

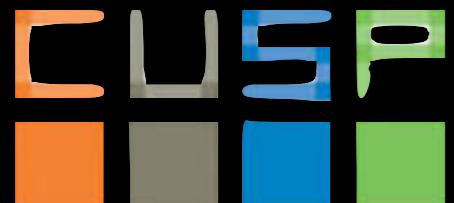
Urban Informatics

Fall 2018

dr. federica bianco fbianco@nyu.edu



@fedhere



Urban Informatics

Dr. federica bianco fbianco@nyu.edu

Office hours: W12-2+M4-530 F10-12 (check)

Office: CUSP1228, NYU Physics 938

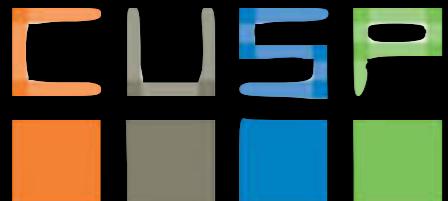
TAs:

Ilyas Habeeb (eve) mih278@nyu.edu

office hours: Tu5-630/F4-530

Fu Shang (mor) fs1520@nyu.edu

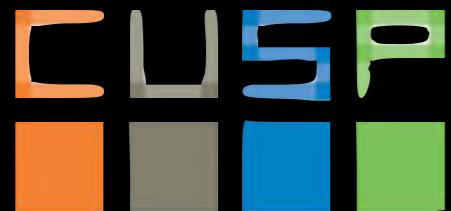
office hours: Tu10-12F3-4

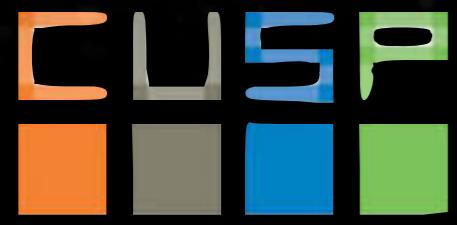


Urban Informatics

Dr. federica bianco fbianco@nyu.edu,
astrophysicist

[http://blogs.teradata.com/international/
sciences-loss-gain-data-science/](http://blogs.teradata.com/international/sciences-loss-gain-data-science/)



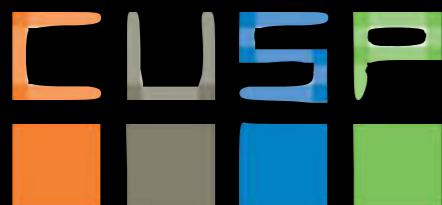


Urban Observatory

Data Science Institute - Inaugural event

what is data science? we have been using data in science the whole time, but with the volume, rate, and complexity of the current data we have to worry about things that we would neglect until now: what happens if our data has errors, what happens if we have missing data?

Lue Rossi, Mathematical Sciences Chairperson, UD
(astrophysicists have always worried about that)

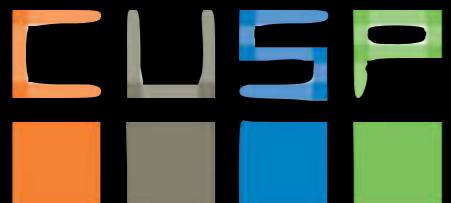


Understand Your Data

Urban Informatics

Class website:

serv.cusp.nyu.edu/~fbianco/PUI2018

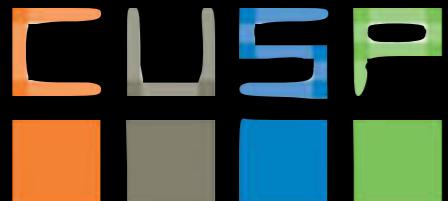


Urban Informatics

Class: 3 hours, lecture + lab

Grade

- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final



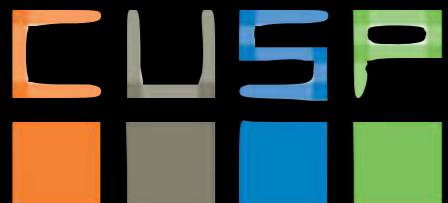
Urban Informatics

Class: 3 hours, lecture + lab

Grade

- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final

*from beginning of class to 5 minutes past the hour (be on time!)
questions on previous class material AND READING ASSIGNMENTS*



Urban Informatics

Class: 3 hours, lecture + lab

Grade

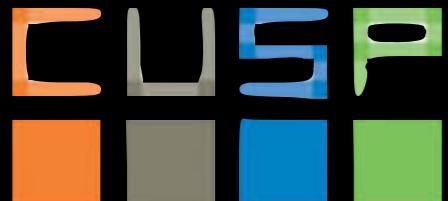
- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final

ask questions

answer questions

get up and code

extra credit assignments



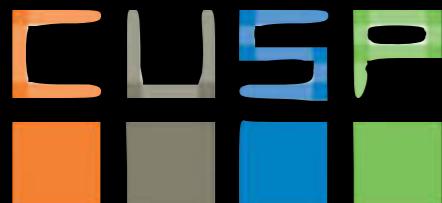
Urban Informatics

Class: 3 hours, lecture + lab

Grade

- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final

Homework projects must be turned in as iPython notebooks by checking them into your github account in the PUI2018_<netID> repo and the project directories HW<hw number>_<netID> (unless otherwise stated). <nyuid> is e.g. fb55



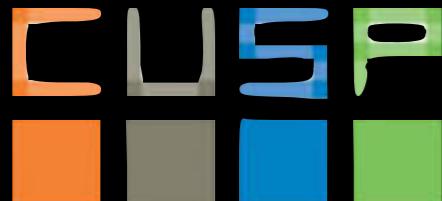
Urban Informatics

Class: 3 hours, lecture + lab

Grade

- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final

I encourage you to work in groups! but as a collaborative project where different group members lead different aspects of the work.
A statement to describing your contribution to the project MUST be included in the README (a la Nature Magazine).



Light echoes reveal an unexpectedly cool η Carinae during its nineteenth-century Great Eruption

A. Rest, J. L. Prieto, N. R. Walborn, N. Smith, F. B. Bianco, R. Chornock, D. L. Welch, D. A. Howell, M. E. Huber, R. J. Foley, W. Fong, B. Sinnott, H. E. Bond, R. C. Smith, I. Toledo, D. Minniti & K. Mandel

Affiliations Contributions

Nature 482, 375–378 (16 Fe
Received 26 August 2011 /
Brief Communication Arisin

Contributions

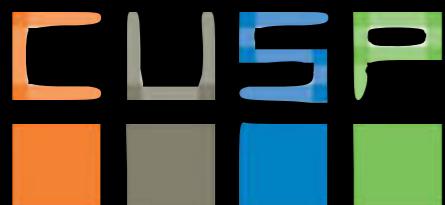
All authors contributed to the drafting of the paper. A.R., N.S. and R.C.S. imaged the area around η Car. A.R. and M.E.H. reduced the imaging data. H.E.B. provided images of the echoes that guided our spectroscopic pointings. J.L.P., R.C., R.J.F. and W.F. obtained the spectra and reduced them. A.R. and J.L.P. performed spectral analysis and interpretation. A.R., N.R.W. and F.B.B. performed spectral classification. F.B.B. and K.M. correlated the spectra. A.R., D.L.W. and B.S. modelled the light echo. I.T. and D.M. provided imaging of η Car. F.B.B. and D.A.H. provided the FTS images, and F.B.B. and A.R. reduced them.

