

# *Dan Foreman-Mackey*

-  *CCPP@NYU*
-  *dfm*
-  *dfm.io*
-  *@exoplaneteer*



*I don't know anything  
about licensing!*

*but...*

*I do know about  
**data science!***



dfm (Dan Foreman-Mackey) X

GitHub, Inc. [US] <https://github.com/dfm>

Apps Betterize! Betterize 2.0 Recent ArXiv New ArXiv Manual:Database lay 52 Recipes K2 Field Descriptions Betterize (with a bet Licensing

Search GitHub Explore Gist Blog Help  dfm + - ⌂ ⚙ ⌁



# Dan Foreman-Mackey

dfm

Developer Program Member

Brooklyn, NY <http://dan.iel.fm>

Joined on Jul 31, 2010

**405** Followers   **97** Starred   **39** Following

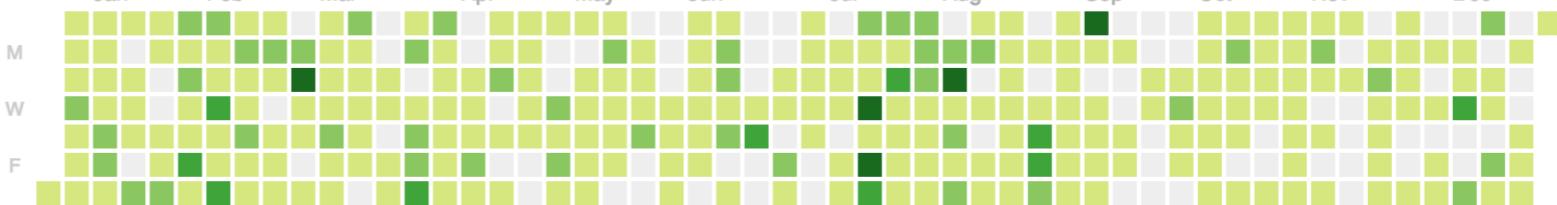
## Popular repositories

 <a href="#">osrc</a>	892 ★
The Open Source Report Card	
 <a href="#">emcee</a>	333 ★
The Python ensemble sampling toolkit for affine...	
 <a href="#">daft</a>	155 ★
Render some probabilistic graphical models u...	
 <a href="#">george</a>	59 ★
Fast Gaussian Processes for regression	
 <a href="#">triangle.py</a>	37 ★
Make a corner plot of samples.	

## Repositories contributed to

 <a href="#">sivaramambika.../HODLR</a>	11 ★
A fast, accurate direct solver and determinant ...	
 <a href="#">RuthAngus/Gyro</a>	0 ★
Calibrating Gyrochronology using Kepler Aster...	
 <a href="#">davidwhogg/smartcard</a>	0 ★
a NASA ADAP proposal to support novel Kepl...	
 <a href="#">benmontet/K2-noise</a>	1 ★
K2 has a lot of correlated noise. Let's find plan...	
 <a href="#">mrtommyb/GP_model_Kepler...</a>	0 ★
modeling a Kepler light curve using a Gaussia...	

## Contributions



Summary of Pull Requests, issues opened, and commits. [Learn more](#).

Less  More 

Year of contributions 1,515 total	Longest streak 27 days	Current streak 4 days
--------------------------------------	---------------------------	--------------------------

GitHub Developer

https://developer.github.com

Apps Betterize! Betterize 2.0 Recent ArXiv New ArXiv Manual:Database lay 52 Recipes K2 Field Descriptions Betterize (with a bet Licensing

# GitHub Developer

API Developers Blog Support Search

## Leverage the power of GitHub in your app.

Get started with one of our guides, or jump straight into the API documentation.

Browse the documentation



Join the GitHub Developer Program.  
The best way to integrate with GitHub. Learn more.



dfm/github-repo-crawler

GitHub, Inc. [US] https://github.com/dfm/github-repo-crawler

Apps Betterize! Betterize 2.0 Recent ArXiv New ArXiv Manual:Database lay 52 Recipes K2 Field Descriptions Betterize (with a bet Licensing

This repository Search Explore Gist Blog Help dfm + - ⚙️ ↗

dfm / **github-repo-crawler** Unwatch 1 Star 0 Fork 1

Crawl all the GitHubz — Edit

8 commits 1 branch 0 releases 1 contributor

branch: master / +

fail gracefully

dfm authored on Aug 7 latest commit 542e780a04

crawler fail gracefully 5 months ago

.gitignore skip forks 6 months ago

LICENSE Initial commit 6 months ago

README.md Initial commit 6 months ago

crawl.py retries 6 months ago

README.md

# github-repo-crawler

Crawl all the GitHubz

Code Issues Pull Requests Wiki Pulse Graphs Settings

HTTPS clone URL <https://github.com/dfm/>

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

Clone in Desktop Download ZIP

GitHub, Inc. [US] https://github.com/dfm/github-repo-crawler

dfm / github-repo-crawler

Crawl all the GitHubz — Edit

8 commits 1 branch 0 releases 1 contributor

branch: master / +

	fail gracefully	dfm authored on Aug 7
	crawler	fail gracefully
	.gitignore	skip forks
	LICENSE	Initial commit
	README.md	Initial commit
	crawl.py	retries

latest commit 542e780a04

README.md

# github-repo-crawler

Crawl all the GitHubz

Unwatch 1 Star 0 Fork 1

Code Issues Pull Requests Wiki Pulse Graphs Settings

HTTPS clone URL https://github.com/dfm/ You can clone with HTTPS, SSH, or Subversion.

Clone in Desktop Download ZIP

**28,569,100**  
**largest public repo ID**

(as of 2014-12-28)

**5,000**  
requests/hour rate limit

a few  
requests/repo required

**1,578,967**  
repositories

("randomly" selected)

**6%**  
of the GitHubz

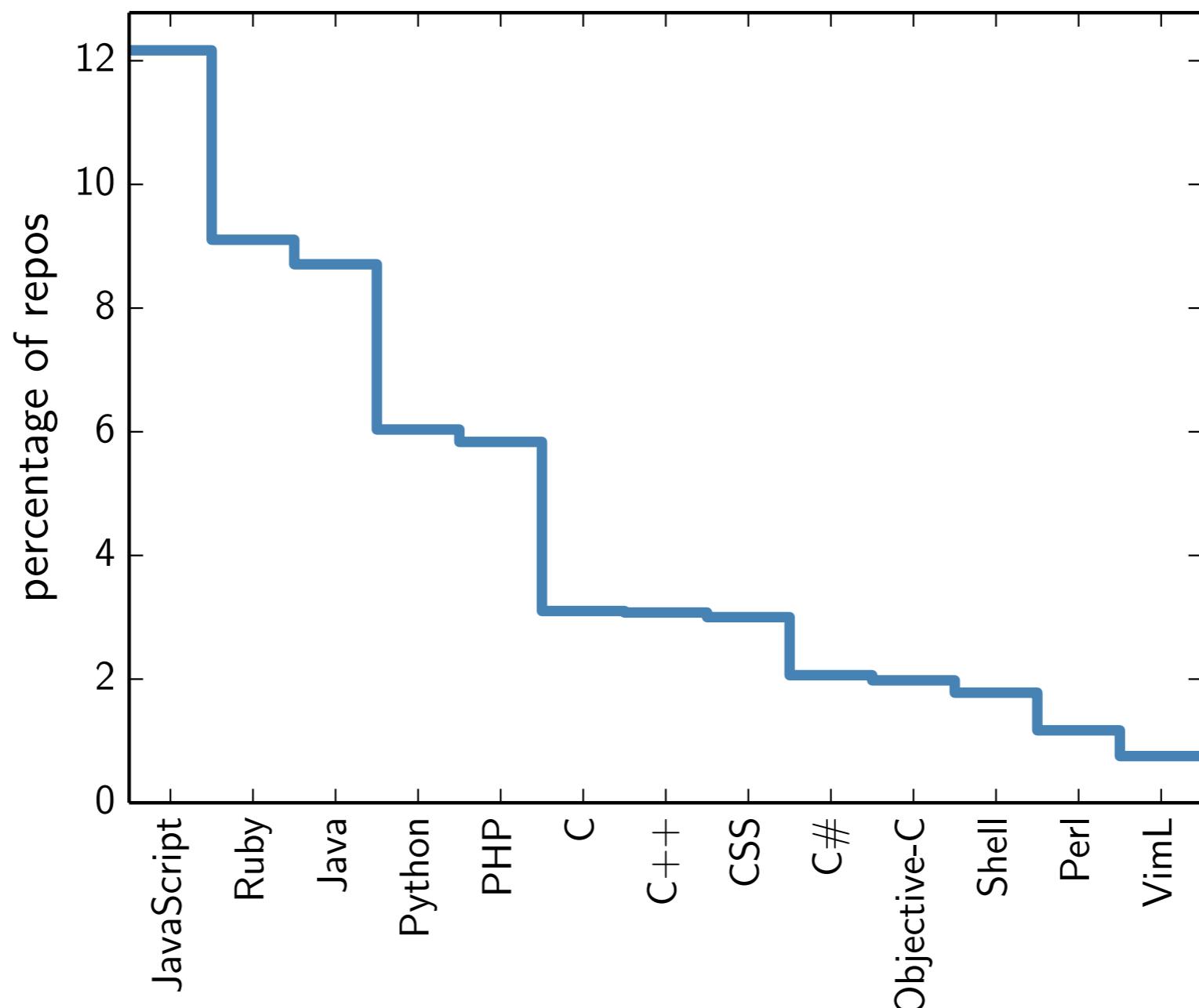
(representative)

