Q1: Word Clouds

1. Run code in R .

Here is the plot.

Figure 1: Example

2. Create new sentence.

I quoted the definition of AI from Wikipedia as an experimental sentence.

“Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by animals and humans. AI research has been defined as the field of study of intelligent agents, which refers to any system that perceives its environment and takes actions that maximize its chance of achieving its goals.”

Here is the plot.

Figure 2: AI

Only one word “intelligence” showed in the plot.

Repeat some words:

“Artificial intelligence (AI) is intelligence demonstrated demonstrated demonstrated by machines machines machines machines machines, as opposed to the natural intelligence displayed by animals and humans. AI AI AI AI AI AI AI research research research research has been defined as the field of study of intelligent agents, which refers to any system that perceives perceives perceives perceives perceives perceives its environment and takes actions that maximize its chance of achieving its goals.”

Here is the plot.

Figure 3: Repeated words

Let min.freq=1, here is the Plot.

Figure 4: min_freq=1

By artificially repeating the frequency of words, the plot shows more words whose frequency is more than there or more. We can see this in Figure 3. Then by reducing the minimum frequency limit, which can be seen in Figure 4, it makes the presentation similar to the first example (Figure 1).

Q2: Google Ngram Viewer

a. “Mark Keane”

The Figure shows four peaks 1820, 1955, 1970, 1975, and 2002. The first one was mentioned in a committee or court report during the 19th century. The second one, after 1955, was mainly derived from U.S. congressional documents. The third one in the 1970s was an executive director of the International City Management Association whose name was Mark Keane, and he was often covered in U.S. Congressional documents. The last one in the near 2000s was Professors Mark Keane at UCD. His publications on computer science and other aspects were published at that time.

b. "Jiahao"

I have a Chinese name. It first appeared mainly in officials mentioned in government documents in China during the 1980s. More authors and companies have published books and reports under this name. after 1990. And when I was a child, I learned the origin of my name from my mother I learned the origin of my name from my mother. That is the year 2000, and many people would name their boys Jiahao during this period due to the Chinese culture. As to the 21st century, the name appears more often in various forms.

c. "Machine Learning"

I pick the word "Machine Learning." I choose the word "Machine Learning." Unexpectedly, I noticed that the first appeared in the mid-19th century. Nevertheless, it was retrieved with the wrong form in the literature, which is not in its present-day meaning. Like "It has no authority or value except in so far as it is a true statement of facts respecting the machine. Learning this law, she learns to operate her machine successfully." Since 1960, there has been a large amount of literature and books about the study of machine learning. Due to the development of computer technology in that period, researching it has made it possible to benefit all areas of life, which is even more essential in modern life.

d. "Artificial intelligence"

I use "Artificial intelligence" to do smoothing. Here are the different figures for various smoothing. Changing the value from 0 to 3, we can easily find its peak in Figure 8. However, after changing the value from 3 to 6, the curve does not seem to shift much.

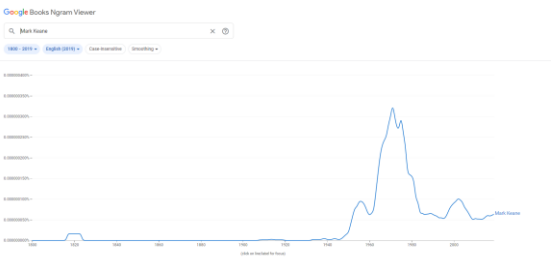

Figure 5: "Mark Keane"

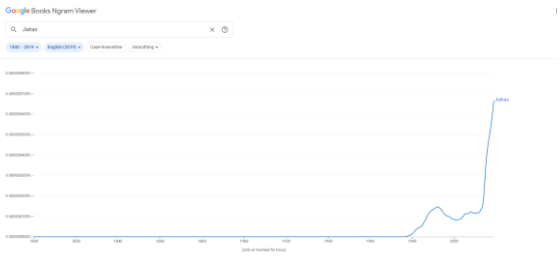

Figure 6: "Jiahao"

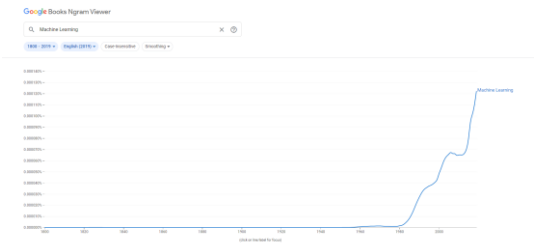

Figure 7: "Machine Learning"

