**R Training Content**

Do you want to recommend using Jupyter notebook R or RStudio:

<http://blog.revolutionanalytics.com/2015/09/using-r-with-jupyter-notebooks.html>

* Section 1: Getting Started with R
* Section 2: Data Types in R
  + Scalars
  + Vectors
  + Matrix
  + Dataframe
  + List
* Section 3: Data Input/Output
  + read.csv, read.table (encoding, errors)
  + readr
  + read\_excel
  + Connecting to Vertica and running SQL queries
* Section 4: Data Manipulation
  + Adding and removing columns
  + Subsetting
  + Merging data frames
  + Sqldf
  + Renaming columns
  + Data type conversion
  + Sorting a dataframe
  + Unique function
* Section 5: Exploratory Data Analysis
  + Summary function
  + Mean,sd,quantile
  + Crosstabs – table, **xtabs**
  + Custom Summary – plyr package (summarize), apply functions
  + Aggregate function
  + Counting NA’s
  + Keyword search functions (grep), stringr,stringi packages, zipf
    - <http://www.r-bloggers.com/basic-text-string-functions-in-r/>
    - <https://en.wikibooks.org/wiki/R_Programming/Text_Processing>
    - <http://www.johnmyleswhite.com/notebook/2009/02/25/text-processing-in-r/>
    - <http://www.johnmyleswhite.com/notebook/2009/02/25/text-processing-in-r/>
    - <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/SvetlanaEdenRFiles/regExprTalk.pdf>
  + Word cloud
  + Keyword search functions (grep), stringr,stringi packages, zipf
  + Visualization to be covered here?
* Section 6: Control Structure and Functions
  + for loop
  + while loop
  + if loop
  + ifelse
* Section 6: Statistical Modeling and Data Mining
  + Train/Test/Validation splitting using runif and sample
  + Binning – Use cut function
  + Classification Models
    - Logistic
    - SVM
    - Decision Tree
    - Grid search
  + Model Validation
    - Confusion Matrix
    - ROC Curves (Use Epi)
  + Linear Regression
  + Clustering Models
    - K-Means
    - Hierarchical Clustering
  + Text mining using tm
* Section 7: Visualization
  + Line chart
  + Scatter plot
  + Histogram
  + Box plots
  + Bar Plot
  + Stacked bar plot
  + Saving a plot
  + qplot, sggplot???
* Section 8: Efficient R Programming
  + Writing functions
  + Avoid loops
  + Write C++ to speed up loops
  + Overcoming memory limitations – R packages for sequential model building