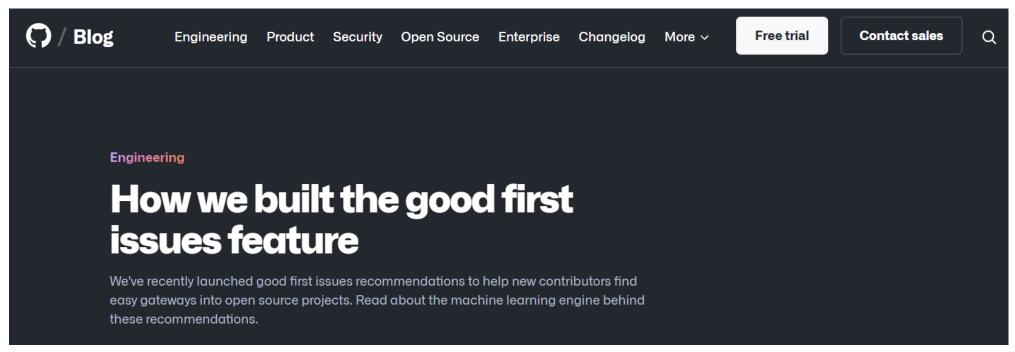
# Good First Issues

Thanks to 何昊



#### What is a Good First Issue?

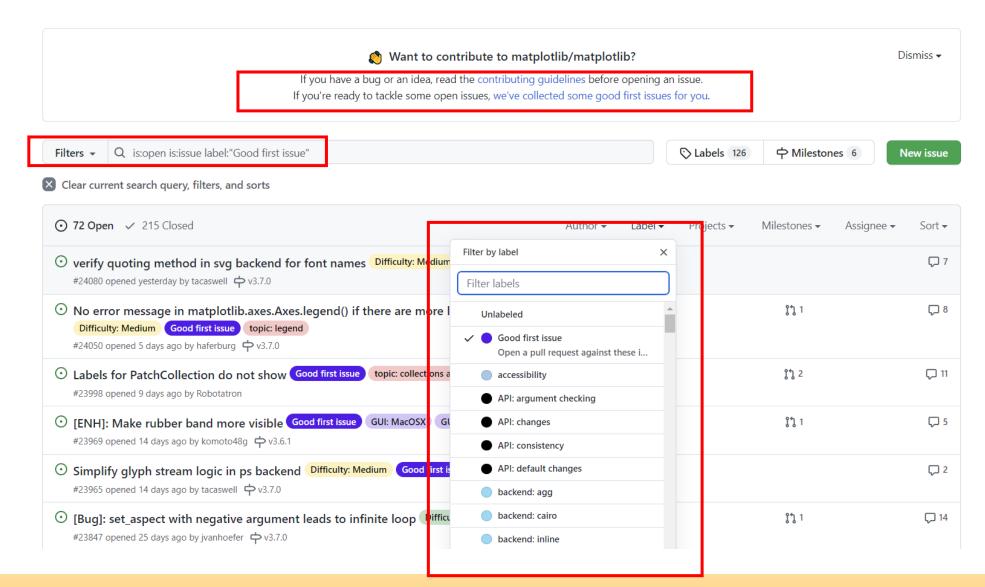
- Good First Issue is an initiative to curate easy pickings from popular projects, so developers who've never contributed to open-source can get started quickly
- Win-win: projects need fresh blood ⇔ newcomers need a starting point



https://github.blog/2020-01-22-how-we-built-good-first-issues/



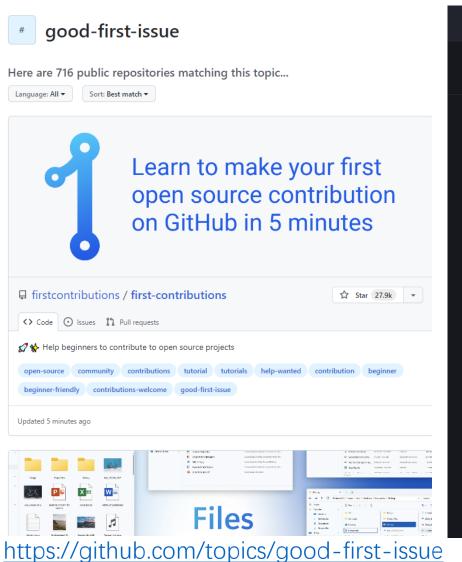
## How to Find Good First Issues in a GitHub Project?

