

Innovation Landscapes: An Analysis of National Science Foundation Research Grants from 1999-2016

Sahar Alhassan, Kaitlin Lynes, Deepa Mehta, Ilya Perepelitsa

December 21, 2015

Introduction

Innovation is often considered to be considered an engine for development. Governments typically support innovation by funding research and development for universities, corporations, and other institutions. (Etzkowitz and Leydesdorff 2000) Recently, the White House approved a bipartisan budget agreement which will allocate even more funding over the next years in order to maintain the United States' leadership in innovation. (Donavan 2015) One of the few agencies that this bipartisan agreement will support is the National Science Foundation (NSF).

NSF is an independent, federal grantmaking agency that awards nearly one fourth of all federal grants in the United States, nearly 11,000 new grants per year, and has a current annual budget of \$7.3 billion (NSF 2015). A majority of these grants support critical research in science and technology. In order to get a sense of how innovation research, as procured by the NSF, is distributed, one can visit the NSF website to locate some summary statistics about grant awards. However, the NSF website provides only basic information about the number of grants awarded and grant allocations by states and institutions. For any serious researcher who is interested in understanding the landscape of federal funding of innovation in the United States, one is met with an aggregate dataset that numbers in the millions of grants awarded since NSF's founding in 1950.

Given the sheer number of grants and the number of variables that denote grant type, this report began with a desire to develop an analytical methodology to understand the distribution of NSF grants and recognize any emergent patterns within these numbers. As such, this report studies NSF grants from the fiscal years 1999 to 2016 and aims to understand the distribution of NSF grants during this period. By developing a subset of the aggregate NSF data, we have been able to study the distribution of grant duration by type of grant (new grants, continuing grants, or revised grants), grants by recipient type (universities, private institutions or corporations, etc.), by state, and by types of funded research.

The broader aim of this exercise in big data analytics is to be able to apply quantitative methods to larger datasets, acquire subsets, and devise collated dataframes that can then be useful in identifying distributional attributes. The findings in this report and this dataframe may be useful to scholars, grantseeking organizations and individuals, as well as state agencies and private institutions to better understanding the landscape of scientific research funding in the United States.

Methodology

As a starting point, our dataset is comprised of queries obtained from the publicly available <https://www.usaspending.gov>. Using multiple queries on the followig page <https://www.usaspending.gov/Pages/TextView.aspx?data=HomeAwardTypeFunding> , the range of fiscal years 1999-2016 was obtained. The variables that were selected for the working dataframe are as follows:

Grant Periods

- fiscal_year - Budget year
- starting_date - Start date of awarded research project

- ending_date - End date of awarded research project
- obligation_action_date - Grant Decision
This variable tells us the date that the award was obligated.

Grant Amounts and Grant Type

- fed_funding_amount - Dollar amount of grant This is the total amount awarded for a particular grant.
- action_type - Type of Grant Award

This variable tells us whether this is a new grant, a continuing grant, or revised grant. A new grant refers to a new award which typically entail a “specific level of support for a specified period of time”, while continuing grants provide specified support for an initial period of time, “with a statement of intent to provide additional support of the project for additional periods, provided funds are available”, while revised grants refer to existing NSF grants that have been amended with new terms (NSF 2015).

Grant Recipient Information

- recipient_name - Recipient Name
This variable denotes the name of the recipient, usually a university or organization.
- recipient_type -Type of Recipient
This variable allows us to identify whether the recipient organization is a higher education institution, corporation, or non-profit, among other categories.
- cfda_program_title - Field of research This refers to the research discipline of a grant.

Grant Recipient Location

- recipient_city_name - Recipient's City
- principal_place_state - Recipient location state
- principal_place_state_code - Recipient location state code These variable allows us to identify the location of the recipient.

