



Current usage and trends in Data Science

Carolyn Harris

April 2024

OUTLINE



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

EXECUTIVE SUMMARY



- More data means the need for more data analysts
- Most data analysts are fluent in multiple languages
- With the advent of AI, data analysts must constantly advance their knowledge and skillset
- There are several indicators to explore to assist both companies and various stakeholders as well as data scientists to be competitive in their respective fields
- We discuss what languages, databases and other skills will help both data analysts and stakeholders to stay current and cutting edge.

INTRODUCTION



- While older languages are still being used, others are coming along and revolutionizing data analysis
- As data accumulates and data warehouses keep growing, these newer languages and database management systems are adapting
- In order to stay competitive in the marketplace, stakeholders want to ensure they are getting the best analysis and hiring the best analysts.
- We will look technology trends in 3 areas:
 - Languages
 - Databases and
 - Analysts

METHODOLOGY



- Dataset was a survey on the website Stack Overflow: <https://stackoverflow.blog/2019/04/09/the-2019-stack-overflow-developer-survey-results-are-in/> under an [ODbL: Open Database License](#).
 - A subset of the data was used, preprocessed and randomized
 - Responses were used to determine current usage as well as future trends
- Dataset: <https://www.kaggle.com/promptcloud/jobs-on-naukricom> under the under a **Public Domain license**.
 - Via API data was collated by languages were most used and number of jobs using these languages in key US cities
 - We explored the correlation between languages and annual salaries as well as locations with the most jobs of that kind

RESULTS

As more and more data are collected and used to drive decision-making, as well as that data becoming ever-more inter-connected, it is unsurprising that older languages are still utilized, and newer languages are used to access that data not only analysis but also creation of visualizations so the results are instantly understood by stakeholders.

Currently, the top 2 languages have to do with web and mobile apps. The trend however is pointing towards being able to access and manipulate the data, communicating directly with the database, such as with the language SQL to interact with the current top types of SQL databases. The trend for databases, likewise, appears to broaden from SQL to some others, such as MongoDB.

In order to remain competitive in the industry, data analysts would benefit from investing time and effort to building their repertoire of various languages.


In order to better understand current conditions and trends data analysts must be well versed in various methods of data collection, and various languages to analyze that data and give a superior analysis work product.

PROGRAMMING LANGUAGE TRENDS

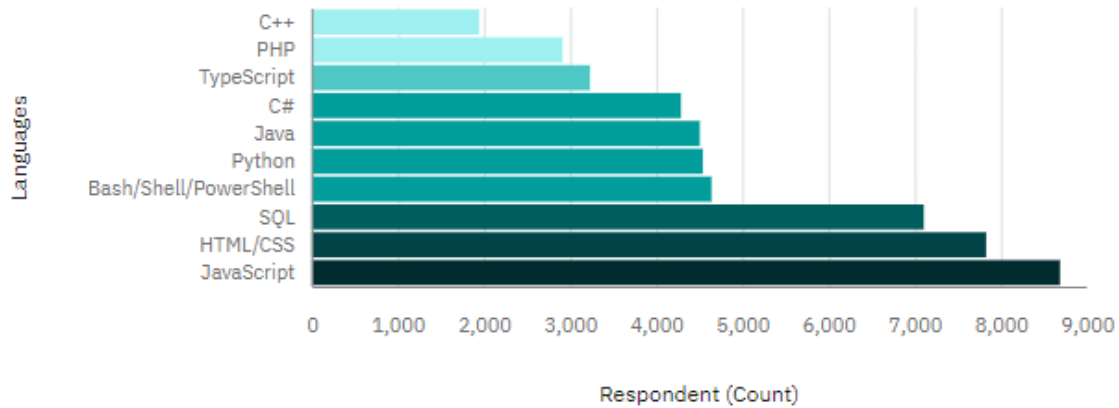
Current Year

Top 10 Programming Languages

Respondent (Count)




