Sentimine: Towards a cregit Plugin to Analyze the Sentiment Behind the Linux Kernel Code

Isabella Ferreira
Polytechnique Montréal

Kate Stewart
Linux Foundation

Daniel German
University of Victoria

Bram Adams
Polytechnique Montréal

Linus Torvalds' temporary leave in September 2018 ...

From: Linus Torvalds <torvalds@linux-foundation.org>
To: Linux Kernel Mailing List <linux-kernel@vger.kernel.org>
Subject: Linux 4.19-rc4 released, an apology, and a maintainership note
Date: Sun, 16 Sep 2018 12:22:43 -0700
Message-ID: <CA+55aFy+Hv9O5citAawS+mVZO+ywCKd9NQ2wxUmGsz9ZJzqgJQ@mail.gmail.com>

This week people in our community confronted me about my lifetime of not understanding emotions. My flippant attacks in emails have been both unprofessional and uncalled for. Especially at times when I made it personal. In my quest for a better patch, this made sense to me. I know now this was not OK and I am truly sorry.

The above is basically a long-winded way to get to the somewhat painful personal admission that hey, I need to change some of my behavior, and I want to apologize to the people that my personal behavior hurt and possibly drove away from kernel development entirely.

I am going to take time off and get some assistance on how to understand people's emotions and respond appropriately.

… sparked a Rich and Diverse **Discussion about Change** in the Community ...

… sparked a Rich and Diverse Discussion about Change in the Community …

THE VERGE TECH REVIEWS SCIENCE ENTERTAINMENT VIDEO FEATURES MORE

REPORT / TECH / LINUX

With Linux's founder stepping back, will the community change its culture?

By Adi Robertson | @thedextriarchy | Sep 21, 2018, 3:00pm EDT

SHARE

Alex Dawson / Flickr

Tesla Model Y S
14th

EXTREMETECH Search Extremetech

Computing Phones Cars Gaming Science Extreme

HOME 〉 COMPUTING 〉 LINUX TORVALDS STEPS BACK FROM LINUX IN FAVOR OF SELF-IMPROVEMENT

Linux Torvalds Steps Back From Linux in Favor of Self-Improvement

By Joel Hruska on September 18, 2018 at 8:07 am | 17 Comments

NETWORKWORLD FROM IDG

Home > Open Source (dupe)

Linux community acts after years of complaints like Sarah Sharp's

Years after Sarah Sharp stepped down from her direct work in the Linux kernel community because of abusive behavior, Linus Torvalds has stepped away and the community adopted a Code of Conduct.

By Jon Gold
Senior Writer, Network World | SEPTEMBER 21, 2018 11:37 AM PT

I'm very disappointed in Linus Torvalds

What happened Linus?

Karel Donk Follow
Sep 22, 2018 · 6 min read

NEW YORKER

News Culture Books Business & Tech Humor Cartoons Magazine Video Podcasts Archive Goings On Subscribe

ELEMENTS

AFTER YEARS OF ABUSIVE E-MAILS, THE CREATOR OF LINUX STEPS ASIDE

By Noam Cohen September 19, 2018

LWN.net
News from the source

Content
Weekly Edition
Archives
Search
Kernel

User: Password: Log in | Subscribe | Register

After Years of Abusive E-mails, the Creator of Linux Steps Aside (The New Yorker)

[Posted September 20, 2018 by jake]

… even after Linus’ comeback, discussions continued

… even after Linus' comeback, discussions continued

THE VERGE TECH REVIEWS SCIENCE ENTERTAINMENT VIDEO FEATURES MORE

WEB TECH LINUX

Linus Torvalds returns to Linux development with new code of conduct in place

Torvalds took a self-imposed break to rethink his controversial treatment of others

By Nick Statt | @nickstatt | Oct 22, 2018, 8:16pm EDT

BBC Sign in News Sport Weather Shop Reel Travel

NEWS

Home Video World US & Canada UK Business Tech Science Stories

Technology

Linus Torvalds returns to head Linux coding community

23 October 2018 Share

FOSSBYTES NEWS GEEK SECURITY HOW TO TOP X REVIEWS VIDEOS

Linus Torvalds Discusses His Return To Linux Development

October 24, 2018

SHARE Facebook Twitter

ZDNet VIDEOS

MUST READ: The Industrial Internet of Things: A guide to deployments

Linus Torvalds is back in charge of Linux

After a few weeks off to reconsider his role in the Linux community, Linus Torvalds is back in the saddle.

By Steven J. Vaughan-Nichols for Linux and Open Source | October 22, 2018 -- 10:45 GMT (03:45 PDT) | Topic: Enterprise Software

code
culture
reconsider
improvement
act
community
discusses
change

Almost half a year later, what did the community learn or change?

Code of Conflict became Code of Conduct :-)

Code of Conflict became Code of Conduct :-)

```
-Code of Conflict
-----------------
-
-The Linux kernel development effort is a very personal process compared
-to "traditional" ways of developing software.  Your code and ideas
-behind it will be carefully reviewed, often resulting in critique and
-criticism.  The review will almost always require improvements to the
-code before it can be included in the kernel.  Know that this happens
-because everyone involved wants to see the best possible solution for
-the overall success of Linux.  This development process has been proven
-to create the most robust operating system kernel ever, and we do not
-want to do anything to cause the quality of submission and eventual
-result to ever decrease.
-
-If however, anyone feels personally abused, threatened, or otherwise
-uncomfortable due to this process, that is not acceptable.  If so,
-please contact the Linux Foundation's Technical Advisory Board at
-<tab@lists.linux-foundation.org>, or the individual members, and they
-will work to resolve the issue to the best of their ability.  For more
-information on who is on the Technical Advisory Board and what their
-role is, please see:
-
-       - http://www.linuxfoundation.org/projects/linux/tab
-
-As a reviewer of code, please strive to keep things civil and focused on
-the technical issues involved.  We are all humans, and frustrations can
-be high on both sides of the process.  Try to keep in mind the immortal
-words of Bill and Ted, "Be excellent to each other."
```

```
+Contributor Covenant Code of Conduct
++++++++++++++++++++++++++++++++++++++
+
+Our Pledge
+==========
+
+In the interest of fostering an open and welcoming environment, we as
+contributors and maintainers pledge to making participation in our project and
+our community a harassment-free experience for everyone, regardless of age, body
+size, disability, ethnicity, sex characteristics, gender identity and
+expression, level of experience, education, socio-economic status, nationality,
+personal appearance, race, religion, or sexual identity and orientation.
+
+Our Standards
+=============
+
+Examples of behavior that contributes to creating a positive environment
+include:
+
+* Using welcoming and inclusive language
+* Being respectful of differing viewpoints and experiences
+* Gracefully accepting constructive criticism
+* Focusing on what is best for the community
+* Showing empathy towards other community members
+
+
+Examples of unacceptable behavior by participants include:
+
+* The use of sexualized language or imagery and unwelcome sexual attention or
+  advances
+* Trolling, insulting/derogatory comments, and personal or political attacks
+* Public or private harassment
```

…, but a Code of Conduct just treats the symptoms, what is the root of the problem?

…, but a Code of Conduct just treats the symptoms, what is the root of the problem?

ci·vil·i·ty

/səˈviladē/

noun

formal politeness and courtesy in behavior or speech.
"I hope we can treat each other with civility and respect"
synonyms: courtesy, courteousness, politeness, good manners, mannerliness, gentlemanliness, chivalry, gallantry, graciousness, consideration, respect, gentility; More

- polite remarks used in formal conversation.
  plural noun: **civilities**
  "she was exchanging civilities with his mother"
  synonyms: polite remark, politeness, courtesy; formality
  "she didn't waste time on civilities"

How civil is the Linux community?

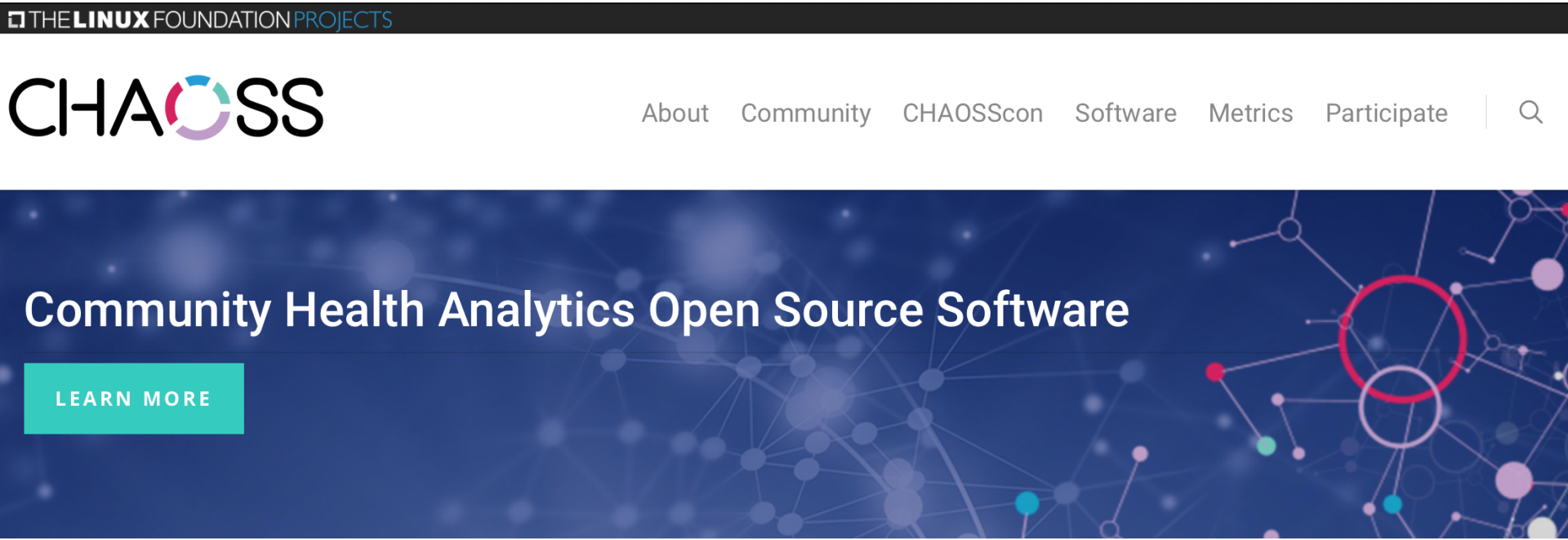
Is it a recent problem?

Is it a problem that is increasing over time?

How does civility change across release cycles?

How is the civility in the Linux community compared to other open source projects?

