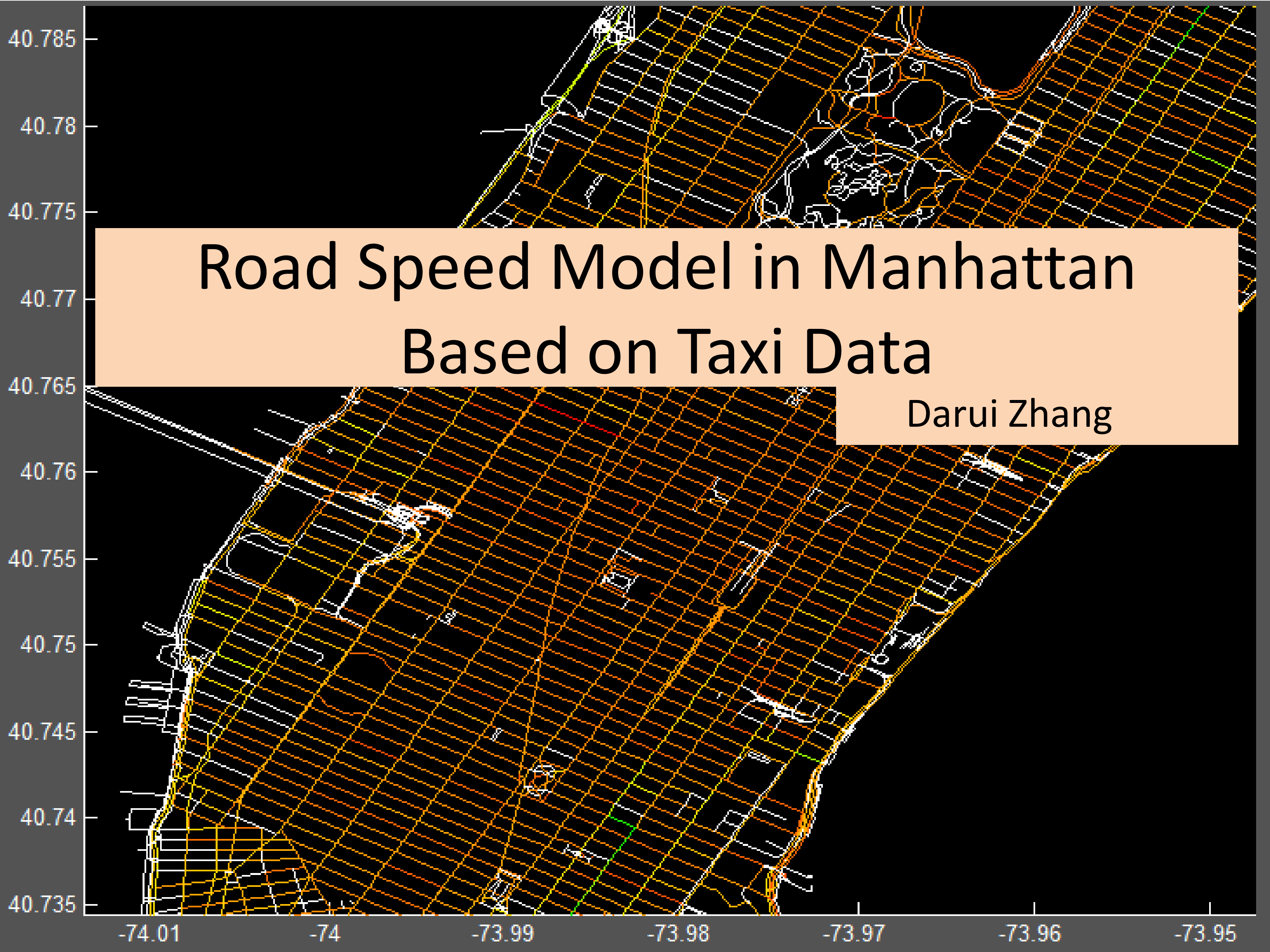
Road Speed Model in Manhattan Based on Taxi Data

