MUSA 611 Final Project Proposal

Xingxing Shu

Problem

Traffic crashes are a severe problem in Philadelphia which takes nearly 100 people's lives and injures around 300 in over 10,0000 incidents every year. On November 7, 2016, Mayor of Philadelphia, James F. Kenny created a plan called Vision Zero Task Force set a target of zero traffic deaths by 2030. Though only one year after this plan launched, Philadelphia had 19% decrease in traffic deaths to 78 people (the lowest since 2012), this number bounced back to 91 in 2018. During last five years from 2013 to 2017, the average traffic deaths per 100,000 residents is 6.06 which is more than twice as New York's. Also, according to report from PennDOT, more than 50% of traffic deaths occurred in just 12% of Philadelphia's streets.

In this project, I will use traffic crashes data from OpenDataPhilly and gather some other relevant data to answer following questions:

• Where do most crashes happened in city?

• What are the characteristics of crashes in Philly?

• Which streets are dangerous and which intersections have the most crashes?

• Where are these dangerous streets or intersections located? Do traffic crashes affect all neighborhoods in Philly equally?

Dataset

Vehicular Crashes Data: This dataset contains crash data in city of Philadelphia from 2011 to 2017 with 77,463 rows of data. Then I selected useful and information-complete columns for latter analysis.

Method

Step 1: My analysis begins with street centerlines of Philadelphia In ArcGIS. First, I need to break each road into separate lines at their intersections and extract all intersection points by topology in ArcGIS. Then I create buffers for these two objects: 15 meters for road centerlines and radius of 10 meters around intersection points. Then I clip the road centerlines file by the intersection buffers to get road lines without roads in intersection buffers to avoid overlapping. To locate which roads or intersections have the most crashes, I spatial joined the crash points to these two shapefiles in ArcGIS;

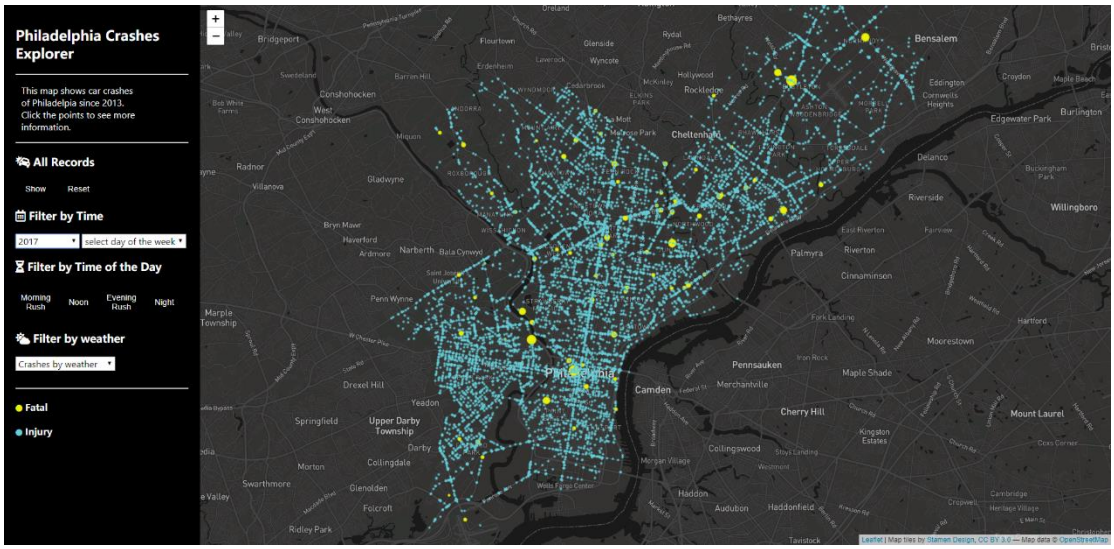
Step 2: Divide the crash raw data from 2011 to 2017 into multiple files in Python. First group is grouped by crash years and months and calculate the total number of injured and fatal people; Second group is grouped and calculate the total number of victims by months and the time of a day to see how it changes through a day. Others are grouped by different characteristics of collisions.

Step 3: Crash data with complete timestamp of 2012 was read in R and I use lubridate package to parse it into the time form I expect: "2012-08-30 06:10:00" and extract

which week of the year the crash was in. Then I need to combine these three data panels together to do further analysis. First is to join crash data with weather data based on the timestamp; Then is to join this new data panel with census data use spatial join.

Design

Since most bicycle and pedestrian crashes don't do significant damage to a vehicle, it's easy to see how many of them simply wouldn't be reported by PennDOT. So, the problem of bicycle and pedestrian crashes could actually be a lot worse than what is actually shown. I'm hopeful that we'll see more data released, perhaps by jurisdictional police departments, which may shed more light on this. It would also be great to combine this with some sort of crowdsourced pedestrian and bicycle crash map where users could report those minor crashes that don't necessitate a police report. That could be a great way to further identify the most dangerous streets and intersections for bicyclists and pedestrians.



Anticipated Difficulties

My first proposal was about to create an app illustrate green view index of community streets in Philly using google street images and perform sentimental analysis of residential tweets of different communities. I tried to download and extract green view of the two selected communities which worked but if I want to expand it to the whole city scale, I need to process more than 1,000,000 images which is a hard work to do in a short time.

As for sentimental analysis on tweets, I collected more than 10,000 tweets in the past week but the biggest problem is tweets with latitude and longitude are only account for less than 1%.

So I had to give this up and found another project to do. Hope you can understand this.

Missing pieces

Hard to tell, JS is so new and sort of difficult to me, I feel like the final application is going to deviate from what I expected ☹