#100DaysOfCode

Group Name

Zooloretto

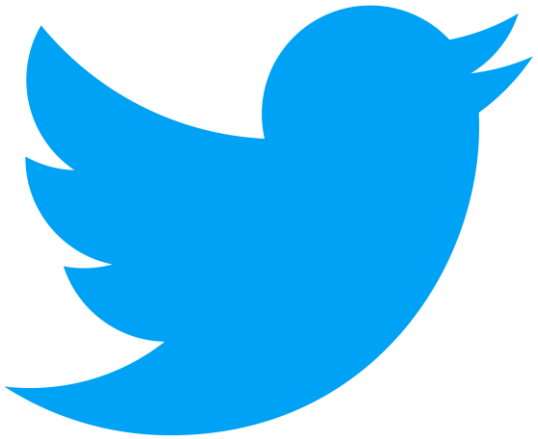
Group Members

CHEN Xiaoqi 20465106(Leader) / CAI Runlin 20426550 / GUO Yuju 20465769 / HUANG Zefei 20449496

AIDM7390 Data Mining and Knowledge Discovery for Digital Media

“Are you struggling with code?”

AGENDA



- WHY DO WE WANT TO DO
- HOW DO WE START
- WHAT DID WE FIND
- CONCLUTION

“What is the most popular programming language?”

INTRODUCTION

AI has become a hot topic in society. More and more people pay attention to learning programming, and even self-learning code, hoping to hold the opportunity of computer science trend to improve their thinking ability and find a high-paying career.

INTRODUCTION

#100DayOfCode

1. Code minimum an hour every day for the next 100 days.
2. Tweet you are progress every day with the #100DaysOfCode hashtag.

INTRODUCTION

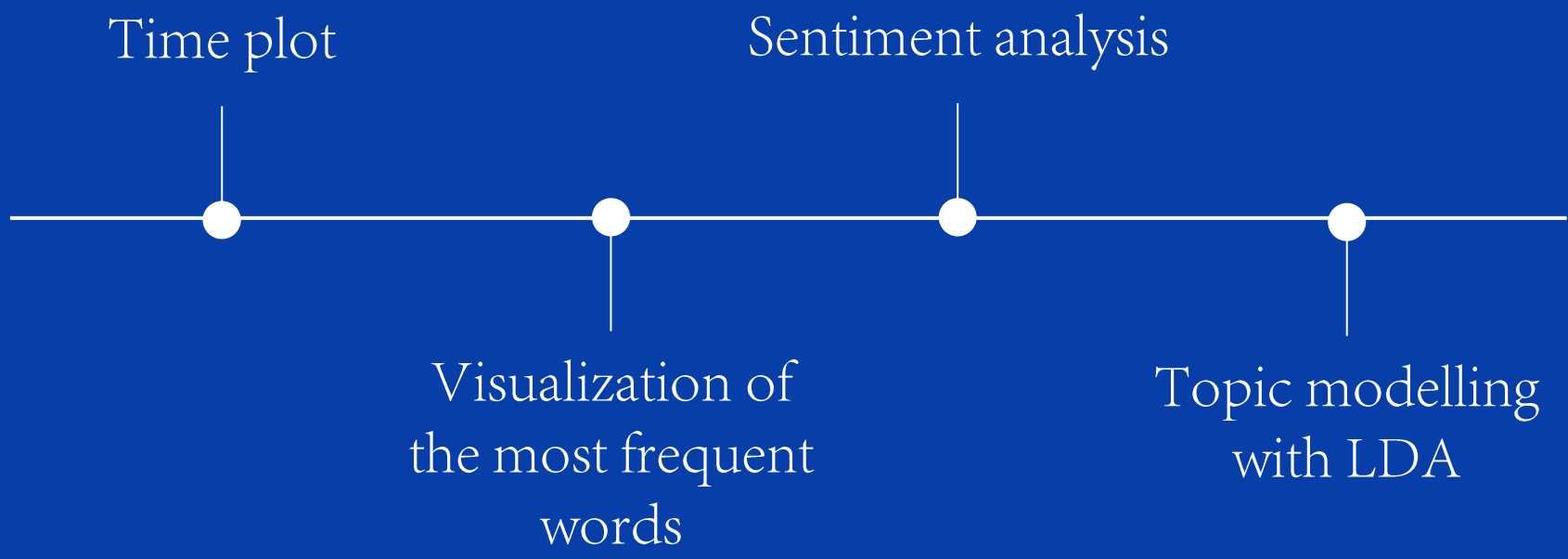
#100DaysOfCode is like a large community on which people record their daily coding learning and moods.

