Evolution of Stack Overflow Discussions
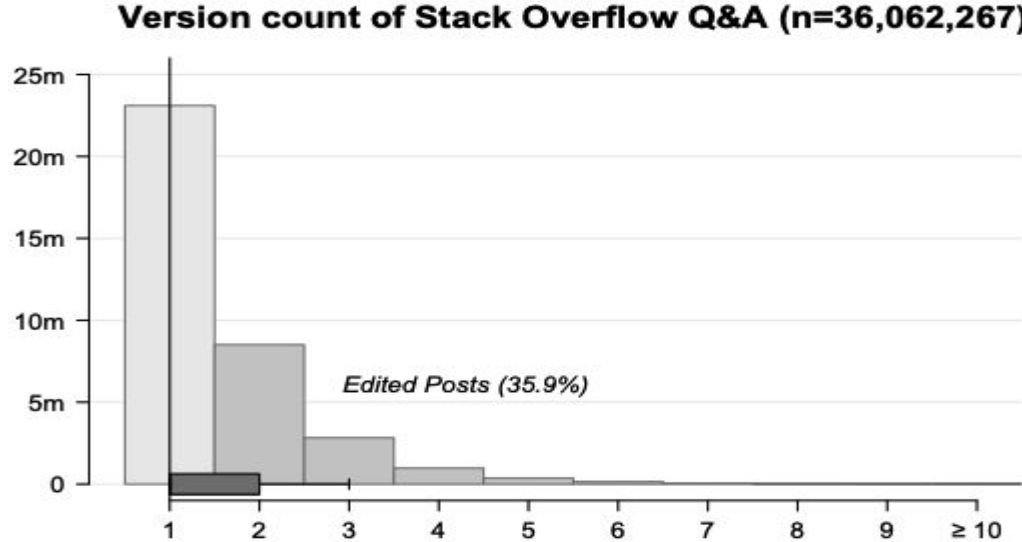
Using Sentimental Analysis on Comments in Stack Overflow Posts

Presented by Norbert Eke and Saraj Manes
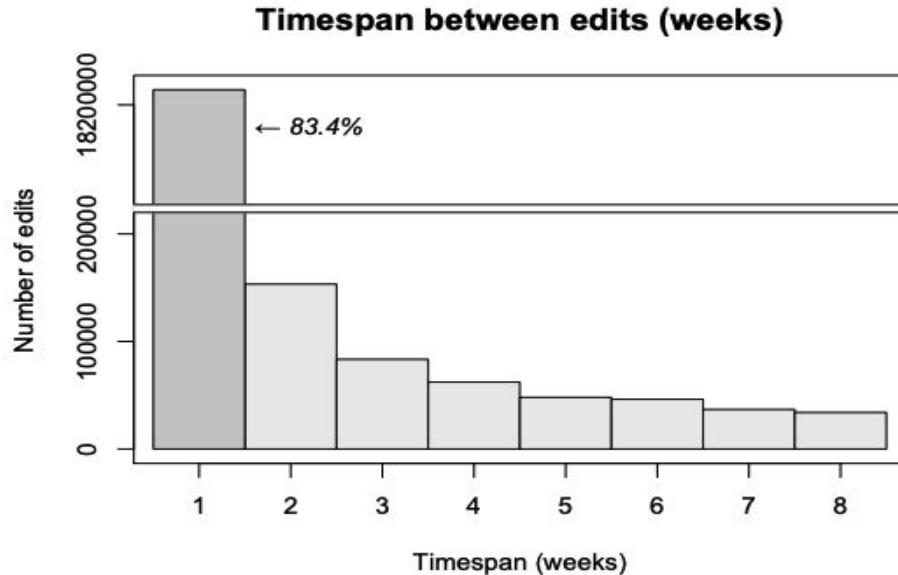
Motivation

- SOTorrent 2018 : Reconstructing and Analyzing the Evolution of Stack Overflow Posts.

