

Group J: NYC Real Estate

Progress book

How to connect 3 datasets

- Link through PUMA codes
- Subset construction/ ACS comparisons for years 2009 to 2019

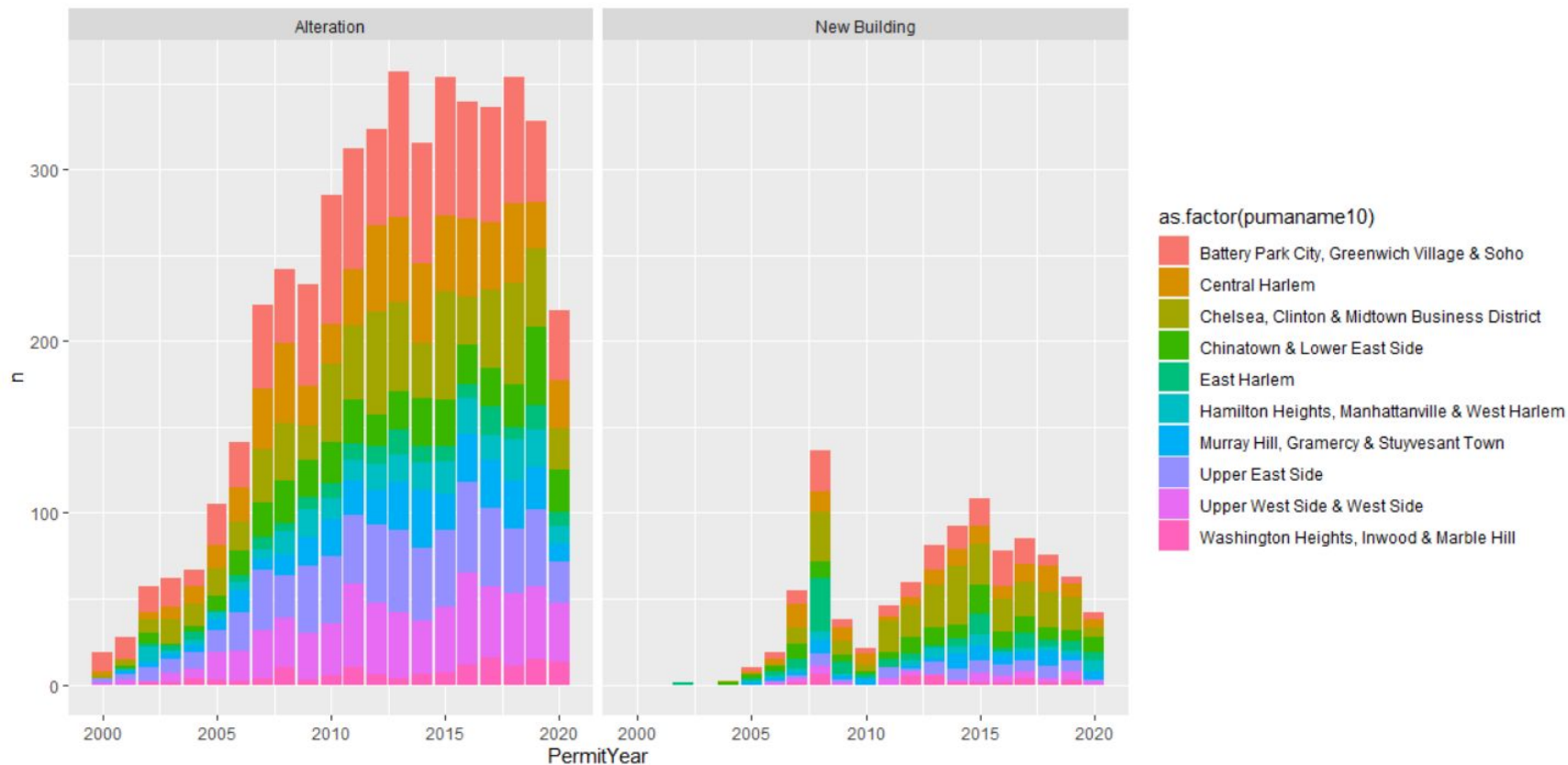
Home Values

Goal

Draw relationship between construction and home prices, ownership

- Examined the type of construction that should be included and filtered for only long-term residential (removed hotels)

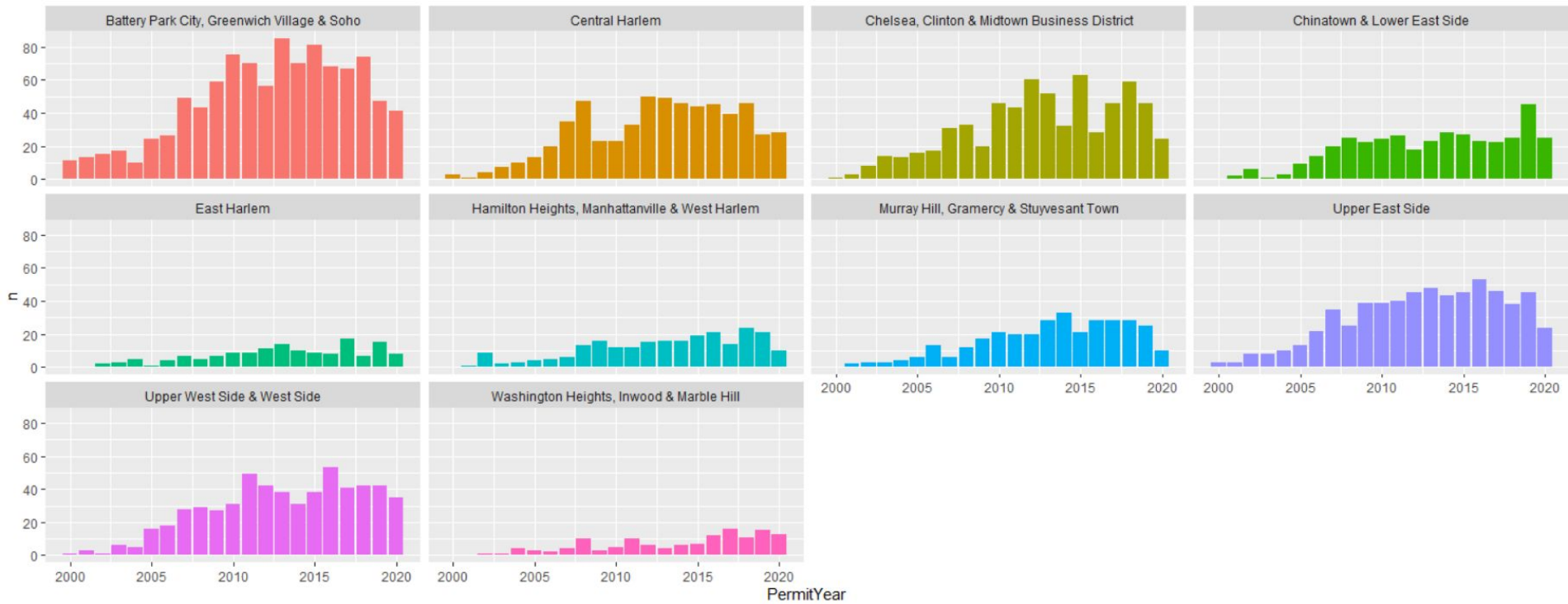
What do permit data look like?



