

# GIS 3 Lab 3

Ryan (Yigong) Wang

Work through the “Linking R to the Web (Links to an external site.)” lab by Singleton et al 2018, from their Urban Analytics text. The full lab is available as an Rmarkdown file or html file. Go through as much as you can.

Using what you’ve learned, generate an Rmarkdown file that uses an online dataset. Using data from the lab is fine, or you can find your own data. Load, extract, clean, and plot the data as a well-designed map. The file you generate should highlight key features throughout this process, and end with a carefully made map. It should be (at least slightly) different from the plots generated in the lab. Render your file as a pdf and/or html, publish on your Github, and link here.

Note:

following the premise of the lab, the data you use **should be linked directly from the Web** not loaded in as a CSV file from your system if you develop an html file, please include the exact link to the html file on your Github. It makes it slightly easier for grading purposes. Thanks!

This lab will practice skills learned in Chapters 5. In this lab, I will use a spatial dataset directly linked from the web to do some spatial analyses.

For this assignment, I will look at the film permits of the City of Chicago and see what the most filmed places in this city are. First, we download the dataset from the Chicago open data portal:

## # Data Preparation

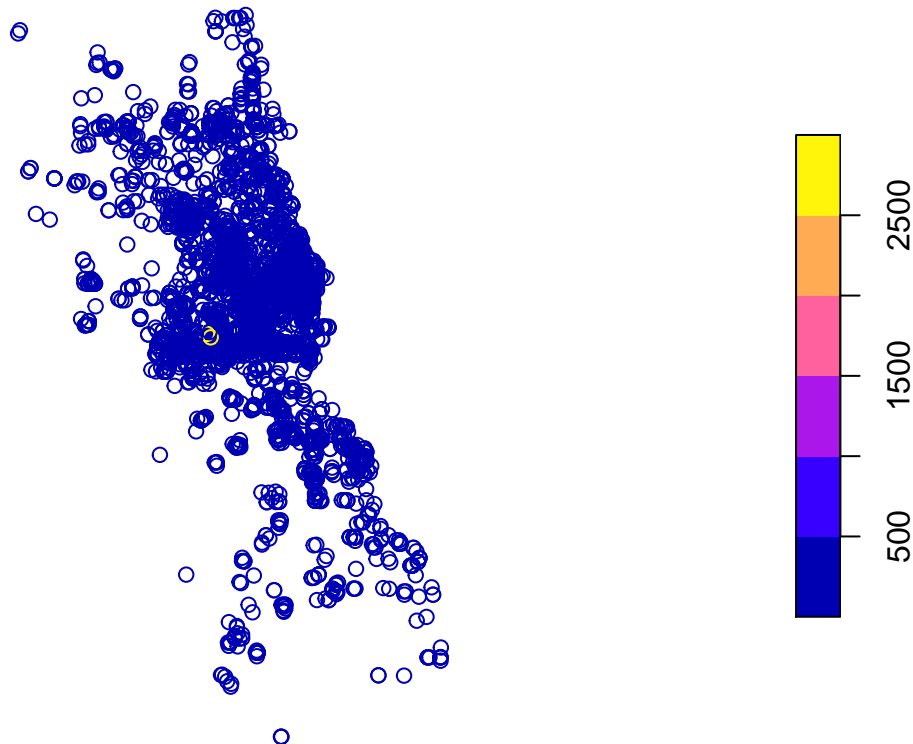
This dataset describes applications to the Chicago Department of Transportation for permits under its jurisdiction where the work type is “Filming.” This should give us a good comprehensive look at where the filming locations are distributed. The time period of this dataset is from 2015 - current (Oct 2019.)

Next up we extract the useful portion of this dataset. For this assignment, we want to look at all the film permits in 2018, and we select all the values where the APPLICATIONSTARTDATE is in 2018.

It is lucky that we do not have any unrecorded location data points, so we do not need to select out any incomplete data entries. We then select the variables needed (**totalfees**, **location.latitude**, and **location.longitude**.)

We then create a point layer (I add the “TOTALFEES” attribute as I want to see if there are geographical differences in the fees charged for film permits):

## Fees for 2018 Chicago Filming Permits



From this point map, we can see that the downtown and central areas are very densely filled with filming applications. Note that the fees turn out to be mostly minimal with only a couple of high numbers. (We also have about 10% missing values as those are mostly waived fee events (583 out of 5620 observations do not have fee data.)) Hence this attribute is not significant. A quick stats summary also confirms this point.

