

LABORATORY NOTEBOOK, COMPUTATIONAL DOCUMENT, REPRODUCIBLE ARTICLE EMACS/ORG-MODE: ONE RING TO RULE THEM ALL?

Arnaud Legrand



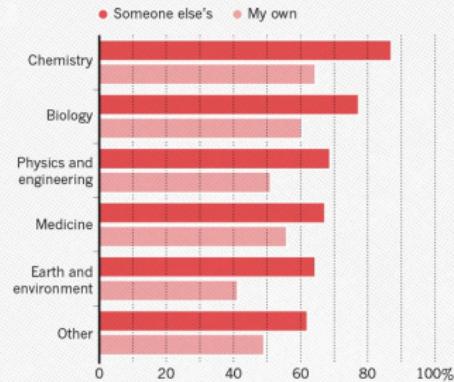
MaiMosine, GRICAD, SARI network, June 2023



REPRODUCIBILITY "CRISIS": SOCIO-TECHNICAL CHALLENGES

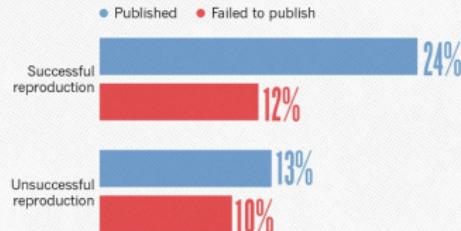
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.



1,500 scientists lift the lid on reproducibility,

Nature, May 2016

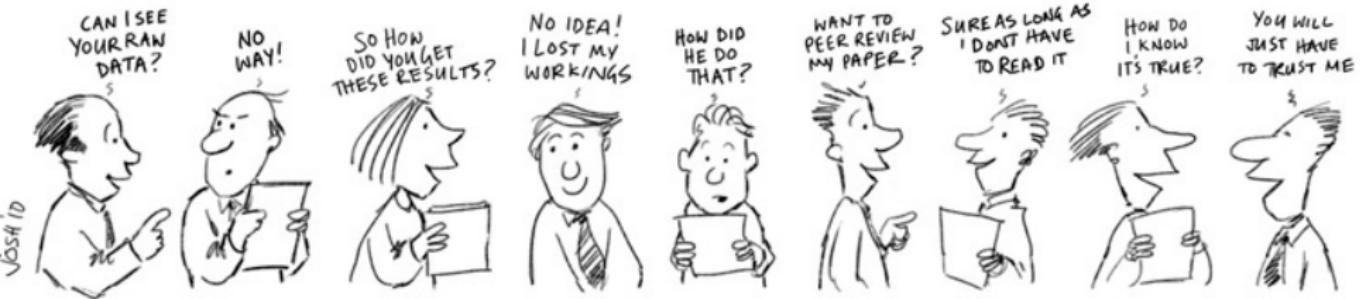
Social causes

- Fraud, conflict of interest (pharmaceutic, ...)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1M+ articles per year!
- Emerging practices: DORA/Plan S/COARA, DMP and FAIR data, artefact evaluation, reproducibility badges, reproducibility challenges, open reviews, ...

Methodological/technical causes

- The many biases (apophenia, confirmation, hindsight, experimenter, ...): bad designs
- Selective reporting, weak analysis (statistics, data manipulation mistakes, computational errors)
- Lack of information, code/raw data unavailable

NO TRANSPARENCY NO CONSENSUS



DIFFERENT REPRODUCIBILITY CONCERNS IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

DIFFERENT REPRODUCIBILITY CONCERNS IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical chaos, parallel architectures

DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

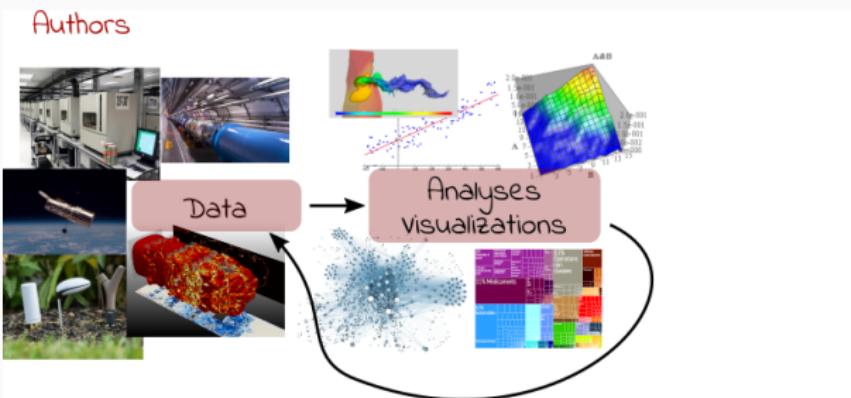
Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical chaos, parallel architectures

