

HackerRank Dataset

Data Manipulation Technique

Kostiantyn Reva

Intro

- HackRank helps developers: practice coding, prepare for interviews, and get hired [1]
- They make reports about state of IT industry based on data (surveys) from their customers [2]
- One of this is: **survey of 25,000 professionals and students on the state of developer skills [3]**
- Tags: demographics, survey analysis, HackerRank, software developers

HackerRank Dataset

Date Range: 2017/10/16 – 2017/10/31

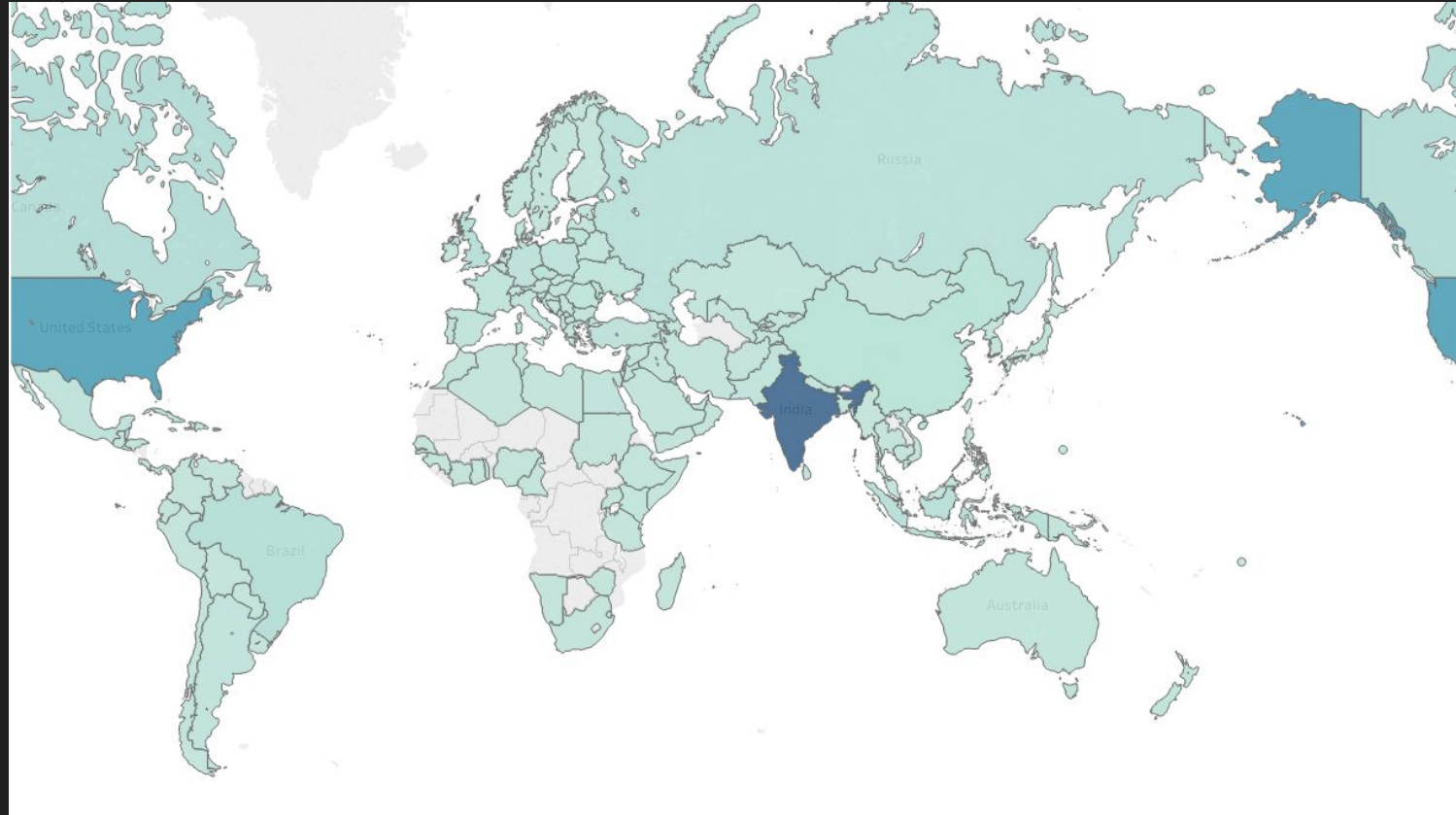
Observations	25090
Variables	251
File Size	79MB

