

Visualizing NYC 2017 House Sales Data

Zhengqing (James) Chen

NYCDSA Cohort 12

01/31/2018

Outline

- 1 Project Objectives
- 2 Data (Source, Pre-Processing, Transformation)
- 3 Shiny App Demonstration
- 4 Additional Observations

Project Objectives

- To visualize NYC house sales pattern, *borough-wise*
In particular:
 - Transaction volumes (number of houses sold)
 - House prices
- To find additional insights from the above visualization
- To practice using R, Shiny, Leaflet

Part II. Data

Data Source and Pre-Processing

- Source

- NYC Dept. of Finance, 2017 properties rolling sales for all tax classes
- Raw data set: five files (M, BX, B, Q and SI)
`http://www1.nyc.gov/site/finance/taxes/
property-rolling-sales-data.page`

- Pre-Processing

- All the usual data cleaning
- For simplicity, all sales $< \$100,000$ filtered out
- Grid granularity: Zipcode-wise ((*Lat.*, *Long.*) added to dataset)

Data Transformation

Boro-wise Normalization Normalize a quantity taken in zip code $\in \text{Boro}$ with respect to μ_{Boro} and σ_{Boro} . Essentially z-scores.

(Two examples next)

Transformation: Boro-wise Normalization (1/2)

Example 1 Normalized volume ZV in Brooklyn, January
Say Zipcode = 11234 \in Brooklyn

$$ZV(\text{Zip} == 11234) = \frac{V(\text{Zip} == 11234) - \bar{V}_{\text{Bklyn., Jan.}}}{\sigma_{\text{Bklyn., Jan.}}}$$

Note

- $\bar{V}_{\text{Bklyn., Jan.}}, \sigma_{\text{Bklyn., Jan.}}$ from

$$\{V_{\text{Jan.}}(\text{Zip}_i) : \text{Zip}_i \in \text{Brooklyn}\}$$

- $ZV(\text{Zip}_i) == 0$: $V(\text{Zip}_i)$ at average level in Brooklyn, January
 - ⊕ $ZV(\text{Zip}_i)$: $V(\text{Zip}_i)$ above average
 - ⊖ $ZV(\text{Zip}_i)$: $V(\text{Zip}_i)$ below average

Transformation: Boro-wise Normalization (2/2)

Example 2 Normalized price ZP in Manhattan, April
Say Zipcode = 10023 \in Manhattan

$$ZP(\text{Zip} == 10023) = \frac{P(\text{Zip} == 10023) - \bar{P}_{\text{Manh., Apr.}}}{\sigma_{\text{Manh., Apr.}}}$$

Note

- $\bar{P}_{\text{Manh., Apr.}}, \sigma_{\text{Manh., Apr.}}$ from

$$\{P_{\text{Apr.}}(\text{Zip}_i) : \text{Zip}_i \in \text{Manhattan}\}$$

- $ZP(\text{Zip}_i) == 0$: $P(\text{Zip}_i)$ at average level in Manhattan, April
 - ⊕ $ZP(\text{Zip}_i)$: $P(\text{Zip}_i)$ above average
 - ⊖ $ZP(\text{Zip}_i)$: $P(\text{Zip}_i)$ below average

Boro-wise Normalization: Why? (1/2)

Philosophy: Everything is relative

- 1 Normalized volume ZV : sharp anomaly pattern *within* a borough

Consider: $(\dots, -1, 0, 1, \dots)$ sharper contrast than $(\dots, \mu - \sigma, \mu, \mu + \sigma, \dots)$

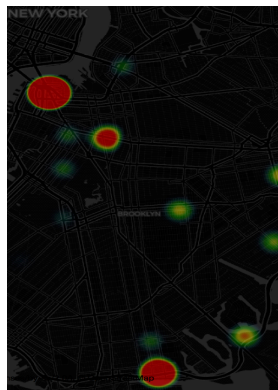
(Example: next slide)

