Lab 2: Our World in Data

Statistics allow us to see the world in a new way. Statistics can act like a telescope, or a microscope, helping us to see things that would otherwise be beyond our reach. This lab is adapted from the work of Hans Rosling, Professor of International Health at Karolinska Institute, Sweden. The data from this lab comes from his team's website: https://ourworldindata.org/ and his findings are presented in his book "Factfulness."
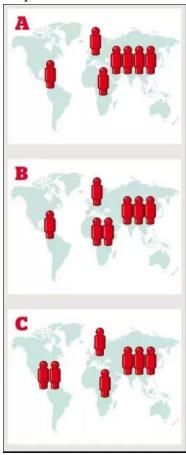
In this lab, you will load data into R from csv files, calculate summary statistics, and make plots, histograms, and boxplots with this data. For this lab, you will create a document, e.g. in Word, or a pdf document, which includes your answers to the questions below, as well as the graphs that you will create.

Warm Up

Before we start this lab, please put your answers to the following questions in your document. Put down the answers you expect are most likely to be correct, without googling for them.

1. In all low-income countries across the world today, how many girls finish primary school?
   a. 20 percent
   b. 40 percent
   c. 60 percent
2. In the last 20 years, the proportion of the world population living in extreme poverty (less than US$2 per day) has…
   a. Almost doubled
   b. Remained more or less the same
   c. Almost halved
3. Worldwide, 30-year-old men have spent 10 years in school, on average. How many years have women of the same age spent in school?
   a. 9 years
   b. 6 years
   c. 3 years
4. How many people in the world have access to electricity?
   a. 20 percent
   b. 50 percent
   c. 80 percent
5. What is the average life expectancy of the world today?
   a. 50 years

      b. 60 years

      c. 70 years

6. There are 2 billion children in the world today, aged 0 to 15 years old. How many children will there be in the year 2100, according to the United Nations?

      a. 4 billion

      b. 3 billion

      c. 2 billion

7. There are roughly 7 billion people in the world today. Which one of the three maps shows best where they live? Each figure represents 1 billion people.

8. How many of the world's 1-year-old children today have been vaccinated against some disease?

      a. 20 percent

      b. 50 percent

      c. 80 percent

Analyzing the Data

1. In all low-income countries across the world today, how many girls finish primary school?

    a. 20 percent

    b. 40 percent

    c. 60 percent

Begin by downloading the data file:

    "primary-completion-rate-of-relevant-age-group.csv"

You can load this data into R with the command:

```
primary<-read.csv(file.choose())
```

This is a very large data set, containing lots of data from countries all over the world. If we want to see the top 166 rows of this data set we just type:

```
primary
```

If we want R to tell us the overall size of the data set we type:

```
dim(primary)
```

We just want the overall average results in low-income countries, so we type:

```
primarylow<-subset(primary,Entity=="Low income")
```

Now, let's plot how the primary school completion rate for girls has evolved:

```
plot(primarylow$Year,primarylow$Female)
```

Copy this graph into your document. (On my Windows computer, I use the "Zoom" button to get a good image of the graph, then right click on the image and choose "copy image.")

Write a sentence explaining what this graph means about the answer to question 1.

2. In the last 20 years, the proportion of the world population living in extreme poverty (less than US$2 per day) has...

    a. Almost doubled

    b. Remained more or less the same

    c. Almost halved

Download the data file: "world-population-in-extreme-posverty-absolute.csv"

Load this data into R:

```
poverty<-read.csv(file.choose())
```

This file contains columns listing the worldwidetotal number of people not in extreme poverty and the total number of people living in extreme poverty. However, we want to know the proportion of people, the fraction of the total. So, first let's create a new column which will be the total number of people:

```
poverty$Total = poverty$NotInEP + poverty$InEP
```

Now, we need to create a column with the proportion living in extreme poverty:

```
poverty$prop = poverty$InEP / poverty$Total
```

Finally, we create a plot of how this has changed:

```
plot(poverty$Year,poverty$prop)
```

We can add lines connecting the data points:

```
plot(poverty$Year,poverty$prop,type="o")
```

Copy this graph into your document and write a sentence explaining what this graph means about the answer to question 2.

3. Worldwide, 30-year-old men have spent 10 years in school, on average. How many years have women of the same age spent in school?
   a. 9 years
   b. 6 years
   c. 3 years

Download the data file: "gender-ratios-for-mean-years-of-schooling.csv"

```
ratio<-read.csv(file.choose())
```

The "Ratio" column shows the gender ratio, female to male. Therefore, if this ratio is 100%, this means that women have had an equal number of years of school. If this ratio is 50% this means that woman have had half as many years of school.
Make a plot of the data from all regions.

```
plot(gender$Year,gender$Ratio)
```

Explain your findings and use them to answer the question.

4. How many people in the world have access to electricity?
   a. 20 percent
   b. 50 percent
   c. 80 percent

Download the data file: "number-of-people-with-and-without-electricity-access.csv"

```
elect<-read.csv(file.choose())
```

Create a column which contains the total population for each line:

```
elect$Total=elect$With+elect$Without:
```

Create a column which contains the proportion with electricity access:

```
elect$Prop=elect$With/elect$Total
```

Create a subset containing just overall world totals:

```
electworld<-subset(elect,Entity=="World")
```

Plot out how this proportion has changed over time:

```
plot(electworld$Year,electworld$Prop)
```

Let's dig a little deeper and explore what this distribution looked like in the past and now. Create a subset containing the proportion with electricity access was in the

nations of the world in 1990:

```
electthen=subset(elect,Year==1990)
```

Make a histogram of this distribution:

```
hist(electthen$Prop)
```

Make a boxplot of this distribution:

```
boxplot(electthen$Prop)
```

Calculate a five number summary of this distribution:

```
summary(electthen$Prop)
```

Next, create a subset for the year 2015 and make a corresponding plots and summary statistics.

Copy the various plots and statistics into your document, answer the question and explain your findings.

5. What is the average life expectancy of the world today?
   a. 50 years
   b. 60 years
   c. 70 years

Download the data file: "life-expectancy-globally-since-1770.csv"
Load this data into R.
Create a subset with just the overall world average (`Entity=="World"`).
Make a plot of how the overall world life expectancy has evolved over time, copy this into your document, explain your findings, and use them to answer the question.

6. There are 2 billion children in the world today, aged 0 to 15 years old. How many children will there be in the year 2100, according to the United Nations?
   a. 4 billion
   b. 3 billion
   c. 2 billion

Download the data file: "younger-than-15-projection-until-2100-un.csv"
Load this data into R.
Create a subset with just the overall world totals (`Entity=="World"`).
Make a plot of the "Projection" changes over the next century, copy this into your document, explain your findings, and use them to answer the question.

7. There are roughly 7 billion people in the world today. Which one of the three maps shows best where they live? Each figure represents 1 billion people.

Download the data file: "world-population-by-world-regions-post-1820.csv"
Load this data into R.
Create a subset with just the current numbers (`Year==2019`).
We can make a bar graph of this data. I named the subset "popnow" so:

```
barplot(popnow$Population)
```

Copy this graph into your document, explain your findings, and use them to answer the question.

8. How many of the world's 1-year-old children today have been vaccinated against some disease?
    a. 20 percent
    b. 50 percent
    c. 80 percent

Download the data file: "share-of-vaccinated-one-year-olds-globally.csv"
Load this data into R.
Make plots that allow us to answer this question, and explain your findings.

Conclusions

Write a paragraph explaining your overall conclusions and findings. What does this data tell us about the overall state of world development? Discuss anything you find interesting or surprising.