for each **repository**, download:

```
{  
  "name": "emcee",  
  "full_name": "dfm/emcee",  
  "language": "Python",  
  "fork": false,  
  "subscribers_count": 39,  
  "forks": 117,  
  "stargazers_count": 414,  
  "open_issues": 35,  
  ...  
}
```

+ **README\***, **LICENSE\***, **COPYING\***

in the top directory



63

languages more  
popular than IDL

(only?)

**1,578,967**  
repositories

**263,288**

**are licensed**

**17%**  
are licensed

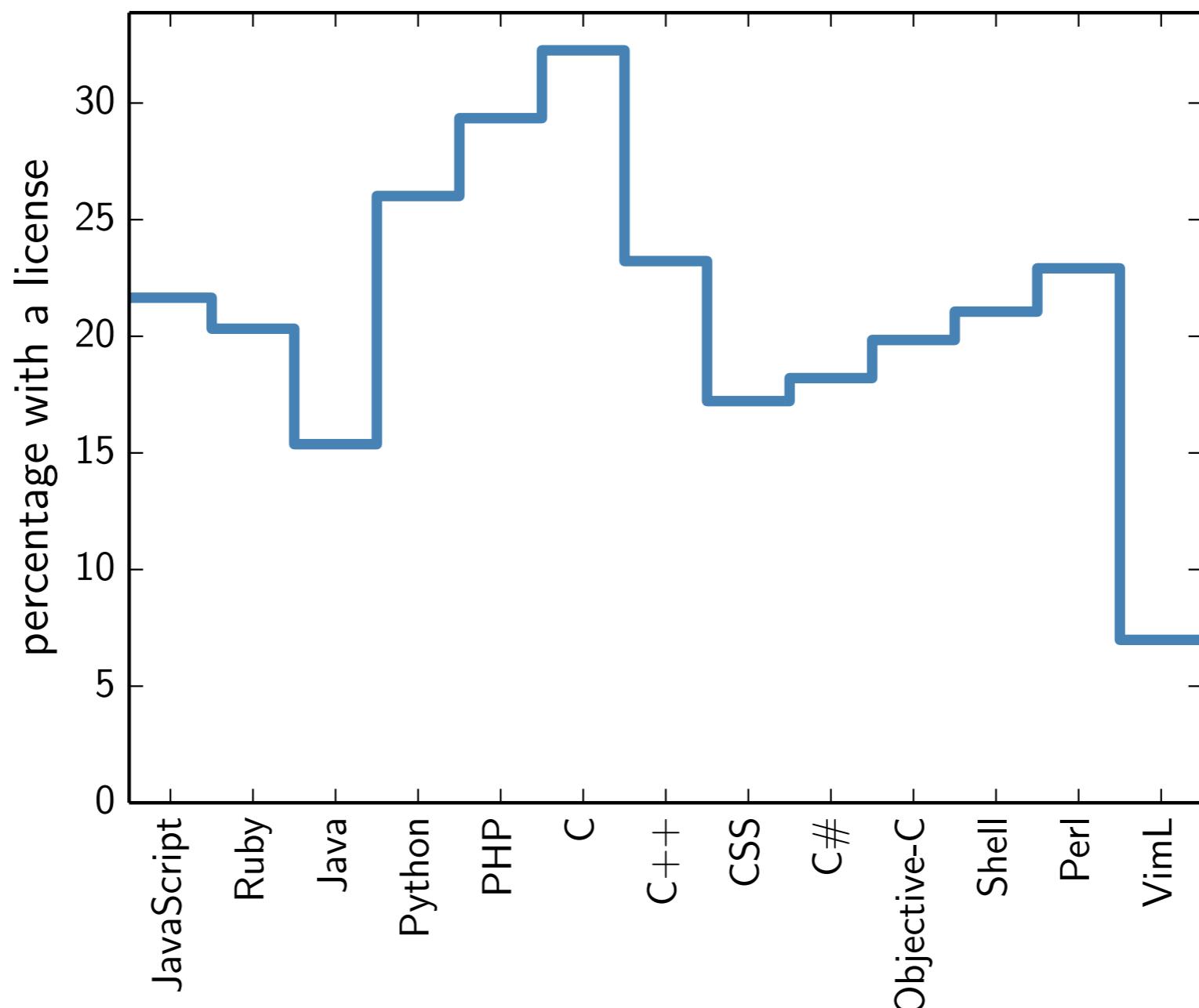
that's **up from 15%** in 2012!

[softwarefreedom.org/resources/2013/lcs-slides-aaronw](http://softwarefreedom.org/resources/2013/lcs-slides-aaronw)

