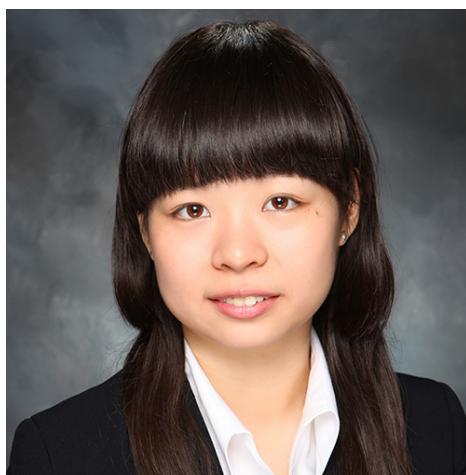


The Signals that Potential Contributors Look for When Choosing Open-source Projects



Sophie Qiu



Lily Li



Susmita Padala



Anita Sarma



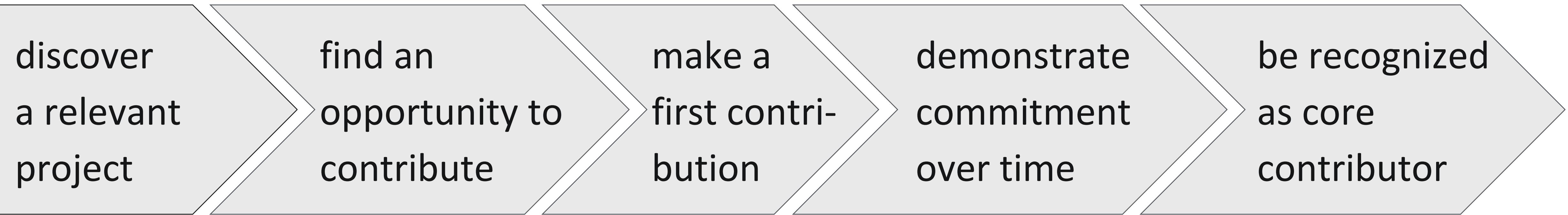
Bogdan Vasilescu

Goal: Increase participation in Open Source

- To sustain open source projects that need constant supply of effort
- To provide opportunities to learn / develop skills / showcase talent
- To increase diversity
- ...

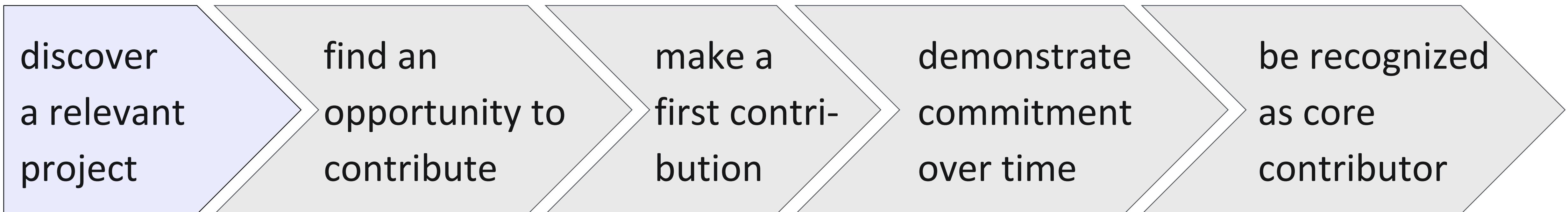


Participation in Open Source has a long history of scholarship



Yet, little is known about how people choose which project to contribute to

Especially for people who contribute for fun, to learn, etc.



Project discovery
is especially relevant on
transparent platforms
like GitHub

GitHub users rely on signals for many inferences

(Dabbish et al., 2012)

The screenshot shows a GitHub repository page for 'charts'. The top navigation bar includes 'Watch' (1), 'Star' (2), and 'Fork' (2) buttons, all circled in red. Below the navigation bar are tabs for 'Code', 'Issues 17', 'Pull requests 1', 'Projects 6', 'Wiki', 'Security', and 'Insights'. A banner at the top states 'Data visualization and charting library for angular deployments.' with a link to <https://charts.piknik.co>. The repository details section shows '667 commits', '2 branches', '20 releases', '2 contributors', and 'MIT' license, with the '667 commits' count circled in red. At the bottom, a list of recent commits includes 'Update angular deps' (2 years ago), '.github' (2 years ago), '.vscode' (4 months ago), 'apps' (4 months ago), and 'build' (2 years ago). The 'Clone or download' button is also circled in red.

Data visualization and charting library for angular deployments. <https://charts.piknik.co>

angular data-visualization d3 charts pioneer-charts pioneer-code

667 commits 2 branches 20 releases 2 contributors MIT

Branch: master ▾ New pull request Create new file Upload files Find file Clone or download ▾

Latest commit 667 on Jun 18

Commit	Description	Date
Update angular deps		2 years ago
.github	Add meta	2 years ago
.vscode	Update document project	4 months ago
apps	Fix table header bug	4 months ago
build	Fox docs build	2 years ago

GitHub users rely on signals for many inferences

(Trockman et al., 2018)

Status



What's included

Within the download you'll find the following directories and files, logically grouping common assets and providing both compiled and minified variations. You'll see something like this:

How do people use signals
when choosing a GitHub
project to contribute to?

Many signals used when choosing projects

Quality	README
	Contributing guideline
Community	Impolite language
	Issue template
Popularity	Stars
	Recent commits

Many signals used when choosing projects

Quality	README
	Contributing guideline
Community	Impolite language
	Issue template
Popularity	Stars
	Recent commits

The screenshot shows a GitHub repository page for 'PioneerAngular-Charts'. At the top, there's a file icon followed by 'README.md'. Below it, the repository name 'PioneerAngular-Charts' is displayed with a 'build passing' badge. Further down, there are download statistics ('downloads 5.8k') and an npm badge ('npm v1.0.2'). A 'Quick start' section contains instructions: 'To get started, it is highly recommended that you [Read the Getting Started guide](#), [review the configuration](#), and examples, and more.' Below this, there's a code block for running tests ('npm test') and another for installing dependencies ('npm install @pioneerangular/charts'). At the bottom, there's a 'Issues and feature requests' section.

Many signals used when choosing projects

Quality	README
	Contributing guideline
Community	Impolite language
	Issue template
Popularity	Stars
	Recent commits

Contributing

Please read through our [contributing guidelines](#). Included are instructions for development.

Moreover, if your pull request contains JavaScript patches or fixes, they should conform to the [Code Guide](#), maintained by [Mozilla](#).

Editor preferences are available in the [editor config](#) for easy usage:
<https://editorconfig.org/>.

Many signals used when choosing projects

Quality	README
	Contributing guideline
Community	Impolite language
	Issue template
Popularity	Stars
	Recent commits

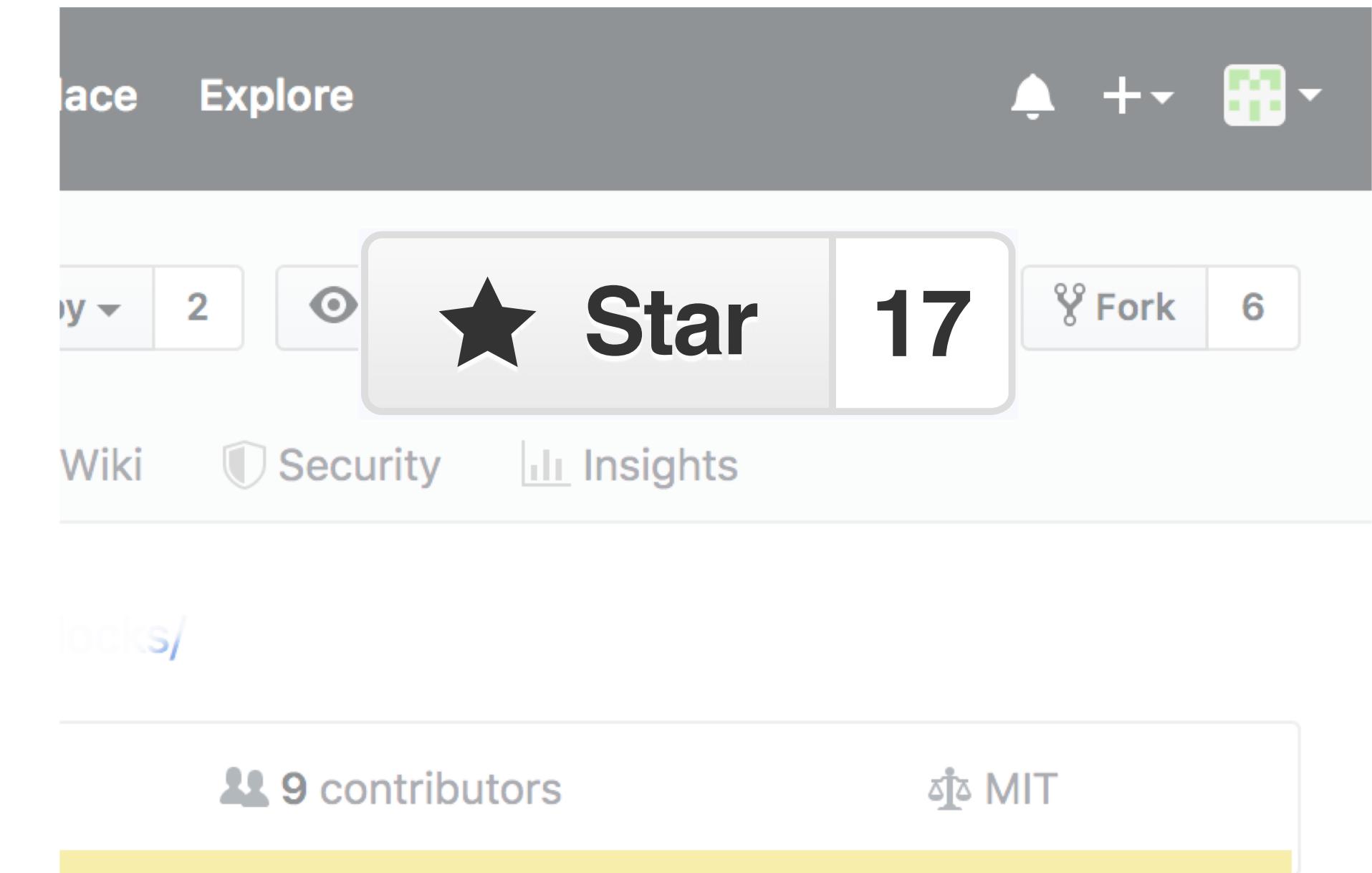
build	YES! FUCK YOU WEBPACK
dist	Precompile assets
docs	Experimenting with component data tables
public	Rederubfne
src	Move redirects again