Civility is not (yet) part of CHAOSS



Initial idea: Measuring civility through sentiment

Initial idea: Measuring civility through sentiment

Politeness

Initial idea: Measuring civility through sentiment

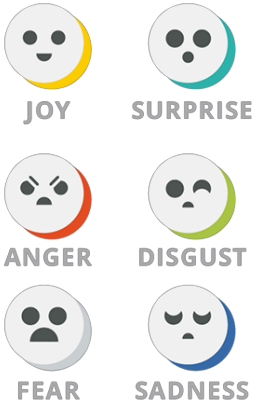
Politeness

Sentiment

Initial idea: Measuring civility through sentiment

Politeness

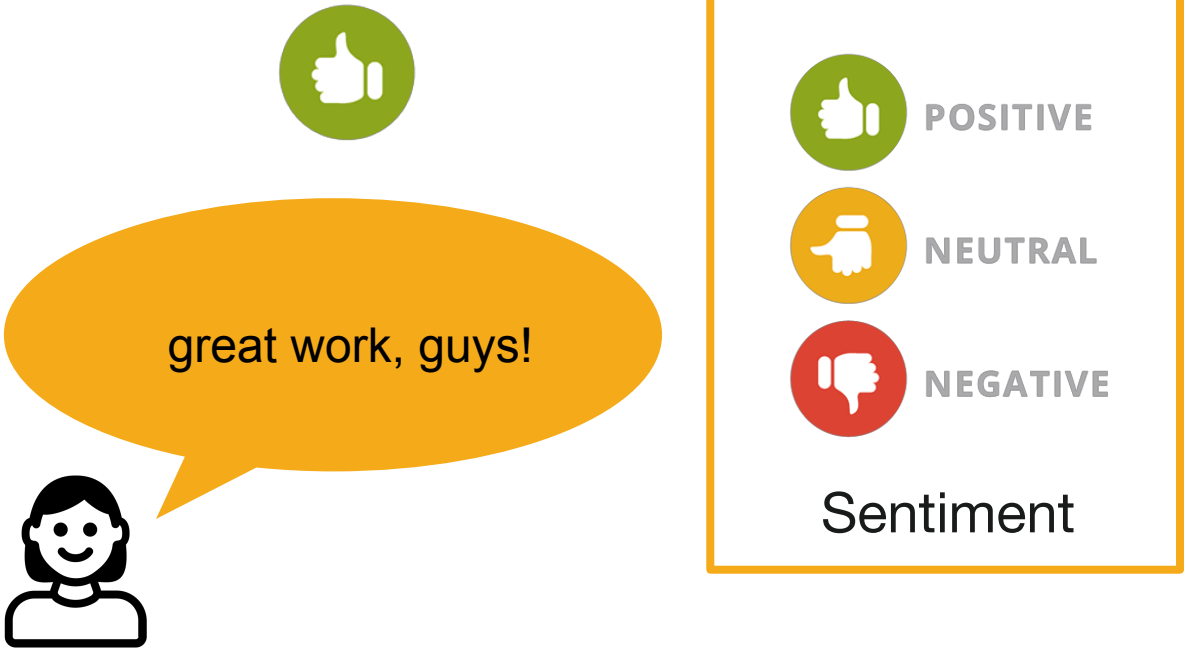
Sentiment



Emotion

Initial idea: Measuring civility through sentiment

Initial idea: Measuring civility through sentiment



Initial idea: Measuring civility through sentiment

great work, guys!

WTF, a package refactoring and class renaming in one patch?

Initial idea: Measuring civility through sentiment

great work, guys!

WTF, a package refactoring and class renaming in one patch?

Good civility

Bad civility

Challenges

Challenges

Identify the **target of the sentiment**

- Is the sentiment about the **person** (civility) or about the **code** (review)?

Challenges

Identify the **target of the sentiment**

- Is the sentiment about the **person** (civility) or about the **code** (review)?

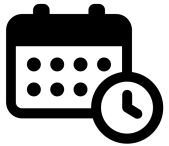
Identify the right **granularity** to analyze sentiments

- Individual e-mails vs. e-mail threads?
- Subsystem vs. the whole kernel?
- Week vs. month vs. release cycle?

Can we measure differences in sentiment **between releases**?

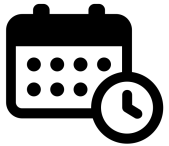
Can we measure differences in sentiment **between releases**?

Can we measure differences in sentiment **across time**?



Can we measure differences in sentiment **between releases**?

Can we measure differences in sentiment **across time**?



Can we measure differences in sentiment **between maintainers and non-maintainers**?

Approach

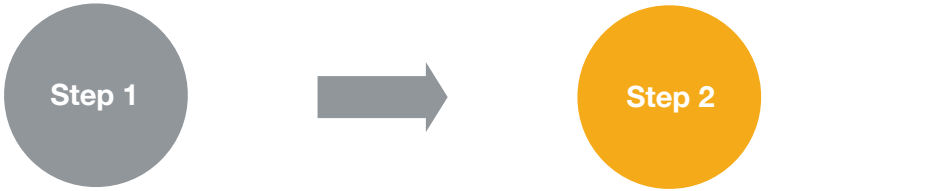
Approach

Step 1

Extract LKML data

Public-inbox e-mails from Jan 2017 to Jan 2019

Approach



Extract LKML data
Public-inbox e-mails from Jan 2017 to Jan 2019

Group e-mails by threads
Mailboxminer tool

Approach

Extract LKML data
Public-inbox e-mails from Jan 2017 to Jan 2019

Group e-mails by threads
Mailboxminer tool

Filter threads
>= 2 e-mails in a thread
ignore e-mails that have patch

Approach

Approach

Calculate metrics

Release cycle

Month, week, weekday

Maintainer vs. non-maintainers

Approach

Calculate metrics

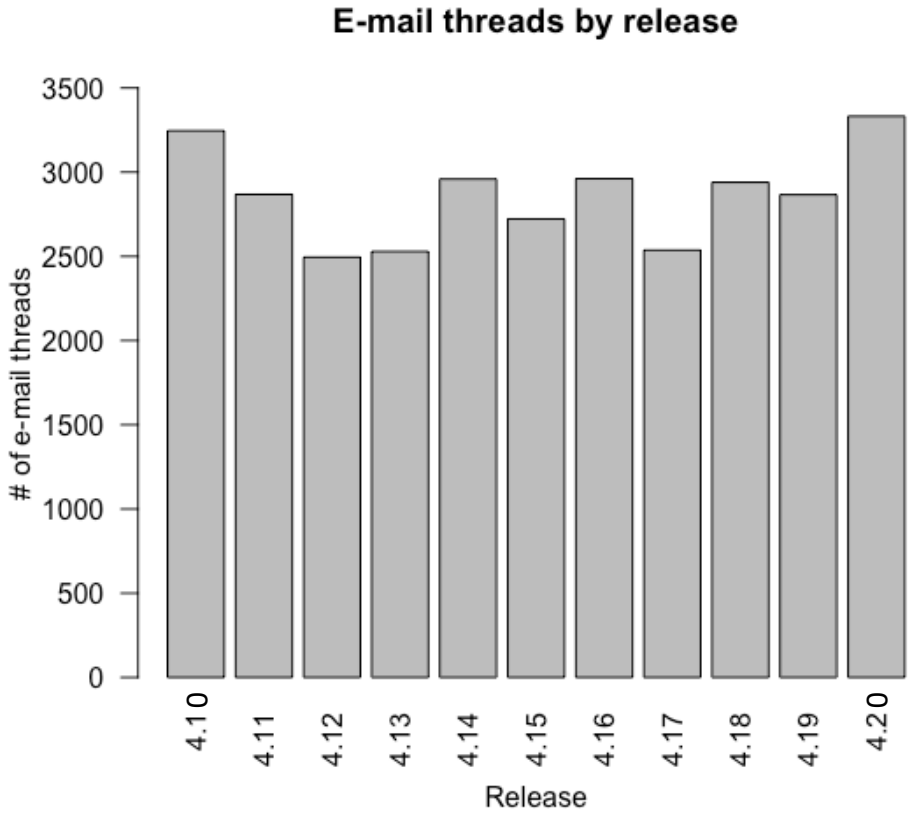
Release cycle

Month, week, weekday

Maintainer vs. non-maintainers

Results

A lot of communication across releases...



Phase 2 in the release cycle has more discussion

Phase 2 in the release cycle has more discussion

January 28	4.15 stable release	Phase 1
February 11	4.16-rc1, **merge** **window** closes	
February 18	4.16-rc2	Phase 2
February 25	4.16-rc3	
March 4	4.16-rc4	
March 11	4.16-rc5	
March 18	4.16-rc6	
March 25	4.16-rc7	Phase 3
April 1	4.16 stable release	

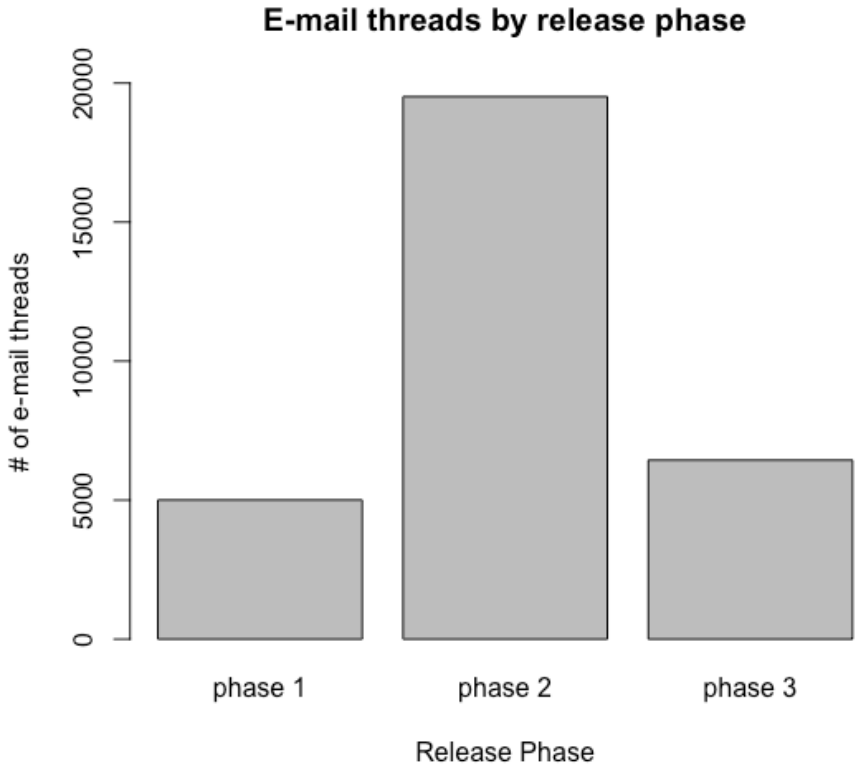
Phase 2 in the release cycle has more discussion

January 28	4.15 stable release
February 11	4.16-rc1, **merge** **window** closes
February 18	4.16-rc2
February 25	4.16-rc3
March 4	4.16-rc4
March 11	4.16-rc5
March 18	4.16-rc6
March 25	4.16-rc7
April 1	4.16 stable release