Artificial Intelligence most of the above 😊

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



DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

Biology, Oncology sample provenance, clinical trials \rightsquigarrow standardized protocols

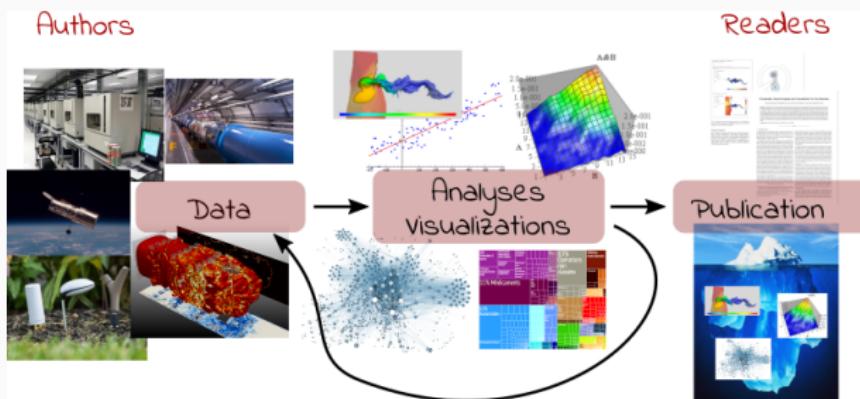
Psychology, Nutrition HARKING, p-hacking \rightsquigarrow pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical chaos, parallel architectures

Artificial Intelligence most of the above 😊

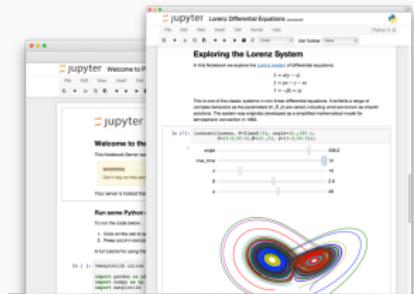
The processing steps between raw observations and findings have gotten increasingly numerous and complex



Reproducible Research = Bridging the Gap by working Transparently

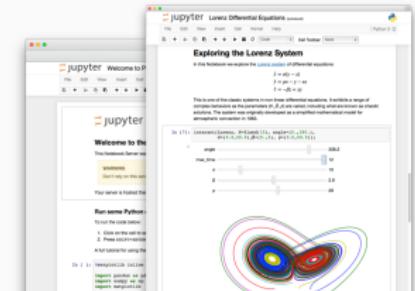
REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

Computation provenance: notebooks and workflows

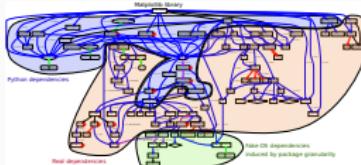


REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

Computation provenance: notebooks and workflows

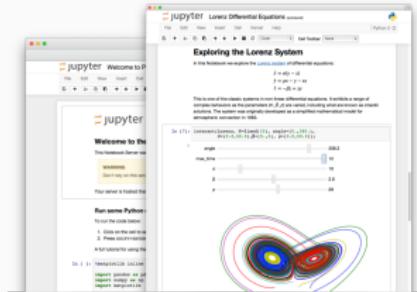


Software environments



REPRODUCIBILITY ISSUES RELATED TO THE USE OF COMPUTERS

Computation provenance: notebooks and workflows



Software environments



Sharing and Archiving



GOOD PRACTICE #1

TAKING NOTES AND DOCUMENTING

What your research supposedly looks like:

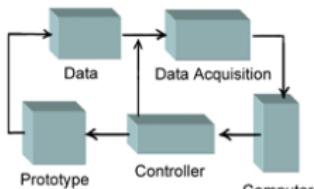


Figure 1. Experimental Diagram

What your research *actually* looks like:

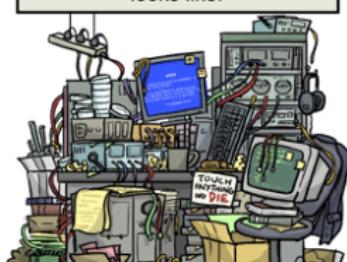


Figure 2. Experimental Mess

FRUSTRATION AS AN AUTHOR/REVIEWER



Author (*Calls for a Journal*)

- I thought I used the same parameters but **I'm getting different results!**
- The new student wants to compare with **the method I proposed last year**
- My advisor asked me whether I took care of setting this or this but **I can't remember**
- The damned fourth reviewer asked for a major revision and wants me to change Figure 3. **Which code and which data set did I use?**
- **It worked yesterday!** 6 months later: Why did I do that?



Author (Calls for a *Journal*)

- I thought I used the same parameters but **I'm getting different results!**
- The new student wants to compare with **the method I proposed last year**
- My advisor asked me whether I took care of setting this or this but **I can't remember**
- The damned fourth reviewer asked for a major revision and wants me to change **Figure 3**. Which code and which data set did I use?
- **It worked yesterday!** 6 months later: Why did I do that?

Reviewer (Calls for a *Reproducible Article*)

- As usual, there is **no confidence interval**, I wonder about the variability and whether the difference is **significant** or not
- That can't be true, I'm sure **they removed some points**
- Why is this graph in logscale? **How would it look like otherwise?** I'm not even **sure** of what this value means. If only I could access the generation script

TOOL 1: COMPUTATIONAL NOTEBOOKS/LITERATE PROGRAMMING

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

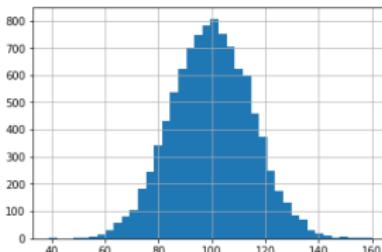
3.141592653589793

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

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

- In [1]:** `# Un document computationnel`
Output: Mais ordinateur m'indique que π vaut "approximativement"
- In [1]:** `from math import *`
Output: 3.141592653589793
- In [2]:** `Mais calculé avec la méthode des aiguilles de Buffon (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtient d'autant 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=np.pi/2)
2/sum((x+np.sin(theta))>1)/N`
Output: 3.14371986944998765
- In [3]:** `On peut inclure des formules mathématiques comme $\sqrt{2/\pi} \exp(-x^2/2)$ 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()`
Output: A histogram showing a bell-shaped distribution centered at 100, with x-axis ranging from 40 to 160 and y-axis ranging from 0 to 800.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

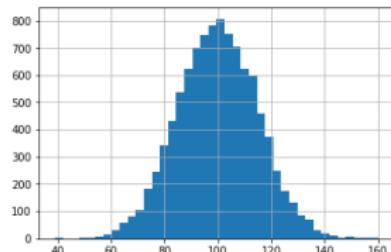
Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtiendrait comme approximation :

```
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=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/LITERATE PROGRAMMING

Document initial dans son environnement

Jupyter example_pl (modified)

Un document computationnel

Mais 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 obtient aussi 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=np.pi/2)
2*(sum((x+np.sin(theta))>1))/N
```

```
Out[2]: 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... ☺).

```
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

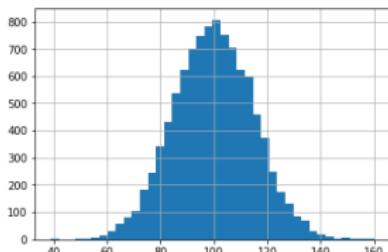
3.141592653589793

Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtient 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/LITERATE PROGRAMMING

Document initial dans son environnement

A screenshot of a Jupyter Notebook interface. The top bar shows 'jupyter example_pi' and 'Python 3'. The notebook has three cells:

- In [1]:** `# Un document computationnel`. The output shows the value of pi as 3,141592653589793.
- In [2]:** A code cell with imports for numpy and random, calculating pi using Buffon's needle method. The output shows the result 3.1437198694098765.
- In [3]:** A code cell with imports for numpy, matplotlib, and random, creating a histogram of 100,000 random numbers. The output is a histogram plot centered around 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

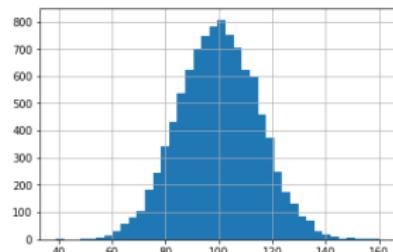
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
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/LITERATE PROGRAMMING

Document initial dans son environnement

Un document computationnel

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

Mais calculé avec la [méthode des aiguilles 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 $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).

```
In [3]:  
%matplotlib inline  
import matplotlib.pyplot as plt  
  
mu, sigma = 100, 15  
x = mu + sigma*np.random.randn(10000)  
  
plt.hist(x, 99)  
plt.grid(True)  
plt.show()
```

Résultats

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

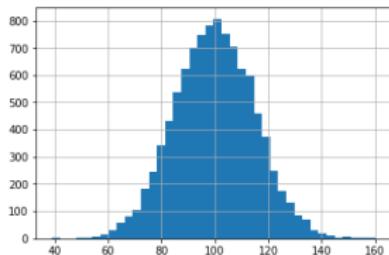
Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtiendrait comme approximation :

```
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

```
In [1]:  
from math import *  
print(pi)  
3,141592653589793
```

Mais calculé avec la `_methodes_ des éimpulles de Buffon` (https://fr.wikipedia.org/wiki/Algille_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 $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des dessins qui n'ont rien à voir avec π (si ce n'est une constante de normalisation... ☺).

```
In [3]:  
%matplotlib inline  
import matplotlib.pyplot as plt  
  
mu, sigma = 100, 15  
x = mu + sigma*np.random.randn(10000)  
  
