

Secrets Behind the Boost of Companies*

Four main factors that contributes to the boost are found.

Zhiyue Yu

28 April 2022

Abstract

Businesses and companies are one of the main roles that maintain the daily operation of our society. To figure out the secrets behind the boost of companies, the mostly-cleaned data is collected from the data world, which contains the data set that describes entrepreneurship and company activities. By fitting a multi-linear regression model, we find that there are four factors that influence the growth of a company, including the states they are in, the industry they are in, the number of years the companies have been founded and the continuous growth. Our finding provides insights and principles for companies for further growth, including but not just limit to year 2019.

1 Introduction

Businesses and companies are one of the main roles that maintain the daily operation of our society. Regardless of the history and industry of the companies, the purpose of them is to solve social problems and improve people's life, it is true both for large companies and for start-ups. The larger a company grows, the bigger social problems it can solve.

In this paper, we aim to find out the factors that boost the growth of a company by applying data visualization and linear regression model based on the data we obtained. Usually, the capital a company owns and the number of employees it has have not much relation with the growth of the company, which most of the people thought they are. In the end of the paper, we concluded the four factors that lead to the boost of a company.

In Section 2, data are cleaned and a multi-linear regression model is fitted with the data. The mostly-cleaned data is collected from the data world("Historical Kauffman Index Microdata" (2022)), which contains the data set that describes entrepreneurship and company activities. The original data is from Inc.com("Inc5000" (2022)), which ranks the top 5000 fast growing companies in the US. All the other related company information can be found in the hidden links by clicking at the company names.

The result shows that there are a number of factors that will influence the growth of a company, including the States they are in, the industry they are in, the number of years the companies have been founded and its long-term continuous growth. For each aspect, a comprehensive analysis is conducted in Section 5. Our finding provides insights and principles for companies for further growth, including but not just limited to the year 2019. We believe that the specific reason for a successful company varies, but the logic and rules behind it are consistent.

*Code and data are available at: <https://github.com/gggeorgeeyu/304final>.

Table 1: Preview of the data

name	state	revenue	growth	industry	wnow	yf	yol	wpre
Freestar	AZ	36.9 Million	36680.39	Marketing	40	2015	1	5
FreightWise	TN	33.6 Million	30547.93	Transportation	39	2015	1	8
Cece’s Veggie Co.	TX	24.9 Million	23880.49	Food & Beverage	190	2015	1	10
LadyBoss	NM	32.4 Million	21849.89	To C products	57	2014	1	2
Perpay	PA	22.5 Million	18166.41	Retail	25	2014	1	6
Cano Health	FL	271.8 Million	14183.41	Health	742	2009	1	18

2 Data

2.1 Data background and the use of R

In the paper, we run our analysis in R (R Core Team 2020). To read in data, we used `readr` package written by (Wickham, Hester, and Bryan 2022). To clean data and visualize data, we used `tidyverse` package written by (Wickham et al. 2019), `ggplot2` written by (Wickham 2016) and `dplyr` written by (Wickham et al. 2022).

The data set we used for research and study is collected from the data world(“Historical Kauffman Index Microdata” (2022)), which contains the data set that describes entrepreneurship and company activities. Data world is the enterprise data catalog for the modern data stack, it is build for the next generation for data discovery and research. The original data is from Inc.com(“Inc5000” (2022)), which ranks the top 5000 fast growing companies in the US.

2.2 Data cleaning and overview

In table 1, it shows the top 6 rows of the whole table. We can see that it has 9 columns, listing different characteristics of the companies. We first have the name of the company, where the state the company locates, the total revenue it gained in 2019, its total growth calculated in percentage, the industry it is in. The variable “wnow” is the number of workers at present, and the variable “wpre” is the previous number of workers in the company. Variable “yf” is the year the company founded and “yol” is the total number of years the company has been on the list.

In order to find out the factors that influence the boost of a company, we removed unnecessary variables from the raw data in the data manipulation process, such as the url link to the company, the profile link, the metro and the city that company locates.

2.3 Distribution of the companies in states

To have an overview of the background of the data, we first want to see the distribution of the companies in terms of the states. By looking at the bar chart(Figure 1), we can see that the top 5 states that have the highest number of companies are state California(CA), Texas(TX), Florida(FL), New York(NY), Virginia(VA), in which California has approximately 710 companies, ranking the first among all.

