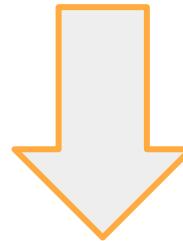


# Cours 6 : Visualisation avancée des données

Gauthier Gidel  
18 septembre 2024

# Pourquoi la visualisation de données

- Travailler avec des données complexes (e.g., de grande dimension, avec de nombreux points de données.)
- Nous voulons développer des intuitions.
- Nous sommes bons avec la vision! (notre cerveau est doué pour la reconnaissance visuelle des formes).
- Peut représenter des images complexes.



La visualisation des données places les données dans notre monde visuel (2D, 3D ou plus) pour que nous les « intuitions ».

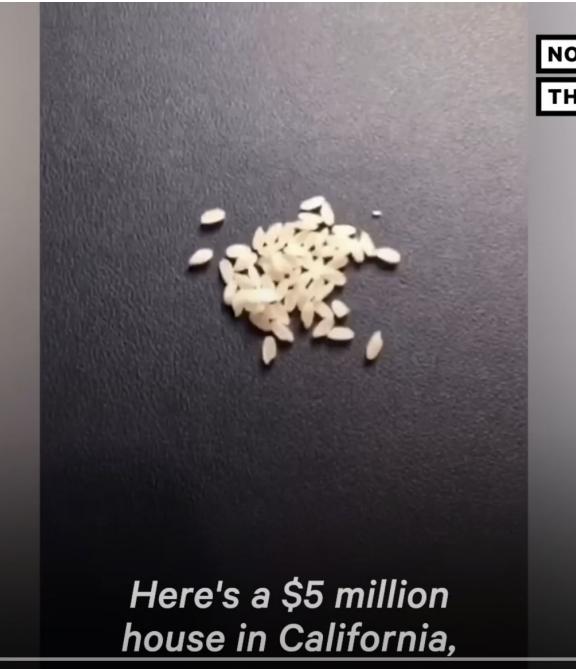
Nous ne sommes pas si bons avec les ordres de  
grandeur



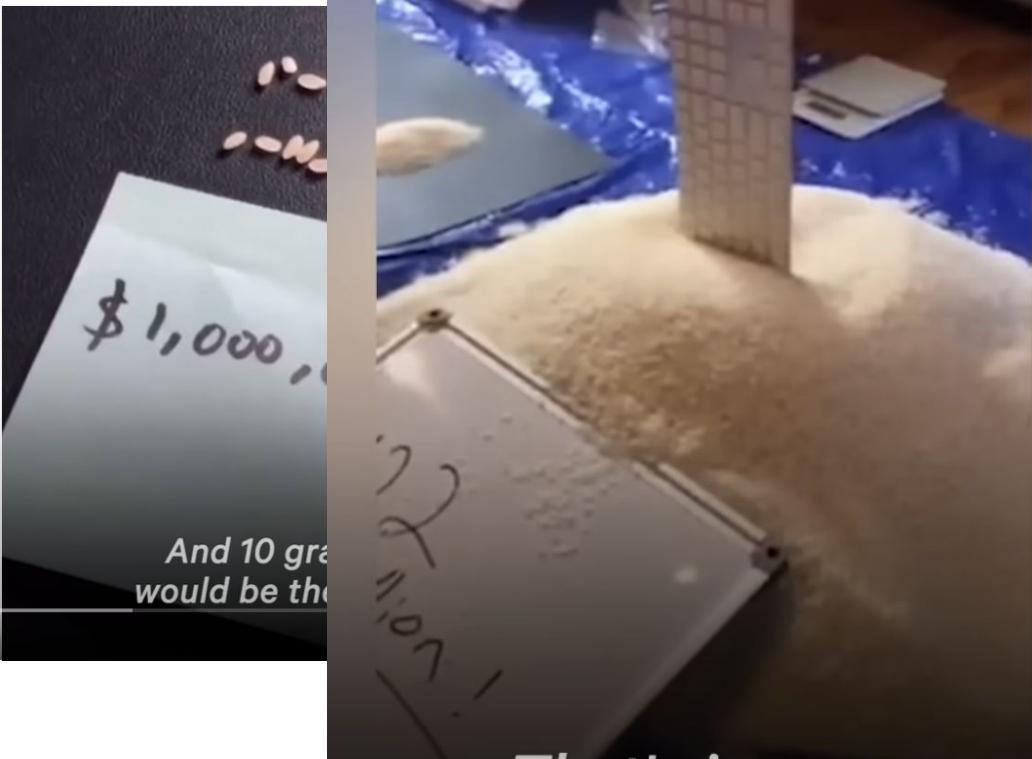
Nous ne sommes pas si bons avec les ordres de grandeur



# Nous ne sommes pas si bons avec les ordres de grandeur



Nous ne  
grandes



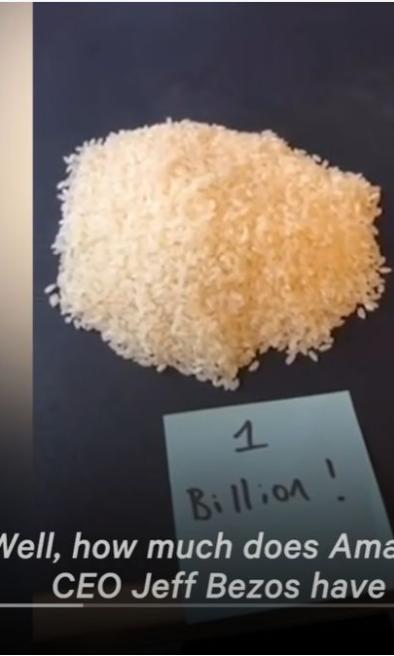
And 10 grams  
would be the

NOW THIS c les ordres de

NOW THIS

NOW THIS

'Well, how much does Amazon  
CEO Jeff Bezos have



NOW  
THIS

# Deux principaux types de visualisation

- **Exploration de données :**
  - obtenir des informations sur vos données.
  - Principalement pour soi-même.
  - Pour trouver ce que vous pouvez faire avec ces données.
  
- **Présentation des données:**
  - Présenter les résultats.
  - Convaincre les autres (et vous-même) qu'un certain résultat est vrai.
  - Plus que convaincant, il a un effet de persuasion visuel.

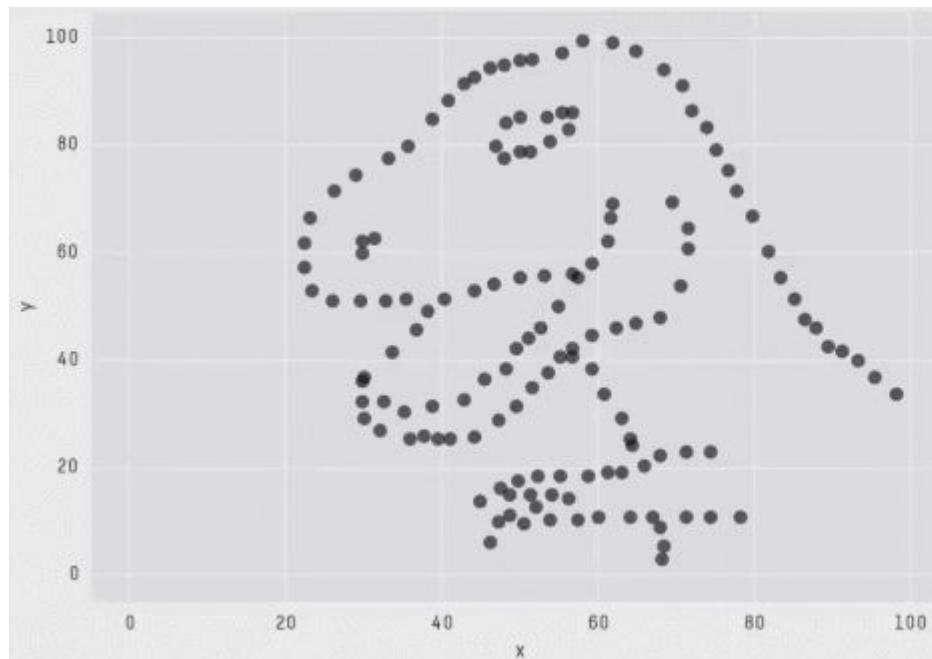
# Plan

1. Présentation (faite)
2. Exploration de données
3. Présentation des données
4. Visualisations avancées
5. Conclusion

# Exploration des données

# Visualisez toujours vos données !

- Les quantités quantitatives sont importantes. Mais ils ne décrivent pas tout !



X Mean: 54.2659224  
Y Mean: 47.8313999  
X SD : 16.7649829  
Y SD : 26.9342120  
Corr. : -0.0642526

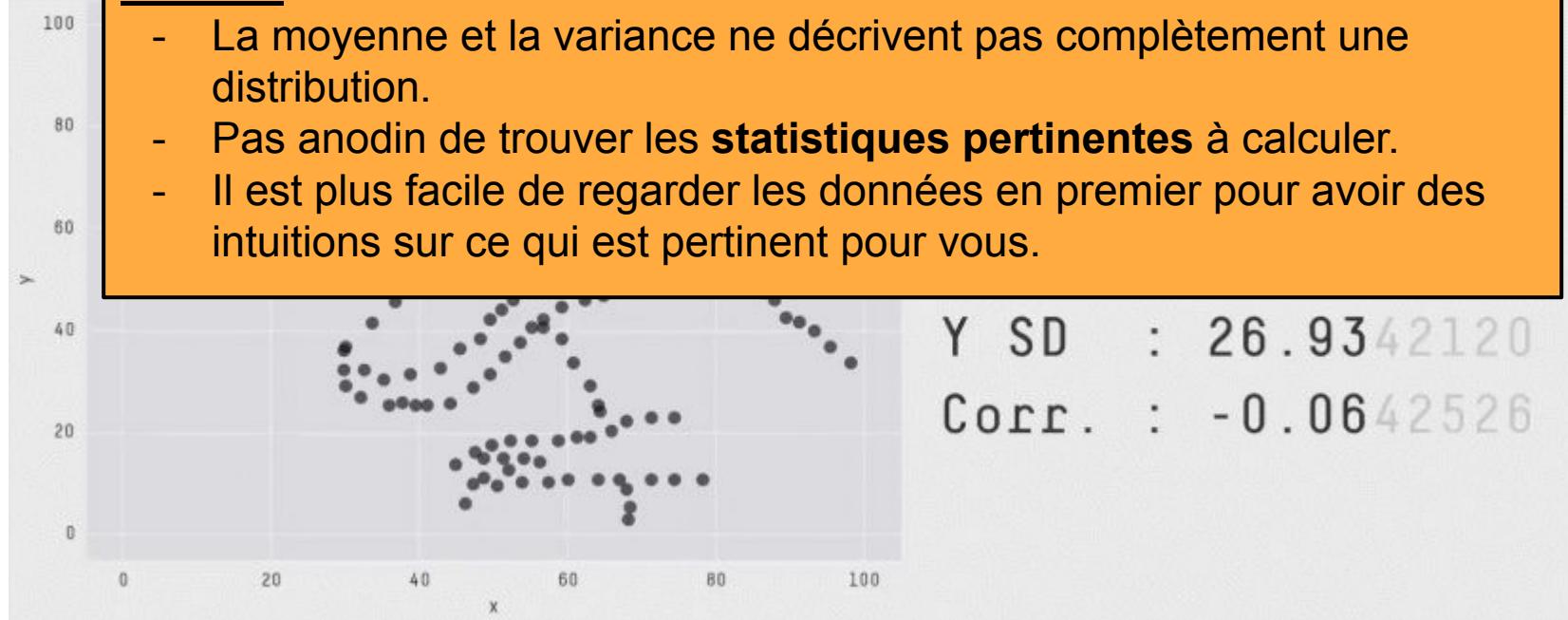
Crédit : @AlbertoCairo

# Visualisez toujours vos données !

- Les quantités quantitatives sont importantes. Mais ils ne décrivent pas tout !

## À retenir:

- La moyenne et la variance ne décrivent pas complètement une distribution.
- Pas anodin de trouver les **statistiques pertinentes** à calculer.
- Il est plus facile de regarder les données en premier pour avoir des intuitions sur ce qui est pertinent pour vous.



# Importance de la visualisation

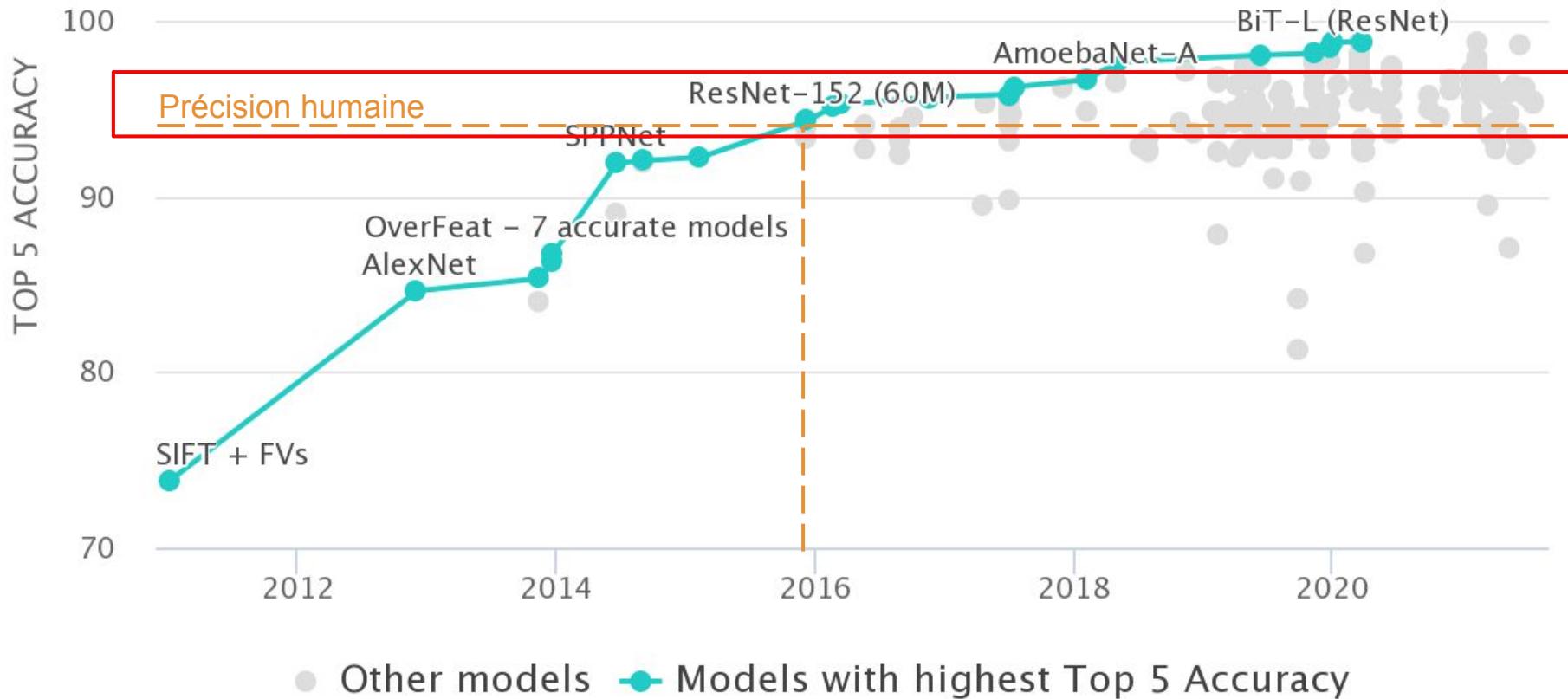
Avant d'exécuter une analyse ou d'essayer des modèles d'apprentissage automatique, vous devez visualiser vos données pour s'assurer que :

- vos données aient du sens (aucune erreur de traitement)
- le problème soit "réalisable"
- avoir un projet.

