# Rating Charities w/ Data Science

Presented by: Roadhouse (Randy, Lawrence, Vimal, Andrew, Richie)

February 9th, 2019

# What is Charity Navigator


- ***Charity Navigator*** is the “largest and most-utilized charity evaluator in America.” Charity Navigator  helps guide intelligent giving by evaluating the Financial Health, Accountability and Transparency of over 9,000 charities.
- Charity Navigator accepts no advertising or donations from the organizations it evaluates.
- Charity Navigator is a non-profit 501 (c) (3) public charity itself, which depends on support from individuals, corporations and foundations that believe it provides a much-needed service to America's charitable givers.

# A look at the data available

Sign up or view our technical documentation here

### Rating Tier

- Search 1.6 million nonprofits by keyword(s) and other criteria
- Verify a charity's IRS standing and deductibility status
- Browse by cause areas*
- Display overall star ratings*
- Display mission statement*

**Pricing and Usage Limits**

- Free
- 1,000 hits/day

**Data Fields Included**

- Employee Identification Number (EIN)
- Charity Name
- Mission*
- Tag Line*
- IRS Classification
- IRS Subsection
- Foundation Status
- Deductibility
- Charity Navigator URL*
- Category*
- Cause*
- Street Address
- City
- State
- Zip
- Active Advisories
- Removed Advisories
- Current Rating (star image)*

### Content Tier

- **All the features included in the Rating Tier, plus**
- Browse premium content (Hot Topics, Top Ten Lists)
- Display detailed ratings for individual charities*
- Display accountability metrics for individual charities*
- Access historical ratings for individual charities*

**Pricing and Usage Limits**

For Content Tier pricing and usage limits, please email us at api@charitynavigator.org.

**Data Fields Included**

- **All the fields included in the Rating Tier, plus**
- Numeric ratings and scores*
- Form 990 data
- Financial performance metrics*
- Accountability tests*
- Representatives (CEO, chairman, etc.)*
- Lists (Top Ten Lists, Hot Topics)*

**Applies to rated organizations only.*

CHARITY NAVIGATOR

**How Is the Overall Score and Star Rating Calculated?**

In our two-dimensional rating system, the overall score is not a sum but rather a measurement of the distance of two component scores from a perfect score of 100 and 100. The smaller the distance to the perfect score, the better the overall score:

$$100 - \frac{\sqrt{(100 - \text{Financial})^2 + (100 - \text{A\&T})^2}}{2}$$

Although the calculation for Overall Score differs from the calculation for Financial Health and Accountability & Transparency, the Overall Rating has the same rating table.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Overall Rating:** | ★★★★ | ★★★☆ | ★★☆☆ | ★☆☆☆ | **0 Stars** | **Donor Advisory** |
| **Overall Score:** | ≥ 90 | 80 - 90 | 70 - 80 | 55 - 70 | < 55 | N/A |

# Questions we asked of this data

- Do charities with higher contributions have higher ratings?
- Do charities in certain categories have higher/lower ratings? (Arts, Humanitarian causes, Medical Research, etc.)
- Are some states more charitable than others?

# Hypothesis

Charitable Organizations that receive higher contributions have higher Ratings when compared to smaller sized charities.

# Pulling from the API itself

In [47]:
```
def createOrganizationSearchURL(size_search_params=default_search_params):
    query_params = [f"{param}={value}" for (param, value) in size_search_params.items()]
    query_string = "&".join(query_params)
    url = f"{base_organization_url}?{query_string}"
    return url
```

```
search_params = default_search_params

for page_num in range(1, 20):
    print(page_num)
    search_params["pageNum"] = page_num
    print(search_params)
    organizations_url = createOrganizationSearchURL(search_params)
    response = requests.get(organizations_url)
    if (response.status_code != 200):
        break;
    else:
        result_df = pd.DataFrame(response.json())
        results.append(result_df)

total_results = pd.concat(results)
total_results.to_csv("All_Organizations.csv")
```

# Issues With the Raw Data

- Irrelevant Data
- Raw JSON was stored so relevant data were embedded in dict strings
- Data was repeated
- Other run-of-the-mill issues such as renaming and reindexing the columns.

