Lecture 6: pandas with pop music, shell, Markdown

LING 1340/2340: Data Science for Linguists

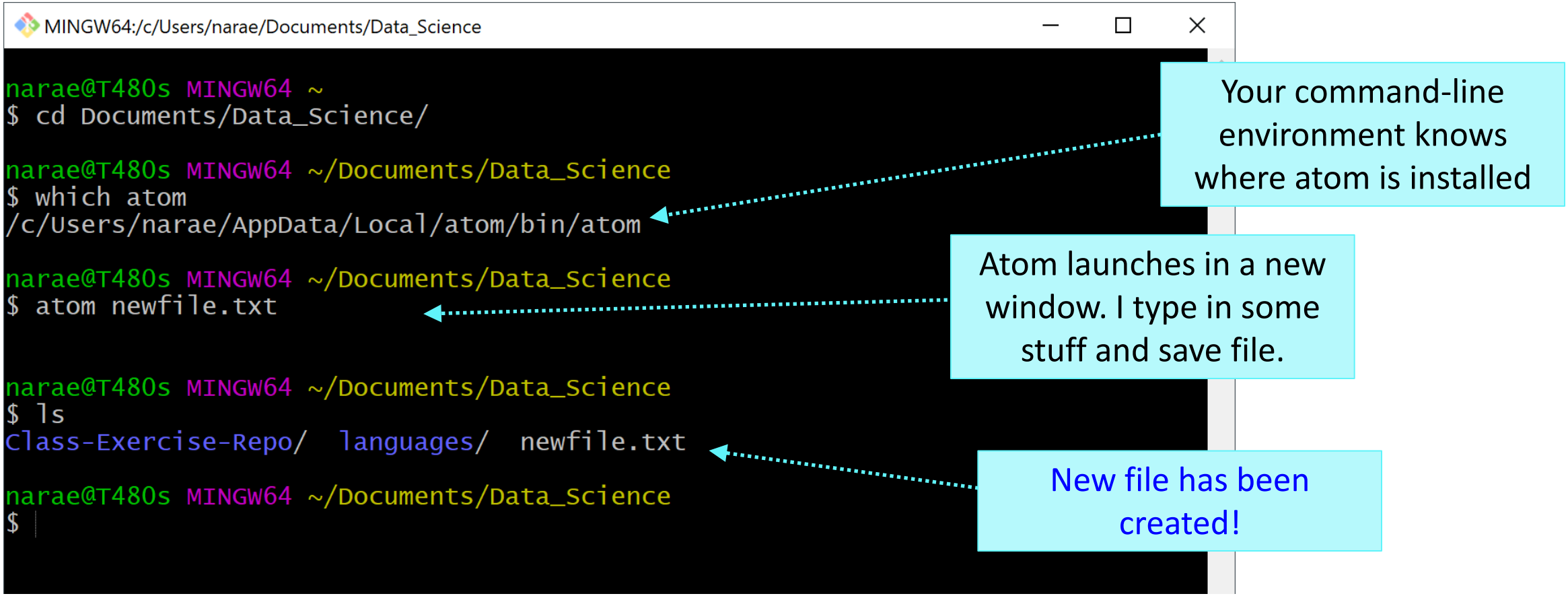
Na-Rae Han

Objectives

- Tools
  - Command-line operations, shell (bash, zsh)
  - Markdown
- pandas library hands-on practice, continued

