Understanding US Non-Profit Efficacy

a charitynavigator.org web scraping exploration

US Non-Profit Sector Overview

The US non-profit sector accounts for $390B per year and 10% of the national workforce, so it is critical we understand how to allocate funds effectively

There are 1.6 million tax-exempt 501(c) organizations in the US that qualify as non-profits

Non-profits received $390 billion in 2016, equivalent to 2% of the US GDP or the size of Tencent, the world's 8th largest company

The US Non-profit sector provides 11.4 million jobs and employees 10% of the workforce, behind only the retail and manufacturing sectors

Key Questions for Exploration

The non-profit sector presents many questions for both donors / volunteers and those who run charities, a subset of which we aim to answer through this analysis

DONORS / VOLUNTEERS	CHARITIES
Which non-profit categories get the most funding? Which get the least?	What financial indicators should I look at to improve the effectiveness of my non-profit?
What indicators should I look at to help me decide what charity to donate to for maximum impact?	How should I brand my non-profit within a category to receive required funding to drive impact?
Which non-profit categories get the most funding? Which get the least?	How much should leadership and staff be compensated to attract talent while also ensuring fund sustainability?

Approach: CharityNavigator Overview

Charity Navigator is an organization with an aim of increasing access to impact philanthropy, and its website contians a broad set of useful data on US charities

OUR MISSION

MAKE IMPACTFUL PHILANTHROPY EASIER FOR ALL

OUR VISION

A WORLD WHERE PHILANTHROPY IS INCLUSIVE, VIBRANT, AND EFFECTIVE

CharityNavigator.org is a website that has consolidated information on 9,000+ charities, with basic data on the rest of the 1.6 million non-profits in the US

Charity Navigator collects a broad set of information on charities, including financials, programs, and other items to develop a holistic view of each non-profit
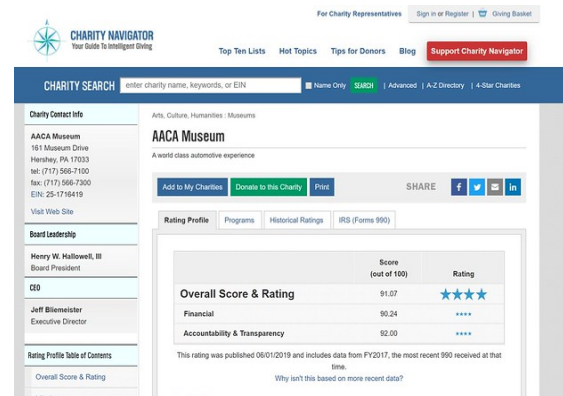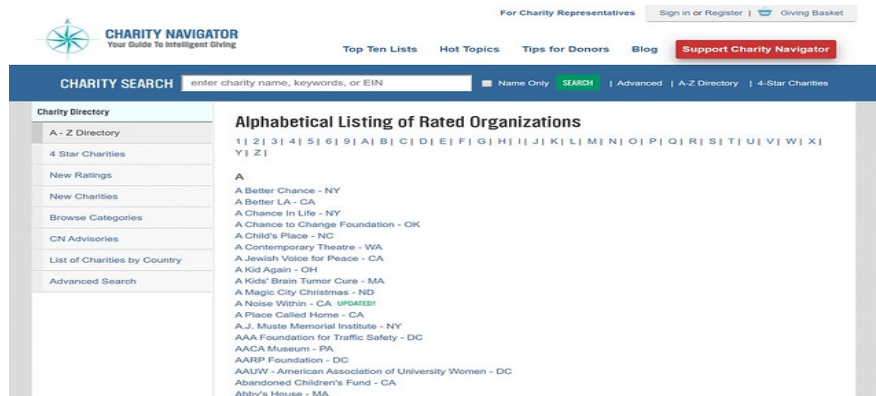
Information on each charity is aggregated into two scores (financial and accountability / transparency) which influence an overall 1-4 Star rating for each charity

Approach: Scraper Details

The developed scraper for the Charity Navigator website leverages Scrapy to crawl through all directories and pull information in each charity in the system

1) Directory   **2) Sub-Directories**   **3) Charity Detail Pages**


```python
def parse(self, response):
    # Find urls for the directory pages
    # that make up the set of charities
    directory_urls = response.xpath('//*
[@class="letters"][1]/a/@href').extract()
    for url in directory_urls:
        yield Request(url=url, callback=
            self.parse_directory_page)
```

```python
def parse_directory_page(self, response):
    # Find all the urls for charity pages
    # within the directory page
    charity_urls = response.xpath('//
    div[@class="mobile-padding charities"]
/a/@href').extract()
    for url in charity_urls:
        yield Request(url=url, callback=
            self.parse_charity_page)
```

```python
def parse_charity_page(self, response):
    # first scrape the information that's on all the charity pages, regardless of whether it is rated or not
    name = response.xpath('//h1[@class="charityname"]/text()').extract_first().strip()
    tagline = response.xpath('//h2[@class="tagline"]/text()').extract_first().strip()
    [category_l1, category_l2] = response.xpath('//p[@class="crumbs"]/text()').extract_first().strip().split(" :

    # Initialize a new CharityItem instance for each charity.
    item = CharityItem()
    item['name'] = name
    item['tagline'] = tagline
    item['category_l1'] = category_l1
    item['category_l2'] = category_l2

    location_line_flag = 1 * (re.search('([a-zA-Z]+)(\s?)[,][\s][A-Z]{2}[\s]([\d]{5})', re.sub('[\r\n\t\xa0]+',
    location = re.split('[\r\n\t\xa0,]+', response.xpath('//div[@id="leftnavcontent"]/div/p[1]/text()').extract(
    if '' in location:
        location.remove('')
    if len(location) == 3:
        location_city = location[0].strip()
        location_state = location[1].strip()
        location_zip = location[2].strip()
        item['location_city'] = location_city
        item['location_state'] = location_state
        item['location_zip'] = location_zip

    # the rest of the fields can only be filled in if the charity has a rating, so can check if the container obj
    # get the part of the page with the information we want using the xpath
```

Results Summary

The resulting data scraped includes information on 9,114 charities accounting for a total of $113B in charitable contributions

Scraped data on 9,114 charities across 11 different categories

Available charity data accounts for $113 billion in annual charitable contributions, equivalent to 29% of the total US contributions across <1% of organizations

Key Summary Statistics:
- Mean Overall Score (0-100 Scale): 86.83
- Mean Overall Rating (0-4 Star Scale): 3.22
- Breakdown of Overall Rating by Stars:
  - 4 Stars: 3,516 (39%)
  - 3 Stars: 3,854 (42%)
  - 2 Stars: 1,001 (11%)
  - 1 Star: 233 (3%)
  - 0 Stars: 56 (1%)
- Largest Charity (Cont.): Feeding America ($2.8B)
- Smallest Charity (Cont.): Center for Leadership Innovation ($23K)
- Average Contributions: $17.2M

