Title: NYC Property Pulse: Sales, Rates & Tweets

Abstract:
This project aims to conduct a comprehensive analysis of New York City's real estate market using property sales data. We will investigate temporal trends in sales volume and pricing, examining how major economic events and seasonal patterns influence market behavior. The research will further categorize properties by type and structural characteristics to determine how these factors correlate with sale prices. By incorporating mortgage rate data from Freddie Mac and consumer price index (CPI) information, the study will contextualize real estate trends within broader economic conditions. This multidimensional approach will yield insights valuable to various stakeholders, including potential homebuyers, real estate investors, urban planners, and policymakers. The findings will contribute to a deeper understanding of the complex dynamics driving NYC's property market and provide evidence-based guidance for decision-making in one of the world's most significant real estate environments.

Our project aims to study the following intended questions:

- Visualize by time and location: How have the number of property sales and average sale prices changed over time? And how does the average sale price vary across boroughs?
- Visualize by property type: What is the price distribution across different property types?
- Visualize by property age: What is the age distribution of sold properties? Or do newer buildings sell for higher prices?
- Visualize by property size: What size of property is more popular? Or how does the price per square foot vary by borough?
- Visualize with mortgage rate: Does mortgage rate affect consumers' preferences for home purchases (e.g., location, size, etc.)?
- Visualize with Consumer Price Index: Does CPI affect consumers' preferences for home purchases (e.g., location, size, etc.)? In this case, we can indirectly understand how major economic events affect real estate sales
- Visualize with homebuyer sentiment: How do homebuyers' comments on social media affect their preferences for home purchases (e.g., location, size, etc.)?

Techniques: ggplot2, ggmap, interaction, NLP

Data Description:

- Data Set: https://www.nyc.gov/site/finance/property/property-rolling-sales-data.page
    - Contains historical data on property sales in New York City.
    - Includes information on the property type (e.g., residential, commercial), sale price, location (borough, neighborhood), and property characteristics (e.g., square footage, number of units).

    - Data is updated regularly, showing transaction records of real estate properties across various periods.
- Data Set: https://fred.stlouisfed.org/series/MORTGAGE30US
    - Provides the average interest rate on 30-year fixed-rate mortgage loans in the U.S.
    - Collected by Freddie Mac and updated regularly (typically weekly).
    - Essential for analyzing trends in housing affordability and the relationship between interest rates and home prices.
    - A widely used economic indicator for the housing market and mortgage loan dynamics.
- Data Set: https://www.bls.gov/cpi/
    - Measures the average change in prices paid by urban consumers for a variety of goods and services over time.
    - Used as a key indicator for inflation and changes in the cost of living.
    - Includes multiple categories of consumer goods and services (e.g., food, housing, medical care, transportation).
    - Essential for economic analysis, policy-making, and understanding the purchasing power of consumers.
- Data Set: Twitter
    - We will try to extract text data from homebuyers on Twitter and process it using NLP for further visualization.

Visualization:

- Map
    - We would use ggmap to illustrate differences in property sale prices across NYC boroughs or neighborhoods.
    - Use varying depths of color (choropleth) to represent different average sale price ranges or total sales volume in each area.
- Line Chart (Time Series Analysis)
    - For each borough (or the city), track the number of property sales or average sale price over time.
    - Illustrate how sales volume or prices change month to month or year over year. We could group by borough or property type to see comparative trends across different segments.
- Bar Chart
    - For different groups of properties (residential vs. commercial, or different property sizes), we would compare sales counts or average prices.
    - Plot each group side by side to reveal any significant differences in sales activity or pricing across these categories.
- Point Chart (Scatter Plot)

- For each property, we could plot square footage on one axis and sale price on the other.
- The point chart allows us to see whether larger properties consistently sell for higher prices. We could add a trend line or compare separate scatter plots for different boroughs.
- An alternative way is using a side-by-side bar chart that could also compare average price per square foot across multiple boroughs or property types.