



# L'abondance et ses revers

Que faire des données numériques en sciences sociales?

Les rencontres de la statistique appliquée, février 2021



**Les big data et les sciences sociales, allégorie**







**Expliciter les enjeux du débat**

**Des big data aux données numériques**

**Saisir les reconfigurations**



# Une triple promesse

## Empirique

- Reposer d'anciennes questions
- En aborder de nouvelles

# Une triple promesse

## L'étude de la formation du couple au XXIème siècle

Age

Ville

Genre, âge demandé

Présentation de soi

Détails personnels

Goûts, pratiques culturelles

Onlines 88 622 Recherche Daily 6 Echanges Les Events Le Mag mb435 20%

Mail 56 Visite 10 Chat 3 Flash 41 Favors 6 Daily 6 Recevez votre rapport de personnalité gratuit Abonnez-vous !

Retour Suivant

43 ans  
PARIS, Ile-de-France (France)  
Je cherche une femme de 30 à 40 ans

Ma taille : 169 cm Statut : --  
Cheveux : châtains Profession : pub/média ...  
Silhouette : normale Hobbies : bricolage, ...  
Enfants : aucun Activités sportives : automobile, boxe, ...  
Mon caractère : conciliant Je fume : oui, occasionnellement

Ecrire Chatter  
Flasher

Quelques mots sur moi

D'un naturel calme, ouvert d'esprit, patient et passionné. J'aime croquer la vie sans me la compliquer.  
Keep it simple & enjoy!!

Je suis quelqu'un de passionné, honnête et entreprenant. J'apprécie certaines valeurs à mon sens fondamentales telles que: la tolérance, la bienveillance, le courage, l'honnêteté, l'ouverture d'esprit...  
La vulgarité me fait fuir...  
Je ne recherche rien de très précis hormis de rencontrer de belles personnes ayant des valeurs communes dans un esprit sympathique, détendu et agréable...

Mon profil

Statut : --  
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 : --

Mes films  
eley\_for\_U n'a pas encore ajouté ses films préférés.  
Demandez lui ce qu'il aime

Mes musiques  
eley\_for\_U n'a pas encore ajouté ses musiques préférés.  
Demandez lui ce qu'il aime

Mes livres  
eley\_for\_U n'a pas encore ajouté ses livres préférés.  
Demandez lui ce qu'il aime

Répondez à mon Quiz  
eley\_for\_U n'a pas encore de Quiz

Envoyez des messages en illimité !  
Abonnez-vous

Des célibataires en affinité  
Multipliez vos chances de rencontres

Terminez votre test >>

Garmin Forerunner 910XT  
Le seul appareil GPS tout-en-un capable d'enregistrer des données.



# Une triple promesse

## Méthodologique

Une distinction classique: échantillon / population

Et si vous étiez capable de collecter aussi rapidement l'échantillon et la population

Quand les données sont numériques et bien organisées, c'est possible.