Class: 3 hours, lecture + lab

Grade

- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final

I encourage you to work in groups! but as a collaborative project where different group members lead different aspects of the work.
A statement to describing your contribution to the project MUST be included in the README (a la Nature Magazine).



Example of a README.md for a PUI homework: missing the README.md costs you 10% of the grade!

The README.md is a MarkDown (md) file. The syntax of a MarkDown is rather simple: <https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>. Also the MD syntax can be used in Jupyter notebook cells to include text (not code) that is automatically formatted (which you will need to do over and over...)

CitiBike HW - v1

Question

Are CitiBike's easing commuter journeys across the East River?

Hypothesis

- H0: The probability of a citibike subscriber crossing the East River in a given month is independent of whether the trip is taken during rush hour
- H1: The probability of a citibike subscriber crossing the East River in a given month is not independent of whether the trip is taken during rush hour

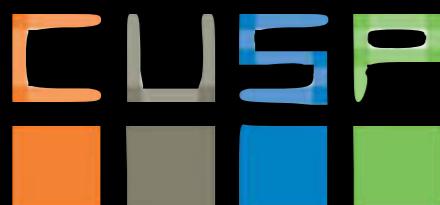
Project work balance

hypothesis generation

Max, Arno, Clayton discussed and equally shared hypothesis generation. Max had the original idea of looking at bridges as he is an avid CitiBike user

Tasks

1. Clayton is tagging trips as cross east river or not
2. Max is defining historic hours as "on peak" or "not on peak"
3. Arno completes a chi-square test of our hypothesis



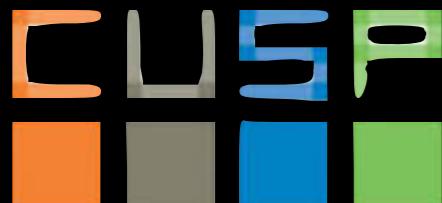
Urban Informatics

Class: lecture + lab

Grade

- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final

After the midterm projects we might have code reviewed by your peers.
We'll have 1 multi-week homework project from proposal to peer review.
<https://blog.fogcreek.com/increase-defect-detection-with-our-code-review-checklist-example/>

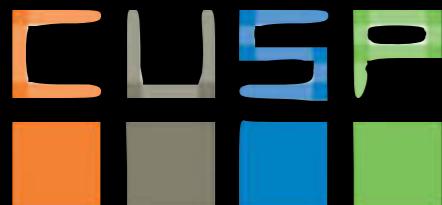


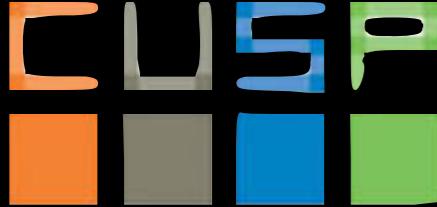
Class: lecture + lab
Grade

- 5% on pre-class question
- 10 % class performance and participation
- 25 % homework
- 25 % midterm
- 35 % final

Midterm and Final will include aspects of the work developed in the homework sessions.

Failing to actively participate in the homework will result in not being able to get the Midterm and Final done.





Class: Grade

- 5%
- 10
- 25
- 25
- 35

This repository Search Pull requests Issues Gist

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Wiki Pulse Graphs Settings

No description or website provided. — Edit

61 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

fedhere committed on GitHub Update README.md Latest commit a183019 2 minutes ago

HW1_fb55 Update README.md 21 hours ago

Lab1_fb55 Delete github_create_repo_cmds.md 21 hours ago

PEP8MinimalRequirements.md Update PEP8MinimalRequirements.md an hour ago

README.md Update README.md 2 minutes ago

PUI2016_fb55

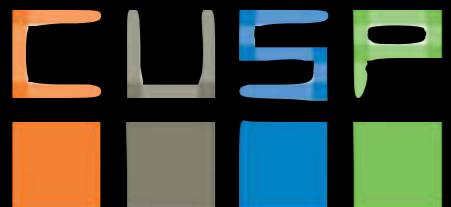
This repository contains the assignments for NYU CUSP Principles of Urban Informatics 2016. Check here for the new assignments, and for the solutions to be posted.

GRADING GUIDELINES

- Each HW must be turned in as a directory in PUI2016_<netID>.
- The directory HW<hw_number>_<netID> must have a README.md which states the student's participation. No penalty if the student declares not to have had any contribution but to have just followed and learned. However missing the README.md, missing the statement about who the student worked with and what they did, or inconsistencies between the statements of students within the group that cannot be easily reconciled by asking will cost them 10% of the grade.
- Each assignment turned in as a notebook must have rendered plots with axis labels and captions. Each missing/non rendered plot, or plot without axes labels or caption will cost 10% of the grade.
- The notebook must be executables: the TA must download the notebook and run it cell by cell without errors. If

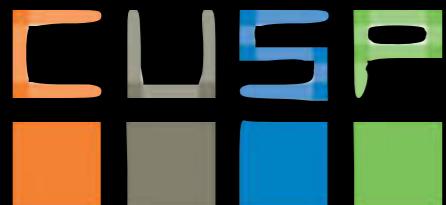
Urban Informatics

GOALS



The workflow of a data driven project

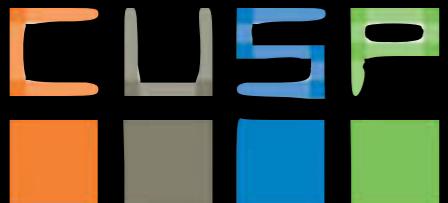
- IDEA
- dataset
 - define ideal data
 - figure out best data available
 - figure out if you can get new data
 - obtain data (including policy issues + technical issues)
- data handling
 - joining databases
 - formatting data
- exploratory data analysis
 - machine learning (clustering? dimensionality reduction?)
- statistics
 - models (regression)
 - prediction
 - validation (simulations)
- interpretation
- presentation
 - visualization
 - write a paper!



I: Good scientific practice
& work flow

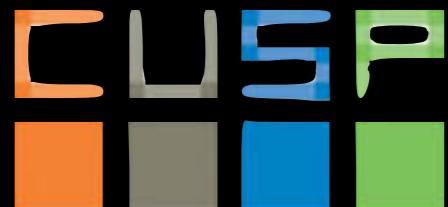
The workflow of a data driven project

- IDEA
- dataset
 - define ideal data
 - figure out best data available
 - figure out if you can get new data
 - obtain data (including policy issues + technical issues)
- data handling
 - joining databases
 - formatting data
- exploratory data analysis
 - machine learning (clustering? dimensionality reduction?)
- statistics
 - models (regression)
 - prediction
 - validation (simulations)
- interpretation
- presentation
 - visualization
 - write a paper or give a talk



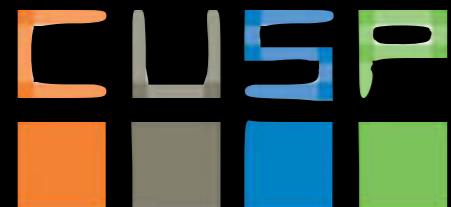
I: Good scientific practice
& work flow

The philosophical side of things



I: Good scientific practice
& work flow

what is a scientific theory?

