

# Trademark Report July 24

*Aniket Kesari*

*July 24, 2017*

## Data Source

Based on the Trademarks Database, I constructed a new dataset that has information about the number of instances for each level of number of prior registration, per year. So, for example, this dataset tells us that in the year 1884, there were 8 instances of registered trademarks having 1 prior registration. Here I display the first 20 lines of the dataset to give an idea of its structure. “Year” refers to the calendar year, “prior\_regs” refers to the number of prior registrations a trademark had, “n” is the number of instances, and “prop” is the relative frequency of per year.

```
setwd("C:/Users/Owner/Dropbox/Trademark Rebrand Project")
library(tidyverse)
TrademarkFreq <- read_csv("Frequency Table of Prior Marks by Year.csv")
head(TrademarkFreq, n=20)
```

```
## # A tibble: 20 × 5
##       X1 Year prior_regs     n      prop
##   <int> <int>      <int> <int>    <dbl>
## 1     1  1823          1     2 1.00000000
## 2     2  1824          1     1 1.00000000
## 3     3  1832          1     1 1.00000000
## 4     4  1875          1     1 1.00000000
## 5     5  1884          1     8 1.00000000
## 6     6  1885          1     8 1.00000000
## 7     7  1886          1    14 1.00000000
## 8     8  1887          1     7 1.00000000
## 9     9  1888          1    12 1.00000000
## 10    10  1889          1    10 1.00000000
## 11    11  1890          1    13 1.00000000
## 12    12  1891          1    29 1.00000000
## 13    13  1892          1    52 1.00000000
## 14    14  1893          1    96 1.00000000
## 15    15  1894          1    68 0.97142857
## 16    16  1894          2     2 0.02857143
## 17    17  1895          1    74 1.00000000
## 18    18  1896          1    75 1.00000000
## 19    19  1897          1    51 1.00000000
## 20    20  1898          1    39 1.00000000
```

## Descriptive Statistics

Here are descriptive statistics for the entire dataset. The summary divides each column into minimum, 1st quantile, Median, Mean, 3rd Quantile, and Maximum. Both the prior registration (prior\_regs) and number of instances (n) have extreme observations, making the Mean a poor measure in this case.

```
print(summary(TrademarkFreq))
```

```
##           X1           Year      prior_regs           n
## Min.      : 1.0    Min.      :1823    Min.      : 1.00    Min.      : 1
## 1st Qu.: 426.5    1st Qu.:1946    1st Qu.: 4.00    1st Qu.: 44
## Median : 852.0    Median :1988    Median : 10.00   Median : 134
## Mean     : 852.0    Mean     :1973    Mean     : 15.02   Mean     : 5485
## 3rd Qu.:1277.5    3rd Qu.:2000    3rd Qu.: 20.00   3rd Qu.: 711
## Max.     :1703.0    Max.     :2017    Max.     :121.00   Max.     :319107
##
##      prop
## Min.      :0.0000233
## 1st Qu.:0.0007087
## Median :0.0050025
## Mean     :0.0810335
## 3rd Qu.:0.0332545
## Max.     :1.0000000
```

