

**Southern Methodist University**

**EMIS 7357 Fall 2018**

**Analytics for Decision Support Assignment 3**

**Prediction of Advertisements Click-Through Rate**

By

BEICHEN HU

Student ID: 47809766

October 31, 2018

**Executive Summary**

The purpose of this report is to use analytic method bearing out the project which is a large data set made by The US Patent and Trademark Office (USPTO) for the “Cancer Moonshot”. By using the data from the project, it gives some solution for different requirements by this assignment. This assignment use Tableau to create different graph together. Therefore, it is easier to compare with the different values and categories in one graph and select the best one as the answer.

Table of Contents

[1. The evolution of patent documents 3](#_Toc528884185)

[1.1. The evolution of patent documents over time 3](#_Toc528884186)

[1.2. The evolution of patent documents by decades 4](#_Toc528884187)

[2. Patent Application Grant Time 5](#_Toc528884188)

[2.1. The distribution of the years 5](#_Toc528884189)

[2.2. The distribution of the decades 6](#_Toc528884190)

[3. Patent family 7](#_Toc528884191)

[4. Classification Appeared in Most 8](#_Toc528884192)

[5. NIH Funding 21](#_Toc528884193)

[5.1. The number of patent documents in NIH Funding 21](#_Toc528884194)

[5.2. The categories of NIH Funding 22](#_Toc528884195)

[5.3. The organization receive NIH the most often 24](#_Toc528884196)

[5.4. University of Wisconsin-Madison 24](#_Toc528884197)

[6. FDA Approval 26](#_Toc528884198)

[6.1. The number of FDA Approval 26](#_Toc528884199)

[6.2. FDA drugs the most often 27](#_Toc528884200)

[6.3. The common code 29](#_Toc528884201)

[7. ESBRIET 31](#_Toc528884202)

[7.1. Esbriet (Pirfenidone) 31](#_Toc528884203)

[7.2. Financial data of Esbriet 31](#_Toc528884204)

[7.3. The most lucrative category 32](#_Toc528884205)

[8. FDA Analysis 34](#_Toc528884206)

[8.1. The company having the most patent documents 34](#_Toc528884207)

[8.2. FDA companies’ categories 40](#_Toc528884208)

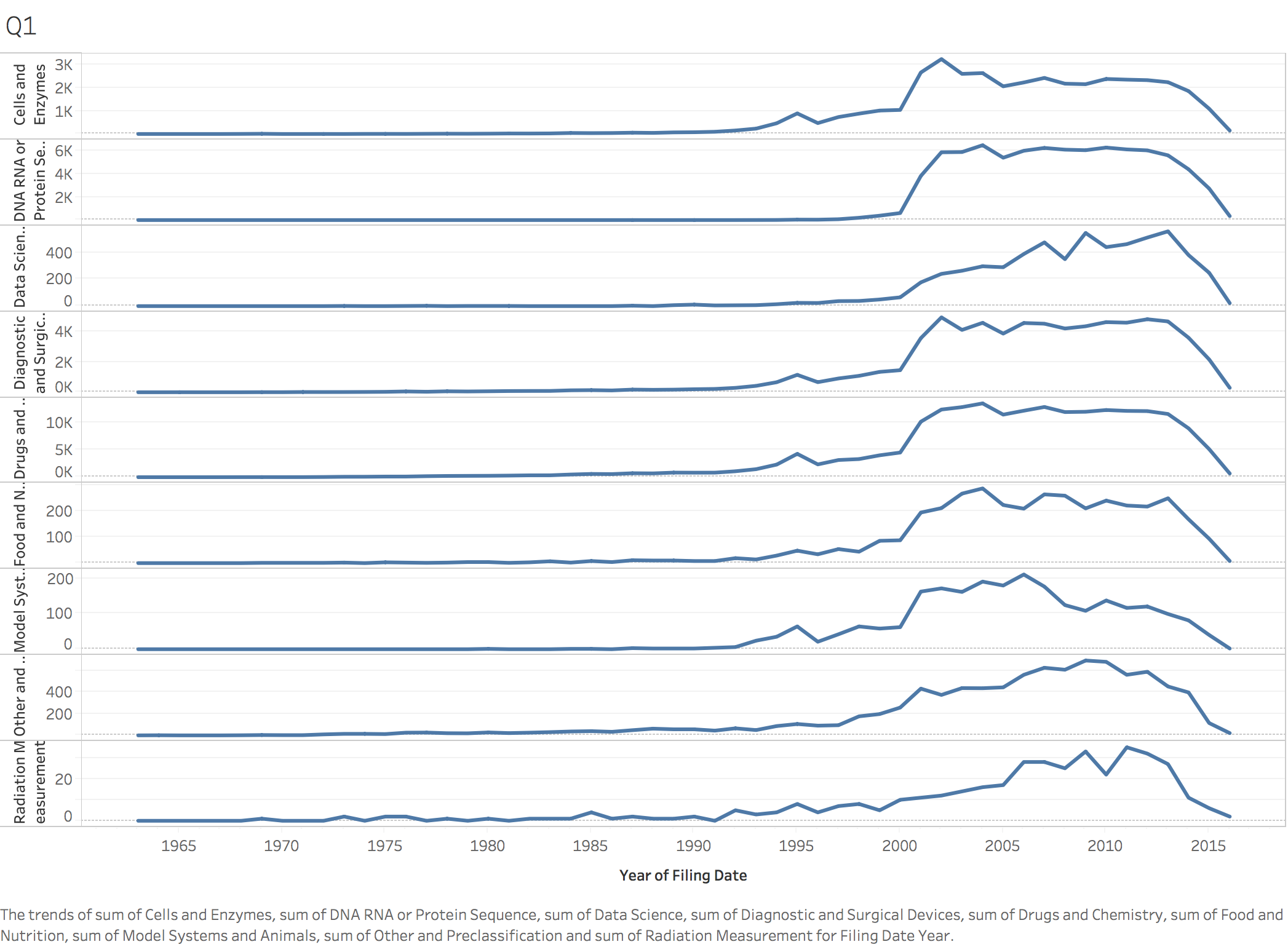
[9. FDA Approval different strategies 45](#_Toc528884209)

[10. The title of patent document 47](#_Toc528884210)

[Reference 49](#_Toc528884211)

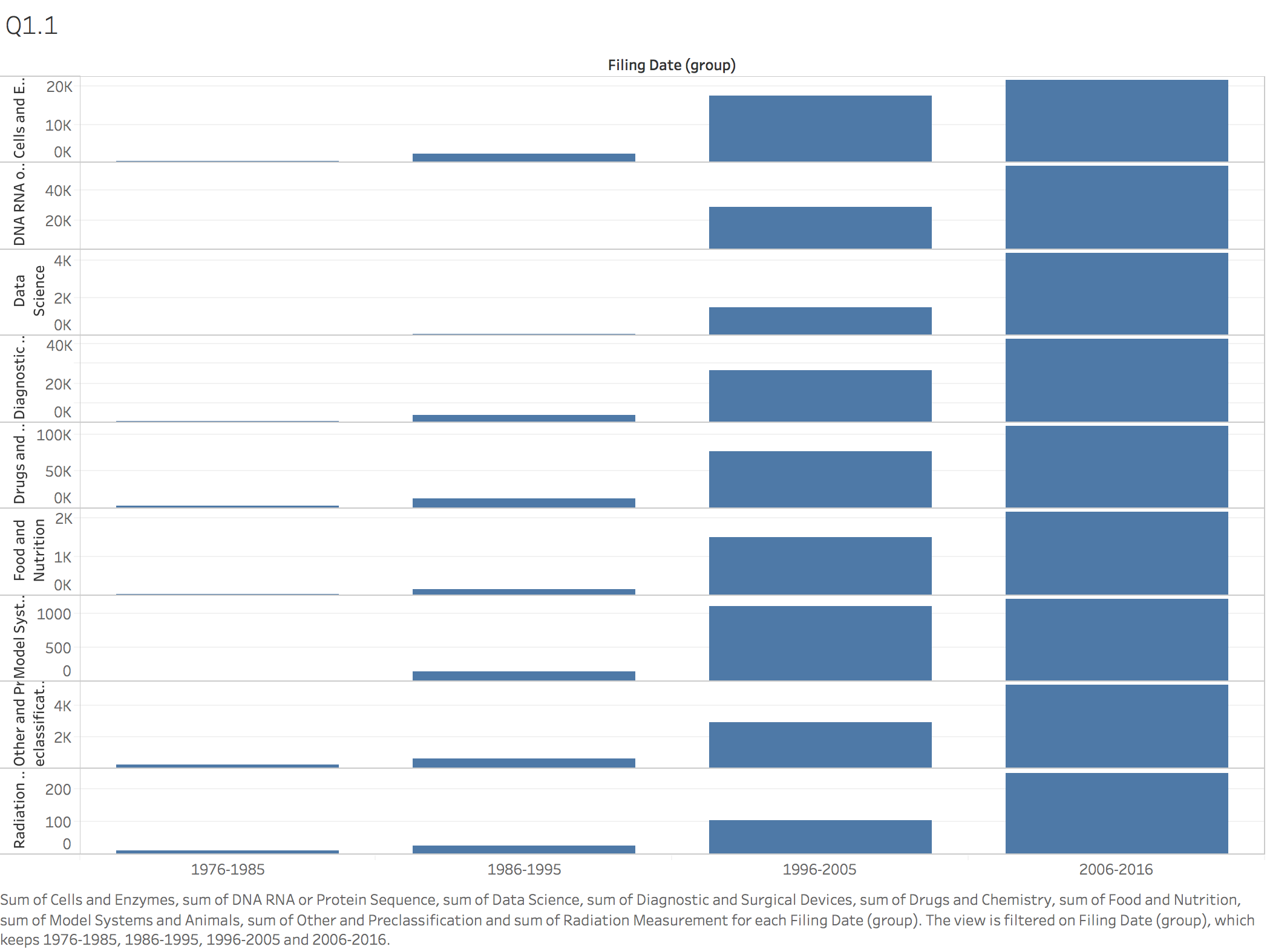
# The evolution of patent documents

## **The evolution of patent documents over time**



According to the different categories this picture shows about the evolution of the patent documents over time. Therefore, we can see that, before 2000, the number of patent documents were still low and depressive. However, after 2000, all categories of the patent documents became prosperous including the drugs and chemical, data science, Diagnostic and Surgical Devices, Radiation Measurement, Food and Nutrition, Model Systems and Animals, Cells and Enzymes, Other and Pre-classification, DNA RNA or Protein Sequence. But at recent time, in 2013, the trend of patent documents fell down in all categories again.

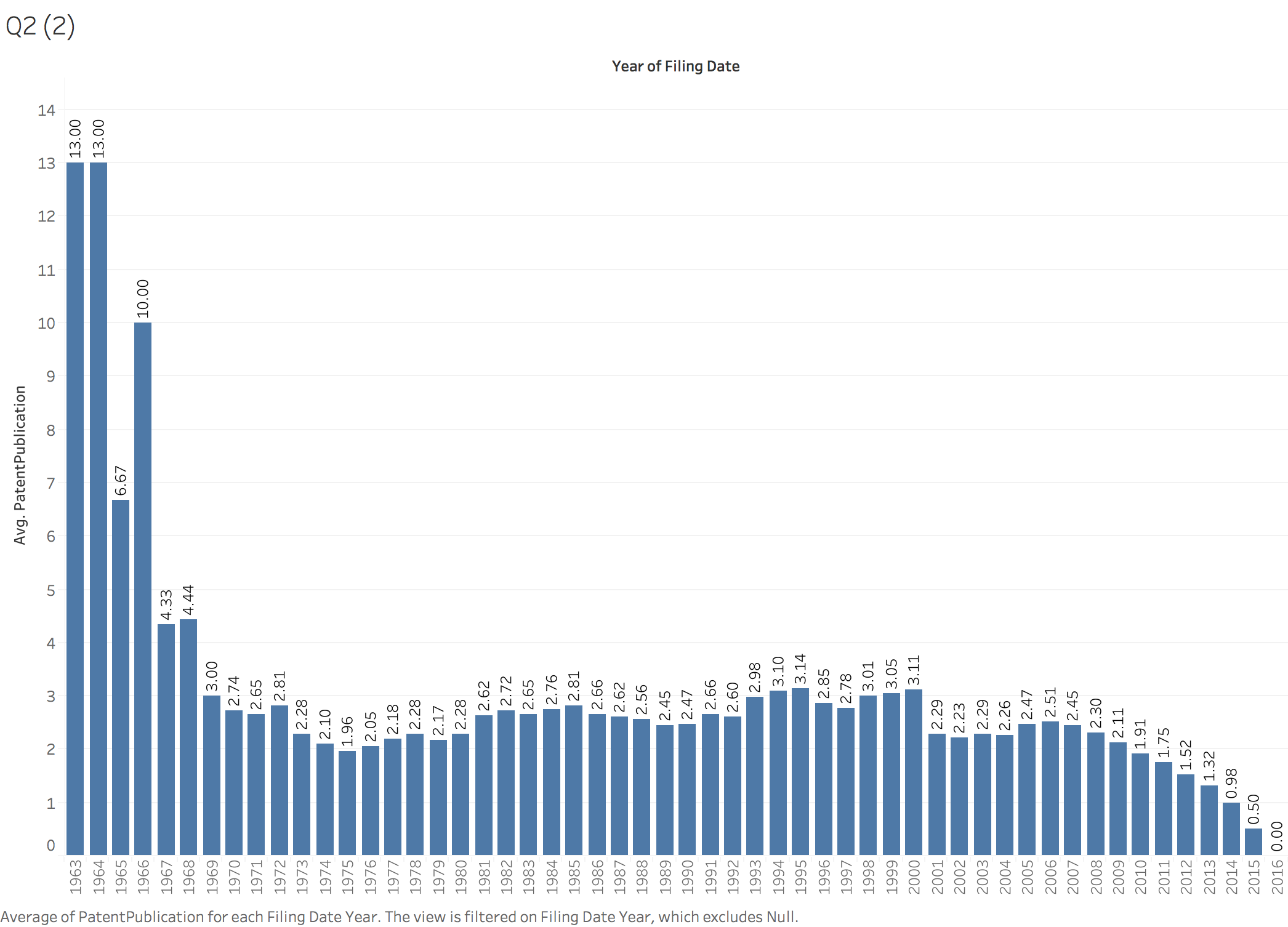
## **The evolution of patent documents by decades**



From this graph, we can see the total evolution of patent documents by decades keeps increasing. And I grouped the filing data into different decades, they are 1976-1985, 1986-1995, 1996-2005 and 2006-2016 together. However, if we explicate in detail, comparing with four decades together, we will find out that the Data Science and Radiation Measurement continue grow. In another word, the last decade increasing is higher than before several decades, so that it appears a trend of continuing to grow. But for other categories, almost just arise but the trend of the evolution have declined.

