BU Law | Patent Authorship Team 2

Team Members:

Seonghoo Kim

Melitza Cuellar

Ji Zhao

Aastha Pal

1. Introduction

Founded in 1872, Boston University School of Law is a top-tier law school with a faculty recognized nationally for exceptional teaching and preeminent scholarship. You can explore virtually any area of the law in 200+ classes and seminars, 20+ clinics and externships, and 20 study abroad opportunities. BU Law offers a full-time JD degree, five LLM & master's degrees, and 17 dual degrees. With the support of a global network of nearly 24,000 alumni and a robust Career Development & Public Service Office, our graduates achieve remarkable career success. BU Law is located in the heart of Boston and housed in the ultra-modern Sumner M. Redstone Building and Law Tower.

Professor Jordana Goodman from the Boston University School of Law was interested in the involvement of women in the entire process of patent applications, starting from the inventors, assignees, to the inventors. A study by the UK's Intellectual Property Office (IPO) found that female inventors accounted for under 13% of patent applications. Almost 70% of all patent applications had all-male teams or individual male inventors, and a study by Yale found that women may be less likely to be granted for their applications. This prompted Professor Goodman to dive deeper into the history of women in the entire patent application process. Given the task of analyzing this information, we gathered patent applications sent to the United States Patent and Trademark Office (USPTO) from the years 2000 to 2020 to analyze and come up with our own findings.

Some of the key questions we aim to answer with this analysis are:

1. **How has the proportion of female inventors changed over the years and what is causing such a low proportion?**
2. **Is there greater gender diversity of attorneys on domestic-based applications than foreign-based applications?**
3. **Do foreign-based applications seem to "give up" more claims than domestic applications?**
4. **Are there some assignees who are more popular and what is the proportion of female inventors in companies with the highest patents?**

`

2. Data and code description

We extracted the data from the Patent Examination Research Dataset (PatEx) | USPTO which contains detailed information on nearly 11.9 million publicly-viewable provisional and non-provisional patent applications to the USPTO and over 4.6 million Patent Cooperation Treaty (PCT) applications. It is based on data that OCE downloaded from the Patent Examination Data System (PEDS) in April, 2021. The PEDS data are sourced from Public PAIR.

Original Datasets:
We used the following csv files from the Patent Examination Research Dataset: application_data, all_inventors, attorney_agent, patent_claims_stats and foreign priority.

Clean Datasets:
The clean datasets can be found in the Cleaned Data folder which has application data, inventors data, and attorney/agent data for patent applications from 2000 to 2020 with the gender classification feature.

Code:
The code for the clean datasets and the generated graphs used in our analysis can be found in the Project Code folder in the form of multiple ipynb files. The documentation to follow along with the code can be found in the Documentation for Project Code file.

Tools:
- Gender-Guesser module in Pypi
- WGND 2.0 - World Gender Name Dictionary.

3. Analysis

a. Proportion of Female Inventors Within Patent Applications

Question: How has the proportion of female inventors changed over the years and what is causing such a low proportion?

We start our analysis by looking at the fundamental blocks of who plays a role in patent applications: inventors. With close to 10 million patent applications filed in the years 2000 to 2020, every single patent application had at least one inventor with most of the time having multiple inventors per application. With the millions of inventors who were given credit for their work in their patent applications, we wondered how many of them were female inventors, and if there was any improvement in the proportion of female inventors from 2000 to 2020. Therefore, we plotted a heatmap of the top 30 countries that submitted the most number of patent applications to USPTO and their female inventor ratio. This is shown in the figure below:

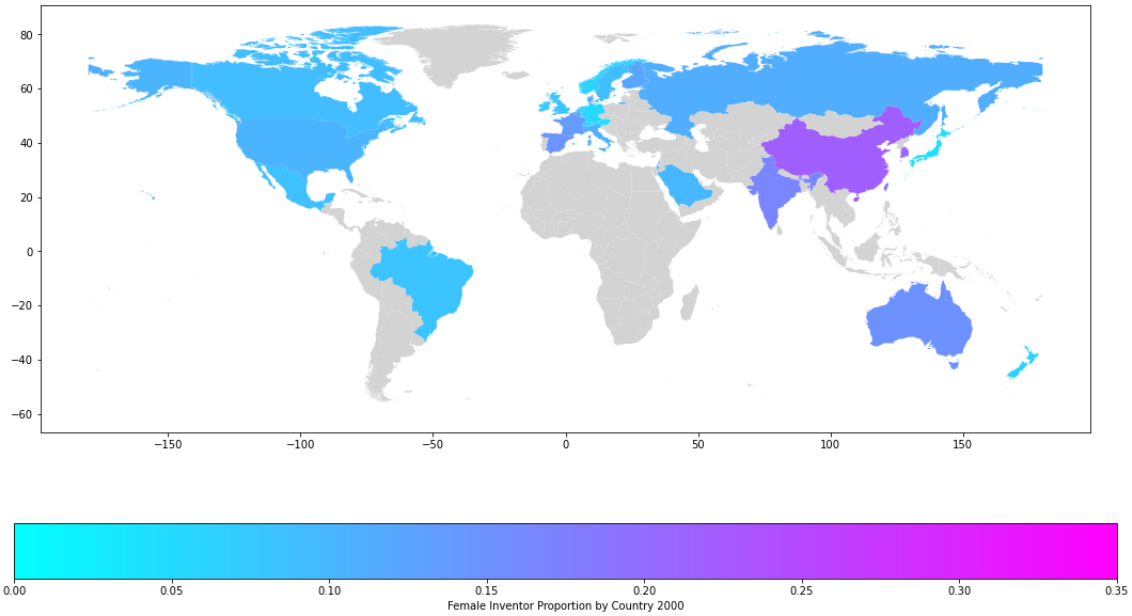

Figure 1. Female inventor proportions based on USPTO patent applications 2000

As seen from this heatmap, we can see that most countries are failing to have over 15% of their inventors be female. Knowing the increased importance of female representing the industry as a whole, we plotted the female inventor proportions 20 years later as seen in the figure below:

