

Seth Stephens-Davidowitz

# **MINDENKI HAZUDIK**

**AZ VAGY, AMIRE KLIKKELSZ**

FORDÍTOTTA SZIEBERTH ÁDÁM

**Seth Stephens-Davidowitz**

# **MINDENKI HAZUDIK**

**AZ VAGY, AMIRE KLIKKELSZ**

ATHENAEUM

A fordítás alapjául szolgáló mű  
Seth Stephens-Davidowitz: Everybody Lies

Copyright © 2017 by Seth Stephens-Davidowitz  
All rights reserved.

Hungarian translation @ Szieberth Ádám, 2019

Minden jog fenntartva.

Kiadta az Athenaeum Kiadó,  
az 1795-ben alapított Magyar Könyvkiadók  
és Könyvterjesztők Egyesülésének tagja.  
1086 Budapest, Dankó u. 4–8.  
Tel.: 1-235-5030  
[www.athenaeum.hu](http://www.athenaeum.hu)  
[www.facebook.com/athenaeumkiado](https://www.facebook.com/athenaeumkiado)  
[athenaeumkiado2012@gmail.com](mailto:athenaeumkiado2012@gmail.com)

ISBN 978 963 293 868 4

Felelős kiadó: Szabó Tibor Benjámín  
Felelős szerkesztő: Besze Barbara  
Szerkesztette: Filippov Gábor  
Műszaki vezető: Drótos Szilvia  
Borítóterv: Földi Andrea  
Nyomdai előkészítés: Tóth Viktor

*Anyunak és Apunak*



## Előszó

A filozófusok már régóta törik a fejüket a „cerebroszkóp” nevű mitikus készüléken, amely képes lenne képernyőre vetíteni egy ember gondolatait; a társadalomtudósok pedig legalább ilyen régóta kutakodnak olyan eszközök után, amelyek képesek feltárni az emberi természet működését. Kísérleti pszichológusi pályám során különböző módszerek jöttek divatba, majd lettek ejtve, én pedig kipróbáltam mindet: az értékelőskálákat, a reakcióidő, illetve a pupilatágulás mérését, a funkcionális agyi képalkotást, sőt még az epilepsziásokba beépített elektródákat is – a páciensek szívesen múltatták az időt egy-egy nyelvi kísérlet alanyai-ként, amíg a következő rohamukat várták...

Mindazonáltal ezek közül egyik módszer sem enged korlátlan betekintést az elme működésébe. A problémát az a brutális „vagy-vagy” dilemma okozza, hogy egy emberi gondolat rendszerint komplex állítás: nem csupán annyit gondolunk, hogy „valami oroszokról szolt”, mint Woody Allen, miután gyorsolvasással átfutotta a *Háború és békét*. Csakhogy a tudós számára igen kemény dió egy teljes fényében ragyogó, keszekusza, többdimenziós állítás elemzése. Persze ha valaki kiönti a szívét, érzékeljük a „tudatfolyam” végtelen gazdagságát, de a monológ korántsem ideális adathalmaz, ha tesztelni akarunk egy hipotézist.

Ha viszont azt tartjuk szem előtt, hogy könnyen kvantifikálható adatokat kapjunk, például a különböző szavak elhangzása utáni reakcióidőt vagy a képek által kiváltott bőrreakciót mérjük, akkor készíthetünk ugyan statisztikát, de a megismerés komplex szövetét egyetlen számmá pépesítjük. Még a legszofisztikáltabb agyi képalkotó módszerek is csak azt árulják el, hogy miként „nyúlik el” egy gondolat a háromdimenziós térben, de azt már nem, hogy miből áll össze az a bizonyos gondolat.