Si vous ne **pouvez pas identifier une tendance** ou faire une prédition pour votre ensemble de données, **un algorithme automatisé ne le fera pas non plus.**

obtenir des performances « surhumaines » dans un domaine est un problème difficile qui prend des années à résoudre (pour chaque problème). Des performances sous-surhumaines ont été réalisées bien avant cela.

## Résultats à la pointe de la technologie (SOTA) pour Imagenet dans le temps.



# Méthode (proposée) pour exploration de données

# Commencez par une question

- Vous devez d'abord poser une question.

Exemple:

- puis-je utiliser mes données pour prédire [quelque chose] ?
  - Mes données sont-elles propres ?
  - Comment puis-je augmenter mon jeu de données ?
  - Ai-je des valeurs aberrantes ?
- 
- Le but d'une visualisation est de répondre (ou d'en donner un aperçu) à cette question.
  - Moteur de votre recherche.

# L'interactivité est importante

"L'interaction est la clé des applications ML"

Peut **explorer plus rapidement**. (ex. google maps)

Mieux vaut construire des **intuitions** (peut jouer avec les données et voir comment elles réagissent)

# Les données sont imparfaites

Je vais répéter cela plusieurs fois.

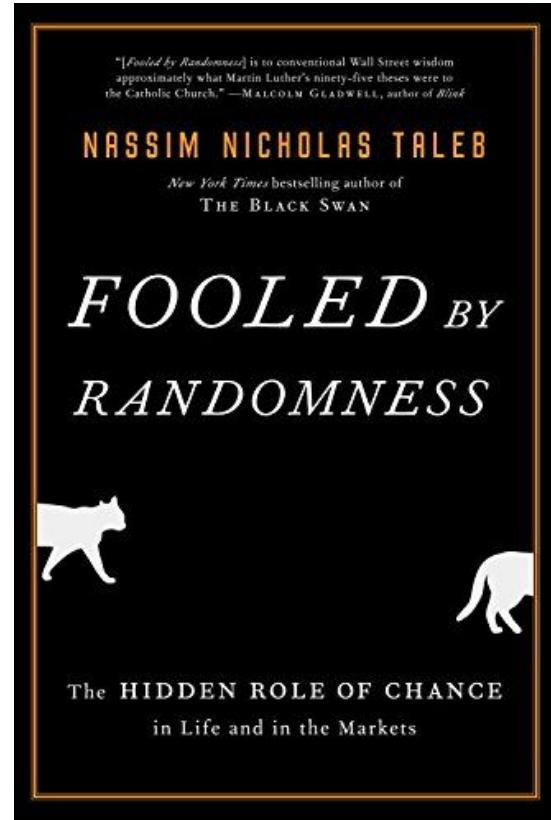
↳ discerner bruit  
& réalité

Vous pouvez être "dupé par le hasard"

Voir le cours sur les tests statistiques.

La visualisation de l'incertitude est aussi importante que la visualisation du phénomène que vous souhaitez illustrer.

(titre en français: Le Hasard sauvage)



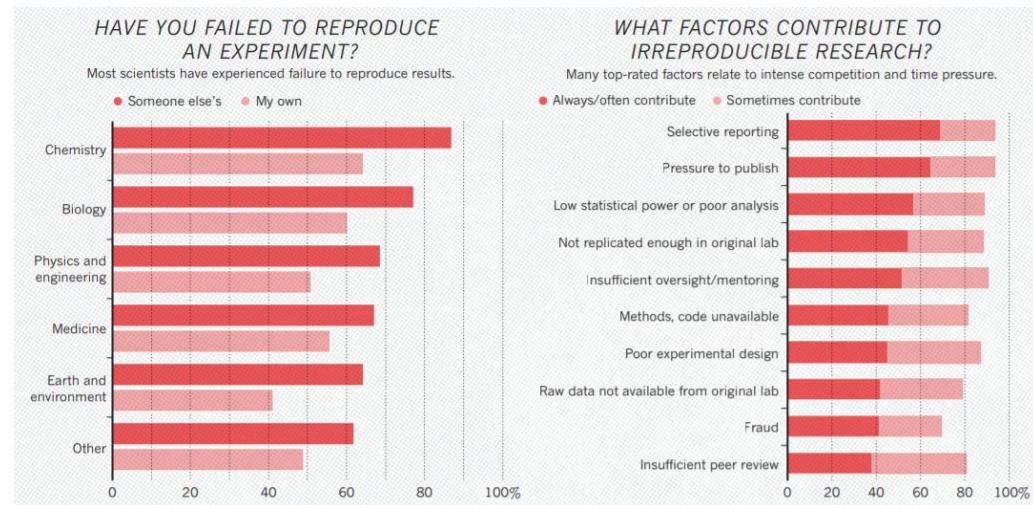
# Reproductibilité

Aléas dans les données.

Important de s'assurer que vos visualisations et ses conclusions sont reproductibles.

*"La reproductibilité, c'est comme se brosser les dents. Une fois que vous l'apprenez, cela devient une habitude."*

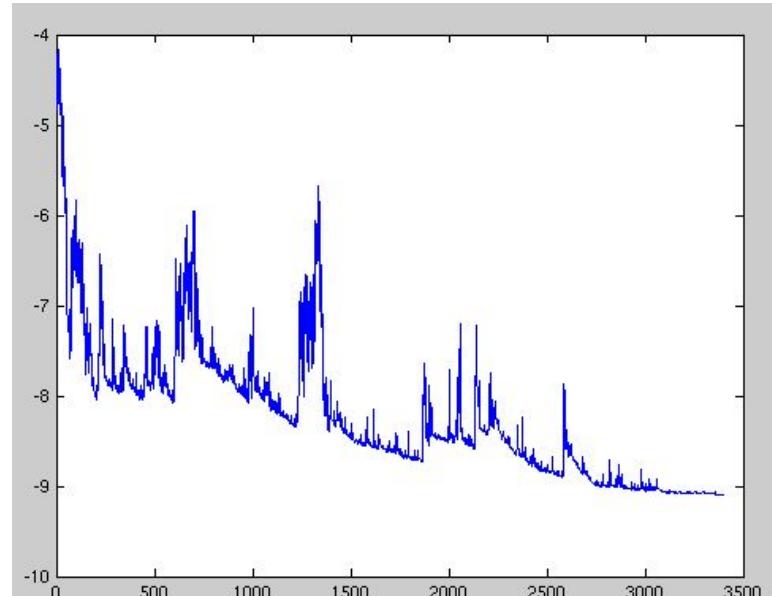
*Irakli Loladzé*



Crise de reproductibilité en science (2016)

# Le temps est un élément important du ML

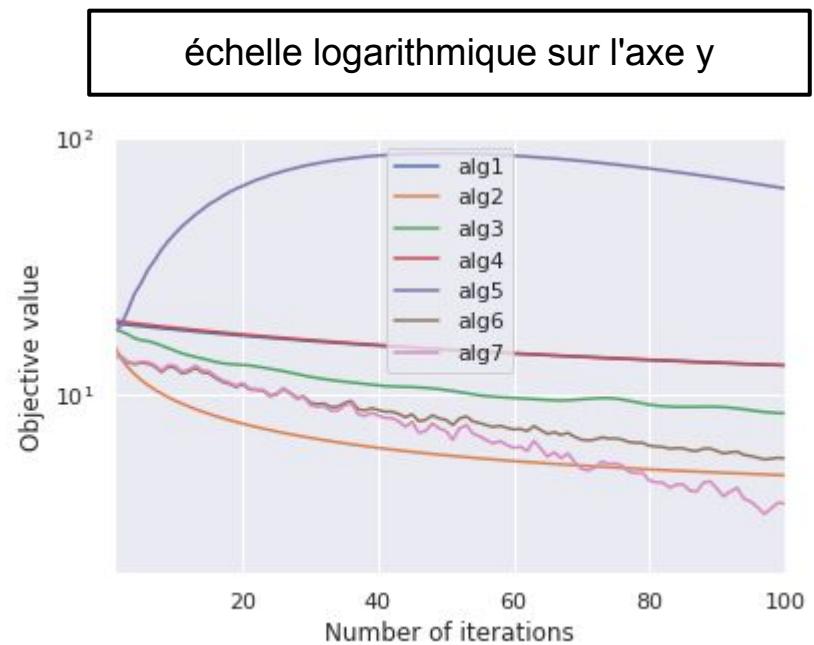
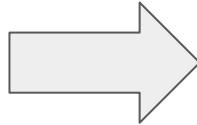
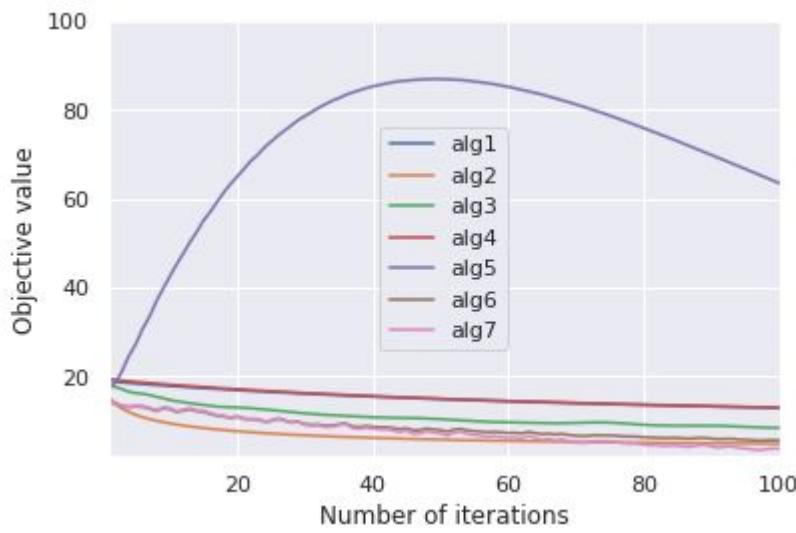
- Question potentielle : dois-je m'entraîner davantage ? **Non, on peut s'arrêter**
- Nous sommes bons en extrapolation.
- Bien que nous puissions être trompés.



source : wikipédia sur SGD

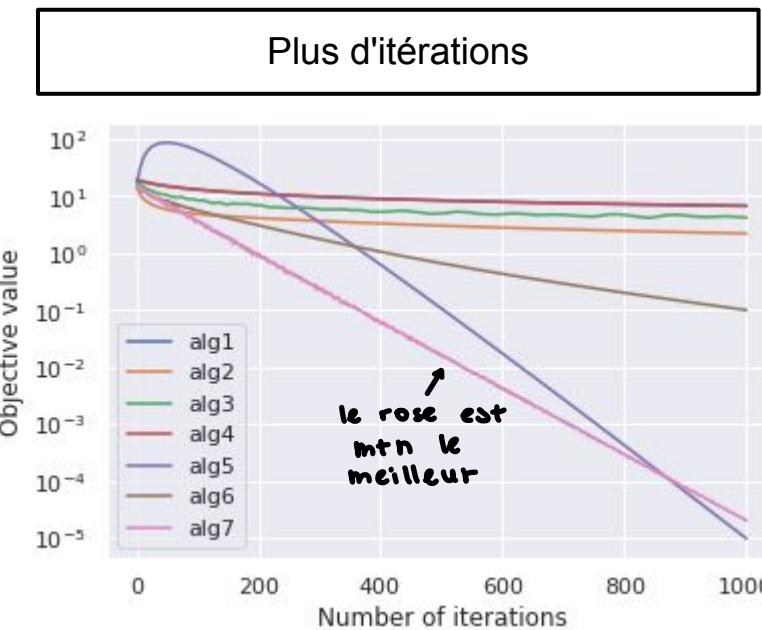
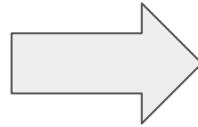
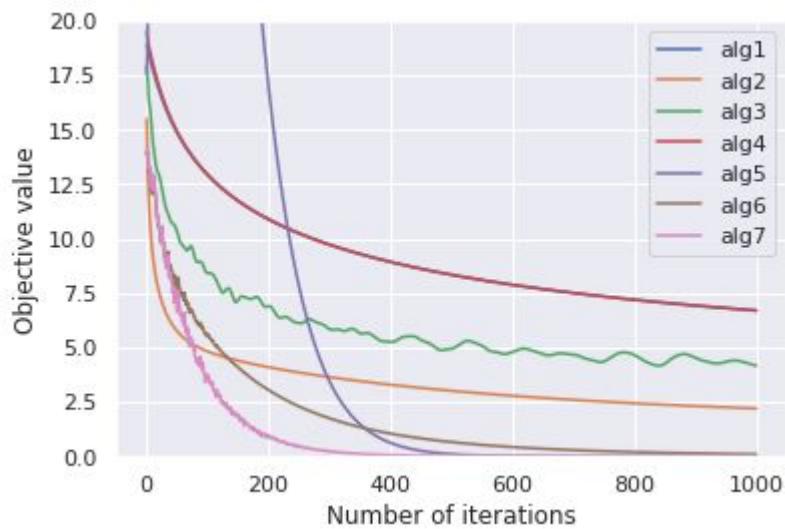
# Le temps est un élément important du ML

- Question potentielle : dois-je m'entraîner davantage ? **Oui**
- Nous sommes bons en extrapolation.
- Bien que nous puissions être dupé.



