Nonprofits in the United States

Elisabeth Zenger

December 18, 2018

Executive summary

I am emphasizing in nonprofit management, and I have been working with nonprofits throughout the state for the past year. The sheer number of nonprofits amazes me, so I decided to use the database of all nonprofits/charities registered with the IRS to find out a little more about them. My main questions are how many nonprofits exist, what purpose do they serve, and which nonprofits are most successful? I answered these questions through three different visuals that focused on (1) the number of nonprofits in each state; (2) the most common names of nonprofits; and (3) the type of nonprofits generating the most revenue in Utah.

I learned that the four states with the most nonprofits are the same four states with the highest populations: California, Texas, New York, and Florida. While this is hardly surprising, I think it emphasizes the fact that nonprofits try to fill gaps and remedy social problems, and more issues exist in highly populated areas.

The exploration of nonprofit names highlighted the fact that most nonprofits exist to build and strengthen the community. Of the fifteen most common words found in nonprofit names, 13 are associated with communities—-whether it's through education, civic engagement, or simply bringing groups of people with similar interests together.

The visualization of nonprofits in Utah with the highest revenue reinforced the community building idea. Utah has 40 nonprofits with revenues over 20 million dollars. Most are associated with higher education or health. I was surprised to find that the Utah Food Bank, Best Friends Animal Society, and Utah Symphony were among them.

The final graphic brings the three visualizations together to show where US nonprofits are and what they do. I tried to highlight the social mission of most nonprofits, as well as some local organizations that are extremely successful. The process of cleaning and creating each graphic will be discussed in the following sections.

Data background

The data came from "The Exempt Organization Business Master File Extract" (EO BMF) which includes cumulative information on tax-exempt organizations. The data is current as of August 14, 2017.

There are six data files separated by region. Each data file has information about 28 variables, including each nonprofit's name, EIN, address, income, assets, and revenue amounts. The first three data files contain data for the United States and the last three include international data.

Data cleaning

After downloading the libraries, I compiled the three US dataframes into one all_regions dataframe. I then created a clean data set with only six variables and a column that calculated the total number of nonprofits.

Next, I prepared the data for the state heatmap. I used the clean data to create a new object that included only the state abbreviation and the corresponding number of nonprofits. That dataset was joined to the state map dataset, which had been filtered to remove Alaska, Hawaii, and the Dominican Republic.

```
library(sf)
library(tidyverse)
library(scales)
library(plotly)


region_1 <- read_csv("data/us-charities-and-nonprofits/eo1.csv")
region_2 <- read_csv("data/us-charities-and-nonprofits/eo2.csv") %>%
  mutate(CLASSIFICATION = as.character(CLASSIFICATION))
region_3 <- read_csv("data/us-charities-and-nonprofits/eo3.csv")
state_maps_with_data <- state_maps %>%
  left_join(nonprofits_by_state, by = c("STUSPS" = "STATE"))
```

Individual figures

Figure 1

The first figure is a pretty straightforward heat map of the United State. I chose to do a heatmap because it is intuitive and familiar to readers. The darker the state, the more nonprofits there are.

The figure exhibits the qualities of great visualizations in the following ways:

Truthful: The map shows an accurate picture of where nonprofits are located in the United States. The data is a compilation of over a million nonprofits and helps the viewer to understand in a glance which states support the most nonprofits.

Functional: The graphic summarized the massive datasets in an easily understood way.

Beautiful: The color scheme is monochromatic and pleasing to the eye. The dark colors represent larger numbers of nonprofits, which is intuitive. Fonts and gridlines were changes in illustrator to enhance the beauty of the figure.

Insightful/Enlightening: The map helps viewers gain a general idea of where nonprofits are in the United States and how many nonprofits there are in each state. This information could be beneficial if you are hoping to work in the field or if you are hoping to start a nonprofit. The map brings up important considerations like market competition, supports, and networking.

