STAT 763 Homework 6 Comparing Software

By Matt Allen

The three technologies that we used this semester were R/RStudio, Tableau and Python. This paper will compare them based on the following criteria: Data Import, Data Manipulation, Data Visualization, Data Reporting, Basic Modeling, and Advanced Modeling. Broadly speaking, Tableau makes everything more user friendly by providing user interfaces whereas R and Python are code based.

The three software packages can import data. R and Tableau have their own formats like .Rdata for R and .TDE for Tableau. If you have a data source, it is likely that all the technologies can handle it, and R and Tableau have their own data formats. Tableau will make it easier to connect to data sources through a defined User Interface, whereas R and Python will require writing code to import data.

Data Visualization is an area where Tableau is arguably both flexible and has decreased need for coding capability. Data Visualization is the heart of Tableau where Python and R have had data visualization packages built for them. R does have core-plotting functions, but the best plotting capabilities come from outside the core like ggplot. Python has third party libraries that provide for Data Visualization. Both R and Python are code based data visualization, and Tableau is more user friendly by making data visualization possible through dragging and dropping. This is a strength for Tableau. However, R and Python being code based makes it easier to reuse and extend visualizations.

When it comes to Data Reporting, all three can handle most every form of output HTML, Word, PDF etc… Tableau has its own proprietary server for sharing dashboards. R has Shiny or R Studio Server, which is partly free. Python has Jupyter Notebooks for easy sharing. Tableau Server makes it easy for Analysts to publish their dashboards. For R and Python, there is required a greater degree of technical maturity.

All three have Data Manipulation capabilities provided. Tableau has built in aggregation and data joining capabilities similar to SQL, and by default leaves out missing values. R and Python can also handle aggregation and data joining capabilities through third party libraries. Further, Python and R have third libraries that can handle sophisticated data imputation methods that is lacking in Tableau.

For Basic Modeling, Tableau can do descriptive statistics like mean, median, standard deviation etc… However, it is primarily a Data Visualization software and I would not consider it a Statistical Software. Further, Tableau does not have advanced modeling capabilities. Python is a general-purpose programming language that happens to have many useful libraries for Statistics and Machine Learning for basic and advanced modeling. Designed for Statistical Analysis and Modeling, R makes it a little easier to use when compared to Python. Python has innovative libraries like Tensor Flow. R also has advanced modeling capability, but Python may have a slight edge.

	R/R Studio	Tableau	Python
Data Import	High	High	High
Data Manipulation	High	Medium	High
Data Visualization	High	High	High
Data Reporting	Medium	High	Low
Basic Modeling	High	Low	High
Advanced Modeling	High	Low	High

Table 1. Technology Capability Ranked High to Low

In addition to the capabilities, documentation on the technologies is different. Tableau primarily has proprietary forums where you would find help. Whereas through forums like Stack Overflow and GitHub an open source community supports Python and R help.

It is possible to use all three together by using strengths of each. Possible workflow of using all three might be using Python for data import and cleaning, R for statistical analysis and Tableau for data visualization. If I had to choose one, I would pick Python, because it is a general-purpose programming language, and has the greatest extensibility. It also happens to have many useful libraries for Statistics and Machine Learning. If I wanted to make a quick and easy visualization for a dashboard, I would use Tableau. Along a similar line, R is definitely superior to Tableau for doing statistical analysis with less overhead than Python. Python and R require coding ability. Tableau requires little or no coding ability, but extensibility is lost. Increased coding ability needed will also increase flexibility and reusability, and decreased coding ability required by Tableau has in general the opposite effect.