

Lecture 5: pandas with pop music

LING 1340/2340: Data Science for Linguists

Na-Rae Han

Objectives

- ▶ Housekeeping
 - ◆ Using DataCamp courses, downloading csv files
- ▶ Setting up your command-line environment
- ▶ pandas library hands-on practice

- ▶ Tools:
 - ◆ Command line (bash and zsh)
 - ◆ Markdown
 - ◆ Text editor: Atom

DataCamp: download exercise .csv files

▶ <https://learn.datacamp.com/courses/intermediate-python-for-data-science>

- ▶ All .csv data files used in DataCamp exercise can be downloaded!
- ▶ Posted on course dashboard. Scroll down!
- ▶ You can download them and replicate exercises in your own Jupyter Notebook.

The screenshot shows the course page for 'Intermediate Python'. At the top, it says 'INTERACTIVE COURSE' and 'Intermediate Python'. There are two buttons: 'Continue Course' (green) and 'Bookmark' (white with a bookmark icon). Below these, it lists course details: '4 hours', '18 Videos', '87 Exercises', '577,851 Participants', and '7,400 XP'. A yellow callout bubble points to the 'DATASETS' section, which lists 'Gapminder', 'Cars', and 'BRICS'. The 'Loops' chapter is highlighted with a purple bar, and its progress is shown as 6%. The 'Case Study: Hacker Statistics' chapter is shown below with 21% progress.

Download datasets and work locally!

4 Loops 6%

There are several techniques you can use to repeatedly execute Python code. While loops are like repeated if statements, the for loop iterates over all kinds of data structures. Learn all about them in this chapter.

VIEW CHAPTER DETAILS **Continue Chapter**

5 Case Study: Hacker Statistics 21%

PREREQUISITES
Introduction to Python

DATASETS

- Gapminder
- Cars
- BRICS

Launching apps via command line

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

```
narae@T480s MINGW64 ~/Documents/Data_Science
$ which atom
/c/Users/narae/AppData/Local/atom/bin/atom

narae@T480s MINGW64 ~/Documents/Data_Science
$ atom newfile.txt
```

Your command-line environment knows where atom is installed

Atom launches in a new window, with a new file

- ▶ 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, etc.
 - ◆ 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

MINGW64:/c/Users/Jane Eyre

```
Jane Eyre@X1Yoga MINGW64 ~  
$ cat .bash_profile  
alias grep='grep --color=auto'  
alias chrome="C:/Program\ Files\ \ (x86\)/Google/Chrome/Application/chrome.exe"  
  
Jane Eyre@X1Yoga MINGW64 ~  
$ source .bash_profile  
  
Jane Eyre@X1Yoga MINGW64 ~  
$
```

