

# EVERYBODY LIES

BIG DATA, NEW DATA,  
AND WHAT THE INTERNET  
CAN TELL US ABOUT WHO  
WE REALLY ARE



SETH STEPHENS-DAVIDOWITZ

FOREWORD BY STEVEN PINKER

# Tabla de contenido

Página del título

Prefacio

Dedicación

Contenido

Introducción: los contornos de una revolución

PARTE I: DATOS, grandes y pequeños

1. Su Gut defectuoso

PARTE II: LOS PODERES DE DATOS BIG

2. ¿Fue Freud, ¿verdad?

3. Los datos Reimagined

Organismos como Data

Palabras como Data

Imágenes como datos

4. suero de la verdad Digital

La verdad sobre el sexo

La verdad sobre el odio y prejuicio

La verdad sobre Internet

La verdad sobre el abuso infantil y el Aborto

La verdad sobre sus amigos de Facebook

La verdad sobre sus clientes

¿Podemos con la verdad?

5. Ampliar la imagen

Lo que realmente está pasando en nuestras condados, ciudades y pueblos?

Cómo llenamos nuestros minutos y horas

nuestros Doppelgangers

Historias de datos

6. Todo el mundo es un laboratorio

El ABC de las pruebas A / B

Cruel, pero la aclaración-experimentos de la naturaleza

PARTE III: DATOS BIG: Manipular con cuidado

7. Big Data, Big Schmata? Lo que no puede hacer

La maldición de la dimensionalidad

El énfasis excesivo en lo que es medible

8. Los datos Mo, Mo problemas? Lo que no debemos hacer

El peligro de las corporaciones Empowered

El peligro de los gobiernos Facultados

Conclusión: ¿Cuántas personas Acabado Libros?

[Expresiones de gratitud](#)

[notas](#)

[Índice](#)

[Sobre el Autor](#)

[Derechos de autor](#)

[Acerca del Editor](#)

# **EVERY**

 **DEY ST.**  
*An Imprint of WILLIAM MORROW*

# **BODY**

---

# **LIES**

---

Big Data, New Data,  
and What the Internet  
Reveals About  
Who We Really Are

**SETH**  
**STEPHENS-DAVIDOWITZ**

# DEDICACIÓN



*Para mamá y papá*

# CONTENIDO

---

[Cubrir](#)

[Página del título](#)

[Dedicación](#)

[Prólogo de Steven Pinker](#)

[Introducción: los contornos de una revolución](#)

## **PARTE I: DATOS, grandes y pequeños**

---

[1. Su Gut defectuoso](#)

## **PARTE II: LOS PODERES DE DATOS BIG**

---

[2. ¿Fue Freud, ¿verdad?](#)

[3. Los datos Reimagined](#)

[Organismos como Data](#)

[Palabras como Data](#)

[Imágenes como datos](#)

[4. suero de la verdad Digital](#)

[La verdad sobre el sexo](#)

[La verdad sobre el odio y prejuicio](#)

[La verdad sobre Internet](#)

[La verdad sobre el abuso infantil y el Aborto](#)

[La verdad sobre sus amigos de Facebook](#)

[La verdad sobre sus clientes](#)

[¿Podemos con la verdad?](#)

[5. Ampliar la imagen](#)

[Lo que realmente está pasando en nuestras condados, ciudades y pueblos?](#)

[Cómo llenamos nuestros minutos y horas](#)

[nuestros Doppelgangers](#)

[Historias de datos](#)

[6. Todo el mundo es un laboratorio](#)

[El ABC de las pruebas A / B](#)

[Cruel, pero la aclaración-experimentos de la naturaleza](#)

### **PARTE III: Big data: Manipular con cuidado**

---

[7. Big Data, Big Schmata? Lo que no puede hacer](#)

[La maldición de la dimensionalidad](#)

[El énfasis excesivo en lo que es medible](#)

[8. Los datos Mo, Mo problemas? Lo que no debemos hacer](#)

[El peligro de las corporaciones Empowered](#)

[El peligro de los gobiernos Facultados](#)

[Conclusión: ¿Cuántas personas Acabado Libros?](#)

[Expresiones de gratitud](#)

[notas](#)

[Índice](#)

[Sobre el Autor](#)

[Derechos de autor](#)

[Acerca del Editor](#)



## PREFACIO

---

**mi**esde filósofos especularon sobre un “cerebroscopio,” un dispositivo mítico que mostraría pensamientos de una persona en una pantalla, los científicos sociales han estado buscando herramientas para exponer el funcionamiento de la naturaleza humana. Durante mi carrera como psicólogo experimental, los diferentes han entrado y salido de la moda, y los he probado todos escalas de calificación, tiempos de reacción, dilatación de la pupila, de neuroimagen funcional, incluso los pacientes con epilepsia con electrodos implantados que estaban dispuestos a pasar las horas en un experimento de lenguaje a la espera de tener una convulsión.

Sin embargo, ninguno de estos métodos proporciona una vista sin obstáculos en la mente. El problema es una solución de compromiso salvaje. Los pensamientos humanos son proposiciones complejas; a diferencia de Woody Allen Guerra de lectura rápida y Paz, no nos limitamos a pensar “Se trataba de algunos rusos.” Pero proposiciones en todo su esplendor multidimensional enredada son difíciles para un científico para analizar. Por supuesto, cuando la gente vierten sus corazones, aprehendemos la riqueza de su corriente de conciencia, pero monólogos no somos un conjunto de datos ideal para probar hipótesis. Por otro lado, si nos concentramos en medidas que son fácilmente cuantificables, como el tiempo de reacción de la gente a las palabras, o su respuesta de la piel a las imágenes, podemos hacer las estadísticas, pero hemos hecho puré el complejo de la textura de la cognición en un solo número.

Como si el compromiso entre manejabilidad y la riqueza no fuera suficiente, los científicos de la naturaleza humana se molieron por la Ley de pequeños números-Amos Tversky y Daniel Kahneman nombre para el error de pensar que los rasgos de una población se reflejarán en ninguna muestra, no importa que tan pequeño. Incluso los científicos más cálculos aritméticos tienen intuiciones lamentablemente defectuosos unos cuantos temas que uno realmente necesita en un estudio antes de que uno abstracta lata lejos de las peculiaridades azar y golpes y generalizar a todos los estadounidenses, por no hablar de los Homo sapiens. Es todo el iffier cuando la muestra es recogida por conveniencia, como por ejemplo, ofreciendo cerveza dinero a los estudiantes de segundo año en nuestros cursos.

Este libro trata de una nueva forma de estudiar la mente. Grandes Los datos de las búsquedas en Internet y otras respuestas en línea no son un cerebroscopio, pero Seth Stephens-Davidowitz muestra que ofrecen una mirada sin precedentes en la psique de las personas. En la privacidad de sus teclados, la gente confesar las cosas más extrañas, a veces (como en los sitios de citas o busca asesoramiento profesional), ya que tienen consecuencias en la vida real, en otras ocasiones, precisamente porque no tienen consecuencias: la gente puede desahogarse de algún deseo o miedo sin una persona real reaccionar con consternación o peor. De cualquier manera,

las personas no son sólo presionando un botón o girando un botón, pero teclear cualquiera de billones de secuencias de caracteres para explicar sus pensamientos en toda su inmensidad explosiva, combinatoria. Mejor aún, que establecen estas huellas digitales en una forma que es fácil de agregar y analizar. Vienen de todas las clases sociales. Pueden tomar parte en experimentos discretos que pueden variar los estímulos y tabular las respuestas en tiempo real. Y felizmente suministran estos datos en números enormes.

*Todos mientenes* más que una prueba de concepto. Una y otra vez mis ideas preconcebidas acerca de mi país y de mi especie se convirtieron al revés por los descubrimientos de Stephens-Davidowitz. ¿De dónde apoyo inesperado de Donald Trump viene? Cuando Ann Landers pidió a sus lectores en 1976 si se lamentaban de tener hijos y se sorprendió al descubrir que la mayoría lo hizo, fue engañado por ella una muestra no representativa, auto seleccionado? Es el Internet el culpable de que la crisis de forma redundante con nombre de finales de los años 2010, la “burbuja filtro”? Lo que desencadena los crímenes de odio? ¿Las personas buscan chistes para animar a sí mismos? Y aunque me gusta pensar que nada me puede sorprender, me sorprendió en abundancia por lo que el Internet revela acerca de la sexualidad humana, incluyendo el descubrimiento de que todos los meses un cierto número de mujeres búsqueda “humping animales de peluche.

Todo el mundo va a disfrutar de todo el mundo miente. Con la curiosidad inagotable y un ingenio entrañable, Stephens-Davidowitz apunta a un nuevo camino para las ciencias sociales en el siglo XXI. Con esta ventana infinitamente fascinante en obsesiones humanas, que necesita un cerebroscopio?

-Steven Pinker, 2017

# INTRODUCCIÓN

---

## Los contornos de una revolución

**S**urely que perdería, dijeron.

En las primarias republicanas de 2016, los expertos concluyeron que la votación Donald Trump no tenía ninguna posibilidad. Después de todo, Trump había insultado a una variedad de grupos minoritarios. Las encuestas y sus intérpretes nos dijeron que pocos estadounidenses acreditados de esos atropellos.

La mayoría de los expertos electorales en el momento pensaron que Trump perdería en la elección general. Demasiados votantes dijeron que se deje intimidar por su forma y puntos de vista.

Pero en realidad había algunas pistas que Trump en realidad podría ganar tanto las primarias y la elección-en Internet en general.

Soy un experto de datos de Internet. Cada día, el seguimiento de los rastros digitales que las personas dejan ya que hacen su camino a través de la web. De los botones o teclas que Haga clic o toque, trato de entender lo que realmente queremos, lo que realmente va a hacer, y lo que realmente somos. Voy a explicar cómo me inicié en este camino inusual.

La historia comienza -y esto parece como hace siglos, con la elección presidencial de 2008 y una cuestión largamente debatida en las ciencias sociales: ¿Qué importancia tiene el prejuicio racial en Estados Unidos?

Barack Obama estaba funcionando como el primer candidato presidencial afroamericano de un partido importante. Ganó-con bastante facilidad. Y las encuestas sugieren que la raza no fue un factor en cómo votaron los estadounidenses. Gallup, por ejemplo, llevó a cabo numerosas encuestas antes y después de la primera elección de Obama. Su conclusión? Los votantes estadounidenses en gran medida no les importaba que Barack Obama era negro. Poco después de la elección, dos profesores bien conocidos en la Universidad de California, Berkeley estudiaron minuciosamente a través de otros datos basados en encuestas, utilizando técnicas de minería de datos más sofisticados. Llegaron a una conclusión similar.

Y así, durante la presidencia de Obama, esto se convirtió en la sabiduría convencional en muchas partes de los medios de comunicación y en grandes sectores de la academia. Las fuentes

que los medios de comunicación y los científicos sociales han utilizado durante más de ochenta años para entender el mundo que nos dijeron que la abrumadora mayoría de los estadounidenses no le importaba que Obama era negro cuando se juzga si debería ser su presidente.

Este país, siempre manchada por la esclavitud y la segregación, parecía finalmente haber dejado de juzgar a la gente por el color de su piel. Esto parece sugerir que el racismo en las últimas en Estados Unidos. De hecho, algunos expertos incluso declararon que vivimos en una sociedad post-racial.

En 2012, yo era un estudiante graduado en economía, perdido en la vida, quemado en mi campo, y con confianza, incluso arrogante, que tenía una muy buena comprensión de cómo funciona el mundo, de lo que la gente pensaba y se preocupaba en el siglo veintiuno. Y cuando se trata de este tema de los prejuicios, me he dejado de creer, sobre la base de todo lo que había leído en la psicología y la ciencia política, que el racismo explícito se limita a un pequeño porcentaje de los estadounidenses de la mayoría de ellos republicanos conservadores, la mayoría de ellos que viven en el sur profundo.

Entonces, me encontré con Google Trends.

Google Trends, una herramienta que se publicó con poca fanfarria en 2009, indica a los usuarios la frecuencia con cualquier palabra o frase se ha buscado en diferentes lugares en diferentes momentos. Se anuncia como una herramienta que permite, tal vez divertido amigos para discutir qué celebridad era la más popular o lo que la moda era repentinamente caliente. Las primeras versiones incluyen una advertencia lúdico que la gente “no quiere escribir su tesis doctoral” con los datos, lo que me motivó de inmediato a escribir mi tesis con él.\*

En ese momento, los datos de búsqueda de Google no parecen ser una fuente adecuada de la información para la investigación académica “grave”. A diferencia de las encuestas, los datos de búsqueda de Google no fue creado como una manera de ayudar a entender la psique humana. Google se inventó para que las personas puedan aprender sobre el mundo, no lo que los investigadores podrían aprender acerca de las personas. Pero resulta que los rastros que dejamos en nuestra búsqueda de conocimiento de Internet son tremendamente revelador.

En otras palabras, la búsqueda de la gente para obtener información es, en sí mismo, la información. Cuándo y dónde buscar datos, citas, chistes, lugares, personas, cosas, o ayudar, resulta que nos puede decir mucho más acerca de lo que realmente piensan, realmente deseo, realmente temer, y realmente hacer lo que nadie podría tener adivinado. Esto es especialmente cierto ya que a veces la gente no tanto la consulta como Google confie en él: “Yo odio a mi jefe” “estoy borracho” “Mi padre me golpeó.”.

El acto cotidiano de escribir una palabra o frase en una caja blanca compacta, rectangular deja un pequeño rastro de la verdad que, cuando se multiplica por millones, finalmente revela realidades profundas. La primera palabra que escribió en Google Trends era “Dios.” He aprendido que los estados que conforman la mayoría de las búsquedas de Google mencionan “Dios” fueron Alabama, Mississippi y Arkansas-Cinturón de la Biblia. Y esas búsquedas son más frecuentes los domingos. Nada de lo cual fue sorprendente, pero era intrigante que los datos de búsqueda podrían revelar un patrón tan claro. Probé “Knicks”, que resulta que es buscado en

Google las más importantes en la ciudad de Nueva York. Otra obviedad. A continuación, he escrito en mi nombre. “Lo sentimos,” Google Trends me informó. “No hay suficiente volumen de búsquedas” para mostrar los resultados. Google Trends, supe, proporcionará datos sólo cuando mucha gente hace la misma búsqueda.

Pero el poder de las búsquedas de Google no es que ellos nos pueden decir que Dios es muy popular en el Sur, los Knicks son muy populares en la ciudad de Nueva York, o que no soy muy popular en cualquier lugar. Cualquier encuesta se podría decir eso. El poder en datos de Google es que la gente cuenta las cosas gigantes del motor de búsqueda puede ser que no lo cuentes a nadie.

Tomemos, por ejemplo, el sexo (un tema que va a investigar con mayor detalle más adelante en este libro). Las encuestas no se puede confiar en que nos diga la verdad sobre nuestra vida sexual. He analizado los datos de la Encuesta Social General, que se considera una de las fuentes más influyentes y con autoridad para obtener información sobre los comportamientos de los estadounidenses. De acuerdo a la encuesta, cuando se trata de relaciones heterosexuales, las mujeres dicen que tienen relaciones sexuales, en promedio, cincuenta y cinco veces al año, utilizando un condón 16 por ciento de las veces. Esto se suma a alrededor de 1,1 mil millones de condones usados por año. Pero los hombres heterosexuales dicen que usan 1,6 mil millones de condones al año. Esos números, por definición, tendría que ser el mismo. Entonces, ¿quién está diciendo la verdad, los hombres o las mujeres?

Ninguno de los dos, resulta. De acuerdo con Nielsen, la información y la medición compañía global que rastrea el comportamiento del consumidor, menos de 600 millones de condones se venden cada año. Así que todo el mundo es la mentira; la única diferencia es en qué medida.

La mentira es, de hecho, muy extendida. Hombres que nunca han estado casadas afirman utilizar un promedio de veintinueve condones por año. Esto añadiría hasta más que el número total de condones que se venden en los Estados Unidos para las personas casadas y solteras combinados. Las personas casadas probablemente exagera la cantidad de sexo que tienen, también. En promedio, los hombres casados menores de sesenta y cinco encuestas dicen que tienen sexo una vez a la semana. Sólo el 1 por ciento dice que han pasado el último año sin sexo. Las mujeres casadas informan que tienen un poco menos el sexo, pero no mucho menos.

búsquedas de Google dan una mucho menos animado y, argumento, mucho más precisa y en la imagen del sexo durante el matrimonio. En Google, la mayor queja acerca de un matrimonio no está teniendo sexo. Las búsquedas de “matrimonio sin sexo” son tres veces y medio más común que el “matrimonio infeliz” y ocho veces más común que el “matrimonio sin amor.” Incluso las parejas casadas se quejan con frecuencia algo acerca de no tener relaciones sexuales. Google busca “relación sin sexo” están en segundo lugar solamente a las búsquedas de “relación abusiva.” (Estos datos, que debería destacar, es todo presentado de forma anónima. Google, por supuesto, no reporta datos sobre búsquedas de cualquier individuo en particular).

Y las búsquedas de Google presentaron una imagen de América que era totalmente distinta a la utopía post-racial esbozado por las encuestas. Recuerdo la primera vez que escribimos “negro” en Google Trends. Llámame ingenuo. Pero teniendo en cuenta el grado de toxicidad es la

palabra, estaba totalmente convencida que esto es una búsqueda de bajo volumen. Boy, estaba equivocado. En los Estados Unidos, la palabra “negro”, o su plural, “negros” -era incluyen en aproximadamente el mismo número de búsquedas como la palabra “migraña (s)”, “economista”, y “Lakers.” Me preguntaba si búsquedas de letras de rap fueron sesgando los resultados? Nop. La palabra usada en canciones de rap es casi siempre “negro (s).” Entonces, ¿cuál fue la motivación de los estadounidenses en busca de “negro”? Con frecuencia, los que buscaban chistes burlones afroamericanos. De hecho, el 20 por ciento de las búsquedas con la palabra “negro” también incluye la palabra “chistes.

Había millones de estas búsquedas cada año. Un gran número de estadounidenses eran, en la intimidad de sus propios hogares, haciendo preguntas sorprendentemente racistas. El más investigaba, más preocupante la información tiene.

En la primera noche de la elección de Obama, cuando la mayoría de los comentarios se centraron en elogio de Obama y el reconocimiento de la naturaleza histórica de su elección, aproximadamente uno de cada cien búsquedas de Google que incluían la palabra “Obama” también incluyó “KKK” o “negro (s).” Tal vez que no suena tan alto, pero piensa en los miles de razones no racista a Google este joven forastero con una familia encantadora punto de apoderarse de trabajo más potente del mundo. En la noche electoral, las búsquedas y las inscripciones para Stormfront, un sitio nacionalista blanco con sorprendentemente alta popularidad en los Estados Unidos, fueron más de diez veces superior a la normal. En algunos estados, hubo más búsquedas de “presidente negro” de “primer presidente negro”.

Había una oscuridad y odio que estaba oculto a las fuentes tradicionales, pero era bastante evidente en las búsquedas que las personas realizan.

Esas búsquedas son difíciles de conciliar con una sociedad en la que el racismo es un factor pequeño. En 2012 sabía de Donald J. Trump sobre todo como un hombre de negocios y la demostración de la realidad intérprete. No tenía más idea que nadie más que él, cuatro años más tarde, ser un candidato serio. Pero esas búsquedas feas no son difíciles de conciliar con el éxito de un candidato que en sus ataques a los inmigrantes, en sus enojos y resentimientos-juega a menudo peores que las inclinaciones de las personas.

Las búsquedas de Google también nos dijeron que gran parte de lo que pensamos acerca de la ubicación del racismo estaba mal. Las encuestas y la sabiduría convencional colocan racismo moderno principalmente en el sur y en su mayoría entre los republicanos. Pero los lugares con las tasas más altas de búsqueda racistas incluyen el estado de Nueva York, Pensilvania occidental, el este de Ohio, Michigan e Illinois industrial rural, junto con Virginia Occidental, el sur de Louisiana y Mississippi. La verdadera división, los datos de búsqueda de Google sugiere, no era del Sur frente al Norte; que era Este contra Oeste. Usted no recibe este tipo de cosas mucho al oeste del Mississippi. Y el racismo no se limita a los republicanos. De hecho, las búsquedas racistas no fueron mayores en lugares con un alto porcentaje de los republicanos que en lugares con un alto porcentaje de demócratas. Google busca, en otras palabras, ayudó a dibujar un nuevo mapa de racismo en los Estados Unidos, y este mapa se veía muy diferente de

lo que puede haber adivinado. Los republicanos en el sur pueden ser más propensas a admitir al racismo. Pero un montón de Demócratas en el Norte tienen actitudes similares.

Cuatro años más tarde, este mapa podría resultar muy importante para explicar el éxito político de Trump.

En 2012, yo estaba usando este mapa del racismo que había desarrollado a través de Google busca volver a evaluar con exactitud el papel que juega la raza de Obama. Los datos estaba claro. En algunas partes del país con un alto número de búsquedas racistas, Obama hizo sustancialmente peor que John Kerry, el candidato demócrata a la presidencia blanca, tenía cuatro años antes. La relación no se explica por cualquier otro factor sobre estas áreas, incluyendo los niveles de educación, la edad, asistencia a la iglesia, o la propiedad de armas. búsquedas racistas no predijeron bajo rendimiento para cualquier otro candidato demócrata. Sólo para Obama.

Y los resultados implicaron un gran efecto. Obama perdió aproximadamente 4 puntos porcentuales a nivel nacional sólo de racismo explícito. Esto era mucho más alta que se podría haber esperado basado en las encuestas. Barack Obama, por supuesto, fue elegido y reelegido presidente, ayudado por unas condiciones muy favorables para los demócratas, pero tuvo que superar un poco más de todo aquel que se basa en datos tradicionales fuentes, y eso fue casi todo el mundo, se había dado cuenta. Había suficientes racistas para ayudar a ganar una primaria o inclinar una elección general en un año no tan favorable a los demócratas.

Mi estudio fue rechazado inicialmente por cinco publicaciones académicas. Muchos de los revisores, si va a perdonar un poco de descontento, dijo que era imposible creer que tantos estadounidenses albergaba tales racismo vicioso. Esto simplemente no encajaba con lo que la gente había estado diciendo. Además, las búsquedas de Google parecía un extraño conjunto de datos tales.

Ahora que hemos asistido a la inauguración del presidente Donald J. Trump, mi hallazgo parece más plausible.

Cuanto más he estudiado, más me he enterado de que Google tiene una gran cantidad de información que se pierde por las encuestas que pueden ser útiles en la comprensión-entre muchos, muchos otros temas, una elección.

No hay información sobre que en realidad va a salir a votar. Más de la mitad de los ciudadanos que no votan diga encuestas inmediatamente antes de una elección que se proponen, sesgando nuestra estimación de participación, mientras que las búsquedas de Google para “cómo votar” o “dónde votar” semanas antes de una elección puede predecir con exactitud el cual partes del país van a tener una gran actuación en las urnas.

También podría suceder que la información sobre quién van a votar a favor. Realmente podemos predecir qué personas candidato va a votar por sólo sobre la base de lo que buscan? Claramente, no podemos estudiar qué candidatos se buscan con más frecuencia. Muchas personas la búsqueda de un candidato, ya que lo aman. Un número similar de personas búsqueda de un candidato, ya que lo odian. Dicho esto, Stuart Gabriel, profesor de finanzas en la

Universidad de California, Los Ángeles, y he encontrado una pista acerca de qué manera sorprendente, la gente está planeando para votar. Un gran porcentaje de las búsquedas relacionadas con las elecciones contiene consultas con los nombres de ambos candidatos. Durante las elecciones de 2016 entre Trump y Hillary Clinton, algunas personas buscaron “encuestas Trump Clinton”. Otros buscaron aspectos más destacados de la “debate Clinton Trump”. De hecho,

Hemos encontrado que estas búsquedas aparentemente neutrales en realidad nos pueden dar algunas pistas a cuál de los candidatos a una persona soporta.

¿Cómo? El orden en el que aparecen los candidatos. Nuestra investigación sugiere que una persona es significativamente más propensos a poner el candidato que apoyan por primera vez en una búsqueda que incluye los nombres de ambos candidatos.

En las tres elecciones anteriores, el candidato que apareció por primera vez en más búsquedas recibió la mayoría de votos. Más interesante, el orden en que se buscaron los candidatos fue predictivo de qué manera se iría un estado particular.

El orden en el que se buscan candidatos también parece contener información que las encuestas se puede perder. En las elecciones de 2012 entre Obama y el republicano Mitt Romney, Nate Silver, el estadístico virtuoso y periodista, predijo con exactitud el resultado en los cincuenta estados. Sin embargo, se encontró que en los estados que aparecen antes de que Obama Romney en las búsquedas con mayor frecuencia, Romney hizo realidad mejor que la plata había predicho. En los estados que más frecuentemente aparecen Obama antes de Romney, Obama hizo mejor que la plata había predicho.

Este indicador podría contener información que las encuestas se pierda porque los votantes están mintiendo a sí mismos o incómodo revelar sus verdaderas preferencias en las encuestas. Tal vez si ellos afirmaron que estaban indecisos en 2012, pero estaban buscando constantemente por “boca de Romney Obama”, “debate Romney Obama”, y “elección Romney Obama,” que estaban planeando votar por Romney todo el tiempo.

Así qué Google predecir Trump? Bueno, todavía tenemos mucho trabajo que hacer, y voy a tener que ser acompañado por un montón más investigadores-antes de que sepamos mejor manera de utilizar los datos de Google para predecir resultados de las elecciones. Esta es una nueva ciencia, y sólo tenemos unas pocas elecciones para los que existe estos datos. Desde luego, no estoy diciendo que estamos en el punto o será en el punto en el que podemos echar a los sondeos de opinión por completo como una herramienta para ayudar a predecir las elecciones.

Pero sin duda alguna había presagios, en muchos puntos, en el Internet que Trump podría hacer mejor que las encuestas predecían.

Durante la elección general, no hubo indicios de que el electorado podría ser una favorable para Trump. Negros estadounidenses dijeron encuestas que resultarían en gran número para oponerse Trump. Sin embargo, Google busca información sobre la votación en áreas muy negros eran camino hacia abajo. El día de elección, Clinton se verían perjudicados por la baja participación negro.



Incluso hubo indicios de que los votantes indecisos supuestamente iban camino de Trump. Gabriel y yo nos pareció que había más búsquedas de “Trump Clinton” que “Clinton Trump” en estados clave en el Medio Oeste que se esperaba que Clinton para ganar. De hecho, Trump debía su elección al hecho de que él bruscamente superó sus urnas allí.

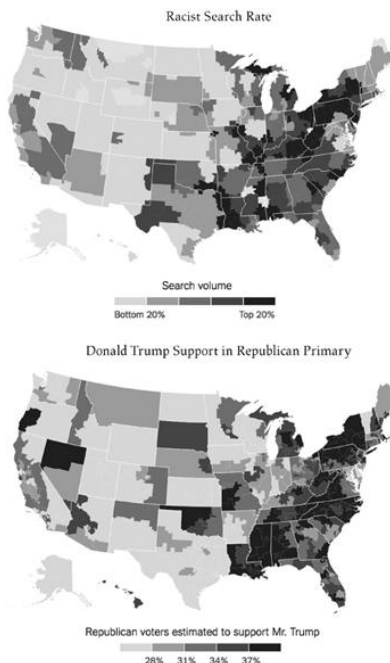
Pero la pista importante, yo diría, que Trump podría resultar un exitoso candidato en las primarias, para empezar, era todo lo que el racismo es un secreto que mi estudio Obama había descubierto. Las búsquedas de Google reveló una oscuridad y odio entre un número significativo de los estadounidenses de que los expertos, desde hace muchos años, se perdió. búsqueda de datos revelaron que vivimos en una sociedad muy diferente de la que académicos y periodistas, basándose en encuestas, pensamos que vivimos en. El estudio reveló una rabia desagradable, miedo, y generalizada de que estaba esperando un candidato para dar voz a la misma.

La gente con frecuencia se encuentran a sí mismos ya los demás. En 2008, los estadounidenses dijeron encuestas que ya no se preocupaban por la raza. Ocho años más tarde, eligieron como presidente Donald J. Trump, un hombre que retweeted una falsa afirmación de que los negros son responsables de la mayoría de los asesinatos de los americanos blancos, defendió a sus seguidores para el desbaste hasta un Negro Vidas Matters manifestante en uno de sus mítines y dudado en repudiar el apoyo de un ex líder del Ku Klux Klan. El mismo racismo oculto que duele Barack Obama ayudó a Donald Trump.

Al principio de las primarias, Nate Silver famoso afirmó que no había prácticamente ninguna posibilidad de que Trump ganaría. A medida que progresaban las primarias y se hizo cada vez más claro que Trump tenía un amplio apoyo, Plata decidió mirar los datos para ver si podía entender lo que estaba pasando. ¿Cómo podría posiblemente ser Trump haciendo tan bien?

La plata se dio cuenta de que las zonas donde Trump realiza mejor hechas para un mapa extraño. Trump se desempeñó bien en partes del noreste y el medio oeste industrial, así como la del Sur. Se lleva a cabo sobre todo en el Oeste peor. Plata buscó variables para tratar de explicar este mapa. ¿Era el desempleo? Fue la religión? Fue propiedad de armas? Fue tasas de inmigración? Fue oposición a Obama?

Plata encontró que el factor que mejor se correlaciona con el apoyo de Donald Trump en las primarias republicanas era que la medida había descubierto cuatro años antes. Las áreas que apoyaron Trump en los números más grandes fueron las que hizo la mayoría de las búsquedas de Google para el “negro”.



He pasado casi todos los días de los últimos cuatro años el análisis de datos de Google. Esto incluyó un período como un científico de datos a Google, que me contrató después de enterarse de mi investigación racismo. Y sigo para explorar estos datos como un escritor y periodista opinión de datos para el New York Times. Las revelaciones han mantenido que viene. Enfermedad mental; la sexualidad humana; abuso infantil; aborto; publicidad; religión; salud. No es exactamente temas pequeños, y este conjunto de datos, que no existía hace un par de décadas, ofrecen nuevas perspectivas sorprendentes en todos ellos. Economistas y otros científicos sociales siempre están buscando nuevas fuentes de datos, así que permítanme ser franco: Ahora estoy convencido de que las búsquedas de Google son el conjunto de datos más importante que se haya recogido en la psique humana.

Este conjunto de datos, sin embargo, no es la única herramienta de internet ha entregado para la comprensión de nuestro mundo. Pronto me di cuenta de que hay otras minas de oro digital. He descargado todas Wikipedia, pored través de los perfiles de Facebook, y se rasparon Stormfront. Además, Pornhub, uno de los sitios pornográficos más grandes de Internet, me dio sus datos completos sobre las búsquedas, visualización de vídeo de personas anónimas de todo el mundo. En otras palabras, he tomado una inmersión profunda en lo que ahora se llama Big Data. Además, he entrevistado a docenas de otros académicos, periodistas de datos, y los empresarios-que también están explorando estos nuevos reinos. Muchos de sus estudios se discutirán aquí.

Pero en primer lugar, una confesión: no voy a dar una definición precisa de lo que es Big Data. ¿Por qué? Debido a que es un concepto inherentemente impreciso. ¿Qué tan grande es grande? Son 18,462 observaciones de datos pequeños y 18,463 observaciones Big Data? Yo prefiero

tener una visión inclusiva de lo que califica: mientras que la mayoría de los datos que jugar con es a través de Internet, voy a discutir otras fuentes, también. Estamos viviendo una explosión en la cantidad y la calidad de todo tipo de información disponible. Gran parte de la nueva información fluye de Google y las redes sociales. Parte de ella es un producto de la digitalización de la información que antes estaba oculto en armarios y archivos. Parte de ella es por el aumento de los recursos dedicados a la investigación de mercado.

Entonces, ¿por qué es tan poderoso Big Data? Piense en toda la información que se encuentra dispersa en línea en un determinado día tenemos un número, de hecho, sólo por la cantidad de información que hay. En un día normal en la primera parte del siglo XXI, los seres humanos generan 2,5 millones de billones de bytes de datos.

Y estos bytes son pistas.

*Una mujer se aburre en un jueves por la tarde. Ella Googles para algunos más limpios “chistes divertidos.” Ella revisa su correo electrónico. Ella firma en Twitter. Ella Googles “chistes de negros.”*

*Un hombre está sintiendo azul. Él Googles para los “síntomas de la depresión” y “historias de depresión.” Juega un juego de solitario.*

*Una mujer ve el anuncio de su amiga conseguir enganchado en Facebook. La mujer, que es única, se hace con el amigo.*

*Un hombre toma un descanso de buscar en Google acerca de la NFL y la música rap a hacer una pregunta al motor de búsqueda: “¿Es normal tener sueños acerca de los hombres que se besan?”*

*Una mujer hace clic en una historia BuzzFeed mostrando los “15 gatos más lindos.”*

*Un hombre ve la misma historia sobre gatos. Pero en su pantalla se llama “15 gatos más adorables.” Él no hace clic.*

*Una mujer de Google “¿Es mi hijo un genio?”*

*Un hombre Googles “cómo conseguir a mi hija a perder peso.”*

*Una mujer está en unas vacaciones con sus seis mejores amigas. Todos sus amigos siguen diciendo lo bien que lo están pasando. Ella se escapa a Google “soledad cuando está lejos de su marido.”*

*Un hombre, el marido de la mujer anterior, está en unas vacaciones con sus seis mejores amigos varones. Él se escapa a Google para escribir “señales de que su mujer lo engaña.”*

Algunos de estos datos incluirán información que de otro modo no se admitió a nadie. Si agregamos todo, mantenerlo en el anonimato para asegurarse de que nunca se sabe acerca de los miedos, deseos y comportamientos de los individuos específicos, y añadir un poco de la ciencia de datos, comenzamos a conseguir una nueva mirada a los seres humanos comportamientos, sus deseos, sus naturalezas. De hecho, a riesgo de sonar grandioso, he llegado a creer que los nuevos datos cada vez más disponibles en nuestra era digital se expandirán radicalmente nuestra comprensión de la humanidad. El microscopio nos mostró que hay más de una gota de agua del estanque de lo que pensamos que vemos. El telescopio nos mostró que hay más hacia el cielo nocturno de lo que pensamos que vemos. Y los nuevos datos digitales, ahora nos muestra que

hay más a la sociedad humana de lo que pensamos que vemos. Puede ser de nuestro microscopio o telescopio posible importante de decisiones era,

Hay otro riesgo al hacer tales declaraciones, no sólo suena grandioso, sino también de moda. Muchas personas han estado haciendo grandes demandas sobre el poder de grandes volúmenes de datos. Sin embargo, han sido cortos en la evidencia.

Esto ha inspirado a los escépticos grandes volúmenes de datos, de los cuales hay también muchos, para descartar la búsqueda de bases de datos más grandes. “No estoy diciendo que no hay información de Big Data”, ensayista y estadístico Nassim Taleb ha escrito. “Hay un montón de información. El problema central de la cuestión es que la aguja entra en un pajar cada vez más grande.”

Uno de los objetivos principales de este libro, por tanto, es proporcionar la prueba que falta de lo que puede hacerse con grandes volúmenes de datos, cómo podemos encontrar las agujas, si se quiere, en esos pajares más y más grandes. Espero proporcionar suficientes ejemplos de grandes volúmenes de datos que ofrece nuevos conocimientos sobre la psicología y el comportamiento humano para que usted comenzará a ver los contornos de algo verdaderamente revolucionario.

“Espera, Seth,” se le puede decir ahora mismo. “Usted está prometiendo una revolución. Estás encerrar poético sobre estos grandes, nuevos conjuntos de datos. Sin embargo, hasta el momento, se ha usado todo este sorprendente, notables, impresionantes, información innovadora que decirme básicamente dos cosas: hay un montón de racistas en los Estados Unidos, y la gente, especialmente los hombres, exagerar la cantidad de sexo que tienen “.

Admito veces los nuevos datos se limita a confirmar lo obvio. Si cree que estos hallazgos eran obvias, espere hasta que llegue al capítulo 4, donde os muestro clara, la evidencia irrefutable de las búsquedas de Google que los hombres tienen una enorme preocupación e inseguridad en todo-esperar a que su tamaño-pene.

Hay, nos atreveríamos a afirmar algún valor en probar cosas es probable que haya tenido sospechosos, pero por lo demás poca evidencia de. Ante la sospecha de que algo es una cosa. Demostrando que es otro. Pero si todos los grandes datos podría hacer es confirmar sus sospechas, no sería revolucionario. Afortunadamente, Big Data puede hacer mucho más que eso. Una y otra vez, los datos me muestra el mundo funciona exactamente de la manera opuesta como lo hubiera adivinado. He aquí algunos ejemplos que puede encontrar más sorprendente.

Se podría pensar que una de las principales causas del racismo es la inseguridad económica y la vulnerabilidad. Es posible, naturalmente, sospechar, entonces, que cuando las personas pierden sus puestos de trabajo, aumenta el racismo. Pero, en realidad, ni búsquedas racistas ni la pertenencia a Stormfront se eleva cuando el desempleo hace.

Se podría pensar que la ansiedad es más alta en las grandes ciudades sobreeducados. El neurótico urbana es un estereotipo famosa. Sin embargo, las búsquedas de Google que reflejan la ansiedad, tales como “síntomas de ansiedad” o -tend “ansiedad ayuda” a ser mayor en lugares con menores niveles de educación más bajos, ingresos medios, y donde una porción mayor de la población vive en zonas rurales. Hay mayores tasas de búsqueda para la ansiedad en las zonas rurales, el estado de Nueva York que en la ciudad de Nueva York.

Se podría pensar que un ataque terrorista que mata a decenas o cientos de personas de forma automática sería seguido por masiva, la ansiedad generalizada. El terrorismo, por definición, se supone que inculcar un sentido de terror. Miré a las búsquedas de Google que reflejan la ansiedad. Probé la cantidad de estas búsquedas se elevaron en un país en los días, semanas y meses después de cada ataque terrorista europeo o americano desde 2004. Así, en promedio, ¿cuánto se elevan las búsquedas relacionadas con la ansiedad? Ellos no lo hicieron. En absoluto.

Se podría pensar que la gente busca chistes más a menudo cuando están tristes. Muchos de los más grandes pensadores de la historia han afirmado que volvemos al humor como una liberación de dolor. Humor mucho tiempo se ha pensado como una manera de hacer frente a las frustraciones, el dolor, las decepciones inevitables de la vida. Como Charlie Chaplin dijo, “La risa es la tónica, el relieve, la tregua al dolor.”

Sin embargo, las búsquedas de chistes son más baratos los lunes, el día en que las personas informan que son más infeliz. Ellos son los más bajos en los días nublados y lluviosos. Y ellos se desploman después de una tragedia mayor, como cuando dos bombas mataron a tres heridos y cientos durante el Maratón de Boston 2013. Las personas son en realidad más propensos a buscar chistes cuando las cosas van bien en la vida que cuando no lo son.

A veces, un nuevo conjunto de datos revela un comportamiento, el deseo, o la preocupación de que nunca habría siquiera considerado. Existen numerosas tendencias sexuales que entran en esta categoría. Por ejemplo, ¿sabía usted que en la India el número uno de búsqueda que comienza con “mi marido quiere. . .”Es ‘mi marido me quiere amamantar a él’? Este comentario es mucho más común en la India que en otros países. Por otra parte, las búsquedas de porno para representaciones de las mujeres en periodo de lactancia hombres son cuatro veces más alta en la India y Bangladesh que en cualquier otro país del mundo. Desde luego, nunca hubiera sospechado que antes de ver los datos.

Además, si bien el hecho de que los hombres están obsesionados con su tamaño del pene puede no ser demasiado sorprendente, la inseguridad corporal mayor para las mujeres, tal como se expresa en Google, es realmente sorprendente. Sobre la base de estos nuevos datos, el equivalente femenino de preocuparse por el tamaño de su pene puede-hacer una pausa para crear suspenso-preocuparse de si su vagina huele. Las mujeres representan casi tantas búsquedas que expresan preocupación por sus genitales como los hombres no preocuparse por ellos. Y la principal preocupación mujeres expresan es su olor y cómo podrían mejorarlo. Desde luego, no sabía que antes de ver los datos.

A veces los nuevos datos revelan diferencias culturales que nunca había contemplado. Un ejemplo: las muy diferentes formas en que los hombres de todo el mundo responden a sus esposas estar embarazada. En México, las principales búsquedas sobre “mi mujer embarazada” incluyen “Frases de Amor Para Mi esposa Embarazada” (palabras de amor a mi mujer embarazada) y “Poemas para Mi esposa Embarazada” (poemas para mi mujer embarazada). En Estados Unidos, las búsquedas más habituales incluyen “mi mujer está embarazada y ahora qué” y “mi mujer está embarazada ¿qué hago.”

Pero este libro es más que una colección de hechos extraños o estudios de una sola vez, aunque habrá un montón de ellos. Debido a que estas metodologías son tan nuevas y sólo se van a poner más potente, yo estaba a cabo algunas ideas sobre cómo funcionan y lo que les hace innovador. También voy a reconocer las limitaciones de grandes volúmenes de datos.

Algunos de entusiasmo por el potencial de la revolución datos se han perdido. La mayoría de los enamorados de los grandes datos fluyen sobre cómo inmensa estos conjuntos de datos se pueden obtener. Esta obsesión por el tamaño de conjunto de datos no es nueva. Antes de Google, Amazon y Facebook, antes de que existiera la frase “Big Data”, se celebró una conferencia en Dallas, Texas, en “conjuntos de datos grandes y complejos.” Jerry Friedman, un profesor de estadística en Stanford que era un colega mío cuando yo trabajaba en Google, recuerda que la conferencia 1977. Un distinguido estadístico se levantaba para hablar. Se explicaría que había acumulado una increíble, sorprendentes cinco gigabytes de datos. La próxima distinguido estadístico se levantaba para hablar. Él comenzaría, “El último orador tenía gigabytes. Eso no es nada. Tengo terabytes.” El énfasis de la charla, en otras palabras, era la cantidad de información que podría acumularse, no lo que esperaba hacer con él, o qué preguntas responder planeado. “Me pareció divertido, en el momento”, dice Friedman, que “lo que se suponía que era impresionado con lo grande que es el conjunto de datos. Todavía sucede “.

Demasiado muchos científicos hoy en día se están acumulando datos conjuntos masivos de datos y nos dicen muy poco de importancia, por ejemplo, que los Knicks son populares en Nueva York. Muchas empresas se están ahogando en datos. Tienen un montón de terabytes, pero pocas ideas principales. El tamaño de un conjunto de datos, creo, es con frecuencia sobrevalorado. Hay una sutil pero importante explicación para esto. Cuanto más grande es un efecto, menor será el número de observaciones necesarias para verlo. Sólo tiene que tocar una estufa caliente una vez para darse cuenta de que es peligroso. Es posible que tenga que beber café miles de veces para determinar si tiende a darle un dolor de cabeza. ¿Qué lección es más importante? Claramente, la estufa caliente, el cual, debido a la intensidad de su impacto, se presenta con tanta rapidez, con tan pocos datos.

De hecho, las empresas más inteligentes grandes volúmenes de datos a menudo son la tala de sus datos. En Google, las decisiones importantes se basan en sólo una pequeña muestra de todos sus datos. No siempre se necesita un montón de datos para encontrar información importante. Necesita los datos correctos. Una de las principales razones de que las búsquedas de Google son tan valioso no es que hay tantos de ellos; es que la gente es tan honesto en ellos. La gente miente a los amigos, amantes, médicos, encuestas, y de ellos mismos. Pero en Google podrían compartir información embarazosa, sobre, entre otras cosas, sus matrimonios sin sexo, sus problemas de salud mental, sus inseguridades, y su animosidad hacia los negros.

Lo más importante, para expresar ideas de grandes volúmenes de datos, lo que tiene que hacer las preguntas correctas. Así como no se puede apuntar un telescopio al azar en el cielo de la noche y tienen que descubrir Plutón para usted, usted no puede descargar un montón de datos y tienen que descubrir los secretos de la naturaleza humana para usted. Hay que mirar al prometer búsquedas lugares-Google que comenzar “mi marido quiere. . .” En la India, por ejemplo.

Este libro va a mostrar cómo grandes volúmenes de datos que se utiliza mejor y explicar en detalle por qué puede ser tan poderoso. Y en el camino, también aprenderá acerca de lo que yo y otros ya han descubierto con él, incluyendo:

- > ¿Cuántos hombres son homosexuales?
- > ¿Funciona la publicidad?
- > ¿Por qué era estadounidense Pharoah un gran caballo de carreras?
- > Está sesgado los medios de comunicación?
- > Son acto fallido real?
- > ¿Quién engaña en sus impuestos?
- > ¿Importa donde vaya a la universidad?
- > ¿Se puede vencer el mercado de valores?
- > ¿Cuál es el mejor lugar para criar a los niños?
- > ¿Qué hace que una historia sea viral?
- > ¿Qué debe hablar en una primera cita si quieres un segundo?

... y mucho, mucho más.

Pero antes de llegar a todo eso, tenemos que discutir una pregunta más básica: ¿por qué necesitamos los datos en absoluto? Y para ello, voy a presentar a mi abuela.

# PARTE I

---

## DATOS, grandes y pequeños



## Su intestino FAULTY

**Y**o f estás treinta y tres años de edad y ha asistido a un par de Gracias en una fila sin una fecha, es probable que surja el tema de la elección de pareja. Y casi todo el mundo va a tener una opinión.

“Seth necesita una chica loca, como él,” dice mi hermana.

“¡Estás loco! Se necesita una chica normal, para equilibrar sacarlo,” dice mi hermano.

“Seth no es una locura”, dice mi madre.

“¡Estás loco! Por supuesto, Seth es una locura “, dice mi padre.

De repente, mi tímida abuela, de voz suave y tranquila a través de la cena, habla. Los agresivos, voces Nueva York van en silencio, y todos los ojos se centran en la pequeña anciana con el pelo corto de color amarillo y sigue siendo un rastro de acento de Europa del Este. “Seth, se necesita una buena chica. No es muy bonito. Muy inteligente. Bueno con las personas. Social, por lo que va a hacer las cosas. Sentido del humor, porque tiene un buen sentido del humor “.

¿Por qué el comando de los consejos de esta anciana tal atención y respeto en mi familia? Bueno, mi ochenta y ocho años de edad, abuela ha visto más que todos los demás en la mesa. Ella observó más matrimonios, muchos que trabajaron y muchos que no lo hicieron. Y a lo largo de las décadas, se ha catalogado las cualidades que hacen de las relaciones exitosas. En esa mesa de Acción de Gracias, por esa pregunta, mi abuela tiene acceso al mayor número de puntos de datos. Mi abuela es Big Data.

En este libro, quiero desmitificar la ciencia de datos. Nos guste o no, los datos está desempeñando un papel cada vez más importante en la vida de todos-y su papel va a ser más grande. Periódicos ahora tienen secciones completas dedicadas a los datos. Las empresas tienen equipos con la exclusiva tarea de analizar sus datos. Los inversores dan arranques decenas de millones de dólares si se pueden almacenar más datos. Incluso si nunca aprende cómo ejecutar una regresión o calcular un intervalo de confianza, que se va a encontrar una gran cantidad de datos en las páginas que lee, las reuniones de trabajo que asistan, los rumores se oye al lado de los enfriadores de agua que beber.

Muchas personas están ansiosos sobre este desarrollo. Ellos se sienten intimidados por los datos, fácilmente perdido y confundido en un mundo de los números. Ellos piensan que una comprensión cuantitativa del mundo es para un grupo selecto de prodigios-izquierda brained, no

para ellos. Tan pronto como se encuentran con los números, que están dispuestos a pasar la página, poner fin a la reunión, o cambiar la conversación.

Pero he pasado diez años en el negocio de análisis de datos y he tenido la suerte de trabajar con muchos de los mejores personas en el campo. Y una de las lecciones más importantes que he aprendido es la siguiente: La buena ciencia de datos es menos complicado que la gente piensa. La mejor ciencia de datos, de hecho, es sorprendentemente intuitivo.

Lo que hace que la ciencia intuitiva de los datos? En su esencia, la ciencia de datos se trata de la detección de patrones y predecir cómo una variable afectará a otra. La gente hace esto todo el tiempo.

Piensa en lo que mi abuela me dio consejos sobre relaciones. Se utilizó la base de datos grande de relaciones que su cerebro ha subido más de un siglo, cerca de la vida en las historias que ha escuchado de su familia, sus amigos, sus conocidos. Ella limita su análisis a una muestra de relaciones en las que el hombre tenía muchas cualidades que tengo, un temperamento sensible, una tendencia a aislarse, un sentido del humor. Ella se concentró en las cualidades esenciales de la clase mujer de cómo estaba, lo inteligente que era, lo bonita que era. Ella correlaciona estas cualidades fundamentales de la mujer con una cualidad clave de la relación, ya sea que era una buena. Por último, se informó de sus resultados. En otras palabras, vio cómo los patrones y predijo una variable afectará a otra. La abuela es un científico de datos.

Usted es un científico de datos, también. Cuando era un niño, Ha notado que cuando lloraste, tu madre te dio la atención. Esa es la ciencia de datos. Cuando alcanzado la edad adulta, se ha notado que si se quejan demasiado, la gente quiere salir contigo menos. Esa es la ciencia de datos, también. Cuando la gente pasar el rato contigo menos, se ha notado, es menos feliz. Cuando esté menos feliz, es menos amigable. Cuando esté menos amigable, la gente quiere salir contigo incluso menos. Ciencia de los datos. Ciencia de los datos. Ciencia de los datos.

Porque la ciencia de datos es tan natural, los mejores estudios de grandes volúmenes de datos, que he encontrado, pueden ser entendidos por casi cualquier persona inteligente. Si no se puede entender un estudio, el problema es probablemente con el estudio, no con usted.

Quiere una prueba de que la ciencia grandes de datos tiende a ser intuitivo? Recientemente me encontré con un estudio que puede ser uno de los más importantes realizados en los últimos años. También es uno de los estudios más intuitivo que he visto nunca. Quiero que piensen no sólo acerca de la importancia del estudio, pero la forma natural y la abuela como es.

El estudio fue por un equipo de investigadores de la Universidad de Columbia y Microsoft. El equipo quería saber qué síntomas predecir el cáncer de páncreas. Esta enfermedad tiene una tasa de sólo cinco años de supervivencia baja un 3 por ciento, pero la detección temprana puede duplicar las posibilidades de un paciente.

método de los investigadores? Utilizaron datos de decenas de miles de usuarios anónimos de Bing, el motor de búsqueda de Microsoft. Que codifican un usuario como si hubiera sido recientemente dado un diagnóstico de cáncer de páncreas en base a búsquedas inconfundibles, como “acaban de diagnosticar un cáncer de páncreas” o “Me dijeron que tengo cáncer de páncreas, lo que puede esperar.”

A continuación, los investigadores analizaron las búsquedas de los síntomas de salud. Compararon ese pequeño número de usuarios que más tarde se informó de un diagnóstico de cáncer de páncreas con aquellos que no lo hicieron. ¿Qué síntomas, en otras palabras, predijo que, en unas pocas semanas o meses, un usuario se le informa que un diagnóstico?

Los resultados fueron sorprendentes. En busca de dolor de espalda y la piel amarillenta y luego resultó ser un signo de cáncer de páncreas; en busca de dolor que acaba de regresar solo hecho que sea poco probable que alguien tenía cáncer de páncreas. Del mismo modo, la búsqueda de la indigestión y el dolor abdominal fue entonces evidencia de cáncer de páncreas, mientras que la búsqueda por sólo la indigestión y sin dolor abdominal significaba una persona era poco probable que tenerlo. Los investigadores pudieron identificar 5 a 15 por ciento de los casos con casi ningún falso positivo. Ahora, esto no puede sonar como un gran precio; pero si usted tiene cáncer de páncreas, incluso un 10 por ciento de probabilidades de una posible duplicación sus posibilidades de supervivencia se sentiría como un golpe de suerte.

El artículo que detalla este estudio sería difícil para los no expertos para hacer plenamente sentido de. Incluye un montón de jerga técnica, tales como la prueba de Kolmogorov-Smirnov, el significado de eso, tengo que admitir, me había olvidado. (Es una manera de determinar si un modelo se ajusta correctamente los datos.)

Sin embargo, tenga en cuenta la forma natural e intuitiva de este notable estudio está en su nivel más fundamental. Los investigadores observaron a una amplia gama de casos médicos y trataron de conectar los síntomas de una enfermedad en particular. Usted sabe quién más utiliza esta metodología para tratar de averiguar si una persona tiene una enfermedad? Los esposos y las esposas, las madres y los padres, y las enfermeras y los médicos. Basado en la experiencia y el conocimiento, que intentan conectarse fiebres, dolores de cabeza, escurrimiento nasal, dolores de estómago y de diversas enfermedades. En otras palabras, los investigadores de Columbia y Microsoft escribieron un estudio pionero mediante la utilización de la metodología natural, obvio que todo el mundo usa para hacer diagnósticos de salud.

Pero espera. Vamos a reducir la velocidad aquí. Si la metodología de la mejor ciencia de datos es con frecuencia natural e intuitiva, como reclamo, esto plantea una pregunta fundamental sobre el valor de grandes volúmenes de datos. Si los seres humanos son naturalmente científicos de datos, si la ciencia de datos es intuitiva, ¿por qué necesitamos las computadoras y el software estadístico? ¿Por qué necesitamos la prueba de Kolmogorov-Smirnov? No podemos utilizar nuestro instinto? ¿No podemos hacerlo como abuela ya, al igual que las enfermeras y los médicos hacen?

Esto nos lleva a una discusión intensificado después del lanzamiento del libro más vendido de Malcolm Gladwell Blink, que exalta la magia de instintos de las personas. Gladwell cuenta las historias de personas que, basándose únicamente en sus entrañas, puede decir si una estatua es falso; si un jugador de tenis fallará antes de que golpea la bola; la cantidad de un cliente está dispuesto a pagar. Los héroes de parpadeo no se ejecutan las regresiones; no calculan los intervalos de confianza; que no se ejecutan las pruebas de Kolmogorov-Smirnov. Pero en general, hacer predicciones notables. Muchas personas han apoyado de manera intuitiva defensa

de la intuición de Gladwell: confían en sus entrañas y sentimientos. Los fans de Blink podrían celebrar la sabiduría de mi abuela dar consejos sobre relaciones sin la ayuda de ordenadores. Los aficionados de parpadeo pueden ser menos propensos a celebrar mis estudios o los otros estudios descritos en este libro, las cuales utilizan las computadoras. Si Big Data-del tipo de ordenador, en lugar de los de tipo abuela es una revolución, tiene que demostrar que es más poderoso que nuestra intuición sin ayuda, que, como Gladwell ha señalado, a menudo puede ser notable.

El estudio de Columbia y Microsoft ofrece un claro ejemplo de la ciencia rigurosa la información y ordenadores que nos enseñan cosas nuestro intestino por sí sola nunca pudo encontrar. Este es también un caso en que el tamaño de los asuntos de conjunto de datos. A veces no existe experiencia suficiente para nuestro intestino sin ayuda para aprovechar. Es poco probable que usted o sus amigos cercanos o miembros de la familia, ha visto suficientes casos de cáncer de páncreas para desentrañar la diferencia entre la indigestión seguido por dolor abdominal en comparación con la indigestión solo. De hecho, es inevitable, ya que el conjunto de datos de Bing se hace más grande, que los investigadores van a recoger los patrones muchos más sutiles en la sincronización de los síntomas, por esta y otras enfermedades-que incluso los médicos podrían perderse.

Por otra parte, mientras que nuestro intestino por lo general nos puede dar una buena sensación general de cómo funciona el mundo, con frecuencia no es preciso. Necesitamos datos para agudizar la imagen. Considere, por ejemplo, los efectos del clima en el estado de ánimo. Es probable que adivinar que la gente es más probable que se sienta más sombrío en un día de 10 grados que en un día de 70 grados. De hecho, esto es correcto. Sin embargo, es posible que no adivinar qué tan grande un impacto esta diferencia de temperatura puede hacer. Busqué las correlaciones entre las búsquedas de Google de un área para la depresión y una amplia gama de factores, incluyendo las condiciones económicas, niveles de educación y asistencia a la iglesia. Invierno el clima inundó todo el resto. En los meses de invierno, los climas cálidos, como el de Honolulu, Hawaii, tienen 40 por ciento menos de búsquedas de depresión que los climas fríos, como el de Chicago, Illinois. ¿Qué tan importante es este efecto? Una lectura optimista de la eficacia de los antidepresivos se encontraba con que los fármacos más eficaces disminuyen la incidencia de la depresión por sólo un 20 por ciento. A juzgar por los números de Google, un movimiento de Chicago a Honolulu sería por lo menos dos veces tan eficaz como la medicación para su depresión invernal.\*

A veces nuestro intestino, cuando no esté guiada por análisis informático cuidado, puede ser totalmente equivocado. Podemos quedar cegados por nuestras propias experiencias y prejuicios. De hecho, a pesar de que mi abuela es capaz de utilizar sus décadas de experiencia para dar un mejor consejo de la relación que el resto de mi familia, que todavía tiene algunos puntos de vista dudosos sobre lo que hace una relación pasada. Por ejemplo, se ha enfatizado con frecuencia para mí la importancia de tener amigos comunes. Ella cree que este fue un factor clave en el éxito de su matrimonio: se pasaba las noches más cálidas con su marido, mi abuelo, en su pequeño patio trasero en Queens, Nueva York, sentados en sillas de jardín y cotilleando con su apretado grupo de vecinos.

Sin embargo, con el riesgo de lanzar mi propia abuela debajo del autobús, la ciencia datos sugieren que la teoría de la abuela está mal. Un equipo de científicos de la computación ha analizado recientemente el mayor conjunto de datos se haya reunido jamás en las relaciones humanas-Facebook. Se miraron un gran número de parejas que eran, en algún momento, “en una relación”. Algunas de estas parejas se quedaron “en una relación”. Otros cambian su estado a “single”. Tener un núcleo común de los amigos, las los investigadores encontraron, es un fuerte predictor de que una relación no durará. Tal vez salir todas las noches con su pareja y el mismo pequeño grupo de personas no es una buena cosa; círculos sociales separadas pueden ayudar a hacer las relaciones más fuertes.

Como se puede ver, nuestra intuición solo, cuando nos alejamos de los ordenadores y vamos con nuestro intestino, a veces puede sorprender. Pero puede cometer grandes errores. La abuela puede haber caído en una trampa cognitiva: se tiende a exagerar la importancia de nuestra propia experiencia. En la jerga de los científicos de datos, se pondera nuestros datos, y damos demasiada importancia a un punto de datos en particular: a nosotros mismos.

La abuela estaba tan concentrado en su noche schmoozes con el abuelo y sus amigos que ella no pensó lo suficiente sobre otras parejas. Ella se olvidó de tener plenamente en cuenta su hermano-en-ley y su esposa, que chitchatted mayoría de las noches con un grupo pequeño, pero consistente de amigos lucharon con frecuencia y se divorció. Ella se olvidó de tener plenamente en cuenta mis padres, su hija y su hijo-en-ley. Mis padres van por caminos separados muchas noches- mi padre a un club de jazz o juego de pelota con sus amigos, a mi madre a un restaurante o al teatro con sus amigos; sin embargo, siguen siendo felizmente casado.

Cuando se basa en nuestro intestino, sino que también puede ser lanzado fuera por la fascinación humana básica con el dramático. Tenemos la tendencia a sobreestimar la prevalencia de cualquier cosa que hace para una historia memorable. Por ejemplo, cuando se le preguntó en una encuesta, la gente constantemente rango tornados como causa más común de muerte que el asma. De hecho, el asma causa cerca de setenta veces más muertes. Las muertes por asma no destacan y no hacen las noticias. Las muertes por tornados hacen.

Estamos a menudo mal, en otras palabras, sobre cómo funciona el mundo cuando nos basamos sólo en lo que oímos o personalmente experiencia. Si bien la metodología de la buena ciencia de datos es a menudo intuitiva, los resultados son con frecuencia contrario a la intuición. la ciencia de datos tiene un sentido proceso de detección de patrones y hacer humano natural e intuitiva de ellos, y lo inyecta con esteroides, lo que potencialmente nos muestra que el mundo funciona de una manera completamente diferente a la forma en que pensamos que lo hizo. Eso es lo que ocurrió cuando estudiaba los predictores de éxito baloncesto.

Cuando era un niño pequeño, tenía un sueño y sólo un sueño: quería llegar a ser un científico economista y datos. Solo bromeo. Quería desesperadamente ser un jugador de baloncesto profesional, para seguir los pasos de mi héroe, Patrick Ewing, todas las estrellas centro de los Knicks de Nueva York.

A veces sospecho que dentro de cada científico de datos es un niño tratando de averiguar por qué sus sueños de la infancia no se hizo realidad. Por lo tanto, no es sorprendente que recientemente he investigado lo que se necesita para hacer la NBA. Los resultados de la investigación fueron sorprendentes. De hecho, demuestran una vez más la buena ciencia de datos puede cambiar su visión del mundo, y cómo contrario a la intuición los números pueden ser.

La pregunta en particular Miré es la siguiente: ¿Es más probable para que sea en la NBA Si creces pobres o de clase media?

La mayoría de la gente se adivina la primera. La sabiduría convencional dice que crecen en circunstancias difíciles, tal vez en los proyectos con una sola mamá, adolescente, ayuda a fomentar la unidad necesaria para alcanzar los niveles más altos de este deporte altamente competitivo.

Esta opinión fue expresada por William Ellerbe, un entrenador de baloncesto de la escuela en Filadelfia, en una entrevista con Sports Illustrated. “Suburban niños tienden a jugar por el gusto de hacerlo,” dijo Ellerbe. “Niños Inner-City mira el baloncesto como una cuestión de vida o muerte.” Yo, por desgracia, criado por padres casados en los suburbios de Nueva Jersey. LeBron James, el mejor jugador de mi generación, nació pobre a un adolescente de dieciséis años de edad, madre soltera en Akron, Ohio.

De hecho, una encuesta en Internet he realizado sugiere que la mayoría de los estadounidenses piensa lo mismo entrenador Ellerbe y pensé: que la mayoría de jugadores de la NBA crecen en la pobreza.

¿Es correcta esta sabiduría convencional?

Veamos los datos. No hay una fuente de datos completa sobre los aspectos socioeconómicos de jugadores de la NBA. Pero por ser detectives de datos, mediante la utilización de los datos de un montón de [sources-basketball-reference.com](http://sources-basketball-reference.com), [ancestry.com](http://ancestry.com), el censo de los EEUU, y otros, podemos averiguar lo que los antecedentes familiares es en realidad más propicia para hacer la NBA. Este estudio, se dará cuenta, utiliza una variedad de fuentes de datos, algunos de ellos más grandes, algunos de ellos menores, algunos de ellos en línea, y algunos de ellos fuera de línea. Tan emocionante como algunas de las nuevas fuentes digitales son, un buen científico de datos no está por encima de la consulta de fuentes anticuadas si pueden ayudar. La mejor manera de obtener la respuesta correcta a una pregunta es combinar todos los datos disponibles.

El primer dato relevante es el lugar de nacimiento de cada jugador. Por cada condado en los Estados Unidos, he grabado como nacieron muchos hombres negros y blancos en la década de 1980. entonces grabé cuántos de ellos llegaron a la NBA. He comparado este con el ingreso medio de la casa de un condado. También controlé para la demografía racial de un condado, ya que -y esto es un tema para un conjunto de otros hombres-libro negro son aproximadamente cuarenta veces más propensas que los hombres blancos para llegar a la NBA.

Los datos nos dice que un hombre tiene un sustancialmente mejor oportunidad de llegar a la NBA si él nació en un condado rico. Un niño negro nacido en uno de los condados más ricos de los Estados Unidos, por ejemplo, es más del doble de probabilidades de que la NBA que un niño negro nacido en uno de los condados más pobres. Para un niño blanco, la ventaja de haber nacido

en uno de los condados más ricos en comparación con haber nacido en uno de los más pobres es del 60 por ciento.

Esto sugiere, en contra de la sabiduría convencional, que los pobres son en realidad poco representados en la NBA. Sin embargo, estos datos no es perfecta, ya que muchos condados ricos en los Estados Unidos, tales como el Condado de Nueva York (Manhattan), también incluyen los barrios pobres, como Harlem. Por lo tanto, todavía es posible que una infancia difícil le ayude a tomar la NBA. Todavía necesitamos más pistas, más datos.

Así que he investigado los antecedentes de los jugadores de la NBA de la familia. Esta información se encuentra en las noticias y en las redes sociales. Esta metodología fue bastante lenta, por lo que limita el análisis a los jugadores cien afroamericanos NBA nacidos en la década de 1980 que obtuvieron el mayor número de puntos. En comparación con el hombre negro promedio en los Estados Unidos, superestrellas de la NBA fueron un 30 por ciento menos probabilidades de haber nacido de una madre adolescente o una madre soltera. En otras palabras, los antecedentes de los mejores jugadores de raza negra de la familia de la NBA también sugieren que un fondo cómodo es una gran ventaja para alcanzar el éxito.

Dicho esto, ni el nacimiento de datos a nivel de condado, ni los antecedentes familiares de una muestra limitada de jugadores da información perfecta sobre la infancia de todos los jugadores de la NBA. Así que todavía no estaba totalmente convencido de que los dos padres, las familias de clase media a producir más estrellas de la NBA que monoparentales, las familias pobres. Cuanto más datos que puede lanzar en esta pregunta, mejor.

Entonces recordé un punto más datos que pueden proporcionar pistas para decirle fondo de un hombre. Se sugirió en un artículo de dos economistas, Roland freidora y Steven Levitt, que el nombre de una persona negro es una indicación de su origen socioeconómico. Freidora y Levitt estudiaron los certificados de nacimiento en California en la década de 1980 y se encontró que, entre los afroamericanos, las madres pobres, sin educación, e individuales tienden a dar a sus hijos nombres diferentes de hacer la clase media, educados y padres casados.

Los niños de entornos favorecidos son más propensos a recibir nombres comunes, tales como Kevin, Chris y Juan. Los niños de hogares difíciles en los proyectos son más propensos a recibir nombres únicos, tales como Knowshon, Uneek, y Breionshay. niños afro-americanos nacidos en la pobreza son casi dos veces más propensos a tener un nombre que se le da a ningún otro niño nacido en el mismo año.

¿Qué pasa con los nombres de los jugadores negros de la NBA? ¿Suenan más como de clase media o los negros pobres? Mirando el mismo período de tiempo, los jugadores de la NBA nacido en California tenían la mitad de probabilidades de tener nombres únicos como el macho negro promedio, una diferencia estadísticamente significativa.

¿Conoce a alguien que piensa que la NBA es una liga para los niños del gueto? Dile que acaba de escuchar de cerca a la siguiente partida en la radio. Dile que se tenga en cuenta la frecuencia con Russell regatea a Dwight y luego trata de deslizar la bola más allá de los brazos extendidos de Josh y en manos de espera de Kevin. Si la NBA realmente fuera una liga llena de hombres negros pobres, sonaría muy diferente. Habría mucho más hombres con nombres como LeBron.

Ahora, nos hemos reunido tres piezas diferentes de pruebas: el condado de nacimiento, el estado civil de las madres de los máximos goleadores, y los primeros nombres de los jugadores. Ninguna fuente es perfecta. Pero los tres apoyan la misma historia. Mejor estatus socioeconómico significa una mayor probabilidad de hacer la NBA. La sabiduría convencional, en otras palabras, es errónea.

Entre todos los afroamericanos nacidos en la década de 1980, aproximadamente el 60 por ciento tenía padres no casados. Pero calculo que entre los afroamericanos nacidos en esa década que llegó a la NBA, una mayoría significativa tenía padres casados. En otras palabras, la NBA no se compone sobre todo de los hombres con antecedentes como el de LeBron James. Hay más hombres como Chris Bosh, criado por dos padres en Texas que cultivaban su interés en aparatos electrónicos, o Chris Paul, el segundo hijo de una familia de clase media en Lewisville, Carolina del Norte, cuya familia se unió a él en un episodio de Family Feud en 2011.

El objetivo de un científico de datos es entender el mundo. Una vez que encontramos el resultado contrario a la intuición, podemos utilizar la ciencia más datos para ayudar a explicar por qué el mundo no es lo que parece. ¿Por qué, por ejemplo, qué los hombres de clase media tienen una ventaja en el baloncesto en relación con los pobres? Hay por lo menos dos explicaciones.

En primer lugar, porque los pobres tienden a terminar más corto. Los estudiosos han sabido por mucho tiempo que el cuidado de la salud y la nutrición de la infancia juegan un papel importante en la salud del adulto. Es por esto que el hombre común en el mundo desarrollado es ahora cuatro pulgadas más alto que un siglo y medio atrás. Los datos sugieren que los estadounidenses de origen pobre, debido a la atención de la salud-vida temprana y la nutrición más débiles, son más cortos.

Los datos también nos puede decir el efecto de la altura al llegar a la NBA. Sin duda, intuyó que ser alto puede ser de ayuda a un aspirante a jugador de baloncesto. Sólo contrastar la altura del jugador de béisbol típico en la cancha para el ventilador típica en las gradas. (El promedio de los jugadores de la NBA es 6'7" ; el hombre americano promedio es de 5'9" ).

¿Cuánto importa la altura? jugadores de la NBA veces fib un poco acerca de su altura, y no hay ninguna lista de distribución completa de la altura de los hombres estadounidenses. Pero trabajar con una estimación matemática aproximada de lo que esta distribución podría ser similar y propios números de la NBA, es fácil de confirmar que los efectos de la altura son enormes, tal vez incluso más de lo que podríamos haber sospechado. Calculo que cada pulgada adicional duplica más o menos probabilidades de llegar a la NBA. Y esto es cierto en toda la distribución de la altura. A 5'11" el hombre tiene el doble de probabilidades de llegar a la NBA como 5'10" hombre. A 6'11" el hombre tiene el doble de probabilidades de llegar a la NBA como 6'10" hombre. Parece ser que, entre los hombres de menos de seis pies de alto, sólo uno de cada dos millones de llegar a la NBA. Entre los más de siete pies de altura, yo y otros hemos estimado,

Datos, tendrá en cuenta, aclara por qué se descarriló mi sueño de ser estrella de baloncesto. No fue porque fui criado en los suburbios. Fue porque soy 5'9" y blanco (sin mencionar lento).



Además, soy perezoso. Y tengo poco vigor, la forma de tiro horrible, y de vez en cuando un ataque de pánico cuando el balón se pone en la mano.

Una segunda razón por la que los niños procedentes de entornos difíciles pueden tener dificultades para hacer que la NBA es que a veces carecen de ciertas habilidades sociales. Utilizando los datos de miles de niños en edad escolar, los economistas han descubierto que la clase media, las familias con los dos padres son en promedio sustancialmente mejor a criar a los niños que están confiando, disciplinado, persistente, enfocados y organizados.

Entonces, ¿cómo lo hacen escasas habilidades sociales descarrilar una carrera en el baloncesto de otra manera prometedora?

Veamos la historia de Doug Wrenn, uno de los prospectos de baloncesto con más talento de la década de 1990. Su entrenador de la universidad, Jim Calhoun en la Universidad de Connecticut, que ha entrenado futuro del Juego de las estrellas, afirmó Wrenn saltó el más alto de cualquier hombre que jamás había trabajado. Pero Wrenn tuvo una educación desafiante. Fue criado por una madre soltera en Blood Alley, uno de los barrios más duros en Seattle. En Connecticut, se enfrentó constantemente con los que le rodean. Él se burlaba de los jugadores, entrenadores cuestionar, y usar ropa suelta, en violación de las reglas del equipo. También tuvo problemas legales, que robó los zapatos en una tienda y espetó a los agentes de policía. Calhoun finalmente tuvo suficiente y le dio una patada fuera del equipo.

Wrenn tuvo una segunda oportunidad en la Universidad de Washington. Pero hay, también, una incapacidad para llevarse bien con la gente que le descarrilar. Luchó con su entrenador durante el tiempo de juego y selección de tiro y se dio inicio a este equipo también. Wrenn se fue seleccionado en el draft por la NBA, rebotó ligas alrededor inferiores, a vivir con su madre, y fue finalmente encarcelado por asalto. “Mi carrera ha terminado”, dijo Wrenn el Seattle Times en 2009. “Mis sueños, mis aspiraciones son más. Doug Wrenn está muerto. Ese jugador de baloncesto, ese tipo está muerto. Se acabó.” Wrenn tenía el talento no sólo para ser un jugador de la NBA, pero para ser un grande, incluso un jugador de leyenda. Pero nunca se desarrolló el temperamento para mantenerse incluso en un equipo de la universidad. Tal vez si hubiera tenido una vida temprana estable, que podría haber sido el próximo Michael Jordan.

Michael Jordan, por supuesto, también tuvo un salto vertical impresionante. Además de un gran ego y la competitividad, una intensa personalidad a veces que no era muy diferente de Wrenn. Jordan podría ser un niño difícil. A la edad de doce años, fue expulsado de la escuela por luchar. Pero tenía al menos una cosa que le faltaba Wrenn: una, la crianza estable de clase media. Su padre era un supervisor de equipos para General Electric, su madre un banquero. Estos ayudaron a navegar por su carrera.

De hecho, la vida de Jordan está llena de historias de su familia guiándolo lejos de las trampas que un gran talento competitivo puede caer. Después de Jordan fue expulsado de la escuela, su madre respondió llevándolo con ella para trabajar. No se le permitió dejar el coche y en su lugar tenía que sentarse en los libros de lectura de estacionamiento. Después de que fue seleccionado por los Bulls de Chicago, sus padres y hermanos se turnaban para visitar para asegurarse de que evita las tentaciones que vienen con la fama y el dinero.

La carrera de Jordan no terminó como Wrenn de, con una cita de lectura poco en el Seattle Times. Se terminó con un discurso sobre la inducción al Salón de la Fama que fue visto por millones de personas. En su discurso, Jordan dijo que trató de mantenerse “centrado en las cosas buenas de la vida, ya sabes cómo la gente te percibe, cómo se los respeta. . . la forma en que son percibidos públicamente. Hacer una pausa y pensar en las cosas que haces. Y que todos ellos procedentes de mis padres “.

Los datos nos dicen Jordan tiene toda la razón para agradecer a su clase media, los padres casados. Los datos nos dicen que en las familias en peores condiciones, en las comunidades más desfavorables, hay talentos de la NBA de nivel que no están en la NBA. Estos hombres tenían los genes, tenía la ambición, pero nunca se desarrolló el temperamento para convertirse en grandes estrellas de baloncesto.

Y nadie lo que podríamos intuir el bienestar en circunstancias tan desesperada que el baloncesto parece “una cuestión de vida o muerte” no ayuda. Historias como la de Doug Wrenn pueden ayudar a ilustrar esto. Y los datos así lo demuestra.

En junio de 2013, LeBron James fue entrevistado en la televisión después de ganar su segundo campeonato de la NBA. (Ha ganado ya que un tercio.) “Estoy LeBron James,” anunció. “A partir de Akron, Ohio. Desde el interior de la ciudad. Ni siquiera se supone que debo estar aquí.” Twitter y otras redes sociales estallaron con la crítica. ¿Cómo podía una persona tan sumamente dotado, identificados a partir de una edad joven y absurdo como el futuro del baloncesto, pretensión de ser un perdedor? De hecho, cualquier persona de un entorno difícil, sin importar su destreza atlética, tiene las probabilidades en su contra. Los logros de James, en otras palabras, son aún más excepcional de lo que parecen ser a primera. Datos de prueba que, también.

## PARTE II

---

### LOS PODERES DE DATOS BIG

## FUE FREUD DERECHO?

**Y**O Recientemente vi una persona que camina por una calle se describe como un “penistrian.” Atrapaste eso, ¿verdad? Un “penistrian” en lugar de un “peatón”. Lo vi en un gran conjunto de datos de errores tipográficos que se cometen. Una persona ve a alguien caminando y escribe la palabra “pene”. Tiene que significar algo, ¿verdad?

Hace poco supe de un hombre que soñaba con comer un plátano mientras caminaba hacia el altar para casarse con su esposa. Lo vi en una gran base de datos de registro de los sueños de personas en una aplicación. Un hombre imagina casarse con una mujer, mientras que comer un alimento en forma fálica. Eso también tiene que significar algo, ¿verdad?

Tenía razón Sigmund Freud? Desde sus teorías por primera vez a la atención del público, la respuesta más honesta a esta pregunta sería un encogimiento de hombros. Fue Karl Popper, el filósofo austriaco-británico, que hizo más claro este punto. Popper afirmaba célebremente que las teorías de Freud no eran falsable. No había manera de probar si eran verdaderas o falsas.