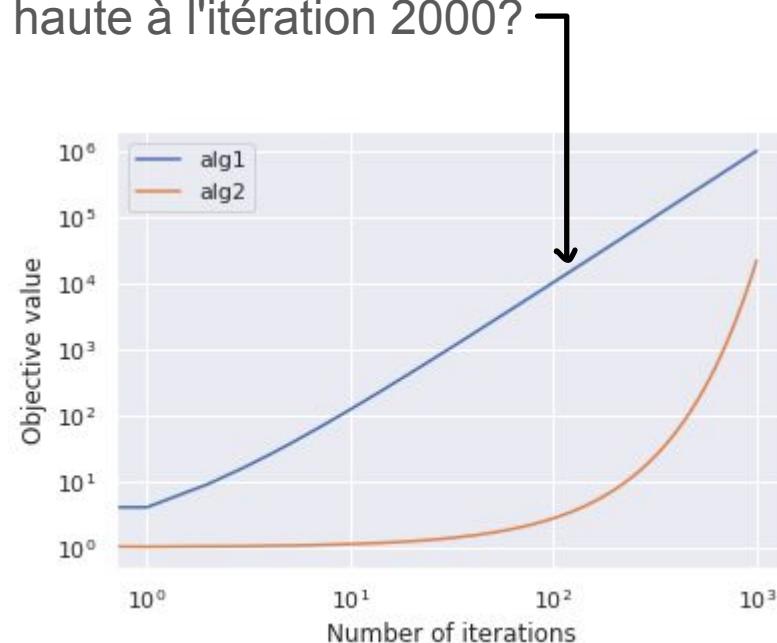
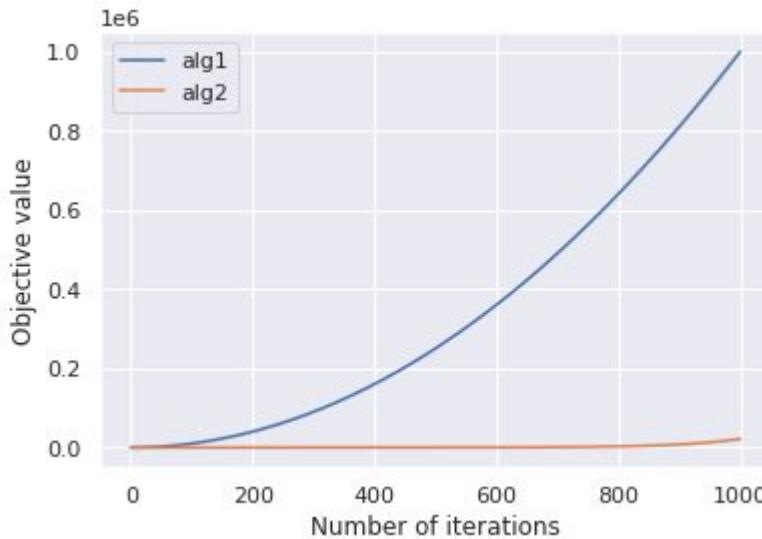
# Le temps est un élément important du ML

- Question potentielle : dois-je m'entraîner davantage ? **Oui**
- Nous sommes "bons" en extrapolation.
- Bien que nous puissions être dupé.



# Choisir la bonne échelle

- Nous sommes bons pour extrapolier des lignes!
- Pas le reste
- Question: Quelle sera la courbe la plus haute à l'itération 2000?



# Choisir la Bonne échelle

Un nénuphar sur un lac:

- Sa taille double chaque jour.
- On veut **intervenir lorsque la moitié du lac est couvert**
- On sait que cela prend **30 jours pour couvrir la totalité du lac.**
- Quand va-t-on intervenir?

A. 1er jour B. 2eme jour

C. 15eme jour D. 28eme jour

E. 29eme jour



# Choisir la Bonne échelle

Un nénuphar sur un lac:

- Sa taille double chaque jour.
- On veut **intervenir lorsque la moitié du lac est couvert**
- On sait que cela prend **30 jours pour couvrir la totalité du lac.**
- Quand va-t-on intervenir?
  - A. 1er jour
  - B. 2eme jour
  - C. 15eme jour
  - D. 28eme jour
  - E. 29eme jour**



# Choisir la Bonne échelle

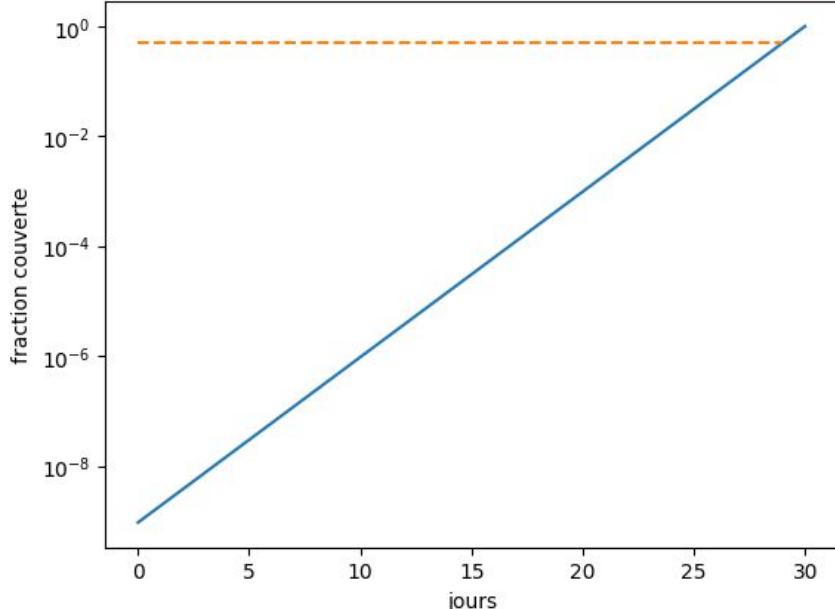
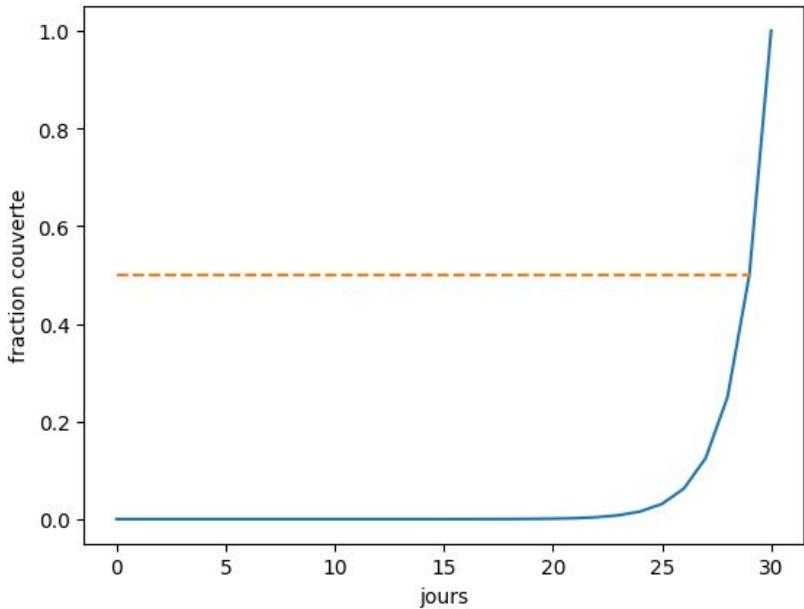
Un nénuphar sur un lac:

- Sa taille double chaque jour.
- On veut **intervenir lorsque la moitié du lac est couvert**
- On sait que cela prend **30 jours pour couvrir la totalité du lac.**
- A quelle fraction sera recouverte le 15eme jour?



115015104

# Choisir la Bonne échelle



115015104



OPEN

## Impact of COVID-19 forecast visualizations on pandemic risk perceptions

Lace Padilla<sup>1</sup>✉, Helia Hosseinpour<sup>1,5</sup>, Racquel Fygenson<sup>2,5</sup>, Jennifer Howell<sup>3,5</sup>, Rumi Chunara<sup>2,5</sup> & Enrico Bertini<sup>2,4,5</sup>

People worldwide use SARS-CoV-2 (COVID-19) visualizations to make life and death decisions about pandemic risks. Understanding how these visualizations influence risk perceptions to improve pandemic communication is crucial. To examine how COVID-19 visualizations influence risk perception, we conducted two experiments online in October and December of 2020 ( $N = 2549$ ) where we presented participants with 34 visualization techniques (available at the time of publication on the CDC's website) of the same COVID-19 mortality data. We found that visualizing data using a cumulative scale consistently led to participants believing that they and others were at more risk than before viewing the visualizations. In contrast, visualizing the same data with a weekly incident scale led to variable changes in risk perceptions. Further, uncertainty forecast visualizations also affected risk perceptions, with visualizations showing six or more models increasing risk estimates more than the others tested. Differences between COVID-19 visualizations of the same data produce different risk perceptions, fundamentally changing viewers' interpretation of information.

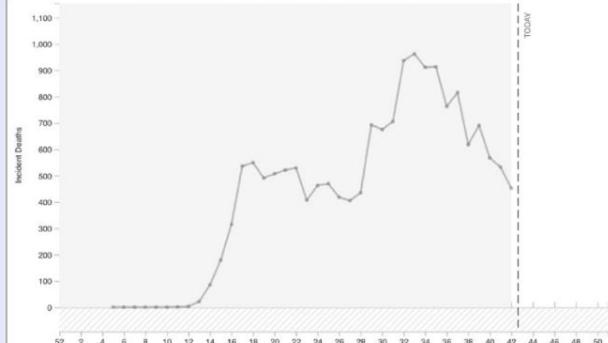
# Changement de perception entre incident and cumulative death

Incident deaths in California

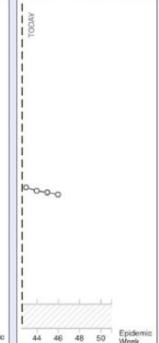
Cumulative deaths in California

## Control Groups

A Historical COVID-19 data (no forecast)

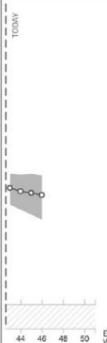


B Mean  
(no uncertainty)

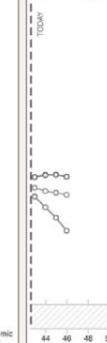


## Summary Visualizations

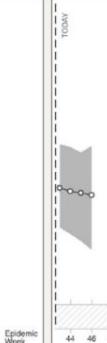
C Mean +  
50%CI



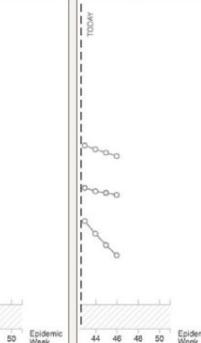
D 3 Models  
Spanning 50%CI



E Mean +  
95%CI



F 3 Models  
Spanning 95%CI



## Distributional Visualizations

G 6 Models



H All Models



Example stimuli (generated in Oct. 2020)

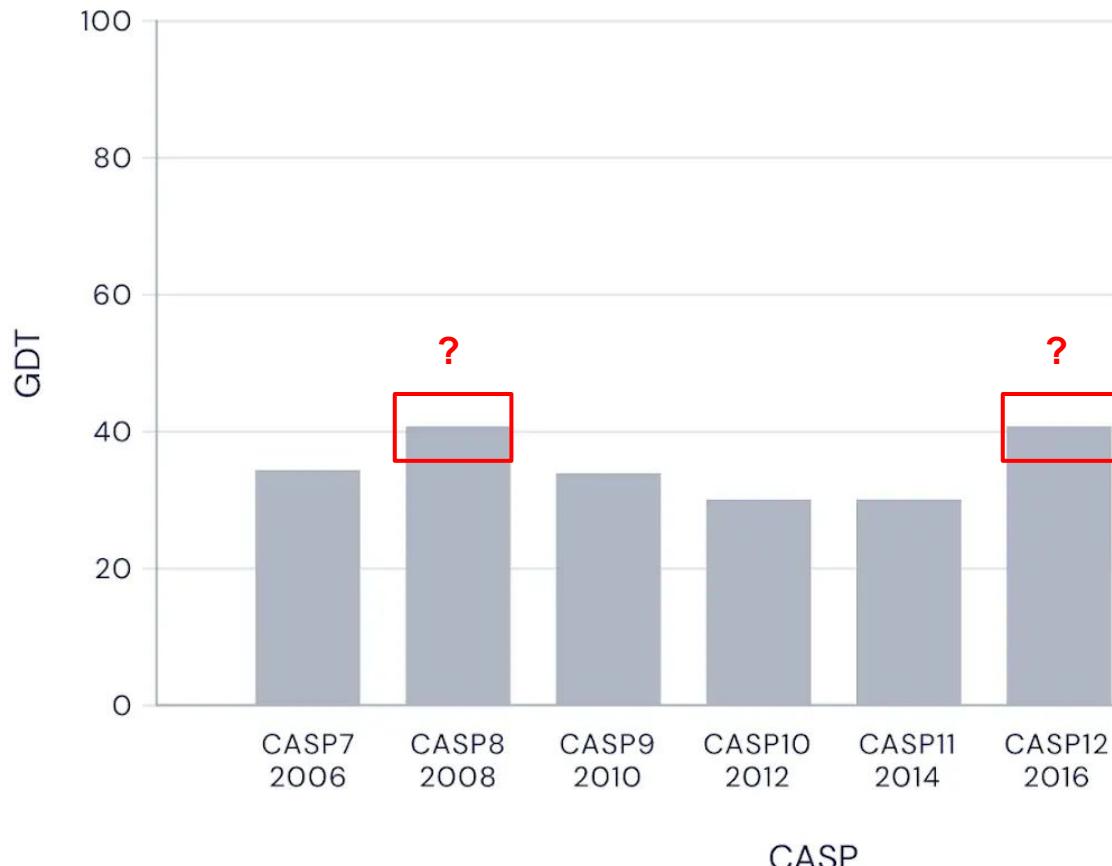
Two-week COVID-19 predictions

# Présentation des données

La visualisation des données peut être plus convaincante que les données brutes

"Nous avons amélioré l'état de l'art de 100 %"

## Median Free-Modelling Accuracy



Source : [www.deepmind.com](http://www.deepmind.com)

## Median Free-Modelling Accuracy



Message fort ici :

- Amélioration significative

Mais aussi:

- La communauté stagnait.
- Le problème est difficile.
- Donne une idée de la variance des résultats.

# Méthode (proposée) pour présentation des données

# Répondre à une question par visualisation

- La question peut être complexe mais doit être claire.

Exemples:

- Quels sont les meilleurs algorithmes pour prédire le repliement des protéines
  - Quelle est l'interaction entre les prix Nobel et la qualification des diplômés entre 1901 et 2012.
  - Quelles sont les performances de l'algorithme X sur la tâche Y par rapport à la ligne de base.
  - Les variables X et Y sont-elles corrélées ?
- De préférence : Une réponse par visualisation
  - Exemple :
    - Oui X et Y sont corrélés
    - BetaFold a apporté une amélioration significative au concours CASP pour le repliement des protéines.

