**Week 8 Discussion**

You have now learned the basics for R and Python. Of the two languages which do you think is more robust? Why? Which do you think is easier to use? Why? Which professions in the data science and analytics industry do you think would benefit from either language? Your answer should be a minimum of three paragraphs. Please cite all sources. Respond to three other student’s posts.

After learning the basics of R and Python, I believe Python to be more robust. While R is a very sound choice for statistical testing, probability distributions, and some data analysis, Python is a stronger choice for a few reasons. First, Python is powerful because of the object-oriented nature built into various objects. For example, say we want to view the first row of data in a data set.

In R, this could look similar to the following (assuming nba is a dataset that has already been read into the environment):

head(nba, 1)

Now, the same expression in Python would look appear as below:

nba.head(1)

As can be seen, Python is more object-oriented in this example. In Python, Head is a method on the dataframe object, whereas R has a separate head function (1).

Additionally, I believe Python is easier to use. Going back to the example above, the Python dot notation syntax can easily be read and is clear, concise, and to the point. Take another example into consideration.

Find the mean of the same dataset above in R:

library(purrr)

library(dplyr)

nba (is.numeric) (mean, na.rm = TRUE)

Now, the same calculation in Python:

nba.mean()

Again, as can be seen above, the syntax in Python is easier to read and will find the mean of each column in the dataset by default. Also, as evidenced above, Python has more functionality in its “main” libraries for analytical tasks such as calculating the mean of a given dataset, whereas R has a greater diversity of packages, but is also more fragmented in the functionality it presents (1). However, there are some cases where R can be easier to accomplish certain tasks. One of those tasks would be statistical analysis and modeling. While Python contains libraries such as statsmodels, R has a lot of built in support for this functionality. An example of this is the summary() function in R, which has summary statistics regarding datasets baked directly into the language.

Finally, I think that the use of Python and R should and does vary across professions and industries. Since Python is a general-purpose language with a wide variety of use-cases, it’s primary user base ranges from programmers and developers to data scientists, data engineers, and more (2). Python can support the building of entire applications, while also providing strong support for data analysis, machine learning, and model development. Conversely, R is used primarily for data analysis and statistics, and therefore the primary user group for the language are statisticians, researchers, and academics (2). R has an extensive array of packages and can provide statisticians with the ability to quickly access the statistical summary of datasets, build models to test against experiments, and create visualizations to support models.

In conclusion, I believe Python to be more robust and easier to use. R also has applicability in the data science space, however its use is more limited in scope to statistical cases. Based on my career goals, I believe I will probably encounter both, but will more heavily use Python in my day-to-day work.

Resources

1. Paruchuri, V. (2023, March 6). R vs python for data analysis - an objective comparison. Dataquest. <https://www.dataquest.io/blog/python-vs-r/>
2. *R vs python for Data Science & Machine Learning: A comprehensive comparison*. EPAM Anywhere. (n.d.). https://anywhere.epam.com/en/blog/r-vs-python