Mac users & zsh:
.zprofile
is the config file

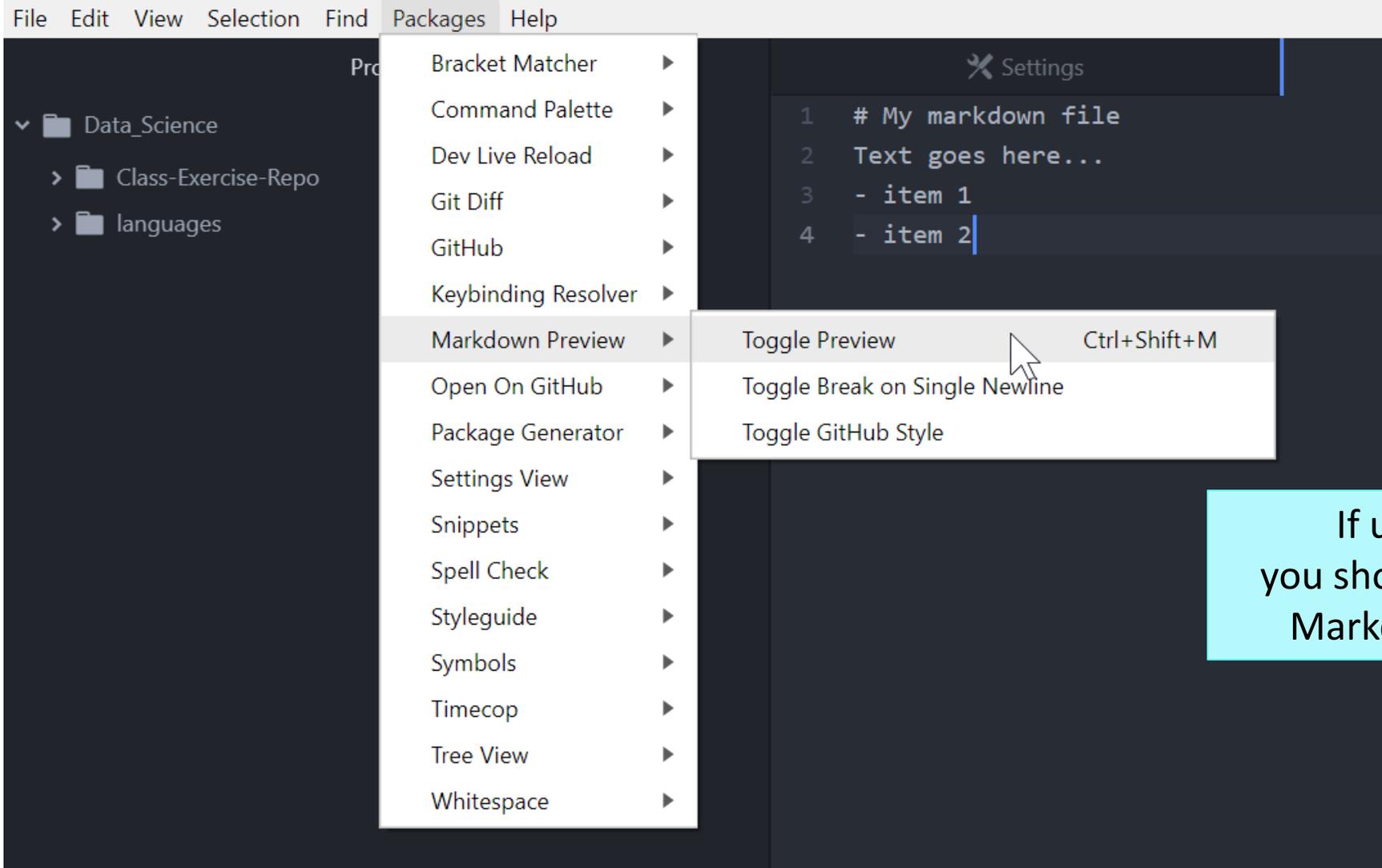
"/" as directory
separator

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

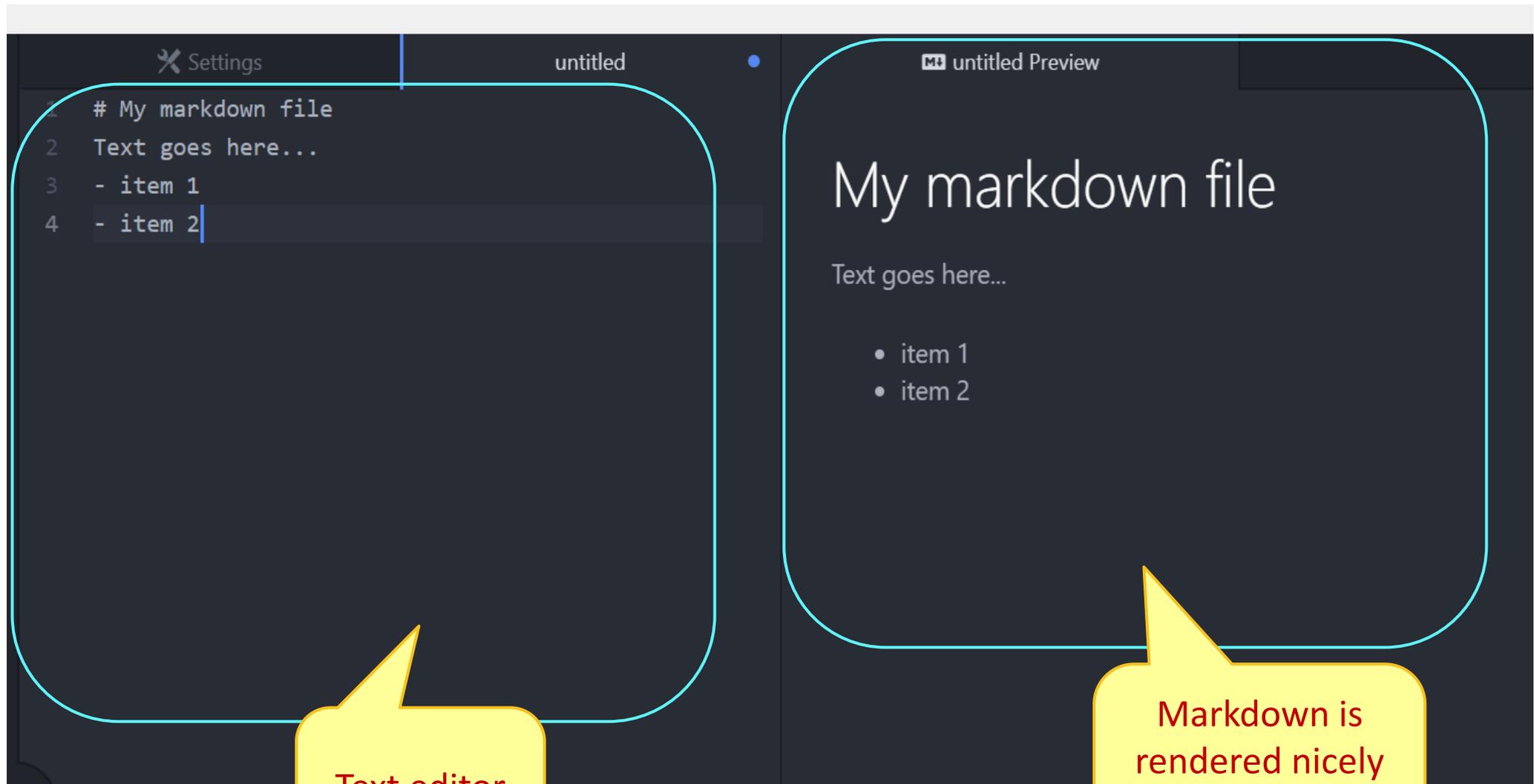
So that aliasing takes effect
immediately (without having
to log back in)

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:
 - ◆ install "**Markdown viewer**" **extension** on Chrome (link on Learning Resources page)