## Avoir un récit clair

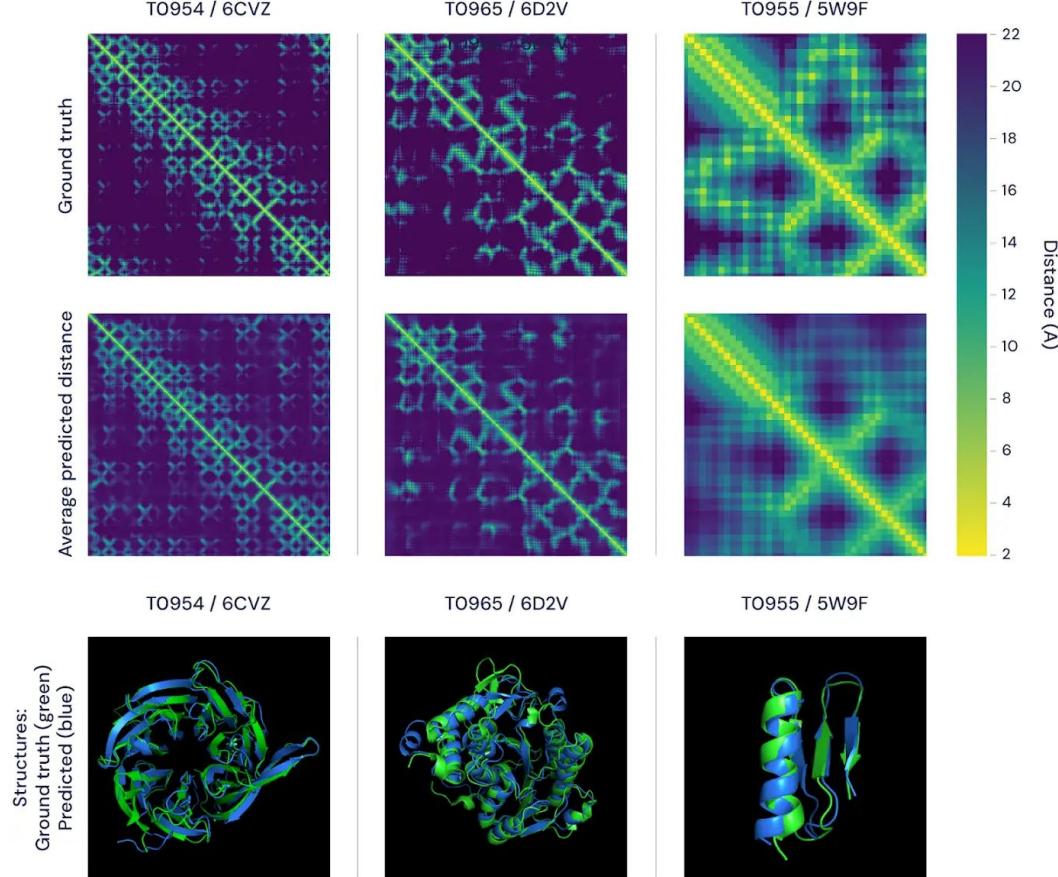
- Votre visualisation n'est pas seule.
- Il soutient un récit.
- Vous devriez poser une question claire.
- Motivez-le et donnez le contexte.
- Donnez une réponse claire... appuyée par la visualisation.

# Ayez un : "Comment lire ceci"

- Des choses simples:
  - Avoir un axe étiqueté
  - Avoir un titre
  - Décrivez à quoi la valeur fait référence (est supérieure ou inférieure mieux)
  - Grandes polices de caractère !
- Légendez votre visualisation :
  - N'ayez pas peur d'être redondant
  - Important lorsque vous affichez la visualisation dans des diapositives (vous devez le dire au lieu de l'écrire)

# Exemple de légende

**Légende :** La figure du haut présente les matrices de distance pour trois protéines. La luminosité de chaque pixel représente la distance entre les acides aminés dans la séquence comprenant la protéine - plus le pixel est brillant, plus la paire est proche. La rangée du haut montre les distances réelles déterminées expérimentalement et, dans la rangée du bas, la moyenne des distributions de distance prédites par AlphaFold. Il est important de noter que ceux-ci correspondent bien à l'échelle mondiale et locale. Les panneaux du bas représentent la même comparaison à l'aide de modèles 3D, présentant les prédictions d'AlphaFold (bleu) par rapport aux données de vérité au sol (vert) pour les trois mêmes protéines.

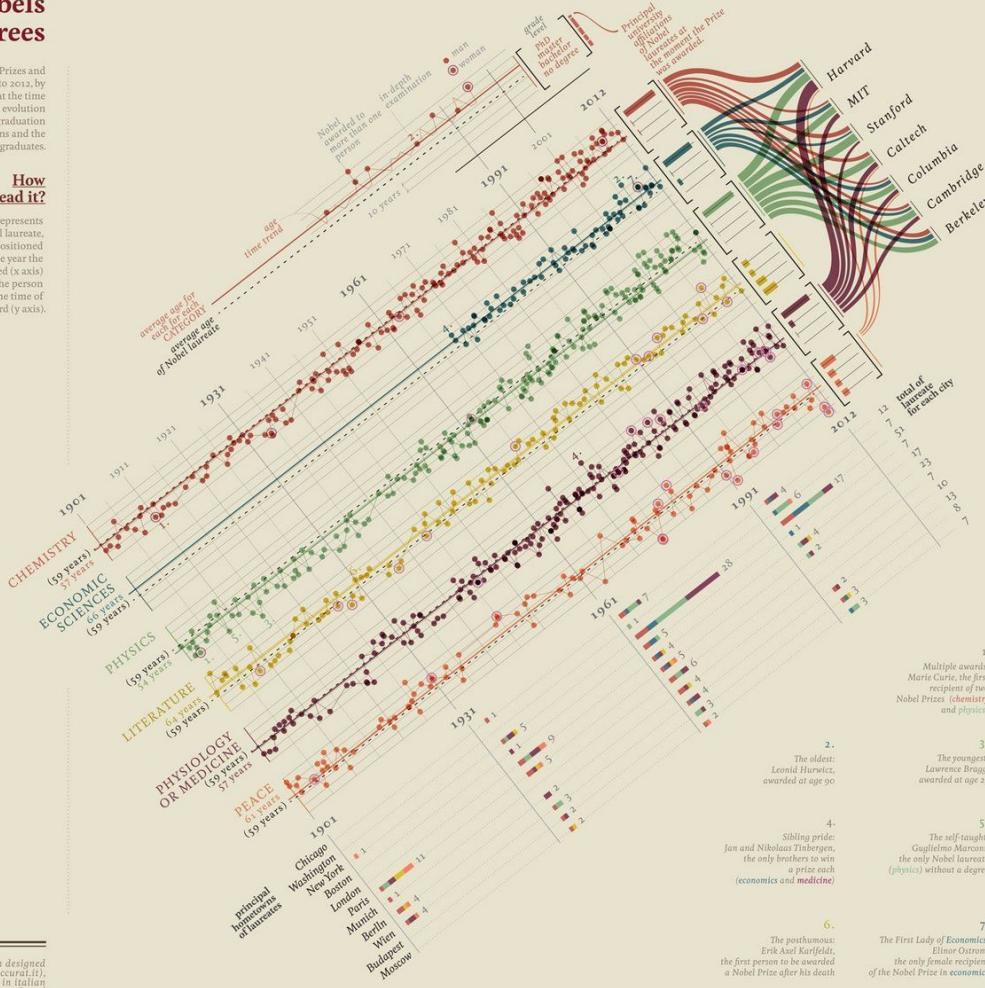


# Nobels no degrees

This visualization explores Nobel Prizes and graduate qualifications from 1901 to 2012 by analysing the age of recipients at the time prizes were awarded, average age evolution through time and among categories, graduation grades, main university affiliations and the principal hometowns of the graduates.

## How to read it?

Each dot represents a Nobel laureate, each recipient is positioned according to the year the prize was awarded ( $x$  axis) and age of the person at the time of the award ( $y$  axis).

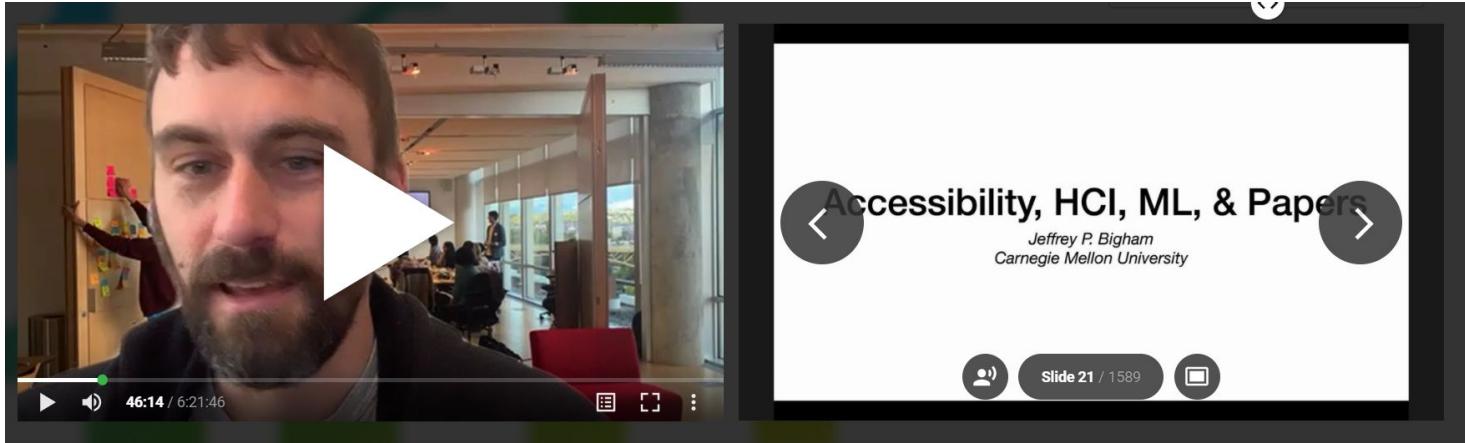


# Pensez à l'accessibilité

Au moins :

- visualisation adaptée pour les daltoniens.
- des graphiques agrandissables (pas .png, préférez les formats vectoriels comme .pdf)

L'accessibilité va au-delà de ça:

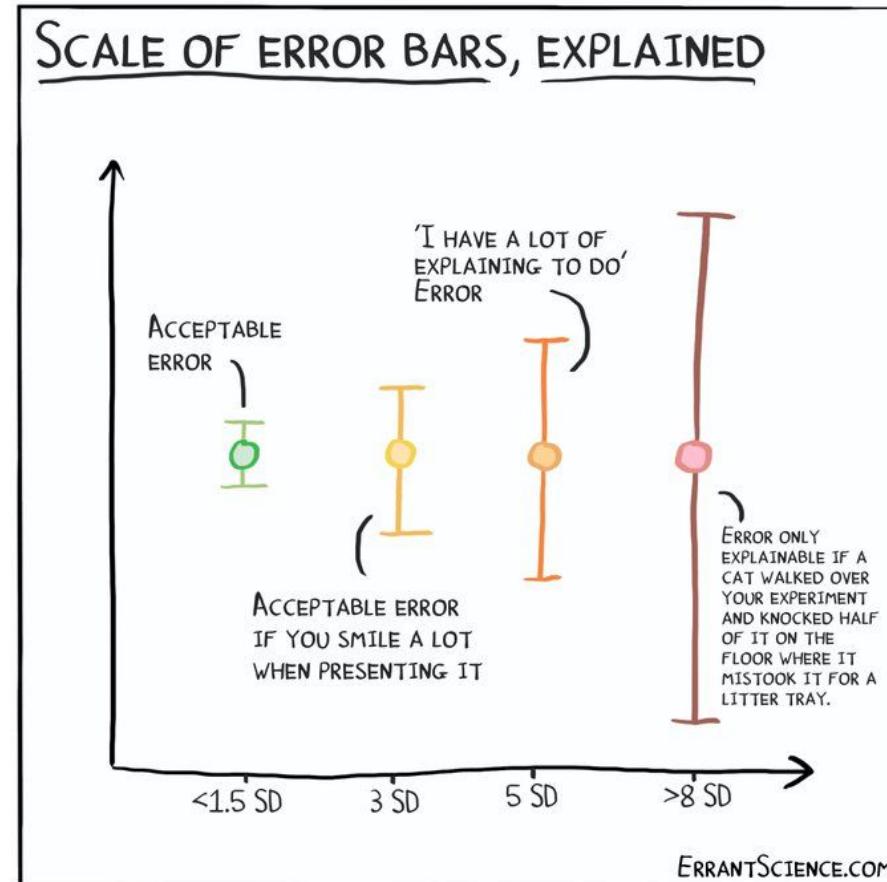


# Citez vos sources !

- Éthique académique.
- **Très important.**
- Citez les données que vous avez utilisées d'où provient la visualisation si vous l'avez utilisée de quelque part.

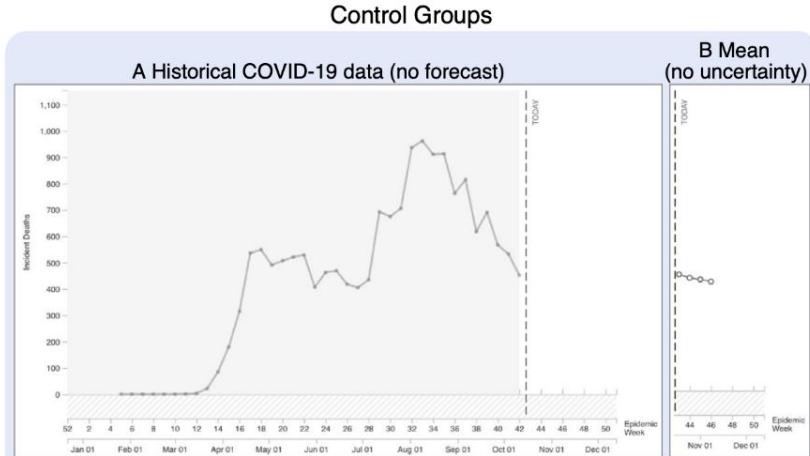
# La barre d'erreur (ou équivalent) est importante

- Le retour du « hasard sauvage»
- Tu ne veux pas te tromper
- Ne trompe pas les autres.
- Crise de reproductibilité en ML
- Les barres d'erreurs sont le **strict minimum**.

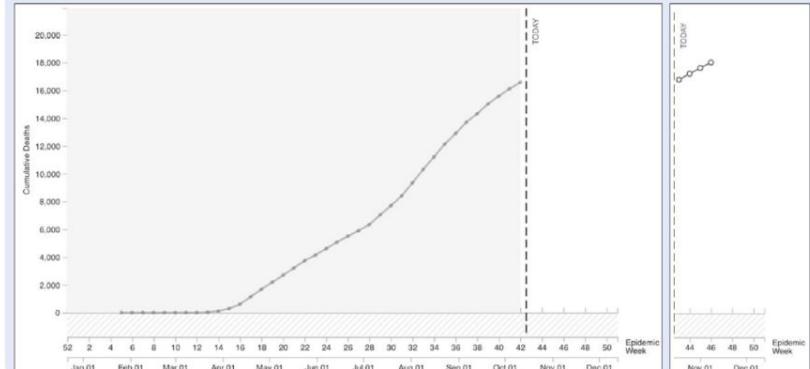


Meilleur visualisation de l'incertitude!