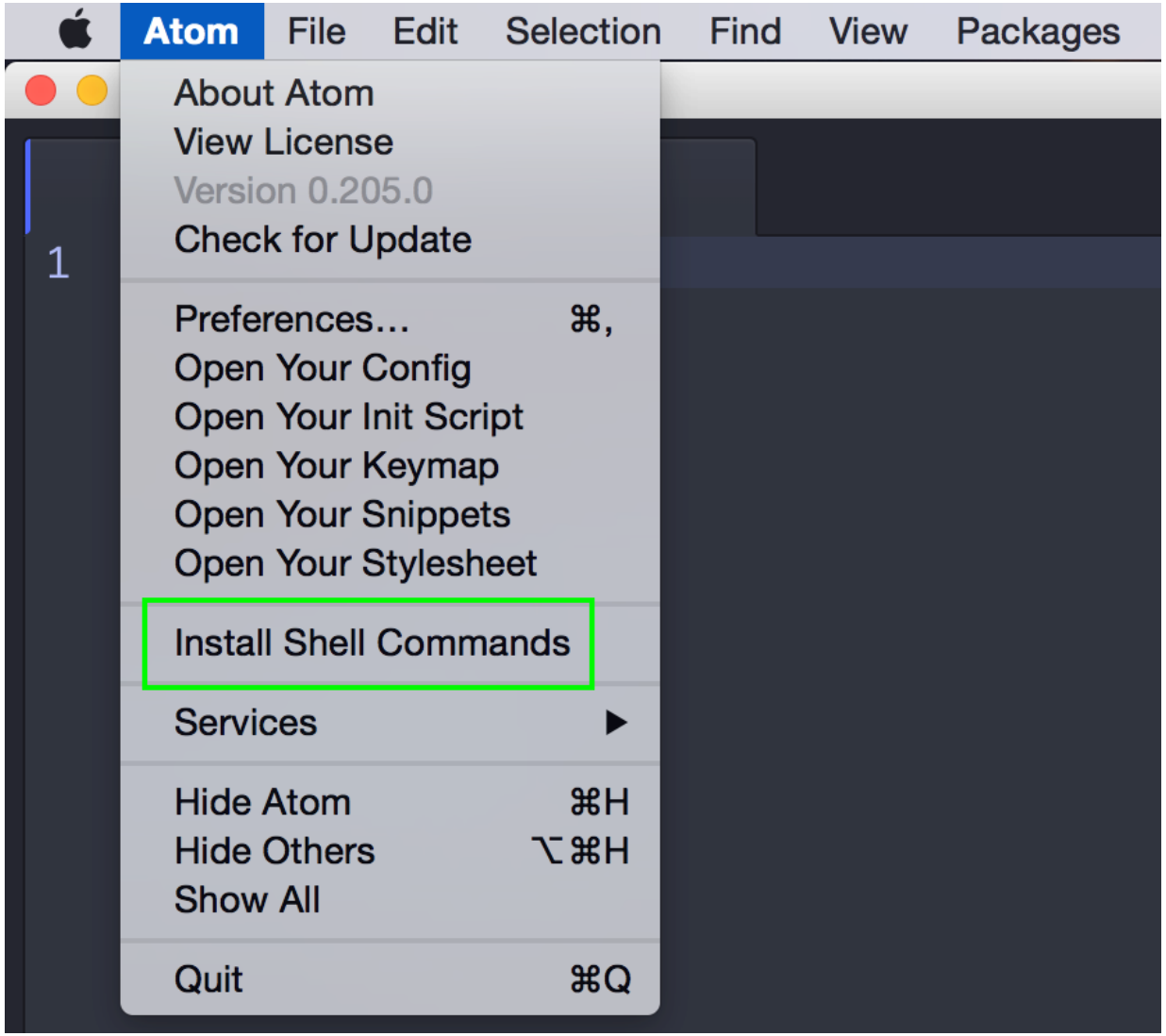
Your text editor in shell

- You should be able to launch your text editor from shell and create a new text file in the directory.



Mac users: configure Atom for shell

- https://stackoverflow.com/questions/22390709/how-to-open-atom-editor-from-command-line-in-os-x
  - "Install Shell Commands"
  - After this, you can launch atom directly from your Terminal (bash or zsh shell).

