RepoDash

A Command Line Github Issue Metrics Dashboard

Laurence Molloy

PyData

laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

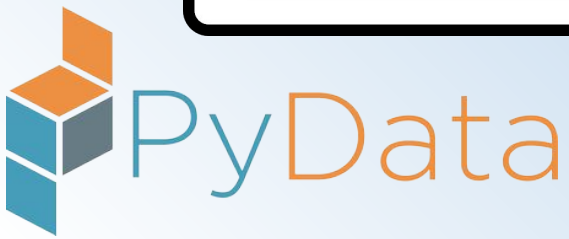
About Me

Previously...

- *Mathematician*
- *Speech Tech Engineer*
- *IT Manager*

Now...

- *Perl / Python Developer*
- *Data Scientist*

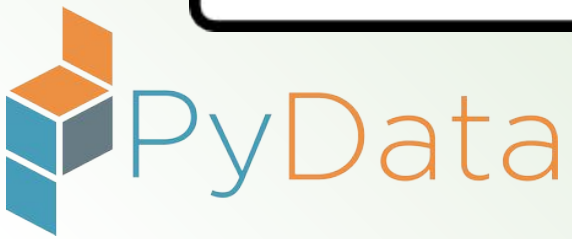


laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

The Historic Need

- IT Manager in a tech startup
- Legacy system support & maintenance
- Fossil source control
- Monthly management reporting needs
- Limited time / resource



laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

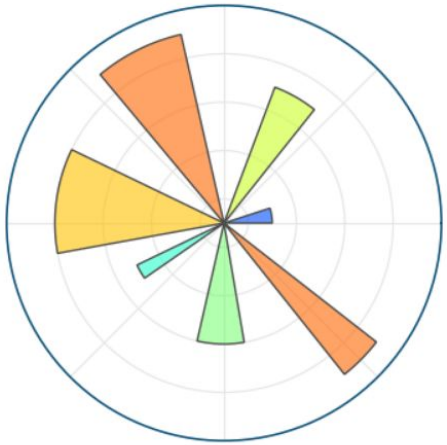
Recent Inspiration

NUMFOCUS
OPEN CODE = BETTER SCIENCE

ABOUT PROJECTS PROGRAMS

Now Hiring: Matplotlib Research Software Engineering Fellow

by Gina Helfrich, Ph.D. (Communications Director, NumFOCUS) | Nov 22, 2019 | Blog



The problem:

- **Backlog of 1200 issues**
- **Backlog of 300 PRs**

Requirements of the post:

