Workshop 3 Instructions: Rmd reports

Introduction

Your task will be to create a short R Markdown-based report looking at the distribution of ANY **CATEGORICAL variable** in the India TB pathways dataset. For example, you could look at the distribution of `tb_status` or `treatment_type`.

The report must contain:

a. One plot of your data created with {ggplot2}/{esquisse}

b. A summary table of your data created with {flextable} or another table package (e.g. {reactable}).

Steps

1) After opening up the downloaded RStudio Project, knit the `example_report_exercise.Rmd` file to see an example of what a good report looks like.

2) Now **create a new Rmd** file (not R script!) in the *"rmd"* folder and give it an appropriate name (e.g., tb_report_YOUR_NAME.Rmd).

3) **Copy the YAML** header below into your new Rmd file.

```
---
title: "India TB Pathways"
output: prettydoc::html_pretty
author: "YOUR NAME HERE"
---
```

4) Make a new code chunk for package loading and **load your needed packages**. You will need to load `tidyverse`, `here`, `janitor`, `esquisse`, and `flextable`.

   - Tip: You can use Ctrl+Alt+I to insert a new code chunk in your rmd.
   - Tip: You should use `echo = FALSE` and `message = FALSE` in this code chunk to prevent the package loading messages from being printed in your report.

5) Under an appropriate *section heading,* **add the following text description** of the India TB dataset:

The India TB Pathways dataset analyzes pre-diagnostic costs and healthcare-seeking behavior for 880 new adult tuberculosis patients in Vellore, Tamil Nadu.

6) Make a **new code chunk** in which you import the India TB dataset from your data folder. Call it `india_tb`. You should use `read_csv()` and `here()` to load the data. (Consult the course textbook online if you need a refresher)

   - Tip: After importing, you can use the `clean_names()` function from the `janitor` package to clean the variable names. This removes spaces and special characters from the variable names, making them easier to work with. See the example report for an example of how to do this.

7) Once your dataset is loaded, inspect the dataframe and **choose one categorical variable for your analysis**.

- Tip: Inspect your dataframe using functions taught in our previous lessons such as `head()`, `summary()`, and `glimpse()` . If you use `View()` in your script, remember to remove it before knitting your report, as it will interrupt the knitting process.

8) Under an appropriate section heading, **create an informative graph of your chosen variable** using {esquisse} and {ggplot2}. To launch esquisse on the `india_tb` dataframe, you would use the code chunk below:

```
esquisse::esquisser(india_tb)
```

- After you have finished using esquisse, MAKE SURE you remove the `esquisse::esquisser()` function from your Rmd file. Otherwise, every time you knit your Rmd file, esquisse will open up and get stuck!

9) **Create a table** using {janitor} and {flextable} to summarize the distribution of your chosen variable.
    - Below is an example of how to create a frequency table using `tabyl()` and then format it with `flextable()`, using the `iris` dataframe:

```
freq_df <- iris %>% tabyl(Species)
flextable(freq_df)
```

```
## Warning: fonts used in `flextable` are ignored because the `pdflatex` engine is
## used and not `xelatex` or `lualatex`. You can avoid this warning by using the
## `set_flextable_defaults(fonts_ignore=TRUE)` command or use a compatible engine
## by defining `latex_engine: xelatex` in the YAML header of the R Markdown
## document.
```

Species	n	percent
setosa	50	0.3333333
versicolor	50	0.3333333
virginica	50	0.3333333

10) Add a short text description explaining what your plot and table show.

11) The final and most important part - **knit your report**!
    - Tip: Before you knit, make sure you delete or comment out any instances of `esquisser()` or `View()`. This is because functions that open new windows will interrupt the knitting process. You may have to try several times before successfully knitting your report. Be patient and ask instructors for help if you get stuck.

OPTIONAL CHALLENGE:

Use inline code to highlight specific numbers from the data. You can refer to the *"example_report_exercise.Rmd"* for hints on how to do this.

Submitting Workshop 3 Exercise

To submit your work, please upload your own **Rmd file** and **HTML file** to their respective submission pages.

Note: Before uploading and submitting your files, make sure the Rmd knits successfully and you are getting the desired HTML output.