Final Exam

Echo Nattinger (en6235)

7/8/2021

GITHUB REPO LINK: <https://github.com/enatt/Exam3>

knitr::opts\_chunk$set(message = FALSE)

rm(list=ls(all=TRUE))

# Loading and preparing data  
library(WDI)  
library(tidyverse)  
female\_lfp = WDI(country="all",  
 indicator="SL.TLF.CACT.FE.ZS",   
 start=2010, end=2015,   
 extra=FALSE, cache=NULL)  
  
female\_lfp = female\_lfp%>%  
 rename(flfp = SL.TLF.CACT.FE.ZS)  
  
flfp\_no\_NAs = na.omit(female\_lfp, select=c("flp"))  
# Collapsing data  
collapsed\_flfp = flfp\_no\_NAs%>%  
 group\_by(country)%>%  
 summarize(average = mean(flfp),  
 iso2c = iso2c)  
  
# Which countries have average participation rates less than 15%  
q5 = collapsed\_flfp %>%  
 filter(average < 15)  
q5

## # A tibble: 30 x 3  
## # Groups: country [5]  
## country average 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

# Libraries  
library(sp)  
library(sf)

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

library(geosphere)  
library(devtools)

## Loading required package: usethis

library(rnaturalearthhires)  
library(ggsflabel)

##   
## Attaching package: 'ggsflabel'

## The following objects are masked from 'package:ggplot2':  
##   
## geom\_sf\_label, geom\_sf\_text, StatSfCoordinates

library(mapdata)

## Loading required package: maps

##   
## Attaching package: 'maps'

## The following object is masked from 'package:purrr':  
##   
## map

library(rio)  
library(tidyverse)  
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(googlesheets4)  
library(varhandle)  
library(ggrepel)  
library(rgeos)

## 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

##   
## Attaching package: 'viridis'

## The following object is masked from 'package:maps':  
##   
## unemp

library(mapview)  
library(rnaturalearth)  
library(rnaturalearthdata)  
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(Imap)

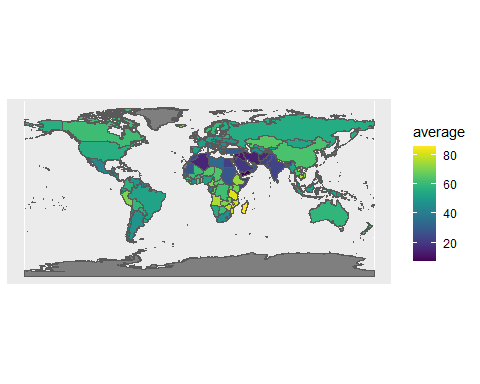
##   
## Attaching package: 'Imap'

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

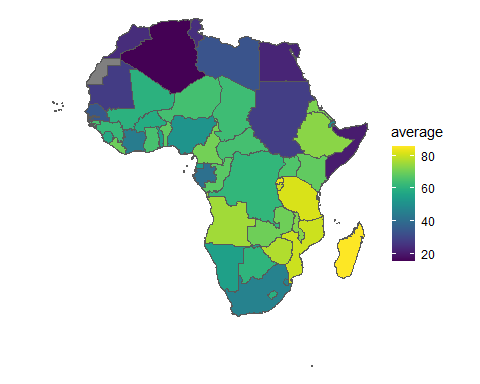
# Loading map  
# Adding border data to female participation rate data  
world\_borders = st\_read("World\_Borders.shp", stringsAsFactors=FALSE)

## Reading layer `World\_Borders' from data source   
## `C:\Users\RNatt\Desktop\Exam 3\World\_Borders.shp' using driver `ESRI Shapefile'  
## 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)  
  
# Changing the name of borders to make merging easier  
setnames(borders, "NAME", "country")  
setnames(borders, "ISO2", "iso2c")  
# Merging  
data\_merged = left\_join(borders, collapsed\_flfp, by=c("iso2c"))  
data\_merged = st\_sf(data\_merged)  
  
# Getting map of the world  
worldmap = ne\_countries(scale="large", returnclass="sf")  
  
# Mapping the data  
ggplot()+  
 geom\_sf(data=worldmap)+  
 geom\_sf(data=data\_merged, aes(fill=average))+  
 scale\_fill\_viridis(option="viridis")

 Based on the above map, regions in Southern Africa have high amounts of female participation in the workforce.

