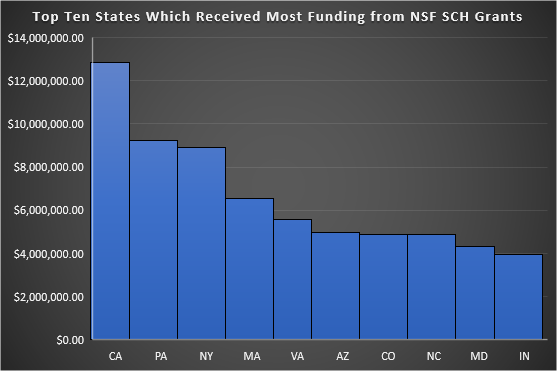
Jared Palmer

**Section One: Distribution of NSF SCH Funds among States**

The top ten states that received the most funding from the National Science Foundation (NSF) for Smart and Connected Health projects are ordered from greatest to least as follows: California ($12,826,821), Pennsylvania ($9,245,234), New York ($8,896,487), Massachusetts ($6,528,166), Virginia ($5,582,868), Arizona ($4,969,154), Colorado ($4,880,159), North Carolina ($4,860,288), Maryland ($4,310,800), and Indiana ($3,947,792).



**Section Two: SCH Organizations**

Out of the 133 organizations that received funding from the National Science Foundation (NSF), these five organizations received the most funding in terms of dollars. The University of California-San Diego received the most funding with $5,174,242, Indiana University received the second most with $4,877,286, University of Memphis with the third most at $4,752,384, Carnegie-Mellon University in fourth with $4,050,717 and in fifth we have Johns Hopkins University with $3,808,313.

In this pie chart, we can see how the funds are distributed to the organizations in the top five of funds received from the NSF.

Additionally, in our next pie chart, we can see that the top five most funded organizations are not receiving an unusually large portion of the overall funds given out by the National Science Foundation.

Out of the 295 total funded projects, these are the organizations with the most projects funded by the National Science Foundation (NSF): Indiana University (10), Georgia Tech Research Corporation (7), University of Southern California (7), Arizona State University (6), Carnegie-Mellon University (6), University of Minnesota-Twin Cities (6), & University of Virginia Main Campus (6).

This pie chart shows the share of funded projects by the top seven organizations. As we can see, the organizations at the top are very close when it comes to how many projects were funded by the NSF. The top seven organizations seen here account for only 17% of the total projects funded by the NSF in this dataset. The next pie chart shows the comparison between the organizations with the most projects funded by the NSF versus the total number of projects funded.

**Section Three: How is Texas Doing on SCH?**

When it comes to the amount of funding received from the National Science Fund (NSF) on Science and Connected Health projects, Texas is ranked 11th overall out of the states included in this dataset. Texas has received $3, 551, 266 in funding from the NSF for various Science and Connected Health projects through universities such as the University of Texas-Austin, University of Texas-Arlington, University of Texas-Tyler, University of North Texas, Texas A&M, and the University of Texas-Dallas. When it comes to the number of projects funded by the NSF for Science and Connected Health, Texas had 14 projects funded out of 201 total projects.

**Section Four: An Association Analysis of NSF SCH Projects**

An association rules analysis was completed on the Abstract value of the data set. The results that were returned after the analysis was completed showed several terms that can be found together in over half of the abstracts. The association analysis was set to look for patterns with a minimum support of 0.5 and a minimum confidence of 0.6. Below shows a table of terms listed in descending order based upon support. As you can see, there are several terms that are present in over half of the abstracts listed in this dataset: “project”, “research”, “health”, and “data”. Additionally, there were a few terms that were found in slightly less than 50% of the abstracts but still had a high confidence percentage. These terms are “develop”, “based”, and “system”.

|  |  |  |  |
| --- | --- | --- | --- |
| **PREMISE** | **CONCLUSION** | **SUPPORT** | **CONFIDENCE** |
| project | research | 0.6825938566552902 | 0.8264462809917356 |
| research | project | 0.6825938566552902 | 0.8333333333333334 |
| research | health | 0.6143344709897611 | 0.75 |
| health | research | 0.6143344709897611 | 0.8411214953271028 |
| project | health | 0.6006825938566553 | 0.7272727272727273 |
| health | project | 0.6006825938566553 | 0.822429906542056 |
| project | data | 0.590443686006826 | 0.7148760330578512 |
| data | project | 0.590443686006826 | 0.8781725888324873 |
| research | data | 0.5460750853242321 | 0.6666666666666666 |
| data | research | 0.5460750853242321 | 0.8121827411167513 |
| health | data | 0.5085324232081911 | 0.6962616822429907 |
| health | project, research | 0.5085324232081911 | 0.6962616822429907 |
| project, research | health | 0.5085324232081911 | 0.745 |
| data | health | 0.5085324232081911 | 0.7563451776649747 |
| research, health | project | 0.5085324232081911 | 0.8277777777777777 |
| project, health | research | 0.5085324232081911 | 0.8465909090909091 |
| based | project | 0.5085324232081911 | 0.9141104294478528 |

