Lecture 22 - Markdown and Jupyter

Updates

- practice exam distributed this week to give you a feel for the format
- pick group project topics (statistics, modeling, bioinformatics) on Friday
- NO reading or quiz next Monday

Has anyone noticed the issues tab on GitHub?

- issues are a useful way to track issues with code or map out tasks for code development

- seems like a good idea for group projects

Has anyone noticed what file extension Github readme files use?

What is Markdown?

- a newish, simple markup language
- it allows one to write a file in plain text, but generate documents (html, presentations, pdfs) that have formatted text
- when written the goal was to be simpler and easier to read than html and Rich Text Format
- all of the handouts and presentations for class this semester were written in Markdown

Markdown is the basis for growing documentation platforms in R and Python

- Folks have developed the capacity to include chunks of code into these documents that can be executed with the output woven into the document. This allows for the ability to make "scripts" that generate nicely formatted reports, presentations, and even manuscripts.

- `R Markdown` is accessible in RStudio

- `Jupyter notebooks` are essentially the `Python` version of `R Markdown`, and are included with Anaconda

Markdown basics

In markdown text can be easily formatted inline. For example:

surrounding text with a single * makes text *italics*

surrounding text in ** makes text **bold**

you can make super$^{\text{scripts}}$ by surrounding text in ^

Markdown is quite good with equations, which are surrounded by $

$ y = mx+b$

$\frac{dN}{dt} = rN(1 - \frac{N}{K})$

Markdown basics

We can also make headings and subheadings with different numbers of #; in this case less is more

Ordered and unordered lists and even tables are easy too

Creating a R Markdown document

RStudio makes this quite easy. You can also make R Notebooks, which are intended to be more interactive.

Creating a Jupyter Notebook

This can be launched in Windows from the Start menu and from the terminal in OSX/Unix.

This runs in a web browser, but is not online, to help with portability across platforms.

As it is a "notebook" it can be interactive, but you can also export the notebook as a pdf or presentation.

This blog post has a few useful tips beyond what is covered in the cheatsheets:

`http://datascience.ibm.com/blog/markdown-for-jupyter-notebooks-cheatsheet/`

Challenge

1. Make a short "How to" document describing the process and showing the code and results for creating and filling a 4x4 dataframe with the numbers 1 to 16.

2. Make another "How to" document describing the process and showing the code and results for creating a scatter plot of a random set of 20 numbers between 0 and 50 (**x**) and a second set of 20 numbers (**y**) that are linearly related to x with a slope of 5 and an intercept of 20 with normally distributed error (standard deviation of 5).