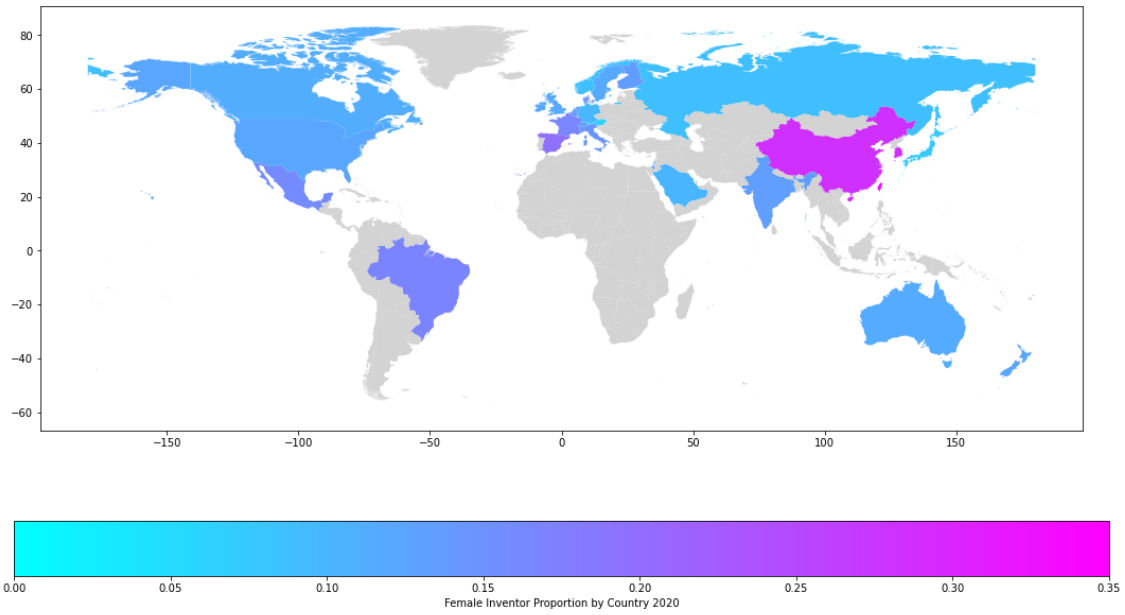

Figure 2. Female inventor proportions based on USPTO patent applications 2020

This time we can see that most of the top 30 countries had an increase in the proportion of female inventors in their patent applications, but the countries were still hovering between 10 to 15%. As a whole, the female inventor proportions started at around 9.5% in 2000 to 12.5% in 2020.

An outlier from this data is the fact that Korea and China have strangely high female inventor ratios compared to the rest of the world. A part of this problem was due to the fact that Korean and Chinese names are based off of characters, which may sound the same but have completely different meanings. Therefore, two homophonic names may be both feminine or masculine depending on the meaning of the name, which is still being researched to find the optimal solution to this problem. Another reason for the high proportions were due to government policies set by the two countries to incentivize female inventors. For example the Korean government in 2002 implemented the Act on Fostering and Supporting Women in Scientists and Technicians to increase female proportion in higher education. Moreover, the Korea Women Inventors Association was created in 1999 to provide resources to female inventors to foster new patents. Meanwhile in China, female researchers are given priority for talent recruitment and research funding grants, making it more accessible for female inventors to be involved in the patent process. This pressure from government policies as well as parental influence from these policies to keep women in STEM explain the high proportion of female inventors in these two countries.

After looking at the top 30 countries, we wanted to see what was responsible for the low female inventor proportions. Therefore, we first looked at the highest classification possible of patent categories, which consisted of Design, Plant, and Utility patents. We plotted the number of female inventors per year for the three patent categories to see if there was a specific category that was stagnant or was decreasing. This was plotted below.

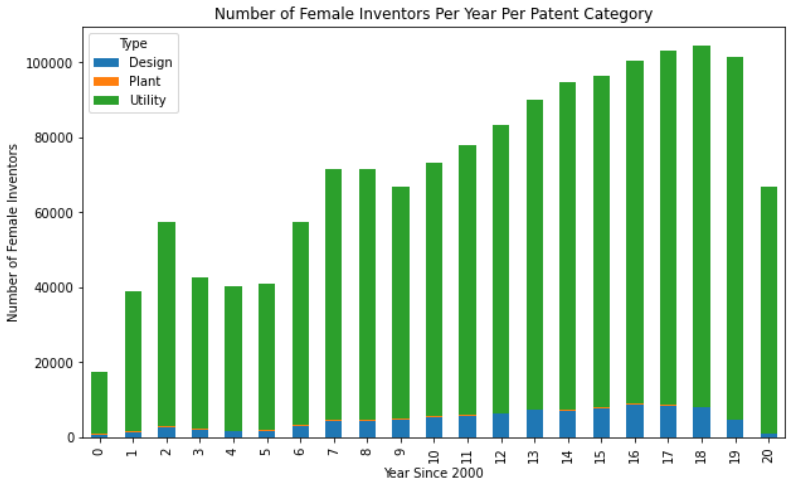

Figure 3. Number of female inventors per year based on USPTO patent applications

This plot shows that generally the trend of patents seemed to line up with the general increase in patent applications, meaning that although there is an increase in female inventors, it is increasing at a similar rate as male inventors. Since Utility patents seemed to be the main source of patent applications, we decided to look into Utility patents in particular, where we were curious as to which industries had the largest number of female inventors. Therefore, we decided to plot pie charts for the years that had a full year of data from the past and the present. The pie charts are shown below.

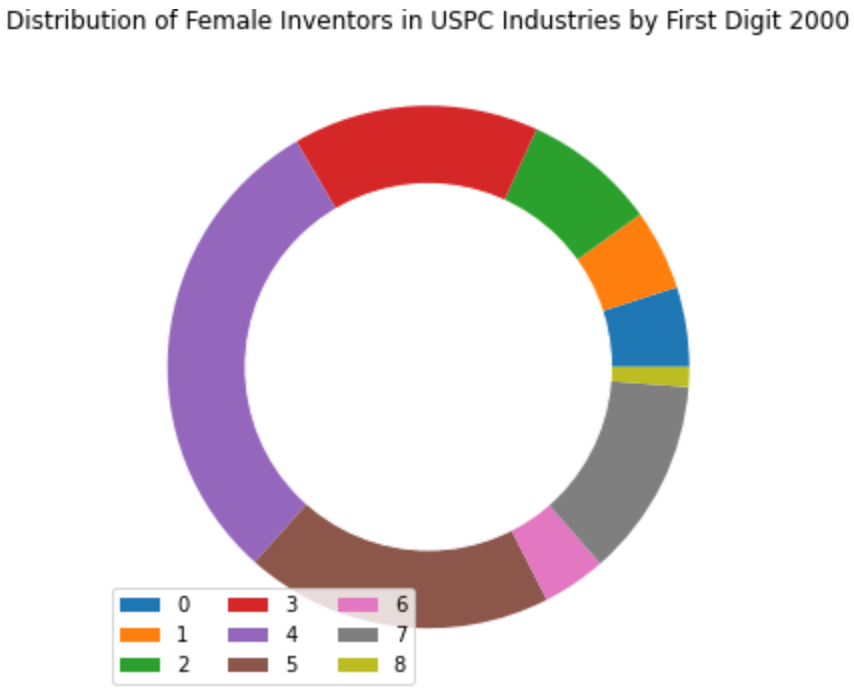

Figure 4. Distribution of female inventors in USPC industries by first digit 2000