The following modifications were conducted to the existing variables: 1. conversion of non-numeric variable into categorical variables 2. dates were transformed to be treated as dates by R 3. duplicates of variable values were merged to be treated as identical values 4. cfda_programm_title values were replaced by more concise Continue, New and Revision

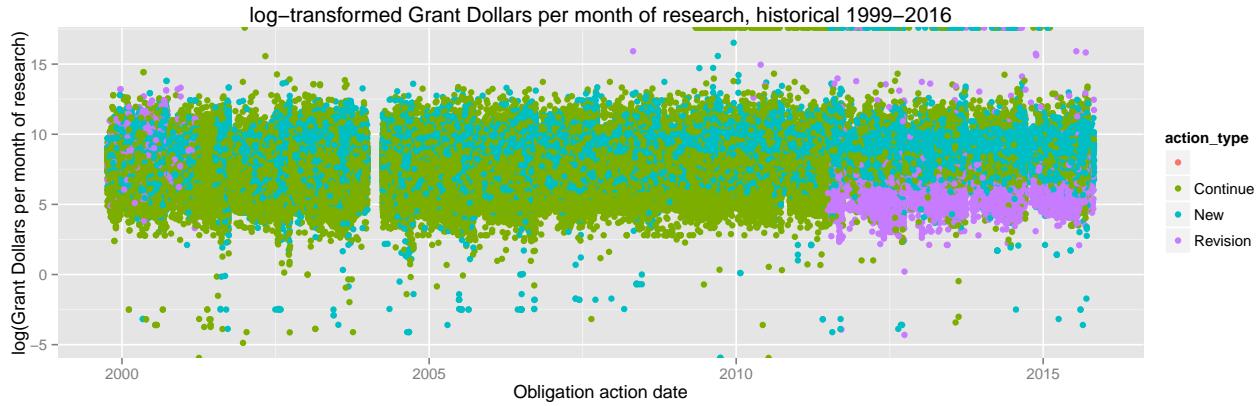
The following variables were added as new columns * res_duration - This is the duration of research obtained by calculating the difference between the ending and starting dates of research. Such values were divided by 30 to obtain duration in months.

- costmon - The variable fed_funding_amount was divided by res_duration to establish the common base for comparing amounts allocated to different recipients.
- monstar, mondec - Month of research start and obligation action date respectively.
- logcost - Log-transformed costmon calculated for plotting these data.

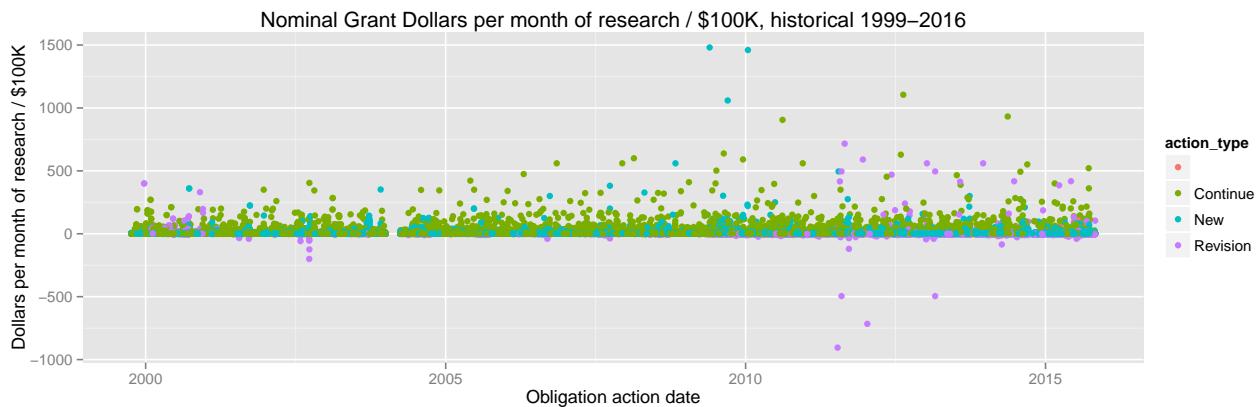
The resulting dataset consists of 410,292 observations of 17 variables. Various subsets were created to calculate cumulative amounts granted to states, programs and by months. Subsets of top recipient states was also created for similar purposes.

