Extracting Insights from FBO.Gov data - Part 2

Earlier this year, Sunlight foundation filed a lawsuit under the Freedom of Information Act. The lawsuit requested solication and award notices from FBO.gov. In November, Sunlight received over a decade's worth of information and posted the information on-line for public downloading. I want to say a big thanks to Ginger McCall and Kaitlin Devine for the work that went into making this data available.

In the first part of this series, we looked at the data and munging the data into a workable set. Once I had the data in a workable set, I created some heatmap charts of the data looking at agencies and who they awarded contracts to. Those didn't really work out all that well. Actually they sucked. They didn't represent the data well at all. The data was too sparsely related so I was forced to show more data than really made sense. In part two of this post, we will create some bubble charts looking at awards by Agency and also the most popular Awardees.

The Data
For this analysis, we are going to use all of the data going back to 2000. We have six data files that we will join together, filter on the 'Notice Type' field and then calculate the sum and count values by the awarding agency and by the awardee.

Pig Script
Again, I created a Mortar project to process the data. The pig script loads the six tab delimited files, filters the files by '(noticeType == 'Award Notice' AND contractAwardAmount IS NOT NULL)', does a union and then calculates the count and sum values. I used a small piece of Python code to clean up the award amount field.

The results are written out to a delimited file. Looking at the results, there were 465 agencies that awarded a total of 428,937 contracts. Again, this is for all of the available data (2000-present). For awardees, there were over 192,000 unique names receiving awards.

Visualization
For this visualization, I wanted to try a bubble chart. The bubble chart will allow me to visualize the number of awards, the size of awards and a relative size based on the number of awards.

For the bubble chart, I am using HighCharts, a javascript library for creating charts. Since this is a personal, non-commercial website, there is no charge to use the product.

Here is a sample of the results for agency awards:

Agency	Number of Awards	Sum of Award Amount
DLA Acq		

Naval S
Comma
Army Co
Air Forc
U.S. Arr

And here's a sample of the results for awardees:

Agency

KAMPI
INC.-7Z
COMPO
CANAL
PA 1903
OSHKO
CORPO
OSHKO
2307 OR
WI 5490
PIONEE
INC.-66
INDUST
MARINE
11735-5
BELL BO
PROJEC
TX 7912

Let's look at a simple bar chart of the Agency Awards:

Now, let's look at the bubble chart of the same data:

Here is a simple bar chart of award recipients:

And a bubble chart showing awards by recipients:

Analysis

There really isn't a lot of surprise in which agencies are awarding contracts. 8 of the top 10 are in the defense industry. Similarily the top recipient of contracts was Kampi Components, which supplies factory replacement spare parts to the United States military. Number two was Oshkosh Corporation, who are manufacturers of severe heavy duty all wheel drive defense or military trucks, aircraft or emergency rescue and firefighting. The one outlier is NIH which didn't award a large number of contracts but awarded large contracts.

Next steps will be to look at the types of awards and see how to display those in a meaningful fashion.