Yigong Wang (Ryan)
GIS 3 Data Report
Dr. Marynia Kolak
Due May 24, 2020

Data Report - The Chicagoland Toolkit

Introduction
This project aims to build upon a R Package in development by Clyde Schwarb[1] from the previous GIS 3 class. Clyde's package contains many aspects of data for the city of Chicago and beyond.

Part I - The R Package
The majority of the datasets are from the Chicago Data Portal. The ideal situation would be to include as many of the datasets in the portal as possible, but this will be a long-term development task. For the sake of this project, I will include the following datasets:

Improving the Data Source
Clyde has downloaded all the data and pulls all the data locally within the package. This is fine for the current stage of the package, but once the package is developed further to include more datasets from the portal, API access needs to be implemented to control for the size of the package. Using API also has the benefit of being able to get up-to-date data. Unless specified, the data cited in this section are all current or of the current iteration.
Datasets to be included in the package include (All accessed through SODA API):

- Census Tracts 2010 (soon to be updated to 2020): Accessed here as a `geojson` file;
- Boundaries - Community Areas (current): Accessed here as a `geojson` file;
- Boundaries - Police Beats (current): Accessed here as a `geojson` file;
- Boundaries - Police Districts (current): Accessed here as a `geojson` file;
- Boundaries - Wards (2015-): Accessed here as a `geojson` file;
- Boundaries - Zoning Districts (current): Accessed here as a `geojson` file;
- Affordable Rental Housing Developments (points) (2019): Accessed here as a `json` file;
- Chicago Public Schools - School Progress Reports SY1617 (points, with information regarding community areas): Accessed here as a `json` file;
- CTA - Ridership - Monthly Day-Type Averages & Totals: Accessed here as a `json` file;
- CTA - 'L' (Rail) Stations: Accessed here as a `zip` file (points, with community area information) (will have to figure out a way to unzip and manage this dataset, possibly a new R script);
- Census Data - Selected socioeconomic indicators in Chicago, 2008 – 2012 (by community area): Accessed here as a `json` file. This dataset is outdated, but is the only one that shows community-area level data. It is possible to use CPS data directly from census bureau or IPUMS, but they are not as detailed to the community area level;

---

[1] I have emailed Clyde Schwarb regarding the Chicago data package, but he has not responded to me just yet. This data report is based on the information I gathered from the existing package.

- Divvy Bicycle Stations (points, with community area info): Accessed here as a `json` file;
- Tax Increment Financing (TIF) Projects (points): Accessed here as a `geojson` file;
- 311 Service Requests - Vacant and Abandoned Buildings (points, with community area information): Accessed here as a `geojson` file;

The above datasets are all original sets included in Clyde's Package, and I linked online API locations, and will update in the code. I also included an additional dataset below which I think is quite important:

- Crimes - 2001 to Present (points, with temporal information, can aggregate by community area, police wards, beats): Accessed here as a `json` file;
- Divvy Trips (points, with beginning and end point info): Accessed here as a `json` file;

More datasets will be added after the completion of the first stage of project.

Documentation

One of the other things I wish to do to improve on the data package is to write up more documentation and vignettes for this package. Information on how to access each dataset and its respective variables, and how to do some basic visualization in R will be provided.

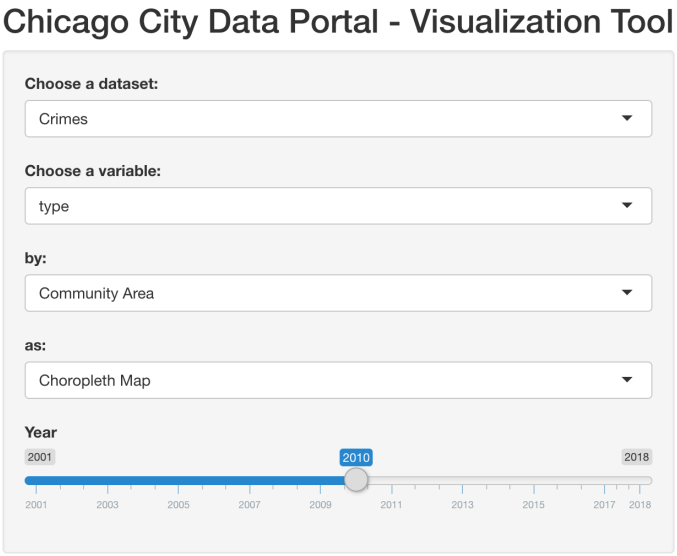
Functionalities

Clyde has done a great job consolidating all the data and did some R data wrangling. I plan to update the base function with all online sources.

Since I plan to include even more datasets in this project after the initial build, I will also include a R-package/City Portal crosswalk codebook, namely a document detailing what each dataset includes and how to find the R dataset name corresponding to the dataset on the data portal.

Part II - The RShiny App

The RShiny app is meant to be included as a demonstration and template for potential users of the package to see what visualizations are possible, and to provide an introductory ESDA tool for further research. Clyde has managed to create an app that has some basic information about CPS tiers and CPS Attendance. I plan to expand the app a lot more, including every single dataset that I have included above.



The sidebar will contain three items: dataset, variable, and visualize by & as. The final result would look something like the image shown on the right. The "Crimes" dataset would contain one extra option as the dataset does include data from 2001 to 2018. I would then build a "Year" functionality to display that information as well.

References

Chicago Data Portal. (2020). Retrieved 21 May 2020, from https://data.cityofchicago.org.

Schwarb, C. (2020). Chicago Data Package. GitHub. Retrieved 21 May 2020, from https://github.com/cschwab1/chicagodatapackage.