Freud podría decir que la persona que escribe de una “penistrian” estaba revelando un deseo sexual posiblemente reprimida. La persona podría responder que no estaba revelando nada; que podría haber hecho tan fácilmente un error tipográfico inocente, como “pedaltrian.” Sería una-dijo, ella dijo situación. Freud podría decir que el caballero soñando con comer un plátano en su día de la boda estaba pensando en secreto de un pene, revelando su deseo de casarse con un hombre de verdad en lugar de una mujer. El caballero podía decir que acaba de pasar a ser el sueño de un plátano. Él podría haber sido tan fácilmente soñando con comer una manzana mientras caminaba hacia el altar. Sería-dijo, él-dijo. No había manera de poner la teoría de Freud a una prueba real.

Hasta ahora, eso es.

la ciencia de datos hace que muchas partes de Freud falsable-pone muchas de sus famosas teorías de la prueba. Vamos a empezar con símbolos fálicos en sueños. El uso de un gran conjunto de datos registrados de los sueños, podemos fácilmente observar la frecuencia con que aparecen los objetos en forma fálica. La comida es un buen lugar para enfocar este estudio. Se manifiesta en muchos sueños, y muchos alimentos tienen la forma de falsos-plátanos, pepinos, perros calientes, etc. A continuación, podemos medir los factores que podrían hacer que soñar

más sobre ciertos alimentos que otras, la frecuencia con que se comen, lo sabroso la mayoría de la gente a encontrar, y, sí, ya sean de naturaleza fálica.

Podemos probar si dos de los alimentos, los cuales son igualmente populares, pero uno de los cuales tiene la forma de un falo, aparecen en los sueños en diferentes cantidades. Si los alimentos en forma de falo no son más propensos a ser soñado que otros alimentos, a continuación, los símbolos fálicos no son un factor significativo en nuestros sueños. Gracias a los grandes datos, esta parte de la teoría de Freud puede de hecho ser falsable.

He recibido los datos de las Sombras, una aplicación que pide a los usuarios grabar sus sueños. Codifiqué los alimentos incluidos en decenas de miles de sueños.

En general, lo que nos hace soñar con los alimentos? El predictor principal es la frecuencia con que los consumimos. La sustancia que más soñaba es el agua. Los veinte primeros alimentos incluyen pollo, pan, bocadillos, y el arroz-notablemente anti-freudiana.

El segundo factor de predicción de la frecuencia con un alimento aparece en los sueños es cómo las personas les resulta sabrosas. Los dos alimentos que soñamos con mayor frecuencia son el chocolate y la pizza sobre todo un-freudiana pero famoso sabroso.

Entonces, ¿qué acerca de los alimentos con forma fálica? ¿Se cuelan en nuestros sueños con frecuencia inesperado? Nop.

Los plátanos son la segunda fruta más común para aparecer en los sueños. Pero también son la segunda fruta más consumida. Así que no necesitamos a Freud para explicar la frecuencia con la que soñamos plátanos. Los pepinos son el séptimo vegetal más común para aparecer en los sueños. Ellos son el séptimo vegetal de mayor consumo. Así que de nuevo su forma no es necesario explicar su presencia en nuestra mente mientras dormimos. Los perritos calientes se soñaban con mucha menos frecuencia que las hamburguesas. Esto es cierto incluso de controlar por el hecho de que la gente come más hamburguesas que los perros.

En general, el uso de un análisis de regresión (un método que permite a los científicos sociales de separar el impacto de múltiples factores) a través de todas las frutas y verduras, me encontré con que el ser de un alimento en forma de falo no darle más probabilidades de aparecer en los sueños que sería cabría esperar por su popularidad. Esta teoría de Freud es falsable y, al menos según mi mirada a los datos, falsos.

A continuación, considere actos fallidos. El psicólogo hipótesis de que usamos nuestros errores en las formas en que misspeak o miswrite a revelamos nuestros deseos subconscientes, con frecuencia sexual. Podemos utilizar grandes volúmenes de datos para probar esto? Aquí está una manera: ver si nuestros errores, nuestros resbalones-magra en la dirección de la traviesa. Si nuestros deseos sexuales enterrados cuelan en nuestros recibos, debe haber un número desproporcionado de errores que incluyen palabras como “pene”, “pene” y “sexo”.

Es por esto que he estudiado un conjunto de datos de más de 40.000 errores de escritura recogidos por los investigadores de Microsoft. El conjunto de datos incluye errores que las personas hacen pero luego inmediatamente correctas. En estas decenas de miles de errores, había un montón de individuos que cometen errores de una especie sexual. No fue la ya mencionada “penistrian.” También hubo alguien que escribe “sexurity” en lugar de “seguridad” y “gallos” en

lugar de “rocas”. Pero también hubo un montón de resbalones inocentes. La gente escribió de “pindows” y “fegetables”, “aftermoons” y “refridersators”.

Así fue el número de resbalones sexuales inusual?

Para probar esto, he utilizado por primera vez el conjunto de datos Microsoft para modelar cómo la gente cambia con frecuencia erróneamente determinadas letras. He calculado la frecuencia con la que sustituyen a con una s, AG con una h. Entonces creé un programa informático que cometió errores en la forma en que la gente hace. Podríamos llamarlo Bot error. Bot error reemplazado en con un s con la misma frecuencia que los seres humanos en el estudio que hizo Microsoft. Se reemplazó con un ag h tan a menudo como lo hicieron. Y así. Me encontré con el programa en las mismas palabras de personas habían conseguido mal en el estudio de Microsoft. En otras palabras, el robot trató de deletrear “peatón” y “rocas”, “Windows” y “refrigerador”. Pero cambió un R con al tan a menudo como la gente hace y escribió, por ejemplo, “tocks.” Se cambiaron una r ac con tanta frecuencia como lo hacen los humanos y escribieron “gallos”.

Entonces, ¿qué podemos aprender de la comparación de Bot error con los humanos normalmente por descuido? Después de hacer unos pocos millones de errores, sólo de colocar mal las letras en las formas en que los seres humanos hacen, Bot error había cometido numerosos errores de carácter freudiana. Es mal escrito “concha” como “sexshell”, “lápiz labial” como “lipsdick” y “afortunado” como “fuckiest”, junto con muchos otros errores similares. Y aquí está la clave Bot punto de error, que por supuesto no tiene un subconsciente, era igual de propensos a cometer errores que podrían ser percibidos como personas reales sexual como eran. Con la advertencia, ya que los científicos sociales les gusta decir, que es necesario que haya más investigación, esto significa que los errores de orientación sexual no son más propensas que los seres humanos hacen lo que se puede esperar por azar.

En otras palabras, para que las personas cometen errores tales como “penistrian”, “sexurity,” y “gallos”, no es necesario tener algún tipo de conexión entre los errores y lo prohibido, algo de teoría de la mente donde la gente revela sus deseos secretos a través sus errores. Estas tiras de los dedos pueden explicarse completamente por la frecuencia típica de los errores tipográficos. La gente hace un montón de errores. Y si haces suficientes errores, el tiempo de empezar a decir cosas como “lipsdick”, “fuckiest” y “penistrian.” Si un tipo de mono durante el tiempo suficiente, con el tiempo se escriben “ser o no ser”. Si una persona digita el tiempo suficiente, que con el tiempo se escribe “penistrian.”

La teoría de Freud de que los errores revelan nuestro subconsciente quiere es falsable de hecho-y, de acuerdo con mi análisis de los datos, falso.

Big Data nos dice un plátano es siempre sólo un plátano y un sólo un mal escrito “penistrian” “peatón”.

Así fue Freud destino fuera de totalmente en todas sus teorías? No exactamente. Cuando llegué por primera vez acceso a los datos pornhub, me encontré con una revelación que no me pareció al menos algo freudiana. De hecho, esta es una de las cosas más sorprendentes que he encontrado

sin embargo, durante mis investigaciones de datos: un número impresionante de personas que visitan los sitios de pornografía principales están buscando representaciones de incesto.

De las cien primeras búsquedas por hombres en Pornhub, uno de los sitios de pornografía más populares, dieciséis están buscando videos de incesto de temática. advertencia esta feria se va a poner un pequeño gráfico: incluyen “hermano y hermana”, “paso madre folla hijo”, “madre e hijo”, “madre folla hijo” y “verdadero hermano y hermana” La pluralidad de. búsquedas incestuosas masculinos son para escenas con las madres y los hijos. ¿Y mujeres? Nueve de los cien principales búsquedas por mujeres en Pornhub son videos de incesto con temas, y la imaginería característica similar, aunque con el sexo de cualquier padre y el niño que se menciona generalmente invierte. De este modo la pluralidad de búsquedas incestuosas realizadas por las mujeres son para escenas con padres e hijas.

No es difícil de localizar en estos datos, al menos, un débil eco de complejo de Edipo de Freud. Se planteó la hipótesis de un deseo casi universal en la infancia, que más tarde se reprime, por relaciones sexuales con los padres del sexo opuesto. Si sólo el psicólogo vienés había vivido el tiempo suficiente para convertir sus habilidades analíticas a los datos pornhub, donde el interés por los padres del sexo opuesto parece estar confirmada por adultos con gran claridad es reprimida y poco.

Por supuesto, los datos Pornhub no nos pueden decir con certeza que las personas fantasean acerca al ver esos videos. ¿Son realmente imaginando tener relaciones sexuales con sus propios padres? búsquedas de Google pueden dar más pistas que hay un montón de gente con tales deseos.

Considere todas las búsquedas del tipo “yo quiero tener relaciones sexuales con mi. . .” La forma del número para completar esta búsqueda es ‘madre’. En general, más de tres cuartas partes de las búsquedas de esta forma son incestuosa. Y esto no se debe a la redacción particular. Las búsquedas de la forma “Me atraen. . .”, Por ejemplo, son aún más dominado por las admisiones de los deseos incestuosos. Ahora admito-a riesgo de decepcionar Herr Freud-que estos no son particularmente comunes búsquedas: unos pocos miles de personas cada año en los Estados Unidos que admiten una atracción a su madre. Alguien también tendría que dar la noticia a Freud que las búsquedas de Google, como se discutirá más adelante en este libro, a veces sesgar hacia lo prohibido.

Pero aún. Hay un montón de lugares inapropiados que tienen las personas que yo hubiera esperado que se han mencionado con más frecuencia en las búsquedas. ¿Jefe? ¿Empleado? ¿Estudiante? ¿Terapeuta? ¿Paciente? El mejor amigo de la esposa? El mejor amigo de la hija? ¿La hermana de la esposa? esposa del mejor amigo? Ninguno de estos deseos confesados puede competir con la madre. Tal vez, en combinación con los datos Pornhub, que realmente significa algo.

Y la afirmación general de Freud de que la sexualidad puede ser moldeada por experiencias de la infancia se apoya en los datos de Google y Pornhub, lo que revela que los hombres, al menos, conservan un número desmesurado de las fantasías relacionadas con la infancia en otros lugares. Según las búsquedas de las esposas de sus maridos, algunos de los mejores fetiches de los

hombres adultos son el deseo de usar pañales y con ganas de ser amamantados, en particular, como se explicó anteriormente, en la India. escenas de sexo explícito Además, animación porno de dibujos animados con los personajes de los programas más populares entre los adolescentes varones, ha alcanzado un alto grado de popularidad. O considerar las ocupaciones de las mujeres más buscados de manera frecuente en el porno por los hombres. Los hombres que son 18-24 años de edad búsqueda con mayor frecuencia para las mujeres que son niñeras. Al igual que los hombres de 25-64 años de edad. Y los hombres de 65 años. Y para los hombres de todas las edades, profesor y animador son a la vez entre los cuatro primeros. Claramente, los primeros años de la vida parecen jugar un papel enorme en las fantasías de los hombres adultos.

Todavía no he podido utilizar todos estos datos sin precedentes sobre la sexualidad adulta de averiguar exactamente cómo se forman las preferencias sexuales. Durante las próximas décadas, otros científicos sociales y que serán capaces de crear nuevas teorías falsables sobre la sexualidad adulta y probarlos con los datos reales.

Ya puedo predecir algunos temas básicos que sin duda serán parte de una teoría basada en los datos de la sexualidad adulta. Está claro que no va a ser la historia idéntica a la que Freud dijo, con sus particulares bien definidas, etapas y universales de la infancia y la represión. Sin embargo, basado en mi primer vistazo a los datos pornhub, estoy absolutamente seguro de que el veredicto final sobre la sexualidad adulta contará con algunos de los temas clave que Freud enfatizó. Infancia jugará un papel importante. Lo mismo ocurrirá con las madres.

Es probable que hubiera sido imposible analizar Freud de esta manera hace diez años. Sin duda hubiera sido imposible hace ochenta años, cuando Freud estaba todavía vivo. Así que vamos a pensar a través de por qué estas fuentes de datos ayudaron. Este ejercicio puede ayudar a entender por qué los grandes datos es tan poderoso.

Recuerde, nosotros hemos dicho que sólo tener montones y montones de datos por sí mismo no genera automáticamente puntos de vista. tamaño de los datos, por sí mismo, está sobrevalorado. ¿Por qué, entonces, es Big Data tan poderoso? ¿Por qué va a crear una revolución en la forma en que vemos a nosotros mismos? Hay, reclamo, cuatro poderes únicos de grandes volúmenes de datos. Este análisis de Freud proporciona una buena ilustración de ellos.

Usted puede haber notado, para empezar, que nos estamos tomando en serio la pornografía en esta discusión de Freud. Y vamos a utilizar los datos de la pornografía con frecuencia en este libro. Sorprendentemente, los datos de la pornografía es raramente utilizada por los sociólogos, la mayoría de los cuales son cómodos confiando en los conjuntos de datos de encuestas tradicionales que han construido sus carreras en. Pero un momento de reflexión muestra que el uso generalizado de los datos de búsqueda y vistas que viene con la pornografía y-es el desarrollo más importante en nuestra capacidad para comprender la sexualidad humana en, también. . . De hecho, es probablemente el más importante jamás. Se trata de datos que Schopenhauer, Nietzsche, Freud y Foucault habría babeado terminado. Estos datos no existía cuando estaban vivos. No existía hace un par de décadas. Existe ahora. Hay muchas fuentes de datos únicas, en una gama de temas, que nos dan las ventanas en las zonas sobre las que



podríamos previamente simplemente adivinar. Ofreciendo nuevos tipos de datos es el primer poder de grandes volúmenes de datos.

Los datos de pornografía y los datos de búsqueda de Google no son sólo nueva; son honestos. En la era pre-digital, escondieron sus pensamientos embarazosos de otras personas. En la era digital, que todavía los esconden de otras personas, pero no a través de Internet y en sitios particulares tales como Google y PornHub, que protegen su anonimato. Estos sitios funcionan como una especie de verdad digitales suero por lo tanto, nuestra capacidad de descubrir una fascinación generalizada con el incesto. Big Data nos permite ver finalmente lo que la gente realmente quiere y realmente lo hacen, no lo que dicen que quieren y dicen que hacen. Proporcionar datos honestos es la segunda potencia de grandes volúmenes de datos.

Porque ahora hay tantos datos, no hay información significativa sobre incluso pequeños trozos de una población. Podemos comparar, por ejemplo, el número de personas que sueñan con los pepinos frente a los que sueñan con tomates. Lo que nos permite hacer zoom en pequeños subgrupos de personas es la tercera potencia de grandes volúmenes de datos.

Big Data tiene uno más impresionante potencia-uno que no se utilizó en mi estudio rápido de Freud pero podría ser en un futuro una: nos permite emprender experimentos rápidos y controlados. Esto nos permite probar la causalidad, no sólo correlaciones. Este tipo de pruebas son utilizados principalmente por las empresas ahora, pero van a ser una herramienta poderosa para los científicos sociales. Permitiéndonos hacer muchos experimentos causales es la cuarta potencia de grandes volúmenes de datos.

Ahora es el momento de desembalar cada uno de estos poderes y explorar exactamente por qué importa grandes volúmenes de datos.

## DATOS reimagined

**UNA**t 6 de la mañana del viernes de cada mes en particular, las calles de la mayor parte de Manhattan será en gran parte desierta. Las tiendas que bordean estas calles estarán cerradas, sus fachadas cubiertas por puertas de seguridad de acero, los apartamentos por encima oscuras y en silencio.

Los pisos de Goldman Sachs, la institución de banca de inversión global en el bajo Manhattan, por el contrario, serán iluminadas, sus ascensores tomar miles de trabajadores a sus escritorios. Las 7 am la mayoría de estos escritorios estarán ocupados.

No sería injusta en cualquier otro día para describir esta hora en esta parte de la ciudad como sueño. En esta mañana de viernes, sin embargo, habrá un zumbido de la energía y la emoción. En este día, la información que masivamente impactará el mercado de valores se establece en llegar.

Minutos después de su lanzamiento, esta información será reportada por los sitios de noticias. Segundos después de su lanzamiento, esta información será discutido, debatido, y se diseca, en voz alta, de Goldman y cientos de otras empresas financieras. Pero gran parte de la acción real en las finanzas en estos días sucede en milésimas de segundo. Goldman y otras firmas financieras pagaron decenas de millones de dólares para obtener acceso a los cables de fibra óptica que redujeron el tiempo de la información viaja de Chicago a Nueva Jersey por sólo cuatro milésimas de segundo (de 17 a 13). Las empresas financieras tienen algoritmos en lugar de leer la información y el comercio sobre la base de lo-todo en cuestión de milisegundos. Después de esta información crucial se libera, se moverá el mercado en menos tiempo de lo que le lleva a abrir y cerrar los ojos.

Entonces, ¿qué es este dato crucial que es tan valioso para Goldman y numerosas otras instituciones financieras?

La tasa de desempleo mensual.

La tasa, sin embargo, que tiene un impacto tan profundo en el mercado de valores que las instituciones financieras han hecho todo lo posible para maximizar la velocidad con la que recibir, analizar y actuar en consecuencia, es a partir de una encuesta telefónica que el Bureau of Labor Statistics lleva a cabo y la información es unas tres semanas-o 2 mil millones de milisegundos de edad en el momento en que se libera.

Cuando las empresas están gastando millones de dólares para astillar un milisegundo el flujo de información, que podría parecerle más que un poco extraño que el gobierno necesita tanto tiempo para calcular la tasa de desempleo.

De hecho, conseguir estos números críticos a cabo antes era uno de los programas principales de Alan Krueger cuando asumió como presidente del Consejo de Asesores Económicos del Presidente Obama en 2011. No tuvo éxito. “O el BLS no cuenta con los recursos”, ha concluido. “O que están atrapados en el pensamiento del siglo XX.”

Con el gobierno claramente no aumentaron el ritmo en el corto plazo, hay una manera de obtener al menos una medida aproximada de las estadísticas de desempleo a un ritmo más rápido? En esta era de alta tecnología, cuando casi todos los haga clic en cualquier ser humano hace en Internet se registra en algún lugar, ¿realmente tenemos que esperar semanas para averiguar cuántas personas están sin trabajo?

Una posible solución se inspiró en la obra de un ex ingeniero de Google, Jeremy Ginsberg. Ginsberg dio cuenta de que los datos de salud, al igual que los datos de desempleo, fue puesto en libertad con un retraso por parte del gobierno. Los Centros para el Control y Prevención de Enfermedades toma una semana para conocer los datos de la gripe, a pesar de que los médicos y hospitales se beneficiarían de tener los datos mucho más pronto.

Ginsberg sospecha que las personas enfermas con la gripe es probable que hagan búsquedas relacionadas con la gripe. En esencia, se informarían sus síntomas a Google. Estas búsquedas, pensó, podría dar una medida razonablemente precisa de la tasa actual de la gripe. De hecho, las búsquedas como “síntomas de la gripe” y “dolores musculares” han demostrado ser indicadores importantes de la rapidez con la gripe se está extendiendo.\*

Mientras tanto, los ingenieros de Google crearon un servicio, Google se correlacionan, que da a los investigadores fuera de los medios para experimentar con el mismo tipo de análisis a través de una amplia gama de campos, no sólo la salud. Los investigadores pueden tomar cualquier serie de datos que se están siguiendo en el tiempo y ver lo que las búsquedas de Google se correlacionan más con ese conjunto de datos.

Por ejemplo, el uso de Google se correlacionan, Hal Varian, economista jefe de Google, y fueron capaces de demostrar qué búsquedas Canción más estrechamente precios de la vivienda. Cuando los precios de la vivienda están subiendo, los estadounidenses tienden a buscar frases tales como “80/20 de la hipoteca”, “constructor de casas nuevas,” y “tasa de apreciación.” Cuando los precios de la vivienda están cayendo, los estadounidenses tienden a buscar frases tales como “venta corta proceso “,’hipoteca bajo el agua’y’alivio de la deuda de la hipoteca perdón.’

Así las búsquedas de Google pueden ser utilizados como una prueba de fuego para el desempleo de la misma forma que puedan para precios de la vivienda o la gripe? Podemos decir, simplemente por lo que las personas están buscando en Google, ¿cuántas personas están en paro, y podemos hacerlo mucho antes de que el gobierno coteja los resultados de la encuesta?

Un día, puse la tasa de desempleo de Estados Unidos desde 2004 hasta 2011 en Google Correlación.

De los billones de búsquedas de Google durante ese tiempo, ¿qué te parece que resultó ser lo más firmemente conectado al desempleo? Se pueden imaginar “oficina de desempleo”, o algo similar. Eso era alto, pero no en la parte superior. “Nuevos trabajos”? También alto, pero tampoco en la parte superior.

El más alto durante el período que han buscado y estos términos Shift-era “Slutload.” Así es, la búsqueda más frecuente fue de un sitio pornográfico. Esto puede parecer extraño a primera vista, pero la gente en paro, presumiblemente, tienen un montón de tiempo en sus manos. Muchos están atrapados en casa, solo y aburrido. Otro de los altamente correlacionado búsquedas de éste en el ámbito PG-es “Solitario Spider.” Una vez más, no es sorprendente para un grupo de personas que supuestamente tienen una gran cantidad de tiempo en sus manos.

Ahora, no estoy argumentando, con base en éste análisis, que el seguimiento “Slutload” o “Spider Solitaire” es la mejor manera de predecir la tasa de desempleo. Las desviaciones específicas que utilizan los parados pueden cambiar con el tiempo (en un momento, “Rawtube”, un sitio de pornografía diferente, fue uno de los más fuertes correlaciones) y ninguno de estos términos particulares por sí mismo atrae a algo parecido a una pluralidad de los parados. Pero he encontrado que por lo general una mezcla de búsquedas relacionadas con la desviación-puede realizar un seguimiento de la tasa de desempleo y sería una parte de la mejor modelo de predicción de la misma.

Este ejemplo ilustra la primera potencia de grandes volúmenes de datos, la reinención de lo que califica como de datos. Con frecuencia, el valor de grandes volúmenes de datos no es su tamaño; es que se le puede ofrecer nuevos tipos de información para estudiar, información que nunca antes habían sido recogidos.

Antes de que Google no había información disponible sobre ciertas actividades de ocio la venta de entradas de películas de, por ejemplo, que podría dar algunas pistas en cuanto a cómo las personas tienen mucho tiempo en sus manos. Pero la oportunidad de saber cuánto se está jugando al solitario o está siendo observado porno es nueva y potente. En este caso, estos datos podría ayudarnos a medir más rápidamente cómo la economía está haciendo, al menos hasta que el gobierno aprende a conducir y cotejar una encuesta con mayor rapidez.

La vida en el campus de Google en Mountain View, California, es muy diferente de la de la sede de Manhattan de Goldman Sachs. A las 9 am oficinas de Google son casi vacío. Si alguna trabajadores están alrededor, es probable que sea para desayunar para panqueques libre de plátano y arándanos, claras de huevo revueltos, pepino agua filtrada. Algunos empleados podrían estar fuera de lugar: en una reunión fuera del sitio en Boulder o Las Vegas o tal vez en un viaje de esquí libre a Lake Tahoe. Alrededor de la hora del almuerzo, se llenarán las canchas de voleibol de arena y campos de fútbol de césped. El mejor burrito que he comido estaba en el restaurante mexicano de Google.

¿Cómo puede una de las compañías tecnológicas más grandes y más competitivas en el mundo aparentemente tan relajado y generoso? Google aprovechada grandes volúmenes de datos de una manera que ninguna otra compañía siempre tiene que construir un flujo de dinero automatizado.

La compañía juega un papel crucial en este libro puesto que las búsquedas de Google son, con mucho, la fuente dominante de grandes volúmenes de datos. Pero es importante recordar que el éxito de Google se basa en la misma colección de un nuevo tipo de datos.

Si eres lo suficientemente viejo para haber utilizado Internet en el siglo XX, es posible que recuerde los diversos motores de búsqueda que existían en ese entonces-MetaCrawler, Lycos, Altavista, para nombrar unos pocos. Y es posible que recuerde que estos motores de búsqueda eran, como mucho, ligeramente fiable. A veces, si tienes suerte, se las arreglaron para encontrar lo que quería. A menudo, no lo harían. Si escribía “Bill Clinton” en los motores de búsqueda más populares en la década de 1990, los resultados principales, se incluye un sitio aleatorio que se acaba de proclamar “Bill Clinton Sucks” o un sitio que contó con una mala broma Clinton. Apenas la información más relevante sobre el entonces presidente de los Estados Unidos.

En 1998, Google presentó. Y sus resultados de búsqueda eran sin duda mejor que los de cada uno de sus competidores. Si escribía “Bill Clinton” en Google en 1998, se le dio a su sitio web, la dirección de correo electrónico de la Casa Blanca, y las mejores biografías del hombre que existía en el Internet. Google parecía ser magia.

Lo que había fundadores de Google, Sergey Brin y Larry Page, hecho de manera diferente?

Otros motores de búsqueda ubicados por sus usuarios los sitios web que incluían más frecuentemente la frase para los que buscaban. Si estaba buscando información sobre “Bill Clinton”, los motores de búsqueda encontrarían, en toda la Internet, los sitios web que tenía la mayoría de las referencias a Bill Clinton. Había muchas razones este sistema de clasificación era imperfecto y una de ellas fue que era fácil de jugar con el sistema. Un sitio de broma con el texto “Bill Clinton, Bill Clinton, Bill Clinton, Bill Clinton, Bill Clinton” escondido en algún lugar en su página sería puntuación superior a la web oficial de la Casa Blanca.\*

Lo que Brin y Page hicieron fue encontrar una manera de grabar un nuevo tipo de información que era mucho más valioso que un simple recuento de las palabras. Los sitios web suelen serían, cuando se habla de un tema, vincular a los sitios que pensaban que eran más útiles en la comprensión de este tema. Por ejemplo, el New York Times, si se menciona Bill Clinton, podría permitir a los lectores que han hecho clic en su nombre, que se enviará a la página oficial de la Casa Blanca.

Cada sitio web la creación de uno de estos enlaces, en cierto sentido, dando su opinión de la mejor información sobre Bill Clinton. Brin y Page se pudieron combinar todas estas opiniones sobre todos los temas. Podría crowdsource las opiniones de The New York Times, millones de servidores de listas, cientos de bloggers, y todos los demás en el Internet. Si una gran cantidad de la gente piensa que el vínculo más importante para “Bill Clinton”, fue su sitio web oficial, esta fue probablemente la página web que la mayoría de la gente en busca de “Bill Clinton” querría ver.

Este tipo de enlaces eran datos que otros motores de búsqueda ni siquiera consideró, y eran muy predictivo de la información más útil sobre un tema determinado. El punto aquí es que Google no dominó búsqueda simplemente mediante la recopilación de más datos que los demás. Lo hicieron mediante la búsqueda de un mejor tipo de datos. Menos de dos años después de su

lanzamiento, Google, impulsado por su análisis de enlaces, creció hasta convertirse en el más popular motor de búsqueda de internet. Hoy en día, Brin y Page están juntos valor de más de \$ 60 mil millones.

Al igual que con Google, por lo que con todos los demás tratando de utilizar los datos para entender el mundo. La revolución de grandes volúmenes de datos es menos sobre la recogida de más y más datos. Se trata de recoger los datos correctos.

Pero Internet no es el único lugar donde se puede recoger nuevos datos y dónde obtener los datos correctos puede tener resultados profundamente perturbadores. Este libro es en gran medida de cómo los datos en la web que nos pueden ayudar a comprender mejor la gente. En la siguiente sección, sin embargo, no tiene nada que ver con datos de la web. De hecho, no tiene nada que ver con la gente. Pero sí ayuda a ilustrar el punto principal de este capítulo: el valor enorme de datos nuevos y no convencionales. Y los principios que nos enseña son útiles en la comprensión de la revolución de datos basada en digital.

## ÓRGANOS como datos

---

En el verano de 2013, un caballo de color marrón rojizo, de un tamaño superior a la media, con una melena negro, se sentó en una pequeña granja en el estado de Nueva York. Fue uno de 152 caballos de un año de edad en August Fasig-Tipton Select Yearling venta en Saratoga Springs, y uno de diez caballos mil años de edad, siendo subastado ese año.

hombres y mujeres ricas, cuando desembolsar una gran cantidad de dinero en un caballo de carreras, quieren el honor de elegir el nombre del caballo. Así, el caballo de color marrón rojizo todavía no tenía un nombre y, como la mayoría de los caballos en la subasta, se hace referencia en lugar de su número de granero, 85.

Era poco lo que hizo N° 85 se destacan en esta subasta. Su pedigrí era buena, pero no excelente. Su padre (padre), Pioneerof [sic] del Nilo, era un caballo de carreras superior, pero otros niños de Pioneerof el Nilo no había tenido mucho éxito en competición. También hubo dudas en función de cómo se veía No. 85. Tenía un rasguño en su tobillo, por ejemplo, que algunos compradores preocupados de que podría ser evidencia de una lesión.

El propietario actual de No. 85 fue un magnate de la cerveza egipcia, Ahmed Zayat, que había llegado al norte del estado de Nueva York en busca de vender el caballo y comprar algunos otros.

Al igual que casi todos los propietarios, Zayat contrató a un equipo de expertos para ayudarlo a elegir qué caballos para comprar. Pero sus expertos eran un poco diferentes a los utilizados por casi todos los otros propietarios. Los expertos típicas de caballos que verías en un evento como este eran hombres de mediana edad, muchos de Kentucky o Florida rural con poca educación, pero con un fondo de la familia en el negocio de los caballos. Los expertos de Zayat, sin embargo, vino de una pequeña empresa llamada JCA. El jefe de la JCA no era un hombre a

caballo de la vieja escuela. El jefe de la JCA, en cambio, era Jeff Seder, un hombre excéntrico, nacido en Filadelfia con una pila de grados de Harvard.

Zayat había trabajado con JCA antes, por lo que el proceso era familiar. Después de unos días de la evaluación de los caballos, el equipo de Seder volvería a Zayat con cinco o menos caballos que recomendaban comprar para reemplazar No. 85.

Esta vez, sin embargo, era diferente. El equipo de Seder regresó a Zayat y le dijo que eran incapaces de cumplir con su petición. Ellos simplemente no podían recomendar que comprar cualquiera de los otros 151 caballos ofrecidos a la venta ese día. En su lugar, ofrecieron una petición inesperada y casi desesperada. Zayat absoluta y positivamente no podía vender el caballo N° 85. Este caballo, JCA declaró, era no sólo el mejor caballo de la subasta; él era el mejor caballo del año y, posiblemente, la década. “Vender su casa”, el equipo le imploró. “No vender este caballo.”

Al día siguiente, con poca fanfarria, caballo N° 85 fue comprado por \$ 300.000 para un hombre que se hace llamar Incardo Bloodstock. Caballos pura sangre, se reveló más tarde, era un seudónimo utilizado por Ahmed Zayat. En respuesta a las peticiones de Seder, Zayat había vuelto a comprar su propio caballo, una acción casi sin precedentes. (Las reglas de la subasta impidieron Zayat desde la simple eliminación de los caballos de la subasta, necesitando por lo tanto la transacción seudónimo.) Sesenta y dos caballos en la subasta se vende a un precio más alto que el caballo N° 85, con dos ir a buscar más de \$ 1 millón cada uno .

Tres meses más tarde, Zayat finalmente eligió un nombre para el N° 85: Pharoah americano. Y dieciocho meses más tarde, en un 75 grados sábado por la noche en las afueras de la ciudad de Nueva York, América del faraón se convirtió en el primer caballo en más de tres décadas para ganar la Triple Corona.

¿Qué le dijo Jeff Seder sabe de caballos N° 85 que al parecer conocía a nadie más? ¿Cómo este hombre de Harvard obtener tan buenos en la evaluación de los caballos?

Me encontré por primera vez con Seder, que tenía entonces sesenta y cuatro, en una tarde de junio abrasador en Ocala, Florida, más de un año después de la American Pharoah Triple Corona. El evento fue un escaparate de una semana para caballos de dos años de edad, que culmina en una subasta, no muy diferente a la 2013 evento donde Zayat compró su propio caballo.

Seder tiene un auge, la voz de Mel Brooks similar, una cabeza llena de cabello, y un rebote discernible en su paso. Llevaba tirantes, pantalones caqui, una camisa de color negro con el logotipo de su compañía en él, y un audífono.

Durante los próximos tres días, me contó su historia y la vida cómo llegó a ser tan bueno en la predicción de los caballos. No era una ruta directa. Después de graduarse magna cum laude y Phi Beta Kappa de Harvard, Seder pasó a obtener, también de Harvard, un título de abogado y un título en negocios. A los veintiséis años de edad, trabajaba como analista de Citigroup en la ciudad de Nueva York, pero se sentía infeliz y quemada. Un día, sentado en el atrio a las nuevas oficinas de la empresa en Lexington Avenue, se encontró el estudio de un gran mural de un campo abierto. La pintura le recordaba a su amor por el campo y su amor por los caballos. Se fue

a casa y se miró en el espejo con su traje de tres piezas en. Entonces supo que no estaba destinado a ser un banquero y él no estaba destinado a vivir en la ciudad de Nueva York. A la mañana siguiente, renunció a su trabajo.

Seder se trasladó a Pennsylvania rural y se encaminó a través de una variedad de puestos de trabajo en el sector textil y medicina deportiva antes de dedicar su vida a tiempo completo a su pasión: predecir el éxito de los caballos de carreras. Los números en las carreras de caballos son ásperas. De los mil y dos años de edad, los caballos se exhiben en una subasta de Ocala, uno de los de la nación de mayor prestigio, tal vez cinco serán terminar ganando una carrera con un bolso significativa. ¿Qué pasará con los otros 995 caballos? Aproximadamente un tercio va a resultar demasiado lento. Otro tercio será lesionarse-más porque sus miembros no pueden soportar la enorme presión de galopar a toda velocidad. (Cada año, cientos de caballos mueren en los circuitos americanos, sobre todo debido a la rotura de las piernas.) Y el tercio restante tendrán lo que se podría llamar el síndrome de Bartleby. Bartleby en extraordinario cuento de Herman Melville, deja de funcionar y responde a cada petición de su empleador hace con “yo preferiría no hacerlo.” Muchos caballos, al principio de sus carreras de competición, al parecer vienen a darse cuenta de que no tienen que correr si no se sienten como él. Pueden comenzar una carrera en funcionamiento rápidamente, pero, en algún momento, van a simplemente reducir la velocidad o detener la ejecución en conjunto. ¿Por qué correr alrededor de un óvalo tan rápido como sea posible, especialmente cuando sus pezuñas y corvejones duelen? “Preferiría no hacerlo”, deciden. (Tengo una debilidad por Bartlebys, caballo o humano). especialmente cuando sus pezuñas y corvejones duelen? “Preferiría no hacerlo”, deciden. (Tengo una debilidad por Bartlebys, caballo o humano). especialmente cuando sus pezuñas y corvejones duelen? “Preferiría no hacerlo”, deciden. (Tengo una debilidad por Bartlebys, caballo o humano).

Con las probabilidades en contra de ellos, ¿cómo pueden los propietarios escoge un caballo rentable? Históricamente, la gente ha creído que la mejor manera de predecir si un caballo tendrá éxito ha sido analizar su pedigrí. Ser un experto en caballos significa ser capaz de recitar de un tirón todo lo que nadie podría querer saber acerca de un caballo padre, madre, abuelos, abuelas, hermanos y hermanas. Agentes anuncian, por ejemplo, que un caballo grande “llegó a su tamaño de forma legítima” si la línea de su madre tiene un montón de grandes caballos.

Hay un problema, sin embargo. Mientras pedigrí sí importa, todavía puede explicar solamente una pequeña parte del éxito de un caballo de carreras. Tenga en cuenta la trayectoria de hermanos completos de todos los caballos nombrados caballo del año, el más prestigioso premio anual de carreras. Estos caballos tienen la mejor-la pedigríes historia familiar posible idénticos como los caballos de la historia universal. Aún así, más de las tres cuartas partes no ganan una carrera importante. La forma tradicional de predecir el éxito de caballo, los datos nos dice, deja un amplio margen de mejora.

En realidad no es tan sorprendente que pedigrí no es tan predictiva. Pensar en los seres humanos. Imagínesse un propietario de la NBA que compró su futuro equipo, como de diez años de edad, en función de sus genealogías. Se habría contratado a un agente para examinar Earvin Johnson III, hijo de “Magic” Johnson. “Tiene buen tamaño, hasta el momento,” un agente podría



decir. “Es legítimo tamaño, desde la línea de Johnson. Él debe tener una gran visión, el desinterés, el tamaño y la velocidad. Parece ser sociable gran personalidad,. paseo confiado. De buen ver. Esta es una gran apuesta.”Por desgracia, catorce años después, este propietario tendría un 6'2” (abreviatura de una pelota de jugador profesional) blogger de moda para E! Earvin Johnson III podría ser de gran ayuda en el diseño de los uniformes, pero que probablemente son de poca ayuda en la cancha.

Junto con el blogger de moda, un dueño de la NBA que eligió un equipo como muchos propietarios de elegir caballos probablemente apoderarse de Jeffrey y Marcus Jordan, ambos hijos de Michael Jordan, y ambos de los cuales resultó jugadores universitarios mediocres. Buena suerte contra los Cavaliers de Cleveland. Ellos son guiados por LeBron James, cuya madre es 5'5” . O imaginar un país que eligió a sus líderes sobre la base de sus genealogías. Nos gustaría ser guiados por gente como George W. Bush. (Lo siento, no pudo resistir.)

agentes caballo hacer uso de otra información, además de pedigrí. Por ejemplo, analizan los aires de dos años de edad y examinar visualmente los caballos. En Ocala, me pasaba horas conversando con varios agentes, que era el tiempo suficiente para determinar que hubo poco acuerdo sobre lo que en realidad estaban buscando.

Añadir a estas contradicciones e incertidumbres rampantes el hecho de que algunos compradores de caballos tienen lo que parece que los fondos infinitos, y se obtiene un mercado con bastante grandes ineficiencias. Hace diez años, caballo N° 153 era un niño de dos años que corrió más rápido que cualquier otro caballo, veía hermosa a la mayoría de los agentes, y tenía un pedigrí-una maravillosa descendiente de Northern Dancer y la Secretaria, dos de los más grandes de todos los caballos de carreras hora. Un multimillonario irlandés y un jeque de Dubai, tanto querían comprar. Se metieron en una guerra de ofertas que se convirtió rápidamente en un concurso de orgullo. Como cientos de hombres y mujeres a caballo aturdidos miraban, las ofertas fueron yendo a más y más alto, hasta los dos años de edad, caballo finalmente se vendió por \$ 16 millones, con mucho, el precio más alto jamás pagado por un caballo. Caballo N° 153, que se le dio el nombre de The Green Monkey, corrió tres carreras, ganó apenas \$ 10.000,

Seder nunca tuvo ningún interés en los métodos tradicionales de evaluación de los caballos. Él estaba interesado sólo en los datos. Él planeó para medir diversos atributos de los caballos de carreras y ver cuál de ellos correlacionado con su desempeño. Es importante señalar que Seder elaboró su plan de la mitad de una década antes de la World Wide Web fue inventada. Sin embargo, su estrategia se basa mucho en la ciencia de datos. Y las lecciones de su historia son aplicables a cualquier persona utilizando grandes volúmenes de datos.

Durante años, la búsqueda de Seder no produjo nada más que frustración. Se midió el tamaño de los orificios nasales de los caballos, la creación de la primera y mayor conjunto de datos del mundo en tamaño de ventana de la nariz del caballo y las ganancias eventuales. tamaño de la fosa nasal, se encontró, no predijo el éxito caballo. Dio caballos electrocardiogramas para examinar sus corazones y cortar las extremidades de los caballos muertos para medir el volumen de sus músculos de contracción rápida. Una vez agarró una pala fuera de un granero para determinar el tamaño de los excrementos de los caballos, en la teoría de que el derramamiento de

exceso de peso antes de un evento puede ralentizar un caballo hacia abajo. Nada de esto se correlacionó con éxito en competición.

Entonces, hace doce años, obtuvo su primer gran éxito. Seder decidió medir el tamaño de los órganos internos de los caballos. Dado que esto era imposible con la tecnología existente, construyó su propia ultrasonido portátil. Los resultados fueron notables. Se encontró que el tamaño del corazón, y en particular el tamaño del ventrículo izquierdo, fue un predictor del éxito masivo de un caballo, la variable más importante. Otro órgano que importaba era el bazo: caballos con pequeñas bazos ganaron prácticamente nada.

Seder tenía un par de golpes. Se digitalizó miles de videos de caballos al galope y se encontró que ciertos aires se correlacionaron con éxito pista de carreras. También descubrió que algunos caballos de dos años de edad, sibilancias después de ejecutar un octavo de milla. Tales caballos a veces se venden por hasta un millón de dólares, pero los datos de Seder le dijeron que los niños con sibilancias casi nunca resultan. Él por lo tanto asigna un asistente para sentarse cerca de la línea de meta y eliminar a las sibilantes.

De alrededor de mil caballos en la subasta Ocala, más o menos cada diez pasar todas las pruebas de Seder. Ignora por completo pedigrí, con excepción de lo que influirá en el precio de un caballo se venderá por. “Pedigree nos dice un caballo podría tener una muy pequeña posibilidad de ser grande”, dice. “Pero si puedo ver que está muy bien, ¿qué me importa cómo llegó hasta allí?”

Una noche, Seder me invitó a su habitación en el hotel Hilton en Ocala. En la sala, me habló de su infancia, su familia y su carrera. Me mostró fotos de su esposa, su hija y el hijo. Me dijo que era uno de los tres estudiantes judíos en su instituto de Filadelfia, y que cuando entró en él era 4'10” . ( Creció en la universidad de 5'9” ) Me habló de su caballo favorito: Pinky Pizwaanski. Seder compra y se nombró a este caballo después de un jinete homosexuales. En su opinión, Pinky, el caballo, siempre daba un gran esfuerzo, incluso si él no era el más exitoso.

Por último, me mostró el archivo que incluye todos los datos que había registrado en el hoyo 85, el archivo que llevó a la mayor predicción de su carrera. ¿Estaba regalando su secreto? Tal vez, pero dijo que no le importaba. Más importante para él que la protección de sus secretos estaba siendo dado la razón, mostrando al mundo que estos veinte años de craqueo extremidades, palear caca, y un ensamble ultrasonidos habían valido la pena.

He aquí algunos de los datos sobre el caballo N° 85:

<b>NO. 85 (más tarde AMERICAN Pharoah) PERCENTILES COMO UNA AÑOS</b>	
	<b>PERCENTILES</b>
Altura	56
Peso	61
Árbol genealógico	70

<i>Ventrículo izquierdo</i>	
-----------------------------	--

Allí estaba, cruda y clara, la razón de que Seder y su equipo habían llegado a ser tan obsesionado con el número 85. Su ventrículo izquierdo estaba en el percentil 99.61st!

No sólo eso, sino que todos sus otros órganos importantes, incluyendo el resto de su corazón y el bazo, eran excepcionalmente grande también. En términos generales, cuando se trata de carreras, Seder había encontrado, cuanto más grande es el ventrículo izquierdo, mejor. Sin embargo, un ventrículo izquierdo tan grande como esto puede ser un signo de enfermedad si los otros órganos son pequeñas. En América del faraón, todos los órganos principales eran más grandes que el promedio, y el ventrículo izquierdo fue enorme. Los datos gritaban que No. 85 era un un caballo 1-in-100.000 o incluso de uno en un millón.

¿Qué pueden aprender de los científicos de datos de proyectos del Seder?

En primer lugar, y quizás lo más importante, si se va a tratar de utilizar nuevos datos para revolucionar un campo, lo mejor es entrar en un campo donde los viejos métodos son pésimo. Los agentes de caballos de raza obsesionada con los cuales vencieron Seder montón izquierda del margen de mejora. Lo mismo hicieron los motores de búsqueda de cantidad de palabras obsesionada que Google latido.

Una debilidad de intento de Google para predecir la gripe a partir de datos de búsqueda es que ya se puede predecir la gripe muy bien simplemente usando los datos de la semana pasada y un simple ajuste estacional. Hay todavía un debate acerca de la cantidad de datos de búsqueda se suma a ese modelo simple, potente. En mi opinión, las búsquedas de Google tienen más condiciones de salud de medición promesa para los que los datos existentes es más débil y, por tanto, algo así como Google ETS puede resultar más valioso en el largo plazo de la gripe de Google.

La segunda lección es que, cuando se trata de hacer predicciones, usted no necesita preocuparse demasiado acerca de por qué sus modelos de trabajo. Seder no podía explicar plenamente para mí por qué el ventrículo izquierdo es tan importante para predecir el éxito de un caballo. Tampoco pudo explicar precisamente por el valor del bazo. Tal vez los cardiólogos y los hematólogos caballo un día va a resolver estos misterios. Pero por ahora no importa. Seder está en el negocio de las predicciones, no el negocio explicación. Y, en el negocio de las predicciones, sólo tiene que saber que algo funciona, no es por eso.

Por ejemplo, Wal-Mart utiliza los datos de las ventas en todas sus tiendas para saber qué productos de dejar de lado. Antes del huracán Frances, una tormenta destructiva que golpeó el sudeste en 2004, Walmart sospecha, correctamente, que los hábitos de compra de la gente puede cambiar cuando una ciudad está a punto de ser golpeado por una tormenta. Se estudiaron minuciosamente a través de los datos de ventas de los huracanes anteriores para ver lo que la gente quiere comprar. Una respuesta importante? Fresa Pop-Tarts. Este producto se vende siete veces más rápido de lo normal en los días previos a un huracán.

Sobre la base de su análisis, Walmart tenía camiones cargados con la fresa Pop-Tarts dirigirse Interstate 95 hacia tiendas en la trayectoria del huracán. Y, en efecto, éstos Pop-Tarts vendió bien.

¿Por Pop-Tarts? Probablemente debido a que no requieren refrigeración o cocinar. ¿Por qué la fresa? Ninguna pista. Pero cuando los huracanes golpean, las personas recurren a fresa Pop-Tarts, aparentemente. Así que en los días previos a un huracán, Walmart ahora abastece regularmente sus estanterías con cajas y más cajas de fresa Pop-Tarts. La razón de la relación, no importa. Pero la relación en sí lo hace. Tal vez un día los científicos de alimentos se darán cuenta de la asociación entre los huracanes y pastas tostadas rellenos de mermelada de fresa. Pero, a la espera de alguna de estas explicaciones, Walmart todavía tiene que almacenar sus estantes con fresa Pop-Tarts cuando los huracanes se están acercando y guardar las golosinas Rice Krispies para los días soleados.

Esta lección también es evidente en la historia de Orley Ashenfelter. Lo Seder es a caballos, Ashenfelter, economista de Princeton, puede ser al vino.

Un poco más de una década, Ashenfelter fue frustrado. Había estado comprando una gran cantidad de vino tinto de la región de Burdeos de Francia. A veces, este vino era delicioso, digno de su alto precio. Muchas veces, sin embargo, fue una decepción.

¿Por qué, se preguntó Ashenfelter, estaba pagando el mismo precio para el vino que resultó de manera tan diferente?

Un día, Ashenfelter recibió un consejo de un amigo periodista y conocedor de vinos. No había hecho una forma de averiguar si un vino sería bueno. La clave, el amigo de Ashenfelter le dijo, era el clima durante la temporada de crecimiento.

El interés de Ashenfelter se despertó. Se fue en una búsqueda para averiguar si esto era cierto y que podía comprar consistentemente mejor vino. Descargó treinta años de datos meteorológicos en la región de Burdeos. También recoge los precios de subasta de vinos. Las subastas, que se producen muchos años después de que el vino se vendió originalmente, se dirá cómo el vino resultó.

El resultado fue sorprendente. Un gran porcentaje de la calidad de un vino podría explicarse simplemente por el clima durante la temporada de crecimiento.

De hecho, la calidad de un vino puede ser roto abajo a una fórmula simple, que podríamos llamar la primera ley de la viticultura:

$$\text{Precio} = 12,145 + 0,00117 \text{ invierno la temperatura promedio de precipitaciones} + 0,0614 \text{ temporada de crecimiento} - 0.00386 \text{ cosecha de lluvia.}$$

Así que ¿por qué la calidad del vino en la región de Burdeos trabajo como este? ¿Cómo se explica la Primera Ley de Viticultura? Hay alguna explicación para el vino fórmula calor de Ashenfelter y riego principios son necesarios para uvas maduren adecuadamente.

Sin embargo, los detalles precisos de su fórmula predictiva van mucho más allá de cualquier teoría y probablemente nunca se entienden completamente, incluso por expertos en el campo.

¿Por qué un centímetro de lluvia invernal añadir, en promedio, exactamente 0,1 centavos al precio de una botella totalmente madurado de vino tinto? ¿Por qué no 0.2 centavos? 0.05 ¿Por qué no? Nadie puede responder a estas preguntas. Pero si hay 1.000 centímetros de lluvia adicional en un invierno, usted debe estar dispuesto a pagar un adicional de \$ 1 para una botella de vino.

De hecho, Ashenfelter, a pesar de no saber exactamente por qué su regresión funcionó exactamente como lo hizo, lo utilizó para comprar los vinos. Según él, “Que funcionaba de maravilla.” La calidad de los vinos que bebió mejorado notablemente.

Si su objetivo es predecir el futuro, lo que vino el buen gusto, qué productos se venden, que los caballos se ejecutarán rápido que no es necesario que preocuparse demasiado acerca de por qué su modelo funciona exactamente como lo hace. Acaba de obtener los números de la derecha. Esa es la segunda lección de la historia del caballo de Jeff Seder.

La última lección que aprender de intento exitoso de Seder para predecir un posible ganador de la Triple Corona es que usted tiene que ser abierto y flexible en la determinación de lo que cuenta como datos. No es como si los agentes de caballos de antaño eran ajenos a los datos antes de Seder llegó. Se examinaron los tiempos de carrera y cuadros genealógicos. El genio de Seder fue la búsqueda de datos donde otros no habían visto antes, al considerar las fuentes no tradicionales de datos. Para un científico de datos, una perspectiva nueva y original puede pagar.

### PALABRAS como datos

---

Un día en 2004, dos jóvenes economistas con experiencia en medios de comunicación, a continuación, Ph.D. estudiantes de la Universidad de Harvard, estaban leyendo acerca de una decisión judicial reciente en Massachusetts legalización del matrimonio entre homosexuales.

Los economistas, Matt Gentzkow y Jesse Shapiro, notaron algo interesante: dos periódicos empleadas lenguaje sorprendentemente diferente para informar de la misma historia. The Washington Times, que tiene una reputación de ser conservador, encabezó la historia: “Los homosexuales 'casan' en Massachusetts.” The Washington Post, que tiene una reputación de ser liberal, informó de que había habido una victoria para “parejas del mismo sexo”.

No es de extrañar que las diferentes organizaciones de noticias pueden inclinar en direcciones diferentes, que los periódicos pueden cubrir la misma historia con un enfoque diferente. Durante años, de hecho, Gentzkow y Shapiro habían sido ponderando si podrían utilizar su formación para ayudar a entender la economía parcialidad de los medios. ¿Por qué algunas organizaciones de noticias parecen tener una visión más liberal y otros uno más conservador?

Pero Gentzkow y Shapiro en realidad no tienen alguna idea de cómo podrían abordar esta cuestión; no podían averiguar cómo podían medir sistemática y objetivamente la subjetividad de los medios.

Lo Gentzkow y Shapiro encontraron interesante, entonces, acerca de la historia del matrimonio homosexual no era que las organizaciones de noticias diferían en su cobertura; que

era la forma en la cobertura de los periódicos difería-todo se redujo a un cambio distinto en la selección de palabras. En 2004, “los homosexuales”, como se usa en el Washington Times, era una forma anticuada y despectivo para describir a las personas homosexuales, mientras que “las parejas del mismo sexo”, como se usa en el Washington Post, hicieron hincapié en que las relaciones homosexuales eran más que otra forma de romance.

Los investigadores se preguntaron si el lenguaje podría ser la clave para entender el sesgo. ¿Se liberales y conservadores usan consistentemente diferentes frases? ¿Podrían las palabras que utilizan los periódicos en las historias pueden convertir en datos? Lo que esto puede revelar acerca de la prensa de Estados Unidos? Podríamos averiguar si la prensa era liberal o conservador? Y podríamos averiguar por qué? En 2004, estos no eran preguntas ociosas. Los mil millones de palabras en los periódicos estadounidenses ya no estaban atrapados en papel de periódico o microfilm. Algunos sitios web ahora se registran cada palabra incluida en cada historia de casi todos los periódicos en los Estados Unidos. Gentzkow y Shapiro podrían raspar estos sitios y probar rápidamente la medida en que el lenguaje podría medir el sesgo periódico. Y, al hacer esto, podrían agudizar nuestra comprensión de cómo funciona la prensa.

Sin embargo, antes de describir lo que encontraron, dejemos por un momento la historia de Gentzkow y Shapiro y su intento de cuantificar la lengua en los periódicos, y discutir cómo los investigadores, a través de una amplia gama de campos, han utilizado este nuevo tipo de palabras de datos : para entender mejor la naturaleza humana.

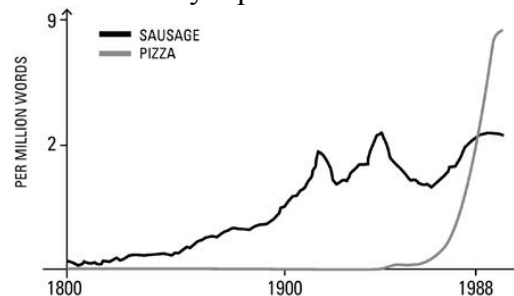
El lenguaje ha, por supuesto, siempre ha sido un tema de interés para los científicos sociales. Sin embargo, el lenguaje generalmente se requiere el estudio de la lectura atenta de los textos, y girando enormes franjas de texto en datos no era factible. Ahora, con las computadoras y la digitalización, la tabulación de las palabras a través de conjuntos masivos de documentos es fácil. Así, el lenguaje se ha convertido en objeto de un análisis de datos grandes. Los enlaces que Google utilizados estaban compuestos de palabras. Así son las búsquedas de Google que estudio. Palabras ofrecen con frecuencia en este libro. Pero el lenguaje es tan importante para la revolución grandes volúmenes de datos, que merece su propia sección. De hecho, se está utilizando tanto ahora que hay todo un campo dedicado a ella: “texto como datos.”

Un desarrollo importante en este campo es Google N-gramas. Hace algunos años, dos jóvenes biólogos, Erez Aiden y Jean-Baptiste Michel, tenían sus asistentes de investigación contando las palabras una por una en textos antiguos y polvorientos para tratar de encontrar nuevas ideas sobre cómo ciertos usos de las palabras extendido. Un día, Aiden y Michel oído acerca de un nuevo proyecto de Google de digitalizar una gran parte de los libros del mundo. Casi inmediatamente, los biólogos comprendieron que esta sería una manera mucho más fácil de entender la historia de la lengua.

“Nos dimos cuenta de nuestros métodos eran tan irremediabilmente obsoleto,” Aiden dijo a la revista Discover. “Estaba claro que no se podía competir con este gigante de la digitalización.” Por lo que decidieron colaborar con la compañía de búsquedas. Con la ayuda de los ingenieros de Google, crearon un servicio que busca a través de los millones de libros digitalizados para una

determinada palabra o frase. A continuación, los investigadores le dirán la frecuencia con esa palabra o frase aparecieron en todos los años, desde 1800 hasta 2010.

Entonces, ¿qué podemos aprender de la frecuencia con la que las palabras o frases aparecen en los libros en diferentes años? Por un lado, se aprende por el lento crecimiento de la popularidad de la salchicha y la relativamente reciente y rápido crecimiento de la popularidad de la pizza.



Pero hay lecciones mucho más profundo que eso. Por ejemplo, Google N-gramas nos puede enseñar cómo se formó la identidad nacional. Un ejemplo interesante se presenta en el libro de Aiden y de Michel, *Uncharted*.

En primer lugar, una pregunta rápida. ¿Cree que los Estados Unidos es actualmente un país unido o un país dividido? Si usted es como la mayoría de la gente, se diría que Estados Unidos está dividido en estos días debido al alto nivel de polarización política. Puede ser que incluso dicen que el país es casi tan dividida como lo ha sido nunca. América, después de todo, es ahora un código de colores: estados rojos son republicano; estados azules son Democrática. Pero, en *Uncharted* Aiden y Michel nota fascinante punto, los datos que revela cuánto más dividida que una vez fue los Estados Unidos. El punto de datos es la gente utilizar el lenguaje para hablar sobre el país.

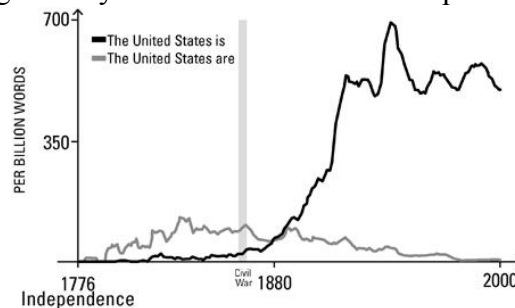
Tenga en cuenta las palabras que utilicé en el párrafo anterior, cuando hablé de cómo dividir el país es. Escribí: “Estados Unidos está dividido.” Me he referido a los Estados Unidos como un sustantivo singular. Esto es natural; es la gramática correcta y el uso estándar. Estoy seguro de que ni se dio cuenta.

Sin embargo, los estadounidenses no siempre hablan de esta manera. En los primeros días del país, los estadounidenses se refiere a los Estados Unidos el uso de la forma plural. Por ejemplo, John Adams, en su 1799 del Estado de la Unión, se refirió a “los Estados Unidos en sus tratados con su Britanic Majestad.” Si mi libro fueron escritos en 1800, habría dicho, “Los Estados Unidos están divididos.” Esta pequeña diferencia uso ha sido durante mucho tiempo una fascinación para los historiadores, ya que sugiere que hubo un punto en que estados Unidos dejó de pensar en sí mismo como una colección de estados y comenzó a pensar de sí mismo como una nación.

Así que cuando sucedió esto? Historiadores, *Uncharted* nos informa, no han sido nunca seguro, ya que no ha habido manera sistemática para probarlo. Pero muchos han sospechado durante mucho tiempo fue la causa de la Guerra Civil. De hecho, James McPherson, ex

presidente de la American Historical Association y un ganador del premio Pulitzer, señaló sin rodeos: “La guerra marcó una transición de los Estados Unidos a un sustantivo singular”.

Pero resulta McPherson estaba mal. Google dio N-gramas Aiden y Michel manera sistemática para comprobar esto. Podían ver la frecuencia con libros estadounidense usó la frase “Los Estados Unidos son. . .” Frente a “Los Estados Unidos es. . .” Para cada año en la historia del país. La transformación fue más gradual y no aceleró hasta bien después de la Guerra Civil.



Quince años después de la Guerra Civil, había aún más usos de “Los Estados Unidos están. . .” Que “Estados Unidos es. . .”, Que muestra el país estaba todavía dividida lingüísticamente. Las victorias militares ocurren más rápido que los cambios en la forma de pensar.

Esto en cuanto a cómo se une un país. ¿Cómo un hombre y una mujer se unen? Las palabras pueden ayudar, también.

Por ejemplo, podemos predecir si un hombre y una mujer irán a una segunda cita en base a la forma en que hablan en la primera cita.

Esto fue demostrado por un equipo interdisciplinario de Stanford y Northwestern científicos: Daniel McFarland, Dan Jurafsky, y Craig Rawlings. Los investigadores estudiaron a cientos de personas que se citan velocidad heterosexuales y trataron de determinar lo que predice si van a sentir una conexión y quieren una segunda fecha.

Primero se utilizaron datos tradicional. Le pidieron a personas que se citan para su altura, peso y aficiones y se prueban cómo estos factores correlacionados con alguien reportar una chispa de interés romántico. Las mujeres, en promedio, prefieren a los hombres que son más altas y comparten sus aficiones; los hombres, en promedio, prefieren a las mujeres que son más delgados y comparten sus aficiones. Nada nuevo.

Pero los científicos también recogieron un nuevo tipo de datos. Instruyeron a las personas que se citan a tomar grabadoras con ellos. a continuación, se digitalizaron las grabaciones de las fechas. Los científicos pudieron así codificar las palabras utilizadas, la presencia de la risa, y el tono de voz. Se podría poner a prueba tanto como los hombres y las mujeres señalaron que estaban interesados y cómo los socios ganaron ese interés.

Entonces, ¿Por qué los datos lingüísticos nos dicen? En primer lugar, cómo un hombre o una mujer transmite que él o ella está interesada. Una de las formas en que un hombre de señales que



le atrae es obvia: se ríe de los chistes de una mujer. Otra es menos obvio: al hablar, que limita el alcance de su lanzamiento. Hay investigaciones que sugieren una voz monótona se ve a menudo por las mujeres como masculinas, lo que implica que los hombres, tal vez inconscientemente, exageran su masculinidad cuando les gusta una mujer.

Los científicos descubrieron que una mujer señala su interés mediante la variación de su lanzamiento, hablando en voz más baja, y tomar turnos más cortos hablando. También hay importantes indicios sobre el interés de una mujer basada en las palabras particulares que utiliza. Una mujer es poco probable que sea interesada cuando ella usa palabras de cobertura y frases tales como “probablemente” o “supongo.”

Amigos, si una mujer está cubriendo sus declaraciones sobre cualquier tema, si ella “sorta” le gusta su bebida o “algo” se siente frío o “probablemente” tendrá otro entremés-usted puede apostar que ella es “sorta” “un poco” “probablemente” no en ti.

Una mujer es probable que estén interesados cuando habla de sí misma. Resulta que, para un hombre que mira a conectar, la palabra más bella se puede escuchar de la boca de una mujer puede ser “I”: es una señal de que se siente cómodo. Una mujer también es probable que estén interesados si se utiliza frases auto-marcado como “Ya saben?” Y “quiero decir.” ¿Por qué? Los científicos observaron que estas frases invitan a la atención del oyente. Son amables y cálidos y sugieren una persona está buscando para conectar, ya sabes lo que quiero decir?

Ahora, ¿cómo pueden los hombres y las mujeres se comunican con el fin de conseguir una cita interesado en ellos? Los datos nos dicen que hay un montón de maneras que un hombre puede hablar de aumentar las posibilidades de que una mujer le gustan. Las mujeres como hombres que siguen su ejemplo. Quizás no sea sorprendente, una mujer tiene más probabilidades de reportar una conexión si un hombre se ríe de sus chistes y mantiene la conversación sobre temas que introduce en lugar de cambiar constantemente el tema a los que él quiere hablar.\*Las mujeres también les gustan los hombres que expresan apoyo y simpatía. Si un hombre dice: “Eso es increíble!” O “Eso es realmente bueno”, es significativamente más propensos a reportar una conexión a una mujer. Del mismo modo, si se utiliza frases como “Eso es duro” o “Usted debe estar triste.”

Para las mujeres, hay algunas malas noticias aquí, ya que los datos parecen confirmar una verdad desagradable en hombres. Conversación sólo desempeña un pequeño papel en la forma en que responden a las mujeres. Apariencia física triunfa sobre todas las cosas para predecir si un hombre informa de una conexión. Dicho esto, hay una palabra que una mujer puede utilizar para al menos mejorar ligeramente las probabilidades que un hombre le gusta y lo que es uno que ya hemos discutido: “I.” Los hombres son más propensos a hacer clic con una mujer que habla de sí misma . Y como se ha señalado anteriormente, una mujer es también más propensos a reportar una conexión después de una fecha en la que habla de sí misma. Por lo tanto, es una gran señal, en una primera cita, si hay un debate de fondo sobre la mujer. La mujer señala su comodidad y probablemente es consciente de que el hombre no está acaparando la conversación. Y el hombre le gusta que la mujer se está abriendo.

Por último, hay un claro indicador de problemas en una transcripción fecha: un signo de interrogación. Si hay un montón de preguntas sobre una fecha, es menos probable que tanto el hombre como la mujer se reportarán una conexión. Esto parece contrario a la intuición; se podría pensar que las preguntas son una señal de interés. Pero no es así en la primera cita. En una primera cita, la mayoría de las preguntas son signos de aburrimiento. “¿Cuáles son sus aficiones?” “¿Cuántos hermanos y hermanas tienes?” Estos son los tipos de cosas que la gente dice cuando los puestos de conversación. Una gran primera fecha puede incluir una sola pregunta al final: “¿Quieres salir conmigo otra vez” Si esta es la única pregunta en la fecha, la respuesta es probable que sea “Sí”.

Y los hombres y mujeres no sólo hablan de manera diferente cuando están tratando de atraer a la otra. Se habla de manera diferente en general.

Un equipo de psicólogos analizaron las palabras usadas en cientos de miles de mensajes de Facebook. Midieron la frecuencia con cada palabra es utilizada por hombres y mujeres. Podrían entonces declarar que son las palabras más masculinas y femeninas más en el idioma Inglés.

Muchas de estas preferencias de palabras, por desgracia, eran evidentes. Por ejemplo, las mujeres hablan de “comercial” y “mi pelo” con mucha más frecuencia que los hombres. Los hombres hablan de “fútbol” y “Xbox” con mucha más frecuencia que las mujeres. Es probable que no necesita un equipo de psicólogos analizar grandes volúmenes de datos que le diga eso.

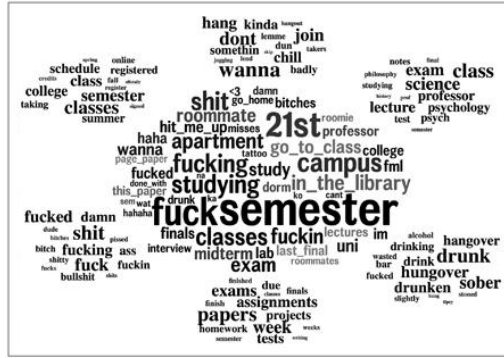
Algunos de los hallazgos, sin embargo, eran más interesantes. Las mujeres usan la palabra “mañana” con mucha más frecuencia que los hombres, tal vez porque los hombres no son tan grandes en el pensamiento a continuación. La adición de la letra “o” de la palabra “por lo que” es uno de los rasgos lingüísticos más femeninas. Entre las palabras más utilizadas en forma desproporcionada por las mujeres son “Soo”, “tan”, “tan”, “increíblemente”, y “taaaaaaaaaaaaaaaaaan”.

Tal vez fue mi exposición de los niños a las mujeres que no tenían miedo a tirar el ocasional f-bomba. Pero siempre pensé maldición era un rasgo de igualdad de oportunidades. No tan. Entre las palabras que se usan con más frecuencia por los hombres que en las mujeres son “mierda”, “mierda”, “folla”, “mierda”, “puta” y “hijos de puta”.

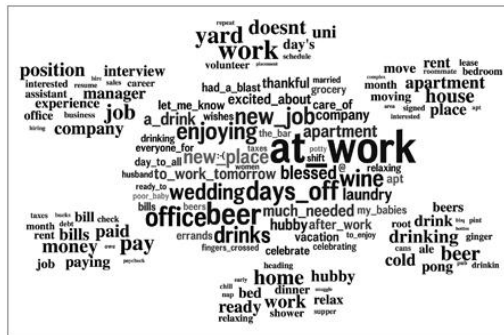
A continuación se muestran nubes de palabras palabras que se usan en su mayoría por hombres y los que se utilizan en su mayoría por mujeres. El más grande parece una palabra, el uso más que de palabra se inclina hacia ese género.

Los machos





23- a 29 años de edad



De 30 a 65 años de edad



Una nueva y poderosa herramienta para el análisis de texto es algo que se llama análisis de los sentimientos. Ahora, los científicos pueden estimar lo feliz o triste un pasaje particular de texto es.

¿Cómo? Equipos de científicos han pedido a un gran número de personas para codificar decenas de miles de palabras en el idioma Inglés como positivos o negativos. Las palabras más positivas, de acuerdo con esta metodología, incluyen “feliz”, “amor”, y “impresionante.” Las palabras más negativos incluyen “triste”, “muerte” y “depresión”. De este modo, han construido un índice de la el estado de ánimo de un enorme conjunto de palabras.

Usando este índice, se puede medir el estado de ánimo promedio de palabras en un pasaje de texto. Si alguien escribe “Estoy feliz y en el amor y el sentimiento, impresionante” análisis de los sentimientos que codificar ese texto como extremadamente feliz. Si alguien escribe “estoy triste pensando en toda la muerte y la depresión del mundo,” análisis de sentimientos sería codificar ese texto como extremadamente triste. Otros fragmentos de texto serían en algún punto intermedio.

Entonces, ¿qué se puede aprender cuando codificar el estado de ánimo de texto? Facebook científicos de datos han demostrado una posibilidad emocionante. Pueden estimar felicidad nacional bruta de un país cada día. Si los mensajes de estado de las personas tienden a ser positivo, el país se asume feliz para el día. Si tienden a ser negativo, el país se asume triste para el día.

Entre los hallazgos de los datos de Facebook de los científicos: la Navidad es uno de los días más felices del año. Ahora, yo era escéptico de este análisis, y soy un poco escéptico de todo este proyecto. En general, creo que mucha gente está en secreto triste en la Navidad porque se sienten solos o luchar con su familia. De manera más general, tiendo a no confiar en las actualizaciones de estado de Facebook, por razones que discutiremos en el siguiente capítulo, a saber, nuestra propensión a mentir acerca de nuestras vidas en las redes sociales.

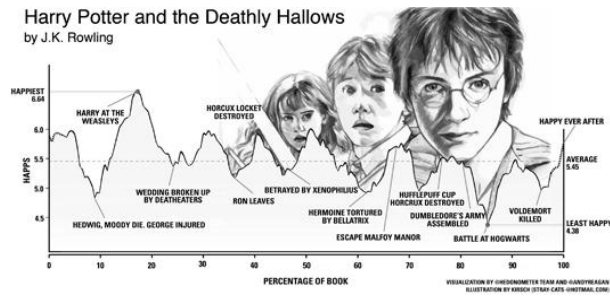
Si está solo y triste en la Navidad, es lo que realmente quiere preocuparse por todos sus amigos mediante la publicación acerca de lo infeliz que eres? Sospecho que hay muchas personas que pasan una Navidad sin alegría que todavía Colocar en Facebook sobre lo agradecidos que están por su “increíble vida maravillosa, impresionante, y feliz.” A continuación, obtener codifican como un aumento sustancial de felicidad nacional bruta de los Estados Unidos. Si vamos a codificar realmente felicidad nacional bruta, debemos utilizar más fuentes que sólo las actualizaciones de estado de Facebook.

Dicho esto, el hallazgo de que la Navidad es, en conjunto, un motivo de alegría parece legítimamente para ser verdad. Google busca para la depresión y las encuestas de Gallup también nos dice que la Navidad es uno de los días más felices del año. Y, al contrario de un mito urbano, suicidios caen durante las fiestas. Incluso si hay algunas personas tristes y solitarios en Navidad, hay muchos otros más alegres.

En estos días, cuando las personas se sientan a leer, la mayoría de las veces es para examinar las actualizaciones de estado en Facebook. Sin embargo, hubo un tiempo, no hace mucho tiempo, los seres humanos leen cuentos, a veces en los libros. El análisis de sentimientos nos puede enseñar mucho, también.

Un equipo de científicos, dirigido por Andy Reagan, ahora en la Universidad de California en Berkeley School de la información, descargar el texto de miles de libros y guiones de cine. A continuación, podría codificar cómo era feliz o triste cada punto de la historia.

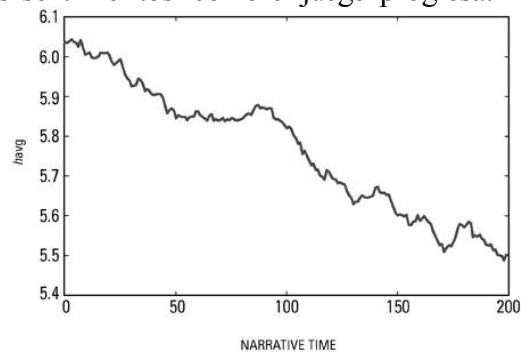
Consideremos, por ejemplo, el libro de Harry Potter y las reliquias de la muerte. A continuación, a partir de ese equipo de científicos, es cómo el estado de ánimo de la historia cambia, junto con una descripción de los puntos de la trama clave.



Tenga en cuenta que las muchas subidas y bajadas en el estado de ánimo que el análisis de los sentimientos detecta corresponden a eventos clave.

La mayoría de las historias tienen estructuras más simples. Tomemos, por ejemplo, la tragedia de Shakespeare rey Juan. En este juego, nada sale bien. Se pidió al Rey Juan de Inglaterra a renunciar a su trono. Él es excomulgado por desobedecer al Papa. Estalla la guerra. Su sobrino muere, tal vez por el suicidio. Otras personas mueren. Por último, John es envenenado por un monje descontento.

Y aquí es el análisis de los sentimientos como el juego progresa.

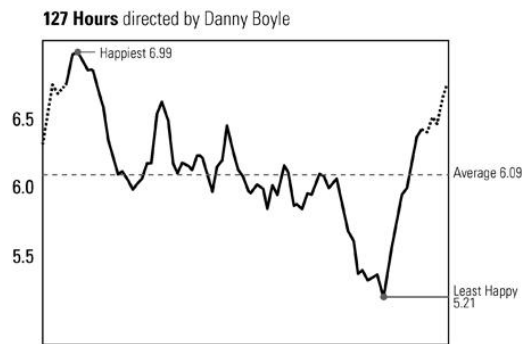


En otras palabras, sólo de las palabras, el equipo fue capaz de detectar que las cosas van de mal en peor a peor.

O considerar la película 127 horas. Un resumen de la trama básica de esta película es la siguiente:

Un alpinista va al Parque Nacional de Canyonlands de Utah para ir de excursión. Se hace amigo de otros excursionistas, pero a continuación formas de piezas con ellos. De repente, él se desliza y golpea una roca suelta, que atrapa la mano y la muñeca. , Intenta varias fugas, pero cada uno de ellos falla. Él se deprime. Por último, se amputa el brazo y se escapa. Él se casa, se inicia una familia, y sigue subiendo, aunque ahora se asegura de dejar una nota cada vez que se apaga.

Y aquí es el análisis de los sentimientos como avanza la película, de nuevo por el equipo de científicos de Reagan.



Entonces, ¿qué podemos aprender del estado de ánimo de miles de estas historias?

Los científicos de la computación encontraron que un gran porcentaje de historias encajan en una de las seis estructuras relativamente simples. Están, tomando prestado un gráfico del equipo de Reagan:

Rags a Riches (Rise)

Riches a Rags (caída)

El hombre en un agujero (caída, después se levantan)

Ícaro (subida, entonces caer)

Cinderella (subida, a continuación, caen, a continuación, se elevan)

Edipo (caída, a continuación, se levanta, luego caer)

Puede haber pequeños giros y NOT pone capturado por este simple esquema. Por ejemplo, 127 horas se ubica como un hombre en un agujero historia, a pesar de que hay momentos en el camino hacia abajo cuando los sentimientos mejoran temporalmente. La gran estructura general de la mayoría de las historias encaja en una de las seis categorías. Harry Potter y las reliquias de la muerte es una excepción.

Hay un montón de preguntas adicionales que podríamos contestar. Por ejemplo, ¿tiene la estructura de historias cambiado a través del tiempo? Han conseguido historias más complicado a través de los años? Diferencian las culturas en los tipos de historias que cuentan? ¿Qué tipos de historias hacen gente como la mayoría? ¿Los diferentes estructuras de las historias atraen a los hombres y las mujeres? ¿Qué pasa con las personas de diferentes países?

En última instancia, el texto como datos nos puede dar una visión sin precedentes en lo que el público realmente quiere, que pueden ser diferentes de lo que los autores o ejecutivos creen que quieren. Ya hay algunos indicios que apuntan en esta dirección.

