Data Science Methods for Clean Energy Research

Week 2, Lecture 1: Data Visualization

January 4, 2017

UNIVERSITY of WASHINGTON

Outline

- Any questions from last time?
- Make sure you join #dsmcer channel in our Slack team
- No new reading for W2L1
- Data visualization
- Visualization of materials and molecules
- Discussion of HW2

What do you use for plotting?

> Open #dsmcer in Slack for quick poll

Basic principles for data visualization: 1) learning by poor example

> Some really terrible graphs

Basic principles for data visualization: 2) community standards and basic principles

- **Understanding expectations of your peers and, especially, your advisor for data presentation is important**
  - **Some communities have very specific guidelines on presentation of error/uncertainty (esp. in life science)**
  - **Some communities rarely show error/uncertainty**
- **Do no evil with your graphics**
  - **There is a fine line between** changing a visualization **to show a scientific conclusion and** implying a scientific conclusion **by changing a visualization**
  - **How do you guard against the latter?**

Introducing matplotlib

- Motivating good data management and conditioning

- Lets plot some data!

Molecular / materials visualization

- Any commonalities between decisions/principles about data visualization in 2D/3D and the visualization of materials and molecules?

- Basic skill of being able to draw 3D structures of molecules and generate extended periodic structures of materials is useful

- There are a lot of widely used free (and for purchase) programs

Molecular / materials visualization

- **The program** Avogadro **is widely used for both molecules and materials – many tutorials and large user support base**
- UCSF Chimera **is also excellent (mostly biomolecular – but renders beautiful images)**
- **The program** VMD **is excellent for rendering images and creating visualizations (given an input structure)**

Discussion of HW2

- Will be entirely GitHub based (instructions coming Wednesday)
- Mostly data visualization -- practice with Python and Matplotlib. Also a small component of molecular visualization

Preview of probability and statistics (next two lectures)

- Probably will not assign HW3 until next Wed

UNIVERSITY *of* WASHINGTON