Version count of Stack Overflow Q&A (n=36,062,267)

Edited Posts (35.9%)

Motivation

- SOTorrent 2018 : Time Span between edits.

Timespan between edits (weeks)

- "If we focus on the comments, we see that 64.4% of them happened on a day where the post had either been created or edited."

Motivation

Questions:

Is there any relation between Comments and edits?
Does one cause the other?

Do edits cause more number of comments
OR
Do comments lead to edits?

Our Previous Work

How Often and What StackOverflow Posts Do Developers Reference in Their GitHub Projects?

- What type of posts are most popular/referred in GitHub projects
- How many versions for referred posts

Data Selection

- Initial data set is same as from our previous paper/work (21k SO Posts)

- Further, selection is based on following characteristics:
  - Post should be an *accepted Answer* on SO
  - Post should have been edited *at least 5 times*
  - Post should have *at least 10 comments*

- Resultant Data Set:
  - 684 SO Posts, referred in GH Projects
  - 14283 number of comments, ~21 average

Research Questions

RQ1: For all edited posts on SO, **what is the overall sentiment of comments w.r.t to edit timeline**? Does the overall sentiment in discussions improve with edits?

RQ2: If **quality of SO answers improves** with edits, **does the sentiment of comments indicate this trend?**
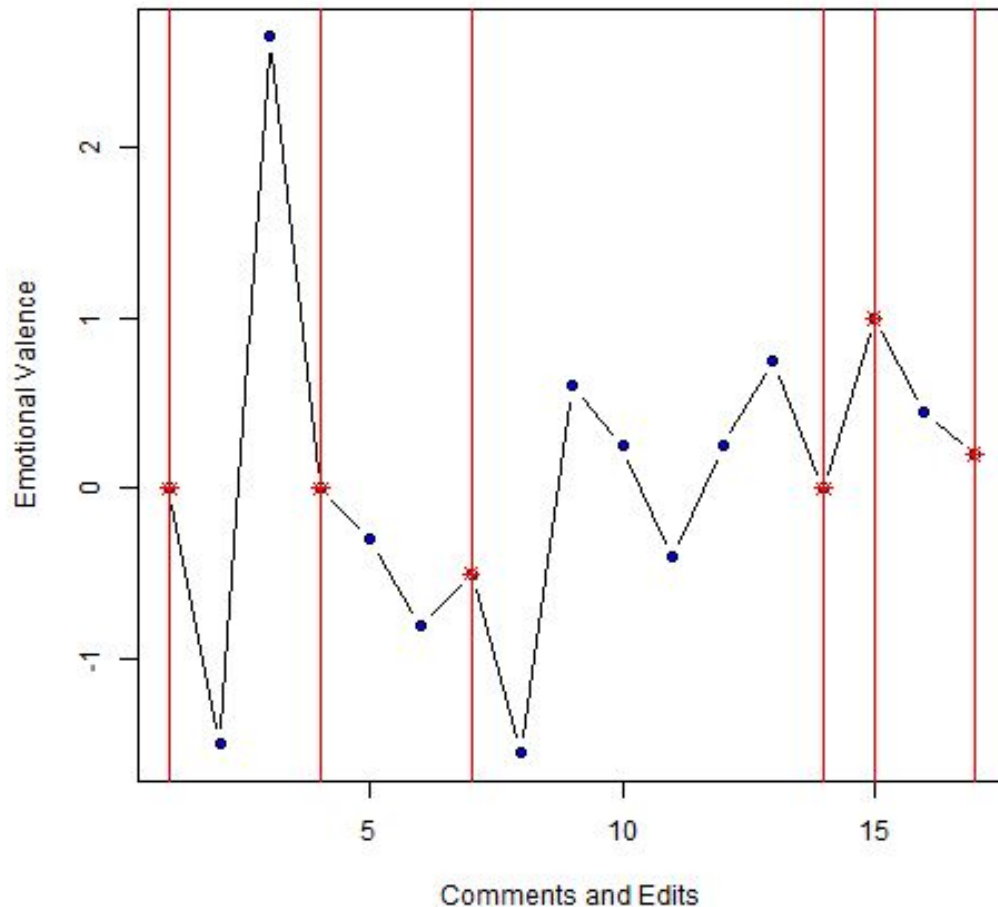
RQ3: How does **sentiment of SO discussions vary with topics?**

(RQ4): Knowing if a post is reliable/stable (unlikely to be edited in future) is important for software engineers. How **could it be detected if a post will likely be edited** in the future ?