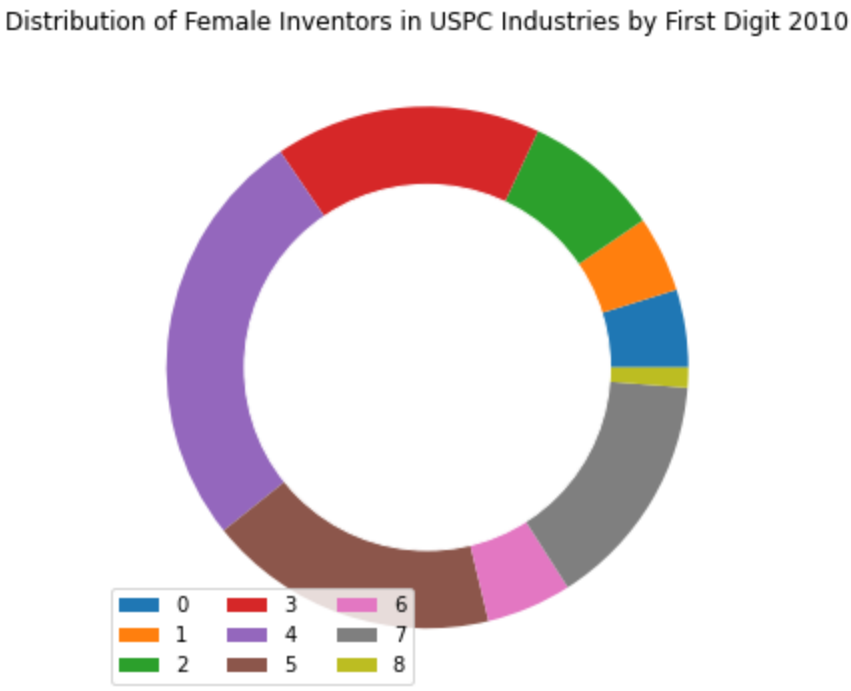

Figure 5. Distribution of female inventors in USPC industries by first digit 2010

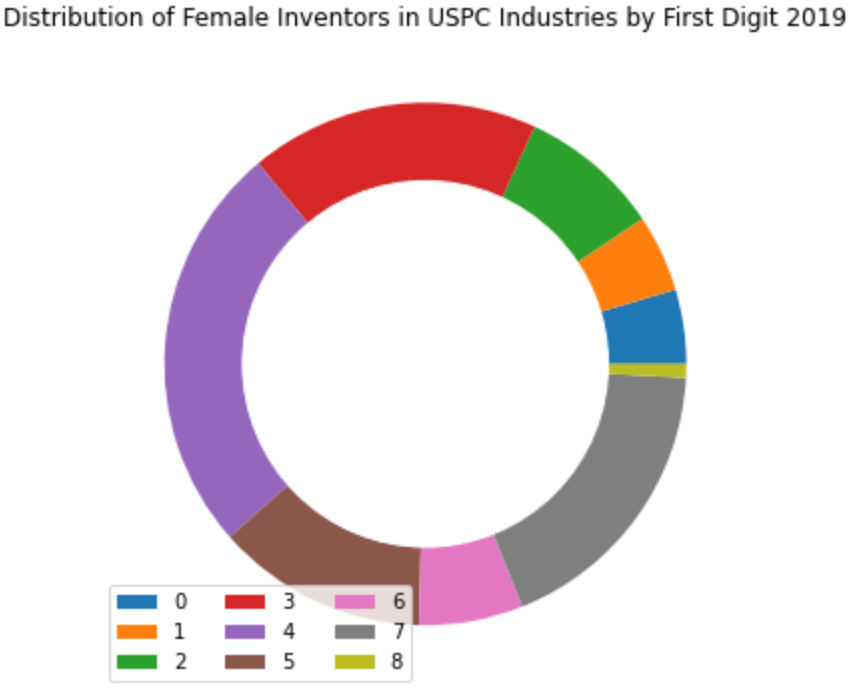

Figure 6. Distribution of female inventors in USPC industries by first digit 2019

We can see that the percentage of female inventors in the 500 category patents decreased compared to the rest of the industry, while 700 category patents increased in the past 19 years. 400 and 300 category patents remained similar while being the biggest industries. Looking at the 400 category patents, we can

see that they are related to chemistry, earth sciences, and biology, which are all fields of higher education research. Looking at the proportion of female inventors that were involved in these patents, we can see that in 2019, there were 24,545 female inventors over a total of 145,960 inventors in total, resulting in a percentage of 16.8%. This is quite alarming in the underrepresentation of female inventors in patent applications compared to the number of females in this industry, as the US Bureau of Labor Statistics in 2021 showed that out of the 728,000 people employed in scientific research and development services, 44.7% were female. This raises the question of why females are underrepresented in patent applications, even though they are almost half of the workforce in these industries.

b. Attorneys/agents' Gender Ratios for Domestic vs Foreign Patent Applications

Question: Is there greater gender diversity of attorneys on domestic-based applications than foreign-based applications?

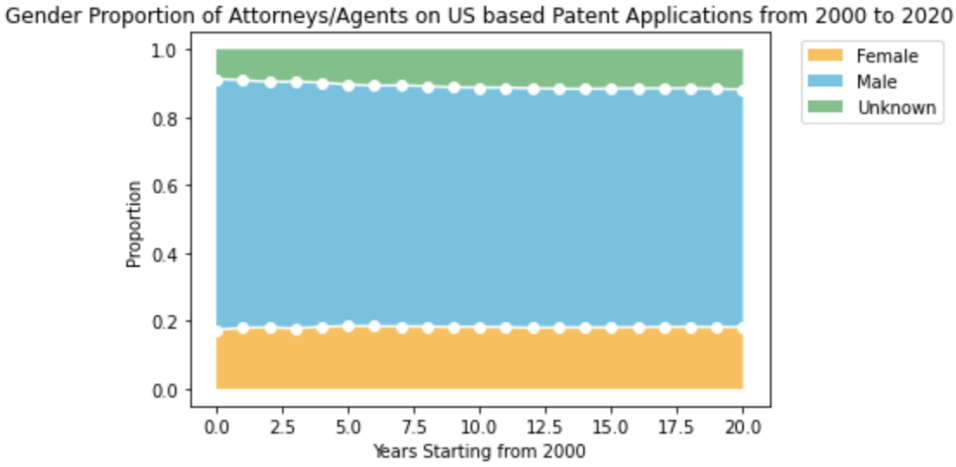

Figure 7. Gender proportion of attorneys/agents on US based patent applications (2000 - 2020)

