Exam 3

Aakansha Bagepally

07/08/21

```
#clear the environment
rm(list = ls(all = TRUE))

#load packages
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(rio)
library(WDI)
library(googlesheets4)
library(labelled)
library(data.table)
```

```
##
## Attaching package: 'data.table'
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
## The following object is masked from 'package:purrr':
##
##     transpose
```

```
library(varhandle)
library(ggrepel)
library(geosphere)
library(rgeos)
```

```
## Loading required package: sp
## rgeos version: 0.5-5, (SVN revision 640)
##  GEOS runtime version: 3.8.0-CAPI-1.13.1
##  Linking to sp version: 1.4-5
##  Polygon checking: TRUE
```

```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
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(raster)
```

```
##
## Attaching package: 'raster'
```

```
## The following object is masked from 'package:data.table':
##
##     shift
```

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

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(sp)
library(Imap)
```

```
##
## Attaching package: 'Imap'
```

```
## The following object is masked from 'package:purrr':
##
##     imap
```

```
library(devtools)
library(sf)
```

```
## Linking to GEOS 3.8.0, GDAL 3.0.4, PROJ 6.3.1
```

```
library(rnaturalearthhires)
library(mapview)
library(rnaturalearth)
library(rnaturalearthdata)
library(ggsflabel)
```

```
##
## Attaching package: 'ggsflabel'

## The following objects are masked from 'package:ggplot2':
##
##     geom_sf_label, geom_sf_text, StatSfCoordinates
```

```
#load data
female_lfp = WDI(country="all", indicator="SL.TLF.CACT.FE.ZS", start=2010, end=2015, extra=FALSE, cache=

#renaming to flfp
female_lfp = female_lfp %>% rename(flfp = SL.TLF.CACT.FE.ZS)

#collapsing flfp by mean by country
collapsed_flfp = female_lfp %>% group_by(country) %>% summarize(mean_flfp = mean(flfp, na.rm = TRUE), is
```

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

```
#coutries with flfp % under 15 during 2010-2015
under_15 = collapsed_flfp %>% filter(mean_flfp < 15)
under_15
```

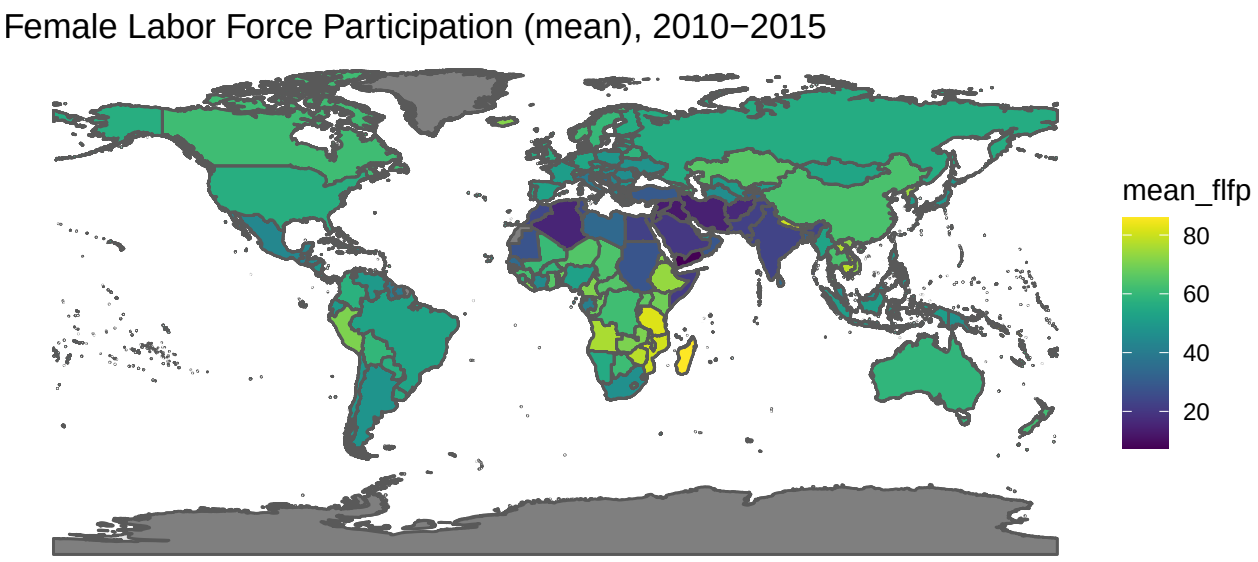
```
## # A tibble: 30 x 3
## # Groups:   country [5]
##    country            mean_flfp iso2c
##    <chr>                  <dbl> <chr>
##  1 Iran, Islamic Rep.      14.5 IR
##  2 Iran, Islamic Rep.      14.5 IR
##  3 Iran, Islamic Rep.      14.5 IR
##  4 Iran, Islamic Rep.      14.5 IR
##  5 Iran, Islamic Rep.      14.5 IR
##  6 Iran, Islamic Rep.      14.5 IR
##  7 Iraq                    12.7 IQ
##  8 Iraq                    12.7 IQ
##  9 Iraq                    12.7 IQ
## 10 Iraq                    12.7 IQ
## # ... with 20 more rows
```

```
#map of collapsed data
world_borders = st_read("World_Borders.shx",stringsAsFactors=FALSE)
```

```
## Reading layer `World_Borders' from data source `/cloud/project/World_Borders.shx' using driver `ESRI
## Simple feature collection with 246 features and 11 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -180 ymin: -90 xmax: 180 ymax: 83.6236
## Geodetic CRS:  WGS 84
```

