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

```
library(cdlTools)
library(sp)
library(maps)
```

These are coding solutions generated by Chatgpt.

1. Stack overflow suggestion

```r
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_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)
```

2. Suggestion with package *rnaturalearth*.

```r
# Load required packages
library(rnaturalearth)
library(sf)

# 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)
```

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

3. Suggestion with *tigris* package. (Final version. Our group will be using this solution to generate county variable.)

```r
# retrieve data frame from tigris package that contains US counties infomation and geometr
county <- tigris::counties(cb = TRUE)
# the goal is to join 2 data frame with geometric attributes.
# but first, convert LAT data into geometric data frame with coordinates
# the transformation require separate longtitude and latitude information
data <- separate(data, coordinate, into = c("lat", "lon"),
                 sep = ",\\s*", remove = FALSE)
# check is there any NAs
data <- data %>%
  mutate(
    lat = as.numeric(lat),
    lon = as.numeric(lon))
na_summary <- data %>%
  summarise_all(~ sum(is.na(.)))
# yes, there is one in lon column.
data$lon[2830] = -85.73642799999999
# transform LAT data into geometric data frame
data <- st_as_sf(data, coords = c("lon", "lat"))
# matching same projection
st_crs(data) <- st_crs(county)
# joining 2 geometric data frames
data <- st_join(data, county)
# housekeeping and make LAT data looks clean and organized
vars <- names(LAT)
data <- data %>%
  select(c(vars, NAME, GEOID)) %>%
  rename(county = NAME)
```

In general, Chatgpt is helpful in terms of offering basic ideas of how we can acheive it. It shows that in order to generate a variable that contains US county names with Lon/Lat data, the basic idea is to match and join the LAT data with a data frame that has US county information. Without specific instruction, however, Chatgpt would generate complicated solution with redundant codes and rarely used packages. Because our group has team members who took QGIS prior to the course and are familiar with *sf* package. In this case, with the basic ideas and reference of packages that can retrieve US county information from Chatgpt, It

would be more convenient and helpful than just simply relying on Chatgpt to generate the solution from scratch.

Question 3

```
knitr::kable(head(data))
```

Timestamp	Employer	Local	Industry	Number of Stations	Address	City	State	Zip Code	Latitude, Longitude	Approximate Number of Participants	Start Date	End Date	Duration	Amount	Unit	Worker Demonstration	Strike Aesthetic	Source	Notes	county	GEOID	Geometry
2022-10-22 10:55:33	Johnson Controls	United Auto Workers (UAW)	1872	Manufacturing	NA	631 S. Richland Avenue	York	PA	17403	39.94923, -76.74407	2022-09-28	NA	NA	NA	NA	Strike	Pay, Paid Time Off (PTO)	https://www.yorkcounty.gov/news/2020/1-controls-workers-strike-pay-increase-vacation-flexibility-york-county-hopewell/6263404002/; https://www.bloomberglaw.com/product/2dd1-4a43-bc1b-1bf2fda081dd	NA	42133	42133	POINT (-76.74407 39.94923)
2022-10-22 10:55:46	Bannister House	Service Employees International Union (SEIU)	1199 NE	Healthcare and Social Assistance	1451	135 Dodge Street	Providence	RI	02905	41.81363, -71.42516	2022-10-23	2022-10-23	2	2	Days	Strike	Pay, COVID-19 protocols, Health and safety	https://www.fnca.org/wp-content/uploads/2020/11/October-2020-Work-Stoppages.pdf; https://www.golocalprov.com/news/Battle-at-Bannister-Frontline-Workers-Strike-at-Nursing-Home	NA	44007	44007	POINT (-71.42516 41.81363)

Time	Labor Organization	Establishment Type	Local	Industry	Number of Bargaining Unit Workers	Address	City	State	Zip Code	Approximate Number of Participants	Start Date	End Date	Duration	Strike Action	Worker Demands	Source	Notes	Longitude	Latitude	GEOID
2022-10-22 10:55:56	Allina Abbott Northwestern Hospital	Service Employees International Union (SEIU)	SEIU Healthcare Minnesota	Health Care and Social Assistance	1011	800 Minnesota E. 28th Street	Minneapolis	MN	55407	2500	2022-10-10	2022-10-16	7 Days	Strike	Pay, COVID-19 protocols, Health and safety	https://www.fmcs.gov/wp-content/uploads/2020/11/October-2020-Work-Stoppages.pdf; https://www.bizjournals.com/twincities/news/2022/10/11/allina-workers-on-strike-abbott-st-francis.html	NAICS 622110	-93.26202	44.95256	GEOINT (-93.26202 44.95256)
2022-10-22 10:56:04	Youngstown State University	Education Association - NEA	Youngstown Education Association	National Service Unions	337	Youngstown State University Plaza	Youngstown	OH	44555	1074	2022-10-10	2022-10-16	7 Days	Strike	Pay, Sharecropper, IP rights	https://www.fmcs.gov/wp-content/uploads/2020/11/October-2020-Work-Stoppages.pdf; https://www.wkbn.com/news/local-news/ysu-and-faculty-union-reach-an-understanding-strike-is-over/; https://www.mahoningmatters.com/local-news/ysu-oea-membership-votes-to-approve-fact-finders-report-2761974	NAICS 611310	-80.64817	41.10741	GEOINT (-80.64817 41.10741)

Timestamp	Employer	Labor Organization	Local	Industry	Bargaining Unit Size	Number of Locations	Address	City	State	Zip Code	Latitude	Longitude	Approximate Number of Participants	Start Date	End Date	Duration Amount	Duration Unit	Strike or Protest	Authorized	Worker Demands	Source	Notes	county	GEOID	geometry
2021-10-22 11:31:27	Gahanna-Jefferson Public School District	Ohio Education Association - NEA	Gahanna-Jefferson Education Association	Educational Services	572	1	160 S. Hamilton Road	Gahanna	Ohio	43230	39.97246	-82.874563	572	2020-10-16	2020-10-18	[illegible]	Days	Strike	Y	COVID-19 protocols, Health and safety	https://www.dispatch.com/story/news/ed… contract-ended-gahanna-jefferson-teachers-union-strike-released/3710537001/; https://www.fmcs.gov/wp-content/uploads/2020/11/October-2020-Work-Stoppages.pdf	NA	Franklin	39049	POINT (-82.87456 39.97247)
2021-10-22 11:31:14	AstenJohnson	International Association of Machinists (IAM)	District 10	Manufacturing	87	1	6480 W. College Avenue	Appleton	Wisconsin	54914	44.26392	-88.50545	87	2021-10-15	2021-10-26	11	Days	Strike	Y	Healthcare	https://www.bloomberglaw.com/product/… 3a59-44d3-a468-156c6439abd1; https://www.postcrescent.com/story/mon… machinist-union-strike-settled-astenjohnson-greenville/6040228002/	NA	Outagamie	55087	POINT (-88.50545 44.26392)