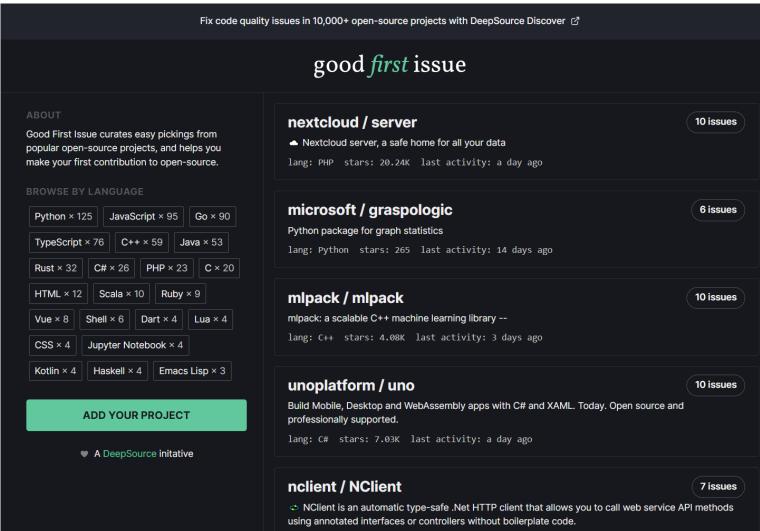


- Other Labels:
  - Help Wanted
  - Up for Grabs
  - ...



## How to Explore Good First Issues over GitHub?





https://goodfirstissue.dev/

《开源软件技术》课程

#### Notes about Good First Issues



#### • Pay attention to the nuances of different labels (checking project documentation!)

#### Help Wanted 👄

Items marked with the help wanted label need to ensure that they meet these criteria:

- Low Barrier to Entry It should be easy for new contributors. Documentation on how that type of change should be made should already exist.
- **Clear Task** The task is agreed upon and does not require further discussions in the community. Call out if that area of code is untested and requires new fixtures.

API and CLI behavior should be decided and included in the OP issue, for example: "The new command syntax is svcat unbind NAME [--orphan] [--timeout 5m] ", with expected validations called out.

- **Goldilocks priority** The priority should not be so high that a core contributor should do it, but not too low that it isn't useful enough for a core contributor to spend time reviewing it, answering questions, helping get it into a release, etc.
- **Up-To-Date** Often these issues become obsolete and have already been completed, are no longer desired, no longer make sense, or have changed priority or difficulty.

A good example of a Help Wanted issue description can be found here: kubernetes/test-infra#21356 (comment).

These commands can be used with GitHub issues to manage the help wanted label:

- /help : Adds the help wanted label to an issue.
- /remove-help : Removes the help wanted label from an issue. If the good first issue label is present, it is removed as well.

#### Good First Issue

Items marked with the <code>good first issue</code> label are intended for *first-time contributors*. It indicates that members will keep an eye out for these pull requests and shepherd it through our processes.

New contributors should not be left to find an approver, ping for reviews, decipher prow commands, or identify that their build failed due to a flake. It is important to make new contributors feel welcome and valued. We should assure them that they will have an extra level of help with their first contribution.

After a contributor has successfully completed one or two <code>good first issue</code> items, they should be ready to move on to <code>help wanted</code> items.

