Funding by Keywords and Program Types for IMLS Discretionary Funds (1996-2014)

Amanda Lee

27 October 2015

The purpose of this project is to, by matching funding amounts granted by the Institute of Museum and Library Services (IMLS) with program types and keywords, identify and possibly predict priorities for this independent US federal agency. While IMLS sets its agenda and presents its goals to the public, I purports that funding amounts are a truer reflection of what issues take precedence for any organization.

Background

Established in 1996 by the Museum and Library Services Act (MLSA), the Institute of Museum and Library Services is the main source of federal support for libraries and museums in the United States. In fiscal year 2015, the agency's budget was $227,860,000. From this amount, $28,724,000 was allotted to museums through the Museum Services Act and $180,909,000 was allotted through the Library Services Technology Act, which includes $154,848,000 for grants to states. (See Appendix A). The President has request $237,428,000 from Congress for 2016.

The data used in this endeavor is titled "Administrative Discretionary Grants FY 1996-2014" and the result of the agency's authorization to carry out and publish analyses of the impact of museum and library services. The data is sourced from and produced by IMLS itself. This dataset is produced each year, so there is another dataset that ends in 2013 available for the public to download and one ending in 2015 can be expected as well. For the 13,594 records, there are 36 features which contain information about the organization and project that was awarded the discretionary funds, and the program through which the funds were granted. (See Appendix B).

Analysis

The rest of this paper will be between the two exercises which are:

1. Funding Amounts for Program Types
2. Funding Amounts for Keywords

The processes are ultimately very similar and I will try to avoid unnecessarily repeating explanations.

1. Funding Amounts for Program Types

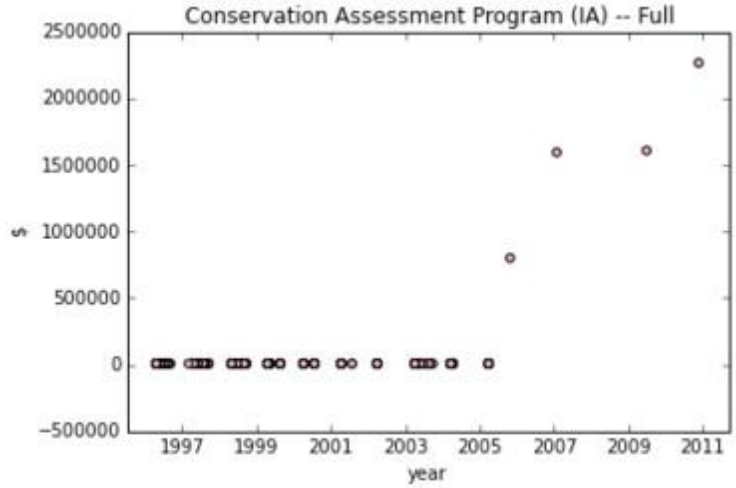
This exercise plots dollar amounts over time for each of the 18 program types.

There are 18 programs through which awardees receive funding and each program name has a matching two-letter designation called the program type. "National Leadership Grants" is an exception in that it encompasses 10 program types (different National Leadership Grants programs). Program types specify the subject matter somewhat. For instance, programs could pertain to library services, museum conservation or the administration of either, or community outreach or exposure to minority cultures. Two-letter program types were preferred in this exercise over the lengthier and spelling-error-prone program name.

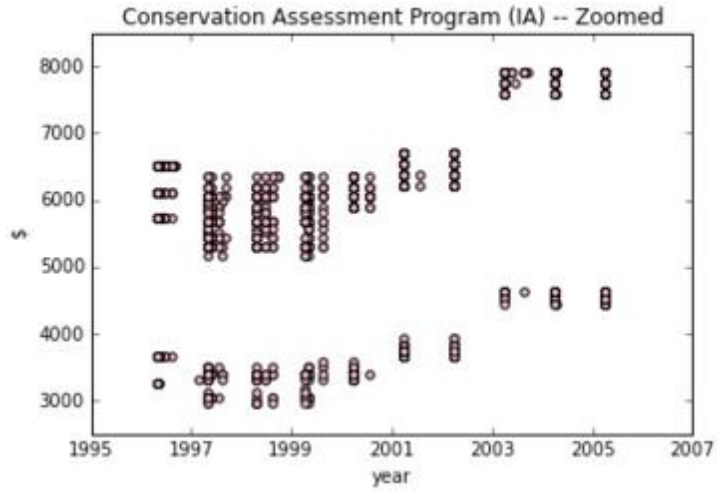
There are several features that contain funding amounts requested or originally awarded, but to determine how much a project is worth to IMLS, I chose the "Total Amount Awarded and Disbursed." These amounts are stored as a string preceded by a dollar sign. I created a new list that contains the items in this series without the dollar sign and converted into a float. This new list was appended to the dataset and named "totAward_float."

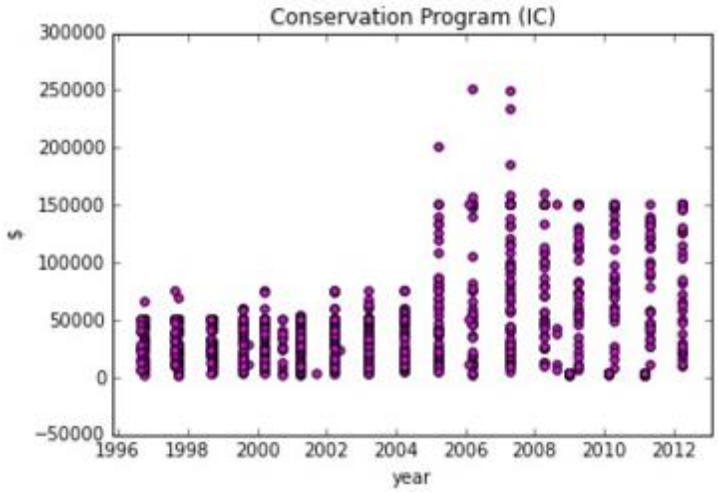
The "Award Date" feature contained both strings and floats, which were NaN values. It had much fewer NaN values than either "Award Period from" and "Award Period To," however. Otherwise it would have been interesting to measure how much an awarded amount is expected to last which speaks to the duration, scope, and perhaps impact of a project. Using the datetime module in a for loop, string values were converted to datetime values and NaN values were ignored. This was also appended to the dataset as "date_asTime."

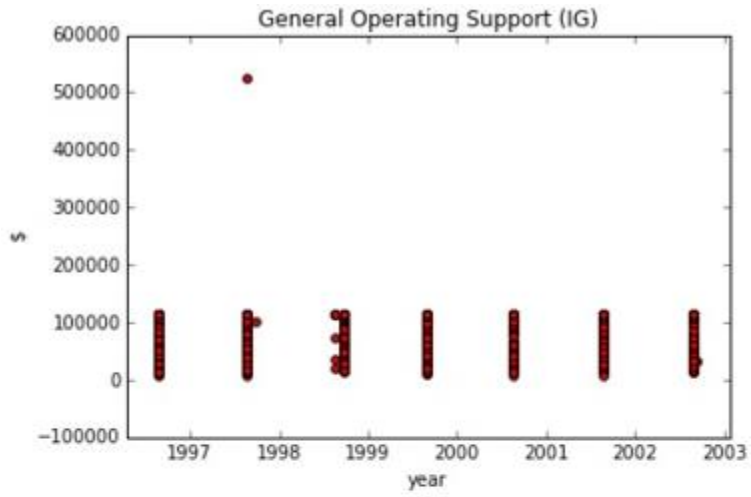
Plotting NaN values in datetime produced points in the 17th Century and augumented the scale of some scatter plots so it became necessary to create a new dataset without NaN values in the "date_asTime" column. This new dataset had 13 less records.



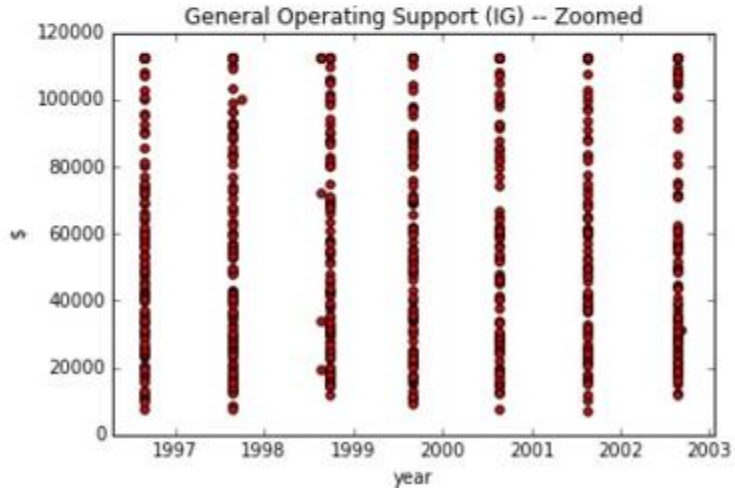

```
<matplotlib.text.Text at 0x10999b4d0>
```

