

Seth Stephens-Davidowitz

TODO EL MUNDO MIENTE

Lo que internet y el *bigdata*
pueden decir de nosotros

Prólogo de Steven Pinker

Traducción de Martín Schifino

 PAIDÓS®

Diseño de portada: Ivan Castillo / Pulso.Inc
Ilustraciones de portada: © Shutterstock / durantelallera

Título original: *Everybody Lies*

© Seth Stephens-Davidowitz

La presente edición ha sido licenciada a Ediciones Culturales Paidós por los propietarios de los derechos mundiales en español, Capitán Swing (Madrid), por mediación de Oh!Books Agencia Literaria.

© Martin Schifino, por la traducción

La presente traducción ha sido licenciada a Ediciones Culturales Paidós por el propietario de los derechos, Capitán Swing (Madrid), por mediación de Oh!Books Agencia Literaria.

Derechos reservados

© 2019, Ediciones Culturales Paidós, S.A. de C.V.

Bajo el sello editorial PAIDÓS M.R.

Avenida Presidente Masarik núm. 111, Piso 2

Colonia Polanco V Sección

Delegación Miguel Hidalgo

C.P. 11560, Ciudad de México

www.planetadelibros.com.mx

www.paidos.com.mx

Primera edición en formato epub: mayo de 2019

ISBN: 978-607-747-694-8

Primera edición impresa en México: mayo de 2019

ISBN: 978-607-747-689-4

No se permite la reproducción total o parcial de este libro ni su incorporación a un sistema informático, ni su transmisión en cualquier forma o por cualquier medio, sea éste electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin el permiso previo y por escrito de los titulares del *copyright*.

La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (Arts. 229 y siguientes de la Ley Federal de Derechos de Autor y Arts. 424 y siguientes del Código Penal).

Si necesita fotocopiar o escanear algún fragmento de esta obra diríjase al CeMPro (Centro Mexicano de Protección y Fomento de los Derechos de Autor, <http://www.cempro.org.mx>).

Impreso en los talleres de Litográfica Ingramex, S.A. de C.V.

Centeno núm. 162, colonia Granjas Esmeralda, Ciudad de México

Impreso y hecho en México - *Printed and made in Mexico*

..... Índice

PRÓLOGO DE STEVEN PINKER	9
INTRODUCCIÓN: EL PERFIL DE UNA REVOLUCIÓN	15

PRIMERA PARTE **DATOS GRANDES, DATOS PEQUEÑOS**

1. MALAS CORAZONADAS	39
----------------------	----

SEGUNDA PARTE **LAS CAPACIDADES DE LOS MACRODATOS**

2. ¿TENÍA RAZÓN FREUD?	59
3. DATOS REINVENTADOS	69
Los cuerpos como datos	76
Las palabras como datos	87
Las imágenes como datos	108
4. SUERO DE LA VERDAD DIGITAL	115
La verdad sobre el sexo	122
La verdad sobre el odio y el prejuicio	137
La verdad sobre internet	148
La verdad sobre el abuso infantil y el aborto	153
La verdad sobre tus amigos de Facebook	158
La verdad sobre tus clientes	161
¿Podemos soportar la verdad?	165

5. MÁS DE CERCA	1 71
¿Qué ocurre realmente en nuestros condados, ciudades y pueblos?	176
¿En qué ocupamos los minutos y las horas?	192
Nuestros dobles	199
Historias de datos	206
6. TODO EL MUNDO ES UN LABORATORIO	2 09
El ABC de las pruebas A/B	211
Los experimentos crueles, pero iluminadores, de la naturaleza	2 21

TERCERA PARTE

MACRODATOS: PRECAUCIÓN

7. ¿MACRODATOS, MACROCHASCOS? LO QUE NO SE PUEDE HACER CON ELLOS	2 43
La maldición de la dimensionalidad	246
El énfasis excesivo en lo medible	2 51
8. ¿MÁS DATOS, MÁS PROBLEMAS? LO QUE NO DEBERÍAMOS HACER	2 57
El peligro del poder de las corporaciones	257
El peligro del poder de los gobiernos	264
CONCLUSIÓN: ¿CUÁNTAS PERSONAS ACABAN LOS LIBROS?	2 71
AGRADECIMIENTOS	2 85

PRÓLOGO

Desde que los filósofos especularon sobre la posibilidad de construir un «cerebroscopio», aparato mítico que proyectaría los pensamientos de una persona en una pantalla, los científicos sociales han buscado herramientas que permitieran sacar a la luz los mecanismos de la naturaleza humana. A lo largo de mi carrera como psicólogo experimental, se pusieron de moda y se olvidaron unas cuantas, y las he probado todas: escalas de calificación, tiempos de reacción, dilatación pupilar, neuroimagen funcional, incluso electrodos implantados en el cerebro de pacientes epilépticos que parecían muy bien dispuestos a participar en un experimento lingüístico mientras esperaban a tener convulsiones.