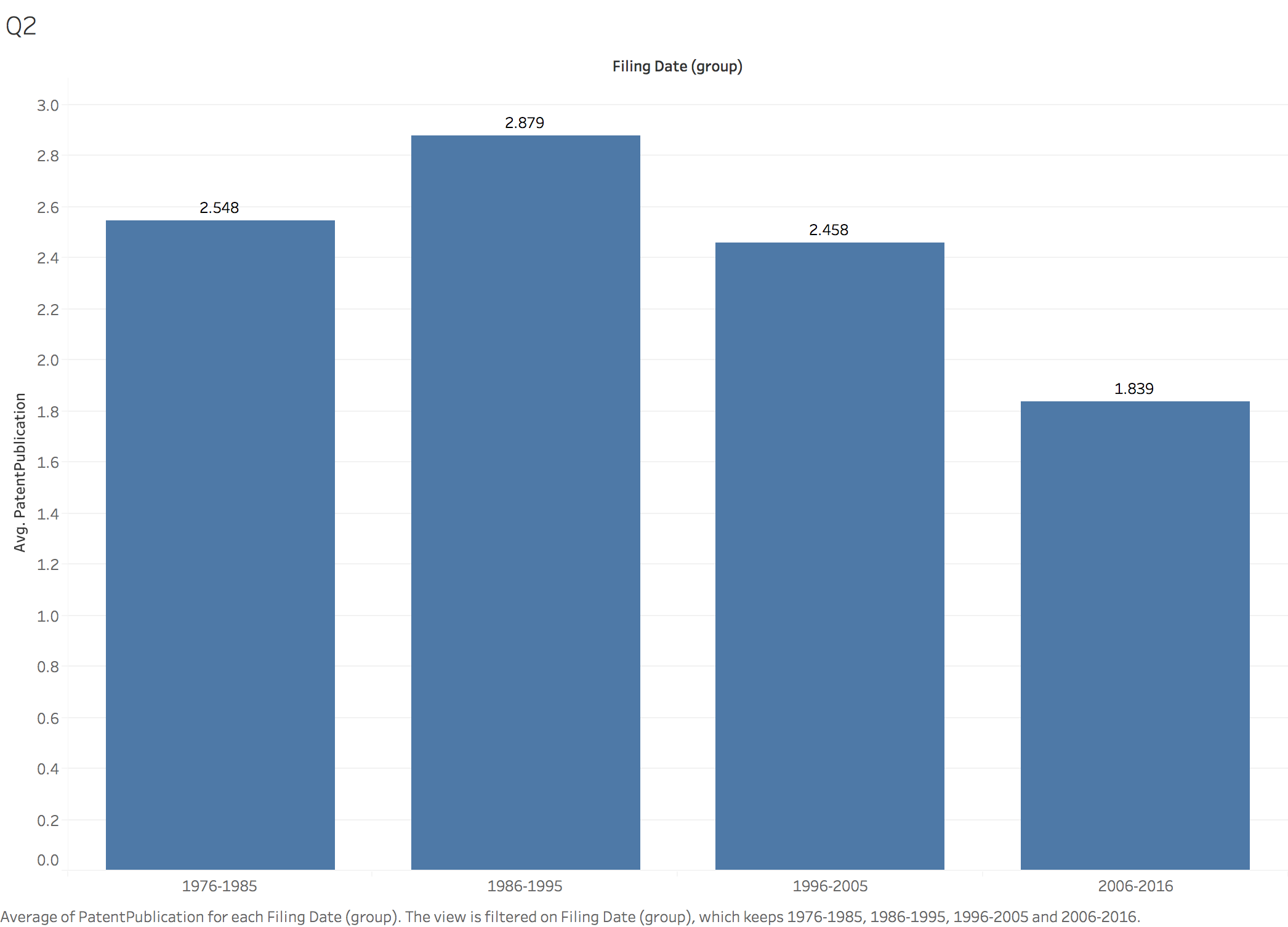
1. Patent Application Grant Time

## **The distribution of the years**



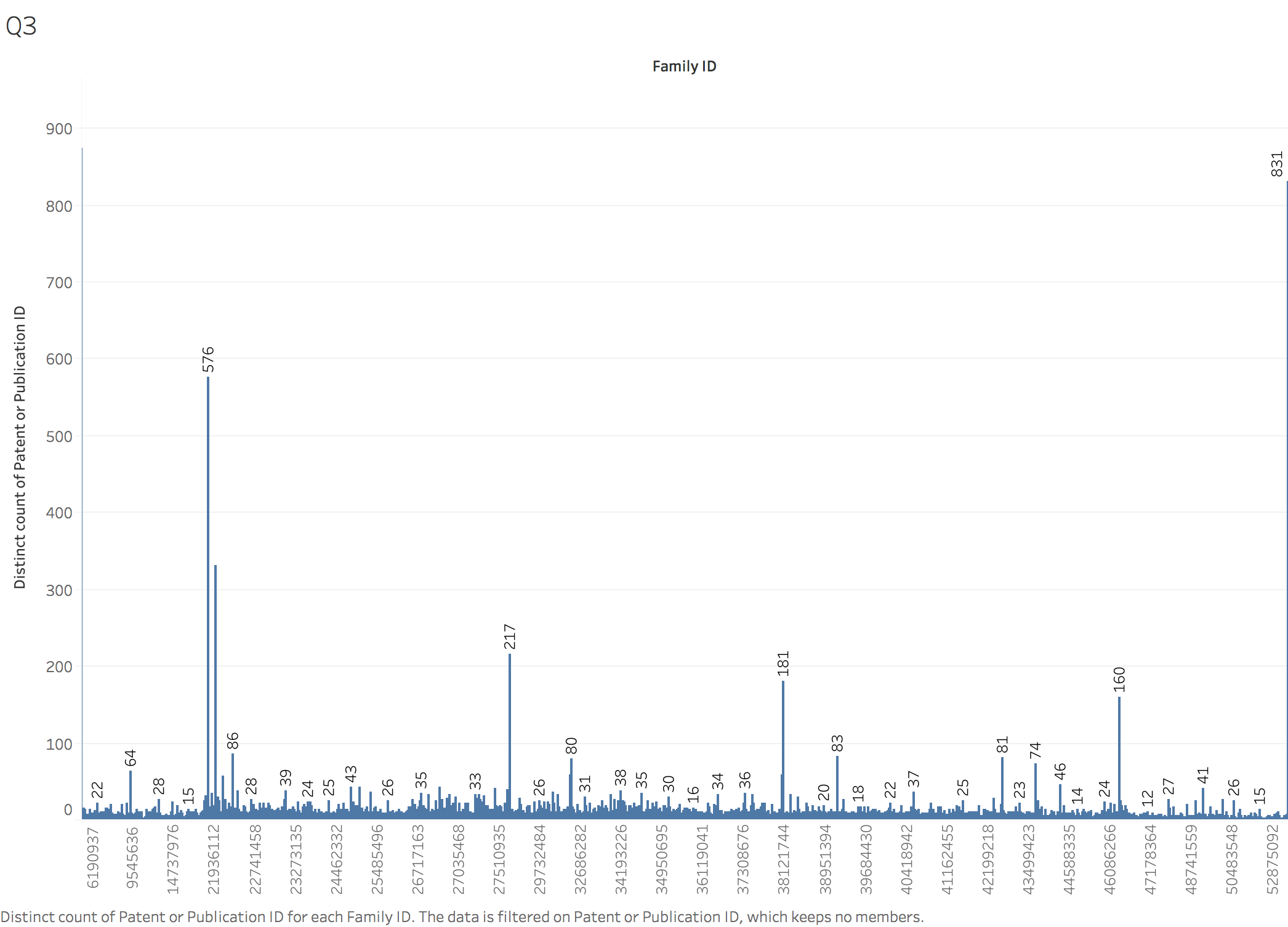
At first, I have to calculate the total years between the patent documents filing data and granted date. So, there is formula that [DATEDIFF ('year', [Filing Date], [Grant or Publication Date])] can calculate the time the patent documents take. And from the graph, the whole trend of distribution of time they take is decreased. And in 1965, the trend fell fastest.

## **The distribution of the decades**



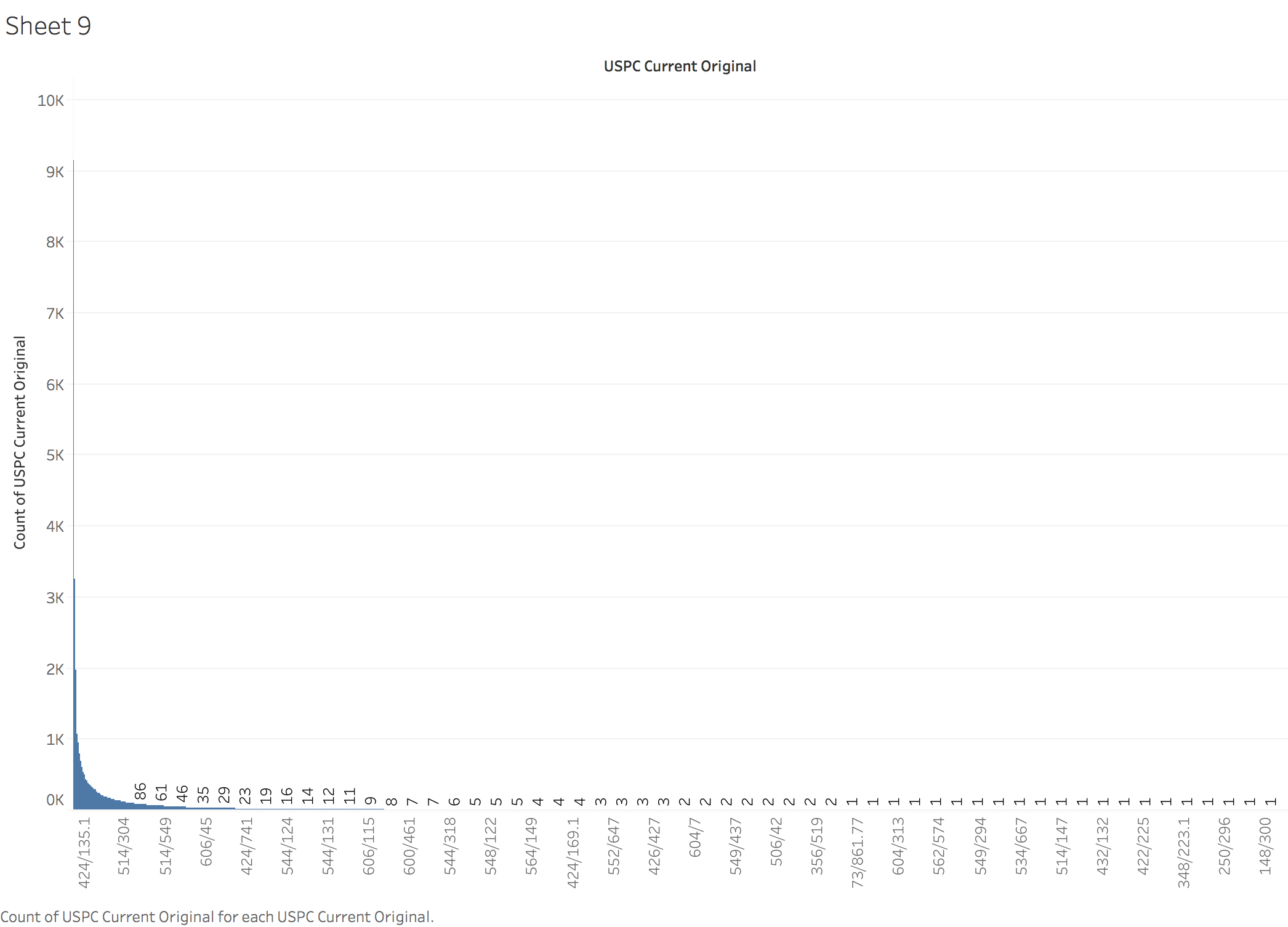
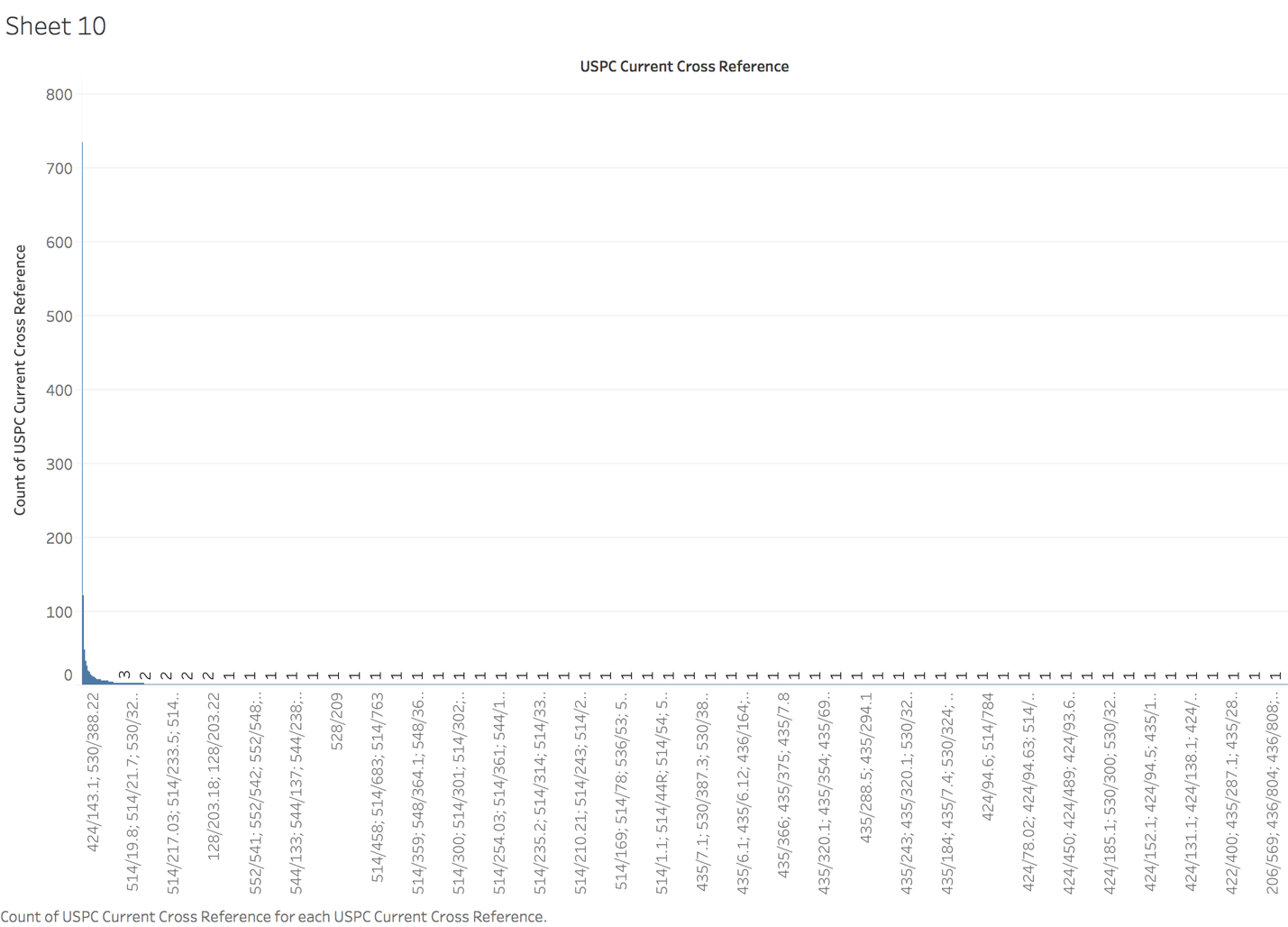
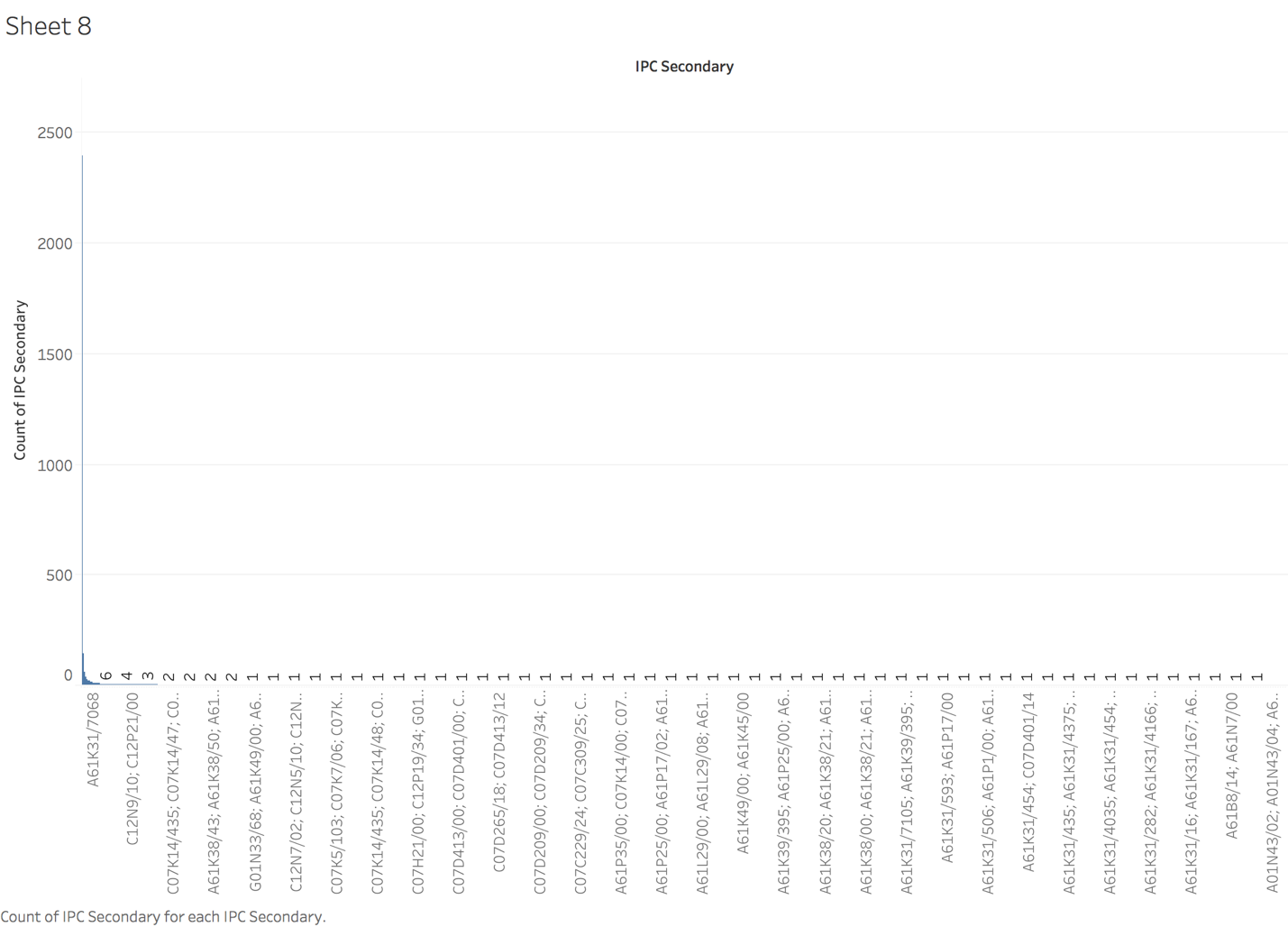
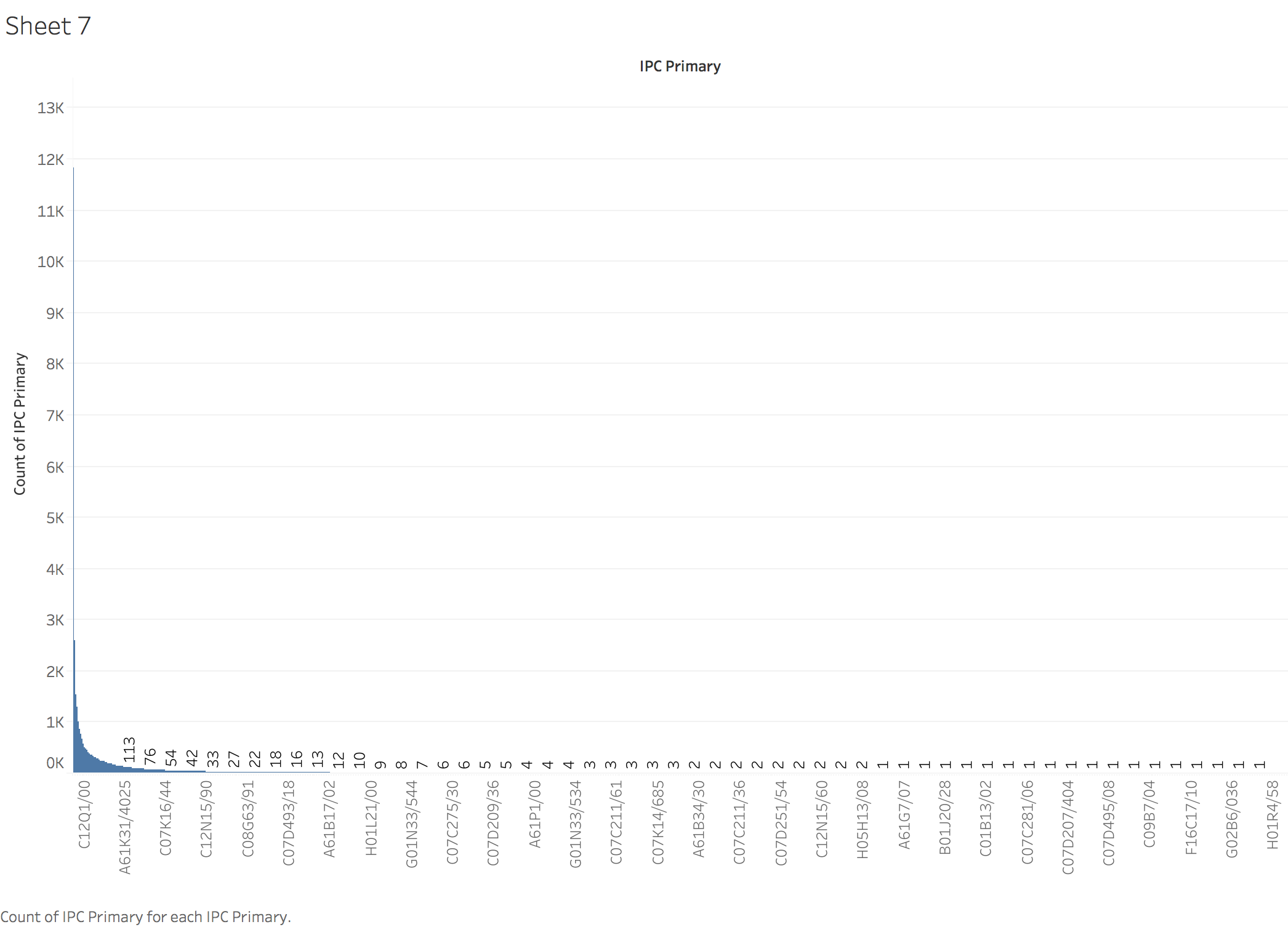
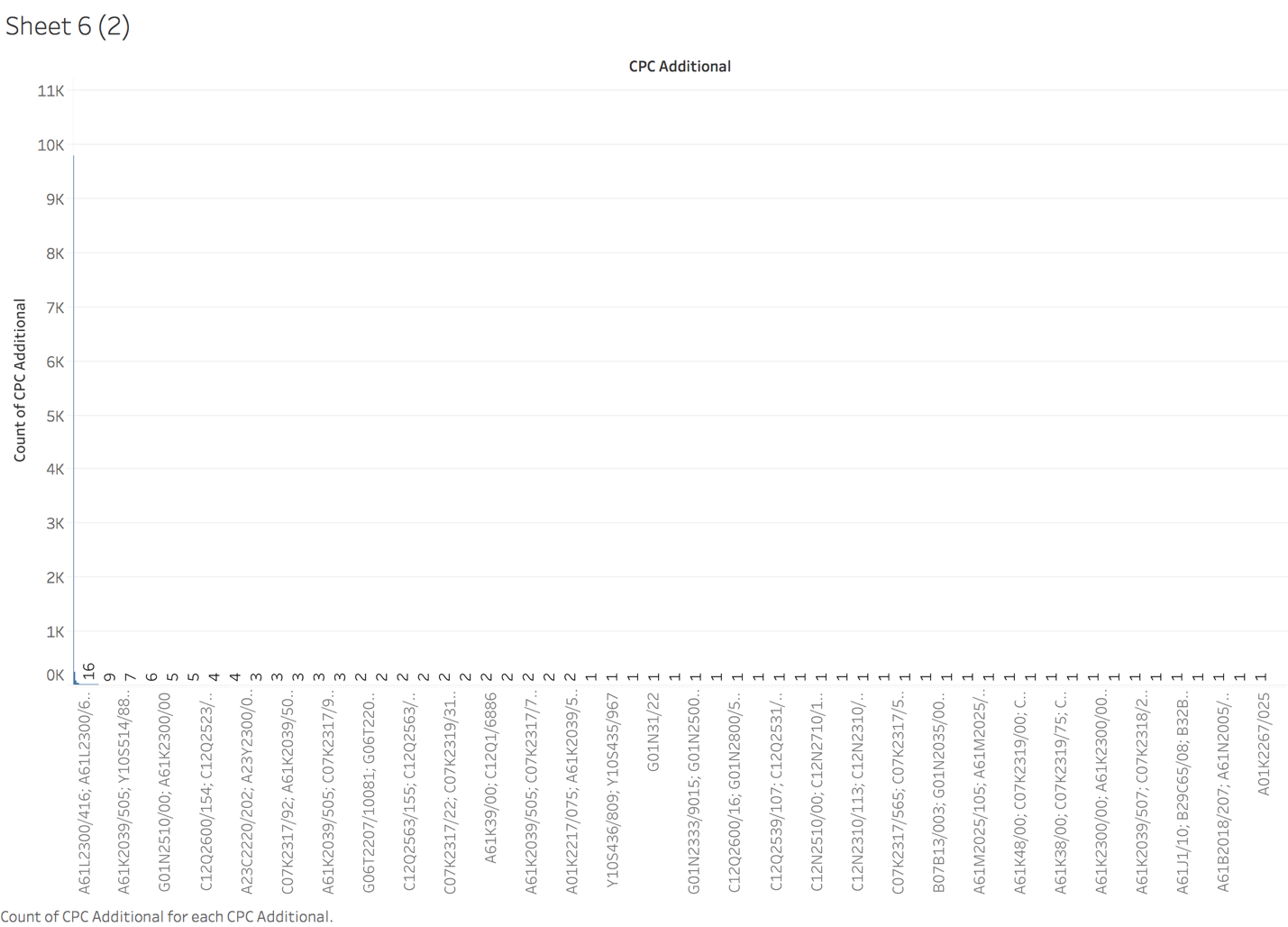
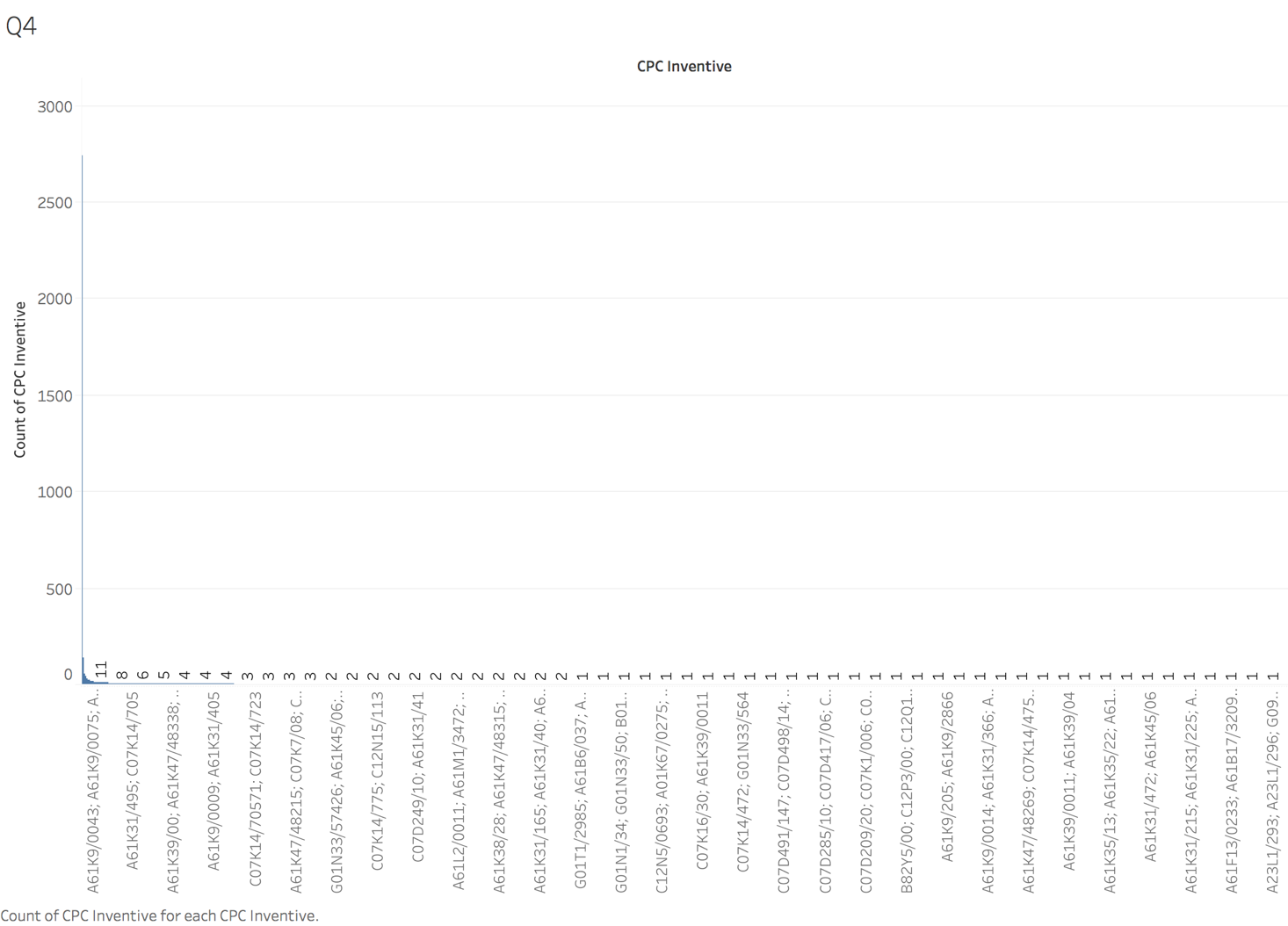
When I group the year into decades, I find that the distribution is increased at first decade. Although it is raised at beginning, it falls down obviously at last three decades. So, it means that the time of patent document granted have reduction a lot. It does not need cost a long time for waiting the patent documents passed.