2.4 Distribution of the companies in industries

In the bar chart(Figure 2), we can see that the top 5 industries where boosting companies lie in are Advertising & Marketing, Business products & services, Software, Constructions and Health, in which Business products & services ranks the number one, containing around 500 companies. From the ranking, we can tell that four out of five industries are fictitious economies, rather than real economies, which is a sign that products and services have more potential to boost and gain rapid growth in a short period of time.

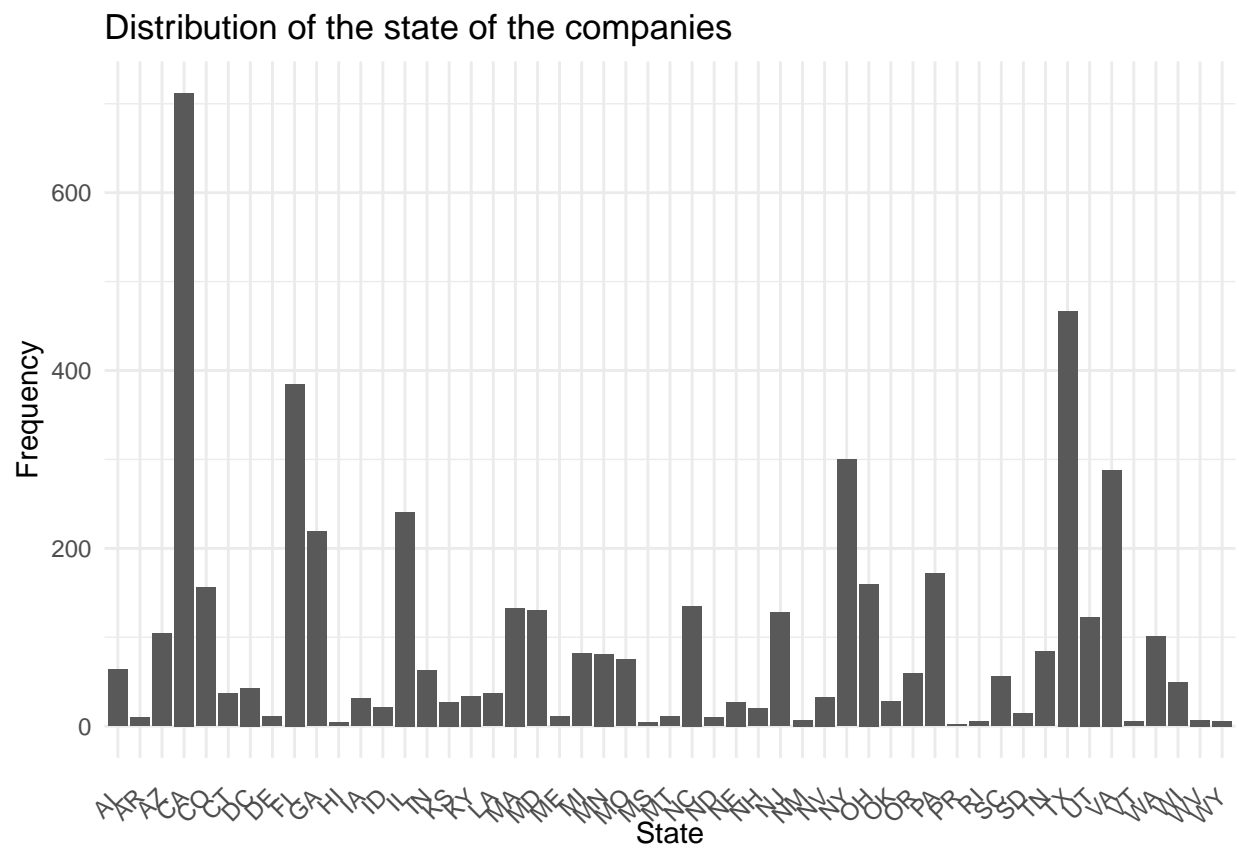