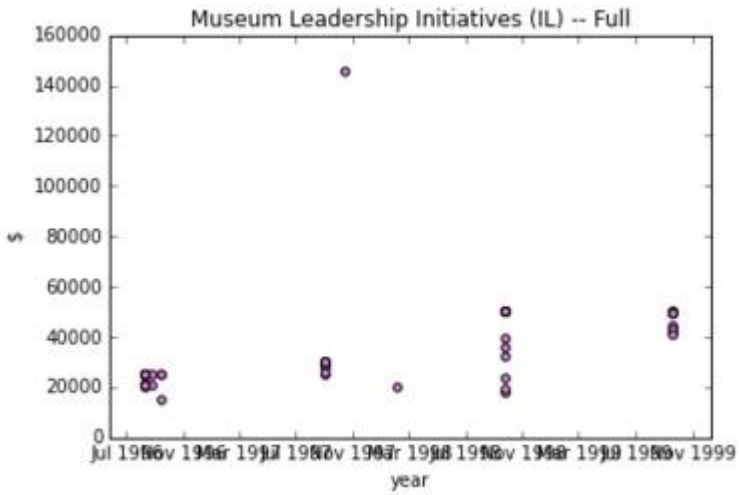




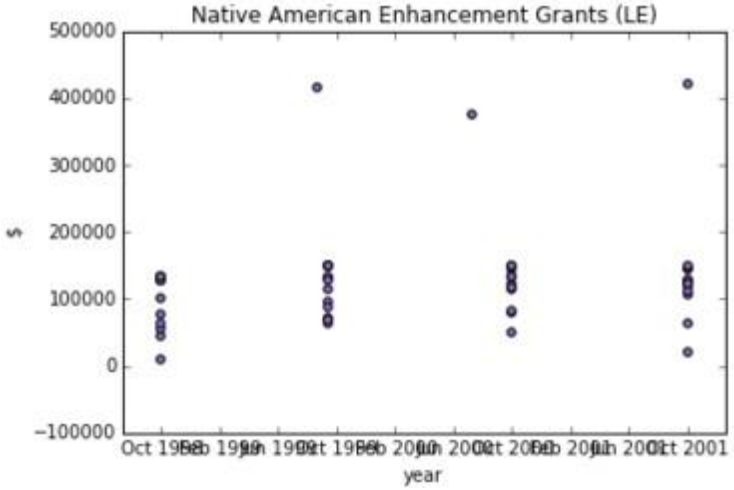

General Operating Support (IG)
600000
500000
400000
300000
200000
100000
0
−100000
$
1997
1998
1999
2000
2001
2002
2003
year

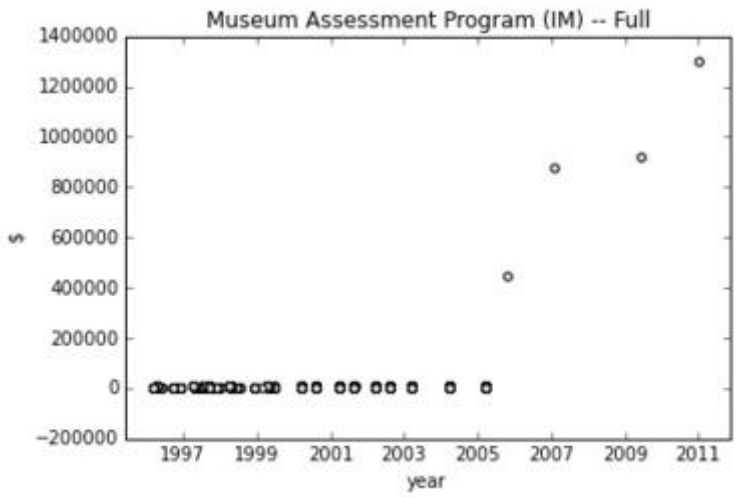
```
(0, 120000)
```


General Operating Support (IG) -- Zoomed
120000
100000
80000
60000
40000
20000
0
$
1997
1998
1999
2000
2001
2002
2003
year

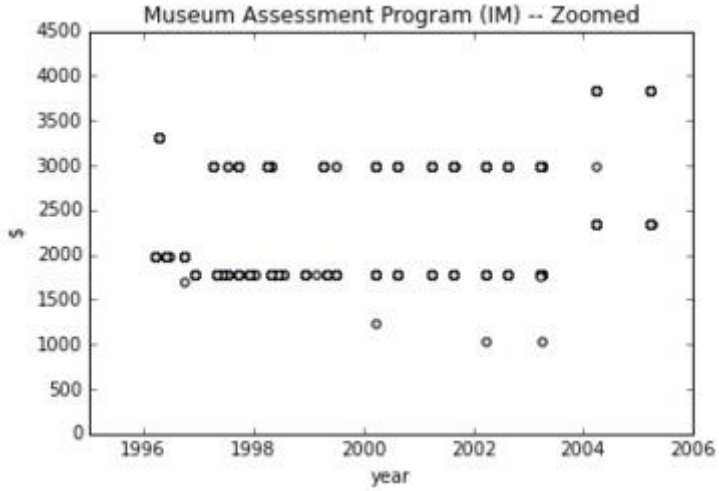


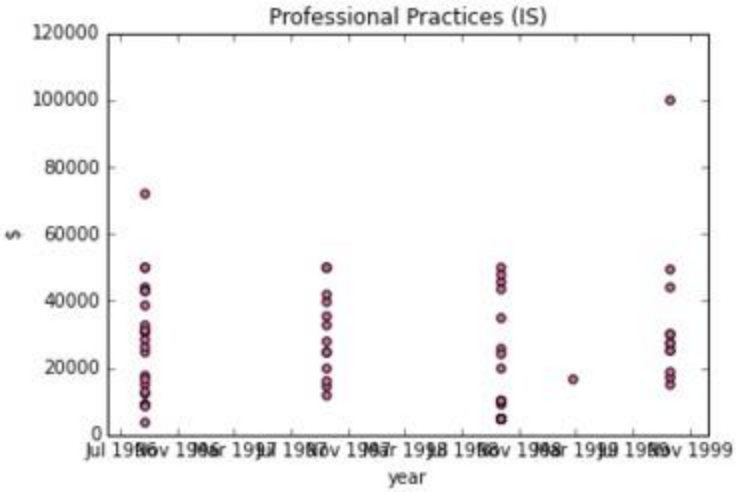

```
<matplotlib.text.Text at 0x10b01c190>
```

