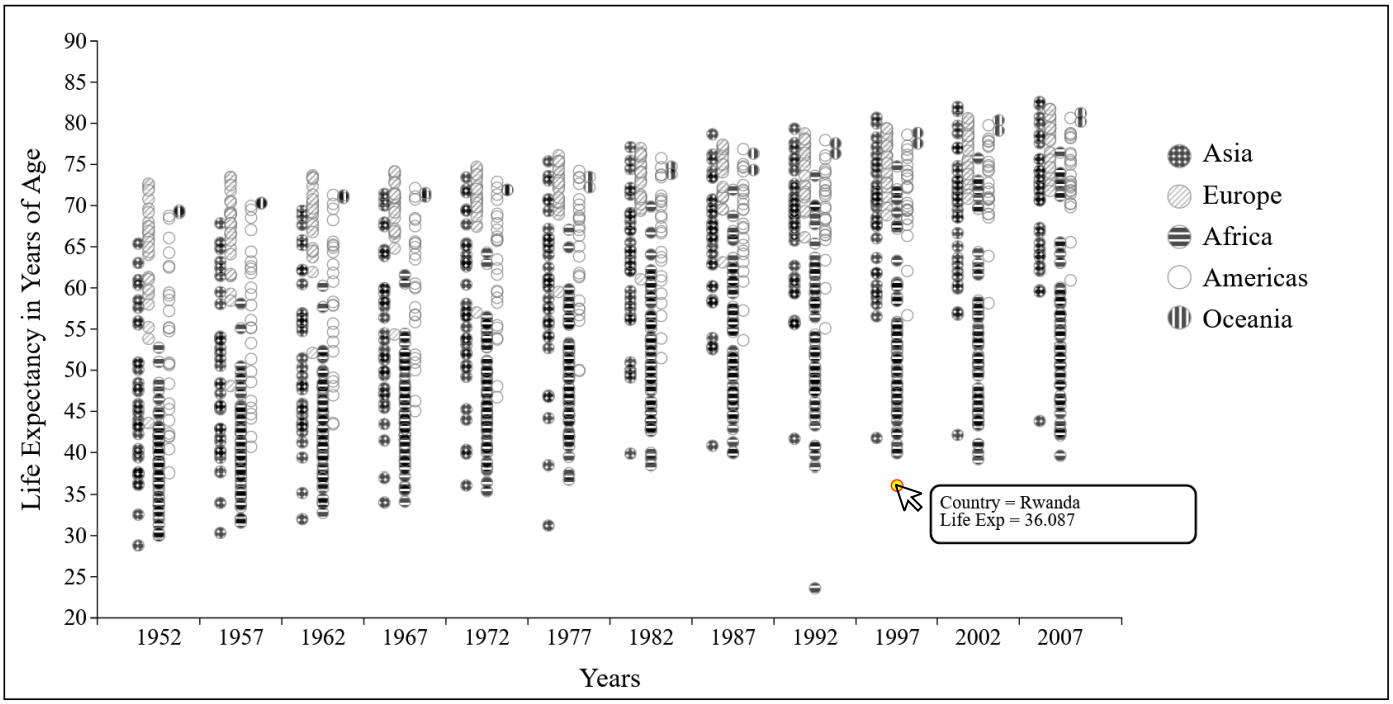
VARIANT 1

This variant used colour to encode continents as colour as an encoding channel is great for ordinal data with limited range (only 5 continents). Since all countries have data for every year, points corresponding to each continent were staggered a bit in the x axis within each year's axis to space the points out enough for the colours to be visible without heavy overlapping. Of the 3 variants experimented with as part of this question, this one using 5 distinct colours to encode the 5 continents with the added tweaks to position of each country's point as per continent, is most effective.