# Patent family

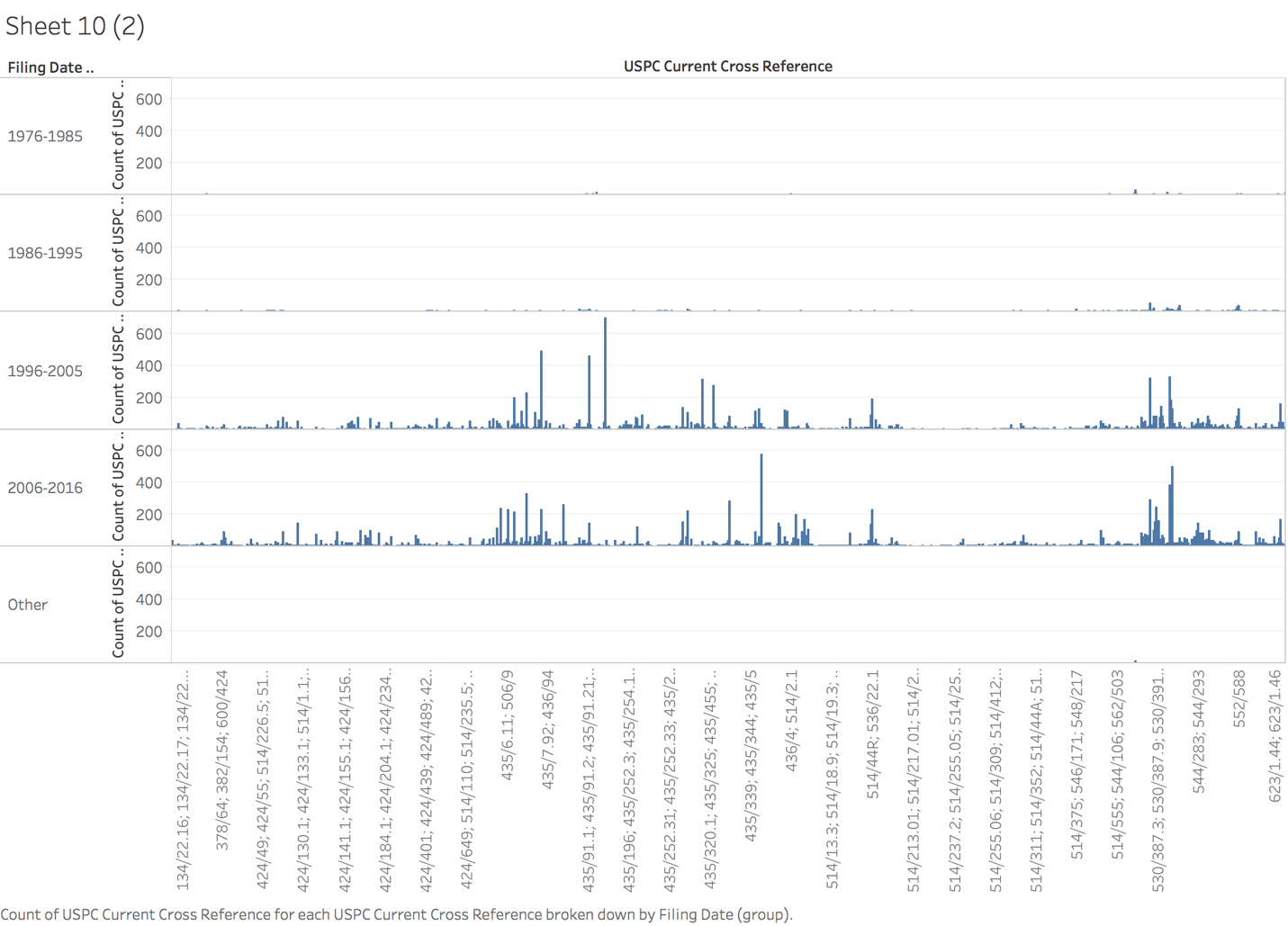
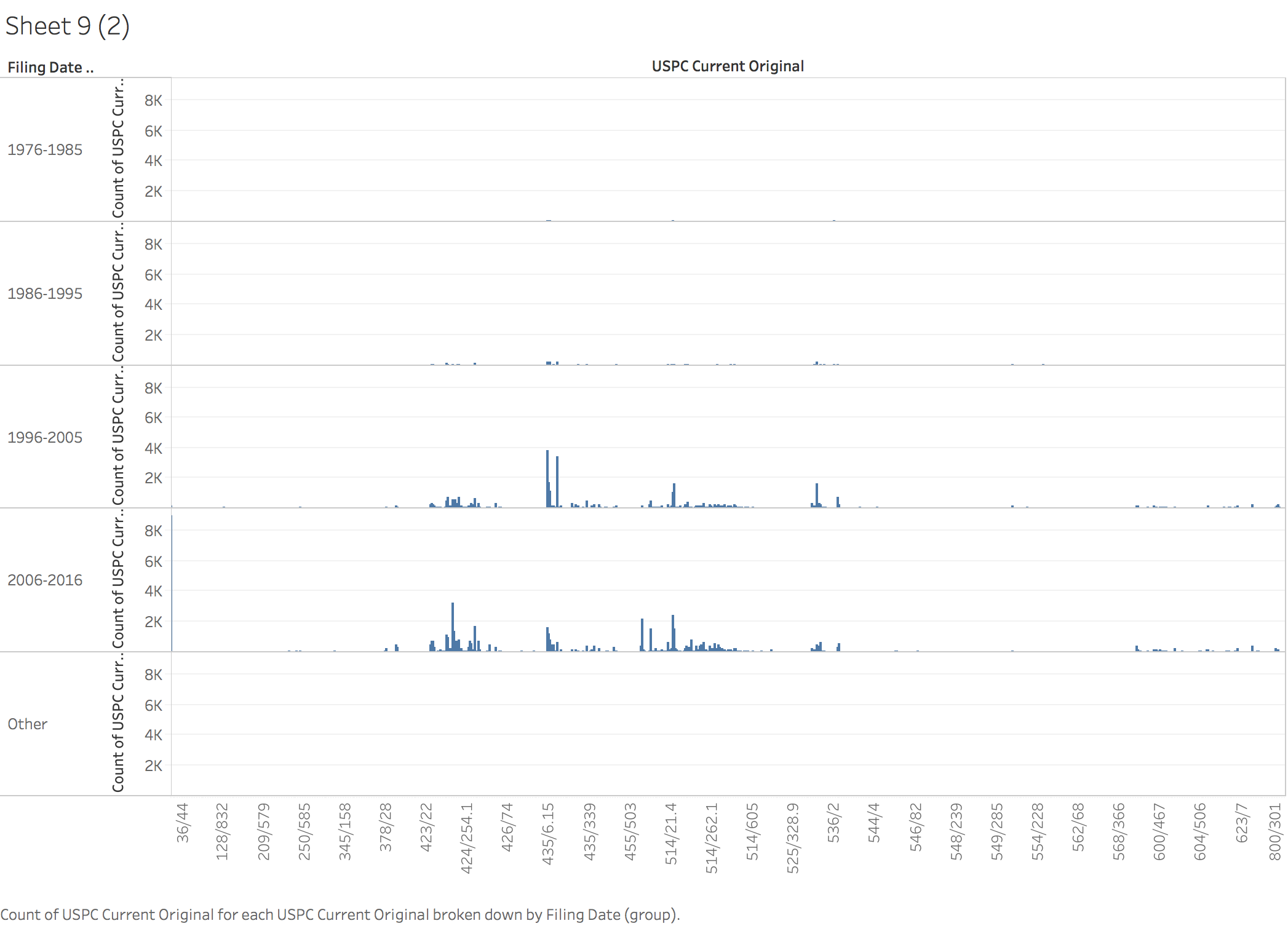
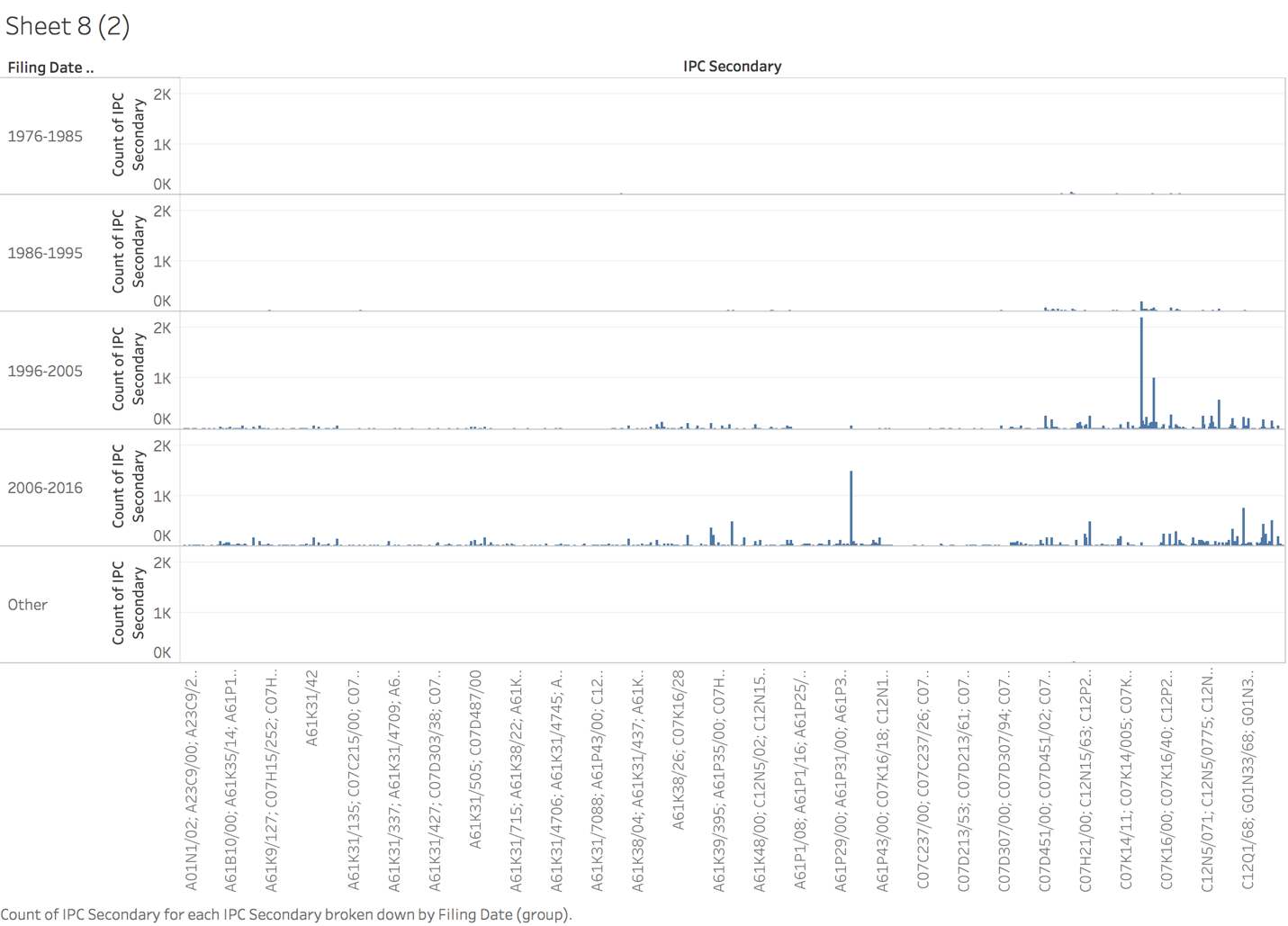
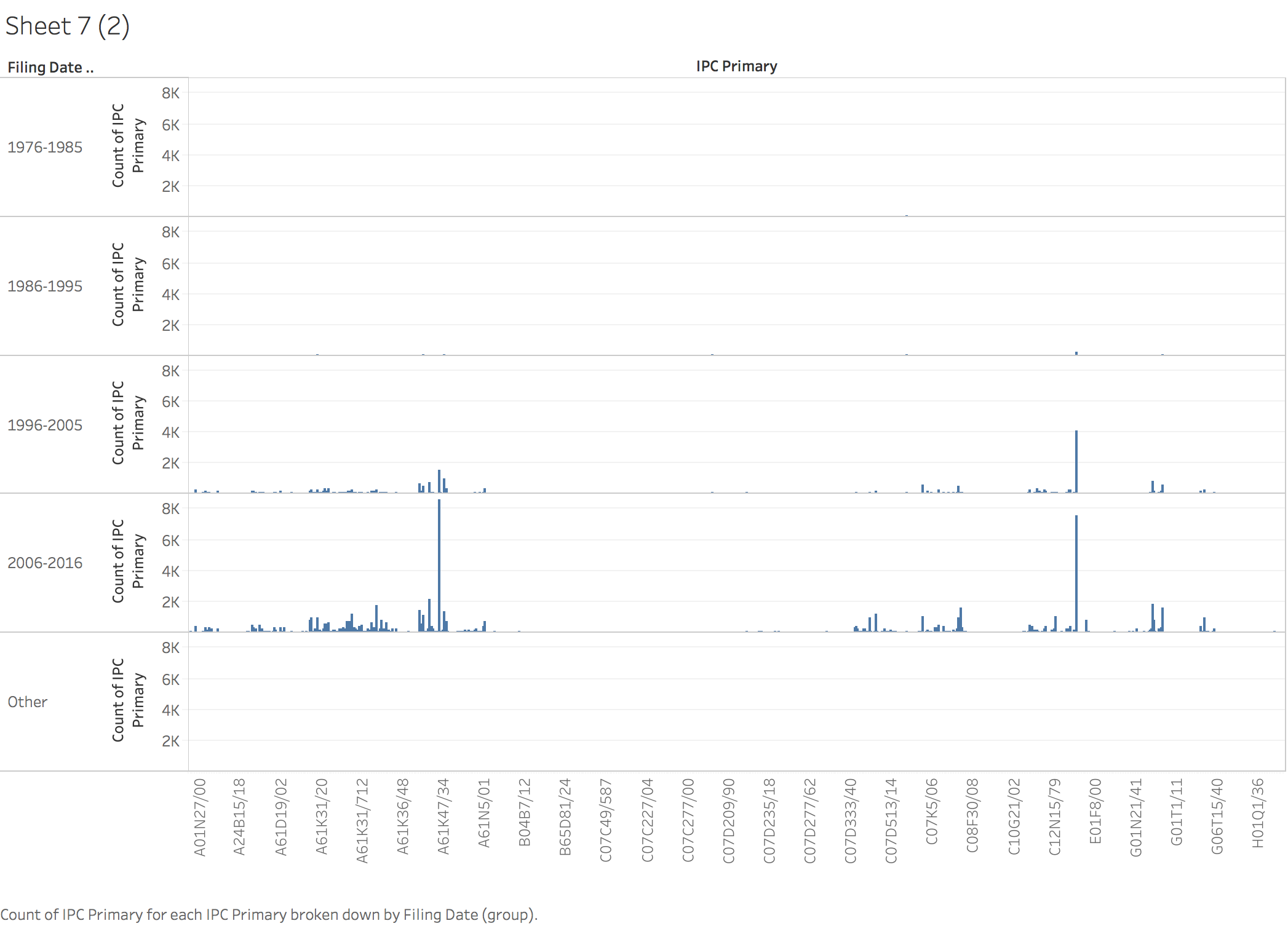
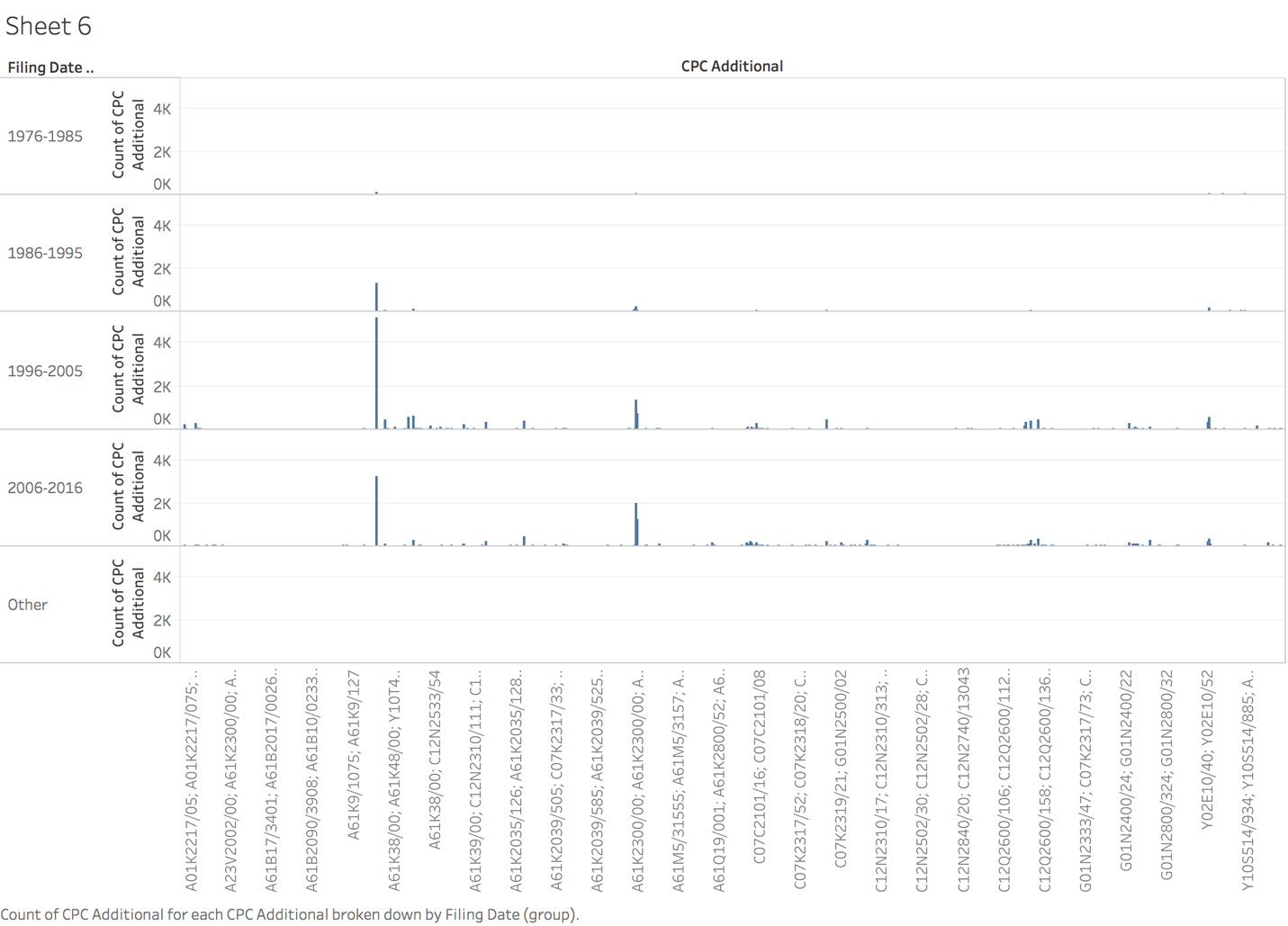
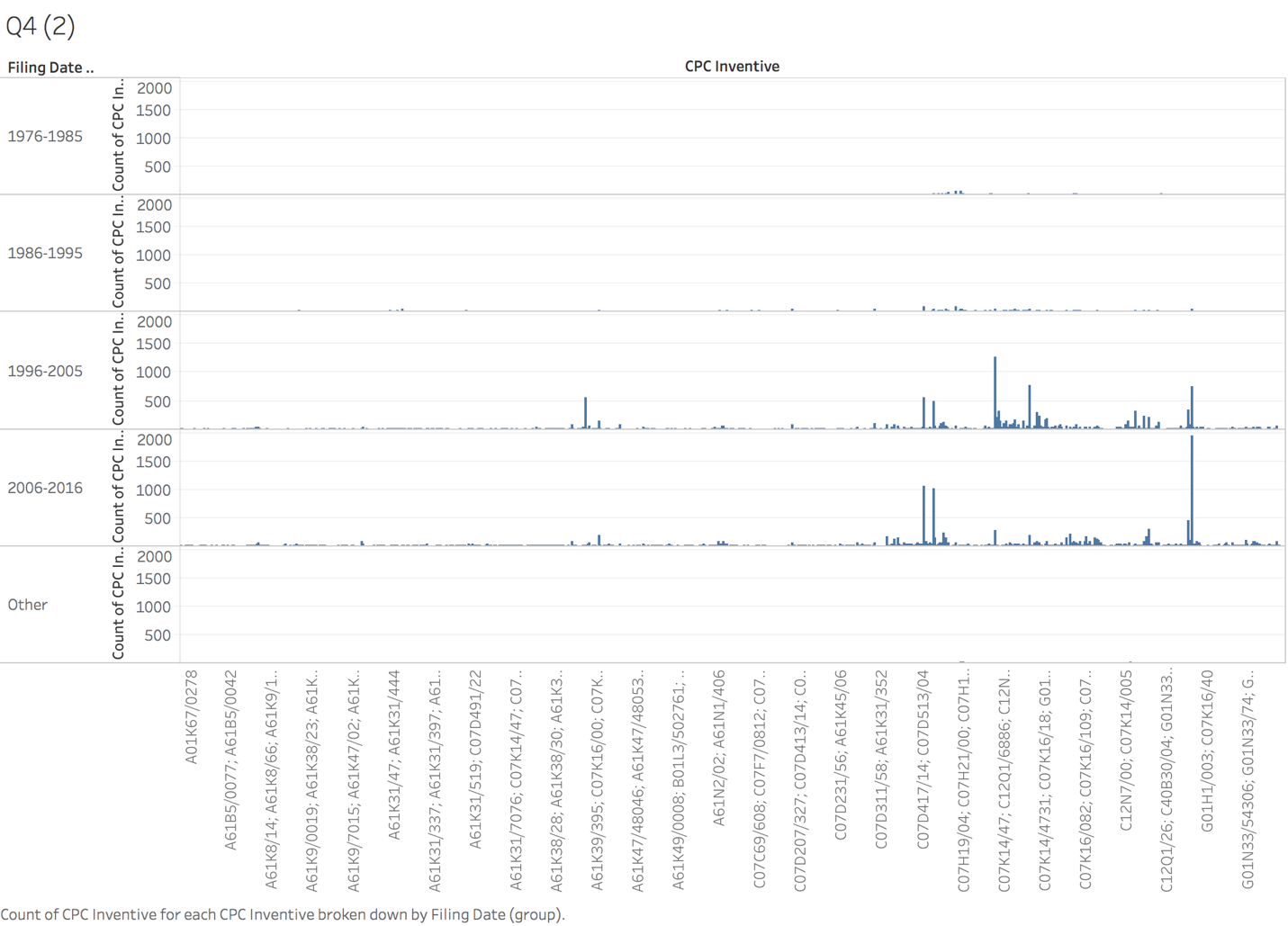


