

Most Valued Data Science Skills

Nfn Tenzin Dakar

New Oo Mon (Nina)

Crystal Quezada



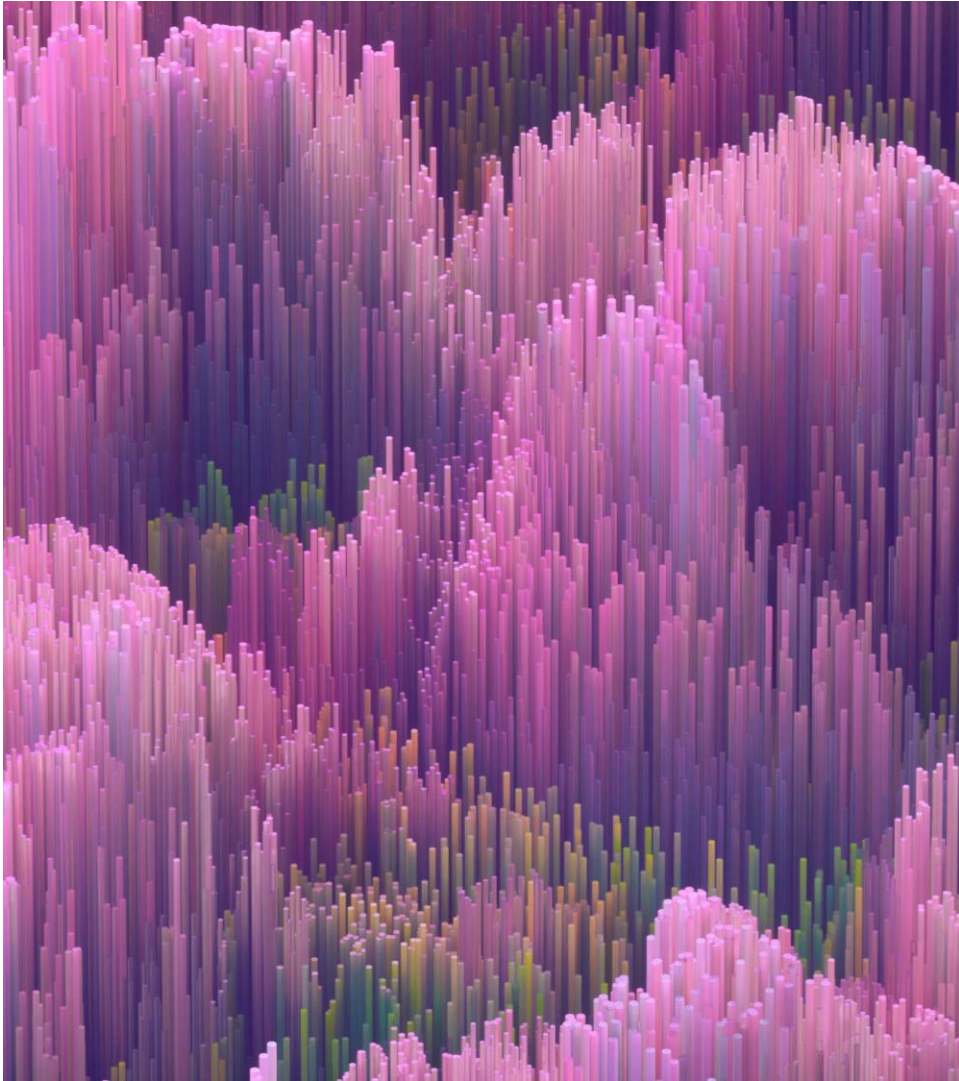
Agenda

Part 1 - Introduction

Part 2 - Our Data

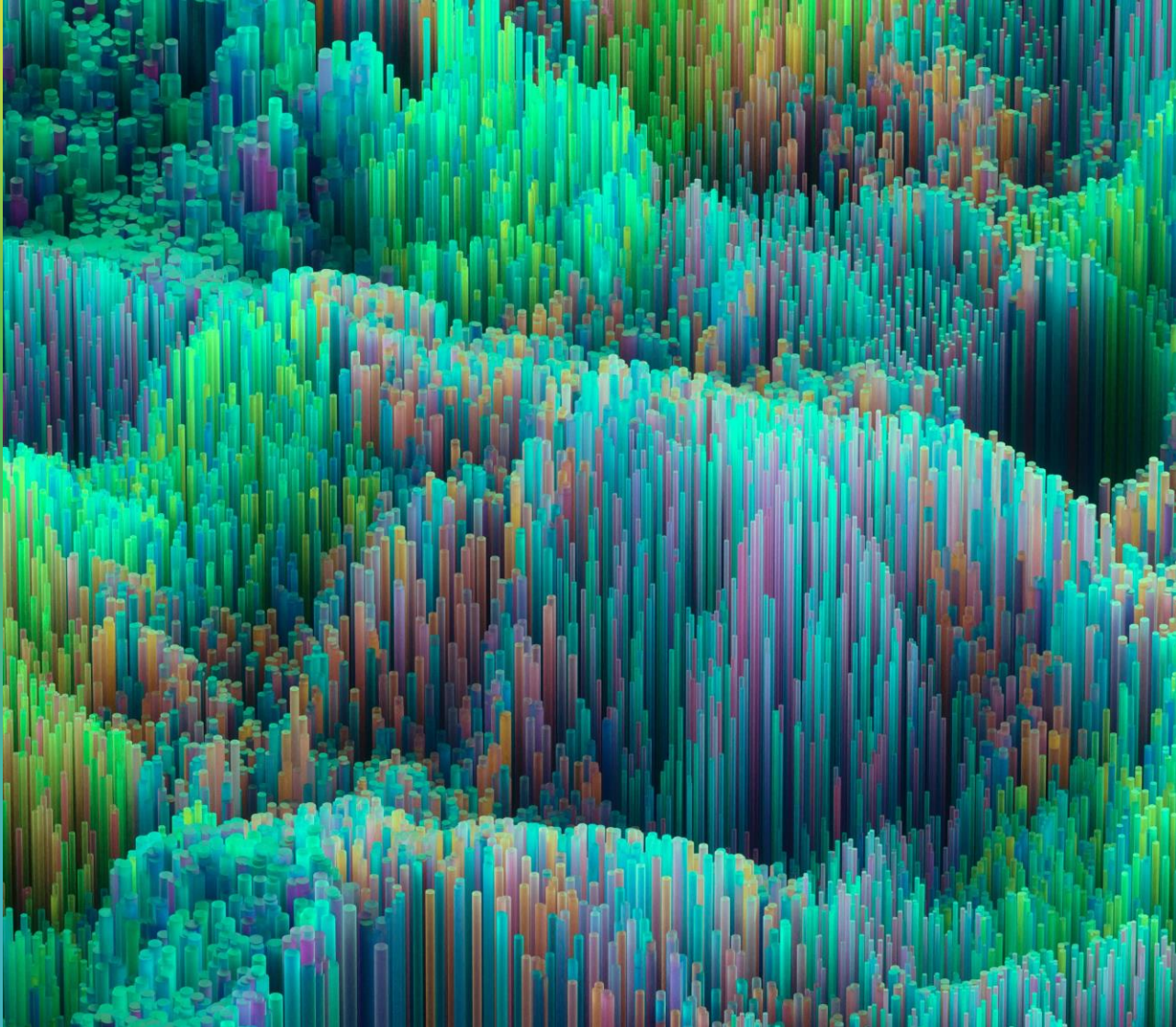
Part 3 - Data Analysis

Part 4 - Conclusion



Introduction

- + We are DATA DOMINATION. For this project, Zoom and text messages were our main source of communication. We also used RStudio Cloud with R Markdown published on RPubS. Every file was shared in a single GitHub repo. Finally, we've created an entity-relationship diagram (ERD) using Mysql DataBase.