Sentiment Valence Plots for Edits and Comments

Comments and Edits Sentiment Plot for PostId = 1152321

Corresponding post on SO

You can do this via the information_schema tables. For example:

```sql
SELECT
    tc.table_schema,
    tc.constraint_name,
    tc.table_name,
    kcu.column_name,
    ccu.table_schema AS foreign_table_schema,
    ccu.table_name AS foreign_table_name,
    ccu.column_name AS foreign_column_name
FROM
    information_schema.table_constraints AS tc
    JOIN information_schema.key_column_usage AS kcu
      ON tc.constraint_name = kcu.constraint_name
      AND tc.table_schema = kcu.table_schema
    JOIN information_schema.constraint_column_usage AS ccu
      ON ccu.constraint_name = tc.constraint_name
      AND ccu.table_schema = tc.table_schema
WHERE tc.constraint_type = 'FOREIGN KEY' AND tc.table_name='mytable';
```

share improve this answer

edited Dec 26 '18 at 9:00
Bear Brown
12.7k ● 8 ● 24 ● 46

answered Jul 20 '09 at 8:28
ollyc
3,187 ● 1 ● 12 ● 8

301

and discussions on answer..

8   table_name='mytable' should be tc.table_name='mytable' or else it throws an ambiguous error — intrepion Jul 15 '11 at 23:50

13   +1, very helpful. To make the query more robust it should probably join on constraint_schema as well, since it's possible for two schemas to have constraints with the same name. Something like: `FROM information_schema.table_constraints AS tc JOIN information_schema.key_column_usage AS kcu USING (constraint_schema, constraint_name) JOIN information_schema.constraint_column_usage AS ccu USING (constraint_schema, constraint_name)` — EMP Aug 26 '11 at 6:41

6   This breaks when there are several columns in a constraint, doesn't it? There seems to be no proper way to associate pk columns with fk columns using information_schema BTW. — fionbio Jun 1 '12 at 18:54

3   It indeed breaks with more than one column in constraint. For Postgres, there is a way of getting this information from the pg_catalog schema. See my answer below. — martin Jun 8 '12 at 14:07

6   The query is wrong. It assumes that constraint names cannot repeat, which is false. Constraints with the same name can exist in different namespaces. You are using constraint_name to make the join. Also joinning on both constraint_name and schema name won't work since you are not sure the two constraints are the same. The only option is going for pg_constraints, pg_class etc. using oids to join. Postgres' ANSI catalog is only there for compliance but it's flawed. pg_catalog is the way to go. The correct answer is here dba.stackexchange.com/questions/36979/retrieving-all-pk-and-fk — Tulains Córdova Nov 24 '15 at 18:18 ✎

  For some reason this works but is very slow for me, it takes 5-6 seconds to run on a schema with 1 table and 2 constraints — Alexandru Severin Jul 1 '16 at 12:19

  Doesnt work if your user is not superuser. — Jan 'splite' K. Mar 26 '17 at 0:33

  w/ schemas: `SELECT tc.constraint_name, tc.table_schema||'.'||tc.table_name table_name , kcu.column_name, ccu.table_schema||'.'||ccu.table_name AS foreign_table_name, ccu.column_name AS foreign_column_name FROM information_schema.table_constraints AS tc JOIN information_schema.key_column_usage AS kcu ON tc.constraint_name = kcu.constraint_name JOIN information_schema.constraint_column_usage AS ccu ON ccu.constraint_name = tc.constraint_name WHERE constraint_type = 'FOREIGN KEY'` ~ — gerardw Jun 1 '17 at 11:40 ✎