We want to explore what aspects of code the public is interested in. We hope to explore what programming are those people who are interested in computer science learning or what programming they want to learn, and what specific content they are interested in.

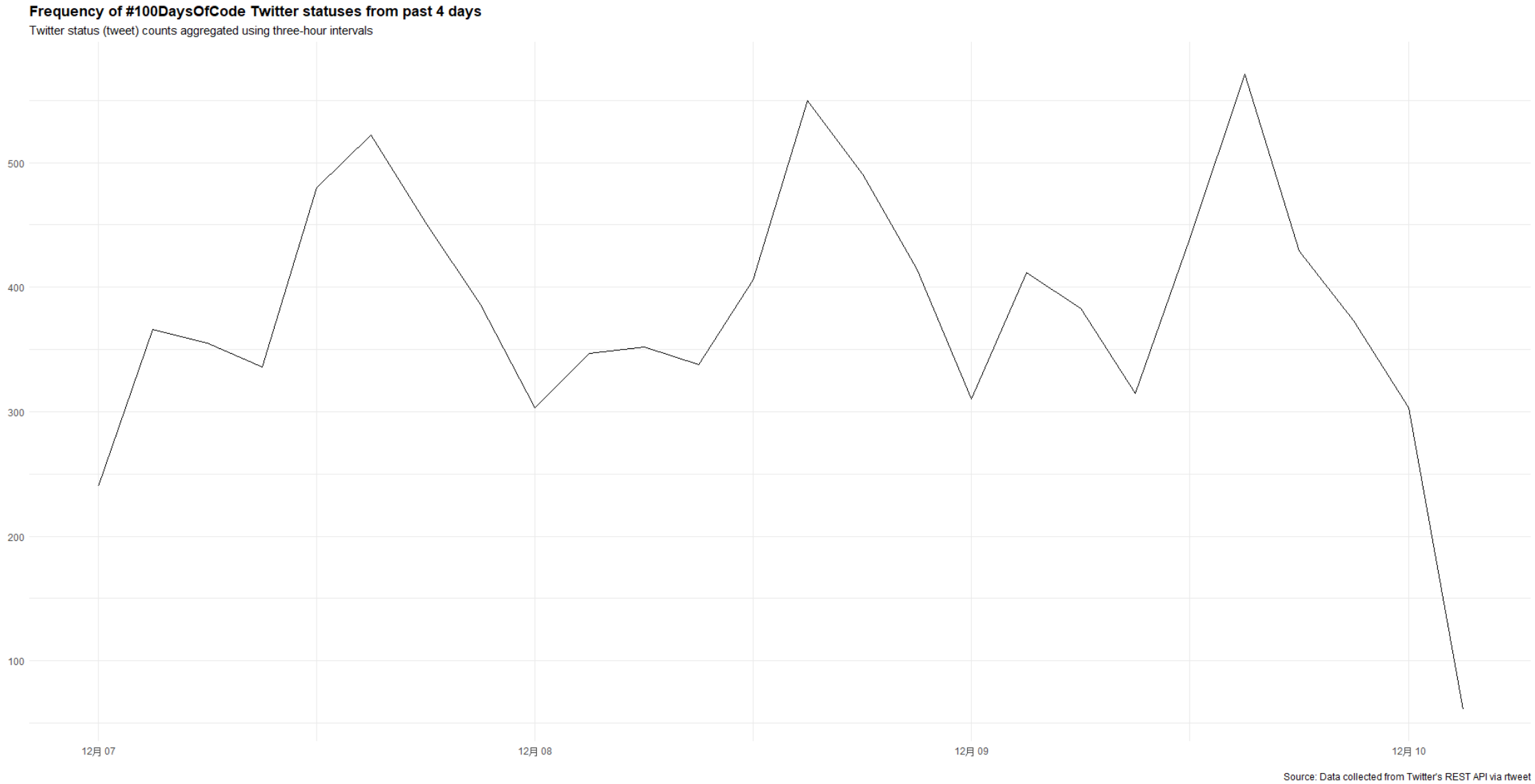
PROCEDURES

```
install.packages('rtweet')
library('rtweet')

token <- create_token(app = 'Zooloretto',
                      consumer_key = 'BpRSq1TOabc2RsBOlRV3HQraB',
                      consumer_secret = 'NdSZNKMgcSS23lvXV4zHPWaqmpO4DptBLh777bK1c7wbVTDPul',
                      access_token = '1327201654018490368-IuKsF3YAZmMCInO87m5iqVdu4MnjdM',
                      access_secret = '7hA7xB6N49fcBmv8ux2ui0aQQiLQZteikrjZoS1PWVfkt',
                      set_renv = TRUE)

Code<- search_tweets('#100DaysOfCode',n=10000,include_rts = FALSE,lang='en')
Code
```

