TSCI 5050:
Introduction to Data Science

Alex F. Bokov, Ph.D.
bokov@uthscsa.edu

Convenience is not a Luxury

Our Toolbox

- git
- bash, and various cool things that live in that ecosystem.
- R
- SQL

Preventing Common Command Line Headaches

1. Often, it matters whether what you type is in upper case or lower case. To be on the safe side, always use lower case.
2. Whitespace usually matters-- it usually separates commands or arguments. Avoid using it in file names.
   a. Often you can include extra whitespace to make your code more readable. Learn when/where.
3. Punctuation almost always matters
   a. #*%^&|+-@;:,.<>=/~\
   b. Except for _ which is usually safe.
   c. If you are typing one of those symbols, make sure you understand what it means.
4. The following symbols almost always have to have a matching pair somewhere in your code: ( ) { } [ ] " " ' '
5. If you need something to be treated as a piece of text (aka character string) rather than a command, you will need to quote it, usually with " " or ' '
6. There is no "Desktop". There is a hierarchy of folders (aka directories) and you need to learn your way around it before you do anything else.
7. The [up] and [down] arrow keys (command history) are your friends! Use them wherever possible.
8. The [tab] key (command completion) is your friend. Use it whenever possible.

Organizing Your Work

1. Use lots of comments in your code.
2. Give your files, functions, and variables short but meaningful names. Try to pick a naming convention and be consistent with it.
   - Except for loop variables. For those, use a doubled letter-- e.g. ii, jj, xx, yy, etc.
3. DRY = "Don't Repeat Yourself"
4. Do not hard-code values. Assign them to variables wherever possible.
   - Do your variable assignments at the beginning of your code.
5. Have clearly defined sections in your code (sometimes even separate files) so it's easier to find things later.
6. Use git!!!

Philosophy

- Machines are consistent.
- Humans are not.
- Therefore, logically consistent outcomes are often counter-intuitive to humans. This, in a nutshell, is why data science is difficult.