Which Programming Language Should I Master to Earn the Highest Salaries?

In

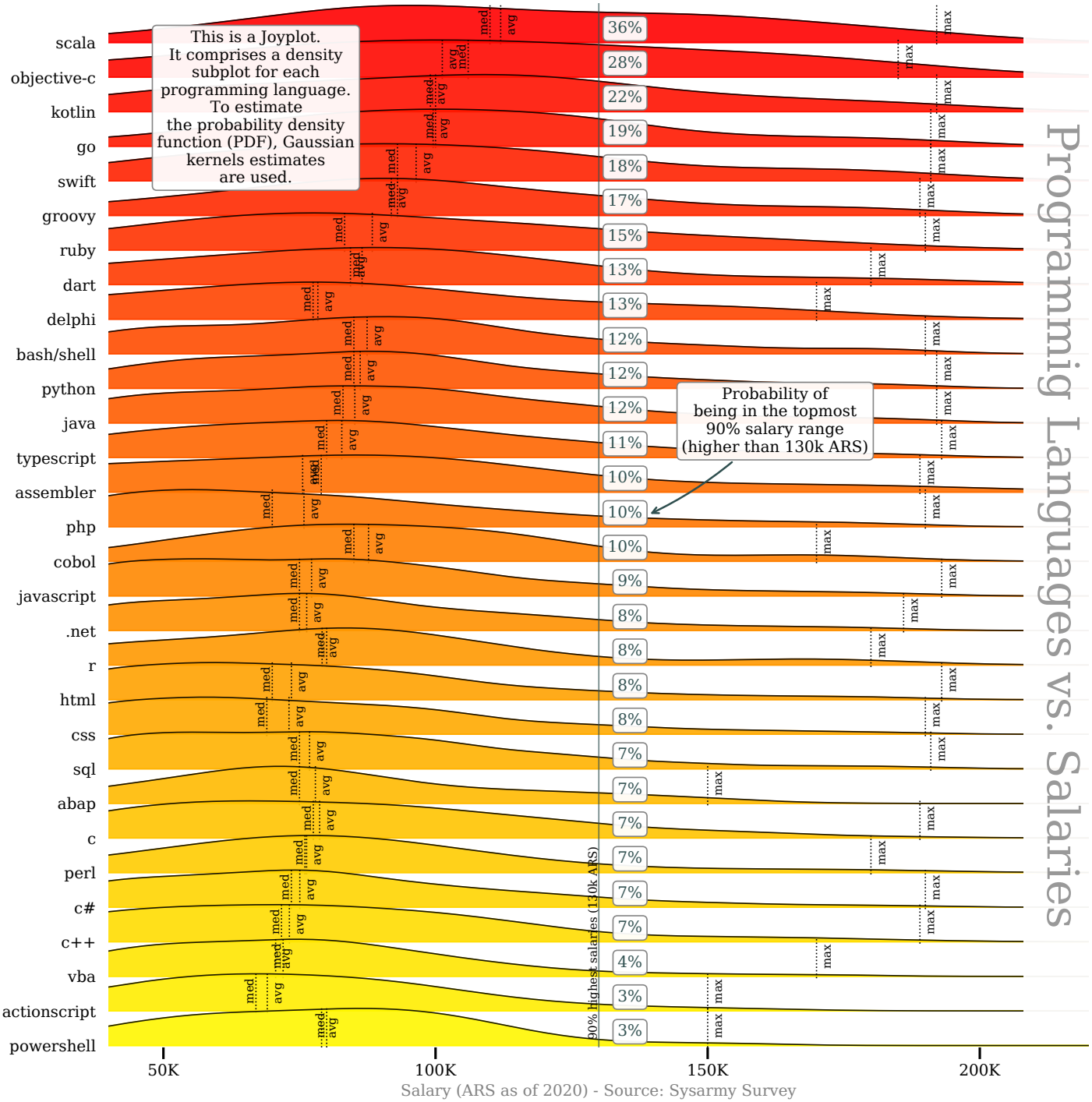
Argentina

By group 4 (Ana, Pablo, Juan, Carlos) of the Data Science Degree (FAMAF)

This is a technical report about the relation of salaries and programming languages used among software industry employees in Argentina as of 2020. The source of the dataset used for the analysis is a SysArmy survey publicly available at this link.

We care about the probability of earning the highest salaries (asuming these are within the topmost 90% of the survey entries), given that the employee is coding a specific programming language. In other words, we compute the conditional probability X in which an employee can earn a salary within the 90-percentile of the best salaries given that he/she masters the language Li. This is P(X|Li) = P(X∩Li)/P(Li), for each language i part of the most frequently used languages.

To this end, we cure the dataset from SysArmy removing outliers salaries below Q.25 - 1.5 IQR and higher than Q.75 + 1.5 IQR (25% and 75% quartiles and Inter-quartile range). We also keep the 30 most frequently used languages on the survey. Each language is considered independently (i.e., a single entry with multiple languages is considered separately). The result is as follows.



The analysis performed shows that coding in Scala unlocks the maxiumum chance (36%) of earning the topmost salaries within Argentina, considering theis dataset from 2020. We obseve that Scala programmers are 8% more likely to earn salaries within the 90-percentile. Modern languages such as Go, Ruby and Python are above 12% of probability of earning the highest salaries. Interestingly, classical languages such as C and C++ are at the bottom of the list, suggesting the lowest salaries among the samples.