**EMIS 7357 Fall 2018 - Analytics for Decision Support**



Assignment 3:

The Cancer Data from the Moonshot dataset 2016

By:

Jaime García

November 4th, 2018

# Table of Contents

Executive Summary……………………………………………………………………….………3

Question 1…………………………………………………………………………………………4

Question 2…………………………………………………………………………………………6

Question 3…………………………………………………………………………………………7

Question 4…………………………………………………………………………………………8

Question 5…………………………………………………………………………………………9

Question 6………………………………………………………………………………………..12

Question 7………………………………………………………………………………………..15

Question 8………………………………………………………………………………………..17

Question 9………………………………………………………………………………………..20

Question 10……………………………………………………………………………..………..21

Conclusions...……………...……………………………………………………………………..24

References………………………………………………………………………………………..25

**Executive** **Summary**

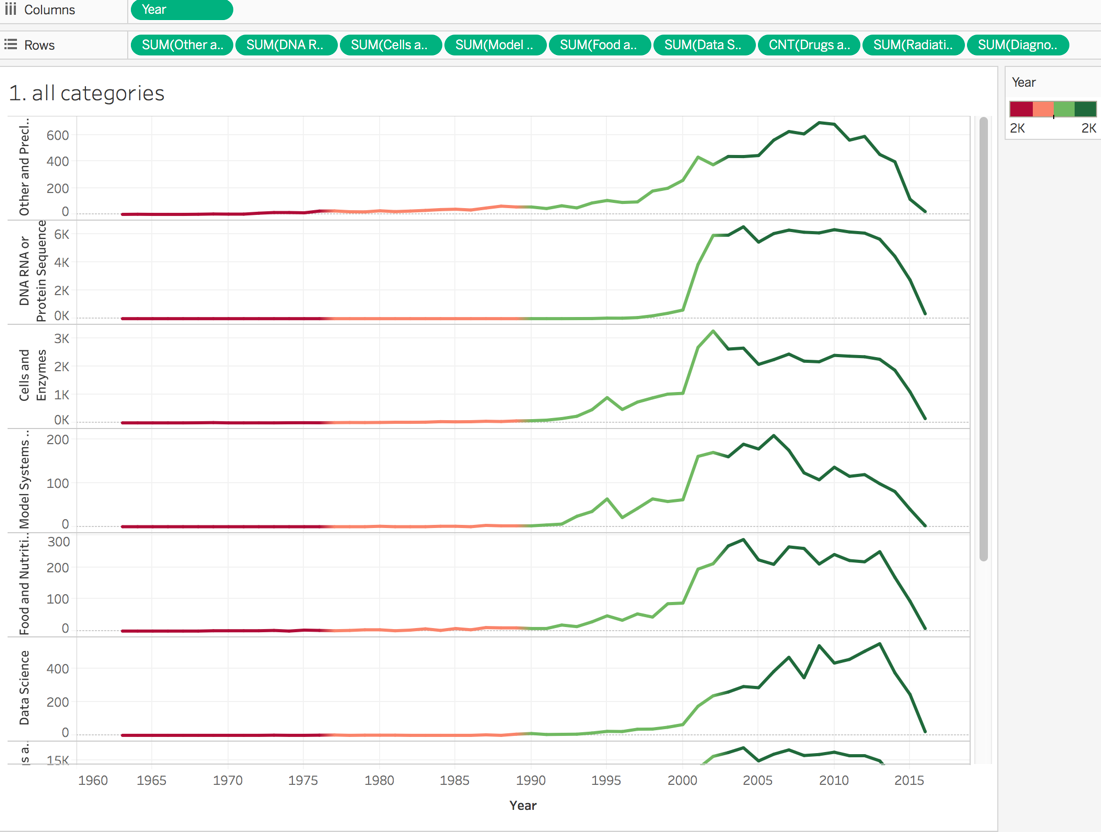
This paper, which is Assignment 3 from the EMIS 7357 course, is based on a dataset provided by the Cancer Moonshot Research, which is a coalition of various institutes with the goal of finding solutions and vaccine-based immunotherapies against cancer. Former president Obama, in his 2016 State of the Union Address, mentioned to Former Vice President Biden to lead a new, national “Cancer Moonshot” to accelerate efforts to prevent, diagnose, and treat cancer. By leveraging decades of scientific understanding from the study and care of cancer, creating and aggregating immensely powerful datasets, and developing unprecedented science and technological capabilities, the commitment to end cancer has brought immense resources.

This dataset pools the resources of multinational pharmaceuticals, biotechnology companies, academic centers, oncologists, and governmental data, intended to creating access to the public under the wills of exploring mechanism for the war against cancer. The goal of this assignment is to explore the dataset of these immunological combinations and develop Tableau visualizations that will enable to create historical analyses of how these protocols have been developed over the period of forty years.

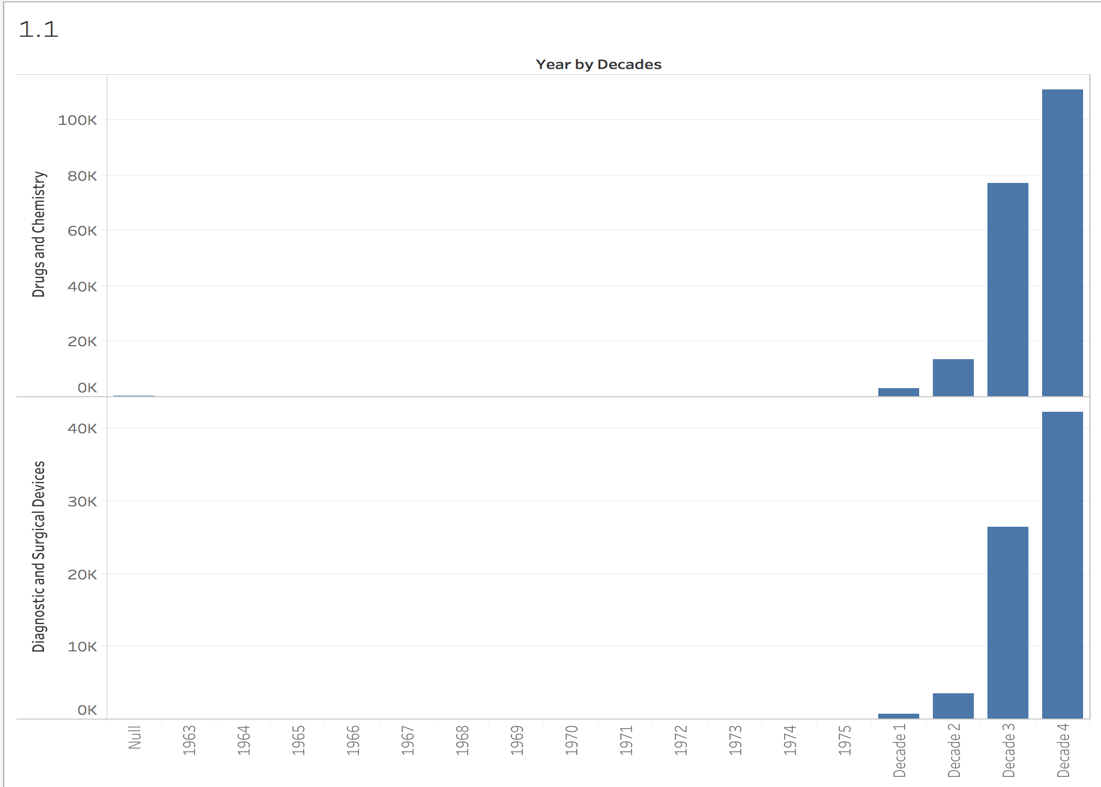
**Questions**

1. *Show the evolution of patent documents over time (column D) by category (column M through U). Also consider the evolution by decades (1976-1985, 1986-1995, 1996-2005, 2006-2016). Which categories do you think will continue to grow, or do you think it is not possible to answer?*

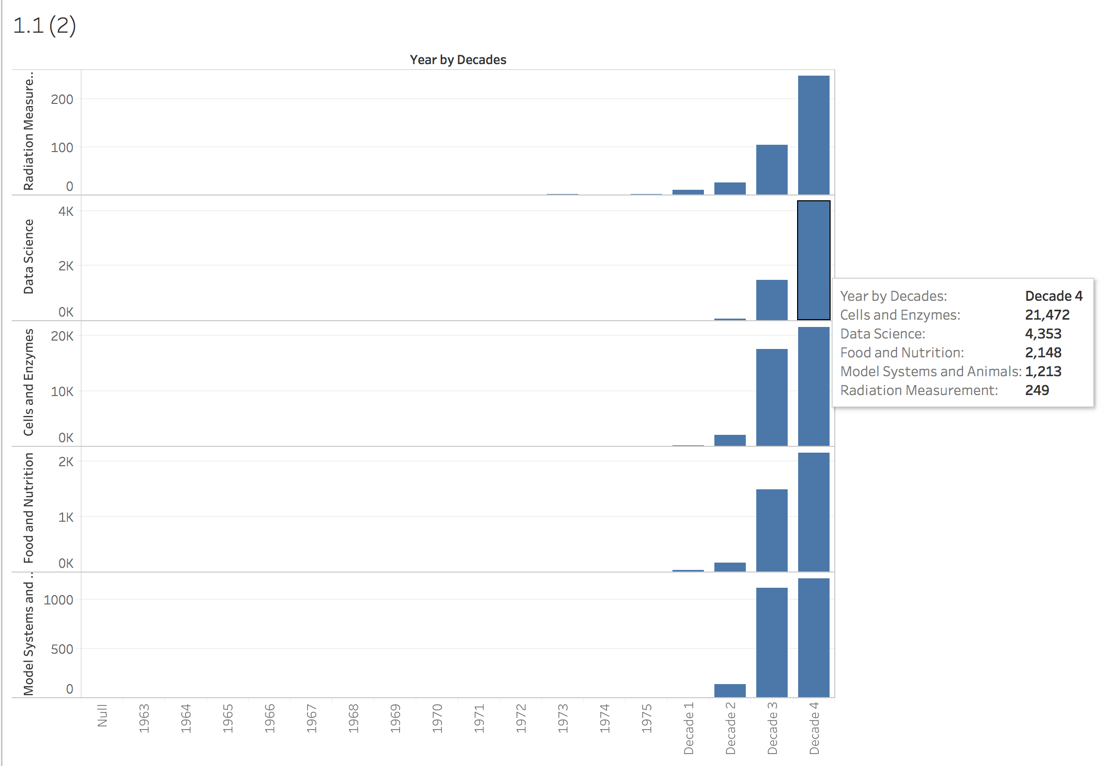
The following graph shows the patent document over. The largest positive slope is found between the years of 2000 and 2005. The colors of the charts are classified by decades. Red (1976-1985), Orange (1986-1995), Lime-Green (1996-2005) and Dark-Green (2006-2016). And they are determined as decade 1,2,3, and respectively. Therefore decade 4 shows the largest number of patents over time. There is a trend for continuing growth for all classifications. The largest slope is associated with Data science from decade 3 to decade 4.



