



Gender, Education, Database and Language Preferences in Programming Today

Jihee Chu

August 3, 2023

OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
 - Visualization - Charts
 - Dashboard
- Discussion
 - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY



- The data shows programming to be a male dominated field.
- The top 2 languages used today are also the 2 most sought after languages to learn
- There has been a shift in database popularity
- People in the programming field tend to be highly educated on average

INTRODUCTION



We take a look at survey data gathered from programmers to reveal insights into the demographics, education level and programming language preference of programmers today. We also note coming changes database preferences.

METHODOLOGY

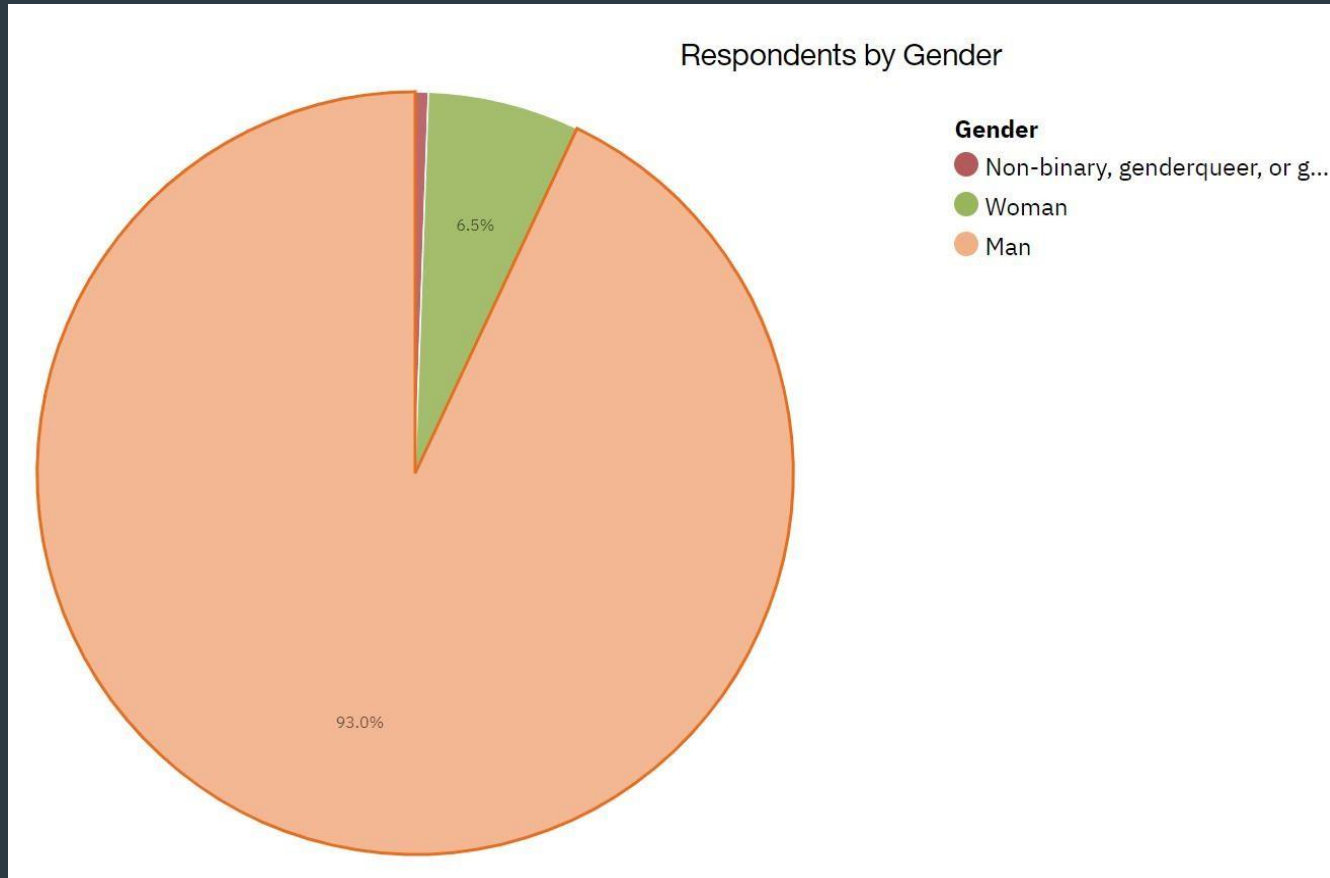


- This Data has been gathered through a survey conducted by StackOverflow.
- The data has been analyzed primarily with graphical analysis through a dashboard in IBM cloud.
- The Pandas module in Python was also used to call statistical functions returning correlation, mean, and unique value counts.

RESULTS

In the following sections we will see graphical data from survey respondents regarding their age, gender, programming language/database preference and use, education level and sex.

