

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

The National Science Foundation is

Methodology

The dataset used in this study is comprised of queries obtained from <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 data frame are:

recipient_name	Recipient name
recipient_city_name	Name of recipient's location city
recipient_type	Type of recipient ranging from non-profits and corporations to higher educational institutions
action_type	New grant , Continuing grant or revised
fed_funding_amount	Dollar amount granted
obligation_action_date	Grant decision made
starting_date	Start of research
ending_date	End of research
principal_place_state	Recipient location state
cfda_program_title	Field of research
fiscal_year	Budget year
principal_place_state_code	Recipient location state code

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 identcal values
4. cfda_programm_title values werereplaced by more concise Continue, New and Revision

Following variables were added as new columns

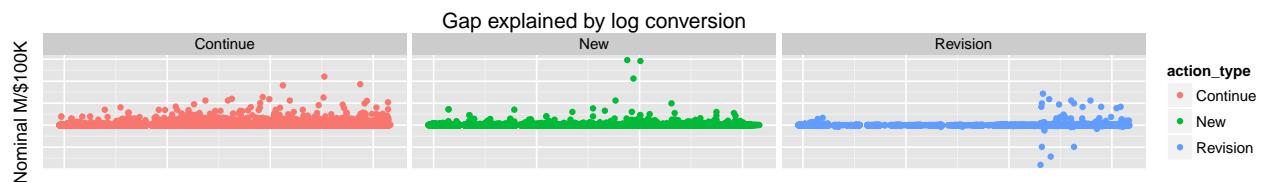
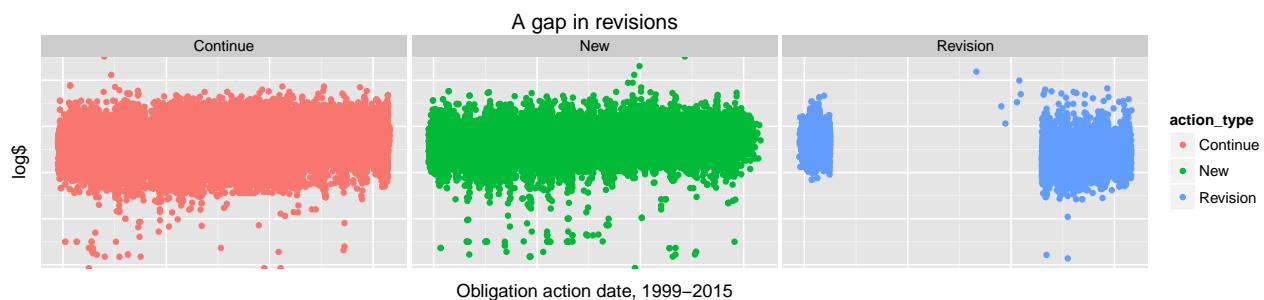
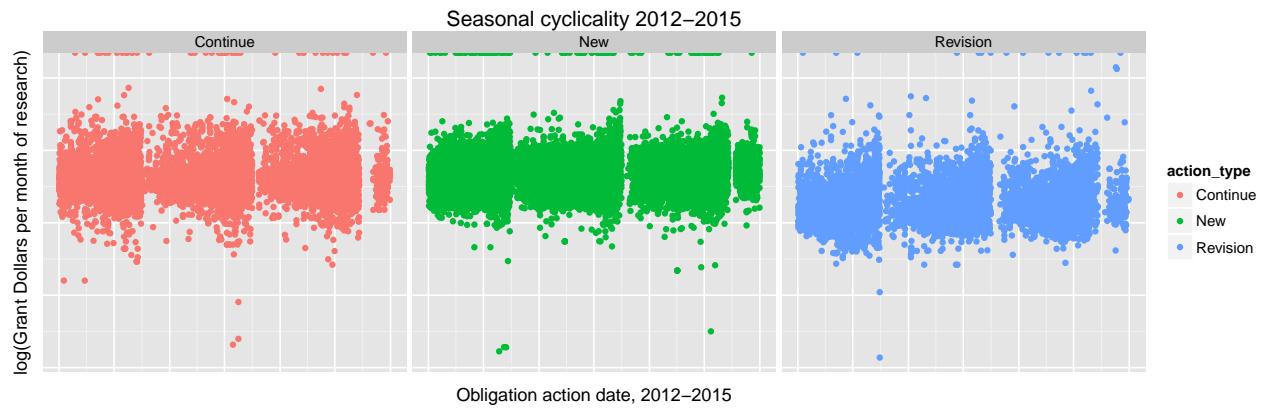
1. res_duration - duration of research obtained by calculating the difference between the ending and starting dates of research. Such values were divided by 30 to obtain duration in months
2. costmon - fed_funding_amount was divided by res_duration to establish the common base for comparing amounts alocated to different recipients
3. monstar, mondec - month of research start and obligation action date respectively
4. logcost - log-transformed costmon calculated for plotting

The resulting dataset consists of 410292 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.

