

GOOD PRACTICES IN THE LAB

RESEARCH DOCUMENTATION AND ELECTRONIC NOTEBOOKS

Arnaud Legrand



Neurocampus Open Science Workshop
October 2023



WHAT'S EXPECTED FROM SCIENCE?



Goal

- **Describe** and **explain** **Nature** (human organizations and artificial objects as well)

WHAT'S EXPECTED FROM SCIENCE?



Goal

- **Describe** and **explain** **Nature** (human organizations and artificial objects as well)
- Find **solutions** to **society** needs and issues
 - Feed technological development and **industry**: health, energy, communications, ...
 - Enlighten public decisions: educate citizens, politics, regulation, ...

WHAT'S EXPECTED FROM SCIENCE?



Goal

- **Describe** and **explain Nature** (human organizations and artificial objects as well)
- Find **solutions** to **society** needs and issues
 - Feed technological development and **industry**: health, energy, communications, ...
 - Enlighten public decisions: educate citizens, politics, regulation, ...

What else distinguishes science from other human activities?

WHAT'S EXPECTED FROM SCIENCE?



Goal

- **Describe** and **explain** **Nature** (human organizations and artificial objects as well)
- Find **solutions** to **society** needs and issues
 - Feed technological development and **industry**: health, energy, communications, ...
 - Enlighten public decisions: educate citizens, politics, regulation, ...

What else distinguishes science from other human activities?

Method

SCIENTIFIC CONSENSUS



NO TRANSPARENCY NO CONSENSUS



NO TRANSPARENCY NO CONSENSUS



Enlightening the society requires moral/methodological/technical warranties

MERTON'S FOUR NORMS OF SCIENCE (1940)

Universalism Scientific **validity** is independent of the sociopolitical status/personal attributes of its participants (origin, gender, sexuality, religion, etc.)

Communism/communalism All scientists should have common ownership of scientific goods (intellectual property), to promote collective collaboration; **secrecy is the opposite of this norm**

- We say “Newton’s law” to remember that Newton made the original discovery, but not because he has any property of this law

Disinterestedness Scientific institutions act for the **benefit of a common scientific enterprise**, rather than for the personal gain of individuals

Organized skepticism Scientific claims should be exposed to **critical scrutiny** (including **reproduction**) before being accepted

MERTON'S FOUR NORMS OF SCIENCE (1940)

Universalism Scientific **validity** is independent of the sociopolitical status/personal attributes of its participants (origin, gender, sexuality, religion, etc.)

Communism/communalism All scientists should have common ownership of scientific goods (intellectual property), to promote collective collaboration; **secrecy is the opposite of this norm**

- We say “Newton’s law” to remember that Newton made the original discovery, but not because he has any property of this law

Disinterestedness Scientific institutions act for the **benefit of a common scientific enterprise**, rather than for the personal gain of individuals

Organized skepticism Scientific claims should be exposed to **critical scrutiny** (including **reproduction**) before being accepted

Merton was not saying “*you should follow these norms*”, but rather
“*this is the behavior belief I am observing in the scientific community*”

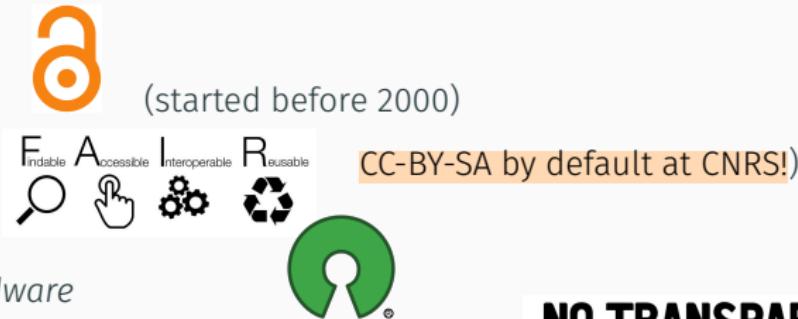
TRANSPARENT RESEARCH ~ OPEN SCIENCE ?

Plan National pour la Science Ouverte (BSN ↔ CoSO)

- CNRS, Inria, INRAE, ...
- Many flavors: *Citizen Science*

Main pillars:

1. Open access
2. Open data
3. Open source
 - Open hardware
4. Open methodology (**Reproducible Research**)
 - Open-notebook science
 - Open science infrastructures
5. Open peer review (avoid **collusion**)
6. Open educational resources



**NO TRANSPARENCY
NO CONSENSUS**



TRANSPARENT RESEARCH ~ OPEN SCIENCE ?

Plan National pour la Science Ouverte (BSN ↔ CoSO)

- CNRS, Inria, INRAE, ...
- Many flavors: *Citizen Science*

Main pillars:

1. Open access
2. Open data
3. Open source
 - Open hardware
4. Open methodology (**Reproducible Research**)
 - Open-notebook science
 - Open science infrastructures
5. Open peer review (avoid **collusion**)
6. Open educational resources



**NO TRANSPARENCY
NO CONSENSUS**



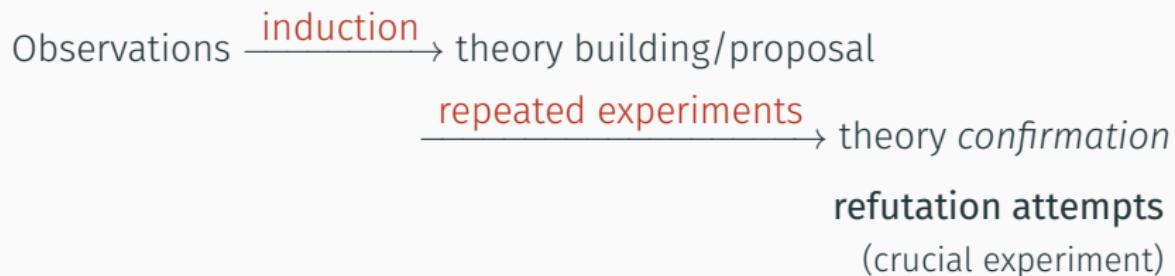
Obviously **making code/data available for the reproduction of results from published papers** has become the new norm

THE SCIENTIFIC METHOD

CLAUDE BERNARD (1865) AND KARL POPPER (1934)

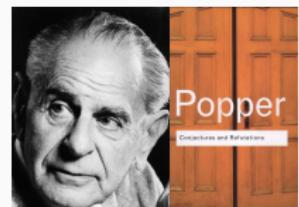


CLAUDE BERNARD (1865) AND KARL POPPER (1934)

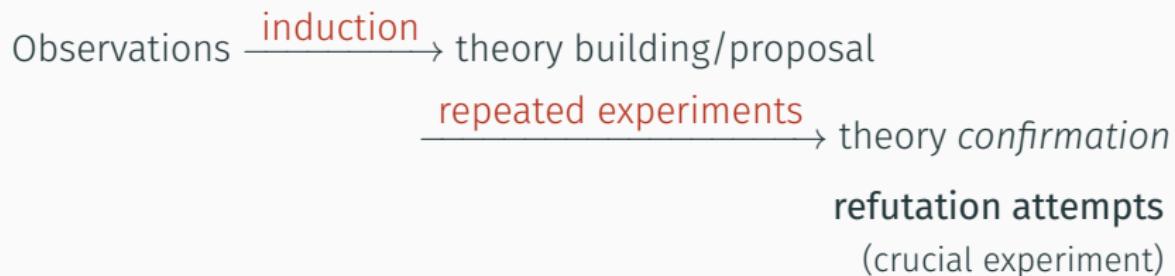


1934: Karl Popper puts the notions of **falsifiability** and **crucial experiment** as the **hallmark of science**

- If no experiment can be set up to **disprove** your theory, it is not science
- Good experiments **discriminate good theories from bad ones**
- **Non-reproducible** single occurrences are of no significance to science

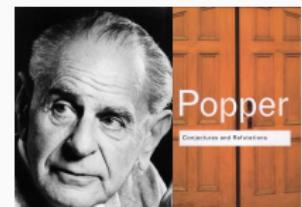


CLAUDE BERNARD (1865) AND KARL POPPER (1934)



1934: Karl Popper puts the notions of **falsifiability** and **crucial experiment** as the **hallmark of science**

- If no experiment can be set up to **disprove** your theory, it is not science
- Good experiments **discriminate good theories from bad ones**
- **Non-reproducible** single occurrences are of no significance to science



An ideal rather than the norm

Popper's proposal works well for Physics from the 18th century but is not so simple for many other domains (theory of evolution, astrophysics, particle physics, biology, anthropology)

AND IN PRACTICE ?