### Incident deaths in California

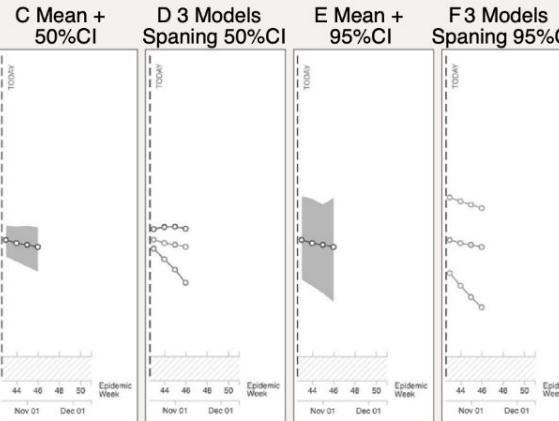


### Cumulative deaths in California



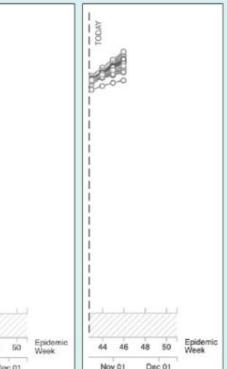
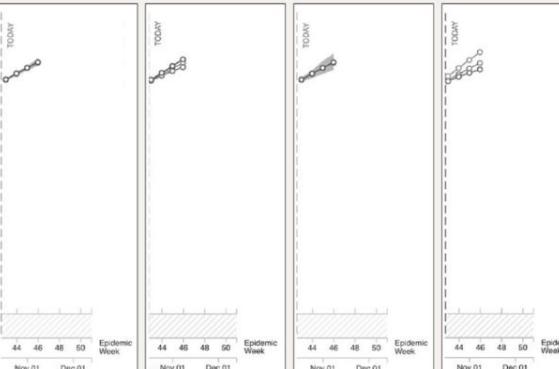
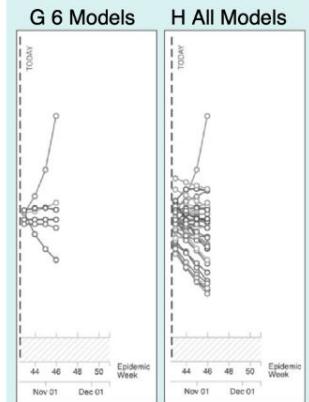
Example stimuli (generated in Oct. 2020)

### Summary Visualizations



Two-week COVID-19 predictions

### Distributional Visualizations



La performance moyenne est parfois non représentative.  
 Surtout en RL avec des performances multi-modales.  
 La même expérience peut mener à différent optima.

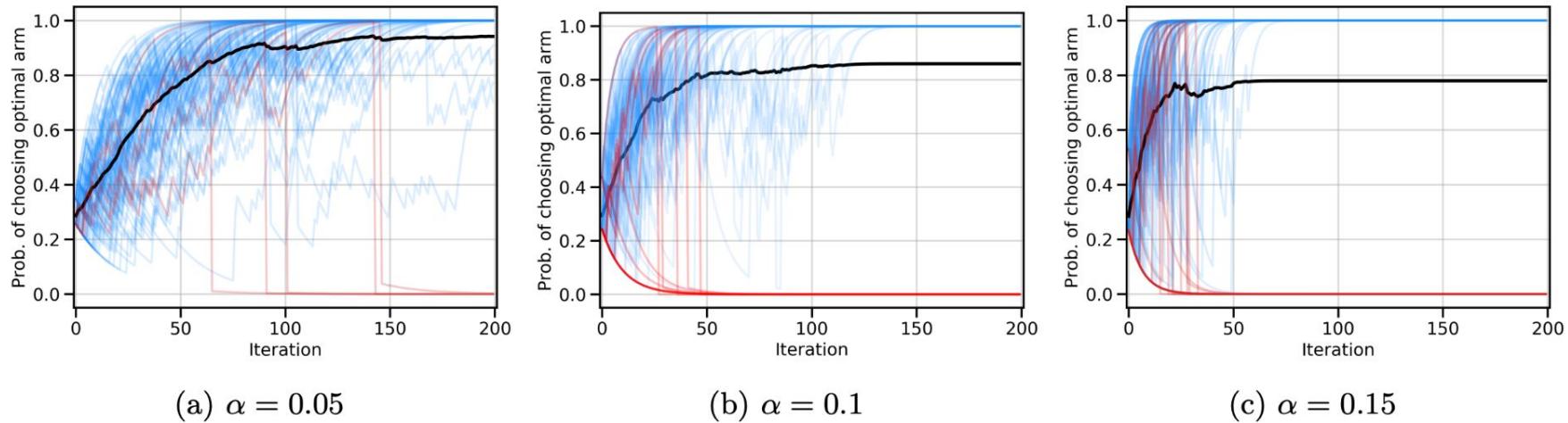


Figure 2: Learning curves for 100 runs of 200 steps, on the two-arm bandit, with baseline  $b = -1$  for three different stepsizes  $\alpha$ . *Blue*: Curves converging to the optimal policy. *Red*: Curves converging to a suboptimal policy. *Black*: Avg. performance. The number of runs that converged to the suboptimal solution are 5%, 14% and 22% for the three  $\alpha$ 's. Larger  $\alpha$ 's are more prone to getting stuck at a suboptimal solution but settle on a deterministic policy more quickly.

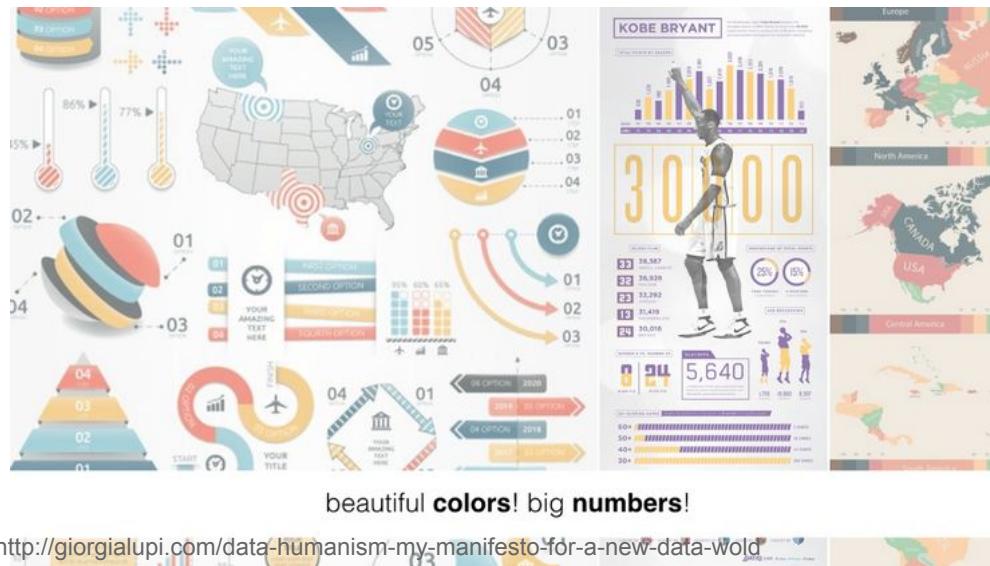
# Visualisation avancée des données

# Clause de non-responsabilité

- Pour créer des visualisations avancées, il est très important de **maîtriser les bases.**
- Compromis inhérent entre **complexité et clarté.**
- Faire de telles visualisations avancées peut prendre **beaucoup de temps.**
- De plus en plus utilisé dans la communauté. (Bien que souvent par de grandes équipes)

## La visualisation avancée des données n'est pas une infographie

- Il ne s'agit pas de couleurs flashy et d'accumulations de GROS CHIFFRES et de statistiques.
  - L'esthétique est importante mais pas le but ultime.

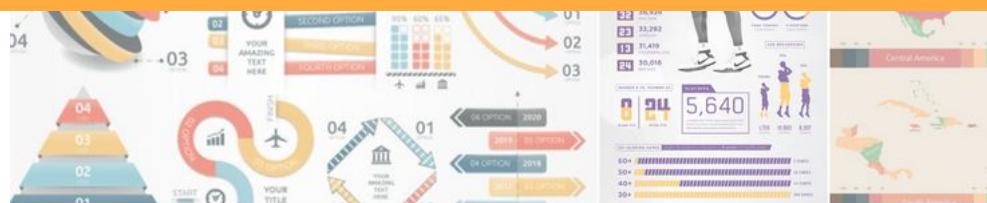


# La visualisation avancée des données n'est pas une infographie

- Il ne s'agit pas de couleurs flashy et d'accumulations de GROS CHIFFRES et de statistiques.
- L'esthétique est importante mais pas le but ultime.



Les données sont complexes.  
La visualisation avancée doit relier les nombres



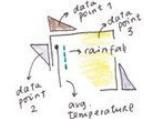
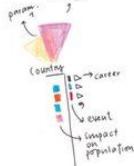
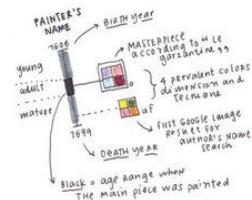
beautiful colors! big numbers!

Source : <http://giorgialupi.com/data-humanism-my-manifesto-for-a-new-data-world>

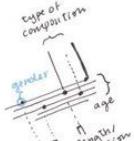
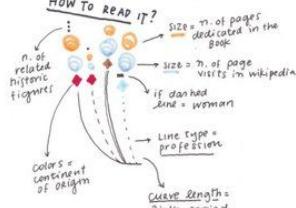
# Embrasser la complexité

- "Les êtres humains aiment les histoires et la narration est l'un des moyens les plus efficaces et les plus puissants d'apprendre, de découvrir et de diffuser des informations. Votre projet doit pouvoir transmettre un message et résumer facilement un récit convaincant. »
- La visualisation avancée: exprimer un récit complexe
- Certain niveau de complexité visuelle -> Compromis inhérent entre clarté et complexité.
- **Contexte et "Comment le lire"** encore plus important ici (car non standard)
- **Avantage :** (contrairement au texte), narration non linéaire !

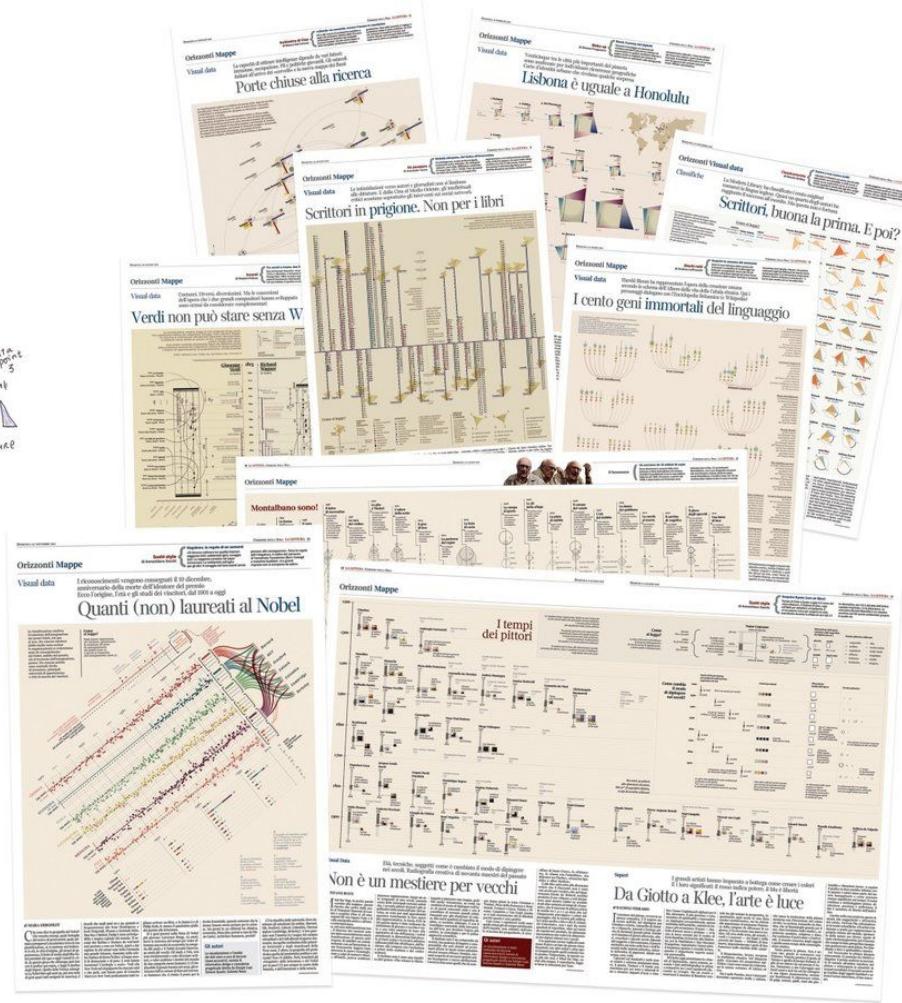
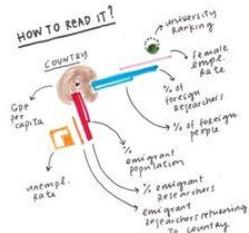
### HOW TO READ IT?



### HOW TO READ IT?



### HOW TO READ IT?



# Nobels no degrees

This visualization explores Nobel Prizes and graduate qualifications from 1901 to 2012 by analysing the age of recipients at the time prizes were awarded, average age evolution through time and among categories, graduation grades, main university affiliations and the principal hometowns of the graduates.

## How to read it?

Each dot represents a Nobel laureate, each recipient is positioned according to the year the prize was awarded ( $x$  axis) and age of the person at the time of the award ( $y$  axis).

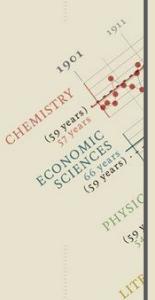


## Nobels no degrees

This visualization explores Nobel Prizes and graduate qualifications from 1901 to 2012, by analysing the age of recipients at the time prizes were awarded, average age evolution through time and among categories, graduation grades, main university affiliations and the principal hometowns of the graduates.

### How to read it?

Each dot represents a Nobel laureate, each recipient is positioned according to the year the prize was awarded (x axis) and age of the person at the time of the award (y axis).



# Nobels no degrees

This visualization explores Nobel Prizes and graduate qualifications from 1901 to 2012, by analysing the age of recipients at the time prizes were awarded, average age evolution through time and among categories, graduation grades, main university affiliations and the principal hometowns of the graduates.

