

# Open Source Development Around the World: A Comparative Study

---

発表者：ダニエル

Thaís Mombach<sup>1</sup>, Marco Tulio Valente<sup>1</sup>, Cuiting Chen<sup>2</sup>,  
Magiel Bruntink<sup>2</sup>, Gustavo Pinto<sup>3</sup>

<sup>1</sup>Federal University of Minas Gerais, Brazil; <sup>2</sup>Software Improvement Group,  
Netherlands; <sup>3</sup>Federal University of Pará, Brazil

# 背景

---

- Little is known about how the growth of open source software is distributed globally
- Characteristics of open source projects developed in different countries are not well understood.
- Example: The Linux kernel has contributions from over 15,000 developers across at least 71 countries.
- Modern platforms like GitHub enable geographically dispersed developers to collaborate on projects.

# 目的

---

- 20 か国の 2,648 のオープンソース プロジェクトを分析します
- 国ごとのプロジェクト数を明らかにします
- 各国のプロジェクトの人気と主要なプログラミング言語を評価します。
- 国のプロジェクト数と GDP を相関させます

# 関連研究

---

## 社会調査:

- オープンソース コミュニティ: これらのコミュニティの特徴とダイナミクス
- オンボーディングと保持: 新しい貢献者の統合と保持に影響を与える要因
- ジェンダー バイアス: オープンソース環境におけるジェンダー バイアスの調査

## 人気と魅力:

- オープンソース プロジェクトの人気と魅力の理由に関する研究
- オープンソース コミュニティにおける多様性と包括性の障壁に対処する
- より公平な社会の構築へのオープンソースの貢献

# 調査方法/ 研究デザイン

---

## プロジェクトの選択基準:

- 大規模なユーザーと開発者を抱える人気プロジェクトに焦点を当てます。
- GitHub 上のトップ 10,000 のオープンソース プロジェクトをstarsの数で選択 (2017 年 11 月現在)。

## 人気指標:

- GitHub のスターは、プロジェクトの人気の代理として使用されます。

## 地理的位置の決定:

- 地理的位置は、GitHub アカウントのメタデータから取得されます。
- 例: aserg-ufmg はブラジルのBelo Horizonteにあるため、プロジェクト aserg-ufmg/jscity はブラジルのものとみなされます。

# 調査方法/研究デザイン

## 位置データの制限:

- 3,311 プロジェクト (33%) には空の場所フィールドがあり、削除されました

## 非ソフトウェア プロジェクトの削除:

- ソフトウェア システムとみなされない 645 のプロジェクト (プログラミング書籍、チュートリアルなど) が削除されました。
- 削除は、cloc ツールを使用したコード行数に基づいて行われました

**cloc**

*Count Lines of Code*

<https://github.com/AlDanial/cloc>

cloc は、多くのプログラミング言語のソース コードの空白行、コメント行、および物理行をカウントします。

# 調査方法/ 研究デザイン

---

## 手動検査:

- 最初の作成者は、追加の 2,518 プロジェクト (42%) について手動で場所を検査し、国に関連付けました
- 例: 英国として手動で定義された場所「Broadcasting House, London」

## 残りの未解決プロジェクト:

- 656 プロジェクト (11%) の場所は曖昧または国以外でした (例: 「The Earth」または「Canada & France」)

# 調査方法/ 研究デザイン

## 最終的なデータセット:

- 77 か国に分散された 5,388 プロジェクトの国を特定しました。
- 保守性メトリクスの手動構成要件のため、米国を分析から削除しました。
- 次の 20 か国に焦点を当て、数と地理的分布を最大化します。
- 最終調査には 2,648 のオープンソース プロジェクトが含まれていました。

