IBiS Computational Biology Bootcamp

September 9 - 11, 2014

1 to 5 PM

Pancoe 1401

You guys are guinea pigs.

Basic introduction

Feedback

Help each other!

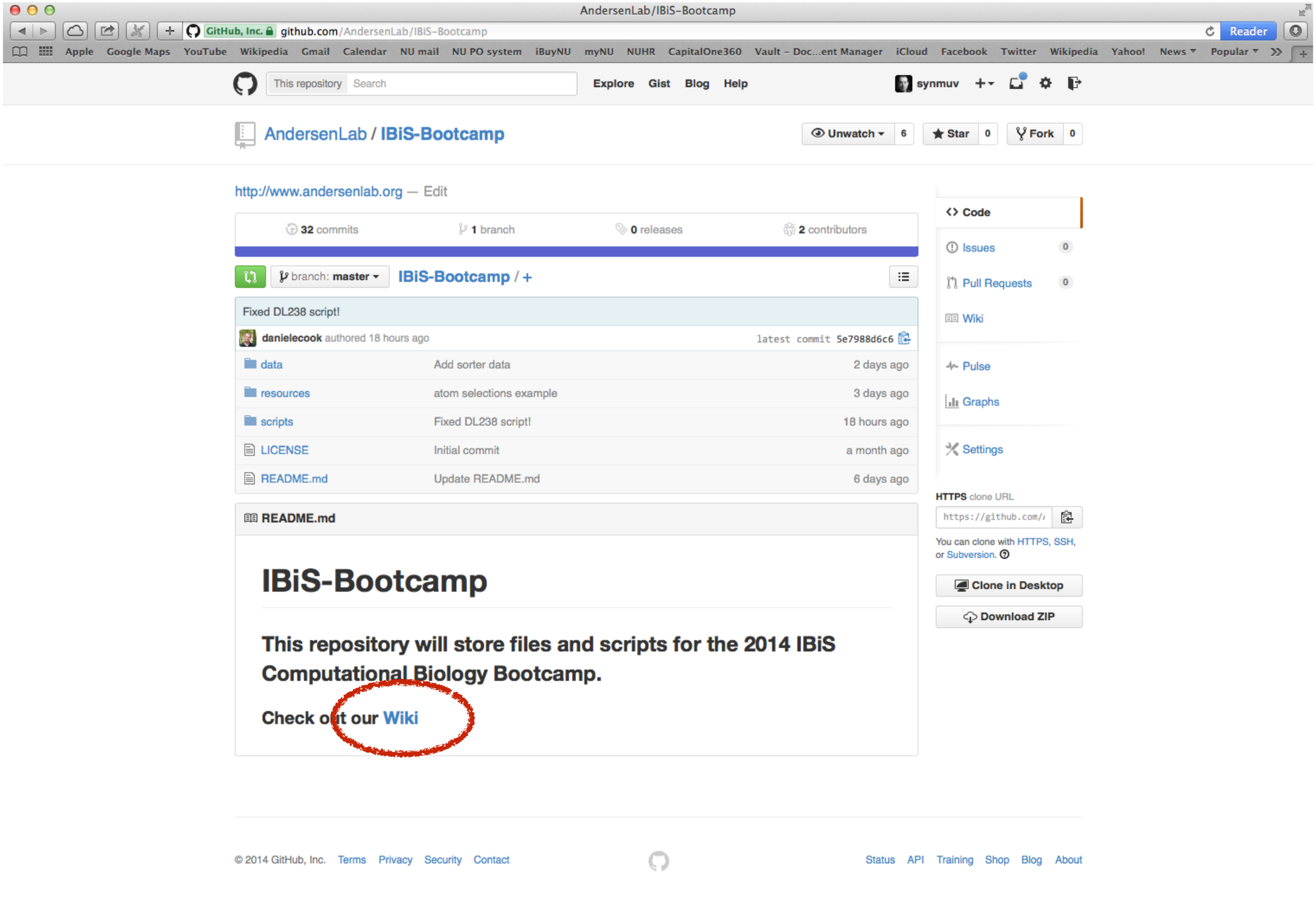
Unsolicited advice about computational work

Don’t be afraid

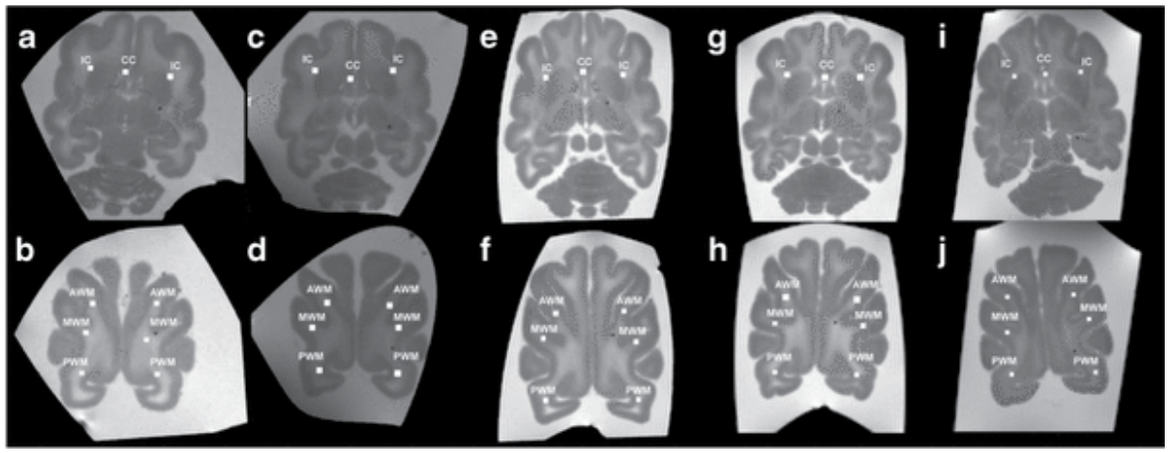
It will be rewarding

Stop using Excel

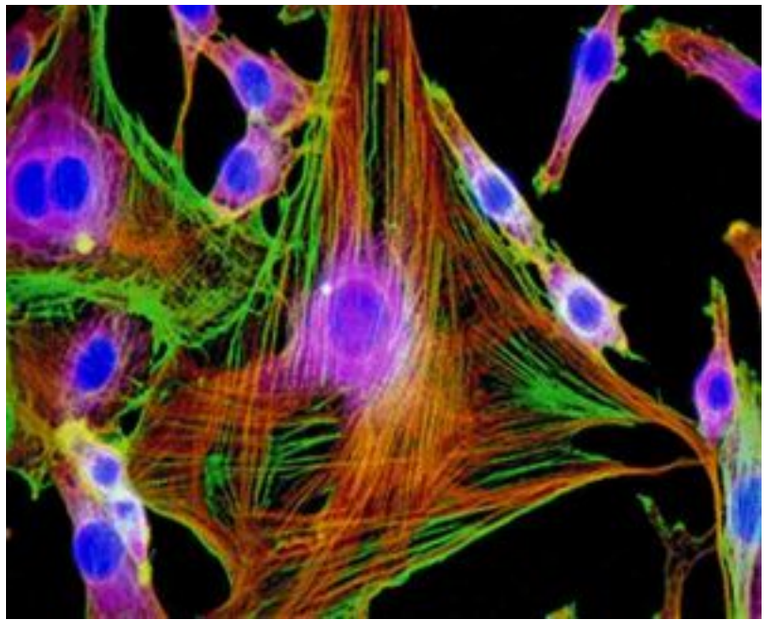
Use our wiki


AndersenLab/IBiS-Bootcamp
GitHub, Inc. github.com/AndersenLab/IBiS-Bootcamp
Reader
Apple Google Maps YouTube Wikipedia Gmail Calendar NU mail NU PO system iBuyNU myNU NUHR CapitalOne360 Vault - Doc...ent Manager iCloud Facebook Twitter Wikipedia Yahoo! News Popular
This repository Search
Explore Gist Blog Help
synmuv
AndersenLab / IBiS-Bootcamp
Unwatch 6
Star 0
Fork 0
http://www.andersenlab.org — Edit
32 commits
1 branch
0 releases
2 contributors
branch: master
IBiS-Bootcamp / +
Fixed DL238 script!
danielecook authored 18 hours ago
latest commit 5e7988d6c6
data Add sorter data 2 days ago
resources atom selections example 3 days ago
scripts Fixed DL238 script! 18 hours ago
LICENSE Initial commit a month ago
README.md Update README.md 6 days ago
README.md
IBiS-Bootcamp
This repository will store files and scripts for the 2014 IBiS Computational Biology Bootcamp.
Check out our Wiki
Code
Issues 0
Pull Requests 0
Wiki
Pulse
Graphs
Settings
HTTPS clone URL
https://github.com/
You can clone with HTTPS, SSH, or Subversion.
Clone in Desktop
Download ZIP
© 2014 GitHub, Inc. Terms Privacy Security Contact
Status API Training Shop Blog About

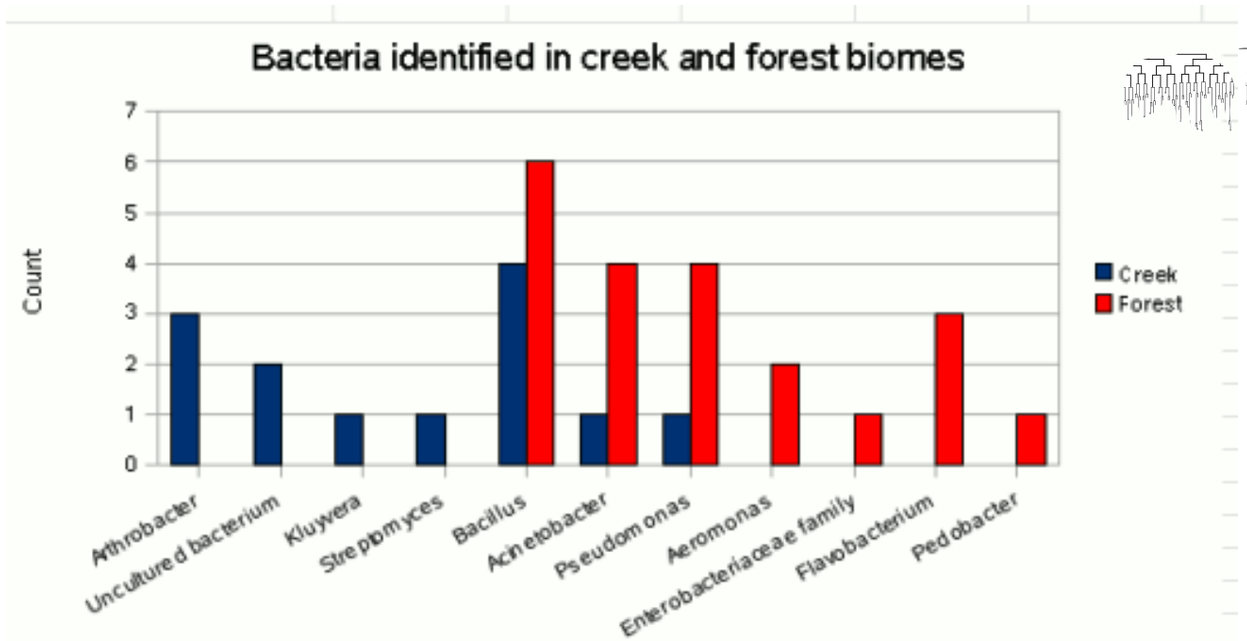
The days of descriptive biology are [almost] over



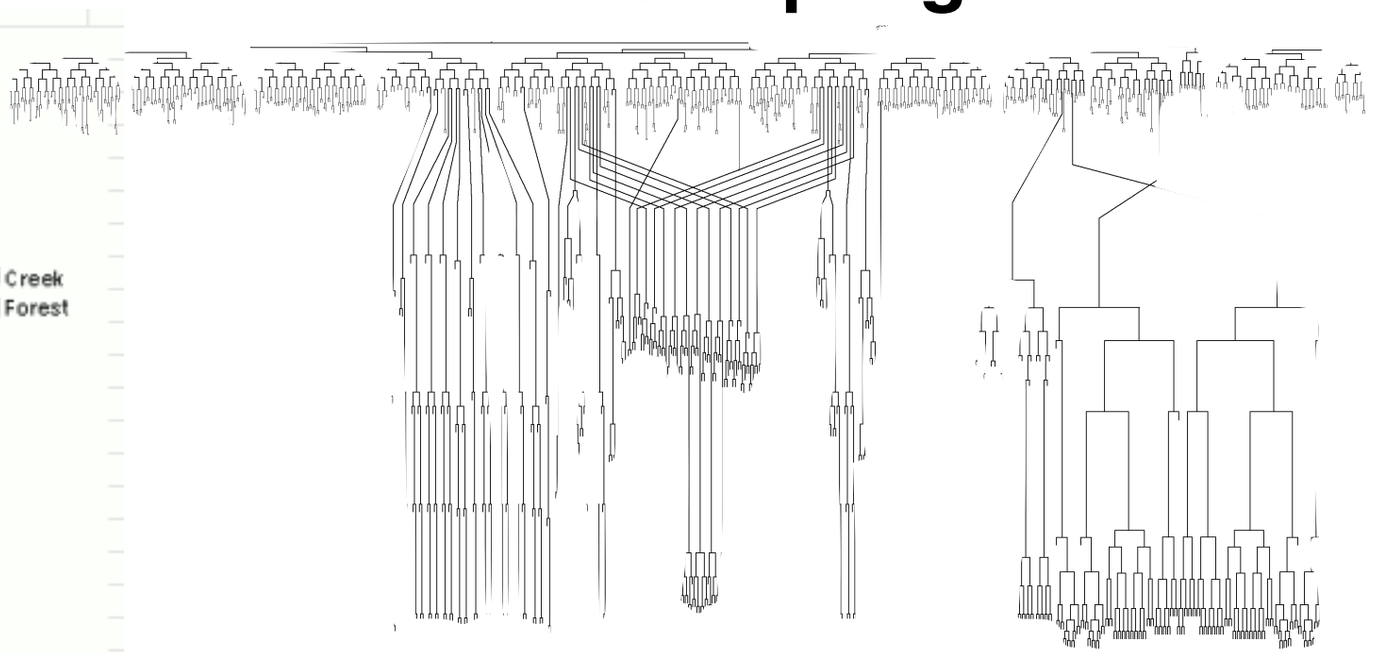
Reproducibilty?