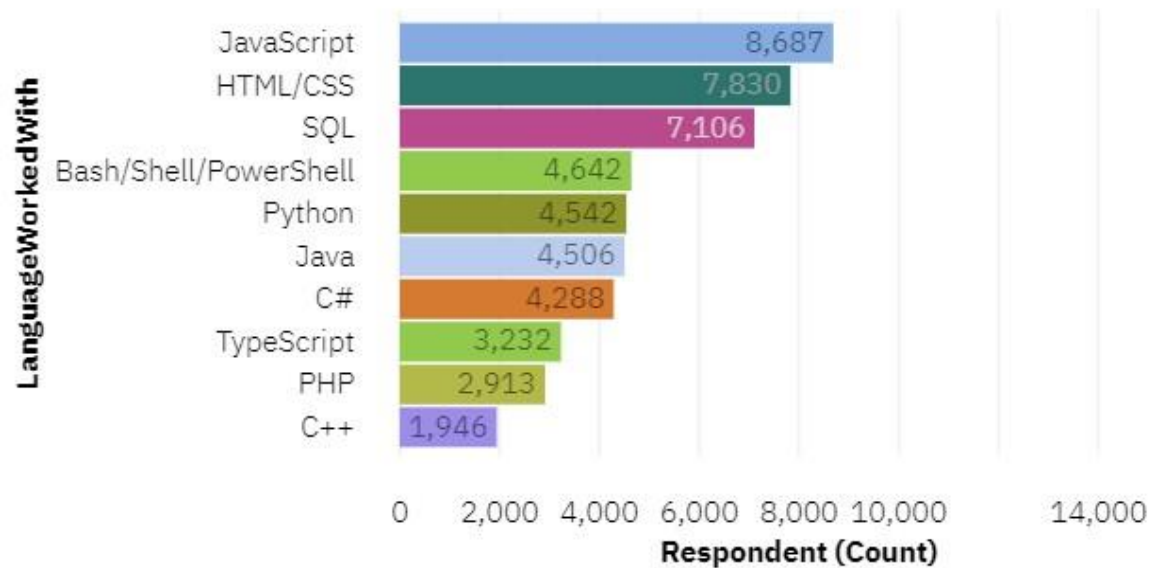
Males in Programming



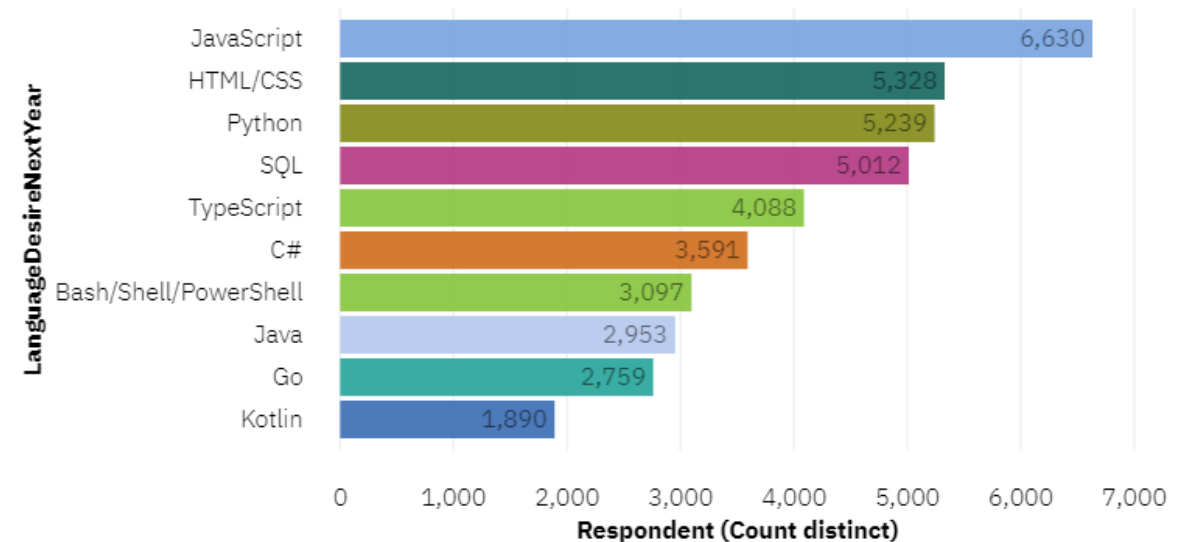
□ Over 90% of survey respondents identified as male. Of those males over 68% are of white or European descent followed by 10% south Asian males and 4% Hispanic or Latino males.

