

Seth Stephens-Davidowitz  
Préface de Steven Pinker

# TOUT LE MONDE MENT...

(et vous aussi!)

Best-seller du  
*New York Times*

**Internet et le Big Data :  
ce que nos recherches  
Google disent vraiment  
de nous**



« *Infiniment fascinant* »

Steven Pinker, auteur de *La Part d'ange en nous*

A L I S I O

# TOUT LE MONDE MENT... (et vous aussi!)

**Les parents favorisent-ils les garçons plutôt que les filles ? L'école que l'on fréquente influence-t-elle réellement notre réussite ? Mentons-nous souvent sur notre sexualité ?**

En ce début de <sup>xxi</sup> siècle, nos navigations Internet génèrent 8 millions de giga-octets de données chaque jour. Une quantité stupéfiante d'informations qui en disent long sur chacun d'entre nous, car c'est à notre moteur de recherche que nous confions nos questions et nos pensées les plus intimes.

S'appuyant sur les recherches Google de millions d'Américains, Seth Stephens-Davidowitz nous livre des révélations fascinantes sur des sujets aussi divers que la politique, le racisme, la violence, notre sexualité, nos relations familiales, le sport...

À la lumière de ce « sérum de vérité numérique » qu'est le Big Data, il nous invite à lever le voile sur nous-mêmes et à prendre conscience de nos préjugés... pour mieux changer.

« Une plongée inédite dans la psyché humaine [...].  
Infiniment fascinant. »

**Steven Pinker, auteur du best-seller *La Part d'ange en nous***

« Une exploration incroyablement intelligente et malicieuse  
de ce que le Big Data nous révèle de notre quotidien.  
Seth Stephens-Davidowitz est le meilleur conteur  
de données que j'aie jamais rencontré. »

**Steven Levitt, co-auteur du best-seller *Freakonomics***

« Les conclusions empiriques présentées dans  
*Tout le monde ment* sont si intrigantes que ce livre serait  
passionnant même s'il était structuré comme une longue liste. »

**The Economist**

Docteur en économie diplômé de Harvard et ancien salarié de Google, **Seth Stephens-Davidowitz** est le premier à avoir analysé nos recherches Google pour mesurer des phénomènes comme le racisme, l'avortement auto-induit, la maltraitance des enfants, les préférences et l'insécurité sexuelles... Traduit dans 13 pays, *Tout le monde ment* est déjà un best-seller.

**Steven Pinker** est professeur de psychologie à Harvard et reconnu dans le monde entier pour ses recherches sur la cognition et la psychologie du langage. Il est l'auteur de nombreux best-sellers tels que *Comprendre la nature humaine*, finaliste du Prix Pulitzer, ou *La Part d'ange en nous*.

Traduit de l'anglais (États-Unis) par Michel Le Séac'h.

ISBN : 979-10-92928-77-8



24 €

Prix TTC France

9 791092 928778

A L I S I O

design : Célia Cousty  
RAYON : ESSAIS



**Suivez toute l'actualité des éditions Alisio sur le blog :**

[www.alisio.fr](http://www.alisio.fr)

*Alisio est une marque des éditions Leduc.s*

Traduit de l'anglais par Michel Le Séac'h

Édition originale : *Everybody Lies : Big Data, New Data,  
and what the Internet reveals about who we really are,*  
Éditions Fletcher and Company, 2017.

Copyright © Seth Stephens-Davidowitz

Relecture-correction : Chantal Nicolas  
Maquette : Sébastienne Ocampo

© 2018 Alisio,  
une marque des éditions Leduc.s  
29 boulevard Raspail  
75007 Paris – France  
ISBN : 979-10-92928-77-8

Seth Stephens-Davidowitz  
Préface de Steven Pinker

TOUT LE MONDE  
**MENT...**  
(et vous aussi!)

