

Elevators in NYC

Data Science Project by Matthew Girard

Fall 2016



Context

- FiveThirtyEight.com had published a [story](#) about a dataset containing every elevator in New York City along with some analysis (Summer 2016)
- The dataset was [available](#) in Github and I wanted to explore the data for more questions and some *geocoding* experience



FiveThirtyEight

GitHub

Description of the dataset

- The data was received in November 2015 through a Freedom of Information request
- Dataset contains information on 76,088 elevators in New York City – covering all five boroughs
- Data includes the following for each elevator:
 - Street address
 - Number of floors
 - Type of elevator
 - Status
 - Zip code
 - Few other columns with lower quality (e.g., last inspection date)
- Overall – the raw data was of moderate quality. A slightly more 'clean' version of the file with two additional columns (latitude and longitude) was made available on Github, which is the version I used.

Questions explored:

- **Question 1:** How can I filter the dataset by zip code?
- **Challenge:** 11% of the dataset did not have zip code data
- **Approach:** through **geocoding** in **java**, fetch zip codes for elevators with latitude-longitude information (*raw code attached as text document*) for nearly 10% of the data set (~0.2% remained unresolved after geocoding)
- **Question 2:** How does the profile of elevators vary across the different boroughs? Is there a correlation to real-estate prices?
- **Challenge:** finding 'sample' neighborhoods and visualizing the data in the correct format
- **Approach:** use real-estate data from **Zillow**; use **MS Excel** to visualize as charts and run correlation analysis (*raw file attached as MS excel file*)

Updates to FiveThirtyEight analysis

In the original data set, ~11% of elevators did not have Zip code data
Using geocoding, I used geolocation (lat/longitude) to return Zip code vales that made most of this data usable
~0.2% of dataset (165 elevators) remained unresolvable as their geolocation was inconclusive

Original table from FiveThirtyEight story

The 2,000-elevator club

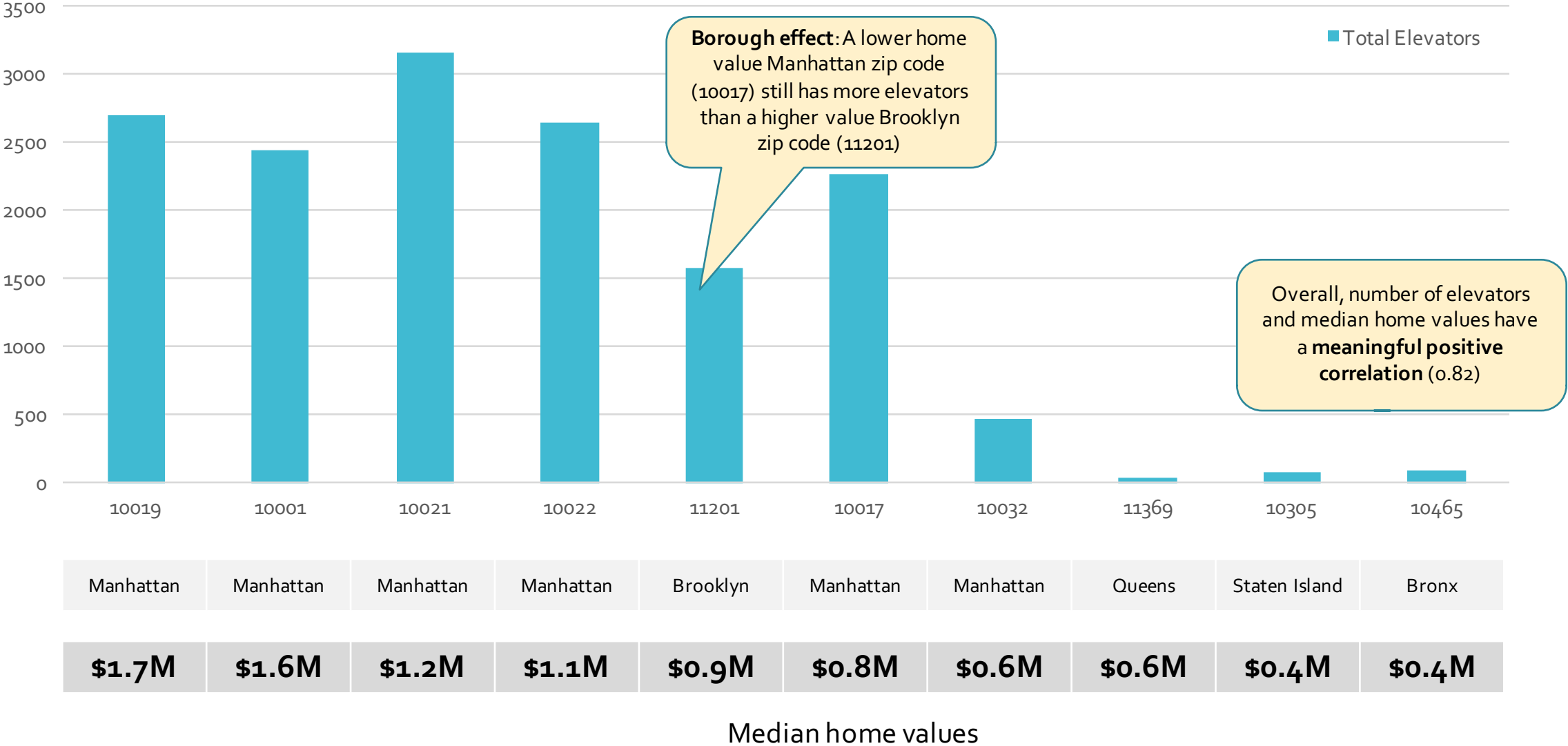
ZIP CODE	NEIGHBORHOOD	ELEVATORS
10021	Upper East Side	3,132
10019	Hell's Kitchen	2,600
10022	Midtown East	2,555
10001	Chelsea	2,294
10017	Midtown East	2,175

SOURCE: NEW YORK CITY DEPARTMENT OF BUILDINGS

Updated number of elevators

→	3,162 (+30)
→	2,696 (+96)
→	2,643 (+88)
→	2,437 (+143)
→	2,264 (+91)

Analysis of total elevators vs. median home values by zip code



The story of 3 neighborhoods – which type of elevator best predicts prosperity?

	10001 Chelsea	11201 Brooklyn Heights	10465 Throgs Neck (Bronx)
Passenger	1 , 877	1 , 269	75
Escalator	288	134	4
Freight	157	125	3
Dumbwaiter	46	26	2
Sidewalk	64	12	1
Other	5	5	2
Total	2 , 437	1 , 571	87

Median home values:	\$1.6M	\$0.9M	\$0.4M
---------------------	---------------	---------------	---------------

Both total elevator count as well as the count of different types of elevators show promise of being highly correlated to the Zip code

But, need to run correlations to decide their relative usefulness

Which type of elevator best predicts neighborhood prosperity?

	Correlation to median home value
Passenger	0.80
Escalator	0.91
Freight	0.91
Dumbwaiter	0.73
Sidewalk	0.75
Other	NA
Total	0.83

Best predictors of neighborhood prosperity – given well-off neighborhoods have retail/office spaces, commercial and condo atriums that typically have an escalator and freight elevators!