Results Summary: Distribution of Scoring

The overall scores for the charities are highly skewed towards positive scores
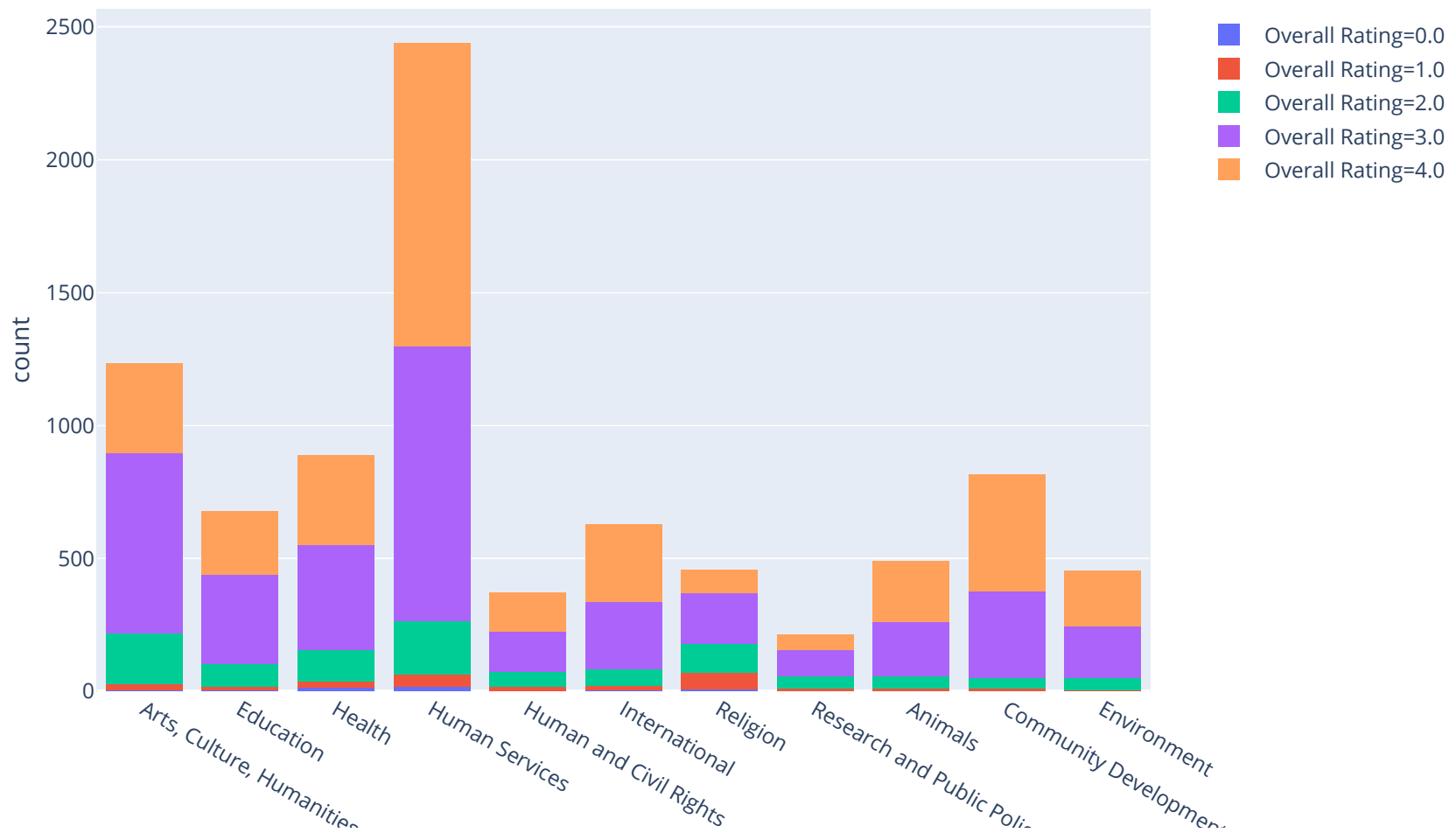
Histogram of Overall Score Broken Down by Overall Rating
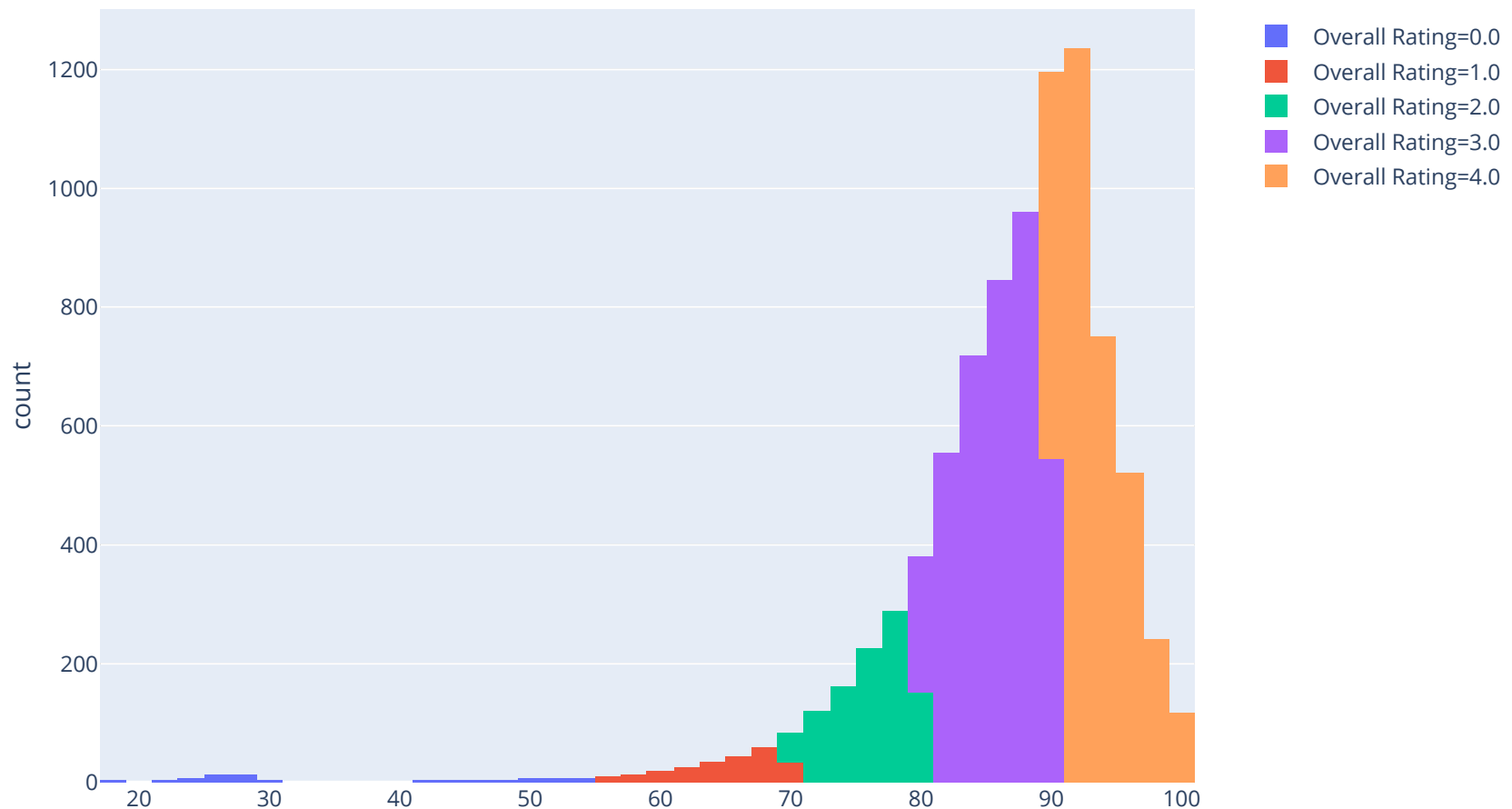
Results Summary: Charities by Category

Charities are spread across 11 categories, with the highest number in the Human Services category...

