

Bilibili数据爬取及分析

企业实训答辩

李林峰，李唯聪，张越，张轶博

目录

01 项目简介

02 项目分工

03 每周必看榜单统计

04 分区热播统计

05 全站视频实时排行榜

06 热点：俄乌冲突

01

项目简介

Python爬虫+数据可视化分析

- 爬虫：request库
- 数据可视化分析：matplotlib库、pyecharts库

Python爬虫+数据可视化分析

- 数据和源码已开源：[GitHub仓库](#)

02

项目分工

项目分工

- 张轶博：
- 李唯聪：
- 李林峰：
- 张越：

03

每周必看榜单统计

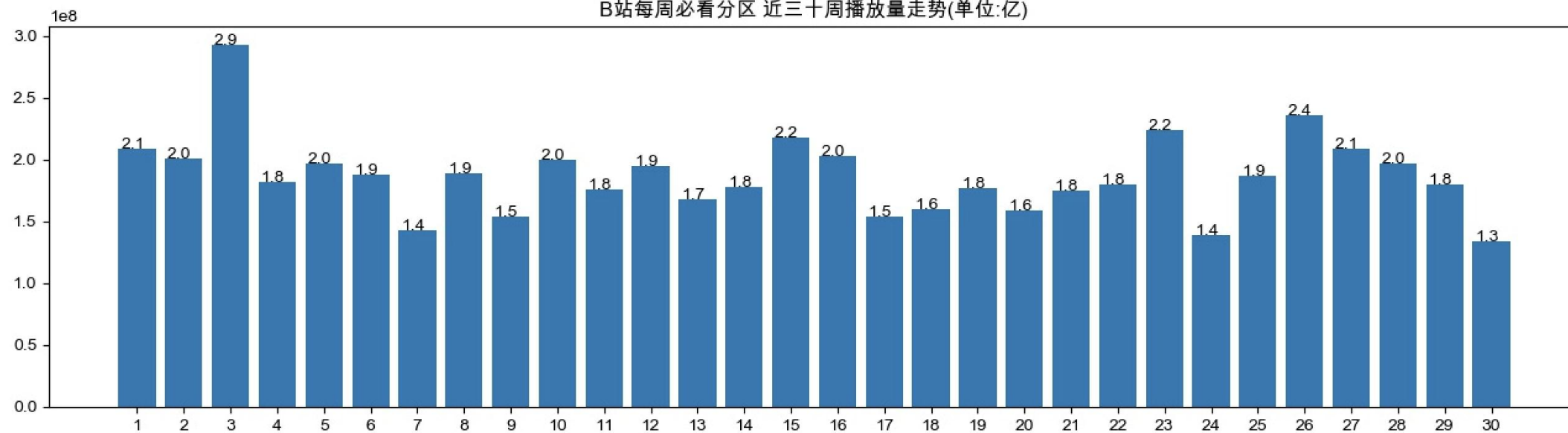
每周必看榜单统计

爬虫：观察b站每周必看榜单，
可通过此url (<https://api.bilibili.com/x/web-interface/popular/series/one?number=153>) 获取
榜单数据，其中number值为榜单期数，通过改变
number值可访问不同期的榜单数据，我们由此抓取了
近30期的榜单每个视频的播放量、弹幕量、收藏量、
点赞量、硬币量、收藏量。

