LAB REPORT - 1
CSE- 564 (VISUALIZATION)
NEW YORK HOUSING DATA BY BUILDING

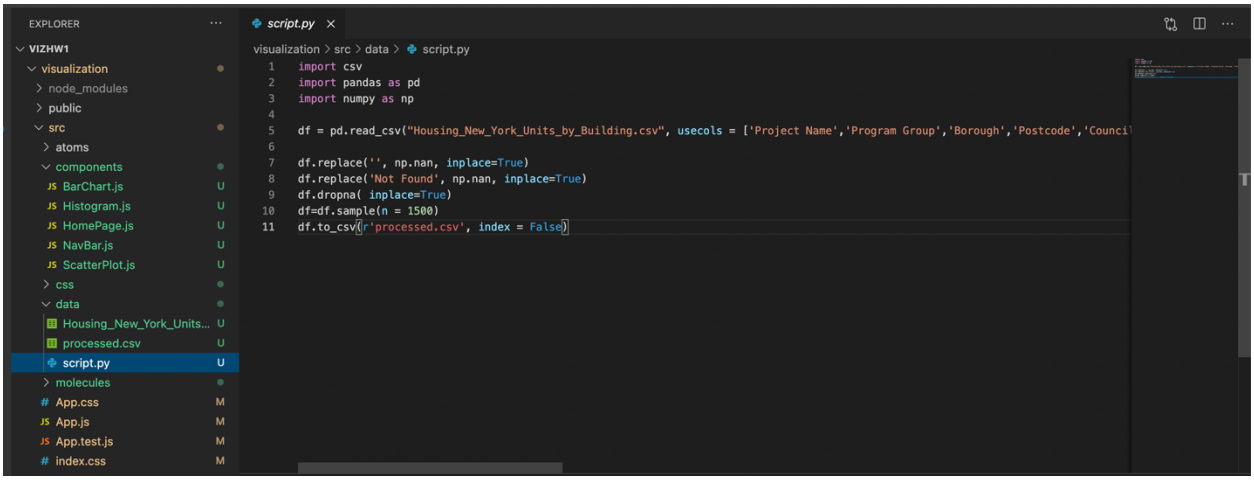
By: Naman Banati

Aim: The project aims to visualize a Data Set by various charts made using ReactJS & d3 library. The charts help in visualizing numeric and categorical attributes using bar-charts, histograms and scatterplots according to specifications in the lab-1 assignment of CSE-564 at Stony Brook University.

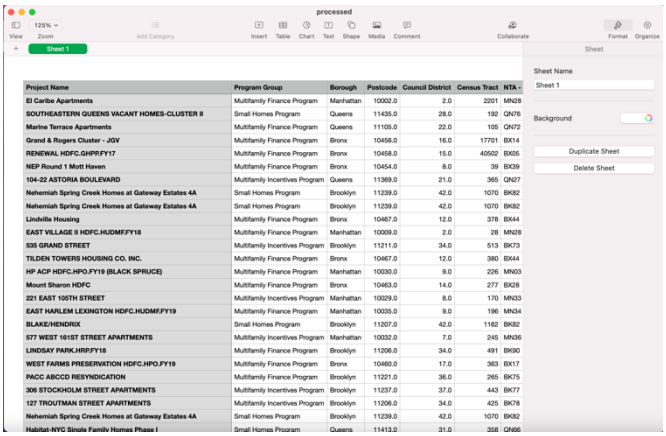
Data Set: The "Housing New York Units by Building" data set was taken from the NYC OpenData web portal and is provided by The Department of Housing Preservation and Development (HPD). HPD provides reports on building units and the features related to them. The Dataset was last updated on November 24th, 2020. The dataset is a good mix of both numerical and categorical variables with 4,732 records of building projects and 42 attributes describing them.

