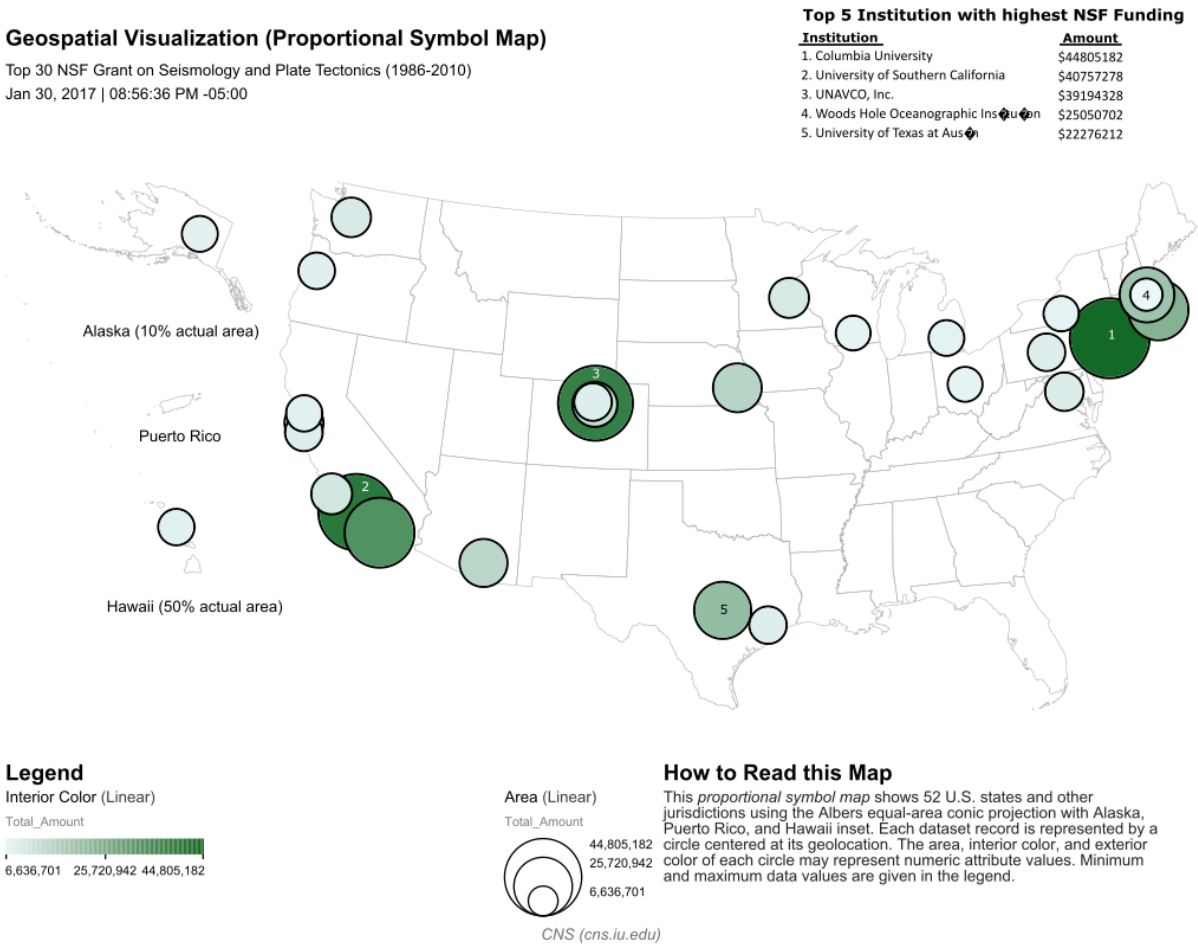
Homework 3

Samvat Rastogi
samrasto@indiana.edu

ILS-Z 637 Information Visualization

Geospatial Visualization (Proportional Symbol Map) – Approach

This document provides the other essential information about the Geospatial visualization which I have made. Below is a snapshot of visualization. A high-resolution copy of same can be found here.



