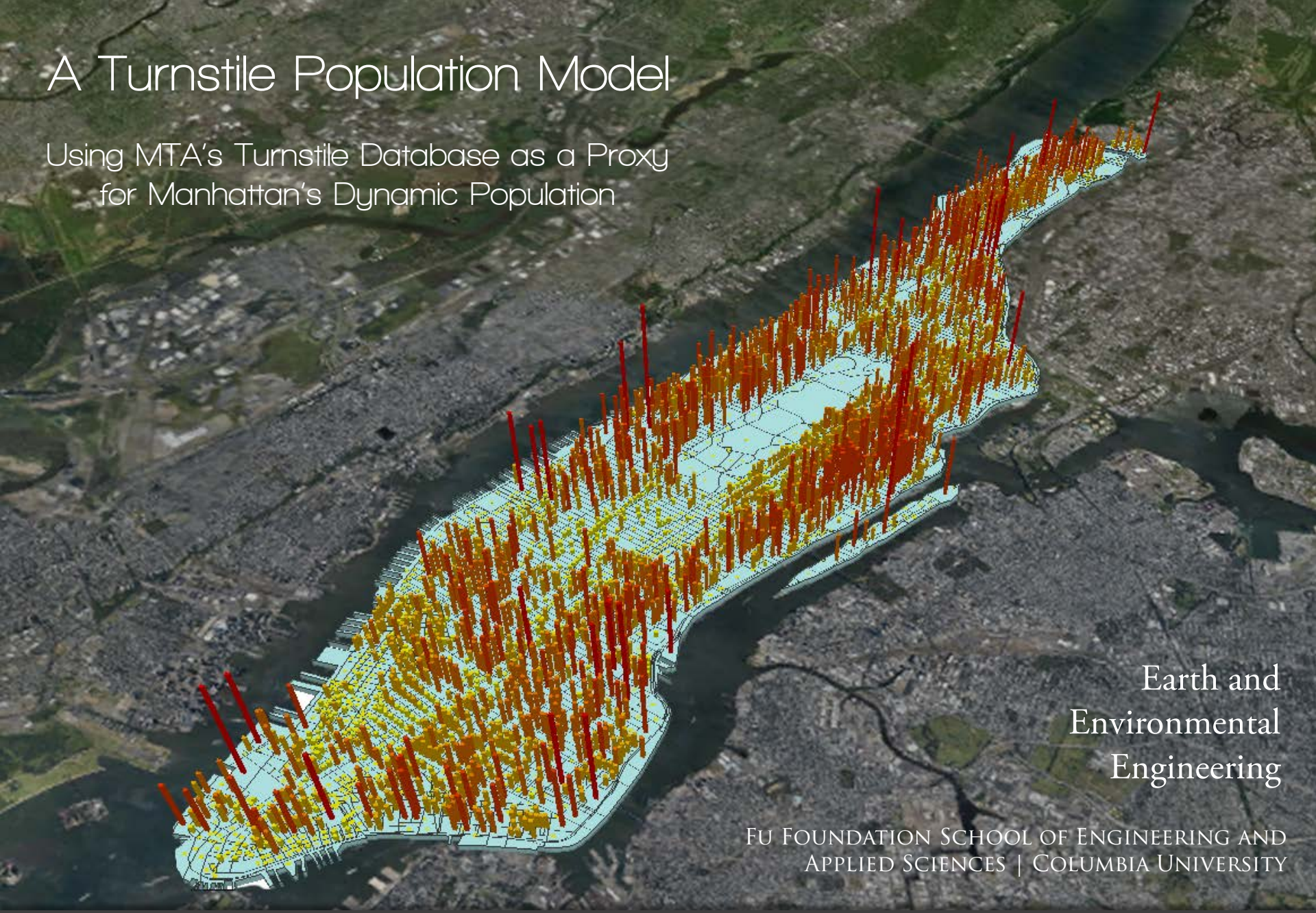
A Turnstile Population Model

Using MTA's Turnstile Database as a Proxy for Manhattan's Dynamic Population

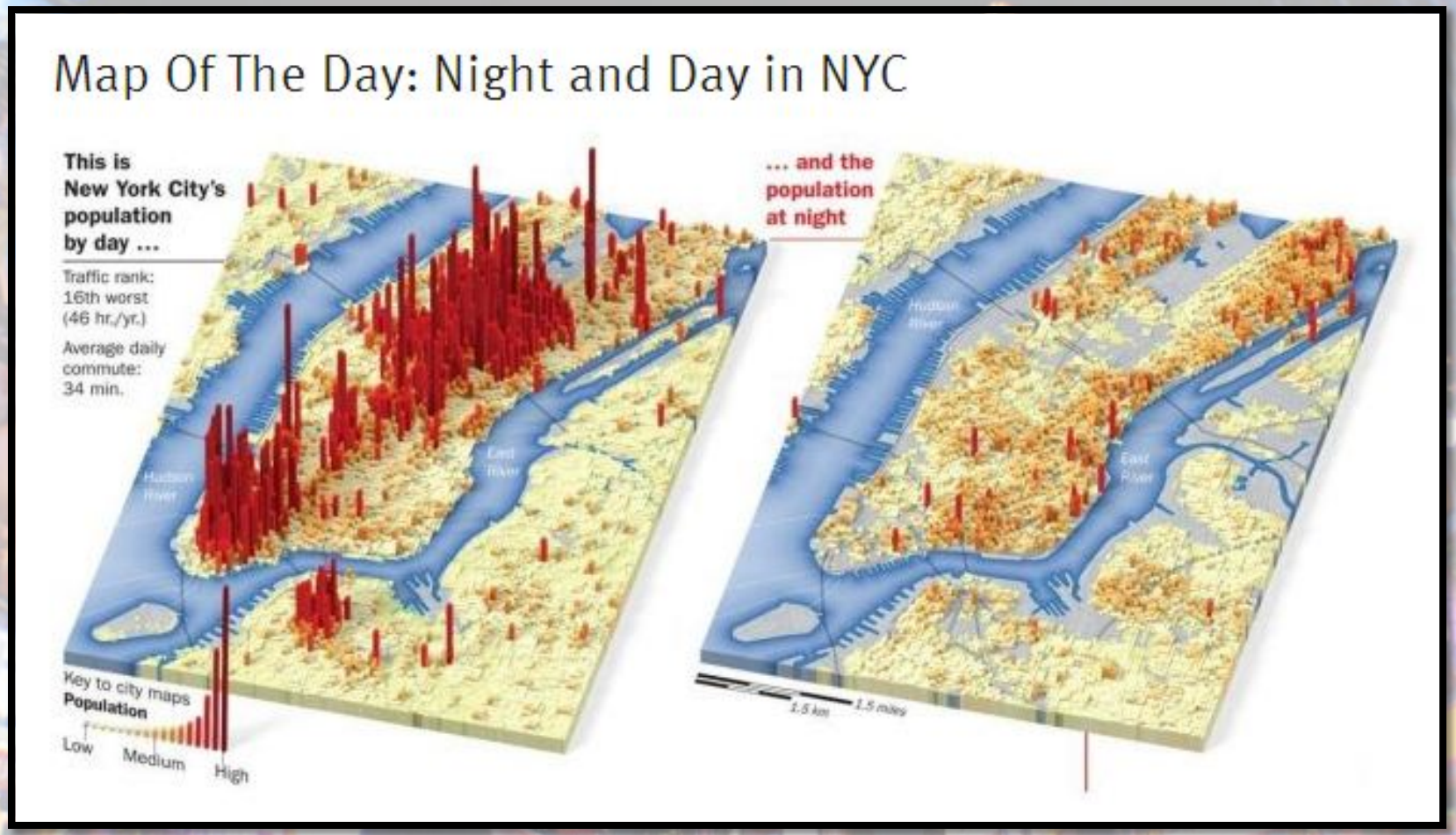


Earth and Environmental Engineering

Fu Foundation School of Engineering and Applied Sciences | Columbia University

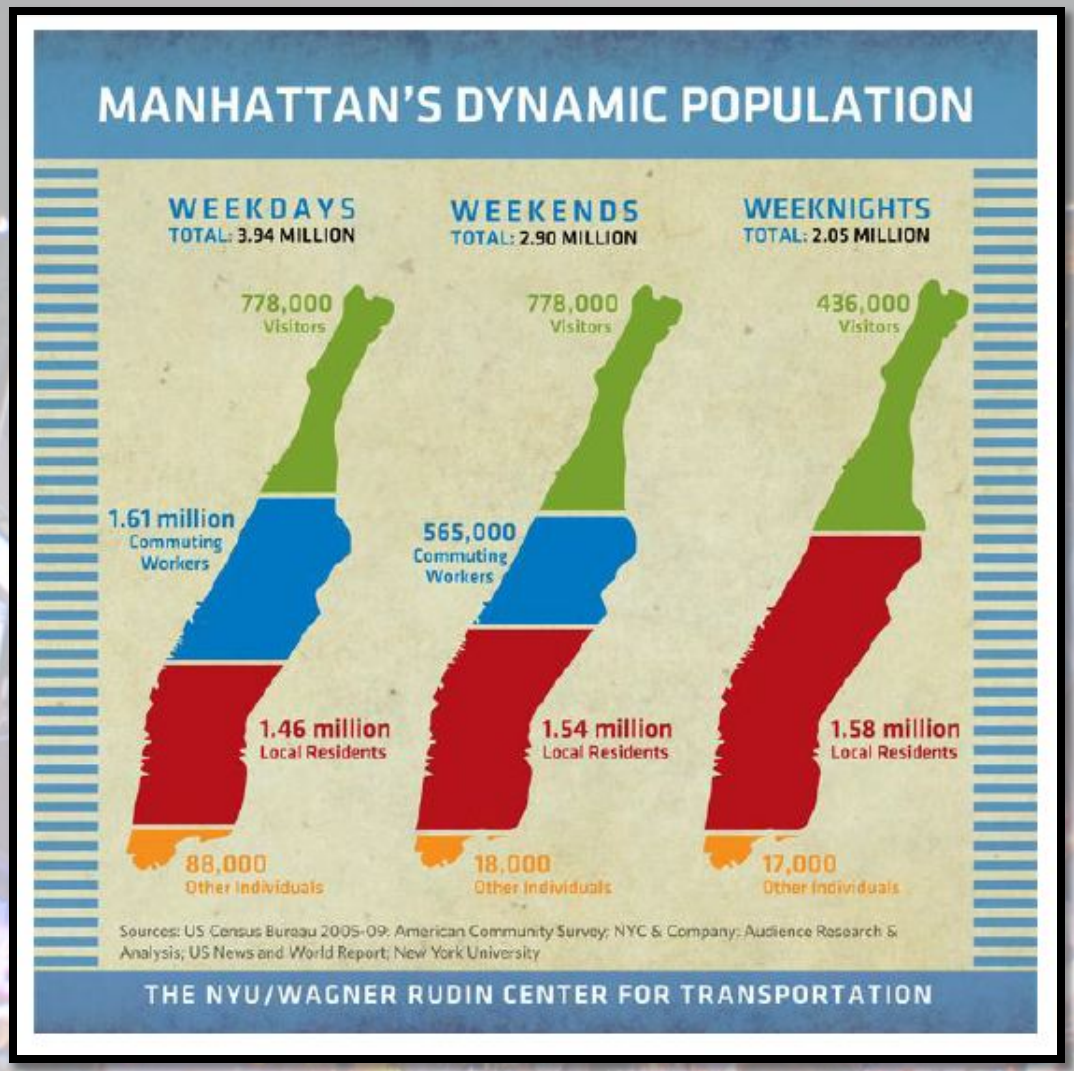
Justin Fung | Industrial Engineering and Operations Research | May, 2013

Existing Dynamic Population Estimates:



Source: LandScan, Time Magazine

Existing Dynamic Population Estimates:



Source: NYU / Wagner Rudin Center for Transportation (2012)

Existing Dynamic Population Estimates:

WEEKDAY DAYTIME POPULATION, 2010

Estimate	Commuting workers	Local residents	Out-of-town visitors	Day-trip Visitors	Hospital users	Commuting students	TOTAL
Upper	1.61 million	1.46 million	404,256	374,223	18,236	70,054	3.94 million

Census defined daytime population = 3.07 million. Based on upper estimate, that figure is an undercount of 940,000

