NIH Shared Instrument Grant (S10) Data Analysis

(FY 2006-2015; One decade; 1096 awards totaling $440.8 M)

PRELIMINARY DATA PREPARATION

1. Download the data from the URL: http://dpcpsi.nih.gov/orip/diic/fy_sig_awards
2. Save the data as text: nih_sig_award.txt
3. Move the file to a Linux machine
4. Sample of text file (`head -10 nih_sig_award.txt`)

   Alabama 7 Awards ($2,458,303)
   University of Alabama at Birmingham
   Robert H. Carter 2006
   Flow Cytometry & High Speed Cell Sorting - $303,810
   Nepalli R. Krishna 2006
   Cryoprobe Accessory for a 600 MHz NMR Spectrometer - $267,500
   Lawrence S. Lamb 2006
   BD Biosciences FACS Canto Flow Cytometer - $174,602
   Andrei V. Stanishevsky 2007
   Imaging/Mapping Fourier Transform Infrared Microscope - $223,747
5. `grep '\$' nih_sig_award.txt >nih_sig_award-short.txt`
6. `grep -v '(\$' nih_sig_award-short.txt>nih_sig_award-shorter.txt`
7. `cut -f 1 -d '$' --complement nih_sig_award-shorter.txt | sort >nih_sig_award-amount.txt`

Line 5 `grep` makes printing only the lines that contain '$' in it (removing other lines)

Line 6 `grep` prints only lines that don't contain '($' in it (removing lines that are state total amounts)

Line 7 `cut` makes field #1 that has the delimiter '$' and prints the complement and sorts (high to low)

Alternatively, the whole thing can be done in one command using pipes

8. `grep '\$' nih_sig_award.txt | grep -v '(\$' | cut -f 1 -d '$' --complement | sort> nih_sig_award-amount.txt`

If we are looking for the State-wide award numbers and totals dollar amounts then,

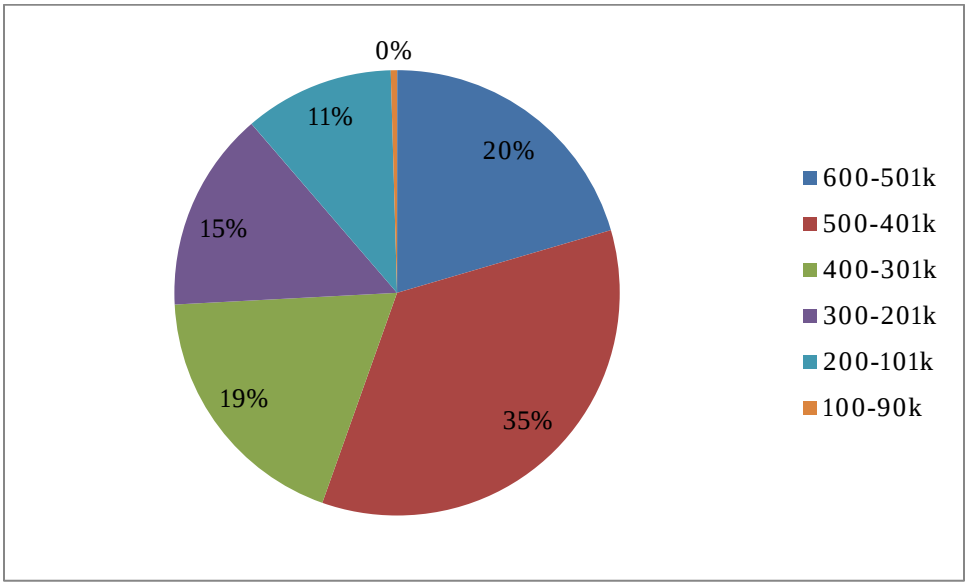
a. `grep '\$' nih_sig_award.txt | grep '(\$' >nih_sig_award-state.txt`

b. Sample of text file (`head -5 nih_sig_award-state.txt`)

Alabama 7 Awards ($2,458,303)
Arizona 9 Awards ($3,378,517)
Arkansas 2 Awards ($700,818)
California 173 Awards ($74,994,165)
Colorado 29 Awards ($10,377,121)

ANALYSIS FOR MID-PRICE RANGE INSTRUMENT FUNDING (FY 2006-15)

Now that we have the decade long data for all S10 grants, using Microsoft Excel I sorted them from highest ($600,000) to the lowest ($91,094) and binned them into six ranges ($600-501k, $500-401k, $400-301k, $300-201k, $200-101k, $100-91k). This data was then plotted and the plot is shown below:



From this graph it is very clear that lion's share of the grants (74%) funded equipment costing >$301k and majority (55%) of the grants funded equipment costing >$401k. Only 11% of the grants funded equipment costing below $200k.