1,946 8,687



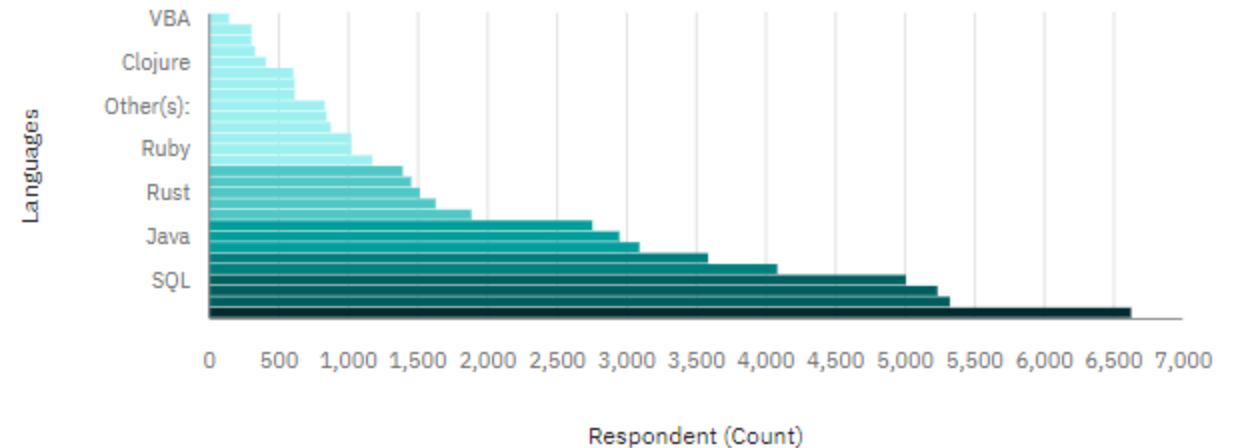
Next Year

Top 10 Languages to Learn in 2025

Respondent (Count)



149 6,630



PROGRAMMING LANGUAGE TRENDS

FINDINGS & IMPLICATIONS

Findings

- Current most used languages are for web/mobile apps
- 2nd set of most used languages refer to creating apps and analyzing data
- The most desired languages to learn involve direct interaction with databases.

Implications

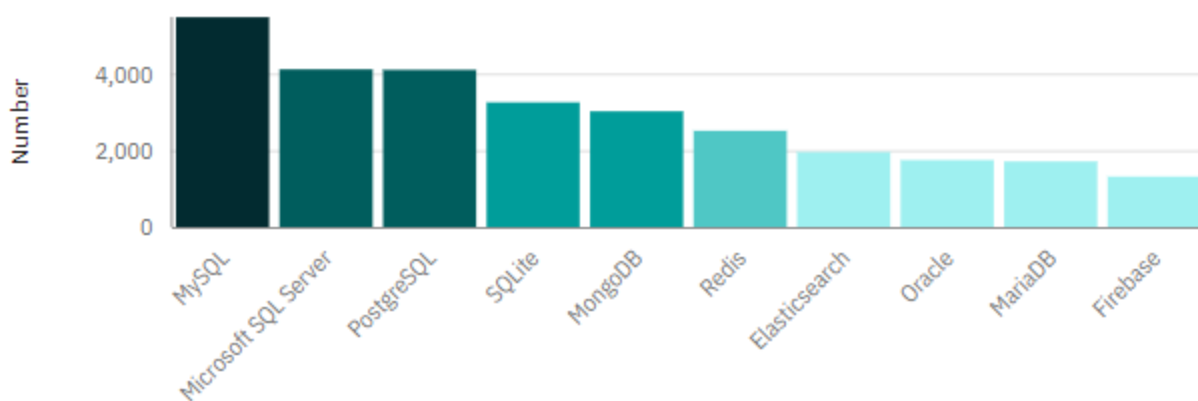
- Consistent creation of applications keep these current languages important
- As the data analysis field grows, knowing several languages is vital
- Trend is towards languages like SQL to interact with databases

DATABASE TRENDS

Current Year

Top 10 Databases

Respondent (Count)



Next Year

Top 10 Databases to use in 2025

Respondent (Count)



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL currently dominates
- SQL most often used
- While there are other database types, they are less used today

Implications

- Knowledge and proficiency with SQL is imperative
- Based on survey responses, there is keen interest in other databases
- While SQL will remain important, MongoDB, Redis and Elasticsearch will become more widely used

