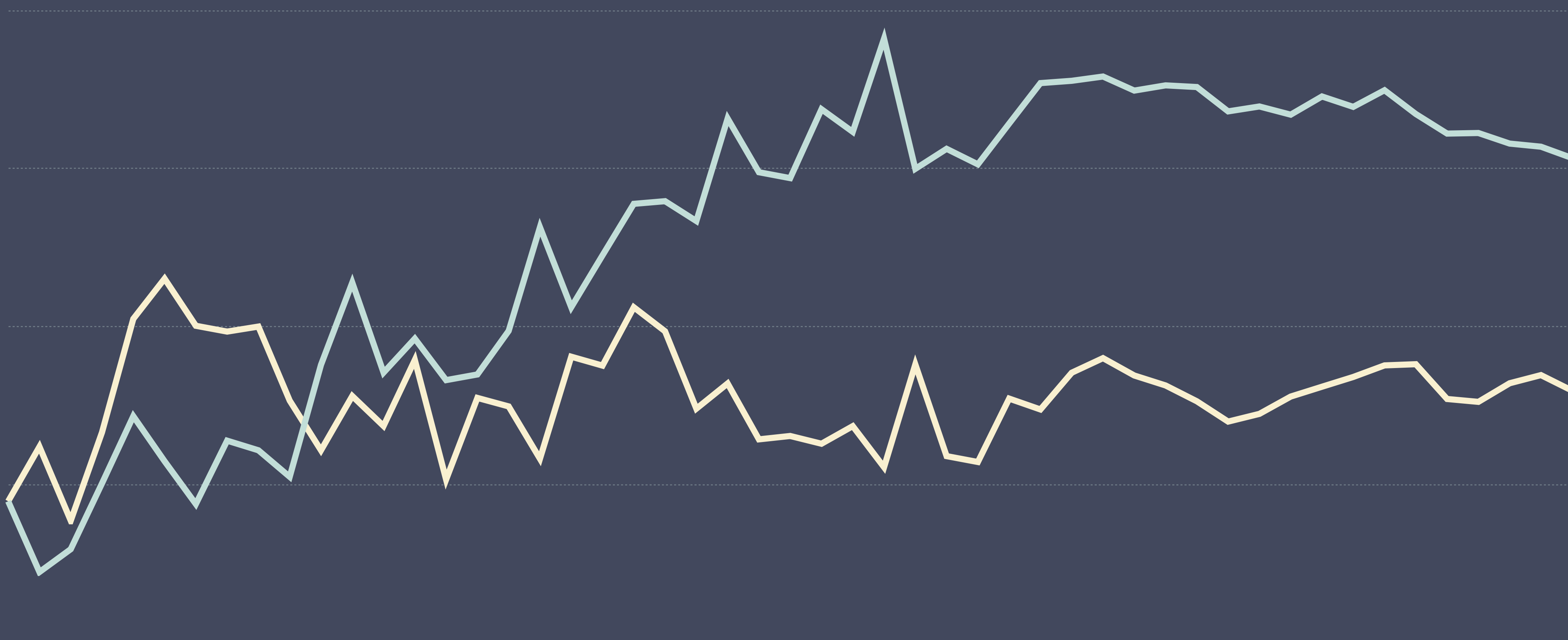


知乎数据那些事儿

李申申

Co-founder & CTO



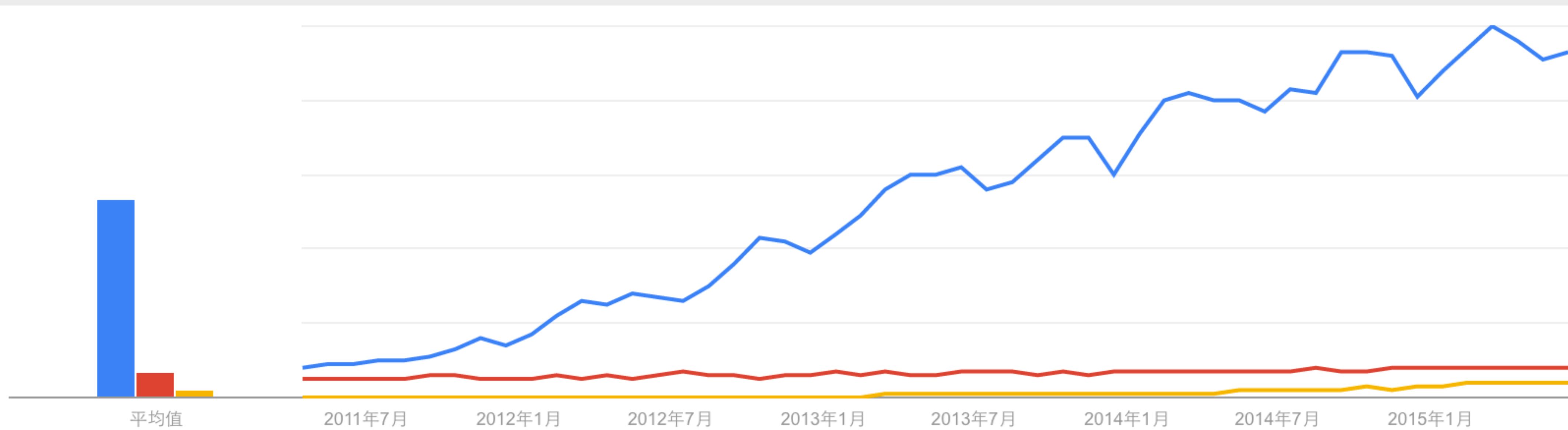
— 大数据

— O2O

big data
搜索字词

Online to offline
搜索字词

Sharing Economy
搜索字词





今天的知乎

29,000,000

注册用户

6,130,000

问题总数

110,000,000

月独立用户

19,190,000

回答总数

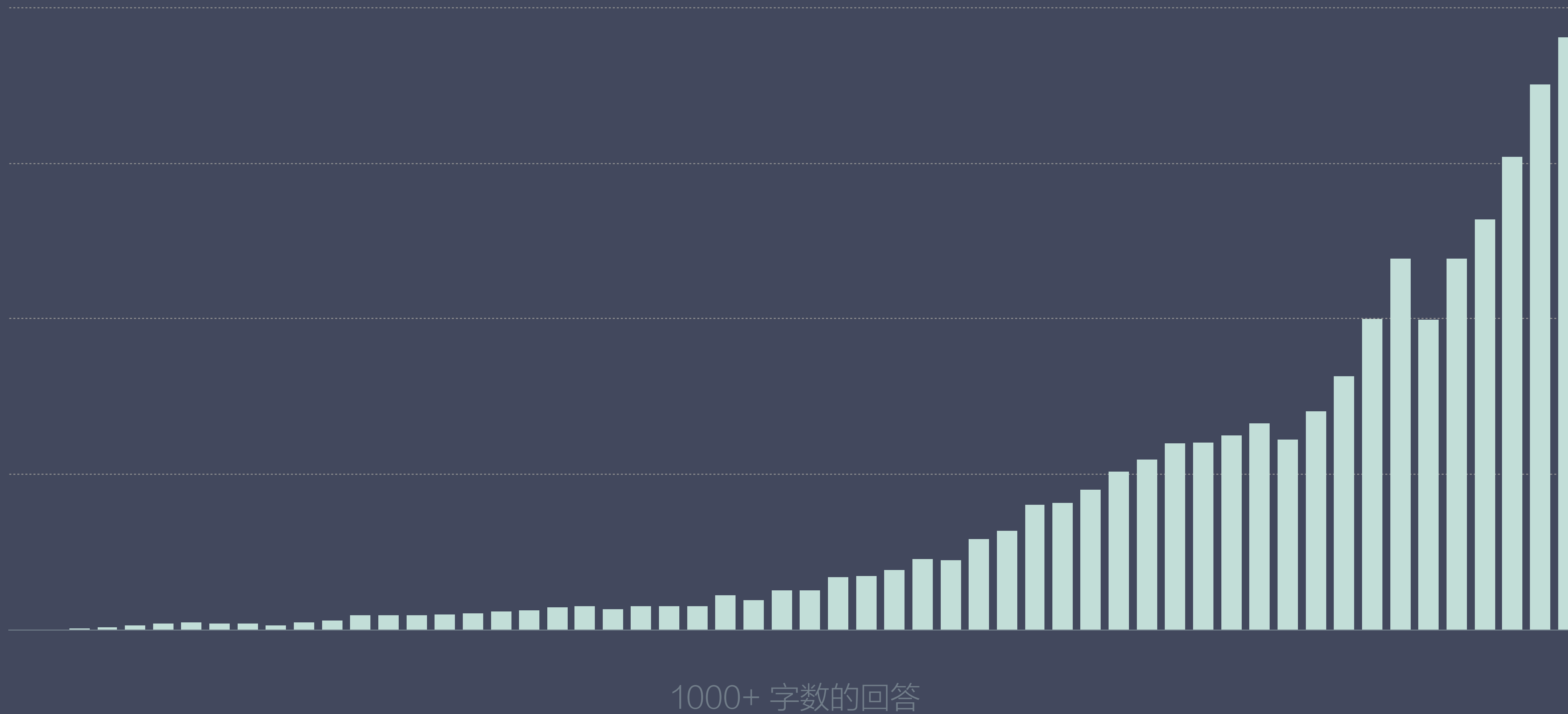
300,000,000

月页面浏览

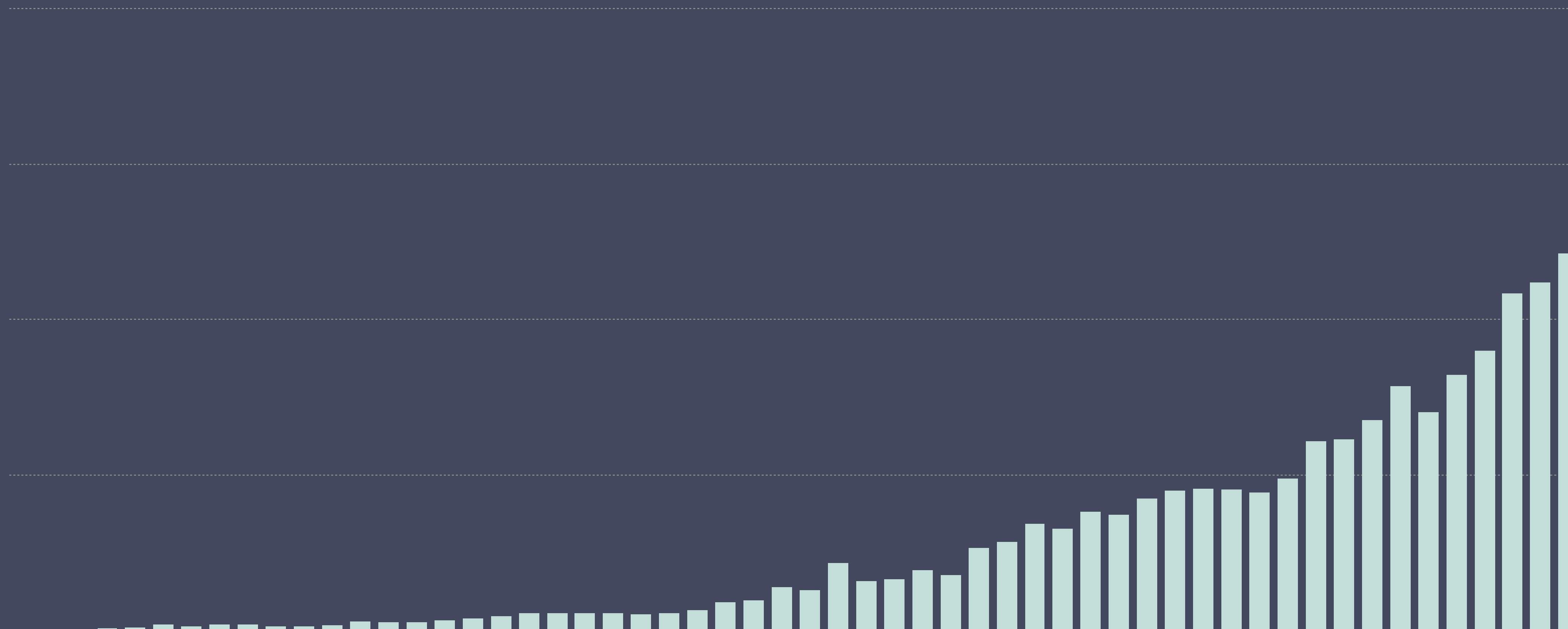
4,129,000,000

全站回答字数

更多的有价值分享



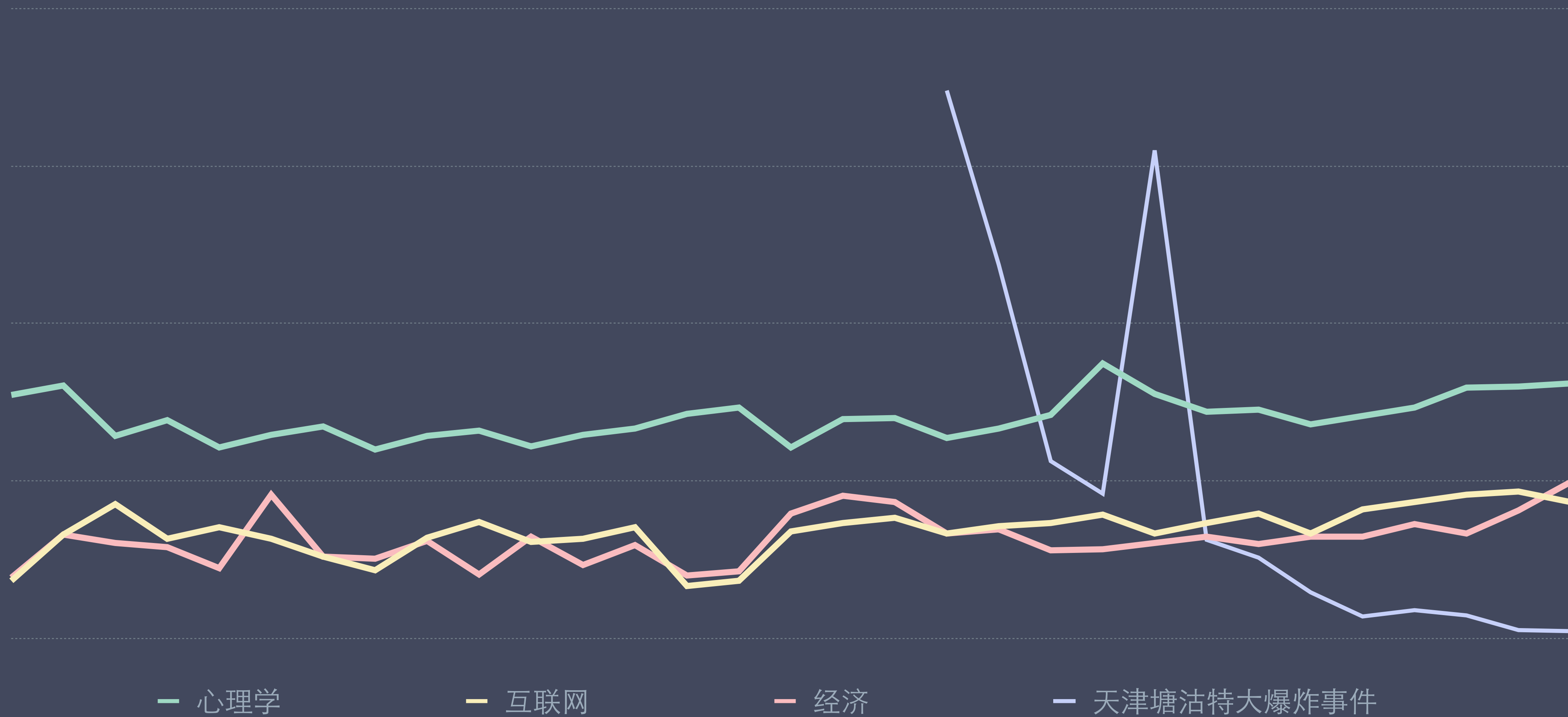
更多的有价值分享



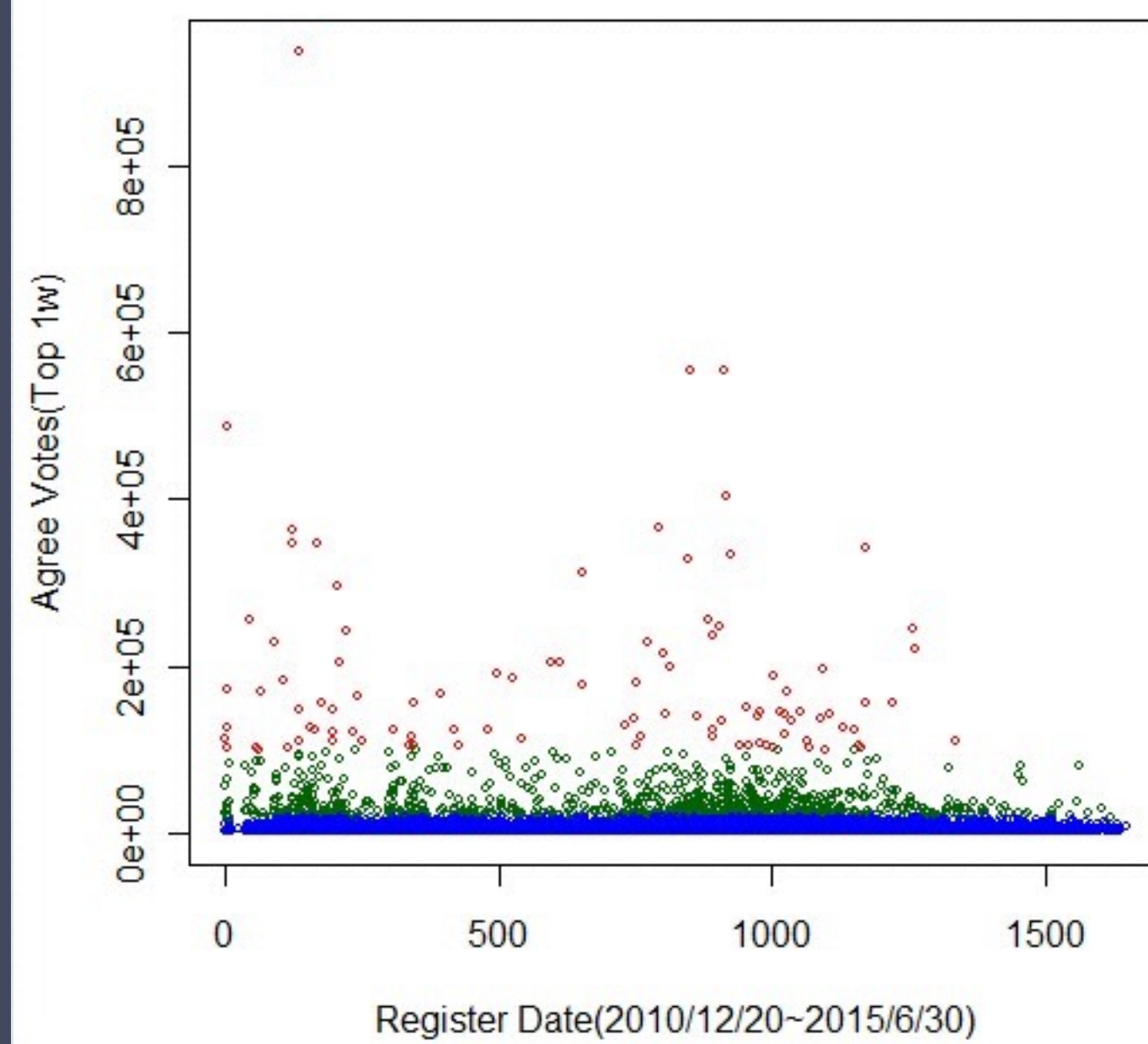
人声鼎沸的知乎广场

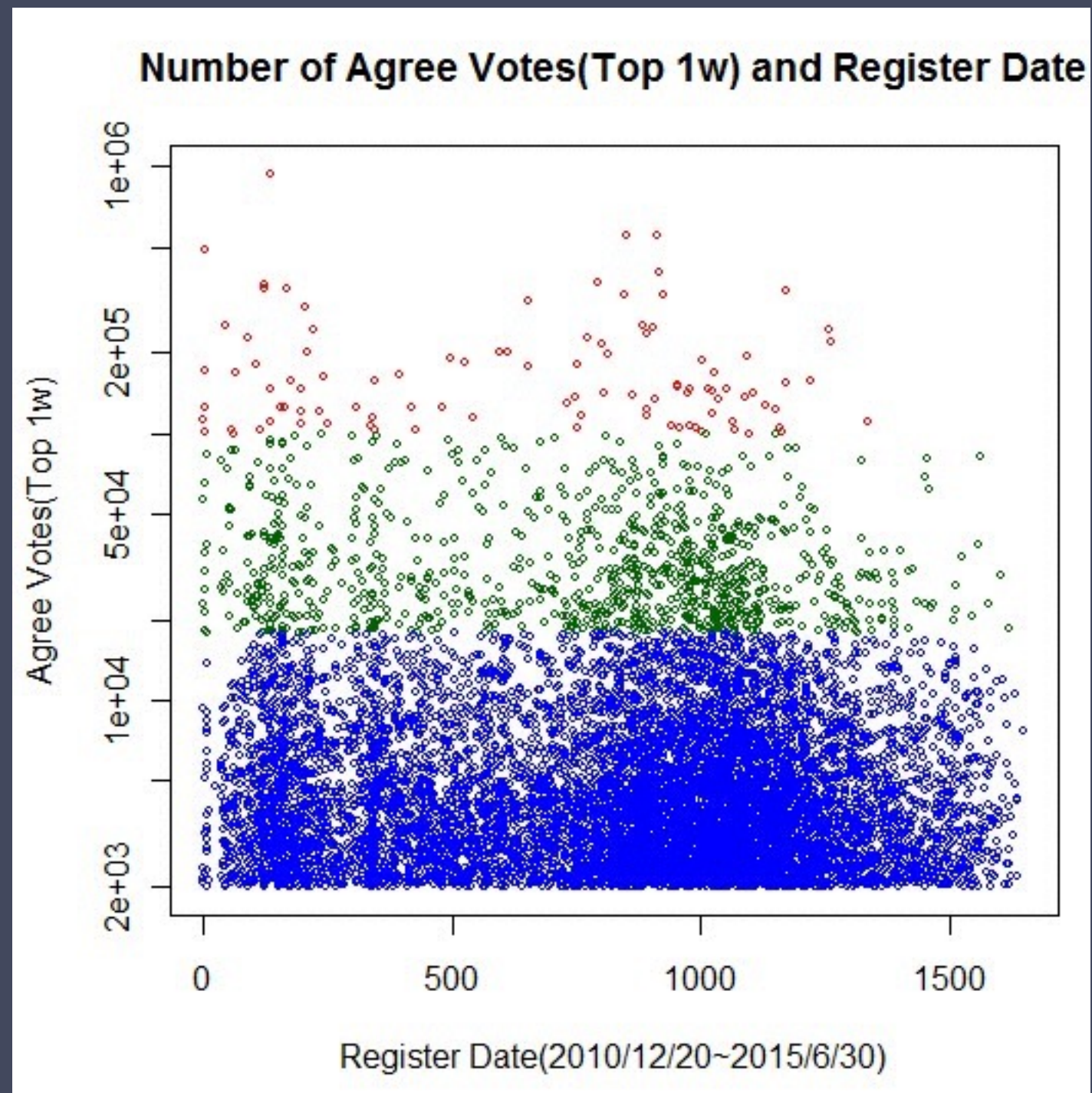
长时间坚持健身可以一定程度上改变容貌吗？ 不同行业的真实工作是怎样的？ 当年的下岗潮是怎么回事？
怎样才能通过光线做建筑设计 蒸汽朋克与赛博朋克的区别在哪里？ 每天花费八小时或以上，在 15 - 20 天内能学会或完成的事有哪些？
高频交易有那些著名的算法？ 国内有哪些冷门但有特色的旅游地点？ 为什么霍金警告我们不要接触外星人？
北大清华或者其他顶尖 985 院校内存在「智商被碾压」吗？ 怎样成为全栈工程师？ 为什么感觉国有企业反而越来越强势了？
金庸小说里有哪些不易发现却很打动人的小细节？ 不同行业的真实工作是怎样的？
梦游的原理是什么，对人体有害么？ 在你的专业里，有什么基础知识是和普通人的认识不相符的？ 手机号为什么必须有归属地？
有哪些人除了自身主业做得好，副业也做得很厉害却不被大多数人所知？ 为什么 2015 年初，上海有卫计委官员呼吁大家生二胎？
如何系统性学习摄影后期调色？

人声鼎沸的知乎广场

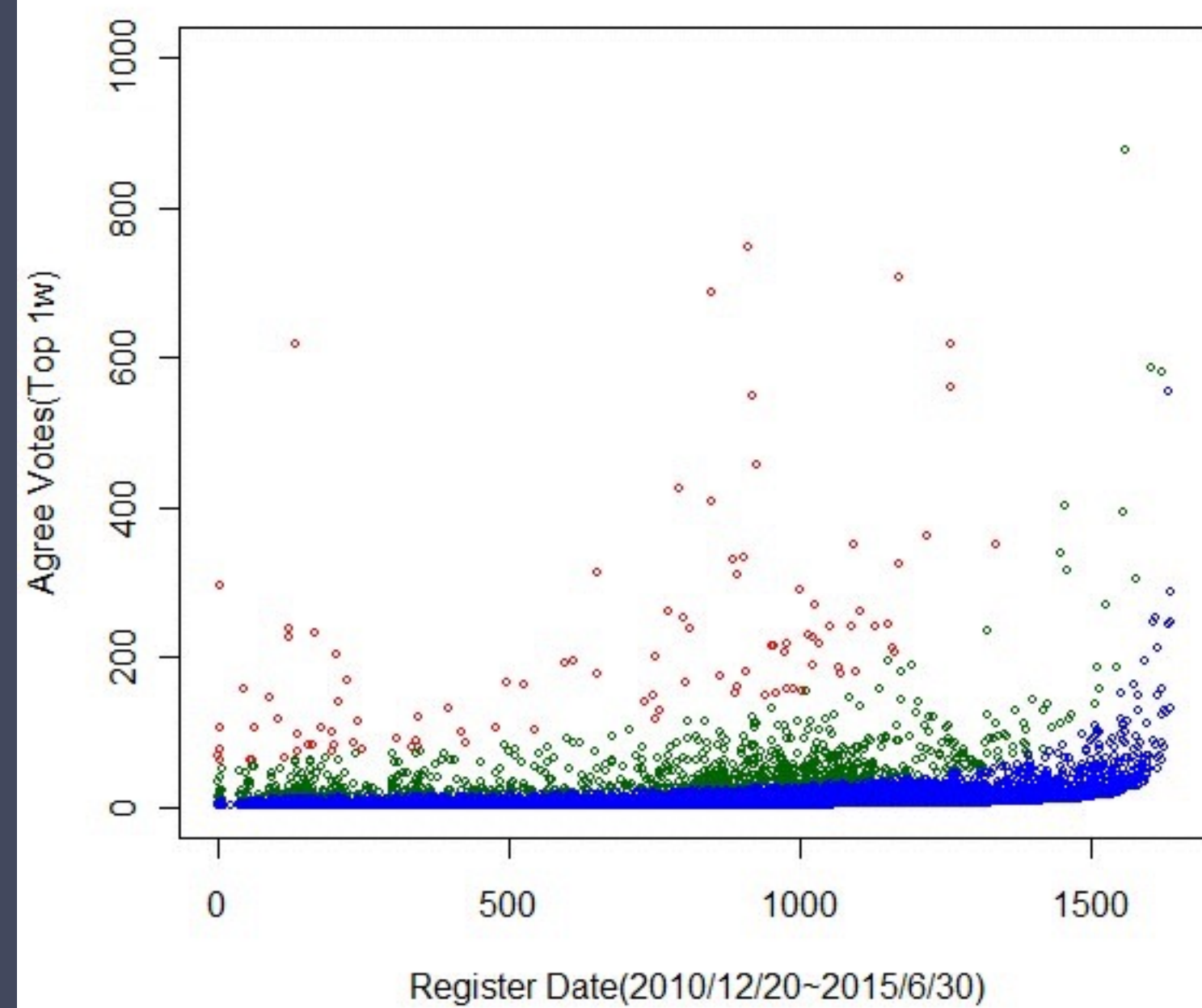


Number of Agree Votes(Top 1w) and Register Date



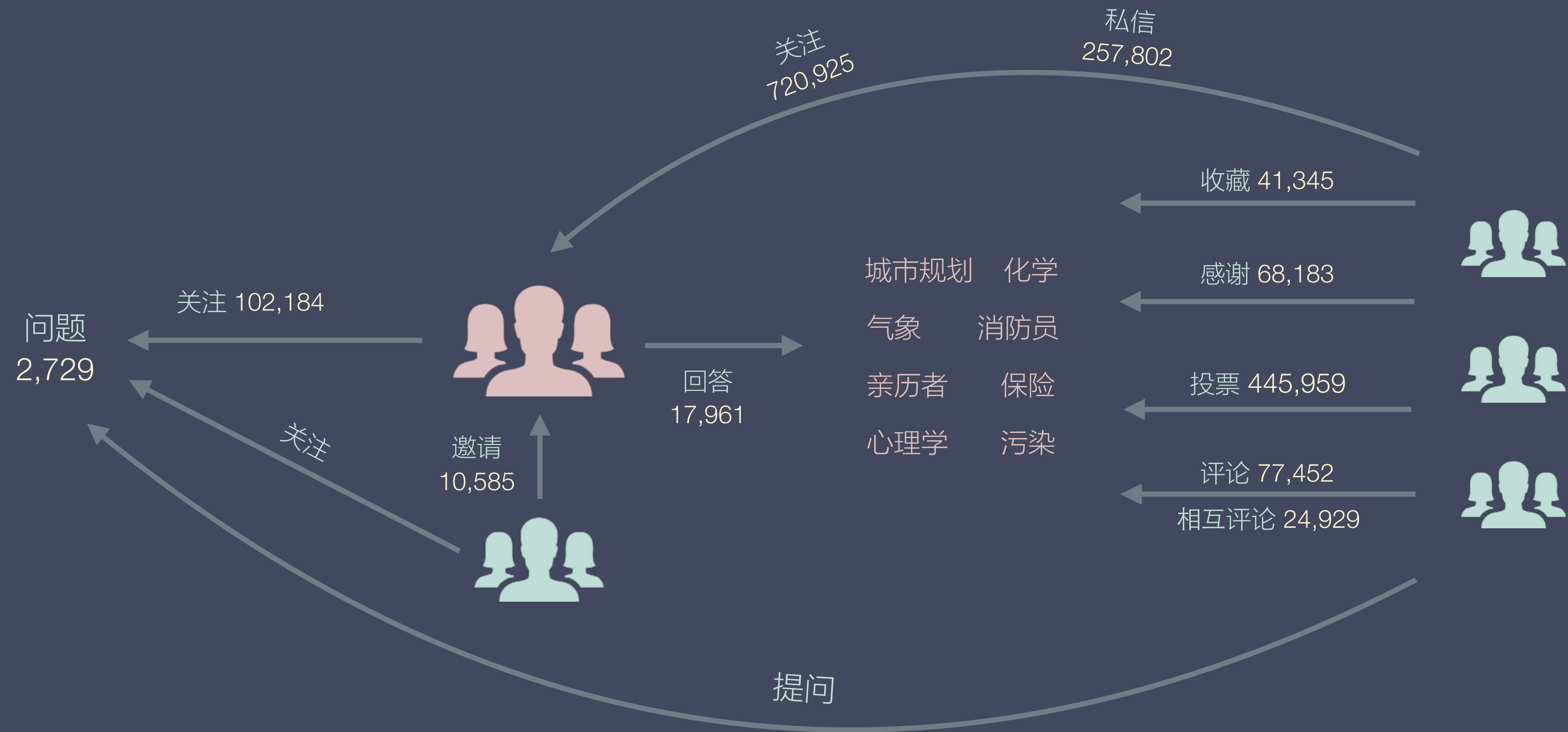


Average Agree Votes Per Day In Top 1w Agree Votes Owners

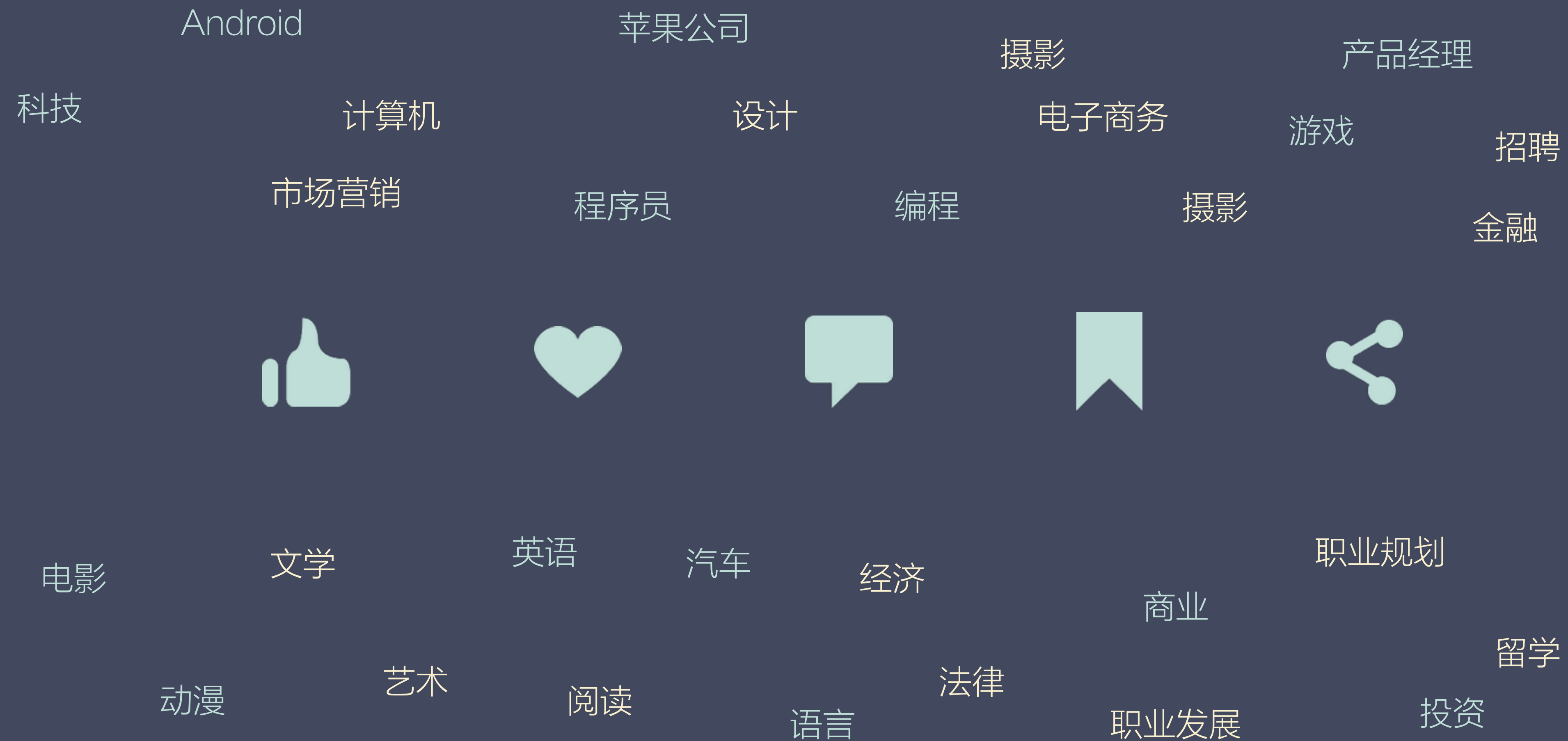


认真，你才会赢

优质内容借助社交行为在知乎「产生和流动」

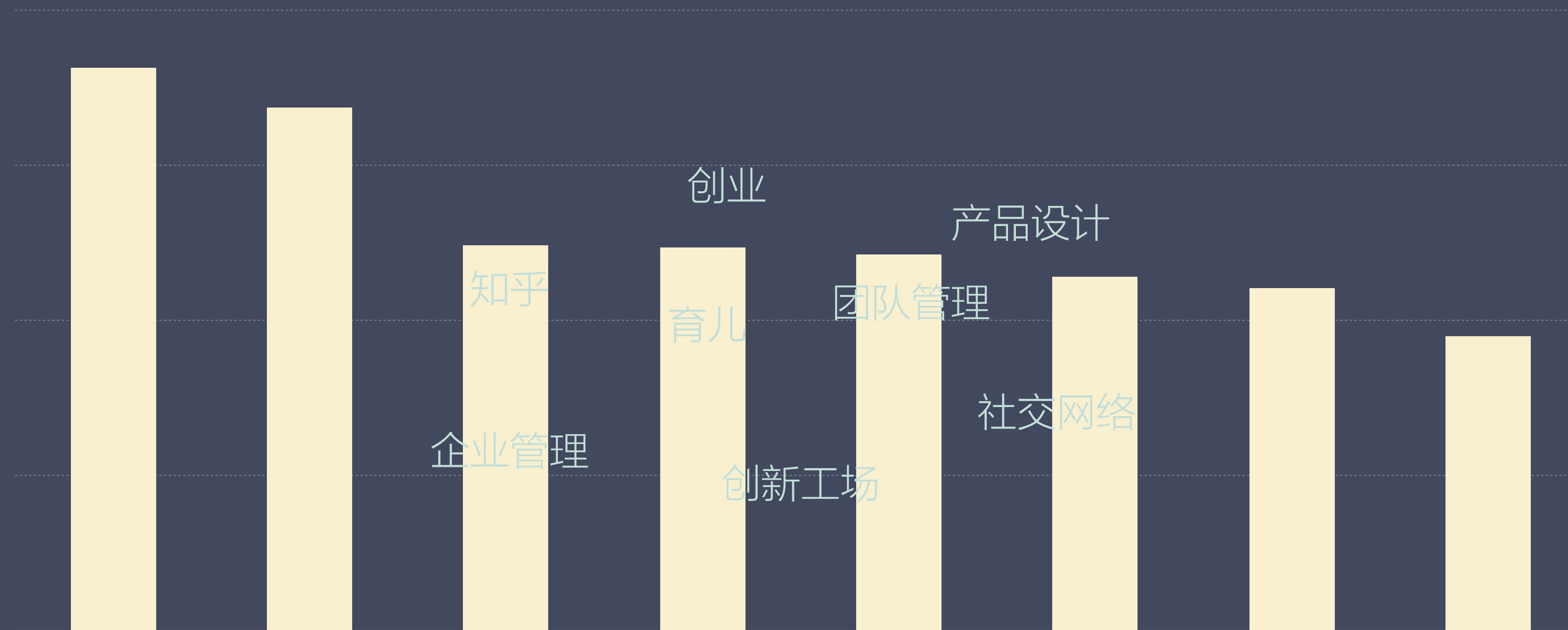


我们基于这些数据做了哪些有价值的工作？



周源

CEO & Co-founder @ Zhihu



周源擅长的话题

答案排序的奥秘与忧愁

排名 = 赞同名 + 赞同 - 反对（加权）

威尔逊得分算法

有较早发布的高票回答导致的回答质量排序无法得到更多曝光

当投票的人比较少时，可以根据已经获得的票数估计这个回答的质量得分，投票的人越多则估计结果越接近真实得分

煽动性高的高票回答，即使得到了更多反对，排序依然靠前

如何评价知乎的回答排序算法？

chenqin
数据帝

51 个回答

看了一下这个算法，根据不同答案能获得让赞同率开始起作用他的赞同比例不低，这个回答

459 人关注 34 条评论

邀请回答

如何看待知乎自 2014 年 12 月起使答排序算法？

maldijie zhang
472

-我觉得让@曹文雯 12赞的答案站第一算法还是合理的，大家去给她点赞啊。告诉你，一个普通用户的反对能抵一千个

jyo gan
194

反对票的权重过大了。只有鸡汤型的，才容易排在上面了，观点鲜明的不论怎么被踩下去。=====刚发出来的(

摇啊摇毛阿毛
5

首先表态：我肥肠支持这个算法。其次对新算法有态度的盆友们，你们啊，说明：我支持这个算法，不是因为我很喜欢

蒋墨梓
119

现在知乎水军伤心了，以前100块买2把高票答案搞下来，现在50块买100个了，而且效果还更好。-----

感谢作者 没有帮助 加入

如何评价知乎的回答排序算法？

黄涛
产品设计 金融 运动 自然科学

这次，我给知乎点 32 个「赞同」——知乎新的回答排序算法

曾加 · 8 月前

「威尔逊得分」算法是对以下回答强有力的遏制：

1、并不切题的抖机灵答案

2、带攻击性的站队答案

3、高关注度用户的低质量答案

今天（2014年12月5日），知乎官方公布新的回答排序算法，具体内容可以参见

知乎如何对回答进行排序？你的一票很重要 - 知乎产品专栏 - 知乎专栏

如何评价知乎的回答排序算法？ - 知乎产

如果你是一个对回答排序有兴趣的人，你会发现这个改革非同小可，仔细看看，你就会发现，很多问题下的回答排序发生了翻天覆地的变化：有些原本高高在上的高票回答突然就降

感谢作者 没有帮助 加入收藏

30 个回答

718 人关注 8 条评论

关注

邀请回答

添加答案

雨亦奇
80

@chenqin @苏莉安 @Gary Wang @曾加 的回答都很好，我希望从social network和median voter theorem角度略加补充，算是抛砖引玉。不过首先

chenqin
163

看了一下这个算法，根据不同的 z_{α} ，只要一个答案能获得让赞同率开始起作用的赞同数，且他的赞同比例不低，这个回答就能排名靠前。 z_{α}

黄涛
162

这个算法中用来做排序依据的得分的更严谨的名称是「二项分布样本的威尔逊置信区间下界」。计算公式是在 1927 年由数学家 Edwin B. Wilson 发

苏莉安
74

谢邀。新算法肉眼可见的效果就是加大了反对票的杀伤力，它对知识型答案影响有限，对世界观型的答案则很容易将正反双方的高票答案都打到下面

孙竟
4

手机上打字太累，见谅。我个人认为知乎的回答排



WUKONG

最新动态

设置



来自 用户界面设计

×

是不是每个设计师都都会有一个自己的素材库？怎么创建并且完善呢。

93

许雅茜，不念倩。



并非专业设计师，只是喜欢自己玩玩。我觉得这个素材库，一方面是设计的一些“部件”（笔刷、高清图、icon矢量图等），一方面是其他设计师的设计作品，用以锻炼自己的审美，找寻新的排版和用色灵感。

这个回答将从素材来源和素材管理两方面来解答。一 素材来源... [显示全部](#)

+ 关注问题 16 条评论 • 作者保留权利



来自 古典音乐

×

怎样欣赏交响乐、歌剧、古典和美声这类音乐？

92

齐观，里赫特即真理

听古典音乐十五年了，占到全部生命的2/3，感觉这个比例会随着年龄和阅历的增长，越来越趋近1。古典音乐是我灵感的源泉，是安居的栖息，更是我面对苦难的庇护所——就像《肖申克的救赎》中，Andy 那句“有莫扎特陪着我”一样——如果离开音乐，或者因为生活... [显示全部](#)

+ 关注问题 15 条评论 • 禁止转载



苏菲 赞同了回答 • 20 分钟前

×

为什么日本人不结婚？

150

食饭游记，影视戏剧文学本科，比较历史学硕士，博士...

这个问题的本质和中国年轻人背负的三个问题有异曲同工之妙，那就是：“你有对象了吗？”“你什么时候买房？”“你一年挣多少钱哪？”这三个问题的本质是什么呢？那就是，一个中国年轻人，到了一定到年纪，就一定要找对象结婚，买房成家以及有一定到收入。大... [显示全部](#)

+ 关注问题 24 条评论 • 作者保留权利



很多人赞同了 覃超 在 覃超帝国兴亡史 - 在希望的田野上 中发布的文章 • 1 天前

×

「Data Driven」从 First Round Capital 的十年投资统计看数据的重要性

38

覃超，曾经沧海难为水，Facebook工程师



这篇文章涉及创业和早期投资。覃超大魔王认为自古以来人类有两种思维模



我的草稿 1



我的收藏



我关注的问题



邀请我回答的问题 2583



所有问题



话题广场



公共编辑动态



社区服务中心

知乎圆桌

[查看全部 »](#)



资金，市场，交易员 还有 12 小时结束



不养熊孩子 还有 12 小时结束



吃得明白 还有 12 小时结束

你可能感兴趣的人

[更多推荐 »](#)



朱炫

闷骚而近妖 微博：大师兄_朱炫
田亮 等 133 人也关注了他 • [关注](#)



刘尧

他也来自 知乎 • [关注](#)



葛巾

一摘使瓜好，再摘令瓜稀。三摘尚自...
顾惜朝 等 258 人也关注了她 • [关注](#)

$$\Sigma = S_e \times W_e \times D_e$$

每周精选 (eDM)

30%

打开率

14%

点击率

相关问题

文本语义 一> 协同过滤

大数据基础上的简单算法比小数据基础上的复杂算法更加有效

今天的知乎团队

34,773 km

骑行总里程数

26.7

平均年龄

2:1

男女比例

19.4%

男生拥有清晰腹肌线的比例

11,000kg

总体重