WEEKEND DAYTIME POPULATION, 2010

Estimate	Commuting workers	Local residents	Out-of-town visitors	Day-trip Visitors	Hospital users	Commuting students	TOTAL
Upper	564,665	1.54 million	404,256	374,223	18,236	0	2.90 million

Census defined weekend daytime population = 2.14 million. Based on upper estimate, that figure is an undercount of 880,000

WEEKNIGHT POPULATION, 2010

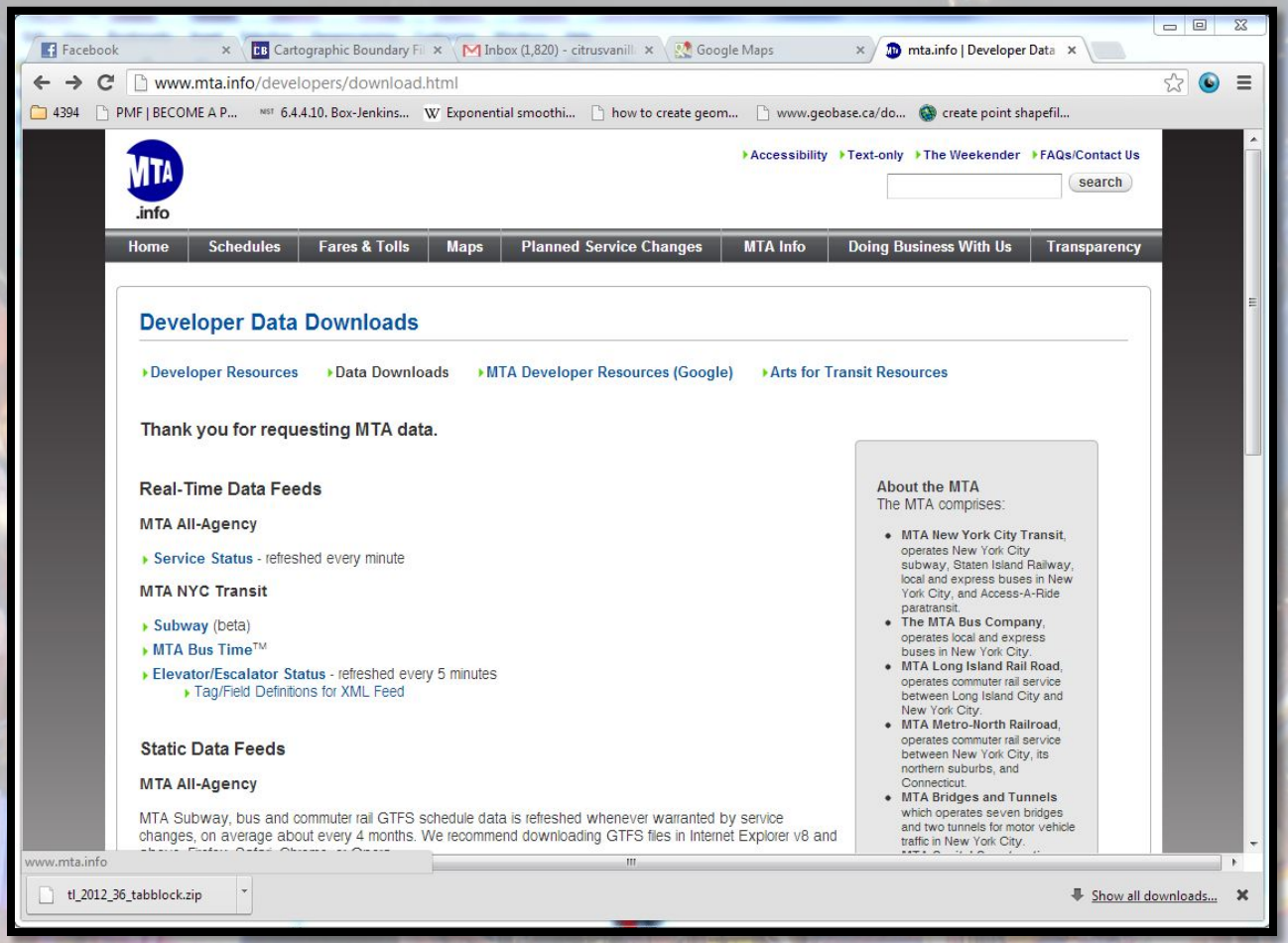
Estimate	Night-shift Commuters	Local residents	Out-of-town visitors	"Night-trip" visitors	Hospital users	Commuting students	TOTAL
Upper	17,747	1.58 million	404,256	31,863	17,260	0	2.05 million

Census defined nighttime population = number of local residents, or 1.62 million. Based on upper estimate, that figure is an undercount of 550,000

Sources: US Census Bureau 2010 American Community Survey, NYC & Company, Audience Research & Analysis, US News and World Report, New York University

Source: NYU / Wagner Rudin Center for Transportation (2012)