General observations

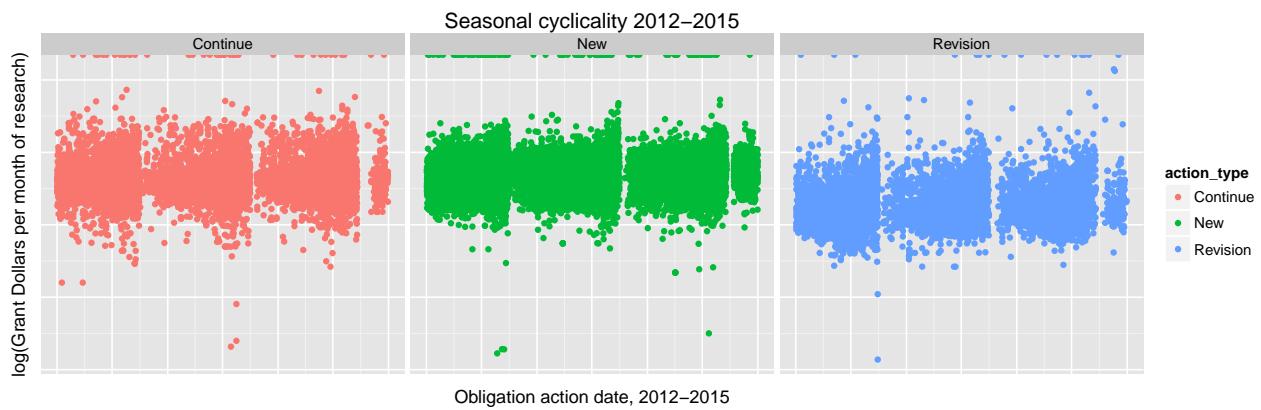
About this plot:



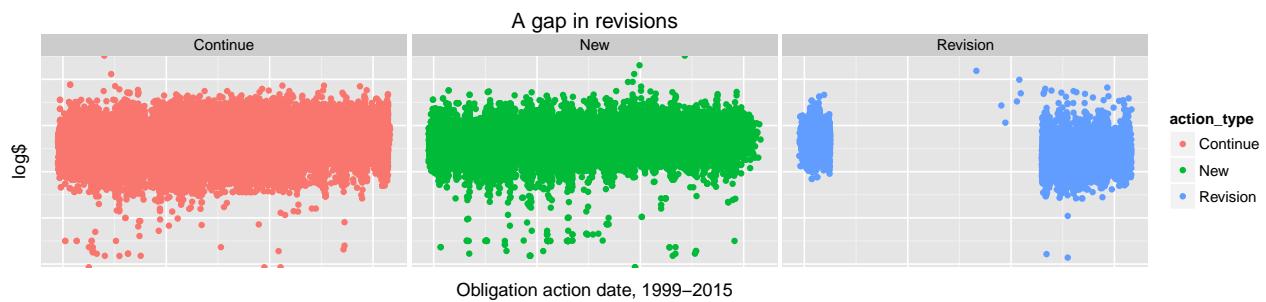
About this plot



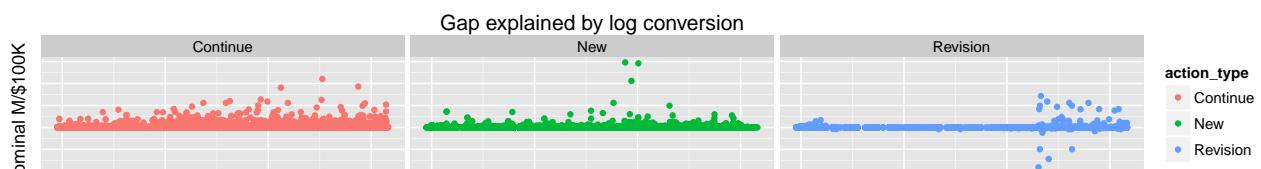
About this Plot



[About this plot](#)