Many signals used when choosing projects

Quality	README	## I'm submitting a... Please search GitHub for a similar issue or PR before submitting <pre><code> [] Regression (a behavior that used to work and stopped working in a new release) [] Bug report [] Feature request [] Documentation issue or request
	Contributing guideline	
Community	Impolite language	<pre><code> [] Regression (a behavior that used to work and stopped working in a new release) [] Bug report [] Feature request [] Documentation issue or request
	Issue template	
Popularity	Stars	<pre><code> [] Regression (a behavior that used to work and stopped working in a new release) [] Bug report [] Feature request [] Documentation issue or request
	Recent commits	

Many signals used when choosing projects

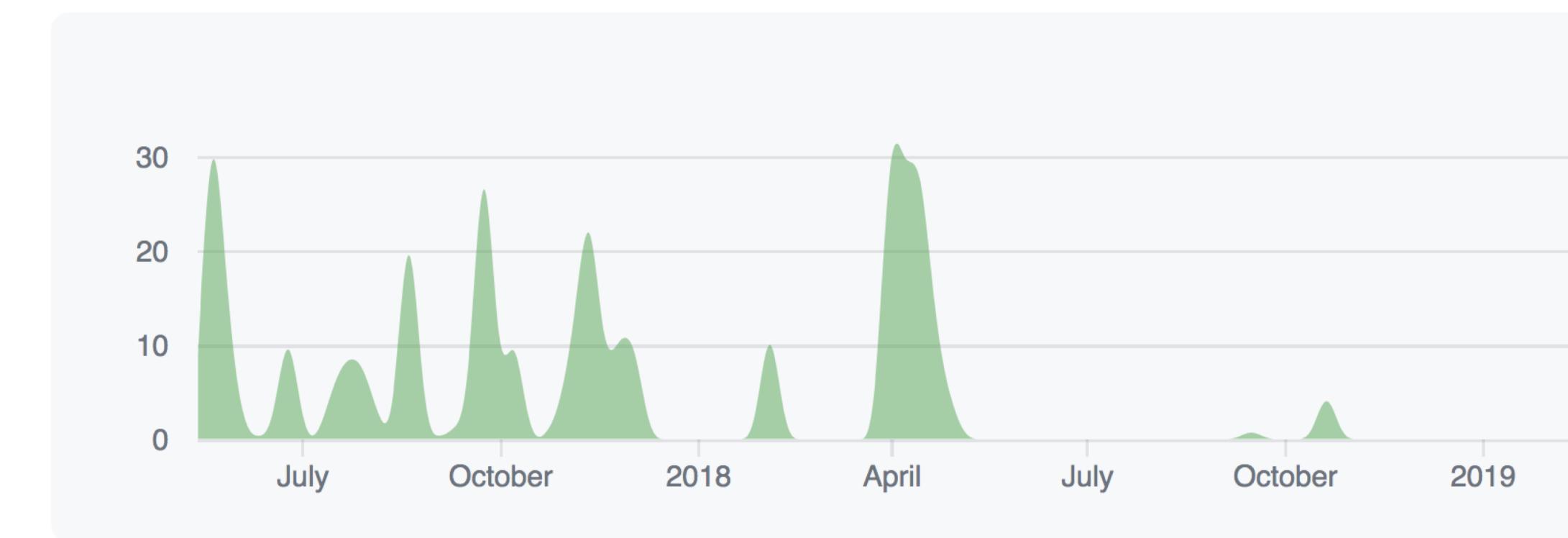
Quality	README
	Contributing guideline
Community	Impolite language
	Issue template
Popularity	Stars
	Recent commits



Many signals used when choosing projects

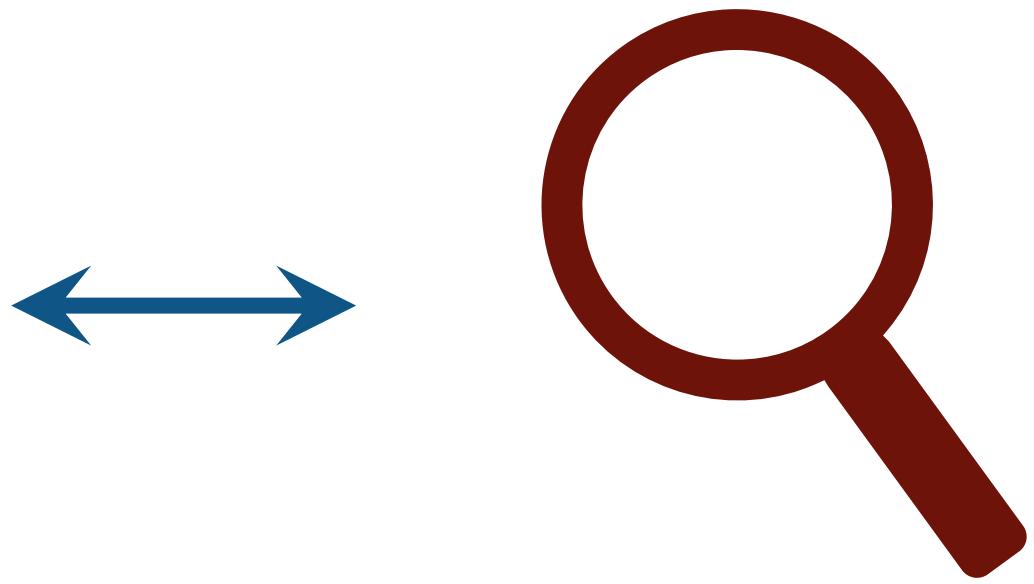
Quality	README
	Contributing guideline
Community	Impolite language
	Issue template
Popularity	Stars
	Recent commits

Contributions to master, excluding merge commits



Methodology

A mixed-method study design - Qualitative

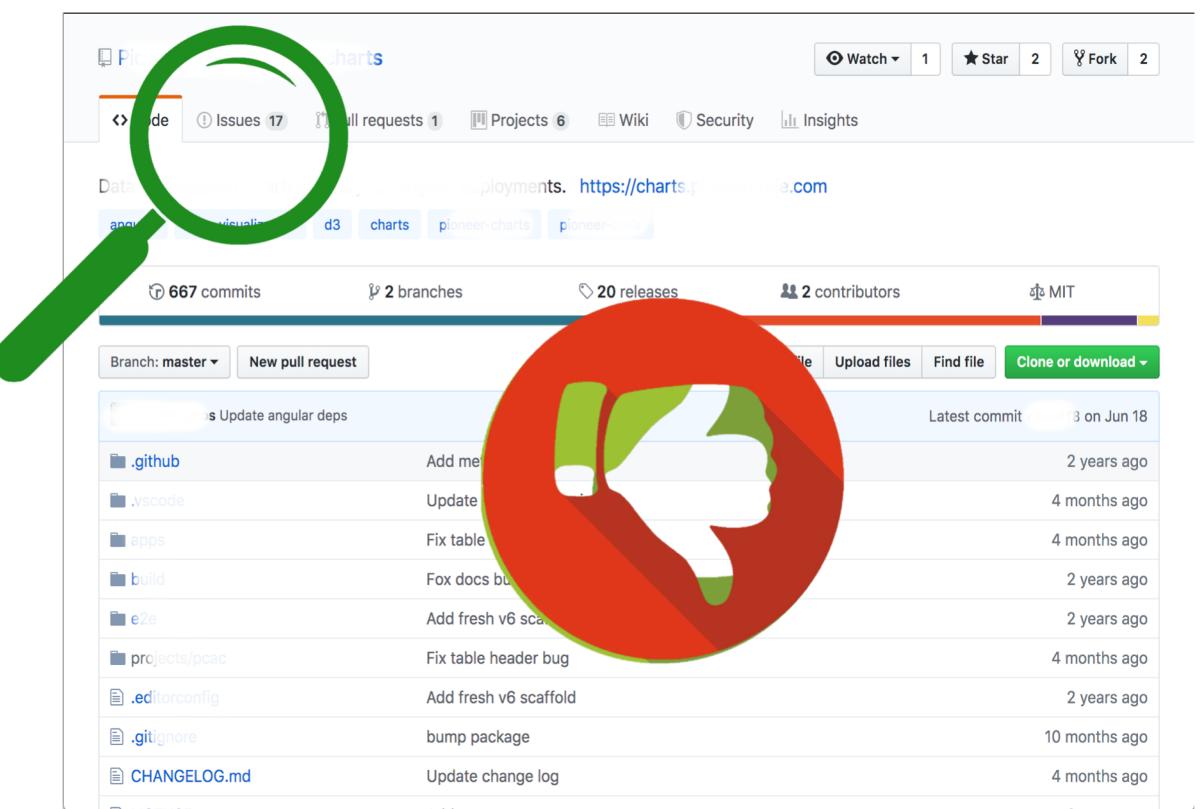
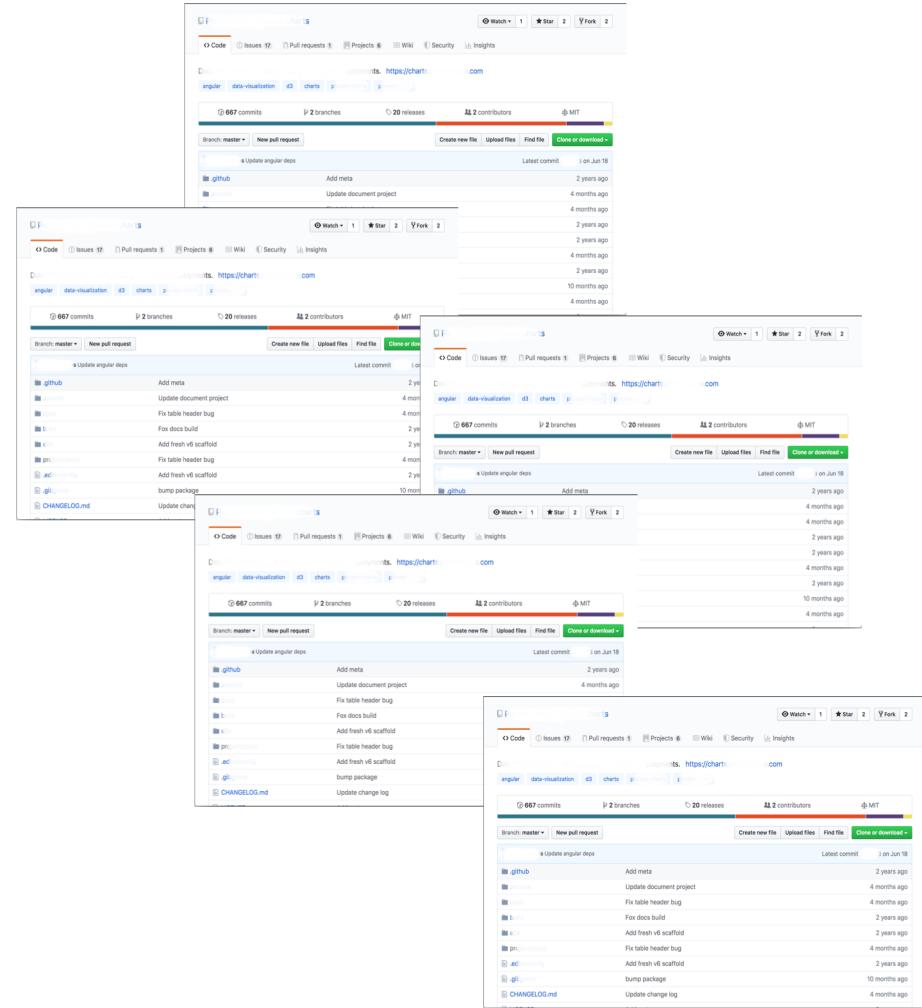