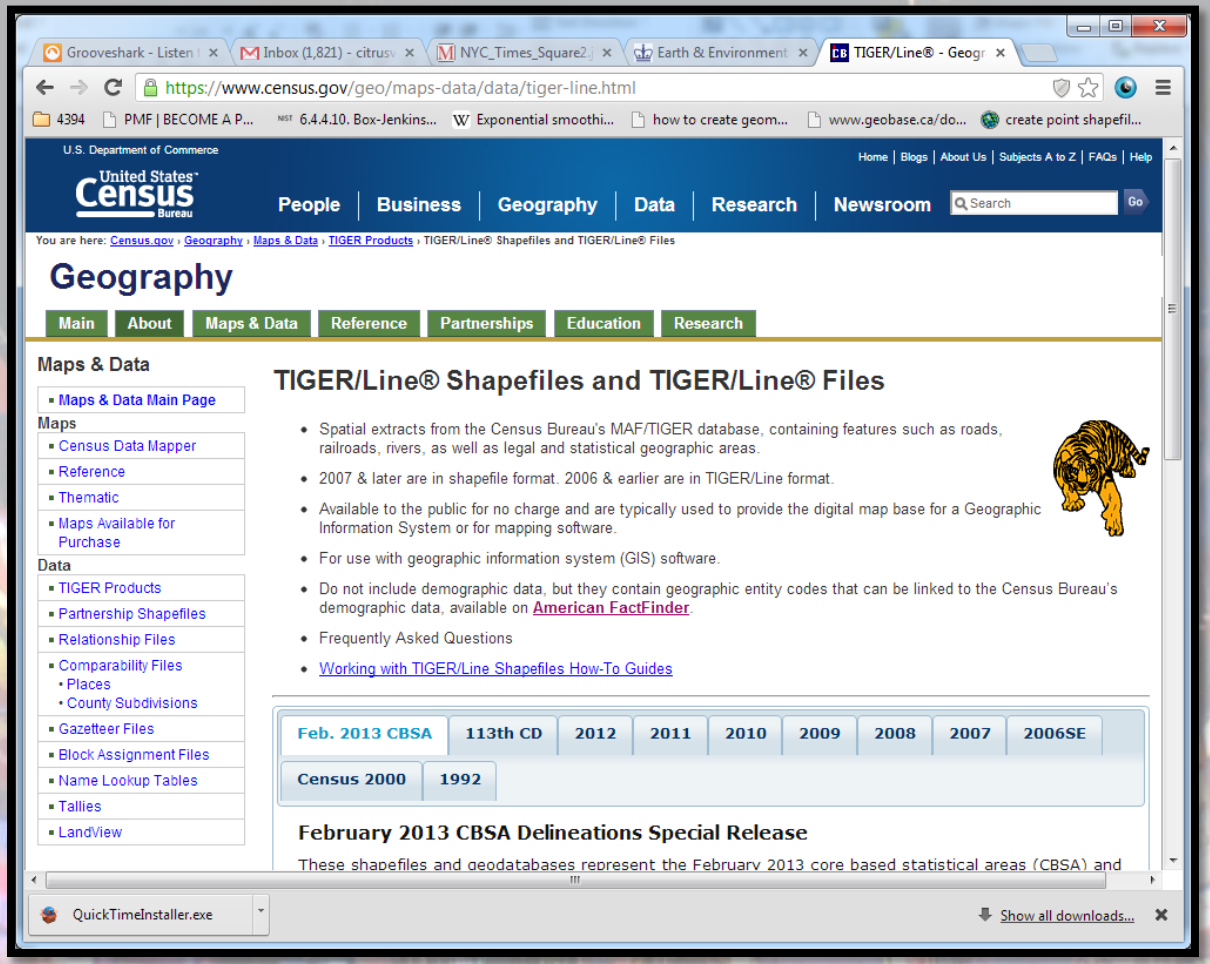
Subway:



Source: MTA Developers Website

ii. Data and Sources

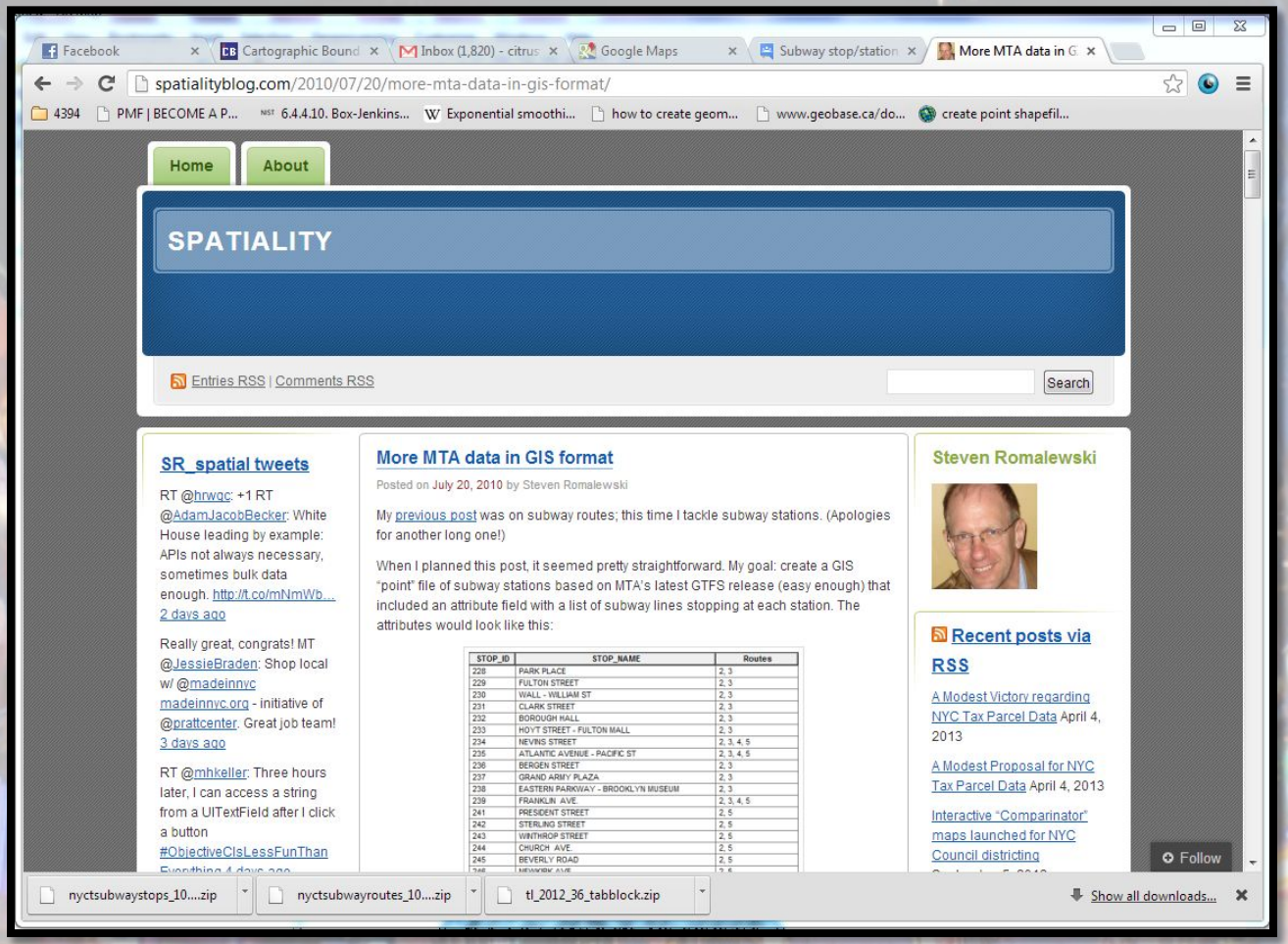
Population:



Source: US Census Bureau TIGER / LINE Database

ii. Data and Sources

Subway:



STOP_ID	STOP_NAME	Routes
228	PARK PLACE	2, 3
229	FULTON STREET	2, 3
230	WALL - WILLIAM ST	2, 3
231	CLARK STREET	2, 3
232	BOROUGH HALL	2, 3
233	HOYT STREET - FULTON MALL	2, 3
234	NEVINS STREET	2, 3, 4, 5
235	ATLANTIC AVENUE - PACIFIC ST	2, 3, 4, 5
236	BERGEN STREET	2, 3
237	GRAND ARMY PLAZA	2, 3
238	EASTERN PARKWAY - BROOKLYN MUSEUM	2, 3
239	FRANKLIN AVE	2, 3, 4, 5
241	PRESIDENT STREET	2, 5
242	STERLING STREET	2, 5
243	WINTHROP STREET	2, 5
244	CHURCH AVE	2, 5
245	BEVERLY ROAD	2, 5