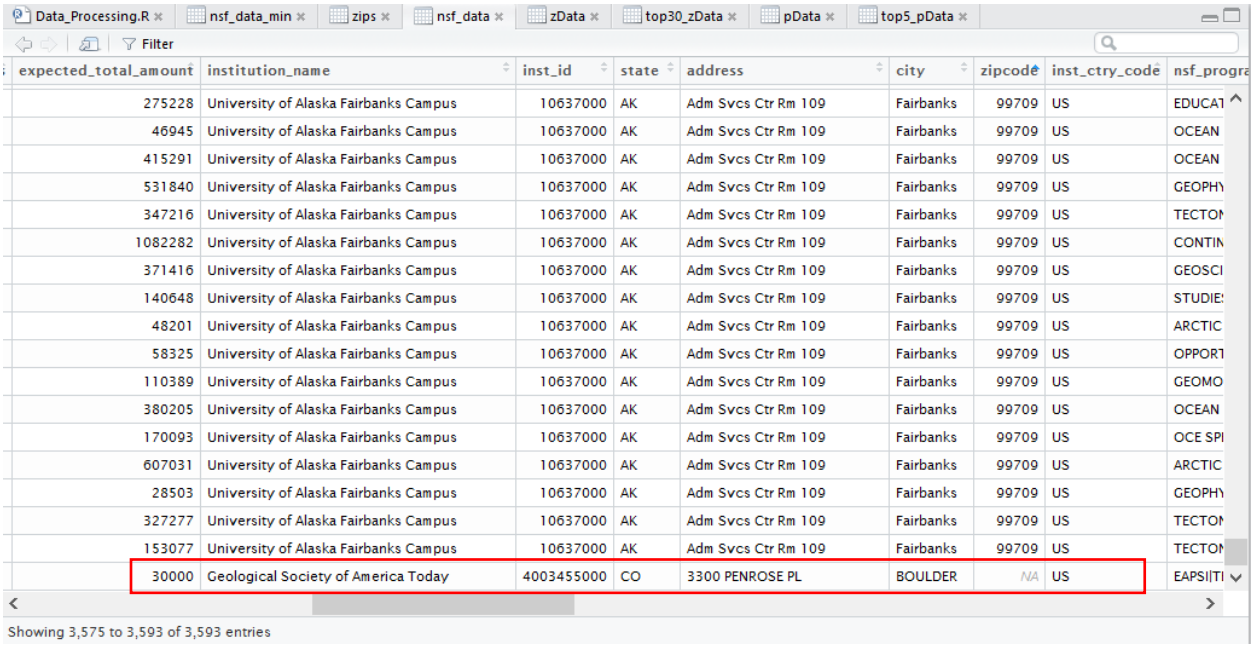
The above visualization is based on the dataset provided in the homework description in Canvas.

The Process:

1. I prepared the Data Processing script in R programming language. The script allows to select the dataset file from the file system and import into dataset.

2. Then, I removed the unnessary attributes or columns from the data and checked for any inconsistency of data.

3. On close observation, I found one inconsistency in the “zipcode” attribute. The zipcode for Institute “Geological Society of America Today” was in format of XXXXX-XXXX and was processed as “NA” value in R data frame. So, I replaced the value of zip code in XXXXX format. (From 80301-9140 to 80301)



Data_Processing.R | nsf_data_min | zips | nsf_data | zData | top30_zData | pData | top5_pData

Filter

expected_total_amount	institution_name	inst_id	state	address	city	zipcode	inst_ctry_code	nsf_progra
275228	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	EDUCAT
46945	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	OCEAN
415291	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	OCEAN
531840	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	GEOPHY
347216	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	TECTON
1082282	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	CONTIN
371416	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	GEOSCI
140648	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	STUDIES
48201	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	ARCTIC
58325	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	OPPORT
110389	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	GEOMO
380205	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	OCEAN
170093	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	OCE SPI
607031	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	ARCTIC
28503	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	GEOPHY
327277	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	TECTON
153077	University of Alaska Fairbanks Campus	10637000	AK	Adm Svcs Ctr Rm 109	Fairbanks	99709	US	TECTON
30000	Geological Society of America Today	4003455000	CO	3300 PENROSE PL	BOULDER	NA	US	EAPSITI

Showing 3,575 to 3,593 of 3,593 entries

4. Then, I aggregated the data based on zip codes and sorted it per the total funding amount. And I exported the top 30 rows in a csv file (geolocation.csv). Similarly, I aggregated the data again using Institution Names and fetched top 5 institutes with highest funding into a csv file.

5. Using generic geocoder in Sci2, I found out the Latitude and longitude against every zip code in geolocation.csv and saved it again as csv file. Using this csv, I generated my visualization and marked the top 5 institutes with highest funding on the map using institutes.csv with latitude and longitudes generated from generic geocoder.