Joshua Rowe
STAT 331
10/17/22

Lab 3 Reflection

To be honest, I'm not sure why I decided to use the name "survey_data" for the dataframe created with the hiphop dataset. In order to maintain consistency in this second attempt at Lab 3, I used the name "hiphop" for the same dataframe. I appreciate the emphasis on understandable name conventions in this class; it's definitely something I could use some work on!

I have addressed your comments for question 1 – in future assignments, I will try to make sure to discuss sampling methods and data coding methods when I am summarizing a dataset.

I also went through my notes to find the case_when() function, which is exactly what I was hoping the if_else() function would be when I first used it! I'm glad I re-discovered case_when() for this resubmission because it seems much more versatile than if_else(), which only allows you to check one condition.

I used the term "ethnic_binary" as the name for my variable which delineated "white" vs "non-white" participants because the term "binary," in its original sense, simply describes something made up of two components or with two unique states (for example a binary star system which would consist of two stars). I can see why this could be confusing from a data science perspective because "binary" can refer to the language of 0's and 1's, but my naming of the ethnic_binary variable is true to its data because it has two distinct states: "white" or "non-white."

My graphs in P8 did not appear to change much for the resubmission, but their data did. I accidentally counted each participant many times in my first attempt because I did not use the distinct() function to ensure that each participant only occupied one row of data. I have remedied this issue in my resubmission.

I also made use of the slice_max() function as opposed to the filter() function!

Finally, I want to address the grade of N that I received on two questions for writing a function to format numbers in in-line R code. Although we have not discussed this in class, we *have* covered how to write such a function; in the readings assigned on day 1 of the course, in the "Reproducibility & Deliverables" coursework, within the "Intro to R Markdown" reading, there is a section titled "27.4.6 Inline Code" which explains how to use R code in your Quarto text. It also suggests creating a function that will format numbers you output from inline code in order to expedite the formatting process. I have attached a screenshot of this section below.

27.4.6 Inline code

There is one other way to embed R code into an R Markdown document: directly into the text, with: `` `r ` ``. This can be very useful if you mention properties of your data in the text. For example, in the example document I used at the start of the chapter I had:

> We have data about `` `r nrow(diamonds)` `` diamonds. Only `` `r nrow(diamonds) - nrow(smaller)` `` are larger than 2.5 carats. The distribution of the remainder is shown below:

When the report is knit, the results of these computations are inserted into the text:

> We have data about 53940 diamonds. Only 126 are larger than 2.5 carats. The distribution of the remainder is shown below:

When inserting numbers into text, `format()` is your friend. It allows you to set the number of `digits` so you don't print to a ridiculous degree of accuracy, and a `big.mark` to make numbers easier to read. I'll often combine these into a helper function:

```
comma <- function(x) format(x, digits = 2, big.mark = ",")
comma(3452345)
#> [1] "3,452,345"
comma(.12358124331)
#> [1] "0.12"
```

It took me many hours to diligently work through this reading material over several days, but the result was that I had a comprehensive set of notes on each topic it covered and had enough memory of it that I knew to go back and reference it when I wanted to add my in-line code to lab 3. I did not cite this information because it was not from an outside source, it was from the assigned coursework; and any other information I added to my own function (namely the "scientific = FALSE" argument) I discovered how to use from the R Help console ("?format"). There was no plagiarism in my lab 3 submission, and I feel obligated to say that it was a little frustrating to have that be suggested when my work was actually the result of thorough notetaking. I understand why you felt the creation of my own function must have come from an outside source, since we never covered how to create one in our lectures. However, since there was no actual plagiarism in my document, I hope that we can work together to remedy the issue of my "grade" for the assignment.