All good first issue items need to follow the guidelines for help wanted items in addition to meeting the following criteria:

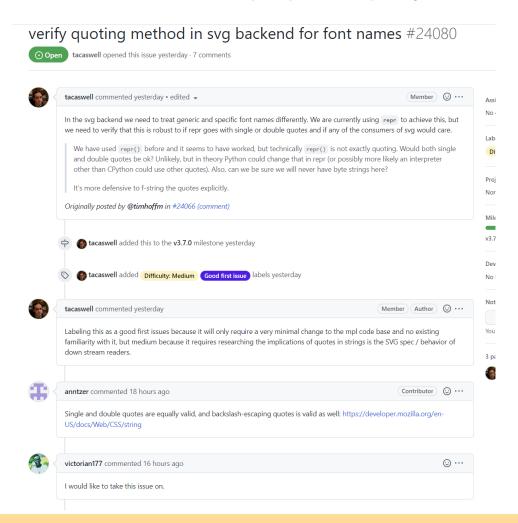
- **No Barrier to Entry** The task is something that a new contributor can tackle without advanced setup or domain knowledge.
- Solution Explained The recommended solution is clearly described in the issue.
- Provides Context If background knowledge is required, this should be explicitly
  mentioned and a list of suggested readings included.
- Gives Examples Link to examples of similar implementations so new contributors have a reference guide for their changes.
- Identifies Relevant Code The relevant code and tests to be changed should be linked in the issue.
- Ready to Test There should be existing tests that can be modified, or existing test
  cases fit to be copied. If the area of code doesn't have tests, before labeling the issue,
  add a test fixture. This prep often makes a great help wanted task!

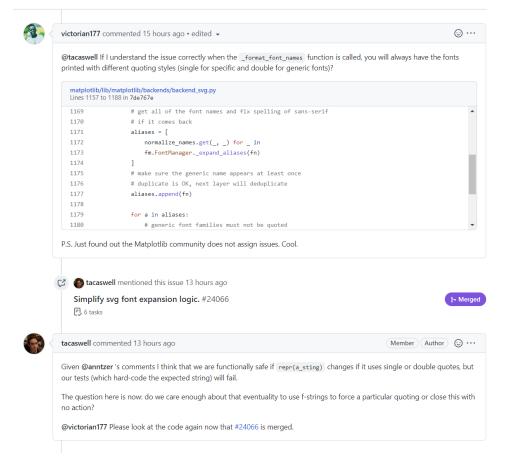
https://www.kubernetes.dev/docs/guide/help-wanted/





Quick takeovers in popular projects





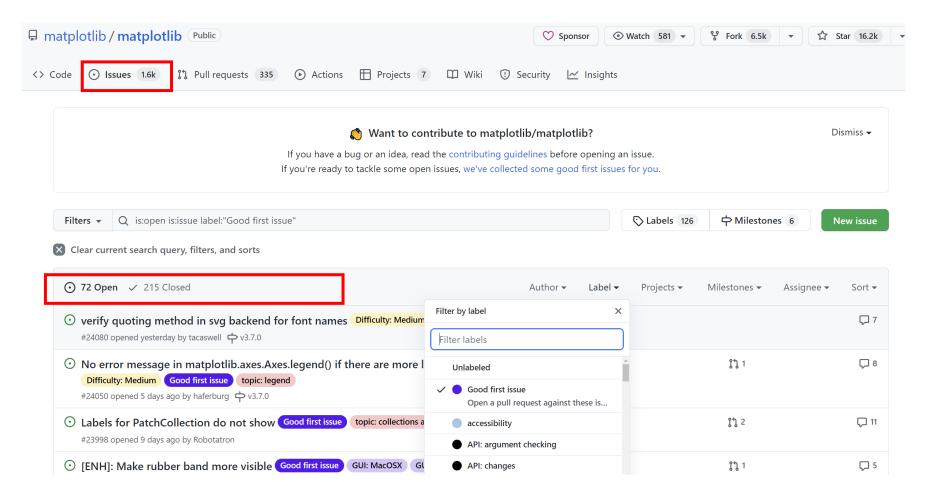
https://github.com/matplotlib/matplotlib/issues/24080