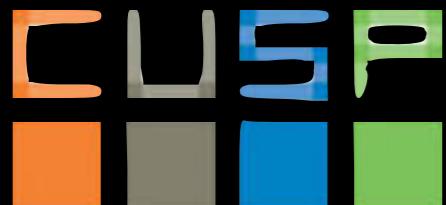


I: Good scientific practice
& work flow

The Demarcation Problem: a scientific theory must be *falsifiable*

My proposal is based upon an *asymmetry* between verifiability and falsifiability; an asymmetry which results from the logical form of universal statements. For these are never derivable from singular statements, but can be contradicted by singular statements.

— Karl Popper, *The Logic of Scientific Discovery*



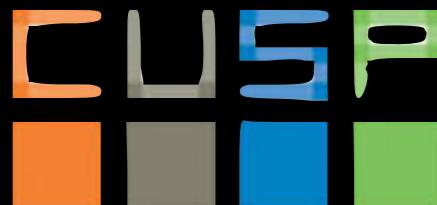
The Demarcation Problem: a scientific theory must be *falsifiable*

My proposal is based upon an *asymmetry* between verifiability and falsifiability; an asymmetry which results from the logical form of universal statements. For these are never derivable from singular statements, but can be contradicted by singular statements.

— Karl Popper, *The Logic of Scientific Discovery*

things can get more complicated though:

most scientific theories are actually based largely on *probabilistic induction* and modern *inductive inference* (Solomonoff, frequentist vs Bayesian methods...)



Ockham's razor: *Pluralitas non est ponenda sine necessitate*

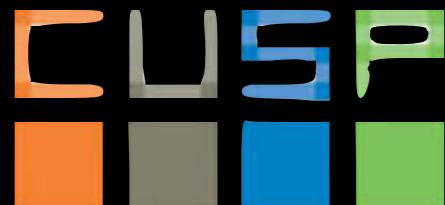
or “the law of parsimony”

William of Ockham (logician and Franciscan friar) 1300ca

but probably to be attributed to John Duns Scotus (1265–1308)

“Complexity needs not to be postulated without a need for it”

“Between 2 theories choose the simpler one”



I: Good scientific practice
& work flow

the earth is round, and it orbits around the sun



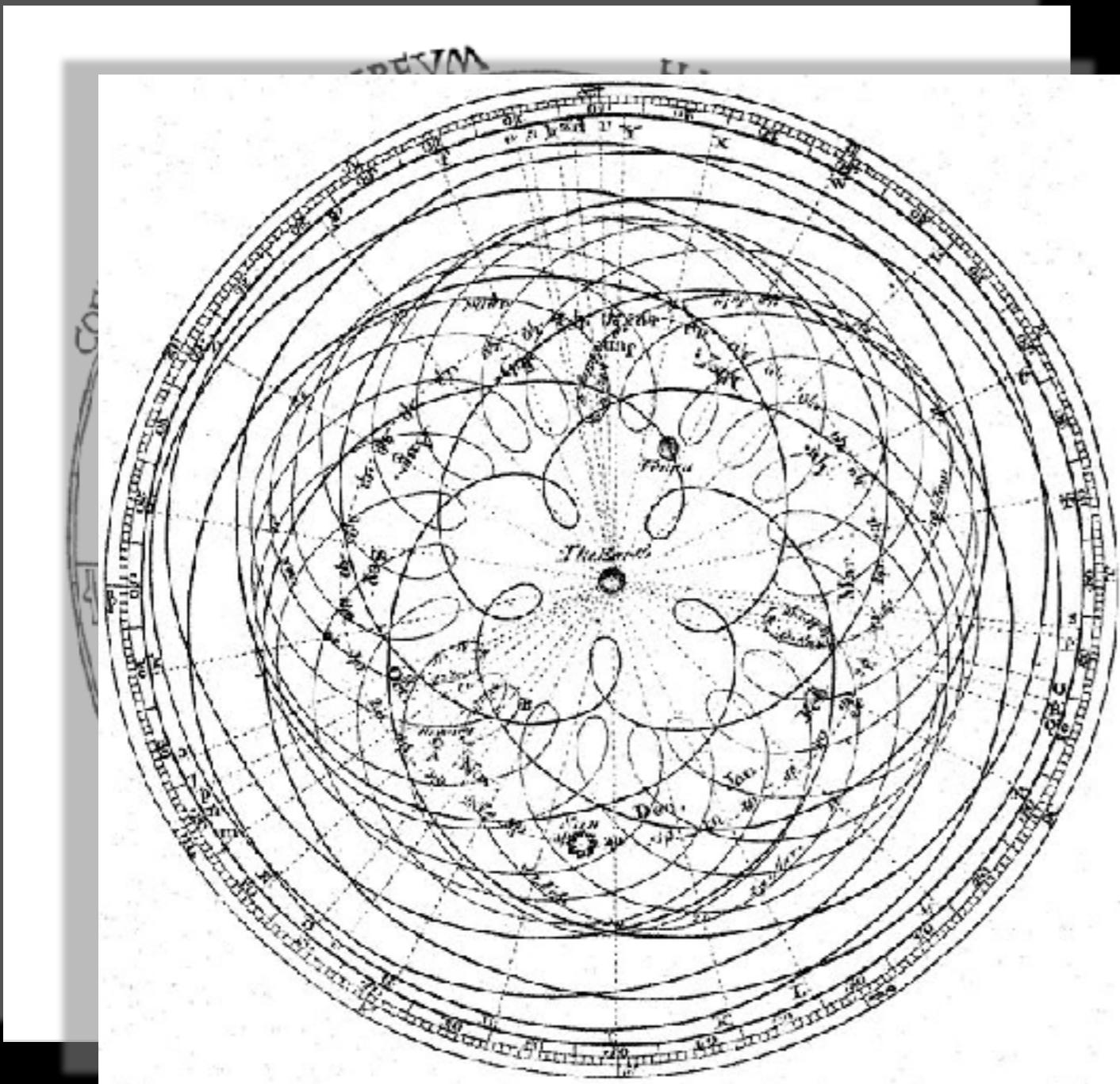
<http://en.wikipedia.org/wiki/File:Ptolemaicsystem-small.png>

Peter Apian, *Cosmographia*, Antwerp, 1524

from Edward Grant, "Celestial Orbs in the Latin Middle Ages", *Isis*, Vol. 78, No. 2. (Jun., 1987).

