

Reproducibility in Computational Experiments

Lucas Mello Schnorr



November 2020



Motivation and Introduction

Public Evidence for a Lack of Reproducibility

- J.P. Ioannidis. *Why Most Published Research Findings Are False* PLoS Med. 2005.
- *Lies, Damned Lies, and Medical Science*, The Atlantic. Nov, 2010
- *Reproducibility: A tragedy of errors*, Nature, Feb 2016.
- Steen RG, *Retractions in the scientific literature: is the incidence of research fraud increasing?*, J. Med. Ethics 37, 2011

The collage includes the following elements:

- Los Angeles Times BUSINESS**: Headline: "Science has lost its way, at a big cost to humanity". Subtext: "Researchers are rewarded for splashy findings, not for double-checking accuracy. So many scientists looking for cures to diseases have been building on ideas that aren't even true."
- Science AAAS NEWS SCIENCE JOURNALS CAREERS MULTIMEDIA COLLECTIONS**: Headline: "Science has lost its way, at a big cost to humanity". Subtext: "Researchers are rewarded for splashy findings, not for double-checking accuracy. So many scientists looking for cures to diseases have been building on ideas that aren't even true."
- nature International weekly journal of science**: Headline: "Announcement: Reducing our irreproducibility". Subtext: "Over the past year, Nature has published a string of articles that highlight the reliability and reproducibility of published research (collected as a special feature on the journal's website)."
- The Economist**: Headline: "Washington's lawyer surplus". Subtext: "How do a nuclear deal with Iran, Investment tips from Nobel economists, Junk bonds are back, The meaning of Sachin Tendulkar".
- HOW SCIENCE GOES WRONG.**: A book cover with a colorful, abstract design.
- TheScientist EXPLORING LIFE. INSPIRING INNOVATION NIH Tackles Irreproducibility**: Headline: "The federal agency speaks out about how to improve the quality of scientific research".

Courtesy V. Stodden
SC, 2015

Newsworthy Stories about Scientific Misconduct

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received \approx \$19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to **57 months imprisonment** for fabricating and falsifying data in HIV vaccine trials **He was also fined US \$7.2 million!**

Newsworthy Stories about Scientific Misconduct

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received ≈ \$19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials He was also fined US \$7.2 million!

Dieterik Stapel Professor, Social Psychology, Univ. Amsterdam, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [...] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

Newsworthy Stories about Scientific Misconduct

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received ≈ \$19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials He was also fined US \$7.2 million!

Dieterik Stapel Professor, Social Psychology, Univ. Amsterdam, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [...] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

Brian Wansink Professor, Psychological Nutrition, Cornell, 2016

When she arrived, I gave her a data set of a self-funded, failed study which had null results. I said "This cost us a lot of time and our own money to collect. There's got to be something here we can salvage because it's a cool (rich & unique) data set." I told her what the analyses should be and what the tables should look like. [...] Every day she came back with puzzling new results, and every day we would scratch our heads, ask "Why," and come up with another way to reanalyze the data with yet another set of plausible hypotheses

17 retracted publications

A Credibility Crisis?

Scientific misconduct is obviously wrong but it's **not new!**

- Every domain has its black sheep
- The publish or perish pressure is a huge pain

Media attention inflates **conspiracy opinions** 😞

Scientific results are worthless. Stop the scientific dictatorship/lobby!

The Battle against Scientific Fraud

CNRS International Magazine



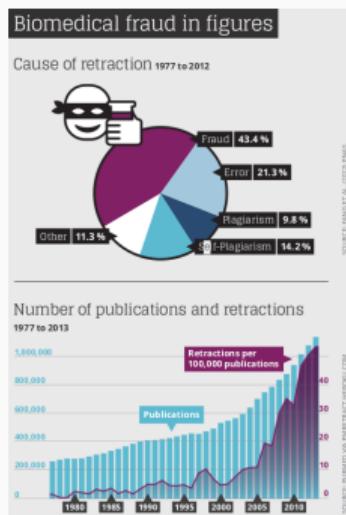
A Credibility Crisis?

Scientific misconduct is obviously wrong but it's **not new!**

- Every domain has its black sheep
- The publish or perish pressure is a huge pain

Media attention inflates **conspiracy opinions** 😞

Scientific results are worthless. Stop the scientific dictatorship/lobby!



Fraud is the (**uninteresting**) visible part of the iceberg

- **Failing to reproduce the results of others is common**

1,500 scientists lift the lid on reproducibility,

Nature, May 2016

- **How so? Why now? Why is this important? What can we do about it?**

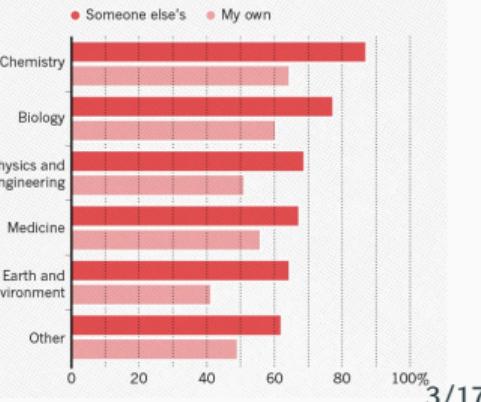
The Battle against Scientific Fraud

CNRS International Magazine



HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

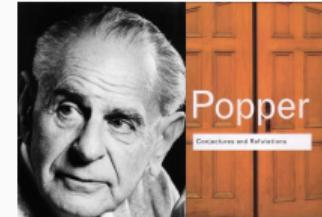
Most scientists have experienced failure to reproduce results.