# Cleaning our data

| | Unnamed: 0 | advisories | category | cause | charityName | charityNavigatorURL | currentRating | donationAddress | ein | irsClassifica |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | {'severity': None, 'active': {'_rapid_links': ... | {'categoryName': 'Research and Public Policy',... | {'causeID': 35, 'causeName': 'Non-Medical Scie... | Mount Desert Island Biological Laboratory | https://www.charitynavigator.org/?bay=search.s... | {'score': 92.61, 'ratingID': 131800, 'publicat... | {'country': None, 'stateOrProvince': 'ME', 'ci... | 10202467 | {'deductib 'Contribut are deduct |
| 1 | 1 | {'severity': None, 'active': {'_rapid_links': ... | {'categoryName': 'Community Development', 'cat... | {'causeID': 42, 'causeName': 'United Ways', 'c... | United Way of Eastern Maine | https://www.charitynavigator.org/?bay=search.s... | {'score': 81.84, 'ratingID': 138738, 'publicat... | NaN | 10211478 | {'deductib 'Contribut are deduct |
| 2 | 2 | {'severity': None, 'active': {'_rapid_links': ... | {'categoryName': 'Health', 'categoryID': 5, 'c... | {'causeID': 14, 'causeName': 'Medical Research... | The Jackson Laboratory | https://www.charitynavigator.org/?bay=search.s... | {'score': 93.16, 'ratingID': 131608, 'publicat... | {'country': None, 'stateOrProvince': 'ME', 'ci... | 10211513 | {'deductib 'Contribut are deduct |
| 3 | 3 | {'severity': None, 'active': {'_rapid_links': ... | {'categoryName': 'Arts, Culture, Humanities', ... | {'causeID': 6, 'causeName': 'Libraries, Histor... | Maine Historical Society | https://www.charitynavigator.org/?bay=search.s... | {'score': 80.31, 'ratingID': 124121, 'publicat... | NaN | 10211530 | {'deductib 'Contribut are deduct |
| 4 | 4 | {'severity': None, 'active': {'_rapid_links': ... | {'categoryName': 'Human Services', 'categoryID... | {'causeID': 16, 'causeName': 'Youth Developmen... | Boys & Girls Clubs of Southern Maine | https://www.charitynavigator.org/?bay=search.s... | {'score': 91.54, 'ratingID': 135033, 'publicat... | NaN | 10211543 | {'deductib 'Contribut are deduct |

# Cleaning Process

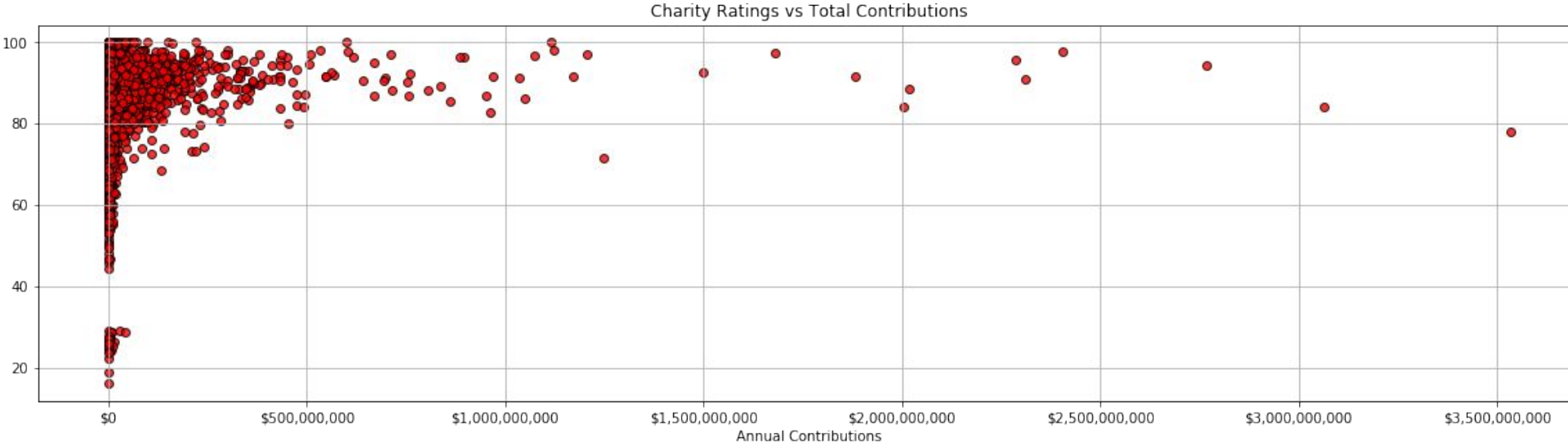
- Trimmed the columns down from 16 to 8, only including the information relevant to our data as well as the charities' missions and causes.
- Created our own functions which utilized the ast.literal_eval function from the abstract syntax tree library to pull out the data from the dictionary strings.
- We determined that the EIN column was the one we would use as a *'Primary Key'* and dropped any repeated values (excluding the first).
- Renamed and reindexed the columns in a way that is much easier to understand.

