Sunday Update
STAT 222: MA Capstone

Matt Boyas

February 9, 2014

This week I completed the following:

1. I completed exercises suggested in class relating to shell scripting and Emacs. I also continued to become comfortable with Git and GitHub by commiting files and pushing them to my GitHub repository as I completed tasks relating to my term project.

2. I spent a large amount of time during the week trying to figure out how to get my XLS dataset file (which also has commas included in string variables) into a CSV or TXT file that can be used with our class software. Normally I would just open Excel and hit *File → Save As CSV* but that is not a very reproducabe action, and I wanted this file format change to be scripted. I tried working with tools and packages for Python, the Shell, and Perl, but either I couldn't get the code to work or I did not have privileges to install the package on my remote workstation. Eventually, I discovered the R package *gdata* that enables you to read XLS files into R, relying on the Perl package that I was not able to directly install and use. Then I wrote a simple R function to read in my primary dataset and save it as a tab-delimited text file that will be compatible with our class analysis tools.

3. I worked through some of Wes McKinney's *Python for Data Analysis* to get myself prepared for class this coming week. I read my tab-delimited dataset (produced by the R code described above) into iPython Notebook. Then I worked with the variable giving the number of deaths for each social conflict episode to create a dummy variable that is 0 if the conflict had no deaths and 1 if the conflict had at least one death (and NaN if the number of deaths were not known).

4. Finally, I spent a little time continuing my literature review of this research topic.