# Case Study 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Cell | Alt Phone | Email1 | Email2 |
| Claudia Woodruff | 817-800-9293 | 817-281-3840 | woodruffc@smu.edu | claudia\_woodruff@yahoo.com |
| Jeff Weltman | 469-605-0387 |  | jweltman@smu.edu |  |

NOT STARTED --> IN PROGRESS --> COMPLETE (Grade ‘B’) --> COMPLETE(Final)

## Data Gathering and Storage

1. Storing files on Github – set up new Repo named **StarsStripesAndBeer**.
   1. Initial *Readme*
   2. Codebook

## Analysis and Results

1. Create Folder “**Analysis**”
   1. Makefile?
   2. *Readme*
      1. What’s in the folder
      2. (any hypotheses formed from initial analysis)
   3. Create Folder “Raw” to put untouched raw data files
      1. *Readme*
         1. where we got it,
         2. where it came from,
         3. how it’s structured
      2. Upload the raw data files
   4. Create Folder “Data”
      1. *Readme*
         1. What is in the folder
      2. Tidy/Clean Data File (from the raw file)
         1. *?? Question - (Do we need? All we had to do is rename a column.)*
         2. For reproducibility’s sake, yes, although you’re right – not a lot of tidying. I also added a step in the Analysis file that changed South Dakota’s IBUs from NA to 0. Since all of their beers were missing IBU info, removing the NAs had the undesired effect of removing **all beer data** for South Dakota. That’s two tidying steps we’ll save future researchers :) I’ve made a note below to remind us to mention the SD change in the report.
      3. Tidy/Clean Merged File
   5. Analysis R files
      1. Created analysis.r – 90% of which is code Jeff wrote. (It may be complete now, if you like the idea of separating out the plotting into a separate r script.)
      2. I started a separate R file, just to keep things modular, which is for plotting. I don’t see how a bar chart is useful (but let’s discuss and ask Dr. T for clarification.) Trying various other types of (ggplot2) plots/graphics.

## Presenting

1. Create Folder “**Presentation**”
   1. Rmarkdown1 – html output (I think html is mandatory, since it is the out-type we’ve been required to produce so far.
      1. We’ll want to mention all the NAs we removed due to missing IBU data– down to 1405 beers after starting with 2410.
      2. Especially mention that we changed South Dakota’s IBU to 0 from NA in order to not remove all beer data for that state.
   2. Rmarkdown2 - (Doc? Slide? (if we have time, maybe doc or slide?)