Week 11 Diary Entry

Dawn Cheung

2023-11-03

R Markdown

1. List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?),

Depended variables will be % population with internet access, internet speed by country, and cost of internet (1GB) by country. Independent variables will be space, time, GINI, GDP, % population in poverty and possibly other forms of inequality like gender inequality. (I believe you're more of an expert on econ indicators than me so im hoping I dont have to explain?)

So basically with that i can use whatever data visualization i want. But I was thinking to do the map (I showed you last session, but with the other depended variables, as separate maps), and scatter plots for GINI, GDP, % population in poverty. I think for time, I'll calculate the rate of "progress" and plot that over the independent variables? (I don't think I'll have the time though). And pie charts for the 3i survey data and percentage of overall world population with internet access!

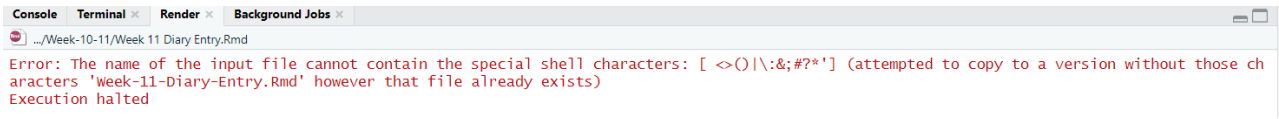
2. How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive) I initially thought to do a slider for time, for the maps. So as you move a slider, the % population with internet access/internet speed by country/cost of internet changes fill of each country changes to the year the slider is on. Might use plot_ly() for this. But that doesn't entirely *show* anything, so not sure yet. I could also do a "tab" function where you sift through the different scatter plots before I highlight which show the largest correlation, to encourage a more exploratory approach to the investigative question.
3. What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

```
library(tibble)
concepts <- tibble(
  weekNumber = list(1, 2, 3, 4, 5, 6, 7, ""),
  conceptTaught = c("knitr::include_graphics()", "Manipulation of csv data", "Variable manipulation", "The concept of filter
ing! Especially for data cleaning & tidying", "Functions, because many of the plotting code is repettitive", "", "The concep
t of interactive pages and copying code but knowing what to change", "how to plot maps, how to make tables"),
)

knitr::kable(concepts, col.names = gsub("[.]", " ", names(concepts)))
```

weekNumber	conceptTaught
1	knitr::include_graphics()
2	Manipulation of csv data
3	Variable manipulation
4	The concept of filtering! Especially for data cleaning & tidying
5	Functions, because many of the plotting code is repettitive
6	
7	The concept of interactive pages and copying code but knowing what to change
	how to plot maps, how to make tables

Include the challenges and errors that you faced and how you overcame them.

```
knitr::include_graphics("includederrors.png")
```




```
#This keeps happening, despite me not having any files names Week-11-Diary-Entry. I overcame this by just ensuring the name
of my Rmd files do not have any spaces in them.


#The following are the placeholders i will put in the website
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.2     ✓ purrr     1.0.2
## ✓ forcats   1.0.0     ✓ readr     2.1.4
## ✓ ggplot2   3.4.4     ✓ stringr   1.5.0
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(maps)
```