Base Fields		Survey questions (247)	
RespondentID	NUM	q1AgeBeginCoding	CHAR
StartDate	DATETIME	q2Age	CHAR
EndDate	DATETIME	q3Gender	CHAR
CountryNumeric2	CHAR	q6LearnCodeUni	CHAR
		q6LearnCodeSelfTaught	CHAR

Questions

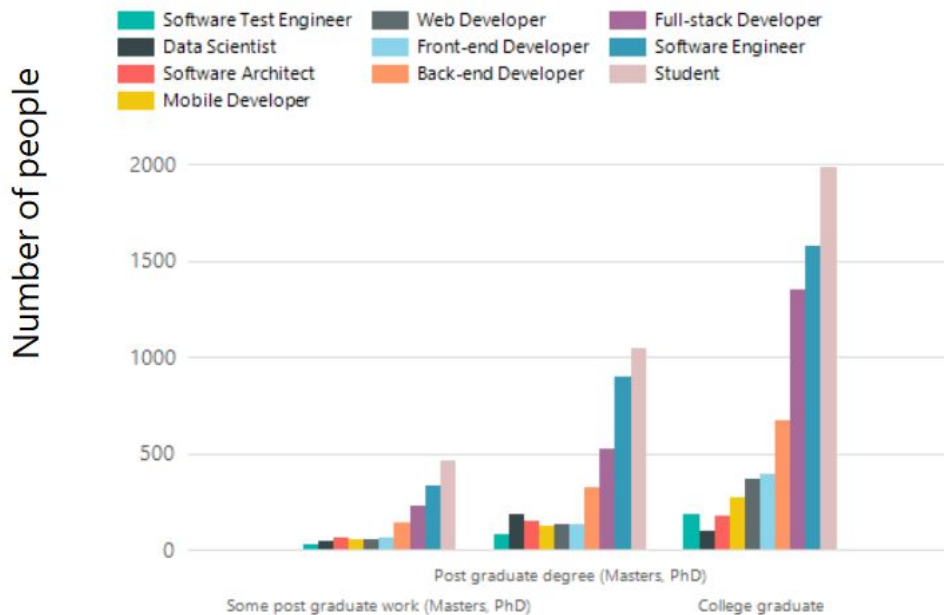
- Covered countries
- Number of developers per country per industry
- Current age of developers
- Distribution of some popular languages