Reproducibility of experimental results is the *hallmark of science*

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

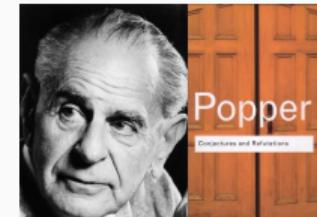
- If no experiment can be set up to **disprove** your theory, it is not science
- Good experiments **discriminate good theories from bad ones**
- **Non-reproducible** single occurrences are of no significance to science



Reproducibility of experimental results is the *hallmark of science*

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
- Biology (every animal does not behave in the same way)
- Spotting a SuperNova
- Anthropology (impact on people from a remote culture)
- Particle Physics (a single LHC)

Reproducibility: a *core value of science*

1. Universality: Science aims for **objective findings**, accessible to anyone

Reproducibility acts as a Universality/Robustness control

2. Incremental: We build on each others work but everybody makes mistakes

Methods, biases, ... How to discriminate sound theories experiments from bad ones? 😊

Reproducibility acts as a **Quality control**

Reproducibility: a *core value of science*

1. Universality: Science aims for **objective findings**, accessible to anyone

Reproducibility acts as a Universality/Robustness control

2. Incremental: We build on each others work but everybody makes mistakes

Methods, biases, ... How to discriminate sound theories experiments from bad ones? 😊

Reproducibility acts as a **Quality control**

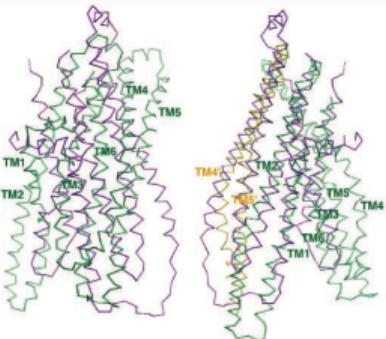
But, **scientific practices have greatly evolved**, in particular since we rely on **computers**

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

– Ben Marwick, The conversation, 2015



How computers broke science



Geoffrey Chang (Scripps, UCSD) works on crystallography and studies the structure of cell membrane proteins.

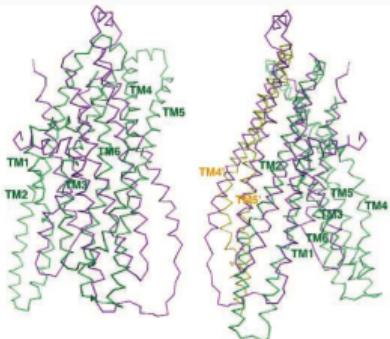
He specialized in structures of **multidrug resistant transporter proteins in bacteria**: MsbA de Escherichia Choli (Science, 2001), Vibrio cholera (Mol. Biology, 2003), Salmonella typhimurium (Science, 2005)

2006: Inconsistencies reveal a programming mistake

A *homemade data-analysis program had flipped two columns of data, inverting the electron-density map from which his team had derived the protein structure.*

5 retractions that motivate **improved software engineering practices** in comp. biology

How computers broke science



Geoffrey Chang (Scripps, UCSD) works on crystallography and studies the structure of cell membrane proteins.

He specialized in structures of **multidrug resistant transporter proteins in bacteria**: MsbA de Escherichia Choli (Science, 2001), Vibrio cholera (Mol. Biology, 2003), Salmonella typhimurium (Science, 2005)

2006: Inconsistencies reveal a programming mistake

A *homemade data-analysis program had flipped two columns of data, inverting the electron-density map from which his team had derived the protein structure.*