Sampling?

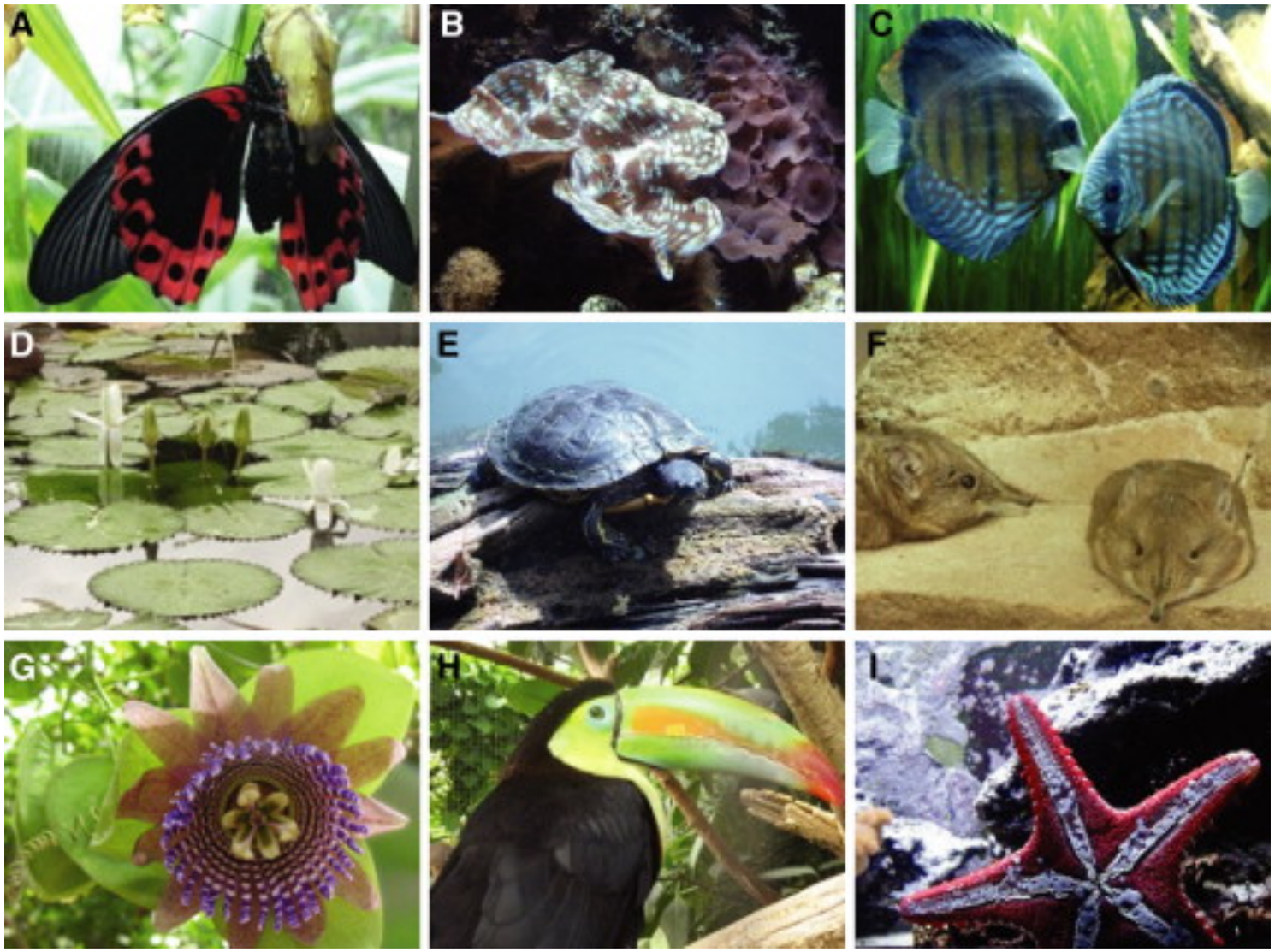


Error?



Automation?

NOW is the most exciting time to be a biologist



Any biologist can collect data.

Few biologists can process and analyze them.

Working with biological data requires the ability to clean data.

The New York Times

For Big-Data Scientists, 'Janitor Work' Is Key Hurdle to Insights

By STEVE LOHR AUG. 17, 2014

Monica Rogati, Jawbone's vice president for data science, with Brian Wilt, a senior data scientist.
Peter DaSilva for The New York Times

FORTUNE

Big data's dirty problem

by Verne Kopytoff @vkopytoff JUNE 30, 2014, 10:58 AM EDT

PHYS.ORG

Science is in a reproducibility crisis: How do we resolve it?

Sep 20, 2013 by Fiona Fidler & Ascelin Gordon, The Conversation

Validation BY THE SCIENCE EXCHANGE NETWORK About Projects Press Contact

Validating key experimental results via independent replication

Learn more »

Reproducibility initiatives will invite increased scrutiny into data cleaning methods.

Course outline

Day #1: Basic command line interface and reproducible research

Day #2: Command line tools and sequence alignment, Intro to R and RStudio

Day #3: R, data manipulation, and plotting

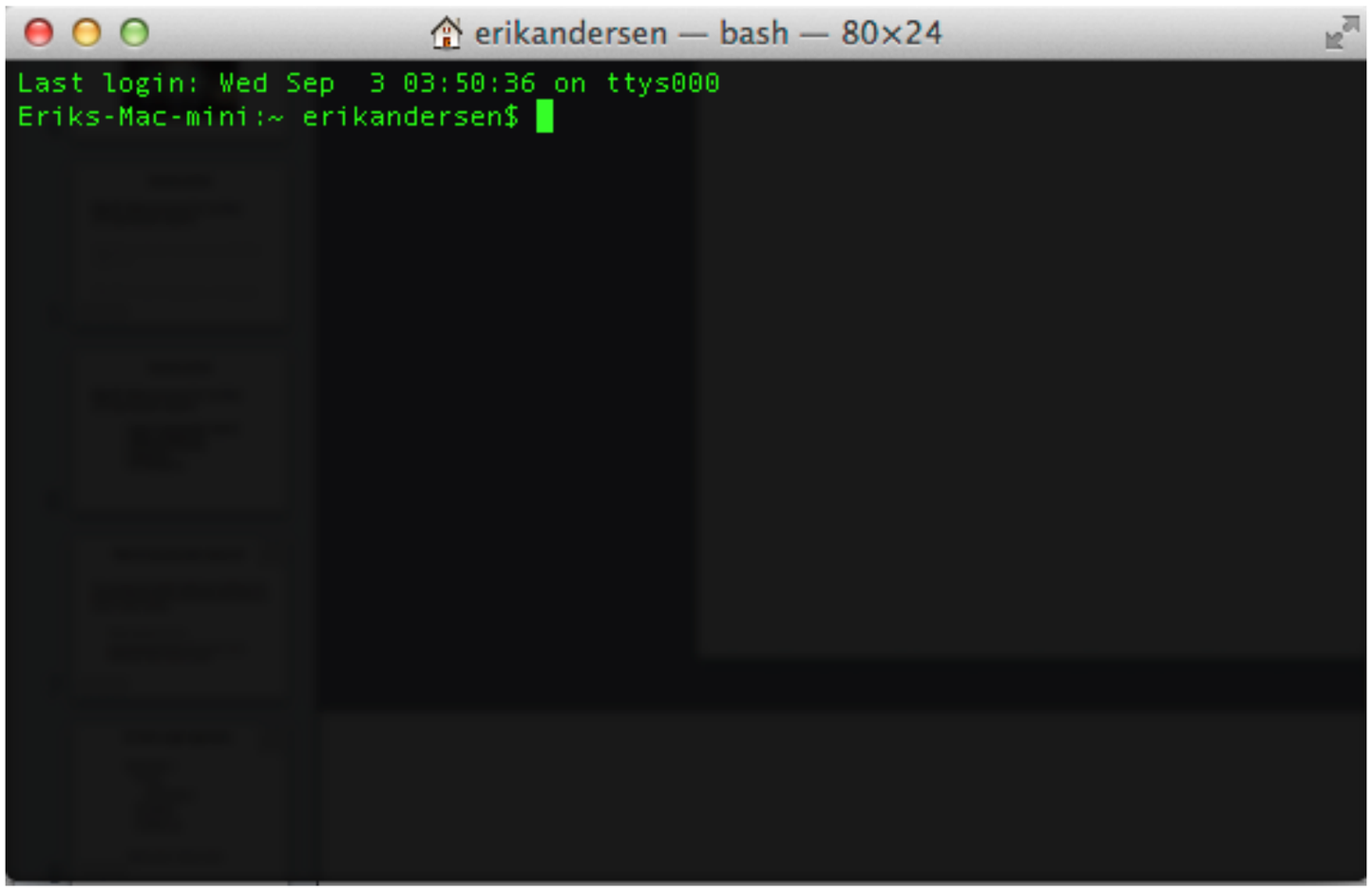
Course outline

Day #1: Basic command line interface and reproducible research

- **Software installation**
- **Intro to reproducible research**
- **Basic command line**
- **Directory structures**
- **Markdown**
- **Git and github**

Notes vs `what you type`

Command-line interface on a Mac



1. Go to Utilities (cmd-shift-u in finder)
2. Find Terminal and move it to your dock
3. Click on Terminal to open a window

Homebrew

The missing package manager for OS X

- Package managers make it easy to manage the software installed on your computer.
- Package ≈ Software ≈ App
- Software can have a lot of dependencies, or *other* required pieces of software. Homebrew takes care of this.
- *1000s* of packages available.

Software installation

Homebrew - a package manager for command line

1. In your browser, go to brew.sh
2. Copy the link at the bottom
3. Go to your Terminal window
4. Paste the link and press enter
5. Follow the prompts to install
6. Run `brew doctor`

Homebrew Commands

Install a package	`brew install package_name`
Uninstall a package	`brew uninstall package_name`
List installed packages	`brew list`
Search for packages	`brew search package_name`
Update Software List	`brew update`

Browse packages at braumeister.org

Homebrew

The missing package manager for OS X

General Purpose software
is already included.

Science Software

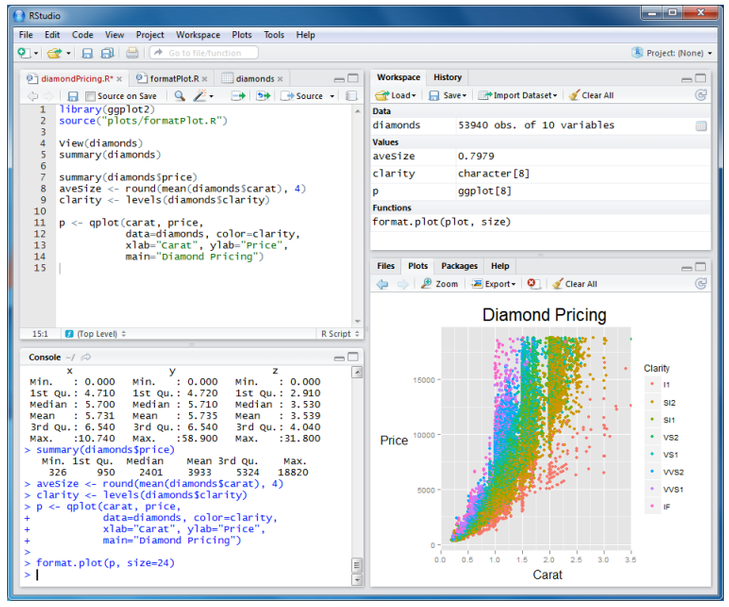
To install science related
software, we have to ‘tap’
into the science-software repository.

```
brew tap homebrew/science
```

Homebrew

Cask

Graphical User Interface (GUI)



“Point and Click”

But Wait! It is possible to manage GUI software with ‘cask’, an extension of homebrew.

```
brew install caskroom/cask/brew-cask
```

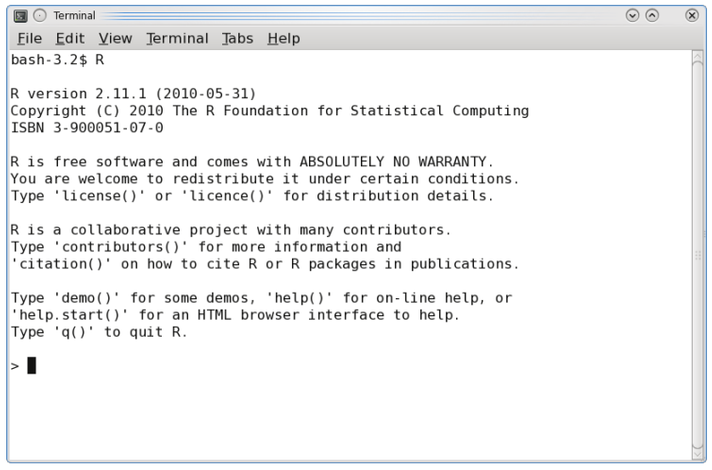
Homebrew

The missing package manager for OS X

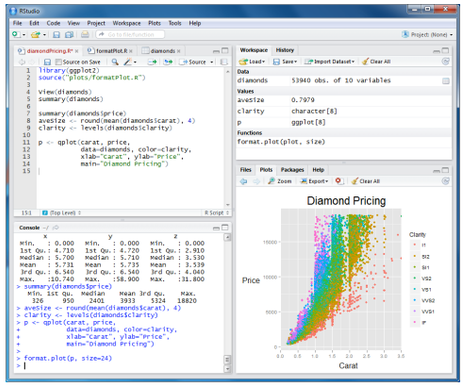
Science Software

General Purpose

Homebrew Cask



Command line software



GUI software

Software installation

Go to www.GitHub.com/AndersenLab/IBiS-Bootcamp/Wiki/Installing-Required-Software

```
#!bin/bash

# This script install the required software!

# Install Homebrew, if you haven't as a part of the lecture
#ruby -e "$(curl -fsSL https://raw.github.com/Homebrew/homebrew/go/install)"
#brew tap homebrew/science

# Cask
#brew install caskroom/cask/brew-cask

# Sequence Analysis
brew install --devel samtools
brew install bcftools
brew install bwa
brew install igv

# wget
brew install wget

# Atom
brew cask install atom --appdir=/Applications

# ql markdown
brew cask install qlmarkdown

# R + Rstudio
brew install R
brew cask install rstudio --appdir=/Applications
```

What is reproducible research?