Darui Zhang



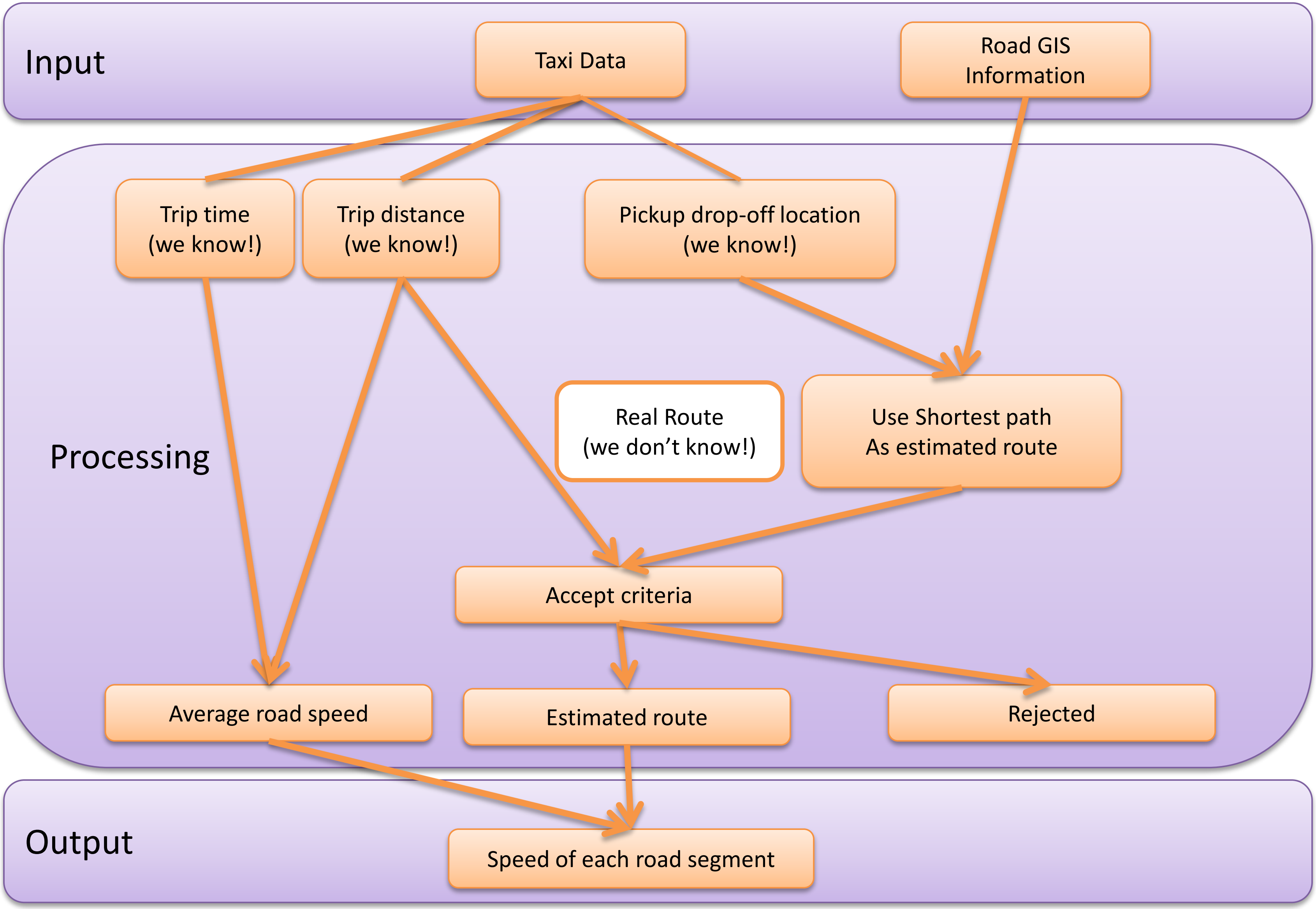
Problem Statement & Challenges

Problem Statement:
We want to use the taxi data we acquired to generate driving speed of different road segments at different time of a day at in Manhattan.