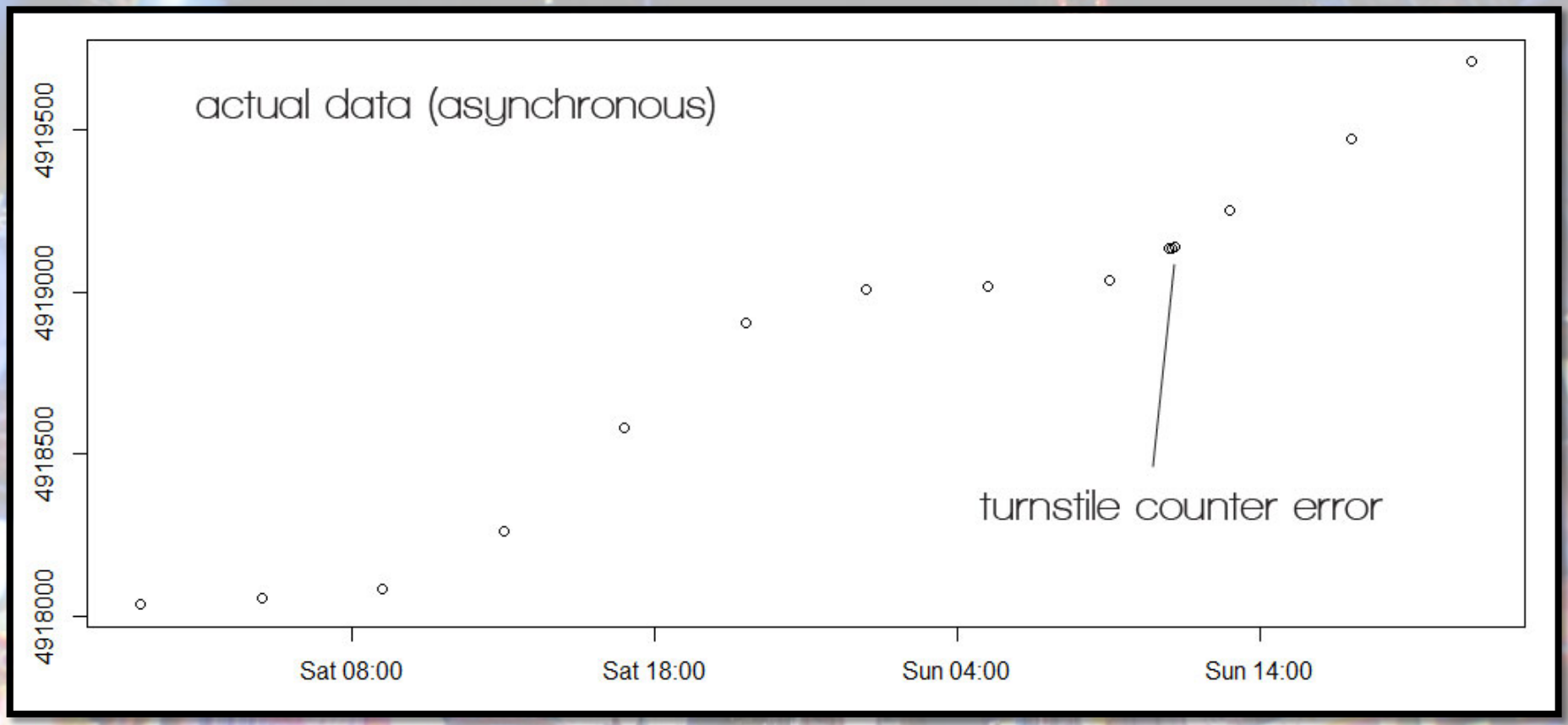
Source: SpatialityBlog.com

ii. Data and Sources

1. Forecasting Model
   - R language script, MTA turnstile Database
2. Geo-spatial Population Assignment
   - GIS

iii. Problem Approach

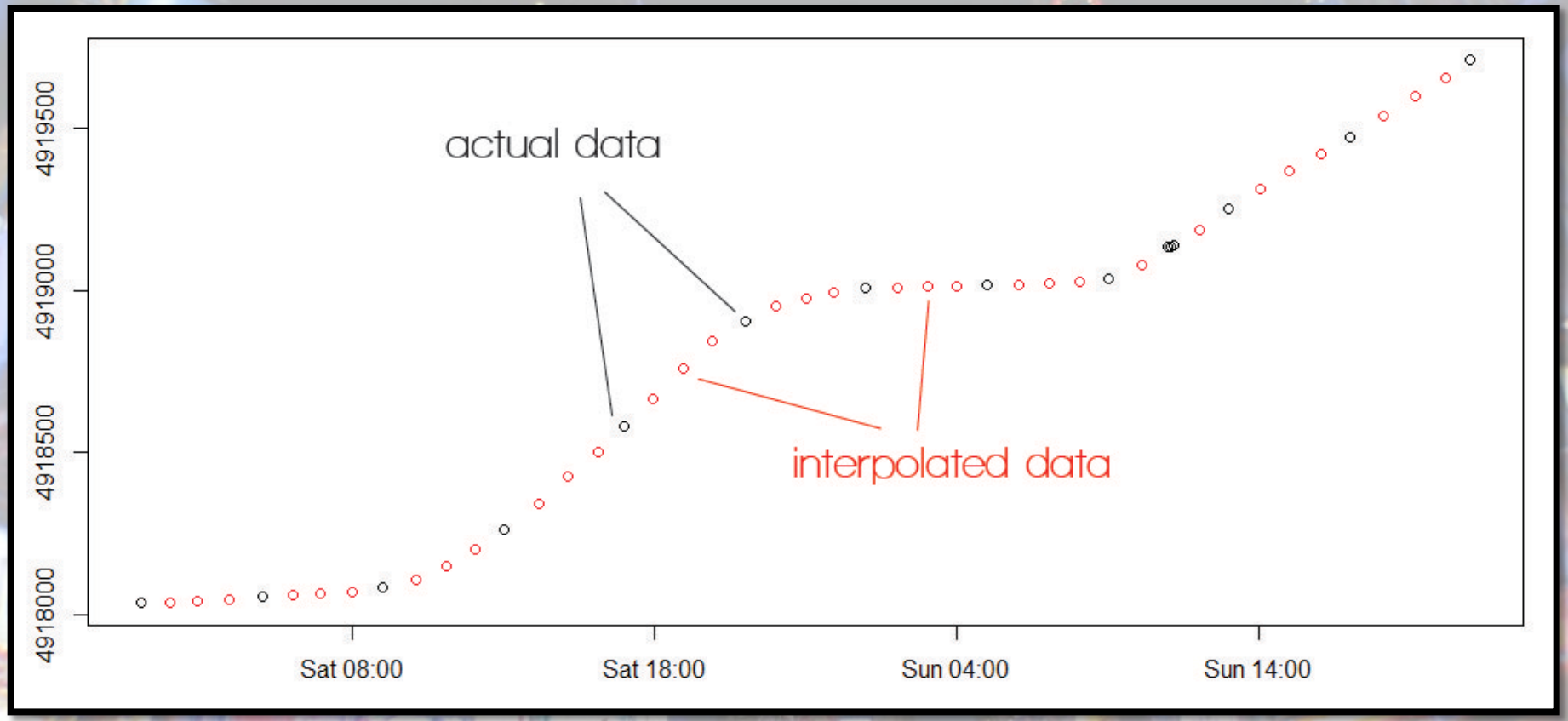
Actual Turnstile Counter Data:



116th and Broadway

iv. Forecast Model

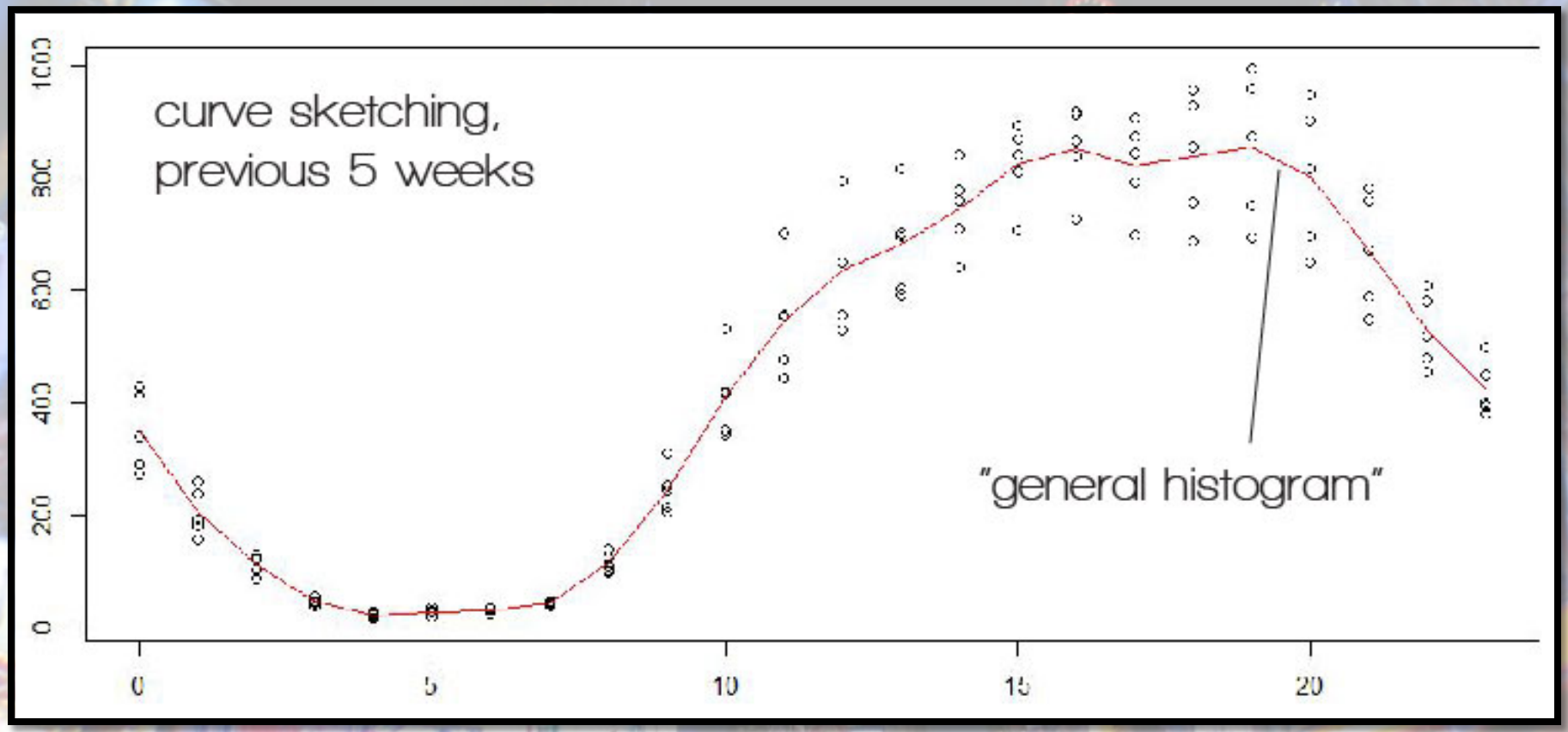
Asynchronous Cubic Monotone Spline Interpolation:



116th and Broadway

iv. Forecast Model

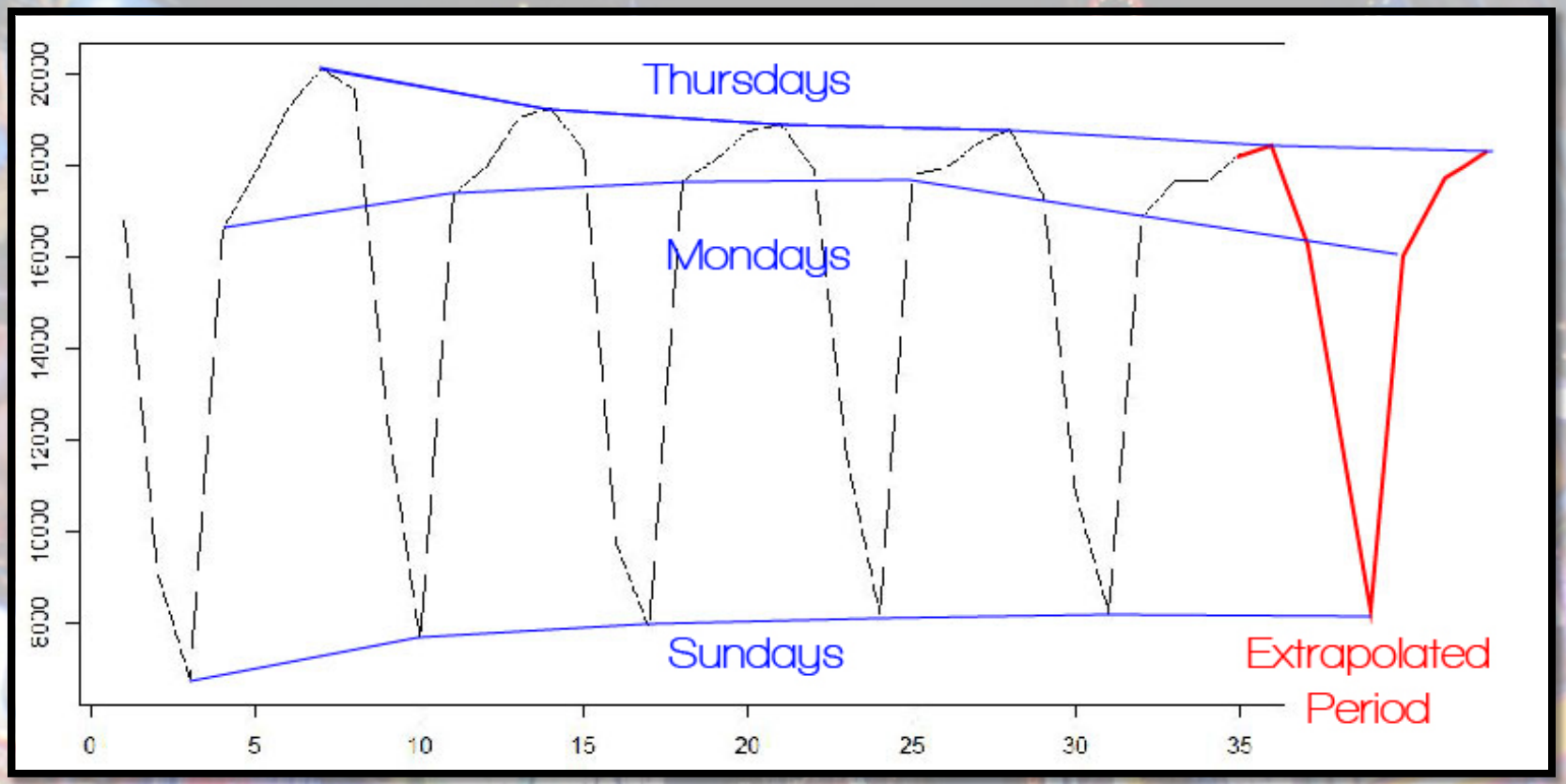
Curve Fitting for Histograms:



Past 5 Saturdays, 116th and Broadway

iv. Forecast Model

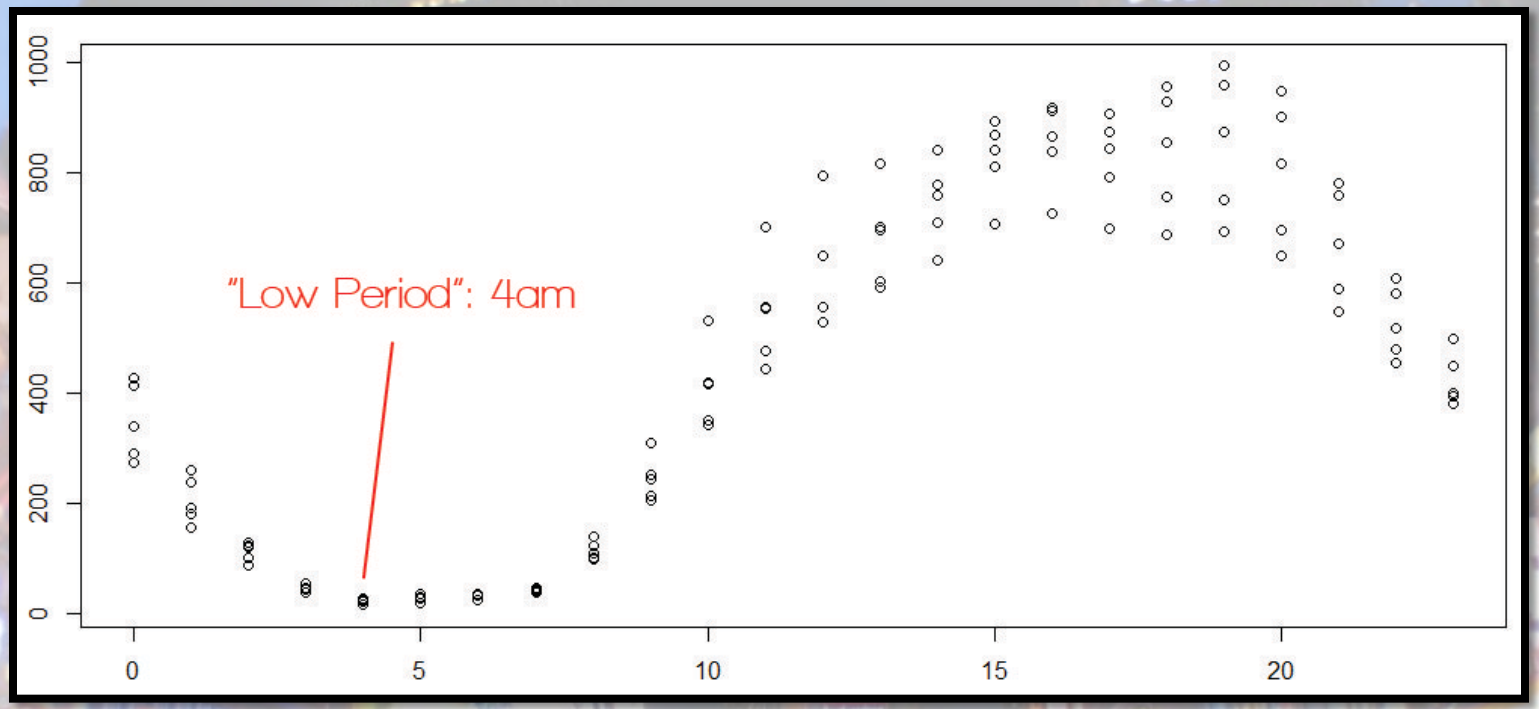
Exponential Smoothing Extrapolation with Trend:



Cumulative Entries for the Past 35 Days, 116th and Broadway

iv. Forecast Model

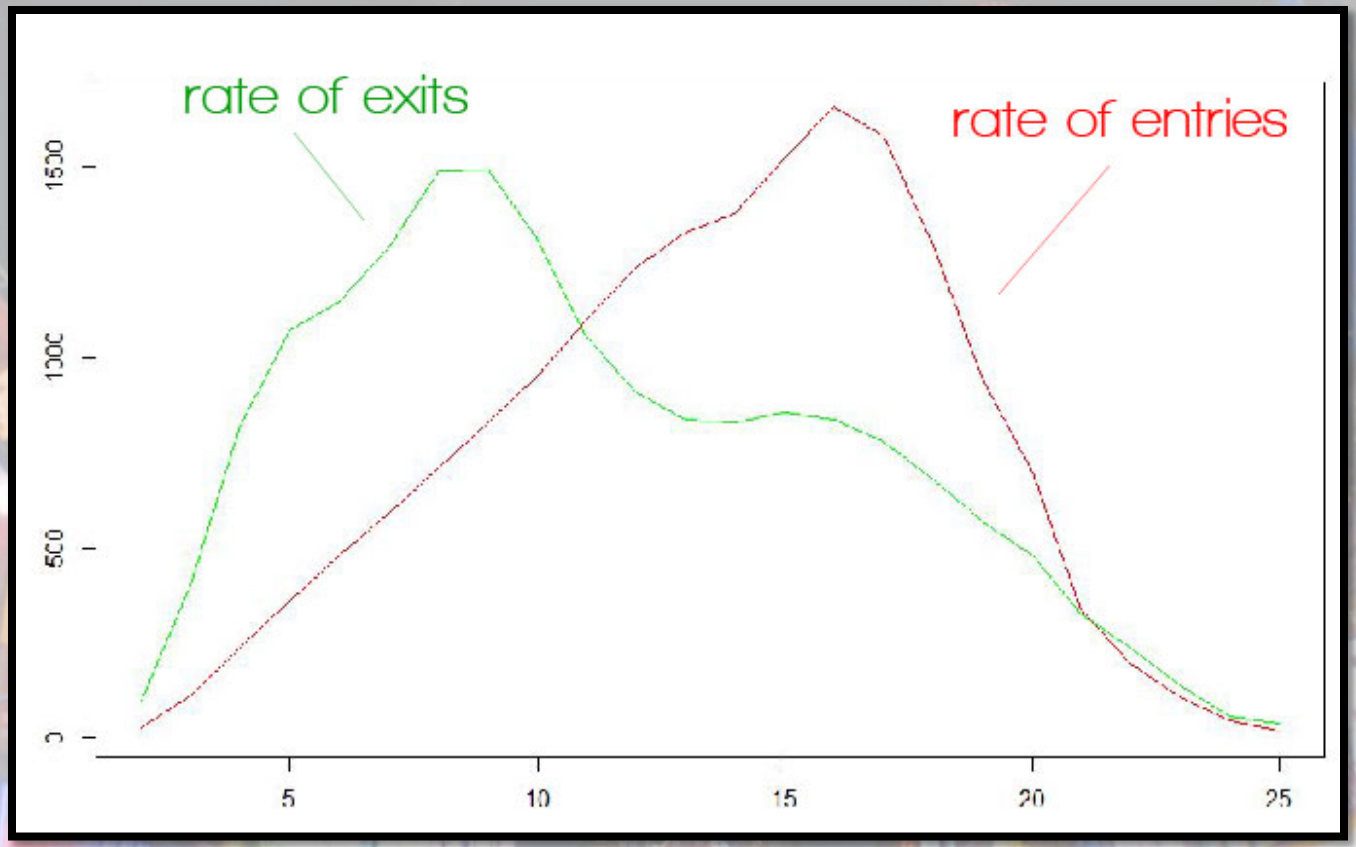
Determining the "low period" ("At what time do we start counting net import/export?"):



Past 5 Saturdays, 116th and Broadway

iv. Forecast Model

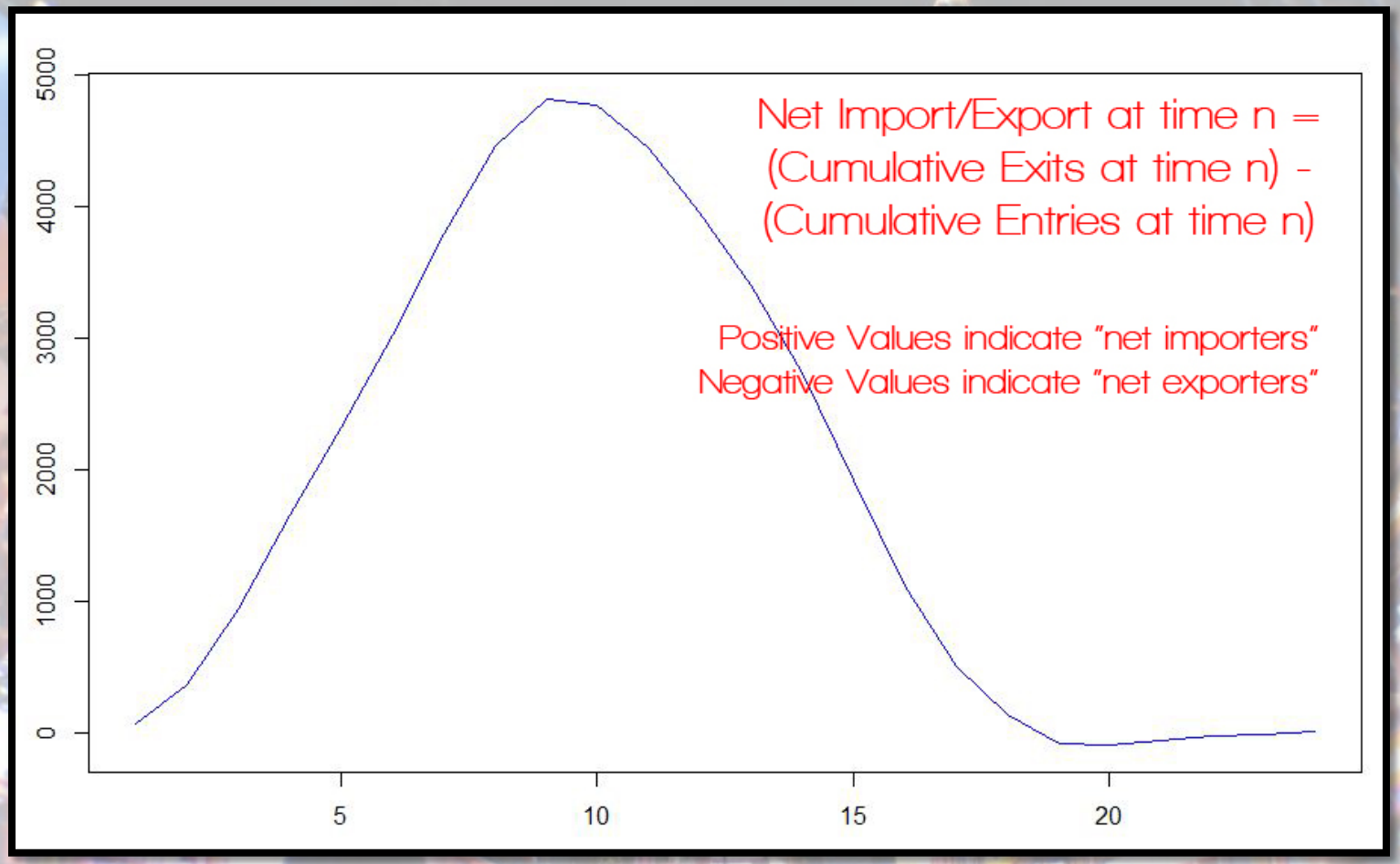
Normalizing the Exit Data:



Comparative Rates of exits and entries, predicted Friday, 116th and Broadway

iv. Forecast Model

Determining Net Import / Exports:



Net Import/Export, predicted Friday, 116th and Broadway

iv. Forecast Model

What the turnstile data can tell us about the dynamic population:

The area around 116th and Columbia is a "net importer", except for the hours of midnight to 4am on weekends in which it is a "net exporter".

