Spatial & Real Estate Data

Team Reveal Estate
Nora Barry, Laura Buchanan, Jackie Gutman

Project Description & Goals

- What factors affect the value of real estate in NYC?
- Subway lines, quality of neighborhood, park land, proximity of restaurants/bars, etc.
- Quantify their importance
- Make a distinction between commercial and residential real estate

Dataset Descriptions

PLUTO: land use, zoning, and geographic data

- Last updated March 2016
- ~70 fields such as assessed lot value, school district info, residential vs. commercial, number of floors, police/fire districts, X and Y coordinates **tax block and lot number**

Dept. of Finance Sales Data: information about properties sold in the past 13 years (2003- 8/2016)

- Fields such as neighborhood, square footage, building type, latitude and longitude, property sales price and date **tax block and lot number**

Additional Data Sources

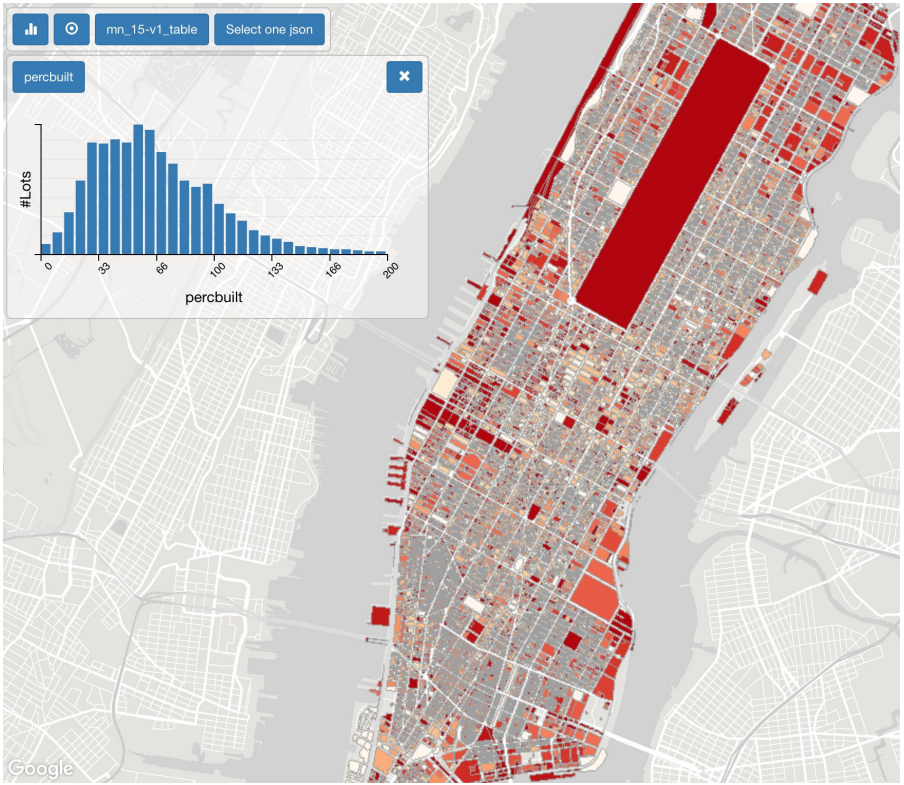
- Will use additional data sources (containing **latitude and longitude**) from NYC Open Data such as:
  - NYPD Crime Data
  - 311 Complaint Data
  - Transportation Data
  - Restaurant/ Liquor License Data

- PLUTO + Dept. of Finance Merged data contains X - Y state coordinates that can be translated to **latitude and longitude**

NYC OpenData

Potential Analysis

- Fit Machine Learning model and extract most important features
- Are these features different for residential/commercial?
- Do limits on building height increase or decrease value?
- Mapping of spatial data (PLUTO) for further analysis



Timeline

- Map PLUTO and Dept. of Finance & Sales data
- Begin feature normalization
- Run simple baseline regression (e.g., historic districts)
- Merge additional urban data
- Investigate differences between commercial and residential data
- Check for abnormal results: what else needs to be cleaned or added?
- Look for hypotheses with visualization tools
- Test new hypotheses

Complete by November 17th ↠ Then focus on paper/poster/presentation