GIS-Clinical Collaboration

Martin Gleason

2/20/2019

The Ask

DCPO Halawa and Data Analyst Loftus asked if it would be possible to begin mapping shooting data to locations within Cook County. Furthemore, both wanted to know if it is possible to utilize Cook County Medical Examiner Office information in addition to the current sources. Both asks are relatively easy tasks and can be done as either a reproducible report or an app/dashboard with R and RMarkdown.. All of the data analysis, including the current tables, forms, and charts, can be moved into R for analysis.

The Lift

Even if the requirements were limitd to do making maps, the following steps would be a part of the reporting:

- Importing the data
- Cleaning the data
- Visualizing the data

Importing the Data

Importing involes either: a) loading a spreadsheet into R or b) accessing an online database with R.

Either set up allows for mapping as long as the data contains an address or GPS coordinates. This document uses data from both methods.

Cleaning Data

Cleaning data is more intensive and requires more subject matter expertise. First, the data must be organized in the following manner:

- Columns are variables
- Rows are observations
- Each cell (intersection of column and row) contains one element
- One group of data per table

Put differently, when the department puts data in a table: What is tracked are columns, what is observed is a row, each cell has one observation, and each table is revolves around one kind of metric or source. This is referred to as tidy data.

Once the data are tidy, missing values are addressed (per business rules), column data types (words, numbers, dates, etc) and order are set, and calculated variables, such as age or risk-level scoring, are created. The tidy format allows for quick and easy analysis of data.

Visiualzing the Data

For example, the 2017 Data set has the following 36 columns: *incident_date, first_name, last_name, dob, type, police_district_arrest, gender, race, ssl, finding, gang, calendar, dcpo, spo, po, home_address, zipcode, police_district_residence, arrests_prior, number_petitions, petitions_with_findings, offense_type,*

jjor_findings, pending_charges, ips_order, IDJJ_bring_back, idjj_commitment, rmis_JTDC_holds_bedstay, rmis_bed_days, total_probation_orders, technical_vops, vop_findings, initial_yasi_risk, number_interventions, interventions, year.

Displaying all of this data in a single column is not useful. However, it can be approached in the following ways:

2017 Dataset: Column Selection

incident_date

dob

type

gender

race

2017-11-19

1999-09-20

SHOOTING

M

BLK

2017-11-26

1999-08-20

HOMICIDE

M

BLK

2017-10-11

1999-06-06

SHOOTING

M

BLK

2017-01-27

1999-03-14

SHOOTING

M

BLK

2017-06-01

1999-03-14

SHOOTING

M

BLK

2017-09-07

1999-09-03

HOMICIDE

M

BLK

2017-11-10

2000-05-12

SHOOTING

M

BLK

2017-10-17

2002-04-20

SHOOTING

M

BLK

2017-04-05

1998-10-23

SHOOTING

M

BLK

2017-03-19

1999-07-20

SHOOTING

M

BLK

2017 Shooting Data Summarized

type

race

gender

Total

HOMICIDE

BLK

F

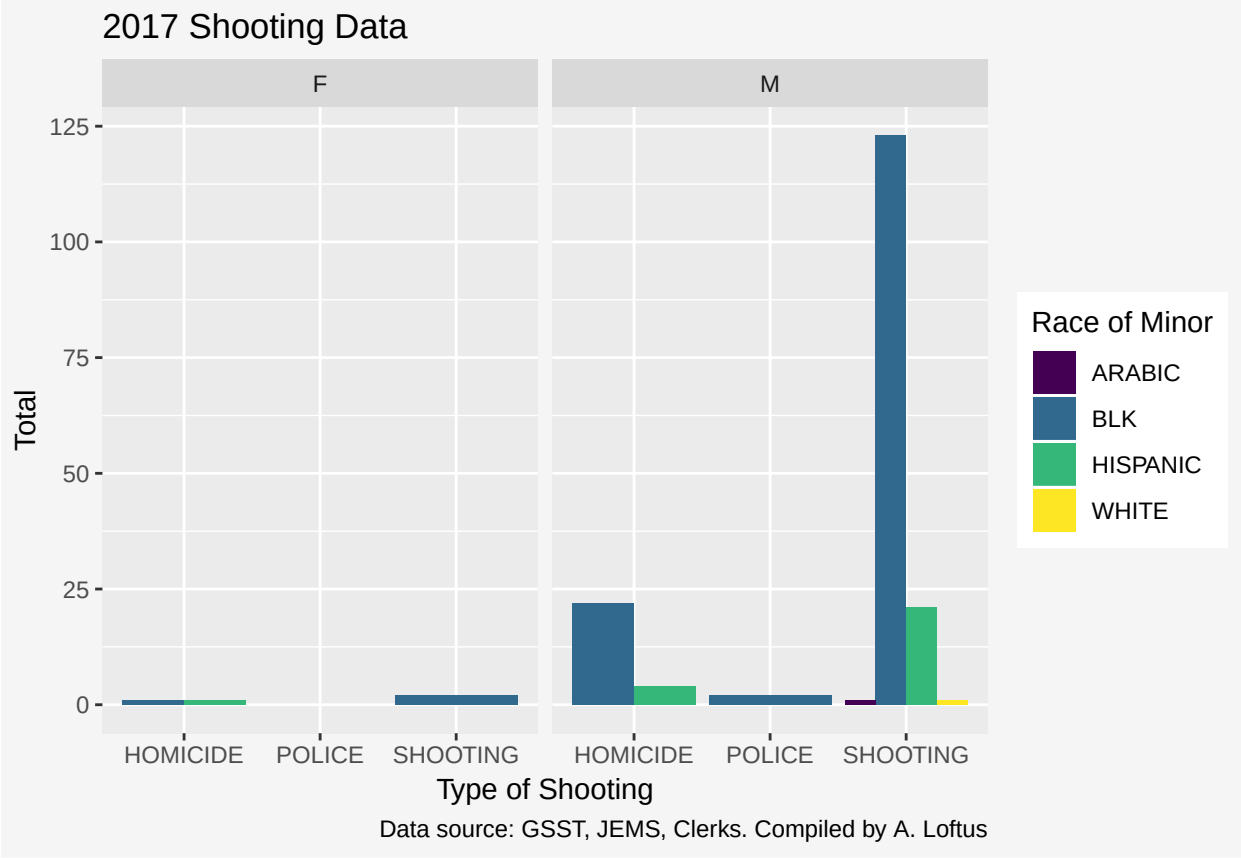
1

HOMICIDE

BLK

M

22
HOMICIDE
HISPANIC
F
1
HOMICIDE
HISPANIC
M
4
POLICE
BLK
M
2
SHOOTING
ARABIC
M
1
SHOOTING
BLK
F
2
SHOOTING
BLK
M
123
SHOOTING
HISPANIC
M
21
SHOOTING
WHITE
M
1



With additional time, the 2018 dataset and the diversion data set can be tidied into a similar "shape" which would allow for analyzing yearly trends in the same graph.

GIS

If a table does not include coordinates, then after the data is tidyed, addresses need to be geocoded. This is a simple process that requires either access to Google Maps or another GIS provider. After it is geocoded, the addresses are represented as points on the map. GIS also allows for the drawing of polygons – features of a map that are of particular interest (such as neighborhood boundaries, forest preservce districts, etc.)