# ANALYZING AND VISUALIZING CAMPAIGN FINANCE DATA

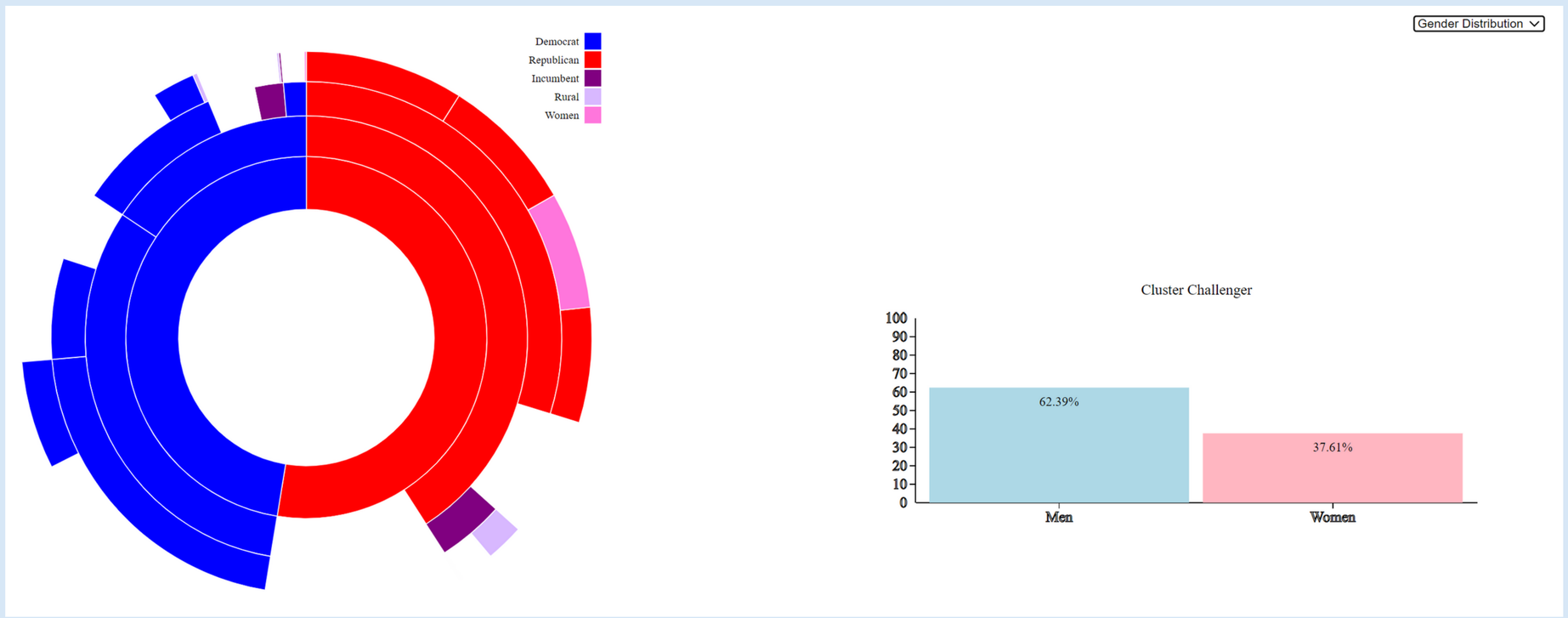
Team 014: Taylor Graham, Niv Balamurali, Sharon Arulpragasam, Beatrice Lee, Federico Reinoso, Kaitlyn Baek

## INTRODUCTION AND MOTIVATION

- We want to take raw data detailing individuals campaign finance contributions and create visualizations of this data to make this information accessible to the public
- Our objective is to categorize donors and recipients into distinct groups based on their characteristics to identify underlying patterns and trends in political donations
- Beyond mere classification, we aim to uncover and analyze trends within these groups