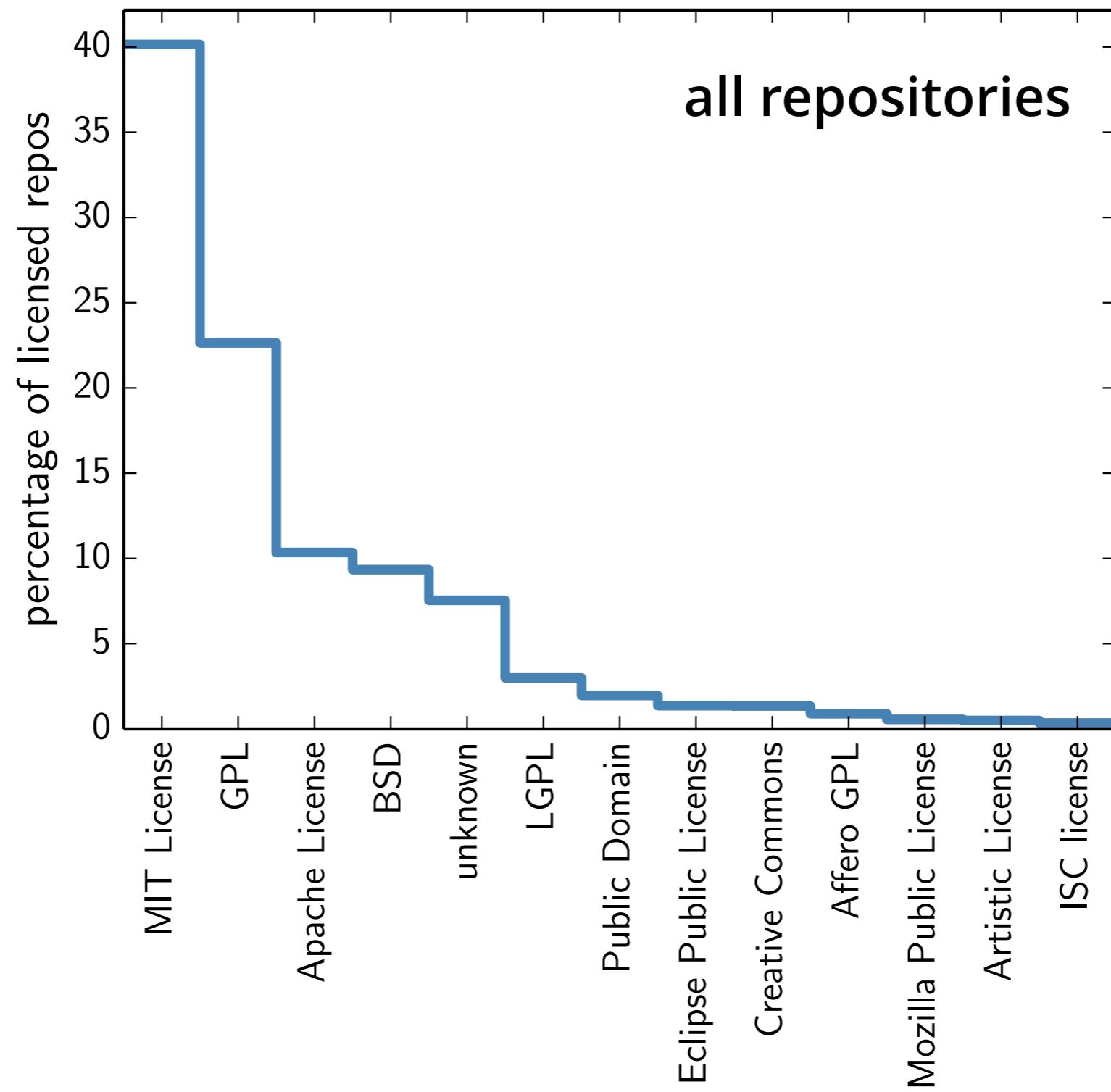
of these,

**28,972**

only mention the  
license in the README



license  
popularity



Dan F-M on Twitter: "Of the x

Twitter, Inc. [US] https://twitter.com/exoplaneteer/status/547056929668075520

Apps Betterize! Betterize 2.0 Recent ArXiv New ArXiv Manual:Database lay 52 Recipes K2 Field Descriptions Betterize (with a bet Licensing

Home Notifications Messages Discover Search Twitter DAMN

**DAMN** Dan F-M @exoplaneteer

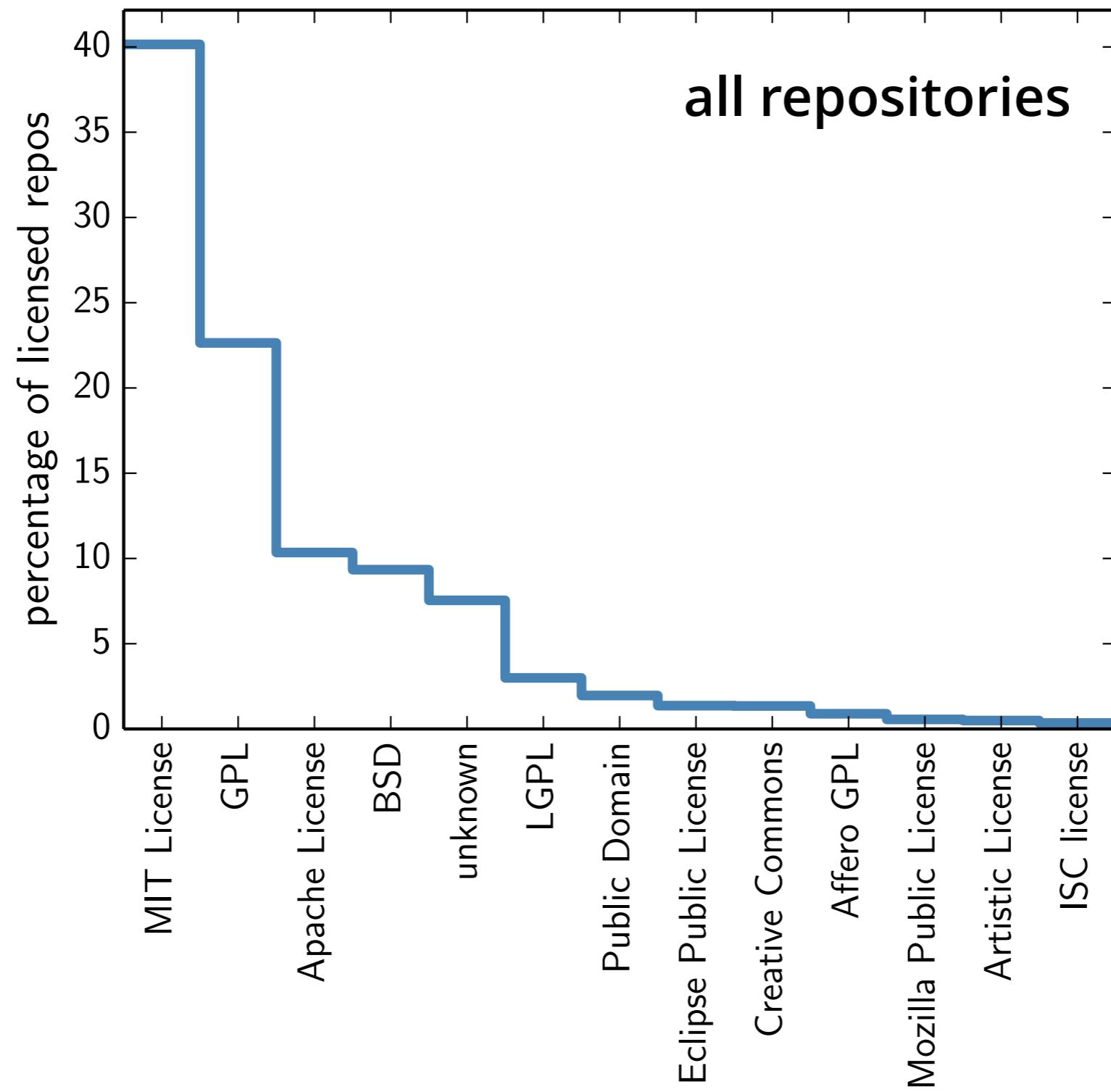
Of the 1.6M randomly chosen repos that I've downloaded off [@github](#), only 15% have licenses & here's the distribution:

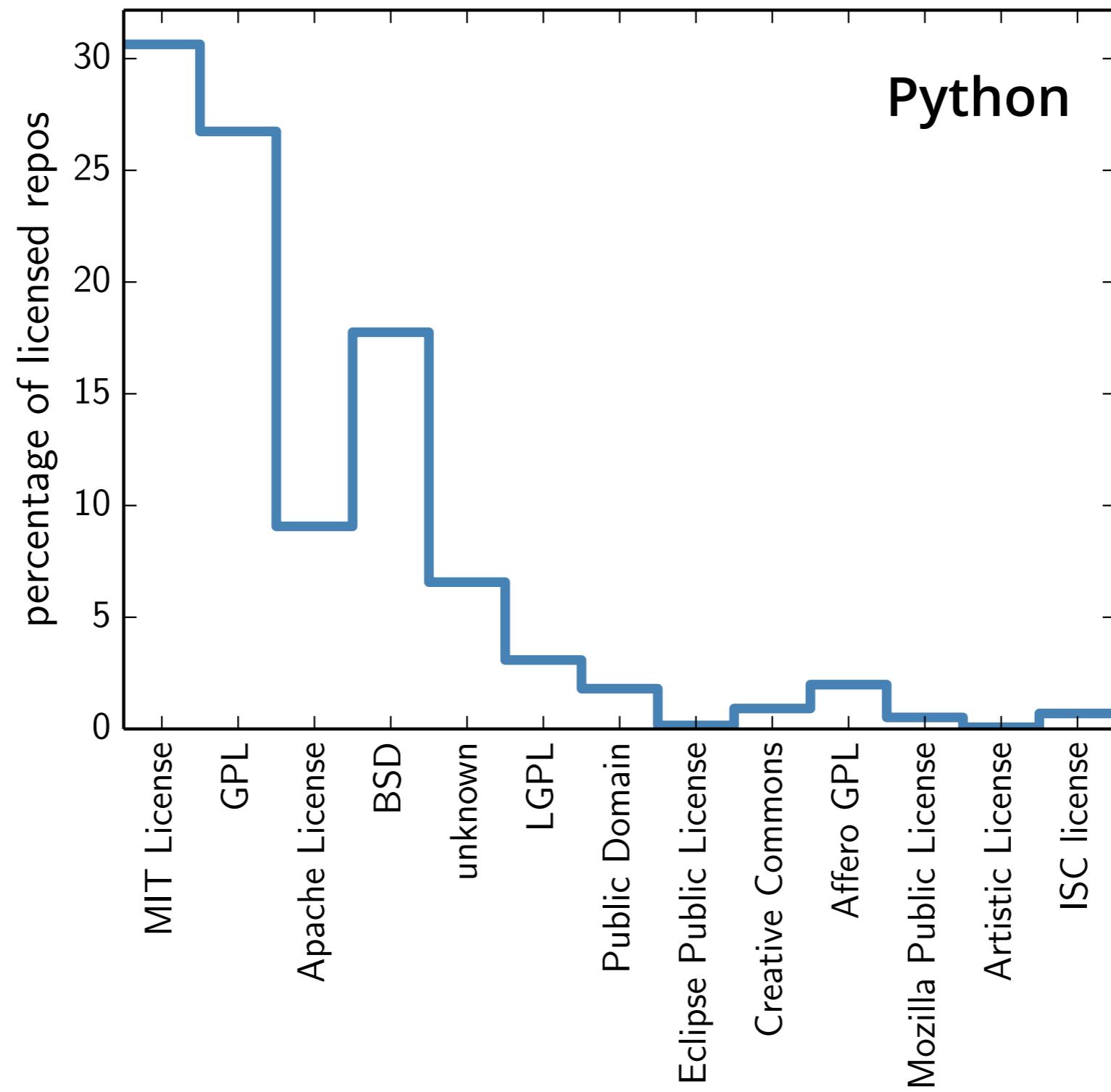
📍 Toronto, Ontario

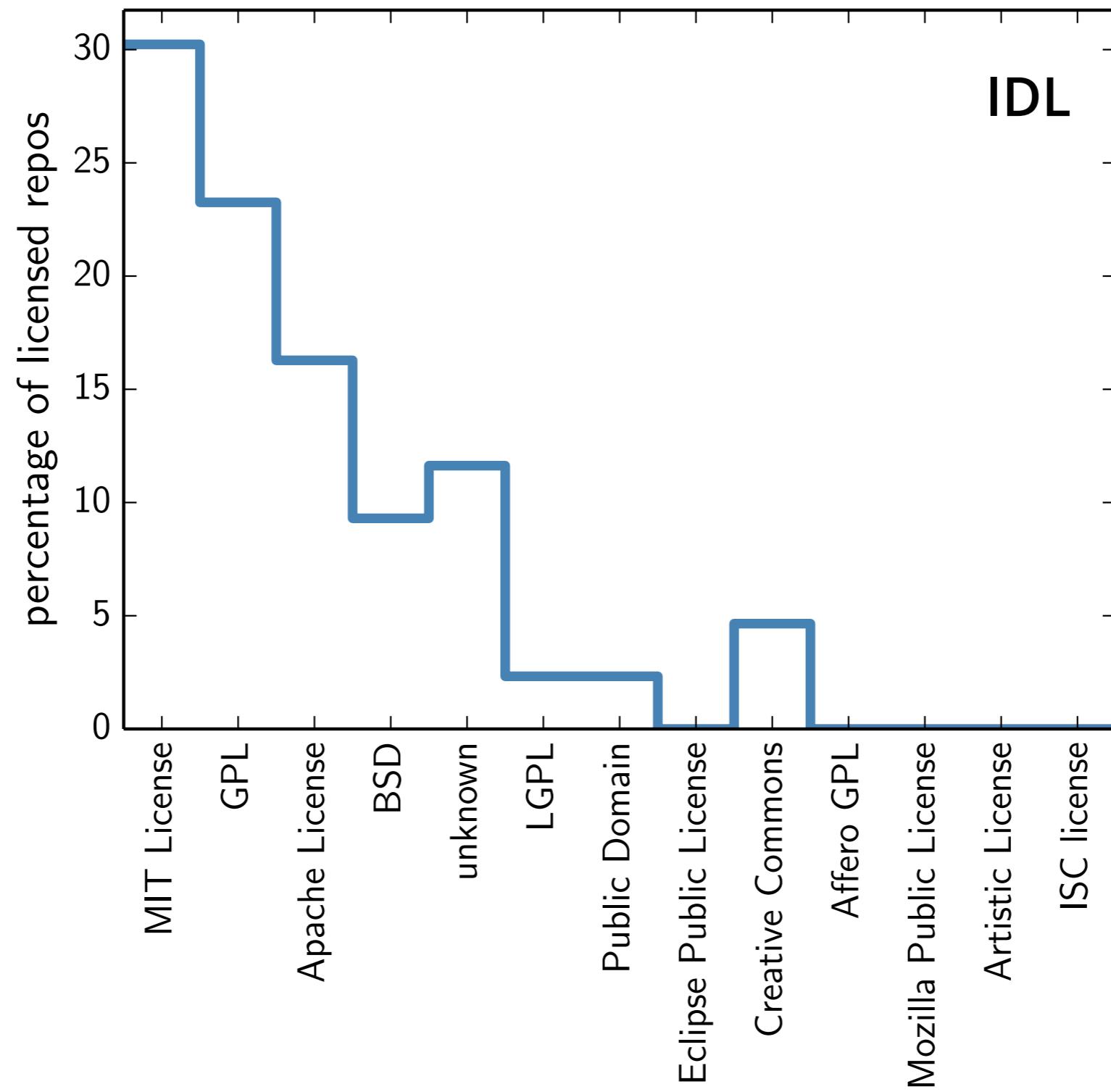
...

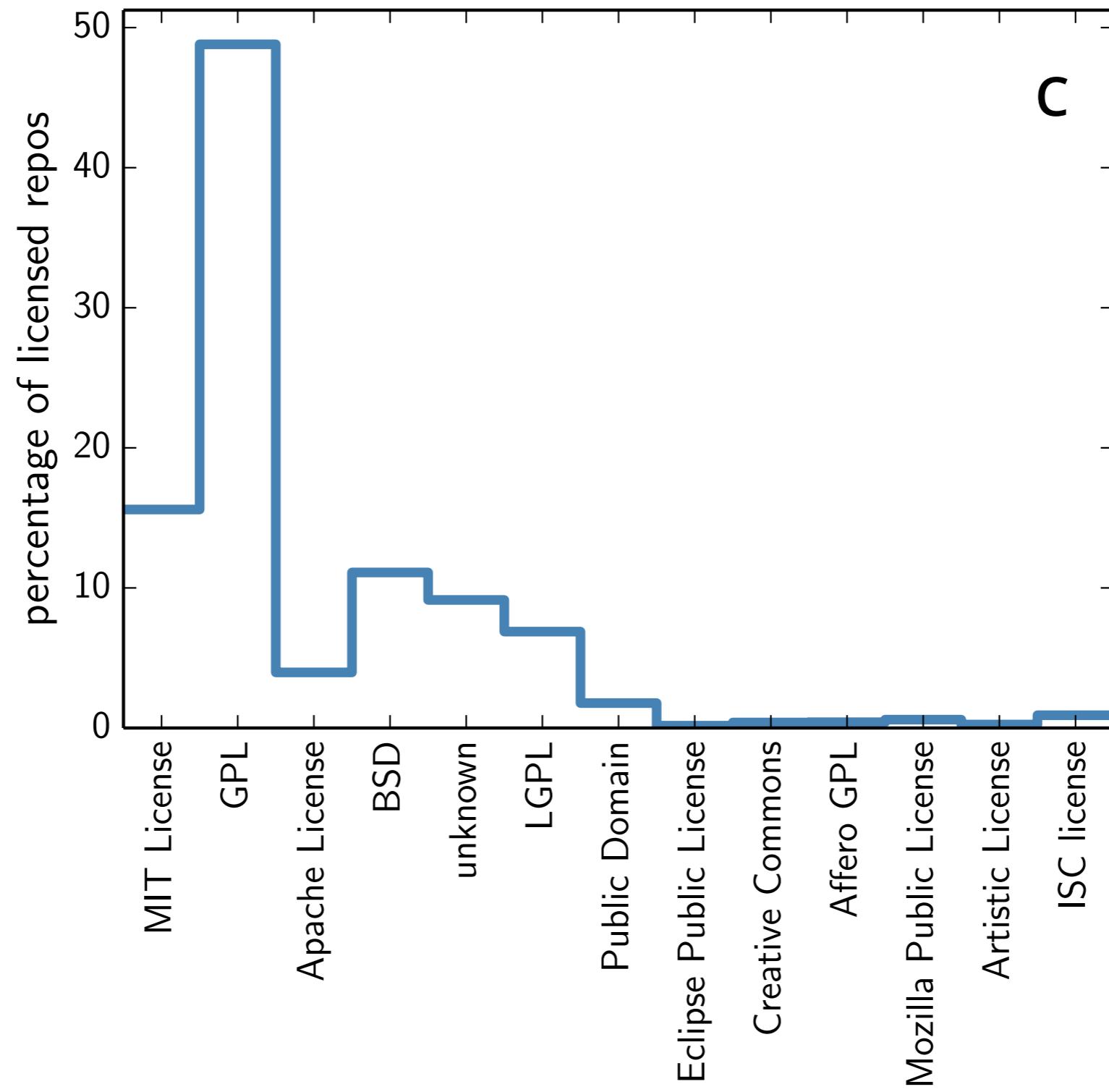
License	Percentage (%)
MIT License	40
CC-BY SA v2.0	~14
CC-BY v2.0	~10
CC-BY v3.0	~9
CC-BY-NC v3.0	~8
CC-BY-NC-SA v3.0	~7
CC-BY-NC-ND v3.0	~6
CC-BY-ND v3.0	~5
CC-BY-NC-ND v4.0	~4
CC-BY-NC-SA v4.0	~3
CC-BY-ND v4.0	~2
CC0 v1.0	~1
Apache License 2.0	~1
ISC license	~1
1.0 Universal	~1
Public Domain	~1

