R Workshop [potential] framework

1. Intro to R
   1. An introduction R studio interface
   2. Basic coding 101
   3. Data manipulation
   4. Uploading data
   5. In-class homework (upload and visualize your own data)
   6. Conclusion – best practices when stumped and resources
2. Dataset manipulation and processing
   1. Foundational skills of data manipulations:
   2. Detecting errors and outliers
   3. Dataset restructuring
   4. Missing data
   5. Miscellaneous data manipulation
   6. getting data out of R
3. Publication quality plots
   1. Plotting multiple plots
   2. Displaying multiple variables in one plot
   3. Stacked area (i.e., proportions)
   4. Density plots
   5. Plotting model outputs
   6. Graphics aesthetics
4. Intro to R
   * 1. Throughout, sprinkle in when, why, and how you ask R for help
     2. Throughout, use base functions to explain what functions are and how functions are simple models based in coding syntax/language
   1. An introduction R studio interface
      1. Soft-wrap
      2. What is a script
   2. Basic coding 101
      1. Syntax intro with examples
         1. Commenting out
         2. R as a calculator
            1. Addition, subtraction, sum, sqrt, square, mean, exponentiate, median, summary
         3. Assigning names
            1. “=” & “<-“
      2. Basic data generation
         1. seq(), colon (i.e., 1:10), rnorm, vectorizing (concatenate)
         2. Dataset building from generated data
            1. data.frame()
   3. Data manipulation
      1. Basic dataset manipulation (use a generated dataframe (see above) or a stock dataset (e.g., ‘ChickWts’ for categorical data, ‘Orange’ for continuous/categorical and/or ‘trees’ for continuous)
         1. Subsetting (this also allows for basic logical operator instruction)
            1. subset() function syntax
            2. using indexing (i.e., data[,c(“column1”, “column2”] etc)
         2. random sampling from a dataset
            1. sample()

replace =TRUE v replace=FALSE

* + 1. Basic dataset summarizing
       1. summary() function
       2. tapply() function
       3. aggregate() function
          1. this alone could take some time
       4. apply() function
          1. the ‘trees’ dataset would be good for this
          2. this would/could also be a great spot to introduce students to basic function building:

st.err <- function(x){

sd(x)/sqrt(length(x))

}

apply(trees,2, st.err)

* 1. Uploading data
     1. read.csv()
     2. Basic plotting
        1. Histograms, scatter, boxplots
        2. Could be a good place to show basic data transformations
           1. Reciprocal, log, sqrt
     3. Simple statistics
        1. Basic regression, t.test
  2. In-class homework:
     1. Upload your own csv file, perform some basic data summaries/subsetting, make a basic plot
  3. Conclusions
     1. Concepts of coding etiquette/good practices
     2. Discussion on resources, talk about the reference card, google.
     3. Explain packages and how to download/upload

1. Dataset manipulation and processing
   * 1. Could use the “tidy data” concept as a foundation
   1. Foundational skills of data manipulations:
      1. Generating data
         1. rnorm, rbinom
      2. Apply function
      3. For loops
      4. Aggregate()
      5. Renaming columns
      6. Reordering columns
      7. Adding new mutated data
         1. i.e., dat$ln\_var = log(dat$var)
      8. reordering factors
      9. Converting among data classes
      10. Convert numeric categorical data into text
          1. E.g., from 0/1 to “absent”/”present”
      11. Working with Dates
          1. Converting to usable date notation
             1. The format() function
          2. How to generate DOY
   2. Detecting errors and outliers
      1. Identifying typos
         1. unique(dat$column)
      2. visualizations for outliers; boxplots
      3. searching for NAs
   3. Dataset restructuring
      1. Merging datasets
         1. cbind()
         2. rbind()
         3. merge()
      2. ‘reshape2’ package
         1. Casting and melting
   4. Missing data
      1. Simple interpolation
   5. Miscellaneous data manipulation
      1. Normalizing data and back transforming
      2. strsplit()
   6. getting data out of R
      1. write.csv()
2. Publication quality plots
   1. Plotting multiple plots
      1. Identical plots size
      2. Different plot sizes using Layout
      3. Plotting through multiple datasets/simulations
         1. Plotting with For Loops
   2. Displaying multiple variables in one plot
      1. Layering datasets in a plot
         1. par(new=TRUE)
         2. alternative axes
      2. contour/heat map plots
      3. 3d plots
   3. Stacked area (i.e., proportions)
   4. Density plots
   5. Plotting model outputs
      * 1. ‘Effects’ package
      1. Linear regression
      2. Logistic regression
   6. Graphics aesthetics
      1. Transparent graphics
         1. gray() & rgb()
            1. R Color Brewer
      2. Legends inside and outside of plot
      3. DOY plots
      4. Adding text to multiple plots
      5. Expression function for notation and Greek
      6. Creating your own color gradient