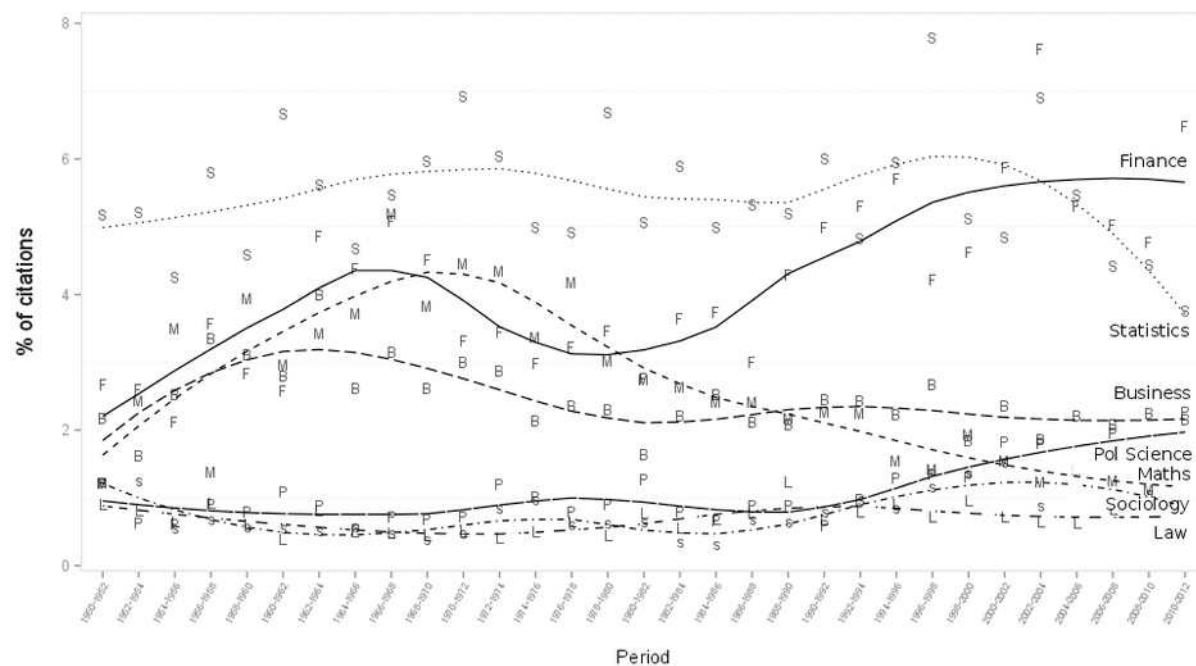


# Une triple promesse

## Méthodologique

Une étude sur l'interdisciplinarité en économie  
En partant des citations

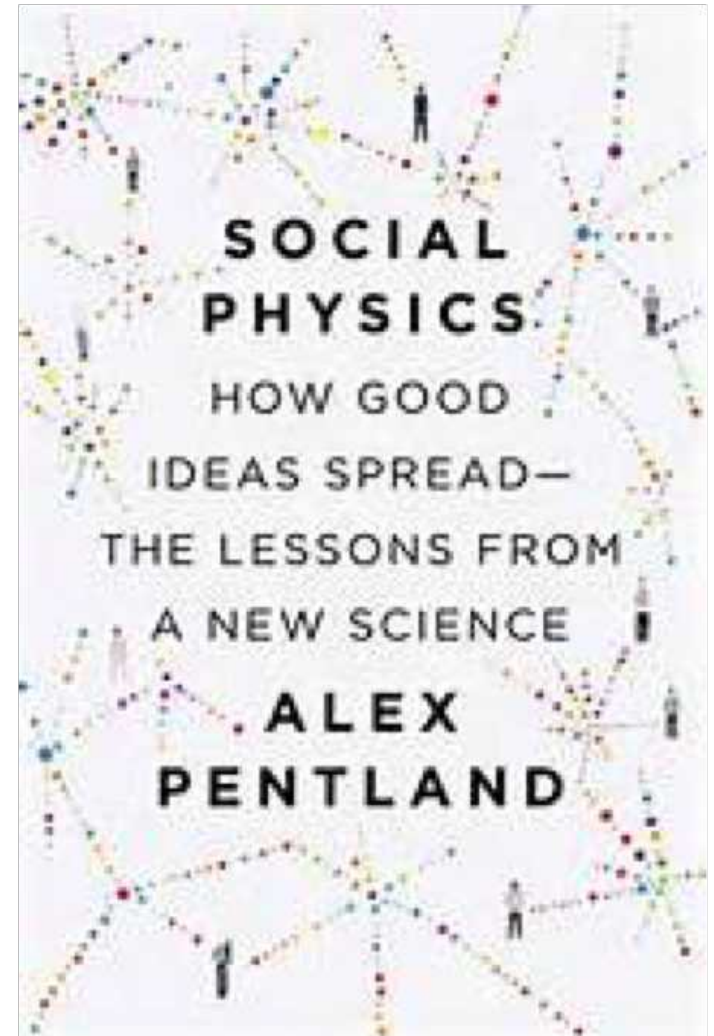
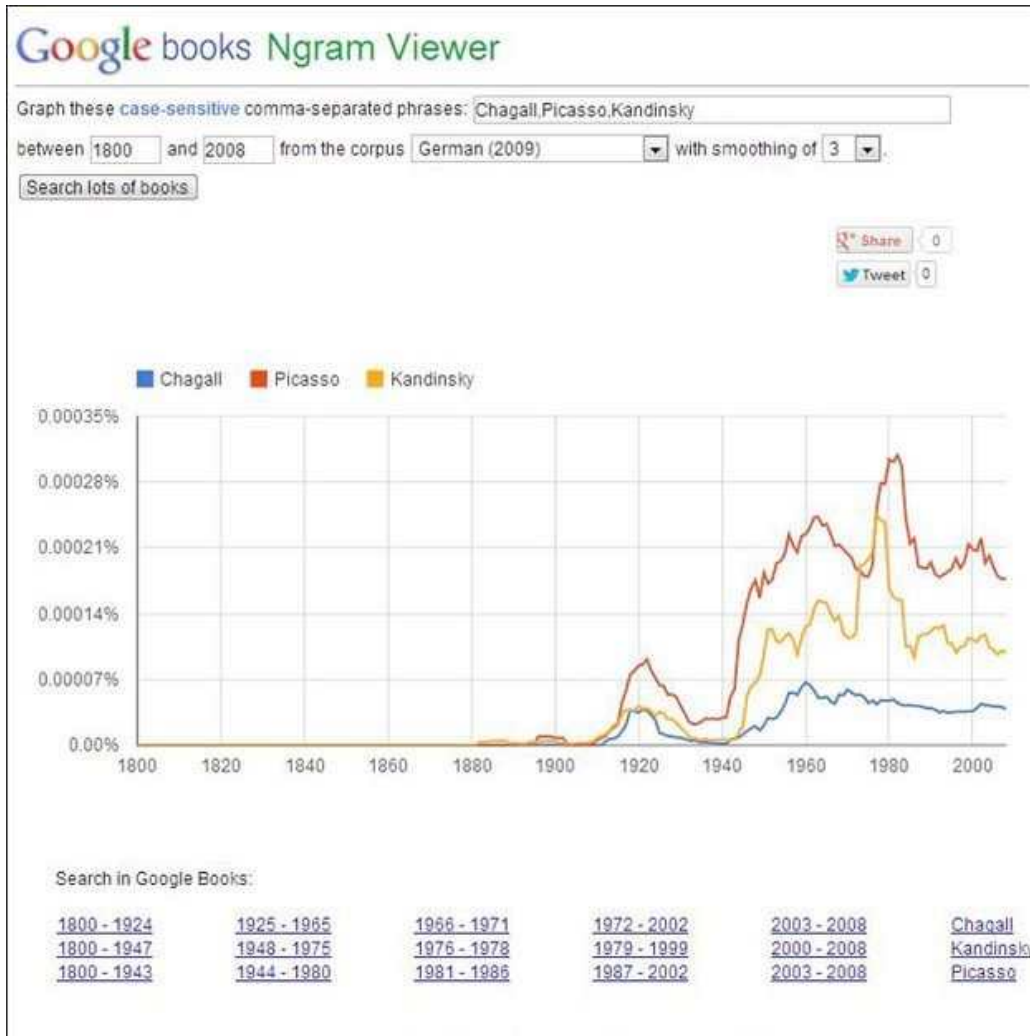
- Données : les 5 principaux journaux, sur un siècle
  - 24,500 articles
  - 600,000 citations



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

# Une triple promesse

## Théorique





# Une triple promesse, **mais** Empirique

# Une triple promesse, mais Empirique

Onlines 68 522 Recherche Daily 6 Echanges Les Events NEW Le Mag NEW mb435 20%

Mail 56 Visite 19 Chat Flash 3 Favoris 41 Daily 6 6 Recevez votre rapport de personnalité gratuit Abonnez-vous !

Retour Suivant

Online Profils similaires

43 ans PARIS, Ile-de-France (France) Je cherche une femme de 30 à 40 ans

Ma taille : 169 cm Statut : --

Cheveux : châtains Profession : pub/média ...

Silhouette : normale Hobbies : bricolage, ...

Enfants : aucun Activités sportives : automobile, boxe, ...

Mon caractère : conciliant Je fume : oui, occasionnellement

Ecrire Flasher Chatter

Quelques mots sur moi

D'un naturel calme, ouvert d'esprit, patient et passionné. J'aime croquer la vie sans me la compliquer. Keep it simple & enjoy!!

