Principles of Quality Data Analysis

1. Write a plan before conducting the analysis.
   1. What datasets are being used?
   2. What variables within those datasets are important?
   3. What variables are going to be created?
   4. What summary statistics or graphs are going to be output?
2. Write a readme in the project directory.
   1. Who - are analysts and the PI?
   2. What - are you doing? Include a quick project summary for context. Which scripts do which tasks?
   3. Where – are the locations of the data, the scripts, supporting documents?
   4. When – did this project start and when was it last updated?
3. Organize folders clearly and cleanly, with obvious names.
   1. data
      1. /derived\_data
      2. /raw\_data
   2. scripts
      1. /sas
      2. /R
      3. /functions (or macros)
   3. reports
   4. documentation and plans
4. Keep meticulous notes.
   1. For each project, keep a Notes and Meeting minutes word document with dates, names, keywords, and a summary/action items from each meeting so you can ctrl-f items later.
5. Stay on top of version control – either Git or use archive folders + dated versions. Git is strongly preferred.