CS 419 Term Project: 2014 NASA Budget

Michael Woffendin, John Zeller

**Abstract**—In this project, we investigated various methods of visualizing 2014 contract budgeting information provided by NASA. Our approach was guided by a particular scenario: informing a congressperson about NASA’s impact on their constituents. The visualization methods were a heatmap, scatterplot, and bar chart.

**Index Terms**—NASA, 2014 budget, contract, information visualization, heatmap, bar chart, scatterplot

# Introduction

NASA funding is set on a yearly cycle. The first step is the President’s Budget proposal, which Congress approves before coming law. Almost 80% of NASA funding is paid out in contracts to companies, offering some product or service to NASA. Naturally, these contracts are not likely to be evenly distributed among the United States, but with upwards of 36,000 contracts and more than $14 billion annually, it can be hard to tell where all that money is going.

# Domain Situation

The target user is an economic advisor for a member of Congress, whose primary duty is to provide data on how funding/policy changes may impact a congressperson’s state/district. The advisor will have to provide an overview of what economic impacts NASA’s 2014 expenditures had on national, state, or local economies. The congressperson will use this overview of the previous year to decide how to vote going forward.

There are several relevant questions that would help reveal any impact:

1. How much money do state/district economies receive from NASA contracts?
2. How much money do companies receive from NASA contracts?
3. What is the project duration for NASA contracts?
4. Do NASA contracts favor a particular size of company?

# Task Identification

Table 1 outlines the tasks associated with each of the aforementioned questions.

|  |  |  |
| --- | --- | --- |
| Question | Task | Target |
| Q1: Locate areas with NASA spending | Analyze: Discover  Search: Browse  Query: Summarize | High: Feature  Low: Value of contracts |
| Q2: Look for patterns company contract values | Analyze: Discover  Search: Browse  Query: Summarize | High: Distribution  Low: Value of contracts |
| Q3: Look for patterns in contract durations | Analyze: Discover  Search: Browse  Query: Summarize | High: Distribution  Low: Contract durations |
| Q4: Look for correlation between company size and contracts awarded | Analyze: Discover  Search: Browse  Query: Identify | High: Correlation  Low: Contract dollars awarded |

Table 1. Task identification table.

## Detailed Task Descriptions

1. At the high level, the user is attempting to identify regions with high numbers of NASA contracts. At the low level, the user must browse through summarized regions to find those that have a high number of contracts.
2. At the high level, the user is attempting to characterize the distribution of contract budgets. At the low level, the user must browse through summarized contract budgets to identify patterns.
3. At the high level, the user is attempting to characterize the distribution of contract durations. At the low level, the user must browse through a summarization of contract durations to identify patterns.
4. At the high level, the user is looking for correlations between company size and contracts awarded. At the low level, the user must browse through correlation values for those in the high range.

# Tables

Four tables are necessary to successfully answer the tasks outlined earlier: contract, company, state, and district.

## Contract Table

The raw data for this project consists of contracts (items) and their attributes. These attributes include company (cat), congressional district (cat), state (cat), annual budget (quant), number of employees (quant), start date (quant), and completion date (quant).

## Company Table

For Q4 we need to know a company’s size to correlate it with awarded contracts. This table can be derived from the contract table when applicable, or from online sources as necessary. The items in this table are companies. The attributes are company name (cat), company size (quant), and awarded contracts (quant)

## State Table

For Q1 we need to know how many contracts are awarded in a given state. This table can be derived from the contracts table. The items in this table are states. The attributes are the state name (cat) and number of contracts (quant).

## District Table

For Q1 we need to know how many contracts are awarded in a given district. This table can be derived from the contracts table. The items in this table are districts. The attributes are the district id (cat) and number of contracts (quant).

# Encodings

The following subsections outline the data encodings chosen to answer each of the aforementioned questions

## Total Value of NASA Contracts by State/District

* Chart: Heatmap by State/District
* District (C): spatial/shape
* Total Contracts Value (Q): color saturation

## Total Value of NASA Contracts by Company

* Chart: Bar
* Company (C): spatial
* Total Contracts Value (Q): length

## Duration of NASA Contracts by State/District

* Chart: Heatmap by State/District
* District (C): spatial/shape
* Duration (Q): color saturation

## Correlation of Company Size and NASA Contracts

* Chart: Scatter
* Company Size (Q): horizontal position
* Number of Contracts (Q): vertical position

# Technique / Justification

The following subsections outline the logic behind the charts chosen in section 5 above.