Dividing the dataset to pre- and post-1989:

```
# Pre-1989
TrademarkFreq %>%
  filter(Year <= 1989) %>%
  summary()
```

```
##           X1           Year      prior_regs           n
## Min.      : 1    Min.      :1823    Min.      : 1.000    Min.      : 1
## 1st Qu.:228    1st Qu.:1932    1st Qu.: 3.000    1st Qu.: 32
## Median :455    Median :1947    Median : 6.000    Median : 91
## Mean     :455    Mean     :1950    Mean     : 8.752    Mean     : 2699
## 3rd Qu.:682    3rd Qu.:1974    3rd Qu.:12.000    3rd Qu.: 513
## Max.     :909    Max.     :1989    Max.     :57.000    Max.     :171761
##
##      prop
## Min.      :0.0000589
## 1st Qu.:0.0047619
## Median :0.0206372
## Mean     :0.1210121
## 3rd Qu.:0.0547445
## Max.     :1.0000000
```

```
# Post-1989
TrademarkFreq %>%
  filter(Year >= 1989) %>%
  summary()
```

```
##           X1           Year      prior_regs           n
## Min.      : 874    Min.      :1989    Min.      : 1.00    Min.      : 3.0
## 1st Qu.:1081    1st Qu.:1995    1st Qu.: 8.00    1st Qu.: 63.0
## Median :1288    Median :2000    Median : 17.00   Median : 200.0
## Mean     :1288    Mean     :2000    Mean     : 22.14   Mean     : 8416.1
## 3rd Qu.:1496    3rd Qu.:2004    3rd Qu.: 30.00   3rd Qu.: 837.5
## Max.     :1703    Max.     :2017    Max.     :121.00   Max.     :319107.0
##
##      prop
## Min.      :0.0000233
## 1st Qu.:0.0002763
## Median :0.0009097
## Mean     :0.0349398
## 3rd Qu.:0.0038888
## Max.     :0.9897454
```

The summary statistics clearly reveal that the number of instances of Trademarks having prior registrations was much higher in the nearly 30 years since 1989 than in the 100 years before that. There are also clearly some extreme outliers in the number of prior registrations as the median is 6 pre-1989 and 17 post-1989, with maxes of 57 and 121 respectively. A cursory look at the data indicates that for any given year, trademarks with one prior registration make up the bulk of registrations.

To visualize the distribution of the number of prior registrations by here, consult this Scatterplot:

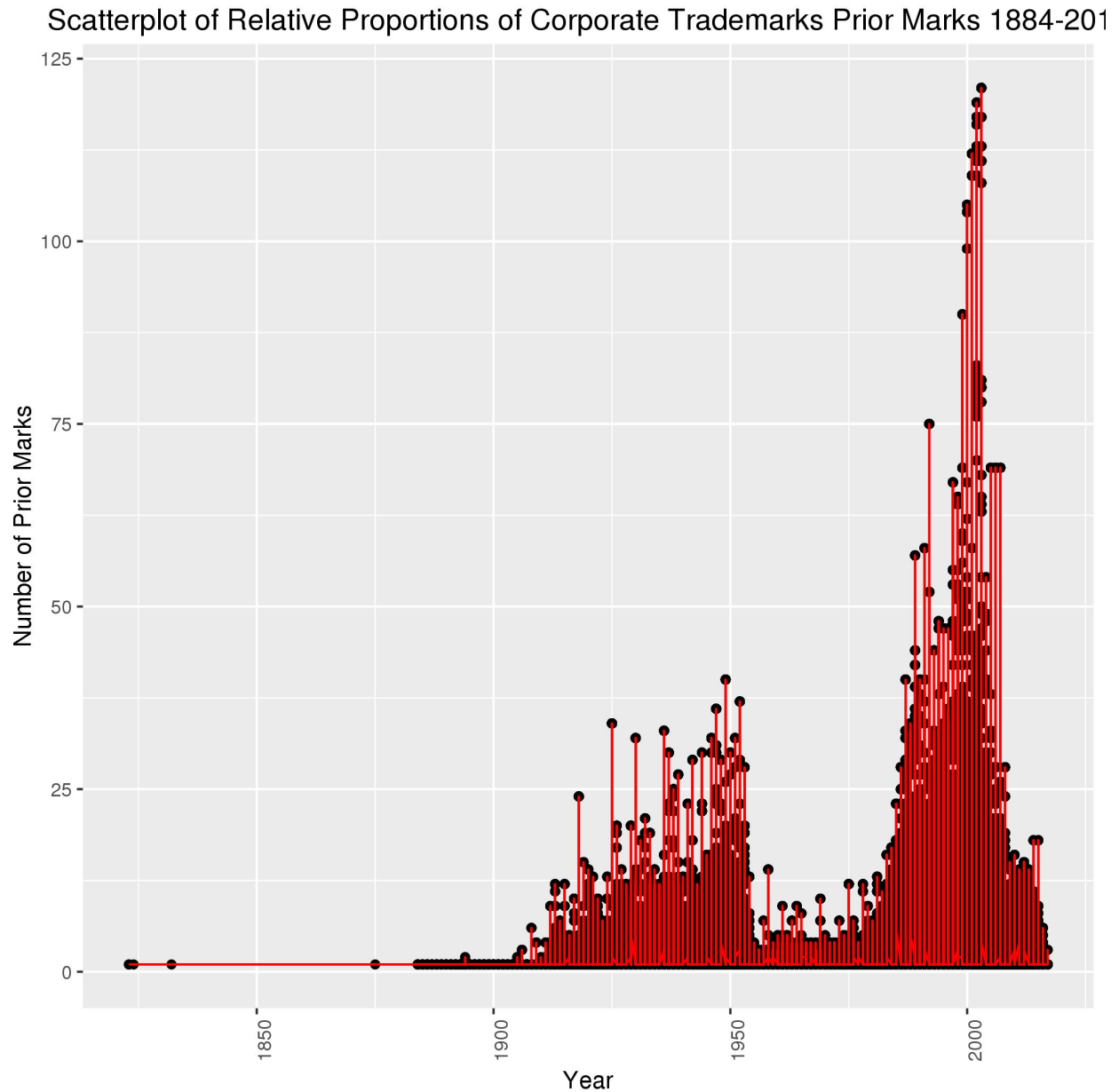


Figure 1: Scatterplot of Prior Registrations by Year

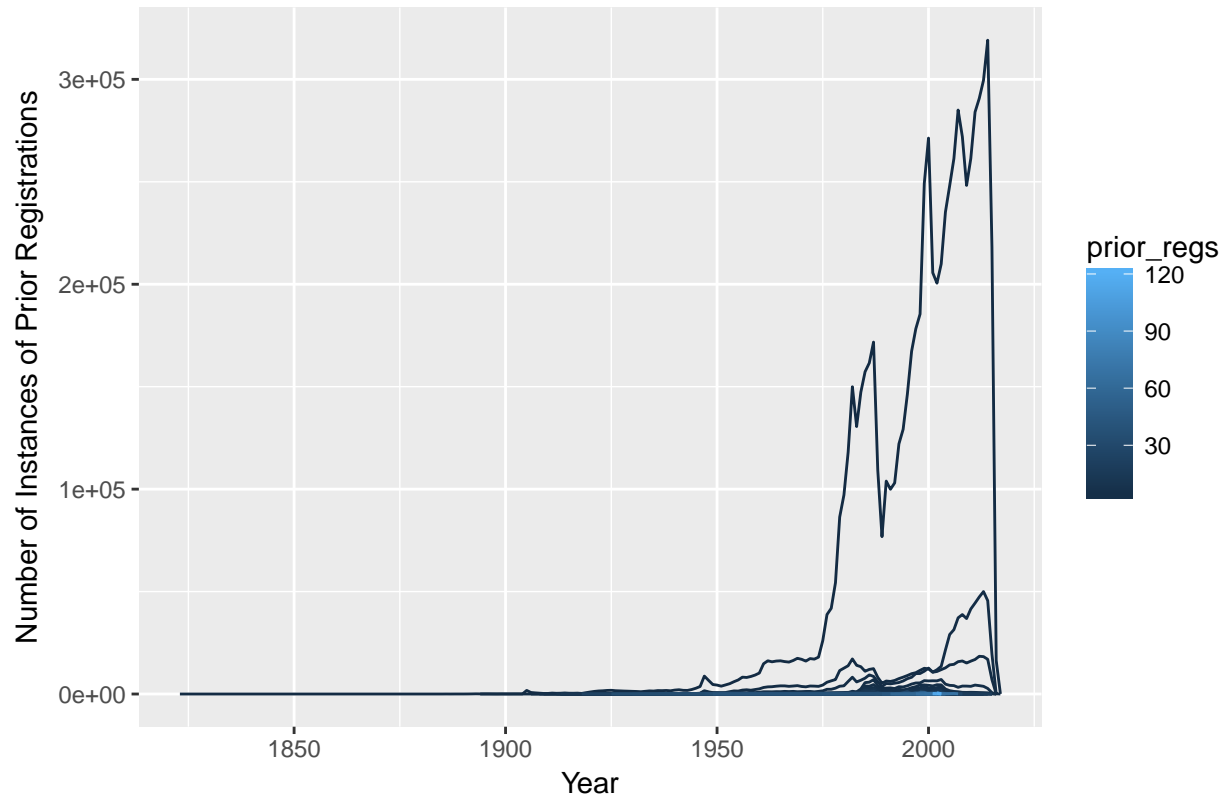
## Growth in Trademark Registrations

Now, to visualize the growth:

```
# All observations
```

```
