Exam

Clay Cavanaugh

6/26/2020

{r setup, include=FALSE} knitr::opts\_chunk$set(echo = TRUE)

```{r include=TRUE} #clear the environment rm(list=ls(all=TRUE))

#loading dataset library(rio) inequality\_data = import(“inequality.xlsx”, which = 1)

head(inequality\_data)

#The dataset is cross sectional because the data is all observed within one time frame. In this case, the data set is pulled from 2015.

```{r include=TRUE} #clear the environment rm(list=ls(all=TRUE))

#loading dataset library(rio) inequality\_data = import(“inequality.xlsx”, which = 1)

#observing data frame head(inequality\_data) View(inequality\_data) summary(inequality\_data)

library(dplyr) library(tidyverse) library(ggplot2)

#subsetting subset(inequality\_data, inequality\_gin) inequality\_data$inequality\_gini

#no clue how to do this so skipping 4 and 5

#6 It would be better to have a low inequality score

#quick peak at data frame head(inequality\_data)

#creating function to remove accent on Belarus accent.remove <- function(s) { old1 <- “ú” new1 <- “u” s1 <- chartr(old1,new1,s) }

inequality\_datacountry)

#confirming function executed properly head(inequality\_data)

#sorting data by lowest inequality\_gini inequality\_data <- inequality\_data[order(inequality\_data$inequality\_gini),]

#viewing head command to confirm head(inequality\_data)

#to show the mean of inequality gini summary(inequality\_data) mean(inequality\_data$inequality\_gini, na.rm=TRUE) ## 10. The Mean Inequality gini is 36.81- both methods show mean

##ifelse low\_inequality < ifelse(test = inequality\_data$inequality\_gini==1, yes = high\_inequality, no = low\_inequality) #did not work

##cross tab library(doBy) summaryBy(inequality\_gini, data=inequality\_data, FUN=c(mean(length)) #cross tab did not work either

#for loop for World Bank World <- c(‘The World Bank’,‘African Development Bank’,‘Bill and Melinda Gates Foundation’), for (i in World){ print(i) }

##14 #I picked GDP per capita because I believe that a country’s overall economic output #will correlate well with the country’s inequality standards.

#importing variable library(WDI) gdp = WDI(indicator=‘NY.GDP.PCAP.KD’, country=all, start=2015, end=2015, extra=FALSE, cache= NULL) ##cannot figure out how to import… WDIsearch(‘gdp’)

dat = WDI(indicator=‘NY.GDP.PCAP.KD’, country=c(‘MX’,‘CA’,‘US’), start= 2015, end =2015) head(dat)

#renaming variable GDP\_per <- dat

#merging data frames merged\_df = left\_join(inequality\_data, GDP\_per, by = c(“country”, “year”)) merged\_df #verifying data frames

#removing missing data on inequality\_gini library(tidyverse) subset(merged\_df, inequality\_gini==“NA”) subset(merged\_df, GDP\_per==“NA”)

#filtering inequality gini scores merged\_df <- merged\_df %>% dplyr::filter(!(inequality\_gini>=30))

#counting how many countries data\_greater\_30 <- merged\_df data\_greater\_30 <- dat\_greater\_30 %>% mutate(country = ifelse(country == country“ai” “yes”, “no”)) ##unsure how to do this

#labelling variables library(labelled) var\_label(data\_greater\_30) <- list(country = “country”, inequality\_gini = “in”, primary\_enroll = “gross primary enrollment by population”, iso2c.x = “ISO-2 country code”, year = “year” ) ```