**Section Five: A Clustering Analysis of NSF SCH Projects**

A clustering analysis was completed on the Abstract value of this dataset. For the sake of this analysis, the “k” was set to “10” and the max runs was set to “4”. Below I will list the 10 clusters with the first 20 terms from each cluster.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Attribute**   |  | | --- | | design | | develop | | developed | | development | | provide | | analysis | | detect | | goal | | health | | monitoring | | sensor | | control | | critical | | detection | | device | | integrated | | methods | | patients | | specific | | time | | **Cluster\_0**   |  | | --- | | 0.5333333333333333 | | 0.5333333333333333 | | 0.5333333333333333 | | 0.5333333333333333 | | 0.5333333333333333 | | 0.4666666666666667 | | 0.4666666666666667 | | 0.4666666666666667 | | 0.4666666666666667 | | 0.4666666666666667 | | 0.4666666666666667 | | 0.4 | | 0.4 | | 0.4 | | 0.4 | | 0.4 | | 0.4 | | 0.4 | | 0.4 | | 0.4 | |
| **Attribute**   |  | | --- | | address | | advance | | assess | | available | | build | | central | | communities | | contribute | | develop | | developed | | developing | | dynamics | | impact | | inform | | make | | power | | process | | project | | research | | social | | **Cluster\_1**   |  | | --- | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | |
| **Attribute**   |  | | --- | | health | | based | | project | | data | | design | | research | | smart | | support | | individuals | | leading | | cost | | effective | | system | | use | | care | | cognitive | | learning | | models | | team | | developed | | **Cluster\_2**   |  | | --- | | 0.875 | | 0.8333333333333334 | | 0.8333333333333334 | | 0.75 | | 0.75 | | 0.75 | | 0.75 | | 0.75 | | 0.7083333333333334 | | 0.7083333333333334 | | 0.6666666666666666 | | 0.625 | | 0.625 | | 0.625 | | 0.5833333333333334 | | 0.5833333333333334 | | 0.5833333333333334 | | 0.5833333333333334 | | 0.5833333333333334 | | 0.5416666666666666 | |
| **Attribute**   |  | | --- | | research | | support | | science | | project | | development | | health | | researchers | | addition | | students | | area | | challenges | | engineering | | experts | | forum | | generation | | multidisciplinary | | participants | | provides | | student | | approaches | | **Cluster\_3**   |  | | --- | | 0.9411764705882353 | | 0.8823529411764706 | | 0.7647058823529411 | | 0.7058823529411765 | | 0.6470588235294118 | | 0.6470588235294118 | | 0.6470588235294118 | | 0.5882352941176471 | | 0.5882352941176471 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.47058823529411764 | |
| **Attribute**   |  | | --- | | develop | | learning | | machine | | project | | assistive | | design | | lack | | potential | | trained | | algorithms | | analysis | | based | | data | | impact | | access | | addition | | associated | | broad | | camera | | develops | | **Cluster\_4**   |  | | --- | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 0.875 | | 0.875 | | 0.875 | | 0.875 | | 0.875 | | 0.75 | | 0.75 | | 0.75 | | 0.75 | | 0.75 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | |
| **Attribute**   |  | | --- | | project | | research | | system | | using | | based | | develop | | data | | time | | support | | health | | systems | | foundation | | impacts | | evaluation | | models | | algorithms | | broader | | development | | human | | award | | **Cluster\_5**   |  | | --- | | 0.9333333333333333 | | 0.8666666666666667 | | 0.7166666666666667 | | 0.7166666666666667 | | 0.6833333333333333 | | 0.6666666666666666 | | 0.65 | | 0.65 | | 0.6166666666666667 | | 0.6 | | 0.5833333333333334 | | 0.5666666666666667 | | 0.55 | | 0.5333333333333333 | | 0.5333333333333333 | | 0.5166666666666667 | | 0.5166666666666667 | | 0.5166666666666667 | | 0.5166666666666667 | | 0.5 | |
| **Attribute**   |  | | --- | | project | | research | | data | | health | | based | | develop | | patients | | students | | time | | provide | | patient | | healthcare | | information | | using | | clinical | | systems | | proposed | | medical | | disease | | improve | | **Cluster\_6**   |  | | --- | | 0.905982905982906 | | 0.8376068376068376 | | 0.7863247863247863 | | 0.7777777777777778 | | 0.5897435897435898 | | 0.5726495726495726 | | 0.5470085470085471 | | 0.5470085470085471 | | 0.5470085470085471 | | 0.5384615384615384 | | 0.5299145299145299 | | 0.5128205128205128 | | 0.5128205128205128 | | 0.5042735042735043 | | 0.48717948717948717 | | 0.47863247863247865 | | 0.4700854700854701 | | 0.46153846153846156 | | 0.452991452991453 | | 0.452991452991453 | |
| **Attribute**   |  | | --- | | health | | technology | | based | | data | | project | | time | | care | | healthcare | | interventions | | key | | research | | behavioral | | environment | | provide | | significant | | social | | understanding | | workshop | | challenges | | computational | | **Cluster\_7**   |  | | --- | | 0.9411764705882353 | | 0.7058823529411765 | | 0.5882352941176471 | | 0.5882352941176471 | | 0.5882352941176471 | | 0.5882352941176471 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.5294117647058824 | | 0.47058823529411764 | | 0.47058823529411764 | | 0.47058823529411764 | | 0.47058823529411764 | | 0.47058823529411764 | | 0.47058823529411764 | | 0.47058823529411764 | | 0.4117647058823529 | | 0.4117647058823529 | |
| **Attribute**   |  | | --- | | health | | research | | healthcare | | participants | | science | | computer | | engineering | | opportunity | | provide | | fields | | related | | social | | challenges | | collaboration | | issues | | researchers | | scientific | | support | | technologies | | work | | **Cluster\_8**   |  | | --- | | 1.0 | | 0.9375 | | 0.875 | | 0.875 | | 0.875 | | 0.8125 | | 0.75 | | 0.75 | | 0.75 | | 0.6875 | | 0.6875 | | 0.6875 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | | 0.625 | |
| **Attribute**   |  | | --- | | data | | project | | research | | information | | develop | | disease | | health | | impact | | methods | | scale | | time | | use | | community | | novel | | students | | system | | approach | | assessing | | based | | evaluation | | **Cluster\_9**   |  | | --- | | 1.0 | | 1.0 | | 1.0 | | 0.8888888888888888 | | 0.7777777777777778 | | 0.7777777777777778 | | 0.7777777777777778 | | 0.7777777777777778 | | 0.7777777777777778 | | 0.7777777777777778 | | 0.7777777777777778 | | 0.7777777777777778 | | 0.6666666666666666 | | 0.6666666666666666 | | 0.6666666666666666 | | 0.6666666666666666 | | 0.5555555555555556 | | 0.5555555555555556 | | 0.5555555555555556 | | 0.5555555555555556 | |