PROGRAMMING LANGUAGE TRENDS

Top 10 Languages Worked With



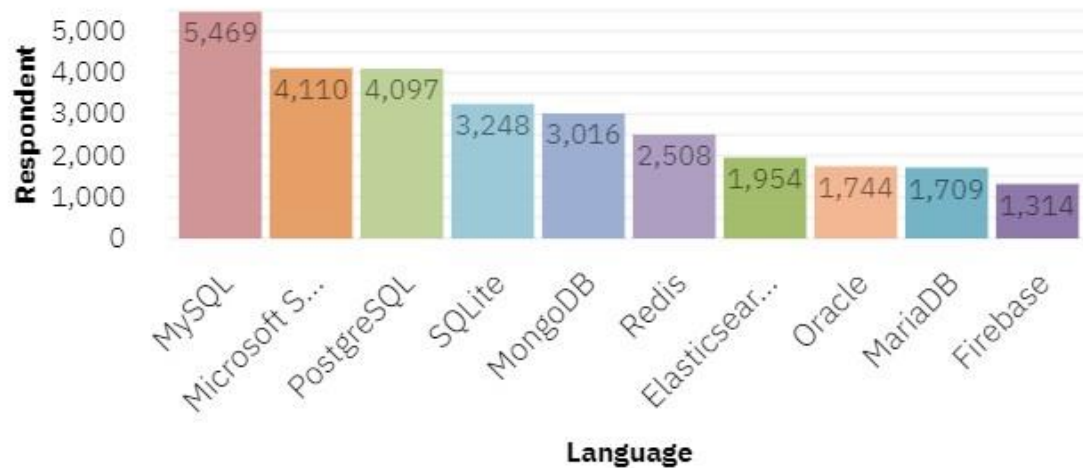
Top 10 Languages Desired to Learn Next Year



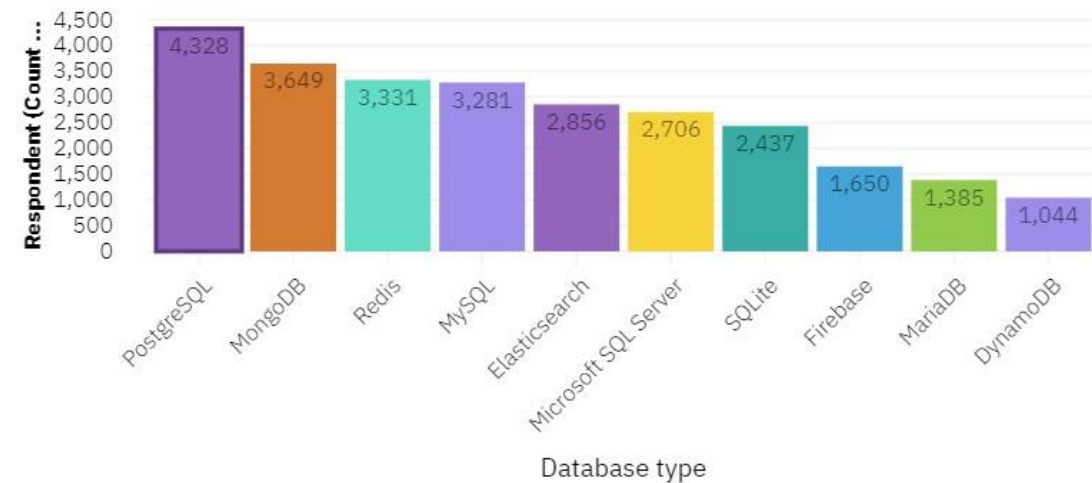
The top 2 programming languages used today are JavaScript and HTML/CSS which are also the top 2 programming languages programmers are looking to learn next year. SQL, Bash/PowerShell and Python are also ranked high as a currently used language and desired future language.

