Exam3Markdown

Evan Shimek

7/8/2021

#1. clear environment  
rm(list=ls(all=TRUE))  
  
  
#2. download WDI data on female labor force participation  
library(WDI)

## Warning: package 'WDI' was built under R version 4.0.5

female\_lfp = WDI(country = "all",  
 indicator =c("SL.TLF.CACT.FE.NE.ZS"),  
 start = 2010, end = 2015, extra = FALSE, cache = NULL)  
  
  
#3. rename the female labor force participation variable flfp  
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.5

##   
## 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

female\_lfp = rename(female\_lfp, flfp = SL.TLF.CACT.FE.NE.ZS)  
  
  
#4. collapse by mean flfp by country  
  
collapsed\_flfp = female\_lfp %>%  
 group\_by(country, iso2c) %>%  
 summarize(flfp = mean(flfp, na.rm=TRUE))

## `summarise()` has grouped output by 'country'. You can override using the `.groups` argument.

#5. show which countries have average female force participation rates   
#for the 2010-2015 period that are less than 15%  
  
collapsed\_flfp$flfp = as.numeric(collapsed\_flfp$flfp)  
  
low\_participation\_countries = collapsed\_flfp %>%  
 filter(flfp < 15)  
  
print(low\_participation\_countries)

## # A tibble: 5 x 3  
## # Groups: country [5]  
## country iso2c flfp  
## <chr> <chr> <dbl>  
## 1 Iran, Islamic Rep. IR 13.8   
## 2 Iraq IQ 12.8   
## 3 Jordan JO 13.9   
## 4 Syrian Arab Republic SY 13.9   
## 5 Yemen, Rep. YE 8.01

#6. map collapsed\_flfp onto a world map, using viridis color scheme  
library(sf)

## Warning: package 'sf' was built under R version 4.0.5

## Linking to GEOS 3.9.0, GDAL 3.2.1, PROJ 7.2.1

library(rnaturalearth)

## Warning: package 'rnaturalearth' was built under R version 4.0.5

library(rnaturalearthhires)  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

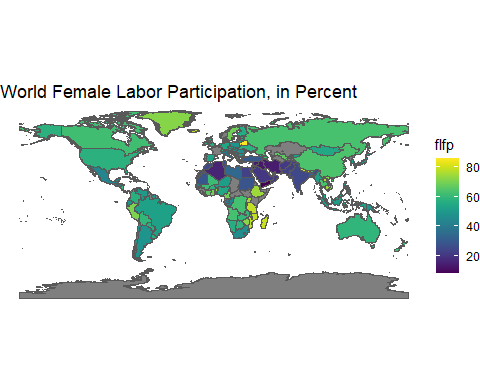
library(viridis)

## Warning: package 'viridis' was built under R version 4.0.5

## Loading required package: viridisLite

## Warning: package 'viridisLite' was built under R version 4.0.5

world <-ne\_countries(scale = "large", returnclass = "sf")  
  
world = rename(world, iso2c = iso\_a2)  
  
merged\_data = left\_join(world, collapsed\_flfp, by="iso2c")  
  
  
collapsed\_flfp\_map = ggplot() +  
 geom\_sf(data = world) +   
 geom\_sf(data = merged\_data, aes(fill=`flfp`)) +   
 scale\_fill\_viridis(option = "viridis") +   
 ggtitle("World Female Labor Participation, in Percent") +   
 theme(plot.title = element\_text(hjust = 0.5)) +   
 theme\_void()   
  
print(collapsed\_flfp\_map)



1. Based on the map, which area of the world has, perhaps surprisingly, a cluster of yellow-colored average female labor force participation rate states, indicating the highest on the scale?

**A:** Based on the map, a few nations in Southeast Africa seem to have higher-than-average female labor force participation (nearly at the top of the scale).

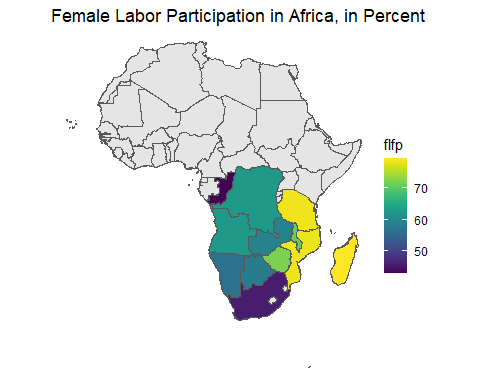
# 8. show the same cluster of states referenced in the previous question   
library(data.table)