### How to read it?

Each dot represents a Nobel laureate, each recipient is positioned according to the year the prize was awarded (x axis) and age of the person at the time of the award (y axis).

Columbia  
Cambridge  
Berkeley

age each city

1901  
1911  
1921  
1931  
1941  
1951  
1961  
1971  
1981  
1991  
2001  
2011

1.  
Multiple awards:  
Marie Curie, the first  
recipient of two  
Prizes (chemistry  
and physics)

3.  
The youngest:  
Lawrence Bragg,  
awarded at age 25

5.  
The self-taught:  
Guglielmo Marconi,  
Nobel laureate  
without a degree

7.  
First female Nobel  
Prize in Economics:  
Elinor Ostrom,  
a female recipient  
of the Nobel Prize in Economics



In the past year, countries with nuclear weapons continued to devote tens of billions of dollars on nuclear modernization. This is despite the alleged Nuclear Risk

# An Expanding Concern

The number of nuclear warheads has fallen from Cold War highs and nuclear testing has all but ceased. Still, nuclear risk is rising as arms control treaties wither and die, the potential for accidental nuclear war remains under-appreciated despite decades of nuclear mishaps, and nuclear weapons countries devote billions of dollars to nuclear modernization. Has the world come full circle and begun a new nuclear arms race?

**Information on nuclear warheads, countries possessing or seeking**

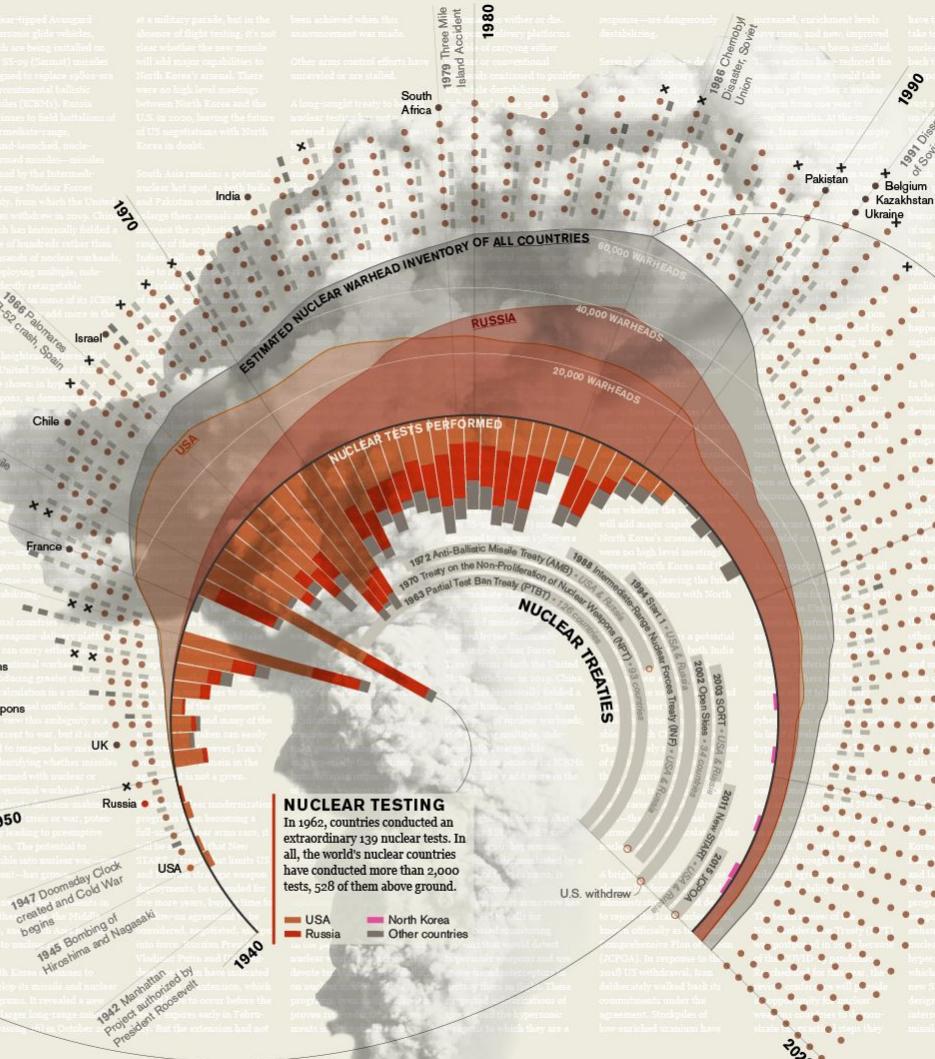
countries possessing or seeking nuclear weapons, nuclear testing, and nuclear diplomatic agreements is arranged on a circular timeline that starts in 1945 (lower left) and moves clockwise toward the present.

prehensive Plan of Action (CPA). In response to the US withdrawal, Iran gradually walked back its commitments under the deal. Stockpiles of enriched uranium have increased, enrichment levels have been improved and centrifuges installed. The IAEA has made no further inspections since October 2018. At the same time, Iran continues to comply with many of the agreement's restrictions, and many of the steps it has taken can easily be reversed within a few months. At the same time, Iran continues to comply with many of the agreement's restrictions, and many of the steps it has taken can easily

military parade, but in the absence of flight testing, it's not clear whether the new missile adds major capabilities to North Korea's arsenal. There were no high-level meetings between North Korea and the United States in 2000, leaving the future of negotiations with North Korea in doubt.

*...and when Milner  
achieved what he  
concerned was made.  
cross control efforts have  
weaved or are stalled.*

**1981**  
with every platinums  
carrying estate  
convention  
continued to provide  
the destination  
  
**Y OF ALL COUNTRIES**



or will commit to the safety of space and sea lanes, or reduce nuclear arsenals on nuclear weapons.

Days ago, the Treaty of Nukashiro was ratified into force and completed. This treaty brings new attention to the threat posed by nuclear weapons, especially the enormous long-term impacts any use of weapons would have. The Treaty of Nukashiro also serves to continue to build institutions of intermediate-range nuclear launched missile systems, missiles banned by the Intermediate-range Nuclear Forces Treaty, from which the United States withdrew in 2019. China, which has historically fielded a force of hundreds rather than thousands of nuclear warheads, is deploying multiple, independently targetable warheads on some of its ICBMs and will likely add more in the coming year.

The heightened interest that France and Russia, both countries situated by a major river, has

## **NUCLEAR INVENTORY**

In 1986, the global inventory of nuclear weapons peaked at more than 64,000, declining dramatically thereafter, largely because of arms control agreements between the United States and the Soviet Union/Russia.

## CURRENT INVENTORY

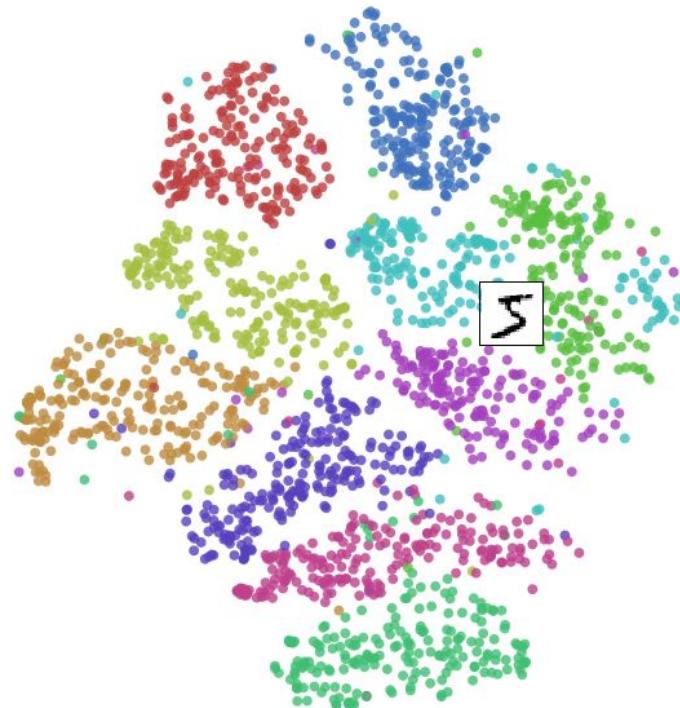
From a high in the 1980s, nuclear arsenals have fallen to about 13,000 warheads in September 2020. Use of even a small fraction of those weapons would kill millions and cause enormous environmental damage, perhaps ending civilization.

# Rendez-le interactif !

## Why Momentum Really Works



# Rendez-le interactif !



A t-SNE plot of MNIST

# Rendez-le interactif !



The Distill logo consists of a white diamond icon followed by the word "Distill" in a white sans-serif font, all contained within a dark blue header bar.

ABOUT PRIZE SUBMIT

## Distill:

- Journal en ligne mettant l'accent sur la visualisation interactive
- Un bon article de blog représente environ **100 heures de travail**.
- Grand succès comme revue de blogs.

July 2, 2021

EDITORIAL

### Distill Hiatus

Editorial Team

After five years, Distill will be taking a break.

---

March 4, 2021

PEER-REVIEWED

### Multimodal Neurons in Artificial Neural Networks

Gabriel Goh, Nick Cammarata †, Chelsea Voss †, Shan Carter, Michael Petrov, Ludwig Schubert, Alec Radford, and Chris Olah

We report the existence of multimodal neurons in artificial neural networks, similar to those found in the human brain.



Three circular illustrations of faces, each with a different abstract, colorful pattern. The first face has a rainbow-like pattern with the words "de a", "fe", "sel", and "ce". The second face has a purple and yellow pattern. The third face has a blue and green pattern with a small figure sitting on it.

---

Nov. 17, 2020

PEER-REVIEWED

### Understanding RL Vision

Jacob Hilton, Nick Cammarata, Shan Carter, Gabriel Goh, and Chris Olah

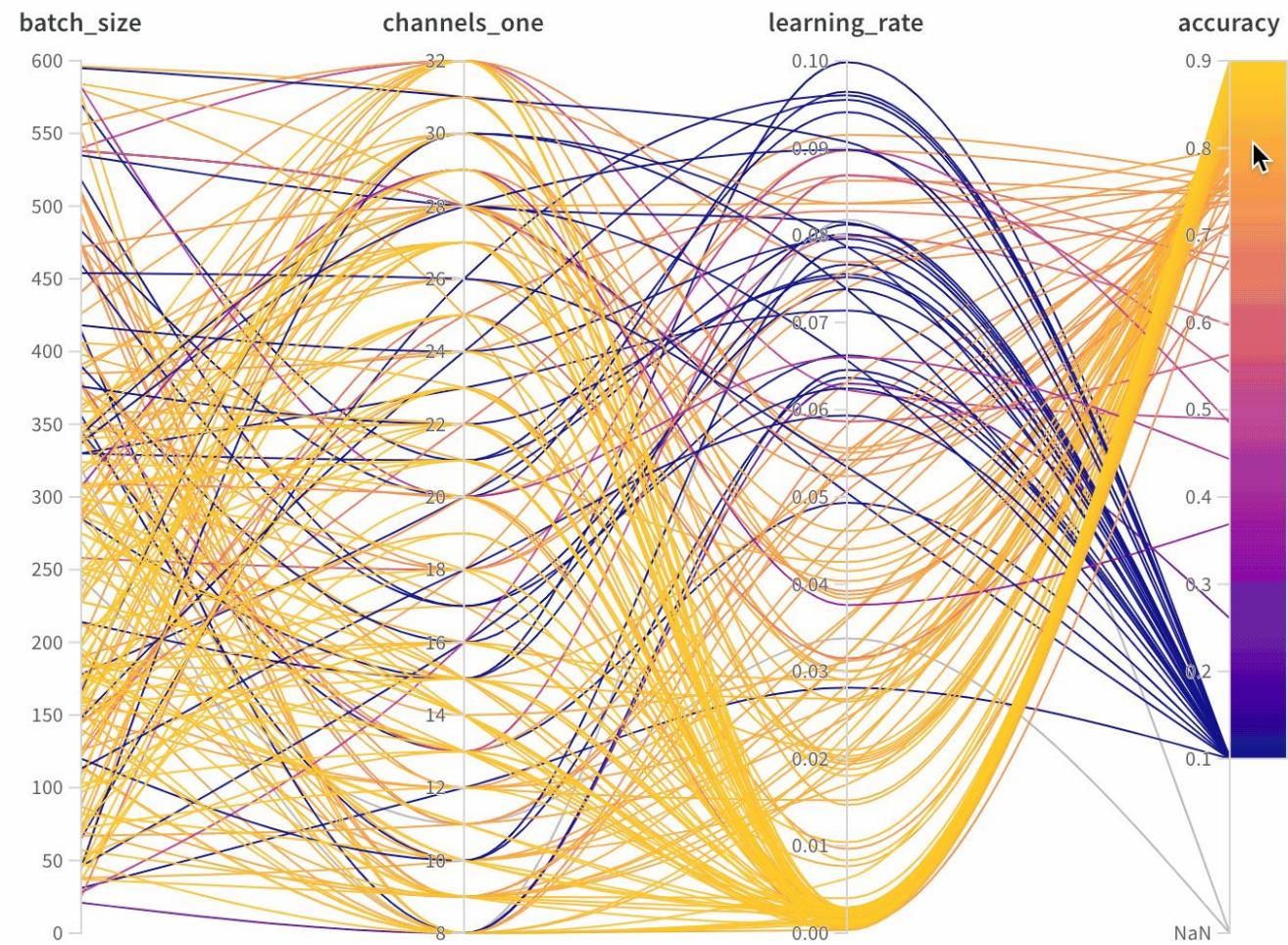
With diverse environments, we can analyze, diagnose and edit deep reinforcement learning models using attribution.



A colorful illustration of a green landscape with a blue robot head in the foreground and a yellow square in the background.