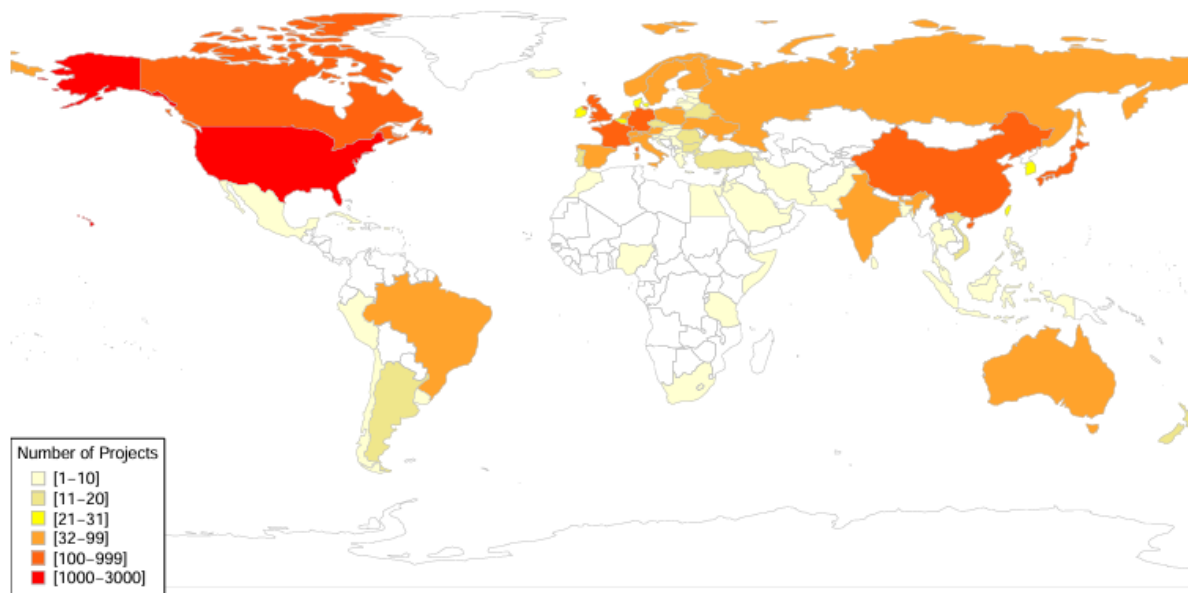


Figure 1: Number of GitHub projects per country

# 調査方法/ 研究デザイン

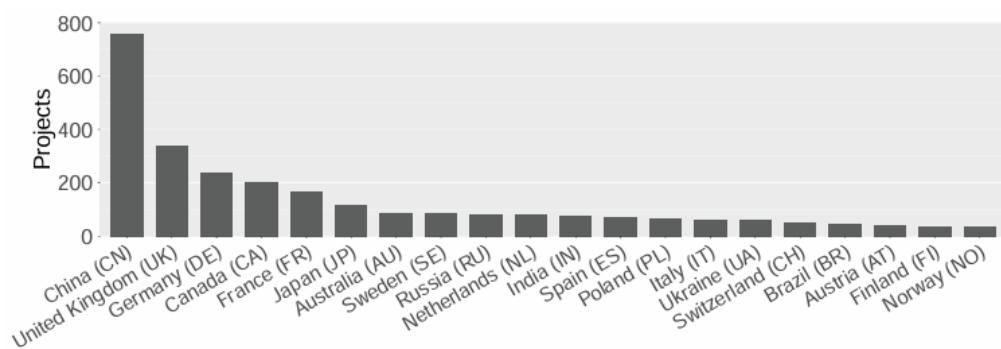


Figure 2: Countries studied in this work

# 調査結果

## GDPとの相関

GDP values obtained from the International Monetary Fund (IMF) for the most recent year available for each country

中国は最大のGDPとプロジェクト数を理由に、視覚化を容易にするために図から除外した。

相関分析:

- プロジェクト数と GDP の間には強い正の単調相関があります。
- スピアマン(Spearman's)の相関係数 ( $\rho$ ) = 0.75。
- p 値 = 0.001。

Growth Domestic Product (GDP)

# 調査結果

## GDPとの相関

- **High Ratio of Projects per GDP:**

- Ukraine: 0.61 projects per billion GDP.
- Sweden: 0.16 projects per billion GDP.
- Finland: 0.15 projects per billion GDP.

- **Low Ratio of Projects per GDP:**

- Japan: 0.02 projects per billion GDP.
- India: 0.03 projects per billion GDP.
- Italy: 0.03 projects per billion GDP.
- Brazil: 0.03 projects per billion GDP.

