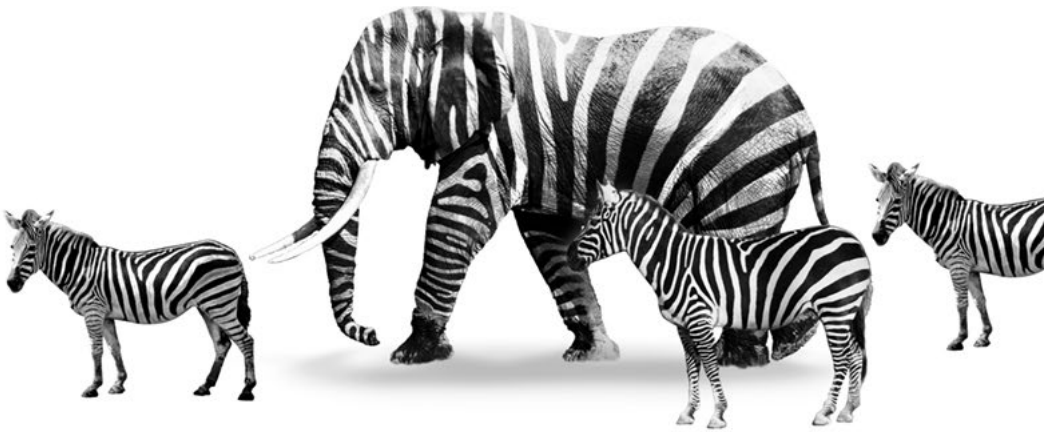
NEW YORK TIMES BESTSELLER

KAŽDÝ KLAME

BIG DATA, NEW DATA

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



SETH STEPHENS-DAVIDOWITZ

Lindeni

Každý klame

Vyšlo aj v tlačovej podobe

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

Lindeni

Seth Stephens-Davidowitz
Každý klame – e-kniha

ALBATROS MEDIA

KAŽDÝ KLAME

SETH STEPHENS-DAVIDOWITZ

KAŽDÝ KLAME

BIG DATA, NEW DATA

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

Lindeni

Mame a otcovi

OBSAH

PREDSLOV

Už od čias, kedy filozofi špekulovali o „cerebroskope" – mýtickom prístroji, ktorý by dokázal zobraziť ľudské myšlienky na obrazovke, spoločenskí vedci hľadali nástroje na odhalenie fungovania ľudskej povahy. Počas mojej kariéry experimentálneho psychológa mnohé z nich spopulárneli 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 spletitej, viacdimenzionálnej 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 obrazy, 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 štúdii 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 ponúkneme 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 rozmernosti. 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 neoľutovali, 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 nenávisti? 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 obsesií?

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 uraził 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 odradilo 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 černoch. 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 spoločenskí vedci využívali na porozumenie sveta viac ako 80 rokov, nám tvrdili, že prevažnú väčšinu Američanov pri zvažovaní prezidentského kandidáta nezaujímal fakt, že Obama bol černoch.

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 arogantný, 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 nezisťujú,

---

* Google Trendy bol zdrojom väčšiny mojich údajov. Keďže však povoľuje iba porovnávanie 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 štúdii v Časopise verejnej ekonomiky „Cena rasistických tendencií voči černošskému kandidátovi: Dôkazy využitia dát Google vyhľadávaní". Dizertáciu, spojivo medzi štúdiou 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 vyhľadávaní 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 vyhľadávaní 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 preháňajú 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 dovolím 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“ vyskytovalo zhruba v rovnakom množstve vyhľadávaní 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 vyhľadávaní so slovom „neger“ taktiež zahŕňalo slovo „vtip“. Iné bežné vyhľadávania obsahovali „hlúpy neger“ a „nenávidím negrov“.

Týchto vyhľadávaní 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 vyhľadávaní, ktoré obsahovalo slovo „Obama“, taktiež obsahovalo „kkk“** alebo „neger (negri)“. Možno sa to nezdá až tak veľa, ale zamyslite 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 vyhľadávaní 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 nepekné vyhľadávania nie je ťažké spojiť s úspechom kandidáta, ktorý – v jeho útokoch na imigrantov, v jeho hneve a odpore – hral na najhoršie ľudské sklony.

