Tidewater Big Data Enthusiasts: Exploring Baby Names Using Social Security Administration (SSA) Number Applications

Chuck Cartledge, PhD

October 31, 2025

Contents

List of Tables

List of Figures

1 Introduction

We examine Social Security Number (SSN) applications as provided by the Social Security Administration (SSA) to see how the self declared first name and gender of the applicant changes over time. The source data includes the first name of the applicant, their gender, and the number of times that name and gender combination occurs per year. We suspect that names tend to be gender specific, but may not be absolutely specific.

We look at the absolute number of applications per year, changes in absolute numbers as a percentage from the previous year, possible economic influences (recessions and depressions) on SSN applications, and how selected names wax and wane over time.

2 Discussion

The Social Security Administration (SSA) does not require a Social Security Number (SSN) for children.

> *"Getting an SSN for your newborn is voluntary, but may be necessary to obtain important services, such as those listed above, for your child. Therefore, getting a number when your child is born is a good idea."*
>
> SSA [5]

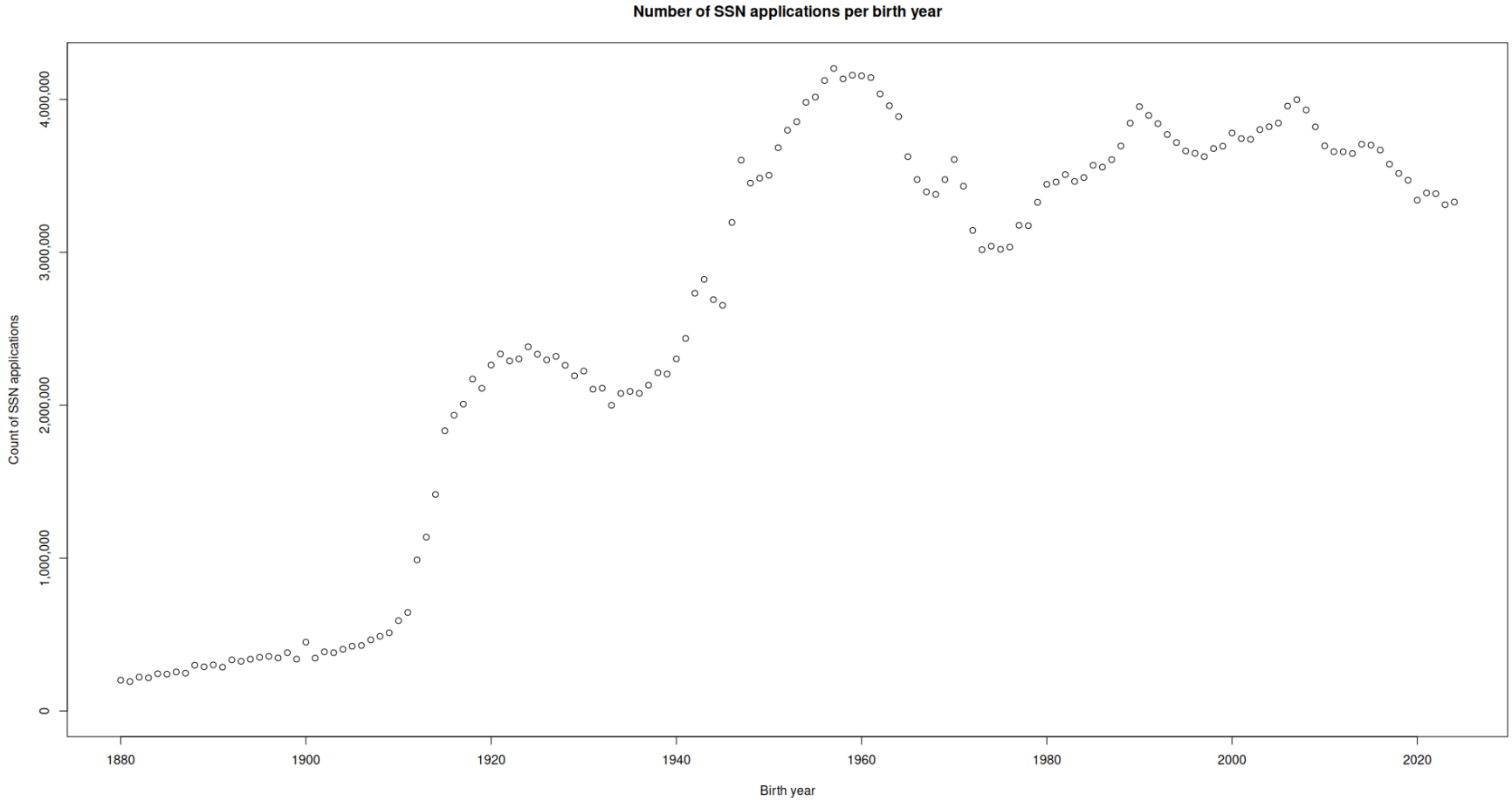
In practical terms, a child needs to have an SSN as soon as practical, as it is difficult to accrue SSA benefits without one. This is currently not too much of an issue now, as many parents apply for an SSN for their newly born child as soon as possible. Things were more difficult when the SSA was established in 1935[3]. Reaching back from the SSA's start date to the year when SSA beneficiaries could have been born, gets us to 1880. The SSA makes available for each year from 1880 to 2024, the first name of each SSN applicant, the gender from the application, and how many times that name and gender combination occurred for that birth year[4]. Thus it is possible to track the number of applications per year, the number of times a name and gender combination per year, and the change in number of applications per year.

We will be exploring how these data change over the years.

2.1 Gross trends

The gross number of SSN applications can be expected to increase as the population increases (see Figure 1), but doesn't.

Figure 1: Number of SSN applications per year. The number of applications seems to follow the population over time, but not always. There are dips in the mid 1920s, the mid 1950s through 1970s, and then smaller/shorter dips in the 1990s, and finally a general downward trend starting in the 2000s.



Taking a different view on the data, and adding the periods of economic depressions and recessions (see Figure 2), there are hints of possible economic and social conditions on SSN applications. The Federal Deposit Insurance Corporation (FDIC) was founded in 1933, in part as a result of the Great Depression. The Securities and Exchange Commission (SEC) founded in 1934, also in part as a reaction to the Great Depression.

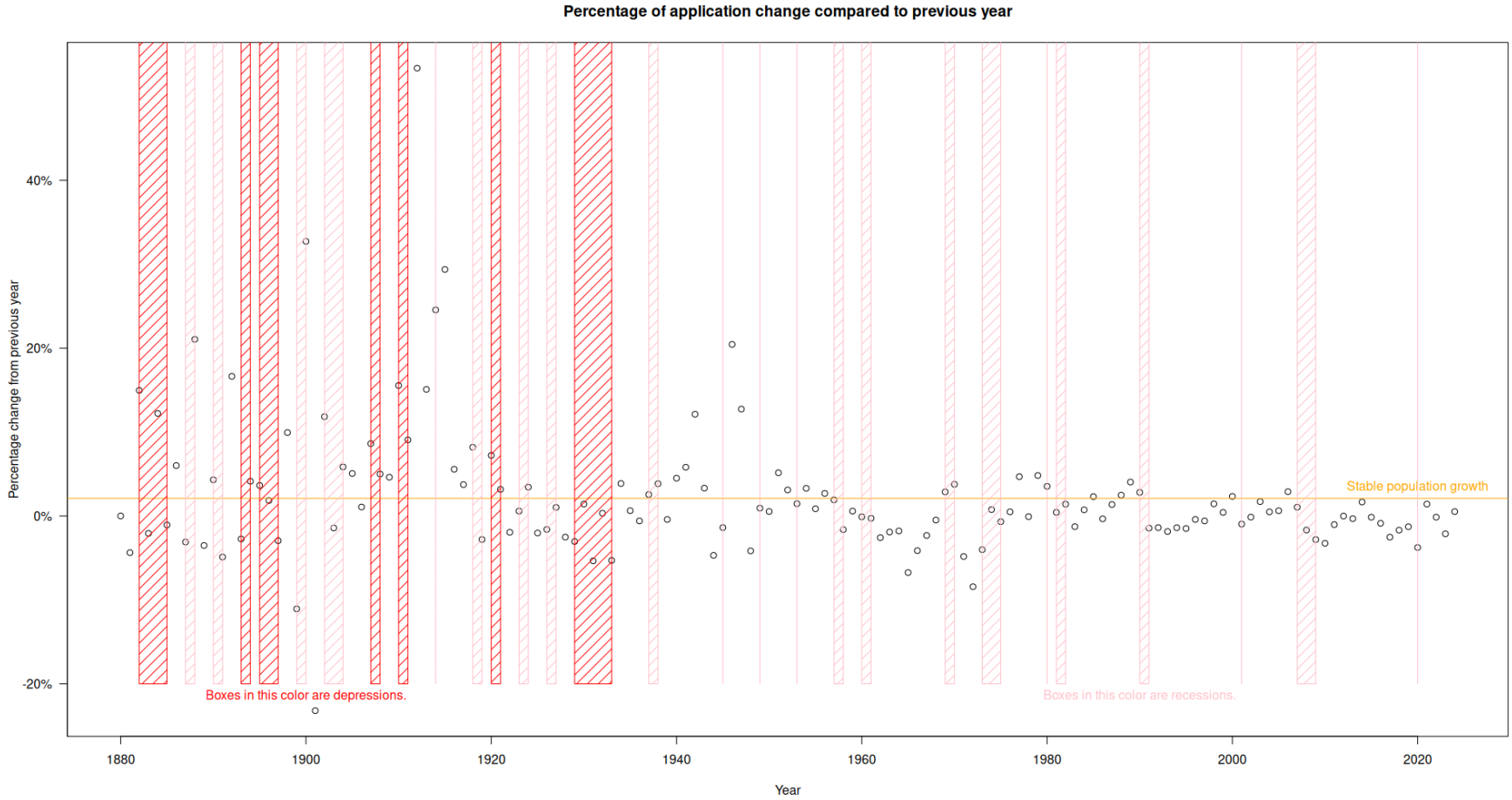
> *"The difference between a recession and a depression is that a depression is much more severe and longer lasting. Depressions are also far more rare than recessions. ... Common Definition of Recession: Two or more consecutive quarters of decline in gross domestic product (GDP) growth. ... An economic depression is similar to a recession, but much more severe and longer lasting. Not only does a depression last longer, but its effects can be far-reaching and linger long after the economy begins to recover"*
>
> Tim Maxwell [2]

A visual inspection (see Figure 2) hints at a relationship between SSN applications and the economic future. The applications could be a lagging indicator because of the time it takes from when a baby is conceived and when the economy turns down. Again, there are hints but nothing that appears really causal.

Figure 2: Number of SSN applications per year and economic downturns. Prior to the founding of the FDIC and SEC, the country went through many depressions and recessions. The combined efforts of the FDIC and SEC may have reduced the likelihood and severity of economic downturns, and that may have an effect on the number of SSN applications, where the applications could be seen as a proxy for a the long term economic future.

Yet another way to look at the yearly applications, is the percentage change in applications from one year to next (see Figure 3). The absolute numbers will change due to changes in the total population and an assumed birth rate. Percentage change may indicate a change in how many children are needed or desired. A positive change (more applications than previously) may indicate a more positive outlook for the future. While a negative change may indicate a less positive outlook. A 2.1% fertility is generally considered the minimum needed to maintain a stable population[1] assuming mortality rates remain constant and migration has no effect. Extended periods below this replacement level may lead to a long term decline in the total population.

Figure 3: Percentage change in number of SSN applications per year and economic downturns. There are hints that an economic downturn may affect the number of applications. The percentage change in applications sort of follows a downturn, but the change is not always absolute. A 2.1% fertility is generally considered the minimum needed to maintain a stable population.



2.2 Name trends

The SSN application data includes the applicants first name, gender, and number of times that name and gender combination was used. We will now start drilling down into the details of the data. Data was not reported for names that appeared less than 5 times. So individuals with names with very small *n*s will not be identified. Also, the names are formatted so that only the first letter is capitalized, and only upper and lower case letters are used.

All names across all years were tabulated. Some names were always reported as female, some always as male, and some reported as both male and female (see Figure 4). For instance, in 1920 there were about 4,645 uniquely female names, 3,867 uniquely male names, and 1,122 names that were reported female and male.

Because the total population changes over time, looking at absolute numbers will give one view, while looking percentages gives another (see Figure 5). The percentage of names that are in one of the three possibilities (only female, only male, or both female and male) appears relatively constant. The percentage of names that are both self reported as female and male appears to be relatively constant at 10%.

Figure 4: Number of unique male, female, and both male and female names per year. All names across all years were tabulated. Some names were always reported as female, some always as male, and some reported as both male and female.

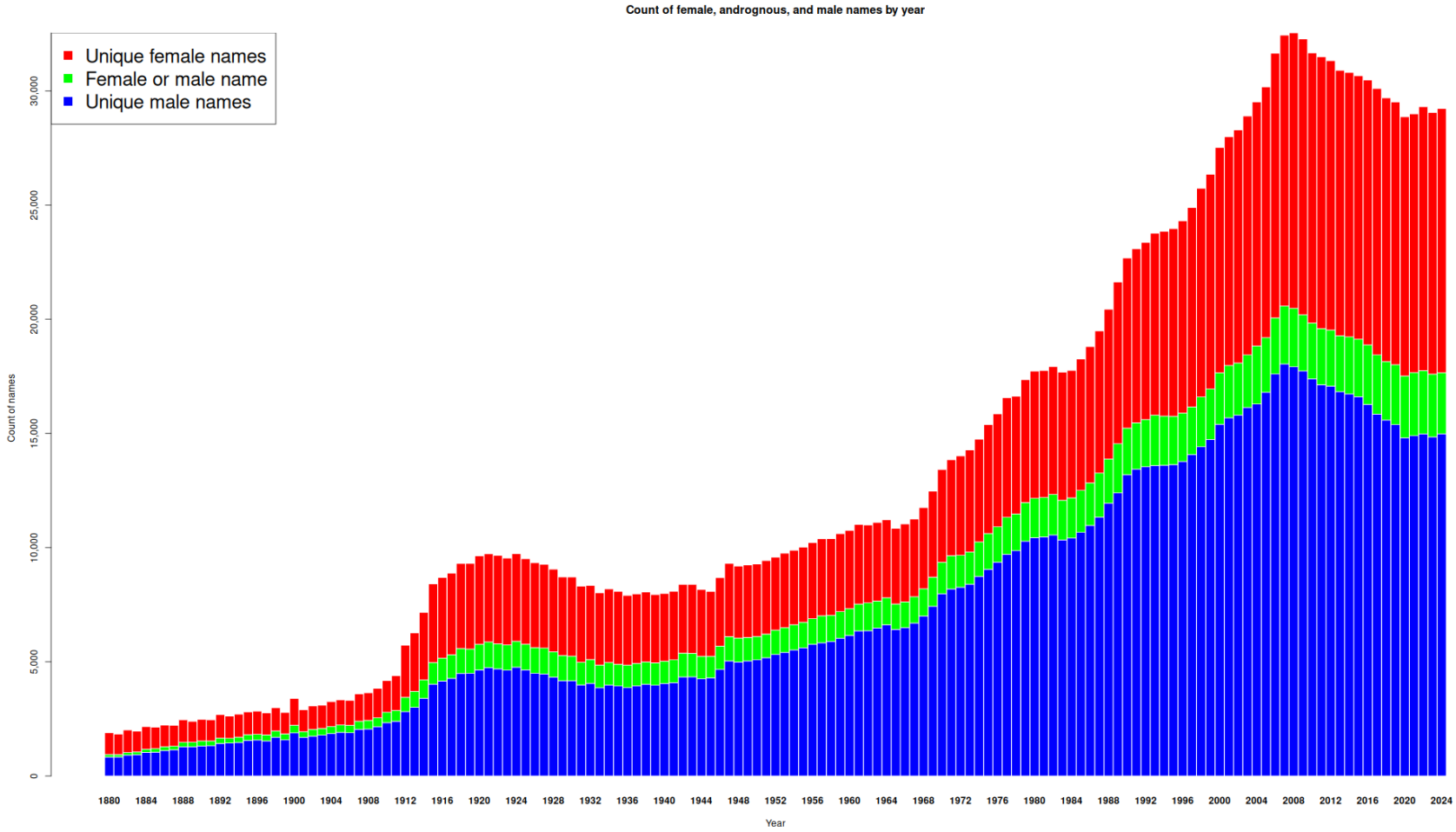
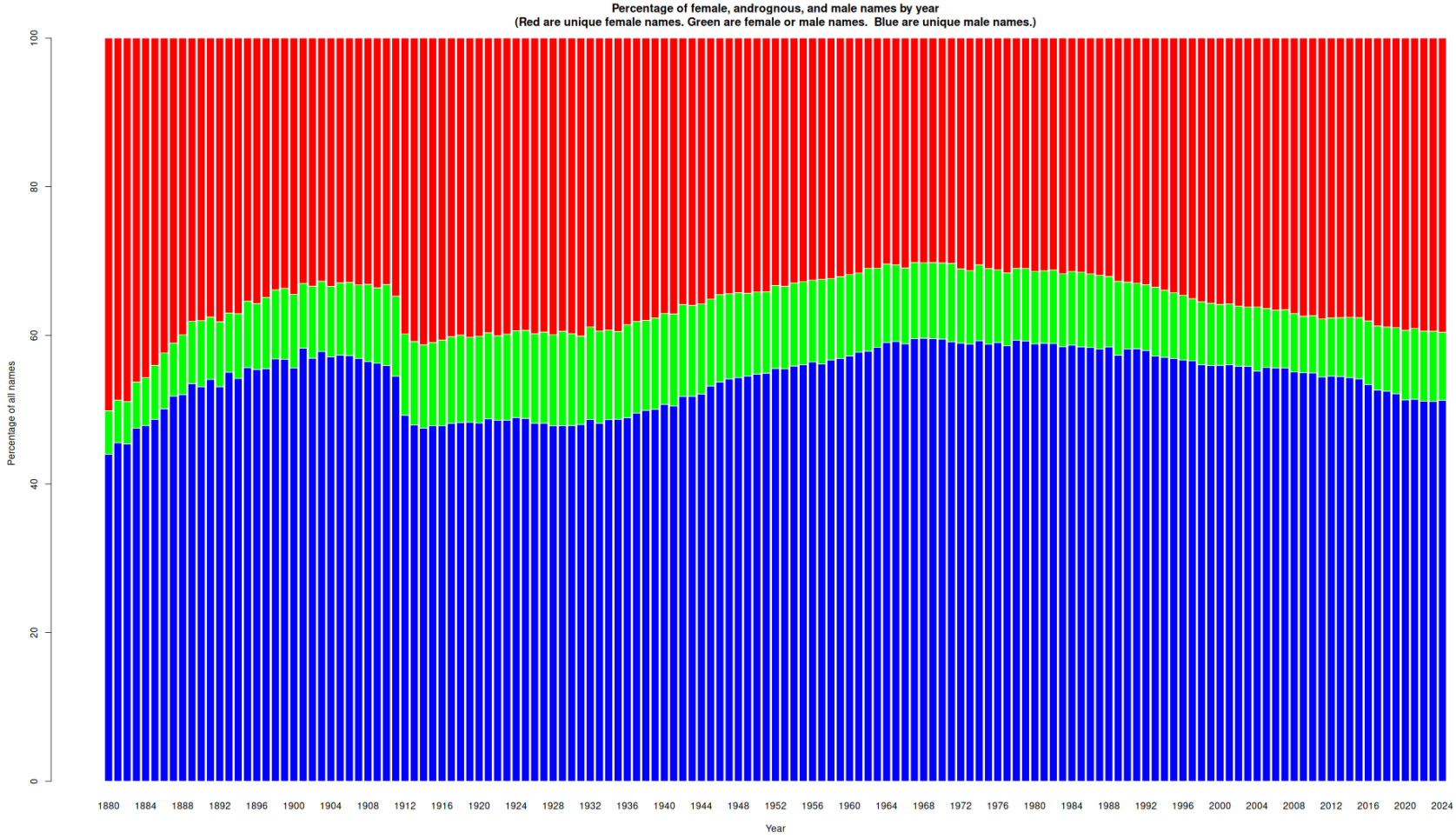


Figure 5: Percentage of number of male, female, and both male and female names per year. Uniquely female names are shown in red. Uniquely male names are in blue. Names that are reported as both female and male are in green. The percentage of names that are in one of the three possibilities (only female, only male, or both female and male) appears relatively constant.



