

# Contributing to pandas

# pandas?

~~팬더가 아닙니다~~

*Flexible and powerful data analysis / manipulation library for Python, providing labeled data structures similar to R data.frame objects, statistical functions, and much more*

– <https://github.com/pandas-dev/pandas>

# 기여한다는 건 무엇일까?

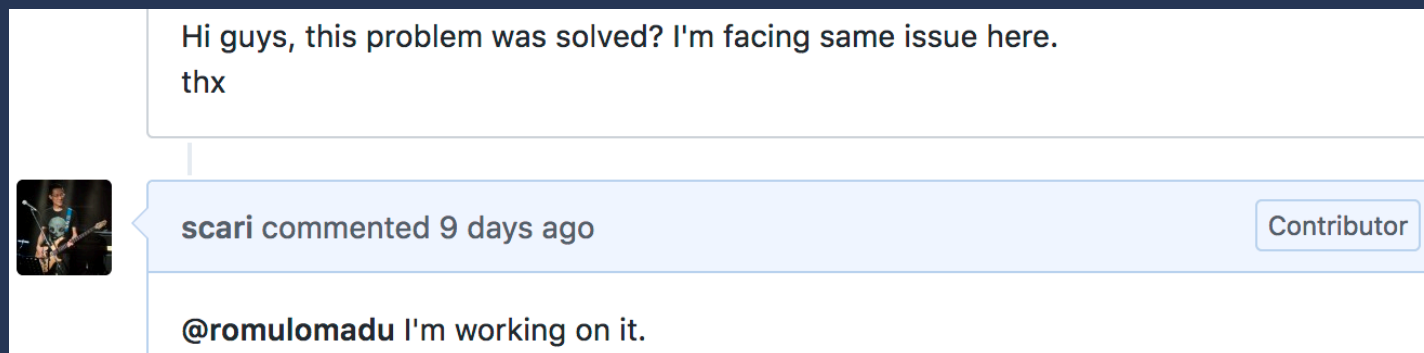
프로젝트에 도움이 되는 "모든" 활동

- 기부
- 도큐먼트
- 이슈 리포트
- 코드 변경

# 왜죠?

기분이 좋잖아요

- 홍익인간
- 애정표현
- Contributor 딱지



# 누가 하나요?

- 17명의 코어 컨트리뷰터
- 전체 컨트리뷰터 수 900명 돌파 (2016.11 500여명)
- ...그리고 당신!

# 시작하기

중요합니다.

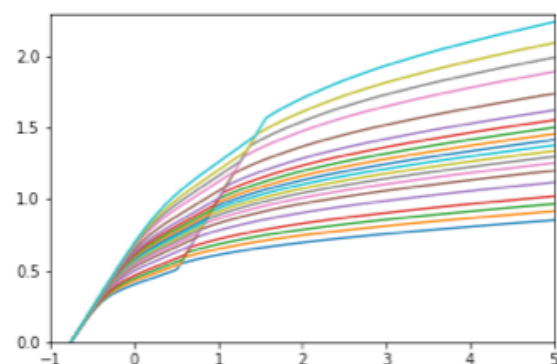
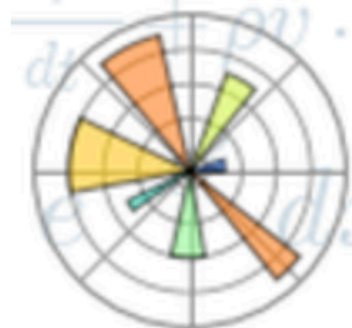
1. 일단, ★ 를 찍어주세요 😊
2. 어떤 식으로 기여할 것인지 정합니다.

# 기부

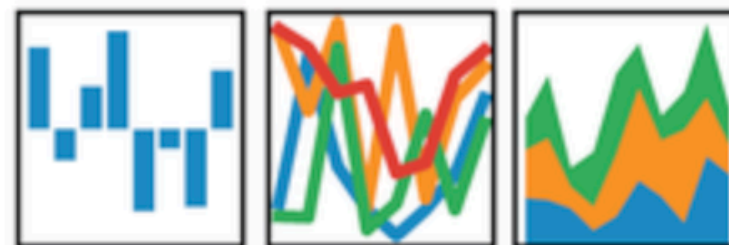
## 가장 쉬운 방법!