Google vyhľadávania taktiež zistili, že sme sa mýlili v miestach výskytu rasizmu. Prieskumy a všeobecná znalosť umiestnila rasizmus predovšetkým na juh a hlavne medzi republikánov. Ale miesta s najvyšším 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 vyhľadávania 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 vašej pôvodnej predstavy. Republikáni na juhu sa skôr priznajú k rasizmu ako mnohí demokrati na severe, aj keď majú takmer rovnaké postoje.

O štyri roky neskôr sa potvrdila dôležitosť tejto mapy pri vysvetľovaní politického úspechu Trumpa. V roku 2012 som využil túto mapu, ktorú som vytvoril pomocou Google vyhľadávaní, na prehodnotenie roly, ktorú zohrávala Obamova rasa. Dáta boli jednoznačné. V častiach krajiny s najvyšším počtom rasistických vyhľadávaní 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 dovolíte 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“ týždne 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. Náš 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, virtuózny štatistik a novinár, 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 nepríjemné 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 černochmi boli oveľa nižšie. V deň volieb by Clintonovej výsledkom poškodila nízka účasť černochov.

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 dovolím tvrdiť, že hlavná indícia, že Trump sa

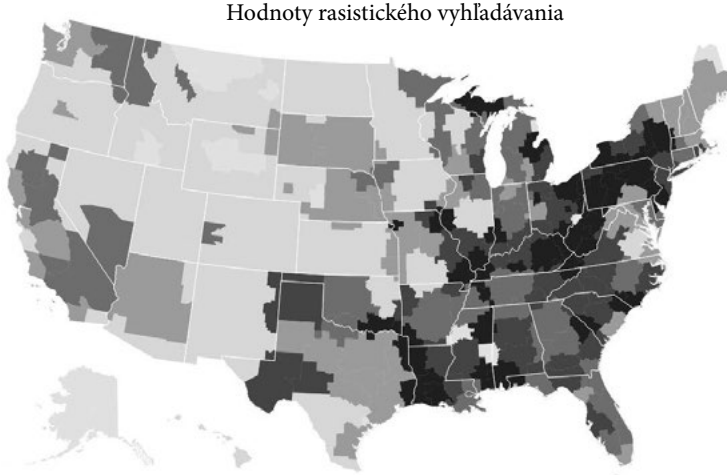
preukáže ako úspešný kandidát – v primárkach – bol v prvom rade tajný rasizmus odhalený štúdiou o Obamovi. Google vyhľadávania ukázali temnotu a nenávisť medzi významným počtom Američanov. A tieto vedcom dlhé roky unikali. Vyhľadávanie dát prezradilo, že žijeme v úplne inej spoločnosti, akú si na základe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

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

V začiatočnej 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?

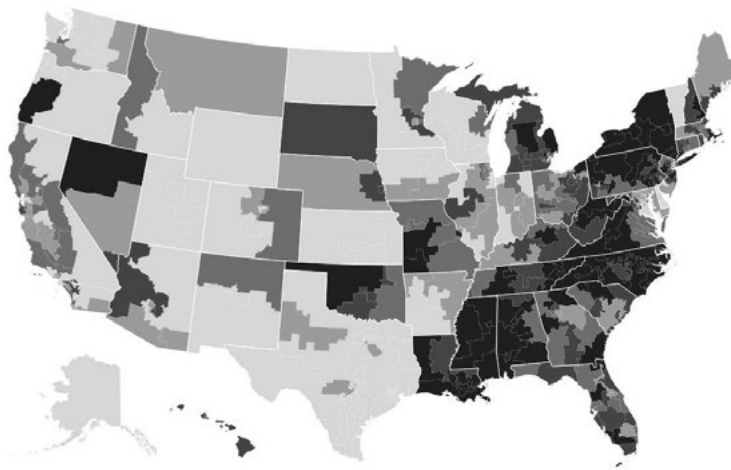
Hodnoty rasistického vyhľadávania



Objem vyhľadávania

Najnižších 20 % Najvyšších 20 %

Podpora Donalda Trumpa v republikánskych primárkach



Odhad republikánskych voličov na podporu Trumpa

28 % 31 % 34 % 37 %

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 kompletné 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