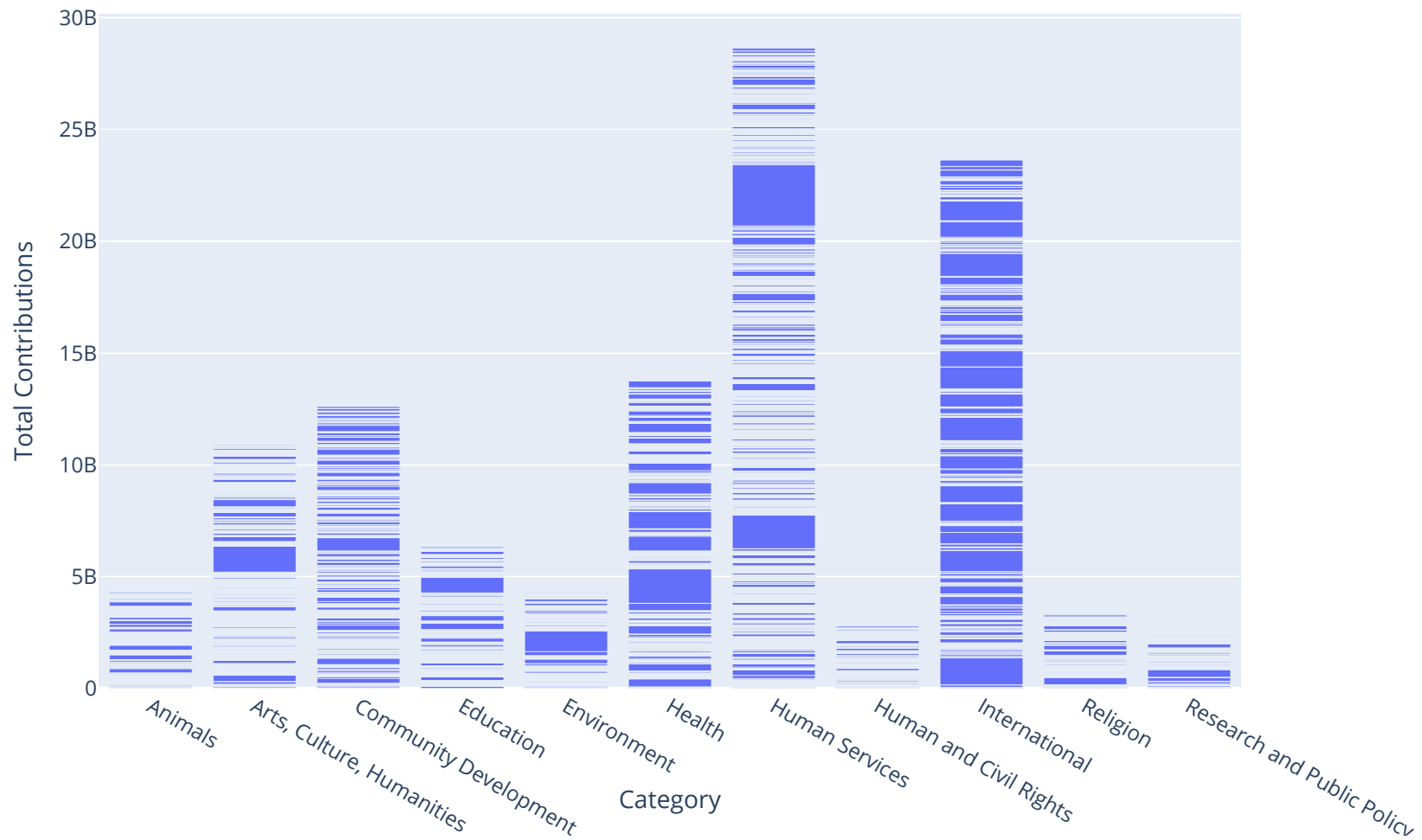Histogram of Charity Category Broken Down by Overall Rating

Results Summary: Charities by Category

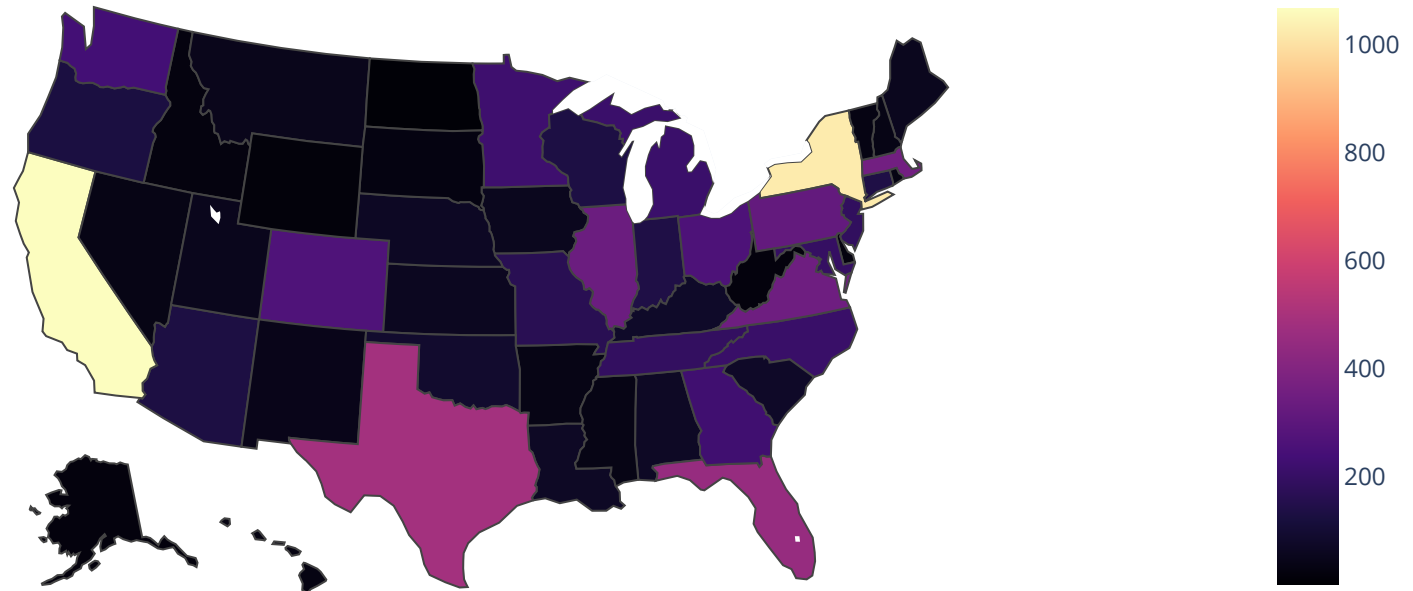...with the Human Services category also accounting for the highest proportion of contributions (~25%)

Bar Chart of Category against Total Contributions

Results Summary: Charities by State

The highest number of charities are present in California and New York...
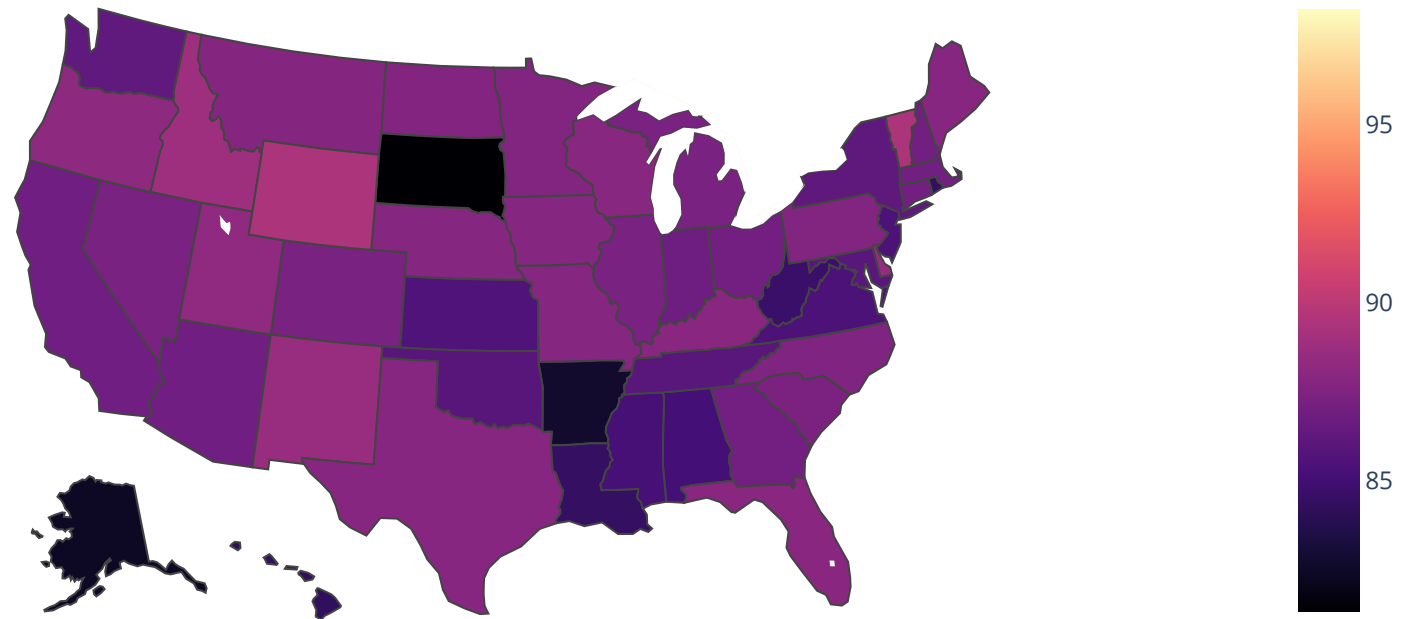
Map of Number of Charities by State

