Assignment Three

HVL DS 2020

Arian Steen Anne Grete Lilleland

30 10 2020

The Report

Our assignment report will include the explanation of the code used in this report, in order to analyze and look at the data imported from 'Gapminder-Systema_Globalis'.

We will aim at answering questions regarding the data, some of the questions are written below:

- What information does the file ddf_concepts.cssv contain?
- What information does the file ddf–entities–geo–country.csv contain?
- What information does the file ddf–entities–geo–un_sdg_region.csv contain?

We will also recreate the variable 'Continent', with the new data. We will only include countries that have a iso3166_1_alpha3 code.

We will also show the graphics/plots used within our report. Our report will include the code writting within the IDE for R 'Rstudio' and work within the 'tidy data' framework as our focuspoint.

We are going to write our report in English as to streamline the workflow between the book 'R for Everyone' our own report and make the repport more accessible.

The Data

The data that we are going to use in this report is taken from the official Gapminder website. Containing local and global statistics from several hundrer sources. Including but not limited to : Geographical data, Income data, age statistics and population,density data.

The Assignment

1 What information does *'ddf_concepts.csv'* contain? :

This file contains certain information collected within the dataset, including the source url and description of the data collected. This file is used to explain the gapminder datafile we are going to use in our report, as such this is purely an explanatory text file that does not contain statistics, numbers or other variables.

The total number of observations within the file is 596 with 17 variable columns, the different description-variables includes name, catalog , short description , url of the source and type of measurements done.

2 What information does *'ddf–entities–geo–country.csv'* contain? :

This file contains certain information about different countries, the different variables include income, religion , region and wheter or not the country is landlocked (Access to a coastline within the borders of the country). This file contains 273 observations, and 21 variable columns.

3 What information does *'ddf–entities–geo–un__sdg__region.csv'* contain? :

This file contains information about different continents and countries, there are 8 regions in this file, each with their own unique color to make it easier to distinguish between them in a graphical setting.

A sample from the file would be : un_europe_and_northern_america this would be the regions of Europe and Northern America, with its own unique color to make displaying the data easier in a graphical setting.

4 Recreating the continent variable.

Recreate the continent variable with the new data. Only include countries that have aiso3166_1_alpha3code. Use data from ddf–entities–geo–country.csv and call this tibble g_c. Let g_c be your main tibble in the following, i.e. add variables to this tibble.

```
g_c <- g_c %>%
  mutate(continent = case_when(
    world_4region == "asia" & un_sdg_region %in% c("un_australia_and_new_zealand", "un_oceania_exc_aust
    world_4region ==  "asia" & !(un_sdg_region %in% c("un_australia_and_new_zealand", "un_oceania_exc_a
    world_4region == "americas" ~ "Americas",
    world_4region == "africa" ~ "Africa",
    world_4region ==  "europe" ~ "Europe")
  ) %>%
 filter(!is.na(iso3166_1_alpha3))
```