

LEARN TO BE DATA DRIVEN

CHARLIE CHEN

APR 26, 2018 @ HKBU

ABOUT THE SPEAKER

My name is Charlie. I am a website/backend programmer who is also interested in Data Science.

- wechat: xhcsster
- charlie.chenshen@gmail.com
- github.com/carolusian

A SIMPLE SURVEY

- Do you know the job title: Data Scientist?
- Do you think Data Scientist sounds cool?
- Do you accept if you are offered be a Data Scientist?

DATA XXX: NOT A GOOD JOB TITLE

I guess you are facing the following realities:

- You are going to **graduate**;
- You may need to **find a job** after graduation;
- Some of you may be interested in **data science related positions**, like data scientist, data engineer, data analyst, data news editor or data visualisation specialist, etc;

However, in my opinion, we shall not think data xxx as good job titles.

DATA XXX: NOT A GOOD JOB TITLE

For most small businesses, a structure with dedicated data team may be too luxury:

- the organisation may **lack enough data**;
- available time or labour in the team can **NOT** be fully utilised;
- or, the work being done are **NOT** directly related to its business.
- this team may **not bring enough value** comparing resources it costs;

WHAT ARE THE SOLUTIONS?

You can find job in big companies where it has sufficient data for you to torture and make it confess;

or, do not think data science as a job:

- learn “driving” does not necessarily mean you have to work as driver.
- **change your mindset;**
- let it be one **skillset** in your work and live;
- learn to be **data-driven**, and try to cultivate the culture in your organisation.