Geocentric models are natural:
from our perspective
we see the Sun
moving, while we stay
still

the earth is round, and it orbits around the sun



Source Encyclopaedia Britannica 1st Edition

Author Dr Long's copy of Cassini, 1777

As observations
improve
this model cannot fit
the data anymore!
not *easily* anyways...

the earth is round, and it orbits around the sun

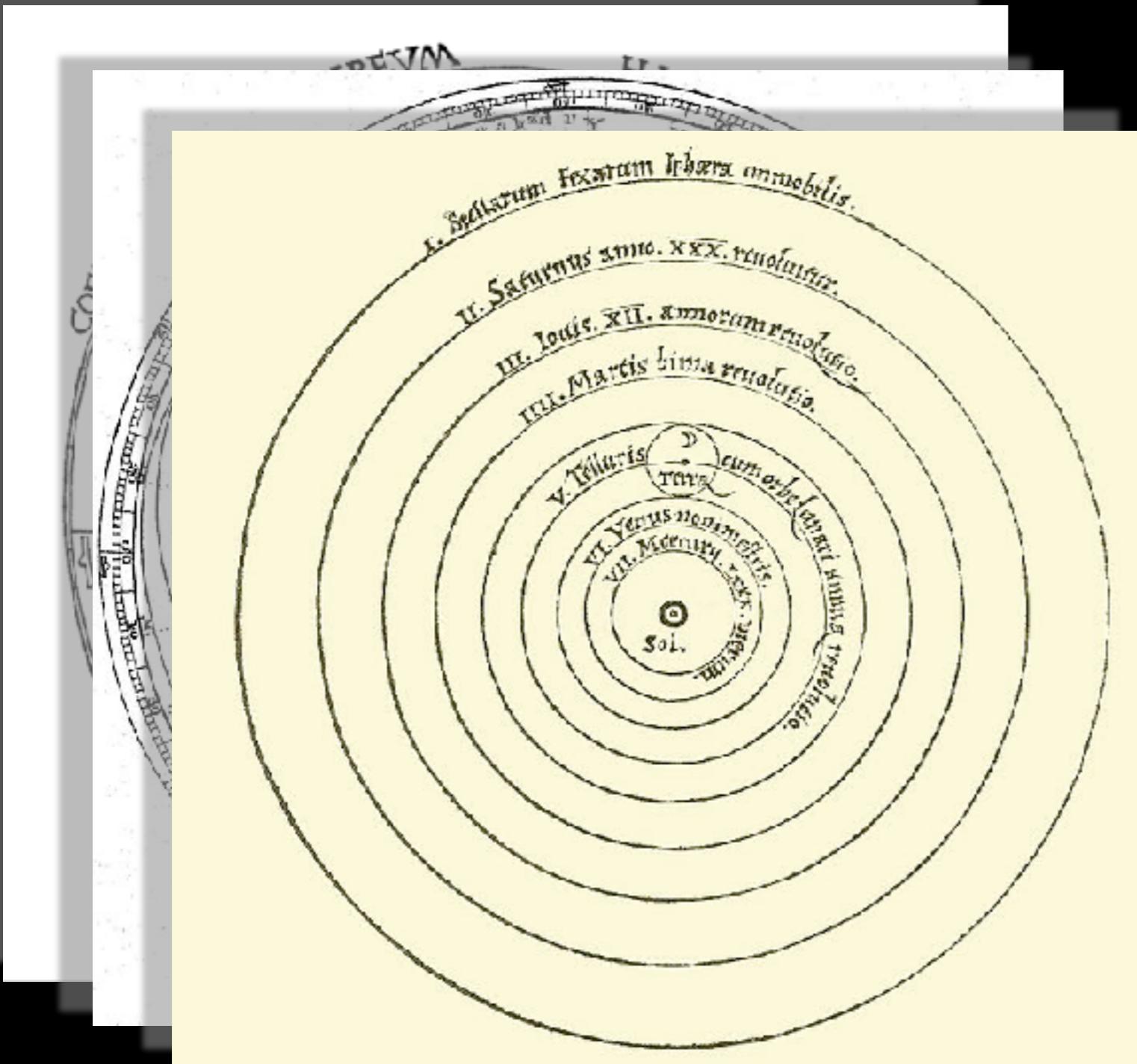


Image of heliocentric model from Nicolaus Copernicus' "De revolutionibus orbium coelestium".

A new model that is much simpler fit the data just as well (perhaps though only until better data comes...)

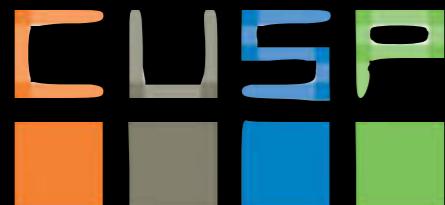
Ockham's razor: *Pluralitas non est ponenda sine necessitate* or the law of parsimony

William of Ockham (logician and Franciscan friar) 1300ca

but probably to be attributed to John Duns Scotus (1265–1308)

“Complexity needs not to be postulated without a need for it”

“Between 2 theories choose the simpler one”

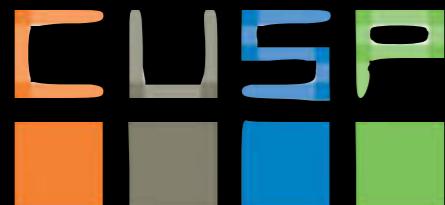


Ockham's razor: *Pluralitas non est ponenda sine necessitate* or the law of parsimony

William of Ockham (logician and Franciscan friar) 1300ca
but probably to be attributed to John Duns Scotus (1265–1308)

“Complexity needs not to be postulated without a need for it”

“Between 2 theories choose the one with fewer parameters!”



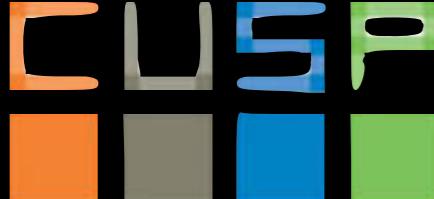
What is the question?

the data speaks, if you know how to listen...