Crucial experiment: what happens when an experiment whose results are incoherent with the theory?

AND IN PRACTICE ?

Crucial experiment: what happens when **an experiment whose results are incoherent with the theory?** Is it *real* ?

- Did I finally manage to break the *dominant theory*?
 - Others will want to confirm by repeating the experiment.
 - Did I make a **mistake**?
 - Whoops
- ~~ Careful **inspection of the experimental process**

AND IN PRACTICE ?

Crucial experiment: what happens when **an experiment whose results are incoherent with the theory?** Is it *real* ?

- Did I finally manage to break the *dominant theory*?
 - Others will want to confirm by repeating the experiment.
 - Did I make a **mistake**?
 - Whoops
- ~~ Careful **inspection of the experimental process**

The role of statistics: what are the chances that I observe this by chance (i.e., there is actually no relation but I observe one)?

- Observation \neq experiments (**HARKING**)
 - Selective reporting (**p-hacking**)
- ~~ **Objectivity?** Failed/successful experiments are often carefully selected...

AND IN PRACTICE ?

Crucial experiment: what happens when **an experiment whose results are incoherent with the theory?** Is it *real* ?

- Did I finally manage to break the *dominant theory*?
 - Others will want to confirm by repeating the experiment.
 - Did I make a **mistake**?
 - Whoops
- ~~ Careful **inspection of the experimental process**

The role of statistics: what are the chances that I observe this by chance (i.e., there is actually no relation but I observe one)?

- Observation \neq experiments (**HARKING**)
 - Selective reporting (**p-hacking**)
- ~~ **Objectivity?** Failed/successful experiments are often carefully selected...

These mistakes require **more than public numbers and raw data**

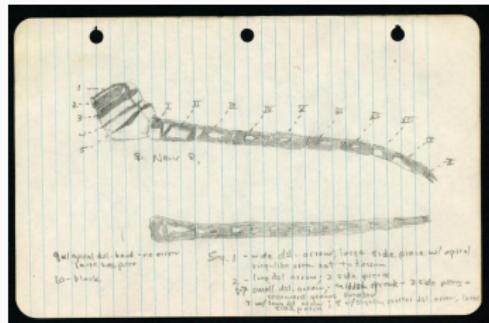
NOTEBOOKS

FIELD AND LABORATORY NOTEBOOKS

Social Sciences, Ecology, Biology



Chemistry, Physics, Biology



Robert Henry Gibbs, Jr.,
ichthyologist (1929 – 1988)



Emil Heinrich du Bois-Reymond,
electrophysiology (1818 – 1896)

FIELD AND LABORATORY NOTEBOOKS

Social Sciences, Ecology, Biology



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Marie-Claude Quidoz, Centre d'Ecologie
Fonctionnelle et Evolutive

Essentially no evolution throughout the last century

Chemistry, Physics, Biology



Contemporary laboratory notebook
in Neurology

PROS AND CONS

Pros

- Intuitive
- Convenient (lightweight)
- Sketching, equations, thoughts
- *Difficult* to temper

Cons

- Attachments
- Large amount of information
- Indexing and Searching
- No sharing nor backup

PROS AND CONS

Pros

- Intuitive
- Convenient (lightweight)
- Sketching, equations, thoughts
- *Difficult* to temper

Cons

- Attachments
- Large amount of information
- Indexing and Searching
- No sharing nor backup

Toward digital notebooks

Electronic Laboratory Notebooks and Laboratory Information Management Systems

- Batch and Sample management
- Instrument integration
- Stock management (e.g., chemical substances)
- Native handling of digital objects

ELN EXAMPLES 1/2: ELABFTW

eLabFTW: Chemistry @ Lyon (web server w. PHP)

Experiments

Filter status Order by Sort 15 Create

Expand all - Select all

Some microscopy experiment
Next step: analyze data
 RUNNING 2020.04.30 project SIMA
video1

Another experiment
 SUCCESS 2020.04.30 Project SIMA

Making fake experiments to do a screenshot
 SUCCESS 2020.04.30 elabweb elabftw.net

Looking at cells
 SUCCESS 2020.04.30

A failed experiment
 FAIL 2020.04.30 sirma

无标题
 SUCCESS 2020.04.30

Senza titolo
 SUCCESS 2020.04.27

lock unlock
 SUCCESS 2020.04.22

mathjax I
 SUCCESS 2020.04.21

analysis of sample #2343 from this company
 NEED TO BE REDONE 2020.04.16 project X

が無効になっています。
 SUCCESS 2020.04.20

Processing of Sample #392A
 SUCCESS 2020.04.16 project X

Qui pariatur aut architecto fuga.
 NEED TO BE REDONE 2020.03.27 labore
minus inventore qui la'g

mathjax test $\frac{d}{dx} x\dot{x} = \dot{x}^2 + x\ddot{x}$
 SUCCESS 2020.04.16

Model test
 RUNNING 2020.04.07 null non sed
omnis sit

NEXT PAGE

ELN EXAMPLES 1/2: eLABFTW

eLabFTW: Chemistry @ Lyon (web server w. PHP)

Experiments Back to listing

2020.04.30 Success Team User

project SIMA video1

Some microscopy experiment

Next step: analyze data

Goal:

Show what an experiment looks like for elabftw.net.

Sample #	Concentration	Absorbance	Density	Something
1	20 µM	0.239	3.4	102
2	30 µM	0.193	4.1	1339

You can have **colors** too.

And use **bold**, *italics*, underline like you would do in a normal text processing application. Have some mathjax: $x^y z$ is x^{y^z} , and $x^y \{y^z\}$ is x^{y^z} .

This is a list:

- List item 1
- List item 2
- Now show an image in the text

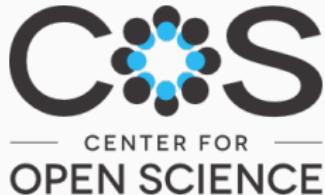


9/23

ELN EXAMPLES 2/2: OSF.io

A **cloud** to ease collaboration at the team/lab level

- Shared documents (articles, data, spreadsheets ...)
- Built-in version control
- Integration with Zotero/Mendeley
- Preprints, DMP, Pre-registration, ...
- Social network



OSPHOME ▾ My Projects Search Support Donate Arnaud Legrand ▾

Infrastructure and Resource ... Metadata Files Wiki Analytics Registrations Contributors Settings

CREP Africa study: A Paradigmatic replication of Rottman and Young (2019) / Infrastructure and Resource Assessment report

1.3MB Public 0 ...

Contributors: Adeyemi Adetula, Patrick F. Forcher, Dana Basnight-Brown, Hans Ijzerman, Arnaud Legrand

Date created: 2021-07-18 08:02 AM | Last updated: 2023-05-10 09:19 PM

Category: Uncategorized

License: CC-BY Attribution 4.0 International

Files

Name	Modified
Infrastructure and Resource Assess...	2021-07-22 02:1...
OSF Storage (United States)	
Infrastructure and Resources ...	2021-07-22 02:1...

Citation

Recent Activity

- Dana Basnight-Brown registered Infrastructure and Resource Assessment report 2023-05-10 09:19 PM
- Adeyemi Adetula made Infrastructure and Resource Assessment report public 2021-10-21 01:00 PM
- Adeyemi Adetula created a view-only link to Infrastructure and Resource Assessment report 2021-10-21 12:56 PM
- Adeyemi Adetula added Arnaud Legrand as contributor(s) to Infrastructure and Resource Assessment report 2021-10-20 04:31 PM
- Adeyemi Adetula created an anonymous view-only link to Infrastructure and Resource Assessment report 2021-10-20 04:31 PM

10/23

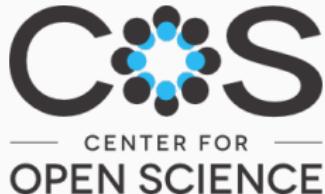
ELN EXAMPLES 2/2: OSF.io

A **cloud** to ease collaboration at the team/lab level

- Shared documents (articles, data, spreadsheets ...)
- Built-in version control
- Integration with Zotero/Mendeley
- Preprints, DMP, Pre-registration, ...
- Social network

But also

- Dropbox, Google drive storage
(privacy? sovereignty?)
- Google/Facebook/ORCID/Institution authentication
- **Locks you in** (huge machinery)
- Nothing done for **computations**



OSPHOME ▾ My Projects Search Support Donate Arnaud Legrand ▾ Infrastructure and Resource ... Metadata Files Wiki Analytics Registrations Contributors Settings

CREP Africa study: A Paradigmatic replication of Rottman and Young (2019) /

Infrastructure and Resource Assessment report

Contributors: Adeyemi Adetula, Patrick S. Forcher, Dana Basnight-Brown, Hans Ijzerman, Arnaud Legrand Date created: 2021-07-18 08:02 AM | Last updated: 2023-05-10 09:19 PM Category: Uncategorized License: CC-By Attribution 4.0 International

Files	
Name	Modified
Infrastructure and Resource Assess...	2021-07-22 02:1...
OSF Storage (United States)	
Infrastructure and Resources ...	2021-07-22 02:1...