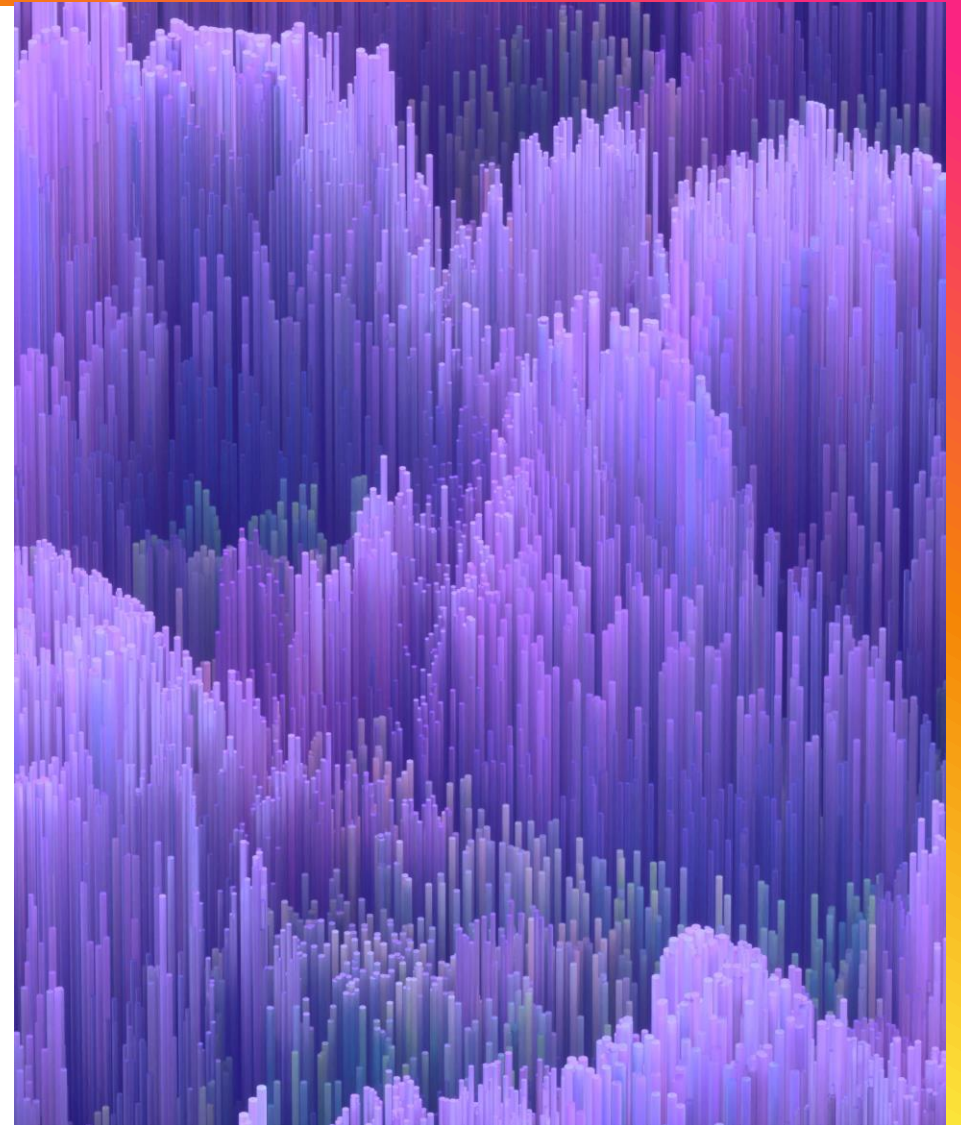
An abstract 3D visualization consisting of a dense field of vertical bars of varying heights and colors. The colors transition from dark blue and purple on the left to bright green and yellow on the right. The bars are arranged in a way that creates a sense of depth and movement, resembling a data landscape or a topographical map. The background is a gradient from yellow at the top to blue at the bottom.

What is the most valued data science skills?