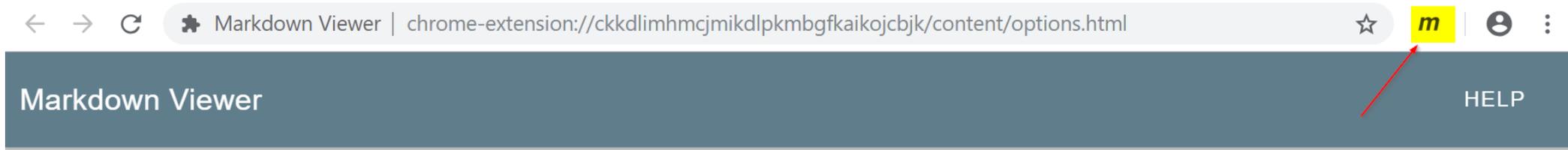


If using **Atom**,
you should first enable
Markdown preview



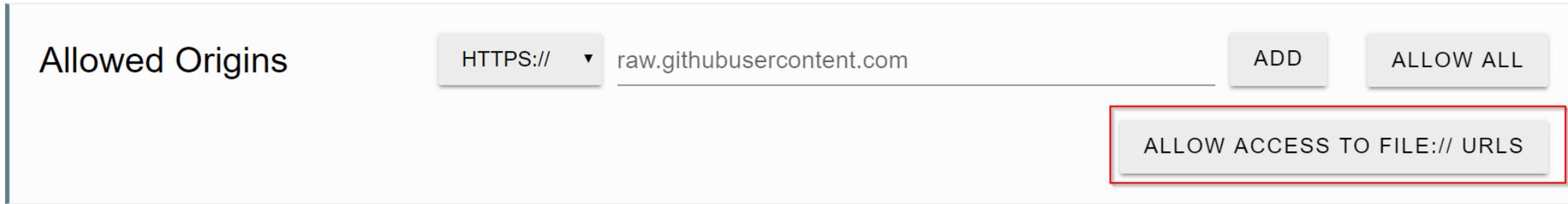
Text editor window

Markdown is rendered nicely in the preview window



Markdown Viewer

HELP



Allowed Origins

HTTPS://

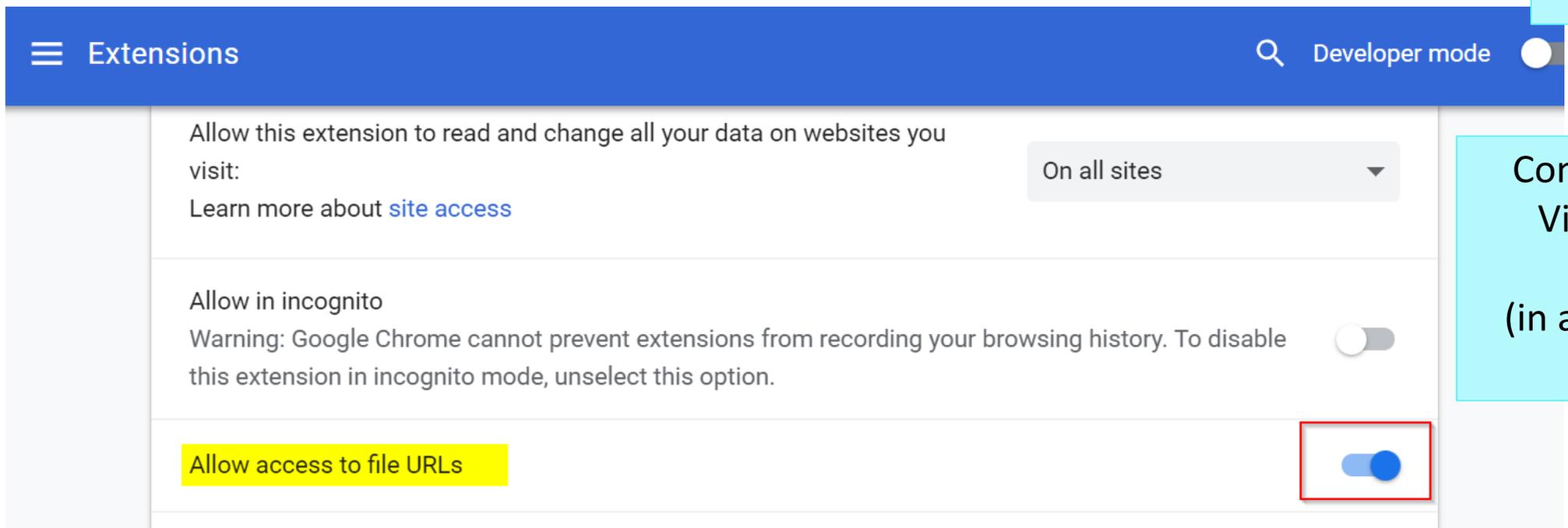
raw.githubusercontent.com

ADD

ALLOW ALL

ALLOW ACCESS TO FILE:// URLs

Chrome browser



Extensions

Developer mode

Allow this extension to read and change all your data on websites you visit:
Learn more about [site access](#)

On all sites

Allow in incognito
Warning: Google Chrome cannot prevent extensions from recording your browsing history. To disable this extension in incognito mode, unselect this option.



