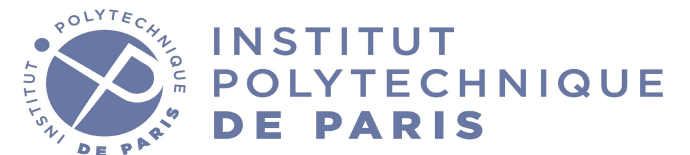


# **WELCOME**

**to SICSS-Paris**

**SICSS**

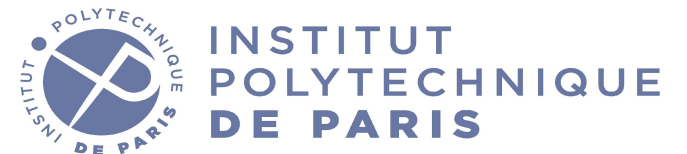


# Logistics

Wifi: SICSS / SicssParis2022

## **This morning:**


- Introduction: What are CSS?
- Presentations/Schedule/Logistics



# What are Computational Social Sciences?

**SICSS**

 CREST

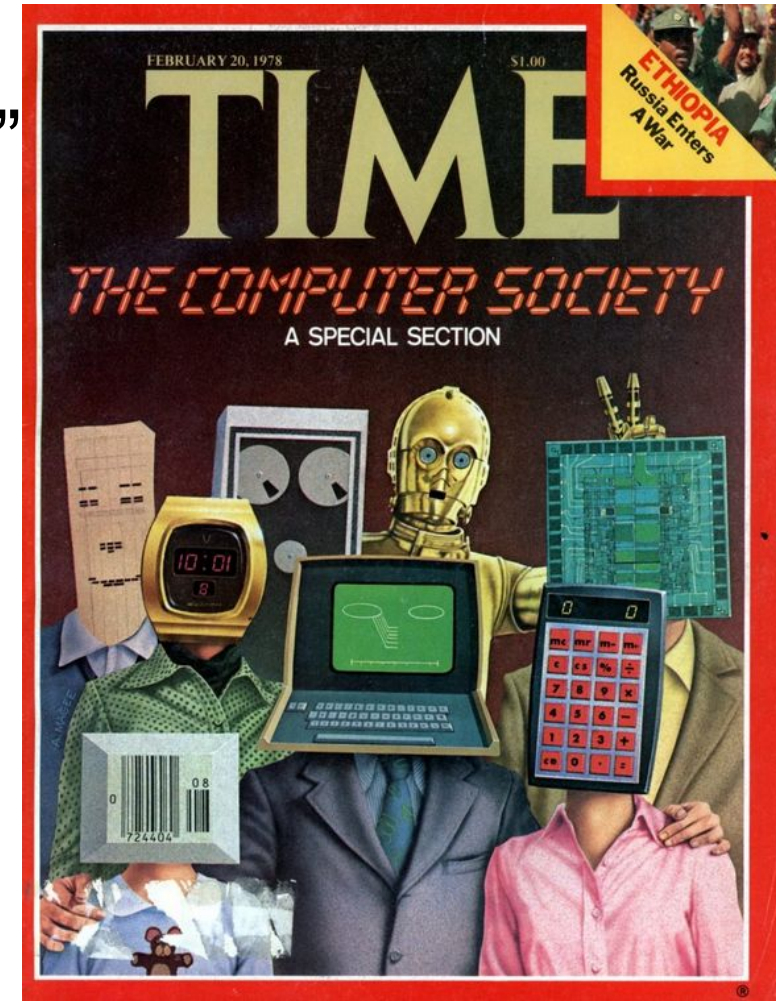
 INSTITUT  
POLYTECHNIQUE  
DE PARIS

# Computational Social Sciences


- Harnessing the power of computers to carry out (social) scientific research
- An interdisciplinary endeavor
- A set of methods more than a field

# Computational Social Sciences

- Computers? Nothing new!  
“Demography has long used computer data”  
(Kashyap, 2021)
- “But since the 2000s, two changes”
  - Data abundance
  - Computer power



# What are Computational Social Sciences?

- The same remark holds true in many disciplines: the combination of more data and more power did create **massive opportunities** for research.
  - CSS are trying to make the most of this new social & technical configuration.
- 