In this graph, I put the family ID as column and patent or publication ID as row. So then let patent ID in distinct, it is easy to obtain the distribution of patent ID according to different family IDs. At last, when the family ID equals to 1000000000000, they have the greatest number of Patent or Publication IDs, totally 831. And you can check the file in tableau.

# Classification Appeared in Most



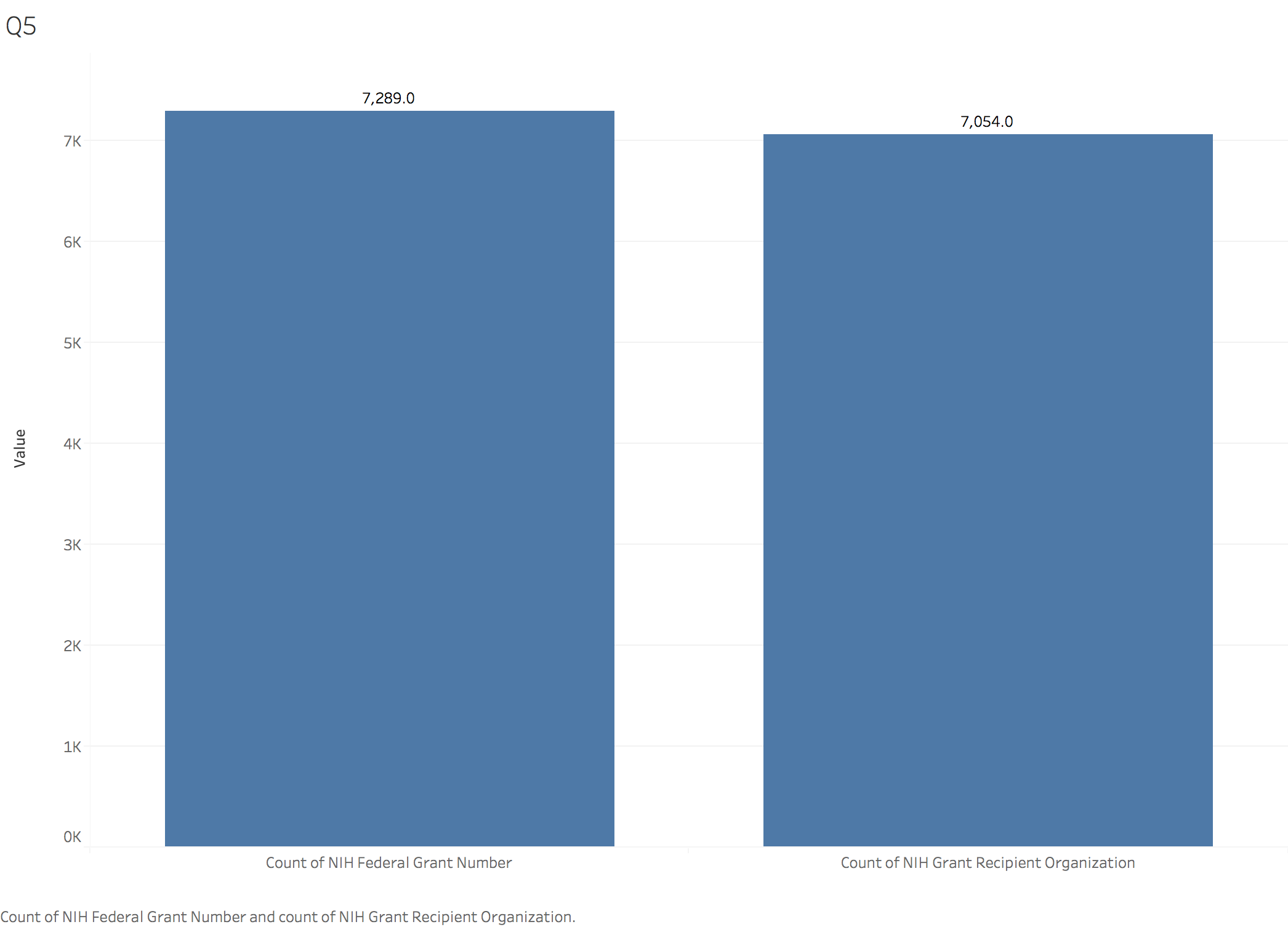
These graphs show the different classifications in specific code appearing patent documents times most often. Therefore, comparing the six graphs totally, the largest value is 11841, the IPC Primary named C12Q1/68.



These graphs show the different classifications in specific code appearing patent documents times most often split by decades. From these graphs, in 1996-2005 and 2006 -2016, all kinds of classifications of patent documents number are increased obviously, I think that is because the treating in cancer developed in that period time.

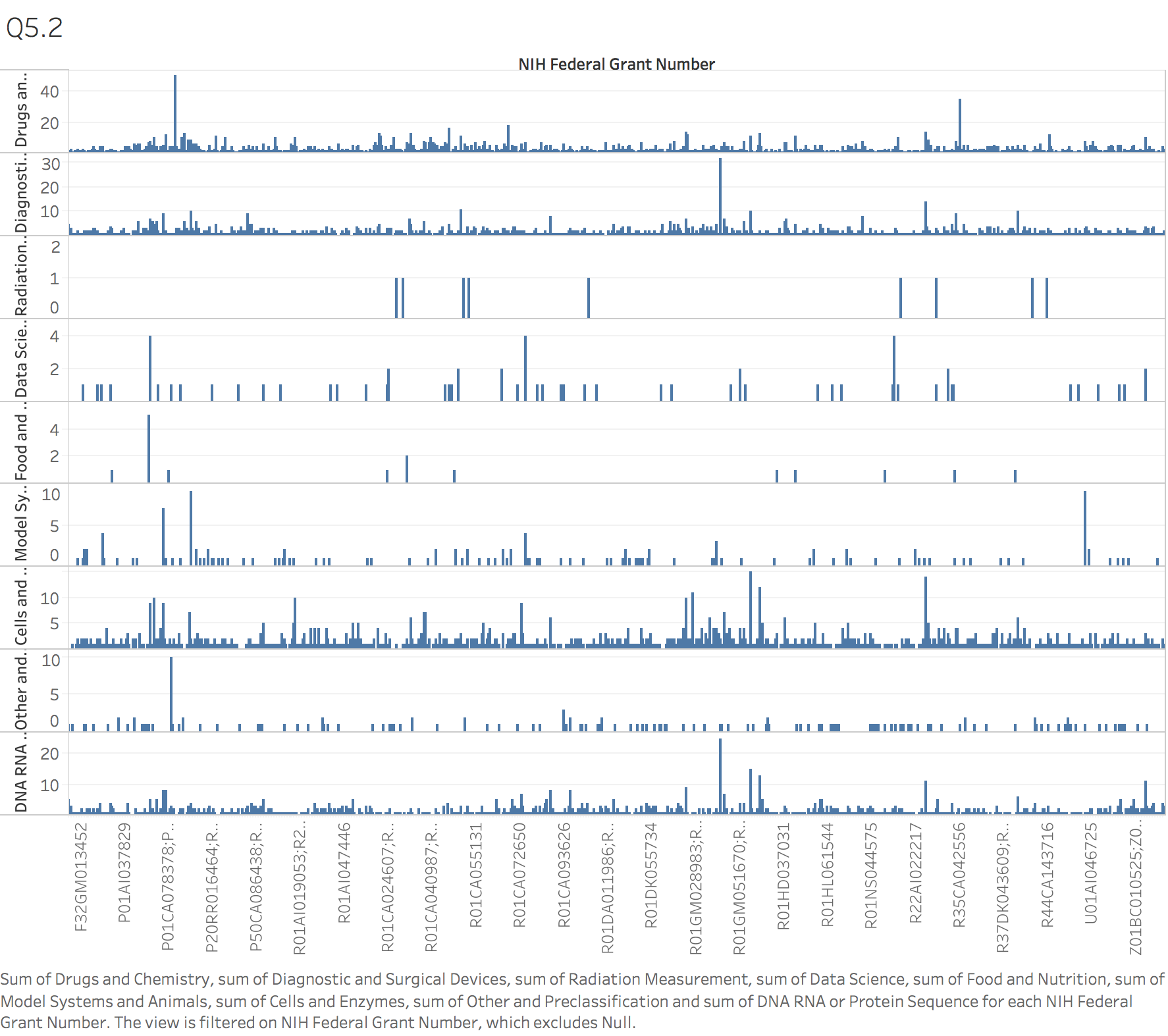
# NIH Funding

# The number of patent documents in NIH Funding

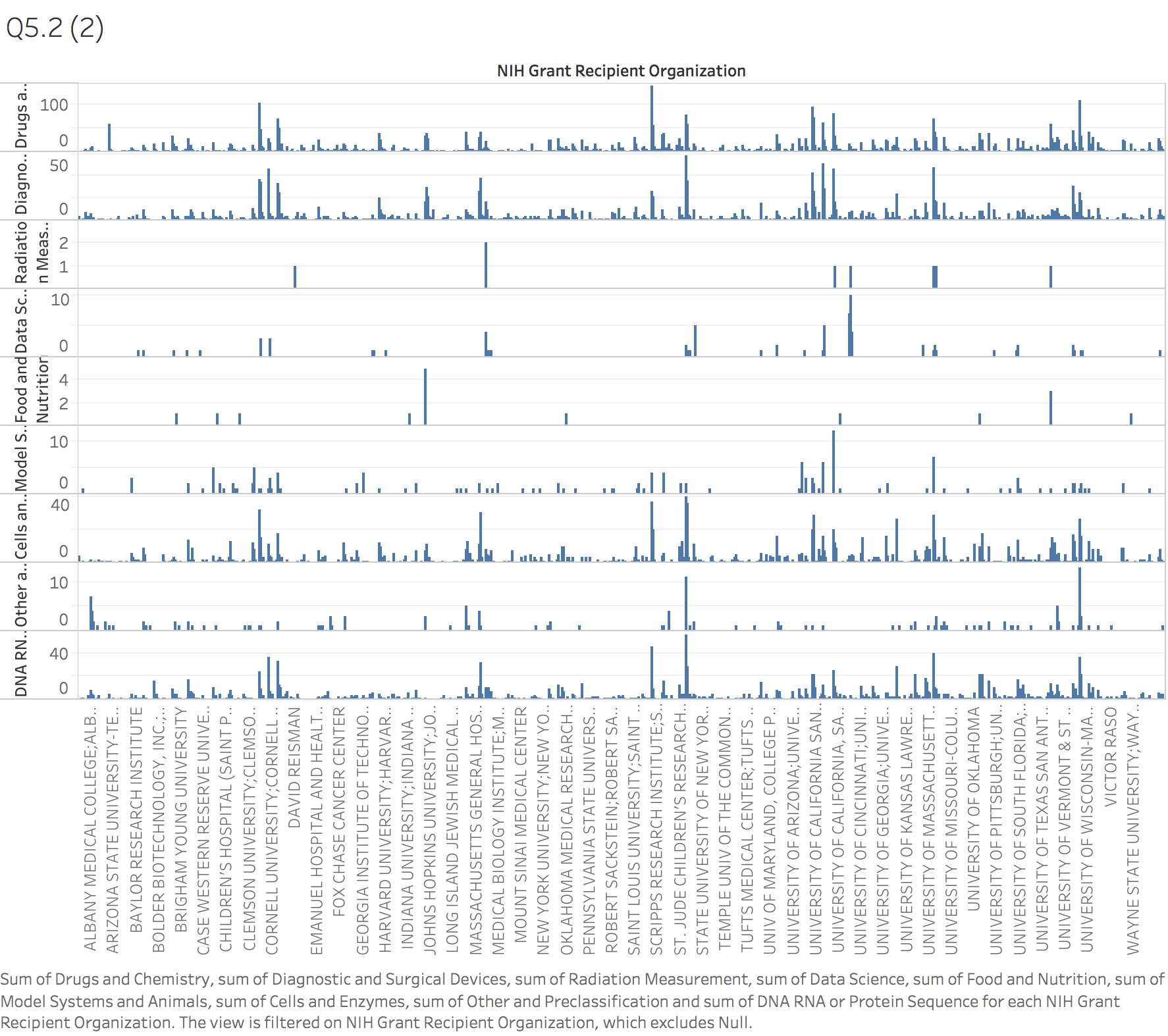


In this graph, we can see that the number of NIH federal grant number is 7289.0 and NIH grant recipient organization is 7054.0.