- NUMFocus
  - <https://www.numfocus.org/open-source-projects/>

IP[y]:  
IPython

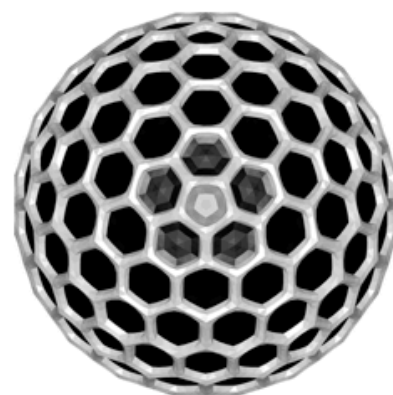


pandas  
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



julia

将軍  
sho gun





# 감사 표시는 돈으로! (**again**)

- donate to pandas
  - [https://www.flipcause.com/widget/give\\_now/MzA3OA==](https://www.flipcause.com/widget/give_now/MzA3OA==)

# 문서화

- <http://pandas.pydata.org/pandas-docs/stable/>
- 문서를 읽다가 잘못된 설명 또는 오타 발견 -> 수정
- DOC : PR 제목에 prefix를 붙여줍니다.
- 꾸르팁 : <https://github.com/lyda/misspell-check>

# 이슈 리포트

- 버그 리포트
- 호환성 문제
- 성능 문제
- 이미 리포트 되었는지 먼저 검색 (중요)

# Last day of month should group with that month #17898



louispotok opened this issue 4 days ago · 2 comments



louispotok commented 4 days ago • edited

Contributor



## Code Sample, a copy-pastable example if possible

```
import pandas as pd
dts = pd.date_range(start='2016-09-01', end='2017-08-31')
df = pd.DataFrame(
    {
        'day': dts,
        "val": pd.np.random.rand(len(dts)),
    }
)
right_cls = df.groupby(pd.Grouper(key='day', freq='4M', closed="right")).size()
left_cls = df.groupby(pd.Grouper(key='day', freq='4M', closed="left")).size()

print(right_cls.index.max())
# pd.Timestamp('2017-09-30 00:00:00')
print(left_cls.index.max())
# pd.Timestamp('2017-12-31 00:00:00')
```

## Problem description

I don't know whether this is a bug or whether I don't understand the expected behavior. I would expect the max index value in both cases to be "2017-08-31", but it looks like it's considering 08-31-2017 to be in September. If I have an observation on 2017-08-31 before midnight, it should still be grouped into the period ending 2017-08-31, right?

## Expected Output

```
# This should work
assert right_cls.index.max() == pd.Timestamp('2017-08-31 00:00:00')
assert left_cls.index.max() == pd.Timestamp('2017-08-31 00:00:00')
```

## Output of `pd.show_versions()`

I'm using pandas 0.20.3 here, but I also checked this on the latest commit and it looks like the behavior persists.

► Details

# 이슈 템플릿

- 재현 가능한 코드 샘플
- 문제 설명
- 기대하는 정상적인 결과
- 사용 환경/버전 명시 `pd.show_versions()`



Write

Preview

AA ▾ B i

“ ” < > ↻

☰ ☰ ☰ ☰ ☰

↶ @ ★

#### Code Sample, a copy-pastable example if possible

```
```python
# Your code here

```
```

#### Problem description

[this should explain **why** the current behaviour is a problem and why the expected output is a better solution.]

**Note:** We receive a lot of issues on our GitHub tracker, so it is very possible that your issue has been posted before. Please check first before submitting so that we do not have to handle and close duplicates!

**Note:** Many problems can be resolved by simply upgrading `pandas` to the latest version. Before submitting, please check if that solution works for you. If possible, you may want to check if `master` addresses this issue, but that is not necessary.

For documentation-related issues, you can check the latest versions of the docs on `master` here:

<https://pandas-docs.github.io/pandas-docs-travis/>

If the issue has not been resolved there, go ahead and file it in the issue tracker.

#### Expected Output

Attach files by dragging & dropping, [selecting them](#), or pasting from the clipboard.

 Styling with Markdown is supported

Submit new issue

# 이슈 말머리

- ENH: 새 기능 또는 개선
- BUG: 버그 수정
- DOC: 문서 업데이트
- TST: 테스트 업데이트
- BLD: 빌드 관련 업데이트
- PERF: 성능 개선
- CLN: 코드 정리

# 코드 변경

- Python, Cython, C
- git과 GitHub, Python에 대한 약간의 지식이 필요
- 내가 뱉은 버그를 수정하고 싶다! -> 이슈 리포트 먼저
- 그냥 코드에 기여하고 싶다! -> 이슈 헌팅