# Doing Research in the Big Data Era

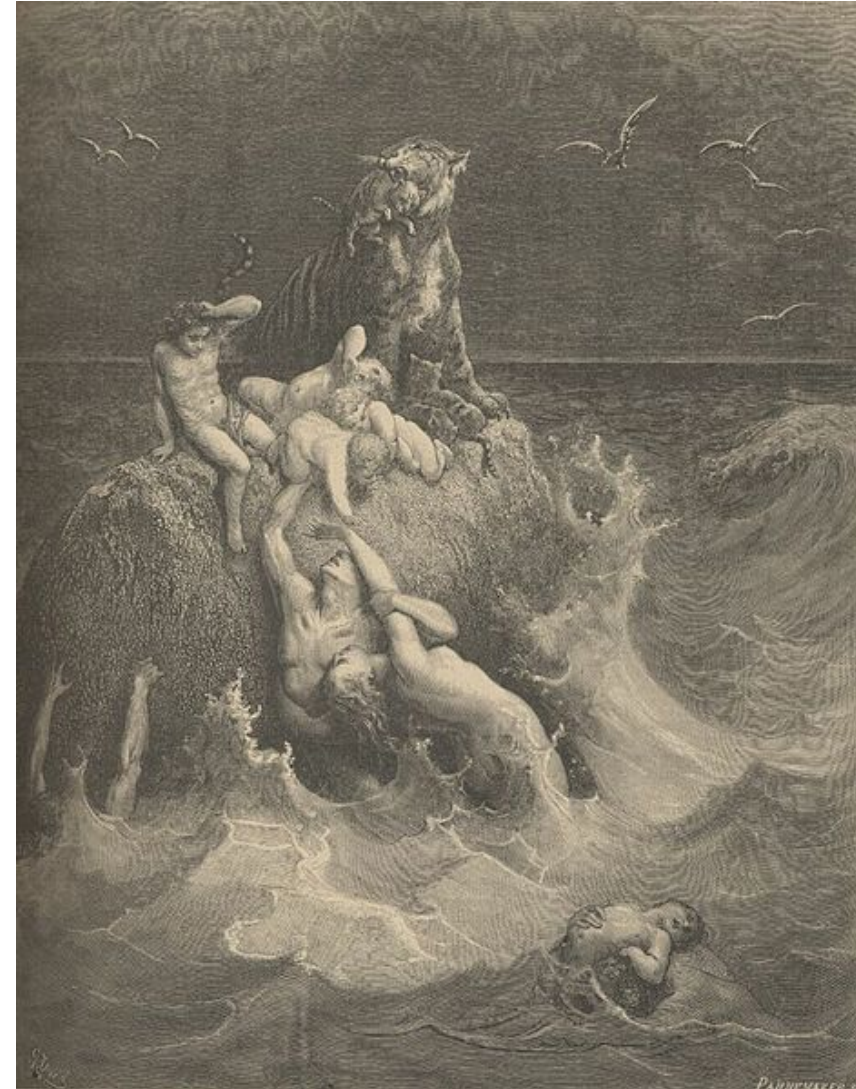
Is more data better? Polarized views





# Doing Research in the Big Data Era

Is more data better? Polarized views





# A Threefold Promise

## Empirical

- Investigating old questions anew
- Asking new questions

# A Threefold Promise

Empirical

Ex. Studying couple formation in the 21<sup>st</sup> century

# A Threefold Promise

Empirical

Ex. Studying couple formation in the 21<sup>st</sup> century

The image shows a screenshot of a French dating website profile. The profile is for a 43-year-old male from Paris, France, who is seeking a female partner aged 30 to 40. The profile includes various sections such as 'Mes films', 'Mes musiques', 'Mes livres', 'Mon profil', and 'Quelques mots sur moi'. Annotations with arrows point to specific parts of the profile:

- Age**: Points to the age '43 ans'.
- City**: Points to the location 'PARIS, Ile-de-France (France)'.
- Gender, age sought for**: Points to the text 'Je cherche une femme de 30 à 40 ans'.
- Presentation of self**: Points to the 'Quelques mots sur moi' section.
- Personal details**: Points to the 'Mon profil' section.
- Cultural taste**: Points to the 'Mes films', 'Mes musiques', and 'Mes livres' sections.

The profile also includes a 'Revenir' button, a 'Suivant' button, and a 'Profils similaires' section. The user's name is 'mb435' and they have a 20% completion rate. The website has a navigation bar with links like 'Onlines', 'Recherche', 'Daily 6', 'Echanges', 'Les Events', 'Le Mag', and 'mb435'.

