Week 5 discussion:

In today’s competitive data science climate, it is important to have as many tools in your toolbelt as possible. Compare and contrast the benefits of using either R or Python for statistical analysis and machine learning. In which cases should you use R? In which cases should you use Python?

In my opinion, it is beneficial to learn and be proficient in both R and Python for statistical analysis. Many corporates still use R from legacy statisticians to run analyses and it would be difficult to switch the method and communicate back to the team. On the other hand, Python is widely used, so being able to adapt to this popular method is advantageous. Because Python is an open-source, being used worldwide, the great benefits are available resources and many times they are free under Open Source Initiative (OSI) license [1,2]. It’s easy to read and write program that comes with full capacity to collect, store, analyze, visualize data and other tasks in web development or machine learning in one language [3]. On the other hand, R usage focuses on statistics and data analysis which has more flexibility to visualize data and is customizable. Javier Canales Luna, an author of datacamp.com blog, wrote an article in Dec 2022, comparing statistics of trends and highlights for popularity, preference, etc… between Python and R, including usages of each in different industries [4]. Interestingly, retain/CPG industries prefer to use R more than Python, while in other industries both are comparable. For machine learning, clearly Python would be preferred due to its flexibility in implementing ML projects and available packages that are continuously being developed and optimized that can extend beyond ML [5]. Python also has a much better integration rate with other languages than R. However, because Python heavily uses computer memory, so for simple analysis, R will be faster than Python [6]. In summary, it really depends on what corporate environment you are in, the applications needed to fulfill your purpose, and the familiarity you have with each language. Both R and Python will serve the purpose.

1. <https://onlinedegrees.sandiego.edu/python-for-data-science/>
2. <https://opensource.org/osd/>
3. <https://towardsdatascience.com/python-vs-r-for-data-science-cf2699dfff4b>
4. <https://www.datacamp.com/blog/python-vs-r-for-data-science-whats-the-difference>
5. <https://www.educative.io/answers/machine-learning-python-vs-r>
6. <https://www.linode.com/docs/guides/pros-and-cons-of-python/>