Lab 1

Xuyi Ren

2025-01-31

Question 1:

Create a new R Markdown file and copy the entirety of this question over to the new file (we will do this for all questions in this lab). Then, proceed with the instructions below.

Between the stars below, do the following:

1. Use two # to create a header that says About Me

About Me

2. Type your first name in bold and your last name in italics **Xuyi** *Ren*

3. Create a bullet point list of the people sitting on either side of you

- Yujin Soga
- Riku
- Sam

4. Create a numbered list of your 3 least favorite animals

   1. snake
   2. fish
   3. lizard

5. Once you have done this, Knit to PDF by clicking the bar of yarn above and verify that everything looks like it should.

Question 2:

Again, copy the entirety of this question into the R Markdown file you created for Question 1.

Let's practice creating vectors and subsetting with a short exercise.

1. First, create an R code chunk between the rows of stars below (Ctrl+Alt+I is quick way to do this)

```r
# R code chunk
```

2. Next, create a vector called x that has all of the numbers from 11 to 20

```r
x <- c(11:20)
```

3. Use square brackets and subsetting to select the first five numbers of this vector.

```r
x[1:5]
```

```
## [1] 11 12 13 14 15
```

Question 3:

For this question, we will be using the HappyPlanet data that we have just looked at:

1. Part A Copy the code above to read the Happy Planet data into your own R Markdown file, saving the dataset to a variable called planet

```
planet <- read.csv("https://collinn.github.io/data/HappyPlanet.csv")
```

2. Part B Looking at the Happy Planet data, explain in one or two sentences what constitutes an observation in this dataset (what is the data being recorded from)

In this dataset an observation represents a specific country or region, with data recorded on various attributes like life expectancy, happiness score, ecological footprint, HPI, GDP per capita, HDI, and population that Each observation provides a insigit of these metrics for a given country or region

3. Part C Using $ to extract columns from the dataset, find the mean life expectancy of all countries in the dataset? (Hint: what functions have we seen already in this lab?)

```
mean(planet$LifeExpectancy)
```

```
## [1] 67.83846
```

4. Part D Are there any variables in this dataset that are stored as a numeric that would be better suited as a categorical variable? Explain your answer

Yes, they are: 1. Region is categorical because it represents geographic or political groupings rather than continuous numeric values. 2. HPI Rank because it represents a ranking system rather than continus numeric measurements, some method like mean does not apply for this kind of variable so it is categorical.