

# NSF Major Research Instrumentation (MRI) Program Data Analysis

(FY 2008-2016; active awards)

## DATA PREPARATION AND ANALYSIS

- Go to the URL: <http://www.nsf.gov/awardsearch/>
- Keep the “Active” box selected
- Search for “MRI”
- There were 767 awards when done in Feb 2016

1. Save the data as text: NSF\_MRI\_Active-FullDataAsText.txt
2. Move the file to a Linux machine
3. Sample of text file (head -10)

```
head -10 NSF-MRI-Active-FullDataAsText.txt
Award Number: 0932272
Title: CPS: Medium: Image Guided Robot-Assisted Medical Interventions
NSF Org: CNS Division Of Computer and Network Systems
Initial Amendment Date: September 15, 2009
Latest Amendment Date: January 23, 2015
Award Instrument: Standard Grant
Program Manager: David Corman
Start Date: September 01, 2009
End Date: February 29, 2016
Awarded Amount to Date: $1,558,492
```

4. `grep 'Awarded Amount' -B2 NSF-MRI-Active-FullDataAsText.txt>nsf-active.txt`
5. `grep -v 'End' nsf-active.txt>nsf-active2.txt`
6. `grep -v '\-' nsf-active2.txt>nsf-active3.txt`

Line 4 prints only lines that contain 'Awarded Amount' in it and two lines (-B2 option) prior to it

Line 5 prints only lines that don't contain 'End' in it (removing 1/3 of the not needed lines)

Line 6 prints only lines that don't contain '--' in it (“\” used to escape the special character “-”)

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

```
grep 'Awarded Amount' -B2 NSF-MRI-Active-FullDataAsText.txt | grep -v
'End' |grep -v '\-' >nsf-active-award-amounts.txt
```

We need to carry out two more steps on this file. The first step is to get the award years-only data we do that by issuing the following command,

```
grep '\$' nsf-active-award-amounts.txt | cut -f 1 -d ',' --complement
>nsf-active-award-years.txt
```

The next step is to get the award dollars-only data we do that by issuing the following command,

```
nsf-active-award-dollars.txt
```

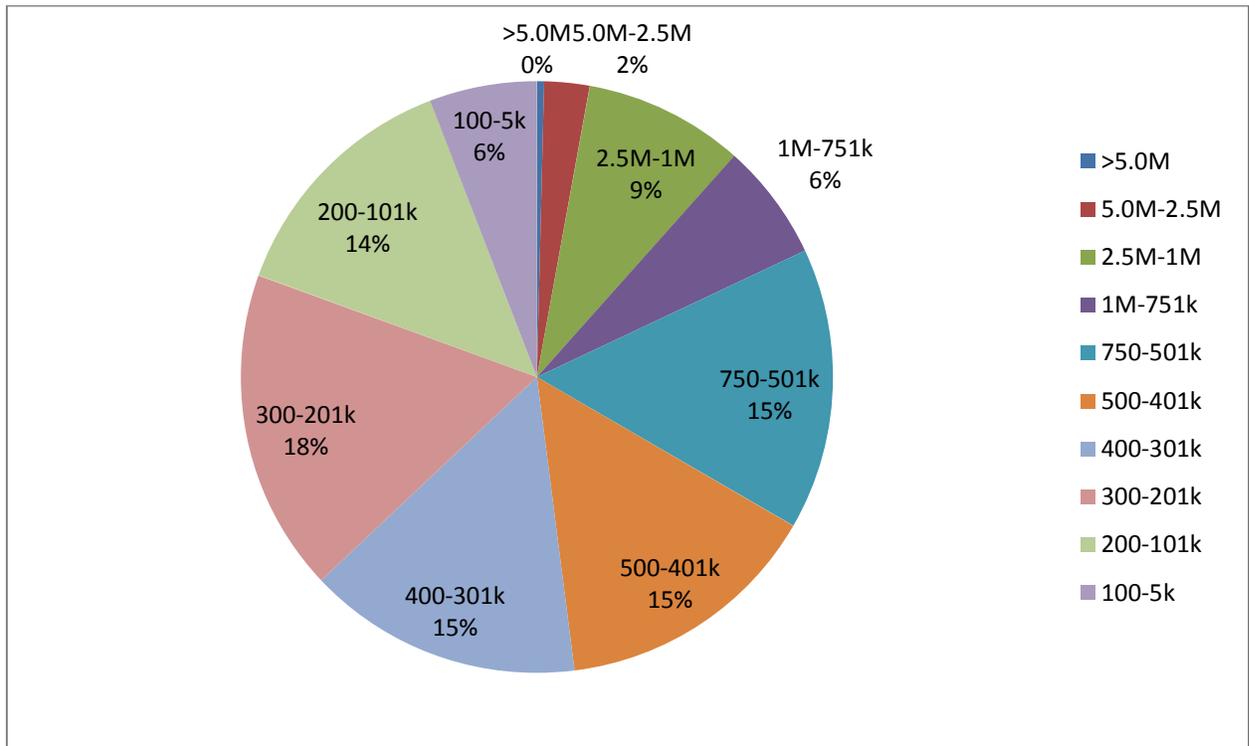
Next is to combine these two data into one file

```
pr -mts nsf-active-award-years.txt nsf-active-award-dollars.txt> nsf-active-award-years+dollars.txt
```

