

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

The screenshot shows a GitHub repository page for `AndersenLab/IBiS-Bootcamp`. The repository has 32 commits, 1 branch, 0 releases, and 2 contributors. The master branch is selected. The commit history includes:

- Fixed DL238 script! by [danielecook](#) 18 hours ago
- Add sorter data by [danielecook](#) 2 days ago
- atom selections example by [danielecook](#) 3 days ago
- Fixed DL238 script! by [danielecook](#) 18 hours ago
- Initial commit by [danielecook](#) a month ago
- Update README.md by [danielecook](#) 6 days ago

The README.md file contains the following text:

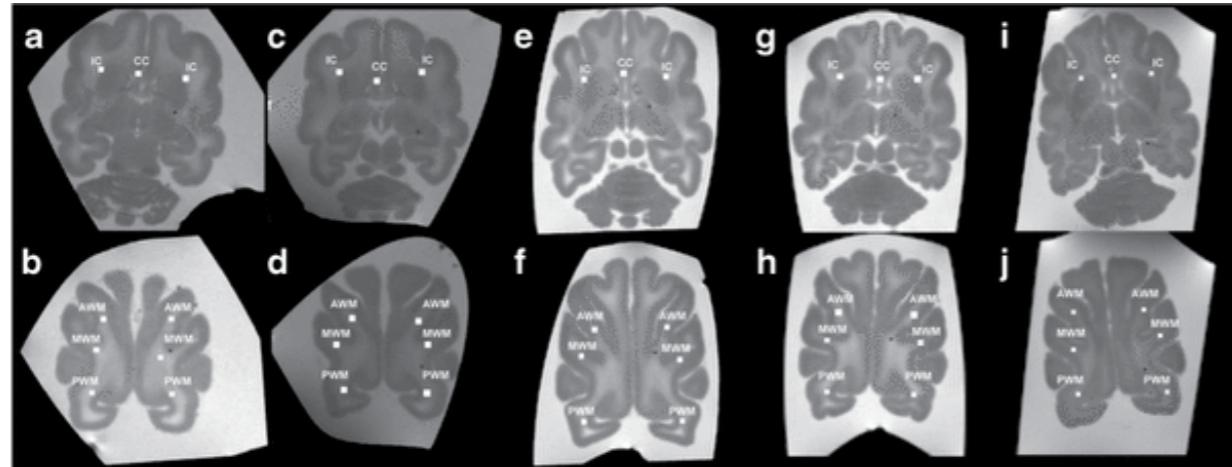
IBiS-Bootcamp

This repository will store files and scripts for the 2014 IBiS Computational Biology Bootcamp.

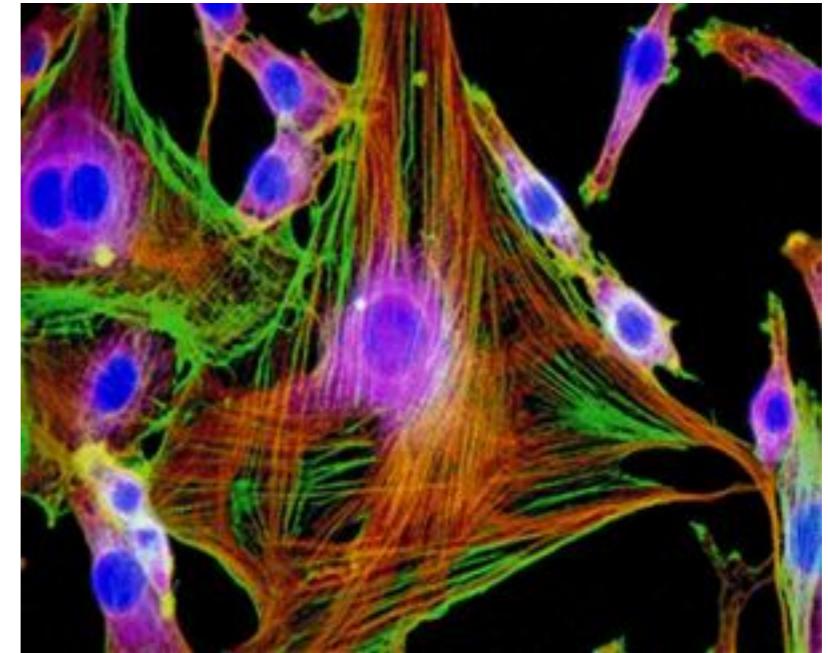
Check out our [Wiki](#)

On the right side, there is a sidebar with links to Code, Issues, Pull Requests, Wiki, Pulse, Graphs, and Settings. It also provides an HTTPS clone URL and options to Clone in Desktop or Download ZIP.

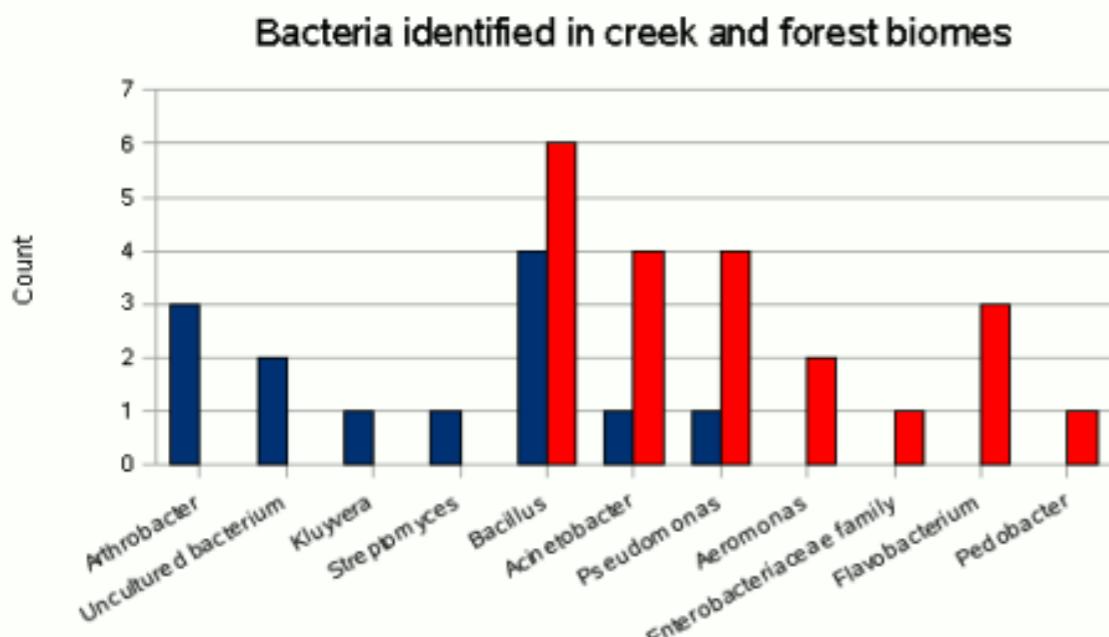
The days of descriptive biology are [almost] over



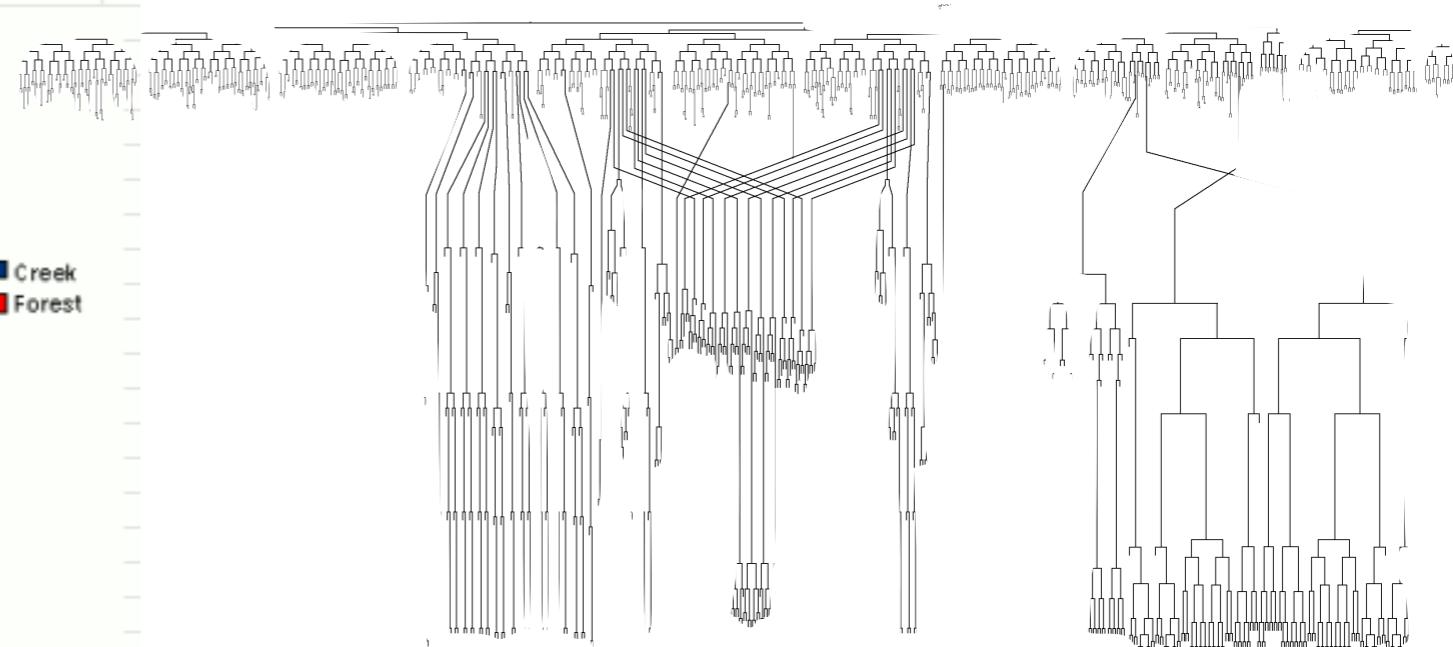
Reproducibility?



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

Reproducibility initiatives will invite increased scrutiny into data cleaning methods.

**Validating key experimental results
via independent replication**

[Learn more »](#)

