

# Welcome to Day 1!

Complex Human Data Summer  
School

Wifi name: Visitor wireless  
(or viswireless)

Username: chd-school

Password: #dsG9y

# Introductions

# Summer School Overview

- Day 0: R bootcamp
- Day 1: Workflow, Google App Engine
- Day 2: Online Experiments
- Day 3: Data wrangling, visualization
- Day 4: Statistics, Probabilistic models
- Day 5: Experience sampling

# Announcements

- Piazza
- Photo today: 4:40 at steps of Melbourne School of Design
- Dinner tonight: 5:00 in the courtyard of the Redmond Barry building

# Day 1: Workflow, Google App Engine

1. Git
2. Project organization

(lunch)

3. Internet and webpages
4. Deploying to Google App Engine

# Why workflow matters

- A good workflow
  - saves time in the long run
  - helps you avoid errors
  - makes it easy for others to reproduce your work

# Replication crisis





# Reproducing computational work

**MAY 04 2015**

**12 COMMENTS**

DATA SHARING, JOURNALS &  
REPLICATION

## LEADING JOURNAL VERIFIES ARTICLES BEFORE PUBLICATION – SO FAR, ALL REPLICATIONS FAILED

The American Journal of Political Science recently announced that every article will be externally verified before it is published. Authors have to provide all data, code and materials which will be verified by an external statistician. I talked to the editor William G. Jacoby about the goals of the new policy, and why no author has provided ‘perfect’ files yet.

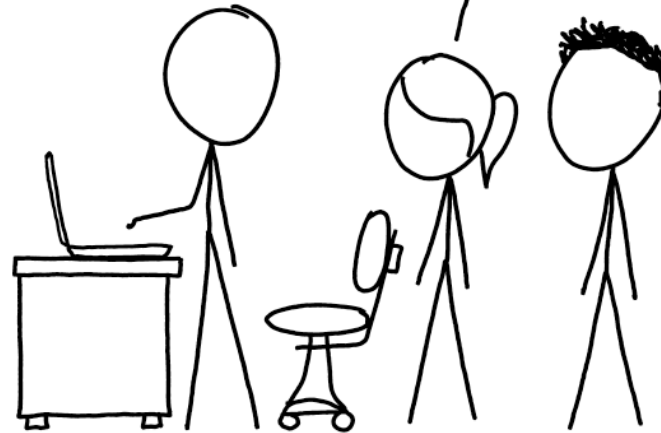
# Version control with Git

- Why use Git?
  - keep a record of what you were doing and thinking during the lifecycle of a project
  - “version control is like an undo command for *everything*”
  - have a backup if your laptop dies or is stolen
  - collaborate on code and writing (optional)
  - release code and data publicly (optional)

THIS IS GIT. IT TRACKS COLLABORATIVE WORK  
ON PROJECTS THROUGH A BEAUTIFUL  
DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZE THESE SHELL  
COMMANDS AND TYPE THEM TO SYNC UP.  
IF YOU GET ERRORS, SAVE YOUR WORK  
ELSEWHERE, DELETE THE PROJECT,  
AND DOWNLOAD A FRESH COPY.



# Exercises (follow tutorial)

## 1. Install Git

- Mac: install Xcode command line tools from the App Store
- Windows: <https://gitforwindows.org/>

## 2. Check Git is installed

## 3. Set user.name and user.email

# GitHub

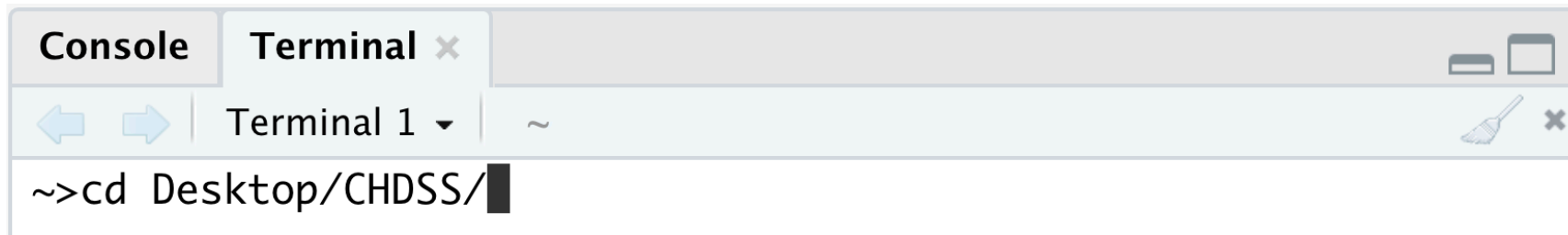
GitHub (and other similar sites) allows you to host repositories remotely.

## Exercise:

- Sign up for a GitHub account. Use a university email if possible, and choose your username wisely!

