

Data Visualization of Chinese Internet Companies

cha-net-vis.yifangbao.me

Bao Yifang

check the digital interactive datavis version on the website

<http://cha-net-vis.yifangbao.me>

O.Abstract

(in Italian)

Le aziende web cinesi sono l'argomento principale della tesi. Attraverso il processo di raccolta e elaborazione dei dati, sono stati ottenuti cinque set di dati finali. Si tratta del set di dati di Alexa tra il 2012 e il 2019 da Wikipedia, il set di dati delle 55 società quotate in borsa di Google Stock e Xueqiu, il set di dati dell'unicorno cinese di ITjuzi, il set di dati della registrazione cinese tra il 1980 e il 2019, il set di dati del collegamento Tencent e la Topologia Alibaba da Wikipedia.

I dati sono stati quindi visualizzati utilizzando diversi modelli visuali al fine di permettere una loro analisi visuale. In questo modo la situazione delle società web cinesi è stata esplorata dal punto di vista globale, storico ed economico.

(in English)

This thesis aims at representing the Chinese Internet companies. Through the process of data collecting and processing, five final datasets were obtained. They were the dataset of Alexa ranking between 2012 and 2019 from Wikipedia, the dataset of Chinese 55 Stock Listed Companies from Google Stock and Xueqiu, the dataset of Chinese unicorns list from ITjuzi, the dataset of Chinese company registration between 1980 and 2019, the dataset of Tencent and Alibaba Topology article on Wikipedia.

The dataset were then visualised using different visual models allowing their visual analysis. These visualisations are meant to explore the current situation of Chinese Internet companies from the global, historical, and economical perspective.

1. Background and Introduction

Today, the Internet has become the primary driver of global economic growth with continuous development in hardware and software facilities, in which Chinese Internet economy had developed rapidly during the recent 30 years. From 1994, the year when China accessed to the Global Internet in all functions, until today, it has been over 20 years. Endless applications and services penetrated every corner of people's lives and work. In recent years, there existed several symbolic titles connected to the Internet, such as the O2O, the VR, the sharing economy, the blockchain, the live, and artificial intelligence [1]. In 2016, the total number of China's digital economy with the Internet as the main component reached 22.6 trillion yuan, surpassing the sum of Japan and the United Kingdom, ranking second in the world. There were various outstanding Internet companies from China, such as Tencent and Alibaba.

The Internet industry would be of its characteristics, differing from the traditional sector. For example, due to the rapid development of Internet technology, the molecular industry needed to be subverted into and adapt to the innovation field at a glimpse of an eye, if not, the company might bankrupt. Besides, in the early developing periods of Internet companies, customers were usually offered free product service. Thus, the debt ratio of the Internet company was generally high, and the losses in profits were a general phenomenon [2]. Therefore, the financing ability was crucial for Internet companies to survive at the initial development period. The Chinese Internet capital market also presented different characteristics at various stages of Internet development, the year of 1994-1998 was the initial period, the year of 1999-2003 was the development period, the year of 2004-2008 was the further growth period, the year of 2009 to 2013 was the mature period, and the year of 2014 to 2018 was the perfecting period [3]. Therefore, it would be fascinating to explore the financing, vertical fields, geographical locations, founders, and other topics of Chinese Internet industry companies.

The aim of the analysis for the Chinese Internet topic was as followed. First, to explore the development of the Chinese Internet industry from a historical perspective. Second, to understand the founders' background and discover the secrets of the individual triumph. Third, to help the entrepreneur students to understand more about the Chinese Internet market.

Fourth, to promote the practitioners in the Internet industry understanding of the Internet industry from another perspective view. Fifth, to assist the financial investors to know the company situations somehow. Sixth, to help foreigners acknowledge the development of China's Internet industry. Besides all the above, Chinese Internet company, as an instance topic, was a practice for my personal data visualization skills. In the thesis, the entire process and design of data visualization would be recorded, as a reference for other design solutions.

In the following sections, we would explore the topics related to Chinese Internet companies from five levels through data visualization, as the Protocol.01 shown. The first section was combined with the dataset of the Alexa ranking to explore the global situations of Internet sites. Section two was based on the dataset of the Chinese Internet Companies, which were listed on Stock Market, to uncover the secrets about the listed companies and the founders. Section three was focused on the dataset of the Chinese Internet unicorns listed on ITjuzi to explore the current status of the unicorns in China. Section four was connected to the dataset of the Chinese Internet historical registration on Github to explore the overall historical states of Chinese Internet companies. The last section was through the Wikipedia network topology data about Alibaba Group and Tencent, the two largest Internet companies in China, to analyzes the reasons behind the enterprise success. Each of these sections included the introduction, the

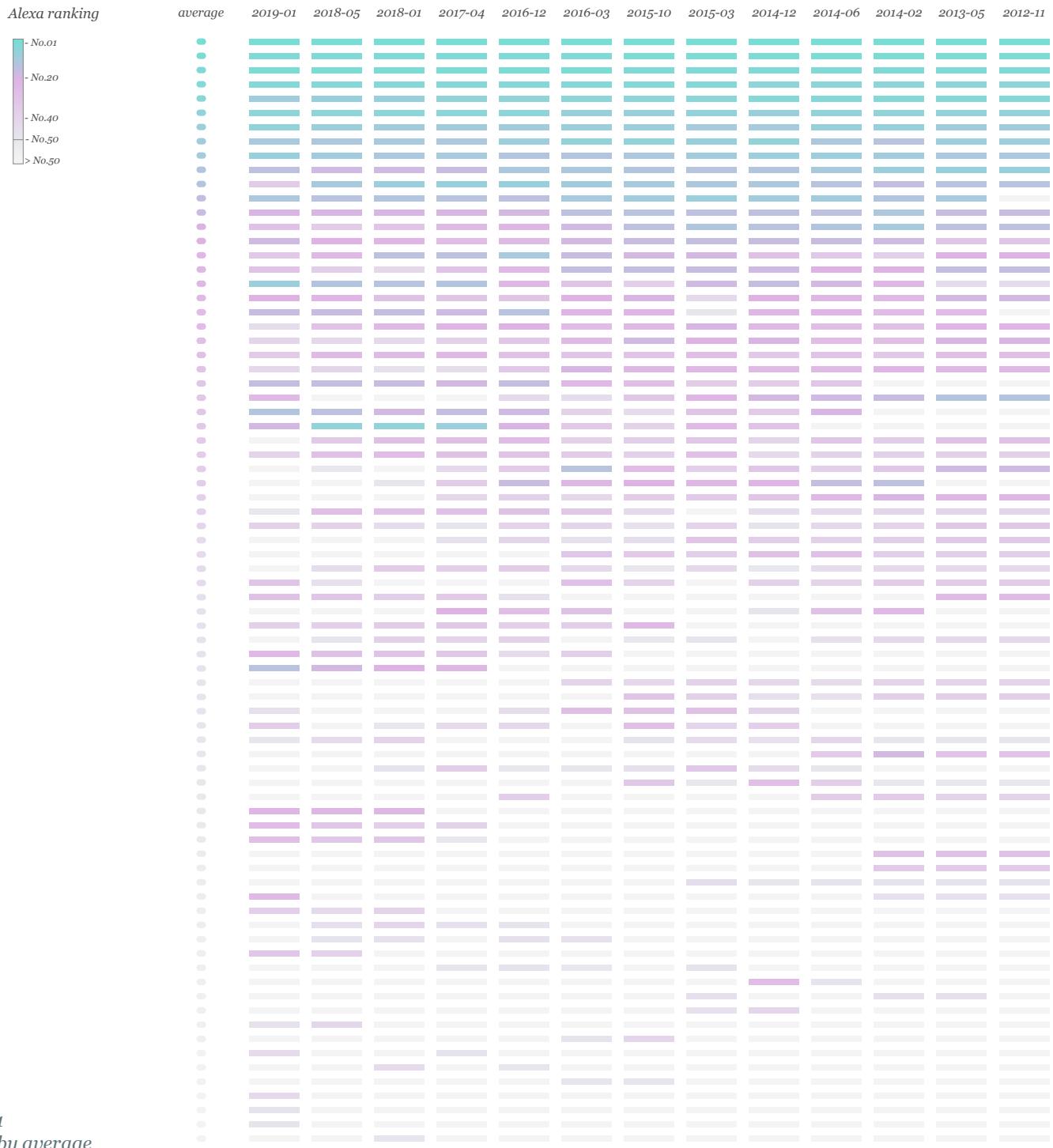
flowchart description of the data visualization process, the description of the data visualization design, the display of the data visualization output, and the insights found through the data visualization. In each workflow diagram, there were four steps of the starting point(query), data collection, data processing, and data visualization.

- 01 How is the Internet companies developed globally for recent ten years? [Alexa Ranking / List of most popular websites](#)
[en.wikipedia.org](#)
- 
- vis.
Site Ranking
vis.
Site Country
with interactive version
- 02 What is the status of Chinese Internet companies, which have been listed on the stock market?
[Chinese Internet Companies , Stock Market](#) / [中国互联网企业，上市公司](#)
[google.com & xueqiu.com](#)
- 
- vis.
Stock Market Type
vis.
Type and Address
vis.
Stock Situation
vis.
Founders
with interactive version
- 03 How about the situation of the Chinese Unicorn Internet Companies?
[Chinese Internet Companies , Unicorn](#) / [中国互联网公司，独角兽](#)
[Itjuzi.com](#)
- 
- vis.
Unicorn Value
vis.
Unicorn Size
with interactive version
- 04 How about the historical registration of Internet companies in China?
[China Internet enterprise registration data](#)
[Github](#)
- 
- vis.
company map
vis.
wordpart
with interactive version
- 05 How about the China's leading Internet companies of Tencent and Alibaba Group?
[Tencent & Alibaba](#)
[en.wikipedia.org](#)
- 
- vis.
Tencent network
vis.
Alibaba network

2. Project Statement

2.1 Data Visualization of Alexa Ranking

To explore the topic of Chinese Internet companies, we should put it into a global view initially. So the first question was ‘How is the Internet companies developed globally for recent ten years?’ To explore the subject could help us to understand the role of China in the Internet world better. Overall, in this section, combined with the dataset of the Alexa ranking, the design logic and the process for interactive visualization of one matrix and one flow chart would be illustrated to explore the current state of the global Internet sites from a historical perspective.



Visualization.01

site

google.com
 facebook.com
 youtube.com
 baidu.com
 yahoo.com
 wikipedia.org
 qq.com
 amazon.com
 taobao.com
 live.com
 google.co.in
 twitter.com
 sina.com.cn
 linkedin.com
 weibo.com
 google.co.jp
 yahoo.co.jp
 tmall.com
 yandex.ru
 vk.com
 google.de
 ebay.com
 google.co.uk
 bing.com
 instagram.com
 blogspot.com
 sohu.com
 reddit.com
 google.fr
 google.com.br
 msn.com
 hao123.com
 wordpress.com
 google.ru
 microsoft.com
 tumblr.com
 pinterest.com
 google.it
 mail.ru
 google.com.hk
 360.cn
 t.co
 google.es
 netflix.com
 jd.com
 paypal.com
 apple.com
 amazon.co.jp
 aliexpress.com
 xvideos.com
 163.com
 imgur.com
 ask.com
 soso.com
 360.cn
 pornhub.com
 twitch.tv
 babylon.com
 googleusercontent.com
 imbd.com
 blogger.com
 alipay.com
 google.ca
 ok.ru
 csdn.net
 stackoverflow.com
 adcash.com
 fc2.com
 alibaba.com
 office.com
 onclickads.net
 livejasmin.com
 google.com.mx
 gmw.cn
 porn555.com
 imbd.com
 Naver.com
 bongacams.com

country

US

China

Russia

Japan

Canada

India

Germany

UK

France

Brazil

Italy

Spain

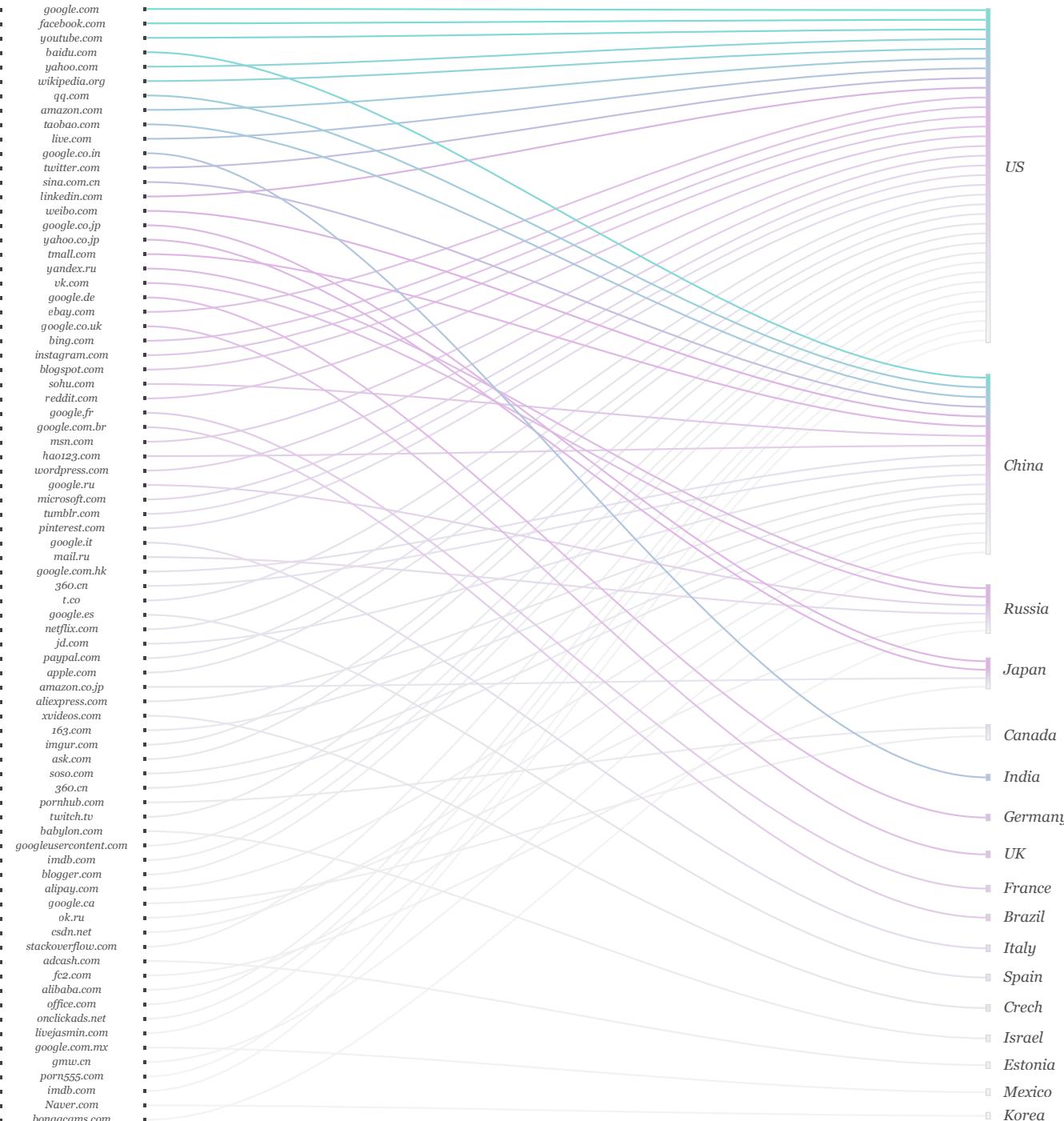
Czech

Israel

Estonia

Mexico

Korea

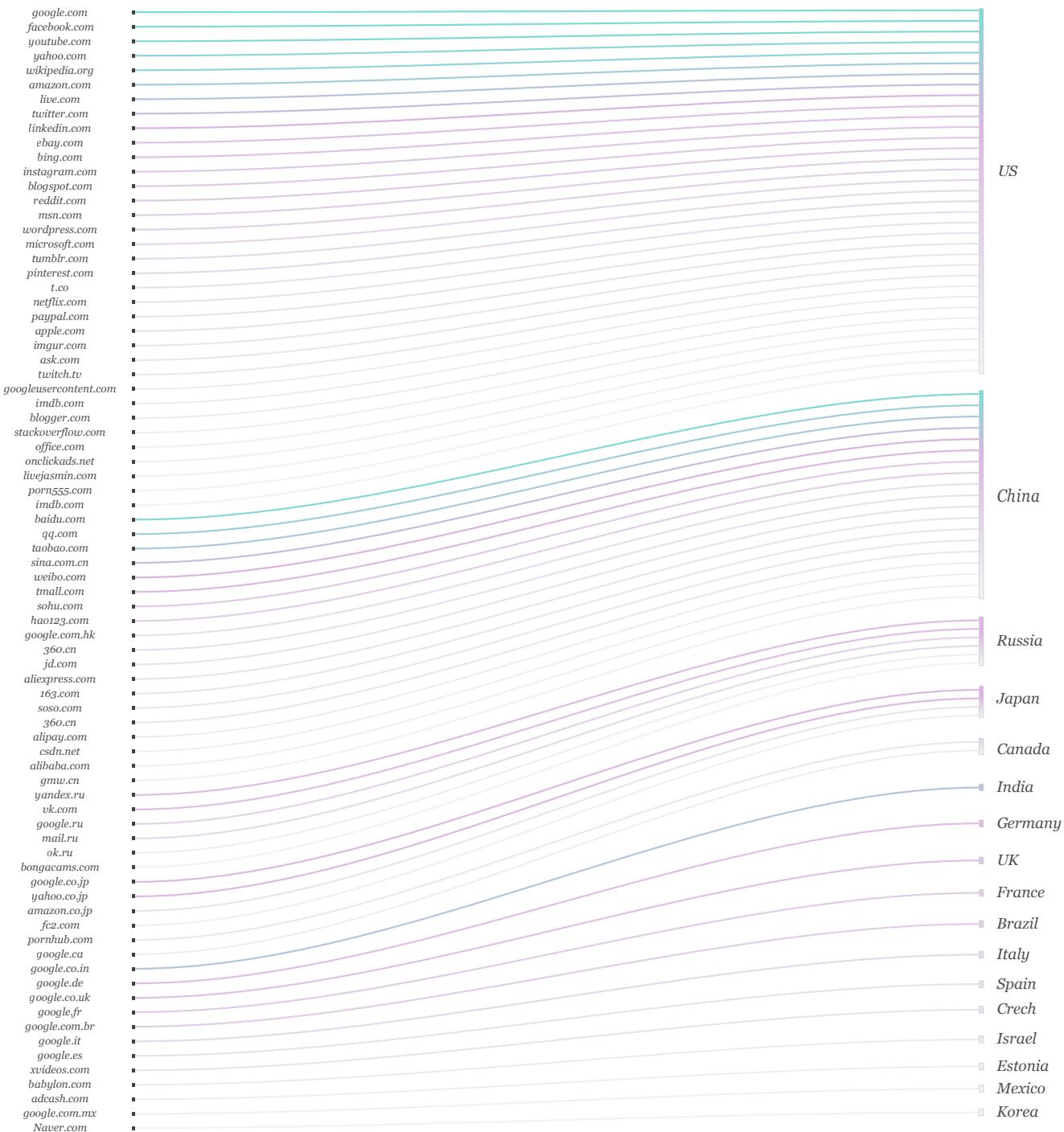




Visualization.01
Alexa ranking by country

site

country



How is the Internet companies developed globally for recent ten years?

Alexa Ranking / List of most popular websites

en.wikipedia.org

retain the
top 50 popular websites and the rankings
during 2012-2019

manually collect the websites' ranking
data in the lists

file2012.xlsx

file2013.xlsx

.....

file2018.xlsx

file2019.xlsx



each file includes the below element
> site:google.com
> country:u.s.
> type:internet service
> rank:1

input the excel files into
[Tableau Prep Builder](#)

to replace the missing data by number
60 to calculate the average rankings

fileall.csv



the file includes the below element
> site:google.com
> country:u.s.
> type:internet service
> ranking2019:1
> ranking2018:1
.....
> ranking2012:1
> average ranking:1

[Data Illustrator](#)

[Raw Graphs](#)

[Adobe Illustrator](#)

[Adobe Illustrator](#)

vis.

vis.

Site Ranking

Site Country



the visualization includes two models
> model 1: ranking by average
> model 2: ranking by country

vis.combine



Interactive Visualization

To explore the topic of Chinese Internet companies, we should put it into a global view initially. So the first question was ‘How is the Internet companies developed globally for recent ten years?’ To explore the subject could help us to understand the role of China in the Internet world better. Overall, in this section, combined with the dataset of the Alexa ranking, the design logic and the process for interactive visualization of one matrix and one flow chart would be illustrated to explore the current state of the global Internet sites from a historical perspective.

For this recipe, as the Protocol.02 shown, through visiting the Wikipedia editing history, the data of Alexa Internet ranking (Lists of most popular websites)[4]was manually collected, which filtered the top 50 websites in each duration from 2012 to 2019, including the elements of the site, the site country, the site type, and the ranking. Then, through Tableau Prep Builder(Tool), one integrated single CSV file (site, country, type, 2019-01 rank, 201805 rank, 2018-01 rank, 2017-04 rank, 2016-12 rank, 2016-03 rank, 2015-10rank, 2015-03 rank, 2014-12 rank, 2014-06 rank, 2014-02 rank, 2013-05 rank, 2012-11 rank) was obtained. Because only the top 50 ranking websites were collected in each duration, there existed some missing data, which means the ranking of these websites exceeded the top 50. For rough estimation for the average ranking, the missing data were replaced by number 60 to calculate.

After data collecting and processing, the next step was to do data visualization design. The final dataset file was imported into Data Illustrator(Tool) to generate a matrix diagram, which was then manually adjusted in Adobe Illustrator(Tool) to get the visualization of site ranking. Besides, the dataset was also imported into Raw Graphs(Tool) to generate a flow diagram, which was then manually adjusted in Adobe Illustrator(Tool) to get the visualization of site and Country. Then, these two visualization were integrated as a whole of Visualization.01 - Alexa ranking. Finally, the SVG files were exported and combined with the D3.js Library(Tool)[5] to get the final interactive visualization on the website.

Besides, the design logic of the visualization would be explained in details here, including the four design aspects: variables, properties, the layout, and the interaction. According to Lev Manovich, ‘Infovis uses graphical primitives such as points, straight lines, curves, and simple geometric shapes to stand for objects and relations between them.’ [6]. Here, the point (rectangle) in Visualization.01 and the line(curve) in Visualization.02 were chosen to represent each website. ‘Infovis uses spatial variables (position, size, shape, and more recently curvature of lines and movement) to represent key differences in the data and reveal most important patterns and relations. Other less important properties of the objects are represented through different visual dimensions – tones, shading, patterns, colors, or transparency of the graphical

elements.' [6] As for this project, the logical design for variables and properties was shown in the Table.01.

<i>variables</i>	<i>properties</i>	<i>value</i>	<i>meaning</i>
point	shape	 rect	each website
point	x position		discrete timeline from 2012 to 2019
point	color		ranking
line	shape		each website
line	color		average ranking
line	position of the ending point		the website country
line	position of the starting point		the website domain

Table.01

The X position of the rectangle illustrated the relative discrete timeline(2019-01 rank, 201805-rank, 2018-01 rank, 2017-04 rank, 2016-12 rank, 2016-03 rank, 2015-10 rank, 2015-03 rank, 2014-12 rank, 2014-06 rank, 2014-02 rank, 2013-05 rank, 2012-11 rank). The gradient color of the rectangle and the line represented the ranking (#76ded5 - ranking No.1, #dfb7e5-ranking, No.20, #e5dbeb-ranking, No.40, #efeff1-ranking >Top 50). The position of curve starting point related to the site, and the position of curve ending point related to the site country. Thus, the corresponding country of the world could be connected to each Internet site through the flow.

As for the layout, the left of the canvas was the visualization about the site ranking, the right of the canvas was the visualization about the site belonging country, and the middle of the canvas was the supplemental text information.

As for the interaction, the audience could hover on each reacts to see the tooltip information in the relevant position, including the elements of ranking, site, country, and description. Besides, clicking the button could adjust the ranking by country or average ranking, shown in the Figure.01.