Je suis quelqu'un de passionné, honnête et entreprenant. J'apprécie certaines valeurs à mon sens fondamentales telles que: la tolérance, la bienveillance, le courage, l'honnêteté, l'ouverture d'esprit... La vulgarité ne fait fuir... Je ne recherche rien de très précis hormis de rencontrer de belles personnes ayant des valeurs communes dans un esprit sympathique, détendu et agréable...

Mon profil

|                          |            |
|--------------------------|------------|
| 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 :    | --         |

Envoyez des messages en illimité ! Abonnez-vous !

Des célibataires en affinité Multipliez vos chances de rencontres

Terminez votre test >>>

Garmin Forerunner 910XT le seul appareil GPS tout-en-un capable d'accumuler des données.

Nombre de NAs

Vous connaissez quelqu'un qui n'aime pas s'amuser, voir ses amies?

Tout le monde est plus jeune, plus grand sur les sites de rencontre

Que faire du signe astrologique en science?

# Une triple promesse, **mais**

## Empirique

L'ère actuelle est riche en données “organiques”  
(Groves, 2011)

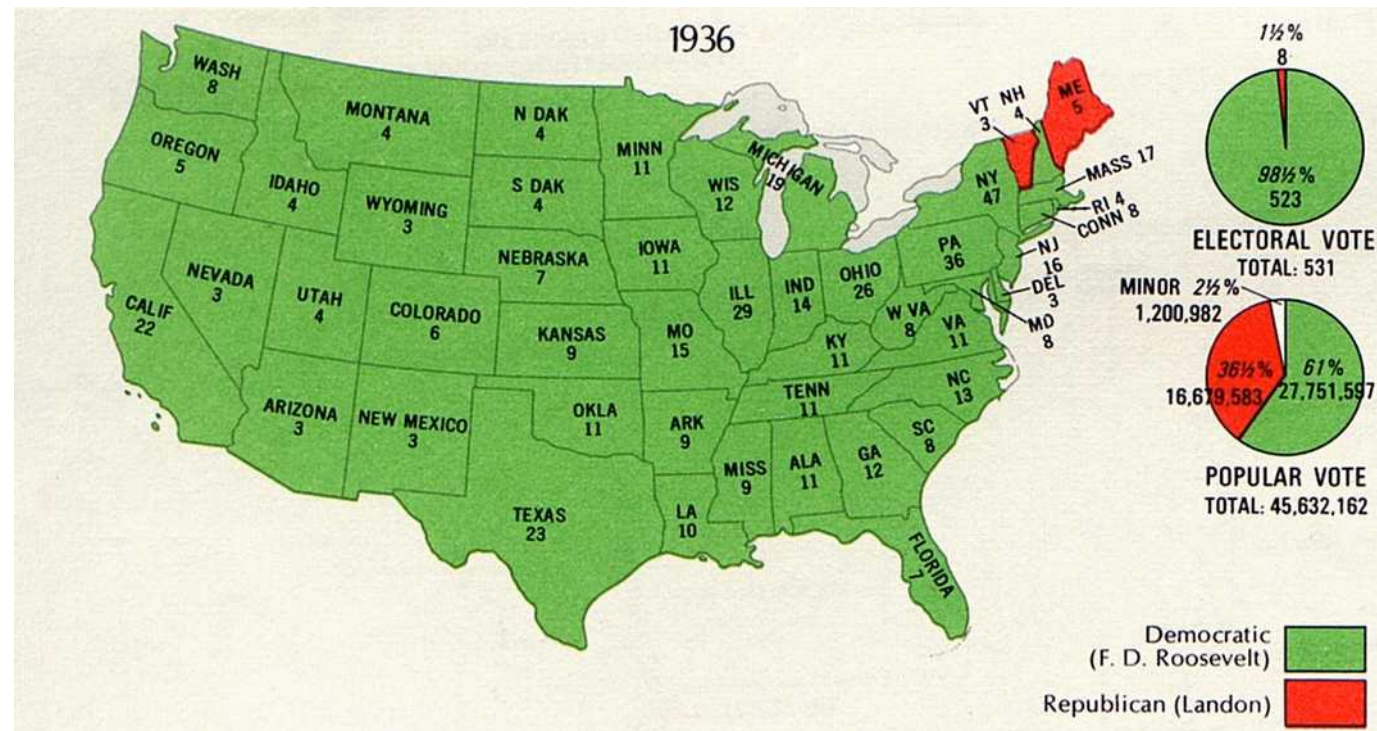
“Organic” vs “Design” data

==> Big data ne veut pas dire “rich” data

# Une triple promesse, **mais**

## Méthodologique

N=All, un cas finalement pas si fréquent



# Une triple promesse, **mais**

## Théorique

- Des promesses qui se font attendre
- Quelques leçons de l'histoire des sciences
  - Le retour de la “physique sociale” nous sauvera t-il?
  - Est-ce plus de données veut dire plus de “théorie”?