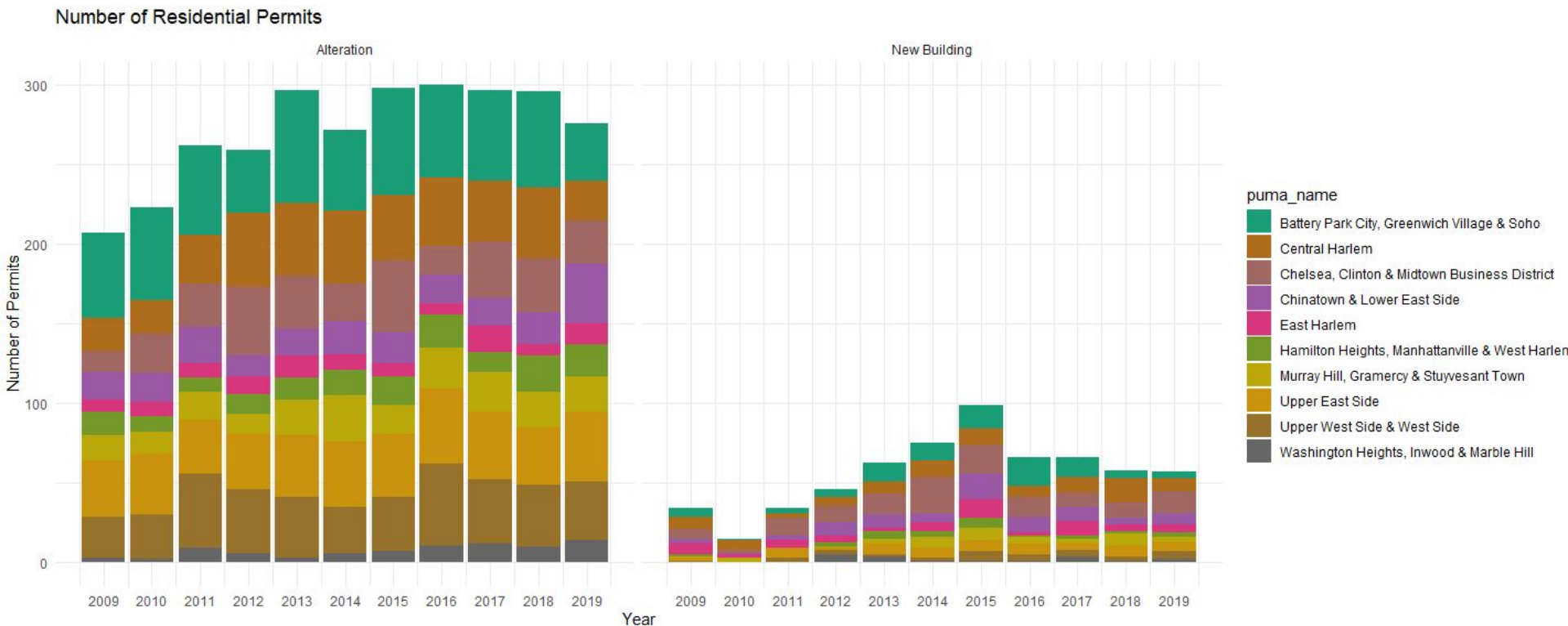
New Construction



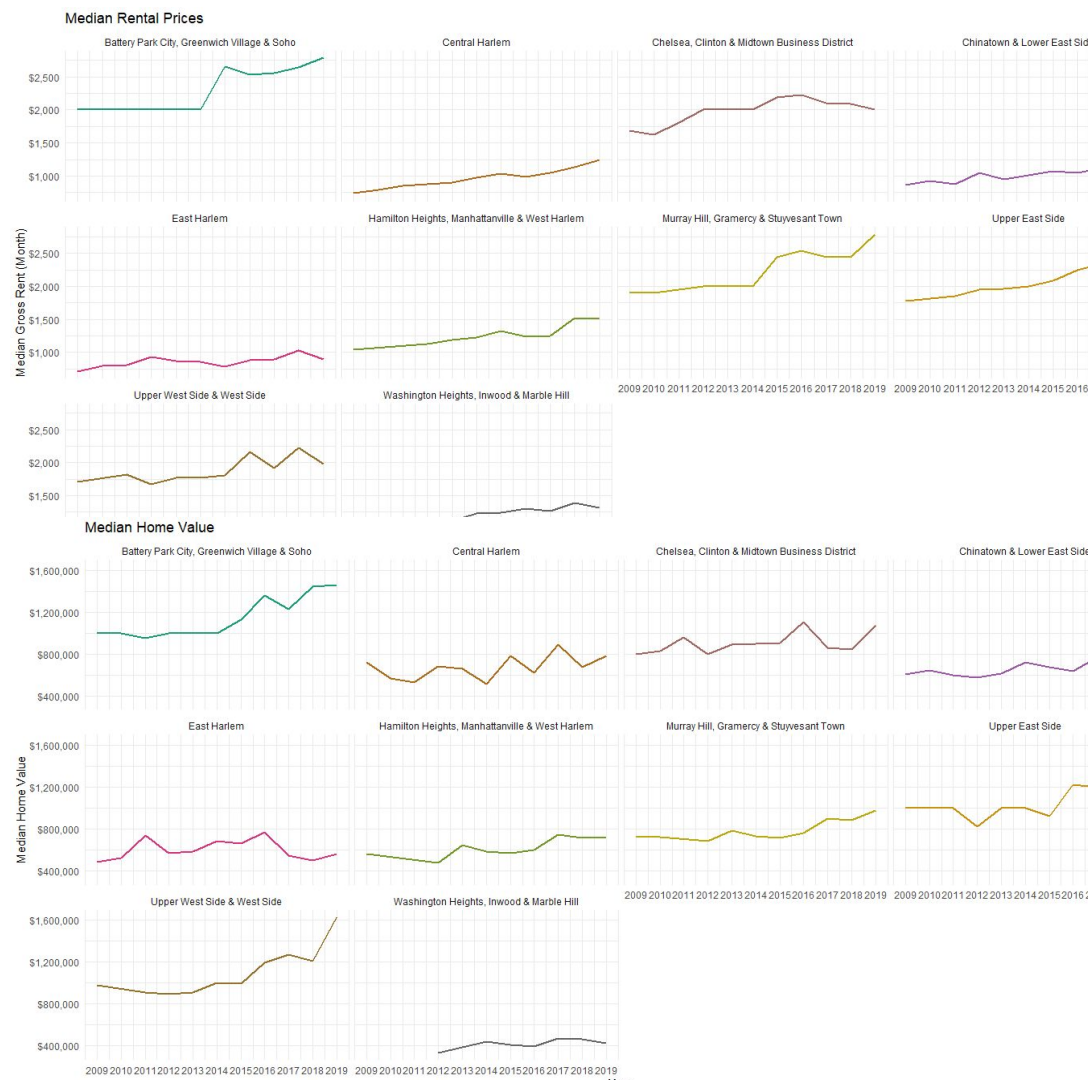
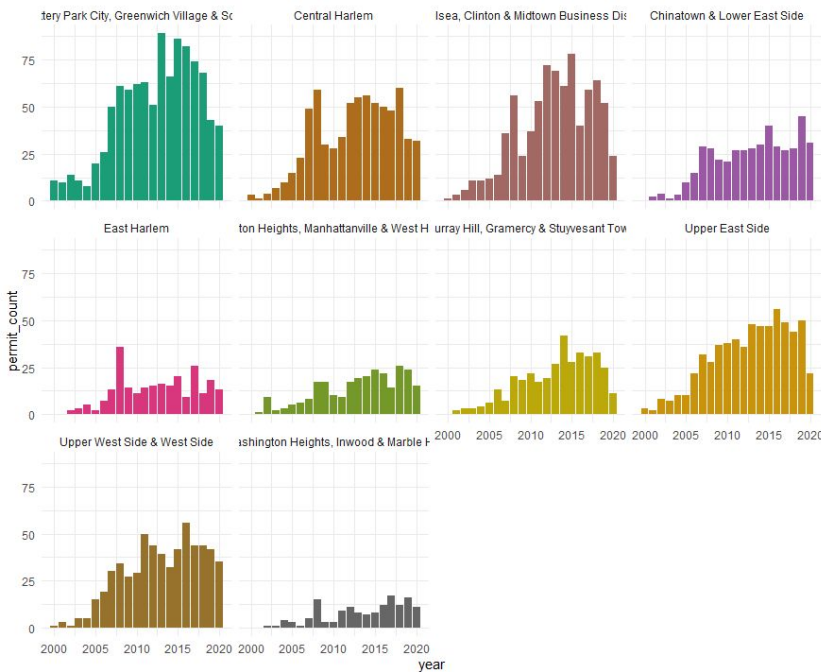
Alteration