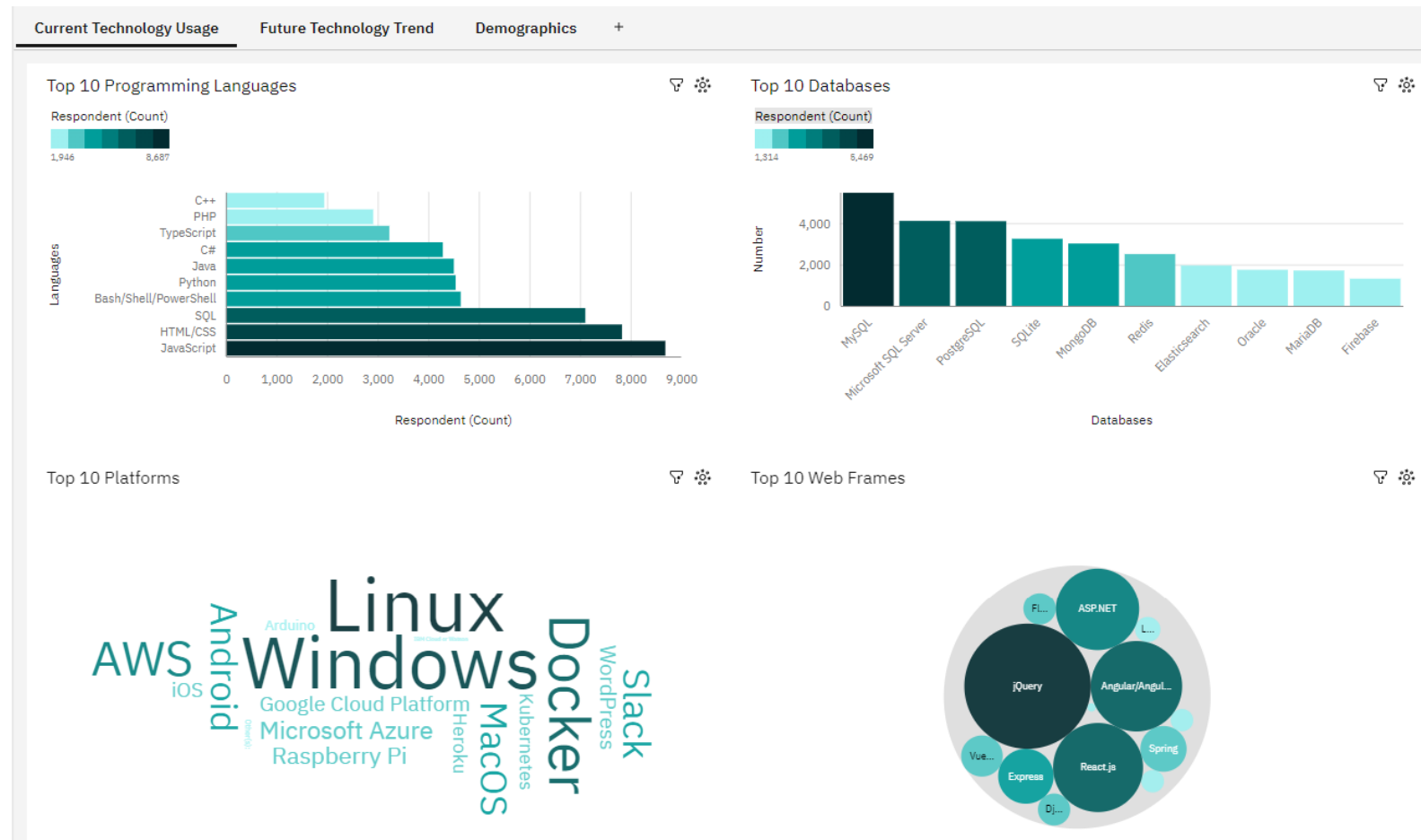
DASHBOARD



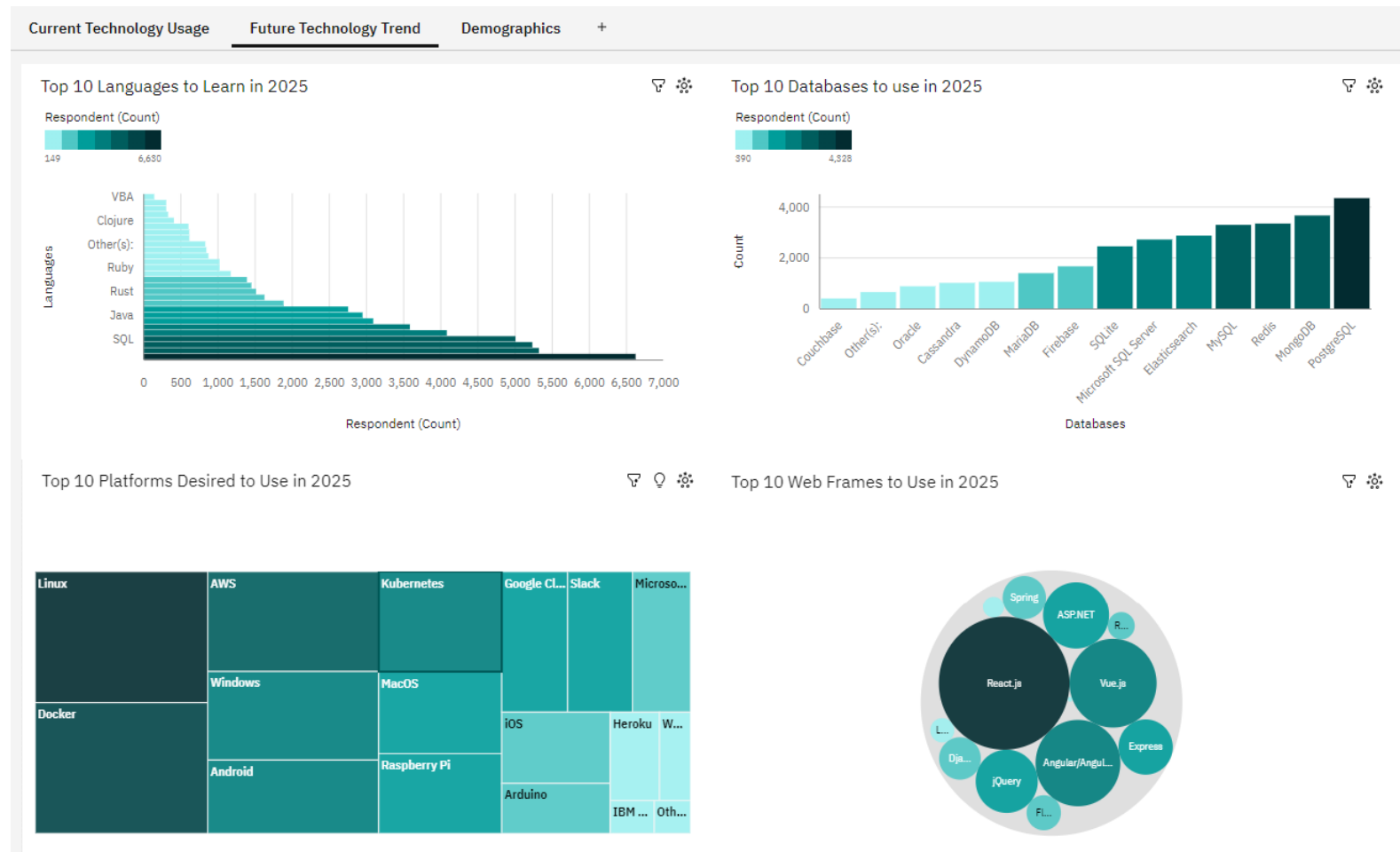
GitHub link to dashboard:

<https://github.com/NutmegAllAlong/IBM-Data-Analysis-Class/blob/b538b1e0c59781875f75e4facb6be8372c55c36f/New%20dashboard.pdf>