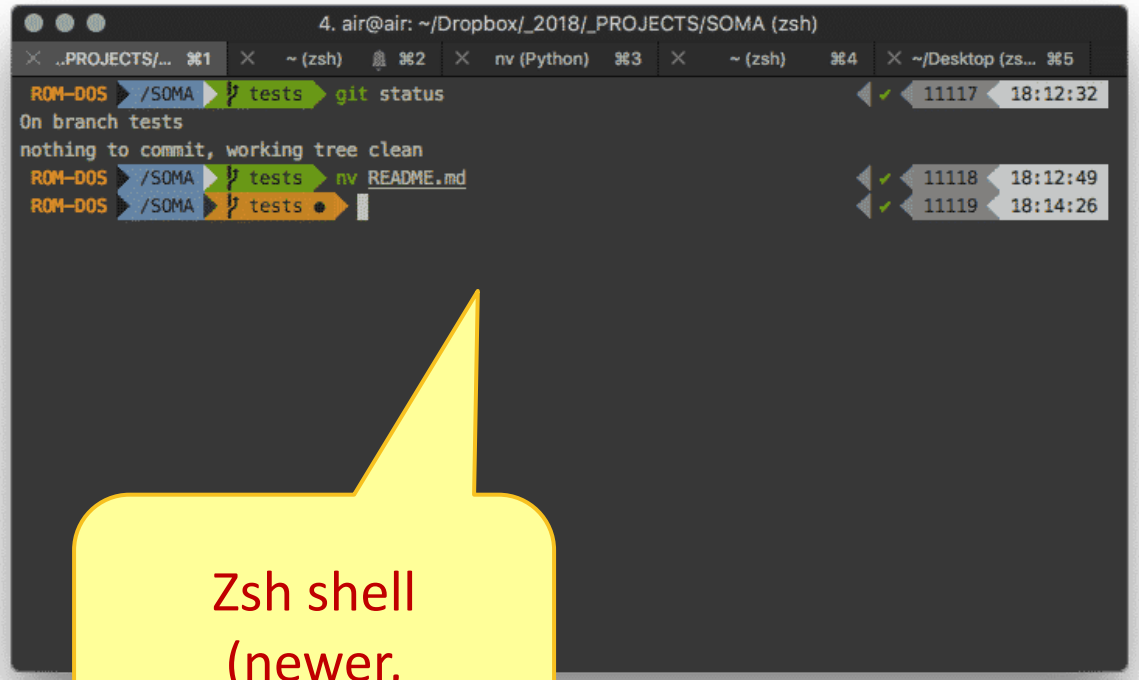


Mac only: Bash vs. Zsh

- **Windows** folks are using Git-bash, which has nice colorized Git output
- **Mac**: new default shell is zsh, older versions will have bash
  - In your terminal, execute `echo $0`

```
mc-130-49-26-56:Documents rhlmc1$ ls
Data-Science-for-Linguists-2019 HW1-Repo
mc-130-49-26-56:Documents rhlmc1$ cd HW1-Repo/
mc-130-49-26-56:HW1-Repo rhlmc1$ git status
On branch master
Your branch is up to date with 'origin/master'.

Untracked files:
  (use "git add <file>..." to include in what will be committed)

        dan/hello.txt

nothing added to commit but untracked files present (use "git add" to track)
mc-130-49-26-56:HW1-Repo rhlmc1$ git log
commit 23df080f51b370eb9dff5794187be4628f801d9f (HEAD -> master, origin/master, origin/HEAD)
Author: Na-Rae Han <naraehan@gmail.com>
Date:   Thu Jan 10 15:14:17 2019 -0500

    script mention changed to file

commit d329b4bc04deac9436ebdea0451893f375a26
Author: Na-Rae Han <naraehan@gmail.com>
Date:   Thu Jan 10 14:57:26 2019 -0500

    repo prepped for perfonal folders,
```

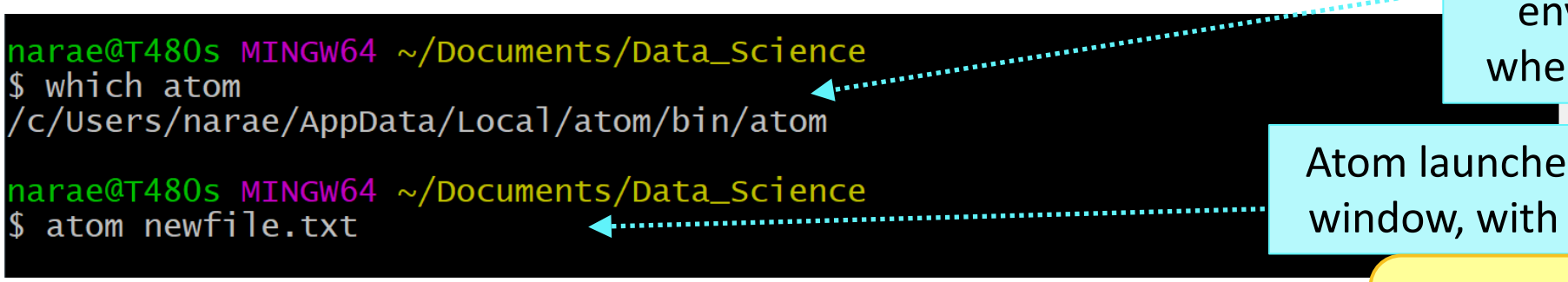
Bash shell (older)



Zsh shell (newer, recommended)

Launching apps via command line

- If `which xxx` works, you can readily launch xxx from shell:



Your command-line environment knows where atom is installed

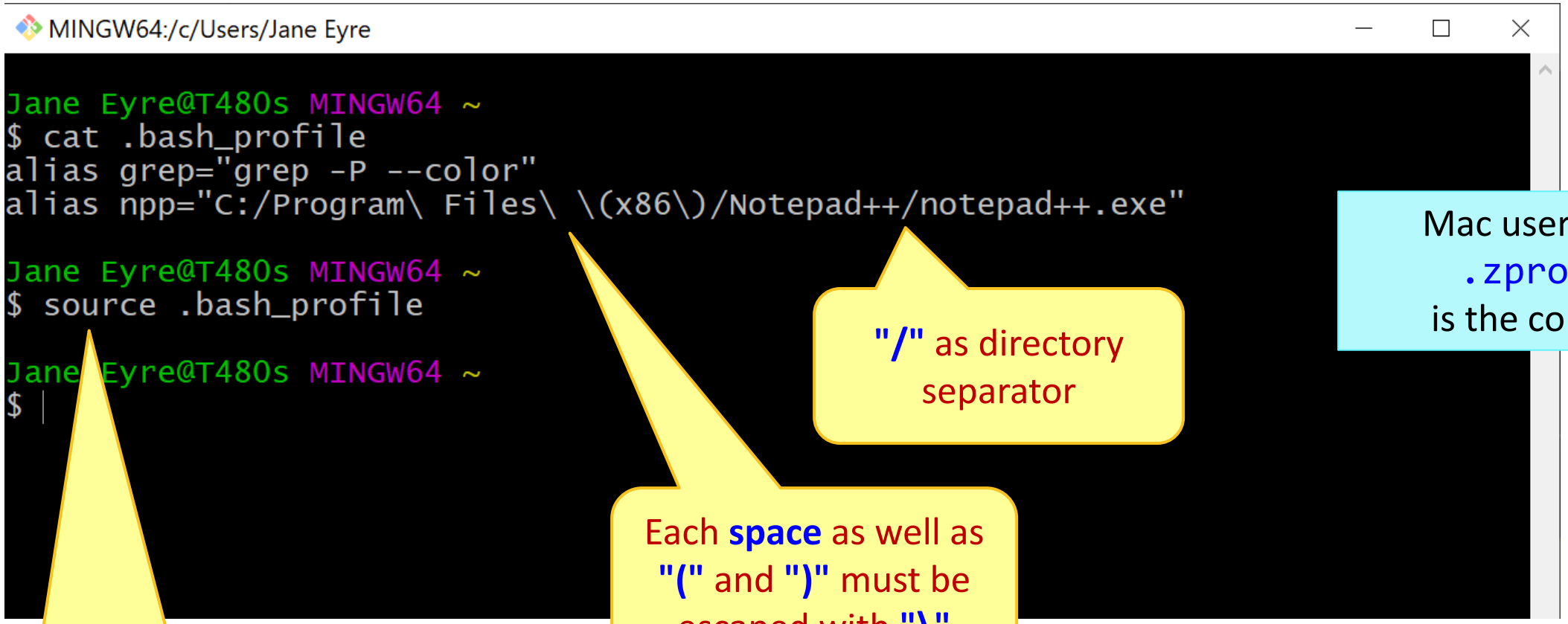
Atom launches in a new window, with a new file

