QMSS5063 Data Visualization Final Project Book

Group E: Yiming Zhang, Paul Sheng, Emily Li, Junhan Chen

About This Project:

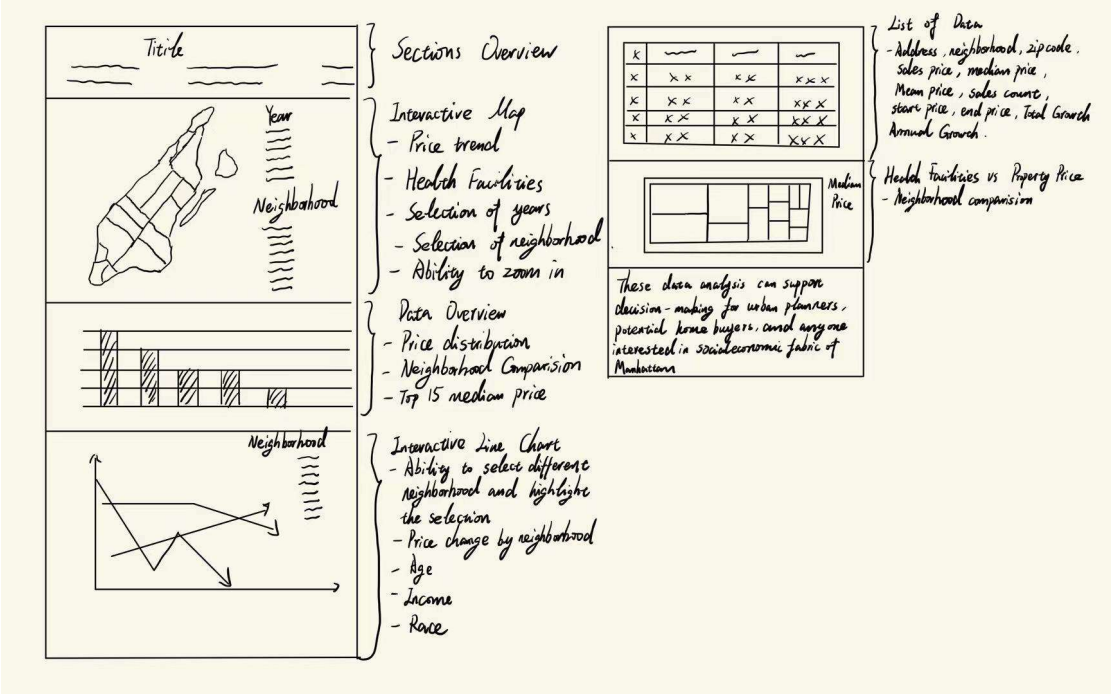
Our project analyzes the dynamics of real estate sales in Manhattan, New York City. Subject to dramatic real estate price fluctuations, we seek to better understand how property values change over time and space, and what underlying factors may influence these changes.

Examining neighborhood-level sales data alongside demographic profiles and proximity to healthcare resources, we present patterns and correlations that potentially shape the real estate market to provide insights on how social infrastructure and community characteristics intersect with property valuation.

This analysis can support decision-making for urban planners, potential homebuyers, and anyone interested in the socioeconomic fabric of New York City, for specifically, Manhattan. To make our findings accessible, we used a variety of visualization tools including geospatial maps, data tables, histograms, line charts, scatter plots, treemaps to illustrate trends in pricing, shifts over time, and neighborhood-level disparities.

Draft:

Our draft detailed a visualization project framework for population migration research. The design includes core modules such as interactive maps, database construction, neighborhood analysis, and price monitoring. I strive to use visualization tools to show the changes in housing prices between different neighborhoods in Manhattan, and explore the internal factors that affect its changes.



Findings:

Our research project revealed significant spatial and temporal patterns by analyzing sales data from 2015-2023. In terms of spatial distribution, downtown and central areas, especially SoHo and Tribeca, continue to maintain the highest median real estate prices. However, Washington Heights and Hamilton Heights have seen home prices increase by more than 100% over time, making them among the more expensive neighborhoods in 2023.

We found that Manhattan's demographic structure has a complex correlation with housing prices. Although there is only a weak negative correlation between the median age of the population and housing prices, income levels show a more obvious positive correlation with housing prices. High-income neighborhoods such as Battery Park City, Greenwich Village, and the Upper East Side continue to maintain high housing prices, while low-income neighborhoods such as East Harlem and Washington Heights have relatively low housing prices. Analysis of racial composition shows that neighborhoods with a higher proportion of whites generally have higher housing prices, but this pattern has changed slightly during the study period, reflecting Manhattan's socioeconomic transformation.

Surprisingly, there is a weak to moderate negative correlation between medical facility density and housing prices. Neighborhoods with more medical facilities tend to have slightly lower median real estate prices, possibly because institutional buildings take up a lot of space, limiting high-end residential development. This finding suggests that the configuration of public infrastructure is more based on population needs and density rather than wealth levels, further revealing the complexity and multidimensional characteristics of Manhattan's real estate market.

Challenges and Limitations:

We encountered several challenges during data collection and analysis. First, when integrating multi-source data sets, inconsistent definitions of geographic boundaries led to incomplete matches between real estate data and demographic data. For example, the income growth pattern of some neighborhoods did not match the growth of housing prices, which may be due to differences in the definition of neighborhood boundaries between different data sources. Second, the American Community Survey was missing data in 2020 due to the pandemic, and we needed to use interpolation or supplementary analysis methods for this period. Finally, it was also challenging to deal with the biased effect of high-value real estate sales on average prices. We used both median and mean statistical methods to mitigate the impact of this bias.

Regarding limitations, first, due to project scope considerations, we only focused on Manhattan rather than the entire New York City, which may have missed the interactive effects of the real estate market in surrounding areas. Second, although we analyzed factors such as age, income, race, and medical facilities, we did not cover other key variables that may affect housing prices, such as educational resources, public transportation accessibility, crime rates, or green space ratios. In addition, correlation analysis only reveals statistical relationships between variables, not causal relationships, and we cannot determine whether demographic characteristics cause changes in housing prices or whether changes in housing prices attract specific population groups.