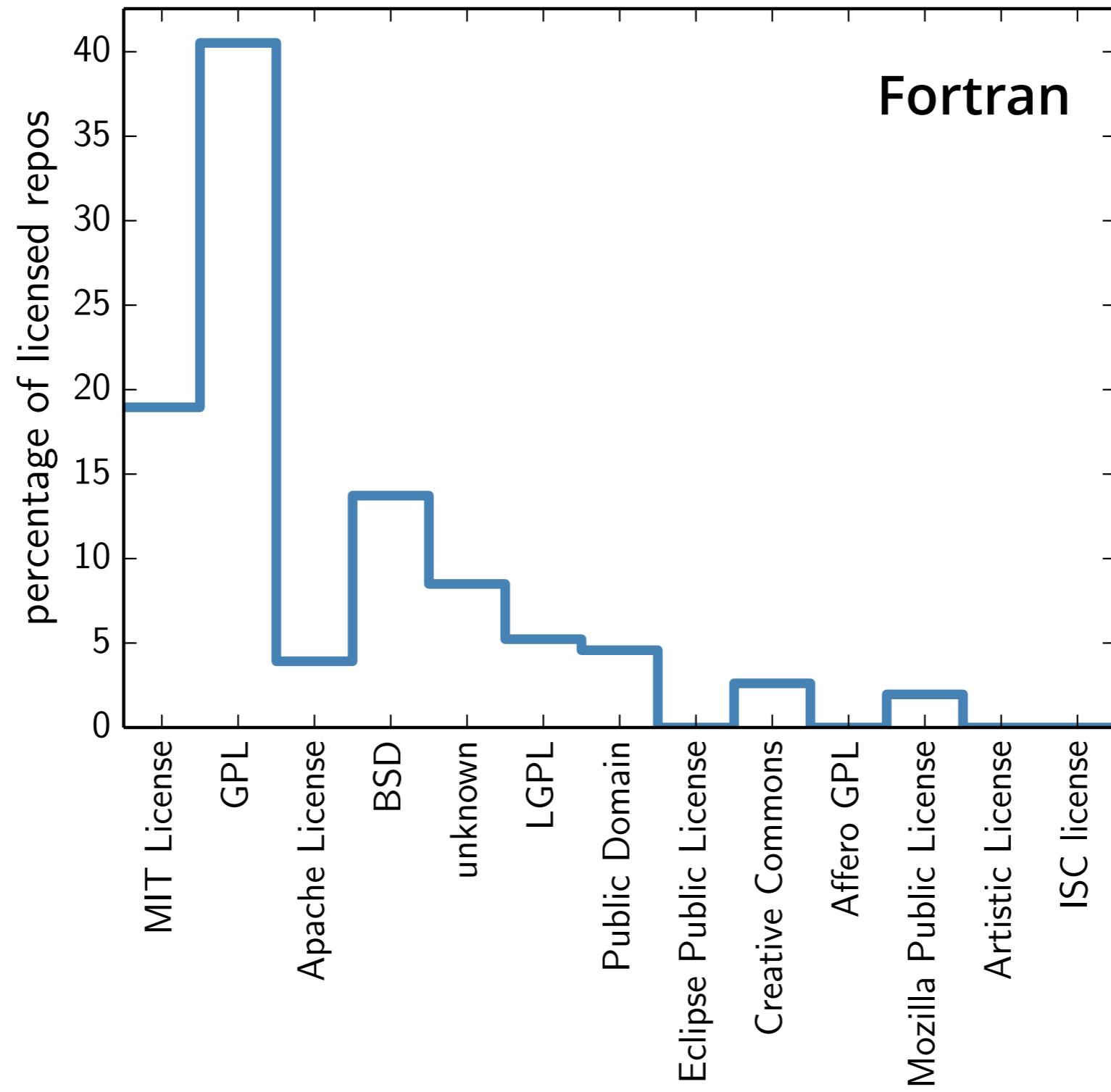
the internet gets really  
**EXCITED**  
about licensing