## **The categories of NIH Funding**



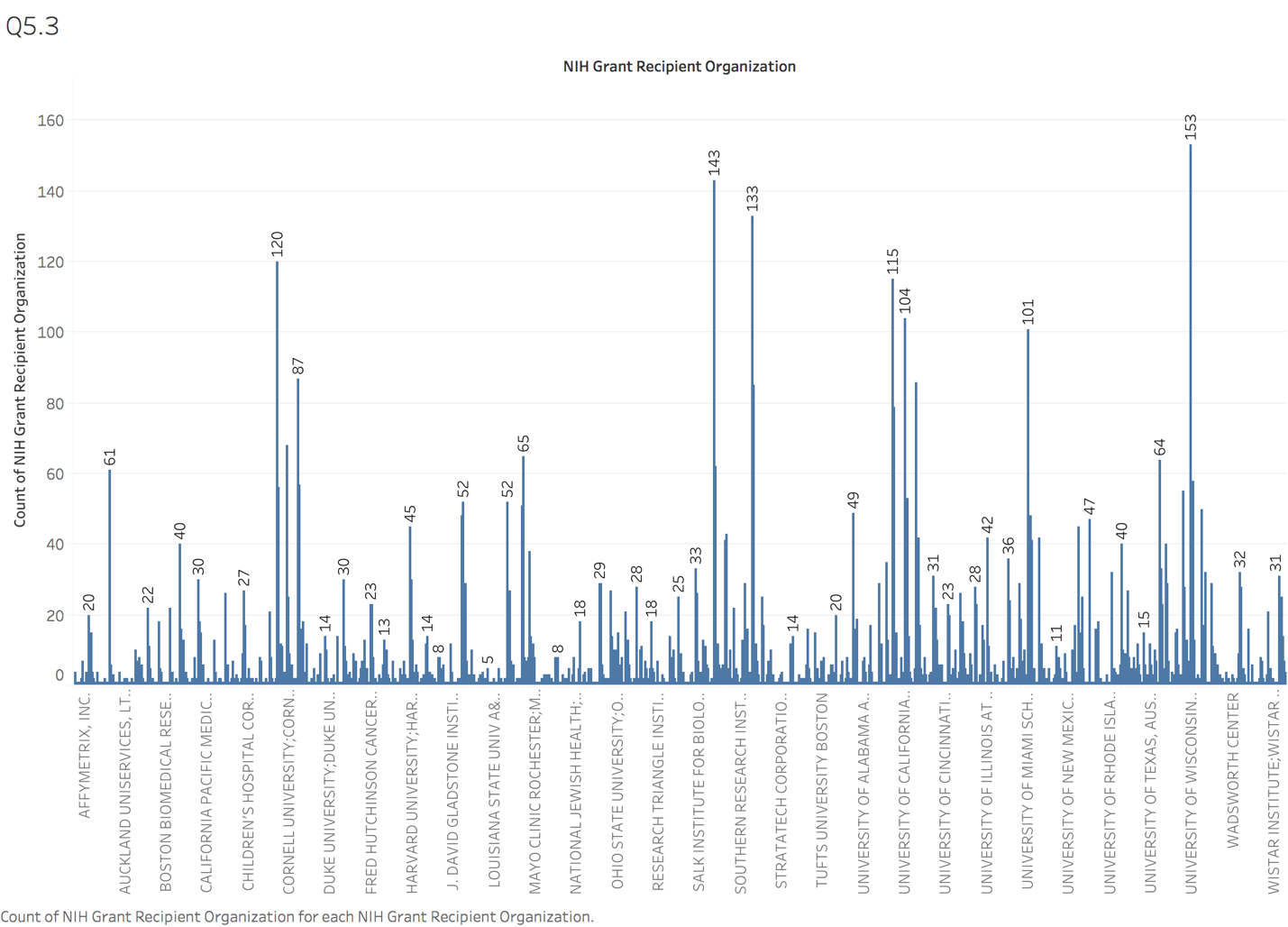
In this graph, it shows the different categories according to the NIH Federal Grant Number.



And in this graph, it shows the different categories according to the NIH Grant Recipient Organization.

It is different, so we can see that in NIH Federal Grant Number, the most often category is Cells and Enzymes. However, in NIH Grant Recipient Organization, the most often category is Cells and Enzymes, Diagnostic and Surgical Devices and Drugs and Chemistry. I am realized that it split into different category because of the organization characteristic.

## **The organization receive NIH the most often**



And the organization named University of Wisconsin-Madison received the NIH funding the most often and the number of that is 153.

## **University of Wisconsin-Madison**

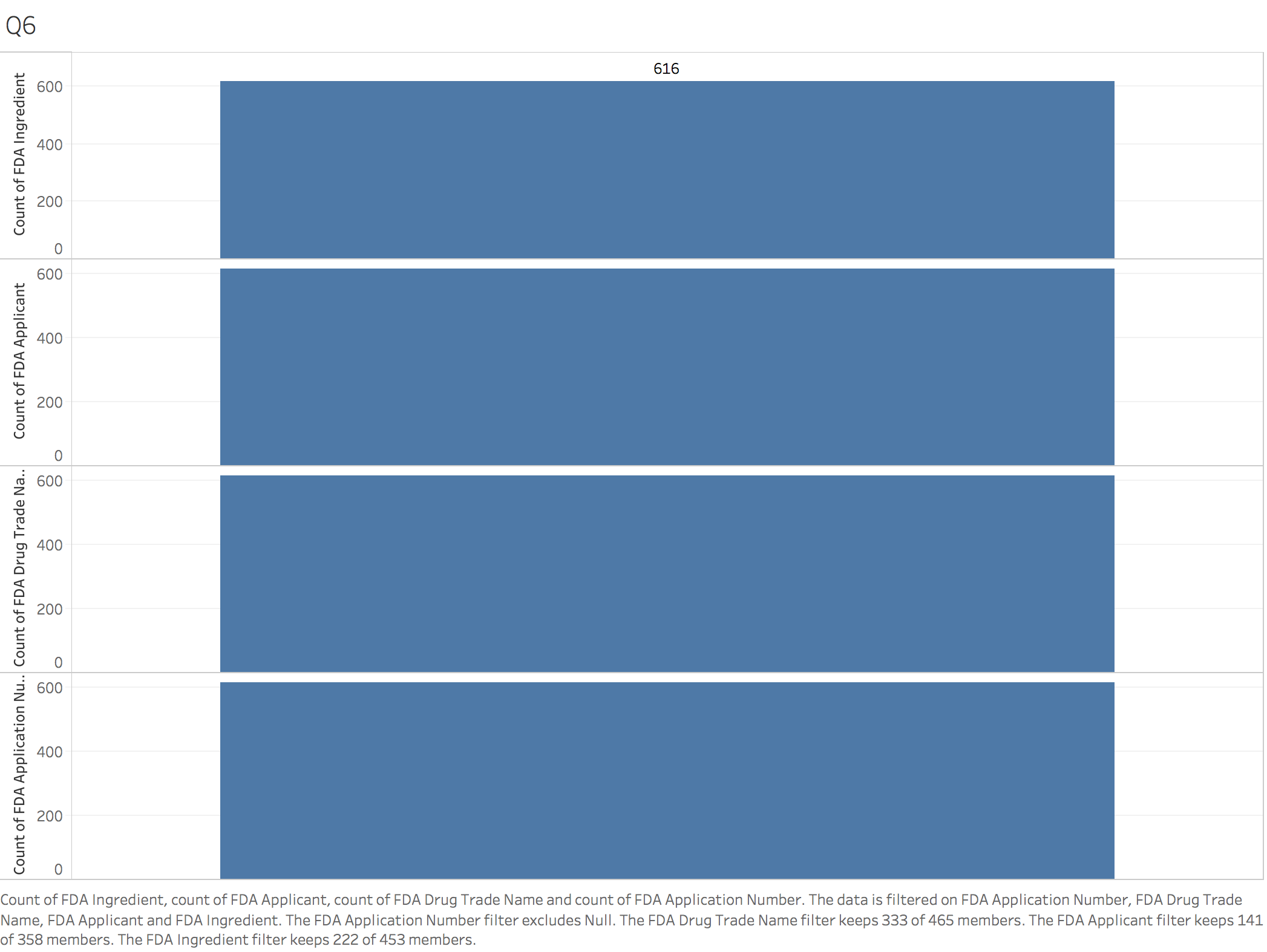


The University of Wisconsin–Madison (also known as University of Wisconsin, Wisconsin, UW, or regionally as UW–Madison, or simply Madison) is a public research university in Madison, Wisconsin. And it was founded at 1848, it is the official state university of Wisconsin. Also, it was the first public university founded in Wisconsin and remains the oldest and largest public university in the state.

It is the largest university college, the College of Letters and Science, enrolls approximately half of the undergraduate student body and is made up of 38 departments and five professional schools that instruct students and carry out research in a wide variety of fields, such as astronomy, economics, geography, history, linguistics, and zoology. The graduate instructional program is classified by Carnegie as "comprehensive with medical/veterinary." In 2008, it granted the third largest number of doctorates in the nation.

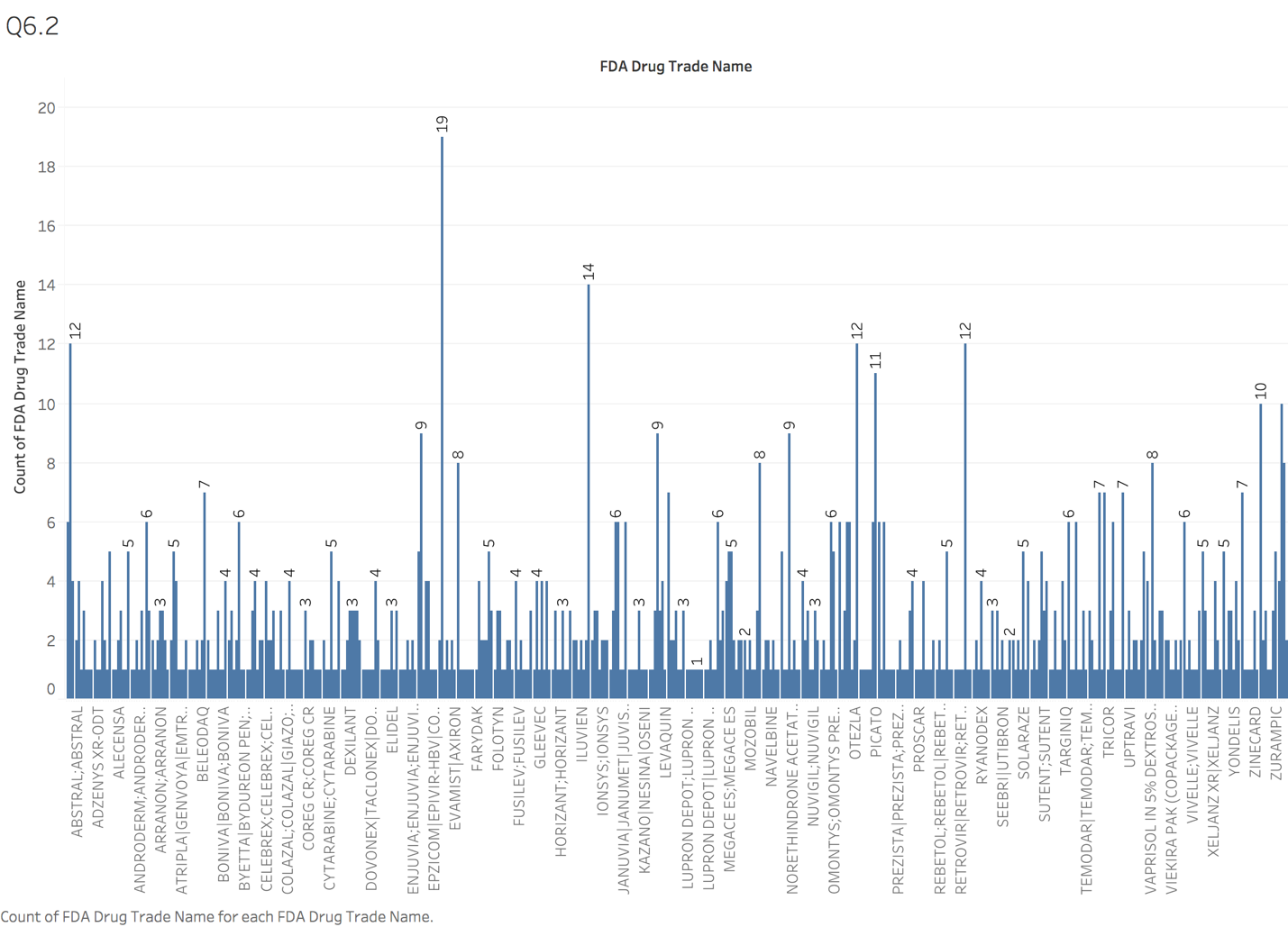
# FDA Approval

## **The number of FDA Approval**

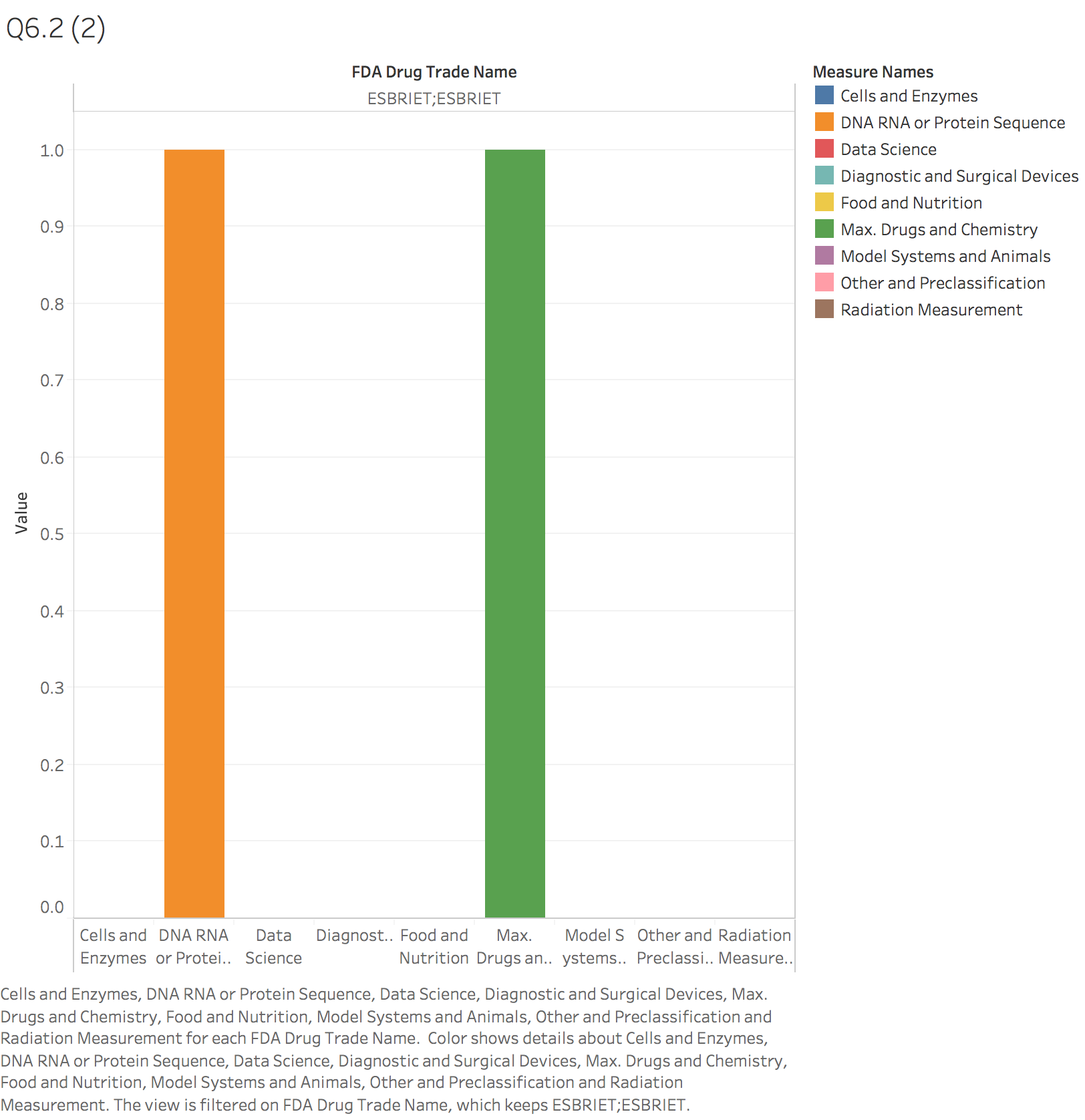


After fliting the null value, the number of patent documents having received FDA approved is 616 together. And there are four elements of FDA, they are FDA Application Number, FDA Drug Trade Name, FDA Approval Date, FDA Applicant and FDA Ingredient. And we just calculate the number, but not the trend of that, so we pick of the FDA approval date out at first.

## **FDA drugs the most often**



From this graph, we select out drugs appeared the most often. Its name is ESBRIET and total number is 19.