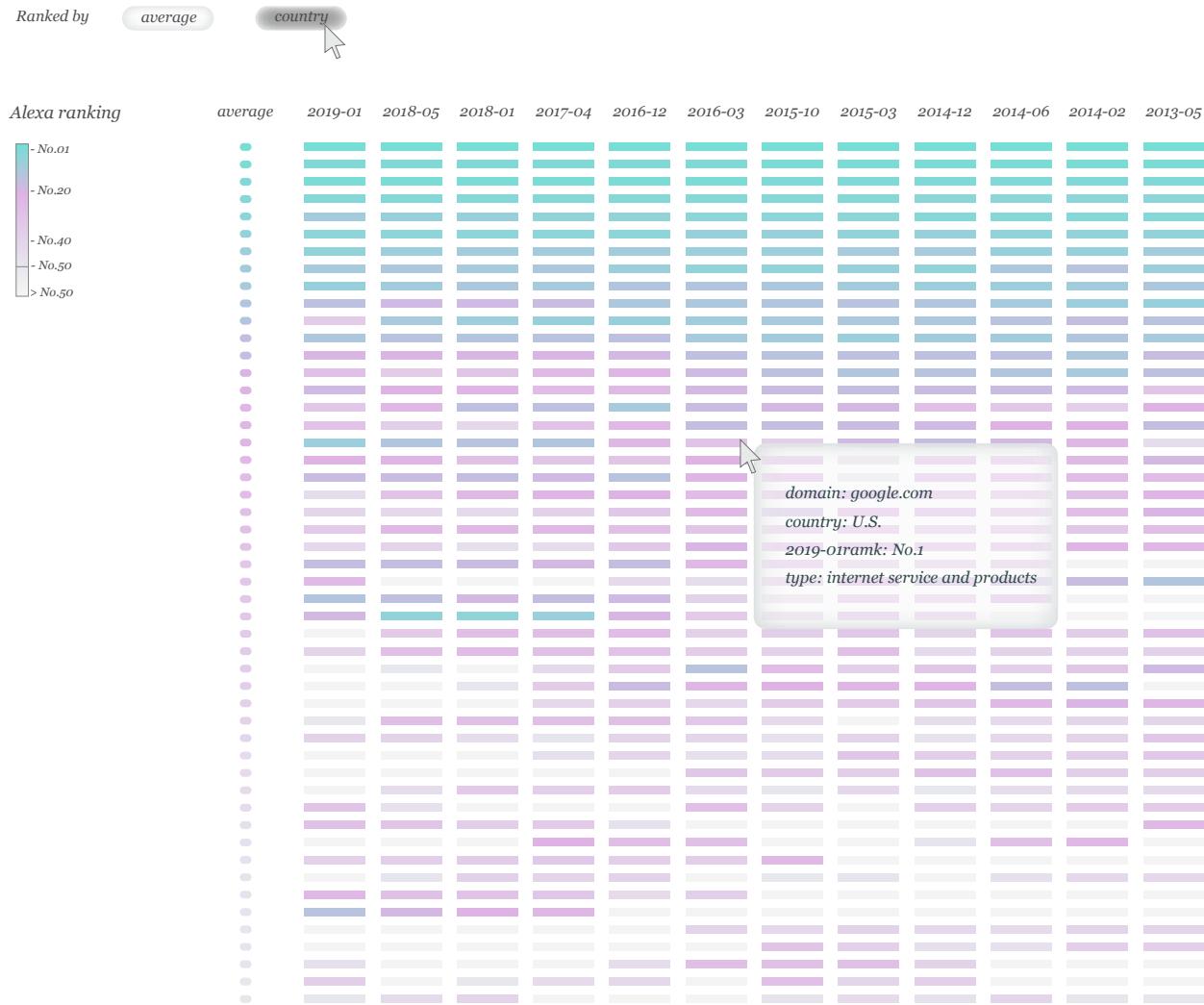
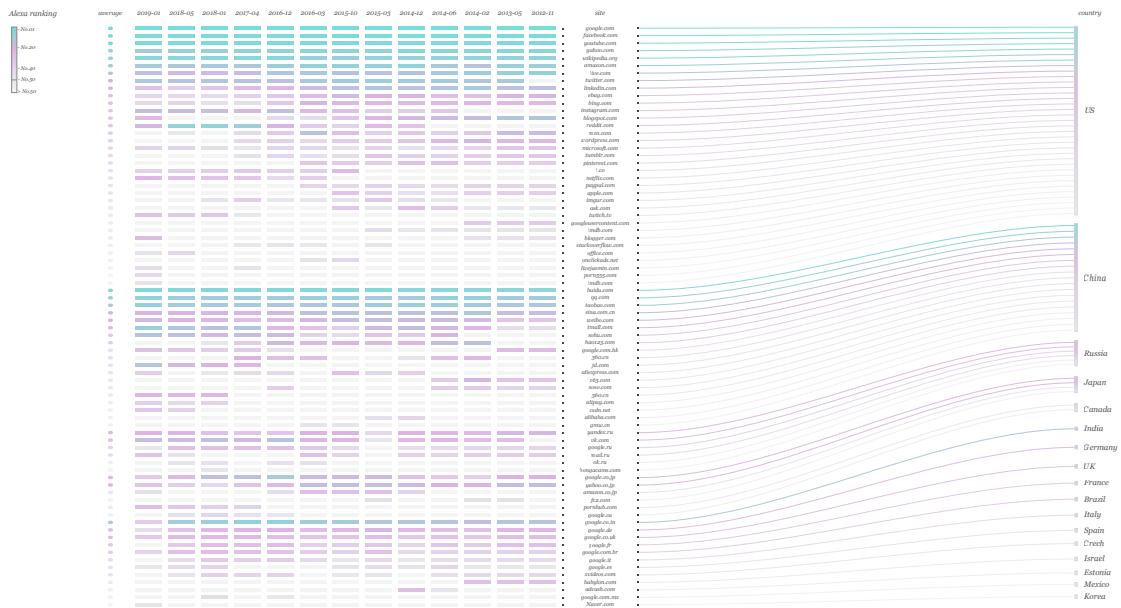
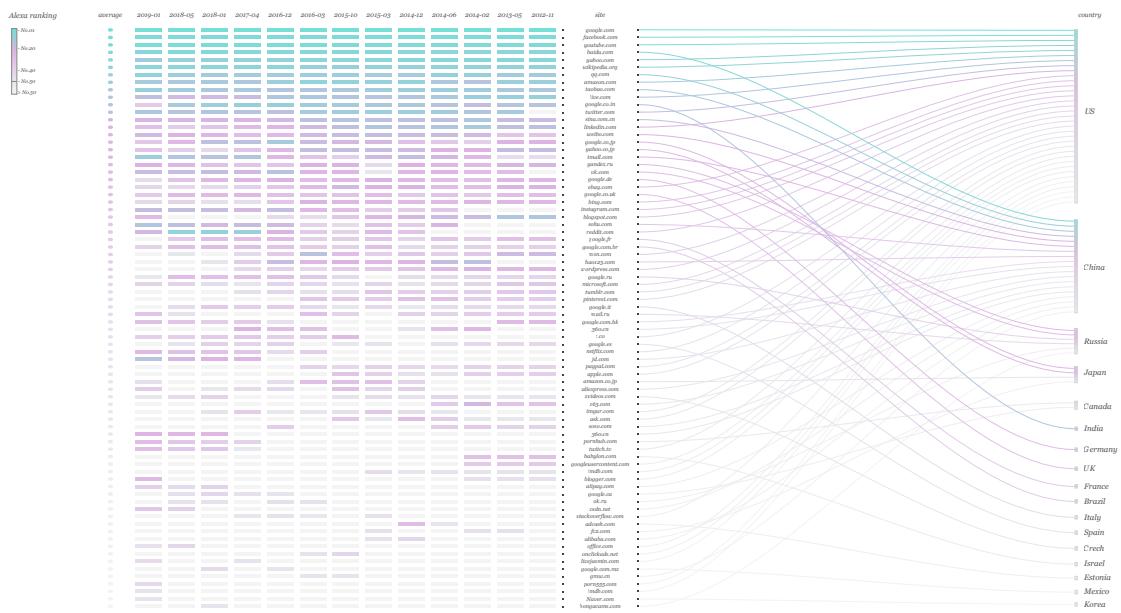


Figure.01

According to Edward Tufte, 'Clarity and excellence in thinking are very much like clarity and excellence in the display of data. When principles of design replicate principles of thought, the act of arranging information becomes an act of insight.'[7] So what insights could be drawn from this visualization?

Overall, the high-traffic website was concentrated on the United States and China, followed by Russia and Japan. However, the U.S. was still ahead of China in terms of the top website quantity and ranking. Europe, the Middle East, Canada, Australia, and other regions showed a weaker Internet developing state from a global view. Besides, the companies with the highest traffic (such as Google, Facebook, Youtube) maintained the leading status, and the companies with the central traffic were relatively volatile, which indicated a form of monopoly in the Internet world.

In details, there were 15 domain sites based on Google, such as google.de and google.fr. Google did occupy an important position in the Internet world. Also, the traffic data of several websites in the porn industry was growing. According to the classification of country, despite that the top websites in China and the U.S. were in a stable condition, causing a huge traffic pressure, there were other websites constantly could find an opportunity to develop and grow.



2.2 Data Visualization of Chinese Internet Companies listed on Stock Market

After understanding the state of global Internet companies, the situation of Chinese Internet companies would be further explored. Since that company is listed on the stock market was a measure of business success, to explore the status of Chinese Internet listed companies and the background of their founders would make sense. So the first question in this section was ‘What is the status of Chinese Internet companies, which have been listed on the stock market?’, moreover, the second question in this section was ‘How about the founders of Chinese Internet companies, which have been listed on the stock market?’ Overall, in this section, combined with the dataset of the Chinese Internet Companies listed on Stock Market, the design logic and the process for interactive visualization of circle canvas would be illustrated to explore the current state of the listed companies and the founders.

What is the status of Chinese Internet companies, which have been listed on the stock market?

Chinese Internet Companies , Stock Market

中国互联网企业，上市公司

google.com.hk

retain the list for
55 names of Chinese Internet Companies
which have been listed on stock market

[file-name.xlsx](#)

the file includes the below element
> company name

search each company's name in

google.com.hk google.com.en en.wikipedia.org ch.wikipedia.org

obatin the relative company's
information

[file-basic.xlsx](#)

each file includes the below element
> company name (en / ch)
> stock code
> stock market
> headquarter
> company keyword
> company foundation date

search stock code in

google.com.hk google.com.en xueqiu.com

[file-stock1.xlsx](#)

each file includes the below element
> max stock price
> max stock date
> min stock price
> min stock date
> issue stock price
> issue stock date
> the highest total stock value
> the highest total stock date
> revenue in 2018(billion CNY)

input the excel files into
[Excel](#)

to caculate the price ratio of
min/issue and max/issue

[file-stock2.xlsx](#)

each file includes the below element
> max stock price
> max stock date
> min stock price
> min stock date
> issue stock price
> issue stock date
> the highest total stock value
> the highest total stock date
> revenue in 2018(billion CNY)
> min/issue price
> max/issue price
> company foundation date

obatin the relative company founder's
information

[file-founders1.xlsx](#)

each file includes the below element
> company name (en / ch)
> founder's name
> birthday
> education background

Raw Graphs

Adobe Illustrator

vis.

Stock Market Type



vis.

Company Type and Address



Adobe Illustrator

Charticulator

vis.

Stock Situation



the visualization
includes various
models

> model 1 min
> model 2 max
> model 3 issue
> model 4 founded
> model 5 growth
> model 6 revenue

Interactive Visualization



Data Visualization

For this recipe, as the protocol.o3 shown, searching ‘Chinese Internet companies on the stock market’ on google.com.hk would obtain the list of 55 Chinese Internet Companies which have been listed on the stock market. Further, through searching the company names on the list on google.com.hk, google.com.en, en.wikipedia.org, and ch.wikipedia.org, the basic dataset file [company name (en/ch), founder, stock code, stock market, headquarter, company keyword, and company foundation date] were obtained. Then company types are manually classified by company keywords.

2.2.1 Visualization for stock market, type, and address

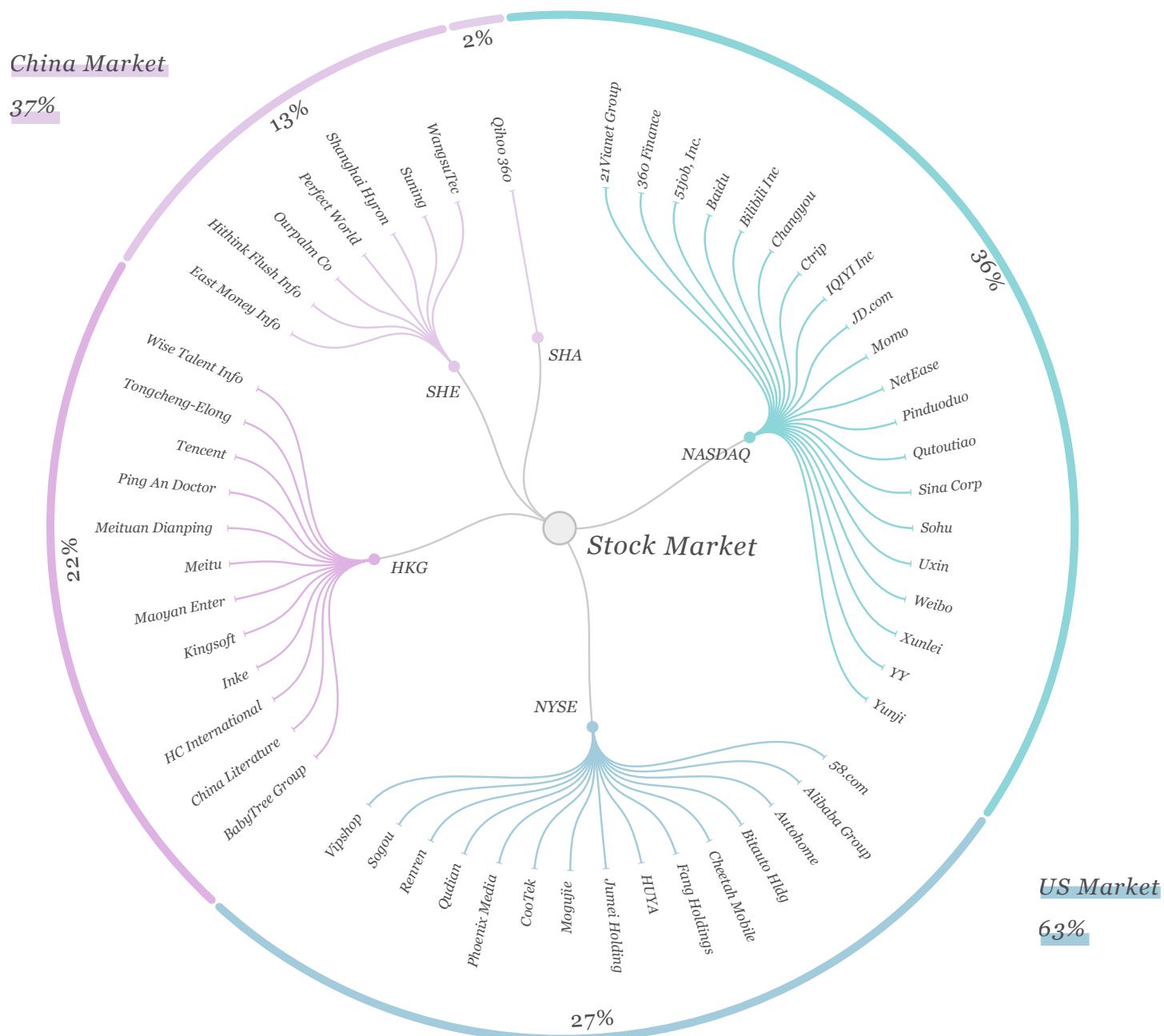
Here, analyzing this basic dataset file could shape some simple data visualization as a starter. First, importing the basic data file into Raw Graphs(Tool) would obtain a Dendrograms(Visualization o2)about the company stock market type, which was then manually adjusted in Adobe Illustrator. Besides, the dataset was also imported into Raw Graphs to generate a flow diagram, and manually adjusted in Adobe Illustrator to shape the diagram(Visualization o3) of the relationship between company type and address.

In the Visualization o2, the color of pink purple represented the China Stock Exchange Market (HKG, SHE, SHA), and the color of blue-green stated the U.S. Stock Exchange Market(NASDAQ, NYSE). The outer ring pie chart presented the percentage of each stock exchange.

Besides, each line connected each company name with the stock exchange market in details, and the text could help to describe and supplement information more clearly. Seen from the illustration, The listed exchanges selected by these companies included NASDAQ, NYSE, HKG, SHE, and SHA.

According to the analysis, 63% of companies chose to register in the US, 37% selected to list in China, and 22% of companies listed in China decided to list in Hong Kong. That is to say, only 15% of companies listed on the mainland of China.

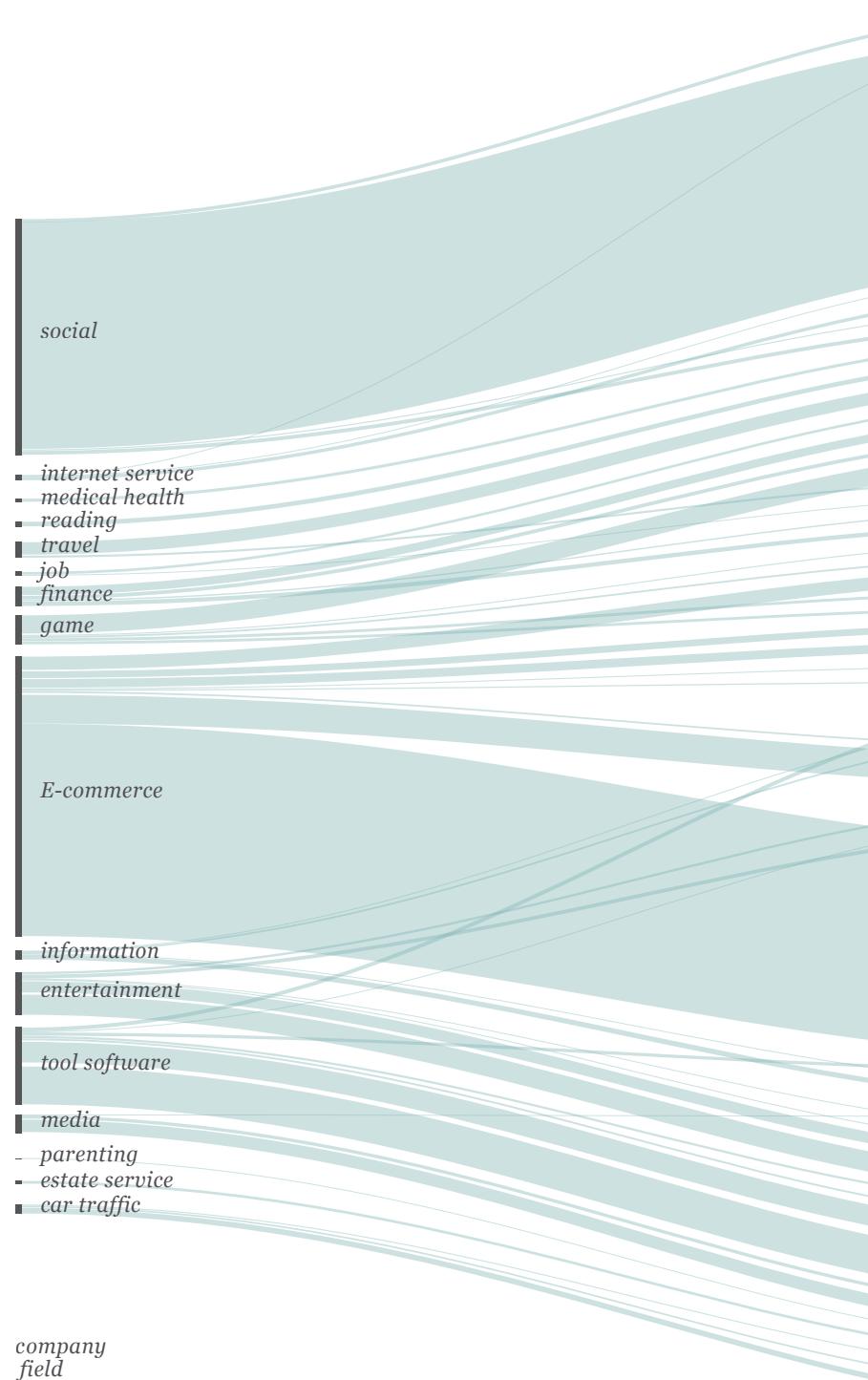
The leading companies, such as Alibaba, Tencent, and Baidu, chose to go public in the U.S. or Hong Kong. The situation depended on various reasons. Firstly, the U.S or H.K stock listing can increase the company reputation. Secondly, Chinese mainland stock listing requirements would be higher than U.S. and H.K, and the U.S stock market risk assessment was lower than the Chinese market.

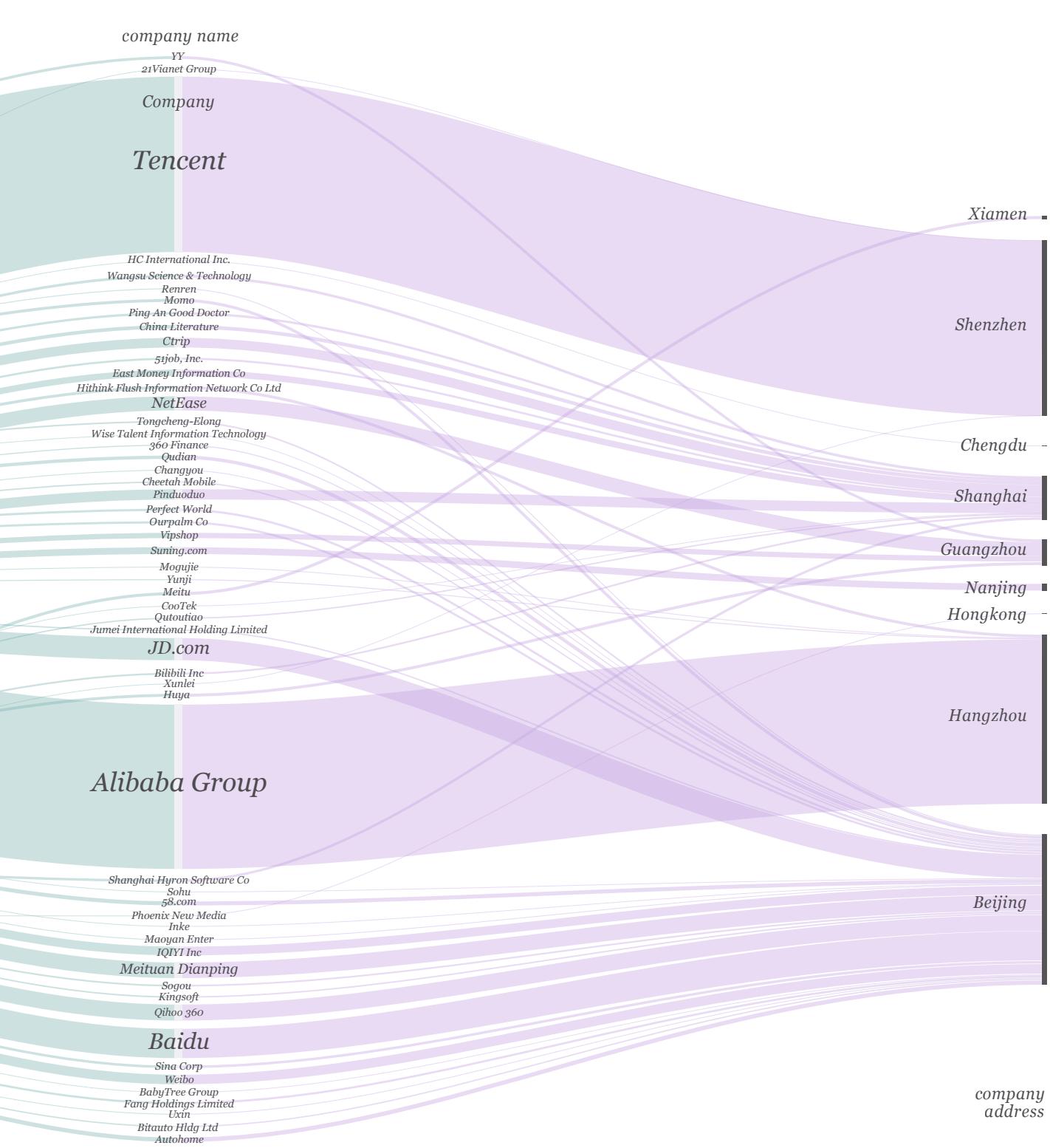


Visualization.03

Relationship between company and address

size: max total stock value





In the Visualization 03, the left side was the master field of the companies, the right side was the company headquarters, and in the middle was shown the company's name by text. Each line represented one company, and the weight of the line showed the max total stock value of the company.

Seen from the illustration, in terms of total stock value, the listed companies based in Shenzhen, Hangzhou, and Beijing ranking before Shanghai, Guangzhou, and Nanjing. In terms of quantity, the number of listed companies based in Beijing was the largest, followed by Shanghai and Guangzhou. Besides, there were many small companies listed in Beijing, and the accumulated value was equivalent to the individual companies of Alibaba or Tencent. Thus, Alibaba and Tencent were two giants Internet company in China. Alibaba occupied the vast majority of Hangzhou's market capitalization. Tencent held the vast majority of Shenzhen's market capitalization. In terms of company field, the total market capitalization from high to low was ranked as electronic commerce, social, tool software, entertainment, and games. In addition to Alibaba and Tencent, other companies such as Baidu, JD, Meituan Dianping, NetEase, Ctrip, Qihoo 360, Weibo, Pinduoduo, and IQIYI were accounted for a relatively high stock market value, since these companies were the leading Internet companies in the relative vertical industry.