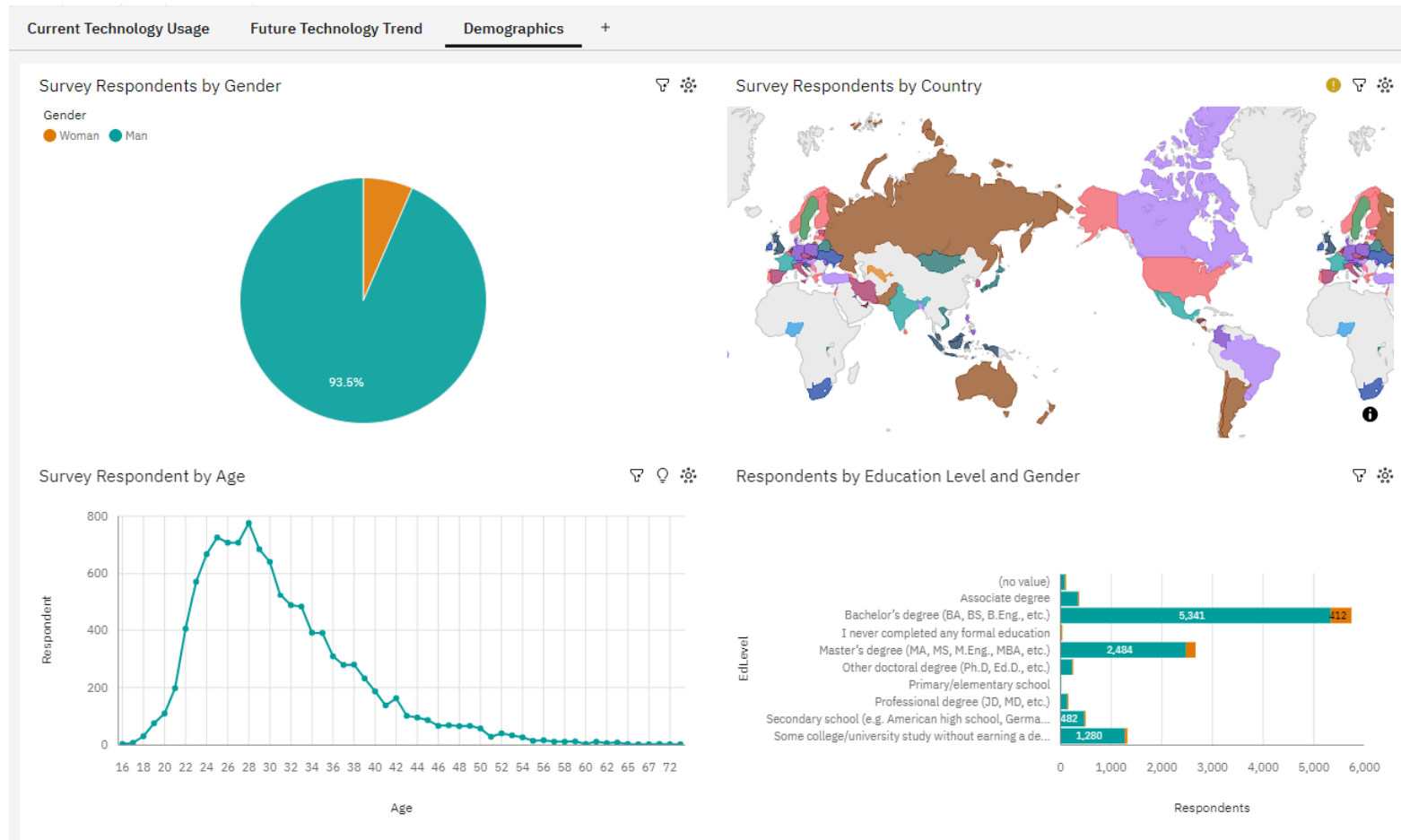
DASHBOARD TAB 1



DASHBOARD TAB 2



DASHBOARD TAB 3



DISCUSSION



- Developers span the globe and as technology is ever-evolving, programmers, developers, and data analysts have an interest in constantly expanding knowledge-base and skillset.
- Paying attention to events, new architecture, trends, and the broader community allows data science to progress.

OVERALL FINDINGS & IMPLICATIONS

Findings

- Majority of respondents were male
- Majority of respondents in the West
- Majority of respondents between 20-35

Implications

- Majority of data science will continue this pattern, and more females are joining this field
- Focus on mathematics, science and logical thinking will enable more people to join this field outside the West.
- This generation grew up with a level of technology never known before so this trend will continue.