Leek&Rodgers 2015 in Science

<http://www.sciencemag.org/content/347/6228/1314.full.pdf>

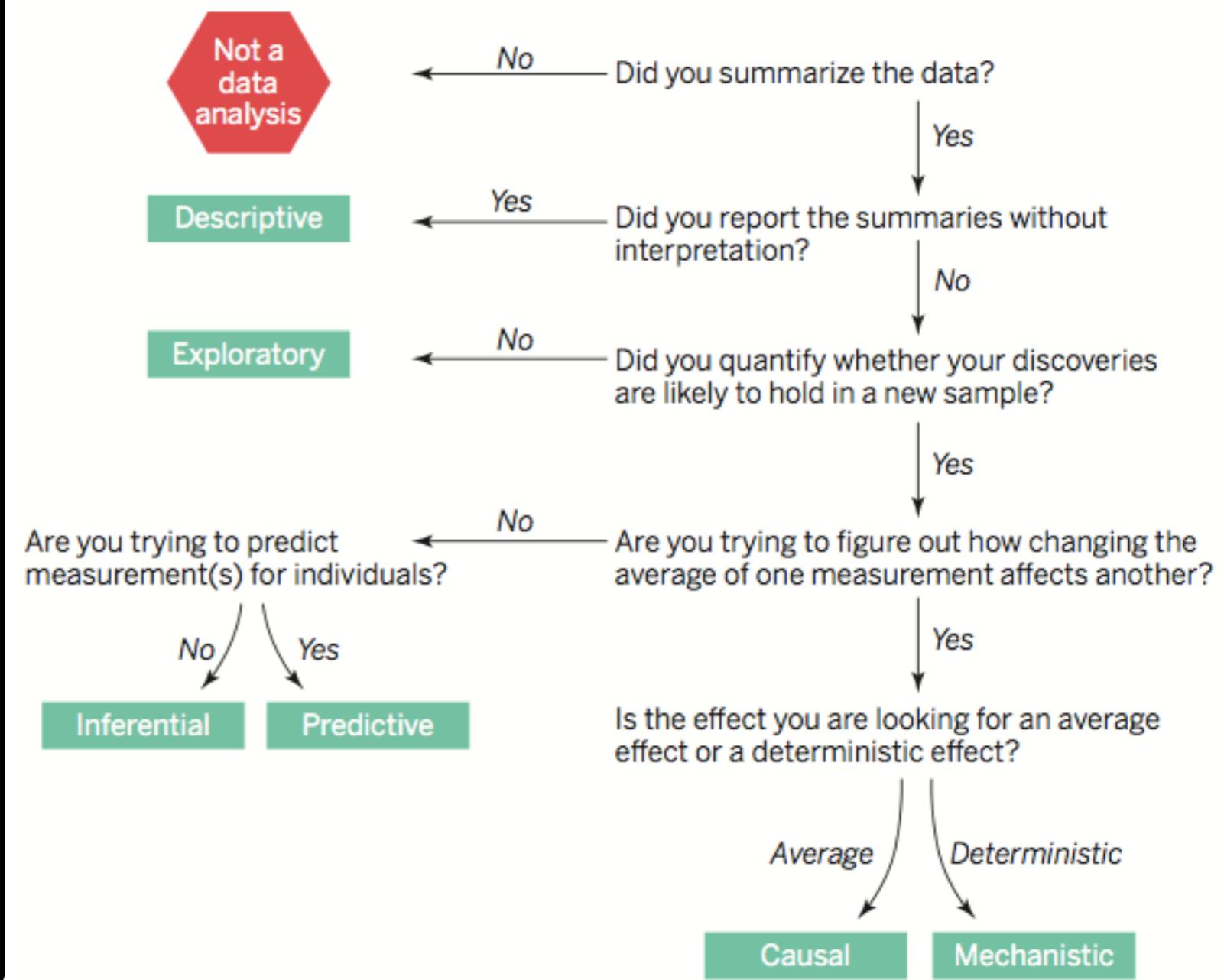
<http://moscow.sci-hub.bz/4d3cf57483ccf211f66cad18440023cd/10.1126%40science.aaa6146.pdf>



I: Good scientific practice
& work flow

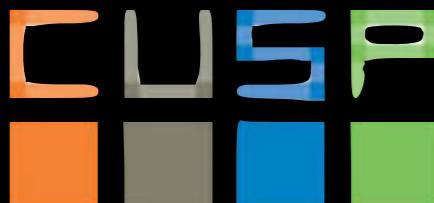
What is the question?

Data analysis flowchart



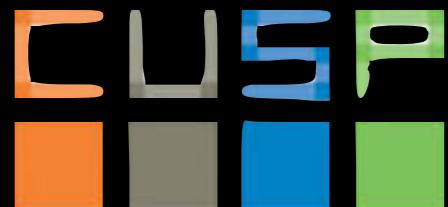
Leek&Rodgers 2015 in Science <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4530023/>
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4530023/>

<http://moscow.sci-hub.bz/4d3cf57483ccf211f66cad18440023cd/10.1126/science.aaa6146.pdf>