## Heatmap

Spatial/shape is the logical encoding for states and districts since they are effective encodings for categorical data and the target audience is already familiar with this encoding technique due to elections. Color saturation is the logical encoding for duration / number of contracts since all other quantitative encodings would distort the categorical encoding.

## Bar

Length is a good encoding for contract budgets since it is a quantitative value. Using spatial encoding to pair up bars with companies works since it is the best categorical encoding.

## Scatter

Since both data are quantitative, position along common axes is the best encoding for expressing them.

# Results

The heatmap turned out great. In addition to the proposed datasets of total dollar value and average duration, we also included total number of contracts as a dataset option. The user can also toggle the map layout between congressional district and state, which makes the data relevant to members of both the House and Senate. Toggling between these layouts and datasets is also easy and fairly quick. Finding outliers, features, and trends based on geography are simple tasks using the heatmap, which indicates it was likely the correct encoding choice. Since our color scheme is in the cool-temperature blue range, highlighting regions using warm colors makes them stand out really well. Currently this is only used on hover, but could easily be expanded to draw attention to other features. Overall, the heatmap proved an effective way to answer several questions.

The scatter plot turned out pretty well also. It’s pretty clear that NASA generally favors hiring many small companies over a few large companies. Finding outliers to this trend is also easy. The only major drawback to the scatter plot is that identifying individual data points within a cluster can prove difficult due to their close proximity. While we implemented the same highlighting on hover as in the heatmap, actually seeing the color change is difficult on the scatterplot due to each dot’s small size.

The bar chart was the least successful and most difficult to implement of the three charts. Due to the number of companies involved, not all data points could be visible at once; hence, some way to restrict the data was necessary. We decided that scrolling might be the best way to accomplish this. After implementing it though, we realized that locating an individual company within the dataset was difficult. In a future version, being able to re-sort the data on various attributes and including a detailed scroll preview might help alleviate this issue.

# Encountered Issues

Since our dataset was larger than most of our classmates’ (over 50,000 rows), we encountered some unique problems related to its size. The first problem had to do with the asynchronous nature of JavaScript; specifically, the heatmap was trying to render before the data finished loading from disk. Normally with D3 this issue wouldn’t occur due to rendering being a data binding callback; however in this particular instance we could not bind our heatmap data in typical D3 fashion due to limitations within our graphing library. To sidestep this issue, we serialized data loading and rendering by using a d3 queue construct.

The second issue we encountered was performance drop, particularly with regard to the heatmap. Switching between the different data sets took several seconds and made the browser temporarily nonresponsive, and the problem became progressively worse with the cumulative number of switches made. We discovered that the accumulation of performance drops was caused by certain chart elements being redrawn instead of replaced during each switch, leading to a memory leak. Addressing the memory leak eliminated unresponsiveness and increasing switch times, but switching data sets still takes a while due to the volume of data.

After viewing some of the data on the charts, we realized that interpreting large numbers is difficult. We therefore switched from using raw dollar values to formatting them with an appropriate unit indicating quantity. For example, $1042612.95 would become $1m, which is much easier to read. Since the use cases for this project are fairly high-level, the tradeoff of precision for legibility is acceptable.

We encountered an issue where all our charts were drawing on top of each other when we put them on the same view. We determined that this was due to a namespace conflict in the name of our SVG; therefore, simply changing the variable names resolved our issue.

# Future Features

While the project reflects all the features showcased in the design documents and demonstration, there are still several improvements that could benefit the user experience. The first is improving interactivity by linking the charts together. For the heatmap, clicking on a district would restrict contracts and companies displayed in the other charts to those that exist within the chosen district. Similarly, selecting a particular company from the bar chart for example would only display information related to that company on the heatmap. This feature would allow for congresspersons to quickly find information pertaining to their districts and supporters with the fewest clicks possible.

Another possible addition to the heatmap would be creating unique color encodings for negative contract values. Presently, very low values and negatives receive the same color encoding of near-white blue. Since the current encodings use cool colors, using a white to red scale would work well in contrast. Providing this new encoding would allow congresspersons to quickly see if NASA negatively impacted their constituents, and therefore allow them to make more rapid informed decisions. As an added bonus, the red, white, and blue color scheme would appeal to the target audience since it’s patriotic.

As a final addition to the heatmap, including state/district names on either the map or tooltips could be beneficial. While this is not a required feature, since the target audience will be able to visually identify their desired area without a label, it would make the map useful to a wider audience.