Exam 2

Anthony Douglas

6/26/2020

## R Markdown

#Github Repo: <https://github.com/AntDouglasJ1/Datascienceclass> #clearing environment and setting working directory

rm(list=ls(all=TRUE))  
setwd("C:/Users/antdo/Documents/R/Exam2")

#libraries

library(tidyverse)

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

## v ggplot2 3.3.1 v purrr 0.3.4  
## v tibble 3.0.1 v dplyr 1.0.0  
## v tidyr 1.1.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

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

library(rio)  
library(doBy)

##   
## Attaching package: 'doBy'

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

library(WDI)  
library(labelled)

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

library(tinytex)

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

#loading inequality dataset

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

## iso2c country inequality\_gini year   
## Length:203 Length:203 Min. :25.40 Min. :2015   
## Class :character Class :character 1st Qu.:31.55 1st Qu.:2015   
## Mode :character Mode :character Median :35.75 Median :2015   
## Mean :36.81 Mean :2015   
## 3rd Qu.:41.12 3rd Qu.:2015   
## Max. :59.10 Max. :2015   
## NA's :123

#Answering question 3. Print in R Markdown before turning in

print("Answer to number 3 is this is a cross-sectional dataset. The data provides a number of variables within a specific snapshot of time.")

## [1] "Answer to number 3 is this is a cross-sectional dataset. The data provides a number of variables within a specific snapshot of time."

#subset for Denmark  
subset(inequality\_data$inequality\_gini, inequality\_data$country == "Denmark")

## [1] 28.2

#Subset for Sweden  
subset(inequality\_data$inequality\_gini, inequality\_data$country == "Sweden")

## [1] 29.2

#Subset for Brazil  
subset(inequality\_data$inequality\_gini, inequality\_data$country == "Brazil")

## [1] 51.9

#Answering question 6, print in R markdown before turning in

print("It is better to have a lower gini score as that indicates less inequality.")

## [1] "It is better to have a lower gini score as that indicates less inequality."

#head command  
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

#Change accent from Belarus w/ function  
remove.accents <- function(s) {  
   
 #one character substitution  
 old1 <- "ú"  
 new1 <- "u"  
 s1 <- chartr(old1,new1,s)  
}  
#remove accents  
inequality\_data$country <- remove.accents(inequality\_data$country)  
  
#Checking Head  
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

#sorting  
inequality\_data <- inequality\_data[order(inequality\_data$inequality\_gini),]  
head(inequality\_data)

## iso2c country inequality\_gini year  
## 161 SI Slovenia 25.4 2015  
## 190 UA Ukraine 25.5 2015  
## 4 BY Belarus 25.6 2015  
## 39 CZ Czech Republic 25.9 2015  
## 92 XK Kosovo 26.5 2015  
## 160 SK Slovak Republic 26.5 2015

#mean  
inequality\_data <- na.omit(inequality\_data)  
inequality\_data\_mean <- mean(inequality\_data$inequality\_gini)  
  
#ifelse  
  
inequality\_data$low\_inequality <- ifelse(inequality\_data$inequality\_gini <= inequality\_data\_mean, yes = "1", no = "0")  
inequality\_data$high\_inequality <- ifelse(inequality\_data$inequality\_gini > inequality\_data\_mean, yes = "1", no = "0")  
#crosstab  
summaryBy(low\_inequality ~ high\_inequality, inequality\_data, FUN=c(mean,length))

## Warning in mean.default(x, ...): argument is not numeric or logical: returning  
## NA  
  
## Warning in mean.default(x, ...): argument is not numeric or logical: returning  
## NA

## high\_inequality low\_inequality.mean low\_inequality.length  
## 1 0 NA 46  
## 2 1 NA 34

#for loop  
  
names <- c("The World Bank", "The African Development Bank", "The Bill and Melinda GatesFoundation")  
for (i in names){  
 print(i)  
}

## [1] "The World Bank"  
## [1] "The African Development Bank"  
## [1] "The Bill and Melinda GatesFoundation"

print("I chose Poverty gap at $1.90 a day (2011 PPP) (%)), code:SI.POV.GAPS")

## [1] "I chose Poverty gap at $1.90 a day (2011 PPP) (%)), code:SI.POV.GAPS"

# Answer to question 14, put into R

#import variable into R

ppp\_data <- WDI(country = "all",indicator =c("SI.POV.GAPS"),  
 start = 2015, end = 2015, extra = FALSE, cache = NULL)  
#merge datasets  
merge\_df <- dplyr::full\_join(inequality\_data,ppp\_data)

## Joining, by = c("iso2c", "country", "year")

merge\_df <- na.omit(merge\_df)  
#filter  
data\_greater\_30 <-   
 merge\_df%>%   
 dplyr::filter((inequality\_gini>30))  
#sum of inequality\_gini from data\_greater\_30  
sumofineq <- function(x) {  
 sum(data\_greater\_30$inequality\_gini, na.rm=TRUE)  
 data\_greater\_30<- sapply(data\_greater\_30, sumofineq)  
data\_greater\_30  
}

#label variables

var\_label(merge\_df) <- list(`country` = "country",  
 `year`= "year",  
 `inequality\_gini`="Gini score",  
 `low\_inequality` = "Score of 1 indicates low inequlality",  
 `high\_inequality` = "Score of 1 indicates high inequlality",  
 `SI.POV.GAPS` = "Poverty gap at $1.90 a day (2011 PPP) (%)",  
 `iso2c`="ISO-3 country code")

#exporting

final\_data <- export(merge\_df, "final\_data.dta")