# Des 'big data' aux données numériques



# Des 'big data' aux données numériques

Archives numérisées  
(Google Books)

Enquêtes massives en  
ligne (questionnaires)



Données de capteurs

Traces de navigation en  
ligne

Bases de données  
(open data, wikileaks,  
d'entreprise)

etc

# Des 'big data' aux données numériques

A partir de quand les données sont-elles massives?



# Des 'big data' aux données numériques

Il faut aller au-delà des “big data”

> La période actuelle a vu se multiplier les **données numériques** (petites ou grandes, nativement numériques ou pas, organiques ou pas, formats différents).

Les mérites de la clarification :

- > Ne pas succomber aux discours prophétiques
- > Des données comme les autres
- > Des transformations déjà en cours



# Reconfigurations

## 1. L'interdisciplinarité, jusqu'où?

- Se former, ou collaborer?
- Reconfigurations disciplinaires : les sciences sociales computationnelles

# Reconfigurations

## 2. Accès et propriété

> Qui peut plonger dans le “data lake”?



> Accès aux données, RGPD: des données abondantes, mais pas forcément disponibles

# Reconfigurations

## 3. Le renouveau de l'intelligence artificielle

Le retour en grâce d'une initiative ancienne

L'IA pour **analyser des données**

- Supervised ↔ Méthodes paramétriques
- Unsupervised ↔ Réduction de dimensionalité

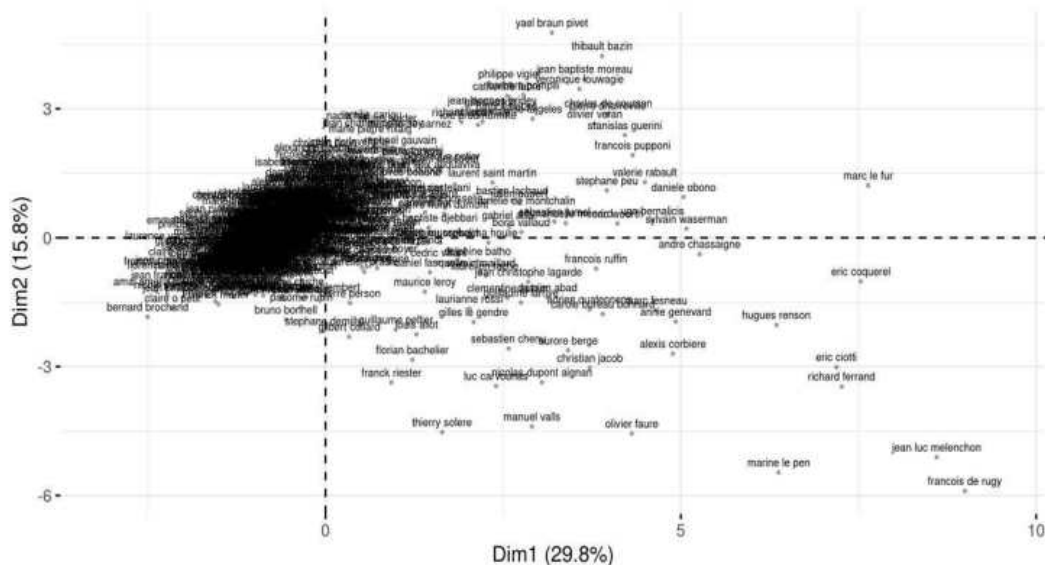
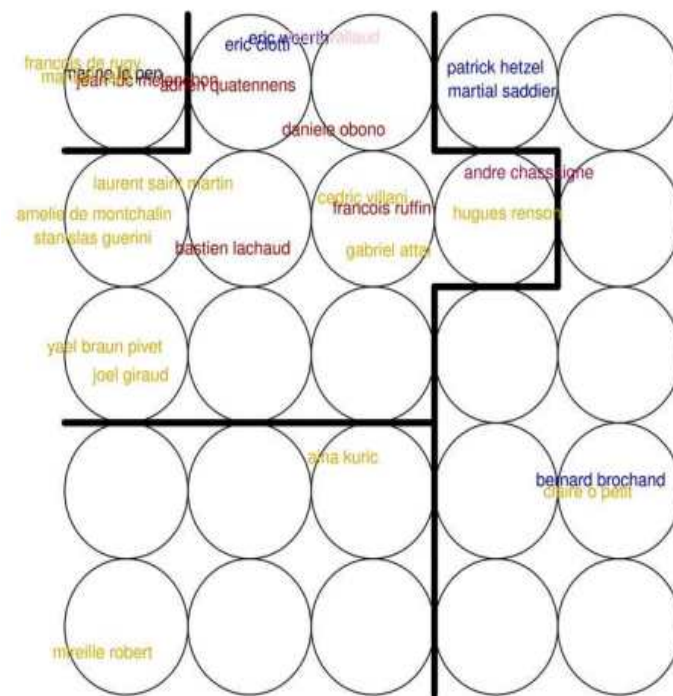