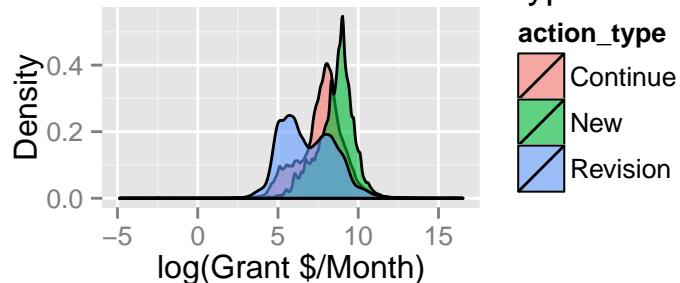
General observations

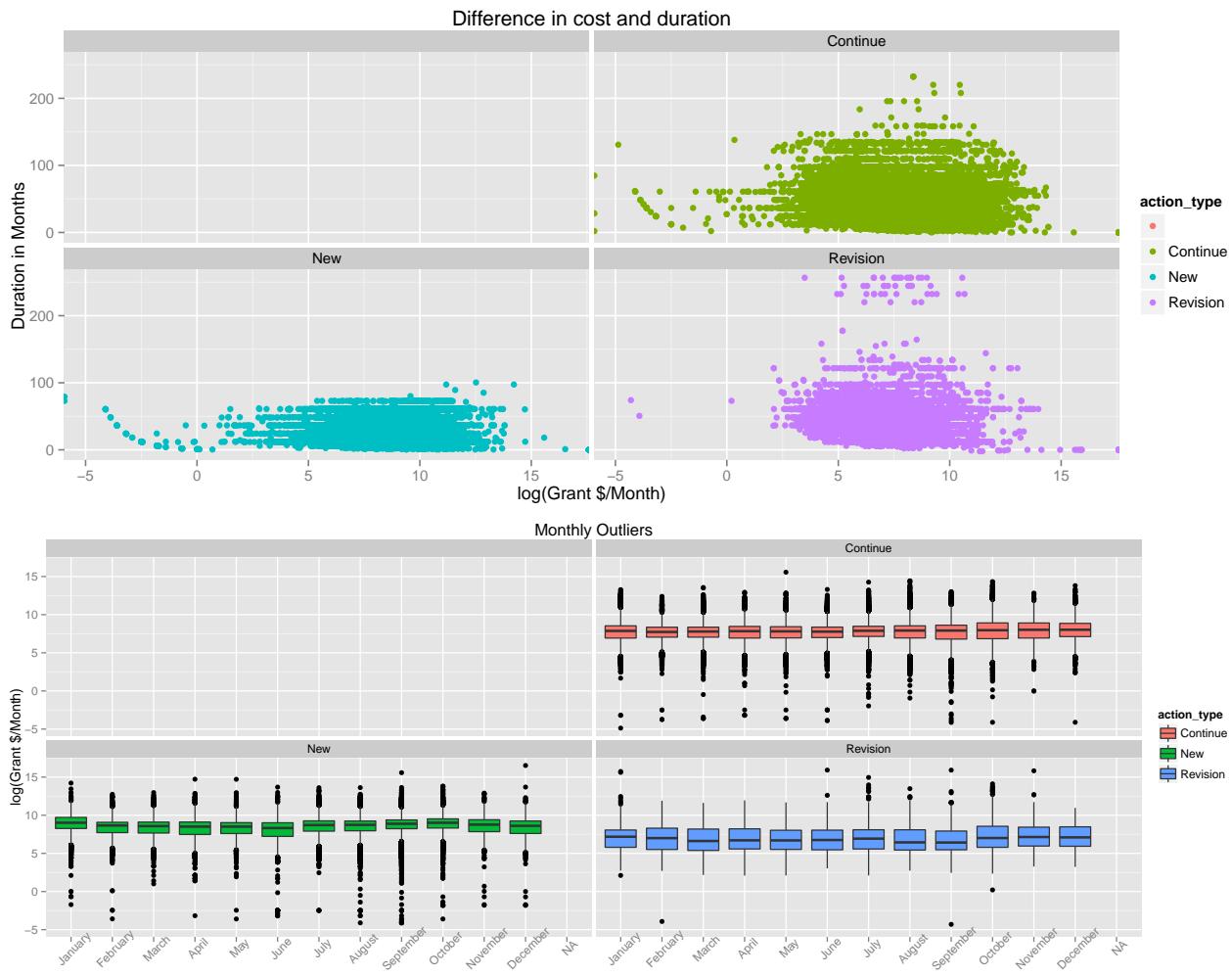