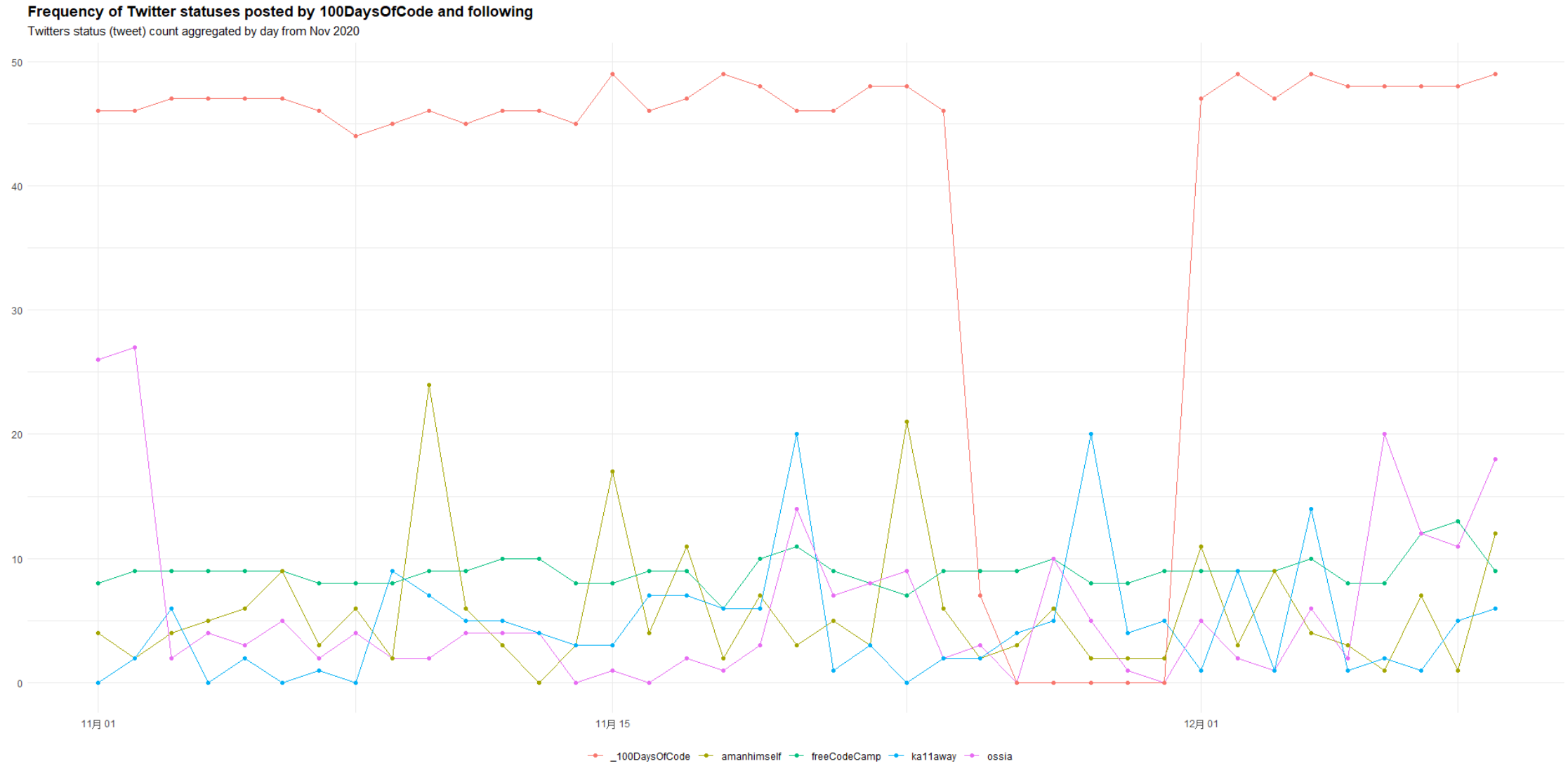
PROCEDURES


Time plot
Visualization of the most frequent words
Sentiment analysis
Topic modelling with LDA

