Exercises for Lecture 12

Statistical Computing with R, 2022-23

Exercise 1

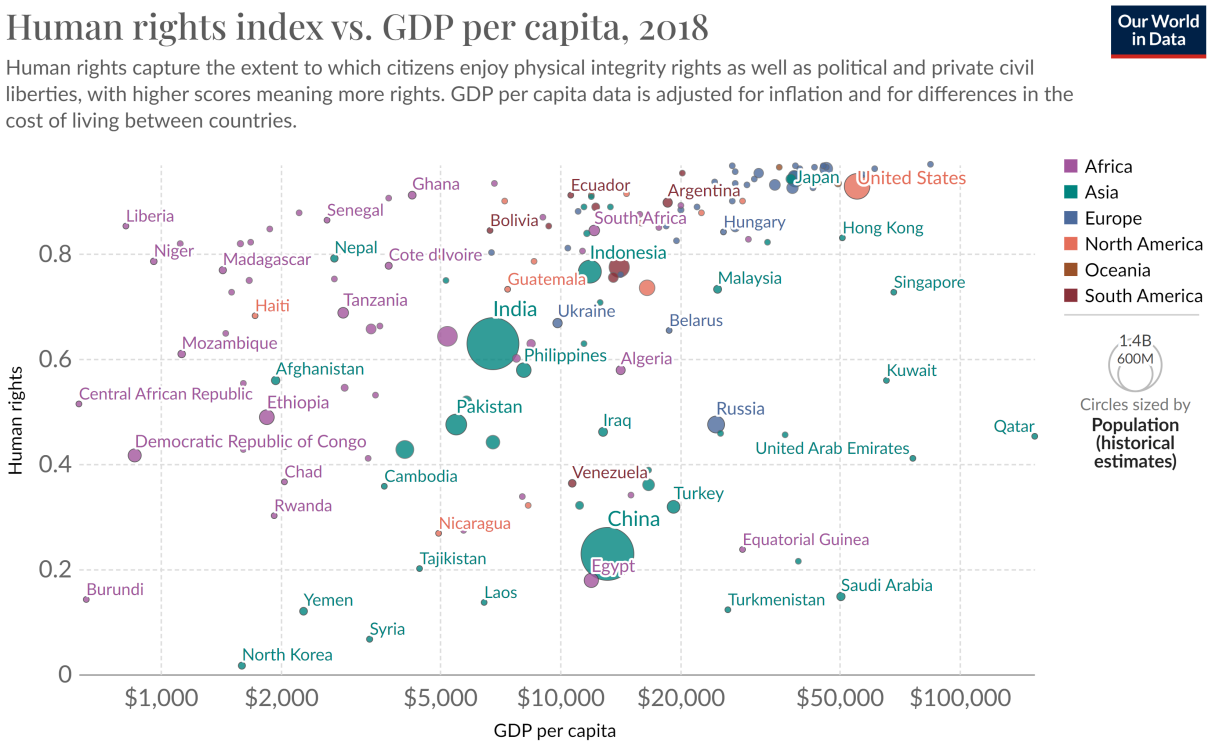
In exercise 5 of lecture 5 you created barplots using the `barplot( )` function, and pie charts using the `pie( )` function. In this exercise, we will use `ggplot2` to create barplots and pie charts.

1. Retrieve the "cleaned" data frame which you obtained at the end of point 3 of exercise 1 from lecture 5. Rename the column names so that they don't contain any invalid characters.
2. Select the 10 most recent opinion polls contained in the data frame retrieved at (1).
3. Use the `summarize( )` function to compute the average estimated share of each party.
4. Which parties have a mean share above 6%? For the rest of the exercise, consider only those parties, plus an additional category named "Others" whose value is 100 minus the sum of the shares of the aforementioned parties.
5. Use the `ggplot( )` and `geom_bar( )` functions to create a barplot that compares the shares of different political parties. If you get stuck, check out the examples that you can find in the R Graph Gallery. Make sure that the title and labels of the chart contain meaningful information.
6. Use `ggplot( )`, `geom_bar( )` and `coord_polar` to create a pie chart that compares the shares of different political parties. If you get stuck, check out the examples that you can find in the R Graph Gallery.

Exercise 2

Our World in Data (https://ourworldindata.org) has a thematic section dedicated to human rights.

1. Go to https://ourworldindata.org/human-rights, locate the section *All our charts on Human Rights* and click on "Human Rights Index vs. GDP per capita". If everything works as it should, you should see a scatter plot that looks similar to the one below:



1. International dollars: International dollars are a hypothetical currency that is used to make meaningful comparisons of monetary indicators of living standards. Figures expressed in international dollars are adjusted for inflation within countries over time, and for differences in the cost of living between countries. The goal of such adjustments is to provide a unit whose purchasing power is held fixed over time and across countries, such that one international dollar can buy the same quantity and quality of goods and services no matter where or when it is spent. Read more in our article: What are Purchasing Power Parity adjustments and why do we need them?

In this exercise, we will try to produce a scatter plot similar to this!

2. At the bottom of the chart, click on **Download** to download the `csv` file that contains the data used to generate the scatter plot.
3. Import the data in `R`.

4. Select observations that refer to the year 2018. Rename the variable containing the "civ_libs_vdem_owid" as "HRI" (Human Rights Index), the one with the GDP per capita as "GDPpc", and the one with the "Population historical estimates" as "Population".
5. Draw a scatter plot of HRI vs GDPpc. Include suitable labels (similar to those used by Our World in Data) for the x and y axis.

In order to be able to add continent information to the chart, you first need to fix the dataset you imported. Note how continent information is mentioned only once per country, specifically for 2015 and not for 2018.

6. Retrieve continent information for each country, and add it to the scatter plot by colouring the dots in the chart by continent.
7. You may notice the presence of two categories that show up in the legend, but not in the chart - one called "Antartica", and one which is an empty label (" "). Fix the input data frame by removing all rows for which either x or y is missing, and redraw the chart.
8. Add population as "bubble size" to the scatter plot. This can be done using:

```
ggplot(x = ..., y = ..., color = , size = ...) +
  geom_point()
```

Exercise 3

In exercise 3 of lecture 8 you drew a boxplot displaying mean heights by continent in 1980 using the `boxplot( )` package. Load the `heights` dataset using

```
library(brolgar)
data(heights)
heights = as.data.frame(heights)
```

1. Select observations from 1980. Use `dplyr::count( )` to check how many countries per continent are left in the data frame.

2. Draw a series of boxplots that compare the mean height of the male population across continents using functions from `ggplot2`:

a. draw both a vertical and an horizontal boxplot, both inclusive of a legend.
b. remove the legend from the two boxplots created at point (a).