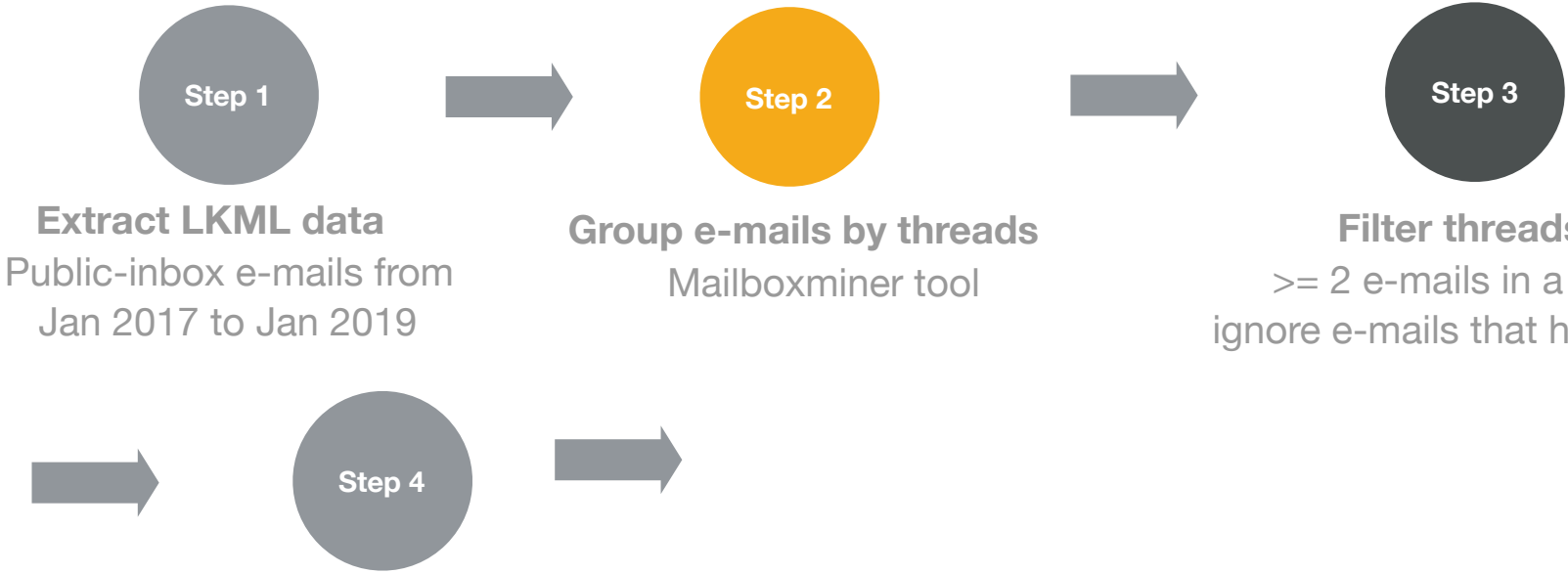
Phase 1

Phase 2

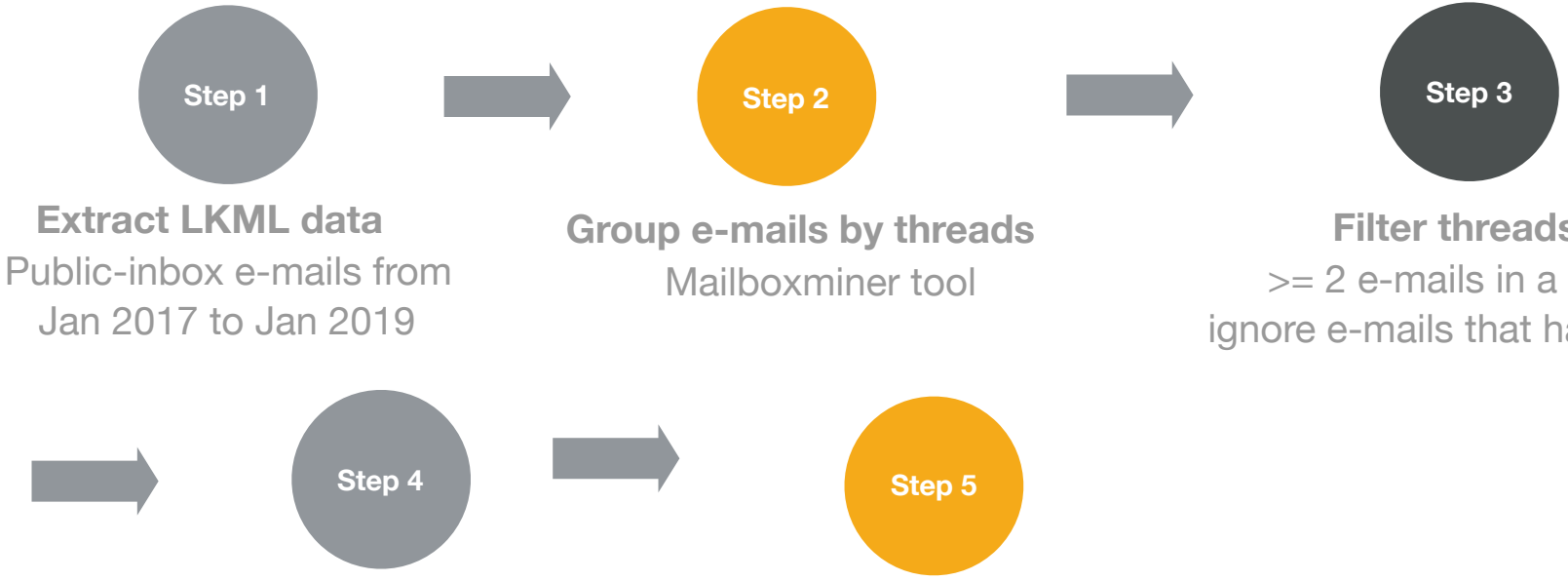
Phase 3



Approach


Step 1
Extract LKML data
Public-inbox e-mails from
Jan 2017 to Jan 2019
Step 2
Group e-mails by threads
Mailboxminer tool
Step 3
Filter threads
>= 2 e-mails in a thread
ignore e-mails that have patch
Step 4
Calculate metrics
Release cycle
Subsystem
Maintainer vs. non-maintainers

Approach


Step 1
Extract LKML data
Public-inbox e-mails from
Jan 2017 to Jan 2019
Step 2
Group e-mails by threads
Mailboxminer tool
Step 3
Filter threads
>= 2 e-mails in a thread
ignore e-mails that have patch
Step 4
Calculate metrics
Release cycle
Subsystem
Maintainer vs. non-maintainers
Step 5
Detect sentiment
Senti4SD