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

```
library(sf)
```

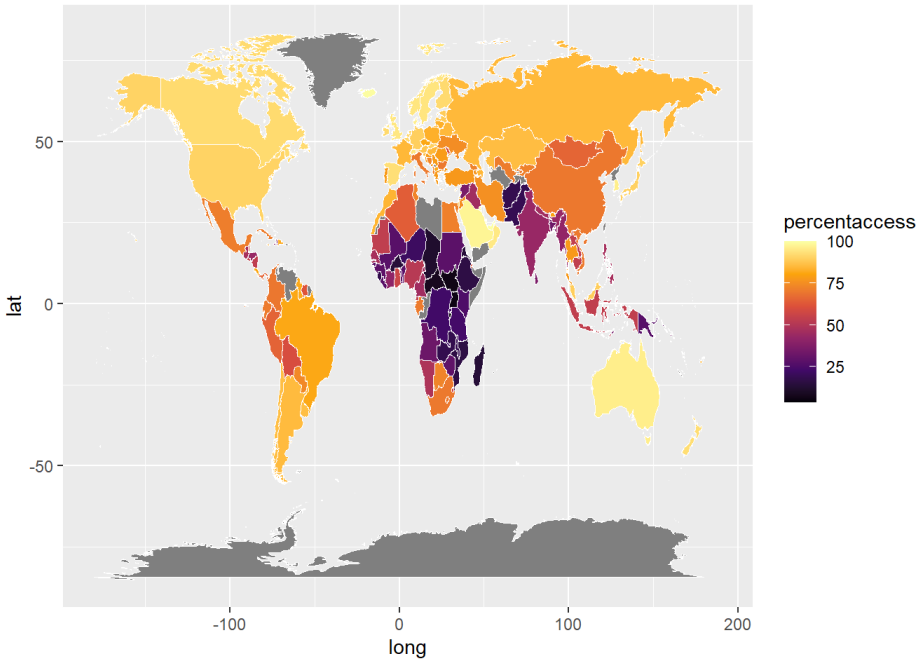
```
## Linking to GEOS 3.11.2, GDAL 3.7.2, PROJ 9.3.0; sf_use_s2() is TRUE
```

```
library(dplyr)
library(countrycode)

# Import the data with coordinates
world_map <- map_data("world")
world_map$region <- countrycode(world_map$region, "country.name", "iso3c")
```