*Chart 1. Evolution of all categories over time classified by colors – representing decades*



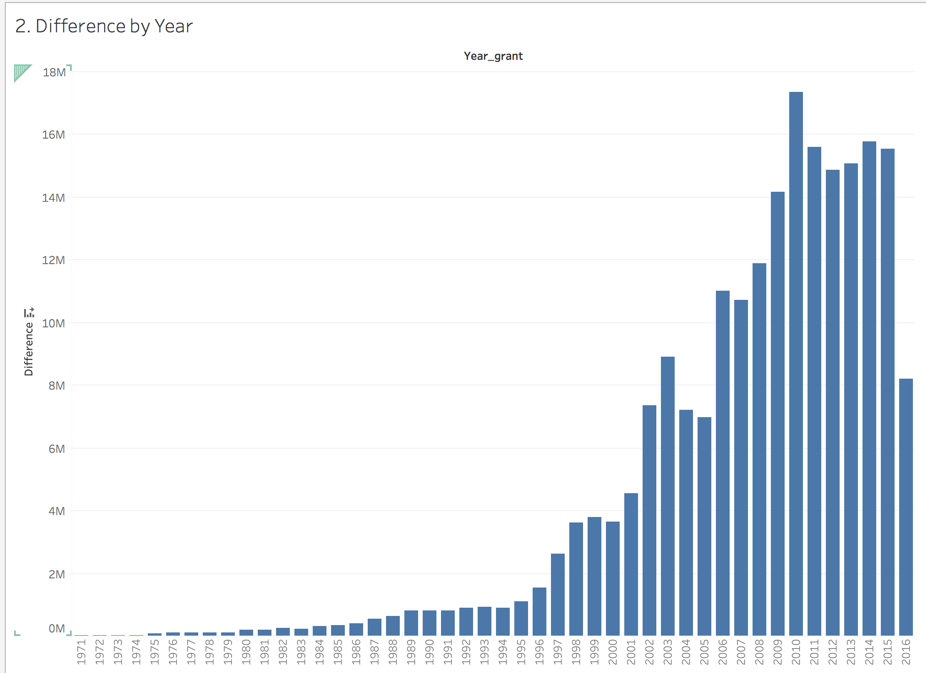
*Chart 2. Decade 4 shows the largest number of documents for the categories in Drug and Surgical Devices, as well for Drug Chemistry (this last one has over 110,000 patents in decade 4)*

**

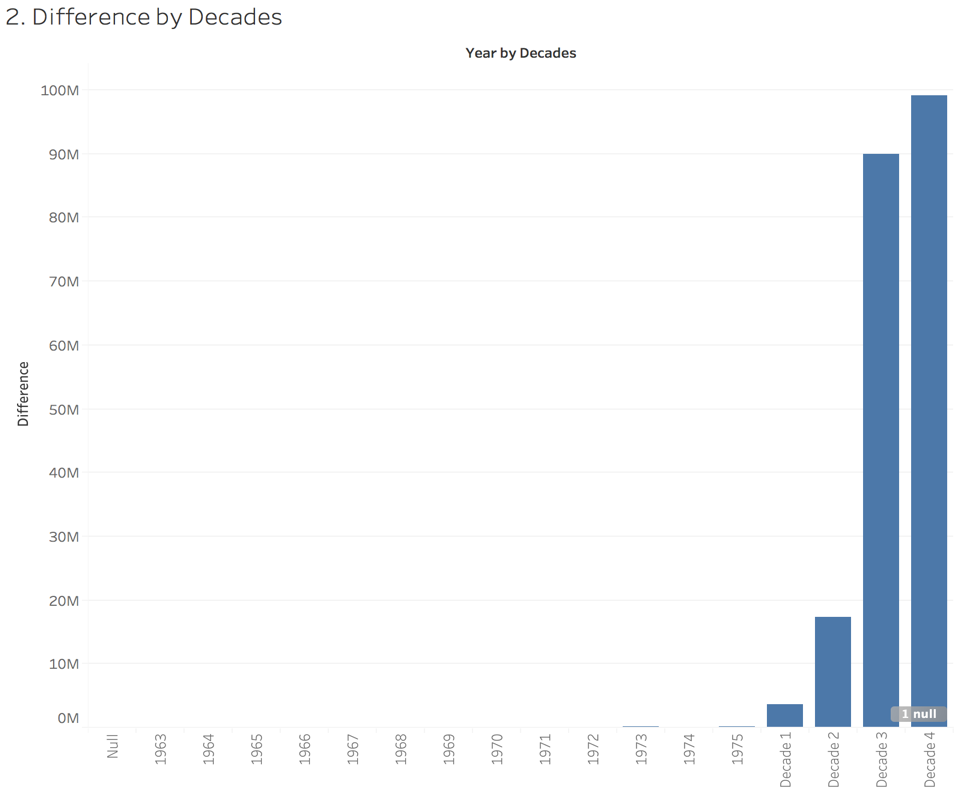
*Chart 3. This chart shows that Cells and Enzymes has 21,472 patent documentation. It also shows the category of Data Science – which is young – has an impressive number of patents (4,353)*

1. *Plot the distribution of the time it takes for a patent application to be granted (difference between filing data and grant/publication data, Columns D and E). Has this changed over the decades?*

The difference of grant date and filing date is the time it takes for an application to be granted. The following chart shows that the time of granted patents has increased drastically due to the fact that more patents have been developed.



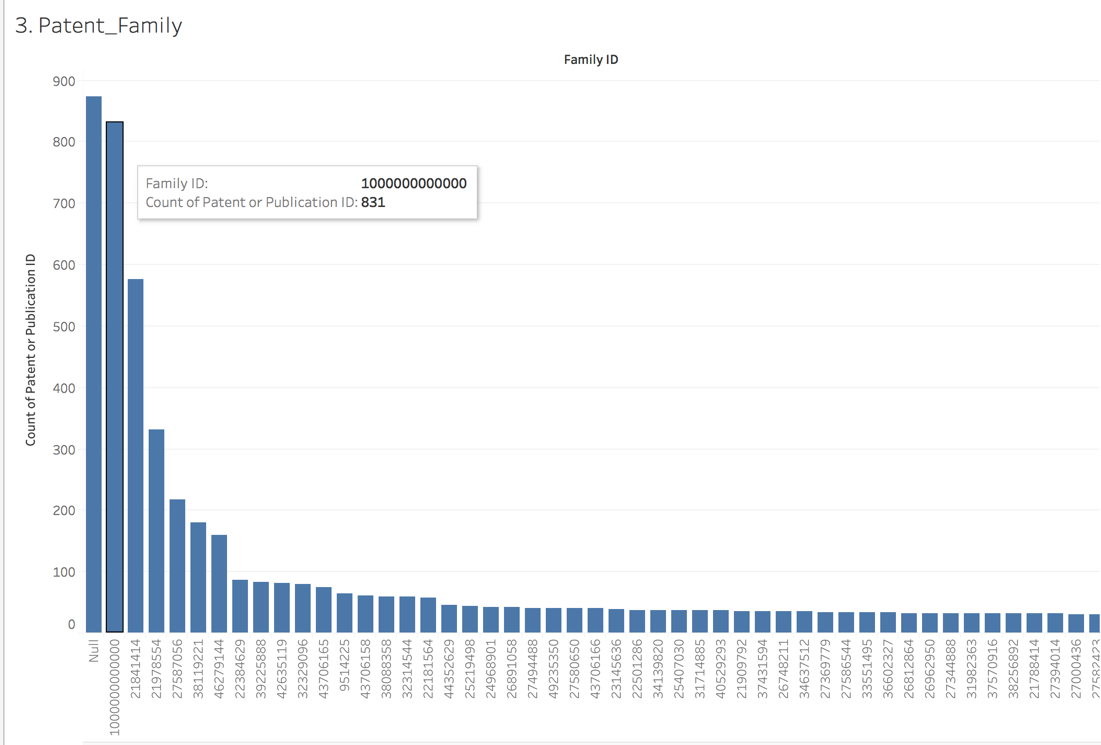
*Chart 4. Year 2010 shows the longest time for patents to be granted. Time has been reduced but not to the level of years prior to 1996.*

**

*Chart 5. This chart shows how Decade 4 is greater than the other decades in term of time for patents to be granted*