Approach

Step 1

Extract LKML data

Public-inbox e-mails from Jan 2017 to Jan 2019

Step 2

Group e-mails by threads

Mailboxminer tool

Step 3

Filter threads

>= 2 e-mails in a thread

ignore e-mails that have patch

Step 4

Calculate metrics

Release cycle

Subsystem

Maintainer vs. non-maintainers

Step 5

Detect sentiment

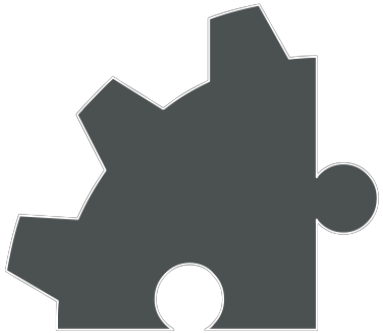
Senti4SD

Answer questions about civility

Senti4SD

Senti4SD

Lexicon-based features



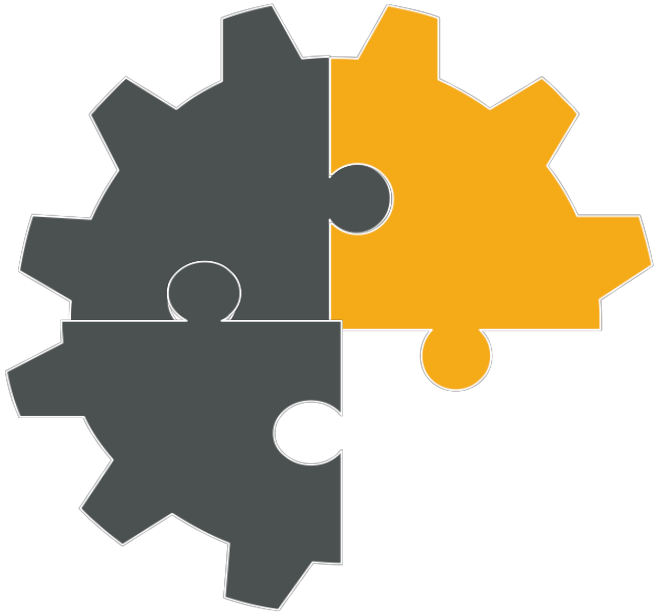
Senti4SD

Lexicon-based features

Keyword-based features

Senti4SD

Lexicon-based features



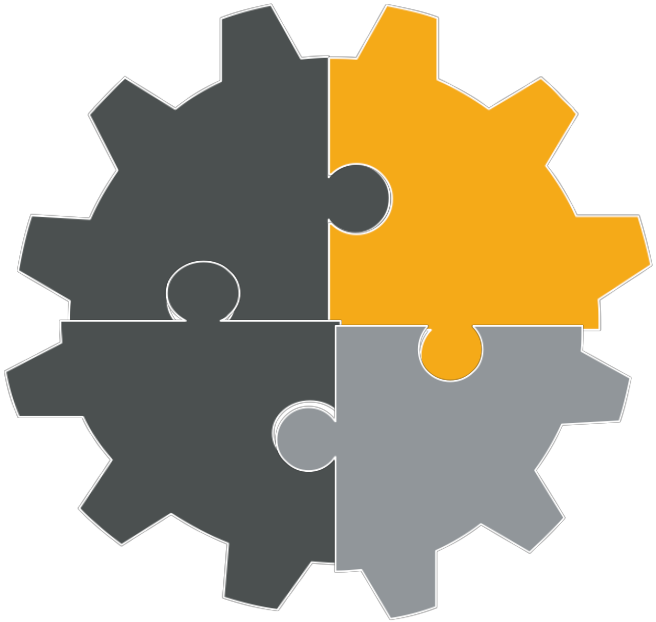
Keyword-based features

Semantic features

Senti4SD

Lexicon-based features

Keyword-based features



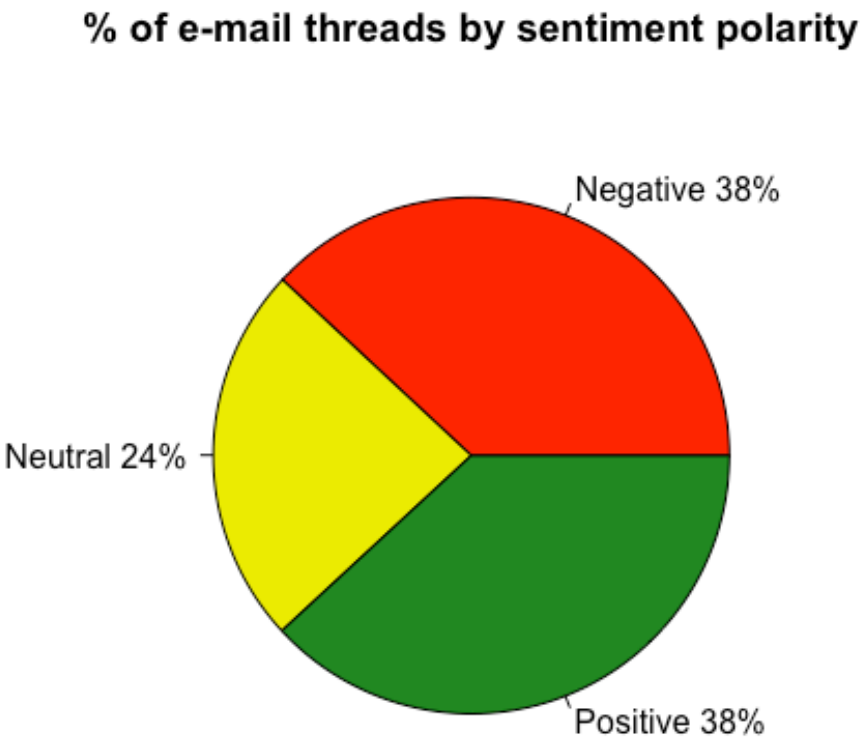
Semantic features

Training the model

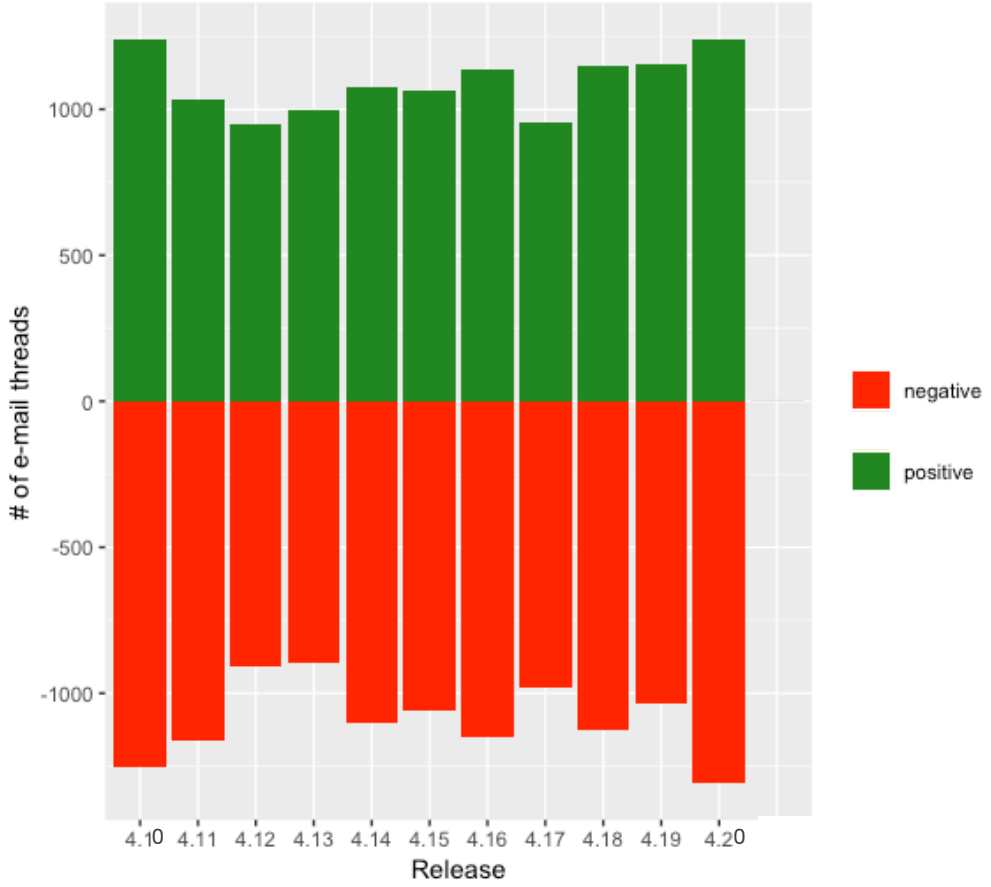
Senti4SD

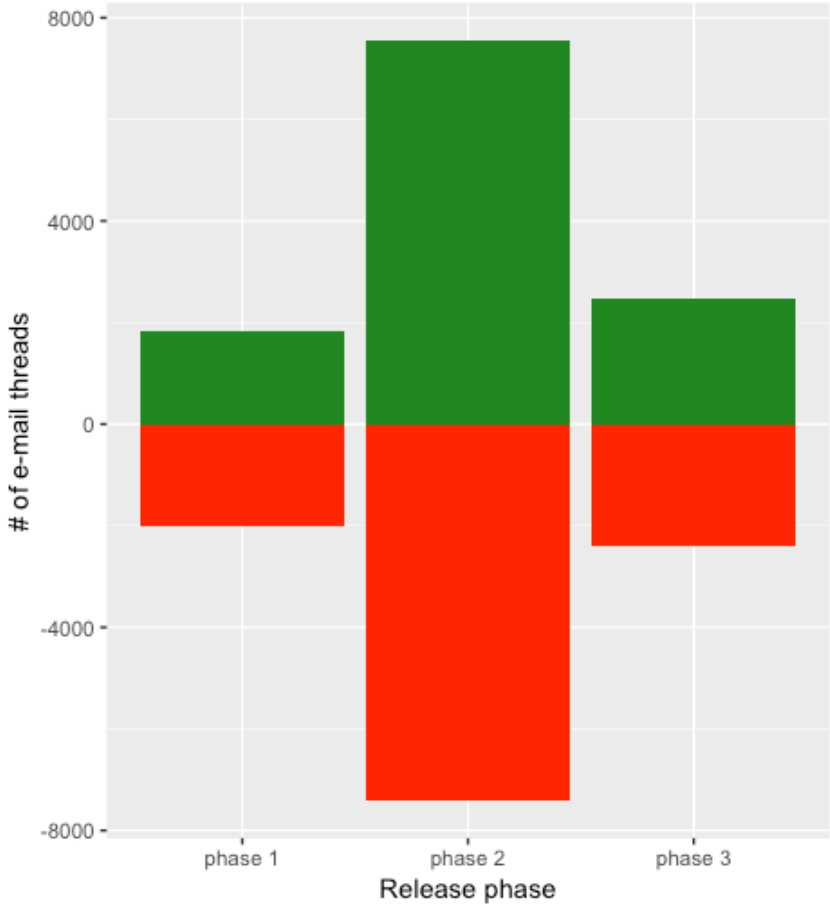
INPUT	OUTPUT
Sweet :) Happy hacking!	positive
I want them to resize based on the length of the data they're showing.	neutral
Do you have jQuery loaded correctly?	neutral
For Python 3 the following will work.	positive
If you're really worried about this, Java is not the language for you	negative
	neutral

Equal proportion of positive/negative e-mail threads

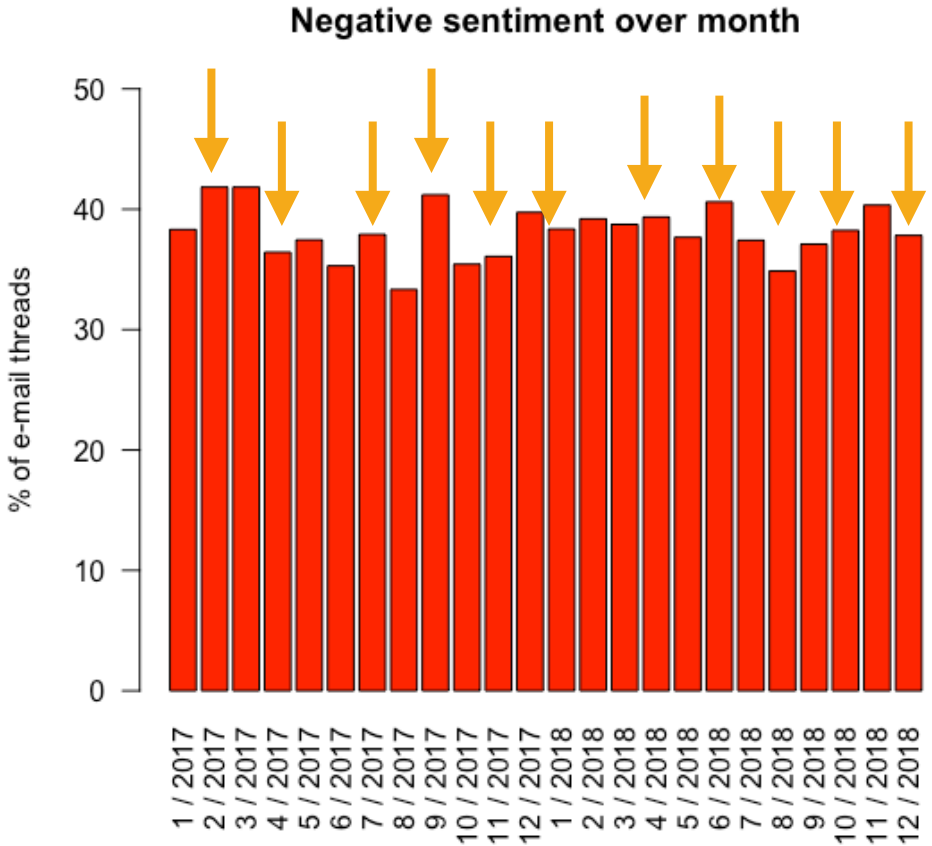


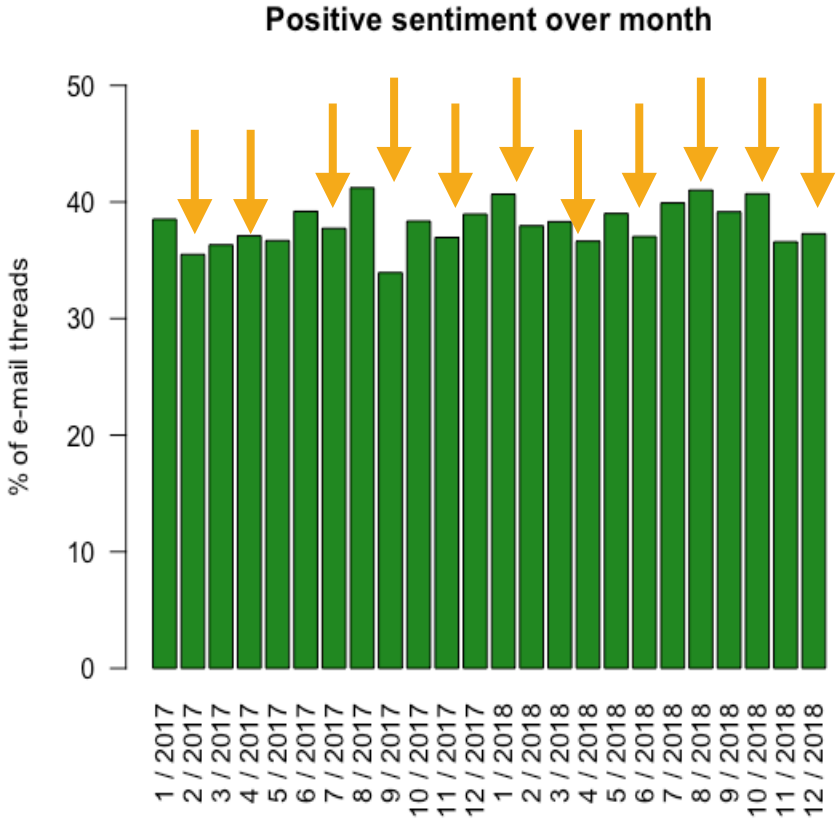
Equal number of positive/negative discussions across releases