Figure 1: Distribution of the state of the companies

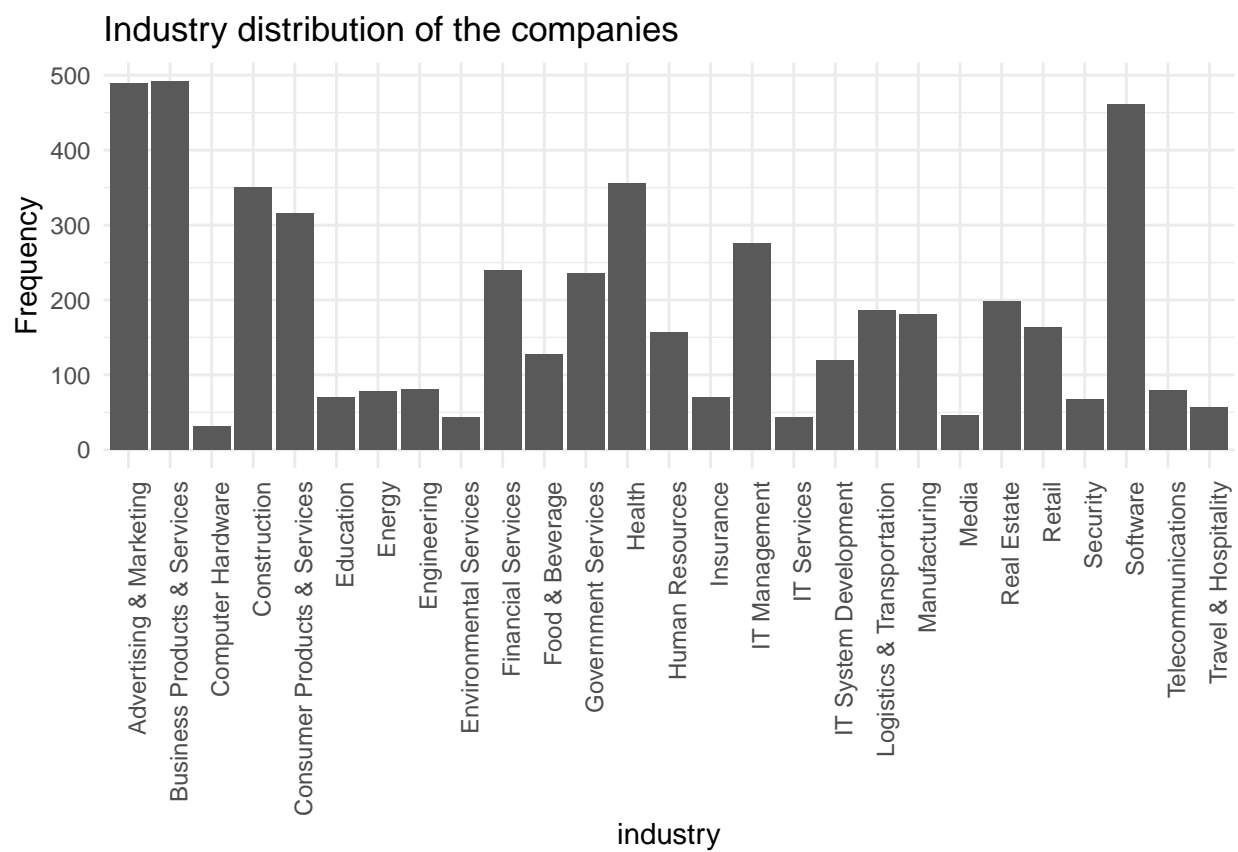


Figure 2: Industry distribution of the companies

2.5 Number of years the companies have founded

As shown in the bar chart (Figure 3), the trend of the occurrence of fast growing companies tends to be exponential over the 150 years from 1869 to 2016. With the development of science and technology, especially the invention of the internet after 1980, the growth boosts much faster than before, which was almost linearly in the past. In 2015, the number of companies ranked the highest among all, reaching around 475 companies.