Sin embargo, ninguno de esos métodos proporciona una visión diáfana de la mente. Y además debemos elegir. Los pensamientos humanos son un problema complejo; a diferencia de Woody Allen tras leer *Guerra y paz* a toda velocidad, no podemos concluir solamente: «Iba sobre unos rusos». Pero es difícil analizar de manera científica el intrincado esplendor multidimensional de ese problema. Sin duda, cuando la gente se desahoga, aprehendemos la riqueza de sus monólogos interiores, pero el fluir de la conciencia no es un conjunto de datos ideal para someter a prueba una hipótesis. Al revés, si nos centramos en parámetros fáciles de cuantificar, como las reacciones personales a las palabras, o la respuesta de la piel a las imágenes, podemos compilar estadísticas, pero trituramos la compleja textura de la cognición para transformarla en un número único. Si bien las sofisticadas técnicas de neuroimagen pueden mostrarnos cómo se distribuye un pensamiento en un espacio tridimensional, no pueden decirnos en qué consiste ese pensamiento.

Como si no fuera problema suficiente tener que optar entre computabilidad y riqueza, los científicos de la naturaleza humana deben vérselas con la «ley de los pequeños números», el nombre que dieron Amos Tversky y Daniel Kahneman a la falacia de pensar que los rasgos de una población se ven reflejados en cualquier porción de ella, por pequeña que sea. Incluso los científicos más expertos en aritmética tienen intuiciones penosamente desacertadas sobre la cantidad de sujetos que se necesitan en un estudio para poder eliminar irregularidades aleatorias y hacer generalizaciones válidas sobre una población determinada, no hablemos ya del *Homo sapiens*. La cosa es aún más discutible cuando las muestras se obtienen por conveniencia, como cuando ofrecemos algo de dinero a nuestros estudiantes de primer curso.

Este libro versa sobre un modo completamente nuevo de estudiar la mente. Los *big data* o macrodatos de las búsquedas de internet y otras huellas en línea no constituyen un cerebroscopio, pero Seth Stephens-Davidowitz demuestra que ofrecen una visión sin precedentes de la psique humana. En la privacidad de sus teclados, la gente confiesa las cosas más peregrinas, a veces (como en los sitios de citas o de asesoramiento profesional) porque tienen consecuencias en la vida real, y otras precisamente porque *no* tienen consecuencias: se puede exponer un deseo o un temor sin riesgo de que una persona de carne y hueso se muestre consternada o cosas peores. En cualquier caso, la gente no solo aprieta un botón o gira un interruptor, sino que consigna trillones de secuencias de caracteres que detallan sus pensamientos en toda su vastedad explosiva y combinatoria. Mejor aún, deja estelas digitales en un formato que puede compilarse y analizarse fácilmente. Son personas de todas las profesiones. Pueden formar parte de experimentos no invasivos que varían los estímulos y tabulan las respuestas en tiempo real. Y proporcionan de buen grado esos datos en cantidades ingentes.

Todo el mundo miente es más que la demostración de un concepto. Una y otra vez, los descubrimientos de Stephens-Davidowitz subvirtieron las ideas preconcebidas que tenía sobre mi país y mi especie. ¿De dónde salió el inesperado apoyo a Donald Trump? En 1976, cuando Ann Landers preguntó a sus lectores si

se arrepentían de haber tenido hijos y descubrió con sorpresa que la mayoría lo hacía, ¿la engañaba una muestra poco representativa y sesgada de quienes habían contestado a la encuesta? ¿Debe culparse a internet de la crisis de fines de la década de 2010 que, de manera redundante, se ha dado en llamar «filtro burbuja»? ¿Cómo se desencadenan los delitos de odio? ¿Busca la gente chistes para alegrarse? También, aunque me gusta creer que nada me impacta, me impactó sobremanera lo que revela internet sobre la sexualidad humana, incluido el descubrimiento de que todos los meses varias mujeres hacen una búsqueda con las palabras «cogerse animales embalsamados». Ningún experimento basado en el tiempo de reacción, la dilatación pupilar o las neuroimágenes funcionales habría descubierto ese dato.

Todo el mundo disfrutará de *Todo el mundo miente*. Con una curiosidad incesante y un ingenio entrañable, Stephens-Davidowitz allana un nuevo camino para las ciencias sociales en el siglo XXI. Con esta fascinante ventana siempre abierta a las obsesiones humanas, ¿quién necesita un cerebroscopio?