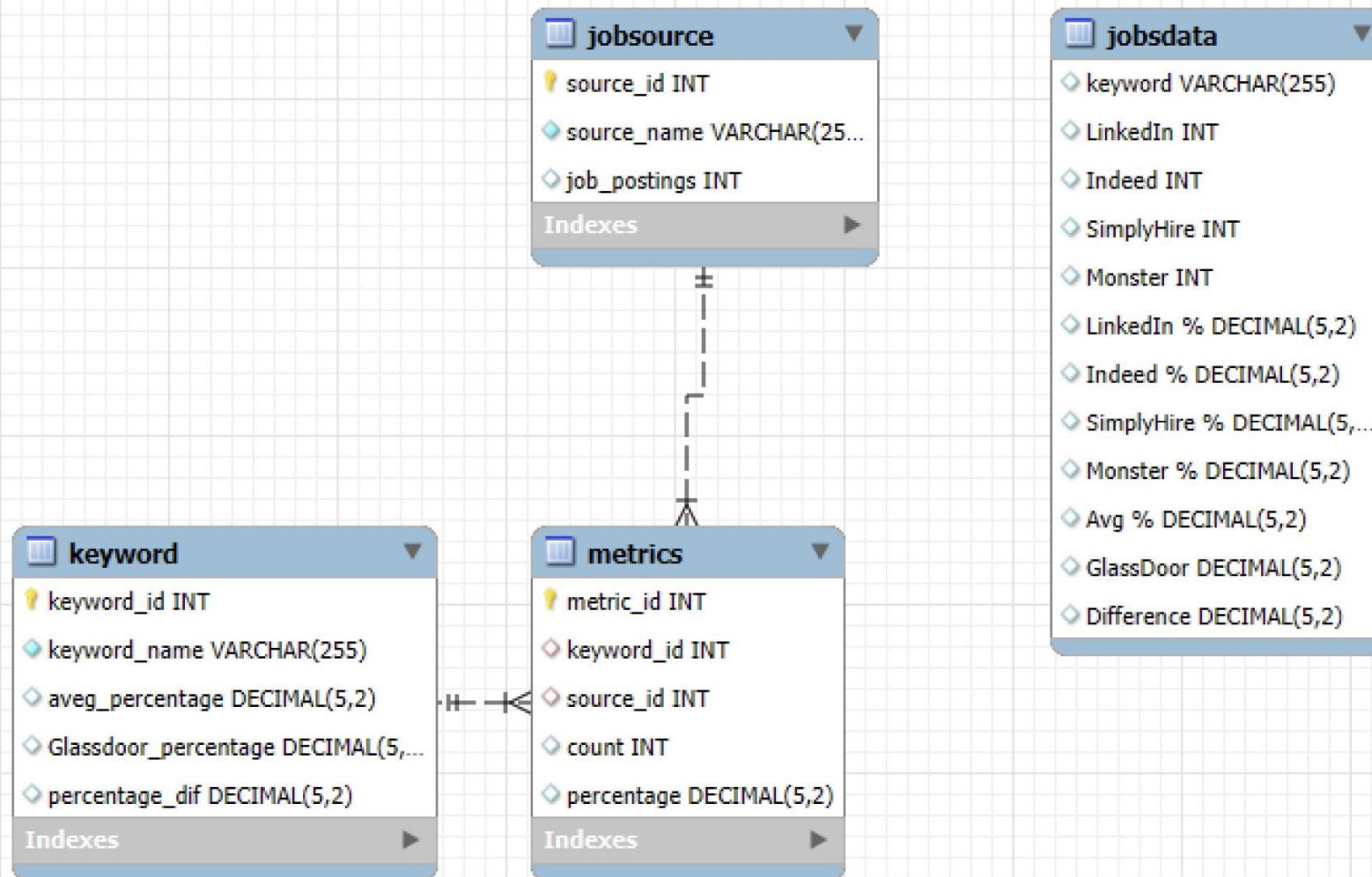
What better way to answer this question than to determine which skills were most in demand by job sources?

Our Data

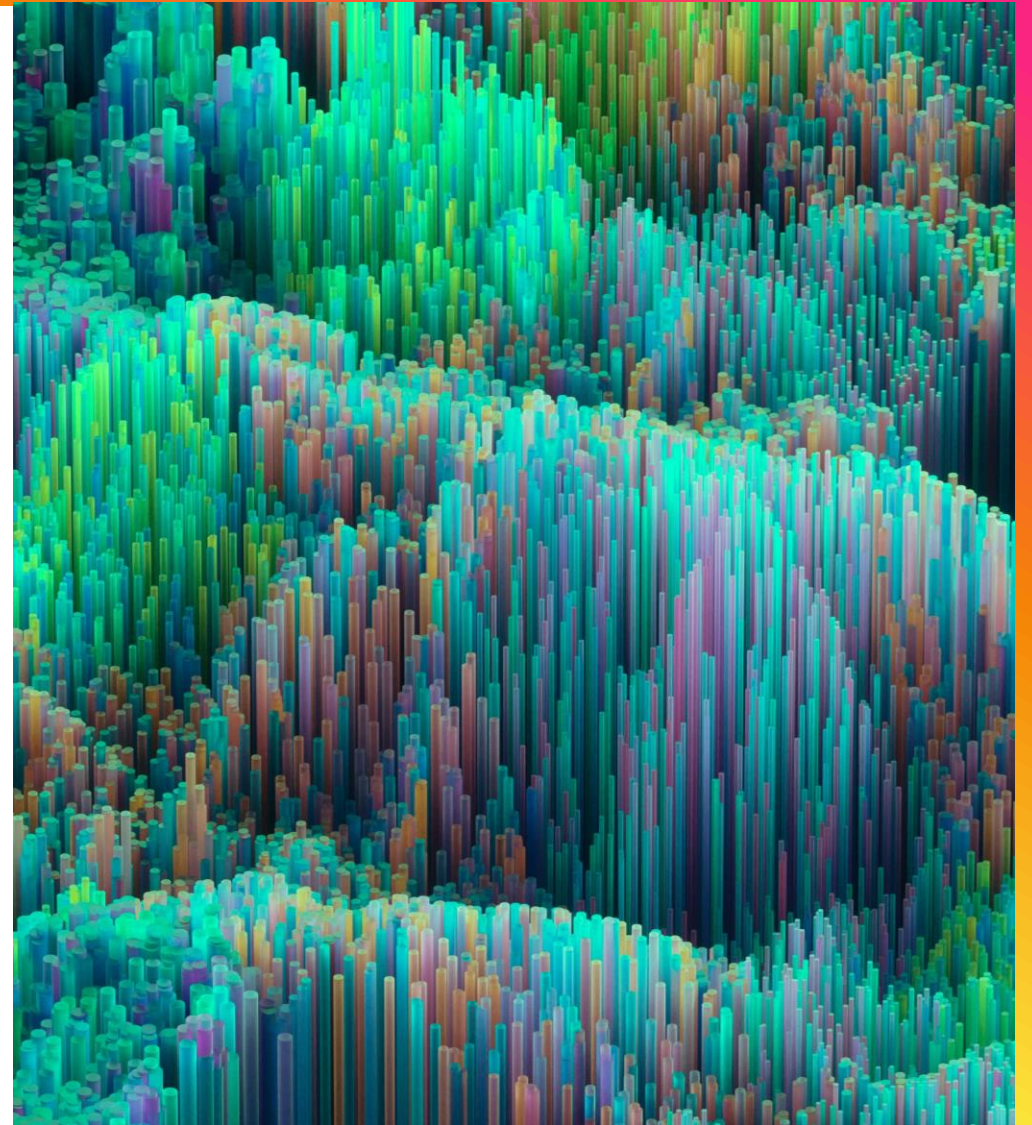
Our dataset reviews job sources, such as LinkedIn, Monster, Indeed, and SimplyHired. The "keywords" are the data skills needed for these hiring sites.