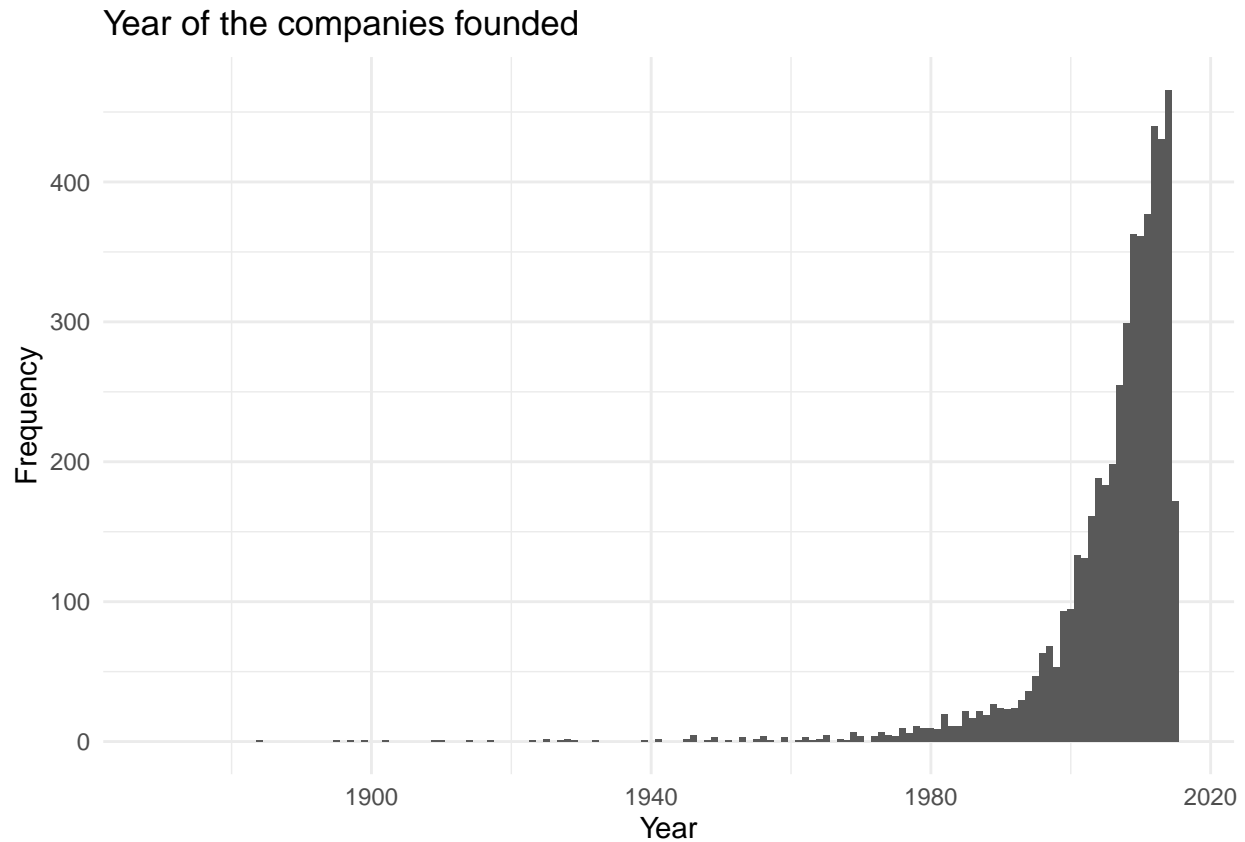


Figure 3: Year of the companies founded

2.6 Total number of years onlist

As shown in the bar chart (Figure 4), the total number of years of companies on-list decreases. Around 2000 companies have only been on the list for once, which may occur after 2015 – the time when a large number of fast growing companies showed up. However, there are still some great companies that have been on the list for 14 years, which is rare and amazing.