**Internet et le Big Data :  
ce que nos recherches  
Google disent vraiment  
de nous**

A L I S I O



*À Maman et Papa*



# SOMMAIRE

<b>PRÉFACE</b>	<b>9</b>
<b>INTRODUCTION</b>	
<b>LES CONTOURS D'UNE RÉVOLUTION</b>	<b>13</b>
<b>PREMIÈRE PARTIE. BIG DATA OU SMALL DATA</b>	<b>35</b>
1. Votre petit doigt se trompe	37
<b>DEUXIÈME PARTIE. LES POUVOIRS DU BIG DATA</b>	<b>55</b>
2. Freud a-t-il vu juste ?	57
3. Les données réimaginées	67
4. Le sérum de vérité numérique	115
5. Zoom avant	175
6. Le monde entier est un labo	217
<b>TROISIÈME PARTIE. BIG DATA : À MANIER AVEC PRÉCAUTION</b>	<b>251</b>
7. Big Data, big n'importe quoi ? Ce que les données ne peuvent faire	253
8. Plus de données, plus de dommages ? Ce qu'on ne devrait pas faire	269
<b>CONCLUSION</b>	
<b>COMBIEN DE GENS LISENT LES LIVRES JUSQU'AU BOUT ?</b>	<b>283</b>
<b>REMERCIEMENTS</b>	<b>297</b>
<b>NOTES</b>	<b>301</b>
<b>INDEX</b>	<b>331</b>



# PRÉFACE

**D**epuis le temps où les philosophes fantasmaient sur le « cérébroscopie », appareil mythique censé afficher les pensées d'un individu sur un écran, les sciences sociales sont en quête d'outils qui montreraient comment fonctionne la nature humaine. Dans ma carrière de psychologue expérimental, j'en ai vu plus d'un devenir à la mode, puis passer de mode. Je les ai tous essayés – échelles de notation, temps de réaction, dilatation de la pupille, neuro-imagerie fonctionnelle, et même électrodes implantées sur des épileptiques qui, dans l'attente d'une attaque, tuaient le temps en expériences sur le langage.

Cependant, aucune de ces méthodes n'offre une vue imprenable sur l'esprit. Le problème tient à un compromis abrupt. Les pensées humaines sont des propositions complexes; quand on lit *Guerre et Paix*, on ne se borne pas à penser comme Woody Allen : « C'est une histoire de Russes. » Mais un scientifique a du mal à analyser ces propositions dans toute leur gloire multidimensionnelle et touffue. Certes, quand des gens ouvrent leur cœur, on appréhende la richesse de leur flux de conscience, mais un monologue ne constitue pas un jeu de données idéal pour tester des hypothèses. Si, en revanche, on s'attache à des mesures aisément quantifiables, comme le temps de réaction à un mot ou la réponse d'un épiderme à des images, on peut en tirer des statistiques, mais la texture complexe de la cognition se trouve réduite à un nombre unique. Les méthodes de

neuro-imagerie, même les plus élaborées, sont capables de dire où une pensée se situe dans un espace en trois dimensions, mais pas en quoi elle consiste.

Comme si ce tiraillement entre souplesse et richesse ne suffisait pas, les chercheurs en nature humaine se heurtent à la loi des petits nombres – le nom donné par Amos Tversky et Daniel Kahneman à l'idée fautive selon laquelle les traits d'une population se retrouveront dans n'importe quel échantillon, si petit fût-il. Même les savants les plus férus de chiffres ont des intuitions lamentablement fausses sur le nombre réel de sujets à inclure dans une étude pour échapper aux bizarreries du hasard et la généraliser à la totalité des Américains, sans parler de l'*Homo sapiens*. À plus forte raison quand on a composé l'échantillon avec les moyens du bord, par exemple en payant une bière à des étudiants de deuxième année.

Ce livre est consacré à une manière totalement nouvelle d'étudier les esprits. Le *Big Data* issu des recherches sur l'Internet et autres réponses en ligne n'est pas un cérébroscopie, mais Seth Stephens-Davidowitz montre qu'il offre un aperçu sans précédent sur le psychisme des gens. Dans l'intimité de leur clavier, ils livrent les confessions les plus étranges, quelquefois (comme sur les sites de rencontres ou dans les recherches de conseils spécialisés) parce qu'elles ont des conséquences dans la vie réelle, d'autres fois précisément parce qu'elles n'en ont pas : on peut avouer un désir ou une crainte quelconque sans susciter la consternation, ou pire, d'une personne en chair et en os. Dans un cas comme dans l'autre, les gens ne se contentent pas d'appuyer sur un bouton ou d'abaisser une manette : ils frappent une séquence de caractères, une parmi les millions de milliards de combinaisons possibles, pour exprimer leurs pensées dans toute leur ampleur exponentielle et combinatoire. Mieux encore, ils déposent ces traces numériques sous une forme qu'il est aisé d'agrèger et d'analyser. Leurs origines sont extrêmement variées. Ils peuvent participer

à des expériences non dérangeantes dans lesquelles on fait varier les *stimuli* et l'on note les réactions en temps réel. Et ils ne demandent pas mieux que de fournir ces données en quantités astronomiques.