Results Summary: Avg Rating by State

...while charities in Wyoming and New Mexico have the highest average ratings

Map of Average Score by State

Charities Analysis Summary

Analysis was conducted to process missing data, analyze the charity mission through NLP, and explore the financial metrics of the charities

Missing Data Exploration and Processing

- Some charities on Charity Navigator do not have ratings or financial information
- Need to understand how many charities are missing information, and whether this skews our data set in any way

Basic NLP on Charity Mission Text

- Application of NLP functions on charities' missions to understand what language is typically used (overall and by category)
- Analysis of relationship between polarity / subjectivity of mission and overall charity score

Analysis of Charity Financials

- Analysis of the relationship between key financial indicators (e.g., funding efficiency, program expense ratio) and overall charity score / rating
- Exploration of leadership compensation at charities and its impacts on financial / charity effectiveness

Missing Value Processing

5% of charities did not have rating or financial information, so the distribution of this missing data was explored to ensure it would not skew the overall results

- Missing values existed across many fields for 454 charities (5%) due to no rating or financial information
- In addition, leader compensation was listed as "not compensated" for many rows, which was updated to $0
- Explored the distribution of missing vs. non-missing rows to ensure missing data did not skew the distribution:

Mission Text: WordClouds by Category
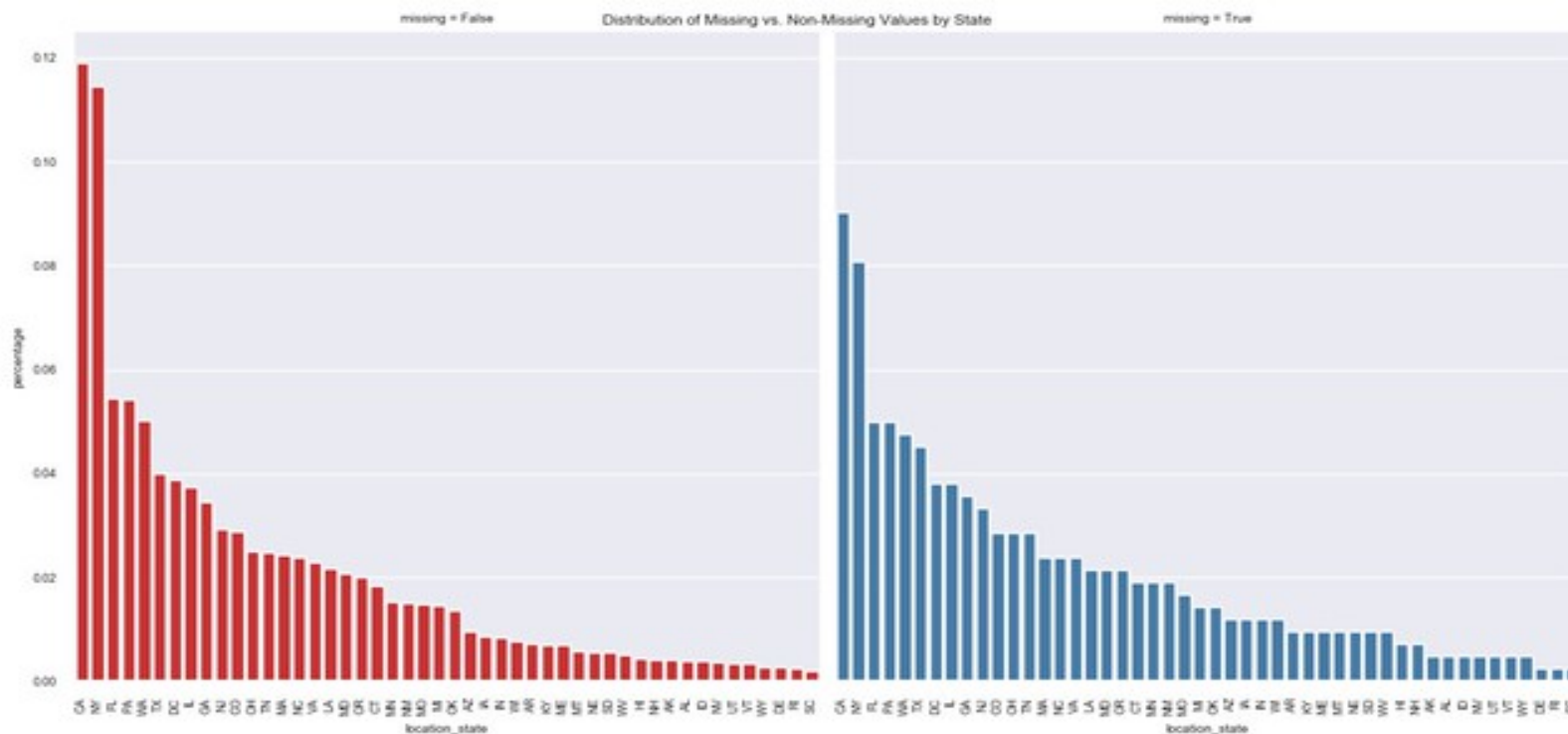
a charitynavigator.org web scraping exploration

Mission Text Wordcloud for All Charities
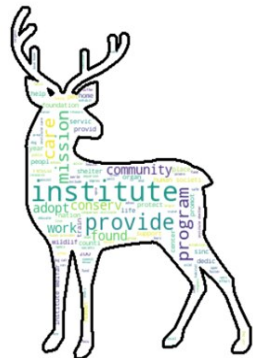

Mission Text Wordcloud for Arts, Culture, Humanities Charities
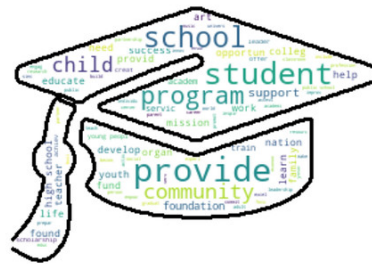

Mission Text Wordcloud for Community Development Charities
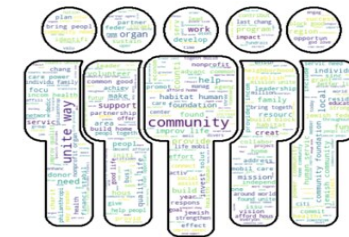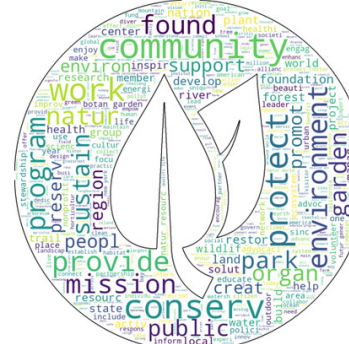