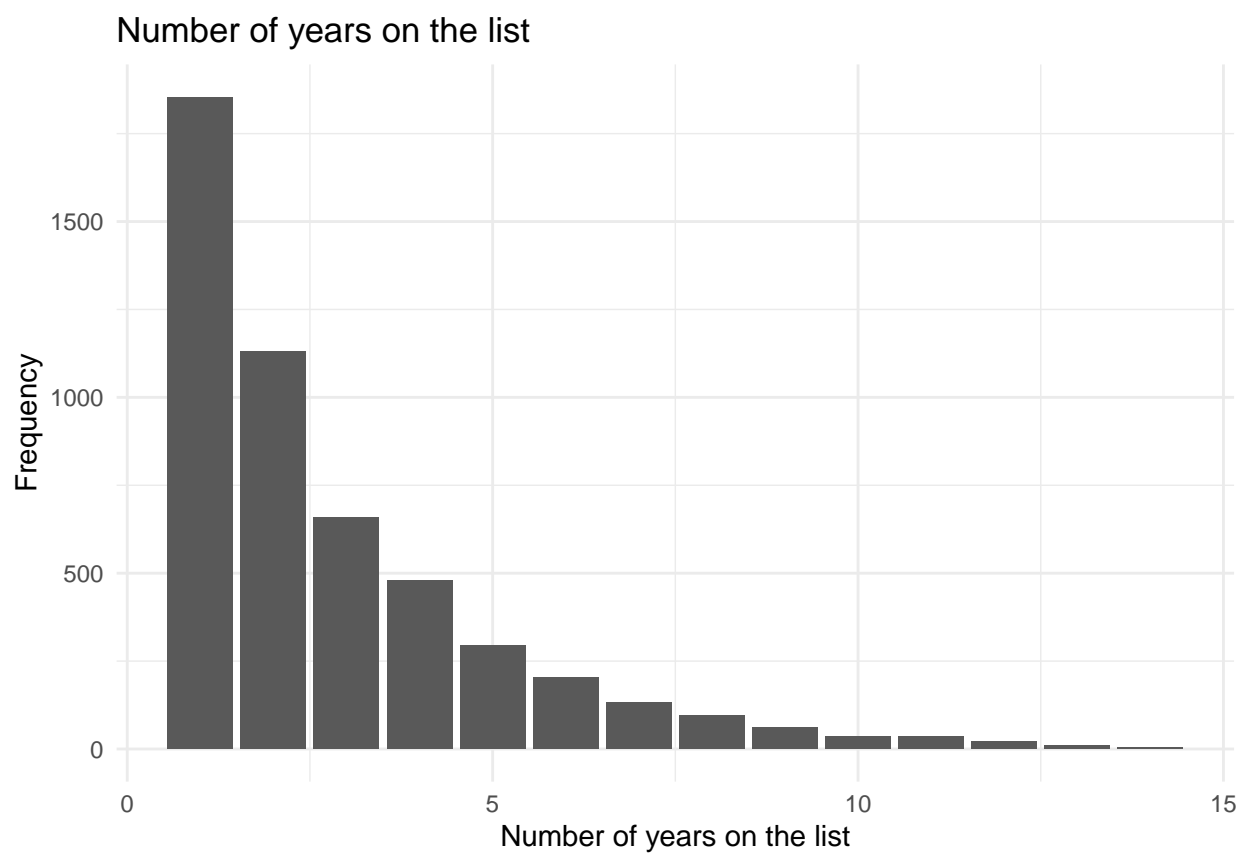


Figure 4: Number of years on the list

3 Model

We carried out the analysis using a multi-linear regression model. We treated the growth of the company calculated in percentage as the response variable. In terms of the predictors, they are the states the companies are in, the industry the companies are in, the current number of workers the companies employ, the number of years the companies have been found, the total number of years the companies have been on the list and the previous number of workers in each company. Therefore, the model deployed is as follows:

$$Y_{growth-in-percentage} = \beta_1 X_{state} + \beta_2 X_{industry} + \beta_3 X_{c-num-workers} + \beta_4 X_{years-founded} + \beta_5 X_{years-onlist} + \beta_6 X_{pre-num-workers} \quad (1)$$

where

$$Y_{growth-in-percentage}$$

is the growth of the company calculated in percentage,

$$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$$

are the coefficients to the corresponding predictors,

$$X_{state}$$

is the state the companies are in,

$$X_{industry}$$

is the industry companies are in,

$$X_{c-num-workers}$$

is the current number of workers the companies employ,

$$X_{years-founded}$$

is the number of years the companies have been found,

$$X_{years-onlist}$$

is the total number of years the companies have been on the list,

$$X_{pre-num-workers}$$

is the previous number of workers in each company.

4 Results

In this section, we will first examine the result of the model, and then have a little discussion of the graph about the number of years founded.

4.1 Result of the linear regression model

After fitting the multi-linear regression model, we find that only a few factors that have the impact on the boost of a company. These factors are state ID and NM, industry Consumer Products & Services, Logistics & Transportation, IT Services, the number of years the company has been founded and the total number of years the company on the list. To categorise, they are the states the companies are in, the industries they are in, the years companies have been founded and the total number of years on list.

As we can see in the table 2, stateNM, number of years founded and the total number of years on the list have very small p values, which are close to zero, meaning that we have very strong evidence to reject the null hypothesis that these factors are not related to the percentage of growth of a company.

To interpret these data, we need to look at the coefficient of each predictor. First, for those who have positive relations, for stateID, if we keep the rest of coefficients the same, every unit increase in the number of companies in state Idaho, the number of growth will increase by 775 per cent. For stateNM, if we keep the rest of coefficients zero, every unit increase in the number of companies in state New Mexico, the number of growth will increase by 2780 per cent. In terms of industries, if we keep the rest of the coefficients zero, when there is an increase of one company in Consumer Products & Services, the growth of the company will increase by 189 per cent. When there is an increase of one company in Logistics & Transportation, the growth of the company will increase by 256 per cent if we keep the rest of the coefficients zero. For the number of years founded, when it increases by 1, the growth of the company will increase by 15 per cent if we keep the rest of the coefficients zero. Now let's look at the predictors with negative relations, when keeping all the other coefficients zero, for industry IT service, when there is one more company in this industry, the growth will decrease by 398 per cent. When there is one more year on the list, the growth of the company will decrease by 76 percent.