CRAP Elements: The color scheme shows contrast between low numbers of nonprofits and high numbers. The color scheme is repeated in the legend and throughout the remaining graphics. The legend is centered here, but in the final version it is aligned on the right. Titles and legends were added in the final version.

```
US_map <- ggplot(state_maps_with_data) +
geom_sf(aes(fill = num_nonprofits), color = NA) +
coord_sf(crs = st_crs(102003)) +
scale_fill_gradient(low = "#E0FFE7", high = "#1C7F33", na.value = "grey98") +
guides(fill = guide_colorbar(title.position = "top",
title.hjust = 0.5,
title = "Nonprofit Density in the United States",
barwidth = 15, barheight = 0.5)) +
theme_void() +
theme(legend.position = "bottom")
US_map
```

Nonprofit Density in the United States
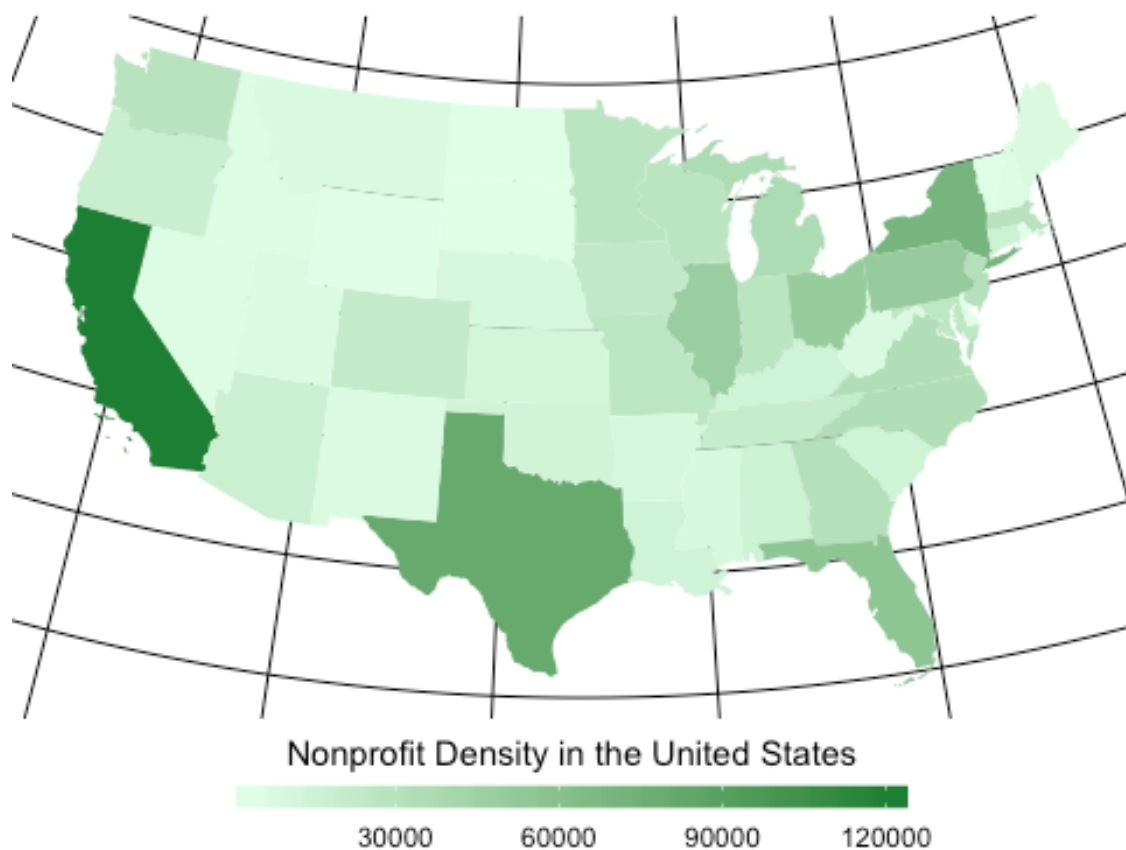
30000   60000   90000   120000

Figure 2

The second figure is a bar graph. The simplicity of the bar graph is helpful for viewers and accurately depicts the frequency of the distributions.

The figure exhibits the qualities of great visualizations in the following ways:

Truthful: The scale starts at zero and the words are ordered from largest to smallest, which helps readers interpret the information quickly and acurately.

Beautiful: The graph is simple and clean. The green matches the previous color scheme, and most of the gridlines are removed. The remaining gridlines serve as a reference point. The graph is flipped so that the bars and text are horizontal, making it easier to read.

Insightful/Enlightening: The findings highlight some interesting trends. First, many values are represented in the most common names, including national pride, religion, education, and communities. The words also highlight roles of nonprofits. Among these are associations, foundations, clubs, societies, and ministries. Taken together, the visualization helps to answer why we have nonprofits and the role they play in our society.