Only 168th (CUMC) and 116th (Columbia) north of 59th Street (Central Park South) are considered "net importers". All other areas are "net exporters".

Rockefeller Center (47-50th St.) is the biggest importer in Manhattan, running a surplus of about 45K people on Wednesday afternoon, about 6-7 times that of 116th.

The Data indicates that the bus terminal at 42nd as well as Penn Station are very large "net exporters", indicating a flaw in the methodology. This can be attributed to people making multi-stop commutes.

Net Import/Export, predicted Friday, 116th and Broadway

iv. Forecast Model

What the turnstile data can tell us about the dynamic population:

Judging by subway activity, which time of the week is the "deadest" for 116th?

- Monday morning from 2-3 am (~30 people use the subway)

Judging by exits from the subway, at what time of the week is 116th (Columbia) adding the most number of people?

- Monday morning from 10-11am (~1,000 more people exit the station than enter per hour)

Judging by exits from the subway, at what time of the week is 116th (Columbia) adding the fewest number of people?

- Late Monday, early Tuesday from 1-3am

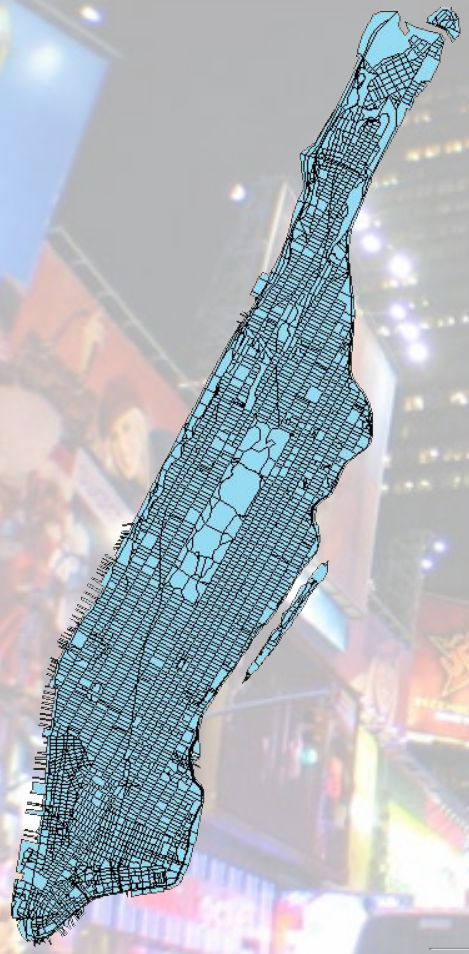
Judging by entries to the subway, at what time of the week are people leaving the area the quickest?

- Thursday afternoon from 6-7p

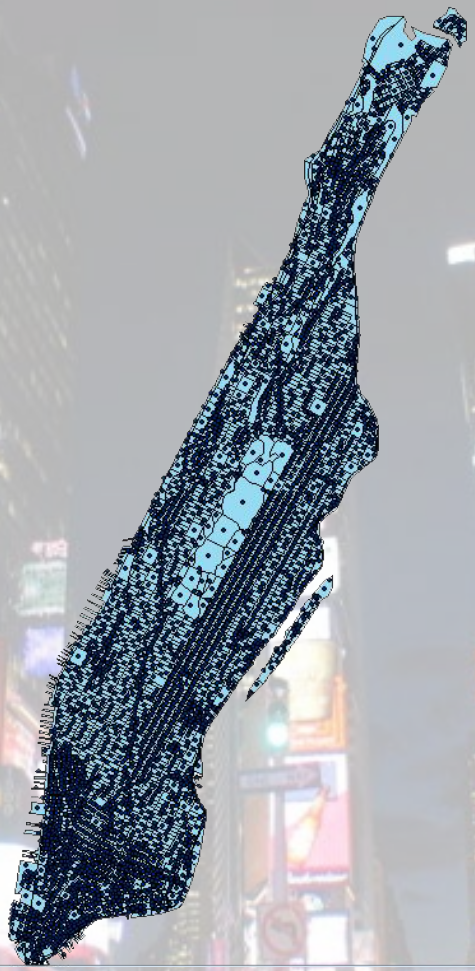
Net Import/Export, predicted Friday, 116th and Broadway

iv. Forecast Model

Attaching Import/Export values to a map:



Blocks with Population



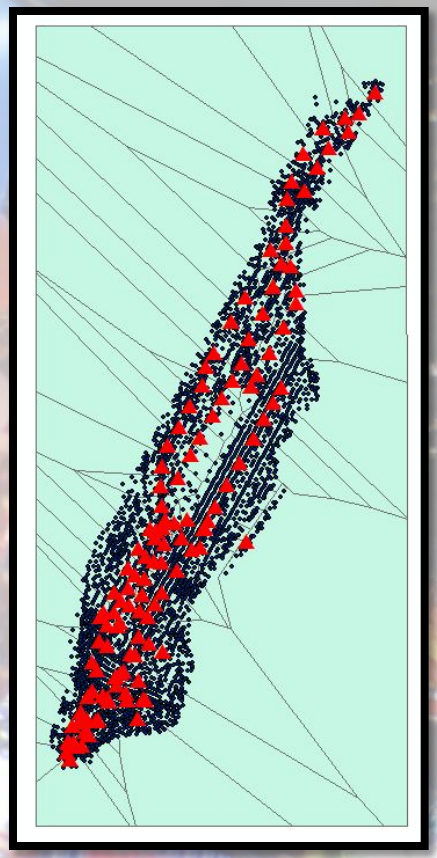
Blocks as Points

Subway Stations

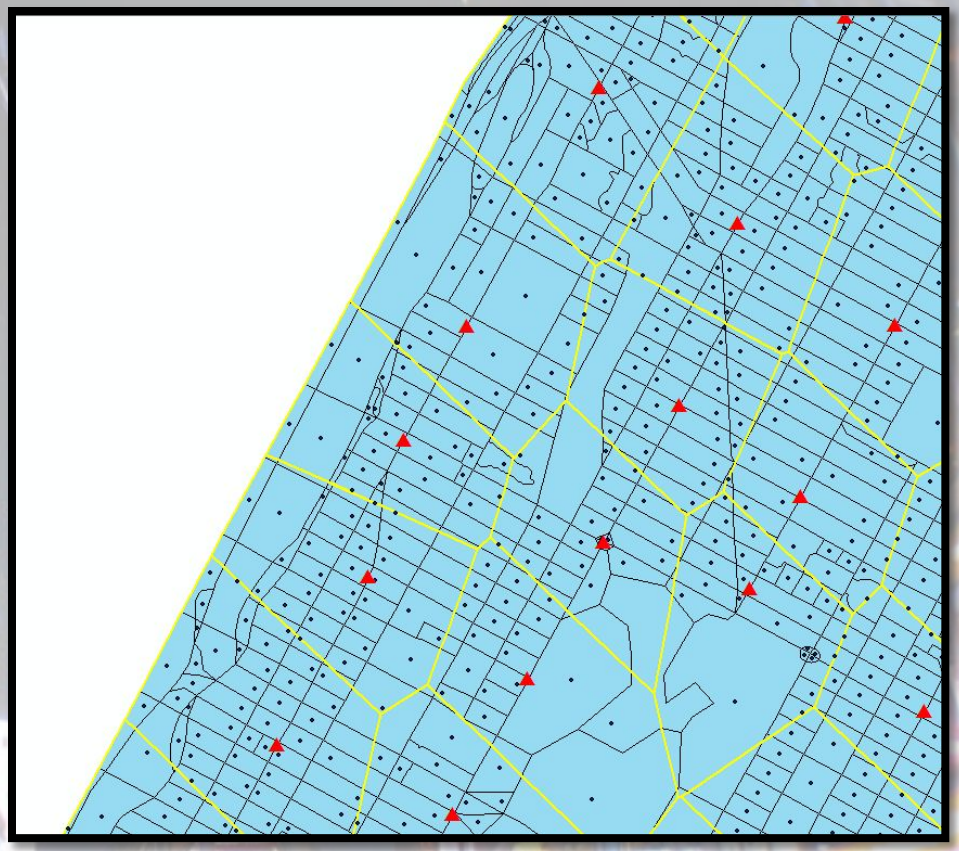
v. Geo-spatial Model

Part 2: Geo-Spatial Assignment

Attaching Import/Export values to a map:



Thiessen Polygons



Close-up of GIS modeling

v. Geo-spatial Model

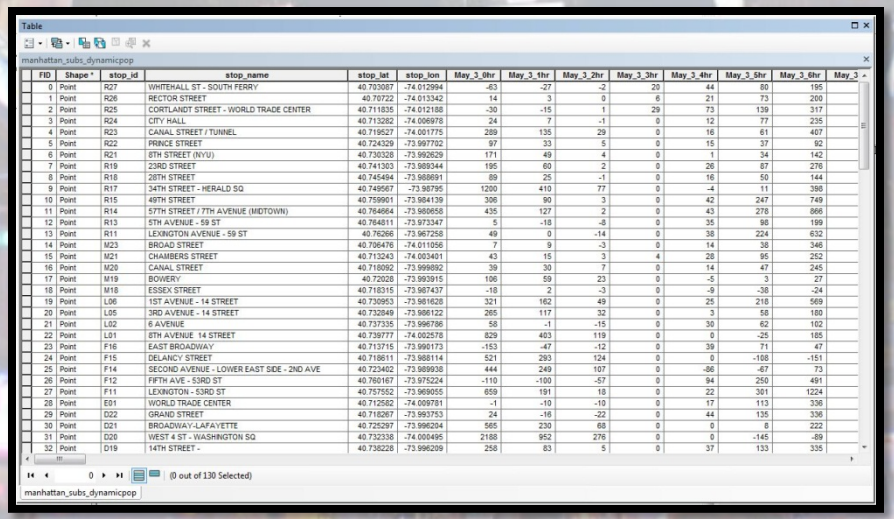
Part 2: Geo-Spatial Assignment

Assigning Blocks to Stations:

1. Join the blocks to the Thiessen Polygons

2. Divide the Import/Export total at a station by the count of joined blocks assigned to that station

3. Add/Subtract the divided import/export total to the block's base population, for every time period (uniform distribution)



Table

manhattan_subs_dynamicpop

FID	Shape *	stop_id	stop_name	stop_lat	stop_lon	May_3_0hr	May_3_1hr	May_3_2hr	May_3_3hr	May_3_4hr	May_3_5hr	May_3_6hr	May_3
0	Point	R27	WHITEHALL ST - SOUTH FERRY	40.703087	-74.012994	-63	-27	-2	20	44	80	195	
1	Point	R26	RECTOR STREET	40.70722	-74.013342	14	3	0	6	21	73	209	
2	Point	R25	CORTLANDT STREET - WORLD TRADE CENTER	40.711835	-74.012188	-30	-15	1	29	73	139	317	
3	Point	R24	CITY HALL	40.713282	-74.006978	24	7	-1	0	12	77	235	
4	Point	R23	CANAL STREET / TUNNEL	40.719527	-74.001775	289	135	29	0	16	61	407	
5	Point	R22	PRINCE STREET	40.724329	-73.997702	97	33	5	0	15	37	92	
6	Point	R21	8TH STREET (NYU)	40.730328	-73.992629	171	49	4	0	1	34	142	
7	Point	R19	23RD STREET	40.741303	-73.989344	195	60	2	0	26	87	276	
8	Point	R18	28TH STREET	40.745494	-73.988691	89	25	-1	0	16	50	144	
9	Point	R17	34TH STREET - HERALD SQ	40.749567	-73.98795	1200	410	77	0	-4	11	398	
10	Point	R15	49TH STREET	40.759901	-73.984139	306	90	3	0	42	247	749	
11	Point	R14	57TH STREET / 7TH AVENUE (MIDTOWN)	40.764664	-73.980658	435	127	2	0	43	278	866	
12	Point	R13	5TH AVENUE - 59 ST	40.764811	-73.973347	5	-18	-8	0	35	98	199	
13	Point	R11	LEXINGTON AVENUE - 59 ST	40.76266	-73.967258	49	0	-14	0	38	224	632	
14	Point	M23	BROAD STREET	40.706476	-74.011056	7	9	-3	0	14	38	346	
15	Point	M21	CHAMBERS STREET	40.713243	-74.003401	43	15	3	4	28	95	252	
16	Point	M20	CANAL STREET	40.718092	-73.999892	39	30	7	0	14	47	245	
17	Point	M19	BOWERY	40.72028	-73.993915	106	59	23	0	-5	3	27	
18	Point	M18	ESSEX STREET	40.718315	-73.987437	-18	2	-3	0	-9	-38	-24	
19	Point	L06	1ST AVENUE - 14 STREET	40.730953	-73.981628	321	162	49	0	25	218	569	
20	Point	L05	3RD AVENUE - 14 STREET	40.732849	-73.986122	265	117	32	0	3	58	180	
21	Point	L02	6 AVENUE	40.737335	-73.996786	58	-1	-15	0	30	62	162	
22	Point	L01	8TH AVENUE 14 STREET	40.739777	-74.002578	829	403	119	0	0	-25	185	
23	Point	F16	EAST BROADWAY	40.713715	-73.990173	-153	-47	-12	0	39	71	47	
24	Point	F15	DELANCY STREET	40.718611	-73.988114	521	293	124	0	0	-108	-151	
25	Point	F14	SECOND AVENUE - LOWER EAST SIDE - 2ND AVE	40.723402	-73.989938	444	249	107	0	-86	-67	73	
26	Point	F12	FIFTH AVE - 53RD ST	40.760167	-73.975224	-110	-100	-57	0	94	250	491	
27	Point	F11	LEXINGTON - 53RD ST	40.757552	-73.969055	659	191	18	0	22	301	1224	
28	Point	E01	WORLD TRADE CENTER	40.712582	-74.009781	-1	-10	-10	0	17	113	336	
29	Point	D22	GRAND STREET	40.718267	-73.993753	24	-16	-22	0	44	135	336	
30	Point	D21	BROADWAY-LAFAYETTE	40.725297	-73.996204	565	230	68	0	0	8	222	
31	Point	D20	WEST 4 ST - WASHINGTON SQ	40.732338	-74.000495	2188	952	276	0	0	-145	-89	
32	Point	D19	14TH STREET -	40.738228	-73.996209	258	83	5	0	37	133	335	

0 (0 out of 130 Selected)

manhattan_subs_dynamicpop

v. Geo-spatial Model

Part 2: Geo-Spatial Assignment

Adjustments and Assumptions:

1. Adjust nighttime population (Census base population) of 1.6 million to NYU's estimate of 2.05 million by multiplying base populations by 30%

2. Account for Population that does not use the subway:
   - assume dynamic geo-spatial distribution of non-subway users matches that of subway users
   - assume 50% of commuters use subway (NYU Wagner) -> multiply by 2
   - adjust import/export values by normalizing the highest total import level for Manhattan as a whole (800k on Wednesday afternoon) to that of the NYU study (1.6 million on a weekday) -> also multiply by 2

v. Geo-spatial Model

Application synthesis:

1. ArcMap
   - Square Buffer block points
2. ArcScene
   - "Extrude" z-value of square blocks
   - "heat-map" for population levels
3. Photoshop
   - frame animation

vi. Visualizing

Predicting Dynamic Population Graphically

vi. Visualizing

Predicting Dynamic Population Graphically

vi. Challenges

1. Normalizing Cumulative Exits to match Cumulative Entries
2. Nighttime population (30% over base census population) is distributed according to the census households
3. Net adjustment factor of 2 for import/export measures
4. Does not account for multi-stop commutes (imports from Jersey, CT, etc...)
5. No PATH, MetroNorth data
6. Mapping turnstile database to Spatiality shapefile results in 130 matched stations, meaning 20 stations are missing
7. Uniform distribution of net importers/exporters from subways to joined blocks

Addressing Assumptions

vi. Future Research

Addressing Assumptions

1. Curve-fitting entries, exits histogram to sigmoidal distribution
2. Fitting Autoregressive Integrated Moving Average (ARIMA) time-series model to account for periodicity, seasonality, and trends in subway usage
3. Using triangular distribution or overlapping buffer for assigning net importers/exporters to blocks
4. Better data, more data
5. Simply make the government enforce the release of anonymous call detail records (CDRs) and cell tower density data from telecomm companies

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

e4009: GIS for Infrastructure
Yuri Gorokhovich
Spring 2013

Justin Fung
Operations Research
jcf2164@columbia.edu

A Turnstile Population Model

Using MTA's Turnstile Database as a Proxy for Manhattan's Dynamic Population

Earth and Environmental Engineering

Fu Foundation School of Engineering and Applied Sciences | Columbia University