Day2 Problem Set

Cecilia Sui

1/11/2022

Day 2 Outline:

1. Tips on Writing Code
2. Loading and Saving Data
3. R Notations
4. Modifying Values

Questions:

1. Get your working directory.

```
x <- 1
x
```

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

```
mean(x)
```

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

2. Create a new directory named Day2, and set your working directory to Day2.
3. List files in your working directory
4. Load the built-in dataset "USArrests" into a dataframe called df, and view the dataframe in the R data viewer.
5. Load the data from the "resources.csv" file, and store it in a dataframe named resources. This dataset is from the paper "Does Oil Hinder Democracy?' ' by Michael L. Ross (2001).
6. Find the dimension of resources. How many observations and how many variables does it have? What are the names of these variables?
7. Can you get the structure of the dataframe?
8. Select the first row and first column of resources.
9. The column regime stores the measurement of a country's level of democracy: -10 (authoritarian) to 10 (democratic). How many countries are between -5 to 5 inclusive according to this measurement?
10. How many NA's does each variable have? If you drop all NA's in the dataset, how many observations are left? (Please save your new dataset without NA's as a new dataframe. Do not modify in place.)
11. The column "oil" tells us the amount of oil exports as a percentage of the country's GDP. Can you find how the average of oil has changed before and after omitting all NA's in the table? Why caused the change in values?
12. Can you fill the NA's in each column using its mean that you found above?

13. The “illit” column shows the percentage of the population that is illiterate. Can you remove the entire column and save the new dataframe as resources? What happens if you attempt to access values in the illit column now after removal?

14. Which country has life expentacy over 76 (variable named life) in year 1980?

15. Can you rename the column “life” to “life_expectancy”?

16. In year 1990, which countries scored a 10 (most democratic) in their regime scores?

17. Add a variable “life” to the data frame which should categorize the level of life expectancy: [0,50) is low,[50,70) is medium, and [70,max] is high. Remember you have already found the maximum of life expectancy above.

18. Can you add a dummy variable that categorizes whether a country is democratic or not? Name the new column democracy, and put 1 if it is a democracy, 0 otherwise. If regime column is 0, you can treat it as a democracy here.

19. Find out which country with low life expectancy in year 1990 has the highest amount of oil exports as a percentage of the country’s GDP? What is its percent of oil exports?

20. Save the dataframe as “new_resources.csv” into the new directory data which you need to create first.