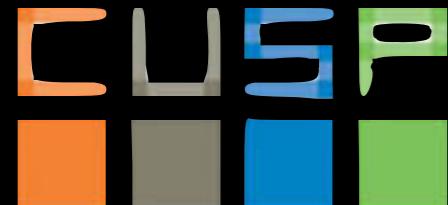
I: Good scientific practice
& work flow

The practical side of things

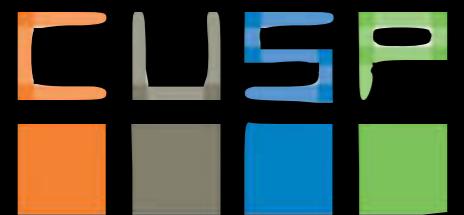


I: Good scientific practice
& work flow

workflow: your environment

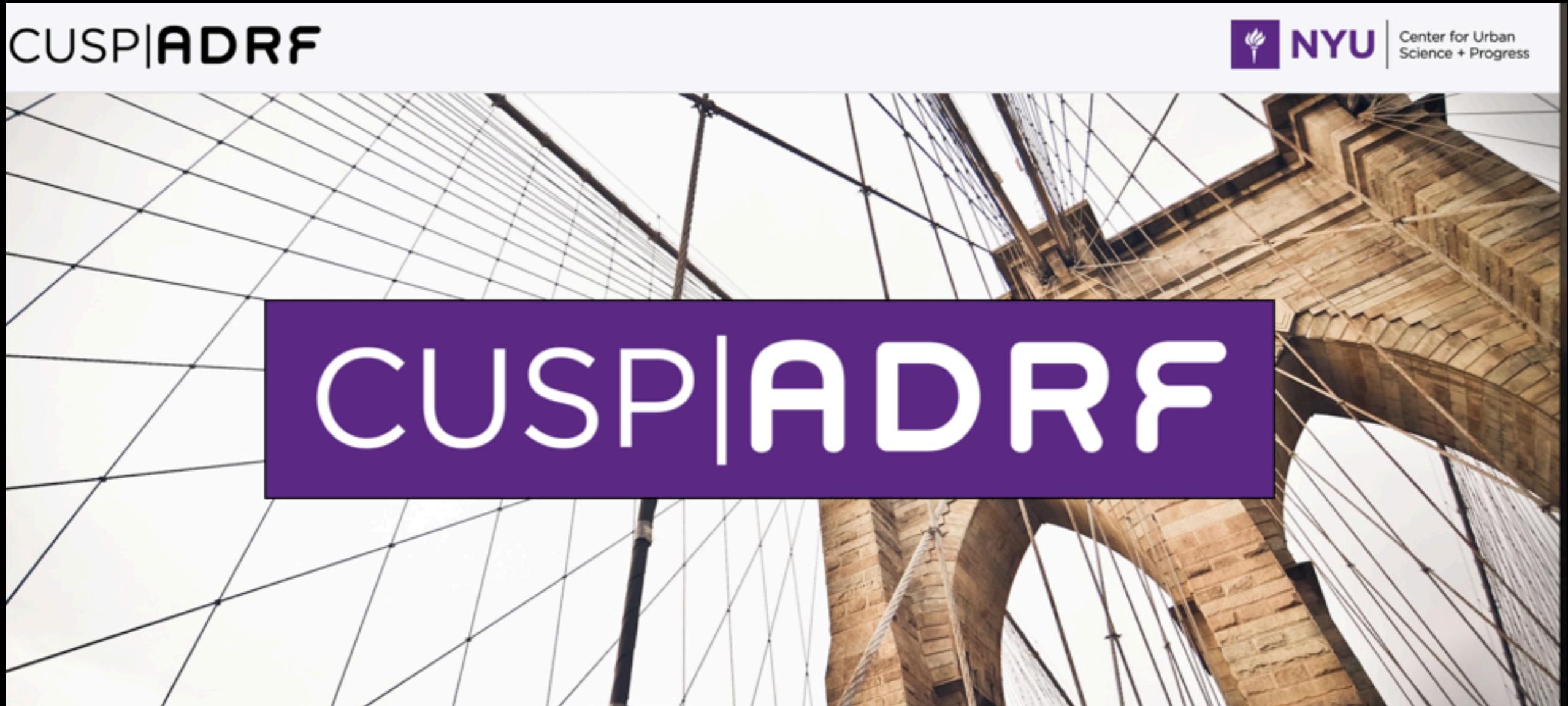


I: Good scientific practice
& work flow

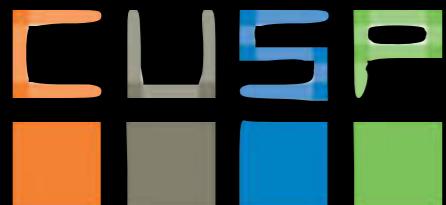


I: Good scientific practice
& work flow

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



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



I: Good scientific practice
& work flow

other solution and environment details...

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

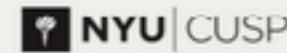


bash:

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

CUSP PUI - CLASS 1

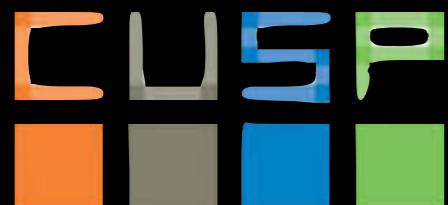
Extension of UCSL bootcamp



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

- Mohit Sharma

CUSP 2018



I: Good scientific practice
& work flow

follow the instructions for the docker

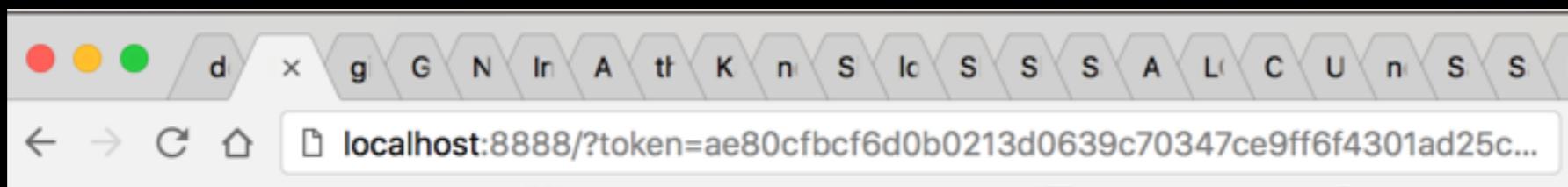
once the following command finishes
docker-compose up

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

open a browser, on the url bar type

localhost:

and then paste the content that you copied there



getting to Code:

Python, iPython, iPython notebooks

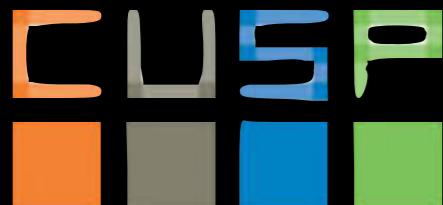
You should be fluent in *at least Python or R*

to be competitive on the job market

In this class we will only work in Python.

All homework should be developed in

python and delivered through github.



I: Good scientific practice
& work flow

getting to Code: Python, iPython, iPython notebooks

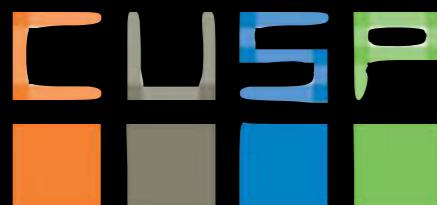
Python 2.7 vs Python 3.0

I will write in Python2.7 for compatibility (e.g. GeoPandas)

I will use the future package

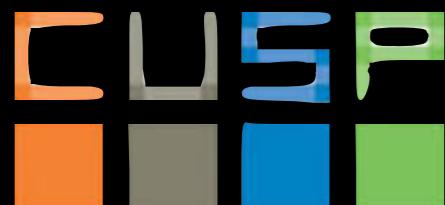
```
from __future__ import print_function  
__author__ = "Federica B. Bianco, CUSP NYU 2016"
```

to make the code forward compatible with Python 3.0
(though some lines of code may be broken in Python 3.0)



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 this menu. At the bottom of the interface, there is a URL bar with the address 'https://jupyterhub.cusp.adrf.cloud/user/fb55/tree?#'. Overlaid on the bottom left of the interface is the text 'access a terminal on ADRF' in red and the URL 'http://cusp.adrf.cloud/terminal' in blue.



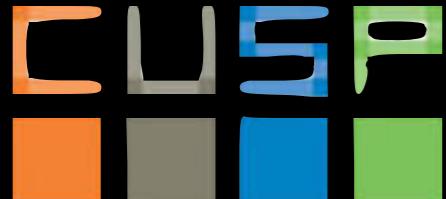
I: Good scientific practice
& work flow

Operating System:

Shell commands & Environmental variables demo



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



I: Good scientific practice
& work flow

Operating System:

Shell commands & Environmental

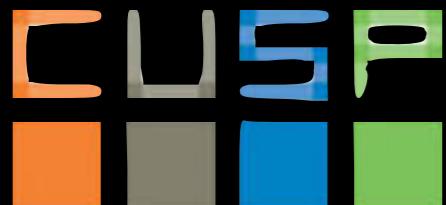
variables demo

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



I: Good scientific practice
& work flow

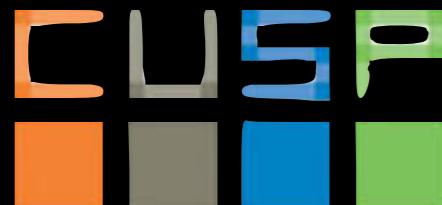
Choosing a text Editor: Integrated or not?

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

vim, vi, pico, nano: other common editors

Jupyter Notebooks: Browser-based interactive computational



environment
demo time!

I: Good scientific practice
& work flow

Choosing a text Editor: Integrated or not?

Emacs:

PyCharm:

demo time!

