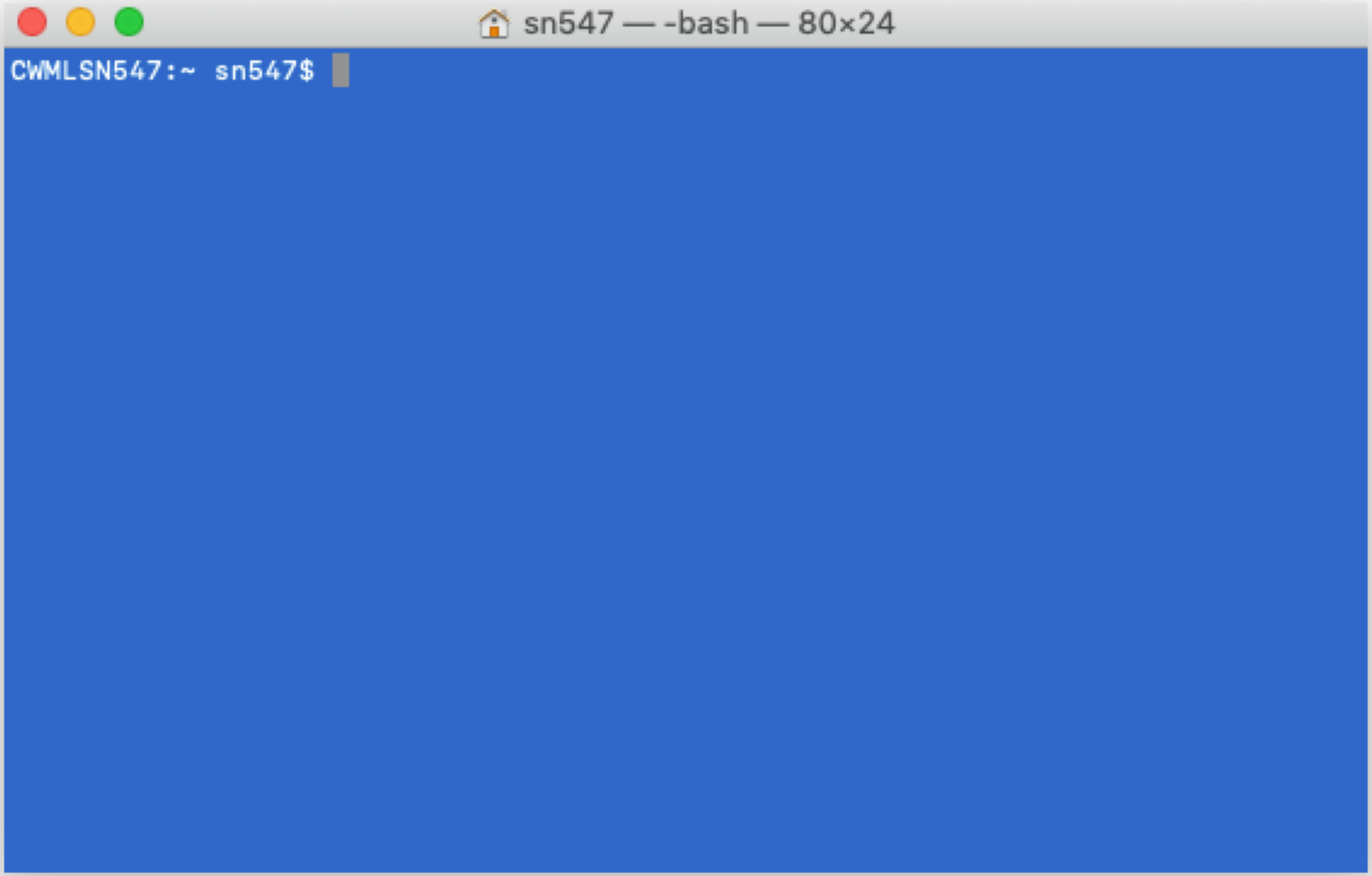
Unix Shell: Working with Data

Fall 2019

Sawyer Newman
Data Librarian for the Health Sciences

Yale *Harvey Cushing / John Hay Whitney Medical Library*

What is the Shell?