1. *Which patent families (Column A) have the most patents (Column B)?*

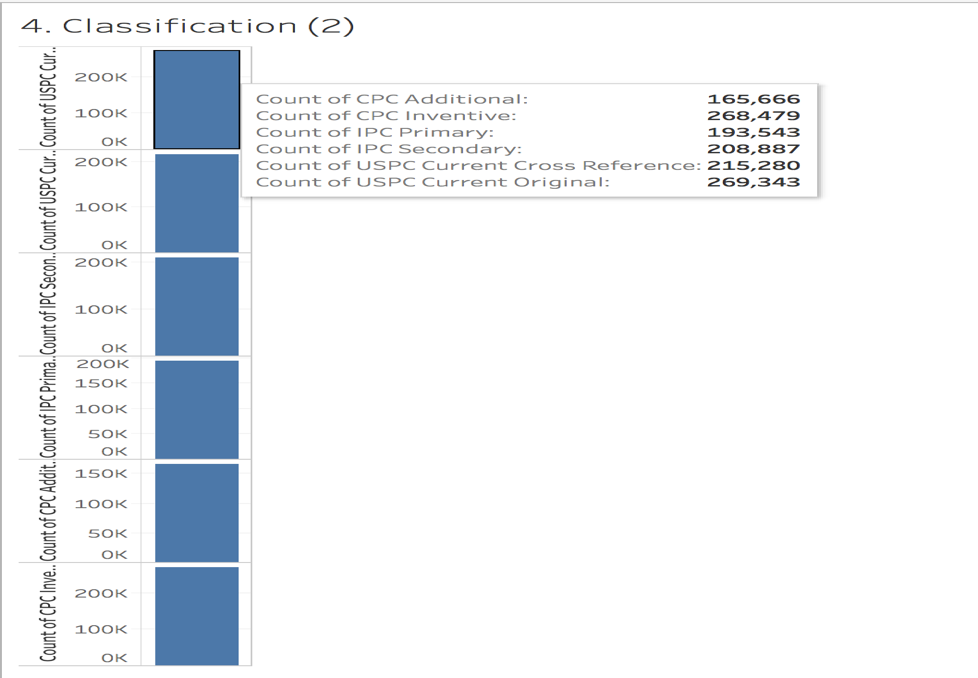
There are more patents publications without family ID (874). But the most patents publications are found family ID 1000000000000 at 831 patents, followed by family ID 21841414.



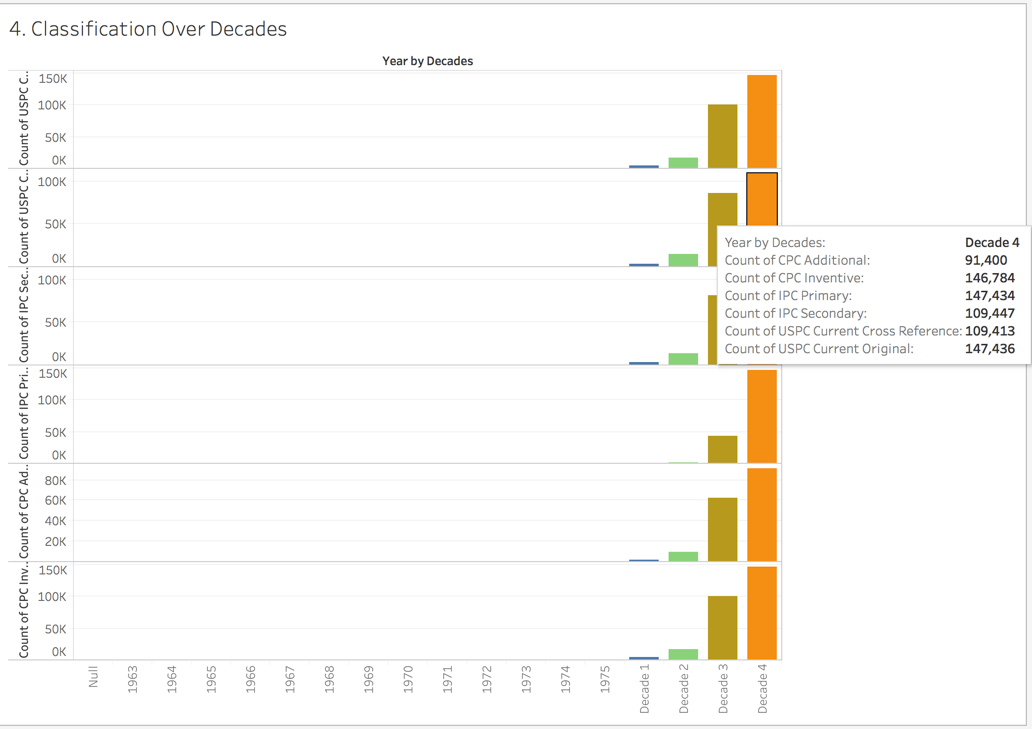
*Chart 6. Number of patents by family ID in descending order*

1. *Which classifications (columns F through K) appears the most often for the whole data set, and if you split the data by decades? Do you see any trend?*

The classification that appears the most often in the dataset is USPC Current Original with over 269,000 counts. Decade 4 (2006 – 2016) has overall the largest numbers, and the USPS Current Original also wins over the classifications in this decade.



*Chart 7. Counts of Classifications in the entire data*

**

*Chart 8. Decade 4 shows the larger counts of each of the classifications, where USPC Current Original has the most with 147,436*

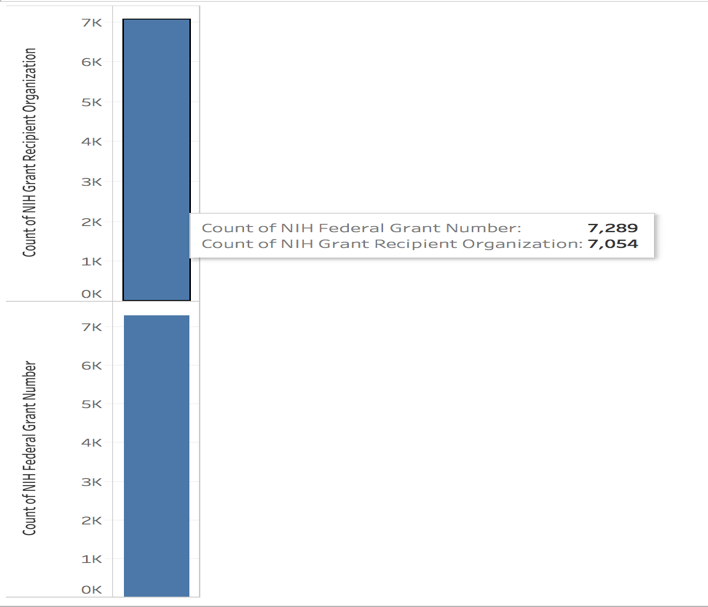
1. *How many patent documents have received NIH funding (Columns V and W)? In which categories were those? Which organizations have received NIH funding most often (as measured by patent documents)? In particular, which academic organizations? Research them online and discuss whether the academic organizations that appear most often in the file are the ones that are best known or not.*

There are categories for NIH funding. One is NIH Federal Grant number and the other is NIH Grant recipient organization. There has been 7,289 federal grant numbers and 7,054 grant recipients for all time.

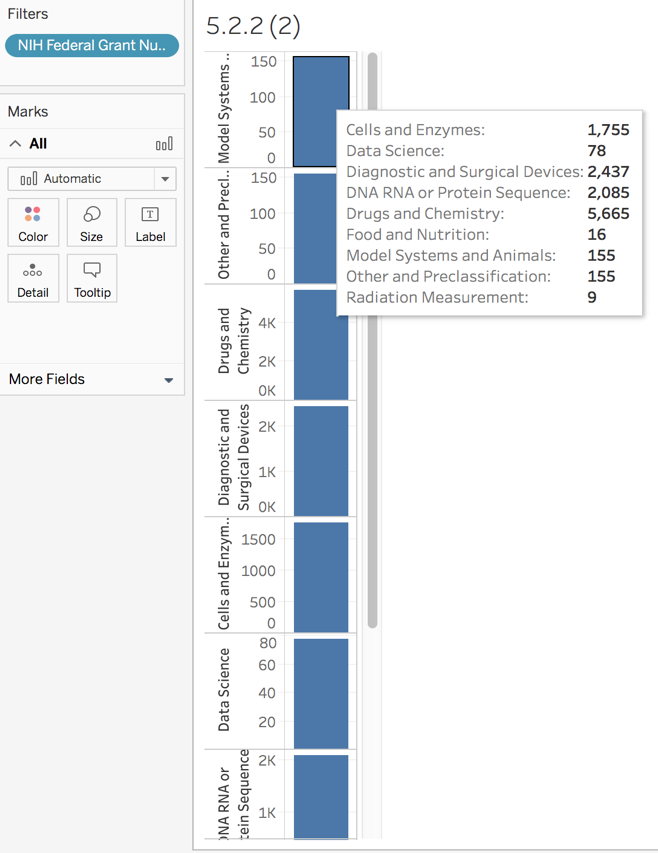
For categories that has received more NIH federal grant has been Diagnostic and Surgical Devices with 2,437 grants.

University of Wisconsin, Scripps Research, Standard, Columbia, and UCSD has received most often funding as measured by patent documents with a number, in particular. Wisconsin with 153 grants.

