

Fiscal Health Package and Geospatial Dashboard: Updates

Christian Maino Vieytes

Fiscal Health Package

- ▶ **Goal:** provide users with automated calculation of critical fiscal health metrics for nonprofit organizations (NPOs) filing PC-990 and 990-EZ tax forms
- ▶ Particular emphasis on those using the NCCS 990 database
- ▶ These data contain a mixture of records from both Form 990 and 990-EZ filers

What is Form 990?

- ▶ Form 990 is for tax-exempt organizations, nonexempt charitable trusts, and section 527 political organizations that file Form 990 to provide the IRS with the information required by section 6033.
- ▶ Yields key information about nonprofit operations in the US
 - ▶ Geographic
 - ▶ Size
 - ▶ Financial
 - ▶ Personnel
- ▶ Extract geographic information from NPO key personnel

Fiscal Health Metrics

Fiscal Health Indicators

Data Wrangling

Metrics -

Debt to Asset Ratio

Definition

Variables

Tabulation

Standardize Scales

Metric Scope

Descriptive Statistics

Save Metrics

Definition

$$DAR = \frac{\text{Total Liabilities}}{\text{Total Assets}}$$

Notes: Variables should be equivalent for 990 and EZ filers in the IRS data.

Variables

Total liabilities, EOY

- On 990: Part X, line 26B
 - SOI PC EXTRACTS: totliabend
 - NCCS Core:
- On EZ: Part II, line 26B

Denominator: Total assets, EOY

- On 990: Part X, line 16B
- On EZ: Part II, line 25B

Fiscal Health Metrics

- ▶ Variables in 990 data come from fields in Form 990 or 990 EZ
- ▶ 20 health metrics total
- ▶ All are simple ratios
- ▶ Metrics vary in the number of fields required for computation of numerator or denominator
 - ▶ As many as 6 fields may be required

Function Composition: The Skeleton

- ▶ **Input:** User supplies column names for variables required in metric
 - ▶ A consideration: variables with “PC”, “EZ”, or “PZ” scope
- ▶ **Return:** a `data.frame` with the transformed (e.g., normalized) or winsorized (i.e., truncated) iterations of the metrics and accompanying plots of their distributions

Function Composition: Nuance

- ▶ Main target audience: researchers harnessing the 990 database
- ▶ Argument defaults: write functions with default column names being those in the
- ▶ Intended result: user supplies `data.frame` and function searches for pre-specified defaults and returns computed metrics or an error warning users of missing columns

Function Composition: More Nuance

- ▶ What about metrics that can be computed from both PC- and EZ-scope variables?
- ▶ **Potential Solution:** write functions that can handle a vector of two character strings for each of the EZ and PC scope variables but can also handle a user supplying a single string for a single column

Next Steps

- ▶ Take written conditionals and can them into helper function to allow program make decisions on what type of inputs are being supplied and what to output
- ▶ Finish building and then install package
- ▶ Publish

Geospatial Dashboard Updates

Overview of Steps

- ▶ Import/wrangle data
 - ▶ Update previously supplied code to incorporate new cycles of 1023-EZ data releases
- ▶ Generate 1023-EZ data shapefile and obtain other cartographic shapefiles to generate some maps (static)
- ▶ Scale up to dashboard

Data Import/Cleaning

- ▶ Time-intensive (slow steps)
- ▶ Two datasets: one containing addresses for NPOs and one for board members
- ▶ 1023-EZ data is easy to access
 - ▶ Fields require cleaning and standardization (in some cases across cycles)
 - ▶ Regex is useful
- ▶ Geocoding:
 - ▶ Census Geocoder: free; use `tidygeocoder` package (first pass)
 - ▶ Google Geocoding API: first 40k queries are free (clean-up)
 - ▶ PO Boxes and failed addresses using zip and city centroids (more clean-up)

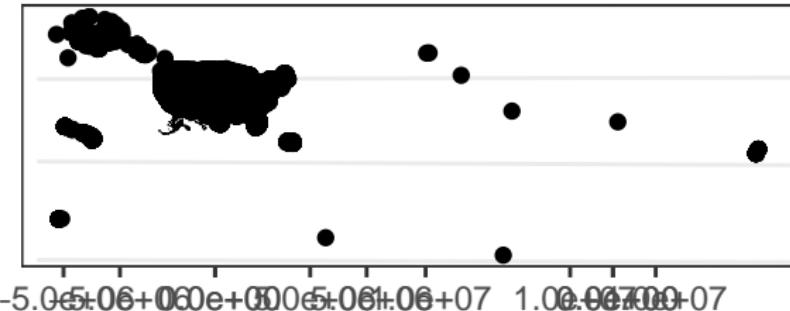
Data Import/Cleaning (cont.)