CONCLUSION



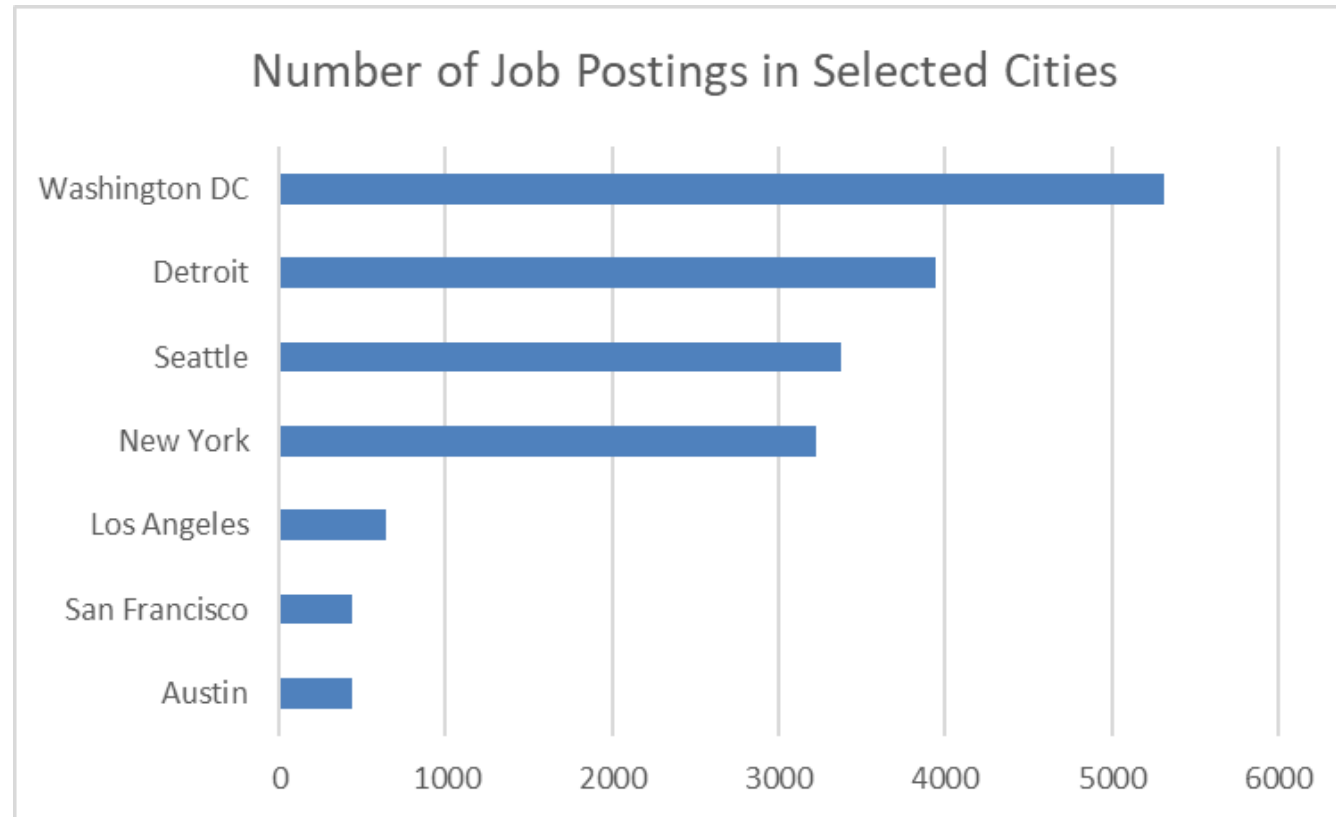
- Those in data science must be in a constant state of continuing education.
- It is necessary to be well-rounded in skillset of languages, web frames and databases.
- Increasingly connected world is allowing greater participation in this field worldwide.
- Current demographics in this field (sex, age, education) is likely to continue.

APPENDIX



- Fig 1 – Number of job postings in US cities for data science jobs
- Fig 2 – Average Annual Salary by Language

JOB POSTINGS



POPULAR LANGUAGES

