

Gapminder exploration

Rebecca Barter

August 31, 2017

1 Introduction

Gapminder is an excellent organization aimed at increasing the use and understanding of statistics on a number of global topics. They collect a variety of data from many sources and aim to produce fact-based statistics reflecting the current state of our world. The data we are exploring throughout this analysis consists of population, life expectancy and GDP information for many countries through time.

The data can be found from <https://raw.githubusercontent.com/resbaz/r-novice-gapminder-files/master/data/gapminder-FiveYearData.csv> if you would like to download it yourself.

Fortunately, the data was already very clean, so we did not conduct any major modifications to the data.

2 Visualizing the gapminder data (ggplot2)

We are interested in exploring life expectancy as a function of GDP. Figure 1 shows a scatterplot of life expectancy versus GDP.

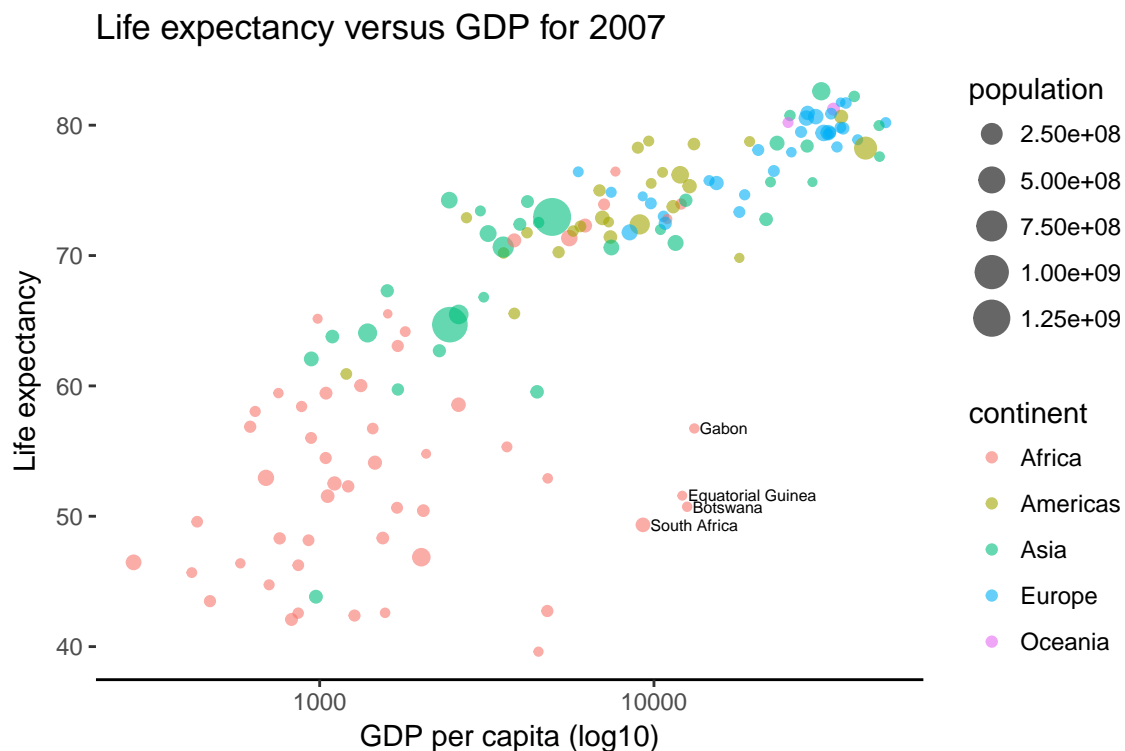


Figure 1: Life expectancy versus GDP for all countries in the year 2007

It certainly appears as though there is some kind of rapid increase in the low GDP range, which slows to a gradual increase in the high GDP range. Several African countries have surprisingly low life expectancy for their GDP.

Next, we explore change in life expectancy over time. Figure 2 shows a series of boxplots, one for each year-continent combination. Each data point corresponds to the life expectancy of a country for the given year in the given continent.

