

SNU: Startup Newbies' User guideviz

Keondo Park*

Graduate School of Data Science
Seoul National Univ.

Woochul Lee †

Graduate School of Data Science
Seoul National Univ.

Wooseok Jang ‡

Graduate School of Data Science
Seoul National Univ.

Yoobin Baik§

Graduate School of Data Science
Seoul National Univ.

Youngjun Son ¶

Graduate School of Data Science
Seoul National Univ.

ABSTRACT

Startup industries are drawing a lot of interests across diverse groups of people - university students, employed workers who dream of starting their own businesses, venture capitalists who are looking for the next unicorn and general public who are always craving to listen to another success story. Information on startups, however, is scattered around and not presented in well-organized manner. We collected information from startup world and presented this information in user-friendly and efficient way with proper visualization. Our work includes visualizations showing funding amount of companies and number of employees by funding stage, hub of startups by industry, strength of certain universities and dominant majors in the startup world.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

1 INTRODUCTION

In recent periods, start-up is booming and more and more people are getting interested in start-ups. Some entrepreneurs start their own companies to solve the worlds' existing problems with their own solutions, and some venture capitalists are looking for the next unicorn start-ups such as Google, Amazon or Facebook. Also general public is fascinated by start-up founders' success story and craving for another success story. Some recently made TV series described the start-up founders' success stories and became one of the most popular TV series of the year. According to CrunchBase, 5,022 new start-ups were founded by 6,581 founders and \$294.8 billion was invested into start-up companies in 2019 across the globe. Start-up companies or industry, however, are rapidly changing and inherently not very well organized, because they have just started. Many prospective entrepreneurs, early stage founders or venture capitalists find it very hard to get relevant information regarding start-ups at a high level. In order to narrow down the gap between high level of interests and lack of well-organized information, we collected and processed relevant data, then visualized it in very efficient, user-friendly and intuitive manner through this project.

Our work is presented in this order: Related works are presented in the second section. Tasks and data abstractions are described in the third section. Implementation of our work is explained in the fourth section. Our conclusion is suggested in the fifth section.

*e-mail: gundo0102@snu.ac.kr

†e-mail: woochulee@snu.ac.kr

‡e-mail: wscwsc92@snu.ac.kr

§e-mail: banca321@snu.ac.kr

¶e-mail: jun041577@snu.ac.kr

1.1 Domain backgrounds

As in other areas, startup industry has its own words. Some of the most basic and widely used terminologies are listed below:

Angel Investors Investors who invest in the very early stage startups and provide them with advice.

Seed Investment The investments made before prototype development is completed. The major purpose objective of this investment is to ensure enough money for the operating expenses. Mainly invested by founders, founders' family or friends, or angel investors.

Series A Investment The investment made after prototype is developed and before large scale marketing is started.

Series B Investment The investment for expanding the market or hiring at a larger scale etc.

Series C/D/... Investment The investment for additional expansion.

IPO(Initial Public Offering) Companies list their shares on stock exchanges so that external investors publicly buy the companies' stock. It is required to disclose companies' management details. Founders or early investors can collect some of early investments by selling their shares to public.

M&A(Merge & Acquisition) Large companies may merge or acquire the share of start-up companies. As in IPO, Founders or early investors can collect some of early investments by selling their shares.

Unicorn start-ups a private company with a valuation over \$1 billion. As of December 2020, there are more than 500 unicorns around the world.

Venture Capital(VC) Venture capital is a form of private equity financing that is provided by venture capital firms or funds to startups that have been deemed to have high growth potential or which have demonstrated high growth.

2 RELATED WORK

Archie Tse [5] suggested 3 rules for visual storytelling: 1. If you make the reader click or do anything other than scroll, something spectacular has to happen. 2. If you make a tooltip or rollover, assume no one will ever see it. If content is important for readers to see, don't hide it. 3. When deciding whether to make something interactive, remember that getting it to work on all platforms is expensive.

TheVC [4] also provides useful information associated with startups in South Korea with visualizations. Line charts for total funding amount or number of employees, bubble charts with timeline to show the amount of capital injected per funding stages for each company are included in its visualizations.

