

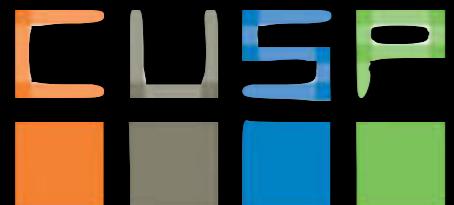
# Urban Informatics

Fall 2018

dr. federica bianco [fbianco@nyu.edu](mailto:fbianco@nyu.edu)

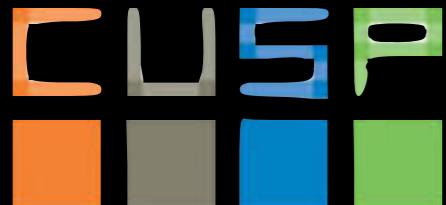


@fedhere



# Summary:

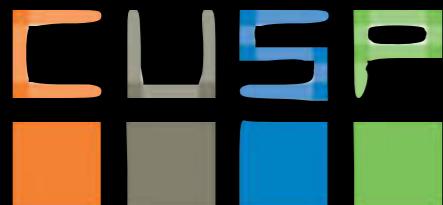
- **Epistemological concepts:**  
falsifiability, law of parsimony
- **Good Scientific Practices:**  
environment set up and environmental variables



## Summary:

- **Epistemological concepts:**  
falsifiability, law of parsimony
- **Good Scientific Practices:**  
environment set up and environmental variables

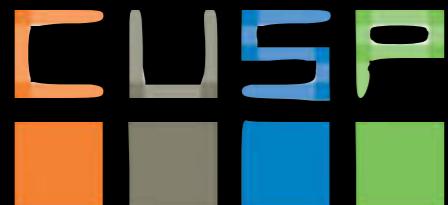
Karl Popper - The Logic of Scientific Discovery:  
*asymmetry* between verifiability and falsifiability



# Summary:

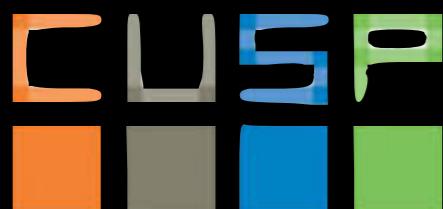
- **Epistemological concepts:**  
falsifiability, law of parsimony
- **Good Scientific Practices:**  
environment set up and environmental variables

Using environmental variables to facilitate reproducibility  
and protect information

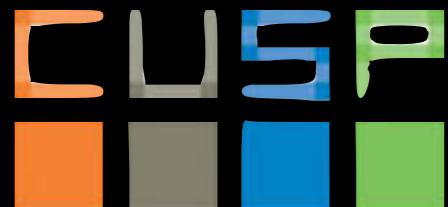


## Summary:

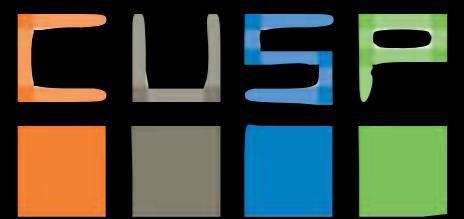
- falsifiability and law of parsimony
- types of scientific questions
- understand how to set up your environment
- basic bash commands
- reproducible research
- work with github
- creating and checking into github a jupyter notebook
- PEP8 and style standards



# *The practical side of things*



II: Reproducibility  
& data types



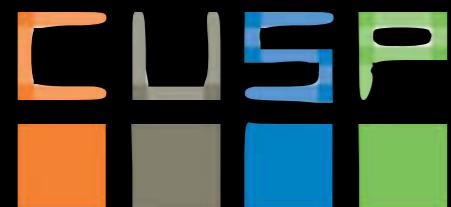
II: Reproducibility  
& data types

# Operating System:

Android, BlackBerry, BSD, Chrome OS, OS X, iOS, QNX, Linux  
Steam OS

Microsoft Windows, Windows Phone, and z/OS.

FreeRTOS, Micrium, and VxWorks



# Operating System:

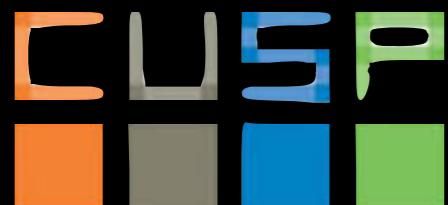
Android, BlackBerry, BSD, Chrome OS, OS X, iOS, QNX, Linux  
Steam OS



Where there is a shell, there is a way.

Microsoft Windows, Windows Phone, and z/OS.

FreeRTOS, Micrium, and VxWorks

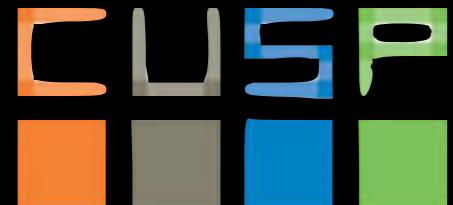


# Operating System:

OS X

Linux **UNIX**

Where there is a shell, there is a way.