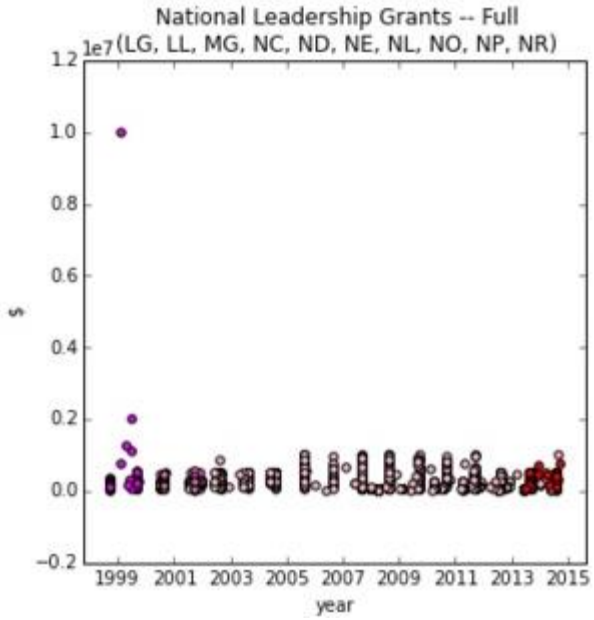


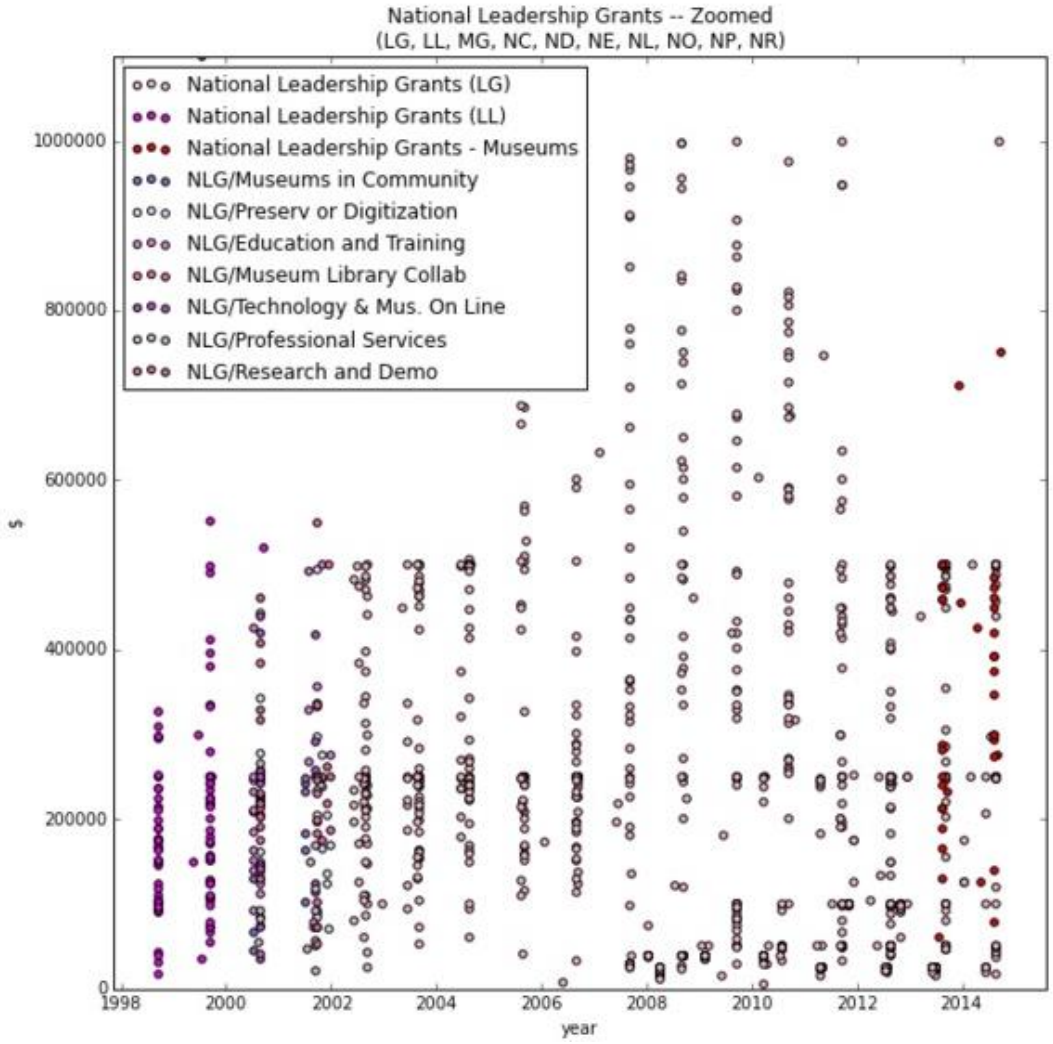


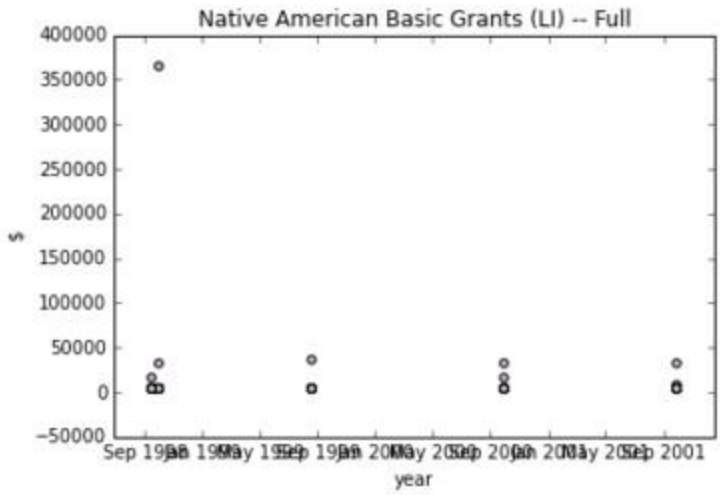
(728294.0, 732312.0)



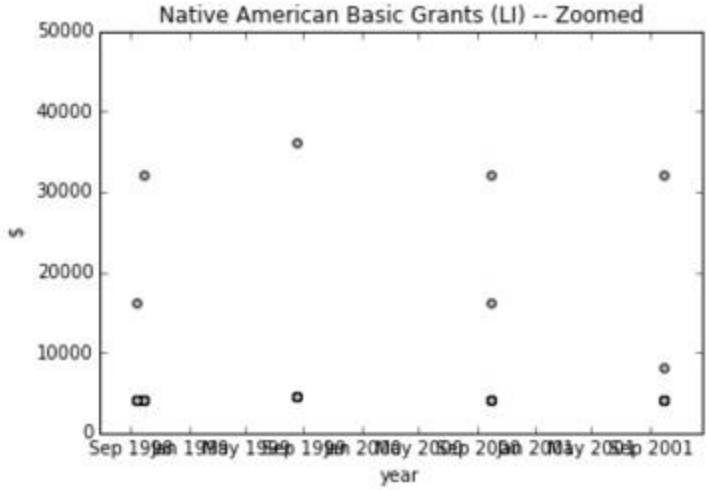


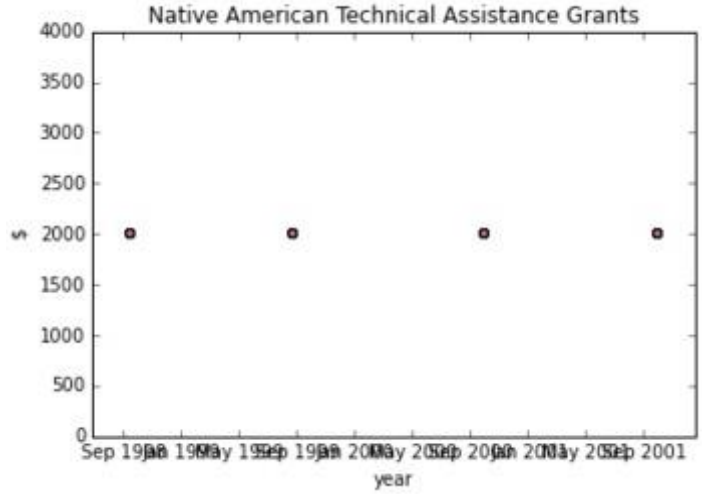

National Leadership Grants -- Full
(LG, LL, MG, NC, ND, NE, NL, NO, NP, NR)
1e7
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
$
1999 2001 2003 2005 2007 2009 2011 2013 2015
year

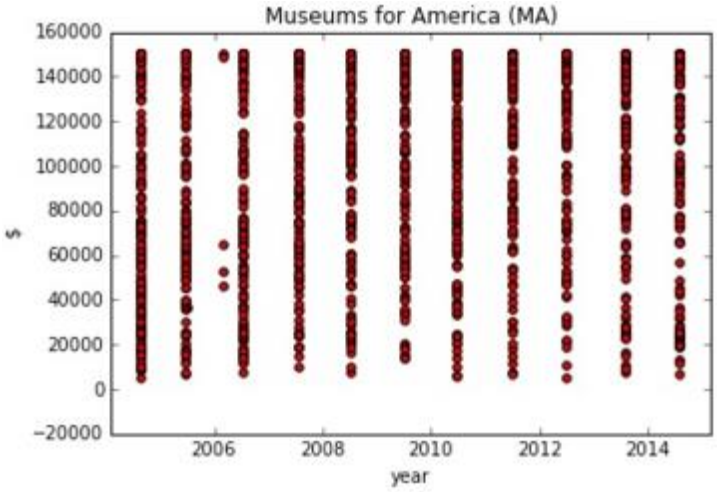

National Leadership Grants -- Zoomed
(LG, LL, MG, NC, ND, NE, NL, NO, NP, NR)
National Leadership Grants (LG)
National Leadership Grants (LL)
National Leadership Grants - Museums
NLG/Museums in Community
NLG/Preserv or Digitization
NLG/Education and Training
NLG/Museum Library Collab
NLG/Technology & Mus. On Line
NLG/Professional Services
NLG/Research and Demo
1000000
800000
600000
400000
200000
0
$
1998 2000 2002 2004 2006 2008 2010 2012 2014
year

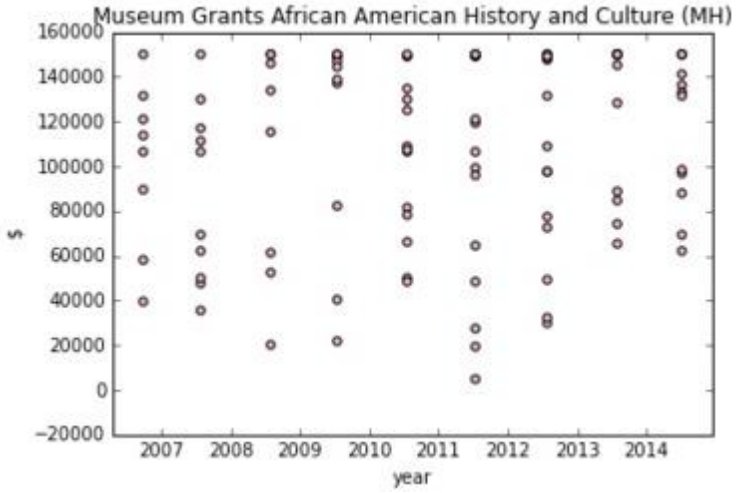

Native American Basic Grants (LI) -- Full
400000
350000
300000
250000
200000
150000
100000
50000
0
−50000
$
Sep 2001
year

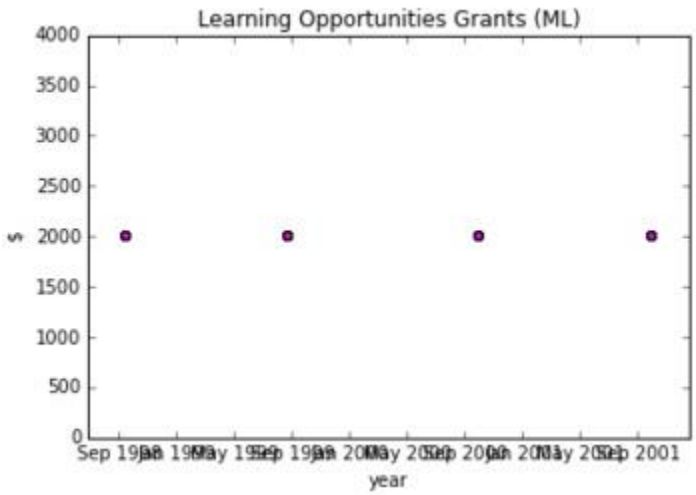
```
(0, 50000)
```

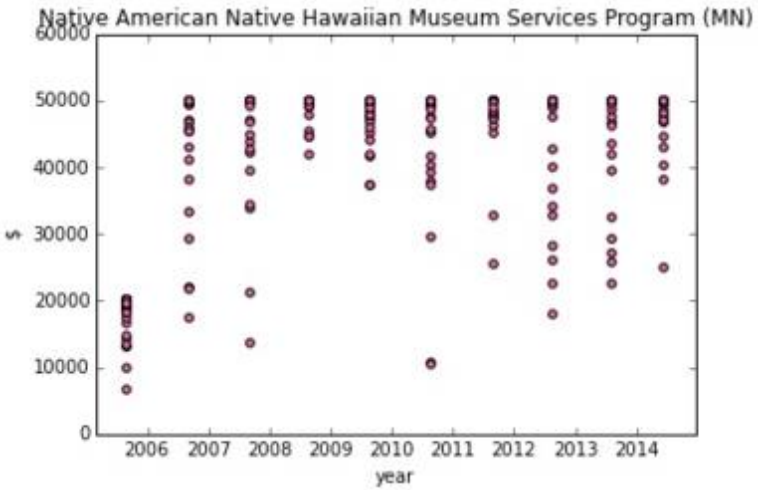

Native American Basic Grants (LI) -- Zoomed
50000
40000
30000
20000
10000
0
$
Sep 2001
year