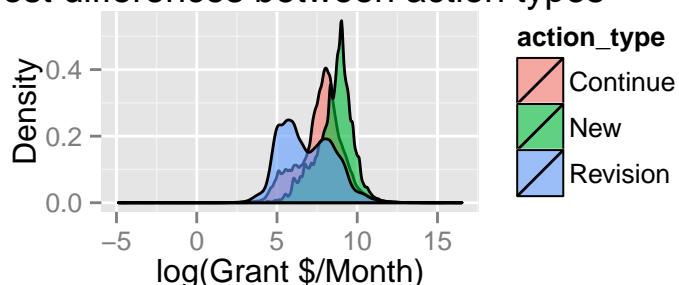


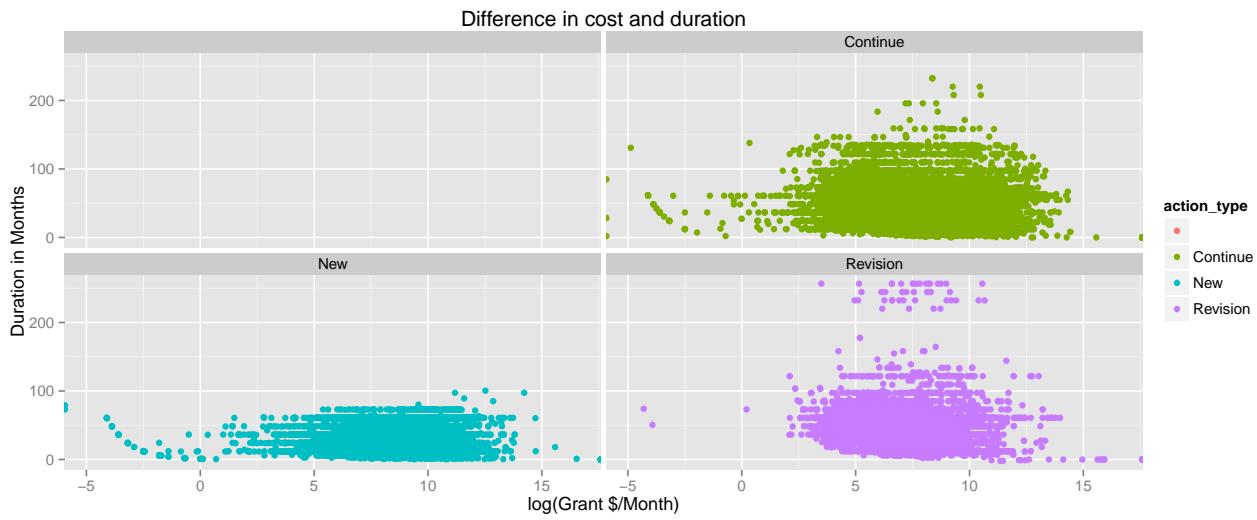
[About this plot](#)



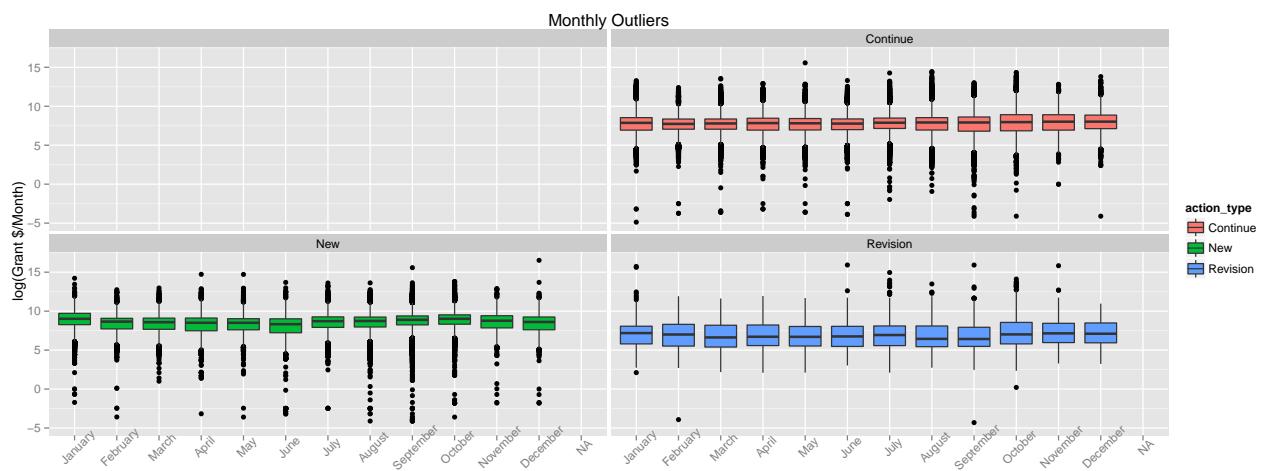
[About this plot](#)