Mintha nem lenne épp elég bajuk a kezelhetőség kont-ra adatgazdagság dilemmájával, az emberi természet kutatóinak a „kis számok törvénye” is komoly fejfájást okoz. Amos Tversky és Daniel Kahneman nevezte el így azt a tévhitet, amely szerint egy populáció jellemző vonásait hűen tükrözi bármely, mégoly csekély egyedszámú minta. A leginkább a számokra összpontosító tudósok intuíciója is csődöt mond, ha azt kell megítélniük, hány alanyra van szükség egy tanulmányhoz annak érdekében, hogy a véletlenszerű egyéni sajátosságoktól és esetlegességektől elrugaszkodva, minden amerikaira, pláne minden *Homo sapiens*re érvényes általánosításokat lehessen levonni. Még kétségesebb a siker, ha a mintavétel során kényelmi szempontok érvényesülnek, például sörpénzt ajánlunk fel a kurzusainkra járó másodéveseknek.

Ez a könyv az emberi elme tanulmányozásának mérően újszerű módjáról szól. Az internetes keresésekből, illetve más online megnyilatkozásokból kinyert, irdatlan méretű adathalmaz, a „Big Data” nem cerebroszkóp, de Seth Stephens-Davidowitz rávilágít, hogy eddig példátlan mélységű bepillantást enged az emberek pszichéjébe. Az emberek a billentyűzet előtti magányukban roppant fura dolgokat ismernek be – olykor (például randioldalakon



vagy ha szakmai tanácsokra keresnek rá) azért, mert ennek van következménye a való életben, máskor pedig éppen azért, mert *nincs* következmény: úgy könnyíthetnek a lelkükön, úgy vallhatják meg vágyaikat, félelmeiket, hogy azzal egyetlen valódi személyből sem váltanak ki döbbenetet, rémületet vagy még rosszabbat. De akár van következmény, akár nincs, az emberek ilyenkor nem csupán megnyomnak egy gombot vagy elfordítanak egy kart, hanem begépelik a sok-sok trillió lehetséges karaktersorozat egyikét, hogy körvonalazzák gondolataikat a maguk robbanásveszélyes, kombinatorikus, mérhetetlen valójában. Sőt mi több, ezeket a digitális nyomokat könnyen összesíthető és analizálható formában terítik. Az élet legkülönbözőbb területeiről jönnek, és közben észrevétlen kísérletek alanyivá válhatnak, amelyek során mind az ingerek módosítása, mind a válaszok csoportosítása valós időben történik. Ők pedig önként és dalolva szolgáltatják ezt az irdatlan mennyiségű adatot.

*A Mindenki hazudik* nem csupán egy elmélet igazolása: Stephens-Davidowitz felfedezései újra meg újra a fejük tetejére állították a hazámmal, illetve az emberi nemmel kapcsolatos prekonceptióimat. Honnan jött Donald Trump váratlanul magas támogatottsága? Amikor Ann Landers 1976-ban azt kérdezte olvasóitól, nem bánták-e meg, hogy gyereket vállaltak, és sokkolta, hogy a többség megbánta, lehet, hogy az önkényesen összeválogatott, nem reprezentatív minta vezette félre? Az internet okolható a kései 2010-es évek kissé redundáns elnevezésű válságáért, a „szűrőbuborékért”? Mi váltja ki a gyűlölet-bűncselekményeket? Azért keresnek vicceket az emberek, hogy felvidítsák magukat? És bár szeretem azt gondolni, hogy engem már semmi sem tud sokkolni, bőven sokkolt az, amit az

internet az emberi szexualitásról felfed: beleértve azt a felfedezést, hogy meghatározott számú nő minden hónapban rákeres arra, hogy „dugás plüssállatokkal”. Ezt az adatot soha nem hozták volna felszínre sem a reakcióidőt vagy a pupillatágulást mérő kísérletek, sem pedig a funkcionális agyi képalkotás.

A *Mindenki hazudik*ot mindenki élvezni fogja. Stephens-Davidowitz lankadatlan kíváncsiságtól fűtött, megnyerően okos könyve új utat mutat a 21. századi társadalomtudománynak. Ennek az emberi rögeszmékre nyíló, végtelenül izgalmas ablaknak a kitárása után ugyan kinek van szüksége cerebrószkópra?

*Steven Pinker, 2017*

# Bevezetés

## Egy forradalom körvonalai

Holtbiztos, hogy veszteni fog – mondták.