CRAP Elements:

Contrast: There is very little contrast in this graph. but the gridlines are very faded, which highlights the data. The distinct edges and lines also give the visualization a clean feel.
Repetition: The color scheme is repeated, as is the font.
Alignment: the text is aligned right, so that it is close to the bars. In the final graphic, the title is aligned left, flush with the text.
Proximity: Addressed in the final graphic.

```r
most_pop_names <- us_nonprofits_clean %>%
  unnest_tokens(word, NAME) %>%
  anti_join(stop_words) %>%
  count(word, sort = TRUE) %>%
  top_n(15) %>%
  mutate(word = fct_inorder(word))

## Joining, by = "word"

## Selecting by n

name_plot <- ggplot(most_pop_names, aes(x = fct_rev(word), y = n)) +
  geom_col(fill = "#1C7F33") +
  coord_flip() +
  scale_y_continuous(labels = scales::comma) +
  labs(y = "Nonprofits", x = NULL, title = "15 Most Frequent Words in Nonprof
it Names") +
  theme_minimal() +
 theme(panel.background = element_rect(fill = "transparent", colour = NA),
       plot.background = element_rect(fill = "transparent", colour = NA),
       axis.ticks.x = element_blank(),
       panel.grid.major.x = element_line(size = 0.5),
       panel.grid.major.y = element_blank(),
       panel.grid.minor = element_blank())
name_plot
```
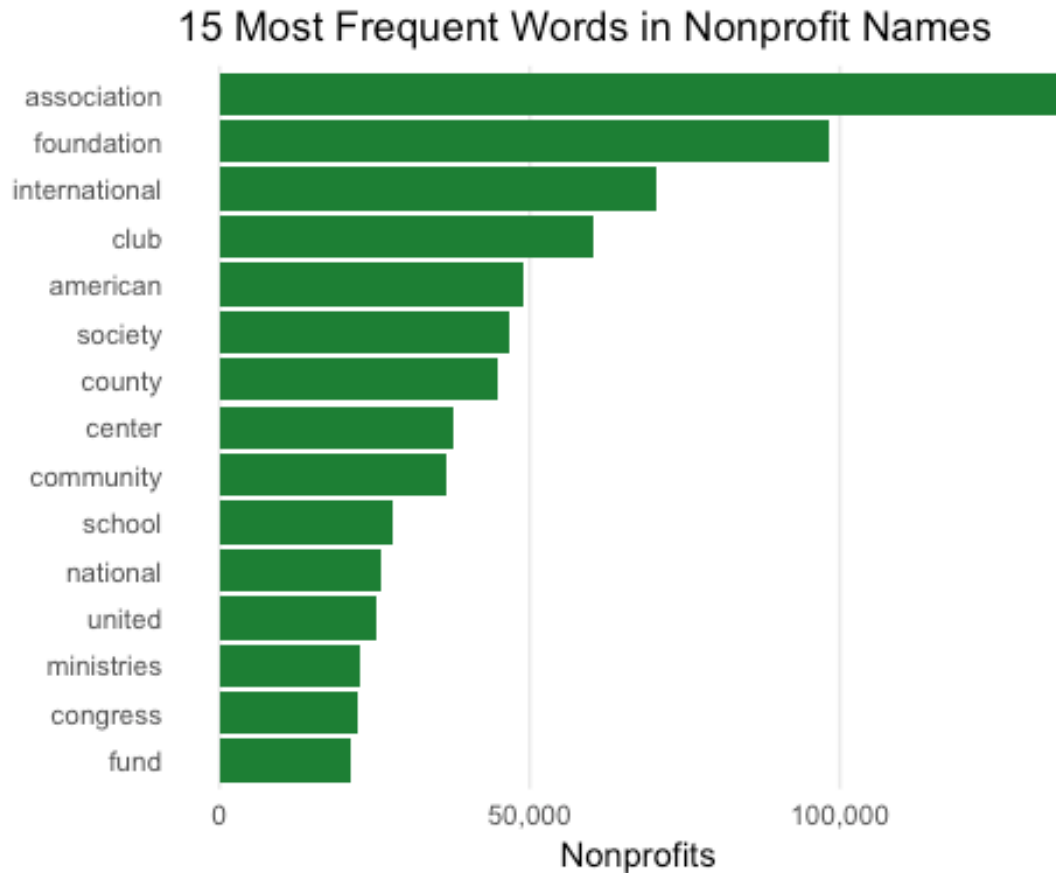
15 Most Frequent Words in Nonprofit Names

Figure 3

Figure 3 is a scatterplot designed to show the nonprofits with the most revenue. I started with data from all 50 states, then narrowed it down to Utah, then again to only nonprofits with revenue greater than $20M, because there was simply too much data and I wanted to be able to highlight individual nonprofits. The annotations are added in the final graphic.