Interviews with
15 developers

Qualitative
analysis



Qualitative: Review and recommend projects



Jane,
Bachelor degree in
Computer Science
First-time contributor to npm

Pre-selected projects

Shown a set of NPM
projects' GitHub pages

Think-aloud interviews

Describe how they
evaluate each project

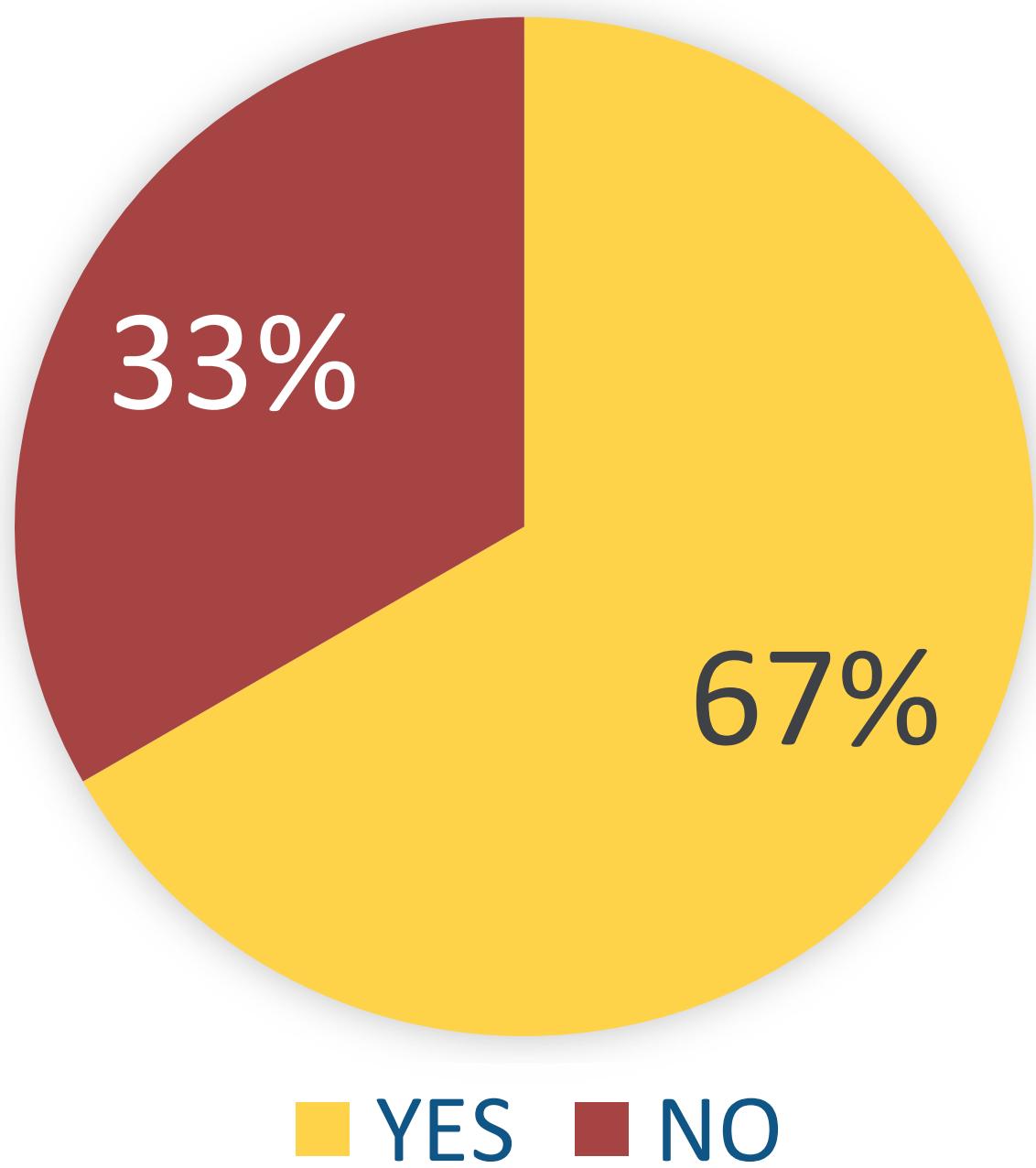
Persona

Decide for each project,
if they would
recommend it to Jane

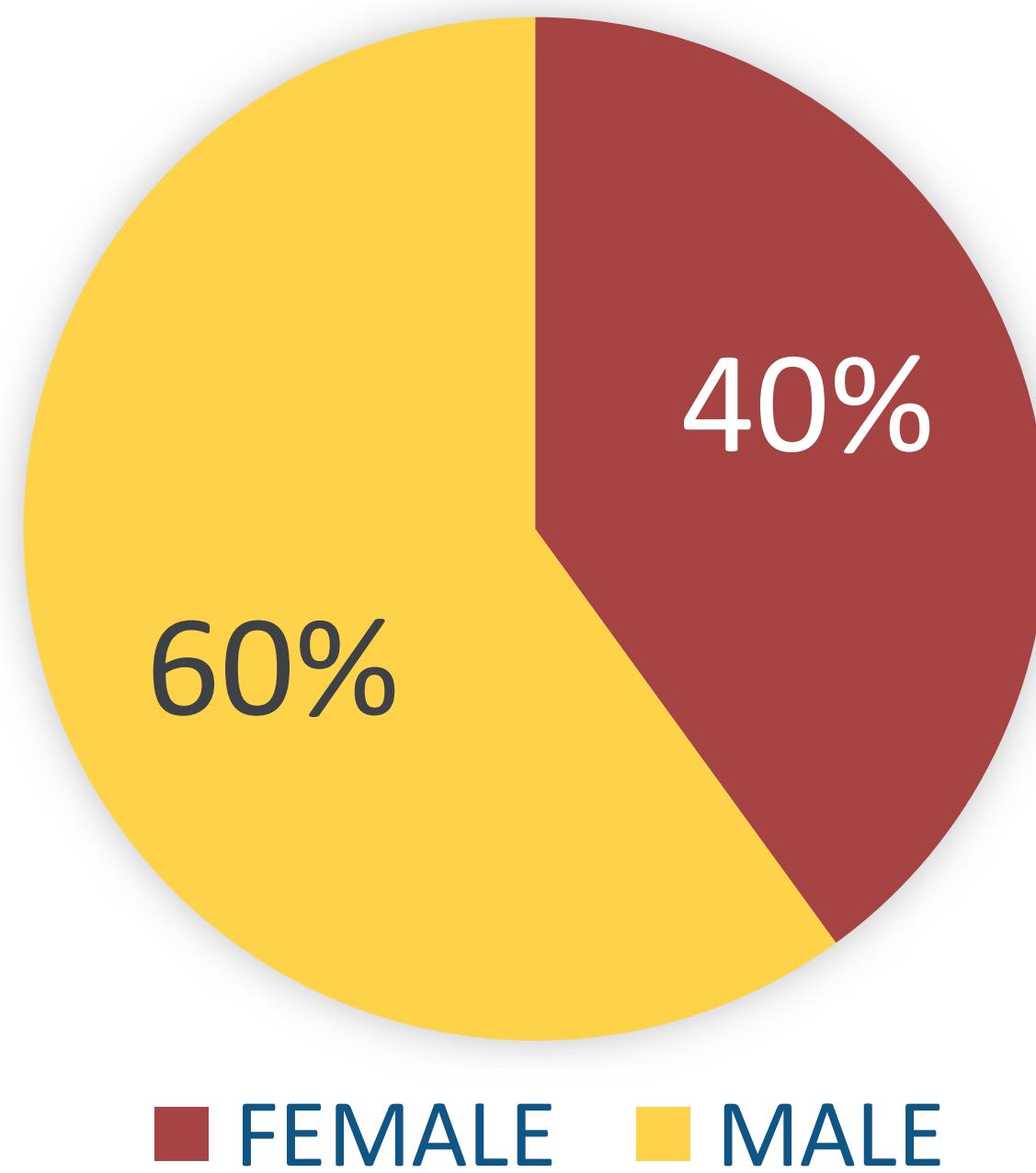


Interviewees came from diverse backgrounds

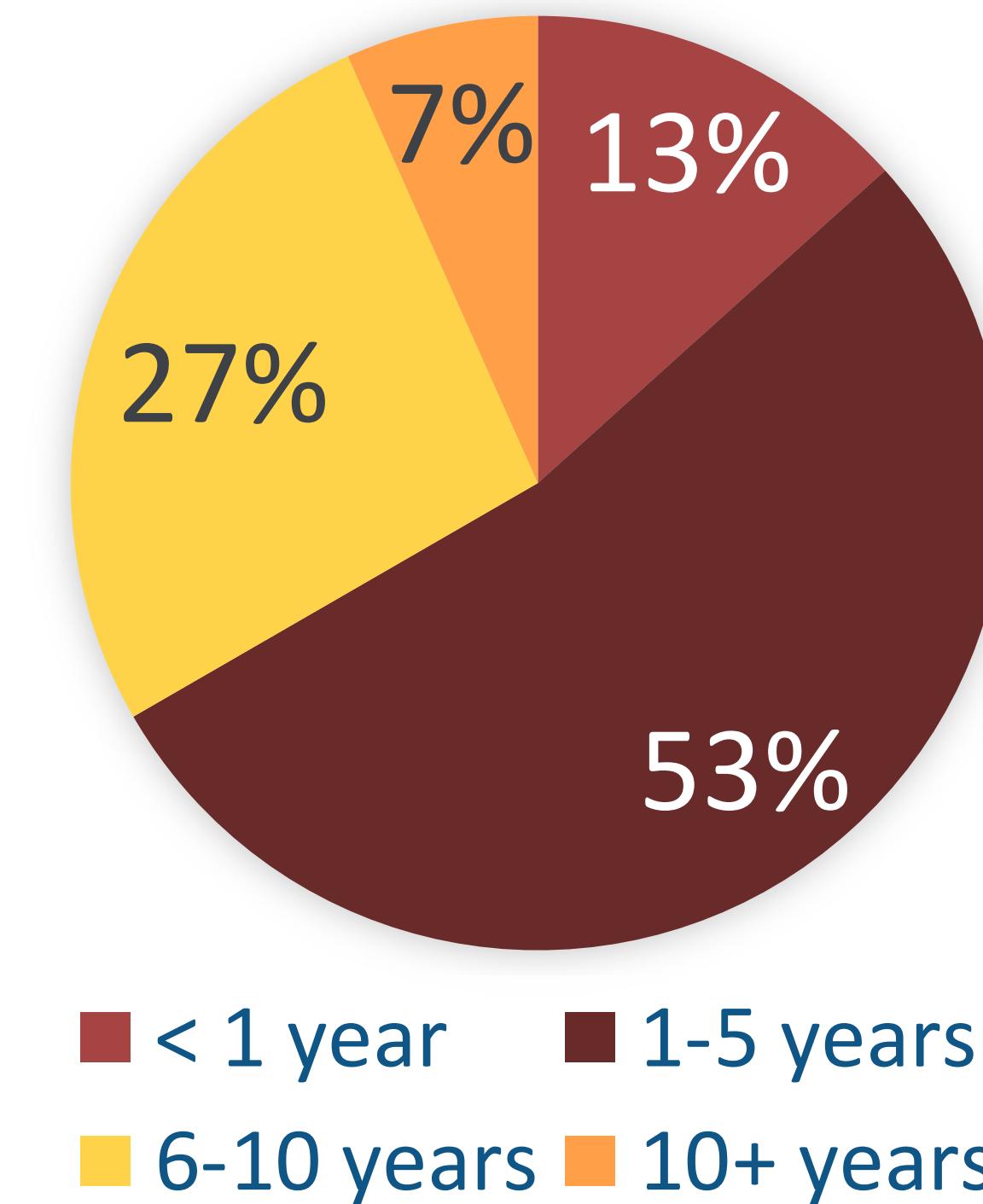
Full-time developers



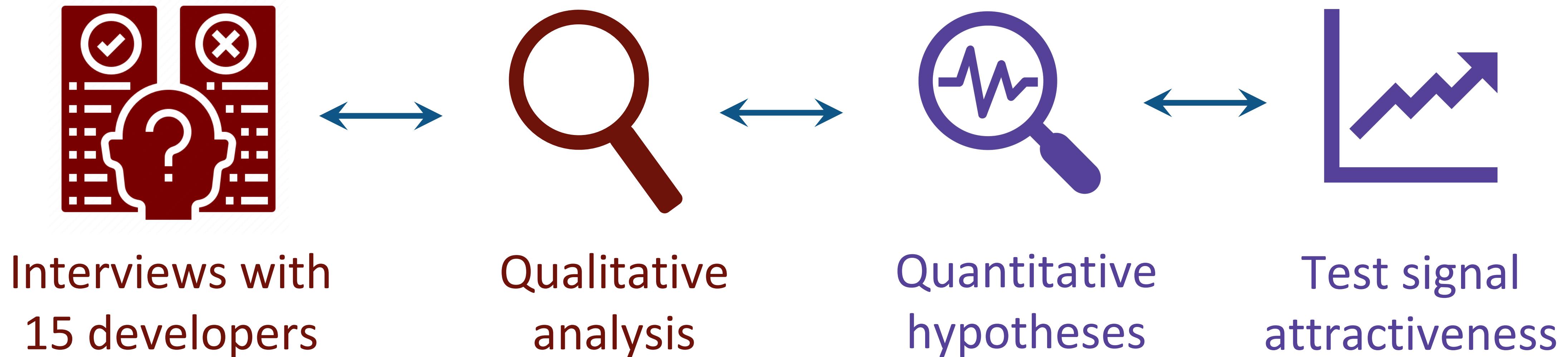
Gender



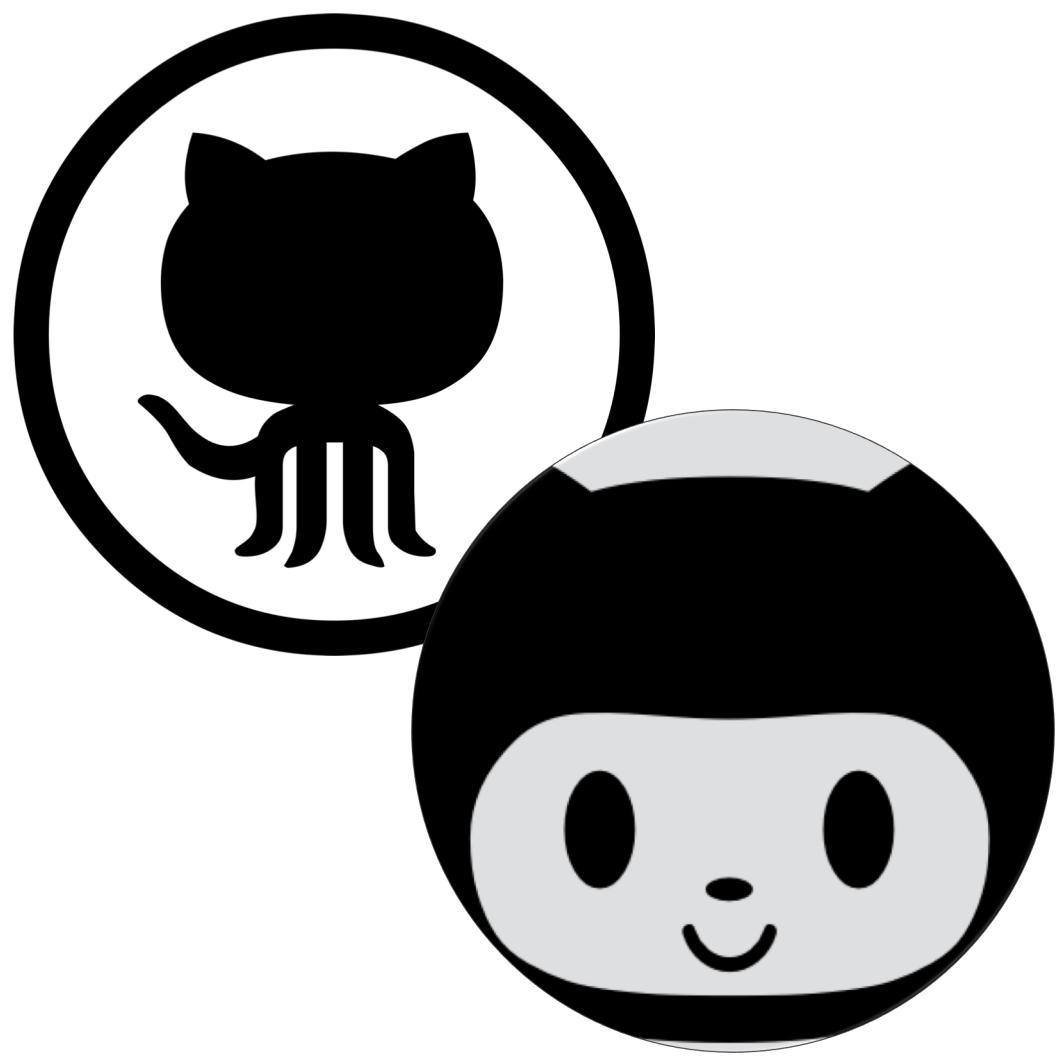
Years of OSS experience



A mixed-method study design - Quantitative



↗ Quantitative: GitHub data mining



Mining trace data

~10,000
npm packages from
GitHub and GHTorrent

Operationalization

14 quantitative
measurements

Logistic regression

Associations between
signals and newcomers

Result highlights: Many signals used when choosing projects