## Warning: package 'data.table' was built under R version 4.0.5

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

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

africa = ne\_countries(continent = 'Africa',  
 scale = "large",  
 returnclass = "sf")  
  
setnames(merged\_data, "sovereignt", "country")  
  
  
africa\_only\_data <-subset(merged\_data, country=="United Republic of Tanzania"|  
 country=="Democratic Republic of the Congo"|country=="Angola"|  
 country=="Mozambique"|country=="Malawi"|country=="Zambia"|  
 country=="Madagascar"|country=="Zimbabwe"|country=="Eswatini"|  
 country=="Namibia"|country=="Botswana"| country=="South Africa"|  
 country=="Republic of the Congo")  
  
  
africa\_flfp\_map = ggplot() +  
 geom\_sf(data = africa) +   
 geom\_sf(data = africa\_only\_data, aes(fill=`flfp`)) +   
 scale\_fill\_viridis(option = "viridis") +   
 ggtitle("Female Labor Participation in Africa, in Percent") +   
 theme(plot.title = element\_text(hjust = 0.5)) +   
 theme\_void()   
  
print(africa\_flfp\_map)



1. In a Shiny app, what are the three main components and their subcomponents?

**A:** A shiny app is comprised of three main components: a UI, a server, and code to execute the shinyapp. The UI requires a style of input and output to be specified, while the server requires the specification of the back-end code that actually transforms the input to the output.

#10. Pull Pdf from Mike's webpage  
library(pdftools)

## Warning: package 'pdftools' was built under R version 4.0.5

## Using poppler version 21.04.0

mytext = pdf\_text(pdf = "https://pdf.usaid.gov/pdf\_docs/PA00TNMJ.pdf")  
  
  
#11. Convert pdf to data frame  
armeniatext = as.data.frame(mytext, stringsAsFactors=FALSE)  
  
  
#12. Tokenize by word, removing stop words  
library(tidytext)

## Warning: package 'tidytext' was built under R version 4.0.5

library(stringr)

## Warning: package 'stringr' was built under R version 4.0.5

library(tidyr)

## Warning: package 'tidyr' was built under R version 4.0.5

armeniatext = armeniatext %>%  
 unnest\_tokens(word, mytext)  
  
data("stop\_words")  
armeniatext = armeniatext %>%  
 anti\_join(stop\_words)

## Joining, by = "word"

#13. Figure out the top 5 most used words  
most\_used = armeniatext %>%  
 count(word, sort=T)  
  
print(most\_used[1:10,])

## word n  
## 1 law 276  
## 2 corruption 242  
## 3 rule 206  
## 4 armenia 195  
## 5 european 105  
## 6 political 102  
## 7 review 94  
## 8 evidence 90  
## 9 2018 86  
## 10 reform 79

#14. Load the billboard Hot 100 webpage  
library(rvest)

## Warning: package 'rvest' was built under R version 4.0.5

hot100exam = rvest::read\_html("https://www.billboard.com/charts/hot-100")  
  
  
#15. Identify all nodes using rvest  
body\_nodes = hot100exam %>%  
 rvest::html\_node("body") %>%  
 rvest::html\_children()  
  
  
#16. Pull data on Rank, Artist, Title, and Last Week  
rank = hot100exam %>%  
 rvest::html\_nodes('body') %>%  
 xml2::xml\_find\_all("//span[contains(@class,   
 'chart-element\_\_rank\_\_number')]") %>%  
 rvest::html\_text()  
  
artist <-hot100exam %>%  
 rvest::html\_nodes('body') %>%  
 xml2::xml\_find\_all("//span[contains(@class,  
 'chart-element\_\_information\_\_artist')]") %>%  
 rvest::html\_text()  
  
title <-hot100exam %>%   
 rvest::html\_nodes('body') %>%  
 xml2::xml\_find\_all("//span[contains(@class,  
 'chart-element\_\_information\_\_song')]") %>%  
 rvest::html\_text()  
  
last\_week <-hot100exam %>%   
 rvest::html\_nodes('body') %>%  
 xml2::xml\_find\_all("//span[contains(@class,  
 'chart-element\_\_meta text--center color--secondary text--last')]") %>%  
 rvest::html\_text()  
  
#16b Save as dataframe, .dta file  
chart\_df = data.frame(rank, artist, title, last\_week)  
library(rio)

## Warning: package 'rio' was built under R version 4.0.5

export(chart\_df, "hot100examchart.dta")

1. What is the link to your GitHub repository?

**A:** <https://github.com/eshimek/Exam3>