EDAV

Tool experience analysis

Next, we wanted to look at the prevalence of tool use among the students from different programs.
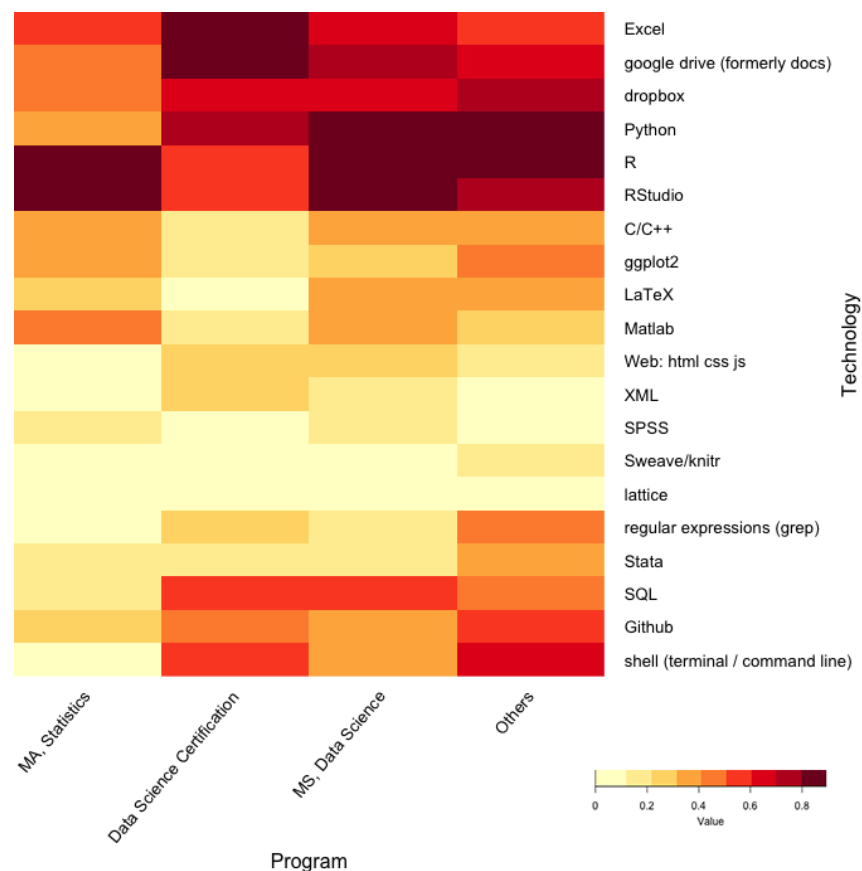

Figure 1: Proportion of students with tool experience among different programs

In the heatmap above, darker colours indicate a higher proportion of the students in that program with experience using tool and lighter colours indicating otherwise. As expected, the use of general tools like Excel, Dropbox, and Google Drive is generally very high across all programs.

- The largest difference is seen between the statistics students and the rest of the class. A very high proportion of statistics students use R and RStudio, but show less use of every other technology, with the exception of Matlab, which is arguably the closest tool to R on the list.

- Next, we see a difference between the Data Science MS and Certificate students. Full time MS students would have encountered Python and R/RStudio during the Fall 2015 coursework, as a result, the three tools are reported as being familiar by nearly everyone in the MS program, which is in contrast to the certificate program. Conversely, the use of tools like Excel and Google Drive, which may be more prevalent in corporate environments, is higher among working professionals in the certificate program compared to the MS program. Likewise for Github, which is a collaboration tool more necessary in the professional environment than in academia.

- Lastly, the profile of students from the other programs (e.g. QMSS, Applied Mathematics, various PhD programs, etc.) look quite similar to that of MS, Data Science students with the exception of higher GitHub and shell use. Curiously, the usage of R and RStudio does not match for these students, which suggests some of these students may be working with R outside the RStudio's IDE environment.

Self-reported confidence levels

Here we explored how confident students from different programs felt about R, RMarkdown, Github, Matlab, and visualization. In the original survey, the students were asked their confidence levels on a scale of *None*, *A little*, *Confident*, and *Expert*. The respective levels were converted to a four-point scale with 0 corresponding to *None* and 4 to *Expert*.

- As we can see on figure 2, the different programs show a noticeable different profiles in skill sets. As expected, Statistics MA students show the highest self-reported confidence with R-related questions. closely followed by the Data Science MS students. While reporting slightly lower numbers for R related categories, Data Science students show much higher confidence with RMarkdown, which is likely due to the use of RMarkdown in the Statistical Inference and Modeling course offered last semester.

- Similar to the findings in the previous *Tool use analysis* section, the Data Science Certificate students show a high self-reported confidence in Github skills in contrast to the masters students, but otherwise report lower confidence with the other skills. Again, as with the previous section, the students from the other programs show a response similar to a combination of all the others.

- Interestingly, Matlab was clumped into the same question as visualization in the questionnaire, making it difficult to interpret the results of the particular question.

Figure 2: Skill set confidence level among different programs