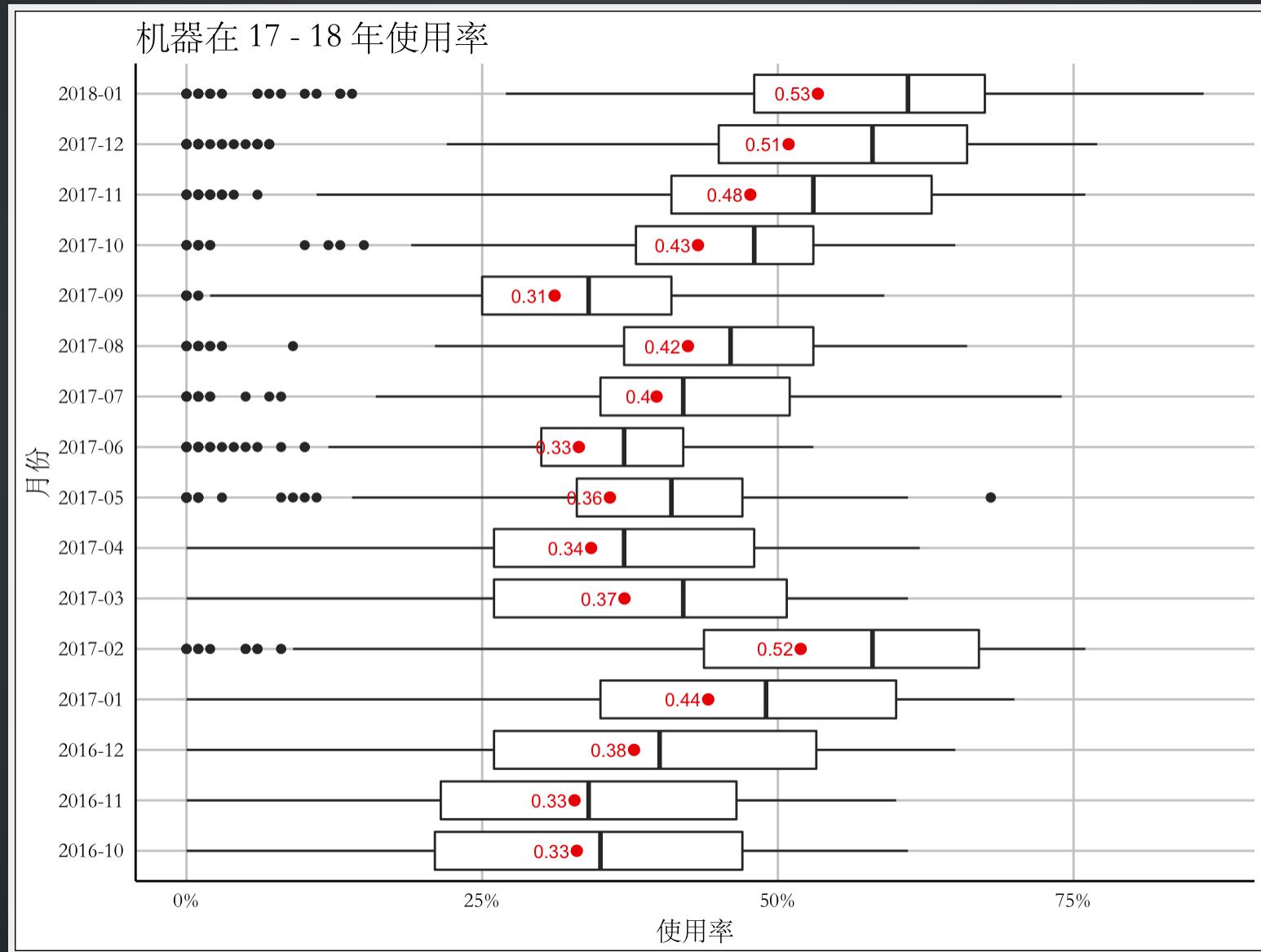
CASE STUDIES

CASE STUDY 1: INTERNET CAFE

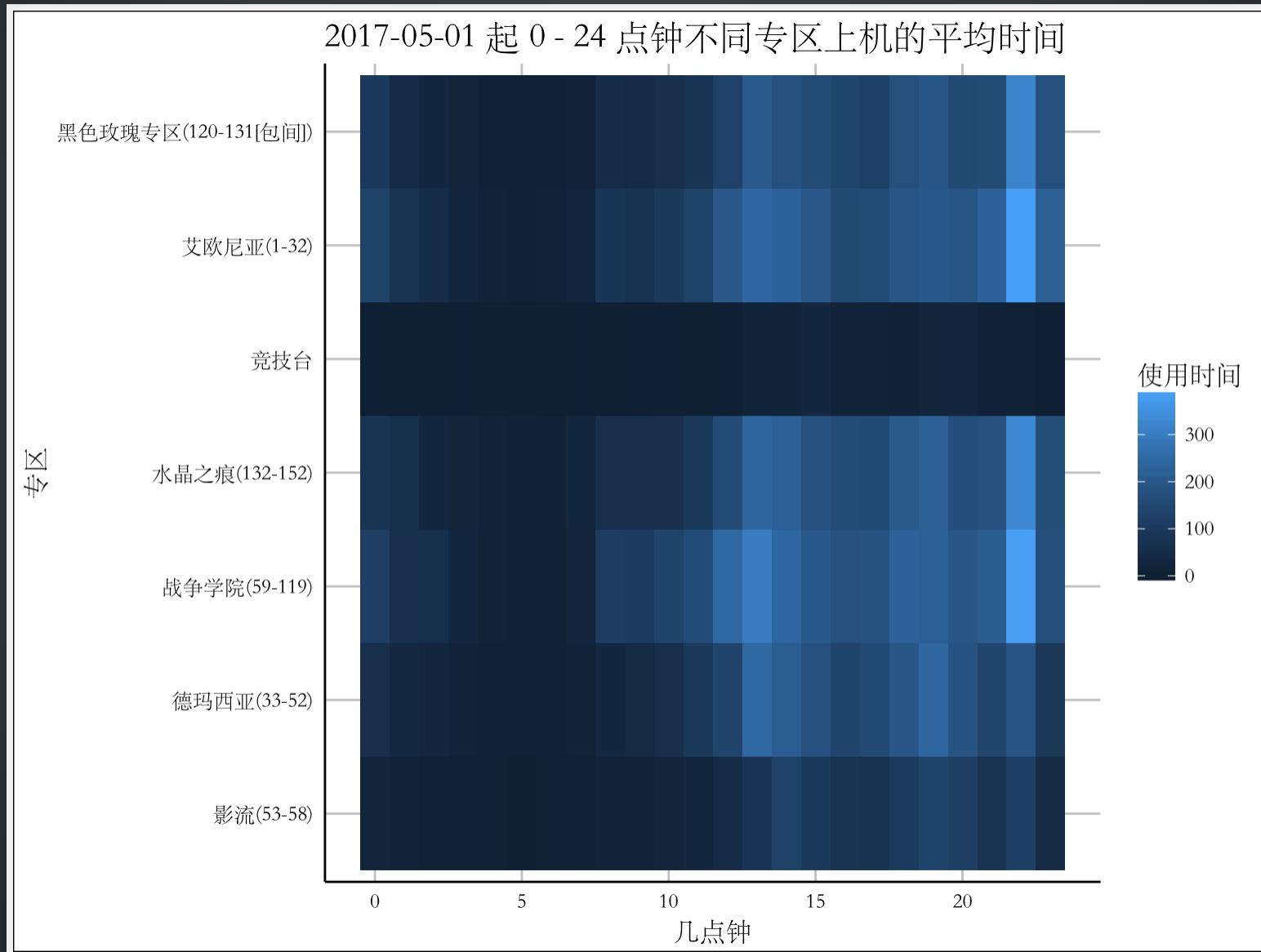
I was in my hometown last week. One day, I went to an internet cafe to play PUBG. The boss talked with me about the general operation state, e.g. volume of customers, profit or not, etc.