《开源软件技术》课程



#### Notes about Good First Issues

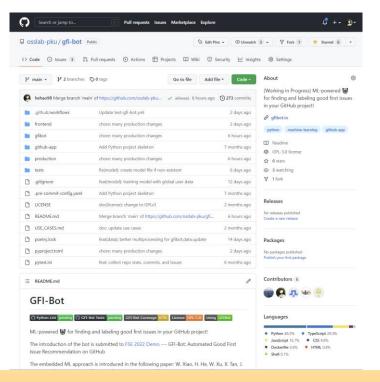
Generally, scarce



#### Some Research about Good First Issues...



- W Xiao, J Li, H He, R Qiu, M Zhou. Personalized First Issue Recommender for Newcomers in Open Source Projects. ASE 2023.
- Hao He, Haonan Su, Wenxin Xiao, Runzhi He, and Minghui Zhou. 2022. GFI-Bot: Automated Good First Issue Recommendation on GitHub. In Proceedings of the 2022 ACM 30th Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/FSE 2022, Singapore, November 14-16, 2022. ACM. <a href="https://hehao98.github.io/files/2022-gfibot.pdf">https://hehao98.github.io/files/2022-gfibot.pdf</a>
- Wenxin Xiao, Hao He, Weiwei Xu, Xin Tan, Jinhao Dong, and Minghui Zhou. 2022. Recommending Good First Issues in GitHub OSS Projects. In Proceedings of the 44th International Conference on Software Engineering, ICSE 2022, Pittsburgh, PA, USA, May 21–29, 2022.
   ACM. https://hehao98.github.io/files/2022-recgfi.pdf
- X Tan, M Zhou, Z Sun. A first look at good first issues on GitHub. FSE 2020.





https://github.com/osslab-pku/gfi-bot https://gfibot.io/

# Add documentation for RESTful APIs #20 Open hehao98 opened this issue 3 days ago · 1 comment hehao98 commented 3 days ago Currently gfibot.data.backend does not have any documentation. It will be helpful to have documentation describing the behavior of each exposed RESTful API. hehao98 added documentation help wanted labels 3 days ago hehao98 commented yesterday — with GFI-Bot Member Author : ... [GFI-Bot] Predicted as Good First Issue with probability 63.27%.

《开源软件技术》课程



## How to identify appropriate GFIs?

# Extraction of Key Information in GFI Descriptions.

Compare the description of GFIs solved by newcomers and unsolved by newcomers.

Table 4: Comparison of keywords in solved GFIs and unsolved GFIs. (■: Solved ■: Unsolved)

Keywords	Comparison	Keywords	Comparison
version	107	map	37
documentation / doc	89	platform	■ 36
return	49	subsystem	■ 34

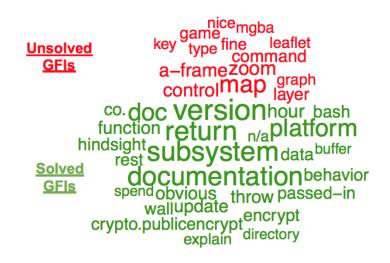


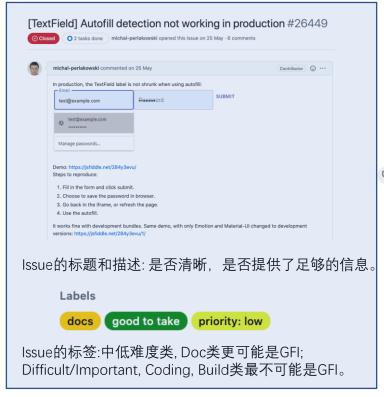
Figure 10: Comparison of solved GFIs and unsolved GFIs. The top 50 keywords with the highest frequency are shown.