plt.hist(x, 99)  
plt.grid(True)  
plt.show()
```

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

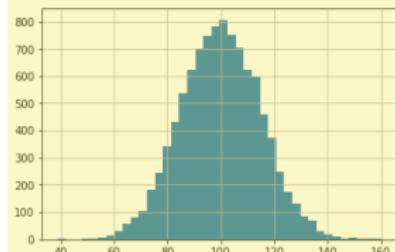
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
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

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/LITERATE 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 using 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 to approximate pi.
- In [3]:** Plots a histogram of random numbers between 0 and 100, showing a bell-shaped distribution centered around 50.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

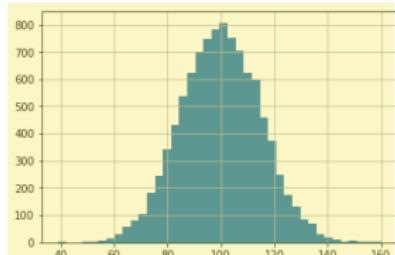
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
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... ☺).



SEVERAL POSSIBLE USAGES

Code Documentation (\neq code comments)

- For developers: explain code organization (data structures, algorithms, modules, class, etc.) and how to contribute
- For users: API, examples, installation, ...

SEVERAL POSSIBLE USAGES

Code Documentation (\neq code comments)

- For developers: explain code organization (data structures, algorithms, modules, class, etc.) and how to contribute
- For users: API, examples, installation, ...

Data Analysis data dependent

- Data transformation/curation
- Exploratory/sequential data analysis
- Document hypothesis, graphs/tables

SEVERAL POSSIBLE USAGES

Code Documentation (\neq code comments)

- For developers: explain code organization (data structures, algorithms, modules, class, etc.) and how to contribute
- For users: API, examples, installation, ...

Data Analysis data dependent

- Data transformation/curation
- Exploratory/sequential data analysis
- Document hypothesis, graphs/tables

Reproducible article (book with code/computations)

- Classical article/book structure with typography constraints
- Code is behind the scene (e.g., figures, tables, numbers)

SEVERAL POSSIBLE USAGES

Code Documentation (\neq code comments)

- For developers: explain code organization (data structures, algorithms, modules, class, etc.) and how to contribute
- For users: API, examples, installation, ...

Data Analysis data dependent

- Data transformation/curation
- Exploratory/sequential data analysis
- Document hypothesis, graphs/tables

Reproducible article (book with code/computations)

- Classical article/book structure with typography constraints
- Code is behind the scene (e.g., figures, tables, numbers)

"Laboratory" notebook Research reporting. Daily notes on

- Experimental parameters, specific configurations
- Meetings, seminars, lectures, ...
- Readings, bibliography
- TODOs, Ideas, Random hacks

SEVERAL POSSIBLE USAGES

Code Documentation (\neq code comments)

- For developers: explain code organization (data structures, algorithms, modules, class, etc.) and how to contribute
- For users: API, examples, installation, ...

Data Analysis data dependent