- ▶ Final Step: Integrated Public Use Microdata Series (IPUMS)
 - ▶ *Input:* supply a .csv of lat/long coordinates
 - ▶ *Return:* tract-level Census data associated with those coordinates
 - ▶ This will tell us something about the communities NPOs and board members reside in (*limitation:* not necessarily the communities they serve)

Map-Making

- ▶ Use `sf` package
- ▶ US county/state shapefiles provided through `urbnmapr` package
- ▶ Transform 1023 EZ data into a shapefile (points not polygons)
- ▶ Upon plotting initial map, we see lingering issues with the data...

Non-Profits in Questionable Locations



Issues:

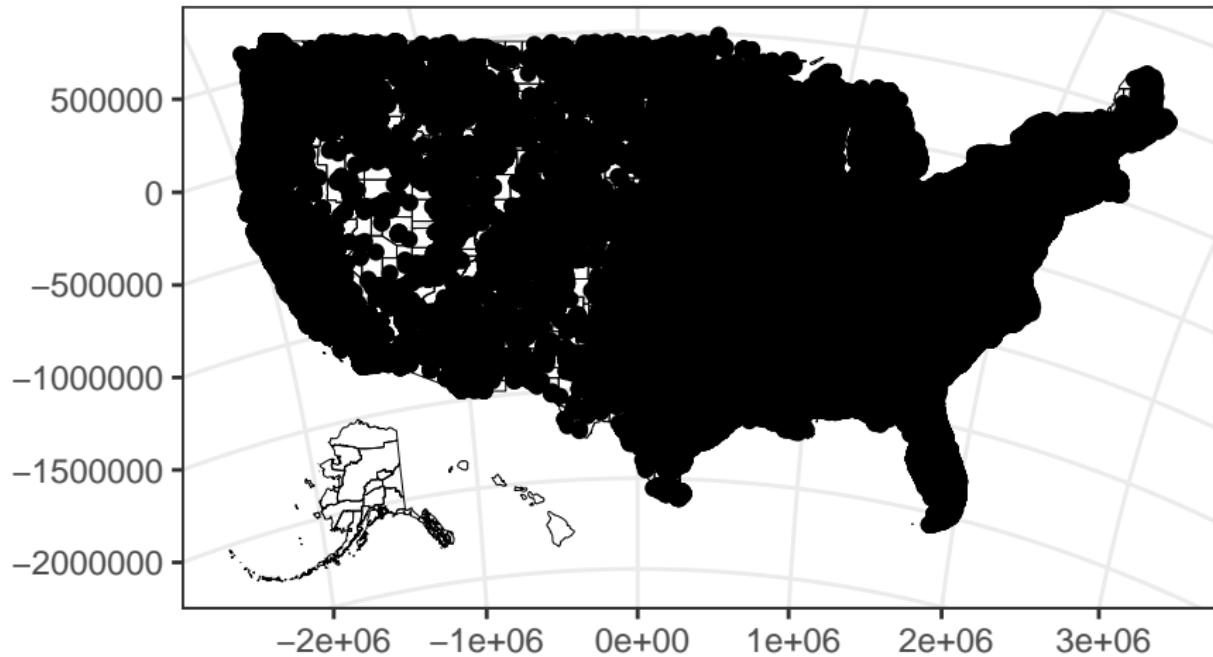
- ▶ The same phenomena persists after excluding Hawaii, Alaska, Puerto Rico, and Guam
- ▶ Projection issues with Cartesian coordinate system for Hawaii and Alaska
- ▶ Loading time is very much an issue using this level of data

