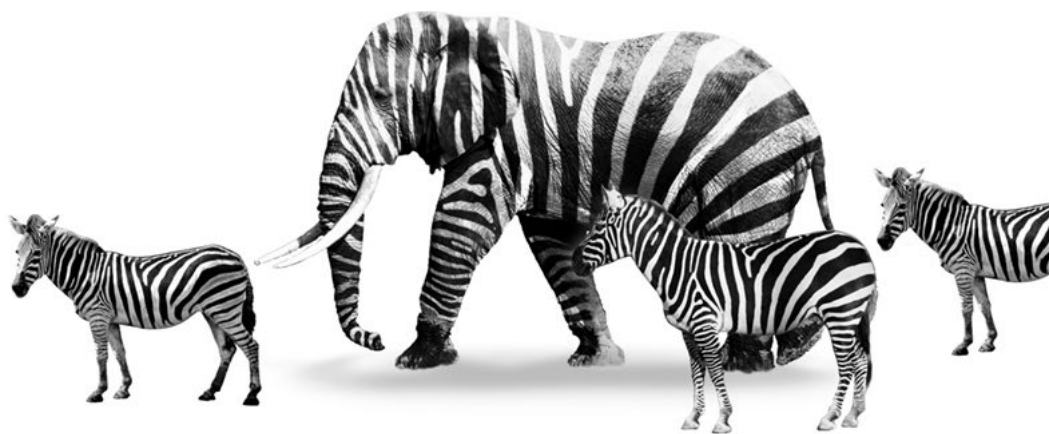


NEW YORK TIMES BESTSELLER

KAŽDÝ KLAME

BIG DATA, NEW DATA

A ČO NÁM INTERNET MÔŽE PREZRADIŤ
O TOM, KTO SKUTOČNE SME



SETH STEPHENS-DAVIDOWITZ

Ljndeni

Každý klame

Vyšlo aj v tlačovej podobe

Objednať môžete na
www.lindeni.sk
www.albatrosmedia.sk

Ljndeni

Seth Stephens-Davidowitz
Každý klame – e-kniha
Copyright © Albatros Media a. s., 2019

Všetky práva vyhradené.
Žiadna časť tejto publikácie nesmie byť rozširovaná
bez písomného súhlasu majiteľov práv.

ALBATROS  **MEDIA**

**KAŽDÝ
KLAME**

SETH STEPHENS-DAVIDOWITZ

KAŽDÝ KLAME

BIG DATA, NEW DATA

A ČO NÁM INTERNET MÔŽE PREZRADIŤ

O TOM, KTO SKUTOČNE SME

Ljndeni

Mame a otcovi

OBSAH

| | |
|------------------------|----|
| Predslov | 11 |
| Úvod: Obrysy revolúcie | 15 |

ČASŤ I **DÁTA, VEĽKÉ A MALÉ** **35**

| | |
|--------------------------------------|-----------|
| 1. Váš chybný pocit v žalúdku | 37 |
|--------------------------------------|-----------|

ČASŤ II **SILA VEĽKÝCH DÁT** **53**

| | |
|-----------------------------|-----------|
| 2. Mal Freud pravdu? | 55 |
|-----------------------------|-----------|

| | |
|----------------------------------|-----------|
| 3. Znovuprehodnotené dáta | 65 |
|----------------------------------|-----------|

| | |
|----------------|----|
| Ľudia ako dáta | 72 |
|----------------|----|

| | |
|----------------|----|
| Slová ako dáta | 82 |
|----------------|----|

| | |
|---------------------|-----|
| Fotografie ako dáta | 103 |
|---------------------|-----|

| | |
|----------------------------------|------------|
| 4. Digitálne sérum pravdy | 109 |
|----------------------------------|------------|

| | |
|---------------|-----|
| Pravda o sexe | 115 |
|---------------|-----|

| | |
|----------------------------------|-----|
| Pravda o nenávisti a predsudkoch | 129 |
|----------------------------------|-----|

| | |
|---|------------|
| Pravda o internete | 139 |
| Pravda o zneužívaní detí a potrate | 144 |
| Pravda o vašich priateľoch na Facebooku | 148 |
| Pravda o vašich zákazníkoch | 151 |
| Dokážeme zniesť pravdu? | 155 |
| 5. Detailný záber | 161 |
| Čo sa skutočne deje v našich krajoch, veľkomestách a mestách? | 167 |
| Ako vyplňame naše minúty a hodiny | 184 |
| Naši dvojníci | 189 |
| Príbehy dát | 197 |
| 6. Celý svet je laboratórium | 199 |
| Abeceda testovania A/B | 201 |
| Kruté, ale poučné pokusy prírody | 212 |
| | |
| ČASŤ III VEĽKÉ DÁTA: ZAOBCHÁDZAJTE OPATRNE | 229 |
| <hr/> | |
| 7. Veľké dáta, veľké haraburdy? Čo nedokážu | 231 |
| Kliatba rozmernosti | 234 |
| Prehnaný dôraz na to, čo môžeme merať | 239 |
| 8. Viac dát, viac problémov? Čo by sme robiť nemali | 243 |
| Nebezpečenstvo oprávnených spoločností | 243 |
| Nebezpečenstvo vlád s autorizáciou moci | 250 |
| Záver: Koľko ľudí dočíta knihy? | 255 |
| Podakovanie | 267 |