2.2.2 Visualization for stock price

After that, to further explore the development of listed companies, the stock price fluctuations would be a better choice. So searching for stock code on Google and Xueqiu, the data in four key points (foundation, issue, max, min) about the date and the stock price was obtained manually. In Excel, through the min price divided the issue price and max price divided by the issue price, the proportion value comparison could reflect the stock benefits of various companies during these years. Thus, the final dataset of stock (company name, company foundation date, max stock price, max stock date, minimum stock price, minimum stock price date, issue stock price, issue stock date, the highest total stock value, the highest total stock date, revenue in 2018(billion CNY), min price/issue price, max price/issue price) was obtained. Importing the dataset into Charticulator and after that adjusting in Adobe Illustrator would get the Visualization.o4 about the company stock situation.

The design logic of the variables and properties would be explained in the Table o2, in which the point (circle) and the line (straight) were chosen to represent each company. The position along the radius of the circle illustrated the continuous timeline dimension from 1990 to 2020. The angle position of the circle showed the discrete company name ranked by alphabet. The

gradient color of the circle in gray blue-green represented the ratio of min price and issue price, and the gradient color of the circle in gray-pink purple represented the ratio of max price and issue price. The purple circle ring represented the issue price. The size of the circles meant the single stock price. The line with a purple color said how many years were spent from foundation date to the issue date for the company. The line with arrows stated four kinds of company growth type: keep increasing, keep decreasing, first increasing then decreasing, and first decreasing and increasing, The gradient color of the circle with borders simplified the Company revenues in 2018 (billion CNY), the darker, the higher. The size of the circle with borders means the max total stock value of the company (billion Dollars).

As for the layout, the traditional horizontal timeline was converted into the center ring timeline, which could concentrate a large amount of data into the center position of canvas to avoid the extra space consuming. Besides, combined with the web interaction, there were different forms to show the visualization for different analytical models including the model of min price, the model of max price, the model of the issue price, the model of foundation, the model of company growth, and the model of revenue and total value. Users can switch between different combinations by clicking the button.

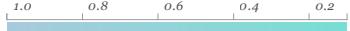
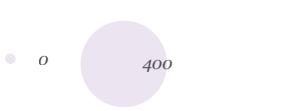
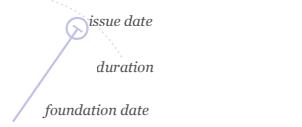
<i>variables</i>	<i>properties</i>	<i>value and meaning</i>
point ●	shape	<i>the issue price of the company being listed on stock</i>
point ●	color	 <i>max price / issue price</i>
point ●	color	 <i>min price / issue price</i>
point ●	color	 <i>company revenue in 2018 (billion CNY)</i>
point ● ● ●	size	 <i>single stock price</i>
point ●	size	 <i>max stock total value</i>
line ━━	length	 <i>the years between the company foundation date and issue date</i>
line ←→	direction	 <i>Company Growth Condition</i>

Table.02

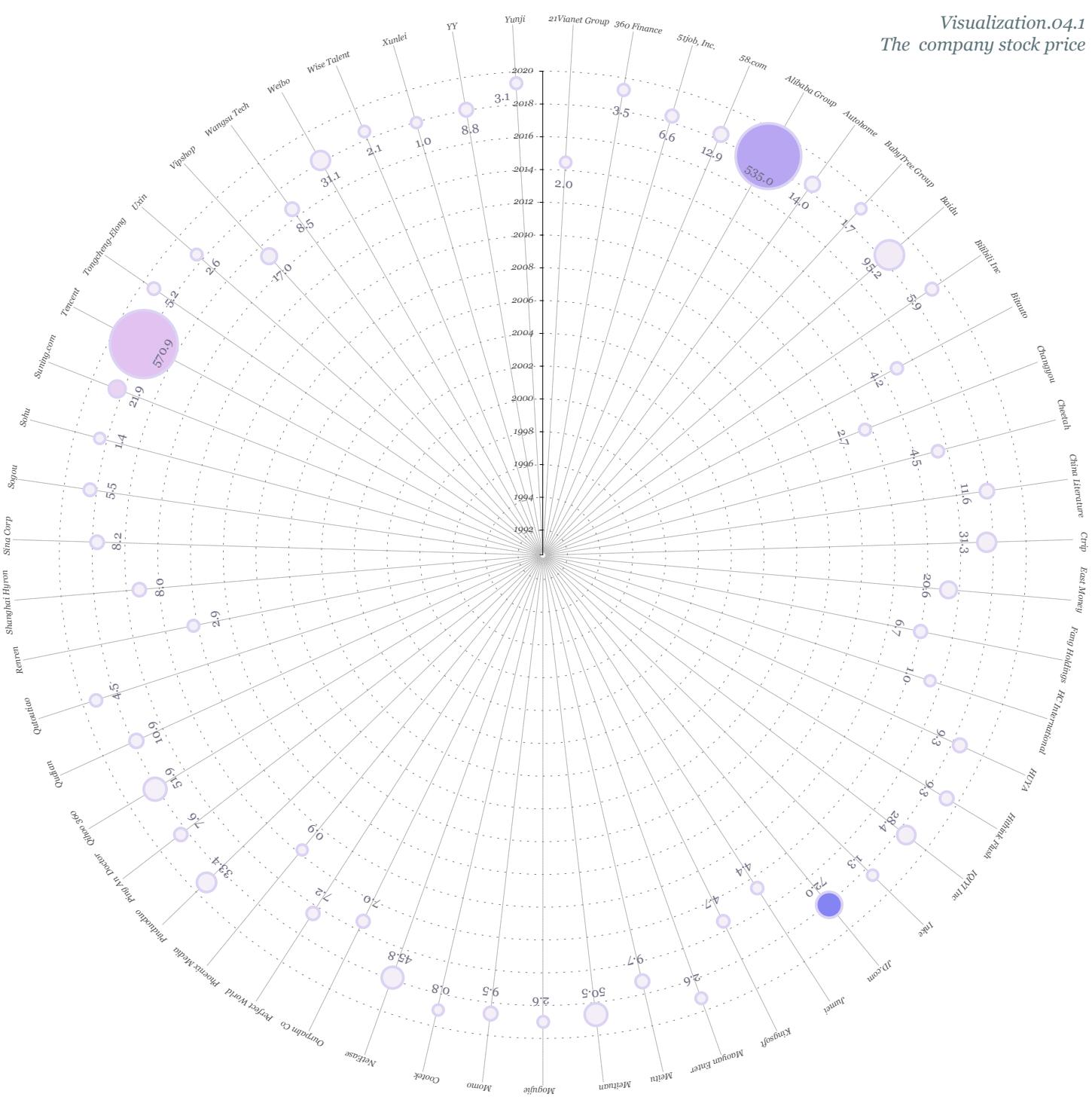
Insights

In the visualization 04.1, the gradient color of the circle with borders simplified the Company revenues in 2018 (billion CNY), the darker, the higher. The size of the circle with borders means the max total stock value of the company (billion Dollars).

According to the illustration, the time, when the market value of these listed companies reached the maximum, was averagely distributed in 2016-2019, which was the golden period of Chinese Internet companies. The highest market value was Tencent in \$570.9 billion, followed by Alibaba in \$535 billion, and the third one was Baidu, \$95.2 billion.

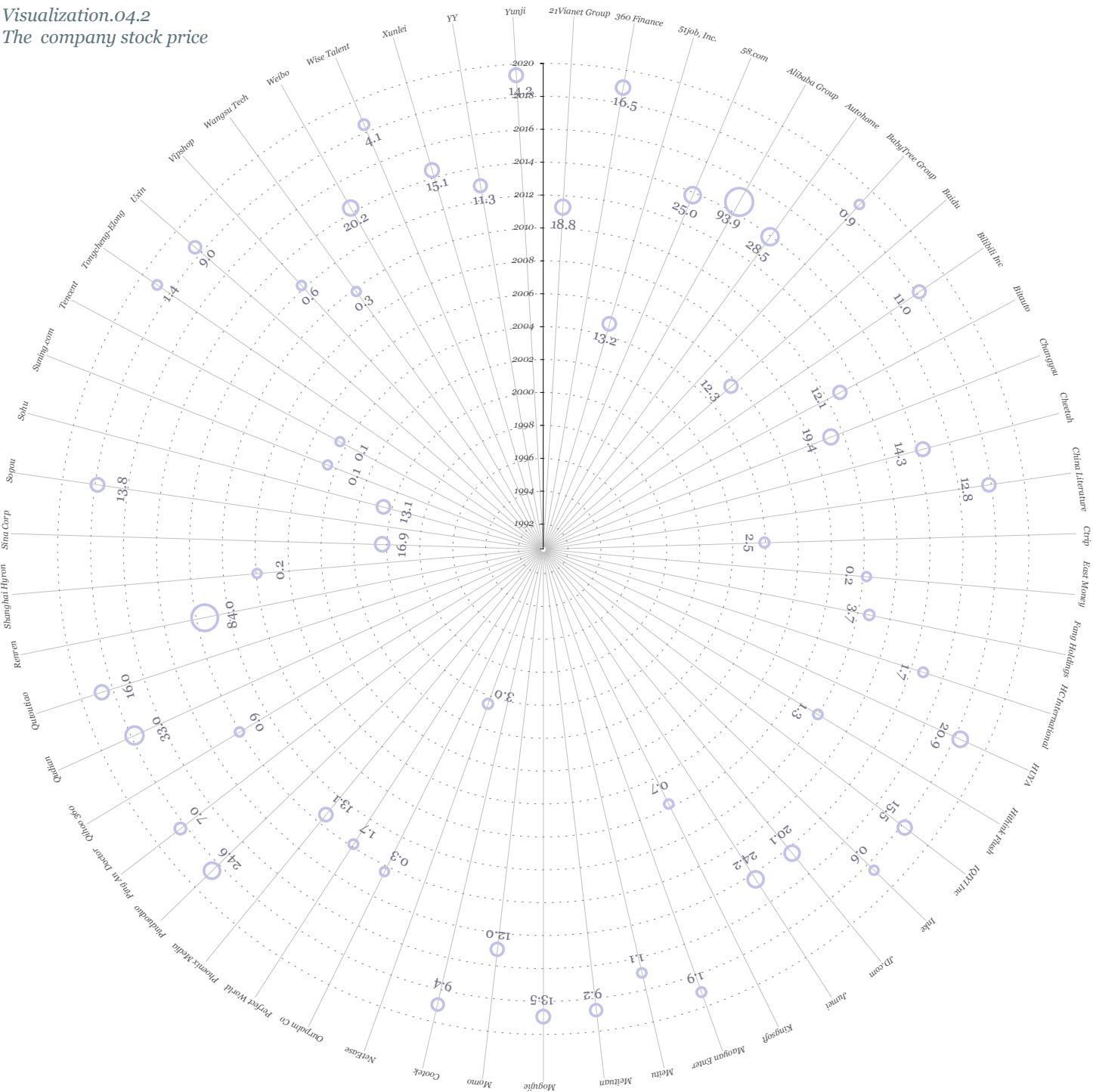
The total stock value of Alibaba or Tencent could overpass 100 times larger than other companies. Besides, the Renren Company, like Facebook in China, was positioned in 2012. Renren used to obtain a proper development but failed at last. The failure of Renren company was an experience worth reflecting. The company founders must consider the particularity of the Chinese market to make the relative right decision.

Visualization.04.1
The company stock price



Visualization.04.2

The company stock price



Insights

In this visualization of visualization o4.2, the purple circle ring represented the issue price. The size of the circle meant the single stock price.

Through the illustration, it could be found that Tencent stock issue price was only 0.1 US dollars, while Alibaba was 93.9 US dollars. As an investor, investing in Tencent must be the right choice. From another opinion, a good company was not necessarily worth investing, which did not necessarily give investors the highest returns — observed from the stock issue date, these companies chose to issue from 2000 to 2019 averagely distributed.

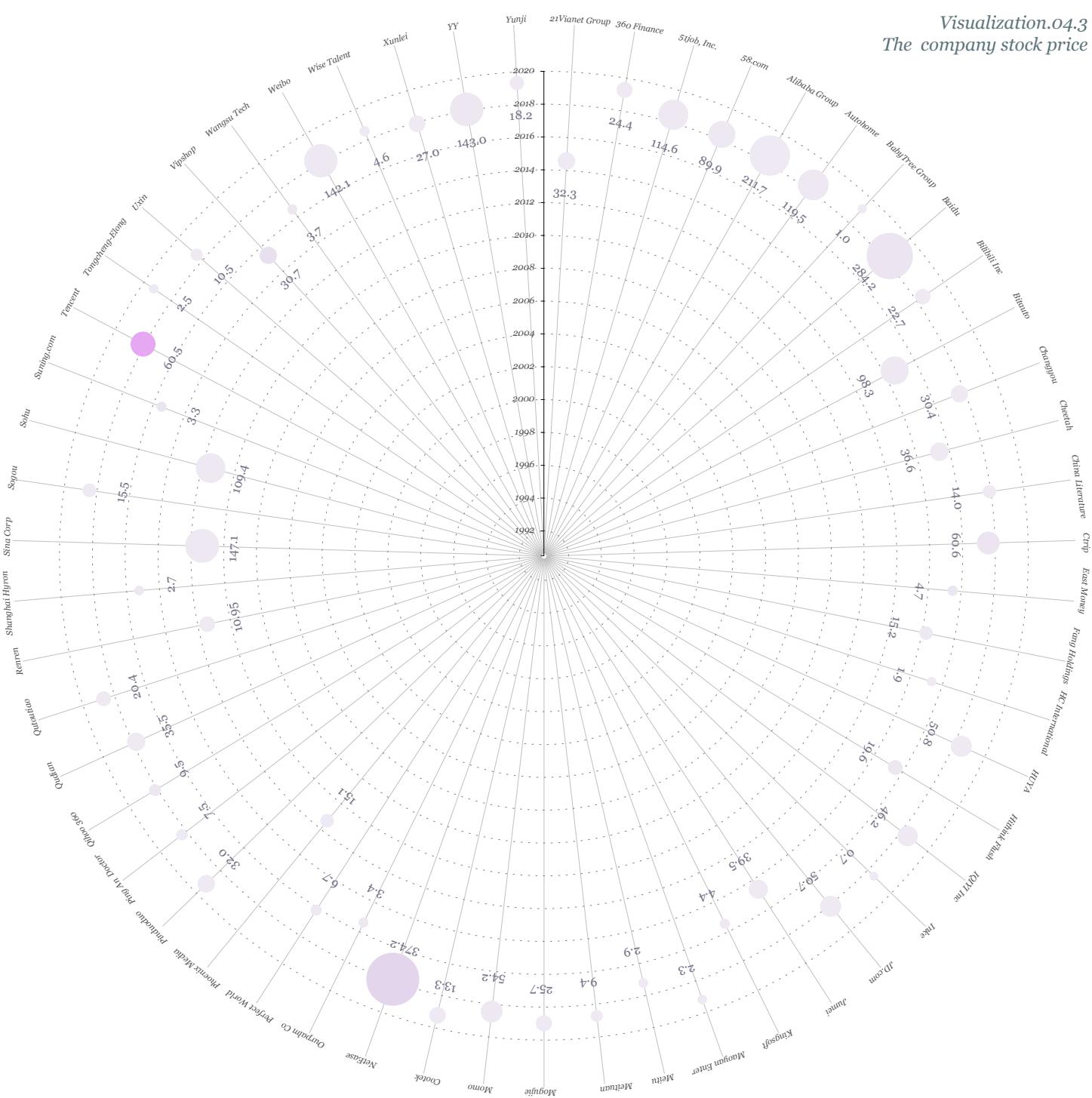
As a whole, the count of stock listed companies was growing with the timeline.

Insights

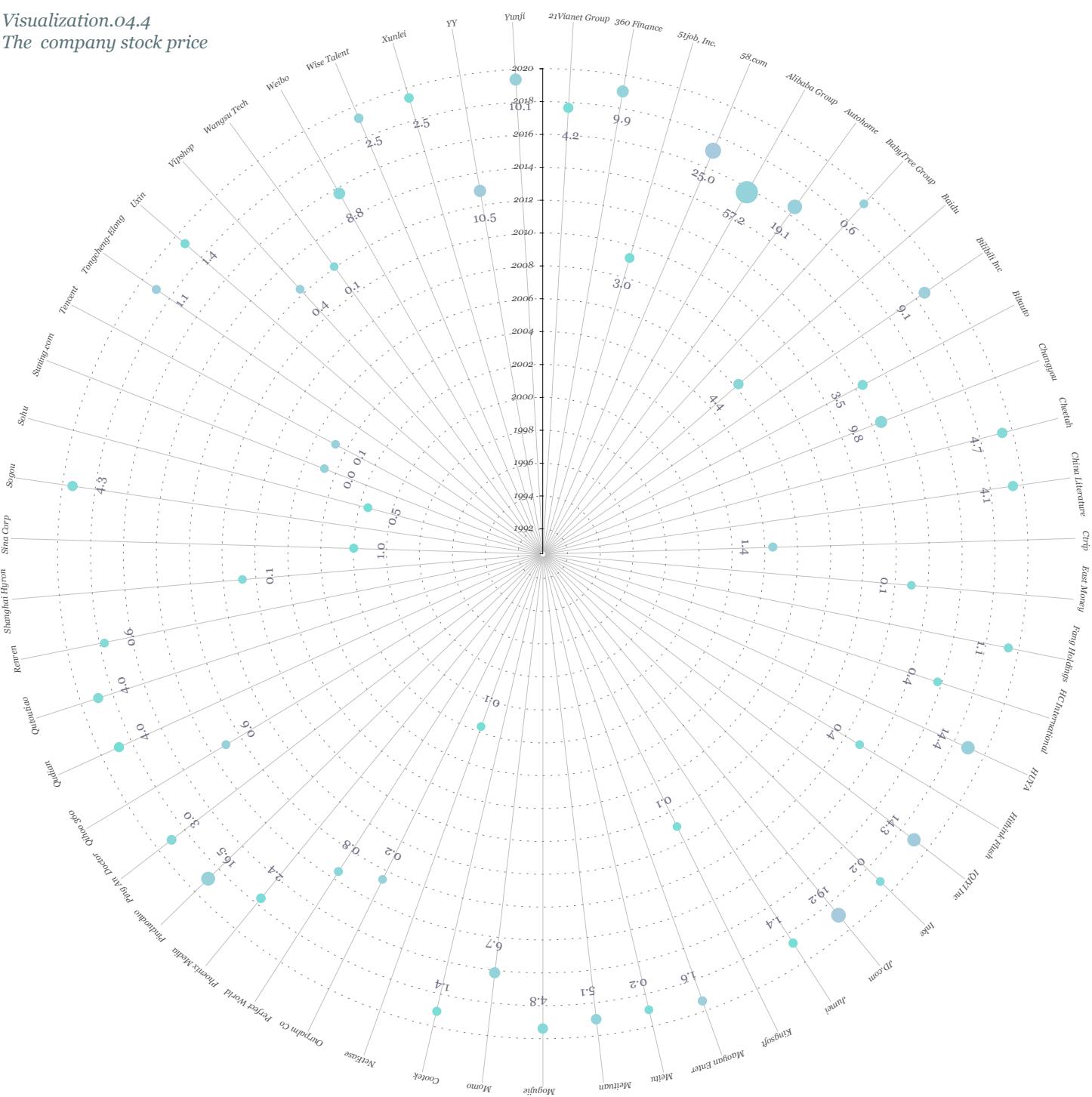
In the visualization 04.3, the gradient color of the circle in gray-pink purple represented the ratio of max price and issue price. The size of the circle meant the single stock price. According to the illustration, in terms of color, Tencent's return on investment eclipsed all other companies.

From historical data, Tencent was the most profitable investing Internet company, with a growth return of up to 600 times. The period when the single stock price reached its highest level concentrated from 2016 to 2019. Besides, the single-stock price and return on investment of some listed companies were still relatively low even though they had reached the highest, which indicated that many listed companies were not worth being invested.

Visualization.04.3 The company stock price



Visualization.04.4 The company stock price



Insights

In the visualization o4.4 , the gradient color of the circle in gray blue-green represented the ratio of min price and issue price, and the size of the circles meant the single stock price.

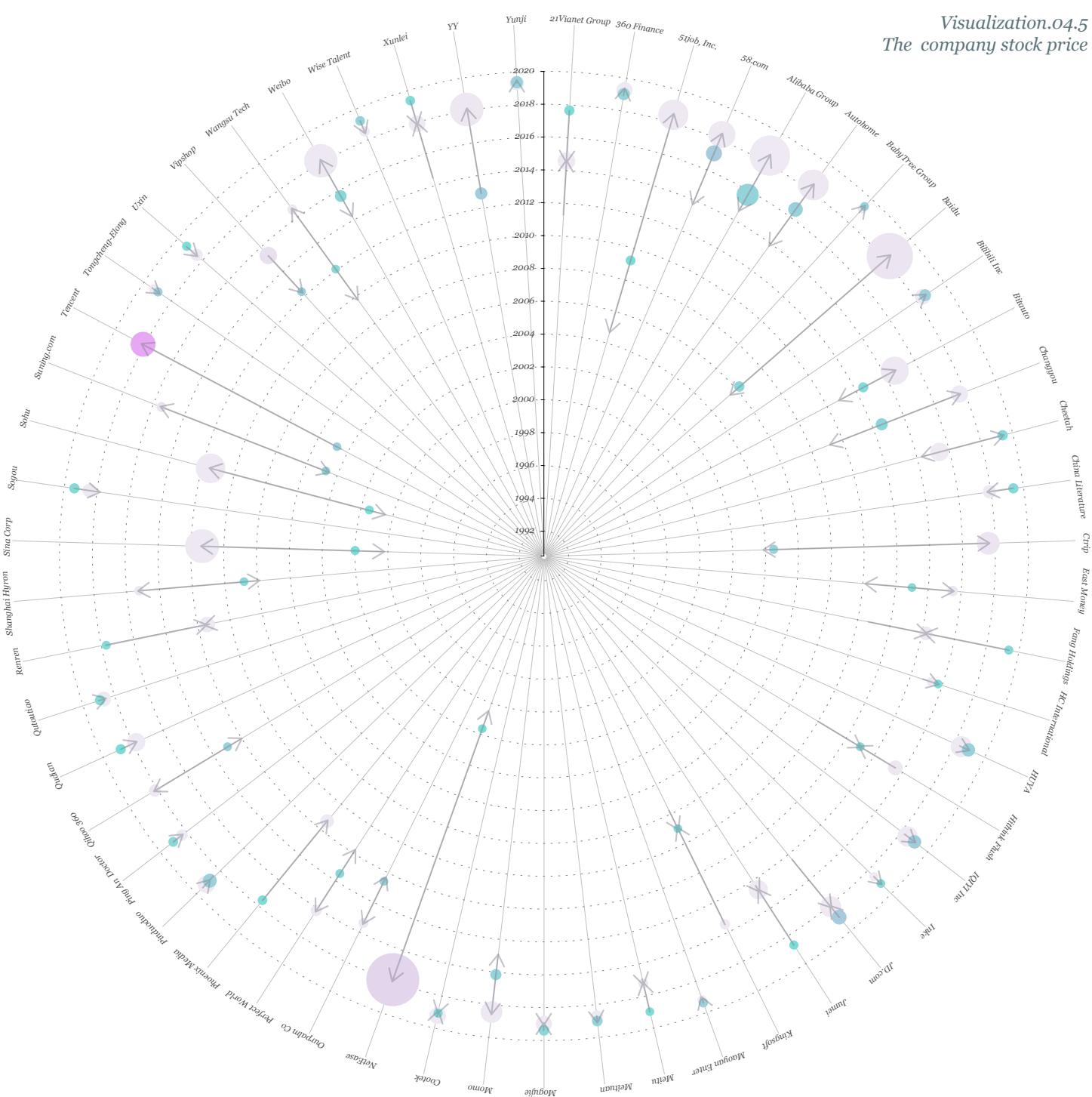
According to the illustration, in terms of color, calculating the ratio of minimum price to issue price indicated the highest loss state of these company stocks in historical data. The overall status showed in green, which meant investing on Internet listed companies of some loss risk. In contrast, as an individual, investing in 58.com, YY, and JD.com was a better choice with less loss.

Insights

In the visualization o4.5 , In combination with the minimum, maximum, and issue prices on canvas and showing the arrows symbol, the growth types would be displayed. The direction of the arrow was from the minimum to the maximum and from minimum to issue.