Citation	
Recent Activity	
Dana Basnight-Brown registered Infrastructure and Resource Assessment report	2023-05-10 09:19 PM
Adeyemi Adetula made Infrastructure and Resource Assessment report public	2021-10-21 01:00 PM
Adeyemi Adetula created a view-only link to Infrastructure and Resource Assessment report	2021-10-21 12:56 PM
Adeyemi Adetula added Arnaud Legrand as contributor(s) to Infrastructure and Resource Assessment report	2021-10-20 04:31 PM
Adeyemi Adetula created an anonymous view-only link to	2021-10-20 04:31 PM

DAMNED COMPUTERS...

How computers broke science – and what we can do about it

*Most modern science is **so complicated**, and most journal articles **so brief**, it's impossible for the article to include details of many important methods and decisions made by the researcher as he analyzed his data on his computer. How, then, can another researcher judge the reliability of the results, or reproduce the analysis?*



– Ben Marwick,
The conversation, 2015

DAMNED COMPUTERS...

How computers broke science – and what we can do about it

*Most modern science is **so complicated**, and most journal articles **so brief**, it's impossible for the article to include details of many important methods and decisions made by the researcher as he analyzed his data on his computer. How, then, can another researcher judge the reliability of the results, or reproduce the analysis?*



– Ben Marwick,
The conversation, 2015

When moving to the digital world, the rigor is often completely lost.

- Point-and-click procedures are rampant but hinder reproducibility.
- **Spreadsheets** are generalized and intensively used in biology:
 - Membrane-Associated Ring Finger (C3HC4) 1,
 - E3 Ubiquitin Protein Ligase → MARCH1 → 2016-03-01 → 1456786800
 - 2310009E13 → 2.31E+19
- And more recently, we had the **COVID tracing** failure.

Not even mentioning **non-mature** software and **blackbox** machine learning procedures

NOTES AND DOCUMENTING

What your research supposedly looks like:

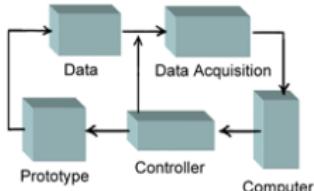


Figure 1. Experimental Diagram

What your research *actually* looks like:

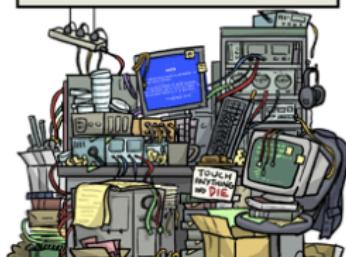


Figure 2. Experimental Mess

JORGE CHAM © 2008

WWW.PHDCOMICS.COM

MAIN CHALLENGES FOR A COMPUTATIONAL SCIENTIST

```
1 my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

MAIN CHALLENGES FOR A COMPUTATIONAL SCIENTIST

```
1 my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and
`-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

MAIN CHALLENGES FOR A COMPUTATIONAL SCIENTIST

```
1 my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and
`-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

Tracking parameters and data

- `*.dat`? Ooh, you ran this in `data/2091293-AJXQ37`?
- Wasn't `mymap.dat` updated since then?
- That was for `foo.csv`. What about `bar.csv`? Is it reproducible?

MAIN CHALLENGES FOR A COMPUTATIONAL SCIENTIST

```
1 my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and
`-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

Tracking parameters and data

- `*.dat`? Ooh, you ran this in `data/2091293-AJXQ37`?
- Wasn't `mymap.dat` updated since then?
- That was for `foo.csv`. What about `bar.csv`? Is it reproducible?

Tracking the process (on short/long term)

- Why did I run this? What did I learn from it? I remember doing this but when?

MAIN CHALLENGES FOR A COMPUTATIONAL SCIENTIST

```
1 my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and
`-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

Tracking parameters and data

- `*.dat`? Ooh, you ran this in `data/2091293-AJXQ37`?
- Wasn't `mymap.dat` updated since then?
- That was for `foo.csv`. What about `bar.csv`? Is it reproducible?

Tracking the process (on short/long term)

- Why did I run this? What did I learn from it? I remember doing this but when?

TOOL 1: COMPUTATIONAL NOTEBOOKS (LITTERATE PROGRAMMING)

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

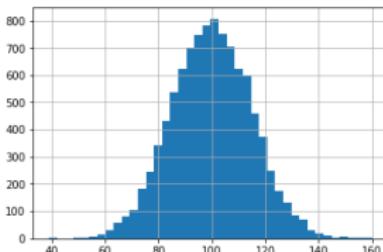
3.141592653589793

Mais calculé avec la méthode des [aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
```

3.1437198694098765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS (LITTERATE PROGRAMMING)

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with several code cells:

- In [1]:** `# Un document computationnel`
Output: Mais ordinateur m'indique que \$pi\$ vaut "approximativement"
- In [1]:** `from math import *
print(pi)`
Output: 3.141592653589793
- In [2]:** `Mais calculé avec la méthode des aiguilles de Buffon (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtientrait comme approximation :`
- In [2]:** `import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/sum((x+np.sin(theta))>1)/N`
Output: 3.1437198694998765
- In [3]:** `On peut inclure des formules mathématiques comme $\\sqrt{1/(\\pi * sign(\\sqrt{2}))} \\exp(-\\exp(-\\left(\\frac{(x-\\mu)^2}{2 * sign(\\sqrt{2})^2}\\right)))$ et des dessins qui n'ont rien à voir avec $\\pi$, avec $\\$!$ (si ce n'est une constante de normalisation...).`
- In [3]:** `%matplotlib inline
import matplotlib.pyplot as plt

mu, sigma = 100, 15
x = mu + sigma*np.random.randn(10000)

plt.hist(x,40)
plt.grid(True)
plt.show()`
Output: A histogram showing a bell-shaped distribution centered around 100, with x-axis from 40 to 160 and y-axis from 0 to 800.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

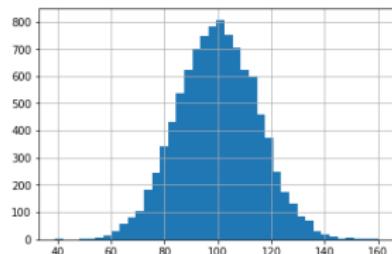
3.141592653589793

Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np  
N = 1000000  
x = np.random.uniform(size=N, low=0, high=1)  
theta = np.random.uniform(size=N, low=0, high=pi/2)  
2/sum((x+np.sin(theta))>1)/N
```

3.1437198694998765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation...).



TOOL 1: COMPUTATIONAL NOTEBOOKS (LITTERATE PROGRAMMING)

Document initial dans son environnement

A screenshot of a Jupyter Notebook interface. The title cell contains the text '# Un document computationnel'. Below it, a text cell says 'Mon ordinateur m'indique que π vaut "approximativement"'. An input cell (In [1]) shows the Python code:

```
from math import *  
print(pi)  
3.141592653589793
```

. An output cell (Out [1]) displays the result: 3.141592653589793. Another text cell explains the calculation using the Buffon's needle method. A second input cell (In [2]) contains code to calculate pi using theBuffon's needle method:

```
import numpy as np  
N = 1000000  
x = np.random.uniform(size=N, low=0, high=1)  
theta = np.random.uniform(size=N, low=0, high=np.pi/2)  
2/(sum((x+np.sin(theta))>1))/N
```

. The output (Out [2]) is 3.14371986944998765. A text cell at the bottom discusses including mathematical formulas and drawings.

In [3]:

```
%matplotlib inline  
import matplotlib.pyplot as plt  
  
mu, sigma = 100, 15  
x = mu + sigma*np.random.randn(10000)  
  
plt.hist(x,40)  
plt.grid(True)  
plt.show()
```

A histogram plot showing a bell-shaped curve centered at 100. The x-axis ranges from 40 to 160, and the y-axis ranges from 0 to 800. The peak of the distribution is at approximately 100, with a frequency of about 800.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

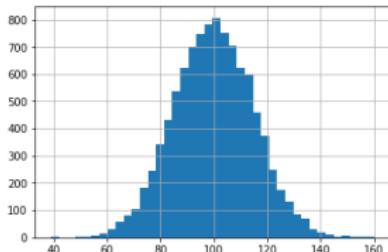
3.141592653589793

Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np  
N = 1000000  
x = np.random.uniform(size=N, low=0, high=1)  
theta = np.random.uniform(size=N, low=0, high=np.pi/2)  
2/(sum((x+np.sin(theta))>1))/N
```

3.14371986944998765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS (LITTERATE PROGRAMMING)

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** `from math import *
print(pi)` → **Out[1]:** 3,141592653589793
- In [2]:** `import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2*(sum((x+np.sin(theta))>1))/N` → **Out[2]:** 3,1437198694098765
- In [3]:** `%matplotlib inline
import matplotlib.pyplot as plt

mu, sigma = 100, 15
x = mu + sigma*np.random.randn(10000)

plt.hist(x,60)
plt.grid(True)
plt.show()` → **Out[3]:** A histogram showing a normal distribution centered at 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

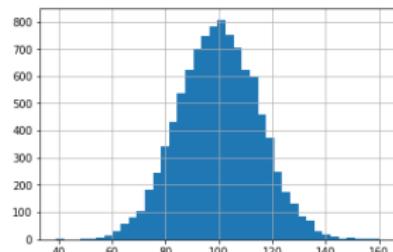
3.141592653589793

Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
import numpy as np  
N = 1000000  
x = np.random.uniform(size=N, low=0, high=1)  
theta = np.random.uniform(size=N, low=0, high=pi/2)  
2*(sum((x+np.sin(theta))>1))/N
```

3.1437198694098765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS (LITTERATE PROGRAMMING)

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** Prints the value of pi (3.141592653589793) and includes a note about calculating pi with Buffon's needle method.
- In [2]:** Generates random points (x, y) and calculates the ratio of points below the unit circle to total points, which approximates pi.
- In [3]:** Plots a histogram of x values from -100 to 100, showing a symmetric bell-shaped distribution centered at 0.

Annotations with red arrows point from the notebook cells to the final document on the right, specifically highlighting the code and its output.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

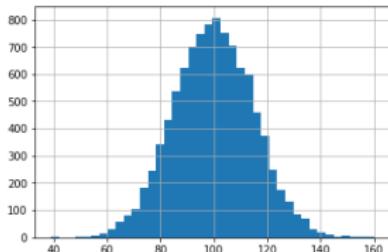
Mais calculé avec la méthode des [aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1))/N
```

3.14371986949098765

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et

des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS (LITTERATE PROGRAMMING)

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** Prints the value of pi (3.141592653589793) and includes a note about calculating pi with Buffon's needle method.
- In [2]:** Generates random points (x, y) and calculates the ratio of points below the unit circle to the total number of points, which is approximately 0.7854.
- In [3]:** Plots a histogram of 100,000 random numbers between 0 and 1, showing a bell-shaped distribution centered around 0.5.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

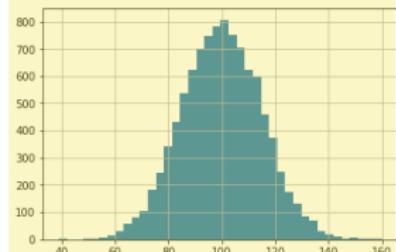
Mais calculé avec la méthode des [aiguilles de Buffon](#), on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
```

3.1437198694098765

Export →

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



TOOL 1: COMPUTATIONAL NOTEBOOKS (LITTERATE PROGRAMMING)

Document initial dans son environnement

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** `# Un document computationnel`
Mon ordinateur m'indique que π vaut approximativement
3,141592653589793
- In [2]:** `import numpy as np`
n = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
- Out[2]:** 3,1437198694098765

Below the notebook is a histogram plot of a normal distribution centered at 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

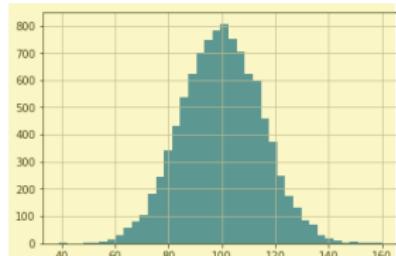
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
```

3.1437198694098765

Export →

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).



QUICK DEMO OF JUPYTER, RSTUDIO, AND ORG-MODE



(python)

Pros Python/R, friendly, portable (web browser, client/server)

- Cons**
- Installation, software dependencies (`minimal-notebook` \approx 440Mb)
 - Limited control on typography (unless using *Rube Goldberg* machines like `quarto`)

QUICK DEMO OF JUPYTER, RSTUDIO, AND ORG-MODE



(python)

Pros Python/R, friendly, portable (web browser, client/server)

- Cons**
- Installation, software dependencies ([minimal-notebook](#) \approx 440Mb)
 - Limited control on typography (unless using *Rube Goldberg* machines like [quarto](#))



RStudio (java/R)

Pros R/Python, friendly, portable, real IDE, Good typography control

- Cons**
- Installation, software dependencies ([rocker/rstudio](#) \approx 550Mb)

QUICK DEMO OF JUPYTER, RSTUDIO, AND ORG-MODE



(python)

Pros Python/R, friendly, portable (web browser, client/server)

- Cons**
- Installation, software dependencies ([minimal-notebook](#) \approx 440Mb)
 - Limited control on typography (unless using *Rube Goldberg* machines like [quarto](#))



(java/R)

Pros R/Python, friendly, portable, real IDE, Good typography control

- Cons**
- Installation, software dependencies ([rocker/rstudio](#) \approx 550Mb)



(emacs-lisp)

Pros R, Python, Perl, Ruby, C, Java, ...

- Pure text, Good typography control, "Lightweight", Programmable editor

Cons

- No default configuration \rightsquigarrow rough and steep learning curve
- Big machinery: [silex/emacs-alpine-ci](#) \approx 240Mb even though [flycheck/emacs-cask](#) \approx 80Mb

WHY AND HOW TO USE A COMPUTATIONAL NOTEBOOK?

The processing steps between raw observations and findings have gotten increasingly numerous and complex

- A **reproducible document** links **raw data** with **final figures**
- What's at stake: **verification** and **reuse** by others

WHY AND HOW TO USE A COMPUTATIONAL NOTEBOOK?

The processing steps between raw observations and findings have gotten increasingly numerous and complex

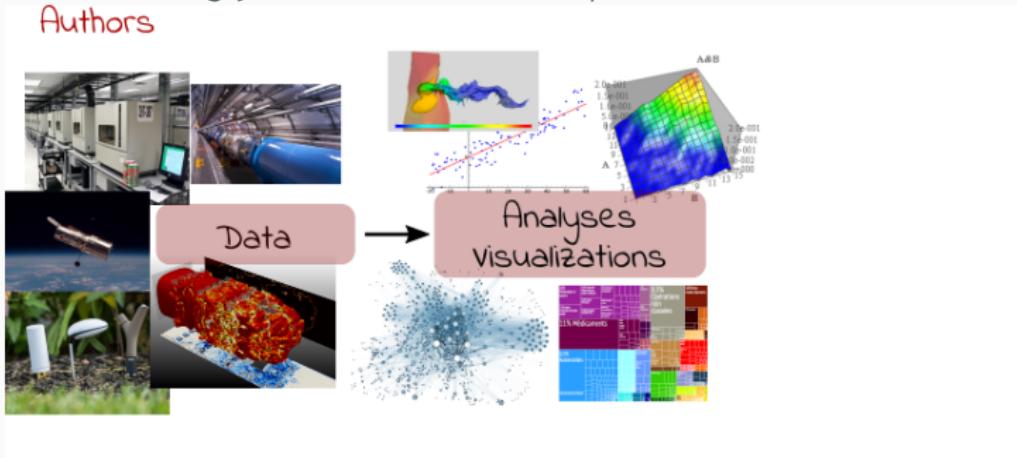
Authors



- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

WHY AND HOW TO USE A COMPUTATIONAL NOTEBOOK?