Considere un estudio realizado por dos profesores de la Escuela Wharton, Jonah Berger y Katherine L. Iyengar, en lo que se comparten tipos de historias. Se probaron si las historias positivas o negativas historias eran más propensos a hacer lista más enviado por correo

electrónico, el New York Times'. Se descargan todos los artículos veces durante un período de tres meses. Utilizando análisis de los sentimientos, los profesores codifican el estado de ánimo de los artículos. Los ejemplos de historias positivas incluyen “con los ojos abiertos nuevas llegadas Enamorarse con la ciudad” y las historias como “rumores Web atadas al suicidio actriz coreana” y ‘Alemania: Alimentador del bebé del oso polar muere’ “Tony Award para la Filantropía.” Demostró , no sorprendentemente, a ser negativo.

Los profesores también tenían información acerca de donde se colocó la historia. Fue en la página principal? En la parte superior derecha? La parte superior izquierda? Y tenían información sobre la historia cuando salió. Tarde martes por la noche? Lunes por la mañana?

Se podría comparar dos artículos: uno de ellos positivos, uno de ellos negativos, que apareció en un lugar similar en el sitio Times y salió en un tiempo similar y ver cuál era más probable que sea enviado por correo electrónico.

Entonces, ¿qué se comparte, artículos positiva o negativa?

artículos positivos. Como concluyen los autores, “El contenido es más probable que se convierta viral más positivo que es.”

Tenga en cuenta que esto parecería que contraste con la sabiduría convencional periodística que las personas se sienten atraídos por las historias violentas y catastróficas. Puede ser cierto que los medios de comunicación dan a la gente un montón de historias oscuras. Hay algo en el dicho sala de redacción “Si sangra, conduce.” El estudio de los profesores de Wharton, sin embargo, sugiere que las personas pueden realmente quieren historias más alegres. Se puede sugerir un nuevo adagio: “Si sonríe, se enviará por correo electrónico,” a pesar de que en realidad no rima.

Tanto para el texto triste y feliz. ¿Cómo se puede averiguar qué palabras son liberales o conservadores? Y ¿qué nos dice esto acerca de los medios de comunicación modernos? Esto es un poco más complicado, lo que nos lleva de nuevo a Gentzkow y Shapiro. Recuerde, ellos fueron los economistas que veían el matrimonio homosexual describen diferentes maneras en dos periódicos diferentes y se preguntaban si podían usar el lenguaje para descubrir sesgo político.

La primera cosa que estos dos jóvenes académicos ambiciosos hicieron fue examinar las transcripciones de las Actas del Congreso. Desde este disco ya estaba digitalizada, podrían descarga cada palabra utilizada por cada congresista demócrata en 2005 y cada palabra usada por cada congresista republicano en 2005. Entonces podrían ver si ciertas frases fueron significativamente más propensos a ser utilizado por los demócratas o los republicanos.

Algunos eran de hecho. Aquí hay algunos ejemplos de cada categoría.

<b>Frases de uso mucho más por la DEMÓCRATAS</b>	<b>Frases de uso mucho más por la REPUBLICANOS</b>
impuesto de propiedad	impuesto a la muerte
Privatizar la seguridad social	la reforma de la seguridad social
Rosa Parks	Saddam Hussein



Derechos de los trabajadores	los derechos de propiedad privada
Gente pobre	Gastos gubernamentales

¿Qué explica estas diferencias en el lenguaje?

A veces, los demócratas y los republicanos utilizan diferentes frases para describir el mismo concepto. En 2005, los republicanos trataron de cortar el impuesto federal sobre la herencia. Ellos tienden a describirlo como un “impuesto de la muerte” (que suena como una imposición sobre el recientemente fallecido). Demócratas lo describió como un “impuesto de estado” (que suena como un impuesto sobre los ricos). Del mismo modo, los republicanos trataron de mover la Seguridad Social en cuentas individuales de retiro. Para los republicanos, esto era una “reforma”. Para los demócratas, se trataba de un sonido más peligroso “privatización”.

A veces las diferencias de idioma son una cuestión de énfasis. Republicanos y demócratas, presumiblemente, ambos tienen un gran respeto por Rosa Parks, el héroe de los derechos civiles. Pero los demócratas habló de su mayor frecuencia. Del mismo modo, los demócratas y los republicanos, presumiblemente, ambos piensan que Saddam Hussein, el ex líder de Irak, fue un malvado dictador. Pero los republicanos lo mencionado varias veces en su intento de justificar la guerra de Irak. Del mismo modo, “los derechos de los trabajadores” y la preocupación por ‘gente pobre’ son principios básicos del Partido Demócrata. “Derechos de propiedad privada” y el corte “gasto público” son principios fundamentales de los republicanos.

Y estas diferencias en el uso del lenguaje son sustanciales. Por ejemplo, en 2005, los republicanos del Congreso utilizó la frase “impuesto de la muerte” 365 veces y “impuesto de estado” sólo 46 veces. Para los demócratas del Congreso, el patrón se invirtió. Ellos usaron la frase “Impuesto de Transferencia” sólo 35 veces y “impuesto de sucesión” 195 veces.

Y si estas palabras nos pueden decir si un congresista es un demócrata o un republicano, los estudiosos se dieron cuenta, también podrían decirnos si un periódico se inclina hacia la izquierda o hacia la derecha. Al igual que los congresistas republicanos podrían ser más propensos a utilizar la frase “impuesto de la muerte” para convencer a la gente para oponerse a ella, periódicos conservadores podrían hacer lo mismo. La relativamente liberal Washington Post usó la frase “impuesto de estado” 13.7 veces más frecuentemente que antes la frase “impuesto a la muerte.” El conservador Washington Times utiliza “impuesto de la muerte” y “impuesto de estado” casi la misma cantidad.

Gracias a las maravillas de internet, Gentzkow y Shapiro podría analizar el lenguaje utilizado en un gran número de periódicos de la nación. Los investigadores utilizaron dos sitios web, newslibrary.com y proquest.com, que en conjunto tenían digitalizadas 433 periódicos. Se contaron entonces con qué frecuencia se utilizaron mil frases carga política en los periódicos con el fin de medir la inclinación política de los periódicos. El periódico más liberal, con esta medida, resultó ser el Philadelphia Daily News; los más conservadores: el Billings (Montana) Gaceta.

Cuando se tiene la primera medida integral de sesgo de los medios de comunicación para una amplia franja de medios de este tipo, puede responder quizás la pregunta más importante acerca de la prensa: ¿por qué algunas publicaciones inclinarse hacia la izquierda y otros no?

Los economistas alojados rápidamente en un factor clave: la política de un área determinada. Si un área es en general liberal, como Filadelfia y Detroit son, el periódico dominante tiende a ser liberal. Si un área es más conservadora, como lo son Billings y Amarillo, Texas, el papel dominante que tiende a ser conservador. En otras palabras, la evidencia sugiere fuertemente que los periódicos se inclinan a dar a sus lectores lo que quieren.

Se podría pensar que el propietario de un documento podría tener alguna influencia sobre la inclinación de su cobertura, pero como regla general, que es dueño de un documento tiene un efecto menor de lo que pensamos sobre su sesgo político. Tenga en cuenta lo que sucede cuando la misma persona o compañía es propietaria de papeles en diferentes mercados. Considere el New York Times Company. Posee lo Gentzkow y Shapiro encuentran para ser la tendencia liberal New York Times, con sede en la ciudad de Nueva York, donde aproximadamente el 70 por ciento de la población es democrático. También poseía, en el momento del estudio, el conservador de tendencia, por su medida, Spartanburg Herald-Journal, en Spartanburg, Carolina del Sur, donde aproximadamente el 70 por ciento de la población es republicano. Hay excepciones, por supuesto: de Rupert Murdoch, News Corporation posee lo que casi todo el mundo encontraría a ser la conservadora Nueva York Post. Pero en general,

El estudio tiene un profundo impacto en la forma en que pensamos acerca de los medios de comunicación. Mucha gente, especialmente los marxistas, han visto como el periodismo estadounidense controlado por la gente rica o corporaciones con el objetivo de influir en las masas, tal vez para empujar a la gente hacia sus puntos de vista políticos. Gentzkow de papel y de Shapiro sugiere, sin embargo, que esto no es la motivación predominante de los propietarios. Los propietarios de la prensa de Estados Unidos, en cambio, están dando principalmente a las masas lo que quieren por lo que los propietarios pueden llegar a ser aún más rico.

Ah, y una última pregunta, una pregunta grande, polémico, y quizás aún más provocativa. ¿Los medios de comunicación estadounidenses, en promedio, inclinada hacia la izquierda o la derecha? Son los medios de comunicación en la media liberal o conservador?

Gentzkow y Shapiro encontraron que los periódicos inclinación a la izquierda. El diario promedio es más similar, en las palabras que utiliza, a un congresista demócrata de lo que es un congresista republicano.

“Aha!” Lectores conservadores pueden estar a punto de gritar, “¡Te lo dije!” Muchos conservadores han larga sospechosos periódicos han sido parciales para tratar de manipular a las masas para apoyar los puntos de vista de izquierda.

No es así, dicen los autores. De hecho, la tendencia liberal está bien calibrado a lo que quieren los lectores de periódicos. La lectura de periódicos, en promedio, se inclina un poco hacia la izquierda. (Tienen datos sobre eso.) Y los periódicos, en promedio, se inclinan un poco más que dar a sus lectores los puntos de vista que exigen.

No hay una gran conspiración. No es sólo el capitalismo.

Los medios de comunicación, Gentzkow y los resultados de Shapiro implican, a menudo operan como cualquier otra industria en el planeta. Al igual que los supermercados averiguar lo que las personas quieren helado y llenar sus estantes con ella, periódicos averiguar qué puntos de

vista y la gente quiere llenar sus páginas con él. “Es sólo un negocio”, Shapiro me dijo. Eso es lo que se puede aprender cuando usted analiza y cuantifica asuntos tan retorcido como noticias, análisis y opinión en sus partes componentes: palabras.

### Pictures como DATOS

---

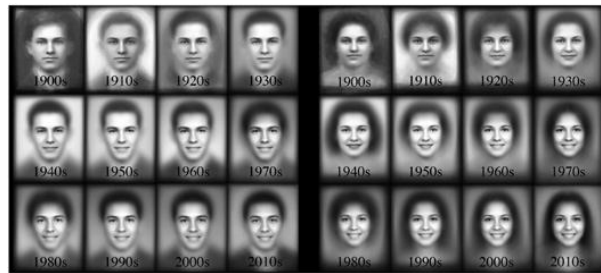
Tradicionalmente, cuando los académicos o empresarios querían los datos, se realizaron encuestas. Los datos se produjeron perfectamente formados, elaborado a partir de los números o casillas marcadas en cuestionarios. Esto ya no es el caso. Los días de datos estructurados, limpio, simple, basados en encuestas se han terminado. En esta nueva era, las huellas que dejamos desordenado a medida que avanzamos por la vida se están convirtiendo en la principal fuente de datos.

Como ya hemos visto, las palabras son datos. Clics son datos. Los enlaces son datos. Los errores tipográficos son datos. Los plátanos de los sueños son los datos. El tono de voz es de datos. Las sibilancias son datos. Los latidos del corazón son los datos. El tamaño del bazo es de datos. Las búsquedas son, argumento, los datos más reveladores.

Imágenes, resulta que son los datos, también.

Así como las palabras, que estaban confinados a los libros y publicaciones periódicas en los estantes polvorientos, ya han sido digitalizados, imágenes han sido liberados de álbumes y cajas de cartón. Ellos también se han transformado en trozos y se liberan en la nube. Y como texto nos puede dar lecciones mostrando historia de EE.UU., por ejemplo, las formas cambiantes gente ha hablado imágenes nos pueden dar lecciones mostrando historia de EE.UU., por ejemplo, las formas cambiantes de personas han planteado.

Considere un ingenioso estudio por un equipo de cuatro científicos de la computación en Brown y Berkeley. Se aprovecharon de un desarrollo digital de la era ordenada: muchas escuelas secundarias han explorado sus anuarios históricos y ponerlos a disposición en línea hecho. A través de Internet, los investigadores encontraron 949 anuarios escaneadas de las escuelas secundarias de Estados Unidos que abarca los años 1905-2013. Esto incluyó decenas de miles de retratos de alto nivel. El uso de software de ordenador, que fueron capaces de crear una cara “promedio” de las imágenes de cada década. En otras palabras, se podría averiguar la ubicación normal y la configuración de la nariz, los ojos, los labios y el pelo de la gente. Estas son las caras medias de todo el siglo pasado, además, desglosados por género:



Darse cuenta de nada? -Americanos y en particular las mujeres-empezó a sonreír. Se fueron de casi cara de piedra en el inicio del siglo XX hicieron incidir un haz para el final.

Entonces ¿por qué el cambio? Hicieron los estadounidenses reciben más feliz?

Nop. Otros estudiosos han ayudado a responder a esta pregunta. La razón es, al menos para mí, fascinante. Cuando las fotografías fueron inventadas por primera vez, la gente pensaba de ellos como pinturas. No había nada más que compararlos. Por lo tanto, sujetos en las fotografías copiadas sujetos en pinturas. Y puesto que la gente sentada en los retratos no pudo contener una sonrisa por las muchas horas de la pintura tomó, adoptaron una mirada seria. Sujetos en las fotografías adoptaron el mismo aspecto.

Lo que finalmente los llevó a cambiar? Negocio, el beneficio y la comercialización, por supuesto. A mediados del siglo XX, Kodak, la compañía de la película y la cámara, se vio frustrado por el número limitado de imágenes de personas estaban tomando e ideó una estrategia para llegar a tomar más. la publicidad de Kodak comenzó a asociar fotos con la felicidad. El objetivo era que la gente en el hábito de tomar una foto cada vez que quería mostrar a los demás lo bien que una que estaban teniendo. Todos esos sonrientes fotos del anuario son el resultado de esa exitosa campaña (como lo son la mayoría de las fotos que se ven en Facebook e Instagram hoy en día).

Pero las fotos como los datos pueden decirnos mucho más que cuando los estudiantes de secundaria empezaron a decir “queso”. Sorprendentemente, las imágenes pueden ser capaces de decirnos cómo la economía está haciendo.

Considere un artículo académico provocativamente titulado: “Midiendo el crecimiento económico del espacio exterior.” Cuando un documento tiene un título como ese, se puede apostar que voy a leerlo. Los autores de este trabajo-J. Vernon Henderson, Adam Storeygard, y David N. Weil-empezar señalando que en muchos países en desarrollo, las medidas existentes del producto interno bruto (PIB) son ineficientes. Esto se debe a que gran parte de la actividad económica ocurren fuera de los libros, y las agencias del gobierno destinado a medir la producción económica de recursos limitados.

idea poco convencional de los autores? Ellos podrían ayudar a medir el PIB basado en la cantidad de luz que hay en estos países por la noche. Consiguieron que la información de las fotografías tomadas por un satélite de la Fuerza Aérea de los Estados Unidos que rodea la tierra catorce veces por día.

¿Por qué podría iluminar la noche sea una buena medida del PIB? Así, en las partes más pobres del mundo, la gente lucha para pagar por la electricidad. Y como resultado, cuando las condiciones económicas son malas, los hogares y aldeas reducirán drásticamente la cantidad de luz que se dejan por la noche.

luz de la noche se redujo drásticamente en Indonesia durante la crisis financiera asiática de 1998. En Corea del Sur, luz de noche aumentó 72 ciento entre 1992 y 2008, que corresponde a una extraordinariamente fuerte desempeño económico durante este período. En Corea del Norte, durante el mismo tiempo, la luz nocturna de hecho disminuyó, lo que corresponde a un pésimo desempeño económico durante este tiempo.

En 1998, en el sur de Madagascar, se descubrió una gran acumulación de rubíes y zafiros. La ciudad de Ilakaka pasó de poco más de una parada de camiones a un importante centro de comercio. No había prácticamente ninguna luz en la noche Ilakaka antes de 1998. En los próximos cinco años, hubo una explosión de luz en la noche.

Los autores admiten que sus datos luz de noche está lejos de ser una medida perfecta de la producción económica. Usted definitivamente no se puede saber exactamente cómo una economía está haciendo precisamente a partir de la cantidad de luz satélites pueden recoger en la noche. Los autores no recomiendan el uso de esta medida en todos los países desarrollados, tales como Estados Unidos, donde los datos económicos existentes es más preciso. Y para ser justos, incluso en los países en desarrollo, se encuentran con que la luz de noche es solamente tan útil como las medidas oficiales. Sin embargo, la combinación de ambos los datos del gobierno con los datos erróneos luz de noche imperfecta da una mejor estimación de cualquiera de las fuentes por sí sola podría proporcionar. Puede, en otras palabras, mejorar su comprensión de las economías en desarrollo utilizando las fotografías tomadas desde el espacio exterior.

Joseph Reisinger, un doctorado de la informática con una voz suave, comparte la frustración de los autores de luz la noche con los conjuntos de datos existentes en las economías en los países en desarrollo. En abril de 2014, notas Reisinger, Nigeria actualizan su estimación de PIB, teniendo en cuenta los nuevos sectores que pueden haber perdido en las estimaciones anteriores. Su PIB estimado era ahora el 90 por ciento más alto.

“Son la mayor economía de África”, dijo Reisinger, alzando la voz lentamente. “No sabemos siquiera lo más básico que nos gustaría saber de este país.”

Quería encontrar una manera de conseguir una mirada más aguda en los resultados económicos. Su solución es un gran ejemplo de cómo re-imaginar lo que constituye datos y el valor de hacerlo.

Reisinger fundó una compañía, Local, que emplea a un grupo de trabajadores en los países en desarrollo, armados con teléfonos inteligentes. El trabajo de los empleados? Para tomar imágenes de idas y venidas que interesantes podrían tener importancia económica.

Los empleados podrían obtener instantáneas fuera de las estaciones de servicio o de cajas de fruta en los supermercados. Se toman fotos de los mismos lugares una y otra vez. Las imágenes se envían de nuevo a Premisa, cuyo grupo de empleados por ordenador segundo científicos-convertir las fotos en datos. Los analistas de la compañía pueden codificar todo, desde la

longitud de las líneas en las estaciones de servicio a cuántas manzanas están disponibles en un supermercado a la madurez de estas manzanas con el precio que aparece en las manzanas bin. Sobre la base de fotografías de todo tipo de actividad, Premisa puede comenzar a armar las estimaciones de la producción económica y la inflación. En los países en desarrollo, las largas colas en las estaciones de servicio son un indicador de problemas económicos. También lo son las manzanas inmaduras o no disponibles. Premisa está en el terreno imágenes de vajillas ayudaron a descubrir la inflación comida allí en 2011 y la deflación de alimentos en 2012,

Premisa vende esta información a los bancos o fondos de cobertura y también colabora con el Banco Mundial.

Al igual que muchas buenas ideas, la premisa de es un regalo que sigue dando. El Banco Mundial fue recientemente interesado en el tamaño de la economía subterránea de cigarrillos en las Filipinas. En particular, querían conocer los efectos de los recientes esfuerzos del gobierno, que incluyeron redadas al azar, para acabar con los fabricantes que producen los cigarrillos sin tener que pagar un impuesto. ingeniosa idea de la premisa? Tomar fotos de las cajetillas de cigarrillos se ven en la calle. Ver cuántos de ellos tienen sellos fiscales, que todos los cigarrillos legítimas. Han encontrado que esta parte de la economía sumergida, aunque grande en 2015, consiguió significativamente menor en 2016. Los esfuerzos del gobierno trabajaron, a pesar de ver algo por lo general tan oculto-cigarrillos ilegales requeridos por los nuevos datos.

Como hemos visto, lo que constituye datos han sido violentamente reimagined en la era digital y una gran cantidad de puntos de vista se han encontrado en esta nueva información. Aprender lo que impulsa el sesgo de los medios de comunicación, lo que hace una buena primera fecha, y cómo las economías en desarrollo están haciendo en realidad es sólo el principio.

No es casual que un montón de dinero también se ha hecho a partir de tales datos nuevos, empezando por los Sres. Decenas de Brin y la página de de mil millones. Joseph Reisinger no ha hecho mal a sí mismo. Los observadores estiman que Premisa ahora está haciendo decenas de millones de dólares en ingresos anuales. Los inversores recientemente vertieron \$ 50 millones a la empresa. Esto significa que algunos inversores consideran Premisa entre las empresas más valiosas en el mundo, principalmente en el negocio de la toma y la venta de fotos, en la misma liga que Playboy.

No es, en otras palabras, el valor de tallas grandes, para los estudiosos y empresarios por igual, en la utilización de todos los nuevos tipos de datos ya está disponible, al pensar en términos generales acerca de lo que cuenta como datos. En estos días, un científico de datos no deben limitar a sí misma a una visión estrecha o tradicional de los datos. En estos días, las fotografías de las líneas de supermercados son datos valiosos. La plenitud de los contenedores de los supermercados es de datos. La madurez de las manzanas es de datos. Fotos del espacio exterior son datos. La curvatura de los labios es de datos. Todo está datos!

Y con todos estos nuevos datos, podemos ver por fin a través de las mentiras de la gente.

## DIGITAL suero de la verdad

**mi**entiras odo el mundo.

La gente miente sobre la cantidad de bebidas que había en el camino a casa. Mienten sobre la frecuencia con que van al gimnasio, cuánto cuestan esos zapatos nuevos, ya sea que leen ese libro. Ellos llamar a los enfermos cuando no lo son. Dicen que van a estar en contacto cuando no lo harán. Dicen que no se trata de que cuando es. Dicen que te amo cuando no lo hacen. Dicen que son felices mientras que en los vertederos. Ellos dicen que les gustan las mujeres, cuando en realidad les gustan los hombres.

La gente miente a los amigos. Se encuentran a Jefes. Se encuentran a los niños. Se encuentran a los padres. Se encuentran a los médicos. Se encuentran a maridos. Se encuentran a las esposas. Se encuentran a sí mismos.

Y absolutamente seguro de mentira con las encuestas.

Aquí está mi breve encuesta para usted:

¿Alguna vez has hecho trampa en un examen? \_\_\_\_\_

¿Alguna vez has fantaseado con matar a alguien? \_\_\_\_\_

¿Estaba tentado a mentir? Muchas personas no reportan comportamientos embarazosos y reflexiones sobre las encuestas. Ellos quieren verse bien, a pesar de que la mayoría de las encuestas son anónimas. Esto se denomina sesgo de deseabilidad social.

Un papel importante en 1950 proporcionó una poderosa evidencia de cómo las encuestas pueden ser víctimas de este sesgo. Los investigadores recogieron los datos, a partir de fuentes oficiales, a los residentes de Denver: ¿qué porcentaje de ellos votaron, dieron a la caridad, y era dueño de una tarjeta de la biblioteca. Se les realizó una encuesta a los residentes para ver si los porcentajes se correspondería. Los resultados fueron, en su momento, impactante. Lo que los residentes informaron a las encuestas fue muy diferente a los datos de los investigadores habían recogido. A pesar de que nadie dio sus nombres, la gente, en gran número, exageraron su estado de registro de votantes, el comportamiento del voto, y las donaciones benéficas.

	REPORTED ON SURVEY	OFFICIAL COUNT
Registered to vote	83%	69%
Voted in last presidential election	73%	61%
Voted in last mayoral election	63%	36%
Have a library card	20%	13%
Gave to a recent Community Chest charitable drive	67%	33%

¿Ha cambiado algo en sesenta y cinco años? En la era de Internet, no ser propietario de una tarjeta de la biblioteca ya no es embarazoso. Pero, mientras que lo que es embarazoso o deseable puede haber cambiado, la tendencia de las personas a engañar a los encuestadores sigue siendo fuerte.

Una reciente encuesta preguntó a los graduados de la Universidad de Maryland diversas preguntas acerca de su experiencia universitaria. Las respuestas se compararon con los registros oficiales. Las personas dieron consistentemente información incorrecta, de manera que hicieron que se vean bien. Menos de 2 por ciento informó que se graduaron con más bajo que un promedio de 2.5. (En realidad, alrededor del 11 por ciento lo hizo.) Y el 44 por ciento dijeron que habían donado a la universidad en el año pasado. (En realidad, alrededor del 28 por ciento lo hizo.)

Y es muy posible que la mentira juega un papel en el fracaso de las encuestas para predecir la victoria de Donald Trump 2016. Las encuestas, en promedio, subestimaron su apoyo alrededor de 2 puntos porcentuales. Algunas personas pueden haber sido vergüenza decir que estaban planeando para apoyarlo. Algunos pueden haber afirmado estar indeciso cuando fueron realmente van camino de Trump todo el tiempo.

¿Por qué la gente desinforman encuestas anónimas? Le pregunté a Roger Tourangeau-, profesor emérito de investigación en la Universidad de Michigan y quizás el mayor experto del mundo en el sesgo de deseabilidad social. Nuestra debilidad por “mentiras blancas” es una parte importante del problema, explicó. “Alrededor de un tercio de las veces, las personas se encuentran en la vida real”, sugiere. “Los hábitos se transfieren a las encuestas”.

Luego está esa manía a veces tenemos de mentir a nosotros mismos. “Hay una falta de voluntad para admitir a sí mismo que, por ejemplo, que eras una metida de pata como un estudiante”, dice Tourangeau-.

Mentirse a uno mismo puede explicar por qué tantas personas dicen que están por encima del promedio. ¿Qué tan grande es este problema? Más del 40 por ciento de los ingenieros de una compañía dijo que están en la parte superior del 5 por ciento. Más del 90 por ciento de los profesores de la universidad dicen que lo hacen el trabajo por encima de la media. Una cuarta parte de los estudiantes de secundaria piensan que están en la parte superior del 1 por ciento en su capacidad de llevarse bien con otras personas. Si está engañando a sí mismo, no puede ser honesto en una encuesta.



Otro factor que juega en nuestra mentir a las encuestas es nuestro fuerte deseo de hacer una buena impresión en el extranjero que realiza la entrevista, si hay alguien que realiza la entrevista, es decir. Como Tourangeau- pone, “Una persona que se parece a su tía favorita entra.. . . ¿Quieres decirle a su tía favorita que usó marihuana el mes pasado?”\* ¿Quieres que admitir que usted no dio dinero a su buena alma mater de edad?

Por esta razón, la más impersonal las condiciones, las personas más honestas serán. Para la obtención de respuestas verdaderas, las encuestas de Internet son mejores que las encuestas telefónicas, que son mejores que las encuestas en persona. La gente se admiten más si son solo que si los demás están en la habitación con ellos.

Sin embargo, en temas sensibles, cada método de encuesta provocará comunicación errónea sustancial. Tourangeau aquí se usa una palabra que a menudo se produce alrededor de los economistas: “incentivo” Las personas no tienen ningún incentivo para contar las encuestas de la verdad.

¿Cómo, por lo tanto, podemos aprender lo que nuestros compañeros humanos realmente están pensando y haciendo?

En algunos casos, existen fuentes de datos oficiales podemos referencia para obtener la verdad. Incluso si la gente miente sobre sus donaciones de caridad, por ejemplo, podemos obtener los números reales de dar en una zona de las mismas organizaciones. Pero cuando estamos tratando de aprender acerca de los comportamientos que no están tabulados en registros oficiales o estamos tratando de aprender lo que la gente está pensando, sus verdaderas creencias, sentimientos y deseos, no hay otra fuente de información, salvo lo que la gente se digne a contar encuestas. Hasta ahora, eso es.

Esta es la segunda potencia de Big Data: ciertas fuentes en línea que la gente a admitir cosas que no admitiría ningún otro lugar. Sirven como un suero de la verdad digital. Piense en las búsquedas de Google. Recuerde que las condiciones que hacen que la gente más honesta. ¿En línea? Comprobar. ¿Solo? Comprobar. No hay persona que administra una encuesta? Comprobar.

Y hay otra gran ventaja de que las búsquedas de Google tienen en lograr que la gente dice la verdad: incentivos. Si usted disfruta de chistes racistas, que tiene cero incentivo para compartir ese hecho no-PC con una encuesta. Usted, sin embargo, tiene un incentivo para buscar los mejores nuevos chistes racistas en línea. Si cree que puede estar sufriendo de depresión, que no tiene un incentivo para admitir esto a una encuesta. Usted tiene un incentivo para solicitar a Google síntomas y posibles tratamientos.

Incluso si usted está mintiendo a sí mismo, sin embargo, Google puede saber la verdad. Un par de días antes de la elección, usted y algunos de sus vecinos puede pensar legítimamente le va a conducir a un lugar de votación y emiten sus votos. Pero, si usted y que no ha buscado alguna información sobre cómo votar o dónde votar, los científicos de datos como yo puedo entender que la participación en su área será realmente baja. Del mismo modo, puede que usted no ha admitido a sí mismo que puede sufrir de depresión, incluso cuando usted está buscando en Google sobre el llanto jags y dificultad para levantarse de la cama. Se podría aparecer, sin

embargo, en las búsquedas relacionados con la depresión de un área que he analizado anteriormente en este libro.

Piense en su propia experiencia a través de Google. Creo que te tiene en ocasiones escrito cosas en ese cuadro de búsqueda que revelan un comportamiento o pensamiento que usted dude de admitir entre gente educada. De hecho, la evidencia es abrumadora de que una gran mayoría de los estadounidenses están diciendo a Google algunas cosas muy personales. Los estadounidenses, por ejemplo, la búsqueda para “porno” más que la búsqueda de “tiempo”. Esto es difícil, por cierto, de conciliar con los datos de la encuesta, ya que sólo alrededor del 25 por ciento de los hombres y el 8 por ciento de las mujeres admiten que ven pornografía.

Es posible que también han notado una cierta honestidad en las búsquedas de Google cuando se mira la forma en que este motor de búsqueda automáticamente trata de completar sus consultas. Sus sugerencias se basan en las búsquedas más comunes que otras personas han hecho. EE.UU. Por lo tanto pistas de auto-completado en lo que la gente está buscando en Google. De hecho, la función de autocompletar puede ser un poco engañoso. Google no sugerir ciertas palabras que considere apropiada, como “gallo”, “mierda” y “porno”. Esto significa que la función de autocompletar nos dice que los pensamientos de la gente de Google son menos mordaz de lo que realmente son. Aún así, algunas cosas sensibles a menudo todavía aparece.

Si escribe “¿Por qué es. . .” Las dos primeras Google completa automáticamente la actualidad son ‘¿Por qué el cielo es azul?’ Y ‘¿Por qué hay un día bisiest?’ Sugiriendo que estas son las dos formas más comunes para completar esta búsqueda. La tercera: “¿Por qué mi caca verde?” Y Google función de autocompletar puede conseguir molesto. Hoy en día, si escribe “¿Es normal que se quiera. . .” La primera sugerencia es ‘matar’. Si escribe “¿Es normal que quiere matar. . .” La primera sugerencia es ‘mi familia’.

¿Necesita más pruebas de que las búsquedas de Google pueden dar una imagen diferente del mundo que la que normalmente vemos? Considere búsquedas relacionadas con lamentos en torno a la decisión de tener o no tener hijos. Antes de decidir, algunas personas temen que podrían tomar la decisión equivocada. Y, casi siempre, la pregunta es si van a lamentar no tener hijos. Las personas son siete veces más propensos a pedir a Google si van a lamentar no tener hijos que si van a lamentar tener hijos.

Después de tomar su decisión, ya sea para reproducir (o adoptar) o no- gente a veces confiesan a Google que rue su elección. Esto puede venir como una especie de shock, pero después de la decisión, los números se invierten. Los adultos con los niños son 3,6 veces más propensos a decirle a Google que se arrepienten de su decisión que los adultos sin hijos.

Una advertencia que debe tenerse en cuenta en este capítulo: Google puede mostrar un sesgo hacia pensamientos indecorosos, pensamientos las personas sienten que no pueden hablar con nadie más. Sin embargo, si estamos tratando de descubrir los pensamientos ocultos, la capacidad de Google para descubrir a cabo puede ser útil. Y la gran disparidad entre lamenta de tener frente a no tener hijos parece estar diciéndonos que el pensamiento impropio en este caso es un tema muy importante.

Hagamos una pausa por un momento para considerar lo que significa incluso hacer una búsqueda como “Lamento tener hijos.” Google se presenta como una fuente de la que podemos buscar información directamente, sobre temas como el clima, que ganaron el partido de anoche, o cuando se erigió la estatua de la libertad. Pero a veces escribimos nuestros pensamientos sin censura en Google, sin mucha esperanza de que será capaz de ayudarnos. En este caso, la ventana de búsqueda sirve como una especie de confesionario.

Hay miles de búsquedas cada año, por ejemplo, por “No me gusta el frío”, “la gente está molesta”, y “estoy triste”. Por supuesto, esos miles de búsquedas en Google de “estoy triste” representan sólo una pequeña de la fracción de los cientos de millones de personas que se sienten tristes en un año determinado. Búsquedas expresan pensamientos, en vez de buscar la información, mi investigación ha encontrado, sólo se hacen por una pequeña muestra de todo el mundo para quien ese pensamiento viene a la mente. Del mismo modo, mi investigación sugiere que los siete mil búsquedas de estadounidenses cada año para “Lamento tener hijos” representan una pequeña muestra de los que han tenido ese pensamiento.

Los niños son, obviamente, una gran alegría para muchos, probablemente la mayoría, la gente. Y, a pesar del miedo de mi madre que “usted y su análisis de los datos estúpida” va a limitar su número de nietos, esta investigación no ha cambiado mi deseo de tener hijos. Pero ese pesar indecoroso es interesante, y otro aspecto de la humanidad que no solemos ver en las bases de datos tradicionales. Nuestra cultura está constantemente inundando con imágenes de familias maravillosas y felices. La mayoría de la gente nunca considerar el tener hijos como algo que podría lamentar. Pero algunos lo hacen. Pueden admitir esto a nadie, excepto Google.

## LA VERDAD SOBRE EL SEXO

---

¿Cuántos hombres estadounidenses son homosexuales? Esta es una pregunta legendaria en la investigación de la sexualidad. Sin embargo, ha sido una de las cuestiones más difíciles para los científicos sociales para responder. Los psicólogos ya no creen de Alfred Kinsey famosa en encuestas que sobremuestreo presos y prostitutas-que el 10 por ciento de los hombres estadounidenses son homosexuales basados en la estimación. encuestas representativas ahora nos dicen alrededor de 2 a 3 por ciento son. Sin embargo, la preferencia sexual ha sido durante mucho tiempo uno de los temas sobre los que la gente ha tendido a mentir. Creo que puedo utilizar grandes volúmenes de datos para dar una mejor respuesta a esta pregunta de lo que nunca hemos tenido.

En primer lugar, más en que los datos de la encuesta. Las encuestas nos dicen que hay muchos más hombres homosexuales en los Estados Unidos tolerantes que intolerantes. Por ejemplo, según una encuesta de Gallup, la proporción de la población que es gay es casi el doble que en Rhode Island, el estado con el mayor apoyo al matrimonio homosexual, de Mississippi, el estado con el apoyo más bajo para el matrimonio homosexual.

Hay dos posibles explicaciones para esto. En primer lugar, los hombres homosexuales nacidos en Estados intolerantes pueden pasar a estados tolerantes. En segundo lugar, los hombres homosexuales en estados intolerantes no podrá divulgar su homosexualidad; que son aún más probable que se encuentre.

Alguna información sobre el número explicación de un gay movilidad puede ser obtenida de otra fuente de datos grandes: Facebook, que permite a los usuarios una lista de lo géneros que están interesados en un 2,5 por ciento de los usuarios masculinos de Facebook que enumeran una de género de interés dicen que están interesados en los hombres; que se corresponde aproximadamente con lo que las encuestas indican. Y Facebook también muestra grandes diferencias en la población homosexual en los estados con alta versus baja tolerancia: Facebook tiene la población homosexual más de dos veces mayor en Rhode Island como en Mississippi.

Facebook también puede proporcionar información sobre cómo las personas se mueven alrededor. Yo era capaz de codificar la ciudad natal de una muestra de usuarios de Facebook abiertamente homosexuales. Esto me permitió estimar directamente el número de hombres homosexuales se mueven fuera de los Estados intolerantes en partes más tolerantes del país. ¿La respuesta? Es evidente que existe una cierta movilidad, de Oklahoma City a San Francisco, por ejemplo. Pero estimo que los hombres a empacar sus CDs Judy Garland y dirigirse a un lugar más abierto de mente puede explicar menos de la mitad de la diferencia en la población homosexual en tolerantes frente a los estados intolerantes.\*

Además, Facebook nos permite centrarnos en los estudiantes de la escuela secundaria. Este es un grupo especial, porque los niños de secundaria rara vez tienen que elegir el lugar donde viven. Si la movilidad explicó las diferencias del estado-por-estado en la población homosexual, estas diferencias no deben aparecer entre los usuarios de la escuela secundaria. Entonces, ¿qué dicen los datos de la escuela secundaria? Hay muchos menos chicos del instituto abiertamente homosexuales en estados intolerantes. Sólo dos de cada mil estudiantes de secundaria masculinos en Mississippi son abiertamente homosexual. Por lo tanto, no es sólo la movilidad.

Si un número similar de hombres homosexuales nacen en cada estado y la movilidad no puede explicar completamente por qué algunos estados tienen muchos más hombres abiertamente homosexuales, el armario debe estar jugando un papel importante. Lo que nos lleva de nuevo a Google, con la que tantas personas han demostrado dispuesto a compartir tanto.

Podría haber una manera de utilizar las búsquedas de pornografía para probar cuántos hombres homosexuales en realidad están en diferentes estados? De hecho, existe. En todo el país, estimo que utilizan los datos de las búsquedas de Google y Google AdWords que alrededor del 5 por ciento de las búsquedas pornográficas masculinas son para el porno gay masculina. (Éstos incluirían las búsquedas de términos tales como “Tubo Rocket,” un sitio pornográfico gay muy popular, así como “porno gay.”)

Y ¿cómo varía esto en diferentes partes del país? En general, hay búsquedas porno más en estados tolerantes en comparación con los estados intolerantes. Esto tiene sentido, dado que algunos hombres homosexuales se mueven fuera de los lugares intolerantes a lugares tolerantes. Pero las diferencias no son tan grandes como las diferencias sugeridas por encuestas o Facebook.

En Mississippi, estimo que el 4,8 por ciento de las búsquedas pornográficas masculinas son para el porno gay, muy superior a las cifras sugeridas por encuestas o Facebook y razonablemente cerca del 5.2 por ciento de las búsquedas de pornografía que son para el porno gay en Rhode Island.

Entonces, ¿cuántos hombres estadounidenses son homosexuales? Esta medida busca de pornografía por hombres más o menos 5 por ciento son del mismo sexo, parece una estimación razonable del verdadero tamaño de la población gay en los Estados Unidos. Y hay otra forma, menos sencillo para llegar a este número. Se requiere un poco de ciencia de datos. Podríamos utilizar la relación entre la tolerancia y la población homosexual. Tengan paciencia conmigo un poco aquí.

Mi investigación preliminar indica que en un estado dado cada 20 puntos porcentuales de apoyo al matrimonio homosexual significa casi una vez y media mayor número de hombres de estado que van a identificarse abiertamente como homosexuales en Facebook. En base a esto, se puede estimar la cantidad de hombres nacidos en un hipotéticamente totalmente tolerantes lugar-donde, por ejemplo, el 100 por ciento de la gente apoyaba el matrimonio-gay sería ser abiertamente homosexual. Mi estimación es alrededor del 5 por ciento sería, que se ajusta a los datos de la pornografía busca muy bien. Lo más cerca que podríamos tener que crecer en un entorno totalmente tolerantes son los chicos de secundaria en Área de la Bahía de California. Cerca del 4 por ciento de ellos son abiertamente homosexual en Facebook. Eso parece en línea con mi cálculo.

Debo señalar que todavía no he sido capaz de llegar a una estimación de la atracción del mismo sexo para las mujeres. Los números de pornografía son menos útiles aquí, ya que muchas menos mujeres miran pornografía, por lo que la muestra menos representativa. Y de los que lo hacen, incluso las mujeres que se sienten atraídas principalmente a los hombres en la vida real parecen disfrutar viendo porno lesbiana. Totalmente 20 por ciento de los videos vistos por las mujeres en Pornhub son lesbianas.

El cinco por ciento de los hombres estadounidenses ser homosexual es una estimación, por supuesto. Algunos hombres son bisexuales; algunos, sobre todo cuando no se está seguro de jóvenes de lo que son. Obviamente, no se puede contar esto como, precisamente, como puede ser que el número de personas que votan o asistir a una película.

Pero una de las consecuencias de mi estimación es clara: una gran cantidad de hombres en los Estados Unidos, particularmente en los estados de intolerancia, todavía están en el armario. No revelan sus preferencias sexuales en Facebook. Ellos no lo admiten en encuestas. Y en muchos casos, pueden incluso estar casado con la mujer.

Resulta que las esposas a sus maridos sospechan de ser gay con bastante frecuencia. Ellos demuestran que la sospecha en la búsqueda sorprendentemente común: “¿Es mi marido gay” “Gay” es 10 por ciento más propensos a completar las búsquedas que comienzan “es mi marido. . .” De la palabra en segundo lugar, ‘hacer trampa’. Es ocho veces más común que ‘un alcohólico’ y diez veces más común que la ‘depresión’.

Tal vez lo más revelador, búsquedas en duda la sexualidad de un marido son mucho más frecuentes en las regiones menos tolerantes. Los estados con el mayor porcentaje de mujeres que piden esta pregunta son de Carolina del Sur y Luisiana. De hecho, en veintiuno de los veinticinco estados en los que se hacen con más frecuencia a esta pregunta, el apoyo al matrimonio gay es más bajo que el promedio nacional.

Google y los sitios de pornografía no son los únicos recursos de datos útiles cuando se trata de la sexualidad masculina. Hay más datos disponibles en grandes volúmenes de datos en lo que significa vivir en el armario. Analicé anuncios en Craigslist para los hombres que buscan “encuentros casuales”. El porcentaje de estos anuncios que buscan relaciones esporádicas con hombres tiende a ser mayor en los estados menos tolerantes. Entre los estados con los porcentajes más altos son de Kentucky, Louisiana y Alabama.

Y para aún más de un vistazo en el armario, volvamos a los datos de búsqueda de Google y obtener un poco más granular. Una de las búsquedas más comunes realizados inmediatamente antes o después de “porno gay” es “prueba de homosexuales.” (Estas pruebas suponen a decir a los hombres si son o no son homosexuales.) Y las búsquedas de “prueba de homosexuales” son aproximadamente dos veces más frecuente en el estados menos tolerante.

¿Qué significa para ir y venir entre la búsqueda de “porno gay” y la búsqueda de “prueba gay”? Presumiblemente, sugiere una bastante confusa si no mente torturada. Es razonable sospechar que algunos de estos hombres tienen la esperanza de confirmar que su interés en el porno gay en realidad no quiere decir que sean homosexuales.

Los datos de búsqueda de Google no permite que podamos ver el historial de búsqueda de un usuario determinado en el tiempo. Sin embargo, en 2006, AOL lanzó una muestra de las búsquedas de sus usuarios a los investigadores académicos. Estas son algunas de las búsquedas de un usuario anónimo en un período de seis días.

Viernes 03:49:55	picks homosexuales libres [sic]
Viernes 03:59:37	selecciones homosexuales vestuario
Viernes 04:00:14	selecciones de los homosexuales
Viernes 04:00:35	selecciones de sexo gay
Viernes 05:08:23	un concurso gay larga
Viernes 05:10:00	una buena prueba gay
Viernes 05:25:07	pruebas de los homosexuales para un hombre confundido
Viernes 05:26:38	pruebas de los homosexuales
Viernes 05:27:22	¿Estoy pruebas homosexuales

Viernes 05:29:18	selecciones de los homosexuales
Viernes 05:30:01	Hombres desnudos picos
Viernes 05:32:27	selecciones de hombres desnudos gratis
Viernes 05:38:19	selecciones calientes de sexo gay
Viernes 05:41:34	hombre caliente sexo a tope
Miércoles 13:37:37	¿Estoy pruebas homosexuales
Miércoles 13:41:20	pruebas de los homosexuales
Miércoles 13:47:49	hombre caliente sexo a tope
Miércoles 13:50:31	Gay vidio sexo libre [sic]

Esto sin duda se lee como un hombre que no se siente cómodo con su sexualidad. Y los datos de Google nos dice que todavía hay muchos hombres como él. La mayor parte de ellos, de hecho, viven en estados que son menos tolerantes a las relaciones del mismo sexo.

Para un aspecto aún más de cerca a la gente detrás de estos números, le pregunté a un psiquiatra en Mississippi, que se especializa en ayudar a los hombres homosexuales encerrados, si alguno de sus pacientes podría querer hablar conmigo. Un hombre se acercó. Me dijo que era un profesor retirado, de unos sesenta años, y casado con la misma mujer desde hace más de cuarenta años.

Hace unos diez años, abrumado con el estrés, vio que el psiquiatra y finalmente reconoció su sexualidad. Siempre ha sabido que estaba atraído por los hombres, dice, pero pensó que esto era algo universal y que todos los hombres simplemente se escondieron. Poco después de comenzar la terapia, tuvo su primer y único encuentro sexual gay, con un alumno suyo que era de casi treinta años, una experiencia que él describe como “maravillosa”.

Él y su esposa no tienen relaciones sexuales. Dice que se sentiría culpable cada vez que termina su matrimonio o salir con un hombre abiertamente. Se lamenta de prácticamente cada una de sus decisiones más importantes de la vida.

El profesor retirado y su esposa irán otra noche sin amor romántico, sin sexo. Pese a sus avances, la persistencia de la intolerancia provocará millones de otros estadounidenses a hacer lo mismo.

Puede que no sea sorprendida al saber que el 5 por ciento de los hombres son homosexuales y que muchos permanecen en el armario. Ha habido momentos en los que habrían sido sorprendido mayoría de la gente. Y todavía hay plazas donde mucha gente se sorprendería también.

“En Irán no tenemos homosexuales como en su país”, Mahmud Ahmadineyad, entonces presidente de Irán, insistió en 2007. “En Irán no tenemos este fenómeno.” Del mismo modo, Anatoly Pakhomov, alcalde de Sochi, Rusia, poco antes de su ciudad fue sede de los Juegos Olímpicos de invierno de 2014, dijo de las personas homosexuales, “no tenemos en nuestra ciudad.” sin embargo, el comportamiento de Internet revela un interés significativo en el porno gay en Sochi e Irán.

Esto plantea una pregunta obvia: ¿existen intereses sexuales comunes en los Estados Unidos hoy en día que todavía se consideran impactante? Depende de lo que se considera común y la facilidad con impactado eres.

La mayoría de las principales búsquedas en Pornhub no son sorprendentes, que incluyen términos como “adolescente”, “trío”, y “mamada” para los hombres, frases como “la toma apasionado amor”, “chupar el pezón,” y “hombre comiendo coño” para mujer.

Al salir de la corriente principal, los datos Pornhub nos dice acerca de algunos fétiches que usted podría no haber existido jamás imaginado. Hay mujeres que buscan “manzanas anal” y “follar animales de peluche.” Hay hombres que buscan “fetichismo mocos” y “crucifixión desnuda.” Pero estas búsquedas son raras sólo el diez por cada mes, incluso en este gran sitio porno .

Otro punto relacionado que se vuelve bastante claro al revisar los datos Pornhub: hay alguien ahí fuera para todo el mundo. Las mujeres, como es lógico, a menudo buscan chicos “altos”, chicos “oscuras” y “chicos guapos”. Pero también a veces la búsqueda de chicos “cortas”, “pálido” chicos y chicos “feo”. Hay mujeres que buscan chicos “discapacitados”, “gordito con pene pequeño”, y “anciano grasa fea.” Frecuentemente, los hombres buscan mujeres “finas”, las mujeres con “tetas grandes”, y las mujeres con el pelo “rubio” . Pero también a veces la búsqueda de la mujer “grasa”, las mujeres con “tetas pequeñas”, y las mujeres con “pelo verde.” Hay hombres que buscan mujeres “calvas”, las mujeres “enano”, y las mujeres con “no hay pezones.” Estos datos pueden ser animando a los que no son de alto, moreno y guapo o delgada, de grandes pechos, y la rubia.\*

¿Qué pasa con otras búsquedas que son a la vez común y sorprendente? Entre las 150 búsquedas más comunes de los hombres, el más sorprendente para mí son los incestuosos traté en el capítulo sobre Freud. Otros objetos poco discutido de deseo de los hombres son “travestis” (77a más común de búsqueda) y “abuelita” (número 110 de búsqueda más común). En general, aproximadamente el 1,4 por ciento de las búsquedas Pornhub de los hombres son para las mujeres con penes. Alrededor de 0,6 por ciento (0,4 por ciento para los hombres menores de treinta y cuatro) son para los ancianos. Sólo 1 de cada 24.000 búsquedas Pornhub por los hombres son explícitamente para preadolescentes; que puede tener algo que ver con el hecho de que Pornhub, por razones obvias, prohíbe todas las formas de pornografía infantil y que posee es ilegal.

Entre las principales búsquedas Pornhub por las mujeres es un género de la pornografía que, te lo advierto, moleste a muchos lectores: el sexo con la violencia contra las mujeres. Totalmente 25 por ciento de las búsquedas hembra para la pornografía recta enfatizar el dolor y / o la



humillación de la mujer-“anal dolorosa llorando”, “vergüenza pública” y “brutal orgía extrema”, por ejemplo. Cinco por ciento busca sexo no consensuado “violación” o “forzados” sexo-a pesar de que estos videos están prohibidos en Pornhub. Y las tasas de búsqueda de todos estos términos son al menos dos veces más común entre las mujeres como entre los hombres. Si hay un género de porno en el que la violencia se ejerce contra una mujer, mi análisis de los datos muestra que casi siempre se apela de manera desproporcionada a las mujeres.

Por supuesto, cuando se trata de llegar a un acuerdo con esto, es muy importante recordar que hay una diferencia entre la fantasía y la vida real. Sí, de la minoría de mujeres que visitan Pornhub, hay un subgrupo que busca, sin éxito, para las imágenes de violación. Para decir lo obvio, esto no significa que las mujeres quieren ser violadas en la vida real y ciertamente no hace que sea menos horrible violación un crimen. Lo que los datos del porno nos dice es que a veces la gente tiene fantasías que deseen que no tenían y que nunca pueden mencionar a los demás.

Armarios no son sólo los repositorios de fantasías. Cuando se trata de sexo, la gente guarda muchos secretos acerca de lo mucho que tienen, por ejemplo.

En la introducción, he observado que los estadounidenses informan que usan mucho más preservativos que se venden cada año. Por lo tanto, se podría pensar que esto significa que sólo están diciendo que usan preservativos con más frecuencia durante el sexo de lo que realmente hacen. La evidencia sugiere que también exageran la frecuencia con que están teniendo relaciones sexuales, para empezar. Alrededor del 11 por ciento de las mujeres entre las edades de quince y cuarenta y cuatro dicen que son sexualmente activos, no está embarazada, y no usar la anticoncepción. Incluso con supuestos relativamente conservadoras sobre el número de veces que tienen relaciones sexuales, los científicos esperarían que el 10 por ciento de ellos se quede embarazada cada mes. Pero esto ya sería más que el número total de embarazos en los Estados Unidos (que es 1 de cada 113 mujeres en edad fértil). En nuestra cultura obsesionada con el sexo puede ser difícil admitir que se acaba de no tener que mucho.

Pero si usted está buscando la comprensión o consejo, que tiene, una vez más, un incentivo para decirle a Google. En Google, hay dieciséis veces más quejas acerca de un cónyuge no querer sexo que aproximadamente un socio casado no estar dispuesto a hablar. Hay cinco y medio veces más quejas acerca de la pareja de hecho sin querer el sexo de una pareja de hecho negarse a texto espalda.

Y las búsquedas de Google sugieren un culpable sorprendente para muchas de estas relaciones sin sexo. Hay el doble de muchas quejas de que un novio no va a tener relaciones sexuales que una novia no va a tener relaciones sexuales. Por el momento, el número uno de búsqueda queja acerca de un novio es “Mi novio no va a tener relaciones sexuales conmigo.” (Las búsquedas de Google no están desglosados por género, pero, puesto que el análisis anterior, dijo que el 95 por ciento de los hombres son rectas, nos puede adivinar que no demasiadas búsquedas “novio” vienen de los hombres).

¿Cómo debemos interpretar esto? ¿Realmente esto implica que los novios retienen el sexo más que amigas? No necesariamente. Como se mencionó anteriormente, las búsquedas de Google

pueden estar sesgados a favor de las personas de la materia se tensa hablando. Los hombres pueden sentirse más cómodos al hablar con sus amigos acerca de la falta de interés sexual que las mujeres están diciendo a sus amigos acerca de su novio o de su novia. Sin embargo, incluso si los datos de Google no implican que los novios son realmente dos veces más propensos a evitar las relaciones sexuales como amigos, sí sugiere que los novios evitan las relaciones sexuales es más común que la gente deja en.

Google datos también sugieren una razón las personas pueden estar eludiendo las relaciones sexuales con tanta frecuencia: enorme ansiedad, con gran parte de ella fuera de lugar. Comience con las ansiedades de los hombres. No es ninguna novedad que los hombres se preocupan por lo bien dotado que son, pero el grado de esta preocupación es bastante profunda.

Hombres Google más preguntas acerca de su órgano sexual que cualquier otra parte del cuerpo: más de alrededor de los pulmones, el hígado, los pies, las orejas, la nariz, la garganta y el cerebro combinado. Los hombres realizan más búsquedas de cómo hacer que sus penes más grandes que la forma de afinar una guitarra, hacer una tortilla, o cambiar un neumático. Googled principal preocupación de los hombres acerca de los esteroides no es si pueden dañar su salud, pero si tomar ellos podrían disminuir el tamaño de su pene. Arriba Pregunta Googled de los hombres en relación con la forma de su cuerpo o de la mente cambiarían a medida que envejecían era si su pene podría hacerse más pequeño.

Nota al margen: Una de las preguntas más comunes de Google con respecto a los genitales de los hombres es “¿Qué tan grande es mi pene?” Que los hombres recurren a Google, en lugar de una regla, con esta pregunta es, en mi opinión, una expresión por excelencia de nuestra era digital . \*

¿Le importa a las mujeres sobre el tamaño del pene? En raras ocasiones, según las búsquedas de Google. Por cada búsqueda mujeres constituyen aproximadamente el falo de un socio, los hombres hacen aproximadamente 170 búsquedas sobre su propia. Es cierto que en las raras ocasiones en las mujeres hacen expresar preocupaciones sobre el pene de su pareja, es con frecuencia acerca de su tamaño, pero no necesariamente que es pequeño. Más del 40 por ciento de las quejas sobre el tamaño del pene de la pareja dicen que es demasiado grande. “Dolor” es la palabra más Googled utilizado en las búsquedas con la frase “\_\_\_ durante el acto sexual.” ( “Sangrado”, “pipí”, “llorando” y “pedos” completan los cinco primeros.) Sin embargo, sólo el 1 por ciento del de los hombres búsquedas que buscan cambiar su tamaño del pene están buscando información sobre cómo hacerlo más pequeño.

la segunda más común cuestión sexual masculino es cómo hacer que sus encuentros sexuales más tiempo. Una vez más, las inseguridades de los hombres no parecen coincidir con los intereses de las mujeres. Hay más o menos el mismo número de búsquedas que piden cómo hacer un clímax novio más rápido clímax más lentamente. De hecho, las mujeres han preocupación más común relacionada con el orgasmo de un novio no es acerca de cuándo sucedió, pero por qué no está sucediendo en absoluto.

No es frecuente hablar de problemas de imagen corporal cuando se trata de hombres. Y si bien es cierto que el interés general en la apariencia personal sesga femenino, que no es tan desigual

como estereotipos sugerirían. De acuerdo con mi análisis de Google AdWords, que mide los sitios web de las personas que visitan, el interés por la belleza y la forma física es del 42 por ciento hombres, la pérdida de peso es del 33 por ciento hombres, y la cirugía estética es de 39 por ciento hombres. Entre todas las búsquedas con “cómo” relacionado con los senos, alrededor del 20 por ciento preguntar cómo deshacerse de los senos de hombre.

Pero, incluso si el número de hombres que carecen de confianza en sus cuerpos es más alto que la mayoría de la gente podría pensar, las mujeres todavía dejan atrás cuando se trata de la inseguridad acerca de cómo se ven. Entonces, ¿qué puede ser esto suero de la verdad digitales revelar acerca de las dudas de las mujeres? Cada año en los Estados Unidos, hay más de siete millones de búsquedas que buscan en los implantes mamarios. Las estadísticas oficiales nos dicen que alrededor de 300.000 mujeres pasan por el procedimiento al año.

Las mujeres también muestran una gran inseguridad sobre sus traseros, aunque muchas mujeres se han recientemente-flop flip sobre qué es lo que no les gusta de ellos.

En 2004, en algunas partes de los Estados Unidos, la búsqueda más común en relación con el cambio de uno a tope era cómo hacer que sea más pequeño. El deseo de hacer la propia parte inferior más grande se concentró mayoritariamente en áreas con grandes poblaciones negras. A partir de 2010, sin embargo, el deseo de colillas más grandes creció en el resto de los Estados Unidos. Este interés, si no la distribución posterior en sí, se ha triplicado en cuatro años. En 2014, hubo más búsquedas preguntando cómo hacer que su trasero más grande que la más pequeña en todos los estados. En estos días, por cada cinco búsquedas que buscan en los implantes de mama en los Estados Unidos, hay un mirar en los implantes de glúteos. (Gracias, Kim Kardashian!)

Hace cada vez mayor preferencia de las mujeres para las preferencias de los hombres más grandes a los partidos de fondo? Curiosamente, sí. Búsquedas “gran trasero porno”, que también se utilizan para concentrarse en las comunidades negras, se han disparado recientemente en popularidad en los Estados Unidos.

¿Qué más quieren los hombres en el cuerpo de una mujer? Como se mencionó anteriormente, y como la mayoría encontrarán salta a la vista, los hombres muestran una preferencia por los pechos grandes. Alrededor del 12 por ciento de las búsquedas pornográficas no genéricas están buscando grandes pechos. Esto es casi veinte veces mayor que el volumen de búsquedas de pornografía pequeña de mama.

Dicho esto, no está claro que esto significa que los hombres quieren que las mujeres reciban implantes mamarios. Aproximadamente el 3 por ciento de la pornografía grandes búsquedas de mama dice explícitamente que quieren ver los pechos naturales.

Google busca sobre uno de los implantes de mama y esposa están divididos en partes iguales entre pedir cómo persuadir a ella para obtener implantes y perplejidad en cuanto a por qué los quiere.

O considerar la búsqueda más común acerca de los pechos de una novia: “Me encanta las tetas de mi novia.” No está claro lo que los hombres tienen la esperanza de encontrar desde Google al realizar esta búsqueda.

Las mujeres, como los hombres, tienen preguntas acerca de sus genitales. De hecho, tienen casi tantas preguntas acerca de sus vaginas que los hombres tienen acerca de sus penes. preocupaciones de las mujeres sobre sus vaginas son a menudo relacionados con la salud. Pero al menos el 30 por ciento de sus preguntas dedicarse a otras preocupaciones. Las mujeres quieren saber cómo afeitarse, apretarlo, y que tenga mejor sabor. Una preocupación sorprendentemente común, como se habló anteriormente, es cómo mejorar su olor.

Las mujeres son más frecuentemente preocupados de que sus vaginas huelen a pescado, seguido por el vinagre, la cebolla, el amoníaco, el ajo, el queso, el olor corporal, orina, pan, blanqueo, heces, sudor, metal, pies, basura, y la carne podrida.

En general, los hombres no hacen muchas búsquedas de Google que involucran los genitales de la pareja. Los hombres hacen más o menos el mismo número de búsquedas sobre la vagina de una novia como las mujeres sobre el pene de un novio.

Cuando los hombres de la búsqueda Sobre la vagina de su pareja, por lo general es para quejarse de lo que las mujeres se preocupan más: el olor. En su mayoría, los hombres están tratando de encontrar la manera de decirle a una mujer sobre un mal olor sin herir sus sentimientos. A veces, sin embargo, las preguntas de los hombres con respecto al olor revelan sus propias inseguridades. El hombre a veces piden formas de usar el olfato para detectar el engaño-Si huele como los condones, por ejemplo, o el semen de otro hombre.

¿Qué debemos hacer de todo esto la inseguridad secreto? Es evidente que hay algunas buenas noticias aquí. Google nos da razones legítimas para preocuparse menos que nosotros. Muchos de nuestros miedos más profundos acerca de cómo nuestras parejas sexuales nos perciben son injustificadas. A solas, en sus computadoras, sin incentivo para mentir, socios se revelan a ser bastante nonsuperficial y de perdón. De hecho, todos estamos tan ocupados juzgar nuestros propios cuerpos que hay poca energía de sobra para juzgar a otras personas.

También es probable que una conexión entre dos de las grandes preocupaciones reveladas en las búsquedas en Google sexuales: la falta de sexo y una inseguridad sobre la propia atracción sexual y el rendimiento. Tal vez estos están relacionados. Tal vez si nos preocupa menos sobre el sexo, tendríamos más de lo mismo.

¿Qué más pueden búsquedas de Google decirnos sobre el sexo? Podemos hacer una batalla de los sexos, para ver quién es más generoso. Tome todas las búsquedas que buscan maneras de mejorar en la práctica de sexo oral sobre el sexo opuesto. Qué buscan los hombres para obtener más consejos o las mujeres? ¿Quién es sexualmente más generosos, hombres o mujeres? Las mujeres, duh. La suma de todas las posibilidades, estimo que la relación es de 2: 1 a favor de las mujeres en busca de consejo sobre cómo realizar mejor sexo oral a su pareja.

Y cuando los hombres buscan consejos sobre cómo dar sexo oral, que con frecuencia no se buscan maneras de complacer a otra persona. Los hombres hacen la mayor cantidad de búsquedas que buscan maneras de realizar sexo oral a sí mismos como lo hacen cómo dar a una mujer un orgasmo. (Este es uno de mis favoritos en hechos de datos de búsqueda de Google.)

**[LA VERDAD sobre el odio Y PREJUICIO](#)**

---

El sexo y el romance son casi los únicos temas envuelta en la vergüenza y, por lo tanto, no son los únicos temas sobre los que las personas guardan secretos. Muchas personas son, por una buena razón, inclinado a mantener sus prejuicios a sí mismos. Supongo que se podría llamar progreso que muchas personas hoy en día sienten que serán juzgados si admiten que juzgar a otras personas sobre la base de su origen étnico, orientación sexual o religión. Pero muchos estadounidenses todavía lo hacen. (Esta es otra sección, advertir a los lectores, que incluye material perturbador.)

Esto se puede ver en Google, donde los usuarios a veces hacen preguntas tales como “¿Por qué las personas de raza negra grosera?” O “¿Por qué los Judios mal?” A continuación, en orden, son las cinco palabras negativas más utilizados en las búsquedas sobre diversos grupos.

	1.	2.	3.	4.	5.
<b>AFRICAN AMERICANS</b>	rude	racist	stupid	ugly	lazy
<b>JEWS</b>	evil	racist	ugly	cheap	greedy
<b>MUSLIMS</b>	evil	terrorists	bad	violent	dangerous
<b>MEXICANS</b>	racist	stupid	ugly	lazy	dumb
<b>ASIANS</b>	ugly	racist	annoying	stupid	cheap
<b>GAYS</b>	evil	wrong	stupid	annoying	selfish
<b>CHRISTIANS</b>	stupid	crazy	dumb	delusional	wrong

Unos patrones entre estos estereotipos se destacan. Por ejemplo, los afroamericanos son el único grupo que se enfrenta a un estereotipo “grosero”. Casi todos los grupos es víctima de un estereotipo “estúpida”; los únicos dos que no lo son: Judios y musulmanes. El estereotipo “mal” se aplica a Judios, musulmanes y homosexuales, pero la gente no negros, mexicanos, asiáticos y cristianos.

Los musulmanes son el único grupo estereotipado como terroristas. Cuando un musulmán estadounidense hace el juego a este estereotipo, la respuesta puede ser instantánea y vicioso. los datos de búsqueda de Google nos pueden dar un vistazo minuto a minuto en tales erupciones de rabia odio como combustible.

Considere lo que sucedió poco después del tiroteo masivo en San Bernardino, California, el 2 de diciembre de 2015. Esa mañana, Rizwan Malik Farook y Tashfeen entraron en una reunión de compañeros de trabajo de Farook armados con pistolas semiautomáticas y rifles semiautomáticos y asesinados catorce personas. Esa noche, literalmente minutos después de los medios de comunicación informó por primera vez uno de nombre que suena musulmana de los tiradores, un número preocupante de los californianos habían decidido lo que querían hacer con los musulmanes: los mata.

La parte superior de búsqueda de Google en California con la palabra “musulmanes” en ella en ese momento era “matar a los musulmanes.” Y, en general, los estadounidenses buscaron la

frase “matar a los musulmanes” con la misma frecuencia que ellos buscaron “receta de martini,” “síntomas de la migraña”, y “Cowboys lista.” En los días siguientes al ataque de San Bernardino, para todos los estadounidenses que se ocupan de la ‘islamofobia’, otro fue en busca de ‘matar a los musulmanes.’ Aunque las búsquedas de odio fueron de aproximadamente el 20 por ciento de todas las búsquedas sobre los musulmanes antes el ataque, más de la mitad de todo el volumen de búsqueda sobre los musulmanes se convirtió en odio en las horas que siguieron.

Y este minuto a minuto los datos de búsqueda nos puede decir lo difícil que puede ser para calmar esta rabia. Cuatro días después del tiroteo, el entonces presidente Obama dio un discurso en horario estelar para el país. Él quería tranquilizar a los estadounidenses que el gobierno podría tanto detener el terrorismo y, quizás más importante, tranquilizar esta peligrosa islamofobia.

Obama hizo un llamamiento a nuestros mejores ángeles, hablando de la importancia de la inclusión y la tolerancia. La retórica era poderoso y en movimiento. El diario Los Angeles Times alabó a Obama por “[advertencia] en contra de permitir que el miedo nublar nuestro juicio.” The New York Times calificó el discurso tanto “duro” y “calma”. El sitio web de Think Progreso elogió como “una herramienta necesaria del buen gobierno, orientada a salvar las vidas de los musulmanes estadounidenses.” el discurso de Obama, en otras palabras, se consideró un éxito importante. Pero era?

datos de búsqueda de Google sugiere lo contrario. Junto con Evan SOLTAS, a continuación, en Princeton, examiné los datos. En su discurso, dijo el presidente, “Es responsabilidad de todos los estadounidenses de todas las religiones, para rechazar la discriminación.” Pero búsquedas llamando a los musulmanes “terroristas”, “malo”, “violentos” y “malo” se duplicó durante y poco después del discurso. El presidente Obama también dijo: “Es nuestra responsabilidad para rechazar pruebas religiosas en que se admite en este país.” Pero búsquedas negativas sobre los refugiados sirios, un grupo de mayoría musulmana entonces buscando desesperadamente un refugio seguro, aumentó 60 por ciento, mientras que las búsquedas preguntar cómo para ayudar a los refugiados sirios cayó un 35 por ciento. Obama pidió a los estadounidenses que “no hay que olvidar que la libertad es más fuerte que el miedo.” Sin embargo, las búsquedas de “matar a los musulmanes” se triplicó durante su discurso. De hecho,

En otras palabras, Obama parecía decir todas las cosas correctas. Todos los medios de comunicación tradicionales felicitó a Obama por sus palabras de curación. Pero los nuevos datos de internet, que ofrecen suero de la verdad digital, sugirieron que el discurso en realidad contraproducente en su objetivo principal. En lugar de calmar la multitud enojada, como todo el mundo pensó que estaba haciendo, los datos de Internet nos dice que Obama en realidad se inflama. Cosas que creemos que están trabajando pueden tener el efecto contrario del que esperamos. A veces necesitamos datos de Internet para corregir nuestro instinto para darnos una palmada en la espalda.

Entonces, ¿qué debería haber dicho Obama para acabar con esta particular forma de odio actualmente tan virulenta en América? Vamos a circundamos volver a eso más adelante. Ahora vamos a echar un vistazo a una vena secular de perjuicio en los Estados Unidos, la forma de odio que de hecho se destaca por encima del resto, el que ha sido el más destructivo y el tema de la

investigación que comenzó este libro. En mi trabajo con los datos de búsqueda de Google, el hecho más revelador único que he encontrado en relación con el odio en Internet es la popularidad de la palabra “negro”.

Ya sea en singular o en plural, la palabra “negro” se incluye en siete millones de búsquedas estadounidenses cada año. (Una vez más, la palabra usada en canciones de rap es casi siempre “negro” no “negro”, así que no hay impacto significativo desde el hip-hop letras para tener en cuenta.) Las búsquedas de “negro chistes” son diecisiete veces más común que las búsquedas de “chistes kike”, “Gook chistes,” “chistes spic”, “chistes chink”, y “marica chistes” combinado.

Cuando son las búsquedas de “nigger (s)” o “-” chistes de negros -la mayoría común? Siempre que los afroamericanos son en las noticias. Entre los períodos en que tales búsquedas eran más alto fue el período inmediatamente posterior del huracán Katrina, cuando la televisión y los periódicos mostraron imágenes de los negros desesperados en Nueva Orleans que luchan por su supervivencia. También se dispararon durante la primera elección de Obama. Y las búsquedas de subida “chistes de negros”, en promedio, un 30 por ciento en el día de Martin Luther King Jr.

La ubicuidad de esta aterradora insulto racial pone en duda algunos conocimientos actuales de racismo.

Cualquier teoría del racismo tiene que explicar un gran rompecabezas en Estados Unidos. Por un lado, la inmensa mayoría de los estadounidenses negros piensan que sufren de prejuicios y tienen una amplia evidencia de la discriminación en las identificaciones policiales, entrevistas de trabajo, y las decisiones del jurado. Por otro lado, muy pocos norteamericanos blancos admitirá a ser racista.

La explicación dominante entre los politólogos recientemente ha sido que esto se debe, en gran parte, a los prejuicios implícitos generalizada. Los norteamericanos blancos puede significar también, según esta teoría, pero tienen un sesgo subconsciente, que influye en su tratamiento de los americanos negros. Académico inventó una manera ingeniosa para la prueba de tal sesgo. Se llama el Test de Asociación Implícita.

Las pruebas han demostrado consistentemente que toma la mayoría de las personas milisegundos ya asociar caras negras con palabras positivas, tales como “bueno”, que con palabras negativas, tales como “horrible.” Para caras blancas, el patrón se invierte. El tiempo extra que se necesita es una prueba de que alguien implícita de prejuicio un perjuicio a la persona puede incluso no ser conscientes.

Hay, sin embargo, una explicación alternativa para la discriminación de que los afroamericanos se sienten blancos y niegan: oculta el racismo explícito. Supongamos que hay un racismo consciente razonablemente generalizada de que las personas son muy conscientes, pero a los que no confesará-ciertamente no en una encuesta. Eso es lo que los datos de búsqueda parece estar diciendo. No hay nada implícito sobre la búsqueda de “chistes de negros.” Y es difícil imaginar que los estadounidenses están buscando en Google la palabra “negro” con la misma frecuencia que “migraña” y “economista” sin racismo explícito tener un impacto importante en los afroamericanos. Antes de los datos de Google, que no tienen una medida

convinciente de esta animosidad virulenta. Ahora hacemos. Estamos, por lo tanto, en condiciones de ver lo que explica.

En él se explica, como se explicó anteriormente, ¿por qué el total de votos de Obama en 2008 y 2012 estaban deprimidos en muchas regiones. También se correlaciona con la diferencia de salarios blanco-negro, como un equipo de economistas informó recientemente. Las áreas que había encontrado hacen las búsquedas más racistas, en otras palabras, pagan de menos personas de raza negra. Y luego está el fenómeno de la candidatura de Donald Trump. Como se ha señalado en la introducción, cuando Nate Silver, el gurú de votación, buscó la variable geográfica que se correlaciona más fuertemente con el apoyo en las primarias republicanas de 2016 Trump, lo encontró en el mapa de racismo que había desarrollado. Esa variable fue búsquedas de “nigger (s).”

Los estudiosos han puesto recientemente en conjunto una medida del estado-por-estado de los prejuicios implícitos que obstaculicen la gente negra, que me ha permitido comparar los efectos del racismo explícito, según lo medido por las búsquedas de Google, y el sesgo implícito. Por ejemplo, he probado la cantidad de cada trabajaron contra Obama en sus dos elecciones presidenciales. Utilizando el análisis de regresión, he encontrado que, para predecir donde Obama obtuvo un rendimiento inferior, racistas búsquedas de Google de un área explican mucho. el rendimiento de un área de pruebas de asociación implícita-añadió poco.

Para ser provocador y para fomentar una mayor investigación en esta área, que me puso sucesivamente la siguiente conjetura, listo para ser probado por los estudiosos en una amplia gama de campos. La explicación principal de la discriminación contra los afroamericanos hoy no es el hecho de que las personas que estén de acuerdo en participar en experimentos de laboratorio hacen asociaciones subconscientes entre las palabras negativas y las personas de raza negra; es el hecho de que millones de americanos blancos continúan haciendo cosas como búsqueda de “chistes de negros.”

La discriminación de los negros experimentan regularmente en los Estados Unidos parece estar alimentado más ampliamente por explícita, si oculta, la hostilidad. Sin embargo, para otros grupos, el prejuicio subconsciente puede tener un impacto más fundamental. Por ejemplo, yo era capaz de utilizar Google busca encontrar evidencia de prejuicio implícito en contra de otro segmento de la población: las niñas.

¿Y quién, puede solicitarle, sería albergar prejuicios contra las niñas?

Sus padres.

No es de extrañar que los padres de niños pequeños a menudo son excitados por la idea de que sus hijos podrían ser dotados. De hecho, de todas las búsquedas de Google a partir “¿Es mi hijo de 2 años de edad,” la siguiente palabra más común es “dotado”. Pero esta pregunta no se hace igualmente sobre los niños y las niñas. Los padres son dos veces y media más probabilidades de preguntar “¿Es dotado mi hijo?” Que “Es mi hija dotado?” Los padres muestran un sesgo similar al utilizar otras frases relacionadas con la inteligencia de que puedan temor de decirlo en voz alta, como, “mi hijo es un genio?”



Son los padres den cuenta de las legítimas diferencias entre niños y niñas? Tal vez los niños pequeños son más propensos que las niñas a usar grandes palabras o de otra manera muestran signos objetivos de la superdotación? Nop. En todo caso, es todo lo contrario. A edades jóvenes, niñas consistentemente han demostrado tener un vocabulario más amplio y utilizar oraciones más complejas. En las escuelas estadounidenses, las niñas son un 9 por ciento más probabilidades que los niños estén en programas dotados. A pesar de todo esto, los padres que buscan alrededor de la mesa aparecen para ver chicos más dotados que en niñas.\*De hecho, en cada término de búsqueda relacionados con la inteligencia que probé, incluyendo las que indican su ausencia, los padres eran más propensos a estar preguntando por sus hijos en lugar de sus hijas. También hay más búsquedas de “es mi hijo detrás” o “tonto” que las búsquedas comparables para las hijas. Sin embargo, las búsquedas con palabras negativas como “detrás” y “estúpida” están sesgados menos específicamente hacia los hijos que las búsquedas con palabras positivas, tales como o “dotados” “genio”.

¿Cuáles son entonces las preocupaciones primordiales de los padres con respecto a sus hijas? En primer lugar, todo lo relacionado con la apariencia. Considere las preguntas sobre el peso de un niño. Los padres Google “Es mi hija sobrepeso?” Aproximadamente el doble de la frecuencia que Google “¿Es mi hijo tiene sobrepeso?” Los padres son dos veces más propensos a preguntar cómo llegar a sus hijas para bajar de peso, ya que son a preguntar cómo llegar a sus hijos a hacer lo mismo. Al igual que con talento, este sesgo de género no está basado en la realidad. Alrededor del 28 por ciento de las niñas tienen sobrepeso, mientras que el 35 por ciento de los niños son. A pesar de que las escalas miden los niños con sobrepeso que en las niñas, los padres ven o se preocupan acerca de las niñas con sobrepeso-mucho más frecuencia que los niños con sobrepeso.

Los padres también son una vez y media más probabilidades de preguntar si su hija es bello que si su hijo es guapo. Y son casi tres veces más propensos a preguntar si su hija es feo que si su hijo es feo. (¿Cómo se espera que Google para saber si un niño es bello o feo es difícil de decir.)