—STEVEN PINKER, 2017

El perfil de una revolución

Claro que iba a perder, dijeron.

En las primarias republicanas de 2016, los analistas de los sondeos concluyeron que Donald Trump no tenía ninguna posibilidad de ganar. Al fin y al cabo, había denigrado a numerosas minorías. Los sondeos y sus intérpretes nos dijeron que pocos estadounidenses veían con buenos ojos esos desplantes.

Por entonces, la mayoría de los analistas de sondeos pensaban que Trump perdería en unas elecciones generales. Sus modales y opiniones —dijeron— disgustaban a demasiados votantes.

Pero en realidad había indicios de que Trump podía ganar las elecciones primarias y las generales; estaban en internet.

Soy experto en datos de internet. Todos los días rastreo las huellas digitales que se dejan al transitar por la red. A partir de los botones o teclas que apretamos, intento entender qué queremos en realidad, qué haremos en realidad y quiénes somos en realidad. Permítaseme explicar cómo me inicié en esta atípica profesión.

La historia comienza —y parece que fue hace siglos— con las elecciones presidenciales de 2008 y con una cuestión muy discutida en las ciencias sociales: ¿hasta qué punto son importantes los prejuicios raciales en los Estados Unidos?

Barack Obama fue el primer candidato afroamericano de un partido importante. Ganó las elecciones con bastante facilidad. Y ya los sondeos sugerían que la cuestión racial no era un factor determinante en la manera de votar de los estadounidenses. La organización Gallup, por ejemplo, hizo varios sondeos antes y después de que Obama fuera elegido por primera vez. ¿Sus conclusiones?

En general, a los votantes estadounidenses les daba igual que Obama fuese negro.¹ Poco después de la elección, dos profesores muy conocidos de la Universidad de California, Berkeley, examinaron datos adicionales obtenidos en encuestas mediante técnicas de extracción de datos más sofisticadas.² Llegaron a una conclusión similar.

Y así, durante la presidencia de Obama, aquella conclusión se convirtió en una opinión generalizada para muchos medios de comunicación y en amplios sectores universitarios. Las fuentes que los medios y los científicos sociales habían utilizado durante más de ochenta años para comprender el mundo nos decían que a la mayoría de los estadounidenses le daba igual que Obama fuera negro a la hora de decidir si debía ser su presidente.

Los Estados Unidos, tanto tiempo manchados por la esclavitud y las leyes de segregación racial, por fin parecían haber dejado de juzgar a la gente por el color de su piel. Por lo visto, en los Estados Unidos el racismo tenía los días contados. De hecho, algunos comentaristas incluso declararon que vivíamos en una sociedad posracial.³

En 2013 yo era un estudiante de posgrado en economía, perdido en la vida y hastiado de mi campo; estaba seguro, incluso hasta la petulancia, de entender bastante bien el funcionamiento del mundo, los deseos y las preferencias de la gente en el siglo XXI. Ante la cuestión de los prejuicios, me permitía creer, sobre la base de mis lecturas en materia de psicología y ciencias políticas, que el racismo explícito se limitaba a un porcentaje restringido de estadounidenses, en su mayoría republicanos conservadores que vivían en el sur profundo.

¹ Katie Fretland, «Gallup: Race Not Important to Voters», *The Swamp*, *Chicago Tribune*, junio de 2008.

² Alexandre Mas y Enrico Moretti, «Racial Bias in the 2008 Presidential Election», *American Economic Review* 99, no. 2 (2009).

³ En la emisión del 12 de noviembre de 2009, Lou Dobbs dijo en su programa que vivíamos en una «sociedad pospartisana, posracial». En la emisión del 27 de enero de 2010, Chris Matthews dijo en su programa que el presidente Obama era «al parecer posracial». Para más ejemplos, véase Michael C. Dawson y Lawrence D. Bobo, «One Year Later and the Myth of a Post-Racial Society», *Du Bois Review: Social Science Research on Race* 6, no. 2 (2009).

Y entonces descubrí Google Trends.

Google Trends, una herramienta que se lanzó con poco ruido en 2009, informa a los usuarios de la frecuencia con que cualquier palabra o frase se ha buscado en distintos momentos en distintos sitios. Se promocionó como una herramienta divertida, que quizá permitiera a los grupos de amigos debatir sobre qué persona famosa era más popular o qué se había puesto súbitamente de moda. Las primeras versiones incluían el travieso aviso de que «nadie desearía escribir su tesis de doctorado con esos datos», lo que de inmediato me motivó a usarlos para escribir mi tesis de doctorado.⁴

Por entonces, los datos de búsqueda en Google no parecían ser una fuente adecuada de información para realizar investigaciones académicas «serias». A diferencia de los sondeos, los datos de búsqueda en Google no se crearon con el fin de entender la psique humana. Google se inventó para que la gente pudiera averiguar cosas sobre el mundo, no para que los investigadores pudieran averiguar cosas sobre la gente. Pero lo cierto es que las huellas que dejamos al buscar conocimientos en internet son sumamente reveladoras.