Challenges:

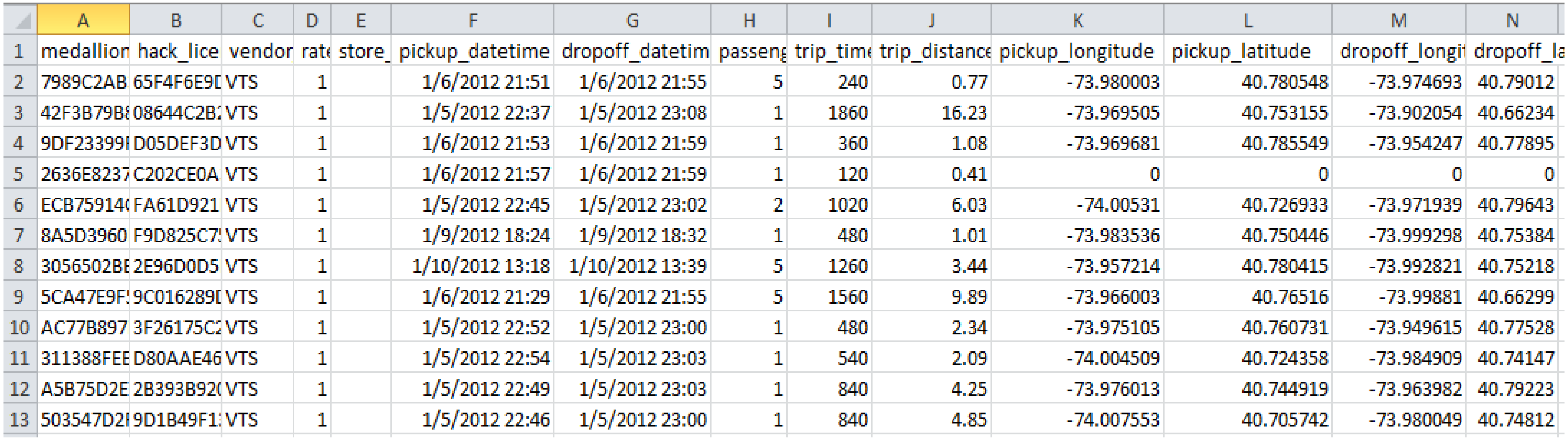
1. Generate estimated routes from GPS positions.
2. Set up acceptance criteria
3. Process large amount of data
4. Data visulization

Approach



Input
Taxi Data
Road GIS Information
Processing
Trip time (we know!)
Trip distance (we know!)
Pickup drop-off location (we know!)
Real Route (we don't know!)
Use Shortest path As estimated route
Accept criteria
Average road speed
Estimated route
Rejected
Output
Speed of each road segment