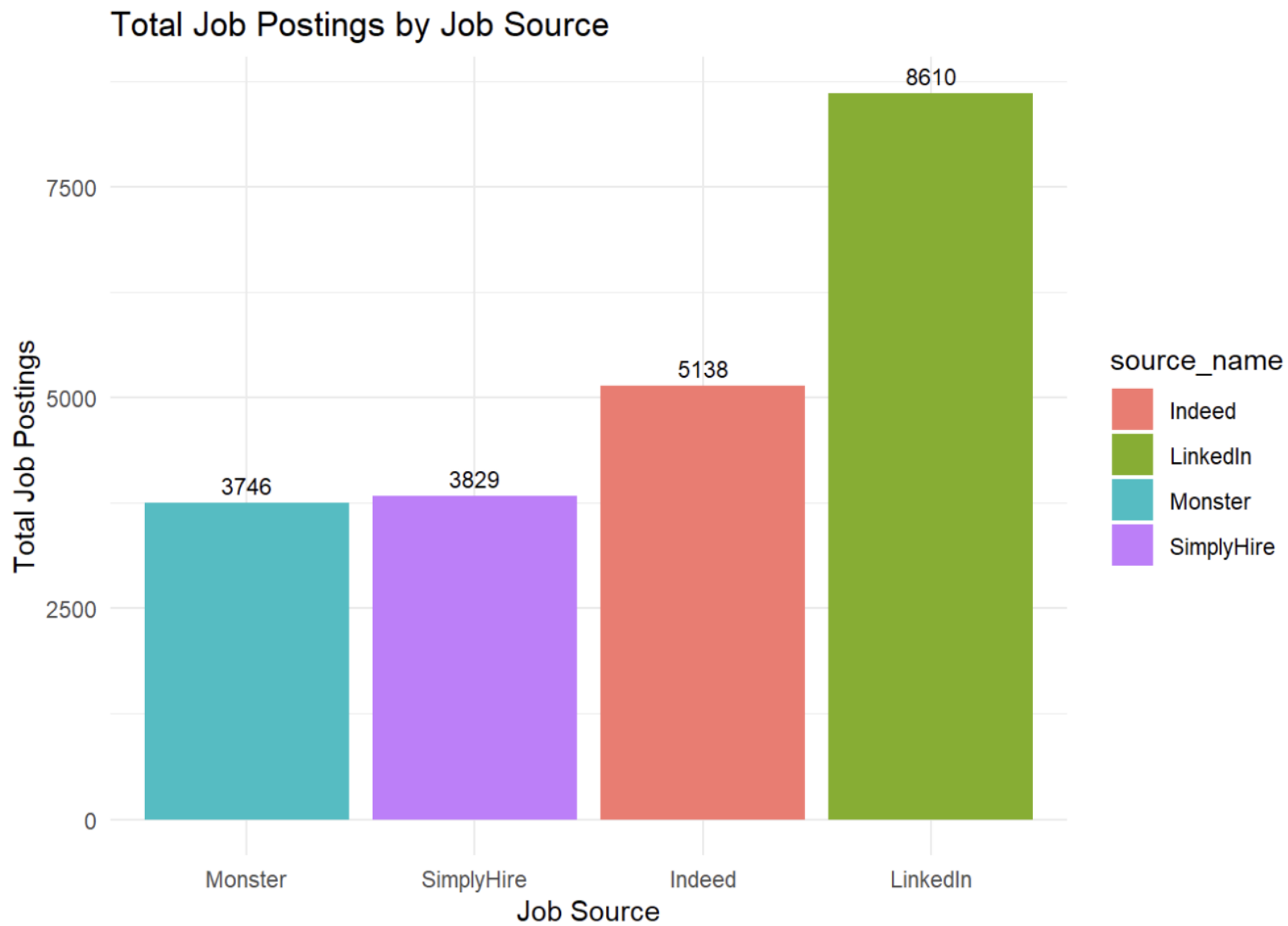
Keyword	LinkedIn	Indeed	SimplyHired	Monster	LinkedIn %	Indeed %	SimplyHired %	Monster %	Avg %	GlassDoor Self Reported % 2017	Difference
Python	6,347	3,818	2,888	2,544	74%	74%	75%	68%	73%	72%	1%
R	4,553	3,106	2,393	2,365	53%	60%	62%	63%	60%	64%	-4%
SQL	3,879	2,628	2,056	1,841	45%	51%	54%	49%	50%	51%	-1%
Spark	2,169	1,551	1,167	1,062	25%	30%	30%	28%	29%	27%	2%
Hadoop	2,142	1,578	1,164	1,200	25%	31%	30%	32%	30%	39%	-9%
Java	1,944	1,377	1,059	1,002	23%	27%	28%	27%	26%	33%	-7%
SAS	1,713	1,134	910	978	20%	22%	24%	26%	23%	30%	-7%
Tableau	1,216	1,012	780	744	14%	20%	20%	20%	19%	14%	5%
Hive	1,182	830	637	619	14%	16%	17%	17%	16%	17%	-1%
Scala	1,040	739	589	520	12%	14%	15%	14%	14%		