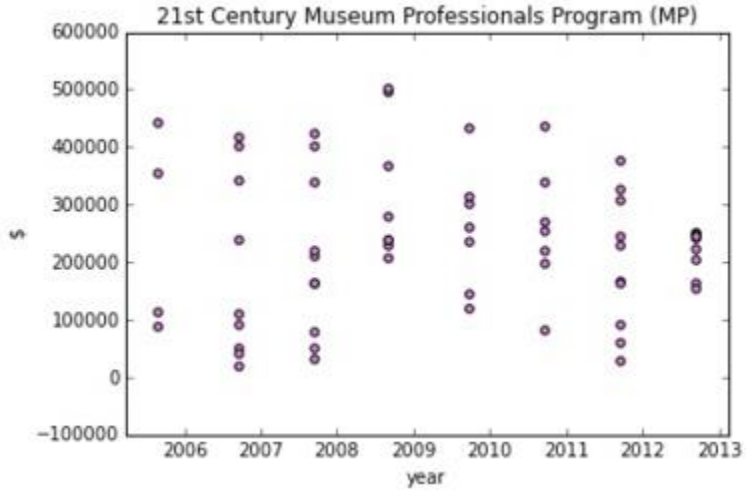

Native American Technical Assistance Grants
4000
3500
3000
2500
2000
1500
1000
500
0
$
Sep 2001
year

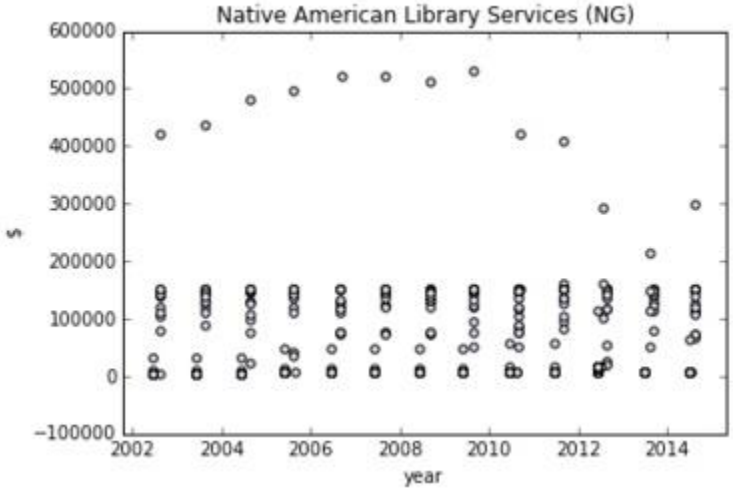

Museums for America (MA)
160000
140000
120000
100000
80000
60000
40000
20000
0
−20000
$
2006
2008
2010
2012
2014
year

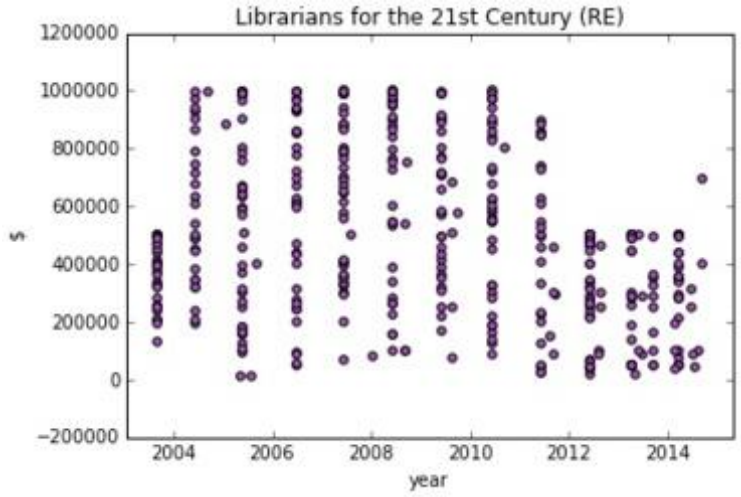

Museum Grants African American History and Culture (MH)
160000
140000
120000
100000
80000
60000
40000
20000
0
−20000
$
2007
2008
2009
2010
2011
2012
2013
2014
year

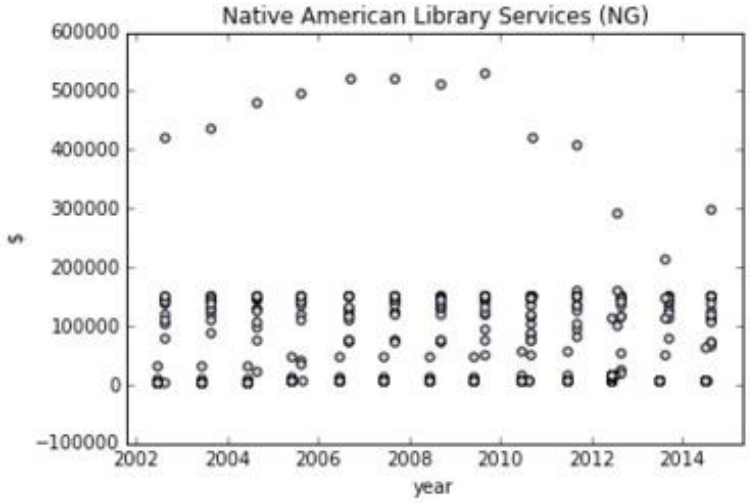

Learning Opportunities Grants (ML)
4000
3500
3000
2500
2000
1500
1000
500
0
$
Sep 2001
year

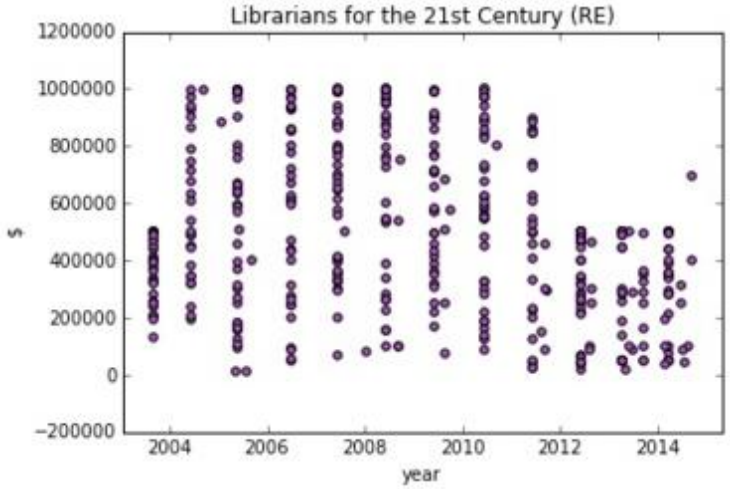

Native American Native Hawaiian Museum Services Program (MN)
60000
50000
40000
30000
20000
10000
0
$
2006
2007
2008
2009
2010
2011
2012
2013
2014
year

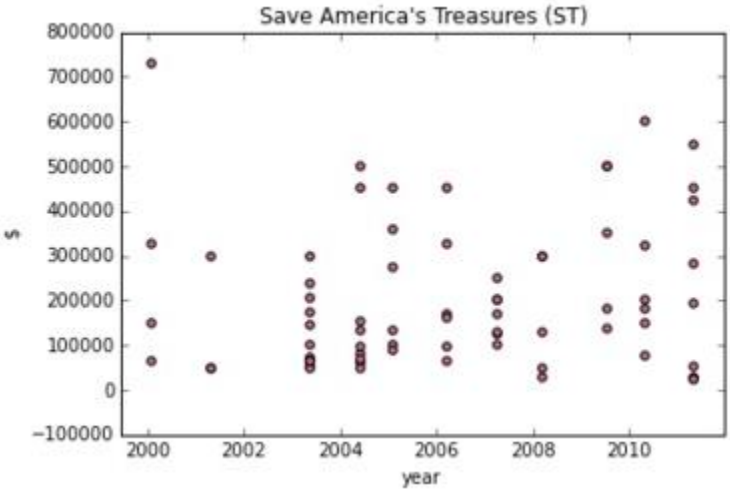

21st Century Museum Professionals Program (MP)
600000
500000
400000
300000
200000
100000
0
−100000
$
2006
2007
2008
2009
2010
2011
2012
2013
year


Native American Library Services (NG)
600000
500000
400000
300000
200000
100000
0
−100000
$
2002
2004
2006
2008
2010
2012
2014
year


Librarians for the 21st Century (RE)
1200000
1000000
800000
600000
400000
200000
0
−200000
$
2004
2006
2008
2010
2012
2014
year


Native American Library Services (NG)
600000
500000
400000
300000
200000
100000
0
−100000
$
2002
2004
2006
2008
2010
2012
2014
year





The data's "stacked" appearance is probably due to IMLS program's application periods and subsequent award dates. There also appears to be a range of funding amounts each award period, with some "grand prize" amounts for some of the programs ("Native American Library Services (NG)," "Native American Enhancement Grants (NE)," for example). Some programs grant awards sparsely, showing only 4 award dates with similar amounts. For these programs, using the "Award Period" start and end dates may have been the more demonstrative features.

2. Funding Amounts for Keywords

This exercise is a method to find keywords and plots funding amounts over time for those selected keywords.

The stop_words module was used in the "Project Title" column. This feature describes the name of the project awarded discretionary funds. I split each "title" string and eliminated stop words provided by the module, returning a list in the string's place. The result is a list of lists.

The list of lists was appended to the dataset as "titleList." Additionaly, titleList is counted then sorted by counts.

The goal of having keywords is to identify trends in subject matter that does not duplicate the plotting by program type. Not surprisingly, words like "museum," "library," and "program" appear at the top and add little value. Although these and other words like "using," "center," "national," "American," "access," and "strengthening" are not stop words, they also add little value but occur frequently. I also discovered that, for instance, "Art" might have 50 counts, but combined with "art," "arts," or "Arts," I might find 250 records for projects about art.