Sentiment Valence Plots for Edits and Comments

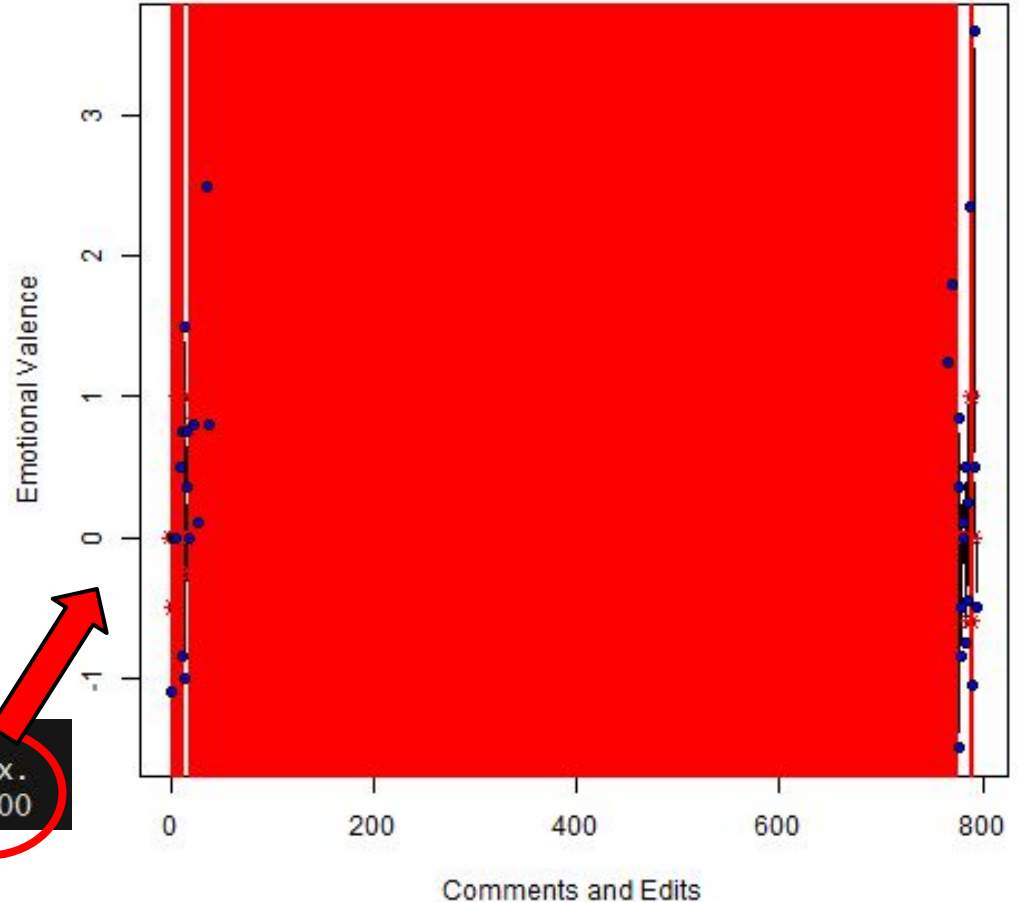Comments and Edits Sentiment Plot for PostId = 14220323