Permit data for just residential?



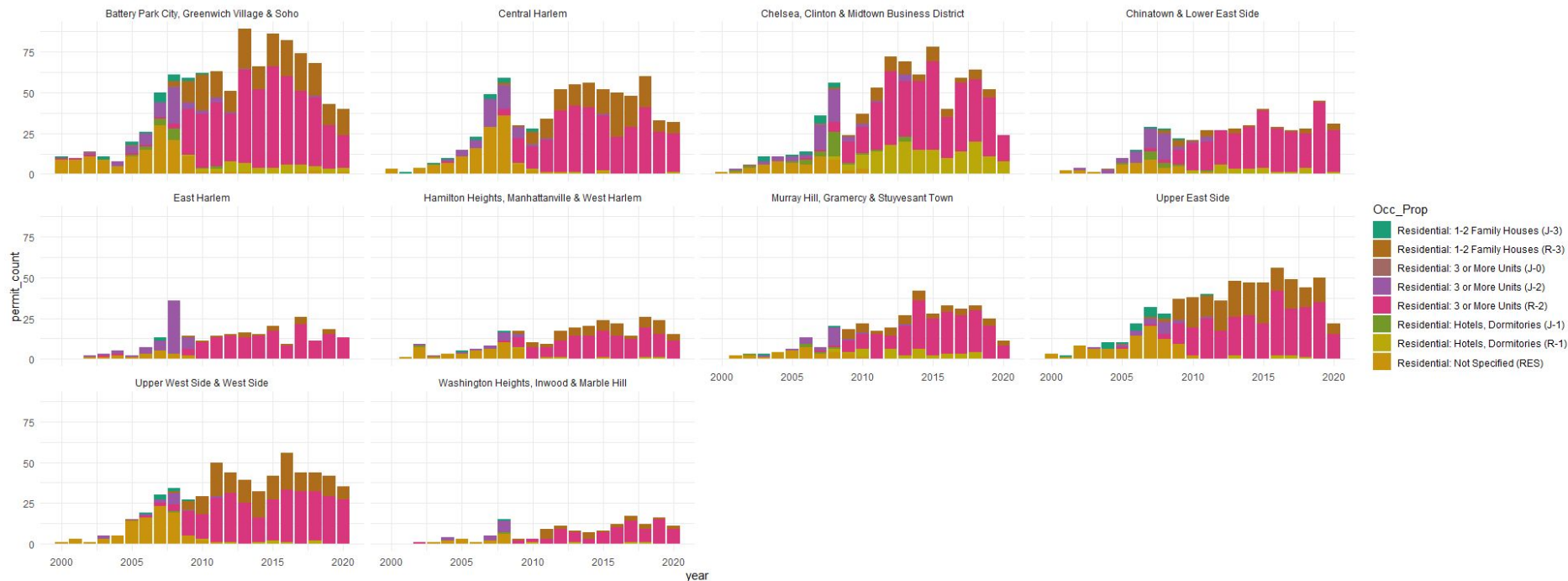
Residential Permit v prices



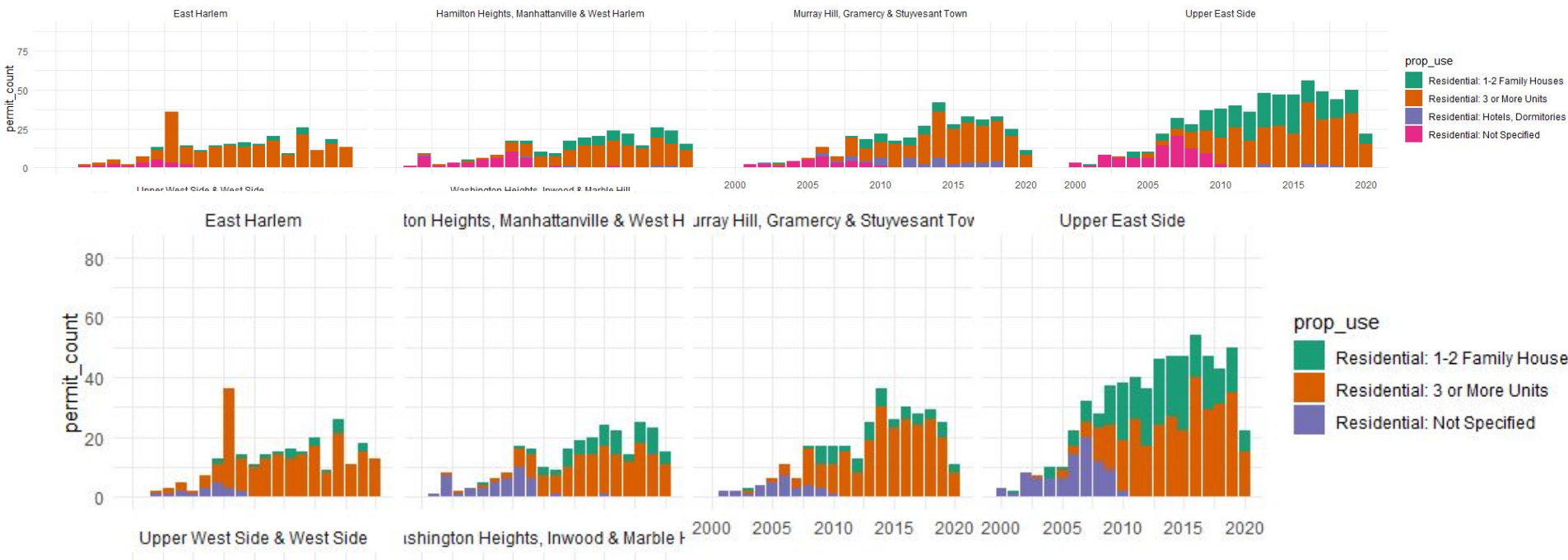
Residential Permit Type

Collapse R vs J types of permits

2008 Code	OCCUPANCY DESCRIPTION	1968 Code
R-1	occupied transiently (for less than one month)	J-1