# Exercise: Cloning a repository

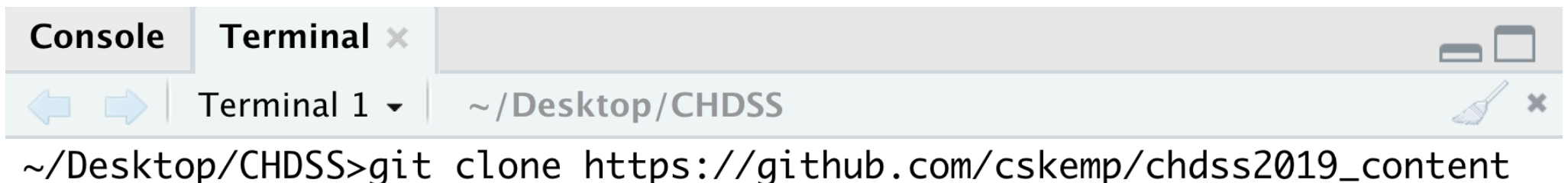
1. If you didn't already, make a folder called CHDSS on your desktop.
2. Open a terminal and change into your CHDSS folder



A screenshot of a terminal window. The title bar shows 'Console' and 'Terminal x'. The terminal content shows the prompt '~>' followed by the command 'cd Desktop/CHDSS/' and a cursor. Navigation arrows and a dropdown menu labeled 'Terminal 1' are visible on the left. Icons for a folder and a close button are on the right.

```
~>cd Desktop/CHDSS/
```

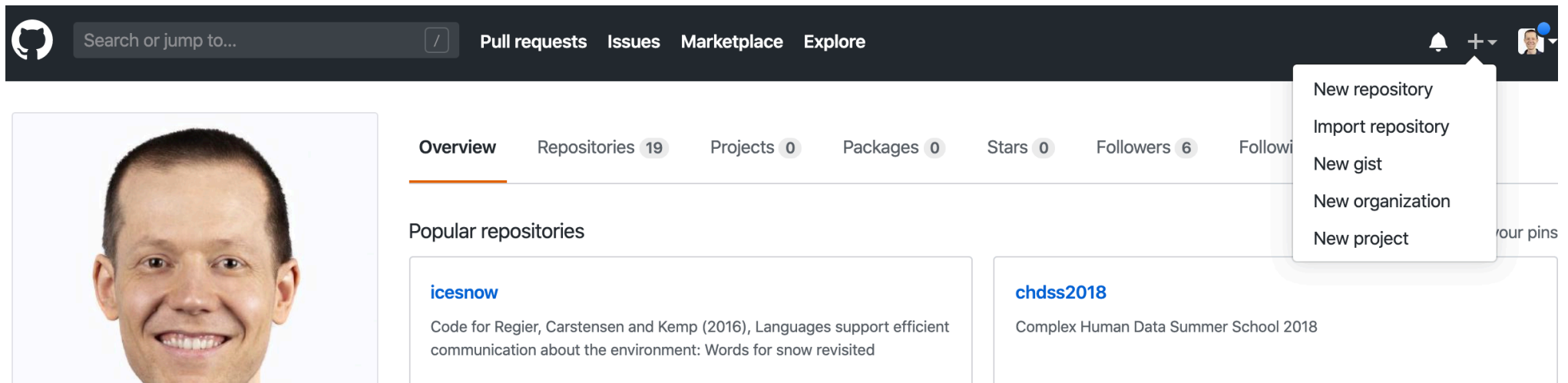
3. Clone the repository for summer school content



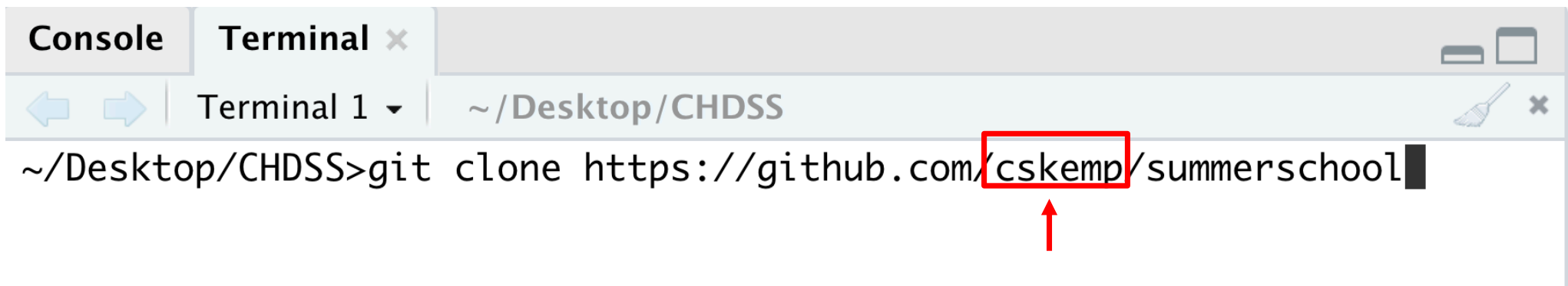
A screenshot of a terminal window. The title bar shows 'Console' and 'Terminal x'. The terminal content shows the prompt '~/Desktop/CHDSS>' followed by the command 'git clone https://github.com/cskemp/chdss2019\_content'. Navigation arrows and a dropdown menu labeled 'Terminal 1' are visible on the left. Icons for a folder and a close button are on the right.

```
~/Desktop/CHDSS>git clone https://github.com/cskemp/chdss2019_content
```