Mission Text Wordcloud for Animals Charities

Mission Text Wordcloud for Education Charities

Mission Text Wordcloud for Environment Charities
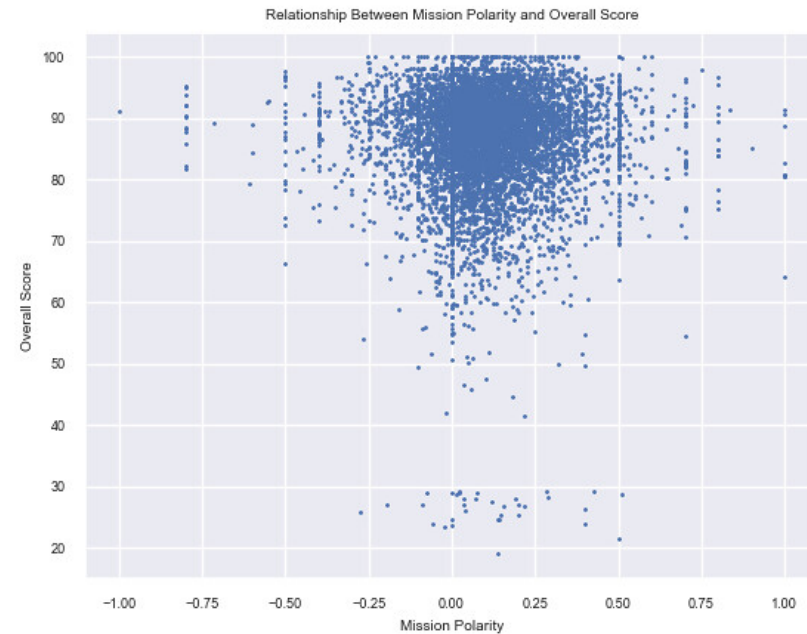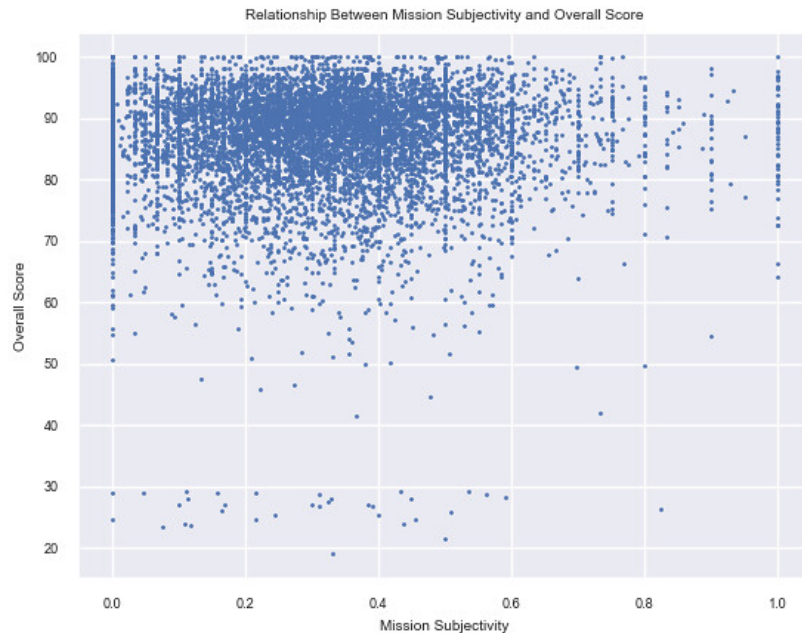
Mission Text: WordClouds by Category

WordClouds were developed to explore the common language in charities' mission statements (overall and by category) (2/2)

Mission Text Wordcloud for All Charities

Mission Text Wordcloud for Human and Civil Rights Charities

Mission Text Wordcloud for Research and Public Policy Charities

Mission Text Wordcloud for Health Charities

Mission Text Wordcloud for Human Services Charities

Mission Text Wordcloud for International Charities

Mission Text Wordcloud for Religion Charities

Mission Text: Polarity / Subjectivity

a charitynavigator.org web scraping exploration

Polarity and subjectivity were
calculated from the text in the
mission for each charity...

Relationship Between Mission Polarity and Overall Score

Relationship Between Mission Subjectivity and Overall Score

...and neither was found to have a
meaningful impact on score, although
charities with close to zero polarity
were more likely to have low scores

Financials: Charity Scale Distribution

Charities in the data set have wide variance in scale, from ~$100K-$1B in revenue

Histogram of Log Base 10 of Total Revenue Broken Down by Overall Rating

Legend:
- Overall Rating=0.0
- Overall Rating=1.0
- Overall Rating=2.0
- Overall Rating=3.0
- Overall Rating=4.0

count

Log Base 10 of Total Revenue
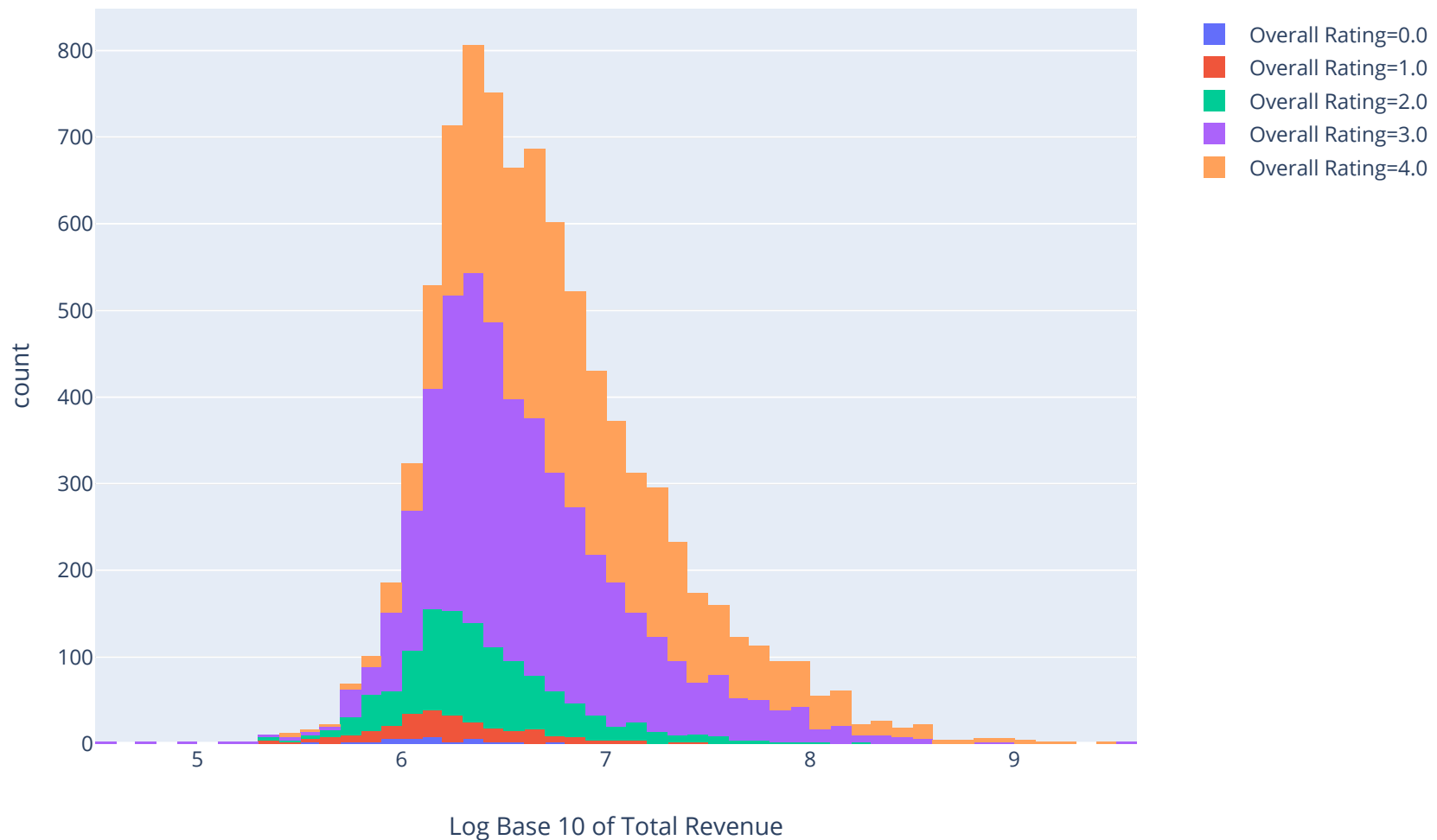
Financials: Program Expense Ratio