PREDSLOV

Už od čias, kedy filozofi špekulovali o „cerebroskope“ – mýtikom prístroji, ktorý by dokázal zobrazit' ľudské myšlienky na obrazovke, spoločenski vedci hľadali nástroje na odhalenie fungovania ľudskej povahy. Počas mojej kariéry experimentálneho psychológa mnohé z nich spopulárnili a potom vyšli z módy. Ja som ich vyskúšal všetky – posudzovacie stupnice, reakčný čas, rozšírenie zrenice, funkčné zobrazovanie mozgu, dokonca aj epileptických pacientov s implementovanými elektródami, ktorí si ochotne krátili čas jazykovým experimentom čakajúc na záchvat.

Avšak ani jedna z týchto metód neposkytuje neobmedzený pohľad do ľudskej mysle. Problém je v hrubom kompromise. Ľudské myšlienky sú zložité. Na rozdiel od Woodyho Allena*, ktorý po absolvovaní rýchlokurzu prečítal *Vojnu a mier*, si nepomyslíme len: „Bolo to o nejakých Rusoch.“ Lenže myšlienky v celej ich spleťtej, viac-dimenzionálnej nádhere vedec dokáže analyzovať len s ťažkosťami. Samozrejme, keď si ľudia vylievajú srdce, chápeme bohatosť prúdu ich vedomia, ale monológy nie sú ideálnym súborom dát na testovanie hypotéz. Na druhej strane, ak sa zameriame na ľahko vyčísliteľné

* Woody Allen ako vtip na účet kurzov rýchločítania vyhlásil: Absolvoval som kurz rýchločítania, kedy prstom prebehneš stred stránky, a bol som schopný prečítať *Vojnu a mier* za dvadsať minút. Je to o Rusku, pozn. prekladateľa.

kroky, akými sú napríklad reakčný čas ľudí na slová alebo reakcia ich kože na obrázky, môžeme z toho vytvoriť štatistiku, ale zjednodušili sme tým zložitú textúru poznania na jednoduché číslo. Dokonca aj najsofistikovanejšie metódy zobrazovania mozgu nám dokážu povedať, ako je myšlienka rozprestretá v 3D priestore, ale nie to, čo myšlienka obsahuje.

Ako keby kompromis medzi poddajnosťou a bohatosťou myšlienok nebol dostatočne komplikovaný, vedci zaoberajúci sa ľudskou povahou sú otrávení Zákonom malých čísiel – názov, ktorým Amos Tversky a Daniel Kahneman opísali falošný úsudok, že charakterové vlastnosti populácie budú odzrkadlené v každej, akokoľvek malej vzorke. Dokonca aj vedci zaoberajúci sa číslami majú žalostne chybné poznatky o tom, aké množstvo predmetov je skutočne potrebných v štúdiu predtým, než z nej môžeme vyňať náhodné bizarnosti a nezrovnalosti, aby sme ich potom mohli zovšeobecniť na všetkých Američanov, nehovoriac o *homo sapiens*. Je to ešte oveľa horšie, ak je vzorka vybraná ako sa nám to hodí, napríklad, keď v našom kurze ponúkne peniaze na pivo druhákom vysokej školy. Táto kniha je o úplne novom spôsobe štúdia ľudskej mysle. Big data (veľké dáta) z internetových vyhľadávačov a iných online výsledkov nie sú síce cerebroskopom, ale Seth Stephens-Davidowitz poukazuje na to, že ponúkajú neobvyklý náhľad do ľudskej psychiky. V súkromí vlastných klávesníc sa ľudia priznávajú k neobyčajným veciam. Niekedy preto, že majú reálne dôsledky (na zoznamkách alebo vo vyhľadávačoch odborných rád), inokedy práve preto, že žiadne dôsledky nemajú – ľudia sa môžu priznať k tajným želaniam alebo strachom bez toho, aby niekoho reálne vyľakali alebo ohromili. Každopádne, ľudia nielen stláčajú klávesy alebo otáčajú gombíkmi, ale vpísaním ktorejkoľvek z triliónu sekvencií znakov vypovedajú svoje myšlienky v ich výbušnej, spletitej rozmerosti. A čo je ešte lepšie, zanechávajú tak digitálne stopy vo forme, v ktorej ich je možné ľahko zhromaždiť a analyzovať. Tieto stopy pochádzajú zo všetkých sfér života. Zúčastňujú

sa tak nenápadného experimentu, ktorý mení podnety a usporadúva odpovede v reálnom čase a oni spokojne dodávajú tieto údaje v gargantuovských číslach.