```
borders = st_transform(world_borders, "+proj=longlat +ellps=WGS84 +datum=WGS84")
rm(world_borders)
world <-ne_countries(scale = "large", returnclass = "sf")
world = rename(world, iso2c = iso_a2)
setnames(borders, "ISO2", "iso2c")
merge_data = left_join(borders, collapsed_flfp, by = "iso2c")
merge_data = st_sf(merge_data)
flfp_map = ggplot() + geom_sf(data = world) +
  geom_sf(data = merge_data, aes(fill=mean_flfp)) +
```

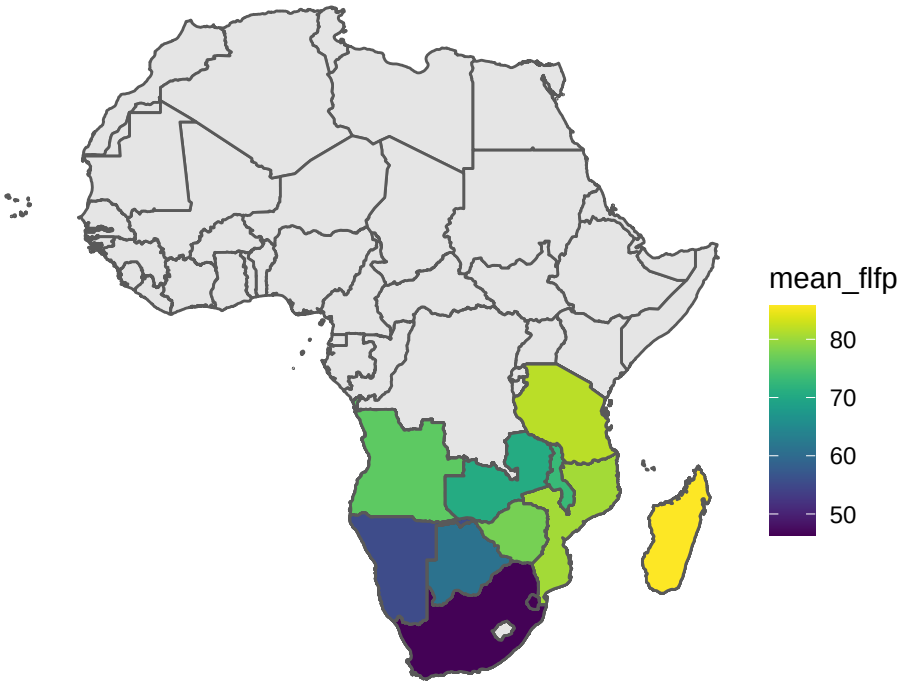
```
  scale_fill_viridis(option = "viridis") +
  ggtitle("Female Labor Force Participation (mean), 2010-2015")+
  theme(plot.title = element_text(hjust = 0.5)) +
  theme_void()
flfp_map
```



The map displayed above shows that some countries in the South of Africa have, perhaps surprisingly, high percentages of female labor force participation (shown in yellow).

```
#Mapping Africa

africa <-ne_countries(continent ='africa',scale = "large",returnclass = "sf")
africa_data <- subset(merge_data, country == "Namibia"|
                      country=="Tanzania"|
                      country=="Madagascar"|country=="Botswana"|
                      country=="Zimbabwe"|country=="Angola"|
                      country=="Malawi"|country=="Zambia"|
                      country=="Mozambique"|country=="Eswatini"|
                      country=="South Africa"|country=="Republic of the Congo")
africa_map = ggplot() + geom_sf(data=africa) + geom_sf(data=africa_data, aes(fill=mean_flfp))+
  scale_fill_viridis(option="viridis") +theme_void()
africa_map
```



A Shiny app has 3 main components: a UI, a server, execution with code. UI requires input and output, a server needs code to direct and use fucntions to output.

```
#pulling and converting pdf
library(pdftools)
```

```
## Using poppler version 0.86.1
```

```
library(tidytext)
library(stringr)
library(tidyr)
pdf_USAID = pdf_text(pdf = "https://pdf.usaid.gov/pdf_docs/PA00TNMJ.pdf")
armeniatext = as.data.frame(pdf_USAID, stringsAsFactors=FALSE)

#tokenize and remove stop words
armeniatext <- armeniatext %>% unnest_tokens(word, pdf_USAID)
data(stop_words)
armeniatext <- armeniatext %>%anti_join(stop_words)
```

```
## Joining, by = "word"
```

```
#top five most used words
armenia_5 <- armeniatext %>% count(word, sort = TRUE)
print(armenia_5[1:5,])
```

```
##         word   n
## 1        law 276
## 2 corruption 242
## 3       rule 206
## 4    armenia 195
## 5   european 105
```

```
#loading Billboard Hot 100
library(rvest)
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```
library(xml2)
library(rio)
hot100exam <-read_html("https://www.billboard.com/charts/hot-100")

#identify nodes
body_nodes <- hot100exam %>%
  html_node("body") %>% html_children()

#pulling 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()

lastweek <-hot100exam %>%
  rvest::html_nodes('body') %>%
  xml2::xml_find_all("//span[contains(@class,
                     'chart-element__meta text--center color--secondary text--last')]") %>%
  rvest::html_text()

hot100dataframe <- data.frame(rank, artist, title, lastweek)
export(hot100dataframe, "hot100exam.dta")
```

link to Git Repo: https://github.com/akshab/exam3