The figure exhibits the qualities of great visualizations in the following ways:

Truthful: The scale starts at $30M, which could be confusing. However, I state that it is only nonprofits with more than $20M in revenue in the final graphic title. The chart is not designed to be exact, but to give a general idea of the large revenue amount and to highlight individual organizations.

Beautiful: The plot is very simple (no gridlines or background color). The same green is used for the dots.

Insightful/Enlightening: The results are both expected and surprising. Cities small cities like Loa have nonprofits with tons of revenue. The universities in Salt Lake were pretty expected, but I was interested to see some of the other nonprofits (like the ski association or the animal friend's association) in the >$20M plot. The visual is enlightening because it

helps you gain an understanding of who the most successful are (at least with revenue). I would love to do some further research into organizational structure, members, and volunteers.

CRAP Elements:

Contrast: The distance between the very high and high revenue nonprofits provides the contrast.
Repetition: The green color is repeated, as well as the same simple style and background.
Alignment: The text is aligned right, so that it is close to the dots. Proximity: In the final graphic, some of the the largest nonprofit names are added next to their dot. In addition, a green background is added to help the reader visualize all of the dots together.

```
Utah_Revenue <- us_nonprofits_clean %>%
  filter(STATE %in% "UT") %>%
  filter(!CITY %in% "SALT LAKE CITY") %>%
  filter(REVENUE_AMT > 20000000)

plot_revenue <- ggplot(Utah_Revenue, aes(x = REVENUE_AMT, y = factor(CITY, le
vels = rev(levels(factor(CITY)))))) +
  geom_point(shape = 16, size = 3, color = "#1C7F33", show.legend = FALSE) +
  scale_x_log10(labels=comma) +
  theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(
)) +
  theme(axis.ticks= element_blank(), axis.title.y = element_blank(), axis.tit
le.x = element_blank())
plot_revenue
```
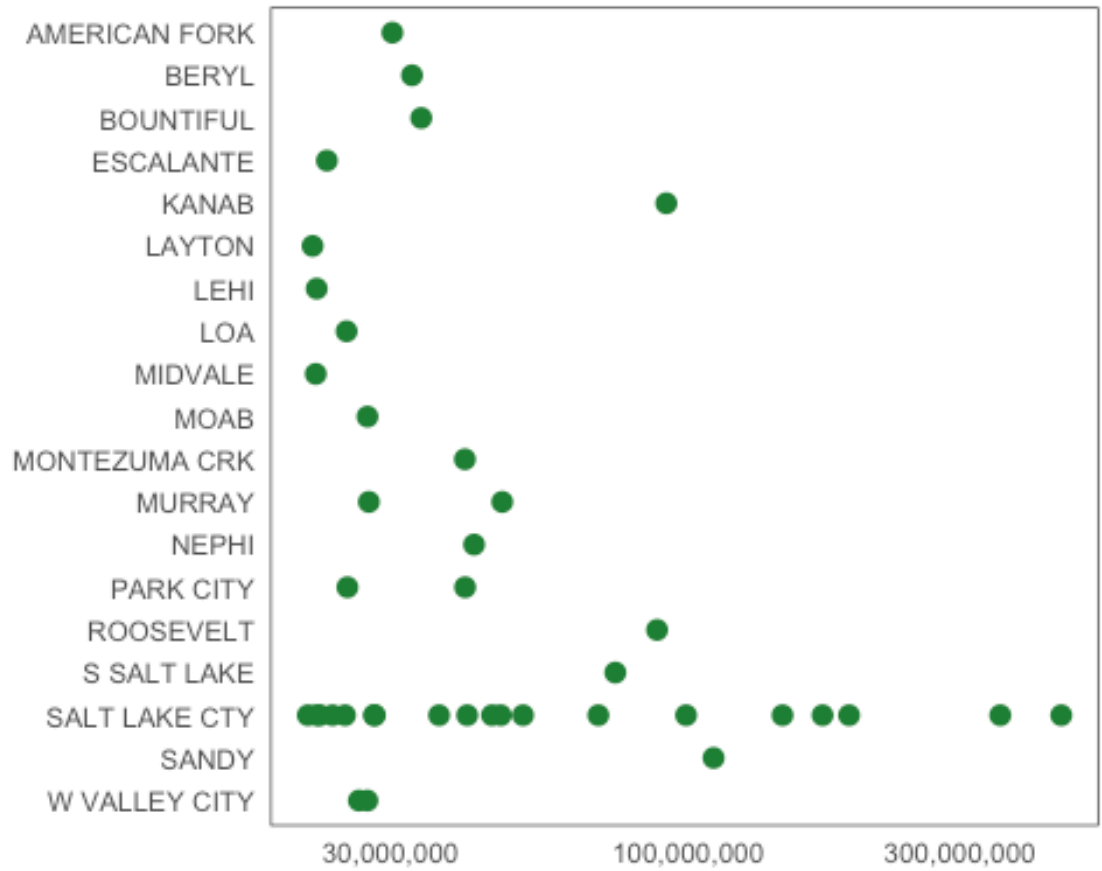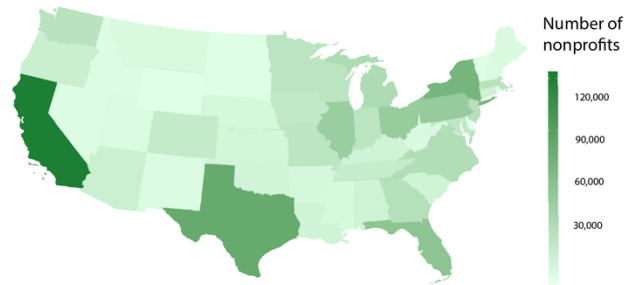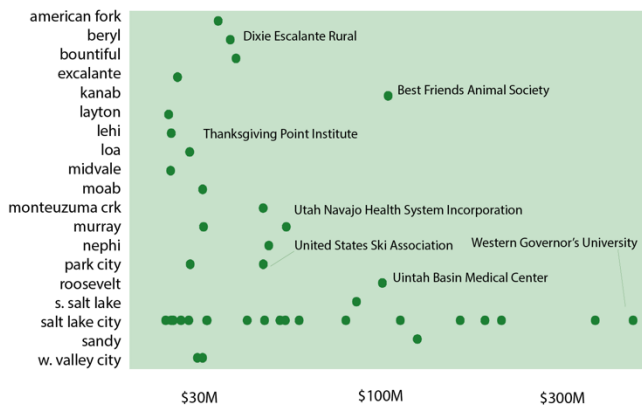
Final figure