Každý klame je viac ako len dôkazom konceptu. Zakaždým tak Stephens-Davidowitz svojimi objavmi úplne prevrátil moje predsudky o mojej krajine a mojom druhu. Odkiaľ sa zjavila nečakaná Trumpova podpora? Alebo zistenie Ann Landers, ktorá, keď sa v roku 1976 opýtala svojich čitateľov, či neolutovali, že mali deti, a šokovane zistila, že väčšina z nich to skutočne oľutovala – bola len pomýlená nereprezentatívnou, samo-vybranou vzorkou? Je možné viniť internet za „filtračnú bublinu“, túto zbytočne označenú krízu roku 2010? Čo zapríčiňuje zločin z nenávisť? Vyhľadávajú ľudia vtipy, aby sa rozptýlili? Aj napriek tomu, že si rád myslím, že ma už nemôže nič zaskočiť, bol som veľmi šokovaný tým, čo internet odhaľuje o ľudskej sexualite – zahŕňajúc objav, že každý mesiac určité množstvo žien vyhľadáva „sex s plyšákmi“. Žiadny experiment využívajúci reakčný čas, rozšírenie zrenice či funkčné zobrazenie mozgu by nikdy nedokázal prísť na takýto poznatok.

Každému sa bude páčiť kniha *Každý klame*. S nepoľavujúcou zvedavosťou a milým dôvtipom Stephens-Davidowitz poukazuje na nový smer spoločenskej vedy v dvadsiatom prvom storočí. Kto už len potrebuje cerebroskop s týmto nekonečne fascinujúcim oknom do ľudských obsesí?

Steven Pinker, 2017

ÚVOD

OBRYSY REVOLÚCIE

Povedali, že určite prehrá.

V primárnych republikánskych voľbách v roku 2016 odborníci na prieskum usúdili, že Trump nemá šancu. Trump predsa urazil niekoľko menšinových skupín. Prieskumy verejnej mienky a ich vyhodnocovači tvrdili, že iba zopár Američanov súhlasilo s takými urážkami.

Väčšina volebných odborníkov si v tom čase myslela, že Trump všeobecné voľby prehrá. Príliš veľa voličov sa vyjadrilo, že ich odradiť jeho správanie a názory.

V skutočnosti sa však vyskytli určité náznaky, že Trump môže naozaj oboje voľby, primárne aj všeobecné, vyhrať – na internete.

Som expert na internetové dáta. Každý deň sledujem digitálne stopy, ktoré ľudia zanechávajú, keď surfujú po internete. Z tlačidiel alebo kľúčov, ktoré stlačíme alebo ktorými klikneme, sa snažím porozumieť tomu, čo v skutočnosti chceme, čo v skutočnosti spravíme a kto v skutočnosti sme. A teraz vám vysvetlím, ako som sa na túto nezvyčajnú dráhu dostal.

Môj príbeh sa začal – a zdá sa to už byť večnosť – prezidentskými voľbami v roku 2008 a dlhodobo debatovanou otázkou spoločenskej vedy: Aký význam majú rasové predsudky v Amerike?

Barack Obama bol prvým afroamerickým volebným kandidátom významnej strany a vyhral pomerne ľahko. Prieskumy naznačovali, že rasa nebola faktorom ovplyvňujúcim hlasovanie Američanov. Gallup*, napríklad, uskutočnil mnohé prieskumy pred a po Obamových prvých voľbách. A ich záver? Amerických voličov prevažne nezaujímalo, že Barack Obama bol černochoch. Krátko po voľbách si dvaja známi profesori z Kalifornskej univerzity Berkley preštudovali iné dáta z prieskumov použitím sofistikovanejších techník – a prišli k podobnému záveru. A tak sa počas Obamovho prezidentského úradovania toto stalo bežnou vedomosťou v mnohých médiách a širokých okruhoch. Zdroje, ktoré médiá a spoločenský vedci využívali na porozumenie sveta viac ako 80 rokov, nám tvrdili, že prevažnú väčšinu Američanov pri zvažovaní prezidentského kandidáta nezaujímalo fakt, že Obama bol černochoch.

Zdalo sa, že táto krajina, dlho hanobená otroctvom a zákonmi Jima Crowa** konečne prestala posudzovať ľudí podľa farby pokožky, čo zdanlivo naznačovalo, že rasizmus v Amerike pomaly vymiera. Niektorí odborníci dokonca vyhlásili, že žijeme v spoločnosti, v ktorej nezáleží na tom, akej ste rasy.

V roku 2012 som bol postgraduálnym študentom ekonomiky, cítil som sa stratený, vyhorený vo svojom obore, a sebaistý až arrogantný, presvedčený o tom, že rozumiem fungovaniu sveta. Dokonca aj tomu, čo si ľudia v dvadsiatom prvom storočí mysleli a na čom im záležalo. Takže keď prišlo na problém tohto predsudku, dovolil som si veriť, že všetko, čo som sa dočítal v psychológii a politológii, poukazovalo na to, že otvorený rasizmus bol obmedzený len na malé percento Američanov – väčšinou na konzervatívnych republikánov, najmä z ďalekého juhu.

A potom som objavil Google Trendy.

* Prieskumná agentúra, pozn. prekladateľa.

** Štátne a miestne zákony, ktoré presadzovali rasistické rozdelenie južnej časti Spojených štátov, pozn. prekladateľa.

Google Trendy, nástroj, ktorý bol nenápadne spustený v roku 2009, hovorí užívateľom, ako často bolo nejaké slovo alebo fráza vyhľadávané na rôznych miestach v rôznych časoch. Bol propagovaný ako prostriedok na zábavu – asi aby umožnil kamarátom diskutovať o celebritách alebo móde. Prvotná verzia obsahovala žartovné upozornenie, že ľudia „nebudú chcieť napísať vašu PhD dizertáciu“ s týmito dátami, čo ma okamžite motivovalo k napísaniu mojej dizertácie.*

Google vyhľadávač nebol v tom čase považovaný za vierohodný zdroj informácií pre „seriózny“ akademický výskum.