But! Your shell could become unusable until you close app. To avoid, put & at the end.

- If not, you need additional configuration
  - Configure your bash or zsh, add app alias
    - In your **home directory**, open `.bash_profile` in a text editor (`nano` works)
      - Mac users (zsh): open `.zprofile` instead
    - Find the full path for your text editor, chrome, or any app you want
    - Add this line: `alias chrome="/path/to/file"`
    - Save file. Back in command line, execute: `source .bash_profile`
  - **Mac OS users** can use the built-in `open` command:
    - http://osxdaily.com/2007/02/01/how-to-launch-gui-applications-from-the-terminal/

No space around **=** !!

Adding app aliases to your Bash shell



"/" as directory separator

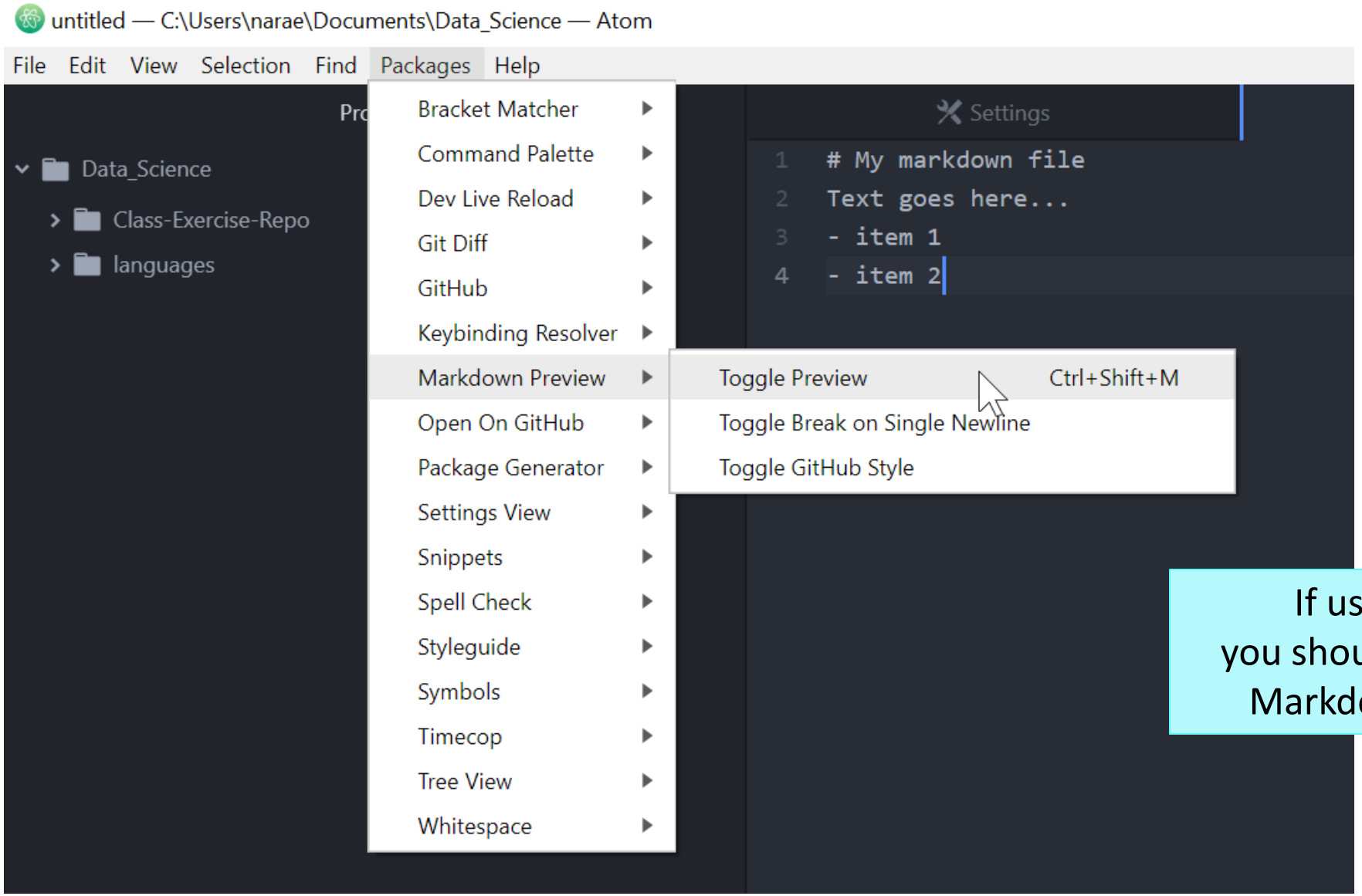
Each **space** as well as "(" and ")" must be escaped with "\"