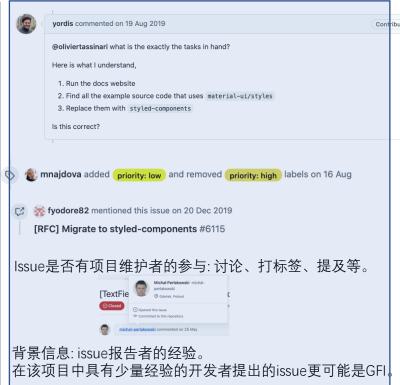
Important information in descriptions:

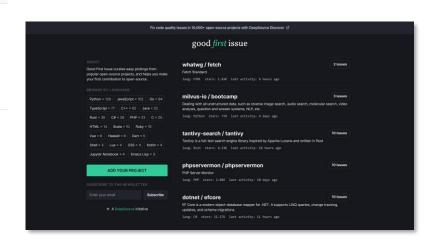
- Detailed environment of the version, subsystem, and platform where the issue occurs
- The actual return of the issue.
- The expected return of the issue.
- Documentations.

# 自动化新手任务推荐: GFI-Bot

AI-Powered智能推荐: 为开源项目issue自动打标签(即判断其是否"新手"任务)







https://github.com/osslab-pku/gfi-bot

**模型**(f): 回归模型(XGBoost等): p=f(x) 训练集: 历史issue,令被新人解决的issue的p=1,其余issue的p=0

测试点: 任一未被分配的open issue

输出(p): 一个issue是GFI的概率



## 探索开源贡献的过程: 以我自己为例 - 何昊

- 我最近主要在深度使用什么样的开源项目,而且对改进他们很有兴趣?
  - Python数据分析与机器学习相关: pandas, scikit-learn, matplotlib, seaborn...
  - GitHub爬虫: PyGitHub
- 我有没有自己遇到一些想要去改进的地方?
  - PyGitHub可太多了
  - 其他这些好像没有我自己会遇到的使用问题(非常成熟)
- 有没有good first issue? 活不活跃?
  - PyGitHub好像不太活跃。。。其他的都很活跃
  - pandas, scikit-learn, matplotlib有good first issue
- 能不能找到我想贡献的good first issue?
  - 简单的good first issue被抢占的也太快了
  - Matplotlib倒是会给一些medium难度的issue标记为good first issue,感觉可以试试!
- 开始了解Matplotlib的贡献流程与开发环境...

# 探索开源贡献的过程: 寻找贡献指南



#### https://matplotlib.org/devdocs/devel/contributing.html

#### Do I really have something to contribute to Matplotlib?

100% yes. There are so many ways to contribute to our community.

When in doubt, we recommend going together! Get connected with our community of active contributors, many of whom felt just like you when they started out and are happy to welcome you and support you as you get to know how we work, and where things are. Take a look at the next sections to learn more.

#### Contributor incubator

The incubator is our non-public communication channel for new contributors. It is a private gitter room moderated by core Matplotlib developers where you can get guidance and support for your first few PRs. It's a place you can ask questions about anything: how to use git, github, how our PR review process works, technical questions about the code, what makes for good documentation or a blog post, how to get involved in community work, or get "pre-review" on your PR.

To join, please go to our public gitter community channel, and ask to be added to '#incubator'. One of our core developers will see your message and will add you.

#### New Contributors meeting

Once a month, we host a meeting to discuss topics that interest new contributors. Anyone can attend, present, or sit in and listen to the call. Among our attendees are fellow new contributors, as well as maintainers, and veteran contributors, who are keen to support onboarding of new folks and share their experience. You can find our community calendar link at the Scientific Python website, and you can browse previous meeting notes on github. We recommend joining the meeting to clarify any doubts, or lingering questions you might have, and to get to know a few of the people behind the GitHub handles ③. You can reach out to @noatamir on gitter for any clarifications or suggestions. We <3 feedback!