Na rozdiel od prieskumov, dáta z Google vyhľadávaní neboli vytvorené za účelom porozumenia ľudskej psychiky. Google bol vynájdený na to, aby ľudia mohli získavať informácie o svete, a nie na to, aby výskumníci mohli získavať informácie o ľuďoch. Výsledkom však je, že stopy, ktoré po sebe zanechávame pri vyhľadávaní poznatkov na internete, sú nesmierne odhaľujúce.

Inými slovami, spôsob, akým ľudia vyhľadávajú informácie, je sám o sebe informáciou. Kedy a kde vyhľadávajú fakty, citáty, vtipy, miesta, osoby, veci alebo pomoc, nám dokáže povedať oveľa viac o tom, čo si skutočne myslia, želajú, čoho sa boja a čo v skutočnosti robia. O to viac, že ľudia až tak často na internete veci nezistujú,

* Google Trendy bol zdrojom väčšiny mojich údajov. Keďže však povoľuje iba porovnanie relatívnej frekvencie rôznych vyhľadávaní a nezaznamenáva skutočné množstvá akéhokoľvek vyhľadávania, zvyčajne som ho doplnil Google AdWords, ktorý presne uvádza frekvenciu každého vyhľadávania. Vo väčšine prípadov som mohol navyše upresniť celkový obraz pomocou vlastného Trendy-podloženého algoritmu, ktorý popisujem vo svojej dizertácii „Eseje využívajúce Google dáta“ a v mojej štúdiu v Časopise verejnej ekonomiky „Cena rasistických tendencií voči černošskému kandidátovi: Dôkazy využitia dát Google vyhľadávaní“. Dizertáciu, spojivo medzi štúdiom a kompletným vysvetlením dát a kódov použitých v originálnom výskume prezentovaných v tejto knihe, môžete nájsť na mojej webovej stránke sethsd.com.

skôr sa mu zverujú: „Neznášam môjho šéfa“, „Som opitý“, „Otec ma udrel“.

Každodenný úkon vpísania slova alebo frázy do kompaktného, hranatého, bieleho rámčeka zanecháva malé stopy pravdy, ktoré po vynásobení miliónmi za sebou nakoniec zanechávajú významné fakty. Prvé slovo, ktoré som napísal do Google Trendov, bolo „Boh“. Zistil som, že štáty, ktoré vykonávajú najviac vyhľadávanií s heslom „Boh“, boli Alabama, Mississippi a Arkansas – štáty biblického pásu. A najviac sa tieto vyhľadávania diali v nedeľu. Nebolo to prekvapením, ale bolo zvláštne, že výskumy dát dokážu odhaliť takýto jednoznačný príklad. Vyskúšal som „Knicks“, ktorý je, pochopiteľne, najviac vyhľadávaný v New Yorku. Ďalšia samozrejmosť. Potom som zadal svoje meno. „Je nám ľúto“, informovali ma Google Trendy. „Nedostatočný vyhľadávací objem“ na zobrazenie výsledkov. Zistil som, že Google Trendy zabezpečí dáta iba vtedy, ak je rovnaké vyhľadávanie vykonávané veľkým množstvom ľudí.

Sila Google vyhľadávanií však nie je v tom, že nám dokážu povedať, že Boh je populárny dole na juhu, Knicks sú populárni v New Yorku, alebo že ja nie som nikde populárny. To vám mohol povedať hociktorý prieskum. Sila Google dát je v tom, že ľudia tomuto obrovskému internetovému vyhľadávaču povedia veci, ktoré by nepovedali nikomu inému.

Vezmite si, napríklad, sex (predmet, ktorému sa budem oveľa podrobnejšie venovať v tejto knihe neskôr). Prieskumom sa nedá veriť, že nám povedia pravdu o našom sexuálnom živote. Analyzoval som dáta zo všeobecných sociálnych prieskumov, ktoré sú považované za najvplyvnejšie a aj najspoľahlivejšie zdroje informácií o správaní sa Američanov. Čo sa týka heterosexuálneho sexu, ženy podľa tohto prieskumu tvrdia, že majú sex v priemere päťdesiatpäťkrát do roka, z toho 16 percent s použitím kondómu. To zodpovedá 1,1 miliarde

* Basketbalový tím v New Yorku, pozn. prekladateľa.

kondómov použitých za rok. Ale heterosexuálni muži tvrdia, že ročne použijú 1,6 miliardy kondómov. Kto teda vraví pravdu, muži alebo ženy?

Zistilo sa, že ani jedna strana. Podľa globálnej informačnej a vymeriavacej spoločnosti Nielsen, ktorá sleduje správanie zákazníkov, sa ročne predá menej ako 600 miliónov kondómov. Takže klamú všetci, rozdiel je len v tom, ako veľmi.