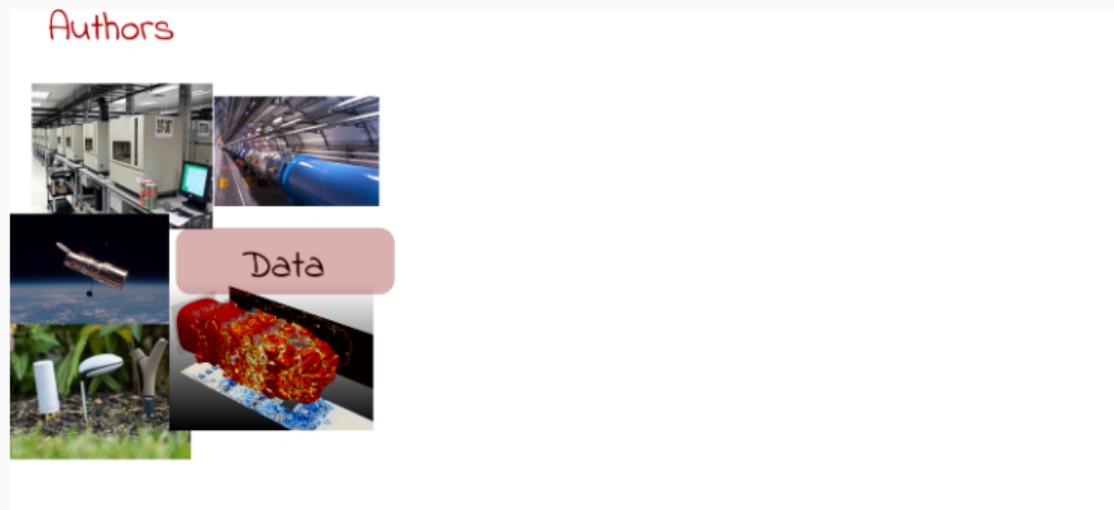
5 retractions that motivate **improved software engineering practices** in comp. biology

There is **worse!**

- The generalized and intensive use of **spreadsheets (COVID tracing)**
- Relying on **black box** statistical methods is infinitely easier than understanding them
 - Learning and Data Analytics frameworks are nuclear weapons
- Numerical errors and **software environment** unawareness

