Command line and data essentials

Bernease Herman^{1,3}, Joseph Hellerstein^{1,3}, David Beck^{1,2,3}, Daniel Jones³

¹eScience Institute ²Chemical Engineering ³Computer Science

September 27, 2019



Survey the room

Raise your hand if...

Your primary computer's operating system is:

Linux or Mac

Windows 10

Earlier version of Windows than 10

You're inexperienced / never used a Unix command line

You've used Unix command line, but not made Bash script

You're experienced in Unix command line, Bash scripts





Agenda

- 1. Prior lecture today: Course introduction
- 2. Installation and setup instructions
- 3. Hands-on practice in using Unix command line
- 4. Hands-on practice creating Bash scripts





Confirm access to Unix command line

Linux/Mac: "Terminal"

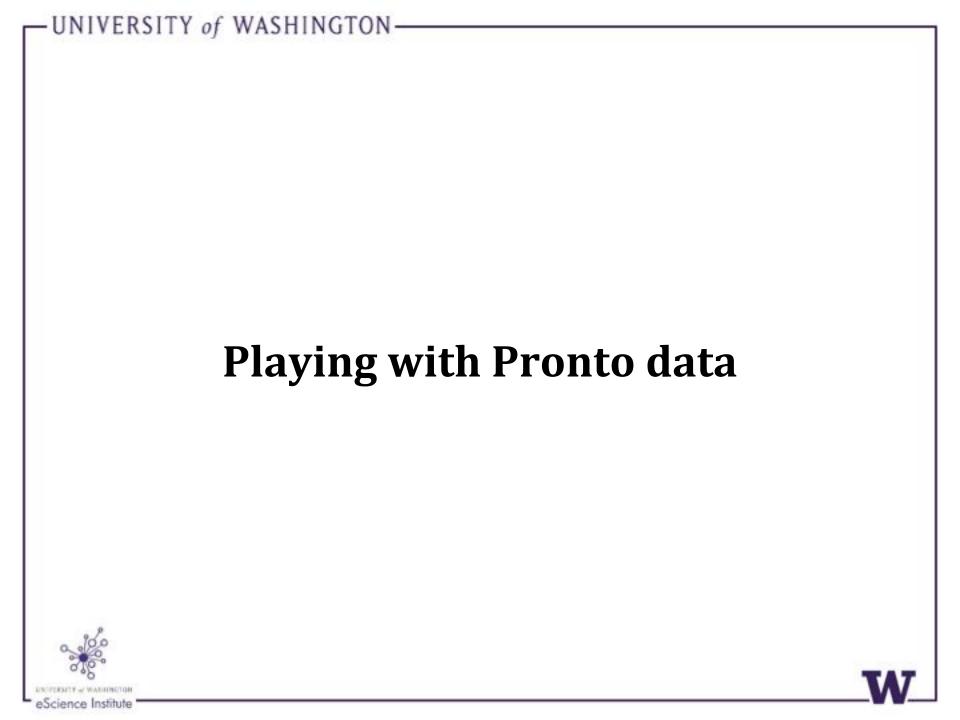
Windows 10: Linux subsystem, e.g. "Ubuntu"

Windows Pre-10: "Git Bash" or "Cygwin"





Introduction to Unix command line	3
-UNIVERSITY of WASHINGTON-	





Open Data

Here you'll find Pronto's trip data for public use. Whether you're a designer, developer, or just plain curious, feel free to bring it to life!

The Data

Each trip is anonymized and includes:

- Bike number
- · Trip start day & time
- · Trip end day & time
- · Trip start station
- Trip end station
- · Rider Type: Annual Member or Short-Term (24-Hour or 3-Day) Pass Holder
- · Annual Member trips will also include the member's gender and year of birth

The data set also includes:

- Weather information per day (using 98101 zip code)
- · Bike and dock availability per minute per station

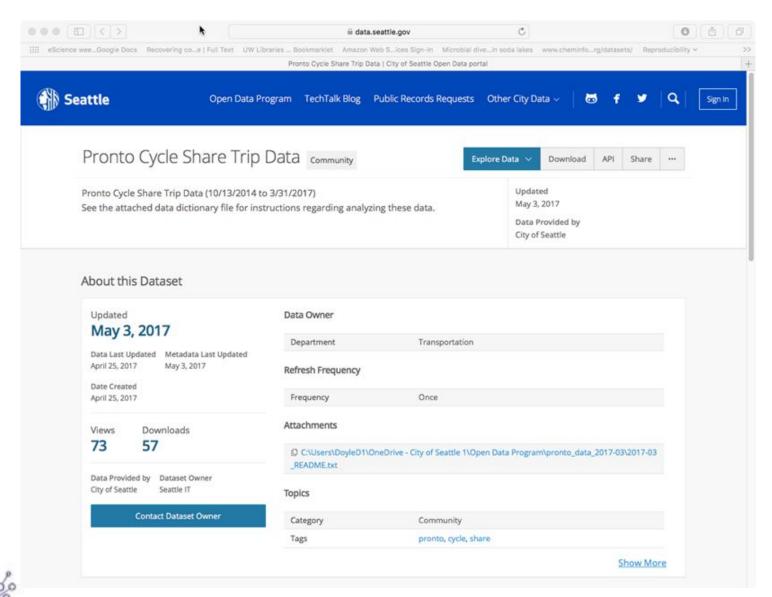
Click the buttons on the right side of the page to download the available data sets.

Additionally, you can always use our live JSON feed.





UNIVERSITY of WASHINGTON-

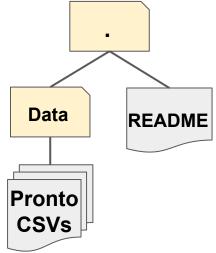






Activity: Storing data in directories

Create directory structure, download Pronto data from internet, place data into directory



Link:

https://data.seattle.gov/api/views/tw7j-dfaw/rows.csv?accessType=DOWNLOAD



Variable	Data Type	Units
trip_id	Int64	
starttime	datetime64	
stoptime	datetime64	
bikeid	string	Coded (e.g., "SEA00298")
tripduration	float	Seconds
from_station_name	string	Address
to_station_name	string	Address
from_station_id	string	Coded (e.g., "PS-04")
to_station_id	string	Coded (e.g., "PS-04")
usertype	string	Coded (e.g., "Annual Member")
gender	string	Coded (e.g., "Male")



Fields in Pronto data



UNIVERSITY of WASHINGTON-

Introduction to pipes, filters, and Bash scripts