## DATA

- We sourced our raw data in the form of a csv file from OpenSecrets, a non-profit organization that tracks and publishes data on campaign finance and lobbying, and the original file contains approximately 2 million rows, with each row representing one donation
- We filtered the data by eliminating candidates that received only one donation and donors that only donated once during the election cycle, condensing multiple donations from a donor to the same candidate into a total donation amount, and removing total donations less than $2000
- In order to begin our algorithm we needed to transform our data into an undirected graph where nodes represent donors or recipients and edges represent a total donation amount between a donor and recipient pair




# APPROACH

## ALGORITHM

**Spectral Characterization**

- We utilized spectral characterization to create an eigen-spectrum that we analyzed for signs of hierarchical structure in our data
- The eigen-spectrum revealed signs of a hierarchical structure involving hierarchy levels of 2, 4, 6, and 8 clusters

**Spectral Clustering using Fuzzy C-Means**

- Using the results from spectral analysis we processed our data further using a spectral clustering technique to identify cluster shapes
- The clustering algorithm we used was fuzzy c-means which outputs membership percentages for each cluster for every data point
- Fuzzy clustering allows us to see overlap in clusters and avoid an oversimplification of our data

**Decision Tree**

- In order to find characteristics that defined our clusters we tried to use decision tree classification to identify defining characteristics between our clusters
- The decision tree produced results that resulted in overfitting and we could not achieve a test accuracy over approximately 65%

## VISUALIZATION

- Utilized a zoomable sunburst visualization that allowed us to create fuzzy clusters denoted by color
- The zoomable feature allowed levels to be split into hierarchies
- A secondary visualization that appears on hovering over a cluster details important features of each cluster such as: gender ratios, employment status, and political party

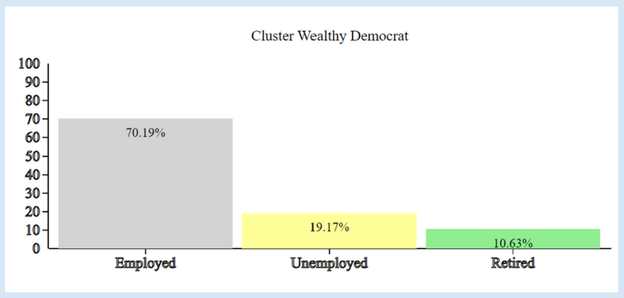


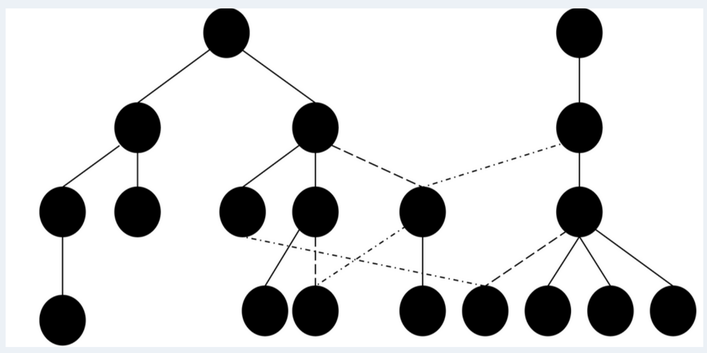

Fig 1: An example our secondary visualization that represents employment information about the cluster Wealthy Democrat

## EXPERIMENTS

- Our goal for this experiment was to uncover underlying trends and patterns in campaign finance data and make these results easily accessible to voters.
- Utilizing hierarchical clustering we aimed to categorize donors and recipients based on their characteristics, hoping to reveal insights into why people donate to political campaigns and the effects of these contributions on political outcomes.
- Our experiment was designed to answer questions such as:
  - “Are there any underlying structures in the campaign financial data in Georgia?”
  - “Are there any interesting relationships between sub communities and hierarchies in campaign financial data?”
  - “Is there any relation between this structure and the success of candidates?”
  - “What features within our data are the most relevant indicators of donation behavior?”

## RESULTS

- We were able to determine a hierarchical structure utilizing fuzzy c-means clustering and Jaccard similarity scoring



- The above structure was identified by our algorithm. Each circle represents a cluster center of the data. Solid lines represent links while dashed lines represent weak links between clusters with the darker the weight, the stronger the connection
- However, we were unable to conclude the significance of each cluster and what the determining characteristic of each cluster was