How to Prepare a Spreadsheet for Importation Into R

Hips Don't Lie Brown Scholar Internship

Before importing a data spreadsheet into R, we must adequately clean the spreadsheet so that R can read it properly. Follow these steps to prepare our primate data for use in R:

1. **Fix Column/Row Names:**
    - All column/row names must be fixed to fit R naming conventions.
    - The first row of your spreadsheet represents the name of each column. These are your *variables*.
    - Remove ALL blank spaces, special symbols, duplicate names, and empty cells in column names. Make sure you do not accidentally have any blank spaces at the end of a name !
      Ex:
      Bad name: Species N@me
      Good name: Species_Name
    - Do not start names with a number.
      Ex:
      Bad name: 100m Dash
      Good name: Dash_100m
    - The first column of your spreadsheet may represent the name of each row. If so, follow the steps above for these names, too.
2. **Deal with Empty Cells:**
    - Delete any fully empty rows or columns from your spreadsheet.
    - Replace any empty cells (missing observations) with NA (meaning "not available").
      Note: R should do this for you when you import the spreadsheet, but doing this before importing into R ensures that R reads these empty cells correctly.
3. **Remove Formatting:**
    - Remove any and all formatting from your spreadsheet. This includes fonts, color, boldface, italics, etc.
    - Remove extra comments or notes from the spreadsheet.
4. **Final Tweaks:**
    - You will use your column/variable names a lot in R. Try to make these relatively uniform, short, and understandable. Think: typing morph.dat$Abduction is much easier than morph.dat$Abduction_Of_The_Hip_And_Leg .
    - Remove any useless or repetitive data to make your data as easy to use as possible.
    - Remember that R is case sensitive. "Name" and "NAME" are different.
    - If you have dates in your data, use all four digits for the year.
      Ex:

Bad date: 01/01/19
Good date: 01/01/2019

5. **Save the File:**
    - Save your spreadsheet as a .csv (comma separated values) file. This is not the default file format in Excel, Google Sheets, Numbers, etc., but .csv files work best with R.
      Note: If your spreadsheet file has multiple sheets (tabs), saving the spreadsheet as a .csv file will only save one sheet (whichever one you are active on at the time of saving).
6. **Open R/R Studio and Import Your .csv File!**
    - Use the R functions read.csv() and file.choose() to choose and import your .csv file from your computer.
      Ex:
      data.name <- read.csv(file.choose())