2.3 Individual name trends

A collection of names were selected based on idle curiosity to see how their reported gender in absolute and percentage terms changed over time. The names and links to their images (see Table 1).

Table 1: First names, their absolute and percentage plots.

Name	**Absolute Plot**	**Percentage Plot**
Candy	Figure 6	Figure 7
Charles	Figure 8	Figure 9
Colleen	Figure 10	Figure 11
Cynthia	Figure 12	Figure 13
Geoffrey	Figure 14	Figure 15
Jennifer	Figure 16	Figure 17
John	Figure 18	Figure 19
Lane	Figure 20	Figure 21
Lee	Figure 22	Figure 23
Liam	Figure 24	Figure 25
Lois	Figure 26	Figure 27
Lori	Figure 28	Figure 29
Louie	Figure 30	Figure 31
Marjorie	Figure 32	Figure 33
Mary	Figure 34	Figure 35
Mitchell	Figure 36	Figure 37
Peggy	Figure 38	Figure 39
Vera	Figure 40	Figure 41
Vlasta	Figure 42	Figure 43

Figure 6: Candy, number of occurrences per year. The name appears to be on the decline and relative obscurity in the next 10 – 20 years. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

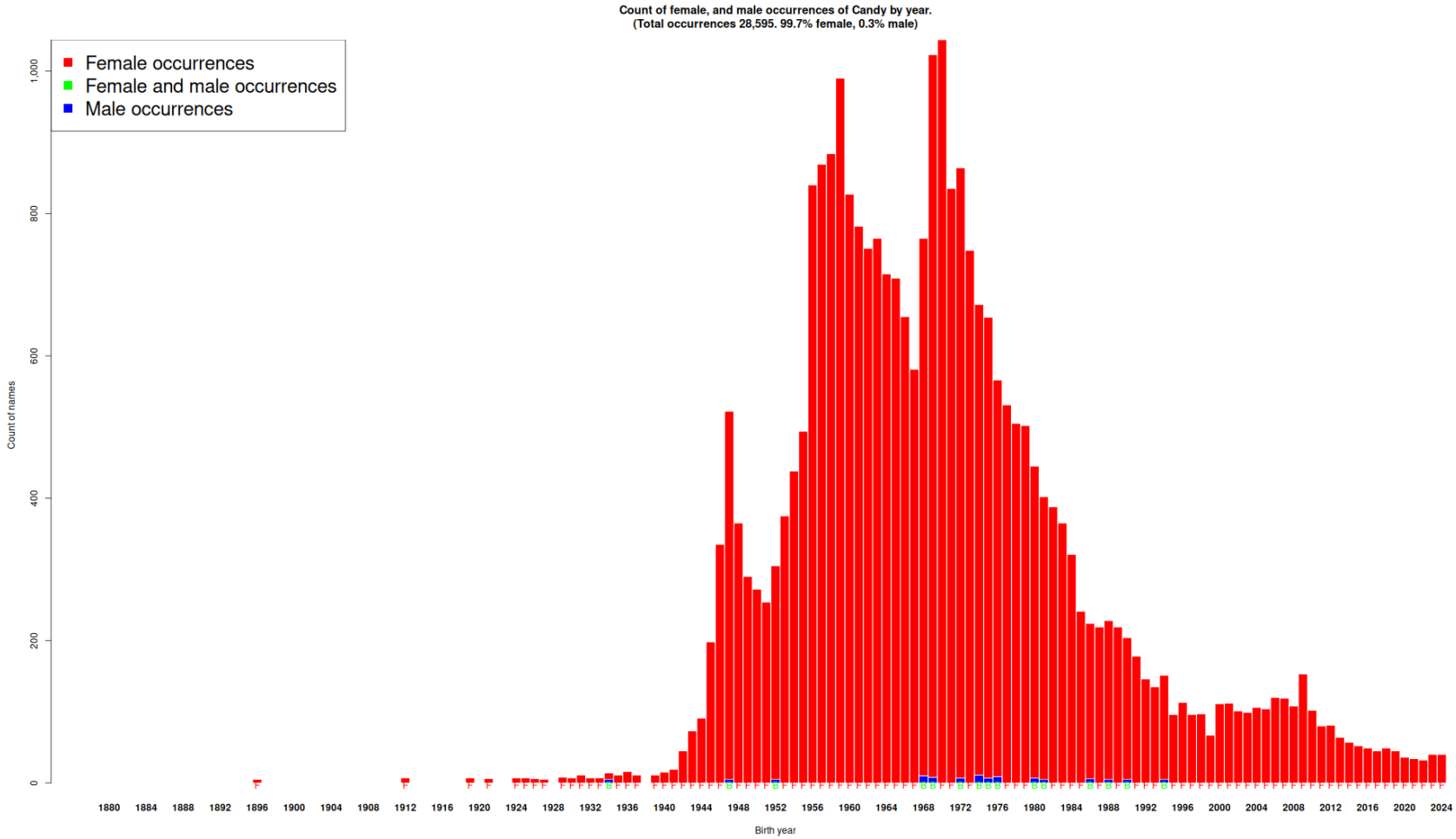


Figure 7: Candy, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

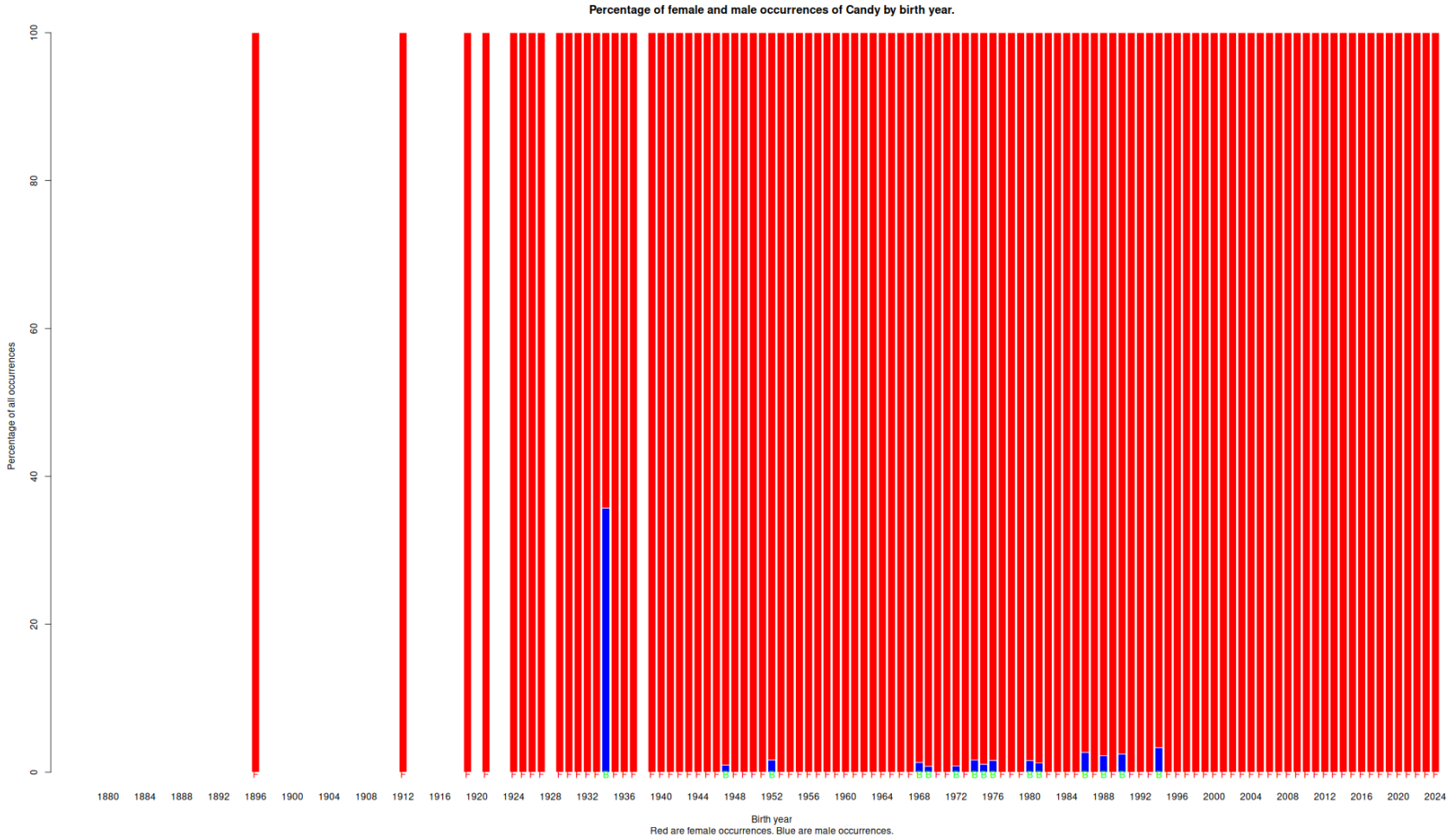


Figure 8: Charles, number of occurrences per year. The name's heyday was in the 1950s and is now slowly declining. Below each column has a red "F", or a blue "M", or a green "B" at it's base meaning all the data is in the column is Female, Male, or Both respectively.

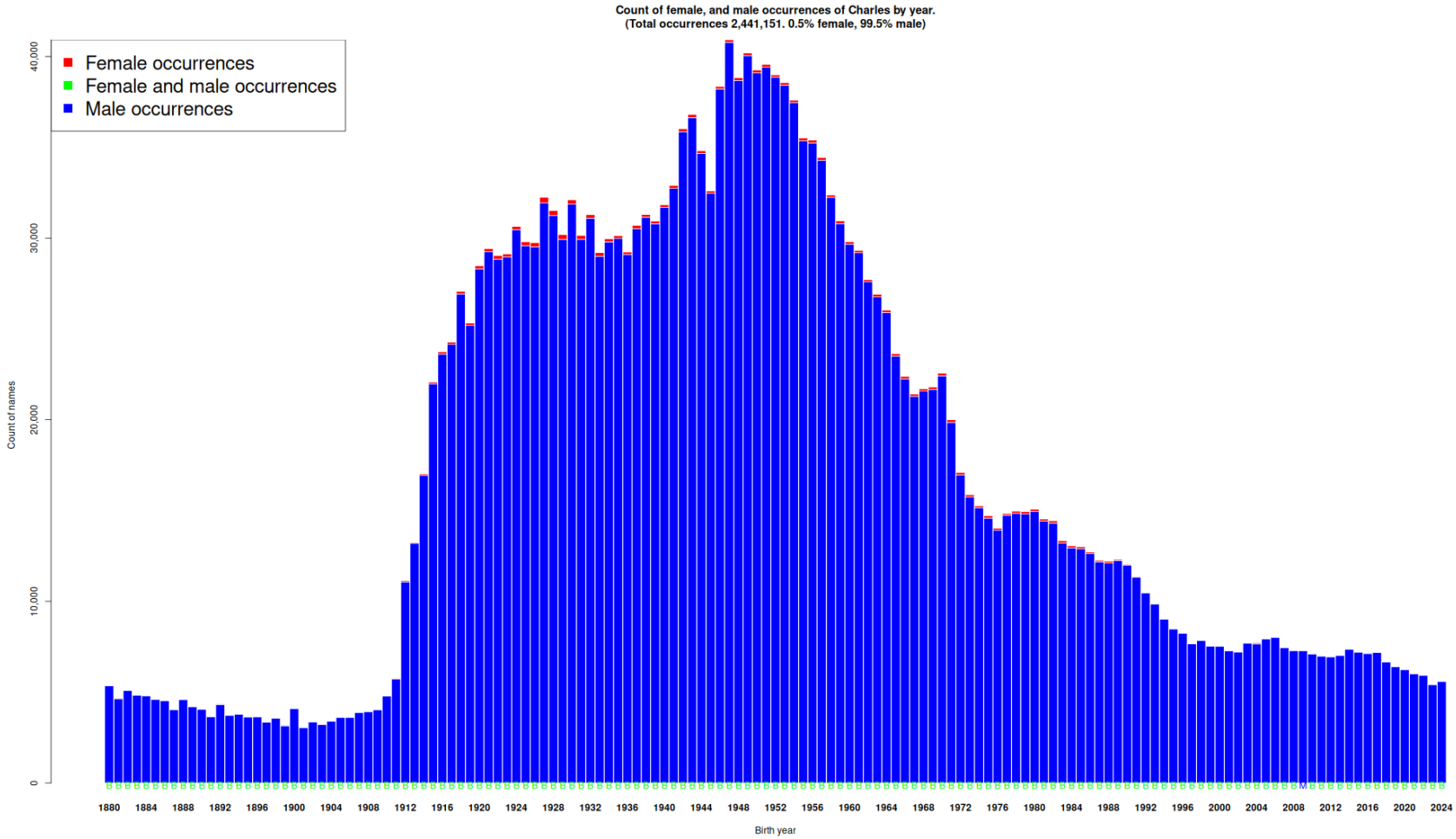


Figure 9: Charles, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

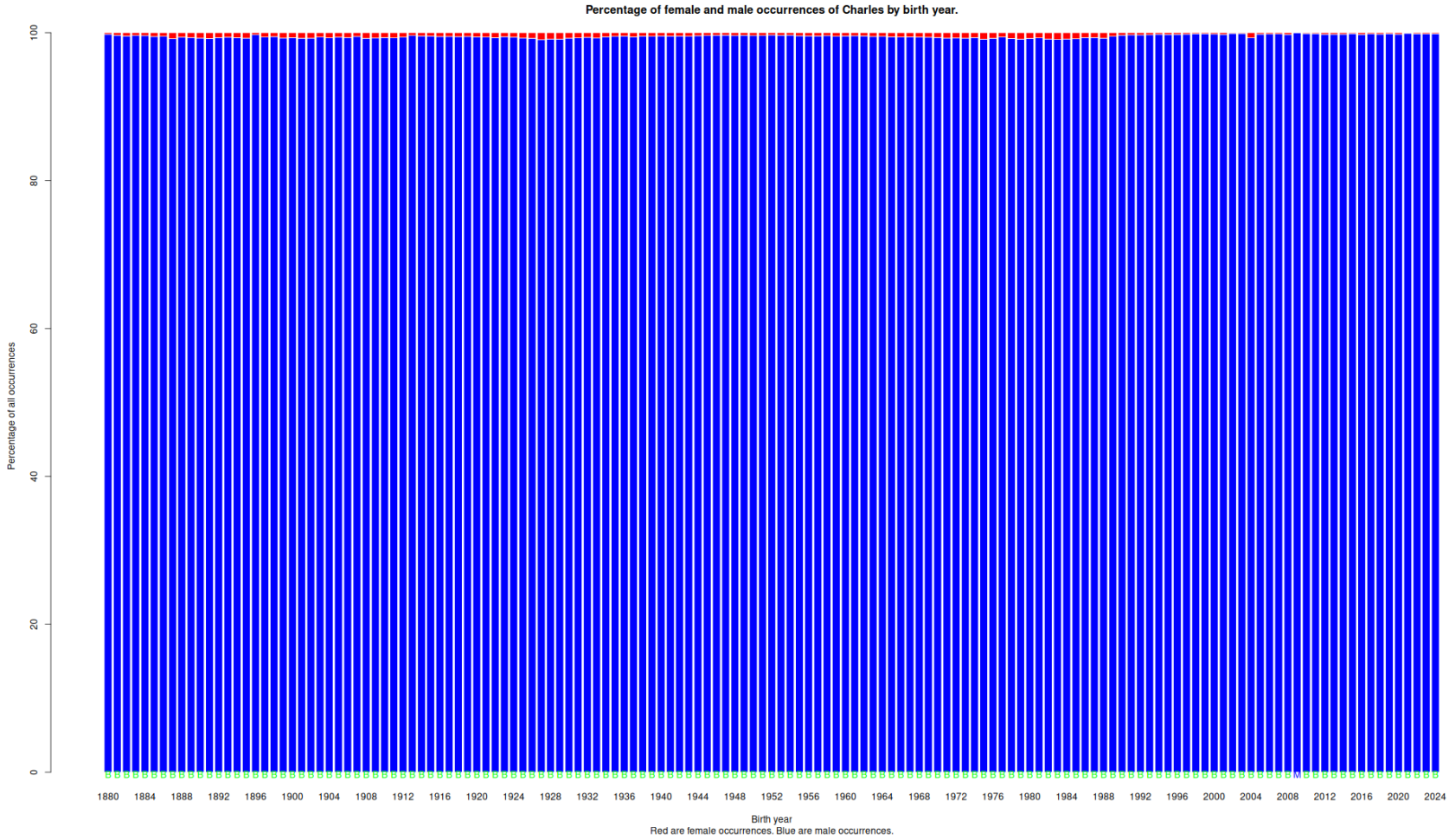


Figure 10: Colleen, number of occurrences per year. The name appears to be on the decline and relative obscurity in the next 10 – 20 years. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