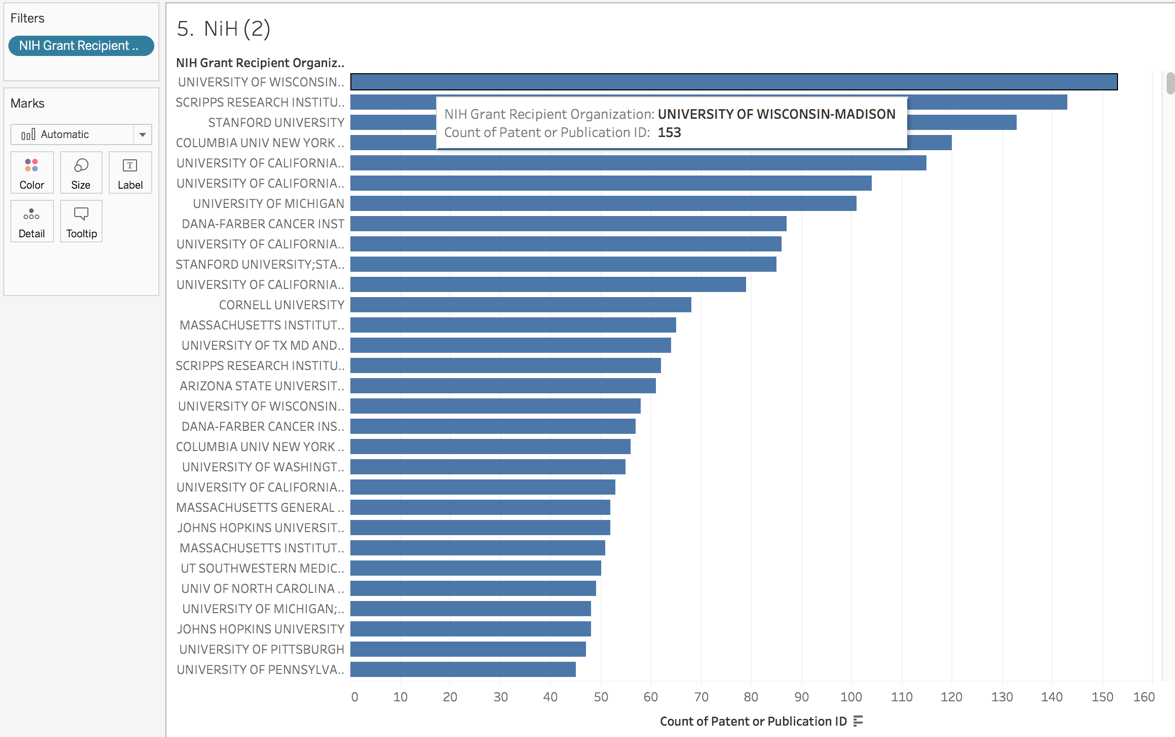
According to the UCSF Cancer research institute, JHU, UCSF, Michigan, Pennsylvania, Pittsburgh, Washington, Stanford, Duke, North Carolina, and Yale appears as the top NIH funding institution for 2016. Within this list, Stanford appears the dataset.



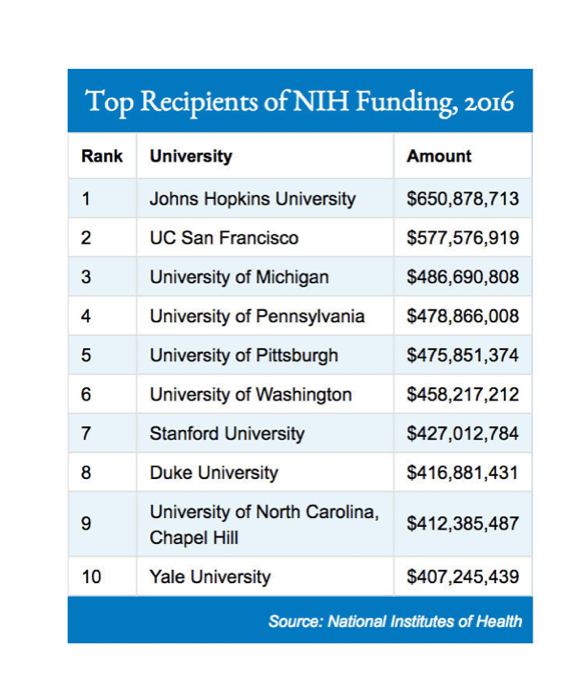
*Chart 9. The number of patents receiving Federal Grant (NIH funding) is 7,289 and 7,054 for Grant Recipient organizations*

**

*Chart 10. Diagnostic and Surgical Devices has received more Federal Grants that any other category*



*Chart 11. University of Wisconsin, Scripps Research and Stanford have received most funding by document patent respectively*



*Chart 12. This table shows the Top recipients of grants by the NIH in 2016. Not only Cancer research appears here but all the categories of science and medicine. Courtesy of ‘UCSF Is Top Public Recipient of NIH Research Funding‘*

1. *How many patent documents have received FDA approval (meaning columns X and beyond aren’t empty)? For the patents with FDA approval (those with data in columns X and beyond), which FDA drugs (column Y) have the most patents? What category do they belong to? What codes in column F through K do they have in common, if any (read the Word file on the Cancer Moonshot website for descriptions) or not?*

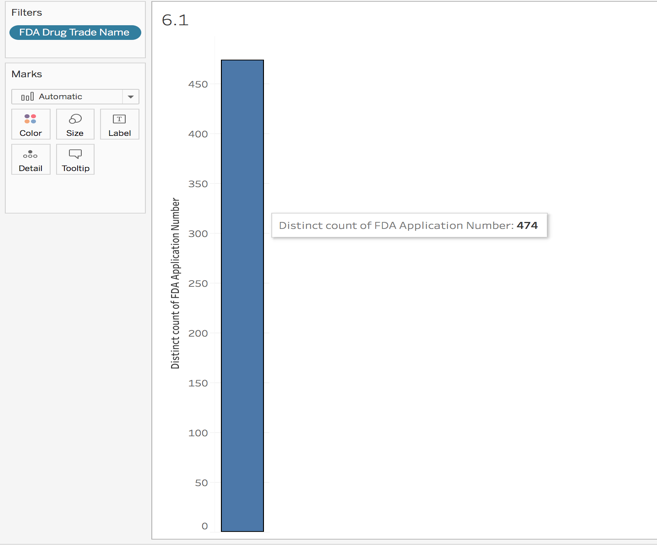
474 patents received FDA approval from the period of 1976 to 2016

ESBRIET has the most patents with 19 publications.

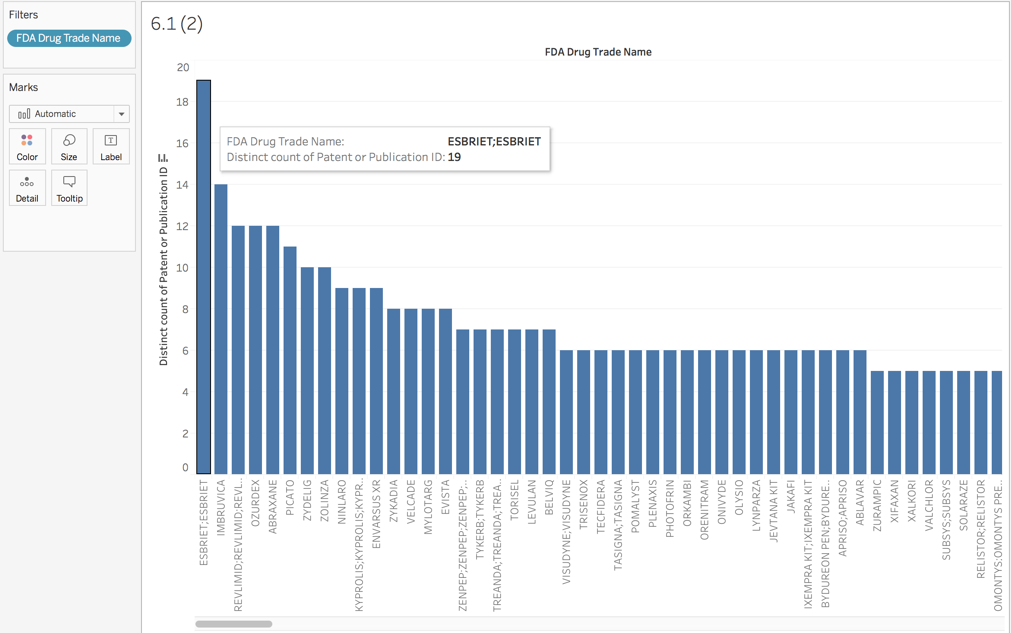
CPC Additional Code (Cooperative Patent Classification) = A61K2300/00, with 22 applications

According to the site patent Base (patbase.com)

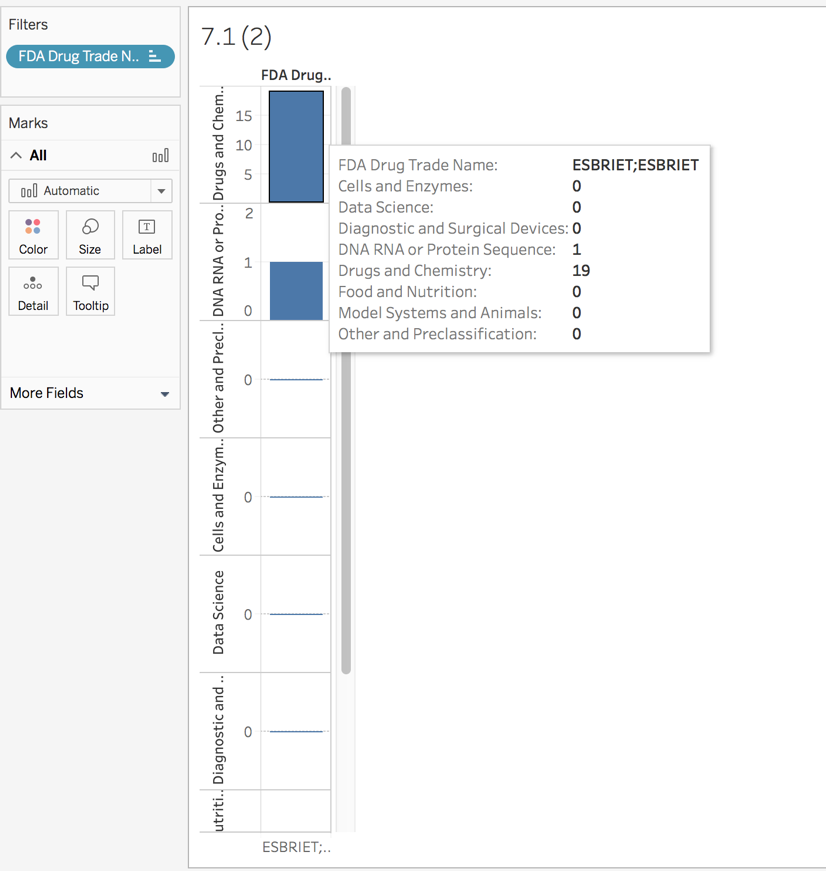
|  |  |  |
| --- | --- | --- |
| **A**: | HUMAN NECESSITIES | [http://www.patbase.com/images/anal2_bar.png](http://www.patbase.com/classSnapshot/public/?class=A61K2300/00&system=CPC) |
| **A61**: | HEALTH; AMUSEMENT; MEDICAL OR VETERINARY SCIENCE; HYGIENE | [http://www.patbase.com/images/anal2_bar.png](http://www.patbase.com/classSnapshot/public/?class=A61K2300/00&system=CPC) |
| **A61K**: | PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES (devices or methods specially adapted for bringing pharmaceutical products into particular physical or administering forms A61J3/00; chemical aspects of, or use of materials for deodorization of air, for disinfection or sterilization, or for bandages, dressings, absorbent pads or surgical articles A61L{; compounds per seC01, C07, C08, C12N; }; soap compositions C11D{; microorganisms per seC12N}) | [http://www.patbase.com/images/anal2_bar.png](http://www.patbase.com/classSnapshot/public/?class=A61K2300/00&system=CPC) |
| **A61K2300/00**: | Mixtures or combinations of active ingredients, wherein at least one active ingredient is fully defined in groups A61K31/00�-�A61K41/00 |  |



