Assignment 2

Group 2

2024-02-18

Question 1

Our group discussed that we need to expand the LAT data set with multiple locations. This is because, once we done that,the unite of analysis in the cleaned data set would be per labor action and per location. Each row will only contain one geographical location. In this case, we can build correlations to find which country may have higher possibilities to have a certain type of labor action. Moreover, in this case, we can combine geo-econ data into each row, such as GDP, working population etc. according to each location. If we did not separate the locations, some rows may have multiple locations and also more than one geo-econ data.

```
LAT <- readxl::read_xlsx(here("Data Raw/Labor action tracker data 12.4.23.xlsx"))

data <- LAT %>%
  mutate(
    coordinate = ifelse(
      `Number of Locations` > 1,
      strsplit(as.character(`Latitude, Longitude`), ";\\s*"),
      `Latitude, Longitude`
    ))%>%
  unnest(coordinate)
```

Question 2

You can add options to executable code like this

```
#stack overflow suggestion
latlong2county <- function(pointsDF) {
    # Prepare SpatialPolygons object with one SpatialPolygon
    # per county
```

```
    counties <- map('county', fill=TRUE, col="transparent", plot=FALSE)
    IDs <- sapply(strsplit(counties$names, ":"), function(x) x[1])
    counties_sp <- map2SpatialPolygons(counties, IDs=IDs,
                     proj4string=CRS("+proj=longlat +datum=WGS84"))

    # Convert pointsDF to a SpatialPoints object
    pointsSP <- SpatialPoints(pointsDF,
                    proj4string=CRS("+proj=longlat +datum=WGS84"))

    # Use 'over' to get _indices_ of the Polygons object containing each point
    indices <- over(pointsSP, counties_sp)

    # Return the county names of the Polygons object containing each point
    countyNames <- sapply(counties_sp@polygons, function(x) x@ID)
    countyNames[indices]
}

data <- read_xlsx(here("Labs/Data Raw/Labor action tracker data 12.4.23.xlsx"))

#Clean Data using Junyi's Code
data_clean <- data %>%
  mutate(
    Address = strsplit(as.character(Address), ";\\s*"),
    City = strsplit(as.character(City), ";\\s*"),
    `Latitude, Longitude` = strsplit(as.character(`Latitude, Longitude`), ";\\s*")
  ) %>%
  unnest(Address, City, `Latitude, Longitude`)



data_latlong <- separate(data_clean, `Latitude, Longitude`, into = c("x", "y"), sep = ",",

data_latlong <- data.frame(x=data_latlong$x, y=data_latlong$y)


coordinates(data_latlong) <- c("x", "y")
proj4string(data_latlong) <- CRS("+proj=longlat +datum=WGS84")

latlong2county(data_latlong)
```

```r
#chatgpt suggestion
# Load required packages
library(rnaturalearth)
library(sf)
#devtools::install_github("ropensci/rnaturalearthhires")

# Load the US counties data
us_counties <- ne_states(country = "united states", returnclass = "sf")
# Sample DataFrame with latitude and longitude columns
df <- data.frame(Latitude = c(40.7128, 34.0522, 51.5074),
                 Longitude = c(-74.0060, -118.2437, -0.1278))

# Convert DataFrame to SpatialPointsDataFrame
coordinates(df) <- c("Longitude", "Latitude")
proj4string(df) <- CRS("+proj=longlat +datum=WGS84")

# Convert SpatialPointsDataFrame to sf object
df_sf <- st_as_sf(df)

# Set the projection to match with US counties data
st_crs(df_sf) <- st_crs(us_counties)

# Perform spatial join to find US counties for each point
result <- st_join(df_sf, us_counties)

print(result)
```

```r
#chatgpt suggestion

# Load required packages
library(tigris)
library(sf)

# Load US counties data
us_counties <- counties(cb = TRUE, resolution = "20m")

# Sample DataFrame with latitude and longitude columns
df <- data_latlong

# Convert DataFrame to SpatialPointsDataFrame
coordinates(df) <- c("Longitude", "Latitude")
```

```r
proj4string(df) <- CRS("+proj=longlat +datum=WGS84")

# Convert SpatialPointsDataFrame to sf object
df_sf <- st_as_sf(df)

# Set the projection to match with US counties data
st_crs(df_sf) <- st_crs(us_counties)

# Perform spatial join to find US counties for each point
result <- st_join(df_sf, us_counties)

print(result)
```

```r
county <- tigris::counties(cb = TRUE)
```

```
|
|                                                                      |   0%
|
|                                                                      |   1%
|
|=                                                                     |   1%
|
|=                                                                     |   2%
|
|==                                                                    |   2%
|
|===                                                                   |   4%
|
|====                                                                  |   6%
|
|=====                                                                 |   7%
|
|=====                                                                 |   8%
|
|======                                                                |   8%
|
|======                                                                |   9%
|
|=======                                                               |   9%
|
```

```
|=======                                                |  10%
|
|=======                                                |  11%
|
|========                                               |  11%
|
|========                                               |  12%
|
|=========                                              |  12%
|
|=========                                              |  13%
|
|==========                                             |  14%
|
|============                                           |  17%
|
|============                                           |  18%
|
|==============                                         |  18%
|
|==============                                         |  19%
|
|===============                                        |  19%
|
|===============                                        |  20%
|
|===============                                        |  21%
|
|================                                       |  21%
|
|================                                       |  22%
|
|=================                                      |  22%
|
|=================                                      |  23%
|
|=================                                      |  24%
|
|==================                                     |  24%
|
|==================                                     |  25%
|
|====================                                   |  25%
```

```
|
|==================                                                | 26%
|
|===================                                               | 26%
|
|===================                                               | 27%
|
|====================                                              | 28%
|
|====================                                              | 28%
|
|=====================                                             | 29%
|
|======================                                            | 29%
|
|======================                                            | 30%
|
|=======================                                           | 31%
|
|=======================                                           | 31%
|
|========================                                          | 32%
|
|========================                                          | 32%
|
|=========================                                         | 33%
|
|=========================                                         | 34%
|
|==========================                                        | 34%
|
|==========================                                        | 35%
|
|===========================                                       | 36%
|
|===========================                                       | 36%
|
|============================                                      | 37%
|
|============================                                      | 38%
|
|============================                                      | 38%
|
```

```