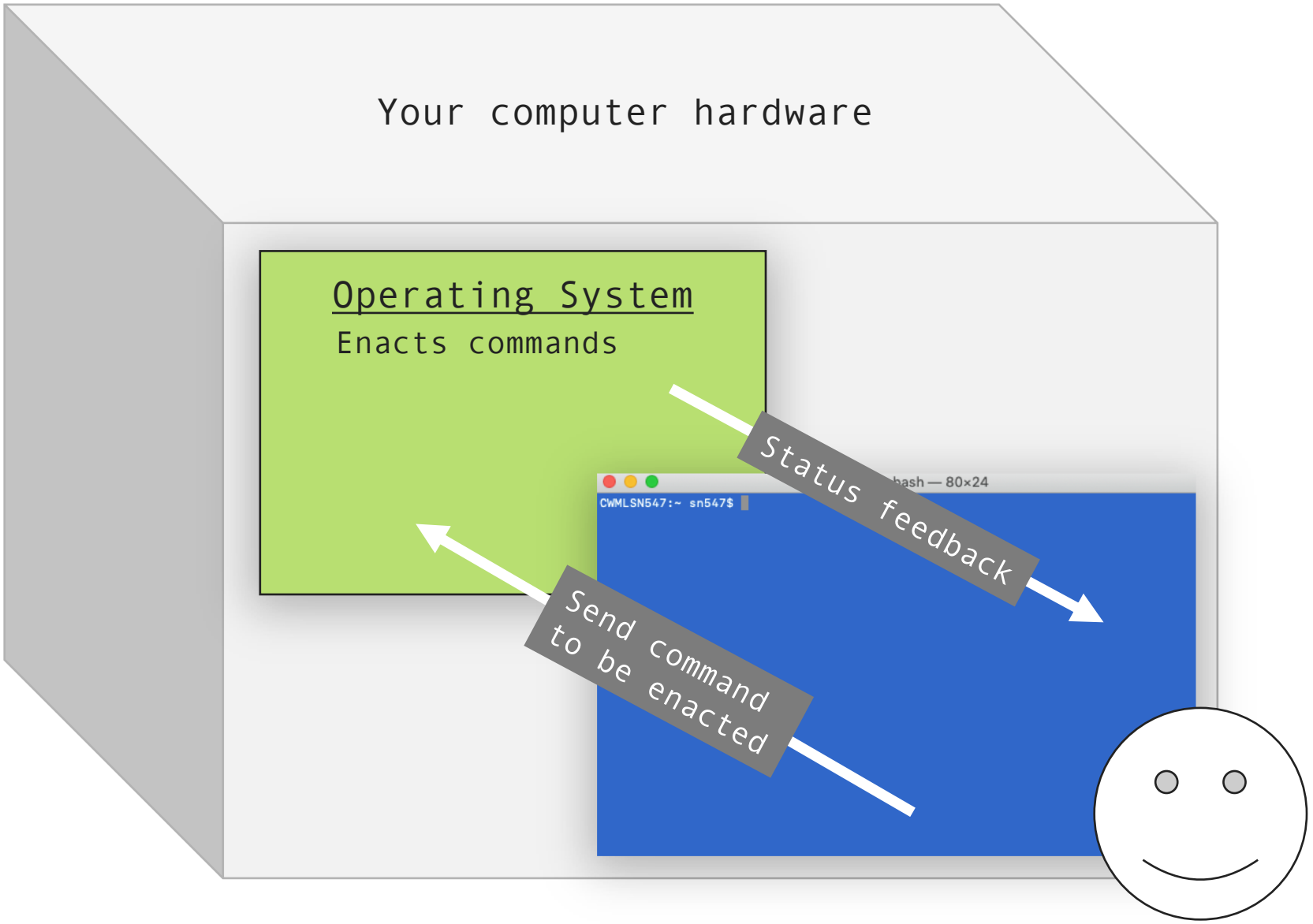


Shell Uses

- Obtain data from websites, APIs, databases, and spreadsheets
- Clean tabular and hierarchical data formats
- Explore, analyze and visualize data
- Manage your data workflow
- Create reusable tools
- Leverage python and R across your data

Relevant Terms

- Command-line tools: Typed command prompts
- Terminal: The application where commands are typed
- Shell: Where the command is sent to by the terminal to be interpreted
- Bash: Bourne Again SHell
- Operating system: Executes commands and connects hardware


Your computer hardware
Operating System
Enacts commands
Status feedback
Send command to be enacted
CWMLSN547:~ sn547$
User
Types commands

Why are we using the Shell?

- HPC clusters at Yale clusters run on Linux
- 95% of the top 500 supercomputers are running Linux

References for Unix/Linux commands

- Unix shell commands for windows: link
- Search for Unix command explanations: explainshell.com
- Find bugs in your shell scripts: shellcheck.net

Continue your learning

- Yale access to Linkedin Learning training videos
- Yale Center for Research Computing training videos
- Yale Center for Research Computing Practical Introduction to Linux
- O'Reilly Textbook Data Science at the Command Line (Jeroen Janssens 2019)

Research Data Services @ CWML

Data Management Plans

Data Tools & Software

Data Policy Guidance

Find Datasets

Data Storage

Best Practices & Definitions

Data Support Groups at Yale

Consultations & Drop-Ins

medicaldata@yale.edu

Demo topics

- File navigation
- Creating new files and folders
- Moving files
- Copying files
- READMe files
- Finding data within files
- Merging and collecting data from files