TrademarkFreq %>% ggplot() + geom_line(aes(x=Year, y=n, colour=prior_regs, group=prior_regs)) + xlab("Y
```

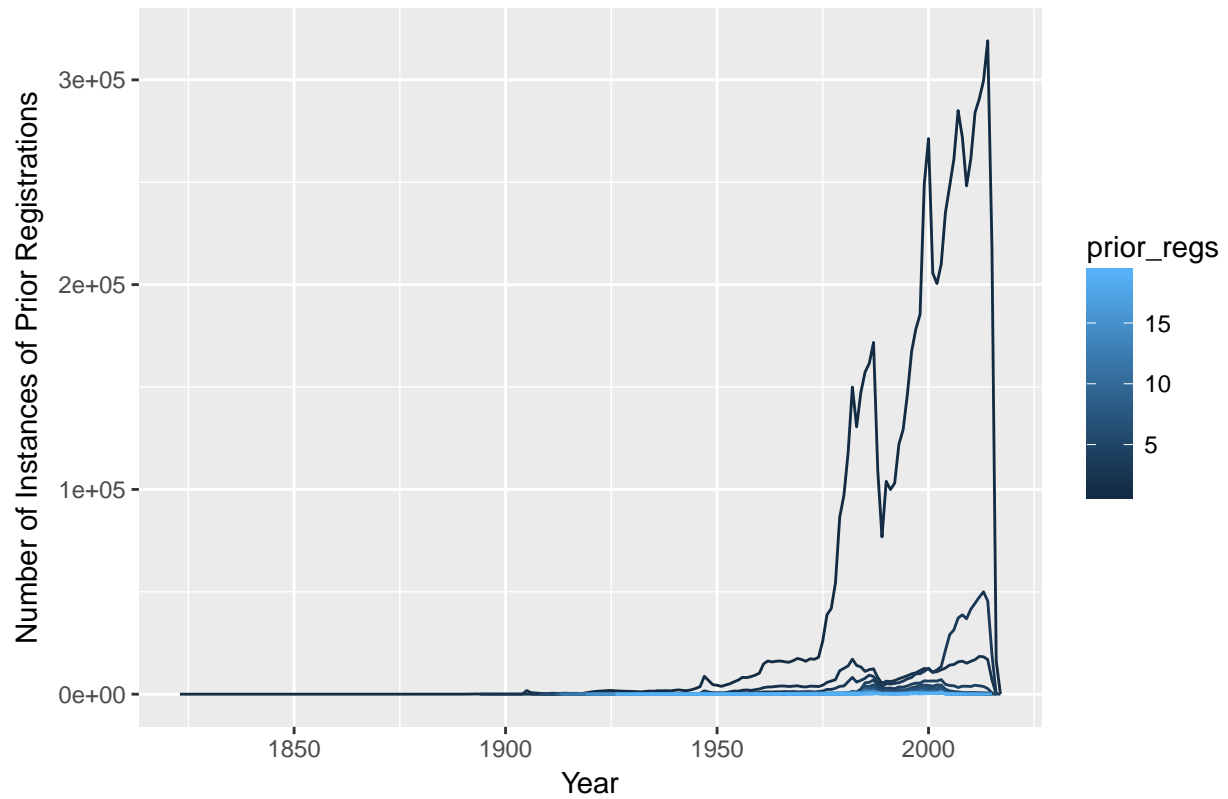
Line Graph of All Data



```
# Take out extreme number of prior registrations (75th percentile)
```