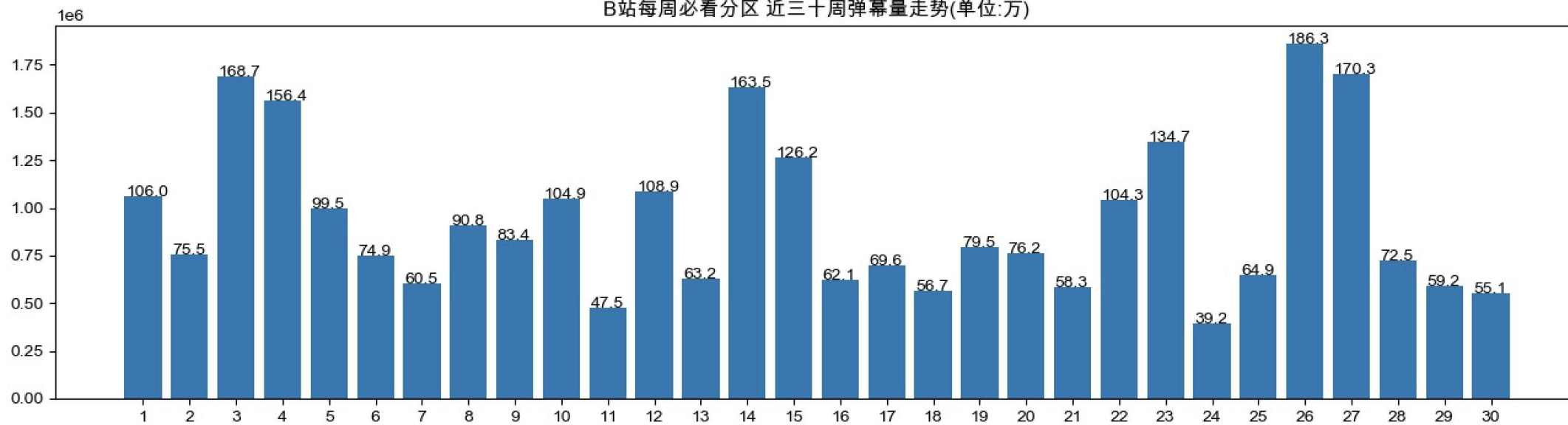


```
owner: {mid: 300073277, name: "哔哩哔哩"},
pic: "http://i2.hdslb.com/bfs/archive/efb",
pubdate: 1644717617
rcmd_reason: "全体起立！！歌坛灭霸我腾叔发
▶ rights: {bp: 0, elec: 0, download: 0, mov:
season_type: 0
short_link: "https://bili2233.cn/BV12L4y1:
short_link_v2: "https://bili2233.cn/BV12L
▼ stat: {aid: 851417393, view: 11593893, dan
aid: 851417393
coin: 109385
danmaku: 7154
dislike: 0
favorite: 34424
his_rank: 2
like: 348554
now_rank: 0
reply: 5590
share: 29281
view: 11593893
```

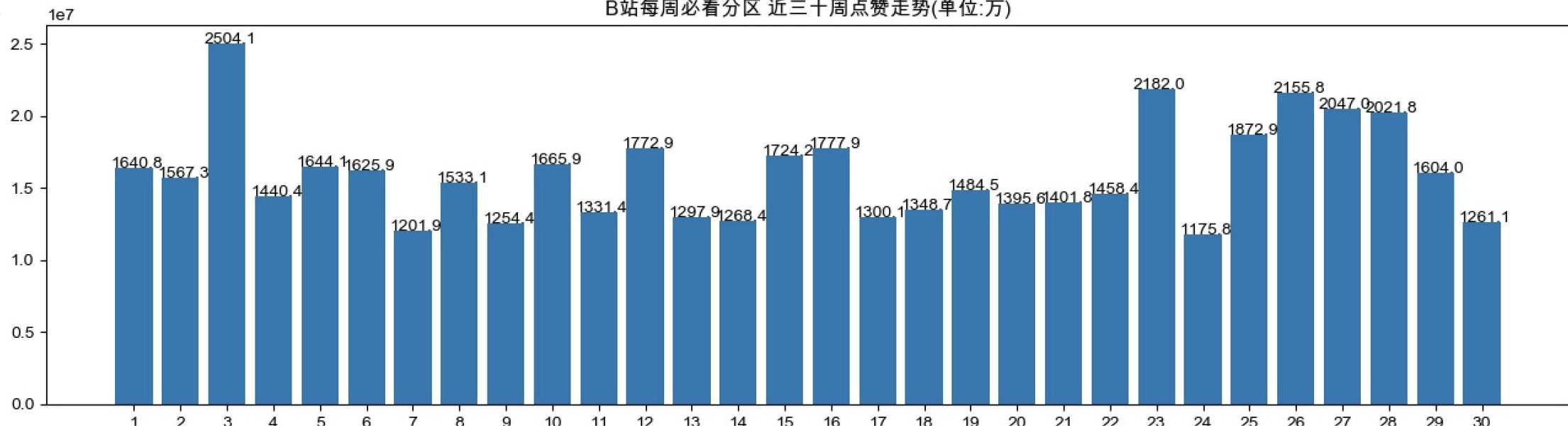
B站每周必看分区 近三十周播放量走势(单位:亿)