# A Threefold Promise

## Methodological

- A Classic Distinction: Sample, and Population
- But what if you could collect all data in the time it takes to collect a sample?
- If your data is digitized, well-organized, this could be possible

# A Threefold Promise

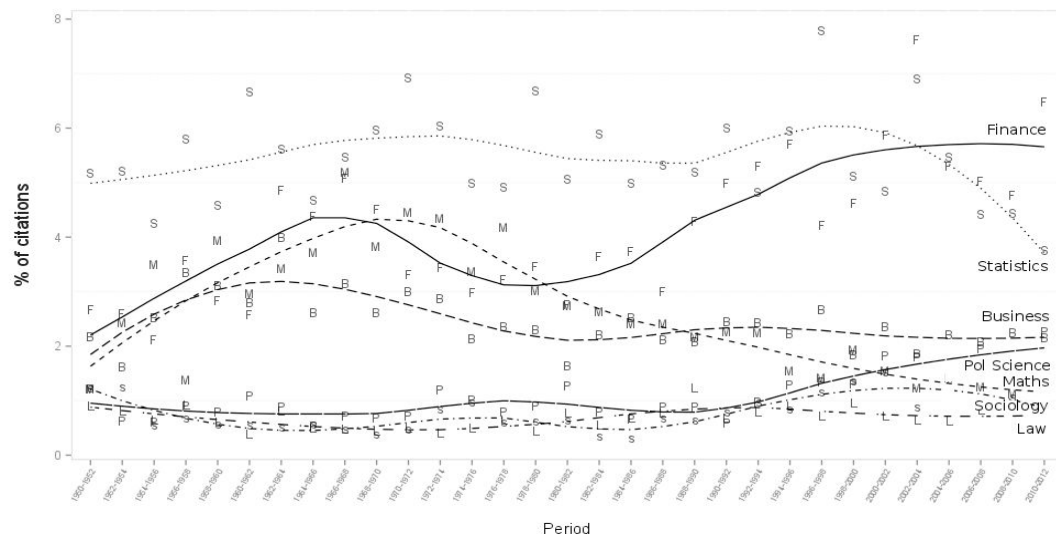
## Methodological

Ex. A study on interdisciplinarity in Economics

Idea: Looking at citations

Data: 5 journals, over a century

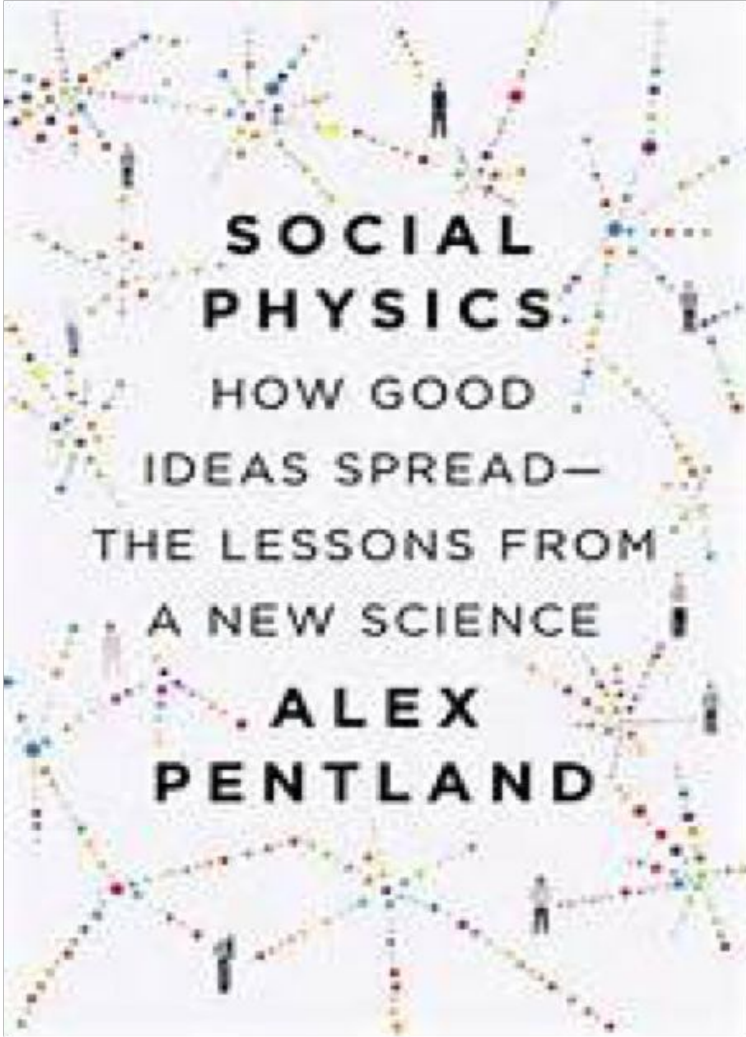
- 24,500 articles
- 600,000 references



in Fourcade, Ollion & Algan, "The Superiority of Economists," *Journal of Economic Perspectives*, 2015.