- **Classify & prioritise**
- **Help to reduce the backlog**

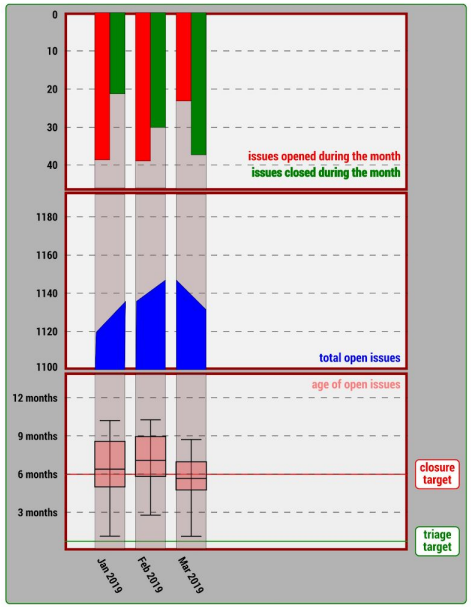
PyData

laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

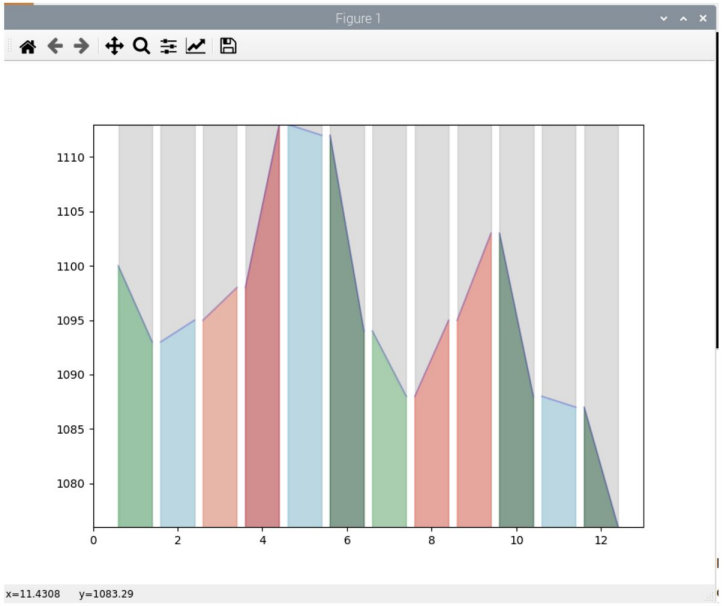
09 January 2020

The Development Journey

Inkscape Wireframe



Early Matplotlib Prototype



PyData

laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

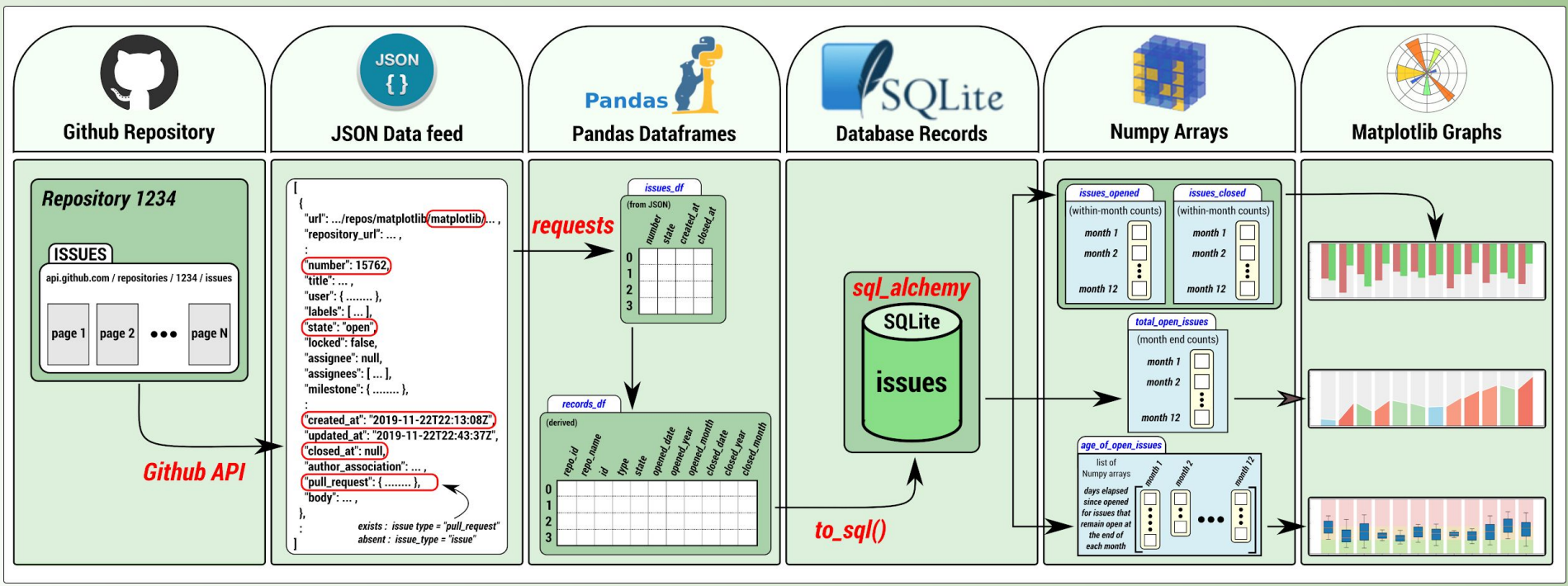
Basic Dashboard Design

PyData

laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

Integration with Github



PyData

laurencemolloy.github.io/RepoDash

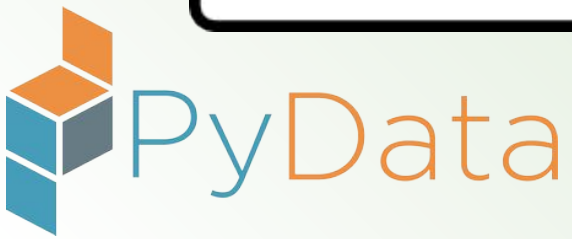
@MolloyLaurence

laurence.molloy@gmail.com

09 January 2020

Enhancements

- Authentication
- Different projects
- Different timeframes
- Command line control
- Documentation
- ‘current’ open issue count
- Issue labelling (triage) & assignment



laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

python3 RepoDash.py OPTIONS

-h, --help	Show this help message and exit
-a, --authtoken TOKEN	Github personal access token
-u, --user USER	Github user name
-r, --repo REPO	Github repo name
-t, --type { *issue* , *pr* }	Issue type, one of ['issue' , 'pr']
-m, --months MONTHS	analysis timespan in months
-d, --refdate REFDATE	reference date ('YYYY-MM')
-f, --firstpage FIRSTPAGE	first page of issues to request
-c, --pagecount PAGECOUNT	number of pages to request
-o, --offsetmonth { *opened* , *closed* }	which month to offset issue closure in
-p, --datapath DATAPATH	path for SQLite db file
-i, --info	write environment info to debug_info.txt file