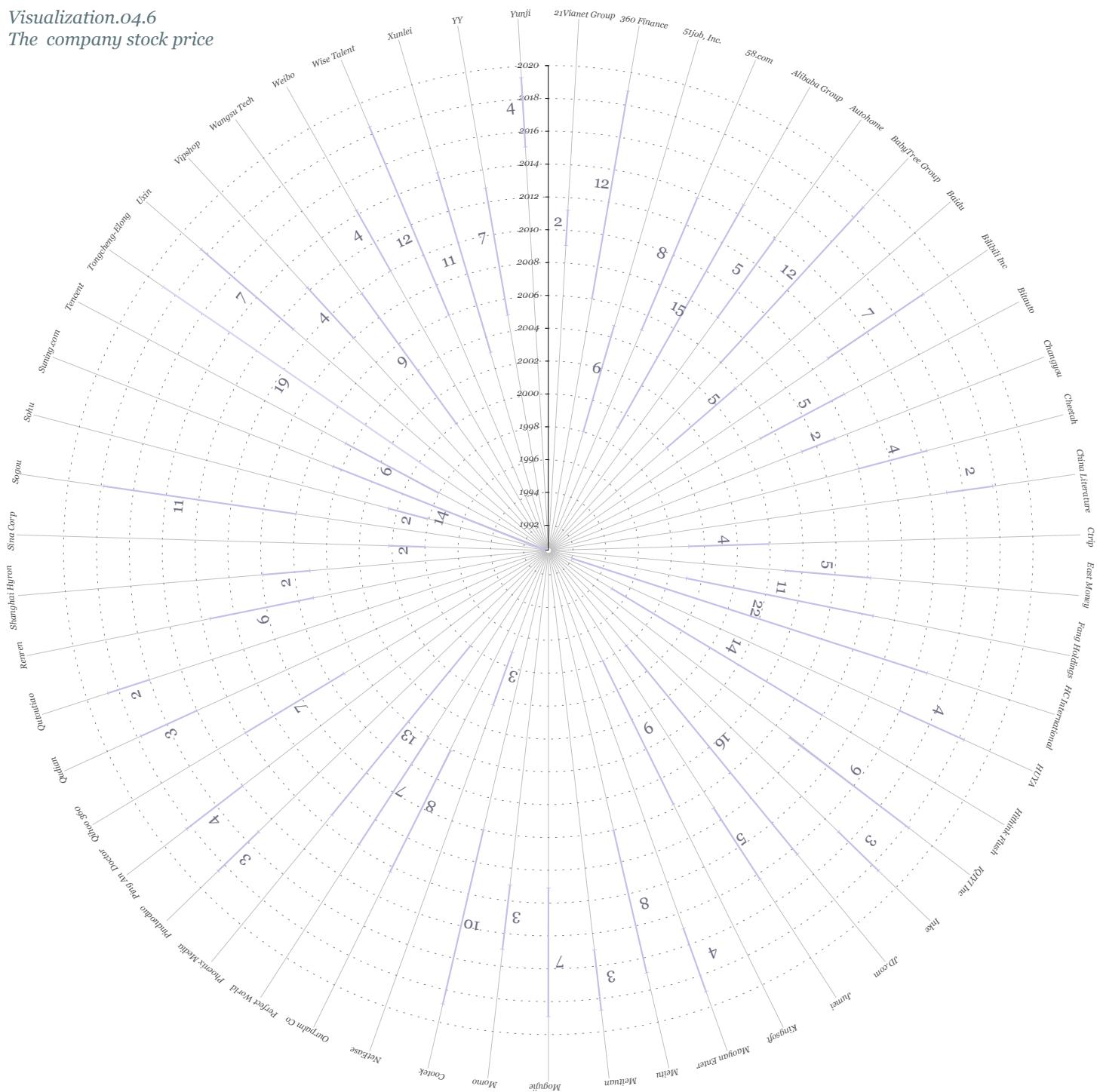
There were the growth types of the four types of companies: increasing first then decreasing, first decreasing then increasing, increasing all the time, and decreasing all the time. The visualization illustrated that most companies belonged to the type of decreasing first then increasing, such as Netease, 51job, and so on. Besides, the longer the company established, the higher the probability of decreasing first then increasing. In other words, if a company could survive the market competition for an extended period, the stock market would show more trust on them.

Visualization.04.5 The company stock price



Visualization.04.6

The company stock price



Insights

In this visualization 04.6 ,the line with a purple color said how many years were spent from foundation date to the issue date for the company.

It was discovered that Suning.com was the first established company in the Internet industry, which was in 1990. Such company was not native to the Internet, but combining Internet services in the process of gradual development to empower the company. Besides, about the company establishment time, the shortest period was two years, the longest was 22 years. Many early-stage companies tended to go public with longer than 5-7 years, and many new companies in recent years often experience a short period of 1-2 years.

How about the founders of Chinese Internet companies, which have been listed on the stock market?

Chinese Internet Companies , Stock Market

中国互联网企业，上市公司

google.com.hk

retain the list for
55 names of Chinese Internet Companies
which have been listed on stock market

file-name.xlsx

the file includes the below element
> company name

search each company's name in

google.com.hk google.com.en en.wikipedia.org ch.wikipedia.org

obatin the relative company founder's
information

file-founders1.xlsx

each file includes the below element
> company name (en / ch)
> founder's name
> birthday
> education background

combined the file-basic.xlsx and calculate in
[Excel](#)

file-founders2.xlsx

the file includes the below element
> company name (en / ch)
> founder's name
> the foundation age
> the stock listed age
> age in 2019
> highly educated in bachelor degree,
> higher education
> financial background
> computer background

Chartificulator

[Adobe Illustrator](#)

vis.

Founders

the visualization includes various models

- > model 1: ranking by foundation age
- > model 2: ranking by age in 2019
- > model 3: ranking by stock listed age
- > model 4: highly educated in bachelor degree
- > model 5: higher education besides bachelor
- > model 6: computer background
- > model 7: financial/management background

Interactive Visualization



2.2.3 Visualization for founders

Of course, analyzing the founders of these listed companies would be interesting, which might reveal the secrets of successors somehow. Searching the names of each founder in Google and Wikipedia could obtain the relative founder's information including the birthday and education background. The final dataset (company name, founder's name, the foundation age, the stock listed age, the age in 2019, highly educated in bachelor degree,higher education, financial or management educated background, computer educated background) was shaped by using Excel(Tool) to clean and organize. Then, importing the dataset into Charticulator(Tool) and thereafter adjusting in Adobe Illustrator(Tool) would get the Visualization.05 about the company stock founders, shown in the protocol.04.

The design logic of the Visualization.05 for variables and properties would be explained in the Table 03, in which the point (circle) and the line(straight) represented each company founders. The circle with outermost position meant the computer education background, the circle with slightly outer position meant the commercial education background, the circle with a smaller size in the slightly inner position meant the higher education backgrounds besides bachelor, and circle with a ring shape in the innermost position meant the advanced education background in famous universities of China. The color of the circles in purple meant true, in green meant false. The line with a green color meant the duration between 2019 and the year when the company was listed on the stock market. The line with a purple color meant the duration between the year when the company was listed on the stock market and the year when the company was established.

As for the layout, the timeline was also changed into the center ring timeline to avoid the extra space consuming. The interior of the ring was the age element of the founders, and the external point represented the educational background of the founders. Each radius axis represented the situation of each founder, and the whole could be contrasted in one single canvas. Combined with the web interaction, there were different forms by sorting model (sorting by age in

2019, stock listed age, highly educated in bachelor degree, higher education besides bachelor, computer background or financial/management background.)

The founder's Age

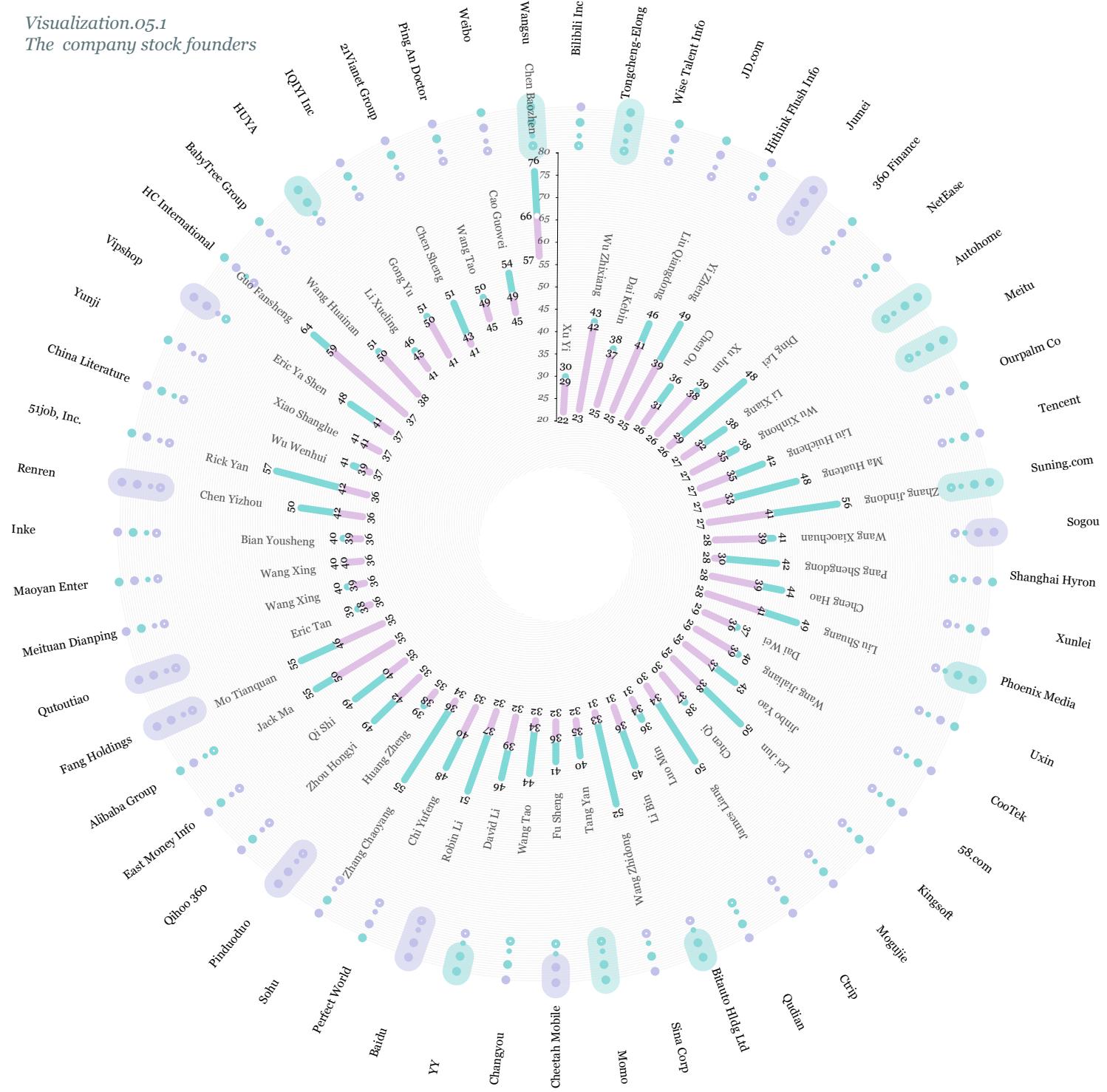


The founder's Education Backrgound

<input type="radio"/>	computer education	<input checked="" type="radio"/>	false	<input checked="" type="radio"/>	true
<input type="radio"/>	financial/management education	<input checked="" type="radio"/>	false	<input checked="" type="radio"/>	true
<input type="radio"/>	higher education besides bachelar	<input checked="" type="radio"/>	false	<input checked="" type="radio"/>	true
<input type="radio"/>	high education in university	<input checked="" type="radio"/>	false	<input checked="" type="radio"/>	true

Table.03

Visualization.05.1 *The company stock founders*



Insights

From the perspective of education background in visualization 05.1, there were a small number of successful entrepreneurs who did receive a high education in bachelor degree and a higher education of master degree. At the same time, they had a computer and business education background, such as Wang Xing(from the Meituan Dianping Inc), Eric Tan(from the Qutoutiao Inc), Huang Zheng(from the Pinduoduo Inc), and Robin Li (from the Baidu Inc). In contrast, there were a small number of entrepreneurs of relatively lower educational background, who could still create some excellent listed companies, such as Wu Xinhong (from the Meitu Inc), Li Xiang (from the Autohome Inc), and Tang Yan (from the Momo Inc).

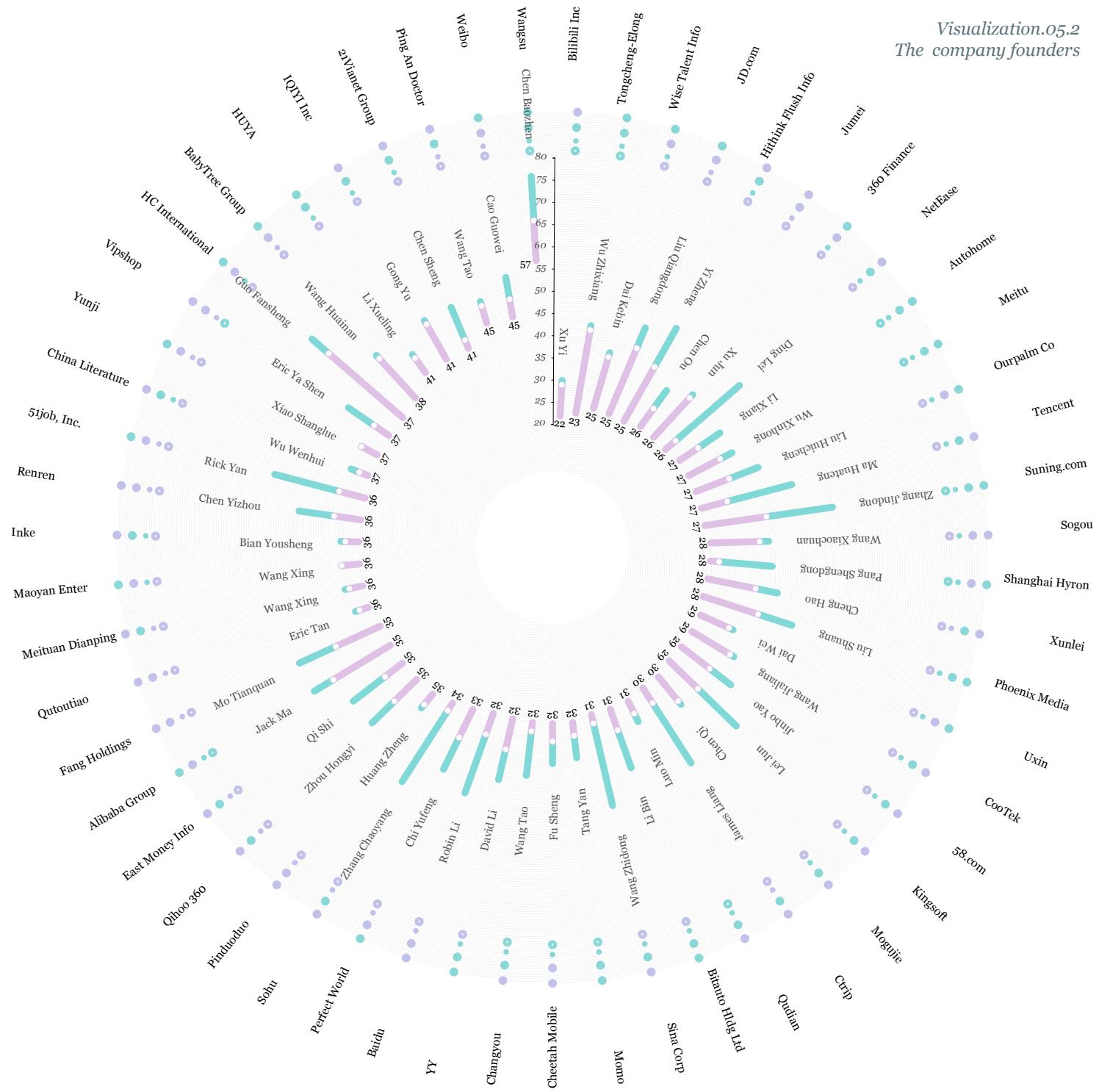
From the perspective of age, the entrepreneurs started their business from 20 to 60 years old. When the company listed on the stock market, the entrepreneur's age ranged from about 30 to 70 years old. Nowadays, as reputational entrepreneurs, they ranged from 30 to 80 years old. It illustrated that, for entrepreneurs, entrepreneurship was their lifelong dedications.

Insights

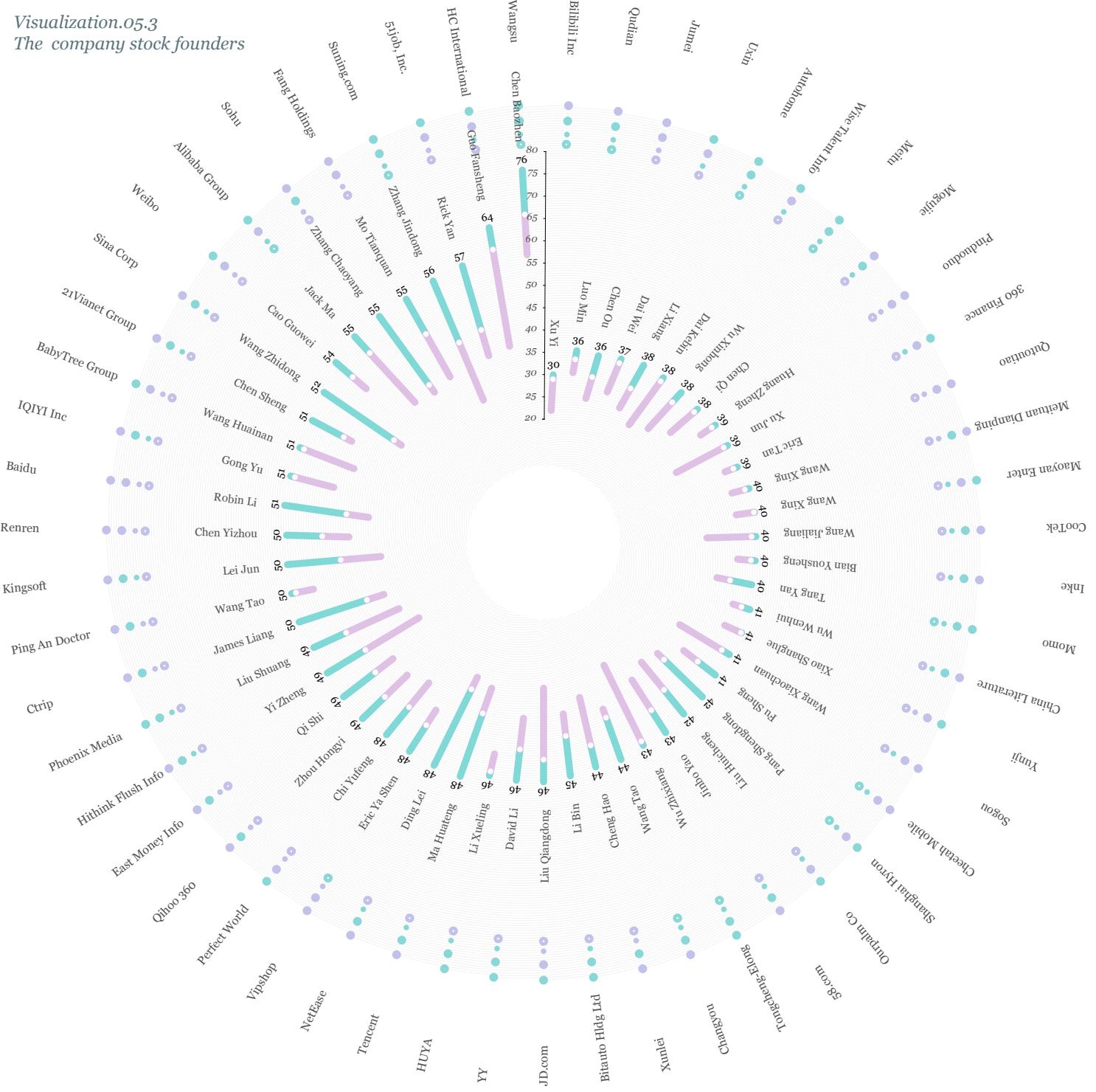
Ranking by the founders' age in visualization 05.2 When the companies established, most entrepreneurs were 25-30 years old. The youngest entrepreneur, XuYi, founded the Bilibili website at the age of 22, which was a UGC(user Generating Content) website similar to youtube.

Therefore, touching the punch light in the vertical field could shape an opportunity to succeed in business. However, later, Bilibili was acquired, and Xuyi was forced to leave, which showed that the founder's industry experience and educational background would affect individual development. In contrast with the oldest entrepreneurs, Chen Baozhen, 57-year-old, the investment and founder of Wangsu Technology Inc, who was also the only woman on this list.

Visualization.05.2 *The company founders*



Visualization.05.3 The company stock founders



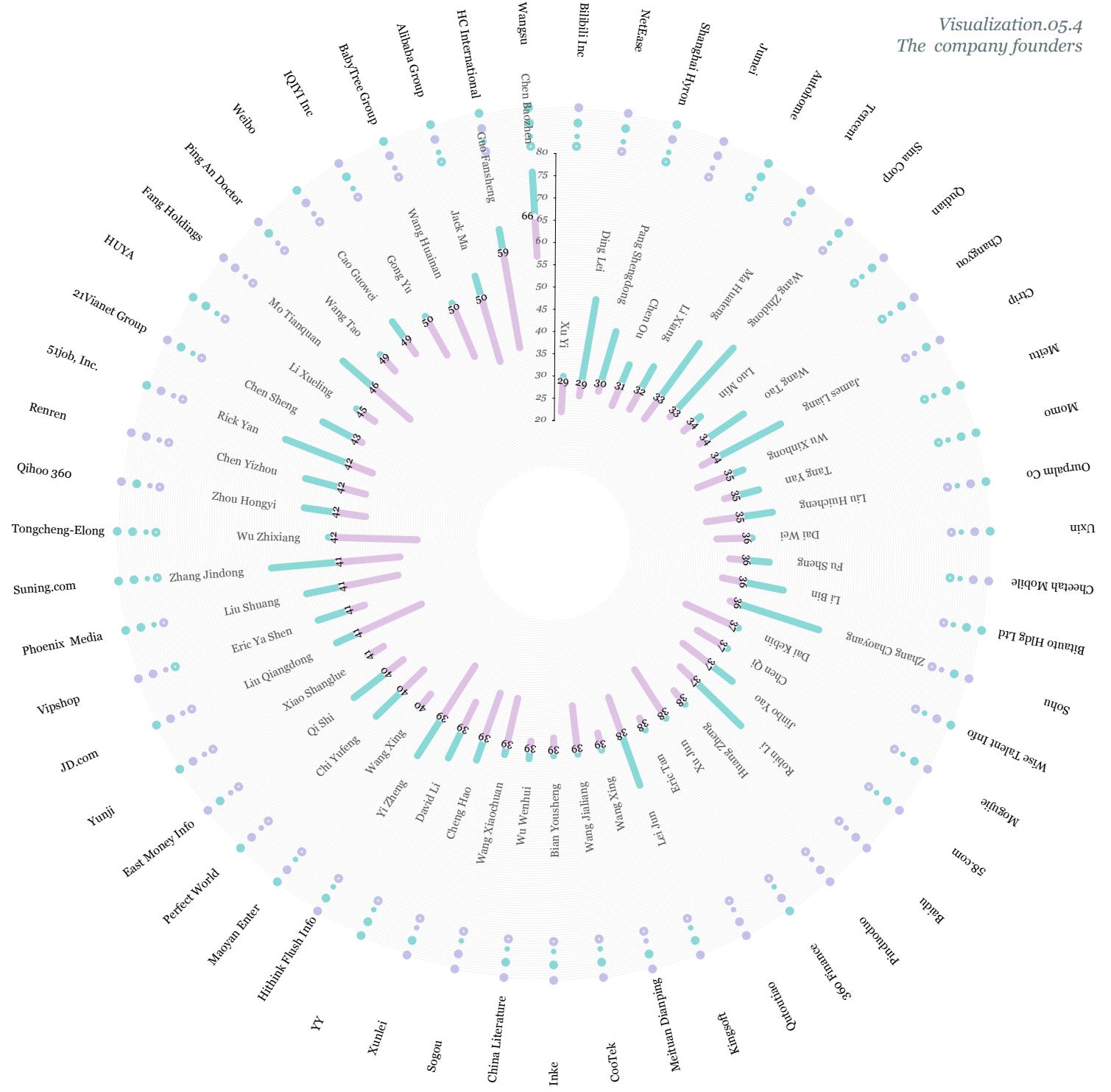
Insights

Ranking by the age of the founders in 2019 in visualization 05.3, it was concentrated between 40 and 60 years old, and the distribution was relatively even. In them, many entrepreneurs still maintained a young mindset, embracing the changes of the Internet industry with an open mind, and shaping the company transformation strategy. This illustration told us that to be an entrepreneur and to be successful, the industry experience and time-consuming depended.