Taxi Data



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	medallion	hack_lice	vendor	rate	store_	pickup_datetime	dropoff_datetim	passeng	trip_time	trip_distance	pickup_longitude	pickup_latitude	dropoff_longit	dropoff_la
2	7989C2AB	65F4F6E9D	VTS	1		1/6/2012 21:51	1/6/2012 21:55	5	240	0.77	-73.980003	40.780548	-73.974693	40.79012
3	42F3B79B8	08644C2B2	VTS	1		1/5/2012 22:37	1/5/2012 23:08	1	1860	16.23	-73.969505	40.753155	-73.902054	40.66234
4	9DF23399	D05DEF3D	VTS	1		1/6/2012 21:53	1/6/2012 21:59	1	360	1.08	-73.969681	40.785549	-73.954247	40.77895
5	2636E8237	C202CE0A	VTS	1		1/6/2012 21:57	1/6/2012 21:59	1	120	0.41	0	0	0	0
6	ECB75914	FA61D921	VTS	1		1/5/2012 22:45	1/5/2012 23:02	2	1020	6.03	-74.00531	40.726933	-73.971939	40.79643
7	8A5D3960	F9D825C7	VTS	1		1/9/2012 18:24	1/9/2012 18:32	1	480	1.01	-73.983536	40.750446	-73.999298	40.75384
8	3056502BE	2E96D0D5	VTS	1		1/10/2012 13:18	1/10/2012 13:39	5	1260	3.44	-73.957214	40.780415	-73.992821	40.75218
9	5CA47E9F	9C016289	VTS	1		1/6/2012 21:29	1/6/2012 21:55	5	1560	9.89	-73.966003	40.76516	-73.99881	40.66299
10	AC77B897	3F26175C2	VTS	1		1/5/2012 22:52	1/5/2012 23:00	1	480	2.34	-73.975105	40.760731	-73.949615	40.77528
11	311388FEE	D80AAE46	VTS	1		1/5/2012 22:54	1/5/2012 23:03	1	540	2.09	-74.004509	40.724358	-73.984909	40.74147
12	A5B75D2E	2B393B92	VTS	1		1/5/2012 22:49	1/5/2012 23:03	1	840	4.25	-73.976013	40.744919	-73.963982	40.79223
13	503547D2	9D1B49F1	VTS	1		1/5/2012 22:46	1/5/2012 23:00	1	840	4.85	-74.007553	40.705742	-73.980049	40.74812

Contains: (accuracy)

F: pickup date and time
G: drop-off date and time
H: number of passengers
I: trip time (10 seconds)
J: trip distance (0.01 mile)

K: pickup longitude (0.000001 deg)
L: pickup latitude (0.000001 deg)
M: pickup longitude (0.000001 deg)
N: pickup latitude (0.000001 deg)

Road GIS Data

1. Find road gis data. (.shp)
   Source: OpenStreetMap http://download.geofabrik.de/north-america.html

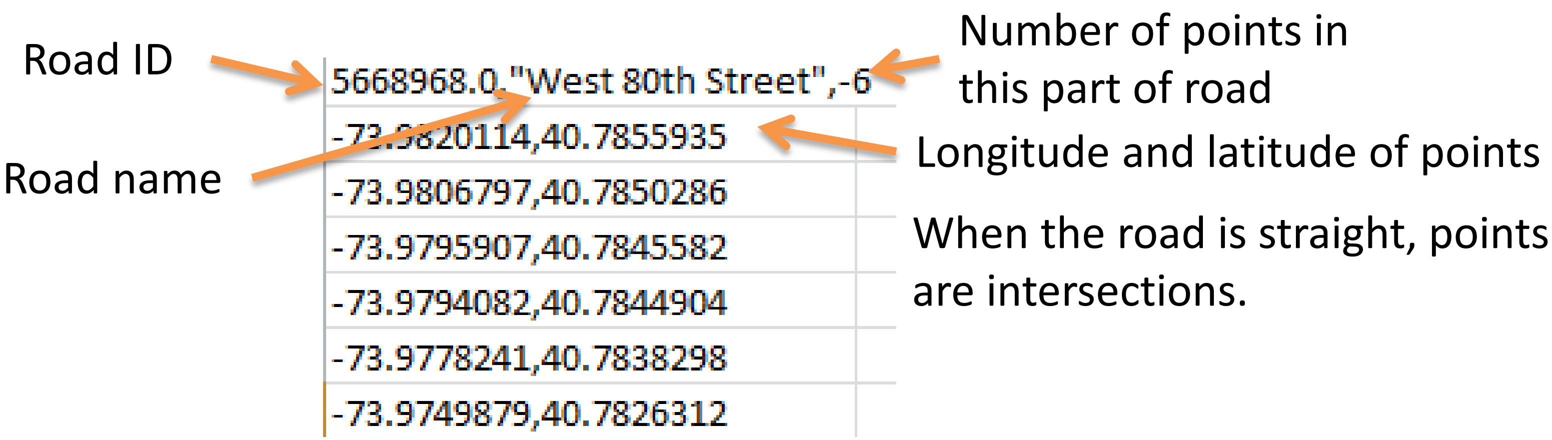
2. Use QGIS to select Manhattan as the “area of interest”.