The concept that scientific claims are published with data and software code so that others can verify and build on those findings.

Please experience more at:

www.coursera.org Reproducible research course
GitHub.com Help and training pages

It’s time to get organized…

```
MyProject/
  data/
    raw/
    processed/
  scripts/
  results/
  readme.md
```

`MyProject/data/raw/`

Basic directory commands for UNIX

`man` = read the manual for a command

`ls` = list the contents of a directory

`pwd` = display the present working directory

`cd` = change directory

`mkdir` = make a directory

`rmdir` = remove a directory

Further experiences:
http://freeengineer.org/learnUNIXin10minutes.html
http://www.ee.surrey.ac.uk/Teaching/Unix/
or just google: “unix command line tools”

Keyboard shortcuts for UNIX

Press *tab* to complete a filename

Navigation shortcuts
ctrl + a : go to beginning of line
ctrl + e : go to end of line
alt + <- (->) : skip between delimiters
alt + delete : delete previous word
command + up (down) arrow: beginning (or end) of text file

Selection shortcuts
option + shift + <- (->): highlight text to next delimiter
shift + up (down) arrow : highlight previous (next) line
command + shift <- (->) : highlight to beginning (or end of line)
command + shift + up (down) arrow : highlight to beginning (or end of text)

http://help.filemaker.com/app/answers/detail/a_id/6493/~/text-keyboard-shortcuts-(mac-os)

Let's each make the directory structure using the command line.

```
MyProject/
   data/
      raw/
      processed/
   scripts/
   results/
   mistakes_happen/
```

`man` = read the manual for a command

`ls` = list the contents of a directory

`pwd` = display the present working directory

`cd` = change directory

`mkdir` = make a directory

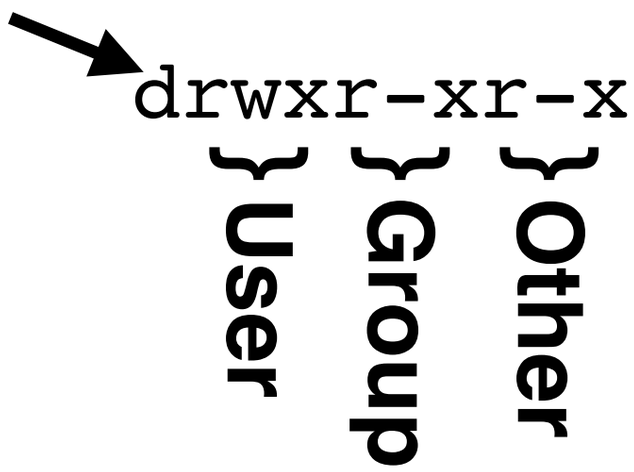
`rmdir` = remove a directory

`cd ..` allows you to go back one directory in tree

Permissions (modes) matter.

Go to your `data` directory, type `ls -la`

Filetype (- , d, or c)



`r` = read
`w` = write
`x` = execute

`chmod` changes the mode

Symbolic Notation	Octal Notation	English
`----------`	0000	no permissions
`---x--x--x`	0111	execute
`--w--w--w-`	0222	write
`--wx-wx-wx`	0333	write & execute
`-r--r--r--`	0444	read
`-r-xr-xr-x`	0555	read & execute
`-rw-rw-rw-`	0666	read & write
`-rwxrwxrwx`	0777	read, write, & execute

What about that readme.md file?

md = markdown

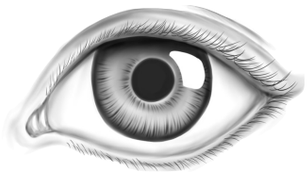
Markdown is the emerging standard for software documentation.

Easily written, read, and translated.

Portable to HTML, PDF, etc.

We need files that the computer and we can read.

Binary files
are read by computers



Markdown files

CSV or TSV files

Plain text files
are read by computers
and us

MS files
are read by
us

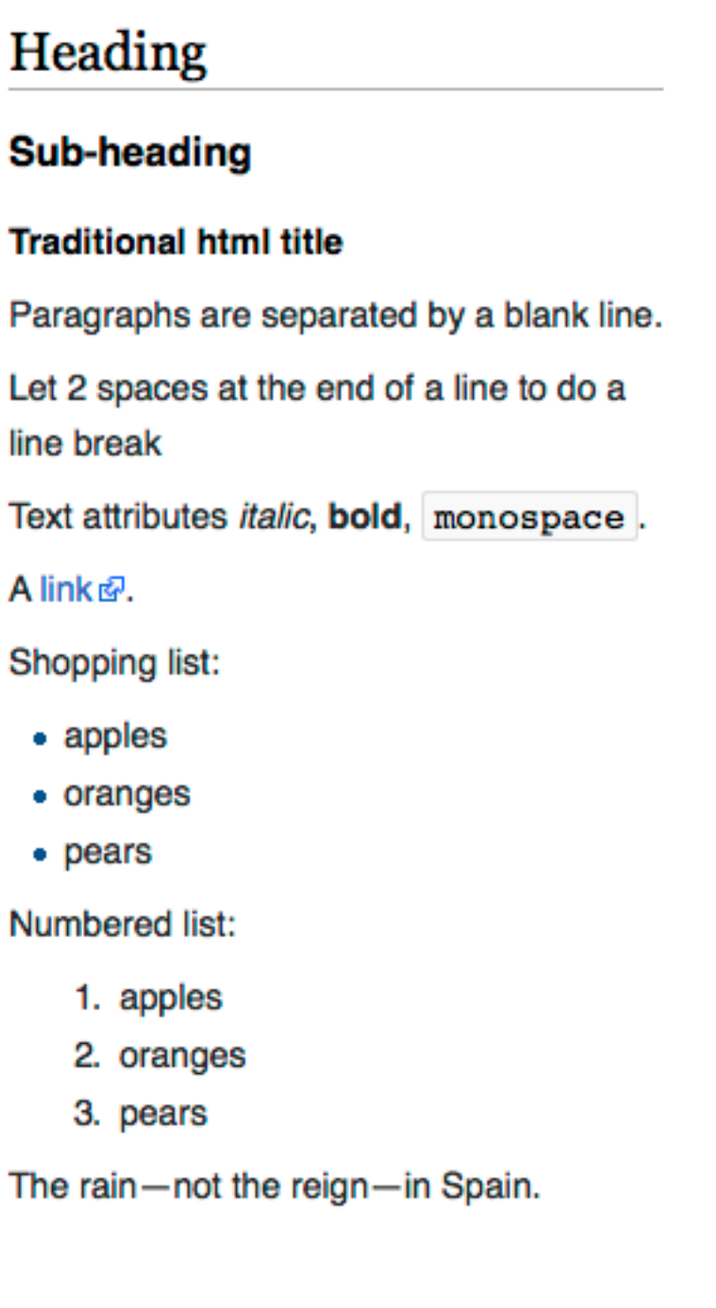
What about that readme.md file?

md = markdown

what you type

```
Heading
=======

Sub-heading
-----------

h3. Traditional html title

Paragraphs are separated
by a blank line.

Let 2 spaces at the end of a line to do a
line break

Text attributes *italic*,
**bold**, `monospace`.

A [link](http://example.com).
<<<   No space between ] and (  >>>

Shopping list:

  * apples
  * oranges
  * pears

Numbered list:

  1. apples
  2. oranges
  3. pears

The rain---not the reign---in
Spain.
```

what is rendered



Heading

Sub-heading

Traditional html title

Paragraphs are separated by a blank line.

Let 2 spaces at the end of a line to do a line break

Text attributes *italic*, **bold**, `monospace`.

A link.

Shopping list:

- apples
- oranges
- pears

Numbered list:

1. apples
2. oranges
3. pears

The rain—not the reign—in Spain.

Let’s make readme.md from the command line

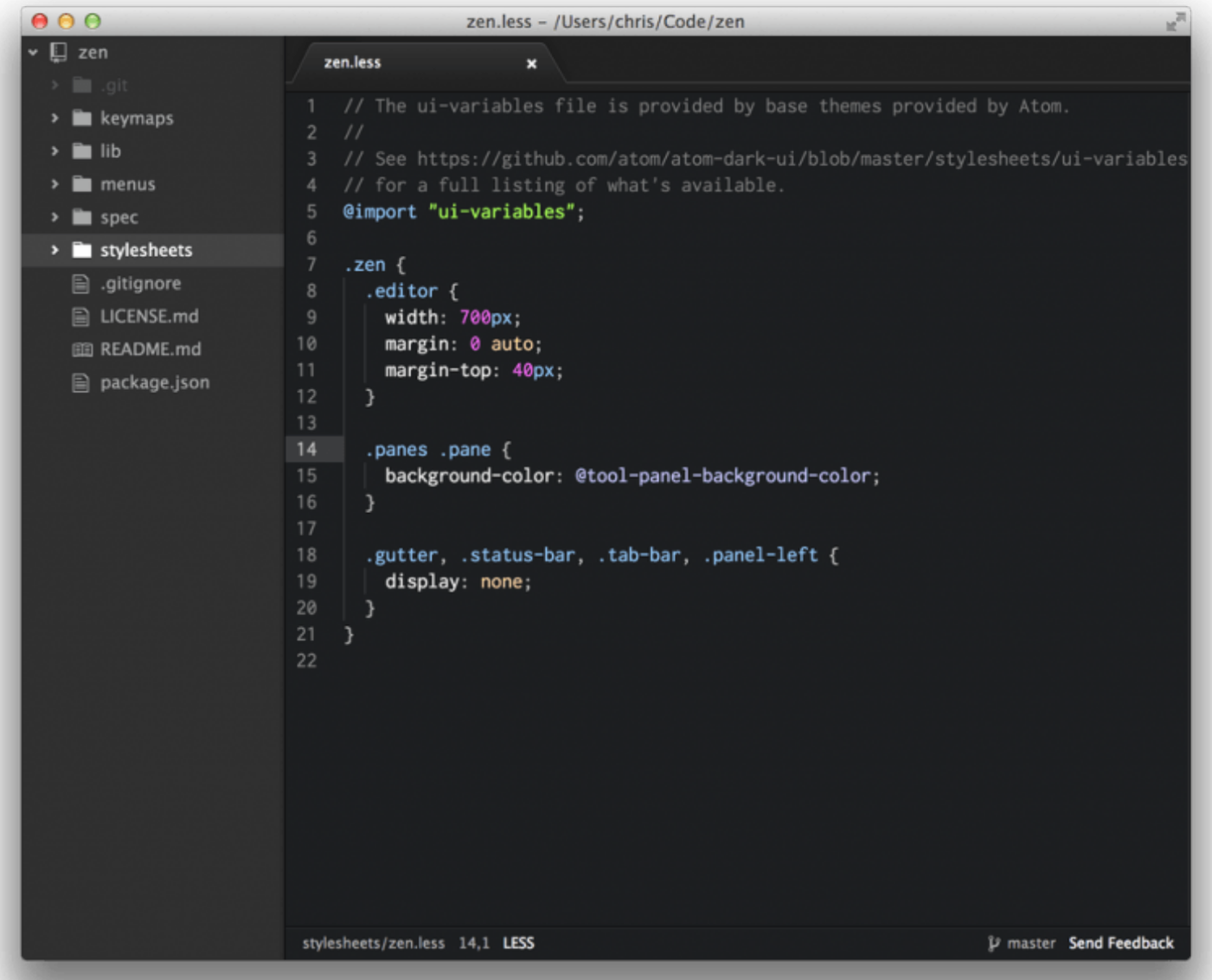
1. Go to MyProject directory
2. Type `nano readme.md`
3. Type:

```
This example *project* will teach us **a lot** about basic computational biology.
```

4. Press control O to save. Press enter.
5. Press control X to exit.

Further experiences:
https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet
http://daringfireball.net/projects/markdown/basics