Link to Dataset: https://data.cityofnewyork.us/Housing-Development/Housing-New-York-Units-by-Building/hg8x-zxpr

Processing of Raw Data and Features/Attribute selected: The data was reduced horizontally as well as sampled vertically using a Python script to extract only meaningful data and reduce the number of samples. Samples with incomplete information were cleaned and then 1500 samples with 15(+1 for backtracking to project names if needed in future) features/attributes were taken randomly as suggested by the assignment. The output csv is named as processed.csv and is read by the React Application. The python script is as follows:



Python Script



Project Name	Program Group	Borough	Postcode	Council District	Census Tract	NTA -
El Caribe Apartments	Multifamily Finance Program	Manhattan	10002.0	2.0	2201	MN28
SOUTHEASTERN QUEENS VACANT HOMES-CLUSTER II	Small Homes Program	Queens	11435.0	28.0	192	QN76
Marine Terrace Apartments	Multifamily Finance Program	Queens	11105.0	22.0	105	QN72
Grand & Rogers Cluster - JGV	Multifamily Finance Program	Bronx	10456.0	16.0	17701	BX14
RENEWAL HDFC.GHPP.FY17	Multifamily Finance Program	Bronx	10458.0	15.0	40502	BX05
NEP Round 1 Mott Haven	Multifamily Finance Program	Bronx	10454.0	8.0	39	BX39
104-22 ASTORIA BOULEVARD	Multifamily Incentives Program	Queens	11369.0	21.0	365	QN27
Nehemiah Spring Creek Homes at Gateway Estates 4A	Small Homes Program	Brooklyn	11239.0	42.0	1070	BK82
Nehemiah Spring Creek Homes at Gateway Estates 4A	Small Homes Program	Brooklyn	11239.0	42.0	1070	BK82
Lindville Housing	Multifamily Finance Program	Bronx	10467.0	12.0	378	BX44
EAST VILLAGE II HDFC.HUDMF.FY18	Multifamily Finance Program	Manhattan	10009.0	2.0	28	MN28
535 GRAND STREET	Multifamily Incentives Program	Brooklyn	11211.0	34.0	513	BK73
TILDEN TOWERS HOUSING CO. INC.	Multifamily Finance Program	Bronx	10467.0	12.0	380	BX44
HP ACP HDFC.HPO.FY19 (BLACK SPRUCE)	Multifamily Finance Program	Manhattan	10030.0	9.0	226	MN03
Mount Sharon HDFC	Multifamily Finance Program	Bronx	10463.0	14.0	277	BX28
221 EAST 105TH STREET	Multifamily Incentives Program	Manhattan	10029.0	8.0	170	MN33
EAST HARLEM LEXINGTON HDFC.HUDMF.FY19	Multifamily Finance Program	Manhattan	10035.0	9.0	196	MN34
BLAKE/HENDRIX	Small Homes Program	Brooklyn	11207.0	42.0	1162	BK82
577 WEST 161ST STREET APARTMENTS	Multifamily Incentives Program	Manhattan	10032.0	7.0	245	MN36
LINDSAY PARK.HRFFY18	Multifamily Finance Program	Brooklyn	11206.0	34.0	491	BK90
WEST FARMS PRESERVATION HDFC.HPO.FY19	Multifamily Finance Program	Bronx	10460.0	17.0	363	BX17
PACC ABCCD RESYNDICATION	Multifamily Finance Program	Brooklyn	11221.0	36.0	265	BK75
306 STOCKHOLM STREET APARTMENTS	Multifamily Incentives Program	Brooklyn	11237.0	37.0	443	BK77
127 TROUTMAN STREET APARTMENTS	Multifamily Incentives Program	Brooklyn	11206.0	34.0	425	BK78
Nehemiah Spring Creek Homes at Gateway Estates 4A	Small Homes Program	Brooklyn	11239.0	42.0	1070	BK82
Habitat-NYC Single Family Homes Phase I	Small Homes Program	Queens	11413.0	31.0	358	QN66

Processed Data Set

Feature/Attribute selection: The following attributes were selected from the 42 attributes available in the dataset to provide most useful and meaningful interpretations –

Categorical Attributes:[1]

1. Program Group - The project group/category for a building project.
2. Borough - The Borough where the building is located.
3. PostCode - The Zip code where the building is located.
4. Council District - The NYC Council District where the building is located.
5. Census Tract – The 2010 U.S. Census Tract number where the building is located.
6. NTA - Neighborhood Tabulation Area – The NYC Neighborhood Tabulation Area number where the building is located.
7. Reporting Construction Type – Either a 'new construction' or 'preservation'
    - 'New construction' – New affordable housing units are created, or down payment assistance is given to a homeowner to access an affordable home.
    - 'Preservation' - Existing buildings receive physical rehabilitation and/or financial operating assistance is for existing and future tenants.
8. Prevailing wage status - Whether the project is subject to any prevailing wage requirements, such as Davis Bacon.