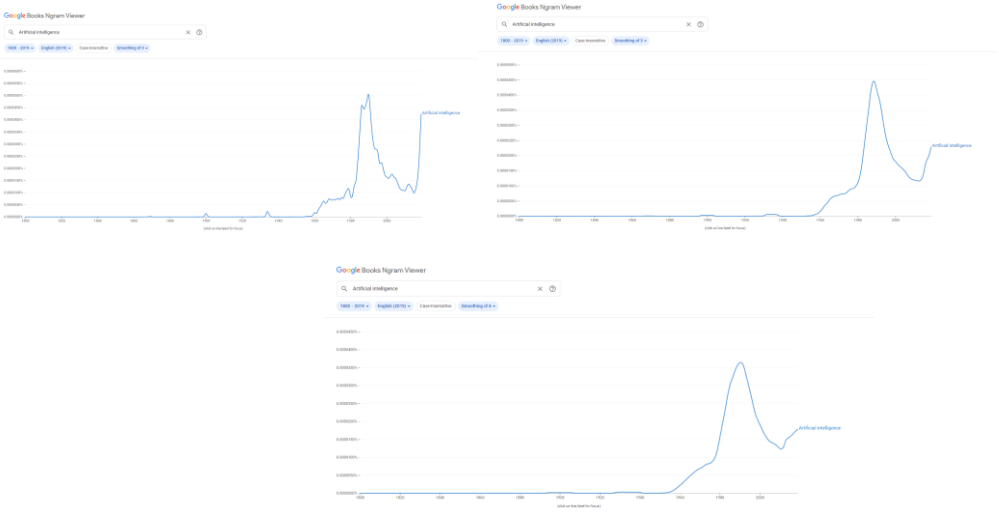

Figure 8: “Artificial intelligence”

e. “Dota”, “League of Legends” and “Dota 2”

I choose the terms of these three most popular MOBA games, which all are multiplayer real-time strategy games. Dota began in 2003, and As the number of players continued to grow, from this time onwards, many books on gameplay and strategy began to appear. Dota 2 is the sequel to it, which was published in 2013. However, they all have relatively smooth rising curves. Furthermore, League of Legends was developed by borrowing from Dota. Since it is much simpler than Dota to play, it has attracted more people to try it. Since its release in 2009, it has accumulated more players, and more strategy books and videos on League of Legends have appeared on the internet. Nowadays, the related gaming industry has been spawned around these games. Moreover, its search index also exceeds the first two and is still on the rise.

f. “zoom_NOUN”, “zoom_VERB”

I choose to compare the word “zoom” in different wordings. Throughout a period, the noun form of zoom is cited more frequently than its verb form. Nevertheless, zoom should also be associated with the company zoom meeting, and in recent years I think the actual frequency curve should be very steep compared to the verb form of zoom rather than being flat in the Figure 10.

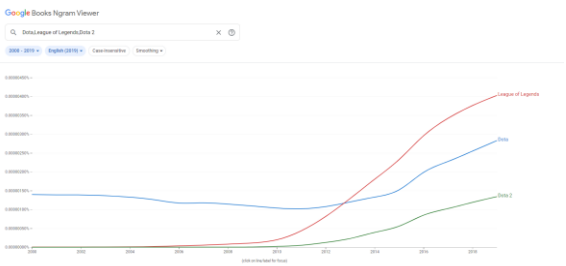

Figure 9: Moba games

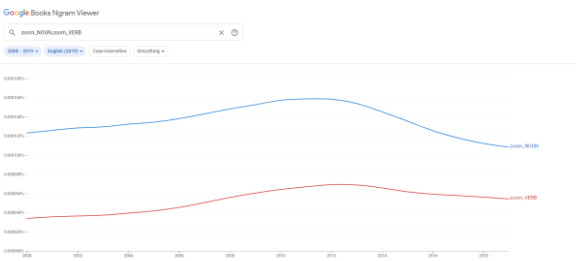

Figure 10: “Zoom”

g. “Computer”

I think it was the invention of the computer. In 1946, the first computer created in the USA was used for military ballistic calculations, which took human arithmetic to another level.

Computers could simulate more scenarios to the point where nuclear testing is done today. The internet, a term that came into being as information dissemination continued to evolve in computers, is even more indispensable in today's daily life. In a sense, it was used to connect the whole world. This has created an era of rapid information expansion and high sharing. Furthermore, people now prefer to read on the screen than in the traditional reading mode. This should probably be the image of what humanity has not envisioned in 500 years.