ATOM

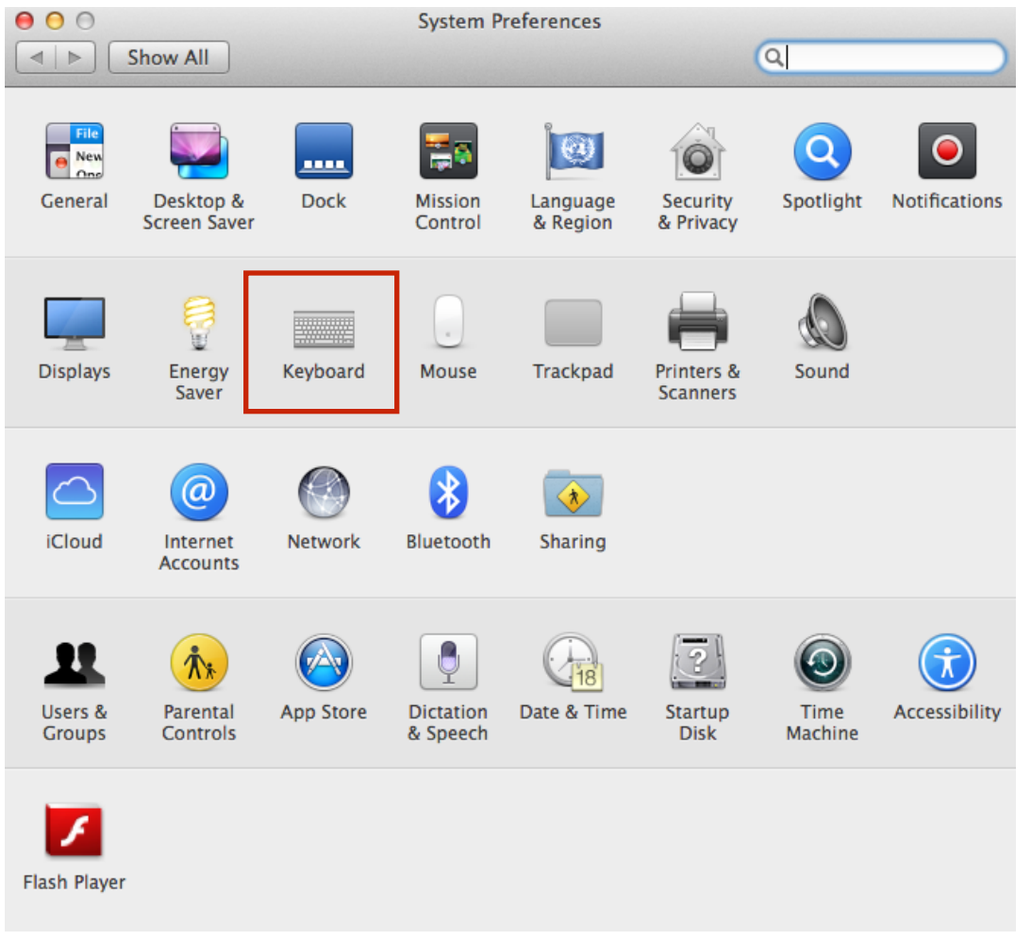


A plain text editor.
Very different from
Microsoft Word.

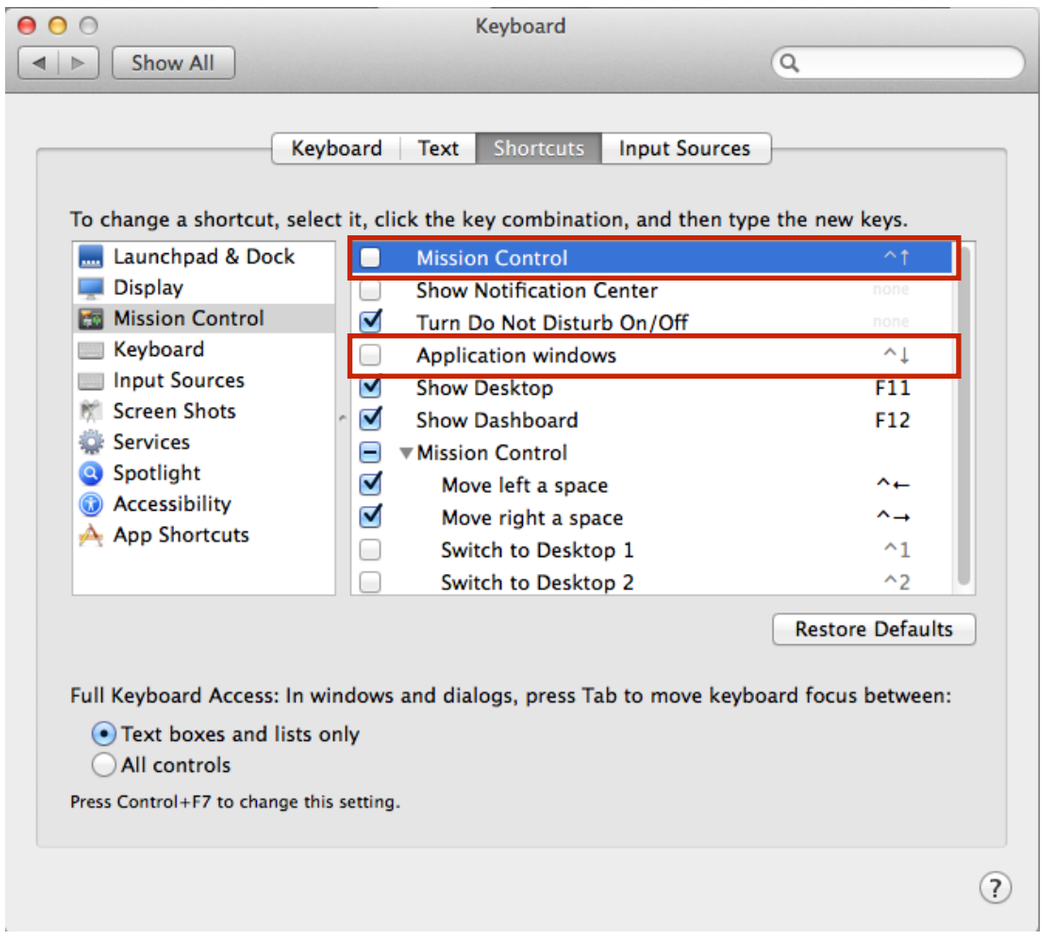
ATOM

1)

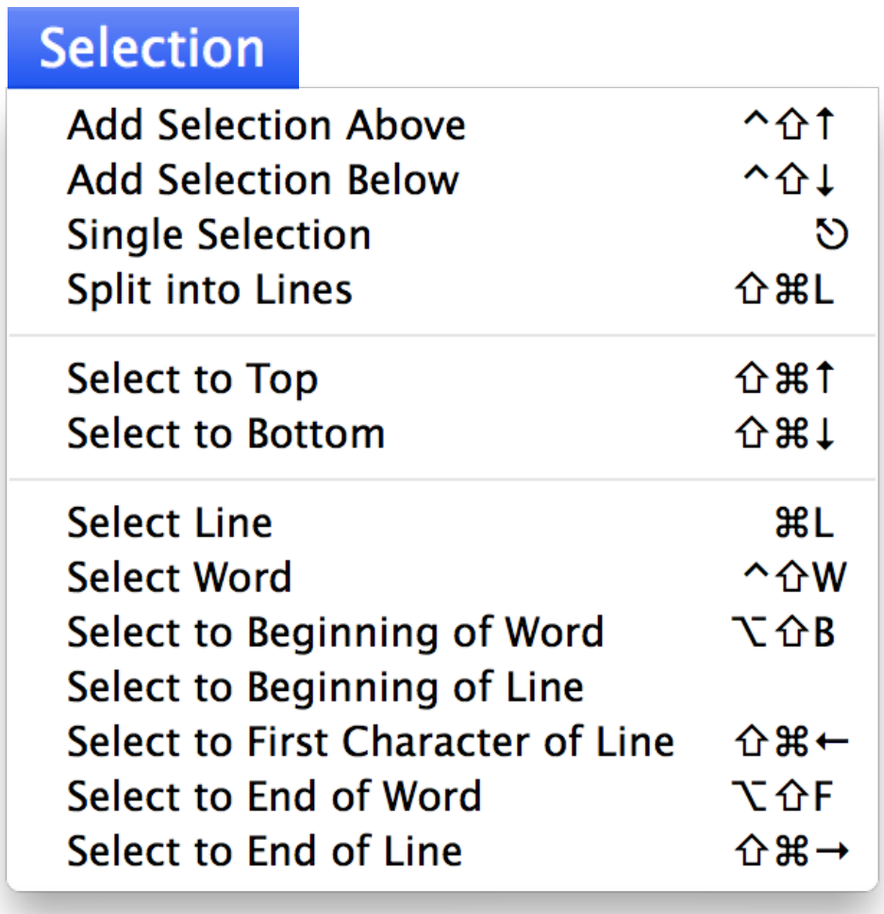
2)



3)



uncheck these two shortcuts

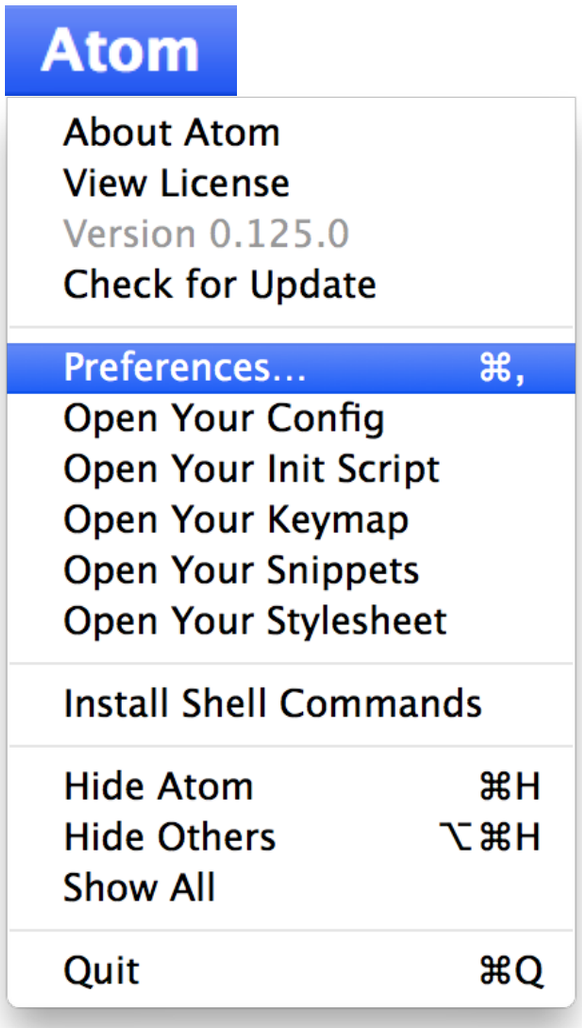


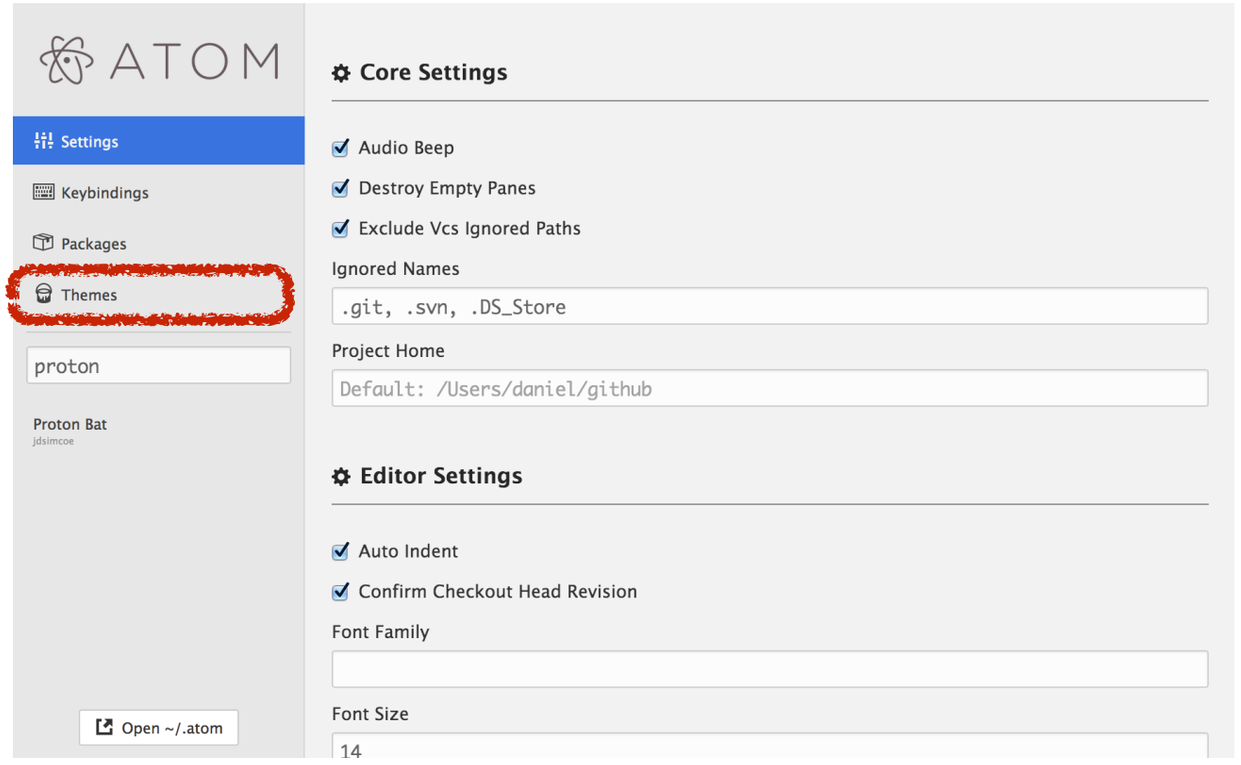
Open IBiS_bootcamp/resources/atom_selections_example.txt

Try the exercises, then try some of the options in the selection menu.