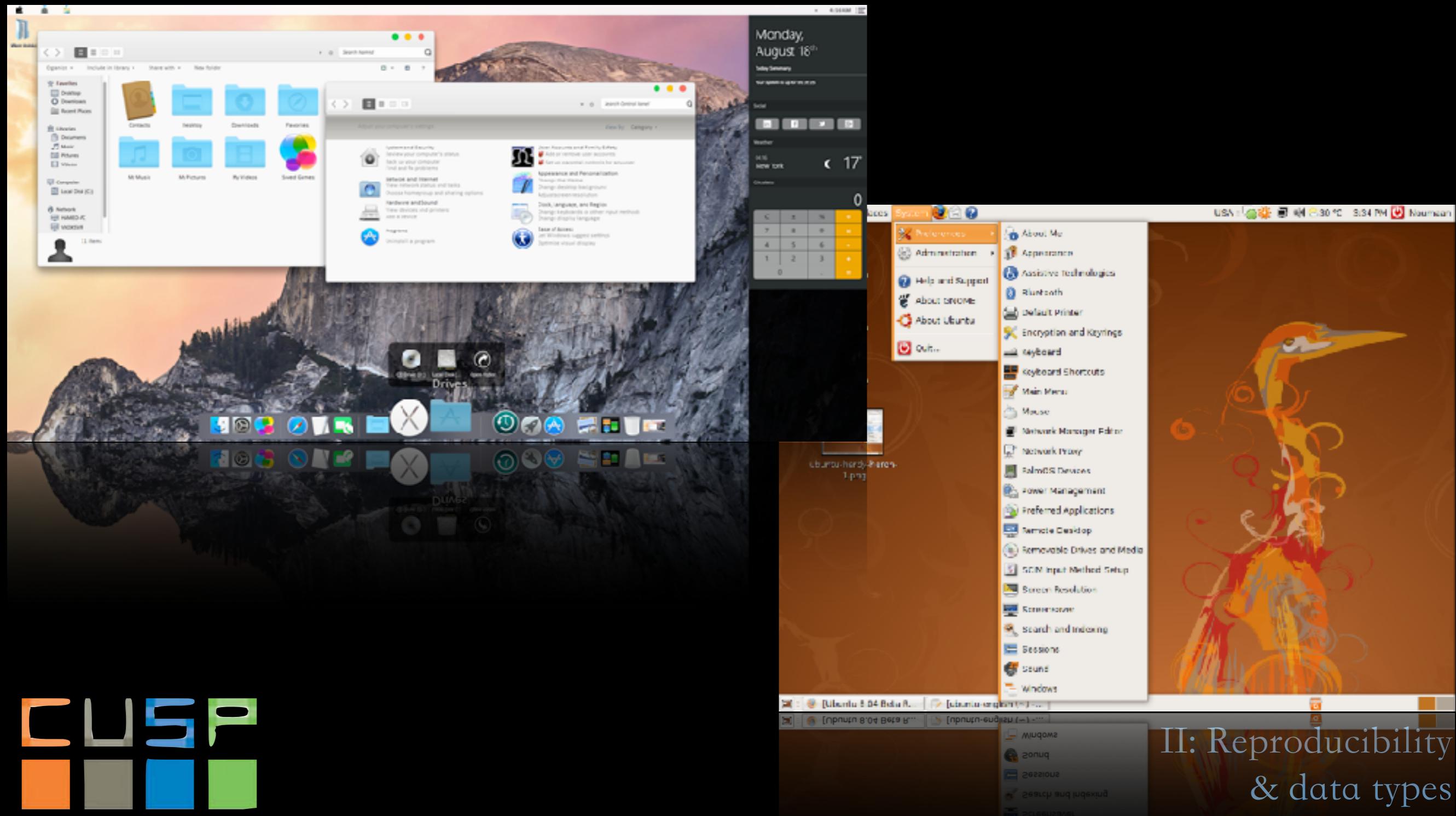


II: Reproducibility  
& data types

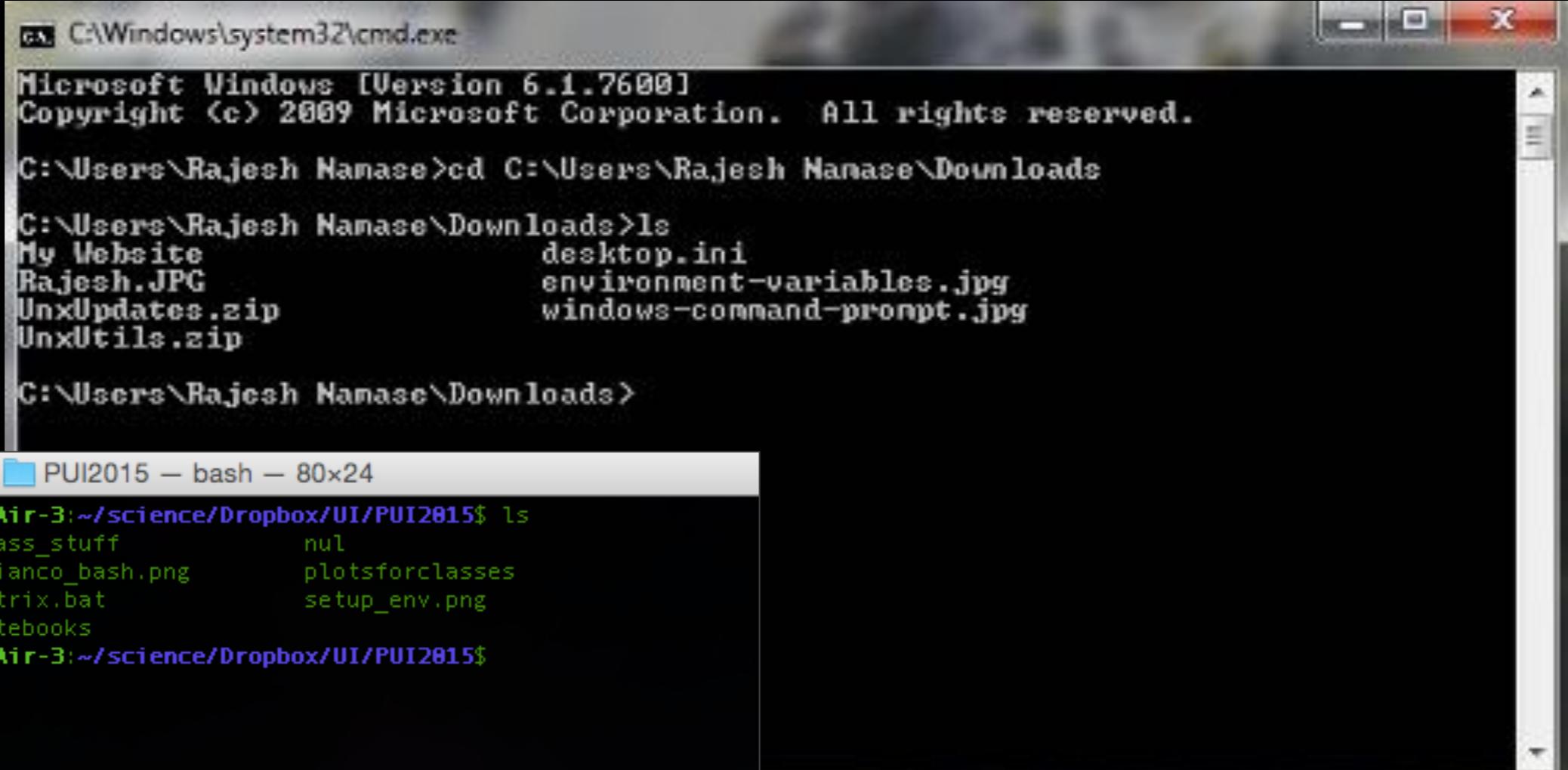
# Operating System:

OS X

Linux



# Operating System:

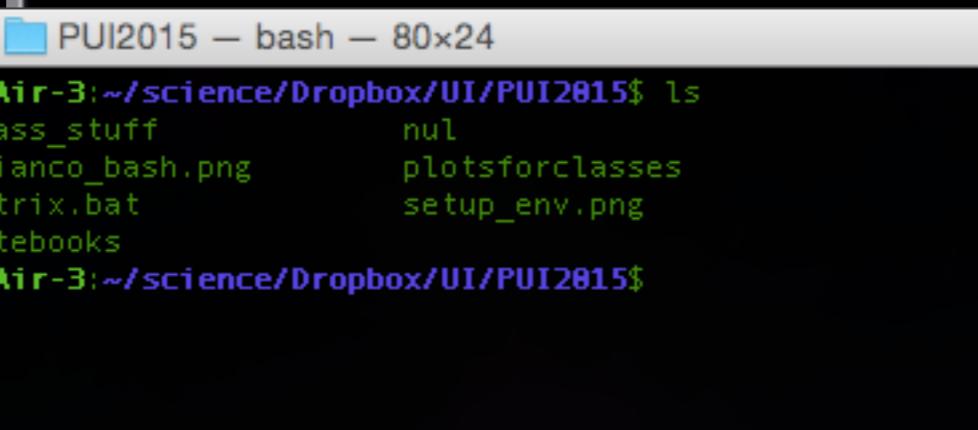


```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 6.1.7600]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Users\Rajesh Namase>cd C:\Users\Rajesh Namase\Downloads

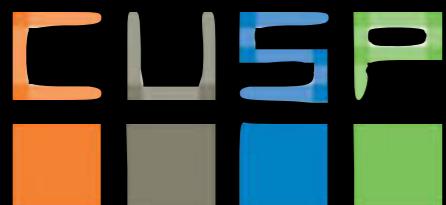
C:\Users\Rajesh Namase\Downloads>ls
My Website                               desktop.ini
Rajesh.JPG                                environment-variables.jpg
UnxUpdates.zip                           windows-command-prompt.jpg
UnxUtils.zip

C:\Users\Rajesh Namase\Downloads>
```

```
PUI2015 – bash – 80x24
fbianco@Federicas-MacBook-Air-3:~/science/Dropbox/UI/PUI2015$ ls
CityChallenge      class_stuff        nul
PUI2015labs        fbianco_bash.png   plotsforclasses
README.md          matrix.bat       setup_env.png
class_slides       notebooks
```

<https://speakerdeck.com/62gerente/bash-introduction>



# other solution and environment details...

<https://slides.com/mohitsharma44/cusp-pui-2018/live#>

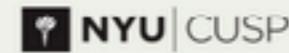


bash:

```
fb@fbMacBook-2: 12:24:05:~$computemc
#####
# CUSP RESEARCH COMPUTING FACILITY #
# CUSP-RCF #
#####
##### Welcome to NYU CUSP gateway server #####
# All connections are monitored and recorded #
# Disconnect IMMEDIATELY if you are not an authorized user! #
#
# WARNING-----WARNING-----WARNING #
# PLEASE DO NOT UPLOAD OR CREATE ANY FILE IN YOUR LOCAL DIRECTORY #
# ON THIS SERVER (gateway server), ALL FILES OR FOLDERS ARE DELETED #
# WITHOUT WARNING. #
#
# The gateway server is only used as a secure door to CUSP RESEARCH #
# COMPUTING FACILITY environments #
# Please go to https://datahub.cusp.nyu.edu to learn more about #
# data transfer (upload/download) to/from your CUSP home directory#
# THANK YOU FOR YOUR COOPERATION #
# CUSP IT email:cusp.it@nyu.edu #
#####
Load key "/Users/fb/.ssh/id_rsa": Invalid Format
fbalanceggw.cusp.nyu.edu's password: []
```