*Tout le monde ment* est davantage qu'une démonstration de faisabilité. À de nombreuses reprises, les découvertes de Stephens-Davidowitz ont bouleversé mes idées reçues sur mon pays et mon espèce. D'où sont venus les renforts inattendus de Donald Trump ? Quand Ann Landers, en 1976, ayant interrogé ses lecteurs, a eu la surprise de découvrir qu'ils regrettaient en majorité d'avoir des enfants, a-t-elle été induite en erreur par un échantillon autosélectionné non représentatif ? Faut-il reprocher à l'Internet la crise de la fin des années 2010 si souvent appelée « bulle de filtres » ? Qu'est-ce qui provoque les crimes de haine ? Recherche-t-on des blagues pour se remonter le moral ? Et même si j'aime à penser que rien ne peut me choquer, je l'ai été plus d'une fois par ce que l'Internet révèle de la sexualité humaine – en particulier quand j'ai découvert que, chaque mois, un certain nombre de femmes font une recherche sur « baiser avec des animaux en peluche ». Jamais une expérience à base de temps de réaction, de dilatation des pupilles ou de neuro-imagerie fonctionnelle n'aurait pu révéler cela.

Tout le monde adorera *Tout le monde ment*. Avec sa curiosité toujours en éveil et son humour bon enfant, Stephens-Davidowitz montre une nouvelle voie aux sciences sociales du XXI<sup>e</sup> siècle. Puisqu'une fenêtre aussi fascinante s'est ouverte sur les obsessions humaines, qui a encore besoin d'un cérébroscopie ?

Steven Pinker, 2017



# INTRODUCTION

## Les contours d'une révolution

Il va perdre, c'est sûr, disaient-ils.

Les experts en sondages étaient formels : Donald Trump n'avait pas une chance aux élections primaires républicaines de 2016. Il avait insulté différents groupes minoritaires. Rares sont les Américains qui approuvent ce genre d'outrage, affirmaient les sondages et ceux qui les interprétaient.

La plupart des experts sondeurs pensaient à l'époque que Trump serait battu lors de l'élection présidentielle. Trop d'électeurs potentiels se disaient rebutés par ses manières et ses opinions.

Pourtant, l'éventualité d'une victoire de Trump aux primaires et à l'élection générale se lisait dans certains indices – sur l'Internet.

Je suis expert en données internet. Tous les jours, j'observe les pistes numériques laissées par les gens qui parcourent le Web. D'après les touches que nous frappons ou les boutons sur lesquels nous cliquons, j'essaie de comprendre ce que nous faisons réellement et qui nous sommes vraiment. Permettez-moi de raconter comment je me suis engagé dans cette voie peu commune.

L'histoire commence – et cela paraît déjà très lointain – avec l'élection présidentielle américaine de 2008 et une question depuis longtemps débattue par les sciences sociales : à quel point le préjugé racial est-il important en Amérique ?

Barack Obama a été le premier Afro-Américain désigné comme candidat à la présidentielle par l'un des grands partis. Il a été élu, assez aisément. Et les sondages semblaient montrer que la question raciale n'avait pas influencé le choix des électeurs américains. Gallup, par exemple, a conduit de nombreuses enquêtes d'opinion avant et après la première élection d'Obama. Il a conclu que pour les électeurs américains, dans l'ensemble, la couleur de peau de Barack Obama n'avait pas d'importance<sup>1</sup>. Peu après l'élection, deux professeurs réputés de l'université de Californie à Berkeley ont épluché d'autres données de sondages à l'aide de techniques de *data-mining* plus élaborées<sup>2</sup>. Ils sont parvenus à une conclusion similaire.

Pendant la présidence Obama, la cause a donc été considérée comme acquise dans une bonne partie des médias et des milieux universitaires : d'après les sources utilisées depuis plus de 80 ans par les médias et les sciences sociales pour comprendre le monde, le fait qu'Obama soit noir ne dérangeait pas l'écrasante majorité des Américains quand ils se demandaient s'ils devaient en faire leur président.

Longtemps souillé par l'esclavage et les lois Jim Crow, le pays paraissait avoir enfin cessé de juger les gens d'après la couleur de leur peau. Cela semblait indiquer que le racisme était à bout de souffle en Amérique. En fait, certains gourous déclaraient même que nous vivions dans une société post-raciale<sup>3</sup>.

En 2012, étudiant en économie égaré dans l'existence et consumé par ma discipline, je me sentais confiant, faraud même : je comprenais plutôt bien comment le monde fonctionnait, ce que les gens pensaient et ce dont ils se souciaient au XXI<sup>e</sup> siècle. Et en ce qui concerne la question du préjugé, je m'autorisais à croire, en fonction de tout ce que j'avais lu en

psychologie et en science politique, que le racisme explicite était limité à un faible pourcentage des Américains – républicains conservateurs en majorité et habitant pour la plupart le Sud profond.