Cost differences between action types

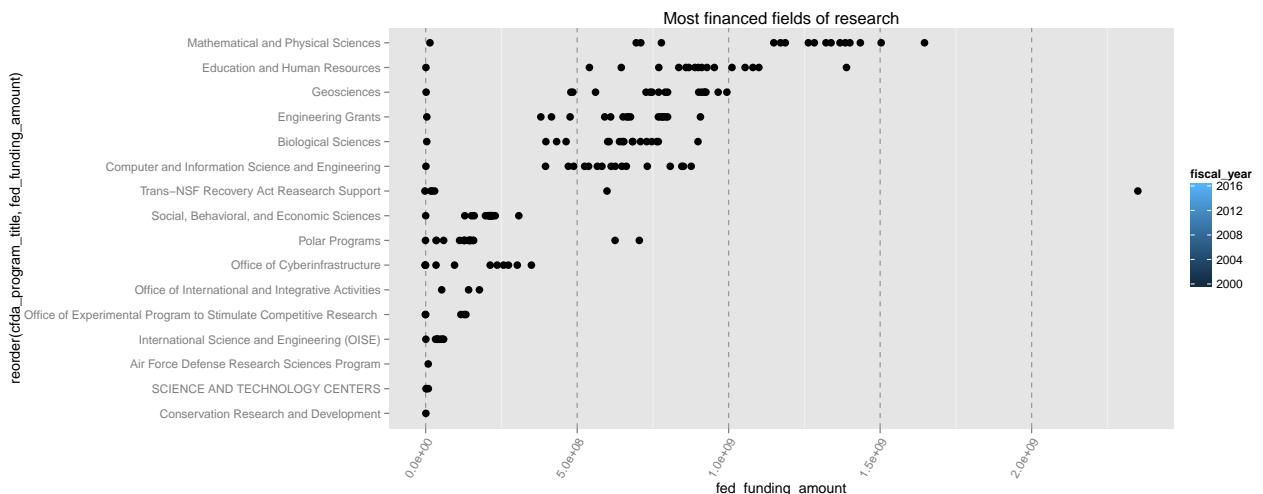