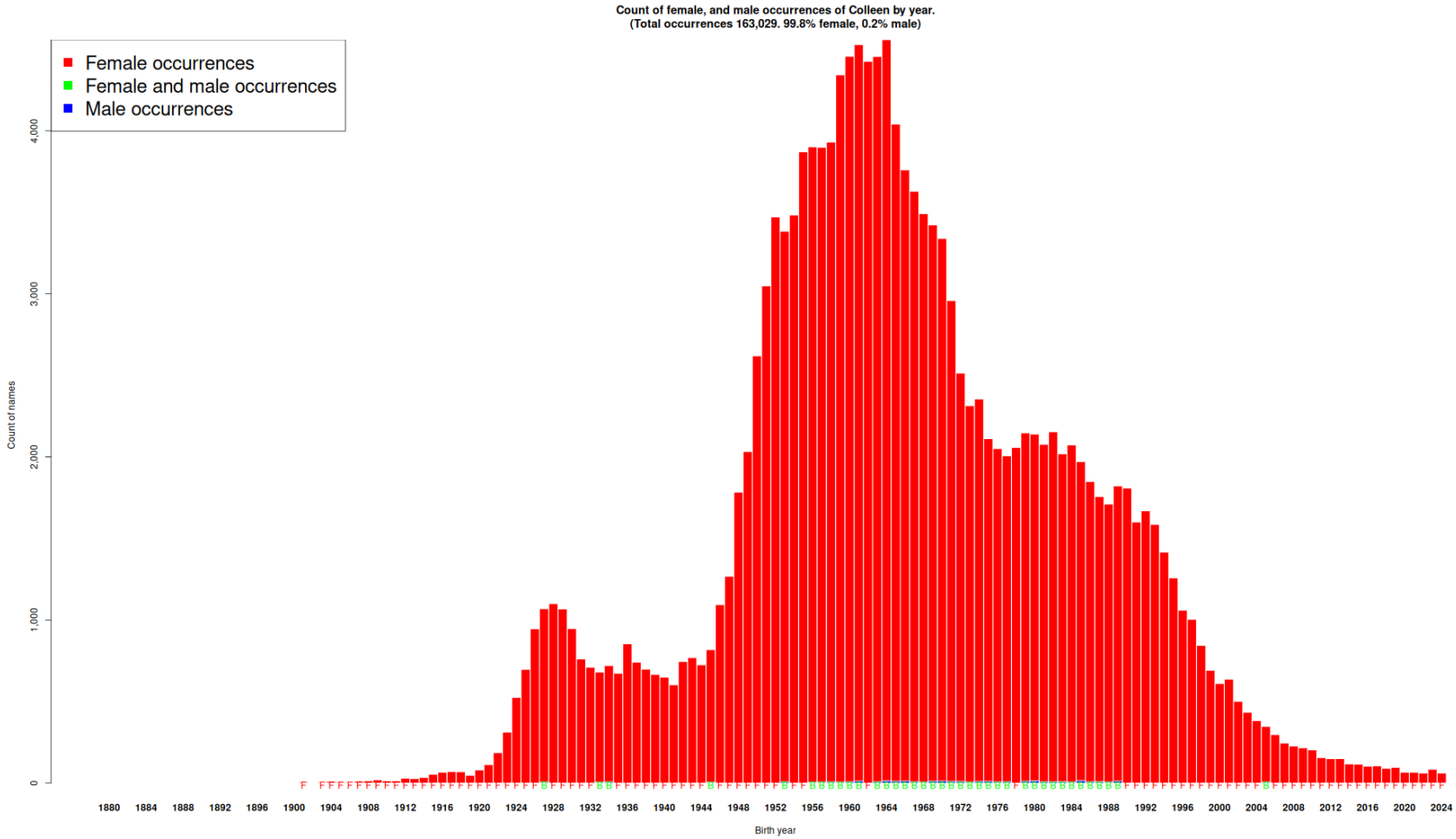


Figure 11: Colleen, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

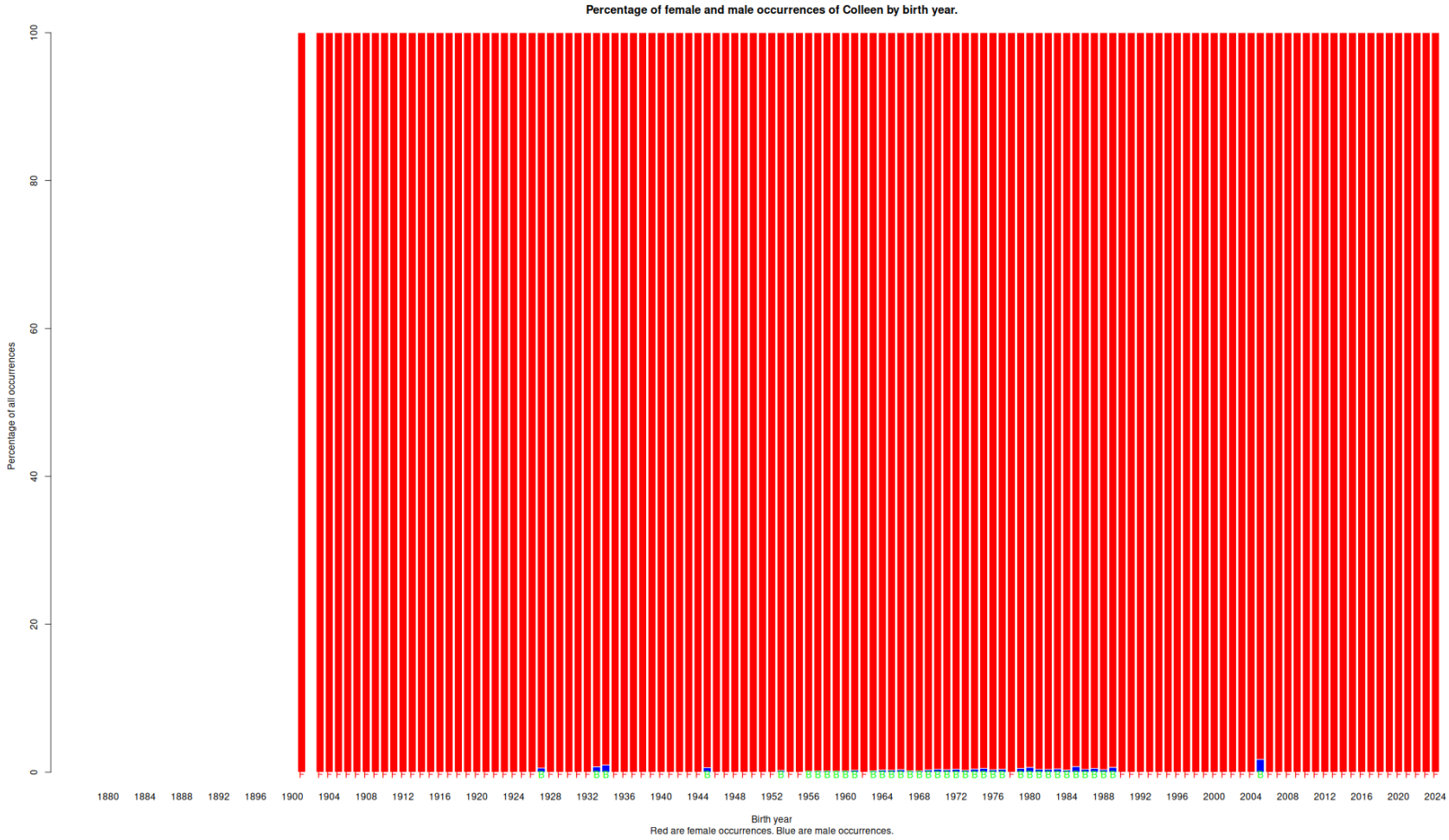


Figure 12: Cynthia, number of occurrences per year. The name appears to be on the decline and relative obscurity in the next 10 – 20 years. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

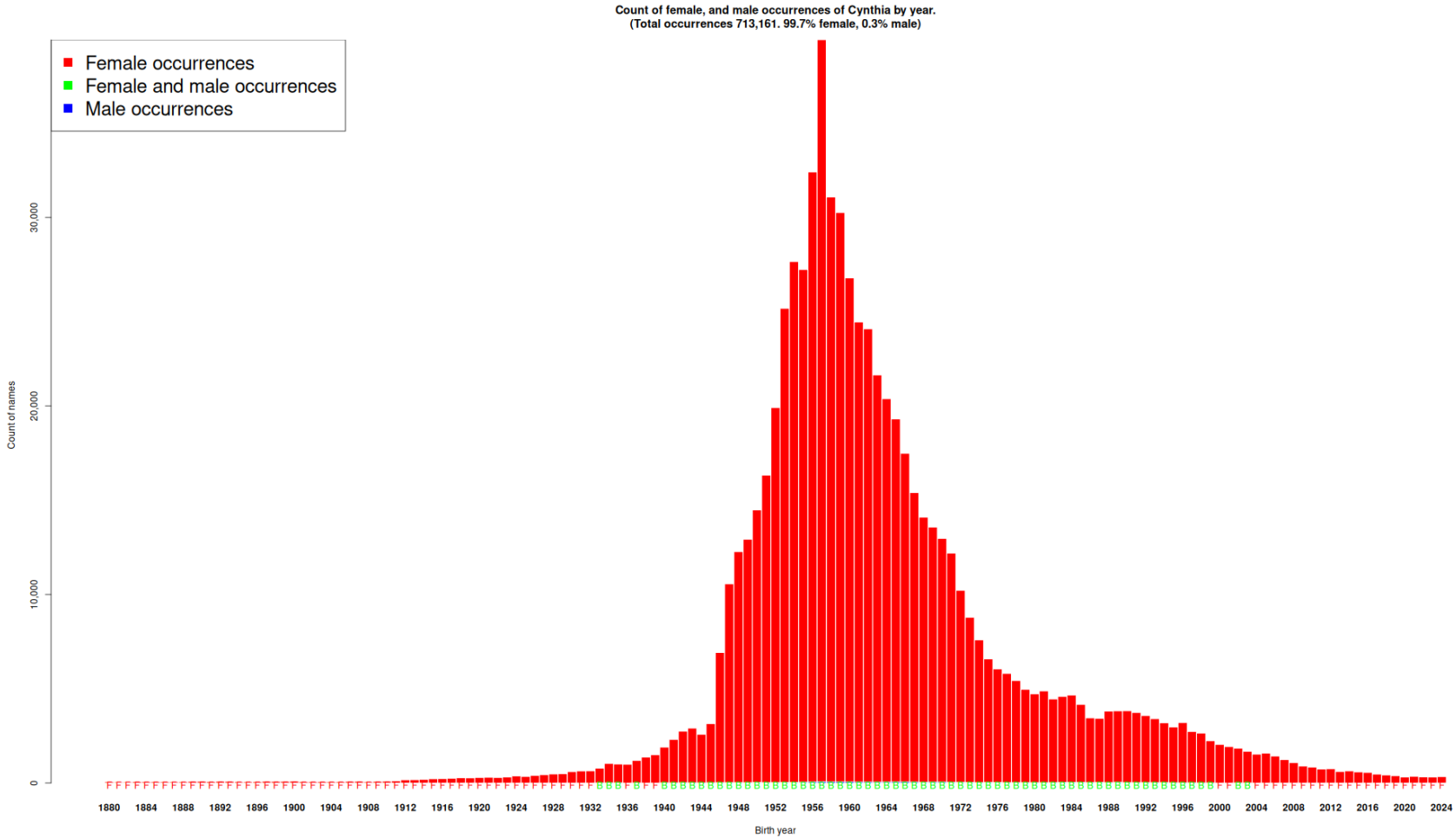


Figure 13: Cynthia, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

Figure 14: Geoffrey, number of occurrences per year. The name appears to be on the decline and relative obscurity in the next 10 – 20 years. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

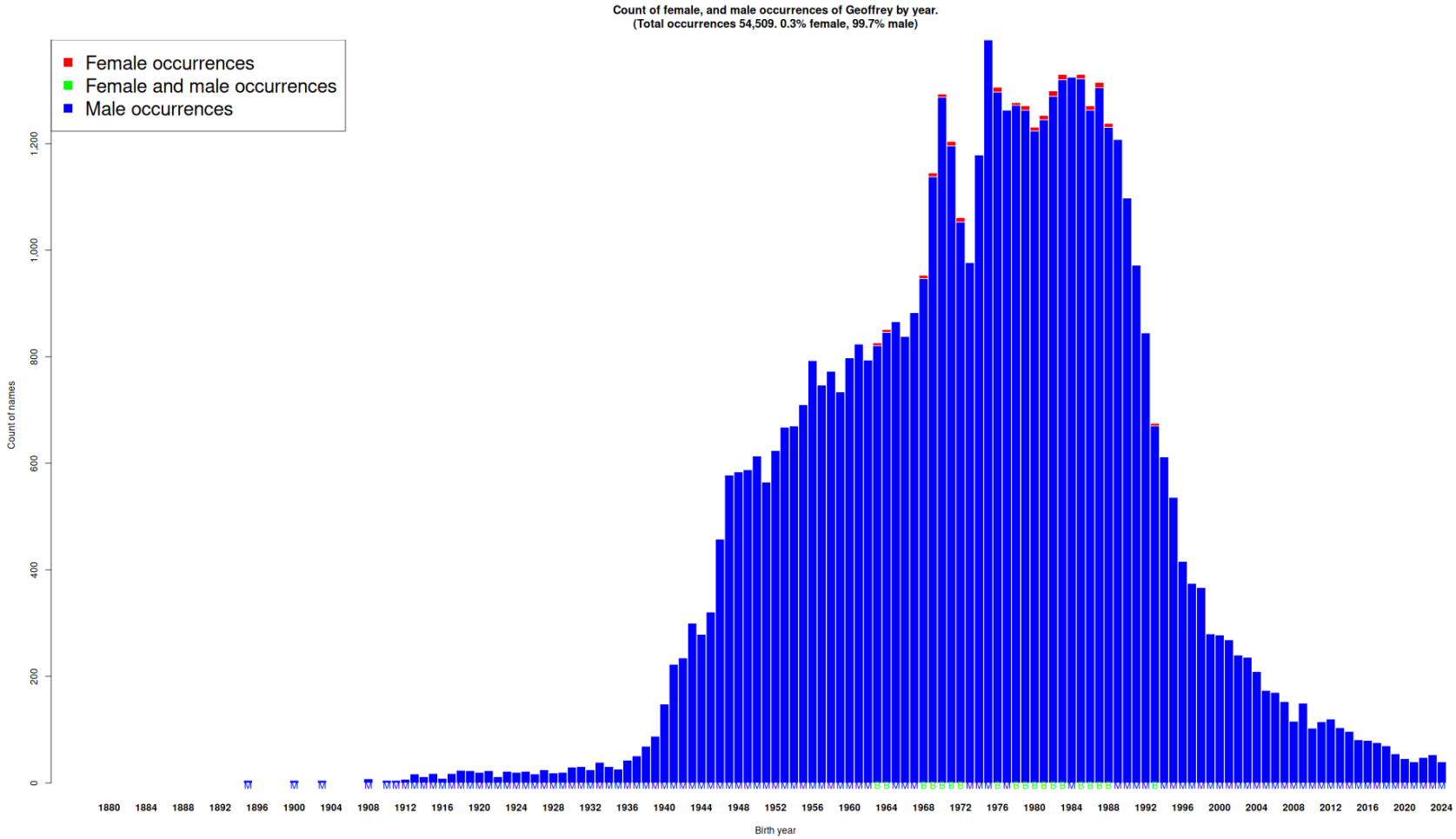


Figure 15: Geoffrey, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

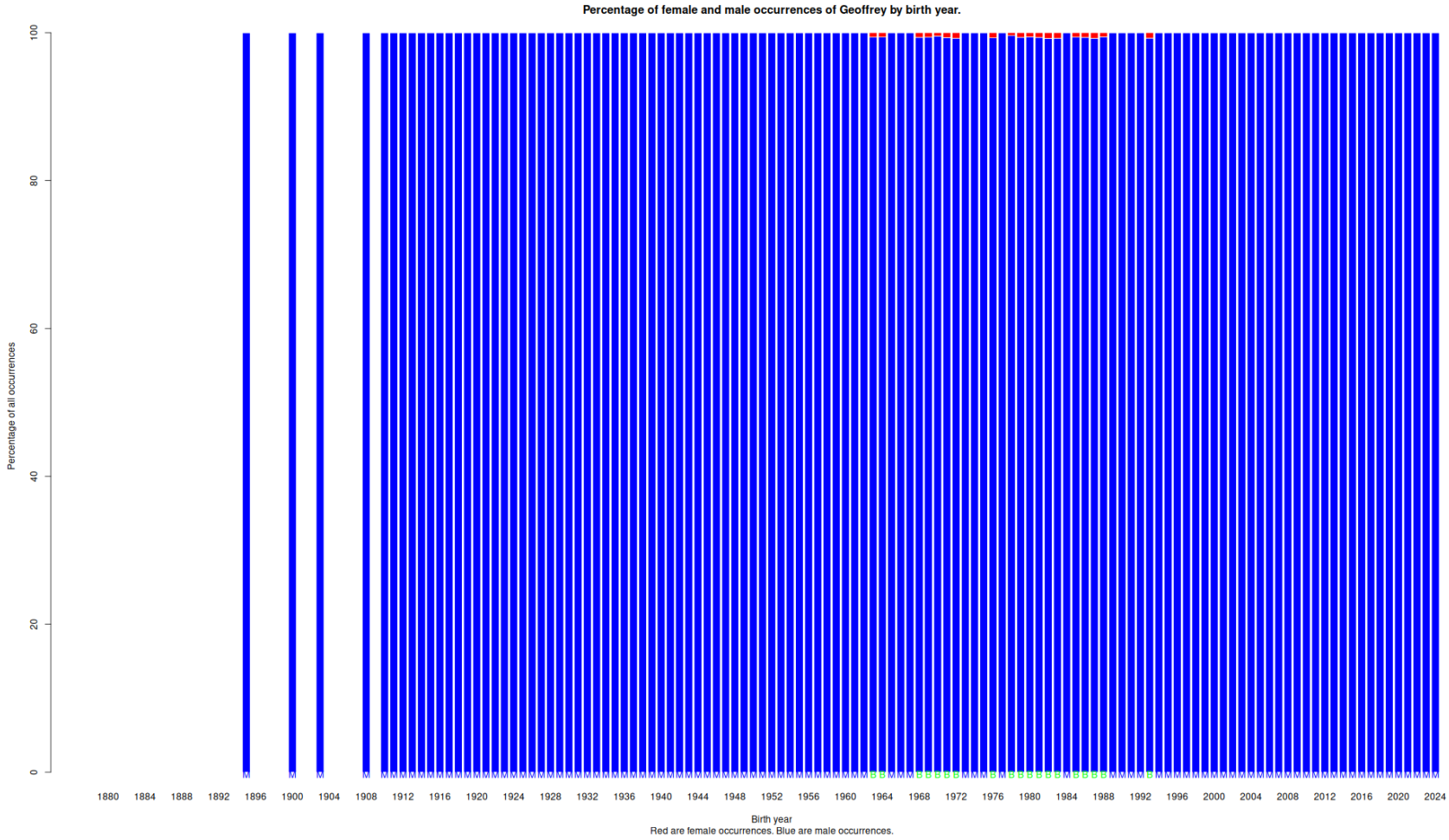


Figure 16: Jennifer, number of occurrences per year. The name appears to be on the decline and relative obscurity in the next 10 – 20 years. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