Puis j'ai découvert Google Trends.

Google Trends, outil lancé sans tambour ni trompette en 2009, indique avec quelle fréquence un mot ou une phrase quelconque a fait l'objet d'une recherche en différents endroits à différents moments. Il était présenté comme un outil ludique – qui alimenterait par exemple des débats entre amis sur la célébrité la plus populaire ou sur la toute dernière mode. Dans ses premières versions, un avertissement signalait sur le ton de la blague qu'il valait mieux « vous abstenir d'écrire votre thèse de doctorat » à partir de ses données. Ce qui m'a aussitôt décidé à les utiliser pour rédiger ma propre thèse\*.

À l'époque, les données de recherche de Google n'étaient pas considérées comme une source d'informations convenable pour des travaux académiques « sérieux ». Au contraire des sondages, les recherches de Google n'étaient pas créées de manière à faciliter la compréhension du psychisme humain. Google a été inventé pour que les gens puissent connaître le monde, pas pour que les chercheurs puissent connaître les gens. Mais il s'avère que les traces que nous laissons en recherchant des connaissances sur l'Internet sont extrêmement révélatrices.

Autrement dit, les recherches d'informations sont en soi des informations. Le lieu et le moment où leurs auteurs

---

\* Mes données proviennent en grande partie de Google Trends. Cependant, celui-ci ne permet que de comparer la fréquence relative de différentes recherches, sans indiquer le nombre absolu des recherches. Je le complète en général par Google AdWords, qui rapporte exactement la fréquence de toutes les recherches. Dans la plupart des cas, j'ai aussi pu accentuer le tableau à l'aide de mon propre algorithme de calcul de tendances, décrit dans ma thèse « *Essays Using Google Data* », et dans mon article publié par le *Journal of Public Economics*, « *The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data* ». Ma thèse, un lien vers l'article et une explication complète des données et du code utilisé dans toute la recherche originale présentée dans ce livre sont disponibles sur mon site web, [www.sethsd.com](http://www.sethsd.com).

recherchent des faits, des citations, des blagues, des lieux, des personnes, des choses ou de l'aide peuvent en dire bien plus qu'on ne le croirait sur ce qu'ils pensent, ce qu'ils désirent ou ce qu'ils craignent en réalité. C'est d'autant plus vrai que certaines personnes, parfois, adressent à Google des confidences plus que des questions : « je hais mon patron », « j'ai trop bu », « papa m'a tapé ».

Dactylographier un mot ou une phrase dans un petit rectangle blanc : ce geste de tous les jours laisse une mince trace de vérité qui, répétée des millions de fois, finit par révéler des réalités profondes. Le premier mot que j'ai saisi dans Google Trends était « God » (Dieu). J'ai appris que les États américains d'où provenait le plus grand nombre de recherches mentionnant « Dieu » étaient l'Alabama, le Mississippi et l'Arkansas – la « Bible Belt ». Et ces recherches intervenaient le plus souvent le dimanche. Cela n'avait rien d'étonnant ; plus étonnante en revanche était la régularité révélée par les données. J'ai essayé « Knicks »\*, qui se trouve être le mot le plus « googlé » à New York City. Même résultat évident. Puis j'ai saisi mon propre nom. « Désolé », m'a répondu Google. « Données insuffisantes. » Google Trends, ai-je ainsi appris, ne fournit de données que lorsque de nombreuses personnes effectuent la même recherche.

Mais l'intérêt des recherches Google n'est pas qu'elles soient capables de nous dire que Dieu est populaire dans le sud des États-Unis, les Knicks à New York et moi-même nulle part. N'importe quel sondage vous en dirait autant. Leur intérêt est que les gens disent au moteur de recherche géant des choses qu'ils ne diraient probablement à personne d'autre.

Prenez par exemple le sexe (un sujet qui sera davantage approfondi plus loin). Impossible de compter sur les sondages pour nous apprendre la vérité sur notre vie sexuelle. J'ai analysé les données du General Social Survey<sup>4</sup>, considéré

---

\* Knicks : équipe de basket-ball new-yorkaise (NdT).

comme l'une des sources d'informations les plus influentes et les plus dignes de foi sur les comportements des Américains. Selon cette enquête, à la rubrique « sexe hétérosexuel », les femmes disent faire l'amour en moyenne cinquante-cinq fois par an, en utilisant un préservatif dans 16 % des cas. Soit à peu près 1,1 milliard de préservatifs utilisés par an. Mais les hommes hétérosexuels disent utiliser 1,6 milliard de préservatifs par an. Par définition, les nombres devraient être identiques. Qui donc dit la vérité, les hommes ou les femmes ?