And the DNA RNA or protein sequence and Drugs and Chemistry are belonged to this kind of drug.

## **The common code**



From this picture, we find out some common code in ESBRIET drugs.

In CPC Inventive, it has several code in common and they are A61K 31/496; 31/4412; 31/4418; 31/496.

In CPC Additional, it has the common code, it is A61K 2330/00

IPC Primary is A61K 31/497

IPC Secondary is A61K 31/435

USPC Current Original is 514/253.07

USPC Current Cross Reference is 514/277

So, according to the document from the Cancer Moonshot website for descriptions, A61K means the Cancer antigens (Medicinal) in IPC Classification and 514 means the Cancer (Drugs) in USPC Classification. Therefore, we can see that the ESBRIET’s codes are almost focus on the cancer antigens and drugs.

# ESBRIET

## **Esbriet (Pirfenidone)**

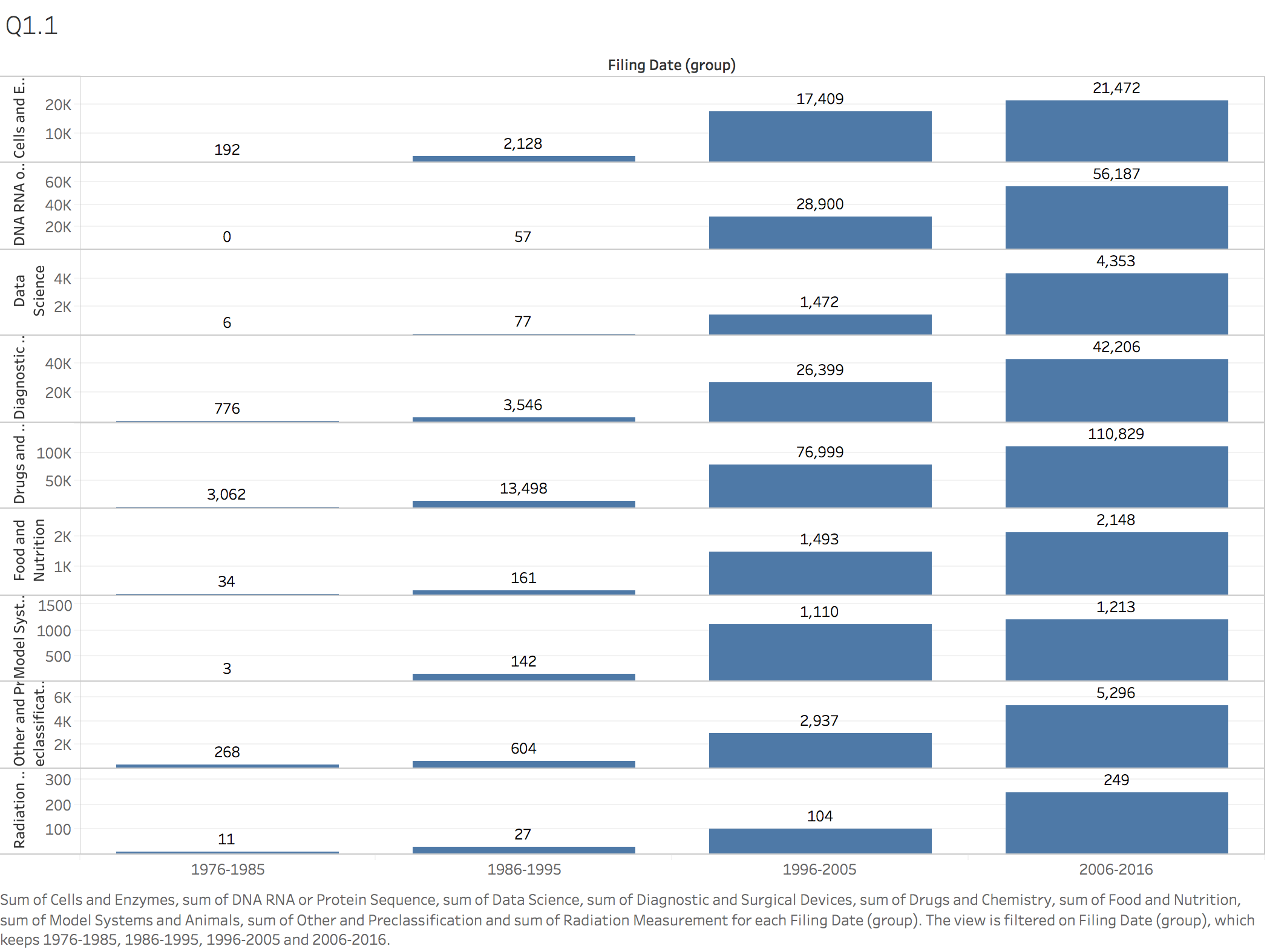
Pirfenidone is a medication used for the treatment of idiopathic pulmonary fibrosis. It works by reducing lung fibrosis through downregulation of the production of growth factors and procollagens I and II.

## **Financial data of Esbriet**

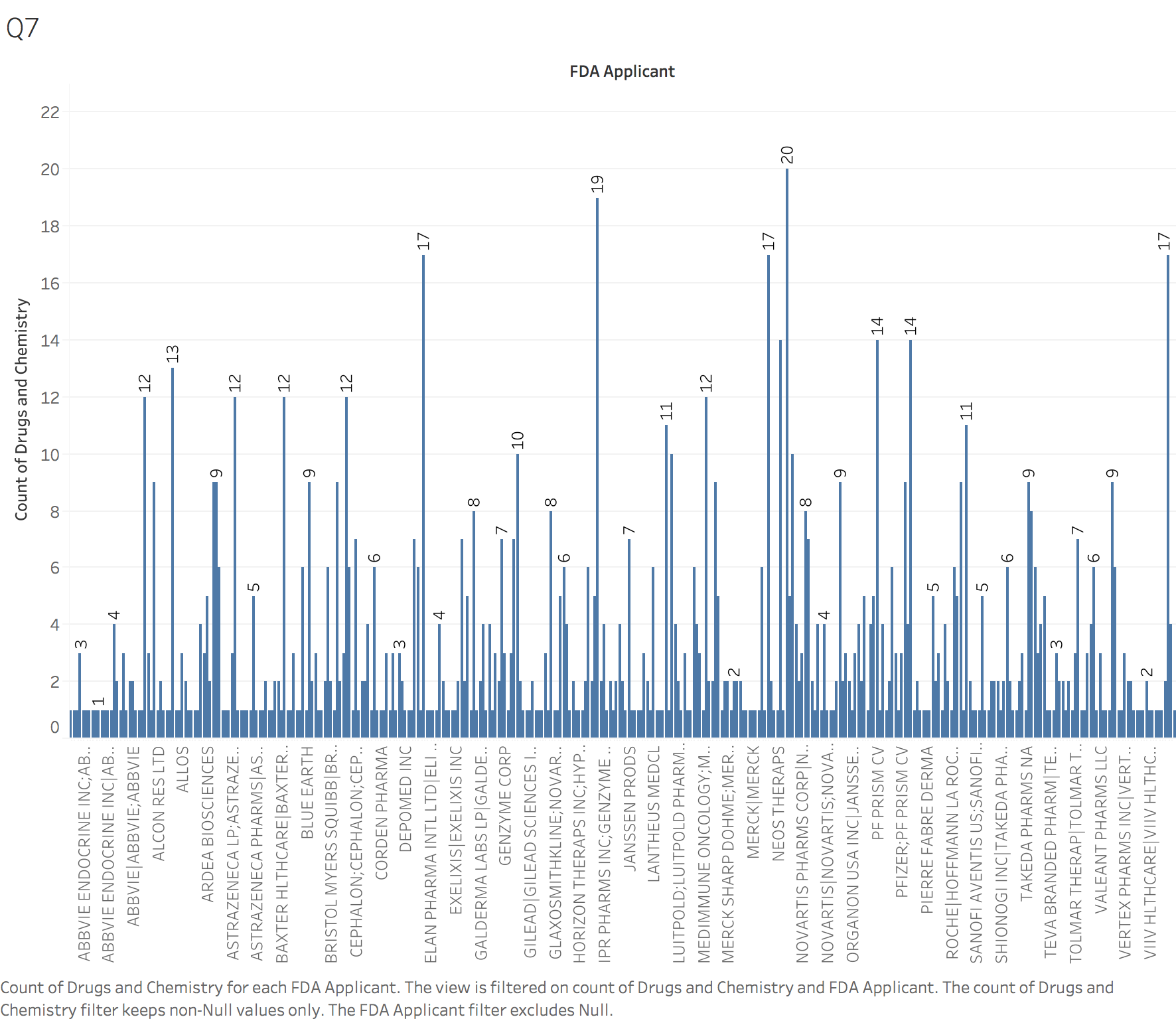
This Esbriet price is based on using the [**Drugs.com discount card**](https://www.drugs.com/discount-card/) which is accepted at most U.S. pharmacies. The cost for Esbriet oral capsule 267 mg is around $9,701 for a supply of 270 capsules, depending on the pharmacy.

FDA approved Esbriet for treatment of idiopathic pulmonary fibrosis in 2014. The drug, whose revenue stood at $48 million in 2014, garnered nearly $856 million in 2017. Peak sales could range from $3-4 billion. The drug was developed by Intermune, which was acquired by Roche.

## The most lucrative category



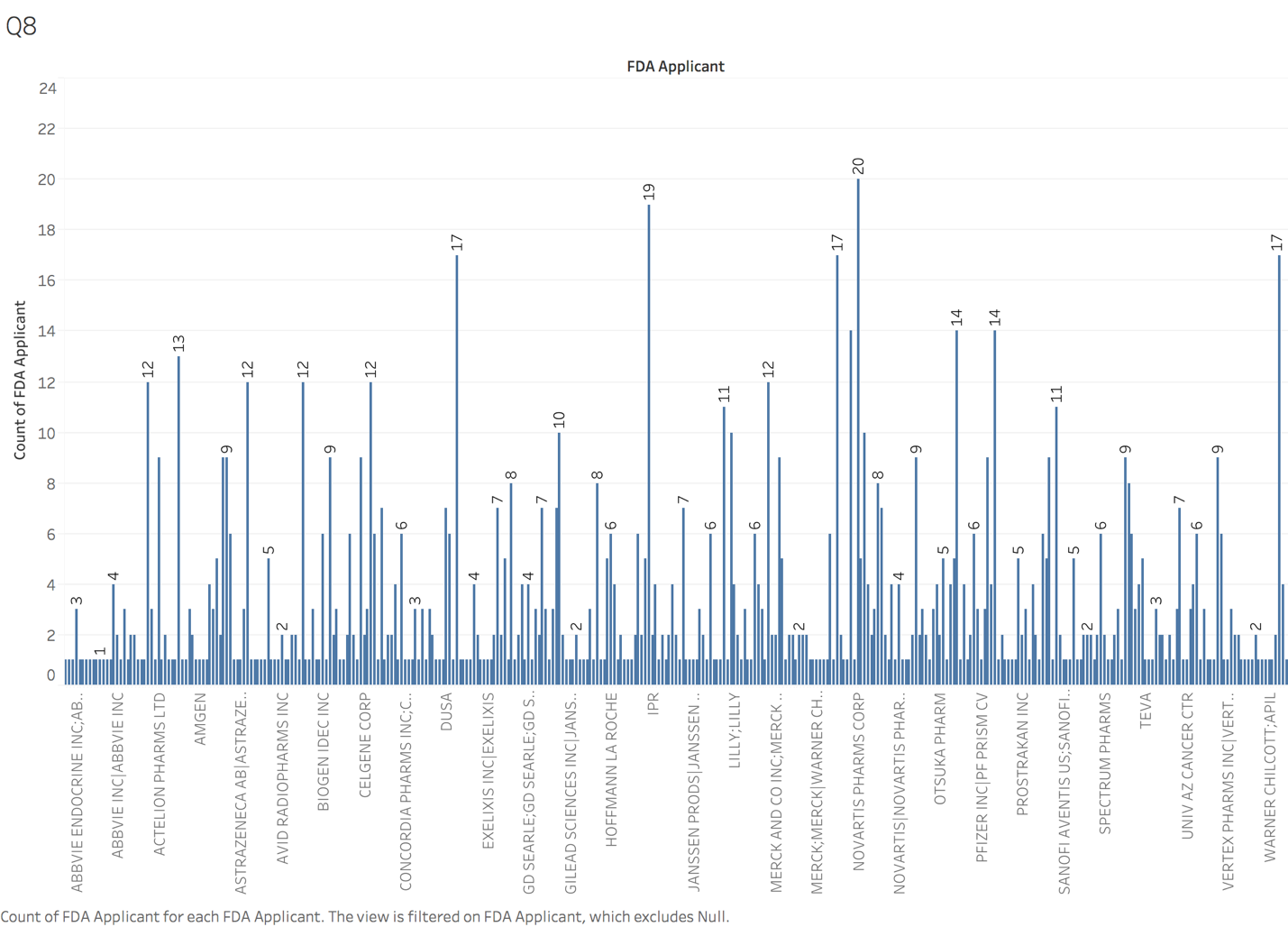
So, I decide to use the largest number of the patent documents as the lucrative category, it means that the more patent documents people apply, the more value and lucrative they have. Then we can see Drugs and Chemistry has the largest patent number.



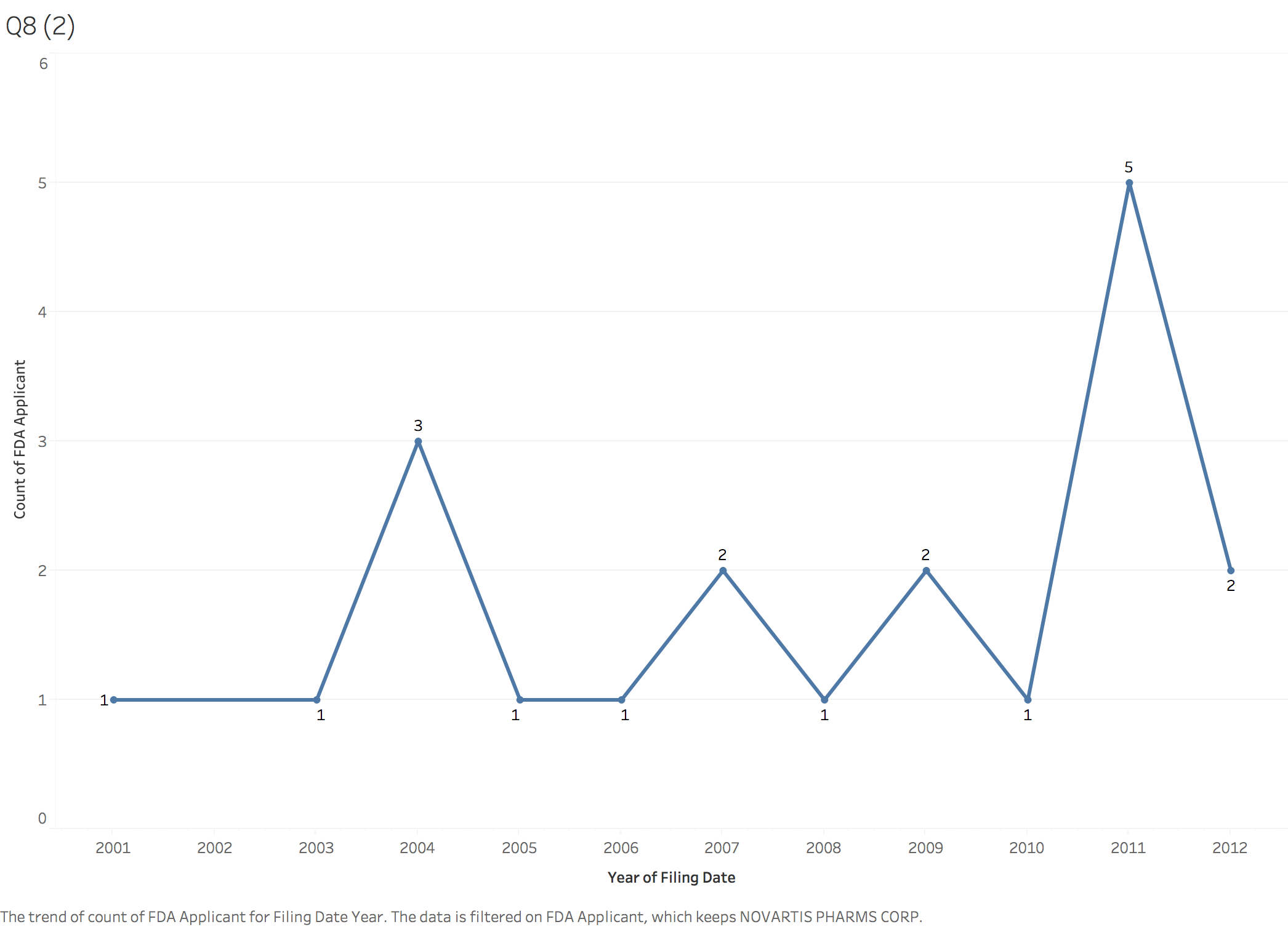
And this graph shows that the companies focus on the Drugs and Chemistry category. The company named NOVARTIS PHARMS CORP focus on the most lucrative category, and they apply for the patent document number is 20.

# FDA Analysis

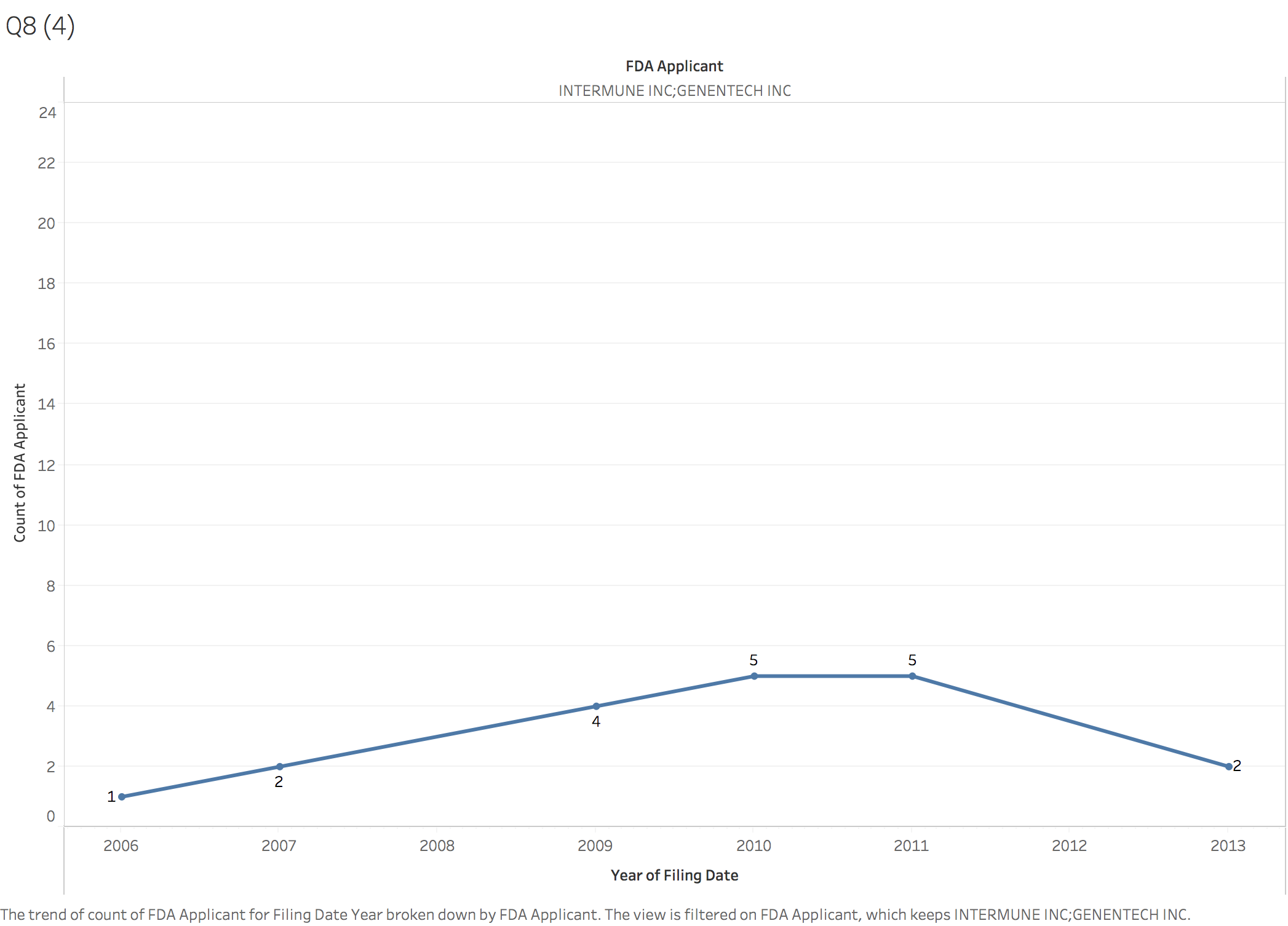
## **The company having the most patent documents**



From the graph, I pick out the company having the most number patent documents. So, the company named NOVARTIS PHARMS CORP and the number of patent documents is 20. Besides that, there are also have two company having 19 patent documents named INTERMUNE INC; GENENTECH INC, but it is lower than NOVARTIS PHARMS CORP. And there are three companies having 17 patent documents totally, they are EISAI INC, MILLENNIUM PHARMS, WYETH PHARMS INC.