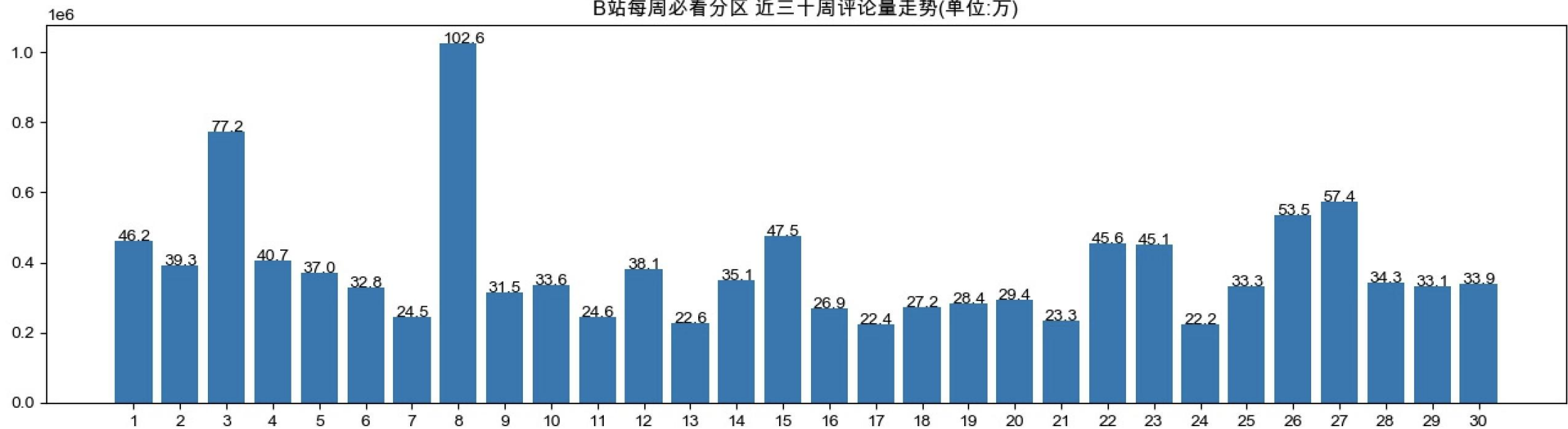
B站每周必看分区 近三十周弹幕量走势(单位:万)



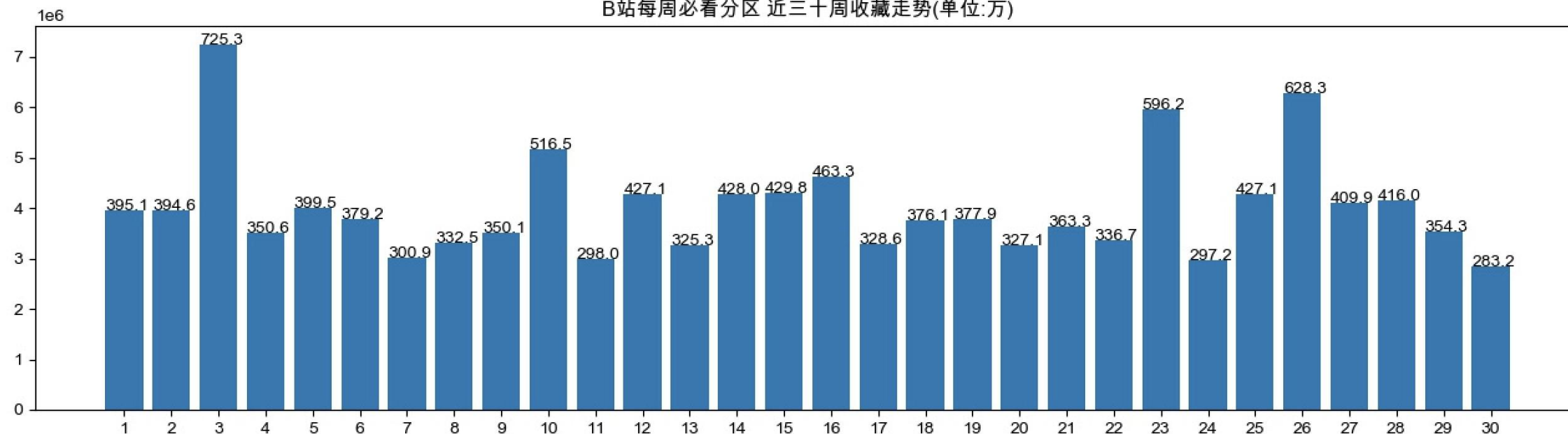
B站每周必看分区 近三十周点赞走势(单位:万)