# Cleaning our data (part two)

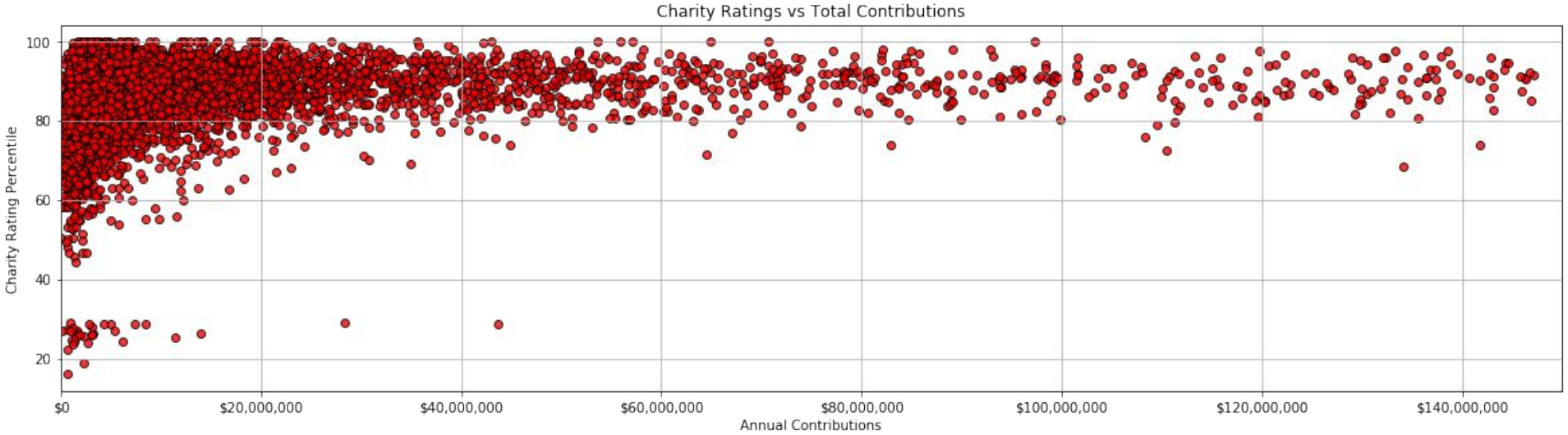
## Cleaned Data

| | Charity | ein | Mission | Cause | Category Name | Rating | State | Contributions |
|---|---|---|---|---|---|---|---|---|
| 0 | Mount Desert Island Biological Laboratory | 10202467 | The MDI Biological Laboratory is a rapidly gro... | {'causeID': 35, 'causeName': 'Non-Medical Scie... | Research and Public Policy | 92.61 | ME | 12924245.0 |
| 1 | United Way of Eastern Maine | 10211478 | Working with the communities we serve, the Uni... | {'causeID': 42, 'causeName': 'United Ways', 'c... | Community Development | 81.84 | ME | 2438155.0 |
| 2 | The Jackson Laboratory | 10211513 | The Jackson Laboratory, founded in 1929, is an... | {'causeID': 14, 'causeName': 'Medical Research... | Health | 93.16 | ME | 331241336.0 |
| 3 | Maine Historical Society | 10211530 | Founded in 1822, the Maine Historical Society ... | {'causeID': 6, 'causeName': 'Libraries, Histor... | Arts, Culture, Humanities | 80.31 | ME | 2219692.0 |
| 4 | Boys & Girls Clubs of Southern Maine | 10211543 | For more than 100 years, Boys & Girls Clubs of... | {'causeID': 16, 'causeName': 'Youth Developmen... | Human Services | 91.54 | ME | 3859985.0 |

