

# Exploratory Data Analysis of the City of Seattle's Participatory Budgeting Program

**Katrina Ernst, MPA Candidate, Evans School of Public Policy & Governance**  
**Prepared for the course: Computational Thinking for Government Analytics**  
**March 2018**

## Introduction

I chose to explore the data for the City of Seattle's 2017 participatory budgeting program, called "Your Voice, Your Choice". This was not data that I was familiar with before this project. The code that I used to explore the data is available in the R markdown MASTER.

## About Participatory Budgeting

Participatory budgeting is used around the world to engage residents in civic decision-making. Seattle first started its participatory budgeting program with a youth-focused initiative in 2015. The program was expanded to include an initiative open to all Seattle residents in 2017.

The 2017 participatory budgeting program - Your Voice, Your Choice - began by soliciting project proposals for neighborhood improvements. The program received almost 900 proposals. Those proposals were evaluated by teams, so that each City Council District, of which there are seven, voted on its top ten proposals. Voting commenced and the projects receiving the highest votes up to the maximum allowed \$285,000 budget per Council District were funded. The program data was made available on the City's Open Data platform: [data.seattle.gov](https://data.seattle.gov).

## Why Look at the Data?

I thought this program would be interesting to explore because participatory budgeting is meant to be a way to encourage civic engagement. The program was structured to equally distribute funding, so exploring whether the variable Council District correlated to the variable Project Status, which would tell us if it was funded or not, is not interesting. However, I thought there would still be some interesting things to discover through exploratory data and text analysis. I chose to do some text analysis because the data provided includes several fields with long, unique character strings, rather than levels. I also attempted a cluster analysis and mapped the data by Council Districts.

## Findings

### Project Type by Council District

I thought that one of the interesting things this data might reveal is something about specific neighborhood improvement needs. There is a variable in the data named "Project Type", so I looked at Project Type by Council District and ran a Pearson's chi-squared test to see if the differences between the means was significant.  $p = 0.01166714$  so the null hypothesis is not rejected and it appears that Project Type is not associated with Council District. However, that "Chi-squared approximation may be incorrect" was noted in the output.

## Project Status by Average Score

I thought it would also be interesting to explore if Average Score made a difference for whether or not a project was funded. Since a committee of residents scored each project, but every resident was invited to vote, did voters perceive a difference in the higher versus lower scored top ten projects and vote to be funded the highest among the top ten? I looked at a box plot to see if there was a notable difference.

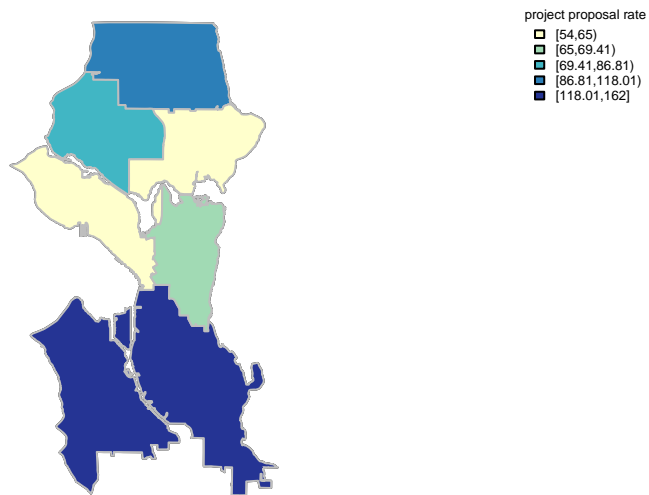


The box plot reveals some interesting things. Namely, there was a project that scored low enough to be an outlier, but was still funded for implementation. It would be interesting to know what about that project appealed to voters.

## Mapping Number of Project Proposals by Council District

I also mapped the quantity of project proposals received by Council District. The map may tell us a couple of different things. It may be telling us that outreach to certain council districts was higher than others and thus resulted in more proposals. Or, it could be that the northern and southern districts are more populous. Or, it could be telling us that there really is more need in those Council Districts. Further research is needed to determine which of these is true. If the latter is true than this has implications for policy and decision-making at the City.

## Project Proposal Rate by Council District



## Questions for Future Research

It would be interesting to do further text analysis with this data. I was able to generate the top ten words used in the “Who Benefits?” field. Since project proposers fill in this field, there is variability. I wonder if there is any correlation between the terms used in this field and the score that the committee gives the project, or if it correlates to the final projects funded for implementation. This data seems rich for exploring perceived needs and benefits for City projects.