Numeric Features: [1]

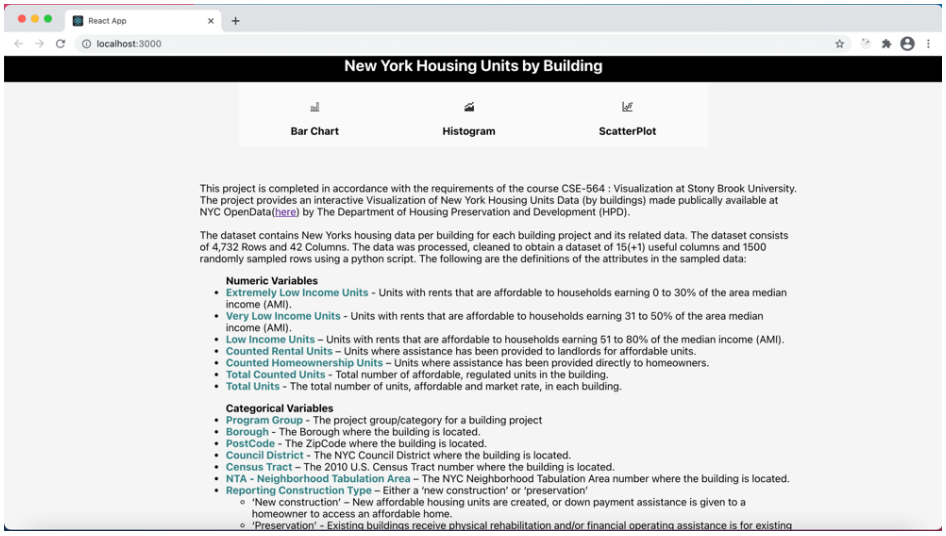
1. Extremely Low Income Units - Units with rents that are affordable to households earning 0 to 30% of the area median income (AMI).
2. Very Low Income Units - Units with rents that are affordable to households earning 31 to 50% of the area median income (AMI).
3. Low Income Units – Units with rents that are affordable to households earning 51 to 80% of the median income (AMI).
4. Counted Rental Units – Units where assistance has been provided to landlords for affordable units.
5. Counted Homeownership Units – Units where assistance has been provided directly to homeowners.
6. Total Counted Units - Total number of affordable, regulated units in the building.
7. Total Units - The total number of units, affordable and market rate, in each building.

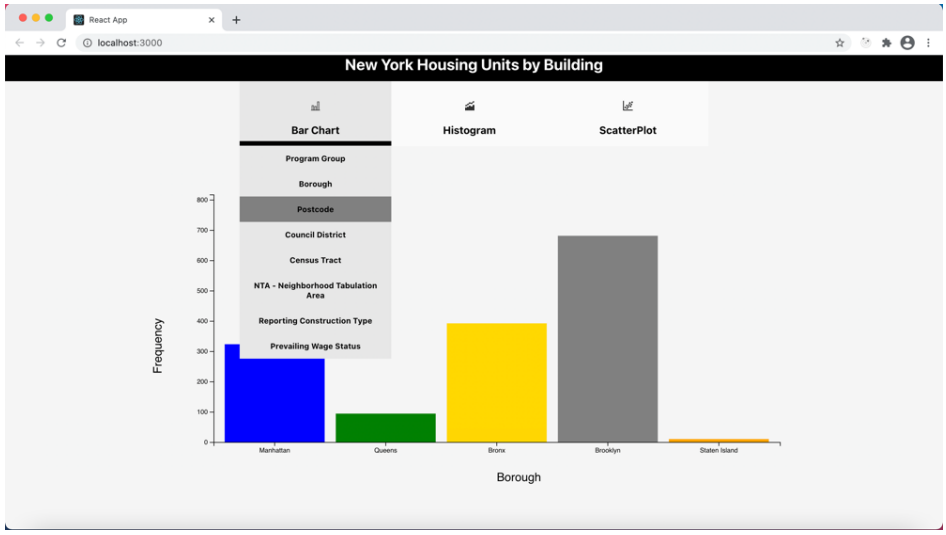
Deploying the React Application: The Application has been made following standard industry practices of modularity and reusability of code. Each element on the application has been divided into component, molecules and atoms depending upon their usage to ensure modularity and code reusability. To run the application, navigate into the "visualization" folder.