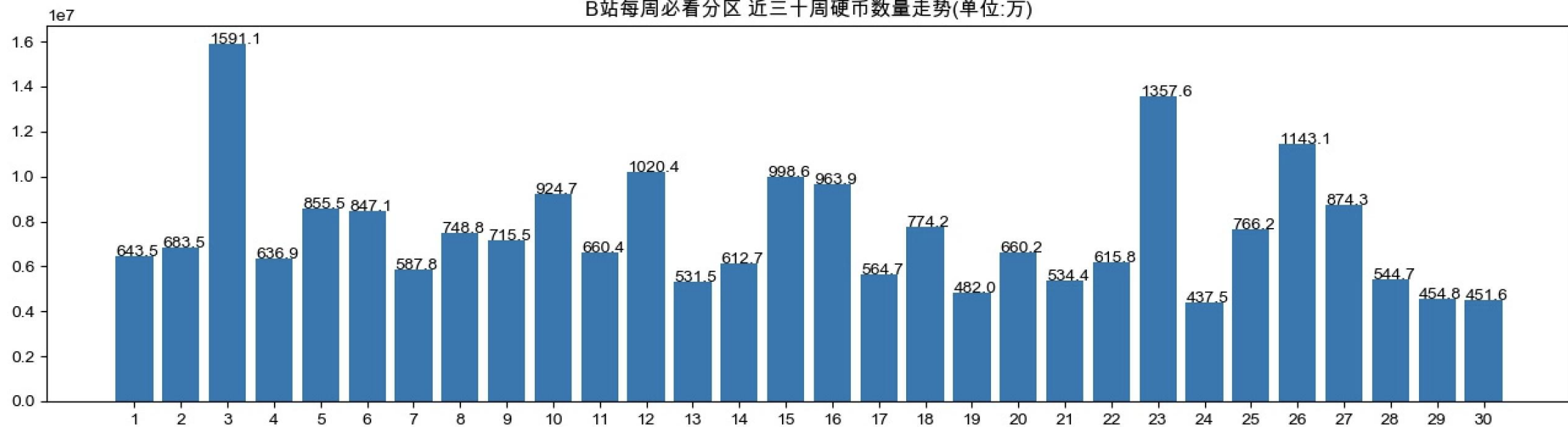
B站每周必看分区 近三十周评论量走势(单位:万)



B站每周必看分区 近三十周收藏走势(单位:万)



B站每周必看分区 近三十周硬币数量走势(单位:万)



04

分区热播统计

分区热播统计——数据

- 时间区间：20220202~20220302
- 抽取每个分区一个月内，播放量top(50)的视频

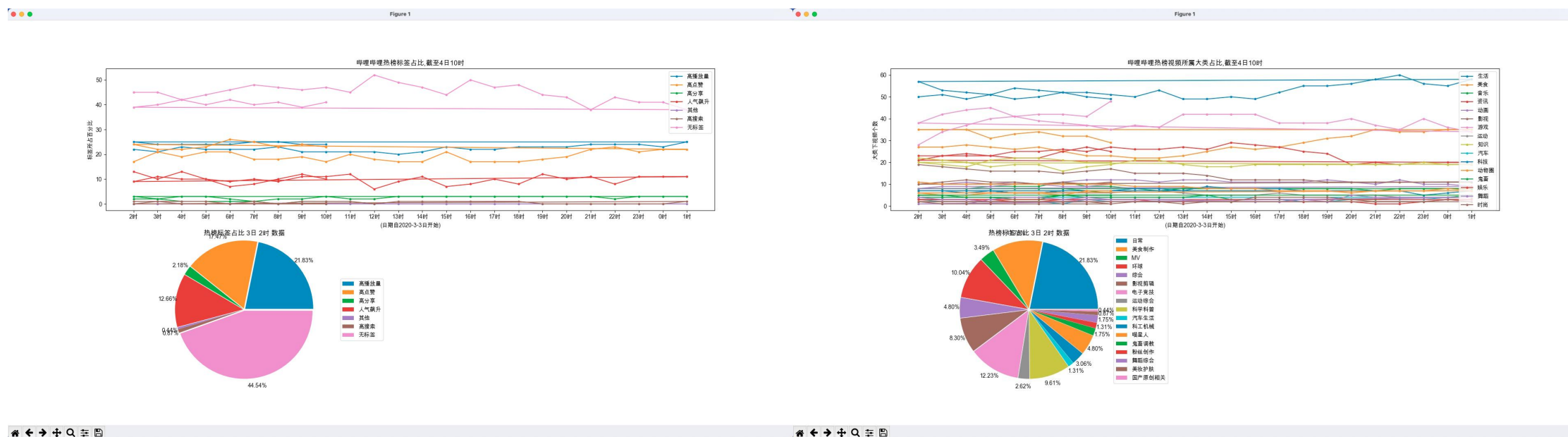
分区热播统计——可视化

05

全站视频实时排行榜可视化

全站视频实时排行榜可视化

- 从全站热播定时爬取热播视频列表
- 爬取时长间隔为五分钟，从20220303到20220304
- 对于热播视频列表，重点关注其中“视频所处分区占比”和“热播推荐标签”
- 通过折线图展示以上两组数据在时间上的变化
- 以饼图的形式展示指定时间点的比例情况



06

热点：俄乌冲突

热点：俄乌冲突

数据获取：利用b站关键词搜索网页，搜索俄乌，再对html数据进行解析获得。搜索url

(https://search.bilibili.com/all?keyword=俄乌&from_source=nav_suggest_new0&page=1) 其中，keyword为搜索关键字，page为页数，b站上限一共可搜索获得50页视频数据。

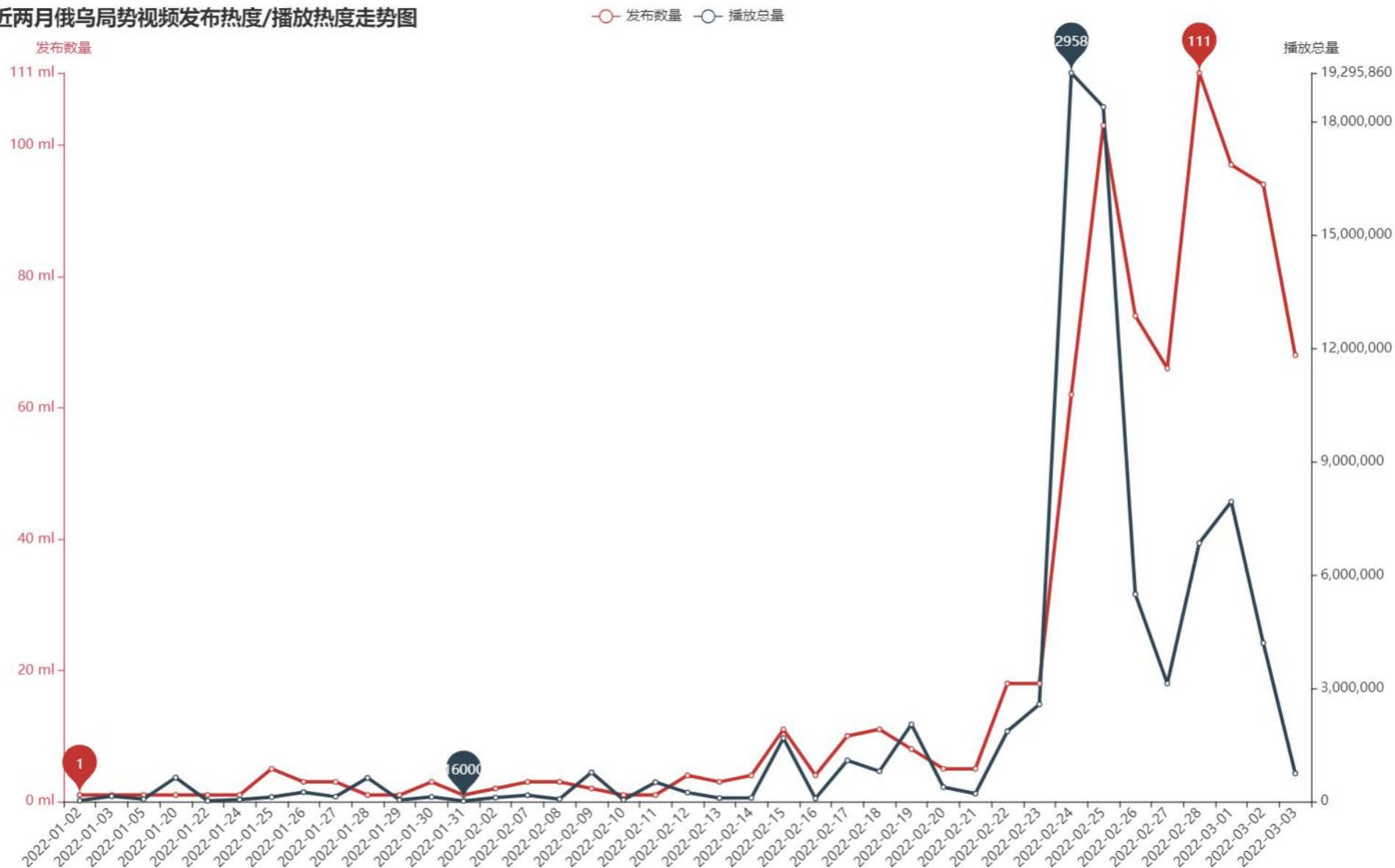
可视化：

- 1、统计近两个月俄乌视频每天的总播放量和视频发布数量，以获得热度趋势。
- 2、统计近两个月俄乌视频发布数量在每个分区的占比
- 3、统计近两个月俄乌视频在中每个分区中的总播放量

```
<!-->
<ul type="video" class="video-list clearfix">
  <li class="video-item matrix">
    <a href="//www.bilibili.com/video/BV1WZ4y1k7aZ?from=search&seid=17305165014339299071" title="【卢克文工作室】奇怪的俄乌冲突，背后暗藏怎样的大国博弈？" target="_blank" class="img-ancho
r">...</a>
    <div class="info">
      <div class="headline clearfix">...</div>
      <div class="des hide"> 这一次俄乌冲突，到最后得益的，可能是美国和俄罗斯，但失去最多的，还是乌克兰这个傀儡国家。 </div>
      <div class="tags">
        <span title="观看" class="so-icon watch-num">...</span>
        <span title="弹幕" class="so-icon hide">...</span>
        <span title="上传时间" class="so-icon time">...</span>
        <span title="up主" class="so-icon">...</span>
      </div>
    </li>
  </ul>
```

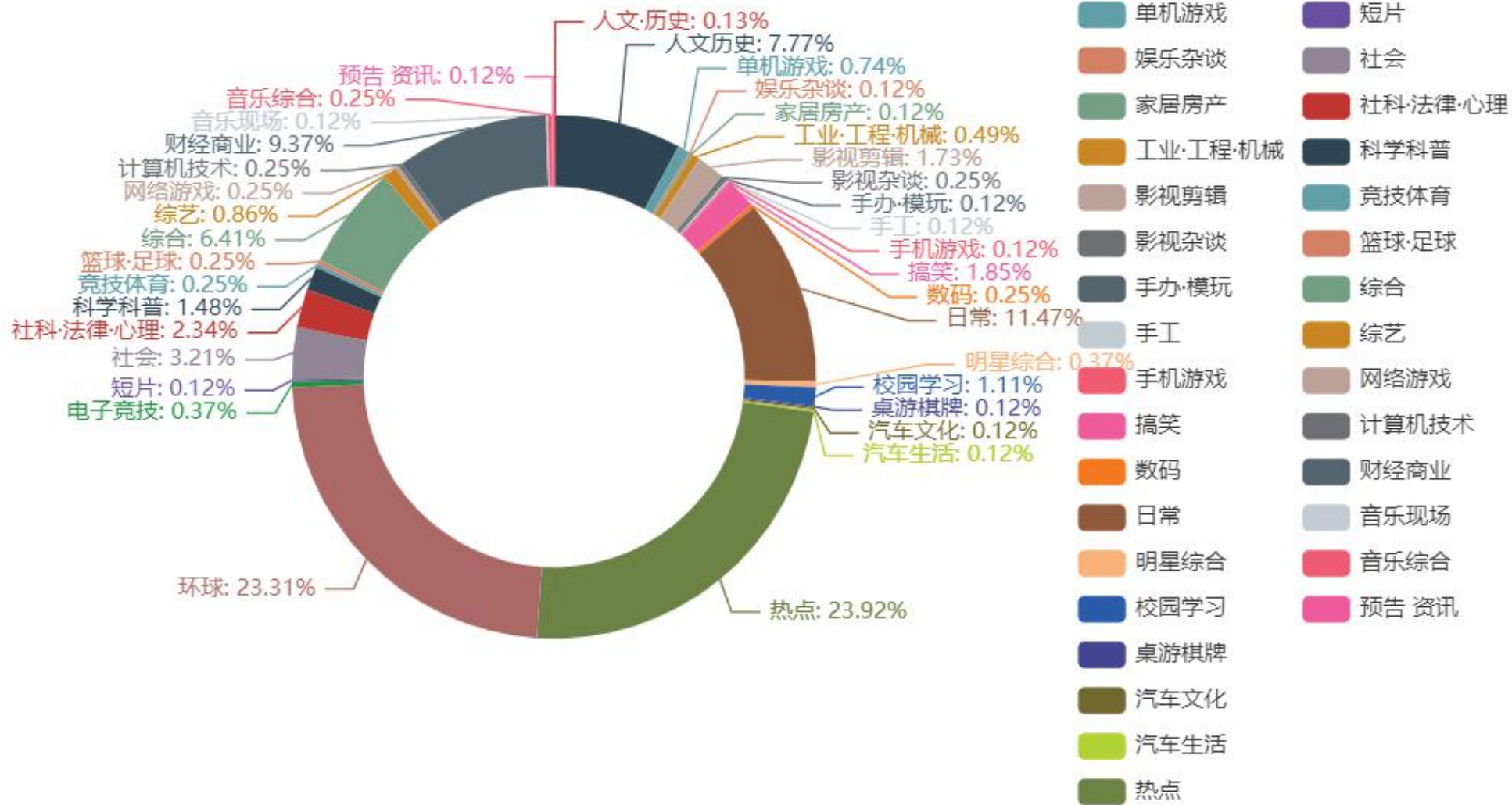
热点：俄乌冲突

近两月俄乌局势视频发布热度/播放热度走势图



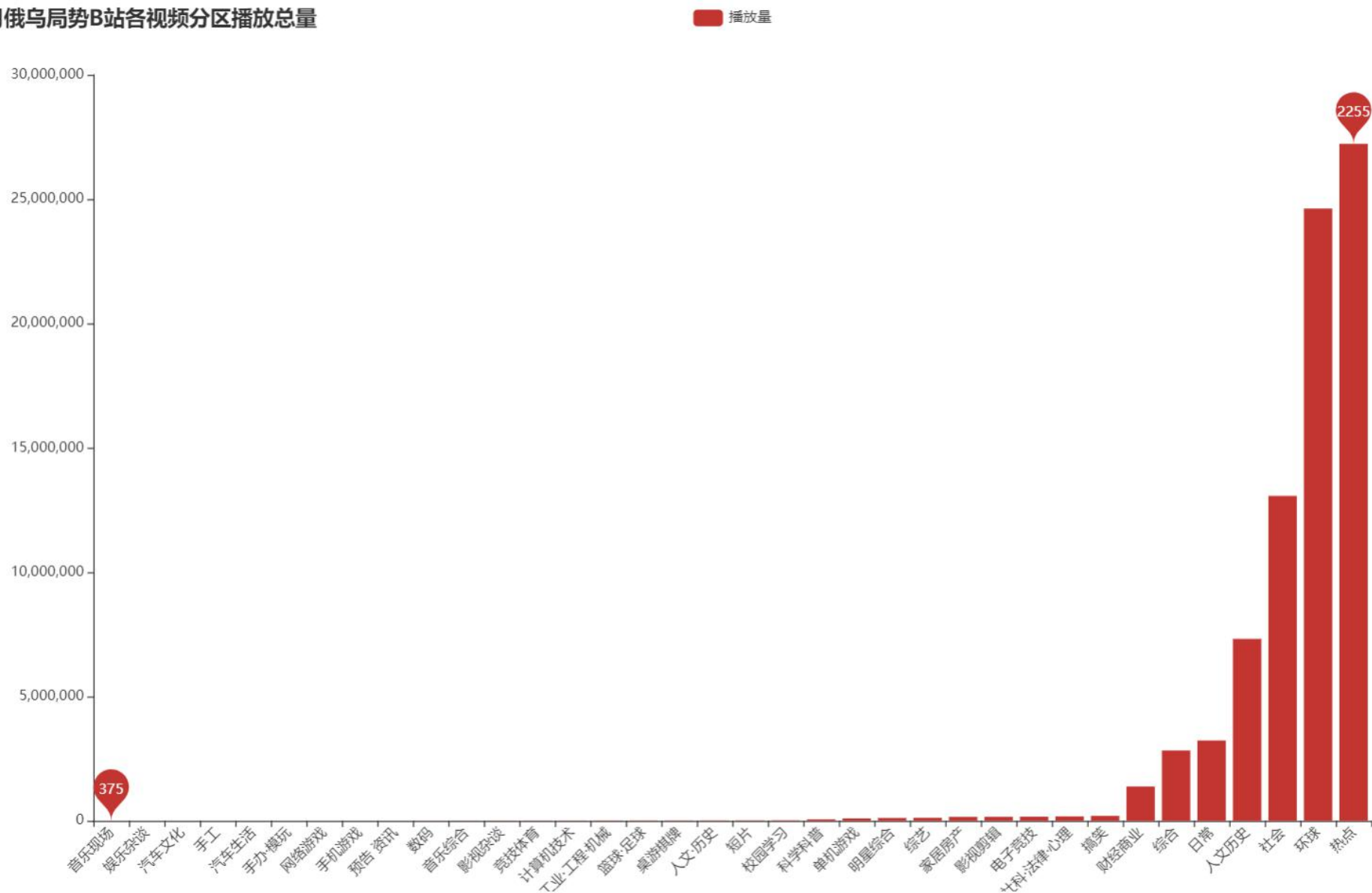
热点：俄乌冲突

近两月俄乌局势B站各视频分区发布数量



热点：俄乌冲突

近两月俄乌局势B站各视频分区播放总量



回顾

- 时间紧促
- 在有限的时间内爬取到尽可能多的有效数据。
- 没有针对反爬虫机制进行代码改进
- 体验 入门

Q&A

THANK YOU