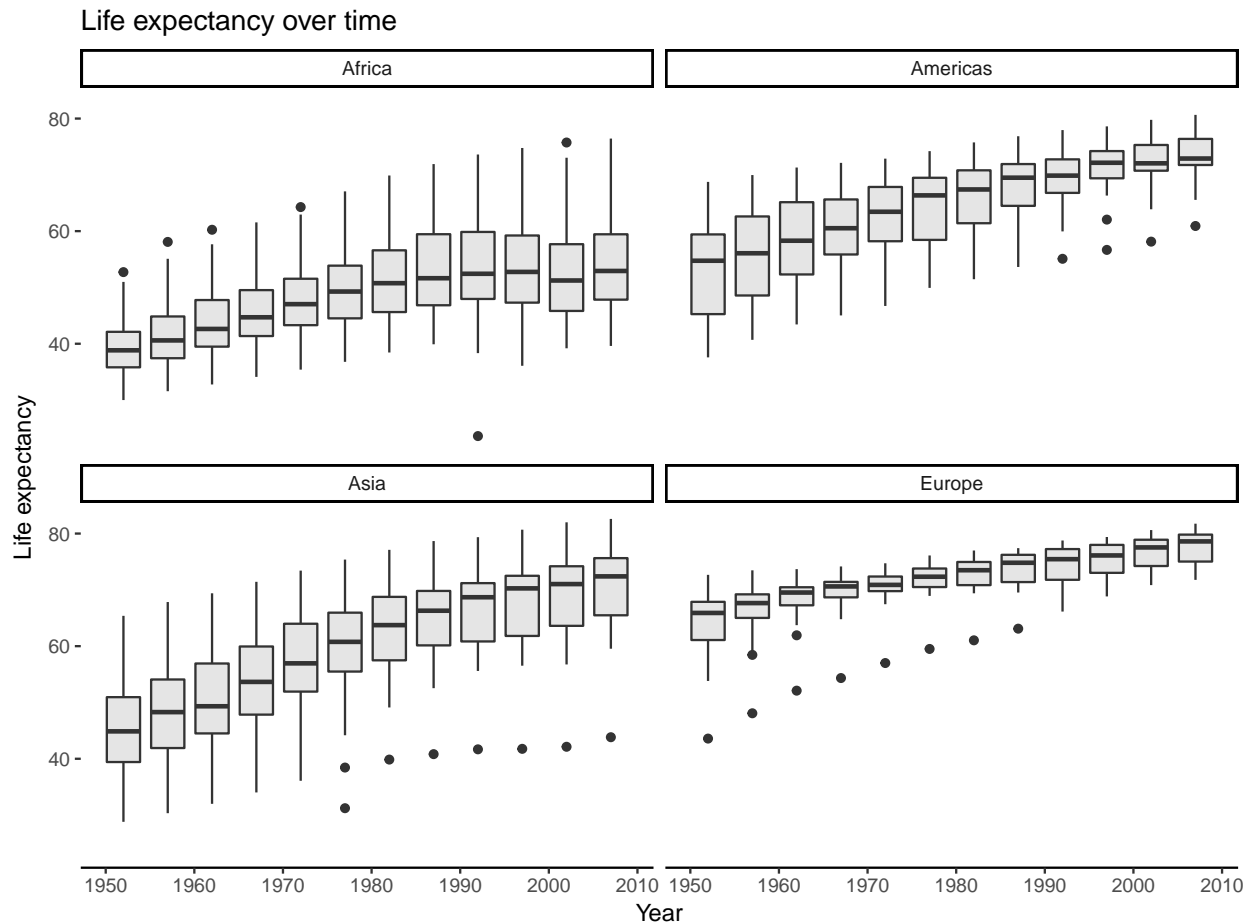


Figure 2: Life expectancy over time

We see that the life expectancy increased in Africa from 1950 up until the 1990s but has stayed fairly constant with a median of around 50 years since the 1990s. The Americas, Asia, and Europe on the other hand, have experienced continued growth.

2.1 Comparing GDP across continents (dplyr)

Table 1 compares GDP per capita across continents.

continent	countries	mean	SD
Oceania	2	29810.19	6540.99
Europe	30	25054.48	11800.34
Asia	33	12473.03	14154.94
Americas	25	11003.03	9713.21
Africa	52	3089.03	3618.16

Table 1: A table displaying the mean and standard deviation of GDP per capita in 2007 for each continent

Clearly Oceania is leading the way in terms of GDP per cap! Go Aussies!

Next, we want to ask about raw GDP (i.e. overall GDP for each country, rather than standardized by per capita). Table 2 shows the average total GDP for each continent for 2007.

continent	countries	mean	SD
Americas	25	777	2573
Asia	33	628	1344
Europe	30	493	678
Oceania	2	404	424
Africa	52	46	92

Table 2: A table displaying the mean and standard deviation of GDP (in billions) in 2007 for each continent

I didn't get a chance to write up examples about tidyr, but feel free to follow the example at <http://swcarpentry.github.io/r-novice-gapminder/14-tidyr/>.