Dealing with Data Homework

Getting Started

We're going to use the gapminder dataset here.

Load the **dplyr**, **readr**, and **stringr** packages, and read the gapminder data into R using the `read_csv()` function (n.b. `read_csv()` is *not* the same as `read.csv()`). Assign the data to an object called `gm`. Run `gm` to display it.

Problem set

Use **dplyr** and/or **stringr** to address the following questions:

1) How many unique countries are represented per continent?
2) Which European nation had the lowest GDP per capita in 1997? (*Hint:* `head(n=1)`)
3) According to the data available, what was the average life expectancy across each continent in the 1980s?
4) What 5 countries have the highest total GDP over all years combined? (*Hint:* GDP per capita is simply GDP divided by the total population size. To get GDP back, you'd mutate to calculate GDP as the product of GDP per capita times the population size. Also, `head(n=5)`)
5) What countries and years had life expectancies of *at least* 80 years? *N.b. only output the columns of interest: country, life expectancy and year (in that order).*
6) What 10 countries have the strongest correlation (in either direction) between life expectancy and per capita GDP?
7) Which combinations of continent (besides Asia) and year have the highest average population across all countries? *N.b. your output should include all results sorted by highest average population.*
8) Which three countries that start with the letter M have had the most consistent population estimates (i.e. lowest standard deviation) across the years of available data?
9) ***Bonus!*** Which observations indicate that the population of a country has *decreased* from the previous year **and** the life expectancy has *increased* from the previous year? See the vignette on window functions.