En general, los padres parecen más propensos a usar palabras positivas en preguntas sobre los hijos. Ellos son más propensos a preguntar si un hijo es “feliz” y menos propensos a preguntar si un hijo está “deprimido”.

liberales lectores pueden imaginar que estos errores son más frecuentes en las partes conservadoras del país, pero no se encontró ninguna evidencia de ello. De hecho, no he encontrado una relación significativa entre cualquiera de estos sesgos y la composición política o cultural de un estado. Tampoco hay evidencia de que estos sesgos han disminuido desde 2004, año en que los datos de búsqueda de Google es la primera disponible. Al parecer, este sesgo contra las niñas es más generalizada y profundamente arraigada de lo que había cuidado a creer.

El sexismo no es el único lugar en nuestros estereotipos sobre los prejuicios pueden ser apagado.

Vikingmaiden88 es de veinte y seis años de edad. A ella le gusta leer la historia y escribir poesía. Su cotización es la firma de Shakespeare. He recogido todo esto de su perfil y mensajes en Stormfront.org, odio sitio en línea más popular de Estados Unidos. También aprendí que

Vikingmaiden88 ha contado con el contenido en el sitio del periódico donde trabajo, el New York Times. Ella escribió un mensaje entusiasta sobre una característica particular Times.

Recientemente he analizado decenas de miles de tales perfiles del frente de tormenta, en la que los miembros registrados pueden introducir su ubicación, fecha de nacimiento, intereses y otras informaciones.

Stormfront fue fundada en 1995 por Don Negro, ex líder del Ku Klux Klan. Sus más populares “grupos sociales” son “Unión de los nacionalsocialistas” y “aficionados y seguidores de Adolf Hitler.” Durante el año pasado, según Quantcast, aproximadamente 200.000 a 400.000 estadounidenses visitaron el sitio cada mes. Un informe reciente del Southern Poverty Law Center vinculado casi un centenar de asesinatos en los últimos cinco años a los miembros del frente de tormenta registrados.

miembros del frente de tormenta no se quien lo hubiera imaginado.

Ellos tienden a ser jóvenes, al menos de acuerdo a las fechas de nacimiento de auto-reporte. La edad más común en la que la gente se une al sitio tiene diecinueve años. Y cuatro veces más de diecinueve años de edad se inscriban de cuarenta años de edad. Los usuarios de Internet y de redes sociales se inclinan joven, pero no casi tan joven.

Los perfiles no tienen un campo para el género. Pero miré a todos los puestos y perfiles completos de una muestra aleatoria de los usuarios de América, y resulta que se puede trabajar el género de la mayoría de los miembros: Calculo que alrededor del 30 por ciento de los miembros del frente de tormenta son mujeres.

Los estados con el mayor número de miembros per cápita son de Montana, Alaska y Idaho. Estos estados tienden a ser abrumadoramente blanco. ¿Quiere decir esto que crecer con poca diversidad fomenta el odio?

Probablemente no. Por el contrario, puesto que esos estados tienen una mayor proporción de los blancos no judíos, que tienen más miembros potenciales de un grupo que ataca Judíos y no blancos. El porcentaje de público objetivo de Stormfront que une es en realidad mayor en las zonas con más minorías. Esto es particularmente cierto cuando se mira a los miembros del frente de tormenta que son dieciocho años y el más joven y por lo tanto no se eligen a sí mismos en el que viven.

Entre este grupo de edad, California, un estado con una de las mayores poblaciones minoritarias, tiene una tasa de afiliación del 25 por ciento más alto que el promedio nacional.

Uno de los grupos sociales más populares en el sitio es “en apoyo del antisemitismo.” El porcentaje de miembros que se unen a este grupo se correlaciona positivamente con la población judía de un estado. Nueva York, el estado con la mayor población judía, tiene membresía superior a la media per cápita en este grupo.

En 2001, Dna88 unió Stormfront, describiéndose a sí mismo como un “buen aspecto, racialmente conscientes” de treinta años de edad, desarrollador de Internet que vive en “la ciudad Judio York.” En los próximos cuatro meses, escribió más de doscientos mensajes, como “judía crímenes contra la humanidad”y‘Blood Money judía’, y la gente dirigidos a un sitio web, jewwatch.com, que pretende ser una‘biblioteca académica’sobre‘la criminalidad sionista’.

miembros del frente de tormenta se quejan de las minorías que hablan idiomas diferentes y la comisión de delitos. Pero lo que me pareció más interesante fueron las quejas sobre la competencia en el mercado de las citas.

Un hombre que se hace llamar William Lyon Mackenzie King, después de un ex primer ministro de Canadá, que sugirió una vez que “Canadá debería seguir siendo el país del hombre blanco”, escribió en 2003 que se esforzaba por “contener” su “rabia” después de ver una mujer blanca “llevar alrededor de su media negro feo niglet mestizo “. en su perfil, Whitepride26, un estudiante de cuarenta y un años de edad, en Los Ángeles, dice: ‘No me gusta los negros, latinos, y en ocasiones los asiáticos, sobre todo cuando los hombres encuentran más atractivas’ que “una mujer blanca.”

Ciertos acontecimientos políticos juegan un papel. El día que registró el mayor incremento individual en calidad de miembro en la historia de Stormfront, con diferencia, fue de 5 noviembre de 2008, el día después de que Barack Obama fue elegido presidente. Hubo, sin embargo, no aumenta el interés por Stormfront durante la candidatura de Donald Trump y sólo una pequeña subida inmediatamente después de haber ganado. Trump montó una ola de nacionalismo blanco. No hay evidencia de que aquí se creó una ola de nacionalismo blanco.

La elección de Obama dio lugar a un aumento en el movimiento nacionalista blanco. La elección de Trump parece ser una respuesta a eso.

Una cosa que no parece tener importancia: la economía. No hubo relación entre el registro de membresía mensual y tasa de desempleo de un estado. Estados afectados de manera desproporcionada por la Gran Recesión no vieron aumento comparativo de las búsquedas de Google para Stormfront.

Pero quizás lo más interesante-y sorprendente-fueron algunos de los temas de conversación miembros del frente de tormenta tienen. Son similares a aquellos de mis amigos y yo hablamos acerca. Tal vez fue mi propia ingenuidad, pero me hubiera imaginado nacionalistas blancos que habitan un universo diferente de la de mis amigos y yo. En cambio, tienen largos hilos alabando Juego de Tronos y discutir los méritos comparativos de los sitios de citas en línea, como PlentyOfFish y OkCupid.

Y el hecho clave que muestra que los usuarios del frente de tormenta se habitan universos similares a gente como yo y mis amigos: la popularidad de The New York Times entre los usuarios del frente de tormenta. No se trata sólo VikingMaiden88 colgando alrededor del sitio Times. El sitio es popular entre muchos de sus miembros. De hecho, cuando se comparan los usuarios del frente de tormenta a las personas que visitan el sitio Yahoo News, resulta que la multitud Stormfront es dos veces más probable que visite nytimes.com.

Los miembros de un sitio de odio haberle sido notificado el nytimes.com oh-tan-liberal? ¿Cómo podría ser esto posible? Si un número considerable de miembros del frente de tormenta obtener sus noticias de nytimes.com, significa que nuestra sabiduría convencional sobre nacionalistas blancos es incorrecto. También significa nuestra sabiduría convencional acerca de cómo funciona el internet es incorrecto.

## LA VERDAD SOBRE LA INTERNET

---

El Internet, la mayoría todos están de acuerdo, está impulsando estadounidenses aparte, causando la mayoría de las personas a refugiarse en sitios dirigidos a personas como ellos. Así es como Cass Sunstein de la Escuela de Derecho de Harvard, describió la situación: “Nuestro mercado de las comunicaciones se está moviendo rápidamente [hacia una situación en la que] la gente se limitan a sus propios puntos de vista liberales viendo y leyendo en su mayoría o sólo los liberales; moderados, los moderados; conservadores, conservadores; Neonazis, los neonazis “.

Este punto de vista tiene sentido. Después de todo, el Internet nos da un número virtualmente ilimitado de opciones entre las que podemos consumir las noticias. Puedo leer lo que quiera. Usted puede leer lo que quiera. VikingMaiden88 puede leer lo que quiera. Y la gente, si se deja a sus propios dispositivos, tienden a buscar los puntos de vista que confirman lo que ellos creen. Por lo tanto, sin duda, el Internet debe ser la creación de segregación política extrema.

Hay un problema con este punto de vista estándar. Los datos nos dicen que simplemente no es cierto.

La evidencia en contra de esta pieza de la sabiduría convencional proviene de un estudio de 2011 por Matt Gentzkow y Jesse Shapiro, dos economistas cuyo trabajo hemos comentado anteriormente.

Gentzkow y Shapiro recogieron datos sobre el comportamiento de navegación de una amplia muestra de los estadounidenses. Su conjunto de datos incluyó también la ideología de auto-informó de sus temas: si la gente se consideraban más liberal o conservador. Ellos utilizan estos datos para medir la segregación política en Internet.

¿Cómo? Se realizó un experimento interesante.

Supongamos que se tomaron muestras al azar a dos estadounidenses que resultan tanto estar visitando la misma página web de noticias. ¿Cuál es la probabilidad de uno de ellos será liberal y otro conservador? ¿Con qué frecuencia, en otras palabras, hacer liberales y conservadores “se encuentran” en sitios de noticias?

Pensar en esto más, supongamos que los liberales y los conservadores en el Internet nunca consiguieron sus noticias en Internet desde el mismo lugar. En otras palabras, los liberales visitaron sitios web exclusivamente liberales, conservadores exclusivamente las conservadoras. Si este fuera el caso, las posibilidades de que dos estadounidenses en un sitio de noticias dado tienen puntos de vista opuestos políticos sería del 0 por ciento. El Internet sería perfectamente segregado. Liberales y conservadores nunca se mezclan.

Supongamos, por el contrario, que los liberales y los conservadores no difieren en absoluto de cómo llegaron sus noticias. En otras palabras, un liberal y un conservador eran igualmente propensos a visitar cualquier sitio de noticias en particular. Si este fuera el caso, las posibilidades de que dos estadounidenses en un sitio web de noticias dado tienen puntos de vista políticos opuestos serían aproximadamente el 50 por ciento. El Internet sería perfectamente desegregada. Liberales y conservadores se mezclan perfectamente.

Entonces, ¿qué nos dicen los datos? En los Estados Unidos, de acuerdo con Gentzkow y Shapiro, las posibilidades de que dos personas que visitan el mismo sitio de noticias tienen diferentes puntos de vista políticos es alrededor del 45 por ciento. En otras palabras, el Internet es mucho más cercana a la perfecta eliminación de la segregación de la segregación perfecta. Liberales y conservadores son “encuentro” entre sí en la web todo el tiempo.

Lo que realmente pone la falta de separación en internet en perspectiva es en comparación a la segregación en otras partes de nuestras vidas. Gentzkow y Shapiro pudieron repetir su análisis por diversas interacciones offline. ¿Cuáles son las posibilidades de que dos miembros de la familia tienen diferentes puntos de vista políticos? Dos vecinos? Dos colegas? ¿Dos amigos?

Utilizando datos de la Encuesta Social General, Gentzkow y Shapiro encontró que todos estos números fueron menores que las posibilidades de que dos personas en el mismo sitio web de noticias tienen diferentes políticas.

#### **Probabilidad de que alguien que ha conocido HA puntos de vista opuestos POLÍTICOS**

---

En un sitio de noticias	45,2%
compañero de trabajo	41,6%
Vecino Desconectado	40.3
Miembro de la familia	37%
Amigo	34,7%

En otras palabras, es más probable encontrarse con alguien con puntos de vista opuestos en línea que no esté conectado.

¿Por qué no es el Internet más segregadas? Hay dos factores que limitan la segregación política en Internet.

En primer lugar, algo sorprendente, la industria de las noticias de Internet está dominado por unos pocos sitios masivos. Por lo general pensamos de la Internet como apelando a la periferia. De hecho, hay sitios para todos, independientemente de sus puntos de vista. Hay lugares para el aterrizaje favor de las armas y anti-armas cruzados, los derechos de cigarros y activistas moneda del dólar, anarquistas y nacionalistas blancos. Pero estos sitios en conjunto representan una pequeña fracción del tráfico de noticias de internet. De hecho, en 2009, cuatro sitios-Yahoo News, noticias de AOL, msnbc.com, y cnn.com-recolectaron más de la mitad de los puntos de vista de las noticias. Yahoo News sigue siendo el sitio de noticias más popular entre los estadounidenses, con cerca de 90 millones de visitantes únicos mensuales -o algunos 600 veces la audiencia de Stormfront. sitios de medios masivos como Yahoo Noticias apelación a un público amplio políticamente diversa,.

La segunda razón por la Internet no es todo lo que es segregada que muchas personas con fuertes opiniones políticas visitan los sitios del punto de vista opuesto, aunque sólo sea para enojarse y discutir. adictos a la política no se limitan únicamente a los sitios dirigidos a ellos. Alguien que visita ThinkProgress.org y moveon.org-dos sitios extremadamente liberales es-es

más probable que el usuario medio de Internet para visitar foxnews.com, un sitio inclinada hacia la derecha. Alguien que visita rushlimbaugh.com o glennbeck.com-dos sitios es extremadamente conservadores más probable que el usuario medio de Internet para visitar nytimes.com, un sitio más liberal.

Gentzkow y de Shapiro estudio se basó en los datos en 2004-09, relativamente temprano en la historia de Internet. Internet podría haber crecido más compartimentada desde entonces? Haga que los medios de comunicación social y, en particular, Facebook alterado su conclusión? Está claro que si nuestros amigos tienden a compartir nuestros puntos de vista políticos, el aumento de las redes sociales debe significar un aumento de cámaras de eco. ¿Derecha?

Una vez más, la historia no es tan simple. Si bien es cierto que los amigos de la gente en Facebook son más propensos que los que no comparten sus puntos de vista políticos, un equipo de los datos científicos-Eytan bakshy, Salomón Messing y Lada Adamic, han encontrado que una cantidad sorprendente de la información de la gente consigue en Facebook viene de las personas con puntos de vista opuestos.

¿Cómo puede ser esto? No nuestros amigos tienden a compartir nuestros puntos de vista políticos? De hecho, lo hacen. Pero hay una razón crucial que Facebook puede conducir a una mayor diversidad discusión política que fuera de línea socialización. La gente, en promedio, tienen sustancialmente más amigos en Facebook que lo hacen fuera de línea. Y estos vínculos débiles facilitados por Facebook son más propensos a ser personas con puntos de vista políticos opuestos.

En otras palabras, Facebook nos expone a débil conocimiento social, conexiones de la escuela secundaria, el tercer primo loco, el amigo del amigo del amigo que clase de, clase de, tal vez sabe. Estas son personas que nunca podría jugar a los bolos con o para una barbacoa con. Es posible que no invitarles a una cena. Pero lo hace amigo de Facebook ellas. Y usted ve sus enlaces a artículos con vistas que nunca podría haber considerado lo contrario.

En suma, el Internet realmente une a las personas de diferentes puntos de vista políticos juntos. El liberal promedio puede pasar su mañana con su marido liberal y niños liberales; su tarde con sus compañeros de trabajo liberales; su viaje rodeado de liberales pegatinas; la noche con sus compañeros de clase de yoga liberales. Cuando ella llega a casa y peruses algunos comentarios conservadores en cnn.com o recibe un vínculo de Facebook de su conocido republicano de la escuela secundaria, esto puede ser su mayor exposición conservadora del día.

Probablemente nunca encuentro nacionalistas blancos en mi cafetería favorita en Brooklyn. Pero VikingMaiden88 y yo frecuentan el sitio del New York Times.

## **LA VERDAD sobre el abuso infantil Y ABORTO**

---

El Internet nos puede dar una visión de no sólo actitudes perturbadoras, sino también los comportamientos perturbadores. De hecho, los datos de Google pueden ser eficaces para

alertarnos a las crisis que se pierden por todas las fuentes habituales. La gente, después de todo, a su vez a Google cuando están en problemas.

Considere el maltrato infantil durante la Gran Recesión.

Cuando esta gran recesión económica comenzó a finales de 2007, muchos expertos eran naturalmente preocupados por el efecto que podría tener en los niños. Después de todo, muchos padres se estresan y deprimido, y estos son los principales factores de riesgo de maltrato. El abuso infantil podría dispararse.

A continuación, los datos oficiales entraron, y parecía que la preocupación era infundada. Niño agencias de servicios de protección informaron que estaban recibiendo un menor número de casos de abuso. Además, estas gotas eran más grande de los estados que fueron más afectados por la recesión. “Las predicciones de fatalidad-and-penumbra no se han hecho realidad,” Richard Gelles, un experto en bienestar infantil en la Universidad de Pensilvania, dijo a la Associated Press en 2011. Si, como contrario a la intuición, ya que puede haber parecido, el abuso infantil parecía tener caído en picado durante la recesión.

Pero el abuso infantil realmente gota con tantos adultos sin trabajo y extremadamente angustiados? Tenía problemas para creer esto. Así que di vuelta a los datos de Google.

Resulta que algunos niños hacen algunas búsquedas trágicas y desgarradoras, en Google, tales como “mi mamá me pegaba” o “mi padre me golpeó.” Y estas búsquedas presentar una imagen angustiosa diferente y de lo que sucedió durante esta vez. El número de búsquedas como esta se disparó durante la Gran Recesión, siguiendo de cerca la tasa de desempleo.

Esto es lo que creo que pasó: era la notificación de casos de abuso infantil que declinaron, no el abuso infantil. Después de todo, se estima que sólo un pequeño porcentaje de casos de abuso infantil son reportados a las autoridades de todos modos. Y durante una recesión, muchas de las personas que tienden a reportar casos de abuso infantil (maestros y oficiales de policía, por ejemplo) y manejar los casos (protección infantil trabajadores del servicio) son más propensos a ser con exceso de trabajo o fuera del trabajo.

Hubo muchas historias durante la recesión económica de las personas que tratan de informar de los casos posibles de cara a largos tiempos de espera y renunciar.

De hecho, hay más pruebas, esta vez no desde Google, que el abuso infantil en realidad aumentó durante la recesión. Cuando un niño muere debido al abuso o negligencia tiene que ser informado. Tales muertes, aunque rara, sí aumentó en los estados que fueron más afectados por la recesión.

Y hay alguna evidencia de Google que más personas sospechaban el abuso en las áreas más afectadas. El control de las tasas anteriores a la recesión y las tendencias nacionales, los estados que habían sufrido comparativamente las mayores tasas de búsqueda más tenido un abuso y abandono infantil. Por cada aumento de un punto porcentual en la tasa de desempleo, hubo un aumento del 3 por ciento asociada de la tasa de búsqueda de “abuso de menores” o “abandono de los niños.” Presumiblemente, la mayoría de estas personas nunca reportado con éxito el abuso, ya que estos estados tenían el mayor gotas en la presentación de informes.

Las búsquedas por el sufrimiento aumentan niños. La tasa de mortalidad infantil pico. Las búsquedas por personas que sospechan abuso suben en los estados más afectados. Sin embargo, la presentación de informes de los casos se pone. Una recesión parece ser la causa más niños a decirle a Google que sus padres están golpeando o golpearlos y más personas a sospechar que ven el abuso. Sin embargo, las agencias de exceso de trabajo son capaces de manejar un menor número de casos.

Creo que es seguro decir que la gran recesión hizo abuso de niños es peor, a pesar de las medidas tradicionales no lo demostró.

En cualquier momento sospecho personas pueden estar sufriendo de los libros ahora, me dirijo a los datos de Google. Uno de los beneficios potenciales de estos nuevos datos, y saber cómo interpretarlo, es la posibilidad de ayudar a las personas vulnerables que podrían pasar por alto por las autoridades.

Por eso, cuando el Tribunal Supremo estaba buscando hace poco sobre los efectos de las leyes por lo que es más difícil conseguir un aborto, di vuelta a los datos de consulta. Sospeché mujeres afectadas por esta legislación podría estudiar la manera off-the-libros de interrumpir un embarazo. Lo hicieron. Y estas búsquedas fueron más altas en los estados que habían aprobado leyes que restringen abortos.

Los datos de búsqueda aquí es a la vez útil y preocupante.

En 2015, en los Estados Unidos, había más de 700.000 búsquedas de Google que buscan en abortos autoinducidos. En comparación, hubo alrededor de 3,4 millones de búsquedas para las clínicas de aborto de ese año. Esto sugiere que un porcentaje significativo de mujeres que consideran el aborto han contemplado hacerlo ellos mismos.

Las mujeres buscaron, alrededor de 160.000 veces, de maneras de conseguir píldoras abortivas a través canales- no oficial “comprar píldoras abortivas en línea” y “píldoras abortivas libres.” Se pidió a Google sobre el aborto por las hierbas como el perejil o por la vitamina C. Había alrededor de 4.000 búsquedas que buscan para obtener instrucciones sobre abortos percha, incluyendo alrededor de 1.300 para la frase exacta “cómo hacer un aborto percha.” también había unos pocos cientos de mirar en aborto a través de blanqueo de una de útero y la perforación de uno de estómago.

Lo que impulsa el interés en aborto autoinducido? La geografía y el calendario de las búsquedas de Google apuntan a un posible culpable: cuando es difícil conseguir un aborto oficial, las mujeres buscan en los enfoques off-the-libros.

las tasas de la búsqueda de aborto autoinducido eran bastante estable desde 2004 hasta 2007. Ellos comenzaron a subir a finales de 2008, coincidiendo con la crisis financiera y la recesión que siguió. Se llevaron un gran salto en 2011, saltando del 40 por ciento. El Instituto Guttmacher, una organización de derechos de reproducción, señala a 2011 como el comienzo de la reciente represión del país sobre el aborto; se promulgaron noventa y dos disposiciones estatales que restringen el acceso al aborto. Mirando por la comparación en Canadá, que no ha



visto una ofensiva contra los derechos reproductivos, no hubo un aumento comparable en las búsquedas de abortos autoinducidos durante este tiempo.

El estado con el mayor índice de búsquedas de Google para abortos autoinducidos es Mississippi, un estado con aproximadamente tres millones de personas y, ahora, sólo una clínica de aborto. Ocho de los diez estados con las tasas más altas de búsqueda para abortos autoinducidos son considerados por el Instituto Guttmacher ser hostil o muy hostil al aborto. Ninguno de los diez estados con las tasas más bajas de búsqueda para el aborto autoinducido están en una u otra categoría.

Por supuesto, no podemos saber de Google busca en cuántas mujeres con éxito se dan abortos, pero la evidencia sugiere que un número significativo de mayo. Una manera de iluminar este es comparar los datos de aborto y el parto.

En 2011, el último año con datos completa aborto a nivel estatal, las mujeres que viven en estados con pocas clínicas de aborto tenía muchos menos abortos legales.

Comparación de los diez estados con la mayoría de las clínicas de aborto per cápita (una lista que incluye a Nueva York y California) a los diez estados con la menor cantidad de las clínicas de aborto per cápita (una lista que incluye Mississippi y Oklahoma). Las mujeres que viven en estados con las clínicas de aborto menor cantidad tenían 54 por ciento menos de abortos legales- una diferencia de once abortos por cada mil mujeres entre las edades de quince y cuarenta y cuatro. Las mujeres que viven en los estados con las clínicas de aborto también tuvieron menor número de nacimientos vivos más. Sin embargo, la diferencia no fue suficiente para compensar el menor número de abortos. Había seis más nacidos vivos por cada mil mujeres en edad fértil.

En otras palabras, no parece que haya habido algunos embarazos que faltan en algunas partes del país en el que era más difícil de conseguir un aborto. Las fuentes oficiales no nos dicen lo que les pasó a esos cinco nacimientos que faltan por cada mil mujeres en los estados en los que es difícil conseguir un aborto.

Google ofrece algunas buenas pistas.

No podemos confiar ciegamente en los datos del gobierno. El gobierno nos puede decir que el abuso infantil o el aborto ha caído y los políticos puede celebrar este logro. Pero los resultados que piensan que estamos viendo puede ser un artefacto de fallas en los métodos de recolección de datos. La verdad puede ser diferente y, a veces, mucho más oscuro.

### [La verdad sobre sus amigos de Facebook](#)

---

Este libro trata de grandes volúmenes de datos, en general. Sin embargo, este capítulo ha hecho hincapié en su mayoría las búsquedas de Google, que he sostenido revelar un mundo oculto muy diferente de la que creemos ver. También lo son otras fuentes de datos grandes suero de la verdad como digital, así? El hecho es que muchas fuentes de datos grandes, como Facebook, son a menudo lo contrario de suero de la verdad digital.

En las redes sociales, como en las encuestas, no tiene incentivo para decir la verdad. En las redes sociales, mucho más que en las encuestas, tiene un gran incentivo para hacer quedar bien. Su presencia en línea no es anónima, después de todo. Usted está cortejando a una audiencia y decirle a sus amigos, familiares, colegas, conocidos y desconocidos que se encuentre.

Para ver cómo los datos extraídos de las redes sociales puede ser sesgada, tenga en cuenta la relativa popularidad del *Atlántico*, una revista mensual respetado, culta, frente al *National Enquirer*, una chismosa, la revista menudo-sensacional. Ambas publicaciones tienen circulaciones promedio similares, la venta de unos pocos cientos de miles de ejemplares. (The *National Enquirer* es una semana, por lo que realmente vende más copias totales.) También hay un número comparable de las búsquedas de Google para cada revista.

Sin embargo, en Facebook, aproximadamente 1,5 millones de personas, ya sea como el *Atlántico* o discutir los artículos del *Atlántico* en sus perfiles. Sólo alrededor de 50.000 como el *Enquirer* o analizar su contenido.

### ***ATLÁNTICOVS. National Enquirer* POPULARIDAD comparado por DIFERENTES FUENTES**

---

Circulación	Aproximadamente 1 del <i>Atlántico</i> por cada 1 <i>National Enquirer</i>
-------------	--

Búsquedas de Google	1 <i>Atlántico</i> por cada 1 <i>National Enquirer</i>
---------------------	--

Le gustan de Facebook	27 <i>Atlántico</i> por cada 1 <i>National Enquirer</i>
-----------------------	---

Para la evaluación de popularidad revista, datos de circulación es la realidad del terreno. Google los datos se aproxima a ella. Y Facebook de datos está sesgada aplastante contra el tabloide de mala calidad, por lo que es el peor de los datos para determinar lo que la gente realmente le gusta.

Y al igual que con las preferencias de lectura, por lo que con la vida. En Facebook, mostramos nuestro ser cultivadas, y no nuestro verdadero ser. Yo uso los datos de Facebook en este libro, de hecho, en este capítulo, pero siempre con esta advertencia en mente.

Para obtener una mejor comprensión de lo que se pierde de medios sociales, volvamos a la pornografía por un momento. En primer lugar, tenemos que hacer frente a la creencia común de que Internet está dominada por el carbón. Esto no es cierto. La mayoría de los contenidos en Internet es nonpornographic. Por ejemplo, de los diez sitios web más visitados, ninguno es pornográfico. Así que la popularidad de la pornografía, mientras que enormes, no debe ser exagerada.

Sin embargo, dicho esto, tomar un vistazo de cerca a lo que nos gusta y compartir pornografía deja claro que Facebook, Instagram y Twitter sólo proporcionan una ventana limitada de lo que

es realmente popular en Internet. Hay grandes subconjuntos de la web que operan con gran popularidad pero poca presencia social.

El video más popular de todos los tiempos, a partir de este escrito, es Psy de “Gangnam Style”, un video de la música pop tonto que satiriza coreanos de moda. Se ha visto alrededor de 2,3 mil millones de veces sólo en YouTube desde su debut en 2012. Y su popularidad es clara, no importa en qué sitio se encuentra. Se ha compartido a través de diferentes plataformas de medios sociales decenas de millones de veces.

El video pornográfico más popular de todos los tiempos puede ser “gran cuerpo, gran sexo, gran mamada.” Ha sido visto más de 80 millones de veces. En otras palabras, por cada treinta puntos de vista de “Gangnam Style”, se ha producido alrededor de al menos una vista de “gran cuerpo, gran sexo, gran mamada.” Si los medios sociales nos dio una visión precisa de los videos de personas vieron “Gran cuerpo, gran sexo, gran mamada” debe ser publicado millones de veces. Sin embargo, este video ha sido compartida en los medios de comunicación social sólo unas pocas docenas de veces y siempre por las estrellas porno, y no por los usuarios promedio. La gente claramente no sienten la necesidad de anunciar su interés en este video a sus amigos.

Facebook es digital presumir-a-mis-amigos-sobre-cómo-buena-mi-vida-es suero. En Facebook mundo, el adulto promedio parece estar felizmente casada, de vacaciones en el Caribe, y leyendo el Atlántico. En el mundo real, una gran cantidad de gente está enojada, en las líneas de cajas de los supermercados, mira a escondidas en el National Enquirer, haciendo caso omiso de las llamadas telefónicas de su cónyuge, el cual no han dormido con en años. En Facebook mundo, la vida familiar parece perfecto. En el mundo real, la vida familiar es desordenado. En ocasiones puede ser tan complicado que un pequeño número de personas, incluso lamentan tener hijos. En Facebook mundo, parece que cada adulto joven está en una fiesta fresca sábado por la noche. En el mundo real, la mayoría son solo en casa, espectáculos binge-viendo en Netflix. En Facebook mundo, una novia publica veintiséis imágenes felices de su escapada con su novio. En el mundo real,

VERDAD DIGITAL	MENTIRAS DIGITAL
<ul style="list-style-type: none"><li>• Búsquedas</li><li>• Puntos de vista</li><li>• Los clics</li><li>• Swipes</li></ul>	<ul style="list-style-type: none"><li>• mensajes de medios de comunicación social</li><li>• Le gusta los medios de comunicación social</li><li>• Los perfiles de citas</li></ul>

## [LA VERDAD ACERCA DE SUS CLIENTES](#)

---

En la madrugada del 5 de septiembre de 2006, Facebook introdujo una importante actualización de su página de inicio. Las primeras versiones de Facebook sólo habían permitido a los usuarios

a hacer clic en los perfiles de sus amigos para aprender lo que estaban haciendo. El sitio web, considerado un gran éxito, tenía en el momento 9,4 millones de usuarios.

Pero después de meses de duro trabajo, los ingenieros habían creado algo que ellos llamaron “News Feed”, que proporcionaría a los usuarios información actualizada sobre las actividades de todos sus amigos.

Los usuarios informaron de inmediato que odiaban News Feed. Ben Parr, un estudiante de Northwestern, creó “Estudiantes contra Facebook suministro de noticias.” Dijo que “servicio de noticias es demasiado espeluznante, también acosador-esque, y una característica que tiene que ir.” A los pocos días, el grupo tenía 700.000 haciéndose eco de los miembros el sentimiento de Parr. Una Universidad de Michigan menor al diario de Michigan, “Estoy muy descolocada por el nuevo Facebook. Me hace sentir como un acosador “.

David Kirkpatrick cuenta esta historia en su cuenta autorizada de la historia de la página web, El efecto Facebook: La historia interna de la empresa que es conectar el mundo. Se dobla la introducción de News Feed “la mayor crisis de Facebook se ha enfrentado jamás.” Pero Kirkpatrick informa que cuando entrevistó a Mark Zuckerberg, cofundador y director de la empresa de rápido crecimiento, el CEO se inmutó.

¿La razón? Zuckerberg tenía acceso a suero de la verdad digitales: Los números en los clics y visitas a Facebook de la gente. Como Kirkpatrick escribe:

Zuckerberg, de hecho, sabía que la gente le gusta el News Feed, no importa lo que decían en los grupos. Tenía los datos que lo demuestran. La gente se pasan más tiempo en Facebook, en promedio, que antes de News Feed puso en marcha. Y estaban haciendo más allí, dramáticamente más. En agosto, los usuarios han visualizado 12 mil millones de páginas en el servicio. Pero en octubre, con Feed Noticias en curso, que consideraban 22 mil millones.

Y eso no era toda la evidencia a disposición de Zuckerberg. Incluso la popularidad viral del grupo de alimentación anti-News era evidencia del poder de Noticias. El grupo fue capaz de crecer tan rápidamente, precisamente debido a que muchas personas habían oído que sus amigos se habían unido y se enteraron de esto a través de su Servicio de Noticias.

En otras palabras, mientras que las personas se unían en una protesta pública grande sobre lo infeliz que estaban a punto de ver todos los detalles de sus vidas amigos en Facebook, que venían de vuelta a Facebook para ver todos los detalles de sus amigos vidas. News Feed se quedó. Facebook ahora tiene más de mil millones de usuarios activos al día.

En su libro de cero a uno, Peter Thiel, de los primeros inversores en Facebook, dice que las grandes empresas se basan en secretos, ya sea secretos acerca de la naturaleza o secretos acerca de las personas. Jeff Seder, como se discute en el Capítulo 3, se encontró el secreto natural que dejó el tamaño del ventrículo predijo el rendimiento del caballo. Google ha encontrado el secreto natural de lo poderoso que la información contenida en los enlaces puede ser.

Thiel define “secretos acerca de las personas” como “cosas que la gente no sabe acerca de ellos mismos o de las cosas que se ocultan porque no quieren que los demás sepan.” Este tipo de negocios, en otras palabras, se basan en mentiras de la gente.

Se podría argumentar que todos Facebook se funda en un secreto desagradable acerca de las personas que Zuckerberg aprendió mientras que en Harvard. Zuckerberg, al comienzo de su segundo año, creó un sitio web para sus compañeros de estudios llamado Facemash. Siguiendo el modelo de un sitio llamado “Am I Sexy o no ?,” Facemash presentaría imágenes de dos estudiantes de Harvard y luego otros estudiantes han juzgar que estaba más guapo.

El sitio del segundo año fue recibida con indignación. El Harvard Crimson, en un editorial, acusó joven Zuckerberg de “catering a la peor cara” de las personas. grupos hispanos y afroamericanos lo acusan de sexismo y el racismo. Sin embargo, antes de que los administradores de Harvard cerrar el acceso a Internet, sólo de Zuckerberg unas pocas horas después de que el sitio fue fundado-450 personas habían visitado el sitio y votado 22.000 veces en diferentes imágenes. Zuckerberg había aprendido un secreto importante: las personas pueden afirmar que están furiosos, pueden condenar algo tan desagradable, y sin embargo, todavía van a hacer clic.

Y aprendió una cosa más: para todas sus profesiones de seriedad, responsabilidad y respeto por la privacidad, la gente, incluso los estudiantes de Harvard de los demás, tenía un gran interés en la evaluación de las miradas de la gente. Los puntos de vista y los votos que le dijeron. Y más tarde, ya que resultó ser demasiado controvertido Facemash-tomó este conocimiento de cómo las personas interesadas podrían estar en hechos superficiales por los demás que tipo de lo conocieron y lo controlaron en la empresa de mayor éxito de su generación.

Netflix aprendió una lección similar al principio de su ciclo de vida: no confían en lo que la gente te dice; confían en lo que hacen.

Originalmente, la empresa permitió a los usuarios crear una cola de las películas que quería ver en el futuro, pero no tienen tiempo para el momento. De esta manera, cuando tenían más tiempo, Netflix podría recordarles de esas películas.

Sin embargo, Netflix notó algo extraño en los datos. Los usuarios estaban llenando sus colas con un montón de películas. Sin embargo, días más tarde, cuando se les recordaba de las películas en la cola, que rara vez se hace clic.

¿Cual fue el problema? Pedir a los usuarios lo que las películas que planean ver en pocos días, y van a llenar la cola con las películas que se aspira, intelectuales, tales como documentales de la Segunda Guerra Mundial blanco y negro o películas extranjeras graves. Unos días más tarde, sin embargo, van a querer ver las mismas películas que por lo general quieren ver: comedias románticas o películas lowbrow. Las personas estaban mintiendo constantemente a sí mismos.

Ante esta disparidad, Netflix dejó de pedir a la gente para decirles lo que querían ver en el futuro y comenzaron a construir un modelo basado en millones de clics y vistas de clientes similares. La compañía comenzó a saludar a sus usuarios con listas sugeridas de películas basadas no en lo que decía estar a favor sino en lo que los datos dijeron que era probable que ver. El resultado: clientes visitaron Netflix con más frecuencia y observaron más películas.

“Los algoritmos que saben mejor que tú mismo”, dice Xavier Amatriain, ex científico de datos en Netflix.

#### THE OUTSIZE VALUE OF IGNORING WHAT PEOPLE TELL YOU

WHAT PEOPLE SAY	REALITY	IPSO FACTO ...
They don't want to stalk their friends.	There is little in this world they want more than to keep up with and judge their friends.	Mark Zuckerberg, cofounder of Facebook, is worth \$55.2 billion.
They don't want to buy products that are produced in sweatshops.	They will buy nice, "reasonably priced" products.	Phil Knight, cofounder of Nike, is worth \$25.4 billion.
They want to listen to news in the morning.	They want to hear about midgets having sex with porn stars in the morning.	Howard Stern is worth \$500 million.
They have no interest in reading about bondage, dominance, and sadomasochism.	They want to read about BDSM between a young college graduate and a business magnate.	<i>50 Shades of Gray</i> has sold 125 million copies.
They want politicians to outline their policy positions.	They want politicians to spare them the details but seem tough and self-assured.	Donald Trump

### ¿PODEMOS CON LA VERDAD?

Usted puede encontrar partes de este capítulo deprimente. suero de la verdad Digital ha revelado un interés permanente en juzgar a las personas basándose en su apariencia; la continuación de la existencia de millones de hombres homosexuales encerrados; un porcentaje significativo de las mujeres fantasear acerca de la violación; animosidad generalizada contra los afroamericanos; un abuso infantil oculto y la crisis aborto autoinducido; y un brote de rabia islamófobo violento que sólo se agravó cuando el presidente hizo un llamamiento a la tolerancia. cosas no exactamente alegre. A menudo, después de dar una charla en mi investigación, la gente viene a mí y me dicen, "Seth, todo es muy interesante. Pero es tan deprimente ".

No puedo pretender que no hay una oscuridad en algunos de estos datos. Si la gente nos dice constantemente lo que creen que queremos oír, por lo general nosotros se nos diga lo que es más reconfortante que la verdad. suero de la verdad digital, en promedio, nos muestran que el mundo es peor de lo que hemos pensado.

Qué necesitamos saber esto? Aprender acerca de las búsquedas de Google, datos de pornografía, y que hace clic en lo que no podría hacer pensar, "Esto es muy bueno. Podemos entender lo que realmente somos."Es posible que en lugar de pensar:"Esto es horrible. Podemos entender lo que realmente somos ".

Pero la verdad ayuda, y no sólo para Mark Zuckerberg o demás en busca de atraer clics o clientes. Hay al menos tres formas en que este conocimiento puede mejorar nuestras vidas.

En primer lugar, no puede haber consuelo en saber que no está solo en sus inseguridades y el comportamiento vergonzoso. Puede ser bueno saber que otros son inseguros sobre sus cuerpos. Es probable que sea agradable para muchas personas, especialmente aquellos que no están teniendo tanto sexo-conocer todo el mundo no está fornicando como conejos. Y puede ser valiosa para un chico de secundaria en Mississippi que ha chocado en el quarterback saber que, a pesar del bajo número de hombres abiertamente homosexuales a su alrededor, un montón de otros sienten el mismo tipo de atracción.

Hay otra zona y uno que aún no he discutido en las búsquedas de Google puede ayudar a mostrar que no está solo. Cuando eras joven, un profesor puede haber dicho que, si usted tiene una pregunta, debe levantar la mano y pedir, porque si usted está confundido, otros son, también. Si fueras como yo, que ignoró el consejo de su profesor y se sentó en silencio, miedo de abrir la boca. Sus preguntas eran demasiado tonto, pensaste; todos los demás eran más profundas. Los datos de Google, recopilados de forma anónima nos pueden decir de una vez por todas cómo la derecha fueron nuestros maestros. Un montón de preguntas básicas, sub-profundas están al acecho en otras mentes, también.

Considere las preguntas más estadounidenses tuvieron durante 2014 el Estado de la Unión de Obama. (Ver la foto de color al final del libro.)

## **USTED no es único que se pregunta: TOP en Google PREGUNTAS DURANTE EL ESTADO DE LA UNIÓN**

---

¿Qué edad es Obama?

¿Quién está sentado junto a Biden?

¿Por qué es Boehner puesta una corbata verde?

¿Por qué es de color naranja Boehner?

Ahora, es posible leer estas preguntas y pensar que hablan mal de nuestra democracia. Para estar más preocupado por el color de la corbata de alguien o su tono de piel en lugar del contenido del discurso del presidente no refleja bien en nosotros. No saber que John Boehner, el presidente de la Cámara de Representantes, es también no dice mucho para nuestro compromiso político.

Yo prefiero en lugar de pensar en cuestiones tales como la demostración de la sabiduría de nuestros maestros. Estos son los tipos de preguntas que la gente por lo general no plantean, porque suenan demasiado tonto. Pero mucha gente ellos- y Google tienen.

De hecho, creo que grandes volúmenes de datos puede dar una actualización de-siglo XXI a una famosa cita de autoayuda: “Nunca compare a su interior a los exteriores de todos los demás.”

Una actualización de datos grandes puede ser: “Nunca comparar las búsquedas de Google a los mensajes de los medios sociales de todos los demás.”

Compárese, por ejemplo, la forma en que las personas describen sus maridos en las redes sociales públicos y en las búsquedas anónimas.

## **PERSONAS Mejores formas de describir su ESPOSOS**

Medios de Comunicación Social PUESTOS	BÚSQUEDAS
el mejor	homosexual
mi mejor amigo	un idiota
increíble	increíble
la mayor	molesto
muy lindo	media

Como vemos los mensajes de medios sociales de otras personas, pero no sus búsquedas, que tienden a exagerar el número de mujeres piensan constantemente sus maridos son “los mejores”, “el más grande”, y “tan lindo”. Tenemos la tendencia a minimizar el número de las mujeres piensan que sus maridos son “un tirón”, “molesto” y “significa”. Mediante el análisis de los datos anónimos y agregados, todos podemos entender que no somos los únicos que se encuentran el matrimonio y la vida, difícil. Podemos aprender a dejar de comparar nuestras búsquedas de mensajes de medios sociales de todos los demás.

El segundo beneficio del suero de la verdad digital es que nos alerta a las personas que están sufriendo. La Campaña de Derechos Humanos me ha pedido que trabajar con ellos para ayudar a educar a los hombres en ciertos estados de la posibilidad de salir del armario. Ellos están buscando a utilizar los datos de búsqueda anónimos y agregados de Google para ayudarles a decidir dónde es mejor orientar sus recursos. Del mismo modo, las agencias de servicios de protección infantil han puesto en contacto conmigo para conocer en qué partes del país puede ser mucho más abuso de menores que están grabando.

Uno de los temas de extrañar que también se puso en contacto sobre: los olores vaginales. Cuando empecé a escribir sobre esto en el New York Times, de todos los lugares, lo hacía en un tono irónico. La sección me hizo, y otros, risa.

Sin embargo, cuando más tarde explorado algunos de los paneles de mensajes que surgen cuando alguien hace estas búsquedas incluyen numerosos puestos de chicas jóvenes convencidos de que sus vidas se arruinaron debido a la ansiedad sobre el olor vaginal. No es un chiste. expertos en educación sexual me han contactado, preguntando cómo pueden incorporar algunos de los mejores datos de Internet para reducir la paranoia entre las jóvenes.

Aunque me siento un poco fuera de mi fondo sobre todas estas cuestiones, que son serios, y creo que la ciencia de datos puede ayudar.

El más poderoso-valor final y, creo que, en este suero de la verdad digital es de hecho su capacidad para conducirnos de los problemas a las soluciones. Con una mayor comprensión, podemos encontrar maneras de reducir la oferta mundial de actitudes desagradables.

Volvamos al discurso de Obama sobre la islamofobia. Recordemos que cada vez que Obama argumentó que las personas deben respetar los musulmanes más, la misma gente que estaba tratando de alcanzar hicieron más enfurecido.

las búsquedas de Google, sin embargo, revelan que hubo una línea que no desencadenar el tipo de respuesta entonces presidente Obama hubiera querido. El dijo, “los musulmanes



estadounidenses son nuestros amigos y nuestros vecinos, nuestros compañeros de trabajo, nuestros héroes deportivos y, sí, son nuestros hombres y mujeres uniformados, que están dispuestos a morir en defensa de nuestro país.”

Después de esta línea, por primera vez en más de un año, la parte superior sustantivo Googled después de “musulmán” no era “terroristas”, “extremistas” o “refugiados”. Fue “atletas”, seguido de “soldados”. Y en efecto, los “atletas” mantienen el primer lugar durante un día completo después.

Cuando sermoneamos enojo de la gente, la búsqueda de datos implica que su furia puede crecer. Pero sutilmente provocar la curiosidad de la gente, dando nueva información, y ofreciendo nuevas imágenes del grupo que está alimentando su rabia puede convertir sus pensamientos en diferentes direcciones, más positivos.

Dos meses después de que el discurso original, Obama dio otro discurso televisado el islamofobia, esta vez en una mezquita. Tal vez alguien en la oficina del presidente había leído de mi columna Times, que discutió lo que había funcionado y lo que no SOLTAS y. Por el contenido de este discurso fue notablemente diferente.

Obama pasó poco tiempo insistiendo en el valor de la tolerancia. En su lugar, se centró mayoritariamente en provocar la curiosidad de la gente y cambiar sus percepciones de los musulmanes estadounidenses. Muchos de los esclavos de África eran musulmanes, Obama nos dijo; Thomas Jefferson y John Adams tenían sus propias copias del Corán; la primera mezquita en suelo estadounidense fue en Dakota del Norte; un musulmán estadounidense rascacielos en Chicago diseñado. Obama volvió a hablar de los atletas musulmanes y miembros de servicios armados, pero también habló de los agentes de policía y los bomberos musulmanes, profesores y médicos.

Y mi análisis de las búsquedas de Google sugiere que este discurso fue más exitoso que el anterior. Muchas de las búsquedas de odio, rabiosos contra los musulmanes cayeron en las primeras horas después del discurso del presidente.

Hay otras maneras posibles de utilizar los datos de búsqueda para aprender lo que provoca, o reduce, el odio. Por ejemplo, podemos ver cómo cambian las búsquedas racista después de un quarterback negro se redactó en una ciudad o cómo sexista búsquedas cambian después de una mujer es elegido para un cargo. Podríamos ver cómo el racismo responde a la policía de proximidad o de cómo el sexismo responde a las nuevas leyes de acoso sexual.

Aprender de nuestros prejuicios subconscientes también puede ser útil. Por ejemplo, todos podríamos hacer un esfuerzo extra para deleitar en la mente de niñas pequeñas y muestran una menor preocupación por su apariencia. Google datos de búsqueda y otras fuentes de la verdad en el Internet nos dan una visión sin precedentes de los rincones más oscuros de la psique humana. Esto es, a veces, lo admito, difícil de enfrentar. Pero también puede resultar beneficioso. Podemos utilizar los datos para luchar contra la oscuridad. La recopilación de datos ricos en los problemas del mundo es el primer paso hacia la fijación de ellos.

## ZOOM DE

**METRO**y hermano, Noé, es cuatro años menor que I. La mayoría de la gente, al primer encuentro con nosotros, nos encuentre extrañamente similar. Ambos hablar en voz demasiado alta, se están quedando calvos de la misma manera, y tienen gran dificultad para mantener nuestros apartamentos ordenado.

Pero hay diferencias: Cuento peniques. Noah compra el mejor. Me encanta Leonard Cohen y Bob Dylan. Para Noé, que es la torta y Beck.

Quizás la diferencia más notable entre nosotros es nuestra actitud hacia el béisbol. Estoy obsesionado con el béisbol y, en particular, mi amor de los Mets de Nueva York siempre ha sido una parte fundamental de mi identidad. Noah encuentra béisbol increíblemente aburrido, y su odio por el deporte ha sido durante mucho tiempo una parte fundamental de su identidad.\*



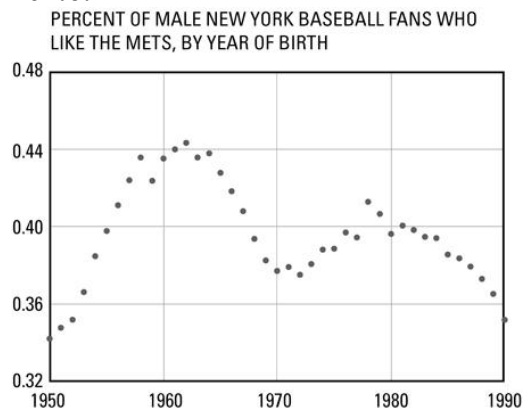
Seth Stephens-Davidowitz Béisbol-o- Phile



¿Cómo pueden dos tipos con tales genes similares, planteadas por los mismos padres, en la misma ciudad, tener tales sentimientos opuestos sobre béisbol? Lo que determina los adultos nos hemos convertido? Más importante aún, ¿qué pasa con Noé? Hay un campo cada vez mayor dentro de la psicología del desarrollo que las bases de datos masivas adultos minas y los correlaciona con los acontecimientos clave de la infancia. Puede ayudar a hacer frente a esta y otras preguntas relacionadas. Podríamos llamar a este aumento del uso de grandes volúmenes de datos para responder a cuestiones psicológicas Gran Psych.

Para ver cómo funciona esto, vamos a considerar un estudio que realicé sobre cómo experimenta la infancia influencia que el equipo de béisbol apoyas-o si usted apoya cualquier equipo en absoluto. Para este estudio, he utilizado los datos de Facebook en “me gusta” de los equipos de béisbol. (En el capítulo anterior he señalado que los datos de Facebook pueden ser profundamente engañosa sobre temas sensibles. Con este estudio, estoy asumiendo que nadie, ni siquiera un ventilador de Phillies, se avergüenza de reconocer un interés enraizamiento en un equipo en particular en Facebook.)

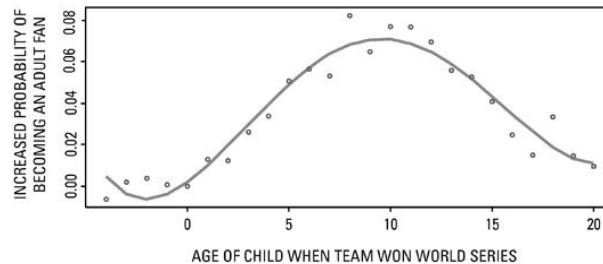
Para empezar, he descargado el número de hombres de todas las edades que “al igual que” cada uno de los dos equipos de béisbol de Nueva York. Aquí son el porcentaje que son fanáticos de los Mets, por año de nacimiento.



Cuanto mayor sea el momento, los más fanáticos de los Mets. La popularidad del equipo se eleva y cae luego sube y baja de nuevo, con los Mets siendo muy popular entre los nacidos en 1962 y 1978. Supongo que los aficionados al béisbol podría tener una idea de lo que está pasando aquí. Los Mets han ganado dos Series Mundiales: en 1969 y 1986. Estos hombres eran más o menos siete u ocho años de edad cuando los Mets ganaron. Así, un gran predictor de Mets fans, para los niños, por lo menos, es si los Mets ganaron una Serie Mundial cuando eran alrededor de la edad de siete u ocho.

De hecho, podemos extender este análisis. He descargado la información en Facebook que muestra cómo muchos aficionados de todas las edades “como” cada uno de una amplia selección de equipos de Grandes Ligas.

He descubierto que hay también un número inusualmente alto de fans masculinos Baltimore Orioles nacidos en 1962 y ventiladores Pittsburgh Pirates varones nacidos en 1963. Esos hombres eran de ocho años de edad, los niños cuando estos equipos fueron campeones. De hecho, el cálculo de la edad del fandom pico para todos los equipos que he estudiado, a continuación, encontrar la manera de edad habrían sido estos ventiladores, me dio esta tabla:



Una vez más vemos que el año más importante en la vida de un hombre, a los efectos de cimentar su equipo de béisbol favorito como un adulto, es cuando es más o menos ocho años de edad. En general, cinco a quince es el período clave para ganar a un niño. Ganadora cuando un hombre es diecinueve o veinte años es aproximadamente un octavo tan importante en la determinación de que él arrancará a como ganar cuando es de ocho. Para entonces, él ya le encantará ya sea un equipo para la vida o no lo hará.

Se puede preguntar, ¿qué pasa con los fans de las mujeres de béisbol? Los patrones son mucho menos agudo, pero el pico de edad parece ser veintidós años de edad.

Este es mi estudio favorito. Se refiere a dos de mis más queridos temas: béisbol y las fuentes de mi descontento adulto. Estaba firmemente enganchado en 1986 y he estado sufriendo a lo largo de enraizamiento para los Mets-desde entonces. Noé tuvo el buen sentido de nacer cuatro años más tarde y se salvó de este dolor.

Ahora, el béisbol no es el tema más importante en el mundo, o por lo que mi Ph.D. asesores me dijeron en repetidas ocasiones. Sin embargo, esta metodología podría ayudar a hacer frente a cuestiones similares, incluyendo cómo la gente desarrolla sus preferencias políticas, tendencias sexuales, gustos musicales, y hábitos financieros. (Me gustaría que particularmente interesado en los orígenes de las ideas locas de mi hermano en los dos últimos temas.) Mi predicción es que vamos a encontrar que muchos de nuestros comportamientos e intereses adultos, incluso los que consideramos fundamental para lo que somos, puede haber explicado por los hechos arbitrarios de cuando nacimos y lo que estaba ocurriendo en ciertos años clave, mientras que éramos jóvenes.

De hecho, algunos trabajos se ha hecho ya en el origen de las preferencias políticas. Yair Ghitza, científico en jefe de Catalist, una empresa de análisis de datos, y Andrew Gelman, politólogo y estadístico de la Universidad de Columbia, trataron de probar la idea convencional

de que la mayoría de la gente empieza a cabo liberal y cada vez más conservadora a medida que envejecen. Esta es la opinión expresada en una famosa cita a menudo se atribuye a Winston Churchill: “Cualquier hombre que es menor de 30 años, y no es un liberal, no tiene corazón; y el hombre que es más de 30, y no es un conservador, no tiene cerebro “.

Ghitza y Gelman estudiaron minuciosamente a través de sesenta años de datos de la encuesta, tomando ventaja de más de 300.000 observaciones sobre las preferencias de voto. Encontraron, contrariamente a la afirmación de Churchill, que los adolescentes a veces se inclinan liberal y conservadora a veces se inclinan. Al igual que los de mediana edad y los ancianos.

Estos investigadores descubrieron que los puntos de vista políticos de hecho se forman de una manera no muy diferente a la forma en que nuestras preferencias de equipos deportivos hacen. Hay un período crucial que imprime en las personas para la vida. Entre las edades clave de catorce y veinticuatro años, numerosos estadounidenses formar sus puntos de vista sobre la base de la popularidad del actual presidente. Un popular demócrata republicano o impopular influirá en muchos adultos jóvenes a ser republicanos. Un demócrata republicano o popular impopular pone este grupo impresionable en la columna Democrática.

Y estos puntos de vista, en estos años clave, tendrán, en promedio, durar toda la vida.

Para ver cómo funciona esto, comparar los estadounidenses nacidos en 1941 y los nacidos una década más tarde.

Los del primer grupo de la mayoría de edad durante la presidencia de Dwight D. Eisenhower, un republicano popular. A principios de 1960, a pesar de ser menor de treinta años, esta generación inclina fuertemente hacia el partido republicano. Y los miembros de esta generación han inclinado constantemente republicano, ya que han envejecido.

Los estadounidenses nacidos diez años más tarde-baby boom mayoría de edad durante las presidencias de John F. Kennedy, un demócrata muy popular; Lyndon B. Johnson, un demócrata inicialmente populares; y Richard M. Nixon, un republicano que finalmente renunció en desgracia. Los miembros de esta generación han inclinado liberal toda su vida.

Con todos estos datos, los investigadores fueron capaces de determinar el solo año más importante para el desarrollo de puntos de vista políticos: dieciocho años de edad.

Y encontraron que estos efectos de impresión son sustanciales. Su modelo estima que la experiencia Eisenhower dio lugar a un fuerte estímulo vida útil de 10 puntos porcentuales para los republicanos entre los estadounidenses nacidos en 1941. El Kennedy, Johnson, Nixon y la experiencia dio a los demócratas una ventaja de 7 puntos porcentuales entre los estadounidenses nacidos en 1952.

Me he dejado en claro que yo soy escéptico de datos de la encuesta, pero estoy impresionado con el gran número de respuestas se examinan aquí. De hecho, este estudio no podría haberse hecho con una pequeña encuesta. Los investigadores necesitan los cientos de miles de observaciones, agregados de muchas encuestas, para ver cómo cambian las preferencias con la edad.

tamaño de los datos también fue crucial para mi estudio de béisbol. Que necesitaba para hacer un zoom no sólo de los aficionados de cada equipo, pero en las personas de todas las edades. Se

requieren millones de observaciones para hacer esto y Facebook y otras fuentes digitales suelen ofrecer este tipo de números.

Aquí es donde la grandeza de los grandes datos realmente entra en juego. Se necesita una gran cantidad de píxeles en una foto con el fin de ser capaz de acercarse con claridad en una pequeña parte de ella. Del mismo modo, se necesita una gran cantidad de observaciones en un conjunto de datos con el fin de ser capaz de acercarse con claridad en un pequeño subconjunto de los datos, por ejemplo, la popularidad de los Mets entre los hombres nacidos en el año 1978. Una pequeña encuesta de un par de miles de personas no tendrá una muestra suficientemente grande de tales hombres.

Esta es la tercera potencia de Big Data: Big Data nos permite acercarnos de manera significativa en los pequeños segmentos de un conjunto de datos para obtener nuevos conocimientos sobre lo que somos. Y podemos hacer zoom en otras dimensiones además de la edad. Si tenemos suficientes datos, podemos ver cómo la gente en los pueblos y ciudades en particular se comporta. Y podemos ver cómo la gente lleva en hora por hora o incluso minuto a minuto.

En este capítulo, el comportamiento humano obtiene su primer plano.

### [Lo que realmente está pasando en nuestros condados, ciudades y pueblos?](#)

---

En retrospectiva, es sorprendente. Pero cuando Raj Chetty, entonces profesor en la Universidad de Harvard, y un pequeño equipo de investigación primero tiene una bodega de un lugar de gran conjunto de datos todos los registros de impuestos de los estadounidenses desde 1996, que no eran de él vendría determinado nada. El IRS había entregado los datos porque pensaban que los investigadores podrían ser capaces de utilizarlo para ayudar a aclarar los efectos de la política fiscal.

La primera intentó Chetty y su equipo utilizarlo este Big Data, de hecho, a numerosos callejones sin salida. Sus investigaciones sobre las consecuencias de las políticas de impuestos estatales y federales alcanzado su mayoría las mismas conclusiones a todos los demás tenía sólo mediante el uso de encuestas. Tal vez las respuestas de Chetty, utilizando los cientos de millones de puntos de datos del IRS, eran un poco más precisos. Sin embargo, conseguir las mismas respuestas que todos los demás, con un poco más de precisión, no es un logro importante de las ciencias sociales. No es el tipo de trabajo que las principales revistas están ansiosas por publicar.

Además, organizar y analizar todos los datos del IRS era mucho tiempo. Chetty y su equipo de ahogarse en datos estaban tomando más tiempo que los demás para encontrar las mismas respuestas que todos los demás.

Estaba empezando a parecerse a los escépticos de grandes datos eran correctos. Usted no necesita datos para cientos de millones de estadounidenses a entender la política fiscal; una encuesta de diez mil personas fue mucho. Chetty y su equipo estaban desanimados comprensible.

Y luego, por último, los investigadores se dieron cuenta de su error. “Big Data no se trata sólo de hacer lo mismo que habría hecho con las encuestas, excepto con más datos,” explica Chetty. Pedían pequeñas preguntas de datos de la recogida masiva de datos que habían sido entregados. “Big Data realmente debería permitir el uso de diseños completamente diferentes de lo que tendría con una encuesta”, añade Chetty. “Se puede, por ejemplo, hacer zoom en las geografías.”

En otras palabras, con datos sobre cientos de millones de personas, Chetty y su equipo podrían detectar patrones entre las ciudades, pueblos y barrios, grandes y pequeñas.

Como estudiante graduado en Harvard, yo estaba en una sala de reuniones cuando Chetty presentó sus resultados iniciales utilizando los registros de impuestos de cada estadounidense. Los científicos sociales se refieren en su trabajo a las observaciones en el número de puntos de datos que tienen. Si un científico social está trabajando con una encuesta de ochocientas personas, decía, “Tenemos ochocientas observaciones.” Si él está trabajando con un experimento de laboratorio con setenta personas, decía, “Tenemos setenta observaciones.”

“Tenemos uno coma dos mil millones de observaciones”, dijo Chetty, con cara seria. El público se rió con nerviosismo.

Chetty y sus colaboradores comenzaron, en esa sala de reuniones y luego en una serie de documentos, que nos dan importantes nuevos conocimientos sobre cómo funciona Latina.

Considere esta pregunta: ¿es América una tierra de oportunidades? ¿Tiene un tiro, si sus padres no son ricos, a enriquecerse a sí mismo?

La forma tradicional de responder a esta pregunta es mirar a una muestra representativa de los estadounidenses y comparar esto con datos similares de otros países.

Estos son los datos de una variedad de países en materia de igualdad de oportunidades. La pregunta: ¿cuál es la probabilidad de que una persona con los padres en el 20 por ciento de la distribución de los ingresos alcanza el 20 por ciento superior de la distribución del ingreso?

#### **PROBABILIDADES A PERSONA CON padres pobres se hará rico (países seleccionados)**

---

Estados Unidos	7.5
Reino Unido	9.0
Dinamarca	11.7
Canadá	13.5

Como se puede ver, América no marca bien.

Pero este simple análisis no alcanza la verdadera historia. El equipo de Chetty zoom en la geografía. Encontraron que las probabilidades difieren una enorme cantidad dependiendo de donde en los Estados Unidos de que nacieras.

#### **PROBABILIDADES A PERSONA CON padres pobres se hará rico (partes seleccionadas)**

---

## de la Estados Unidos)

---

San Jose, CA	12.9
Washington DC	10.5
<i>Estados Unidos Promedio</i>	7.5
Chicago, IL	6.5
Charlotte, NC	4.4

En algunas partes de los Estados Unidos, la posibilidad de que un niño pobre éxito es tan alta como en todos los países desarrollados en el mundo. En otras partes de los Estados Unidos, la posibilidad de que un niño pobre éxito es menor que en cualquier país desarrollado del mundo.

Estos patrones nunca serían vistos en un pequeño estudio, que sólo podría incluir unas pocas personas en Charlotte y San José, y que por tanto, se les impide hacer zoom como este.

De hecho, el equipo de Chetty podría acercar aún más. Porque tenían tanto datos de datos en cada uno de los estadounidenses-podrían incluso hacer zoom sobre los pequeños grupos de personas que se mudaron de ciudad en ciudad para ver cómo eso podría haber afectado sus perspectivas: los que se trasladó desde la ciudad de Nueva York a Los Ángeles , Milwaukee a Atlanta, San José a Charlotte. Esto permitió poner a prueba la causalidad, no sólo de correlación (una distinción Voy a discutir en el próximo capítulo). Y, sí, trasladándose a la ciudad justo en los años formativos de uno hecho una diferencia significativa.

Así es América una “tierra de oportunidades”?

La respuesta no es ni sí ni no. La respuesta es: algunas partes son, y algunas partes no lo son.

Como escriben los autores, “Estados Unidos está mejor descrito como una colección de las sociedades, algunas de las cuales son 'tierras de oportunidad' con altas tasas de movilidad a través de generaciones, y otras en las que algunos niños se escapan de la pobreza.”

Entonces, ¿qué es lo que tiene partes de los Estados Unidos, donde existe una alta movilidad de ingresos? Lo que hace que algunos lugares mejores para igualar el campo de juego, de permitir que un niño pobre que tiene una muy buena vida? Las áreas que gastan más en educación proporcionan una mejor oportunidad a los niños pobres. Los lugares con más personas religiosas y el delito menor va mejor. Lugares con mucha gente negra hacen peor. Curiosamente, esto tiene un efecto sobre no sólo los niños negros, pero en los niños blancos que viven allí también. Los lugares con gran cantidad de madres solteras hacen peor. Este efecto también es válido no sólo para los niños de madres solteras, pero para los niños de padres casados que viven en lugares con gran cantidad de madres solteras. Algunos de estos resultados sugieren que los compañeros de un niño pobre importan. Si sus amigos tienen un fondo difícil y pocas oportunidades, se puede luchar más para escapar de la pobreza.

Los datos nos dicen que algunas partes de América son mejores para dar a los niños la oportunidad de escapar de la pobreza. Entonces, ¿qué lugares son los mejores en dar a las personas la oportunidad de escapar de la Parca?



Nos gusta pensar en la muerte como el gran ecualizador. Nadie, después de todo, puede evitarlo. No es la pobre ni el rey, el hombre sin hogar ni Mark Zuckerberg. Todos mueren.

Pero si los ricos no pueden evitar la muerte, los datos nos dicen que ahora pueden retrasarlo. Las mujeres estadounidenses en el 1 por ciento de los ingresos viven, en promedio, diez años más que las mujeres americanas en la parte baja del 1 por ciento de los ingresos. Para los hombres, la diferencia es de quince años.

¿Cómo varían estos patrones en diferentes partes de los Estados Unidos? ¿Su esperanza de vida varía según el lugar donde vive? ¿Es esta variación diferente para las personas ricas y pobres? Una vez más, haciendo zoom sobre la geografía, el equipo de Raj Chetty encontró las respuestas.

Curiosamente, para los estadounidenses más ricos, la esperanza de vida apenas se ve afectada por el lugar donde viven. Si tiene excesos de dinero, se puede esperar para que sea más o menos ochenta y nueve años como una mujer y ochenta y siete años como un hombre. La gente rica de todo el mundo tiende a desarrollar hábitos saludables, en promedio, más ejercicio, comer mejor, fuman menos, y son menos propensos a sufrir de obesidad. Los ricos pueden permitirse el tapiz rodante, los aguacates orgánicos, las clases de yoga. Y pueden comprar estas cosas en cualquier rincón de los Estados Unidos.

Para los pobres, la historia es diferente. Para los estadounidenses más pobres, la esperanza de vida varía enormemente dependiendo del lugar donde viven. De hecho, la vida en el lugar correcto puede agregar cinco años a la esperanza de vida de una persona pobre.

Entonces, ¿por qué algunos lugares parecen permitir a los empobrecidos a vivir mucho más tiempo? ¿Qué clase de atributos ciudades donde vive la gente pobre la proporción más larga?

Aquí hay cuatro atributos de una ciudad-tres de ellos no se correlacionan con la esperanza de vida de la gente pobre, y uno de ellos lo hace. Ver si se puede adivinar cuál asuntos.

### **¿QUÉ hace que la gente pobre en una ciudad viven mucho más tiempo?**

---

La ciudad tiene un alto nivel de religiosidad.

La ciudad cuenta con bajos niveles de contaminación.

La ciudad tiene un mayor porcentaje de residentes cubiertos por el seguro médico.

Una gran cantidad de personas ricas viven en la ciudad.

La primera de tres religión, el medio ambiente, la salud y el seguro-no se correlacionan con vidas más largas para los pobres. La variable que sí importa, según Chetty y los otros que trabajaron en este estudio? ¿Cuánta gente rica vive en una ciudad. Las personas más ricas en una ciudad significa los pobres no viven más tiempo. La gente pobre de la ciudad de Nueva York, por ejemplo, viven mucho más tiempo que la gente pobre de Detroit.

¿Por qué es la presencia de los ricos un potente predictor de la esperanza de vida de los pobres tales? Una hipótesis -y esto es especulativo fue propuesta por David Cutler, uno de los autores del estudio y uno de mis asesores. comportamiento contagiosa puede ser la conducción algo de esto.

Hay una gran cantidad de investigaciones que muestran que los hábitos son contagiosos. Así que las personas pobres que viven cerca de los ricos pueden recoger una gran cantidad de sus hábitos. Algunos de estos hábitos -por ejemplo, pretenciosa vocabulario HAYA probable que afectan a la salud. Otras-trabajan fuera sin duda tendrá un impacto positivo. De hecho, las personas pobres que viven cerca de la gente rica hacer más ejercicio, fuman menos, y son menos propensos a sufrir de obesidad.

Mi estudio favorito personal por el equipo de Raj Chetty, que tenía acceso a esa gran colección de datos del IRS, fue su investigación sobre por qué algunas personas hacen trampa en sus impuestos, mientras que otros no lo hacen. Al explicar este estudio es un poco más complicado.

La clave es saber que hay una manera fácil para los trabajadores autónomos con un niño para maximizar el dinero que reciben por parte del gobierno. Si informar que recibió ingreso gravable de exactamente \$ 9.000 en un año dado, el gobierno va a escribir un cheque por \$ 1.377, esa cantidad representa el Crédito por Ingreso, una subvención para complementar los ingresos de los trabajadores pobres, menos sus impuestos sobre la nómina. Informar nada más que eso, y sus impuestos de nómina va a subir. Reportar menos que eso, y deja caer el Crédito por Ingreso. Una utilidad gravable de \$ 9.000 es el punto dulce.

Y, ¿no lo sabes, \$ 9.000 es el ingreso gravable más comunes reportados por los trabajadores autónomos con un niño.

¿Estos americanos ajustar sus horarios de trabajo para asegurarse de que los ingresos obtenidos perfecto? Nop. Cuando estos trabajadores fueron auditados-a-azar ocurrencia muy rara se encontró que casi siempre hicieron ninguna parte cerca de \$ 9,000 que ganaban sea sustancialmente menor o sustancialmente más.

En otras palabras, hicieron trampa en sus impuestos al pretender que hicieron la cantidad que les daría el cheque más gordo del gobierno.

Entonces, ¿cómo era típico de este tipo de fraude fiscal y que entre los autónomos con un hijo con la mayor probabilidad de cometer ella? Resulta, Chetty y sus colegas informaron que había grandes diferencias entre los Estados Unidos en lo común que este tipo de engaño era. En Miami, entre la gente en esta categoría, un sorprendente 30 por ciento informó que hicieron \$ 9.000. En Filadelfia, a sólo 2 por ciento lo hizo.

Lo que predice que va a engañar? ¿Qué hay en los lugares que tienen el mayor número de los tramposos y los que tienen números más bajos? Podemos correlacionar las tasas de hacer trampa con otros datos demográficos a nivel de ciudad y resulta que hay dos predictores fuertes: una alta concentración de personas en la zona que califican para el Crédito por Ingreso y una alta concentración de profesionales de impuestos en el barrio.

¿Qué indican estos factores? Chetty y los autores no tenían una explicación. El motivador clave para hacer trampa en sus impuestos de este modo era de la información.

La mayoría de los contribuyentes de un niño autónomos simplemente no sabían que el número mágico para recibir un cheque grande y gordo del gobierno fue de \$ 9.000. Pero vivir cerca de

otras personas que puedan, ya sea de sus vecinos o de impuestos asisters-drásticamente el aumento de las probabilidades que iban a aprender acerca de él.

De hecho, el equipo de Chetty encontró aún más evidencia de que el conocimiento condujo este tipo de trampas. Cuando los estadounidenses se movieron de un área donde esta variedad de fraude fiscal fue de baja a un área donde fue alta, aprendieron y adoptaron el truco. A través del tiempo, el engaño propagación de una región a través de los Estados Unidos. Como un virus, la evasión fiscal es contagiosa.

Ahora parar por un momento y pensar en cómo revela este estudio es. Se demostró que, cuando se trata de averiguar quién va a hacer trampa en sus impuestos, la clave no es la determinación de que es honesto y que es deshonesto. Se está determinando que sabe cómo hacer trampa y que no lo hace.

Así que cuando alguien le dice que nunca se harían trampa en sus impuestos, hay una buena posibilidad de que ellos son-lo adivinaste-mentira. La investigación de Chetty sugiere que muchos si sabía cómo.

Si quieres hacer trampa en sus impuestos (y no estoy recomendando esto), usted debe vivir cerca de los profesionales de impuestos o vivir cerca de los tramposos de impuestos que le puede mostrar el camino. Si usted quiere tener hijos que son de fama mundial, por dónde debe vivir? Esta capacidad para hacer un zoom en los datos y obtener realmente granular puede ayudar a responder a esta pregunta, también.

Tenía curiosidad donde los estadounidenses más exitosas vienen, así que un día decidí descargar Wikipedia. (Se puede hacer ese tipo de cosas hoy en día.)

Con un poco de codificación, que tenía un conjunto de datos de más de 150.000 estadounidenses considerados por los editores de Wikipedia que ser lo suficientemente notable como para justificar una entrada. El conjunto de datos incluye el condado de nacimiento, fecha de nacimiento, ocupación y género. Me fundí con los datos de nacimientos a nivel de condado recogidos por el Centro Nacional de Estadísticas de Salud. Por cada condado en los Estados Unidos, he calculado las probabilidades de hacer en Wikipedia si nació allí.

Está siendo perfilado en Wikipedia un marcador significativo del logro notable? Es cierto que existen limitaciones. editores de Wikipedia sesgan hombres jóvenes, lo que puede sesgar la muestra. Y algunos tipos de notoriedad no son particularmente dignos. Ted Bundy, por ejemplo, tasas de una entrada de Wikipedia porque mató a decenas de mujeres jóvenes. Dicho esto, yo era capaz de eliminar los criminales sin afectar a los resultados mucho.

Limité el estudio de los baby boomers (los nacidos entre 1946 y 1964), ya que han tenido casi toda una vida para convertirse en notable. Aproximadamente uno de cada 2.058 generación del baby boom nacidos en Estados Unidos fueron consideradas lo suficientemente notable como para justificar una entrada de Wikipedia. Alrededor del 30 por ciento lo hizo a través logros en el arte o entretenimiento, el 29 por ciento a través del deporte, un 9 por ciento a través de la política, y 3 por ciento en el mundo académico o de la ciencia.

El primer hecho sorprendente que noté en los datos fue la enorme variación geográfica en la probabilidad de convertirse en un gran éxito, al menos en términos de Wikipedia. Sus posibilidades de lograr notoriedad dependían de dónde nació altamente.

Aproximadamente uno de cada 1.209 nacidos generación del baby boom en California llegaron a Wikipedia. Sólo uno de cada 4.496 nacidos generación del baby boom en Virginia Occidental hicieron. Acercar por condado y los resultados se vuelven más revelador. Aproximadamente uno de cada 748 baby boom nacidos en el condado de Suffolk, Massachusetts, donde se encuentra Boston, llegaron a Wikipedia. En algunos otros condados, la tasa de éxito era veinte veces menor.

¿Por qué algunas partes del país parecen ser mucho mejor en produciendo que mueven los hilos de Estados Unidos? Examiné de cerca los condados. Resulta que casi todos ellos encajan en una de dos categorías.

En primer lugar, y esto me sorprendió, muchos de estos condados contenían una ciudad universitaria de tamaño considerable. Casi cada vez que vi el nombre de un condado que no había oído hablar de la parte superior de la lista, al igual que Washtenaw, Michigan, descubrí que estaba dominado por una ciudad universitaria clásica, en este caso, Ann Arbor. Los condados adornado por Madison, Wisconsin; Athens, Georgia; Columbia, Missouri; Berkeley, California; Chapel Hill, Carolina del Norte; Gainesville, Florida; Lexington, Kentucky; e Ithaca, Nueva York, están todos en el top 3 por ciento.

¿Por qué es esto? Parte de ella es bien puede ser debido a la reserva genética: hijos e hijas de los profesores y estudiantes graduados tienden a ser inteligentes (un rasgo que, en el juego de gran éxito, puede ser sumamente útil). Y, de hecho, tener más graduados universitarios en un área es un fuerte predictor del éxito de las personas nacidas allí.

Pero hay más probable es algo más en juego: la exposición temprana a la innovación. Uno de los campos en ciudades de la universidad tienen más éxito en la producción de los mejores perros es la música. Un niño en una ciudad universitaria estará expuesto a conciertos únicos, estaciones de radio inusuales, y tiendas de discos, incluso independientes. Y esto no se limita a las artes. ciudades de la universidad también se incuban más de su parte esperada de notables empresarios. Tal vez la exposición temprana al arte y las ideas de vanguardia les ayuda, también.

El éxito de las ciudades de la universidad no sólo el cruce de las regiones. Atraviesa carrera. Afroamericanos estuvieron poco representados notablemente en la Wikipedia en campos no atletas, especialmente los negocios y la ciencia. Esto, sin duda, tiene mucho que ver con la discriminación. Sin embargo, un pequeño condado, donde la población de 1950 fue del 84 por ciento negro, producido notables generación del baby boom a una velocidad cerca de los de los más altos condados.

De menos de 13.000 boomers nacidos en el condado de Macon, Alabama, quince llegaron a la Wikipedia, o uno de cada 852. Cada uno de ellos es negro. Catorce de ellos eran de la ciudad de Tuskegee, el hogar de la Universidad de Tuskegee, una universidad históricamente negro fundada por Booker T. Washington. La lista incluye a jueces, escritores y científicos. De hecho, un niño negro nacido en Tuskegee tenía la misma probabilidad de ser notable en un campo fuera

de los deportes como un niño blanco nacido en algunos de los más anotadores, ciudades de la universidad de mayoría blanca.