Maureen Stone [3] suggested some principles of color design in visualization. Especially, how three dimensions of color - hue, value, and chroma - could be used for visualizations. It was emphasized high contrast layer and low contrast layer could be used to separate information and pastel colors are the best choice in most cases.

Some early pioneers of startup industry wrote their story and advice [2] [1]. They helped us to select and fine-tune visualization topics in which users are most interested.

3 TASKS AND DATA ABSTRACTIONS

3.1 Tasks

3.1.1 Users

Our main anticipated users are those who are interested in their own startups. This may include those who are planning on starting their own startup, those who have already started one, or those who are simply intrigued by this new culture.

3.1.2 Tasks

The main tasks of this project is to answer the following key questions:

Funding by Stages It is no surprise that one of the most important factor in business is money, so naturally funding becomes the fuel on which startup runs. It is important to make sure that the investment the company is receiving is not far from the standards. Through this key question, users will be able to have a better grasp on how much funding is expected for each stage, or even compare the total amount of their investment with the average of each industry.

Number of Employees by Stages Man power demanded may differ between industries. While some industries may require as many employees as possible at the start of the company, other industries feel the need to start hiring more employees at a certain point of time. For example, the size of most transformation industry's workforce proliferates after Series C, while the workforce of healthcare maintains a small size until IPO. Users may use this as a guideline to navigate the size of human resources.

Hub of Start-ups Networking is one of the most important virtues emphasized by many successful startups, and no place is better for networking than hub of startups. With better understanding of the hub of start-ups, users will be able to organize a list of locations for their new headquarter.

Founding Member It is important to choose the best candidate for the founding member of the company. One may become a huge asset or may become a liability. When evaluating the potential of a candidate, the university they have graduated from and the major they have studied may be influential. An insight on the strength of certain universities and the dominant majors in the startup world may help.

3.2 Data

3.2.1 Dataset and Data Collection

Data for the CB rank top 1000 companies were collected from the website of Crunchbase(<https://www.crunchbase.com>) using Python crawlers. CB rank stands for Crunchbase rank, which reflects the crunchbase's own ranking algorithm. The processed data for each company is saved in csv format and includes 6 columns as in (Table 1). The educational background data for individuals are collected in csv format as well.

We also collected information of listed companies from their annual report. For example, total assets, the size of listed companies' workforce or revenue could be found, processed and used in the visualizations.

Table 1: Columns for processed data

Column name	
1	Name
2	Industries
3	Headquarter location
4	Last funding type
5	Number of employees
6	Total asset and Total Funding Amount

Table 2: Industry classification

Industry	
1	Software
2	Financial Services
3	E Commerce
4	Health Care
5	AI
6	Security
7	Manufacturing
8	Education
9	Transportation
10	Food
11	Others

3.2.2 Data Processing

Data Grouping There are 492 industrial groups in the raw data. It is too subdivided and hard to understand. It is also not ideal number in terms of visualization. Therefore, we grouped them into 11 industrial groups. (Table 2) Similarly, we reclassified the major of founders into 5 groups. (Table 3)

Data interpolation There are NaN values in the total asset data and the missing values are filled using linear interpolation.

Currency unit All currencies other than US dollars are converted into US dollars.

4 IMPLEMENTATION

4.1 Design Rationale

Scrollytelling Our visualizations are organized in scrollytelling way. With sheer mouse scroll, users can easily follow our analysis in the form of visualizations with supplemental sentences and interact with the visuals where they are interested.

Color code per industry Throughout all visualizations, we assigned different color to different industry where the visualization includes industry segmentation. This is expected to help users to find the industry of industry quickly and intuitively.

Funding by Stages Funding by Stages consists of two axes. First, each funding stage is shown on X axis. It contains each funding stages from Seed and Series A rounds to Exit rounds like M&A or IPO. Second, each funding amount is shown on Y axis in 1M US dollars. Lastly, funding amounts of each industry is