väčšina dát, s ktorými sa zahrávam, je z internetu, neskôr budem hovoriť aj o ďalších zdrojoch. Žijeme v čase informačnej explózie rôznej kvality a kvantity. Väčšina nových informácií prichádza z Googlu a zo sociálnych médií. Niektoré sú produktom digitalizácie informácií, ktoré boli predtým skryté v kabinetoch a fascikloch. Niektoré sú výsledkom zvýšenia zdrojov zameraných na prieskum trhu. Niektoré zo štúdií v tejto knihe vôbec nepoužívajú obrovské súbory dát, ale používajú iba nový a kreatívny prístup k dátam – prístup, ktorý je kritický v tejto ére prekypujúcich informácií.

Tak prečo sú big data také mocné? Zamyslime sa nad všetkými informáciami, ktoré sú za deň roztrúsené po internete – v skutočnosti vieme, máme čísla o tom, koľko informácií sa tam vyskytuje. Za priemerný deň na začiatku dvadsiateho prvého storočia ľudské bytosti vytvoria 2,5 milióna biliónov bajtov dát.

A tieto bajty sú záchytnými bodmi.

Žena sa vo štvrtok poobede nudí. Tak si zadá do Googlu „slušné zábavné vtipy". Skontroluje si svoj email. Prihlási sa na Twitter. Vyhľadá si na Googli „vtipy o negroch".

Mužovi je smutno. Vyhľadá si cez Google „symptómy depresie" a „depresívne príbehy". Zahrá si jednu hru solitér.

Žena uvidí oznam o zásnubách svojej kamarátky na Facebooku. Žena, ktorá je slobodná, svoju kamarátku zablokuje.

Muž si prestane vyhľadávať na Googli NFL, počúvať rapovú hudbu a opýta sa vyhľadávača otázku: „Je normálne, že sa mi sníva o bozkávaní mužov?"*

*Žena klikne na BuzzFeed** heslo odkazujúce na „15 najchutnejších mačiek".*

Muž vidí ten istý príbeh o mačkách, ale na jeho obrazovke sa volá „15 rozkošných mačiek". On naň neklikne.

---

* National Football League – Národná liga amerického futbalu – pozn. prekladateľa

** Americká stránka so správami a zábavou, pozn. prekladateľa

Žena si vyhľadá cez Google „Je môj syn génius?"

Muž si vyhľadá cez Google „Ako prinútiť moju dcéru schudnúť?"

Žena je na dovolenke so šiestimi najlepšími kamarátkami. Všetky jej kamarátky jej hovoria, že sa veľmi dobre bavia. Ona sa vyparí na Google a hľadá „Osamelá, keď som preč od manžela".

Manžel predošlej ženy je na dovolenke so šiestimi kamarátmi. Vyparí sa na Google a vyťuká „znamenia, že vás žena podvádza".

Niektoré z týchto dát budú zahŕňať informácie, ktoré by zvyčajne nikdy nikomu nepriznali. Keď ich všetky zhromaždíme, necháme ich v anonymite, aby sme sa poistili, že nikdy nebudeme vedieť o strachoch, túžbach a správaní určitých konkrétnych ľudí, a pridáme k nim vedecké dáta, dopracujeme sa k novému pohľadu na ľudské bytosti – ich správaniu, túžbam, ich povahe. Napriek riziku, že budem znieť pompézne, skutočne som začal veriť, že dáta, ktoré sú čoraz viac prístupné v našej digitálnej ére, radikálne rozšíria naše porozumenie ľudstvu. Mikroskop nám ukázal, že v kvapke vody z jazera je toho oveľa viac, ako dokážeme vidieť. Teleskop nám ukázal, že na nočnej oblohe je toho oveľa viac, ako dokážeme vidieť. A nové digitálne dáta nám teraz ukazujú o ľudskej spoločnosti viac, ako sme si mysleli, že vidíme. Môžu byť mikroskopom alebo teleskopom našej éry umožňujúcim dôležité, dokonca revolučné nahliadnutia.