PRELIMINARY DATA PREPARATION FOR FY 2015 ALONE FROM DECADE DATA

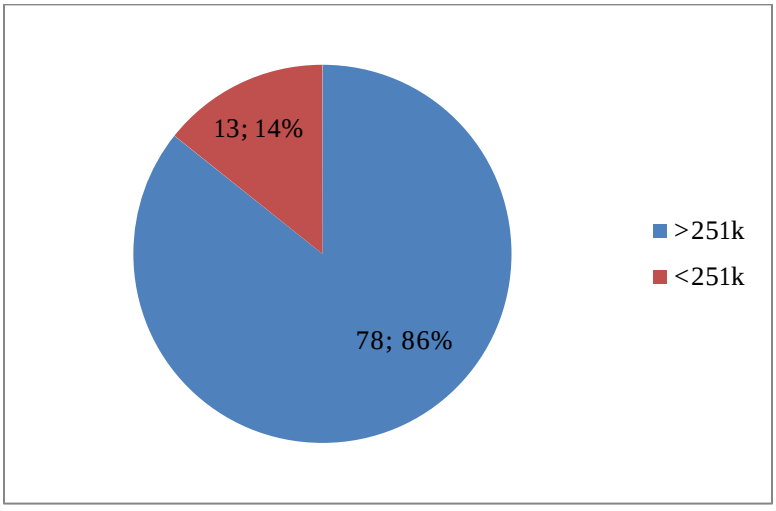
1. Download the data from the URL: http://dpcpsi.nih.gov/orip/diic/fy_sig_awards
2. Save the data as text: nih_sig_award.txt
3. Move the file to a Linux machine
4. Sample of text file (`head -10 nih_sig_award.txt`)
   Alabama 7 Awards ($2,458,303)
   University of Alabama at Birmingham
   Robert H. Carter 2006
   Flow Cytometry & High Speed Cell Sorting - $303,810
   Nepalli R. Krishna 2006
   Cryoprobe Accessory for a 600 MHz NMR Spectrometer - $267,500
   Lawrence S. Lamb 2006
   BD Biosciences FACS Canto Flow Cytometer - $174,602
   Andrei V. Stanishevsky 2007
   Imaging/Mapping Fourier Transform Infrared Microscope - $223,747
5. `grep '2015' –A1 nih_sig_award.txt`
6. `grep '2015' -A1 nih_sig_award.txt |grep '\$' | wc`
7. `grep '2015' -A1 nih_sig_award.txt |grep '\$' |cut -f 1 -d '$' --complement |sort >nih_sig_award-2015-amounts.txt`

Line 5 prints only lines that contain '2015', the year we are interested in it (removing other lines)

Line 6 prints only line containing '2015' and one line afterwards (-A option). The second grep then only prints line containing '$' amount

Line 7 cuts field #1 that has the delimiter '$' and prints the complement and sorts low to high and writes to a file called nih_sig_awards-2015-amounts.txt

Now that we have the 2015 for all S10 grants, using Microsoft Excel I sorted them from highest ($600,000) to the lowest ($91,094) and binned them into six ranges ($600-501k, $500-401k, $400-301k, $300-201k, $200-101k, $100-91k). In Excel I used the function like the one shown here (`=COUNTIF(A2:A92,"<=600000")-COUNTIF(A2:A92,"<=500001")`) to get $600-501k awards. This data was then plotted and the plot is shown below:



Thayumanasamy Somasundaram
Florida State University | tsomasundaram@fsu.edu | biophysics.fsu.edu/soma
Ver. 3: Mar 15, 2016. Ver. 1: January 11, 2016