Dicho de otro modo, la búsqueda de información por personas es, a su vez, información. El cuándo y el dónde se buscan hechos, citas, chistes, lugares, nombres, cosas o ayuda, en efecto, pueden decirnos mucho más sobre los deseos, pensamientos y temores reales de lo que se creía. Y eso es especialmente cierto porque la gente a veces no solo pide datos a Google, sino que también se confiesa al buscador: «Odio a mi jefe», «Estoy borracho», «Mi padre me ha pegado».

⁴ Google Trends me ha proporcionado muchos de los datos que utilizo aquí. Sin embargo, como solo permite comparar la frecuencia relativa de distintas búsquedas, pero no informa sobre el número absoluto de una búsqueda particular, por lo general lo he complementado con GoogleAdWords, que informa de la frecuencia exacta con que se hace una búsqueda. En casi todos los casos también he podido enfocar mejor la imagen con ayuda de mi propio algoritmo basado en Trends, que describo en mi tesis «Essays Using Google Data», así como en mi artículo «The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data», publicado en el *Journal of Public Economics*. La disertación, un vínculo al artículo y una explicación completa de los datos y el código utilizados en las investigaciones originales llevadas a cabo para este libro pueden consultarse en mi sitio web, sethsd.com.

La acción cotidiana de escribir una palabra o frase en una caja compacta y rectangular blanca deja un pequeño rastro de verdad que, multiplicado por millones, acaba revelando realidades profundas. La primera palabra que ingresé en el Google Trends fue «Dios». Descubrí que los estados en los que se hacían más búsquedas con la mención de «Dios» eran Alabama, Misisipi y Arkansas, los estados mayormente evangélicos. Además, esas búsquedas son más frecuentes los domingos. Nada de ello era sorprendente, pero el hecho de que los datos de búsqueda pudieran revelar un patrón tan claro me intrigó. Probé con «Knicks», el equipo de basketbol neoyorquino, que casualmente se busca sobre todo en Nueva York. Otra obviedad. Después escribí mi nombre. «Tu búsqueda no tiene suficientes datos para mostrar resultados», me informó Google Trends. Google Trends, descubrí, arrojaba datos solo cuando mucha gente hacía una misma búsqueda.

Pero las búsquedas en Google no solo tienen la capacidad de decirnos que Dios es popular en el sur de los Estados Unidos, los Knicks en Nueva York y yo en ningún sitio. Cualquier sondeo diría lo mismo. Los datos de Google son poderosos porque la gente le cuenta al gigantesco motor de búsqueda cosas que no le contarían a nadie más.

Pensemos, por ejemplo, en el sexo (un tema que investigaré con lujo de detalles más adelante en este libro). Los sondeos no son fiables a la hora de decirnos la verdad sobre nuestra vida sexual. Analicé los datos de la General Social Survey, que se considera una de las fuentes más influyentes y fidedignas de información sobre las conductas de los estadounidenses.⁵ De acuerdo con los datos de esa encuesta, cuando se trata de relaciones heterosexuales, las mujeres dicen en promedio que tienen relaciones sexuales 55 veces por año y utilizan un preservativo el 16% de las veces. Eso da unos 1 100 millones de preservativos usados por año. Pero los hombres heterosexuales declaran usar 1 600 millones de preservativos al año. Ambas cifras, por definición, debe-

⁵ Los detalles de todos estos cálculos pueden encontrarse en mi sitio web, sethsd.com, en los archivos csv etiquetados «Sex Data». Los datos de General Social Survey pueden encontrarse en <http://gss.norc.org/>

rían ser iguales. ¿Quién dice la verdad, los hombres o las mujeres?

En realidad, nadie. De acuerdo con Nielsen, la compañía global de gestión de información acerca de lo que ve y compra el consumidor, por año se venden menos de 600 millones de preservativos.⁶ Así que todo el mundo miente: la cuestión es cuánto.

