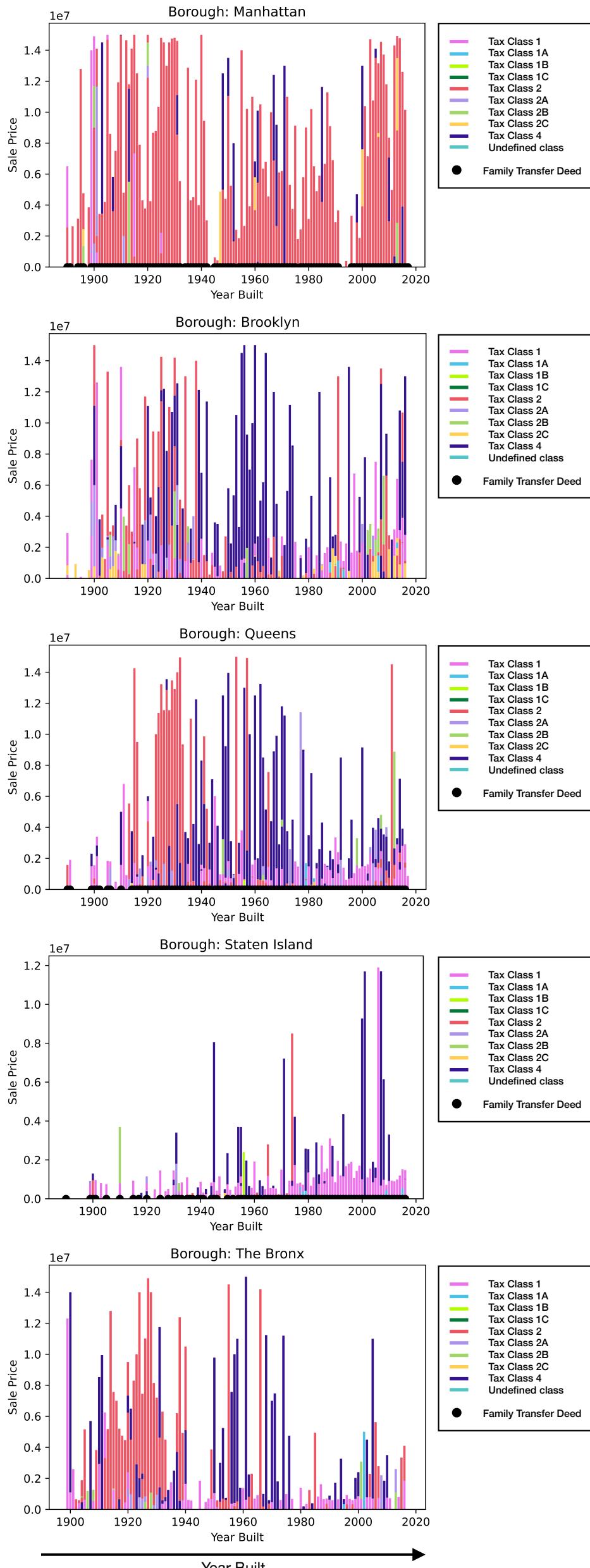


New York City House Sale Price (2016-2017)

Relation between the Sale Price and Year Built of houses across 5 Boroughs in New York City



Visualization Design:

- **Data Attributes Utilized:**
 - **Borough** (Nominal)
 - **Tax Class At Present** (Nominal)
 - **Sale Price** (Quantitative)
 - **Year Built** (Quantitative)
- **Type of Visualization:**
 - **Bar Chart (Color-coded)**
 - **Metrics:**
 - **X-Axis:** Year Built (**SHARED AXIS** - Trellis Plots)
 - **Y-Axis:** Sale Price
 - **Color-coding:** Tax Class At Present
- **Purpose / Intent of this visualization:**
 - To clearly present the relation between Sale Price of a House and the Year it was built, in a particular Borough in New York City.
 - To effectively recognize the price trends in accordance with the age of the construction. For example, in Staten Island, it is evident from the graph-3 that the prices of newly-built houses are significantly higher compared to old houses in the borough.
 - The Tax Class At Present has been color-coded in order to better understand the distribution of Tax class applicable through different boroughs. For Example, it is Evident from the graph of Manhattan Borough that the majority falls under the Tax Class : 2 at present. This information may be helpful to the New York Finance Department.
 - To easily compare market prices for a house built in a specific Year across different Boroughs, which may be useful from the perspective of Real Estate Investors.
- **Data Cleaning Practices:**
 - **Year-Built:** Years prior to 1850 have not been utilized for this visualization. There were entries with Year Built as 1111, and 1680. Entries containing both theses values have been pruned as the New York City itself was founded officially in 1898. Accordingly, the Department of Finance's data for the building built in 1680 might not be accurate. Moreover, the entries with year 1111 are most likely noise in the data. Moreover, as the data presented pertains to be for the Year 2016-2017, both the above mentioned years are seemingly irrelevant in this prospect as currently, the scope of study would be recent times in order to infer from the data and reach a decision.
 - **Sale-Price:** The Sale Price has been considered till 14 Million USD. It is necessary to note that the purpose of this visualization, as mentioned previously, is to understand the trend and classify numerical data through visualizations. As for values above 14 Million dollars, they basically represent outliers as they are unusually expensive properties (such as celebrity homes) which may affect the User's perception of this visualization by distracting the User to an astronomical Sale value.
 - **Tax Class at Present:** Some entries in the Tax Class column were empty. However, it doesn't make sense to remove such entries due to missing Tax class value as it contributes to the Sale Price and Year Built graph trend which makes the Visualization more accurate. Moreover, a Tax Class can be added later to the data and can be easily changed in the Visualization via color-change of the bar and still there will not be any change in the trend/values of the position of data points in the graph.
- **Design Practices:**
 - **Bar Chart:** There are multiple ways to visualize this data, such as a scatter plot or a line plot. However, a line plot isn't a good option as it will not generate an insightful visualization for such high quantity of data. Also, a scatter plot could provide insights to a certain extent; however, deriving trend from such dispersed and high amount of data will not be a good option due to the complexity of data. Accordingly, a bar chart is used to ease the understanding of trends in data and to infer from the data in a glance.
 - **Multiple Charts (one for each Borough):** The main constraint of this assignment is to plot the data using a **Single Image**. Accordingly, in order to simplify the plotting of data and to improve the effectiveness and expressibility of this data visualization, multiple bar charts have been plotted having a **common axis (Year Built)**, one for each Borough, in order to satisfy the Single Image requirement and simultaneously produce an effective visualization which is easy to infer from.
- **Tools:**
 - **Programming Language:** Python
 - **Libraries:** Matplotlib, NumPy, Pandas
- **Improvements:**
 - As evident, specific data in terms of % share is not accurately possible to infer, especially for the Tax Class data. A pie-chart can be introduced which demonstrates the share of each class, if the Single Image constraint is waived.