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

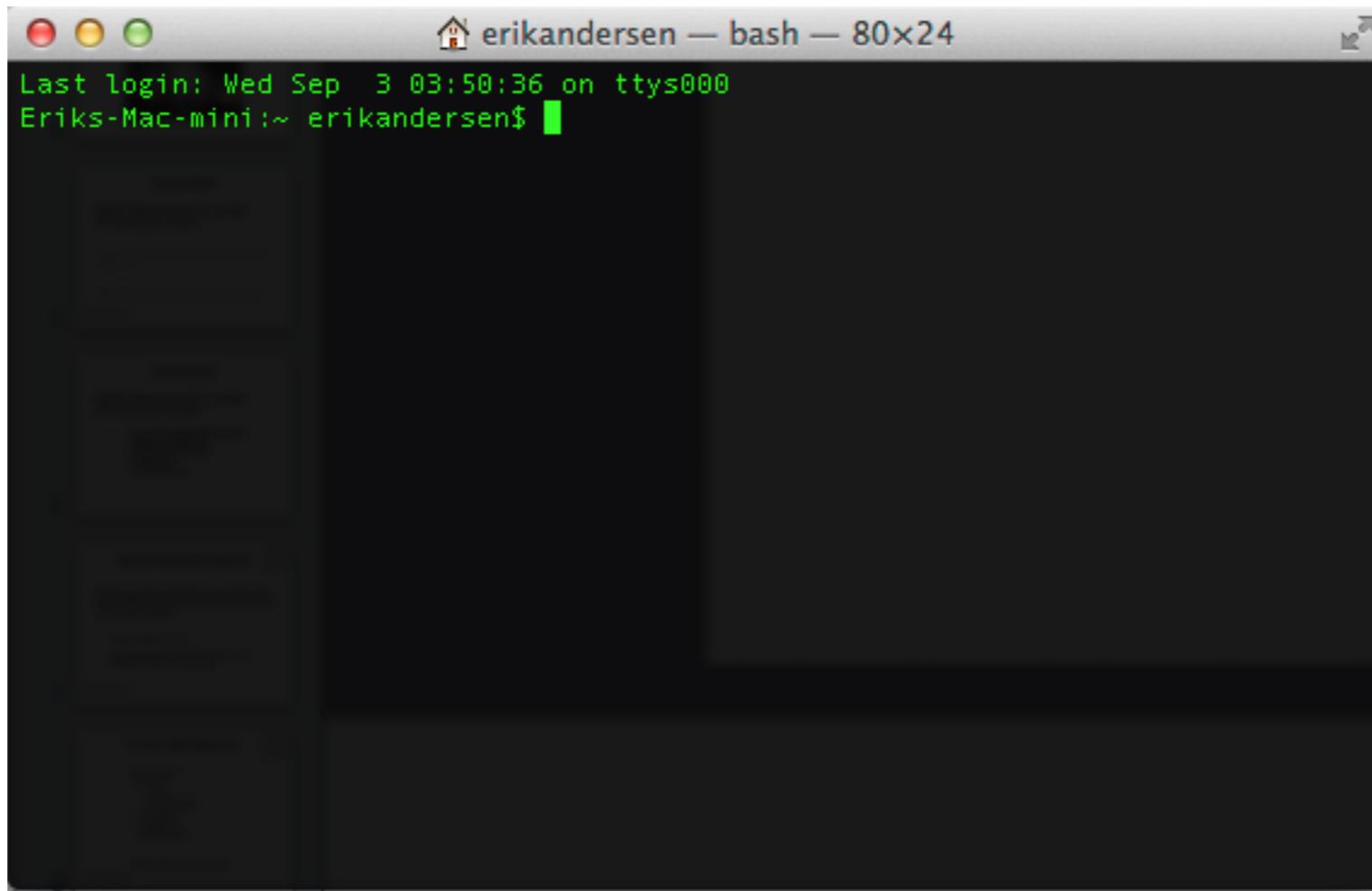
Course outline

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



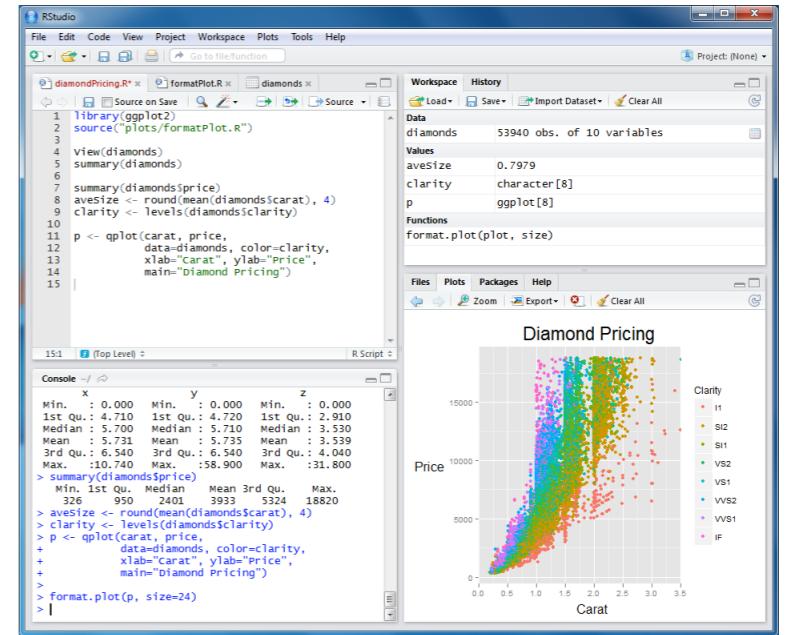
General Purpose software
is already included.

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

Science Software

`brew tap homebrew/science`

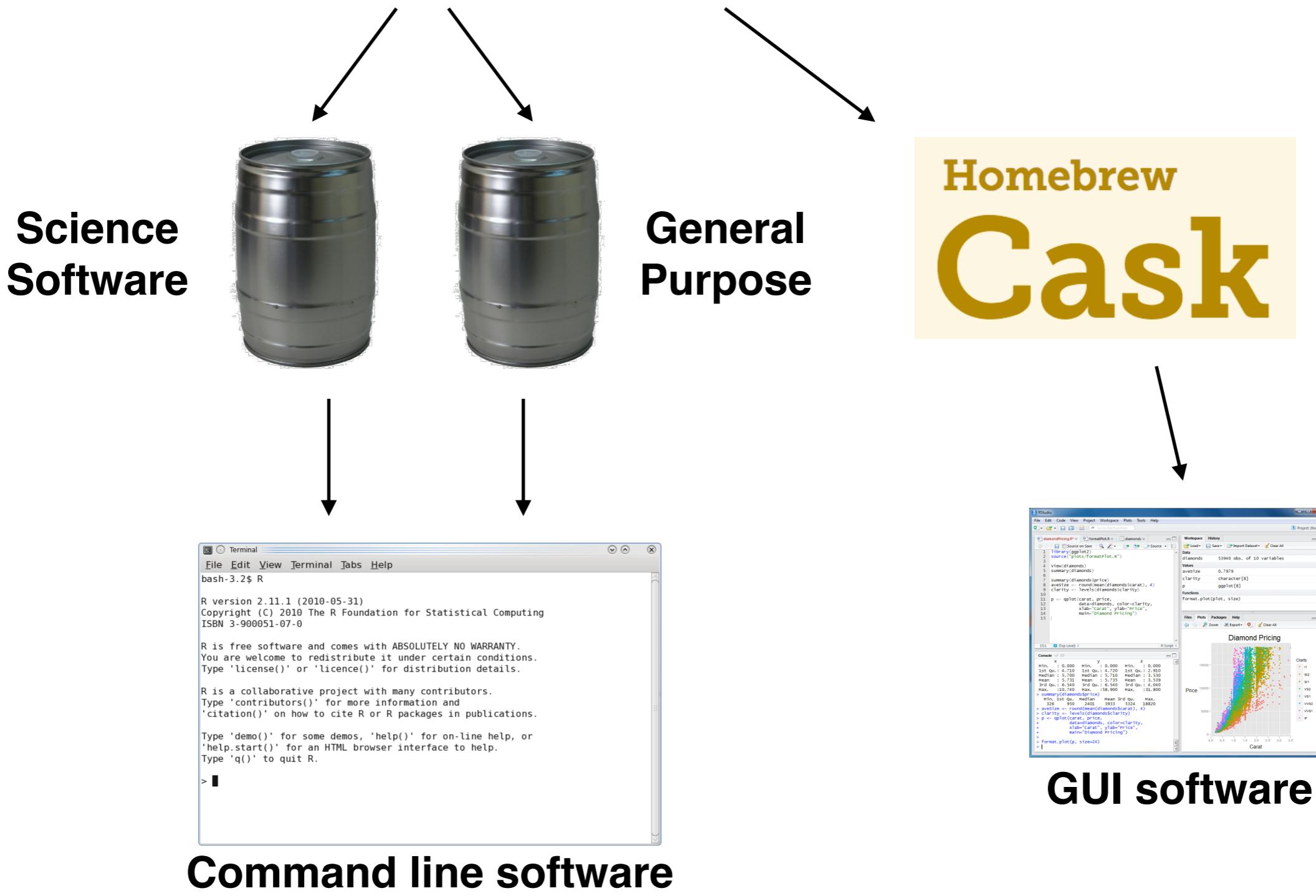
Graphical User Interface (GUI)



“Point and Click”

