Geospatial Analysis in R

Kim Kreiss

Introduction

Brief re-intro to spatial data and mapping

Kernel Density Estimation

Global and Local Moran's I

Getis-Ord / G*

Wrap-up

Introduction

Why R?

- Reproducibility and Transparency
- Flexibility and Customization
- Integration with Statistical Analysis
- Geospatial Visualization

Today

- We will analyze crime data in Chicago together:
  - Combine point data and boundary file
  - Generate informative visualizations
  - Apply statistical techniques to identify hot and cold spots
- This is an interactive workshop, where you are provided with code and data to run locally:
  - `01-geospatial-workshop.Rmd`
  - `2016-2020-chicago-crime.csv`
  - `Boundaries - Neighborhoods.geojson`
  - `Boundaries - Census Tracts - 2010.geojson`

Brief re-intro to spatial data and mapping

Spatial Data

- Most commonly point data, boundary data, grid or raster data
  - point data: longitude and latitude of a unit of interest
  - boundary data: polygon/shape data delineating a geographical area
  - grid or raster data: usually squares/rectangles over the Earth's surface with specific attributes
- Uses Coordinate Reference System (CRS)

Mapping

- Common application is to show point data within a region or geographic boundaries of interest, ie, are homicides in Chicago concentrated in certain neighborhoods?

```
# Load required libraries
library(sf)
library(spdep)
library(spatstat)
library(tidyverse)
library(colorspace)
```

```r
path <- here::here()
crime_data <- read.csv(paste0(path,"/data/2016-2020-chicago-crime.csv"))

neighborhoods <- st_read(
  paste0(path,
         "/data/Boundaries - Neighborhoods.geojson"),
  quiet=T)

census_tracts <- st_read(
  paste0(path,
         "/data/Boundaries - Census Tracts - 2010.geojson"),
  quiet=T)

homicide <- crime_data %>%
  filter(Primary.Type=="HOMICIDE")

### double check CRS and refernece
```

Kernel Density Estimation

Global and Local Moran's I

Getis-Ord / G*

Wrap-up

Wrap-up

```
path <- here::here()
crime_data <- read.csv(paste0(path, "/data/2016-2020-chicag

neighborhoods = st_read(paste0(path,"/data/Boundaries - Nei
#neighborhoods = st_read(paste0(path,"/data/Boundaries - Co

homicide <- crime_data %>%
  filter(Primary.Type=="HOMICIDE")

### double check CRS and refernece
```

```
chicago_map <- neighborhoods %>%
  ggplot() +
  geom_point(data = homicide, aes(x=Longitude, y=Latitude),
  geom_sf(aes(geometry=geometry), fill="transparent", linew
  theme(axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        rect = element_blank(),
```

Next steps:

- descriptions/slides for each type of analysis
- consider using different geographical boundaries, ie census tracts, highlight modifiable area unit problem + parameters for different methods
- resources on spatial regression, future directions, etc.
- bigger data/computation: plug for Eric's workshop!!!

package dependencies

some other demographic/neighborhood characteristics that are associated with higher crimes, etc.

add in population data for the neighborhoods and census tracts

if you have other high frequency data, ie cell phone data or all crime data as benchmark for computation time spatial heterogeneity ??

source: https://michaelminn.net/tutorials/r-crime/