

# **City of Boston 311 Service Request Data Analysis**

---

by the BU MSSP 2015-16 Cohort



**Big Picture**

# Context

**MS in Statistical Practice (MSSP) Program:** *A new BU masters program, with focus on producing holistically trained statisticians to work in an integrated fashion in the modern data-science environment.*

At the heart of the program is the Statistics Practicum -- where “it all comes together”.

The Bos 311 data were used this semester in the Practicum as the data set upon which to focus skills development, practice of principles, and team building.

End product was a comprehensive analysis and corresponding analytics platform, focused on the topic of *Quality of Service*.

# Overview of Presentation

1. Preliminaries: Data Cleaning and Manipulation
2. Exploratory Data Analysis of QoS Metrics
3. Revisiting Efficiency
  - a. The Role of Target Times
  - b. Impact of Demographics
4. Concluding Remarks

# **Data Cleaning and Manipulation**

# Data Cleaning

The 311 Service Request dataset from the City of Boston was downloaded and imported into R. The following steps were then taken:

## 1. Filters / exclusions / removals:

- Filter data to select only requests from 2014.
- Exclude requests
  - generated by city employees (Source field: City Worker App and Employee Generated); or
  - where open date is later than the closed date
- Remove
  - incomplete records;
  - requests with Closure Reasons containing “invalid” or “duplicate”;
  - requests that were closed within 15 minutes of opening.

# Data Manipulation

2. Four key quality-related metrics were defined:

- **Request Volume** [count] = total number of requests
- **Duration** [hours] = (Closed Time) - (Open Time)
- **On-Time Rate** [ratio] = (Number of On -Time Requests) / (Total Number of Requests)
- **Efficiency** [ratio] = (Target Time - Open Time) / (Closed Time - Open Time)

*Note: Metric summaries typically used the median because of its resistance to outliers & heavy tails.*

3. Aggregation: Data were, for some analyses, regrouped by department and neighborhood.

# Additional Data Sources: Demographics

## Data from the 2010 Census (Census Tract Level):

- Total Population
  - Total White Population, Total Black Population, Total Asian Population, Total Hispanic Population, Total Male Population, Total Female Population, Total Population over 18
  - All of these were turned into percentages of the total population
- Area (Square Miles)
  - Dividing total population by area led to population density
- Median Age

## Data from 2013 American Community Survey (Also Census Tract Level)

- Median Income



# **Exploratory Data Analysis**

# Exploration Using Shiny

We conducted a variety of exploratory data analyses of the Bos311 data.

In order to facilitate *exploration*, *communication* and *sharing* of our results, we incorporated final versions of select analyses into an *interactive Shiny Application in R*.

Focus on 3 Visualizations

1. Scatter Plots
2. Bubble Charts
3. Maps

[Link to Shiny Application](#)

# Revisiting Efficiency

# Closer Look at Efficiency

*Efficiency* was defined objectively using open time, target time, and closed time, but how you define efficiency matters!

To investigate *efficiency* further, we looked at 2 things:

## 1. Impact of Target Time

- $(\text{Closed Time} - \text{Open Time}) / r * (\text{Target Time} - \text{Open Time})$
- Finding: Differential impact of varying relative target times

## 2. Impact of Demographics

- Regression analysis allows for notion of efficiency, adjusted for demographics.
- Finding: Modest effects.

[Link back to Shiny Application](#)

# Adjusting for Demographics?

It might be argued that departments have their efficiency affected differentially due to interaction of

1. their particular services (e.g., parks versus schools), and
2. demographics (e.g., neighborhood age, income, density)

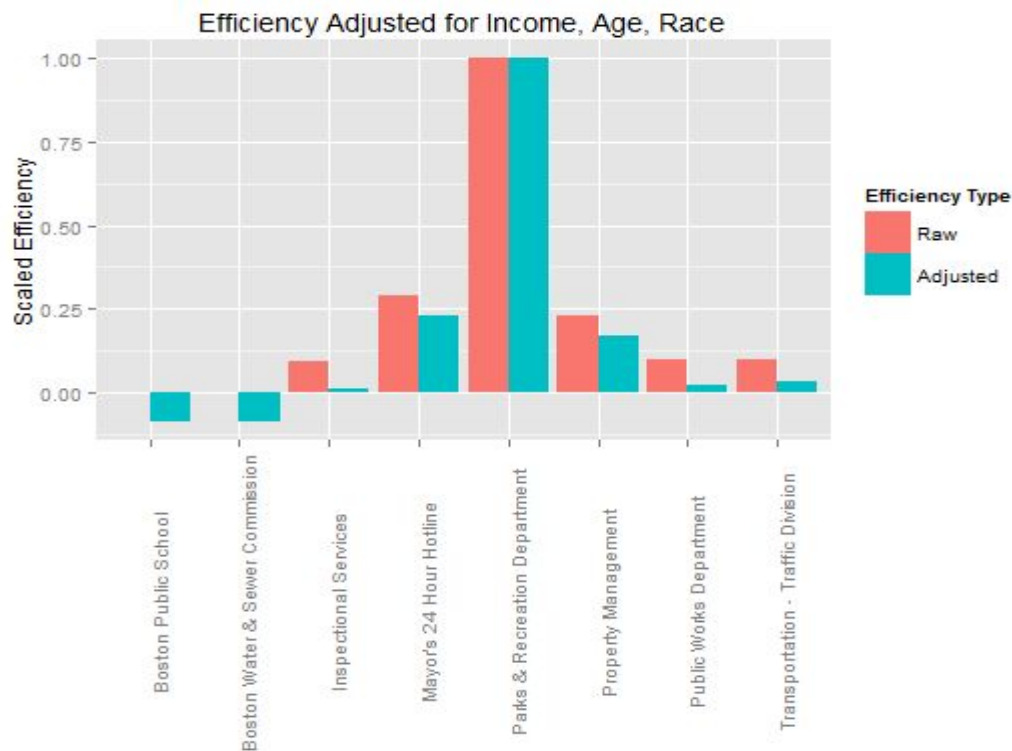
To assess this possibility, we used regression analysis to adjust for demographics, comparing