R-2	more than 2 dwelling units on a long term basis (for a month or more)	J-2
R-3	not more than 2 apartments on a long term basis (for a month or more)	J-3

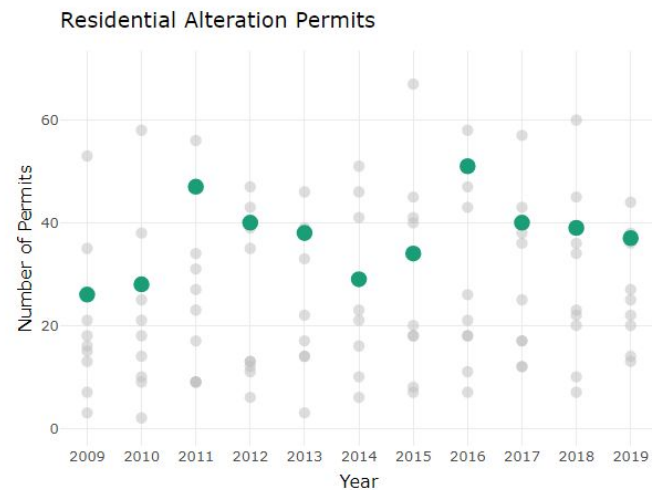
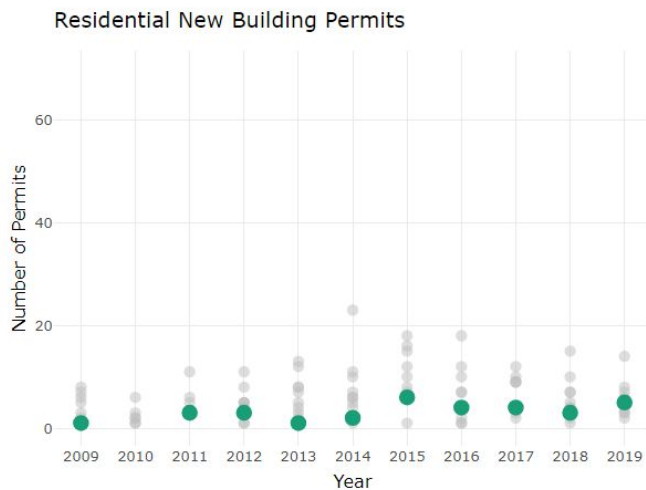
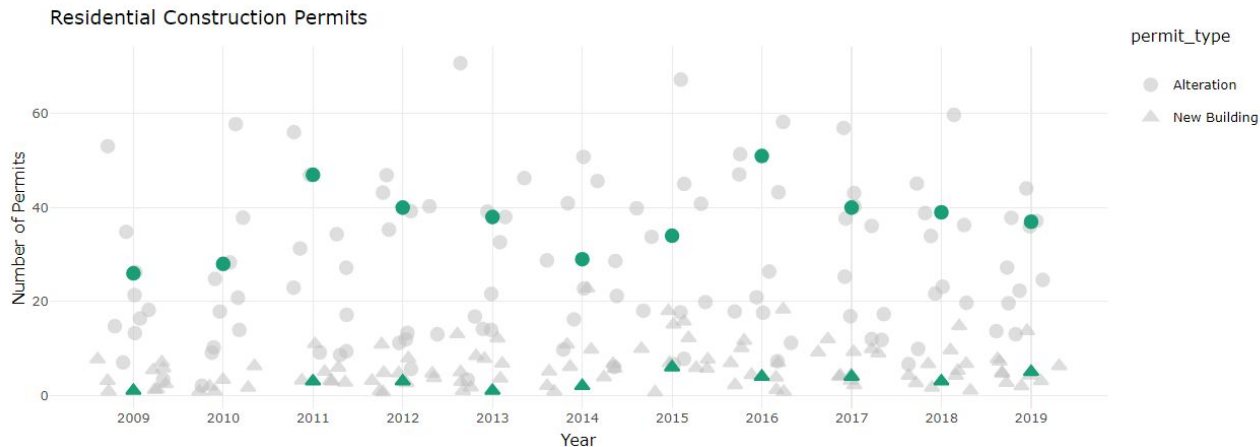