- Data transformation/curation
- Exploratory/sequential data analysis
- Document hypothesis, graphs/tables

Reproducible article (book with code/computations)

- Classical article/book structure with typography constraints
- Code is behind the scene (e.g., figures, tables, numbers)

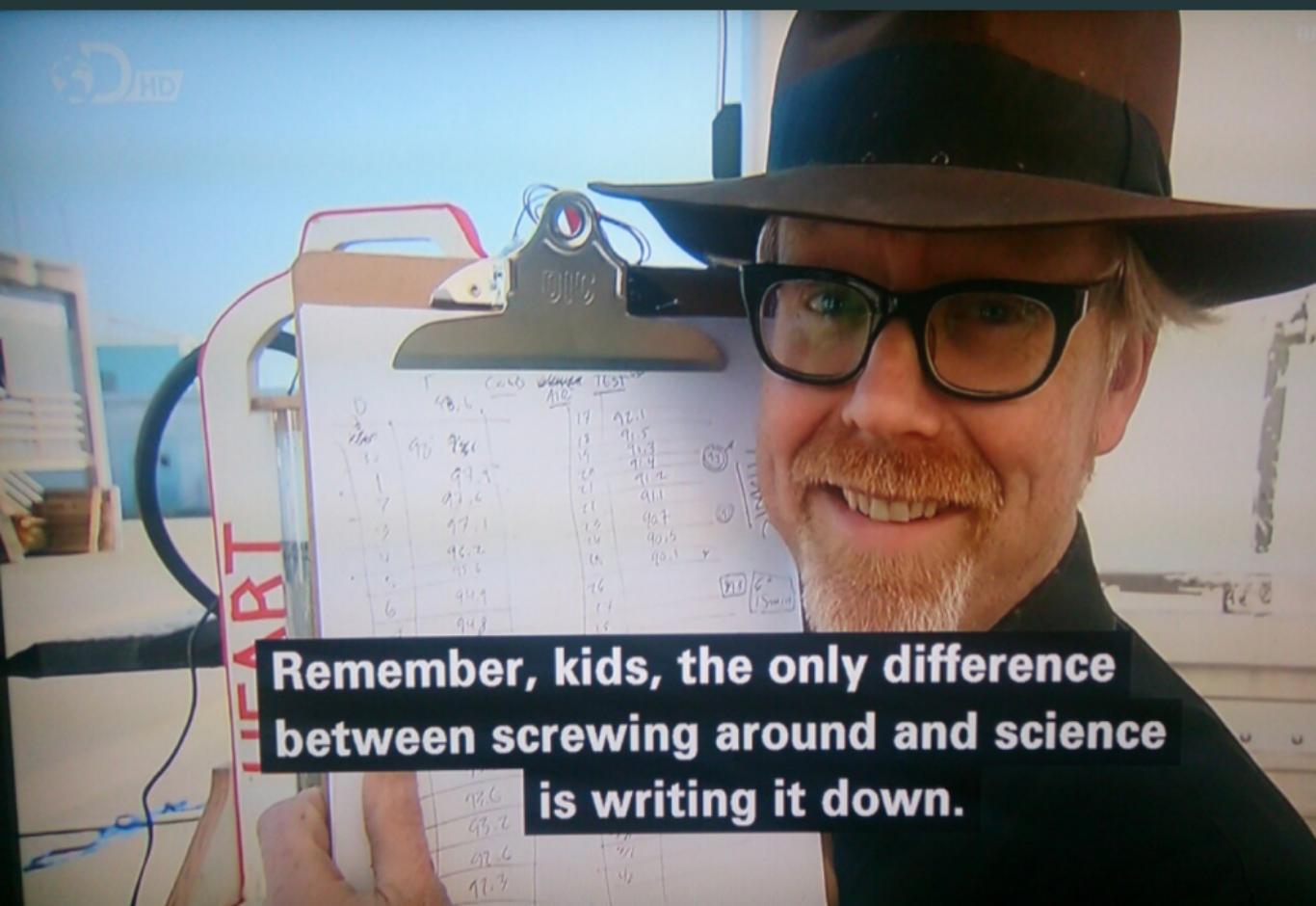
"Laboratory" notebook Research reporting. Daily notes on

- Experimental parameters, specific configurations
- Meetings, seminars, lectures, ...
- Readings, bibliography
- TODOs, Ideas, Random hacks

Teaching material, Presentations

Dynamic documents, Websites, ...

LABORATORY NOTEBOOKS, COMPUTATIONAL DOCUMENTS



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

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 <https://quarto.org/>)

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 <https://quarto.org/>)



Studio[®] (java/R)

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

- Cons**
- Installation, software dependencies (`rocker/rstudio` \approx 550Mb)
 - Limited control on typography (unless using monsters 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 <https://quarto.org/>)



(java/R)

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

- Cons**
- Installation, software dependencies (`rocker/rstudio` \approx 550Mb)
 - Limited control on typography (unless using monsters like `quarto`)



(emacs-lisp)

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

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

Cons

- No default configuration \leadsto rough and steep learning curve
- Big machinery: `silex/emacs-alpine-ci` \approx 240Mb even though `flycheck/emacs-cask` \approx 80Mb

WHICH TOOL FOR WHICH USAGE ?

A biased opinion

	Jupyter
Coding	
Data Analysis	
Articles	
Lab Notebook	
Slides	
Dynamic docs	
Websites	

WHICH TOOL FOR WHICH USAGE ?

A biased opinion

	Jupyter	Rstudio
Coding	😢	😊 (R)
Data Analysis	😎	😎
Articles	😢	😊
Lab Notebook	😢	😐
Slides	😐	😊
Dynamic docs	😎	😎
Websites		

WHICH TOOL FOR WHICH USAGE ?

A biased opinion

	Jupyter	Rstudio	Org-mode
Coding	:(:((R)	:(
Data Analysis	:)	:)	:)
Articles	:(:)	:(
Lab Notebook	:(:)	:(
Slides	:	:)	:)
Dynamic docs	:)	:)	:(
Websites			

PANDORA'S BOX



COMMON ISSUES

1. Format evolution through time (increasing complexity)

COMMON ISSUES

1. Format evolution through time (increasing complexity)
2. Environment control. Uuh ???



COMMON ISSUES

1. Format evolution through time (increasing complexity)
2. Environment control. Uuh ???



3. Scalability

Simplifying Color Names by Web Image Processing

This project aims to find the most suitable colors for the selected colors in a color palette. It uses a neural network to predict the most suitable colors based on the input colors.

The main idea is to use a neural network to predict the most suitable colors for a given set of colors. This is achieved by training the neural network on a large dataset of color palettes and their corresponding suitable colors. The neural network then takes a set of colors as input and outputs the most suitable colors based on the learned patterns.