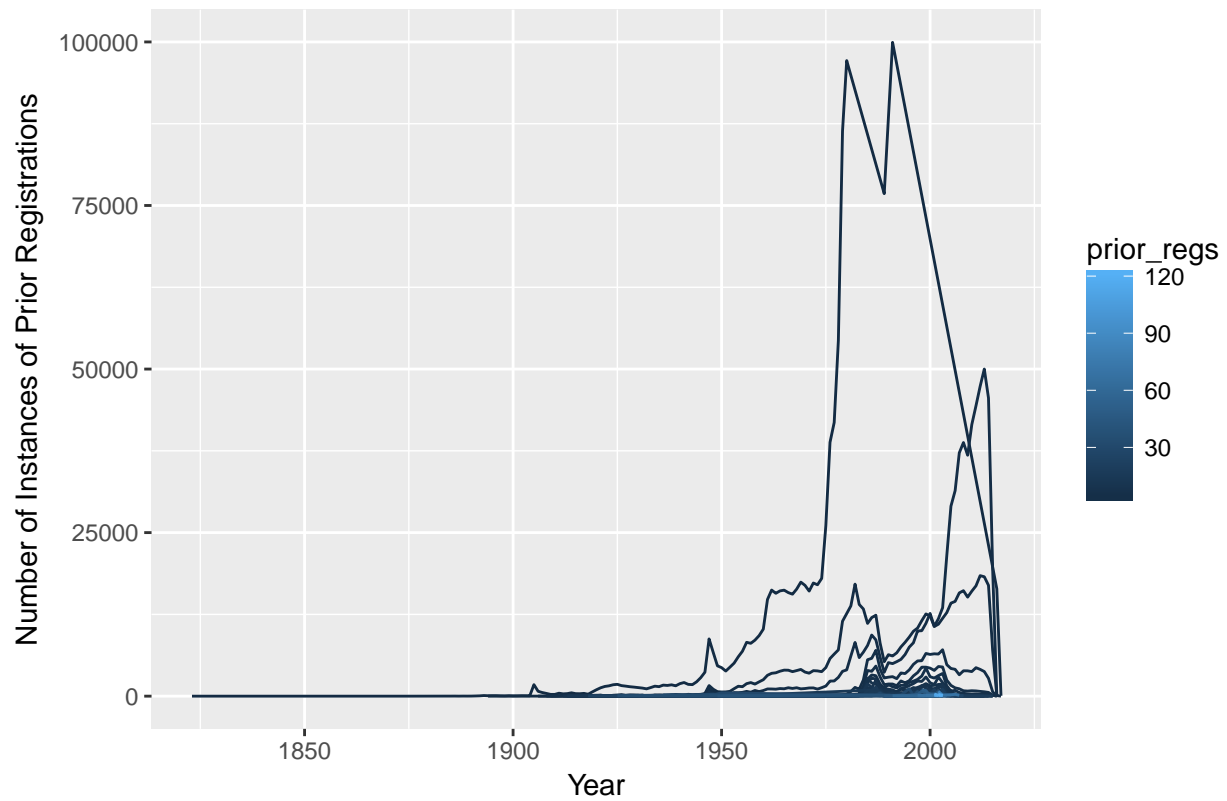
```
TrademarkFreq %>% filter(prior_regs < 20) %>% ggplot() + geom_line(aes(x=Year, y=n, colour=prior_regs, g
```