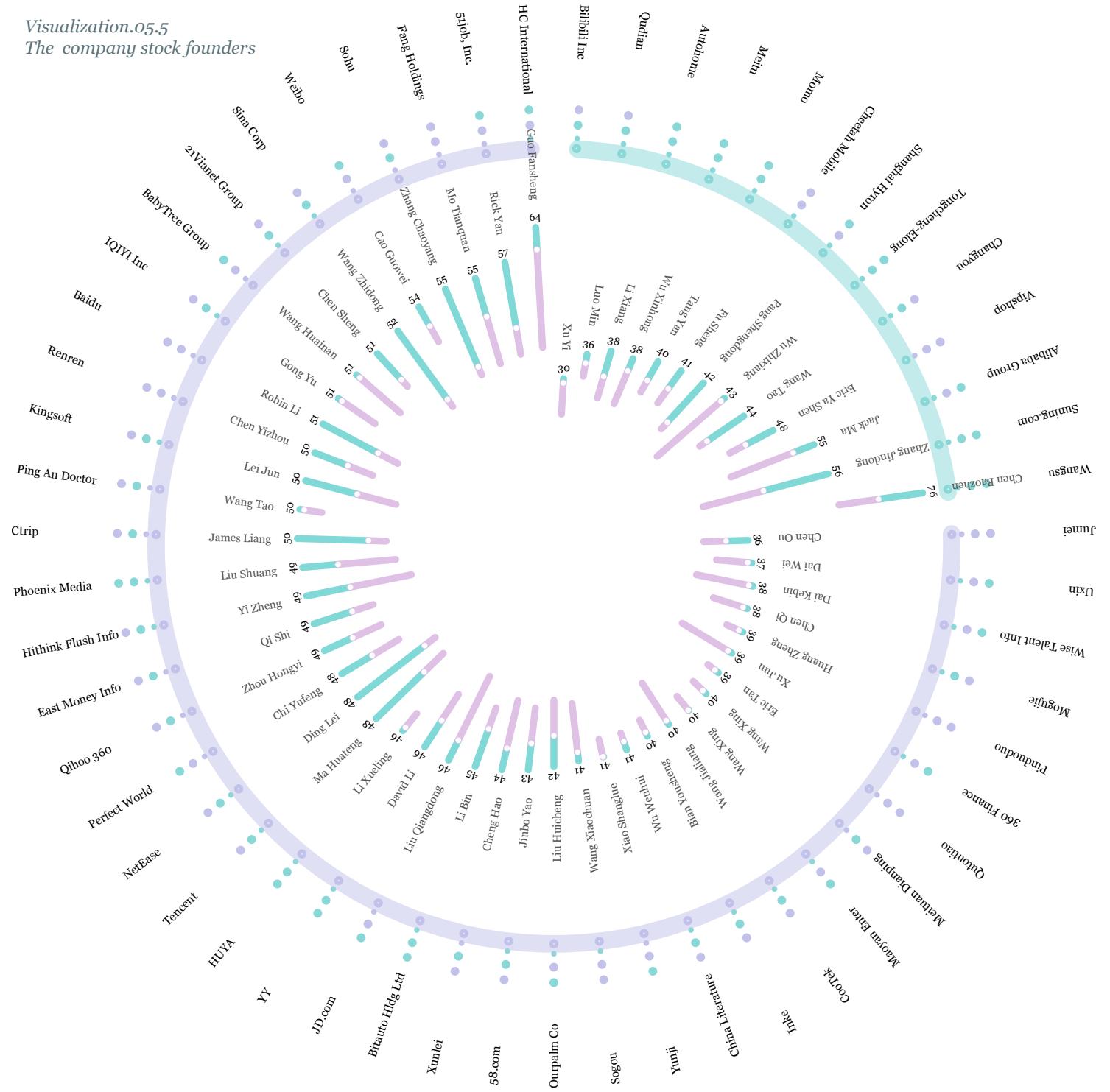
Insights

Ranking by the age of the founders when the company was listed on the stock market in visualization 05.4 , most entrepreneurs concentrated between the ages of 35 and 45, but the time length between the establishment-time and issue-time varied from company to company. Some founders went through a few years, while some founders went through for more than a decade. Entrepreneurship was not an easy task, which took time to settle and needed the founders accumulating the industry experience.

Visualization.05.4 The company founders



Visualization.05.5 The company stock founders



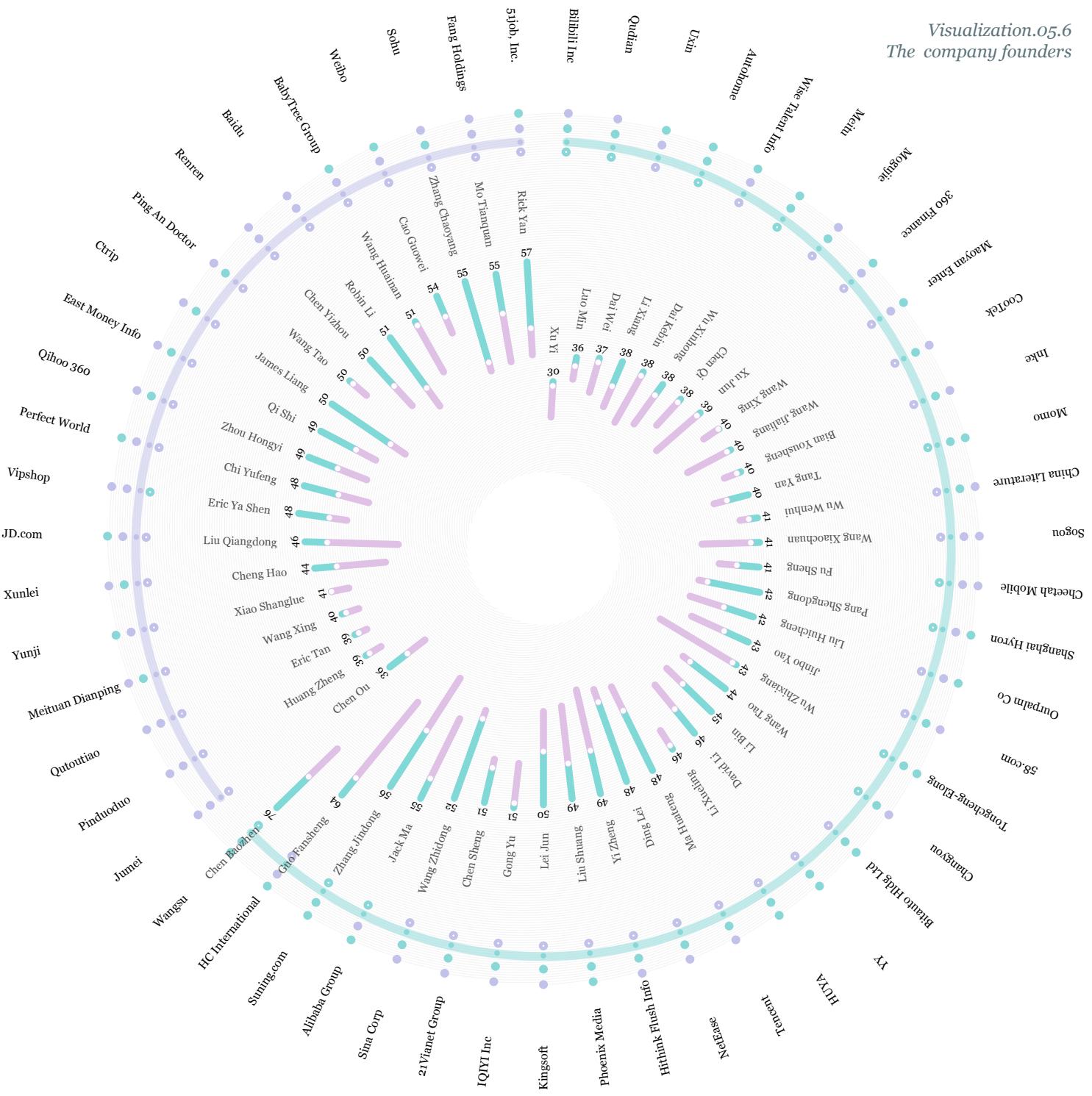
Insights

Classifying by whether the founders received a bachelor degree in an advanced educational university in China in visualization 05.5, 75% of entrepreneurs did. The entrepreneurship in the Internet industry was different from traditional Chinese manufacturing in the past, in which people with higher education had the potential to succeed. However, there were also a few people, such as Ma Yun, the founder of Alibaba. Although he did not own one bachelor degree at a prestigious university, he still created a great world-class company. This illustrated that the vision of the entrepreneurs and education diplomas did not necessarily have a necessary connection.

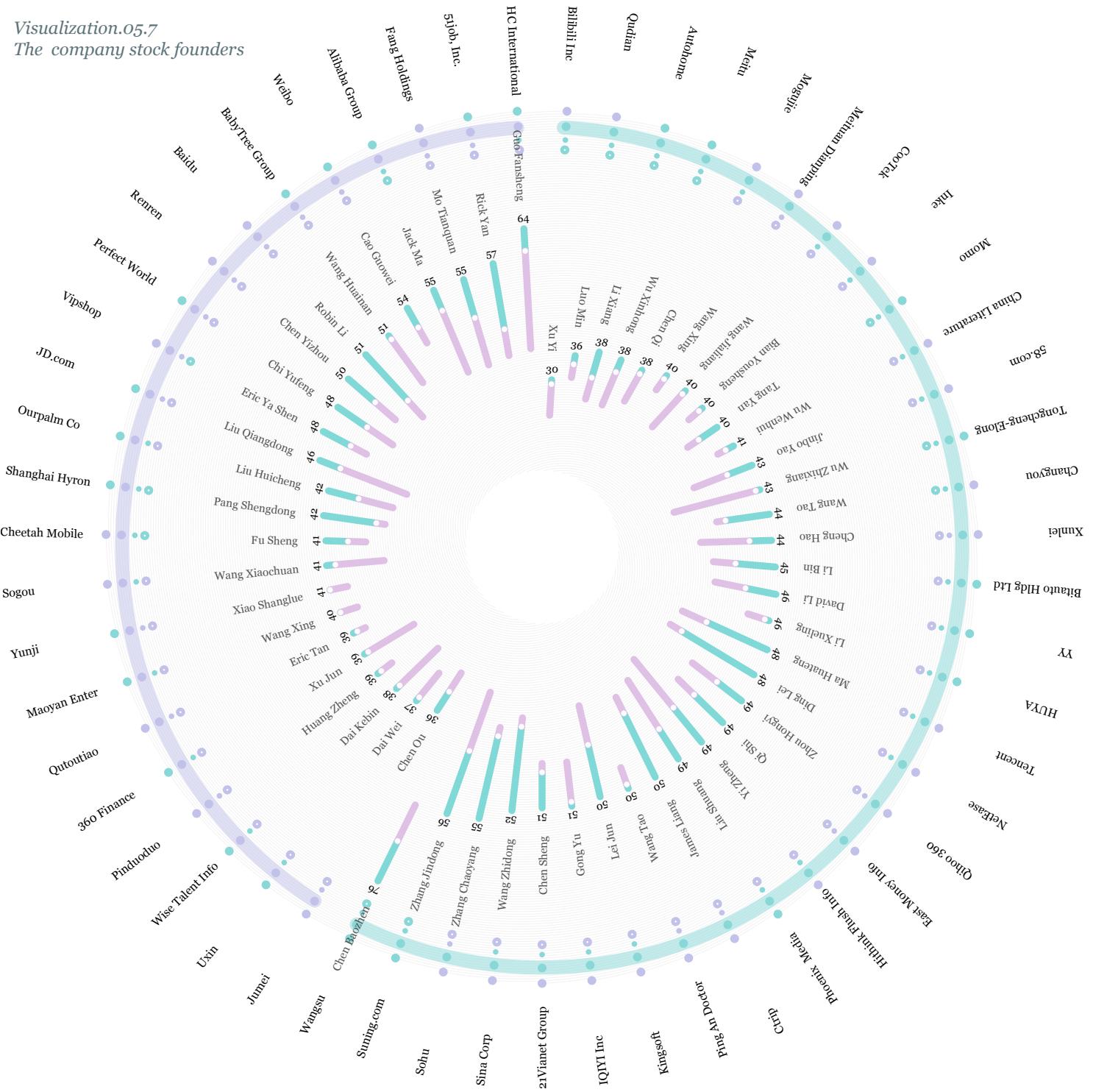
Insights

Classifying by whether the founders received a master degree for a higher-level education in university in visualization 05.6?, most of the entrepreneurs did not, which indicated that whether the entrepreneur had a high level of education background was not necessarily related to whether the entrepreneur established a successful company.

Visualization.05.6 The company founders



Visualization.05.7 The company stock founders



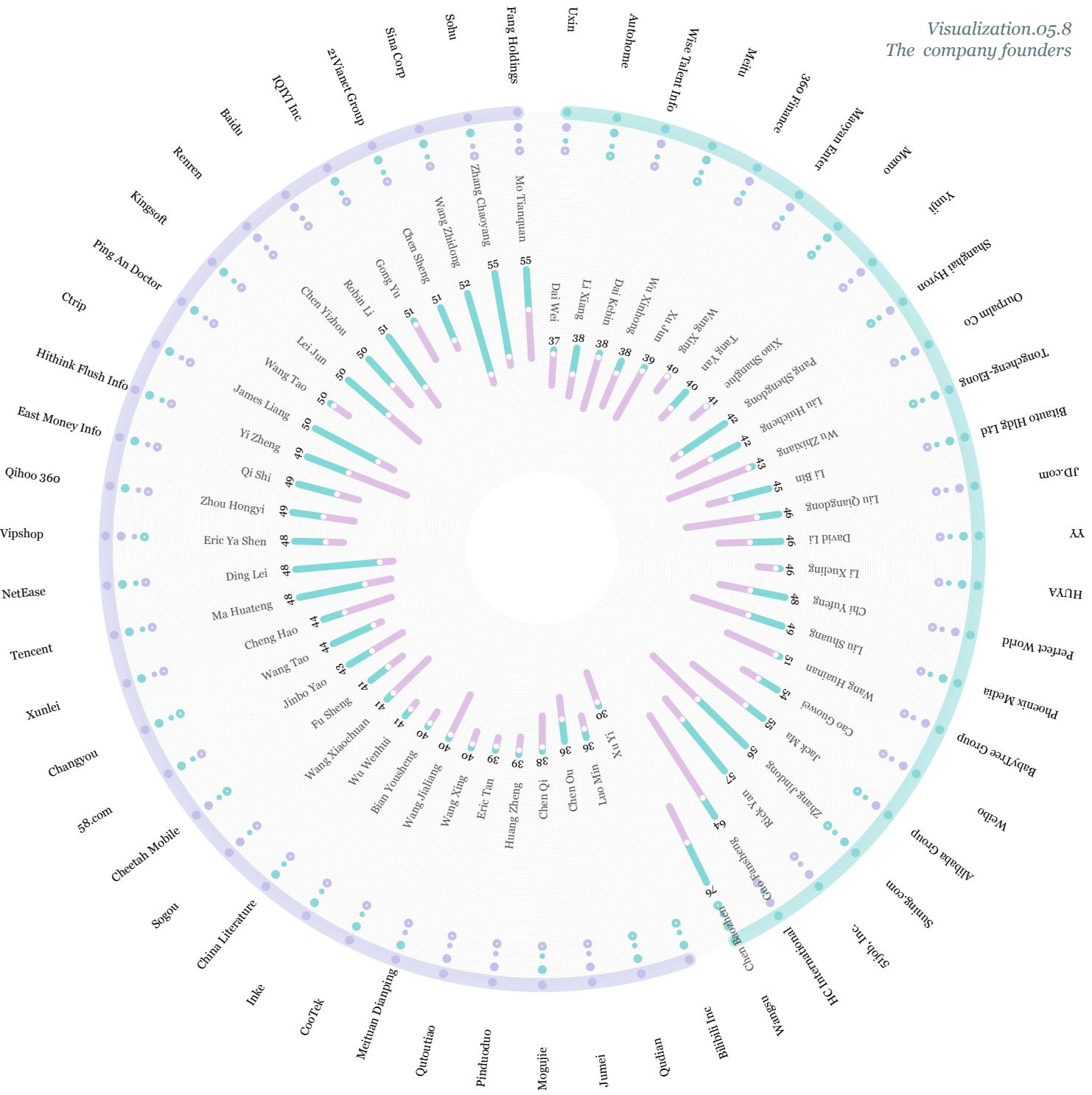
Insights

Operating a successful business and obtaining profits required knowledge and assistance from business administration. As well, the entrepreneurs would also often play the role of a manager. By sorting out whether the founders had a background in business administration education in visualization 05.7, more than half do not. At this time, the co-founder CMO(Chief of Management) was required to join.

Insights

Internet as a high-tech industry, computer technology development was a key factor. By sorting out whether the founders had a computer technology background, more than half did in visualization 05.3. If the company founders with no computer technology background, the co-founder CTO(Chief of Technology)was needed.

Visualization.05.8 The company founders



2.4 Data Visualization of Unicorn Internet Companies in China

'A unicorn is a privately held startup company valued at over \$1 billion.' [8] Apart from the stock listed companies, many unicorn companies, such as Ant Financial, Toutiao, and Didi, also had an enormous impact on the Chinese Internet market today. Based on light industrial assets but heavy services, Internet company development required a large amount of financing today. So the company valuation became an important indicator to measure the company's development. So the question in this section was 'How about the situation of the Chinese Unicorn Internet Companies?' Overall, in this section, combined with the dataset of the Chinese Internet unicorns listed on ITjuzi[9], the design logic and the process for interactive visualization of matrix plots would be illustrated to explore the current state of the unicorns in China.

How about the situation of the Chinese Unicorn Internet Companies?

Chinese Internet Companies , Unicorn

中国互联网公司，独角兽

itjuzi.com

retain the list for information of 224 companies in different fields which have been listed as the unicorn companies in Chinese Internet

[unicorn-health.xlsx](#)

[unicorn-hardware.xlsx](#)

.....

[unicorn-car.xlsx](#)

[unicorn-finance.xlsx](#)

manually check the data at
qichacha.com



each file includes the below element

- > company type
- > name (CH)
- > latest financing time
- > total financing valuation (100 million\$)
- > latest financing round
- > key words
- > company founding time
- > company Size(person)
- > company based address

input the excel files into
[Tableau Prep Builder](#)



to intergrate the files into one list and to calculate how many years since each company was established

[unicorn.csv](#)



the file includes the below element

- > company type
- > name (CH)
- > latest financing time
- > total financing valuation (100 million\$)
- > latest financing round
- > key words
- > company founding time
- > company Size(person)
- > company based address
- > founding year

[Charticulator](#)

[Adobe Illustrator](#)

vis.

[Unicorn Value](#)



the visualization includes two models

- > model 1: presenting the note of key points
- > model 2: without presenting the note of key points

vis.

[Unicorn Size](#)



Interactive Visualization



Interactive Visualization



The professional database of itjuzi.com was found through the search engine, containing the list of Chinese Internet unicorns. By checking with the Qichacha[10], the official database for Chinese enterprise, the unicorn dataset files of Chinese Internet companies in different fields were obtained. Through the Tableau Prep Builder for cleaning and organizing, the final dataset included the element of company type, name (CH), latest financing time, total financing valuation (100 million\$), latest financing round, keywords, company founding time, company Size(person), company based address, and founding year.

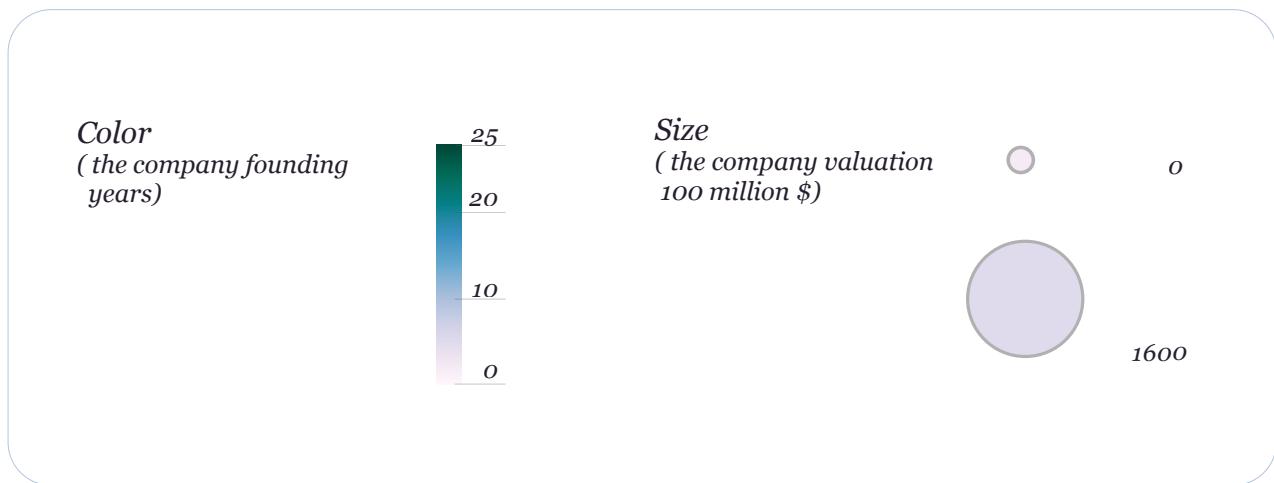
Then, importing the dataset file into Charticulator(Tool) would obtain one visualization o6 about the relationship between unicorn valuation, founding year, unicorn address, and unicorn field. Besides, another visualization o7 was about the relationship among unicorn valuation, founding year, unicorn financial round, and unicorn scale, shown as the protocol o5.

The design logic of the visualizations for variables and properties would be explained in Table o4, in which the point (circle) was chosen to represent each company. In the visualization of unicorn.valuation, the X position of the circle illustrated the different fields(education, e-commerce, enterprise service, entertainment media, tool software, agricultural, car traffic, estate service, finance, game, hardware, local life, logistics, marketing, medical health, social,

sports, and travel), and the Y position illustrated the different company addresses (Anhui, Beijing, Chongqing, Guangzhou, Guizhou, Kongkong, Hubei, Jiangsu, Shandong, Shanghai, Sichuan, Zhejiang), which were arranged in alphabetical order.

The gradient color of the circle represented the founding company year, the darker the longer. The size of the circle represented the company valuation, the bigger the higher. Therefore, a matrix bubble chart was formed. Besides, the summary data of each axis was also displayed respectively on the leftmost and lowermost sides. In the visualization of unicorn.size, the horizontal and vertical coordinates were different, in which the X position represented the company scale, and the Y position represented the company latest financing round. Besides, in these two unicorn visualizations, the audience could hover on each circle to see the tooltip for the company details.

This visualization o6 the illustration was a data visualization of Chinese unicorn companies. One circle represented one unicorn company. The area of the circle was the valuation of the Unicorn Company in latest duration, and the color of the circle represented the establishment duration of the unicorn company. The darker the color, the longer the establishment duration of the unicorn company. From the illustrations, these unicorn companies mainly concentrated in



Chinese economically developed regions, such as Beijing, Shanghai, Zhejiang, and Guangdong, which could also see from the illustration same visualization o3 (that the four provinces obtained the highest total stock market value of listed technology companies).

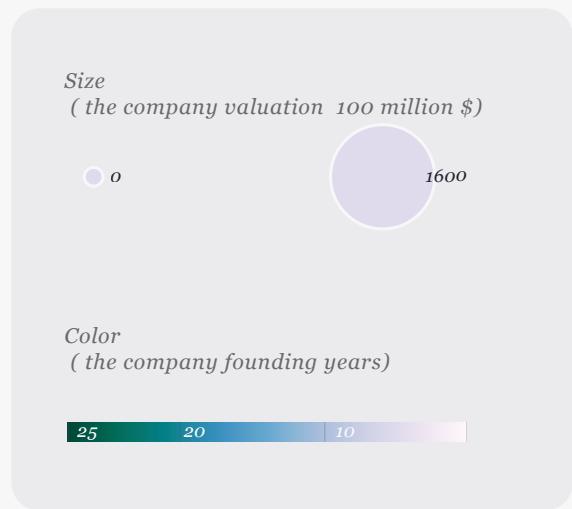
These listed Internet companies also played an essential role in venture capital institutions, and they invested in a massive number of unicorn companies for startups in China. The large listed companies reserved a large number of talents for these unicorn companies, which was one of the key factors contributing to the unicorn growth. Besides, the unicorn company in Beijing presented a similar status as the listed companies – in small scale but numerous count.