DATABASE TRENDS

Top 10 Databases Worked With

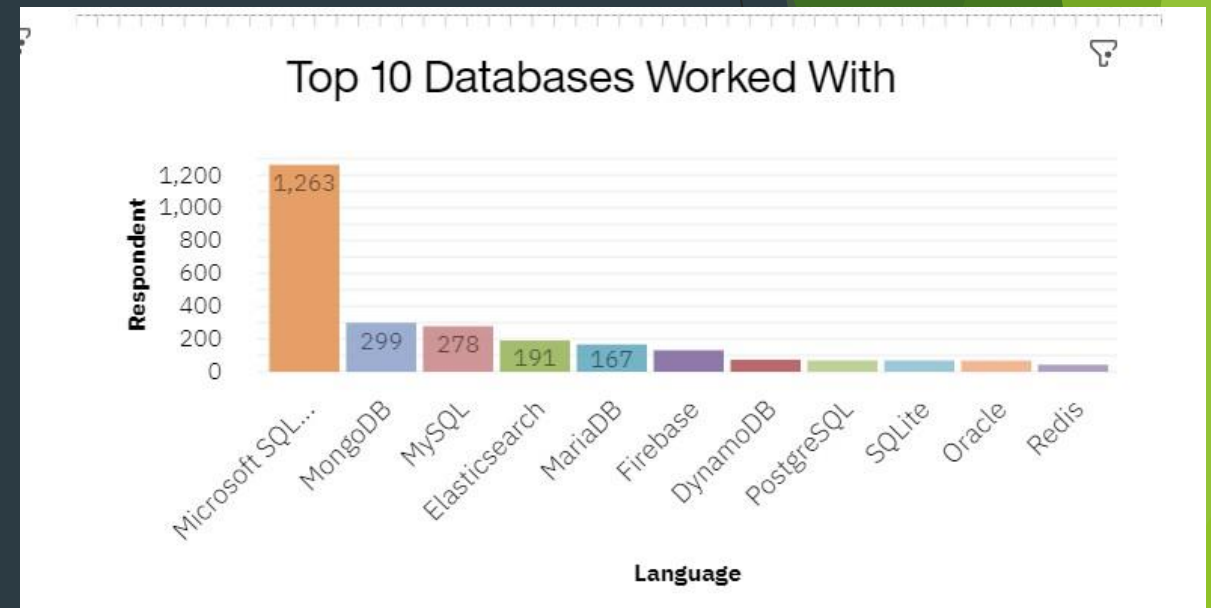
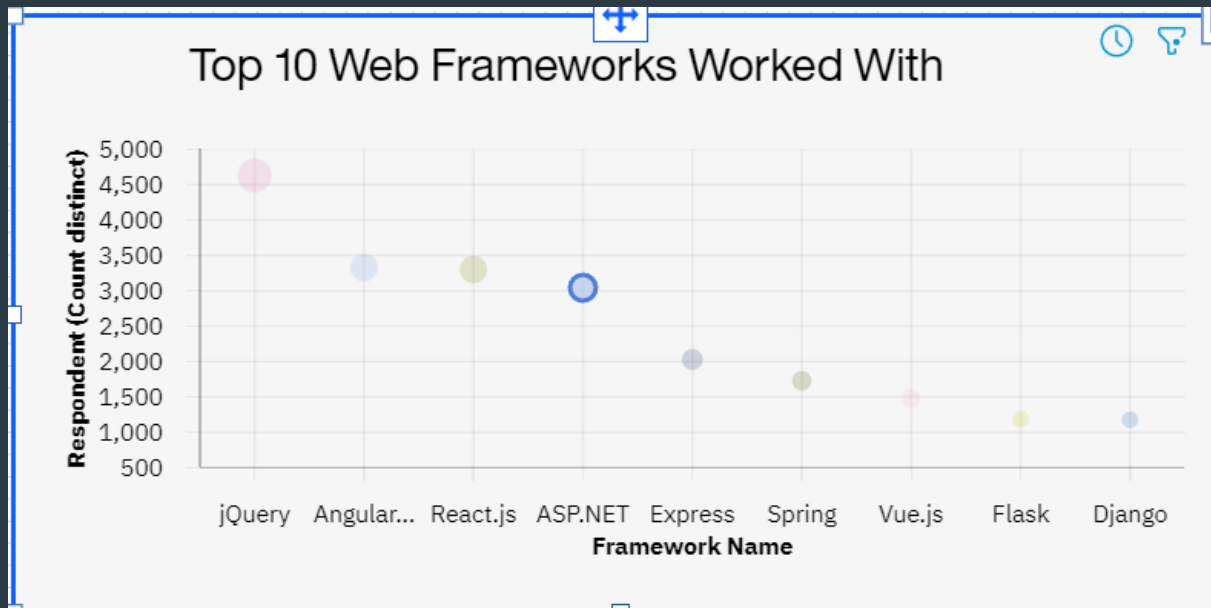