Table 3: Major classification

Major	
1	Engineering
2	Economics
3	Business
4	Natural Sciencee
5	Etc

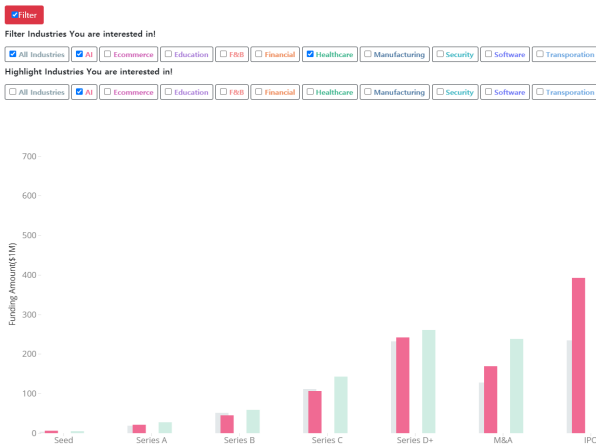


Figure 1: Funding Amount of each industries

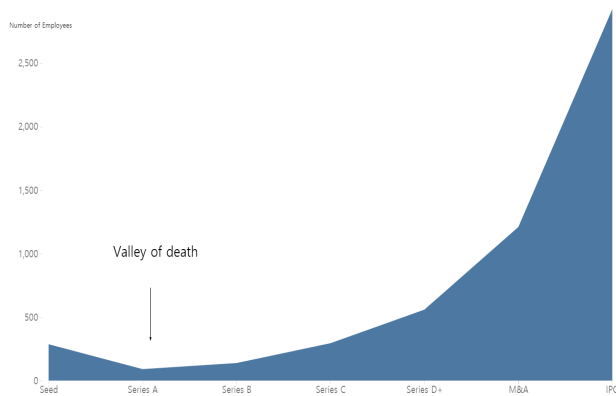


Figure 2: Area chart for all industries

differentiated by 10 different colors with slightly low opacity to compare each value efficiently.

Furthermore, user can utilize a filter option as well. When users click a filter button, user can filter and highlight industries in interest. When it comes to highlight option, it increases opacity of selected bar to make the funding amount come in sight. The visualization is presented in Figure 1.

Growth of startup companies The number of employees is a natural indicator of the companies' growth. In this visual, the size of workforce of startup companies is presented by funding stage. Funding stage is an ordinal attribute and the number of employees is a quantitative attribute, so area chart is used for this visual. Moreover, funding stage is placed at x-axis in the order.

The growth is presented for all industries combined in the first visual. Scrolling down, the growth of companies in each industry is presented as a separate area chart. Buttons for each industry is located above the visual so that the user can select the industry in which the user is interested. As in other visualizations, color code for each industry is used. The visualization is presented in Figure 2 Figure3.

Strength of certain Universities Strength of certain universities contains number of startup members pertaining to each univer-



Figure 3: Area chart for individual industry

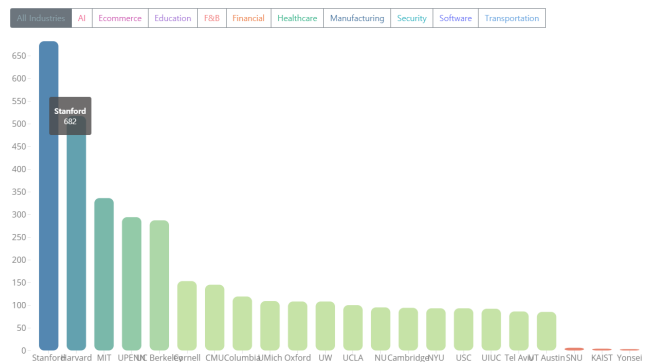


Figure 4: Bar chart for dominant Universities

sity. Each bar colored with blue and green like colors. Color gradation shows how many people in each university belong to each industry to make user understand the trend easily. There are three design elements in the visualization. First, it has mouser over tool-tip which can show the exact number of people belong to the university. Second, user can click and filter each industry with industries bar above. Lastly, the bar moves dynamically with time duration option whenever the button is clicked. The visualization is presented in Figure 4.

Dominant Majors in the Startup World Dominant Majors in the Startup World shows the ratio of each major by industries with pie chart. User can click and filter each industry to find optimal results. When users click each industry, pie chart moves and changes dynamically to show the result with highlight. Since there are a lot of major names, we gathered and re-grouped each major in about 5 major groups to minimize color confusion. The visualization is presented in Figure 5.