Sentiment Valence Plot
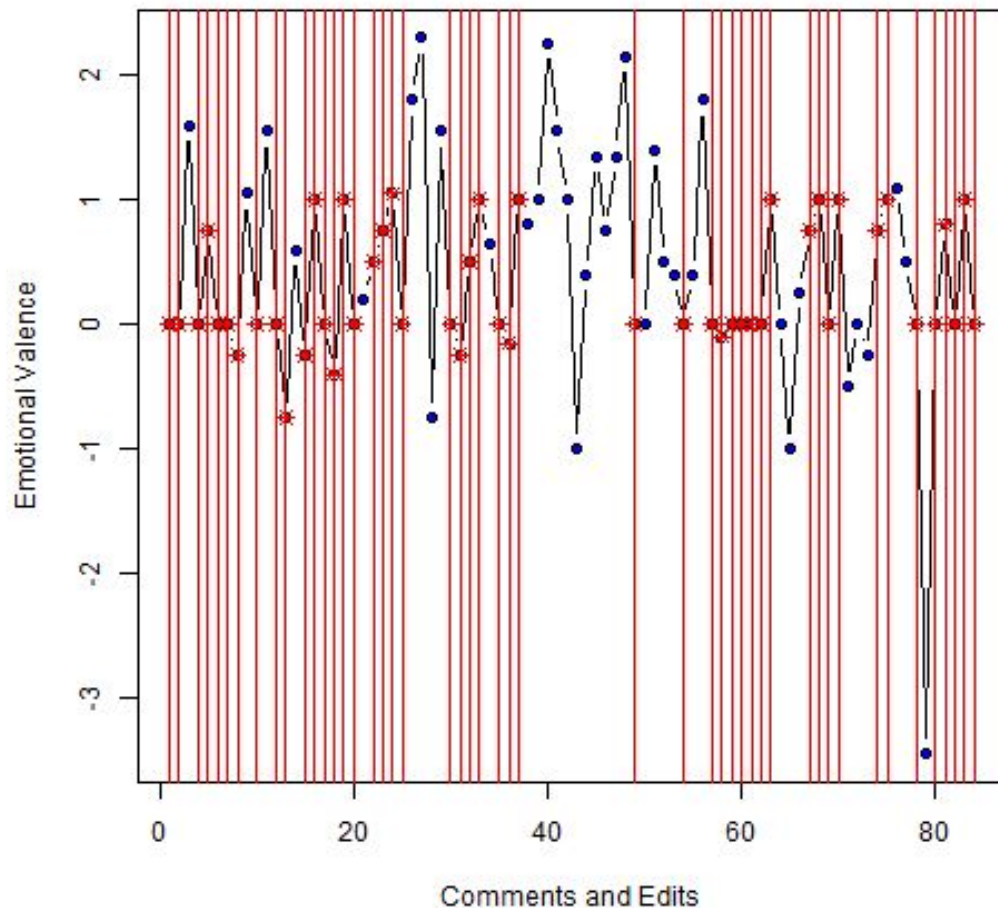
Max Number of Edits Outlier Example

Comments and Edits Sentiment Plot for PostId = 4673436

```
summary(numberOfEdits)
 Min. 1st Qu. Median    Mean 3rd Qu.    Max.
 1.00    6.00    8.00   11.87   13.00  761.00
```