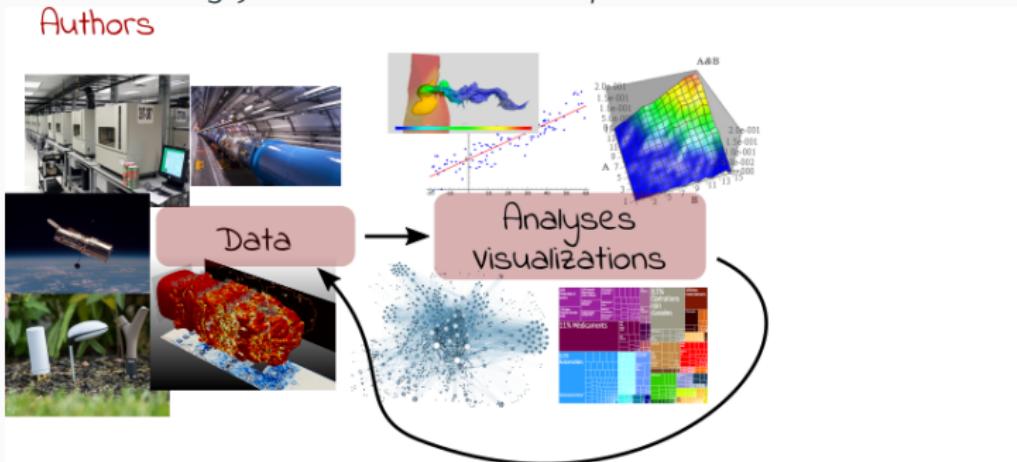
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A **reproducible document** links **raw data** with **final figures**
- What's at stake: **verification** and **reuse** by others

WHY AND HOW TO USE A COMPUTATIONAL NOTEBOOK?

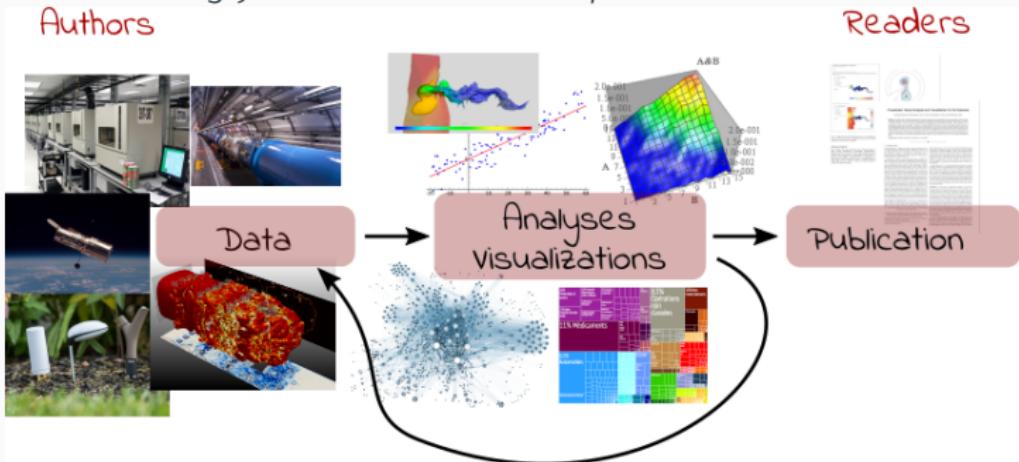
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A **reproducible document** links **raw data** with **final figures**
- What's at stake: **verification** and **reuse** by others

WHY AND HOW TO USE A COMPUTATIONAL NOTEBOOK?

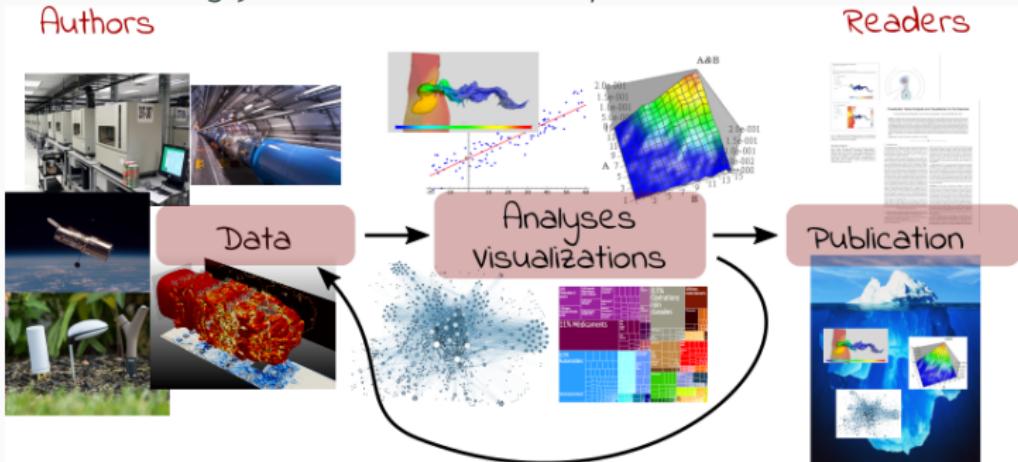
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A **reproducible document** links **raw data** with **final figures**
- What's at stake: **verification** and **reuse** by others

WHY AND HOW TO USE A COMPUTATIONAL NOTEBOOK?

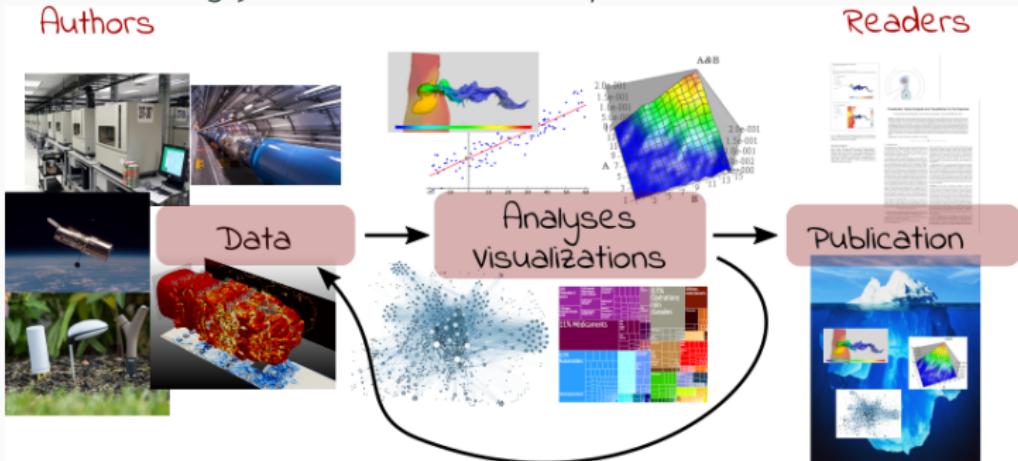
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A **reproducible document** links **raw data** with **final figures**
- What's at stake: **verification** and **reuse** by others

WHY AND HOW TO USE A COMPUTATIONAL NOTEBOOK?