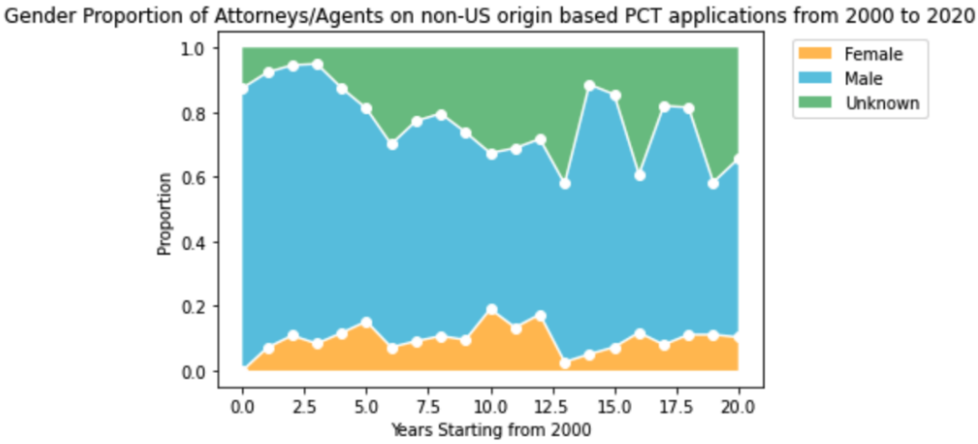

Figure 8. Gender proportion of attorneys/agents on non-US based patent applications (2000 - 2020)

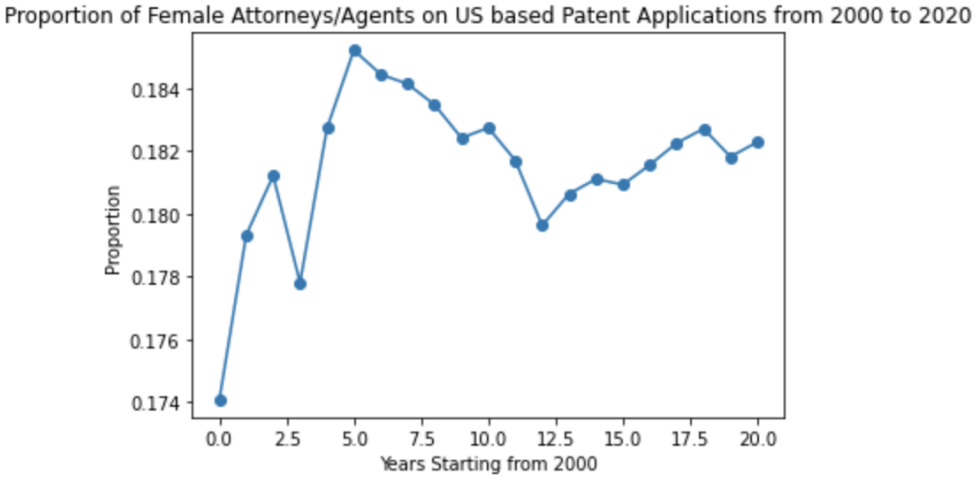

Figure 9. Proportion of attorneys/agents on US origin based PCT applications (2000 - 2020)

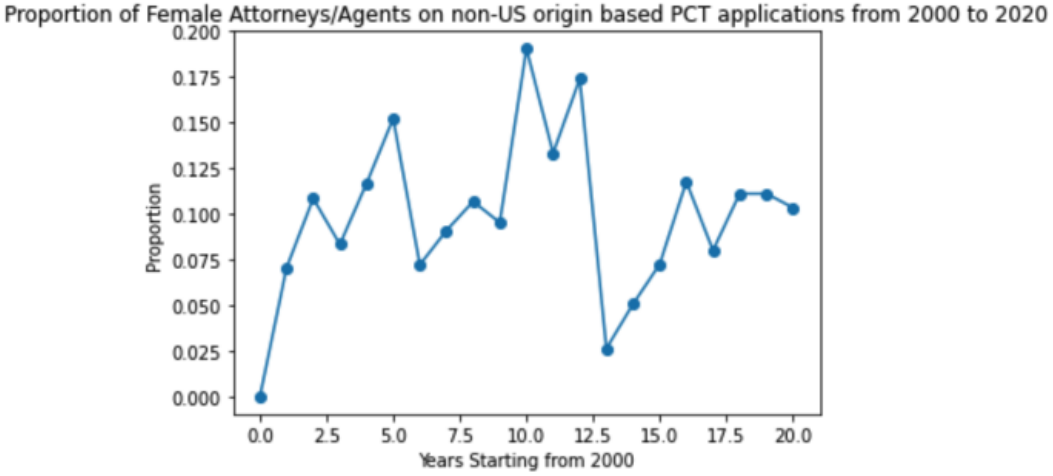

Figure 10. Proportion of attorneys/agents on non-US origin based PCT applications (2000 - 2020)

We decided to group attorneys and agents together because there are few attorneys/agents on foreign applications (non-US origin based PCT applications). There were only 1,727 first names given for attorneys and agents on foreign applications (16,574 foreign applications out of the approximate 10 million patent applications shared by the USPTO over the past 20 years) versus approximately 215 million attorneys/agents first names listed on domestic applications (applications filed with the USPTO and US-origin PCT applications). As shown in the graphs above, female attorneys and agents made up about 20% of all attorneys and agents from 2000 to 2020. The proportion of female attorneys and agents on foreign applications fluctuates throughout the years. The proportion of about 20% in 2010 dropped to 3% in 2013 and went up to 10% in 2020. Domestic applications had a greater gender diversity of attorneys than foreign-based applications. However, there were more names unassigned to a gender for attorneys on foreign-based applications so this might not be the case. Furthermore, the foreign applications captured by the USPTO office were those looking to register for a patent in the US and therefore, were only few in number.

c. Claims numbers of foreign-based applications and domestic applications

Question: Do foreign-based applications seem to "give up" more claims than domestic applications?