## CUSP PUI - CLASS 1

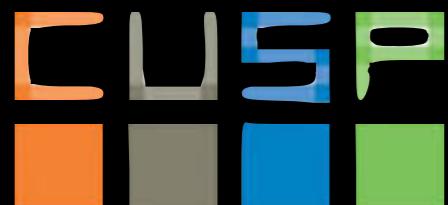
Extension of UCSL bootcamp



<https://slides.com/mohitsharma44/cusp-pui-2018/live#>

- Mohit Sharma

CUSP 2018



II: Reproducibility  
& data types

follow the instructions for the docker

once the following command finishes

**docker-compose up**

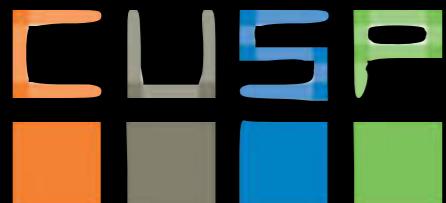
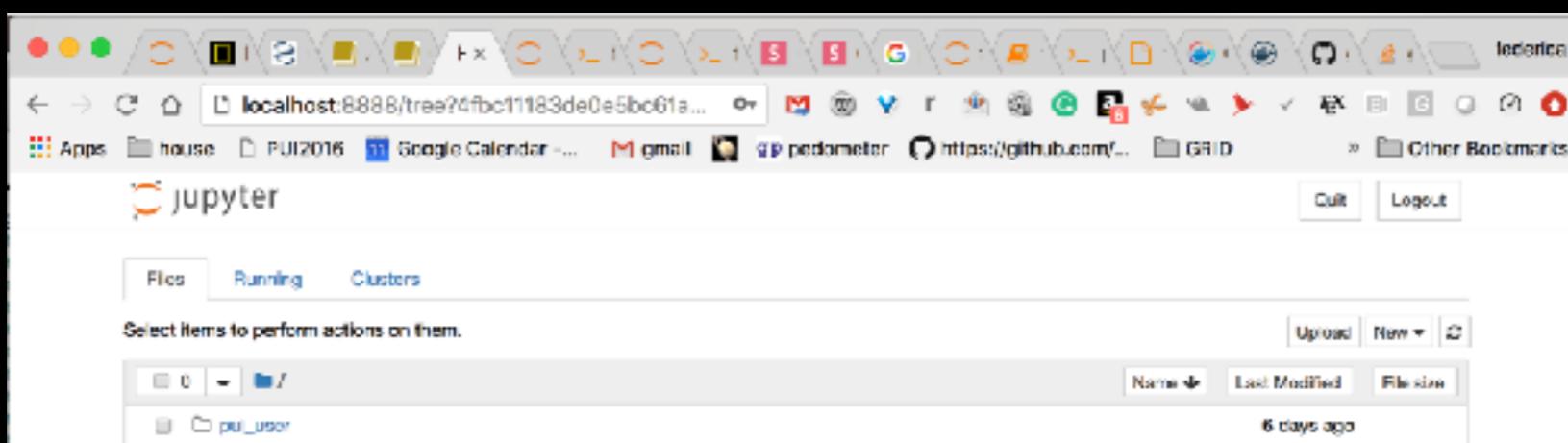
you will see a line on your terminal similar to this  
**(but with a different token)**

```
ucsl-container | Copy/paste this URL into your browser when you connect for the first time,  
ucsl-container | to login with a token: [REDACTED]  
ucsl-container | http://(c99137fcfed81 or 127.0.0.1):8888/?token=4fb...  
[REDACTED]
```

open a browser, on the url bar type

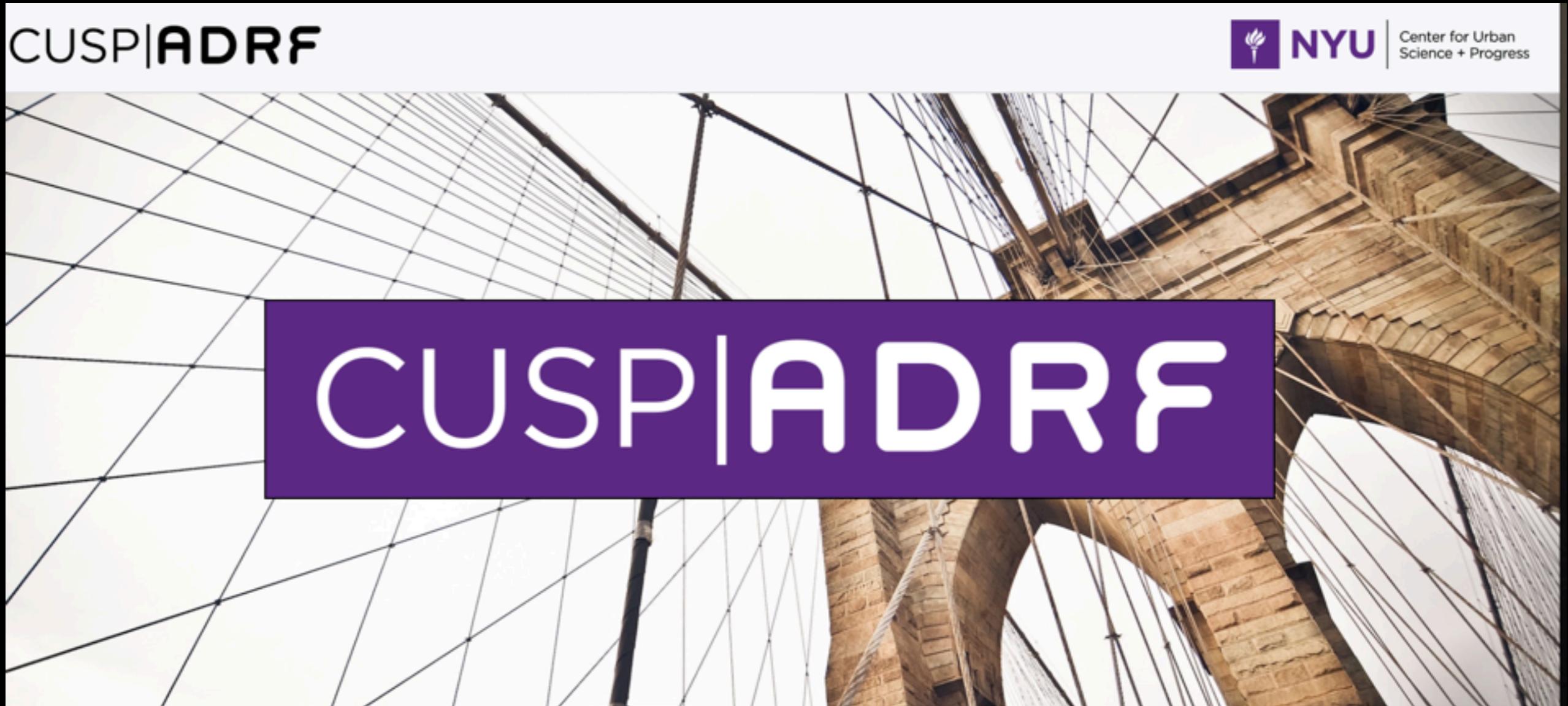
*localhost:*

and then paste the content that you copied there

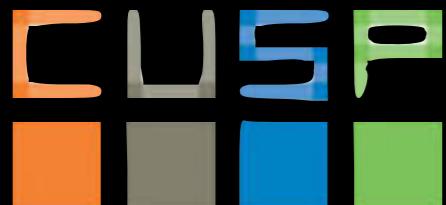


II: Reproducibility  
& data types

# CUSP ADRF with relevant urban data and VEs containing the appropriate setup



<http://cusp.adrf.cloud/documentation>



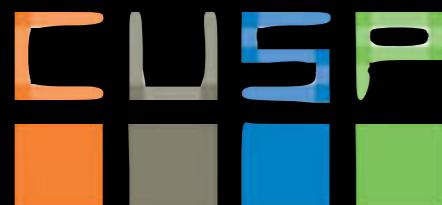
II: Reproducibility  
& data types

# You have a *bash* shell on ADRF

The screenshot shows the CUSP|ADRF web interface. At the top, there is a navigation bar with 'Logout' and 'Control Panel' buttons. Below the navigation bar, there are three tabs: 'Files' (selected), 'Running', and 'Clusters'. A message 'Select items to perform actions on them.' is displayed above a file tree. The file tree shows a root directory '/' containing 'projects', 'pui2018\_fb55', and 'Untitled.ipynb'. On the right side, there is a 'New' button with options for 'Upload', 'New', and a dropdown menu. A red oval highlights the 'Terminal' option in the dropdown menu. Below the interface, the text 'access a terminal on ADRF' is displayed in red, followed by the URL 'http://cusp.adrf.cloud/terminal'.

access a terminal on  
ADRF

<http://cusp.adrf.cloud/terminal>

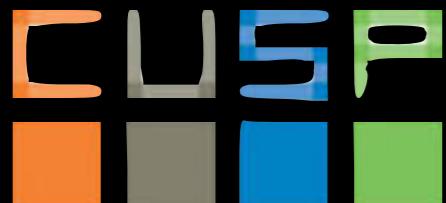


## Operating System:

# Shell commands & Environmental variables demo



<https://speakerdeck.com/62gerente/bash-introduction>



II: Reproducibility  
& data types

# Operating System:

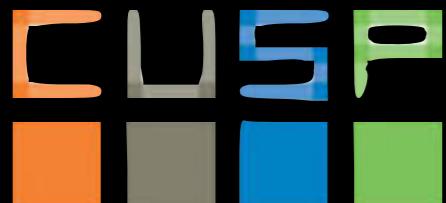
## Shell commands & Environmental

### variables demo

pwd  
ls  
mkdir  
cd  
touch  
cp  
mv  
rm  
less  
echo  
head/tail  
top  
ps  
bg  
chmod  
chown  
grep

ssh  
scp  
rsync  
df  
du  
  
alias  
export  
Wildcards  
**I/O Redirection**  
**Standard Output**  
**Standard Input**  
**Pipes**  
**Listing your processes**  
**Killing a process**