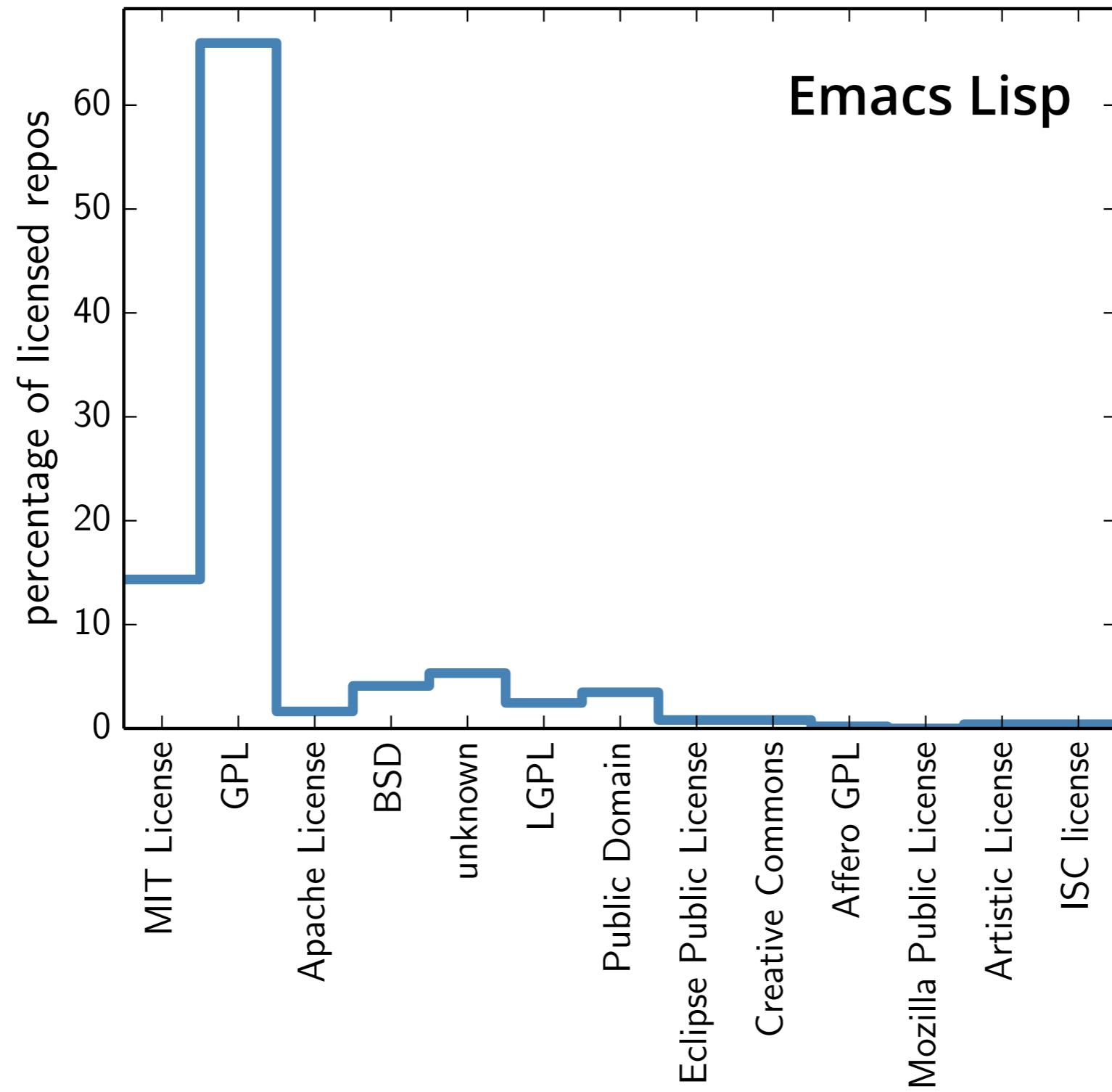


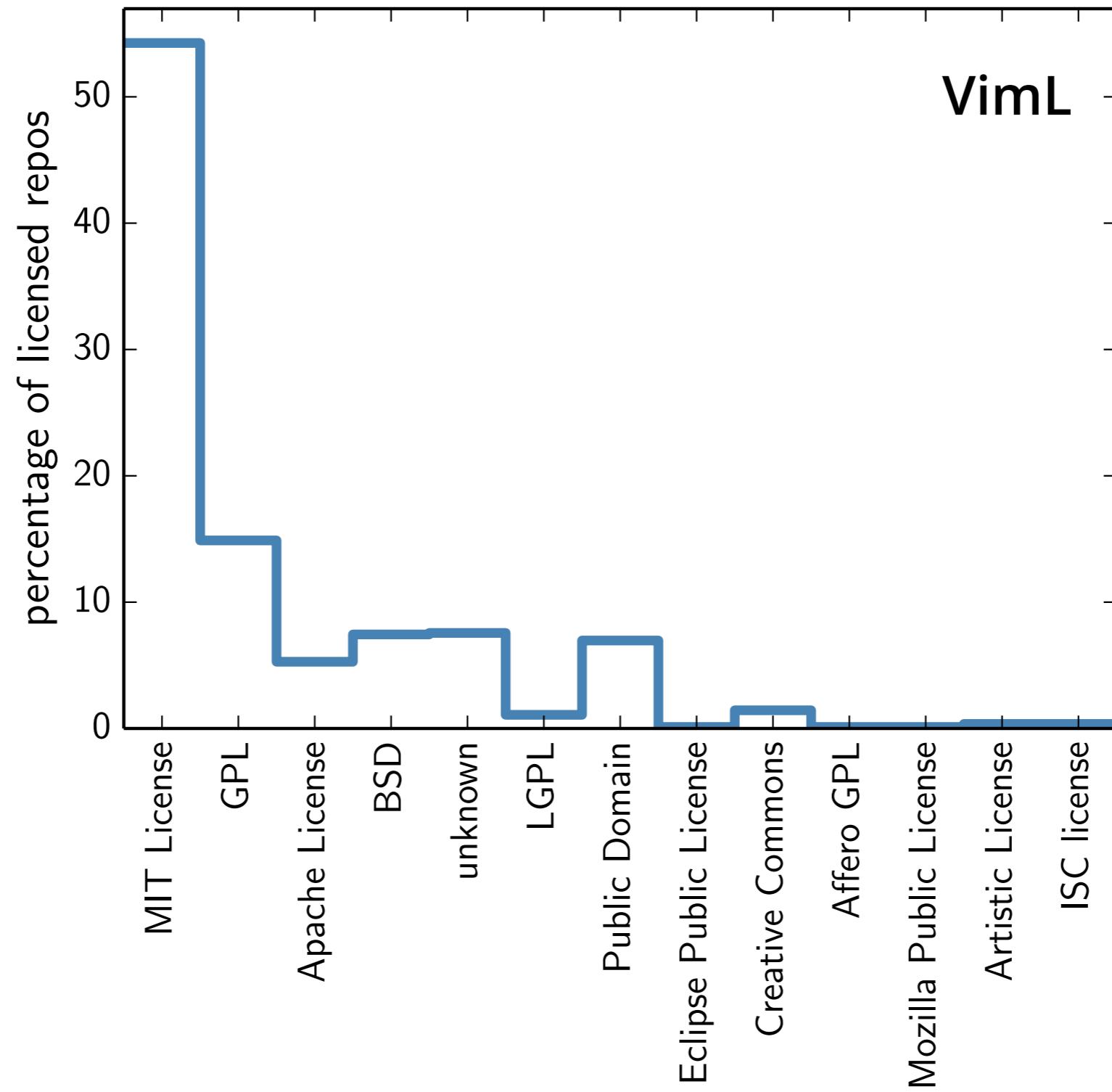


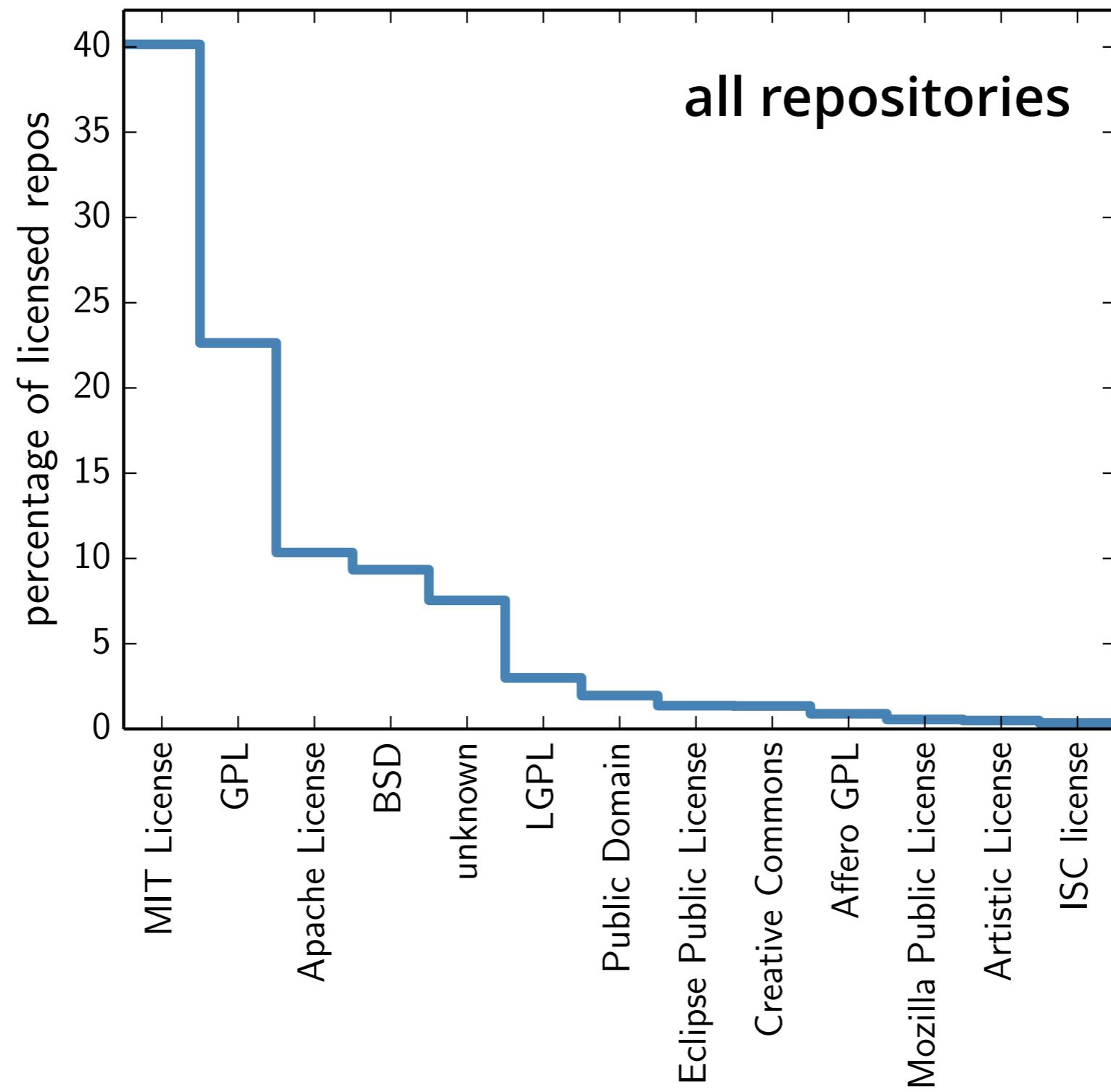


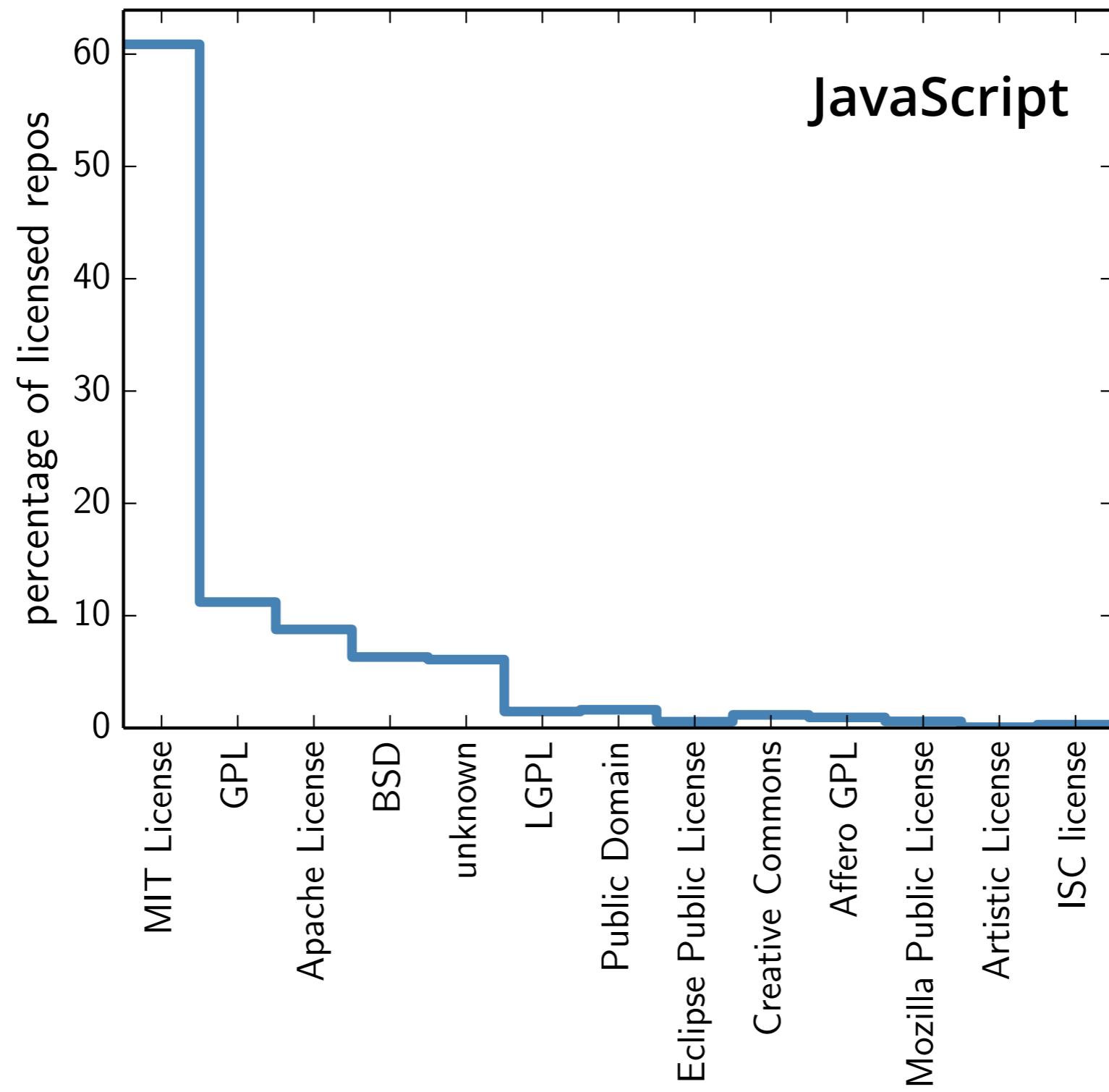


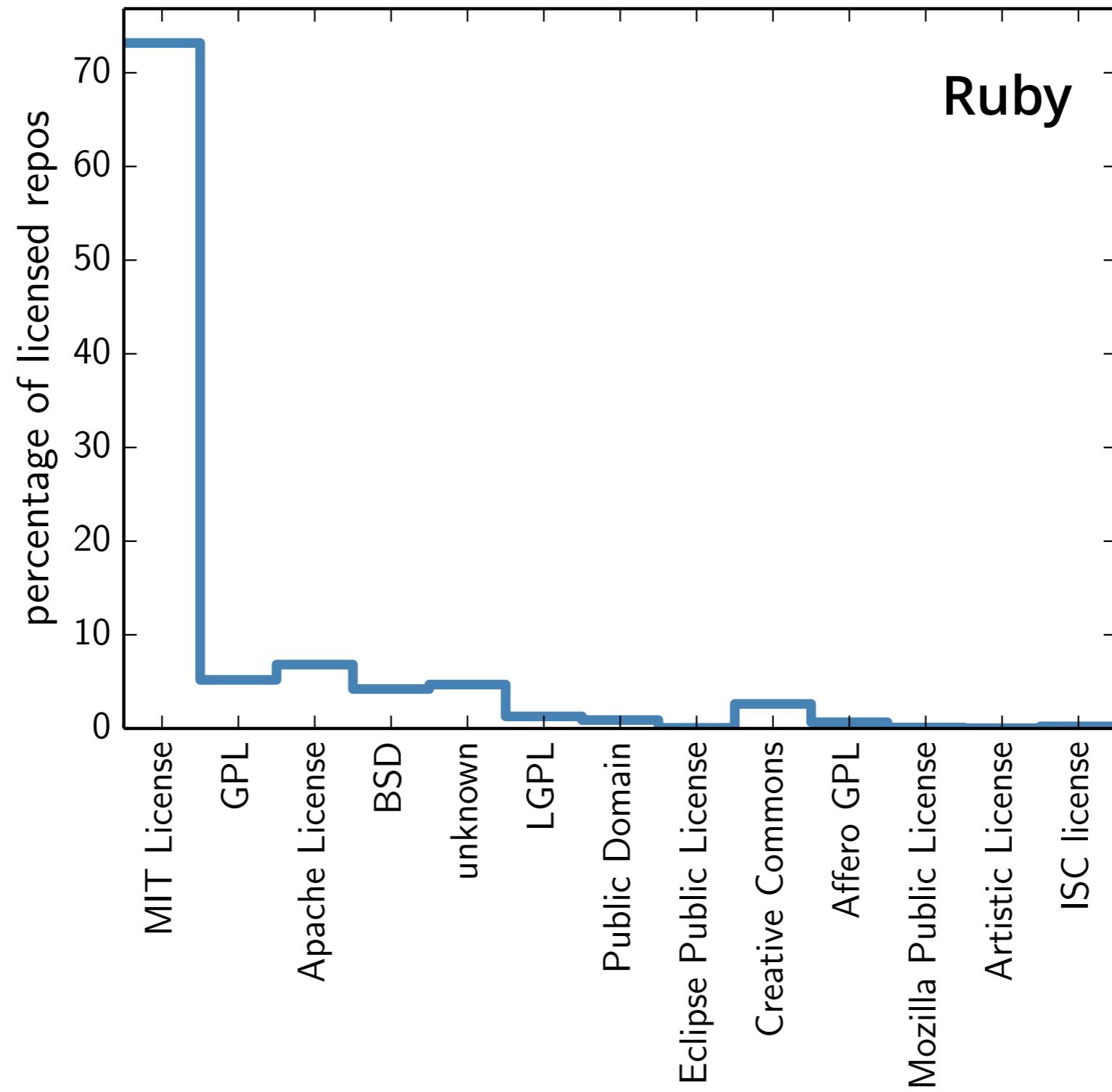