El segundo atributo más probable que los nativos de un condado exitosa fue la presencia en ese condado de una gran ciudad. Haber nacido en el Condado de San Francisco, el condado de Los Ángeles o Nueva York todas puestas entre las probabilidades más altas de lo que es la Wikipedia. (Estoy en un grupo de cinco condados de la Ciudad de Nueva York juntos porque muchas entradas de la Wikipedia no especificaron una ciudad de nacimiento.)

Las áreas urbanas tienden a estar bien provisto de modelos de éxito. Para ver el valor de estar cerca de los profesionales de éxito de una embarcación cuando son jóvenes, comparar la ciudad de Nueva York, Boston y Los Ángeles. Entre los tres, la ciudad de Nueva York produce periodistas notables en la tasa más alta; Boston produce científicos notables en la tasa más alta; y Los Ángeles produce actores notables en la tasa más alta. Recuerde, estamos hablando de personas que nacieron allí, no personas que se mudaron allí. Y esto es así incluso después de restar las personas con padres notables en ese campo.

condados suburbanos, a menos que contenían principales ciudades de la universidad, realizan mucho peor que sus contrapartes urbanas. Mis padres, al igual que muchos boomers, se alejaron de las aceras atestadas calles que en sombra de los árboles este caso desde Manhattan hasta el condado de Bergen, Nueva Jersey, para elevar sus tres hijos. Esta era potencialmente un error, al menos desde la perspectiva de tener hijos notables. Un niño nacido en la ciudad de Nueva York es 80 por ciento más propensos a hacerlo en Wikipedia que un niño nacido en el condado de Bergen. Estos son sólo correlaciones, pero sí sugieren que crecen cerca de las grandes ideas es mejor que crecer con un gran patio.

La cruda efectos identificados aquí podrían ser aún más fuerte si tuviera mejores datos sobre los lugares vivieron durante toda la infancia, ya que muchas personas crecen en diferentes condados de aquel en el que nacieron.

El éxito de las ciudades de la universidad y las grandes ciudades es sorprendente cuando se acaba de ver los datos. Pero también profundicé más profundamente para llevar a cabo un análisis empírico más sofisticado.

Si lo hace, mostró que había otra variable que es un fuerte predictor de una persona de asegurar una entrada en Wikipedia: la proporción de inmigrantes en su condado de nacimiento. Cuanto mayor es el porcentaje de residentes nacidos en el extranjero en un área, mayor es la proporción de niños nacidos allí que van a notable éxito. (Toma eso, Donald Trump!) Si dos lugares tienen las poblaciones urbanas y universitarios similares, el que tiene más inmigrantes producirá estadounidenses más prominentes. ¿Cómo se explica esto?

Mucho de esto parece ser atribuible directamente a los hijos de los inmigrantes. Hice una búsqueda exhaustiva de las biografías de los cien más famosos del baby boom blancos, de acuerdo con el Massachusetts Institute of Technology proyecto Panteón, que también está trabajando con los datos de Wikipedia. La mayoría de ellos eran artistas. Al menos trece tenían madres nacidas en el extranjero, entre ellos Oliver Stone, Sandra Bullock, y Julianne Moore. Esta tasa es más de tres veces superior a la media nacional durante este período. (Muchos tenían

padres que eran inmigrantes, incluyendo a Steve Jobs y John Belushi, pero estos datos era más difícil de comparar con los promedios nacionales, ya que la información sobre los padres no siempre se incluye en los certificados de nacimiento.)

¿Qué pasa con las variables que no afectan el éxito? Uno que he encontrado más que un poco sorprendente fue la cantidad de dinero un estado gasta en educación. En estados con porcentajes similares de sus residentes que viven en zonas urbanas, el gasto en educación no se correlacionó con los índices de producción de notables escritores, artistas o líderes de negocios.

Es interesante comparar mi estudio Wikipedia a uno de los estudios del equipo de Chetty discutidos anteriormente. Recordemos que el equipo de Chetty estaba tratando de averiguar qué áreas son buenos para permitir que las personas lleguen a la clase media alta. Mi estudio estaba tratando de averiguar qué áreas son buenos para permitir que las personas lleguen a la fama. Los resultados son notablemente diferentes.

Gastar mucho en la educación ayuda a los niños lleguen a la clase media alta. Hace poco para ayudarles a convertirse en un escritor, artista o líder empresarial notable. Muchos de estos enormes éxitos odiaba la escuela. Algunos abandonaron.

Ciudad de Nueva York, el equipo de Chetty encontrado, no es un buen lugar para criar a un niño si usted quiere asegurarse de que llegue a la clase media alta. Es un gran lugar, mi estudio encontró, si quieres darle una oportunidad a la fama.

Cuando nos fijamos en los factores que impulsan el éxito, la gran variación entre los condados comienza a tener sentido. Muchos condados combinan todos los ingredientes principales para el éxito. Volver, de nuevo, a Boston. Con numerosas universidades, se cuece en ideas innovadoras. Es un área urbana con muchas personas muy exitosos ofreciendo Alevín ejemplos de cómo hacerlo. Y dibuja un montón de inmigrantes, cuyos hijos se ven obligados a aplicar estas lecciones.

¿Qué pasa si un área tiene ninguna de estas cualidades? ¿Está destinado a producir menos superestrellas? No necesariamente. Hay otro camino: la especialización extrema. Condado de Roseau, Minnesota, un pequeño condado rural con pocos extranjeros y no hay grandes universidades, es un buen ejemplo. Aproximadamente 1 de cada 740 personas nacidas aquí lo hizo en Wikipedia. Su secreto? Los nueve eran jugadores de hockey profesional, sin duda ayudado por la juventud de clase mundial del condado y programas de hockey de la escuela secundaria.

Así es el punto aquí, suponiendo que no está tan interesado en la crianza de un hockey de estrella para pasar a Boston o Tuskegee si usted quiere dar a sus hijos en el futuro la mayor ventaja? No puede hacer daño. Pero hay lecciones más grandes aquí. Por lo general, los economistas y sociólogos se centran en cómo evitar los malos resultados, como la pobreza y la delincuencia. Sin embargo, el objetivo de una gran sociedad no es sólo para dejar un menor número de personas que están detrás; que es ayudar a tantas personas como sea posible para destacarse realmente. Tal vez este esfuerzo para hacer un zoom sobre los lugares donde nacieron cientos de miles de los más famosos estadounidenses nos puede dar algunas estrategias iniciales: fomentan la inmigración, que subvencionan las universidades, y el apoyo a las artes, entre ellas.

Por lo general, estudio de los Estados Unidos. Así que cuando pienso en hacer zoom por la geografía, pienso en hacer zoom sobre nuestras ciudades y pueblos-de ver lugares como el condado de Macon, Alabama, y el condado de Roseau, Minnesota. Pero otra enorme y sigue creciendo-ventaja de datos de Internet es que es fácil de recoger los datos de todo el mundo. A continuación, podemos ver cómo difieren los países. Y científicos de datos tienen la oportunidad de puntillas en la antropología.

Uno de los temas un tanto aleatoria Hace poco explorado: ¿cómo jugar el embarazo a cabo en diferentes países de todo el mundo? Examiné las búsquedas de Google por las mujeres embarazadas. Lo primero que encontré fue una sorprendente similitud en los síntomas físicos de la que se quejan las mujeres.

Probé la frecuencia con síntomas diferentes se buscaron en combinación con la palabra “embarazada”. Por ejemplo, con qué frecuencia es “embarazada” buscado en conjunción con “náuseas”, “dolor de espalda,” o “estreñimiento”? síntomas de Canadá eran muy cercanos a los de los Estados Unidos. Los síntomas en países como Gran Bretaña, Australia y la India eran todos más o menos similar, también.

Las mujeres embarazadas de todo el mundo al parecer también anhelan las mismas cosas. En los Estados Unidos, la parte superior de búsqueda de Google en esta categoría es “ansia de hielo durante el embarazo.” El próximo cuatro son sal, dulces, fruta y comida picante. En Australia, los antojos no difieren casi nada: la lista cuenta con sal, dulces, chocolate, hielo y fruta. ¿Qué sobre la India? Una historia similar: la comida picante, dulces, chocolate, sal, y el helado. De hecho, los cinco primeros son muy similares en todos los países que miraba.

La evidencia preliminar sugiere que ninguna parte del mundo se ha topado con una dieta o el medio ambiente que cambia drásticamente la experiencia física del embarazo.

Pero los pensamientos que rodean el embarazo definitivamente difieren.

Comience con preguntas sobre lo que las mujeres embarazadas pueden hacer de manera segura. Las preguntas más habituales en los Estados Unidos: las mujeres embarazadas pueden “comer camarones”, “beber vino”, “tomar un café”, o “tomar Tylenol”?

Cuando se trata de este tipo de preocupaciones, otros países no tienen mucho en común con los Estados Unidos o entre sí. Si las mujeres embarazadas pueden “tomar vino” no está entre las diez preguntas en Canadá, Australia o Gran Bretaña. preocupaciones de Australia están principalmente relacionados con el consumo de productos lácteos durante el embarazo, especialmente el queso crema. En Nigeria, donde el 30 por ciento de la población utiliza Internet, la cuestión es si la parte superior las mujeres embarazadas pueden beber agua fría.

Son estas preocupaciones legítimas? Depende. Hay una fuerte evidencia de que las mujeres embarazadas tienen un mayor riesgo de listeria de queso sin pasteurizar. Se han establecido vínculos entre el consumo excesivo de alcohol y los resultados negativos para el niño. En algunas partes del mundo, se cree que beber agua fría puede dar a su bebé la neumonía; No conozco ningún apoyo médico para esto.

Las enormes diferencias en las preguntas planteadas en todo el mundo son causados probablemente por el abrumador flujo de información procedente de fuentes dispares en cada

país: los estudios científicos legítimos, estudios científicos tan-tan, cuentos de viejas, y la charla barrio. Es difícil que las mujeres saben en qué concentrarse, o lo que a Google.

Podemos ver otra diferencia clara cuando nos fijamos en las principales búsquedas para “cómo \_\_\_\_ durante el embarazo?” En los Estados Unidos, Australia y Canadá, la búsqueda de arriba es “cómo prevenir las estrías durante el embarazo.” Pero en Ghana, India y Nigeria, la prevención de las estrías no es ni siquiera entre los cinco primeros. Estos países tienden a estar más preocupados por cómo tener relaciones sexuales o la forma de dormir.

**TOP FIVE SEARCHES (IN ORDER) FOR “HOW TO \_\_\_\_ DURING PREGNANCY”**

UNITED STATES	INDIA	AUSTRALIA	BRITAIN	NIGERIA	SOUTH AFRICA
prevent stretch marks	sleep	prevent stretch marks	lose weight	have sex	have sex
lose weight	do sex	lose weight	prevent stretch marks	lose weight	lose weight
have sex	have sex	avoid stretch marks	avoid stretch marks	make love	prevent stretch marks
avoid stretch marks	sex	sleep	sleep	stay healthy	sleep
stay fit	take care	have sex	have sex	stop vomiting	stop vomiting

**TOP FIVE SEARCHES BEGINNING WITH “CAN PREGNANT WOMEN \_\_\_\_?”**

UNITED STATES	eat shrimp	drink wine	drink coffee	take Tylenol	eat sushi
BRITAIN	eat prawns	eat smoked salmon	eat cheesecake	eat mozzarella	eat mayonnaise
AUSTRALIA	eat cream cheese	eat prawns	eat bacon	eat sour cream	eat feta cheese

NIGERIA	drink cold water	drink wine	drink coffee	have sex	take moringa (edible plant)
SINGAPORE	drink green tea	eat ice cream	eat durian	drink coffee	eat pineapple
SPAIN	eat pâté	eat jamón	take paracetamol (pain relief)	eat tuna	sunbathe
GERMANY	fly	eat salami	go in sauna	eat honey	eat mozzarella
BRAZIL	dye their hair	take Dipirona (pain relief)	take paracetamol	ride a bike	fly



Hay, sin duda, más que aprender de hacer zoom sobre los aspectos de la salud y la cultura en diferentes rincones del mundo. Pero mi análisis preliminar sugiere que grandes volúmenes de datos nos dirá que el ser humano es mucho menos potente que nos dimos cuenta cuando se trata de trascender nuestra biología. Sin embargo, nos encontramos con muy diferentes interpretaciones de lo que significa todo esto.

### CÓMO llenamos nuestros minutos y horas

---

“Las aventuras de un joven cuyos intereses principales son la violación, la ultra-violencia y Beethoven”.

Así fue como se anunciaba controvertido naranja mecánica de Stanley Kubrick. En la película, el joven protagonista de ficción, Alex DeLarge, cometió actos de violencia impactantes con desprendimiento de refrigeración. En una de las escenas más notorios de la película, que violó a una mujer mientras está cantando a todo pulmón “Cantando bajo la lluvia”.

Casi inmediatamente, hubo informes de incidentes de imitación. De hecho, un grupo de hombres violó a una chica de diecisiete años de edad, mientras que canta la misma canción. La película fue cerrado en muchos países europeos, y algunas de las escenas más impactantes fueron retirados para una versión mostrada en América.

Hay, de hecho, muchos ejemplos de la vida real imita al arte, con hombres aparentemente hipnotizados por lo que acababan de ver en la pantalla. Una muestra de los colores de la película de pandillas fue seguido por un tiroteo violento. Una proyección de la película banda de New Jack City fue seguido por disturbios.

Tal vez lo más inquietante, cuatro días después de la publicación de El tren del dinero, los hombres utilizan un líquido más ligero para encender una cabina de peaje de metro, que imitan casi perfectamente una escena en la película. La única diferencia entre el incendio de ficción y en el mundo real: En la película, el operador escapó. En la vida real, quemó hasta la muerte.

También hay alguna evidencia de los experimentos psicológicos que los sujetos expuestos a una película violenta reportarán más ira y la hostilidad, incluso si no imitan precisamente una de las escenas.

En otras palabras, anécdotas y experimentos sugieren películas violentas pueden incitar a comportamientos violentos. Pero lo grande que un efecto es lo que realmente tienen? ¿Estamos hablando de uno o dos asesinatos cada década o cientos de muertes cada año? Anécdotas y experimentos no pueden responder a esto.

Para ver si podía grandes volúmenes de datos, dos economistas, Gordon Dahl y Stefano DellaVigna, se fusionaron tres grandes conjuntos de datos para los años 1995 a 2004: Los datos del FBI crimen por hora, números de taquilla, y una medida de la violencia en todas las películas de los niños- in-mind.com.

La información que utilizaban era completa, todas las películas de los crímenes cometidos en cada hora en ciudades de Estados Unidos. Podría llegar a ser importante.

La clave de su estudio fue el hecho de que algunos fines de semana, la película más popular fue una violenta de una sola Anibal o amanecer de los muertos, por ejemplo, mientras que en otros fines de semana, la película más popular fue no violenta, como Novia a la fuga o Toy Story .

Los economistas podrían ver exactamente cuántos asesinatos, violaciones y agresiones se cometieron en los fines de semana cuando una película violenta prominente fue puesto en libertad y que para comparar el número de asesinatos, violaciones, asaltos y hubo fines de semana cuando un prominente película pacífica fue puesto en libertad.

Entonces, ¿qué fue lo que encontraron? Cuando vio la película violenta, hizo aumento del crimen, ya que algunos experimentos sugieren? ¿O siendo el mismo?

Los fines de semana con una película violenta populares, los economistas se encuentran, la delincuencia se redujo.

Has leído bien. Los fines de semana con una película violenta populares, cuando millones de estadounidenses fueron expuestos a imágenes de hombres matar a otros hombres, la delincuencia se redujo significativamente-.

Cuando se obtiene un resultado esta extraña e inesperada, su primer pensamiento es que usted ha hecho algo mal. Cada autor fue cuidadosamente sobre la codificación. Sin errores. Su segundo pensamiento es que hay alguna otra variable que va a explicar estos resultados. Comprobaron si el tiempo del año afectó los resultados. No lo hizo. Se recogieron datos sobre el tiempo, pensando que tal vez de alguna manera esto conducía a la relación. No lo era.

“Verificamos todas nuestras suposiciones, todo lo que estábamos haciendo”, Dahl me dijo. “No hemos podido encontrar nada malo.”

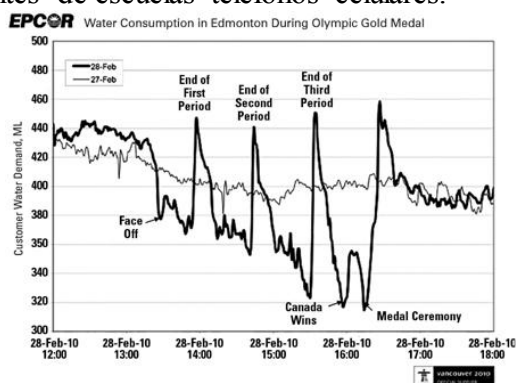
A pesar de las anécdotas, a pesar de las pruebas de laboratorio, y tan extraño que pareciera, mostrando una película violenta de alguna manera provocó una gran caída en el crimen. ¿Cómo podría ser esto posible?

La clave para calcular hacia fuera para Dahl y DellaVigna estaba utilizando su Big Data para hacer un zoom más cerca. Los datos de encuestas proporcionaron información que tradicionalmente era anual o, en el mejor tal vez mensual. Si estamos realmente suerte, podríamos obtener datos para un fin de semana. En comparación, ya que cada vez hemos estado utilizando conjuntos de datos completos, en lugar de las encuestas con muestras pequeñas, hemos sido capaces de casa en la hora y hasta el minuto. Esto nos ha permitido aprender mucho más sobre el comportamiento humano.

A veces, las fluctuaciones en el tiempo son divertida, si no temblar la tierra. EPCOR, una empresa de servicios públicos en Edmonton, Canadá, informó los datos de consumo de agua minuto a minuto durante el partido por la medalla de oro olímpica de hockey 2010 entre los Estados Unidos y Canadá, que se estima que el 80 por ciento de los canadienses observaba. Los datos nos dicen que poco después de cada período terminó, el consumo de agua se disparó. Inodoros a través de Edmonton fueron enrojecimiento claridad.

búsquedas de Google también pueden ser descompuestos por el minuto, que revela algunos patrones interesantes en el proceso. Por ejemplo, las búsquedas de “juegos” no bloqueados se

disparan a las 8 de la mañana del lunes a viernes y se mantienen altos a 3 de la tarde, sin duda, en respuesta a los intentos de bloquear el acceso a juegos móviles en la propiedad de la escuela sin la prohibición de los estudiantes de escuelas teléfonos celulares.



las tasas de búsqueda para “tiempo”, “oración” y “noticias” pico antes de las 5:30 AM, la evidencia de que la mayoría de la gente se despierta mucho antes de lo que hago. las tasas de pico de la búsqueda para “suicidio” a las 12:36 am y se encuentran en los niveles más bajos a las 9 am, la evidencia de que la mayoría de las personas son mucho menos miserable en la mañana de lo que soy.

Los datos muestran que las horas entre 2 y 4 h Las prime time para las grandes preguntas: ¿Cuál es el significado de la conciencia? No existirá libre? ¿Hay vida en otros planetas? La popularidad de estas preguntas nocturnas puede ser el resultado, en parte, del consumo de cannabis. las tasas de la búsqueda de “cómo rodar una articulación” pico entre 1 y las 2 am

Y en su gran conjunto de datos, Dahl y DellaVigna podían ver cómo el crimen cambió rápidamente por hora en esas películas los fines de semana. Encontraron que el descenso de la criminalidad cuando se muestran, en relación con otras películas violentas populares los fines de semana, se inició en la tarde-noche. El crimen era menor, es decir, antes de que las escenas violentas incluso empezó, cuando los aficionados al teatro pueden acaban de entrar.

Se puede adivinar por qué? Pensar, en primer lugar, acerca de quién es probable que optar por asistir a una película violenta. Es hombres jóvenes -en particular jóvenes, hombres agresivos.

Piense, al lado, acerca de dónde delitos tienden a estar comprometido. Pocas veces en una sala de cine. Ha habido excepciones, más notablemente un disparo premeditado 2012 en un teatro de Colorado. Pero, en general, los hombres van a teatros desarmados y se sientan en silencio.

Ofrecer a los jóvenes, hombres agresivos la oportunidad de ver a Aníbal, y van a ir al cine. Ofrecer a los jóvenes, hombres agresivos Novia a la fuga como su opción, y que se hará un pase y en lugar de salir, tal vez a un bar, un club o un salón de billar, donde la incidencia de crímenes violentos es mayor.

películas violentas mantienen a las personas potencialmente violentos de las calles.

Puzzle resuelto. ¿Derecha? No del todo todavía. Había una extraña cosa más en los datos. Los efectos comenzaron justo cuando las películas empezaron a aparecer; Sin embargo, no se

detuvieron después de que la película terminó y el teatro cerrado. En las noches donde las películas violentas estaban mostrando, el crimen fue menor bien entrada la noche, desde la medianoche y las 6 AM

Incluso si el crimen fue menor, mientras que los hombres jóvenes estaban en la sala de cine, ¿no debería levantarse después de salir y ya no estaban preocupados? Ellos sólo habían visto una película violenta, lo que hace que la gente experimentos dicen más enojado y agresivo.

¿Puede pensar en ninguna explicación de por qué el crimen aún se redujo después de la película terminó? Después de mucho pensar, los autores, que eran expertos en delitos, tuvo otro momento “Aha”. Sabían que el alcohol es un importante contribuyente a la delincuencia. Los autores se habían sentado en las salas de cine suficientes para saber que prácticamente no hay salas de cine en los Estados Unidos sirven licor. De hecho, los autores encontraron que los delitos relacionados con el alcohol se desplomaron en horas nocturnas después de las películas violentas.

Por supuesto, Dahl y resultados de Della Vigna eran limitadas. No podían, por ejemplo, probar el mes de salida, con una duración en efectos para ver el tiempo que el descenso de la criminalidad podría durar. Y sigue siendo posible que la exposición constante a las películas violentas en última instancia conduce a más violencia. Sin embargo, su estudio no puso el impacto inmediato de las películas violentas, que ha sido el tema principal en estos experimentos, en perspectiva. Tal vez una película violenta hace influir en algunas personas y hacerlos inusualmente enojado y agresivo. Sin embargo, ¿sabes lo que sin duda influye en la gente en una dirección violento? Salir con otros hombres y beber potencialmente violentas.\*

Esto tiene sentido ahora. Pero no tenía sentido antes Dahl y Della Vigna comenzaron el análisis de montones de datos.

Un punto más importante que se hace evidente cuando nos acercamos: el mundo es complicado. Medidas que tomemos hoy pueden tener efectos distantes, la mayoría de ellos no deseado. Las ideas se propagan a veces lentamente; otras veces de manera exponencial, como los virus. Las personas responden de manera impredecible a los incentivos.

Estas conexiones y relaciones, estas sobretensiones y se hincha, no se pueden rastrear con pequeñas encuestas o métodos tradicionales de datos. El mundo, sencillamente, es demasiado complejo y demasiado rico para pocos datos.

## [NUESTROS doppelgangers](#)

---

En junio de 2009, David “Big Papi” Ortiz parecía que estaba hecho. Durante la década anterior media, Boston había caído en el amor con su toletero dominicano con la sonrisa amable y dientes con huecos.

Había hecho cinco juegos consecutivos de las Estrellas, ganó un premio de MVP, y ayudó a poner fin a la sequía de campeonato de ochenta y seis años de Boston. Sin embargo, en la temporada de 2008, a la edad de treinta y dos años, sus números cayeron. Su promedio de bateo

había caído 68 puntos, su porcentaje en base de 76 puntos, su porcentaje de slugging de 114 puntos. Y al inicio de la temporada 2009, los números de Ortiz fueron cayendo más.

Así es como Bill Simmons, un periodista deportivo y apasionado fan de Boston Red Sox, describió lo que ocurría en los primeros meses de la temporada 2009: “Está claro que David Ortiz ya no sobresale en el béisbol. . . . Beefy bateadores son como las estrellas porno, luchadores, los centros de la NBA y esposas trofeo: Cuando se va, se va “ventiladores grandes deportes confían en sus ojos, y los ojos de Simmons le dijo Ortiz había terminado.. De hecho, Simmons predijo que iba a ser enviado a la banca o en libertad poco.

Se Ortiz realmente terminado? Si usted es el gerente general de Boston, en 2009, ¿se lo cortó? De manera más general, ¿cómo podemos predecir cómo un jugador de béisbol llevará a cabo en el futuro? Incluso de manera más general, ¿cómo podemos utilizar grandes volúmenes de datos para predecir lo que la gente va a hacer en el futuro?

Una teoría que lo llevará lejos en la ciencia de datos es la siguiente: mirar lo que sabermetricians (aquellos que han utilizado los datos para el estudio de béisbol) han hecho y espera que se extienda a otras áreas de la ciencia de datos. El béisbol fue uno de los primeros campos con conjuntos de datos completos sobre casi todo, y un ejército de gente inteligente que estén dispuestos a dedicar su vida a dar sentido a esos datos. Ahora, casi todos los campos está allí o llegar allí. Béisbol es lo primero; cualquier otro campo siguiente. Sabermetrics come el mundo.

La forma más sencilla de predecir el futuro de un jugador de béisbol es suponer que seguirá actuando como él es actualmente. Si un jugador ha luchado durante los últimos 1,5 años, se puede adivinar que va a luchar por los próximos 1,5 años.

Por esta metodología, Boston debería haber cortado David Ortiz.

Sin embargo, puede haber información más relevante. En la década de 1980, Bill James, que la mayoría considera el fundador de sabermetrics, hizo hincapié en la importancia de la edad. Los jugadores de béisbol, James se encontró, alcanzó su punto máximo a principios-en torno a la edad de veintisiete años. Equipos tendían a ignorar hasta qué punto los jugadores disminuyen a medida que envejecen. Que pagó en exceso por el envejecimiento de los jugadores.

Por esta metodología más avanzada, Boston sin duda debería haber cortado David Ortiz.

Pero este ajuste edad podría perder algo. No todos los jugadores siguen el mismo camino por la vida. Algunos jugadores podrían alcanzar su punto máximo a los veintitrés años, otros a los treinta y dos. jugadores cortos pueden envejecer diferente a jugadores altos, reproductores de grasa de jugadores flacos. los estadísticos de béisbol encontró que había tipos de jugadores, cada uno siguiendo un camino diferente envejecimiento. Esta historia fue aún peor para Ortiz: “bateadores encarnada” de hecho lo hacen, en promedio, pico temprano y colapsan poco más de treinta años.

Si Boston considera su pasado reciente, su edad y su tamaño, deben, sin duda, han cortado David Ortiz.

Luego, en 2003, estadístico Nate Silver introdujo un nuevo modelo, al que llamó PECOTA, para predecir el rendimiento del jugador. Se demostró que es el mejor y, también, el más fresco.

Plata buscó doppelgangers jugadores. Así es como funciona. Construir una base de datos de todos los jugadores de las Grandes Ligas nunca, más de 18.000 hombres. E incluir todo lo que sabe acerca de esos jugadores: su altura, edad y posición; su casa se ejecuta, promedio de bateo, camina, y ponches por cada año de sus carreras. Ahora, encontrar los veinte jugadores de béisbol que se ven más similar a Ortiz justo hasta ese momento de su carrera, los que jugaron como lo hizo cuando tenía 24, 25, 26, 27, 28, 29, 30, 31, 32 y 33 . En otras palabras, encuentre sus doppelgangers. Y luego ver cómo progresaban carreras Doppelgangers' de Ortiz.

Una búsqueda doppelganger es otro ejemplo de acercamiento. Se hace un zoom en el pequeño subgrupo de personas más similares a una persona determinada. Y, al igual que con todos los zoom, se pone mejor cuanto más datos que tiene. Resulta, doppelgangers de Ortiz dieron una predicción muy diferente para el futuro de Ortiz. doppelgangers de Ortiz incluidos Jorge Posada y Jim Thome. Estos jugadores comenzaron sus carreras un poco lento; tenido estallidos sorprendentes en sus treinta años, con un poder de clase mundial; y luego tuvo problemas en sus treinta y pocos años.

Plata y luego predijo cómo Ortiz haría en función de cómo estos doppelgangers terminaron haciendo. Y esto es lo que encontró: que recuperaron su poder. Para las esposas trofeo, Simmons puede tener razón: cuando se va, se va. Pero para doppelgangers de Ortiz, cuando se fue, volvió.

La búsqueda doppelganger, la mejor metodología vez ha usado para predecir el rendimiento del jugador de béisbol, dijo Boston debe ser paciente con Ortiz. Y Boston de hecho era paciente con su toletero de envejecimiento. En 2010, el promedio de Ortiz se elevó a 0.270. Bateó 32 jonrones e hizo el equipo de las estrellas. Esto comenzó una serie de cuatro juegos consecutivos de Estrellas por Ortiz. En 2013, bateando en su tercera posición tradicional en la alineación, a la edad de treinta y siete años, Ortiz bateó .688 como Boston venció a San Luis, 4 juegos a 2, en la Serie Mundial. Ortiz fue elegido MVP de la Serie Mundial.\*

Tan pronto como terminé de leer el enfoque de Nate Silver para predecir la trayectoria de los jugadores de béisbol, inmediatamente me puse a pensar acerca de si yo podría tener un doppelganger, también.

Doppelganger búsquedas son prometedores en muchos campos, no sólo el atletismo. Pude encontrar a la persona que comparte la mayoría de los intereses conmigo? Tal vez si me pareció la persona más parecido a mí, podríamos pasar el rato. Tal vez él sabría algunos restaurantes que nos gustaría. Tal vez me podría introducir a cosas que tenía ni idea de que podría tener una afinidad por.

Una búsqueda doppelganger zoom sobre los individuos e incluso en los rasgos de los individuos. Y, al igual que con todos los zoom, se hace más agudo cuanto más datos que tiene. Supongamos que yo busqué para mi doppelganger en un conjunto de datos de una decena de personas. Podría encontrar a alguien que comparte mi interés por los libros. Supongamos que yo busqué para mi doppelganger en un conjunto de datos de un millar de personas. Que podría encontrar a alguien que tenía una cosa para los libros de física populares. Pero supongo que busqué para mi doppelganger en un conjunto de datos de cientos de millones de personas. Entonces podría ser capaz de encontrar a alguien que era realmente, realmente similar a mí.

Un día, fui doppelganger caza en las redes sociales. El uso de todo el corpus de los perfiles de Twitter, busqué las personas en el planeta que tienen los intereses más comunes conmigo.

Por supuesto que puede decir mucho acerca de mis intereses de quien sigo en mi cuenta de Twitter. En general, sigo a unas 250 personas, mostrando mis pasiones para los deportes, la política, comedia, ciencia y folksingers judías morosos.

Entonces, ¿hay alguien ahí fuera en el universo que sigue toda 250 de estas cuentas, mi gemelo Twitter? Por supuesto no. Doppelgangers no son idénticos a nosotros, sólo es similar. Tampoco hay alguien que sigue a 200 de las cuentas que sigo. O incluso 150.

Sin embargo, yo eventualmente encontrar una cuenta que siguió a un increíble 100 de las cuentas que sigo: Country Music Radio Hoy. Eh? Resulta, Country Music Radio Hoy fue un bot (que ya no existe) que siguió a 750.000 perfiles de Twitter con la esperanza de que seguirían espalda.

Tengo una ex-novia que sospecho que sacar partido de este resultado. Una vez me dijo que era más como un robot que un ser humano.

Bromas aparte, mi conclusión inicial de que mi doppelganger era un bot que siguió a 750.000 cuentas al azar no hacer un punto importante acerca de las búsquedas doppelganger. Para una búsqueda doppelganger para ser verdaderamente precisa, no desea encontrar a alguien que simplemente le gustan las mismas cosas que le gustan. También desea encontrar a alguien que no le gusta las cosas que no le gustan.

Mis intereses son evidentes no sólo de las cuentas que sigo, pero de los que decido no seguir. Estoy interesado en los deportes, la política, comedia, y la ciencia no es un alimento, la moda o el teatro. Mi siguiente espectáculo que me gusta Bernie Sanders, pero no Elizabeth Warren, Sarah Silverman, pero no Amy Schumer, el New Yorker, pero no el Atlántico, mis amigos Noah Popp, Emily Sands, y Josh Gottlieb pero no es mi amigo Sam Asher. (Lo siento, Sam. Sin embargo, su cuenta de Twitter es una repetición.)

De los 200 millones de personas en Twitter, que tiene el perfil más parecido a mí? Resulta que mi doppelganger es Vox escritor Dylan Matthews. Esta era una especie de decepción, a los efectos de mejorar mi consumo de medios, como ya sigo Matthews en Twitter y Facebook y compulsivamente leer sus mensajes Vox. Así que aprender fue mi doppelganger en realidad no ha cambiado mi vida. Pero sigue siendo bastante frío para conocer a la persona más parecida a ti en el mundo, sobre todo si se trata de alguien que admiras. Y cuando termino este libro y dejar de ser un ermitaño, tal vez Matthews y me pueden pasar el rato y hablar de los escritos de James Surowiecki.

La búsqueda doppelganger Ortiz fue ordenado para los aficionados al béisbol. Y mi búsqueda doppelganger fue entretenido, al menos para mí. Pero ¿qué otra cosa pueden revelar estas búsquedas? Por un lado, las búsquedas doppelganger han sido utilizados por muchas de las empresas más grandes de Internet para mejorar dramáticamente sus ofertas y la experiencia del usuario. Amazon utiliza algo así como una búsqueda doppelganger para sugerir qué libros te pueden gustar. Ellos ven lo que las personas similares a usted seleccionar y basan sus recomendaciones al respecto.

Pandora hace lo mismo en la selección de las canciones que puede que desee escuchar. Y así es como Netflix se da cuenta de las películas que te pueden gustar. El impacto ha sido tan profunda que cuando el ingeniero Amazon Greg Linden doppelganger originalmente introducido busca de predecir las preferencias de los lectores de libros, la mejora en las recomendaciones era tan buena que el fundador de Amazon, Jeff Bezos, se puso de rodillas y gritó: ‘No soy digno!’ a Linden.

Pero lo que es realmente interesante acerca de las búsquedas doppelganger, teniendo en cuenta su poder, no es la forma en que comúnmente están siendo usados ahora. Es la frecuencia con la que no se utilizan. Hay grandes áreas de la vida que se pueden mejorar enormemente por el tipo de personalización permiten a estas búsquedas. Participa en nuestra salud, por ejemplo.

Isaac Kohane, un científico informático e investigador médico en Harvard, está tratando de llevar este principio a la medicina. Quiere organizar y recopilar toda nuestra información de la salud para que en lugar de utilizar una talla única para todos, los médicos pueden encontrar pacientes como usted. Luego se puede emplear, diagnósticos y tratamientos más focalizados más personalizadas.

Kohane considera una extensión natural para el campo de la medicina y ni siquiera una particularmente radical. “¿Cuál es el diagnóstico?”, Se pregunta Kohane. “Un diagnóstico es realmente una declaración que usted comparte propiedades con las poblaciones estudiadas previamente. Cuando se diagnostica con un ataque al corazón, Dios no lo quiera, digo que haya una fisiopatología que he aprendido de otras personas significa que usted ha tenido un ataque al corazón “.

El diagnóstico es, en esencia, una especie primitiva de búsqueda doppelganger. El problema es que los conjuntos de datos médicos utilizan para hacer sus diagnósticos son pequeñas. En estos días el diagnóstico se basa en la experiencia de un médico con la población de pacientes que él o ella ha tratado y tal vez complementados con trabajos académicos de pequeñas poblaciones que otros investigadores han encontrado. Como hemos visto, sin embargo, para una búsqueda doppelganger para conseguir realmente bueno, tendría que incluir muchos más casos.

Aquí es un campo donde algunos grandes volúmenes de datos podría ayudar realmente. Entonces, ¿qué está tomando tanto tiempo? ¿Por qué no se utiliza ya ampliamente? El problema radica en la recogida de datos. La mayoría de los informes médicos todavía existen en el papel, enterrado en los archivos, y para aquellos que están informatizadas, a menudo se encerraron en formatos incompatibles. A menudo tenemos mejores datos, Kohane señala, en el béisbol que en salud. Pero las medidas simples que recorrer un largo camino. Kohane habla repetidamente de “fruta madura”. Cree, por ejemplo, que solamente crear un conjunto de datos completa de los cuadros de altura y peso de los niños y las enfermedades que pudieran tener sería revolucionario para la pediatría. senda de crecimiento de cada niño podría entonces ser comparado con senda de crecimiento de cualquier otro niño. Una computadora podría encontrar a los niños que estaban en una trayectoria similar y la bandera automáticamente los patrones preocupantes. Podría detectar la altura de un niño de estabilizarse antes de tiempo, lo que en ciertos escenarios es probable que apunte a una de las dos causas posibles: hipotiroidismo o un tumor cerebral. El diagnóstico



precoz en ambos casos sería un gran beneficio. “Estos son aves raras”, según Kohane, “uno de cada diez mil tipo de eventos. Los niños, por lo general, son saludables. Creo que podríamos diagnosticarlos antes, al menos un año antes. El cien por ciento, lo que pudimos “.

James Heywood es un empresario que tiene un enfoque diferente para hacer frente a las dificultades que vinculan los datos médicos. Se creó un sitio web, PatientsLikeMe.com, donde los individuos pueden informar de sus propios su información: enfermedades, tratamientos y efectos secundarios. Él ya ha tenido mucho éxito trazar los cursos para distintas enfermedades pueden tomar y cómo se comparan con nuestra comprensión común de ellos.

Su objetivo es reclutar suficientes personas, que cubren las condiciones suficientes, para que las personas puedan encontrar su doppelganger salud. Heywood espera que usted puede encontrar gente de su edad y sexo, con su historia, la presentación de informes síntomas similares a la suya, y ver lo que ha funcionado para ellos. Eso sería un tipo muy diferente de la medicina, de hecho.

## **HISTORIAS DE DATOS**

---

En muchos sentidos, el acto de acercamiento es más valioso para mí que los hallazgos particulares de un estudio en particular, ya que ofrece una nueva forma de ver y hablar de la vida.

Cuando las personas se enteran de que soy un científico de datos y un escritor, a veces se compartirán algún hecho o encuesta conmigo. A menudo me encuentro estos datos aburrida-estática y sin vida. No tiene ninguna historia que contar.

Del mismo modo, los amigos han tratado de conseguir que me uno a ellos en la lectura de novelas y biografías. Pero estos tienen poco interés para mí también. Siempre me encuentro preguntando, “¿Eso sucederá en otras situaciones? ¿Cuál es el principio más general?” Sus historias se siente pequeña y no representativa.

Lo que he tratado de presentar en este libro es algo que, para mí, es como ninguna otra cosa. Se basa en los datos y números; es ilustrativo y de largo alcance. Y sin embargo, los datos es tan rica que se puede visualizar las personas por debajo de ella. Cuando nos acercamos a cada minuto del consumo de agua de Edmonton, veo a la gente levantarse de su sofá al final del período. Cuando nos acercamos a la gente en movimiento de Filadelfia a Miami y comenzar a hacer trampa en sus impuestos, veo estas personas hablan con sus vecinos en su complejo de apartamentos y aprender sobre el truco de impuestos. Cuando nos acercamos a los fanáticos del béisbol de todas las edades, veo a mi propia infancia y la niñez y millones de hombres adultos sin dejar de llorar sobre un equipo que les ganó a cuando tenían ocho años de mi hermano.

A riesgo de sonar una vez más grandioso, creo que los economistas y los datos científicos presentados en este libro están creando no sólo una nueva herramienta, sino un nuevo género. Lo que he tratado de presentar en este capítulo, y gran parte de este libro, son datos tan grande y tan rico, lo que nos permite ampliar de forma estrecha que, sin limitarnos a cualquier ser humano en particular, no representativo, aún podemos decir compleja y historias sugerentes.

## TODO EL MUNDO ES UN LABORATORIO

February 27, 2000, comenzó como un día ordinario en campus de Mountain View de Google. El sol brillaba, los ciclistas se pedalea, los masajistas fueron masajes, los empleados fueron hidratante con agua de pepino. Y luego, en este día ordinario, unos ingenieros de Google tuvieron la idea de que desbloquea el secreto que hoy impulsa gran parte de la Internet. Los ingenieros encontraron la mejor manera de conseguir que haga clic, volviendo, y permanecer en sus sitios.

Antes de describir lo que hicieron, tenemos que hablar sobre la correlación frente a la causalidad, es un gran problema en el análisis de datos, y una que no hemos abordado todavía adecuadamente.

Los medios de comunicación nos bombardean con estudios basados en la correlación aparentemente cada día. Por ejemplo, se nos ha dicho que aquellos de nosotros que beber una cantidad moderada de alcohol tienden a estar en una mejor salud. Esa es una correlación.

¿Quiere esto decir beber una cantidad moderada va a mejorar la propia salud, una relación causal? Tal vez no. Podría ser que la buena salud hace que la gente a beber una cantidad moderada. Los científicos sociales llaman a esta causalidad inversa. O podría ser que hay un factor independiente que hace que tanto el consumo moderado y buena salud. Tal vez de pasar mucho tiempo con los amigos conduce tanto el consumo moderado de alcohol y la buena salud. Los científicos sociales llaman a este sesgo de omisión de variables.

¿Cómo, entonces, podemos establecer con mayor precisión la causalidad? El patrón oro es un experimento aleatorio y controlado. Así es como funciona. Se divide aleatoriamente en dos grupos de personas. Uno de ellos, el grupo de tratamiento, se le pide que hacer o tomar algo. El otro, el grupo de control, no lo es. A continuación, verá cómo responde cada grupo. La diferencia en los resultados entre los dos grupos es su efecto causal.

Por ejemplo, para probar si el consumo moderado hace que una buena salud, es posible escoger al azar a algunas personas a beber un vaso de vino al día durante un año, al azar elegir otros de no beber alcohol durante un año, y luego comparar la salud informado de ambos grupos. Dado que las personas se asignaron aleatoriamente a los dos grupos, no hay razón para esperar que un grupo tendría una mejor salud inicial o, socializaron más. Usted puede confiar en que los efectos del vino son causales. , experimentos controlados aleatorizados son la evidencia más

confiable en cualquier campo. Si una pastilla puede pasar un experimento aleatorizado, controlado, se puede dispensar a la población general. Si no puede pasar esta prueba, no lo hará en los estantes de las farmacias.

experimentos aleatorios cada vez se han utilizado en las ciencias sociales. Esther Duflo, economista francés en el MIT, ha llevado la campaña por un mayor uso de experimentos en economía del desarrollo, un campo que trata de descubrir las mejores maneras de ayudar a las personas más pobres del mundo. Considere el estudio de Duflo, con sus colegas, de manera de mejorar la educación en la India rural, donde más de la mitad de los estudiantes de secundaria no puede leer una oración simple. Una posible razón estudiantes luchan tanto es que los profesores no se presentan de forma coherente. En un día determinado en algunas escuelas de la India rural, más del 40 por ciento de los maestros están ausentes.

La prueba de duflo? Ella y sus colegas dividieron al azar en dos grupos de escuelas. En uno (el grupo de tratamiento), además de su salario base, los maestros se les pagó una cantidad del 50 pequeñas rupias, o alrededor de \$ 1,15 por cada día que se presentaron a trabajar. En el otro, se le dio ningún pago adicional para la asistencia. Los resultados fueron notables. Cuando se pagaron profesores, ausentismo de los maestros se redujo a la mitad. el rendimiento de Student también ha mejorado sustancialmente, con los mayores efectos en las niñas. Al final del experimento, las niñas en las escuelas donde los maestros se pagaron para venir a clase eran 7 puntos porcentuales más probabilidades de ser capaz de escribir.

Según un artículo de The New Yorker, cuando Bill Gates aprendió de la obra de Duflo, estaba tan impresionado que le dijo: “Tenemos que financiar ti.”

### [El ABC de las pruebas A / B](#)

---

Así experimentos aleatorios son el estándar de oro para demostrar la causalidad, y su uso se ha extendido a través de las ciencias sociales. Lo que nos lleva de nuevo a las oficinas de Google el 27 de febrero de 2000. ¿Qué hacer Google en ese día que revolucionó internet?

En ese día, unos ingenieros decidieron llevar a cabo un experimento en el sitio de Google. Se dividieron al azar en dos grupos de usuarios. El grupo de tratamiento se muestra veinte enlaces en las páginas de resultados de búsqueda. El grupo de control se demostró el usual diez. Los ingenieros compararon la satisfacción de los dos grupos en función de la frecuencia con que volvieron a Google.

Esta es una revolución? No parece tan revolucionaria. Yo ya se ha señalado que los experimentos aleatorios han sido utilizados por las compañías farmacéuticas y los científicos sociales. ¿Cómo puede copiarlos ser tan importante?

El punto clave de este y se dio cuenta rápidamente por los ingenieros de Google-es que los experimentos en el mundo digital tienen una ventaja enorme en relación con experimentos en el mundo real. Tan convincente como fuera de línea experimentos aleatorios pueden ser, también son muchos recursos. Para el estudio de Duflo, las escuelas tuvieron que ser contactado,

financiación que se había arreglado, algunos maestros tuvieron que ser pagado, y todos los estudiantes tuvieron que ser probado. experimentos fuera de línea pueden costar miles o cientos de miles de dólares y tomar meses o años para conducir.

En el mundo digital, los experimentos aleatorios pueden ser barato y rápido. No es necesario para reclutar y pagar a los participantes. En su lugar, se puede escribir una línea de código para asignarles al azar a un grupo. No es necesario que los usuarios llenar encuestas. En su lugar, se pueden medir los movimientos del ratón y los clics. No es necesario codificar manualmente y analizar las respuestas. Se puede construir un programa para esto automáticamente por usted. Usted no tiene que ponerse en contacto con nadie. Ni siquiera tiene que decir a los usuarios que forman parte de un experimento.

Esta es la cuarta potencia de Big Data: hace que los experimentos aleatorios, que pueden encontrar verdaderos efectos causales, mucho, mucho más fácil de llevar a cabo en cualquier momento y en cualquier lugar, más o menos, siempre y cuando esté en línea. En la era del Big Data todo el mundo es un laboratorio.

Esta visión se extendió rápidamente a través de Google y luego el resto de Silicon Valley, donde los experimentos controlados aleatorios han cambiado de nombre “pruebas A / B.” En 2011, los ingenieros de Google corrieron siete mil pruebas A / B. Y este número solamente está en aumento.

Si Google quiere saber cómo conseguir más gente haga clic en los anuncios en sus sitios, pueden tratar dos tonos de azul en un tono más anuncios-para el grupo A, otro para el Grupo B. A continuación, Google puede comparar las tasas de clic. Por supuesto, la facilidad de estas pruebas puede conducir a exceso de uso. Algunos empleados sintieron que era porque las pruebas sin esfuerzo, Google se overexperimenting. En 2009, un diseñador frustrado se retiró después de Google pasó por cuarenta y un marginalmente diferentes tonos de azul en las pruebas A / B. Pero la postura de este diseñador a favor del arte sobre la investigación de mercado obsesivo ha hecho poco para detener la propagación de la metodología.

Facebook ahora dirige un millar de pruebas A / B por día, lo que significa que un pequeño número de ingenieros de inicio de Facebook experimentos más controlados y aleatorizados en un día dado que toda la industria farmacéutica se inicia en un año.

Las pruebas A / B se ha extendido más allá de las grandes empresas de tecnología. Un ex empleado de Google, Dan Siroker, trajo esta metodología para la primera campaña de Barack Obama presidencial, que puso a prueba B-diseños de página de inicio A /, lanzamientos de correo electrónico y formularios de donación. Entonces Siroker comenzó una nueva compañía, Optimizely, que permite a las organizaciones para llevar a cabo pruebas rápidas A / B. En 2012, Optimizely fue utilizado por Obama, así como a su oponente, Mitt Romney, para maximizar inscripciones, voluntarios y donaciones. También es utilizado por empresas tan diversas como la revista Netflix, TaskRabbit, y Nueva York.

Para ver cómo las pruebas puede ser valiosa, considere cómo Obama lo utilizó para conseguir más personas que se dedican a su campaña. la página de inicio de Obama incluyó inicialmente una imagen del candidato y un botón debajo de la imagen que invita a la gente a “Registrarse”.



Fue esta la mejor manera de saludar a la gente? Con la ayuda de Siroker, el equipo de Obama podría probar si una imagen y botón diferente podrían hacer que más gente registro en alguna. Serían más gente haga clic en la página de inicio si en lugar aparece una imagen de Obama con una cara más solemnes? Serían más gente haga clic en el botón si en vez dijo “Join Now”? El equipo de Obama mostró usuarios diferentes combinaciones de imágenes y botones y se mide cuántos de ellos hace clic en el botón. Ver si se puede predecir el botón ganar imagen y ganar.

Probado fotos



botones Probado



La ganadora fue la imagen de la familia de Obama y el botón “Más información”. Y la victoria fue enorme. Mediante el uso de esta combinación, el equipo de campaña de Obama estima que tiene un 40 por ciento más de la gente a inscribirse, la compensación de la campaña más o menos \$ 60 millones en fondos adicionales.

Combinación ganadora



Hay otro gran beneficio al hecho de que todas estas pruebas estándar de oro se puede hacer tan barato y fácil: nos libera más lejos de nuestra confianza en nuestra intuición, que, como se señaló en el Capítulo 1, tiene sus limitaciones. Una razón fundamental para la importancia de A / B testing es que las personas son impredecibles. Nuestra intuición falla a menudo para predecir cómo responderán.

Era correcta en el sitio web óptima de Obama su intuición?

Aquí hay algunas pruebas más para su intuición. Los / B pruebas titulares Boston Globe A para averiguar cuáles obtienen la mayor cantidad de gente haga clic en una historia. Trata de adivinar los ganadores de estos pares:

**ONE OF THESE HEADLINES WAS WAY BETTER  
THAN THE OTHER IN GETTING CLICKS.**

	HEADLINE A	HEADLINE B
1.	Can the SnotBot drone save the whales?	Can this drone help save the whales?
2.	Of course "deflated balls" is a top search term in Massachusetts	This top Mass. Google search term is pretty embarrassing
3.	Hookup contest at heart of St. Paul rape trial	No charges in prep school sex scandal
4.	Woman makes bank off rare baseball card	Woman makes \$179,000 off rare baseball card
5.	MBTA projects annual operating deficit will double by 2020	Get ready: the MBTA's deficit is about to double
6.	How Massachusetts helped win you the right to birth control access	How Boston University helped end "crimes against chastity"
7.	When the first subway opened in Boston	Cartoons from when the first subway opened in Boston
8.	Victim and family in prep-school rape trial blame toxic culture	Victim and family in prep-school rape trial releases statement
9.	Guy in "Free Brady" hat is only one able to foil Miley Cyrus prank	Pats fan gets an eyeful for recognizing an undercover Miley Cyrus

Terminados sus conjeturas? Las respuestas están en **negrita** a continuación.

	HEADLINE A	HEADLINE B	WINNER?
1.	<b>Can the SnotBot drone save the whales?</b>	Can this drone help save the whales?	53% more clicks for A
2.	Of course "deflated balls" is a top search term in Massachusetts	<b>This top Mass. Google search term is pretty embarrassing</b>	986% more clicks for B

	HEADLINE A	HEADLINE B	WINNER?
3.	Hookup contest at heart of St. Paul rape trial	<b>No charges in prep school sex scandal</b>	108% more clicks for B
4.	<b>Woman makes bank off rare baseball card</b>	Woman makes \$179,000 off rare baseball card	38% more clicks for A
5.	MBTA projects annual operating deficit will double by 2020	<b>Get ready: the MBTA's deficit is about to double</b>	62% more clicks for B
6.	How Massachusetts helped win you the right to birth control access	<b>How Boston University helped end "crimes against chastity"</b>	188% more clicks for B
7.	<b>When the first subway opened in Boston</b>	Cartoons from when the first subway opened in Boston	33% more clicks for A
8.	Victim and family in prep-school rape trial blame toxic culture	<b>Victim and family in prep-school rape trial releases statement</b>	76% more clicks for B
9.	Guy in "Free Brady" hat is only one able to foil Miley Cyrus prank	<b>Pats fan gets an eyeful for recognizing an undercover Miley Cyrus</b>	67% more clicks for B

Mi predicción es que tienes más de la mitad derecha, tal vez considerando lo que tendría que hacer clic en. Pero es probable que no adivinas todos ellos correctamente.

¿Por qué? ¿Qué le pasa? ¿Qué conocimientos sobre el comportamiento humano ¿os faltó? ¿Qué lecciones puede aprender de sus errores?

Por lo general, preguntas como éstas después de hacer malas predicciones.

Pero mira lo difícil que es para sacar conclusiones generales de los titulares Globe. En la primera prueba del título, el cambio de una sola palabra, "esto" a "SnotBot," condujo a una gran victoria. Esto podría sugerir más detalles ganan. Pero en el segundo titular, "desinfla las bolas", el término detallada, pierde. En el cuarto titular: "hace banco" supera el número \$ 179.000. Esto podría sugerir términos de argot ganan. Pero el término del argot "concurso de conexión" pierde en el tercer titular.

La lección de las pruebas A / B, en gran medida, es que tener cuidado de enseñanzas generales. Clark Benson es el CEO de ranker.com, un sitio de noticias y entretenimiento que se basa en gran medida en las pruebas A / B para elegir titulares y diseño del sitio. "Al final del día, no se puede asumir nada", dice Benson. "Prueba literalmente todo."

Las pruebas rellena los huecos en nuestra comprensión de la naturaleza humana. siempre van a existir estas brechas. Si supiéramos, basado en nuestra experiencia de vida, ¿cuál sería la respuesta, la prueba no sería de valor. Pero no lo hacemos, por lo que es.

Otra razón pruebas A / B es tan importante es que los cambios aparentemente pequeños pueden tener grandes efectos. Como Benson dice, "Estoy constantemente sorprendido con factores menores, menores que tengan un valor desmesurado en las pruebas."

En diciembre de 2012, Google cambió sus anuncios. Se añadió una flecha que apunta hacia la derecha rodeado de un cuadrado.





Observe lo extraño es esta flecha. Apunta hacia la derecha para absolutamente nada. De hecho, cuando estas flechas aparecieron por primera vez, muchos clientes de Google fueron críticos. ¿Por qué fueron añadiendo flechas sin sentido para el anuncio, se preguntaban?

Bueno, Google es de protección de sus secretos comerciales, por lo que no dicen exactamente lo valioso que eran las flechas. Pero ellos dicen que estas flechas habían ganado en las pruebas A / B. La razón es Google les añadió que consiguieron mucha más gente a hacer clic. Y esto, a cambio aparentemente menor sentido hizo Google y sus socios de anuncios montones de dinero.

Entonces, ¿cómo puede usted encontrar estos pequeños retoques que producen ganancias de tallas grandes? Usted tiene que probar un montón de cosas, incluso muchos que parecen triviales. De hecho, los usuarios de Google han señalado en numerosas ocasiones que los anuncios se han cambiado un poquito sólo para volver a su forma anterior. Se han convertido, sin saberlo, los miembros de los grupos de tratamiento en A / B pruebas, pero a costa única de ver estas ligeras variaciones.

#### Centrado experimento (no funcionó)

**Best Selling iPad 2 Case**  
The ZAGGmate™ - Tough Aluminum Case  
with built in Bluetooth Keyboard  
www.zagg.com

#### Experimento estrella verde (no funcionó)

**Foster's Hollywood Restaurant Reviews, Madrid, Spain ...**  
www.tripadvisor.co.uk > ... > Madrid > Madrid Restaurants ~ TripAdvisor  
★★★★ Rating: 3 - 118 reviews  
Foster's Hollywood, Madrid: See 118 unbiased reviews of Foster's Hollywood, rated  
3 of 5 on TripAdvisor and ranked #3647 of 6489 restaurants in Madrid

#### Experimento nueva fuente (no funcionó)

**Live Stock Market News**  
Free Charts, News and Tips from UTVI Experts. Visit us  
Today!  
UTVI.com/Stocks

Las variaciones anteriores nunca llegaron a las masas. Perdieron. Pero eran parte del proceso de selección de ganadores. El camino hacia una flecha puede hacer clic está lleno de estrellas feos, posicionamientos defectuosas, y las fuentes de reclamo.

Puede ser divertido para adivinar lo que hace que la gente haga clic. Y si usted es un demócrata, podría ser bueno saber que la prueba se llevó a Obama más dinero. Pero hay un lado oscuro en las pruebas A / B.

En su excelente libro irresistible, Adam Alter escribe sobre el aumento de las adicciones conductuales en la sociedad contemporánea. Muchas personas están descubriendo aspectos de internet cada vez más difícil de apagar.

Mi favorito conjunto de datos, búsquedas de Google, nos puede dar algunas pistas sobre lo que la gente encuentra más adictivo. Según Google, la mayoría de las adicciones siguen siendo las personas que han luchado durante muchas décadas las drogas, el sexo y el alcohol, por ejemplo. Pero Internet está empezando a hacer sentir su presencia en la lista, con “porno” y “Facebook” ahora entre los diez mejores adicciones reportados.

### **ADICCIONES TOP NOTIFICADOS A GOOGLE, 2016**

---

drogas	Azúcar
Sexo	Amor
Pornografía	Juego
Alcohol	Facebook

Las pruebas A / B puede estar jugando un papel en hacer que Internet sea tan maldito adictivo.

Tristán Harris, un “especialista en ética de diseño”, fue citado en irresistible para explicar por qué las personas tienen un momento muy difícil resistirse a ciertos sitios de Internet: “Hay miles de personas en el otro lado de la pantalla cuya tarea es la de romper el auto -Reglamento usted tiene “.

Y estas personas están utilizando la prueba A / B.

A través de pruebas, Facebook puede darse cuenta de que hacer un botón en particular un color en particular hace que la gente a volver a su sitio más a menudo. Para que cambien el botón para ese color. Entonces pueden darse cuenta de que una fuente en particular hace que la gente a volver a su sitio más a menudo. Para que cambien el texto a esa fuente. Entonces pueden darse cuenta de que la gente enviando un correo electrónico en un momento determinado se les harán volver a su sitio más a menudo. Por lo que un correo electrónico de personas en ese momento.

Muy pronto, Facebook se convierte en un sitio optimizado para maximizar la cantidad de tiempo que la gente en Facebook. En otras palabras, encontrar suficientes ganadores de pruebas A / B y usted tiene un sitio adictivo. Es el tipo de retroalimentación que las compañías de cigarrillos nunca tuvieron.

Las pruebas A / B es cada vez más una herramienta de la industria del juego. Como se discute Alter, World of Warcraft A / B pruebas diferentes versiones de su juego. Una misión puede pedirle que matar a alguien. Otro podría pedirle que salvar algo. Los diseñadores de juegos pueden dar diferentes muestras de diferentes misiones jugadores y luego ver cuáles mantienen más personas jugando. Que podrían encontrar, por ejemplo, que la misión que le ha pedido que salvar a una persona tiene la gente a regresar el 30 por ciento más a menudo. Si resultan ser muchas, muchas misiones, empiezan a encontrar más y más ganadores. Estos 30 por ciento de

victorias sumar, hasta que tengan un juego que mantiene a muchos hombres adultos escondidos en el sótano de sus padres.

Si está un poco perturbado por esto, yo estoy con vosotros. Y vamos a hablar un poco más sobre las implicaciones éticas de este y otros aspectos de grandes volúmenes de datos cerca del final de este libro. Sin embargo, para mejor o peor, la experimentación es ahora una herramienta crucial en la caja de herramientas de los científicos de datos. Y hay otra forma de experimentación sentado en esa caja de herramientas. Se ha utilizado para hacer una variedad de preguntas, incluyendo si los anuncios de televisión realmente funcionan.

### CRUEL-pero esclarecedor-EXPERIMENTOS DE LA NATURALEZA

---

Sus 22 enero 2012, y los New England Patriots están jugando los Baltimore Ravens en el juego de campeonato de la AFC.

Hay un minuto por jugarse en el juego. Los Ravens están abajo, pero tienen la pelota. Los próximos sesenta segundos determinarán qué equipo jugará en el Super Bowl. Los próximos sesenta segundos ayudarán legados jugadores sello. Y el último minuto de este juego va a hacer algo que, para un economista, es mucho más profundo: los últimos sesenta segundos ayudarán finalmente nosotros, de una vez por todas decir, ¿Funcionan los anuncios?

La idea de que los anuncios mejorar las ventas es obviamente crucial para nuestra economía. Pero es exasperantemente difícil de probar. De hecho, este es un ejemplo de libro de exactamente lo difícil que es distinguir entre correlación y causalidad.

No hay duda de que los productos que se anuncian la mayoría también tienen las mayores ventas. Twentieth Century Fox gastó \$ 150 millones de la comercialización de la película Avatar, que se convirtió en la película más taquillera de todos los tiempos. Pero, ¿cuánto de los \$ 2.7 mil millones en la venta de entradas de Avatar fue debido a la pesada de marketing? Parte de la razón por la 20th Century Fox gastó tanto dinero en la promoción era de suponer que sabían que tenían un producto deseable.

Empresas creen que conocen la eficacia de sus anuncios. Los economistas son escépticos que realmente hacen. Universidad de Chicago profesor de economía Steven Levitt, colaborando al mismo tiempo con una empresa de electrónica, era insuficiente cuando la empresa trató de convencerlo de que sabían lo mucho que trabajaron sus anuncios. ¿Cómo, se preguntó Levitt, podían estar tan seguros?

La empresa explicó que, cada año, en los días precedentes día del padre, que la rampa encima de sus gastos de anuncios de televisión. Efectivamente, todos los años, antes del día de padre, tienen las más altas ventas. Uh, tal vez eso es sólo porque una gran cantidad de niños comprar la electrónica por sus padres, sobre todo para los regalos del día de padre, independientemente de la publicidad.

“Tienen la causalidad totalmente al revés”, dice Levitt en una conferencia. Al menos que pudieran tener. No sabemos. “Es un problema muy difícil”, añade Levitt.

Tan importante como este problema es resolver, las empresas son reticentes a realizar experimentos rigurosos. Levitt trató de convencer a la empresa de electrónica para llevar a cabo un experimento aleatorio, controlado con precisión para aprender cómo es eficaz eran sus anuncios de televisión. Dado que las pruebas A / B no es posible en la televisión, sin embargo, esto requeriría de ver lo que sucede sin publicidad en algunas áreas.

Así es como la firma respondió: “¿Estás loco? No podemos dejar de anunciarse en veinte mercados. El director general nos mataría.” Eso terminó la colaboración de Levitt con la empresa.

Lo que nos lleva de nuevo a los patriotas de este juego-Ravens. ¿Cómo pueden los resultados de un partido de fútbol ayudarnos a determinar los efectos causales de la publicidad? Bueno, no nos puede decir a los efectos de una campaña publicitaria en particular de una empresa en particular. Pero puede dar evidencia sobre los efectos medios de publicidad de muchas campañas de gran tamaño.

Resulta que hay un experimento de publicidad oculta en juegos como este. Así es como funciona. Por el momento se juegan estos juegos de campeonato, las empresas han comprado, y producido, sus anuncios de Super Bowl. Cuando las empresas deciden qué anuncios se publiquen, no saben qué equipos jugarán en el juego.

Pero los resultados de las eliminatorias tendrán un enorme impacto en la realidad que observa el Super Bowl. Los dos equipos que califiquen en última instancia, traerán consigo una enorme cantidad de espectadores. Si Nueva Inglaterra, que juega cerca de Boston, gana, muchas más personas en Boston verán el Super Bowl de la gente en Baltimore. Y viceversa.

Para las empresas, es el equivalente de una moneda al aire para determinar si decenas de miles de personas adicionales en Baltimore o Boston serán expuestos a su anuncio, un giro que va a pasar después de que sus puntos son adquiridos y producidos.

Ahora, de vuelta al campo, donde Jim Nantz en la CBS está anunciando los resultados finales de este experimento.

Aquí viene Billy Cundiff, para empatar el juego, y, con toda probabilidad, enviarlo a tiempo extra. Los dos últimos años, dieciséis de dieciséis años en tiros de campo. Treinta y dos yardas para empatar el partido. Y la patada. ¡Estar atento! ¡Estar atento! No es bueno. . . . Y los patriotas toman la rodilla y ahora a tomar el viaje a Indianápolis. Se dirigen a la Super Bowl cuarenta y seis.

Dos semanas más tarde, el Super Bowl XLVI anotaría una cuota de audiencia del 60,3 en Boston y una cuota de 50,2 en Baltimore. Sesenta mil personas más en Boston observaban los 2012 anuncios.

Al año siguiente, los mismos dos equipos se enfrentarán por el campeonato de la AFC. Esta vez, Baltimore ganaría. Las exposiciones de anuncios adicionales para los anuncios del Super Bowl 2013 serían vistos en Baltimore.

	2012 SUPER BOWL RATINGS (BOSTON PLAYS)	2013 SUPER BOWL RATINGS (BALTIMORE PLAYS)
Boston	56.7	48.0
Baltimore	47.9	59.6

Hal Varian, economista jefe de Google; Michael D. Smith, economista de Carnegie Mellon; y he utilizado estos dos juegos y todos los otros Super Bowls 2004-2013 para probar si-y, si es así, ¿cuánto-Super Bowl anuncios de trabajo. Específicamente se analizó si cuando una empresa anuncia una película en el Super Bowl, que ven un gran salto en la venta de entradas en las ciudades que tenían mayor audiencia para el juego.

Que de hecho lo hacen. La gente en ciudades de equipos clasificados para el Super Bowl asistir a películas que se anuncian durante el Super Bowl a una tasa significativamente más alta que hacer los de las ciudades de los equipos que se acaba de perder la calificación. Más gente en esas ciudades vieron el anuncio. Más gente en esas ciudades decidieron ir a la película.

Una explicación alternativa podría ser que el tener un equipo en el Super Bowl hace más propenso a ir a ver películas. Sin embargo, hemos probado un grupo de películas que tenían presupuestos similares y fueron puestos en libertad en momentos similares pero que no hacen publicidad en el Super Bowl. No hubo aumento de la asistencia en las ciudades de los equipos de Super Bowl.

De acuerdo, como es de suponer, los anuncios que funcionan. Esto no es demasiado sorprendente.

Pero no se trata sólo de que funcionen. Los anuncios eran increíblemente eficaz. De hecho, cuando vimos por primera vez los resultados, doble y triple y cuádruple ellos seleccionados para asegurarse de que tenían razón, porque los rendimientos eran tan grandes. La película promedio en nuestra muestra pagó cerca de \$ 3 millones para un espacio publicitario Super Bowl. Consiguieron \$ 8,3 millones en el aumento de la venta de entradas, un retorno de 2,8 a 1 en su inversión.

Este resultado fue confirmado por otros dos economistas, Wesley R. Hartmann y Daniel Klapper, que de forma independiente y anterior se le ocurrió una idea similar. Estos economistas estudian la cerveza y los anuncios de bebidas gaseosas se ejecutan durante el Super Bowl, al mismo tiempo que la utilización de las mayores exposiciones de anuncios en las ciudades de equipos clasificados. Encontraron un retorno de 2.5 a 1 a la inversión. Tan caro como estos anuncios del Super Bowl son, nuestros resultados y los de ellos sugieren que son tan eficaces en la elevando la demanda que las empresas están realmente pagando menos dramática para ellos.

Y ¿qué significa todo esto para nuestros amigos de vuelta en la empresa de electrónica Levitt había trabajado? Es posible que los anuncios del Super Bowl son más rentables que otras formas de publicidad. Pero al menos nuestro estudio sugiere que la publicidad Día todo lo del Padre es probablemente una buena idea.

Una virtud del experimento Super Bowl es que no era necesario asignar a nadie intencionalmente a grupos de tratamiento o de control. Sucedió en base a los rebotes de la suerte en un partido de fútbol. Sucedió, en otras palabras, de forma natural. ¿Por qué es una ventaja? Debido a que los experimentos no naturales, controlados al azar, mientras que super-potente y más fácil de hacer en la era digital, todavía no son siempre posibles.

A veces no podemos actuar juntos en el tiempo. A veces, como con la empresa de electrónica que no quería realizar un experimento en su campaña publicitaria, estamos demasiado invertido en la respuesta a probarlo.

A veces los experimentos son imposibles. Supongamos que usted está interesado en cómo un país responde a la pérdida de un líder. ¿Va a la guerra? No deje de funcionar su economía? No hace nada mucho cambio? Obviamente, no podemos matar a un número significativo de presidentes y primeros ministros y ver qué pasa. Eso sería no sólo es imposible, sino inmoral. Las universidades han acumulado, durante muchas décadas, Juntas de Revisión Institucional (IRB) que determinan si un experimento propuesto es ético.

Así que si queremos conocer los efectos causales en un determinado escenario y no es ético o no factible hacer un experimento, ¿qué podemos hacer? Podemos utilizar lo que los economistas definitorios de la naturaleza lo suficientemente amplia para incluir juegos de fútbol-llaman experimentos naturales.

Para mejor o peor (bueno, claramente peor), hay un componente aleatorio enorme a la vida. Nadie sabe con certeza qué o quién está a cargo del universo. Pero una cosa está clara: el que se está ejecutando el show-las leyes de la mecánica cuántica, Dios, un niño lleno de granos en su ropa interior que simula el universo en su ordenador que, ella, o él no va a través de la aprobación del IRB.

experimentos de la naturaleza en nosotros todo el tiempo. Dos personas se dispararon. Una bala se detiene justo antes de un órgano vital. El otro no. Estos mala suerte son los que hacen la vida injusta. Pero, si le sirve de consuelo, las malas roturas hacen la vida un poco más fácil para los economistas para estudiar. Los economistas utilizan la arbitrariedad de la vida para probar los efectos causales.

De cuarenta y tres presidentes de Estados Unidos, dieciséis han sido víctimas de atentados graves, y cuatro han muerto. Las razones por las que algunos vivían eran esencialmente aleatoria.

Comparación de John F. Kennedy y Ronald Reagan. Ambos hombres tenían balas dirigidas directamente por sus partes del cuerpo más vulnerables. La bala de JFK explotó su cerebro, causándole la muerte poco después. La bala de Reagan detuvo centímetros cortos de su corazón, permitiendo a los médicos para salvar su vida. Reagan vivió, mientras JFK murió, sin rima o la suerte razón justa.

Estos atentados contra la vida de los líderes y la arbitrariedad con la que vivir o morir es algo que sucede en todo el mundo. Comparar Akhmad Kadyrov, de Chechenia, y Adolf Hitler, de Alemania. Ambos hombres han sido pulgadas de distancia de una bomba en pleno funcionamiento. Kadyrov murió. Hitler había cambiado su horario, terminó dejando la sala-bomba unos minutos antes de coger un tren, y por lo tanto sobrevivido.

Y podemos usar de naturaleza fría aleatoriedad para matar Kennedy, pero no de Reagan y ver lo que sucede, en promedio, cuando es asesinado el líder de un país. Dos economistas, Benjamin F. Jones y Benjamin A. Olken, fue lo que hizo. El grupo de control aquí es cualquier país en los años inmediatamente después de un asesinato, por ejemplo cerca de perder, los Estados Unidos a mediados de la década de 1980. El grupo de tratamiento es cualquier país en los años inmediatamente después de un asesinato-completado por ejemplo, los Estados Unidos a mediados de la década de 1960.

Entonces, ¿cuál es el efecto de tener a su líder asesinado? Jones y Olken encontraron que los asesinatos exitosos alteran drásticamente la historia del mundo, teniendo en países radicalmente diferentes caminos. Un nuevo líder hace que los países anteriormente pacíficos para ir a la guerra y previamente países en guerra para conseguir la paz. Un nuevo líder hace que los países económicamente pujantes para comenzar a reventar y los países para iniciar auge económico que revienta.

De hecho, los resultados de este experimento natural basada en el asesinato derrocaron unas pocas décadas de sabiduría convencional sobre cómo funcionan los países. Muchos economistas se inclinaron hacia previamente la opinión de que los líderes eran en gran parte mascarones impotentes empujado por fuerzas externas. No es así, según el análisis del experimento de la naturaleza Jones y de Olken.

Muchos no considerarían este examen de intentos de asesinato de los líderes mundiales un ejemplo de grandes volúmenes de datos. El número de líderes asesinados o casi asesinados en el estudio era ciertamente pequeña al igual que el número de guerras que recibieron o no resultan. Los conjuntos de datos económicos necesarios para caracterizar la trayectoria de una economía eran grandes, pero en su mayor parte son anteriores a la digitalización.

No obstante, este tipo de experimentos naturales, aunque ahora se utilizan casi exclusivamente por los economistas son potentes y tendrán una importancia creciente en una época con más y mejores y más grandes conjuntos de datos. Esta es una herramienta que científicos de datos no van a renunciar a mucho.

Y sí, como debería estar claro ahora, los economistas están jugando un papel importante en el desarrollo de la ciencia de datos. Al menos me gustaría pensar que sí, ya que era mi entrenamiento.

¿Dónde más podemos encontrar en experimentos-naturales es decir, situaciones en el curso de los acontecimientos al azar coloca a las personas en los grupos de tratamiento y control?

El ejemplo más claro es una lotería, por lo que los economistas ellos no les gusta jugar con ellos, que encontramos irracional, pero el estudio de ellos. Si una pelota de ping-pong con un tres en que sube a la superficie, señor Jones será rico. Si se trata de una bola con un lugar de seis, el Sr. Johnson será.

Para probar los efectos causales de las ganancias inesperadas monetarias, los economistas comparan los que ganan loterías para los que compran boletos, pero pierde. Estos estudios

generalmente han encontrado que ganar la lotería no te hace feliz en el corto plazo, pero lo hace en el largo plazo.\*

Los economistas también pueden utilizar la aleatoriedad de las loterías para ver cómo cambia la vida de uno cuando un vecino se hace rico. Los datos muestran que si su vecino gana la lotería puede tener un impacto en su propia vida. Si su vecino gana la lotería, por ejemplo, que son más propensos a comprar un coche caro, como un BMW. ¿Por qué? Casi con toda seguridad, los economistas sostienen, la causa son los celos después de que su vecino más rico compró su propio coche caro. Atribuírsele a la naturaleza humana. Si el señor Johnson ve al señor Jones conducir un BMW nuevo, el Sr. Johnson quiere uno, también.

Por desgracia, el Sr. Johnson a menudo no pueden permitirse este BMW, por lo que los economistas encontraron que los vecinos de ganadores de la lotería son significativamente más propensos a ir a la quiebra. Mantenerse al día con los vecinos, en este caso, es imposible.

Pero los experimentos naturales no tienen que ser explícitamente al azar, como las loterías. Una vez que comience a buscar el azar, lo ves en todas partes, y se puede utilizar para entender cómo funciona nuestro mundo.

Los médicos son parte de un experimento natural. De vez en cuando, el gobierno, por razones esencialmente arbitrarios, cambia la fórmula que utiliza para reembolsar a los médicos para los pacientes de Medicare. Los médicos en algunos condados ven sus cargos por ciertos procedimientos de aumento. Los médicos en otros condados ven sus tasas bajar.

Dos economistas-Jeffrey Clemens y Joshua Gottlieb, un antiguo compañero mío-probaron los efectos de este cambio arbitrario. ¿Los médicos siempre dan a los pacientes el mismo cuidado, la atención que consideran más necesario? ¿O están impulsados por incentivos financieros?

Los datos muestran claramente que los médicos pueden estar motivados por incentivos monetarios. En condados con mayores reembolsos, algunos médicos ordenan sustancialmente más de los procedimientos más-mejor reembolsadas-cirugías de cataratas, colonoscopias, y resonancia magnética, por ejemplo.

Y luego, la gran pregunta: hacer sus pacientes les va mejor después de conseguir todo esto un cuidado especial? Clemens y Gottlieb informaron sólo “pequeños impactos para la salud.” Los autores no encontraron ningún efecto estadísticamente significativo sobre la mortalidad. Dar más fuertes incentivos financieros a los médicos a ordenar algunos procedimientos, este experimento natural sugiere, y algunos ordenará más procedimientos que no tienen mucha diferencia para la salud de los pacientes y no parecen prolongar sus vidas.

Los experimentos naturales pueden ayudar a responder preguntas de vida o muerte. También pueden ayudar con preguntas que, a algunos jóvenes, se sienten como de vida o muerte.