### essential commands



II: Reproducibility  
& data types

# Choosing a text Editor: Integrated or not?

**vi, nano:** basic, no-window “light weight” editors

**Emacs:** “extensible, customizable, self-documenting real-time display editor”

customizable via the .emacs file

can be run without tunneling with `emacs -nw`

the command sequences are tricky (customize emacs for python

<https://realpython.com/blog/python/emacs-the-best-python-editor/>)

**PyCharm, Sublime, Brackets:** Integrated Development

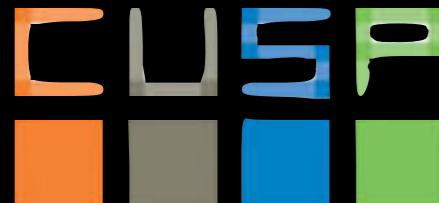
Environment: checks syntax and standard compliancy

can be run directly within the developing window

recognizes the syntax for other codes (Django)

integrated with version control

**Jupyter Notebooks:** Browser-based interactive computational environment



demo time!

II: Reproducibility  
& data types

# Choosing a text Editor: Integrated or not?

**Emacs:**

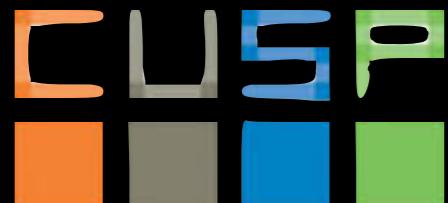
**PyCharm:**

demo time!

⌚ jupyter

**vi, pico, nano:**

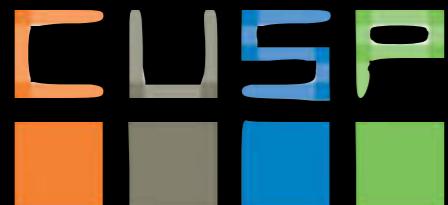
**Jupyter Notebooks:**



II: Reproducibility  
& data types

# Choosing a text Editor: Integrated or not?

nano: follow the instructions!

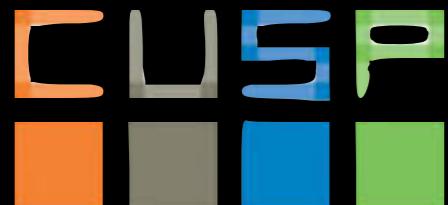


II: Reproducibility  
& data types

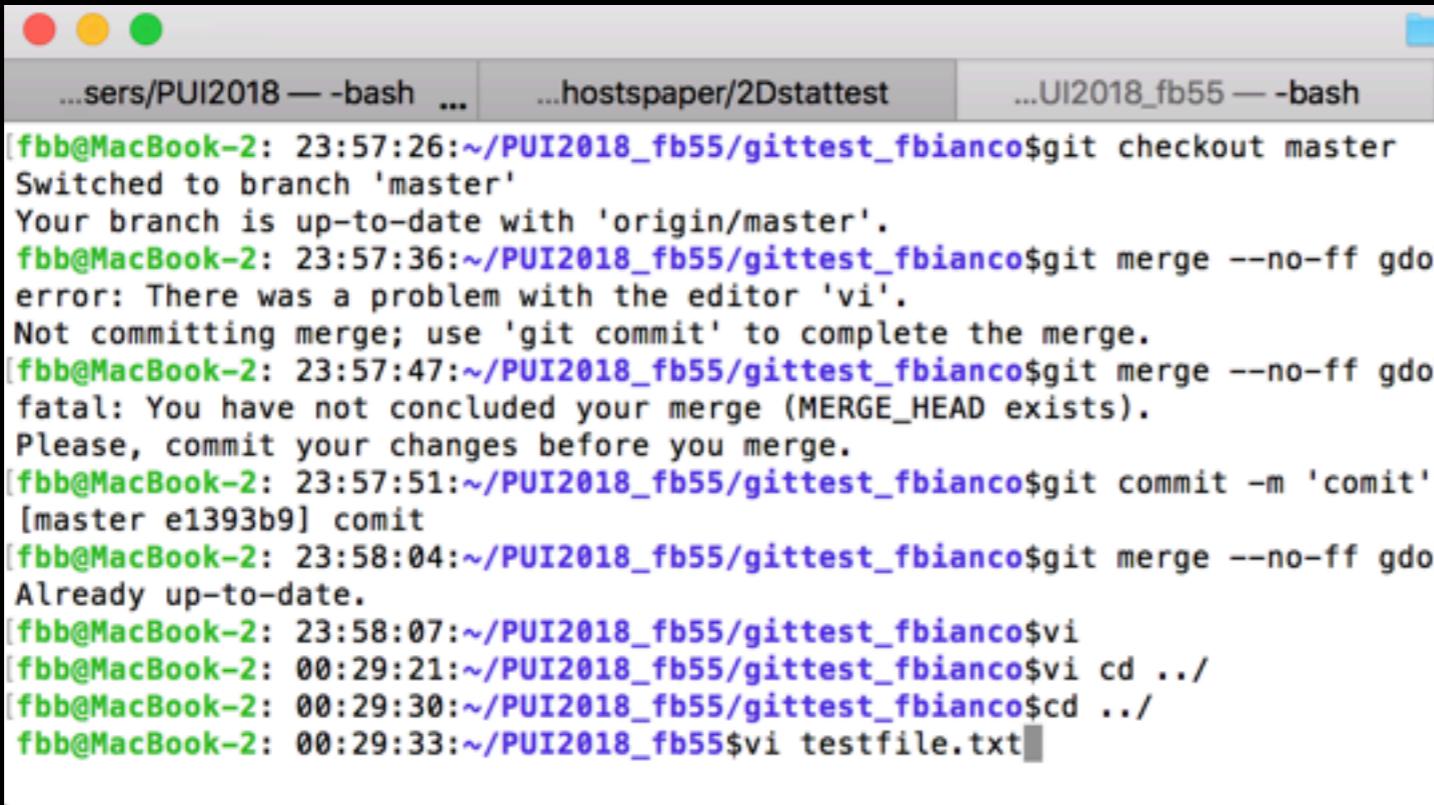
# Choosing a text Editor: Integrated or not?

nano: follow the instructions!

vi: you may need this to edit GitHub messages



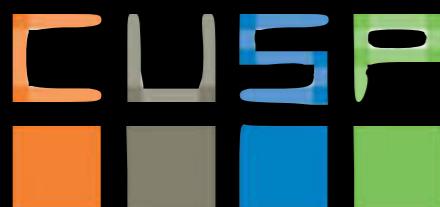
# Choosing a text Editor: Integrated or not?



```
[fbb@MacBook-2: 23:57:26:~/PUI2018_fb55/gittest_fbianco$git checkout master
Switched to branch 'master'
Your branch is up-to-date with 'origin/master'.
[fbb@MacBook-2: 23:57:36:~/PUI2018_fb55/gittest_fbianco$git merge --no-ff gdob
error: There was a problem with the editor 'vi'.
Not committing merge; use 'git commit' to complete the merge.
[fbb@MacBook-2: 23:57:47:~/PUI2018_fb55/gittest_fbianco$git merge --no-ff gdob
fatal: You have not concluded your merge (MERGE_HEAD exists).
Please, commit your changes before you merge.
[fbb@MacBook-2: 23:57:51:~/PUI2018_fb55/gittest_fbianco$git commit -m 'comit'
[master e1393b9] comit
[fbb@MacBook-2: 23:58:04:~/PUI2018_fb55/gittest_fbianco$git merge --no-ff gdob
Already up-to-date.
[fbb@MacBook-2: 23:58:07:~/PUI2018_fb55/gittest_fbianco$vi
[fbb@MacBook-2: 00:29:21:~/PUI2018_fb55/gittest_fbianco$vi cd ...
[fbb@MacBook-2: 00:29:30:~/PUI2018_fb55/gittest_fbianco$cd ...
[fbb@MacBook-2: 00:29:33:~/PUI2018_fb55$vi testfile.txt]
```