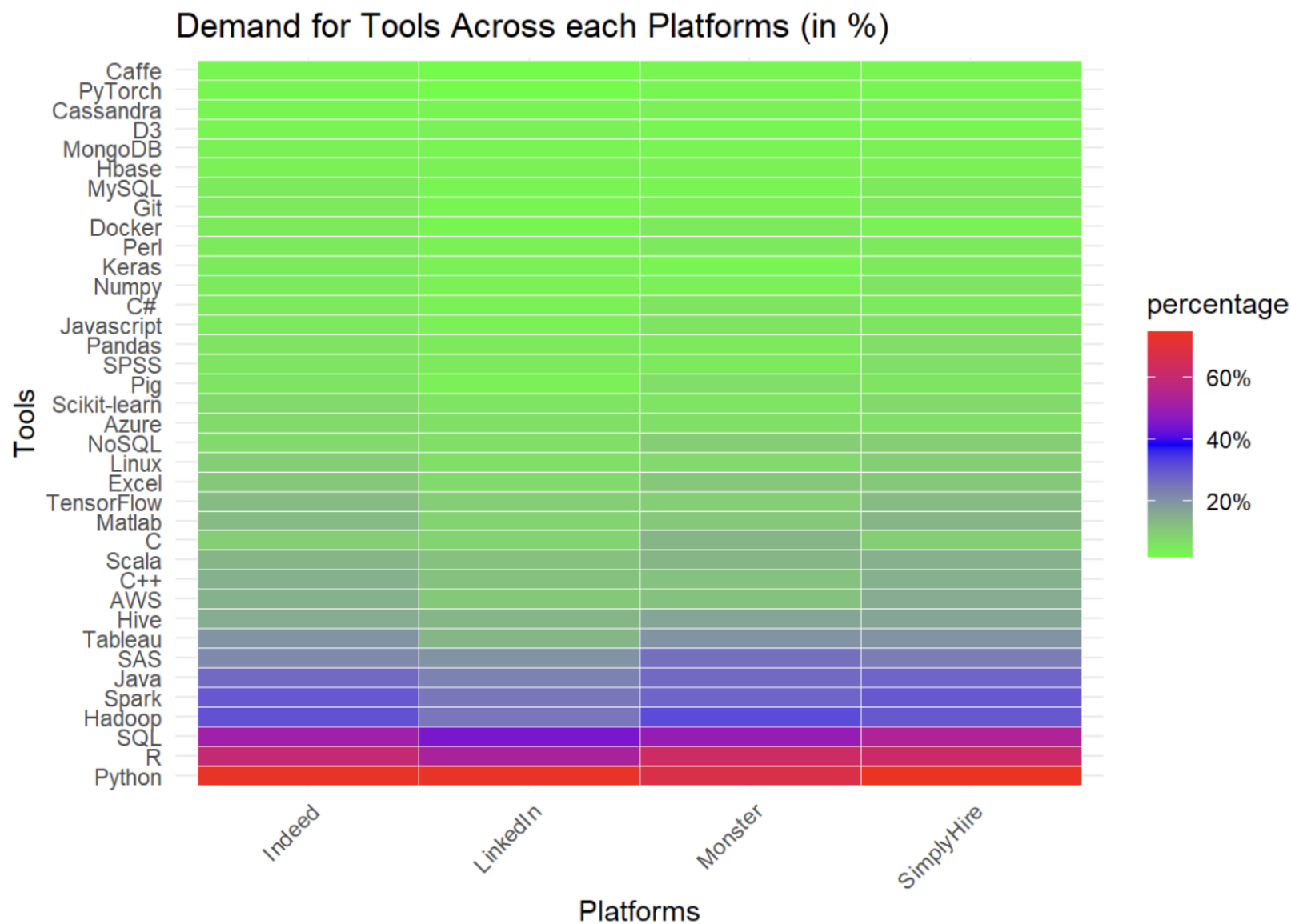
Data Analysis



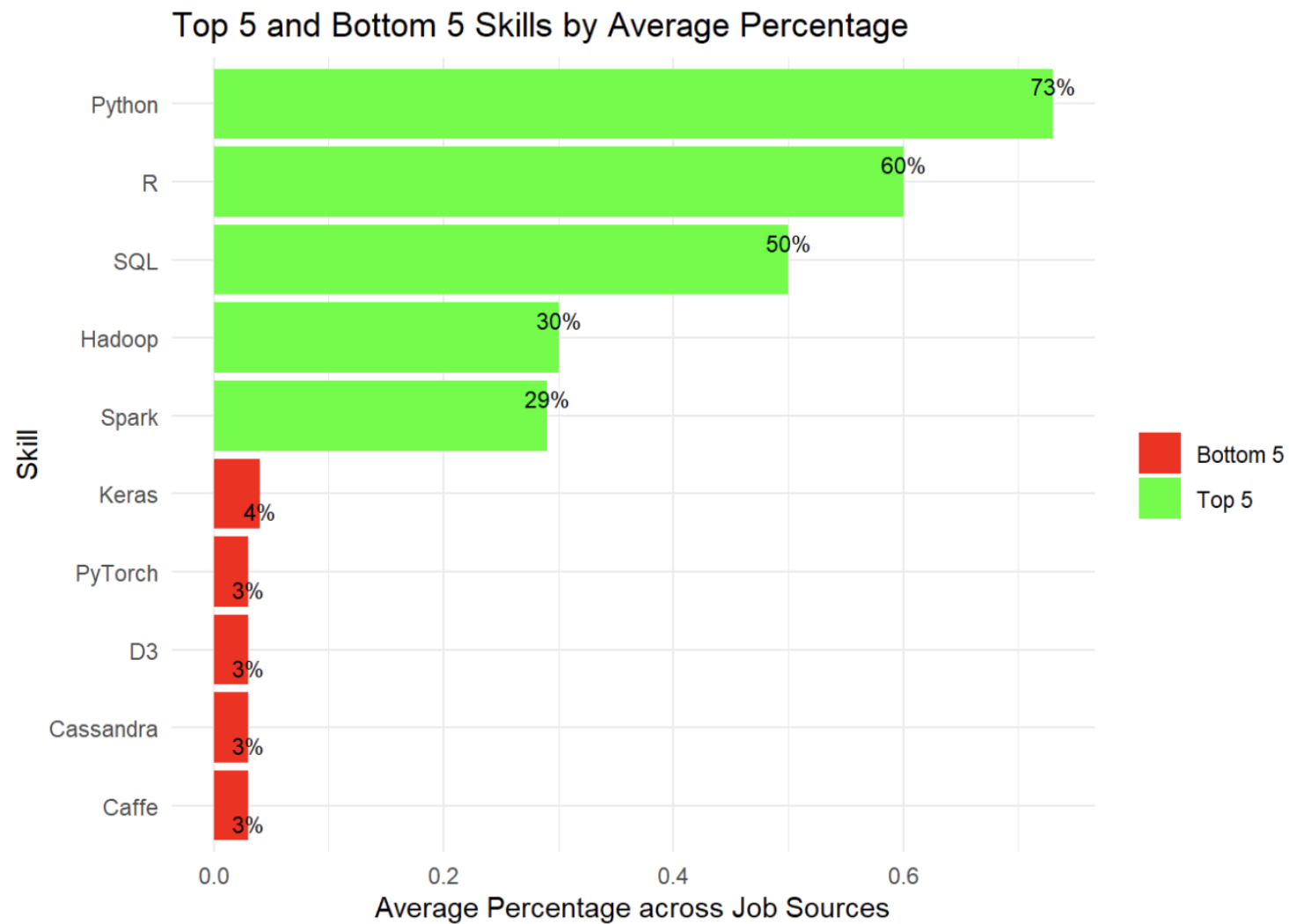
Plot 1



Plot 2

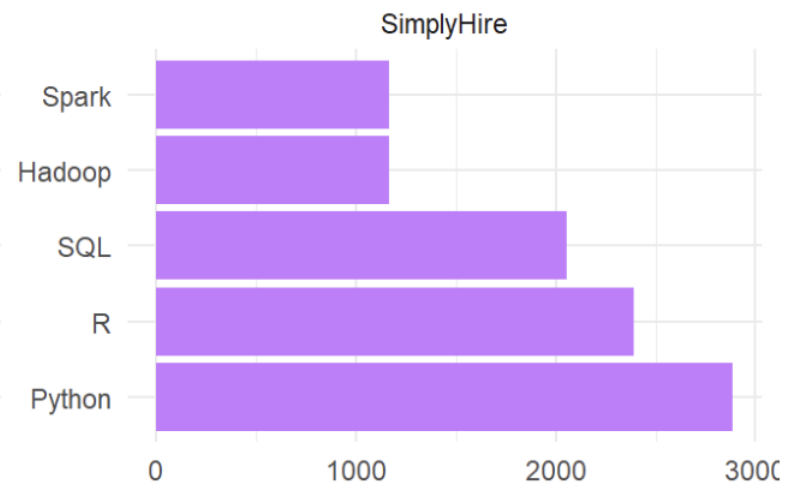