As seen on the plot



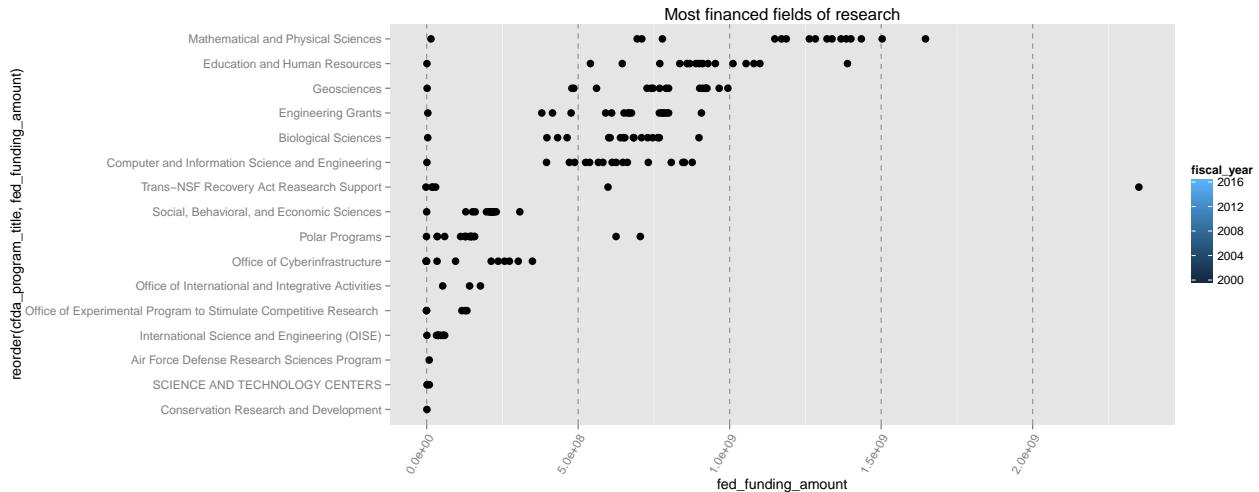


Cost differences between action types