and

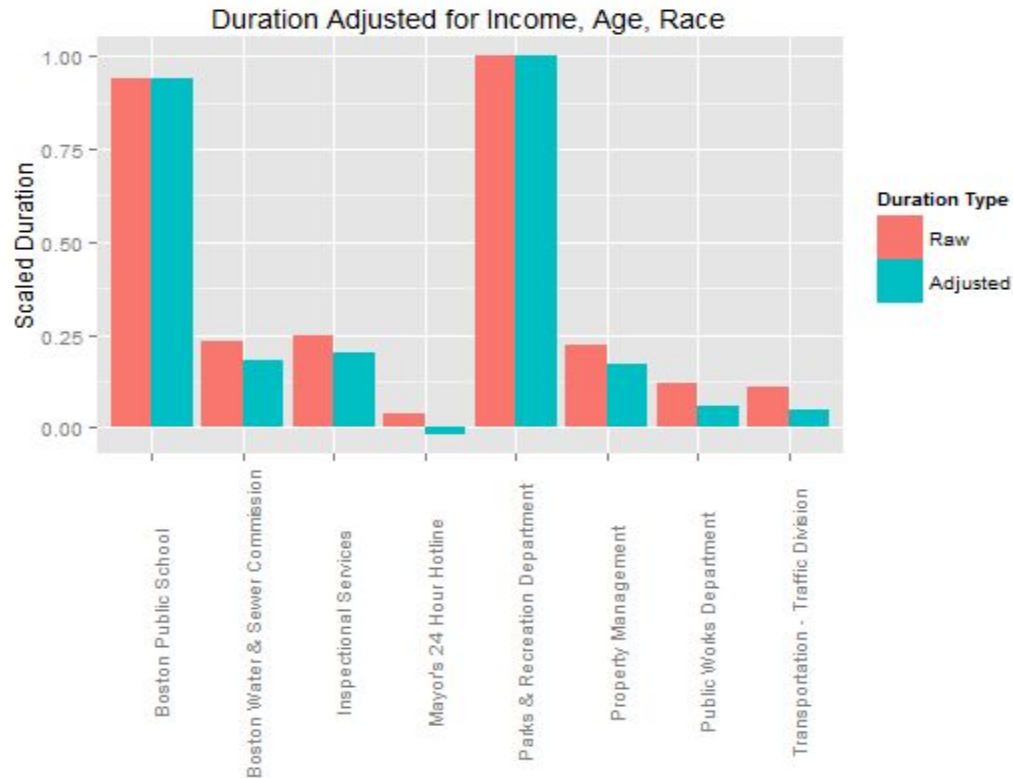
QoS Metric ~ Department

QoS Metric ~ Department + Demographics

# Efficiency and Demographics



# Duration and Demographics



**Take-Home Points**



# Takeaways To Think About

1. Visualizations and interactive Shiny environment allow for rich and simultaneous analysis of multiple, interacting quality metrics and variables.
2. Definition of high-level concepts from low-level data needs care.
  - E.g., “Efficiency” may reflect differences in budget/manpower, difficulty of requests, over-estimation of target time, etc.
3. Confounding of efficiency with demographics appears to be modest.
4. Data on “Effectiveness” of work done would be a useful complement

# Contributors

- Students
  - Qian Fang, Frank Giron, Brian Hart, Jing Liu, Zhuojin Lyu, Jeffrey Paadre, Hankui Peng, Rhea Roy, Jiayuan Shi, Yulan Sun, Su Yang, Jiapeng Yuan, Hanxi Zhang, Xiaoyi Zhang, Lina Zhou, Yitong Zhou, Jingshu Wu
- Teaching Fellows
  - Aleksandrina Goeva and Clementine Mottet
- Professors
  - Dr. Eric Kolaczyk and Dr. Haviland Wright
- Program Coordinator
  - Marisa DiSarno