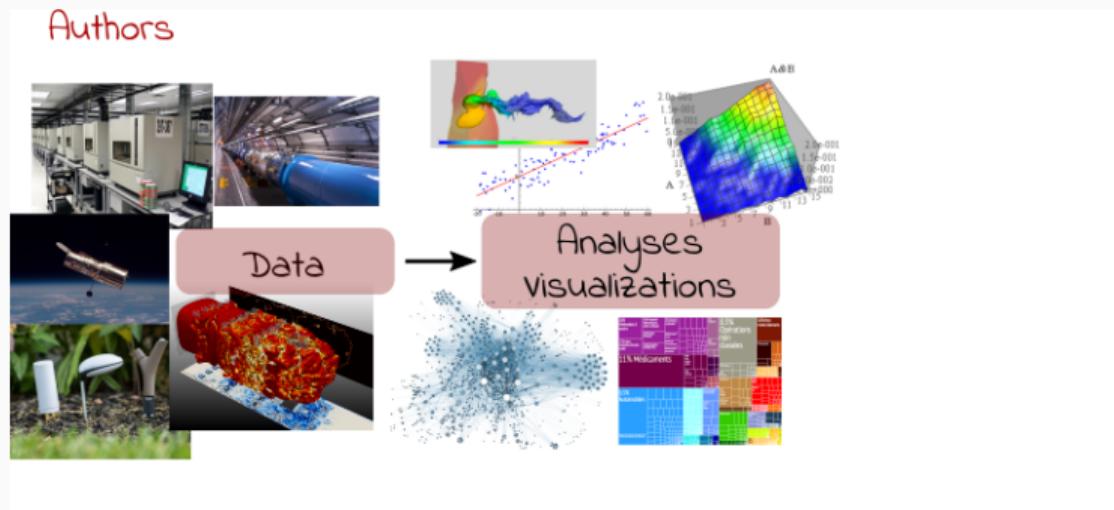
Modern Science

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



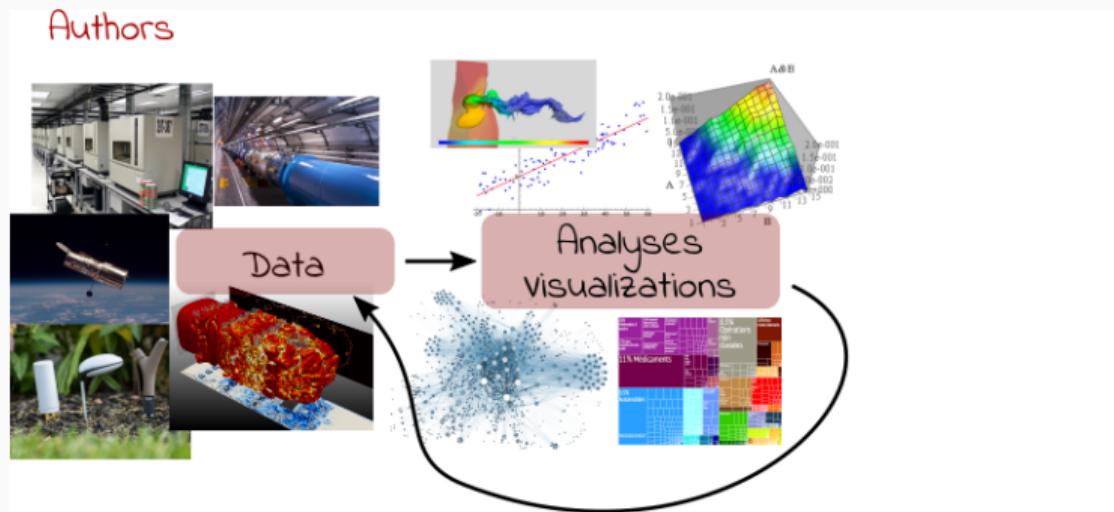
Modern Science

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



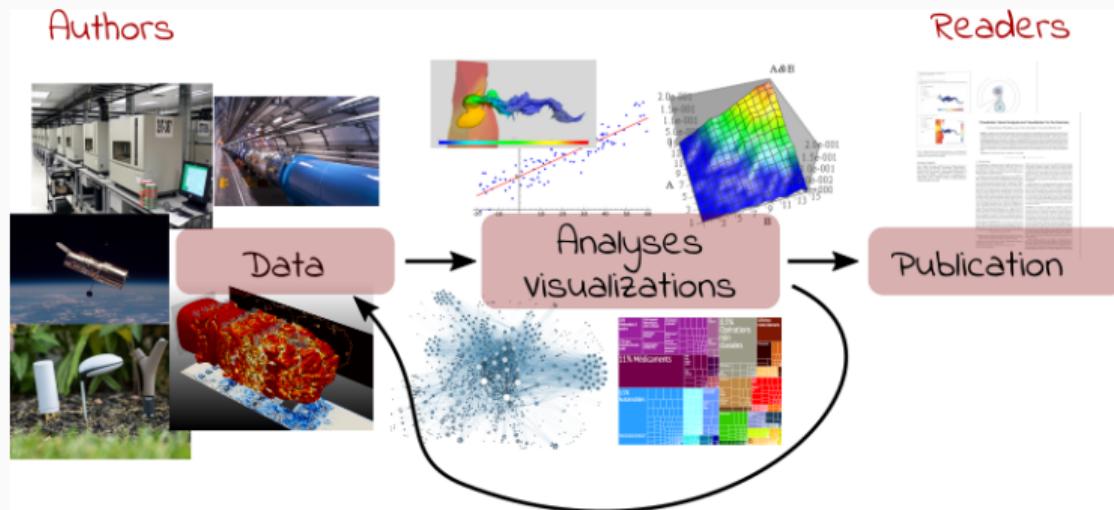
Modern Science

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



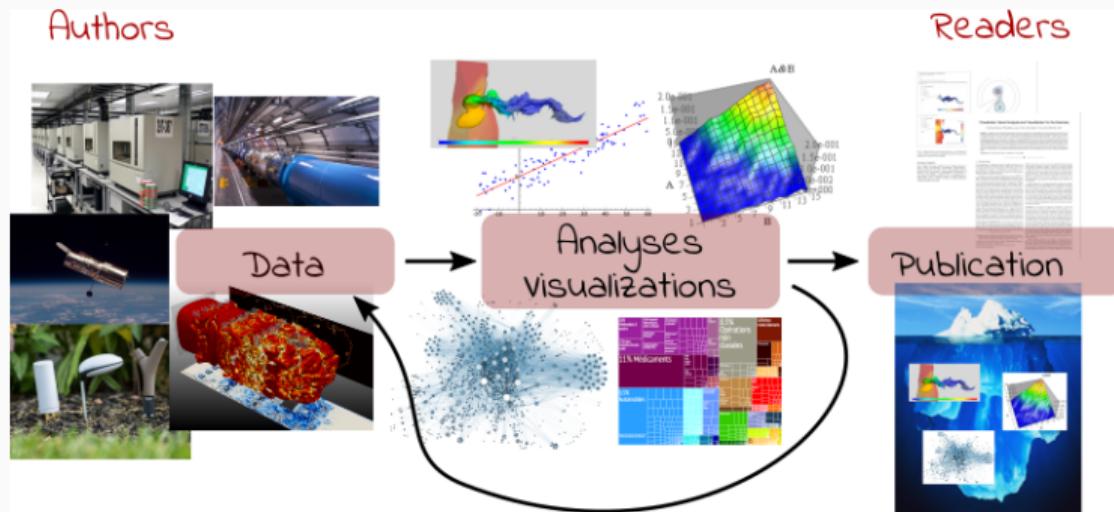
Modern Science

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



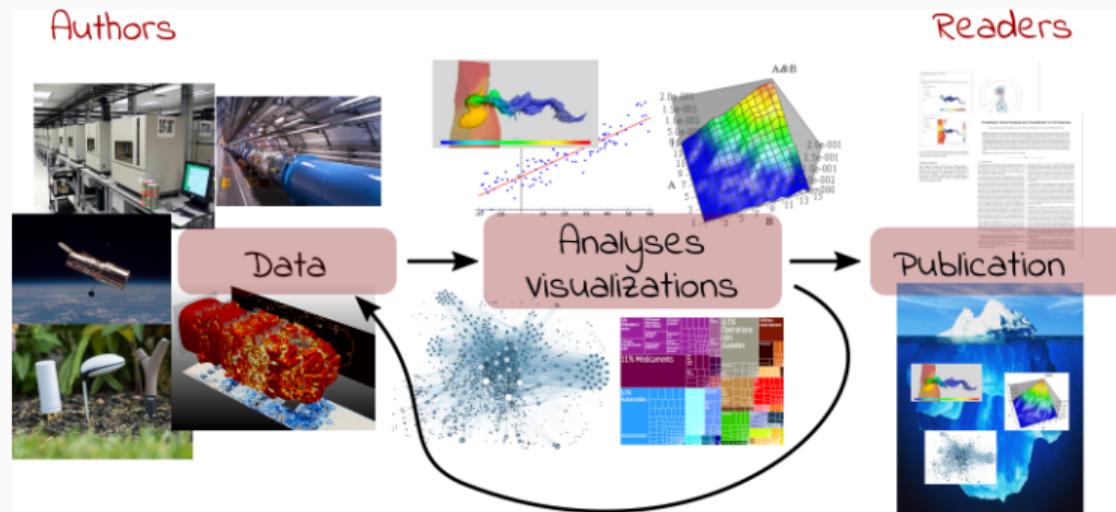
Modern Science

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



Modern Science

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

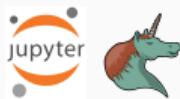
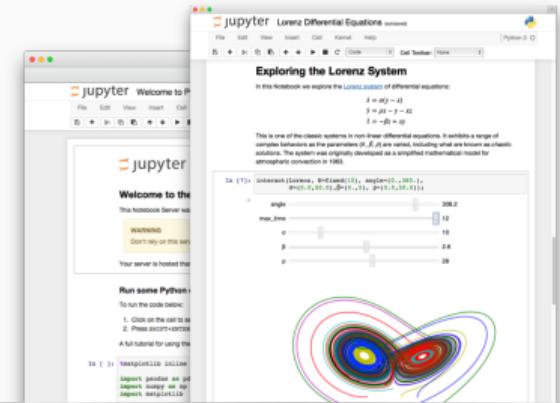


Reproducible Research = Bridging the Gap by working Transparently

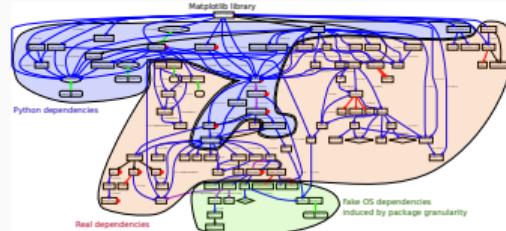
Tools

Existing Tools, Emerging Standards

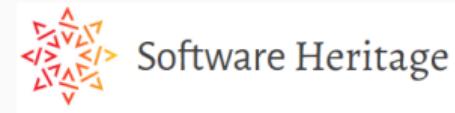
Notebooks and workflows



Software environments



Sharing platforms



Tool 1: Computational Notebooks/Literate Programming

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

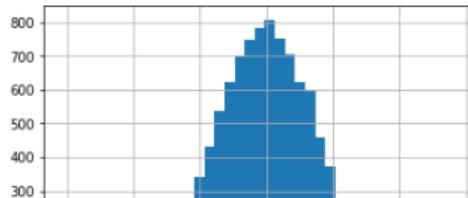
3.141592653589793

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/Literate Programming

Document initial dans son environnement

jupyter example_pi (automated)

File Edit View Insert Cell Kernel Widgets Help Hide Code Python 3

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

In [1]:

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

Mais calculé avec la [méthode des aiguilles de Buffon](#) (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtiendrait comme approximation :

In [2]:

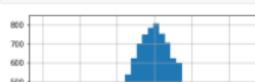
```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2*(sum((x==np.sin(theta))>1))/N
```

Out[2]: 3.1437198694098765

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

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()
```



Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

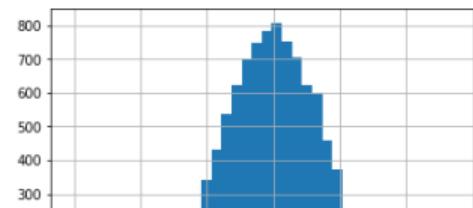
Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtiendrait comme approximation :

```
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 $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et

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



Tool 1: Computational Notebooks/Literate Programming

Document initial dans son environnement

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

- In [1]:** A Markdown cell containing "# Un document computationnel". It also contains a block of text: "Mon ordinateur m'indique que \$\pi\$ vaut *approximativement*". Below this is a code cell:

```
from math import *
print(pi)
3.141592653589793
```
- In [2]:** A code cell:

```
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
```

Its output is: **Out[2]:** 3.1437198694098765
- In [3]:** A code cell:

```
%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()
```

Its output is a histogram plot.