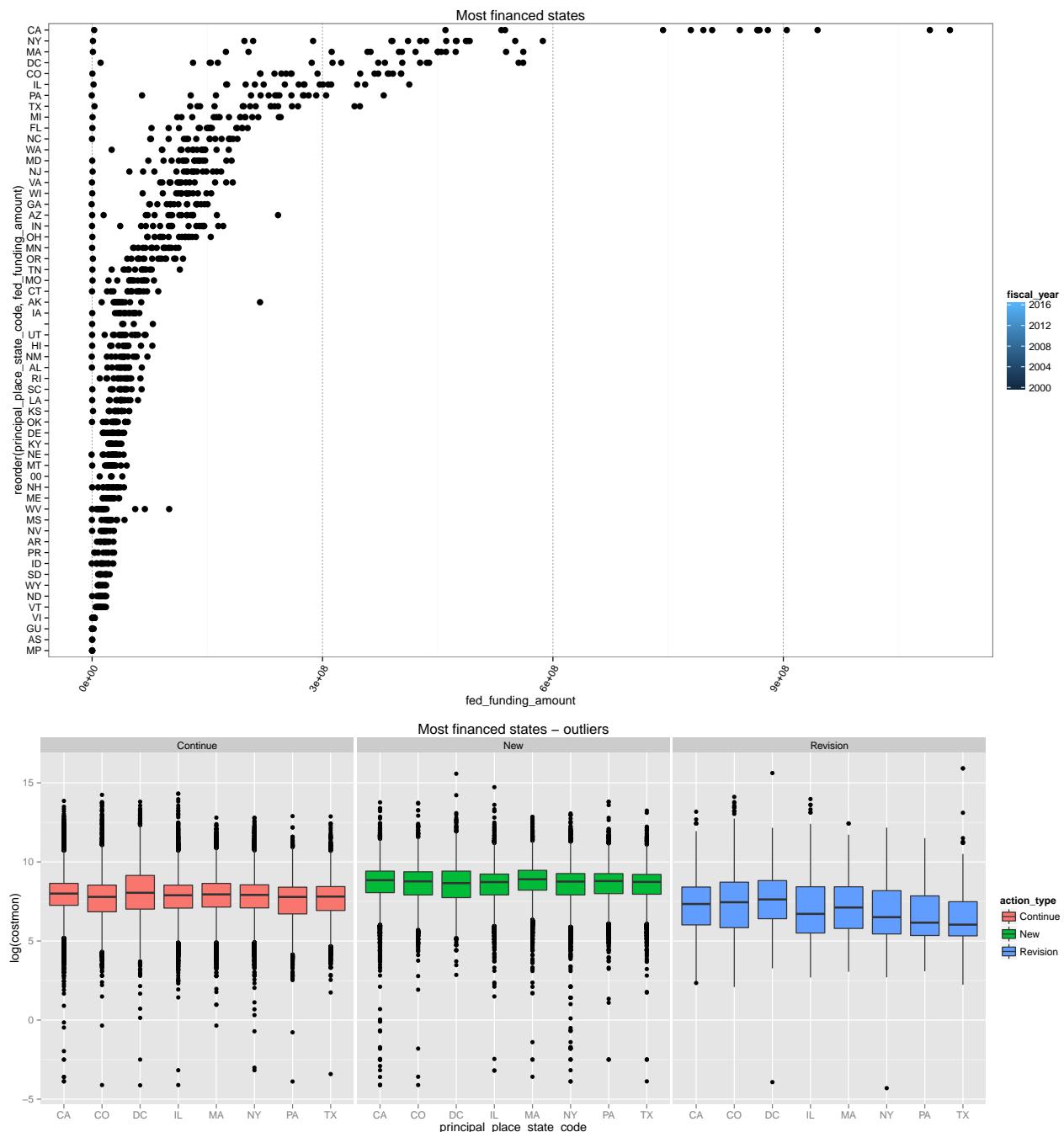




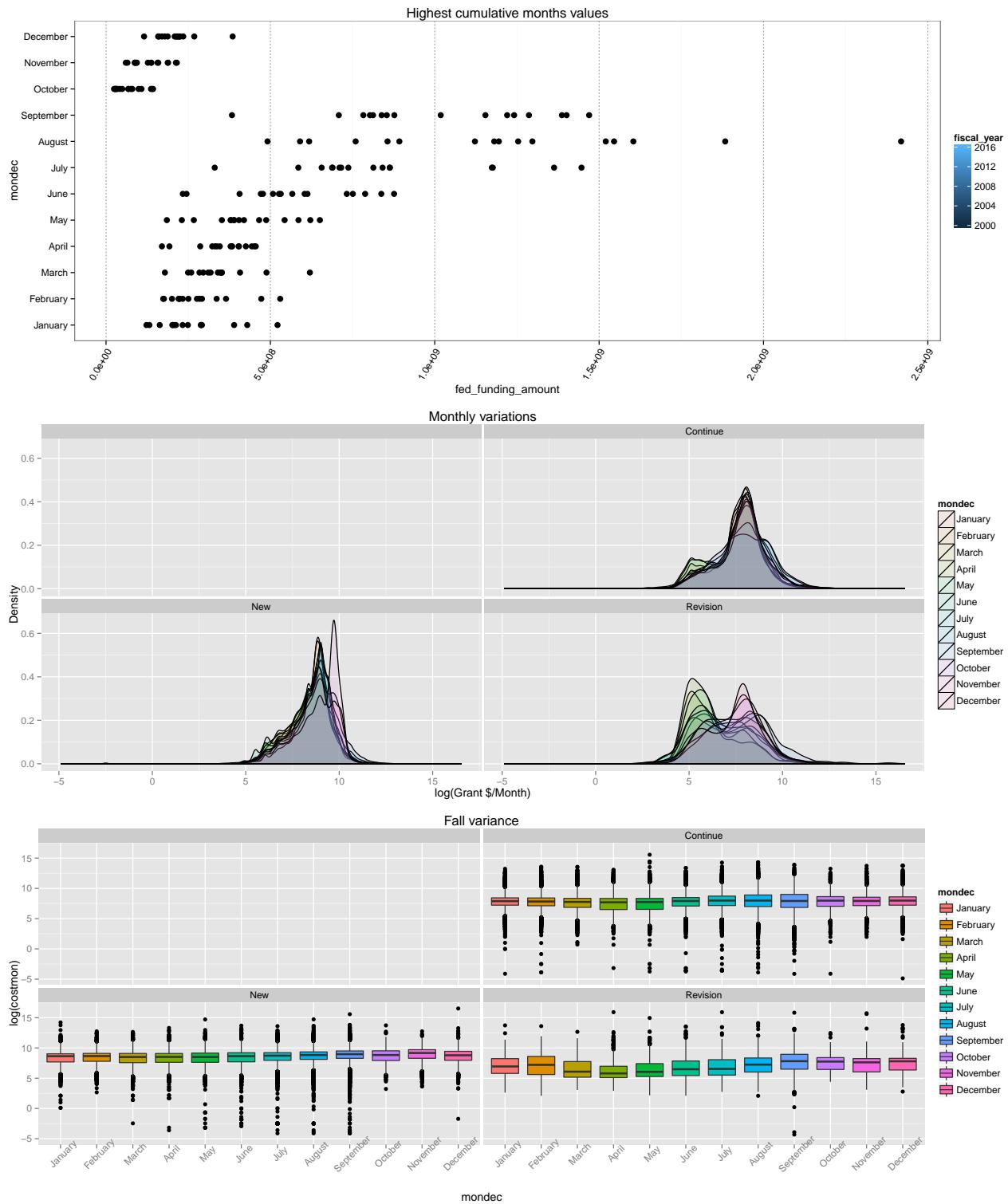
Fields of research