During the conversation, the boss expressed his willingness to optimise the operation of the cafe. He acknowledged me that every internet cafe has its **POS system**, and the **data is exportable**. The boss asked me whether there is anything we can do using those data.

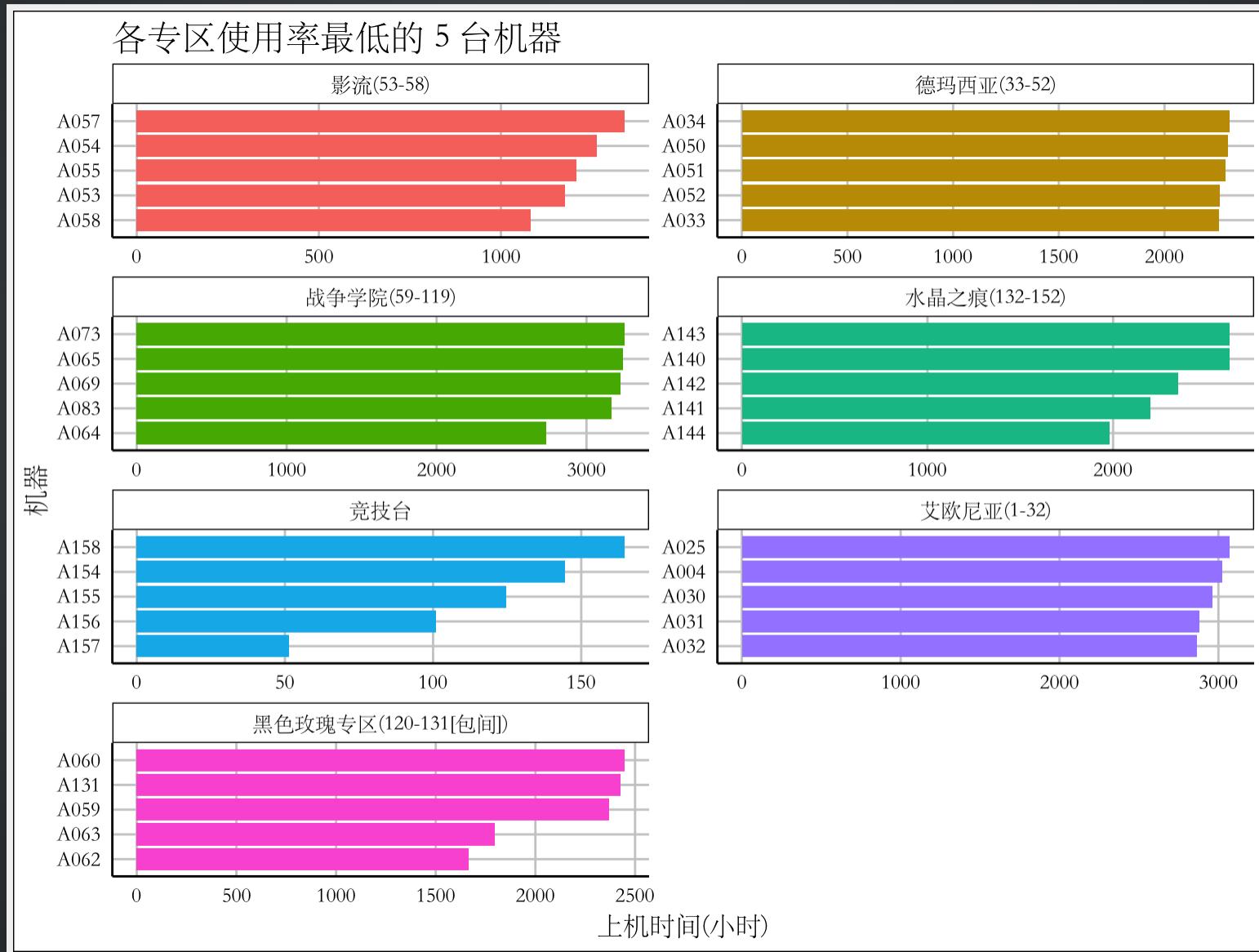
USAGE OF MACHINES



USAGE OF MACHINES IN DIFFERENT AREAS



WHICH MACHINES PERFORM THE WORST



WHAT WE HAVE LEARNED IN CASE 1

- A internet cafe does not need to hire a fulltime data scientists.
- But this does not denied the fact that there will be huge benefit if the boss has a data-driven mindset.

CASE STUDY 2: “DATA-INSIGHTS” PROJECT

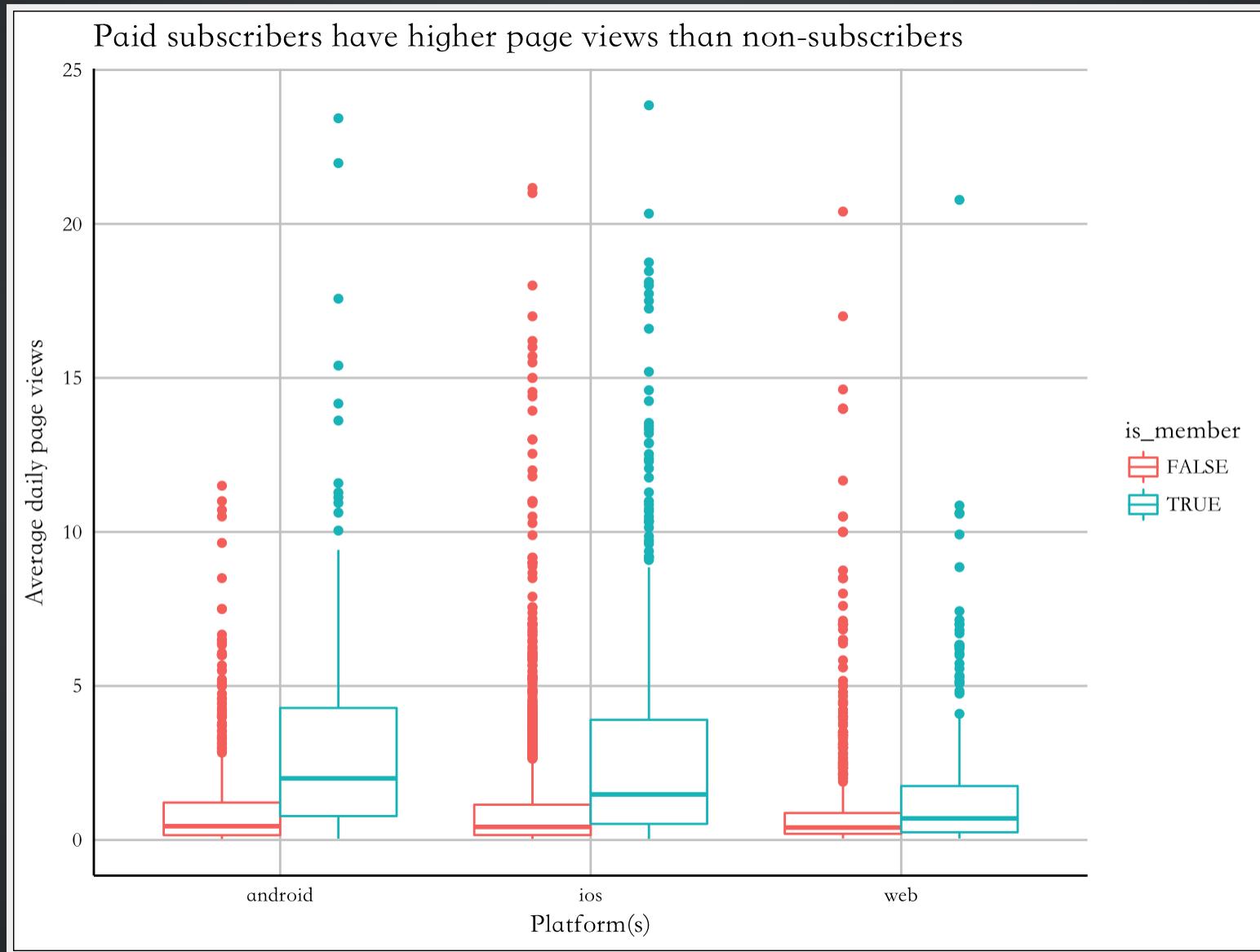
The data team was dissolved.

However, I launched a personal project called
data-insights.

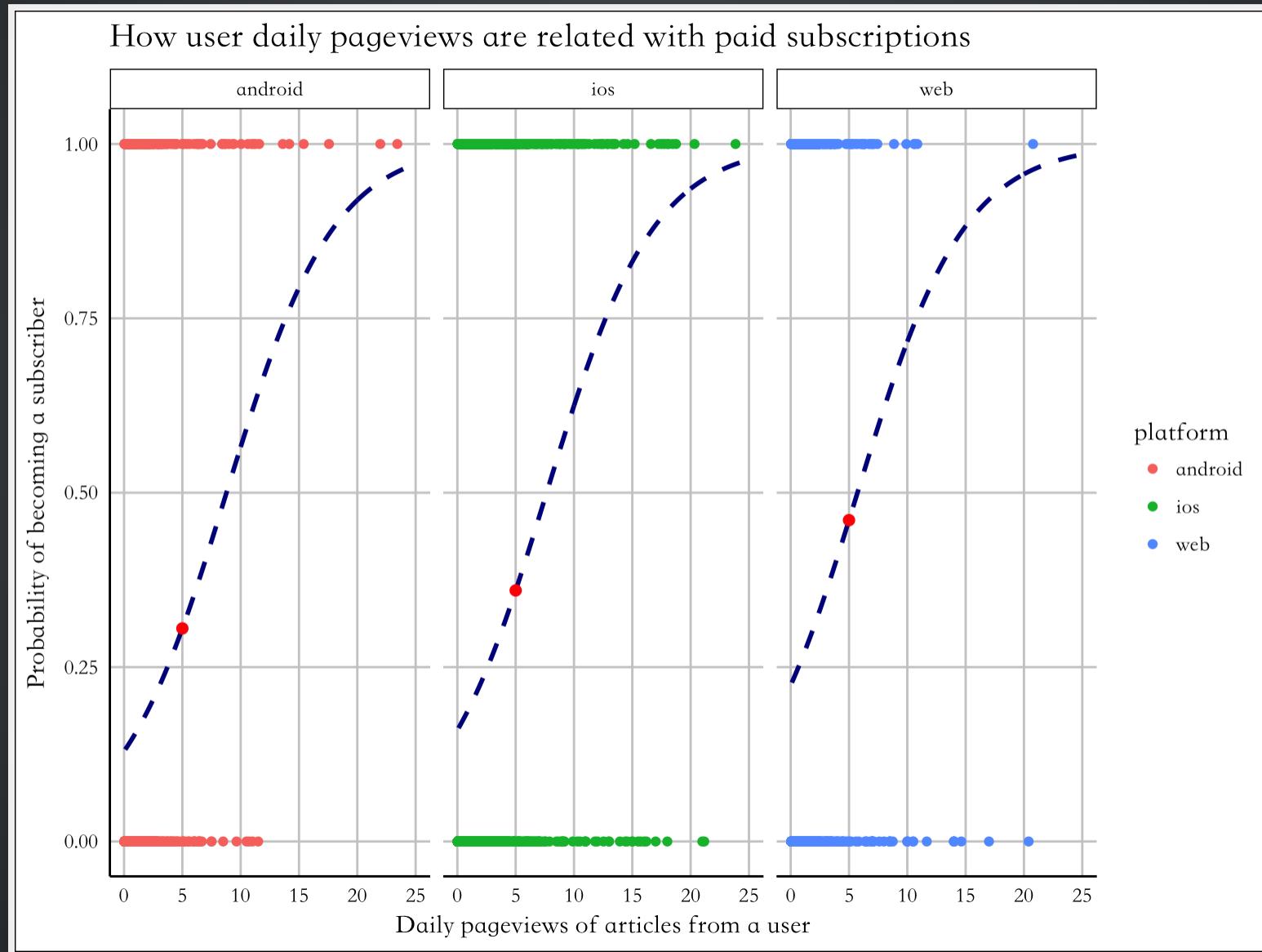
CASE STUDY 2: “DATA-INSIGHTS” PROJECT

- Data Insights: 付費會員的累計增長/取消/流失率
- Data Insights: 會員的“存活率”
- Data Insights: 年付恢復原價後，我們為什麼營收更少了
- Data Insights: 找到合理價格
- ...
- Data Insights: 讀者平均每天讀多少文章就會付費

DATA-INSIGHTS: PAGE VIEWS AND SUBSCRIPTION



DATA-INSIGHTS: PAGE VIEWS AND SUBSCRIPTION



WHAT WE HAVE LEARNED IN CASE 2

- You don't have to have data scientists to be data driven
- Find the incentive to involve more people

CASE STUDY 3: DATA DRIVEN INVESTMENTS

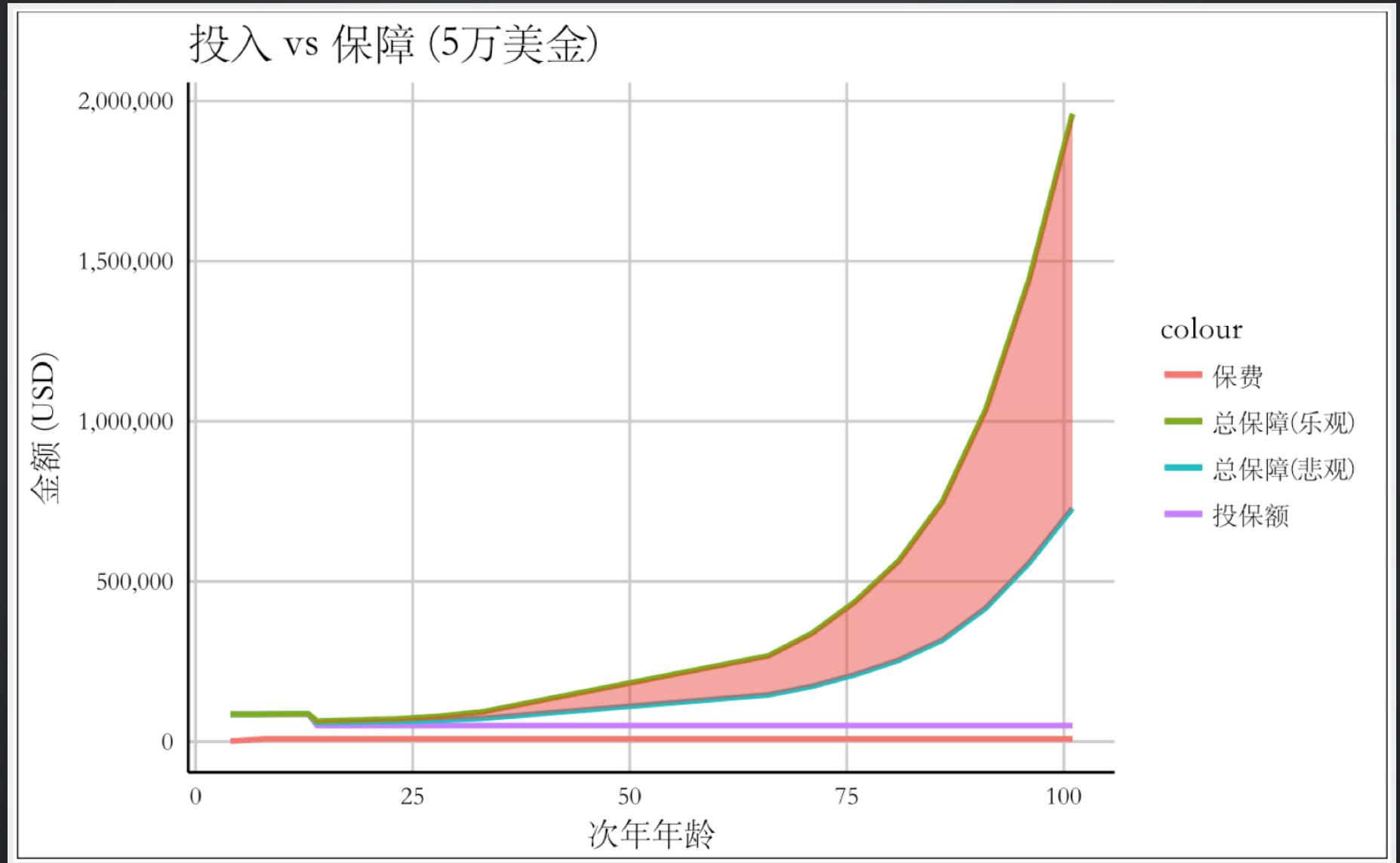
After several years of work, you may start to thinking about invest your saving into e.g. stocks, funds, insurance or even bitcoin.

This is not investment advice in this section!

CASE STUDY 3: INSURANCE

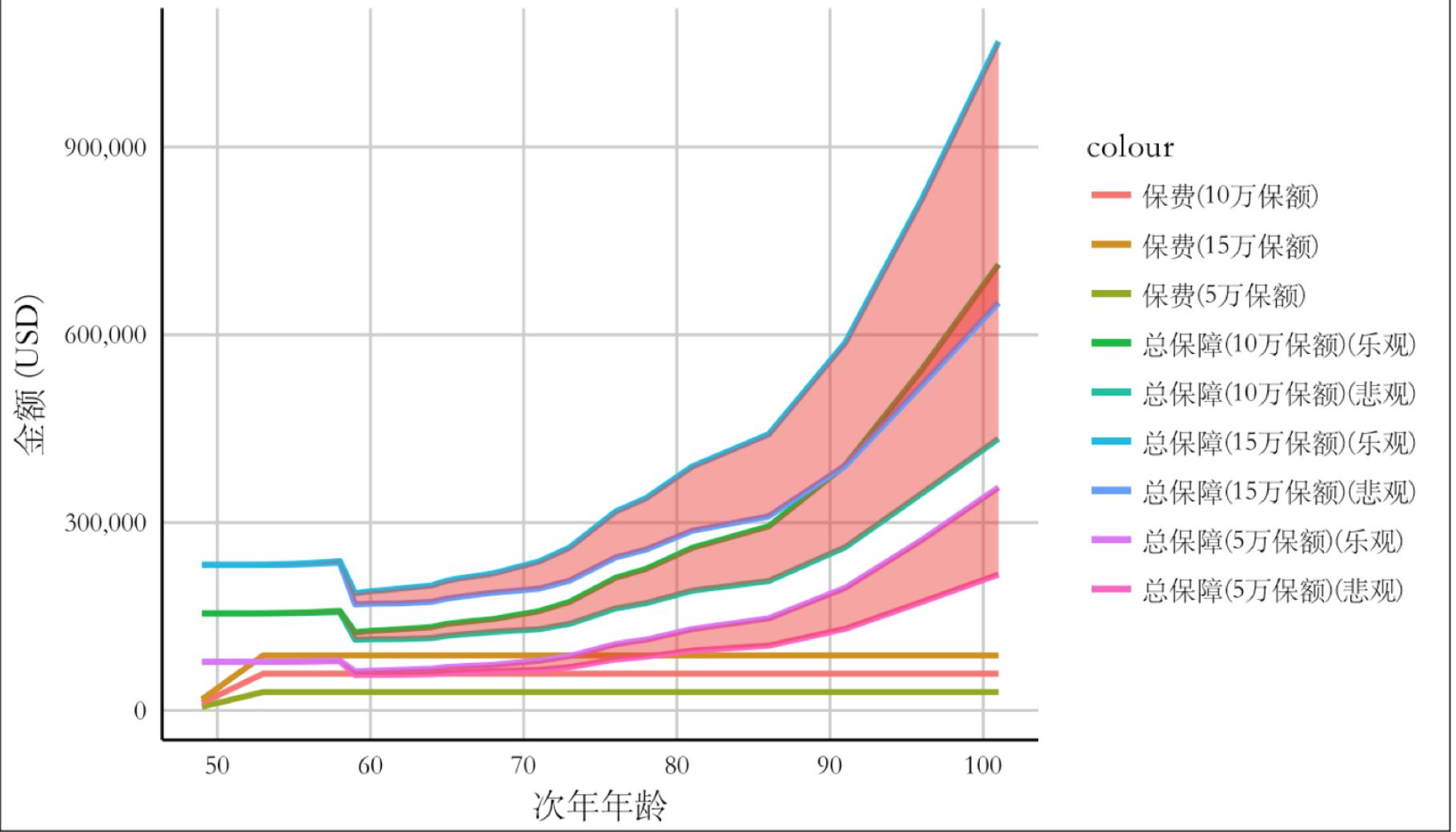
保單年度終結	繳付保費總額	身故保障 / 嚴重疾病保障					
		保證金額 ^b (A)	悲觀情景		樂觀情景		
			非保證特別紅利 (B)	總額 (A) + (B)	非保證特別紅利 (C)	總額 (A) + (C)	
1	4,110	225,000	0	225,000	0	225,000	
2	8,220	225,000	0	225,000	0	225,000	
3	12,330	225,000	0	225,000	0	225,000	
4	16,440	225,000	0	225,000	0	225,000	
5	20,550	225,000	2,750	227,750	3,324	228,324	
6	24,660	225,000	4,175	229,175	5,114	230,114	
7	28,770	225,000	6,165	231,165	7,580	232,580	
8	32,880	225,000	8,481	233,481	10,527	235,527	
9	36,990	225,000	11,124	236,124	13,925	238,925	
10	41,100	225,000	13,220	238,220	16,902	241,902	
11	45,210	225,000	15,213	240,213	19,871	244,871	
12	49,320	225,000	16,839	241,839	22,574	247,574	
13	53,430	225,000	18,219	243,219	25,139	250,139	
14	57,540	225,000	19,872	244,872	28,059	253,059	
15	61,650	225,000	21,657	246,657	31,205	256,205	
16	65,760	150,000	63,359	213,359	92,897	242,897	
17	69,870	150,000	68,729	218,729	104,483	254,483	
18	73,980	150,000	70,637	220,637	118,094	268,094	
19	78,090	150,000	73,391	223,391	132,428	282,428	
20	82,200	150,000	76,250	226,250	139,067	289,067	
25	82,200	150,000	85,050	235,050	100,620	289,620	

CASE STUDY 3: INSURANCE



CASE STUDY 3: INSURANCE

不同投保额所对应的保障(5万/10万/15万美金为例)



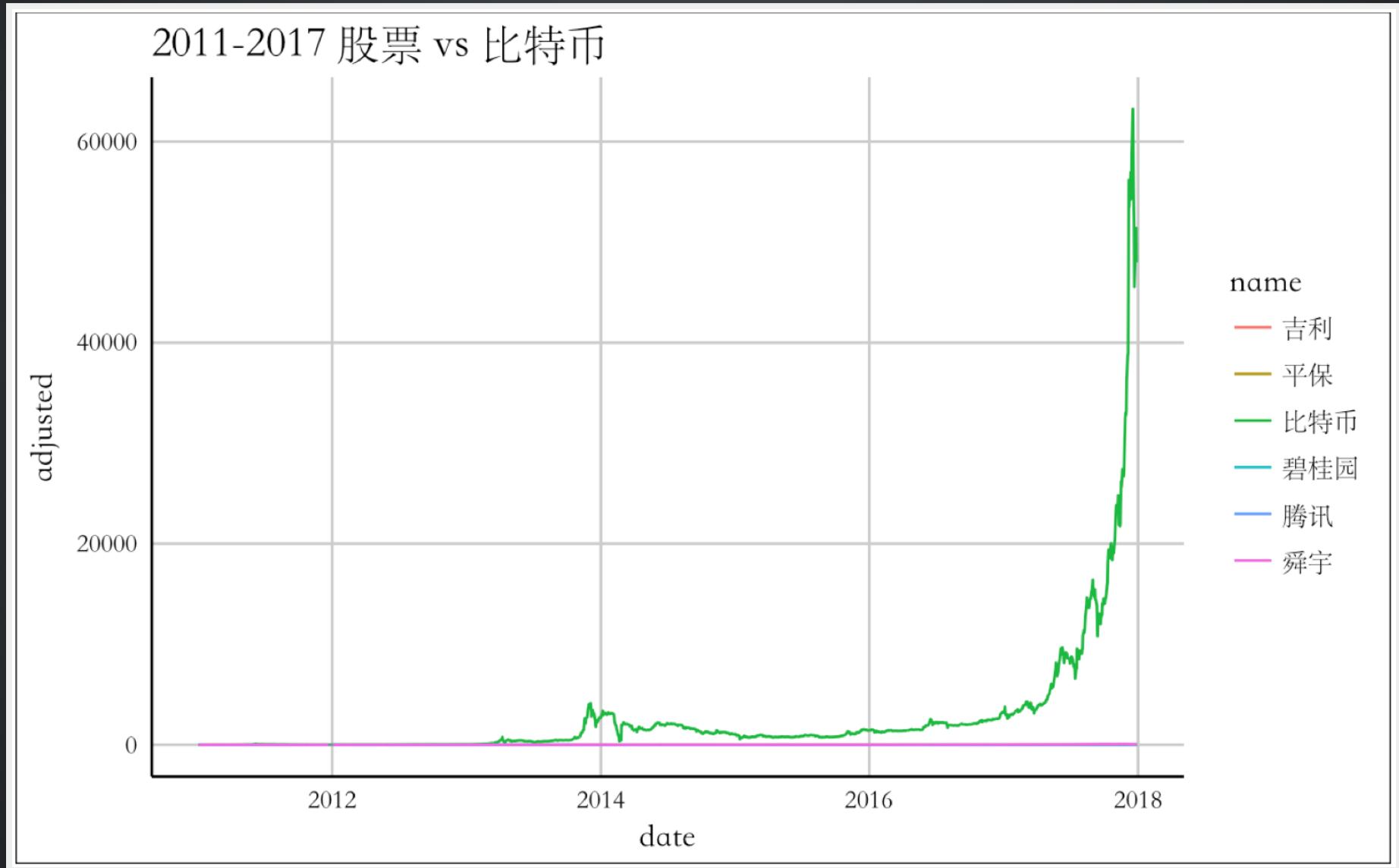
CASE STUDY 3: STOCKS (2017)



CASE STUDY 3: STOCKS VS BITCOIN (2017)



CASE STUDY 3: STOCKS VS BITCOIN (FROM 2011)



A BOOK I RECOMMEND TO YOU

R for Data Science by Hadley Wickham

Best book available introducing “tidyverse”

A BLOG I RECOMMEND TO YOU

Variance Explained by David Robinson

I started to learn R after reading this blog

A POEM FROM A MATHEMATICIAN

數缺形時少直觀，形少數時難入微；
數形結合百般好，隔離分家萬事休。

by 華羅庚

A STUDY GROUP



THANKS