tHub messages

type **vi <filename>** on your terminal + hit *Return*

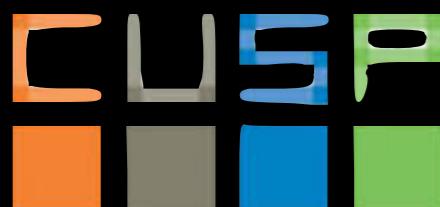


# Choosing a text Editor: Integrated or not?



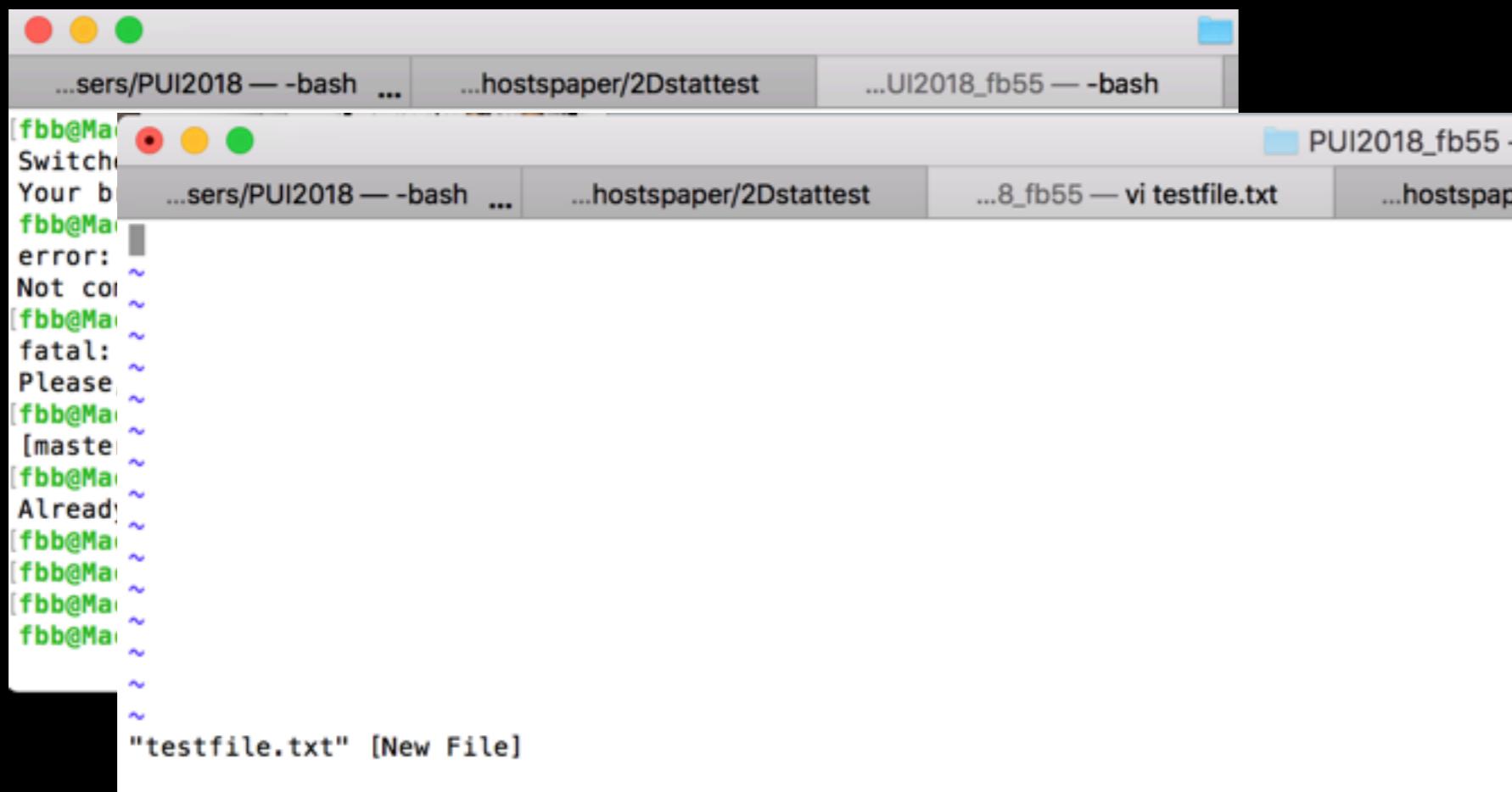
```
[fbb@MacBook-Air ~] Switched to a new terminal window.
Your branch is up-to-date with 'origin/main'.
[fbb@MacBook-Air ~] error:
Not connected.
[fbb@MacBook-Air ~] fatal:
Please try again.
[fbb@MacBook-Air ~] [master]
[fbb@MacBook-Air ~] Already exists.
[fbb@MacBook-Air ~] [fbb@MacBook-Air ~] [fbb@MacBook-Air ~] "testfile.txt" [New File]
```

type *i* for insert and insert text.



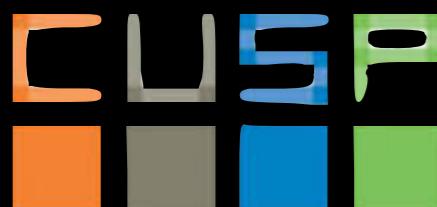
## Choosing a text Editor: Integrated or not?

other commands: *r* for replace one character      *y* to yank a line  
*ESC* + ...      *R* to replace endless characters      *p* to paste below the line  
                  *u* to undo      *P* to paste above the line

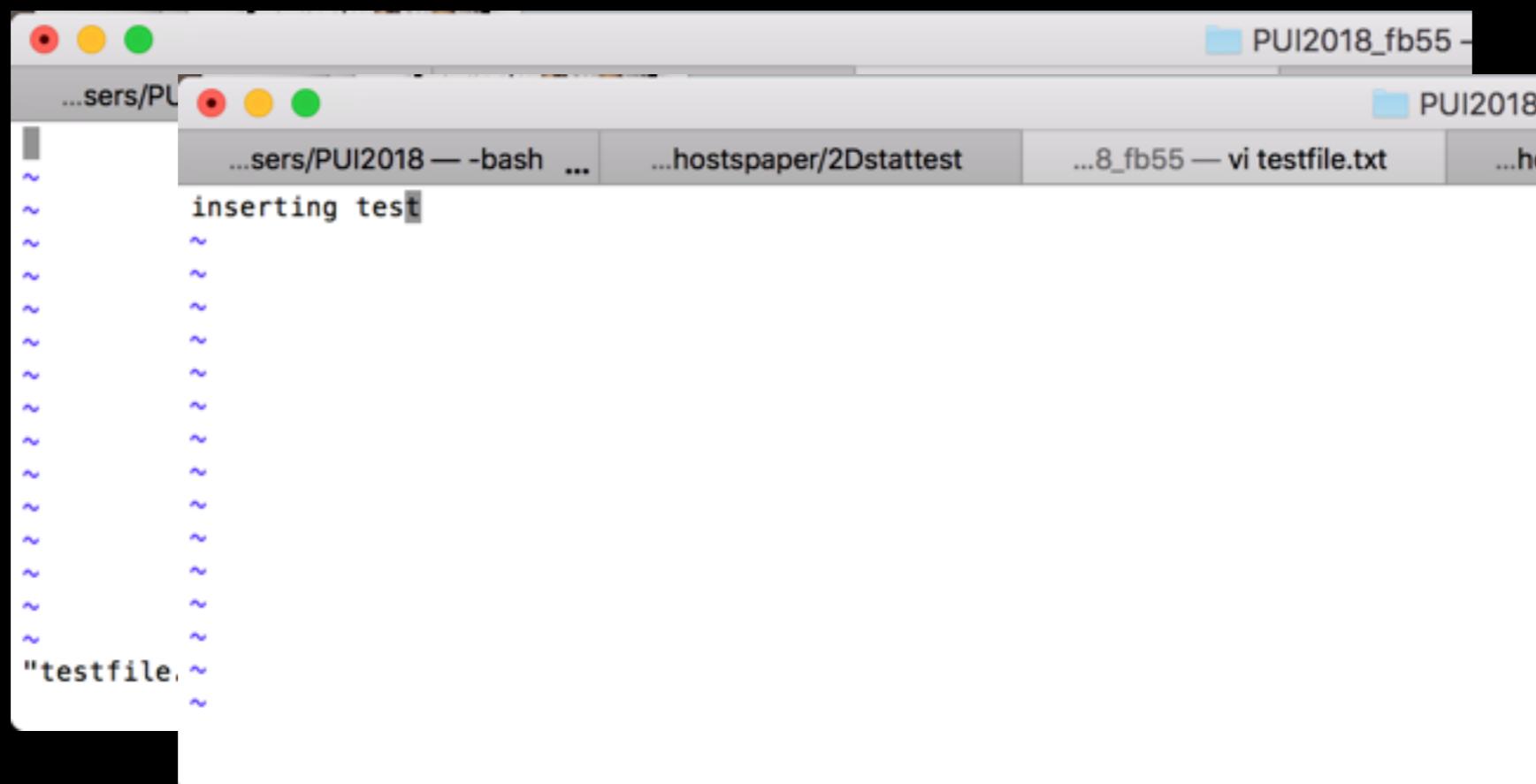


messages

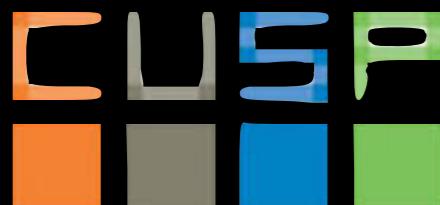
type *i* for insert and insert text.