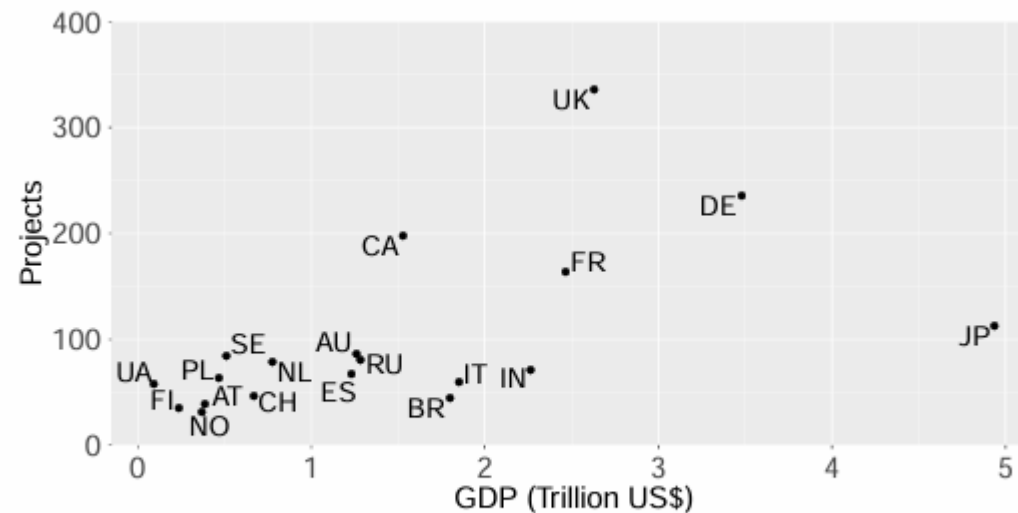


Figure 3: GDP vs number of projects (excluding China, to ease visualization)