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

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



Software installation

Go to [www.GitHub.com/AndersenLab/IBiS-Bootcamp/Wiki/Installing-Required-Software](https://www.github.com/AndersenLab/IBiS-Bootcamp/Wiki/Installing-Required-Software)

```
1  #!/bin/bash
2
3  # This script install the required software!
4
5  # Install Homebrew, if you haven't as a part of the Lecture
6  #ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/homebrew/go/install)"
7  #brew tap homebrew/science
8
9  # Cask
10 #brew install caskroom/cask/brew-cask
11
12 # Sequence Analysis
13 brew install --devel samtools
14 brew install bcftools
15 brew install bwa
16 brew install igv
17
18 # wget
19 brew install wget
20
21 # Atom
22 brew cask install atom --appdir=/Applications
23
24 # ql markdown
25 brew cask install qlmarkdown
26
27 # R + Rstudio
28 brew install R
29 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)

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

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

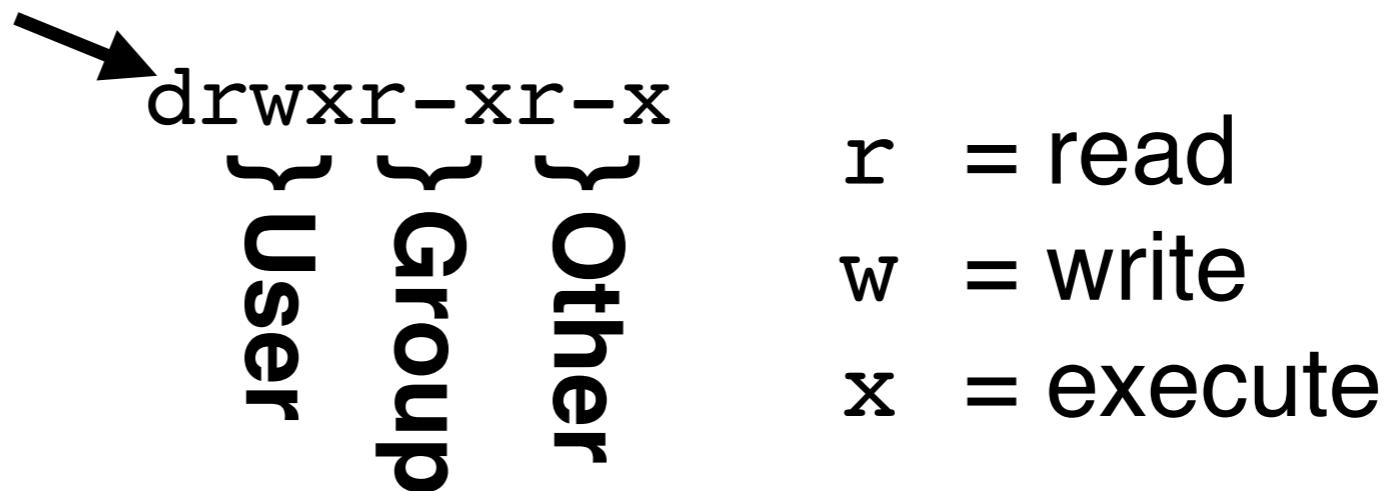
man	= read the manual for a command
ls	= list the contents of a directory
pwd	= display the present working directory
cd	= change directory
mkdir	= make a directory
rmdir	= remove a directory

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)



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

`chmod` changes
the mode

What about that `readme.md` file?

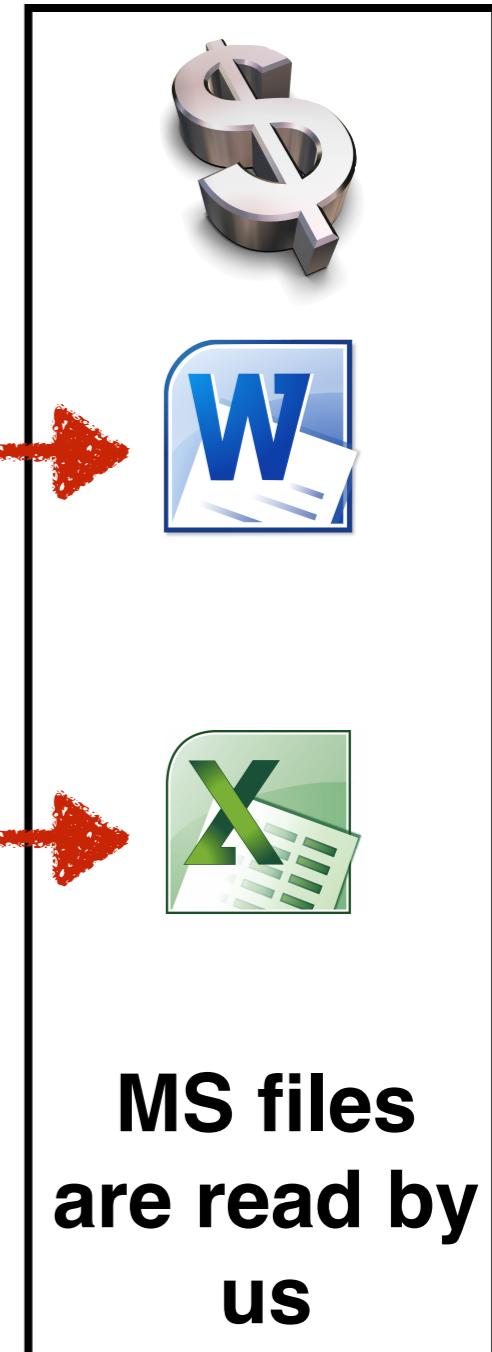
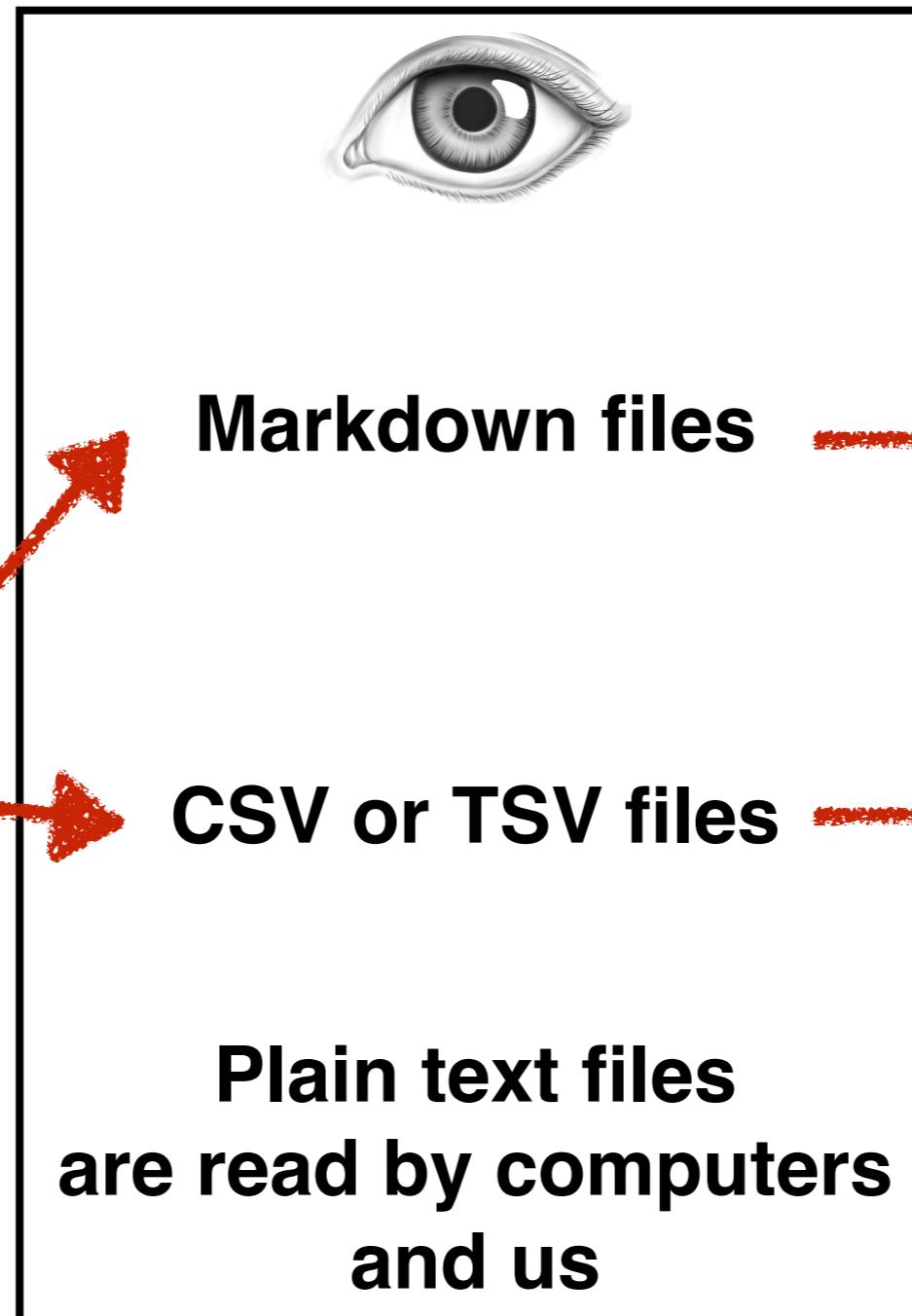
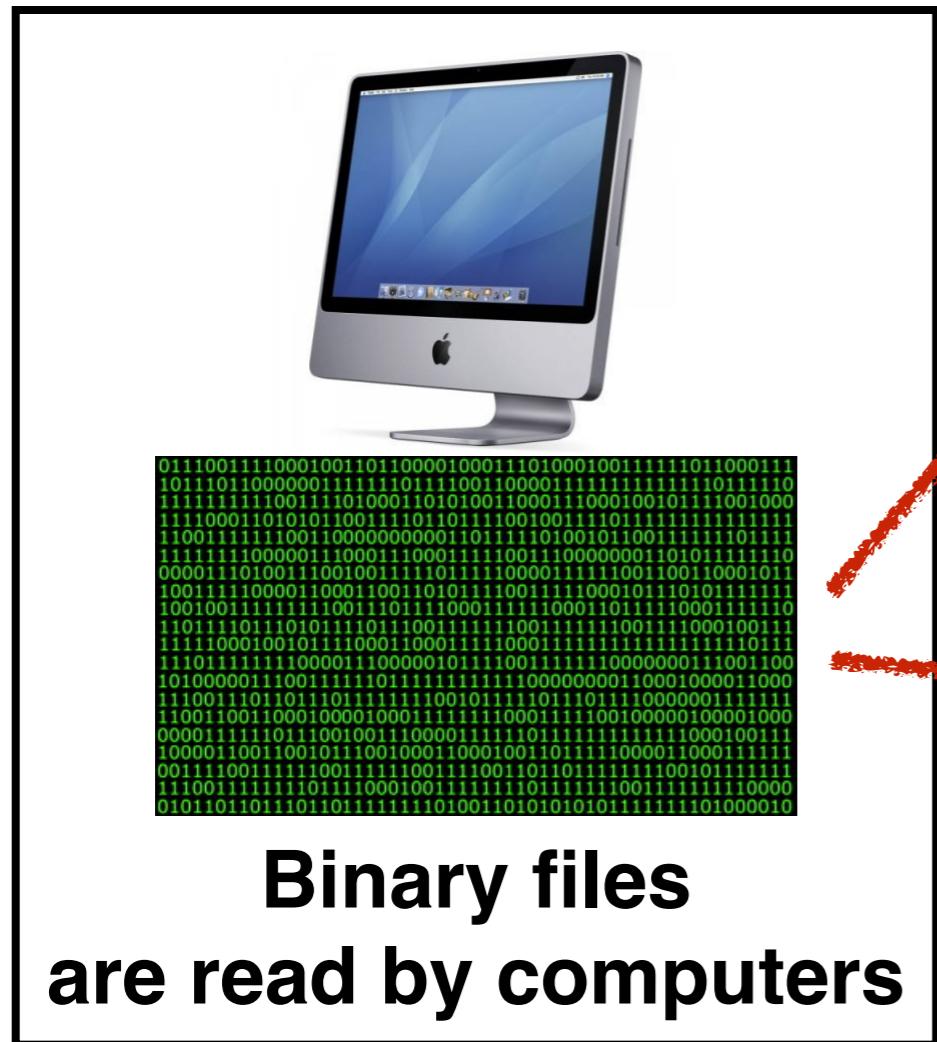
`md` = markdown

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.



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

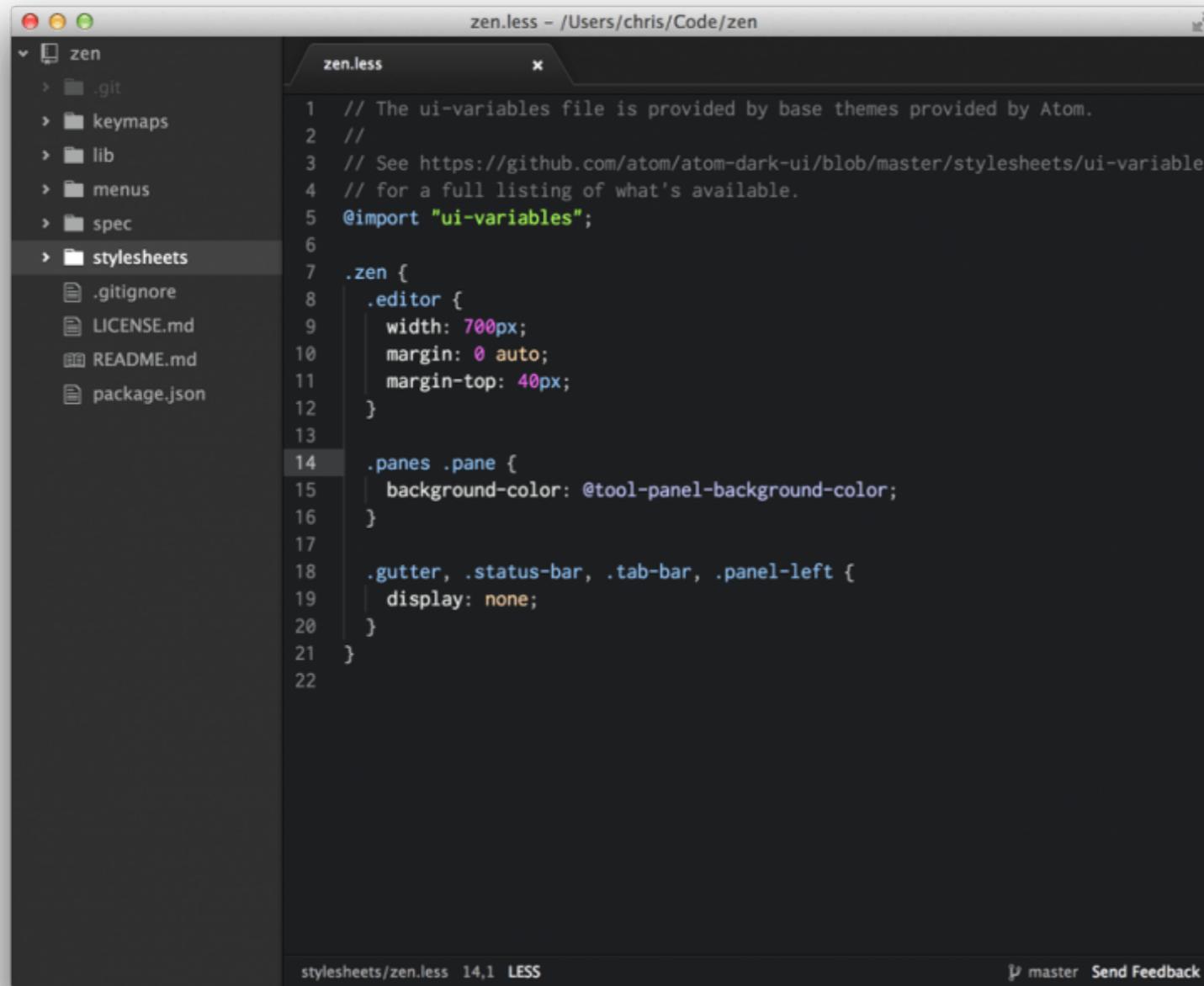
1. Go to MyProject directory
2. Type `nano readme.md`
3. Type:

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>



A screenshot of the Atom text editor interface. The window title is "zen.less - /Users/chris/Code/zen". On the left is a sidebar with a tree view of the project structure, showing "zen" as the root folder with subfolders ".git", "keymaps", "lib", "menus", "spec", and "stylesheets". The "stylesheets" folder is currently selected. The main editor area displays the content of the "zen.less" file. The code is as follows:

```
1 // The ui-variables file is provided by base themes provided by Atom.
2 //
3 // See https://github.com/atom/atom-dark-ui/blob/master/stylesheets/ui-variables
4 // for a full listing of what's available.
5 @import "ui-variables";
6
7 .zen {
8     .editor {
9         width: 700px;
10        margin: 0 auto;
11        margin-top: 40px;
12    }
13
14     .panes .pane {
15         background-color: @tool-panel-background-color;
16     }
17
18     .gutter, .status-bar, .tab-bar, .panel-left {
19         display: none;
20     }
21 }
22
```

The status bar at the bottom shows "stylesheets/zen.less 14,1 LESS".

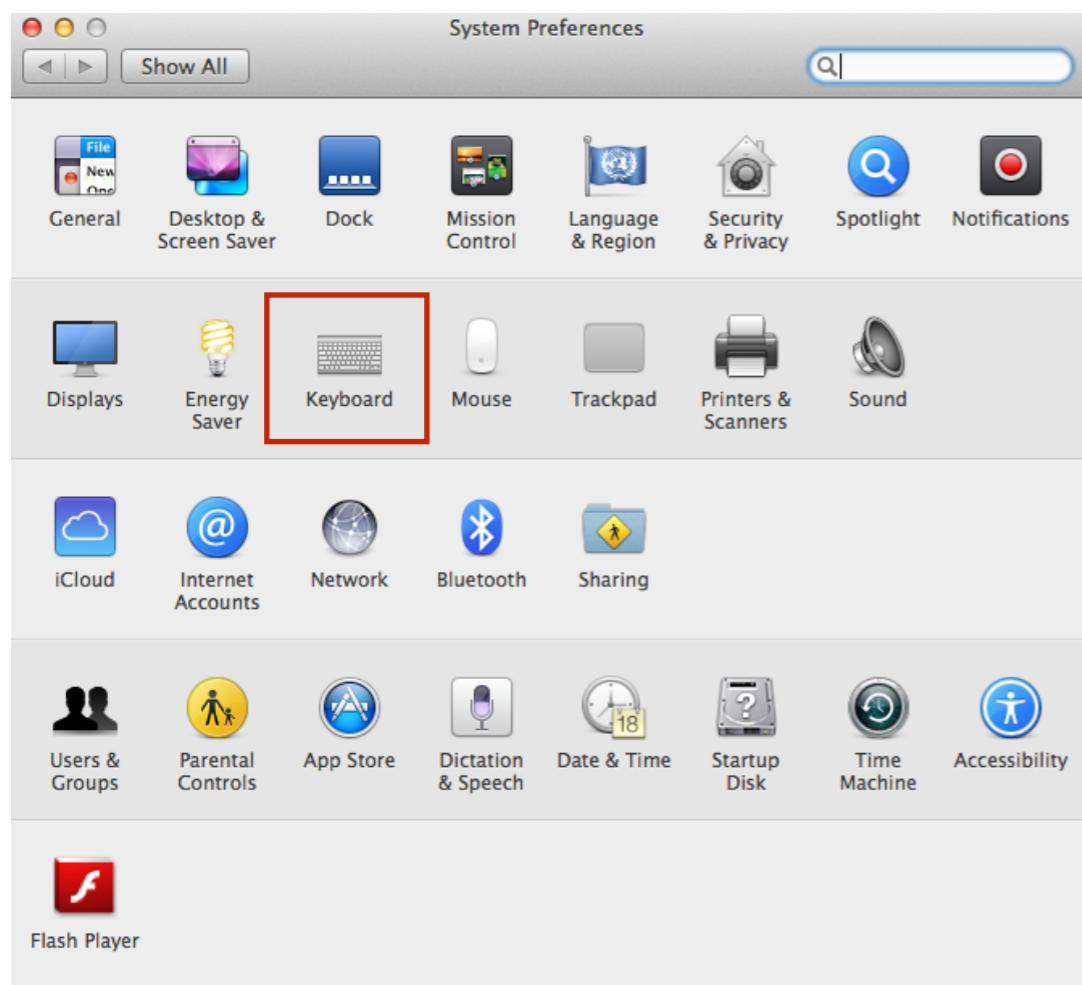
A plain text editor.
Very different from
Microsoft Word.



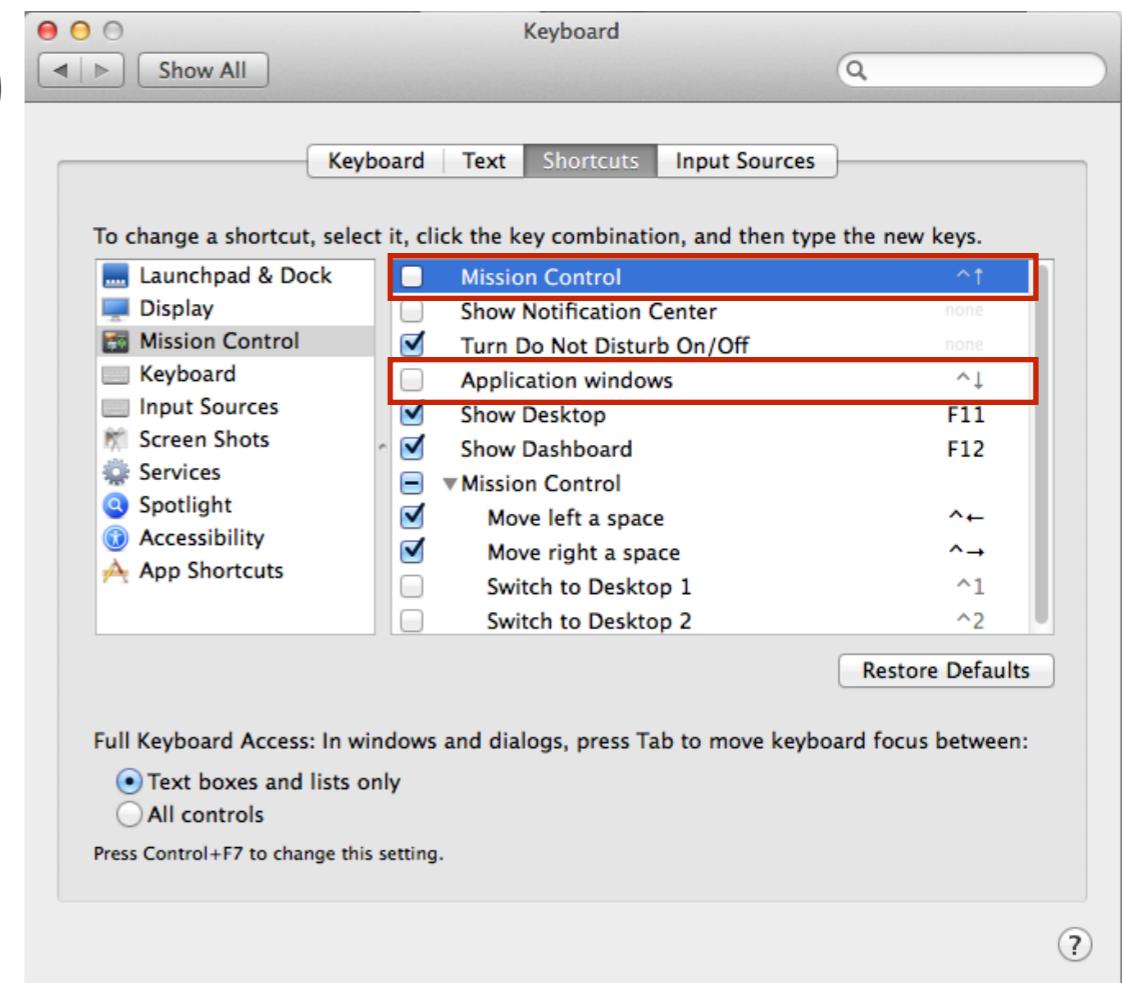
1)



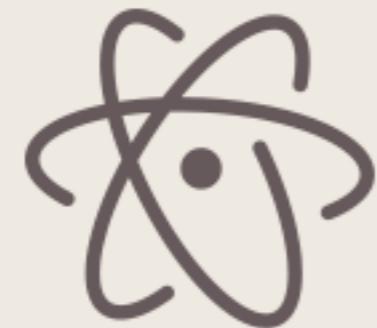
2)



3)



uncheck these two shortcuts



ATOM

Selection

Add Selection Above	⌃⇧↑
Add Selection Below	⌃⇧↓
Single Selection	◎
Split into Lines	⇧⌘L
Select to Top	⇧⌘↑
Select to Bottom	⇧⌘↓
Select Line	⌘L
Select Word	⌃⇧W
Select to Beginning of Word	⌥⇧B
Select to Beginning of Line	⌥⇧B
Select to First Character of Line	⇧⌘←
Select to End of Word	⌥⇧F
Select to End of Line	⇧⌘→

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!

The image displays two screenshots of the Atom application. On the left, the Atom menu is shown with the following options:

- About Atom
- View License
- Version 0.125.0
- Check for Update
- Preferences... ⌘,** highlighted with a blue background
- Open Your Config
- Open Your Init Script
- Open Your Keypad
- Open Your Snippets
- Open Your Stylesheet
- Install Shell Commands
- Hide Atom ⌘H
- Hide Others ⌘H
- Show All
- Quit ⌘Q

An arrow points from the "Preferences..." option in the menu to the "Themes" section in the settings interface on the right. The "Themes" section is circled in red. The settings interface shows the following sections:

- Core Settings**