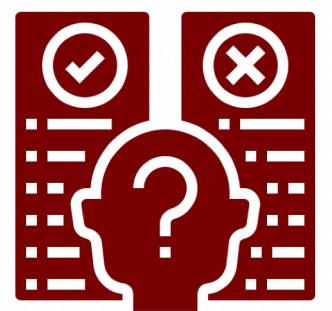
		Interviews	Logistic Regression
Quality	README		
	Contributing guideline		
Community	Impolite language		
	Issue template		
Popularity	Stars		
	Recent commits		

Three main classes of signals

Quality

Community

Popularity



Interviews:

“README is a project’s welcome mat”

Data Visualization Charting library for Angular deployments. <https://github.com/pioneer-charts/pioneer-charts>

#angular #data-visualization #d3 #charts #pioneer-charts #pioneer-charts

667 commits

2 branches

20 releases

2 contributors

MIT

Branch: master

New pull request

Create new file

Upload files

Find file

Clone or download

Update angular deps	Latest commit ↗ on Jun 18
.github	Add meta
.vscode	Update document project
tslint.json	Update to Angular 8
README.md	4 months ago

README.md

Playground Examples build passing

downloads 5.8k npm v1.0.2

Quick start

To get started, it is highly recommended that you [Read the Getting Started](#) page for information on setup, configuration and examples, and more.