3. Export road layer to “.bna” and “.csv” format file.

4. use “.bna” and “.csv” to generate “edges” and “vertices” which enable us to use shortest path algorithm to find the estimated route.

Road GIS Data

.bna file



```
5668968.0,"West 80th Street",-6
-73.9820114,40.7855935
-73.9806797,40.7850286
-73.9795907,40.7845582
-73.9794082,40.7844904
-73.9778241,40.7838298
-73.9749879,40.7826312
```

Road ID

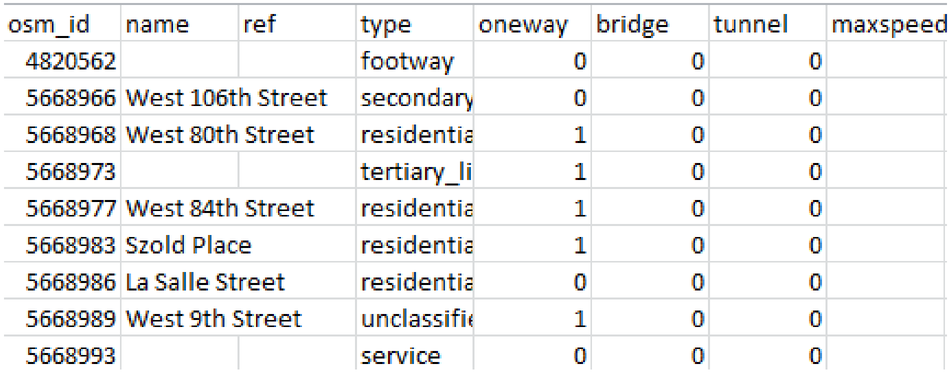
Road name

Number of points in this part of road

Longitude and latitude of points

When the road is straight, points are intersections.

.csv file



osm_id	name	ref	type	oneway	bridge	tunnel	maxspeed
4820562			footway	0	0	0	
5668966	West 106th Street		secondary	0	0	0	
5668968	West 80th Street		residentia	1	0	0	
5668973			tertiary_li	1	0	0	
5668977	West 84th Street		residentia	1	0	0	
5668983	Szold Place		residentia	1	0	0	
5668986	La Salle Street		residentia	0	0	0	
5668989	West 9th Street		unclassifie	1	0	0	
5668993			service	0	0	0	

.csv file contains information of road id, road name, type of road, whether is oneway, bridge, tunnel and max speed

Shortest Path Algorithm

Dijkstra's algorithm:

The algorithm is to find the path with lowest cost between vertices.
In our problems, vertices are intersections. Edges are road segments that connect two vertices.

Resources:
Dijkstra's shortest path algorithm
YouTube http://www.youtube.com/watch?v=87_1K2GQFdU
Matlab command:
http://www.mathworks.com/help/bioinfo/ref/graphshortestpath.html

Acceptance Criteria

Errors data come from three aspects:

1. Resolution of trip distance data which is of 0.01 mile.
2. Distance form pick up and drop-off point to the nearest intersections or vertices, because our algorithm can only search trip distances form vertex to vertex.
3. Others factors that we cannot control, such as drivers might not take the shortest path, or passengers want to drop off a friend along the way.

Acceptance criteria involves two aspects:

1.Absolution value should be within a bound which is the data resolution plus the length of pickup and drop-off road segments. This account for errors in the first and second category