# Initial Scatter Plot



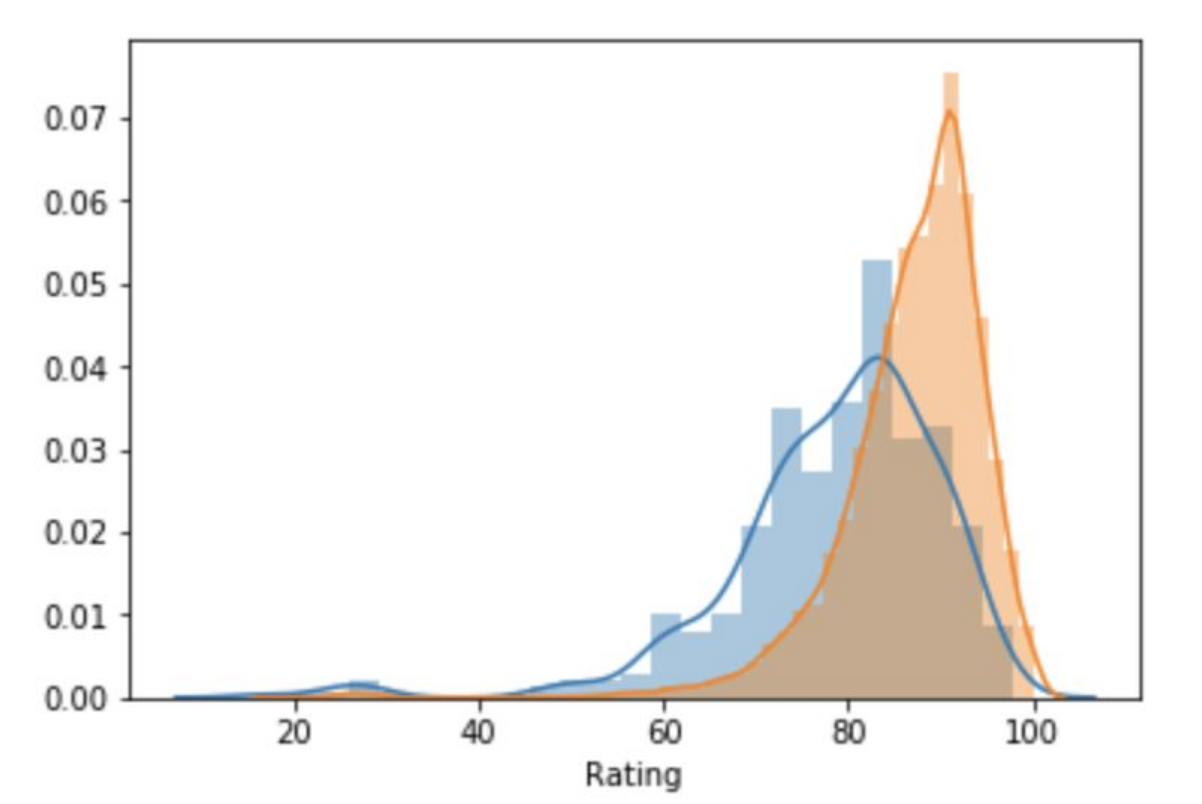
# Reformatted Scatter Plot



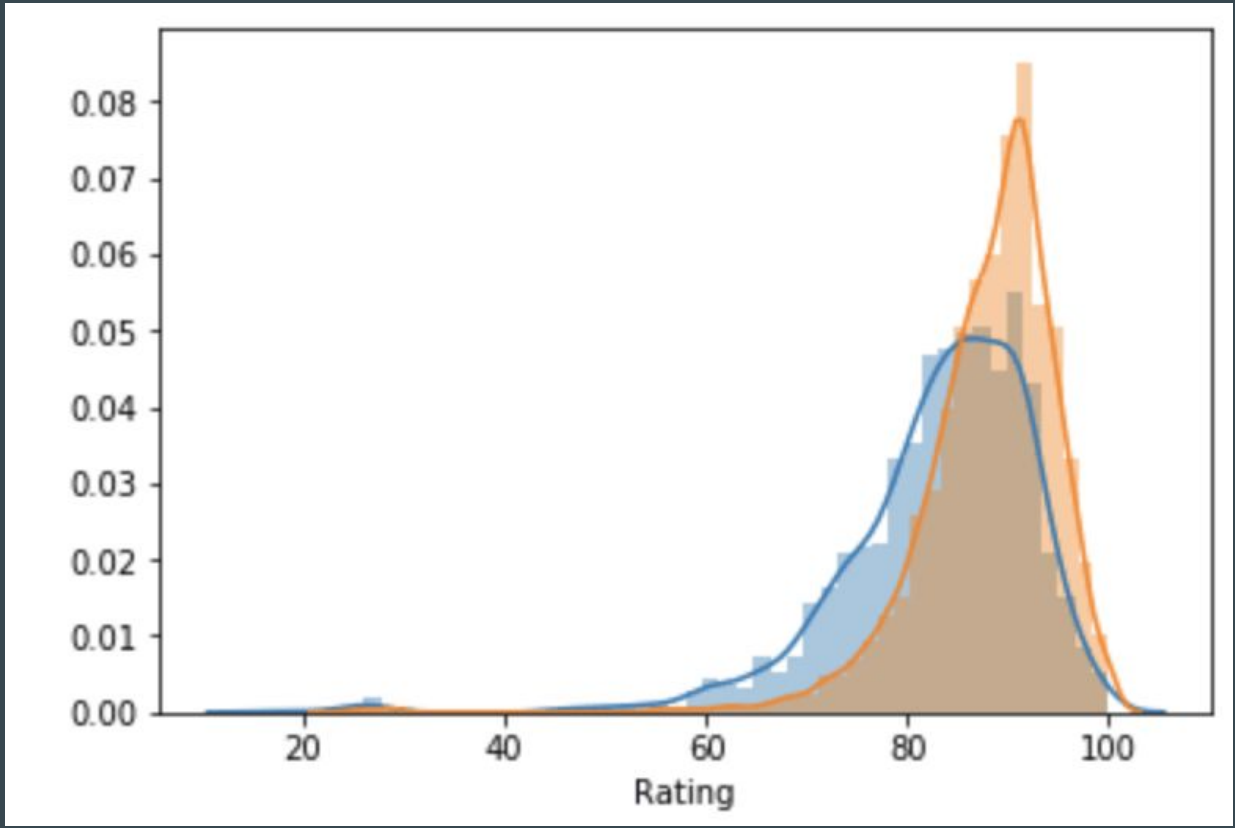
# Results

- Upon visualizing the contributions against rating, we saw that there was no positive linear correlation between them. The distribution of data points did not follow a line.
- There was however a significant difference (pvalue=4.84e-96 comparing below a million) between lower percentiles and upper percentiles of contributions when performing anovas and ttests on their ratings.
- The best cutoff point (Done with t-tests comparing edges at percentiles in a range of 0 to 95 by 5) was the 30th percentile at $2.6 million with a p-value of 3.32e-149.
    - Average Rating below $2.6 Million was 83.1 compared to 88.1 above $2.6 Million