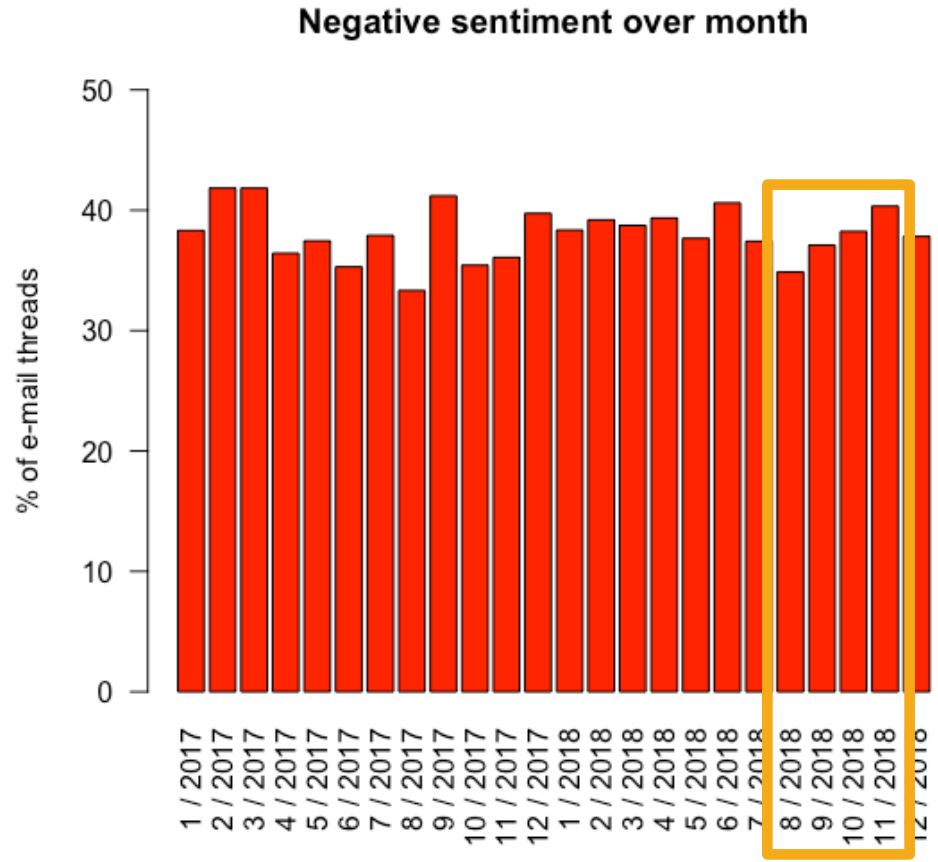


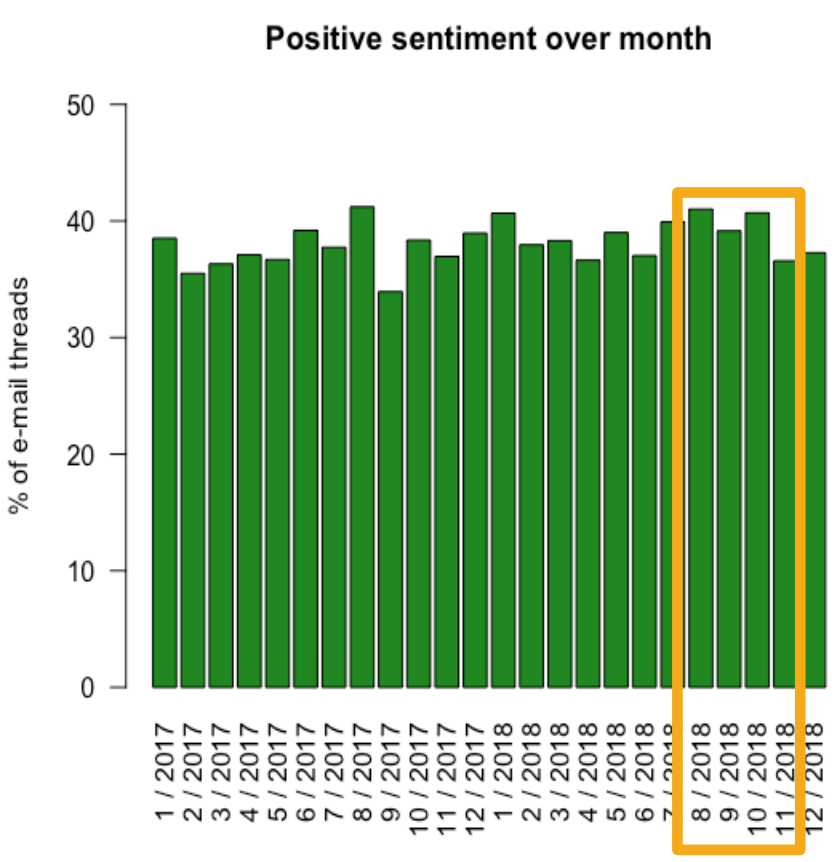
Specific events might change the sentiment…



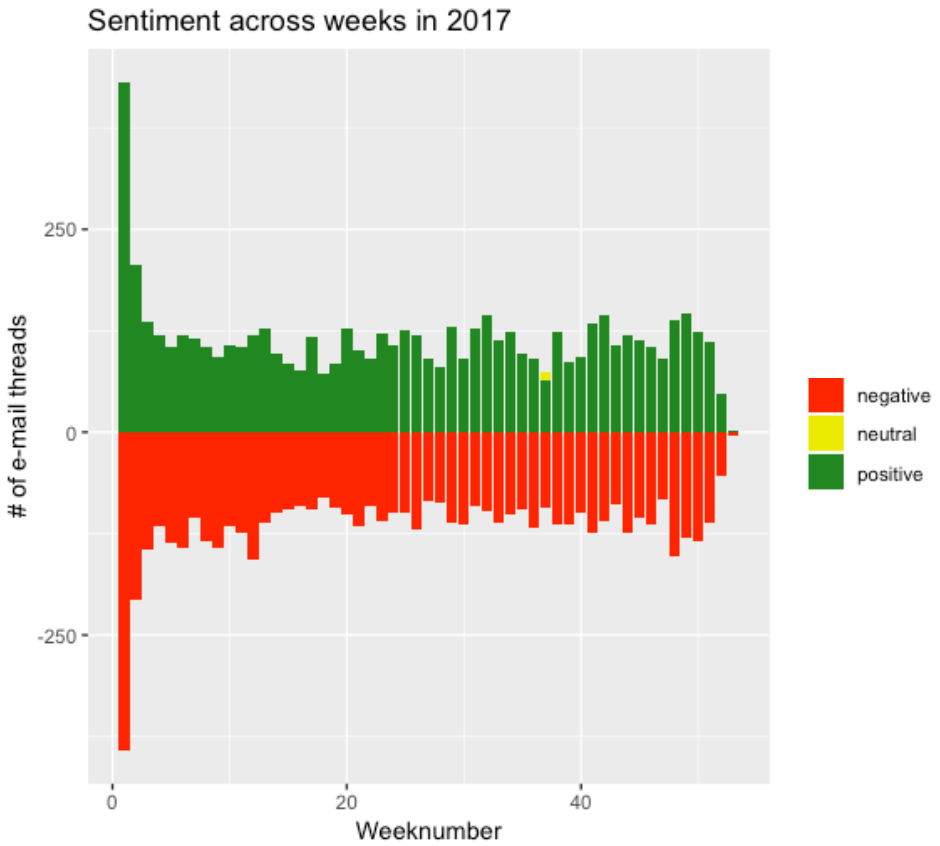


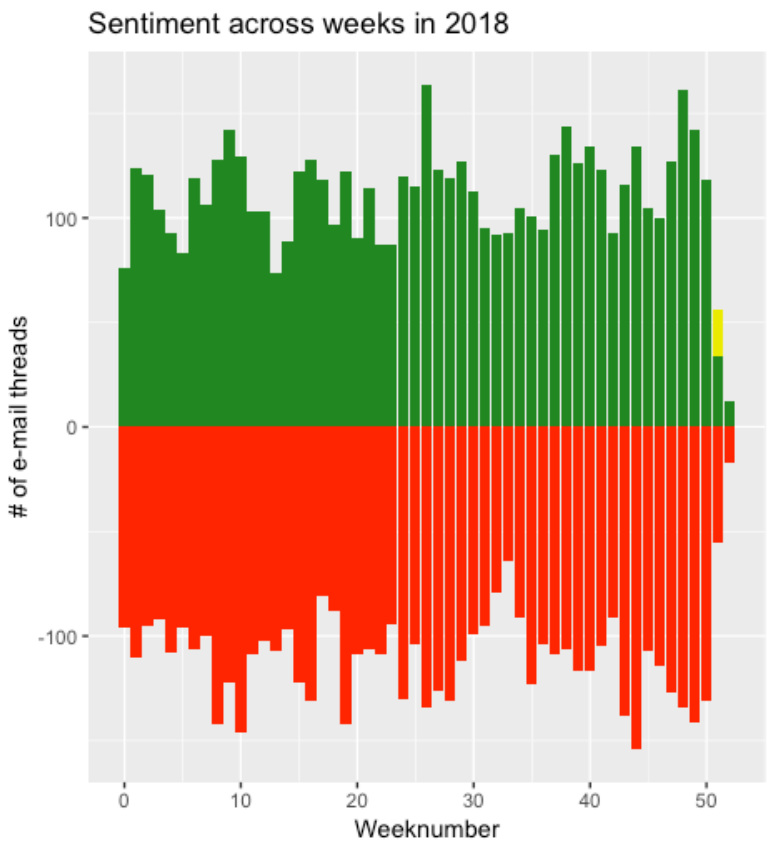
Specific events might change the sentiment…