2. Relative error less than 10% which account for the errors in the third category.

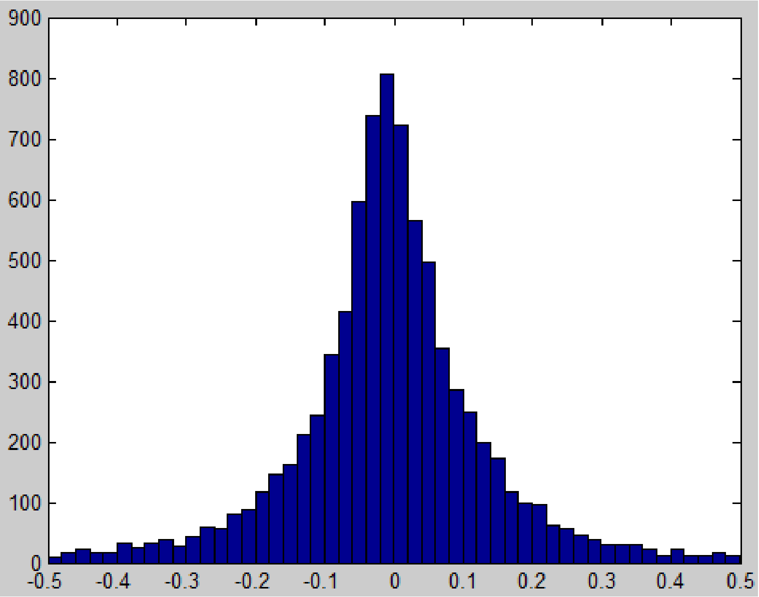
Driving Speed Model

The driving speed during rush hours and other times of a day can be very different. We divide time segments every half an hour.

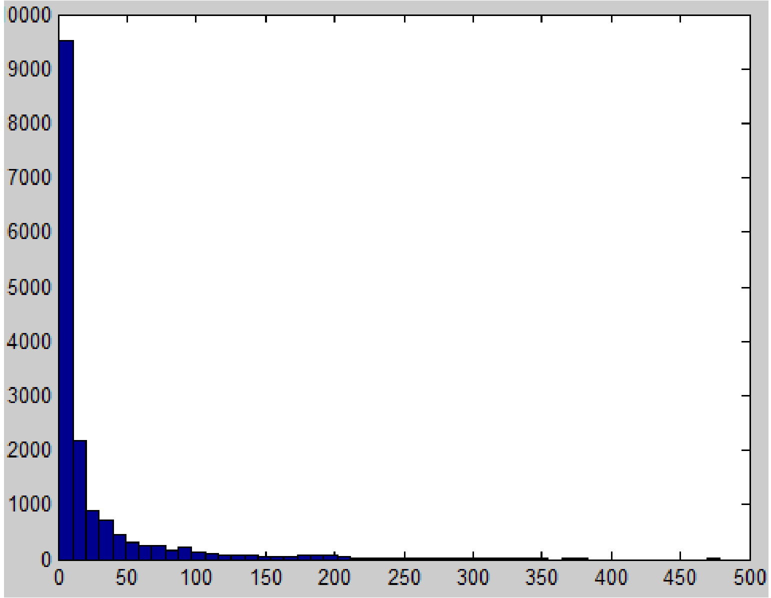
We assume a taxi travels at a same speed during a trip. So we put the speed in “cell” which is assigned for this specific road segment and time. If the a trip is made during two time sections, we will put the speed in both time segment.

The average value of each “cell” represent the driving speed of a road segment and time segment.