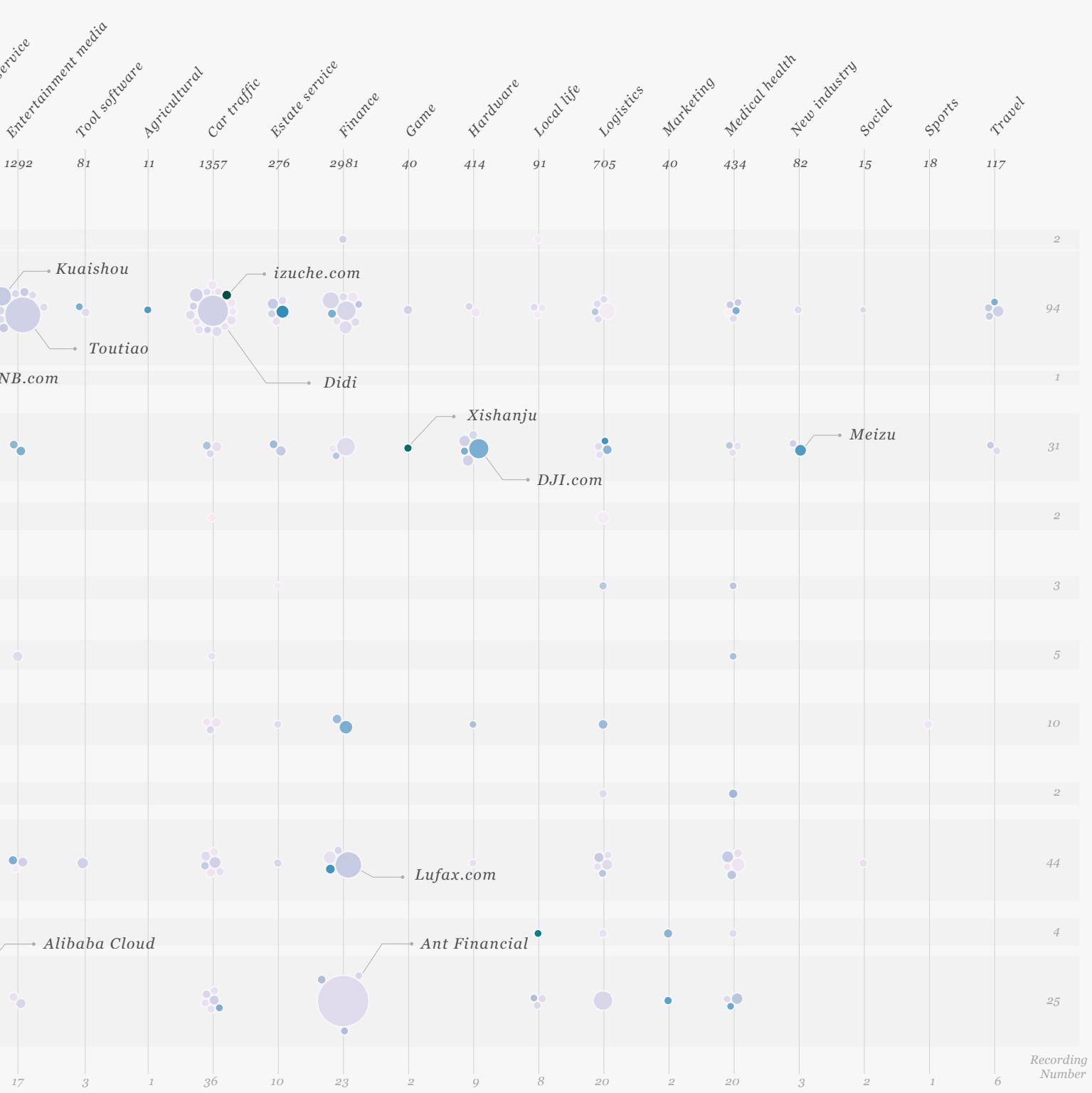
One of the most interesting points was that in the eastern of China, such as Guizhou province, where the economic development was relatively backward, two unicorn companies were born in there. It was firmly related to the Guizhou industrial policy. In the past two years, Guizhou Province was vigorously developing on the digital economy. The local government invested heavily in big data and cloud computing as the core industries, providing enterprises with excellent infrastructure and preferential policies. With the cool climate, abundant hydropower, and low reserve prices, many large companies deployed their data centers into Guizhou. These advantages could contribute to breeding a unicorn company.

From the focusing field of the unicorns, which are the most highly valued, the industrial clusterings varied from field to field. For example, DJI, the highest valuation unicorn company in Guangdong Province, belonged to the hardware field. Moreover, Guangdong Province had far more unicorn companies in the hardware field than other regions. It could be inferred that the industrial chain of Guangdong in the hardware manufacturing field was the most complete. While the headquarter of Alibaba located in Hangzhou, the two super unicorns shaped by Alibaba, named Ant Financial and Alibaba Cloud, contributed to the economy of Zhejiang Province. Besides, as the financial capital of China, Shanghai spawned Internet finance companies like Lufax.

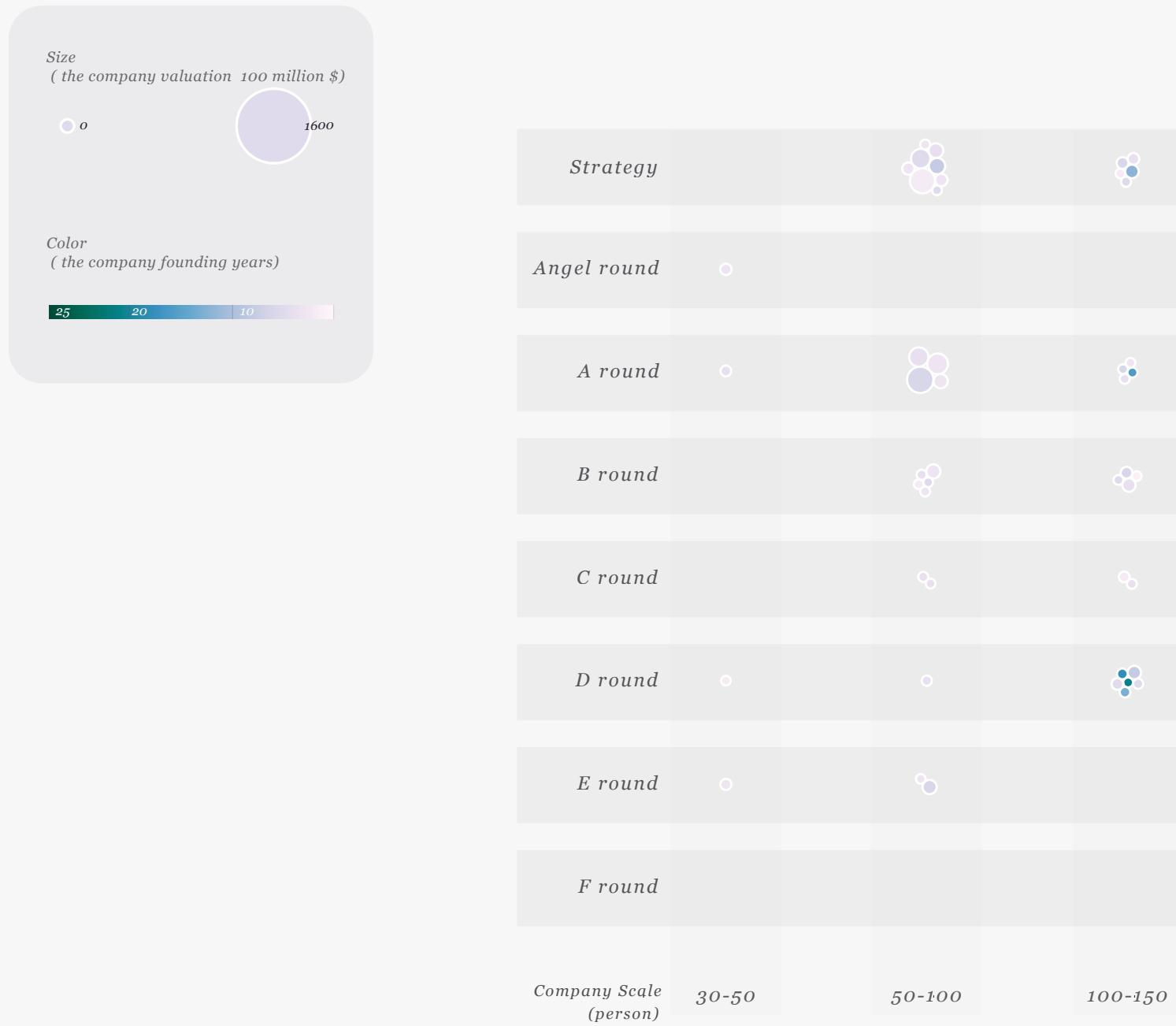
This visualization 07 was about the relationships among company financial round, and the scale of the company, the company establishment duration, and the company valuation data. It indicated that the size of the company personnel was positively related to company valuation. Besides, the scale of the most unicorn companies concentrated in the 100-1000 range. It can be seen from the company founding time and financing round that many unicorn companies quickly reached the D round of financing in recent years. On the one hand, Chinese science and technology sector was developing rapidly in the current decade, with the rapid development of venture capital industry, which might lead to the unicorns obtaining many rounds of financing for survival. From another perspective, many new unicorns developed rapidly, which were able to overpass the D round of funding in a short time.

The Unicorns Of Chinese Internet Companies





The Unicorns Of Chinese Internet Companies





2.4 Data Visualization of China Internet Company Registration Data

The leading companies of the Chinese Internet (including the stock listed companies and unicorns) were explored, but they could not fully represent the overall development of Chinese Internet companies. Therefore, if we could collect data of the industrial and commercial registration in Chinese enterprises from the source, and then filter out the registration data of Chinese Internet companies, it would be beneficial to analyze the macroscopic situation of the Chinese Internet. Through searching on Google, fortunately, the relevant database for Chinese industrial and commercial enterprise registration of Chinese mainland from 1978 to 2019 was found [11] So the question in this section was ‘How about the historical registration of Internet companies in China?’ Overall, in this section, combined with the dataset of the Chinese Internet registration on Github, the design logic and the process for interactive visualization would be illustrated to explore the historical states of Chinese Internet companies.

How about the situation of the Chinese Unicorn Internet Companies?

Chinese Internet Companies , Unicorn

中国互联网公司，独角兽

itjuzi.com

retain the list for
information of 224 companies in different fields
which have been listed as the unicorn companies in Chinese Internet

unicorn-health.xlsx

unicorn-hardware.xlsx

.....

unicorn-car.xlsx

unicorn-finance.xlsx

manually check the data at
qichacha.com



each file includes the below element

- > company type
- > name (CH)
- > latest financing time
- > total financing valuation (100 million\$)
- > latest financing round
- > key words
- > company founding time
- > company Size(person)
- > company based address

input the excel files into
[Tableau Prep Builder](#)

to intergrate the files into one list
and to calculate how many years
since each company was established

unicorn.csv



the file includes the below element

- > company type
- > name (CH)
- > latest financing time
- > total financing valuation (100 million\$)
- > latest financing round
- > key words
- > company founding time
- > company Size(person)
- > company based address
- > founding year

[Charticulator](#)

[Adobe Illustrator](#)

vis.

Unicorn Value



the visualization includes two models

- > model 1: presenting the note of key points
- > model 2: without presenting the note of key points

vis.

Unicorn Size



Interactive Visualization



Interactive Visualization



Shown in protocol.06, after forking and downloading the database, the entire database for 34 provinces in China from 1978 to 2019 was obtained, each file included the elements of company name, province, registered date, registered capital, representative legal person, and company business scope description. It is focusing on Internet industry, by filtering the data by keywords ‘computer, network, software, communication, E-commerce service’ of the business scope and then inserting the file into Mysql for quickly query and access. After selecting useful fields of the data, the final Internet company CSV file was obtained, which included the elements of company name, registration date, registration capital, and registration province, registration scope, and company business scope description. The representative legal person was deleted because it was less relevant, and the registered capital was deleted because it cannot reflect the economic situation of the company truly. Besides, after the cleaning, this dataset contained more than 140,000 data.

Combining the business scope description with the dimensions of timeline, the development of the entire Internet industry in China might be incepted. Due to the fact that each company scope description was very long, through Python NLP library “Jieba”, the description frequency dataset file was obtained, through the process to break the business scope to tokens, remove stop words, count the business word frequency grouped by years, and to filter the word

frequency ranked in top 50. In Visual Studio Code to format and export the final dataset, including dimensions of word, frequency, year. Then, importing the dataset file into Raw Graphs would obtain one visualization o8 about the word frequency from 1980 to 2019, and another visualization o9 about the word frequency from 1980 to 1990.

As far as the design of data visualization was concerned, in this two visualizations of word frequency flow, the X position illustrated the timeline, and the flow width illustrated the word frequency, the wider, the higher. In the visualization of word frequency from 1980 to 2019, the gradient color also illustrated the word frequency, the greener the higher. The layout was a horizontal flow chart along with the timeline, and the size of the text was adjusted according to the size of each flow line. Thus, the interaction for this visualization was designed reasonably for zooming in and out to check the words of tiny size. Moreover, by clicking the button could select the two browsing models of all years or from 1980 to 1990, which also achieved the effect of magnifying the invisibility in 1980-1990 at the all year visualization.

The visualization o8 illustration was a word cloud chart based on the company business scope of newly registered Internet companies in China from 1980 to 2019, which filtered the top 50-word frequency for the statistics. As the illustration showed, the flowing size contrasts of word

frequency statistics among each year were huge, and it could be seen that Chinese Internet companies developed rapidly since 2000, which reach the peaks between 2016 and 2018. In the Chinese market, a large number of Chinese companies were registered and collapsed every day. From the word cloud, the three words of technology, service, and computer were the highest-frequency words, as technology companies described their business scope most frequently. The Internet companies attached considerable weight to technology role, with emphasizing Internet for customer service to obtain profits, which was different from the traditional manufacturing industry. Besides, other keywords such as e-commerce, design, and culture can be seen in various Internet company scopes. The Internet services in China became more humanity from offering infrastructure services at the very beginning to emphasizing the multi-dimensions of design, user experience, culture identity, and other diversity. In 2008, the word flow size declined, which implied that the number of newly registered technology Internet companies declined. This period happened to be the financial crisis in the United States, and Chinese enterprises were also affected. Then, there was a small climax in around 2010, when Apple released the iPhone 4, which set off a wave of mobile Internet climax. China just caught up with this wave with some a series of measure policies for the financial crisis. After that, it could be seen that the number of company registrations during this period increased significantly.

Due to the rapid development of the Chinese Internet and Communication industry in the past 40 years, in contrast, data from 1980 to 1990 became invisible. By individually extracting the data from 1980 to 1990 for the word cloud statistics of the Chinese Internet company registration scopes on visualization.09, the visualization for this period was obtained. It could be found that the China Computer and Electronic Telecommunications Industry started developing in 1980 after the Chinese policy of reform and opening. During this period, the Chinese electronic information-related industries mainly concentrated on computers, various hardware companies, and related fields and service industries, as the keyword ‘engineering, operating and manufacturing’ shown in the visualization.

After importing the Internet company registration data file(company name, registration date, registration province) into Tableau to calculate how many companies were registered on each year of each province, the visualization was embodied by mapping the counting data by color and positioning by provinces on the map. Through Adobe Illustrator, the final visualization was demonstrated by manually modifying. As far as the design of data visualization was concerned, the overall layout was a matrix map by the time dimension, which could be called ‘small multiples.’ The color of each area represented the registration counting, the more red, the more company counting. The position of each area was determined by the latitude and longitude of

the relative province on the map. Besides, a dark background was designed to highlight the static display. Users could browse and compare data between above and below by sliding the mouse wheel.

What needed to be informed here for the visualization 10 was that the dataset was of some inaccuracy, since the data source was shared on open-source platform Github despite an official organization. Thus, the purpose of this visualization was to emphasize the form and method of visualization design. Overall, through a macro perspective, the Chinese Internet industry developed rapidly during the 40 years from 1980 to 2019. Every year, many new Internet companies registered in various provinces across all over the country since 2004, and the registration counting showed a trend of increasing over time.

Visualization.08
word cloud

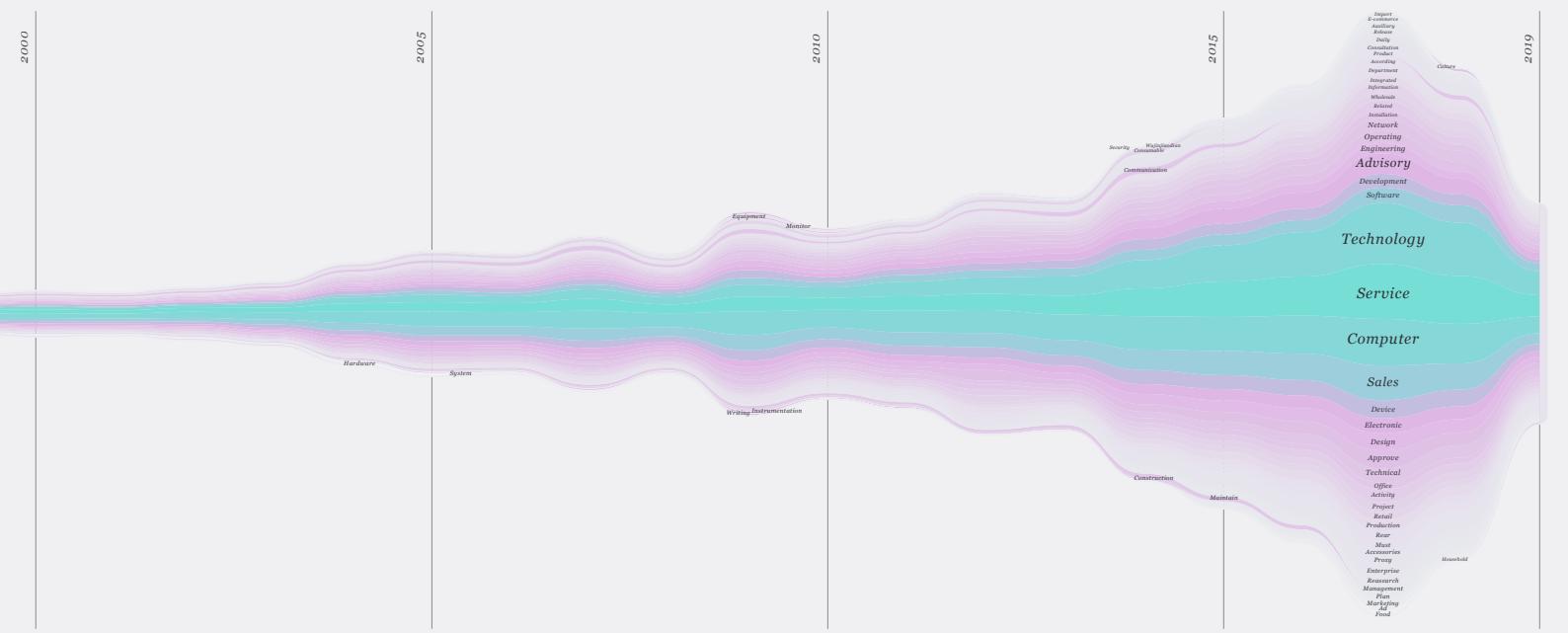
1980

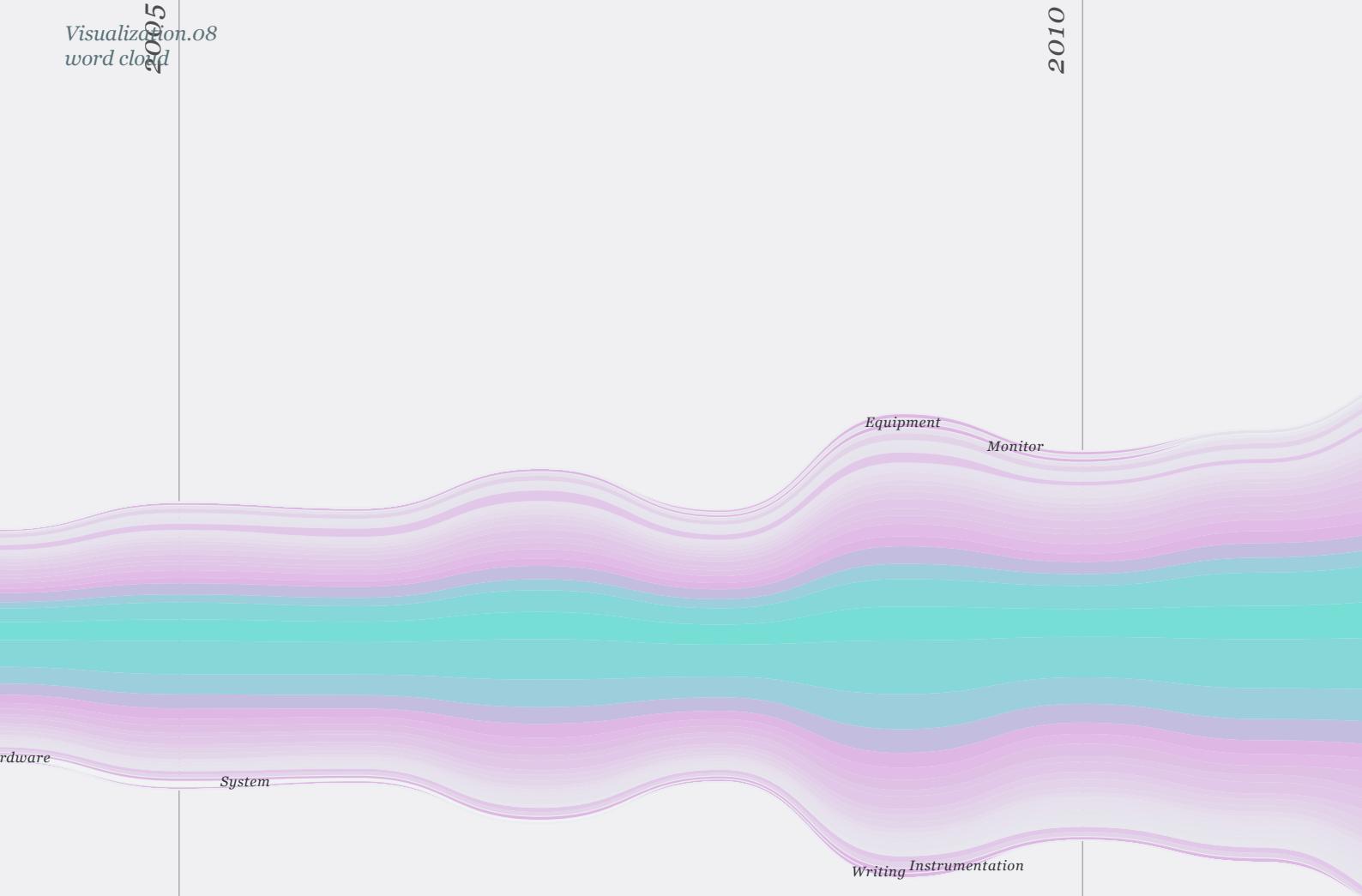
1985

1990

1995

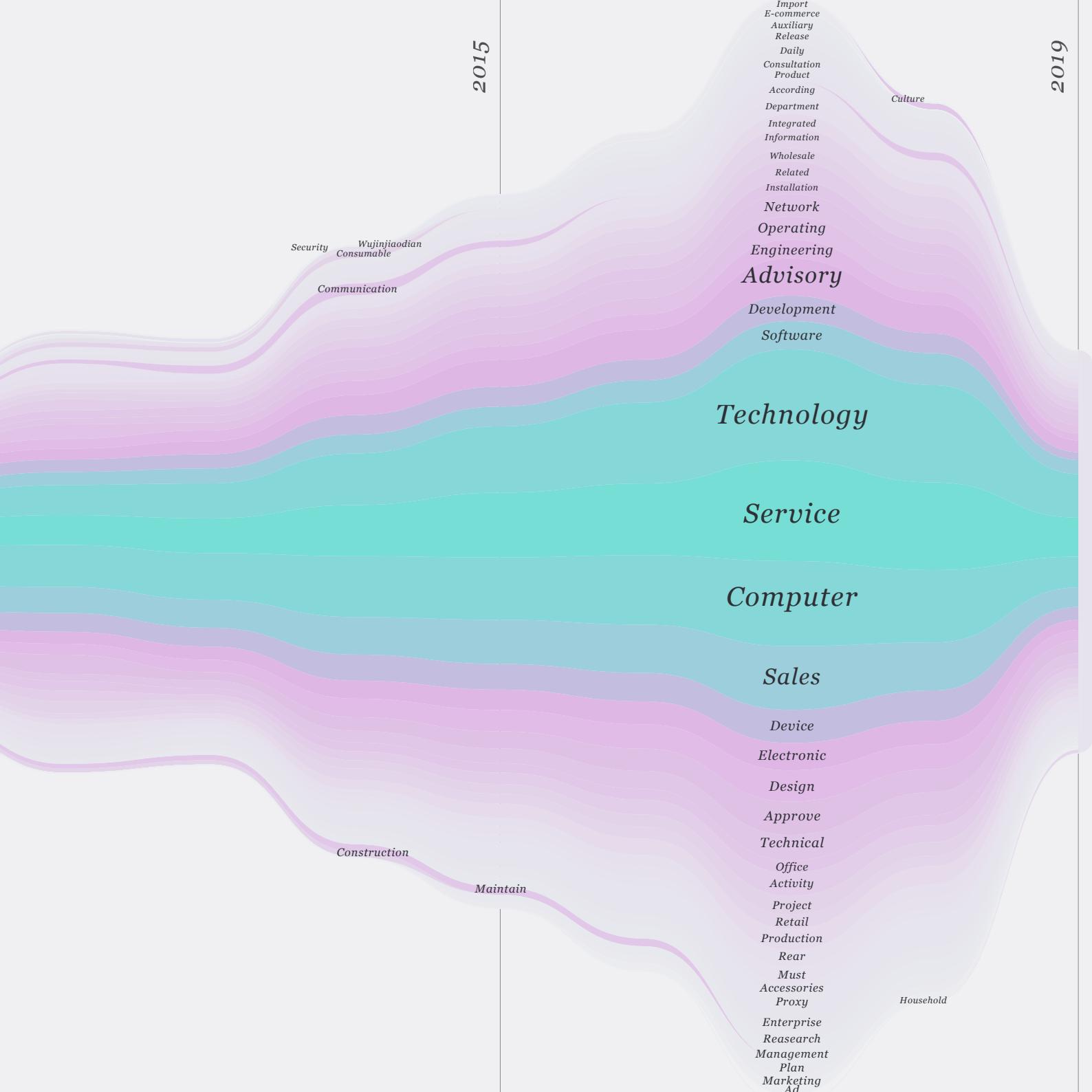
Automation External Mechanical Lease
Manufacturing Building Produce Material
Business

