

Geog 4/6300: Lab 2

Spatial statistics and probability distributions

Due: Wednesday, Oct. 4

Value: 30 points

Overview: This lab covers two main topics: basic spatial statistics and probability distributions. Your answers to the lab questions should be typed in the response template and turned in using the Assignment Dropbox on the ELC site.

Part 1: mapping the data and its distribution

The lab folder on Google Drive has three files related to crime data for Spokane, Washington.

- “Spokane_crimes_all” has data for every crime committed in from 2000-2015.
- “SpokanePrecincts_data” has precinct boundaries, crime counts, the 2010 population of blocks in those precincts. Note: this is saved as a geojson file, which is similar to a shapefile but more compact.

You can load each file using the commands below, using `st_as_sf` to convert the crimes to spatial data

```
library(tidyverse)
library(sf)
spokane_crimes_all<-read_csv("https://github.com/jshannon75/geog4300/raw/master/Labs/Lab%20_%20Point%20Data/spokane_crimes_all.geojson")
spokane_crimes_sf<-st_as_sf(spokane_crimes_all,coords=c("Long","Lat"),crs=4326)

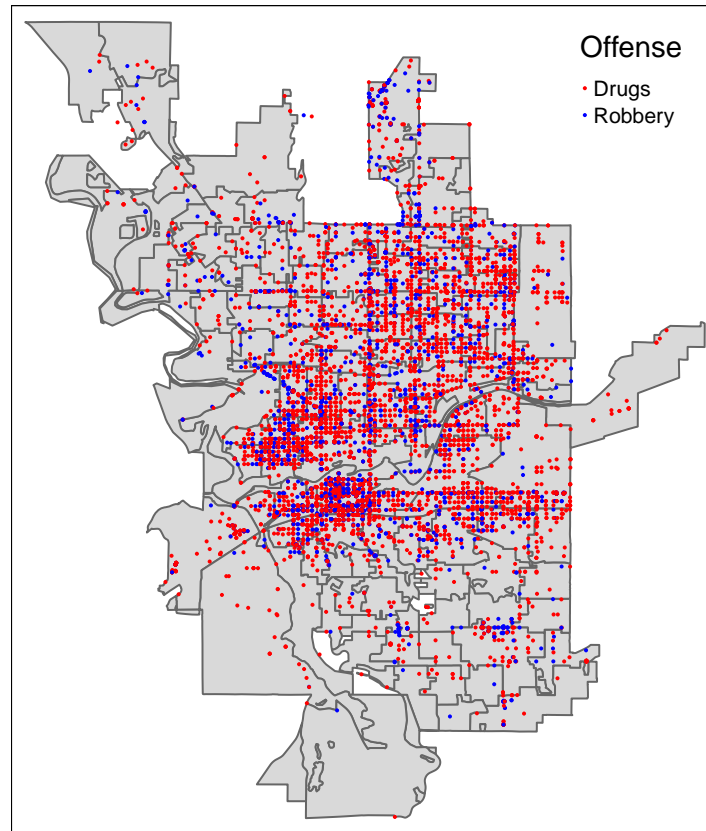
spokaneprecincts_data<-st_read("https://github.com/jshannon75/geog4300/raw/master/Labs/Lab%20_%20Point%20Data/spokaneprecincts_data.geojson")

## Reading layer `SpokanePrecincts_data' from data source `https://github.com/jshannon75/geog4300/raw/master/Labs/Lab%20_%20Point%20Data/spokaneprecincts_data.geojson'
## Simple feature collection with 117 features and 14 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -117.5215 ymin: 47.5871 xmax: -117.3039 ymax: 47.7588
## epsg (SRID):    4326
## proj4string:     +proj=longlat +datum=WGS84 +no_defs
```

Question 1 (3 points): *Using the `Offense` variable and working with the `spokane_crimes_all` data, select just murder and arson as crimes. Using `tmap`, create a map that shows the location of all reports in the this data frame and the precinct boundaries. See script 6-2 for a model to work from or the example for below.*

```
library(tmap)
tmap_mode("plot")
sample_crimes<-spokane_crimes_sf %>% filter(Offense=="Drugs" | Offense=="Robbery")

tm_shape(spokaneprecincts_data)+
  tm_polygons()+
tm_shape(sample_crimes) +
  tm_dots("Offense",palette=c("red","blue"))
```



Question 2 (2 points): Create a new data frame with the mean centers for each type of crime listed in the *spokane_crime_sf* data frame.