A 2016-os republikánus előválasztás előtt a közvéleménykutatók arra jutottak, hogy Donald Trumpnak az égvilágon semmi esélye. Trump mégiscsak számos kisebbségi csoportot megsértett. A felmérések és azok értelmezői azt mondták nekünk, hogy az efféle durvaságot csak kevés amerikai tartja elfogadhatónak.

Akkoriban a szakértők nagy többsége úgy vélte, hogy Trump elveszíti az elnökválasztást. A szavazni készülőkhöz túl sokan állították azt, hogy elriasztja őket a modora és a nézetei.

Pedig valójában voltak előjelei annak, hogy Trump mind az előválasztást, mind az elnökválasztást megnyerheti. Bizony, voltak – az interneten.

Internetes adatszakerítő vagyok. Nap mint nap követem a weben járó-kelő emberek digitális lábnyomait. A lenyomott-leütött gombok, billentyűk segítségével próbálom megérteni, mit is akarunk valójában, mit is teszünk valójában, kik is vagyunk valójában. Hadd magyarázzam el, hogy vágtam neki ennek a szokatlan útnak.

A történet a 2008-as elnökválasztással (mintha már ezer éve lett volna!), illetve egy, a társadalomtudósok által rég-

óta vitatott kérdéssel kezdődik. A kérdés így szól: mennyire jelentős Amerikában a faji előítélet?

Barack Obama volt minden idők első, valamely nagy párt által indított afroamerikai elnökjelöltje. Győzött – meglehetősen könnyedén. És a közvélemény-kutatási eredmények arra utaltak, hogy a faji hovatartozás nem játszott szerepet abban, hogyan szavaztak az amerikaiak. A Gallup például számos felmérést végzett Obama első megválasztása előtt és után. Az eredmény? Az amerikai szavazókat nagyrészt hidegen hagyta, hogy Barack Obama fekete. Nem sokkal az elnökválasztás után a Berkeley Egyetem két híres professzora más, szintén felméréseken alapuló adatokba ásta bele magát, jóval szofisztikáltabb adatbányászati technikákat alkalmazva. Ők is hasonló következtetésre jutottak.

Így aztán Obama elnöksége alatt a média nagy részében, illetve tudományos körökben ez közkeletű vélekedésé vált. Azok a források, amelyekre a médiamunkások és a társadalomtudósok már vagy nyolcvan éve támaszkodtak a világ megértésében, azt mondták nekünk, hogy az amerikaiak túlnyomó többségét nem érdekli, hogy Obama fekete, amikor azt latolgatják, hogy ő legyen-e az elnökük.

Úgy tűnt, ez az ország, amelyet oly sokáig szennyezett a rabszolgaság és a Jim Crow-törvények,\* végre felhagyott azzal, hogy a bőrük színe alapján ítélje meg az embereket. Úgy tűnt, a rasszizmus az utolsókat rúgja Amerikában. Sőt, egyes megmondóemberek odáig mentek, hogy kijelentették: posztrasszista társadalomban élünk.

\* Jim Crow-törvények: az amerikai polgárháborút követően számos déli államban életbe lépő, diszkriminatív jogszabályok, amelyek a formális jogegyenlőség ellenére, a szövetségi állam jóváhagyásával egészen az 1960-as évekig fenntartották a fehérek és afroamerikaiak szegregációját, többek között a közoktatásban, a tömegközlekedésben és a hivatalokban. (A szerk.)

2012-ben posztgraduális közgazdászhallgató voltam – elveszve az életben, kiegészve a szakmámban, abban a magabiztos, sőt beképzelt hitben, hogy meglehetősen jól értem a világ működését, hogy mi jár az emberek fejében, mi foglalkoztatja őket a 21. században. Ami pedig az előítéleteség problémáját illeti, pszichológiai és politikatudományi olvasmányaim alapján elhittem magammal, hogy a nyílt rasszizmus az amerikaiak csekély hányadára korlátozódik – akiknek többsége konzervatív, a „Deep South”-on, vagyis a legdélebbi amerikai államokban élő republikánus.

Aztán felfedeztem a Google Trendset.