Las mentiras están muy extendidas. Los hombres que nunca han estado casados afirman usar un promedio de 29 preservativos por año. Eso ascendería a una cifra mayor que el total de los preservativos vendidos en los Estados Unidos a los casados y solteros juntos. También la gente casada, con toda probabilidad, exagera el número de relaciones sexuales que tiene. En promedio, los hombres casados de menos de 65 años dicen que tienen relaciones una vez por semana. Solo el 1% de ellos dice que lleva un año sin tener relaciones sexuales. Las mujeres casadas dicen que tienen menos relaciones, aunque no mucho menos.

Las búsquedas en Google proporcionan un retrato mucho más vívido —y, a mi entender, más exacto— del sexo en el matrimonio. En Google, la queja más frecuente vinculada al matrimonio es la falta de relaciones sexuales. Las búsquedas de «matrimonio sin sexo» son tres veces y media más numerosas que «matrimonio infeliz» y ocho veces y media más que «matrimonio sin amor». Incluso las parejas que no están casadas se quejan con bastante frecuencia de que no tienen suficientes relaciones. Las búsquedas en Google de «relación sin sexo» solo van por detrás de «relación abusiva». (Estos datos, debo recalcar, se presentan siempre de forma anónima. Por supuesto, Google no aporta información sobre las búsquedas de ningún individuo particular).

Por otra parte, las búsquedas en Google presentaban un retrato de los Estados Unidos sumamente diferente de la utopía posracial esbozada en los sondeos. Recuerdo la primera vez que escribí «nigger» en Google Trends. Se dirá que soy un ingenuo. Pero dado lo incendiaria que es la palabra, estaba bastante convencido de que la búsqueda no tendría muchos datos. Me equivocaba por completo. En los Estados Unidos, la palabra «nigger»

⁶ Datos suministrados al autor.

—o su plural, «*niggers*»— sumaba más o menos el mismo número de búsquedas que «migraña(s)», «economista» y «Lakers». Me pregunté si las búsquedas de letras de rap sesgaban los resultados. Pero no. La grafía utilizada en las canciones de rap es casi siempre «*nigga(s)*» (no «*nigger(s)*»). Así pues, ¿por qué los estadounidenses buscaban «*nigger*»? A menudo, buscaban chistes que se burlaran de los afroamericanos. De hecho, el 20% de las búsquedas con la palabra «*nigger*» también incluía la palabra «chistes». Otras búsquedas frecuentes incluían «*niggers* estúpidos» y «odio a los *niggers*».

Había millones de esas búsquedas por año. En la privacidad de sus hogares, un importante número de estadounidenses hacían indagaciones escandalosamente racistas. Cuanto más investigaba, más perturbadora resultaba ser la información.

La noche en que Obama fue elegido por primera vez, cuando muchos de los comentarios se centraron en elogiar a Obama y destacar el carácter histórico de las elecciones, alrededor de 1 de cada 100 búsquedas en Google de la palabra «Obama» incluía también «KKK» o «*nigger(s)*». Puede parecer una cifra no muy alta, pero piénsense en los miles de motivos no racistas que había para buscar en Google el nombre de aquel inesperado político joven con una familia encantadora que estaba por tomar posesión del cargo más poderoso del mundo. La noche de las elecciones, las búsquedas de Stormfront y las inscripciones en ese sitio nacionalista blanco que goza de una sorprendente popularidad en los Estados Unidos fueron más de diez veces mayores de lo normal.⁷ En algunos estados, hubo más búsquedas de «presidente *nigger*» que de «primer presidente negro».⁸

La maldad y el odio quedaban ocultos en las fuentes tradicionales de información, pero saltaban a la vista en las búsquedas que hacía la gente en internet.

⁷ Datos de Google Trends analizados por el autor. También reuní datos sobre todos los miembros de Stormfront, como se menciona en Seth Stephens-Davidowitz, «The Data of Hate», *New York Times*, 13 de julio de 2014, SR4. Los datos pertinentes pueden descargarse de sethsd.com, en la sección de datos titulada «Stormfront».

⁸ Datos de Google Trends analizados por el autor. Entre los estados en los que esto es cierto figuran Kentucky, Luisiana, Arizona y Carolina del Norte.

Esas búsquedas son difíciles de conciliar con una sociedad en la que el racismo es un factor pequeño. En 2012, yo conocía a Donald J. Trump sobre todo como hombre de negocios y animador de *reality shows*. Ni a mí ni a nadie se le habría pasado por la cabeza que, cuatro años después, sería un candidato presidencial serio. Pero esas horribles búsquedas no son difíciles de compaginar con el éxito de un candidato que, con sus ataques a los inmigrantes y sus muestras de ira y resentimiento, a menudo alimenta las peores inclinaciones de la gente.