# A Threefold Promise

## Theoretical



Science

Current Issue First release papers

HOME > SCIENCE > VOL. 331, NO. 6014 > QUANTITATIVE ANALYSIS OF CULTURE USING MILLIONS OF DIGITIZED I

RESEARCH ARTICLE

## Quantitative Analysis of Culture Using Millions of Digitized Books

JEAN-BAPTISTE MICHEL, YUAN KUI SHEN, AVIVA PRESSER AIDEN, ADRIAN VERES, MATTHEW K. GRAY, THE GOOGLE BOOKS TEAM, DAN CLANCY, EREZ LIEBERMAN AIDEN +5 authors [Authors Info & Affiliations](#)

SCIENCE • 16 Dec 2010 • Vol 331, Issue 6014 • pp.176-182 • DOI: 10.1126/science.1199644

2,535 1,124

### Abstract

We constructed a corpus of digitized texts containing about 4% of all books ever printed. Analysis of this corpus enables us to investigate cultural trends quantitatively. We survey the vast terrain of ‘culturomics,’ focusing on linguistic and cultural phenomena that were reflected in the English language between 1800 and 2000. We show how this approach can provide insights about fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of new technology, the pursuit of fame, censorship, and historical epidemiology. Our work extends the boundaries of rigorous quantitative inquiry to a wide array of phenomena spanning the social sciences and the humanities.



# A Threefold Promise **BUT**

Lots and lots of NA

Who doesn't like reading, hiking, seeing their friends?

Everyone's taller and younger than they should be

The screenshot shows a profile on a French dating site. The user is a 43-year-old man from Paris, France, seeking a woman aged 30-40. His profile includes various attributes: height (169 cm), hair (brun), build (normal), no children, and a conciliatory personality. He lists hobbies like car and boxing. The profile also features sections for favorite movies, music, and books, as well as a 'Mon profil' section with more personal details like nationality (French), language (French), and eye color (blue). The interface includes navigation tabs at the top and various promotional banners on the right side.

Attribute	Value
Age	43 ans
Location	PARIS, Ile-de-France (France)
Seeking	Je cherche une femme de 30 à 40 ans
Ma taille	169 cm
Cheveux	brun
Silhouette	normale
Enfants	aucun
Mon caractère	conciliant
Statut	---
Profession	pub/média ...
Hobbies	bricolage, ...
Activités sportives	automobile, boxe, ...
Je fume	oui, occasionnellement

Attribute	Value
Statut	---
Ma nationalité	française
Mon origine ethnique	européenne
Je parle	français
Je vis	---
Ma taille	169 cm
Mon poids	---
Ma longueur de cheveux	courts
Mes yeux	bleus
Mon style	---
Mon aspect physique	---