Why do we have so many nonprofits?

The United States supports a staggering number of nonprofits. In fact, more than 1.6 million nonprofits and charities were registered with the IRS as of August 2017.
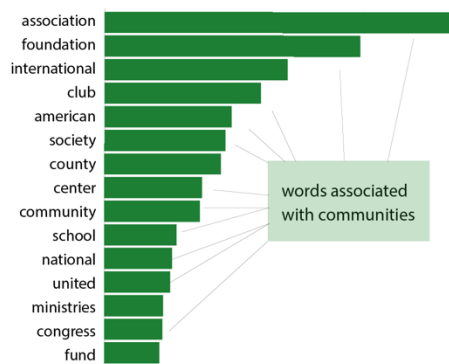
As illustrated in the map on the right, nonprofits are most dense in the states with the highest populations—not a bit shocker. What's a little more interesting is why they exist. Based on the data, I propose that strengthening communities is one of their main purposes.

Number of nonprofits

120,000

90,000

60,000

30,000

utah nonprofits bringing in more than $20M in revenue

- american fork
- beryl — Dixie Escalante Rural
- bountiful
- excalante
- kanab
- layton
- lehi — Thanksgiving Point Institute
- loa
- midvale
- moab
- monteuzuma crk — Utah Navajo Health System Incorporation
- murray
- nephi — United States Ski Association — Western Governor's University
- park city
- roosevelt
- s. salt lake
- salt lake city
- sandy
- w. valley city

Best Friends Animal Society

Uintah Basin Medical Center

$30M $100M $300M

the 15 most common words in nonprofit names

- association
- foundation
- international
- club
- american
- society
- county
- center
- community
- school
- national
- united
- ministries
- congress
- fund

words associated with communities

I'm probably a little biased after all the time that went into this final figure, but I think the final effect is quite charming. The infographic catches the eye and provides a lot of accessible information. I chose to keep the green color scheme going, which ties all of the pieces together. Green is a calming color, and I think it goes well with the nonprofit vibe.

I followed the proximity principle by having space between each graphic. The scatterplot was a little empty for the space, so I put the green background behind it to complete the unit. I made sure to line up the two bottom plots (titles and bars). I also kept the same font size for the titles and axis labels. A title just did not look right on top of the map, so I meshed it in with the legend title and talked about it in the paragraph blurb. I tried to repeat the lowercase letters throughout the visualization, as well as the left alignment.

Overall, I think it conveys a truthful message and draws attention to the story that I was trying to tell.