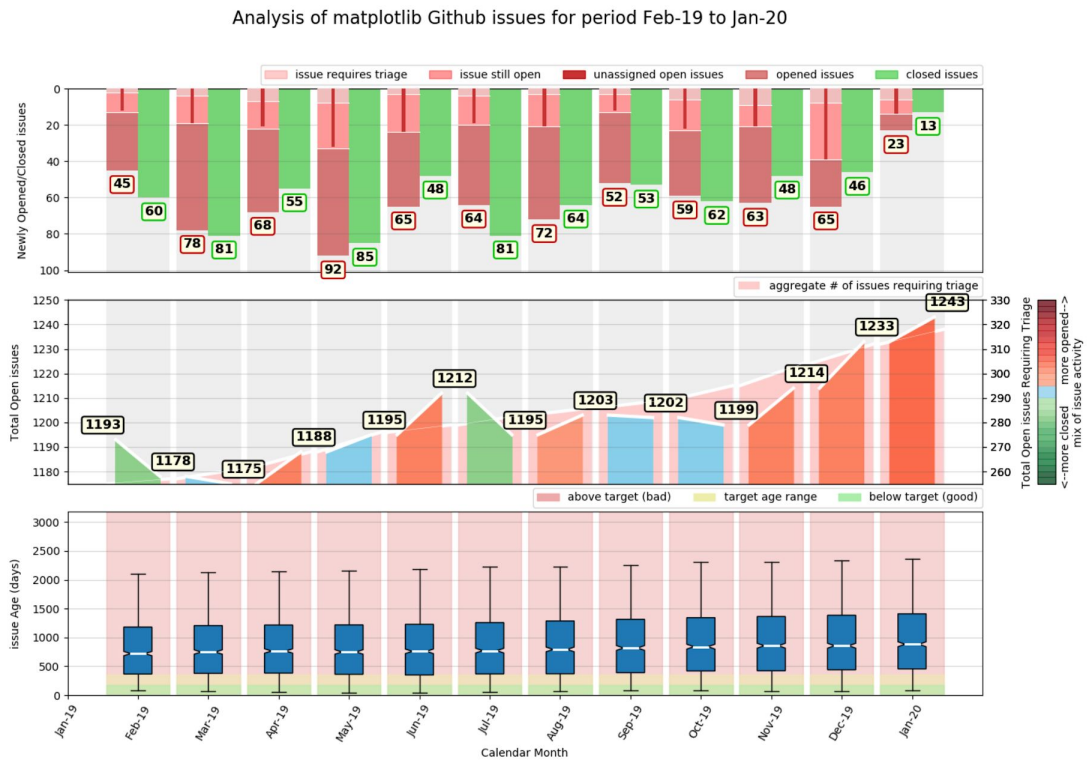
PyData

laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

Enhanced Dashboard Design

RepoDash.py -u matplotlib -r matplotlib -f 1 -c 162 -m 12 -d '2020-01'



laurencemolloy.github.io/RepoDash

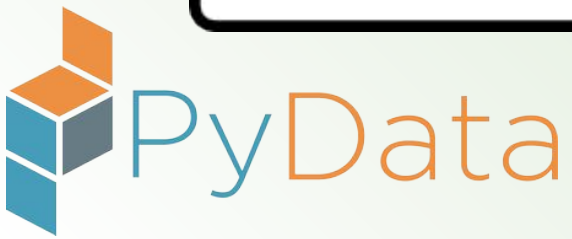
@MolloyLaurence

laurence.molloy@gmail.com

09 January 2020

Limitations

- Efficiency
- Encapsulation of plotting code
- More metrics
- Dashboard configuration / optionality
- Private repos on corporate GitLabs
- Bugs



laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

Performance Testing

Time taken to process all pages of issues

Matplotlib	**162 pages**	**~10 minutes**
Pandas	**389 pages**	**~30 minutes**
Kubernetes	**870 pages**	**~70 minutes**

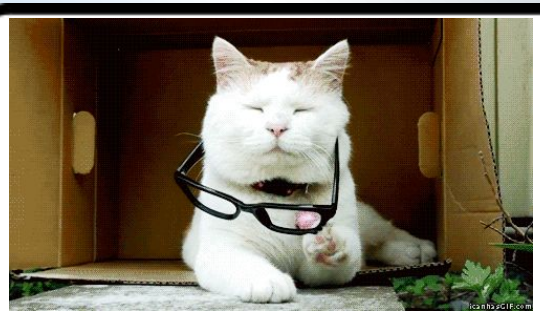
Remembering current progress can save time

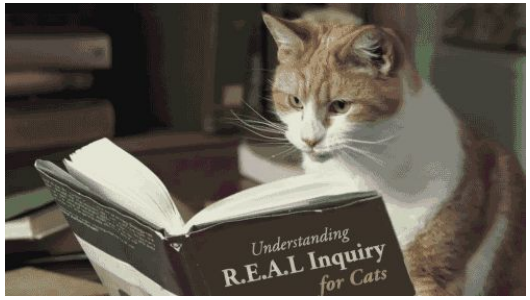
PyData

laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020

Any Questions?





CATch me afterwards
I'm PURRfectly approachable

(paws for applause)

Documentation & Repository

laurencemolloy.github.io/RepoDash

PyData

laurencemolloy.github.io/RepoDash
@MolloyLaurence
laurence.molloy@gmail.com

09 January 2020