Las búsquedas en Google también demostraban que en gran medida nos equivocábamos sobre la ubicación del racismo en Estados Unidos. Los sondeos y la opinión popular situaban el racismo moderno sobre todo en el sur y mayormente entre republicanos. Pero entre los lugares con mayores tasas de búsquedas racistas figuraban el norte del estado de Nueva York, el oeste de Pensilvania, el este de Ohio, la zona industrial de Michigan y la zona rural de Illinois, así como Virginia Occidental, el sur de Luisiana y Misisipi. La verdadera línea divisoria, según sugerían los datos de búsquedas en Google, no estaba entre el sur y el norte; estaba entre el este y el oeste. Uno no encontraba estas cosas muy al oeste de Misisipi. Y el racismo no se limitaba a los republicanos. De hecho, las búsquedas racistas no eran más numerosas en lugares con un alto porcentaje de republicanos que en lugares con un alto porcentaje de demócratas. Dicho de otro modo, las búsquedas en Google ayudaban a trazar un nuevo mapa del racismo en los Estados Unidos, y ese mapa tenía un aspecto diferente del que cualquiera hubiera creído. Los republicanos afincados en el sur podían ser más proclives a admitir el racismo. Pero muchos demócratas del norte albergaban actitudes similares.

Cuatro años más tarde, dicho mapa sería clave para explicar el éxito político de Trump.

En 2012, usé el mapa del racismo desarrollado con las búsquedas en Google para reevaluar exactamente qué papel había desempeñado la raza de Obama. Los datos eran claros. En las zonas del país con un alto número de búsquedas racistas, Obama salió mucho peor parado que, cuatro años antes, John Kerry, el candidato

presidencial demócrata blanco. Ningún otro factor en esas áreas, incluidos los niveles de educación, edad, asistencia a misa o posesión de armas, explicaba la diferencia. Las búsquedas racistas no predecían un mal resultado de ningún otro candidato demócrata. Solo el de Obama.

Y en los resultados iba implícito un efecto a gran escala. Obama perdió alrededor de 4 puntos porcentuales en todo el país solo debido al racismo explícito. Aquella era una cifra mucho más alta de la que cabía esperar de acuerdo con los sondeos. Barack Obama, por supuesto, fue elegido y reelegido presidente con la ayuda de condiciones muy favorables para los demócratas, pero tuvo que superar bastantes más obstáculos de los que era consciente cualquiera que dependiera de las fuentes de datos tradicionales (y eso incluía a casi todo el mundo). Existían suficientes personas racistas como para ayudar a ganar unas primarias o inclinar la balanza en unas elecciones generales en un año que no fuese tan favorable para los demócratas.

Al principio, mi estudio fue rechazado por cinco publicaciones académicas.⁹ Muchos de mis homólogos revisores, si me disculpan apuntar la discrepancia, dijeron que era imposible creer que tantos estadounidenses albergaran un racismo tan horrendo. Sencillamente, ello no encajaba con lo que se decía por entonces. Además, las búsquedas en Google parecían un conjunto de datos muy extraño.

Ahora que hemos presenciado la toma de posesión del presidente Donald J. Trump, mi conclusión parece más plausible.

Cuanto más estudio el tema, más descubro que Google posee una gran cantidad de información que los sondeos omiten y que nos puede ayudar a entender —entre muchísimas otras cosas— unas elecciones.

⁹ El artículo acabó publicándose como Seth Stephens-Davidowitz, «The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data», *Journal of Public Economics* 118 (2014). En él pueden encontrarse más detalles sobre la investigación. Además, los datos pueden encontrarse en mi sitio web, sethsd.com, en la sección de datos titulada «Racism».

Hay información sobre quiénes irán a votar. Más de la mitad de los ciudadanos que no votan dicen en los sondeos inmediatamente anteriores a unas elecciones que tienen la intención de hacerlo, lo que sesga nuestra estimación de la participación, mientras que las búsquedas hechas en Google con las frases «cómo votar» o «dónde votar» semanas antes de las elecciones pueden predecir con exactitud en qué partes del país se acudirá en masa a las urnas.

Incluso puede haber información sobre por quién se votará. ¿En serio podemos predecir por qué candidatos votará la gente en función de lo que busca? Claramente, no podemos limitarnos a estudiar qué candidatos se buscan con más frecuencia. Muchas personas buscan a un candidato porque lo adoran. Otras buscan a un candidato porque lo odian. Dicho eso, Stuart Gabriel, un profesor de finanzas en la Universidad de California, Los Ángeles, y yo hemos encontrado una sorprendente clave para averiguar de qué manera planea votar la gente. Un gran porcentaje de las búsquedas relacionadas con las elecciones incluyen preguntas sobre los nombres de dos candidatos. Durante la contienda de 2016 entre Trump y Hillary Clinton, había gente que buscaba «sondeos Trump Clinton». Otra buscaba los platos fuertes del «debate Trump Clinton». De hecho, el 12% de las búsquedas con «Trump» también incluían la palabra «Clinton». Más de un cuarto de las búsquedas vinculadas a «Clinton» incluían también la palabra «Trump».