One of the questions we want to answer in the project is: do foreign-based applications seem to "give up" more claims than domestic applications? For this question, a deeper understanding is still needed. In this part, we began to explore this question.

In the previous study, we processed the "all_inventors" dataset. A gender column and a year column were added to the dataset. Our analysis was also based on this dataset. The difference was that other datasets were applied to do this analysis. Patent numbers were extracted from "application_data.csv" and the numbers of claims were from 'patent_claims_stats.csv'.

Firstly, the numbers of claims and patent numbers were extracted from file 'patent_claims_stats.csv'. The claims numbers were averaged and were merged into the "all_inventor" dataset by the patent id. However, only a portion of the data could be matched, so we used this set of data for a preliminary analysis. Secondly, the countries of inventors were divided into the "US" group and "Foreign" group to seek some differences among the different country-basement applications. However, the 'patent_claims_stats.csv' only contained data before 2015, the data after 2015 was still unknown.

The following figure showed the trends of average claims numbers of "Foreign" and "Domestic" ('US') applications.

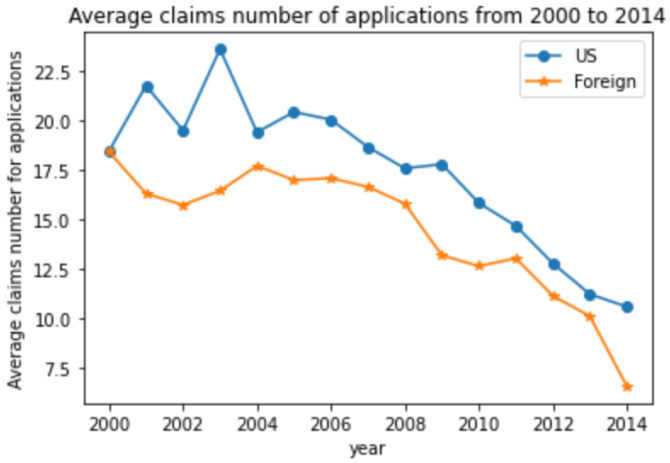

Figure 11. The average claims number of applications from 2000-2014

As shown in the above figure, the average claims number showed a downward trend in both "Foreign" and "Domestic" groups with respect to time. It showed that people gradually tend to use fewer claims. In general, the "US" curve was always above the "Foreign" curve which indicated that foreign-based applications might "give up" more claims than domestic applications. The difference between the two had nothing to do with time.

To have a deeper understanding of the distribution of the difference in the average claims number, we firstly ranked the average claims numbers of different countries within various years and plotted the graph as follows.

The following figure showed the trend of the top 10 countries (ranked by the average claim number) and the US. We found that Russia had a significant uptick in 2003-2004. However, it followed a sharp decrease. Overall, at the national level, the downward trends of different countries were not particularly obvious.

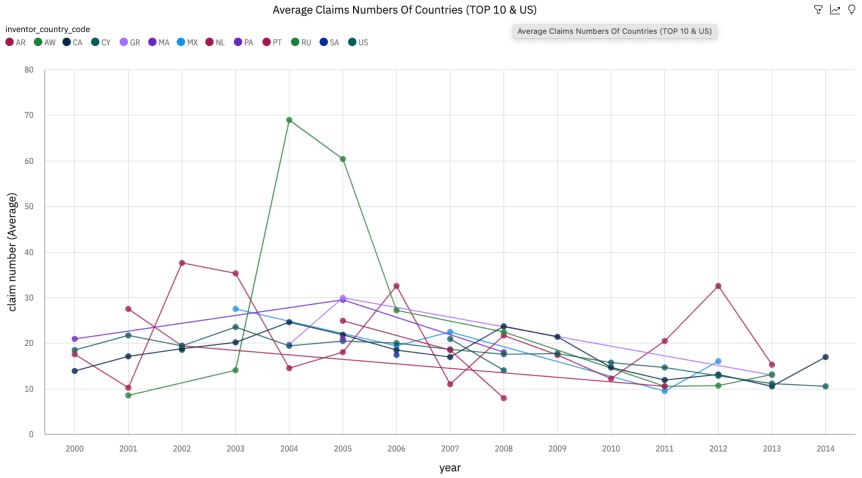

Figure 12. The average claims numbers of countries (Top 10 & US)