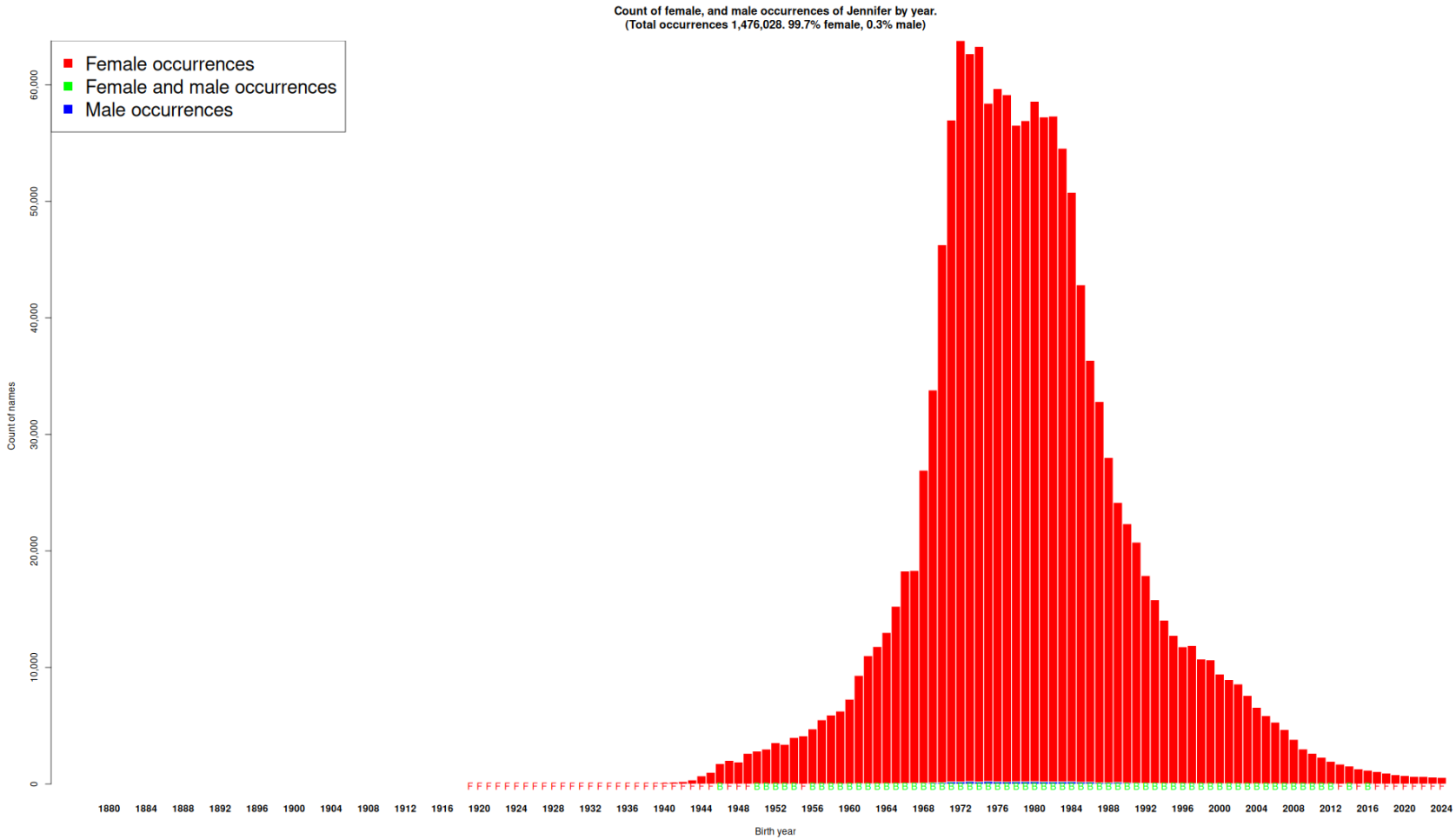


Figure 17: Jennifer, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

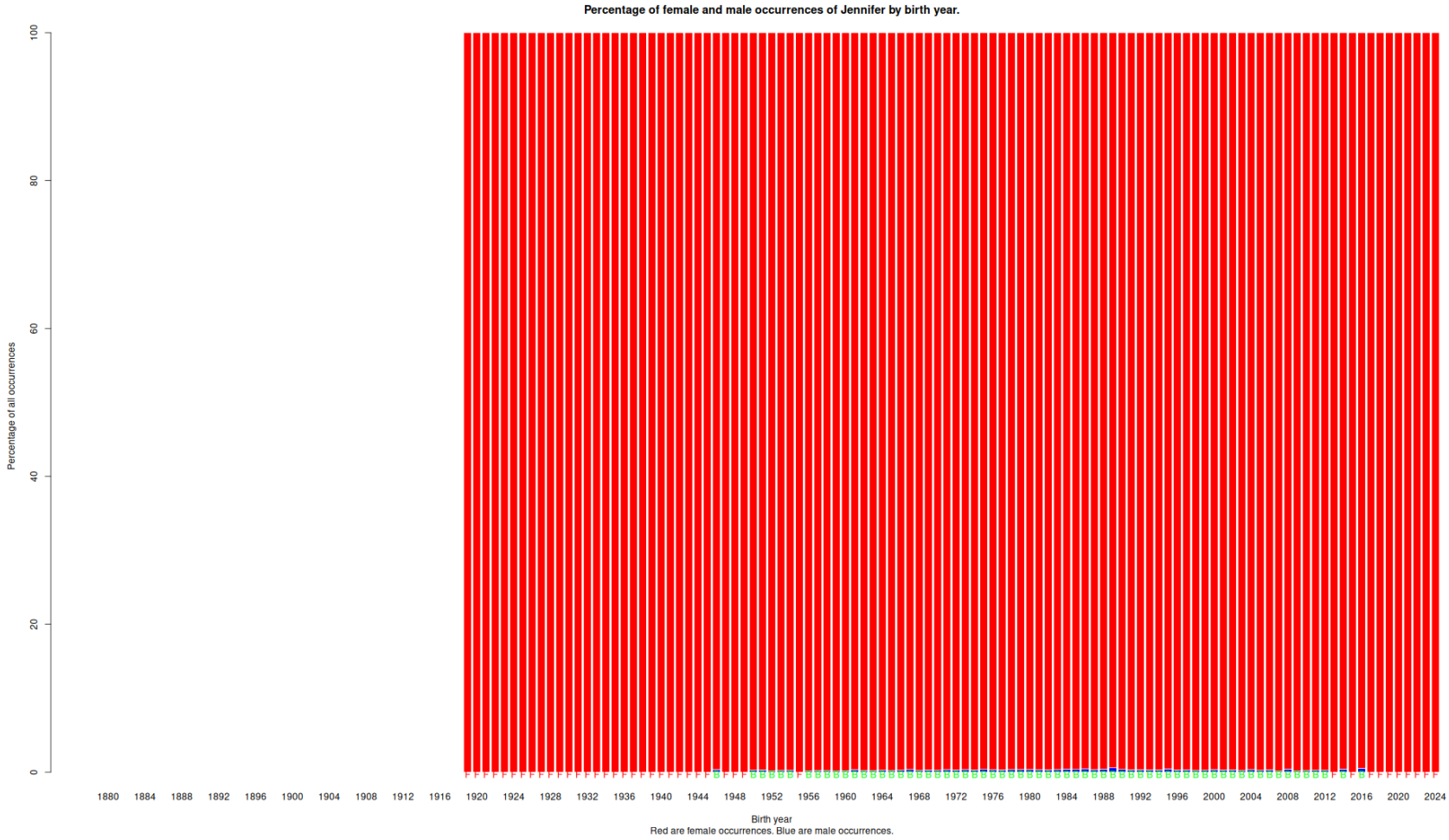


Figure 18: John, number of occurrences per year. Interesting shape of the name's occurrence over the years. It still remains strong. Below each column has a red "F", or a blue "M", or a green "B" at it's base meaning all the data is in the column is Female, Male, or Both respectively.

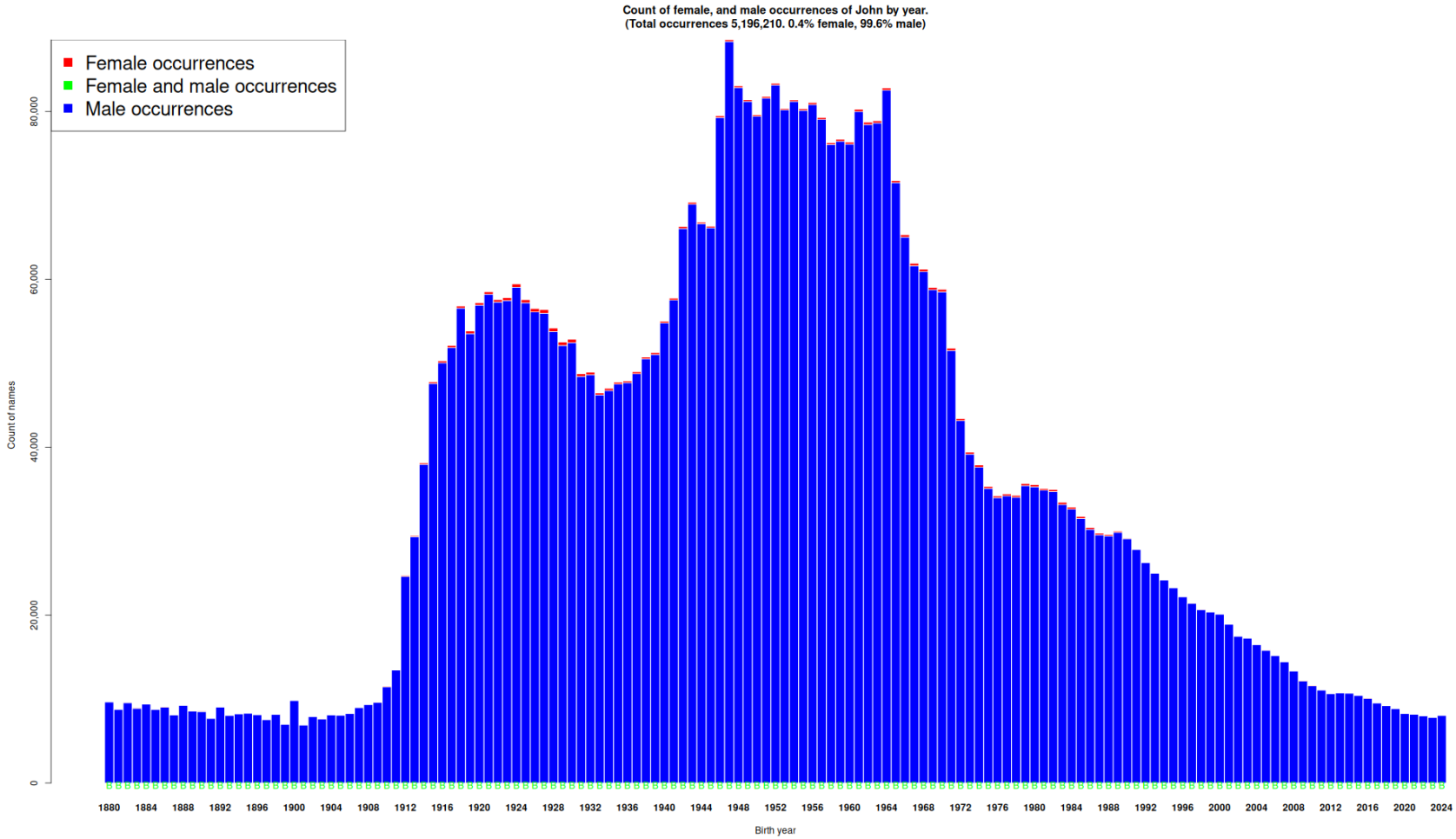


Figure 19: John, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

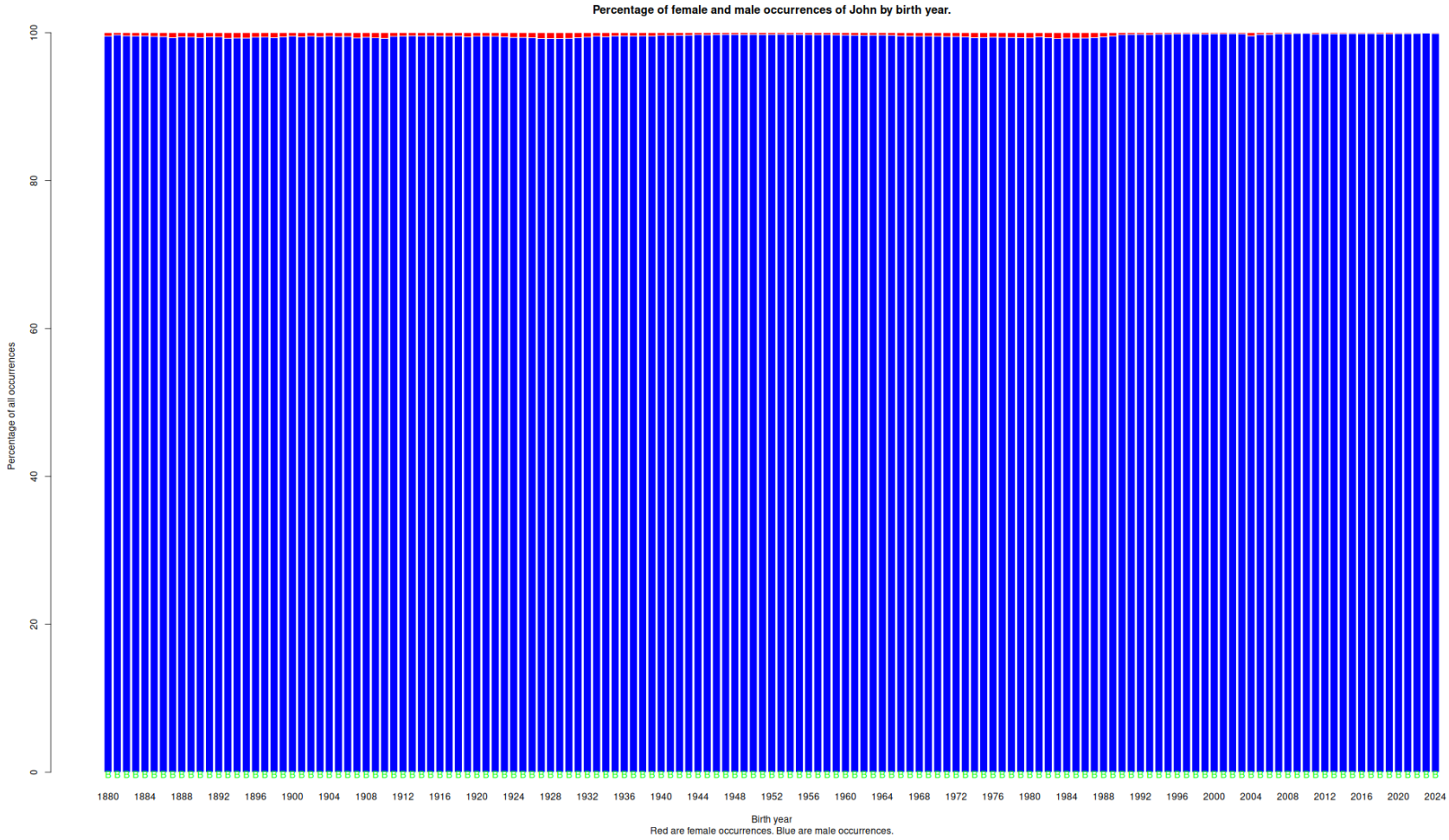


Figure 20: Lane, number of occurrences per year. The name is gaining strength. Interesting that both genders are using the name. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

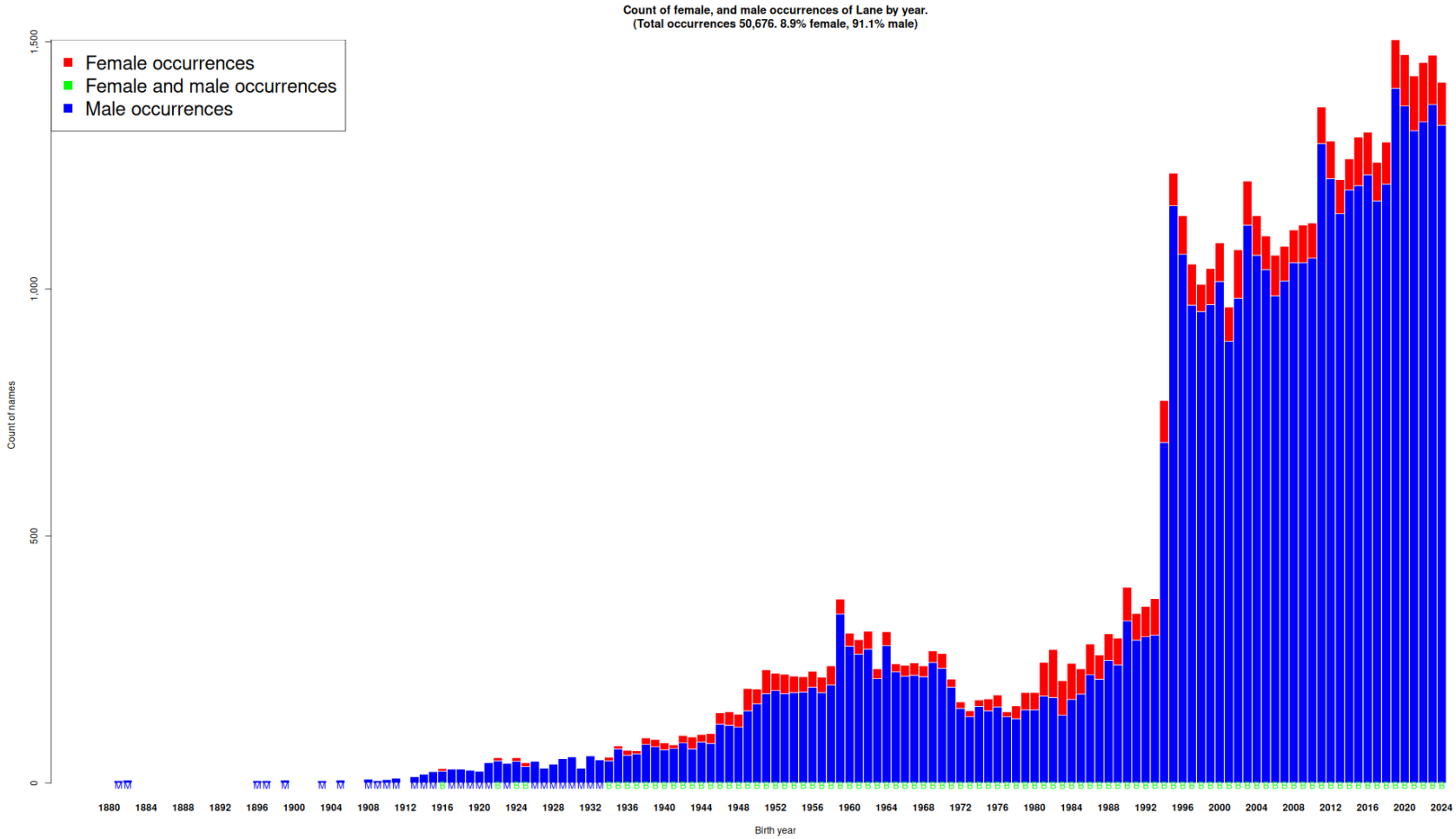


Figure 21: Lane, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

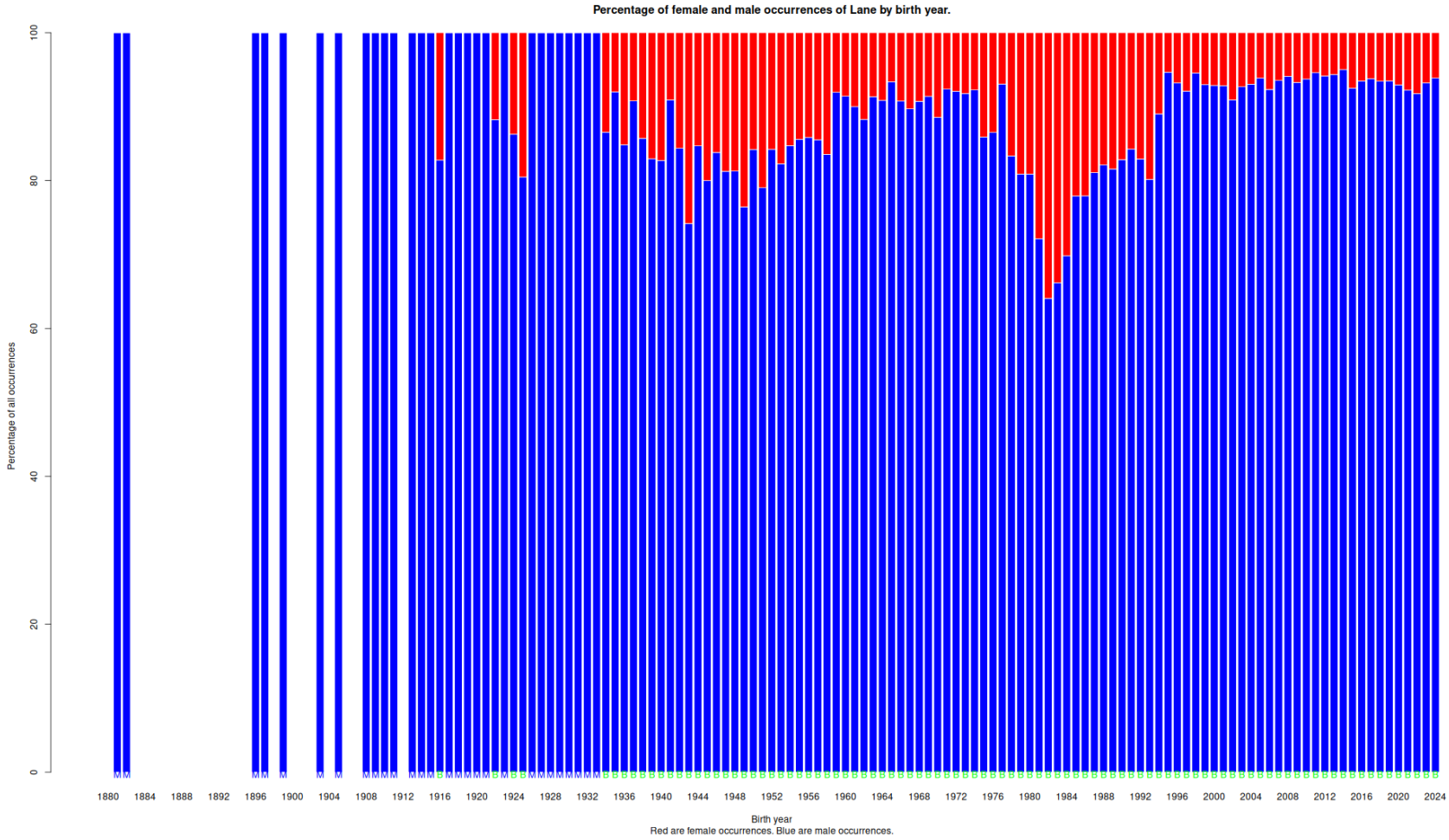


Figure 22: Lee, number of occurrences per year. The name peaked and now appears steady. Below each column has a red "F", or a blue "M", or a green "B" at it's base meaning all the data is in the column is Female, Male, or Both respectively.