#### Issues for new contributors

While any contributions are welcome, we have marked some issues as particularly suited for new contributors by the label good first issue. These are well documented issues, that do not require a deep understanding of the internals of Matplotlib. The issues may additionally be tagged with a difficulty. Difficulty: Easy is suited for people with little Python experience. Difficulty: Medium and Difficulty: Hard require more programming experience. This could be for a variety of reasons, among them, though not necessarily all at the same time:

- The issue is in areas of the code base which have more interdependencies, or legacy code.
- It has less clearly defined tasks, which require some independent exploration, making suggestions, or follow-up discussions to clarify a good path to resolve the issue.
- It involves Python features such as decorators and context managers, which have subtleties due to our implementation decisions.

In general, the Matplotlib project does not assign issues. Issues are "assigned" or "claimed" by opening a PR; there is no other assignment mechanism. If you have opened such a PR, please comment on the issue thread to avoid duplication of work. Please check if there is an existing PR for the issue you are addressing. If there is, try to work with the author by submitting reviews of their code or commenting on the PR rather than opening a new PR; duplicate PRs are subject to being closed. However, if the existing PR is an outline, unlikely to work, or stalled, and the original author is unresponsive, feel free to open a new PR referencing the old one.



## 探索开源贡献的过程: 配置开发环境

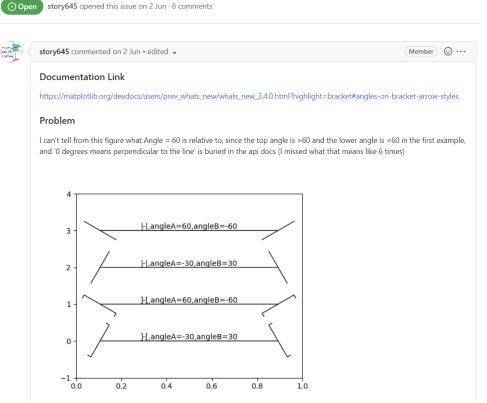
https://matplotlib.org/devdocs/devel/development\_setup.html

```
(base) hehao@DESKTOP-3IVKUR7:/mnt/d/GitHub/matplotlib$ conda env create -f environment.yml
Collecting package metadata (repodata.json): done
Solving environment: done
 ==> WARNING: A newer version of conda exists. <==
  current version: 4.12.0
  latest version: 4.14.0
Please update conda by running
   $ conda update -n base -c defaults conda
Downloading and Extracting Packages
pytest-forked-1.4.0
bzip2-1.0.8
                 9 KB
                  484 KB
toml-0.10.2
                  18 KB
platform linux -- Python 3.10.6, pytest-7.1.3, pluggy-1.0.0
rootdir: /mnt/d/GitHub/matplotlib, configfile: pytest.ini, testpaths: lib
plugins: forked-1.4.0, rerunfailures-10.2, cov-3.0.0, timeout-2.1.0, xdist-2.5.0
collected 9082 items / 1 skipped
lib/matplotlib/tests/test afm.py ......
lib/matplotlib/tests/test agg.py .....
lib/matplotlib/tests/test agg filter.py .s
lib/matplotlib/tests/test animation.py ...ssssssss .... sssssss s.s.....
lib/matplotlib/tests/test api.py .....
lib/matplotlib/tests/test arrow patches.py .ss......
lib/matplotlib/tests/test artist.py ...ss...ss....
                                                                                                                                                                         3%]
5%]
                                                                                                                                                                        10%]
lib/matplotlib/tests/test backend bases.py .....
```



# 探索开源贡献的过程:寻找可以解决的issue

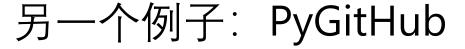
- https://github.com/matplotlib/matplotlib/issues/23176
- https://github.com/matplotlib/matplotlib/issues/23626