# Choosing a text Editor: Integrated or not?

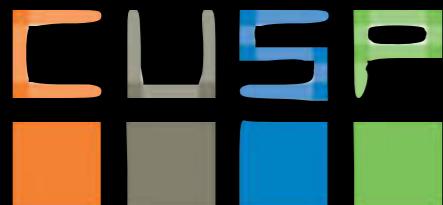
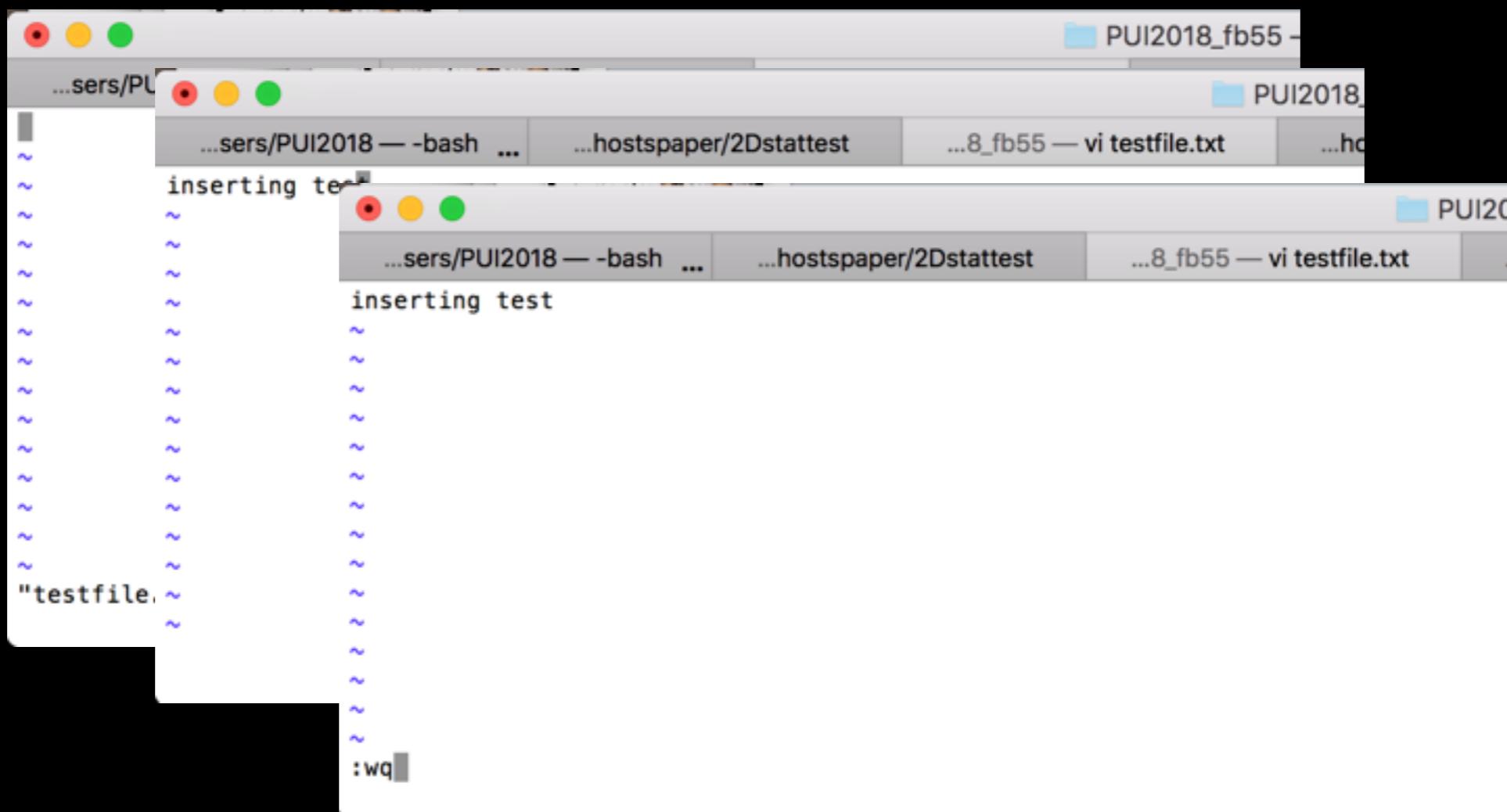


type **i** for insert and insert text.



II: Reproducibility  
& data types

# Choosing a text Editor: Integrated or not?



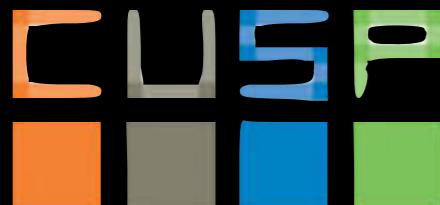
type *ESC* + `:wq` to save and quit what you entered

II: Reproducibility & data types

# Choosing a text Editor: Integrated or not?

```
fbb@MacBook-2: 23:57:36:~/PUI2018_fb55/gittest_fbianco$ git merge --no-ff gdobler-master
error: There was a problem with the editor 'vi'.
Not committing merge; use 'git commit' to complete the merge.
[fbb@MacBook-2: 23:57:47:~/PUI2018_fb55/gittest_fbianco$ git merge --no-ff gdobler-master
fatal: You have not concluded your merge (MERGE_HEAD exists).
Please, commit your changes before you merge.
[fbb@MacBook-2: 23:57:51:~/PUI2018_fb55/gittest_fbianco$ git commit -m 'comit'
[master e1393b9] comit
[fbb@MacBook-2: 23:58:04:~/PUI2018_fb55/gittest_fbianco$ git merge --no-ff gdobler-master
Already up-to-date.
[fbb@MacBook-2: 23:58:07:~/PUI2018_fb55/gittest_fbianco$ vi
[fbb@MacBook-2: 00:29:21:~/PUI2018_fb55/gittest_fbianco$ vi cd ../
[fbb@MacBook-2: 00:29:30:~/PUI2018_fb55/gittest_fbianco$ cd ../
[fbb@MacBook-2: 00:29:33:~/PUI2018_fb55$ vi testfile.txt
[fbb@MacBook-2: 00:30:04:~/PUI2018_fb55$ vi testfile.txt
[fbb@MacBook-2: 00:30:41:~/PUI2018_fb55$
fbb@MacBook-2: 00:30:56:~/PUI2018_fb55$
```

hit Return

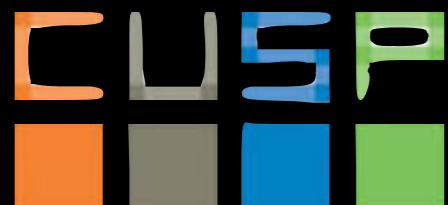


# Choosing a text Editor: Integrated or not?

nano (and pico): follow the instructions!

vi: you may need this to edit GitHub messages

emacs: you may need this to edit GitHub messages



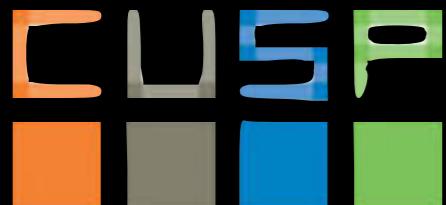
# Choosing a text Editor: Integrated or not?

nano: follow the instructions!

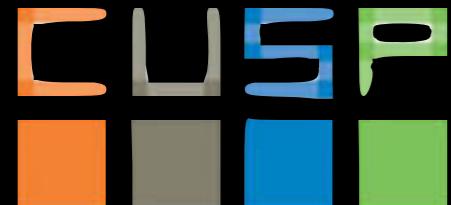
vi: you may need this to edit GitHub messages

emacs: integrated environment

Sublime: integrated environment



# reproducible research

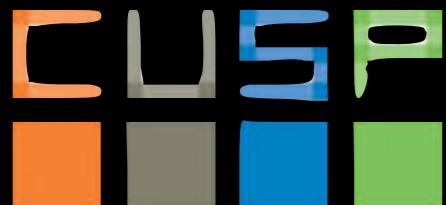


II: Reproducibility  
& data types

## Reproducible research means:

all numbers in a data analysis can be recalculated exactly (down to stochastic variables!) using the **code** and **raw data** provided by the analyst.