# Exercise: make your own repository



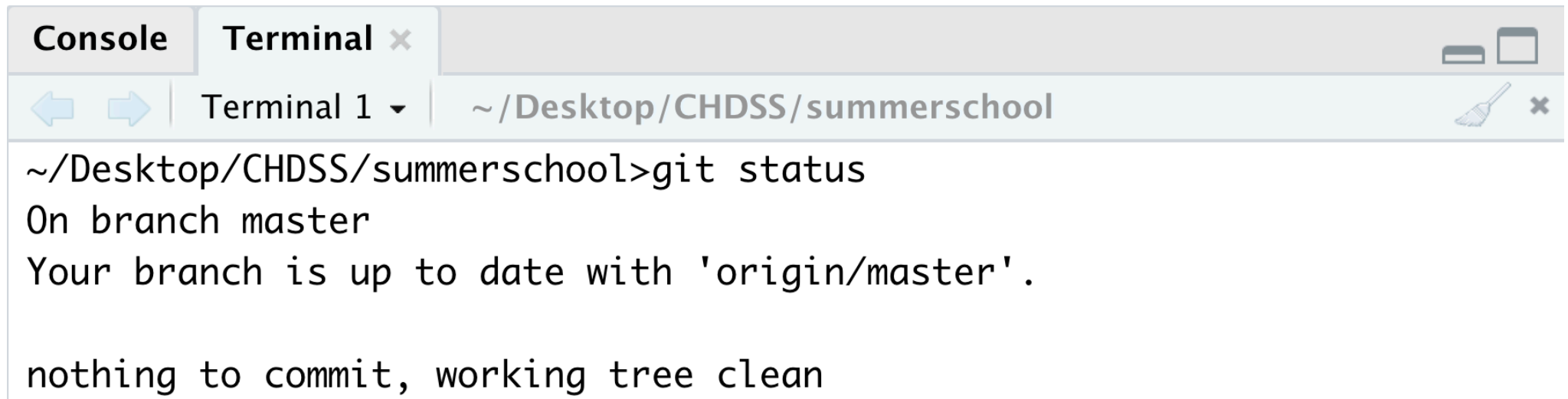
When you're done, clone your repository:



**Replace with your own username**

# Exercises

## 1. Check the status of your repository

A screenshot of a terminal window. The title bar shows 'Console' and 'Terminal x'. Below the title bar, there's a tab labeled 'Terminal 1' and a path '~ / Desktop / CHDSS / summerschool'. The terminal content shows the command 'git status' being executed, with the output: 'On branch master', 'Your branch is up to date with 'origin/master'.', and 'nothing to commit, working tree clean'.

```
~/Desktop/CHDSS/summerschool>git status
On branch master
Your branch is up to date with 'origin/master'.

nothing to commit, working tree clean
```

2. Edit README.md and play around with markdown
3. Check the status again



# Exercises

1. > git add -all
2. > git status
3. > git commit -m"updated  
README.md"
4. > git status
5. > git push
6. > git status

Also check the online GitHub page for your repository to confirm that your edits were uploaded successfully

# Collaborating using GitHub

- The tutorial shows how to add a collaborator so that both of you can work on the same set of files.
- Exercise (later):
  - Find a partner and ask her to add you as a collaborator on her repository. Clone her repository, make a change, and push your edit to her repository.

# Using the issue tracker

## Rules

**(A) Manage tasks with a task management system.**

**(B) E-mail is not a task management system.**

(Gentzkow + Shapiro, Code and Data for the social sciences: a practitioner's guide)

Filters ▾

Labels 9

Milestones 0

New issue

<input type="checkbox"/>	3 Open ✓ 4 Closed	Author ▾	Labels ▾	Projects ▾	Milestones ▾	Assignee ▾	Sort ▾
<input type="checkbox"/>	<b>Missing information in dice experiments</b> #6 opened on Aug 7 by perfors						1
<input type="checkbox"/>	<b>Standardise preprocessing scripts</b> #5 opened on Aug 1 by djnavarro						3
<input type="checkbox"/>	<b>Data analysis scripts</b> #2 opened on Jul 27 by perfors						

# Using the issue tracker

## **Rules**

**(A) Manage tasks with a task management system.**

**(B) E-mail is not a task management system.**

(Gentzkow + Shapiro, Code and Data for  
the social sciences: a practitioner's guide

- **Exercise:**
  - Add an issue to your repository.
  - Add a comment to the issue
  - Add a final comment and close the issue