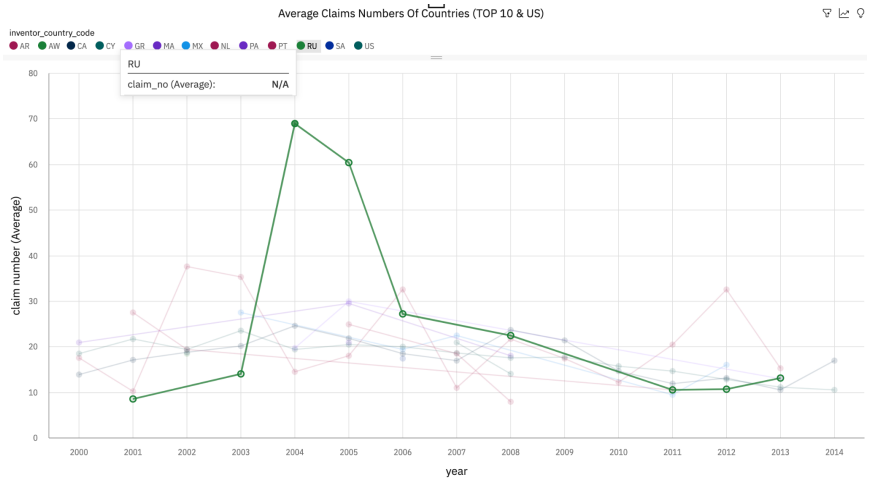

Figure 13. The average claims numbers of Russia

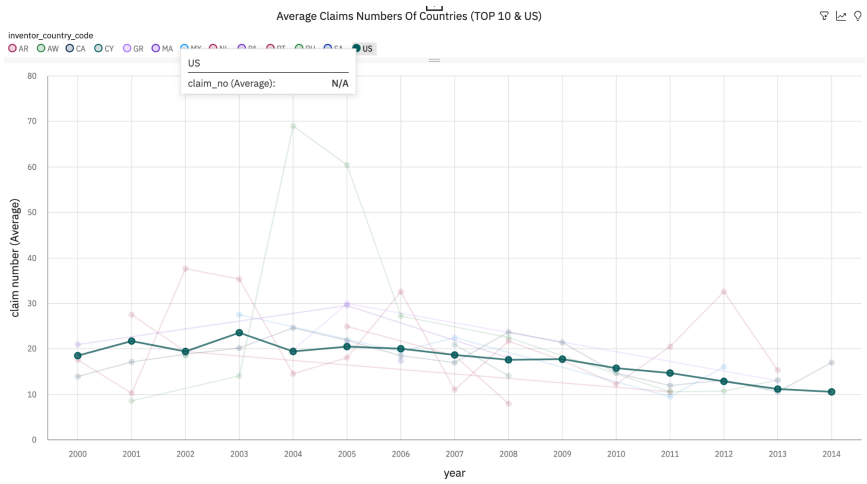

Figure 14. The average claims numbers of the United States

The average claims number didn't change a lot as time went by. This may be related to the relatively perfect patent application system in the United States, as well as the higher level of scientific research. In the data from 2000 -to 2014, compared to other countries, the United States didn't not stand out in terms of the claims number. So, the right answer to the question - "Do foreign-based applications seem to 'give up' more claims than domestic applications?"- foreign countries seem to not "give up" more claims than domestic applications at the national level.

Next, we analyzed at the overall level, regardless of time. The following map showed the distribution of different average claims numbers in countries. We could also see from the chart that the United States didn't rank very high compared to other countries. Georgia had the largest average claims number, but it may be because of the tiny data in Georgia. Pakistan had the lowest average claims number. There was no significant difference between developed countries.

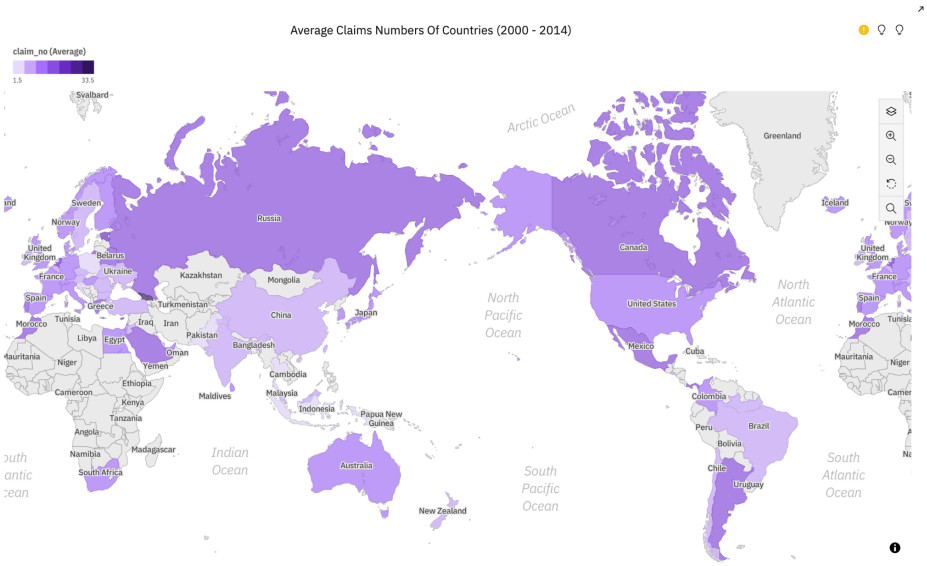

Figure 15. The distribution of different average claims numbers in countries (2000-2014)

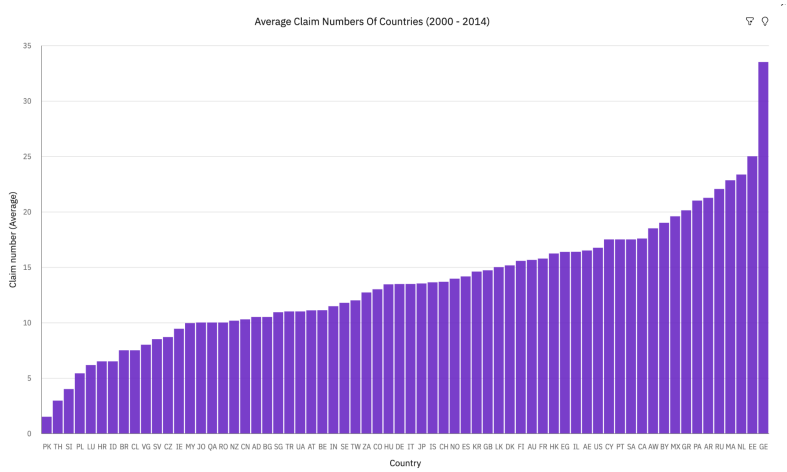

Figure 16. The data of different average claims numbers in countries (2000-2014)

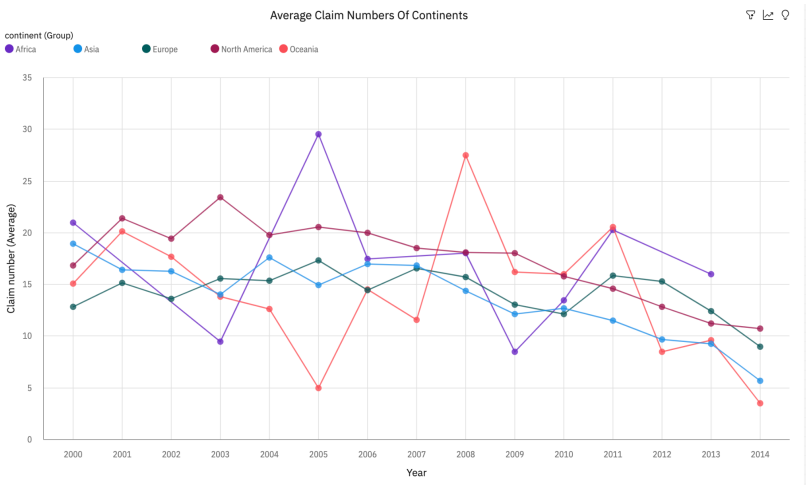

Figure 17. The different average claims numbers of continents by years (2000-2014)

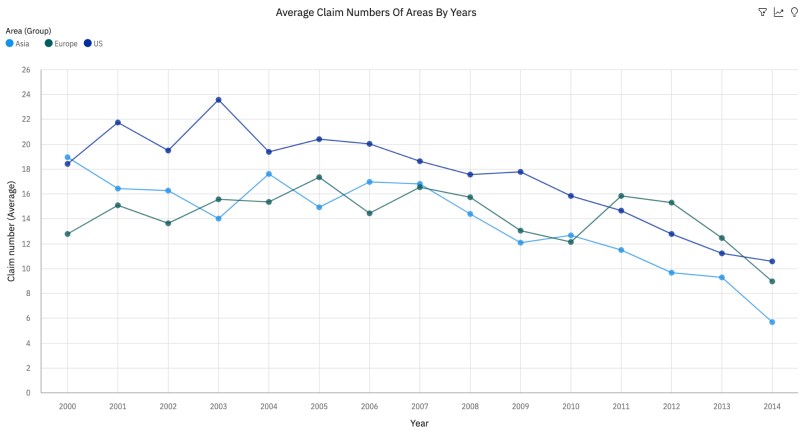

Figure 18. The different average claims numbers of areas by years (2000-2014)