So that aliasing takes effect immediately (without having to log back in)

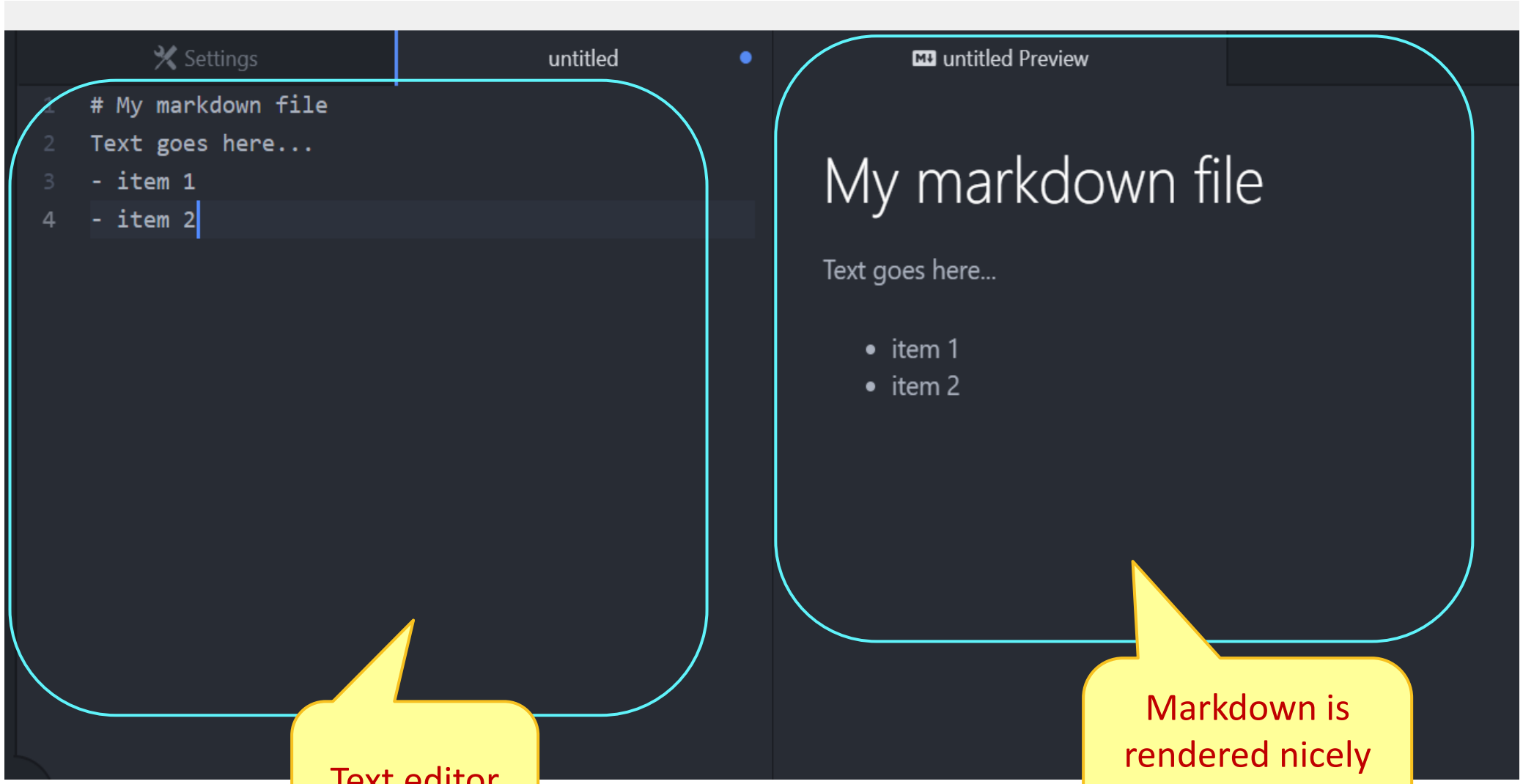
Mac users & zsh:
`.zprofile`
is the config file