A Google Trends, amelyet különösebb csinnadratta nélkül indítottak útjára 2009-ben, olyan eszköz, amelynek segítségével megtudhatjuk, milyen gyakorisággal kerestek rá az emberek egy-egy szóra, kifejezésre különböző helyeken és különböző időpontokban. Szórakoztató termékként reklámozták, amely arra való, hogy az ember megvitassa a barátaival, melyik celeb a legnépszerűbb, vagy mi jött hirtelen divatba. A legkorábbi verziókhöz egy vicces intelmet is mellékeltek, amely szerint „doktori disszertációt azért nem kellene írni” ezen adatok alapján: ami azonnal arra sarkallt, hogy ezek alapján írjam meg a disszertációm.\*

\* Adataim nagy részének forrása a Google Trends. Mivel azonban így csak a különböző keresések relatív gyakoriságát tudom összevetni, de egy-egy adott keresés abszolút száma nem derül ki belőle, az adatbázist általában kiegészítettem a Google AdWordszel, amely arról is tájékoztat, hogy pontosan milyen gyakran kerestek rá valamire. Az esetek többségében tovább tudtam élesíteni a képet a saját, Google Trends-alapú algoritmusom segítségével, amelynek leírása *Essays Using Google Data (Értekezések Google-adatok felhasználásával)* című disszertációmban, valamint a *Journal of Public Economics*ban publikált, *The Cost of Racial Animus on a Black Candidate: Evidence Using Google Search Data* (Mibe kerül egy fekete jelöltnek a fajgyűlölet? Bizonyítás Google-keresési adatok alapján) című tanulmányomban olvasható. A disszertáció, a cikk linkje, valamint az e könyvben bemutatott, eredeti kutatásaimhoz felhasznált adatok és kódolás részletes magyarázata megtalálható a honlapomon: [sethsd.com](http://sethsd.com).

A Google-keresések adatai akkoriban még nem tűntek megfelelő, azaz „komoly” tudományos kutatás céljára használható információforrásnak. A felmérésekkel ellentétben a Google-keresések adatait nem az emberi psziché megértését segítő eszköznek szánták. A Google-t arra találták ki, hogy az emberek mindenfélét megtudjanak a világról, nem arra, hogy a kutatók tudjanak meg mindenfélét az emberekről. De mint kiderült, a nyomok, amelyeket a tudást hajhászva hagyunk magunk után a neten, elképesztő mértékben árulkodók.

Más szóval: az emberek információkeresése önmagában is információ. Mint kiderült, az, hogy mikor és hol keresnek rá tényadatokra, idézetekre, viccekre, helyekre, személyekre, tárgyakra, illetve mikor és hol kérnek segítséget, sokkal többet árul el arról, hogy mit gondolnak valójában, mire vágynak valójában, mitől félnek valójában és mit tesznek valójában, mint azt bárki hitte volna. Különösen igaz ez azt figyelembe véve, hogy az emberek időnként nem is annyira faggatják a Google-t, mint inkább a bizalmukba fogadják: „Gyűlölöm a főnökömet.” „Részeg vagyok.” „Apám megütött.”

Az a mindennapos aktus, hogy egy-egy szót vagy kifejezést bepötyögünk egy kompakt, fehér téglalapba, az igazság apró nyomát hagyja maga után, amely idővel, megszorozódva, milliós nagyságrendben feltárja az élet legvalódibb tényeit. Az első szó, amelyet beírtam a Google Trendsbe, az „Isten” volt. Megtudtam, hogy a legtöbb olyan Google-keresés, amelyben szerepelt az „Isten” szó, Alabama, Mississippi és Arkansas államokhoz, vagyis a Bibliaövezethez\* köthető. Megtudtam továbbá, hogy az ilyen

\* Az Egyesült Államok déli, döntően konzervatív, főként neoprotestáns felekezetekhez tartozó területei. (A szerk.)

keresések vasárnaponként a leggyakoribbak. Mindez egyáltalán nem volt meglepő, azt viszont már érdekesnek találtam, hogy a keresési adatok ilyen tisztán kirajzolódó mintázatot tudnak feltárni. Ezután megpróbálkoztam a „Knicks” szóval, és kiderült, hogy arra New York Cityben keresnek rá a legtöbben. Itt sincs semmi látnivaló. Aztán beírtam a saját nevemet. „Sajnáljuk. Nincs elegendő mennyiségű keresés” – tájékoztattott a Google Trends. Máris megtudtam tehát, hogy a Google Trends csak akkor szolgáltat adatokat, ha nagy tömegek keresnek rá ugyanarra.