`npm install @pioneer-charts/pioneer-charts`

`npm`

Bugs and feature requests

Have a bug or a feature request? Please first read the [issue guidelines](#) and search for existing and closed issues. If your



“the README is most important. It should describe without having to navigate away from that page the key information people need”

Project [Issues](#) [Pull requests](#) [Build](#) [passing](#)

downloads 5.8k npm v1.0.2

Quick start

To get started, it is highly recommended that you [Read the Getting Started](#) page for information on setup, configuration and examples, and more.

```
npm install @picocss/picocss@next-peer-charts  
npm
```

Bugs and feature requests

Have a bug or a feature request? Please first read the [issue guidelines](#) and search for existing and closed issues. If your problem or idea is not addressed yet, [please open a new issue](#).

Documentation

Picocss's documentation, of which is [included in this repo](#), is built with [Angular](#) and publicly hosted on GitHub Pages at <https://carnegiemellon-university.github.io/picocss/>. The docs may also be run locally.

Contributing

Please read through our [contributing guidelines](#). Included are directions for opening issues, coding standards, and notes on development.

Editor preferences are available in the [editor config](#) for easy use in common text editors. Read more and download plugins at <http://editorconfig.org/>.



A nice README should have: a table of contents

README.md



Bootstrap

Sleek, intuitive, and powerful front-end framework for faster and easier web development.

[Explore Bootstrap docs »](#)

[Report bug](#) · [Request feature](#) · [Themes](#) · [Blog](#)

Table of contents

- [Quick start](#)
- [Status](#)
- [What's included](#)
- [Bugs and feature requests](#)
- [Documentation](#)
- [Contributing](#)
- [Community](#)



A nice README should have: descriptions and instructions

README.md

Project Issues Pull requests Artifacts build passing

downloads 5.8k npm v1.0.2

Quick start

To get started, it is highly recommended that you [Read the Getting Started](#) page for information on setup, configuration and examples, and more.

`npm install @picocss/picocss@next`

`npm`

Bugs and feature requests

Have a bug or a feature request? Please first read the [issue guidelines](#) and search for existing and closed issues. If your problem or idea is not addressed yet, [please open a new issue](#).

Documentation

Picocss's documentation, of which is [included in this repo](#), is built with [Angular](#) and publicly hosted on GitHub Pages at <https://carnegiemellon-strudel.s3-us-west-2.amazonaws.com/>. The docs may also be run locally.

Contributing

Please read through our [contributing guidelines](#). Included are directions for opening issues, coding standards, and notes on development.

Editor preferences are available in the [editor config](#) for easy use in common text editors. Read more and download plugins at <http://editorconfig.org/>.



A nice README should have: contact information

Bugs and feature requests

Have a bug or a feature request? Please first read the [issue guidelines](#) and search for existing and closed issues. If your problem or idea is not addressed yet, [please open a new issue](#).

Documentation

This project's documentation, of which is included in this repo, is built with [Angular](#) and publicly hosted on GitHub Pages at <https://strudel.cs.cmu.edu/>. The docs may also be run locally.

Contributing

Please read through our [contributing guidelines](#). Included are directions for opening issues, coding standards, and notes on development.

Editor preferences are available in the [editor config](#) for easy use in common text editors. Read more and download plugins at <http://editorconfig.org/>.

Creator

Chad Morris

- [Twitter](#)
- [GitHub](#)



A nice README should have: code quality badges

Status

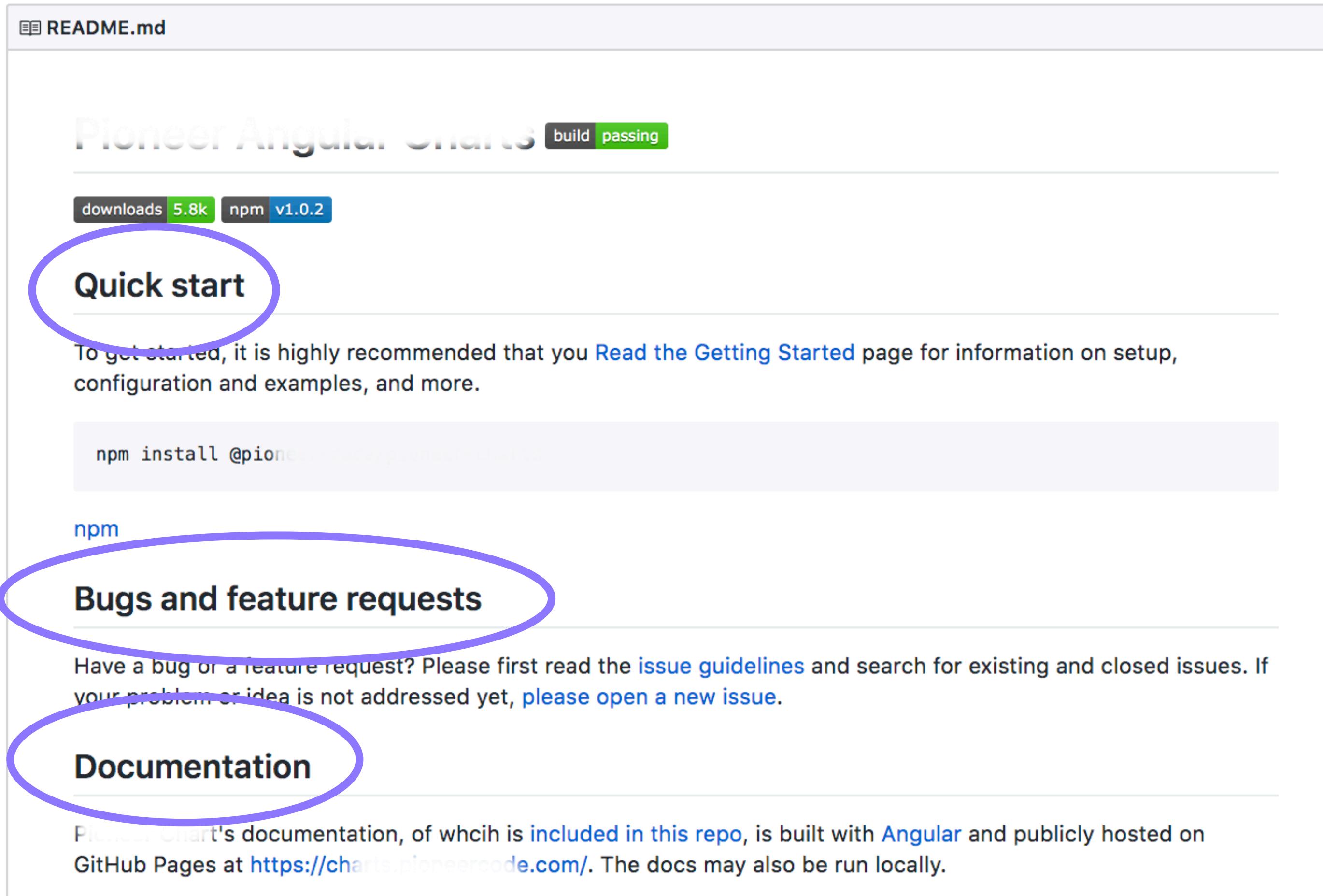
 Tests passing  npm v4.3.1  gem v4.3.1  meteor twbs:bootstrap  dev dependencies up to date  coverage 94%  CSS gzip size 21.2 kB

What's included

Within the download you'll find the following directories and both compiled and minified variations. You'll see something



Logistic regression: Counting the number of headers in README



The screenshot shows the contents of a README.md file for a GitHub repository named "Pioneer Angular Charts". The file includes sections for "Quick start", "Bugs and feature requests", and "Documentation". Three sections are highlighted with blue circles: "Quick start", "Bugs and feature requests", and "Documentation".

Quick start

To get started, it is highly recommended that you [Read the Getting Started page](#) for information on setup, configuration and examples, and more.

```
npm install @pioneer/chart
```

Bugs and feature requests

Have a bug or a feature request? Please first read the [issue guidelines](#) and search for existing and closed issues. If your problem or idea is not addressed yet, [please open a new issue](#).

Documentation

Pioneer Chart's documentation, of which is [included in this repo](#), is built with [Angular](#) and publicly hosted on GitHub Pages at <https://charts.pioneercode.com/>. The docs may also be run locally.

Logistic regression: A well-structured README correlates to higher likelihood of attracting newcomers

README.md

Pioneer Angular Charts build passing

downloads 5.8k npm v1.0.2

Quick start

To get started, it is highly recommended that you [Read the Getting Started page](#) for information on setup, configuration and examples, and more.

```
npm install @pioneer/chart
```

npm

Bugs and feature requests

Have a bug or a feature request? Please first read the [issue guidelines](#) and search for existing and closed issues. If your problem or idea is not addressed yet, [please open a new issue](#).