Ni les uns ni les autres, en fait. Selon Nielsen, société internationale d'informations et de mesures d'audience qui suit le comportement des consommateurs, il se vend chaque année moins de 600 millions de préservatifs<sup>5</sup>. Donc, tout le monde ment, la seule différence étant dans les proportions.

Le mensonge est partout. Les hommes qui n'ont jamais été mariés disent utiliser en moyenne vingt-neuf préservatifs par an. Ce qui, acheteurs mariés et célibataires confondus, ferait au total plus que le nombre total de préservatifs vendus aux États-Unis. Il est probable que les gens mariés exagèrent eux aussi leur vie sexuelle. En moyenne, les hommes mariés de moins de 65 ans déclarent aux sondeurs faire l'amour une fois par semaine. Un pour cent seulement dit ne l'avoir pas fait du tout dans l'année écoulée. Les femmes mariées disent pratiquer le sexe un peu moins souvent, juste un peu.

Les recherches Google brossaient un tableau bien moins dynamique – et selon moi beaucoup plus exact – du sexe dans le mariage. Sur Google, la récrimination numéro un à l'égard du mariage est l'absence de sexualité. Les recherches sur « sexless marriage » (mariage sans sexe) sont trois fois et demie plus fréquentes que celles sur « unhappy marriage » (mariage pas heureux) et huit fois plus que celles sur « loveless marriage » (mariage sans amour). Même les couples non mariés se plaignent assez fréquemment de ne pas faire l'amour. « Sexless relationship » (relation sans sexe) n'est devancé que

par « abusive relationship » (relation violente). (Ces données, il faut le souligner, sont toutes présentées de manière anonyme. Google ne publie évidemment pas la moindre donnée sur les recherches effectuées par une personne désignée.)

Et les recherches Google brossaient un tableau de l'Amérique extrêmement différent de l'utopie post-raciale esquissée par les sondages. Je me souviens de la première fois où j'ai tapé « nigger » (nègre) dans Google Trends. Naïf que j'étais ! Vu l'obscénité du mot, je m'attendais à un volume de recherches faible. J'avais tort, ô combien. Aux États-Unis, le mot « nigger » – ou son pluriel « niggers » – apparaissait dans les recherches à peu près aussi souvent que les mots « migraine(s) », « économiste » et « Lakers »\*. Les résultats auraient-ils pu être faussés par les morceaux de rap ? Non. Le mot utilisé par les rappeurs est presque toujours « nigga(s) ». Alors, quelle est la motivation des Américains qui font une recherche sur « nigger » ? Souvent, ils sont en quête de blagues moquant les Afro-Américains. En fait, 20 % des recherches comprenant le mot « nigger » comprenaient aussi le mot « jokes » (blagues). « Stupid niggers » (nègres stupides) et « I hate niggers » (je déteste les nègres) étaient fréquents aussi.

Ces recherches se chiffraient par millions chaque année. Dans l'intimité de leur demeure, un grand nombre d'Américains lançaient des recherches d'un racisme choquant. Plus j'enquêtais, plus l'information devenait dérangeante.

Le soir de la première élection d'Obama, alors que la plupart des commentaires chantaient les louanges du nouveau président et soulignaient le caractère historique de cette élection, à peu près 1 % des recherches Google contenant le mot « Obama » incluait aussi « kkk » ou « nigger(s) ». Cela ne paraît peut-être pas beaucoup, mais songez aux milliers de raisons non racistes de googler ce jeune *outsider* doté d'une famille charmante qui

---

\* Lakers : équipe de basket-ball de Los Angeles (NdT).

allait occuper le poste le plus puissant du monde. Le soir de l'élection, les recherches sur Stormfront et les visites sur ce site nationaliste blanc étonnamment populaire aux États-Unis ont été dix fois plus nombreuses qu'à l'ordinaire<sup>6</sup>. Dans certains États, il y a eu plus de recherches sur « nigger president » que sur « first black president » (premier président noir)<sup>7</sup>.

Les recherches effectuées par les gens révélaient manifestement une noirceur et une haine que les sources traditionnelles ne distinguaient pas.