[Doc]: Bracket ArrowStyle Angle Unclear #23176

[Bug]: unintended space between comma and digit when using useMathText = New ISSUE True together with comma as decimal separator #23626 Open der-zi opened this issue on 16 Aug · 4 comments der-zi commented on 16 Aug • edited 🕶 ⊕ … No one assigned Bug summary When I turn on "useMathText" in matplotlib in the intention to replace 1e6 by  $\times 10^6$ , the formatting of the commas in the tick labels themselves get wrong. Before they are correct in the form "3,5", with "useMathText" they become "3,5", with (half?) a whitespace between the comma and the following digit. 4.5 Projects 4,0 -Milestone No milestone Development 1,5 2,0 2,5 3,0 No branches or pull requests Notification Customize ∆ Subscribe 4,0 -You're not receiving notifications from this thread. 3 participants # 6 0,5 1,0 1,5 2,0 2,5 3,0 Code for reproduction

14





 $\Box$  7

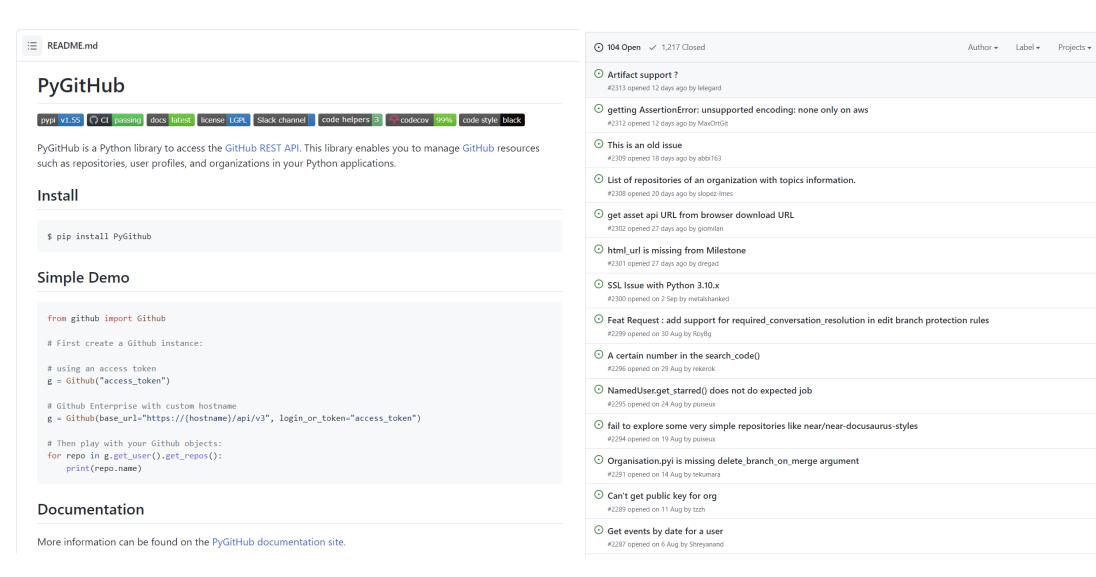
 $\Box$  1

□ 2

 $\Box$  1

 $\Box$  1

Milestones ▼ Assignee ▼



## 一些参考资料



- https://opensource.guide/
  - 关于如何参与开源的各种指南
- <a href="https://github.blog/2020-01-22-browse-good-first-issues-to-start-contributing-to-open-source/">https://github.blog/2020-01-22-browse-good-first-issues-to-start-contributing-to-open-source/</a>
  - GitHub官方对good first issue功能的介绍
- https://github.com/topics/good-first-issue
  - GitHub提供的good first issue主题页面
- https://goodfirstissue.dev/
  - 一个列出有good first issue的GitHub项目的网站
- https://gfibot.io/
  - 对一些GitHub项目推荐good first issue (来自我们实验室)

## 课堂练习



- Current Research Topics on newcomer onboarding
  - What are really good "good first issues"?
  - What good first issues are good for retaining newcomers?

• 你认为什么样的issue会是适合你解决的?