Documentation

Pioneer Chart's documentation, of which is [included in this repo](#), is built with [Angular](#) and publicly hosted on GitHub Pages at <https://charts.pioneercode.com/>. The docs may also be run locally.

Other README signals in the paper





Interviews:

Contributing file “suggests they have some experience with handling new contributors”

.test.eslintrc.js	configuring
.travis.yml	Removed travis node version
AUTHORS	changed (from AUTHORS) Fix npm user
CONTRIBUTING.md	changed (from CONTRIBUTING.md) use yarn start to l
LICENSE	added (from LICENSE) e): include 2018 in
README.md	changed (from README.md) remove greenkeeper

To get started, it is highly recommended that you [Read the Getting Started](#) page for information on setup, configuration and examples, and more.

Bugs and feature requests

Documentation

Contributing

Please read through our [contributing guidelines](#). Included are directions for opening issues, coding standards, and notes on development.

Editor preferences are available in the [editor config](#) for easy use in common text editors. Read more and download plugins at <http://editorconfig.org/>.



A nice contributing guideline should be: prominent

README.md

downloads 5.8k npm v1.0.2

Quick start

To get started, it is highly recommended that you [Read the Getting Started page](#) for information on setup, configuration and examples, and more.

```
npm install @piaw/picard --save-dev  
npm
```

Bugs and feature requests

Have a bug or a feature request? Please first read the [issue guidelines](#) and search for existing and closed issues. If your problem or idea is not addressed yet, [please open a new issue](#).

Documentation

This project's documentation, of which is [included in this repo](#), is built with [Angular](#) and publicly hosted on GitHub Pages at <https://piaw.github.io/picard/>. The docs may also be run locally.

Contributing

Please read through our [contributing guidelines](#). Included are directions for opening issues, coding standards, and notes on development.

Editor preferences are available in the [editor config](#) for easy use in common text editors. Read more and download plugins at <http://editorconfig.org/>.



A nice contributing guideline should: explain the GitHub jargon

Using the issue tracker

The [issue tracker](#) is the preferred channel for [bug reports](#), [features requests](#) and [submitting pull requests](#), but please respect the following restrictions:

- Please **do not** derail or troll issues. Keep the discussion on topic and respect the opinions of others.



A nice contributing guideline should:
mention ways to contribute other than code

Using the issue tracker

The [issue tracker](#) is the preferred channel for [bug reports](#), [features requests](#) and [submitting pull requests](#), but please respect the following restrictions:

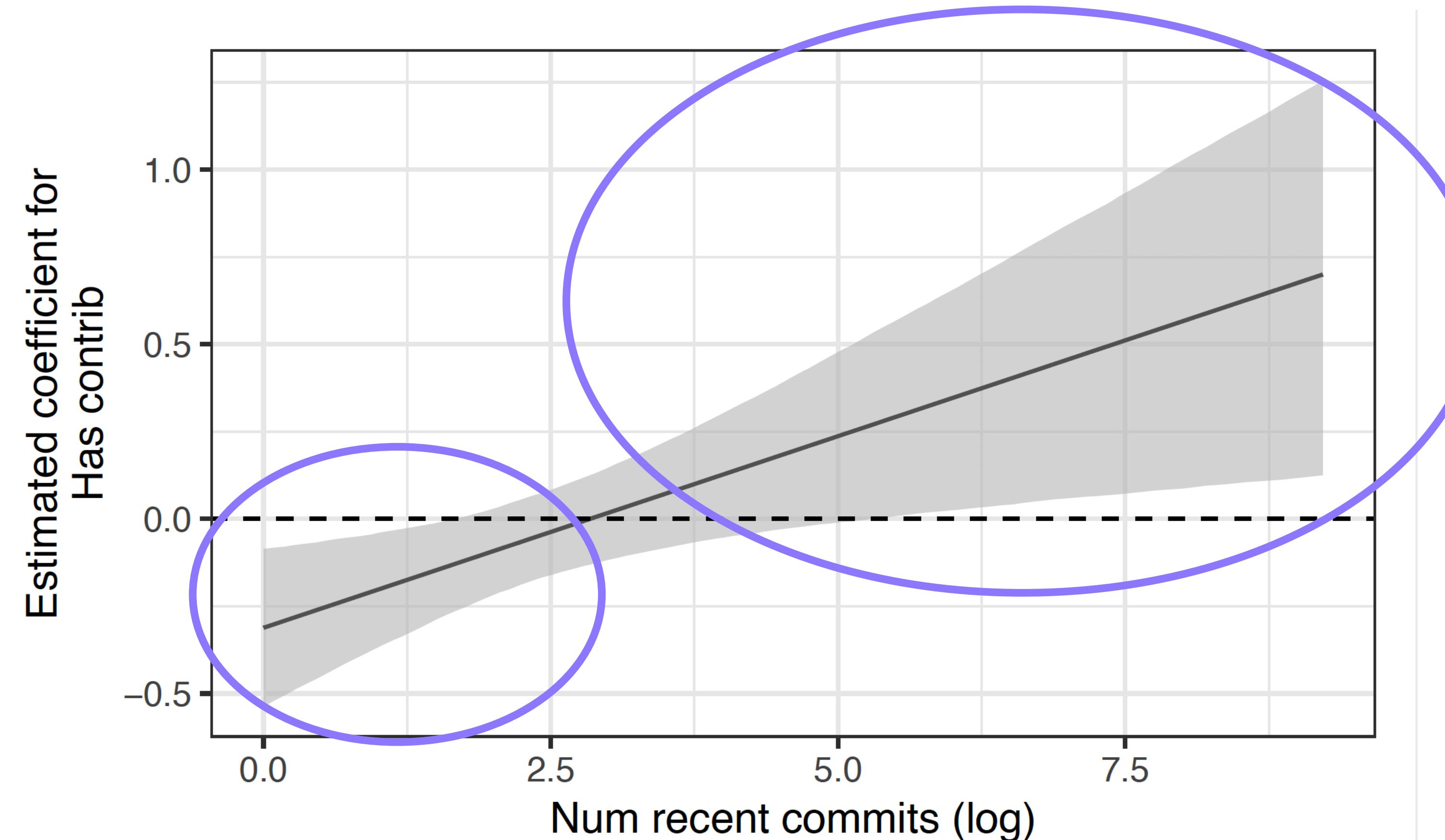
- Please **do not** derail or troll issues. Keep the discussion on topic and respect the opinions of others.

Logistic regression: The presence of contributing guidelines correlates with lower likelihood of attracting newcomers

📄 .test.eslintignore	t (commit): add node_modules
📄 .test.eslintrc.js	ch (commit): configure
📄 .travis.yml	Removed travis node version
📄 AUTHORS	ch (commit): Fix npm use
📄 CONTRIBUTING.md	c (commit): use yarn start to l
📄 LICENSE	d (commit): include 2018 in
📄 README.md	c (commit): remove greenkeeper



Logistic regression: Contributing Guidelines and amount of project activity interact: help vs overhead



Three main classes of signals

Quality

Community

Popularity



Interviews:

The use of language signals the community's openness



qipeng commented on Mar 7

Collaborator + 😊 ...

Hi @g3rm4n0x , could you share a minimal code snippet with the text input that produces this error, so we can more easily diagnose it? Thanks in advance!



Contributing

1. Search the issues, if it is not already there, add one.
2. Fork the repository
3. Code Code Code
4. Submit a pull request
5. Wait for some nice guy to review and merge

Logistic regression: Measuring impoliteness in conversations

lol dude, that's not how this works.

@dolphin and I have clearly explained in [#1809](#) why we won't add this to Compose. Opening a new issue won't change that.

 13





Logistic regression: Impolite language correlates with lower likelihood of attracting new contributors



!Only for experienced contributors!

First-time GitHub
contributors:





Interviews:

Having a template signals that “there’s a good structure for contributing to [this project]”

```
## I'm submitting a...
```

Please search GitHub for a similar issue or PR before submitting

```
<pre><code>
```

Regression (a behavior that used to work and stopped working in a new release)

Bug report

Feature request

Documentation issue or request

(<https://github.com/twbs/bootstrap/blob/master/>)

Bug reports must include:

- Operating system and version (Windows, macOS)
- Browser and version (Chrome, Firefox, Safari, IE)
- [Reduced test case] (<https://css-tricks.com/reduced-test-cases/>)
- [CodePen] (<https://codepen.io/>) or [JS Bin] (<https://jsbin.com/>)

Logistic regression: Having a template correlates with higher likelihood of attracting new contributors

Branch: master ▾ [Logistic regression](#) / .github / ISSUE_TEMPLATE /

 c'haosaur and Ximilicur Fix issue template links to the contributing guidelines

..

bug.md	Fix issue template links to the contributing
bug_report.md	Fix issue template links to the contributing
feature.md	Fix issue template links to the contributing
feature_request.md	Fix issue template links to the contributing



!Only for experienced contributors!

First-time GitHub
contributors:

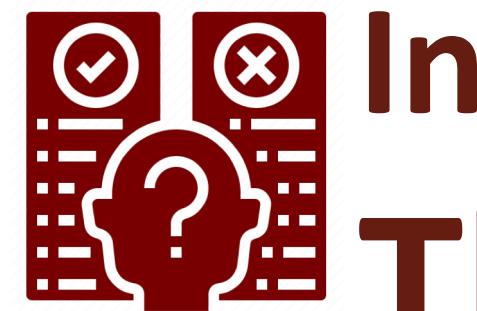


Three main classes of signals

Quality

Community

Popularity



Interviews:

The number of stars signals popularity

A screenshot of a GitHub repository page. At the top, there is a navigation bar with a search bar, pull requests, issues, marketplace, and explore links. Below the navigation bar, there is a large button with a star icon and the word "Star" followed by the number "17". To the right of this button, there is a "Fork" button with the number "6". Below the main header, there are sections for code, issues (28), pull requests (2), actions, projects, and releases. A summary bar at the bottom shows 2,700 commits, 95 branches, 2,093 releases, 9 contributors, and an MIT license. There are buttons for creating a new pull request, cloning the repository, and viewing the latest commit.