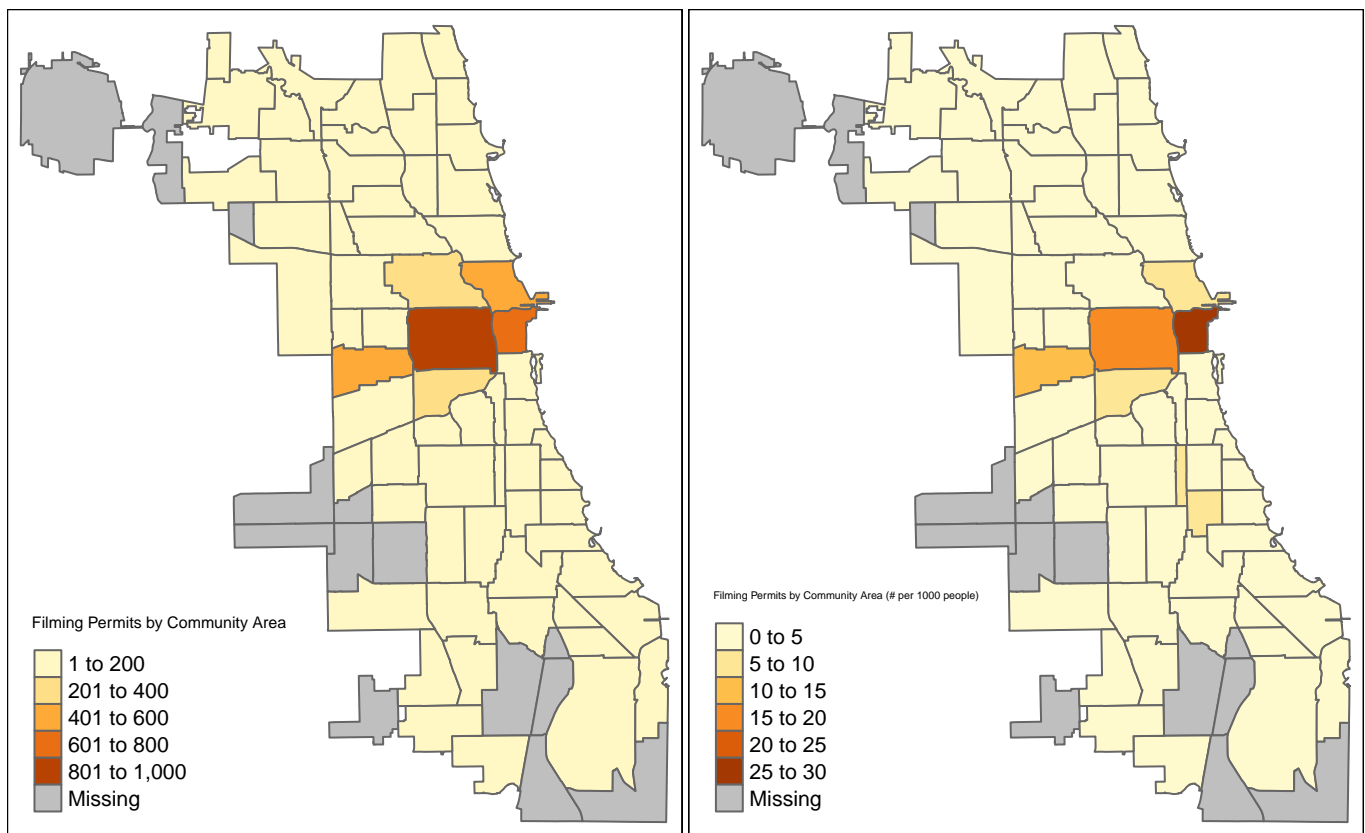
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.0	25.0	25.0	51.6	50.0	3000.0	583

## Community Area Data

Next up we can add the community area attribute to our dataset by getting the community area information from the following source, and we will then spatial join this community boundary to our location information in our selected 2018 data.:

At this point, we can do a choropleth map showing the number of filming permits in each community areas. I am also interested in adjusting this data keeping the community area population number in mind (we will calculate the filming permits per 1000 people in each community area,) and here we use the data from the 2010 census:

We will then parse this PDF file to gather the community area population data, and then calculate the required statistic (filming permits per 1000 people.) Finally, we can update our map with regards to this population information. We now get the number of filming permits per 1000 people (the map on the left is generated with data not adjusted with population number, and the right one is adjusted):



## Discussion

In this assignment, we plot a point map, and two choropleth maps using the Chicago filming permit data. From the results, we can see that the areas with the most filming is the central area of Chicago (Loop, Near North Side, Near West Side, and North Lawndale being the most filmed in terms of numbers, with Near West Side topping the chart.) When we adjust this data with population number, we see that the Loop now stands out as the most popular place for filming, which makes sense. Near West Side is still popular despite being surpassed by the loop. We can also see that some communities have missing information, and one peculiar example is O'Hare, as the airport in intuition should receive some amounts of filming. I suspect this might be due to the information source. This data set is from the Department of Transportation, and O'Hare is an airport, which might not be under the jurisdiction of such administration (It might be the FAA or the airport administration itself.) Overall, we can visualize the filming popularity in Chicago in a clear and concise manner through this method.