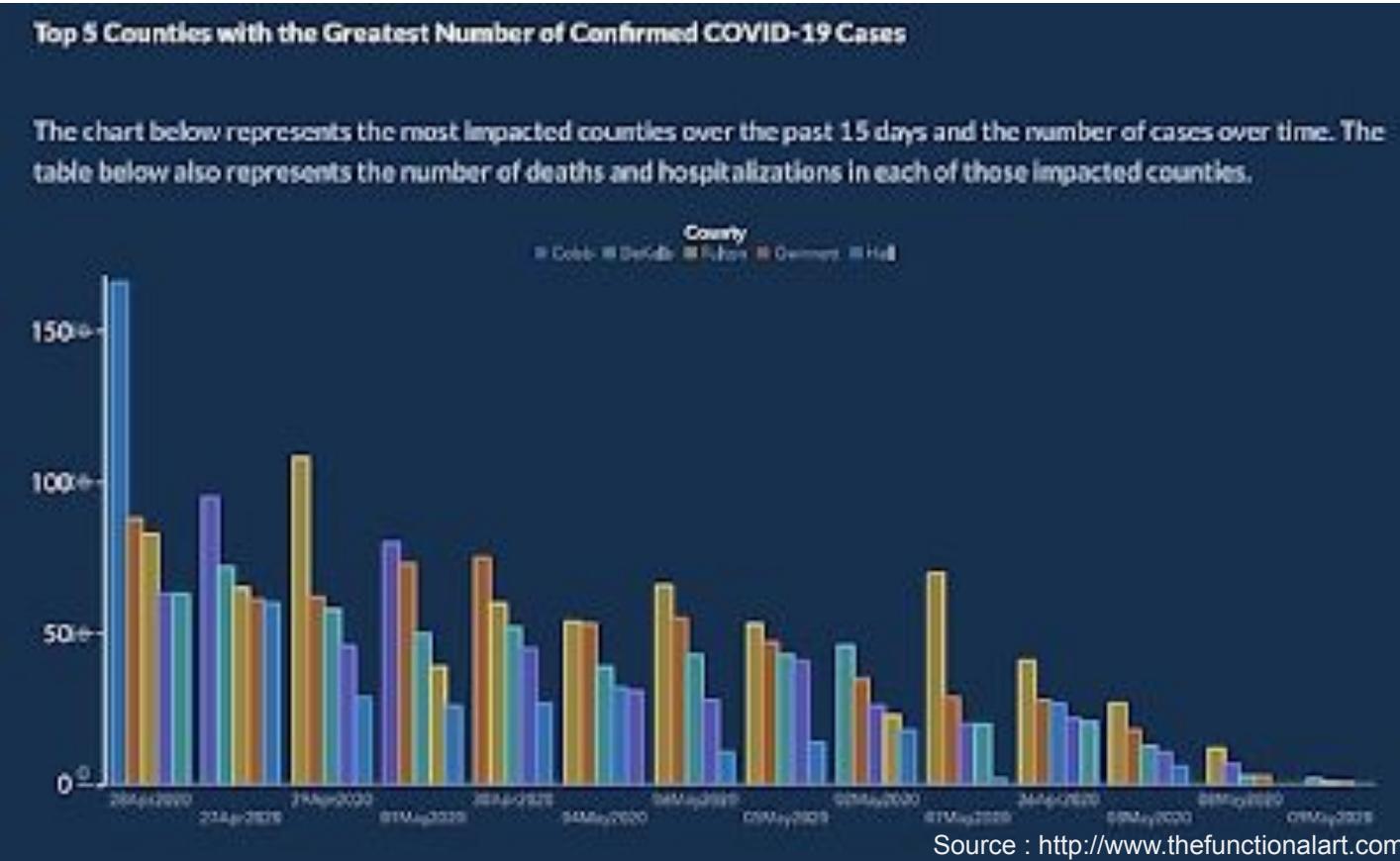
# Recherche d'hyper paramètres

Interactivité ⇒  
un must pour  
l'exploration !!!

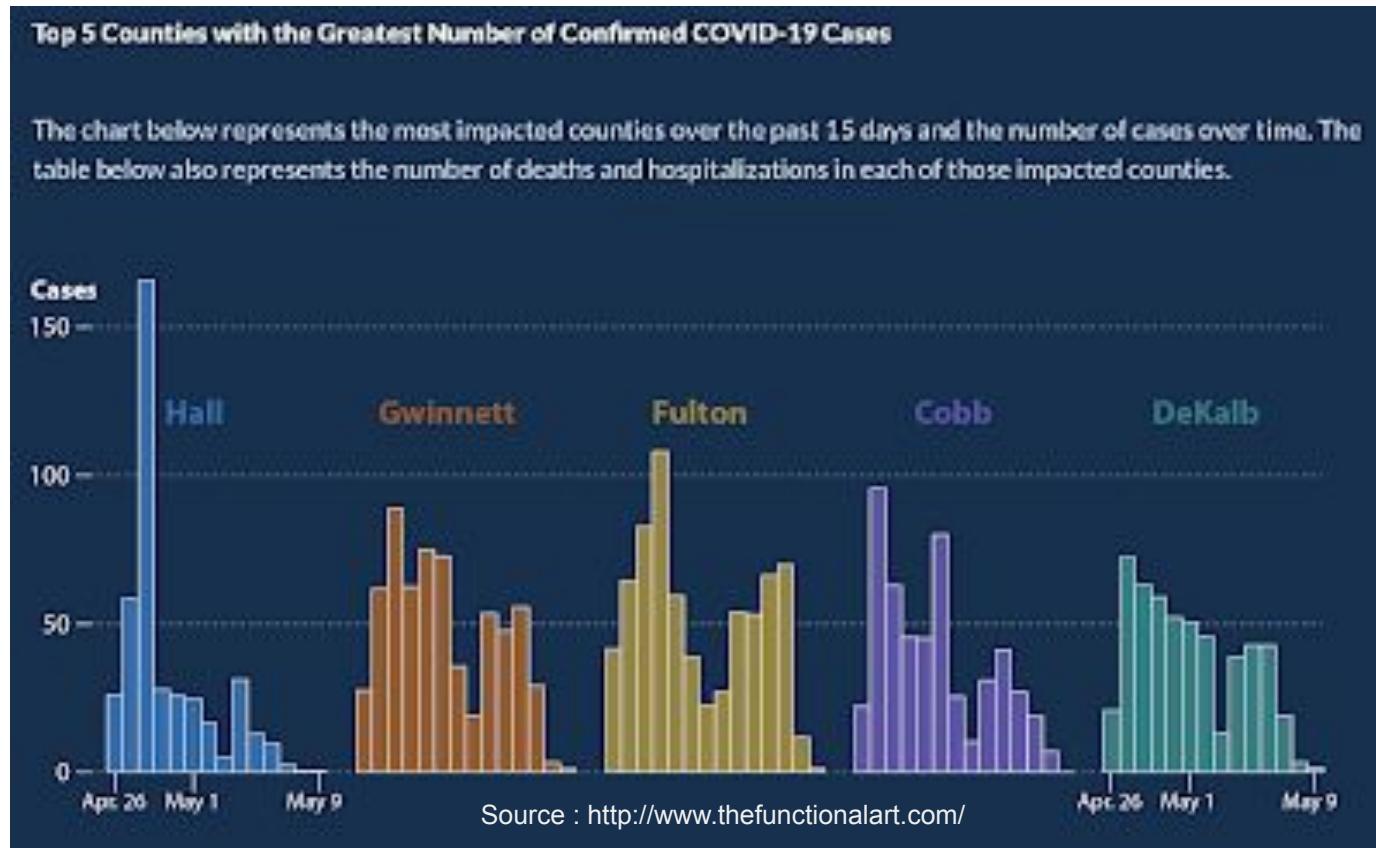


Un grand pouvoir implique de grandes responsabilités

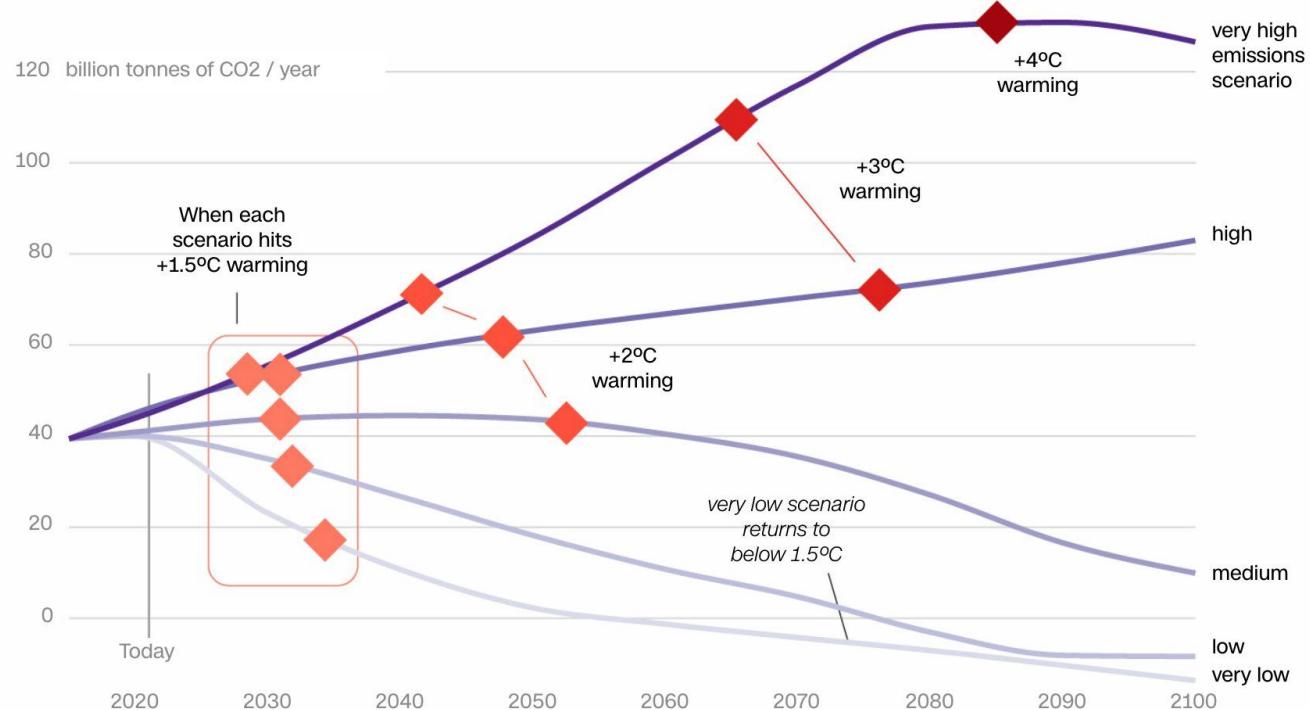
# Vos choix comptent beaucoup !



# Vos choix comptent beaucoup !



# Scénarios d'émissions du GIEC



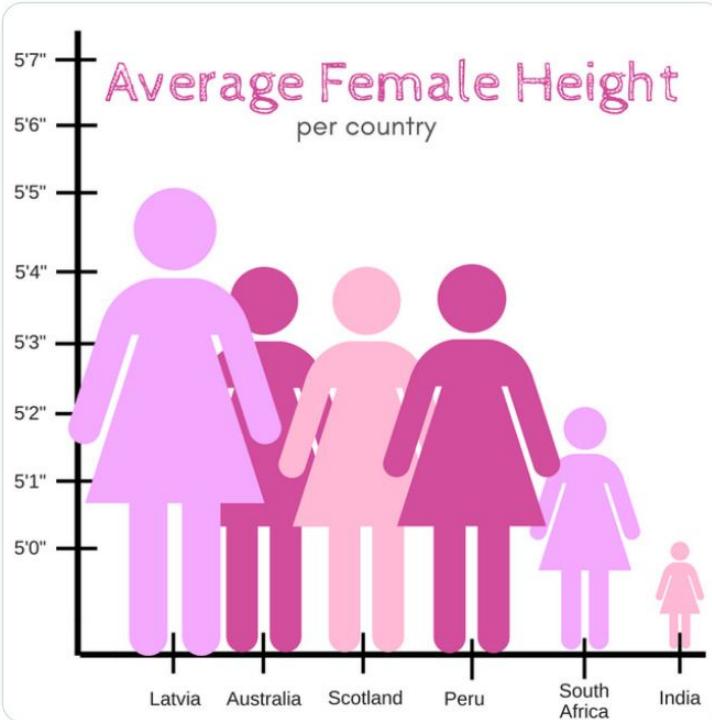
# Dangers des visualisations de données



**Sabah Ibrahim**  
@reina\_sabah



As an Indian woman, I can confirm that too much of my time is spent hiding behind a rock praying the terrifying gang of international giant ladies and their Latvian general don't find me



Quel est le problème ici???

10:58 PM · Aug 6, 2020

i

104.6K

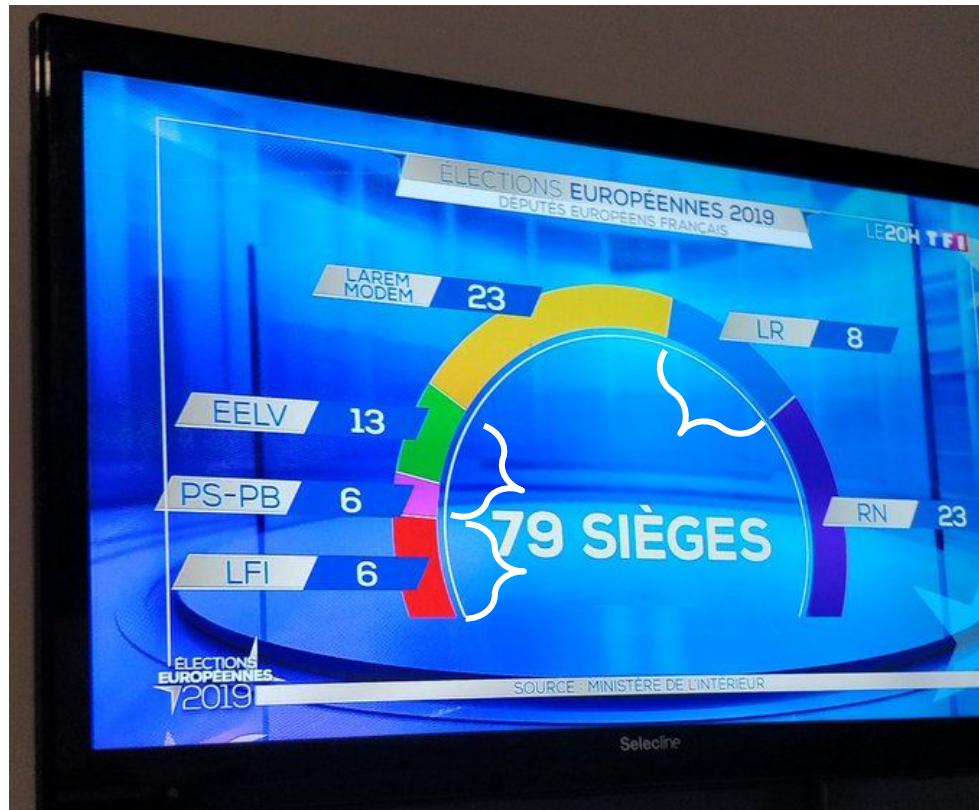
# La visualisation peut mentir !

Source : TF1

Ici : faute professionnelle.