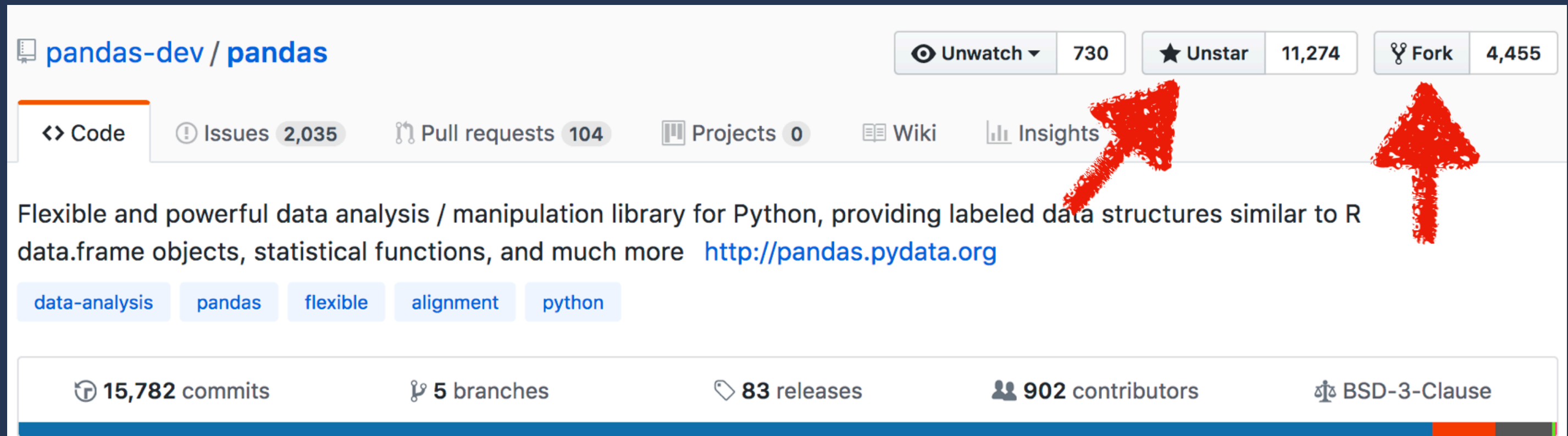
# 이슈 헌팅

- good first issue <- 여기부터 시작
- 난이도
  - Difficulty Novice / Intermediate / Advanced
- 노가다
  - Effort Low / Medium / High
- 짜릿함
  - Performance
- More on <https://github.com/pandas-dev/pandas/labels>

# Fork

1. Star 를 눌렀는지 확인 😊

2. Fork 버튼 클릭



The screenshot shows the GitHub repository page for `pandas-dev / pandas`. At the top right, there are buttons for `Unwatch` (730), `Unstar` (11,274), and `Fork` (4,455). Below these, a navigation bar includes `Code`, `Issues` (2,035), `Pull requests` (104), `Projects` (0), `Wiki`, and `Insights`. The repository description states: "Flexible and powerful data analysis / manipulation library for Python, providing labeled data structures similar to R data.frame objects, statistical functions, and much more <http://pandas.pydata.org>". Below the description are tags: `data-analysis`, `pandas`, `flexible`, `alignment`, and `python`. At the bottom, statistics are listed: 15,782 commits, 5 branches, 83 releases, 902 contributors, and the BSD-3-Clause license. Two red arrows are drawn on the image: one pointing to the `Unstar` button and another pointing to the `Fork` button.

# Clone

내 저장소를 **clone** 후, 공식 저장소를 **upstream** 리모트로 추가

```
git clone git@github.com:your-user-name/pandas.git pandas-yourname  
cd pandas-yourname  
git remote add upstream git://github.com/pandas-dev/pandas.git
```

# 빌드 환경 /w Anaconda

pandas\_dev 라는 이름으로 새 환경을 생성, 의존 패키지를 설치

```
conda create -n pandas_dev python=3 --file ci/requirements_dev.txt
```

주의: **Anaconda3-5.0.0** 버전에서는 개발 환경 설치에 실패합니다.

**4.4.0** 버전을 사용하거나 `-c conda-forge` 옵션을 추가로 지정합니다.

# 빌드

```
$ source activate pandas_dev  
$ python setup.py build_ext --inplace
```

pandas 소스트리 루트에서 Python을 실행하고,

```
import pandas as pd  
pd.show_versions()
```

INSTALLED VERSIONS

```
-----  
commit: 6f03530c5a4e662eb65791730e8b7edd4ea16983  
python: 3.6.2.final.0  
python-bits: 64  
OS: Darwin  
OS-release: 17.0.0
```

# pdb.set\_trace()

- pdb: <https://docs.python.org/3/library/pdb.html>
- Traceback을 보고 수상쩍은 위치에 breakpoint

```
import pdb  
pdb.set_trace()
```

- 범위를 좁혀가며 R&D. Reset and Debug!

# Test

- 기존 테스트를 참조하여 변경 내용 테스트를 추가
  - `./pandas/tests/`
- 이슈에서 재현 가능 코드를 참고
- 테스트 함수 안에 이슈 번호 추가 GH1234

# Linting

- PEP8
- CI에서 스타일 에러가 걸리므로 미리 체크

```
git diff master -u -- '*.py' | flake8 --diff
```



# What's New 추가

- 다음 릴리즈 노트에 포함되도록 이번 변경 사항을 직접 기록  
`doc/source/whatsnew/v0.xx.xx.txt`

# PR 만들기

1. 보내기 전에 commit squash
2. 내 저장소로 먼저 push
3. GitHub 내 저장소 페이지에서 Pull Request
4. PR에 빠진 내용이 없는지 확인
5. 발사!

# 스프린트

해커톤과는 다르다! 해커톤과는!!<sup>1</sup>



<sup>1</sup> <https://www.pycon.kr/2017/program/tutorials/>

# 기타 읽을 거리

- <https://opensource.guide/>
- <https://naver.github.io/OpenSourceGuide/book/index.html>
- <https://github.com/pandas-dev/pandas/blob/master/doc/source/contributing.rst#working-with-the-code>