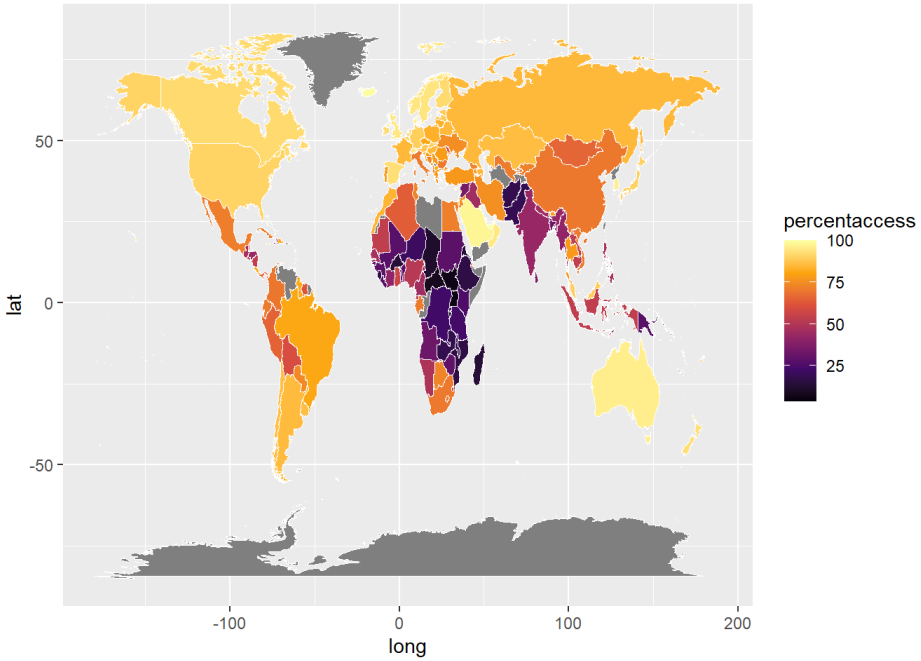
```
## Warning: Some values were not matched unambiguously: Ascension Island, Azores, Barbuda, Bonaire, Canary Islands, Chagos A
rchipelago, Grenadines, Heard Island, Kosovo, Madeira Islands, Micronesia, Saba, Saint Martin, Siachen Glacier, Sint Eustati
us, Virgin Islands
```

```
accesstonet <- read.csv("accesstointernet.csv")
netspeed <- read.csv("worldwideInternetSpeed.csv") #ISO2 code used in dataset!
netprice <- read.csv("worldwideMobileDataPricing2019to2023.csv")

#TO MAKE A FUNCTION FOR THE SWEET BROWNIE POINTS
#data cleaning for access to net
imstupid <- accesstonet %>%
   pivot_longer(
    cols = X1990:X2022,
    names_to = "Xyear",
    values_to = "percentaccess") %>%
  rename(region = Country.Code)

#data cleaning for speed of net
imhungry <- netspeed %>%
   pivot_longer(
    cols = D2017:D2023,
    names_to = "Dyear",
    values_to = "meandownloadspeed") %>%
  rename(region = Country.Code)

#data cleaning for price of net
imtired <- netspeed %>% #im ok i swear its just 2.30am
   pivot_longer(
    cols = D2017:D2023,
    names_to = "Dyear",
    values_to = "meandownloadspeed") %>%
  rename(region = Country.Code)

#do i get brownie points if i make that a function,, sigh more work for me ig

yearnetspecs <- imstupid %>% filter(Xyear == "X2020")
sotired_map <- left_join(world_map, yearnetspecs, by = "region")

#group = group connects the points in the correct order
ggplot(sotired_map, aes(long, lat, group = group))+
  geom_polygon(aes(fill = percentaccess), color = "white", size = 0.03)+
  scale_fill_viridis_c(option = "B")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

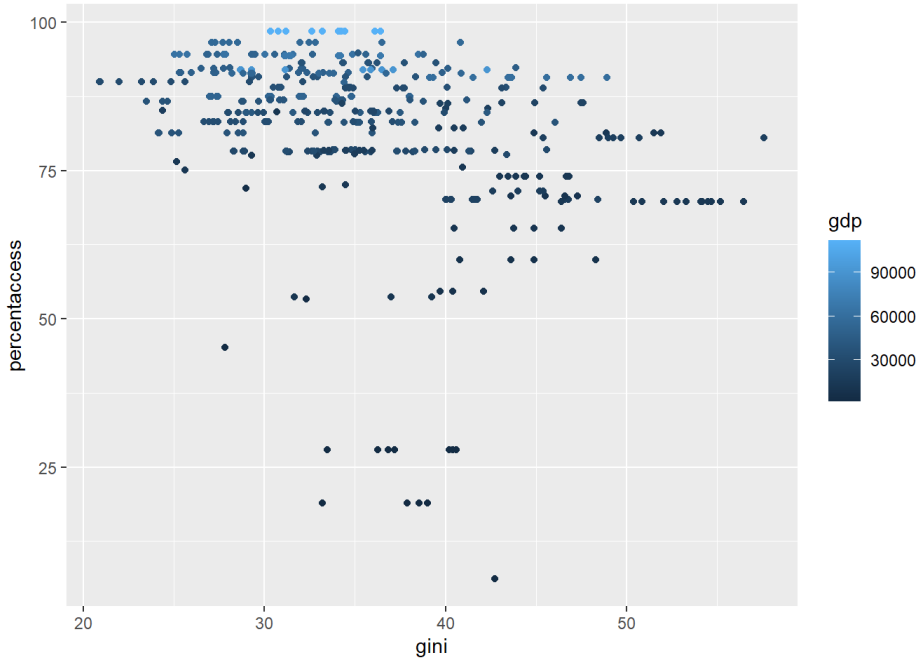



```
ggplot(sotired_map, aes(long, lat, group = group))+
  geom_polygon(aes(fill = percentaccess), color = "white", size = 0.03)+
  scale_fill_viridis_c(option = "B")
```




```
econineq <- read.csv("EconomicInequality.csv")
GINIvaluesyear <- econineq %>%
  filter(year == 2020) %>%
  #select(gini, c3) %>%
  rename(region = c3)

datascatter <- left_join(GINIvaluesyear, yearnetspecs, by = "region")

datascatter %>%
        drop_na(percentaccess) %>%
  ggplot(data = datascatter, mapping = aes(x = gini, y = percentaccess, colour = gdp)) +
  geom_point()
```

```
## Warning: Removed 3 rows containing missing values (`geom_point()`).
```


100
75
50
25
percentaccess
20
30
40
50
gini
gdp
90000
60000
30000