 - Audio Beep
 - Destroy Empty Panes
 - Exclude Vcs Ignored Paths
 - Ignored Names: .git, .svn, .DS_Store
 - Project Home: Default: /Users/daniel/github
- Editor Settings**
 - Auto Indent
 - Confirm Checkout Head Revision
 - Font Family: (empty input field)
 - Font Size: 14



ATOM is customizable!

 **Choose a Theme**

 You can also style Atom by editing [your stylesheet](#)

UI Theme Atom Light → Overall look of program
This styles the tabs, status bar, tree view, and dropdowns

Syntax Theme Proton Bat → Color of text in the editor
This styles the text inside the editor

Overall look of program

Color of text in the editor



ATOM is extensible!

A screenshot of the Atom settings interface. The left sidebar shows navigation options: Settings (highlighted in blue), Keybindings, Packages (highlighted with a red oval), Themes, and a search bar for 'proton'. Below the sidebar, it says 'Proton Bat' and 'jdsimcoe'. At the bottom is a button 'Open ~/atom'. The main area shows 'Core Settings' with checkboxes for 'Audio Beep', 'Destroy Empty Panes', and 'Exclude Vcs Ignored Paths'. It also lists 'Ignored Names' as '.git, .svn, .DS_Store'. Under 'Project Home', it shows 'Default: /Users/daniel/github'. The 'Editor Settings' section includes 'Auto Indent' and 'Confirm Checkout Head Revision', with dropdowns for 'Font Family' and 'Font Size' set to '14'.

ATOM

Settings

Keybindings

Packages

Themes

proton

Proton Bat
jdsimcoe

Open ~/atom

Core Settings

- Audio Beep
- Destroy Empty Panes
- Exclude Vcs Ignored Paths

Ignored Names

.git, .svn, .DS_Store

Project Home

Default: /Users/daniel/github

Editor Settings

- Auto Indent
- Confirm Checkout Head Revision

Font Family

Font Size

14

Install ‘sort lines’



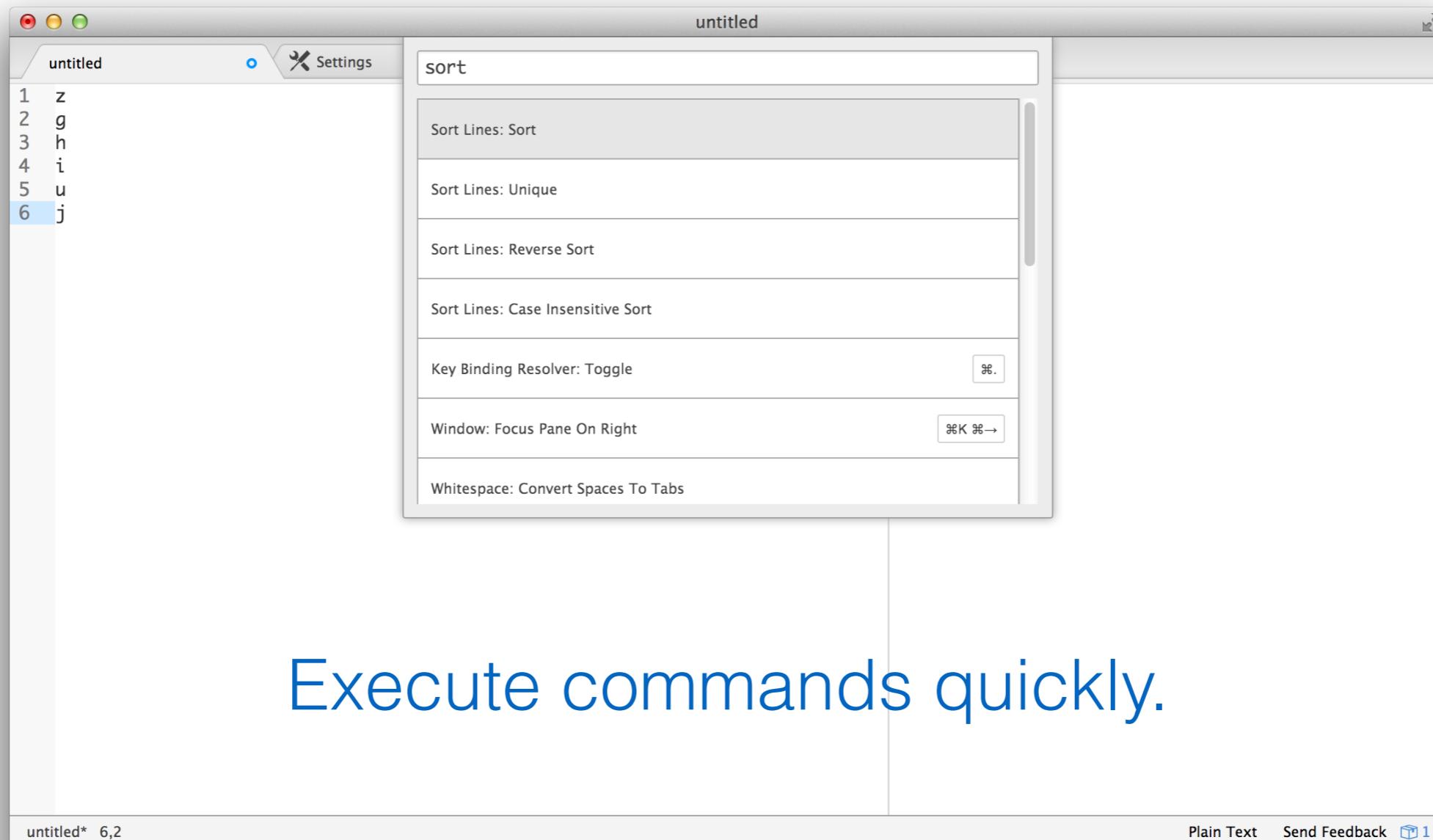
ATOM is extensible!