The neural network architecture consists of three layers: an input layer, a hidden layer, and an output layer. The input layer takes the input colors as input and processes them. The hidden layer performs the actual learning and prediction. The output layer provides the final suitable colors.

The neural network is trained on a large dataset of color palettes and their corresponding suitable colors. The dataset is used to train the neural network to learn the patterns and relationships between the input colors and the output suitable colors.

The neural network is then used to predict the most suitable colors for a given set of colors. The predicted colors are then used to generate a simplified color palette.

Chemistry plane and Uniref50 model results

The following figure shows the chemistry plane and the results of the Uniref50 model. The plane is a 2D grid where each point represents a specific chemical structure. The colors represent the results of the Uniref50 model, showing which structures are more likely to be found in the training data.

Prediction error vs. Training sample variance

The following figure shows the prediction error versus the training sample variance. The plot shows a strong negative correlation, indicating that models with more training samples tend to have lower prediction errors.

Analysis

The following figure shows the analysis of the training data. The plot shows the distribution of training data points across different categories, such as protein families and subfamilies.

Scaling the model

The following figure shows the scaling of the model. The plot shows the relationship between the number of training samples and the prediction error. The error decreases as the number of training samples increases, indicating that the model is learning better.

Training the model

The following figure shows the training of the model. The plot shows the progress of the training process, including the loss function and the accuracy of the model over time.

Conclusion

The following figure summarizes the conclusions of the study. The figure highlights the importance of having a large and diverse training dataset for accurate predictions. It also emphasizes the need for careful model selection and validation.

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
- 3rd Edition: March 2020 – March 2023 (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
- 3rd Edition: March 2020 – March 2023 (15,000+)



MOOC "Advanced RR" planned for Nov. 2023

- 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
- 3rd Edition: March 2020 – March 2023 (15,000+)



MOOC "Advanced RR" planned for Nov. 2023

- 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
- 3rd Edition: March 2020 – March 2023 (15,000+)



MOOC "Advanced RR" planned for Nov. 2023