Interviews:

Three different opinions

“Everyone uses [project X]. If you contribute to it your change is gonna have a **huge impact.**”

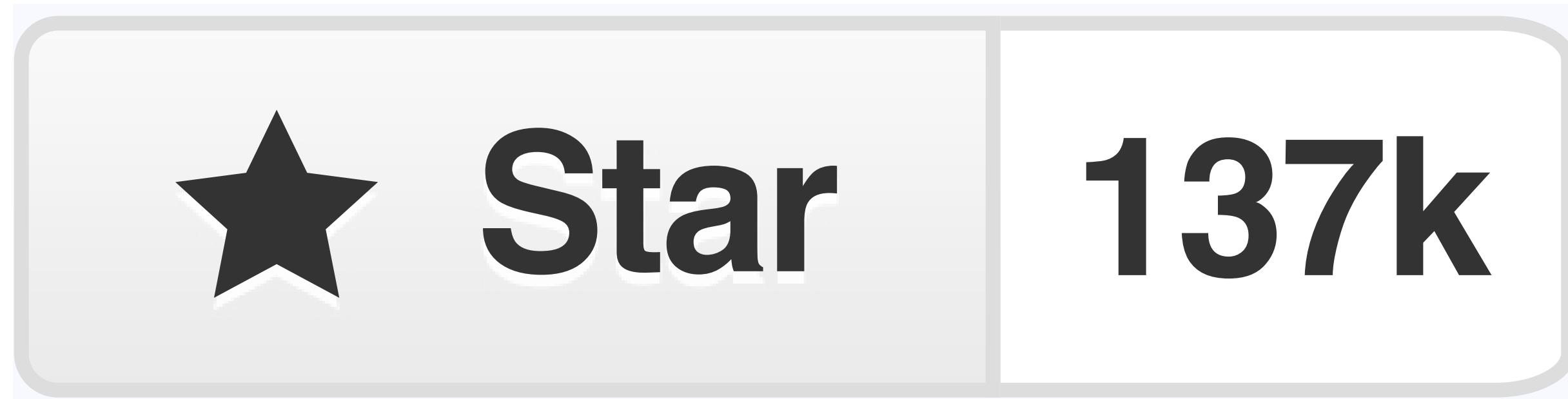


“If it has tons of stars, and there weren’t that many contributors, I would think they weren’t **friendly to new people**”.

“I wouldn’t worry about popularity, because a less popular project gives you **more self-efficacy** that forces you really to look at things, ... try everything out.”



Logistic regression: The number of stars correlates strongly with higher likelihood of attracting new contributors





Interviews:

Recency of commits signals that the repo is active

The screenshot shows a GitHub repository page with the following details:

- Commits:** 667 (circled in red)
- Branches:** 2
- Releases:** 20
- Contributors:** 2
- Licence:** MIT

Actions available on the page include:

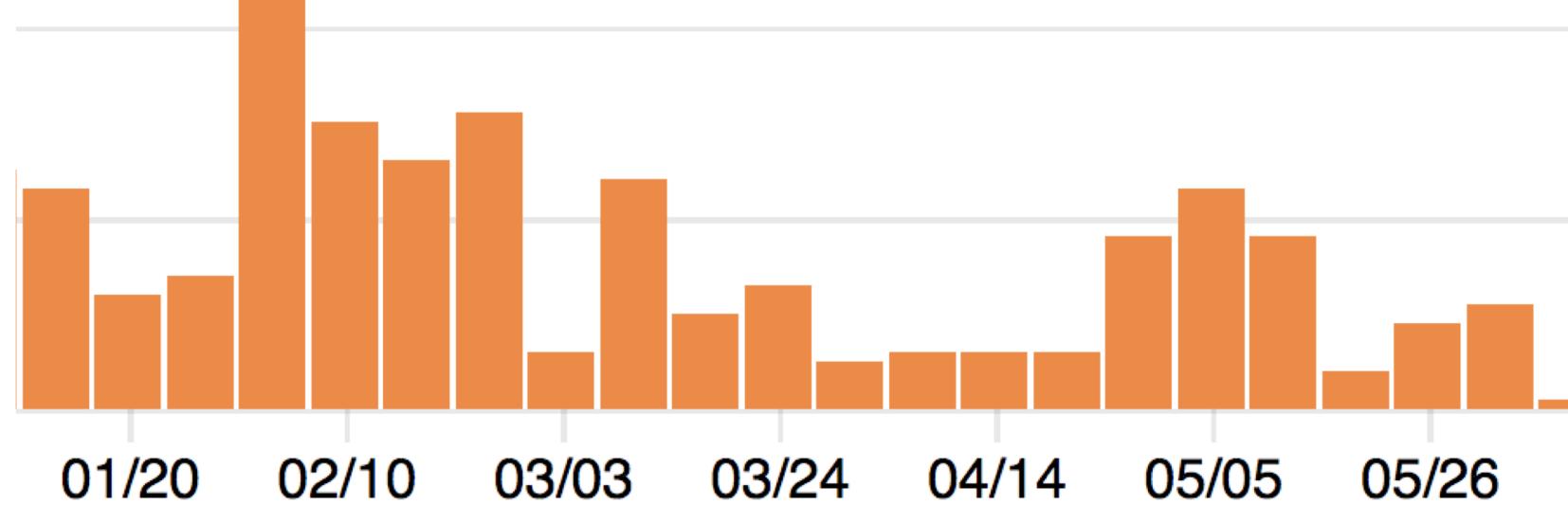
- Branch: master ▾
- New pull request
- Create new file
- Upload files
- Find file
- Clone or download ▾

A recent commit is listed at the bottom:

- Update angular deps
- Latest commit e on Jun 18



Logistic regression: More recent commits correlates with higher likelihood of attracting new contributors



!Only for experienced contributors!

First-time GitHub
contributors:



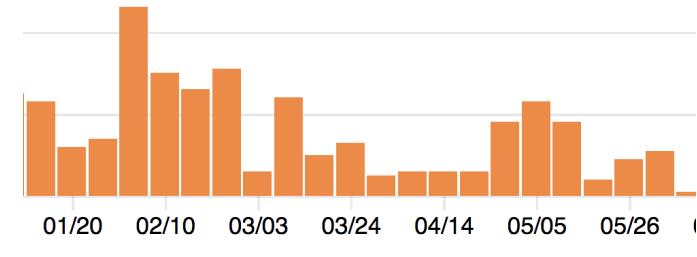
Result highlights: Many signals used when choosing projects

		Interviews	Logistic Regression
Quality	README		
	Contributing guideline		
Community	Impolite language		
	Issue template		
Popularity	Stars		
	Recent commits		

What can we do?

New signals (maybe in the form of badges)

Logistic regression:
More recent commits correlates with higher likelihood of attracting new contributors



!Only for experienced contributors!

First-time GitHub contributors:



maintained no! (as of 2018)

commit activity 457/month

issue resolution

3 h

updated today

Logistic regression:
Impolite language correlates with lower likelihood of attracting new contributors



!Only for experienced contributors!

First-time GitHub contributors:



commitizen friendly

PRs welcome

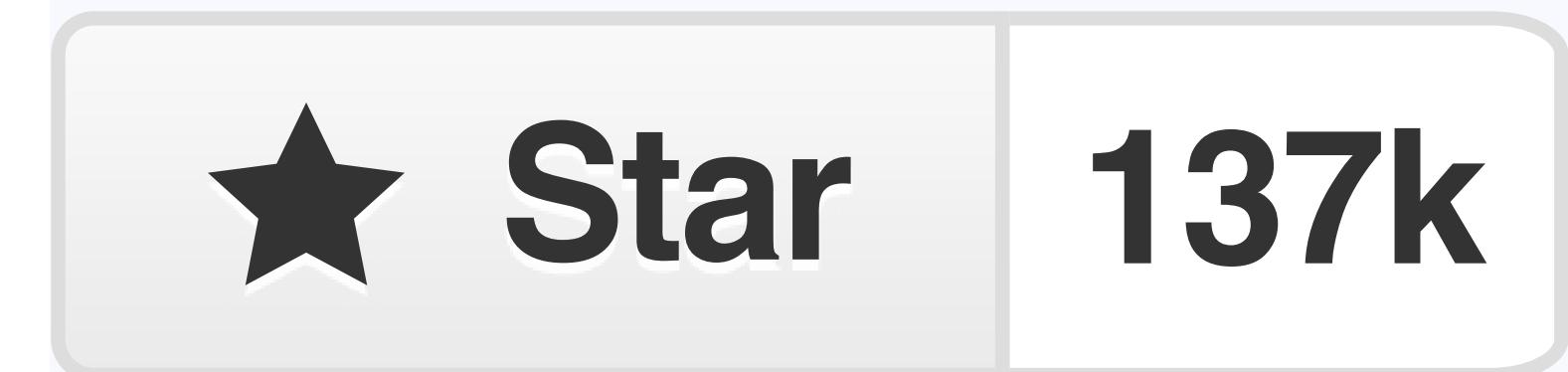
help-wanted

Personalized project recommendation

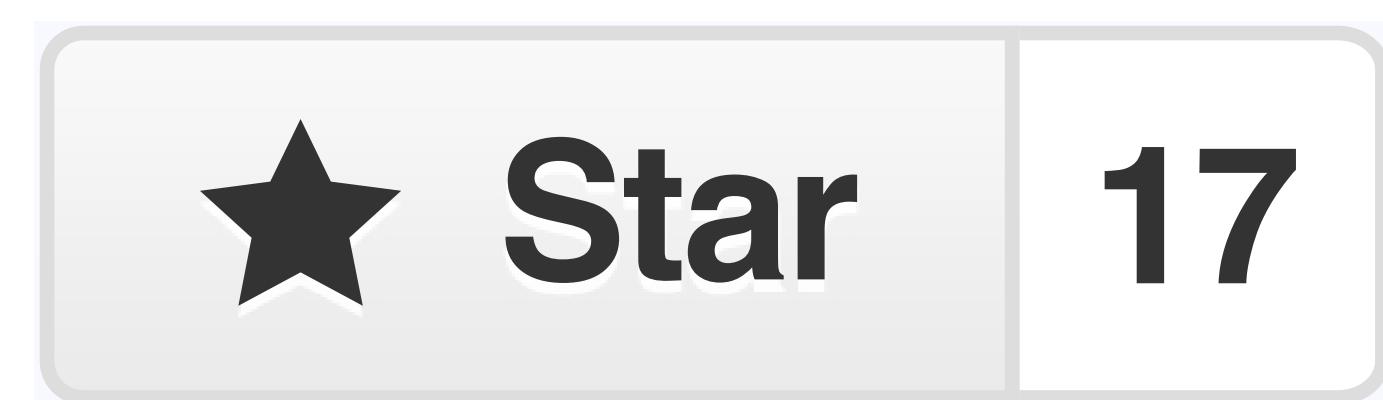
“Everyone uses [project X]. If you contribute to it your change is gonna have **a huge impact.**”

“a less popular project gives you **more self-efficacy** that forces you really to look at things, ... try everything out.”