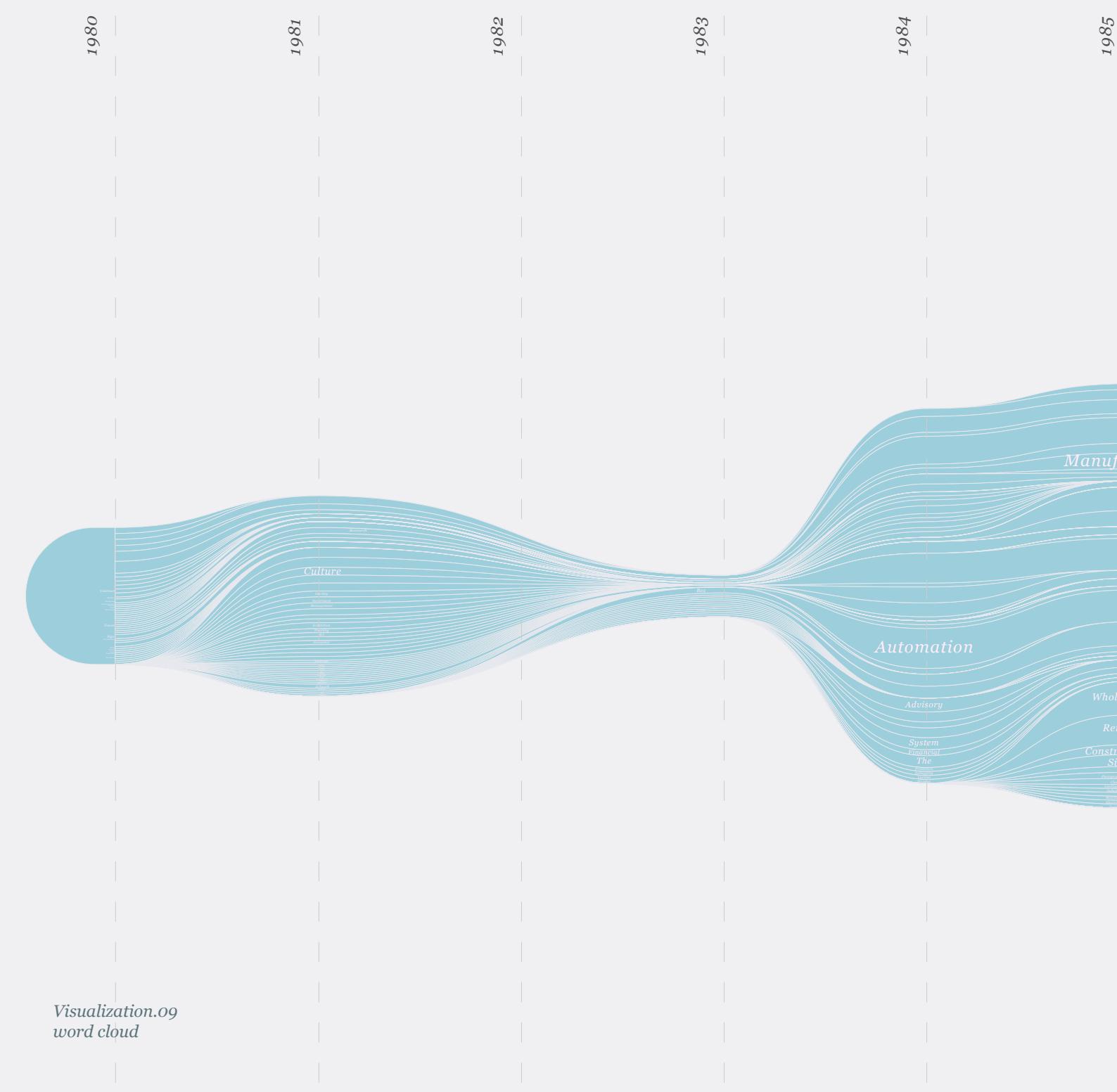


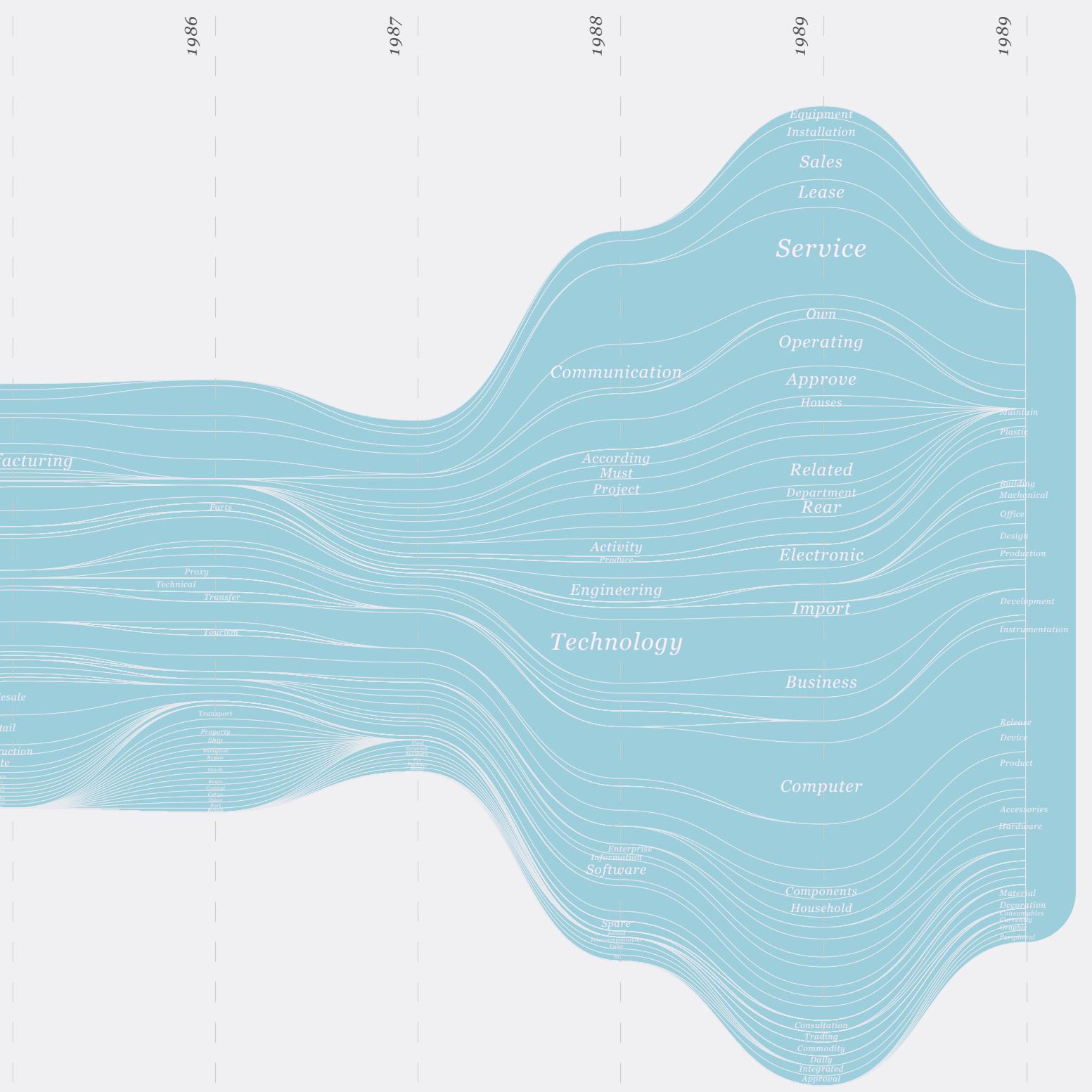


2019

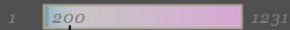
2015







Company Registration Count



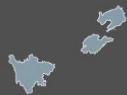
1980

1981

1982

1983

1984



1985

1986

1987

1988

1989



1990

1991

1992

1993

1994



1995

1996

1997

1998

1999



2000



2001



2002



2003



2004



2005



2006



2007



2008



2009



2010



2011



2012



2013



2014



2015



2016



2017



2018



2019



2.4 Data Visualization of Alibaba & Tencent

From the above analysis, it was acknowledged that Alibaba Group and Tencent were the two leading sheep in the Chinese Internet industry, each with a market capitalization of more than 500 billion US dollars. So naturally, to visualize the relationship between the Alibaba or Tencent with the Internet would contribute to the secret of business success. Exactly, Wikipedia was a remarkable link library for the exploration.

How about the China's leading Internet companies of Tencent and Alibaba Group?

Tencent & Alibaba

use Python to implement a crawler to extract the webpage Network topology

Python

assign the root wikipedia page
en.wikipedia.org

analyzing the raw HTML data to extract links
removing the irrelevant links like login, edit,etc.

visit the links on the root page
record the title of each pages, pointer to relation and the distance to the root page by

Breadth-first search (BFS)

Stop at the distance of 2

Tencent.tsv



the file includes
the below elements
> source
> target
> depth

Alibaba.tsv



the file includes
the below elements
> source
> target
> depth

Gephi

ForceAtlas

Gephi

ForceAtlas

Adobe Illustrator

Adobe Illustrator

vis.

Tencent network

vis.

Alibaba network



Shown on protocol 07, the final TSV dataset was obtained by the process of using Python to implement a crawler to extract the webpage Network topology, assigning the root Wikipedia page(Alibaba Group & Tencent), analyzing the raw HTML data to extract links, removing the irrelevant links, using Breadth-first search (BFS) methods to visit the links on the root page, recording the title of each pages, pointing to relation and the distance to the root page, stopping at the distance of 2, and then exporting the Alibaba.tsv file and Tencent.tsv file. The elements in each file were the source, target, and depth. Importing the CSV file into Gephi(Tool), using the Force Atlas layout method could obtain the mesh visualization for Alibaba and Tencent. Since the amount of data was too large, the SVG format file was exported from Gephi, and then put into the Adobe Illustrator for manual deletion and word editing, to display the data better.

As far as the design of data visualization was concerned, static rendering for the massive amount of data became the primary choice. Therefore, the size and color of the corresponding character were manually adjusted according to the number of links, and some unnecessary words were deleted to avoid overlapping. Setting a dark background could highlight the display. The zooming tool was designed for browsing the visualization so that the tiny details could also be observed.

Insights

This visualization 11 was the link topology from Alibaba Group page on Wikipedia. 'Alibaba Group Holding Limited is a Chinese multinational conglomerate holding company specializing in e-commerce, retail, Internet, and technology.' "Alibaba Group".[12] From it, we could quickly obtain the pertinent information by the connected words. For example, the words like Credit Card, Online banking, and Alipay indicated the Alibaba Group was a company specializing in e-commerce and payment. The words like Retail, E-commerce, eBay, Amazon, and Walmart indicated that the Alibaba Group was related to online retailing. In a nutshell, the Alibaba Group scope was more related to business, trade, retail, and online payment. Besides, words like Cloud computing, Linux distribution, and The Pirate Bay stated the Alibaba was also focusing on the Internet technology research and application to empower the middle and small enterprises. There were also someplace words like Malaysia, Iran, Sierra Leone were connected, which might indicate the service plan of Alibaba in some undeveloped regions.

The Pirate Bay

Internet

E-commerce

Retail

Amazon (company)

EBay

Walmart

YG Entertainment

Alibaba

Universal Music Group

Sony Music

NBC

General Electric

Fox

Warner Bros.

MTV

USA

Hulu

ESPN

Turner

CNN

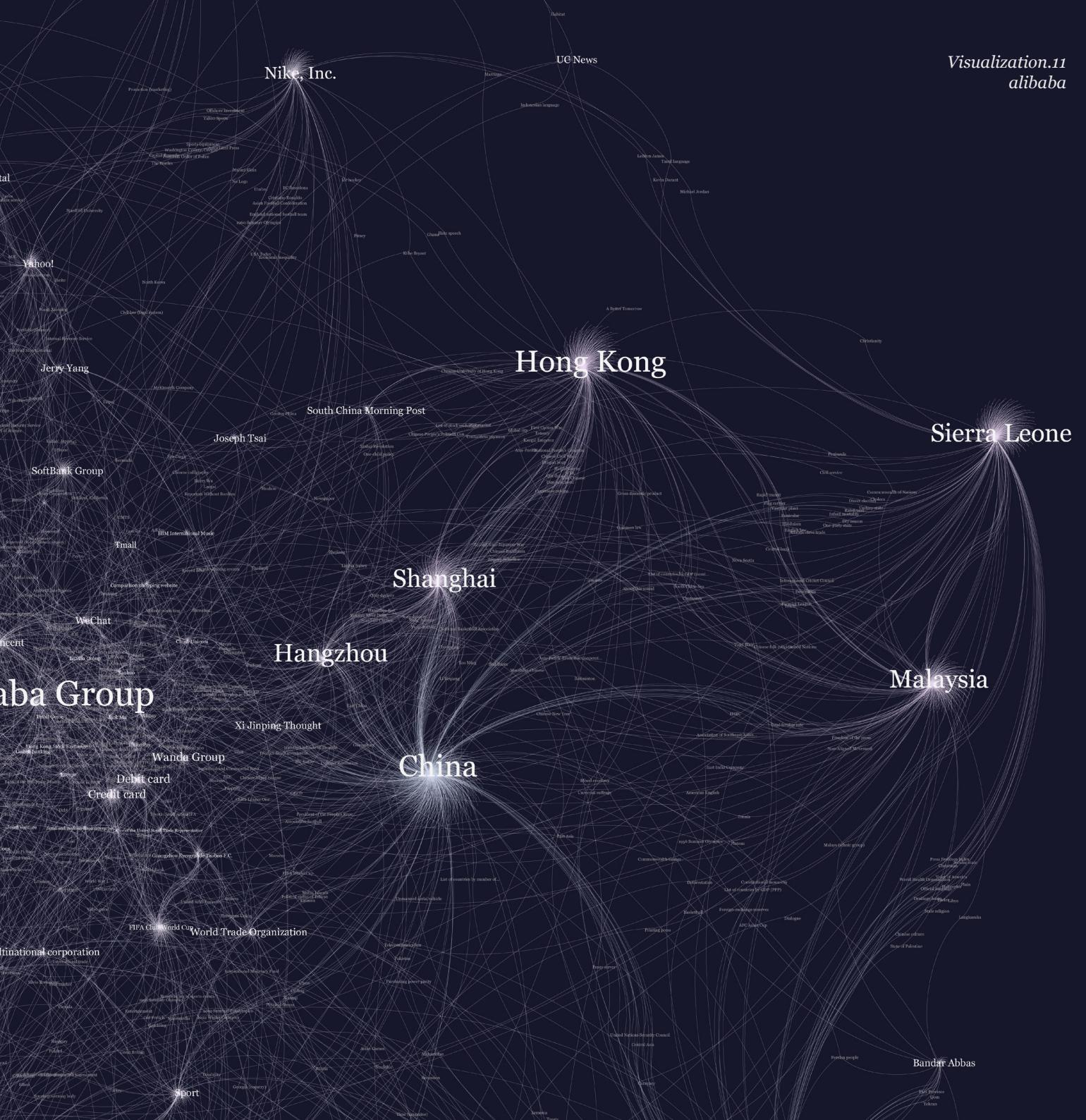
Fox News

Fox Sports

FX

ABC

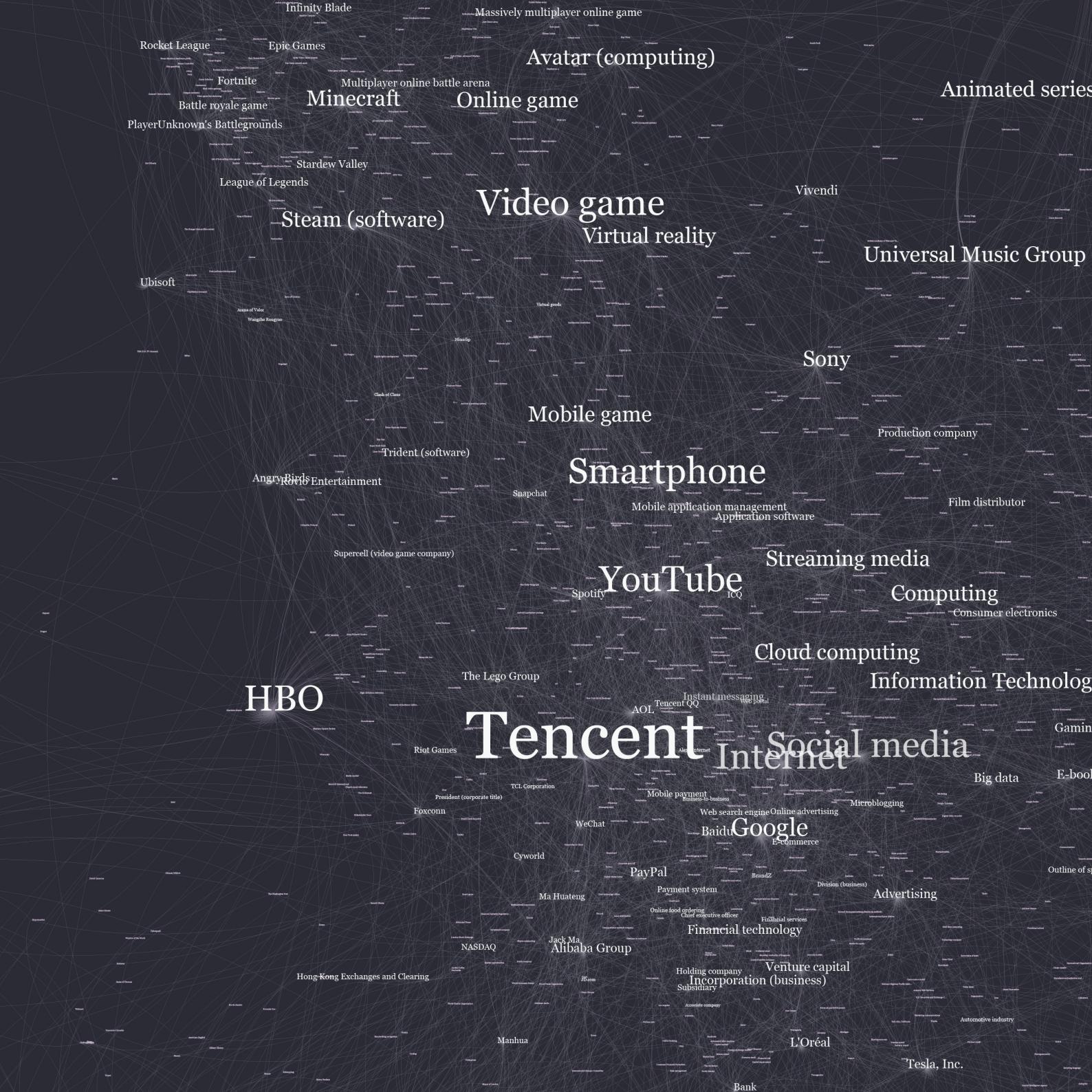
WB

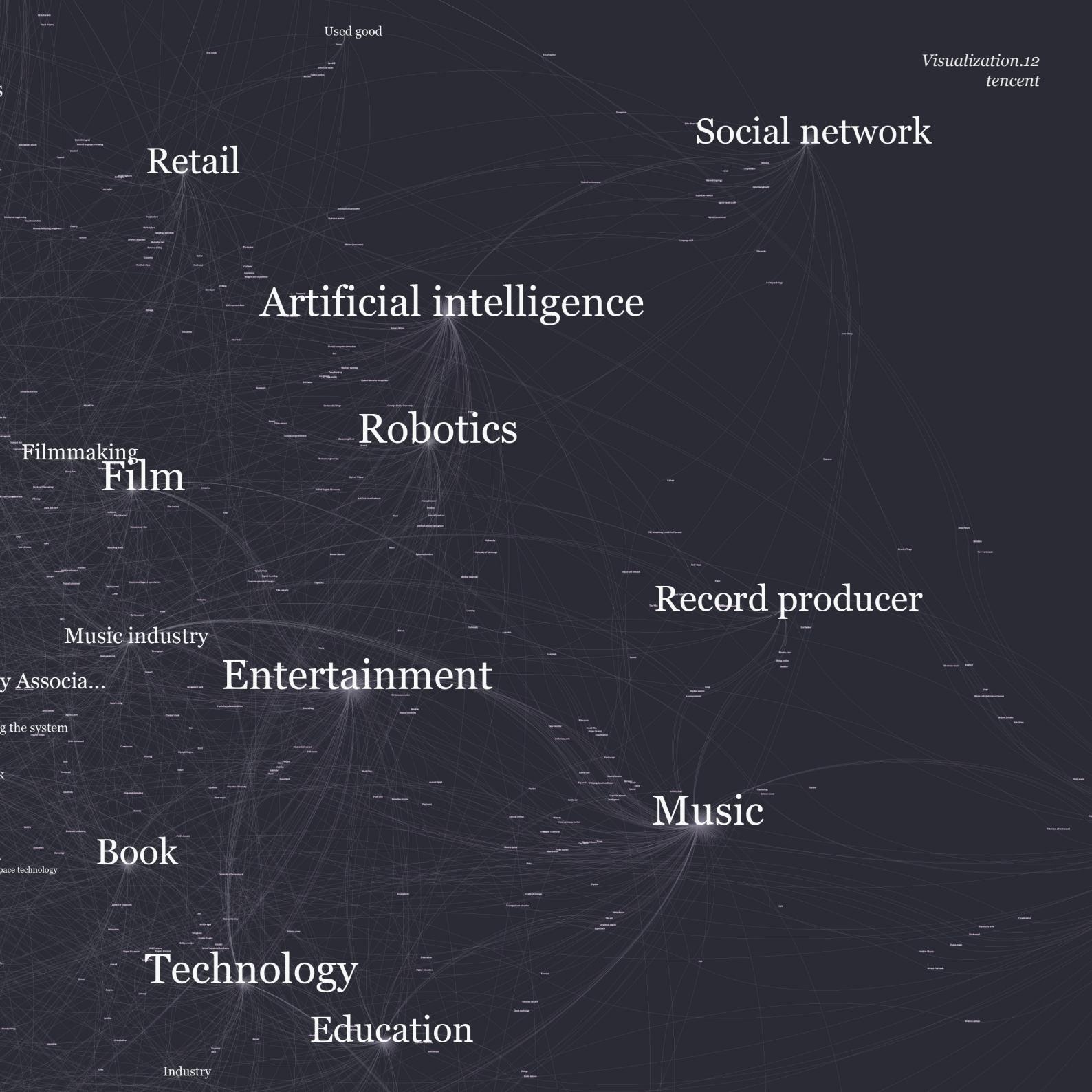


Insights

This data visualization 12 showed the web links from Tencent's wiki page. Looking at the neighbors of Tencent, the keywords was Music, Entertainment, Film, HBO, Book, YouTube, Online Game, Mobile game, QQ, and Wechat, through which, we could roughly sketch the business empire of Tencent, the information market (Instant Message and Social) and the entertainment market including (Game, Video, and Music). In a nutshell, Tencent was a company focusing on providing entertainment and communication tools for the market. At the same time, we could see keywords like Artificial Intelligence and Robotics connected to Tencent, which highlighted the future technology transformation.