These functions and keyboard shortcuts are designed to *save you time*.

ATOM is customizable!





ATOM is customizable!

Choose a Theme

You can also style Atom by editing your stylesheet

UI Theme Atom Light

This styles the tabs, status bar, tree view, and dropdowns

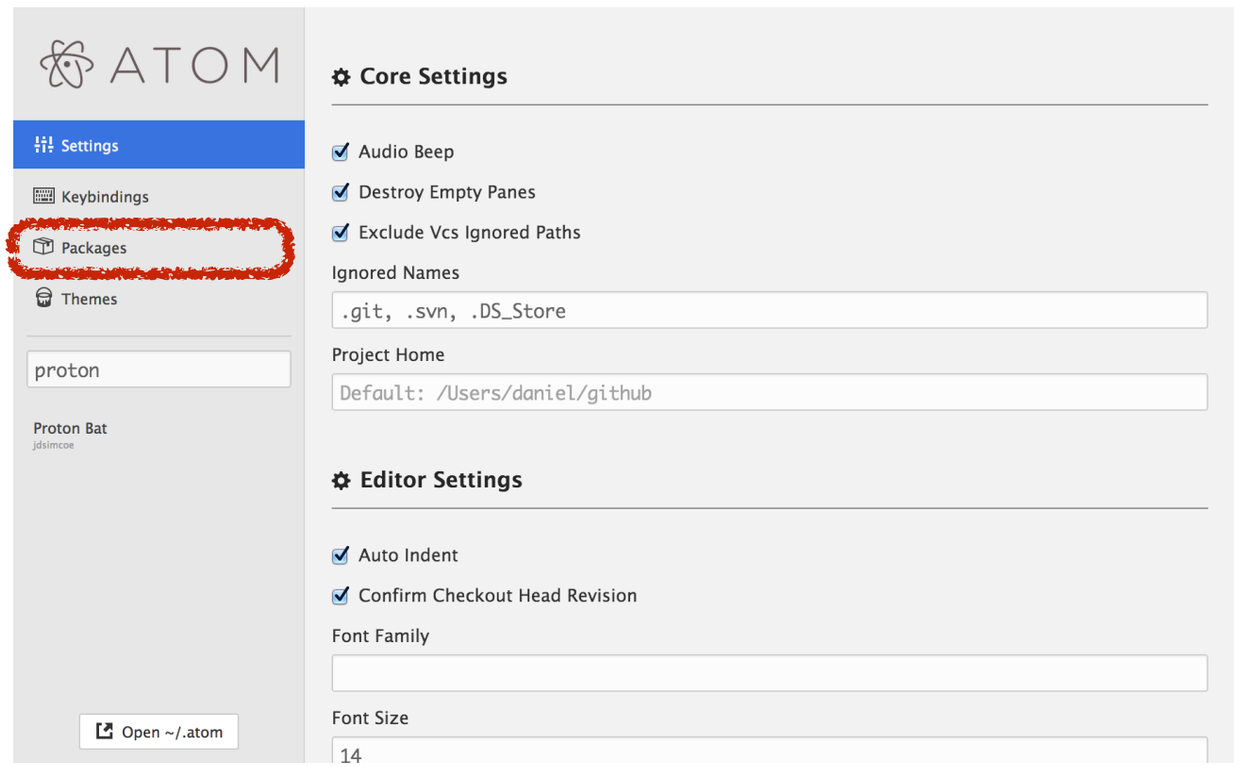
→ Overall look of program

Syntax Theme Proton Bat

This styles the text inside the editor

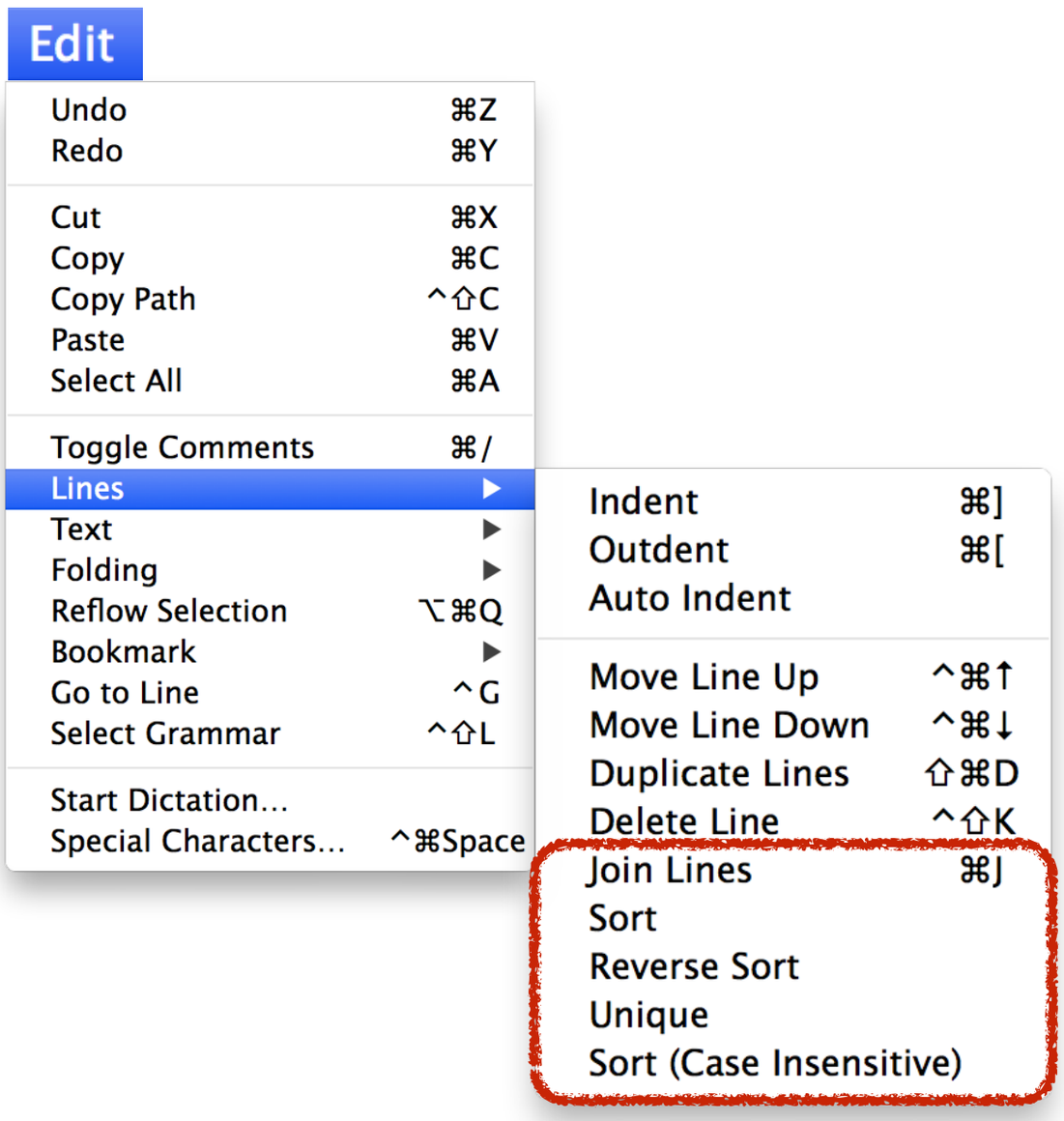
→ Color of text in the editor

ATOM is extensible!



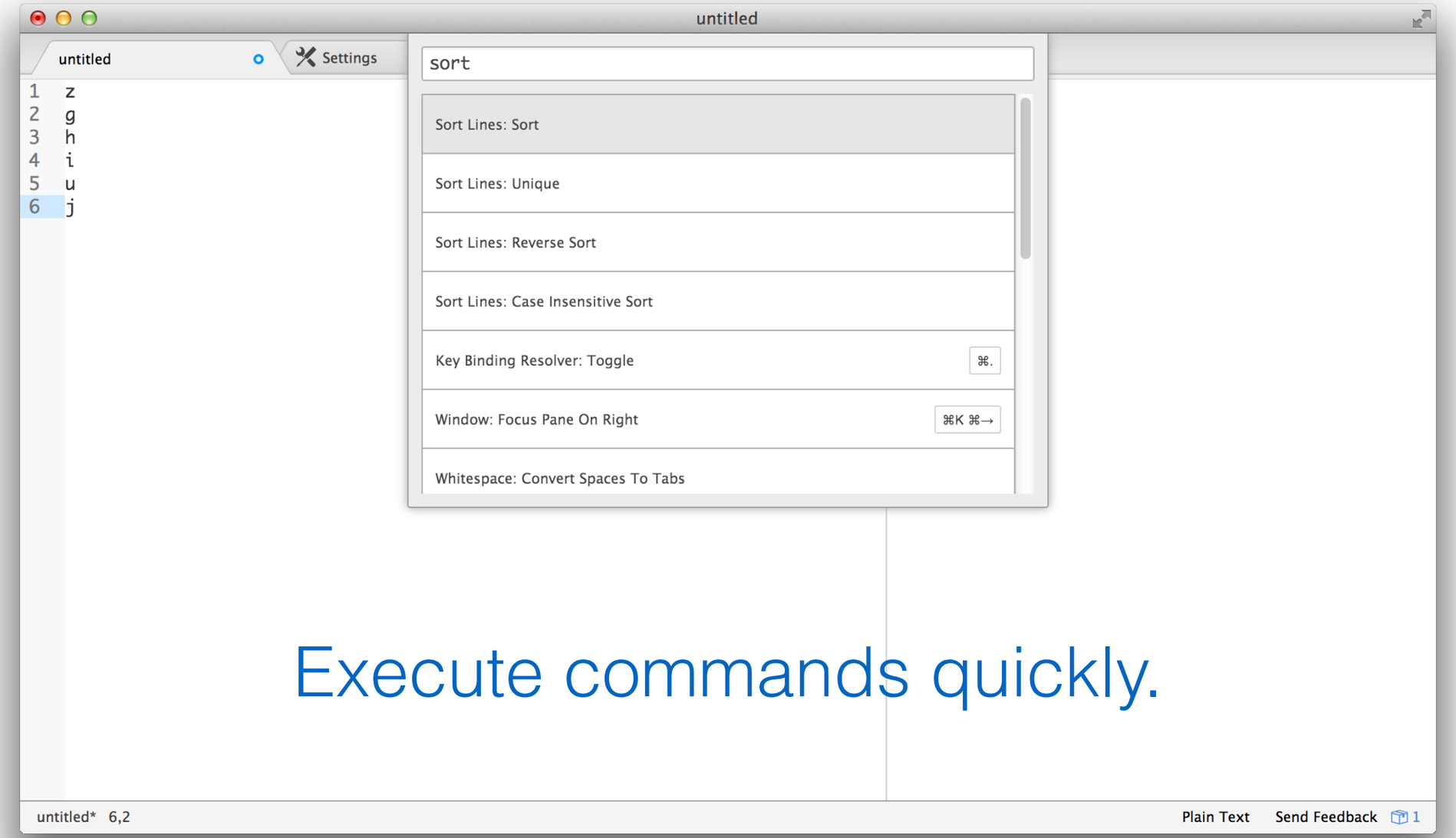
Install ‘sort lines’

ATOM is extensible!



New functions!

ATOM is has “command palette”! Command-shift-p



Let’s make readme.md from ATOM

1. Open Atom
2. Check out the welcome.md
3. Go to the untitled pane.
4. Type:

```
This example *project* will teach us **a lot** about basic computational biology.
```

5. Press command-shift-p
6. Type “markdown” and press enter
7. Check it out!
8. Close the markdown preview
9. Press control-shift-m
10. Check it out!
11. Save as “readme.md” into MyProject folder

Further experiences:
https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet
http://daringfireball.net/projects/markdown/basics

Make an ibis.md file

1. Title: IBiS is awesome
2. “awesome” in bold
3. Make it a first level header with a line underneath it
4. Add normal text below: “Our webpage is here.”
5. Add the link for the IBiS webpage to “here”.
6. Make a list of the top three reasons you love IBiS
7. Add the IBiS image to the bottom