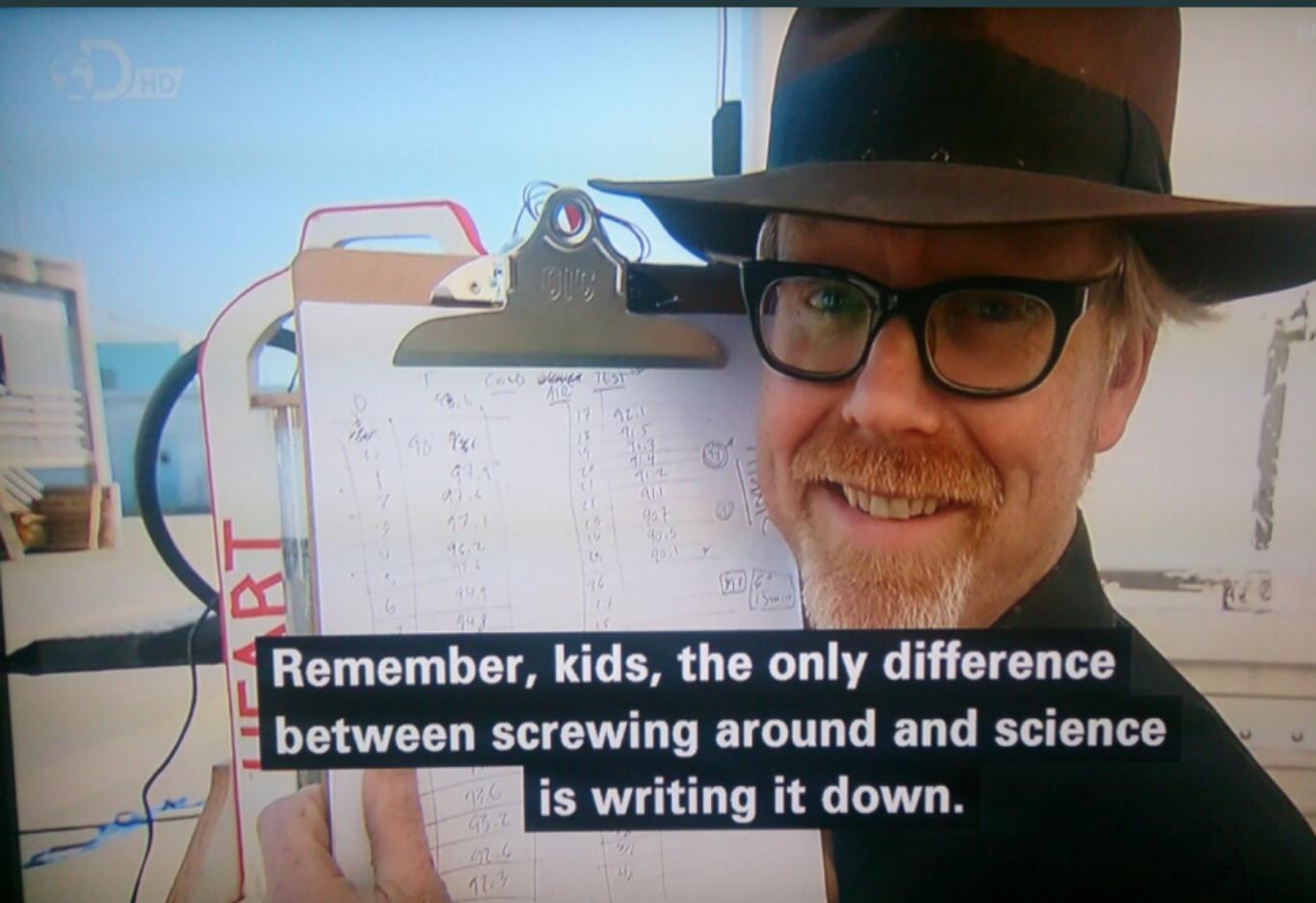
The processing steps between raw observations and findings have gotten increasingly numerous and complex



Reproducible Research = Bridging the Gap by working Transparently

- A **reproducible document** links **raw data** with **final figures**
- What's at stake: **verification** and **reuse** by others

TOOL 2: ELECTRONIC NOTEBOOKS



**Remember, kids, the only difference
between screwing around and science
is writing it down.**

TAKING NOTES

Org-mode and Markdown two simple text formats

- simple formatting and export to more elaborate formats
- Hyperlinks, images, code, etc.

Journal structure

- My journal/notebook (single org file)
- Tom Cornebize's journal (single org file + Jupyter notebooks)

Zettelkasten possible structure

- Org-roam, Zettler, Roam, Obsidian, ...
- Architects, librarians, **gardeners**



TAKING NOTES

Org-mode and Markdown two simple text formats

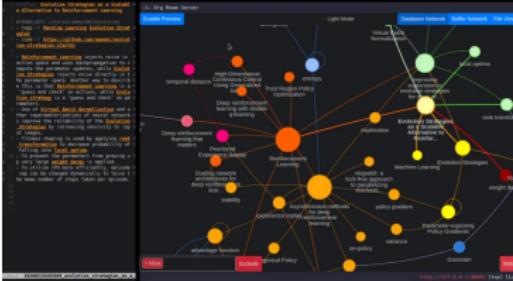
- simple formatting and export to more elaborate formats
- Hyperlinks, images, code, etc.

Journal structure

- My journal/notebook (single org file)
- Tom Cornebize's journal (single org file + Jupyter notebooks)

Zettelkasten possible structure

- Org-roam, Zettler, Roam, Obsidian, ...
- Architects, librarians, **gardeners**



My recommendations: Do not use a fancy cloud-based proprietary tool

- Simple open source **text-based** format
- **Control versions** and backup yourself (e.g., using gitlab, github)
- Single location if possible
- Annotate (tags in a journal, links in a Zettelkasten)

CONCLUSION

REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments

- Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- [Markdown](#), [GitLab](#), [Jupyter](#) / [Rstudio](#) / [OrgMode](#)
- [3rd Edition: March 2020 – March 2024 \(15,000+\)](#)



REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments

- Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- `Markdown`, `GitLab`, `Jupyter` / `Rstudio` / `OrgMode`
- 3rd Edition: March 2020 – March 2024 (15,000+)



MOOC "Advanced RR" planned for Jan. 2024

- Managing data (`FITS/HDF5`, `git annex`)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments

- Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- `Markdown`, `GitLab`, `Jupyter` / `Rstudio` / `OrgMode`
- 3rd Edition: March 2020 – March 2024 (15,000+)



MOOC "Advanced RR" planned for Jan. 2024

- Managing data (`FITS/HDF5`, `git annex`)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

2. Change the norm: make publication practices evolve

3. Incentive: consider RR/open science when hiring/promoting

REPRODUCIBLE RESEARCH = RIGOR AND TRANSPARENCY

Good research requires time and resources

1. Train yourself and your students: RR, statistics, experiments

- Beware of checklists and norms Understand what's at stake

MOOC Reproducible Research: Methodological principles for a transparent science, Inria Learning Lab

- Konrad Hinsen, Christophe Pouzat
- `Markdown`, `GitLab`, `Jupyter` / `Rstudio` / `OrgMode`
- 3rd Edition: March 2020 – March 2024 (15,000+)



MOOC "Advanced RR" planned for Jan. 2024

- Managing data (`FITS/HDF5`, `git annex`)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

2. Change the norm: make publication practices evolve

3. Incentive: consider RR/open science when hiring/promoting

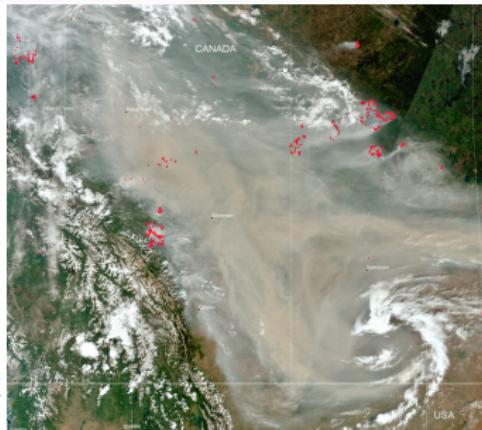
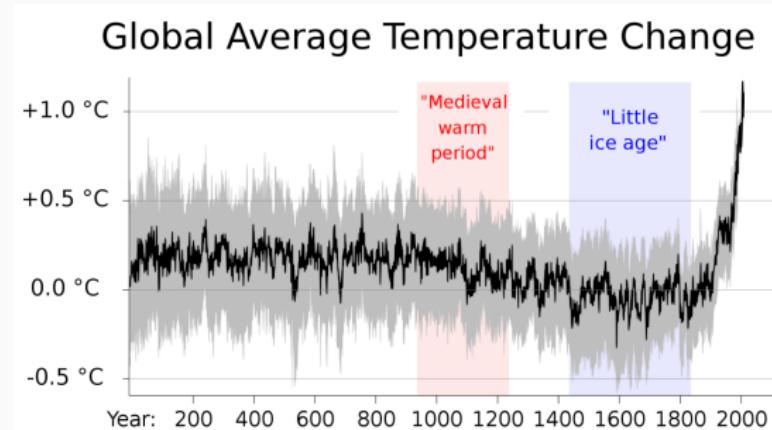
THE SCIENCE IS CLEAR

Why are we
ignoring it?

scientist rebellion

IPCC, IPBES, <https://climate.nasa.gov/>

1. Global climate change is not a future problem



https://en.wikipedia.org/wiki/Global_temperature_record

2023 Alberta wildfires (> 1 Mha)

THE SCIENCE IS CLEAR

scientist rebellion

Why are we ignoring it?