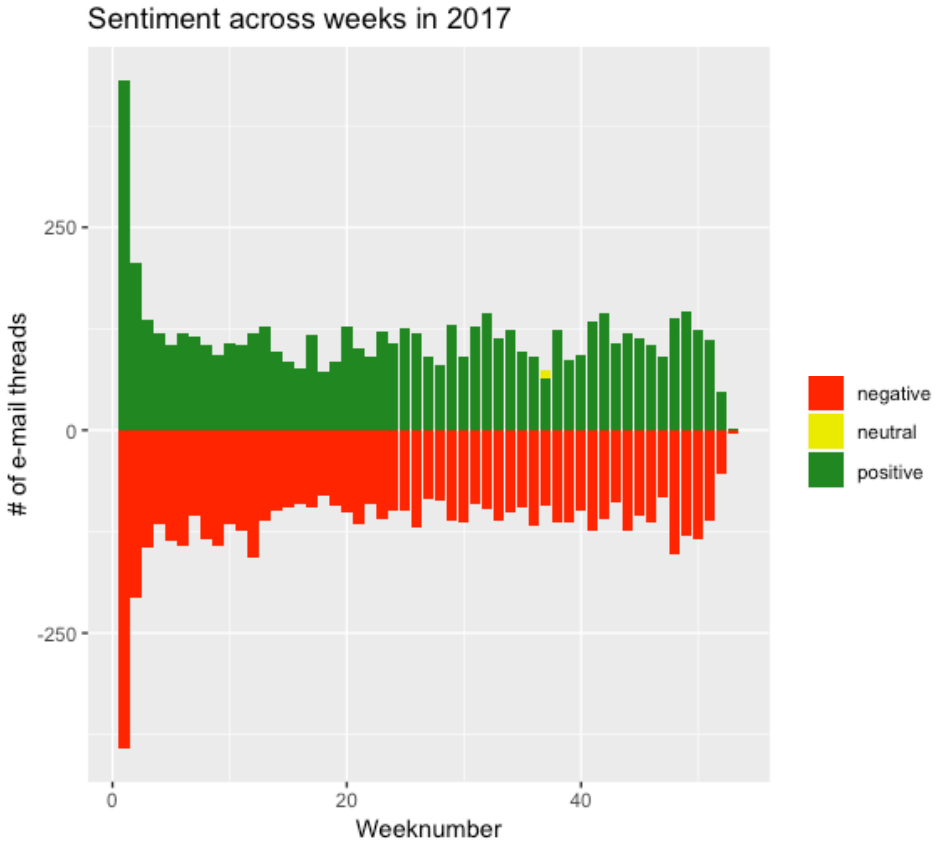


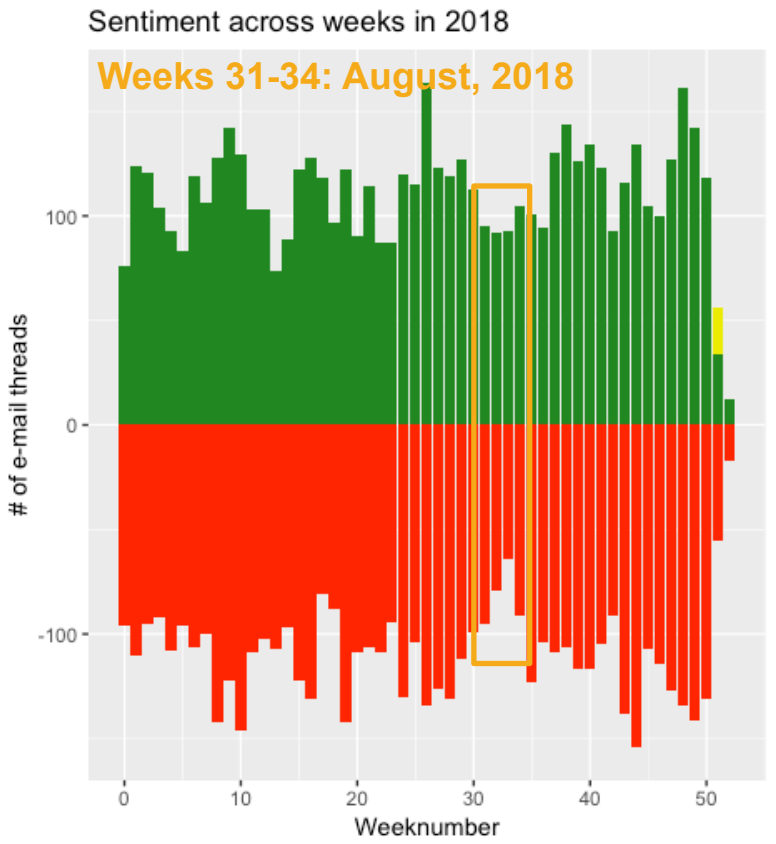
The sentiment fluctuates among weeks



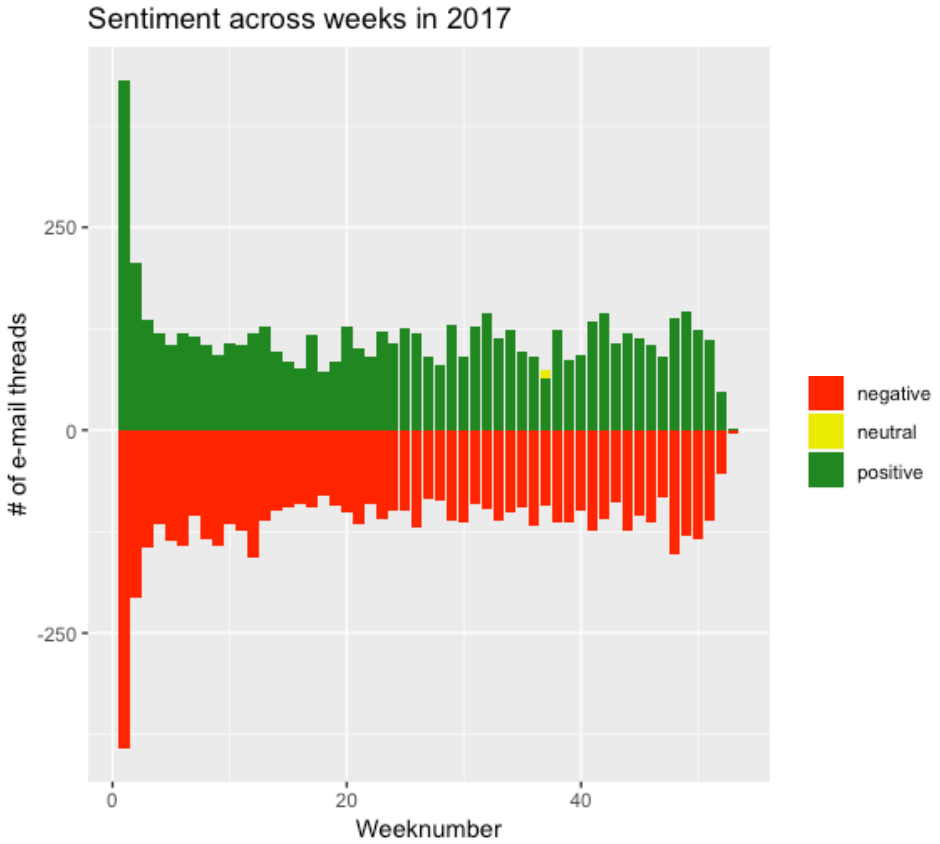


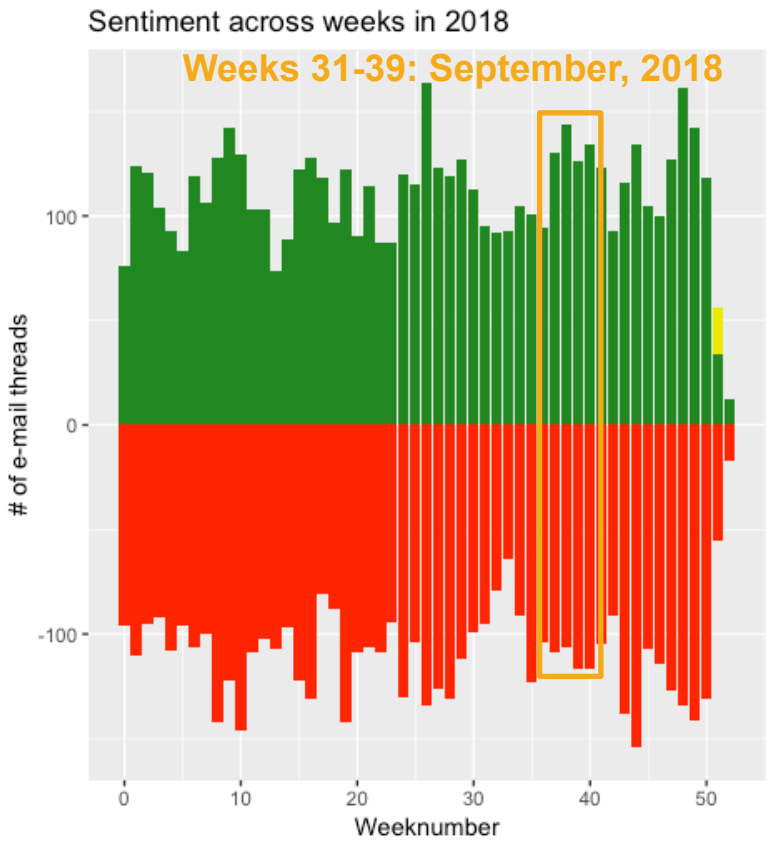
The sentiment fluctuates among weeks



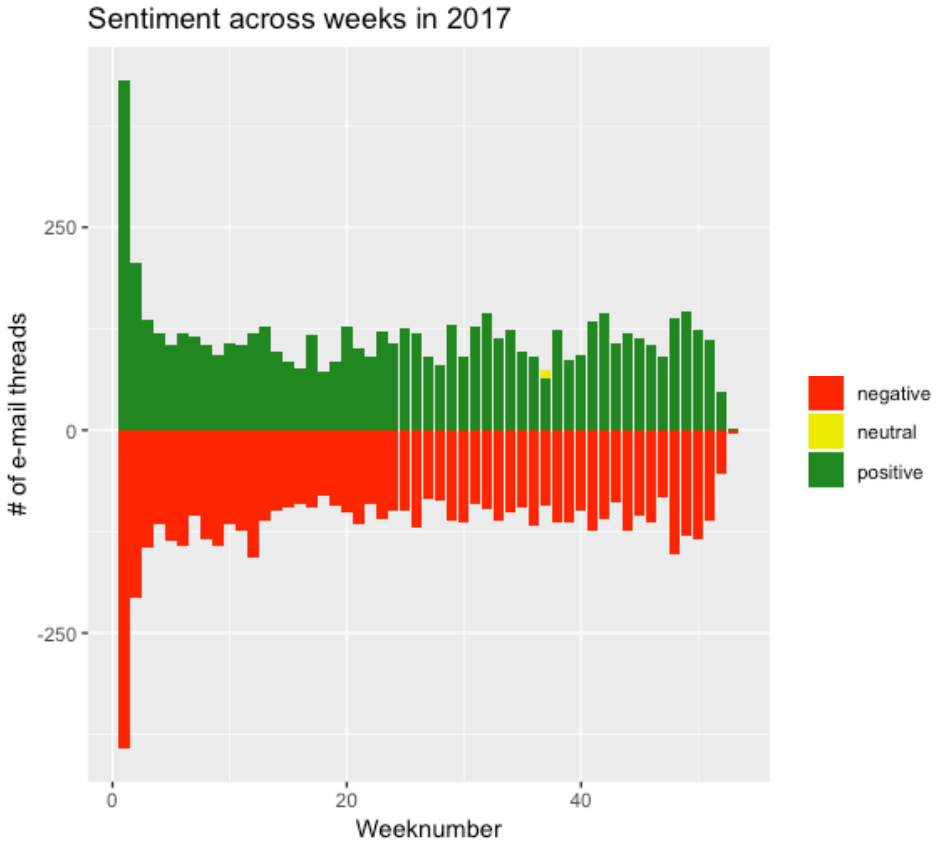


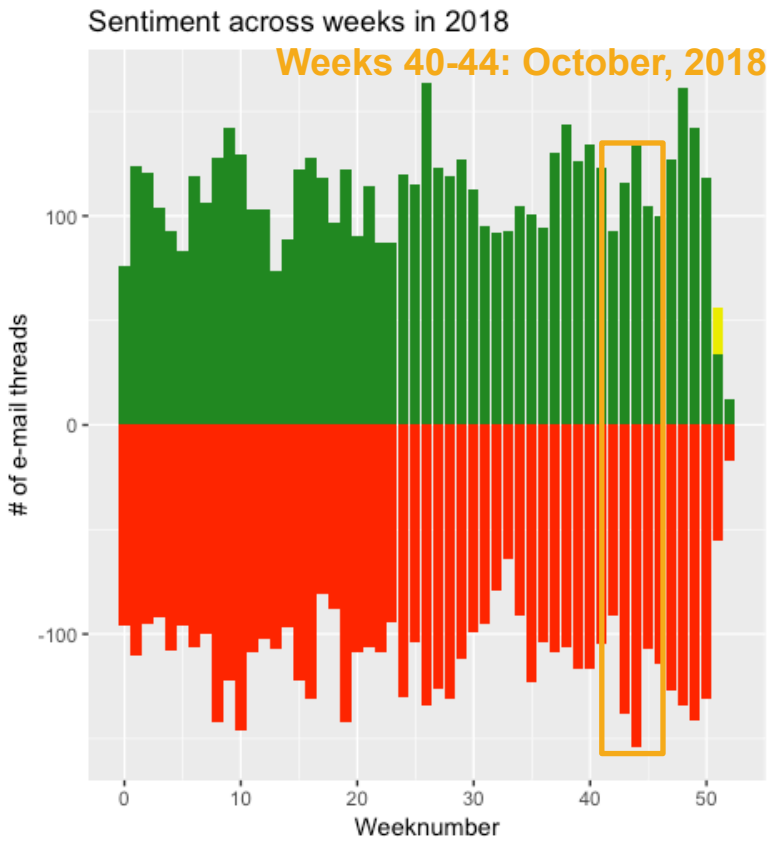
The sentiment fluctuates among weeks



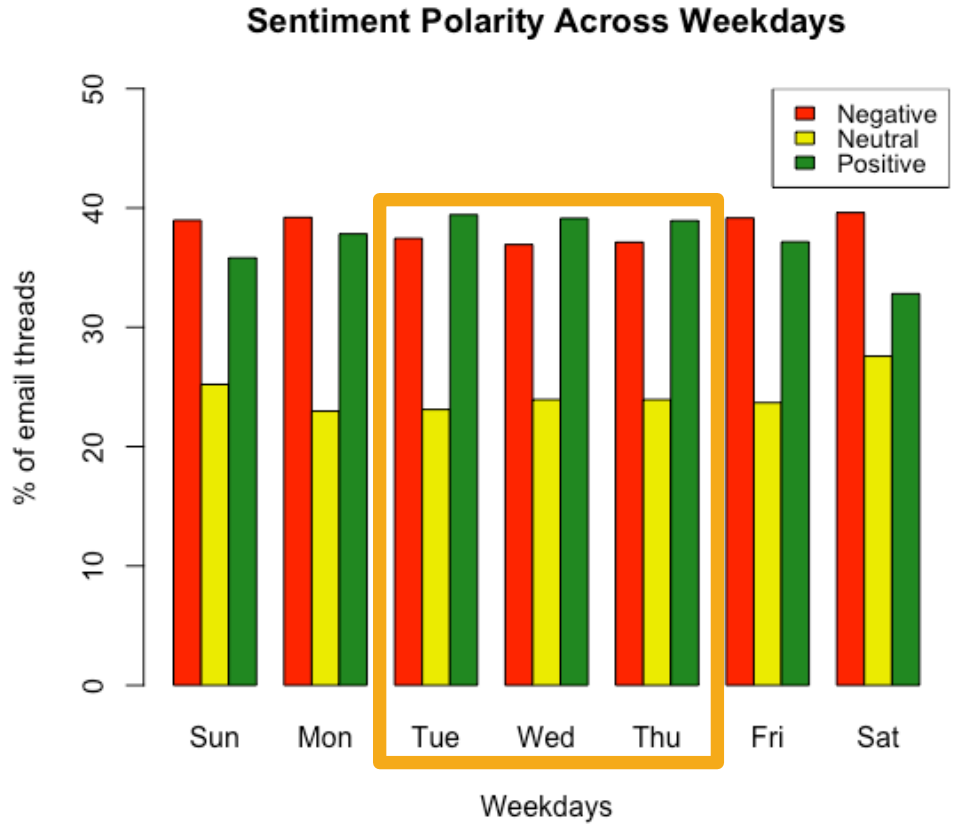


The sentiment fluctuates among weeks

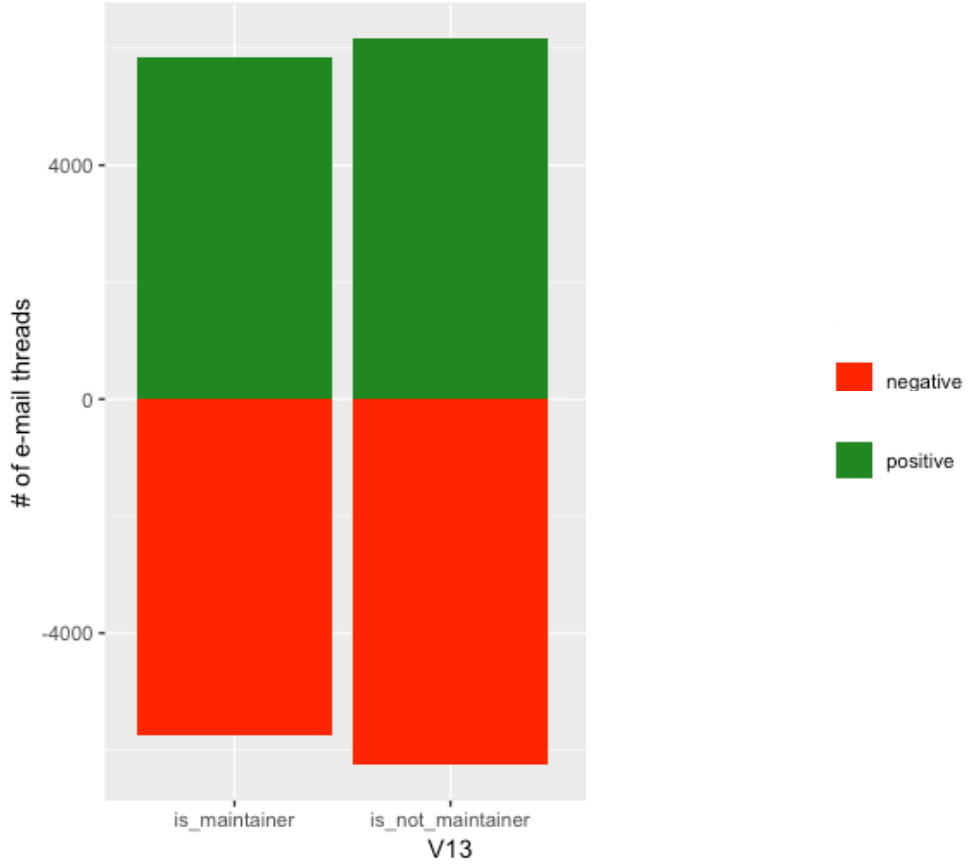




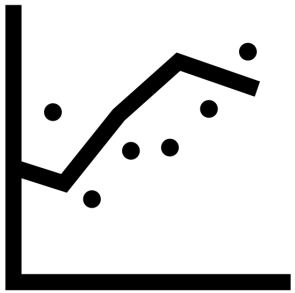
Tuesdays, Wednesdays, and Thursdays have more positive discussions



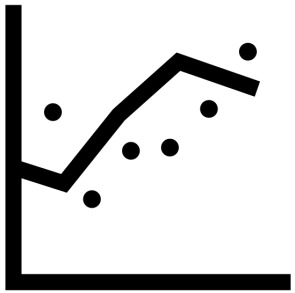
The sentiment does not differ between maintainers and non-maintainers



Challenges and Ongoing Work

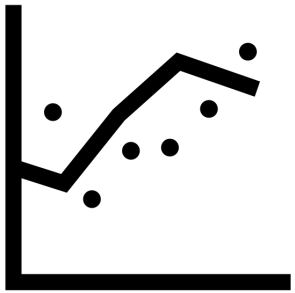