Sentiment Valence Plot

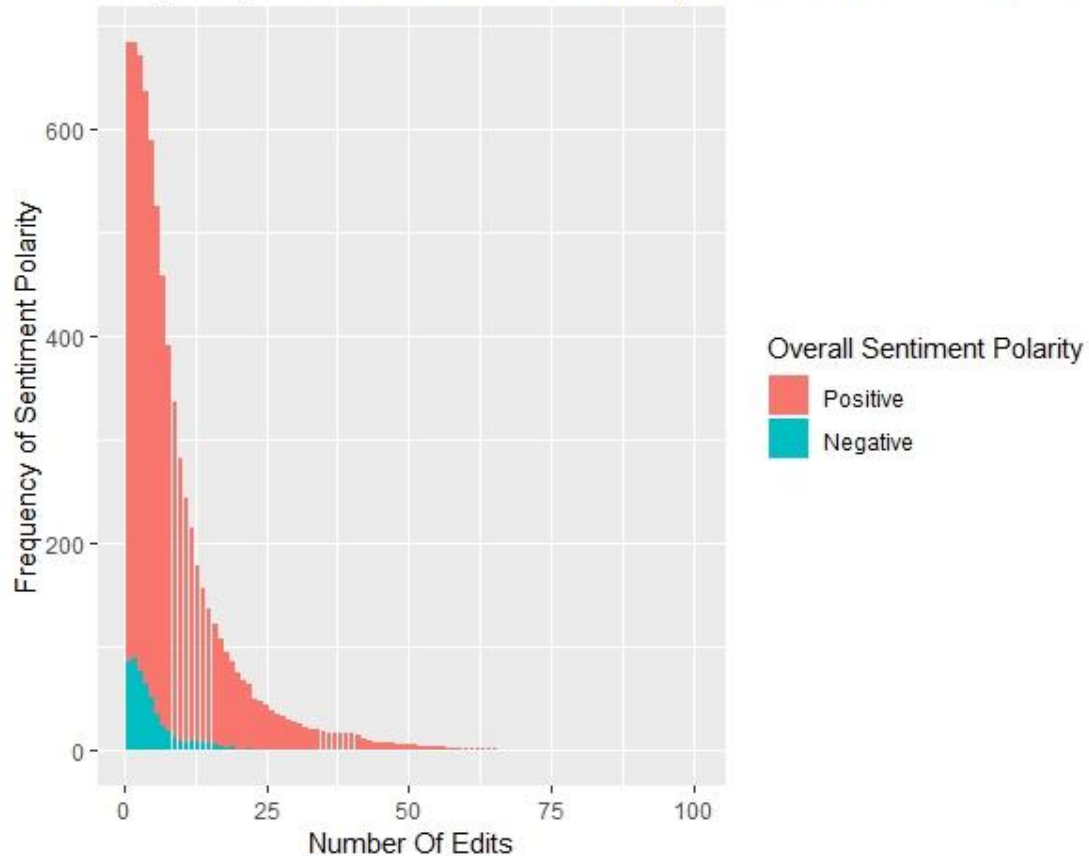# Max Number of Comments Outlier Example

Comments and Edits Sentiment Plot for PostId = 18746930
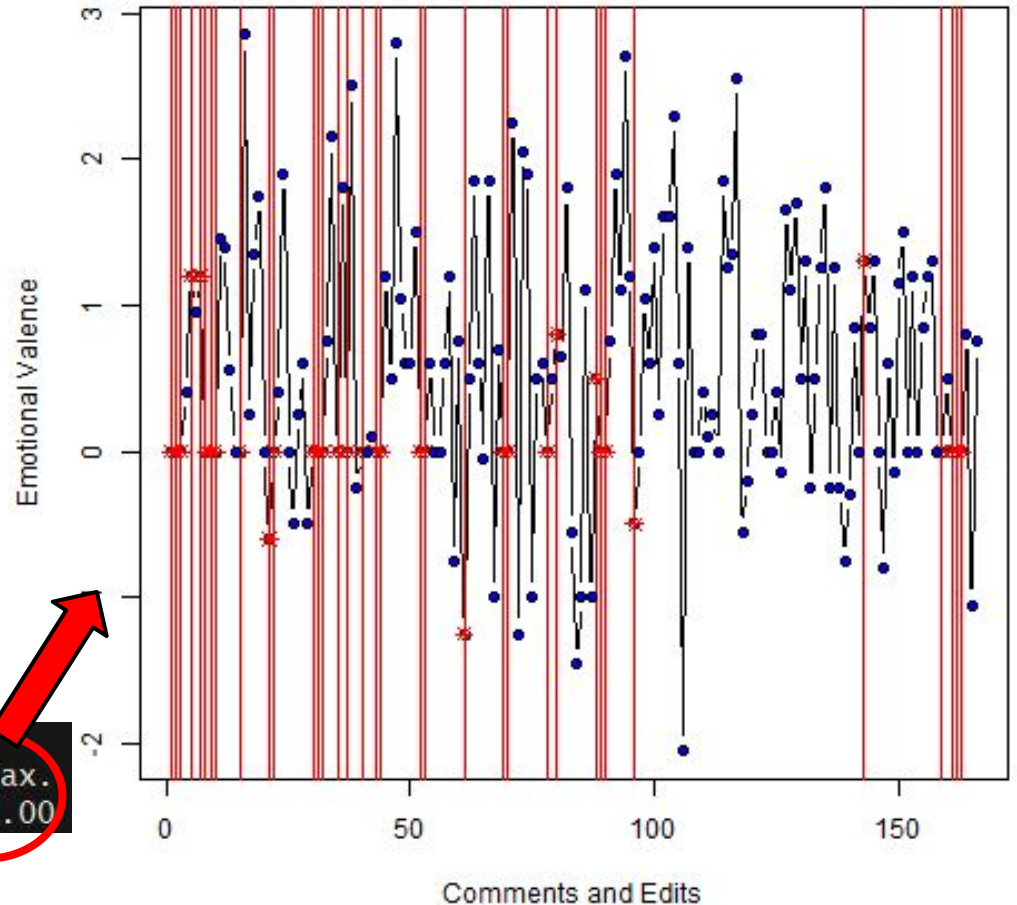
```
summary(numberOfComm)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  11.00   13.00   17.00   20.88   25.00  131.00
```