Atom's extensibility is demonstrated through its customisable menu system. The 'Edit' menu is open, showing various options like Undo (⌘Z), Redo (⌘Y), Cut (⌘X), Copy (⌘C), Copy Path (^⇧C), Paste (⌘V), Select All (⌘A), Toggle Comments (⌘/), Lines (selected), Text, Folding, Reflow Selection (⌥⌘Q), Bookmark, Go to Line (^G), Select Grammar (^⇧L), Start Dictation..., and Special Characters... (^⌘Space). A secondary submenu for 'Lines' is also open, listing Indent (⌘]), Outdent (⌘[), Auto Indent, Move Line Up (^⌘↑), Move Line Down (^⌘↓), Duplicate Lines (^⇧⌘D), Delete Line (^⇧K), Join Lines (⌘J), Sort, Reverse Sort, Unique, and Sort (Case Insensitive). The 'Join Lines' option is highlighted with a red box.

New functions!



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



Execute commands quickly.

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

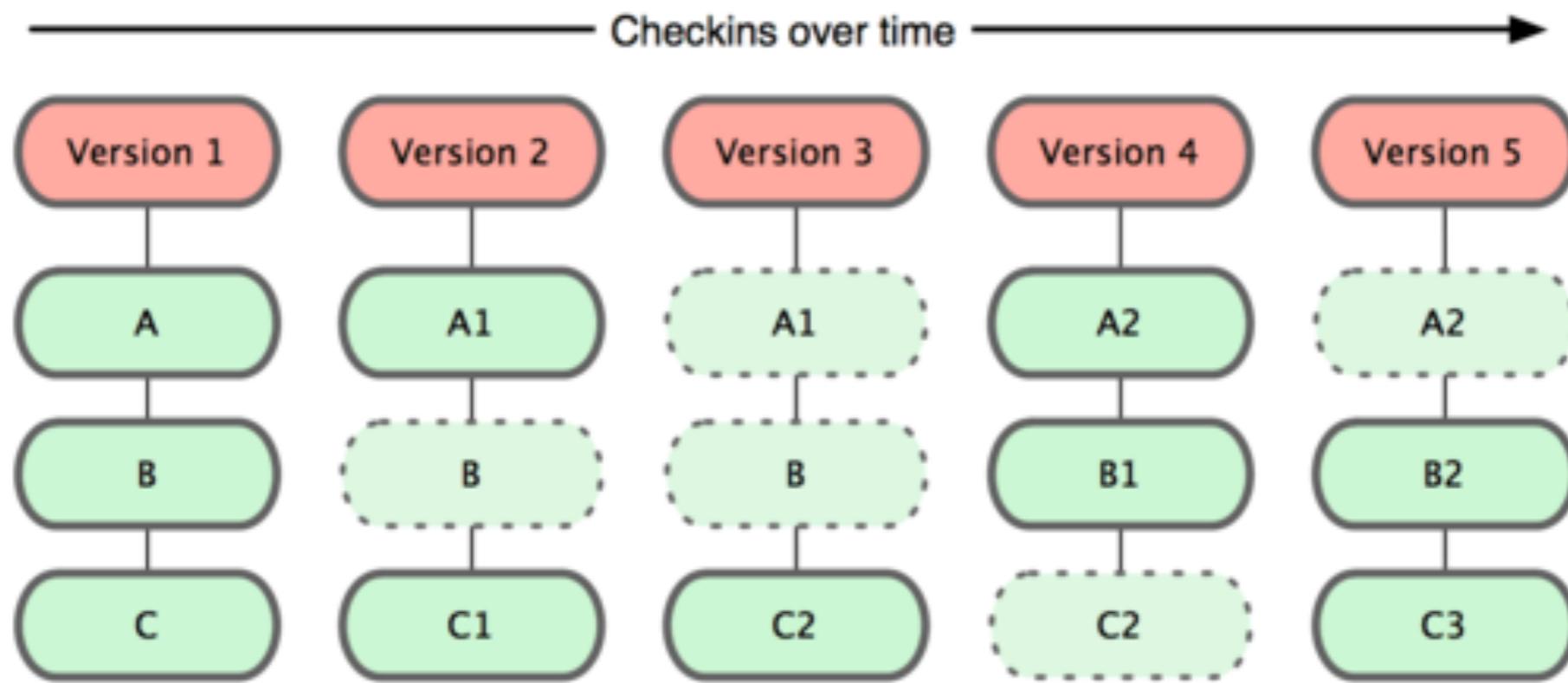
- **Intro to reproducible research**
- **Basic command line**
- **Directory structures**
- **Markdown**
- **Git and github**



git

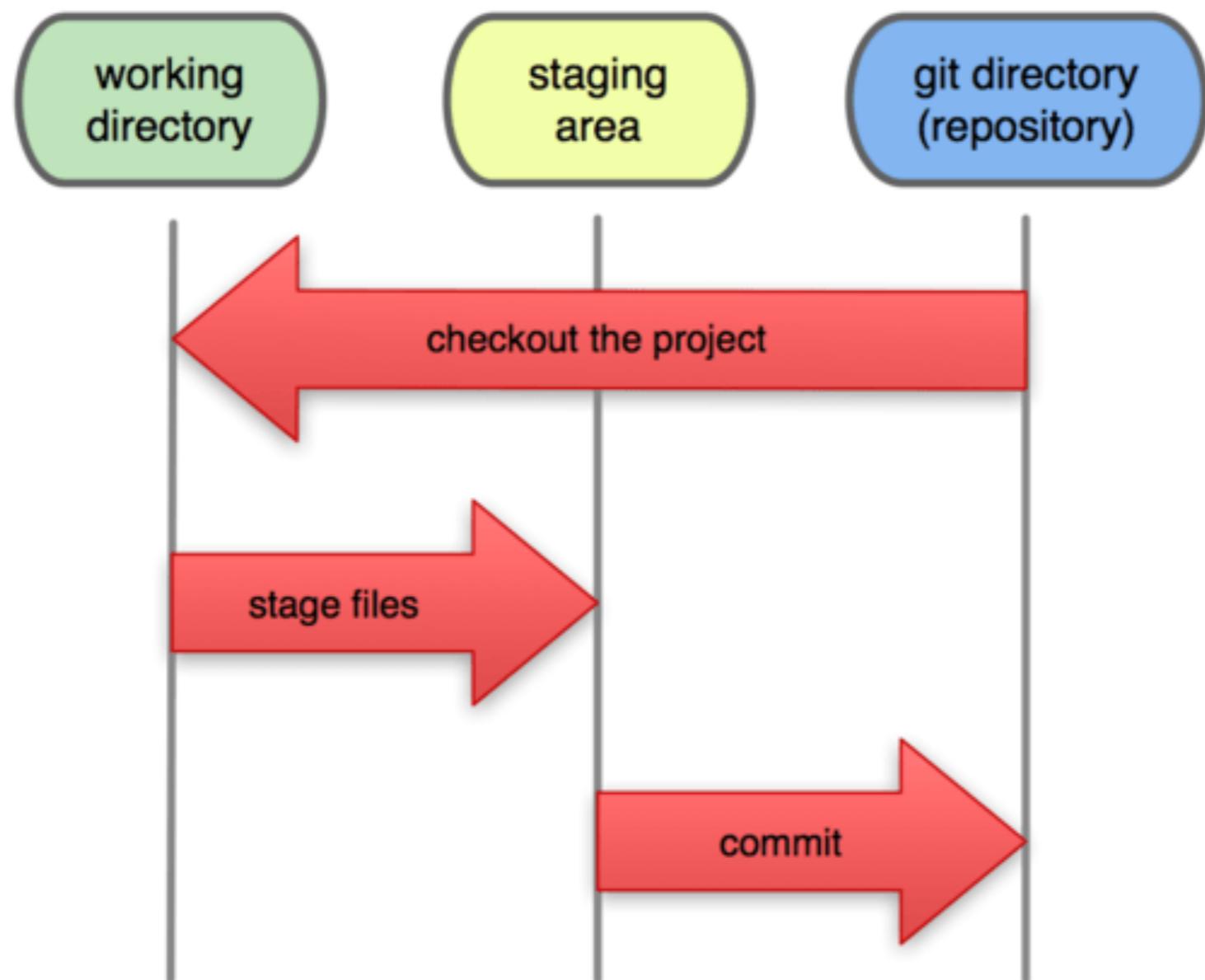
is a version control system

git tracks changes and files over time



git tracks local changes

Local Operations

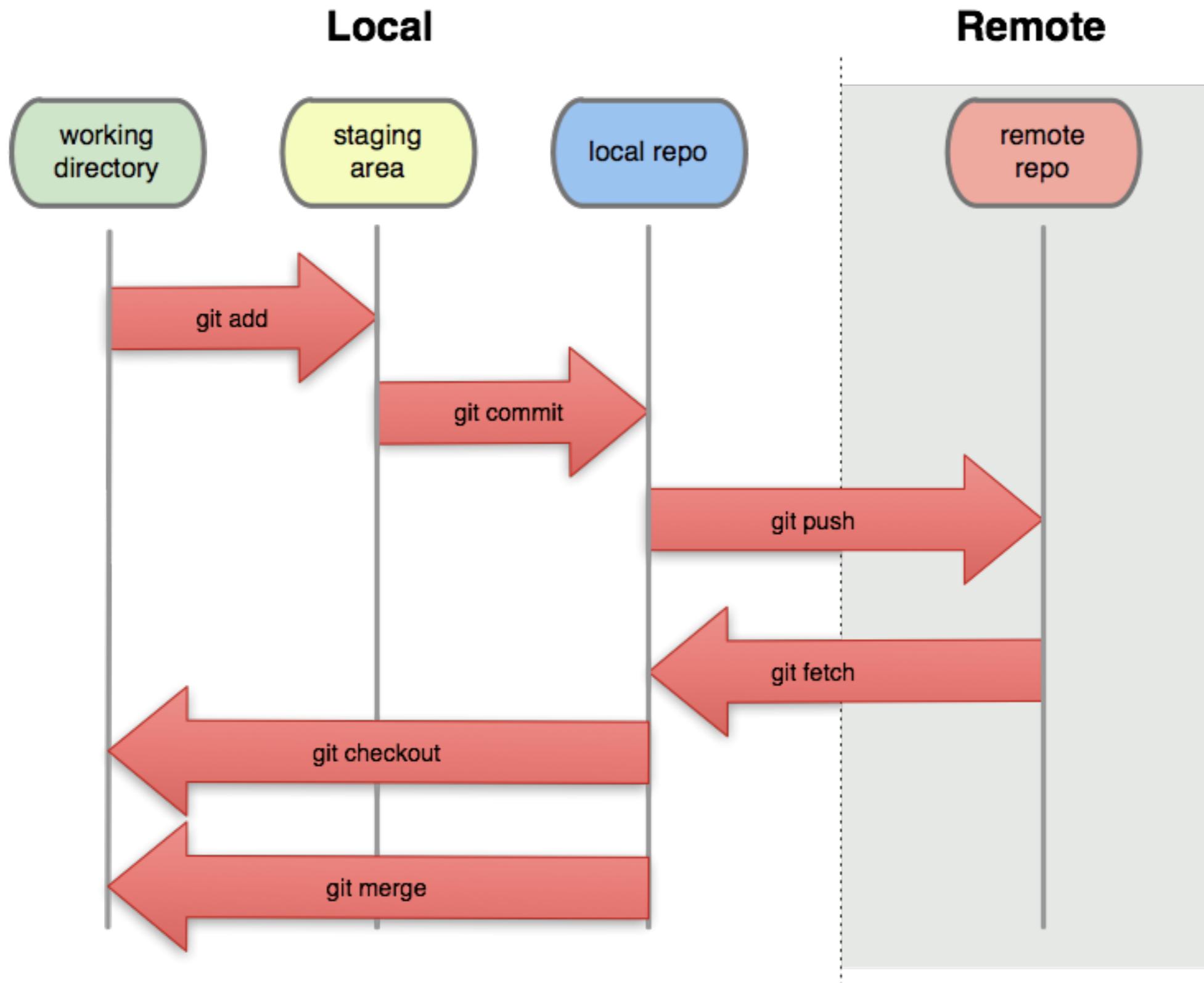




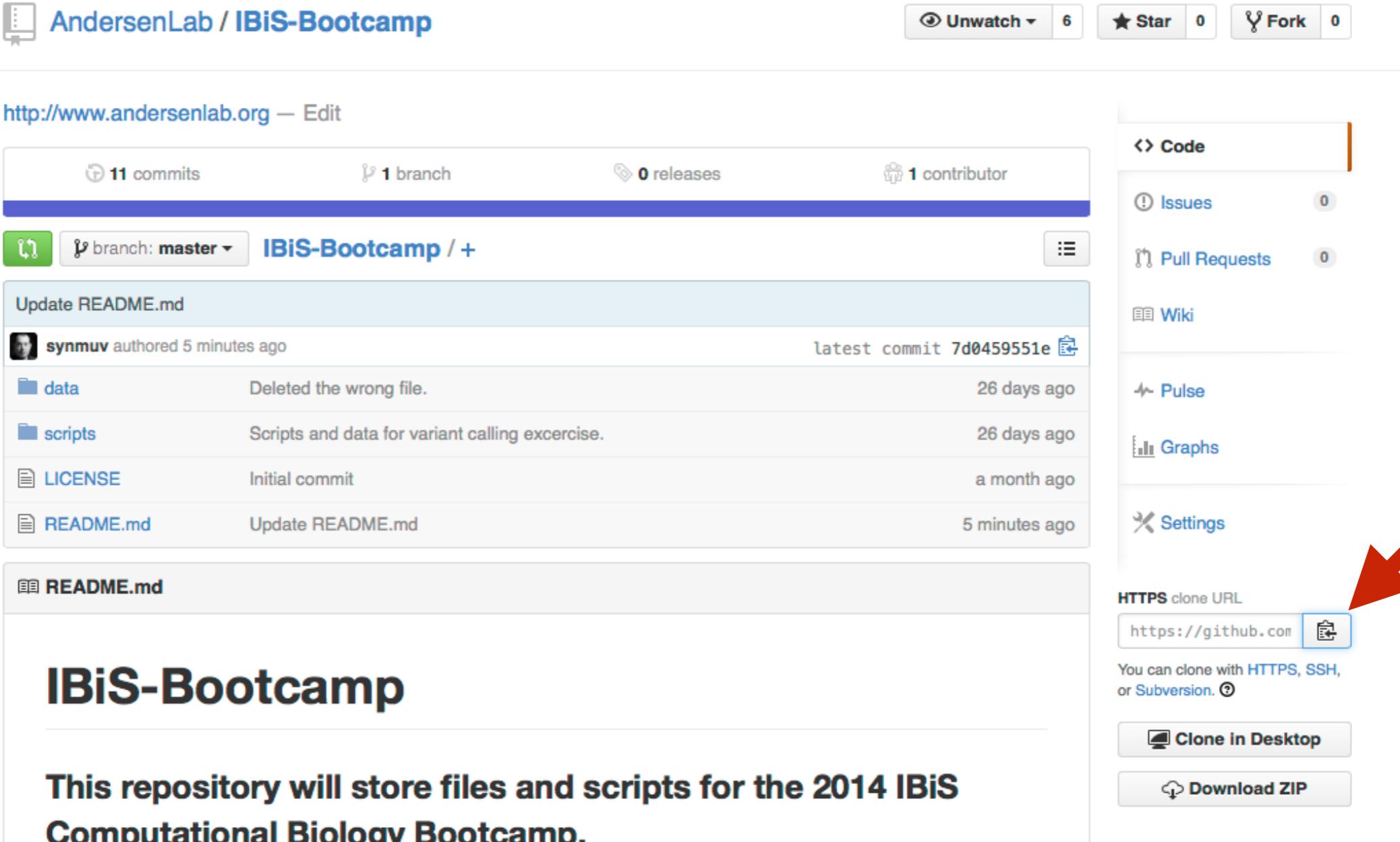
Github makes git super powerful

Check out: www.github.com/andersenlab

github takes git from local to the world



Let's clone the IBiS Bootcamp repository for work tomorrow

A screenshot of a GitHub repository page for "AndersenLab / IBiS-Bootcamp". The page shows 11 commits, 1 branch (master), 0 releases, and 1 contributor. A red arrow points to the "HTTPS clone URL" button at the bottom right of the page.

<http://www.andersenlab.org> — Edit

11 commits 1 branch 0 releases 1 contributor

branch: master + IBiS-Bootcamp / +

Update README.md

 **synmuv** authored 5 minutes ago latest commit `7d0459551e` 

File	Description	Time
data	Deleted the wrong file.	26 days ago
scripts	Scripts and data for variant calling excercise.	26 days ago