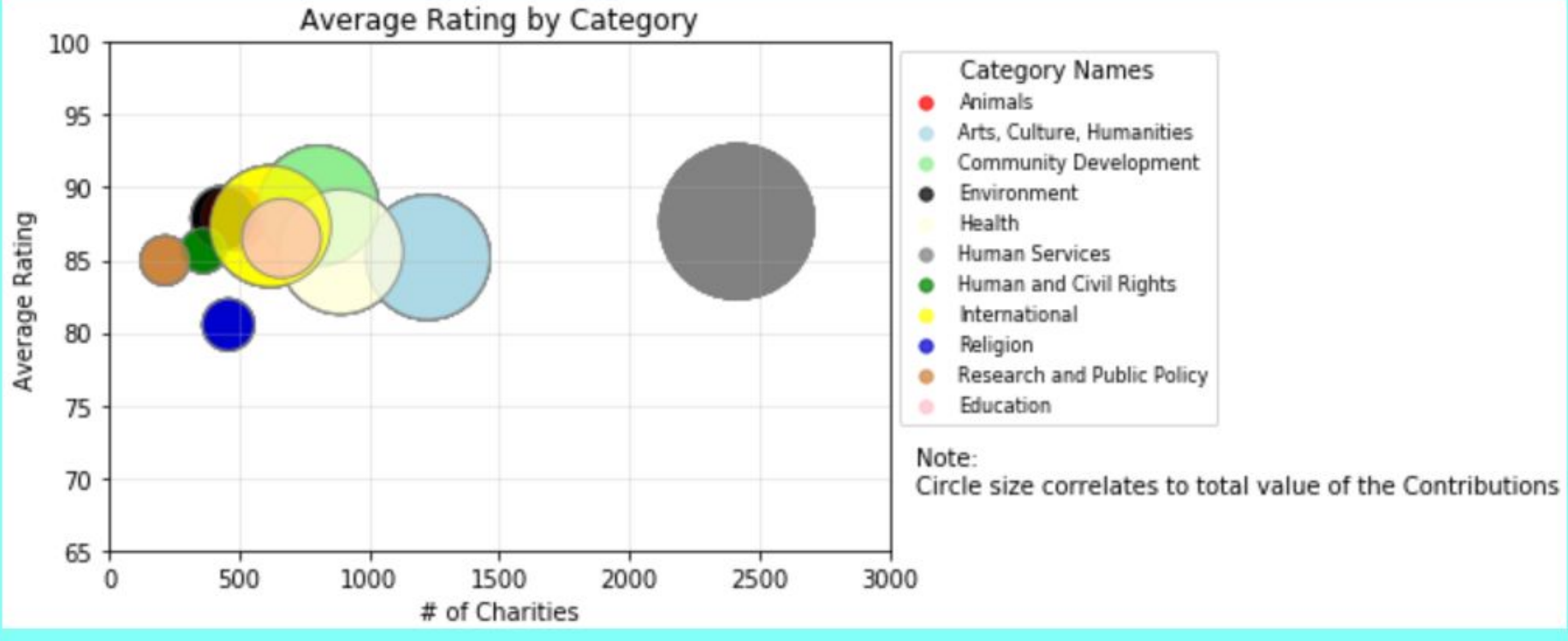
# Ratings below $1,000,000 vs Ratings above