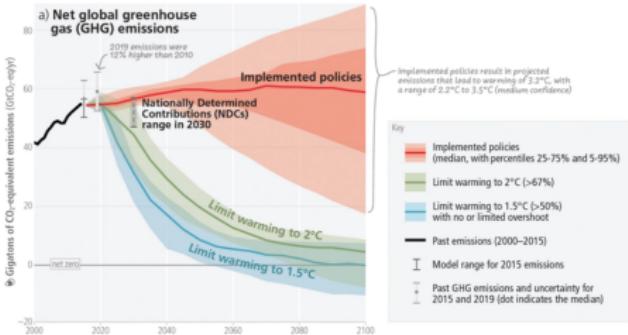


IPCC, IPBES, <https://climate.nasa.gov/>

1. Global climate change is **not** a future problem
2. It is **entirely** due to human activity

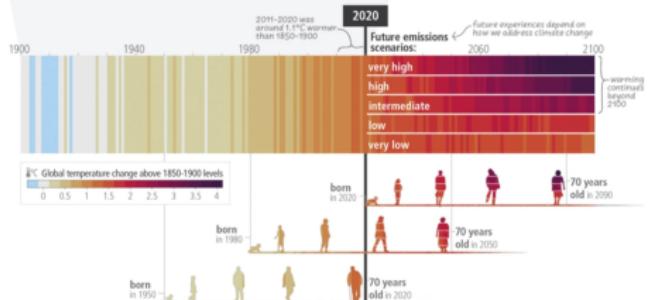
Limiting warming to **1.5°C** and **2°C** involves rapid, deep and in most cases immediate greenhouse gas emission reductions

Net zero: CO₂ and net zero GHG emissions can be achieved through strong reductions across all sectors



Paris Agreement'15 ~ Net Zero by 2050

c) The extent to which current and future generations will experience a hotter and different world depends on choices now and in the near-term



Latest IPCC report

19/23

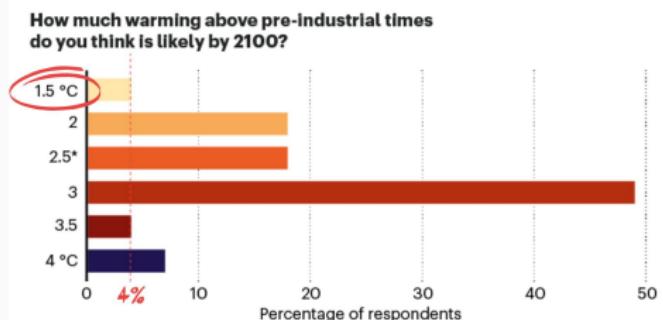
THE SCIENCE IS CLEAR

Why are we
ignoring it?

scientist rebellion

IPCC, IPBES, <https://climate.nasa.gov/>

1. Global climate change is **not** a future problem
2. It is **entirely** due to human activity
3. **9 out of 10 IPCC scientists believe overshoot is likely**

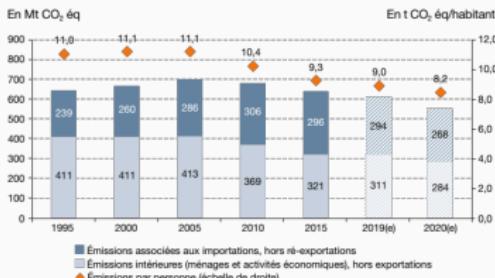


@natu Nature survey, Nov. 2021

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

Put aside biodiversity loss, pollution, freshwater, land system change...

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE



Empreinte carbone moyenne en France
10 tonnes de CO₂e/an/pers.



÷2
d'ici
2030

<2t CO₂e

Objectif d'ici 2050

- de 2 t de CO₂e/an/pers.

+ Faire plus d'activités bas carbone !

Danser, chanter, jardiner, rêver, écire, lire, courir, randonner, planter des arbres, discuter, marcher, méditer, méditer, passer du temps avec ceux qu'on aime, lire...

Bref, inventer nos vies bas carbone désirables !

Par exemple :

0,5 t CO₂e/Annee : à l'entretien régulatoire d'un jardin ou de produits urbains

0,5 t CO₂e/Annee : faire une partie nocturne (80%) de fabrication manuelle sur 30 ans, importer les matériaux et énergie et faire un peu de transports en commun.

0,5 t CO₂e/Annee : faire des expériments dans son jardin ou d'intégrer des arbres, sobriété dans l'usage des produits et énergie

0,2 t CO₂e/Annee : faire une partie nocturne (80%) de fabrication manuelle sur 30 ans, importer les matériaux et énergie et faire un peu de transports en commun.

0,2 t CO₂e/Annee : faire une partie nocturne (80%) de fabrication manuelle sur 30 ans, importer les matériaux et énergie et faire un peu de transports en commun.

0,2 t CO₂e/Annee : faire une partie nocturne (80%) de fabrication manuelle sur 30 ans, importer les matériaux et énergie et faire un peu de transports en commun.



<https://www.nosviesbascarbonne.org/>

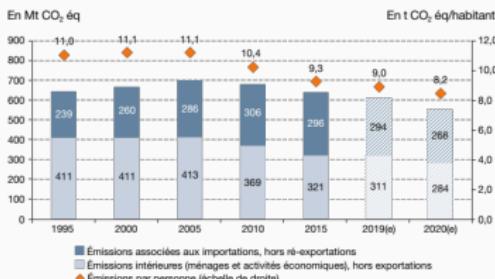
Sources : Kit Inventons nos vies bas carbone (Fév. 2021), Rapport sur l'état de l'environnement en France (Déc. 2020)

INVENTONS
NOS VIES
BAS CARBONE

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

Put aside biodiversity loss, pollution, freshwater, land system change...

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE



(e) = estimations.

Note : l'empreinte carbone porte sur les trois principaux gaz à effet de serre (CO₂, CH₄, N₂O). En 2021, la méthodologie a été ajustée afin de mieux tenir compte de l'évolution des coûts du pétrole brut, du gaz et du charbon. L'ensemble de la série a ainsi été révisé, l'essentiel des ajustements portant sur les émissions importées de CH₄.

Champ : périmètre Kyoto (Île-de-France et outre-mer appartenant à l'UE).

Sources : Citepa ; AIE ; FAO ; Douanes ; Eurostat ; Insee. Traitement : SDES, 2021



Empreinte carbone moyenne en France
10 tonnes de CO₂e/an/pers.



÷2
d'ici
2030

<2t CO₂e

Objectif d'ici 2050

- de 2 t de CO₂e/an/pers.

+ Faire plus d'activités bas carbone !

Danser, chanter, jardiner, rêver, écire, lire, courir, randonner, planter des arbres, discuter, marcher dans la forêt, méditer, admirer, passer du temps avec ceux qu'on aime, lire...

Bref, inventer nos vies bas carbone désirables !

Par exemple :

0,5 t CO₂e : Alimentation : à tendance végétalienne, hypergourmande en sans produits animaux

0,5 t CO₂e : Transport : 2000km en voiture (800km de fabrication amortie sur 30 ans, importations de pétrole et gaz, émissions de CO₂ dans les transports en commun)

0,5 t CO₂e : Consommation : dans un ménage, faire du sport régulièrement, faire diverses expérimenter éco-responsables et intemporelles, se déplacer dans le quartier à vélo ou à pied

0,2 t CO₂e : Logement : Choisir loger sur un emplacement en peripherie, "off-grid" et d'origine locale et durable, faire des travaux de rénovation à la maison, penser à la durabilité

0,2 t CO₂e : Services publics : faire évoluer, développer et améliorer les services publics

INVENTONS
NOS VIES
BAS CARBONE

<https://www.nosviesbascarbonne.org/>

Sources : Kit Inventons nos vies bas carbone (Fév. 2021), Rapport sur l'état de l'environnement en France (Déc. 2020)

French government response

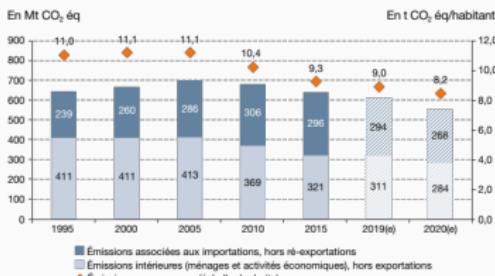
- Verdissement de l'industrie: « pause » sur les normes environnementales
- Loi de programmation militaire (+41%)
- Nous devons préparer la France à une élévation de la température de 4 °C
- Academia ? PEPR 5G, Cloud, NUMPEX, Quantique, IA, Agroécologie et numérique

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

2/2

Put aside biodiversity loss, pollution, freshwater, land system change...

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE



(e) = estimations.

Note : l'empreinte carbone porte sur les trois principaux gaz à effet de serre (CO₂, CH₄, N₂O). En 2021, la méthodologie a été ajustée afin de mieux tenir compte de l'évolution des coûts du pétrole brut, du gaz et du charbon. L'empreinte de la séries a ainsi été révisée, l'essentiel des ajustements portant sur les émissions importées de CH₄.

Champ : périmètre Kyoto (Île-de-France et outre-mer appartenant à l'UE).