Question 3 (3 points): Map the mean centers you created in question 2 using *tmap*. To do so, you'll need to use *st_as_sf* to convert your new data frame to *sf* format. Use *tm_dots* to show the points in different colors, as in the example above (though here you're showing only the mean centers).

Question 4 (2 points) Interpret how geographic differences between the mean centers you calculated are meaningful. What do they tell you about the locations of these crimes?

Question 5 (2 points): The nearest neighbor index for vehicle thefts in this dataset is 0.44 compared to 1.51 for murders. Explain what these two numbers tell you about the distribution of these two point datasets.

Part 2: Crime rates

The precinct data frame has the number of crimes in each precinct. Crimes are commonly reported as crimes/100,000, so you should calculate the rate of any of these crimes rather than showing the raw counts. For this lab, we will focus on reported *burglary*. Calculate the assault rate by creating a new variable in the data, dividing the reported crimes by population and then multiplying by 100,000:

```
[new variable] <- [Burglary variable] / [population variable] * 100000
```

Use the *mutate* command from the *tidyverse* to create this variable.

Question 6 (5 points) Select just the precinct number and burglary rate from the *precincts* data frame. Then create a new variable for this rate using the formula above. Then do the following: 1) Call your dataset once you've created it to show your results. 2) Create a histogram using *ggplot* showing the distribution of these data. 3) Map the rates using *tmap*.

Question 7 (2 points) The burglary rate for the city of Spokane in this dataset is 10,639 per 100,000 residents. Using this figure, calculate the location quotient for burglary in each precinct within the city.

Quadrat analysis

You can also use the spatstat package to create a ppp object for your robbery data.

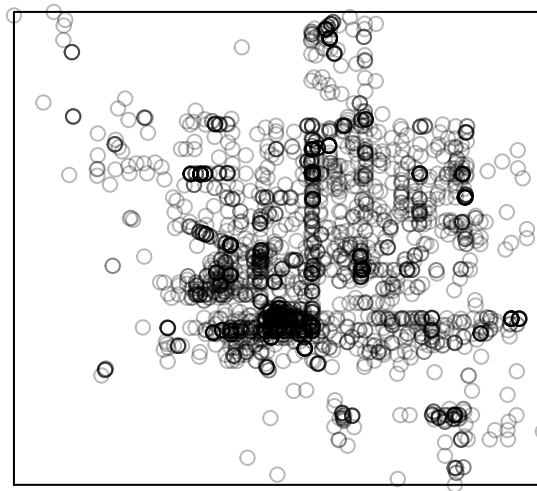
```
library(spatstat)
robberies<-spokane_crimes_all %>% filter(Offense=="Robbery")

long_bounds<-c(min(robberies$Long),max(robberies$Long))
lat_bounds<-c(min(robberies$Lat),max(robberies$Lat))

robberypoints<-ppp(robberies$Long,robberies$Lat,long_bounds,lat_bounds)

## Warning: data contain duplicated points
plot(robberypoints)
```

robberypoints



Question 8 (3 points) Using the script from week 7 in class as a guide, combine the kernel density map and appropriately sized quadrat grid for robberies in Spokane during the study period. You'll need to include "add=TRUE" for the quadratcount function to layer it and change the text color to white as we did in class.

Question 9 (4 points) Based on your analysis in the questions 6 through 8, describe the statistical and geographic distribution of burglary in the city of Spokane in a paragraph. Specifically address any skew in this data and the geographic location of the highest and lowest values for the burglary rate. Which of these visualizations was most helpful to you and why?

Probability

The number of persons in each camping party that reserves a campsite at Dawgsville National Park is assumed to be distributed as a Poisson distribution. The maximum number of persons in a camping party is 6. The mean number of persons in a camping party is 2.43.

Question 10 (2 points): *Calculate the probability of observing each possible count of persons in a camping party for each reservation (including no shows).*

For the past 100 years, a stream close to Athens has been measured at a gauging station. The station measures the volume passing a point in a minute, so the values obtained are in units of cubic feet per minute (or CFM). A USGS scientist has crunched these numbers and determined that the results are normally distributed, with a mean of 35.26 CFM and a standard deviation of 4.61 CFM

Question 11 (2 points): *Calculate the probabilities the stream flow in a given year will exceed 42 and 48 CFM.*