

# Exercises: Preston curve—ggplot2

YSC2210 - DAVis with R

Michael T. Gastner

## Preston curve

### Introduction

In a classic paper, [Preston \(1975\)](#) discussed scatter plots of life expectancy versus national income per capita (see figure 1), where each point represents one country. The term ‘Preston curve’ has since then become a synonym for curves fitted to similar data, usually with the per-capita gross domestic product (GDP), instead of national income, as x-value. [Preston \(1975\)](#) and many others have used untransformed x-values and y-values. For a different take on plotting the data, the Swedish foundation [Gapminder \(2016\)](#) uses a logarithmic scale for income (figure 2). A logarithmic scale makes sense because most economic indicators are right-skewed. I recommend to adopt Gapminder’s approach.

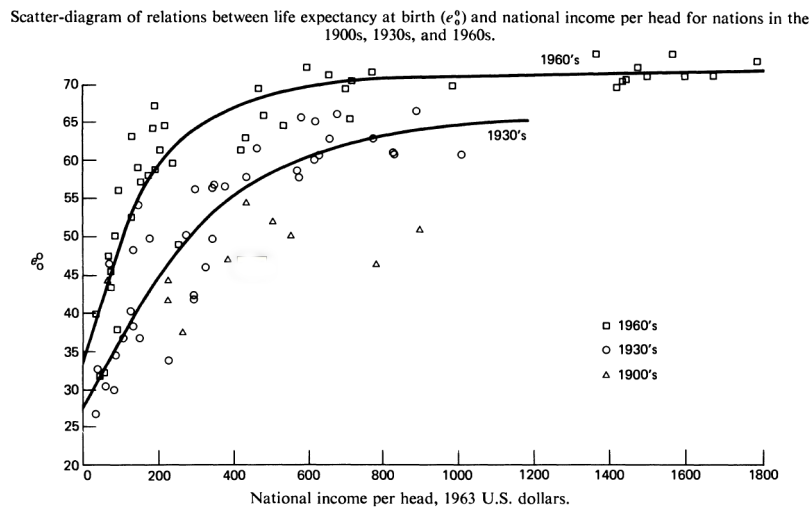
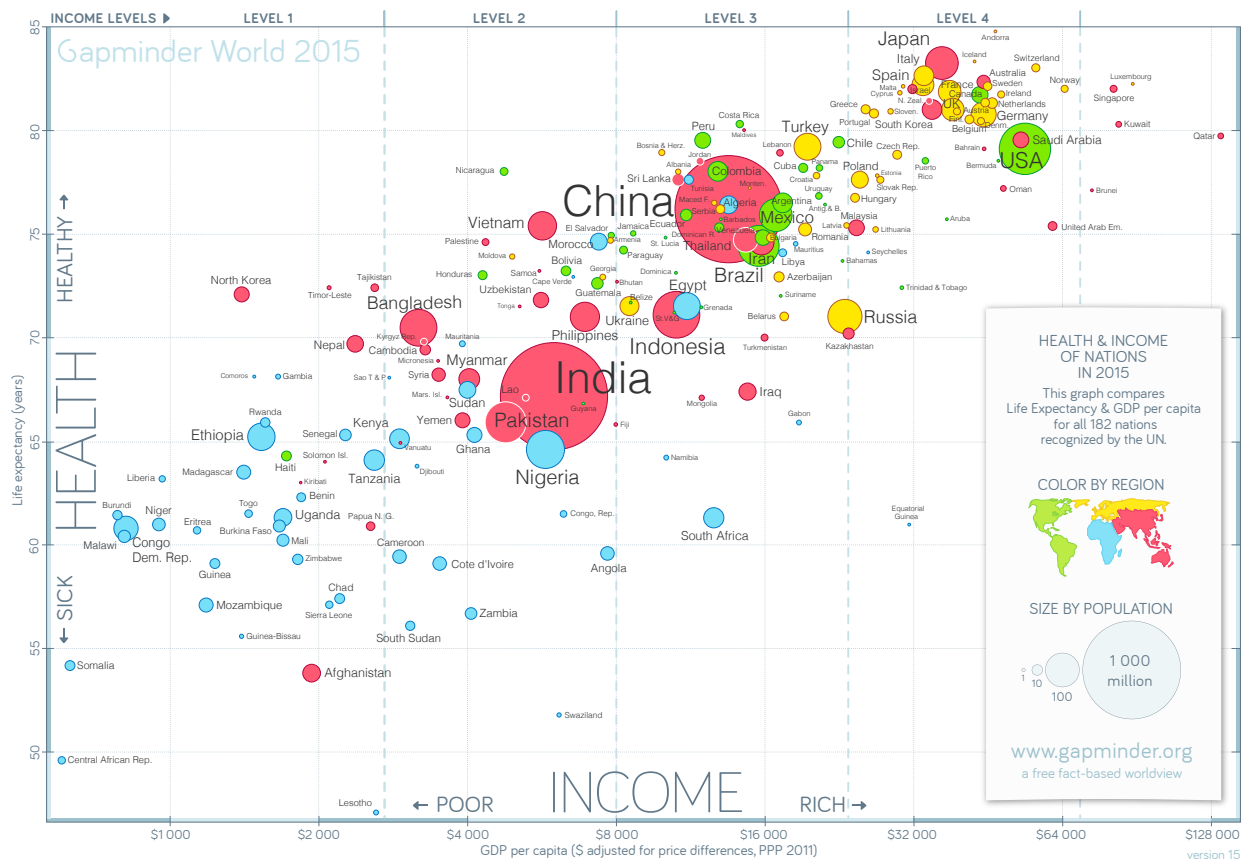


Figure 1: Diagram from [Preston \(1975\)](#).