Pri takýchto tvrdeniach však vyvstáva ďalšie riziko – nielenže budú vyznievať pompézne, ale aj trendovo. Mnohí ľudia vyjadrili veľké tvrdenia o sile veľkých dát, ale bez dostatku dôkazov. To podnietilo mnohých skeptikov veľkých dát, aby prestali vyhľadávať veľké súbory dát. „Netvrdím, že vo veľkých dátach nie sú žiadne informácie," napísal esejista a štatistik Nassim Taleb. „Informácií je dostatok. Problémom – hlavnou otázkou – je však, že táto ihla sa nachádza vo veľmi veľkej kope slamy."

Jedným zo základných cieľov tejto knihy je teda zabezpečiť chýbajúce dôkazy o tom, na čo je možné použiť big data – ako dokážeme

nájsť ihly, teda ak ich vôbec nájdeme – v neustále narastajúcich kopách slamy. Dúfam, že poskytnem dostatočné množstvo príkladov veľkých dát, a budem tak schopný ponúknuť nový náhľad do ľudskej psychológie a správania, aby ste dokázali pochopiť hlavné črty niečoho skutočne prevratného.

Možno si práve teraz hovoríte: „Počkaj, Seth. Ty sľubuješ novú revolúciu. Poeticky tu opisuješ tieto nové big data. Ale doteraz si používal tieto úžasné, výnimočné, dych vyrážajúce, priekopnícke dáta, aby si nám povedal len dve základné veci: v Amerike je veľa rasistov a ľudia, hlavne muži, preháňajú v tom, koľko majú sexu."

Pripúšťam, že niekedy nové dáta iba potvrdzujú niečo, čo je očividné. Ak si myslíte, že tieto výsledky boli jasné, počkajte, až sa dostanete ku kapitole 4. Tu vám jasne ukážem hodnoverné dôkazy z Google vyhľadávaní, že muži sa veľmi obávajú a sú si neistí v otázke – počkajte si na to – veľkosti svojho penisu.

Dovolím si tvrdiť, že určitá hodnota spočíva aj v dokazovaní vecí, ktoré ste už tušili, ale nemali ste dostatok dôkazov.

Vyjadriť podozrenie je jedna vec, dokázať ho je druhá. Ale ak by veľké dáta dokázali potvrdiť vaše podozrenia, tak by to nebolo revolučné. Chvalabohu, veľké dáta dokážu oveľa viac. Z času na čas dáta ukážu, že svet funguje úplne opačným spôsobom, ako som si myslel. Tu sú určité príklady, ktoré vás môžu prekvapiť.

Mohli by ste si myslieť, že hlavným dôvodom rasizmu je ekonomická neistota a zraniteľnosť. Mohli by ste prirodzene očakávať, že keď ľudia stratia prácu, rasizmus sa zvýši. No v skutočnosti sa ani rasistické vyhľadávania ani členstvo v Stormfronte nezvýšia, keď sa zvýši nezamestnanosť.

Mohli by ste si myslieť, že úzkosť je najvyššia v mestách s vyššou vzdelanostnou úrovňou. Pojem mestský neurotik je známym stereotypom. Ale Google vyhľadávania odrážajúce úzkosť – ako napríklad „symptómy úzkosti" alebo „pomoc v úzkosti" – sú skôr vyššie

v miestach s nižším stupňom vzdelania, nižšími príjmami a tam, kde väčšina populácie žije na vidieku. Vyššie hodnoty vyhľadávania úzkosti sa vyskytujú vo vidieckej, severnej časti New Yorku ako v New York City.

Mohli by ste si myslieť, že teroristický útok, ktorý zabil tucty alebo stovky ľudí, bude nasledovaný rozsiahlou úzkosťou. Terorizmus by mal, podľa definície, spôsobiť pocit teroru. Preskúmal som Google vyhľadávania na heslo úzkosť. Testoval som, koľko z týchto vyhľadávaní sa zvýšilo v krajine niekoľko dní, týždňov a mesiacov po každom závažnom európskom alebo americkom teroristickom útoku od roku 2004. Takže o koľko sa v priemere zvýšili vyhľadávania týkajúce sa úzkosti? Nezvýšili sa. Vôbec.

Mohli by ste si myslieť, že ľudia vyhľadávajú vtipy oveľa častejšie, keď sú smutní. Mnohí najväčší myslitelia tvrdili, že sa zameriame na humor, aby sme sa zbavili bolesti. Humor sa už dlho považuje za spôsob, ako sa vysporiadať s frustráciou, bolesťou, neodvratnými sklamaniami života. Ako to povedal Charlie Chaplin: „Smiech je tonikom, úľavou, tabletkou od bolesti."