There were several missed opportunities as well. For one, word pairs could have identified a phrase like "national treasure" which is more specific than "national museum," or "Native American" and "African American" rather than just "American." I probably should have cleaned the lists up before sorting them by making all items lowercase and taking out punctuation marks. I also thought about taking words that occur in the program name out of the titleList for each record.

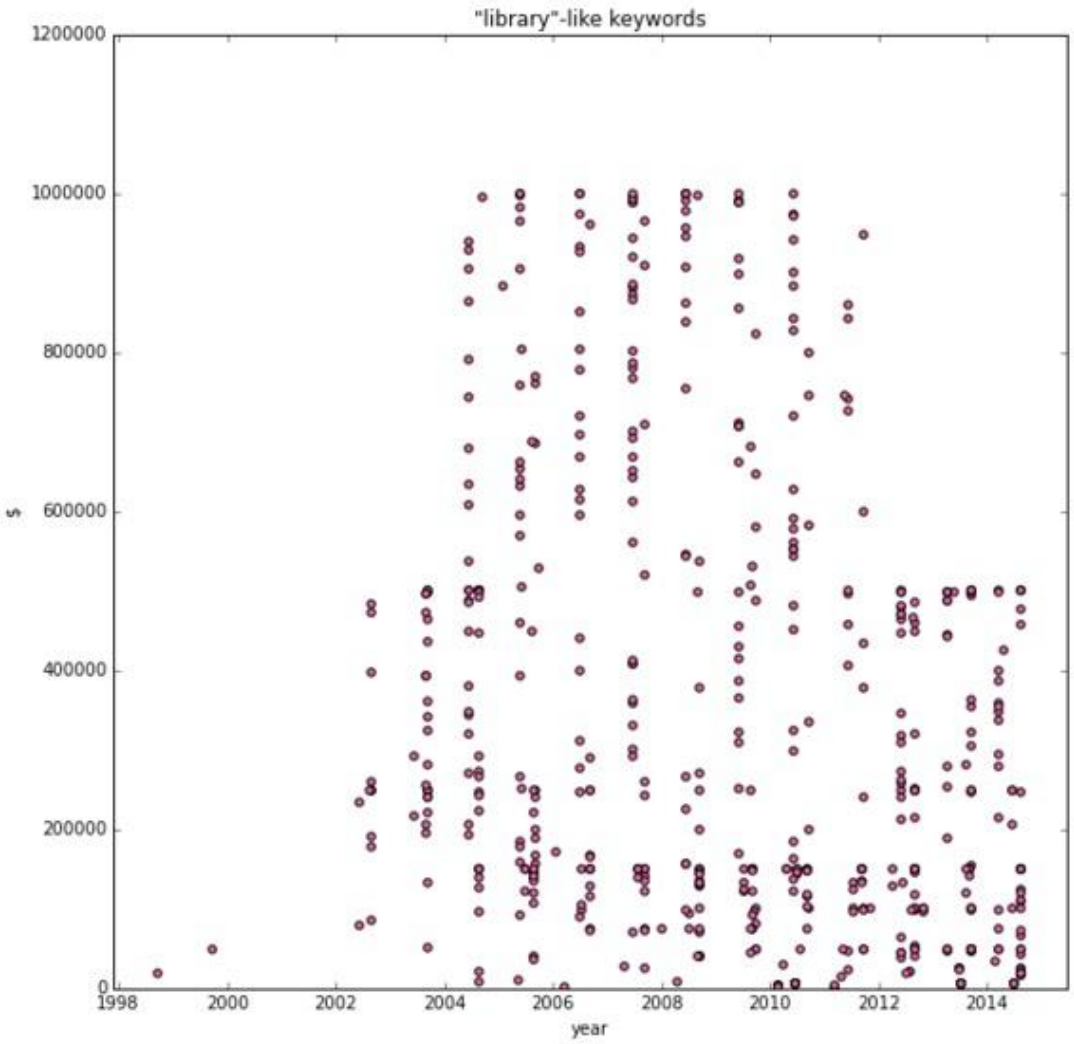
I had done the same for the "Description" features but found that there were many empty records. So although a description is longer and contains maybe a wider range of words to choose from than the "Title" feature, I chose to exclude the project description feature in order to be able to include more records. Every record contains a project title, so none are excluded through this criteria.

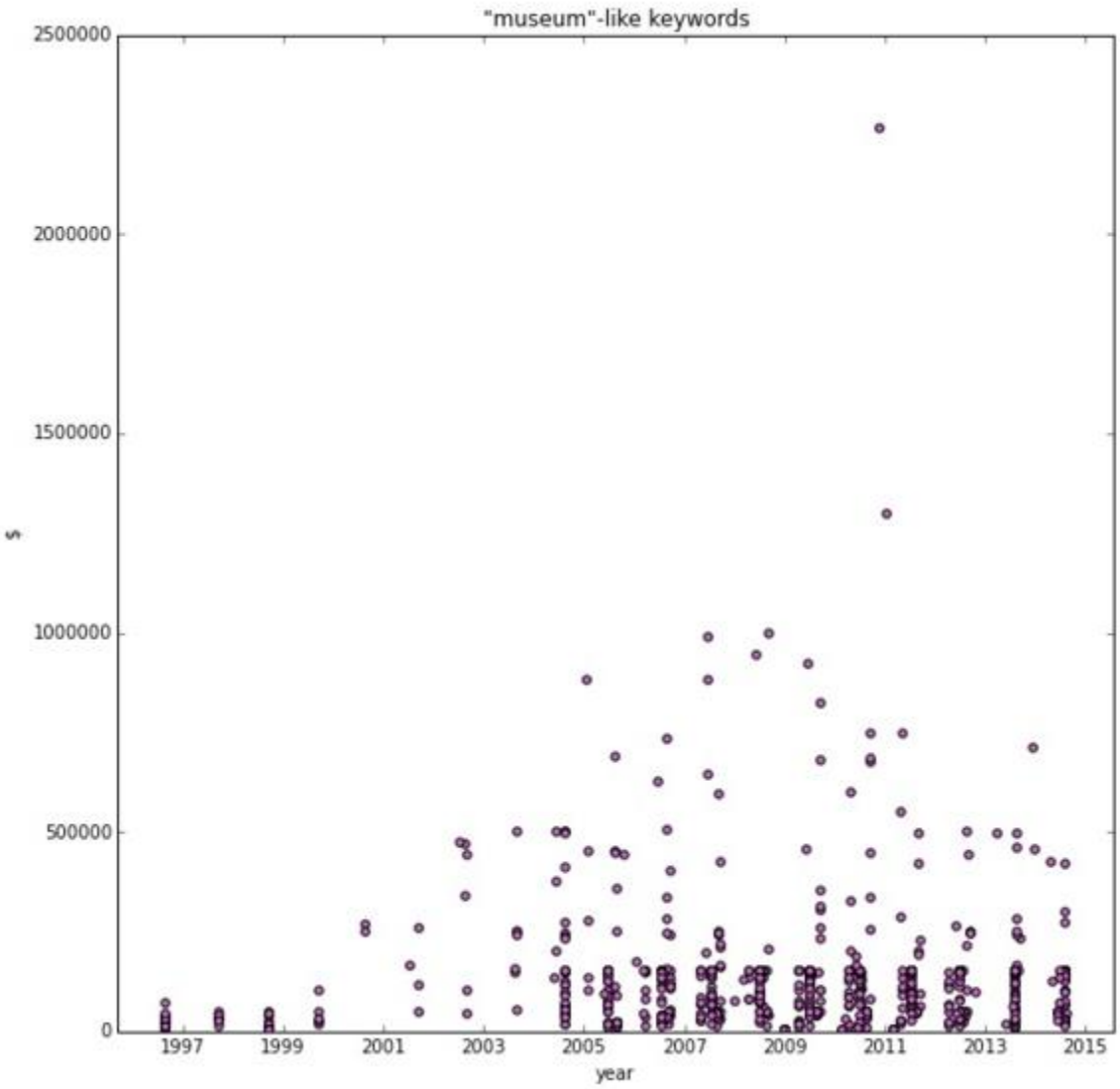
"Museum" and "Library"-like keywords. Next, I created a dummy list for "museum"-words. Because "museum" may appear as "museum," "Museum," I searched for "useum" and created a new list that returns a 1 where such a word appears. For "library," I used "ibrar."

Similar to the above Project Type exercise, I turned award amounts into floats without a dollar-sign, and turned the strings in "Award Date" into datetime objects.