- For the production run, install serve using "sudo npm install -g serve" and then type in "serve -s build".
- For the development run, type in "npm i" to install all react specific libraries. Then type in "npm start" to start the server.

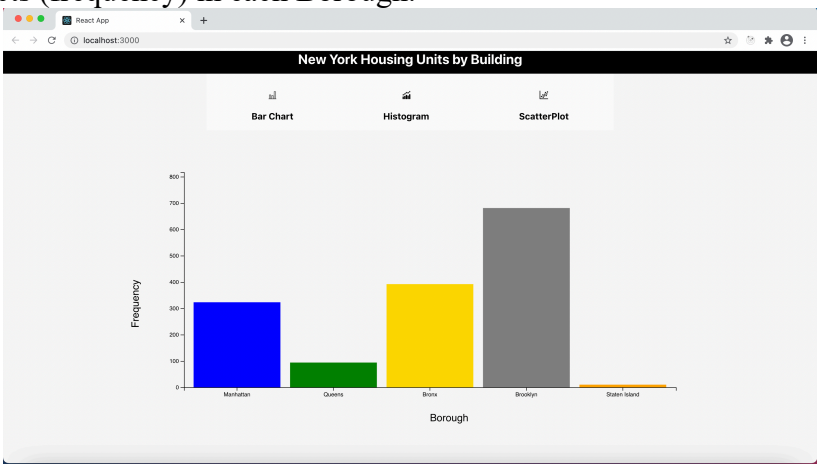
Please have an active internet connection.

Features of the Application: The application has a main menu with the initial welcome screen portraying the meaning of each feature/attribute in the dataset. A user can hover over the options on the main menu to see the various options of attributes available for plotting on each kind of chart. According to the requirements of the assignment, categorical variables are plotted on bar-charts and numerical variables are plotted on histogram and both type of variables are used for the scatterplot.



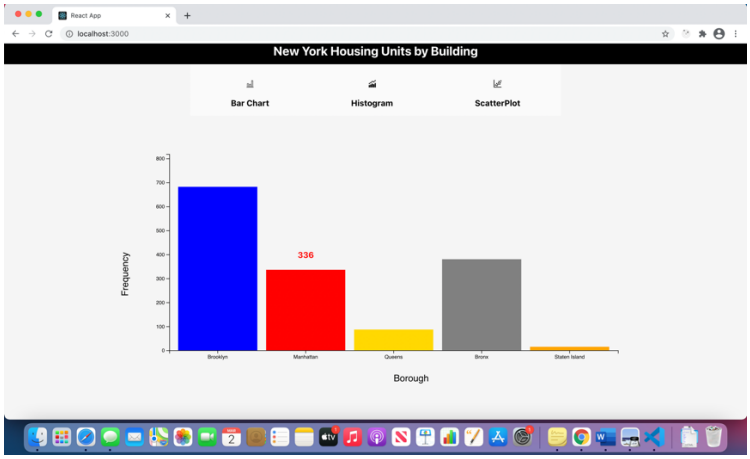


For bar chart and histogram, once a user selects a variable from the options, a chart for that particular variable appear against the frequency. For example, the below image shows the number of housing projects (frequency) in each Borough.



For variables having more than 20 categories, the categories are sorted in decreasing order of frequency and the top 20 categories are shown. The rest of the categories are combined to form a category named “Other” for better visualization.

If a bar is hovered upon, its color changes to red and its value (of frequency) is shown on the top of the bar in red color as per the requirement of the assignment.

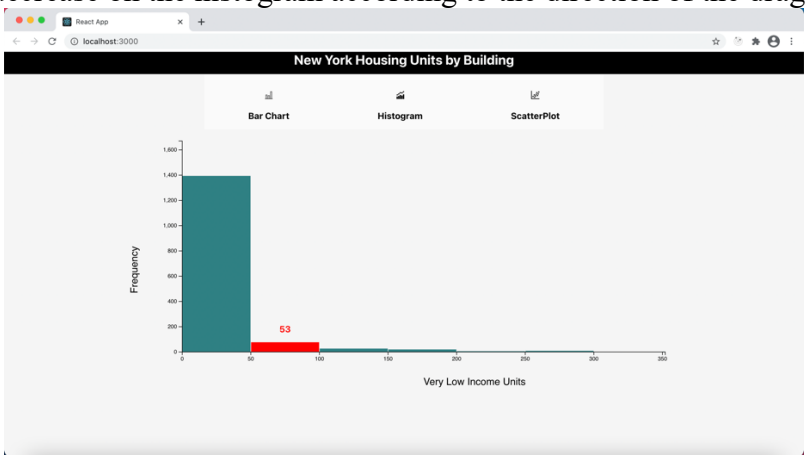


Hovered bar shows frequency (another sample)

