Yetz Lab Assignment 1

Discussion Questions:

1. Discuss the impact that the new emergence of this new research methodology is having on the field when so many researchers only use these GUI based platforms. Who is most effected most as we have higher expectations or our researchers?
   1. I.e. What effect will this have on GUI based users that have been conducting researchers on these platforms for decades? Additionally, how much of an impact will this have on students that learned GUI platforms in their education?
2. The article mentions in section 2.1 Wicherts et al. (2011) article that mentioned why scientists fear exposing their work due to data being associated with weaker evidence and apparent errors in the reported statistical results. Understanding this, are their actions we can take to enhance the sharing of data within the scientific community?

**BONUS!**

The Article mentions that “R is not the most efficient platform for processing massive datasets…” As a note, in 2015 Microsoft purchased “Revolution Analytics” (Who set-out to make R quality for “Big Data”) and now has created R-servers for handling larger sets within our Big Data world… And improving! It really just shows how big of a deal R is! <http://www.cio.com/article/2906456/data-analytics/microsoft-closes-acquisition-of-r-software-and-services-provider.html>

1. Introduction
   1. Overall, there have been extreme problems if reproducibility of analyses when it comes to published results. In this article, they promote the importance of creating “reproducible data analysis”.
   2. Additionally, by correctly archiving data analysis techniques through coding and making it reproducible, we can multiple analyses procedures such as:
      1. Reproduce the results
      2. Apply alternative analysis techniques to the same raw data
      3. Apply the same analyses techniques to other raw data as to apply the analysis techniques to other datasets
   3. Lastly, there are many tools available for this kind of work such as: R, R-Markdown, and the R-package knitr to produce replicable reports.
2. Data Archiving
   1. 2.1 Making Datasets Publicly Available
      1. The limitations and concerns to sharing data publicly are expressed here. Including:
         1. Fear of finding inconsistencies in their evidence,
         2. There may be inconsistencies in the way they reported their own results
         3. IRB concerns
         4. Data collectors want to do their own analyses before having others partake
         5. P-bashing: fear of others making their results go away (i.e identifying other outliers)
         6. Lack of understanding of where to upload data
   2. 2.2 Dataverse Project
      1. What is the Dataverse project?
         1. Open-source web application to help promote the sharing, citing, exploring and allowing all to analyze research data.
      2. 3 pillars of Data publishing
         1. Trusted data repository
         2. Data citation
         3. Information about the data
      3. Metadata = information on the variables
         1. Improves reproducibility
      4. Dataverse data may be accessed through R (through Github)
3. Statistics! R in Psychology
   1. 3.1 Why Use R?
      1. R is an extremely useful tool. Especially in recent years when sample sizes are becoming increasingly large and our analyses techniques are becoming increasingly complex.
      2. Essentially, R has everything you need through its wide array of packages and there are more being created for all sorts of fields everyday.
   2. 3.2 Understanding the R Environment
      1. R-studio is a platform for R that is somewhat more user friendly (but you still need to code!)
      2. As Base R does perform statistical functions, you will be needing to install new packages as your statistical analyses and data manipulation needs become more complex.
   3. 3.3 On Learning R
      1. Learning R have quite the learning curve, however, once you understand certain concepts, you’ll be able to pick it up.
      2. Benefit of R: “one actually needs to know what he or she is doing in order to fit a statistical model”
   4. 3.4 R and Reproducible Research
      1. There are many benefits to R-language based research
         1. Instead of sending tables back and forth, we send code and makes reproducibility easy.
         2. The a-priori and post-hoc distinction within the R-programming language helps to produce a better understanding of what researchers are doing.
         3. We can get much more enhanced reports.
4. Generating Dynamic reports
   1. Latex = word processing tool
   2. Introduction to *Dynamic Report Generation*
      1. Making minor changes in the data, visuals, and outputs through little code changes get immediately reported in the paper.
      2. Minimizes user-error inconsistencies
   3. 4.1 The Latex Markup Language
      1. Latex = It is a markup language (I guess people don’t like it being called a word processing tool)
      2. Has the advantage of making reports seem beautiful
      3. Open source and runs in R!
   4. 4.2 Marrying R and Latex: The knitr package
      1. Knitr combines R code and the Latex environement
      2. Automatically updates reports if any steps to the analyses has changed
      3. This is what is happening when you are “knitting” in R-mardown
   5. 4.3 R Markdown
      1. Report generator through R and uses Markdown programming language
      2. Incorporates R, Markdown, and HTML language.
5. Summary and Discussion
   1. Overall summary of the report
   2. NOTE: All technologies are open source in this article (That means FREE!)
   3. These open source softwares are helping researchers in less-developed countries and allowing much more high-level research.
   4. Every software has its limitations, however, R is really becoming more and more apparent as time moves forward.
   5. R is not effeicient for massive datasets
      1. PYTHON is a good alternative
   6. There are many resources for learning R easily.
   7. Look at the supplementary material!