VISUALIZATION



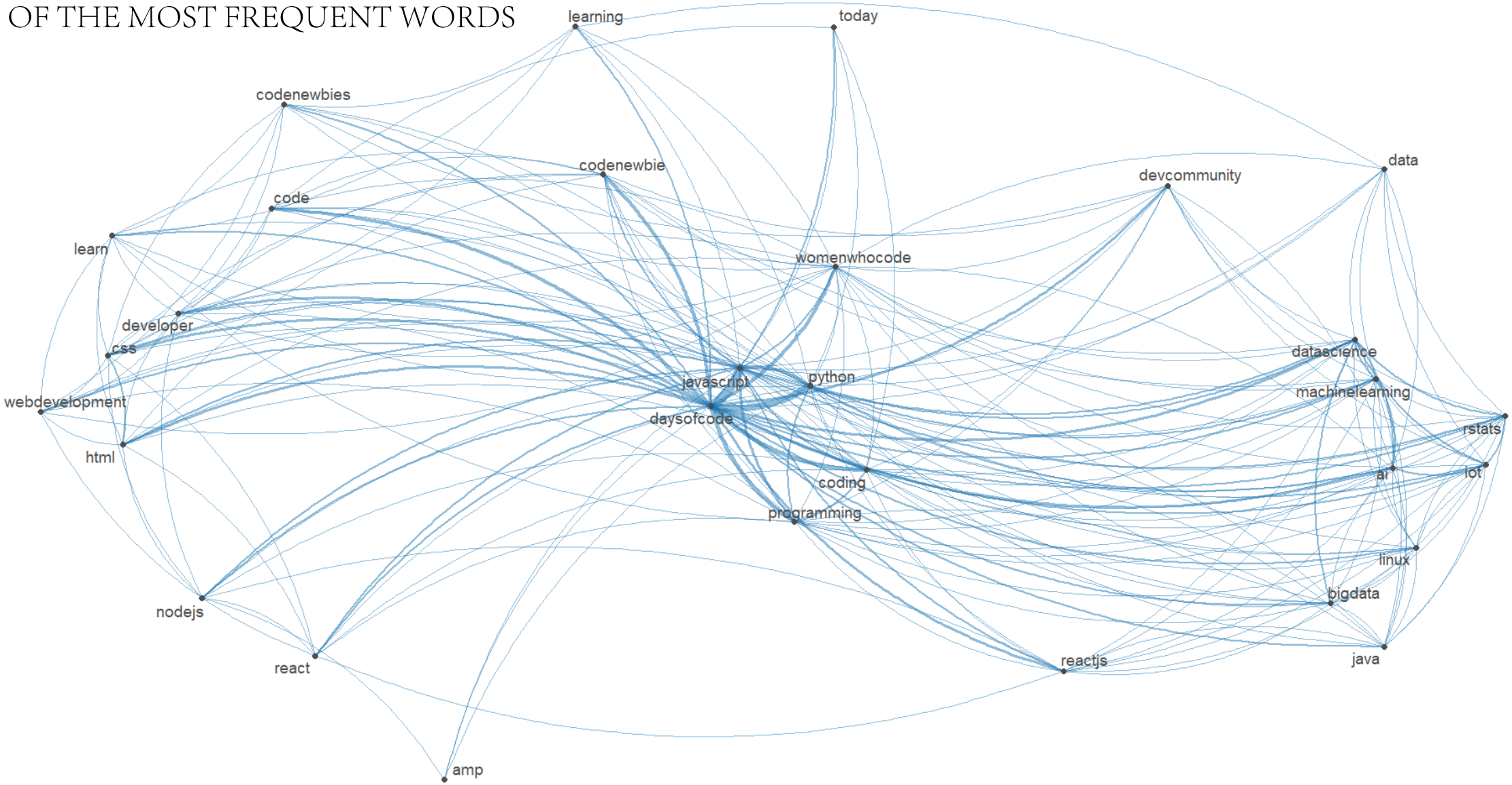
VISUALIZATION



Source: Data collected fron Twitter's REST API via rtweet

VISUALIZATION

OF THE MOST FREQUENT WORDS

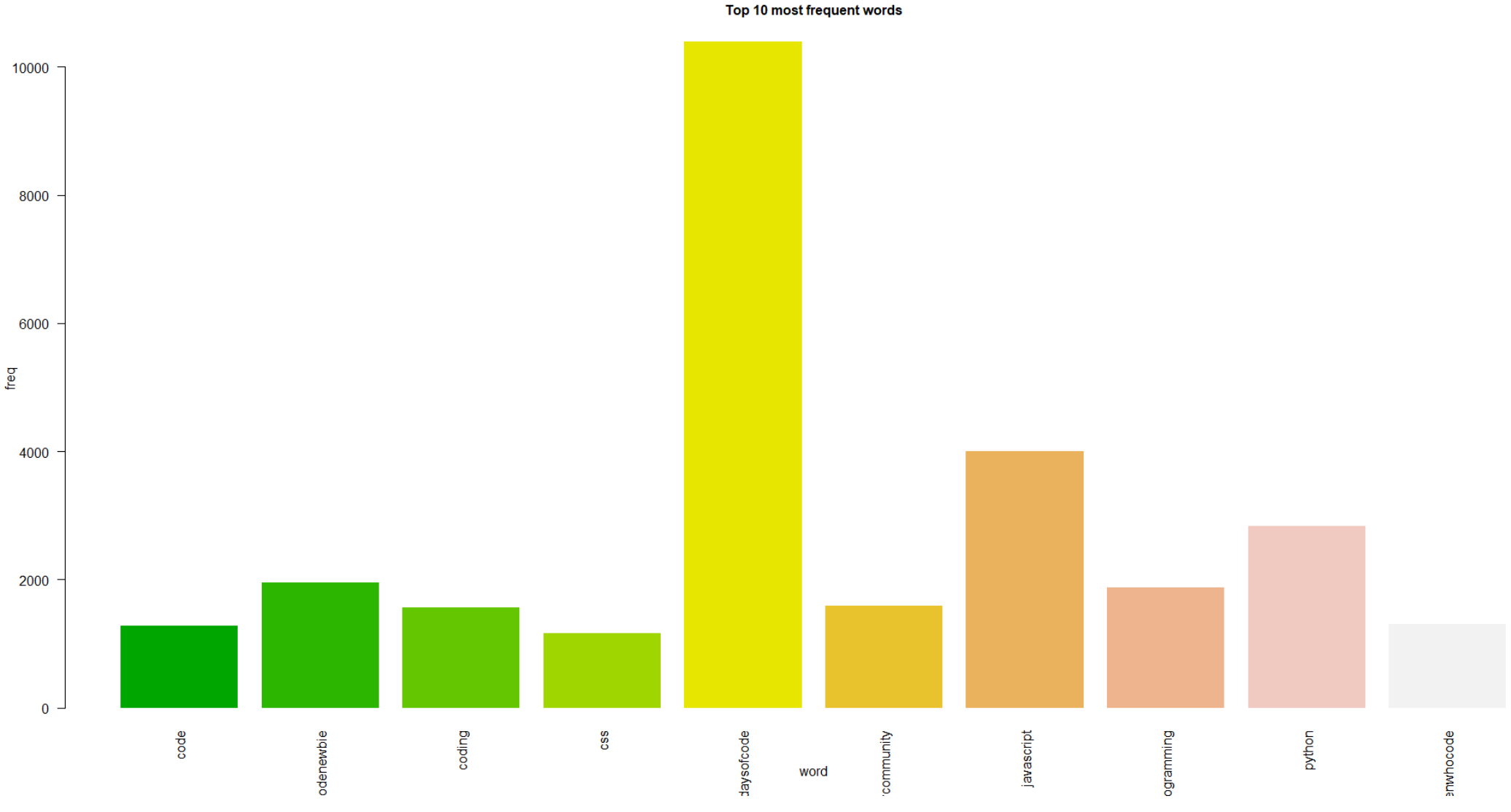


VISUALIZATION

OF THE MOST FREQUENT WORDS

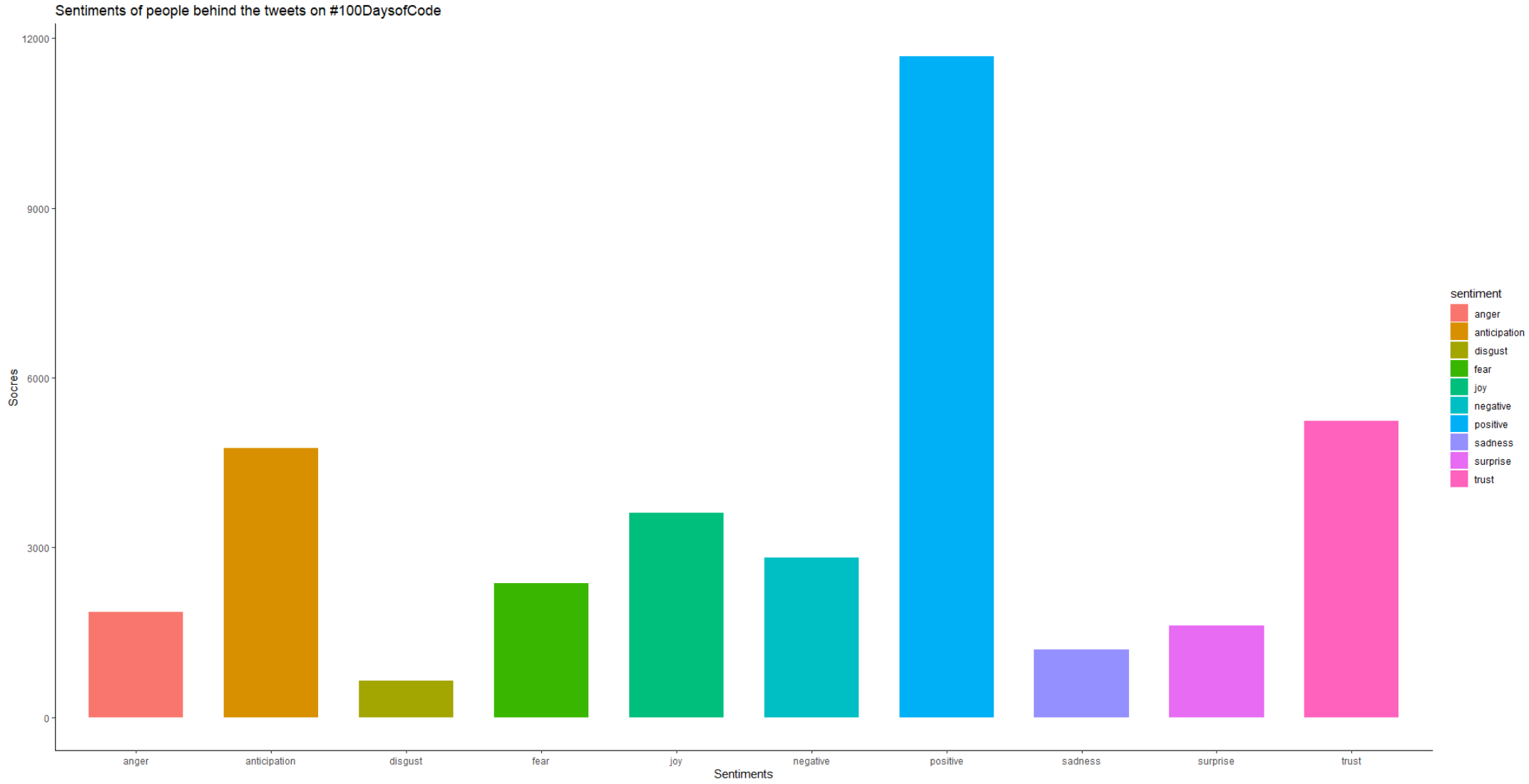
VISUALIZATION