In this graph, it shows the count of FDA Applicant by different years. We can see that NOVARTIS PHARMS CORP appeared in 2001 and started having the FDA patent documents. When 2004, it had 3 patent documents, and when 2011, it had 5 patent documents. But in other time, they just only had 1 or 2 documents every year.



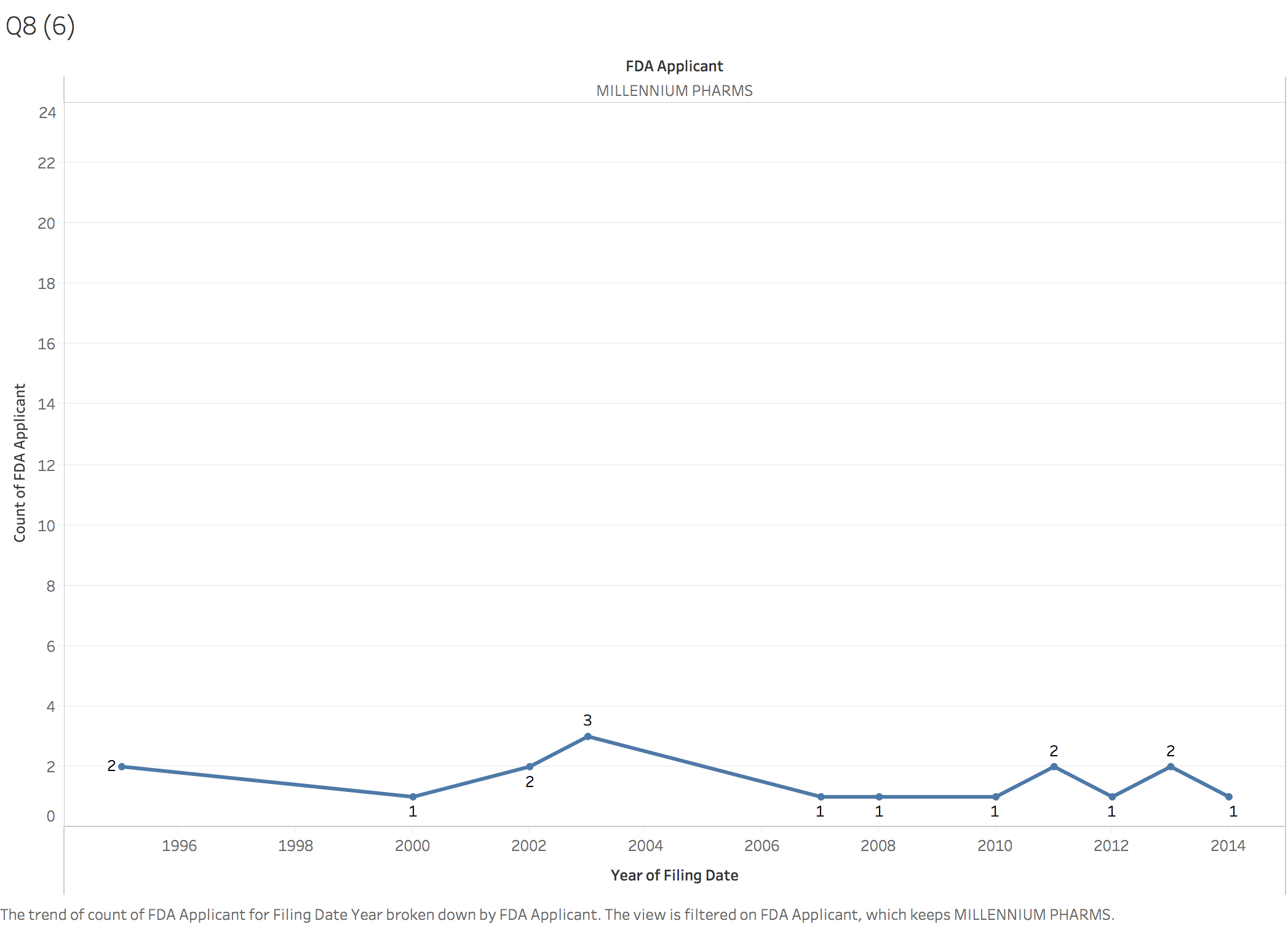
In this graph, it shows the company, INTERMUNE INC; GENENTECH INC, distribution of the FDA applicant documents. It was increased from 2006 and became steady in 2010 and 2011, and then fell down until 2013.



And this is the company named EISAI INC having 17 patent documents. It seems like that being steady from the 1995 to 2012, and the highest points are 2003 and 2012, 3 patent documents.

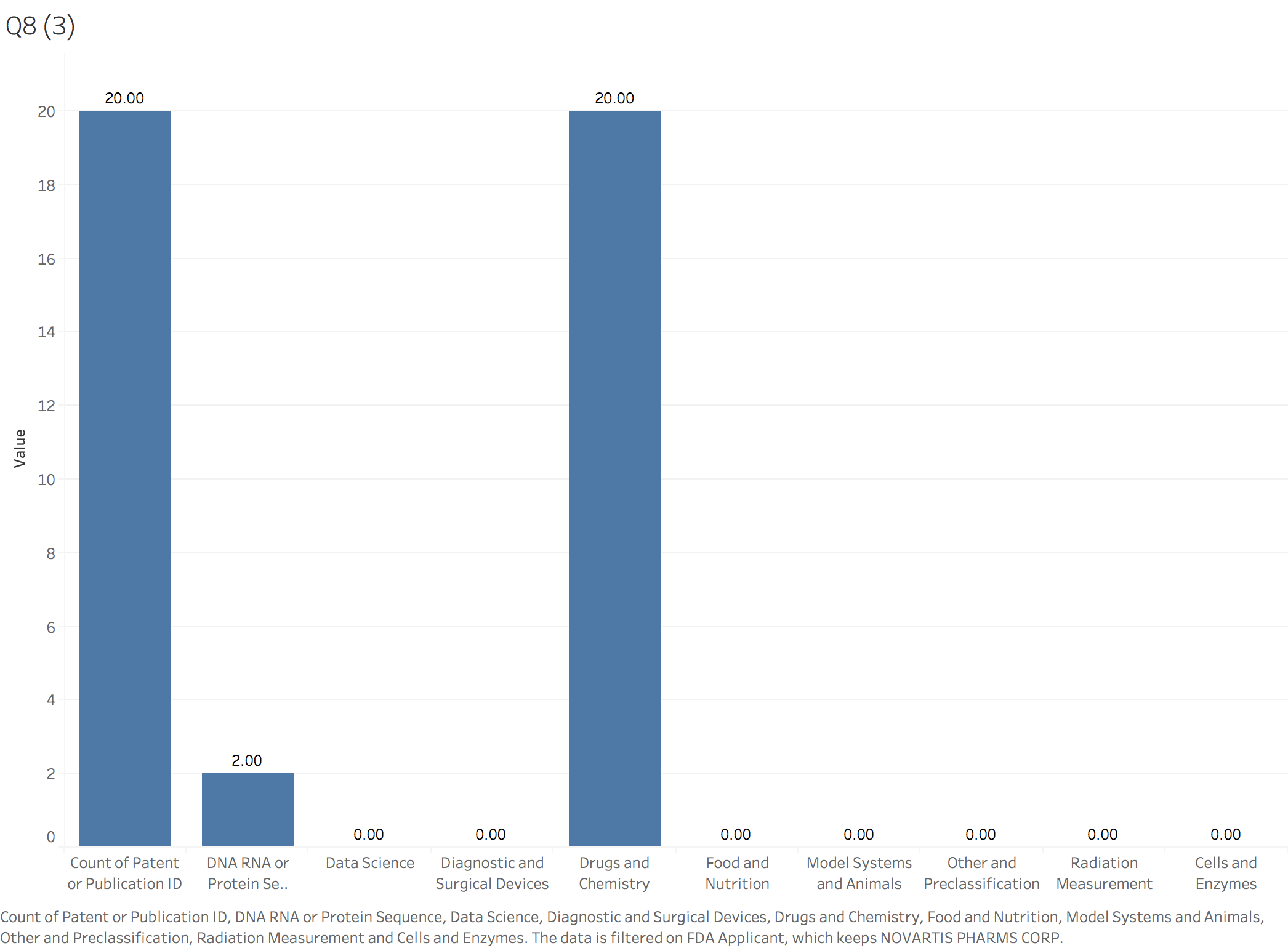


This is the company named WYETH PHARMS INC having 17 patent documents. It is older than others, and had started applying patent documents since 1987, the highest points is 1995, 4 patent documents.

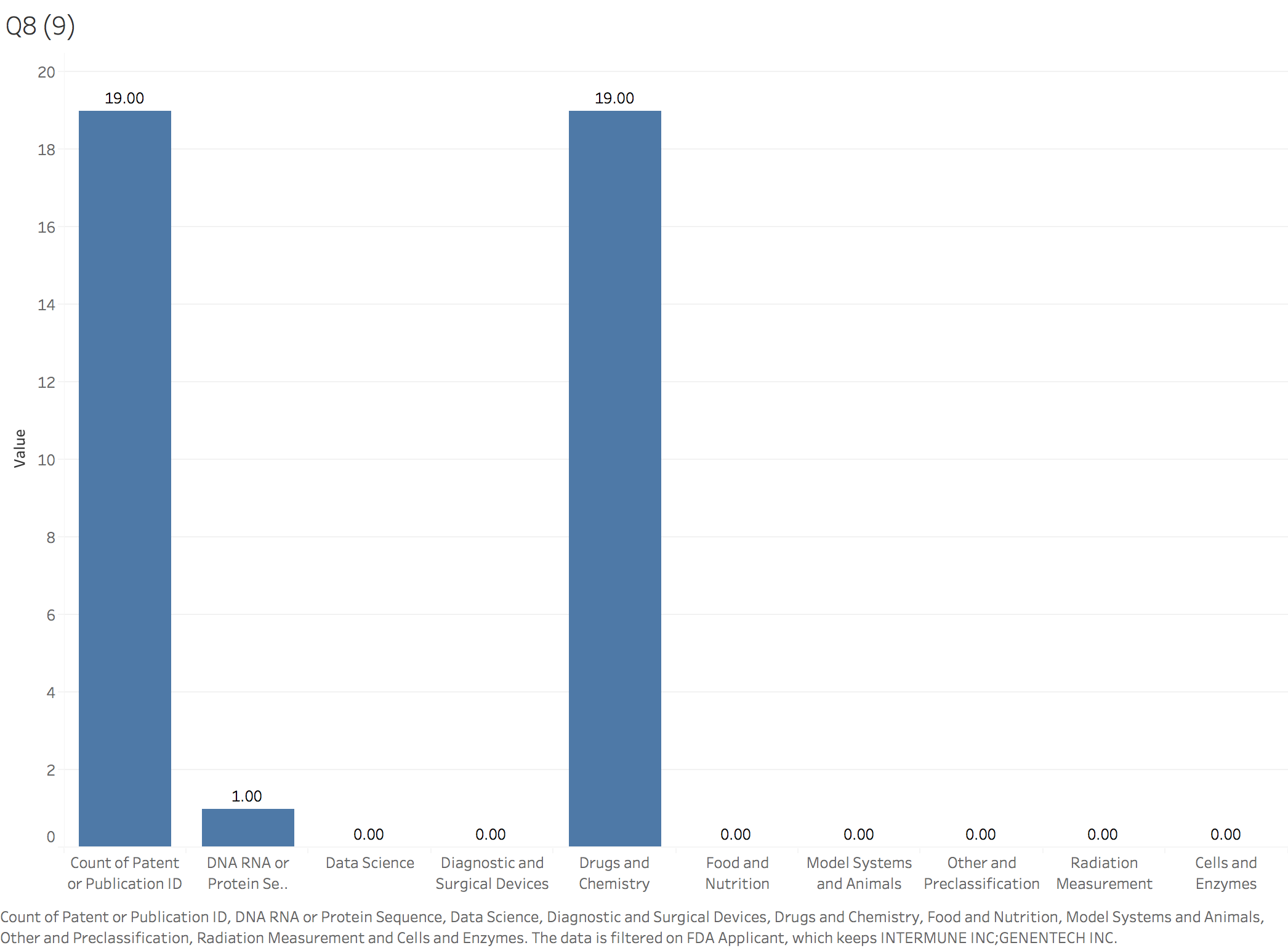


And this is the company named MILLENNIUM PHARMS having 17 patent documents. It similar with the company EISAI INC that being steady from the 1995 to 2014, and the highest points is 2003, 3 patent documents.

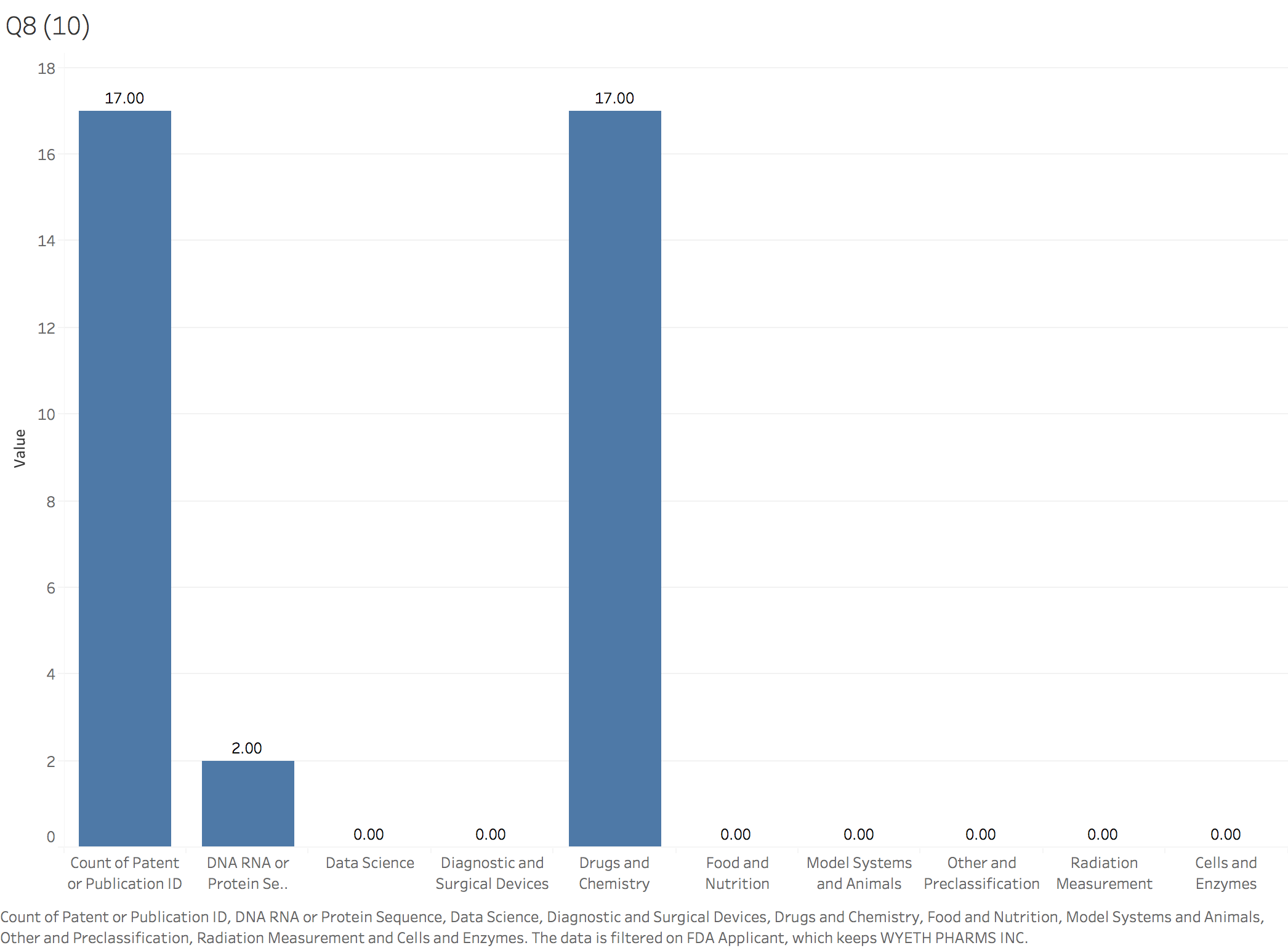
## **FDA companies’ categories**



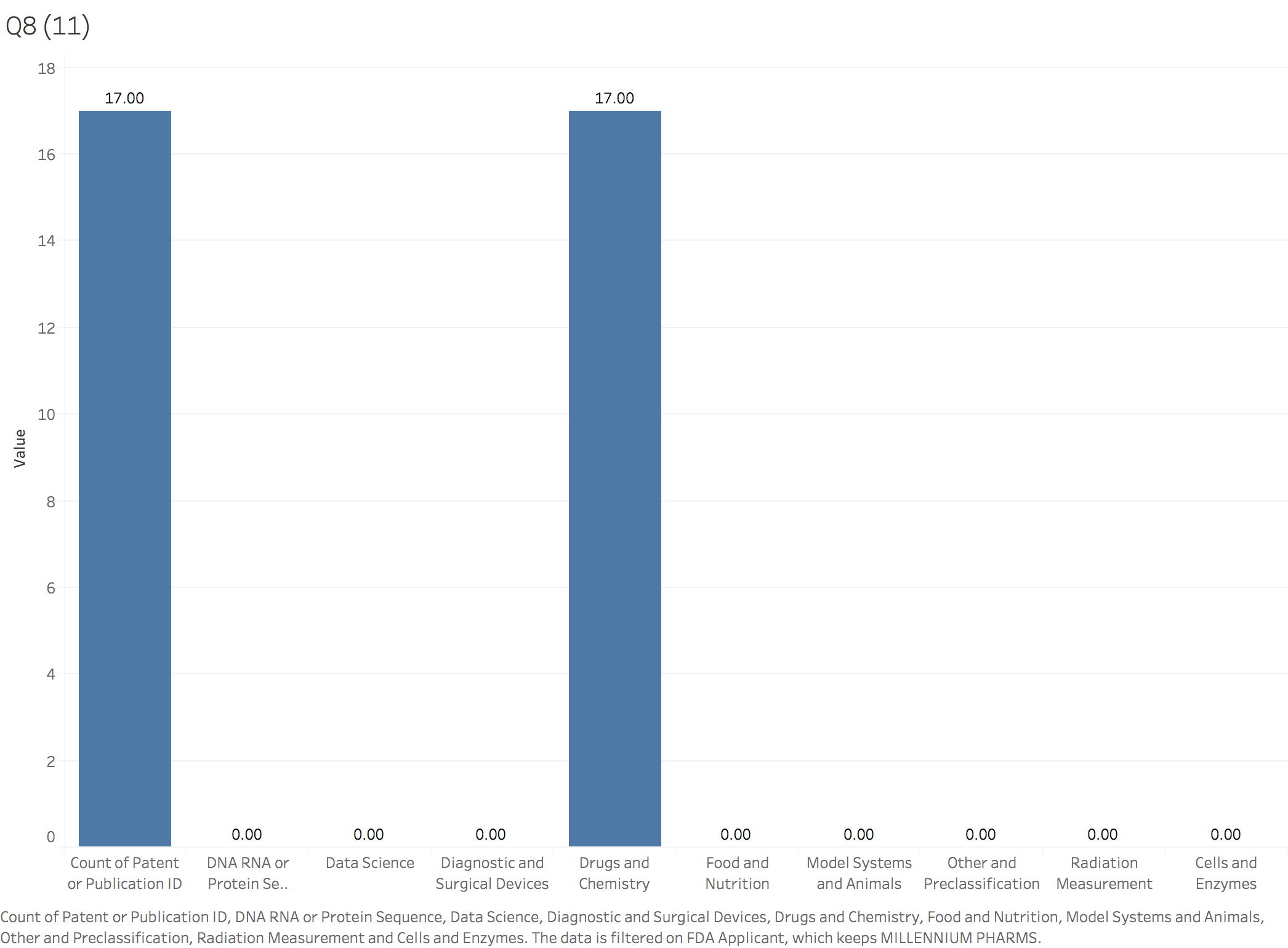
This graph is the NOVARTIS PHARMS CORP patents diversified across categories. The patent categories in total have 2 times of DNA RNA cross the 20 times of Drugs and Chemistry.