Claerbout, J. 1990,  
Active Documents and Reproducible Results, Stanford Exploration Project  
Report, 67, 139



# Reproducible research means:

 PERSPECTIVE | SCIENTIFIC INTEGRITY  
0 What does research reproducibility mean?

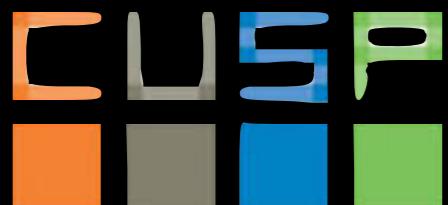
 Steven N. Goodman\*, Daniele Fanelli and John P. A. Ioannidis  
+ See all authors and affiliations

 Science Translational Medicine 01 Jun 2016:  
Vol. 8, Issue 341, pp. 341ps12  
DOI: 10.1126/scitranslmed.aaf5027

Methods reproducibility is meant to capture the original meaning of reproducibility, that is, the ability to implement, as exactly as possible, the experimental and computational procedures, with the same data and tools, to obtain the same results. Results reproducibility refers to what was previously

**assigned reading**

<http://stm.sciencemag.org/content/8/341/341ps12>

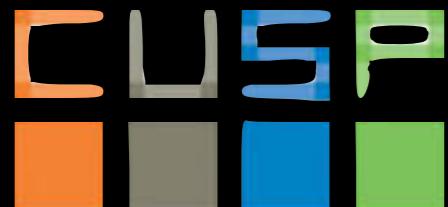


II: Reproducibility  
& data types

## Reproducible research means:

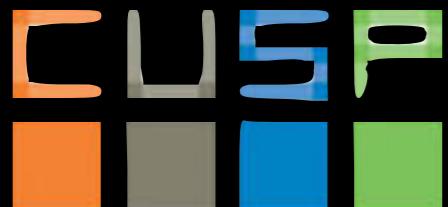
<https://arxiv.org/pdf/1304.2818.pdf>

<http://bayesianblocks.blogspot.com/>



# Reproducible research means:

code      raw data



# Reproducible research means:

code

raw data



<https://github.com/>



## Reproducible research means:

code

raw data



<https://github.com/>

distributed version control system:  
a version of the files on your local computer is  
made also available at a central server.  
The history of the files is saved remotely so  
that any version (that was checked in) is  
retrievable.

Others can access and generate their versions  
of the files enabling collaborative work.

# Reproducible research means:

code

raw data



other version control systems:  
RCS

CVS (Centralized version control system)  
Subversion  
SVN



Git (<https://github.com>)

Mercurial (<https://bitbucket.org/>)

<https://git-scm.com/book/en/v2/Getting-Started-About-Version-Control>

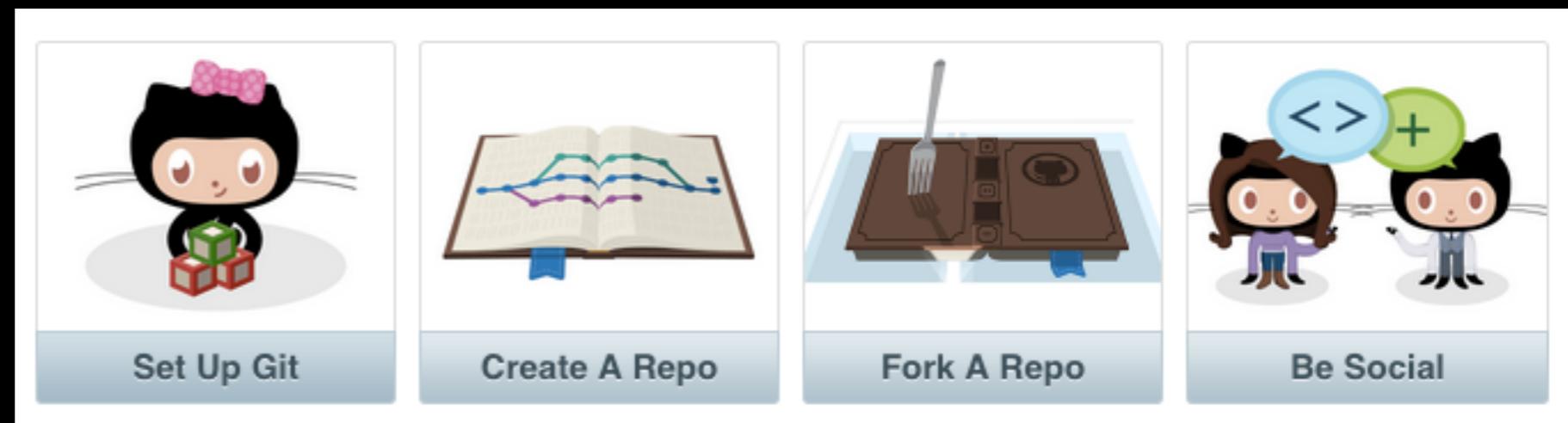
# Reproducible research means:

code

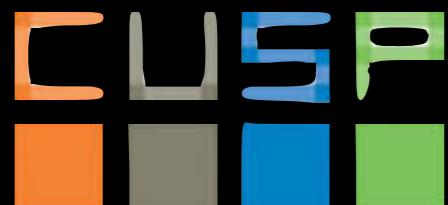
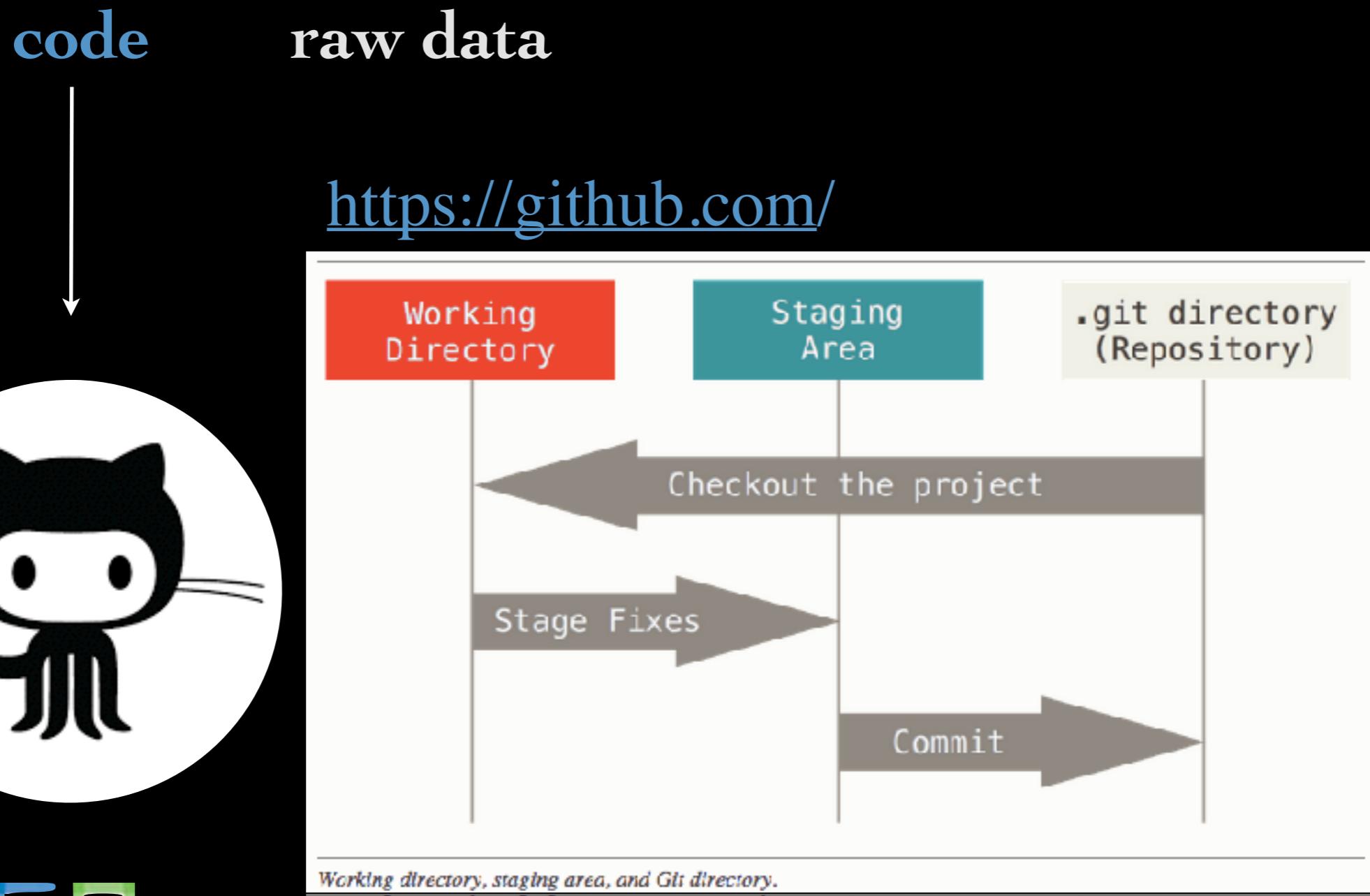
raw data



<https://github.com/>



# Reproducible research means:



# Reproducible research means:

code

raw data



<https://github.com/>

markdowns & standards:

**in order for your research to be  
reproducible it has to be understandable:**

- Paper or slides
- Repository Markdown files
- Understandable (PEP8 compliant) code -  
explicit declare the version of the code!

