Exam 3

Kavya Sethi

7/8/2021

Table of Contents

# 1. Clear the evironment

rm(list=ls(all=TRUE))

# 2. Female Labor Force Participation download

Googled the indicator and got <https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS> as the website.

library(WDI)  
  
female\_lfp = WDI(country = "all", indicator = c("SL.TLF.CACT.FE.ZS"),  
 start = 2010, # start of Female Labor Force Data   
 end = 2015, # end of of Female Labor Force Data   
 extra = FALSE, cache = NULL)

# 3. Rename female labor force participation variable

library(data.table)   
setnames(female\_lfp,"SL.TLF.CACT.FE.ZS", "flfp",skip\_absent = TRUE)

# 4. Collapse female\_lfp by the mean value for flfp for each country.

#collapsed\_flfp <-   
 #female\_lfp  
  
#collapsed\_flfp <-   
 #female\_lfp %>%   
 #group\_by(country,iso2c) %>%   
 #ummarize(flfp = mean(flfp, na.rm=TRUE))  
  
#collapsed\_flfp$year = NULL # Removes year from dataframe

# 5. Show participation rates that are less than 15%

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':  
##   
## between, first, last

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

#filter(collapsed\_flfp,flfp <= 15)

# 6. World Map presenting collapsed\_flfp data.

## Must first run necessary libraries.

library(rio)   
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.5 ✓ purrr 0.3.4  
## ✓ tibble 3.1.2 ✓ stringr 1.4.0  
## ✓ tidyr 1.1.3 ✓ forcats 0.5.1  
## ✓ readr 1.4.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::between() masks data.table::between()  
## x dplyr::filter() masks stats::filter()  
## x dplyr::first() masks data.table::first()  
## x dplyr::lag() masks stats::lag()  
## x dplyr::last() masks data.table::last()  
## x purrr::transpose() masks data.table::transpose()

library(googlesheets4)   
library(labelled)   
library(data.table)  
  
library(varhandle)   
library(ggrepel)   
library(geosphere)   
library(rgeos)

## Loading required package: sp

## rgeos version: 0.5-5, (SVN revision 640)  
## GEOS runtime version: 3.8.1-CAPI-1.13.3   
## Linking to sp version: 1.4-5   
## Polygon checking: TRUE

library(viridis)

## Loading required package: viridisLite

library(mapview)  
library(rnaturalearth)  
library(rnaturalearthdata)  
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:tidyr':  
##   
## extract

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

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

library(sp)  
  
library(sf)

## Linking to GEOS 3.8.1, GDAL 3.2.1, PROJ 7.2.1

library(Imap)

##   
## Attaching package: 'Imap'

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

library(rnaturalearthhires)  
library(ggsflabel)

##   
## Attaching package: 'ggsflabel'

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

## Loading/Transforming Shapefiles

Must first load shape file of the world.

world\_borders <- st\_read('/Users/kavyasethi/Documents/DataScience/Module 11 - GIS/world border shape files')

## Reading layer `World\_Borders' from data source   
## `/Users/kavyasethi/Documents/DataScience/Module 11 - GIS/world border shape files'   
## 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

Now transform shapefiles into WGS84 format.

borders <-st\_transform(world\_borders, "+proj=longlat +ellps=WGS84 +datum=WGS84")  
  
# remove world\_boarders cus no longer needed.   
rm(world\_borders)

## Fix Data for Map

Drop NAs for Map

#collapsed\_flfp = na.omit(collapsed\_flfp,select(flfp))

## Now merge tables of collapsed\_flfp and borders to get lat/long

library(data.table)   
#setnames(collapsed\_flfp,"iso2c","ISO2")  
  
#merged\_data = left\_join(borders, collapsed\_flfp, by=c("ISO2"))

## Now get map of the world

world <- ne\_countries(scale = "large", returnclass = "sf")

## Now Map the Viridis Color Map

#collapsed\_flfp\_map = ggplot() +  
#geom\_sf(data = world) +  
#geom\_sf(data = merged\_data, aes(fill=`flfp`)) + scale\_fill\_viridis(option = "viridis") +  
#ggtitle("female labor force participation, 2010-2015") + theme(plot.title = element\_text(hjust = 0.5)) +  
#theme\_void()  
  
#print(collapsed\_flfp\_map) Printing was taking to long to load.   
  
#ggsave(collapsed\_flfp\_map, filename = "femalelaborforceworldmap.png", width = 6.5, height = 6)

# 7. The southern region of Africa has the largest cluster of yellow (percentage of female labor force participation)

# 8. Use R to show cluster of yellow states.

The instruction don’t detail that it only has to be the cluster that shows up. So looking at color scale, we see upward of 75% are the yellowish states.

library(dplyr)  
  
#filter(collapsed\_flfp,flfp >= 75)

# 9. Shiny components

Shiny has three main components: UI, Server, Excecution

The Subcomponents of UI: Input and Output

# 10. Pull Mike Denly’s PDF

library(pdftools)

## Using poppler version 20.12.1

library(tidyr)   
library(tidytext)   
library(dplyr)   
library(stringr)   
library(ggplot2)  
  
denlypdf = pdf\_text("https://pdf.usaid.gov/pdf\_docs/PA00TNMJ.pdf")

# 11. Convert pdf character vector to dataframe

armeniatext=as.data.frame(denlypdf)  
armeniatext$page=c(1:59)

# 12. Tokenize and Remove Stop Words

R keeps giving me error for code below. Error: by must be supplied when x and y have no common variables. ℹ use by = character()` to perform a cross-join. I am unable to fix the error. I searched online and in notes. I literally copied and pasted the code from the notes. Its not fixable. I just put the code down that I know would work from the basis of the notes. Also the notes pdf for harrypoter/trump analysis wasn’t live, so I was not able to truly test the concept.