Tableau 1 : Indicateurs de représentation des variables sur l'axe 1 de l'ACP capital politique



# Reconfigurations

## 3. Le renouveau de l'intelligence artificielle

Le retour en grâce d'une initiative ancienne

L'IA pour **produire des données**

Ex. retrouver des schèmes journalistiques

Selon une source proche du pouvoir xxxx

Un élu du premier cercle affirme que xxx

En off, un proche du Président déplore que xxx

# Reconfigurations

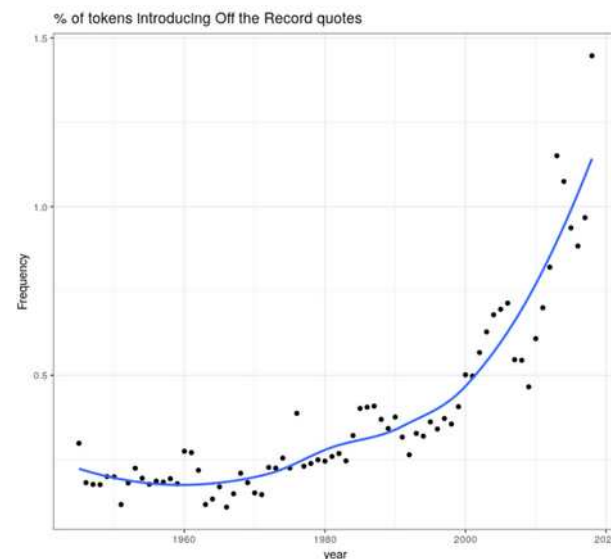
## 3. Le renouveau de l'intelligence artificielle

Le retour en grâce d'une initiative ancienne

L'IA pour **produire des données**

Ex. retrouver des schèmes journalistiques

- Avant: annotation manuelle (et sélective)
- Maintenant: entraînement d'un algorithme, puis généralisation







# Conclusion

“The human quest for knowledge and the academic labor are impeded by the numerous obstacles, the first of which is the abundance of existing works”

# Conclusion

“The human quest for knowledge and the academic labor are impeded by the numerous obstacles, the first of which is the abundance of existing works”



Ibn Khaldun, Muqaddimah (1377  
[1969], p. 414)

# Conclusion

## Références

- Ollion et Boelaert, “Au-delà des big data. Les sciences sociales face à la multiplication des données numériques”, in *Sociologie*, 2015.

## Voir aussi

- Boelaert et Ollion, “The Great Regression. Econometrics, Machine Learning, and the Future of Quantitative Social Sciences”, in *Revue Française de Sociologie*, 2018
- Boelaert & Ollion, “Les sommets du palais. Analyser l’espace social avec des cartes auto-organisatrices”, in *Revue Française de Science Politique*, 2020
- Fourcade, Ollion & Algan, “The Superiority of Economists”, in *Journal of Economic Perspectives*, 2015
- Groves, “Three eras of Survey Research”, *Public Opinion Research*, 2011.
- Ollion et Do, “Learning to Analyze, Learning to Produce”, AI & Social Sciences workshop, Institute for Advanced Studies, 2020.