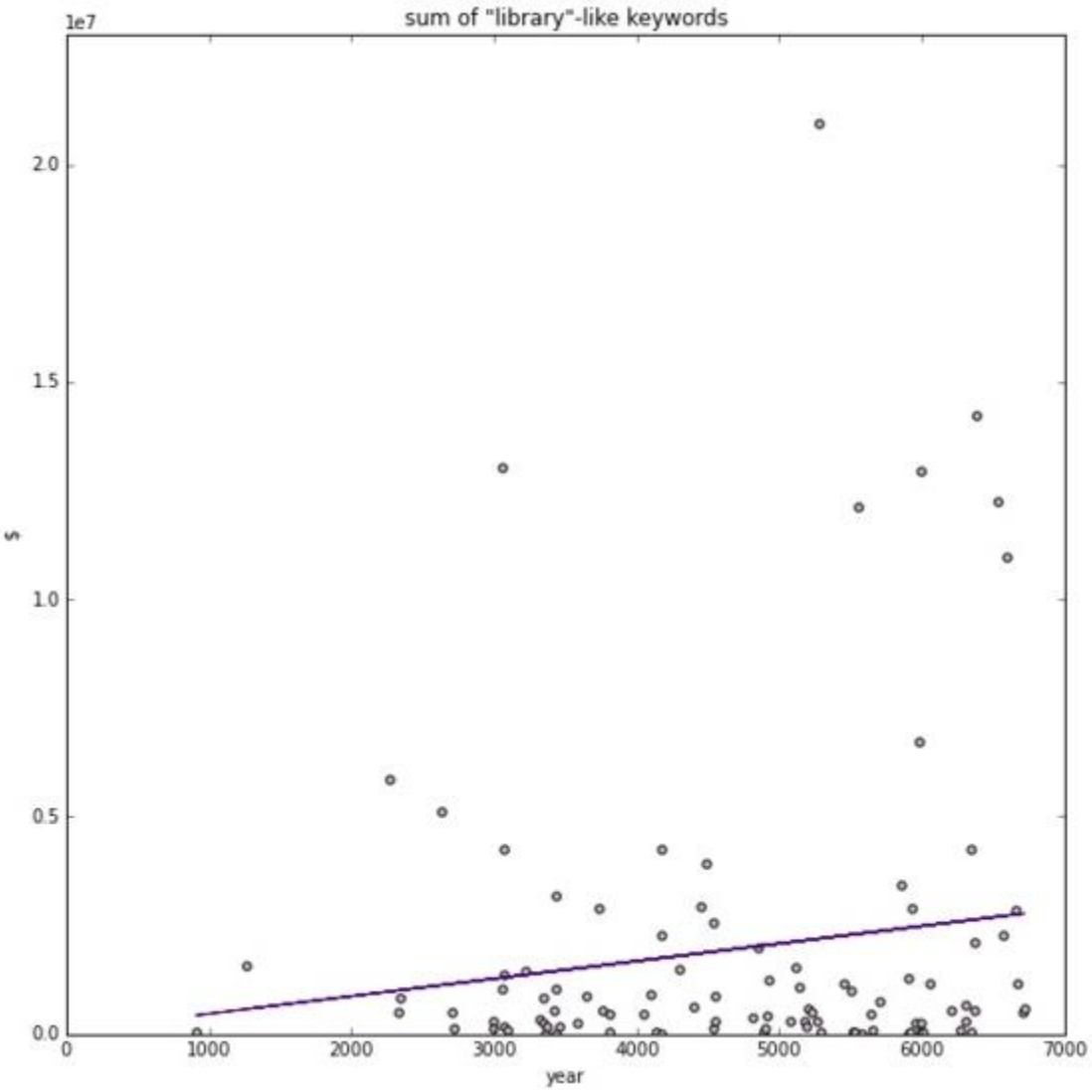
I also added another column that lists the date-time object as days as integers because of some issues I ran into further down the road. This ranges from 0 (first date in 1996) to 6757 (last date in 2014).

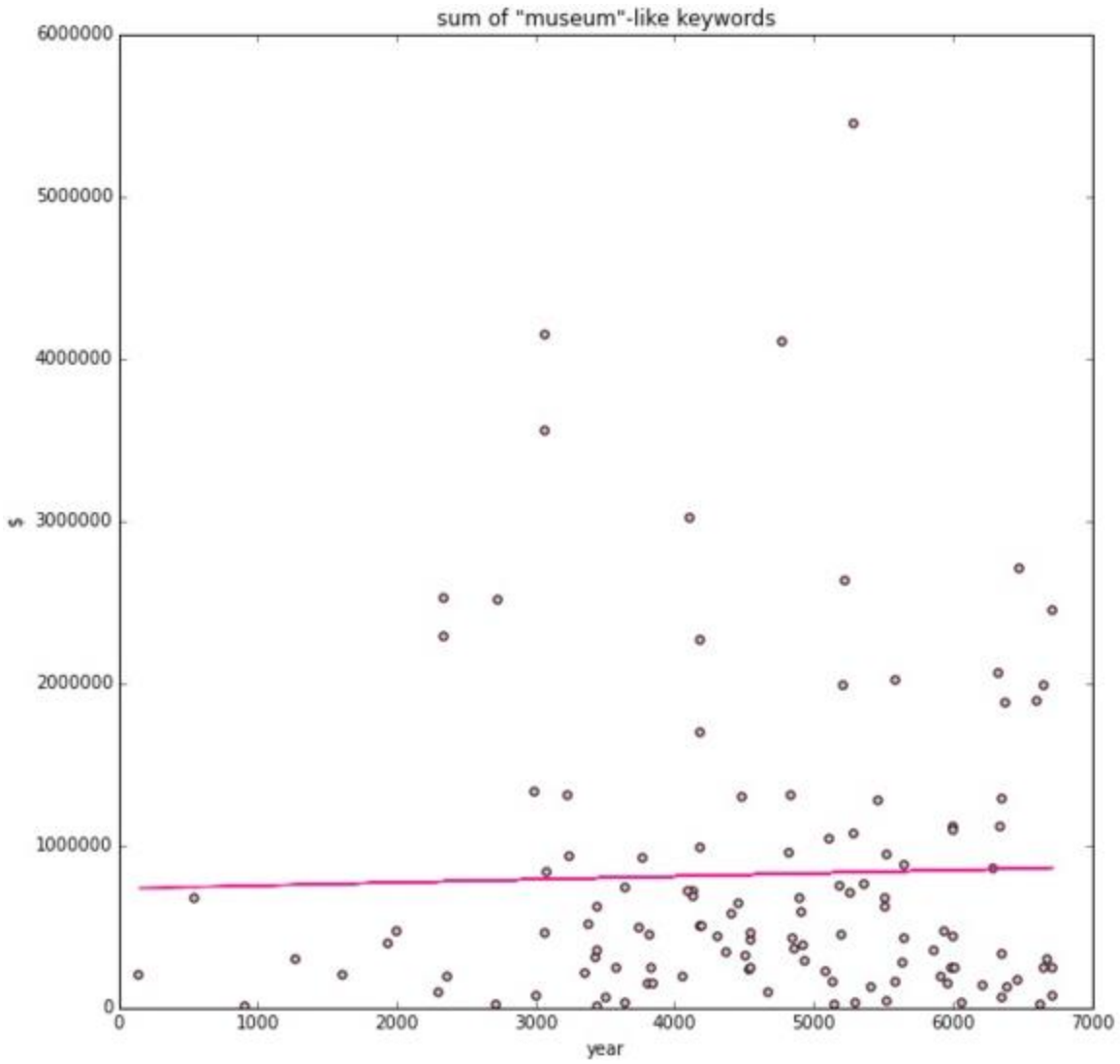
Using the dummy columns for “museum” and “library”-like keywords, I plotted award amounts over time.





I then summed award amounts over time and created a linear regression model for both "museum" and "library." This exercise was inspired by the annual budget which allocates a $150 million to Grants to States through the Library Services Technology Act (LSTA), versus the $28 million through the Museum Services Act. However, it is not known whether those grants to states are included in this dataset, and if they are, under which program, so no such conclusion could be drawn at this time.

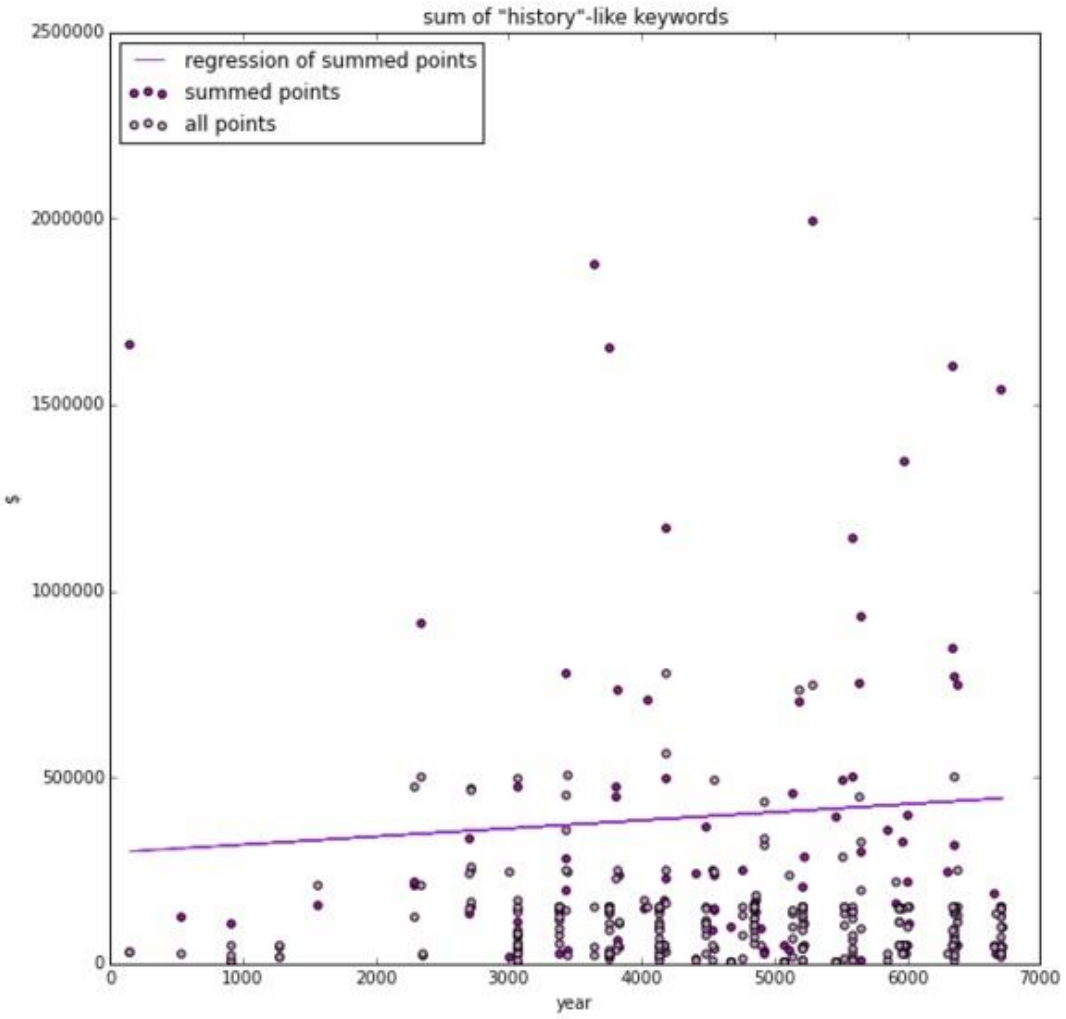

sum of "library"-like keywords
1e7
2.0
1.5
1.0
0.5
0.0
$
0
1000
2000
3000
4000
5000
6000
7000
year