# Distributions of Rating above and below $2.6 million

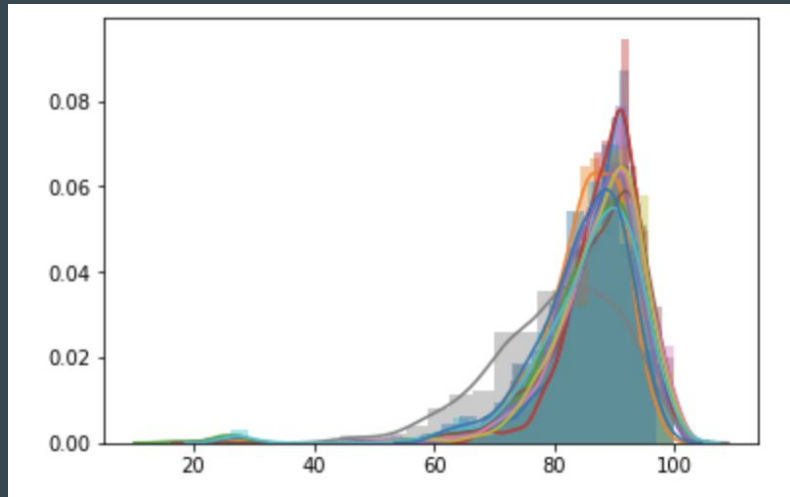


# A deeper dive into our data:



# Category relationship with Ratings

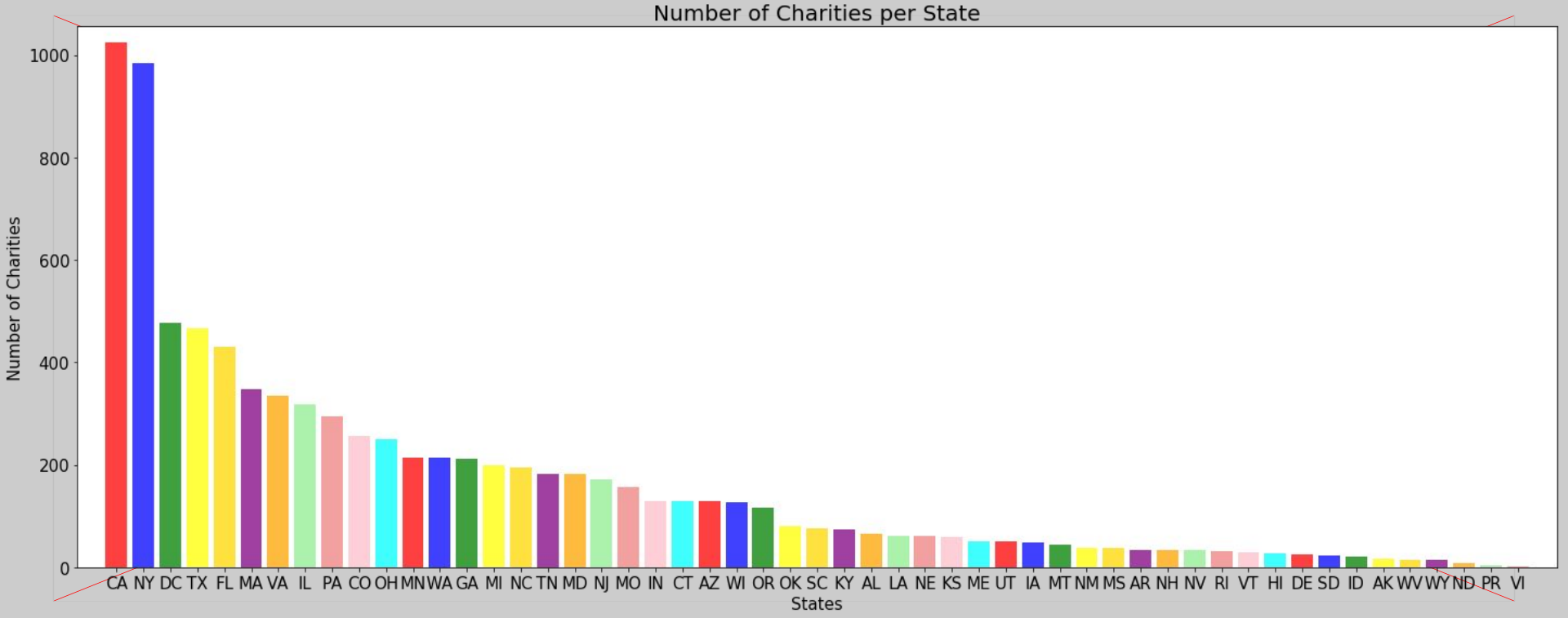
- We saw a noticeably lower rating on charities belonging to the religious category and running an ANOVA, we were able to confirm that there were significant differences between categories.
- We also saw that removing Religion had the largest impact to p-value compared to any other category. From 1.5e-86 to 3.13e-34
- The average rating for Religion was 80.5
    - This is compared to 86.6 generally