POPULARITY



vs



SIZE



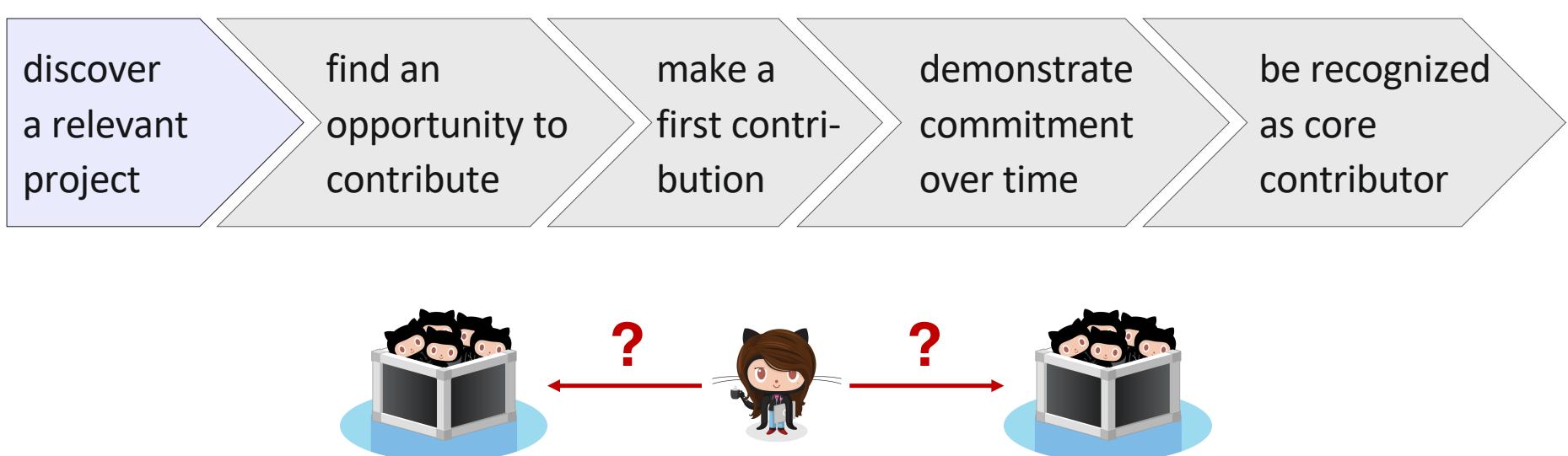
vs



Many signals used when choosing projects

Yet, little is known about how people choose which project to contribute to

Especially for people who contribute for fun, to learn, etc.



Carnegie Mellon University
School of Computer Science STRIDE L

@sophiehsqq

5

Result highlights:
Many signals used when choosing projects

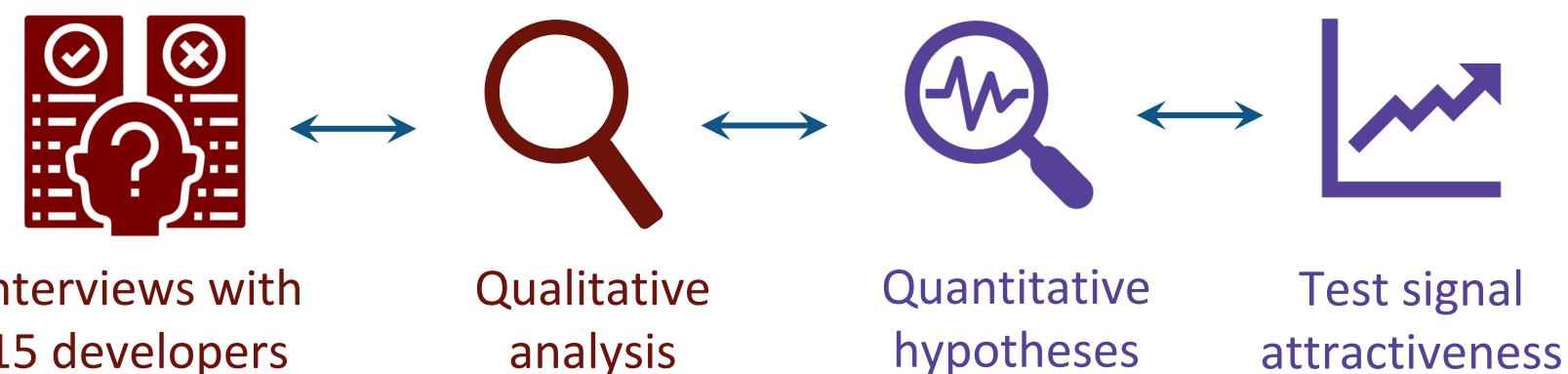
	Interviews	Logistic Regression
Quality	README	
	Contributing guideline	
Community	Impolite language	
	Issue template	
Popularity	Stars	
	Recent commits	

Carnegie Mellon University
School of Computer Science STRIDE L

@sophiehsqq

49

A mixed-method study design

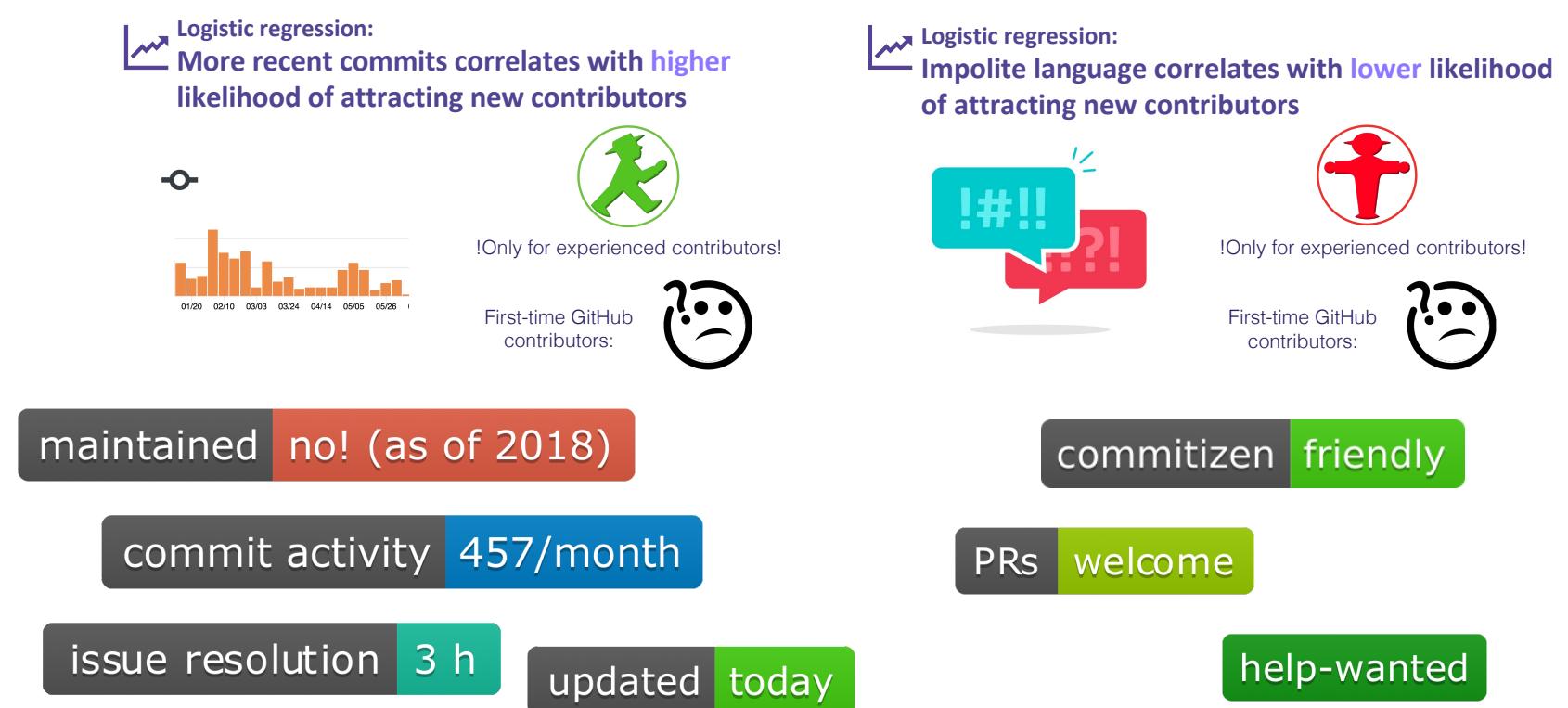


Carnegie Mellon University
School of Computer Science STRIDE L

@sophiehsqq

20

New signals (maybe in the form of badges)



Carnegie Mellon University
School of Computer Science STRIDE L

@sophiehsqq

54

Acknowledgements:



Contact: Sophie Qiu
@sophiehsqq
hsqq@cmu.edu

@asarma
Anita.Sarma@oregonstate.edu

@b_vasilescu
vasilescu@cmu.edu

Code and data:

DOI 10.5281/zenodo.3371186

@sophiehsqq

55