Hub of Startups The Hub of Startups visualization indicates the location of hubs for each industry. First we presented the location of hubs on the map of the world. Though Silicon Valley is known as the global center of startups, it might not be true for some industries. It can be easily checked through the first visualization; users are able to observe locations of hubs for each industry all over the world. From the first visualization, we are able to note that most of the largest hubs are located in the United States. From this observation, we implemented another visualization to provide better perspective, manifest-

All Industries AI Ecommerce Education F&B Financial Healthcare Manufacturing Security Software Transportation

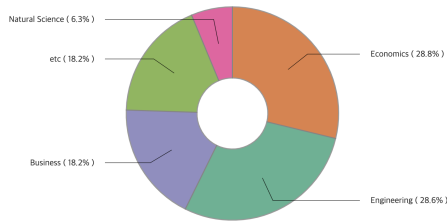


Figure 5: Pie chart for dominant major

All Industries AI Ecommerce Education F&B Financial Healthcare Manufacturing Security Software Transportation

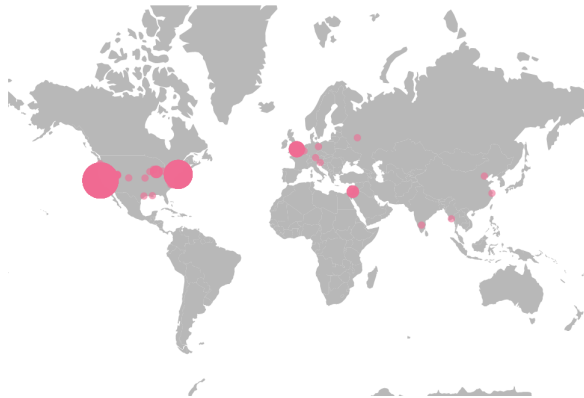


Figure 6: Global Hub of Startups

ing the cities of each hub and the number of startups. The visualization is presented in Figure 6 and Figure7.

4.2 System Overview

We implemented our visualization project mainly using D3, a JavaScript library for producing dynamic, interactive data visualizations in web browsers. In addition, a total of eight visualizations(including word cloud) were created using various tools such as geopy, jquery, topojson, bootstrap, layout.cloud and more. Our project is open-sourced and can be found on <https://github.com/KeondoPark/viz/tree/mergedBranch>.

5 CONCLUSION

In this paper we developed an interactive visualization web page for people who have interest in startup field. We gathered more than 3000 startup data to build a database and used several libraries including d3.js to generate 7 main visualizations. With our visualizations, users can find useful clues about funding amounts by investment rounds, number of employees, dominant universities, dominant majors, and hub of startups by industries. By using our visualization, even people who were not familiar with startups are able to find clues for his/her own venture company or investment funding for venture companies, especially in any industry of interest.

ACKNOWLEDGMENTS

This paper is a part of the term project of the 'Data Analysis and Visualization' class in SNU(Seoul National University). We wish to thank Professor Hwajung Hong and other classmates for giving high quality feedback.

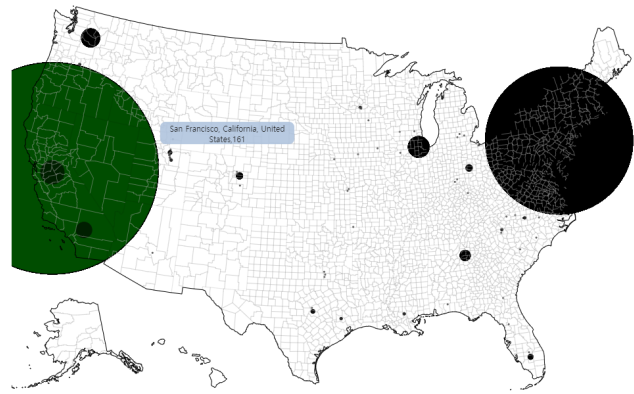


Figure 7: US Hub of Startups

REFERENCES

- [1] S. K. Hwanjin Choi. *Start startup wisely*. Ji And Sun, 2015.
- [2] D. Kwon. *Startup management class by Dokun Kwon*. Wisdom house, 2015.
- [3] M. Stone. Choosing colors for data visualization. 2006.
- [4] TheVC. <https://thevc.kr/>.
- [5] A. Tse. Why we are doing fewer interactives. 2016.