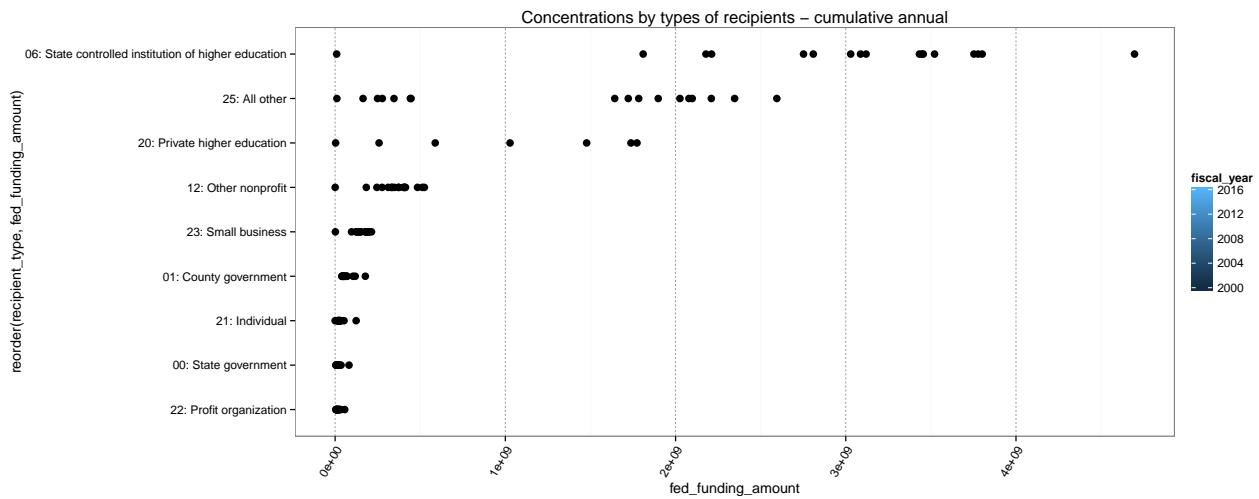
States



Months



Recipient



Discussion

Bibliography