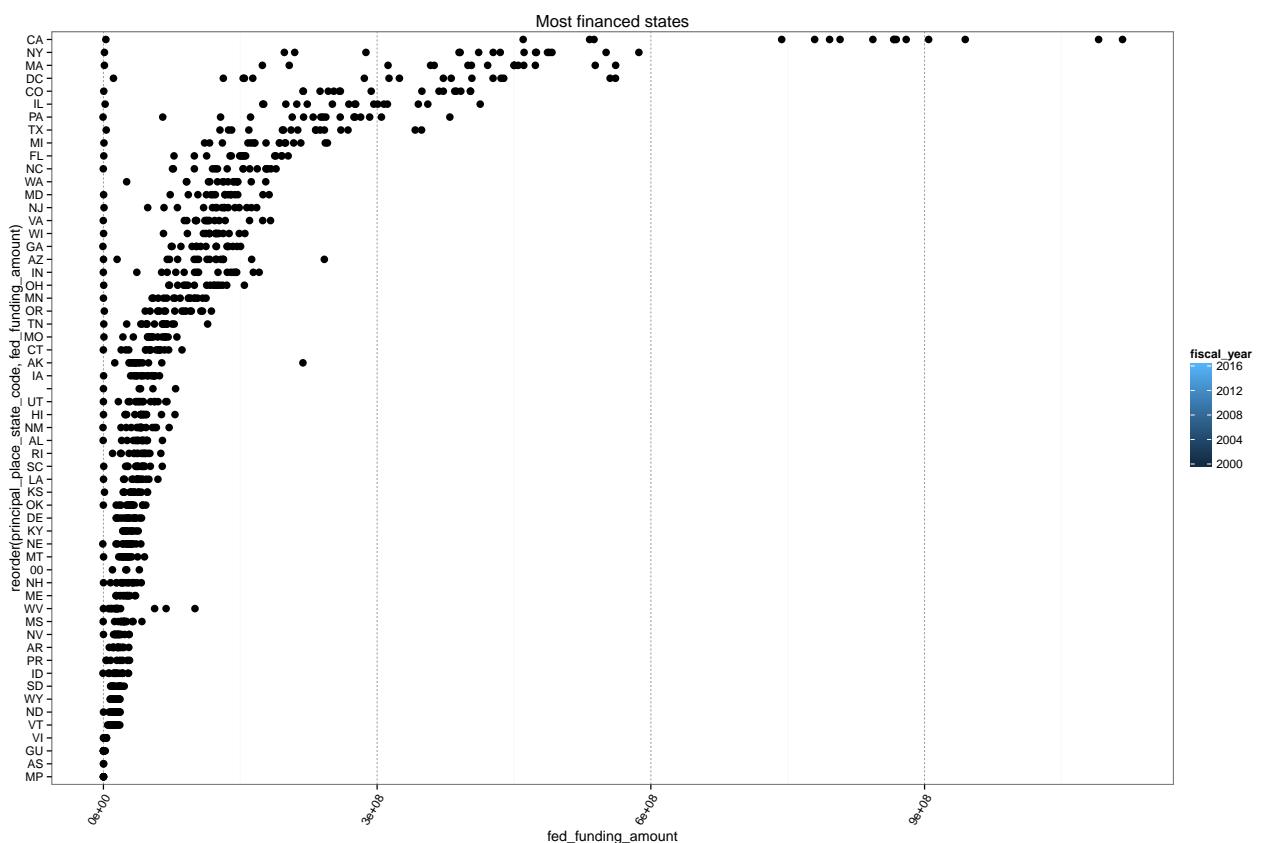
About this plot

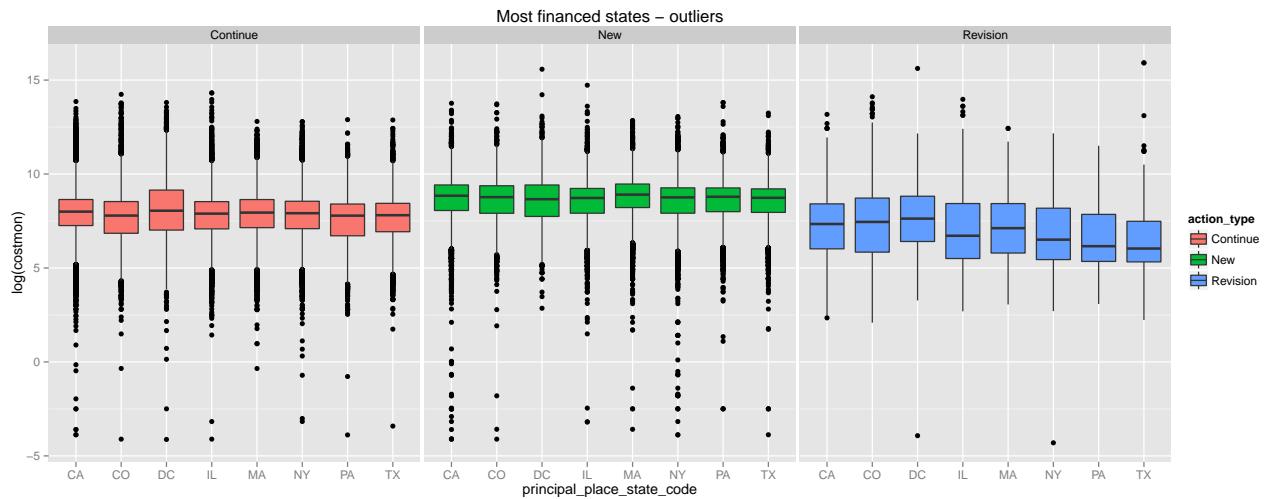


Fields of research

