We collected over 1.9M repositories with OSI-approved licenses as of July 2019. Of these, we analyzed 4.9M repos that have at least one commit.

Although 87 OSI-approved licenses exist, the top 13 licenses contain >99% of OSS repos on GitHub. The most popular ones are:
- MIT: Massachusetts Institute of Technology
- Apache License
- GPL: GNU Public License
- BSD: Berkeley Software Distribution

License regulations vary:
- The MIT license allows developers to use the code for any purpose.
- The GPL license grants the ability to use the code under the stipulation that the derivative work remains open source.

MIT is the most common OSS license (55%). These five licenses (presented above) together comprise about 93% of all OSS on GitHub.

Most repositories have fewer than 5 contributors. On average, 1.94 contributors work on each OSS repository. Only 0.3% of all OSS repositories have more than 50 contributors.

In the data, only ~2% of the users’ sectors can be identified. Users are not required to accurately fill in organization or location information.

We can use the self-reported company field in contributors profile.
- Only 2% of the contributors can be identified.
- OSS is becoming more permissive as businesses contribute more code.

OSS is growing rapidly, 2.350% increase in the number of repositories from 2012 to 2018. Permissive licenses are becoming more common (MIT is the most popular OSS license).

We collect more detailed data on the OSS repositories, including additions and deletions to estimate the developer cost using the lines of code. We obtain contributor emails to improve the sector analysis. We conduct network analysis to study interactions between contributors and OSS projects, and diffusion of OSS innovation.