LICENSE	Initial commit	a month ago
README.md	Update README.md	5 minutes ago

 README.md

IBiS-Bootcamp

This repository will store files and scripts for the 2014 IBiS Computational Biology Bootcamp.

Code Issues Pull Requests Wiki Pulse Graphs Settings

HTTPS clone URL <https://github.com> 

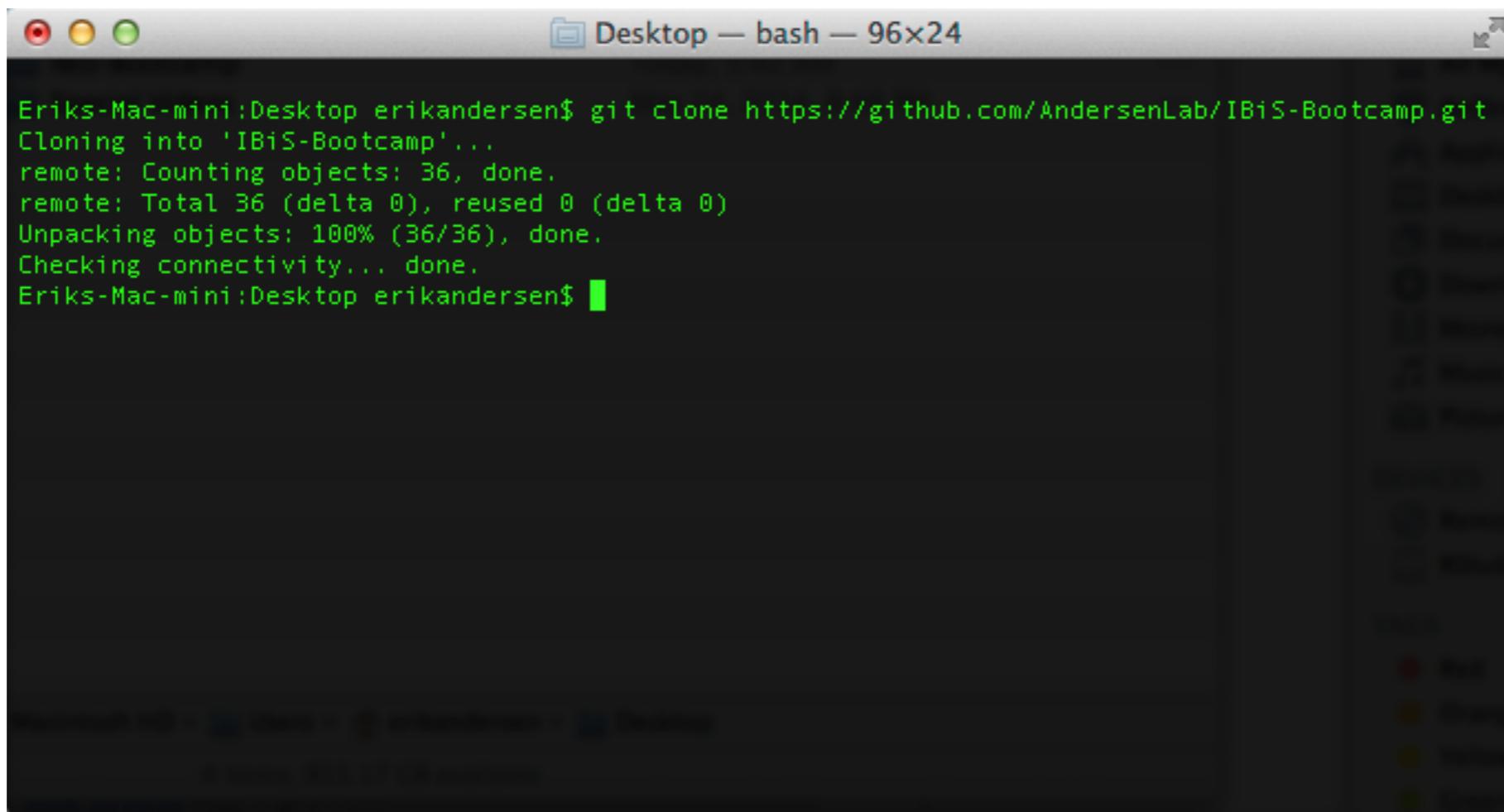
You can clone with [HTTPS](#), [SSH](#), or [Subversion](#). 

 [Clone in Desktop](#)

 [Download ZIP](#)

Let's clone the IBiS Bootcamp repository for work tomorrow

git clone invokes git to clone a repository



Eriks-Mac-mini:Desktop erikandersen\$ git clone https://github.com/AndersenLab/IBiS-Bootcamp.git
Cloning into 'IBiS-Bootcamp'...
remote: Counting objects: 36, done.
remote: Total 36 (delta 0), reused 0 (delta 0)
Unpacking objects: 100% (36/36), done.
Checking connectivity... done.
Eriks-Mac-mini:Desktop erikandersen\$

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`

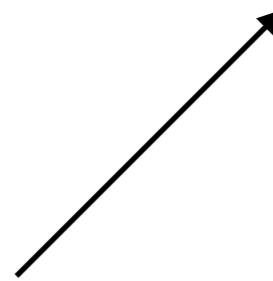
Commit files to local repo `git commit -m "short message"`

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.

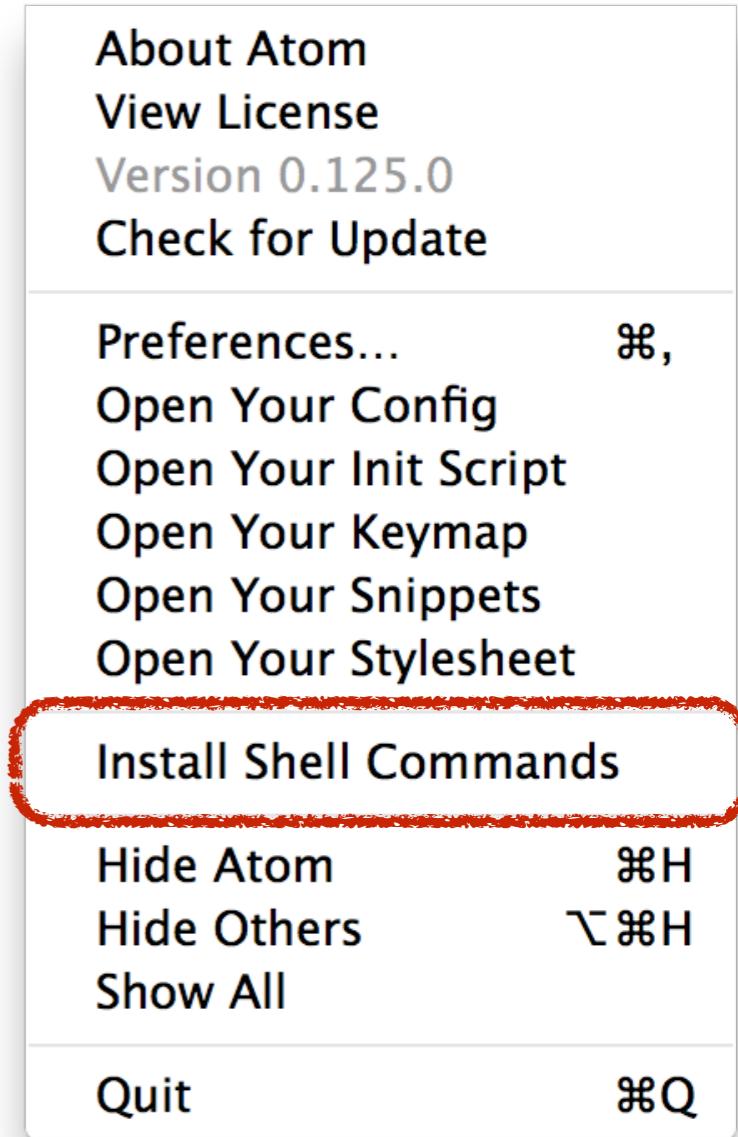
Editing your bash profile

1. Install the atom shell commands.

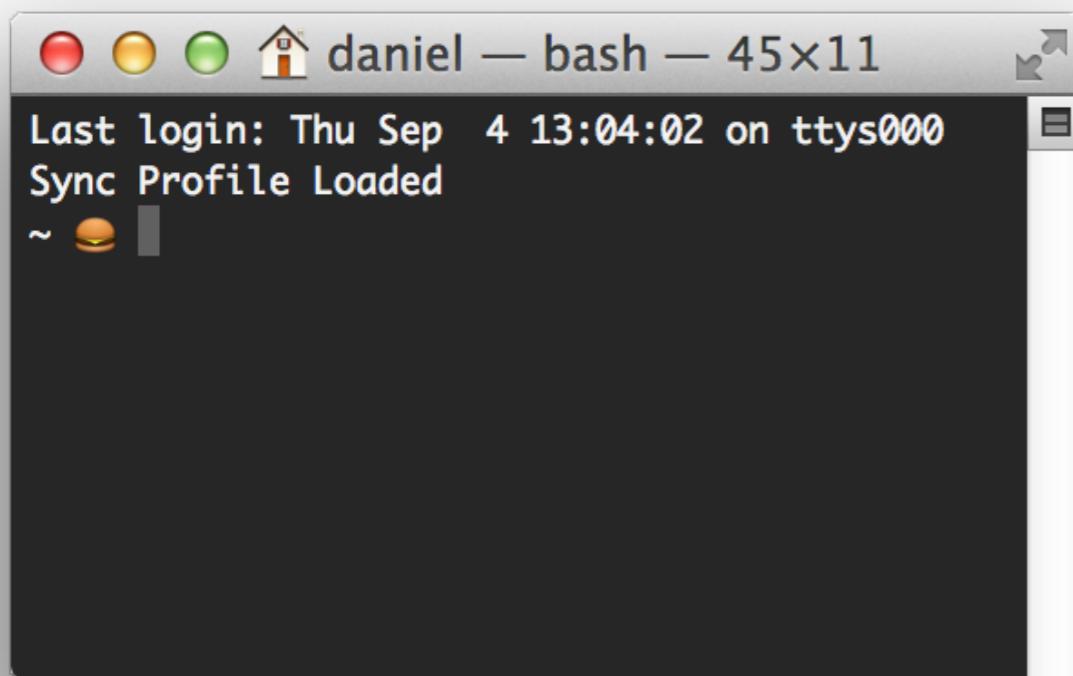
[Atom > Install Shell Commands](#)

2. Type:

`atom .bash_profile`



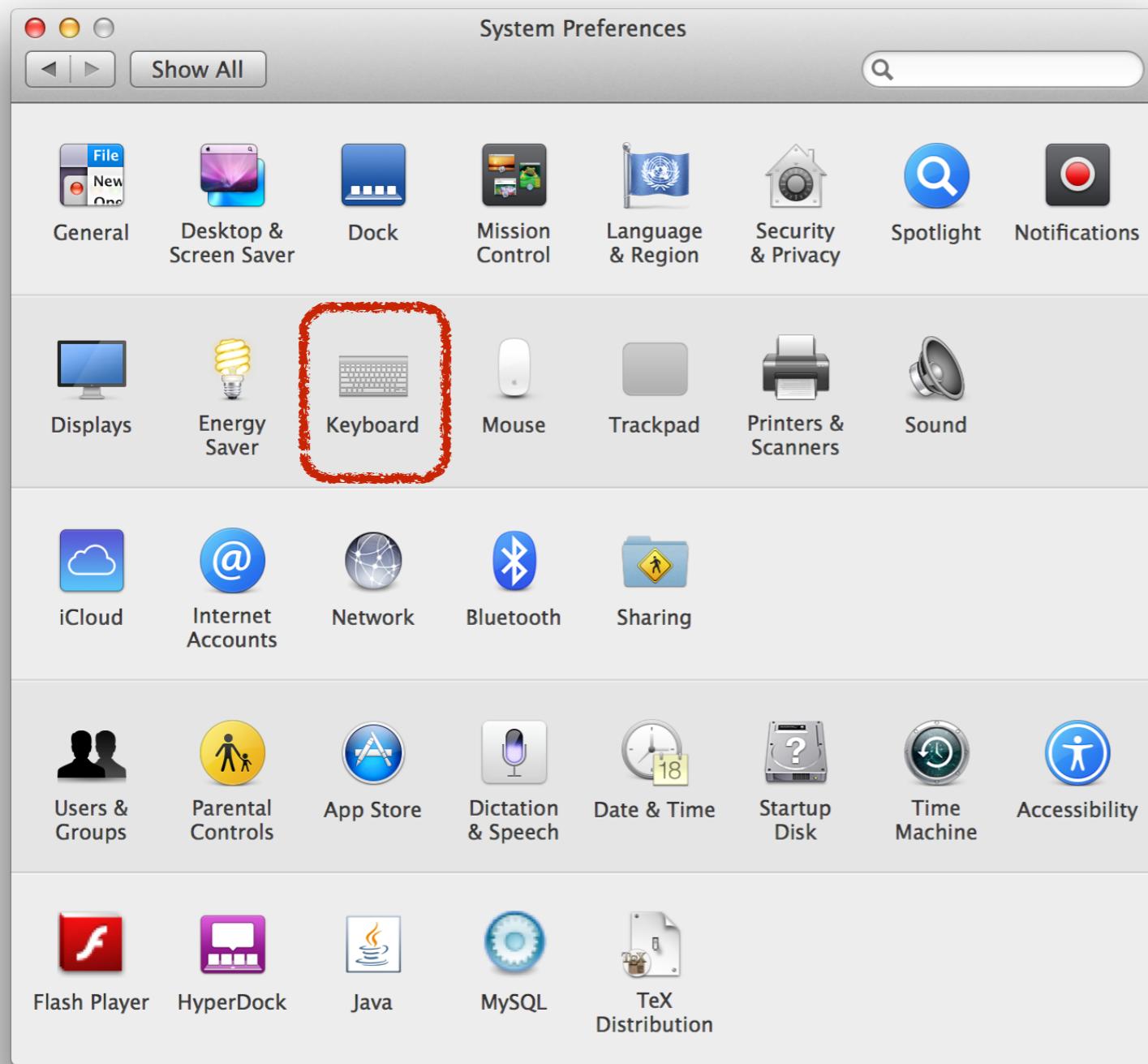
Personalize your Terminal



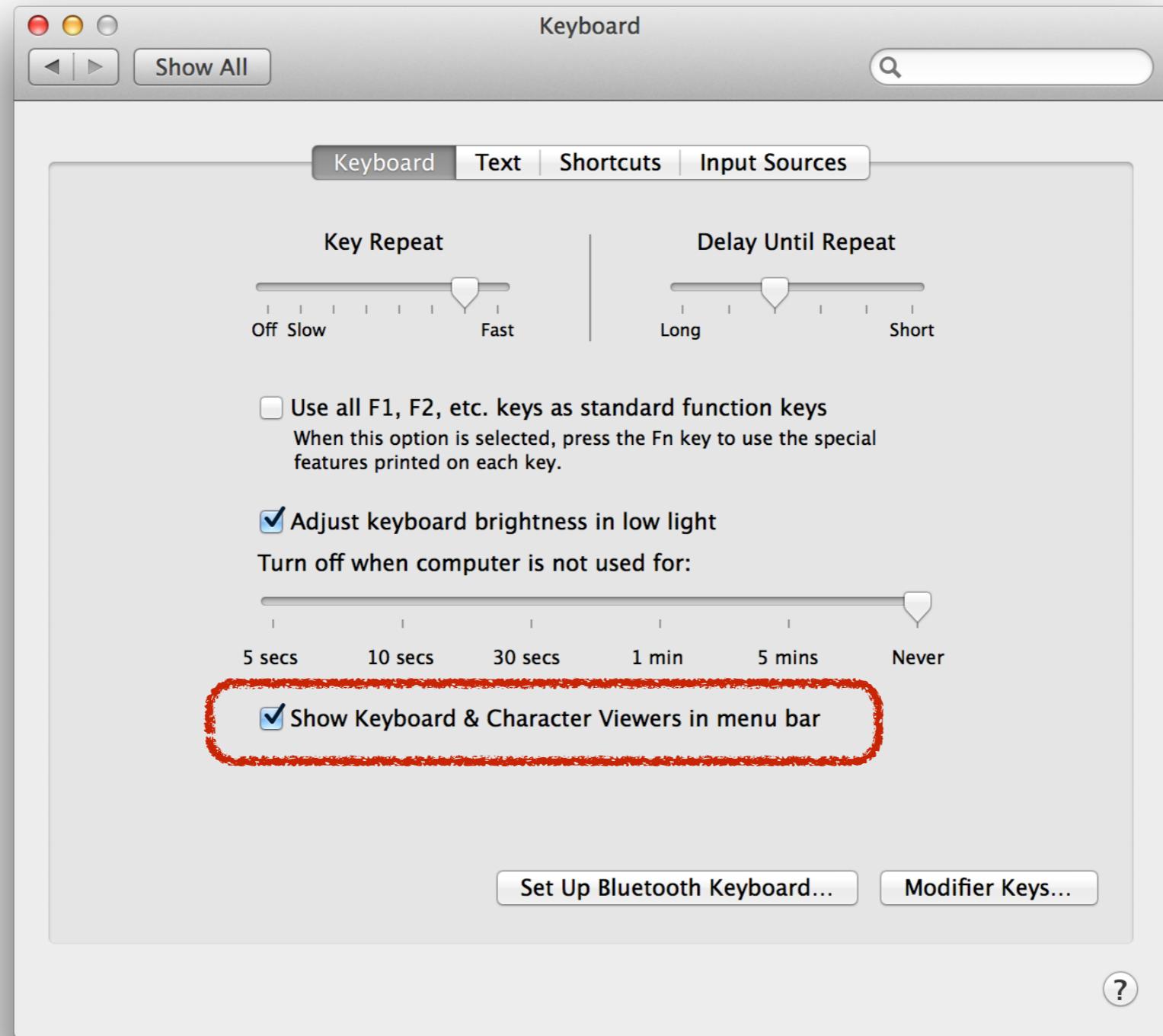
Add an icon to see command
lines



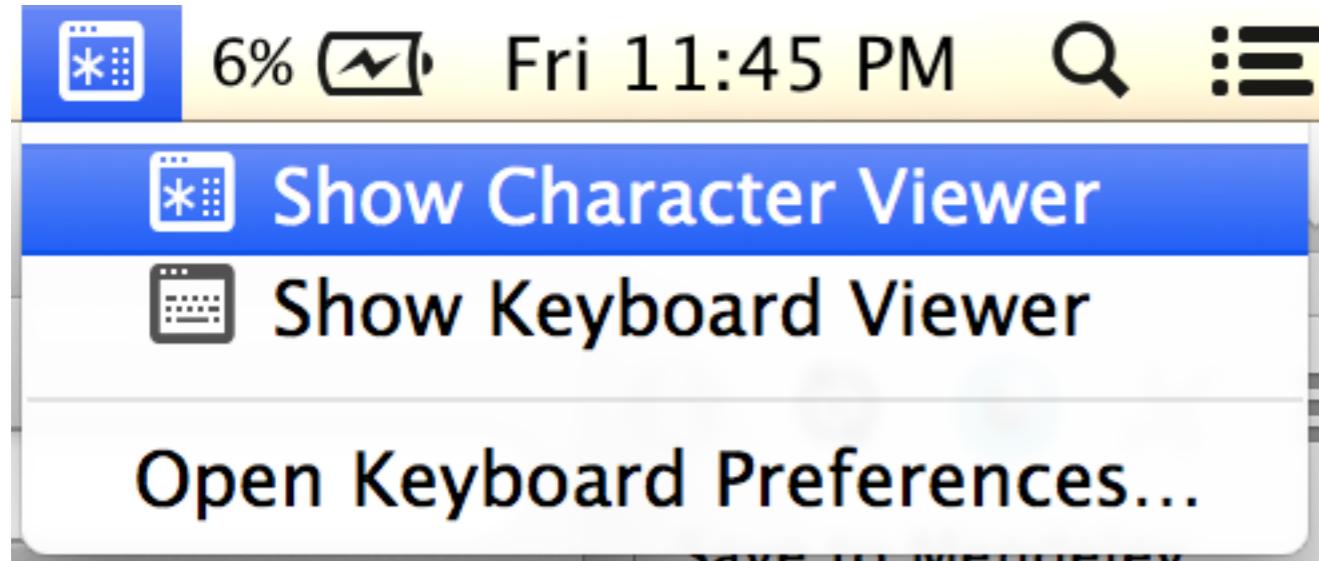
Open System Settings



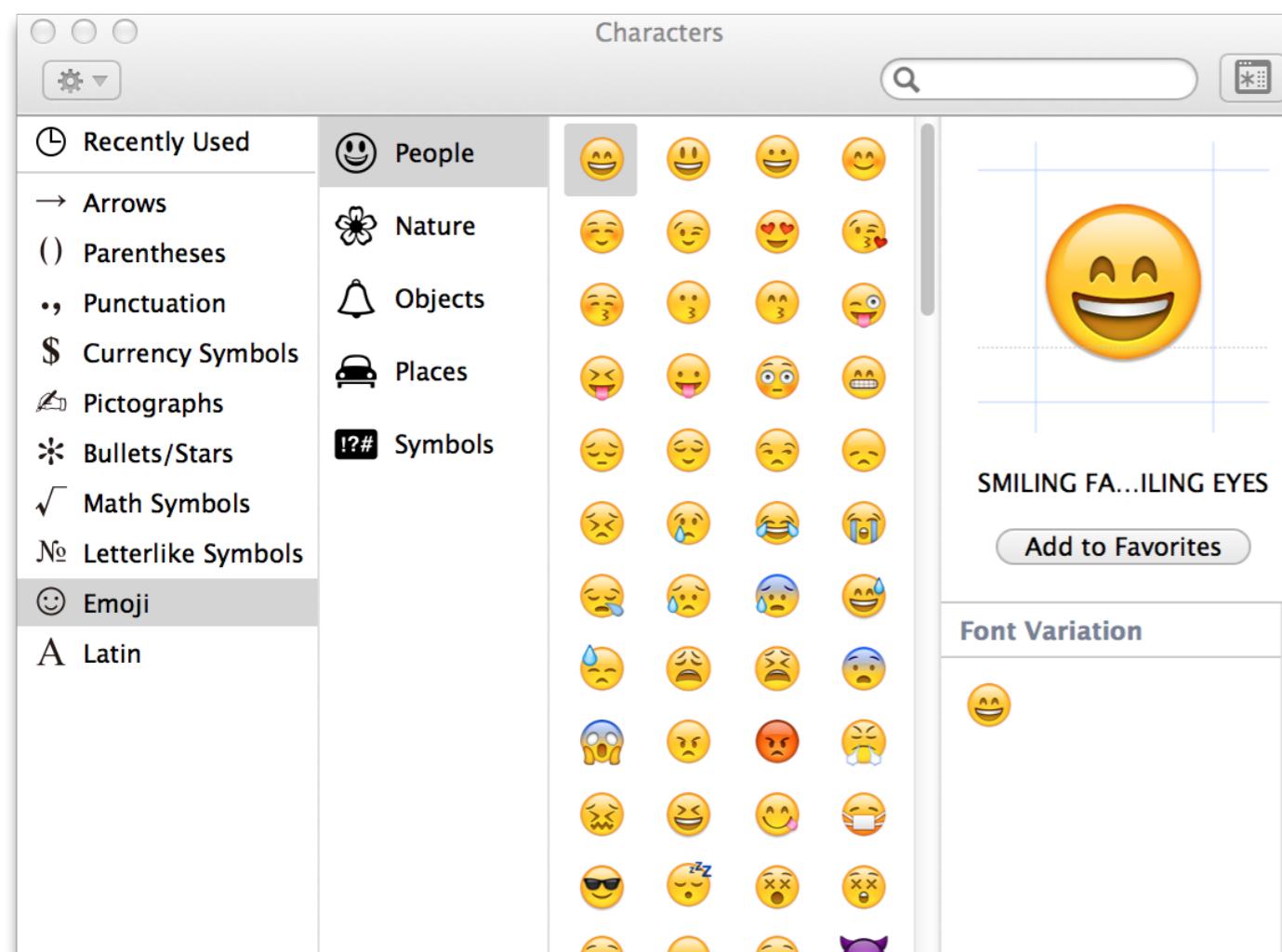
Click “Keyboard”



Check “Show Keyboard & Character Viewers in menu bar”

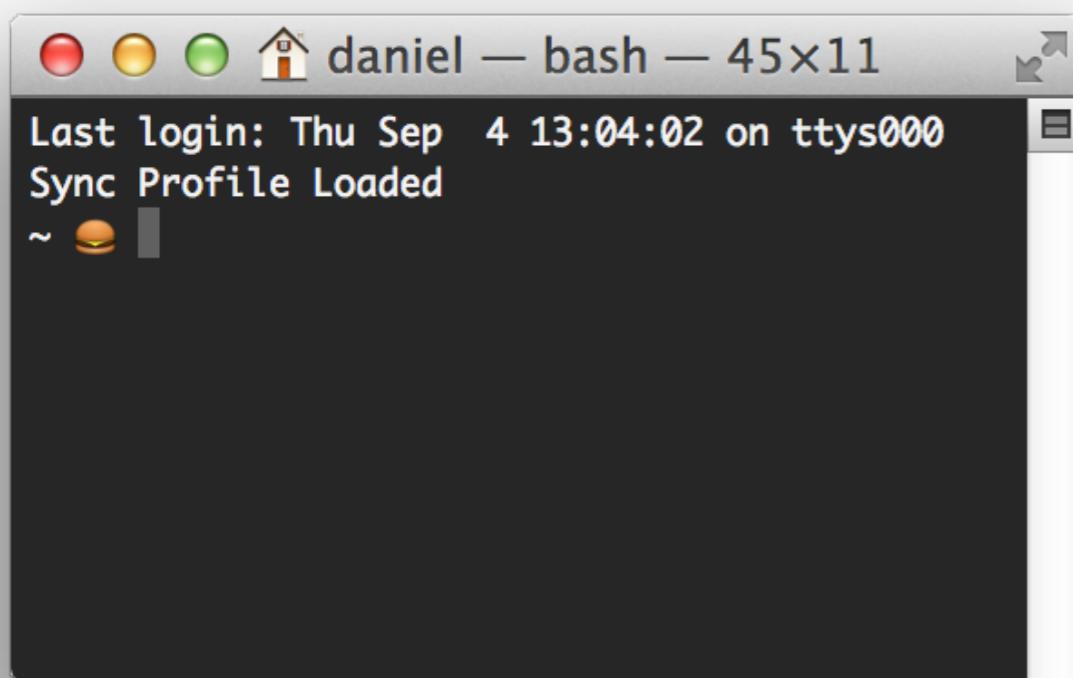


Open the character viewer.



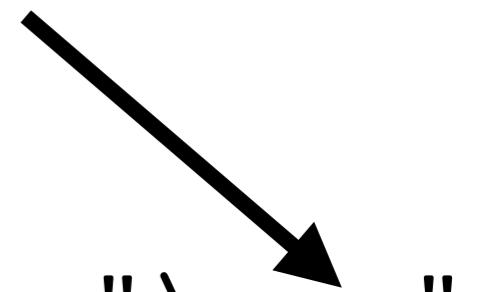
Pick a fun icon!

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.

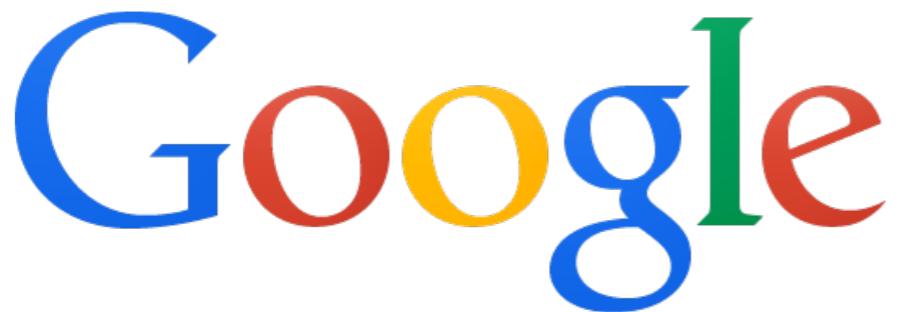
```
new_project() {  
