

Exam Two

Herrington

6/26/2020

```
#Loading the dataset into a dataframe
```

```
#Repo Link Repo Link
```

```
#1 and 2
```

```
library(rio)
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(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
##      order_by
```

```
library(devtools)
```

```
## Loading required package: usethis
```

```
library(remotes)
```

```
##
```

```
## Attaching package: 'remotes'
```

```
## The following objects are masked from 'package:devtools':
```

```
##
```

```
##      dev_package_deps, install_bioc, install_bitbucket, install_cran,
```

```
##      install_deps, install_dev, install_git, install_github,
```

```
##      install_gitlab, install_local, install_svn, install_url,
```

```
##      install_version, update_packages
```

```
## The following object is masked from 'package:usethis':
```

```
##
```

```
##      git_credentials
```

```

library(remote)

## Loading required package: Rcpp
## Loading required package: raster
## Loading required package: sp
##
## Attaching package: 'raster'
## The following object is masked from 'package:dplyr':
##
##      select
library(WDI)
library(data.table)

##
## Attaching package: 'data.table'
## The following object is masked from 'package:raster':
##
##      shift
## The following objects are masked from 'package:dplyr':
##
##      between, first, last
library(countrycode)
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 data.table::between() masks dplyr::between()
## x tidyr::extract()      masks raster::extract()
## x dplyr::filter()       masks stats::filter()
## x data.table::first()   masks dplyr::first()
## x dplyr::lag()          masks stats::lag()
## x data.table::last()    masks dplyr::last()
## x doBy::order_by()      masks dplyr::order_by()
## x raster::select()      masks dplyr::select()
## x purrr::transpose()    masks data.table::transpose()

inequality_data <- import("inequality.xlsx")
attach(inequality_data)

#3 The data is cross sectional, it is all from 2015. What follows are pulled years, all 2015
print(year)

##      [1] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
##     [16] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
##     [31] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
##     [46] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015

```

```
## [61] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [76] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [91] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [106] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [121] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [136] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [151] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [166] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [181] 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015
## [196] 2015 2015 2015 2015 2015 2015 2015 2015 2015
```

#4 Gini scores for Sweden and Denmark

```
ginisub <- subset(inequality_data, select = c("inequality_gini"))
sweGini <- filter(ginisub, country=="Sweden")
denmGini <- filter(ginisub, country=="Denmark")
#printing

print(paste("Sweden:", sweGini, "and Denmark:", denmGini))
```

```
## [1] "Sweden: 29.2 and Denmark: 28.2"
```

#5 Brazil Inequality Gini Score

```
brazGini <- filter(ginisub, country=="Brazil")
print(paste("Brazil:", brazGini))
```

```
## [1] "Brazil: 51.9"
```

#6 It seems to be better to have a generally lower gini score

#7

```
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

```
accent.remove <- function(n)
{
  inequality_data$country[inequality_data$country=="Belarús"] = "Belarus"
}
accent.remove()
```

#9 sort

```
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      Belarús           25.6 2015
```

```
## 39      CZ  Czech Republic      25.9 2015
## 92      XK          Kosovo      26.5 2015
## 160     SK Slovak Republic      26.5 2015
```

#10 Mean gini score

```
mean(inequality_gini, na.rm = TRUE)
```

```
## [1] 36.81375
```

The mean iniequality gini score is 36.81375

#11

```
ifelse(test = inequality_data$inequality_gini > mean(inequality_gini, na.rm = TRUE),
       yes = high_inequality <- 0, no = low_inequality <- 1)
```

```
## [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [26] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0
## [51] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [76] 0 0 0 0 0 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [101] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [126] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [151] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [176] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
## [201] NA NA NA
```

#12 Cross Tabulation

```
#summaryBy(high_inequality ~ low_inequality, data=inequality_data, FUN=c(mean,length))
```

#13

```
orgs <- c('World Bank', 'African Development Bank', 'Bill and Melinda Gates Foundation')
for (i in orgs){
  print(i)
}
```

```
## [1] "World Bank"
```

```
## [1] "African Development Bank"
```

```
## [1] "Bill and Melinda Gates Foundation"
```

#14 I picked income share held by the lowest 20%. I feel that this is a good indicator because the lower that income share is the higher inequality would likely be as wealth would be accumulating above that 20%.

#15

```
WDI_D = WDI(country = "all",
indicator = c("SI.DST.FRST.20"), start = 2015, end = 2015, extra = FALSE, cache = NULL)
```

#16 Rename the variable

```
setnames(WDI_D, "SI.DST.FRST.20", "Income Share Lowest 20%")
```

#17

```
#assign country codes and filter
```

```
WDI_D$country_code = countrycode(sourcevar = WDI_D$iso2c,
origin = "iso2c", destination = "iso3c", warn = TRUE)
```

```
## Warning in countrycode(sourcevar = WDI_D$iso2c, origin = "iso2c", destination = "iso3c", : Some values
```

```
WDI_D <- WDI_D %>% dplyr::filter(!(country_code=="NA"))
```

```
WDI_D <-
```

```
WDI_D %>% dplyr::filter(!(country_code=="NA"))
```

```
#merging
```

```
#merged_df = left_join(x=inequality_data, y=WDI_D, by =c("country_code", "year"))
```

```
#18
```

#23 Stata data set I did not make it here as I encountered some issues. I am still doing this part to prove that I understand how to export

```
export(inequality_data, file = "clean_dataset.dta")
```