# 調査結果

## Popularity

### 中央値:

- 国の星の数の中央値に大きな違いはありません。
- Switzerland: 1,705.5 stars
- Norway: 2,406.5 stars

すべての国に、人気のあるプロジェクトを作成できる開発者がいることを示します。

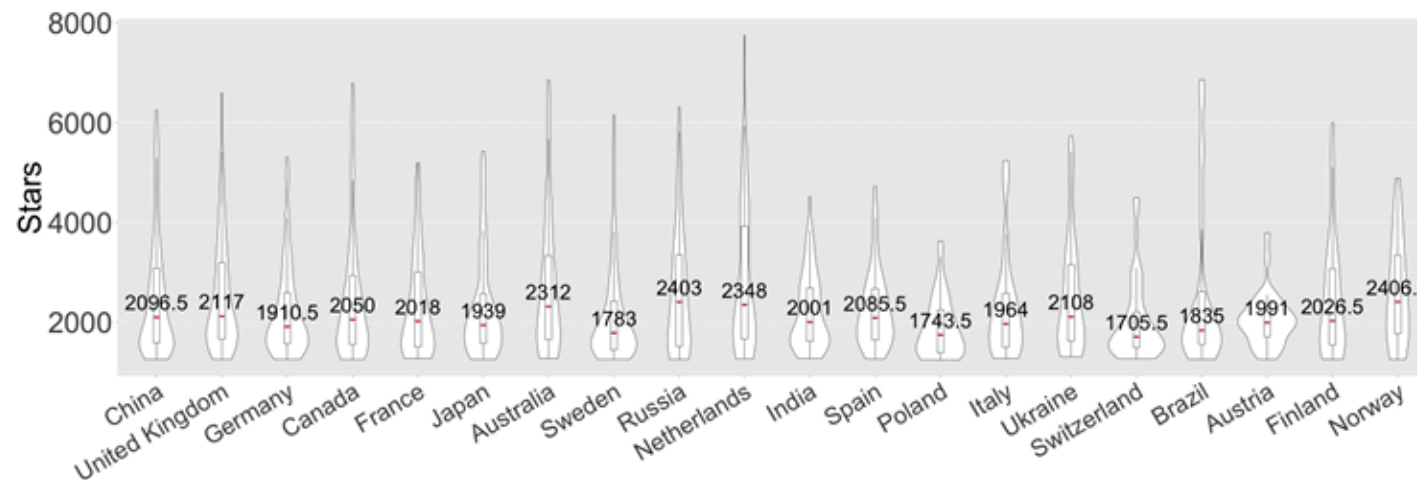


Figure 4: Popularity in terms of the number of stars

# 調査結果

## International vs Domestic Projects

所有者の所在地は、プロジェクトの国を定義するために使用されます。

コア開発者が含まれるには、プロジェクトの少なくとも 5% のコミットが必要です。

プロジェクトの分類:

- **Domestic Projects:** すべてのコア開発者はプロジェクト所有者と同じ国の出身です。
- **International Projects:** 少なくとも 1 人のコア開発者がプロジェクト所有者とは異なる国の出身です。
- **Undefined:** 少なくとも 1 人のコア開発者の国を推測できません。残りの開発者はプロジェクト所有者と同じ国の出身です。

# 調査結果

## International vs Domestic Projects

国内プロジェクトは 19 か国で最も一般的です (例外: Finland)。

国内プロジェクトの割合が最も高い分野:

- China (76%)
- Japan (73%)

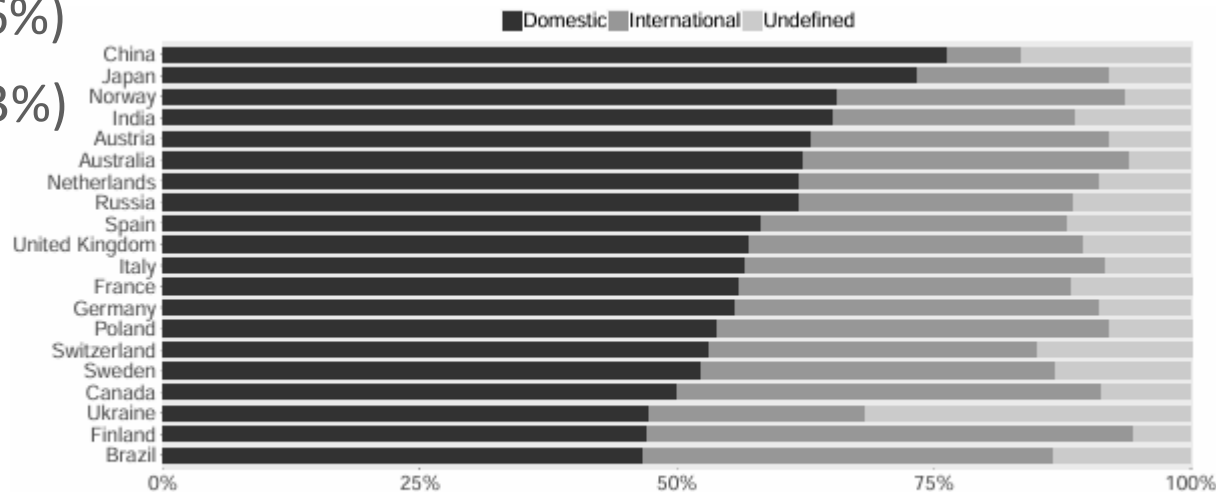


Figure 6: International and domestic projects per country

国際プロジェクトの割合が最も高い分野:

- Finland (47%)
- Canada (41%)
- Brazil (40%)

未定義プロジェクトの割合は、Ukraine (32%) を除き、すべての国で最大 17% です。

## **Popular Projects:**

- Top countries (excluding US):
  - China: 754 projects
  - United Kingdom: 336 projects
  - Germany: 236 projects

Found a positive correlation between the number of projects and a country's GDP

## **Domestic vs. International Projects:**

- Domestic projects (all core developers from the same country) are the most common in 19 out of 20 countries.

# 所感

---

- | 国内プロジェクトの割合が最も高いのはアジア国（例:中国、日本）です。  
これは言語の壁やコーディングスタイルに関係しているのでしょうか?
- | 調査方法をガイドとして使用できるかと考えます
- | なぜ国のGDPは高いのにプロジェクト比率が低いのか