Vyhľadávania vtipov sú však najnižšie v pondelok, v deň, keď ľudia tvrdia, že sú najmenej šťastní. Sú najnižšie v zamračenom a daždivom počasí. A najviac klesnú po vážnej tragédii, ako vtedy, keď dve bomby zabili troch ľudí a zranili stovky počas Bostonského maratónu. Vtipy sú skôr vyhľadávané, keď sa ľuďom darí.

Niekedy nové súbory dát odkryjú správanie, túžby alebo obavy, ktoré by som inak vôbec nevzal do úvahy. Do tejto kategórie spadajú mnohé sexuálne sklony. Napríklad, vedeli ste, že v Indii je číslo jeden vo vyhľadávaní výrazov začínajúcich sa „môj manžel chce..." výraz „môj manžel chce, aby som ho dojčila"? Táto poznámka je oveľa častejšia v Indii než inde na svete. A dokonca pornografické vyhľadávanie žien dojčiacich mužov je štyrikrát vyššie v Indii a Bangladéši ako v iných krajinách sveta. Nepomyslel by som si to predtým, ako som videl dáta.

Okrem toho, ak fakt, že muži sú posadnutí veľkosťou svojho penisu, nie je prekvapujúci, najväčšia neistota žien súvisiaca s ich telom vyjadrená cez Google je skutočne prekvapujúca. Problémom žien, zodpovedajúcim veľkosti penisu u mužov, je podľa nových dát – a tu sa zastavím, aby som zvýšil napätie – obava, či im zapácha vagína. Ženy vyhľadávajú záležitosti týkajúce sa genitálií tak často ako muži. A najväčším problémom je pre ženy ich zápach – a ako ho môžu zlepšiť. Toto som určite netušil.

Niekedy nové dáta odhalia kultúrne rozdiely, ktoré som si nikdy neuvedomil. Takým príkladom je veľmi odlišný spôsob, akým muži z rôznych kútov sveta pristupujú k tomu, že ich manželka je tehotná. V Mexiku vyhľadávania zahŕňajúce výrazy „moja tehotná manželka" obsahovali „frases de amor para mi esposa embarazada" (slová lásky pre moju tehotnú manželku) a „poemas para mi esposa embarazada" (básničky pre moju tehotnú manželku). V Spojených štátoch vyhľadávania najviac zahŕňali „moja manželka je tehotná, čo teraz" a „moja manželka je tehotná, čo mám robiť".

Táto kniha je ale viac než len zbierkou zvláštnych faktov alebo ojedinelých štúdií, aj keď tých sa vyskytuje pomerne dosť. Pretože sú tieto metódy také nové, a stanú sa ešte oveľa mocnejšími, predstavím niektoré myšlienky o tom, ako fungujú a prečo sú také prevratné. A takisto zareagujem na nedostatky veľkých dát.

Časť entuziazmu z revolučného potenciálu dát bol nevhodne zameraný. Väčšina milovníkov veľkých dát sa rozplývala nad tým, aké úžasné môžu tieto súbory dát byť. Posadnutosť veľkosťou súborov dát nie je nová. Predtým ako Google, Amazon a Facebook a termín „big data" existovali, uskutočnila sa v Dallase, v Texase, konferencia na tému „Značné a komplexné súbory dát". Jerry Friedman, profesor štatistiky v Stanforde, ktorý bol mojím kolegom, keď som pracoval v Googli, si pamätá túto konferenciu z roku 1977. Jeden uznávaný štatistik sa postavil a prehovoril. Vysvetľoval, že nazhromaždil úžasných až prekvapujúcich päť gigabajtov dát. Ďalší uznávaný štatistik sa

postavil a povedal: „Predošlý rečník mal gigabajty. To je nič. Ja mám terabajty." Dôraz v ich diskusii kládli inými slovami na to, koľko informácií môžete nazhromaždiť a nie na to, čo s nimi dokážete urobiť, alebo na akú otázku hľadáte odpovede. „V tom čase sa mi zdalo zábavné," povedal Friedman, „že to, čo vás malo skutočne ohromiť, bol fakt, aký veľký súbor dát vlastnia. Stále sa to stáva."