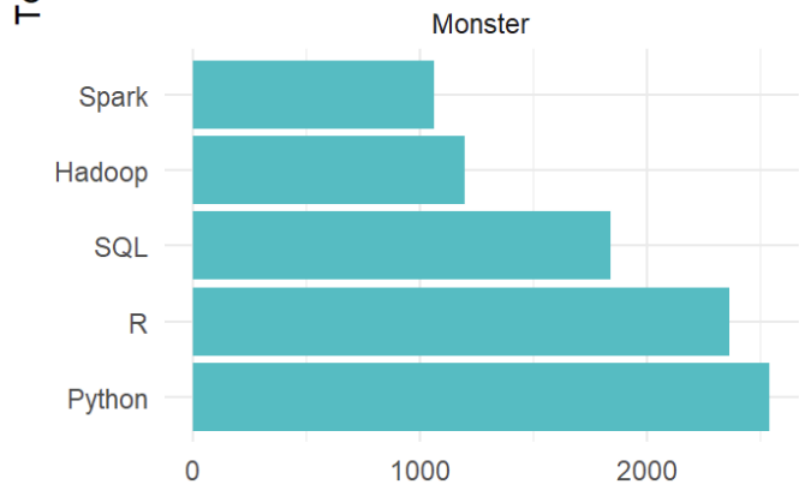
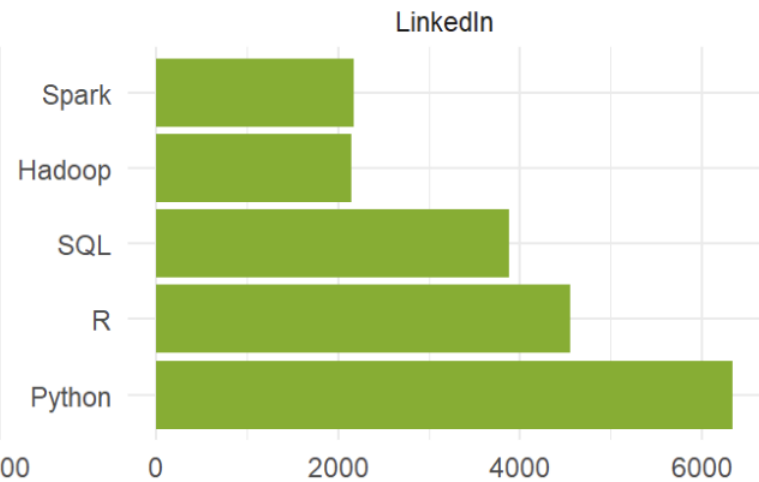
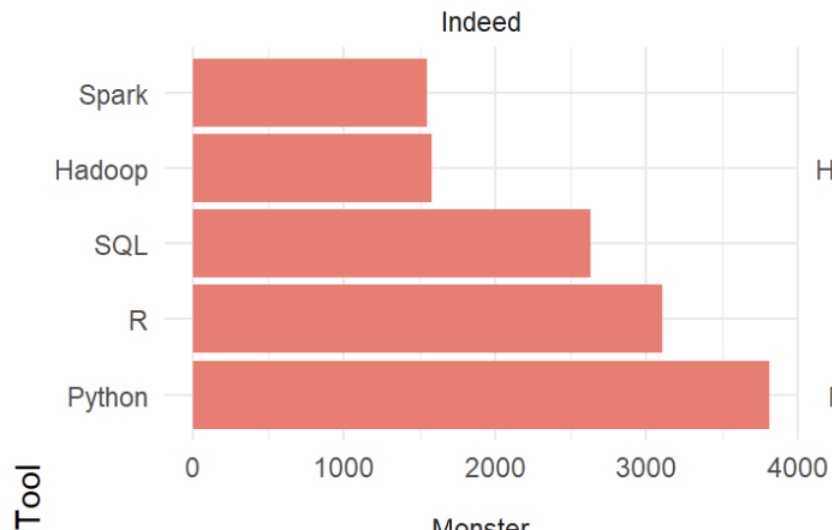