## Reproducible research means:

code

raw data



<https://github.com/>

markdowns & standards:



# Reproducible research means:

code

raw data



<https://github.com/>

markdowns & standards:

A screenshot of a GitHub repository page. The repository is named 'fedhere / PUI2018\_fb55'. The current branch is 'master'. A file named 'githubCreateRepoCmds.md' is displayed. The content of the file is:

This is a markdown file guiding you through the very first steps to create and manage a git repo with github.

Lets start on your bash shell

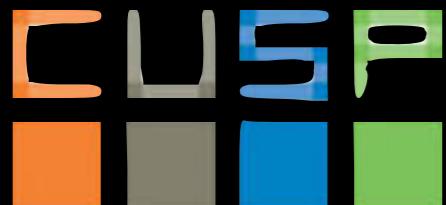
Create a directory

```
finances$ Federicas-MacBook-Air: ~ /science/branex/US_methattract_finance$
```

mkdir gittest\_fobanco

cd gittest\_fobanco

[https://github.com/fedhere/  
PUI2018\\_fb55/blob/master/  
Lab1\\_fb55/  
githubCreateRepoCmds.md](https://github.com/fedhere/PUI2018_fb55/blob/master/Lab1_fb55/githubCreateRepoCmds.md)



II: Reproducibility  
& data types

# Reproducible research means:

code

raw data



<https://github.com/>

markdowns & standards:



PEP8: Python Enhancement Proposals 8

“This document gives coding conventions for the Python code comprising the standard library in the main Python distribution.”

**Readability counts.**

# Reproducible research means:

code

raw data



<https://github.com/>

markdowns & standards:



PEP8: Python Enhancement Proposals 8

Indentation, Tabs or Spaces?, Maximum Line Length, Blank Lines, Source File Encoding, Imports, Whitespace in Expressions and Statements, Comments Bookkeeping, Naming

# Reproducible research means:

code

raw data



<https://github.com/>



a good video tutorial

<https://www.youtube.com/watch?v=ZDR433b0HJY>

## Reproducible research means:

code

raw data



<https://github.com/>

let's make a repo!



# Choosing a text Editor: Integrated or not?

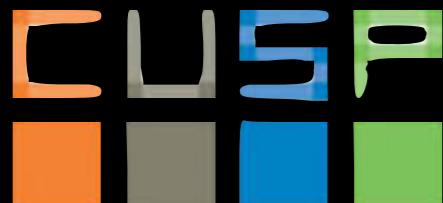
nano (and pico): follow the instructions!

vi: you may need this to edit GitHub messages

emacs: integrated environment

```
conda install -c conda-forge emacs
```

Sublime: integrated environment



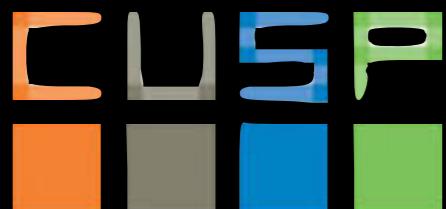
II: Reproducibility  
& data types

# Reproducible research:

## How to share your data

- Share/Reference the source of your raw data
- Share the “tidy” data
- Share the code used to process the data at each step

example time

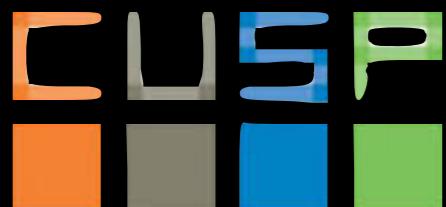


# Reproducible research means:

code      raw data

privacy concerns

in order for your research to be reproducible  
the data you use must be accessible BUT  
THAT IS NOT ALWAYS POSSIBLE:  
CUSP has access to data that has restricted  
access. Share your data when possible!



# Reproducible research means:

code raw data

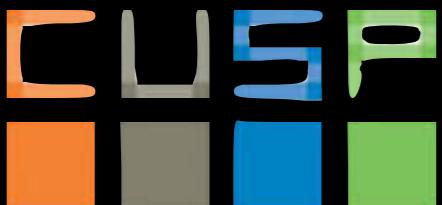
## Remove sensitive data

Some day you or a collaborator may accidentally commit sensitive data, such as a password or SSH key, into a Git repository. Although you can remove the file from the latest commit with `git rm`, the file will still exist in the repository's history. Fortunately, there are other tools that can entirely remove unwanted files from a repository's history. This article will explain how to use two of them: `git filter-branch` and the [BFG Repo-Cleaner](#).

**Danger:** Once you have pushed a commit to GitHub, you should consider any data it contains to be compromised. If you committed a password, change it! If you committed a key, generate a new one.

This article tells you how to make commits with sensitive data unreachable from any branches or tags in your GitHub repository. However, it's important to note that those commits may still be accessible in any clones or forks of your repository, directly via their SHA-1 hashes in cached views on GitHub, and through any pull requests that reference them. You can't do anything about existing clones or forks of your repository, but you can permanently remove all of your repository's cached views and pull requests on GitHub by contacting [GitHub support](#).

<https://help.github.com/articles/remove-sensitive-data/>



II: Reproducibility  
& data types

# Reproducible research means:

code      raw data

**create environmental variables that point to your data directories in your environment.**

They help maintaining security as you do not have to reveal the full path to your data or code to use it  
They facilitate collaboration cause two people can work on a project on their machine, refer to the same variable, and have it associated to a different path, unique to the user.

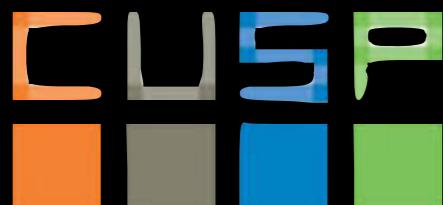
e.g. \$PUIdata on my computer that is /User/fbianco/PUIdata.

\$PUIdata on your computer points to /User/<your\_user\_name>/PUIdata

**create an env variable on the fly with**

```
export PUIdata=<path_to_pui_data_folder>
```

**save the environmental variable in your .bashrc or .bash\_profile file (in your home directory) to have them stored permanently:** open the file and add the line above!

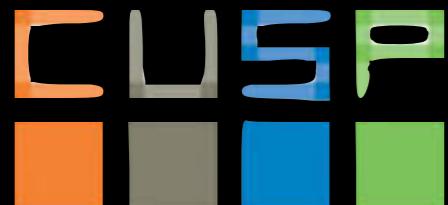


# HOMEWORK

[https://github.com/fedhere/PUI2018\\_fb55/tree/  
master/HW2\\_fb55](https://github.com/fedhere/PUI2018_fb55/tree/master/HW2_fb55)

**HWI\_1 - use GitHub**

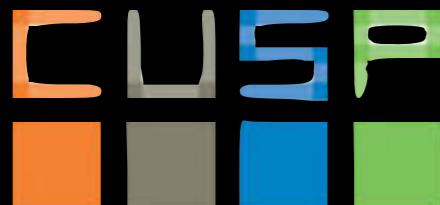
**HWI\_2 - reproducible research (if you did not do it yet)**



II: Reproducibility  
& data types

## Key Concepts:

- falsifiability and law of parsimony
- types of scientific questions
- reproducible research
- PEP8 and style standards
  
- work with github
- understand how to set up your environment
- basic bash commands
- creating and checking into github an ipython notebook



# Resources:

about reproducible research:

<http://stm.sciencemag.org/content/8/341/341ps12> ASSIGNED READING

<https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>

<http://www.pnas.org/content/early/2018/03/08/1802324115>

<https://www.nature.com/news/reality-check-on-reproducibility-1.19961>