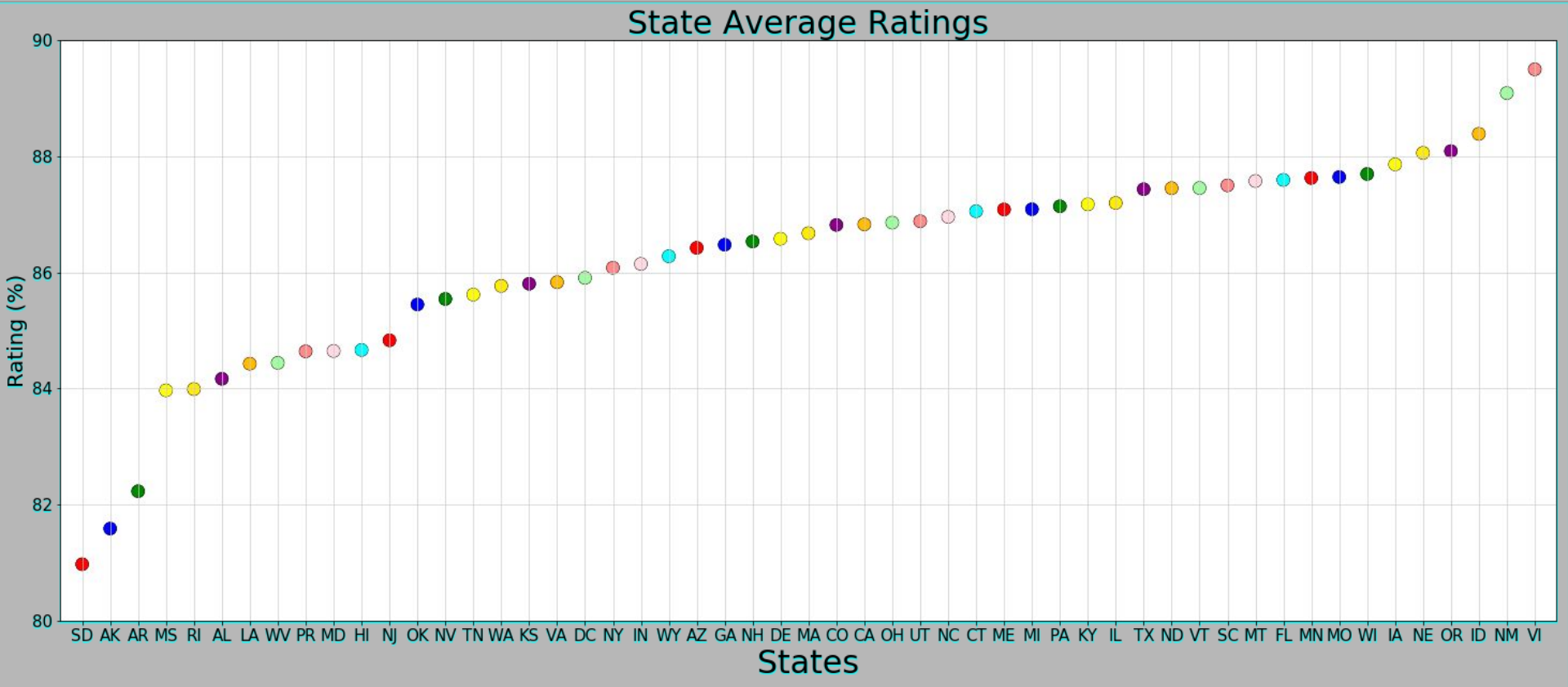
# Questions about the data we saw:

1) Which states/territories/districts have the most charities as their domicile?
2) Did more populous states such as Texas, New York, Florida, or California have higher rated charities on average versus less populated states?

# The Fab Five: CA, NY, DC, TX, & FL



# Conclusion: No relationship between populous states and ratings



# Questions?