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

# 13 Figure Out the top 5 Most Common word in report.

Continued Error because of previous error. I just put down the code of how to do it.

#freq <- armeniatext %>%   
 #count(word, sort = TRUE)  
#head(freq)

# 14 Load Billboard Page

library(rvest)

##   
## Attaching package: 'rvest'

## The following object is masked from 'package:readr':  
##   
## guess\_encoding

library(dplyr)  
  
  
hot100exampage <- "https://www.billboard.com/charts/hot-100"   
hot100exam <- read\_html(hot100exampage)  
  
hot100exam

## {html\_document}  
## <html class="" lang="">  
## [1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-8 ...  
## [2] <body class="chart-page chart-page-" data-trackcategory="Charts-TheHot100 ...

# 15 Identify All the nodes in Webpage

body\_nodes <- hot100exam %>%   
 html\_node("body") %>%   
 html\_children()  
body\_nodes

## {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 ...  
## ...

head\_nodes <- hot100exam %>%   
 html\_node("head") %>%   
 html\_children()  
head\_nodes

## {xml\_nodeset (57)}  
## [1] <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">\n  
## [2] <meta charset="utf-8">\n  
## [3] <meta http-equiv="X-UA-Compatible" content="IE=edge">\n  
## [4] <meta name="viewport" content="width=device-width, initial-scale=1, user ...  
## [5] <title>The Hot 100 Chart | Billboard</title>\n  
## [6] <meta name="title" property="title" content="The Hot 100 Chart">\n  
## [7] <meta name="twitter:site" content="@billboard">\n  
## [8] <meta property="og:site\_name" content="Billboard">\n  
## [9] <meta property="og:type" content="article">\n  
## [10] <link rel="manifest" href="/manifest.json">\n  
## [11] <style>\n .chart-pro-access {\n background-image: url( ...  
## [12] <script type="text/javascript" async="async" data-cfasync="false" src="h ...  
## [13] <link rel="canonical" href="/charts/hot-100">\n  
## [14] <link rel="shortcut icon" href="https://www.billboard.com/assets/1624920 ...  
## [15] <link rel="apple-touch-icon" href="https://www.billboard.com/assets/1624 ...  
## [16] <link rel="dns-prefetch" href="https://www.billboard.com">\n  
## [17] <link rel="dns-prefetch" href="https://www.google-analytics.com/">\n  
## [18] <link rel="dns-prefetch" href="https://www.googletagservices.com/">\n  
## [19] <link rel="dns-prefetch" href="https://www.googletagmanager.com/">\n  
## [20] <link rel="dns-prefetch" href="https://www.instagram.com/">\n  
## ...

# 16 Billboard Ranking This Week including 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()  
  
  
  
chart\_df <- data.frame(rank, artist, title,lastweek)  
  
knitr::kable(  
 chart\_df %>% head(10))

|  |  |  |  |
| --- | --- | --- | --- |
| rank | artist | title | lastweek |
| 1 | BTS | Butter | 1 |
| 2 | Olivia Rodrigo | Good 4 U | 2 |
| 3 | Doja Cat Featuring SZA | Kiss Me More | 4 |
| 4 | Dua Lipa Featuring DaBaby | Levitating | 3 |
| 5 | Ed Sheeran | Bad Habits | - |
| 6 | Silk Sonic (Bruno Mars & Anderson .Paak) | Leave The Door Open | 6 |
| 7 | Justin Bieber Featuring Daniel Caesar & Giveon | Peaches | 5 |
| 8 | Lil Nas X | Montero (Call Me By Your Name) | 8 |
| 9 | The Weeknd & Ariana Grande | Save Your Tears | 7 |
| 10 | Olivia Rodrigo | Deja Vu | 9 |

# 17. GitHUB repo link

<https://github.com/kavyasethi14/KavyaSethi_GOVF355M_Exam_3.git>

# 18. Rmarkdwon keeps giving me an error

I included the pictures in this pdf. So I had to comment every r chunk that included the collapsed\_flfp. I know it works when I individually run the chunks, but rmarkdown isn’t letting it load.

[proof of filter working on question 5](proofoffilterworking.png)

[error on collapsed\_flfp](Rmarkdownerror.png)

Attached proof will also be uploaded to github I also exported all the dataframes from where the error occured. Will uplaod to github. I also exported the map, will uplaod to github.