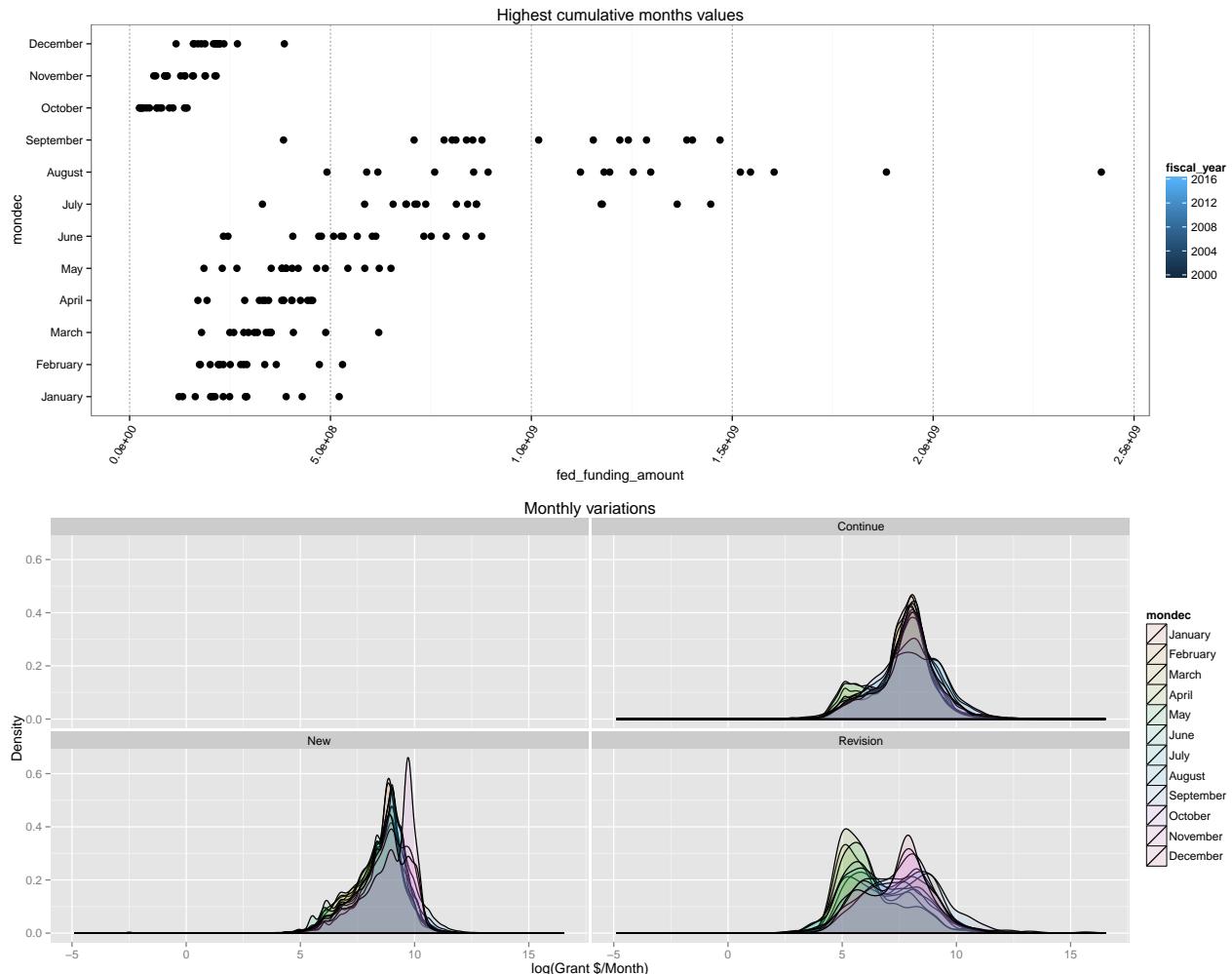


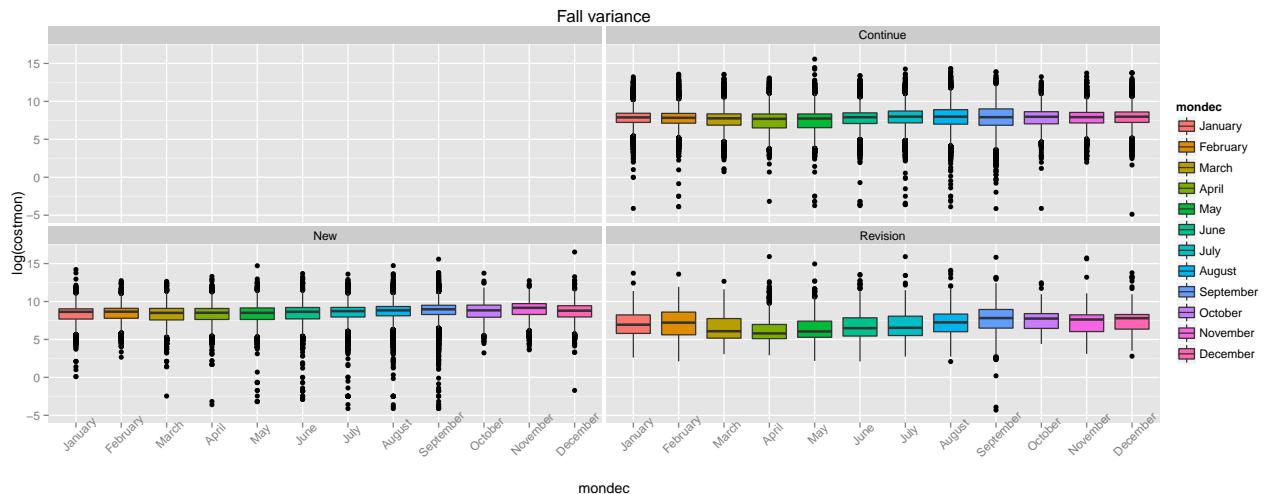