De a Google-keresések hatalma nem abban rejlik, hogy megtudhatjuk: Isten lent, a „mély Délen” népszerű, a New York Knicks kosárlabdacsapata New Yorkban népszerű, én pedig nem vagyok népszerű sehol. Ezt bármilyen felmérésből megtudhatnánk. A Google-adatok igazi erejét az adja, hogy az emberek olyasmiket közölnek a gigantikus keresőmotorral, amit senki másnak nem árulnának el.

Vegyük például a szexet (egy olyan témát, amelyet ebben a könyvben később sokkal részletesebben fogok vizsgálni). Ha a szexuális életünkkel kapcsolatban az igazságra vagyunk kíváncsiak, a felmérésekben nemigen bízhatunk. Elemeztem a chicagói National Opinion Research Center (Országos Közvélemény-kutató Központ) által végzett általános társadalmi felmérés, a General Social Survey adatait, amelyet az amerikaiak viselkedésére vonatkozó egyik legbefolyásosabb, legtekintélyesebb információforrásnak tartanak. A heteroszexuális szexre vonatkozó felmérés szerint a nők azt állítják, hogy évente 55 alkalommal szexelnek, és az esetek 16 százalékában használnak óvszert. Ez összesen évi 1,1 milliárd óvszert jelent. Csakhogy a heteroszexuális férfiak azt állítják, hogy ők 1,6 milliárd óvszert használnak el évente. Márpedig a két számnak magától értetődően egyeznie kellene. Akkor hát ki mond igazat, a férfiak vagy a nők?

Mint kiderült, senki. A Nielsen nevű globális információszolgáltató cég szerint, amely a fogyasztói viselkedést méri, az évente eladott óvszerek száma nem éri el a 600 milliót. Tehát mindenki hazudik: az egyetlen különbség, hogy ki mennyire.

Ami azt illeti, a hazugság széles körben elterjedt szokás. Azok a férfiak, akik soha nem voltak házások, állításuk szerint átlagosan 29 óvszert használnak el évente. Ha ez igaz lenne, az több óvszert jelentene, mint ahányat az Egyesült Államokban a házásoknak és a szingliknek együttvéve eladnak. Valószínűleg a házások is túloznak a szex mennyiségét illetően. A 65 év alatti nők férfiak például azt mondják a közvélemény-kutatóknak, hogy átlagosan hetente egyszer szexelnek. Mindössze egy százalékuk állítja, hogy már egy éve szexmentes életet él. A férjes asszonyok valamivel kevesebb együttlétről számolnak be, de nem sokkal kevesebbről.

A Google-keresések viszont jóval kevésbé élénk – és úgy vélem, jóval pontosabb – képet adnak a házasság alatti nemi életéről. A Google-on az a leggyakoribb panasz a házassággal kapcsolatban, hogy nincs szex. A „szexmentes házasság” keresőkifejezés három és félszer gyakrabban fordul elő, mint a „boldogtalan házasság”, és nyolcszor gyakrabban, mint a „szerelem nélküli házasság”. Még az össze nem házasodott párok is elég sűrűn panaszkodnak a szex hiányára. A Google-keresésekben a „szexmentes kapcsolat” kifejezés gyakoriságát csak a „bántalmazó kapcsolat”-ra való keresések gyakorisága múlja felül. (Hangsúlyoznom kell, hogy ezek az adatok mindig névtelenek. A Google természetesen nem közöl információt egyetlen konkrét személy kereséseiről sem.)