MarkDown

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

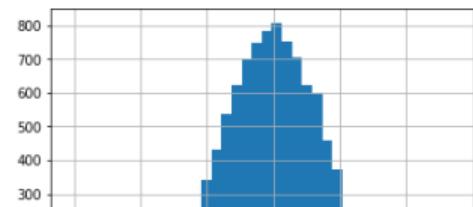
Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtiendrait comme approximation :

```
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/Literate Programming

Document initial dans son environnement

jupyter example_pi (advanced)

File Edit View Insert Cell Kernel Widgets Help Hide Code Python 3

Un document computationnel

Mon ordinateur m'indique que \$\pi\$ vaut *approximativement*

In [1]:

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

Mais calculé avec la méthode des aiguilles de Buffon (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtiendrait 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)
```

Out[2]: 3.1437198694098765

On peut inclure des formules mathématiques comme $\sqrt{1/(2\pi)} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$ et des dessins qui n'ont rien à voir avec π (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()
```



Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

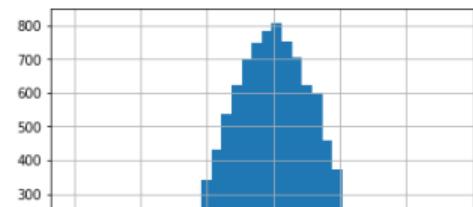
Mais calculé avec la méthode des aiguilles de Buffon, on obtiendrait comme approximation :

```
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/Literate Programming

Document initial dans son environnement

jupyter example_pi (advanced)

File Edit View Insert Cell Kernel Widgets Help Hide Code Python 3

Un document computationnel

Mon ordinateur m'indique que \$\pi\$ vaut *approximativement*

In [1]:

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

Mais calculé avec la méthode des [aiguilles de Buffon] (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtiendrait 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)
```

Out[2]: 3.1437198694098765

On peut inclure des formules mathématiques comme $\sqrt{1/(2\pi)} \exp(-x^2/(2\sigma^2))$ et des dessins qui n'ont rien à voir avec π (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()
```



Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

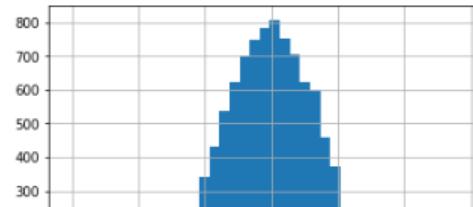
Mais calculé avec la méthode des [aiguilles de Buffon](#), on obtiendrait comme approximation :

```
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/Literate Programming

Document initial dans son environnement



Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

```
In [1]:  
from math import *  
print(pi)  
3.141592653589793
```

Mais calculé avec la méthode des aiguilles de Buffon (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtiendrait 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)
```

Out[2]: 3.1437198694098765

On peut inclure des formules mathématiques comme $\sqrt{1/(2\pi)} \exp(-x^2/2)$ et des dessins qui n'ont rien à voir avec π (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()
```



Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

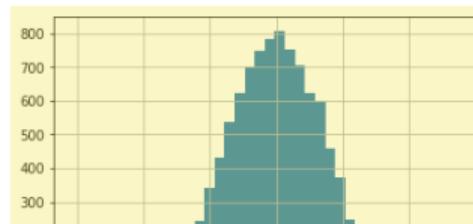
Mais calculé avec la méthode des aiguilles de Buffon, on obtiendrait comme approximation :

```
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...).



Export

Tool 1: Computational Notebooks/Literate Programming

Document initial dans son environnement



Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

```
In [1]:  
from math import *  
print(pi)  
3.141592653589793
```

Mais calculé avec la méthode des aiguilles de Buffon (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtiendrait 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)
```

Out[2]: 3.1437198694098765

On peut inclure des formules mathématiques comme $\sqrt{1/(2\pi)} \exp(-x^2/2)$ et des dessins qui n'ont rien à voir avec π (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()
```



Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

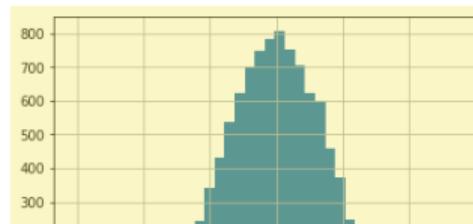
Mais calculé avec la méthode des aiguilles de Buffon, on obtiendrait comme approximation :

```
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...).



Export

Tool 2: Fighting Software Environments Nightmare

What is hiding behind a simple