V skutočnosti je klamanie veľmi rozšírené. Muži, ktorí neboli nikdy ženatí, tvrdia, že používajú v priemere dvadsaťdeväť kondómov ročne. Čo by bolo oveľa viac ako ročné množstvo kondómov, ktoré sa predá v USA ženatým a slobodným mužom dohromady. Manželské páry pravdepodobne tiež prehávajú v tom, koľko majú sexu. Ženatí muži do šesťdesiatpäť rokov v prieskumoch tvrdia, že majú sex v priemere raz za týždeň. Iba jedno percento tvrdí, že za posledný rok nemalo žiadny sex. Vydaté ženy tvrdia, že majú menej sexu, ale nie oveľa menej.

Google vyhľadávania podávajú menej rozmanitý – a ja si dokonca dovoľím tvrdiť, že oveľa presnejší – obraz o sexe v manželstve. Najčastejšou sťažnosťou na Googli je manželstvo bez sexu. Vyhľadávania hesla „manželstvo bez sexu“ sú 3,5-krát častejšie ako „nešťastné manželstvo“ a osemkrát častejšie ako „manželstvo bez lásky“. Dokonca aj nemanželské páry sa akosi často sťažujú, že sú bez sexu. Google vyhľadávania „vzťahy bez sexu“ sú na druhom mieste, tesne pred vyhľadávaniami na „násilné vzťahy“. (Musím však zdôrazniť, že všetky tieto dáta sú prezentované anonymne. Google, samozrejme, nezaznamenáva žiadne dáta o jednotlivých vyhľadávaniach.)

A Google vyhľadávania takisto prezentovali obraz o Amerike, ktorý bol veľmi odlišný od rasistickej utópie načrtnutej v prieskumoch. Pamätám si moment, keď som po prvýkrát zadal heslo „neger“ do Google Trendov. Môžete mi povedať, že som naivný. Ale pri tom, aké toxické je toto slovo, som naozaj očakával, že to bude málo objemné vyhľadávanie. To som sa teda riadne mýlil. V Spojených

štátoch sa slovo „neger“ alebo jeho množné číslo „negri“ vyskytvalo zhruba v rovnakom množstve vyhľadávani ako slovo „migréna“, „ekonóm“ a „Lakers“. Premýšľal som, či vyhľadávania textov rapovej hudby mohli ovplyvniť výsledky. Nie. Slovo použité v rapových pesničkách je takmer vždy „nigga“. Tak čo teda motivovalo Američanov vyhľadávať slovo „neger“? Často hľadali vtipy zosmiešňujúce Afroameričanov. V skutočnosti 20 percent vyhľadávani so slovom „neger“ taktiež zahŕňalo slovo „vtip“. Iné bežné vyhľadávania obsahovali „hlúpy neger“ a „nenávidím negrov“.

Týchto vyhľadávani bolo každý rok milióny. Veľké množstvo Američanov realizovalo v súkromí svojich domovov šokujúce rasistické pátrania. Čím viac som bádal, tým znepokojujúcejšie boli tieto informácie.

V prvú noc Obamovho zvolenia, keď sa väčšina komentárov sústredila na jeho chválospevy a uznanie historickej podstaty volieb, zhruba jedno zo sto Google vyhľadávani, ktoré obsahovalo slovo „Obama“, taktiež obsahovalo „kkk“^{***} alebo „neger (negri)“. Možno sa to nezdá až tak veľa, ale zamyslíte sa nad tými tisíckami nerasistických dôvodov, prečo cez Google vyhľadať tohto mladého outsidera so šarmantnou rodinou, ktorý čoskoro preberie najvplyvnejší post na svete. Počas volebnej noci boli vyhľadávania a prihlásenia na Stormfront, nacionalistické fórum s prekvapujúco vysokou popularitou v Spojených štátoch, viac než desaťkrát vyššie ako obyčajne. V niektorých štátoch sa vyskytlo oveľa viac vyhľadávani hesla „neger prezident“ ako „prvý černošský prezident“.

Objavila sa temnota a nenávisť, ktorá bola ukrytá pred tradičnými zdrojmi, bola však dosť zjavná v ľudských vyhľadávaniach.

Tieto vyhľadávania sa ťažko dávajú do súladu so spoločnosťou, v ktorej je rasizmus nepatrným faktorom. O Donaldovi J. Trumpovi

* Basketbalový tím, pozn. prekladateľa.

** Ku klux klan, pozn. prekladateľa.

som v roku 2012 vedel hlavne ako o podnikateľovi a účinkujúcom v reality šou. Ani mi nenapadlo, rovnako ako mnohým iným, že o štyri roky neskôr sa zrazu stane serióznym kandidátom na prezidenta. Ale tieto nepekne vyhľadávania nie je ťažké spojiť s úspechom kandidáta, ktorý – v jeho útokoch na imigrantov, v jeho hneve a od-pore – hral na najhoršie ľudské sklony.