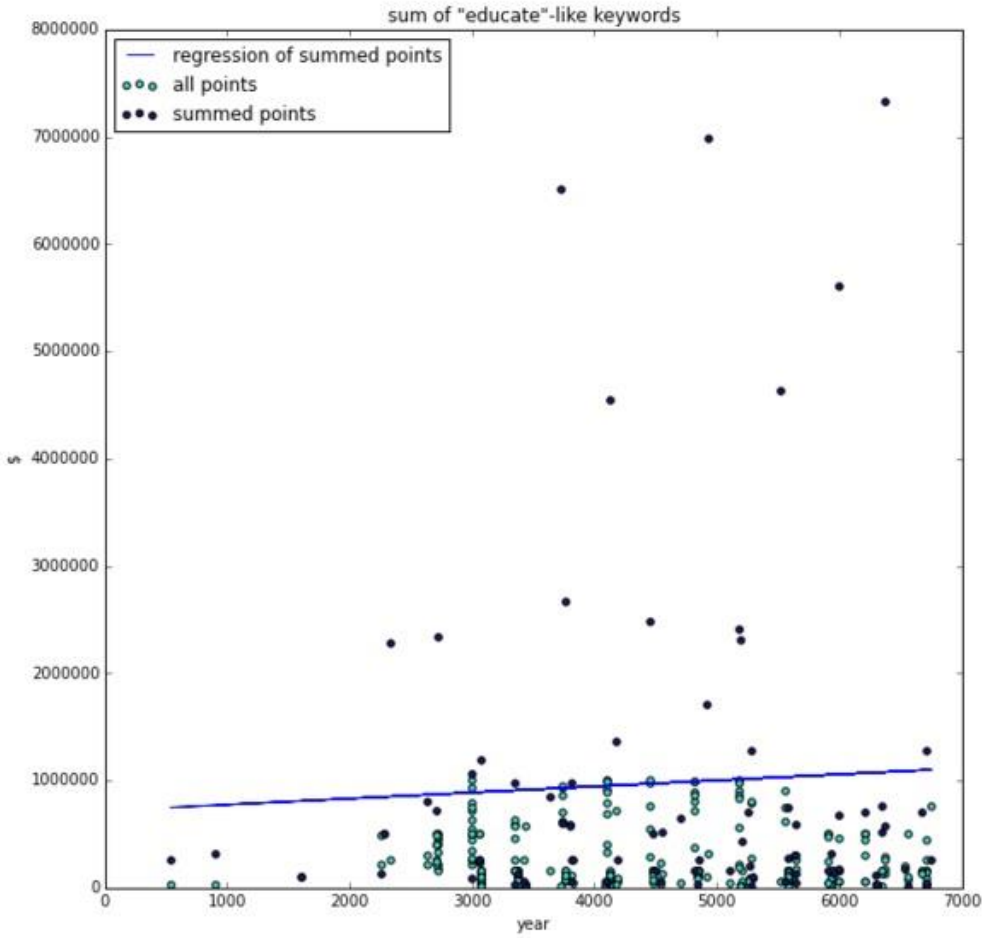
What was interesting is that the first time I ran the regression model for summed “museum”-like keywords, its coefficient was -30, but after running it a few more times while testing my options, it settled on +19.

Other Keywords. I did the same for words that begin with "histor-," "educat-," "data," and "scien-."

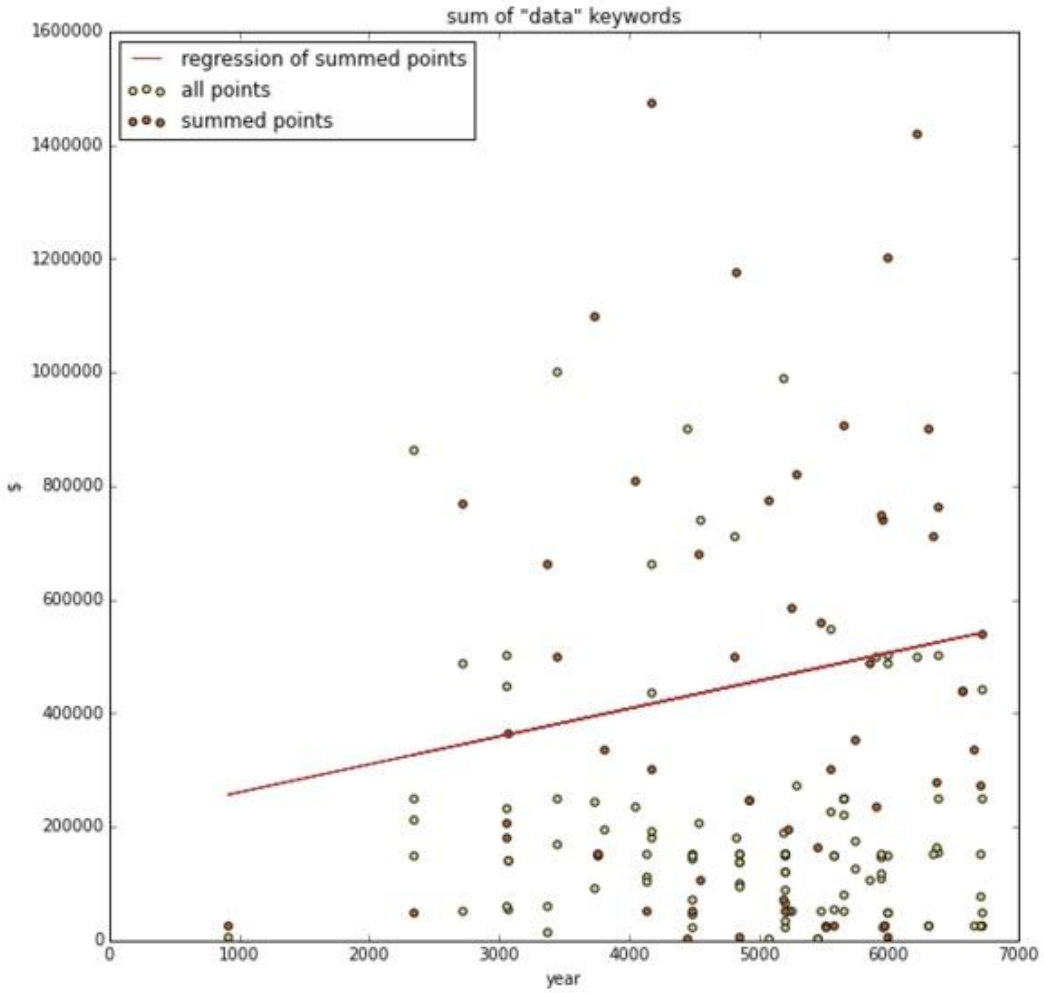
```
History-sum coefficient [ 21.72696447]
History-sum intercept 297807.420492

(0, 2500000.0)
```



```
educate-sum coefficient [ 56.58696295]
educate-sum intercept 718043.399321

(0, 8000000.0)
```

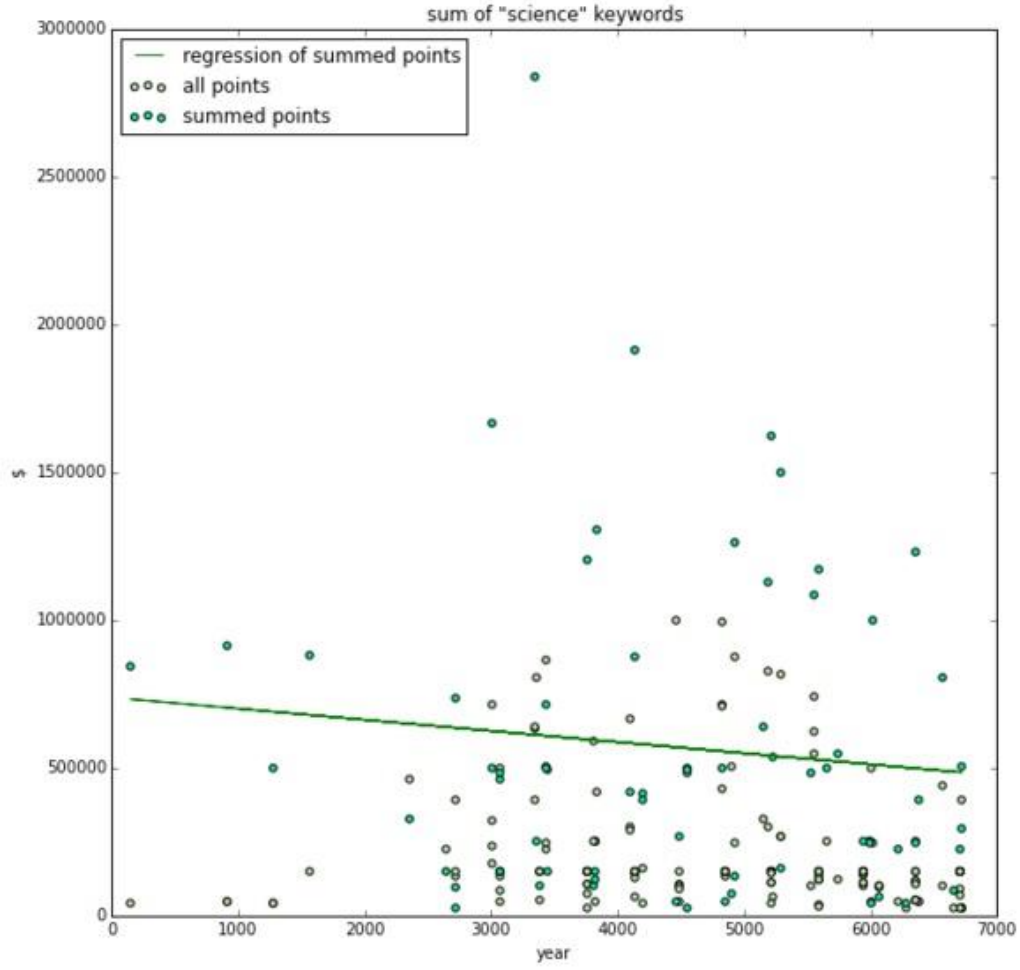


```
data-sum coefficient [ 49.05656834]
data-sum intercept 211644.578635
2015 predicted spending [ 559946.21383885]
2020 predicted spending [ 653153.69368225]

(0, 1600000.0)
```