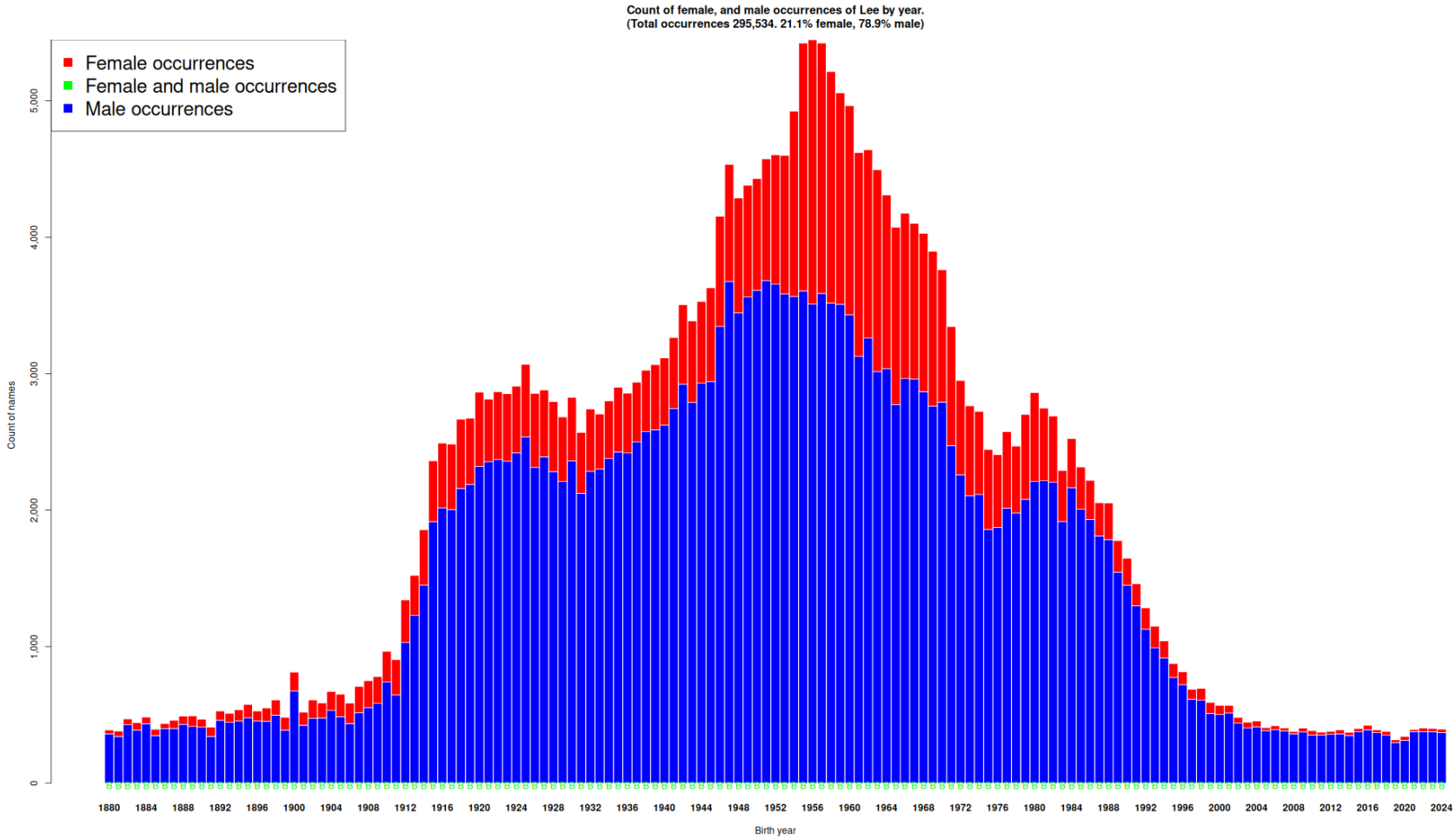


Figure 23: Lee, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

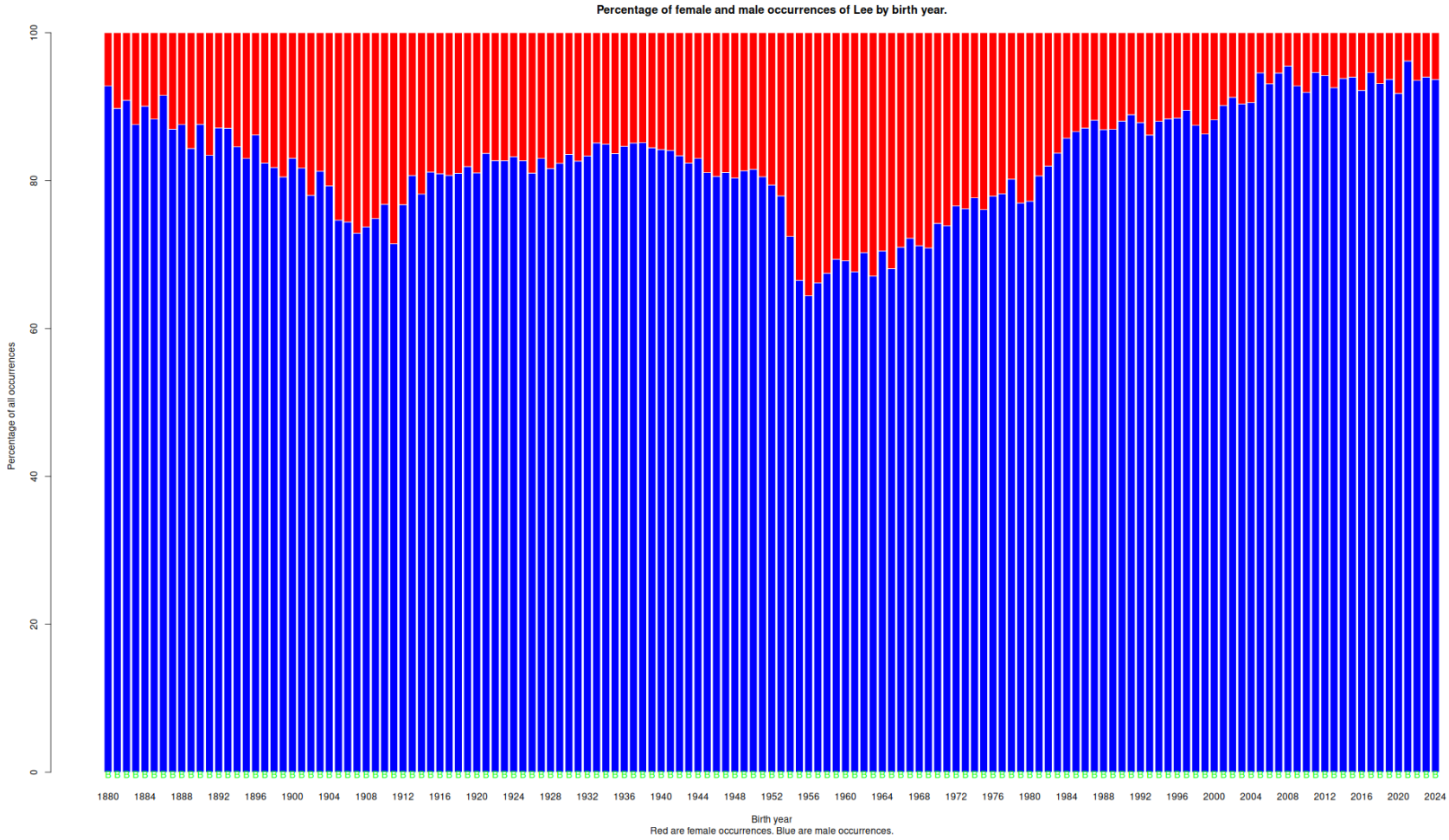


Figure 24: Liam, number of occurrences per year. So many years without any entries, and now it is going strong with males, and a few females. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

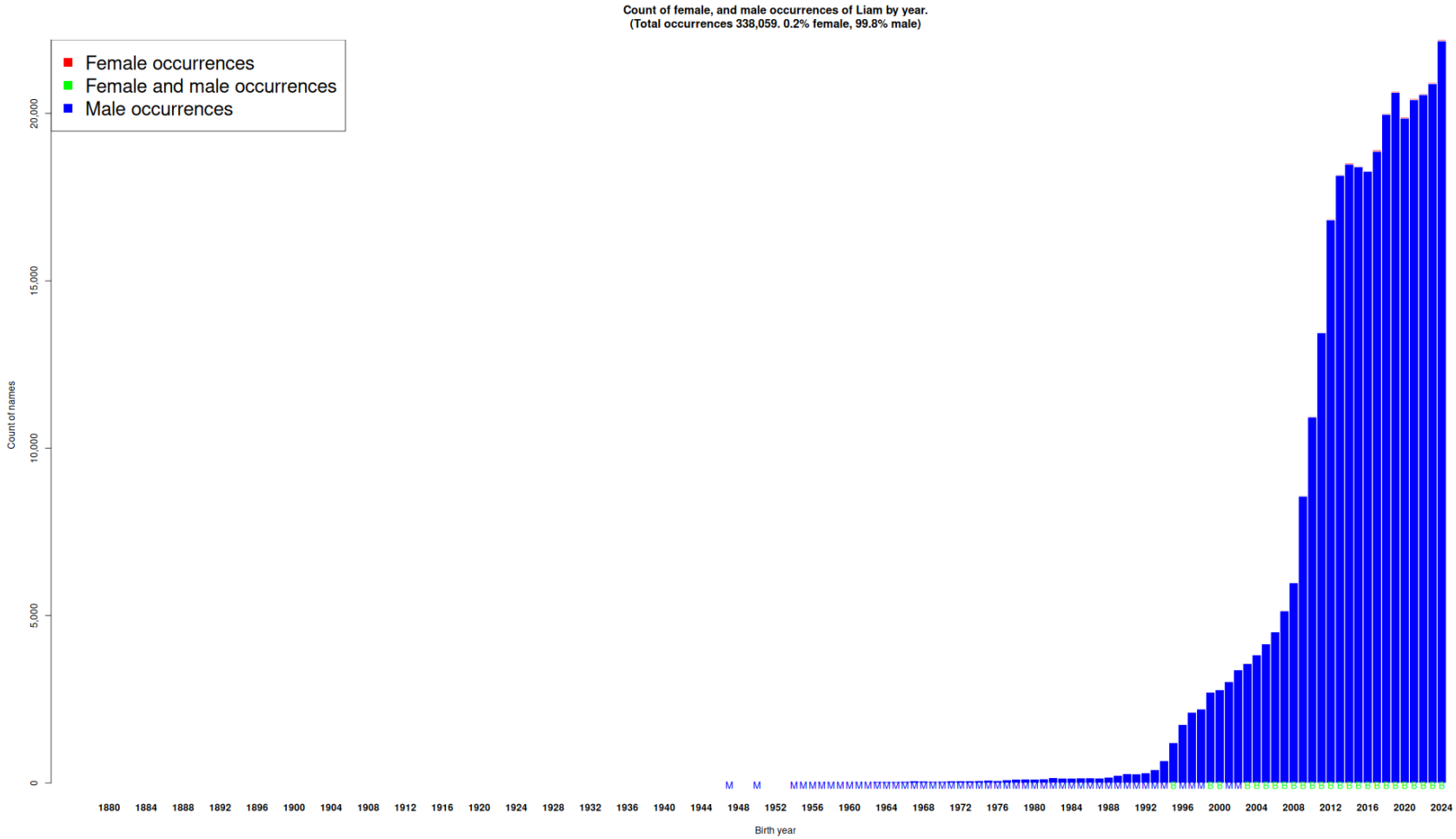


Figure 25: Liam, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

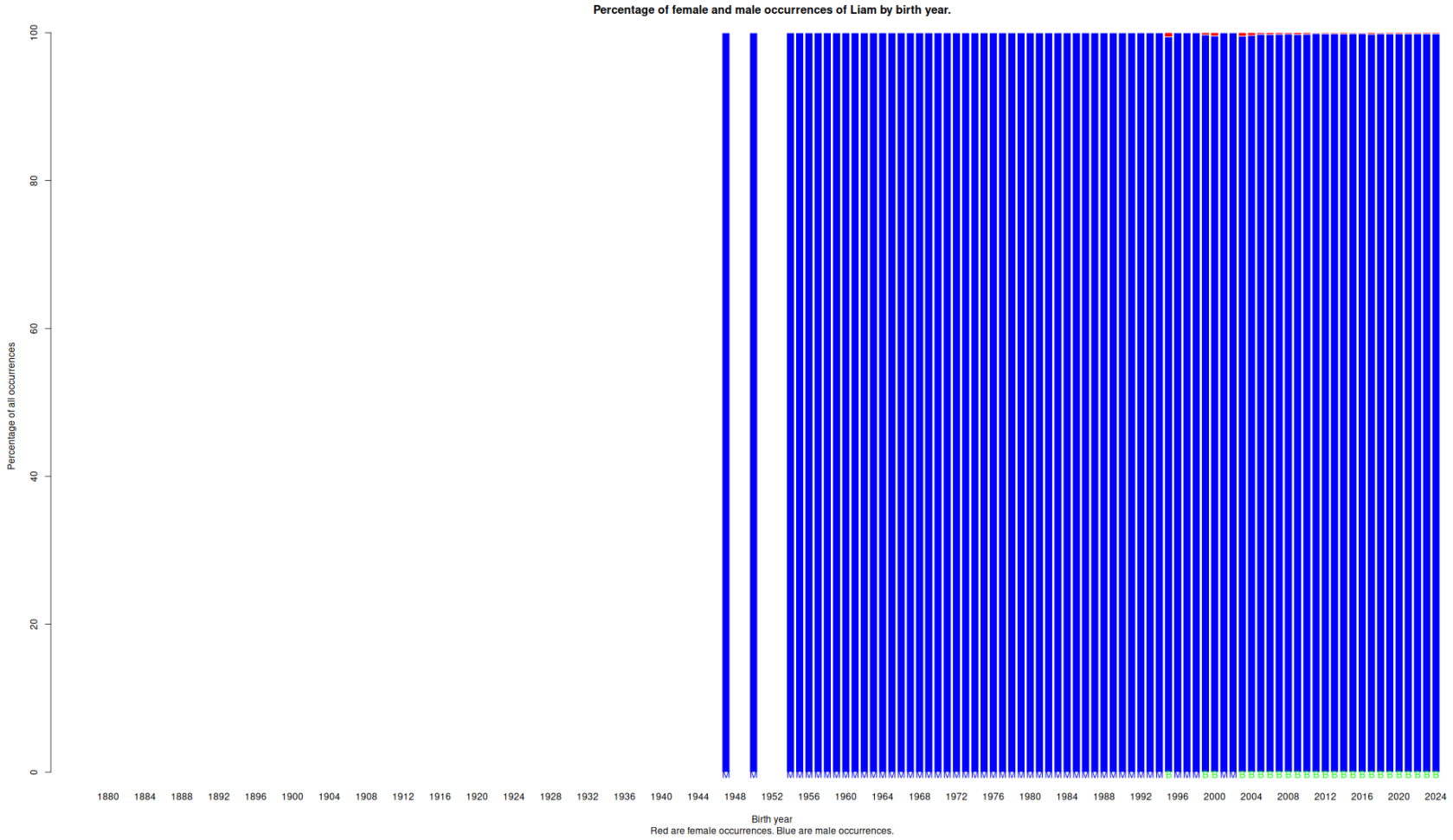


Figure 26: Lois, number of occurrences per year. Way down from it's peak, but still hanging on. Below each column has a red "F", or a blue "M", or a green "B" at it's base meaning all the data is in the column is Female, Male, or Both respectively.