Markdown files (.md)

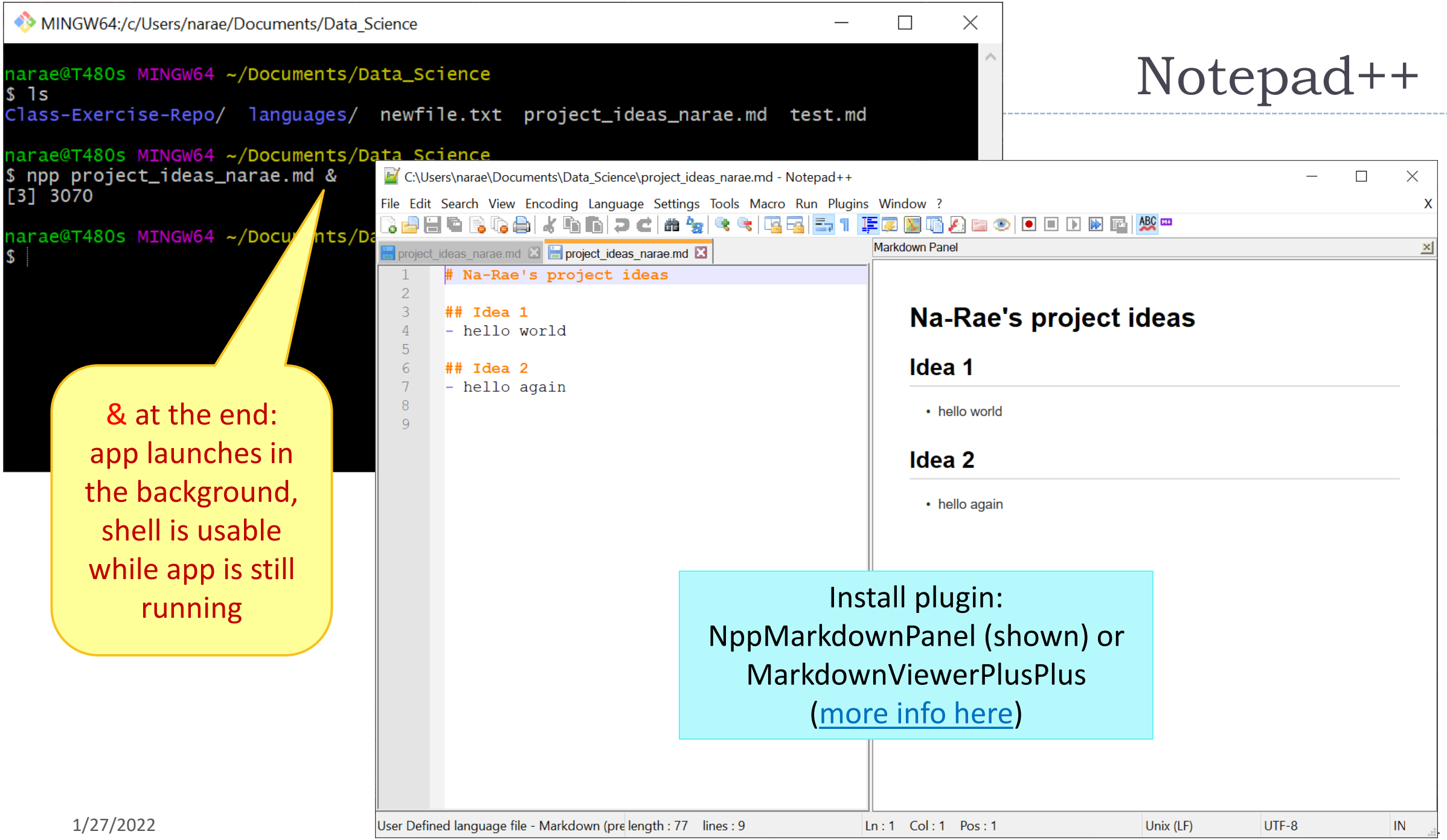
- **Atom** editor has a built-in markdown viewer window
  - Edit windows and Markdown window side-by-side!
- If using other text editor:
  - **Notepad++**: install a md viewer plugin
  - Chrome: install **"Markdown viewer" extension** (link on Learning Resources page)
  - Any other workflow you have?