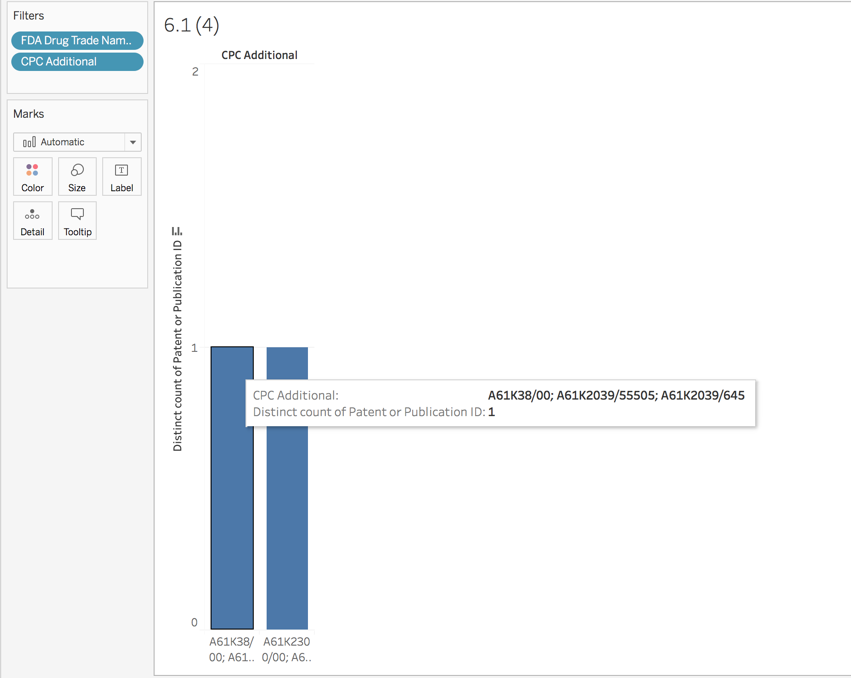
*Chart 13. The count distinct of FDA applications numbers with FDA Drug Trade Name*



*Chart 14. FDA Drug Name with most patent/publications. ESBRIET has 19 and has the most*

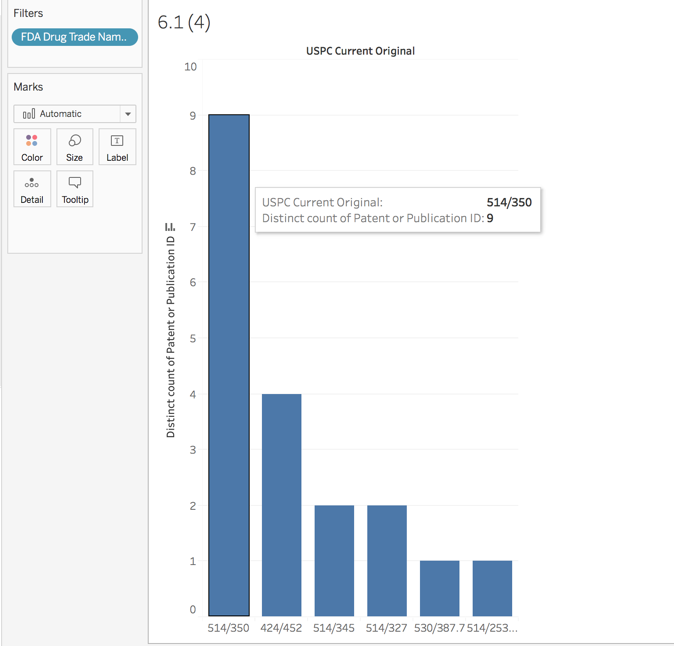


*Chart 15. The category of Drug and Chemistry has the most Trade Name of ESBRIET with 19 patents*



*Chart 16.1. 22 publications are found with the CPC code A61K38/00 with the ESCRIET Trade name*

*(A61K38/00: Medicinal preparations containing peptides (peptides containing beta-lactam rings A61K31/00; cyclic dipeptides not having in their molecule any other peptide link than those which form their ring, e.g. piperazine-2,5-diones, A61K31/00; ergot alkaloids of the cyclic peptide type A61K31/48; containing macromolecular compounds having statistically distributed amino acid units A61K31/74; medicinal preparations containing antigens or antibodies A61K39/00; medicinal preparations characterized by the non-active ingredients, e.g. peptides as drug carriers, A61K47/00)*

**

*Chart 16.2. 22 publications are found with the USPC 514/350 with the ESCRIET Trade name*

*(514/350 C=O bonded directly to the six-membered hetero ring:*

*This subclass is indented under subclass 345. Subject matter wherein a C=O group is bonded directly to the six-membered hetero ring)*

1. *For the FDA Drugs with the most patents in Column Y, research revenue estimates online (price, estimated market size, estimated revenue per year, any financial data you can find online, if any). Which categories (column through U) do you think are the most lucrative and do you see that companies tend to focus on the most lucrative categories or not?*

According to Bloomberg, Esbriet (Pirfenidone), a treatment for [**idiopathic pulmonary fibrosis**](https://pulmonaryfibrosisnews.com/tag/idiopathic-pulmonary-fibrosis/), manufactured by [**Boehringer Ingelheim GmbH**](https://pulmonaryfibrosisnews.com/tag/boehringer-ingelheim/), cost $96,000 per year.

270 capsules can cost up to $10,000.

Within Categories, Drug and Chemistry spend more resources on this type of treatment.

Various corporations use their resources to produce the most expensive Cancer drugs. Here are some examples:

No. 1: Blincyto, $64,260 monthly cost per patient – Amgen’s

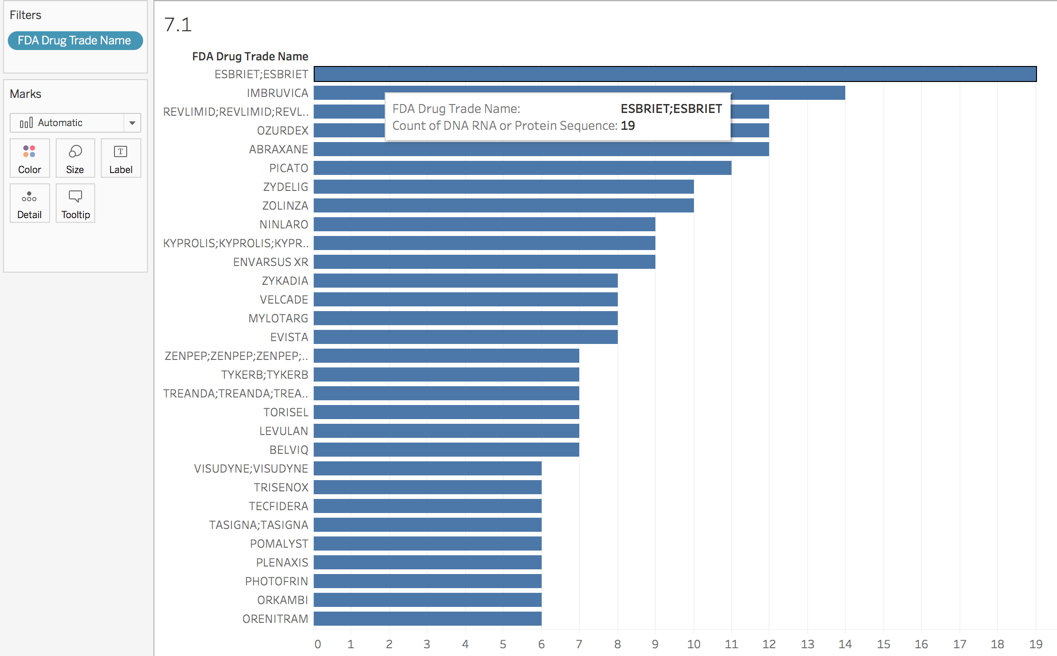
No. 2: Lenvima, $13,945 monthly cost per patient – Eisai’s