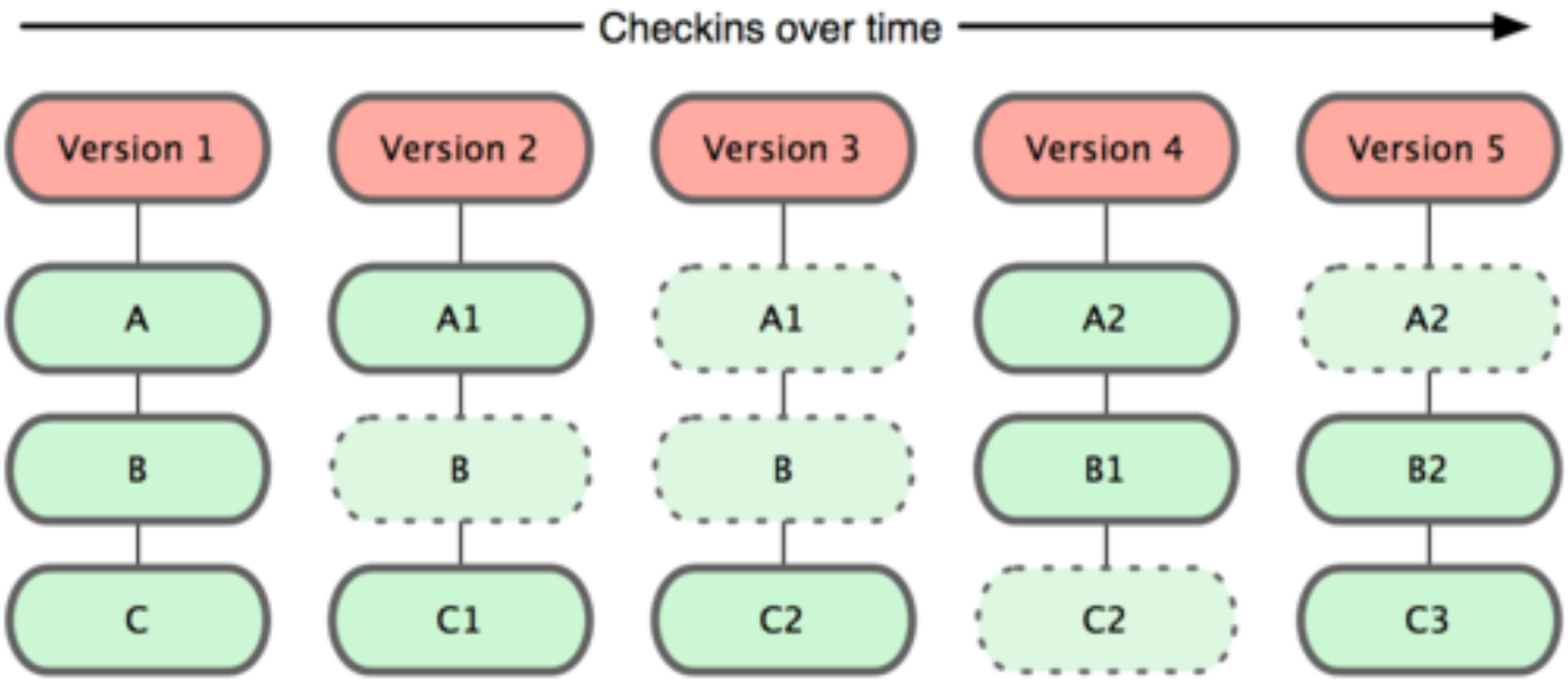
Course outline

Day #1: Basic command line interface and reproducible research

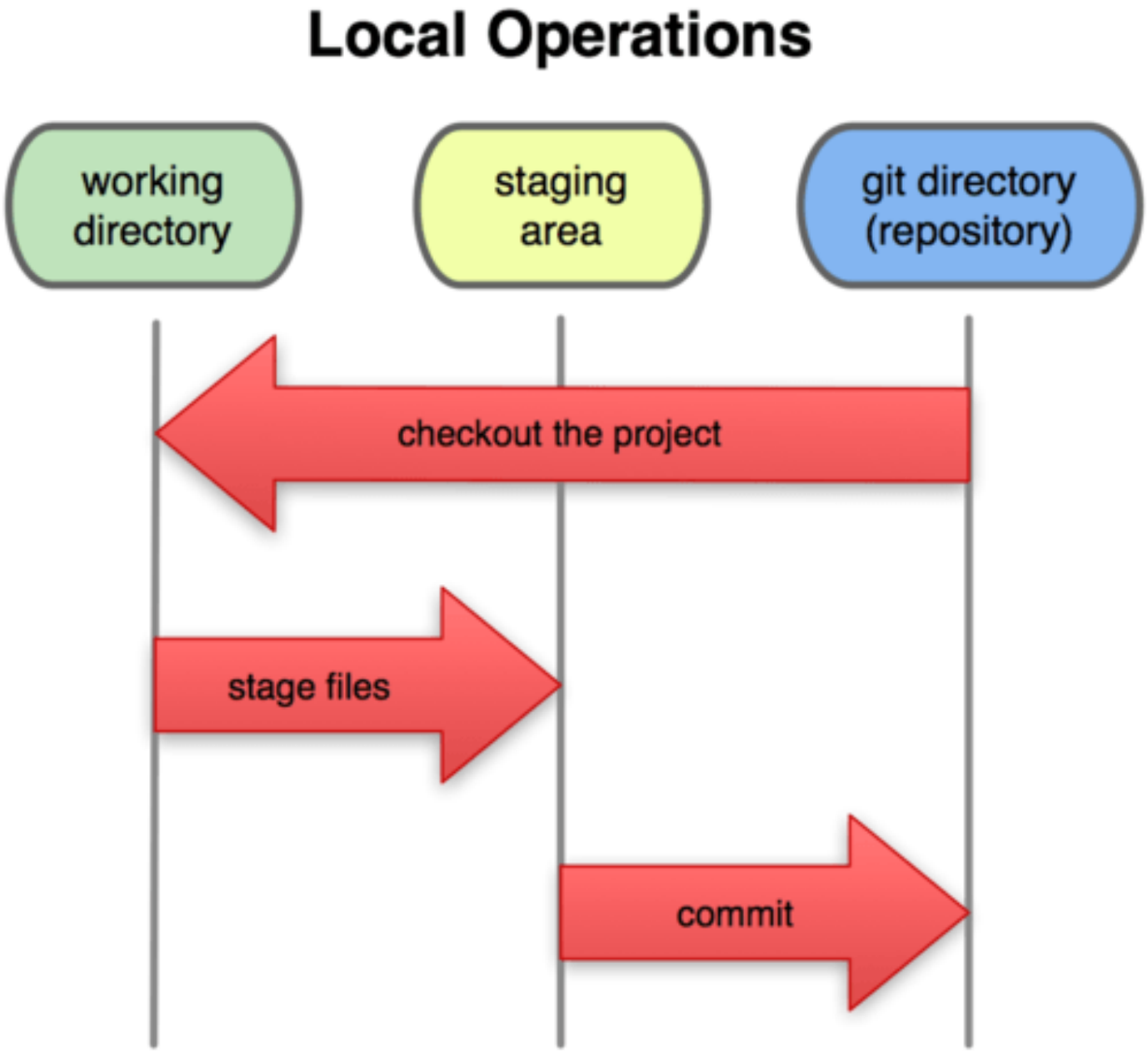
- **Intro to reproducible research**
- **Basic command line**
- **Directory structures**
- **Markdown**
- **Git and github**