- 2 Normalized price ZP : fair price comparisons *across* boroughs
 - House price \$1M implies very differently in Manhattan than Staten Island
 - So does a \$100K fluctuation in Manhattan than Staten Island

Data Transformation: Why? (2/2)

Heat maps: House sales volumes, Brooklyn, May
(Heat map: intensity \propto volume)

Left: Volume; Right: Normalized Volume



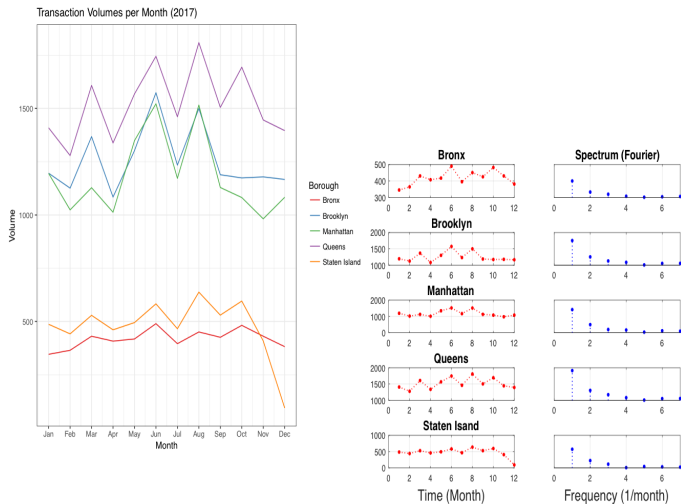
Part III. Shiny Demonstration

`https://jamesczq.shinyapps.io/NYC_House_Sales_2017/`

Part IV. Additional Observations

Additional Observations (1/3)

Time-behavior: Periodicity in transactions



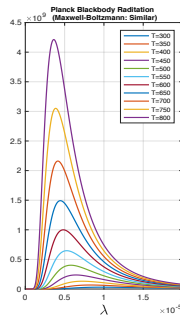
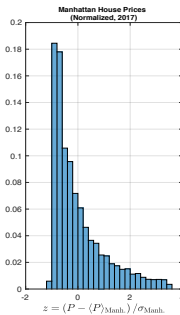
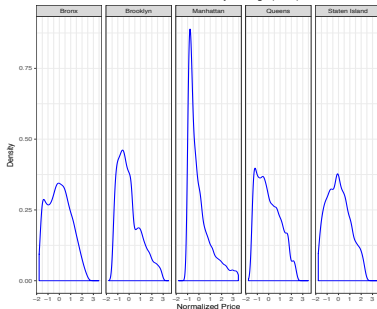
Additional Observations (2/3)

Thermodynamics of NYC real estate business

Postulate: Systems in local thermodynamical equilibria (LTE)

NYC Real Estate Business	Thermodynamics
Location	Temperature

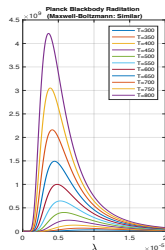
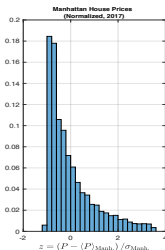
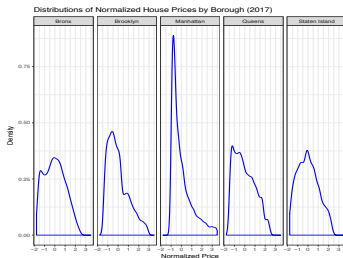
Distributions of Normalized House Prices by Borough (2017)



Additional Observations (3/3)

NYC real estate thermodynamics (systems in LTE)

NYC Real Estate Business	Thermodynamics
Location	Temperature
(Volume, Price)	Phase Space (x, \dot{x})
Multiplicity in (Volume, Price)	Entropy



Thank You!

- Possible improvements:
 - 1 Refine plots and Shiny appearance
 - 2 Incorporate data from past years (2016 and prior)
 - 3 Refine study with different house categories
 - 4 ...
- Questions?
- Thanks for the help with Shiny from Aaron, Drace, Kathryn, and Zeyu
- Thank you!