Line Graph Without Extreme Values of Prior Registrations



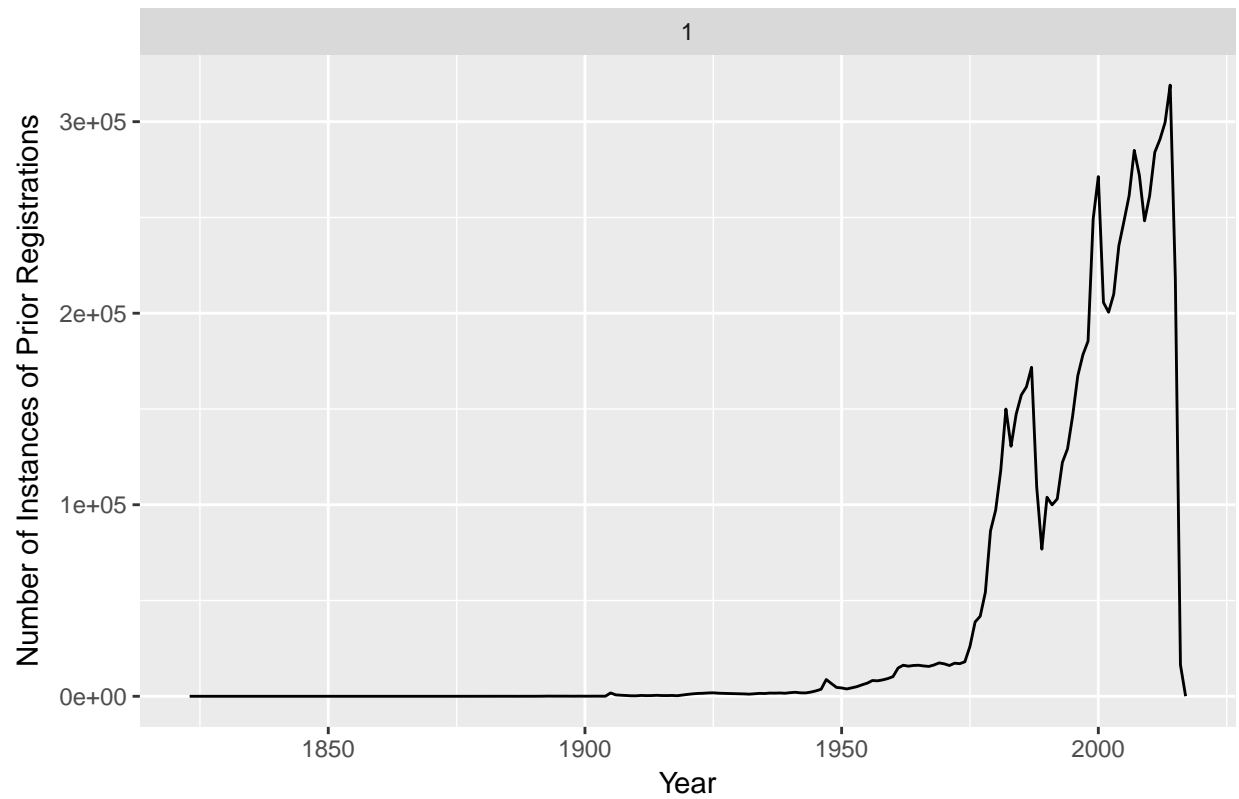
```
# Take out extreme number of instances (n = 300,000)
TrademarkFreq %>% filter(n < 100000) %>% ggplot() + geom_line(aes(x=Year, y=n, colour=prior_regs, group=prior_regs))
```