Google vyhľadávania taktiež zistili, že sme sa mýlili v miestach vý-skytu rasizmu. Prieskumy a všeobecná znalosť umiestnila rasizmus predovšetkým na juh a hlavne medzi republikánov. Ale miesta s naj-vyšším rasistickým vyhľadávaním zahŕňali aj severný New York, západnú Pensylvániu, južné Ohio, priemyselný Michigan a vidiecky Illinois, popri Západnej Virgínii, južnej Louisiane a Mississippi. Skutočným rozdelením, ktoré navrhlo vyhľadávanie dát cez Google, nebolo Juh proti Severu, ale Východ proti Západu. Takýto typ vy-hľadávania nedostanete veľmi na západ od Mississippi. A rasizmus nebol obmedzený len na republikánov. Rasistické vyhľadávania na miestach s vyšším percentom republikánov neboli v skutočnosti o nič vyššie ako na miestach s vysokým percentom demokratov. Google vyhľadávania tak pomohli nakresliť novú mapu rasizmu Spojených štátov – a táto mapa bola veľmi odlišná od našej pôvodnej predstavy. Republikáni na juhu sa skôr priznávajú k rasizmu ako mnohí demokrati na severe, aj keď majú takmer rovnaké postoje.

O štyri roky neskôr sa potvrdila dôležitosť tejto mapy pri vysvet-lovaní politického úspechu Trumpa. V roku 2012 som využil túto mapu, ktorú som vytvoril pomocou Google vyhľadávani, na prehod-notenie roly, ktorú zohrávala Obamova rasa. Dáta boli jednoznačné. V častiach krajiny s najvyšším počtom rasistických vyhľadávani bol Obama na tom oveľa horšie ako biely demokratický kandidát John Kerry pred štyrmi rokmi. Táto súvislosť sa v týchto oblastiach nedala vysvetliť nijakým iným faktorom, zahŕňajúc stupeň vzdelania, vek, návštevu kostola alebo vlastníctvo zbraní. Rasistické vyhľadávania

nepredpovedali taký slabý výkon pri nijakom inom demokratickom kandidátovi. Iba pri Obamovi.

A výsledky naznačovali obrovský efekt. Obama stratil približne 4 percentá v rámci celej krajiny len kvôli explicitnému rasizmu. To bolo oveľa viac ako podľa očakávaní prieskumov. Barack Obama bol teda prirodzene zvolený a znovu zvolený prezident vďaka veľmi priaznivým podmienkam demokratov. Musel však prekonať oveľa viac prekážok ako ktokoľvek iný, kto sa spoliehal na tradičné zdroje dát.

V roku, ktorý nebol príliš priaznivý pre demokratov, sa vyskytlo dosť rasistov, ktorí by pomohli vyhrať primárne alebo vyhrotiť všeobecné voľby.

Moja štúdia bola pôvodne odmietnutá piatimi akademickými časopismi. Mnohí z mojich kolegov, ak mi dovoľíte malú nespokojnosť, povedali, že bolo takmer nemožné veriť, že toľkí Američania stále v sebe prechovávajú rasizmus. Toto jednoducho nebolo v súlade s tým, čo ľudia tvrdili. Okrem toho Google vyhľadávania vyzerajú len ako bizarné súbory dát.

Keď sme sa teraz stali svedkami inaugurácie prezidenta Donalda J. Trumpa, moje zistenia zrazu vyzerajú hodnovernejšie.

Čím viac som študoval, tým viac som prichádzal na to, že Google má veľa informácií, ktoré prieskumom unikli. Tieto môžu byť nápomocné v porozumení volieb – a mnohých iných tém.

Existujú napríklad informácie o tom, kto v skutočnosti pôjde voliť. Viac než polovica obyvateľstva, ktorá nevolí, povie prieskumom tesne pred voľbami, že to majú v úmysle. Tým zmätú náš odhad účasti. Zatiaľ čo Google vyhľadávania „ako voliť“ alebo „kde voliť“ týždeň pred voľbami presne odhadnú, ktoré časti krajiny budú mať veľkú účasť vo voľbách.

Môže sa tam dokonca vyskytnúť informácia o tom, koho budú voliť. Dokážeme skutočne predpovedať, ktorého kandidáta budú

Ľudia voliť na základe ich vyhľadávaní? Samozrejme, že sa nemôžeme zamerať len na to, ktorí kandidáti sú najviac vyhľadávaní. Mnohí ľudia vyhľadávajú kandidáta, pretože ho majú radi. Ale podobné množstvo ľudí vyhľadáva kandidáta, pretože ho nenávidia. Na tomto podklade sme so Stuartom Gabrielom, profesorom financií na Kalifornskej univerzite v Los Angeles, našli prekvapujúce záchytné body o tom, akým spôsobom sa ľudia chystajú voliť. Veľké percento vyhľadávaní týkajúcich sa volieb obsahuje otázky s oboma menami kandidátov. Počas volieb v roku 2016 medzi Trumpom a Hillary Clintonovou niektorí ľudia vyhľadávali „Trump Clintonová prieskumy“. Ostatní vyhľadávali hlavné témy z „debaty Clintonová Trump“. V skutočnosti 12 percent vyhľadávaných otázok s „Trump“ taktiež obsahovalo slovo „Clintonová“. Viac než jedna štvrtina vyhľadávaných otázok s „Clintonová“ takisto obsahovala slovo „Trump“.

Zistili sme, že tieto zdanlivo neutrálne vyhľadávania nám vlastne môžu ponúknuť kľúč k tomu, kto podporuje ktorého kandidáta.

Ako? V poradí, v ktorom sa kandidáti vyskytujú vo vyhľadávaní. Naš prieskum poukazuje na to, že človek má tendenciu uviesť kandidáta, ktorého podporuje, ako prvého pri vyhľadávaní oboch kandidátov.