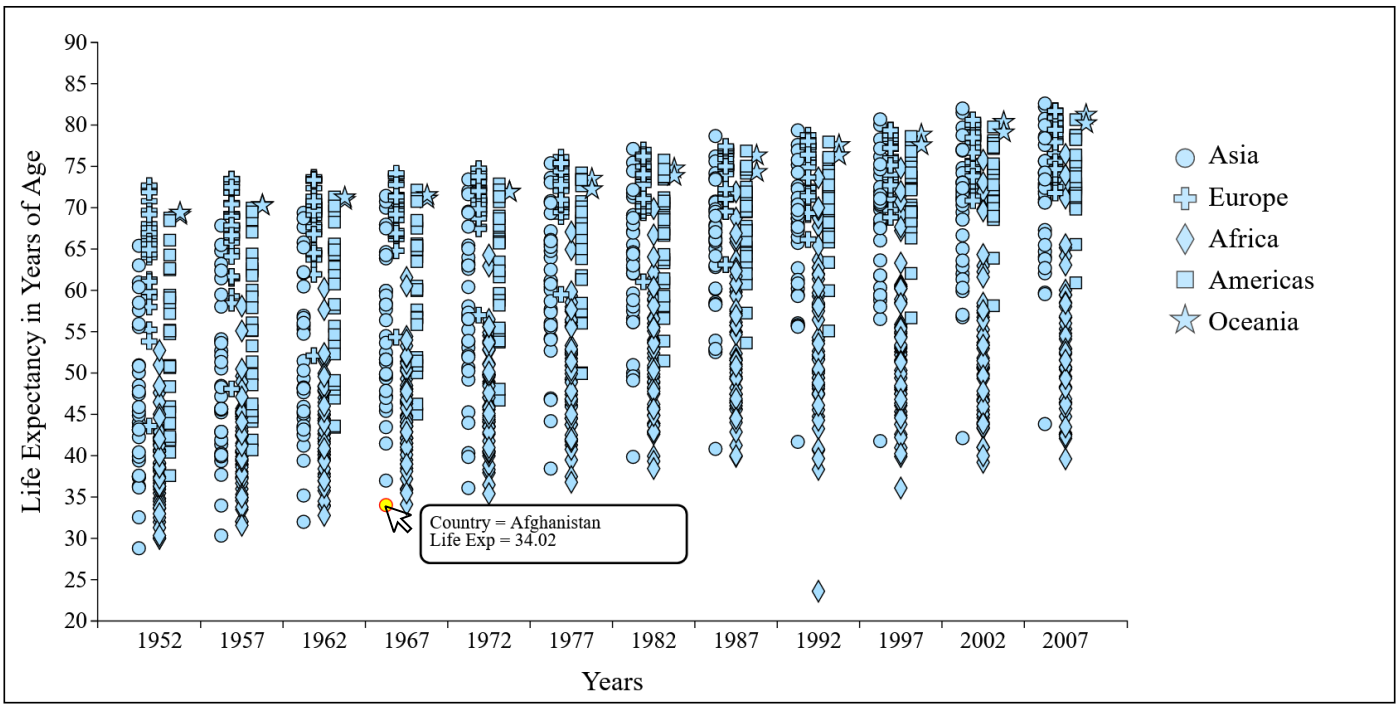
VARIANT 2

This variant encodes continent using the pattern channel. This is not very effective. It's harder to tell the categories apart. The disadvantages of this channel (complex to design, more cognitive effort, more area required) compared to colour w.r.t categorical data encoding is clear here in that this channel requires greater area than what's available here to be effective. Since there are many countries, to keep the points separate, even after position staggering, it's still necessary to keep the circles small. This is what makes it hard to discern the patterns. Much time was spent looking for and tweaking SVG patterns (obtained from: http://iros.github.io/patternfills/sample_d3.html) to create ones that were as different from each other and as visible in the limited area available as possible. Lighter ones were placed interleaved with darker ones and 1 pattern (Americas) was chosen to be an absence of any pattern to further improve readability. SVG patterns were defined within <defs id='mypattern'></defs> tags in the html file and used by setting 'fill' style of circles to 'url(#mypattern)'. Here, stroke (circle outline) was made to be grey instead of black so as not to clash with black patterns within the circles.



VARIANT 3

Of all the variants, this one using shape to encode continents was least legible among all variants. This is because of the high density of points. While the shapes are distinguishable, it's much more cognitively demanding than colour. A light fill colour with black outline and no opacity was chosen to make the shapes as distinguishable as possible. D3.js's symbol generator was used to draw the various shapes as <path></path> elements.