Line Graph Without Extreme Values of Number of Instances



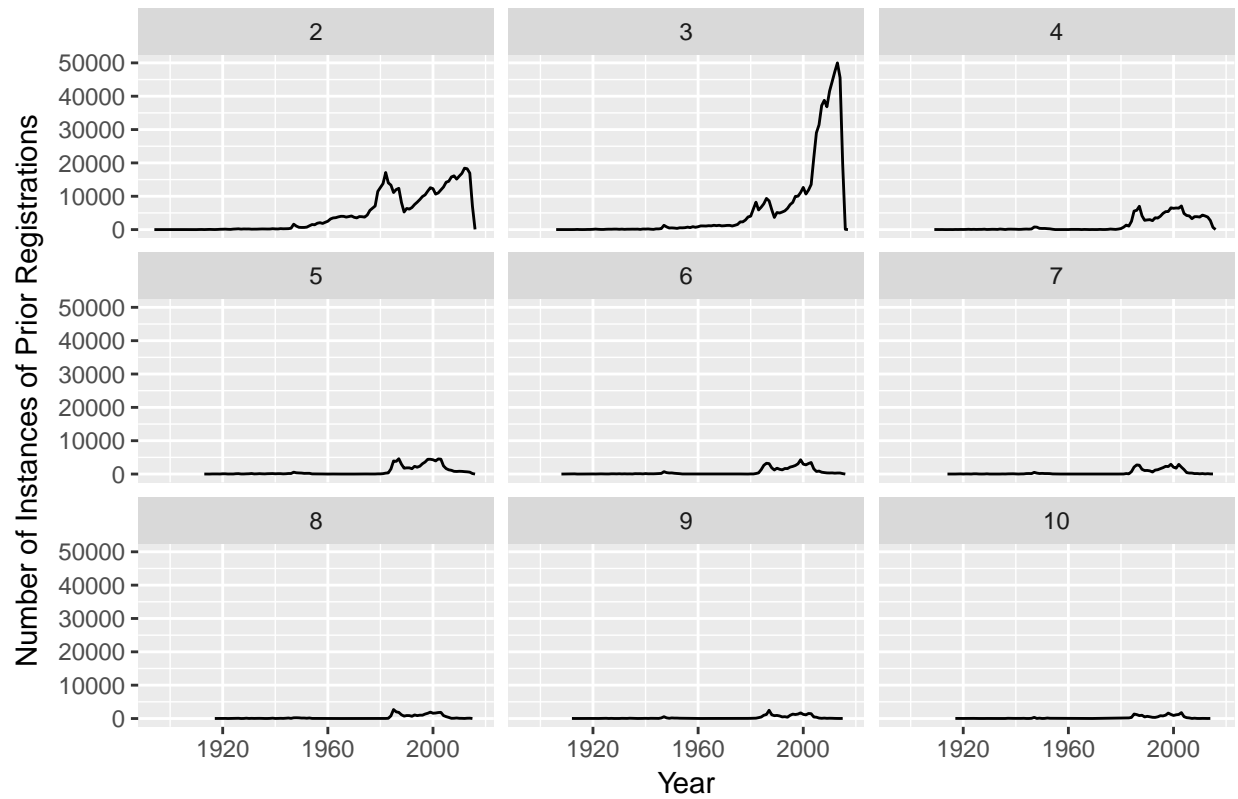
```
# Faceted for prior registrations = 1
TrademarkFreq %>% filter(prior_regs == 1) %>% ggplot() + geom_line(aes(x=Year, y=n, group=1)) + facet_w
```

Line Graph for Prior Registration = 1



```
# Faceted by 1 < prior registrations < 11
TrademarkFreq %>% filter(prior_regs > 1 & prior_regs < 11) %>% ggplot() + geom_line(aes(x=Year, y=n, gr
```