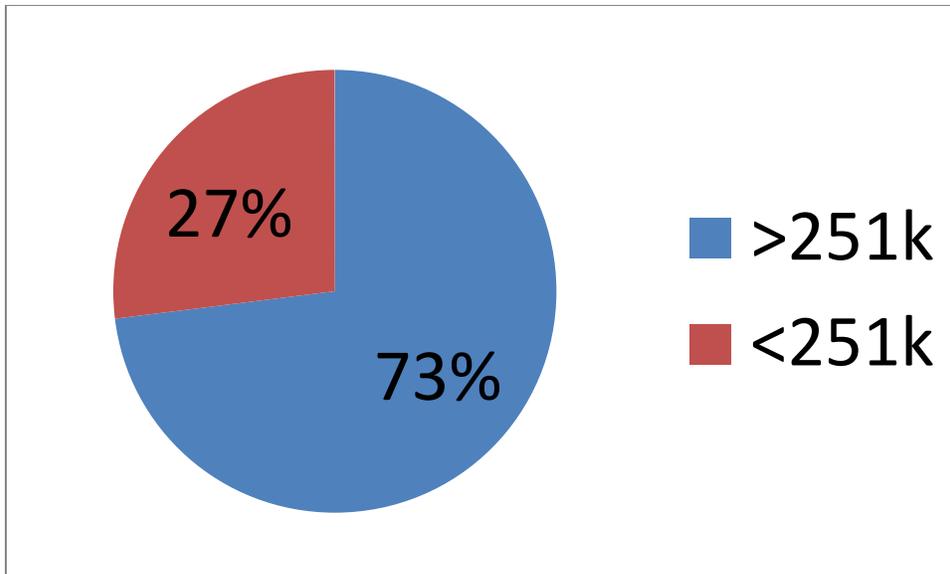
Here the command 'pr' paginates or columnates the file for printing. -m option is for merging two columns, -t is to omit headers and trailers, -s option is for adding 'TAB' as the column separator.

Now bring this file to a Microsoft Windows computer and sort the file with lowest dollar amount on the top and highest dollar amount at the bottom with "sort expanded to year-column" and then binned them into ten ranges (>\$5M, \$5-2.5M, \$2.5-1M, \$1M-751k, \$750-501k, \$500-401k, \$400-301k, \$300-201k, \$200-101k, \$100-5k). Now graph the Excel file with various award ranges as shown below. [I used the following Excel function to bin the data into various dollar ranges to get the number of awards (here it was 49 awards out of the 767, etc.)

```
=COUNTIF(B1:B767, "<=1000000") - COUNTIF(B1:B767, "<=750000")
```



The next graph shows the awards that are above and below \$250k. [I used the following Excel function =COUNTIF(B1:B761, ">250001") to get the number of awards above \$251k (here it was 207 awards out of the 767)].



## HOW MANY AWARDS WERE BIOLOGICAL OR MEDICAL

Once I posted the analysis in mid-February 2016 there was a question about how much of the awards were Biological or Medical research related awards. A quick search of the words “Bio or bio” or “Med or med” of the Award Titles yielded about 50+12 hits out of the 767 awards. But this was qualitative search since not all award title will tell about application of the grant. Here is how I did the quick search (see later for more quantitative search)

```
grep 'Award Number' -A 1 NSF-MRI-Active-FullDataAsText.txt | grep
'Title:'>nsf-active-award-titles.txt
egrep -i 'bio|med' nsf-active-award-titles.txt | wc
62
```

Then I prepared the abstracts only section and did another search. Here is how I did the second search.

```
egrep -i 'biolog|medic' nsf-active-award-abstracts.txt | wc
380
egrep -i 'biologic' nsf-active-award-abstracts.txt | wc
181
egrep -i 'medic' nsf-active-award-abstracts.txt | wc
192
```

So when I searched either for the words “biologic” or “medic” on the abstracts of the grants, it yielded 380 results. Note that these numbers are one occurrence per abstract and NOT multiple occurrence of the same word multiple times in a single grant. So 380 out of 760 is a good chance of getting funded for medical or biological research work.

Thayumanasamy Somasundaram

Florida State University | [tsomasundaram@fsu.edu](mailto:tsomasundaram@fsu.edu) | [biophysics.fsu.edu/soma](http://biophysics.fsu.edu/soma)

Ver. 3: Mar 18, 2016. Ver. Feb 10, 2016