Kandidáti, ktorí sa zjavili ako prví vo viacerých vyhľadávaniach v predošlých troch voľbách, dostali najviac hlasov. Oveľa zaujímavejšie je, že poradie kandidátov vo vyhľadávaniach predpovedalo smerovanie určitého štátu.

Poradie kandidátov, v ktorom sú vyhľadávaní, takisto obsahuje informácie, ktoré môžu prieskumom verejnej mienky uniknúť. V roku 2012 vo voľbách medzi Obamom a republikánom Mittom Romneym, Nate Silver, virtuóznym štatistikom a novinárom, presne predpovedal výsledky vo všetkých päťdesiatich štátoch. Zistili sme však, že v štátoch, v ktorých bol Romney vyhľadávaný pred Obamom častejšie, Romney v skutočnosti zabodoval lepšie než Silver predpovedal. V štátoch,

ktoré častejšie zaradili Obamu pred Romneyho, Obama bodoval lepšie ako Silver predpovedal.

Tento ukazovateľ môže obsahovať viac takých informácií, ktoré prieskumy verejnej mienky nezachytili – jednak preto, že voliči buď klamú sami sebe, alebo je pre nich neprijemné odhaliť ich skutočné preferencie. Možno tým, že v roku 2012 tvrdili, že ešte nie sú rozhodnutí, ale tým, že neustále vyhľadávali „Romney Obama prieskumy“, „Romney Obama debaty“ a „Romney Obama voľby“, vlastne po celý čas plánovali voliť Romneyho.

Takže predpovedal Google Trumpa? Ešte nás čaká veľa práce – a potrebujem, aby sa ku mne pripojilo viac prieskumníkov – aby sme čo najlepšie vedeli použiť Google dáta na predpovedanie výsledkov volieb. Je to nová veda a máme k dispozícii iba zopár volieb s týmito dátami. Určite netvrdím, že sme už v bode – alebo dokonca, že niekedy budeme – kedy budeme môcť úplne vylúčiť prieskumy verejnej mienky ako pomocný nástroj na predpovedanie volieb.

Avšak na internete sa už určite mnohokrát vyskytli náznaky toho, že Trump dosiahne lepšie výsledky, než naznačovali prieskumy verejnej mienky. Počas všeobecných volieb sa vyskytli indície, že voľby dopadnú v prospech Trumpa. Americkí černosi tvrdili prieskumom, že sa zúčastnia hromadne vo veľkých množstvách, aby sa postavili Trumpovi. Ale Google vyhľadávania s volebnými informáciami v oblastiach husto obývaných černochoch boli oveľa nižšie. V deň volieb by Clintonovej výsledkom poškodila nízka účasť černochoch.

Dokonca sa vyskytli náznaky, že údajne nerozhodní voliči podporia Trumpa. S Gabrielom sme zistili, že viac ľudí vyhľadávalo v kľúčových štátoch stredozápadu, kde Clintonová mala vyhrať, „Trump Clintonová“ než „Clintonová Trump“. Trump naozaj vďačil za svoje zvolenie faktu, že v tejto oblasti značne prevýšil prieskumy verejnej mienky. No ja si dovoľím tvrdiť, že hlavná indícia, že Trump sa

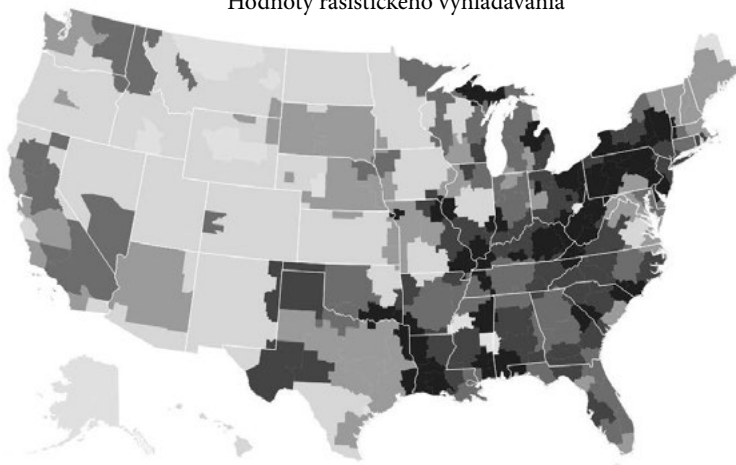
preukáže ako úspešný kandidát – v primárkach – bol v prvom rade tajný rasizmus odhalený štúdiou o Obamovi. Google vyhľadávania ukázali temnotu a nenávisť medzi významným počtom Američanov. A tieto vedcom dlhé roky unikali. Vyhľadávanie dát prezradilo, že žijeme v úplne inej spoločnosti, akú si na základe prieskumov verejnej mienky akademici a novinári predstavovali. Odhalilo ohavný, desivý a rozšírený hnev, ktorý čakal len na kandidáta, ktorý tomu hnevu dá hlas.