Top 10 Databases Desired to Learn Next Year



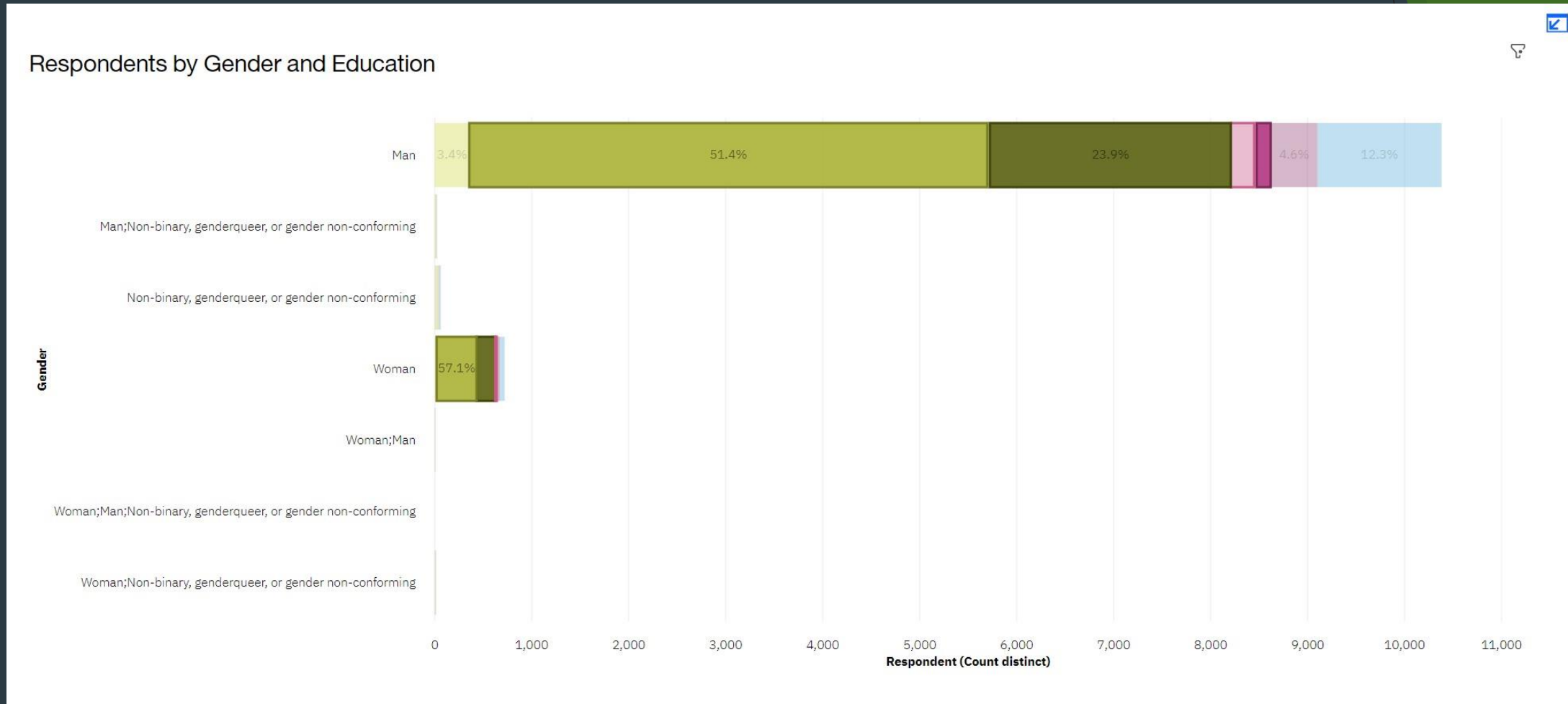
PostgreSQL, MongoDB and Redis are be gaining popularity over current databases worked with, but MySQL is the most worked with database.

DATABASE TRENDS



We can also see that databases worked with are correlated strongly with certain web frameworks. For example ASP.net framework users work with Microsoft SQL Server Databases mainly.

Education Levels of Programmers



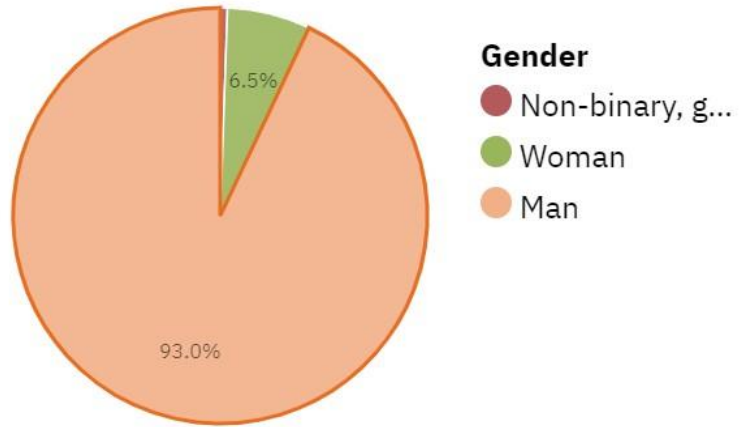
With $\frac{3}{4}$ of Survey respondents holding a 4 year college degree we can see programming is a highly educated field, but a college degree is not a requirement.

DASHBOARD

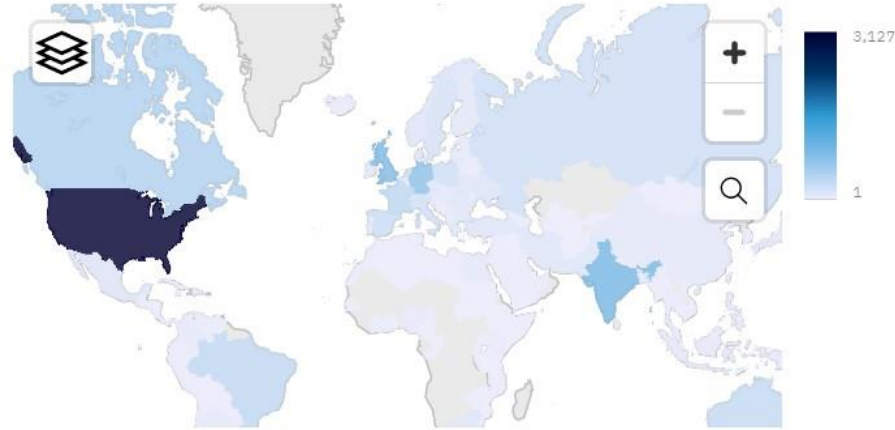
<https://eu-de.dataplatform.cloud.ibm.com/dashboards/e98a244d-8e6e-4ad8-aced-31d2cdc55338/view/0164a12122b5239651d4c8e407ca2b057b31245be0bbd152d2867b490b317197a93a14c7c82a1f5ddc100635a2ee1b50c8>