⌚ jupyter

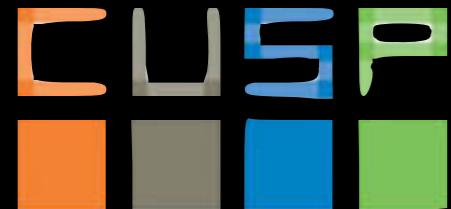
vim, vi, pico, nano: other common editors

Jupyter Notebooks:



I: Good scientific practice
& work flow

reproducible research

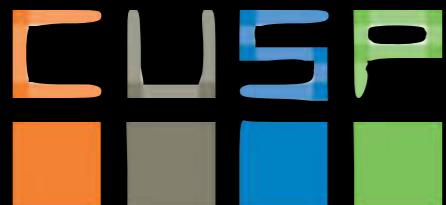


I: Good scientific practice
& work flow

Reproducible research means:

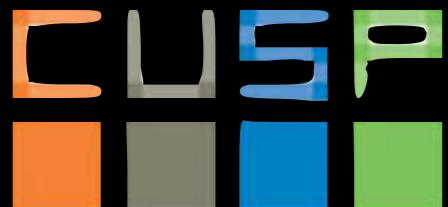
all numbers in a data analysis can be recalculated exactly (down to stochastic variables!) using the **code** and **raw data** provided by the analyst.

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



Reproducible research means:

code raw data



I: Good scientific practice
& work flow

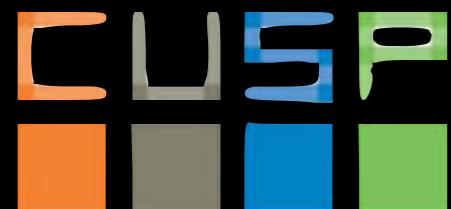
Reproducible research means:

code

raw data



<https://github.com/>



I: Good scientific practice
& work flow

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



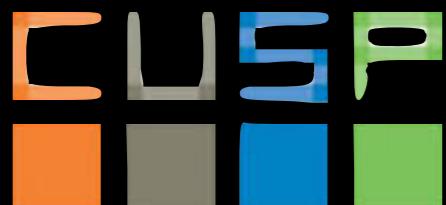
other version control systems:
RCS

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>



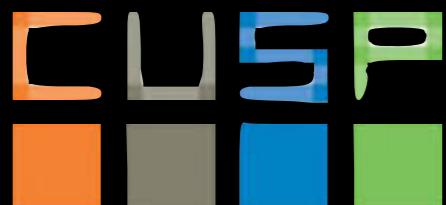
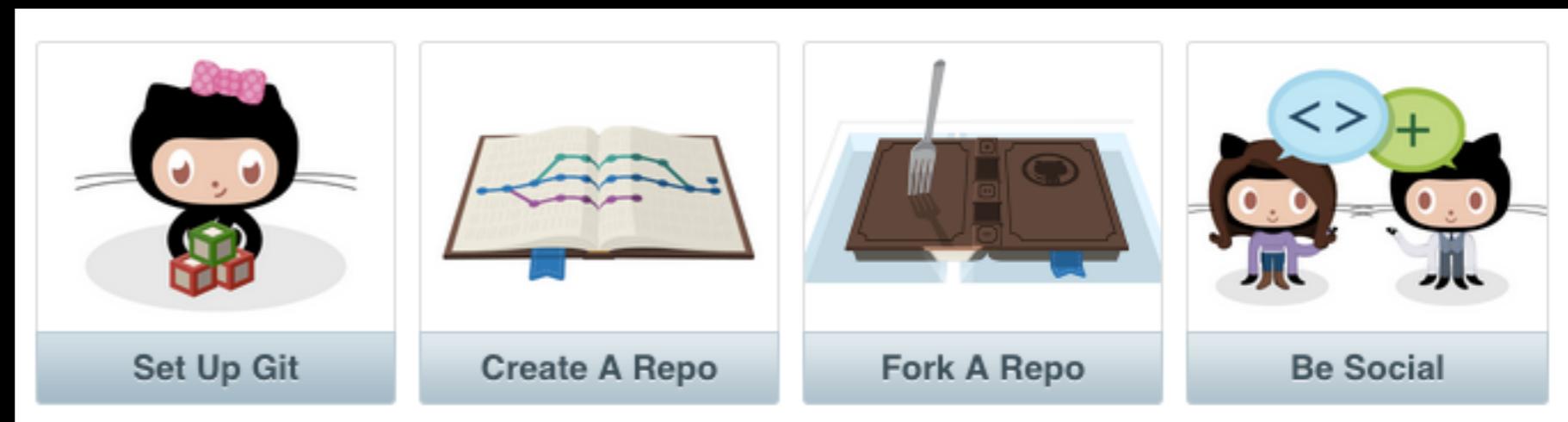
Reproducible research means:

code

raw data



<https://github.com/>



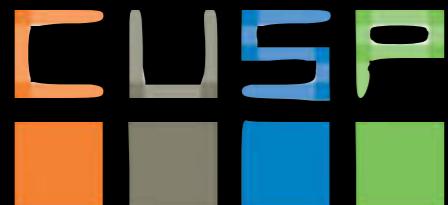
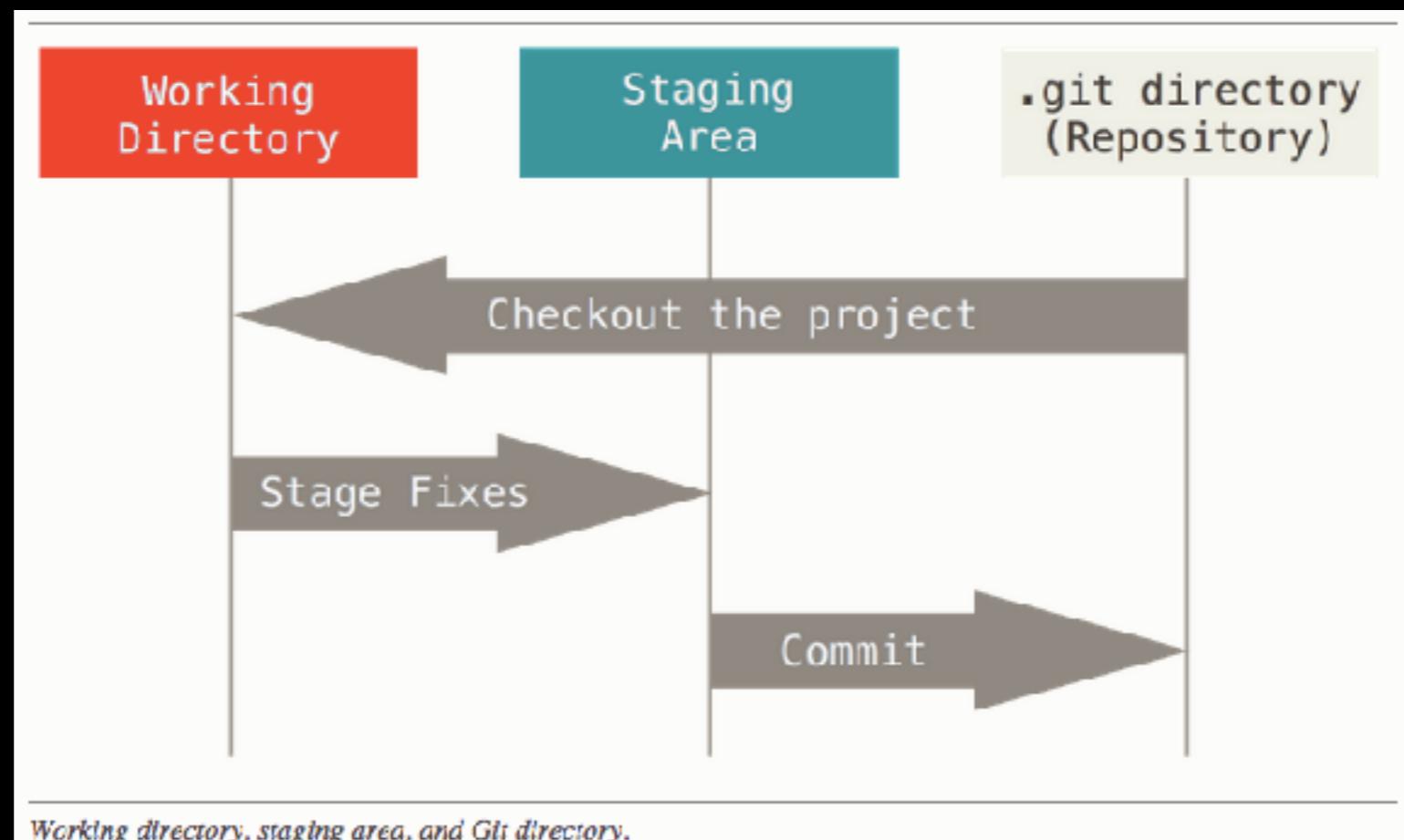
I: Good scientific practice
& work flow

