

## Exam 3

William Hall

7/8/2021

### **clear the environment**

```
rm(list=ls(all=TRUE))
```

### **set working directory where I will download the WDI data**

```
setwd("~/Users/sharonhall/Desktop/Exam 3")
```

### **install and load WDI package (Installed in the packages tab) along with others**

```
library(WDI) library(rio) library(tidyverse) library(data.table)
```

### **searching to see the specific package I need to download**

```
WDIsearch('labor force participation')
```

### **download the data set into the environment through the WDI package**

```
female_lfp = WDI(country="all", indicator=c("SL.TLF.CACT.FE.ZS"), start=2010, end=2015, extra=FALSE, cache=NULL)
```

```
#renaming the female labor force participation variable (indicator) flfp setnames(female_lfp, "SL.TLF.CACT.FE.ZS", "flfp")
```

### **collapsing the data frame so we can get the average flfp from 2010-2015**

```
collapsed_flfp <- female_lfp %>% group_by(iso2c, country) %>% summarize(flfp = mean(flfp, na.rm=TRUE))
```

## making a subset of data to only include countries with flfp < 15%

```
bad_flfp = subset(collapsed_flfp, flfp < 15) # after looking at the subset, we can see Iraq, Iran, Islamic Rep, Jordan, and Syria don't meet the 15% threshold
```

## need to load libraries to download world map

```
library(googlemaps) library(labelled) library(varhandle) library(ggrepel) library(geosphere) library(rgeos) library(viridis) library(mapview) library(rnaturalearth) library(rnaturalearthdata) library(devtools) library(remotes) library(raster) library(sp) library(sf) library(Imap) library(rnaturalearthhires) library(ggsflabel)
```

## load the world borders

```
world_borders <- st_read("/Users/sharonhall/Desktop/Module 11/world border shape files/World_Borders.shp", stringsAsFactors=FALSE)
```

## transform to WGS84 format and remove world borders

```
borders <- st_transform(world_borders, "+proj=latlong +ellps=WGS84 +datum=WGS84") rm(world_borders)
```

## creating missing iso2c data frame and merge with borders. also rename iso2c

```
collapsed_flfp_small <- na.omit(subset(collapsed_flfp, select=c("iso2c", "flfp")))
setnames(collapsed_flfp_small, "iso2c", "ISO2")
merged_data = left_join(borders, collapsed_flfp_small, by=c("ISO2"))
```

## mapping point data. get map of world from the natural earth package

```
world <- ne_countries(scale = "large", returnclass = "sf")
world_flfp_map = ggplot() + geom_sf(data = world) + geom_sf(data = merged_data, aes(fill=flfp)) +
scale_fill_viridis(option = "viridis") + ggtitle("Female Labor Participation Average by Country, 2010-2015")
+ theme(plot.title = element_text(hjust = 0.5)) + theme_void()
print(world_flfp_map)
```

#7. It seems Southeast Africa has the highest rate, an area that is considered underdeveloped and impoverished.

#8. What I would have done is subset it to those specific countries

#9. The three main components in Shiny app are the user interface, the server, and shinyApp which runs the function. The server includes an input and an output, it can't be run without those two items.

```

#10. Need to update libraries so we can pull the PDF library(pdftools) library(tidyr) library(tidytext)
library(stringr)

mytext=pdf_text(pdf = "https://pdf.usaid.gov/pdf_docs/PA00TNMJ.pdf")

mytext

#11. Covert the text to a data frame where each page is an observation armeniatext = as.data.frame(mytext,
stringsAsFactors=FALSE)

colnames(armeniatext)[which(names(armeniatext) == "mytext")] <- "text"

#12. Tokenizing text into words and removing the stop words armeniatext=armeniatext %>%
unnest_tokens(word, text)

armeniatext=armeniatext %>% anti_join(stop_words)

#13. Just need to run a word frequency to see what words come up the most hpfreq <- armeniatext %>%
count(word, sort = TRUE)

head(hpfreq)

#14. Upload the top100 on billboard top 100 library(rvest) library(dplyr)

hot100exam <- "https://www.billboard.com/charts/hot-100" hot100exam <-read_html(hot100page)

hot100exam

```