    mkdir $1  
    touch $1/Readme.md  
    mkdir $1/Data  
    mkdir $1/Data/Raw  
    mkdir $1/Data/Processed  
    mkdir $1/Scripts  
    mkdir $1/Results  
}
```

This function will create a project directory structure.

Example:

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

What to do when you are lost?



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.

The screenshot shows the GitHub Explore page with a light gray header containing the GitHub logo, a search bar, and navigation links for Explore, Gist, Blog, and Help. On the right side of the header are user profile icons and account settings. Below the header, the main navigation bar includes 'Explore GitHub' on the left and 'All' (selected), Showcases, Trending, and Stars on the right. The main content area is titled 'Explore' with the subtitle 'Browse interesting projects, solving all types of interesting problems.' Below this, there are several cards representing different project categories:

- Writing**: A green card with a background pattern of white circles.
- Open Journalism**: A purple card with a background pattern of white circles.
- Projects with great wikis**: A teal card with a background pattern of white waves. Below it, a text snippet reads: "These projects all use GitHub Wikis to share documentation and helpful r...".
- Productivity tools**: A blue card with a background pattern of white squares.
- Web games**: A red card with a background pattern of white triangles. Below it, a text snippet reads: "Who says that coding has to be boring? Have some fun with these open sou...".
- Text editors**: A large green card with a background pattern of white chevrons.
- Projects that power GitHub**: A green card with a background pattern of white triangles.
- Game off winners**: A pink card with a background pattern of white squares.

At the bottom center of the page is a link labeled 'See all ➔'.