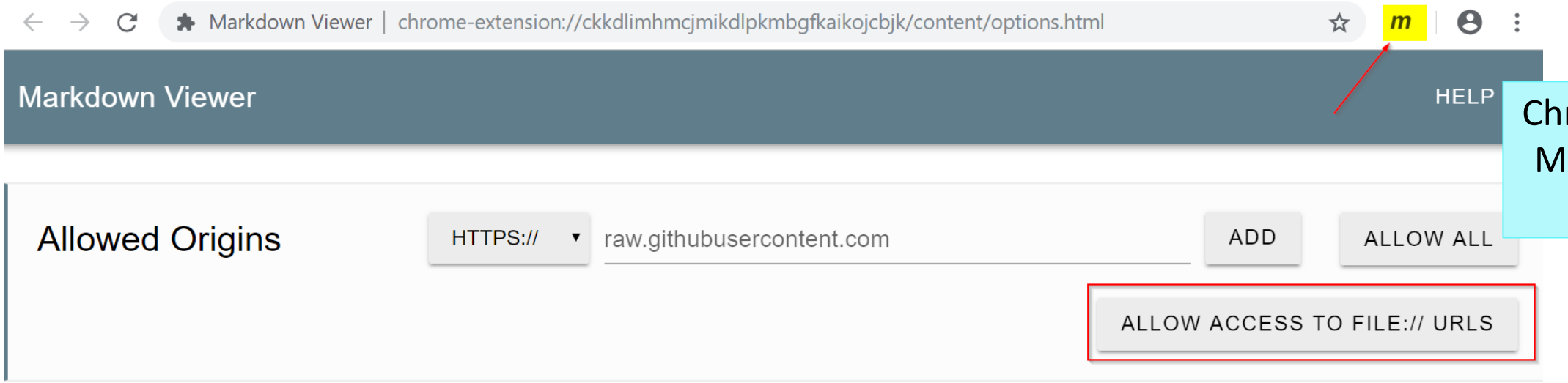


If using **Atom**,
you should first enable
Markdown preview

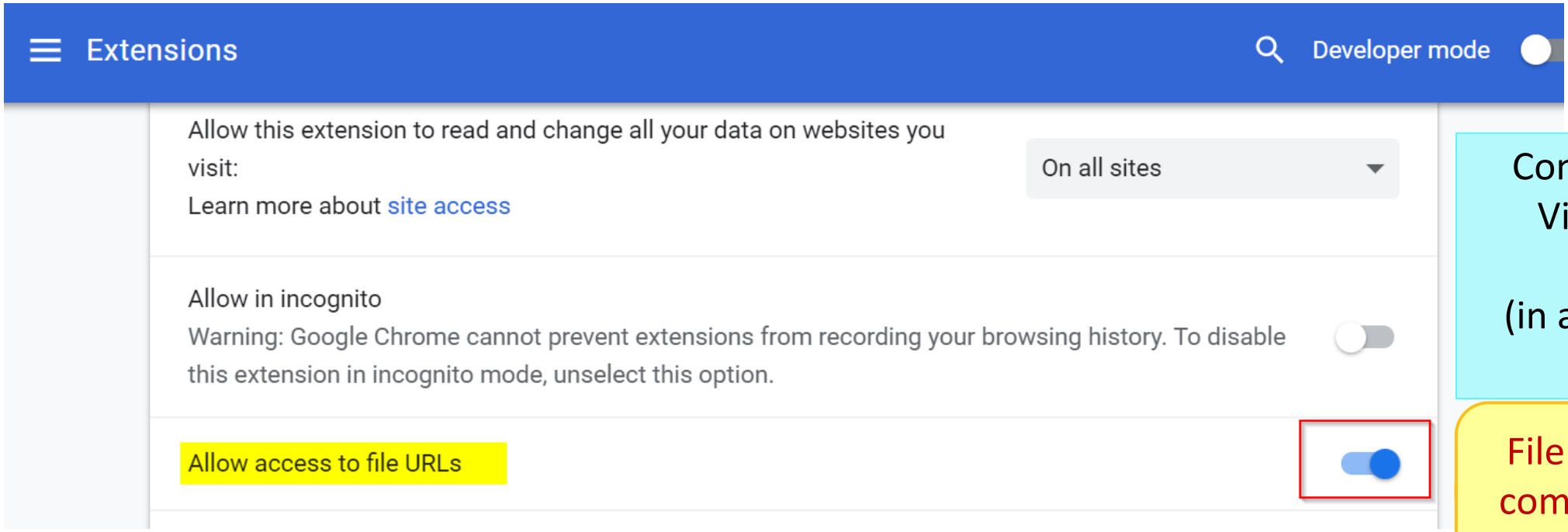

Settings
untitled
My markdown file
Text goes here...
- item 1
- item 2
untitled Preview
My markdown file
Text goes here...
item 1
item 2
Text editor window
Markdown is rendered nicely in the preview window