Príliš veľa dátových analytikov dnes zhromažďuje masívne súbory, ktoré nám oznamujú nepodstatné informácie – ako napríklad, že Knicks* sú populárni v New Yorku. Príliš veľa firiem sa topí v dátach. Vlastnia veľa terabajtov, ale s veľmi málo náhľadmi. Veľkosť súboru dát je mnohokrát preceňovaná, na čo existuje drobné, ale zásadné vysvetlenie. Čím väčší efekt, tým je potrebné menšie množstvo pozorovaní na jeho pochopenie. Na to, aby ste vedeli, že horúci sporák je nebezpečný, sa ho potrebujete chytiť iba raz. Možno budete musieť vypiť kávu tisíckrát, aby ste pochopili, že vám spôsobuje bolesť hlavy. Ktorá lekcia je dôležitejšia? Jednoznačne tá o horúcom sporáku, ktorý sa kvôli sile dôsledku preukáže veľmi rýchlo, aj keď len s použitím malého množstva dát.

V skutočnosti najchytrejšie dátové firmy často redukujú množstvo svojich dát. V Googli sú hlavné rozhodnutia založené len na malom množstve dát. Nie vždy potrebujete obrovské množstvo dát na to, aby ste zistili dôležité fakty. Nevyhnutné sú správne dáta. A hlavným dôvodom, prečo sú Google vyhľadávania také hodnotné, nie je fakt, že ich je tak veľa, ale fakt, že ľudia sú v nich úprimní. Ľudia klamú svojim kamarátom, milencom, lekárom, prieskumom, a aj samým sebe. Ale na Googli sa skôr podelia so zahanbujúcou informáciou, akou je napríklad ich manželstvo bez sexu, psychické problémy, neistoty a nepriateľstvo voči černochom.

Najdôležitejšie je, že ak chcete získať fakty z veľkých dát, musíte klásť tie správne otázky. Presne tak, ako nemôžete nasmerovať

* Newyorský basketbalový tím, pozn. prekladateľa.

teleskop na nočnú oblohu a očakávať, že objavíte Pluto, nemôžete si stiahnuť len veľké množstvo dát a očakávať, že objavíte tajomstvá ľudskej povahy. Musíte hľadať na vhodných miestach – Google vyhľadávania, ktoré začínajú „môj manžel chce...“ v Indii, napríklad.

Táto kniha vám ukáže, ako veľké dáta najlepšie využívať, a detailne vysvetlí, prečo môžu byť také vplyvné. A popritom zistíte to, čo som ja a mnohí iní počas tohto procesu objavil, vrátane:

- Koľkí muži sú homosexuáli?
- Funguje reklama?
- Prečo bol American Pharoah úžasným pretekárskym koňom?
- Sú médiá predpojaté?
- Sú Freudove pošmyknutia skutočné?
- Kto podvádza na daniach?
- Záleží na tom, na akú strednú školu chodíte?
- Môžete poraziť trh s cennými papiermi?
- Kde je najlepšie miesto na výchovu detí?
- Čo spôsobí rýchle šírenie príbehu?
- O čom by ste sa mali rozprávať na prvom rande, ak chcete mať aj druhé?

...a ešte oveľa, oveľa viac. Ale pred tým, ako sa k tomu dostaneme, musíme prediskutovať dosť základnú otázku: načo vôbec potrebujeme dáta. A preto vám predstavím moju starú mamu.

ČASŤ I

DÁTA, VEĽKÉ A MALÉ

1

VÁŠ CHYBNÝ POCIT V ŽALÚDKU

Ak máte tridsaťtri rokov a niekoľkokrát po sebe ste sa zúčastnili večerí Vďakyvzdania bez partnera, je pravdepodobné, že sa bude diskutovať aj na tému o výbere partnera a určite sa k nej všetci vyjadria.

„Seth potrebuje strelené dievča ako je on," vraví moja sestra.

„Ty si strelená! Potrebuje normálne dievča, ktoré ho umierni," vraví môj brat.

„Seth nie je strelený," vraví moja mama.