No. 3: Zykadia, $13,672 monthly cost per patient - Novartis AG’s

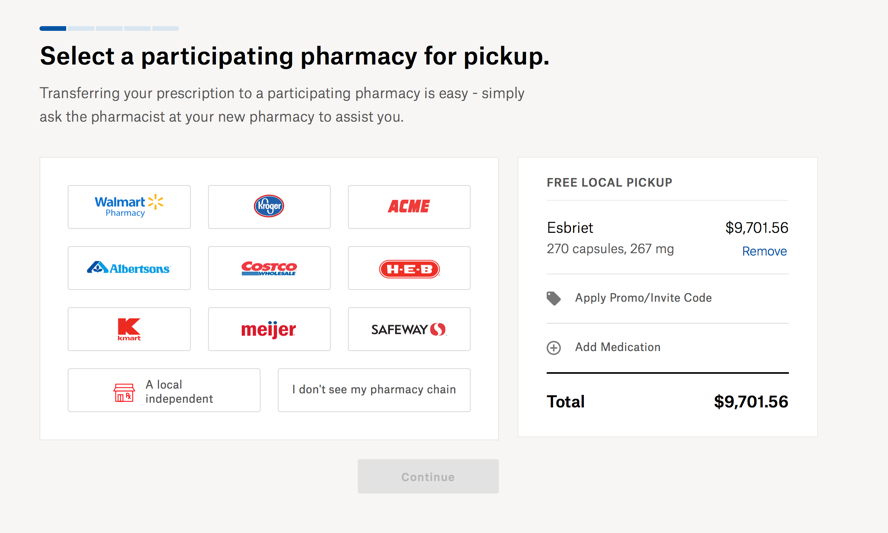
No. 4: Cyramza, $13,256 monthly cost per patient - Eli Lilly's

No. 5: Xofigo, $12,657 monthly cost per patient - Bayer AG's

*Courtesy of The Motley Fool.*



*Chart 17. ESBRIET has the most publications*



*Chart 18. Cost of 270 Esbriet capsulate*

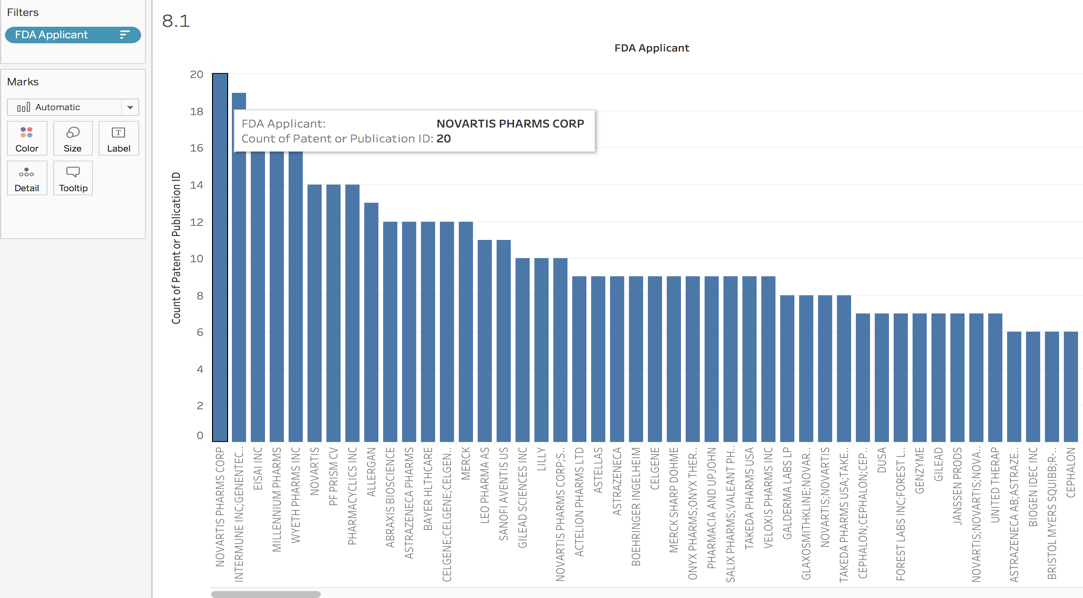
1. *For the patents with FDA approval (those with data in columns X and beyond), which companies (column AA) have the most patent documents and how have those numbers of patent documents evolved over time? For the companies with the largest number of patents (define largest), how are those patents diversified across categories (column M through U)? Which company do you think has the strongest patent portfolio?*

Novatis has the most patent approved by the FDA with 20. Followed by Intermune Inc.

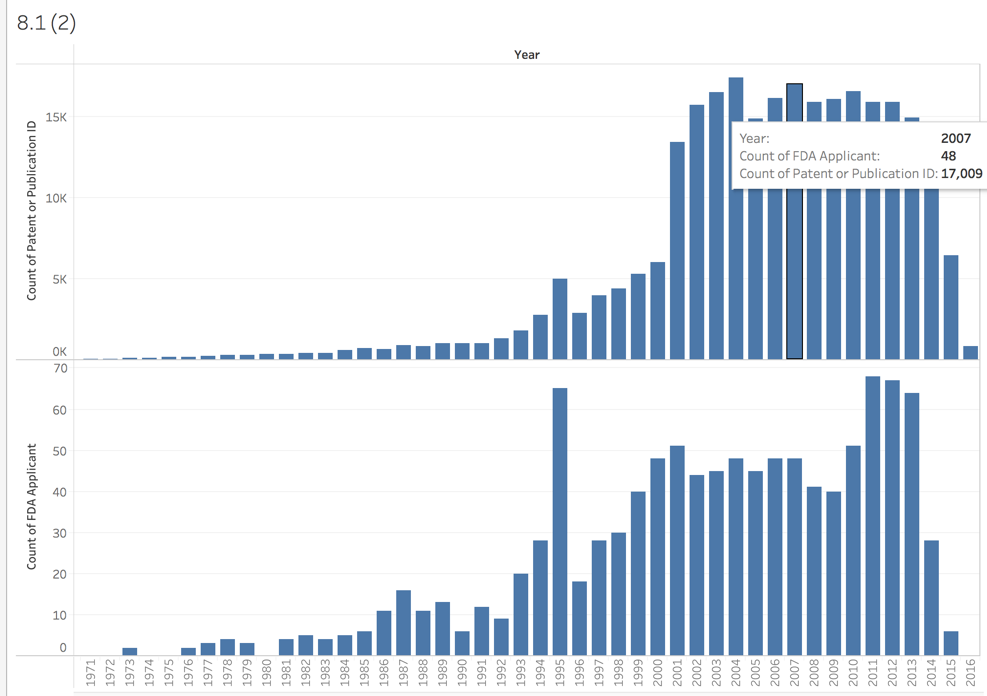
In terms of patent documentation, it has evolved over time, peaking in 2012.

The Top 5 FDA applicants are (largest): Novatis, Intermune Inc, Eisai, Millennium Pharms, Wyeth Pharms.

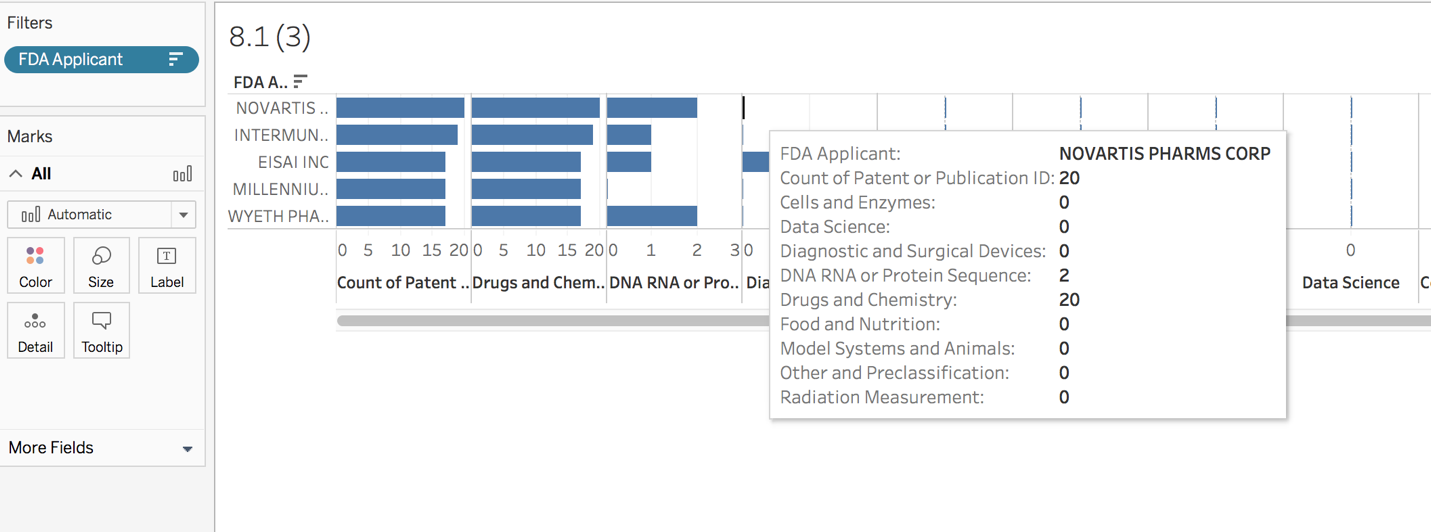
Novatis has the strongest patent portfolio and the most diversified one as well.