Program expense ratio shows a strong relationship with an improved overall score

Relationship Between Program Expense Ratio and Overall Score

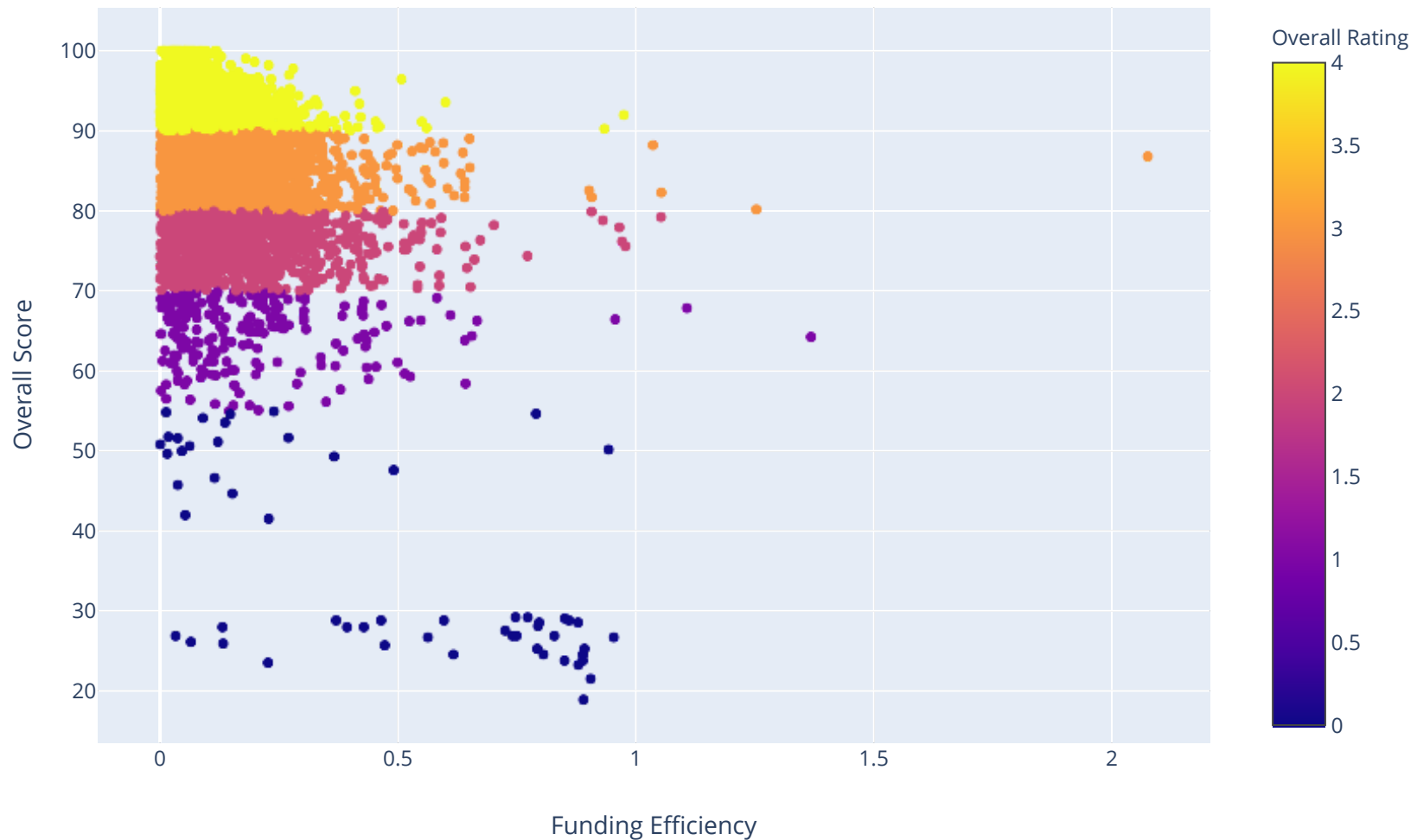**Program expense ratio is the ratio of overall expenses that are spent on delivering programs**
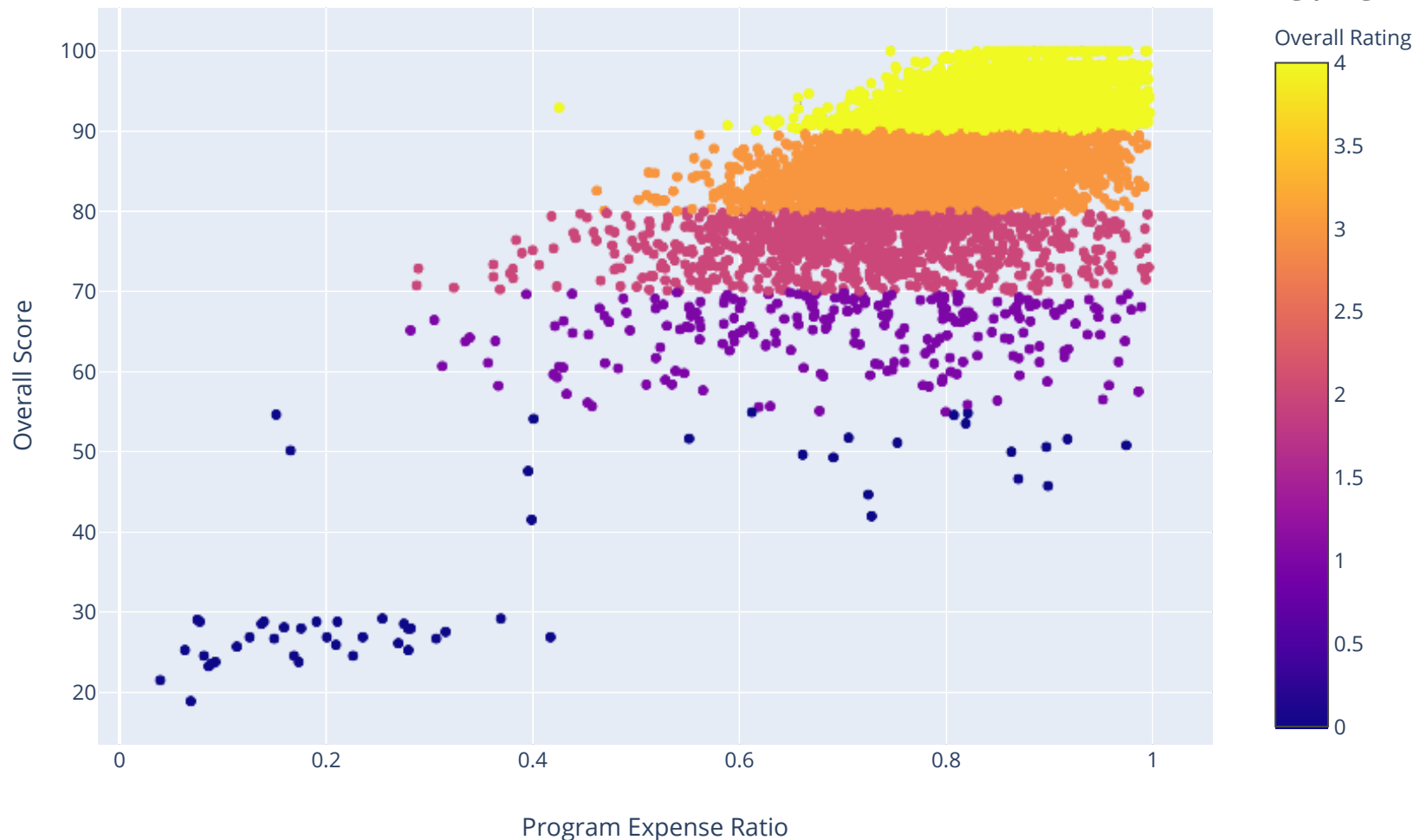
Financials: Fundraising Efficiency