Ludia často klamú – sami sebe a aj druhým. V roku 2008 Američania povedali prieskumom, že už ich vôbec nezaujímá ľudská rasa. O osem rokov neskôr zvolili Donalda J. Trumpa, človeka, ktorý prevzal na Twitteri falošné tvrdenie, že černosi sú zodpovední za väčšinu vražd bielych Američanov. Človeka, ktorý sa zastal svojich stúpcov zodpovedných za výtržníctva počas manifestácie hnutia Black Lives Matter (Na životoch černochoch záleží), a ktorý váhal s odmietnutím podpory predošlému vodcovi Ku-Klux-Klanu. Ten istý skrytý rasizmus, ktorý ublížil Barackovi Obamovi, pomohol Donaldovi Trumpovi.

V začiatkovej fáze primárok Nate Silver slávnostne vyhlásil, že Trump v žiadnom prípade nevyhrá. Ako primárky postupovali a začalo byť jasné, že Trump má širokú podporu, Silver sa rozhodol preskúmať dáta, aby porozumel tomu, čo sa presne deje. Ako bolo možné, že sa Trumpovi tak darilo?

Silver si všimol, že oblasti, kde sa Trumpovi najviac darilo, vytvorili zvláštnu mapu. Darilo sa mu na severovýchode a priemyselnom stredozápade a aj na juhu. Na západe si však počínal nápadne horšie. Silver hľadal parametre na vysvetlenie tejto mapy. Bola to nezamestnanosť? Bola to viera? Bolo to vlastníctvo zbraní? Bolo to množstvo imigrantov? Bola to opozícia voči Obamovi?

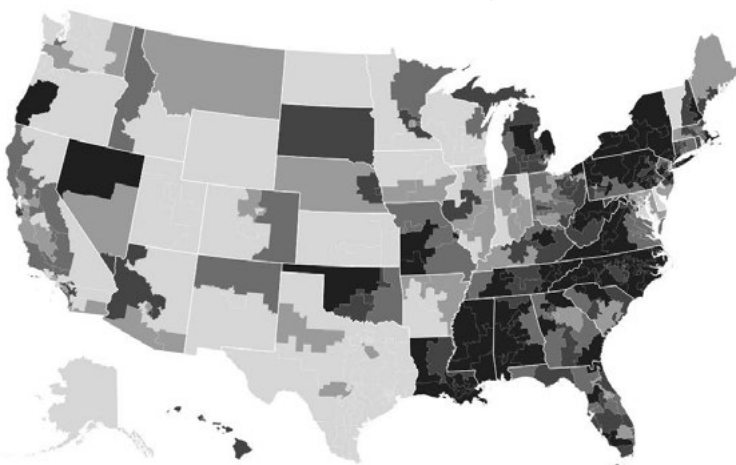
Hodnoty rasistického vyhľadávania



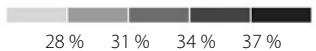
Objem vyhľadávania



Podpora Donalda Trumpa v republikánskych primárkach



Odhad republikánskych voličov na podporu Trumpa



Silver zistil, že jediný faktor, ktorý najlepšie vystihol podporu Donalda Trumpa v republikánskych primárkach, bolo kritérium, ktoré som objavil pred štyrmi rokmi. Oblasti, ktoré vysoko podporovali Trumpa, boli tie, ktoré najviac vyhľadávali na Googli heslo „neger“.

Posledné štyri roky som takmer každý deň trávil analyzovaním Google dát – vrátane práce dátového analytika pre Google, ktorý si ma najal na základe môjho prieskumu rasizmu. Stále pokračujem v tomto analyzovaní dát ako komentátor a dátový žurnalista pre *New York Times*. Odhalenia neustále prichádzajú. Mentálne choroby, ľudská sexualita, zneužívanie detí, potraty, reklama, viera, zdravie. Nejde práve o zanedbateľné témy a tieto súbory dát, ktoré pred niekoľkými desiatkami rokov ešte neexistovali, ponúkli nové prekvapujúce perspektívy vo všetkých z nich. Ekonómovia a ostatní sociológovia vždy poľujú na nové zdroje dát, takže to poviem na rovinu: som presvedčený, že Google vyhľadávania sú tým najdôležitejším súborom dát o ľudskej psychike, ktoré boli kedy zhromaždené.

Tento súbor dát však nie je jediným nástrojom, ktorý internet priniesol pre porozumenie sveta. Čoskoro som si uvedomil, že existujú aj iné zlaté digitálne bane. Stiahol som si Wikipédiu, prezrel som si profily na Facebooku a prebehol Stormfront. Okrem toho mi PornHub, jedna z najväčších pornografických stránok na internete, sprostredkoval kompletne dáta o anonymných vyhľadávaniach a prezeraniach videí po celom svete. Inak povedané, ponoril som sa hlboko do toho, čo dnes nazývame „big data“ (veľké dáta). Dokonca som absolvoval pohovory s tuctom ďalších ľudí – akademikmi, dátovými novinármi a podnikateľmi – ktorí takisto skúmali tieto nové sféry. Mnohým z týchto štúdií sa budeme venovať neskôr.

Ale najskôr priznanie: Nebudem presne definovať, čo sú to big data. Prečo? Pretože je to dosť nejasný koncept. Aké veľké sú veľké? Je 18 462 pozorovaní malými dátami a 18 463 veľkými dátami? Uprednostňujem všeobecný pohľad na to, čo sa takto kvalifikuje. Zatiaľ čo