Noisy data



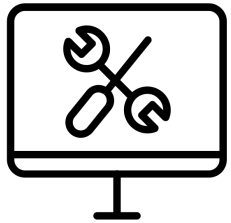
Noisy data

Noisy
tools

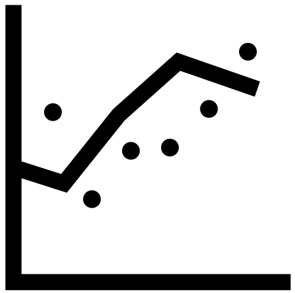


Noisy data

Noisy tools

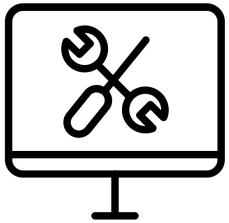


Target of sentiment?

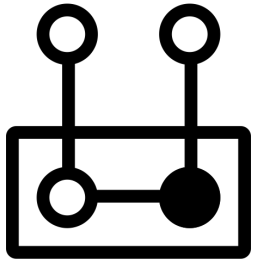


Noisy data

Noisy tools



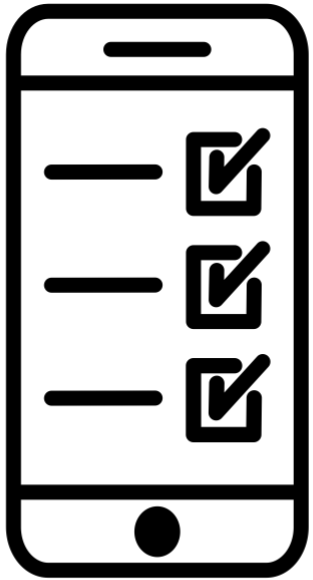
Target of sentiment?



How to act on results?

Let’s do a quick survey?

PollEv.com/isabellaferr986



Can you define the sentiment of the following sentences?

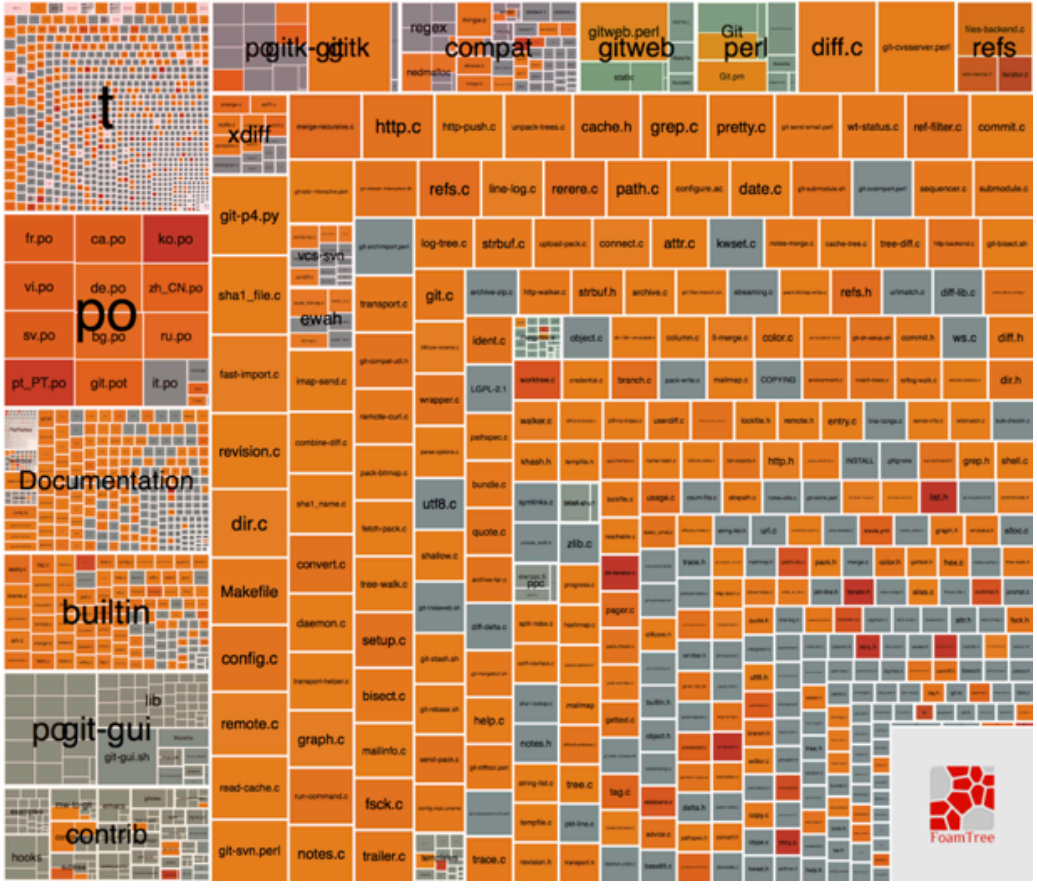
When survey is active, respond at **PollEv.com/isabellaferr986**

0 surveys done

1 survey underway

Start the presentation to see live content. Still no live content? Install the app or get help at **PollEv.com/app**

What kind of tool support could help the community?