```
import matplotlib
```

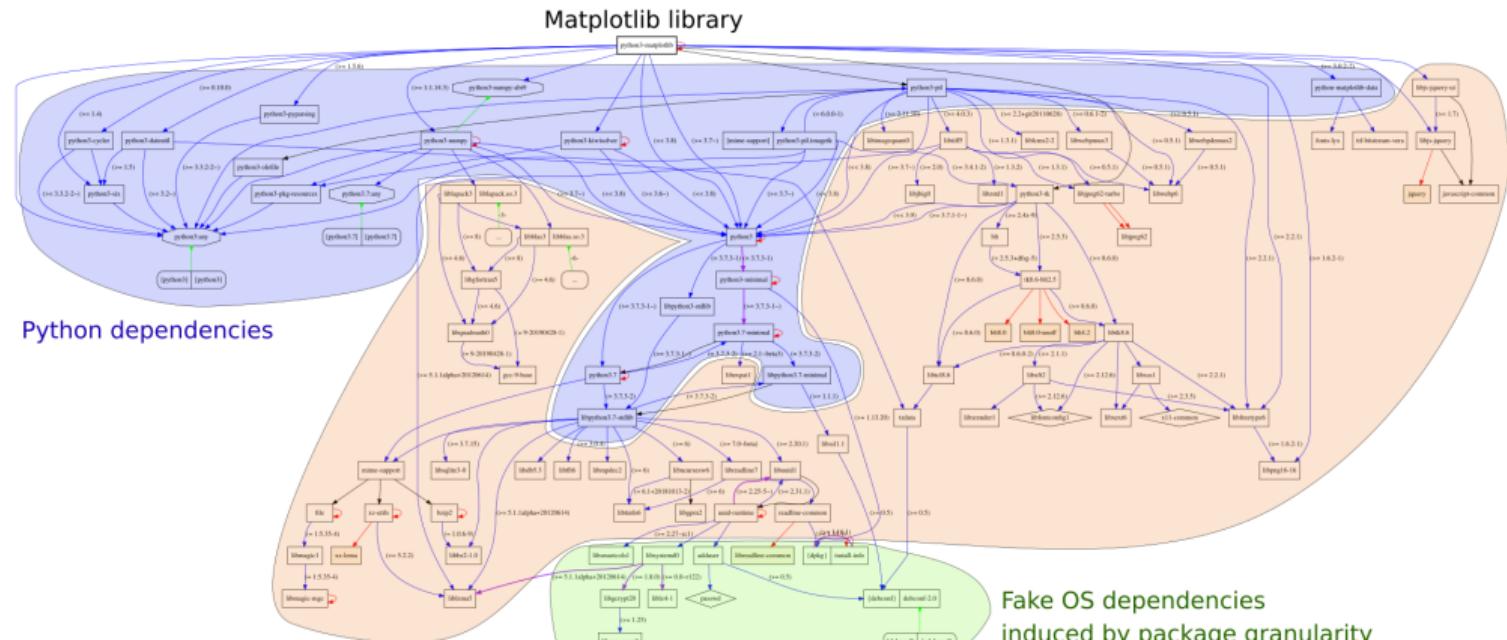
Package: python3-matplotlib
Version: 2.1.1-2
Depends: python3-dateutil, python-matplotlib-data (>= 2.1.1-2),
python3-pyparsing (>= 1.5.6), python3-six (>= 1.10), python3-tz,
libjs-jquery, libjs-jquery-ui, python3-numpy (>= 1:1.13.1),
python3-numpy-abi9, python3 (<< 3.7), python3 (>= 3.6~),
python3-cycler (>= 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)

Tool 2: Fighting Software Environments Nightmare

What is hiding behind a simple

```
import matplotlib
```

Package: python3-matplotlib



Tool 2: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

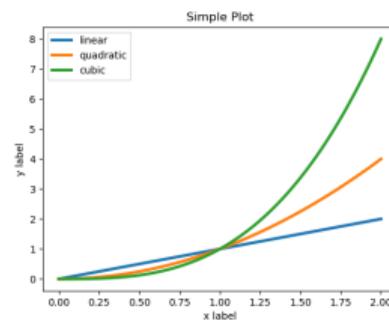
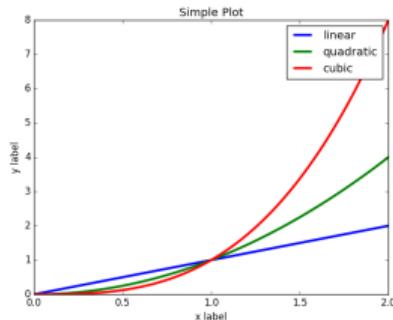
```
python2 -c "print(10/3)"  
python3 -c "print(10/3)"
```

```
3  
3.333333333333335
```

Tool 2: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

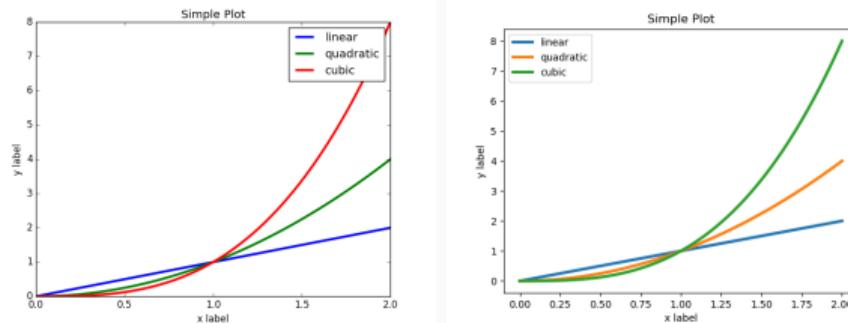
```
python2 -c "print(10/3)"  
python3 -c "print(10/3)"
```



Tool 2: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

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

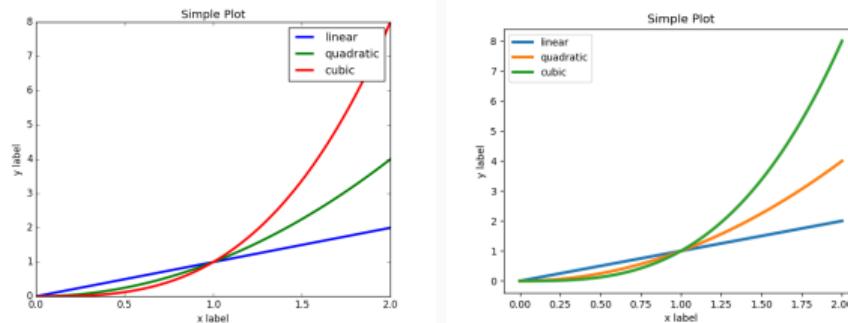


Cortical Thickness Measurements (PLOS ONE, June 2012): *FreeSurfer: differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.*

Tool 2: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

```
python2 -c "print(10/3)"  
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.*



Tool 3: 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

Tool 3: 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



Reproducibility

Different (but Converging) Reproducibility Concerns

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, . . . ?