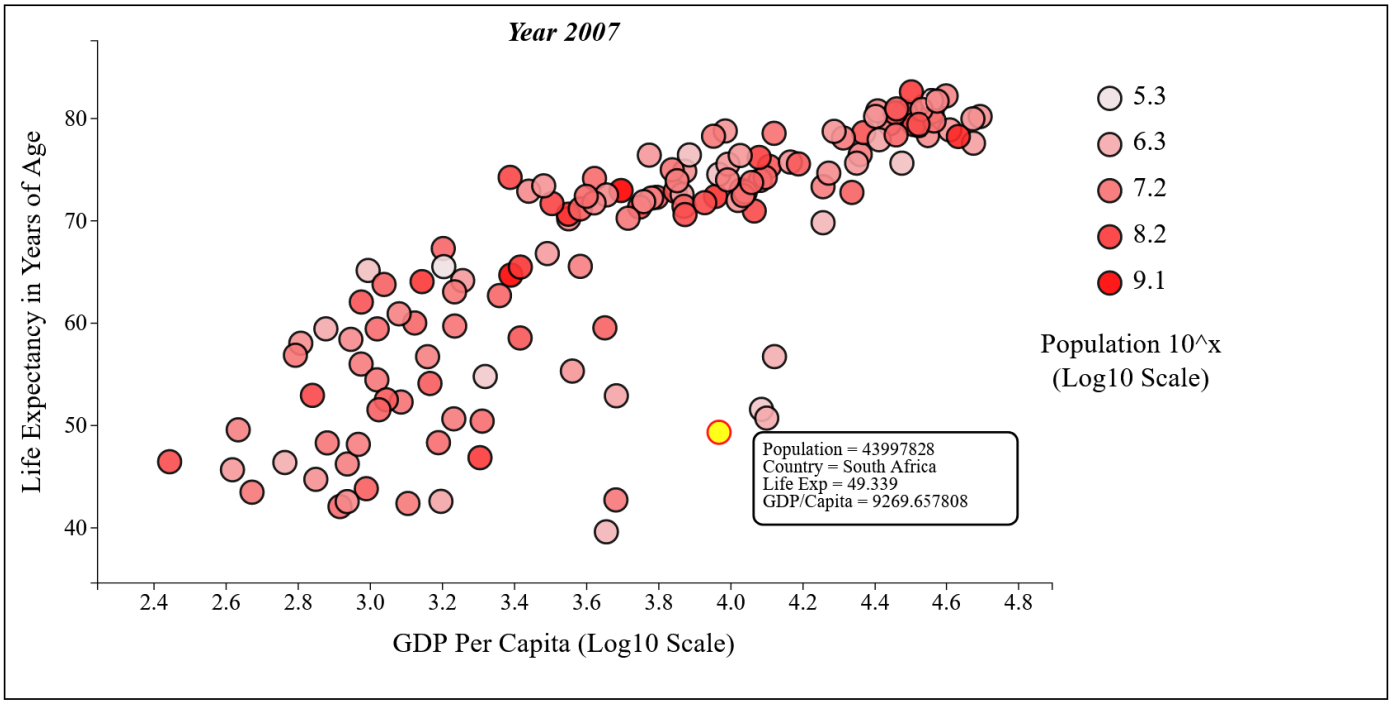


QUESTION 2

VARIANT 1

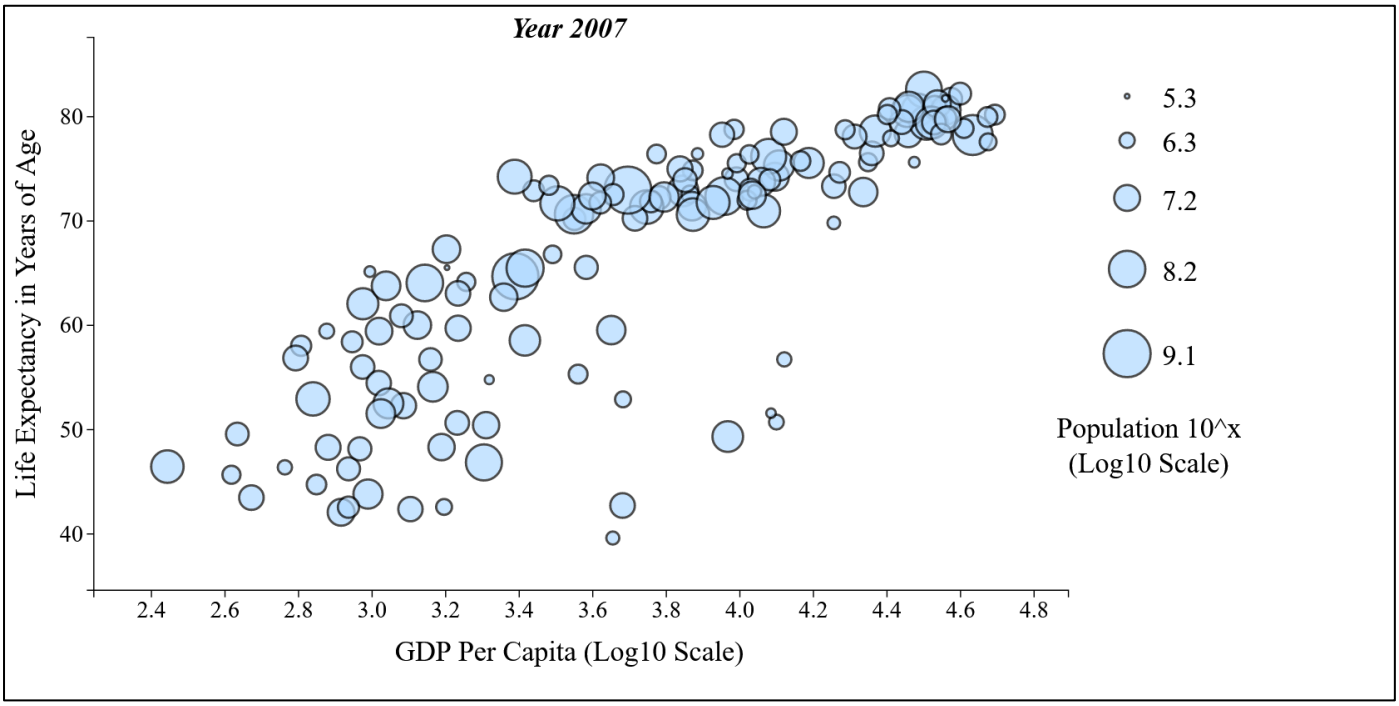
In all 3 variants GDP Per Capita (x axis) and Population values are represented in log base 10 scale as otherwise, the range is too high to visually differentiate/display on the axis comfortably. Only data for the year 2007 is displayed. Here too all 3 variants come with an interactive a tooltip. Since points are clustered, brush select was implemented for all 3 variants such that brush selected points get zoomed in (axes values also change). This selection/zoom can be done multiple times. Clicking on an empty spot (outside any point) results in the chart zooming out to the original view with all points as shown below. For variant 1, population is encoded using brightness. While this is not as effective as using colour for categorical values as seen before, it's still somewhat selective (can detect highest and lowest population easily but intermediate shades are harder to tell apart) given the

choice made here to vary brightness instead of hue. Opacity is kept high (90%) to ensure maximum brightness while still ensuring that overlap among points are evident.