Table 2 Coefficients of the model

Variable names	Estimate	Std. Error.	z value	Pr> z
(Intercept)	-2.949e+04	3.511e+03	-8.400	< 2e-16 ***
stateID	7.754e+02	3.171e+02	2.445	0.0145 *
stateNM	2.780e+03	4.996e+02	5.565	2.76e-08 ***
Consumer Products & Services	1.892e+02	9.143e+01	2.070	0.0385 *
Logistics & Transportation	2.557e+02	1.095e+02	2.335	0.0196 *
IT Services	-3.983e+02	2.002e+02	-1.990	0.0467 *
year_founded	1.503e+01	1.744e+00	8.617	< 2e-16 ***
num_of_years_onlist	-7.628e+01	8.139e+00	-9.372	< 2e-16 ***

4.2 Result of the graph of number of years founded

As we can see in the section 2, the graph of the number of years companies founded shows a clear exponential growth after 2000, and it reaches the highest in 2015. Not only the number of fast-growing companies boosts, the percentage in the growth rises dramatically. This is because of the invention and the widespread use of the internet, thus related industries rise afterward. The occurrence of the internet largely lowers the cost of starting a business and the cost of marketing. People do not have to prepare half of a million to open up an on-site store, they can start a virtual online store with only hundreds of bucks instead, this practice lowers the cost of making a mistake and continuous testing. This is the reason why the number of fast-growing companies rises so fast and they boost so fast.

5 Discussion

This section, we will first conclude this research, then discuss other possible factors that boost a company and the shortcoming of this research.

5.1 Research conclusion

In this paper, we collected data from the data world, aiming to find the factors that contribute to the boost of companies. To achieve this goal, we removed variables not related to the problem we want to research, then we fitted a multi-linear regression model to find out the factors related to the growth of the company by looking at the p-values. Finally, we figure out there are four factors that lead to the boost of the company based on the existing data. They are the states the companies locate, the industries the companies are in, the years the companies have been founded and the number of years they are on-list. The first two factors will vary depending on the specific state companies are in and the specific industries. For the third factor, we can infer that the longer a company exists, the better and more comprehensive operating system they will have, thus becoming a better company to grow fast. For the last factor, the total number of years on-list examines the stability and the continuous growing and optimising of a company. A large number of years on the list does not mean the company has a larger chance to be on the list in the next year, but it does show the company has a great business model and a steady operating system.

5.2 Other possible factors

Apart from the factors that we have investigated in the previous sections of the study, there are a great number of other factors that can lead to the boost of a company. We will discuss additional two in the following subsections:

5.2.1 Customer oriented value

Unlike the value of business in the past, nowadays we focus on the customers, instead of the products itself. This means we are focusing on the problem customers care about, rather than what I want to sell. In the past, the logic behind the business is: now I have a product I developed, how can I sell to the customers? The current logic of the business is: what are the problems customers are facing, so what is the product I can build to solve the problem? With the latter logic, products are much easier to find the product market fit, and thus sell better, which leads to a faster growth of the company.

5.2.2 Social responsibility

Social responsibility is tightly related to the vision of a company. The larger vision a company has, the larger problems they solve, thus the more social responsibility it has to take, because it means more promise and more reliability people will have on the company. For example, a technology company like Tesla will surely have more social responsibility than a random marketing agency company, because Tesla solves larger problems. According to the last subsection, the larger and more critical problems a company solves, the faster it will like to grow.

5.3 Weaknesses and next steps

Although the data world has provided some relative predictors and data for the fast growing companies, the information is still limited, so the result may not be accurate and comprehensive. If we want to look for a more precise conclusion for the reason behind the boost of companies, we need to investigate further for the questions we have discussed in the discussion part.

References

- “Historical Kauffman Index Microdata.” 2022. EWING MARION KAUFFMAN FOUNDATION. <https://www.kauffman.org/historical-kauffman-index/microdata/>.
- “Inc5000.” 2022. INC.COM. <https://www.inc.com/inc5000/2019>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*.