**Section Six: Comments and Conclusions**

After analyzing and reviewing this dataset, I have noticed several items of interest. First, I found it interesting that the organizations that had the most projects funded did not necessarily receive the highest amount of funds from the National Science Foundation. It seems that the National Science Foundation pays close attention to each specific project and what the needs of that specific project are. Once the NSF has this information, they can divvy out the proper amount of funds that each project requires. For example, Indiana ranked first in total projects funded with 10 but only ranked 10th in total funds received.

Next, Texas has a lot of organizations in this dataset that are applying for funding from the National Science Foundation but only a portion of them are centered around Smart and Connected Health. Texas ranked 11th in total projects funded for Smart and Connected Health but I would expect this rank to rise if we were looking at the overall projects funded by the National Science Foundation.

After completing an association rules analysis on the Abstract value of the dataset, I would say that many of the results were what I would expect. A lot of the same terms that are found in the abstracts were terms that we would normally see in professional abstracts that were written for the purpose of getting funding for research projects that are centered around health information. However, there were some results from the clustering analysis that were very interesting.

The clustering analysis resulted in generating 10 clusters in which I listed the first 20 terms from each cluster. For most clusters, it seems that the top terms in each cluster were the same terms that were seen in the association analysis. These terms are: “health”, “research”, “technology”, “project”, “data”, and “develop”. However, there were some clusters that had some terms that were very interesting to me and could be useful to future organizations that are planning on submitting proposals to get funding from the NSF for a project. Clusters 1, 4, 8, and 9 all showed terms that were clustered together closely and were close to the centroid. I believe that many of these terms could hold significance when crafting a future proposal. For instance, cluster\_9 has many terms associated with researching the impact of a disease on a community. Cluster\_8 has terms associated with a collaboration between multiple fields such as health, engineering, and technology. Cluster\_4 seemed to center around the ideas of machine learning, algorithms, and data analysis. Additionally, Cluster\_1 contains several buzzwords that I believe to have no hidden theme but are all powerful terms to use in a proposal.

The purpose of this project was to analyze this dataset to discover any patterns in the way the National Science Foundation funds projects and to see if state, or organization played a big role. From the results of my analysis, I can say that there does not seem to be any obvious pattern in the way the NSF funds projects. It seems that the NSF funds the projects that it sees are worth funding and gives out the appropriate amount of funds depending on the needs of each project. One of the areas of this analysis that I believe is worth revisiting is the clustering analysis of the Abstract value. I was able to complete this analysis and the clustering results were very interesting, but I wonder what the results would be if an additional value was clustered as well. It would be very interesting to see a future analysis done on the Abstract and funds received values to see how certain term usage in the Abstract relates to the amount of funds received from the NSF.