Notepad++


MINGW64:/c/Users/narae/Documents/Data_Science
narae@T480s MINGW64 ~/Documents/Data_Science
$ ls
Class-Exercise-Repo/ languages/ newfile.txt project_ideas_narae.md test.md
narae@T480s MINGW64 ~/Documents/Data_Science
$ npp project_ideas_narae.md &
[3] 3070
narae@T480s MINGW64 ~/Documents/Data_Science
$
& at the end: app launches in the background, shell is usable while app is still running
C:\Users\narae\Documents\Data_Science\project_ideas_narae.md - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
project_ideas_narae.md
project_ideas_narae.md
1 # Na-Rae's project ideas
2
3 ## Idea 1
4 - hello world
5
6 ## Idea 2
7 - hello again
8
9
Markdown Panel
Na-Rae's project ideas
Idea 1
• hello world
Idea 2
• hello again
Install plugin: NppMarkdownPanel (shown) or MarkdownViewerPlusPlus (more info here)
User Defined language file - Markdown (pre length : 77 lines : 9
Ln : 1 Col : 1 Pos : 1
Unix (LF)
UTF-8
IN



Chrome browser w/ Markdown Viewer extension



Configuring Markdown Viewer extension to allow **file URLs** (in addition to the usual https:// etc.)

File can't be opened via command line, however. Open in File Explorer.

pandas practice

Activity 3
50 mins

- 50 Years of Pop Music
  - https://www.kaylinpavlik.com/50-years-of-pop-music/
  - You tried out the CSV file 'billboard_lyrics_1964-2015.csv'

- In Class-Exercise-Repo, `activity2/` folder:
  - The CSV data file is already in the directory.
  - You will find `pop_music_BLANK.ipynb`

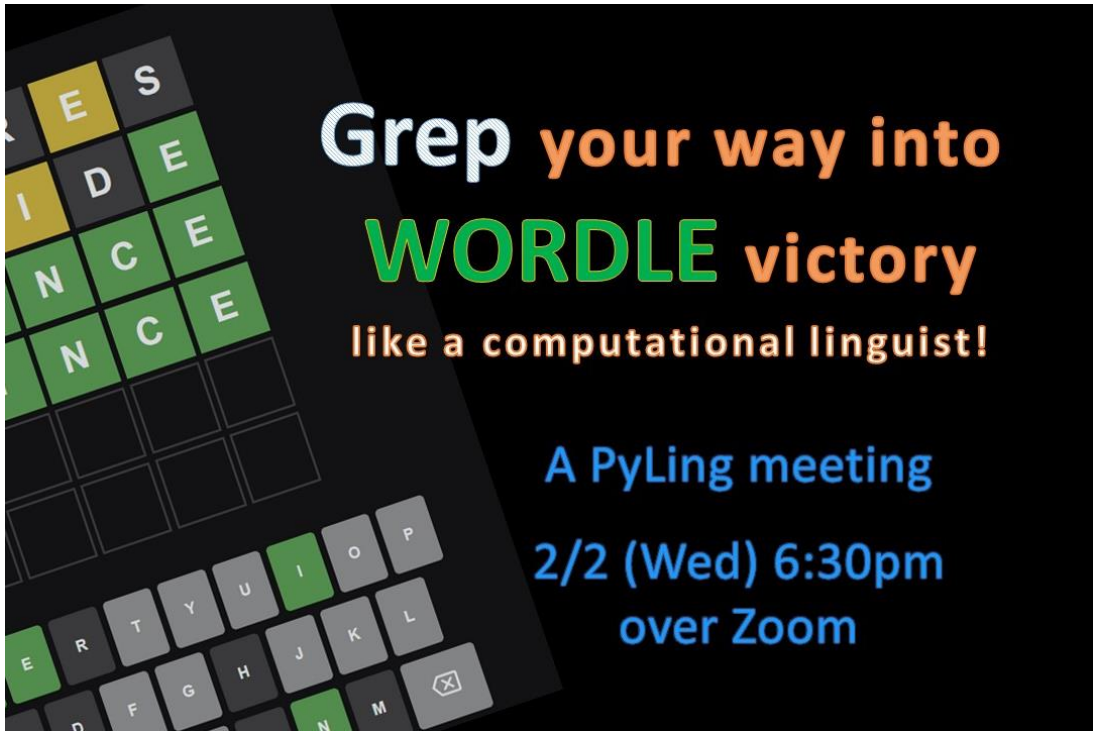
  ← Copy it as `pop_music_YOURNAME.ipynb` and work on it.

  ← You are welcome to commit your version.

`cp file1 file2`
makes a copy of file1 as file2.

Homework 2: Process the ETS Corpus

- https://naraehan.github.io/Data-Science-for-Linguists-2022/hw2
- Corpus distributed via a private GitHub repo "Licensed-Data-Sets"
  - Clone directly to your laptop: no need to fork
- Submission via your own personal & private repo (shared with Na-Rae & the TAs)

Wrapping up

- Coming up: Project ideas
  - Due next Tuesday.
  - Not sure? Come chat with me and the TAs
- Coming up: HW2 Process ETS Corpus
  - Due next Thursday.
  - This one is much bigger than HW1. Start NOW!
- PyLing on WORDLE →
  - Next Wed (2/2) 6:30pm
  - If you haven't received email, let me know