Allow access to file URLs

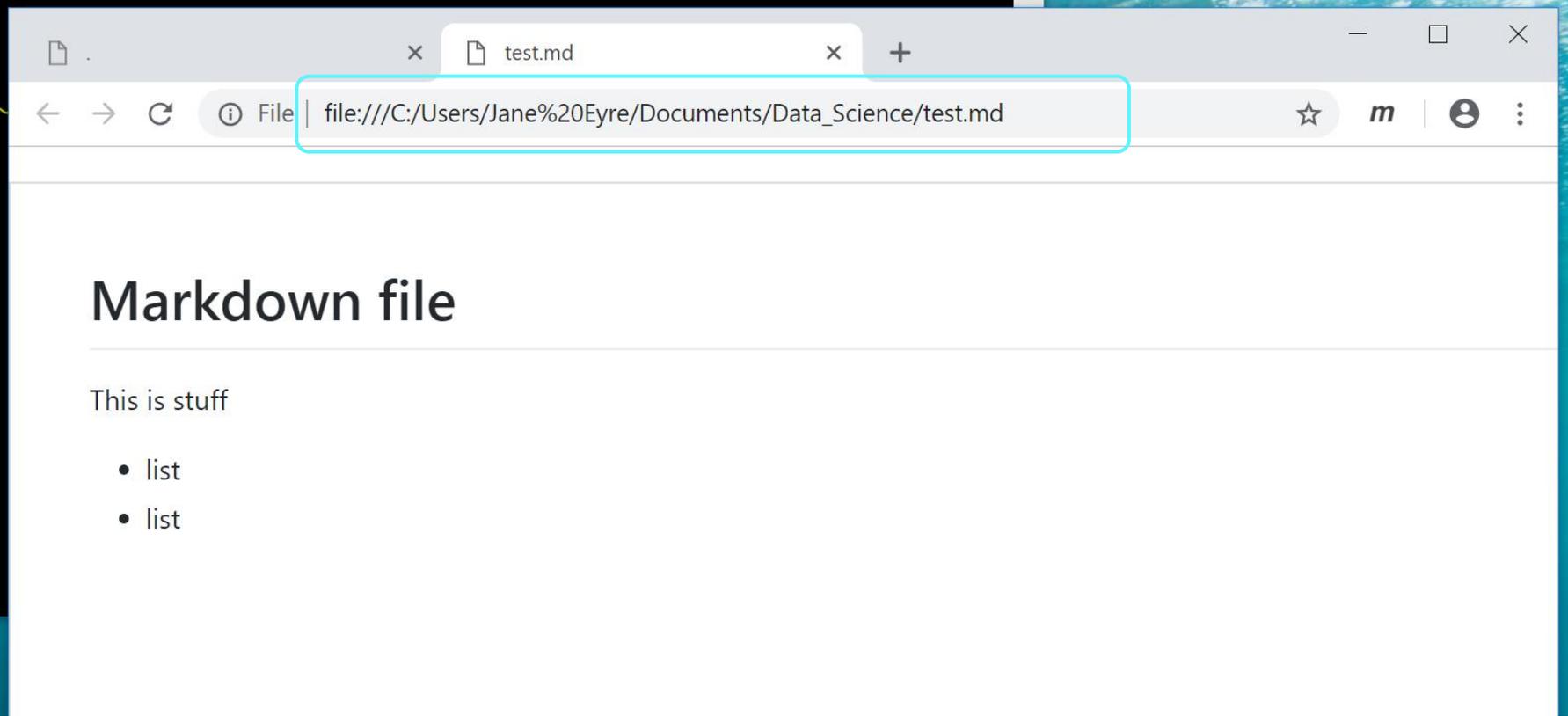


Configuring Markdown Viewer extension to allow **file URLs** (in addition to the usual https:// etc.)

```
MINGW64:/c/Users/Jane Eyre/Documents/Data_Science
Jane Eyre@X1Yoga MINGW64 ~/Documents/Data_Science
$ ls
Class-Exercise-Repo/ HW1-Repo/ languages/ practice-repo/ test.md
Jane Eyre@X1Yoga MINGW64 ~/Documents/Data_Science
$ chrome ./test.md &
[4] 22292
Jane Eyre@X1Yoga MINGW64 ~
$
```

The markdown file I'm trying to open in Chrome

Success!
Adding ./ prefix
(meaning the file is in the current directory)
forces file:// URL



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 ETS Corpus

- ▶ <https://naraehan.github.io/Data-Science-for-Linguists-2021/hw2>
- ▶ Corpus distributed via a private GitHub repo "Licensed-Data-Sets"
 - ◆ Clone directly to your laptop: no need to fork (why?)
- ▶ Submission via your own personal & private repo (shared with Na-Rae & Joey)

Wrapping up

- ▶ **To-do #4**
 - ◆ Continuation of pandas study notes
 - ◆ Try a classmate's CSV
- ▶ **Next class:**
 - ◆ Pop music lyrics, round 2
- ▶ **Coming up: Project ideas**
 - ◆ Due next Tuesday.
 - ◆ Not sure? Come chat with me and Joey
- ▶ **Coming up: HW2 Process ETS Corpus**
 - ◆ Due next Thursday.
 - ◆ This one is much bigger than HW1. Start sooner!