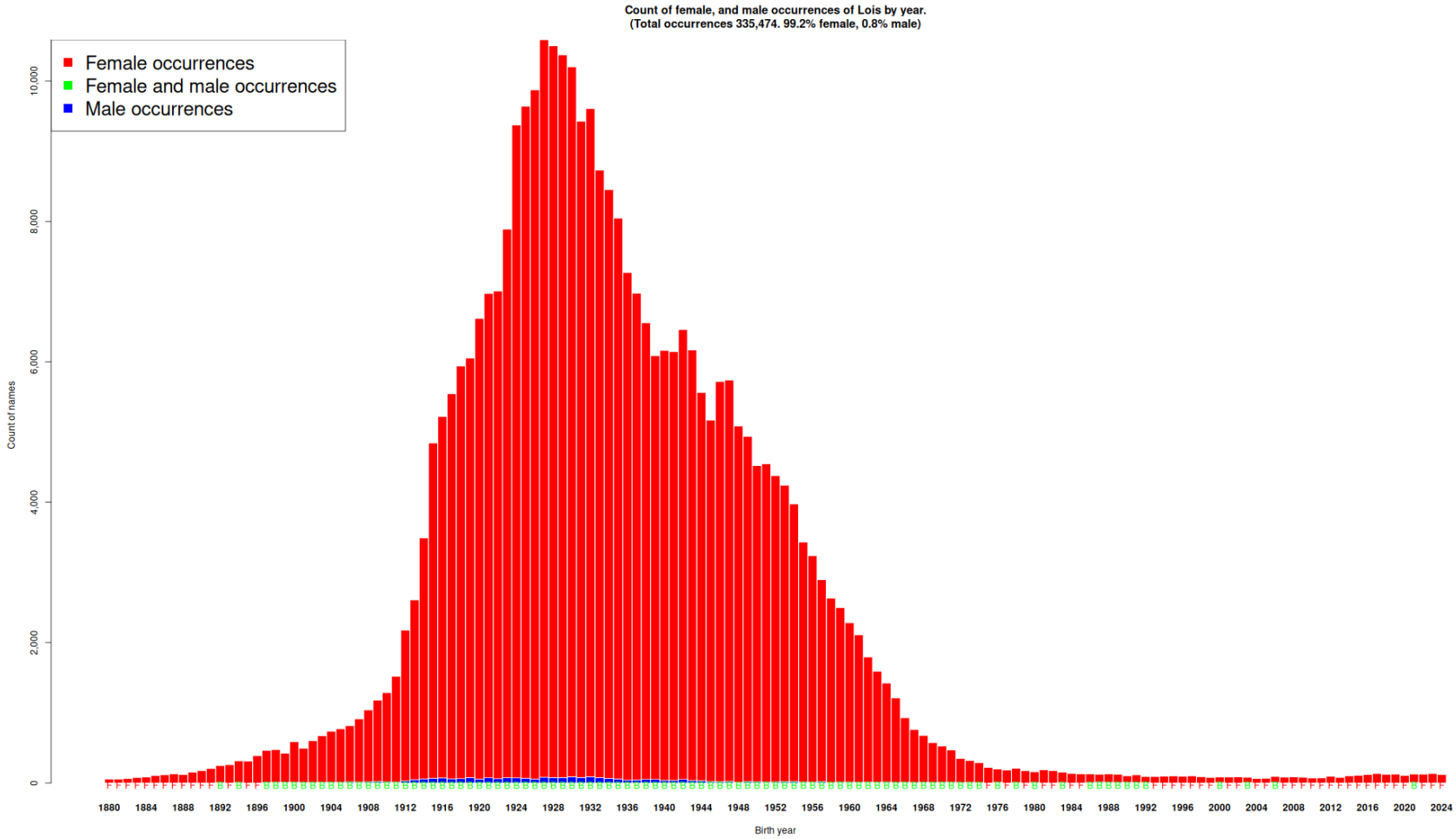


Figure 27: Lois, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

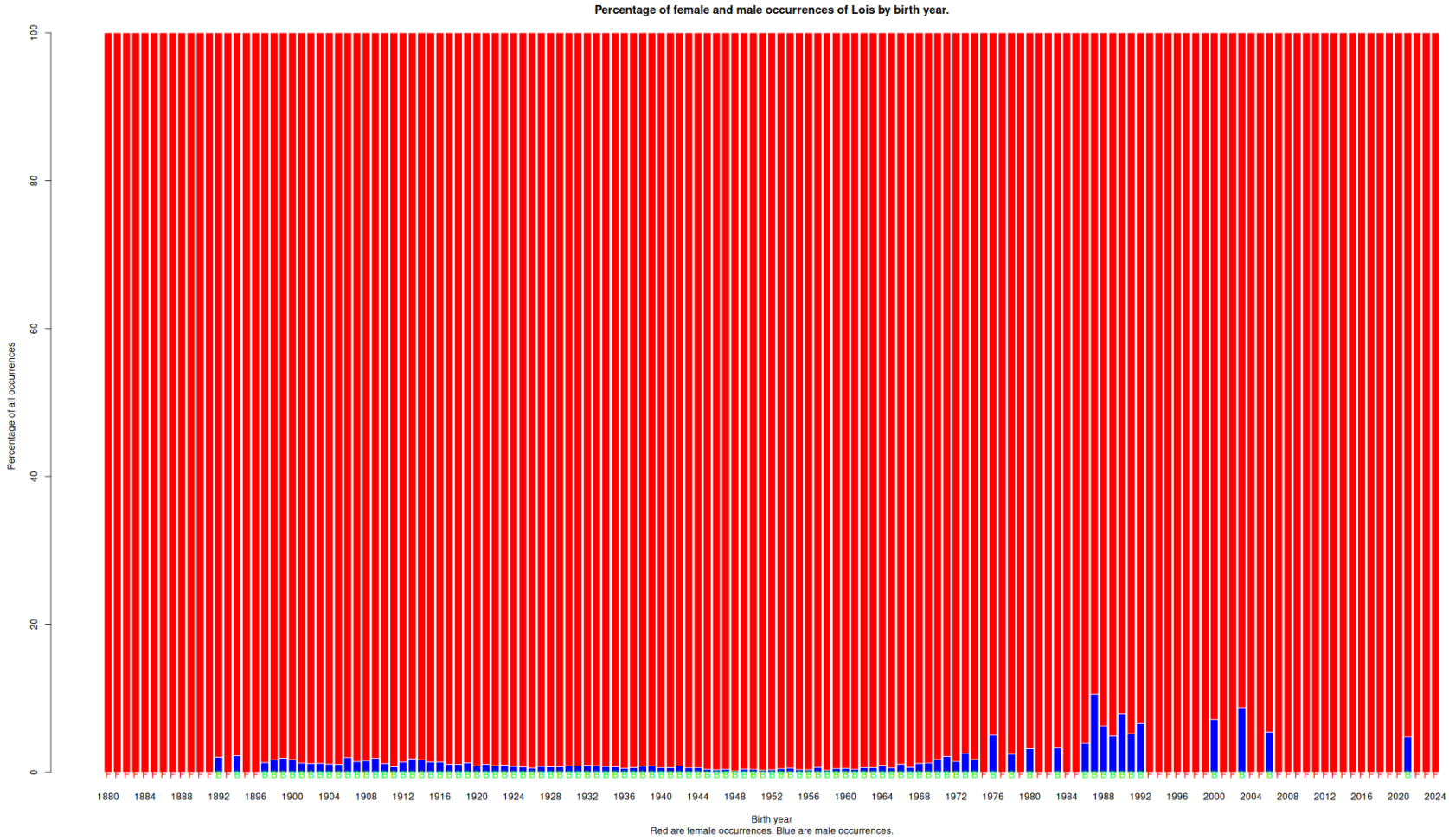


Figure 28: Lori, number of occurrences per year. Way down from it's peak, but still hanging on. Below each column has a red "F", or a blue "M", or a green "B" at it's base meaning all the data is in the column is Female, Male, or Both respectively.

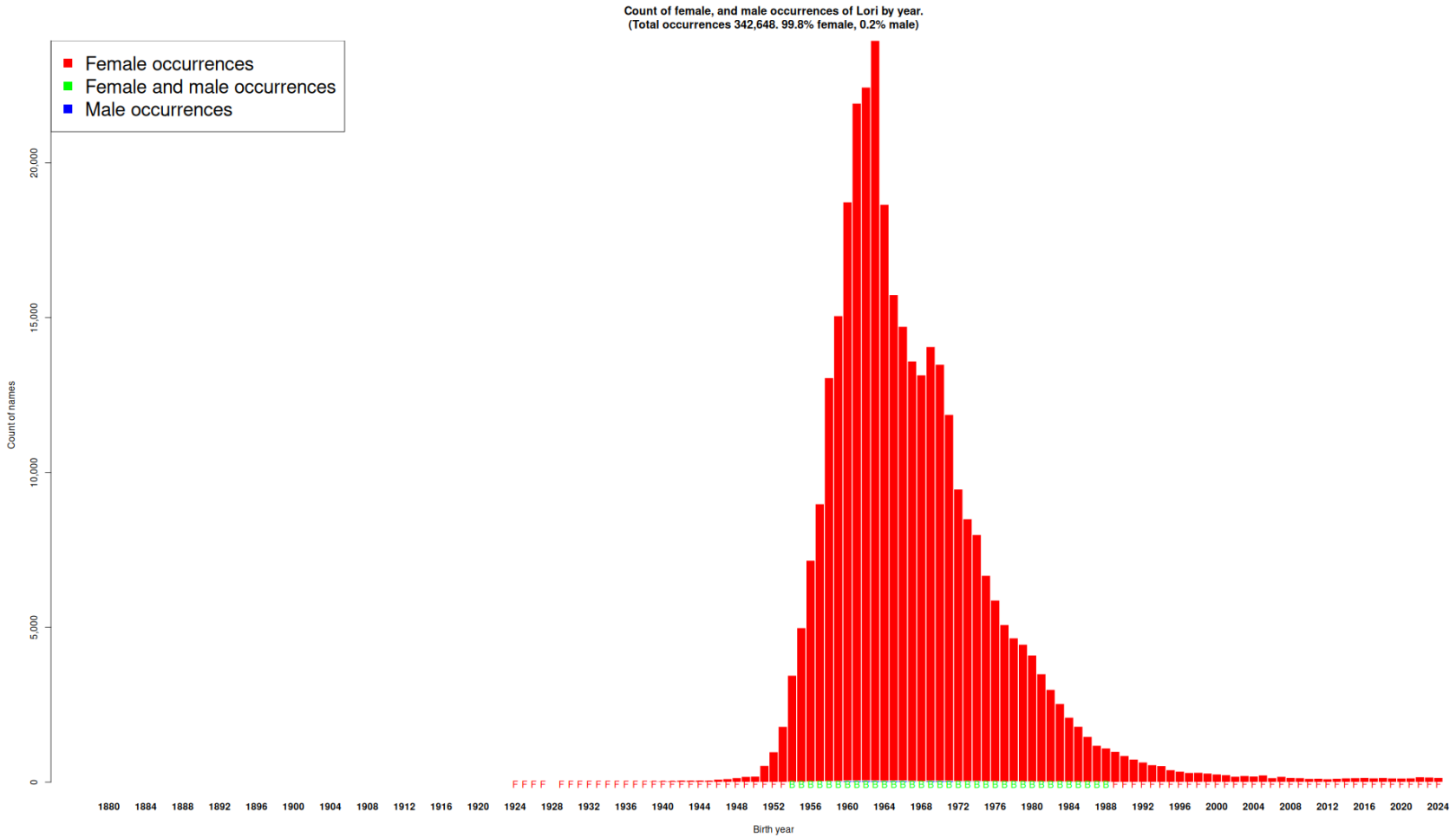


Figure 29: Lori, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

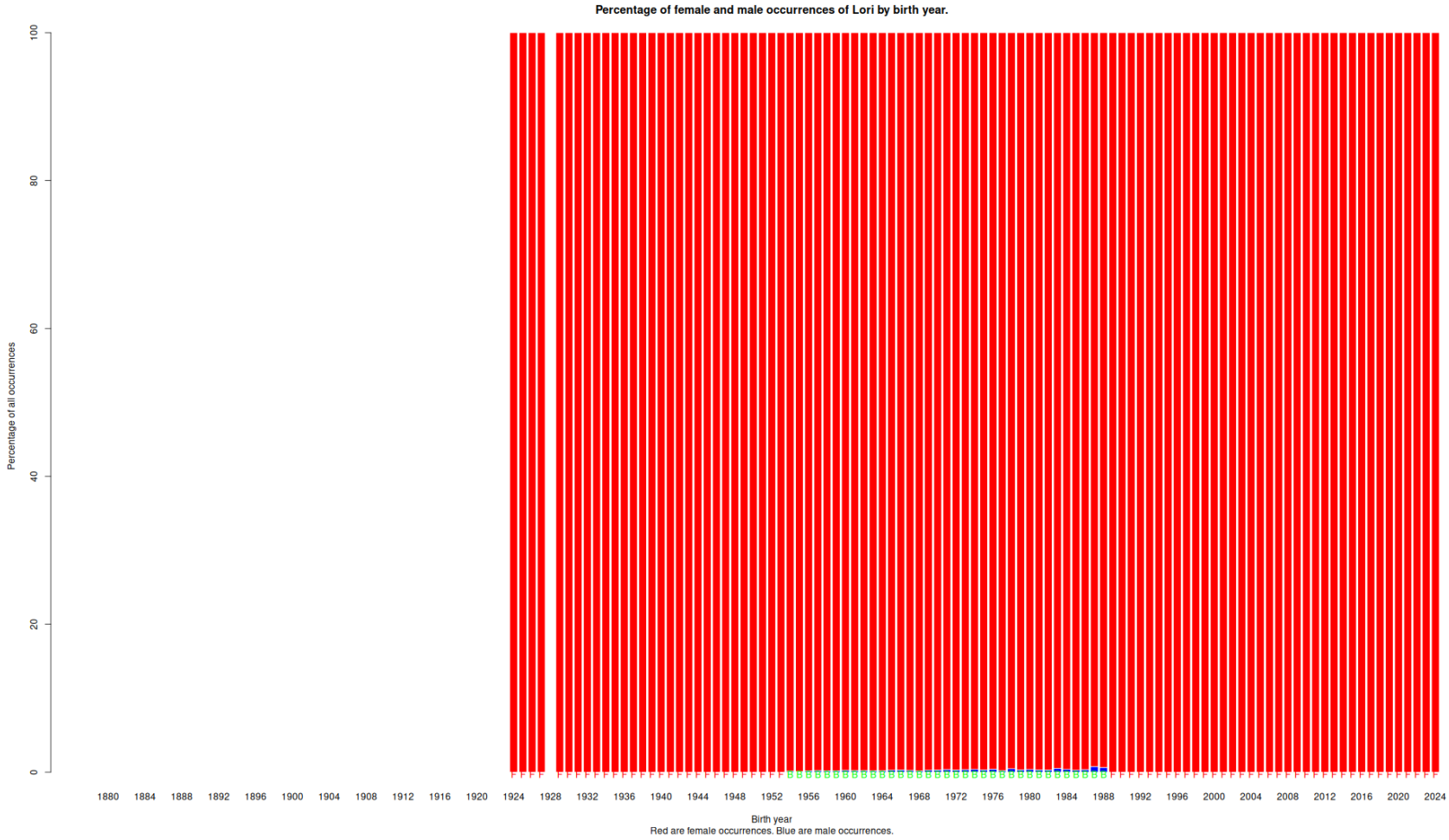


Figure 30: Louie, number of occurrences per year. Peaked in the 1910s and now coming back. Interesting resurgence. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

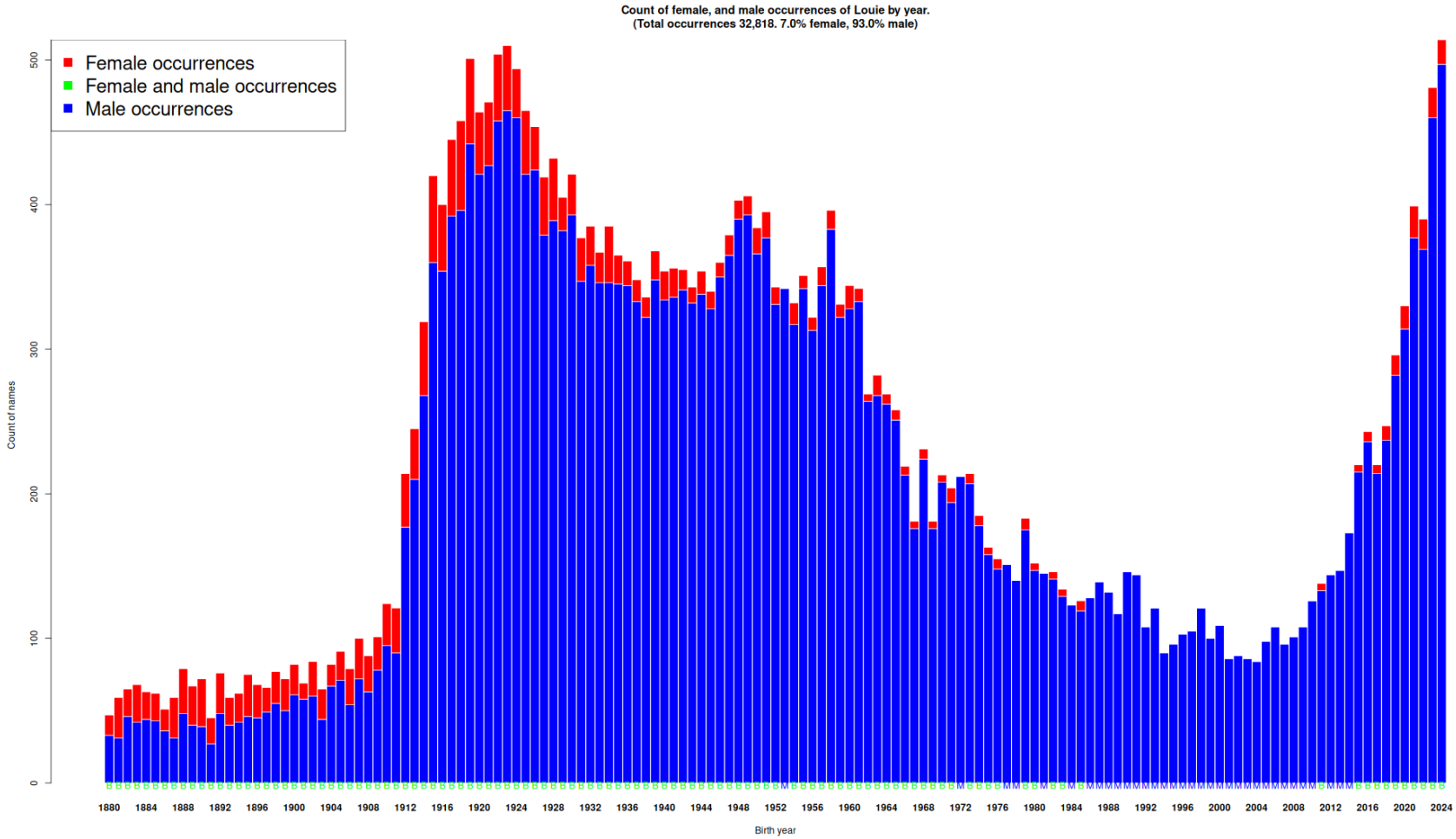


Figure 31: Louie, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

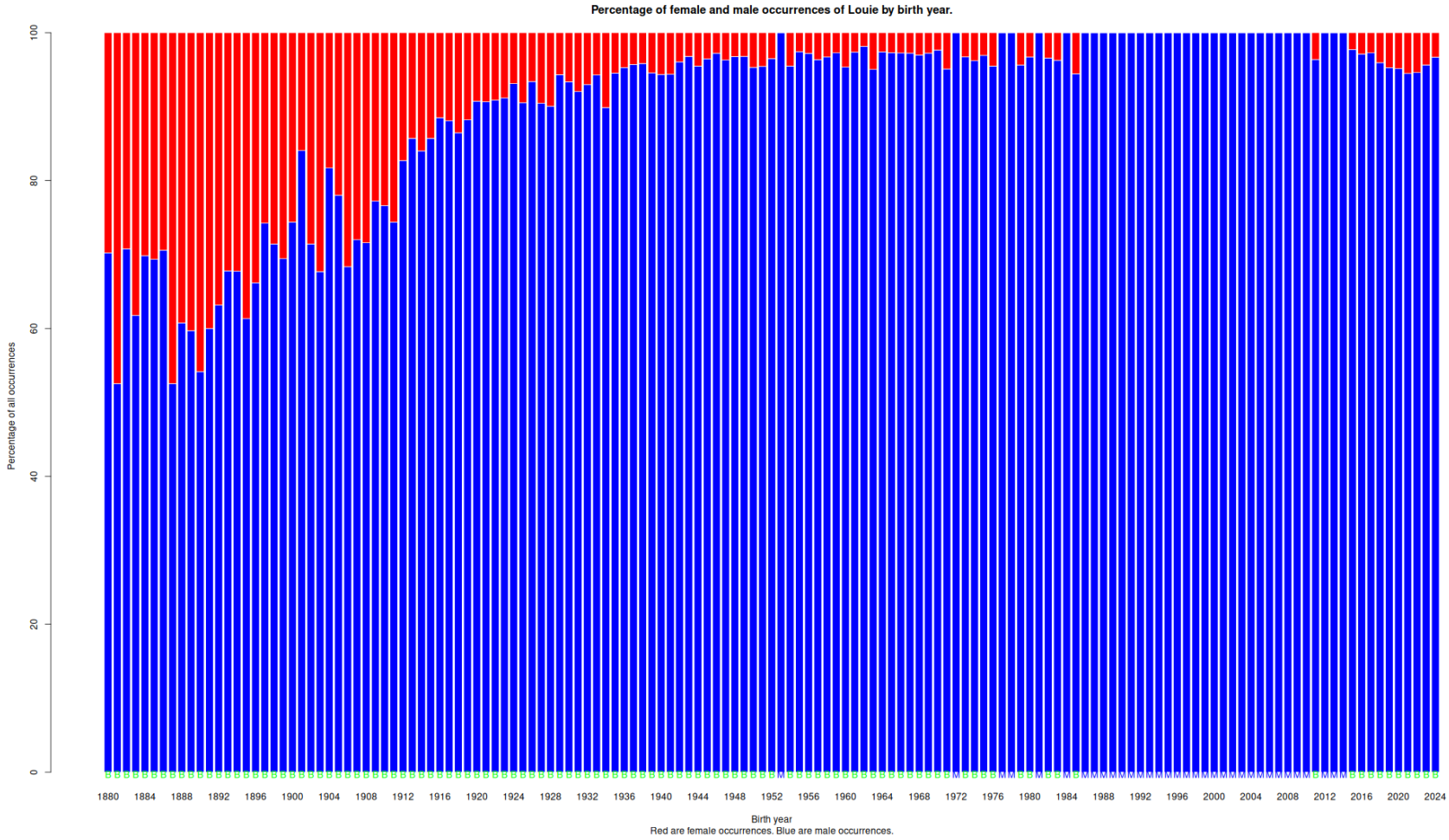


Figure 32: Marjorie, number of occurrences per year. Way down from it's peak, but still hanging on. Below each column has a red "F", or a blue "M", or a green "B" at it's base meaning all the data is in the column is Female, Male, or Both respectively.