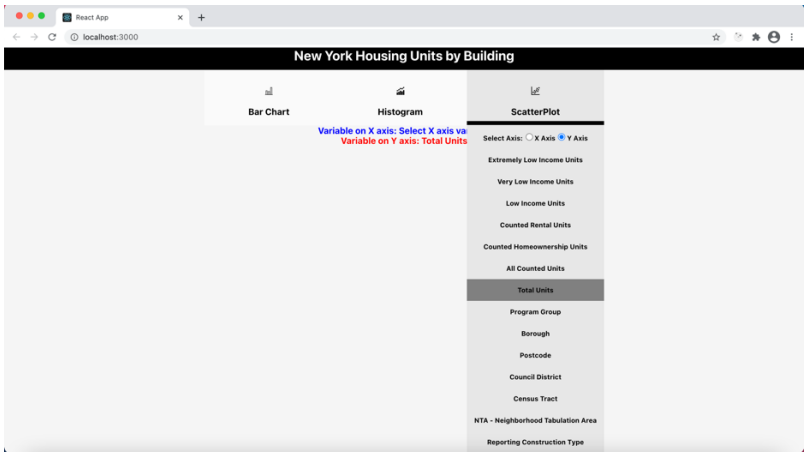
Similarly, a histogram is plotted if a user selects a numeric variable from the histogram dropdown. The histogram has the following feature –

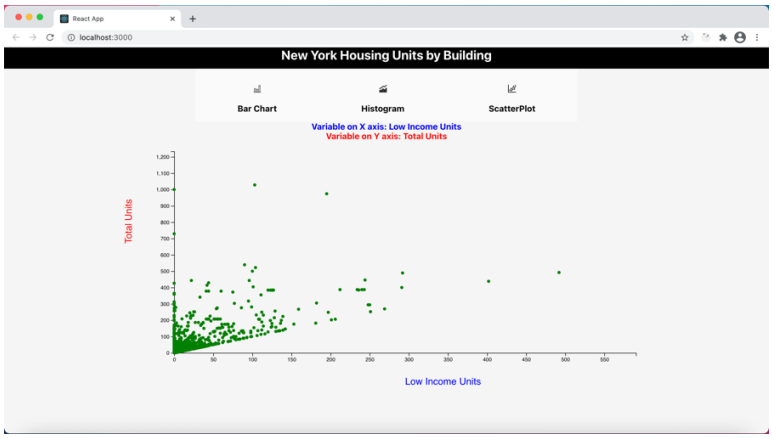
1. If a bar on the histogram is hovered upon, its color changes to red and a red colored text appears over it showing the value (of the frequency) of the bar.
2. The size of the bin also increases slightly to show a magnifying effect when hovered and relaxes when the pointer is moved away.

3. If with a mouse down and drag is performed anywhere on the canvas, the number of bins increase/decrease on the histogram according to the direction of the drag.



To plot a scatterplot between two variables, hover on the scatter plot menu and select a radio button of either of the axis. Then select a variable to be plotted on the axis. The selected variable appears on the screen. Follow the same procedure for the other axis variable. When both variables are selected, a chart is plotted.





When both axes variables are categorical, the points are made to jitter a bit to ensure a better visualization as per the requirement of the assignment.

Importance/Value of Data Set: The data set was selected because it gave a real-world picture of the housing situation in New York. New York (particularly the city) has a housing crisis due to a large population density. Housing in NYC is not very affordable and easy to find, especially for individuals in the low-income bracket. The analysis of this data points to the current situation of housing in NY and also helps future authorities in coming up with new projects targeting the right audience.

Information Sources:

1. Data Set & Rephrased Information - https://data.cityofnewyork.us/Housing-Development/Housing-New-York-Units-by-Building/hg8x-zxpr
2. Images for icons - https://icons8.com/icons/

Link to Project Video:
https://drive.google.com/file/d/12YtFfxua0ayT8G8SCusBmeQJdpOMl-2c/view?usp=sharing