The above figures showed the trend of average claims numbers of continents and areas (US, Asia, Europe). **On the one hand, at the continental level, significant fluctuations occurred in Africa and Oceania. The average claims numbers of other continents presented a downward trend. On the other hand, compared to Asia and Europe, the average claims number in the US was higher almost all the time which was similar to the analysis based on the "US" and "Foreign" groups that foreign-based (Asian and European) applications seem to 'give up' more claims than domestic applications.**

d. MOST POPULAR ASSIGNEE AND GENDER Distribution

Question: Are there some assignees who are more popular? What is the proportion of female inventors in companies with the highest patents?

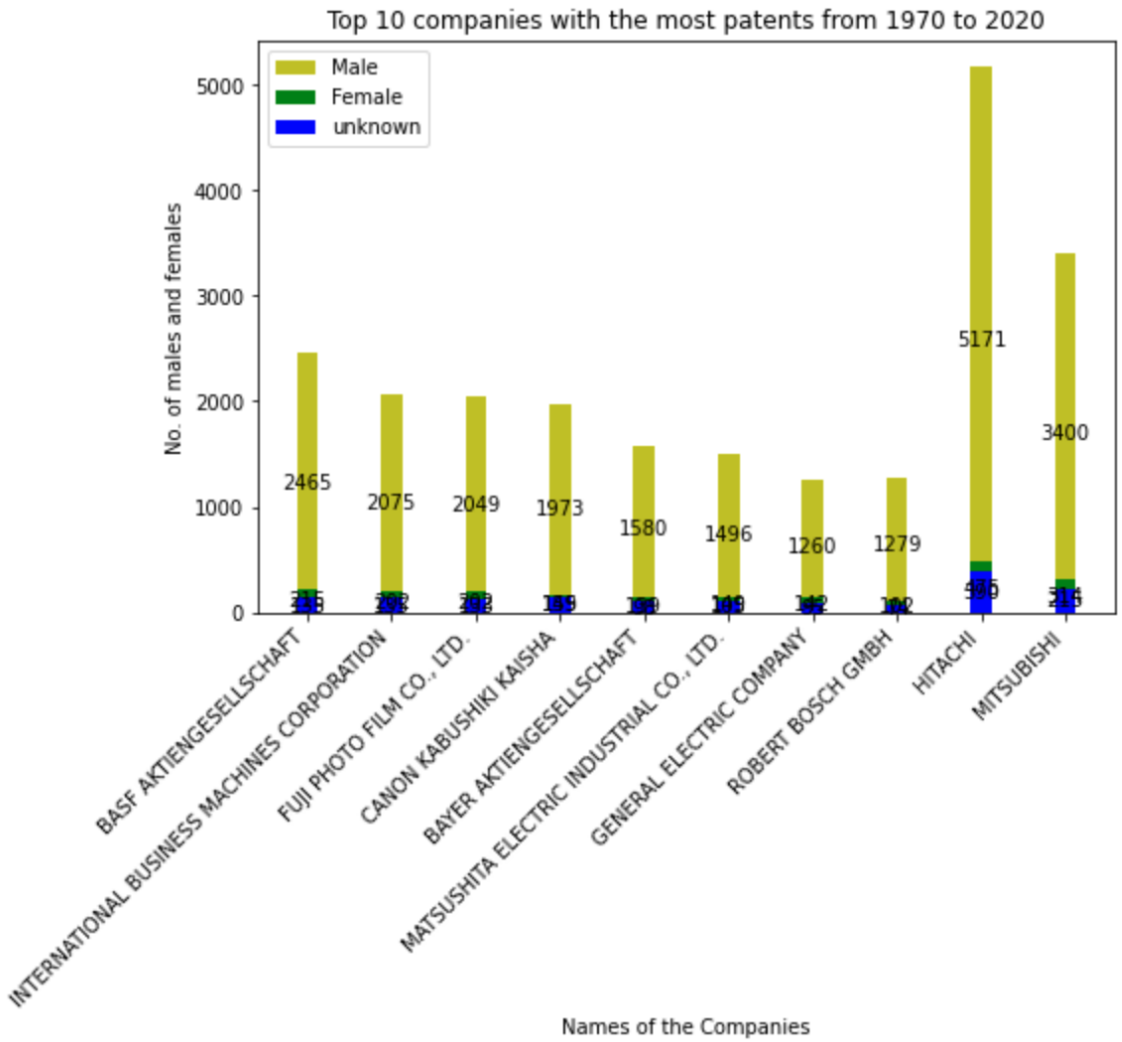

Figure 19. Top10 companies with the most patents (1970 - 2020) (No. of genders)

In this analysis, we looked at assignee and assignor data to look for companies with the most patents. For this analysis, we merged the inventors data with assignee and assignor to get the names of the companies and the gender of inventors. We found that Hitachi Mitsubishi, Basf Aktiengesellschaft and IBM are among the top companies with the highest no. of patents.
These companies are developing electrical, automobile products, and chemical & plastics that have a lot of innovation going on to release more sophisticated products in the market. We see self-driven cars, new chemical formulations for drug repurposing, and electronic products quickly coming up because people are investing time and money to innovate and design better products that lead to innovations and hence they are able to get patents for their novel products.