Fundraising efficiency appears to have no impact on overall score

Relationship Between Funding Efficiency and Overall Score

Funding efficiency is the dollars of fundraising required to generate a dollar in contributions
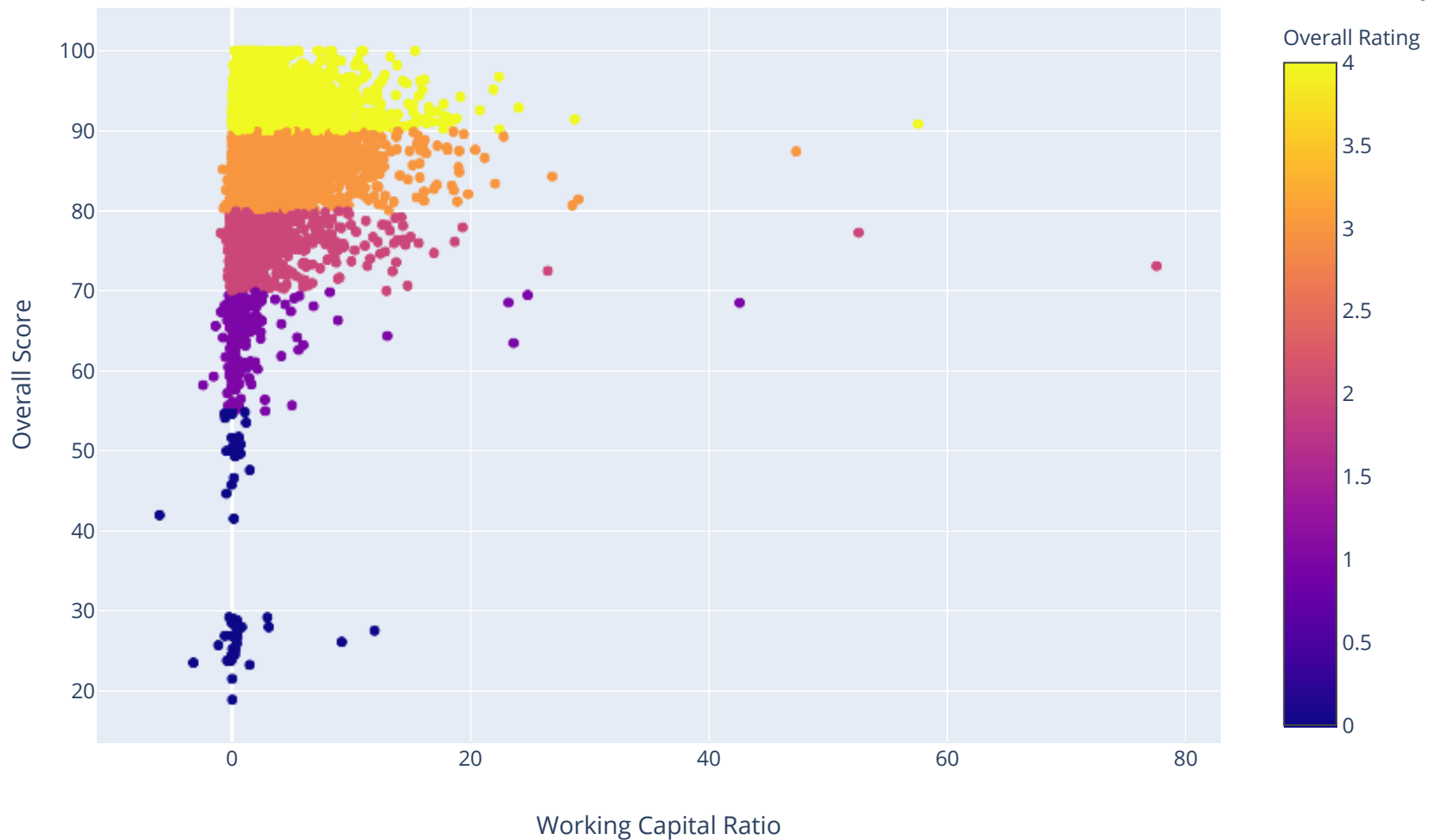
Financials: Working Capital Ratio

Working capital ratio has little impact on the charity's overall score

Relationship Between Working Capital Ratio and Overall Score

Working capital ratio is the number of years that a charity's cash (net assets) would cover all their expenses

Financials: Working Capital Ratio

Working capital ratio has little impact on the charity's overall score

Relationship Between Working Capital Ratio and Overall Score

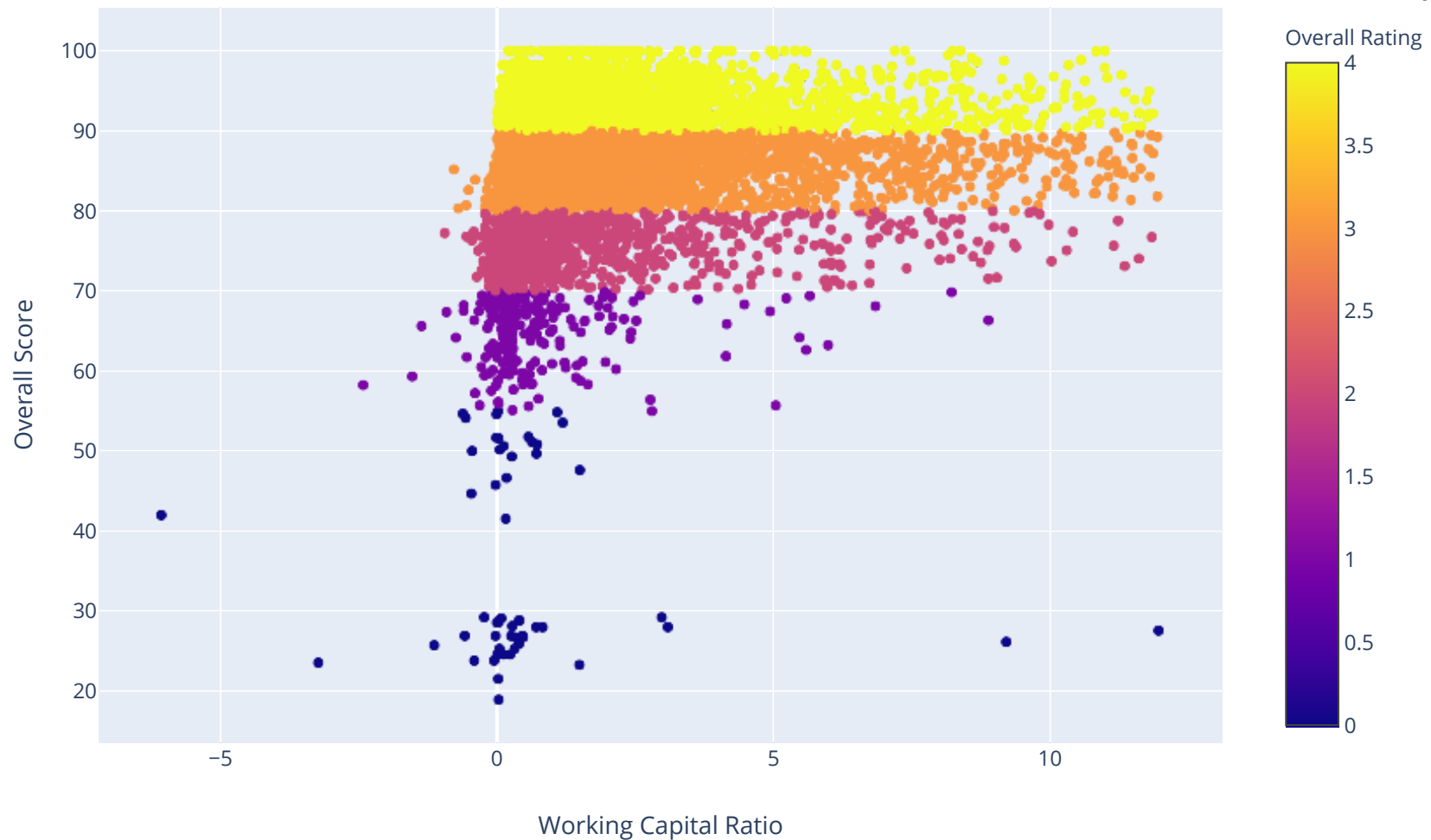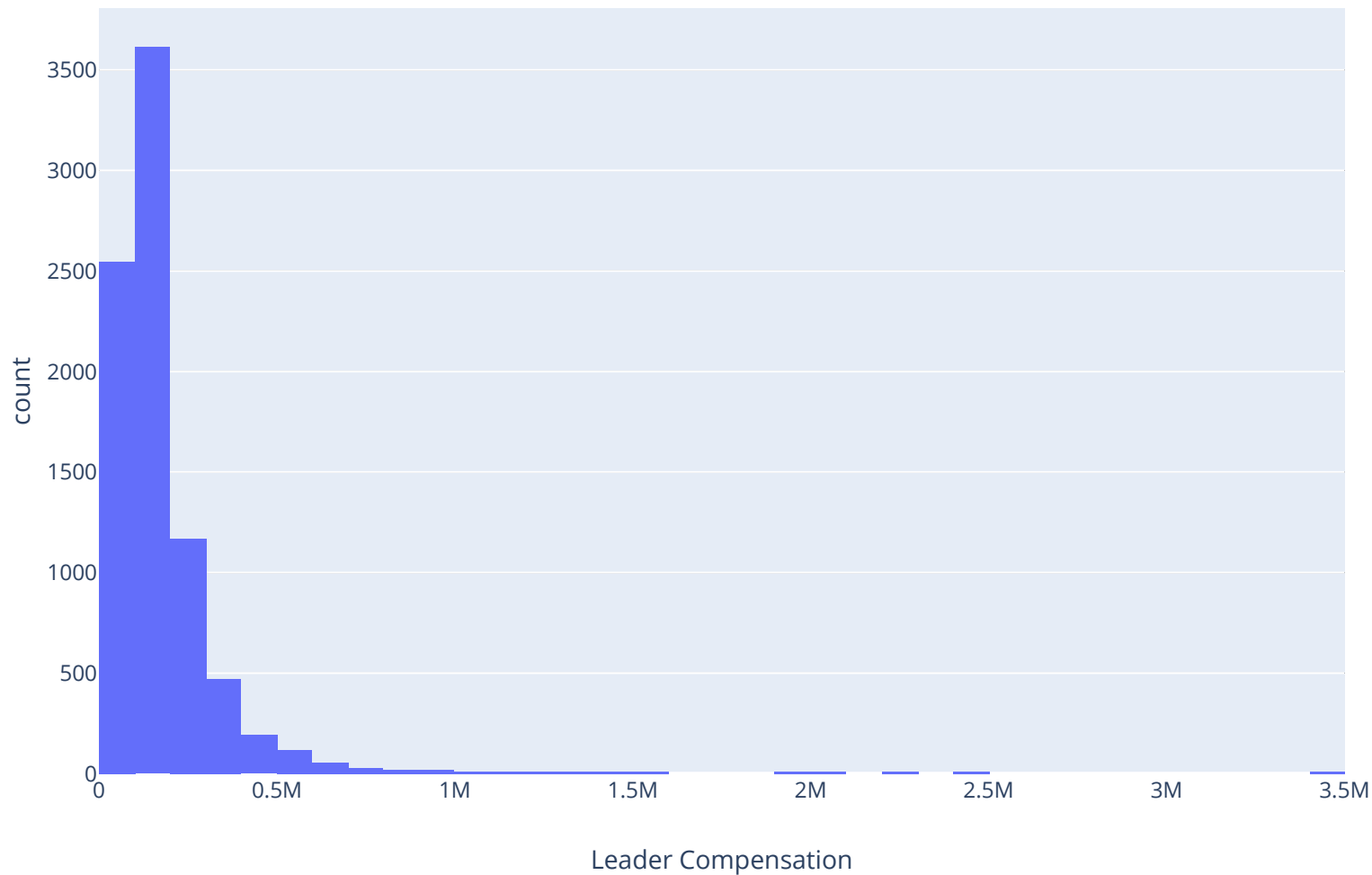**Working capital ratio is the number of years that a charity's cash (net assets) would cover all their expenses**