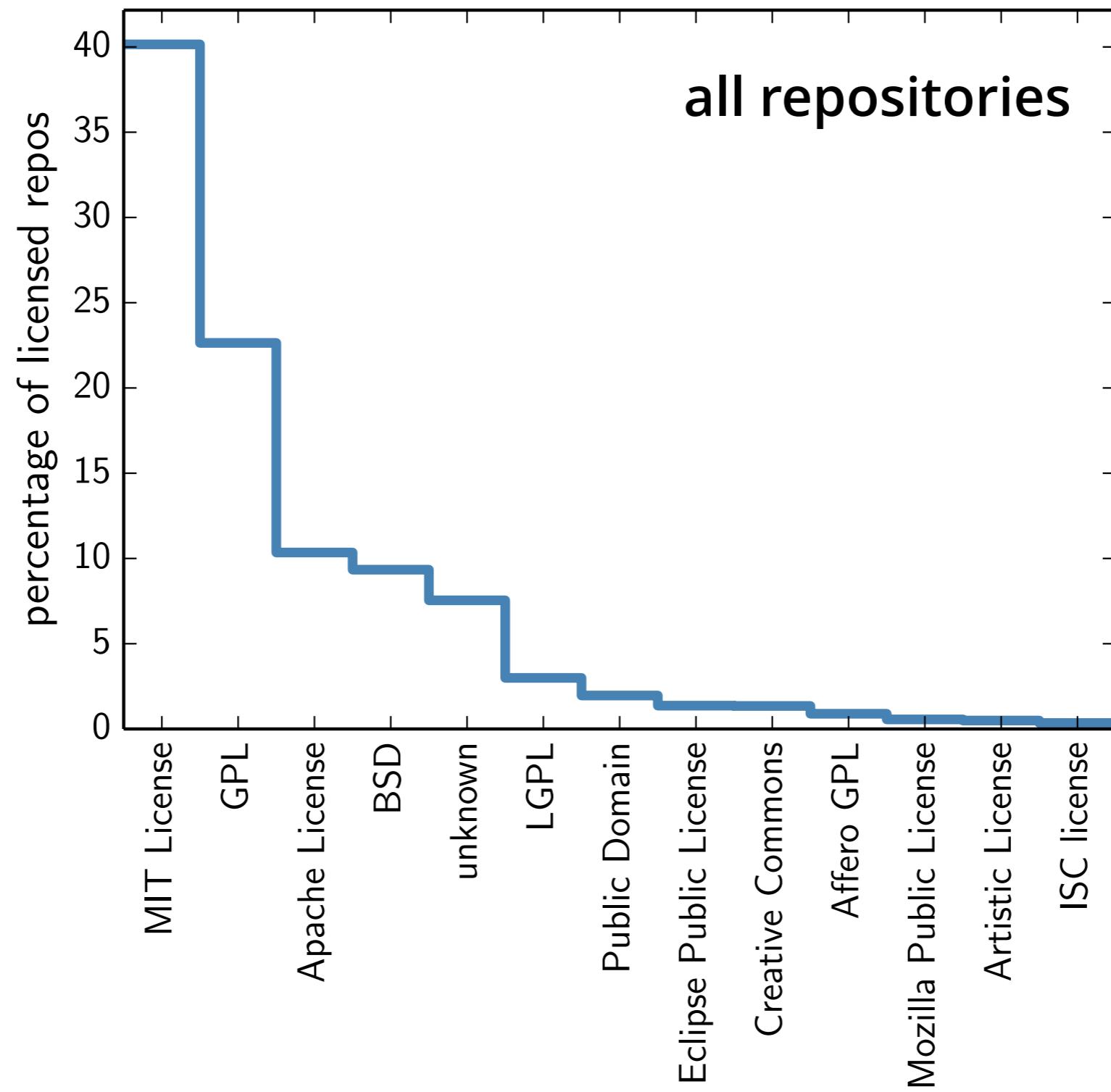


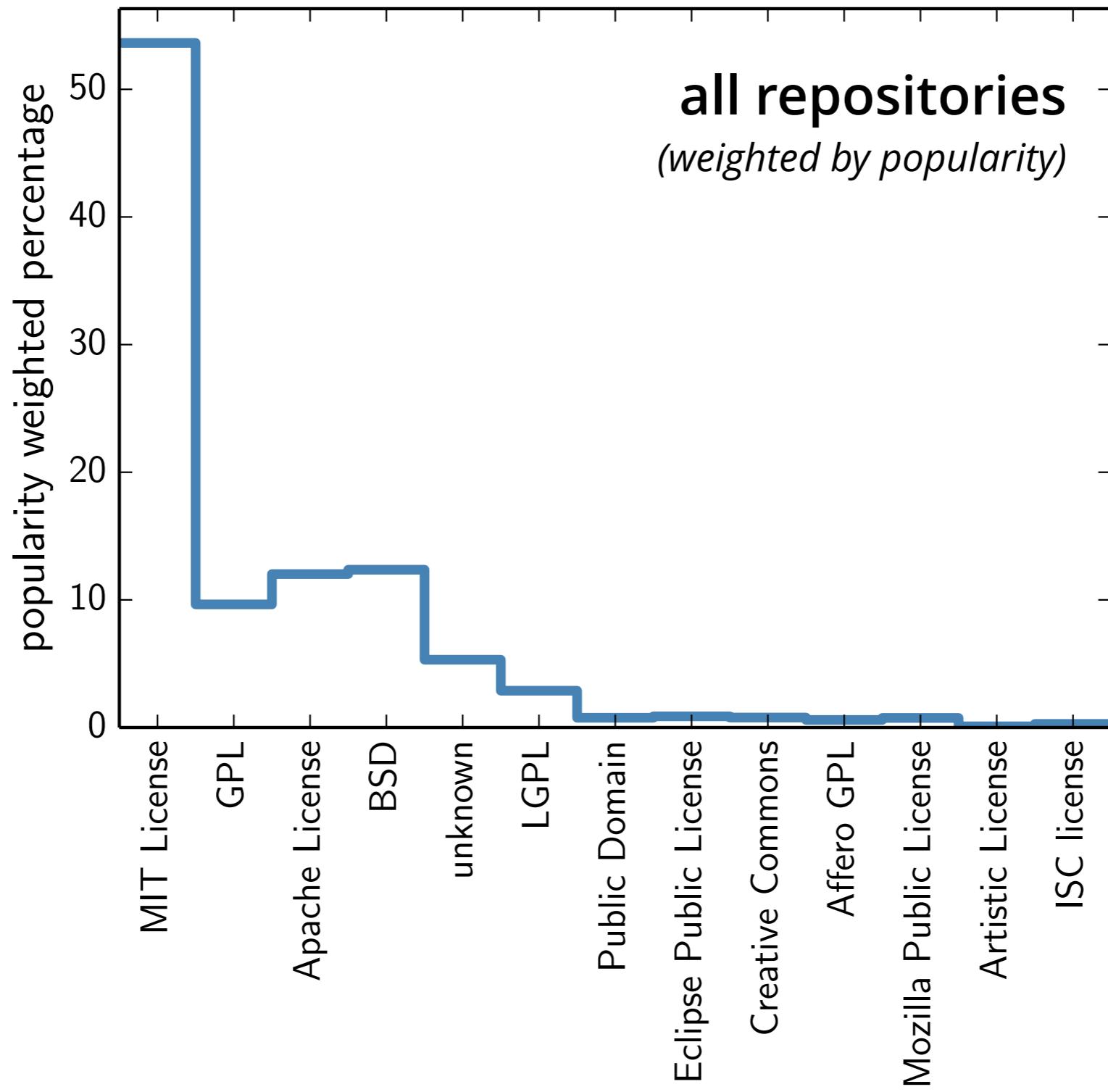














# CAVEATS

I only scraped GitHub



# CAVEATS

I only found top-level  
license files  
*(with the right names)*



# CAVEATS

my README parsing  
was pure hack!

(it's called *data science*)

# TAKE HOME

17%

of repos are licensed

*and*

the kids are permissive

*Thanks to Aaron Williamson for the 2012 data!*