Data Statistics

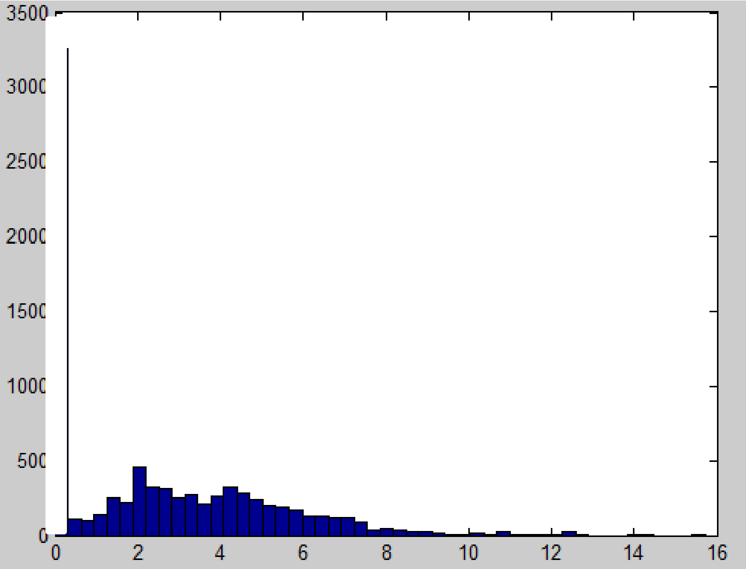


1 Distribution of the difference between estimated and real trip distance

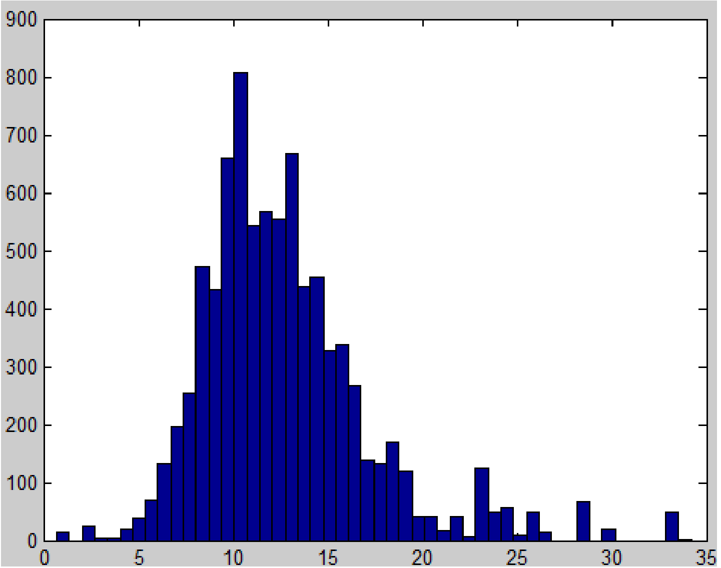


2 Distribution of number of data in road segments.

Data Statistics



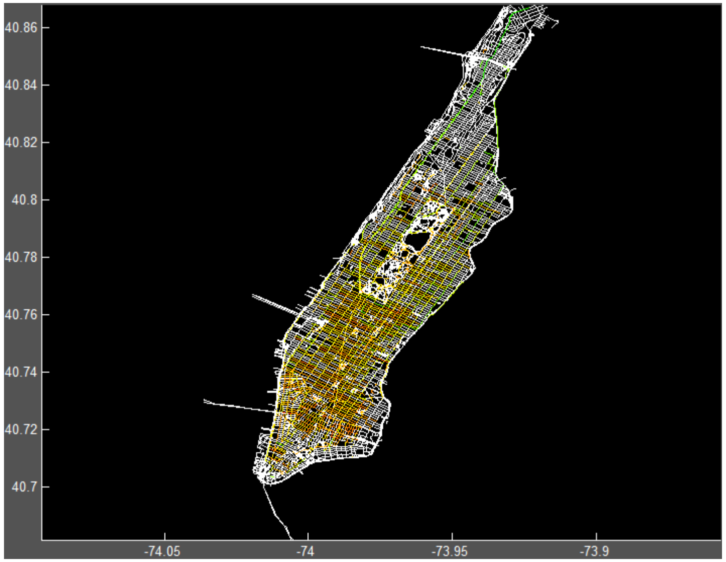
3 Distribution of standard deviation within each cells