Plot 3



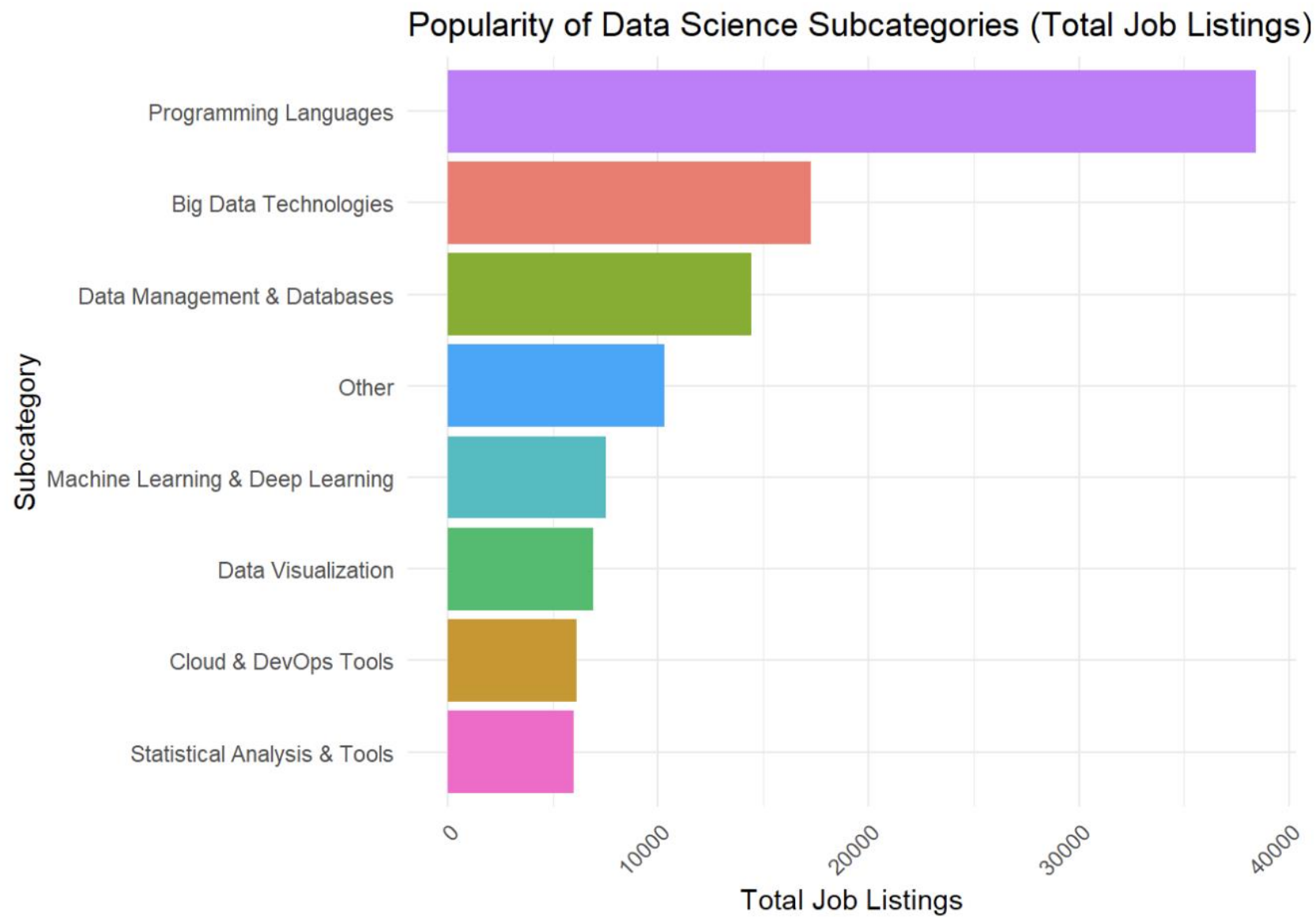
Plot 4

Top Tools by Platform

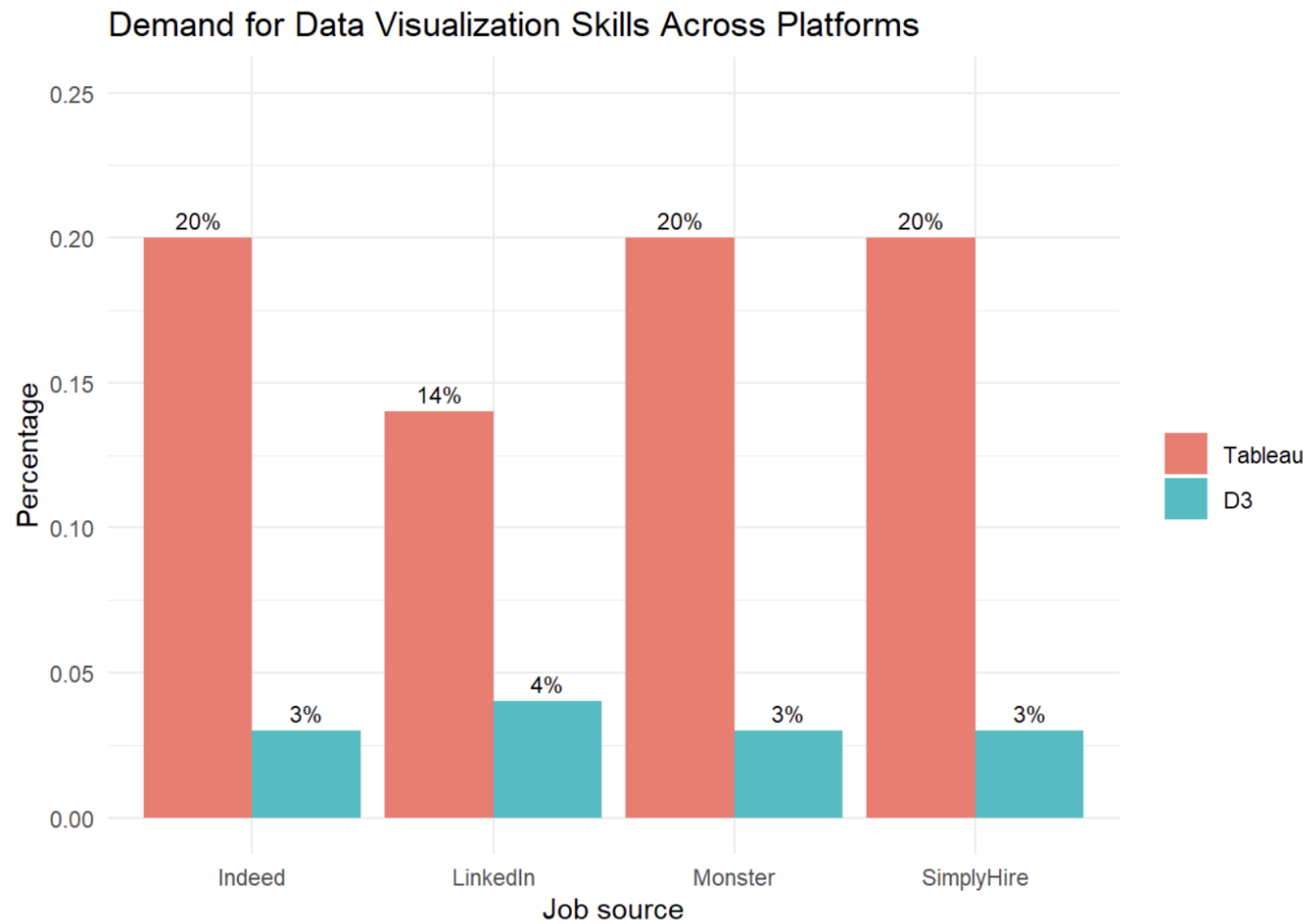


Count

Plot 5

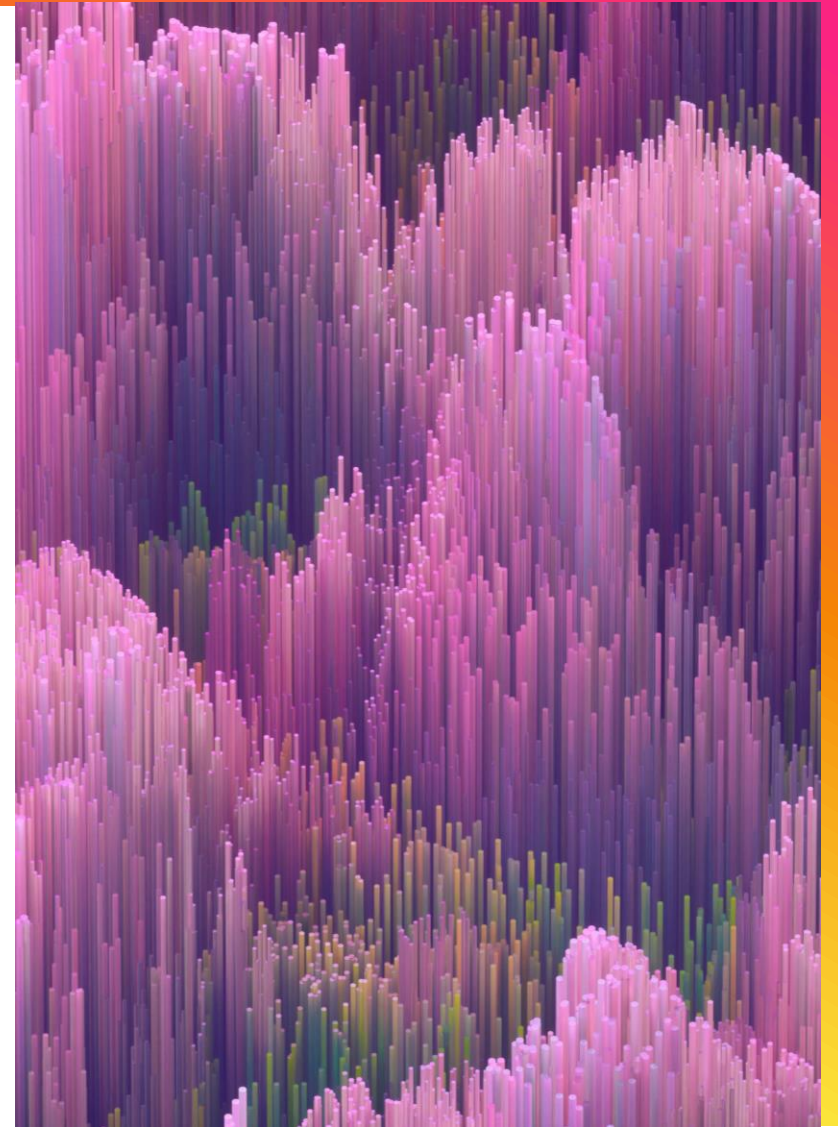


Plot 6



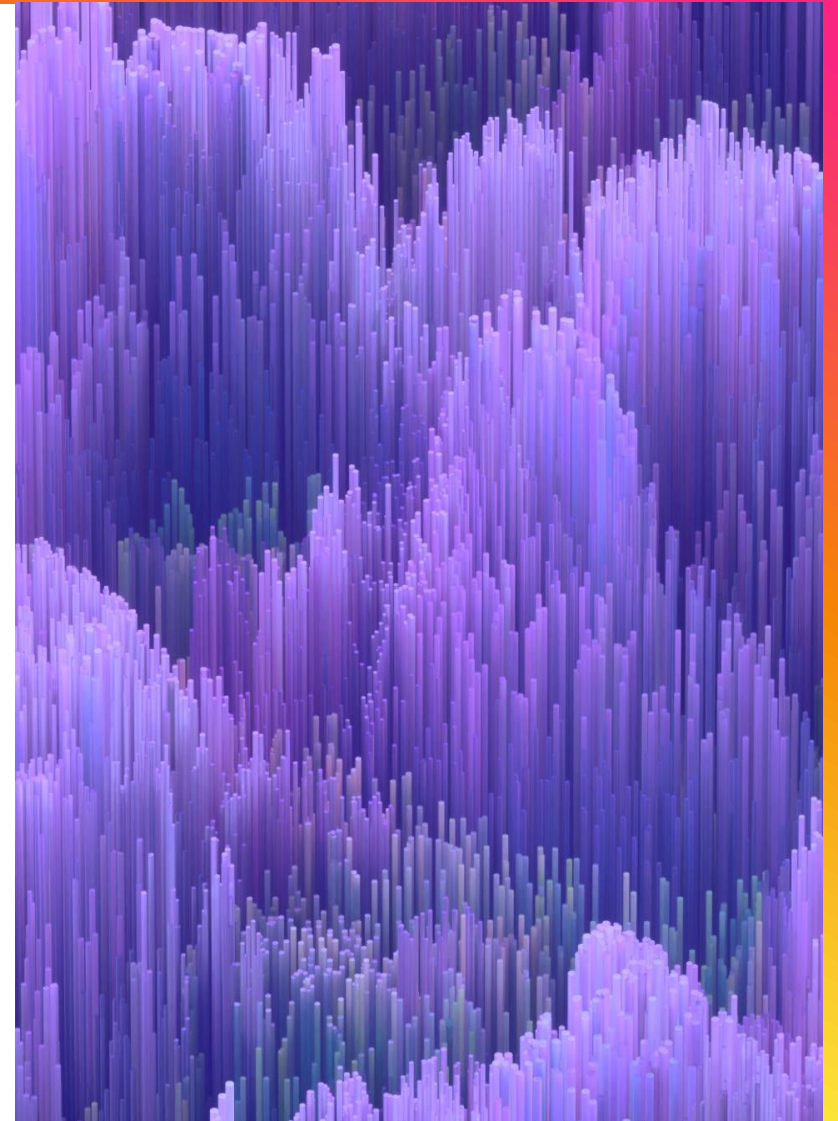
Conclusion

- The analysis shows that Python, R, and SQL are the most in-demand tools across the four major job platforms (SimplyHired, Indeed, LinkedIn, and Monster).
- Python consistently ranks as the top skill, followed closely by R and SQL, indicating strong demand for these core programming languages in data science roles.
- The higher counts observed on LinkedIn suggest that it may be a key platform for job seekers in the data science field.



Conclusion cont.

- Meanwhile, big data tools like Hadoop and Spark are also popular but have fewer listings compared to general-purpose programming languages.
- This trend highlights the importance of Python, R, and SQL in the current job market, making them critical skills for aspiring data scientists to master.
- The job market shows the highest demand for programming languages and big data technologies, while Tableau leads in data visualization skills.





Thank you!

Nfn Tenzin Dakar, New Oo Mon (Nina), Crystal Quezada