OF TOP 10 MOST FREQUENT WORDS



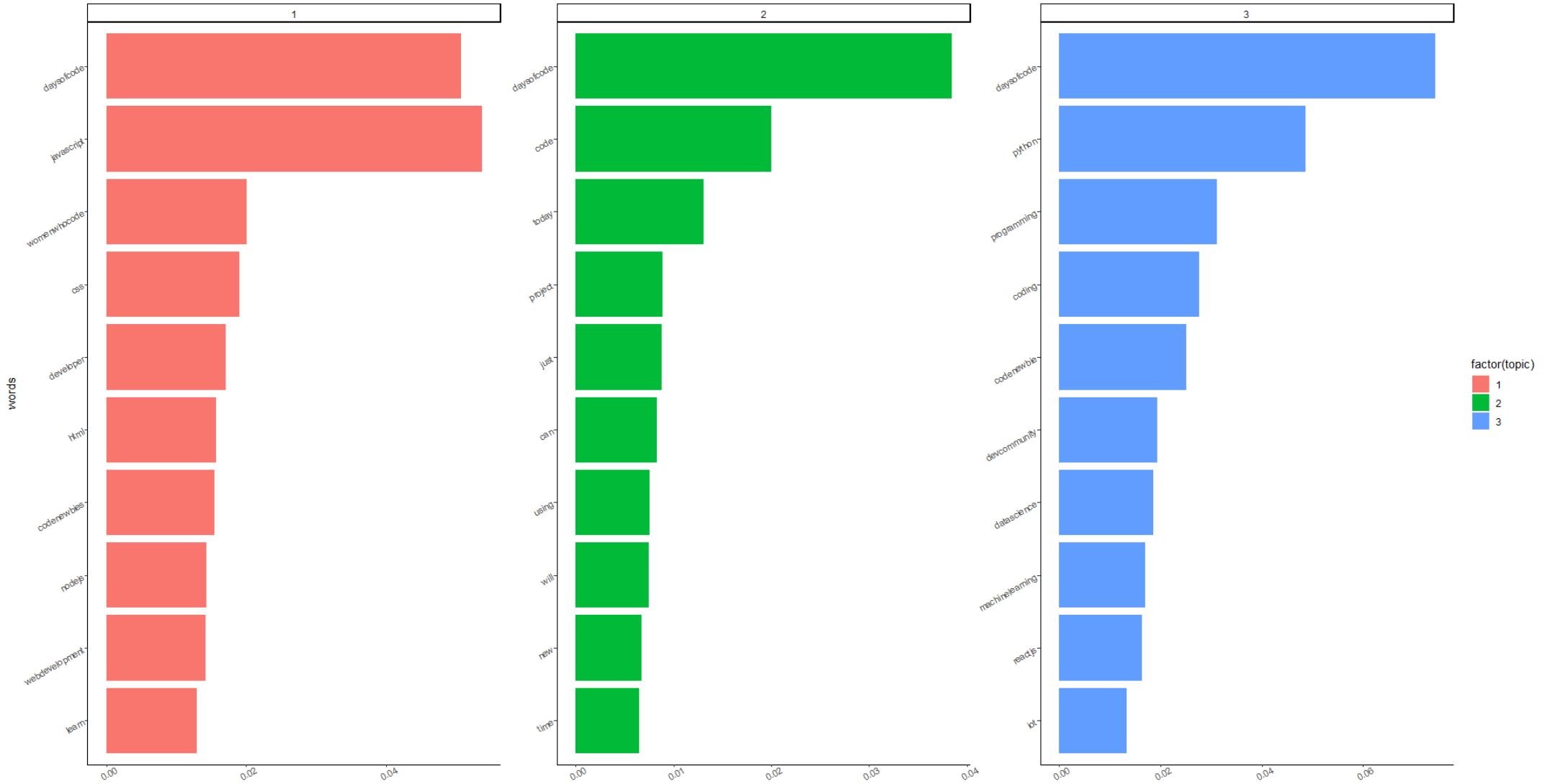
VISUALIZATION

OF SENTIMENT ANALYSIS

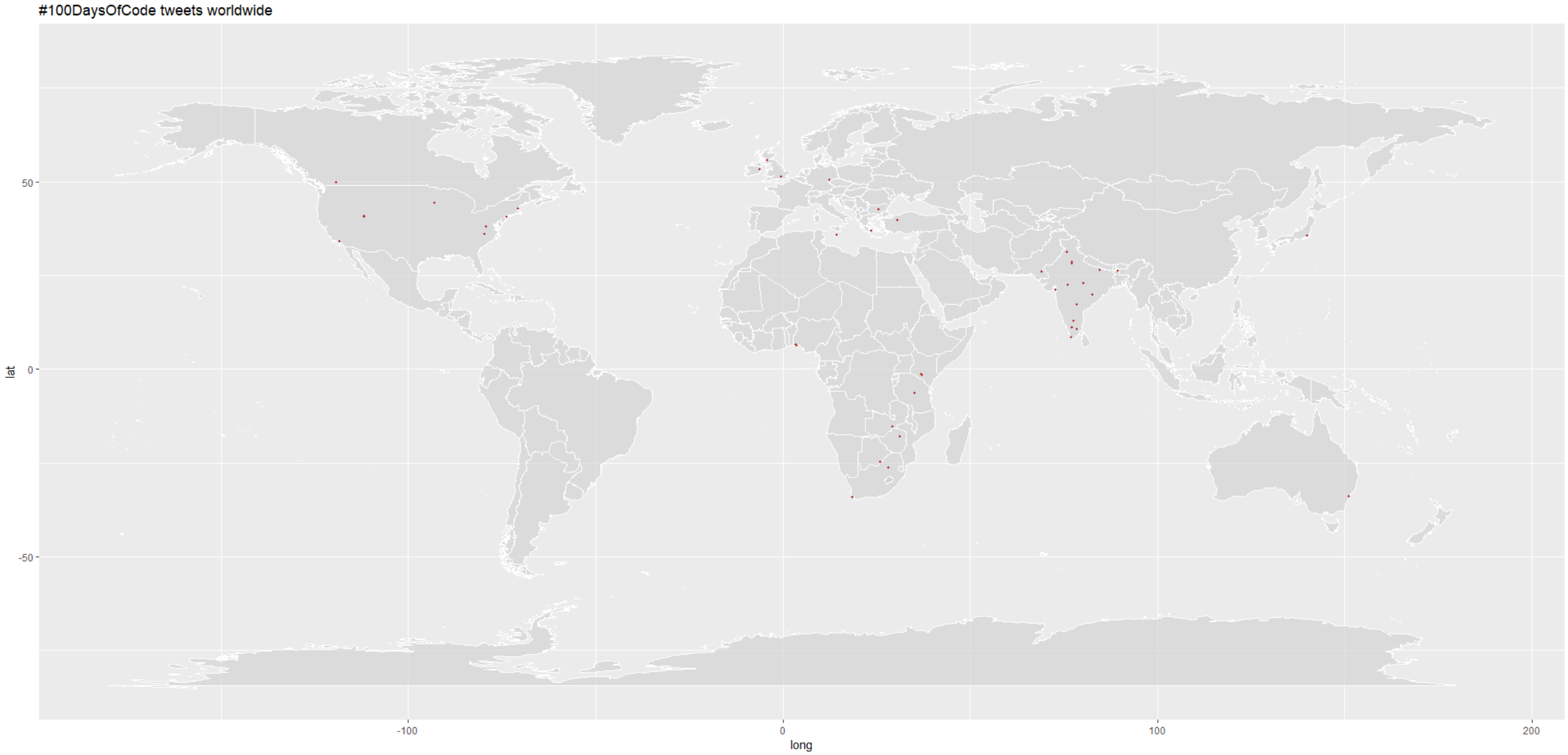


VISUALIZATION

OF TOPIC MODELLING WITH LDA



VISUALIZATION



CONCLUSION

CONCLUSION

Most people are interested in JavaScript and Python, and many of them are working on development, projects, or machine learning.

We found that a lot of female coder also actively tweets on this topic, indicating that the gender gap in the industry may be gradually decreasing.

CONCLUSION

Most of the people involved in hashtags maintain an optimistic and positive attitude towards code or communication. People who are enjoying fun and looking forward to learning than those people in a state of confusion or sadness.

CONCLUSION

· Limitations and Improvements ·

According to the current word cloud, we still cannot judge whether users belong to code learners or workers, nor can we see their specific employment industries.

We need to do single-person Twitter content analysis for more users, so that we can analyze the individual in a comprehensive way, then we can know their learning purpose or employment direction.

“What is the best programming language do you think ? ”

THANKS

Group Project - Zooloretto

AIDM7390 Data Mining and Knowledge Discovery for Digital Media