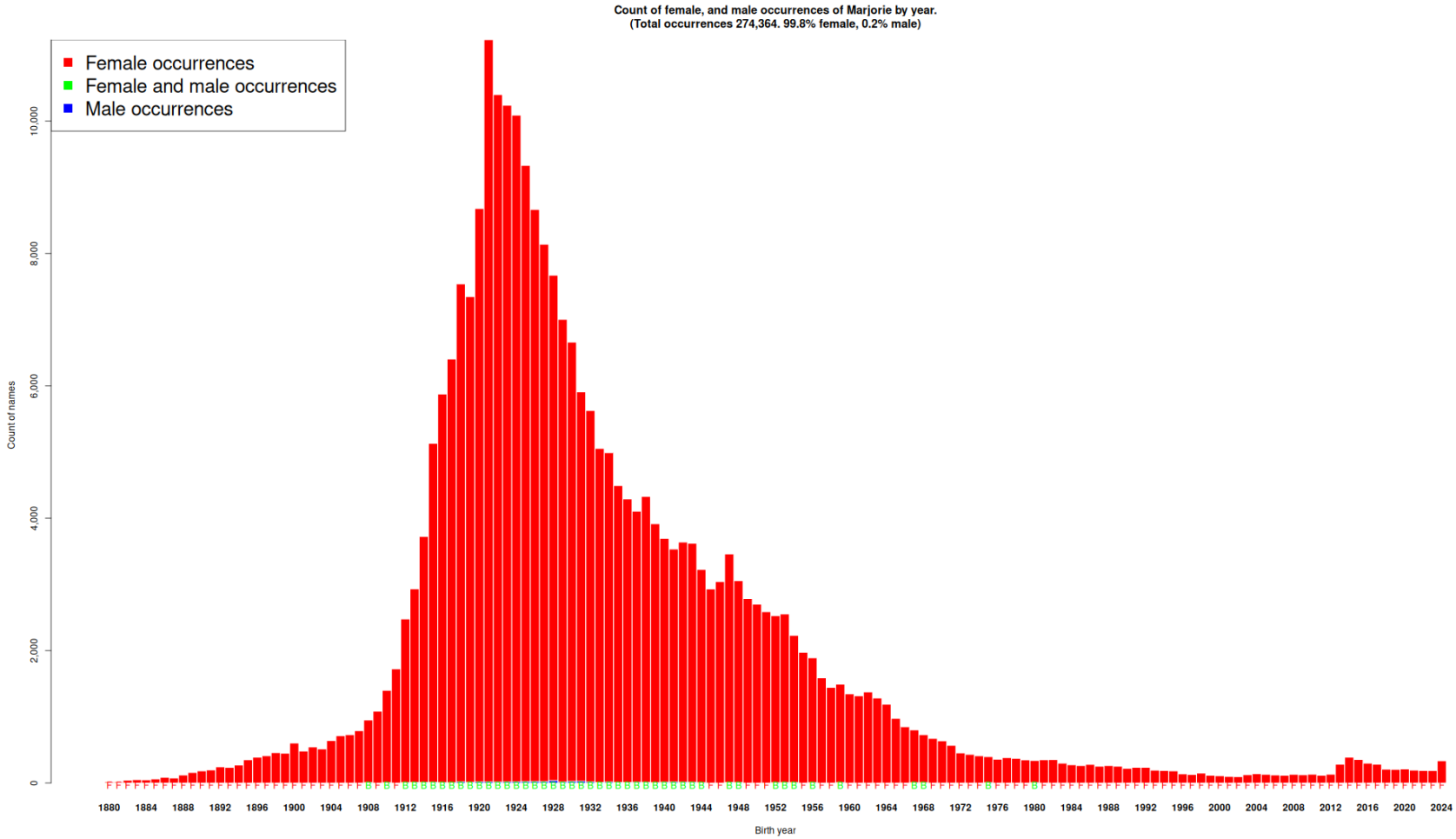


Figure 33: Marjorie, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

Figure 34: Mary, number of occurrences per year. Interesting inter-generational hump at 1910s and 1950s. Still hanging on. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

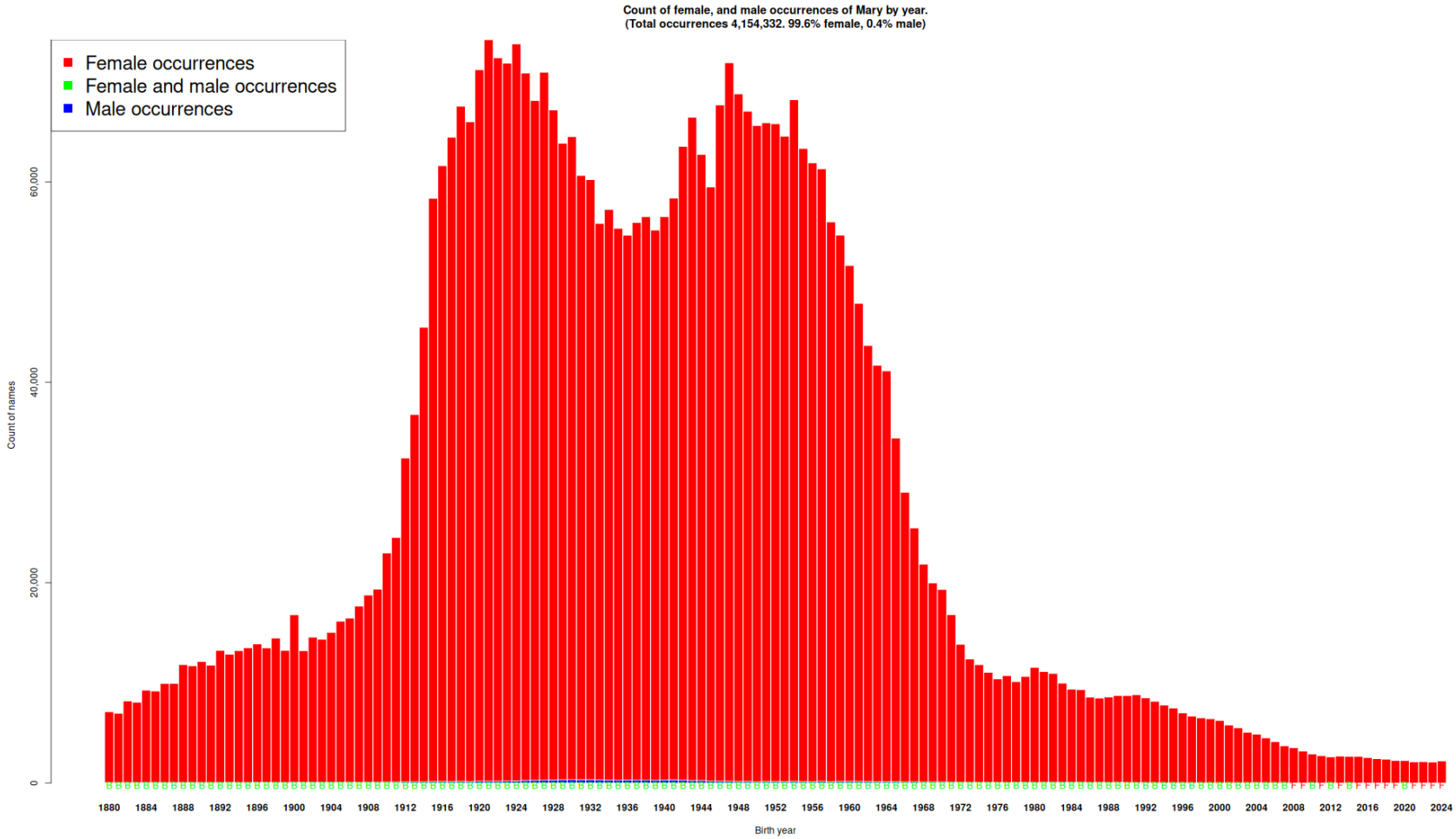


Figure 35: Mary, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

Figure 36: Mitchell, number of occurrences per year. Every 40ish years there is a come back. Interesting. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

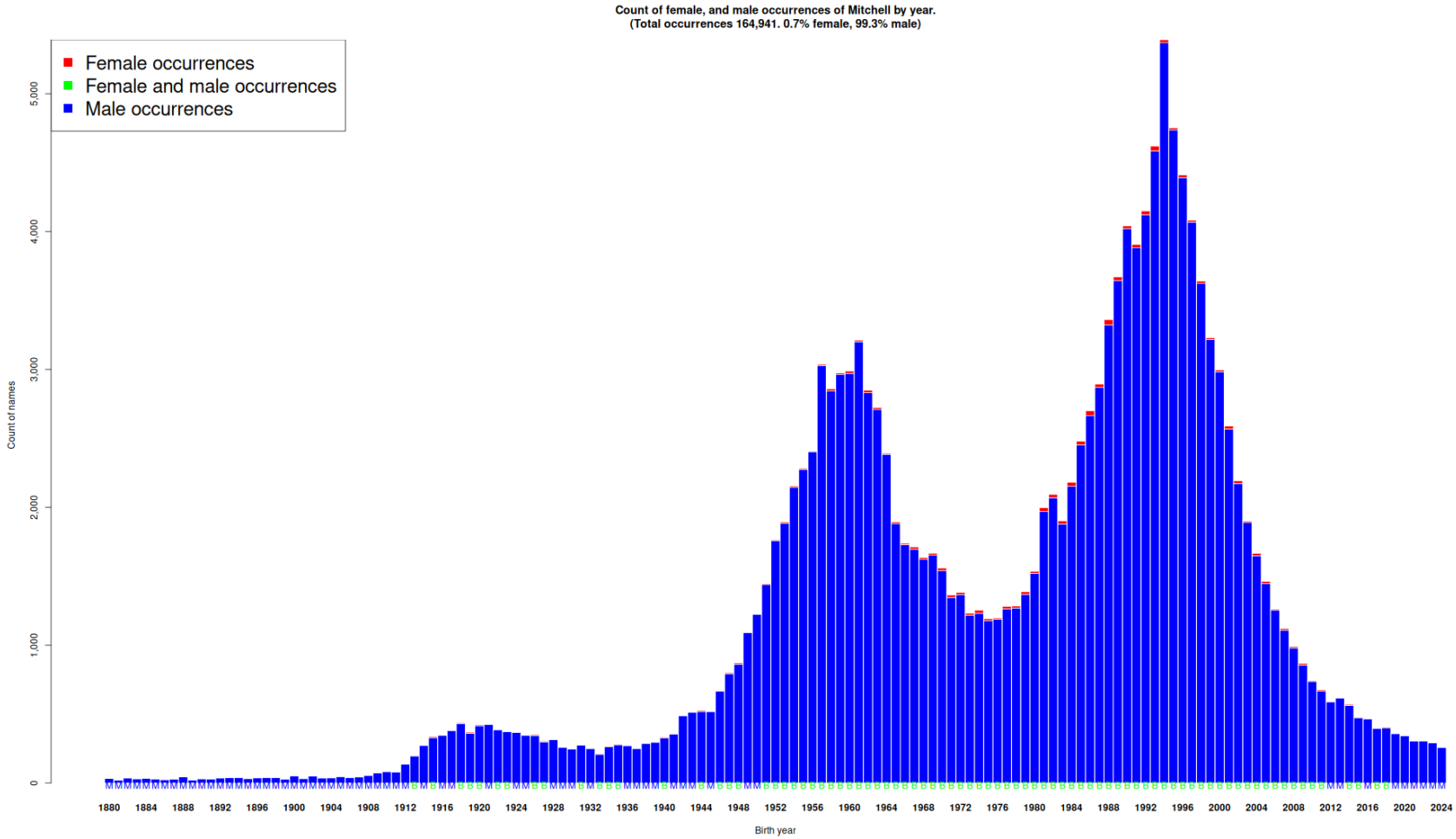


Figure 37: Mitchell, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

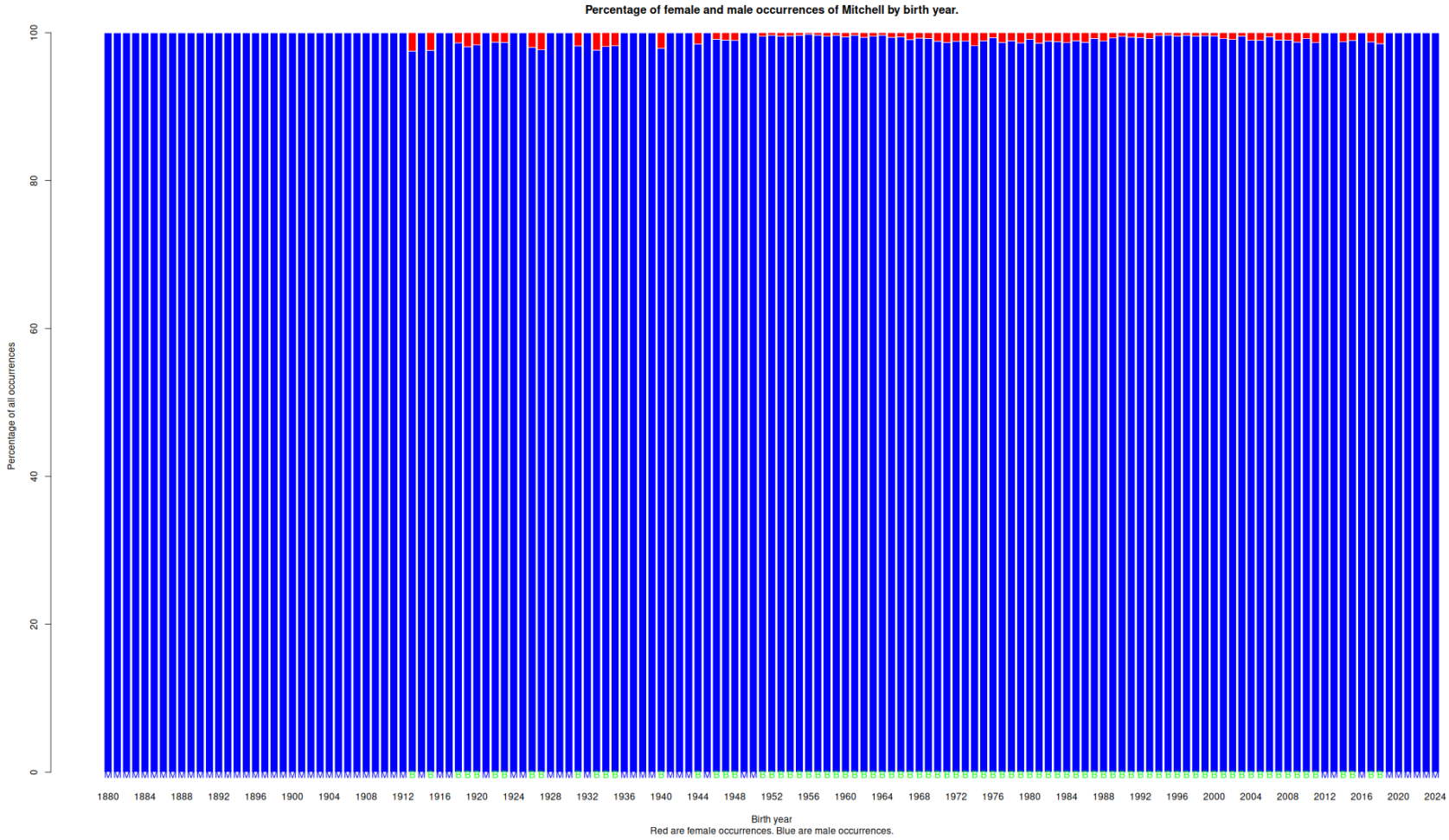


Figure 38: Peggy, number of occurrences per year. Looks like it was good for 30ish years, but now is almost gone. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

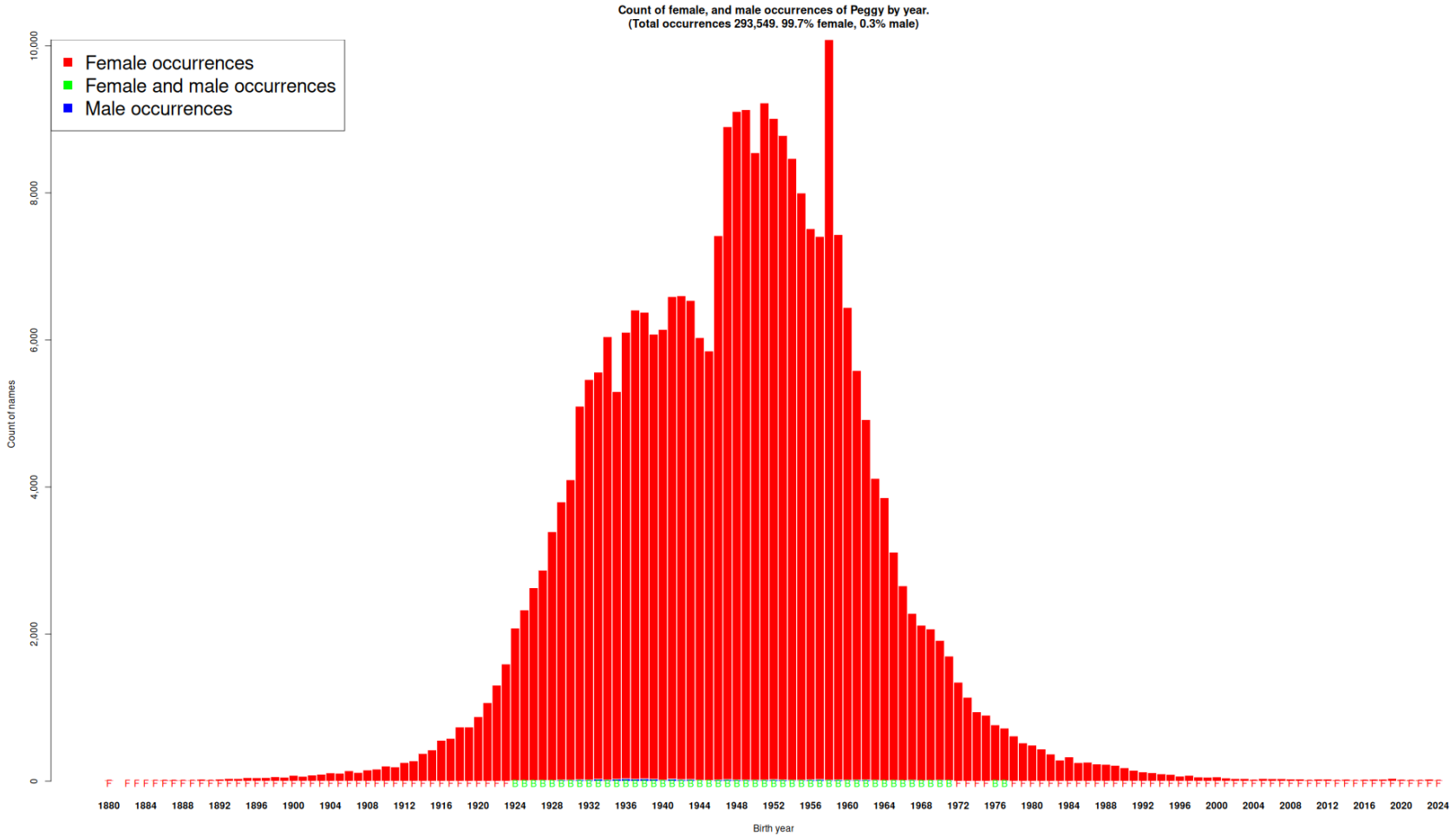


Figure 39: Peggy, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

Figure 40: Vera, number of occurrences per year. Peaked in the 1910s, but now it might be on a comeback. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

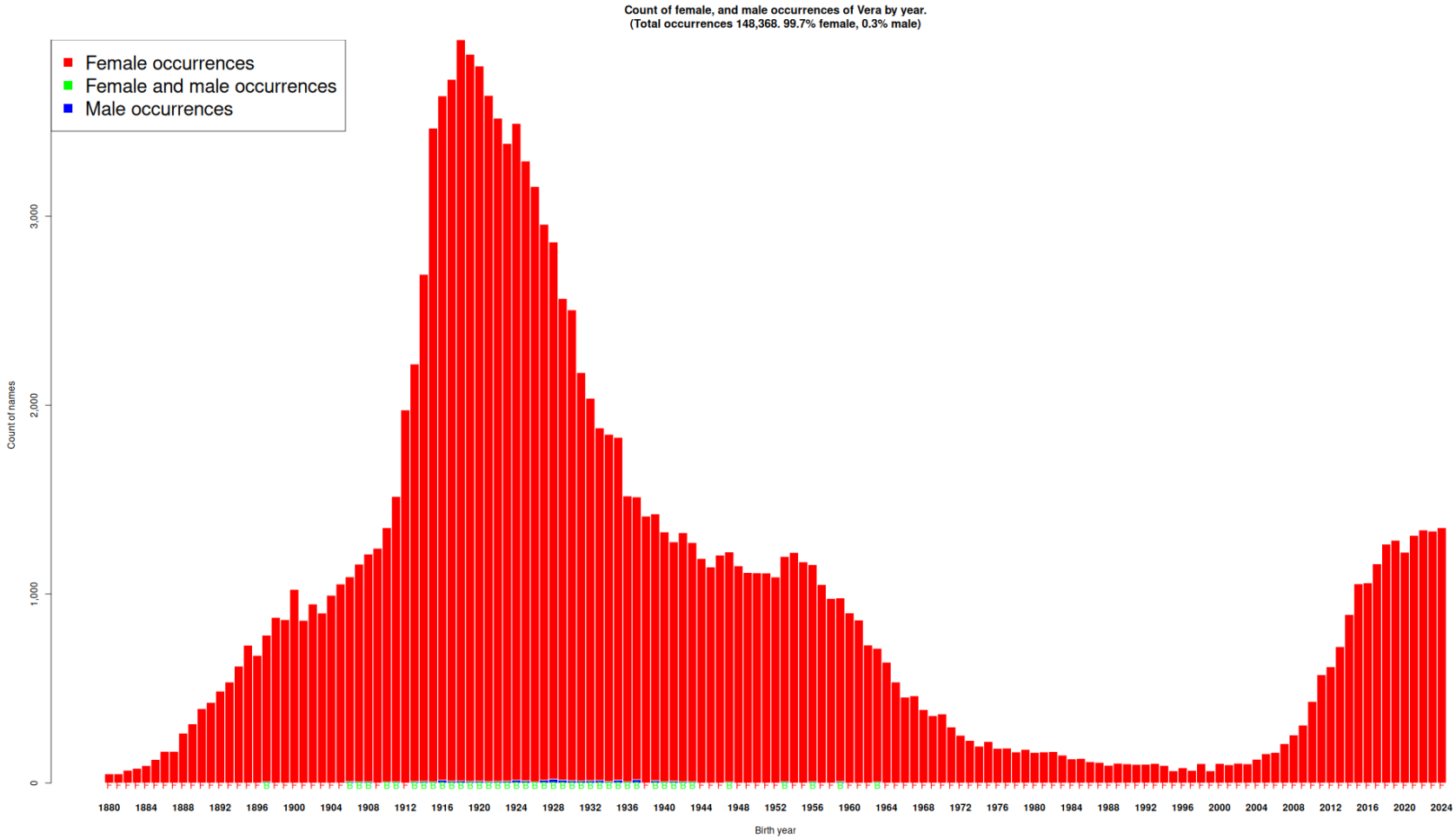


Figure 41: Vera, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

Figure 42: Vlasta, number of occurrences per year. A purely female name to a generation, and now looks to be lost. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.