„Ty si strelená! Samozrejme, že Seth je strelený," vraví môj otec.

A zrazu prehovorí jemným hlasom moja hanblivá stará mama, ktorá bola počas celej večere ticho. Agresívne newyorské hlasy stíchnu a všetky oči sa sústredia na túto drobnú dámu s krátkymi blonďavými vlasmi a nepatrným východoeurópskym prízvukom. „Seth, ty potrebuješ dobré dievča. Nie príliš pekné. Veľmi inteligentné. Také, ktoré to vie s ľuďmi. Spoločenské. So zmyslom pre humor, pretože ty máš dobrý zmysel pre humor."

Prečo rada tejto starej ženy disponuje v mojej rodine takou pozornosťou a rešpektom? Pretože moja osemdesiatosemročná stará mama toho v živote videla a zažila oveľa viac ako ktokoľvek iný za stolom. Bola svedkom mnohých manželstiev, tých, čo fungovali, ale aj tých, čo zlyhali. Počas desaťročí postupne zozbierala poznatky

o kvalitách úspešných vzťahov. Pri stole na Vďakyvzdanie mala tak k tejto otázke najväčšie množstvo údajov. Moja stará mama je zdrojom veľkých dát.

V tejto knihe by som rád odhalil tajomstvá dátovej vedy. Či sa vám to páči alebo nie, dáta zohrávajú v našich životoch veľmi dôležitú úlohu – a tá sa bude len zväčšovať. Noviny teraz majú samostatnú sekciu zameranú na dáta. Firmy majú tímy so špeciálnou úlohou analyzovania vlastných dát. Investori dávajú startupom desiatky miliónov dolárov, ak dokážu uložiť viac dát. Dokonca aj keď sa nikdy nenaučíte, ako dosiahnuť obrat alebo vypočítať interval spoľahlivosti, stretnete sa s mnohými dátami – na stránkach, ktoré čítate; na firemných mítingoch, ktorých sa zúčastnite; v klebetách, o ktorých sa dozviete, keď si budete dopĺňať vodu z dávkovačov pitnej vody.

Mnohí ľudia sú z tohto vývoja nervózni. Sú zastrašení veľkými dátami, ľahko sa v nich stratia a svet čísiel ich mätie. Myslia si, že kvantitatívne porozumenie sveta je len pre zopár vyvolených géniov. Hneď ako sa stretnú s číslami, radšej obrátia stranu, skončia míting alebo zmenia tému.

Ja som však strávil desiatky rokov vo firmách analyzovaním dát, a mal som šťastie pracovať s mnohými špičkovými odborníkmi. Jedna z najdôležitejších lekcií, ktorú som sa naučil, je: Kvalitná dátová veda je menej komplikovaná, ako si ľudia myslia. V skutočnosti najlepšia dátová veda je prekvapivo intuitívna.

Prečo je dátová veda intuitívna? Vo svojej podstate dátová veda je založená na odhaľovaní opakujúcich sa vzorov a predpovedí toho, ako jedna premenná ovplyvňuje druhú. Presne to, čo ľudia vlastne robia celý čas.

Spomeňte si, ako mi moja stará mama radila o vzťahu. Využila širokú databázu vzťahov, ktorú jej mozog nazhromaždil počas takmer storočia jej života, z príbehov rodiny, priateľov a známych. Svoje analýzy obmedzila na vzorku vzťahov, v ktorých muži mali podobné kvality ako ja – citlivý temperament, sklon izolovať sa a zmysel pre

humor. Zamerala sa na kľúčové kvality ženy – aká bola milá, aká bola inteligentná a aká bola pekná, a dala ich do súladu s kľúčovými kvalitami vzťahu – ak bol dobrý. Nakoniec oznámila svoje výsledky. Inými slovami, odpozorovala opakujúci sa vzor a predpovedala, ako jedna premenná ovplyvní druhú. Stará mama je analytička dát.