Financials: Leader Compensation

Compensation varies, but a majority compensate leaders in the 0-$300K range
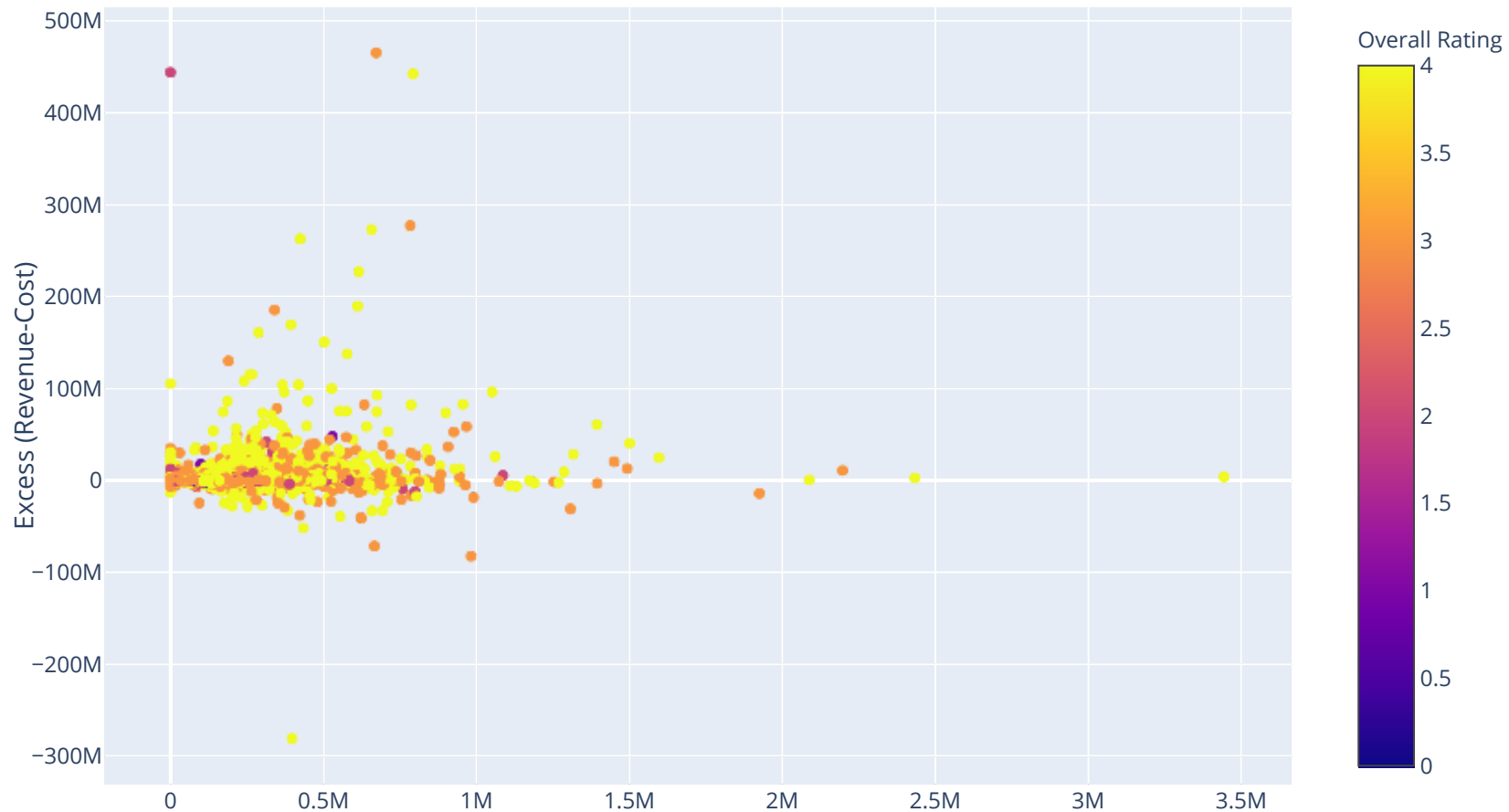
Histogram of Leader Compensation

Leader Compensation (x-axis), _count_ (y-axis)

Financials: Leader Compensation

The hypothesis of charities with highly paid leadership being less financially successful was not proven in the data

Relationship Between Leader Compensation and Excess (Revenue-Cost)

Findings and Future Focus Areas

Several future areas of focus can bring out additional insights from the Charity Navigator website data as well as other non-profit information

FINDINGS	FUTURE FOCUS AREAS

FINDINGS

- Contributions to US charities in the dataset are concentrated in a few categories (Human Services, International, Health)
- Of the financial indicators examined, program expense ratio has the strongest relationship with a high overall score / rating
- Polarity / subjectivity of messaging is not a major indicator of charity rating
- Plotly slideshow editor is not ready for prime time - stick with PPT next time :)

FUTURE FOCUS AREAS

- Analysis of additional fields available in the Charity Navigator data set, including binary attributes of each charity on its website / form 990
- Incorporation of additional data sets to cover greater proportion of US non-profits (currently 29% of funds across <1% of organizations)
- Develop tool to identify recommended high-impact non-profits to donate to based on donation preferences