Exam 2 R Markdown Output

Grace Mappes

6/26/2020

knitr::opts\_chunk$set(echo = TRUE)  
knitr::opts\_chunk$set(error = TRUE)

### call some packages just in case

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## -- Attaching packages ----------------------------------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.1 v purrr 0.3.4  
## v tibble 3.0.1 v stringr 1.4.0  
## v tidyr 1.1.0 v forcats 0.5.0  
## v readr 1.3.1

## -- Conflicts -------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(WDI)

# 1. clear environment

rm(list=ls(all=TRUE))

# 2. call rio and import inequality data set

library(rio)

## Warning: package 'rio' was built under R version 4.0.2

inequality\_data <- import("inequality.xlsx")

# 3. Type of data set

This is a panel data set because there are multiple

# view class of inequality data set  
class(inequality\_data)

## [1] "data.frame"

# 4. show Denmark and Sweden inequality\_gini scores

inequality\_Denmark <- subset(inequality\_data,   
 inequality\_gini == "Denmark")  
inequality\_Sweden <- subset(inequality\_data,   
 inequality\_gini == "Sweden")

# 5. show Brazil inequality\_gini scores

inequality\_Brazil <- subset(inequality\_data, inequality\_gini == "Brazil")

# 6.

Based on my prior answers, it is better to have a lower inequality\_gini score than a higher score.

# 7. View head of data frame

head(inequality\_data)

## iso2c country inequality\_gini year  
## 1 AL Albania 32.9 2015  
## 2 AM Armenia 32.4 2015  
## 3 AT Austria 30.5 2015  
## 4 BY Belarús 25.6 2015  
## 5 BE Belgium 27.7 2015  
## 6 BZ Belize NA 2015

# 8. Write function accent.remove, apply to Belarus, run head again.

Also changed encoding to UTF-8 in global options before running code.

# write function accent.remove  
accent.remove <- function(s) {  
 old1 <- "ú"  
 new1 <- "u"  
 s1 <- chartr(old1, new1, s)  
}  
  
#apply accent.remove to data set  
inequality\_data$country = accent.remove(inequality\_data$country)  
  
#view head of data set to check for accent removal   
head(inequality\_data)

## iso2c country inequality\_gini year  
## 1 AL Albania 32.9 2015  
## 2 AM Armenia 32.4 2015  
## 3 AT Austria 30.5 2015  
## 4 BY Belarus 25.6 2015  
## 5 BE Belgium 27.7 2015  
## 6 BZ Belize NA 2015

# 9. Sort data by lowest inequality\_gini score

# sort data by lowest inequality\_gini score  
inequality\_data <- order(inequality\_data$inequality\_gini)  
  
# view head of sorted data set  
head(inequality\_data)

## [1] 161 190 4 39 92 160

# 10. Mean inequality\_gini score

mean(inequality\_data$inequality\_gini, na.rm=TRUE)

## Error in inequality\_data$inequality\_gini: $ operator is invalid for atomic vectors

As seen on the summary table above, the mean inequality\_gini score is 36.81.

# 11. Create high\_inequality and low\_inequality variables with ifelse.

#write ifelse  
high\_inequality <- ifelse(test = inequality\_data$inequality\_gini > 36.81, yes = 1, no = 0)

## Error in inequality\_data$inequality\_gini: $ operator is invalid for atomic vectors

low\_inequality <- ifelse(test = inequality\_data$inequality\_gini <= 36.81, yes = 1, no = 0)

## Error in inequality\_data$inequality\_gini: $ operator is invalid for atomic vectors

# 12. Cross tab of high\_inequality and low\_inequality

library(doBy)

##   
## Attaching package: 'doBy'

## The following object is masked from 'package:dplyr':  
##   
## order\_by

summaryBy(high\_inequality ~ low\_inequality, data = inequality\_data, FUN = c(mean,length),   
 na.rm=TRUE, keep.names = TRUE)

## Error in rep.int("1", nrow(data)): invalid type (NULL) for 'times' (must be a vector)

# 13 For Loop

First, I will make a vector that includes the names of the organizations.

x <- c("World Bank", "African Development Bank", "Bill and Melinda Gates Foundation")

Next I will write the for loop to print these names.

for (x in 1:x) {  
 print(paste(x))  
}

## Warning in 1:x: numerical expression has 3 elements: only the first used

## Warning: NAs introduced by coercion

## Error in 1:x: NA/NaN argument

# 14 Find variable from World Development Indicators

I chose the variable rural population because there is a stereotype that rural populations are more impoverished, and I would like to test that assertion.