DASHBOARD TAB 1

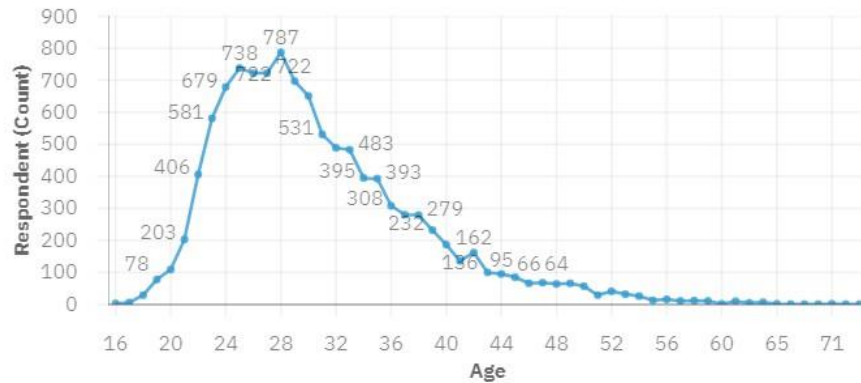
Respondents by Gender



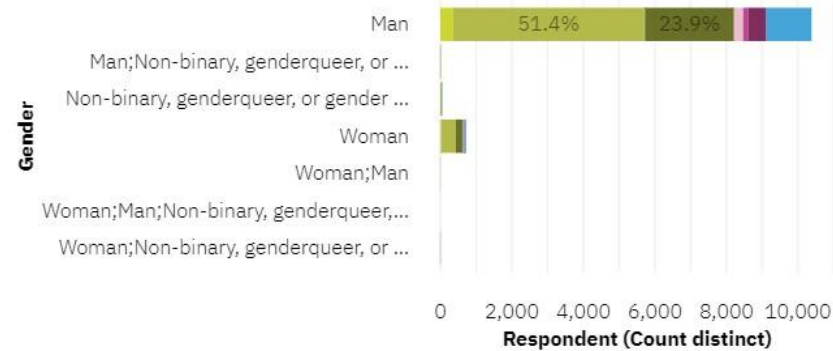
Respondents by Country



Respondents by Age

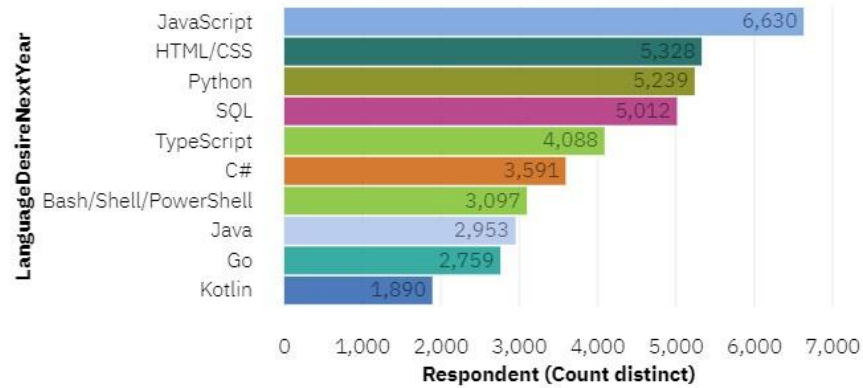


Respondents by Gender and Education

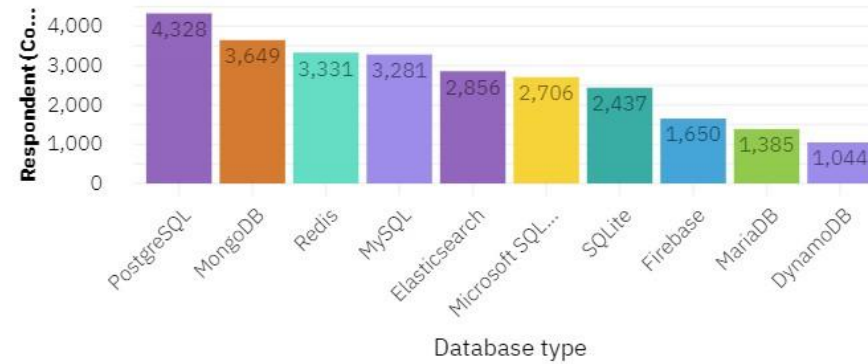


DASHBOARD TAB 2

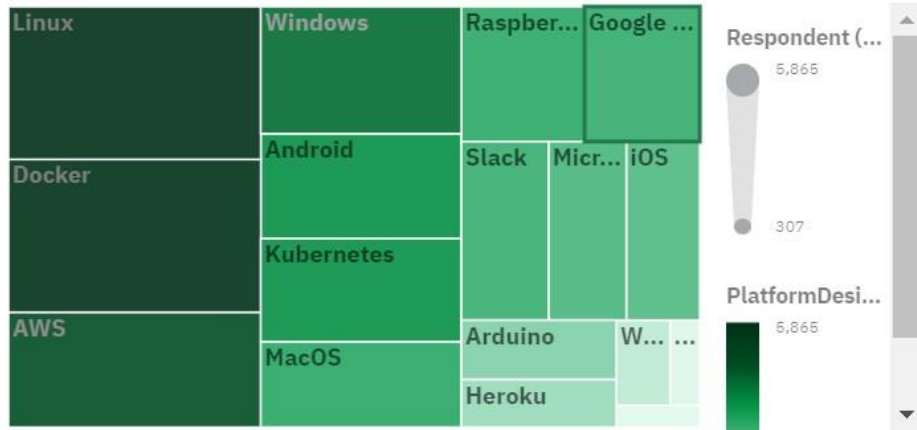
Top 10 Languages Desired to Learn Next Year



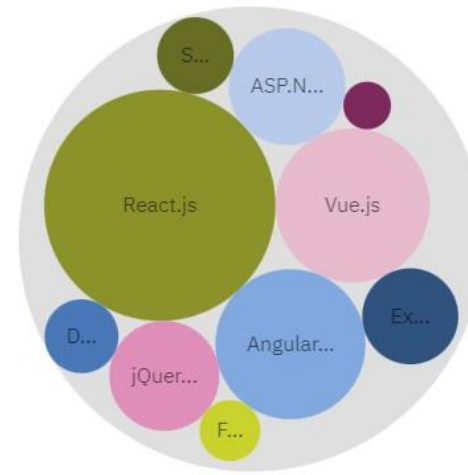
Top 10 Databases Desired to Learn Next Year