Stuyvesant High School (conocido como “Stuy”) está alojado en una de diez pisos, \$ 150 millones de bronceado, edificio de ladrillo con vistas al río Hudson, a pocas cuadras del World Trade Center, en el bajo Manhattan. Stuy es, en una palabra, impresionante. Ofrece cincuenta y cinco clases de Colocación Avanzada (AP), siete idiomas, y electivas en la historia judía, la ciencia ficción y la literatura de origen asiático. Aproximadamente una cuarta parte de sus



graduados son aceptados a una liga de la hiedra o similarmente prestigiosa universidad. Stuyvesant entrenado profesor de física de la Universidad de Harvard Lisa Randall, estratega de Obama, David Axelrod, el actor ganador de un Oscar Tim Robbins, y el novelista Gary Shteyngart. Sus altavoces de graduación han incluido Bill Clinton, Kofi Annan, y Conan O'Brien.

La única cosa más notable que las ofertas y graduados de Stuyvesant es su coste: cero dólares. Es una escuela secundaria pública y, probablemente, el país es mejor. De hecho, un estudio reciente utilizó 27 millones de críticas por 300.000 estudiantes y padres para clasificar cada escuela secundaria pública en los Estados Unidos. Stuy el número uno. No es de extrañar, entonces, que los padres ambicioso, de clase media de Nueva York y su descendencia pueden igualmente ambicioso obsesionarse con la marca del Stuy.

Para Ahmed Yilmaz,\* el hijo de un agente de seguros y maestro en Queens, Stuy era “la escuela secundaria.”

“Trabajo de clase y las familias inmigrantes Stuy ver como una salida”, explica Yilmaz. “Si su hijo va a Stuy, que va a ir a una y veinte de la mejor universidad de fiar,. La familia va a estar bien “.

Entonces, ¿cómo se puede conseguir en Stuyvesant? Usted tiene que vivir en uno de los cinco distritos de la ciudad de Nueva York y puntuación por encima de un cierto número en el examen de admisión. Eso es. No hay recomendaciones, sin ensayo, sin la admisión legado, ninguna acción afirmativa. Un día, una prueba, una puntuación. Si su número está por encima de un cierto umbral, que se encuentra.

Cada mes de noviembre, aproximadamente 27.000 jóvenes de Nueva York rendir el examen de admisión. La competencia es brutal. Menos del 5 por ciento de los que toman la prueba de entrar en Stuy.

Yilmaz explica que su madre había “trabajado su culo fuera” y poner el poco dinero que tenía en su preparación para la prueba. Después de meses que pasan todas las tardes de lunes a viernes y fines de semana completos que se preparan, Yilmaz estaba seguro de que iba a entrar en Stuy. Todavía recuerda el día que recibió el sobre con los resultados. Se perdió por dos cuestiones.

Le pregunté cómo se sentía. “¿Qué se siente como,” él respondió, “para que su mundo se desmorona cuando estás en la escuela secundaria?”

Su premio de consolación era apenas mal estado-Ciencias del Bronx, otra escuela pública exclusivo y altamente clasificado. Pero no fue Stuy. Y Yilmaz sintió Ciencias del Bronx era más una escuela especializada destinada a los técnicos. Cuatro años más tarde, fue rechazado de Princeton. Asistió Tufts y ha barajado través de unas pocas carreras. Hoy es un empleado de un éxito razonable en una empresa de tecnología, aunque dice que su trabajo es “torturantes” y no tan bien compensada como le gustaría.

Más de una década después, Yilmaz admite que a veces se pregunta cómo la vida habría jugado a cabo si hubiera ido a Stuy. “Todo sería diferente”, dice. “Literalmente, todos los que conozco sería diferente.” Se pregunta si Stuyvesant le habría dado lugar a mayores puntuaciones

del SAT, una universidad como Princeton o Harvard (ambos de los cuales se considera significativamente mejor que la de Tufts), y tal vez más lucrativo o cumplimiento empleo.

Puede ser cualquier cosa, desde el entretenimiento a la auto-tortura para los seres humanos que juegan a cabo hipotéticos. ¿Cómo sería mi vida si lo hizo el movimiento en esa chica o ese chico? Si tomara ese trabajo? Si fui a esa escuela? Pero estos que tal sí parecen incontestables. La vida no es un juego de vídeo. No se puede reproducirlo en diferentes escenarios hasta que obtenga los resultados que desea.

Milan Kundera, escritor de origen checo, tiene una cita sustancial acerca de esto en su novela *La insoportable levedad del ser*: “La vida humana se produce sólo una vez, y la razón por la que no podemos determinar cuál de nuestras decisiones son buenas y cuáles malas es que en una dada la situación que podemos hacer sólo una decisión; que no se les concede una segunda, tercera o cuarta vida en la que comparar varias decisiones “.

Yilmaz nunca va a experimentar una vida en la que de alguna manera logró marcar dos puntos más en esa prueba.

Pero tal vez hay una manera que podemos obtener alguna información sobre cómo diferentes su vida puede o no haber sido por hacer un estudio de un gran número de estudiantes de la Escuela Secundaria Stuyvesant.

La metodología contundente, ingenuo sería comparar todos los estudiantes que fueron a Stuyvesant y todos aquellos que no lo hicieron. Podríamos analizar cómo se llevan a cabo en los exámenes AP y SAT, y lo que los colegios que fueron aceptados en. Si hacemos esto, nos encontraríamos con que los estudiantes que fueron a Stuyvesant puntuación mucho más alta en las pruebas estandarizadas y son aceptados a sustancialmente mejores universidades. Pero, como ya hemos visto en este capítulo, este tipo de pruebas, por sí mismo, no es convincente. Tal vez la razón estudiantes Stuyvesant realizan tanto mejor es que Stuy atrae mucho mejores estudiantes en el primer lugar. Correlación aquí no prueba la causalidad.

Para probar los efectos causales de Stuyvesant High School, tenemos que comparar dos grupos que son casi idénticos: uno que recibió el tratamiento Stuy y uno que no lo hizo. Necesitamos un experimento natural. Pero ¿dónde podemos encontrarlo?

La respuesta: los estudiantes, como Yilmaz, que obtuvieron una puntuación muy, muy cerca del punto de corte necesaria para asistir a Stuyvesant.\* Los estudiantes que no alcanzaron el límite son el grupo de control; los estudiantes que acaban de hacer el corte son el grupo de tratamiento.

Hay pocas razones para sospechar que los estudiantes de ambos lados del corte difieren mucho en el talento o la unidad. Lo que, después de todo, hace que una persona acaba de marcar un punto o dos más alto en una prueba que otro? Tal vez el más bajo de puntuación durmió diez minutos demasiado poco o comió un desayuno menos nutritivos. Tal vez el más alto de puntuación se había acordado una palabra particularmente difícil en la prueba de una conversación que tuvo con su abuela tres años antes.

De hecho, esta categoría de experimentos numéricos utilizando-afilados cortes naturales-es tan poderoso que tiene su propio nombre entre los economistas: regresión discontinua. Siempre que hay un número preciso que divide a las personas en dos grupos diferentes-una discontinuidad-

economistas pueden comparar-o regresión-los resultados de las personas muy, muy cerca de la de corte.

Dos economistas, M. Keith Chen y Jesse Shapiro, se aprovecharon de un corte brusco utilizado por las cárceles federales para probar los efectos de las condiciones de reclusión ásperas en el crimen futuro. internos federales en los Estados Unidos se les da una puntuación, basada en la naturaleza de su crimen y su historia criminal. La puntuación determina las condiciones de su estancia prisión. Aquellos con una puntuación lo suficientemente alta irán a un centro penitenciario de alta seguridad, lo que significa un menor contacto con otras personas, menos libertad de movimiento, y probablemente más violencia por parte de los guardias o por otros reclusos.

Una vez más, no sería justo comparar a todo el universo de los presos que fueron a cárceles de alta seguridad a todo el universo de los presos que fueron a prisiones de baja seguridad. prisiones de alta seguridad incluirán más asesinos y violadores, prisiones de baja seguridad más delincuentes de drogas y ladrones de poca monta.

Pero aquellos justo por encima o justo debajo del umbral numérico aguda tenía antecedentes penales prácticamente idénticos y fondos. Este punto uno mísero, sin embargo, significó una experiencia muy diferente de prisión.

¿El resultado? Los economistas encontraron que los presos asignados a las condiciones más duras eran más propensos a cometer crímenes adicionales una vez que dejaron. Las condiciones en las cárceles difíciles, en lugar de disuadirlos de la delincuencia, que se endurecieron y les hizo más violento al regresar al mundo exterior.

Así que lo que hizo un espectáculo “de regresión discontinuidad” de la Escuela Secundaria Stuyvesant? Un equipo de economistas del MIT y de Duke-Atila Abdulkadiroglu, Joshua Angrist y Parag-Pathak realizó el estudio. Se compararon los resultados de los alumnos de Nueva York en ambos lados del corte. En otras palabras, estos economistas analizaron cientos de estudiantes que, como Yilmaz, perdieron Stuyvesant por una o dos preguntas. Ellos los compararon con cientos de estudiantes que tenían un mejor día de la prueba e hicieron Stuy por una o dos preguntas. Sus medidas de éxito fueron las puntuaciones de AP, los resultados del SAT, y las clasificaciones de las universidades a las que finalmente asistieron.

Sus impresionantes resultados se explicaban en el título le dieron el papel: “Ilusión Elite” Los efectos de Stuyvesant? Nulo. Nada. Cero. Bupkus. Los estudiantes de ambos lados del corte terminaron con resultados indistinguibles de AP y los resultados del SAT indistinguibles y asistieron a universidades de prestigio indistintamente.

Toda la razón por la que los estudiantes Stuy lograr más en la vida que los estudiantes no Stuy, los investigadores llegaron a la conclusión, es que los mejores estudiantes asisten a Stuyvesant en el primer lugar. Stuy no causa llevar a cabo mejor en las pruebas de AP, lo hace mejor en sus exámenes SAT, o terminan en una mejor universidad.

“La intensa competencia por los puestos escolares examen de” los economistas escribieron, “no parece estar justificado por la mejora del aprendizaje de un amplio conjunto de estudiantes”.

¿Por qué podría no importa a qué escuela vas a? Algunos más historias pueden ayudar a conseguir la respuesta. Considere dos estudiantes más, Sarah Jessica Kaufmann y Eng, dos jóvenes neoyorquinos que tanto soñó desde muy temprana edad de ir a Stuy. La puntuación de Kaufmann estaba justo en el punto de corte; lo hizo por una pregunta. “No creo que nada podría ser tan emocionante otra vez,” recuerda Kaufmann. La puntuación de Eng estaba justo por debajo del punto de corte; que se perdió por una pregunta. Kaufmann fue a su escuela ideal, Stuy. Eng no lo hizo.

Entonces, ¿cómo sus vidas terminan? ya que ambos han tenido éxito, y gratificante, carreras, como lo hacen la mayoría de las personas que tengan una calificación en el 5 por ciento de todos los neoyorquinos en los exámenes. Eng, irónicamente, disfrutó de su experiencia en la preparatoria más. Ciencias del Bronx, donde asistió, fue la única escuela secundaria con un museo del Holocausto. Eng descubrió que amaba curación y estudió antropología en Cornell.

Kaufmann sentía un poco perdido en Stuy, donde los estudiantes se centraron en gran medida en grados y sintió que había demasiado énfasis en las pruebas, no en la enseñanza. Llamó a su experiencia “definitivamente un poco de todo.” Pero fue una experiencia de aprendizaje. Se dio cuenta, para la universidad, ella sólo se aplicaría a las escuelas de artes liberales, que tenían un mayor énfasis en la enseñanza. Ella fue aceptado en su escuela sueño, la Universidad de Wesleyan. Allí encontró una pasión por ayudar a los demás, y que ahora es un abogado de interés público.

La gente se adapta a su experiencia, y las personas que van a ser exitosas ventajas encontrar en cualquier situación. Los factores que hacen que el éxito son su talento y su unidad. No son que le da a su discurso de graduación u otras ventajas que las escuelas de marca mayor oferta.

Esto es sólo un estudio, y es probable que se debilita por el hecho de que la mayoría de los estudiantes que no alcanzaron el límite Stuyvesant terminaron en otra escuela bien. Sin embargo, existe una creciente evidencia de que, mientras que ir a una buena escuela es importante, hay poca adquirida en ir a la escuela más grande posible.

Tome la universidad. Qué importa si vas a una de las mejores universidades del mundo, como Harvard, o una escuela sólida tal como Penn State?

Una vez más, existe una clara correlación entre el ranking de la escuela de uno y cuánto dinero la gente hace. Tras diez años de su carrera, el promedio de graduados de Harvard hace \$ 123.000. El graduado medio de la Universidad Estatal de Pensilvania hace \$ 87.800.

Sin embargo, esta correlación no implica causalidad.

Dos economistas, Stacy Dale y Alan B. Krueger, que se piensa de una manera ingeniosa para poner a prueba el papel causal de las universidades de élite en el futuro potencial de ingresos de sus graduados. Tenían un gran conjunto de datos que dio seguimiento a una gran cantidad de información sobre los estudiantes de secundaria, incluso, cuando lo aplican a la universidad, donde fueron aceptados a la universidad, donde asistieron a la universidad, sus antecedentes familiares, y sus ingresos como adultos.

Para conseguir un grupo de tratamiento y control, Dale y Krueger compararon los estudiantes con antecedentes similares que fueron aceptados por las mismas escuelas, pero eligieron

diferentes. Algunos estudiantes que se metieron en Harvard asistieron a Penn State-tal vez para estar más cerca de una novia o novio o porque había un profesor que querían estudiar bajo. Estos estudiantes, en otras palabras, eran igual de talento, de acuerdo con los comités de admisión, como los que fueron a Harvard. Pero tenían diferentes experiencias educativas.

Por eso, cuando dos estudiantes, con antecedentes similares, ambos se metieron en Harvard pero eligieron Universidad Estatal de Pensilvania, lo que pasó? Los resultados de los investigadores fueron tan impresionantes como los de Stuyvesant High School. Aquellos estudiantes terminaron con más o menos los mismos ingresos en sus carreras. Si el salario futuro es la medida, los estudiantes aceptados similares a las escuelas de manera similar prestigiosos que deseen asistir diferentes escuelas terminan en aproximadamente el mismo lugar.

Nuestros periódicos están salpicadas de artículos sobre las personas de gran éxito que asistieron a las escuelas de la Ivy League: gente como el fundador de Microsoft, Bill Gates, y los fundadores de Facebook, Mark Zuckerberg y Dustin Moskovitz, todos los cuales asistieron a la Universidad de Harvard. (Por supuesto, todos ellos abandonaron, levantando preguntas adicionales sobre el valor de la educación de la Ivy League.)

También hay historias de personas que fueron talento suficiente para ser aceptado a una escuela de la Ivy League, decidió asistir a una escuela de menor prestigio, y que tenían vidas muy exitosas: la gente como Warren Buffett, que comenzó en la Escuela Wharton de la Universidad de Pennsylvania, una escuela de negocios de la Ivy League, pero transferido a la Universidad de Nebraska-Lincoln porque era más barato, que odiaba Filadelfia, y pensó que eran las clases de Wharton aburrido. Los datos sugieren, las ganancias-sabio, al menos, que la elección de asistir a una escuela de menor prestigio es una decisión bien, por Buffett y otros.

Este libro se llama Everybody Lies. Por esto, quiero decir que la mayoría de la gente miente a los amigos, a las encuestas, y para sí mismos, para verse mejor.

Pero el mundo también nos miente al presentarnos con datos erróneos, engañosos. El mundo nos muestra un gran número de exitosos graduados de Harvard, pero menos graduados con éxito de Penn State, y suponemos que hay una enorme ventaja para ir a Harvard.

Al hacer inteligentemente sentido de experimentos de la naturaleza, podemos correctamente dar sentido a los datos del mundo-para encontrar lo que es realmente útil y lo que no lo es.

Los experimentos naturales se relacionan con el capítulo anterior, también. A menudo requieren hacer zoom sobre los grupos de tratamiento y control: las ciudades en el experimento del Super Bowl, los condados en el experimento de fijación de precios de Medicare, los estudiantes cercanos al límite en el experimento Stuyvesant. Y el zoom, como se discutió en el capítulo anterior, a menudo requiere, conjuntos de datos completos de grandes-del tipo que son cada vez más disponibles como el mundo se digitaliza. Dado que no sabemos cuando la naturaleza elegirá a ejecutar sus experimentos, no se puede establecer una pequeña encuesta para medir los

resultados. Necesitamos una gran cantidad de datos existentes para aprender de estas intervenciones. Necesitamos grandes volúmenes de datos.

Hay un punto más importante hacer acerca de los experimentos de cualquiera de nuestros propios como de la naturaleza-se detalla en este capítulo. Gran parte de este libro se ha centrado en la comprensión del mundo y de cómo cuesta mucho racismo Obama, ¿cuántos hombres son realmente homosexuales, cómo los hombres y las mujeres son inseguras sobre sus cuerpos. Pero estos experimentos controlados o naturales tienen una inclinación más práctico. Su objetivo es mejorar nuestra toma de decisiones, para ayudarnos a aprender las intervenciones que funcionan y las que no lo hacen.

Las empresas pueden aprender cómo conseguir más clientes. El gobierno puede aprender a utilizar el reembolso a los mejores médicos motivar. Los estudiantes pueden aprender lo que las escuelas van a resultar más valiosa. Estos experimentos demuestran el potencial de grandes volúmenes de datos para reemplazar conjeturas, la sabiduría convencional, y las correlaciones de mala calidad con lo que realmente funciona-causalmente.

## PARTE III

---

### Big data: Manipular con cuidado

## **Big Data, BIG SCHMATA? Lo que no puede HACER**

“SETH, Lawrence Summers le gustaría reunirse con usted “, dijo el correo electrónico, de manera algo críptica. Fue a partir de uno de mi doctorado asesores, Lawrence Katz. Katz no me dijo por qué Summers estaba interesado en mi trabajo, aunque más tarde descubrí Katz había sabido todo el tiempo.

Me senté en la sala de espera del despacho de Summers. Después de algún retraso, el ex secretario del Tesoro de los Estados Unidos, el ex presidente de la Universidad de Harvard, y ganador de algunos de los mayores premios de economía, me citó en su interior.

Summers comenzaron la reunión mediante la lectura de mi artículo sobre el efecto del racismo en Obama, que su secretaria había impreso para él. Summers es un lector de velocidad. A medida que lee, de vez en cuando saca la lengua y hacia la derecha, mientras que sus ojos cambian rápidamente izquierda y derecha y abajo de la página. Summers lectura de un documento de ciencias sociales me recuerda a un gran pianista realizar una sonata. Él está tan concentrada que parece perder la pista de todo lo demás. En menos de cinco minutos, había terminado mi papel de treinta páginas.

“Usted dice que las búsquedas de Google para 'negro' sugieren racismo”, dijo Summers. “Eso parece plausible. Ellos predicen que Obama recibe menos apoyo que Kerry. Eso es interesante. Realmente podemos pensar en Obama y Kerry como el mismo?”

“Ellos fueron clasificados como tener ideologías similares por parte de los científicos políticos”, respondí. “Además, no existe una correlación entre el racismo y los cambios en la Casa de votación. El resultado se mantiene fuerte, incluso cuando añadimos los controles de la demografía, la asistencia a la iglesia, y la propiedad de armas.” Así es como hablan los economistas. Yo había crecido animada.

Summers se detuvo y me miró. Se volvió brevemente a la televisión en su oficina, que se sintoniza a la CNBC, entonces me miró de nuevo, y luego miró a la TV, luego a mí. “Está bien, me gusta este trabajo”, dijo Summers. “¿Qué más estás trabajando?”



Los siguientes sesenta minutos puede haber sido el más intelectualmente estimulante de mi vida. Summers y yo hablamos acerca de las tasas de interés y la inflación, la policía y el crimen, las empresas y la caridad. Hay una razón por la que muchas personas que cumplen con los veranos son cautivados. He tenido la suerte de hablar con algunas personas muy inteligentes en mi vida; Summers me pareció el más inteligente. Está obsesionado con las ideas, más que cualquier otra cosa, que parece ser lo que a menudo le mete en problemas. Tuvo que renunciar a su presidencia en Harvard después de sugerir la posibilidad de que parte de la razón de la escasez de mujeres en las ciencias podría ser que los hombres tienen más variación en su coeficiente intelectual. Si se encuentra con una idea interesante, Summers tiende a decir que, incluso si se ofende a algunos oídos.

Ahora era una media hora después de la hora de finalización programada para nuestra reunión. La conversación era intoxicante, pero todavía no tenía idea de por qué estaba allí, ni cuando se suponía que debía salir, ni cómo iba a saber cuando se suponía que debía irse. Tengo la sensación, en este punto, que los veranos mismo pueden haber olvidado por qué había establecido esta reunión.

Y luego le preguntó al millón de dólares, o quizá mil millones de dólares que se trate. “Usted cree que puede predecir el mercado de valores con estos datos?”

Aha. Aquí en el último fue la razón Summers me había citado a su despacho.

Summers no es la primera persona que me haga esta pregunta en particular. Mi padre ha sido generalmente de apoyo de mis intereses de investigación no convencionales. Pero una vez que lo hizo abordar el tema. “El racismo, el abuso infantil, aborto,” dijo. “¿No pueden hacer ningún dinero de esta experiencia de los suyos?” Amigos y otros miembros de la familia han planteado el tema, también. También lo han hecho los compañeros de trabajo y extraños en el Internet. Todo el mundo parece querer saber si puedo utilizar Google búsquedas u otras poblaciones de grandes volúmenes de datos-a recoger. Ahora era el ex secretario del Tesoro de los Estados Unidos. Esto fue más grave.

Así pueden las nuevas fuentes de datos grandes predecir con éxito qué acciones maneras se dirige? La respuesta corta es no.

En los capítulos anteriores hemos hablado de los cuatro poderes de grandes volúmenes de datos. Este capítulo tiene que ver con las limitaciones -tanto de grandes volúmenes de datos lo que no podemos hacer con ella y, en ocasiones, lo que no debemos hacer con él. Y un lugar para empezar es contar la historia del intento fallido por Summers y yo para vencer a los mercados.

En el capítulo 3, observamos que los nuevos datos es más probable que produzca grandes beneficios cuando la investigación existente en un determinado campo es débil. Es una verdad lamentable sobre el mundo que va a tener un tiempo mucho más fácil obtener nuevos conocimientos sobre el racismo, el abuso infantil, o el aborto de lo que va a conseguir una nueva visión, rentable sobre cómo un negocio se está realizando. Esto se debe a los enormes recursos que ya se dedican a la búsqueda de incluso la más ligera ventaja en la medición de rendimiento del negocio. La competencia en las finanzas es feroz. Que ya era una huelga en contra de nosotros.

Summers, que no es alguien conocido por effusing sobre la inteligencia de los demás, estaba seguro de que los fondos de cobertura ya estaban muy por delante de nosotros. Me quedé bastante durante nuestra conversación por el gran respeto que sentía por ellos y cómo muchos de mis sugerencias que estaba convencido de que habían golpeado a nosotros. Con orgullo compartido con él un algoritmo que había ideado que me permitió obtener datos más completos Google Trends. Dijo que era inteligente. Cuando le pregunté si renacimiento, un fondo de cobertura cuantitativa, se habría dado cuenta de que el algoritmo, se rió y dijo: “Sí, por supuesto que habrían dado cuenta de eso.”

La dificultad de mantener el ritmo de los fondos de cobertura no fue el único problema fundamental que Summers y yo corrimos contra en el uso de nuevas grandes conjuntos de datos, para vencer a los mercados.

### La maldición de la dimensionalidad

---

Suponga que su estrategia para predecir el mercado de valores es encontrar una moneda, pero uno que se encuentra a través de cuidadosas pruebas suerte. Aquí tiene su metodología: Se etiquetar mil monedas-1 a 1.000. Cada mañana, durante dos años, le da la vuelta cada moneda, registrar si se trataba de cara o cruz, y luego se tenga en cuenta si el índice Standard & Poor 's se fue arriba o hacia abajo ese día. Usted pore a través de todos sus datos. ¡Y voilá! Usted ha encontrado algo. Resulta que el 70,3 por ciento del tiempo de la moneda salió cara 391 del índice S & P Rose. La relación es estadísticamente significativa, de modo altamente. Que ha encontrado su moneda de la suerte!

Sólo la vuelta de la moneda 391 cada mañana y comprar acciones cada vez que sale cara. Sus días de Target camisetas y cenas de fideos ramen son más. Coin 391 es su boleto a la buena vida! O no.

Te has convertido en otra víctima de uno de los aspectos más diabólicos de Puede golpear cada vez que tenga una gran cantidad de variables (o “dimensiones”) “la maldición de la dimensionalidad.” - en este caso, un millar de monedas-persiguiendo no es que muchos observations- en este caso, 504 días de negociación de esos dos años. Una de esas dimensiones con monedas 391, en este caso, es probable que tenga suerte. Disminuir el número de variables-flip sólo cien monedas, y será mucho menos probable que uno de ellos tenga suerte. Aumentar el número de observaciones tratan de predecir el comportamiento del índice S & P durante veinte años-y monedas tendrán dificultades para mantenerse al día.

La maldición de la dimensionalidad es un problema importante con grandes volúmenes de datos, ya que los nuevos conjuntos de datos con frecuencia nos dan exponencialmente más variables que las fuentes de datos tradicionales, cada término de búsqueda, todas las categorías de pío, etc. Muchas personas que dicen predecir el mercado que utiliza alguna fuente de datos grandes simplemente han sido atrapados por la maldición. Todo lo que realmente han hecho es encontrar el equivalente de la moneda 391.

Tomemos, por ejemplo, un equipo de científicos de la Universidad de Indiana y la Universidad de Manchester que afirmaban que podían predecir qué manera los mercados se van sobre la base de lo que la gente twitteando. Construyeron un algoritmo para codificar los estados de ánimo del mundo del día a día en base a los tweets. Ellos usan técnicas similares para el análisis de los sentimientos discutido en el capítulo 3. Sin embargo, se codificaron no sólo un estado de ánimo, pero muchos estados de ánimo, la felicidad, la ira, la bondad, y mucho más. Encontraron que una preponderancia de los tweets que sugieren la calma, como “Me siento tranquilo,” predice que el Dow Jones Industrial Average es probable que aumente seis días más tarde. Un fondo de cobertura fue fundada para explotar sus hallazgos.

¿Cuál es el problema aquí?

El problema fundamental es que se probaron muchas cosas. Y si se prueba suficientes cosas, sólo por casualidad al azar, uno de ellos será estadísticamente significativa. Ellos probaron muchas emociones. Y probaron cada emoción un día antes, dos días antes, tres días antes y hasta siete días antes de la conducta del mercado de valores que estaban tratando de predecir. Y todas estas variables se utilizan para tratar de explicar tan sólo unos meses de subidas y bajadas de Dow Jones.

La calma seis días antes no era un predictor legítimo de la bolsa de valores. La calma seis días antes era el equivalente de datos grandes de nuestra hipotética de la moneda 391. El fondo de cobertura de tuit fue cerrada un mes después de comenzar debido a rendimientos mediocres.

Los fondos de cobertura que tratan de tiempo de los mercados con los tweets no son los únicos que luchan contra la maldición de la dimensionalidad. También lo son los numerosos científicos que han tratado de encontrar las claves genéticas a lo que somos.

Gracias al Proyecto del Genoma Humano, ahora es posible recoger y analizar el ADN completo de las personas. El potencial de este proyecto parecía enorme.

Tal vez podríamos encontrar el gen que causa la esquizofrenia. Tal vez podríamos descubrir el gen que causa la enfermedad de Alzheimer y de Parkinson y la esclerosis lateral amiotrófica. Tal vez podríamos encontrar el gen que causa-trago-inteligencia. ¿Existe un gen que puede añadir un montón de puntos de CI? ¿Existe un gen que hace que un genio?

En 1998, Robert Plomin, un genetista del comportamiento prominente, afirmó haber encontrado la respuesta. Recibió un conjunto de datos que incluye el ADN y el coeficiente intelectual de cientos de estudiantes. Se comparó el ADN de -aquellos “genios” con CI de 160 o superior para el ADN de aquellos con un CI promedio.

Él encontró una notable diferencia en el ADN de estos dos grupos. Se encuentra en un pequeño rincón del cromosoma 6, un gen desconocido pero de gran alcance que se utilizó en el metabolismo del cerebro. Una versión de este gen, llamado Igf2r, era dos veces más común en genios.

“En primer gen que se va vinculado con alta inteligencia se informa encontramos”, tituló el New York Times.

Es posible pensar en las muchas cuestiones éticas hallazgo de Plomin planteaba. Se debe permitir a los padres a sus hijos para detectar Igf2r? En caso de que se les permita abortar a un

bebé con la variante de bajo coeficiente intelectual? ¿Hay que modificar genéticamente las personas para darles un alto coeficiente intelectual? No Igf2r correlaciona con la raza? ¿Queremos saber la respuesta a esa pregunta? Debe investigar sobre la genética del IQ continuar?

Antes de bioética tuvieron que hacer frente a cualquiera de estas preguntas espinosas, no era una pregunta más básica para los genetistas, incluyendo PLOMIN sí mismo. Era preciso será el resultado? ¿Era realmente cierto que Igf2r podía predecir el coeficiente intelectual? ¿Era realmente cierto que los genios eran dos veces más propensos a llevar a una cierta variante de este gen?

Nop. Pocos años después de su estudio original, Plomin tiene acceso a otra muestra de personas, que también incluía sus puntuaciones de CI y de ADN. Esta vez, Igf2r no se correlacionó con el CI. Plomin, y esto es un signo de un buen científico-retractado de su reclamación.

Esto, de hecho, ha sido un patrón general en la investigación de la genética y el CI. En primer lugar, los científicos informan de que han encontrado una variante genética que predice IQ. A continuación, los científicos obtienen nuevos datos y descubre su afirmación original equivocado.

Por ejemplo, en un artículo reciente, un equipo de científicos, dirigido por Christopher Chabris, examinaron doce reclamaciones prominentes sobre las variantes genéticas asociadas con el CI. Se examinaron los datos de diez mil personas. No podían reproducir la correlación para cualquiera de los doce.

¿Cuál es el problema con todas estas afirmaciones? La maldición de la dimensionalidad. El genoma humano, los científicos saben ahora, se diferencia en millones de formas. Hay, simplemente, demasiados genes para probar.

Si el resultado es suficiente tweets para ver si se correlacionan con el mercado de valores, se encuentra uno que se correlaciona sólo por casualidad. Si prueba suficientes variantes genéticas para ver si se correlacionan con el CI, se encuentra uno que se correlaciona sólo por casualidad.

¿Cómo puede superar la maldición de la dimensionalidad? Tienes que tener un poco de humildad sobre su trabajo y no caer en el amor con los resultados. Usted tiene que poner estos resultados a través de pruebas adicionales. Por ejemplo, antes de apostar sus ahorros en la moneda de 391, que le gustaría ver cómo lo hace durante el próximo par de años. Los científicos sociales llaman a esto una prueba de “fuera de la muestra”. Y los más variables se intente, más humilde que tienen que ser. Los más variables se intente, más dura la prueba fuera de la muestra tiene que ser. También es crucial para mantener un registro de todas las pruebas se intenta. A continuación, puede saber exactamente qué tan probable es que son víctimas de la maldición y cómo escépticos que debería ser de sus resultados. Lo que nos lleva de nuevo a Larry Summers y yo. Así es como tratamos de vencer a los mercados.

La primera idea de Summers era utilizar las búsquedas para predecir las ventas futuras de los productos clave, como iPhones, que podría arrojar luz sobre la evolución futura de las acciones de una empresa, tales como Apple. Hubo efectivamente una correlación entre las búsquedas de

“iPhones” y las ventas de iPhones. Cuando la gente está buscando en Google un montón de “iPhones”, se puede apostar una gran cantidad de teléfonos están siendo vendidos. Sin embargo, esta información ya se incorporó en el precio de las acciones de Apple. Claramente, cuando había un montón de búsquedas en Google de “iPhones”, los fondos de cobertura también habían dado cuenta de que sería un éxito de ventas, independientemente de si se utilizan los datos de búsqueda o de alguna otra fuente.

siguiente idea de Summers fue predecir futuras inversiones en los países en desarrollo. Si un gran número de inversores se va a invertir dinero en países como Brasil o México en un futuro próximo, a continuación, las acciones de las empresas de estos países seguramente subirían. Tal vez podríamos predecir un aumento de la inversión con las búsquedas de Google tales-clave como “invertir en México” o “oportunidades de inversión en Brasil.” Esto resultó ser un callejón sin salida. ¿El problema? Las búsquedas fueron demasiado raro. En lugar de revelar patrones significativos, estos datos de búsqueda saltaron por todo el lugar.

Tratamos las búsquedas de las acciones individuales. Tal vez si las personas estaban buscando “GOOG”, esto significaba que estaban a punto de comprar Google. Estas búsquedas parecían predecir que las existencias se pueden negociar mucho. Pero ellos no pueden predecir si las existencias se aumentará o disminuirá. Una limitación importante es que estas búsquedas no nos dicen si alguien estaba interesado en comprar o vender las acciones.

Un día, con entusiasmo mostró Summers una nueva idea que tenía: búsquedas anteriores para “comprar oro” parecían estar en correlación con los futuros aumentos en el precio del oro. Summers me dijo que debería probarlo en el futuro para ver si seguía siendo preciso. Que dejó de funcionar, tal vez debido a que algunos fondos de cobertura había encontrado la misma relación.

Al final, en unos meses, no hemos encontrado nada útil en nuestras pruebas. Sin lugar a dudas, si hubiéramos buscado una correlación con el comportamiento del mercado en cada uno de los mil millones de términos de búsqueda de Google, nos hemos encontrado uno que funcionaba, sin embargo débilmente. Pero probablemente habría sido sólo nuestra propia moneda de 391.

### [LA énfasis excesivo en lo que es MENSURABLE](#)

---

En marzo de 2012, Zoe Chance, profesor de Marketing de la Universidad de Yale, recibió un pequeño podómetro blanco en su buzón de la oficina en el centro de New Haven, Connecticut. Ella dirigido a estudiar cómo este dispositivo, que mide los pasos que se dan durante el día y le da puntos, como resultado, puede inspirar a hacer más ejercicio.

Lo que sucedió después, mientras se relata en una charla TEDx, era una pesadilla de datos grandes. Oportunidad se obsesionó y adicto a incrementar sus números que ella comenzó a caminar por todas partes, desde la cocina hasta la sala de estar, el comedor, al sótano, en su oficina. Se dirigió temprano en la mañana, tarde en la noche, en casi todas las horas del día veinte mil pasos en un período de veinticuatro hora determinada. Se comprobó sus podómetro cientos de veces al día, y mucho de lo que quedaba de la comunicación humana estaba con otros

usuarios en línea podómetro, discutir estrategias para mejorar los resultados. Ella recuerda poner el contador de pasos en su hija de tres años de mi hija cuando su hija estaba caminando, porque ella estaba tan obsesionado con conseguir el número más alto.

Oportunidad se obsesionó tanto con la maximización de este número que había perdido toda perspectiva. Se olvidó la razón por la que alguien quiere conseguir el número más alto, hacer ejercicio, no tener a su hija caminar unos pasos. Tampoco se complete cualquier investigación académica sobre el podómetro. Finalmente se deshizo del dispositivo después de caer tarde una noche, agotado, mientras que intenta conseguir en más etapas. A pesar de que es un investigador basado en datos de profesión, la experiencia le afectó profundamente. “Me hace escépticos sobre si tener acceso a los datos adicionales es siempre una buena cosa”, dice Chance.

Esta es una historia extrema. Pero apunta a un posible problema con la gente que usa los datos para tomar decisiones. Los números pueden ser seductora. Podemos crecer obsesionado con ellos, y al hacerlo, podemos perder de vista de las consideraciones más importantes. Posibilidad Zoë perdió de vista, más o menos, del resto de su vida.

Incluso los caprichos menos obsesivos con los números pueden tener inconvenientes. Considere el énfasis-siglo XXI en las pruebas en las escuelas estadounidenses y juzgar los profesores en función de cómo anotan sus estudiantes. Si bien el deseo de adoptar medidas más objetivas de lo que sucede en las aulas es legítima, hay muchas cosas que suceden allí que no pueden ser fácilmente capturadas en número. Por otra parte, todas las pruebas que presionaron a muchos profesores para enseñar a los ensayos y peor. Un pequeño número, como se demostró en un artículo de Brian Jacob y Steven Levitt, trampa de plano en la administración de dichas pruebas.

El problema es el siguiente: las cosas que podemos medir con frecuencia no son exactamente lo que nos importa. Podemos medir cómo los estudiantes hacen en preguntas de opción múltiple. No podemos medir fácilmente el pensamiento crítico, la curiosidad, o el desarrollo personal. Sólo se trata de aumentar un solo número, las puntuaciones de las pruebas fáciles de medida o el número de pasos dados en un día Indiferente siempre ayudar a lograr lo que realmente estamos tratando de lograr.

En sus esfuerzos por mejorar su sitio, Facebook se encuentra con este peligro también. La compañía tiene un montón de datos sobre cómo las personas utilizan el sitio. Es fácil ver si una historia de noticias RSS en particular se le gusta, hace clic en, comentado o compartido. Pero, según Alex Peysakhovich, un científico de datos de Facebook con los que he escrito sobre estas cuestiones, ni uno de ellos es un proxy perfecta para las preguntas más importantes: ¿Cuál fue la experiencia de uso del sitio como? ¿La historia conectar al usuario con sus amigos? ¿Se informa ella sobre el mundo? ¿Se hacen reír?

O considerar la revolución de los datos de béisbol en la década de 1990. Muchos equipos comenzaron a utilizar cada vez más intrincados estadísticas en lugar de depender de la antigua exploradores a persona tomar decisiones. Era fácil de medir la ofensiva y pitcheo, pero no fildeo, por lo que algunas organizaciones terminaron subestimar la importancia de la defensa. De hecho, en su libro *La señal y el ruido*, Nate Silver estima que los Atléticos de Oakland, una organización

basada en los datos de perfil en Moneyball, estaban dando hasta ocho a diez victorias por año a mediados de los años noventa, debido a su defensa pésimo.

La solución no es siempre más grandes datos. Una salsa especial es a menudo necesario para ayudar a trabajar mejor Big Data: el juicio de los seres humanos y encuestas pequeñas, lo que podríamos llamar los datos pequeños. En una entrevista con la plata, Billy Beane, gerente general de la entonces A y el personaje principal de la Fortuna, dijo que en realidad había comenzado aumentando su presupuesto de exploración.

Para llenar los vacíos en su banco de datos gigante, Facebook también tiene que tener un anticuado enfoque: pidiendo a la gente lo que piensan. Todos los días, ya que cargan su News Feed, cientos de usuarios de Facebook se presentan con preguntas sobre las historias que ven allí. conjuntos de datos recogidos automáticamente de Facebook (Me gusta, clics, comentarios) se complementan, en otras palabras, por los datos más pequeños ( “¿Quieres ver este post en su Servicio de Noticias?” “¿Por qué?”). Sí, incluso un éxito espectacular organización de datos grandes como Facebook a veces hace uso de la fuente de información más menospreciado en este libro: una pequeña encuesta.

De hecho, debido a esta necesidad de datos pequeños como un suplemento a sus colecciones de pilar-masiva de clics, gustos, y los equipos de datos de puestos de Facebook se ven diferentes de lo que puede imaginar. Facebook emplea a psicólogos sociales, antropólogos, sociólogos y precisamente para encontrar lo que pierden los números.

Algunos educadores, también, son cada vez más conscientes de los puntos ciegos en grandes volúmenes de datos. Hay un creciente esfuerzo nacional para complementar las pruebas de masas con los datos pequeños. encuestas de los estudiantes han proliferado. También lo han hecho encuestas de los padres y las observaciones del maestro, donde otros educadores experimentados ven un maestro durante una clase.

“Los distritos escolares dan cuenta que no deben centrarse exclusivamente en resultados de las pruebas,” dice Thomas Kane, un profesor de educación en la Universidad de Harvard. Un estudio de tres años por la Fundación Bill y Melinda Gates confirma el valor de la educación tanto de datos grandes y pequeños. Los autores analizaron si las observaciones modelos basados en la prueba de puntuación-, encuestas a los estudiantes o maestros eran mejores en el que los profesores de medición aprendizaje de los estudiantes que más ha mejorado. Cuando pusieron las tres medidas en conjunto en una puntuación compuesta, consiguieron los mejores resultados. “Cada medida añade algo de valor”, concluyó el informe.

De hecho, fue justo cuando estaba aprendiendo que muchas operaciones de grandes volúmenes de datos utilizan datos pequeñas para rellenar los agujeros que me presenté en Ocala, Florida, para cumplir con Jeff Seder. Recuerde, él era el gurú del caballo educado en Harvard que utiliza las lecciones aprendidas a partir de un gran conjunto de datos para predecir el éxito de América del faraón.

Después de compartir todos los archivos de computadora y matemáticas conmigo, Seder admitió que no tenía otra arma: Patty Murray.

Murray, como Seder, tiene una gran inteligencia y de élite credenciales-un título de Bryn Mawr. También dejó la ciudad de Nueva York para la vida rural. “Me gustan los caballos más de los seres humanos”, admite Murray. Pero Murray es un poco más tradicional en sus enfoques para la evaluación de los caballos. Ella, al igual que muchos agentes de caballos, caballos examina personalmente, ver cómo caminan, la comprobación de cicatrices y moretones, e interrogar a sus propietarios.

Murray entonces colabora con Seder a medida que recogen los caballos finales que desean recomendar. Murray olfatea problemas con los caballos, los problemas que los datos del Seder, a pesar de ser el conjunto de datos más innovadores e importantes jamás recopilada en caballos, sigue sin ver.

Estoy prediciendo una revolución basada en las revelaciones de grandes volúmenes de datos. Pero esto no quiere decir que sólo puede arrojar datos en cualquier pregunta. Y grandes volúmenes de datos no elimina la necesidad de que todas las otras formas de los seres humanos han desarrollado a lo largo de los milenios para entender el mundo. Ellos se complementan entre sí.



## MO DE DATOS, MO PROBLEMAS? LO QUE NO debe hacer

Algúnas veces, el poder de grandes volúmenes de datos es tan impresionante que da miedo. Se plantea cuestiones éticas.

### EL PELIGRO DE EMPRESAS AUTORIZADAS

---

Recientemente, tres economistas-Oded Netzer y Alain Lemaire, tanto de Columbia, y Michal Herzstein de la Universidad de Delaware-buscaron formas para predecir la probabilidad de que un prestatario pagar un préstamo. Los investigadores utilizaron datos de Próspero, un sitio de préstamos peer-to-peer. Los prestatarios potenciales escribir una breve descripción de por qué necesitan un préstamo y por qué son propensos a hacer bueno en ella, y los posibles prestamistas deciden si les proporciona el dinero. En general, aproximadamente el 13 por ciento de los prestatarios de pagar su préstamo.

Resulta que el lenguaje que utilizan los prestatarios potenciales es un fuerte predictor de la probabilidad de pagar. Y es un indicador importante aunque no controles para otros prestadores de información relevantes fueron capaces de obtener sobre los prestatarios potenciales, incluyendo las calificaciones de crédito y los ingresos.

A continuación se enumeran diez frases Los investigadores encontraron que se utilizan comúnmente para solicitar un préstamo. Cinco de ellos se correlacionan positivamente con pagar el préstamo. Cinco de ellos se correlacionan negativamente con pagar el préstamo. En otras palabras, cinco tienden a ser utilizado por personas de confianza, cinco por personas que no pueden. Ver si se puede adivinar cuál es cuál.

Dios	pagará
promesa	graduado
Libre de deudas	gracias

pago mínimo                      despues del impuesto  
tasa de interés más baja    hospital

Se podría pensar, o al menos de esperanza de que una persona educada, abiertamente religioso que da su palabra estaría entre los más propensos a pagar un préstamo. Pero, de hecho, este no es el caso. Este tipo de persona, los datos muestran, es menos probable que el promedio para cumplir con su deuda.

Estas son las frases agrupadas por la probabilidad de pagar.

### **TÉRMINOS UTILIZADOS EN APLICACIONES DE PRÉSTAMOS POR personas con más probabilidades de pagar**

---

Libre de deudas	pago mínimo
tasa de interés más baja	graduado
despues del impuesto	

### **TÉRMINOS UTILIZADOS EN APLICACIONES DE PRÉSTAMOS POR PERSONAS MÁS probable que por defecto**

---

Dios	gracias
promesa	hospital
pagará	

Antes de discutir las implicaciones éticas de este estudio, vamos a pensar a través, con la ayuda de los autores del estudio, lo que revela sobre las personas. ¿Qué debemos hacer de las palabras en las diferentes categorías?

En primer lugar, vamos a considerar el lenguaje que sugiere que alguien es más probable que hagan sus pagos de préstamos. Frases como “tasa de interés más baja” o “impuesto después de la” indican un cierto nivel de sofisticación financiera por parte del prestatario, así que no es quizás sorprendente que se correlacionan con alguien más propensos a pagar su préstamo. Además, si él o ella habla de los logros positivos, como ser un “graduado” de la universidad y de ser “libre de deudas”, él o ella también es probable que pagar sus préstamos.

Ahora vamos a considerar un lenguaje que sugiere que alguien es poco probable que pagar sus préstamos. En general, si alguien le dice que le devolverá el dinero, no va a devolverle el dinero. El más firme la promesa, es más probable que se rompa. Si alguien escribe “Prometo que pagar, por lo que la ayuda de Dios,” él está entre los menos propensos a reembolso. Apelando a su merced, explicando que él necesita el dinero porque tiene un pariente en el “hospital” -también significa que es poco probable que le reembolse. De hecho, mencionando cualquier miembro de una familia, esposo, esposa, hijo, hija, madre, o padre, es una señal que alguien no se paga de vuelta. Otra palabra que indica por defecto es “explicar”, es decir, si las personas están tratando de explicar por qué van a ser capaces de pagar un préstamo, es probable que no lo hará.

Los autores no tienen una teoría de por qué agradecer a la gente es una prueba de probabilidad por defecto.

En resumen, de acuerdo con estos investigadores, dando un plan detallado de cómo puede hacer sus pagos y la mención de los compromisos que ha mantenido en el pasado son evidencia alguien va a pagar un préstamo. Hacer promesas y apelando a su merced es una clara señal de que alguien va a entrar en default. Independientemente de las razones, o lo que nos dice sobre la naturaleza humana que hacer promesas es un signo seguro de que alguien, en realidad, no hacer algo, los investigadores encontraron la prueba fue una pieza muy valiosa de información para predecir el valor predeterminado. Alguien que menciona a Dios era 2,2 veces más probabilidad de mora. Este fue uno de los indicadores más altos individuales que alguien no pagaría.

Sin embargo, los autores también creen que su estudio plantea cuestiones éticas. Mientras que esto era sólo un estudio académico, algunas empresas informan que utilizan datos en línea en la aprobación de préstamos. Es esto aceptable? ¿Queremos vivir en un mundo en el que las empresas utilizan las palabras que escribimos para predecir si vamos a pagar un préstamo? Es, como mínimo, espeluznante y, muy posiblemente, de miedo.

Un consumidor que busca un préstamo en un futuro próximo podría tener que preocuparse no sólo su historial financiero, sino también su actividad en línea. Y ella puede ser juzgado por factores que parecen absurdas, ya sea que se utiliza la frase “Gracias” o invoca a “Dios”, por ejemplo. Además, ¿qué pasa con una mujer que legítimamente necesita para ayudar a su hermana en un hospital y con toda seguridad, pagar su préstamo después? Parece horrible para castigarla porque, en promedio, las personas que afirman que necesite ayuda para gastos médicos a menudo se han demostrado estar mintiendo. Un mundo que funciona de esta manera empieza a verse muy distópico.

Esta es la cuestión ética: tienen las corporaciones tienen el derecho de juzgar nuestra idoneidad para sus servicios sobre la base de criterios abstractos, pero estadísticamente predictivos no directamente relacionados con dichos servicios?

Dejando atrás el mundo de las finanzas, vamos a ver las implicaciones más grandes en, por ejemplo, las prácticas de contratación. Los empleadores están recorriendo cada vez más las redes sociales al considerar candidatos de trabajo. Eso no puede plantear cuestiones éticas si están en busca de evidencia de hablar mal de los empleadores anteriores o revelar secretos anteriores empleadores. Incluso puede haber alguna justificación para negarse a contratar a alguien cuya Facebook o Instagram mensajes sugerir el uso excesivo de alcohol. Pero lo que si encuentran un indicador aparentemente inofensivo que se correlaciona con algo que les importa?

Investigadores de la Universidad de Cambridge y Microsoft dieron cincuenta y ocho mil usuarios de Facebook en Estados Unidos una serie de pruebas sobre su personalidad e inteligencia. Ellos encontraron que los gustos de Facebook con frecuencia se correlacionan con el CI, extroversión y la escrupulosidad. Por ejemplo, las personas que les gusta Mozart, tormentas eléctricas y papas fritas en Facebook tienden a tener un CI más alto. Los que les gusta motocicletas Harley-Davidson, el grupo de música country Lady Antebellum, o en la página “Amo el ser una mamá” tienden a tener menor coeficiente intelectual. Algunas de estas correlaciones puede ser debido a la maldición de la dimensionalidad. Si prueba suficientes cosas,

algunos se correlacionan de forma aleatoria. Sin embargo, algunos intereses pueden correlacionar legítimamente con el CI.

No obstante, parecería injusto que una persona inteligente que resulta gustaría Harleys no podía conseguir un trabajo acorde con sus habilidades porque él era, sin darse cuenta, lo que indica una baja inteligencia.

Para ser justos, esto no es un problema totalmente nuevo. Las personas siempre han sido juzgados por factores que no están directamente relacionados con el desempeño del trabajo, la firmeza de sus apretones de manos, la pulcritud de su vestido. Pero el peligro de la revolución de datos es que, como se cuantifica más de nuestra vida, estos juicios de proxy puede obtener más esotérica aún más intrusivo. Una mejor predicción puede dar lugar a una discriminación más sutil y más nefasto.

Mejores datos también pueden dar lugar a otro tipo de discriminación, lo que los economistas llaman la discriminación de precios. Las empresas con frecuencia están tratando de averiguar qué precio que deberían cobrar por bienes o servicios. Lo ideal es que quieren cobrar a los clientes el máximo que están dispuestos a pagar. De esta manera, se va a extraer el máximo beneficio posible.

La mayoría de las empresas por lo general terminan recogiendo un precio que todos paguen. Pero a veces son conscientes de que los miembros de un grupo determinado, en promedio, pagar más. Esta es la razón por cines cobran más a los clientes de mediana edad-a la altura de su poder adquisitivo, que a estudiantes o personas mayores y por qué las compañías aéreas a menudo cobran más compradores de última hora. Ellos discriminación de precios.

Big Data puede permitir a las empresas para obtener sustancialmente mejor por el aprendizaje de lo que los clientes están dispuestos a pagar, y por lo tanto excavando ciertos grupos de personas. Las decisiones óptima Grupo fue pionero en el uso de la ciencia de datos para predecir la cantidad de consumidores están dispuestos a pagar por el seguro. ¿Cómo lo hicieron? Utilizaron una metodología que hemos discutido previamente en este libro. Encontraron clientes anteriores más similares a los que actualmente buscan comprar un seguro, y vieron lo alto de una prima que estaban dispuestos a asumir. En otras palabras, corrieron una búsqueda doppelganger. Una búsqueda doppelganger es entretenido si nos ayuda a predecir si un jugador de béisbol volverá a su antigua grandeza. Una búsqueda doppelganger es grande si nos ayuda a curar la enfermedad de alguien. Pero si una búsqueda doppelganger ayuda a una corporación extraer hasta el último centavo de usted? Eso no es tan cool.

El juego es un área en la que la capacidad de hacer zoom sobre los clientes es potencialmente peligroso. Grandes casinos están utilizando algo así como una búsqueda doppelganger para entender mejor sus consumidores. ¿Su meta? Para extraer el máximo beneficio posible, para asegurarse de que más de su dinero que entra en sus arcas.

Así es como funciona. Cada jugador, casinos creen, tiene un “punto de dolor.” Esta es la cantidad de pérdidas que suficientemente asustarán ella para que ella sale de su casino durante un período prolongado de tiempo. Supongamos, por ejemplo, que “punto débil” de Helen es de \$ 3,000. Esto significa que si pierde \$ 3.000, que ha perdido un cliente, tal vez por semanas o

meses. Si Helen pierde \$ 2,999, no va a ser feliz. ¿Quién, después de todo, le gusta perder dinero? Pero no será tan desmoralizado que no va a volver mañana por la noche.

Imagine por un momento que está administrando un casino. E imaginar que Helen ha mostrado para jugar a las máquinas tragaperras. ¿Cuál es el resultado óptimo? Es evidente que usted quiere Helen llegar lo más cerca posible a su “punto débil”, sin cruzarlo. ¿Quieres que pierda \$ 2,999, suficiente que hacer grandes beneficios, pero no tanto que no va a volver a jugar de nuevo pronto.

¿Cómo puedes hacer esto? Bueno, hay maneras de conseguir Helen para detener la reproducción una vez que se ha perdido una cierta cantidad. Usted puede ofrecer sus comidas gratis, por ejemplo. Hacer la oferta tentadora suficiente, y que dejará a las ranuras de la comida.

Pero hay un gran reto con este enfoque. ¿Cómo sabes “punto débil” de Helen? El problema es que las personas tienen diferentes puntos de dolor “” Para Helen, que es \$ 3.000. Para Juan, que podría ser de \$ 2.000. Para Ben, podría ser \$ 26.000. Si se convence a Helen de dejar de jugar cuando perdió \$ 2.000, lo dejó en la mesa de ganancias. Si espera demasiado largo después de que ella ha perdido \$ 3.000 que la ha perdido por un tiempo. Además, Helen no pudo querer decir que su punto de dolor. Ni siquiera se puede saber lo que es ella misma.

Entonces, ¿Qué haces? Si ha llegado hasta aquí en el libro, es probable que pueda adivinar la respuesta. Usted utiliza la ciencia de datos. A aprender todo lo que pueda acerca de un número de sus clientes-su edad, sexo, código postal, y la conducta de juego. Y, a partir de que el juego de comportamiento: sus ganancias, losings, idas y venidas, se calculan-su “punto de dolor.”

A reunir toda la información que sabe acerca de Helen y encontrar jugadores que son similares a sus-sus doppelgangers, más o menos. A continuación, calcular la cantidad de dolor que pueden soportar. Es probablemente la misma cantidad que Helen. De hecho, esto es lo que el casino Harrah hace, la utilización de una firma de almacén de datos grandes, terabyte, para ayudarlos.

Scott Gnau, gerente general de terabyte, explica, en el excelente libro de Super Crunchers, lo que hacen los gerentes del casino cuando ven a un cliente a punto de su punto de dolor: “Ellos salen y dicen: 'Veo que estás teniendo un mal día. Sé que te gusta nuestro asador. Aquí, me gustaría que usted tome su esposa a cenar en nosotros en este momento '.”

Esto podría parecer la altura de la generosidad: una cena gratis de carne. Pero en realidad es su propio beneficio. El casino está tratando de conseguir que los clientes a dejar de fumar antes de que pierdan tanto que van a dejar por un período prolongado de tiempo. En otras palabras, la administración está usando el análisis de datos sofisticados para tratar de extraer la mayor cantidad de dinero de los clientes, a largo plazo, ya que puede.

Tenemos derecho a temer que mejor y un mejor uso de los datos en línea le dará a los casinos, compañías de seguros, prestamistas y otras entidades corporativas demasiado poder sobre nosotros.

Por otra parte, los grandes datos también ha sido permitir a los consumidores a conseguir algunos golpes contra las empresas que los sobrefacturen o entregan productos de mala calidad.

Un arma importante es, sitios como Yelp, que publican reseñas de restaurantes y otros servicios. Un estudio reciente realizado por el economista Michael de Luca, de la Universidad de

Harvard, ha mostrado el grado en que las empresas están a merced de los comentarios de Yelp. Comparando estas revisiones con datos de ventas en el estado de Washington, se encontró con que uno menos estrellas en Yelp hará que los ingresos de un restaurante caída del 5 al 9 por ciento.

Los consumidores también son ayudados en sus luchas con el negocio de los sitios de comparación de compras-como kayak y Booking.com. Como se discutió en Freakonomics, cuando un sitio de Internet comenzó a informar los precios de diferentes compañías estaban cobrando por el seguro de vida, estos precios cayeron dramáticamente. Si una compañía de seguros fue cobrando de más, los clientes lo saben y el uso de otra persona. Los ahorros totales a los consumidores? Mil millones de dólares por año.

Los datos sobre el Internet, en otras palabras, puede decir, que las empresas clientes a evitar y que pueden explotar. También se puede decir a los clientes las empresas que deben evitar y que está tratando de explotarlos. Grandes volúmenes de datos hasta la fecha ha ayudado a ambos lados en la lucha entre los consumidores y las empresas. Hay que asegurarse de que se mantiene una lucha justa.

## **EL PELIGRO DE GOBIERNOS AUTORIZADAS**

---

Cuando su ex novio se presentó en una fiesta de cumpleaños, Adriana Donato sabía que estaba molesto. Ella sabía que estaba loco. Ella sabía que había luchado contra la depresión. A medida que la invitó a dar una vuelta, había una cosa Donato, un estudiante de la zoología de veinte años de edad, no sabía. Ella no sabía que su ex novio, de veintidós años de edad, James Stoneham, había pasado las últimas tres semanas en busca de información sobre cómo matar a alguien y sobre la ley de asesinato, mezclado con la búsqueda de vez en cuando sobre Donato.

Si se hubiera sabido esto, probablemente no habría conseguido en el coche. Es de suponer que no habría sido apuñalado hasta la muerte esa noche.

En la película Minority Report, psíquicos colaboran con los departamentos de policía para detener los crímenes antes de que sucedan. Debería ponerse a disposición de datos grandes a los departamentos de policía para detener los crímenes antes de que sucedan? Donato debería al menos haber sido advertido acerca de las búsquedas presentimiento de su ex-novio? En caso de que la policía ha interrogado Stoneham?

En primer lugar, hay que reconocer que existe una creciente evidencia de que las búsquedas de Google relacionados con la actividad delictiva se correlacionan con la actividad delictiva. Christine Ma-Kellams, la flora o, Ji Hyun Baek, e Ichiro Kawachi han demostrado que las búsquedas de Google relacionadas con el suicidio se correlacionan fuertemente con la tasa de suicidios a nivel estatal. Además, Evan SOLTAS y me han demostrado que las búsquedas de tales islamófobos semanales como “Odio los musulmanes” o “matar a los musulmanes” - correlate con los crímenes de odio contra los musulmanes que semana. Si más personas están haciendo búsquedas diciendo que quieren hacer algo, más gente va a hacer esa cosa.

Entonces, ¿qué debemos hacer con esta información? Uno, simple idea bastante controvertida: podemos utilizar los datos a nivel de zona para asignar recursos. Si una ciudad tiene un enorme aumento en las búsquedas relacionadas con el suicidio, que puede hasta la conciencia del suicidio en esta ciudad. El gobierno de la ciudad o sin fines de lucro pueden ejecutar anuncios que explican donde la gente puede obtener ayuda, por ejemplo. Del mismo modo, si una ciudad tiene un enorme aumento en las búsquedas de “matar a los musulmanes”, departamentos de policía podrían ser aconsejable cambiar la forma en que patrullan las calles. Podrían enviar más agentes para proteger la mezquita local, por ejemplo.

Sin embargo, un paso que debemos ser muy reacios a tomar: ir tras los individuos antes de cualquier crimen ha sido cometido. Esto parece, para empezar, una invasión de la privacidad. Hay un gran salto ético del gobierno que tiene los datos de búsqueda de miles o cientos de miles de personas para el departamento de policía que tiene los datos de búsqueda de un individuo. Hay un gran salto ético de proteger a una mezquita local a saquear la casa de alguien. Hay un gran salto ético de la publicidad prevención del suicidio de bloquear a alguien en un hospital psiquiátrico en contra de su voluntad.

La razón de ser extremadamente cauteloso usando datos a nivel individual, sin embargo, va más allá de la ética pares. Hay una razón de datos también. Es un salto grande para la ciencia de datos para ir de tratar de predecir las acciones de una ciudad para tratar de predecir las acciones de un individuo.

Volvamos al suicidio por un momento. Todos los meses, hay cerca de 3,5 millones de búsquedas de Google en los Estados Unidos relacionadas con el suicidio, y la mayoría de ellos lo que sugiere ideas suicidas e-búsquedas como “suicida”, “cometer suicidio”, y “la forma de suicidio.” En otras palabras, todos los meses, hay más de una búsqueda relacionada con el suicidio por cada cien estadounidenses. Esto trae a la mente una cita del filósofo Friedrich Nietzsche: “La idea de que el suicidio es un gran consuelo: Por medio de ella se obtiene a través de más de una noche oscura” de datos de búsqueda de Google muestra cómo es verdad, es decir, qué tan común es la idea del suicidio es. Sin embargo, todos los meses, hay menos de cuatro mil suicidios en los Estados Unidos. La ideación suicida es muy común. El suicidio no es.

O considerar esas búsquedas islamóforos increíblemente vicioso. En 2015, había aproximadamente 12.000 búsquedas en los Estados Unidos para “matar a los musulmanes.” Hubo 12 asesinatos de musulmanes reportados como crímenes de odio. Es evidente que la gran mayoría de las personas que hacen esta búsqueda aterrador no seguir adelante con el acto correspondiente.

Hay un poco de matemática que explica la diferencia entre la predicción del comportamiento de una persona y predecir el comportamiento de una ciudad. Aquí es un simple experimento mental. Supongamos que hay un millón de personas en una ciudad y una mezquita. Supongamos que, si alguien no busca “matar a los musulmanes,” sólo hay una probabilidad de 1 en 100 millones que va a atacar una mezquita. Supongamos que si alguien hace la búsqueda para “matar a los musulmanes”, esta oportunidad se eleva abruptamente, a 1 de cada 10.000. Supongamos

que la islamofobia se ha disparado y las búsquedas para “matar” a los musulmanes han aumentado de 100 a 1.000.

En esta situación, matemáticas muestra que las posibilidades de una mezquita ser atacado ha subido alrededor de cinco veces, de alrededor del 2 por ciento al 10 por ciento. Pero las posibilidades de una persona que buscaba “matar a los musulmanes” en realidad atacar una mezquita sigue siendo sólo 1 de cada 10.000.

La respuesta adecuada en esta situación no es la cárcel a todas las personas que han buscado “matar a los musulmanes.” Tampoco es para visitar a sus casas. Hay una pequeña posibilidad de que cualquiera de estas personas en particular, incurrirá en delito. La respuesta adecuada, sin embargo, sería la de proteger esa mezquita, que ahora tiene una probabilidad del 10 por ciento de ser atacados.

Está claro que muchas búsquedas terribles nunca se llevan a acciones horribles.

Dicho esto, es al menos teóricamente posible que hay algunas clases de búsquedas que sugieren una probabilidad razonablemente alta de un terrible seguimiento. Es por lo menos teóricamente posible, por ejemplo, que los científicos de datos podrían en un futuro construir un modelo que podría haber encontrado que las búsquedas de Stoneham relacionados con Donato eran causa importante de preocupación.

En 2014, había alrededor de 6000 busca la frase exacta “cómo matar a su novia” y 400 asesinatos de novias. Si todos estos asesinos habían hecho esta búsqueda exacta de antemano, que significaría 1 de cada 15 personas que han buscado “cómo matar a su novia” fue a través de la misma. Por supuesto, muchos, probablemente la mayoría, las personas que asesinaron a sus amigas no hacen esta búsqueda exacta. Esto significaría la verdadera probabilidad de que esta búsqueda en particular condujo al asesinato es menor, probablemente mucho menor.

Pero si los científicos de datos podrían construir un modelo que demostró que la amenaza contra un individuo en particular era, digamos, 1 en 100, lo que se quiere hacer algo con esa información. Al menos, la persona bajo la amenaza podría tener el derecho a ser informado de que hay una probabilidad de 1 en 100, ella será asesinada por una persona en particular.

En general, sin embargo, tenemos que ser muy cautelosos a partir de datos de búsqueda de predecir crímenes a nivel individual. Los datos nos dicen claramente que hay muchas, muchas búsquedas terribles que rara vez conducen a acciones horribles. Y no ha sido, hasta el momento, ninguna prueba de que el gobierno puede predecir una acción horrible, en particular, con alta probabilidad, sólo de examinar estas búsquedas. Así que tenemos que ser muy cautelosos en cuanto a permitir que el gobierno intervenga en el nivel individual sobre la base de datos de búsqueda. Esto no es sólo por razones éticas o legales. Es también, al menos por ahora, por razones científicas datos.



## CONCLUSIÓN

---

### ¿Cuántas personas TERMINAR libros?

**UNA**espués de firmar mi contrato para un libro, que tenía una visión clara de cómo el libro debe ser estructurado. Cerca de la salida, se recordará, he descrito una escena en la mesa de Acción de Gracias de mi familia. Los miembros de mi familia debatieron mi cordura y trataron de averiguar por qué, a los treinta y tres, podría no parecen encontrar a la chica adecuada.

La conclusión de este libro, por tanto, prácticamente se escribió sola. Me gustaría conocer y casarse con la chica. Mejor aún, me gustaría utilizar grandes volúmenes de datos para cumplir con la chica adecuada. Tal vez podría tejer en curiosidades del proceso de cortejo largo. A continuación, la historia sería todos se unen en la conclusión, que se describirían día de mi boda y el doble como una carta de amor a mi nueva esposa.

Por desgracia, la vida no ha coincidido con mi visión. Encerrarme en mi piso y evitando el mundo al escribir un libro, probablemente no ayudó a mi vida romántica. Y, por desgracia, todavía tengo que encontrar una esposa. Más importante, que necesitaba una nueva conclusión.

Me estudió minuciosamente muchas de mis libros favoritos en la búsqueda de lo que hace un gran conclusión. Los mejores conclusiones, llegué a la conclusión, traer a la superficie un punto importante que ha estado allí todo el tiempo, flotando justo por debajo de la superficie. Para este libro, que gran punto es la siguiente: las ciencias sociales se está convirtiendo en una verdadera ciencia. Y esta nueva, la ciencia real está a punto de mejorar nuestras vidas.

En el comienzo de la segunda parte, discutí crítica de Freud de Karl Popper. Popper, he señalado, no creía que la visión de Freud loco del mundo era científica. Pero no he mencionado algo acerca de la crítica de Popper. En realidad, fue mucho más amplio que un simple ataque a Freud. Popper no creía que cualquier científico social fue particularmente científica. Popper era simplemente impresionado con el rigor de lo que estos supuestos científicos estaban haciendo.

Lo que motivó a la cruzada de Popper? Cuando se interactuó con los mejores intelectuales de su época: los mejores físicos, los mejores historiadores, los mejores psicólogos-Popper observó una notable diferencia. Cuando los físicos hablaron, Popper creía en lo que estaban haciendo. Claro, que a veces cometen errores. Claro, que en ocasiones fueron engañados por sus prejuicios subconscientes. Pero los físicos estaban comprometidos en un proceso que se encontraba

claramente verdades profundas sobre el mundo, culminando en la Teoría de la Relatividad de Einstein. Cuando los científicos sociales más famosas del mundo hablaron, por el contrario, Popper pensó que estaba escuchando un montón de jerigonza.

Popper no es la única persona que ha hecho esta distinción. Casi todo el mundo está de acuerdo en que los físicos, biólogos y químicos son verdaderos científicos. Ellos utilizan experimentos rigurosos para encontrar cómo funciona el mundo físico. Por el contrario, muchas personas piensan que economistas, sociólogos, psicólogos y científicos son suaves que tiran alrededor de la jerga sin sentido para que puedan obtener la tenencia.

En la medida en que esto era cierto alguna vez, la revolución de grandes volúmenes de datos ha cambiado. Si Karl Popper estuviera vivo hoy y asistió a una presentación de Raj Chetty, Jesse Shapiro, Esther Duflo, o (sígueme) a mí mismo, tengo la fuerte sospecha que no tendría la misma reacción que tenía en aquel entonces. Para ser honesto, podría ser más probable que la pregunta de si los grandes teóricos de cuerdas de hoy son verdaderamente científica o simplemente participar en gimnasia mental auto-indulgente.

Si una película violenta trata de una ciudad, ¿vaya crimen arriba o hacia abajo? Si más personas están expuestas a un anuncio, ¿más personas utilizan el producto? Si un equipo de béisbol gana cuando un niño es de veinte, ¿va a ser más propenso a la raíz para ellos cuando tiene cuarenta años? Todas estas son preguntas claras con claras sí o no-respuestas. Y en las montañas de datos honestos, podemos encontrarlos.

Esta es la materia de la ciencia, no pseudociencia.

Esto no significa que la revolución de las ciencias sociales vendrá en forma de leyes simples y atemporales.

Marvin Minsky, el fallecido científico del MIT y uno de los primeros en estudiar la posibilidad de la inteligencia artificial, sugirió que la psicología se bajó de la pista al tratar de copiar la física. La física tuvo éxito la búsqueda de leyes simples que llevan a cabo en todos los tiempos y todos los lugares.

Los cerebros humanos, sugirió Minsky, no pueden ser objeto de tales leyes. El cerebro, en cambio, es probable que un sistema complejo de hacks-una parte la corrección de errores en otras partes. La economía y el sistema político pueden ser igualmente compleja.

Por esta razón, la revolución de las ciencias sociales es poco probable que venir en forma de fórmulas ordenadas, tales como  $E = MC^2$ . De hecho, si alguien reclama una revolución de las ciencias sociales sobre la base de una fórmula ordenada, debe ser escéptico.

La revolución, en cambio, vendrá poco a poco, estudio realizado por el estudio, la búsqueda mediante la búsqueda. Poco a poco, vamos a tener una mejor comprensión de los sistemas complejos de la mente humana y la sociedad.

Una conclusión correcta resume, pero también señala el camino a más cosas por venir.

Para este libro, eso es fácil. Los conjuntos de datos que he discutido en este documento son revolucionarios, pero apenas han sido exploradas. Hay mucho más que aprender. Francamente, la inmensa mayoría de los académicos han ignorado la explosión de datos causada por la era

digital. investigadores del sexo más famosos del mundo se pega con el probado y verdadero. Se preguntan unos pocos cientos de temas acerca de sus deseos; no preguntan sitios como Pornhub por sus datos. lingüistas más famosos del mundo analizan textos individuales; ignoran en gran medida los patrones revelados en mil millones de libros. Las metodologías que se enseñan a los estudiantes graduados en psicología, ciencias políticas y sociología han sido, en su mayor parte, al margen de la revolución digital. El amplio terreno, en su mayoría sin explorar abierta por la explosión de datos se ha dejado a un pequeño número de profesores con visión de futuro,

Eso va a cambiar.

Por cada idea que he hablado en este libro, hay un centenar de ideas tan importante listo para ser abordado. La investigación discutida aquí es la punta de la punta del iceberg, un rasguño en el rayado de la superficie.

Entonces, ¿qué viene?

Por un lado, una expansión radical de la metodología que se utilizó en uno de los estudios de salud pública más exitosas de todos los tiempos. A mediados del siglo XIX, John Snow, un médico británico, estaba interesado en lo que estaba causando un brote de cólera en Londres.

Su ingeniosa idea: mapea todos los casos de cólera en la ciudad. Cuando lo hizo, se encontró que la enfermedad se agrupan en gran medida en torno a una bomba de agua en particular. Esto sugiere que la enfermedad se extendió a través del agua infectadas con gérmenes, desmintiendo la idea convencional de que entonces se extendió por el mal aire.

Datos y grande es el zoom que permite-hace que este tipo de estudio fácil. Para cualquier enfermedad, podemos explorar los datos de búsqueda de Google u otros datos de salud digital. Podemos encontrar si hay algunas pequeñas bolsas del mundo donde la prevalencia de esta enfermedad es inusualmente alta o inusualmente baja. A continuación, podemos ver lo que estos lugares tienen en común. ¿Hay algo en el aire? ¿El agua? Las normas sociales?

Podemos hacer esto para las migrañas. Podemos hacer esto para los cálculos renales. Podemos hacer esto para la ansiedad y la depresión y la enfermedad de Alzheimer y el cáncer de páncreas y de la presión arterial alta y dolor de espalda y el estreñimiento y hemorragias nasales. Podemos hacer esto para todo. El análisis que hizo una vez la nieve, que podría ser capaz de hacer cuatrocientas veces (algo partir de este escrito ya estoy empezando a trabajar en).

Podríamos llamar a esta toma de un método simple y la utilización de grandes volúmenes de datos para llevar a cabo un análisis de varios cientos de veces en un corto período de tiempo en la ciencia a escala. Sí, las ciencias sociales y del comportamiento son más definitivamente van a escala. Zoom sobre las condiciones de salud ayudará a estas ciencias escala. Otra cosa que les ayudará a escala: las pruebas A / B. Hablamos de las pruebas A / B en el contexto de las empresas que consiguen los usuarios a hacer clic en los titulares y anuncios, y esto ha sido el uso predominante de la metodología. Sin embargo, las pruebas A / B puede ser utilizado para descubrir las cosas más fundamentales-y socialmente valioso que una flecha que hace que la gente haga clic en un anuncio.

Benjamin F. Jones es un economista de la Northwestern que está tratando de utilizar las pruebas A / B para ayudar mejor a los niños a aprender. Él ha ayudado a crear una plataforma,

EDU STAR, lo que permite a las escuelas para poner a prueba al azar diferentes planes de lecciones.

Muchas empresas están en el negocio del software educativo. Con EDU STAR, los estudiantes inician sesión en un ordenador y están expuestos al azar a diferentes planes de lecciones. Entonces toman pruebas cortas para ver lo bien que han aprendido el material. Escuelas, en otras palabras, aprender lo que funciona mejor software para ayudar a los estudiantes a entender el material.

Ya, como todas las grandes plataformas de pruebas A / B, EDU STAR está dando resultados sorprendentes. Un plan de lección que muchos educadores estaban muy emocionados con el software incluido que utiliza juegos para ayudar a enseñar fracciones estudiantes. Ciertamente, si se convirtió en un juego de matemáticas, los estudiantes tendrían más divertido, aprender más y mejor en los exámenes. ¿Derecha? Incorrecto. Los estudiantes que se les enseñó fracciones mediante un juego probaron peor que aquellos que aprendieron fracciones de una manera más estándar.

Hacer que los niños a aprender más es un emocionante y socialmente beneficioso, el uso de la prueba que fue pionera en Silicon Valley para que la gente haga clic en más anuncios. También lo es hacer que la gente a dormir más.

El estadounidense promedio recibe 6,7 horas de sueño cada noche. La mayoría de los estadounidenses quieren dormir más. Pero alrededor de los rollos 23:00, y está en SportsCenter o YouTube está llamando. Por lo que espera de ojos cerrados. Quijada, una empresa Wearable-dispositivo con cientos de miles de clientes, realiza miles de pruebas para tratar de encontrar intervenciones que ayudan a conseguir sus usuarios que hagan lo que ellos quieren hacer: ir a la cama más temprano.

Quijada anotó una gran victoria mediante un doble objetivo. En primer lugar, pedir a los clientes a comprometerse a un objetivo no-que-ambiciosa. Envíele un mensaje como este: “Parece que usted no ha estado durmiendo mucho en los últimos 3 días. ¿Por qué no te apuntas a llegar a la cama a las 11:30 de esta noche? Sabemos que normalmente se levanta a las 8 AM” A continuación, los usuarios tendrán la opción de hacer clic en ‘Estoy en el’.

En segundo lugar, cuando llega 10:30, quijada enviará otro mensaje: “Decidimos que te apuntas a dormir a las 11:30. Son las 10:30 ahora. ¿Por qué no empezar ahora?”

Quijada encontraron esta estrategia condujo a veintitrés minutos de sueño adicional. No consiguieron los clientes para conseguir realmente a la cama a las 10:30, pero lo hicieron llegar a la cama más temprano.

Por supuesto, cada parte de esta estrategia tuvo que ser optimizado a través de un montón de experimentación. Iniciar el objetivo original demasiado pronto-pedir a los usuarios que se comprometan a ir a la cama a las 23:00, y pocos serán el juego. Pedir a los usuarios a ir a la cama a media noche y poco se ganará.

Quijada utiliza pruebas A / B para encontrar el equivalente del sueño de Google de la flecha apunta hacia la derecha. Pero en lugar de unos pocos clics más para los socios de anuncios de Google, que produce unos minutos más de descanso para los estadounidenses agotados.