DATA SOURCES—INCOME: World Bank's GDP per capita, PPP (2011 international \$). Income of Syria & Cuba are Gapminder estimates. X-axis uses log-scale to make a doubling income show same distance on all levels. POPULATION: Data from UN Population Division. LIFE EXPECTANCY: HME GBD-2015, as of Oct 2016. ANIMATING GRAPH: Go to [www.gapminder.org/tools](http://www.gapminder.org/tools) to see how this graph changed historically and compare 500 other indicators. LICENSE: Our charts are freely available under Creative Commons Attribution License. Please copy share, modify, integrate and even sell them, as long as you mention: "Based on a free chart from www.gapminder.org".

Figure 2: Diagram from Gapminder (2016).

## Objectives

We practise our **ggplot2** skills by making a plot that is comparable to the plot of life expectancy as a function of GDP by the Gapminder Foundation (figure 3).

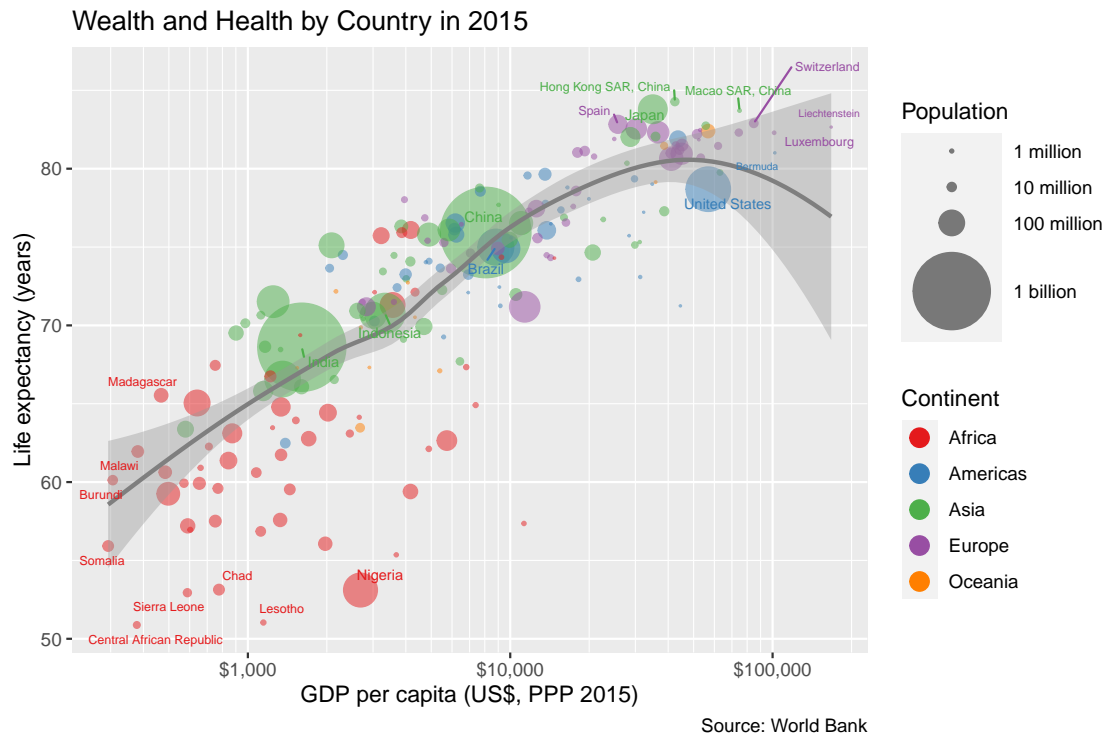


Figure 3: Data from Gapminder (2016) plotted with **ggplot2**.

## Tasks

- (1) Download [https://michaelgastner.com/DAVisR\\_data/life\\_quality.csv](https://michaelgastner.com/DAVisR_data/life_quality.csv).<sup>1</sup> This CSV file contains the columns:

- `country_name`
- `country_code`: standardised 3-letter code (ISO 3166-1 alpha-3)
- `gdp_per_capita` (US\$, PPP 2015)
- `life_expectancy` (in years)
- `pop`: population
- `continent`

These numbers are not exactly the same as those used by Gapminder (2016). Thus, please do not worry if your final plot does not look identical.

Import the CSV as a tibble.

- (2) Make a bubble chart with **ggplot2** where:
- the x-coordinate is the GDP per capita.

<sup>1</sup>These data are based on information available from the World Bank (accessed on 14 February 2022).

- GDP per capita (US\$, PPP 2015): <https://data.worldbank.org/indicator/NY.GDP.PCAP.KD>
- Life expectancy at birth, total (years): <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>
- Population: <https://data.worldbank.org/indicator/SP.POP.TOTL>

- the y-coordinate is the life expectancy.
- the colour indicates the continent.
- the size of the bubble indicates the population.

Change the axis labels and give the plot a title. Give credit to the World Bank as data source in the form of a caption. Make the bubbles semitransparent. (An improvement compared to Gapminder!)

Do not worry about the scales for the coordinates and the bubble areas yet. We will fix them shortly.

## References

- Gapminder (2016). Updated Gapminder World Poster 2015! URL: <https://www.gapminder.org/downloads/updated-gapminder-world-poster-2015/>. Accessed on 2020-11-26.
- Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, **29**(2), 231–248.