**Introduction of Atlanta traffic data**

All the data is acquired from [**https://gdottrafficdata.drakewell.com/**](https://gdottrafficdata.drakewell.com/). It is the data of Georgia Department of Transportation (GDOT). GDOT puts hundreds of sensors in Georgia and the sensors count the number of cars passing by. We get several kinds of data and we will introduce them one by one.

The first one is beats data and there are two files in the beats data file.

The first one is a geojson file. It shows the shape and location of each beats. There are totally 80 beats.

The second one is a csv file. There are 80 rows in the csv. One row contains the information of one beat. The first string in each row indicates for the beat ID and the following strings shows the ID of the sensors which locates in that beat.

The second one is AADT data. AADT means Annual Average Daily Traffic.

Four csv file is included in this part.

The first one is annulized\_statistics.csv. It includes all the information of sensors in whole Georgia. The sensor ID, AADT and position (Lat/Long) are all included in the file.

The second one is Atlanta\_data.csv. It has the sensor ID, AADT and position information of Fulton, Atlanta.

The third one is Atlanta\_data\_plus.csv. It has the information of both Atlanta and several counties around Atlanta.

The fourth one is total\_data\_of\_georgia.csv. It includes the information of sensor ID, AADT and position of whole Georgia.

The third data is row data. It includes the data of 2017 whole year and 2018.1.1-2018.8.19.

The name of each csv file in 2017 is the beat ID. In the file, it is a big matrix. Every row in the file means the sensor in the beat. Column means the traffic data in each time. Every column means one hour. Like in 2017 file, each column means one hour in 2017.

2018 row data file is similar except that the csv file is replaced by txt file. I strongly recommend users to use numpy to process the data.

As only a small number of sensors are open at one time. There are numerous 0 in the matrix. It is a low rank matrix in fact.

Pictures and htmls are visualization.