

DACSS 601: Homework Two: Markdown

audra white

2/19/2021

Homework Two: Markdown

Meredith Rolfe Dec 21, 2020 (Edited Feb 12) 10 points Due Feb 27, 11:59 PM

Your second assignment aims to build your skills for sharing your work in R, and presenting your code and results with external audiences.

For this assignment you will: (1) create your first R Markdown document, and (2) present the number of variables and the number of observations from your Homework One data set.

- For this assignment:
 1. Create an R Markdown document by using the dropdown in RStudio.
 2. Update the document to reflect your particular assignment and data source.
 3. Using your code from Homework One, load your data set.
 4. Write a short paragraph describing your data. This should include printing the number of observations
 5. (hint: the number of rows) and the number of variables (hint: the number of columns).
 6. 5a. Knit the Rmarkdown document to either a pdf (if you have installed latex) or an html file.

Exploring Data Set

Getting the dimensions of the dataset will reveal how long it is (rows) and how many variables there are (columns).

```
## [1] 243 731
```

contains 243 rows and 731 columns, which means there are 731 variables.

The summary for 601_WB_GovTech is quite long because there are hundreds of variables.

A quick review shows that variable with text appear to have a length of 243 rows and the numerical variables print the lowest number in the column. There is a note that 6 rows have been omitted because they've exceeded the maximum number of rows printed on the screen. These rows will need to be reviewed using another function.

Having a broad overview of this data set will be helpful during the data cleaning. I will start by inspecting the data type for each column to determine if the text will be converted to numerical or logical data types.

It's very unlikely that I will be analyzing all 731 variables; reviewing the column names provides a lot of insight into the information this data set contains and which ones I would like to explore further, remove, or define.

601_WB_GovTec is raw data published by the World Bank which uses organization specific indicator codes and evaluation metrics which are defined and explained in the accompanying resource documents.

Putting the first few columns into a table has provided basic information about this data set. The first 3 variables contain data I will not use and can be deleted. Most importantly the country name is in column 4 and is called “Economy”. One of the first things to do when cleaning this data set will be to rename “Economy” to a variable name that I’m able to remember.