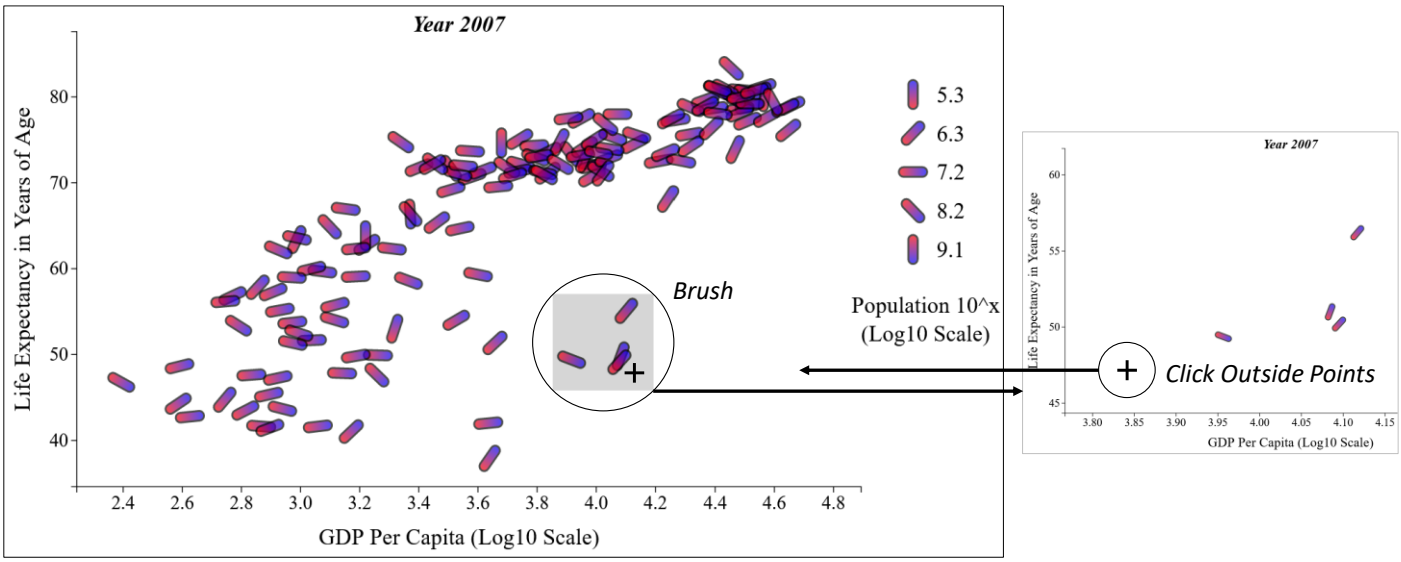
VARIANT 2

This variant encodes population using 2D size (area of circle). This is not as effective as 1D shape would be because difference in area between various intermediate sizes other than largest and lowest is hard to ascertain. The brush zoom feature made this chart much more usable and readable despite overlaps.



VARIANT 3

Here, population is encoded using orientation. Once again, it is difficult to distinguish between middle values. Also, naturally, orientation is not the best channel especially for quantitative values given its limited range and the human brain's limited ability to easily differentiate between various orientations. To make the rotation angles here more visible (0 v/s 180 degrees), the oblong markers are filled with a gradient going from blue to red such that if blue point up then that means lower population and red pointing up means high population. This along with brush zoom and tooltip made this chart the most effective of the three variants. Please find link to interaction video here: https://drive.google.com/file/d/17286kkBTI1olDCjrz5uzyHyGjkba6Ffv/view?usp=sharing.

