Retail Per Tract

Tzu-Hsuan (Jessica) Lin, Koller Adzick, Katherine Li, Allen Hu

2020/11/20

```r
library(tidyverse)
library(dplyr)
library(rgdal)
library(readxl)
```

Read Shape File

```r
hennepin = readOGR(dsn = "map_data", layer = "2010_Census_Tracts")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "D:\UMN\06_6130 Introduction to Business Analytics in R\Hennepin County Live Case\Wrap up\map
## with 299 features
## It has 60 fields
```

Read Retail Data

Read in the retails store data and select only those in Hennepin, MN.

```r
retail_US <- read_csv("SNAP_Store_Locations.csv")
retail_US$Zip5 <- as.numeric(retail_US$Zip5)
retail <- retail_US %>%
  filter(State == 'MN', County == 'HENNEPIN')
```

Convert to spdf

Since retail data uses WSG84 projection as location (the common seen latitude, longitude) and the Hennepin tract geometry uses UTM zone 15 (epsg: 32615) to record the exact location, it is necessary to convert the two into the same CRS (Coordinate Reference System).

```r
# concept reference: https://epsg.io/32615
# concept reference: https://datacarpentry.org/organization-geospatial/03-crs/

# code reference
# http://rstudio-pubs-static.s3.amazonaws.com/19879_7e13ab80d5ed416c8e235bd6bb93cf3e.html

cord.dec <- SpatialPoints(retail[, c("Longitude", "Latitude")],
```

```
                        proj4string = CRS("+proj=longlat"))

lnglat <- spTransform(cord.dec, hennepin@proj4string@projargs)

spdf_retail <- SpatialPointsDataFrame(lnglat, retail)
```

Count stores per tract

To loop over all the tracts in Hennepin county, we used `over()` to calculate the store count for each tract.

Count retailers for each tract

```
tracts <- as.vector(hennepin@data[["GEOID10"]])
tracts <- sort(tracts)

count_vec <- c()

for (t in tracts){
  single_tract <- hennepin[which(hennepin@data$GEOID10 == t), ]

  retail_logical <- over(spdf_retail, single_tract)
  count <- nrow(retail_logical[complete.cases(retail_logical[ , 5:6]),])
  count_vec = c(count_vec, count)
}
```

Convert to df

```
df <- as.data.frame(cbind(tracts, count_vec))
df<- df %>%
  rename(
    tract = tracts,
    count = count_vec)
```

Merge to main dataframe with all other attributes

```
data <- read_csv("2010_Census_Tracts.csv")
data <- data %>%
  select(c("GEOID10", 'NAME10', 'ALAND10', 'TRACT', 'POP_TOTAL', 'POP1_RACE',
           'POP1_WHT', 'POP1_BLK', 'POP1_AMIND', 'POP1_ASN', 'POP1_HAWPA',
           'POP1_OTHR', 'POP2_MORE', 'POP_HSPLAT', 'POP_NO_HSP', 'POPWHNOHSP',
           'TOT_HU', 'TOT_OCCHU', 'TOT_VACHU'))
data$GEOID10 <- as.character(data$GEOID10)
data <- data %>%
  left_join(df, by=c('GEOID10' ='tract'))
```

Store per people (total)

```
data['count'] <- as.numeric(unlist(data['count']))
data['store_per_person'] <- data['count'] / data['POP_TOTAL']
```

Store per people (SNAP)

```
snap <- read_excel('SNAP Summary data for LiveCase fall 2020.xlsx')
snap$tract <- as.character(snap$tract)
snap <- snap %>%
  filter(elig_month >= 2019) %>%
  group_by(tract) %>%
  summarise(sum_snap = sum(people))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
data <- data %>%
  left_join(snap, by=c('GEOID10' = 'tract'))

data['store_per_snap_person'] = data['count'] / data['sum_snap']
```