lab_1_sol

2026-01-22

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
2. Type your first name in bold and your last name in italics
3. Create a bullet point list of the people sitting on either side of you
4. Create a numbered list of your 3 least favorite animals

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

About Me

Vinny *Paris*

- Jokers to the left of me
- Jokers to the right of me

1) Giant Red Headed Texan Centipede
2) Zebra Fish
3) Crappie

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)
2. Next, create a vector called `x` that has all of the numbers from 11 to 20
3. Use square brackets and subsetting to select the first five numbers of this vector.

Note in R, like most programming languages, there are many ways to accomplish any task.

```r
x <- 11:20

x <- x[1:5]
```

Question 3:

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

- **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`

```r
# Use read.csv to pull Happy Planet data
planet <- read.csv("https://collinn.github.io/data/HappyPlanet.csv")
```

- **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)

Each country individually is an observation

- **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?)

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

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

- **Part D** Using [,] to extract columns, what is the median GDP per Capita?

Hint 1: Use the help pages if NA appears (specifically look for something like an na.rm parameter)

Hint 2: To find which column number you need you can run the code colnames(planet) which will list the column names in order

```r
  median(planet[,9], na.rm = TRUE)
```

```
## [1] 6632
```

- **Part E** Are there any variables in this dataset that are stored as a numeric that would be better suited as a factor variable? Explain your answer

Region is numeric and treated as numeric but the different regions almost certainly prefer to geographic location and without any meaningful numeric relationship

Question 4: I said statistics is the study of variability. Please explain that in your own words.

Answers will vary... basically trying to find what is causing the changes we are seeing

Question 5: What is a census? Why are censuses generally considered good? Give two reasons why we do not always want to conduct a census.

Census is whent he entire population has been talked to/measured. They are good because they don't have uncertainty and instead give you the exact answer. They are not always done because it could be too expensive or outright impossible (eg taking a census of beetles in the world)

Question 6: (Healthcare Opinions) In 2009, the PEW research group wanted to learn more about public opinion on the idea of the public option for health coverage. One thing that they wanted to know was the percentage of adult U.S. residents who favored a public option for health coverage in October 2009. In a poll of 1500 randomly selected Adult residents in the United States, they found that 55% of adult residents favored a government health insurance plan to compete with private plans. Source

- What is the point of the study? Why did they do this poll?

They wanted to get an idea of how puplar nationalized healthcare is in the US for adults

- Describe the population in this study:

All adult residents in the US

- Describe the sample in this study:

1500 randomly selected adults

- Describe an observation in this study:

A single adult

- What is the variable of interest in this study? Is it categorical or quantitative?

Views of nationalized healthcare, categorical because they are or are not in favor (which aren't numbers)

- Do you think this data is useful for learning about healthcare opinions in 2026?

Not really, feels like the country has undergone some changes

Question 7: (National household size) The American Community Survey (ACS) conducts yearly surveys. One thing that is of interest is the average household size. In April 2022, the ACS had surveyed 1,980,550 U.S. households and found the average household size to be 2.50. Source

- Describe briefly why the American Community Survey is ran. Hint: you may google the survey to learn more about it and it's goals.

Gives a snapshot of american demographics when the census isn't happening

- Describe the population in this study:

US households

- Describe the sample in this study:

1,9980,550 US households that were surveyed

- Describe an observation in this study:

A single us household

- What is the variable of interest in this study? Is it categorical or quantitative?

Number of people living in the household, it's quantitative

Question 8: (Consulting Problem) Making batteries last longer is an economically profitable area and many experiments occur with them. At Iowa State 64 laptop-style batteries were produced using the same standardized process. The cells were then either "slowly discharged" where the battery is slowly drained of power or "quickly discharged" where the battery is rapidly drained of power. All batteries would then be recharged and the process repeated. This continued for each battery until it fell below an acceptable operating condition (ie wouldn't hold a strong enough charge). The total number of charges/discharges for each battery was recorded.

- Describe an observation in this study:

A single batter

- Describe the sample in this study:

The 64 batteries we were testing

- Describe the population in this study:

All batteries made with this process

- What question do you think the researchers were trying to answer?

Lifetime of batteries given different stressors (answers will vary)

- What are the two variables that are being collected?

Whether it was slowly drained or fastly drained and the number of cycles the battery survived