Ces recherches sont difficiles à concilier avec l'idée d'une société où le racisme serait un facteur peu important. En 2012, je connaissais Donald J. Trump principalement comme un homme d'affaires et l'animateur d'une émission de télé-réalité. Pas plus que quiconque je n'imaginai qu'il serait quatre ans plus tard un candidat sérieux à l'élection présidentielle. Mais il n'est pas difficile de rapprocher ces vilaines recherches et le succès d'un candidat qui – par ses attaques contre les migrants, ses colères et ses rancœurs – s'est souvent adressé aux pires penchants des gens.

Les recherches Google disaient aussi que nos idées sur la géographie du racisme étaient en grande partie fausses. Selon les sondages et les idées reçues, le racisme moderne se situe principalement dans le Sud et surtout chez les électeurs du Parti républicain. Mais parmi les endroits où les taux de recherches racistes étaient les plus élevés figuraient aussi le nord de l'État de New York, l'ouest de la Pennsylvanie, l'est de l'Ohio, la partie industrielle du Michigan et l'Illinois rural, aux côtés de la Virginie occidentale, du sud de la Louisiane et du Mississippi. La vraie fracture, d'après les données des recherches Google, ne passait pas entre le Sud et le Nord, mais entre l'Est et l'Ouest. On ne rencontre pas beaucoup ce genre de choses à l'ouest du Mississippi. Et le racisme ne se limitait pas aux républicains. En fait, les recherches racistes n'étaient

pas plus fréquentes là où le pourcentage de républicains est élevé que là où les démocrates dominent. Autrement dit, les recherches Google aidaient à tracer une nouvelle carte du racisme aux États-Unis – et cette carte était très différente de ce que vous imaginiez peut-être. Les républicains du Sud sont peut-être plus enclins à admettre leur racisme. Mais beaucoup de démocrates du Nord ont des attitudes similaires.

Quatre ans plus tard, cette carte s'avérerait très utile pour expliquer le succès politique de Trump.

En 2012, la carte du racisme que j'avais établie à partir des recherches Google m'a servi à réévaluer le rôle exact joué par la race d'Obama. Les données étaient claires. Dans les parties du pays où les recherches racistes étaient très nombreuses, les résultats d'Obama ont été substantiellement moins bons que ceux du candidat démocrate blanc John Kerry lors de la précédente élection présidentielle, quatre ans auparavant. Aucun autre facteur relatif à ces régions ne pouvait expliquer la relation, que ce soit le niveau d'éducation, l'âge, la fréquentation religieuse ou la détention d'une arme. Les recherches racistes n'annonçaient de mauvais résultat pour aucun candidat démocrate. Sauf pour Obama.

Et les résultats impliquaient un effet large. À lui seul, le racisme explicite a fait perdre à Obama environ 4 points au niveau national. C'était beaucoup plus que ce que l'on aurait pu attendre au vu des sondages. Barack Obama a été élu et réélu président, bien entendu, aidé en cela par certaines conditions très favorables au Parti démocrate, mais il a dû surmonter plus d'obstacles que ne le pensaient ceux qui utilisaient des sources de données traditionnelles – c'est-à-dire à peu près tout le monde. Les racistes étaient assez nombreux pour contribuer à une victoire dans une primaire ou infléchir le sort d'une élection générale en une année moins favorable aux démocrates.

Mon étude a initialement été refusée par cinq journaux académiques<sup>8</sup>. Beaucoup des pairs réviseurs, si vous voulez

bien excuser une petite récrimination, estimaient impossible de croire que tant d'Américains ont affiché un racisme aussi pernicieux. Cela ne collait simplement pas avec ce qu'ils disaient. Et puis, les recherches de Google apparaissaient comme un jeu de données plutôt baroque.

Maintenant que l'on a vu le président Donald J. Trump entrer à la Maison-Blanche, mes découvertes paraissent davantage plausibles.

Plus je creuse la question, plus je constate que Google détient des masses d'informations ignorées des sondages et susceptibles d'aider à comprendre une élection – parmi beaucoup, beaucoup d'autres sujets.

On y trouve des informations sur ceux qui iront effectivement voter. Immédiatement avant une élection, plus de la moitié des abstentionnistes disent aux sondeurs qu'ils ont l'intention d'aller voter. Ils faussent ainsi les estimations de participation, tandis que les recherches Google sur « où voter » ou « comment voter » dans les semaines précédant une élection permettent de prédire exactement dans quelles parties du pays les électeurs seront nombreux à se rendre aux urnes.