Diagnosing The Issue

```
# Constrain lat long
# coordinates and diagnose
# the issue
np %>%
  filter(lat > 50 & State != "AK") %>%
  select(State, City, input_address)

## # A tibble: 6 x 3
##   State    City      input_address
##   <chr>   <chr>      <chr>
## 1 VA     WASILLA    9201 LILY LAKE ROAD, WASILLA, VA, 99623
## 2 AA     ANCHORAGE  11910 WHISPERING SPRUCE CIRCLE, ANCHORAGE, AA, 99516
## 3 NJ     WEST WINDSOR 1398 EDINBURGH DUTCH NECK RD, WEST WINDSOR, NJ, 08552
## 4 WA     TONASKET    31562 HIGHWAY 97, TONASKET, WA, 98885
## 5 AL     ANCHORAGE   721 DEPOT DRIVE, ANCHORAGE, AL, 99501
## 6 AL     ANCHOR POINT 32045 STERLING HWY, ANCHOR POINT, AL, 99556
```

Let's Zoom In

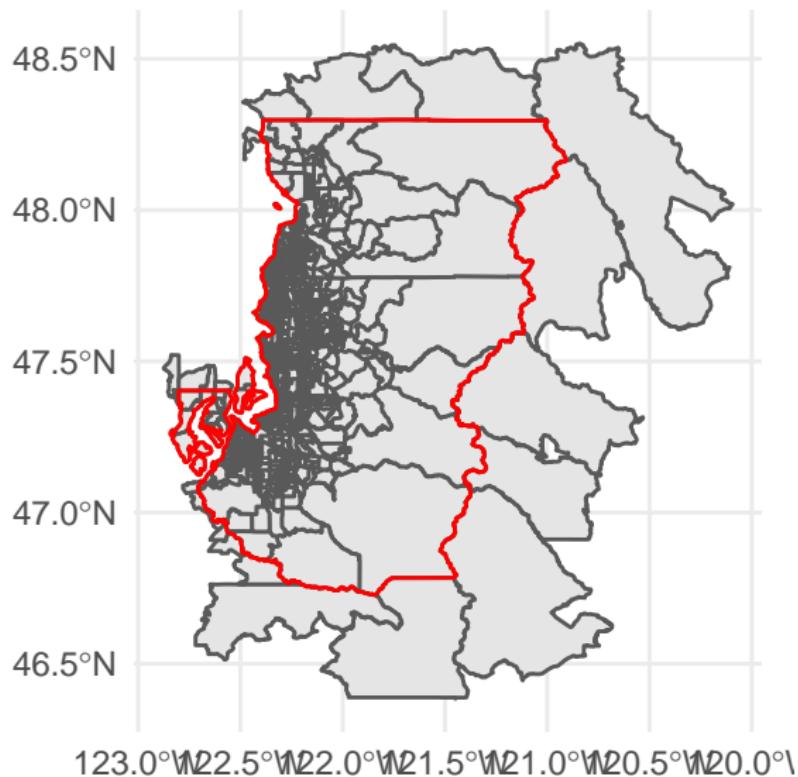


- ▶ **Issue:** Map of points on US map is not very informative
- ▶ **Try:** Density or bubble map
- ▶ **sf subsetting:** Metropolitan Statistical Areas (MSAs)