Line Graph for Prior Registrations Greater than 1 and Less than 10

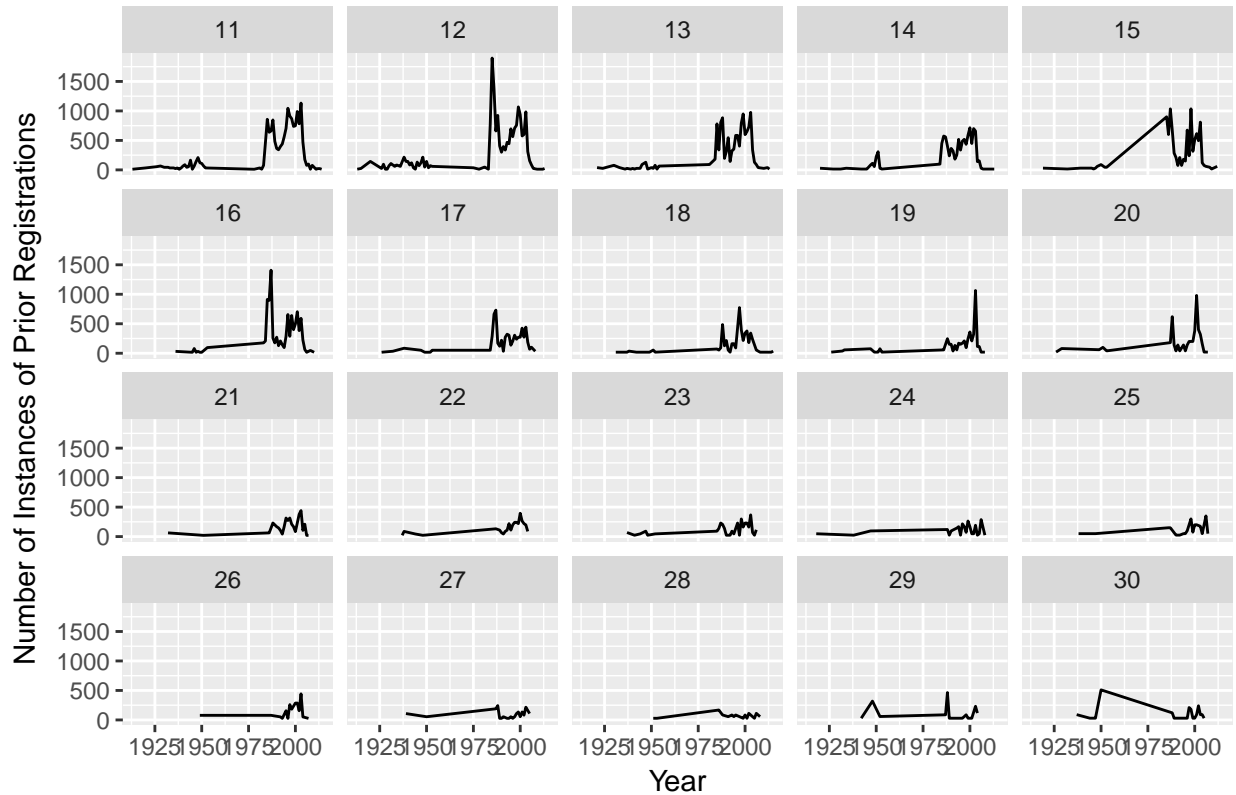


```
# Faceted by 10 < prior registrations < 31
```

```
TrademarkFreq %>% filter(prior_regs > 10 & prior_regs < 31) %>% ggplot() + geom_line(aes(x=Year, y=n, g
```



Line Graph for Prior Registrations Greater than 10 and less than 31



Unsurprisingly, these visualizations generally indicate that the number of trademarks with one prior registration has increased by a large magnitude, with smaller effects as the number of prior registrations grows. That being said, there is generally support for the idea that the number of trademarks with prior registrations has vastly increased over time, with a major spike since 1989 (roughly coinciding with the widespread adoption of technology like cable television, the Internet, etc.)

## Next Steps

These descriptive statistics lend support to the idea that the number of trademark re-brands has increased over time, with a huge increase since 1989. In terms of next steps, we should think about how to better conceptualize the effect of different media forms on the need to rebrand. Otherwise, we still need to fuzzy string match between the SEC list (~8000 observations) and the larger Trademarks dataset to get a better handle on large corporate rebranding.