States



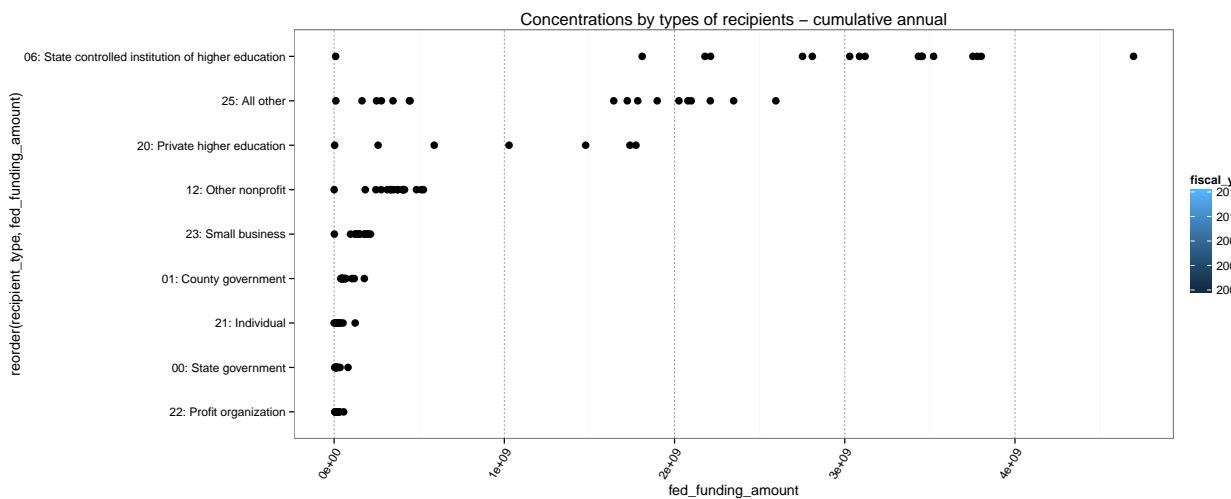


Months





Recipient



Discussion

Bibliography