Peut-être pourrait-on même y trouver des informations sur le choix qu'ils exprimeront. Peut-on vraiment prédire, juste d'après leurs recherches, pour quel candidat les gens vont voter? Clairement, étudier le nombre des recherches ne suffit pas. Beaucoup de gens cherchent à se renseigner sur un candidat parce qu'il leur plaît. Un nombre du même ordre cherchent à se renseigner sur un candidat parce qu'ils le détestent. Cela dit, Stuart Gabriel, professeur de finance à l'université de Californie à Los Angeles, et moi-même avons mis le doigt sur un indice surprenant des intentions de vote. Un pourcentage élevé des requêtes relatives à une élection contient le nom des deux candidats. Au cours de l'élection de 2016 entre Trump et Clinton, certaines personnes ont effectué

une recherche sur « Trump Clinton polls » (sondages Trump Clinton). D'autres sur « Trump Clinton debate » (débat Trump Clinton). En fait, 12 % des requêtes contenant « Trump » contenaient aussi « Clinton ». Plus d'un quart des requêtes contenant « Clinton » incluait aussi « Trump ».

Nous avons découvert que ces recherches apparemment neutres pouvaient en fait révéler quelque chose sur le candidat soutenu par l'électeur.

Comment cela ? D'après l'ordre dans lequel les candidats sont mentionnés. D'après nos travaux, dans une recherche qui comporte les noms des deux candidats, un électeur a nettement plus de chance de placer en premier celui qu'il préfère.

Au cours des trois élections précédentes, le candidat qui a obtenu le plus de voix est celui qui était apparu en premier dans le plus grand nombre de recherches. Plus intéressant encore, l'ordre dans lequel les candidats étaient recherchés était prédictif du résultat au niveau des États.

L'ordre des candidats dans les recherches semble aussi contenir des informations qui peuvent échapper aux sondages. Lors de l'élection de 2012, le journaliste et statisticien virtuose Nate Silver a prédit avec exactitude le résultat de l'affrontement entre Obama et le républicain Mitt Romney dans la totalité des cinquante États américains. Cependant, nous avons découvert que dans les États où Romney précédait le plus souvent Obama dans les recherches, le candidat républicain a en réalité obtenu un résultat meilleur que ne l'avait prévu Silver. Dans les États où Obama précédait le plus souvent Romney, Obama a fait mieux que prévu par Silver.

Cet indicateur pourrait contenir des informations que les sondeurs ne voient pas, soit parce que les électeurs se mentent à eux-mêmes, soit parce qu'ils n'aiment pas révéler leurs vraies préférences aux enquêteurs. S'ils se déclaraient indécis en 2012, mais recherchaient avec insistance « Romney Obama polls », « Romney Obama debate » et « Romney Obama

election », peut-être avaient-ils eu dès le départ l'intention de voter Romney.

Alors, Google a-t-il annoncé Trump ? Eh bien, nous avons encore beaucoup de pain sur la planche – et il faudra que beaucoup d'autres chercheurs y travaillent – avant de savoir utiliser au mieux les données de Google pour prévoir les résultats des élections. Cette science est nouvelle et nous ne disposons de données que pour quelques élections. Je ne prétends certainement pas que nous en sommes – ni même que nous parviendrons un jour – au point où nous pourrions renoncer complètement aux sondages d'opinion pour prévoir le résultat des élections.

Mais il ne fait aucun doute que, à de nombreuses reprises, l'Internet avait présagé pour Trump des résultats supérieurs à ce que les sondages annonçaient.

Lors de l'élection générale, certains indices montraient que l'électorat pouvait lui être favorable. Les Noirs-Américains déclaraient aux sondeurs qu'ils iraient voter en grand nombre pour faire barrage à Trump. Mais les recherches Google sur les modalités du vote dans les quartiers à forte population noire étaient très en retrait. Le jour de l'élection, Hillary Clinton serait handicapée par une faible participation chez les Noirs.

Certains signes montraient même que des électeurs censément indécis penchaient vers Trump. Stuart Gabriel et moi-même avons constaté que les recherches sur « Trump Clinton » l'emportaient sur « Clinton Trump » dans les États clés du Midwest où une victoire de Clinton était attendue. Trump doit son élection à des résultats bien meilleurs que les sondages ne l'avaient annoncé dans ces États.