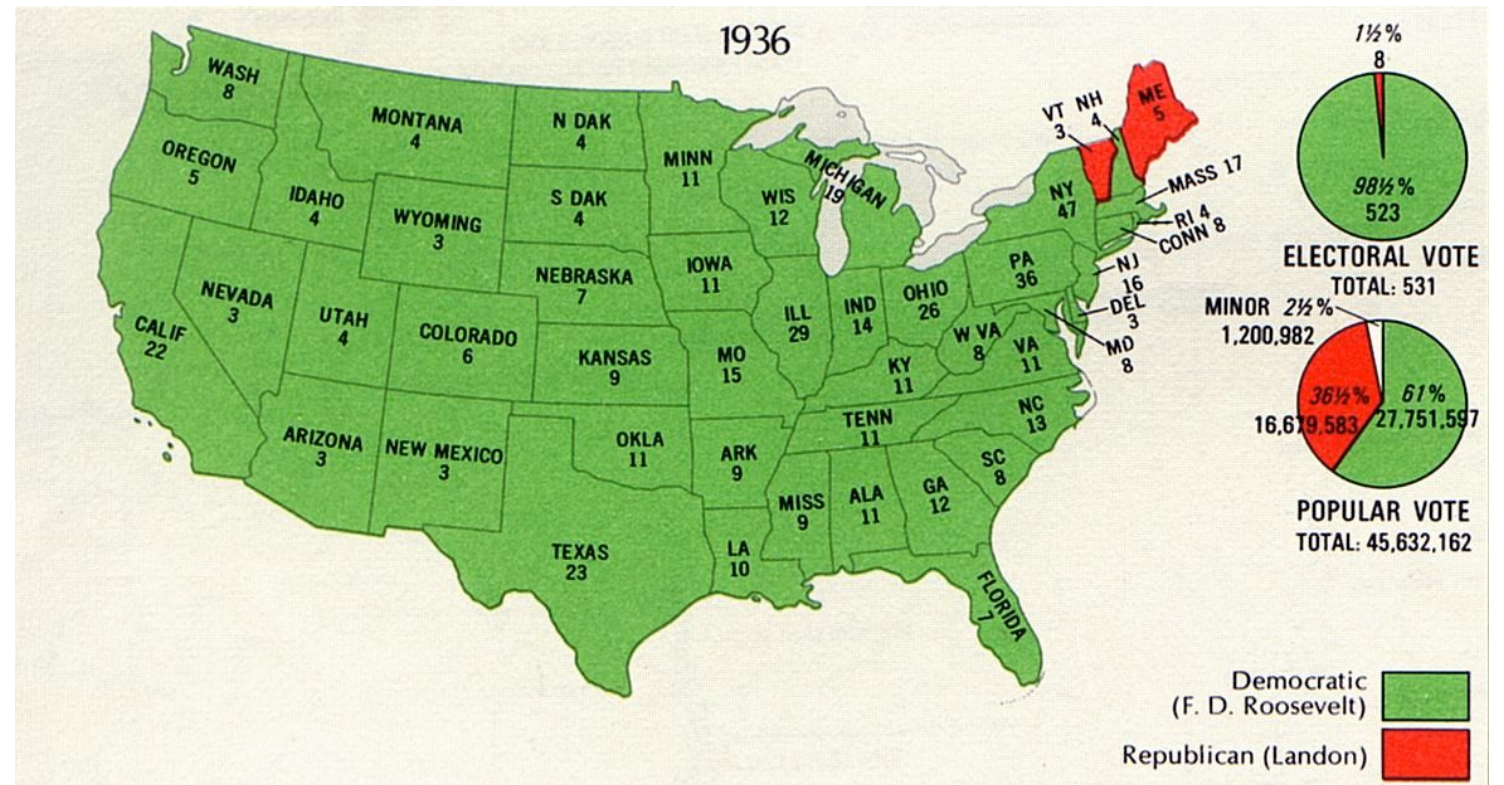
Is the astro sign really useful?



# A Threefold Promise

## Methodological

N = All, a rather rare case



# A Threefold Promise

## Theoretical

- Promises are still in the waiting
- The lessons of the history of science need to be heeded
  - “Social physics”?
  - Does more data lead to more “theory”?



CSS: making the most of the new opportunities, while heeding the lessons of the past.





# Making the Most of these Opportunities

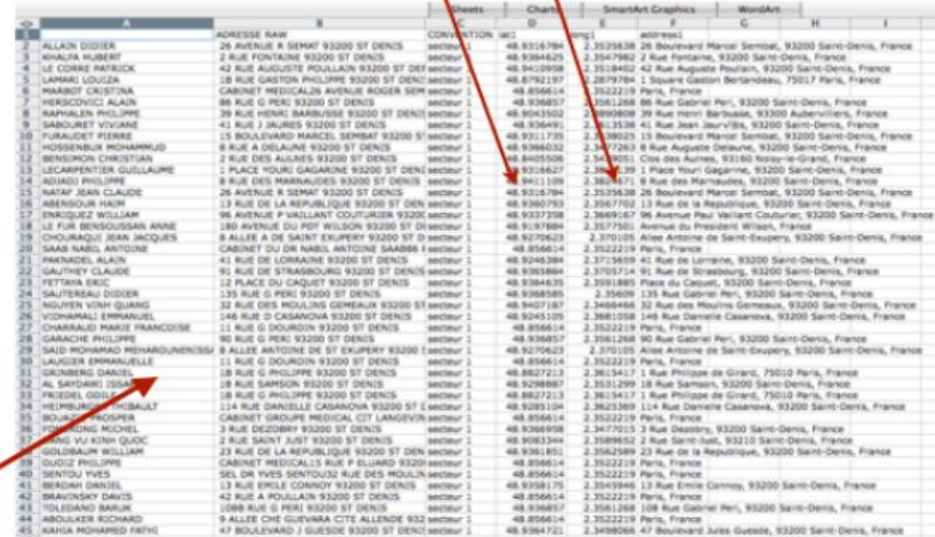
## Collecting data

Ex. Medical density / localization in a given town

Geocoding  
(Lat + Lon)