- 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

4. Prepare the Future:

- Reuse, reuse, reuse!
- Toward **literate experimentation?**
- Shared and controlled testbeds
- How to share Experiments/Simulations ?



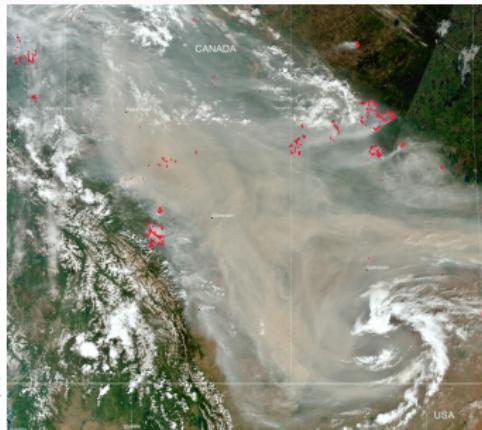
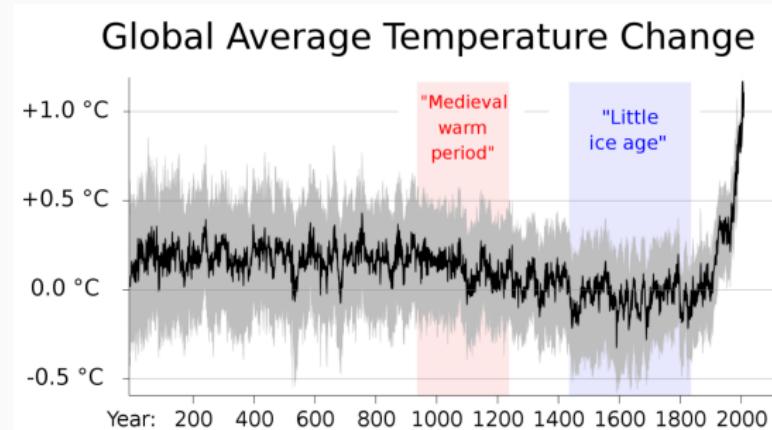
THE SCIENCE IS CLEAR

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?

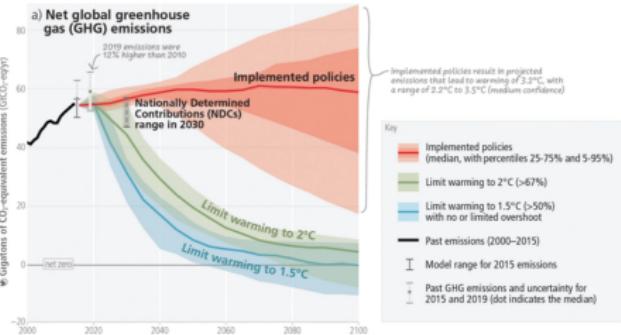


IPCC, IPBES, <https://climate.nasa.gov/>

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

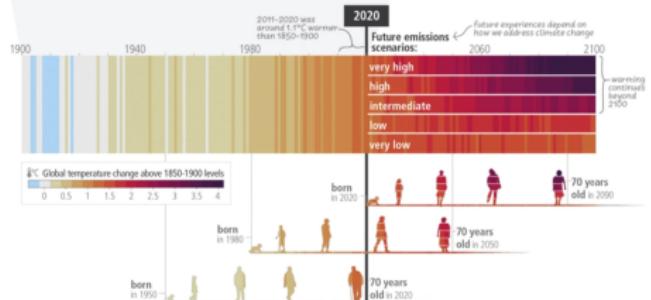
Limiting warming to **1.5°C** and **2°C** involves rapid, deep and in most cases immediate greenhouse gas emission reductions

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

14/15

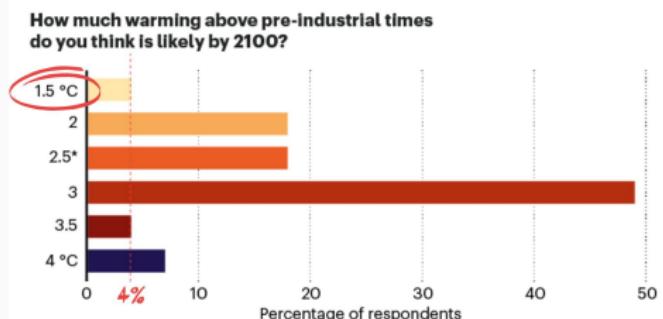
THE SCIENCE IS CLEAR

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**

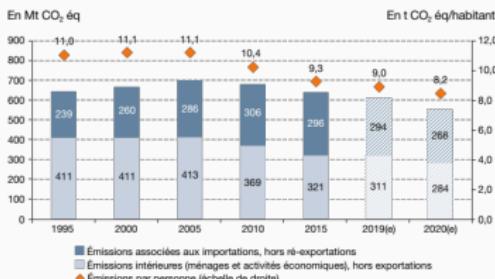


@natu Nature survey, Nov. 2021

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

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

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE



(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 (métropole 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.



$\div 2$
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 en forêt, 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 la maison, préférence à ses produits régionaux

0,5 t CO₂e/Annee : faire une partie route (300km) de votre voiture ancienne sur 30 ans, risquez de faire un peu moins de km dans vos transports en commun.

0,5 t CO₂e/Annee : faire du vélo, prendre le bus ou le métro, faire des expériments éducatifs et informatifs, etc., sobriété dans les déplacements et les achats.

0,2 t CO₂e/Annee : échafaudage sur un rmpc en papier, 100% recyclé et d'un emballage biodégradable, faire un peu moins de km en voiture, etc.

0,2 t CO₂e/Annee : à la maison, passer à la chaleur ou solaire thermique.

0,2 t CO₂e/Annee : faire évoluer, échanger, donner, etc.

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

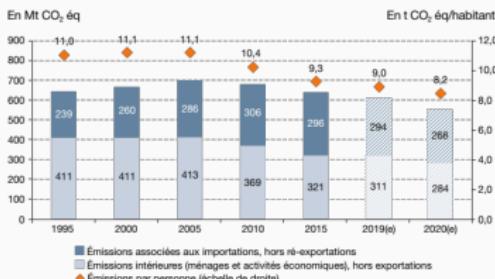
INVENTONS
NOS VIES
BAS CARBONE

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

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 en forêt, méditer, 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)

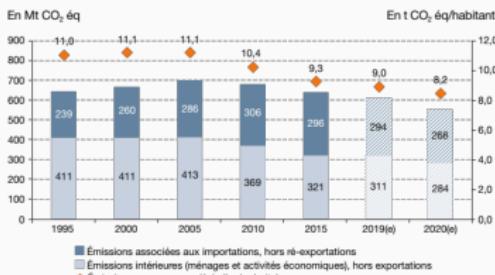
French government response

- 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

THE ELEPHANT IN THE ROOM: CLIMATE CHANGE

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

ÉVOLUTION DE L'EMPREINTE CARBONE DE LA FRANCE

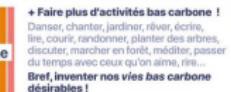


Empreinte carbone moyenne en France
10 tonnes de CO₂e/an/pers.



÷2
d'ici
2030

Objectif d'ici 2050
- de 2 t de CO₂e/an/pers.



Par exemple :



<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

Several scenarios on the table

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