is a version control system

git tracks changes and files over time



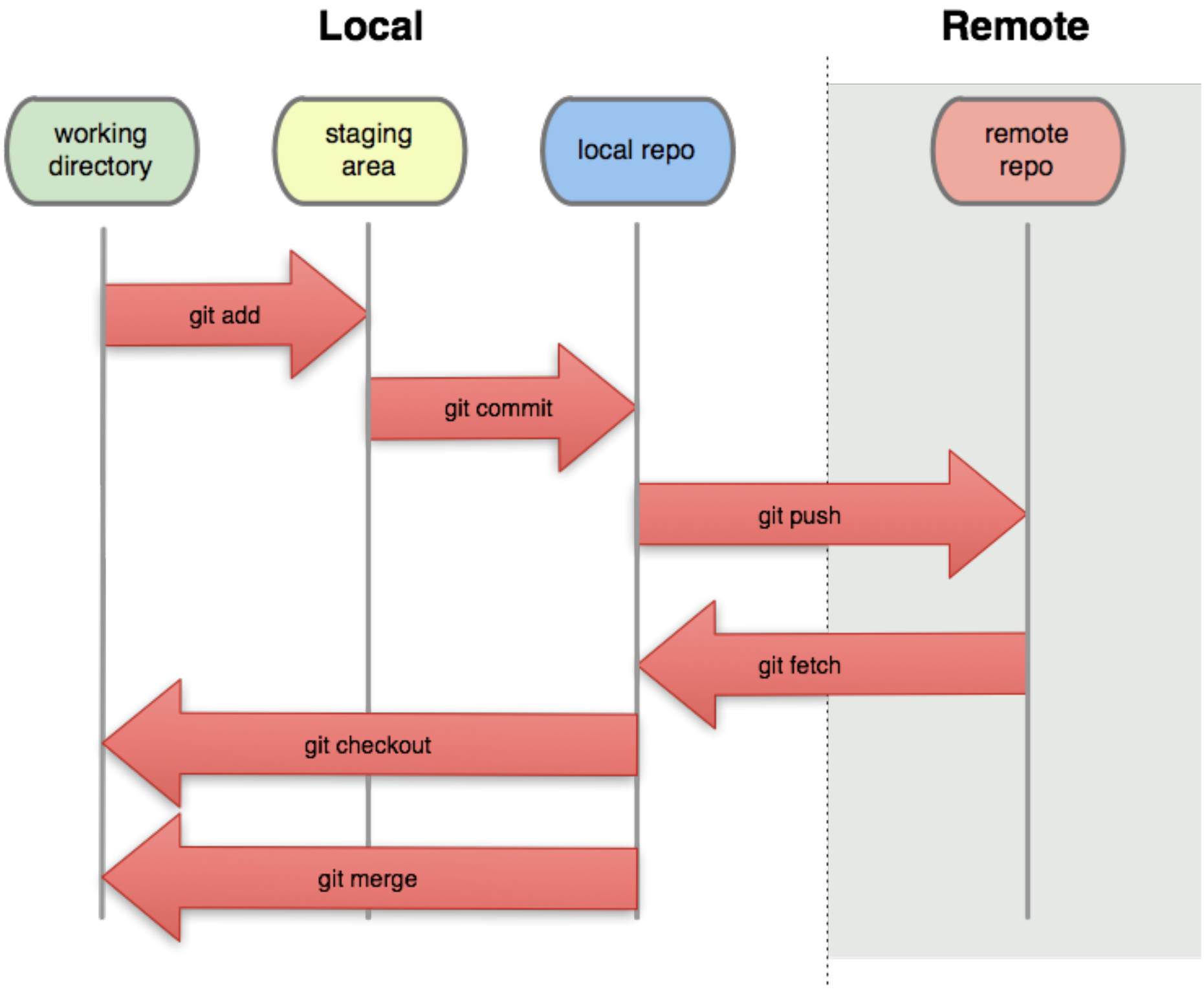
git tracks local changes



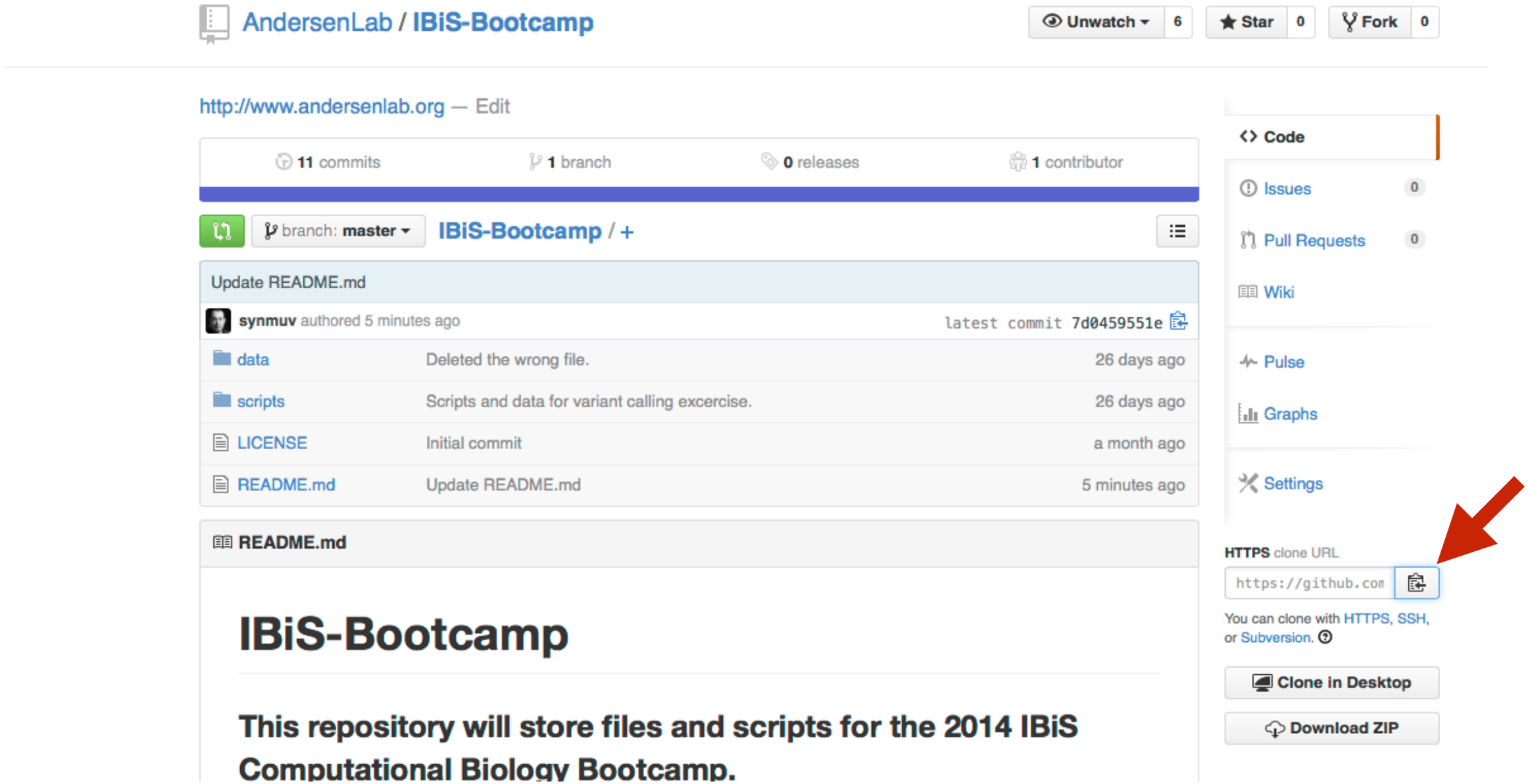
Github makes git super powerful

Check out: www.github.com/andersenlab

github takes git from local to the world

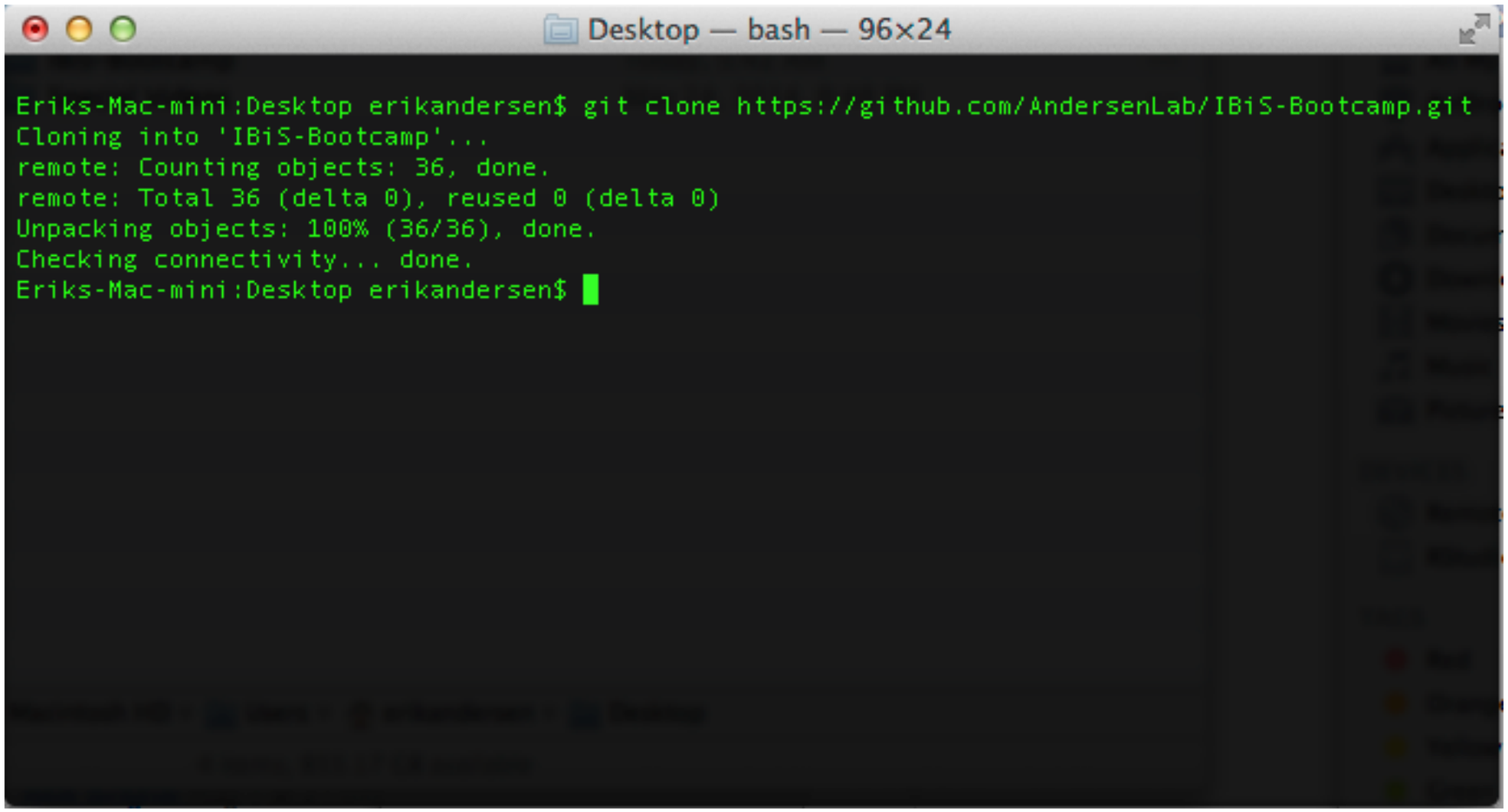


Let's clone the IBiS Bootcamp repository for work tomorrow



Let's clone the IBiS Bootcamp repository for work tomorrow

`git clone` invokes git to clone a repository



1. Go to home directory. Type `cd ~/`
2. Type `git clone` and then paste directory after `clone`.

Git/Github Commands

Move files to staging area `git add [file/s]`

Check status of local repo `git status`

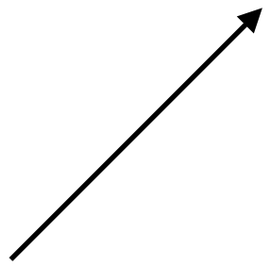
Commit files to local repo `git commit -m “short message”`

Push files to remote repo `git push origin master`

https://training.github.com/kit/downloads/github-git-cheat-sheet.pdf

The command line can be personalized!

.bash_profile



A hidden file in your home directory

- Your *bash profile* is run every time you open your terminal.
- You can use your bash profile to store common settings, functions, modify commands, or personalize your terminal.