[otify](#)  [ICQ](#)

Computing

Consumer electronics

Cloud computing

Information Technology

Coming soon

Environ Biol Fish

Carrefour

HBO

Tencent Internet Social media

Social media

Information

Coming soon

Xi Jinping

Simplified Chinese characters

People's Bank of China

Pinyin

Chinese language Music of China

Chengdu

Shanghai

Shenzhen

Guangzhou

Beijing

San Fr

Asia

North Amer

Record producer

Visualization.12
tencent

Music industry

Entertainment

Associa...

the system

e technology

Book

Technology

Education

Industry

rial vehicle

Health care

Real estate

Cayman Islands

ica

Natural resource management

Agriculture

francisco

Renewable energy

Music

Music of Japan

4. Conclusion

Through the above data visualization analysis, from a macro perspective, the Internet industry in China and the United States were in the leading positions in the world. While the leading Internet companies would have a particular monopoly position nowadays, the leading enterprises on the vertical field still had the opportunity to develop. Besides, E-commerce, social, entertainment, and tools were the main topics about Internet companies today. Of course, there were still many deficiencies in the project. For example, the data visualization interactive design was relatively weak, and collecting and processing dataset existed some inaccurate attribute. More tools for data visualization, such as D3.js, needed to be further explored to generate data visualization by an algorithm rather than manually. Besides, due to the static datasets, the effectiveness of the project was limited to the past time, lacking an exploration for the future.

check the digital interactive datavis version on the website

<http://cha-net-vis.yifangbao.me>

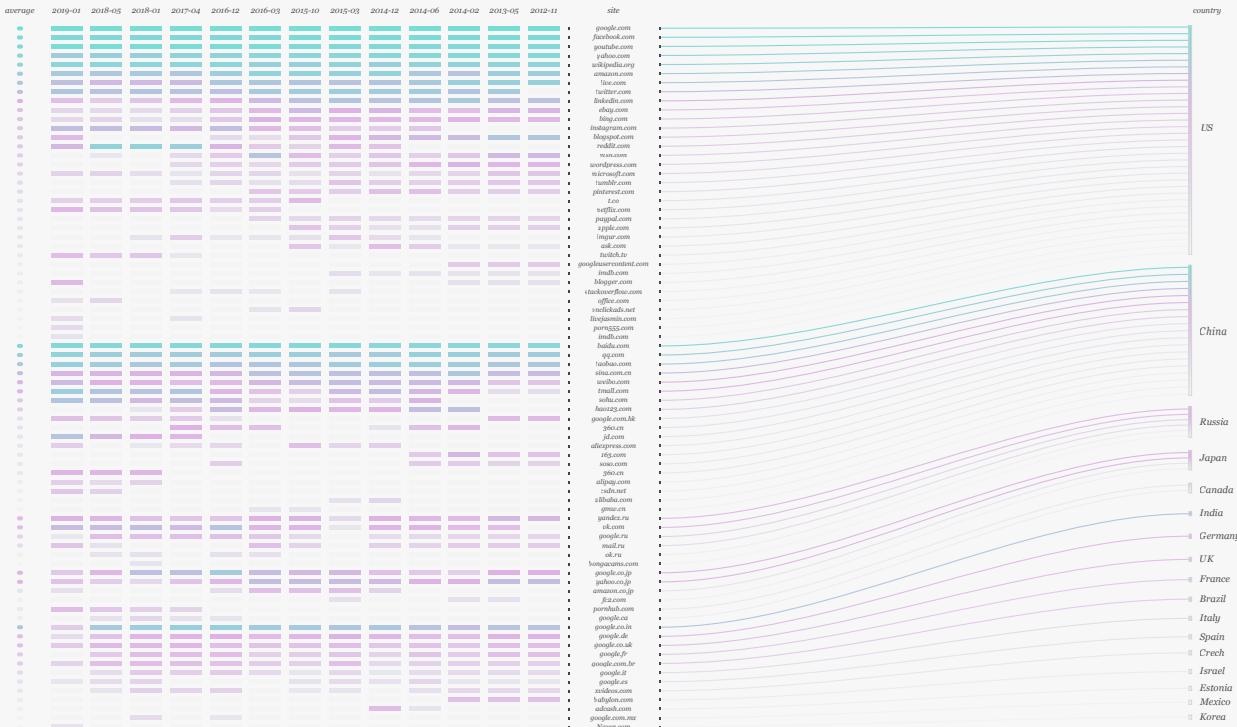
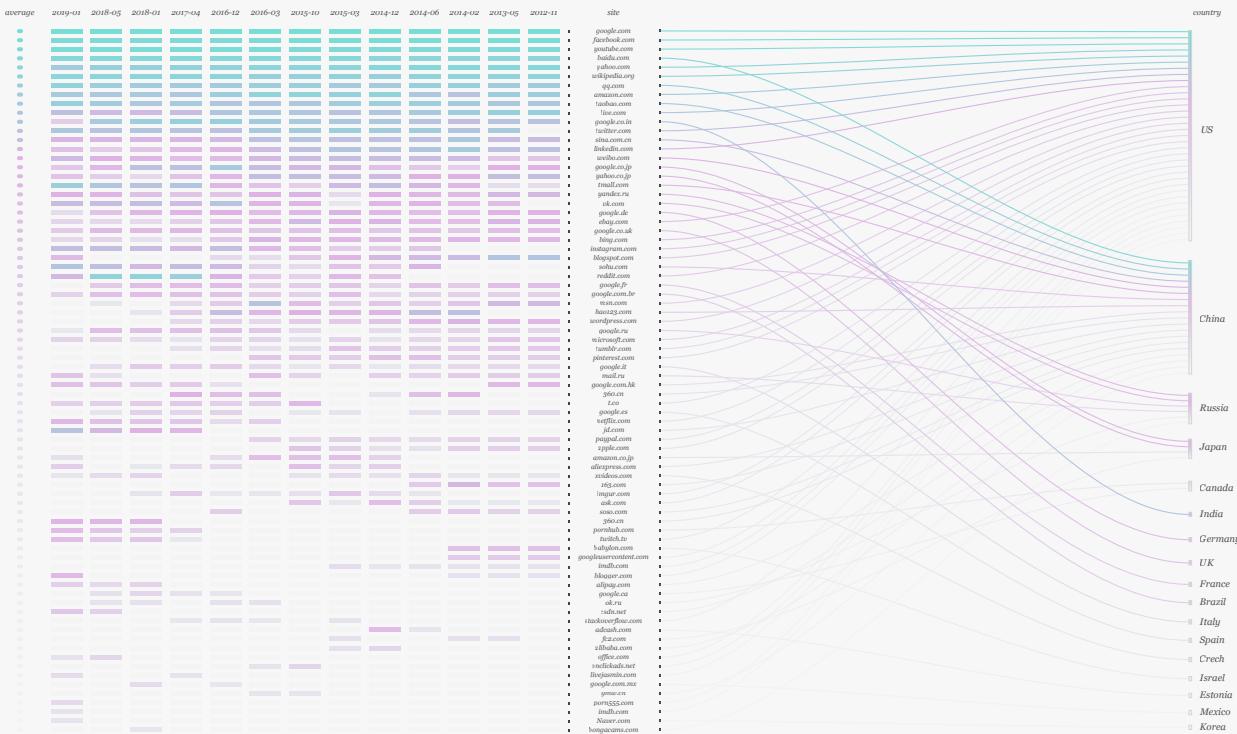
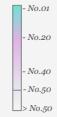
References:

1. Meeker, Mary, and Liang Wu. "Internet trends 2018." (2018).
2. 曾凡 , et al. " 互联网行业公司估值重要性及影响因素初探 ." 上海节能 10 (2018): 11.
3. 李刚 . “中国互联网企业上市简史 .” 互联网经济 , Apr. 2018, pp. 84–95.
4. “List of Most Popular Websites.” Wikipedia, Wikimedia Foundation, 23 May 2019, [en.wikipedia.org/wiki/List_of_most_popular_websites.\]](https://en.wikipedia.org/wiki/List_of_most_popular_websites.)
5. D3.Js - Data-Driven Documents". D3js.Org, 2019, <https://d3js.org/>. Accessed 15 June 2019.
6. Manovich, L. (2011). What is visualization?. Visual Studies, 26(1), 36-49.]
- 7.Tufte, Edward R., et al. "Visual explanations: images and quantities, evidence and narrative." (1998) , count 1]
8. "Unicorn (Finance)". En.Wikipedia.Org, 2019, [https://en.wikipedia.org/wiki/Unicorn_\(finance\)](https://en.wikipedia.org/wiki/Unicorn_(finance)). Accessed 5 June 2019.
9. “中国独角兽公司俱乐部 -IT 桔子数据库 ”. Itjuzi.Com, 2019, <https://www.itjuzi.com/unicorn>. Accessed 5 June 2019.
10. “企查查 - 工商信息查询 - 公司企业注册信息查询 - 全国企业信用信息公示系统 ”. Qichacha.Com, 2019, <https://www.qichacha.com/>. Accessed 5 June 2019.]
11. Liu, Wen. Enterprise-Registration-Data-Of-Chinese-Mainland. 2019, <https://github.com/imhuster/Enterprise-Registration-Data-of-Chinese-Mainland>. Accessed 1 June 2019.
12. En.Wikipedia.Org, 2019, https://en.wikipedia.org/wiki/Alibaba_Group. Accessed 5 June 2019.

From 2012-2019

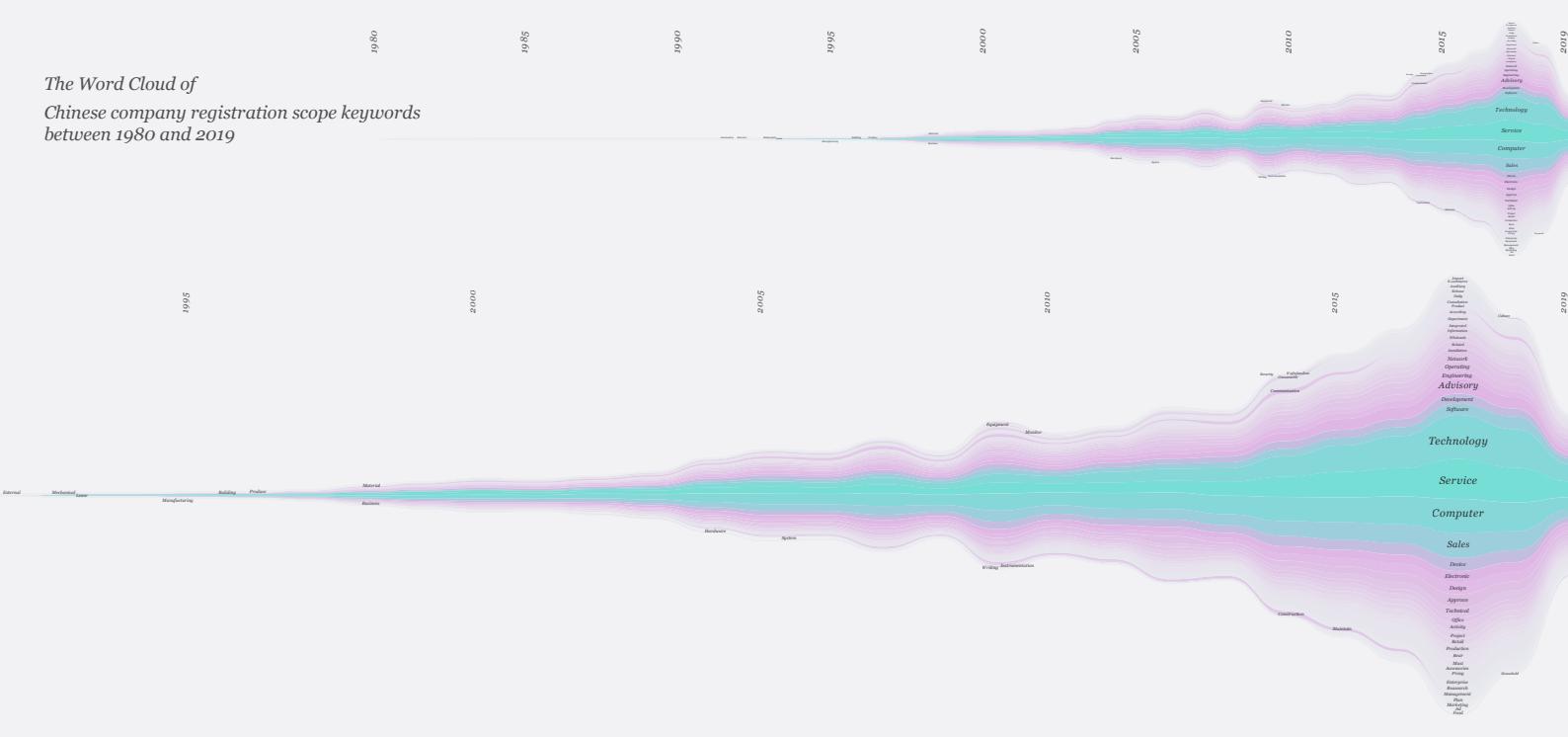
Top 50 Websites of Alexa Ranking

Alexa ranking



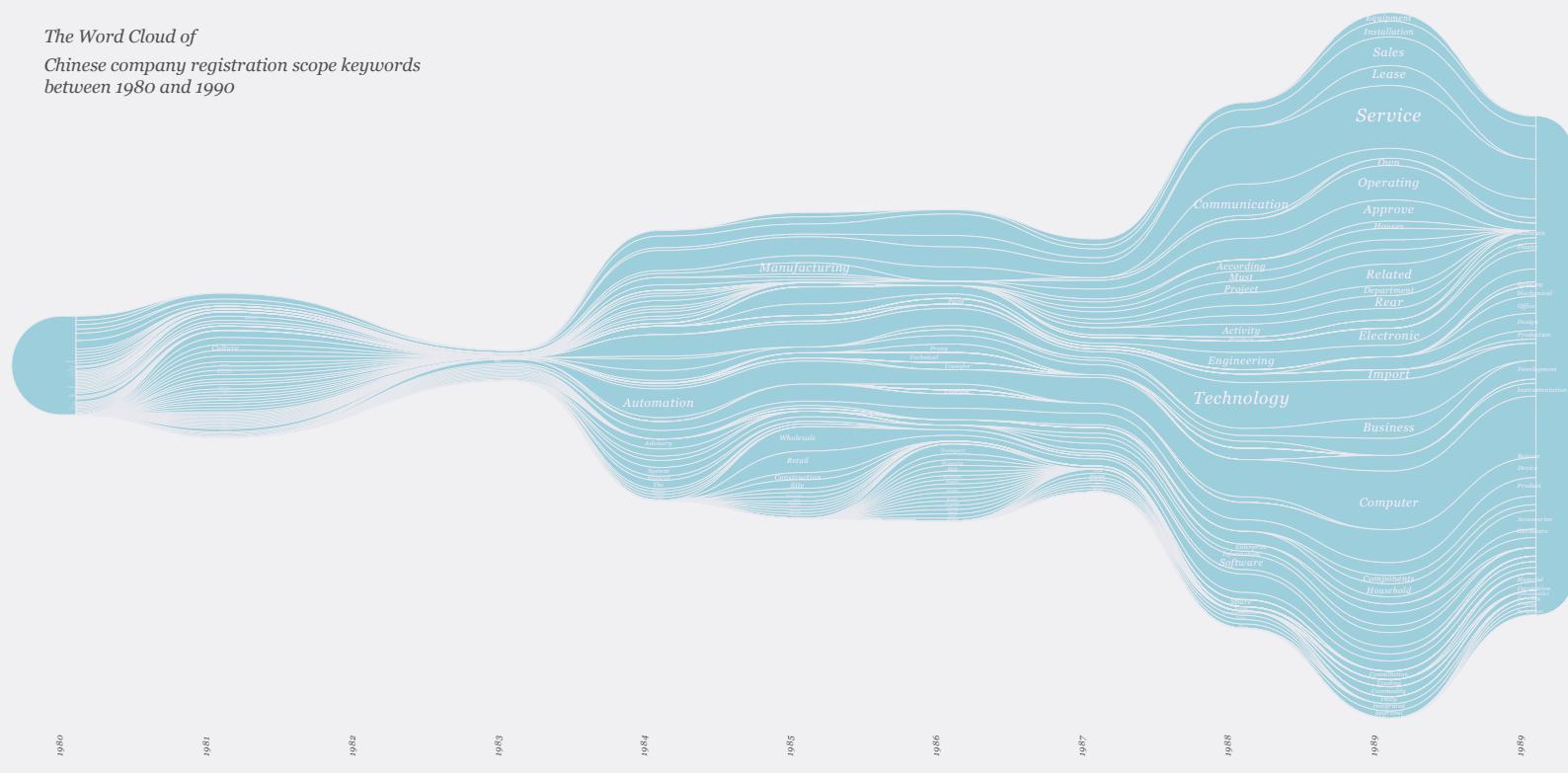
The Word Cloud of

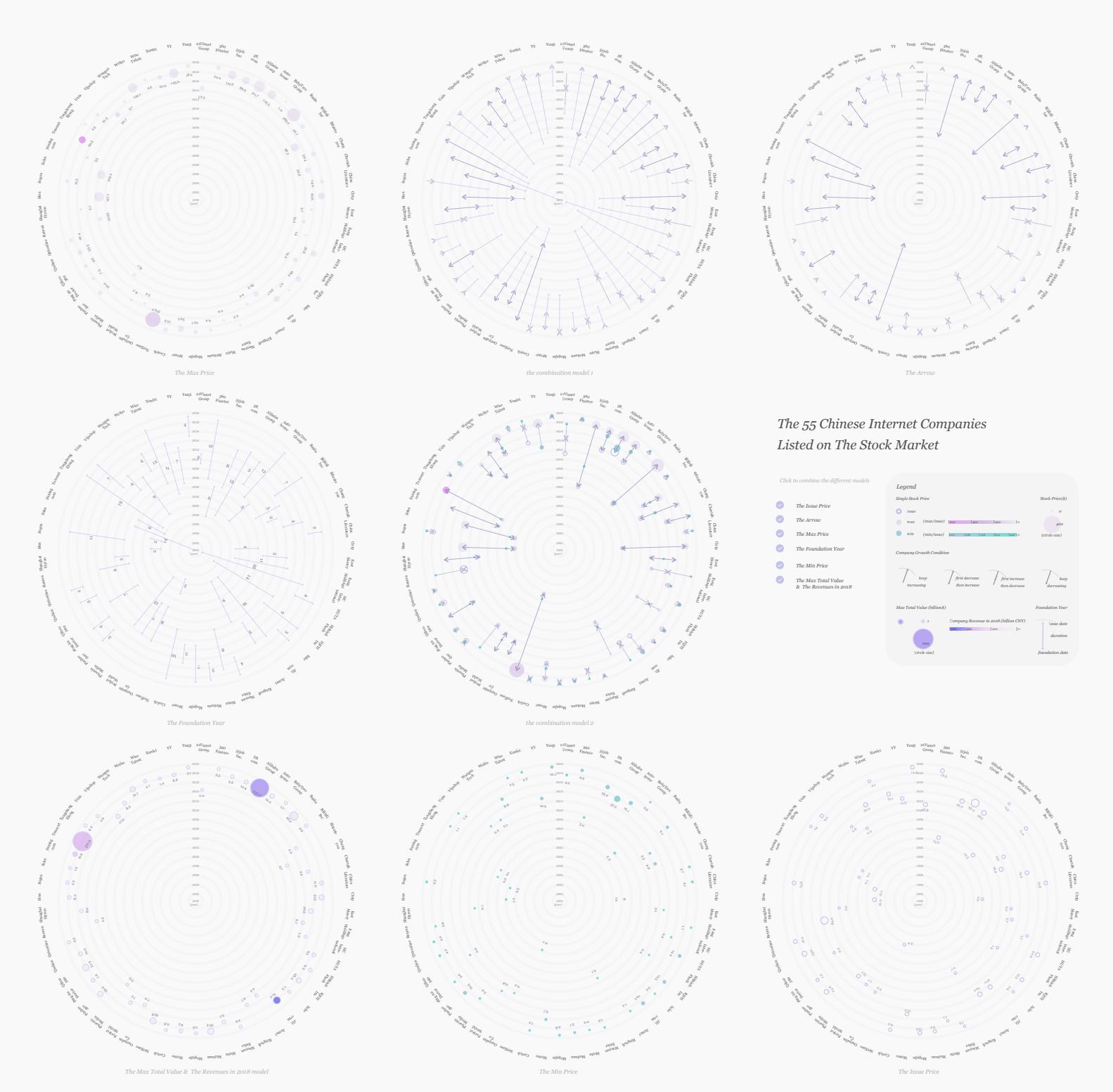
Chinese company registration scope keywords
between 1980 and 2019

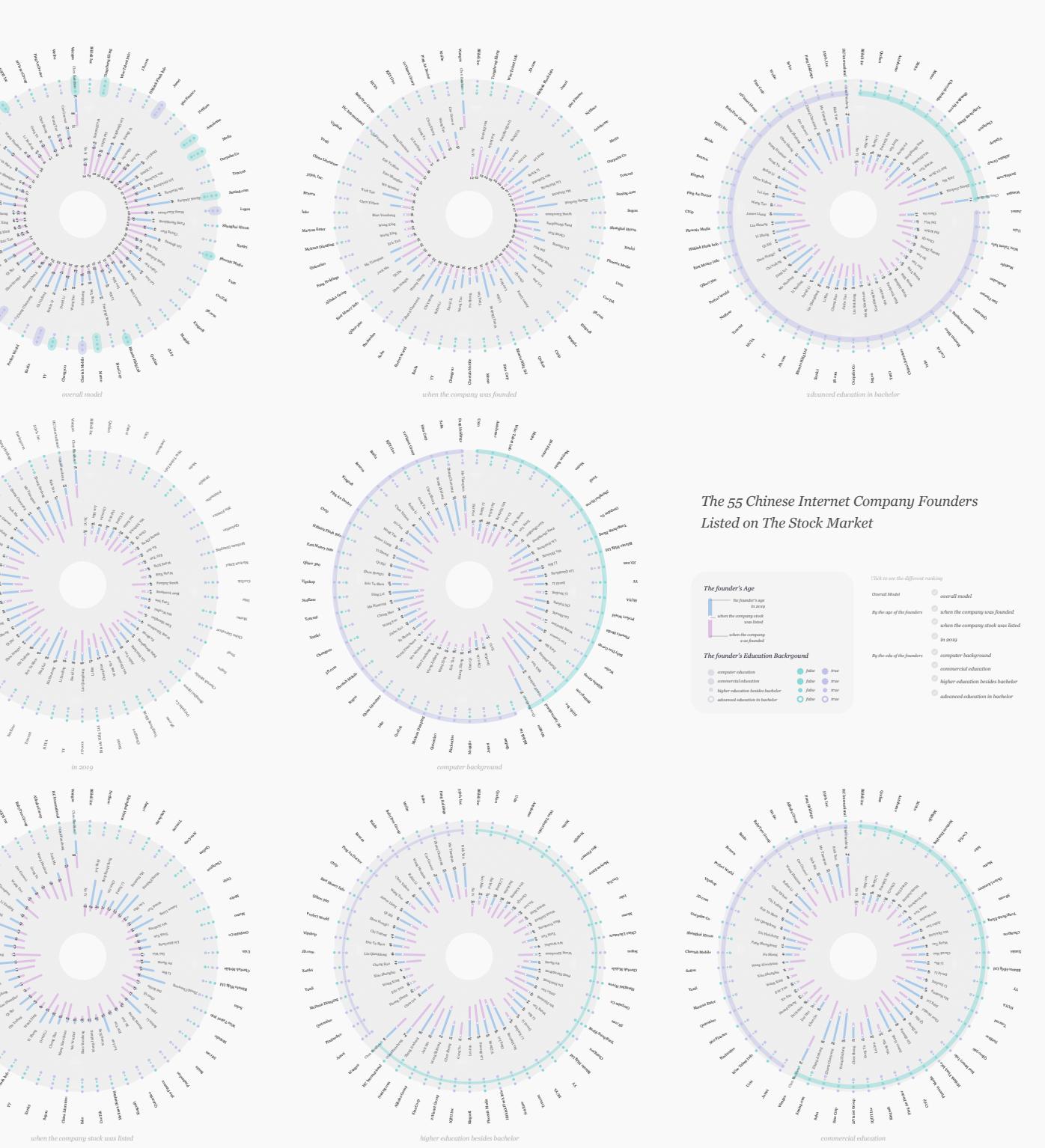


The Word Cloud of

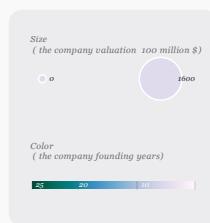
Chinese company registration scope keywords
between 1980 and 1990







The Unicorns Of Chinese Internet Companies



(Model 01)



(Model 02)

1980

1981

1982

1983

1984

1985

1986

1987

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997

1998

1999

2000

2001

2002

2003

2004

2005

2006

2007

2008

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

2019

*The Registration Map of
Chinese Internet registration companies
between 1980 and 2019*

Company Registration Count

1 200 1231

Comic book

Boston Consulting Group

RedGiant
Dragon Fighter Online

Person Interview
Private Development
Point Center
Chess: Skyline
Rocket League
EVE Online
Battle chess game
PlayerUnknown's Battlegrounds
Steam (software)

Avatar (computing)
Minecraft
Online game
Virtual reality

Warner Music Group

Animated series
Vivendi
Universal Music Group

Retail

Social network

Artificial intelligence

Robotics

Record producer

Music of Japan

Mobile game
Smartphone
YouTube
Streaming media
Cloud computing
Information Technology Associa...

Entertainment

Music

Book
Technology
Education

Unmanned aerial vehicle

Health care

Cayman Islands

Natural resource management

Agriculture

Renewable energy

Malaysia

HBO

Tencent Internet

Credit score

Naspers

Carrefour

Courier

Xi Jinping

Chengdu

Shenzhen

Guangzhou

Beijing

Three Kingdoms

Wuhu

Ping An Insurance

Simplified Chinese characters

People's Bank of China

Pinyin

Nanshan District, Shenzhen

China

Asia

North America

San Francisco

Malaysia

Cayman Islands

Real estate

Natural resource management

Agriculture

Renewable energy

Health care

Technology

Education

Book

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

Artificial intelligence

Robotics

Film

Entertainment

Music

Record producer

Social network