Vy ste tiež analytikmi dát. Keď ste boli malí, všimli ste si, že keď ste plakali, mama na vás upriamila pozornosť. To je dátová veda. Keď ste dospeli, zistili ste, že ak sa príliš sťažujete, ľudia s vami trávia menej času. Dátová veda. Všimli ste si, že keď ľudia s vami trávia menej času, ste menej šťastní. Keď ste menej šťastní, ste menej priateľskí. Keď ste menej priateľskí, ľudia s vami chcú tráviť ešte menej času. Dátová veda. Dátová veda. Dátová veda. Pretože Dátová veda je taká prirodzená, najlepšie štúdie veľkých dát, ktoré som našiel, sú zrozumiteľné takmer každému inteligentnému človeku. Ak nedokážete porozumieť štúdii, problém je pravdepodobne v štúdii, a nie vo vás.

Chcete dôkaz o tom, že najlepšia dátová veda má sklon k tomu byť intuitívna? Nedávno som natrafil na štúdiu, ktorú je možné považovať za jednu z najdôležitejších za posledné roky. Je taktiež jednou z najintuitívnejších, akú som kedy videl. Chcem, aby ste sa zamysleli nielen nad jej dôležitosťou – ale aj nad tým, aká je prirodzená a podobná starej mame.

Štúdia bola uskutočnená tímom výskumníkov z Kolumbijskej univerzity a Microsoftu. Tento tím chcel zistiť príznaky, ktoré predpovedajú rakovinu pankreasu. Táto choroba má nízku mieru prežitia prvých piatich rokov – len 3 percentá – ale jej skoré zistenie môže zdvojnásobiť pacientove šance.

Metóda výskumníkov? Využili dáta od desaťtisícov anonymných užívateľov Bingu, internetového vyhľadávača Microsoftu. Klasifikovali užívateľa ako nedávno diagnostikovaného na rakovinu pankreasu na základe neomylných vyhľadávaní ako „práve diagnostikovaný na rakovinu pankreasu“ alebo „Bolo mi oznámené, že mám rakovinu pankreasu, čo môžem očakávať“.

Následne sa výskumníci zamerali na vyhľadávania zdravotných príznakov. Porovnali toto malé množstvo užívateľov, ktorí neskôr oznámili svoju diagnózu, s tými, ktorí ju neoznámili. Inak povedané, ktoré príznaky predpovedali, že o pár týždňov alebo mesiacov používateľ oznámi diagnózu?

Výsledky boli pozoruhodné. Ukázalo sa, že vyhľadávanie bolestí chrbta a následne žltnúcej pokožky boli príznakmi rakoviny pankreasu; ale samotné vyhľadávanie bolesti chrbta ešte neznamenalo, že niekto mal rakovinu pankreasu. Podobne, vyhľadávania na zlé trávenie a následné bolesti žalúdka bolo takisto príznakom rakoviny pankreasu, kým samostatné vyhľadávanie bolesti žalúdka nie. Výskumníci tak dokázali identifikovať 5 – 15 percent prípadov. To ešte nemusí znieť ako úžasné množstvo, ale ak máte rakovinu pankreasu, dokonca aj 10-percentná šanca možného zdvojnásobenia prežitia vám bude pripadať ako šťastie. Práca popisujúca túto štúdiu by bola ťažko zrozumiteľná pre laikov. Obsahuje veľa technického žargónu, ako Kolmogorovov-Smirnovov test, ktorého význam, musím sa priznať, som už zabudol. (Je to spôsob určenia, či model správne zodpovedá dátam.)

Všimnite si však, aká prirodzená a intuitívna je táto mimoriadna štúdia v jej základnej podstate. Výskumníci sa zamerali na širokú škálu zdravotných prípadov a snažili sa ich spojiť so symptómami špecifickej choroby. Viete, kto ešte používa túto metódu na zistenie, či je niekto chorý? Manželia a manželky, matky a otcovia, zdravotné sestričky a lekári. Na základe skúseností a vedomostí sa snažia spojiť horúčky, bolesti hlavy, nádchy, bolesti žalúdka s rozličnými chorobami. Inak povedané, výskumníci z Kolumbijskej univerzity a Microsoftu napísali prelomovú štúdiu využitím prirodzenej, samozrejmej metodiky, ktorú používa každý na určenie zdravotnej diagnózy.

Ale, počkajte! Spomaľme. Ak je metodika najlepšej dátovej vedy často prirodzená a intuitívna, ako tvrdím, tak je potrebné položiť si základnú otázku o hodnote veľkých dát. Ak sú ľudské bytosti od