Countries that were covered by the survey

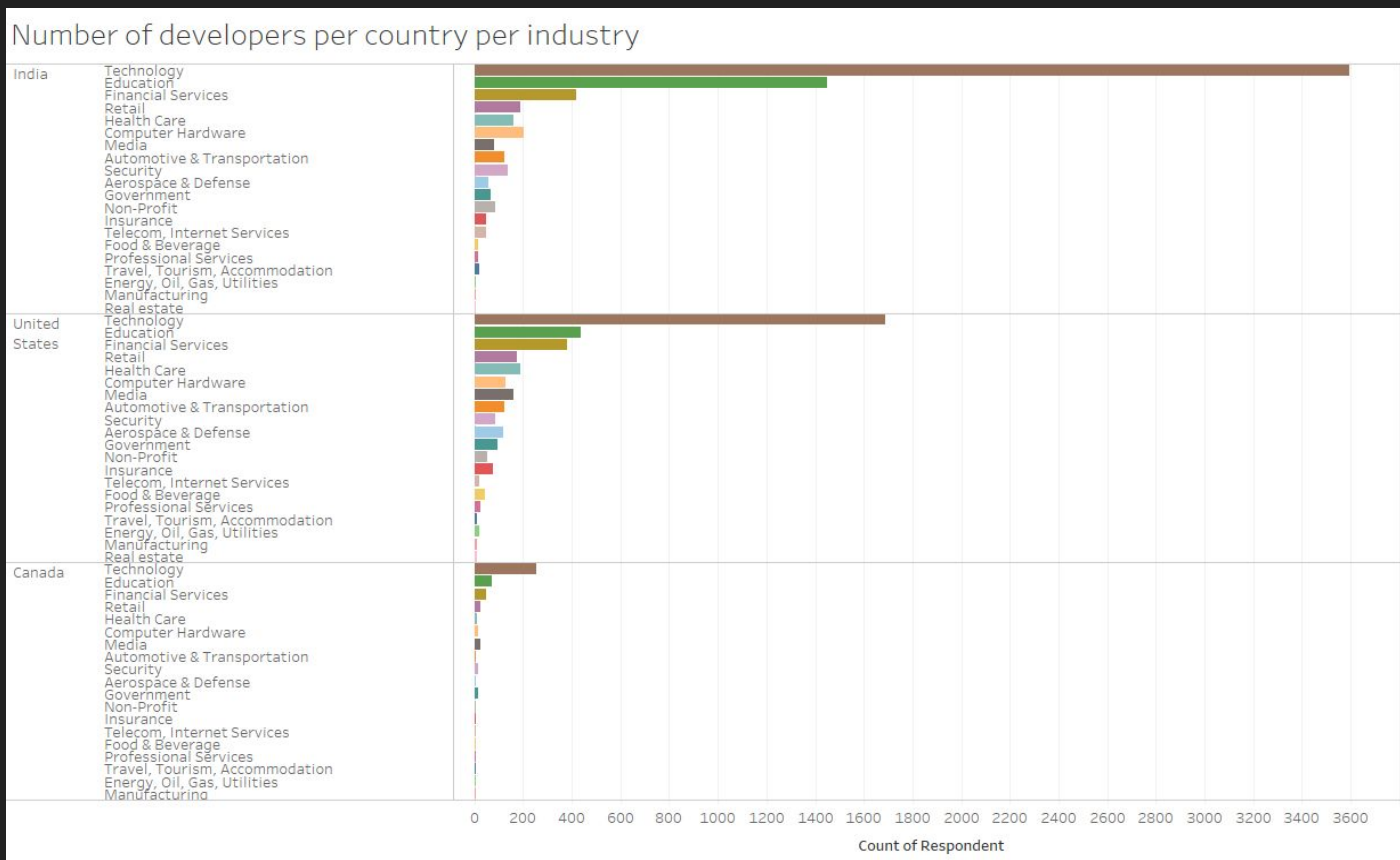


Top 3 education category

Top3 education category and corresponding current roles

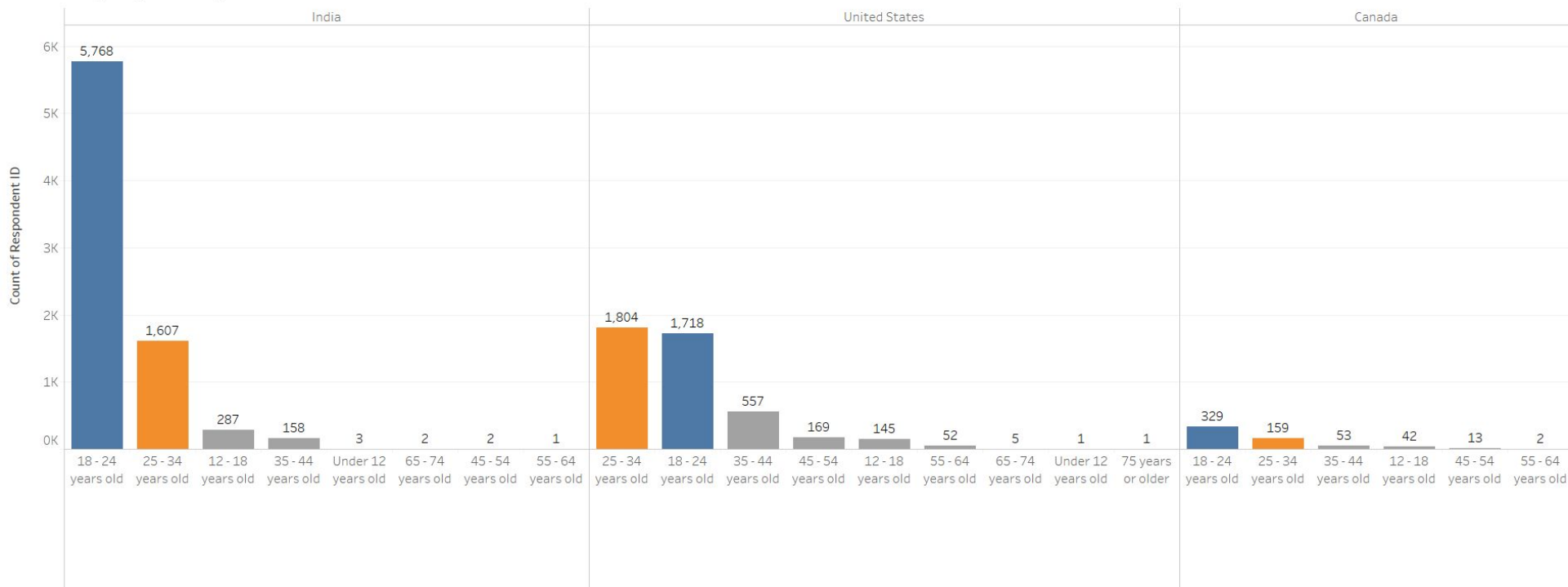


Number of developers per country per industry

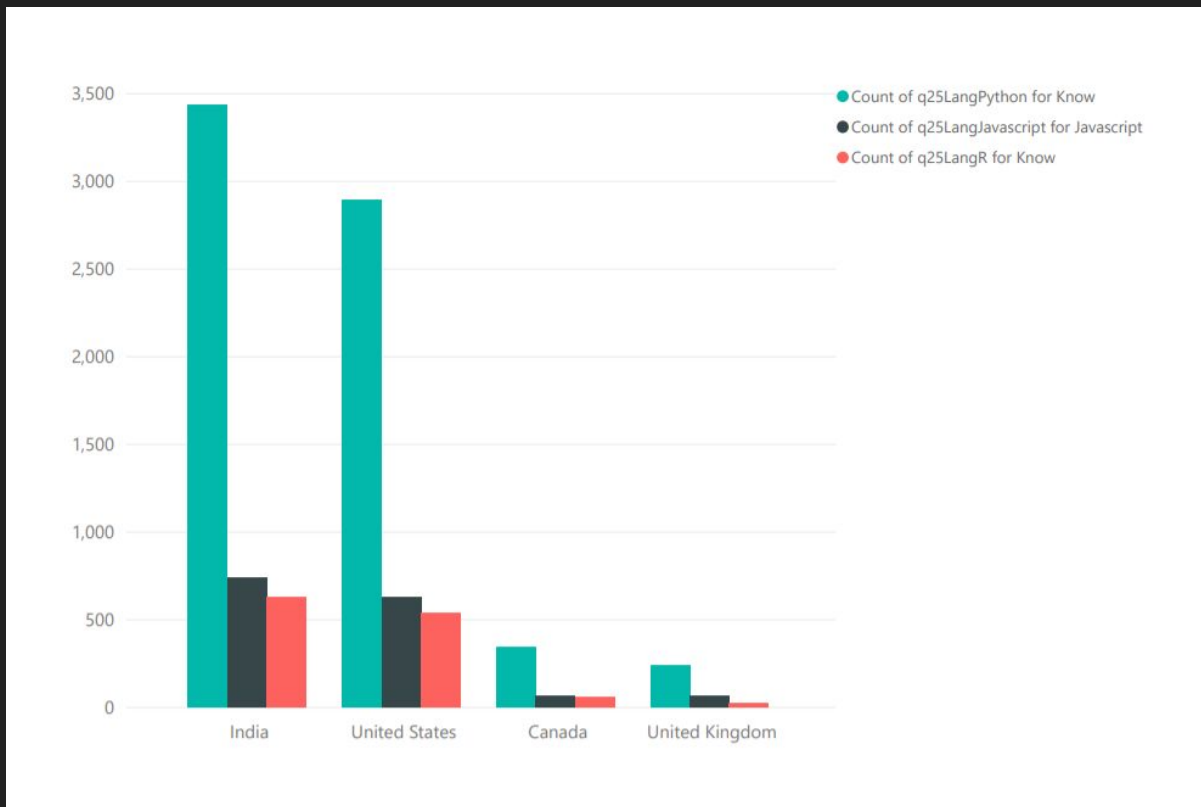


Current age by country

Current age by country



Python vs JavaScript vs R [4]



Summary

- Most developers who participate the survey from India
 - Most of them are students or beginner developers
 - Most of them work in Technology/Education industry
 - They are young: 18-34 categories cover most responses
-
- This survey highly likely does not represent real distribution of developers
 - Seems it's more about newbies/student distribution and their preferences

References

[1] <https://www.hackerrank.com/about-us/>

[2] <https://research.hackerrank.com/>

[3] <https://www.kaggle.com/hackerrank/developer-survey-2018>

[4] <https://www.tiobe.com/tiobe-index/>

Questions

