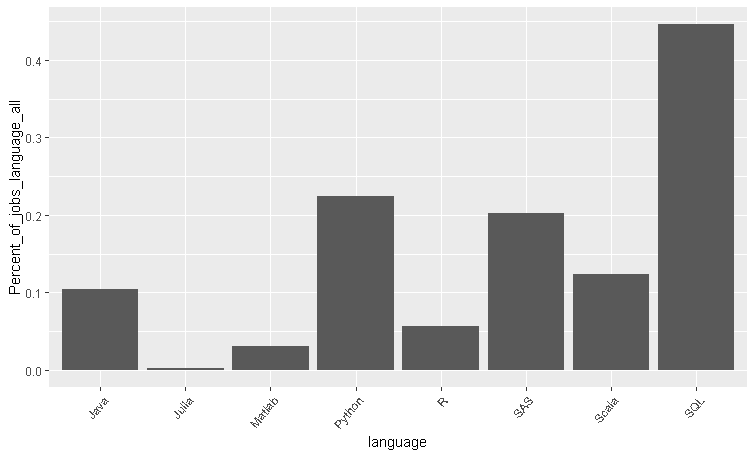
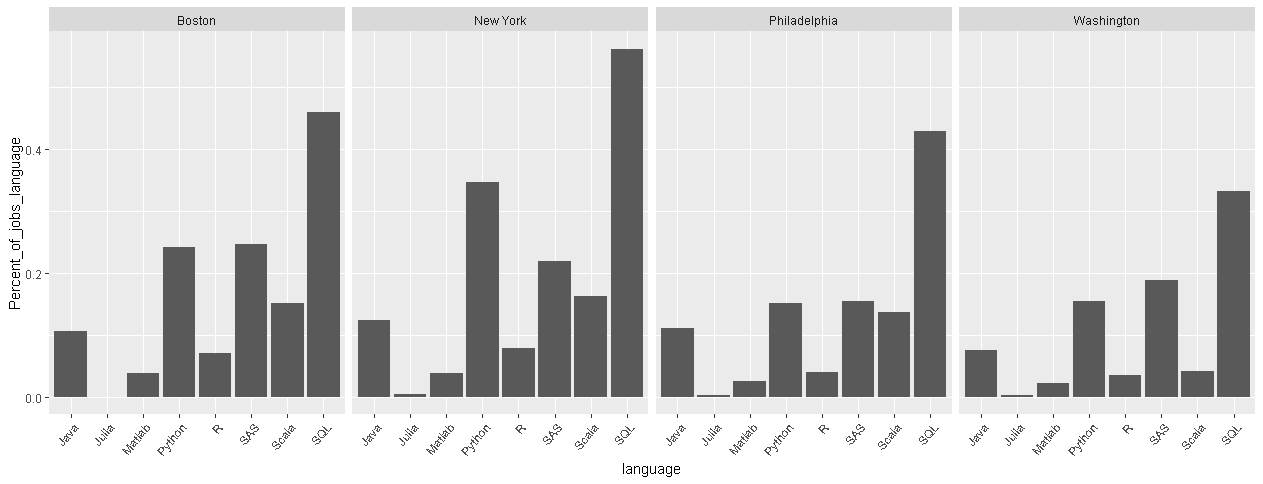
Tanner Sax

Capstone Project - Statistics

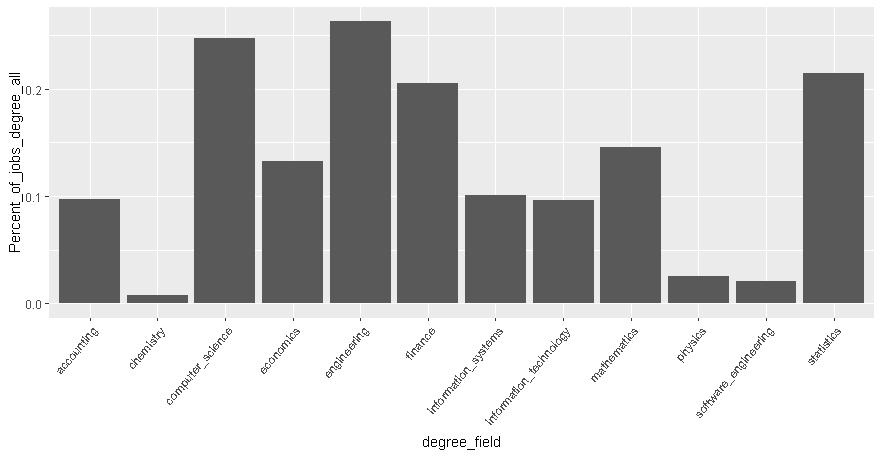
SQL is the most important programming language for data analysts in all four cities, appearing in 44.6 percent of job postings. This is to be expected for a data analyst job. Julia, a very new programming language, is mentioned the least amount of times at 0.28 percent, as expected.



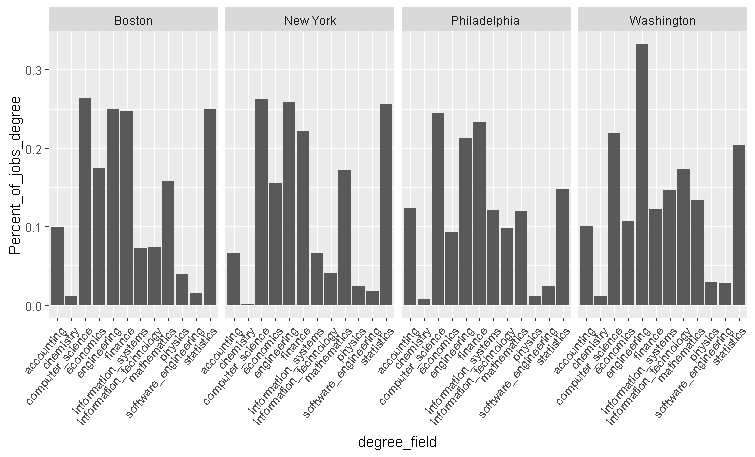
In New York, Boston, Washington D.C., and Philadelphia, SQL is mentioned in 56.22, 46.07, 33.18, and 42.91 percent of job postings, respectively. In New York, Boston, Washington D.C., and Philadelphia, Julia is mentioned in .54, 0, .23, and .34 percent of job postings, respectively. Python appears to be more important in New York than in other cities, appearing in 34.66 percent of job postings there. Scala appears to be less important in Washington D.C. than in other cities, appearing in only 4.16 percent of job postings.



Computer science, engineering, finance, and statistics appear to be among the most popular degree fields. Engineering appears in 26.3 percent of job postings.



Information technology degrees appear more in the job postings in Washington D.C. than in other cities.



For data analyst jobs in the northeast the most preferred programming languages are SQL, Python, and SAS, and the preferred degree fields are computer science, engineering, finance, and statistics. It is more important to know Python for jobs in New York City than for jobs in other cities. It is less important to know Scala for jobs in Washington D.C. than in other cities.

A further analysis could compare each of these top majors with the top programming languages. This could suggest which combination of programming language and degree field is the most advantageous for a data analyst job.