	A	B	C
1	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
2	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
3	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
4	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
5	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
6	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
7	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
8	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
9	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
10	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
11	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
12	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
13	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
14	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
15	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
16	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
17	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
18	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
19	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
20	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
21	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
22	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
23	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
24	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
25	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
26	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
27	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
28	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
29	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
30	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
31	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
32	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
33	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
34	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
35	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
36	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
37	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
38	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
39	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
40	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
41	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
42	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
43	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
44	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1
45	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1



	A	B	C	D	E	F	G	H	I
1	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9324784	2.3525438	26 Boulevard Marcel Sembat, 93200 Saint-Denis, France			
2	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9384425	2.3547982	2 Rue Fontaine, 93200 Saint-Denis, France			
3	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9420958	2.3518452	42 Rue Auguste Poulain, 93200 Saint-Denis, France			
4	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9392187	2.3527954	1 Square Gaston Bertrand, 75017 Paris, France			
5	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.936614	2.3522219	Paris, France			
6	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9368557	2.351268	86 Rue Gabriel Peri, 93200 Saint-Denis, France			
7	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9343502	2.350808	39 Rue Henri Barbusse, 93200 Aubervilliers, France			
8	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9364892	2.3513538	41 Rue Jean Jaurès, 93200 Saint-Denis, France			
9	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9311735	2.3518025	15 Boulevard Marcel Sembat, 93200 Saint-Denis, France			
10	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9384032	2.347263	8 Rue Auguste Delaune, 93200 Saint-Denis, France			
11	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9365506	2.3518051	Cité des Aulnes, 93180 Noisy-le-Grand, France			
12	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9316427	2.3513539	1 Place Youri Gagarine, 93200 Saint-Denis, France			
13	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9411109	2.3521671	8 Rue des Mandoules, 93200 Saint-Denis, France			
14	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9324784	2.3525438	26 Boulevard Marcel Sembat, 93200 Saint-Denis, France			
15	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9365506	2.3518051	13 Rue de la République, 93200 Saint-Denis, France			
16	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.9337358	2.3469157	96 Avenue Paul Vaillant Couturier, 93200 Saint-Denis, France			
17	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
18	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
19	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
20	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
21	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
22	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
23	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
24	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
25	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
26	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
27	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
28	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
29	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
30	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
31	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
32	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
33	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
34	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
35	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
36	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
37	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
38	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
39	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
40	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
41	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
42	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
43	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
44	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			
45	ALLAIN DIDIER	26 AVENUE R DENAT 93200 ST DENIS	secteur 1	48.937884	2.3577501	Avenue du Président Wilson, France			