Starting from “data”-like keywords, I added in two calculations: an approximate spending amount for a day in 2015, and in 2020. Because I have not used any other features besides keywords however, these calculated values are hardly predicted values.

```
data-sum coefficient [-37.68590287]
data-sum intercept 737739.932652
2015 predicted spending [ 470170.02230563]
2020 predicted spending [ 398566.80686083]

(0, 3000000.0)
```



It was interesting to note that total funding for "science," based on the key-word search, seems to have decreased over time. This observation leads me to believe that for some reason or another, "science" became less of a priority for the IMLS. This could be for funding constraints, an administrative wish to focus on issues set by its policy, or perhaps another agency or organization, such as the National Science Foundation, took on priorities that would overlap with IMLS's.

Conclusion

Challenges. My biggest challenge was probably the source of all my challenges, and that was my unfamiliarity with Python. I probably spent more time getting familiar with the data and programming concepts than I did running actual analysis.

Much of the data I wanted to use was stored as strings—descriptions, dollar amounts, ZIP codes, dates. In order to use these, I realized (maybe too late) that I needed to convert these strings into other objects. Working with the data, then, usually meant trying to find an the right type of object that would fit the code's input args.

For instance, ZIP codes were stored as string, but could be five characters or ten characters (XXXXX-XXXX). I thought it would be easy enough to return each one as just the first five and convert that string object into an integer or float, but I ran into errors and ran out of time.

Success. On the flipside, however, overcoming these challenges in regards to making the data work with the problem I was trying to solve was my biggest success. It bred a level of comfort with the problem solving and exploring possible solutions.

If compared to the Data Science course lessons, however, I probably only used the skills learned from the first two classes. Therefore the outcome of this project probably produced more missed opportunities than successes.

Missed Opportunities. I ran out of time (and patience), unfortunately, and missed some opportunities. For one, I would have liked to actually try to train/fit and predict a model with multiple features, especially with ZIP Code and Organization Type. The problems I had with these two were massaging the data into features I could run regression-type problems on.

Even with what I had accomplished, I think I could have turned much of these processes into functions through which I could run many, many more keywords.

I think if I had managed to turn many of the string-features into float-features, I could have tried to predict something. And to this point, I lost any subsequent opportunities.

Next Steps. Knowing what I know now about the dataset, about the features, and with a greater familiarity with the Python language, the next step would be to explore these missed opportunities, especially with more features than just date and award amount. I think the first step before anything, however, would be to rebuild this dataset as float-objects, and clean up much of the processes made and coding syntax used.

Appendix A

IMLS Appropriations History, 2013–2016

(Budget Authority in 000's)

	FY 2013	FY 2014	FY 2015	FY 2016 Request
Museum and Library Services Act (MLSA) General Provisions				
Research, Evaluation, and Data Collection	**1,787**	**1,820**	**1,820**	**1,787**
Library Services Technology Act (LSTA)				
Grants to States	150,000	154,848	154,848	154,500
Native American Library Services	3,667	3,861	3,861	4,063
National Leadership: Libraries	11,377	12,200	12,200	17,500
Laura Bush 21st Century Librarian	10,000	10,000	10,000	10,500
SUBTOTAL, LSTA	**175,044**	**180,909**	**180,909**	**186,563**
Museum Services Act (MSA)				
Museums for America	19,564	20,200	20,200	21,457
Native American/Hawaiian Museum Services	877	924	924	972
National Leadership: Museums	7,468	7,600	7,600	11,168
SUBTOTAL, MSA	**27,909**	**28,724**	**28,724**	**33,597**
African American History Culture Act				
Museum Grants for African American History and Culture	**1,336**	**1,407**	**1,407**	**1,481**
Agency Wide				
Administration	13,745	14,000	15,000	14,000
TOTAL	**219,821**	**226,860**	**227,860**	**237,428**

Appendix B

Data Documentation for Institute of Museum and Library Services Discretionary Grant Awardees FY1996- FY2013

Dataset Description

This dataset of administrative records contain discretionary grant recipients who were awarded funds by the Institute of Museum and Library Services[1] from fiscal year 1996 through fiscal year 2013. Included in this dataset are: the name of the institution, the name of the grant project, the amount of the IMLS grant award, the amount of matching funds the applicant provided and the start and end dates of the grant period. Geographic information, such as the latitude and longitude of the grantees' street addresses, the census tract and block group the grantee institution is located in, and the state and county of the grantee, is also included in the dataset.

Description of Data Fields

LogNumber – Unique identifier for grant application

Institution – Name of organization that applied for grant

Program – Name of IMLS grant program

ProgramType – Two character code that identifies program

Type code	Program name
IA	Conservation Assessment Program
IC	Conservation Program
IG	General Operating Support
IL	Museum Leadership Initiatives
IM	Museum Assessment Program
IS	Professional Practices
LE	Native American Enhancement Grants
LG	National Leadership Grants[2]
LI	Native American Basic Grants
LT	Native American Technical Assistance Grants
MA	Museums for America
MH	Museum Grants African American History and Culture
ML	Learning Opportunities Grants
MN	Native American Native Hawaiian Museum Services Program
MP	21st Century Museum Professionals Program
NG	Native American Library Services
RE	Librarians for the 21st Century
ST	Save America's Treasures[3]

ProjectTitle – Title of awarded project

[1] The Museum and Library Services Act (MLSA) of 1996 established the Institute of Museum and Library Services within the National Foundation on the Arts and Humanities. The new agency combined the Institute of Museum Services, which had been in existence since 1976, and the Library Programs Office, which had been part of the Department of Education since 1956.

[2] This code encompasses all National Leadership Grants including those with log numbers beginning with LG, LL, MG, NC, ND, NE, NL, NO, NP, and NR.

[3] Support for the Save America's Treasures program was provided by the United States Department of the Interior/National Park Service's Historic Preservation Fund.

ProjectType – Sub-categorization of grant program

AwardDate – Date grant was awarded

InstAddr1 – Street address of institution – 1st field

InstAddr2 - Street address of institution – 2nd field

InstAddr3 - Street address of institution – 3rd field

InstCity – Address of institution - City

InstState – Address of institution - State

InstZip - Address of institution – Zip Code

AwardOrig – Total amount originally awarded

AwardTotal – Total amount awarded dispersed

AppMatchAmount – Applicants matching amount

ReqAwardAmount – Requested award amount

ReqMatchAmount – Requested matching amount

Prefix – Title of Principal Investigator (PI)

FirstName – First name of Principal Investigator (PI)

LastName - Last name of Principal Investigator (PI)

OrgType - Type of organization of the applicant institution

FiscalYear - Fiscal year of application

AwardPeriodFr – Award period start date

AwardPeriodTo - Award period end date

OrgUnit – Name of organizational unit

OUAddr1 – Street address of organizational unit – 1st field

OUAddr2 – Street address of organizational unit – 2nd field

OUAddr3 – Street address of organizational unit – 3rd field

OUCity - Address of organizational unit - City

OUState - Address of organizational unit – State

OUZip- Address of organizational unit – Zip

Latitude – latitude of institution's address; this field consists of two integers and six decimal places, with an explicit decimal point, decimal degree format, World Geodetic System Datum 1984; based on geocoded address of institution

Longitude – longitude of institution's address; this field consists of a negative sign, three integers and six decimal places, with an explicit decimal point, decimal degree format, World Geodetic System Datum 1984; based on geocoded address of institution

FIPSState - Federal Information Processing Standards (FIPS) State Code, US 2010 Census; based on geocoded address of institution

FIPSCounty - Federal Information Processing Standards (FIPS) County Code, US 2010 Census; based on geocoded address of institution

CensusTract – US 2010 Census Tract; based on geocoded address of institution

CensusBlock – US 2010 Census Block; based on geocoded address of institution

FIPSMCDCode - Federal Information Processing Standards (FIPS) Minor Civil Division (MCD) code, US 2010 Census; based on geocoded address of institution

FIPSPlaceCode –Federal Information Processing Standards (FIPS) Place code, US 2010 Census; based on geocoded address of institution

CBSACode – Core Based Statistical Area code, 2013 US Office of Management and Budget CBSA; based on geocoded address of institution

MetroDivisionCode – Metropolitan Statistical Area Division code, 2013 US Office of Management and Budget; based on geocoded address of institution

Description – Brief description of awarded grants provided by the IMLS Office of Communications and Government Affairs (OCGA) for fiscal years 1997 to 2013; descriptions are not available for all awarded grants.

Institute of Museum and Library Services

The Institute of Museum and Library Services is the primary source of federal support for the nation's 123,000 libraries and 17,500 museums. Through grant making, policy development, and research, we help communities and individuals thrive through broad public access to knowledge, cultural heritage, and lifelong learning. To learn more about IMLS, please visit www.imls.gov.

Contact Information

Institute of Museum and Library Services
1800 M Street NW, 9th Floor
Washington, DC 20036-5802
202-653-IMLS (4657)
www.imls.gov

Please direct questions to OPRE@imls.gov.