

## Closing Thoughts

Dr. Matthew L. Settles

Genome Center  
University of California, Davis

# 7 Stages to Data Science

1. Define the question of interest
2. Get the data
3. Clean the data
4. Explore the data
5. Fit statistical models
6. Communicate the results
7. Make your analysis reproducible

# Prerequisites

- Access to a multi-core (24 cpu or greater), 'high' memory 64Gb or greater Linux server.
- Familiarity with the 'command line' and at least one programming language.
- Basic knowledge of how to install software
- Basic knowledge of R (or equivalent) and statistical programming
- Basic knowledge of Statistics and model building

## The Bottom Line:

Spend the time (and money) planning and producing **good quality, accurate and sufficient data** for your experiment.

Get to know to your data, develop and test expectations

Result, you'll **spend much less time** (and less money) extracting biological significance and results during analysis.

# Workshop computing reservation

- workshop ACTIVE Saturday Dec 23 11pm, 2017

My recommendation is to follow all of the instructions again, from the beginning on your own and send emails to

[training.bioinformatics@ucdavis.edu](mailto:training.bioinformatics@ucdavis.edu)

And we will be responsive to answering questions