Removed Hotel Permits from Residential

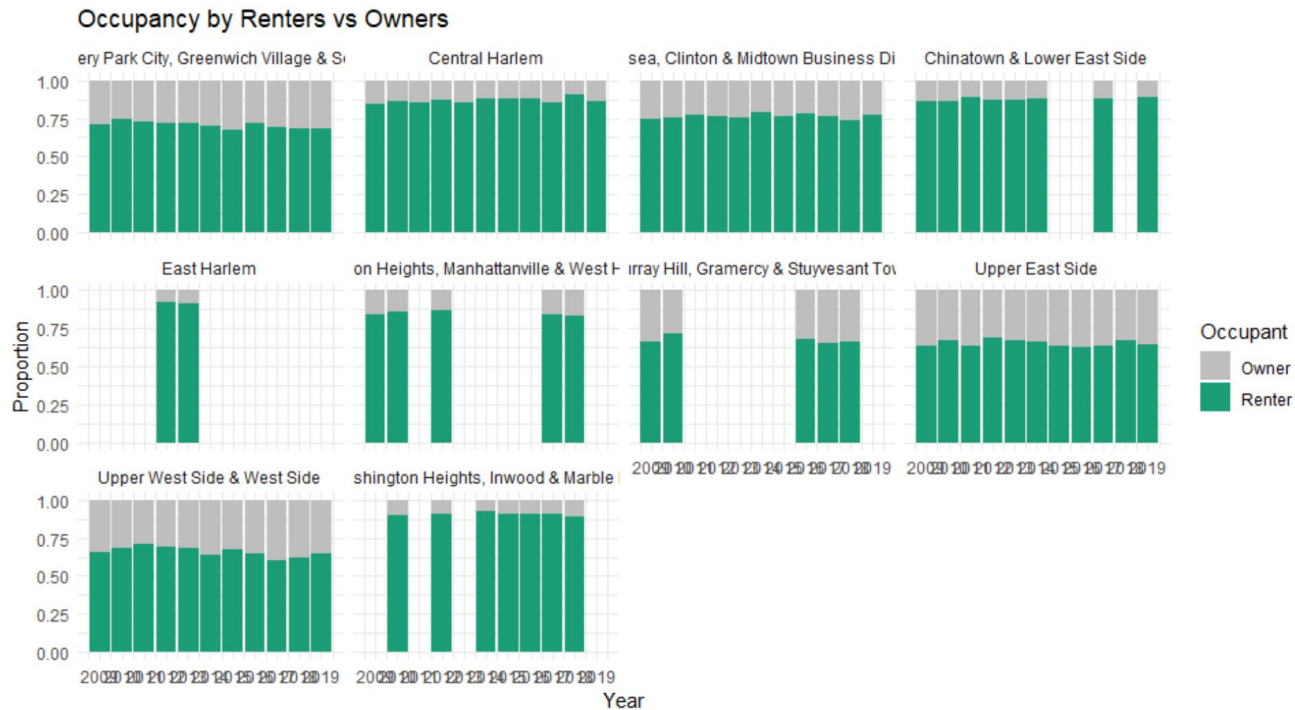


How to show permits

- One plot -- difficult to read
- Two plots with same scale, difficult to read new buildings



Owner occupancy vs renter



Construction

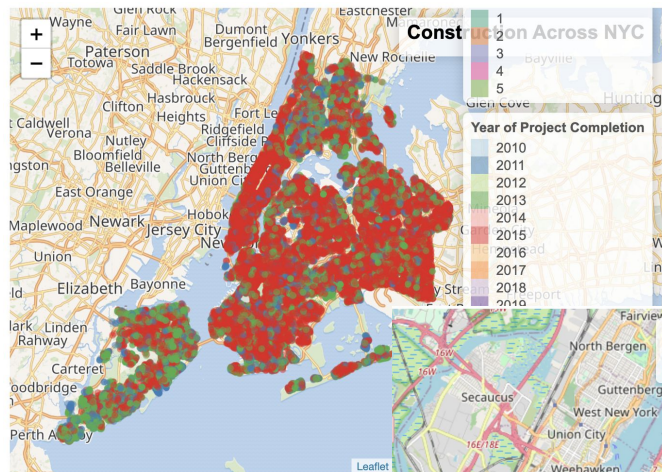
Construction Data Overview

- We ingested construction project post 2010 data to identify PUMA names by neighborhood
- Each construction project can be identified by a unique Job Number- various projects at the same address receive different Job Numbers
- Using permit year instead of completion year gives more insight into the project intention
- Initial Variables selected: Job_Number, Job_Type, ResidFlag, NonresFlag, Job_Status, PermitYear, Landmark, Boro, AddressNum, AddressSt, Occ_Init, Occ_Prop, Job_Desc, Ownership, NTAName10, PUMA2010, Latitude, Longitude)
- Initially thought of including landmark status in map pop-up info but there were too few occurrences for meaningful impact
- Cleaned out null values in permit years and occupancy types and projects that were demolitions
- Grouped projects by permit year groups in the maps to make easier to understand
- Focus on new building projects and building alterations
- Counted the number of each per PUMA neighborhood grouping to create the tree maps
- Differences in the initial occupancy type and proposed occupancy type
- Occupancy type change in the form of from -> to presented opportunity for network graph- used indegree for node size

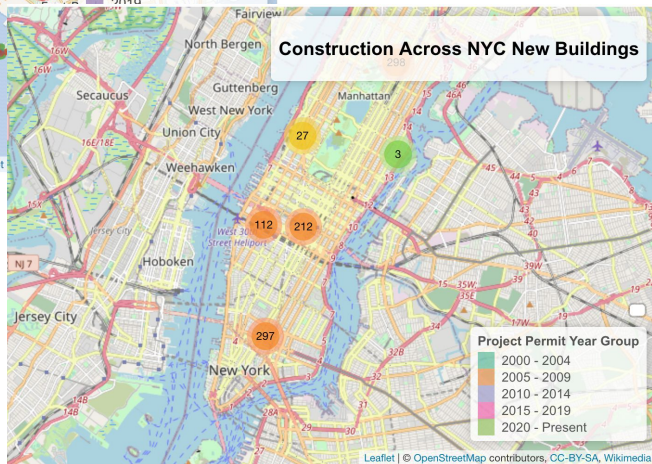
Interactive Maps

- Considered new building projects and building alteration projects, in two separate maps
- The latitude and longitude coordinates of each project provided an opportunity to use a map with exact project locations
- Initially included all five boroughs of NYC but the plot was too crowded and rendered too slowly due to large amounts of data
- Initially split out the projects by permit year, but it was not easy to make visual comparisons between years, so groups of four-year increments were used instead
- Initial maps did not use default zoom on Manhattan which made it more difficult to read the maps

Before



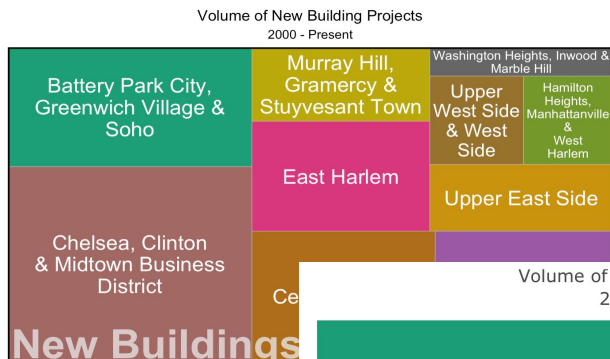
After



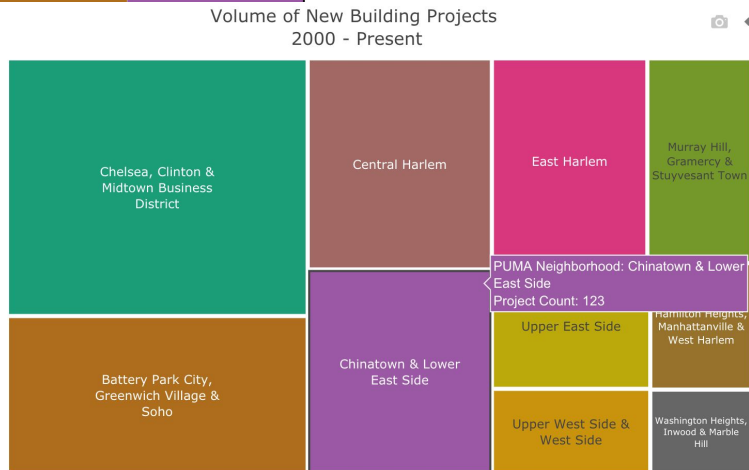
Interactive Treemaps

- Treemaps enabled us to compare the differences in types of projects between each neighborhood group
- Decided to separate into two treemaps, one for each job type, since the aim was to compare the job types between neighborhoods as opposed to overall job type quantity
- Each square within each of the treemap represents a neighborhood grouping, with the size of each square making it simple to visually compare the volume of new building projects with the volume of building alteration projects

Before (static)