Mais à mon avis, le principal indice d'une victoire possible de Trump – et d'abord dans les primaires – était ce racisme secret mis au jour par mon étude sur Obama. Les recherches Google révélaient chez beaucoup d'Américains une noirceur et une haine ignorées des observateurs pendant bien des

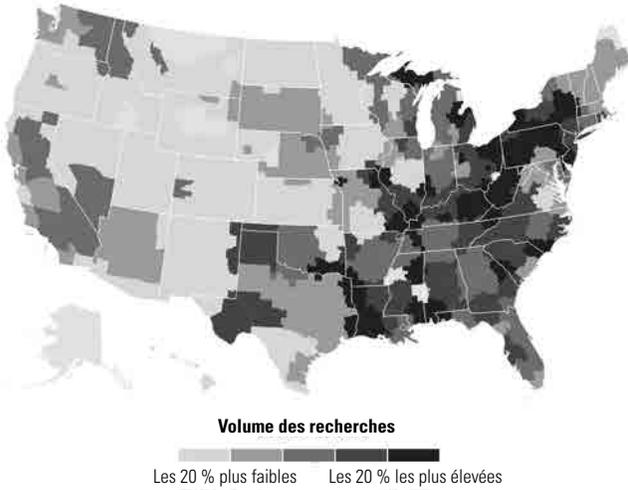
années. Les données montraient que nous vivions dans une société très différente de celle qu'universitaires et journalistes imaginaient sur la foi des sondages. Elles dénotaient un état d'esprit hostile, effrayant et répandu, qui attendait qu'un candidat lui donne l'occasion de s'exprimer.

Les gens mentent souvent – à eux-mêmes et aux autres. En 2008, les Américains disaient aux sondeurs que la race n'avait plus d'importance pour eux. Huit ans plus tard, ils ont élu à la présidence un homme, Donald J. Trump, qui retweetait l'affirmation mensongère que la majorité des meurtres d'Américains blancs étaient commis par des Noirs, qui défendait ses partisans accusés d'avoir brutalisé un manifestant de Black Lives Matters au cours d'une de ses réunions et qui hésitait à refuser le soutien d'un ancien chef du Ku Klux Klan. Le même racisme occulte qui avait nui à Barack Obama a aidé Donald Trump.

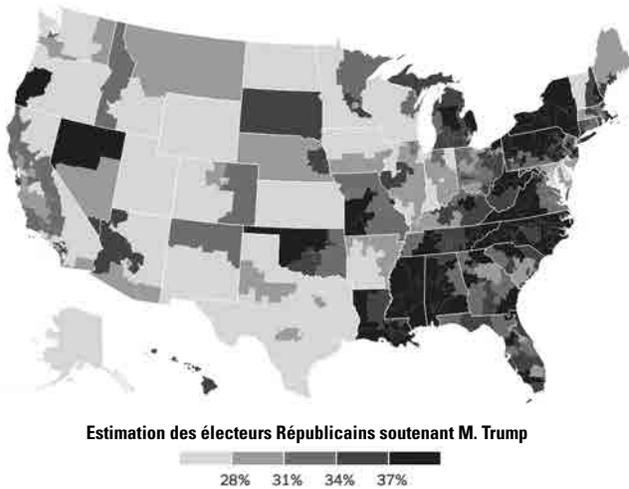
Au début des primaires, Nate Silver avait hautement proclamé que Trump n'avait pratiquement aucune chance de gagner. Puis, quand un vaste mouvement en sa faveur est apparu de plus en plus clair au cours des primaires, il a décidé d'examiner les données pour chercher à comprendre ce qui se passait. Comment Trump pouvait-il réussir aussi bien ?

Silver a remarqué que les zones où Trump obtenait les meilleurs résultats dessinaient une carte étrange. Trump réussissait bien dans des parties du Nord-Est et du Midwest industriel, ainsi que dans le Sud. À l'Ouest, beaucoup moins. Silver a recherché les variables susceptibles d'expliquer cette carte. Le chômage ? La religion ? La détention d'armes ? Le taux d'immigration ? L'opposition à Obama ?

### Taux de recherches racistes



### Partisans de Donald Trump lors de la primaire républicaine



Il a constaté que le facteur unique le mieux corrélé avec le vote en faveur de Trump dans les primaires républicaines était celui que j'avais découvert quatre ans plus tôt<sup>9</sup>. Les zones où Trump avait le plus de partisans étaient celles où les recherches sur le mot « nigger » étaient les plus nombreuses.

Nous espérons que cet extrait  
vous a plu !



**Tout le monde ment... (et vous aussi !)**  
Seth Stephens-Davidowitz



J'achète ce livre

Pour être tenu au courant de nos parutions, inscrivez-vous à la lettre des éditions Leduc.s et recevez des **bonus**, **invitations** et autres **surprises** !

Je m'inscris

Merci de votre confiance, à bientôt !

A L I S I O