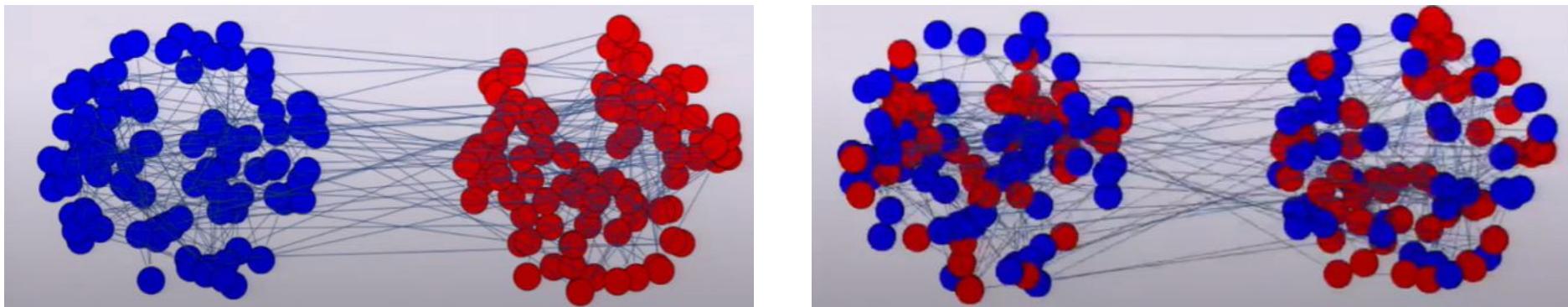
A emporter : toujours confronter data et visualisation !

MAIS...



# On peut se tromper de visualisation.

"Le plus difficile est de trouver un chat noir dans une pièce sombre, surtout s'il n'y a pas de chat" Confusius



Même graphique, même valeur pour le split (11% des arêtes)... Laquelle est la bonne ?

# Avec certains choix, nous influençons certaines conclusions

On peut vouloir conclure :

"Les gens plus riches sont plus disposés à se faire vacciner"

Il faut conclure :

"Il y a une corrélation positive entre le revenu et la vaccination"

Cette corrélation pourrait s'expliquer :

- les personnes âgées sont plus vaccinées (de par la conception des politiques publiques) ont en moyenne un revenu plus élevé.

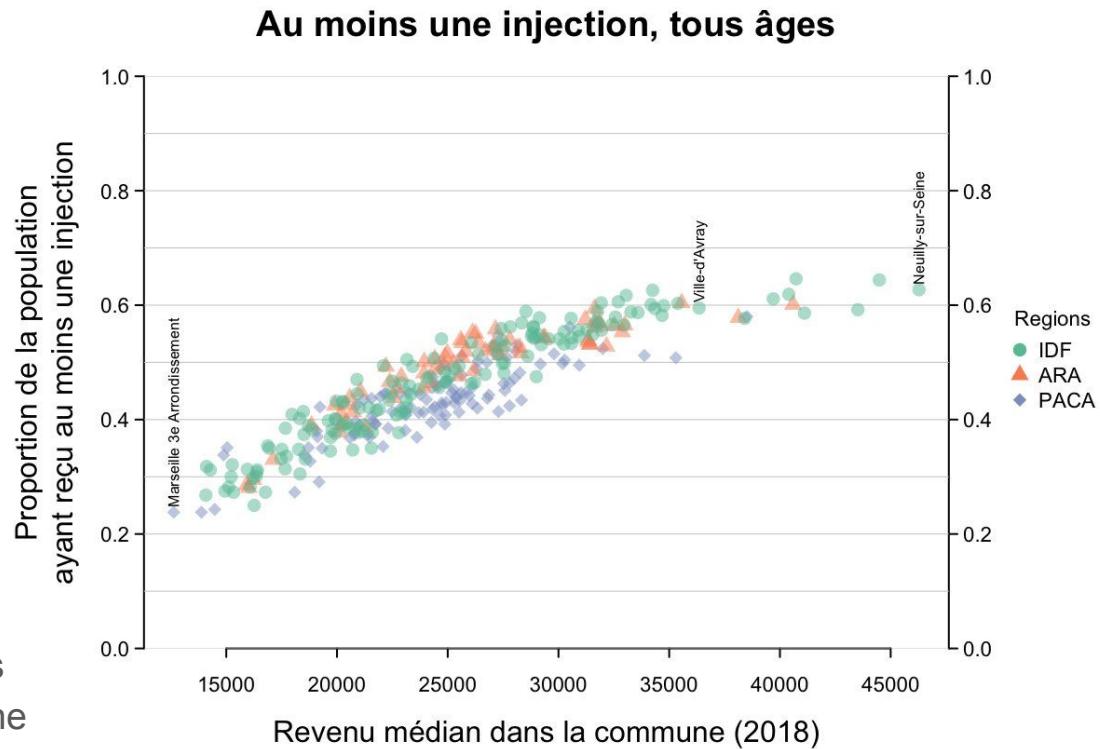


Figure @flodebarre, données vaccination trouvées par @humeursdevictor

Code: [https://github.com/flodebarre/covid\\_vaccination/blob/main/communes.Rmd](https://github.com/flodebarre/covid_vaccination/blob/main/communes.Rmd)

Données vaccination : <https://datavaccin-covid.ameli.fr/explore/dataset/donnees-de-vaccination-par-commune>

Données revenus : <https://www.insee.fr/fr/statistiques/5009236?sommaire=5009255#consulter>

# Deuxième expérience

- Le biais d'âge est éliminé.
- Encore faut-il être prudent dans l'interprétation.
- Nous aurions également besoin de mesures de corrélation telles que le coefficient de corrélation de Spearman. (voir Cours sur les tests statistiques)
- Peut-être une autre métrique pourrait aider à y voir plus clair.

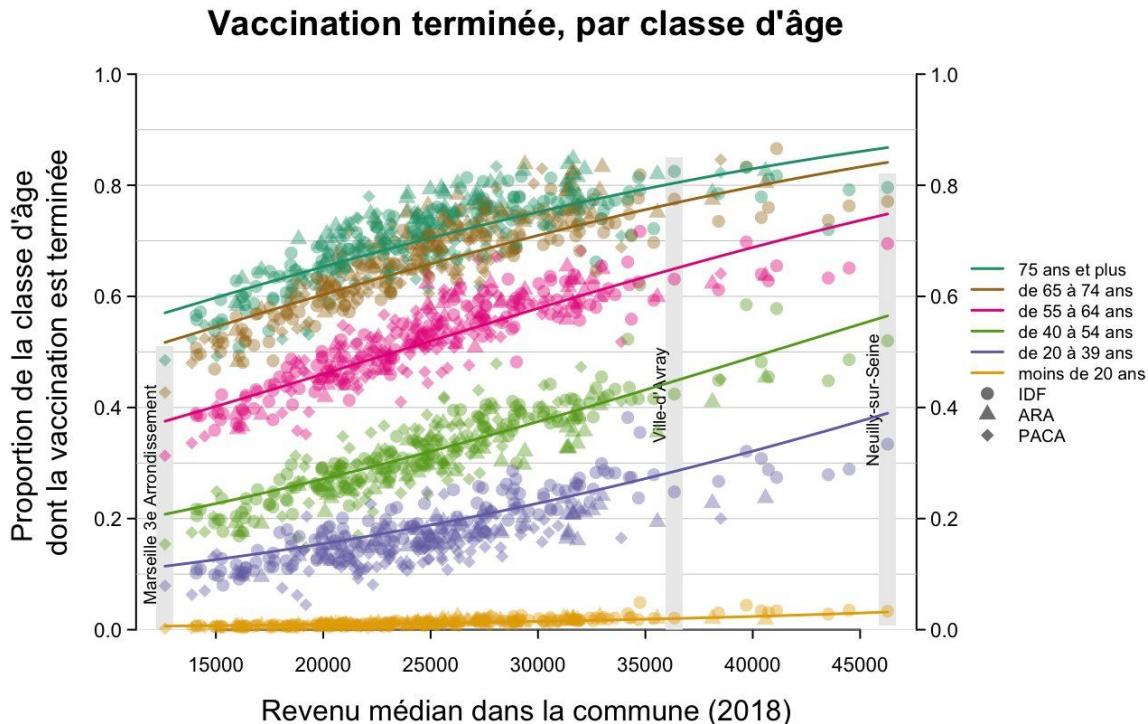


Figure @flodebarre, données vaccination trouvées par @humeursdevictor

Code: [https://github.com/flodebarre/covid\\_vaccination/blob/main/communes.Rmd](https://github.com/flodebarre/covid_vaccination/blob/main/communes.Rmd)

Données vaccination : <https://datavaccin-covid.ameli.fr/explore/dataset/donnees-de-vaccination-par-commune>

Données revenus : <https://www.insee.fr/fr/statistiques/5009236?sommaire=5009255#consulter>

# Troisième expérience

- Une autre métrique.
- La corrélation est un peu moins forte ici.
- Mettre des lignes peut être considéré comme légèrement trompeur car il "cache" la dispersion des données.

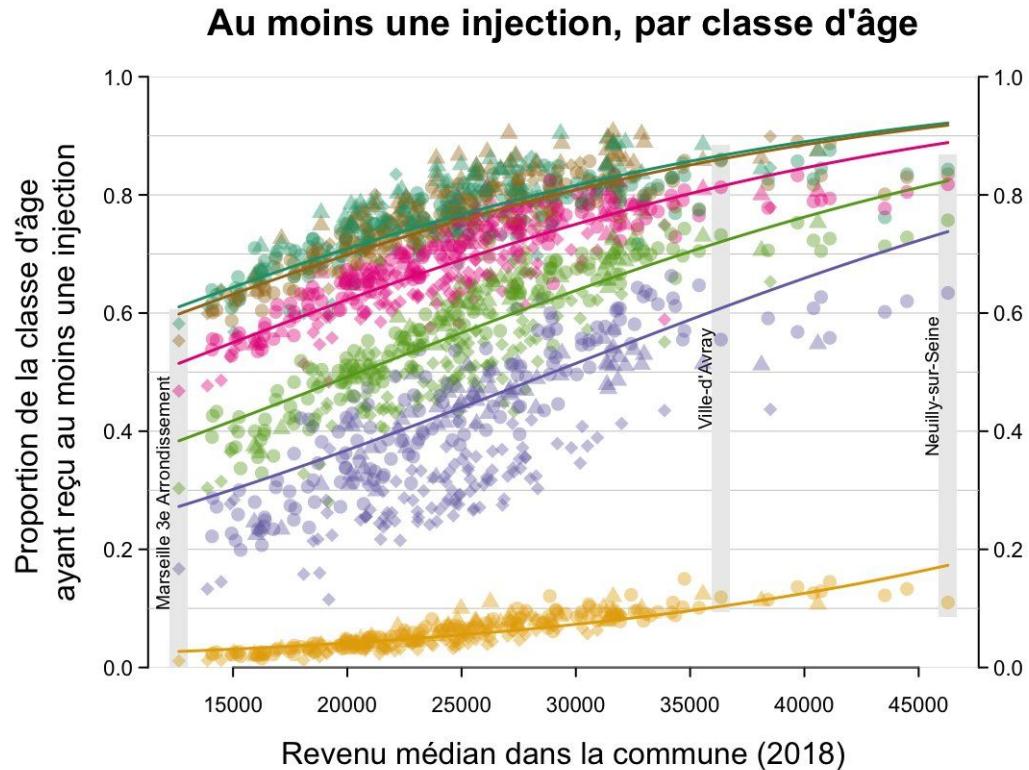


Figure @flodebarre, données vaccination trouvées par @humeursdevictor

Code: [https://github.com/flodebarre/covid\\_vaccination/blob/main/communes.Rmd](https://github.com/flodebarre/covid_vaccination/blob/main/communes.Rmd)

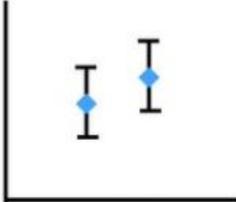
Données vaccination : <https://datavaccin-covid.ameli.fr/explore/dataset/donnees-de-vaccination-par-commune>

Données revenus : <https://www.insee.fr/fr/statistiques/5009236?sommaire=5009255#consulter>

# Un dernier exemple

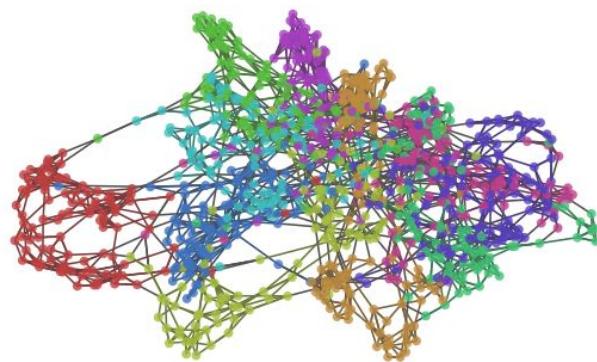
Eugene Vinitsky @EugeneVinitsky · May 6 ...

What if we just didn't bold the overlapping distributions

Normal Science	Machine Learning				
	<table border="1"><tr><td>Their method</td><td><math>25 \pm 100</math></td></tr><tr><td>Our method</td><td><b><math>40 \pm 100</math></b></td></tr></table>	Their method	$25 \pm 100$	Our method	<b><math>40 \pm 100</math></b>
Their method	$25 \pm 100$				
Our method	<b><math>40 \pm 100</math></b>				
Guess we don't have statistical significance	We made the big number <b>bold</b>				

# Exemples en ML

# Visualisation MNIST



# Visualisation des dynamiques d'apprentissage

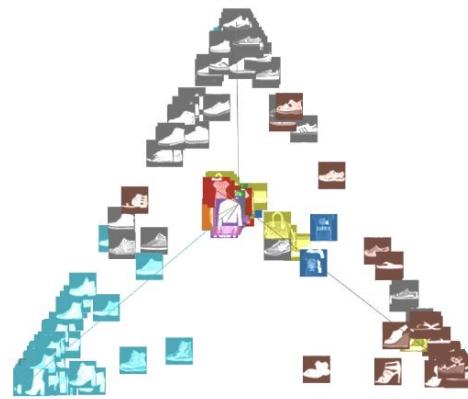
March 16, 2020

PEER-REVIEWED

## Visualizing Neural Networks with the Grand Tour

Mingwei Li, Zheng Zhao, and Carlos Scheidegger

By focusing on linear dimensionality reduction, we show how to visualize many dynamic phenomena in neural networks.



# Conclusion

- Visualisez toujours vos données !
- Exploration de données :
  - Commencez par une question
  - L'interactivité compte !
  - Les visualisations doivent être reproductibles!
- Présentation des données
  - Répondre à une question par visualisation
  - Avoir un récit clair
  - Avoir un "comment lire ceci"
  - Visualisez l'aléatoire ! (Barres d'erreur)

# Réfs utilisées pour construire ce cours

- [Zico Kolter's course on Data Science.](#)
- [Information Visualization Manifesto](#)
- [Data Humanism, The Revolution will be Visualized.](#)
- Previous DS courses [Golnoosh Faradi](#), [Kris Sankaran](#) and [Jhelum Chakravorty](#)