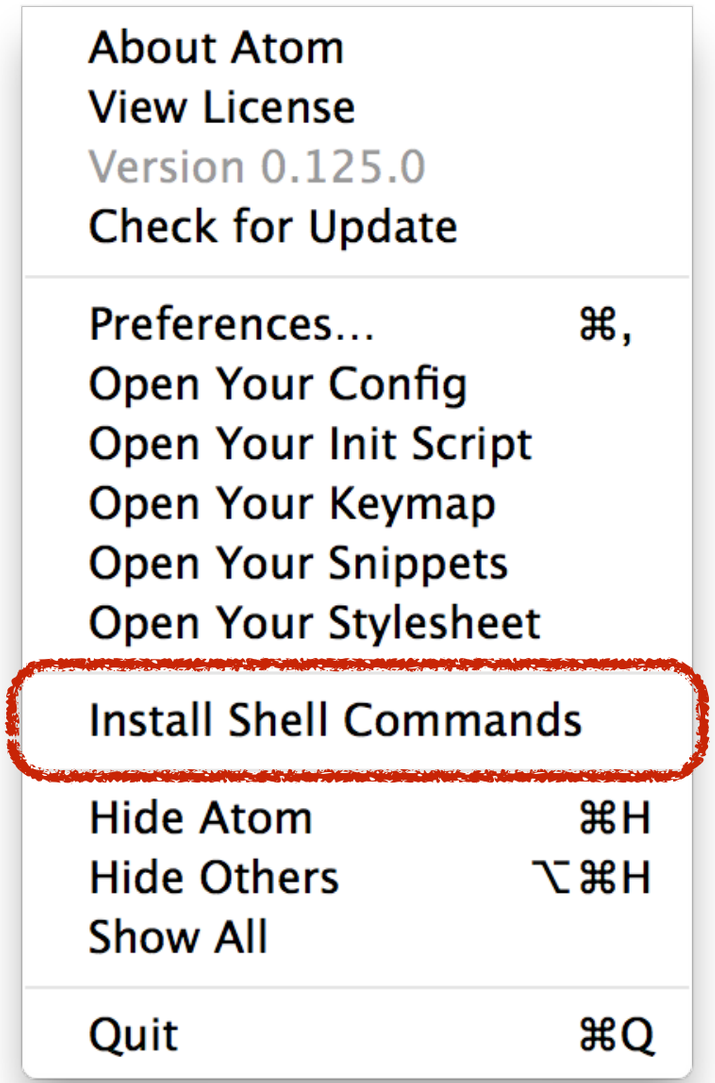
Editing your bash profile

1. Install the atom shell commands.

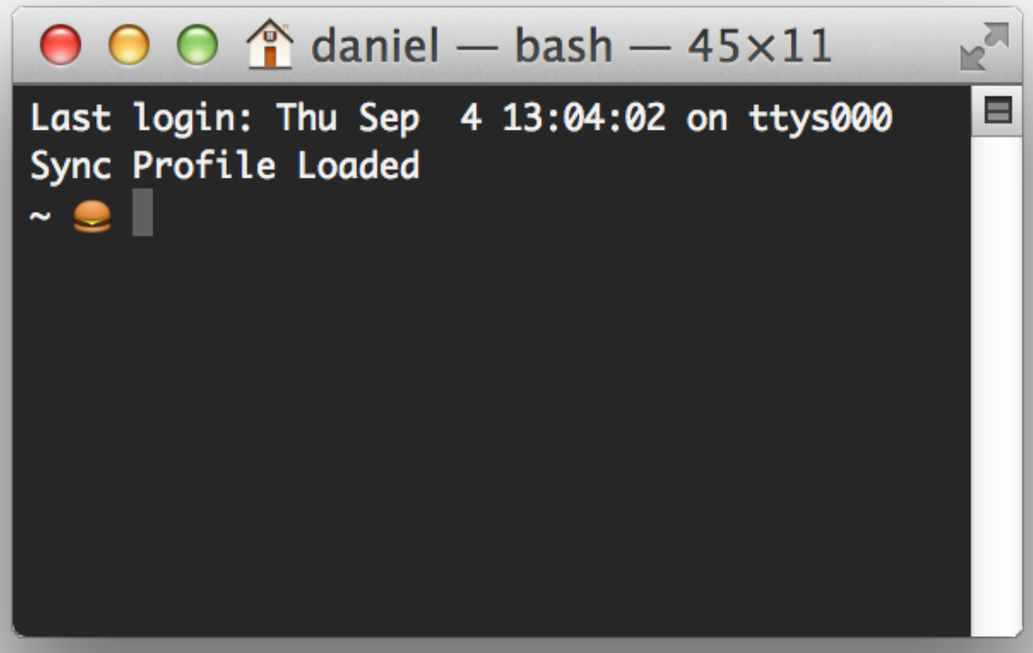
Atom > Install Shell Commands

2. Type:

```
atom .bash_profile
```

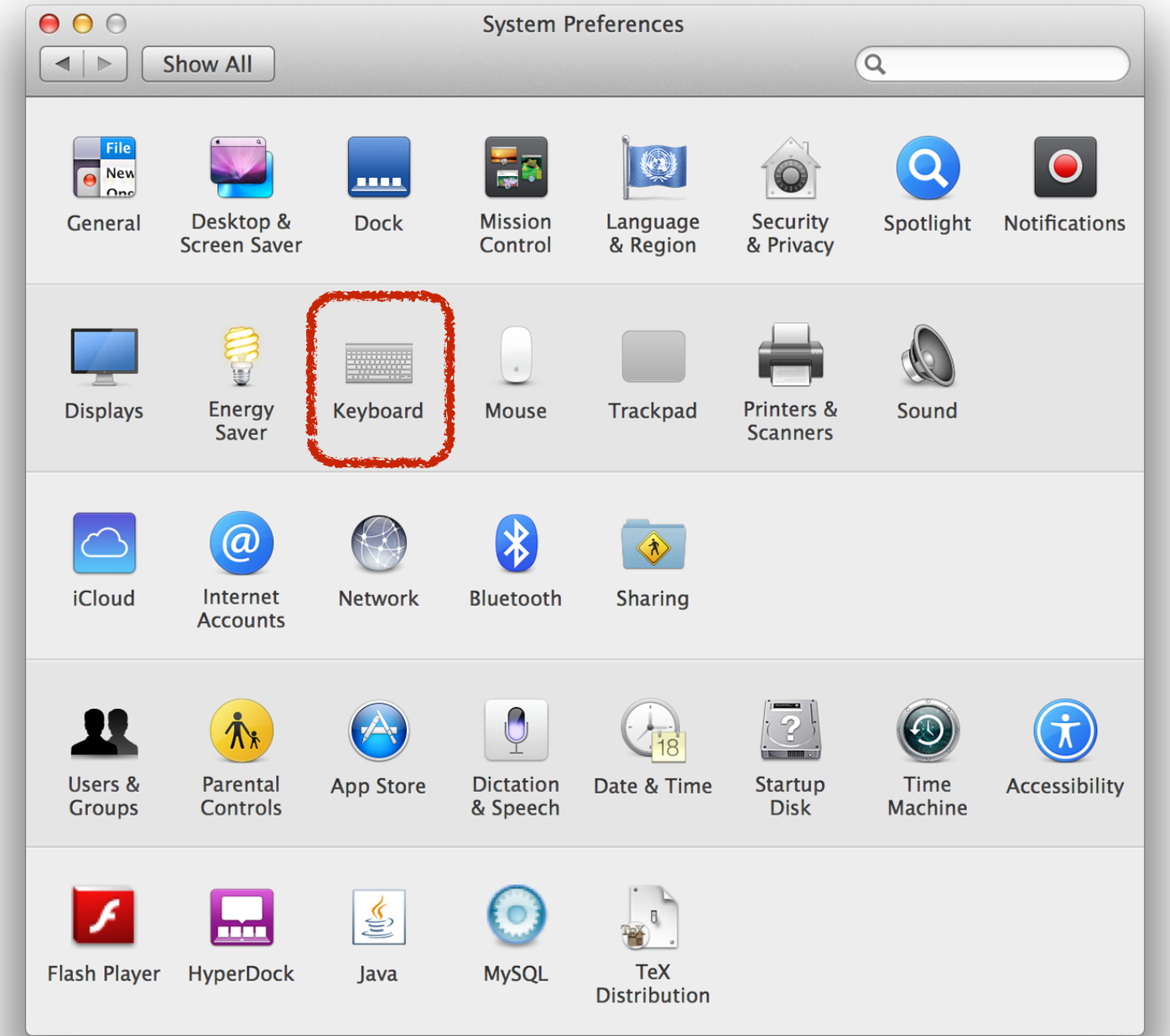


Personalize your Terminal

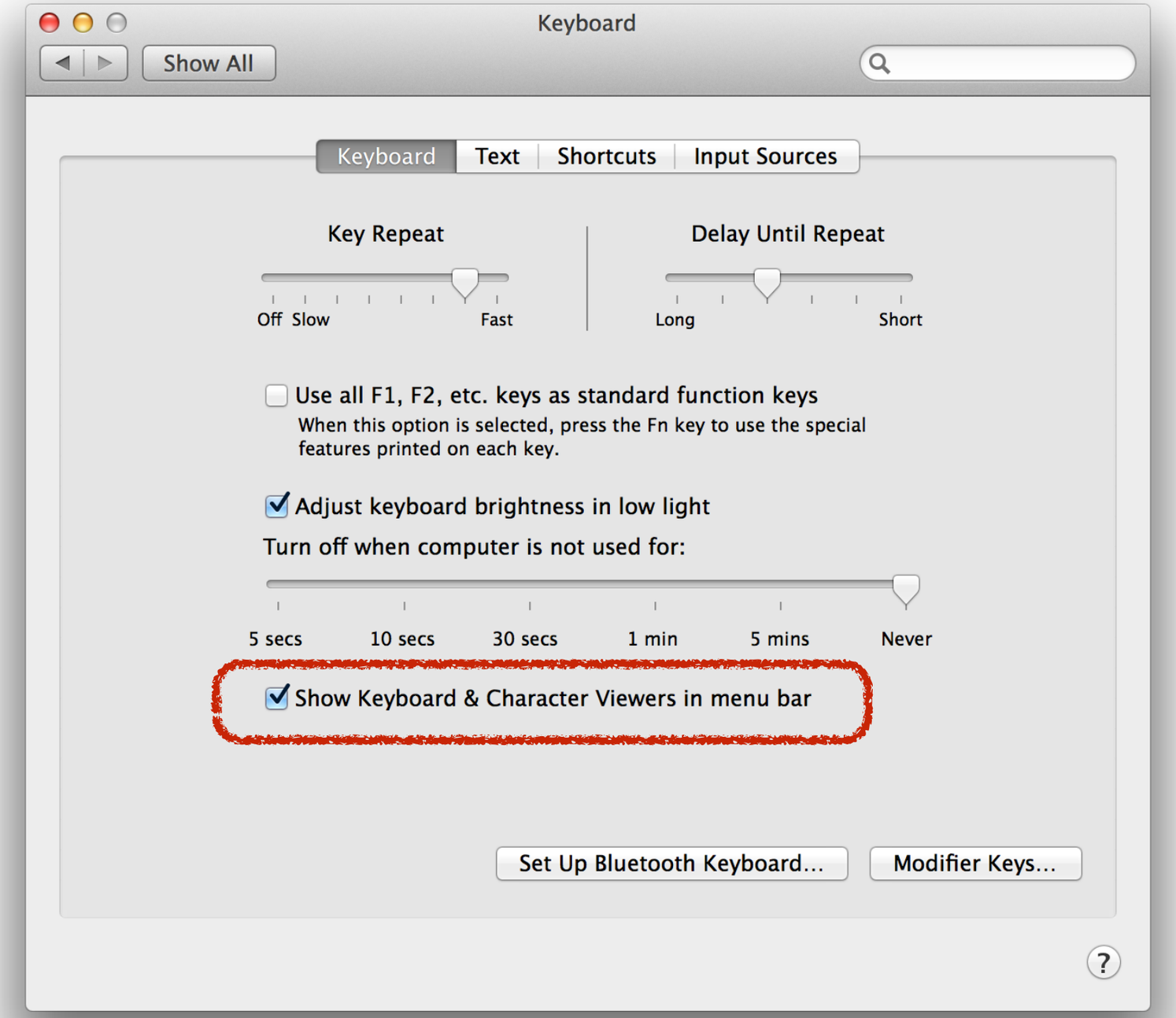


Add an icon to see command lines

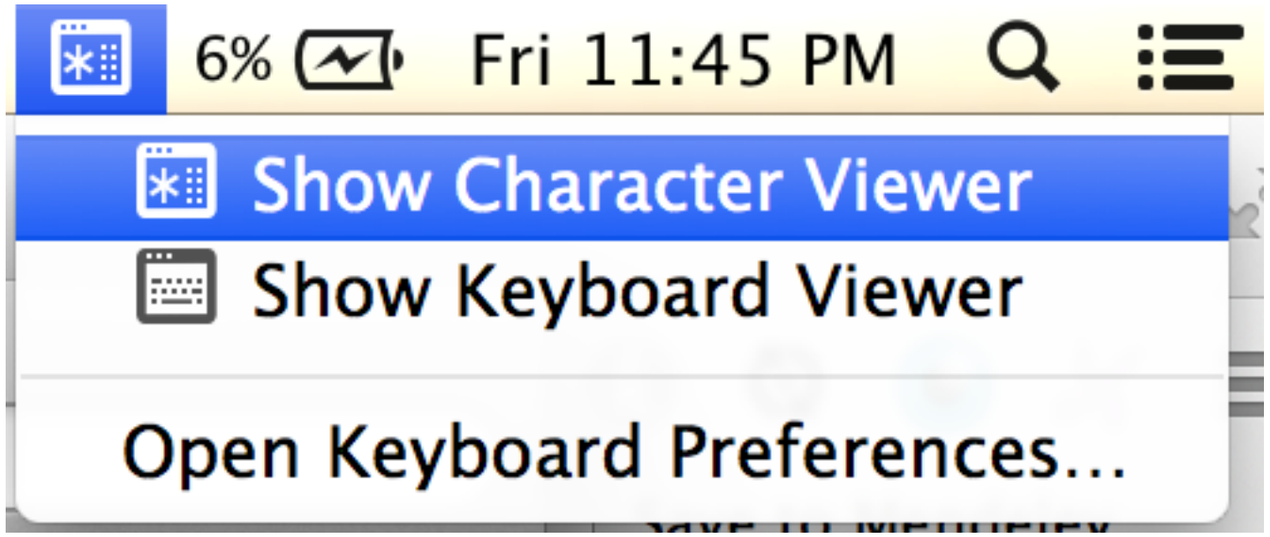
Open System Settings



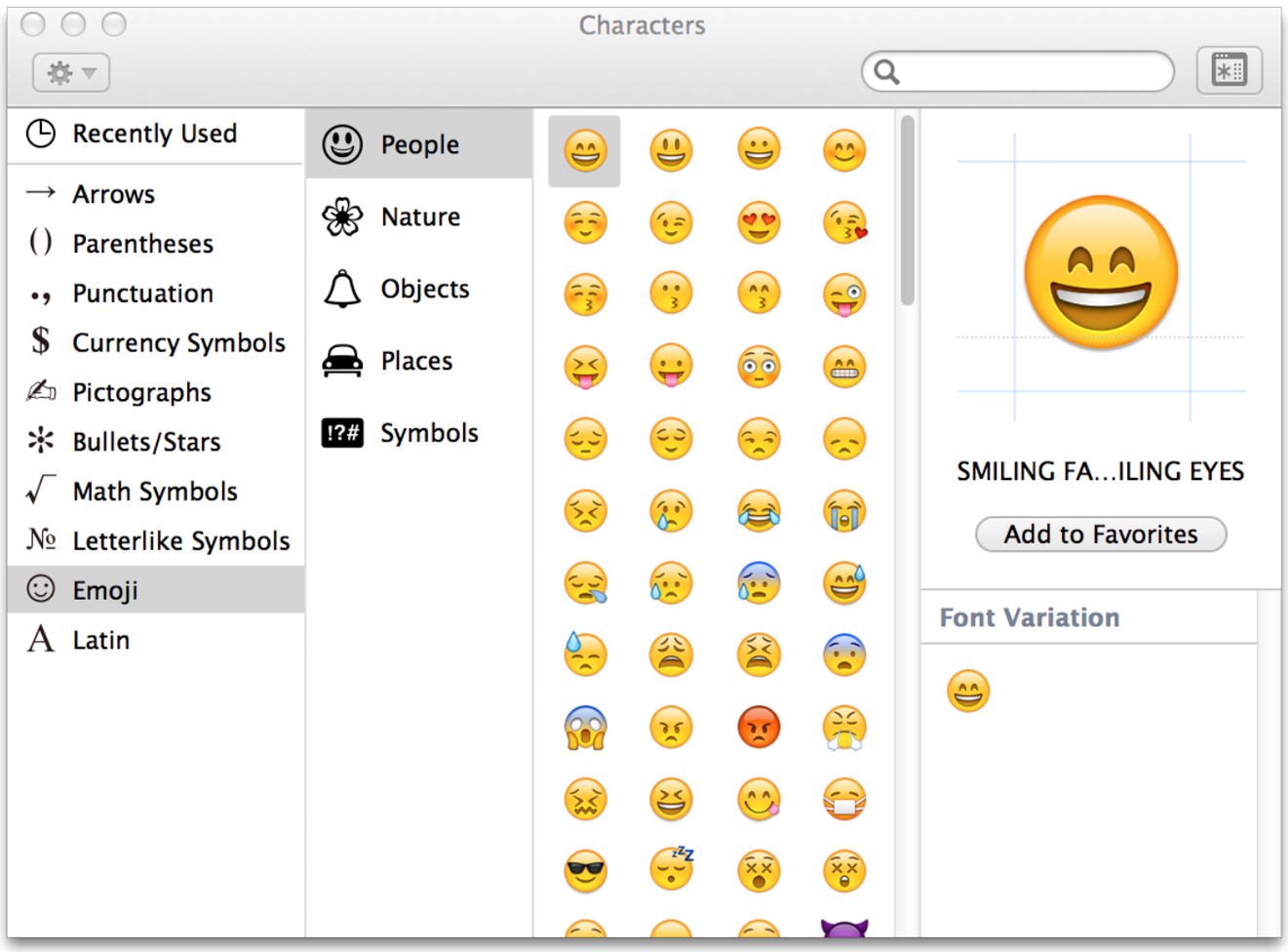
Click “Keyboard”



Check “Show Keyboard & Character Viewers in menu bar”

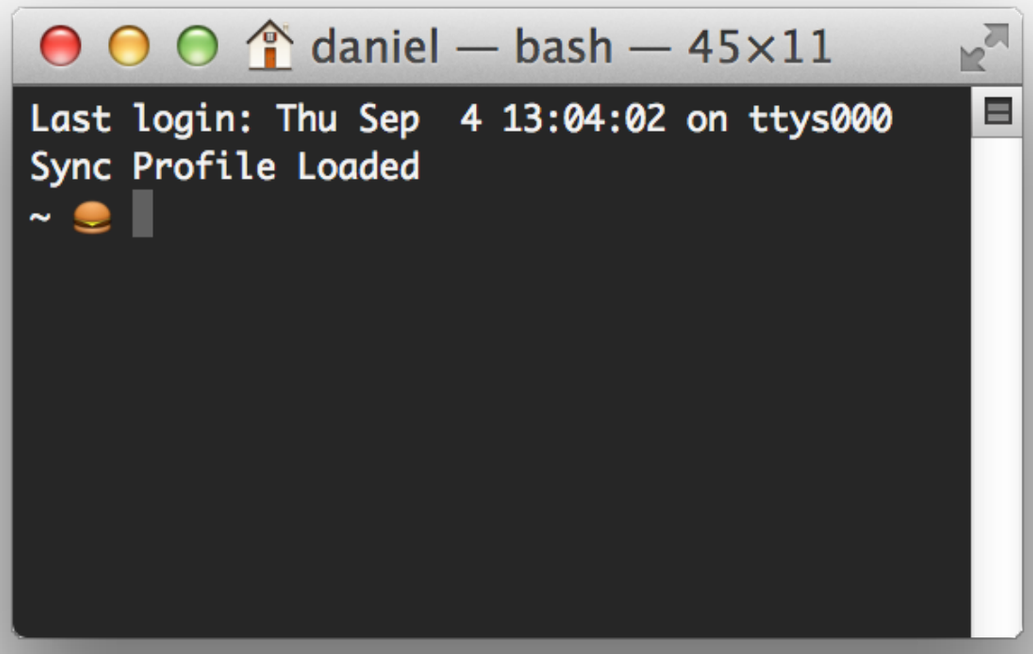


Open the character viewer.



Pick a fun icon!

Personalize your Terminal



Put your cursor here
and double click on the icon

```
export PS1="\w   "
```

Shortcuts can be added with aliases

```
alias refresh="source ~/.bash_profile"
alias dbx="cd ~/Dropbox/"
```

Typing ‘refresh’ will refresh your bash profile.

Typing ‘dbx’ will move you to the dropbox directory.

Functions can be used to execute common tasks.

```
new_project() {
    mkdir $1
    touch $1/Readme.md
    mkdir $1/Data
    mkdir $1/Data/Raw
    mkdir $1/Data/Processed
    mkdir $1/Scripts
    mkdir $1/Results
}
```

This function will create a project directory structure.

Example:

```
new_project chip-seq-analysis
```

What to do when you are lost?

Google

stackoverflow

coursera

CODE ACADEMY

Let’s have some fun!

1. *Cool computer voice:* Type `say` and whatever you want. Make sure your sound is on

2. *Get your fortune:* Type `brew install fortune`. Then type: `fortune`

3. *Download youtube videos:* Type `brew install youtube-dl`.
4. Then type: `youtube-dl cHK428vSIz8`

5. *Watch an asci movie:* Type: `telnet towel.blinkenlights.nl 23`

Day #1 Homework

1. Sign up for github
2. Star our class repository
3. Go and explore github! Find some cool code.