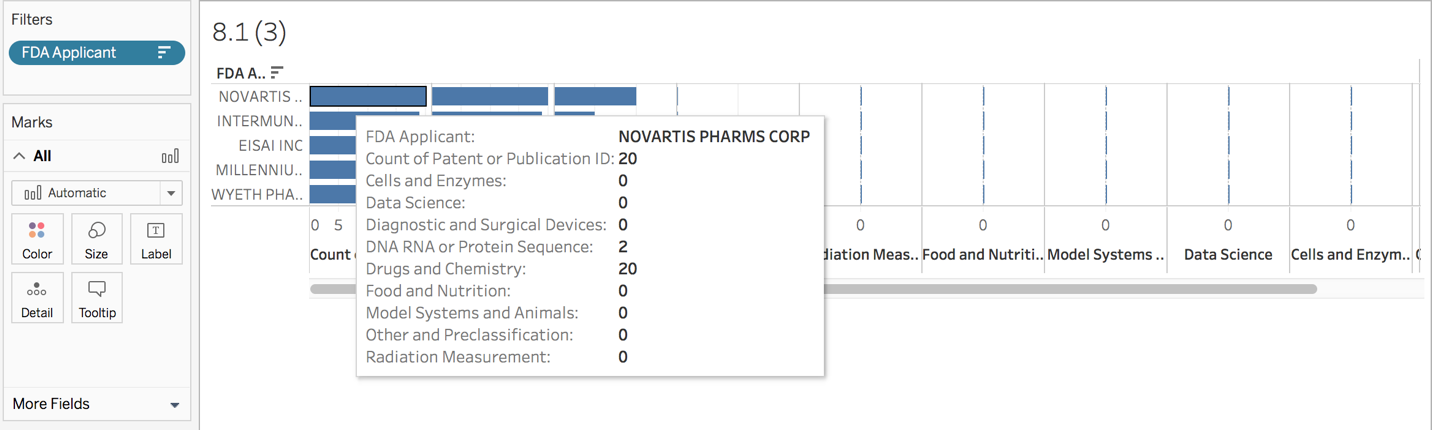
*Chart 19. FDA Applicants ranking*



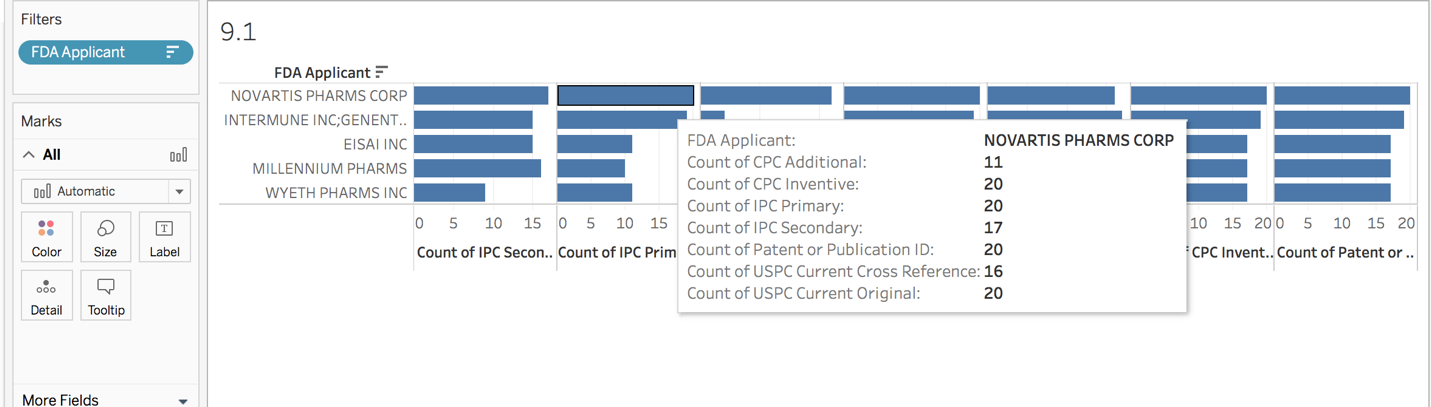
*Chart 20. Patent documents over time – peaking in 2012*



*Chart 21. Top 5 FDA Applicants spent resources in Drug and Chemistry because its profitability*

**

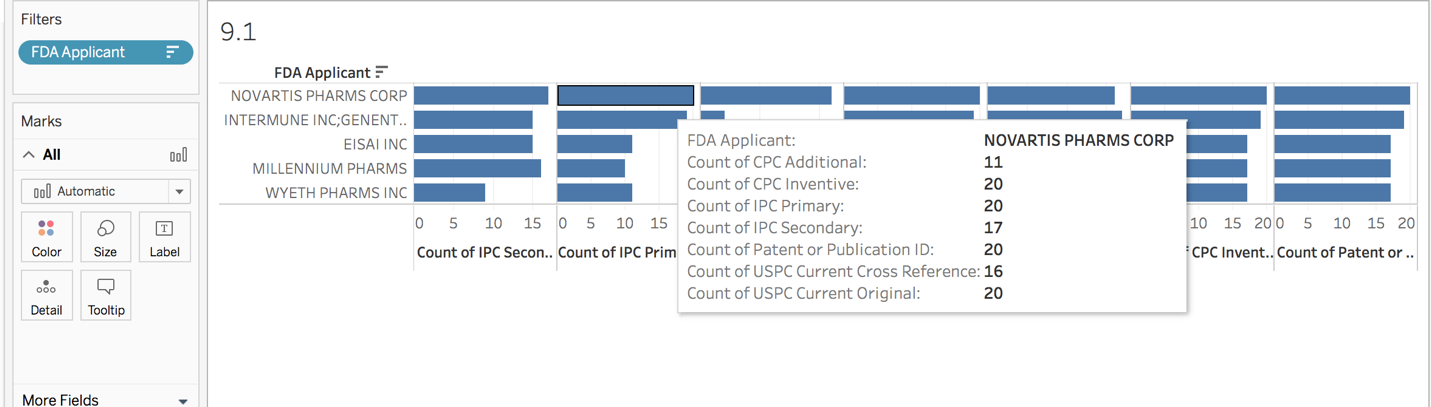
*Chart 22. Novatis Pharms has the strongest portfolio in diversification. 20 patents in Drug and Chemistry and only 2 in DNA RNA or Protein Sequence*

**

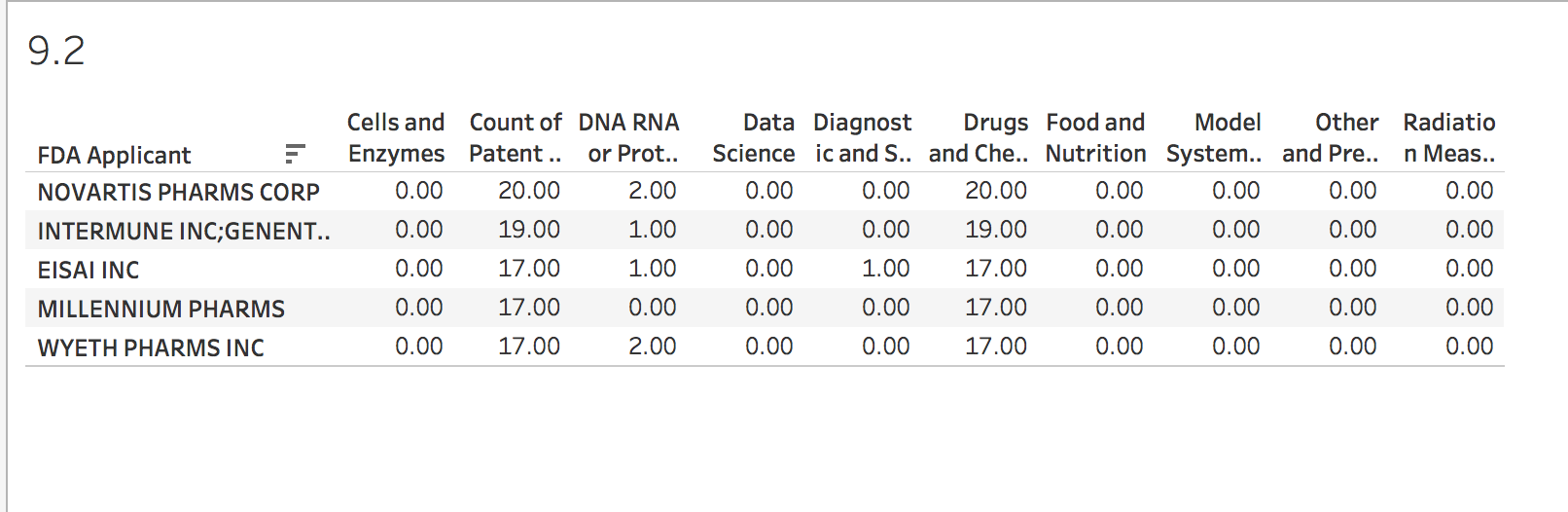
*Chart 23. Novatis Pharms has the strongest patent portfolio*

1. *Do companies that have received FDA approval have differentiated strategies for the war against cancer, meaning, does their portfolio of patents focus more on certain categories (F through K or M through U)?*

Drug and Chemistry is the largest category companies focused on because of profitability. They focus primarily in this category than any other category.

**

*Chart 25 Novatis Pharms has the strongest patent portfolio*

*Chart 26 All Top 5 companies concentrated their research in the Drug and Chemistry Category*

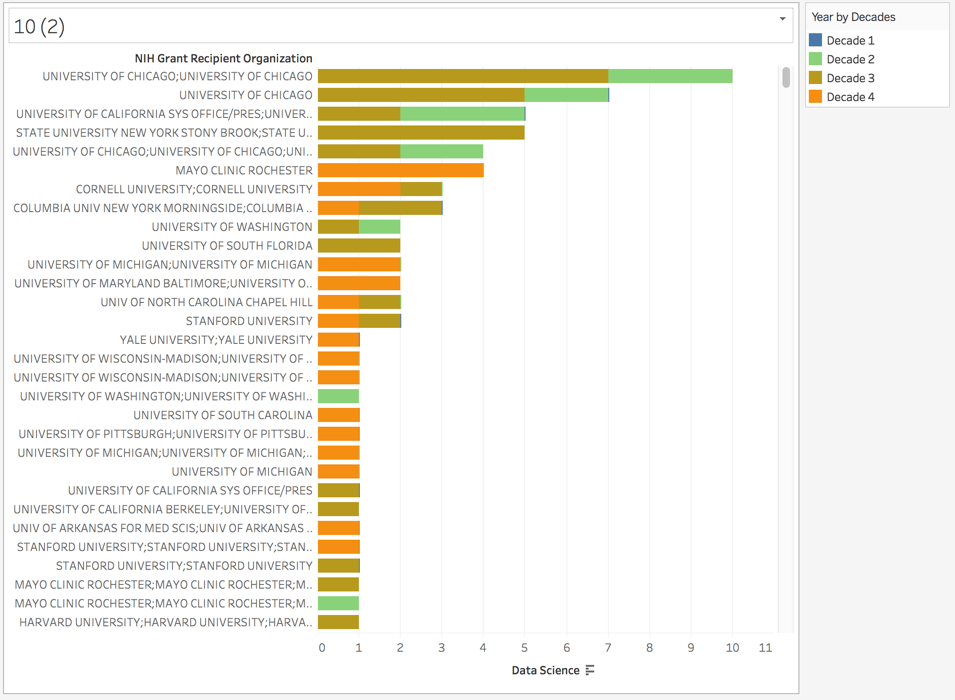
1. *Analyze the dataset in some interesting way not mentioned above.*

There has been a surprising increased in Radiation measurement therapy in the last decade compared to Drug and Chemistry treatment.