De hecho, todo el campo de la psicología podría utilizar las herramientas de Silicon Valley para mejorar dramáticamente su investigación. Estoy esperando con impaciencia el primer documento de la psicología que, en lugar de detallar un par de experimentos realizados con unos pocos estudiantes de licenciatura, muestra los resultados de un millar de pruebas rápidas A / B.

Los días de los académicos que dedican meses a la contratación de un pequeño número de estudiantes de licenciatura para realizar una sola prueba llegará a su fin. En cambio, los académicos utilizarán datos digitales para probar unos pocos cientos o unos pocos miles de ideas en tan sólo unos segundos. Vamos a ser capaces de aprender mucho más en un tiempo mucho menor.

Texto como datos nos va a enseñar mucho más. ¿Cómo se propagan las ideas? ¿Cómo se forman nuevas palabras? ¿Cómo desaparecen las palabras? ¿Cómo se forman los chistes? ¿Por qué son ciertas palabras divertidos y otros no? ¿Cómo se desarrollan dialectos? Apuesto, dentro de los veinte años, vamos a tener profundos conocimientos sobre todas estas cuestiones.

Creo que podríamos considerar la utilización de niños en línea conducta apropiada anónima, como un complemento a las pruebas tradicionales para ver cómo están aprendiendo y desarrollando. ¿Cómo es su ortografía? ¿Se observan signos de dislexia? Están en desarrollo, intereses intelectuales maduros? ¿Tienen amigos? Hay pistas para todas estas preguntas en los miles de pulsaciones de teclas a cada niño hace todos los días.

Y hay otro, la zona no es trivial, donde están llegando muchos más puntos de vista.

En la canción “Shattered”, de los Rolling Stones, Mick Jagger describe todo lo que hace la ciudad de Nueva York, la gran manzana, tan mágico. Risa. Alegría. Soledad. Ratas. Chinchas. Orgullo. Codicia. Personas vestidas con bolsas de papel. Pero Jagger dedica la mayor cantidad de palabras para lo que hace que la ciudad realmente especial: “El sexo y el sexo y el sexo y el sexo.”

Al igual que con la gran manzana, así que con grandes volúmenes de datos. Gracias a la revolución digital, ideas están llegando en la salud. Dormir. Aprendizaje. Psicología. Idioma. Además, el sexo y el sexo y el sexo y el sexo.

Una pregunta que estoy explorando actualmente: cuántas dimensiones de la sexualidad hay? Por lo general pensamos en alguien como homosexuales o heterosexuales. Pero la sexualidad es claramente más complejo que eso. Entre las personas homosexuales y heterosexuales, las personas tienen tipos y tantos hombres como “rubias”, otros “morenas”, por ejemplo. Estas preferencias pueden ser tan fuertes como las preferencias de género? Otra pregunta Estoy buscando en: ¿de dónde preferencias sexuales vienen? Así como podemos averiguar los años clave que determinan el fandom de béisbol o puntos de vista políticos, ahora podemos encontrar los años clave que determinan las preferencias sexuales de adultos. Para obtener estas respuestas, tendrá que comprar mi próximo libro, Todo el mundo tentativamente titulado (fijas) Mentiras.

La existencia de pornografía y los datos que viene con él, es un desarrollo revolucionario en la ciencia de la sexualidad humana.

Se tomó el tiempo para las ciencias naturales para comenzar a cambiar nuestra vida, para crear la penicilina, satélites y computadoras. Puede tomar algún tiempo antes de grandes volúmenes de

datos lleva las ciencias sociales y del comportamiento de los avances importantes en la forma en que nos gusta, aprender y vivir. Pero creo que estos avances están llegando. Espero que vea al menos los contornos de esa evolución de este libro. Espero que, de hecho, que algunos de los que están leyendo este libro ayuda a crear tales avances.

Para escribir correctamente una conclusión, un autor debe pensar por qué escribió el libro en el primer lugar. ¿Qué objetivo está tratando de lograr?

Creo que la razón más grande que escribí este libro es el resultado de una de las experiencias más formativos de mi vida. Usted ve, un poco más de una década atrás, el libro *Freakonomics* salió. El éxito de ventas sorpresa describió la investigación de Steven Levitt, un galardonado economista de la Universidad de Chicago ha mencionado con frecuencia en este libro. Levitt era un “economista delincuentes” que parecía ser capaz de utilizar los datos para responder a cualquier pregunta su mente peculiar que se le ocurrió preguntar: ¿Los luchadores de sumo engañan? ¿Los concursantes en programas de juegos discriminan? ¿Los agentes inmobiliarios le consiguen las mismas ofertas que reciben por sí mismos?

Yo acababa de salir de la universidad, después de haber especialicé en filosofía, con poca idea de lo que quería hacer con mi vida. Después de leer *Freakonomics*, lo sabía. Yo quería hacer lo que hizo Steven Levitt. Yo quería estudiar minuciosamente a través de montañas de datos para averiguar cómo el mundo realmente funcionó. Le gustaría seguir, decidí, y obtener un Ph.D. en economía.

Tanto ha cambiado en los doce años. Se encontró un par de estudios de Levitt haber errores de codificación. Levitt dijo algunas cosas políticamente incorrectas sobre el calentamiento global. *Freakonomics* ha caído en desuso en los círculos intelectuales.

Pero creo que, algunos errores lado, los años han sido amable con el punto más grande Levitt estaba tratando de hacer. Levitt nos decía que una combinación de la curiosidad, la creatividad y datos podría mejorar dramáticamente nuestra comprensión del mundo. Había historias ocultas en los datos que estaban listos para ser contada y esto se ha dado la razón una y otra vez.

Y espero que este libro podría tener el mismo efecto en los demás que *Freakonomics* tenido en mí. Espero que haya alguna persona joven que lee esto ahora que es un poco confusa en lo que quiere hacer con su vida. Si usted tiene un poco de habilidad estadística, una abundancia de la creatividad y la curiosidad, entrar en el negocio de análisis de datos.

Este libro, de hecho, y si puedo ser tan audaz, puede ser visto como más avanzado, *Freakonomics*. Una diferencia importante entre los estudios analizados en *Freakonomics* y los analizados en este libro es la ambición. En la década de 1990, cuando Levitt se hizo un nombre, no había que muchos datos disponibles. Levitt se enorgullecía de ir tras las preguntas peculiar, donde existían datos. Se ignora en gran medida las grandes preguntas donde no existían los datos. Hoy en día, sin embargo, con tantos datos disponibles en casi todos los temas, tiene sentido para ir tras grandes preguntas profundas, que llegan a la esencia de lo que significa ser un ser humano.

El futuro de análisis de datos es brillante. El siguiente Kinsey, yo sospecho fuertemente, será un científico de datos. El siguiente Foucault será un científico de datos. El siguiente Freud será un científico de datos. El siguiente Marx será un científico de datos. El siguiente Salk podría muy bien ser un científico de datos.

De todos modos, esas eran mis intentos de hacer algunas de las cosas que hace a una conclusión correcta. Pero los grandes conclusiones, que se dieron cuenta, hacer mucho más. Mucho más. Una gran conclusión debe ser irónico. Se debe estar en movimiento. Una gran conclusión debe ser profunda y lúdica. Debe ser profundo, buen humor, y triste. Una gran conclusión debe, en una frase o dos, hacer un punto que resume todo lo que ha venido antes, todo lo que está por venir. Debe hacerlo con un único, novedoso punto un giro. Un gran libro debe terminar con una inteligente, divertida, explosión provocativa.

Ahora puede ser un buen momento para hablar un poco acerca de mi proceso de escritura. No soy un escritor especialmente detallado. Este libro es sólo alrededor de setenta y cinco mil palabras, que es un poco corto para un tema tan rico como éste.

Pero lo que me falta de ancho, me invento en la obsesión. Me pasé cinco meses después, y escribió cuarenta y siete borradores de mi primera columna sexo New York Times, que estaba a dos mil palabras. Algunos capítulos de este libro tuvieron sesenta corrientes de aire. Puedo pasar horas para encontrar la palabra correcta para una oración en una nota al pie.

Viví gran parte de mi pasado año como un ermitaño. Sólo yo y mi equipo. Yo vivía en la zona más moderna de la ciudad de Nueva York y fui a aproximadamente Nunca. Esto es, en mi opinión, mi obra magna, la mejor idea que se tiene en mi vida. Y estaba dispuesto a sacrificar lo que fuera necesario para hacer lo correcto. Yo quería ser capaz de defender todas las palabras de este libro. Mi teléfono está lleno de mensajes de correo electrónico que olvide a responder a, e-Vites que nunca abierto, mensajes Bumble ignoré.\*

Después de trece meses de duro trabajo, que finalmente fue capaz de enviar en un proyecto casi completo. Por una parte, sin embargo, faltaba: la conclusión.

Le expliqué a mi editor, Denise, que podría tomar unos cuantos meses. Le dije a sus seis meses fue mi conjetura más probable. La conclusión es, en mi opinión, la parte más importante del libro. Y yo sólo estaba empezando a aprender lo que hace un gran conclusión. Ni que decir tiene, Denise no estaba contento.

Entonces, un día, un amigo mío me envió por correo electrónico un estudio realizado por Jordan Ellenberg. Ellenberg, un matemático de la Universidad de Wisconsin, tenía curiosidad acerca de cuántas personas en realidad terminan los libros. Pensó en una ingeniosa manera de probar que el uso de grandes volúmenes de datos. Amazon informa cuántas personas citan varias líneas en los libros. Ellenberg cuenta de que podía comparar la frecuencia con comillas se destacaron en el comienzo del libro en comparación con el final del libro. Esto daría una guía aproximada de la propensión de los lectores para que sea hasta el final. Por su medida, más del 90 por ciento de los lectores terminó la novela de Donna Tartt El Jilguero. En contraste, sólo el 7 por ciento lo hizo a través de la obra magna del Premio Nobel de Economía Daniel Kahneman,

Pensar rápido. Menos del 3 por ciento, esta metodología en bruto estimado, llegado a la final del economista Thomas Piketty muy discutido y alabado Capital en el siglo 21. En otras palabras, las personas tienden a no terminar tratados por los economistas.

Uno de los puntos de este libro es que tenemos que seguir el Big Data dondequiera que se encuentre y actuar en consecuencia. Puedo esperar que la mayoría de los lectores van a colgar en cada palabra y tratar de detectar patrones que unen las páginas finales a lo que pasó antes. Sin embargo, no importa lo duro que trabajo en pulir mi prosa, la mayoría de la gente va a leer las primeras cincuenta páginas, conseguir algunos puntos, y seguir adelante con sus vidas.

Por lo tanto, concluyo este libro de la única manera apropiada: siguiendo los datos, lo que la gente realmente hace, no lo que dicen. Voy a tomar una cerveza con unos amigos y dejar de trabajar en este maldito conclusión. Muy pocos de ustedes, grandes volúmenes de datos me dice, todavía está leyendo.



# EXPRESIONES DE GRATITUD

---

Este libro fue un esfuerzo de equipo.

Estas ideas fueron desarrolladas mientras era estudiante en Harvard, un científico de datos de Google, y un escritor para el New York Times.

Hal Varian, con quien trabajé en Google, ha sido una gran influencia en las ideas de este libro. Lo mejor que puedo decir, Hal es perpetuamente de veinte años por delante de su tiempo. Su libro Información de Reglas, escrito con Carl Shapiro, básicamente predice el futuro. Y su papel “Predecir el presente”, con Hyunyoung Choi, comenzó en gran medida la revolución grandes volúmenes de datos en las ciencias sociales que se describen en este libro. También es un maestro increíble y amable, como tantos que han trabajado bajo él puede dar fe. Un clásico movimiento Hal es hacer la mayor parte del trabajo sobre un papel que está coautoría con él y luego insistir en que su nombre va delante de él. combinación de Hal del genio y la generosidad es algo que rara vez me he encontrado.

Mi escritura y las ideas desarrolladas bajo Aaron Retica, que ha sido mi editor para cada sola columna del New York Times. Aaron es un gran pensador. De alguna manera él sabe todo acerca de la música, la historia, los deportes, la política, la sociología, la economía, y sólo Dios sabe qué más. Es el responsable de una gran cantidad de lo que es bueno acerca de las columnas del Times que tienen mi nombre en ellos. Otros jugadores del equipo para estas columnas incluyen a Bill Marsh, cuyos gráficos seguir mi golpe de distancia, Kevin McCarthy, y Gita Daneshjoo. Este libro incluye pasajes de estas columnas, reimpresso con permiso.

Steven Pinker, quien amablemente accedió a escribir el prólogo, ha sido durante mucho tiempo uno de mis héroes. Él ha puesto el listón para un libro sobre la ciencia moderna, una exploración de acoplamiento social de los fundamentos de la naturaleza humana, dando sentido a la mejor investigación de una variedad de disciplinas. Ese bar es uno estaré luchando para llegar a toda mi vida.

Mi tesis, de la que ha crecido este libro, fue escrito bajo mis asesores brillantes y pacientes Alberto Alesina, David Cutler, Ed Glaeser, y Lawrence Katz.

Denise Oswald es un poderoso editor. Si usted quiere saber lo bien que su edición es, comparar este proyecto final de mi primer proyecto, en realidad, no se puede hacer eso porque yo no voy a mostrar cada vez que alguien más embarazosa primer borrador. Agradezco también al resto del equipo de Harper Collins, incluyendo a Michael Barr, Lynn Grady, Lauren Janiec, Shelby Meizlik, y Amber Oliver.

Eric Lupfer, mi agente, vio potencial en este proyecto desde el principio, jugó un papel decisivo en la formación de la propuesta, y ayudó a llevarlo a cabo.

Para disfrutar comprobación de los hechos, agradezco Melvis Acosta.

Otras personas de quienes he aprendido mucho en mi vida profesional y académica incluyen Susan Athey, Shlomo Benartzi, Jason Bordoff, Danielle Bowers, David Broockman, Bo Cowgill, Steven Delpome, John Donohue, Bill Gale, Claudia Goldin, Suzanne Greenberg, Shane Greenstein, Steve Grove, Mike Hoyt, David Laibson, AJ Magnuson, Dana Maloney, Jeffrey Oldham, Peter Orszag, David Reiley, Jonathan Rosenberg, Michael Schwarz, Steve Scott, Rich Shavelson, Michael D. Smith, Lawrence Summers, Jon Vaver, Michael Wiggins y Wu Qing.

Doy gracias a Tim Requarth y NeuWrite para ayudar a desarrollar mi escritura.

Para obtener ayuda en la interpretación de los estudios, agradezco a Christopher Chabris, Raj Chetty, Matt Gentzkow, Salomón Messing, y Jesse Shapiro.

Le pregunté a Emma Pierson y Katia Sobolski si podrían dar consejos sobre un capítulo de mi libro. Se decidió, por razones que no entiendo, para ofrecer a leer todo el libro y dar sabios consejos en cada párrafo.

Mi madre, Esther Davidowitz, leer todo el libro en múltiples ocasiones y ha ayudado a mejorar drásticamente la misma. También me enseñó, por ejemplo, que debería seguir mi curiosidad, no importa dónde llevaba. Cuando estaba entrevistando para un trabajo académico, un profesor me interrogó: “¿Qué piensa tu madre de este trabajo que hace?” La idea era que mi madre podría ser avergonzado de que yo estaba investigando el sexo y otros temas tabú. Pero siempre supe que estaba orgullosa de mí para seguir mi curiosidad, dondequiera que se llevó.

Mucha gente lee secciones y ofrecieron comentarios útiles. Doy gracias a Eduardo Acevedo, Coren Apicella, Sam Asher, David Cutler, Stephen Dubner, Christopher Glazek, Jessica Goldberg, Lauren Goldman, Amanda Gordon, Jacob Leshno, Alex Peysakhovich, Noah Popp, Ramón Roullard, Greg Sobolski, Evan SOLTAS, Noah Stephens- Davidowitz, Lauren Stephens- Davidowitz, y Jean Yang. En realidad, Jean era básicamente mi mejor amigo mientras escribía esto, así que darle las gracias por eso, también.

Para obtener ayuda en la recogida de datos, agradezco a Brett Goldenberg, James Rogers, y Mike Williams en MindGeek y Rob McQuown y Sam Miller en Folleto de béisbol.

Estoy muy agradecido por el apoyo financiero de la Fundación Alfred Sloan.

En un momento dado, al escribir este libro, me sentí profundamente atascado, perdido, y cerca de abandonar el proyecto. Entonces fui al campo con mi padre, Mitchell Stephens. En el transcurso de una semana, papá me puso de nuevo juntos. Me llevó a pasear en el que hemos discutido amor, la muerte, el éxito, la felicidad, y la escritura, y luego se sentó conmigo para que pudiéramos ir sobre cada frase del libro. No podría haber terminado este libro sin él.

Todos los errores restantes son, por supuesto, mi propia.

# NOTAS

---

La paginación de esta edición electrónica no coincide con la edición de la que fue creada. Para localizar una entrada específica, por favor utilice herramientas de búsqueda de su lector de libros electrónicos.

## INTRODUCCIÓN

- 2 votantes estadounidenses en gran medida no les importaba que Barack Obama: Katie Fretland, “Gallup: La raza no importante para los votantes,” La ciénaga, Chicago Tribune, junio de 2008.
- 2 Berkeley estudió minuciosamente a través de: Alexandre Mas y Enrico Moretti, “prejuicio racial en la elección presidencial de 2008,” American Economic Review 99, no. 2 (2009).
- 2 sociedad post-racial: En la edición de noviembre 12 de 2009 episodio de su show, Lou Dobbs dijo que vivimos en una En el número de enero 27 de 2010, episodio de su programa, Chris Matthews “la sociedad post-racial postpartidista.” dijo que el presidente Obama fue Para otros ejemplos, véase Michael C. Dawson y Lawrence D. Bobo, “un año más tarde y el mito de una sociedad post-racial,” Du Bois opinión “post-racial por parte de todas las apariencias.”: Ciencias sociales La investigación sobre Carrera 6, no. 2 (2009).
- 5 analicé los datos de la Encuesta Social General: Los detalles de todos estos cálculos se pueden encontrar en mi sitio web, sethsd.com, en el csv etiquetada datos de la Encuesta Social General se pueden encontrar en [http “Sexo de datos.”: // GSS .norc.org /](http://GSS.norc.umd.edu/).
- 5 menos de 600 millones de condones: Datos suministrados al autor.
- 7 búsquedas y inscripciones para Stormfront: Análisis de autor de datos de Google Trends. También me raspó datos sobre todos los miembros de Stormfront, como se discutió en Seth Stephens-Davidowitz, “los datos del odio”, New York Times, 13 de Julio de 2014, SR4. Los datos necesarios pueden descargarse en sethsd.com, en la sección de datos titulado “Stormfront”.
- 7 Búsquedas más para “presidente negro” de “primer presidente negro”: Análisis de autor de datos de Google Trends. Los estados para los cuales esto es cierto como Kentucky, Louisiana, Arizona y Carolina del Norte.

- 9 rechazados por cinco publicaciones académicas: El papel fue finalmente publicado como Seth Stephens-Davidowitz, “el costo de la animosidad racial en un candidato Negro: Evidencia El uso de Google Búsqueda de datos”, *Journal of Public Economics* 118 (2014). Más detalles acerca de la investigación se pueden encontrar allí. Además, los datos se pueden encontrar en mi página web, [sethsd.com](http://sethsd.com), en la sección de datos titulado “Racismo”.
- 13 solo factor que mejor se correlaciona: “más fuerte correlación que he encontrado para el apoyo Trump es búsquedas de Google para la palabra-n. Otros han informado de esto también”(28 de febrero de 2016, pío). Ver también Nate Cohn, “partidarios más fuertes de Donald Trump: Un nuevo tipo de demócrata,” *New York Times*, 31 de Diciembre de 2015, A3.
- 13 Este es el porcentaje de búsquedas de Google que incluyen la palabra “nigger (s).” Tenga en cuenta que, debido a que la medida es como un porcentaje de las búsquedas de Google, no es arbitraria mayor en lugares con grandes poblaciones o lugares que hacen mucho búsquedas. Tenga en cuenta también que algunas de las diferencias en este mapa y el mapa de apoyo Trump tienen explicaciones obvias. Trump perdió popularidad en Texas y Arkansas porque eran los estados de origen de dos de sus oponentes, Ted Cruz y Mike Huckabee.
- 13 Se trata de datos de la encuesta de Cívica Analítica de diciembre de 2015. Los datos reales de votación es menos útil en este caso, ya que está muy influenciada por la primaria, cuando se llevó a cabo y el formato de votación. Los mapas se imprimen con permiso de The New York Times.
- 15 billones de 2,5 millones de bytes de datos: “Traer grandes volúmenes de datos a la empresa”, IBM, <https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html>.
- 17 aguja en un pajar viene cada vez más grande: Nassim Taleb M., “Cuidado con los grandes errores de 'Big Data',” *Wired* 8 de febrero de 2013, <http://www.wired.com/2013/02/big-data-medios-grandes-errores-personas>.
- 18 ni búsquedas ni de miembros racista en Stormfront: He examinado cómo el racismo Internet cambió en partes del país con alta y baja exposición a la gran recesión. Miré a ambos tipos de búsqueda de Google para “negro (s)” y la pertenencia Stormfront. Los datos necesarios pueden descargarse en [sethsd.com](http://sethsd.com), en las secciones de datos titulado “la animosidad racial” y “Stormfront”.
- 18 Sin embargo, las búsquedas de Google que reflejan ansiedad: Seth Stephens-Davidowitz, “cincuenta estados de ansiedad,” *New York Times*, 7 de Agosto, 2016, SR2. Tenga en cuenta, mientras que las búsquedas de Google dan muestras mucho más grandes, este patrón es consistente con la evidencia de las encuestas. Véase, por ejemplo, William C. Reeves et al., “Vigilancia de la enfermedad mental entre los adultos en los Estados Unidos,” *Morbilidad and Mortality Weekly Report Suplemento* 60, no. 3 (2011).

- 18 búsqueda de chistes: Esto se discute en Seth Stephens-Davidowitz, “¿Por qué te ríes?” New York Times, 15 de mayo de 2016, SR9. Los datos necesarios pueden descargarse en sethsd.com, en la sección de datos titulado “Bromas”.
- 19 “mi marido quiere que lo amamantar”: Esto se discute en Seth Stephens-Davidowitz, “¿Qué quieren las mujeres embarazadas” New York Times, 17 de Mayo de 2014, SR6.
- 19 *porn busca representaciones de las mujeres en periodo de lactancia hombres: el análisis de los datos del autor Pornhub.*
- 19 Las mujeres constituyen casi la misma cantidad: Esto se discute en Seth Stephens-Davidowitz, “Búsqueda de Sexo”, New York Times, 25 de enero de 2015, SR1.
- 20 “Poemas para Mi esposa Embarazada”: Stephens-Davidowitz, “¿Qué tienen las mujeres embarazadas Want”
- 21 Friedman dice: Entrevisté Jerry Friedman por teléfono el 27 de octubre de 2015.
- 21 de muestreo de todos sus datos: Hal R. Varian, “Big Data: Nuevos trucos para Econometría”, Journal of Economic Perspectives 28, no. 2 (2.014).

## **CAPÍTULO 1: su intestino FAULTY**

- 26 La mejor ciencia de datos, de hecho, es sorprendentemente intuitivo: Estoy hablando de la esquina de análisis de datos que sé sobre la ciencia-datos que trata de explicar y predecir el comportamiento humano. No estoy hablando de la inteligencia artificial que intenta, por ejemplo, conducir un coche. Estas metodologías, mientras que lo hacen utilizar las herramientas descubiertas del cerebro humano, son menos fáciles de entender.
- 28 lo que predicen los síntomas de cáncer de páncreas: John Paparrizos, Ryan W. White y Eric Horvitz, “La detección de adenocarcinoma pancreático utilizando señales de Web Logs de búsqueda: estudio de viabilidad y resultados”, Journal of Oncology Practice (2016).
- 31 climático invierno inundado todo el resto: Esta investigación se discute en Seth Stephens-Davidowitz, “Dr. Google ahora le,”New York Times, 11 de Octubre de 2013, SR12.
- 32 conjunto de datos más grande jamás reunida en las relaciones humanas: Lars Backstrom y Jon Kleinberg, “Las asociaciones románticas y la dispersión de los lazos sociales: Un Análisis de redes de relación de estado en Facebook”, en Actas de la 17ª Conferencia ACM sobre trabajo cooperativo asistido por computadora y la informática social (2014).
- 33 personas alinean constantemente: Daniel Kahneman, Pensar rápido (Nueva York: Farrar, Straus y Giroux, 2011).
- 33 el asma causa cerca de setenta veces más muertes: Entre 1979 y 2010, en promedio, 55,81 estadounidenses murieron a causa de los tornados y 4216.53 estadounidenses

- murieron a causa de asma. Ver anual nos Killer Estadísticas Tornado, Servicio Meteorológico Nacional, <http://www.spc.noaa.gov/climo/torn/fatalmap.php> y tendencias en Asma morbilidad y mortalidad, Asociación Americana del Pulmón, Unidad de Epidemiología y Estadística.
- 33 Patrick Ewing: Mis vídeos favoritos Ewing son “Top 10 Carrera Obras de Patrick Ewing,” vídeo de YouTube, publicada el 18 de septiembre de, 2015, [https://www.youtube.com/watch?v=Y29gMuYymv8](https://www.youtube.com/watch?v=Y29gMuYymv8;); y “Patrick Ewing Knicks tributo,” vídeo de YouTube, publicada el 12 de mayo de 2006, <https://www.youtube.com/watch?v=8T2l5Emzu-I>.
- 34 “el baloncesto como una cuestión de vida o muerte”: “¿Qué pasó con el deportista blanca” SL Precio, Sports Illustrated, 8 de diciembre de 1997.
- 34 una encuesta en Internet: Esta fue una encuesta de opinión Google llevé a cabo el 22 de octubre de 2013. Le pregunté, “¿Dónde adivinar que la mayoría de los jugadores de la NBA han nacido?” Las dos opciones eran “los barrios pobres” y “de clase media barrios”; 59,7 por ciento de los encuestados eligió “barrios pobres.”
- 36 primer nombre de una persona negro es una indicación de su origen socioeconómico: Roland G. Fryer Jr. y Steven D. Levitt, “las causas y consecuencias de distintivamente negro nombres”, Quarterly Journal of Economics 119, no. 3 (2004).
- 37 Entre todos los afroamericanos nacidos en la década de 1980: Centros para el Control y la Prevención de Enfermedades, “Salud, Estados Unidos, 2009,” Tabla 9, no matrimonial maternidad, por raza detallada y origen hispano de la madre, y la edad materna: Estados Unidos, seleccionados años 1970-2006.
- 37 Chris Bosh. . . Chris Paul: “No es sólo un atleta típico: Delantero de Miami Heat Chris Bosh intereses van mucho más allá de baloncesto,” PalmBeachPost.com, 15 de febrero de 2011, <http://www.palmbeachpost.com/news/sports/basketball/not-just-a-typical-Jock-miami-calor-forward-Chris-b/nLp7Z/>; Dave Walker, “Familia de Chris Paul a competir en 'Family Feud', nola.com 31 de octubre de 2011, [http://www.nola.com/tv/index.ssf/2011/10/chris\\_pauls\\_family\\_to\\_compete.html](http://www.nola.com/tv/index.ssf/2011/10/chris_pauls_family_to_compete.html).
- 38 de cuatro pulgadas más alto: “¿Por qué estamos cada vez más altas como especie?” Scientific American, <http://www.scientificamerican.com/article/why-are-we-getting-taller/>. Curiosamente, los estadounidenses han dejado cada vez más altas. Amanda Cebolla, “¿Por qué los estadounidenses dejaron de crecer más alto?” ABC News 3 de julio de 2016, <http://abcnews.go.com/Technology/story?id=98438&page=1>. He argumentado que una de las razones que ha habido un enorme incremento en jugadores de la NBA nacidos en el extranjero es que otros países están alcanzando a los Estados Unidos de altura. El número de siete pies de altura nacidos en Estados Unidos en la NBA aumentó dieciséis veces 1946-80 como estadounidenses crecieron. Desde entonces se ha estabilizado, ya que los estadounidenses han dejado de crecer. Mientras tanto, el número de siete pies de altura de otros países ha aumentado considerablemente. El mayor aumento de jugadores internacionales, encontré,

- 38 estadounidenses de origen humilde: Carmen R. Isasi et al., “Asociación de la Infancia dificultades económicas con la estatura adulta y la adiposidad adultos entre los hispanos / latinos: El HCHS / SOL Sociocultural auxiliar Estudio,” PLoS One 11, no. 2 (2016); Jane E. Miller y Sanders Korenman, “La pobreza y el estado nutricional de los niños en los Estados Unidos,” American Journal of Epidemiology 140, no. 3 (1994); Harry J. Holzer, Diane Whitmore Schanzenbach, Greg J. Duncan, y Jens Ludwig, “Los costos económicos de la niñez de la pobreza en los Estados Unidos,” Diario de la Infancia y la pobreza 14, no. 1 (2008).
- 38 el hombre americano promedio es de 5'9” : Cheryl D. Fryar, Qiuping Gu, y Cynthia L. Ogden, ‘antropométrica de datos de referencia para niños y adultos: Estados Unidos, 2007-2010,’ Vital y Estadísticas de Salud de la serie 11, sin . 252 (2012).
- 39 algo así como uno de cada cinco llega a la NBA: Pablo S. Torre, “más grande que la vida real,” Sports Illustrated, 4 de julio, 2011.
- 39 de la clase media, las familias con dos padres: Tim Kautz, James J. Heckman, Ron Diris, Bajo Ter Weel, y Lex Borghans, “Fomento y medición de competencias: Mejora de habilidades cognitivas y no cognitivas para promover el curso de la vida Éxito,” Oficina Nacional del Documento de Trabajo de Investigación Económica 20749, 2014.
- 39 Wrenn saltó el más alto: Desmond Conner, “Por Wrenn, del cielo es el límite”, Hartford Courant, el 21 de octubre de 1999.
- 39 Pero Wrenn: la historia de Doug Wrenn se cuenta en Percy Allen, “Ex Washington y O’Dea estrella Doug Wrenn Finds tiempos difíciles,” Seattle Times, 29 de marzo, 2009.
- 40 “*Doug Wrenn está muerto*”: *Ibid.*
- 40 Jordan podría ser un niño difícil: Melissa Isaacson, “Retrato de una leyenda”, ESPN.com 9 de septiembre de 2009, [http://www.espn.com/chicago/columns/story?id=4457017&columnist=isaacson\\_melissa](http://www.espn.com/chicago/columns/story?id=4457017&columnist=isaacson_melissa). Una buena biografía Jordan es Roland Lazenby, Michael Jordan: La Vida (Boston: Back Bay Books, 2015).
- 40 Su padre fue: Barry Jacobs, “alto vuelo Michael Jordan Carolina del Norte Navegando hacia otro NCAA Título:” La gente, 19 de Marzo., 1984
- 40 *La vida de Jordan está llena de historias de su familia guiándolo: Isaacson, “Retrato de una leyenda”.*
- 41 discurso tras la inducción al Salón de la Fama: salón de la fama del habla Enshrinement, video de YouTube de baloncesto de Michael Jordan, publicada el 21 de febrero del 2012, <https://www.youtube.com/watch?v=XLzBMGXfK4c>. El aspecto más interesante del discurso de Jordan no es que él es tan efusiva sobre sus padres; es que todavía se siente la necesidad de señalar desaires desde temprano en su carrera. Tal vez una obsesión de por vida con desaires es necesaria para convertirse en el mejor jugador de baloncesto de todos los tiempos.

41 LeBron James fue entrevistado: “Estoy LeBron James de Akron, Ohio,” vídeo de YouTube, publicada el 20 de junio de 2013, <https://www.youtube.com/watch?v=XceMbPVAggk>.

## CAPÍTULO 2: ERA FREUD DERECHO?

47 un alimento de ser la forma de un falo: codifiqué alimentos como ser en forma de un falo si fueran significativamente más larga que ancha y, en general redonda. Conté pepinos, maíz, zanahoria, berenjena, la calabaza, y plátanos. Los datos y el código se pueden encontrar en [sethsd.com](http://sethsd.com).

48 errores recogidos por los investigadores de Microsoft: El conjunto de datos se pueden descargar en <https://www.microsoft.com/en-us/download/details.aspx?id=52418>. Los investigadores pidieron a los usuarios de Amazon Mechanical Turk para describir las imágenes. Se analizaron los registros de pulsaciones y observaron cada vez que alguien corrija una palabra. Más detalles se pueden encontrar en Yukino Baba y Hisami Suzuki, “¿Cómo se generan y los errores de ortografía corregida? Un estudio de los errores de ortografía corregido y sin corregir Uso de teclas Registros,” Actas de la Quincuagésima Reunión Anual de la Asociación de Lingüística Computacional, 2012. Los datos, código, y una descripción más detallada de esta investigación se puede encontrar en [sethsd.com](http://sethsd.com).

51 *Considere todas las búsquedas del tipo “yo quiero tener relaciones sexuales con mi”:  
La alerta datos completos gráfica es la siguiente:*

### “Quiero tener sexo. . “.

	MENSUAL GOOGLE búsquedas con esta frase exacta
mi mamá	720
mi hijo	590
mi hermana	590
mi primo	480
mi papa	480
mi novio	480
mi hermano	320
mi hija	260
mi amigo	170



mi novia	140
----------	-----

52 de dibujos animados porno: Por ejemplo, “porno” es una de las palabras más comunes incluidos en las búsquedas de Google para varios programas animados muy populares, como se ve a continuación.

**CARTOONS, MEET PORN**  
(MOST COMMON GOOGLE SEARCHES FOR VARIOUS CARTOONS)

<b>family guy porn</b>	watch the simpsons	<b>futurama porn</b>	scooby doo games
family guy episodes	<b>the simpsons porn</b>	futurama leela	scooby doo movie
family guy free	the simpsons online	futurama episodes	<b>scooby doo porn</b>
watch family guy	the simpsons movie	futurama online	scooby doo velma

52 canguros: En base a cálculos del autor, estas son las ocupaciones femeninas más populares en las búsquedas de pornografía por los hombres, desglosados por la edad de los hombres:

**OCCUPATIONS OF WOMEN IN MALE PORN SEARCHES, BY AGE OF MEN**

	18-24	25-64	65+
1.	Babysitter	Babysitter	Babysitter
2.	Teacher	Yoga Instructor	Cheerleader
3.	Yoga Instructor	Teacher	Doctor
4.	Cheerleader	Cheerleader	Teacher
5.	Doctor	Real Estate Agent	Real Estate Agent
6.	Prostitute	Doctor	Nurse
7.	Real Estate Agent	Prostitute	Yoga Instructor
8.	Nurse	Secretary	Secretary
9.	Secretary	Nurse	Prostitute

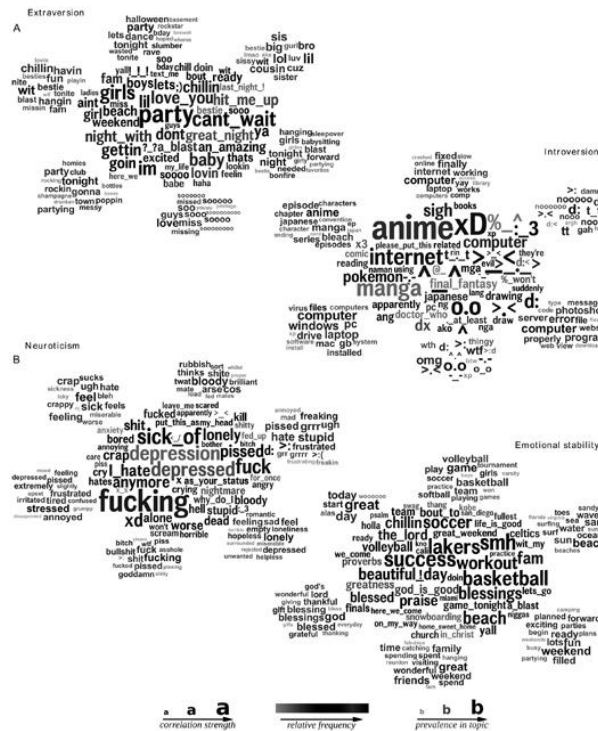
### Capítulo 3: DATOS reimagined

56 algoritmos en su lugar: Mateo Leising, “HFT Tesoro Trading Hurts mercado Cuando es liberado Noticias,” Bloomberg Markets 16 de diciembre de 2014; Nathaniel Popper, “Los robots están viniendo para Wall Street,” New York Times Magazine, 28 de Febrero, 2016, MM56; Richard dedo, “High Frequency Trading: ¿Es una fuerza oscura contra ordinarias operadores humanos y los inversores?” Forbes, 30 de Septiembre de 2013, <http://www.forbes.com/sites/richardfinger/2013/09/30/high-> la frecuencia de comercio-es-que-a-oscuro-fuerza-contra-de humano comerciantes ordinarios-and-inversores / # 50875fc751a6.

56 Alan Krueger: Me entrevistó a Alan Krueger por teléfono el 8 de mayo de 2015.

- 57 indicadores importantes de la rapidez con la gripe: El documento inicial fue Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, y Larry Brilliant, “Detección de las epidemias de gripe Uso de consulta de datos de motor de búsqueda,” *Naturaleza* 457, no. 7232 (2009). Las fallas en el modelo original se discutieron en David Lazer, Ryan Kennedy, Gary King, y Alessandro Vespignani, “La parábola de la gripe de Google: Trampas en Los análisis de datos,” *Ciencia* 343, no. 6176 (2014). Un modelo corregido se presenta en Shihao Yang, Mauricio Santillana, y SC Kou, “estimación precisa de la influenza epidemias Uso de Google Búsqueda de datos a través de ARGO,” *Actas de la Academia Nacional de Ciencias* 112, no. 47 (2015).
- 58 que las búsquedas Canción más estrechamente precios de la vivienda: Seth Stephens-Davidowitz y Hal Varian, “Una guía práctica para datos de Google”, fotocopias, 2015. Véase también Marcelle Chauvet, Stuart Gabriel, y Chandler Lutz, “Hipoteca Riesgo defecto: Nueva La evidencia de Internet de consultas de búsqueda,” *Revista de Economía urbana* 96 (2016).
- 60 Bill Clinton: Sergey Brin y Larry Page, “La Anatomía de un Gran Escala hipertextual Web motor de búsqueda,” Séptimo Internacional World Wide Web Conferencia, Abril 14-18, 1998 Brisbane, Australia.
- 61 *los sitios de pornografía: John Battelle, La búsqueda: Cómo Google y sus rivales reescribió las reglas de negocio y transformado nuestra cultura (Nueva York: Penguin, 2005).*
- 61 Crowdsourcing las opiniones: Una buena discusión de esto se puede encontrar en Steven Levy, *En el Plex: ¿Cómo piensa Google, Obras y modela la existencia* (Nueva York: Simon & Schuster, 2011).
- 64 “Si quiere vender”: Esta cita también se incluyó en el Joe drapeado, “Viaje de Ahmed Zayat: Quiebras y Apuestas Grandes”, *New York Times*, 5 de junio de 2015, A1. Sin embargo, el artículo atribuye incorrectamente a la cita Seder. En realidad, fue hecha por otro miembro de su equipo.
- 65 I primero se reunió con Seder: entrevisté Jeff Seder y Patty Murray en Ocala, Florida, del 12 de junio de 2015, hasta el 14 de junio de 2015.
- 66 Aproximadamente un tercio: Los caballos de carreras razones fracasan son estimaciones aproximadas por Jeff Seder, basado en sus años en el negocio.
- 66 cientos de caballos mueren: cuadros Adicional de la lesión Equina Base de Estadísticas de pura sangre, [http://jockeyclub.com/pdfs/eid\\_7\\_year\\_tables.pdf](http://jockeyclub.com/pdfs/eid_7_year_tables.pdf).
- 66 sobre todo debido a la rotura de las piernas: “Programa Postmortem examen,” *California Salud Animal y del sistema Food Laboratory*, 2013.
- 67 Sin embargo, más de las tres cuartas partes no ganan una importante carrera: Avalyn Hunter, “Un caso para la plena Hermanos,” *Bloodhorse* 18 de abril de 2014, [http://www.bloodhorse.com/horse-racing/articles/115014 / a-caso-de-hermanos completos](http://www.bloodhorse.com/horse-racing/articles/115014/a-caso-de-hermanos-completos).

- 67 Earvin Johnson III: Melody Chiu, "EJ Johnson pierde el 50 Lbs. Desde sometidos a cirugía de manga gástrica," *La gente* 1 de octubre de 2014.
- 67 LeBron James, cuya madre es 5'5": Eli Saslow, 'Historias de LeBron, Parte 1,' ESPN.com, 17 de octubre, 2013 Perdido, [http://www.espn.com/nba/story/\\_/Identificación/9825052/cómo-lebron-james-vida-cambiado-de-cuarto-grado-espn-revista](http://www.espn.com/nba/story/_/Identificación/9825052/cómo-lebron-james-vida-cambiado-de-cuarto-grado-espn-revista).
- 68 The Green Monkey: Ver Sherry Ross, "16 Million Dollar Baby," *New York Daily News*, 12 de marzo de 2006, y Jay Privman, "The Green Monkey, que vendió por \$ 16M, Retirado," ESPN.com 12 de febrero 2008, <http://www.espn.com/sports/horse/news/story?id=3242341>. Un video de la subasta está disponible en "caballo, \$ 16 millones" de video de YouTube, publicado 1 noviembre de 2008, <https://www.youtube.com/watch?v=EyggMC85Zsg>.
- 71 debilidad de intento de Google para predecir la gripe: Sharad Goel, Jake M. Hofman, Sébastien Lahaie, David M. Pennock, y Duncan J. Watts, "Predecir el comportamiento del consumidor con la búsqueda web," *Actas de la Academia Nacional de Ciencias* 107, sin . 41 (2010).
- 72 Fresa Pop-Tarts: Constance L. Hays, "Lo que Wal-Mart sabe acerca de los hábitos de los clientes," *New York Times*, 14 de noviembre de 2004.
- 74 "que funcionaba de maravilla": Entrevisté Orley Ashenfelter por teléfono el 27 de octubre 2016.
- 80 estudiaron cientos de personas que se citan velocidad heterosexuales: Daniel A. McFarland, Dan Jurafsky, y Craig Rawlings, "Haciendo la conexión: la vinculación social en situaciones de cortejo," *American Journal of Sociology* 118, no. 6 (2013).
- 82 Leonard Cohen, una vez dio a su sobrino el siguiente consejo para cortejar a las mujeres: Jonathan Greenberg, "lo que aprendí de mi sabio tío Leonard Cohen," *The Huffington Post*, 11 de Noviembre 2016.
- 83 las palabras usadas en cientos de miles de Facebook: H. Andrew Schwartz et al, "La personalidad, género y edad en el lenguaje de los medios sociales: el enfoque de Vocabulario abierto," *PLoS One* 8, no. 9 (2013). En el documento también se desglosa las formas en que se hable en base a la forma en que la puntuación en los tests de personalidad. Esto es lo que encontraron:



- 88 texto de miles de libros y guiones de cine: Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, y Peter Sheridan Dodds, “los arcos emocionales de las historias están dominados por seis formas básicas,” EPJ Ciencia Datos 5, no. 1 (2016).
- 91 ¿Qué tipo de historias se comparten: Jonah Berger y Katherine L. Lecher, “¿Qué hace que los contenidos en línea viral?” Journal of Marketing Research 49, no. 2 (2012).
- 95 ¿Por qué algunas publicaciones mágica izquierda: Esta investigación tiene que ver con la pulpa expuesta en Mateo Gentzkow y Jesse M. Shapiro, “¿Qué hay detrás de Medios inclinación? La evidencia de los Estados Unidos Diarios”, Econometrica 78, núm. 1 (2010). A pesar de que no eran más Ph.D. estudiantes cuando este proyecto comenzó, Gentzkow y Shapiro son ahora los economistas estrella. Gentzkow, ahora profesor en Stanford, ganó la Medalla John Bates Clark 2014, dado que el economista superior debajo de la edad de cuarenta años. Shapiro, ahora profesor en Brown, es un editor de la prestigiosa revista Journal of Political Economy. Su documento conjunto sobre la inclinación de los medios de comunicación es uno de los artículos más citados para cada uno.
- 96 Rupert Murdoch: la propiedad de Murdoch de la conservadora Nueva York Post podría explicarse por el hecho de que Nueva York es tan grande, que puede soportar periódicos de múltiples puntos de vista. Sin embargo, parece bastante claro el Post pierde dinero constantemente. Véase, por ejemplo, Joe Pompeo, “¿Cuánto cuesta el 'New York Post' pierden realmente?” Político 30 de agosto de 2013,

- <http://www.politico.com/media/story/2013/08/how-much-¿Eso-la-nueva-york-post-realidad-perder-001176>.
- 97 Shapiro me dijo: Entrevisté a Matt Gentzkow y Jesse Shapiro el 16 de agosto de 2015, en el Royal Sonesta Boston.
- 98 anuarios escaneadas de las escuelas secundarias de Estados Unidos: Kate Rakelly, Sarah Sachs, Brian Yin, y Alexei A. Efros, “Un siglo de retratos: un registro visual histórico de American High School anuarios”, documento presentado en iccv, 2015. las fotos se han impreso con el permiso de los autores.
- 99 sujetos en las fotografías temas copiados en las pinturas: Ver, por ejemplo, Christina Kotchemidova, “¿Por qué decimos 'queso': La producción de la sonrisa en la fotografía instantánea,” *Estudios Críticos en Medios de Comunicación* 22, no. 1 (2005).
- 100 PIB medida basada en la cantidad de luz que hay en estos países por la noche: J. Vernon Henderson, Adam Storeygard, y David N. Weil, “Midiendo el crecimiento económico del espacio exterior”, *American Economic Review* 102, no. 2 (2012).
- 101 PIB estimado era ahora el 90 por ciento más alto: Kathleen Caulderwood, “Nigeria PIB Saltos 89% como economistas Añadir en telecomunicaciones, Nollywood,” *IBTimes* 7 de abril de 2014, <http://www.ibtimes.com/nigerian-gdp-jumps-89-economistas-add-telecomunicaciones-nollywood-1568219>.
- 101 Reisinger dijo: Entrevisté a Joe Reisinger por teléfono el 10 de junio de 2015.
- 103 \$ 50 millones de dólares: Leena Rao, “SpaceX y Tesla Backer acaba de invertir \$ 50 millones en este inicio,” *Fortune* 24 de septiembre de 2015.

#### **CAPÍTULO 4: DIGITAL suero de la verdad**

- 106 papel importante en 1950: Hugh J. Parry y Helen M. Crossley, “validez de las respuestas a preguntas de la encuesta,” *Public Opinion Quarterly* 14, 1 (1950).
- 106 encuesta preguntó a los graduados de la Universidad de Maryland: Frauke Kreuter, Stanley Presser, y Roger Tourangeau-, “Conveniencia de polarización social en CATI, IVR, y encuestas Web,” *Public Opinion Quarterly* 72 (5), 2008.
- 107 fracaso de las urnas: Para un artículo argumentando que la mentira puede ser un problema al tratar de predecir el apoyo a Trump, véase Thomas B. Edsall, “¿Cuántas personas de apoyo Trump, pero no quiero admitirlo?” *New York Times*, 15 de mayo de, 2016, SR2. Sin embargo, para un argumento de que esto no era un factor grande, ver a Andrew Gelman, “Explicaciones para que impactante Shift 2%,” *Modelización Estadística, Inferencia causal y Ciencias Sociales* 9 de noviembre de 2016, <http://andrewgelman.com/2016/11/09/explicaciones-chocantes-2-shift/>.
- 107 dice Tourangeau-: entrevisté a Roger Tourangeau- por teléfono el 5 de mayo de 2015.
- 107 por lo que muchas personas dicen que están por encima del promedio: Esto se discute en Adam Grant, *Originales: ¿Cómo no conformistas mover el mundo* (Nueva York:

- Viking, 2016). La fuente original es David Dunning, Chip Heath, y Jerry M. Suls, “defectuoso Autoevaluación: Implicaciones para la Salud, Educación, y el lugar de trabajo,” *Ciencia Psicológica en el Interés Público* 5 (2004).
- 108 meterse con encuestas: Anya Kamenetz, “Confundir Investigación 'Travieso' respondedores en los adolescentes,” *nprED* 22 de mayo de 2014, <http://www.npr.org/sections/ed/2014/05/22/313166161/mischievous-respondedores-factor-de-confusion-en-la-investigacion-en-adolescentes>. La investigación original de este artículo se discute es Joseph P. Robinson-Cimpian, “Estimación inexacto de las disparidades Debido a dañosa respondedores,” *Investigador Educativo* 43, no. 4 (2014).
- 110 búsqueda “porn” más que la búsqueda de “tiempo”: <https://www.google.com/trends/explore?date=all&geo=US&q=porn,weather>.
- 110 admiten que miran pornografía: Amanda Hess, “¿Cuántas mujeres no admitir a Pew Eso que ven pornografía” *Slate*, 11 de octubre de 2013, [http://www.slate.com/blogs/xx\\_factor/2013/10/11/pew\\_online\\_viewing\\_study\\_percentage\\_of\\_women\\_who\\_watch\\_online\\_porn\\_is\\_growing.html](http://www.slate.com/blogs/xx_factor/2013/10/11/pew_online_viewing_study_percentage_of_women_who_watch_online_porn_is_growing.html).
- 110 “gallo”, “mierda” y “porno”: Nicholas Diakopoulos “sexo, violencia, y Algoritmos de autocompletar,” *Slate* 2 de agosto de 2013, [http://www.slate.com/articles/technology/future\\_tense/2013/08/words\\_banned\\_from\\_bing\\_and\\_google\\_s\\_autocomplete\\_algorithms.html](http://www.slate.com/articles/technology/future_tense/2013/08/words_banned_from_bing_and_google_s_autocomplete_algorithms.html).
- 111 3,6 veces más propensos a decir que lamentan Google: Calculo, incluyendo varios fraseos, hay cerca de 1.730 estadounidense Google busca en todos los meses diciendo explícitamente que lamentan tener hijos. Hay solamente cerca de 50 que expresa una queja es no tener hijos. Hay alrededor de 15,9 millones de estadounidenses mayores de cuarenta y cinco años que no tienen hijos. Hay alrededor de 152 millones de estadounidenses que tienen los niños. Esto significa, entre la población elegible, las personas con niños son alrededor de 3,6 veces más propensos a expresar un lamento en Google que las personas sin hijos. Obviamente, como se menciona en el texto, pero cabe destacar una vez más, estos confesionarios a Google sólo son hechas por un selecto número pequeño, de las personas, presumiblemente aquellos que se sientan lo suficientemente fuerte lamentan que momentáneamente se olvida que Google no puede ayudar aquí.
- 113 más alto apoyo al matrimonio gay: Estas estimaciones son de Nate Silver, “¿Cómo Dictamen sobre el matrimonio homosexual está cambiando, y lo que significa,” *FiveThirtyEight* 26 de marzo de 2013, [http://fivethirtyeight.blogs.nytimes.com/2013/03/26/cómo-opinión-en-mismo-sexo-matrimonio-se-cambiando-y-lo-que-medias/?\\_r=0](http://fivethirtyeight.blogs.nytimes.com/2013/03/26/cómo-opinión-en-mismo-sexo-matrimonio-se-cambiando-y-lo-que-medias/?_r=0).
- 113 Sobre el 2,5 por ciento de los usuarios masculinos de Facebook que enumeran una de género de interés dicen que están interesados en los hombres: Análisis de autor de los

- anuncios de Facebook de datos. No incluyo usuarios de Facebook que lista de “hombres y mujeres.” Mi análisis sugiere un porcentaje no trivial de usuarios que dicen que están interesados en los hombres y las mujeres interpretan la pregunta de interés en la amistad lugar de interés romántico.
- 115 alrededor del 5 por ciento de las búsquedas pornográficas masculinas son para el porno gay masculina: Como se discutió, Google Trends no se descompone búsquedas por género. Google AdWords se rompe páginas vistas para las diversas categorías de género. Sin embargo, estos datos es mucho menos precisa. Para estimar las búsquedas por género, por primera vez utilizar los datos de búsqueda para obtener una estimación del estado del porcentaje de búsquedas porno por estado. entonces normalizar estos datos por los datos de género de Google AdWords. Otra forma de obtener datos específicos de género está utilizando datos de Pornhub. Sin embargo, Pornhub podría ser una muestra altamente seleccionada, ya que muchas personas homosexuales podrían utilizar en su lugar se centró sólo en los sitios porno gay. Pornhub sugiere que el uso porno gay entre los hombres es más baja que las búsquedas de Google sugerirían. Sin embargo, se confirma que no hay una fuerte relación entre la tolerancia hacia la homosexualidad y uso de la pornografía gay masculina.
- 116 4 por ciento de ellos son abiertamente homosexual en Facebook: Cálculos del autor de los anuncios de Facebook de datos: El 8 de febrero de 2017, cerca de 300 estudiantes de secundaria de sexo masculino en el mercado de los medios de comunicación de San Francisco-Oakland-San José en Facebook dijeron que estaban interesados en los hombres. Aproximadamente, 7.800 dijeron que estaban interesados en las mujeres.
- 119 “En Irán no tenemos homosexuales”: “ ‘Nosotros no tenemos homosexuales en Irán,’ Dice el presidente iraní audiencia de la Ivy League,” DailyMail.com, 25 de septiembre de 2007, <http://www.dailymail.co.uk/news/article-483746/We-dont-gays-Iran-Iranian-president-tells-Ivy-League-audience.html>.
- 119 “No tenemos en nuestra ciudad”: Brett Logiurato, “Sochi Alcalde afirma que hay Nadie gay de la ciudad,” Sports Illustrated 27 de enero de 2014.
- 119 comportamiento de Internet revela un interés significativo en el porno gay en Sochi e Irán: De acuerdo con Google AdWords, hay decenas de miles de búsquedas cada año por “рей porno” (porno gay). El porcentaje de búsquedas de pornografía de porno gay es más o menos similar en Sochi como en los Estados Unidos. Google AdWords no incluye datos para Irán. Pornhub también no reporta datos para Irán. Sin embargo, PornMD estudió sus datos de búsqueda e informó de que cinco de los diez primeros términos de búsqueda en Irán eran para el porno gay. Esto incluyó “amor papá” y “hombre de negocios de hotel” y se ha reportado en Joseph Patrick McCormick, “Encuesta revela búsquedas de Porno Gay son de primera en los países Banning homosexualidad,” PinkNews, <http://www.pinknews.co.uk/2013/03/13/encuesta-revela-un-registro-corporal-para-gay-porn-son-top-en-paises-que-prohiben-la-homosexualidad/>. De acuerdo con Google Trends,

- 122 Cuando se trata de sexo: Stephens-Davidowitz, “Búsqueda de sexo”. Los datos para esta sección se puede encontrar en mi sitio web, sethsd.com, en la sección “Sexo”.
- 122 11 por ciento de las mujeres: Estado actual de anticonceptivos entre las mujeres de 15-44: Estados Unidos, 2011-2013, los Centros para el Control y la Prevención de Enfermedades, [http://www.cdc.gov/nchs/data/databriefs/db173\\_table.pdf# 1](http://www.cdc.gov/nchs/data/databriefs/db173_table.pdf#1).
- 122 10 por ciento de ellos para quedar embarazada cada mes: David Spiegelhalter, “Sexo: Cuáles son las posibilidades?” BBC News, 15 de marzo del 2012, <http://www.bbc.com/future/story/20120313-sex-in-the-ciudad-o-otro-lugar>.
- 122 1 en 113 mujeres en edad fértil: hay aproximadamente 6,6 millones de embarazos cada año y hay 62 millones de mujeres entre las edades de 15 y 44.
- 128 realiza el sexo oral sobre el sexo opuesto: Como se ha mencionado, no sé el sexo de un buscador de Google. Estoy asumiendo que la inmensa mayoría de las búsquedas que buscan la forma de realizar el cunnilingus son los hombres y que la inmensa mayoría de las búsquedas que buscan la forma de realizar sexo oral son por mujeres. Esto es a la vez debido a que la gran mayoría de las personas son rectas y porque no podría ser menos de una necesidad de aprender cómo complacer a una pareja del mismo sexo.
- 128 cinco palabras negativas: Análisis de autor de datos de Google AdWords.
- 130 que se maten: Evan y Seth SOLTAS Stephens-Davidowitz, “The Rise of Hate Search”, New York Times, 13 de diciembre de 2015, SR1. Datos y más detalles se pueden encontrar en mi sitio web, sethsd.com, en la sección “islamofobia”.
- 132 diecisiete veces más común: el análisis de autor de datos de Google Trends.
- 132 Día de Martin Luther King Jr.: Análisis de autor de datos de Google Trends.
- 133 se correlaciona con la diferencia de salarios blanco-negro: Ashwin Rode y Anand Shukla J., “actitudes de prejuicio y resultados del mercado laboral”, fotocopias, 2013.
- 134 Sus padres: Seth Stephens-Davidowitz, “Google, Dime. ¿Es mi hijo un genio?” New York Times, 19 de enero de 2014, SR6. Los datos para búsquedas exactas pueden encontrarse usando Google AdWords. Las estimaciones también se pueden encontrar con Google Trends, mediante la comparación de las búsquedas con las palabras “dotado” e “hijo” frente a “dotado” y “hija”. Compárese, por ejemplo, [https://www.google.com/trends/explore? fecha = all y geo = US y q =% 20son regalado, dotado% 20daughter y](https://www.google.com/trends/explore?fecha=all&geo=US&q=%20son%20regalado,%20dotado%20daughter)  
<https://www.google.com/trends/explore?date=all&geo=US&q=overweight%20son,overweight%20daughter>. Una excepción a la tendencia general de que hay más preguntas acerca de los cerebros y cuerpos hijas hijos es que hay más búsquedas de ‘gordo hijo’ que ‘hija de grasa.’ Esto parece estar relacionado con la popularidad de incesto porno discutido anteriormente.
- 135 niñas son un 9 por ciento más probabilidades que los niños estén en programas para niños dotados: “La equidad de género en la educación: Una fotografía instantánea de datos”, Oficina de Derechos Civiles del Departamento de Educación de Estados



- Unidos, junio de 2012, <http://www2.ed.gov/about/offices/list/ocr/docs/gender-equity-in-education.pdf>.
- 136 Alrededor del 28 por ciento de las niñas tienen sobrepeso, mientras que el 35 por ciento de los niños son: Centro de Recursos de datos para la Salud del Niño y del Adolescente, <http://www.childhealthdata.org/browse/survey/results?q=2415&g=455&a=3879&r=1>.
- 137 perfiles del frente de tormenta: Stephens-Davidowitz “Los datos de odio” Los datos necesarios pueden descargarse en [sethsd.com](http://sethsd.com), en la sección de datos titulado “Stormfront”.
- 139 Stormfront durante la candidatura de Donald Trump: Google búsqueda interés en Stormfront fue similar en octubre de 2016 para los niveles Fue durante octubre de 2015. Esto está en marcado contraste con la situación durante la primera elección de Obama. En octubre de 2008, el interés de búsqueda en Stormfront había aumentado casi un 60 por ciento en comparación con octubre del año anterior. El día después de que Obama fue elegido, las búsquedas de Google para Stormfront habían subido más o menos diez veces. Al día siguiente de Trump fue elegido, búsquedas del frente de tormenta se elevaron alrededor de dos puntos y cinco veces. Esto era más o menos equivalente a la subida del día después de que George W. Bush fue elegido en 2004 y puede reflejar en gran medida el interés de noticias entre los adictos a la política.
- 141 segregación política en Internet: Mateo Gentzkow y Jesse M. Shapiro, “ideológico Segregación online y offline”, *Quarterly Journal of Economics* 126, no. 4 (2011).
- 144 amigos en Facebook: Eytan bakshy, Salomón Messing y Lada A. Adamic, “la exposición a Ideológicamente diversa información y opinión en Facebook,” *Science* 348, no. 6239 (2015). Ellos encontraron que, entre el 9 por ciento de los usuarios activos de Facebook que declaran su ideología, alrededor del 23 por ciento de sus amigos que también declarar una ideología tiene la ideología opuesta y el 28,5 por ciento de las noticias que ven en Facebook es de la ideología opuesta. Estas cifras no son directamente comparables con otros números en la segregación, ya que sólo incluyen la pequeña muestra de los usuarios de Facebook que declaran su ideología. Presumiblemente, estos usuarios son mucho más propensos a ser políticamente activa y asociarse con otros usuarios políticamente activos con la misma ideología. Si esto es correcto, la diversidad entre todos los usuarios será mucho mayor.
- 144 una razón fundamental que Facebook: Otro factor que hace que las redes sociales sorprendentemente diversa es que le da una gran ventaja a los artículos muy populares y ampliamente compartidos, independientemente de su inclinación política. Ver Salomón Messing y Sean Westwood, “exposición selectiva en la era de los medios sociales: Los endosos Trump Partisan Fuente Afiliación al seleccionar News Online,” 2014.
- 144 amigos en Facebook más de lo que hacen en línea: Ver Ben Quinn, “usuarios de redes sociales tienen el doble de amigos en línea como en la vida real,” *The Guardian*, 8 de

- mayo de 2011. Este artículo se describe un estudio de 2011 por la Fibrosis Quística Trust, que encontrado que el usuario medio de la red social tiene 121 amigos en línea en comparación con 55 amigos físicos. Según un estudio de Pew Research 2014, el usuario promedio de Facebook tiene más de 300 amigos. Ver Aaron Smith, “6 nuevos datos sobre Facebook,” 3 de febrero de 2014, <http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/>.
- 144 lazos débiles: Eytan bakshy, Itamar Rosenn, Cameron Marlow, y Lada Adamic, “El papel de las redes sociales en la difusión de información,” Actas de la 21ª Conferencia Internacional sobre el World Wide Web, 2012.
- 145 “predicciones doom-and-penumbra no se han hecho realidad”: “Estudio: Abuso Infantil en decadencia en los Estados Unidos,” Associated Press 12 de diciembre, 2011.
- 146 tenía abuso infantil realmente caer: mimeo Ver Seth Stephens-Davidowitz, “¿Cómo Abuso googlear desenmascara Niño”, New York Times, 14 de julio de 2013, SR5, y Seth Stephens-Davidowitz, “Las víctimas no declarada de una crisis económica”, 2013 .
- 146 frente a largos tiempos de espera y renunciar: “Detener el Abuso Infantil: Se comienza con usted,” The Arizona Republic, 26 de Marzo 2016.
- 147 off-the-libros maneras de terminar un embarazo: Seth Stephens-Davidowitz, “El retorno del bricolaje Aborto,” New York Times 6 de marzo de 2016, SR2. Datos y más detalles se pueden encontrar en mi sitio web, [sethsd.com](http://sethsd.com), en la sección “auto-inducido Aborto.”
- 150 circulaciones promedio similares: Alianza para el auditado Media, revistas de consumo, <http://abcas3.auditedmedia.com/ecirc/magtitlesearch.asp>.
- 151 En Facebook: Cálculos del autor, el 4 de octubre de 2016, mediante administrador de anuncios de Facebook.
- 151 diez sitios web más visitados: “Lista de los sitios web más populares,” Wikipedia. De acuerdo con Alexa, que mide el comportamiento de navegación, como de 4 de septiembre, 2016, el sitio porno más popular fue xvideos, y este fue el sitio más popular 57a. De acuerdo con SimilarWeb, como de 4 de septiembre, 2016, el sitio porno más popular fue xvideos, y este fue el sitio 17 el más popular. Los diez primeros, según Alexa, son Google, YouTube, Facebook, Baidu, Yahoo !, Amazon, Wikipedia, Tencent QQ, Google India, y Twitter.
- 153 *En la madrugada del 5 de septiembre de 2006: Esta historia es de David Kirkpatrick, El efecto Facebook: La historia interna de la empresa que es conectar el mundo (Nueva York: Simon & Schuster, 2010).*
- 155 *grandes empresas se basan en secretos: Peter Thiel y Blake Masters, cero a uno: Notas sobre Startups, o cómo construir el futuro (Nueva York: La Corona Publishing Group, 2014).*
- 157 dice Xavier Amatriain: Entrevisté Xavier Amatriain por teléfono el 5 de mayo de 2015.

- 159 *top preguntas estadounidenses tuvieron durante 2014 de Obama: Análisis de autor de datos de Google Trends.*
- 162 esta vez en una mezquita: “El Presidente habla en la Sociedad Islámica de Baltimore,” video de YouTube, publicado 3 de febrero de, 2016,  
<https://www.youtube.com/watch?v=LRRVdVqAjdw>.
- 163 de odio, búsquedas rabiosos contra los musulmanes cayeron en las primeras horas después del discurso del presidente: Análisis de autor de datos de Google Trends. Las búsquedas de “matar a los musulmanes” eran más bajas que el mismo periodo de una semana antes. Además, las búsquedas que incluyen “musulmanes” y uno de los cinco mejores palabras negativas sobre este grupo eran más bajas.

## Capítulo 5: Uso del zoom EN

- 166 cómo la infancia experimenta influencia que el equipo de béisbol que admite: Seth Stephens-Davidowitz, “Ellos gancho en el que cuando se es joven,” New York Times, 20 de abril de 2014 SR5. Datos y el código de este estudio se pueden encontrar en mi sitio web, sethsd.com, en la sección “béisbol”.
- 170 el solo año más importante: Yair Ghitza y Andrew Gelman, “La Gran Sociedad, Revolución de Reagan, y generaciones de Presidente Votar”, manuscrito inédito.
- 173 Chetty explica: Entrevisté Raj Chetty por teléfono el 30 de julio de 2015.
- 176 escapar de la parca.: Raj Chetty et al, “La asociación entre ingresos y esperanza de vida en los Estados Unidos, 2001-2014”, JAMA 315, no. 16 (2016).
- 178 comportamiento contagioso puede ser la conducción algo de esto: Julia Belluz, “la desigualdad del ingreso está saltando lejos en la esperanza de vida de los estadounidenses,” vox.com 11 de abril, el 2016.
- 178 por qué algunas personas hacen trampa en sus impuestos: Raj Chetty, John Friedman, y Emmanuel Saez, “Uso de diferencias en conocimientos a través de barrios para descubrir los impactos de la EITC a las Ganancias”, American Economic Review 103, no. 7 (2013).
- 180 decidí descargar Wikipedia: Esto es de Seth Stephens-Davidowitz, “La geografía de la fama,” New York Times, 23 de Marzo de 2014, SR6. Los datos se pueden encontrar en mi sitio web, sethsd.com, en la sección “Wikipedia Tasa de natalidad, por el Condado.” Para obtener ayuda para la descarga y el condado de codificación de nacimiento de cada participante Wikipedia, agradezco a Noah Stephens-Davidowitz.
- 183 una gran ciudad: Para obtener más pruebas sobre el valor de las ciudades, ver a Ed Glaeser, Triunfo de la Ciudad (Nueva York: Penguin, 2011). (Glaeser fue mi asesor en la universidad.)
- 191 *muchos ejemplos de la vida real imita al arte: David Levinson., ed, Enciclopedia de Crimen y castigo (Thousand Oaks, CA: Sage, 2002).*

- 191 *los sujetos expuestos a una película violenta reportarán más ira y hostilidad*: Craig Anderson y otros, “La influencia de los medios de Violencia sobre la juventud,” *Ciencia Psicológica en el Interés Público* 4 (2003).
- 192 Los fines de semana con una película violenta populares: Gordon Dahl y Stefano DellaVigna, “no Película Violencia aumentan el crimen violento?” *Quarterly Journal of Economics* 124, no. 2 (2009).
- 195 búsquedas de Google también pueden ser descompuestos por el minuto: Seth Stephens-Davidowitz, “Días de nuestras vidas digitales”, *New York Times*, 5 de julio de 2015, SR4.
- 196 alcohol es un importante contribuyente a la delincuencia: Anna Richardson y Tracey Budd, “Los adultos jóvenes, el alcohol, la delincuencia y desórdenes,” *Comportamiento Criminal y Salud Mental* 13, no. 1 (2003); Richard A. Scribner, David P. MacKinnon, y James H. Dwyer, “el riesgo de violencia agresiva y Alcohol disponibilidad en el condado de Los Ángeles,” *American Journal of Public Health* 85, no. 3 (1995); Dennis M. Gorman, Paul W. Speer, Paul J. Gruenewald, y Erich W. Labouvie, “espaciales Dinámica de alcohol disponibilidad, estructura de vecindad y el crimen violento,” *Revista de Estudios sobre el Alcohol* 62, no. 5 (2001); Tony H. Grubestic, William Alex Pridemore, Dominique A. Williams, y Loni Philip-Tabb, “Densidad alcohol de salida y la violencia: El papel de los minoristas arriesgado y Gastos Relacionados con el Alcohol” *Alcohol y Alcoholismo* 48, no. 5 (2013).
- 196 dejándose los cuatro de sus hijos jugar al fútbol: “Ed McCaffrey sabía Cristiano McCaffrey sería bueno desde la puesta en 'The manada ',” vídeo de YouTube, Publicado 3 de diciembre de de 2015, <https://www.youtube.com/watch?v=boHMmp7DpX0>.
- 197 análisis de montones de datos: Los investigadores han descubierto más de la utilización de estos datos de criminalidad divididos en pequeños incrementos de tiempo. ¿Un ejemplo? quejas de violencia doméstica suben inmediatamente después de la selección de fútbol de una ciudad pierde un juego que se espera que gane. Ver la tarjeta de David y Gordon B. Dahl, “La violencia intrafamiliar y Fútbol: El Efecto de las señales emocionales inesperados en el comportamiento violento”, *Quarterly Journal of Economics* 126, no. 1 (2011).
- 197 Así es como Bill Simmons: Bill Simmons, “Es difícil decir adiós a David Ortiz,” *ESPN.com* 2 de junio de 2009, <http://www.espn.com/espnmag/story?id=4223584>.
- 198 *¿Cómo podemos predecir cómo un jugador de béisbol llevará a cabo en el futuro: Esto se discute en Nate Silver, la señal y el ruido: ¿Por qué tantas predicciones fallan-Pero algunos no lo hacen (Nueva York: Penguin, 2012).*
- 199 “bateadores carnosos” de hecho lo hacen, en promedio, pico temprano: Ryan Campbell, “¿Cómo Will Prince Fielder Edad” 28 de octubre de 2011 <http://www.fangraphs.com/blogs/how-will-prince-fielder-años/>.

- 199 doppelgangers de Ortiz: Esta información fue proporcionado amablemente por Rob McQuown del Folleto de béisbol.
- 204 Kohane pregunta: Entrevisté a Isaac Kohane por teléfono el 15 de junio de 2015.
- 205 James Heywood es un empresario: Entrevisté James Heywood por teléfono el 17 de agosto de 2015.