# Making the Most of these Opportunities

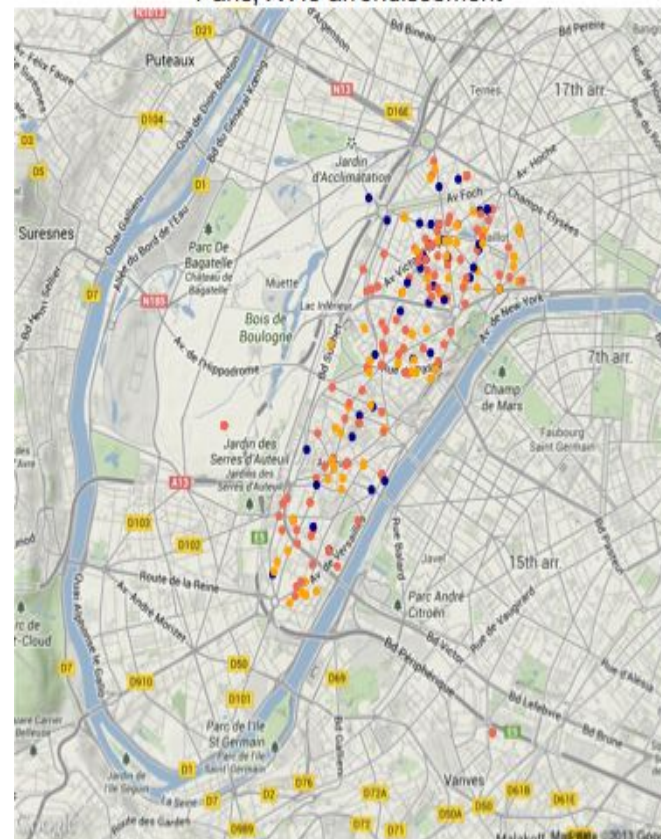
## Collecting data

Ex. Medical density / localization in a given town

Médecins généralistes en exercice libéral  
Saint-Denis (93)



Médecins généralistes en exercice libéral  
Paris, XVI<sup>e</sup> arrondissement





# Making the Most of these Opportunities

## Collecting data, sometimes in real time

### Nowcasting

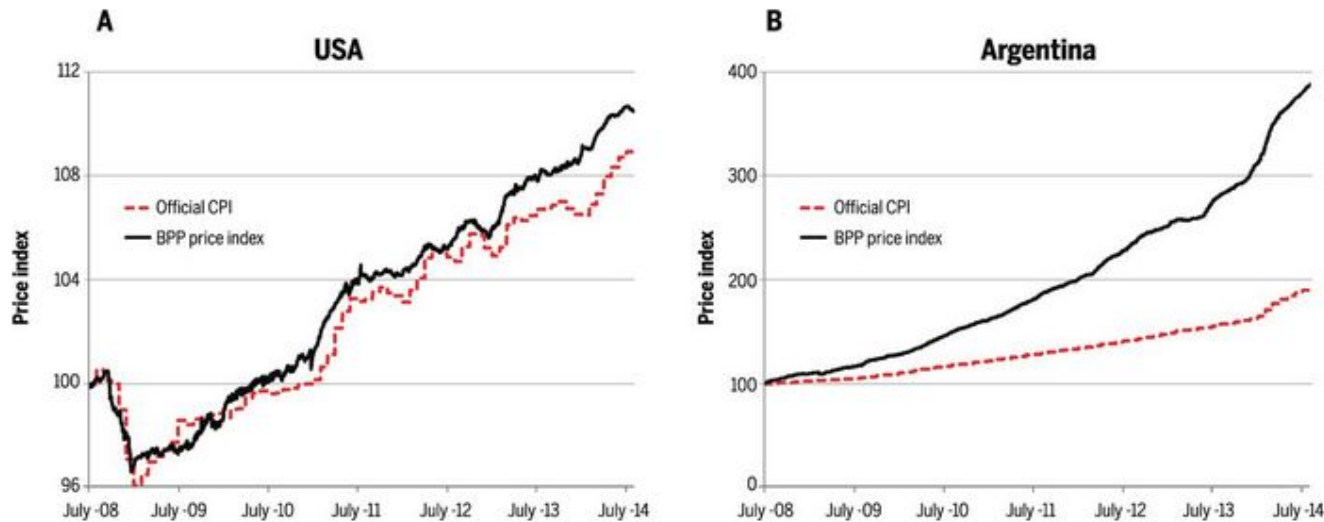


Fig. 2 BPP price index.

(Cavallo & Rigobon, Billion Prices Project, 2011+)

Science

Current Issue First release papers Archive About Submit ma

HOME > SCIENCE > VOL. 343, NO. 6176 > THE PARABLE OF GOOGLE FLU: TRAPS IN BIG DATA ANALYSIS

POLICY FORUM | BIG DATA

## The Parable of Google Flu: Traps in Big Data Analysis

Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.

DAVID LAZER, RYAN KENNEDY, GARY KING, AND ALESSANDRO VESPIGNANI [Authors Info & Affiliations](#)

SCIENCE • 14 Mar 2014 • Vol 343, Issue 6176 • pp. 1203-1205 • DOI: 10.1126/science.1248506

2,853 1,251



In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for



# Making the Most of these Opportunities

## Collecting data sometimes in real time

Producing data: [Experiments](#)

Science

[Current Issue](#)

[First release papers](#)

[Archive](#)

[About](#) ▼

[Submit manuscript](#)

[HOME](#) > [SCIENCE](#) > [VOL. 311, NO. 5762](#) > [EXPERIMENTAL STUDY OF INEQUALITY AND UNPREDICTABILITY IN AN ARTIFICIAL CULTURAL MARKET](#)