Tree Map of Desired Platforms to Learn

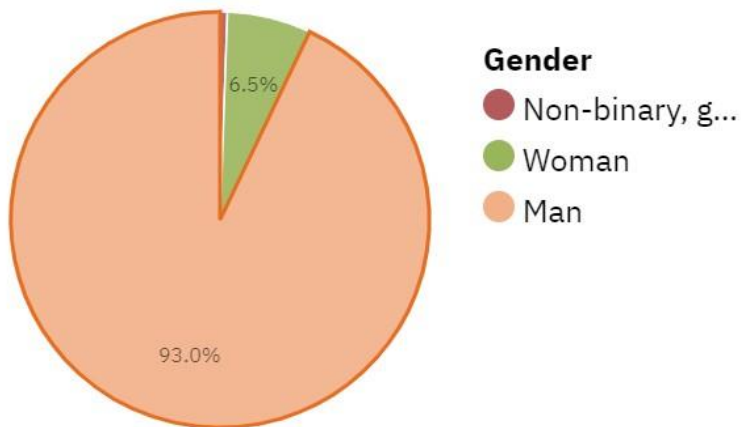


Top 10 WebFrames Desired to Learn

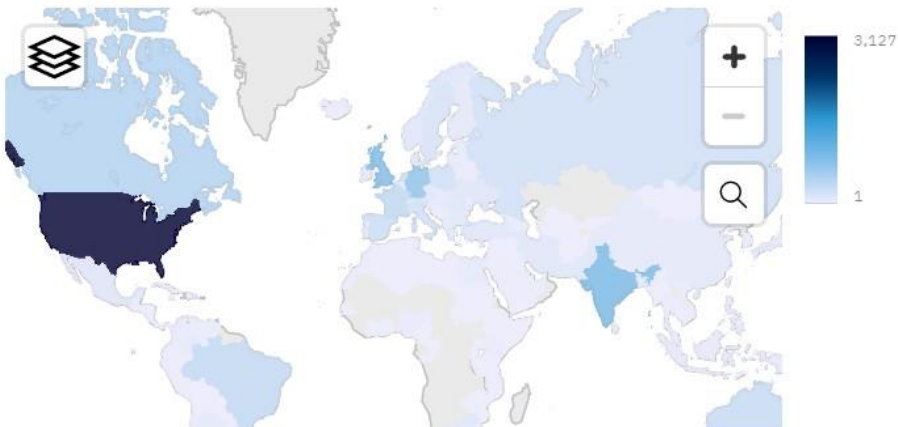


DASHBOARD TAB 3

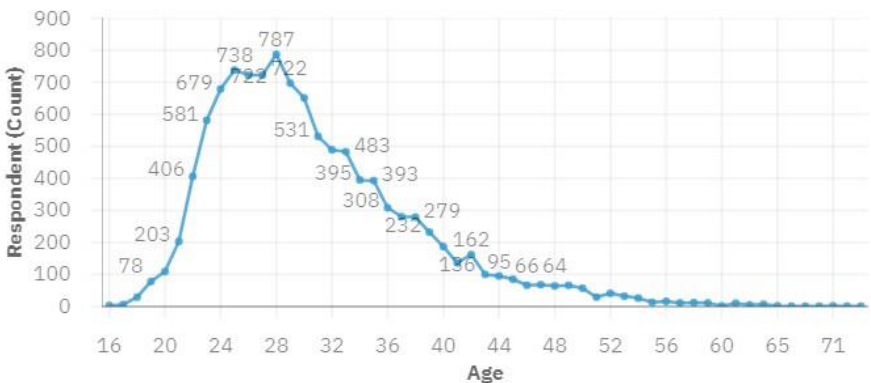
Respondents by Gender



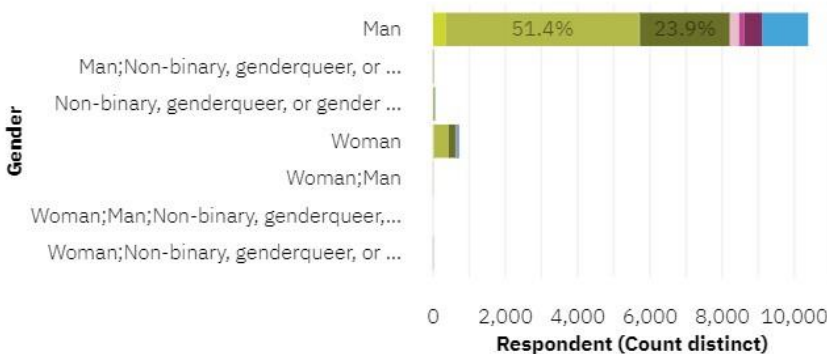
Respondents by Country



Respondents by Age



Respondents by Gender and Education



DISCUSSION

- Stack Overflow is used around the world and has users in 135 countries but the majority of its users are in the United States. Survey respondents show this weight to be largely from the U.S.. This impacts the results of the dataset as it applies to a global audience.



- What reasons are there for so few female programmers?
- what can we do to get more women programming?

OVERALL FINDINGS & IMPLICATIONS

Findings

- Over 90% of programmers are male, with respondents having an average age of 30 overall.
- 3/4s of programmers having a 4 year university degree or higher
- JavaScript and HTML/CSS are the most popular languages learned and desired to learn.
- PostgreSQL is the most popular Database Language to learn and web frameworks correlate to databases

Implications

- More women are needed in the programming fields to increase representation and diversity in the programming field.
- Smart people are programmers! And if you are programming you probably love to learn, since most programmers are long time learners.
- If you were planning on learning a new language JavaScript and HTML/CSS would probably be a safe bet to get a job with since so many people choose to learn them.

CONCLUSION



- ❑ Women are very under-represented in the programming world.
- ❑ If you are programming you likely have a 4 year degree but you don't have to have any degree, there are programmers with no formal education at all.