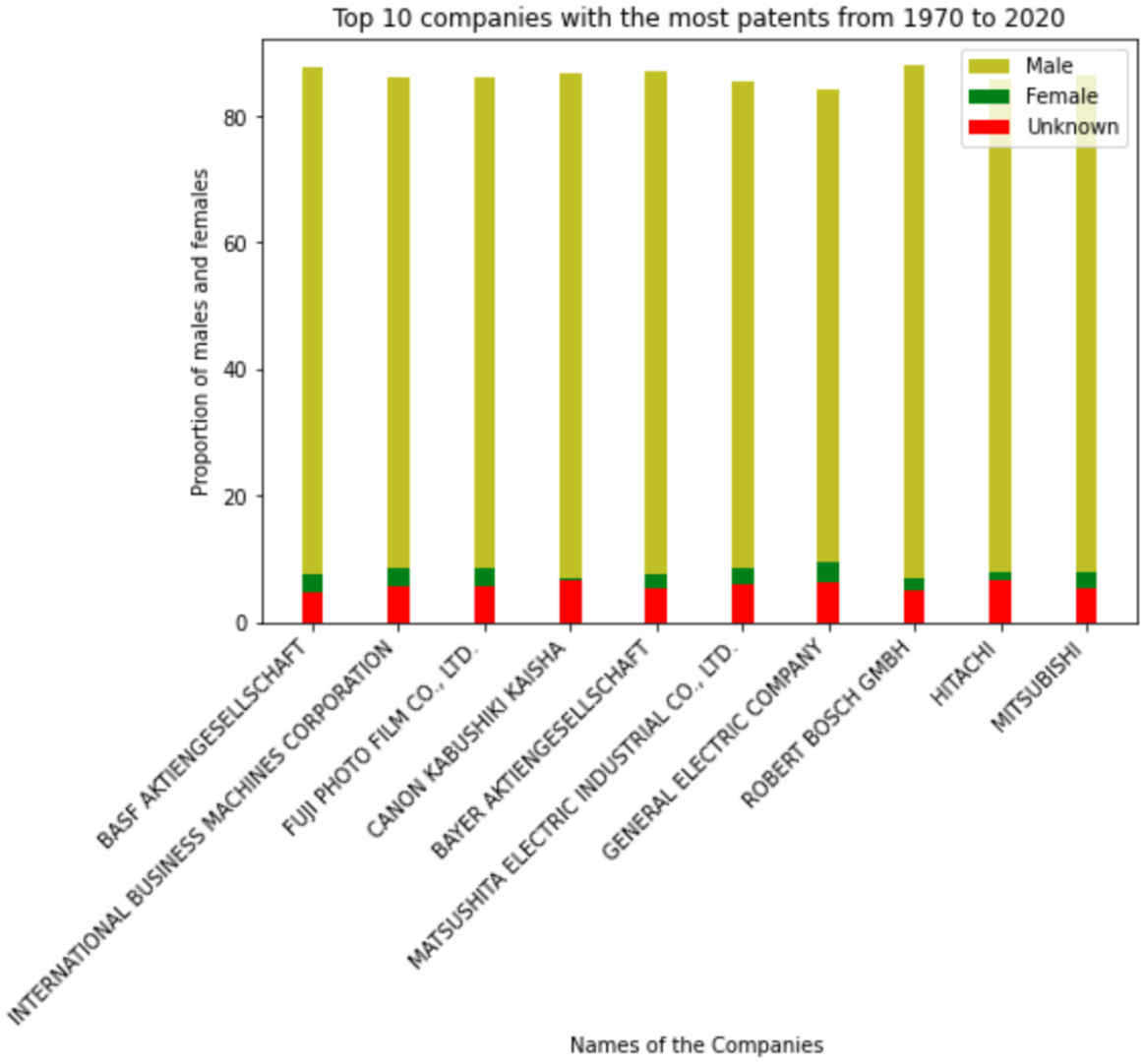

Figure 20. *Top10 companies with the most patents (1970 - 2020) (Gender proportions)*

We further explored the data which has 25 million entries to see proportions of males, females and unknown genders in the 600514 unique companies with the highest number of patents. We found 5 categories in the gender column but we only looked at male female and unknowns in order to find the distribution and compare the male-female proportion. For this analysis, we created a dictionary to get the company names as keys and counts of males, females and unknowns as values.

General Electric company has the highest proportion of women inventors with 9.48% than any other top 10 companies and Hitachi has the largest count of female inventor workforce (with over 475 women). This distribution could be due to the reason that women are increasingly likely to collaborate with other inventors rather than patent alone and more likely to participate on team inventors. Women are also more involved in chemical and pharmaceutical companies than engineering and technology based companies.

Robert bosch GMBH has the highest percentage of men i.e. 87.9% and Hitachi has the largest count of male inventor workforce (with over 5171 men). It could be due to companies like Hitachi, Basf Aktiengesellschaft and IBM working on Automotive systems, Defense systems, supercomputing, electrical products and all which we think are fields majorly dominated by male population therefore they have the high percent of men in the share of company's patents.

To conclude, there are some assignees who are more popular than others and the proportion of females is still far less than male in top 10 companies.

4. Limitations

There are around 10 million patent applications to analyze from, causing a large chunk of time to be spent running the code. In order to figure out the relationships between genders and applications, genders of applicants and attorneys were required. However, our current analysis of gender could only be based on the gender API and a gender dictionary which was not accurate and complete. This led to a large number of missing values and unknown values which might cause great deviation in our analysis results. Moreover, with the data and the fact that there are quite a number of inventor names that we can't fully detect, we have to assume that the data that we do manage to analyze is within the margin of error.

A limitation of looking at utility patents is that we wished we had more detailed information regarding the number of women in the utility industry. Even if there were the highest number of female inventors in biotechnology patents, we are not sure if it truly represents the number of female inventors in the biotechnology industry as they may not be represented as much as other industries. Although certain information regarding the industry was found in studies and articles, there was no compiled data that had an equal matching to the patent application data, giving us a better understanding of how much women are being represented in patent applications.

Big events around the world and in the United States can change the trend of data for a short time or a long time. This can lead to fluctuations in trends and a large number of outliers which could lead to biases and misunderstandings in the process of achieving our analytical goals. To figure out the reason for the results, we still need to find more information. For instance, if we want to know why there is a steeper decreasing trend in gender gap in "Foreign", we should get more details about the data like GDP or the state's investment in scientific research funds. The difficulty is that there are more than 200 countries assigned the label "Foreign", so getting more information will be time-consuming.

In the 25 million applications numbers in assignee and inventors data, there were many NAN values. There are some application numbers with missing gender but having company names which makes it difficult to identify the gender of the inventor of the patent which could lead to some incomplete counts of patents for that particular company. There are missing assignee data as well but the application no. exists and we know the gender but not sure of what company did they assign their assignments, therefore they are not considered in the analysis. If we had the data for all the companies with all the genders of inventors we might have a better result.

5. Acknowledgment

We would like to thank Professor Jordana Goodman for giving us the right guidance every time we asked a question or provided a deliverable. We would also like to thank Professor Lance Galletti for giving us the opportunity to work on this exciting project, and we would finally like to thank Jacqueline Li, Michelle Voong, and BU Spark! for providing us a project to be a part of.