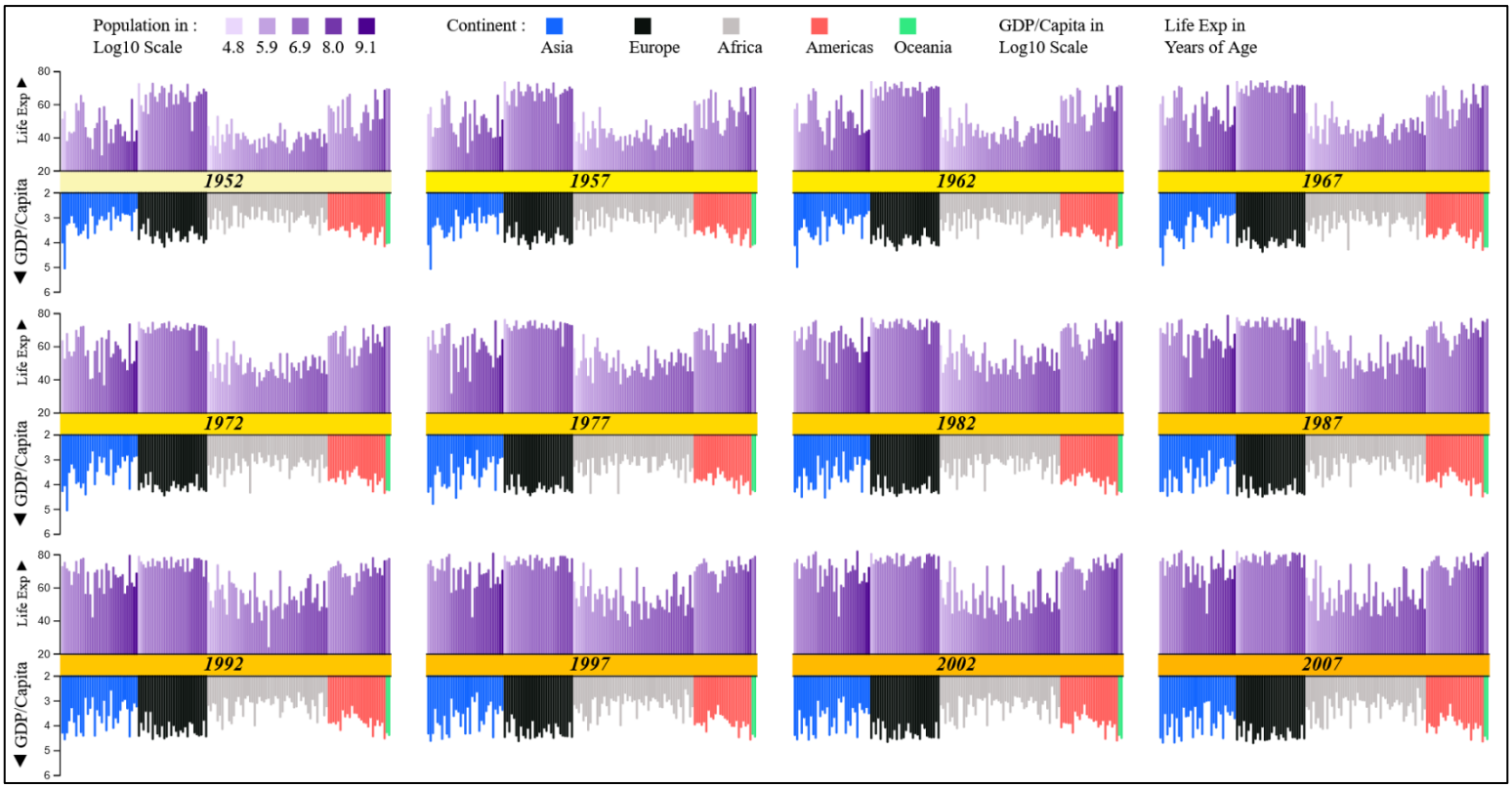


QUESTION 3

Here, **years** are encoded using position on a grid (left to right, top to bottom). This is efficient. **Life expectancy** and **GDP per capita** in Log_{10} scale are encoded using 1D size (lines arranged like a bar graph) which makes the values easily comparable. Brightness (purple) is used to encode **population** in Log_{10} scale. This is not the best option to be able to ascertain precise values but by combining this approach with sorting the bars to be arranged in increasing order of population (per continent), the difference is brightness is made more obvious and countries with high (dark purple toward right) population can be differentiated from those with low population (light purple towards left) fairly easily. Distinct hues were used to efficiently encode **continents**. Country lines corresponding to each continent are grouped together to make them more selective.

Efficiency: Overall, this chart efficiently captures change in data such that comparison is easy. The drawback with this chart is that each year's worth of data is very small (when viewed on half an A4 size paper) which makes it hard to trace exact values like for example, Life Expectancy or GDP/Capita from the y axis to the top of each country line.

Scope for improvement: This biggest drawback regarding size can, however, easily be overcome by encoding years using motion such that instead of displaying every year together at the same time, one year's data can be displayed at a time with a year slider being available that allows one to change underlying data (length of lines) from one year to another (somewhat like this: https://observablehq.com/@uvizlab/d3-tutorial-4-bar-chart-with-transition but over time). Unfortunately, given time constraints, this could not be implemented for this submission. This, along with addition of a tooltip can alleviate the issue of population being encoded using brightness resulting in low range discriminability.


Population in :
Log10 Scale
4.8 5.9 6.9 8.0 9.1
Continent :
Asia
Europe
Africa
Americas
Oceania
GDP/Capita in
Log10 Scale
Life Exp in
Years of Age
Life Exp
GDP/Capita
1952
1957
1962
1967
1972
1977
1982
1987
1992
1997
2002
2007