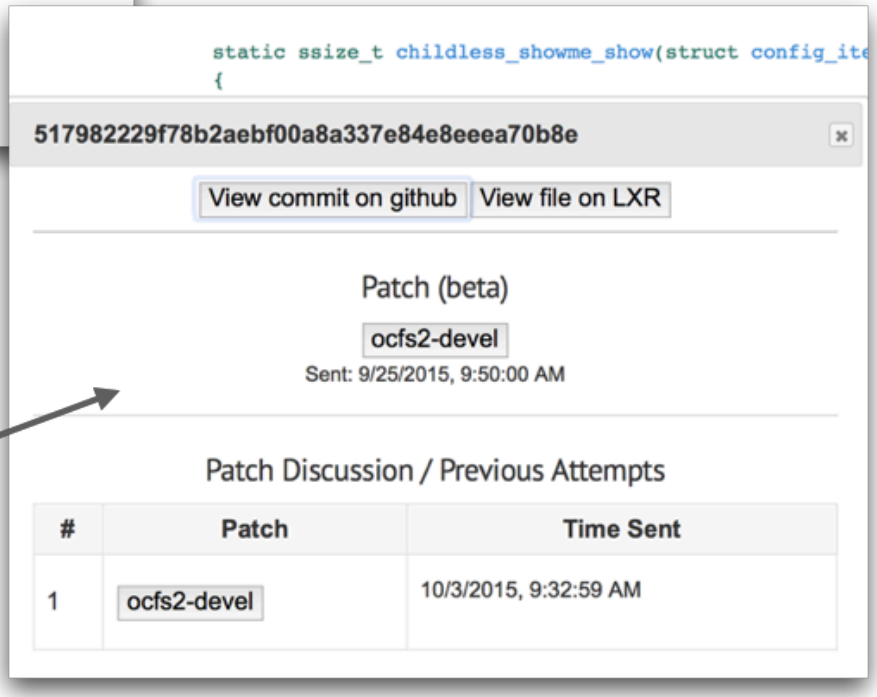


Cregit tracks patches at token level

```
static ssize_t childless_showme_show(struct config_item *item, char *page)
{
        struct childless *childless = to_childless(item);
        ssize_t pos;

        pos = sprintf(page, "%d\n", childless->showme);
        childless->showme++;

        return pos;
}
```



Links to mailing list reviews generated by **email2git**

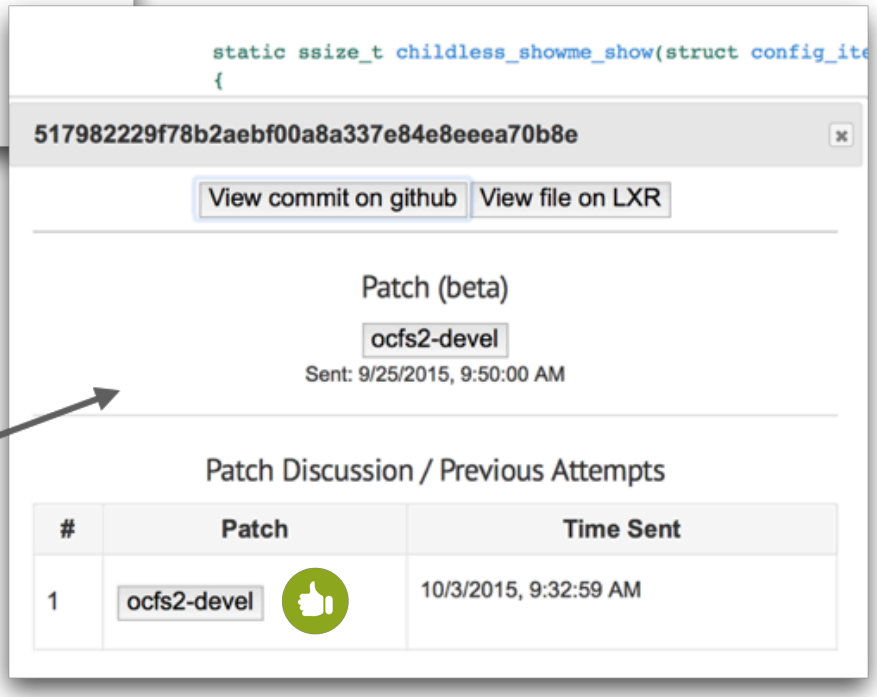
https://cregit.linuxsources.org

Cregit tracks patches at token level

```
static ssize_t childless_showme_show(struct config_item *item, char *page)
{
        struct childless *childless = to_childless(item);
        ssize_t pos;

        pos = sprintf(page, "%d\n", childless->showme);
        childless->showme++;

        return pos;
}
```



Links to mailing list reviews generated by **email2git**

https://cregit.linuxsources.org

Isabella Ferreira

Polytechnique Montréal

MCIS Lab

isabella.ferreira@polymtl.ca

/isabellaferreira

/isabellavieira

@isaferreira_57

…, but a Code of Conduct just treats the Symptoms, what is the Root of the Problem?

ci·vil·i·ty
/səˈvilədē/

noun

formal politeness and courtesy in behavior or speech.
"I hope we can treat each other with civility and respect"
synonyms: courtesy, courteousness, politeness, good manners, mannerliness, gentlemanliness, chivalry, gallantry, graciousness, consideration, respect, gentility; More

- polite remarks used in formal conversation.
  plural noun: **civilities**
  "she was exchanging civilities with his mother"
  synonyms: polite remark, politeness, courtesy; formality
  "she didn't waste time on civilities"

Isabella Ferreira
Polytechnique Montréal
MCIS Lab

isabella.ferreira@polymtl.ca

/isabellaferreira

/isabellavieira

@isaferreira_57

…, but a Code of Conduct just treats the Symptoms, what is the Root of the Problem?

ci·vil·i·ty
/səˈvilədē/

noun

formal politeness and courtesy in behavior or speech.
"I hope we can treat each other with civility and respect"
synonyms: courtesy, courteousness, politeness, good manners, mannerliness, gentlemanliness, chivalry, gallantry, graciousness, consideration, respect, gentility; More

- polite remarks used in formal conversation.
  plural noun: **civilities**
  "she was exchanging civilities with his mother"
  synonyms: polite remark, politeness, courtesy; formality
  "she didn't waste time on civilities"

Noisy data

Noisy tools

Target of sentiment?

How to act on results?

Isabella Ferreira
Polytechnique Montréal
MCIS Lab

isabella.ferreira@polymtl.ca

/isabellaferreira

/isabellavieira

@isaferreira_57

…, but a Code of Conduct just treats the Symptoms, what is the Root of the Problem?

ci·vil·i·ty
/səˈvilədē/

noun

formal politeness and courtesy in behavior or speech.
"I hope we can treat each other with civility and respect"
synonyms: courtesy, courteousness, politeness, good manners, mannerliness, gentlemanliness, chivalry, gallantry, graciousness, consideration, respect, gentility; More

- polite remarks used in formal conversation.
  plural noun: **civilities**
  "she was exchanging civilities with his mother"
  synonyms: polite remark, politeness, courtesy; formality
  "she didn't waste time on civilities"

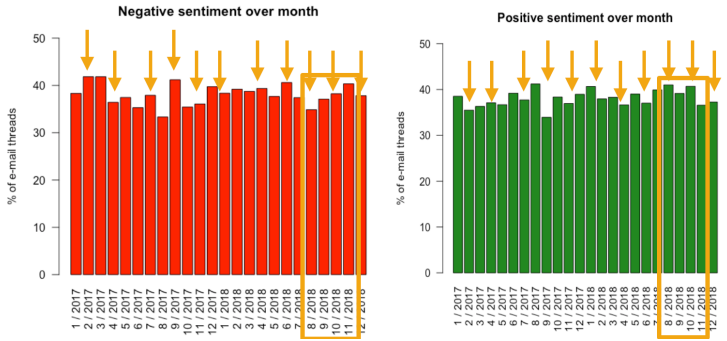
Noisy data

Noisy tools

Target of sentiment?

How to act on results?

Specific events might change the sentiment…



Isabella Ferreira
Polytechnique Montréal
MCIS Lab

isabella.ferreira@polymtl.ca

/isabellaferreira

/isabellavieira

@isaferreira_57

…, but a Code of Conduct just treats the Symptoms, what is the Root of the Problem?

ci·vil·i·ty
/səˈvilədē/

noun

formal politeness and courtesy in behavior or speech.
"I hope we can treat each other with civility and respect"
synonyms: courtesy, courteousness, politeness, good manners, mannerliness, gentlemanliness, chivalry, gallantry, graciousness, consideration, respect, gentility; More

- polite remarks used in formal conversation.
  plural noun: **civilities**
  "she was exchanging civilities with his mother"
  synonyms: polite remark, politeness, courtesy; formality
  "she didn't waste time on civilities"

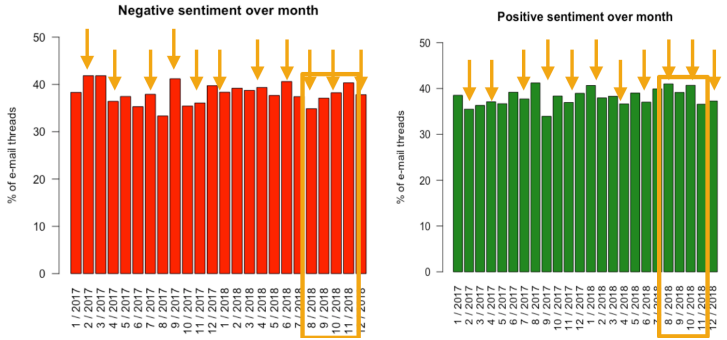
Noisy data

Noisy tools

Target of sentiment?

How to act on results?

Specific events might change the sentiment…



We need your feedback!

Isabella Ferreira
Polytechnique Montréal
MCIS Lab

isabella.ferreira@polymtl.ca

/isabellaferreira

/isabellavieira

@isaferreira_57

…, but a Code of Conduct just treats the Symptoms, what is the Root of the Problem?

ci·vil·i·ty
/səˈvilədē/

noun

formal politeness and courtesy in behavior or speech.
"I hope we can treat each other with civility and respect"
synonyms: courtesy, courteousness, politeness, good manners, mannerliness, gentlemanliness, chivalry, gallantry, graciousness, consideration, respect, gentility; More

- polite remarks used in formal conversation.
  plural noun: **civilities**
  "she was exchanging civilities with his mother"
  synonyms: polite remark, politeness, courtesy; formality
  "she didn't waste time on civilities"

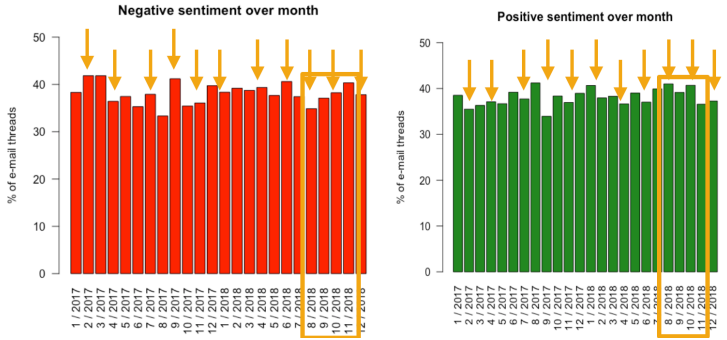
Noisy data

Noisy tools

Target of sentiment?

How to act on results?

Specific events might change the sentiment…



We need your feedback!

SEmotion 2019

Fourth International Workshop on Emotion Awareness in Software Engineering

ICSE 2019 Workshop · 28 May 2019 · Montréal, QC, Canada

Ferreira et al. A Longitudinal Study on the Maintainers' Sentiment of a Large Scale Open Source Ecosystem

Isabella Ferreira
Polytechnique Montréal
MCIS Lab

isabella.ferreira@polymtl.ca

/isabellaferreira

/isabellavieira

@isaferreira_57