Descubrimos que esas búsquedas de apariencia neutral en realidad pueden darnos pistas sobre el candidato al que apoya una persona.

¿Cómo? El orden en que aparecen los nombres. Nuestra investigación sugiere que una persona es notablemente más propensa a poner primero a su candidato preferido en una búsqueda que incluya los nombres de los dos candidatos.

En las tres elecciones anteriores, el candidato que apareció primero en más búsquedas recibió la mayor cantidad de votos. Más interesante aún, el orden en que se buscaba a los candidatos era predictivo del resultado de un estado particular.

El orden en el que se busca a los candidatos también parece incluir información que los sondeos tienden a pasar por alto. Para las elecciones de 2012 disputadas por Obama y el republicano Mitt

Romney, Nate Silver, estadístico y periodista virtuoso, predijo con exactitud los resultados de los cincuenta estados. Con todo, descubrimos que en los estados donde se buscaba con más frecuencia a Romney antes de Obama, Romney tuvo un mejor rendimiento que el predicho por Silver. En los estados en los que se buscaba con más frecuencia a Obama antes de Romney, Obama tuvo un mejor rendimiento que el predicho por Silver.

Ese indicador podría contener información que los sondeos pasan por alto, ya sea porque los votantes se mienten a sí mismos o porque les incomoda revelar sus preferencias reales a los encuestadores. Es posible que, aunque afirmasen en 2012 que estaban indecisos, si buscaban siempre «sondeos Romney Obama», «debate Romney Obama» y «elecciones Romney Obama», planearan votar por Romney desde el comienzo.

¿Predijo Google a Trump? Bueno, aún nos queda mucho trabajo por hacer, y deberán sumárseme a la tarea muchos investigadores antes de que sepamos cómo optimizar los datos de Google para predecir los resultados de unas elecciones. Se trata de una ciencia nueva, y solo hemos contado con estos datos para unas pocas elecciones. Por cierto, con ello no quiero decir que estemos preparados, ni que alguna vez vayamos a estarlo, para prescindir de las encuestas de opinión pública como herramientas que ayudan a predecir las elecciones.

Pero sin duda en internet hubo anuncios, en muchos aspectos, de que Trump sacaría mejores resultados de lo predichos en los sondeos.

Durante las elecciones generales, hubo indicios de que el electorado podría favorecer a Trump. Los estadounidenses negros prometieron en los sondeos que acudirían a las urnas en masa para oponerse a Trump. Pero Google registró muy pocas búsquedas de información sobre cómo votar en las zonas con alta población negra. Y el día de las elecciones, Clinton salió afectada por la baja participación de los negros.

Incluso hubo señales de que los votantes supuestamente indecisos preferían a Trump. Gabriel y yo hallamos que hubo más búsquedas de «Trump Clinton» que de «Clinton Trump» en los estados clave del medio oeste en los que se esperaba que Clinton

ganara. De hecho, Trump ganó las elecciones porque superó con creces las predicciones de los sondeos respecto de esa zona.

Pero la clave principal, a mi entender, de que Trump podía ser un candidato exitoso —para empezar, en las primarias— era el racismo oculto que había descubierto mi estudio sobre Obama. Las búsquedas en Google revelaron que un número importante de estadounidenses siguen albergando un odio y una maldad que los expertos han pasado por alto durante muchos años. Los datos de las búsquedas revelaron que vivíamos en una sociedad muy distinta de la sociedad en la que creían vivir los académicos y periodistas sobre la base de los sondeos. Revelaban una rabia repugnante, aterradora y generalizada que esperaba a que un candidato le diera voz.

Con frecuencia la gente miente: a sí misma y a los demás. En 2008, los estadounidenses dijeron en los sondeos que ya no les importaba la raza. Ocho años después, eligieron presidente a Donald J. Trump, un hombre que retuiteó una afirmación falsa acerca de que los negros son responsables de la mayoría de los asesinatos de estadounidenses blancos, defendió en uno de sus mítines a sus partidarios por haberle dado una paliza a un manifestante de Black Lives Matters y dudó en rechazar el apoyo de un exlíder del Ku Klux Klan. El mismo racismo oculto que había perjudicado a Barack Obama ayudó a Donald Trump.