# Mapping Africa  
africamap = ne\_countries(continent="africa", scale="large", returnclass="sf")  
africadata = subset(data\_merged, iso2c %in% africamap$iso\_a2)  
ggplot()+  
 geom\_sf(data=africamap)+  
 geom\_sf(data=africadata, aes(fill=average))+  
 scale\_fill\_viridis(option="viridis")+  
 theme\_void()



1. A Shiny app involves 1) a user-interface (UI), 2) a server, and 3) executing the Shiny app. Within the UI, you must specify both Inputs and Outputs within a fluidPage. Within the server function, you store directions in objects and use render functions to output what’s stored in those objects. And when you execute the Shiny App, you specify the names of your UI an dserver.

# Loading and preparing PDF  
library(pdftools)  
library(tidytext)  
library(stringr)  
data(stop\_words)  
USAID = pdf\_text("PA00TNMJ.pdf")  
armeniatext = as.data.frame(USAID, stringAsFactors=FALSE)  
armeniatext$page = c(1:59)  
colnames(armeniatext)[which(names(armeniatext)=="USAID")] = "text"  
# Tokenize and remove stop words  
tidyarmenia = armeniatext%>%  
 unnest\_tokens(word, text)%>%  
 anti\_join(stop\_words)  
# Word counts  
armenia\_count = tidyarmenia%>%  
 count(word, sort=TRUE)%>%  
 top\_n(5, n)  
armenia\_count

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

# The top five words are LAW, CORRUPTION, RULE, ARMENIA, and EUROPEAN

library(rvest)  
library(dplyr)  
library(ggplot2)  
library(xml2)  
# Loading webpage  
hot100exam = "https://www.billboard.com/charts/hot-100"  
hot100exam = read\_html(hot100exam)  
# Getting nodes  
hot100nodes = hot100exam%>%  
 html\_nodes("body")%>%  
 html\_children()  
hot100nodes

## {xml\_nodeset (37)}  
## [1] <div class="header-wrapper ">\n<header id="site-header" class="site-head ...  
## [2] <div class="site-header\_\_placeholder"></div>  
## [3] <script>\n var PGM = window.PGM || {};\n PGM.config = PGM. ...  
## [4] <main id="main" class="page-content"><div id="charts" data-page-title="T ...  
## [5] <div class="ad\_desktop dfp-ad dfp-ad-promo " data-position="promo" data- ...  
## [6] <div class="ad-container footerboard footerboard--bottom">\n <div cla ...  
## [7] <footer id="site-footer" class="site-footer"><div class="container foote ...  
## [8] <div class="biz-modal">\n <div class="biz-modal\_\_content">\n < ...  
## [9] <script>\n window.CLARITY = window.CLARITY || [];\n</script>  
## [10] <div class="ad\_clarity" data-out-of-page="true" style="display: none;">< ...  
## [11] <script>\n\n window.top.pageLevelKeys = {};\n window.top.pageAdZon ...  
## [12] <script type="text/javascript" async="async" data-cfasync="false" src="h ...  
## [13] <script type="text/javascript">\n let detectDevice = function() {\n ...  
## [14] <script src="https://cdn.cookielaw.org/opt-out/otCCPAiab.js" type="text/ ...  
## [15] <script>\n\n function loadEUScript(source, attributes = {}) {\n\n ...  
## [16] <script src="https://geolocation.onetrust.com/cookieconsentpub/v1/geo/lo ...  
## [17] <script src="https://www.billboard.com/assets/1624920239/js/vendors\_/art ...  
## [18] <script src="https://www.billboard.com/assets/1624920239/js/vendors\_/clo ...  
## [19] <script src="https://www.billboard.com/assets/1624920239/js/vendors\_/rea ...  
## [20] <script src="https://www.billboard.com/assets/1624920239/js/vendors\_/rea ...  
## ...

# Pulling 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()  
  
lastweek= hot100exam%>%  
 html\_nodes("body")%>%  
 xml\_find\_all("//span[contains(@class, 'chart-element\_\_meta text--center color--secondary text--last')]")%>%  
 html\_text()  
  
hot100df = data.frame(rank, artist, title, lastweek)

library(rio)  
export(collapsed\_flfp, "collapsed\_flfp.dta")