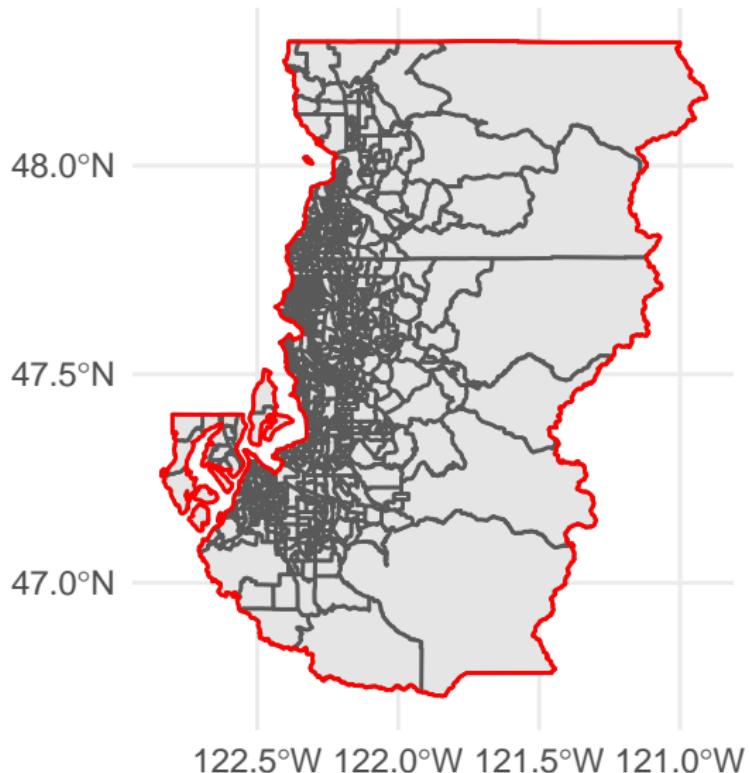
Mapping MSAs

- ▶ Subsetting MSAs and spatial overlay is not so straightforward
- ▶ `tigris` package: `core_based_statistical_areas` for fetching metro- or micropolitan region US Census TIGER shapefiles
- ▶ Tract boundaries within the statistical area are shown
- ▶ *Issue:* subsetting `sf` data returns census tracts grazing MSA borders
- ▶ *Solution:* `tigris` provides tools for addressing spatial overlay (i.e., `st_within`)

Example: Spatial Overlay



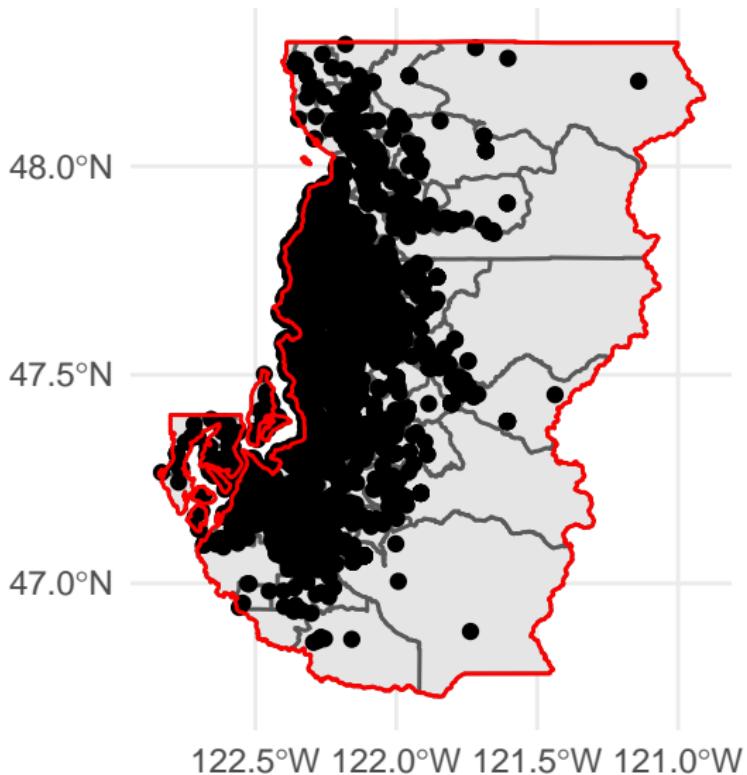
Example: Remove Peripheral Census Tracts at the Cartographic Boundaries with `st_within`



st_within for Subsetting 1023-EZ Data

- ▶ This function will play a key role in subsetting the 1023 data for overlaying on an MSA map
- ▶ Idea is to subset the 1023 EZ shapefiles to those sharing geometries within an MSA