|=========================                                |  39%
|
|==========================                               |  39%
|
|==========================                               |  40%
|
|===========================                              |  41%
|
|===========================                              |  41%
|
|============================                             |  42%
|
|============================                             |  42%
|
|============================                             |  43%
|
|=============================                            |  44%
|
|=============================                            |  45%
|
|=============================                            |  45%
|
|==============================                           |  46%
|
|==============================                           |  47%
|
|==============================                           |  48%
|
|===============================                          |  48%
|
|===============================                          |  49%
|
|===============================                          |  49%
|
|================================                         |  50%
|
|================================                         |  51%
|
|================================                         |  51%
|
|=================================                        |  52%
|
|=================================                        |  52%
```

```
|
|=====================================                                 |  53%
|
|======================================                                |  54%
|
|=======================================                               |  54%
|
|========================================                              |  55%
|
|=========================================                             |  55%
|
|==========================================                            |  56%
|
|===========================================                           |  56%
|
|============================================                          |  57%
|
|=============================================                         |  58%
|
|==============================================                        |  58%
|
|===============================================                       |  59%
|
|================================================                      |  59%
|
|=================================================                     |  60%
|
|==================================================                    |  61%
|
|===================================================                   |  61%
|
|====================================================                  |  62%
|
|=====================================================                 |  62%
|
|======================================================                |  63%
|
|=======================================================               |  64%
|
|========================================================              |  64%
|
|=========================================================             |  65%
|
```

```
|=============================================                         |  65%
|
|=============================================                         |  66%
|
|==============================================                        |  67%
|
|==============================================                        |  68%
|
|==============================================                        |  68%
|
|==============================================                        |  69%
|
|===============================================                       |  69%
|
|===============================================                       |  70%
|
|===============================================                       |  71%
|
|================================================                      |  71%
|
|================================================                      |  72%
|
|================================================                      |  72%
|
|================================================                      |  73%
|
|=================================================                     |  74%
|
|=================================================                     |  75%
|
|=================================================                     |  75%
|
|=================================================                     |  76%
|
|==================================================                    |  77%
|
|==================================================                    |  78%
|
|==================================================                    |  78%
|
|==================================================                    |  79%
|
|===================================================                   |  79%
```

```
|
|=============================================================          |  80%
|
|==============================================================         |  81%
|
|===============================================================        |  81%
|
|================================================================       |  82%
|
|=================================================================      |  82%
|
|==================================================================     |  83%
|
|===================================================================    |  84%
|
|====================================================================   |  84%
|
|=====================================================================  |  85%
|
|====================================================================== |  85%
|
|======================================================================= |  86%
|
|========================================================================|  87%
|
|========================================================================|  88%
|
|========================================================================|  88%
|
|========================================================================|  89%
|
|========================================================================|  89%
|
|========================================================================|  90%
|
|========================================================================|  91%
|
|========================================================================|  91%
|
|========================================================================|  92%
|
|========================================================================|  92%
|
```

```
|=============================================================     |  93%
|
|==============================================================    |  94%
|
|==============================================================    |  94%
|
|===============================================================   |  95%
|
|================================================================  |  95%
|
|================================================================  |  96%
|
|================================================================= |  96%
|
|================================================================= |  97%
|
|================================================================== |  98%
|
|================================================================== |  98%
|
|=================================================================== |  99%
|
|====================================================================|  99%
|
|====================================================================| 100%
```

```
data <- separate(data, coordinate, into = c("lat", "lon"),
                 sep = ",\\s*", remove = FALSE)
data <- data %>%
  mutate(
    lat = as.numeric(lat),
    lon = as.numeric(lon))
data$lon[2830] = -85.73642799999999
na_summary <- data %>%
  summarise_all(~ sum(is.na(.)))

data <- st_as_sf(data, coords = c("lon", "lat"))
st_crs(data) <- st_crs(county)
data <- st_join(data, county)
vars <- names(LAT)
data <- data %>%
  select(c(vars, NAME, GEOID)) %>%
```

```
  rename(county = NAME)
```