ANA 515 Assignment 1 R Markdown

Follow these steps to create the html document using R markdown. If you have not reviewed the Markdown tutorial linked in week 1's content, please do so before starting this assignment.

1. Create a new R markdown file

   YAML:

   a. Include your full name as the author
   b. Include "ANA 515 Assignment 1" in the title
   c. For the output, use a theme (note the indented lines – you must have that):

```
output:
  html_document:
    theme:
      bootswatch: *___*
```

   **Google 'bootswatch themes' and pick one that you like. Put the name of the theme in place of the *___* above. (You may need to install the bslib package)

2. For every code chunk, indicate that you don't want any of the code to show in your output document, but you do want the inline code and graphs.

   The command goes in the brackets at the top of your code chunk, like:

```
```{r  command? = ?}
Your code here
```
```

   Not sure how? Go to https://r4ds.had.co.nz/r-markdown.html

   and check out the section: 27.4.2 Chunk options

   Be sure to note the difference between 'echo=…' and 'include=…'

3. Install the following packages (put the install command in a code chunk)
   a. tidyverse
   b. knitr
   c. bslib

   *If you already have these installed, write the code to install them with a comment symbol in front of it

   And, include the command that loads the packages.

4. Code chunk to get the dataset from GitHub/fivethirtyeight

   Here is the url to the data:
   https://raw.githubusercontent.com/fivethirtyeight/data/master/fifa/fifa_countries_audience.csv

You will use that url to get the data from GitHub to your R Studio working directory

Not sure how? Go to https://beanumber.github.io/sds192/lab-import.html and check out the section: Data from a CSV

5. Code chunk to create a subset of countries

   Create a new dataset with the name of 'smallpops' by using this filter:

   population_share <= 0.7

   Not sure how? Go to https://datasharkie.com/how-to-filter-by-value-in-r/ and check out the section: Part 4 Filter by single value in R

6. Get a summary of the new filtered dataset to better understand the data

   Not sure how? Go to https://www.statology.org/summary-function-in-r/

7. Write inline code that says:

   We have data about [insert the inline code for the number of rows in the original dataset] countries. Only [insert the inline code for the number of rows in the filtered set] have a population share of less than 0.7. The distribution of this is shown below:

   Not sure how? Go to https://rmarkdown.rstudio.com/lesson-4.html

8. Code chunk to graph the distribution of the population_share variable from the filtered dataset (you will need to edit the name of the dataset to match what you called it):

   Use the ggplot command (I'm giving you the code you need. Feel free to enhance the plot):

```
# Population Share
```{r filtered-dist, echo = FALSE}
ggplot(filtered, aes(population_share)) +
geom_histogram(color="white")
```
```

   Add the code to suppress the 'stat_bin' message that appears in the output document. See: https://rmarkdown.rstudio.com/lesson-3.html

9. Code chunk to graph the distribution of the filtered data by gdp_weighted_share

   Use the ggplot function. You can use whatever type of graph you think makes sense for this variable. Don't forget to put a header above your graph (like in the example in #8, # Population Share is a header).

   Not sure how? Check this out: https://r4stats.com/examples/graphics-ggplot2/
   *Suppress the 'stat_bin' message (like you did in #8)

Add any styling/formatting you would like.

Knit your document.

Upload your code and output to the assignment submission box (Not GitHub). BlackBoard does not like the html format, so you will have to Zip it first, then upload the Zipped file.

Here is a peek at what my output document looks like. If you can't get yours to look like this, email me, although your theme may have different formatting.

ANA 515 Assign 1 SP 24

Anna

2023-12-31

We have data about 191 countries. Only 30 have a population share of less than 0.7. The distribution of this is shown below: