

twitterR包入门和应用

肖凯

twitter



提要：

- 什么是Twitter
- twitteR包简介
- 四个应用示例

什么是Twitter

- Twitter（推特）是一个社交网络及微博服务网站；
- 用户可以在任何时间、任何方式发表任何信息，分享给希望获知这些信息的人；
- 在第一时间、任一方式获知你所关注的人（或事）的最新进展。
- 为什么要研究Twitter数据？

Twitter基本术语

- status (推文) : 发布的信息
- follow (关注) : 跟随他人的动作
- following (朋友) : 我关注的人
- followers (粉丝) : 关注我的人

twitterR包简介

- 可以在R中使用twitter API
- 两类主要对象
 - User : `getUser(user, ...)`
 - Status : `searchTwitter(); userTimeline()`

User对象

- 包含的主要信息：

Name ; lastStatus ; statusesCount ;

followersCount ; friendsCount ;

Created ; location

- 可使用的方法:

toDataFrame ; getFollowers ; getFriends

Status对象

- 包含的主要信息：

text ; screenName ; replyToUID

statusSource ; created ; favorited

- 可使用的方法:

toDataFrame ;

应用示例

- 获取某个用户的信息
- 分析某个用户的推文
- 分析包含某个关键词的推文
- 分析某个用户群的特征

示例1：获取某个用户的信息

```
myid <- getUser('xccds')
```

```
> myid$name
```

```
[1] "写长城的诗"
```

```
> myid$lastStatus
```

```
[1] "Unknown: 5 Ways Big Data Is Transforming  
Everyday Life http://t.co/8lcNOPca via @zite"
```

```
> myid$description
```

```
[1] "数是万物的本源"
```

示例1：获取某个用户的信息

```
> myid$statusesCount
```

```
[1] 1114
```

```
> myid$followersCount
```

```
[1] 172
```

```
> myid$friendsCount
```

```
[1] 271
```

```
> myid$created
```

```
[1] "2010-03-14 03:38:36 UTC"
```

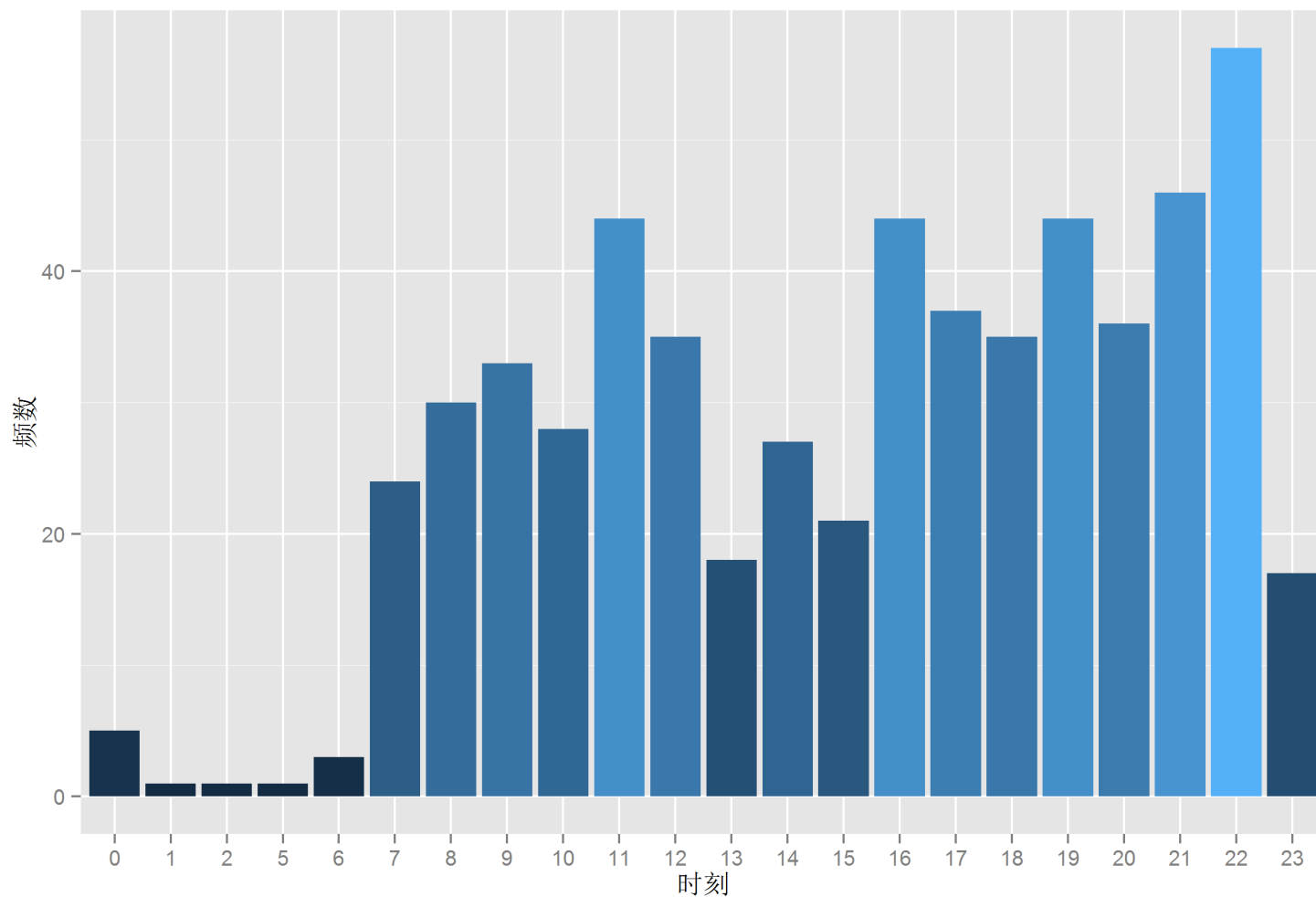
示例1：获取某个用户的信息

```
> xccds <- userTimeline("xccds", n=500)
```

```
> xccds[[1]]$getCreated()
```

```
[1] "2012-10-07 08:31:05 UTC"
```

发推的时间规律



示例2：分析某个用户的推文

从这两个帐户获取空气质量信息

@BeijingAir @CGShanghaiAir

```
airb <- userTimeline("beijingair", n=1000)
```

```
airs <- userTimeline("CGShanghaiAir", n=1000)
```

```
[1] "BeijingAir: 10-15-2012 10:00; PM2.5; 80.0; 159;  
Unhealthy (at 24-hour exposure at this level)"
```

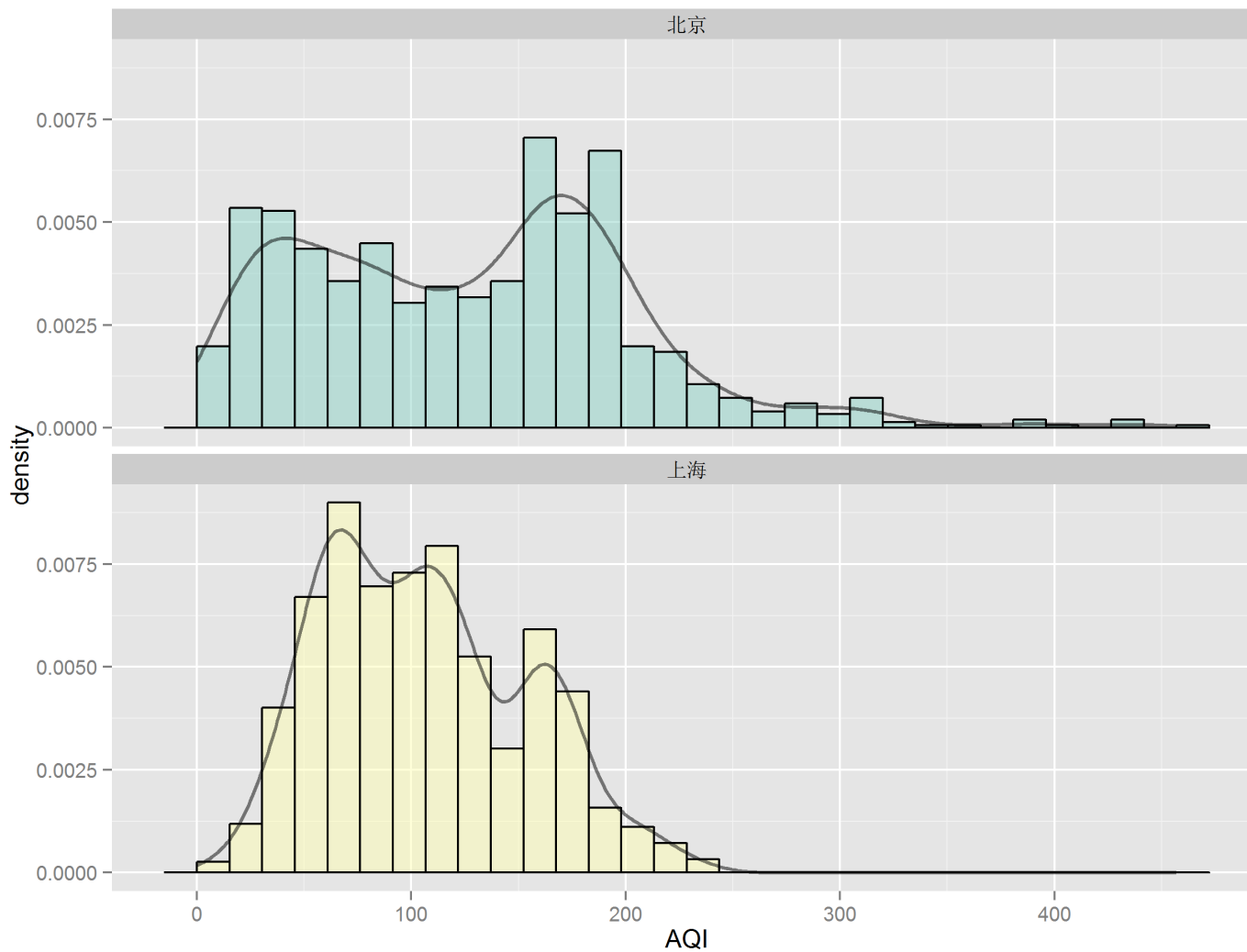
示例2：分析某个用户的推文

time	AQI	type	city	week	clock
10-13-2012 10:00	32	Good	北京	6	10
10-13-2012 09:00	29	Good	北京	6	9
10-13-2012 08:00	32	Good	北京	6	8

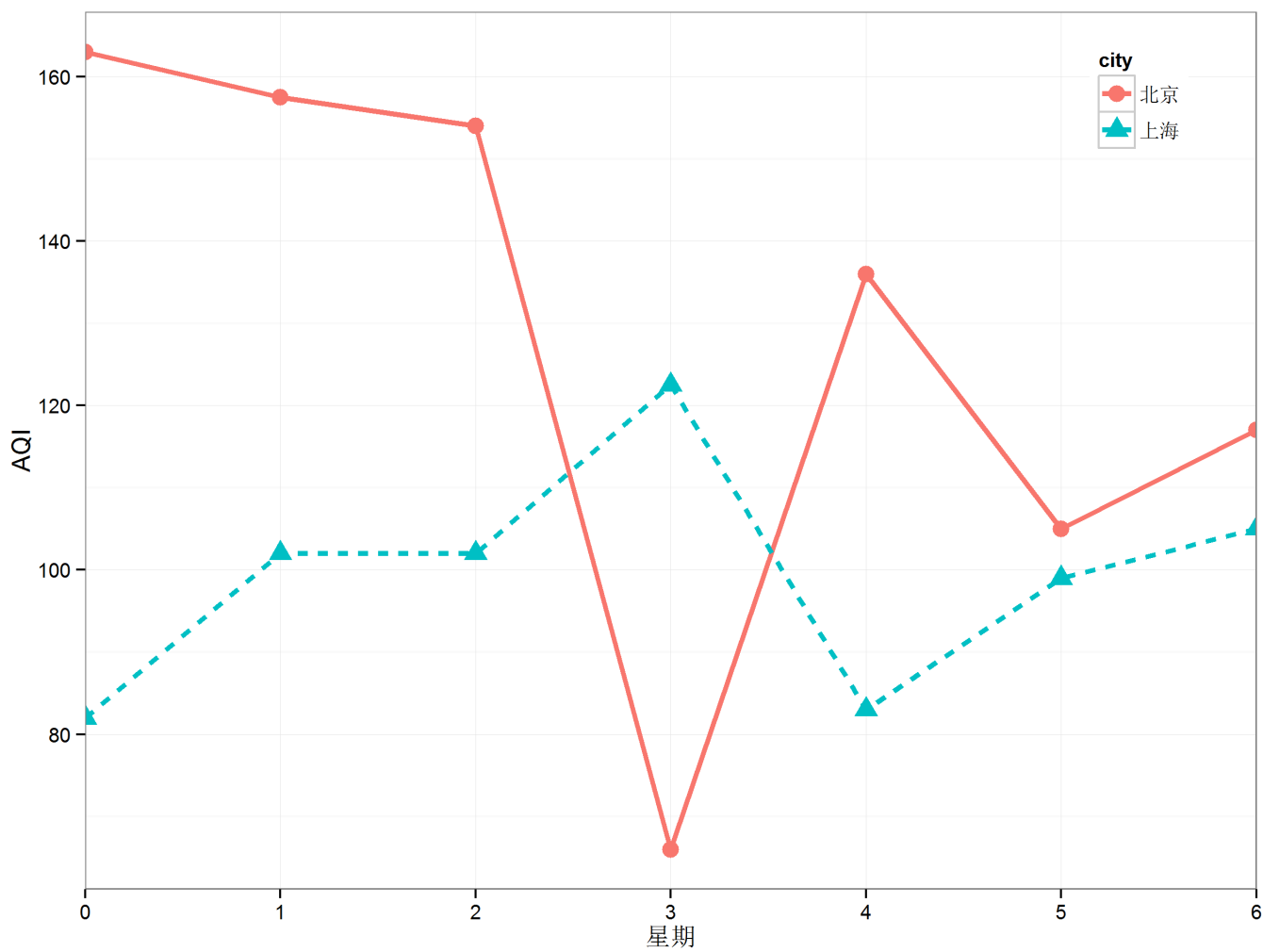
北京和上海AQI比较



北京和上海AQI比较



AQI在不同日子的中位数

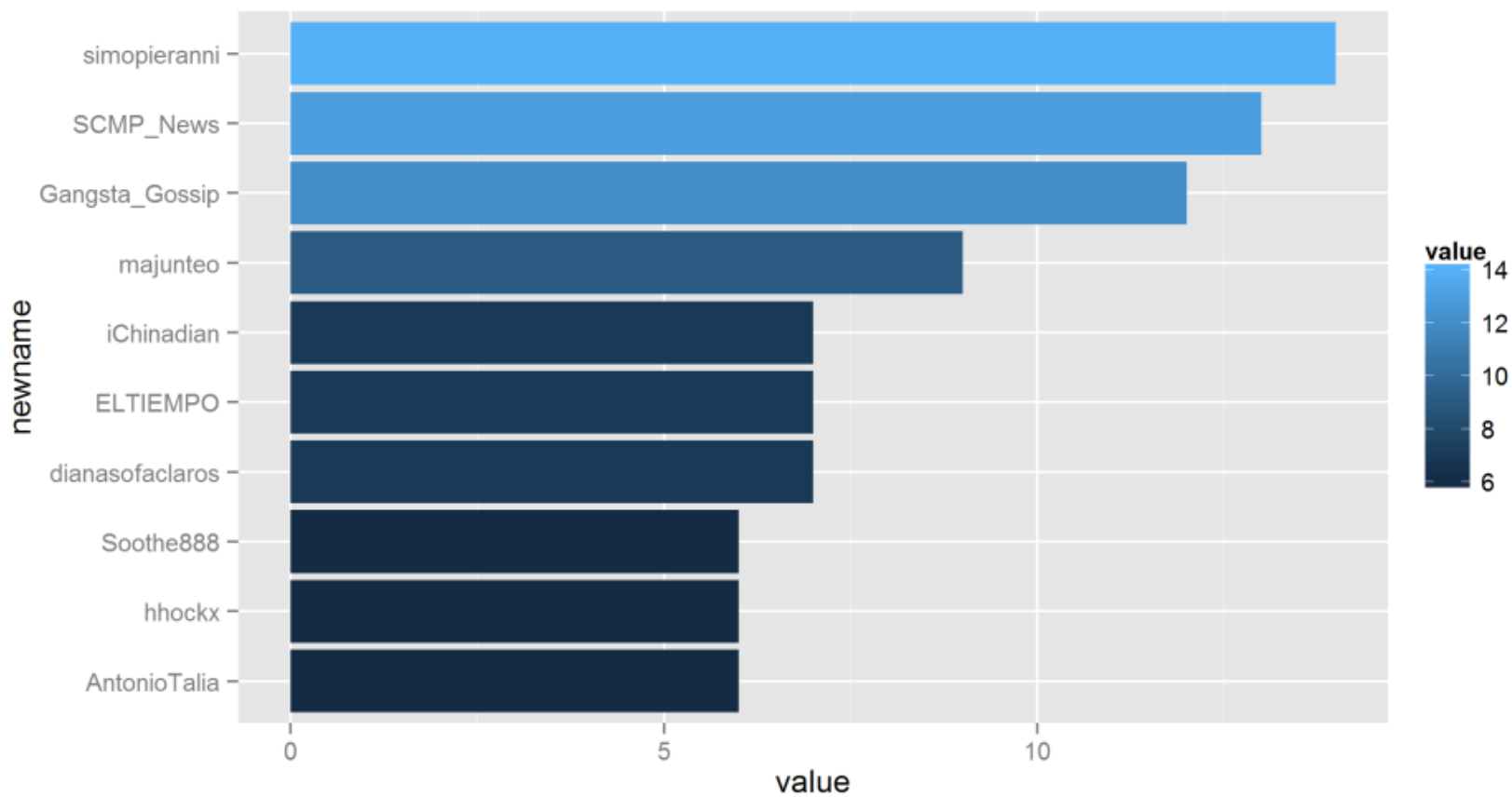


示例3：分析包含某个关键词的推文

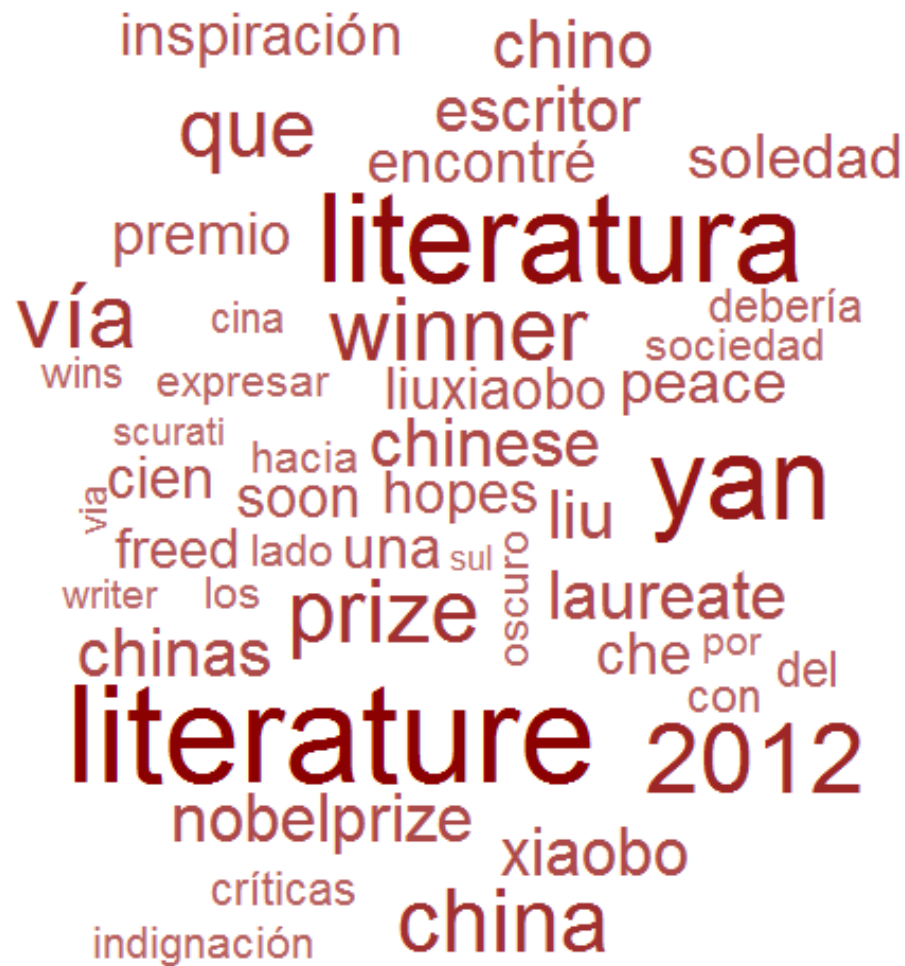
关于莫言获得文学奖，twitter上都有什么样的反应？

```
moyanTweets <- searchTwitter('#moyan', n = 1000)
```

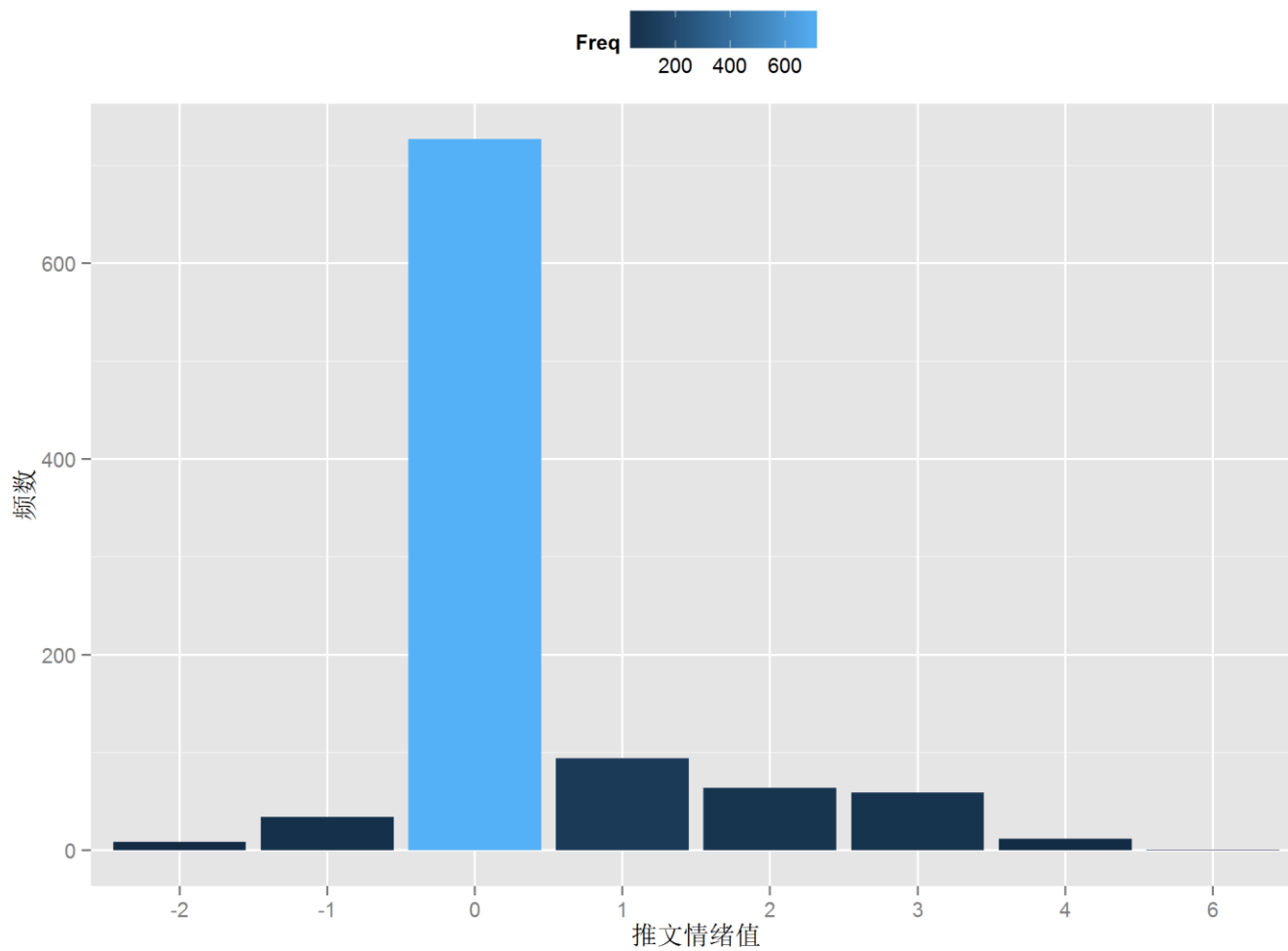
哪些用户关心这个词



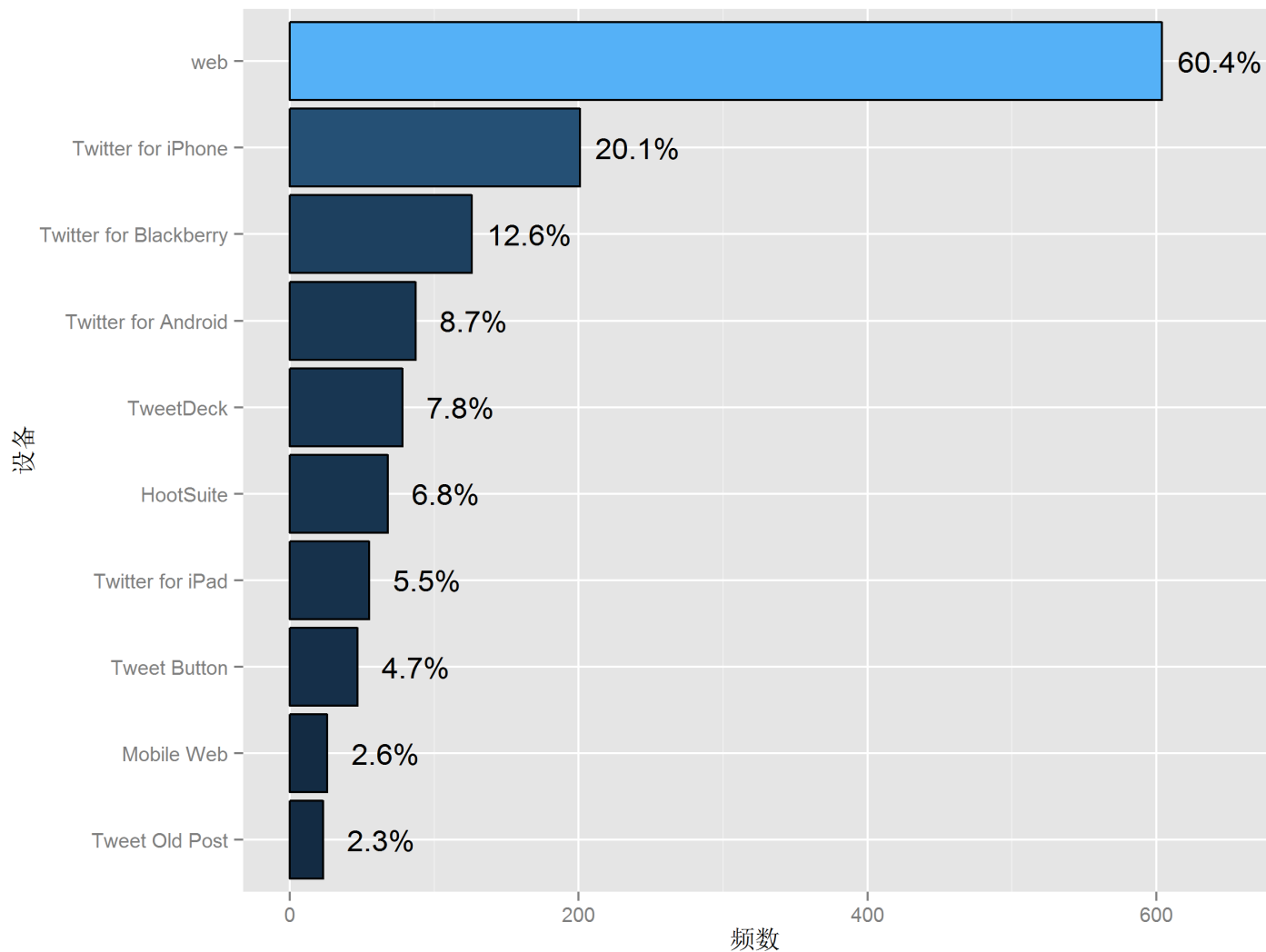
这些推文中包含的内容



分析这些推文的情绪



统计这些推文的发推来源



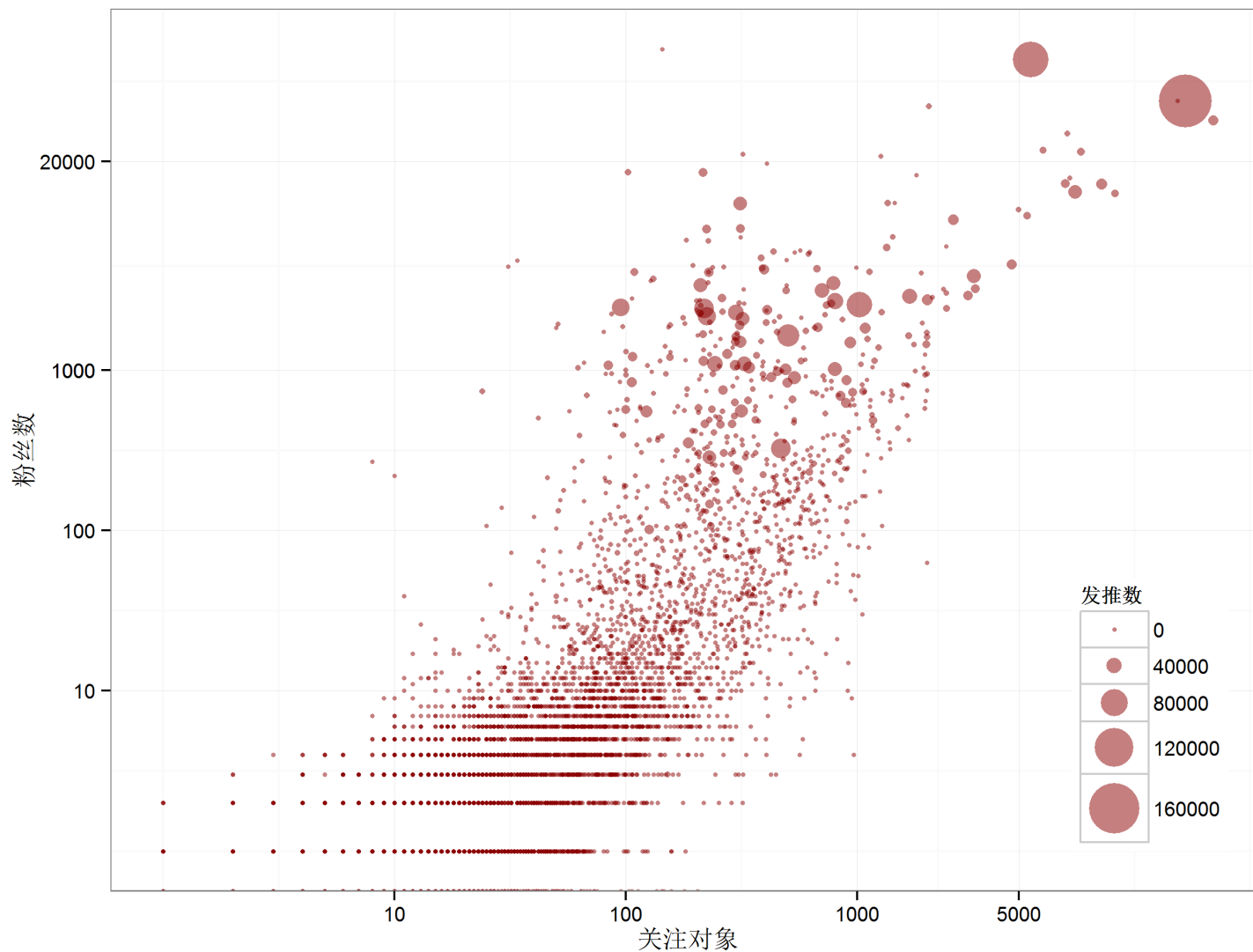
示例4：分析某个用户群的特征

获取ladygaga在twitter上的5000名粉丝

```
ladygaga <- getUser('ladygaga')
```

```
follow.lady <- ladygaga$getFollowers(n=5000)
```

示例4：分析某个用户群的特征



其它

- 研究新浪微博数据可使用李舰的Rweibo包
- 使用twitterR包需要翻墙
- 由于API的限制，最好配合ROAuth包
- 示例代码均可在本人博客或GitHub找到
 - xccds1977.blogspot.com
 - <https://github.com/xccds/my-r-code/blob/master/shanghai.R>
- 欢迎关注@xccds

谢谢