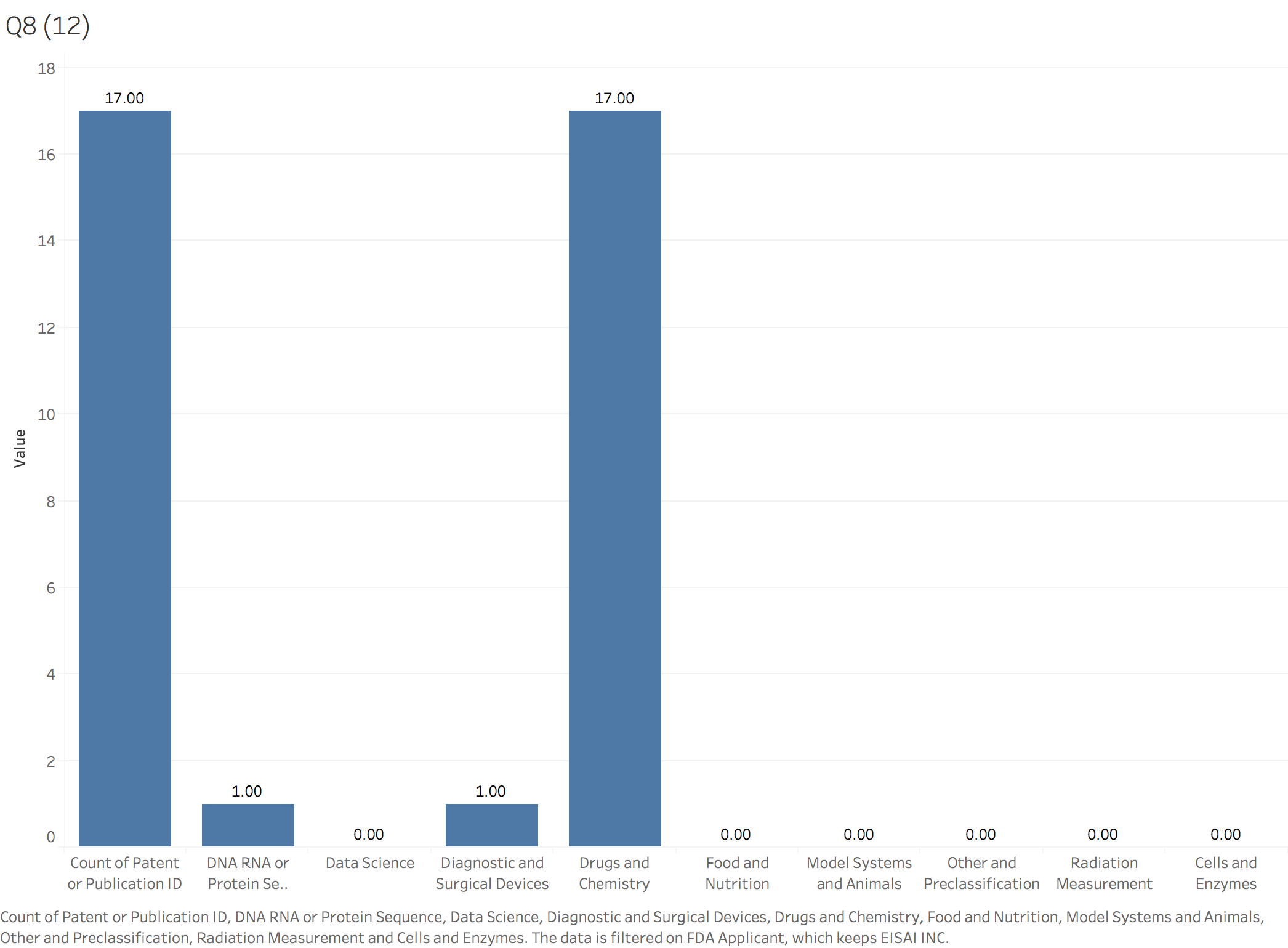
This graph is the INTERMUNE INC; GENENTECH INC patents diversified across categories. The patent categories in total have 1 times of DNA RNA cross the 19 times of Drugs and Chemistry.



This graph is the WYETH PHARMS INC patents diversified across categories. The patent categories in total have 2 times of DNA RNA cross the 17 times of Drugs and Chemistry.



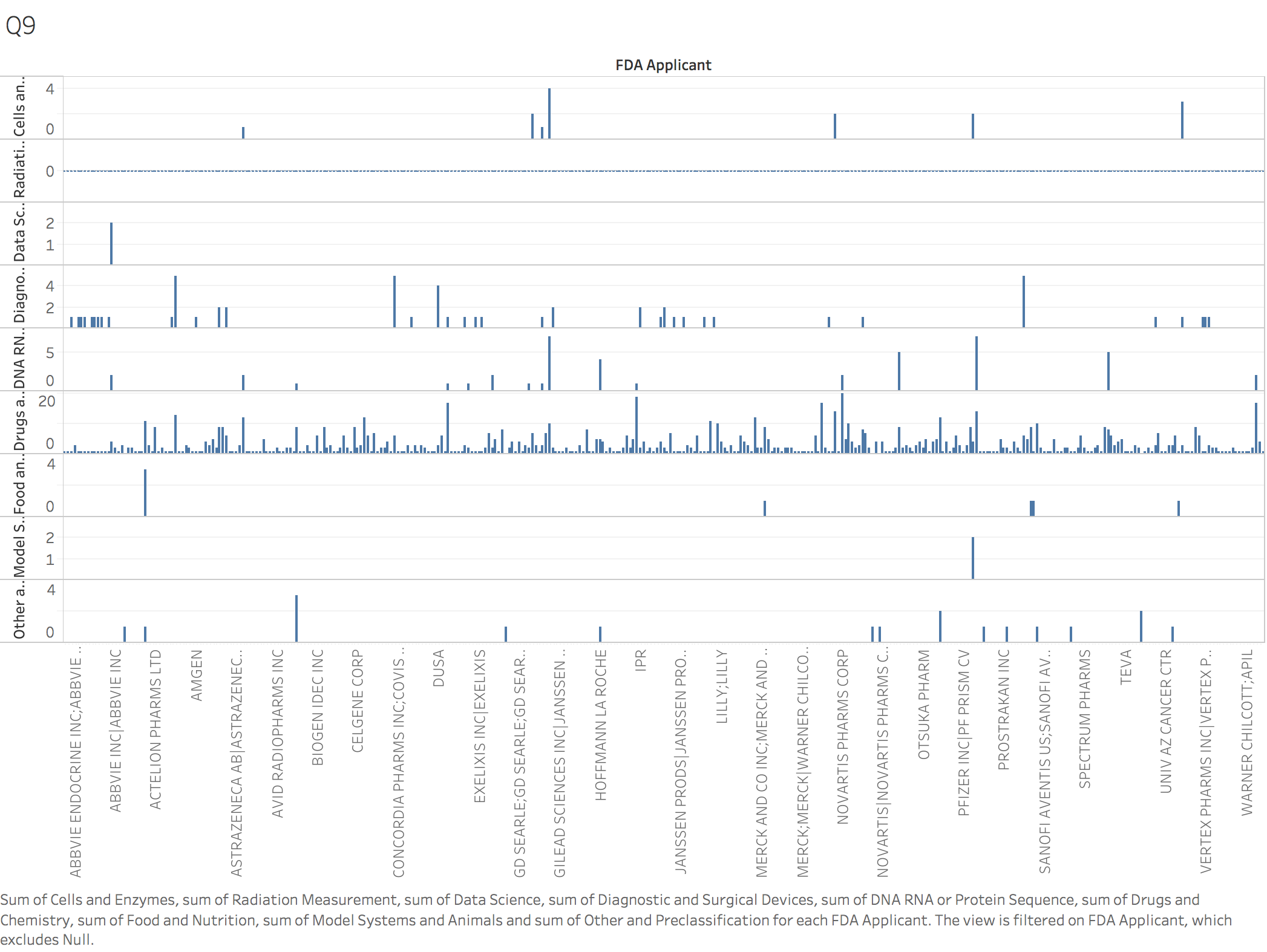
This graph is the MILLENNIUM PHARMS patents diversified across categories. The patent categories in total have the 17 times of Drugs and Chemistry.



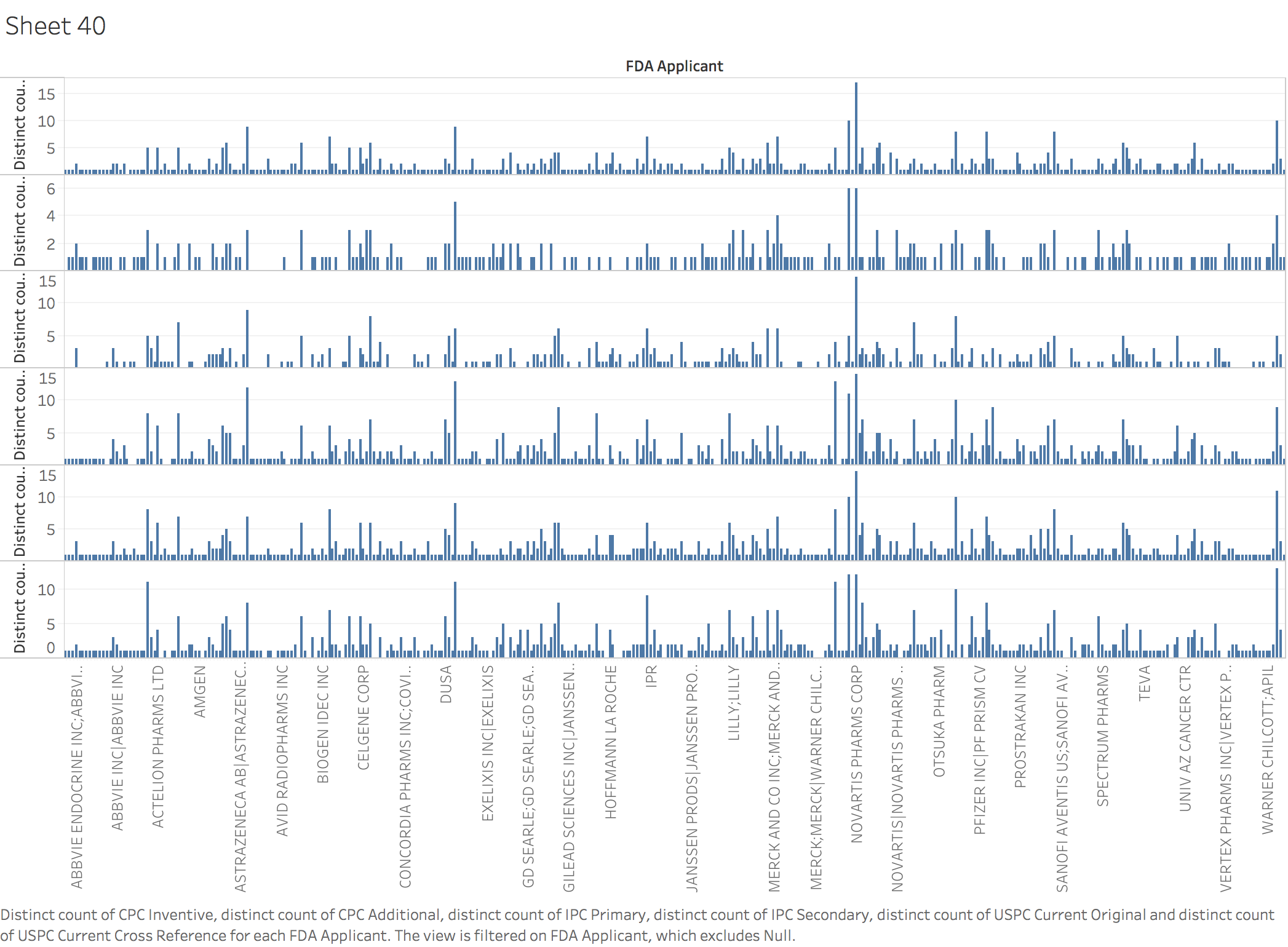
This graph is the EISAI INC patents diversified across categories. The patent categories in total have 1 times of DNA RNA cross the 17 times of Drugs and Chemistry.

In summary, I think EISAI INC has the strongest patent portfolio, because these patent documents in this company involved three types of categories which is more convinced than other companies. Almost companies have two types of patent documents totally.

# FDA Approval different strategies

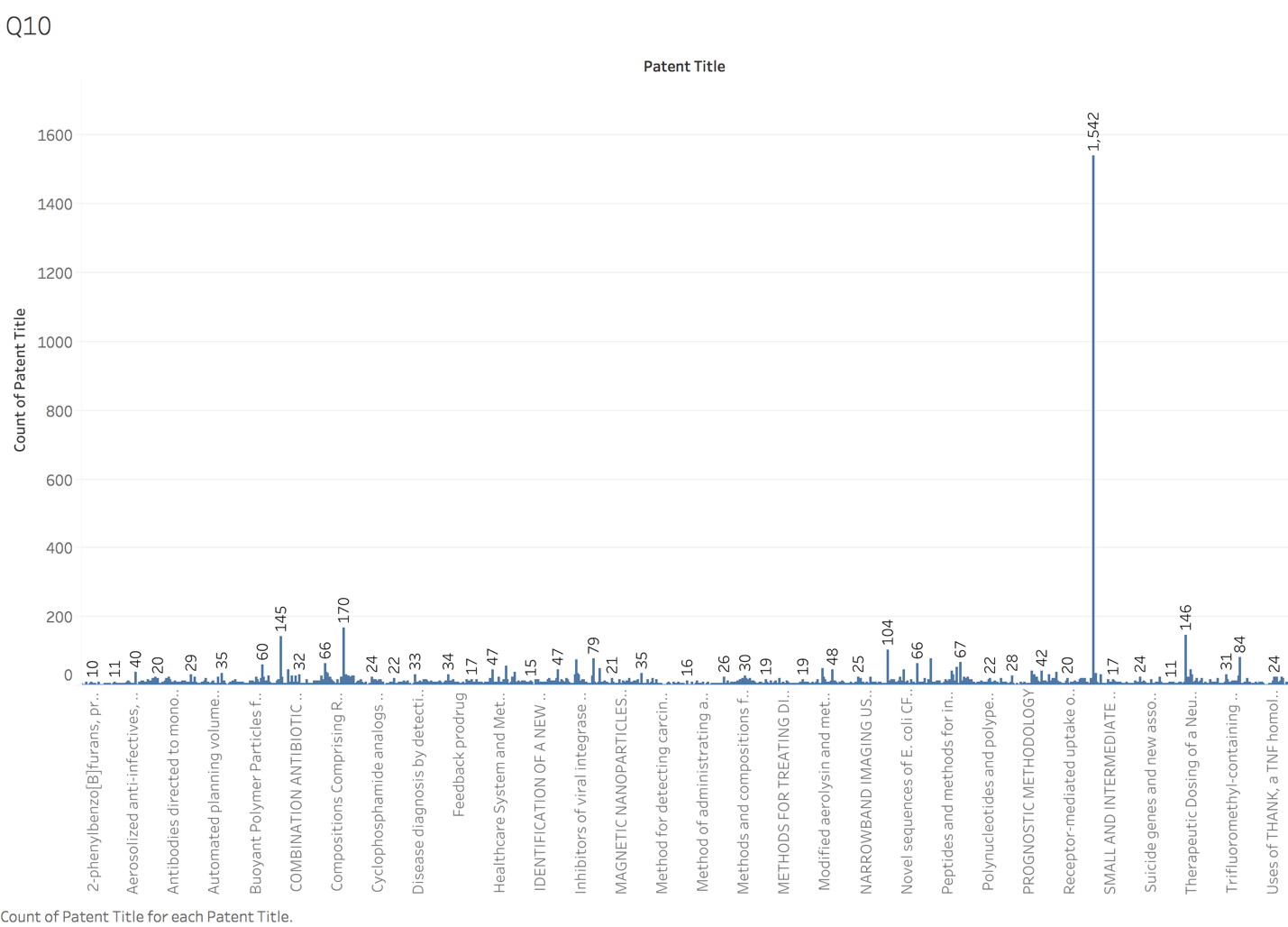


And I put the whole categories together and see that the FDA companies focus on the Drugs and Chemistry more patent documents. Therefore, we can conclude that companies that have received FDA approval have differentiated strategies for the war against cancer, meaning, does their portfolio of patents focus more on Drugs and Chemistry.



In this graph, it shows that the companies that have received FDA approval have the same strategies for the war against cancer based on classification, because of the similar times of patent documents appearing in different classifications.

# The title of patent document



I find in patent title that some titles appearing lots of times. Therefore, I would like to know which patent title appearing the most often. And from the graph, I count the number of patent title appearing times together. Patent Title named “Secreted and transmembrane polypeptides and nucleic acids encoding the same” appearing the most often, and the frequency is 1,542.

# Reference

[1] Trefis, Team. (2018). *Will Roche's Autoimmune Drug Sales Peak In 2018*. Retrieved on November 2, 2018 from

https://www.forbes.com/sites/greatspeculations/2018/03/01/will-roches-autoimmune-drug-sales-peak-in-2018/#4219942578fb

[2] *Esbriet Prices, Coupons and Patient Assistance Programs*. Retrieved on November 2, 2018 from

https://www.drugs.com/price-guide/esbriet