Es bien sabido que, a principios de las primarias, Nate Silver afirmó que Trump no tenía prácticamente ninguna posibilidad de salir ganador. Conforme avanzaban las primarias y se hacía cada vez más evidente que Trump contaba con bases amplias, Silver decidió mirar los datos para tratar de comprender qué estaba sucediendo. ¿Cómo era posible que a Trump le fuera tan bien?

Silver notó que las zonas en las que Trump sacaba los mejores resultados formaban un mapa extraño. Trump tenía bastante éxito en algunas partes del noreste y el medio oeste industrial, así como en el sur. Le iba notablemente peor en el oeste. Silver buscó variables que explicaran ese mapa. ¿El desempleo? ¿La religión? ¿La tenencia de armas? ¿Las tasas de inmigración? ¿La oposición a Obama?

Silver halló que el factor singular que mejor se correlacionaba con el apoyo a Donald Trump en las primarias republicanas era la

medida que yo había descubierto cuatro años antes.¹⁰ Las zonas con un número mayor de apoyos a Trump eran las mismas en las que se hacían la mayor cantidad de búsquedas en Google de «nigger».¹¹

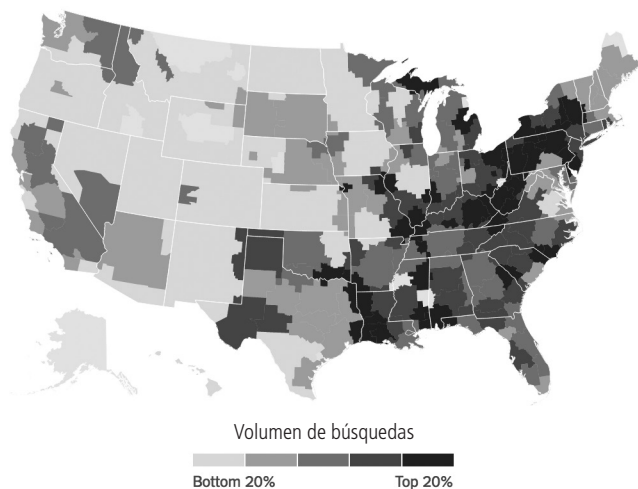
He pasado casi todos los días de los últimos cuatro años analizando los datos de Google. Esa actividad incluyó una temporada en calidad de analista de datos en Google, donde me contrataron tras enterarse de mi investigación sobre el racismo. Sigo examinando los datos como columnista y periodista de datos para *The New York Times*. Las revelaciones no se han agotado. Enfermedad mental, sexualidad humana, abuso infantil, aborto, publicidad, religión, salud. No son temas precisamente baladíes, y este conjunto de datos, inexistente hace dos decenios, ofrece perspectivas nuevas y sorprendentes sobre todos ellos. Los economistas y otros científicos sociales siempre están en busca de nuevas fuentes de datos, así que lo diré sin rodeos: estoy convencido de que actualmente las búsquedas en Google constituyen el conjunto de datos más importante jamás recopilado sobre la psique humana.

Ese conjunto de datos, sin embargo, no es la única herramienta que ha proporcionado internet para entender el mundo. Pronto me di cuenta de que también existen otras minas de oro digitales. Descargué toda la Wikipedia, estudié perfiles de Facebook y me asomé a Stormfront. Además, Pornhub, uno de los sitios pornográficos más grandes de internet, me brindó acceso a todos los datos con que contaban sobre las visualizaciones y búsquedas

¹⁰ «La correlación más fuerte del apoyo a Trump son las búsquedas en Google de la palabra que empieza con n-» (*tweet* del 28 de febrero de 2016). Véase también Nate Cohn, «Donald Trump's Strongest Supporters: A New Kind of Democrat», *New York Times*, 31 de diciembre de 2015, A3.

¹¹ Los mapas muestran el porcentaje de las búsquedas en Google que incluyen la palabra «nigger(s)». Nótese que, como la medida es un porcentaje de las búsquedas en Google, no es arbitrariamente más alta en sitios más poblados o en lugares donde se hacen muchas búsquedas. Nótese también que algunas de las diferencias entre este mapa y el mapa del apoyo a Trump tienen explicaciones obvias. Trump perdió popularidad en Texas y Arkansas porque eran los estados natales de dos de sus oponentes, Ted Cruz y Mike Huckabee. Los datos provienen de encuestas hechas por Civis Analytics en diciembre de 2015. Los datos reales de las votaciones son menos útiles en este caso, porque están muy influenciados por cuándo tuvieron lugar las primarias y por la forma de votar. Los mapas se reproducen con permiso de *The New York Times*.

Tasa de búsquedas racistas



Apoyo a Donald Trump en las primarias republicanas