Reproducible research means:

code raw data



<https://github.com/>



I: Good scientific practice
& work flow

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/>

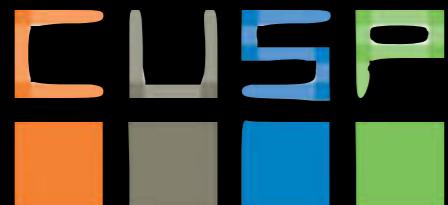
markdowns & standards:



This is 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 as follows:

```
This is a markdown file guiding you through the very first steps to create and manage a git repo with github.  
Let's start on your bash shell  
Create a directory  
$ cd ~/Desktop  
$ git init  
$ touch test.html  
$ git add test.html  
$ git commit -m "Initial commit"  
$ git remote add origin https://github.com/fedhere/PUI2018_fb55.git  
$ git push -u origin master
```

https://github.com/fedhere/PUI2018_fb55/blob/master/Lab1_fb55/githubCreateRepoCmds.md



: Good scientific practice
& work flow

Reproducible research means:

code

raw data



<https://github.com/>

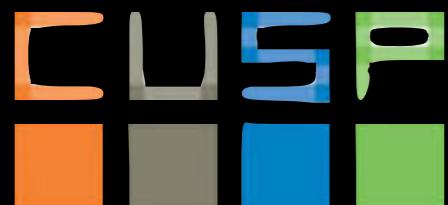
markdowns & standards:



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.



I: Good scientific practice
& work flow

Reproducible research means:

code

raw data



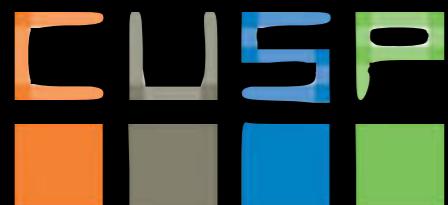
<https://github.com/>

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



I: Good scientific practice
& work flow

Reproducible research means:

code

raw data



<https://github.com/>



a good video tutorial

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

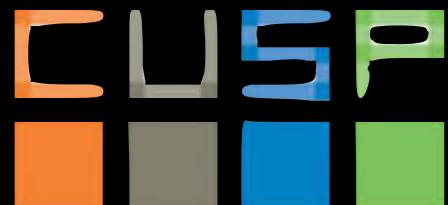
Reproducible research means:

code raw data

↓

<https://github.com/>

let's make a repo!

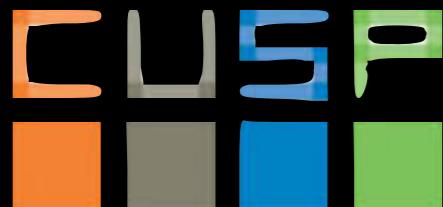


Reproducible research means:

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

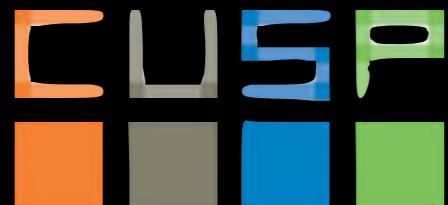
Remove sensitive data

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/>



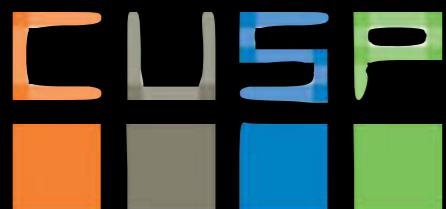
I: Good scientific practice
& work flow

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



I: Good scientific practice
& work flow

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.

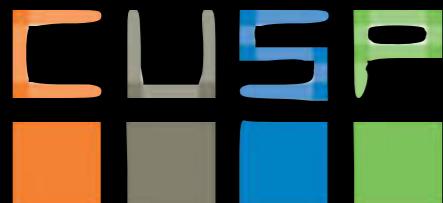
e.g. \$PUIdata on my computer that is /User/fbianco/PUIdata.

\$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!



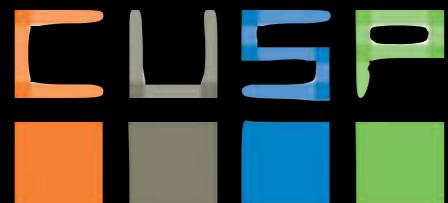
HOMEWORK

https://github.com/fedhere/PUI2018_fb55/tree/master/HW1_fb55

HWI_1 - use GitHub

HWI_2 - set up environment (env variables and aliases)

HWI_3 - reproducible coding (extra credit)

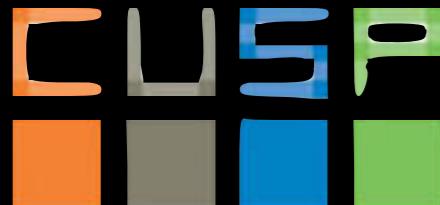


I: Good scientific practice
& work flow

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:

Karl Popper, J. 1934,

The Logic of Scientific Discovery

<http://strangebeautiful.com/other-texts/popper-logic-scientific-discovery.pdf>

Jeff Leek & Rodger Peng. 2015,

What is the Question? ASSIGNED READING

<http://www.sciencemag.org/content/347/6228/1314.summary>

<https://www.d.umn.edu/~kgilbert/ened5560-1/The%20Research%20Question-2015-Leek-1314-5.pdf>

Claerbout, J. 1990,

**Active Documents and Reproducible Results,
Stanford Exploration Project Report, 67, 139**

http://sepwww.stanford.edu/data/media/public/docs/sep67/jon2/paper_html/

Jeff Leek, 2015

The Elements of Data Analytic Style

<https://leanpub.com/dastyle> (\$10.00) and <https://github.com/jtleek/datasharing>

Guido van Rossum, Barry Warsaw, Nick Coghlan, 2001

Proposal Enhancement for Python

<https://www.python.org/dev/peps/pep-0008/>