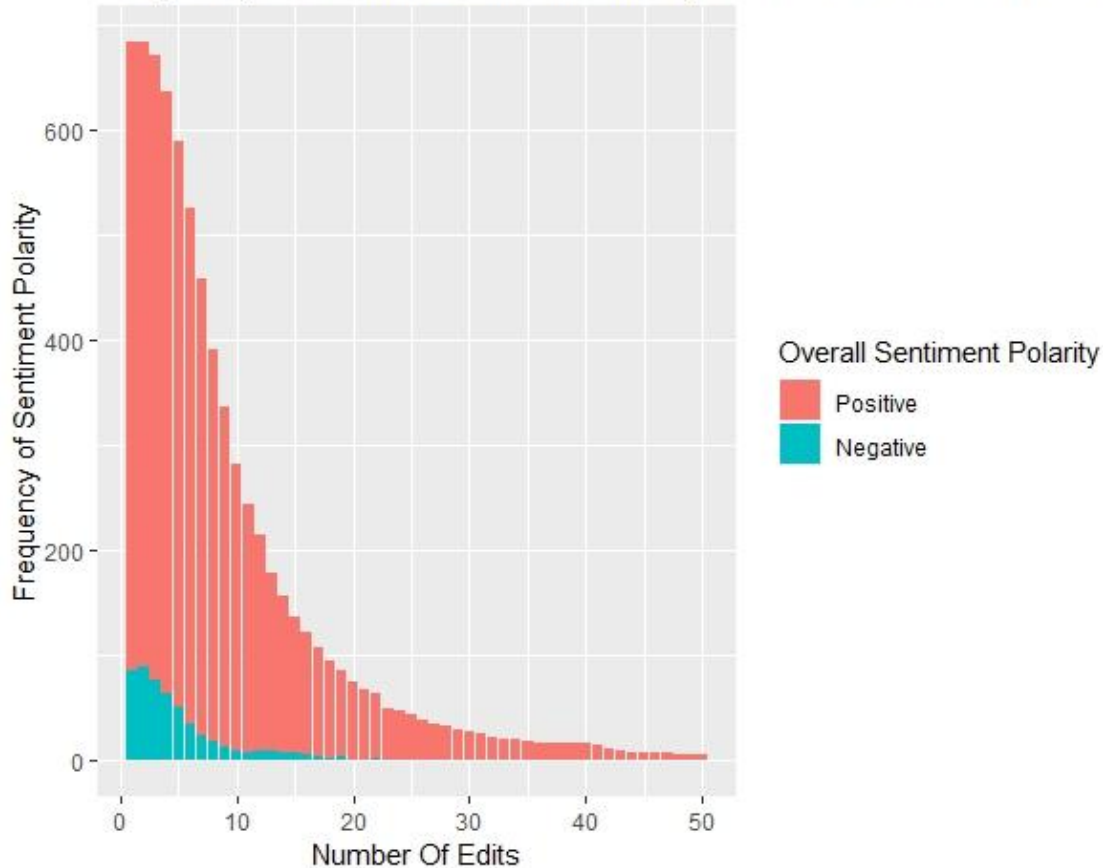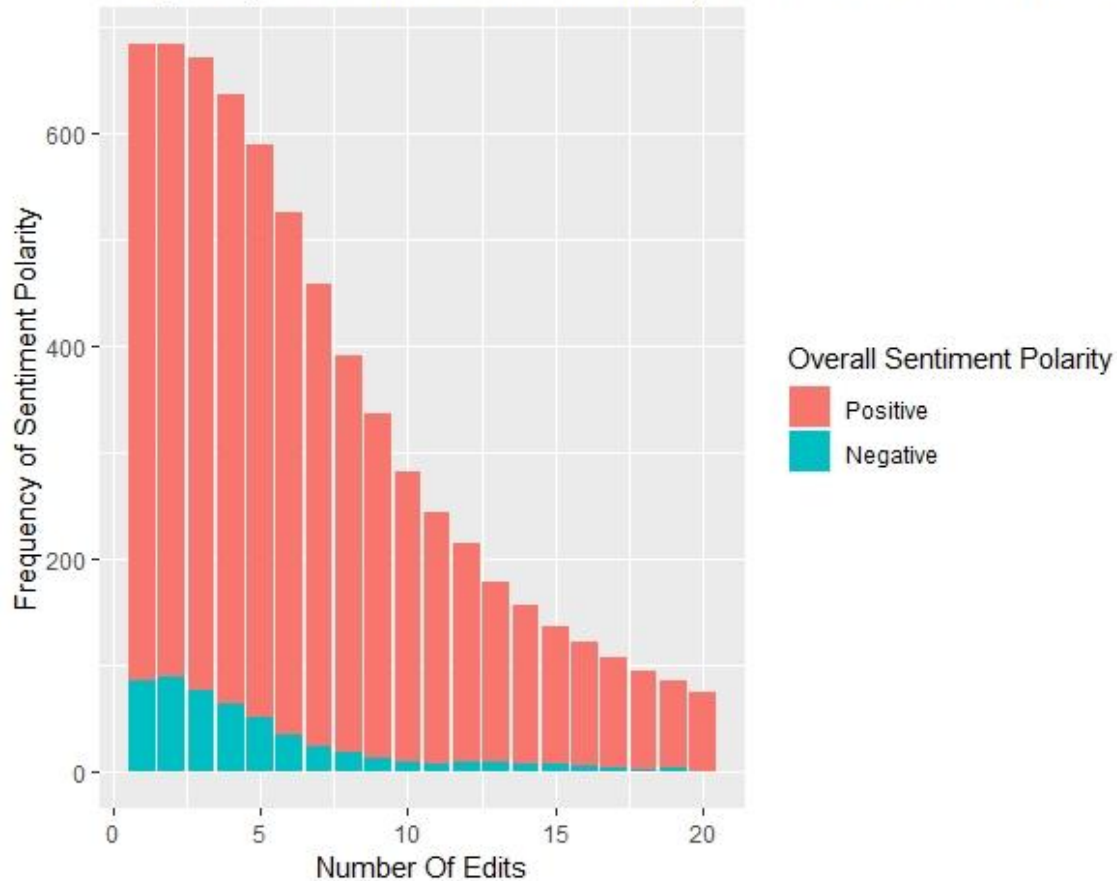# Results - Sentiment Polarity Trends After Edits

Results - Sentiment Polarity Trends After Edits

Frequency of Overall Sentiment Polarity On First 50 edits on an Ans

Results - Sentiment Polarity Trends After Edits

Frequency of Overall Sentiment Polarity On First 20 edits on an Ans

Thank you!

Questions ?