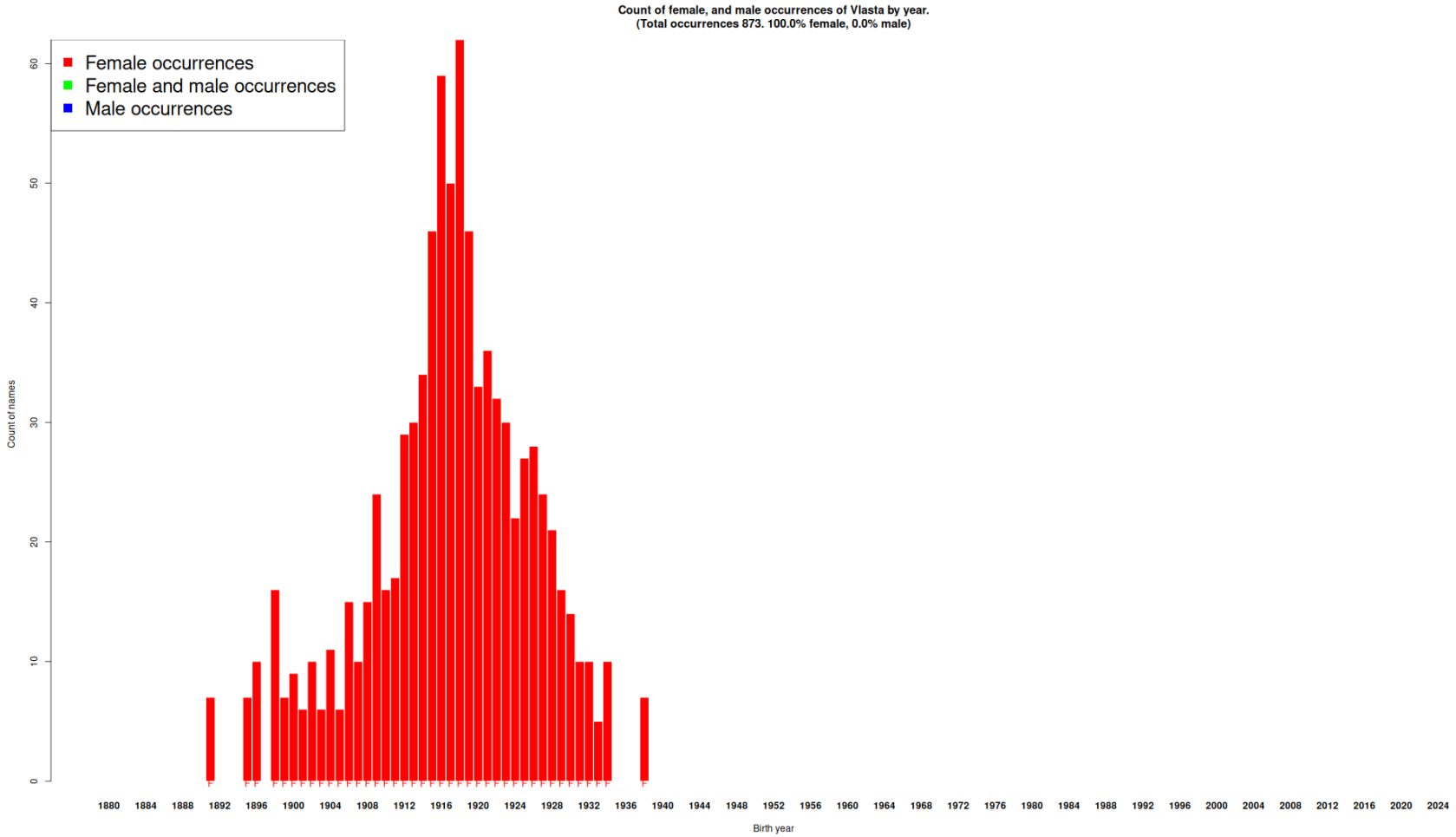
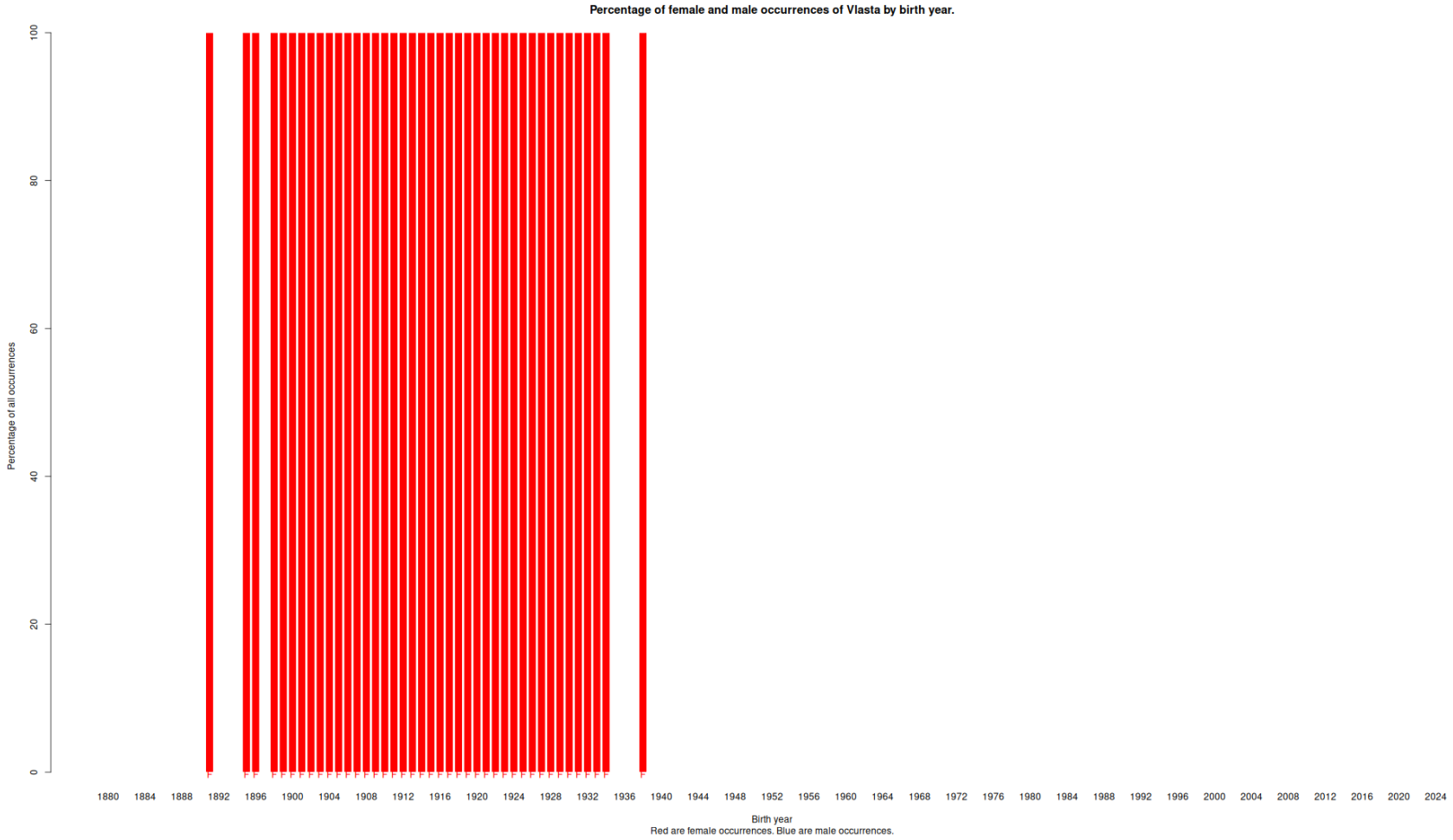


Figure 43: Vlasta, percentage by gender per year. Below each column has a red “F”, or a blue “M”, or a green “B” at it’s base meaning all the data is in the column is Female, Male, or Both respectively.



3 Conclusion

The Social Security Administration (SSA) encourages parents and everyone to get a Social Security Number (SSN) as soon as possible. An application form has to be submitted to get a unique SSN from the SSA. Selected yearly summary data from these application forms have been made available from the SSN for download. The data includes the first name of the applicant, their gender, and how often that name and gender appeared on applications in a particular year. Names that are strictly female or male, or both female and male remains relatively constant for all years reported (1880 through 2024). The percentage of strictly female or male, or both genders is also relatively constant across all years. Examination of these data show how applications and hence births may be affected by economic recessions and depressions.

Detailed examination of 19 selected names shows distinct patterns. Some names wax and wane in popularity on 40 and 60 year cycles. Others live for only a short while and then are lost. Some peak decay to nothingness, or to a low steady number. An exhaustive automated analysis of the data may reveal other trends.

Examination of the names to see if there were trends in name length, or the ratio of vowels to consonants did not reveal anything interesting and was omitted from this report.

The source data, and R script used to analyze the data are embedded in this report.

4 Embedded files

Embedded files may not be extractable using conventional browsers. They are extractable using Adobe Acrobat and Reader software by right-clicking on the paperclip and saving the file to a local directory.

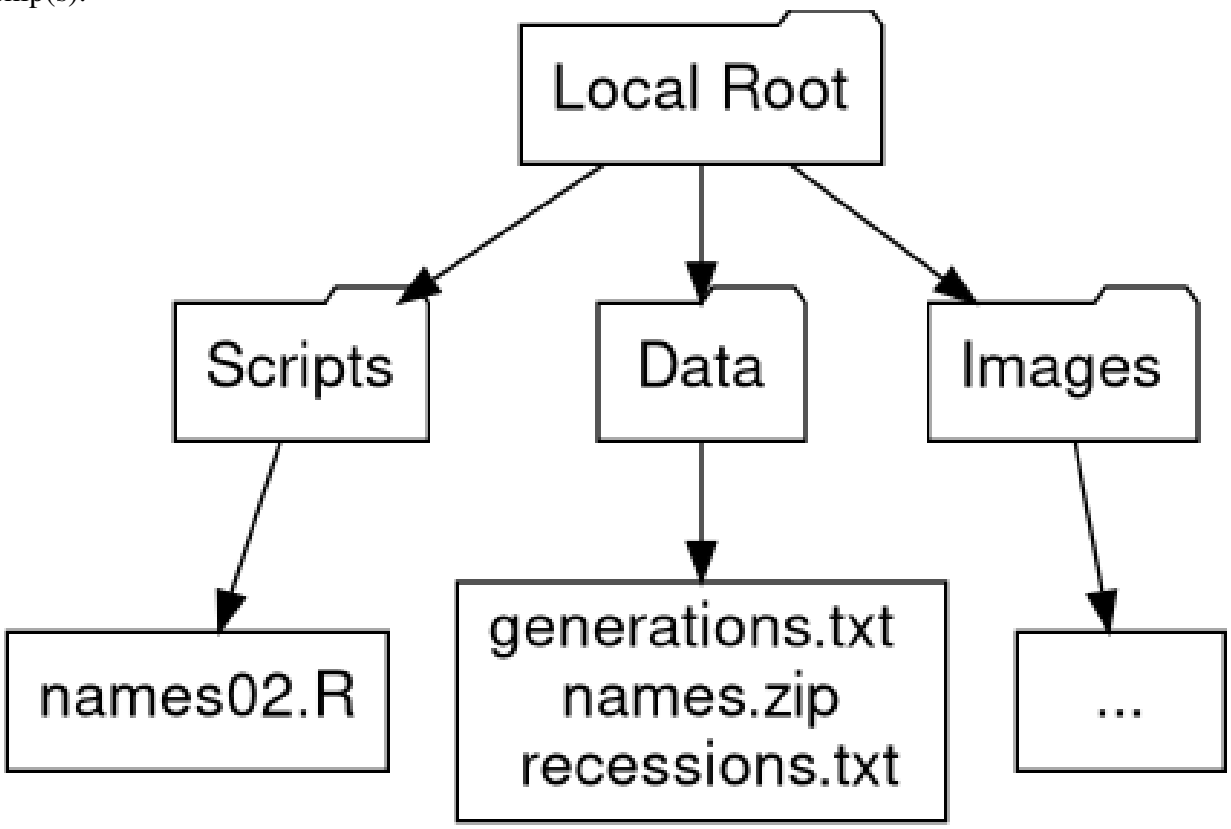
Other techniques to extract the images and attachments from this file are:

- `pdfimages -png ./babyNames.pdf /tmp/Temp/` – will extract all the images in `babyNames.pdf` in the current directory to the `/tmp/Temp` directory as png image files.
- `pdftk babyNames.pdf unpack_files` – will extract all the attachments in `babyNames.pdf` in the current directory to the current directory

4.1 R scripts and data files

The embedded scripts and files were written and tested using a specific file and directory structure (see Figure 44). `names02.R` uses the `setGlobalValue(...)` function to create global key-value pairs to point to the appropriate directory locations. If your installation is different than the assumed relationship, then use `setGlobalValue(...)` inside `main()` to point to the correct locations. Change the `captureImages` to `TRUE` inside `main()` to automatically save images to the `Images` directory. Execute the `main()` function to process all data, and generate all plots.

Figure 44: Assumed directory and file structure. If the relationship between `names02.R` and the other directories/files is different, then use `setGlobalValue(...)` to reflect the new relationship(s).



- names02.R – The R script file used to extract and analyze names from SSA number applications.
- generations.txt – The start and stop years for various generations from 1883 – 2025. `https://en.wikipedia.org/wiki/Generation_X`
- names.zip – Source of SSA Number year, name, and gender data. `https://catalog.data.gov/dataset/baby-names-from-social-security-card-applications-`
- recessions.txt – The start and stop years for recessions and depressions from 1882 – 2020. `https://en.wikipedia.org/wiki/List_of_recessions_in_the_United_States`

4.2 Full sized name image files

1. Candy-byYear-absolute.png – Plot of the absolute number of occurrences of Candy by year.
2. Candy-byYear-percentage.png – Plot of the gender percentages of Candy by year.
3. Charles-byYear-absolute.png – Plot of the absolute number of occurrences of Charles by year.
4. Charles-byYear-percentage.png – Plot of the gender percentages of Charles by year.
5. Colleen-byYear-absolute.png – Plot of the absolute number of occurrences of Colleen by year.
6. Colleen-byYear-percentage.png – Plot of the gender percentages of Colleen by year.
7. Cynthia-byYear-absolute.png – Plot of the absolute number of occurrences of Cynthia by year.
8. Cynthia-byYear-percentage.png – Plot of the gender percentages of Cynthia by year.
9. Geoffrey-byYear-absolute.png – Plot of the absolute number of occurrences of Geoffrey by year.
10. Geoffrey-byYear-percentage.png – Plot of the gender percentages of Geoffrey by year.
11. Jennifer-byYear-absolute.png – Plot of the absolute number of occurrences of Jennifer by year.
12. Jennifer-byYear-percentage.png – Plot of the gender percentages of Jennifer by year.
13. John-byYear-absolute.png – Plot of the absolute number of occurrences of John by year.
14. John-byYear-percentage.png – Plot of the gender percentages of John by year.

15. Lane-byYear-absolute.png – Plot of the absolute number of occurrences of Lane by year.
16. Lane-byYear-percentage.png – Plot of the gender percentages of Lane by year.
17. Lee-byYear-absolute.png – Plot of the absolute number of occurrences of Lee by year.
18. Lee-byYear-percentage.png – Plot of the gender percentages of Lee by year.
19. Liam-byYear-absolute.png – Plot of the absolute number of occurrences of Liam by year.
20. Liam-byYear-percentage.png – Plot of the gender percentages of Liam by year.
21. Lois-byYear-absolute.png – Plot of the absolute number of occurrences of Lois by year.
22. Lois-byYear-percentage.png – Plot of the gender percentages of Lois by year.
23. Lori-byYear-absolute.png – Plot of the absolute number of occurrences of Lori by year.
24. Lori-byYear-percentage.png – Plot of the gender percentages of Lori by year.
25. Louie-byYear-absolute.png – Plot of the absolute number of occurrences of Louie by year.
26. Louie-byYear-percentage.png – Plot of the gender percentages of Louie by year.
27. Marjorie-byYear-absolute.png – Plot of the absolute number of occurrences of Marjorie by year.
28. Marjorie-byYear-percentage.png – Plot of the gender percentages of Marjorie by year.
29. Mary-byYear-absolute.png – Plot of the absolute number of occurrences of Mary by year.
30. Mary-byYear-percentage.png – Plot of the gender percentages of Mary by year.
31. Mitchell-byYear-absolute.png – Plot of the absolute number of occurrences of Mitchell by year.

32. Mitchell-byYear-percentage.png – Plot of the gender percentages of Mitchell by year.

33. Peggy-byYear-absolute.png – Plot of the absolute number of occurrences of Peggy by year.

34. Peggy-byYear-percentage.png – Plot of the gender percentages of Peggy by year.

35. Vera-byYear-absolute.png – Plot of the absolute number of occurrences of Vera by year.

36. Vera-byYear-percentage.png – Plot of the gender percentages of Vera by year.

37. Vlasta-byYear-absolute.png – Plot of the absolute number of occurrences of Vlasta by year.

38. Vlasta-byYear-percentage.png – Plot of the gender percentages of Vlasta by year.

References

[1] J Craig, *Replacement level fertility and future population growth*, https://pubmed.ncbi.nlm.nih.gov/7834459/, 1994.

[2] Tim Maxwell, *What Is the Difference Between a Recession and a Depression?*, https://www.experian.com/blogs/ask-experian/what-is-the-difference-between-a-recession-and-a-depression/, 2025.

[3] SSA Staff, *Social Security History*, https://www.ssa.gov/history/orghist.html.

[4] _____, *Baby Names from Social Security Card Applications - National Data* , https://catalog.data.gov/dataset/baby-names-from-social-security-card-applications-national-data, 2025.

[5] _____, *Social Security Numbers for Children*, https://www.ssa.gov/pubs/EN-05-10023.pdf, 2025.