- ❑ JavaScript and HTML/CSS are very popular programming languages and will be into the future.
- ❑ There is shifting preferences from MySQL to Postgre SQL for database usage.

APPENDIX



```
In [66]: df1 = pd.read_csv (r'm5_survey_data_demographics.csv')  
df =pd.DataFrame(df1)  
df['Age'].mean()
```

```
Out[66]: 30.779137791377913
```

```
In [76]: df1 = pd.read_csv (r'm5_survey_data_demographics.csv')  
df =pd.DataFrame(df1)  
df['Country'].nunique()
```

```
Out[76]: 135
```

APPENDIX



```
In [79]: df1 = pd.read_csv (r'm5_survey_data_demographics.csv')
df = pd.DataFrame(df1)
df['Ethnicity'].value_counts()/sum(df['Ethnicity'].value_counts())

Out[79]: White or of European descent
0.686095
South Asian
0.101744
Hispanic or Latino/Latina
0.048214
East Asian
0.039261
Middle Eastern
0.028630

...
East Asian;Middle Eastern;South Asian
0.000093
Black or of African descent;Hispanic or Latino/Latina;Native American, Pacific Islander, or Indigenous Australian;Multiracial
0.000093
Hispanic or Latino/Latina;South Asian;Multiracial
0.000093
East Asian;Hispanic or Latino/Latina;White or of European descent
0.000093
Hispanic or Latino/Latina;Native American, Pacific Islander, or Indigenous Australian;South Asian;White or of European descent;
Multiracial 0.000093
Name: Ethnicity, Length: 89, dtype: float64
```

The dataset used comes from the following
[source: https://stackoverflow.blog/2019/04/09/the-2019-stack-overflow-developer-survey-results-are-in/](https://stackoverflow.blog/2019/04/09/the-2019-stack-overflow-developer-survey-results-are-in/) under a [ODbL: Open Database License](#).

The dataset is a modified subset of the original survey results.