Q3: Normalisation

I randomly generate ten words list. Furthermore, the frequency of all words is called by a function named RANDOMBETWEEN in Excel, ranging from 0 to 2000. Then I do normalization of the words for four years. The tables are shown in the figure as follows.

RANDBETWEEN(0,2000)						By_Year Normalization					Overall Normalization				
Words	2010	2011	2012	2013	2014	Y_2010	Y_2011	Y_2012	Y_2013	Y_2014	O_2010	O_2011	O_2012	O_2013	O_2014
sun	187	678	257	1814	1819	2.19%	7.51%	2.13%	15.29%	17.73%	0.36%	1.31%	0.50%	3.50%	3.51%
night	1700	598	870	1689	698	19.93%	6.62%	7.20%	14.24%	6.80%	3.28%	1.16%	1.68%	3.26%	1.35%
rain	819	334	1930	1513	1538	9.60%	3.70%	15.98%	12.75%	14.99%	1.58%	0.65%	3.73%	2.92%	2.97%
way	1026	1002	1915	557	1366	12.03%	11.10%	15.85%	4.69%	13.31%	1.98%	1.94%	3.70%	1.08%	2.64%
only	1040	601	1515	729	188	12.20%	6.66%	12.54%	6.14%	1.83%	2.01%	1.16%	2.93%	1.41%	0.36%
matter	259	1642	914	1579	231	3.04%	18.19%	7.57%	13.31%	2.25%	0.50%	3.17%	1.77%	3.05%	0.45%
label	1241	484	1578	646	1696	14.55%	5.36%	13.06%	5.44%	16.53%	2.40%	0.94%	3.05%	1.25%	3.28%
define	1188	1133	271	1222	1807	13.93%	12.55%	2.24%	10.30%	17.61%	2.30%	2.19%	0.52%	2.36%	3.49%
vary	751	1099	899	364	494	8.81%	12.17%	7.44%	3.07%	4.81%	1.45%	2.12%	1.74%	0.70%	0.95%
start	317	1458	1930	1752	425	3.72%	16.15%	15.98%	14.77%	4.14%	0.61%	2.82%	3.73%	3.38%	0.82%
SUM	8528	9029	12079	11865	10262										

Figure 11: Word list, By_Year and Overall Normalization

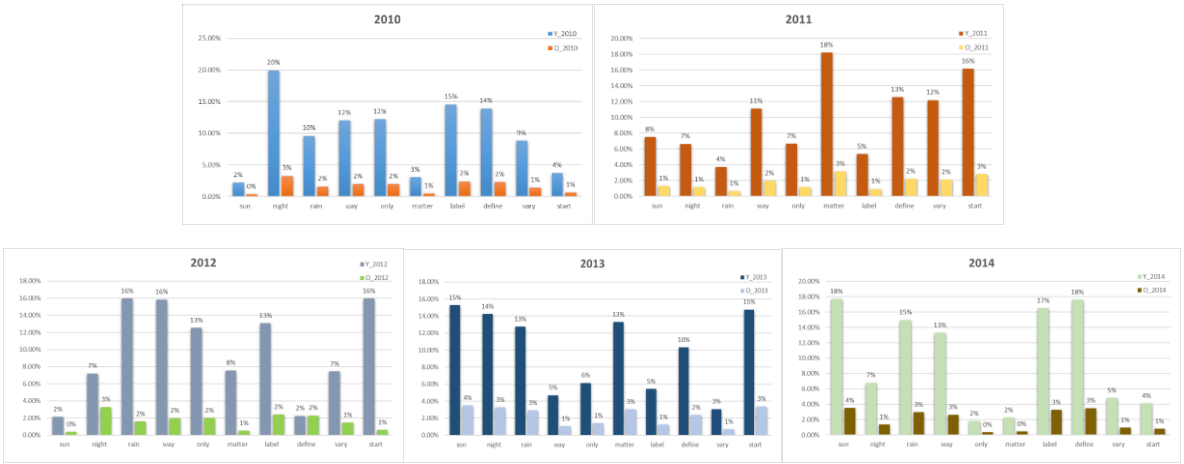

Figure 12: Comparison charts

In these figures, it is clear that they show the variability between Overall Normalization and By-year Normalization. When we want to know how often a word appears in a year, it is evident that we can know which words are the most frequent by By-year Normalization. However, there is slight variation overall when looking at a word over four years.