REPORTS

## Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market

[MATTHEW J. SALGANIK](#), [PETER SHERIDAN DODDS](#), AND [DUNCAN J. WATTS](#)





SCIENCE • 10 Feb 2006 • Vol 311, Issue 5762 • pp. 854-856 • [DOI: 10.1126/science.1121066](#)

# Making the Most of these Opportunities

## Collecting data sometimes in real time

Producing data: [mass collaboration](#)

## Measuring the predictability of life outcomes with a scientific mass collaboration

[Matthew J. Salganik](#) , [Ian Lundberg](#) , [Alexander T. Kindel](#),  [+108](#), and [Sara McLanahan](#)  [Authors Info & Affiliations](#)

Contributed by Sara McLanahan, January 24, 2020 (sent for review October 1, 2019; reviewed by Sendhil Mullainathan and Brian Uzzi)

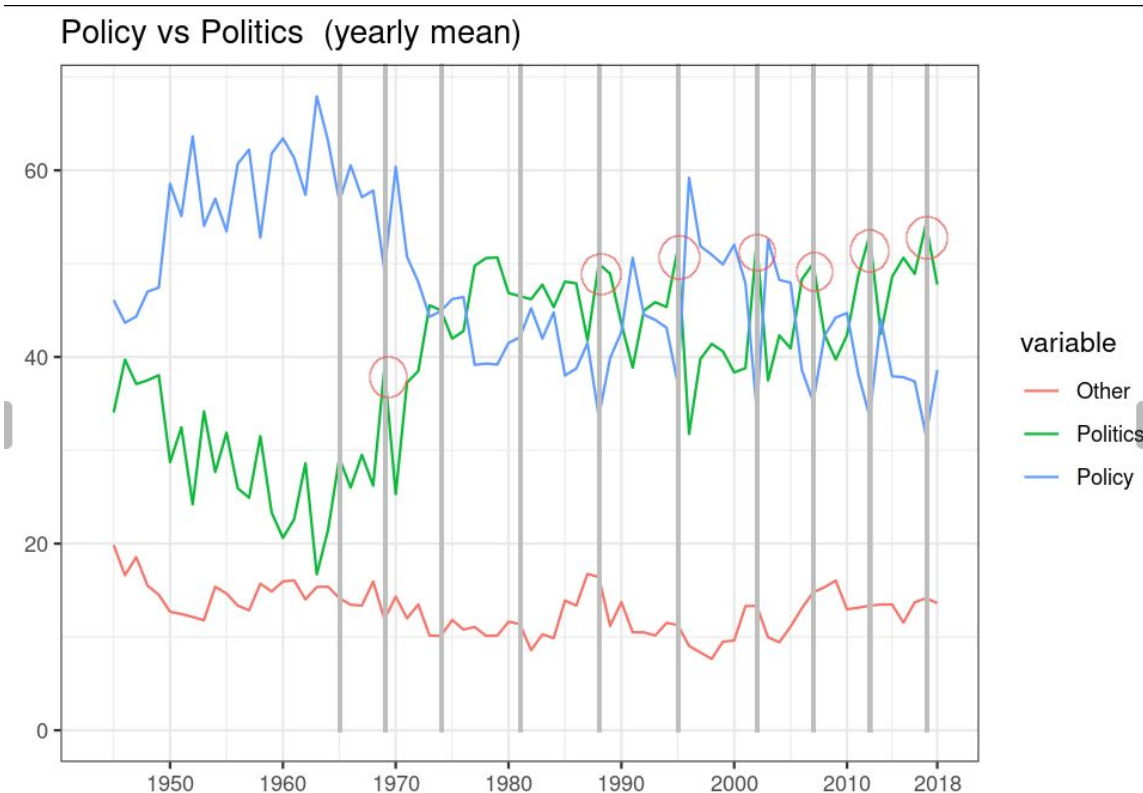
**March 30, 2020** | 117 (15) 8398-8403 | <https://doi.org/10.1073/pnas.1915006117>

---

# Making the Most of these Opportunities

## Collecting data sometimes in real time

Producing data: **artificial intelligence**



Do, Shen & Ollion, « The Augmented Social Scientist »

Using AI to extract sentence-levels forms of narration in politics

# Conclusion

- CSS: making the most of a socio-technological change while avoiding unfettered techno-enthusiasm

- A wide gamut of methods

==> C'est parti!