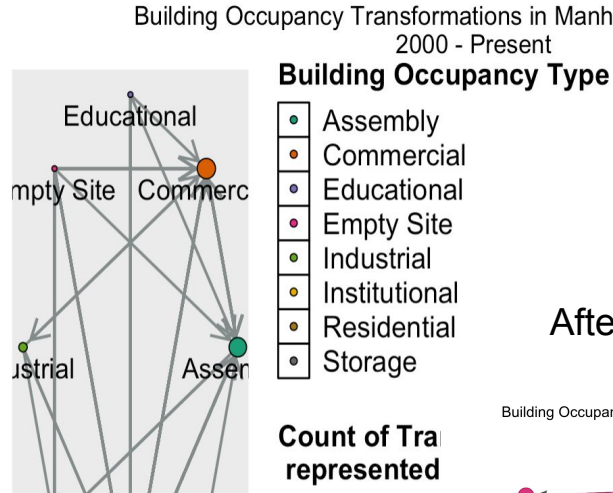
After (interactive)



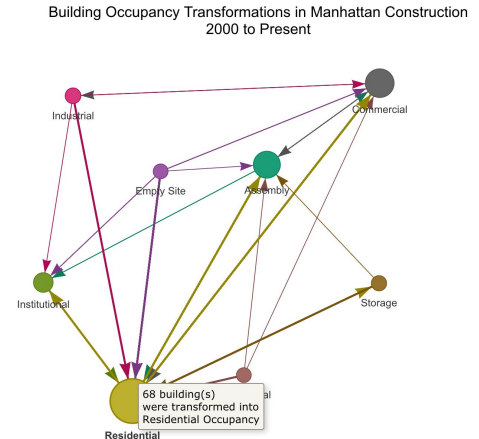
Interactive Network

- As stated in class, network graphs should be used with caution, and only in cases where the data is of appropriate structure
- In the data, all projects had an initial occupancy type, and then a proposed occupancy type, to indicate a change in occupancy type
- This data could be easily converted into a **from** → **to** format as is needed for **directed** networks
- To **avoid self-loops in the network**, projects where the occupancy type remained the same before and after the project were not considered
- However, the graph was not simplified (removal of duplicate edges) to better represent the volume of types of transformations
- To illustrate the most common transformation types, node sizes were determined by the nodes' in-degree measures
- The network was made interactive to allow for further manipulation, with an info box that pops-up that lists the number of transformations

Before (static)



After (interactive)

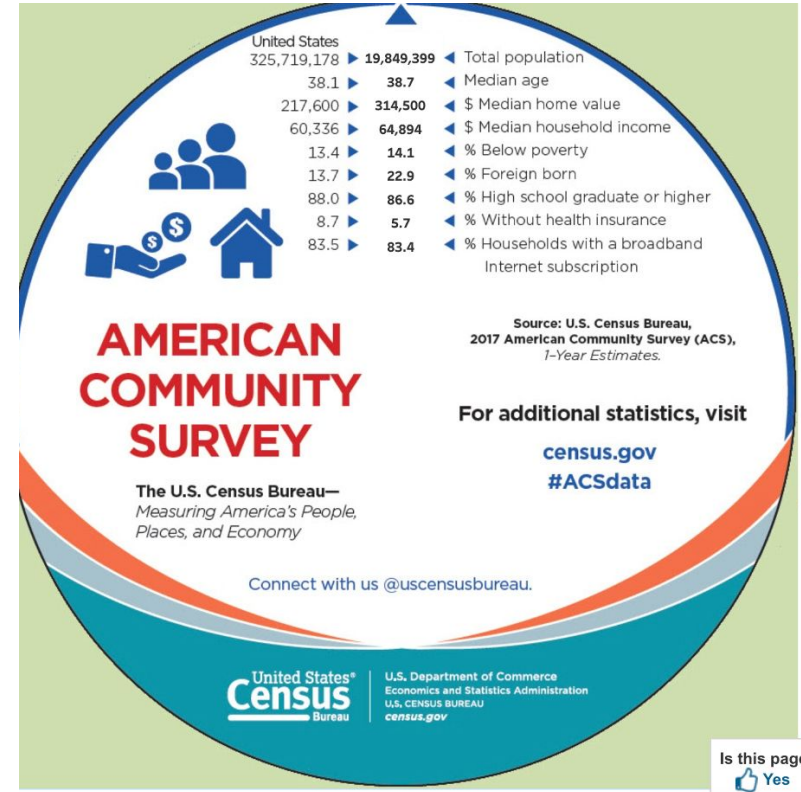


Neighborhood Demographics

Demographics based on ACS data

ACS data:

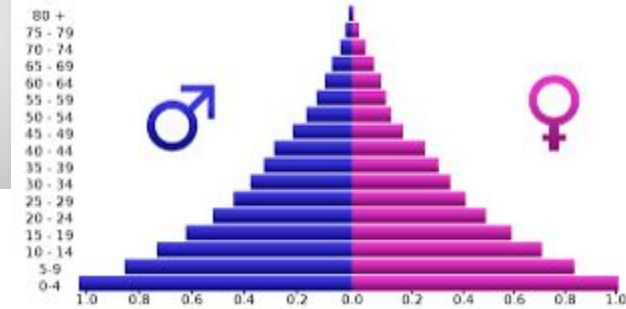
- **American Community Survey (ACS) Data** is the most extensive nationwide survey currently available. From its annual releases, we are able to examine the city's detailed demographic, socioeconomic, and housing characteristics.
- We chose 1-year (2009-2019) and 5-year estimate (2009-2019) data because they are most recent and can represent the most recent changes of city demographics.



Choose the right features to present

I chose **median age, population, household income, gender and where people moved from** as the key stats to present our demographics.

These metrics are high-level metrics that can give readers a great overview of what the population consist of, and relate these attributes to the neighborhood changes over time.



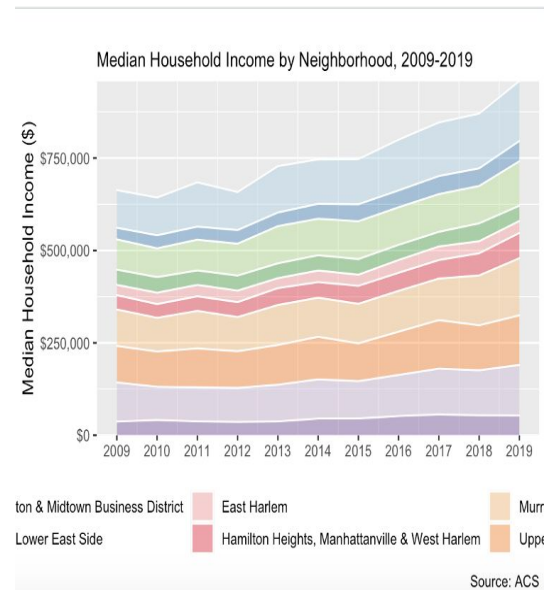
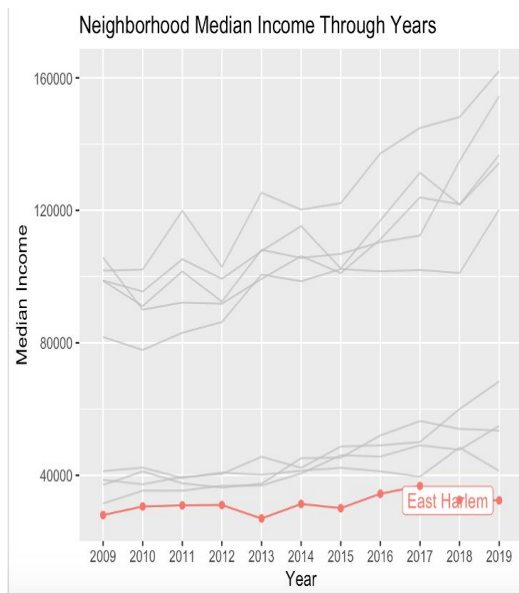
Choose the right visualizations

E.g. for income, I first used the **line chart** and then changed to **stacked chart**.