# 15. Import rural population variable directly into R

# call WDI package to get rural population data  
library(WDI)  
population\_data = WDI(country = "all",  
indicator = c("SP.RUR.TOTL"), # indicator from web  
start = 2015, end = 2015, extra = FALSE, cache = NULL)

# 16. Renaming the variable for clarity

# check current variable of name with summary  
summary(population\_data)

## iso2c country SP.RUR.TOTL year   
## Length:264 Length:264 Min. :0.000e+00 Min. :2015   
## Class :character Class :character 1st Qu.:4.386e+05 1st Qu.:2015   
## Mode :character Mode :character Median :3.567e+06 Median :2015   
## Mean :1.438e+08 Mean :2015   
## 3rd Qu.:2.118e+07 3rd Qu.:2015   
## Max. :3.382e+09 Max. :2015   
## NA's :4

# load data.table package  
library(data.table)

##   
## Attaching package: 'data.table'

## The following object is masked from 'package:purrr':  
##   
## transpose

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

setnames(population\_data, "SP.RUR.TOTL", "rural\_pop")

# 17. Merge the data sets

# merge the data sets  
merged\_df <- dplyr::left\_join(inequality\_data, population\_data)

## Error in UseMethod("left\_join"): no applicable method for 'left\_join' applied to an object of class "c('integer', 'numeric')"

# check for names that end in .x or .y  
names(merged\_df)

## Error in eval(expr, envir, enclos): object 'merged\_df' not found

# 18. Remove missing data

# remove missing data from inequality\_gini  
merged\_df$inequality\_gini <- na.omit(merged\_df$inequality\_gini)

## Error in na.omit(merged\_df$inequality\_gini): object 'merged\_df' not found

# remove missing data from new variable, rural\_pop  
merged\_df$rural\_pop <- na.omit(merged\_df$rural\_pop)

## Error in na.omit(merged\_df$rural\_pop): object 'merged\_df' not found

# 19. Filter data, inequality\_gini > 30

data\_greater\_30 <-  
 merged\_df %>%  
 dplyr::filter(inequality\_gini > 30)

## Error in eval(lhs, parent, parent): object 'merged\_df' not found

# 20. Count sequence “ai” in country names

# 21. Sum of inequality\_gini in data\_greater\_30

# first subset inequality\_gini from data\_greater\_30  
data\_greater\_30\_gini <- subset(data\_greater\_30, select = "inequality\_gini")

## Error in subset(data\_greater\_30, select = "inequality\_gini"): object 'data\_greater\_30' not found

# use apply function to create a new variable with sum  
data\_greater\_30\_gini\_sum <- apply(data\_greater\_30\_gini, 2, sum)

## Error in apply(data\_greater\_30\_gini, 2, sum): object 'data\_greater\_30\_gini' not found

# view sum  
View(data\_greater\_30\_gini\_sum)

## Error in as.data.frame(x): object 'data\_greater\_30\_gini\_sum' not found

# 22. Label merged\_df variables

# install package labelled  
library(labelled)  
  
# label the variables   
var\_label(merged\_df) <- list(`iso2c` = "Country Code",  
`country` = "Country",  
`inequality\_gini` = "Gini Inequality Score",  
`year` = "Year,"  
`rural\_pop' = "Rural Population")

## Error: <text>:9:1: unexpected INCOMPLETE\_STRING  
## 8: `year` = "Year,"  
## 9: `rural\_pop' = "Rural Population")  
## ^

# 23. Save the data as a Stata format

# save the dataset in Stata format with the labels  
library(rio)  
export(merged\_df, file = "final\_data.dta")

## Error in is.data.frame(x): object 'merged\_df' not found