Different (but Converging) Reproducibility Concerns

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, . . . ?

Bad computer/statistic/publication practices "broke science" 😊

- Ensure articles and data are **available**
 - Open Access, Open Data, FAIR and DMP



Different (but Converging) Reproducibility Concerns

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, . . . ?

Bad computer/statistic/publication practices "broke science" 😊

- Ensure articles and data are **available**
 - Open Access, Open Data, FAIR and DMP
- *Ensure the experiment design and analysis is not broken*
 - ~~data torture and hacking, pre-registration, education, . . .~~



Different (but Converging) Reproducibility Concerns

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, . . . ?

Bad computer/statistic/publication practices "broke science" 😊

- Ensure articles and data are **available**
 - Open Access, Open Data, FAIR and DMP
- *Ensure the experiment design and analysis is not broken*
 - ~~data torture and hacking, pre-registration, education, . . .~~
- Explain **how** the computations unfold and **why**
 - Notebooks (that can quickly get out of control)
 - Simple computational **workflows**



Different (but Converging) Reproducibility Concerns

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, . . . ?

Bad computer/statistic/publication practices "broke science" 😊

- Ensure articles and data are **available**
 - Open Access, Open Data, FAIR and DMP
- *Ensure the experiment design and analysis is not broken*
 - ~~data torture and hacking, pre-registration, education, . . .~~
- Explain **how** the computations unfold and **why**
 - Notebooks (that can quickly get out of control)
 - Simple computational **workflows**
- Allow to inspect, rerun, make variations on computations
 - Access to code, data, . . .



Different (but Converging) Reproducibility Concerns

Reproducibility/robustness of the scientific fact, the statistical analysis, the computation, the observation, the process, . . . ?

Bad computer/statistic/publication practices "broke science" 😊

- Ensure articles and data are **available**
 - Open Access, Open Data, FAIR and DMP
- *Ensure the experiment design and analysis is not broken*
 - ~~data torture and hacking, pre-registration, education, . . .~~
- Explain **how** the computations unfold and **why**
 - Notebooks (that can quickly get out of control)
 - Simple computational **workflows**
- Allow to inspect, rerun, make variations on computations
 - Access to code, data, . . . options/parameters, environment, resources?



This requires **first class software engineering practices** instead of building on prototypes
Software factories, Archives, and Provenance Tracking tools

Closure

Changing Research Practices

Soft. Engineering, Statistics, and Reproducible Research in the **curricula**



- Book on RR *Vers une recherche reproduitble: Faire évoluer ses pratiques*
- MOOC on RR (3rd edition Feb. 2020)
- A new "Advanced RR" MOOC (Oct. 2020)
 - Software environment control (Docker)
 - Scientific workflow (snakemake)
 - Managing data (HDF5, archiving)

Manifesto: "*I solemnly pledge*" (WSSSPE, Lorena Barba, FAIR)

1. I will teach my graduate students about reproducibility
2. All our research code (and writing) is under version control
3. We will always carry out verification and validation
4. We will share data, plotting script & figure under CC-BY
5. We will upload the preprint to arXiv at the time of submission of a paper

Changing Publishing Practices

Artifact evaluation and ACM badges



Major conferences

- Supercomputing: Artifact Description (AD) **mandatory**, Artifact Evaluation (AE) still optional, Double blind vs. RR
- NeurIPS, ICLR: open reviews, reproducibility challenge



Joelle Pineau @ NeurIPS'18

- ACM SIGMOD 2015-2019, Most Reproducible Paper Award...

Mentalities are evolving people care, make stuff available, errors are found and fixed

Reproducible Research = Rigor and Transparency

To err is human.

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
- 2. Change the norm:** make publication practices evolve
- 3. Incentive:** consider RR/open science when hiring/promoting



Resources and Acknowledgments

Resources and Acknowledgments

Slides are available here

https://github.com/schnorr/SMPE/raw/master/lectures/talk_20_11_17_WPerformance.pdf

Thanks to

- Arnaud Legrand for sharing slides and knowledge about this topic
- My laboratory students that have understood the importance
 - It may be hard to face/learn new things, but they are essential to evolve!

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

- Konrad Hinsen, Christophe Pouzat, Alexandre Hocquet
- 3rd Edition: March 2020 – March 2021
- MOOC RR "Advanced" planned for 2021