The reason I chose this is because it is easier to compare across different sub groups and easy to see the neighborhood overall income over time.

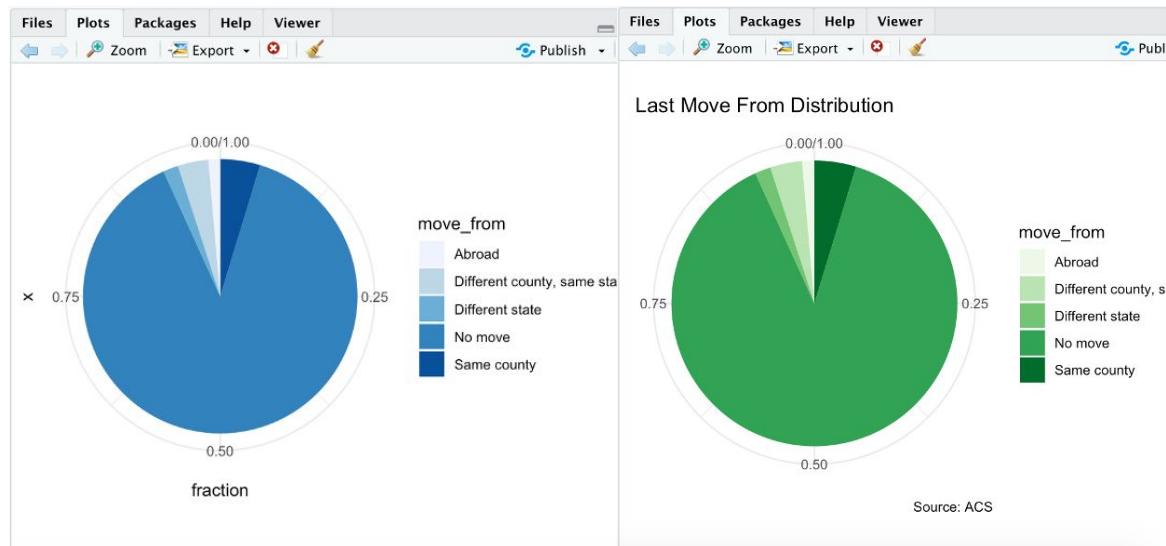
For median age and population, a **line chart** makes more sense to see the changes in time series and compare the changes across groups.

For gender and move from, the focus is to see the composition and a **pie chart** makes more sense.



Refine aesthetics and positioning

I then refined graphs, change graph details e.g. legend position, size, title, theme, color palette etc. to be consistent with other parts of the project.



Fit webpage

Final step: I resized the graphs to fit the web page and other elements on the page more nicely

Explore Neighborhood Demographics

The below graphs show the representative demographics of 10 the neighborhood in Manhattan. These include: median age, population, median income, gender and where they moved from. The demographics painted a vivid picture of what type of people live in and consist of each neighborhood, and how their migration pattern affects the alteration and progression of city's real estate.

Choose Neighborhood:

Chelsea, Clinton & Midtown Business District ▼

Neighborhood Median Age Through Years



Neighborhood Median Population Through Years



Manhattan Construction

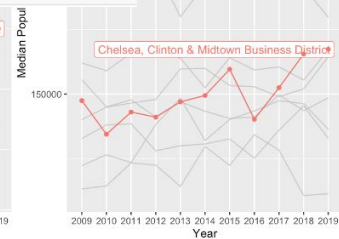
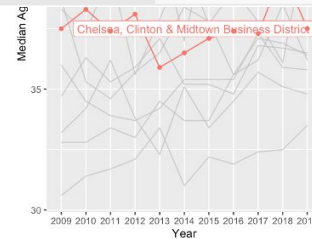
Main

Construction

Home Value

Demographics

Neighborhoods in Words



Neighborhood in Words

Wordcloud - comparison between 2010 and selected year

Version 3

How are neighborhoods described through time?

The four graphs below show how a Manhattan neighborhood is described from 2010 to 2021. The word cloud compares the words that are used to describe a neighborhood in 2010 vs. 2021. The data was collected using the way-back-machine Python API that queries historical versions of a certain webpage stored in Internet Archives.

Select the year to compare against 2010 below and the number of words to use in the comparison.

Chosen Year vs. 2010



Choose Neighborhood:

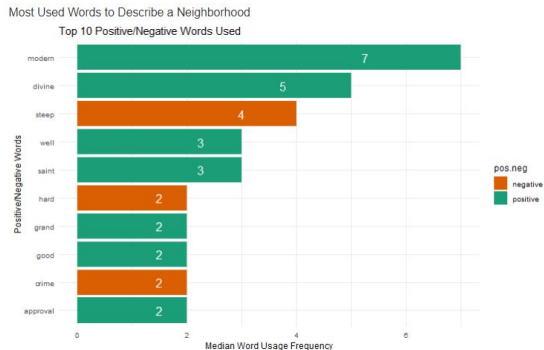
Morningside Heights

Choose a Year

2010 2019 2021

Maximum Number of Words:

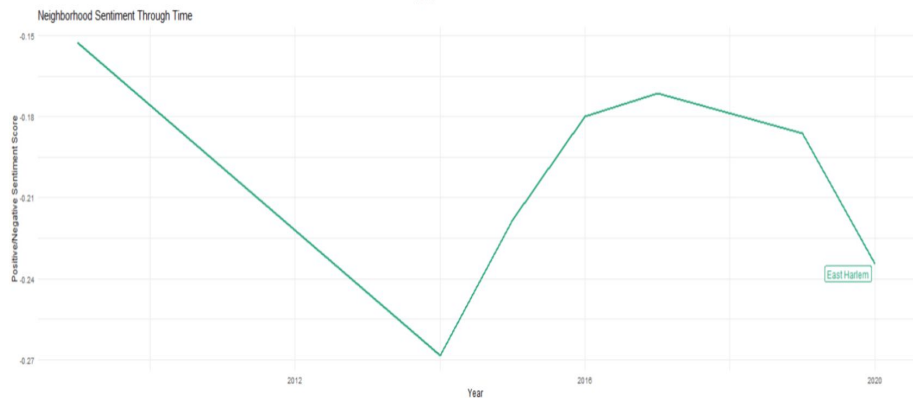
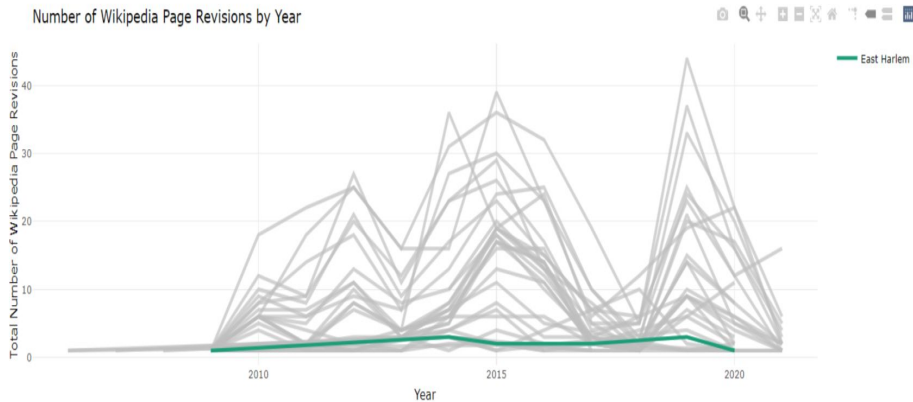
50 100 300



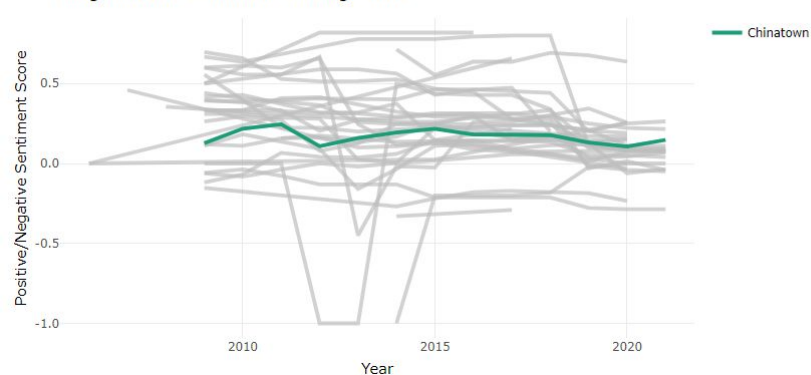
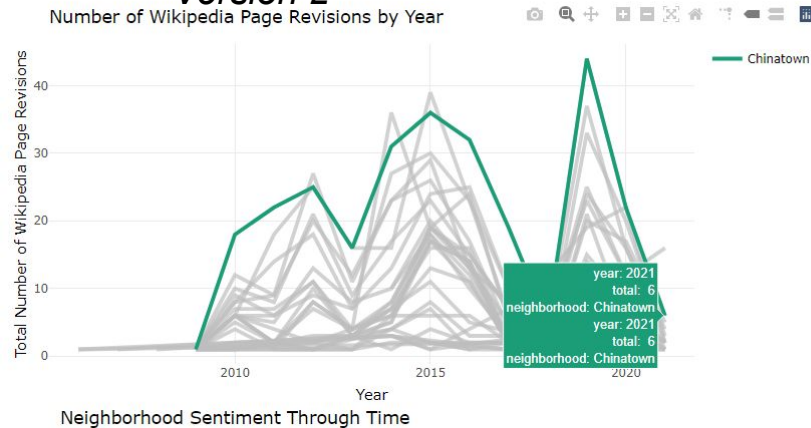
- Added chart on the right with top 10 words used that have positive/negative sentiment and will change depending on selected year
- Added label within charts for word frequency
- Formatted color to match theme
- Reverse ordered words descending by frequency
- Added text to describe page

Graphs Describing Wikipedia Text

Version 1



Version 2

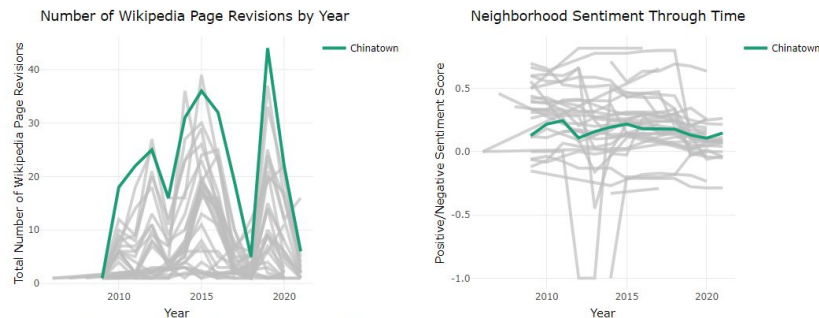


Graphs Describing Wikipedia Text

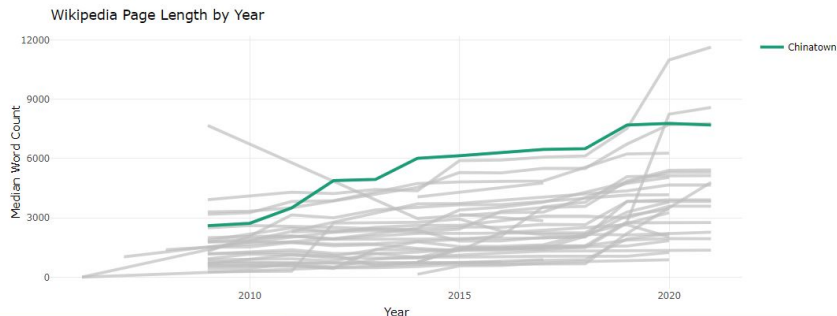
- Added wikipedia page word count by year
- Reformat date x-axis to be clearly legible

Version 3

The below plots show the number of revisions to a neighborhood's wikipedia page through time as well as the sentiment score using a positive/negative dictionary.

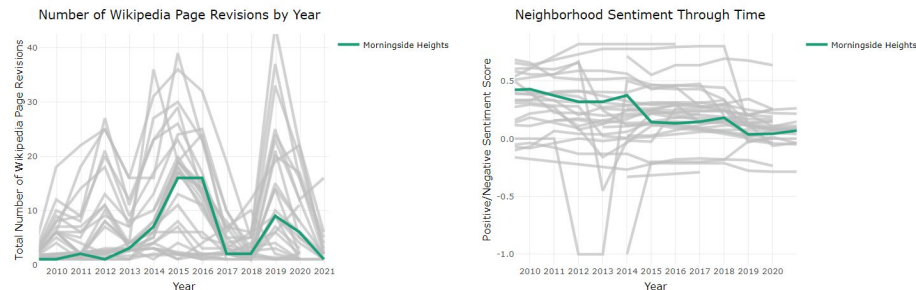


The below shows the number of words used to describe a neighborhood's wikipedia page through time.



Version 4

The below plots show the number of revisions to a neighborhood's wikipedia page through time as well as the sentiment score using a positive/negative dictionary.



The below shows the number of words used to describe a neighborhood's wikipedia page through time.