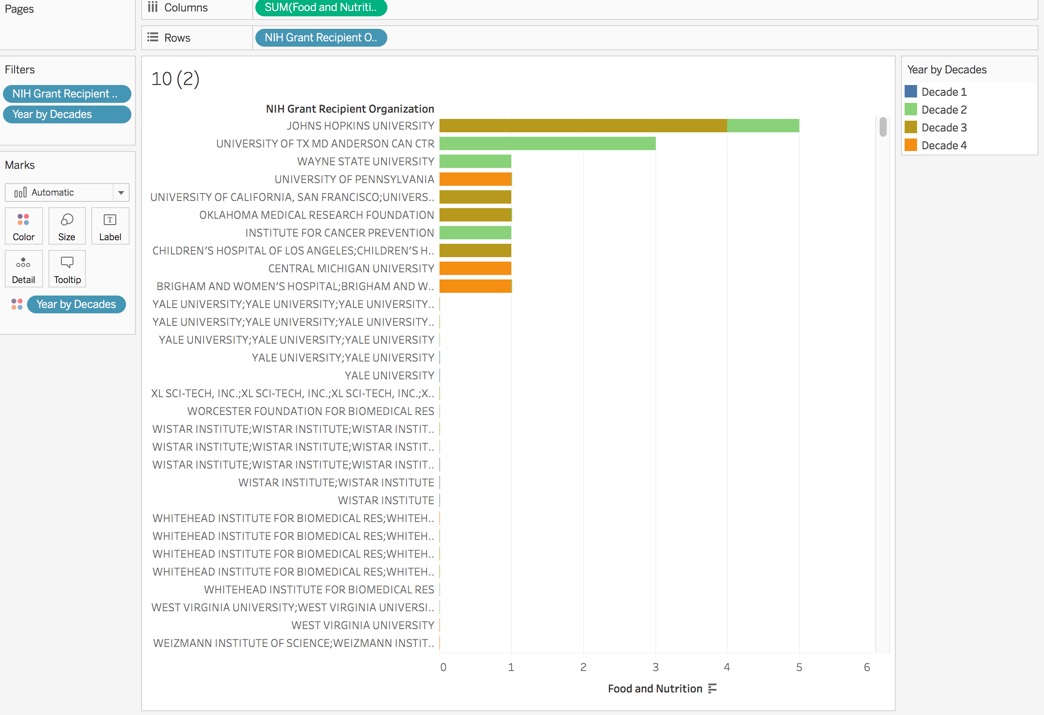
*Chart 27 Drug & Chemistry vs. Radiation Measurement*

University of Chicago has been the leader in Data Science treatment where the NIH has granted its research. However, in the last decade, the Mayo Clinic of Rochester has received more NIH grant in this field that any other institution.



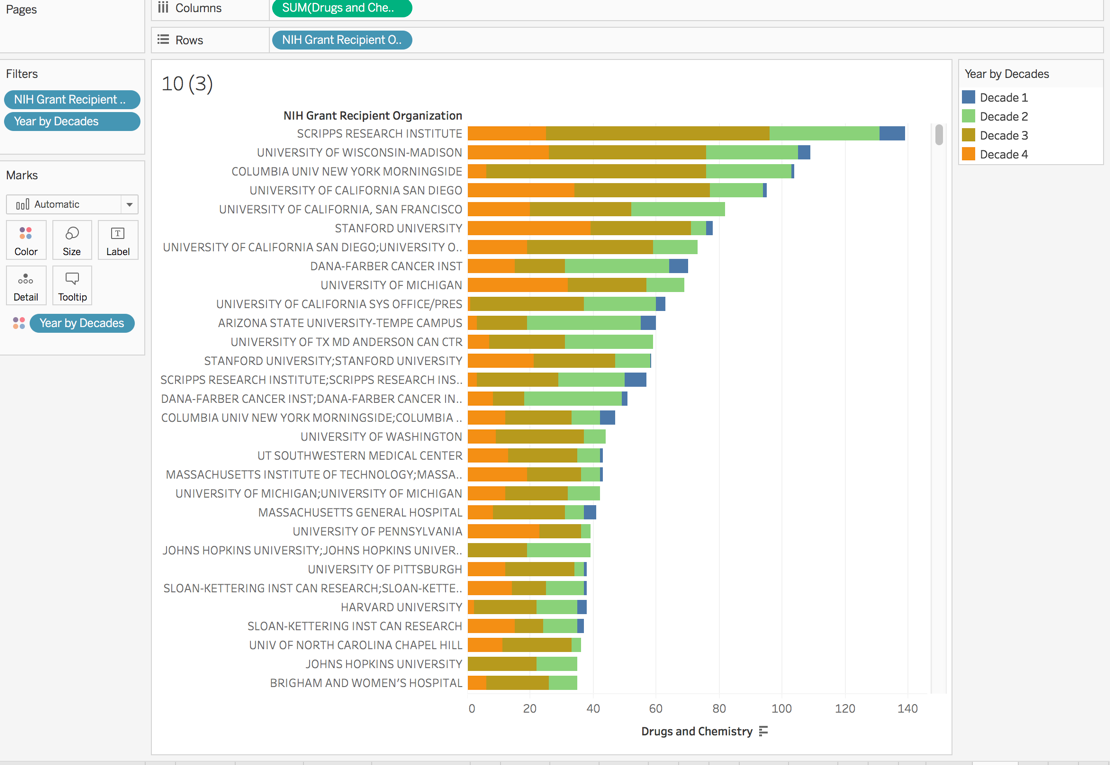
*Chart 28 University of Chicago at best in Data Science research*

On the other hand, John Hopkin University has been in the top whenever Food and Nutrition is about.



*Chart 29. John Hopkin University and the MD Anderson are the champions at Food and Nutrition NIH grant*

And lastly, it can be observed that overall, Universities spend more resources on the category of Drug and Chemistry by observing the number of institutions engaged in this field.



*Chart 30. A Large number of institutions spend enormous resources on Drug and Chemistry research*

**Conclusions**

In this analysis, it was concluded that top institutions in cancer research were identified by selecting the institutions with the highest number of NIH grant recipients and whose patents have been published. This metric partly reflects volume of research for forty-year data, and partly the reach and quality of that output. University of Wisconsin, Scripps Research and Stanford have been in the vanguard of the cutting-edge research by being ranked as the highest research institution respectively. It was also observed that federal grant number P01DK014881 had the largest number of publications. This research category is associated to metabolism and function of fat-soluble vitamins, which was awarded to the University of Wisconsin, and to the project leader named Hector DeLuca.

Part of this investigation also determined that decade 4 (2007 – 20016) had the greatest numbers of time lapse from the moment a patent was filed to when it was granted. This appears to the related to a delay in the process patent applications are developed, and also to the fact that more patents were filed during the decade 4. Moreover, one of the specific insights that can be identified is that although “Drugs and Chemistry” was the #1 category for the overall patent data set, NIH funded patents, and FDA approved patents, the other categories varied significantly in their ranking, implying that the overall industry, NIH, and FDA pursued different priorities in choosing patent filings to approve and fund.

Reviewing these visualizations enables NIH decision makers to gain a better understanding of where cancer research is currently going on. In conclusion, visualizing patent data can be of significant help in understanding where cancer research is currently going on, where FDA approval has been obtained, and where there is a future need for increased funding by NIH.

**References**

Robbins, R. [He vowed to cure cancer. But this billionaire’s moonshot is falling far short of the hype](https://www.statnews.com/2017/02/14/moonshot-soon-shiong-investigation/). STAT, FEB. 14, 2017

Harris, G and Koalata, G. [‘Moonshot’ to Cure Cancer, to Be Led by Biden, Relies on Outmoded View of Disease](https://www.nytimes.com/2016/01/14/health/moonshot-to-cure-cancer-to-be-led-by-biden-relies-on-outmoded-view-of-disease.html), New York Times, Jan. 13, 2016

Research Portfolio Online Reporting Tool. US Department of Health and Human Services. <https://projectreporter.nih.gov/project_info_description.cfm?aid=6050729&icde=0>

DeLuca H. METABOLISM AND FUNCTION OF FAT-SOLUBLE VITAMINS. University of Wisconsin–Madison, United States. Available at: <https://app.dimensions.ai/details/grant/grant.2436052>

Cancer Research Current Trends & Future Directions. Elsevier. Available at: <https://www.elsevier.com/__data/assets/pdf_file/0019/230374/Elsevier-CancerResearchReport.pdf>

Discovering Insights in the Cancer Moonshot Patent Data Set. USPTO. Available at: <https://developer.uspto.gov/visualization/discovering-insights-cancer-moonshot-patent-data-set#comment-form>

Report of the Cancer Moonshot Task Force. The State of the Union. The White House. October 17, 2016. Available at: <https://obamawhitehouse.archives.gov/sites/default/files/docs/final_cancer_moonshot_task_force_report_1.pdf>