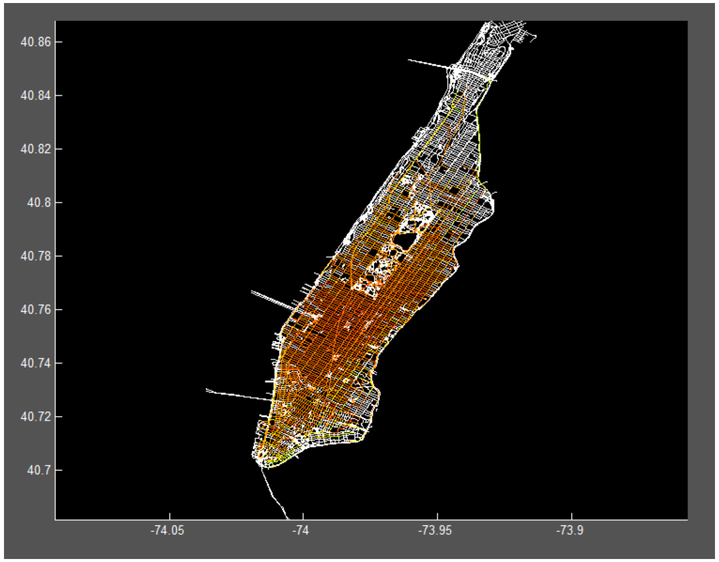
4. Distribution of read segment driving speed

Time lapse

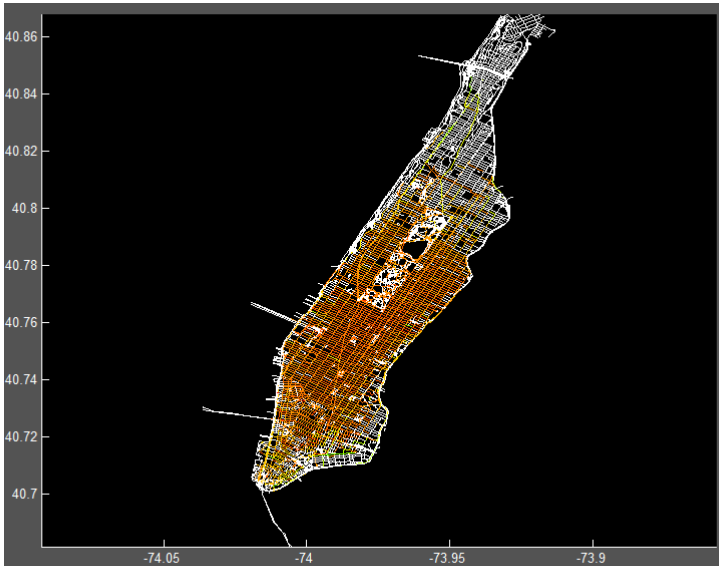
Red: Slow Traffic
Green: Fast Traffic
White: Do not have data



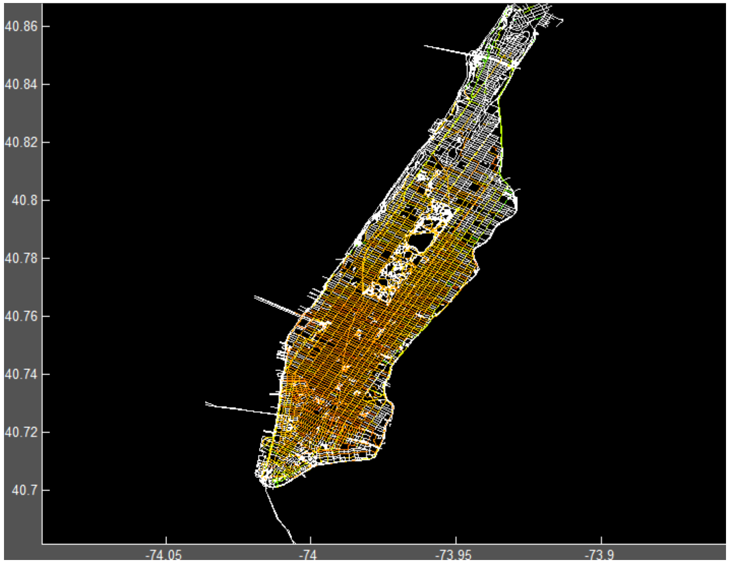
2:00-2:30 am



9:00-9:30 am



6:00-6:30 pm



10:00-10:30 pm