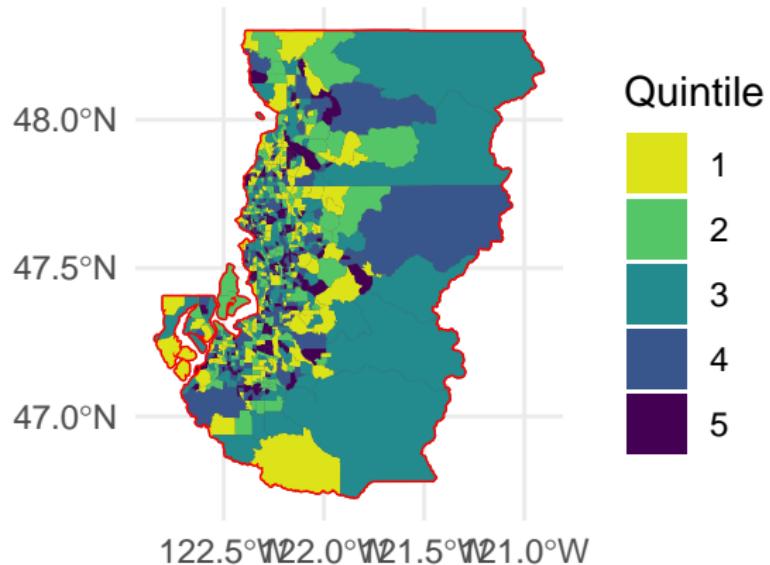
Example



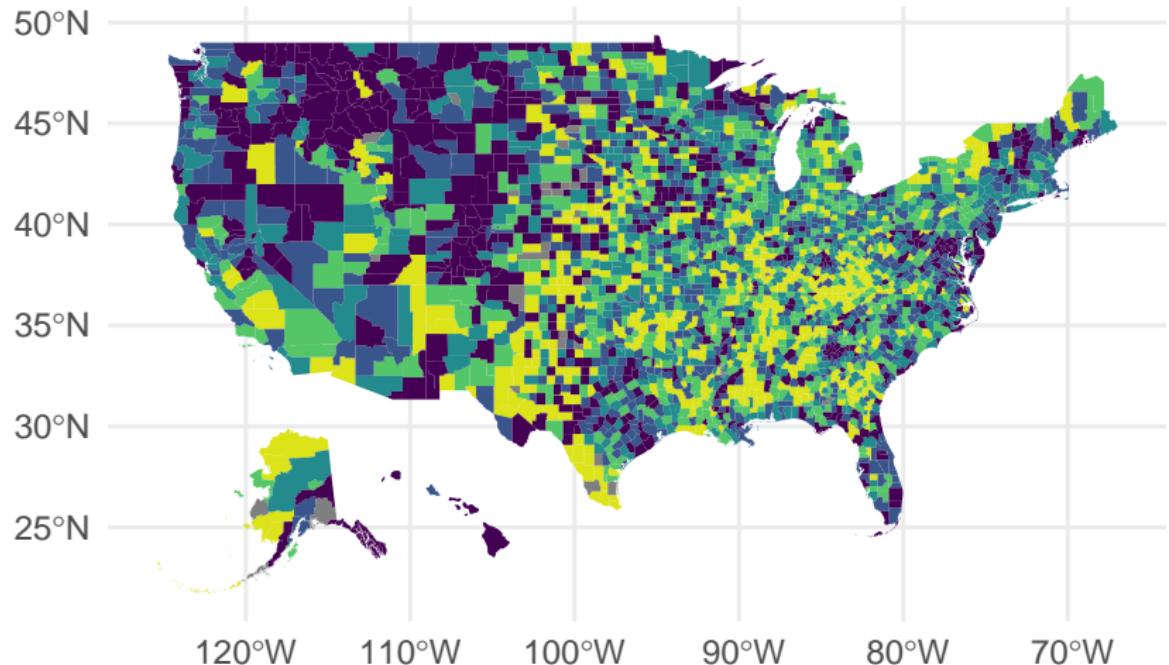
Issues Remain

- ▶ Again, even at this level, dots on a map are not very informative
- ▶ Let's try chloropleths and other approaches for modeling density
- ▶ *Problem:* tract FIPS codes are only available for those addresses geocoded through the Census Geocoder service (we also used Google Geocoder API and other approaches)
- ▶ *Solution:* we can harness data using from the `tigris` package and the `coords_to_fips` function from the `fipio` package

Nonprofit Density (Organizations/1,000 inhabitants)



Landing Page Map



Next Steps

- ▶ Finalize metrics to be used for NPO
- ▶ Begin working with board members data and finalizing metrics with this dataset
- ▶ Scale up to dashboard and get prototype/skeleton in place
- ▶ Conceptualizing widgets that will:
 - i) allow users to zoom into MSAs/other regions
 - ii) select a metric