Sources : Citepa ; AIE ; FAO ; Douanes ; Eurostat ; Insee. Traitement : SDES, 2021



Empreinte carbone moyenne en France
10 tonnes de CO₂e/an/pers.



Objectif d'ici 2050

- de 2 t de CO₂e/an/pers.

+ Faire plus d'activités bas carbone !

Danser, chanter, jardiner, rêver, écire, lire, courir, randonner, planter des arbres, discuter, marcher, méditer, admirer, passer du temps avec ceux qu'on aime, lire...

Bref, inventer nos vies bas carbone désirables !

Par exemple :



<https://www.nosviesbascarbonne.org/>

INVENTONS
NOUS VIES
BAS CARBONE

Sources : Kit Inventons nos vies bas carbone (Fév. 2021), Rapport sur l'état de l'environnement en France (Déc. 2020)

French government response

- Verdissement de l'industrie: « pause » sur les normes environnementales
- Loi de programmation militaire (+41%)
- Nous devons préparer la France à une élévation de la température de 4 °C
- Academia ? PEPR 5G, Cloud, NUMPEX, Quantique, IA, Agroécologie et numérique

Several scenarios on the table

- What will research/CS look like/be used for in such a world?
- Energy optimization/saving ≠ sobriety and frugality

OTHER NIGHTMARES AND TOOLS

NIGHTMARE 1: FIGHTING SOFTWARE ENVIRONMENTS NIGHTMARE

What is hiding behind a simple

1

```
import matplotlib
```

Package: python3-matplotlib
Version: 2.1.1-2
Depends: python3-dateutil, python-matplotlib-data (>= 2.1.1-2),
python3-pyparsing (>= 1.5.6), python3-six (>= 1.10), python3-tz,
libjs-jquery, libjs-jquery-ui, python3-numpy (>= 1:1.13.1),
python3-numpy-abi9, python3 (<< 3.7), python3 (>= 3.6~),
python3-cycler (>= 0.10.0), python3:any (>= 3.3.2-2~), libc6 (>= 2.14), libfreetype6 (>= 2.2.1), libgcc1 (>= 1:3.0), libpng16-16 (>= 1.6.2-1), libstdc++6 (>= 5.2), zlib1g (>= 1:1.1.4)

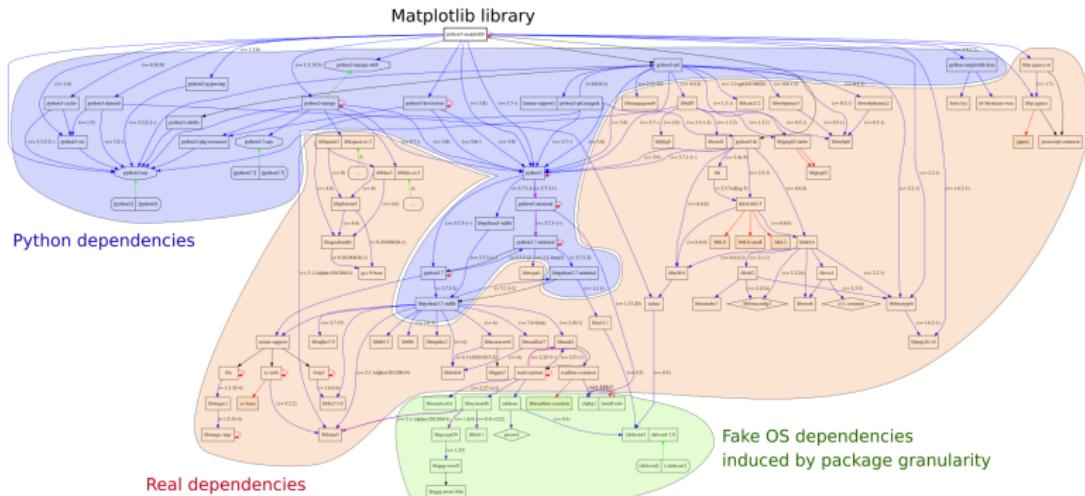
NIGHTMARE 1: FIGHTING SOFTWARE ENVIRONMENTS NIGHTMARE

What is hiding behind a simple

1

```
import matplotlib
```

Package: python3-matplotlib



NIGHTMARE 1: FIGHTING SOFTWARE ENVIRONMENTS NIGHTMARE

Python and its rapidly evolving environment

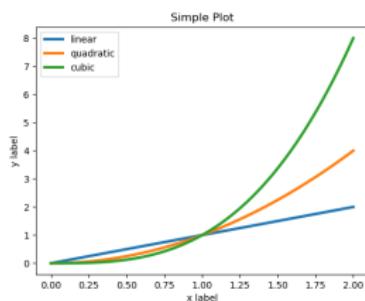
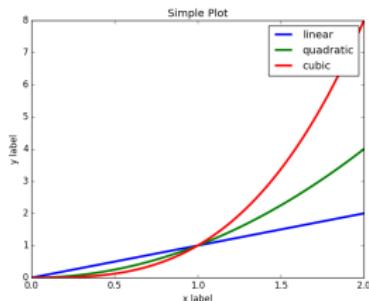
```
1 python2 -c "print(10/3)"  
2 python3 -c "print(10/3)"
```

```
3  
3.333333333333335
```

NIGHTMARE 1: FIGHTING SOFTWARE ENVIRONMENTS NIGHTMARE

Python and its rapidly evolving environment

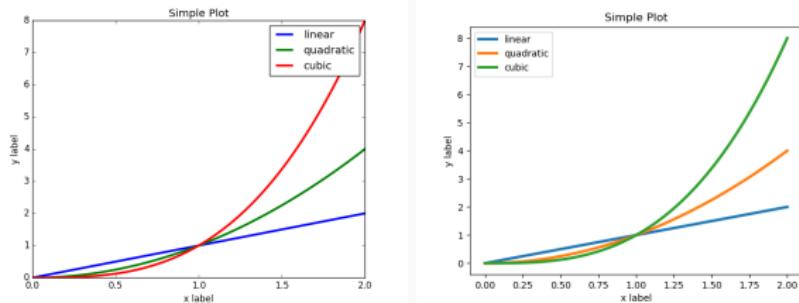
```
1 python2 -c "print(10/3)"  
2 python3 -c "print(10/3)"
```



NIGHTMARE 1: FIGHTING SOFTWARE ENVIRONMENTS NIGHTMARE

Python and its rapidly evolving environment

```
1 python2 -c "print(10/3)"  
2 python3 -c "print(10/3)"
```

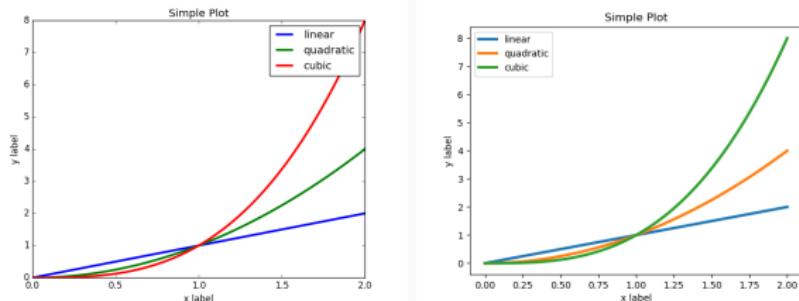


Cortical Thickness Measurements (PLOS ONE, June 2012): *FreeSurfer: differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.*

NIGHTMARE 1: FIGHTING SOFTWARE ENVIRONMENTS NIGHTMARE

Python and its rapidly evolving environment

```
1 python2 -c "print(10/3)"  
2 python3 -c "print(10/3)"
```



Cortical Thickness Measurements (PLOS ONE, June 2012): FreeSurfer: differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.



NIGHTMARE 2: FIGHTING INFORMATION LOSS WITH ARCHIVES

D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), Jan 2003.

The half-life of a referenced URL is approximately 4 years from its publication date.

P. Habibzadeh. *Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals*. Applied Clinical Informatics. 4 (4), 2013

half life ranged from 2.2 years in EMHJ to 5.3 years in BMJ

NIGHTMARE 2: FIGHTING INFORMATION LOSS WITH ARCHIVES

D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), Jan 2003.

The half-life of a referenced URL is approximately 4 years from its publication date.

P. Habibzadeh. *Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals*. Applied Clinical Informatics. 4 (4), 2013

half life ranged from 2.2 years in EMHJ to 5.3 years in BMJ

Article archives arXiv.org HAL
archives-ouvertes.fr

Data archives figshare zenodo

Software Archive Software Heritage



or



= awesome collaborations ≠ archive

THAT'S ALL FOLKS!