## Capítulo 6: TODO EL MUNDO ES UN LABORATORIO

- 207 27 de febrero de 2000: se discute Esta historia, entre otros lugares, en Brian Christian, “La prueba A / B: Dentro de la tecnología que está cambiando las reglas del negocio”, con conexión de cable 25 de abril, 2012, [http://www.wired.com/2012/04/ff\\_abtesting/](http://www.wired.com/2012/04/ff_abtesting/).
- Cuando se pagaron 209 profesores, ausentismo de los maestros se redujo: Esther Duflo, Rema Hanna, y Stephen P. Ryan, “incentivos de trabajo: conseguir que los profesores a venir a la escuela”, *American Economic Review* 102, no. 4 (2012).
- 209 cuando Bill Gates aprendió de la obra de Duflo: Ian Parker, “El laboratorio de la pobreza,” *New Yorker*, 17 de mayo, 2010.
- 211 *Los ingenieros de Google corrieron siete mil pruebas A / B: cristianos*, “La prueba A / B”.
- 211 cuarenta y un marginalmente diferentes tonos de azul: Douglas Bowman, “Adiós, Google,” *Stopdesign* 20 de marzo de 2009, <http://stopdesign.com/archive/2009/03/20/goodbye-google.html>.
- 211 Facebook ahora dirige: Eytan Bakshy, “Grandes Experimentos: Amigo de grandes volúmenes de datos para la toma de decisiones,” 3 de abril de 2014 [https://www.facebook.com/notes/facebook-data-science/big-experiments-big-datas - Amigo-para-hacer-decisiones / 10152160441298859 /](https://www.facebook.com/notes/facebook-data-science/big-experiments-big-datas-Amigo-para-hacer-decisiones/10152160441298859/). Las fuentes de información sobre los estudios farmacéuticos pueden encontrarse en “Cómo se inician muchos ensayos clínicos cada año?” Post Quora, <https://www.quora.com/How-many-clinical-trials-are-started-each-year>.
- 211 Optimizely: Me entrevistaron Dan Siroker por teléfono el 29 de abril de 2015.
- 214 de compensación de la campaña más o menos \$ 60 millones de dólares: Dan Siroker, “¿Cómo Obama recaudó \$ 60 millones mediante la ejecución de un experimento sencillo,” el blog Optimizely 29 de noviembre de 2010, [https://blog.optimizely.com/2010/11/29/how-obama-levantó-60-millones-por-funcionamiento-a-simple-experimento /](https://blog.optimizely.com/2010/11/29/how-obama-levantó-60-millones-por-funcionamiento-a-simple-experimento/).
- 214 Los Boston Globe A / B pruebas titulares: el Boston Globe A tests / B y los resultados se proporcionaron al autor. Algunos detalles sobre la prueba del globo se pueden encontrar en “The Boston Globe: El descubrimiento y optimización de una propuesta de valor de contenido,” Marketing Sherpa archivo de videos,

- <https://www.marketingsherpa.com/video/boston-globe-optimization-summit2>. Esto incluye una conversación grabada entre Peter Doucette del globo y Pamela Markey en MECLABS.
- 217 Benson dice: Entrevisté Clark Benson por teléfono el 23 de julio de 2015.
- 217 añade una flecha que apunta hacia la derecha rodeado de un cuadrado: “Mejora de anuncios de texto en la Red de Display de Google,” Dentro de AdSense, 3 de diciembre de 2012, <https://adsense.googleblog.com/2012/12/enhancing-text-ads-on-google-display.html>.
- 218 clientes de Google eran críticos: Véase, por ejemplo, “grandes flechas que aparecen en los anuncios de google-quita por favor,” Foro Ayuda de DoubleClick Publisher, [https://productforums.google.com/forum/#!topic/dfp/p\\_TRMqWUF9s](https://productforums.google.com/forum/#!topic/dfp/p_TRMqWUF9s).
- 219 *el aumento de las adicciones conductuales en la sociedad contemporánea: Adam Alter, irresistible: The Rise of Addictive Tecnología y el negocio de mantener conectó (Nueva York: Penguin, 2017).*
- Top 219 adicciones informó a Google: Análisis de autor de datos de Google Trends.
- 222 dice Levitt en una conferencia: Este tema se trata en un vídeo que se está presentado en la página de la Oficina Freakonomics Altavoces Harry Walker, <http://www.harrywalker.com/speakers/authors-of-freakonomics/>.
- 225 de cerveza y refrescos publicar anuncios durante el Super Bowl: Wesley R. Hartmann y Daniel Klapper, “Super Bowl anuncio”, manuscrito inédito de 2014.
- 226 un niño lleno de granos en su ropa interior: Para el caso fuerte que probablemente estamos viviendo en una simulación por ordenador, ver Nick Bostrom, “¿Estamos viviendo en una simulación por ordenador?” *Philosophical Quarterly* 53, no. 211 (2003).
- 227 de cuarenta y tres presidentes de Estados Unidos: el personal de Los Angeles Times, “Asesinatos de EE.UU. presidenciales y los intentos”, Los Angeles Times, 22 de enero del 2012, <http://timelines.latimes.com/us-presidential-assassinations-and-attempts/>.
- 227 Comparar John F. Kennedy y Ronald Reagan: Benjamin F. Jones y Benjamin A. Olken, “Haz asesinos realmente cambiar la historia?” *New York Times*, 12 de Abril de 2015, SR12.
- 227 Kadyrov murió: Un video inquietante de que el ataque se puede ver en “desfile sorpresa (Chechenia 2004),” vídeo de YouTube, publicada el 31 de marzo de 2009 <https://www.youtube.com/watch?v=fHWhs5QkfuY>.
- 227 Hitler había cambiado su horario: Esta historia también se discute en Jones y Olken “Haz asesinos realmente cambiar la historia?”
- 228 el efecto de tener a su líder asesinado: Benjamin F. Jones y Benjamin A. Olken, “impredecible? El efecto de los asesinatos en las instituciones y Guerra,” *American Economic Diario: Macroeconomía* 1, no. 2 (2009).
- 229 ganar la lotería no: Este punto se hace en John Tierney, “Cómo ganar la lotería (Felizmente),” *New York Times*, 27 de Mayo de 2014, D5. pieza de Tierney se tratan

- los siguientes estudios: Bénédicte Apouey y Andrew E. Clark, “ganar grande, pero sin sentir mejor? El efecto de los premios de la lotería sobre la salud física y mental,” *Economía de la Salud* 24, no. 5 (2015); Jonathan Gardner y Andrew J. Oswald, “Dinero y Bienestar mental: un estudio longitudinal de victorias Lotería medianas empresas,” *Revista de Economía de la Salud* 26, no. 1 (2007); y Anna Hedenus, “al final del arco iris: Post-Ganar la vida entre los ganadores de la lotería sueca”, manuscrito inédito, pieza 2011. Tierney también señala que el famoso 1978 estudio Philip Brickman, Dan Coates, y Ronnie Janoff-Bulman, “Los ganadores de la lotería y víctimas de accidentes: La felicidad es relativa?” *Journal of Personality and Social Psychology* 36, no. 8 (1978) -que encontró que ganar la lotería no te hace feliz se basó en una muestra pequeña.
- 229 a su vecino ganar la lotería: Véase Peter Kuhn, Peter Kooreman, Adriaan Soetevent, y Arie Kapteyn, “Los efectos de premios de la lotería sobre los ganadores y sus vecinos: Evidencia de la Lotería Holandesa código postal,” *American Economic Review* 101, no. 5 (2011), y Sumit Agarwal, Vyacheslav Mikhed, y Barry Scholnick, “¿Tiene desigualdad causar angustia financiera? La evidencia de ganadores de la lotería y Quiebras vecinos “, papel, 2016 trabajando.
- 229 vecinos de ganadores de la lotería: Agarwal, Mikhed y Scholnick, “la desigualdad también causar angustia financiera?”
- 230 médicos pueden estar motivados por incentivos monetarios: Jeffrey Clemens y Joshua D. Gottlieb, “¿Los incentivos financieros Médicos afectar el tratamiento médico y de salud del paciente?”, *American Economic Review* 104, no. 4 (2014). Tenga en cuenta que estos resultados no significan que los médicos son malos. De hecho, los resultados podrían ser más preocupante si los procedimientos médicos adicionales ordenados cuando se les paga más que ordenar la vida en realidad guardados. Si este fuera el caso, significaría que los médicos necesitan para ser pagado lo suficiente para ordenar los tratamientos que salvan vidas. Clemens y resultados de Gottlieb sugieren, en cambio, que los médicos ordenarán tratamientos que salvan vidas, no importa la cantidad de dinero que se les da a pedirlos. Para los procedimientos que no ayudan a casi nada, los médicos deben ser pagados suficiente para ordenarlos. Otra forma de decir esto: los médicos no prestan demasiada atención a los incentivos monetarios para la materia en peligro la vida; que pagan un montón de atención a los incentivos monetarios por cosas sin importancia.
- 231 \$ 150 millones de dólares: Robert D. McFadden y Eben Shapiro, “Finalmente, una cara para encajar Stuyvesant: una escuela secundaria de alto rendimiento Obtiene un hogar de alto precio,” *New York Times*, 8 de Septiembre., 1992
- Ofrece 231: Las ofertas de cursos están disponibles en la página web de Stuy, <http://stuy.enschool.org/index.jsp>.
- 231 una cuarta parte de sus graduados son aceptados: Anna Bahr, “Cuando las admisiones de la universidad batalla comienza a la edad de 3,” *New York Times*, 29 de julio de

- 2014, <http://www.nytimes.com/2014/07/30/upshot/cuando-la-universidad-admisión-batalla-comienza-por-edad-3.html>.
- 231 Stuyvesant entrenado: Sewell Chan, “Nueva York lazos del equipo de Obama,” *New York Times* 25 de noviembre de 2008; Evan TR Rosenman, “Clase de 1984: Lisa Randall,” *Harvard Crimson* 2 de junio de 2009; “Gary Shteyngart en Stuyvesant: Mi Nueva York,” vídeo de YouTube, publicado 4 agosto del 2010, [https://www.youtube.com/watch?v=NQ\\_phGkC-Tk](https://www.youtube.com/watch?v=NQ_phGkC-Tk); Candace Amos, “30 Estrellas que asistían a escuelas públicas de Nueva York,” *New York Daily News*, 29 de Mayo de 2015.
- 231 Sus altavoces de graduación han incluido: Carl Campanile, “Stuy niños arriba sobre Bubba: Él va a Dirección de graduación de la Escuela de tierra cero,” *New York Post*, 22 de marzo de 2002; Naciones Unidas Comunicado de prensa, “Respuesta elocuente 'Multicultural Tapiz' de Stuyvesant al odio, dice el Secretario General de la graduación de Dirección” 23, de junio de 2004; “El discurso de Conan O'Brien en la clase de Stuyvesant de 2006 Graduación en el Lincoln Center,” vídeo de YouTube, publicado mayo 6, 2012, <https://www.youtube.com/watch?v=zAMkUE9Oxnc>.
- 231 Stuy el número uno: Ver <https://k12.niche.com/rankings/public-high-schools/best-overall/>.
- 232 Menos del 5 por ciento: Pamela Wheaton, “8º grado Obtener ingreso a la Escuela Resultados,” *Insideschools* 4 de marzo de 2016, <http://insideschools.org/blog/item/1001064-8th-graders-get-high-school-admissions-resultados>.
- 235 prisioneros asignados a las condiciones más duras: M. Keith Chen y Jesse M. Shapiro, “Do más duras condiciones penitenciarias reducir la reincidencia? Un enfoque basado en la discontinuidad,” *American Law and Economics Review* 9, no. 1 (2007).
- 236 Los efectos de Stuyvesant? Atila Abdulkadiroğlu, Joshua Angrist y Parag Pathak, “La ilusión Elite: Efectos de logro en Boston y Nueva York escuelas por examen,” *Econometrica* 82, no. 1 (2014). El mismo resultado nulo se encontró de forma independiente por Will Dobbie y Roland G. Fryer Jr., “El impacto de asistir a una escuela con sus compañeros-Obtención de alta: Evidencia de las escuelas por examen ciudad de Nueva York,” *American Journal Económico: Economía Aplicada* 6, sin . 3 (2014).
- 238 graduado medio de Harvard hace: Ver <http://www.payscale.com/college-salary-report/bachelors>.
- 238 estudiantes aceptados similares a las escuelas de manera similar prestigiosos que deseen asistir diferentes escuelas terminan en aproximadamente el mismo lugar: Stacy Berg Dale y Alan B. Krueger, “Estimación de la recompensa de asistir a una universidad más selectiva: una aplicación de selección en observables y no observables”, *Quarterly Journal of Economics* 117, no. 4 (2002).



239 *Warren Buffett: Alice Schroeder, La bola de nieve: Warren Buffett y el negocio de la Vida (New York: Bantam, 2008).*

## **CAPÍTULO 7: DATOS DE GRANDE, GRANDE SCHMATA? Lo que no puede HACER**

- 247 afirmaron que podían predecir en qué dirección: Johan Bollen, Huina Mao y Xiaojun Zeng, “Mood Twitter predice que el mercado de valores”, *Revista de Ciencias de la Computación* 2, no. 1 (2011).
- 248 El fondo de cobertura de tuit fue cerrado: James Mackintosh, “fondo de cobertura que negocian en base a señales de medios sociales no funcionó,” *Financial Times* 25 de mayo de 2012.
- 250 no pudieron reproducir la correlación.: Christopher F. Chabris y otros, “asociaciones genéticas más reportadas con la inteligencia general son probablemente falsos positivos”, *Psychological Science* (2012).
- 252 Zoë Oportunidad: Esta historia se discute en TEDx conversaciones, “Cómo hacer un comportamiento adictivo: Zoë oportunidad en TEDx Mill River,” vídeo de YouTube, publicada el 14 de mayo de 2013, <https://www.youtube.com/watch?v=AHfiKav9fcQ>. Algunos detalles de la historia, como el color del contador de pasos, se precisiones en las entrevistas. Entrevisté Probabilidad por teléfono el 20 de abril de 2015, y por correo electrónico el 11 de julio de 2016, y 8 de septiembre, el 2016.
- 253 Los números pueden ser seductor: Esta sección es de Alex Peysakhovich y Seth Stephens-Davidowitz, “¿Cómo no ahogarse en Números”, *New York Times*, 3 de mayo de 2015, SR6.
- 254 engañado de plano en la administración de esas pruebas: Brian A. Jacob y Steven D. Levitt, “manzanas podridas: Una investigación de la prevalencia y predictores de Maestro de engaño,” *Quarterly Journal of Economics* 118, no. 3 (2003).
- 255, dice Thomas Kane: Entrevisté Thomas Kane por teléfono el 22 de abril de 2015.
- 256 “Cada medida añade algo de valor”: Bill y Melinda Gates, “Garantizar la Feria y medidas fiables de enseñanza eficaz,” [http://k12education.gatesfoundation.org/wp-content/uploads/2015/05/MET\\_Ensuring\\_Fair\\_and\\_Reliable\\_Measures\\_Practitioner\\_Brief.pdf](http://k12education.gatesfoundation.org/wp-content/uploads/2015/05/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf) .

## **Capítulo 8: MO DE DATOS, MO PROBLEMAS? LO QUE NO debe hacer**

- 257 Recientemente, tres economistas: “Las señales de identificación para Préstamo por Defecto en el texto de las solicitudes de crédito Cuando las palabras” *Sweat* 2016 Oded Netzer, Alain Lemaire, y Michal, Herzenstein.

- 258 alrededor de un 13 por ciento de los prestatarios: Peter Renton, “Otro Análisis de Incumplimiento de tarifas en Club de préstamos y prosperar”, 25 de octubre de 2012 <http://www.lendacademy.com/lending-club-prosper-default-rates/>.
- 261 talla de Facebook están correlacionados con frecuencia: Michal Kosinski, David Stillwell, y Thore Graepel, “Rasgos privados y atributos son predecibles a partir de registros digitales de Comportamiento Humano”, PNAS 110, núm. 15 (2013).
- 265 empresas están a merced de los comentarios de Yelp Michael: Luca, “Comentarios, reputación y los ingresos: el caso de tu barrio”, manuscrito no publicado, 2011.
- 266 búsquedas de Google relacionadas con el suicidio: Christine Ma-Kellams, la flora o, Ji Hyun Baek, e Ichiro Kawachi, “Repensar la vigilancia de suicidio: Google búsqueda de datos y las tendencias suicidas diferencialmente estimar el riesgo de Auto Reporte de suicidio consumado,” Ciencia Psicológica Clínica 4, no. 3 (2016).
- 267 3,5 millones de búsquedas de Google: Este utiliza una metodología discutida en mi sitio web en el apartado sobre aborto autoinducido. Comparo búsquedas en Google la categoría de “suicidio” de las búsquedas de “cómo atar un lazo.” Había 6,6 millones de búsquedas en Google de “cómo atar un lazo” en 2015. No eran 6,5 veces más búsquedas en la categoría de suicidio.  $6,5 * 6,6 / 12 \approx 3.5$ .
- 268 12 asesinatos de musulmanes reportados como crímenes de odio: Bridge Initiative equipo, “Cuando la islamofobia vuelve violento: Los Estados Unidos Elección Presidencial 2016,” 2 may, 2016, disponible en <http://bridge.georgetown.edu/when-islamophobia-turns-violent-al-2016-us-presidential-elections/>.

## CONCLUSIÓN

- 272 *Lo cruzado motivado de Popper?: Karl Popper, Conjeturas y refutaciones (Londres: Routledge and Kegan Paul, 1963).*
- 275 asignan todos los casos de cólera en la ciudad: Simon Rogers, “Datos de Periodismo, John Snow: El mapa cólera que cambió el mundo”, The Guardian, 15 de Marzo de 2013.
- 276 Benjamin F. Jones: Jones entrevistó a Benjamin por teléfono el 1 de junio de 2015. Este trabajo también se discute en Aaron Chatterji y Benjamin Jones, “Aprovechamiento de la tecnología para mejorar la educación K-12,” Proyecto Hamilton Documento de Debate, 2012.
- 283 personas tienden a no terminar tratados por los economistas: Jordan Ellenberg, “La mayoría no leído libro de que el verano es. . .” Wall Street Journal 3 de julio de 2014.

# ÍNDICE

---

La paginación de esta edición electrónica no coincide con la edición de la que fue creada. Para localizar una entrada específica, por favor utilice herramientas de búsqueda de su lector de libros electrónicos.

## A / B pruebas

- ABC de 209-21,
- y adicciones, 219-20
- y los titulares Boston Globe, 214-17
- en el mundo digital, 210-19
- a la baja, 219-21
- y la educación / aprendizaje, 276
- y Facebook, 211
- usos futuros de, 276, 277, 278
- y la industria del juego, 220-21
- Google y la publicidad, 217-19
- importancia de, 214, 217
- y la mandíbula, 277
- y la política, 211-14
- y la televisión, 222
- Abdulkadiroglu, Atila, 235-36
- aborto, verdad acerca, 147-50
- Adamic, Lada, 144
- Adams, John, de 78 años
- adicciones
  - y las pruebas A / B, 219-20
  - Ver también la adicción específica*
- publicidad
  - y las pruebas A / B, 217-19
  - efectos causales de, 221-25, 273
  - y los ejemplos de grandes volúmenes de datos búsquedas, 22
  - Google, 217-19
  - y la compañía Levitt-electrónica, 222, 225, 226
  - y películas, 224-25
  - y la ciencia, 273

y juegos de Super Bowl, 221-26  
TV, 221-26  
afroamericano  
y Harvard carmesí editorial sobre Zuckerberg, 155  
ingresos y, 175  
y los orígenes de americanos notables, 182-83  
y la verdad sobre el odio y el prejuicio, 129, 134  
*Ver también* "negro"; raza / racismo  
años  
y los aficionados al béisbol, 165-69, 165-66n  
y la mentira, 108n  
y los orígenes de las preferencias políticas, 169-71  
y predecir el futuro de los jugadores de béisbol, 198-99  
de los miembros del frente de tormenta, 137-38  
y palabras como datos, 85-86  
*Ver también* niños; adolescentes  
Aiden, Erez, 76-77, 78-79  
alcohol  
como la adicción, 219  
y la salud, 207-8  
AltaVista (motor de búsqueda), 60  
Alter, Adam, 219-20  
Amatriain, Xavier, 157  
Amazon, 20, 203, 283  
Pharoah Americana (Horse No. 85), 22, 64, 65, 70-71, 256  
Angrist, Joshua, 235-36  
antisemitismo. ver Judíos  
ansiedad  
datos sobre, 18  
y la verdad sobre el sexo, 123  
AOL, y la verdad sobre el sexo, 117-18  
AOL News, 143  
arte, la vida real como la imitación, 190-97  
Ashenfelter, Orley, 72-74  
Asher, Sam, 202  
Asiáticos, y la verdad sobre el odio y el prejuicio, 129  
hacer las preguntas correctas, 21-22  
asesinatos, 227-28  
*atlántico* revista, 150-51, 152, 202  
Australia, el embarazo en, 189

auto-completar, 110-11, 116  
*Avatar* (Película), 221-22

Bakshy, Eytan, 144

Baltimore Ravens y New England Patriots juegos, 221, 222-24  
béisbol

y la influencia de las experiencias infantiles, 165-69, 165-66n, 171, 206

y demasiado énfasis en la medición, la 254-55

predecir el futuro de un jugador en, 197-200, 200n, 203

y la ciencia, 273

exploración para, 254-55

zoom sobre, 165-69, 165-66n, 171, 197-200, 200n, 203

baloncesto

genealogías y, 67

predecir el éxito en, 33-41, 67

y antecedentes socioeconómicos, 34-41

Beane, Billy, 255

Beethoven, Ludwig von, hacer zoom sobre, 190-91

ciencias de la conducta, y la revolución digital, 276, 279

Belushi, John, 185

Benson, Clark, 217

Berger, Jonás, 91-92

Bezos, Jeff, 203

parcialidad

implícita, 134

lenguaje como clave para la comprensión, 74-76

omisión de variables, 208

subconsciente, 132

*Ver también* odio; perjudicar; raza / racismo

big Data

y la cantidad de información, 15, 21, 59, 171

y hacer las preguntas correctas, 21-22

y experimentos de causalidad, 54, 240

definición de, 14, 15

y dimensionalidad, 246-52

y ejemplos de búsquedas, 15-16

y la expansión de la metodología de investigación, 275-76

y libros de acabado, 283-84

futuro de, 279

Google busca en la fuente dominante de que, el 60

honestidad de, 53-54  
importancia / valor de, 17-18, 29-33, 59, 240, 265, 283  
limitaciones de, 20, 245, 254-55, 256  
poderes de, 15, 17, 22, 53-54, 59, 109, 171, 211, 257  
y predecir lo que la gente va a hacer en el futuro, 198-200  
como revolucionario, 17, 18-22, 30, 62, 76, 256, 274  
datos como correctas, 62  
escépticos de 17  
y pequeños datos, 255-56  
subconjuntos en, 54  
comprensión de, 27-28  
*Ver también tema específico*  
Bill y Melinda Gates, 255  
*Billings* (Montana) Gaceta, y palabras como datos, 95  
Bing (motor de búsqueda), y el estudio de cáncer de páncreas Universidad de Columbia-  
Microsoft, 28, 30  
Negro, Don, 137  
Vidas negro Materia, 12  
*Parpadeo* (Gladwell), 29-30  
Bloodstock, Incardo, 64  
cuerpos, como datos, 62-74  
Boehner, Juan, 160  
Booking.com, 265  
libros  
conclusiones a, 271-72, 279, 280-84  
digitalización, 77, 79  
número de personas que terminan, 283-84  
préstamo de dinero, 257-61  
Bosh, Chris, de 37 años  
*Boston Globe*, y las pruebas A / B, 214-17  
Maratón de Boston (2013), 19  
Boston Red Sox, 197-200  
cerebro, el estudio de Minsky, 273  
Brasil, el embarazo en, 190  
pechos, y la verdad sobre el sexo, 125, 126  
Brin, Sergey, 60, 61, 62, 103  
Gran Bretaña, el embarazo en, 189  
Ciencias del Bronx High School (Nueva York), 232, 237  
Buffett, Warren, 239  
Bullock, Sandra, 185

Bundy, Ted, 181  
Bush, George W., de 67 años  
negocio  
y la comparación de compras, 265  
opiniones de, 265  
*Ver también* corporaciones  
a tope, y la verdad sobre el sexo, 125-26

Calhoun, Jim, 39

Universidad de Cambridge, y el estudio de Microsoft acerca de CI de los usuarios de  
Facebook, 261  
cáncer, pancreática predecir, 28-29, 30  
*Capital en el siglo 21 (Piketty)*, 283  
casinos, y la discriminación de precios, 263-65  
causalidad  
Test A / B y, 209-21  
y la publicidad, 221-25  
y experimentos Big Data, 54, 240  
la universidad y, 237-39  
correlación distinguirse de, 221-25  
y la ética, 226  
y las ganancias inesperadas monetarios, 229  
experimentos naturales y, 226-28  
y el poder de Big Data, 54, 211  
y experimentos controlados aleatorios, 208-9  
revertir, 208  
y el estudio Stuyvesant High School, 231-37, 240  
Centros para el Control y Prevención de Enfermedades, 57  
Chabris, Christopher, 250  
Oportunidad, Zoë, 252-53  
Chaplin, Charlie, 19  
donaciones de caridad, 106, 109  
Chen, M. Keith, 235  
Chetty, Raj, 172-73, 174-75, 176, 177, 178-80, 185, 273  
niños  
abuso de, 145-47, 149-50, 161  
y los beneficios del suero de la verdad digitales, 161  
y la pornografía infantil, 121  
decisiones acerca de tener, 111-12  
datos de altura y peso aproximadamente, 204-5

de inmigrantes, 184-85  
y la distribución del ingreso, 176  
y la influencia de las experiencias infantiles, 165-71, 165-66n, 206  
la inteligencia de, 135  
y los orígenes de americanos notables, 184-85  
prejuicios contra los padres, 134-36, 135N  
apariencia física de, 135-36  
*Ver también* los padres / crianza de los hijos; adolescentes  
el cólera, el estudio sobre la nieve, 275  
Cristianos, y la verdad sobre el odio y el prejuicio, 129  
Churchill, Winston, 169  
economía de cigarrillos, Filipinas, 102  
ciudades  
y el peligro de gobierno facultado, 267, 268-69  
predecir el comportamiento de, 268-69  
zoom sobre, 172-90, 239-40  
Guerra Civil, 79  
Clemens, Jeffrey, 230  
Clinton, Bill, busca, 60-62  
Clinton, Hillary. Ver las elecciones de 2016  
*La naranja mecánica (película)*, 190-91  
cnn.com, 143, 145  
Cohen, Leonard, 82n  
Universidad  
y la causalidad, 237-39  
y los ejemplos de grandes volúmenes de datos búsquedas, 22  
ciudades de la universidad, y los orígenes de americanos notables, 182-83, 184, 186  
*Colores (Película)*, 191  
La Universidad de Columbia, Microsoft estudio de páncreas y cáncer, 28-29, 30  
comparación de compras, 265  
conclusiones  
beneficios de grande, 281-84  
a los libros, 271-72, 279, 280-84  
características de mejor, 272, 274-79  
importancia de, 283  
como se señala manera a más cosas por venir, 274-79  
propósito de, 279-80  
la escritura de, 271-72, 281-84 Stephens-Davidowitz de  
preservativos, 5, 122  
*Registro del Congreso*, y la investigación Gentzkow-Shapiro, 93



- conservadores
- y los orígenes de las preferencias políticas, 169-71
- y los padres prejuicios contra los niños, 136
- y la verdad acerca de la Internet, 140, 141-44, 145
- y palabras como datos, 75-76, 93, 95-96
- los consumidores. Ver clientes / consumidores
- comportamiento contagioso, 178
- conversación, pareja, 80-82
- corporaciones
- golpes contra los consumidores, 265
- peligro de poder, 257-65
- opiniones de, 265
- correlaciones
- causalidad distinguirse de, 221-25
- y predecir el mercado de valores, 245-48, 251-52
- condados, hacer zoom sobre, 172-90, 239-40
- Country Music Radio, 202
- Craigslist, 117
- la creatividad y la comprensión del mundo, 280, 281
- crimen
- alcohol como contribuyente a, 196
- y el peligro de gobierno facultado, 266-70
- y las condiciones de reclusión, 235
- y las películas violentas, 193, 194-95, 273
- Cundiff, Billy, 223
- curiosidad
- y los beneficios del suero de la verdad digitales, 162, 163
- Levitt puntos de vista sobre, 280
- sobre el número de personas que terminan libros, 283-84
- y entender el mundo, 280, 281
- maldición, y palabras como datos, 83-85
- clientes / consumidores
- golpes contra las empresas por, 265
- y la discriminación de precios, 265
- verdad sobre, 153-57
- Cutler, David, 178

Dahl, Gordon, 191-93, 194-96, 196-97n, 197

Dale, Stacy, 238

Dallas, Texas, la conferencia “grandes y complejos conjuntos de datos” (1977) en, 20-21

datos  
cantidad / tamaño de, 15, 20-21, 30-31, 53, 171  
beneficios de la expansión de, 16  
cuerpos como, 62-74  
la recogida de la derecha, 62  
gobierno, 149-50, 266-70  
importancia de, 26  
A nivel individual, 266-70  
como intimidante, 26  
Levitt puntos de vista sobre, 280  
como generador de dinero, 103  
fuentes no tradicionales de, 74  
imágenes como, 97-102, 103  
reinención de lo que califica como, 55-103  
fuentes de, 14, 15  
la velocidad de transmisión, 55-59  
y entender el mundo, 280  
lo que cuenta como, 74  
palabras como, 74-97  
*Ver también* Big Data; Ciencia de los datos; pequeña de datos; datos específicos  
Ciencia de los datos  
como el cambio de vista del mundo, 34  
y los resultados contrarios a la intuición, 37-38  
economistas papel en el desarrollo de, 228  
futuro de, 281  
objetivo de, 37-38  
tan intuitivo, 26-33  
y que es un científico de datos, 27  
citas  
y los ejemplos de grandes volúmenes de datos búsquedas, 22  
apariencia física y, 82, 120n  
y el rechazo, 120n  
y miembros del frente de tormenta, 138-39  
y la verdad sobre el odio y el prejuicio, 138-39  
y la verdad sobre el sexo, 120n  
y palabras como datos, 80-86, 103  
*Amanecer de los muertos (película)*, 192  
la muerte, y las historias memorables, 33  
DellaVigna, Stefano, 191-93, 194-96, 196-97n, 197  
demócratas

principios básicos de 94,  
y los orígenes de las preferencias políticas, 170-71  
y palabras como datos, 93-97  
*Ver persona o elección también específica*  
depresión  
Google busque por, 31, 110  
y gastos de la verdad, 158  
y la mentira, 109, 110  
y los padres prejuicios contra los niños, 136  
países en desarrollo  
economías de, 101-2, 103  
invertir en, 251  
suero de la verdad digitales  
aborto y, 147-50  
y el abuso infantil, 145-47, 149-50  
y clientes, 153-57  
y amigos de Facebook, 150-53  
y gastos de la verdad, 158-63  
y el odio y el prejuicio, 128-40  
y haciendo caso omiso de lo que la gente te dice, 153-57  
incentivos y, 109  
e internet, 140-45  
sexo y, 112-28  
sitios como, 54  
*Ver también*acostado; verdad  
mundo digital, en experimentos aleatorios, 210-19  
dimensionalidad, maldición de, 246-52  
discriminación  
y los orígenes de americanos notables, 182-83  
precio, 262-65  
*Ver también*parcialidad; perjudicar; raza / racismo  
ADN, 248-50  
Dna88 (miembro Stormfront), 138  
los médicos, los incentivos financieros para, 230, 240  
Donato, Adriana, 266, 269  
doppelgangers  
beneficios de, 263  
y la salud, 203-5  
y la caza en las redes sociales, 201-3  
y predecir el futuro de los jugadores de béisbol, 197-200, 200n, 203

y la discriminación de precios, 262-63, 264  
zoom sobre, 197-205  
sueños, símbolos fálicos en, 46-48  
drogas, como la adicción, 219  
Duflo, Esther, 208-9, 210, 273

## Crédito por Ingreso, 178, 179

economistas  
y el número de personas que terminan los libros, 283  
papel en el desarrollo de la ciencia de datos, 228  
científicos como blandas, 273  
*Ver persona también específica*  
Economía / Economía  
complejidad de, 273  
países de desarrollo, 101-2, 103  
de la economía del cigarrillo Filipinas, 102  
e imágenes como datos, 99-102  
y la velocidad de datos, 56-57  
y la verdad sobre el odio y el prejuicio, 139  
*Ver también los economistas; tema específico*  
Edmonton, el consumo de agua en, 206  
EDU STAR, 276  
educación  
y las pruebas A / B, 276  
y la revolución digital, 279  
y énfasis excesivo en la medición, 253-54, 255-56  
en la India rural, 209, 210  
datos pequeñas en, 255-56  
el gasto del estado en 185  
y el uso de comportamiento en línea como complemento a las pruebas, 278  
*Ver también los estudiantes de la escuela secundaria; tests / pruebas*  
Eisenhower, Dwight D., 170-71  
elecciones  
y el orden de las búsquedas, 10-11  
predicciones sobre, 9-14  
participación electoral en, 9-10  
elecciones, 2008  
y las pruebas A / B, 211-12  
racismo en, 2, 6-7, 12, 133, 134  
y la pertenencia Stormfront, 139

elecciones, 2012  
y las pruebas A / B, 211-12  
predicciones sobre, 10  
racismo en, 2-3, 8, 133, 134  
Trump y, 7  
elecciones, 2016  
y la mentira, 107  
mapeo de, 12-13  
encuestas sobre, 1  
la predicción del resultado de, 10-14  
y el racismo, 8, 11, 12, 14, 133  
primarias republicanas para, 1, 13-14, 133  
y la pertenencia Stormfront, 139  
participación electoral en 11  
empresa de electrónica, y la publicidad, 222, 225, 226  
“Ilusión Elite” (Abdulkadiroglu, Angrist y Pathak), 236  
Ellenberg, Jordan, 283  
Ellerbee, William, 34  
Eng, Jessica, 236-37  
medio ambiente, y la esperanza de vida, 177  
EPCOR empresa de servicio público, 193, 194  
JCA, 63-64  
igualdad de oportunidades, hacer zoom sobre, 173-75  
Bot error, 48-49  
ética  
y Big Data, 257-65  
y el peligro de gobierno facultado, 267  
búsquedas y Doppelganger, 262-63  
las empresas habilitadas y, 257-65  
y experimentos, 226  
las prácticas de contratación y, 261-62  
y IQDNA resultados del estudio, 249  
y el reembolso de los préstamos, 257-61  
y la discriminación de precios, 262-65  
y el estudio de CI de los usuarios de Facebook, 261  
Ewing, Patrick, 33  
experimentos  
y la ética, 226  
y la ciencia real, 272-73  
*Véase también el tipo de experimento o experimento específico*

## Facebook

- y las pruebas A / B, 211
- y adicciones, 219, 220
- y las prácticas de contratación, 261
- y haciendo caso omiso de lo que la gente te dice, 153-55, 157
- y la influencia de la infancia experimenta datos, 166-68, 171
- CI de los usuarios de, 261
- estudio de la Universidad de Cambridge-Microsoft de los usuarios de, 261
- “News Feed” de, 153-55, 255
- y énfasis excesivo en la mensurabilidad, 254, 255
- e imágenes como datos, 99
- y “secretos acerca de las personas,” 155-56
- y el tamaño de grandes volúmenes de datos, 20
- y pequeñas de datos, 255
- como fuente de información, 14, 32
- y la verdad acerca de los clientes, 153-55
- verdad sobre amigos en, 150-53
- y la verdad sobre el sexo, 113-14, 116
- y la verdad acerca de la Internet, 144, 145
- y palabras como datos, 83, 85, 87-88
- El efecto Facebook: La historia interna de la empresa que es conectar el mundo*  
(Kirkpatrick), 154
- Facemash, 156
- caras
- negro, 133
- e imágenes como datos, 98-99
- y la verdad sobre el odio y el prejuicio, 133
- Farook, Rizwan, 129-30
- la publicidad del día de padre, 222, 225
- 50 tonos de gris*, 157
- incentivos financieros, para los médicos, 230, 240
- Primera Ley de la viticultura, 73-74
- comida
- y símbolos fálicos en los sueños, 46-48
- predicciones sobre, 71-72
- y el embarazo, 189-90
- fútbol
- y la publicidad, 221-25
- zoom sobre, 196-97n
- Freakonomics* (Levitt), 265, 280, 281

Freud, Sigmund, 22, 45-52, 272, 281  
Friedman, Jerry, 20, 21  
Freidora, Roland, 36

Gabriel, Stuart, 9-10, 11  
Gallup encuestas, 2, 88, 113  
los juegos de azar / industria del juego, 220-21, 263-65  
“Gangnam Style” de vídeo, Psy, 152  
Garland, Judy, 114, 114n  
Gates, Bill, 209, 238-39  
homosexuales  
en el armario, 114-15, 116, 117, 118-19, 161  
y dimensiones de la sexualidad, 279  
y los ejemplos de grandes volúmenes de datos búsquedas, 22  
y gastos de la verdad, 159, 161  
en Irán, 119  
y el matrimonio, 74-76, 93, 115-16, 117  
movilidad de, 113-14, 115  
población de, 115, 116, 240  
y la pornografía, 114-15, 114n, 116, 117, 119  
en Rusia, 119  
estereotipo de, 114n  
encuestas sobre, 113  
adolescentes como, 114, 116  
y la verdad sobre el odio y el prejuicio, 129  
y la verdad sobre el sexo, 112-19  
y esposas sospechas de maridos, 116-17  
las mujeres como, 116  
y palabras como datos, 74-76, 93  
Gelles, Richard, 145  
Gelman, Andrew, 169-70  
género  
y la esperanza de vida, 176  
y los padres prejuicios contra los niños, 134-36, 135N  
de los miembros del frente de tormenta, 137  
*Ver también los homosexuales*  
Encuesta Social General, 5, 142  
genética, y IQ, 249-50  
genitales  
y la verdad sobre el sexo, 126-27

*Ver también* pene; vagina  
Gentzkow, Matt, 74-76, 93-97, 141-44  
geografía  
zoom por, 172-90  
*Ver también* ciudades; condados  
Alemania, el embarazo en, 190  
Ghana, el embarazo en, 188  
Ghitza, Yair, 169-70  
Ginsberg, Jeremy, 57  
amigas, matanza, 266, 269  
las niñas, los padres prejuicio contra los jóvenes, 134-36  
Gladwell, Malcolm, 29-30  
Gnau, Scott, 264  
oro, precio de, 252  
*el Jilguero* (Tartt), 283  
Goldman Sachs, 55-56, 59  
google  
anuncios sobre, 217-19  
y la cantidad de datos, 21  
y libros de digitalización, 77  
Campus de Mountain View de, 59-60, 207  
*Ver también tema específico*  
Google AdWords, 3n, 115, 125  
Google se correlacionan, 57-58  
Google Flu, 57, 57N, 71  
Google N-gramas, 76-77, 78, 79  
búsquedas de Google  
ventajas de usar, 60-62  
auto-completar en, 110-11  
diferenciación con respecto a otros motores de búsqueda de, 60-62  
como suero de la verdad digital, 109, 110-11  
como fuente dominante de grandes volúmenes de datos, 60  
y lo prohibido, 51  
fundación de, 60-62  
y pensamientos ocultos, 110-12  
y la honestidad / plausibilidad de los datos, 9, 53-54  
importancia / valor de, 14, 21  
encuestas en comparación con el 9  
popularidad, 62  
poder de, 4-5, 53-54



y la velocidad de datos, 57-58  
y palabras como datos, 76, 88  
*Ver también grandes volúmenes de datos; búsqueda específica*  
Google ETS, 71  
Google Trends, 3-4, 3n, 6, 246  
Gottlieb, Joshua, 202, 230  
gobierno  
  peligro de poder, 266-70  
  y predecir las acciones de los individuos, 266-70  
  y las cuestiones de privacidad, 267-70  
pasar por, 93, 94  
y la confianza de los datos, 149-50  
y palabras como datos, 93, 94  
“Gran cuerpo, gran sexo, gran mamada” (vídeo), 152, 153  
Gran Recesión y el abuso infantil, 145-47  
El mono verde (caballo N° 153), 68  
El producto interno bruto (PIB), y las imágenes como datos, 100-101  
Felicidad Nacional Bruta, 87, 88  
Instituto Guttmacher, 148, 149

*Hannibal* (Película), 192, 195

  felicidad  
  e imágenes como datos, 99  
  *Ver también* análisis de los sentimientos  
  Casino Harrah, 264  
  Harris, Tristan, 219-20  
  *Harry Potter y las reliquias de la muerte* (Rowling), 88-89, 91  
  Hartmann, Wesley R., 225  
  *Harvard Crimson*, editorial sobre Zuckerberg en, 155  
  La Universidad de Harvard, el ingreso de los graduados de, 237-39  
  odio  
  y el peligro de los gobiernos habilitados, 266-67, 268-69  
  verdad sobre, 128-40, 162-63  
  *Ver también* perjudicar; raza / racismo  
  salud  
  y alcohol, 207-8  
  y comparación de los motores de búsqueda, 71  
  y la revolución digital, 275-76, 279  
  y ADN, 248-49  
  y doppelgangers, 203-5

metodología para el estudio de, 275-76  
y la velocidad de transmisión de datos, 57  
zoom sobre, 203-5, 275  
*Ver también* esperanza de vida  
seguro de salud, 177  
Henderson, J. Vernon, 99-101  
*El Herd con Colin Cowherd*, McCaffrey entrevista en, 196N  
Herzenstein, Michal, 257-61  
Heywood, James, 205  
estudiantes de la escuela  
las pruebas de, 231-37, 253-54  
y la verdad sobre el sexo, 114, 116  
anuarios de la escuela secundaria, 98-99  
prácticas de contratación, 261-62  
Hispanos y Harvard carmesí editorial sobre Zuckerberg, 155  
Hitler, Adolf, 227  
partido de hockey, Olímpico (2010), 193, 194  
Caballo N° 85. Véase estadounidense Pharoah  
Caballo N° 153 (The Green Monkey), 68  
caballos  
y el síndrome de Bartleby, 66  
y los ejemplos de grandes volúmenes de datos búsquedas, 22  
órganos internos, 69-71  
pedigrís de, 66-67, 69, 71  
predecir el éxito de, 62-74, 256  
búsquedas sobre, 62-74  
horas, hacer zoom sobre, 190-97  
la vivienda, el precio de, 58  
Proyecto Genoma Humano, 248-49  
Campaña de Derechos Humanos, 161  
la humanidad, los datos como medio para la comprensión, 16  
Humor / chistes, las búsquedas de, 18-19  
El huracán Frances, 71-72  
El huracán Katrina, 132  
maridos  
esposas descripciones de, 160-61, 160-61n  
esposas y las sospechas sobre la homosexualidad, 116-17  
Hussein, Saddam, 93, 94

haciendo caso omiso de lo que la gente te dice, 153-57

inmigrantes, y los orígenes de americanos notables, 184, 186  
prueba de asociación implícita, 132-34  
incentivos, 108, 109  
incesto, 50-52, 54, 121  
la distribución del ingreso, 174-78, 185  
India  
la educación en las zonas rurales, 209, 210  
el embarazo en, 187, 188-89  
y búsquedas sexo / pornografía, 19  
Universidad de Indiana, y el estudio de la dimensionalidad, 247-48  
individuos, la predicción de las acciones de 266-70,  
influenza, los datos acerca de, 57, 71  
información. Ver datos grandes; datos; pequeña de datos; de origen o de búsqueda específica  
Instagram, 99, 151-52, 261  
Servicio de Impuestos Internos (IRS), 172, 178-80. Ver también los impuestos  
Internet  
como la adicción, 219-20  
en el comportamiento de navegación, 141-44  
como dominada por el carbón, 151  
segregación en, 141-44  
verdad sobre el, 140-45  
*Ver también* Pruebas A / B; medios de comunicación social; sitio específico  
intuición  
y las pruebas A / B, 214  
y los resultados contrarios a la intuición, 37-38  
la ciencia de datos como, 26-33  
y la dramática, 33  
como algo malo, 31, 32-33  
IQ / inteligencia  
y ADN, 249-50  
de los usuarios de Facebook, 261  
y los padres prejuicios contra los niños, 135  
Irán, los homosexuales en, 119  
Guerra de Irak, 94  
*Irresistible* (Alter), 219-20  
islamofobia  
y el peligro de los gobiernos habilitados, 266-67, 268-69  
*Ver también* musulmanes  
las escuelas de la Ivy League

El ingreso de los graduados de, 237-39  
*Ver también la escuela específica*

Jacob, Brian, 254

James, Bill, 198-99

James, LeBron, 34, 37, 41, 67

Quijada, 277

Judios, 129, 138

Ji Hyun Baek, 266

Jobs, Steve, 185

Johnson, Earvin III, 67

Johnson, Lyndon B., 170, 171

Johnson, "Magic", 67

chistes

pareja, 80-81

y la mentira, 109

negro, 6, 15, 132, 133, 134

y la verdad sobre el odio y el prejuicio, 132, 133, 134

Jones, Benjamin F., 227, 228, 276

Jordan, Jeffrey, 67

Jordan, Marcus, 67

Jordan, Michael, 40-41, 67

Jurafsky, Dan, 80

Kadyrov, Akhmad, 227

Kahneman, Daniel, 283

Kane, Thomas, 255

Katz, Lawrence, 243

Kaufmann, Sarah, 236-37

Kawachi, Ichiro, 266

Kayak (sitio web), 265

Kennedy, John F., 170, 171, 227

Kerry, John, 8, 244

*El rey Juan* (Shakespeare), 89-90

King, Martin Luther Jr., 132

King, William Lyon Mackenzie (alias), 138-39

Kinsey, Alfred, 113

Kirkpatrick, David, 154

Klapper, Daniel, 225

Knight, Phil, 157

Kodak, y las imágenes como datos, 99  
Kohane, Isaac, 203-5  
Krueger, Alan B., 56, 238  
Ku Klux Klan, 12, 137  
Kubrick, Stanley, 190-91  
Kundera, Milan, 233

## idioma

y la revolución digital, 274, 279  
énfasis en, 94  
como clave para entender el sesgo, 74-76  
y el reembolso de los préstamos, 259-60  
y los métodos tradicionales de investigación, 274  
y Estados Unidos como unida o dividida, 78-79  
*Ver también las palabras*  
aprendizaje. ver la educación  
Lemaire, Alain, 257-61  
Levitt, Steven, 36, 222, 254, 280, 281. Ver también Freakonomics  
liberales  
y los orígenes de las preferencias políticas, 169-71  
y los padres prejuicios contra los niños, 136  
y la verdad acerca de la Internet, 140, 141-45  
y palabras como datos, 75-76, 93, 95-96  
tarjetas de la biblioteca, y mentiras, 106  
la vida, el arte como imitación, 190-97  
la esperanza de vida, 176-78  
Linden, Greg, 203  
escuchar, y que data, 82n  
préstamos, pagar, 257-61  
*Los Angeles Times*, y el habla Obama sobre el terrorismo, 130  
loterías, 229, 229n  
Luca, Michael, 265  
Lycos (motor de búsqueda), 60  
acostado  
y la edad, 108n  
e incentivos, 108  
y chistes, 109  
a nosotros mismos, 107-8, 109  
y las encuestas, 107  
y la pornografía, 110

prevalencia de, 21, 105-12, 239  
y el racismo, 109  
razones para, 106, 107, 108, 108N  
y los datos de creación de imágenes, 103  
y la búsqueda de información, 5-6, 12  
y el sexo, 112-28  
por Stephens-Davidowitz, 282n  
y las encuestas, 105-7, 108, 108n  
y los impuestos, 180  
y el comportamiento de votación, 106, 107, 109-10  
“Blanco”, 107  
*Ver también* suero de la verdad digital; verdad; tema específico

Ma-Kellams, Christine, 266

Condado de Macon, Alabama, con éxito notable americanos / a, 183, 186-87

Malik, Tashfeen, 129-30

La Universidad de Manchester, y el estudio de la dimensionalidad, 247-48

Instituto de Tecnología de Massachusetts, Panteón de proyecto, 184-85

Matthews, Dylan, 202-3

McCaffrey, Ed, 196-97n

McFarland, Daniel, 80

McPherson, James, de 79 años

mensurabilidad, énfasis excesivo en, 252-56

“Midiendo el crecimiento económico del espacio exterior” (Henderson, Storygard, y Weil),  
99-101

medios de comunicación

sesgo de, 22, 74-77, 93-97, 102-3

y los ejemplos de grandes volúmenes de datos búsquedas, 22

propietarios de, 96

y la verdad sobre el odio y el prejuicio, 130, 131

y la verdad acerca de la Internet, 143

y palabras como datos, 74-77, 93-97

*Ver también la organización específica*

Medicare y los médicos reembolsos, 230, 240

medicina. Ver los médicos; salud

Messing, Solomon, 144

MetaCrawler (motor de búsqueda), 60

Mexicanos, y la verdad sobre el odio y el prejuicio, 129

Michel, Jean-Baptiste, 76-77, 78-79

microsoft

y estudio de la Universidad de Cambridge sobre el coeficiente intelectual de los usuarios de Facebook, 261

Universidad de Columbia estudio de cáncer de páncreas y, 28-29, 30

y los errores de escritura por los buscadores, 48-50

Lechero, Katherine L., 91-92

*Minority Report (película)*, 266

Minsky, Marvin, 273

minutos, hacer zoom sobre, 190-97

*Moneyball*, Un perfil de Oakland en, 254, 255

Moore, Julianne, 185

Moskovitz, Dustin, 238-39

películas

y la publicidad, 224-25

y el crimen, 193, 194-95, 273

violento, 190-97, 273

zoom sobre, 190-97

*Ver película también específica*

msnbc.com, 143

asesinato

y el peligro de gobierno facultado, 266-67, 268-69

*Ver también* violencia

Murdoch, Rupert, 96

Murray, Patty, 256

musulmanes

y el peligro de los gobiernos habilitados, 266-67, 268-69

y la verdad sobre el odio y el prejuicio, 129-31, 162-63

Nantz, Jim, 223

Centro Nacional de Estadísticas de Salud, 181

*National Enquirer* revista, 150-51, 152

identidad nacional, 78-79

experimentos naturales, 226-28, 229-30, 234-37, 239-40

NBA. ver el baloncesto

vecinos y ganancias inesperadas monetarios, 229

Netflix, 156-57, 203, 212

Netzer, Oded, 257-61

Patriotas de Nueva Inglaterra-Baltimore Ravens juegos, 221, 222-24

*New Jack City* (Película), 191

Ciudad de Nueva York, Rolling Stones canción sobre, 278

*Nueva York* revista, y pruebas A / B, 212

Nueva York Mets, 165-66, 167, 169, 171  
*New York Post*, y palabras como datos, 96  
*New York Times*  
Clinton (Bill) se buscará, 61  
y IQDNA resultados del estudio, 249  
y el habla Obama sobre el terrorismo, 130  
primera columna de Stephens-Davidowitz sobre el sexo en, 282  
Stormfront usuarios y, 137, 140, 145  
y la verdad sobre Internet, 145  
tipos de plantas, 92  
historia de los olores vaginales en, 161  
y palabras como datos, 95-96  
New York Times Company, y las palabras de datos como, 95-96  
revista *The New Yorker*  
estudio en dufo, 209  
y la búsqueda de doppelganger Stephens-Davidowitz, 202  
News Corporation, 96  
newslibrary.com, 95  
encuestas de Nielsen, 5  
Nietzsche, Friedrich, 268  
Nigeria, el embarazo en, 188, 189, 190  
"negro"  
y el odio y el prejuicio, 6, 7, 131-34, 244  
chistes, 6, 15, 132, 133, 134  
la motivación para las búsquedas sobre, 6  
y la elección de Obama, 7, 244  
y el poder de los grandes datos, 15  
prevalencia de búsquedas sobre, 6  
y elección de Trump, 14  
luz de la noche, y las imágenes como datos, 100-101  
Nike, 157  
Nixon, Richard M., 170, 171  
números, enamoramiento obsesivo con 252-56,

## Obama, Barack

y las pruebas A / B, 211-14  
la página de inicio de campaña, 212-14  
elecciones de 2008 y, 2, 6-7, 133, 134, 211-12  
elecciones de 2012 y, 8-9, 10, 133, 134, 211-12  
y el racismo en América, 2, 6-7, 8-9, 12, 134, 240, 243-44



Estado de expresión de la Unión (2014), 159-60  
y la verdad sobre el odio y el prejuicio, 130-31, 133, 134, 162-63  
Ocala subasta de caballos, 65-66, 67, 69  
complejo de Edipo, la teoría de Freud de, 50-51  
OkCupid (sitio de citas), 139  
Olken, Benjamin A., 227, 228  
*127 horas* (Película), 90, 91  
Las decisiones óptimas Group, 262  
O, Flora, 266  
Ortiz, David “Big Papi”, 197-200, 200n, 203  
pruebas “fuera de la muestra”, 250-51

Página, Larry, 60, 61, 62, 103

cáncer de páncreas, el estudio de la Universidad de Columbia-Microsoft de, 28-29  
Pandora, 203  
proyecto Panteón (Instituto de Tecnología de Massachusetts), 184-85  
los padres / paternidad  
y el abuso infantil, 145-47, 149-50, 161  
y los ejemplos de grandes volúmenes de datos búsquedas, 22  
y los prejuicios contra los niños, 134-36, 135N  
Parks, Rosa, 93, 94  
Parr, Ben, 153-54  
Pathak, Parag, 235-36  
PatientsLikeMe.com, 205  
patrones, y los datos de la ciencia como intuitiva, 27, 33  
Paul, Chris, de 37 años  
el pago de los préstamos de espalda, 257-61  
modelo PECOTA, 199-200, 200n  
pedigrís  
de jugadores de baloncesto, 67  
de caballos, 66-67, 69, 71  
podómetro, el énfasis en la probabilidad, 252-53  
pene  
y las teorías de Freud, 46  
y símbolos fálicos en los sueños, 46-47  
tamaño de, 17, 19, 123-24, 124n, 127  
“Penistrian”, 45, 46, 48, 50  
Pennsylvania State University, los ingresos de los graduados de, 237-39  
Peysakhovich, Alex, 254  
símbolos fálicos, en los sueños, 46-48

*Philadelphia Daily News*, y las palabras como datos, 95  
Filipinas, la economía de cigarrillos en, 102  
apariencia física  
pareja, 82, 120n  
y los padres prejuicios contra los niños, 135-36  
y la verdad sobre el sexo, 120, 120n, 125-26, 127  
la física, como la ciencia, 272-73  
imágenes, como datos, 97-102, 103  
Pierson, Emma, 160n  
Piketty, Thomas, 283  
Pinky Pizwaanski (caballo), 70  
pizza, información sobre, 77  
PlentyOfFish (sitio de citas), 139  
Plomin, Robert, 249-50  
ciencias políticas, y la revolución digital, 244, 274  
política  
y las pruebas A / B, 211-14  
complejidad de, 273  
y haciendo caso omiso de lo que la gente te dice, 157  
y el origen de las preferencias políticas, 169-71  
y la verdad acerca de la Internet, 140-44  
y palabras como datos, 95-97  
*Ver también* conservadores; demócratas; liberales; Los republicanos  
centro  
las búsquedas de Google, en comparación con el 9  
y la mentira, 107  
fiabilidad de, 12  
*Ver sondeo o tema específico también*  
Pop-Tarts, 72  
Popp, Noah, 202  
Popper, Karl, 45, 272, 273  
PornHub (sitio web), 14, 50-52, 54, 116, 120-22, 274  
pornografía  
como la adicción, 219  
y el sesgo de los medios sociales, 151  
y la lactancia, 19  
dibujos animados, 52  
niño, 121  
y la revolución digital, 279  
y homosexuales, 114-15, 114N, 116, 117, 119

honestidad de datos sobre, 53-54  
y el incesto, 50-52  
en la India, 19  
y la mentira, 110  
vídeos populares, 152  
popularidad, 53, 151  
y el poder de los grandes datos, 53  
los motores de búsqueda para, 61n  
y la verdad sobre el sexo, 114-15, 117  
parados y, 58, 59  
Posada, Jorge, 200  
pobreza  
y la esperanza de vida, 176-78  
y palabras como datos, 93, 94  
*Ver también* la distribución del ingreso  
predicciones  
y la ciencia de datos tan intuitivo, 27  
y las cifras correctas, 74  
y lo que cuenta como de datos, 74  
y lo que frente a por qué funciona, 71  
*Ver también* tema específico  
el embarazo, 20, 187-90  
perjudicar  
implícita, 132-34  
de los padres contra los niños, 134-36, 135N  
subconsciente, 134, 163  
verdad sobre, 128-40, 162-63  
*Ver también* parcialidad; odio; raza / racismo; Frente de tormenta  
Premise, 101-2, 103  
discriminación de precios, 262-65  
las condiciones de reclusión y la delincuencia, 235  
las cuestiones de privacidad, y peligro de gobierno facultado, 267-70  
los derechos de propiedad, y palabras como datos, 93, 94  
proquest.com, 95  
Prosperar (sitio de préstamo), 257  
Psy, video “Gangnam Style” de, 152  
psíquicos, 266  
psicología  
y la revolución digital, 274, 277-78, 279  
como ciencia, 273

la ciencia como blanda, 273  
y los métodos tradicionales de investigación, 274

Quantcast, 137

preguntas  
pidiendo a la derecha, 21-22  
pareja, 82-83

raza / racismo

causas de 18-19,  
elecciones de 2008 y, 2, 6-7, 12, 133  
elecciones de 2012 y, 2-3, 8, 133  
elecciones de 2016 y, 8, 11, 12, 14, 133  
explícita, 133, 134  
y Harvard carmesí editorial sobre Zuckerberg, 155  
y la mentira, 109  
mapa de, 7-9  
y Obama, 2, 6-7, 8-9, 12, 133, 240, 243-44  
y predecir el éxito en el baloncesto, 35, 36-37  
y republicanos, 3, 7, 8  
estudio de, 2-3, 6-7, 12, 14, 243-44 Stephens-Davidowitz de  
y Trump, 8, 9, 11, 12, 14, 133  
y la verdad sobre el odio y el prejuicio, 129-34, 162-63  
*Ver también* musulmanes; "negro"  
experimentos controlados aleatorios  
y las pruebas A / B, 209-21  
y la causalidad, 208-9  
violación, 121-22, 190-91  
Rawlings, Craig, 80  
"Rawtube" (sitio porn), 59  
Reagan, Andy, 88, 90, 91  
Reagan, Ronald, 227  
discontinuidad de regresión, 234-36  
Reisinger, José, 101-2, 103  
relaciones, duradera, 31-33  
la religión, y la esperanza de vida, 177  
Renaissance (fondos de cobertura), 246  
Los republicanos  
principios básicos de 94,  
y los orígenes de las preferencias políticas, 170-71

y el racismo, 3, 7, 8  
y palabras como datos, 93-97  
*Ver persona o elección también específica*  
investigación  
y la expansión de la metodología de investigación, 275-76  
*Ver investigador o investigación también específica*  
opiniones, de las empresas, 265  
“Tubo Rocket” (sitio porn gay), 115  
Rolling Stones, 278  
Romney, Mitt, 10, 212  
Condado de Roseau, Minnesota, con éxito notable americanos / a, 186, 187  
*Novia fugitiva* (Película), 192, 195

sabermetricians, 198-99

San Bernardino, California, rodar en, 129-30  
Arenas, Emily, 202  
ciencia  
y Big Data, 273  
y experimentos, 272-73  
reales, 272-73  
a escala, 276  
suave, 273  
los motores de búsqueda  
diferenciación de Google de otra, 60-62  
para la pornografía, 61n  
fiabilidad de, 60  
número de palabras, 71  
*Ver motor también específica*  
buscadores, por errores de escritura, 48-50  
búsquedas  
palabras negativas utilizadas en 128-29,  
*Ver también la búsqueda específica*  
“secretos acerca de las personas,” 155-56  
Seder, Jeff, 63-66, 68-70, 71, 74, 155, 256  
segregación, 141-44. Véase también el sesgo; discriminación; raza / racismo  
los autónomos y los impuestos, 178-80  
análisis de los sentimientos, 87-92, 247-48  
sexo  
como la adicción, 219  
y los beneficios del suero de la verdad digitales, 158-59, 161

y experiencias de la infancia, 50-52  
preservativos y, 5, 122  
y la revolución digital, 274, 279  
y dimensiones de la sexualidad, 279  
durante el matrimonio, 5-6  
y fétiches, 120  
y Freud, 45-52  
Google busca sobre, 5-6, 51-52, 114, 115, 117, 118, 122-24, 126, 127-28  
y gastos de la verdad, 158-59, 161  
y Harvard carmesí editorial sobre Zuckerberg, 155  
cuánto, 122-23, 124-25, 127  
en la India, 19  
nueva información sobre 19  
oral, 128  
y la apariencia física, 120, 120n, 125-26, 127  
y el poder de los grandes datos, 53  
pregancy y que tiene, 189  
Rolling Stones canción sobre, 278  
y los órganos sexuales, 123-24  
primera columna del New York Times de Stephens-Davidowitz aproximadamente, 282  
y los métodos tradicionales de investigación, 274  
verdad sobre, 5-6, 112-28, 114n, 117  
y los errores de mecanografía, 48-50  
y los genitales de la mujer, 126-27  
*Ver también* incesto; pene; pornografía; violación; vagina  
Shadow (app), 47  
Shakespeare, William, 89-90  
Shapiro, Jesse, 74-76, 93-97, 141-44, 235, 273  
“Shattered” (canción de los Rolling Stones), 278  
hábitos de compra, predicciones sobre, 71-74  
*La señal y el ruido (plata)*, 254  
Plata, Nate, 10, 12-13, 133, 199, 200, 254, 255  
Simmons, Bill, 197-98  
Singapur, el embarazo en, 190  
Siroker, Dan, 211-12  
dormir  
y la revolución digital, 279  
Mandíbula y, 276-77  
y el embarazo, 189  
“Slutload,” 58

datos pequeños, 255-56  
sonrisas, y las imágenes como datos, 99  
Smith, Michael D., 224  
Nieve, Juan, 275  
Sochi, Rusia, los homosexuales en, 119  
medios de comunicación social  
sesgo de los datos de, 150-53  
doppelganger la caza en, 201-3  
y las mujeres de las descripciones de los maridos, 160-61, 160-61n  
*Véase el sitio o tema específico también*  
ciencias sociales, 272-74, 276, 279  
de seguridad social, y palabras como datos, 93  
antecedentes socioeconómicos  
y predecir el éxito en el baloncesto, 34-41  
*Ver también* pedigrees  
sociología, 273, 274  
SOLTAS, Evan, 130, 162, 266-67  
Sudáfrica, el embarazo en, 189  
Southern Poverty Law Center, 137  
España, el embarazo en, 190  
*Spartanburg Herald-Diario* (Carolina del Sur), y las palabras como datos, 96  
especialización, extremo, 186  
velocidad, para la transmisión de datos, 56-59  
“Spider Solitaire,” 58  
Stephens-Davidowitz, Noah, 165-66, 165-66n, 169, 206, 263  
Stephens-Davidowitz, Seth  
ambiciones de, 33  
tumbado por, 282n  
opción para, 25-26, 271 gemela  
motivaciones de, 2  
la obsesión de, 282, 282n  
experiencia profesional de, 14  
y conclusiones escritura, 271-72, 279, 280-84  
Stern, Howard, 157  
bolsa de Valores  
datos para, 55-56  
y los ejemplos de grandes volúmenes de datos búsquedas, 22  
intento de predecir el, 245-48, 251-52 Summers-Stephens-Davidowitz  
Piedra, Oliver, 185  
Stoneham, James, 266, 269

Storegard, Adam, 99-101  
cuentos  
Categorías / tipos de 91-92,  
viral, 22, 92  
y el zoom, 205-6  
*Ver historia también específica*  
Stormfront (sitio web), 7, 14, 18, 137-40  
estrías, y el embarazo, 188-89  
Stuyvesant High School (Nueva York), 231-37, 238, 240  
áreas suburbanas, y los orígenes de americanos notables, 183-84  
éxito / americanos notables  
factores que impulsan, 185-86  
zoom sobre, 180-86  
sufrimiento, y los beneficios de suero de la verdad digitales, 161  
suicidio, y el peligro de gobierno facultado, 266, 267-68  
Summers, Lawrence  
y el estudio de Obama-racismo, 243-44  
y predecir el mercado de valores, 245, 246, 251-52  
reunión con Stephens-Davidowitz de, 243-45  
Sunstein, Cass, 140  
juegos de Super Bowl, la publicidad durante, 221-25, 239  
*Crunchers super (Gnau), 264*  
Tribunal Supremo, y el aborto, 147  
Surowiecki, James, 203  
encuestas  
en persona, 108  
Internet, 108  
y la mentira, 105-7, 108, 108n  
y las imágenes como datos, 97  
escepticismo sobre, 171  
teléfono, 108  
y la verdad sobre el sexo, 113, 116  
y hacer zoom sobre las horas y minutos, 193  
*Ver encuesta o tema específico también*  
refugiados sirios, 131

Taleb, Nassim, 17  
Tartt, Donna, 283  
TaskRabbit, 212  
impuestos



engañando a 22, 178 a 80, 206  
y los ejemplos de grandes volúmenes de datos búsquedas, 22  
y acostado, 180  
y los autónomos, 178-80  
y palabras como datos, 93-95  
zoom sobre, 172-73, 178-80, 206  
maestros, utilizando pruebas para juzgar, 253-54  
adolescentes  
adoptada, 108n  
como homosexuales, 114, 116  
tumbado por, 108n  
y los orígenes de las preferencias políticas, 169  
y la verdad sobre el sexo, 114, 116  
*Ver también* niños  
televisión  
y las pruebas A / B, 222  
la publicidad en, 221-26  
Terabyte, 264  
el terrorismo, 18, 129-31  
tests / pruebas  
de Estudiantes de la Enseñanza, 231-37, 253-54  
y juzgando maestro, 253-54  
y enamoramientos obsesivos con números, 253-54  
el comportamiento en línea como suplemento a, 278  
y pequeños datos, 255-56  
*Ver prueba o estudio también específica*  
Thiel, Pedro, 155  
Piense Progreso (sitio web), 130  
*Pensar rápido (Kahneman)*, 283  
Thome, Jim, 200  
Tourangeau, Roger, 107, 108  
pueblos, hacer zoom sobre, 172-90  
*Historia del juguete* (Película), 192  
Trump, Donald  
elecciones de 2012 y 7 de  
y haciendo caso omiso de lo que la gente te dice, 157  
y la inmigración, 184  
propagadas por temas, 7  
y los orígenes de americanos notables, 184  
encuestas sobre, 1

predicciones sobre, 11-14  
y el racismo, 8, 9, 11, 12, 14, 133, 139  
*Ver también* elecciones, 2016  
verdad  
beneficios de conocer, 158-63  
el manejo de la, 158-63  
*Ver también* suero de la verdad digital; acostado; tema específico  
Universidad de Tuskegee, 183  
Twentieth Century Fox, 221-22  
Twitter, 151-52, 160-61n, 201-3  
los errores de escritura por los buscadores, 48-50

*La insoportable levedad del ser (Kundera), 233*

*Inexplorado* (Aiden y Michel), 78-79  
desempleo  
y el abuso infantil, 145-47  
datos sobre, 56-57, 58-59  
consecuencias no deseadas, 197  
Estados Unidos  
y la Guerra Civil, 79  
tan unida o dividida, 78-79  
Universidad de California, Berkeley, en 2008 el racismo estudio elección en, 2  
Universidad de Maryland, encuesta de graduados de, 106-7  
Áreas urbanas  
y la esperanza de vida, 177  
y los orígenes de americanos notables, 183-84, 186

vagina, olores de, 19, 126-27, 161

Varian, Hal, 57-58, 224  
Vikingmaiden88, 136-37, 140-41, 145  
violencia  
y la ciencia real, 273  
zoom sobre, 190-97  
*Ver también* asesinar  
el registro de votantes, 106  
el número de votantes, 9-10, 109-10  
el comportamiento electoral, y la mentira, 106, 107, 109-10  
*Vox*, 202

Walmart, 71-72

*El Correo de Washington*, y palabras como datos, 75, 94  
*Washington Times*, y palabras como datos, 75, 94-95  
riqueza  
y la esperanza de vida, 176-77  
*Ver también* la distribución del ingreso  
tiempo, y las predicciones sobre el vino, 73-74  
Weil, David N., 99-101  
Weiner, Anthony, 234n  
nacionalismo blanco, 137-40, 145. Véase también Stormfront  
Whitepride26, 139  
Wikipedia, 14, 180-86  
vino, predicciones acerca, 72-74  
esposas  
y las descripciones de los maridos, 160-61, 160-61n  
y sospechas sobre la homosexualidad de los maridos, 116-17  
mujer  
pechos de, 125, 126  
culata de, 125-26  
genitales de 126-27,  
violencia contra, 121-22  
*Ver también las niñas; esposas; tema específico*  
palabras  
y el sesgo, 74-76, 93-97  
y las categorías / tipos de historias, 91-92  
como datos, 74-97  
pareja, 80-86  
y la revolución digital, 278  
y la digitalización de libros, 77, 79  
y el matrimonio homosexual, 74-76  
y análisis de los sentimientos, 87-92  
y Estados Unidos como unida o dividida, 78-79  
derechos de los trabajadores, 93, 94  
Banco Mundial, 102  
World of Warcraft (juego), 220  
Wrenn, Doug, 39-40, 41

Yahoo News, 140, 143  
anuarios, la escuela secundaria, 98-99  
Yelp, 265  
Yilmaz, Ahmed (alias), 231-33, 234, 234n

YouTube, 152

Zayat, Ahmed, 63-64, 65

*Uno a cero* (Thiel), 155

zoom

en el béisbol, 165-69, 165-66n, 171, 197-200, 200n, 203, 206, 239

beneficios de, 205-6

en condados, ciudades y pueblos, 172-90, 239-40

y el tamaño de datos, 171, 172-73

en doppelgangers, 197-205

en la igualdad de oportunidades, 173-75

en el juego, 263-65

en la salud, 203-5, 275

en la distribución del ingreso, 174-76, 185

y la influencia de las experiencias infantiles, 165-71, 165-66n, 206

en la esperanza de vida, 176-78

en minutos y horas, 190-97

y experimentos naturales, 239-40

y el origen de las preferencias políticas, 169-71

en el embarazo, 187-90

historias de, 205-6

el éxito / americanos notables, 180-86

sobre impuestos, 172-73, 178-80, 206

Zuckerberg, Mark, 154-56, 157, 158, 238-39

## **SOBRE EL AUTOR**

---

**Seth Stephens-Davidowitz** es colaborador del New York Times artículo de opinión, profesor visitante en la Escuela Wharton, y un ex científico de datos Google. Recibió una licenciatura en filosofía de Stanford, donde se graduó Phi Beta Kappa, y un doctorado en economía de la Universidad de Harvard. Su investigación, que utiliza nuevas, grandes fuentes de datos para descubrir las conductas y actitudes ocultas, ha aparecido en la Revista de Economía Pública y otras publicaciones de prestigio. Vive en la ciudad de Nueva York.

Descubre grandes autores, ofertas exclusivas, y más en [hc.com](https://www.hc.com).

The logo for Bookperk, featuring the word "Bookperk" in a sans-serif font. The word "Book" is in yellow and "perk" is in white. The logo is centered within a solid black rectangular background.

# Bookperk

[Sign up for Bookperk](#) and get e-book bargains, sneak peeks, special offers, and more—delivered straight to your inbox.

SIGN UP NOW

# DERECHOS DE AUTOR



TODOSMENTIRAS. Copyright © 2017 por Seth Stephens-Davidowitz. Copyright © 2017 por Seth Stephens-Davidowitz. Todos los derechos reservados bajo convenciones internacionales y Pan de Derecho de Autor. Por el pago de las tasas correspondientes, se le ha concedido el derecho no exclusivo, no transferible para acceder y leer el texto de este libro electrónico en la pantalla. Ninguna parte de este texto puede ser reproducido, transmitido, descargado, descompilado, ingeniería inversa, o almacenada o introducida en cualquier sistema de almacenamiento y recuperación de información, en cualquier forma o por cualquier medio, ya sea electrónico o mecánico, conocido ahora o en el futuro inventado, sin el permiso expreso por escrito de HarperCollins libros electrónicos.

PRIMERO EDICIÓN

Diseño de portada por Lisa Amoroso Fotografía de la portada del elefante / cebra © visuales ilimitadas, Inc./Victor Habbick Otros cebras © Shutterstock / Aaron Amat

ISBN 978-0-06-239085-1

EPub Edición Mayo 2017 ISBN 9780062390875

## **Sobre la Editorial**

---

### **Australia**

HarperCollins Publishers (Australia) Pty. Ltd.

Nivel 13, 201 Elizabeth Street

Sydney, NSW 2000, Australia

[www.harpercollins.com.au](http://www.harpercollins.com.au)

### **Canadá**

HarperCollins Canadá

2 Bloor Street East - Piso 20

Toronto, ON M4W 1A8, Canadá

[www.harpercollins.ca](http://www.harpercollins.ca)

### **Nueva Zelanda**

HarperCollins Publishers Nueva Zelanda

Unidad D1, 63 de Apollo Drive

Rosedale 0632

Auckland, Nueva Zelanda

[www.harpercollins.co.nz](http://www.harpercollins.co.nz)

### **Reino Unido**

HarperCollins Publishers Ltd.

1 la calle puente de Londres

Londres SE1 9GF, Reino Unido

[www.harpercollins.co.uk](http://www.harpercollins.co.uk)

### **Estados Unidos**

HarperCollins Publishers Inc.

195 Broadway

Nueva York, NY 10007

[www.harpercollins.com](http://www.harpercollins.com)



- \* Google Trends ha sido una fuente de gran parte de mis datos. Sin embargo, ya que sólo le permite comparar la frecuencia relativa de las diferentes búsquedas, pero no informa sobre el número absoluto de cualquier búsqueda en particular, por lo general he complementado con Google AdWords, que informa exactamente con qué frecuencia se hace cada búsqueda. En la mayoría de los casos también he sido capaz de nitidez de la imagen con la ayuda de mi propio algoritmo basado en las tendencias, que describo en mi tesis doctoral, “Ensayos Usando datos de Google,” y en mi papel Journal of Public Economics, “el costo de la animosidad racial en un candidato Negro: Evidencia el uso de Google Búsqueda de datos” la tesis doctoral, un enlace al papel, y una explicación completa de los datos y el código utilizados en toda la investigación original presentado en este libro están disponibles en mi sitio web, sethsd.com.

- \* La revelación completa: Poco después de terminar este estudio, me mudé de California a Nueva York. Utilizando datos de aprender lo que debe hacer es a menudo fácil. Haciendo realmente es difícil.

- \* Mientras que la versión inicial de la gripe de Google tenía defectos significativos, los investigadores han recalibrado recientemente el modelo, con más éxito.

- \* En 1998, si ha buscado “coches” en un motor de búsqueda más populares pre-Google, que fueron inundados con sitios de pornografía. Estos sitios de pornografía habían escrito la palabra “coches” con frecuencia en letras blancas sobre un fondo blanco para engañar al motor de búsqueda. Entonces consiguieron unos pocos clics adicionales de personas que pretenden comprar un coche, pero se distrajo con la pornografía.

- \* Una teoría que estoy trabajando en: Big Data sólo confirma todo lo que la tarde Leonard Cohen ha dicho. Por ejemplo, Leonard Cohen, una vez dio a su sobrino el siguiente consejo para cortejar a las mujeres: “Escucha bien. A continuación, escuchar un poco más. Y cuando se piensa que haya terminado de escuchar, escuchar un poco más.” Eso parece ser más o menos similar a lo que estos científicos encontraron.

- \* Otra razón para mentir es simplemente meterse con las encuestas. Este es un gran problema para cualquier investigación con respecto a los adolescentes, lo que complica fundamentalmente nuestra capacidad para entender este grupo de edad. Los investigadores encontraron originalmente una correlación entre un adolescente de ser adoptados y una variedad de comportamientos negativos, tales como el uso de drogas, el consumo de alcohol, y faltar a la escuela. En la investigación posterior, se encontraron con esta correlación se explica en su totalidad por el 19 por ciento de los adolescentes adoptados percepción subjetiva de los que no estaban realmente han tomado. seguimiento de la investigación ha encontrado que un porcentaje significativo de adolescentes decirle encuestas que son más de siete pies de altura, pesar más de cuatrocientas libras, o tiene tres hijos. Una encuesta encontró 99 por ciento de los estudiantes que reportaron tener un miembro artificial a los investigadores académicos estaban bromeando.

- \* Para algunos puede ser ofensivo que asocio una preferencia masculina por Judy Garland con una preferencia por tener relaciones sexuales con hombres, ni siquiera en broma. Y luego, no quiero decir que todo o incluso los hombres más-homosexuales tienen una fascinación con las divas. Pero los datos de búsqueda demuestran que hay algo al estereotipo. Calculo que un hombre que busca información sobre Judy Garland es tres veces más propenso a buscar porno gay de la pornografía recta. Algunos estereotipos, grandes volúmenes de datos nos dicen, son verdaderas.

- \* Creo que estos datos también tiene implicaciones sobre la estrategia óptima de citas. Claramente, uno debe ponerse por ahí, son rechazados mucho, y no tomar a pecho. Este proceso permitirá que, con el tiempo, para encontrar el compañero que está más atraído por alguien como tú. Una vez más, no importa lo que parece, estas personas existen. Créeme.



- \* Quería llamar a este libro ¿Qué tan grande es mi pene? Lo que Google Búsquedas nos enseñan sobre la naturaleza humana, pero mi editor me que sería difícil de vender advirtió, que la gente podría ser demasiado avergonzado para comprar un libro con ese título en una librería en el aeropuerto. ¿Estás de acuerdo?

- \* A fin de probar la hipótesis de que los padres tratan a niños de diferentes géneros de manera diferente, estoy trabajando en la obtención de datos de sitios web para padres. Esto incluiría un número mucho mayor de los padres que los que hacen estas búsquedas particulares y específicas.

- \* He analizado los datos de Twitter. Doy gracias a Emma Pierson ayuda descargan. No he incluido descriptores de lo que el marido está haciendo ahora, que son frecuentes en los medios sociales, pero realmente no tendría sentido en la búsqueda. Incluso estas descripciones se inclinan hacia el favorable. Las mejores formas para describir lo que un marido está haciendo ahora en los medios sociales están “trabajando” y “cocina”.

- \* La revelación completa: Cuando estaba hecho de comprobación de este libro, Noah negó que su odio hacia el pasatiempo de América es una parte clave de su personalidad. Él admite que odiaba el béisbol, pero él cree que su bondad, el amor de los niños, y la inteligencia son los elementos básicos de su personalidad y que sus actitudes sobre el béisbol ni siquiera entre los diez primeros. Sin embargo, llegué a la conclusión de que a veces es difícil ver la propia identidad objetiva y, como un observador externo, soy capaz de ver que odiaba el béisbol es de hecho fundamental para quien es Noé, si él es capaz de reconocer o no. Así que lo dejé en.

- \* Esta historia muestra cómo las cosas que parecen malas pueden ser buenas si previenen algo peor. Ed McCaffrey, ex receptor abierto educado en Stanford, utiliza este argumento para justificar dejando que los cuatro de sus hijos jugar al fútbol: “Estos chicos tienen energía. Y, por lo que, si no están jugando al fútbol, que están skate, que están trepando a los árboles, que están jugando al escondite en el patio trasero, que están haciendo paintball. Es decir, que no van a sentarse allí y no hacer nada. Y, así, la manera en que veo que es, bueno, al menos hay reglas dentro del deporte del fútbol. . . . Mis hijos han estado en la sala de emergencias por la caída de las cubiertas, conseguir en accidentes de bicicleta, skate, cayendo de los árboles. Es decir, lo que sea. . . Sí, es un deporte violento choque. Pero, también, mis chicos sólo tienen la personalidad, donde, por lo menos no son de ardilla saltando de montañas y hacer cosas locas como esa. Así, se organiza la agresión, supongo.” El argumento de McCaffrey, hecha en una entrevista en la manada con Colin Cowherd, es uno que nunca había oído antes. Después de leer el periódico Dahl / Della Vigna, tomo el argumento serio. Una ventaja de grandes conjuntos de datos del mundo real, en lugar de los datos de laboratorio, es que pueden recoger este tipo de efectos.

- \* Usted puede decir probablemente por esta parte del libro que tienden a ser cínicos sobre las buenas historias. Quería una historia de sentirse bien aquí, así que estoy dejando mi cinismo a una nota al pie. Sospecho PECOTA acaba de enterarse de que Ortiz era un usuario de esteroides que dejaron de usar esteroides y que empezaron a usarlas de nuevo. Desde el punto de vista de la predicción, en realidad es bastante fresco si PECOTA fue capaz de detectar, pero que se convierte en una historia menos movimiento.

\* Un famoso 1978 de papel que afirmó que ganar la lotería no te hace feliz ha sido ampliamente desacreditada.

\* He cambiado de nombre y algunos detalles.



- \* En la búsqueda de personas como Yilmaz que obtuvieron cerca del punto de corte, me quedé impresionado por el número de personas en sus veinte años a través de sus cincuenta años-que recuerdan esta experiencia para tomar exámenes de sus primeros años de adolescencia y hablan de la falta de un punto de corte en términos dramáticos. Esto incluye al ex congresista y candidato a la alcaldía de la ciudad de Nueva York Anthony Weiner, quien dice que se perdió Stuy por un solo punto. “Ellos no me quieren”, me dijo, en una entrevista telefónica.

\* Dado que todo el mundo miente, se debe cuestionar la mayor parte de esta historia. Tal vez no soy un trabajador obsesivo. Tal vez yo no trabajo extraordinariamente duro en este libro. Tal vez yo, como mucha gente, puedo exagerar lo mucho que trabajo. Tal vez mis trece meses de “trabajo duro” incluído meses completos en los que he hecho ningún trabajo en absoluto. Tal vez yo no vivía como un ermitaño. Tal vez, si marcó mi perfil en Facebook, verías fotos de mí con amigos durante este periodo ermitaño supuesta. O tal vez yo era un ermitaño, pero no fue autoimpuesto. Tal vez pasé muchas noches solo, incapaz de trabajar, esperando en vano que alguien pondría en contacto conmigo. Tal vez nadie Envíenme un Vites a nada. Tal vez nadie me mensajes en Bumble. Todos mienten. Cada narrador es poco fiable.