És a Google-keresések olyan képet festettek Amerikáról, amely meghökkentően elütött a közvélemény-kutatások által felvázolt posztrasszista utópiától. Máig emlékszem a



pillanatra, amikor először írtam be a Google Trendsbe azt, hogy „nigger”. Mondják bátran, hogy naiv vagyok – de abból kiindulva, mennyire mérgező ez a szó, alacsony keresési mennyiségre számítottam. Hű, hát tévedtem! A „nigger” – vagy ennek többes száma, a „niggerek” – nagyjából ugyanannyi keresésben jelent meg az Egyesült Államokban, mint a „migrén”, a „közgazdász” és a „Lakers” szavak. Eszembe jutott, hogy talán a rapszövegek torzították el az eredményt. Hát nem. A rapperek szinte mindig a „nigga” („niggák”) szót használják. Akkor hát mi ösztönözte az amerikaiakat arra, hogy rákeressenek a „nigger”-re? Sokszor az, hogy az afroamerikaiakon gúnyolódó vicceket kerestek. Ami azt illeti, a „nigger” szót tartalmazó keresések csaknem 20 százaléka tartalmazta a „viccek” szót is. Gyakori keresőkifejezés volt még többek között a „buta niggerek”, valamint a „gyűlölöm a niggereket”.

Évente több millió ilyen keresés volt. Otthon, a négy fal között rengeteg amerikai tudakozódik sokkolóan rasszista tartalmak iránt. És minél tovább kutakodtam, annál nyugtalanítóbb információkra bukkantam.

Obama első megválasztásának estéjén, amikor a kommentárok zöme az új elnök méltatásáról szólt, illetve arról, hogy a győzelme történelmi jelentőségű, durván minden századik Google-keresésben, amely tartalmazta az „Obama” szót, szerepelt a „KKK” vagy a „nigger(ek)” is. Ez talán nem tűnik túl nagy aránynak, de gondoljanak csak bele, hány ezer nem rasszista oka lehetett annak, hogy az ember rákeressen erre a fiatal outsiderre, akinek ilyen bájos családja van, és éppen átveszi a világ legnagyobb hatalommal járó munkakörét! A választás estéjén a Stormfront nevű fehér nacionalista weboldalra, amely meglepően népszerű az Egyesült Államokban, tízszer annyian kerestek rá, illetve iratkoztak fel, mint egy átlagos napon. Egyes államokban

többször kerestek rá arra, hogy „nigger elnök”, mint arra, hogy „első fekete elnök”.

Volt ott valami sötétség és gyűlölet, amely a hagyományos források előtt rejtve maradt, de az emberek Google-kereséseiben elég látványosan megjelent.

Az ilyen keresések nehezen békíthetők össze egy olyan társadalom képével, ahol a rasszizmus jelentéktelen tényező. 2012-ben Donald J. Trump elsősorban üzletemberként és valóságshow-házigazdaként élt a fejében. Ugyanúgy nem sejtettem, ahogy senki más, hogy négy évvel később komoly elnökjelölt lesz belőle. Ám a fenti, rusnya Google-kereséseket már igazán nem nehéz összebékíteni egy olyan jelölt sikerével, aki – a bevándorlók elleni támadásaival, indulataival és gyűlölködésével – gyakran játszott rá az emberek legrosszabb hajlamaira.

A Google-keresések azt is elárulták, hogy amit a rasszizmus területi eloszlásáról gondoltunk, többnyire téves volt. A felmérések, illetve a közmegegyezés főként Délen és főként a republikánusok körében helyezte el a modern rasszizmust. Csakhogy az ország azon területei között, ahol a legmagasabb volt a rasszista keresések aránya, Nyugat-Virginia, Dél-Louisiana és Mississippi állam mellett megjelent New York állam északi része, Nyugat-Pennsylvania, Kelet-Ohio, a michigani iparvidék, valamint Illinois vidékies része is. A Google-keresések adatai szerint a valódi választóvonal nem Észak és Dél, hanem Kelet és Nyugat közt húzódott. A Mississippitől nyugatra nemigen találkozik ilyesmivel az ember. És a rasszizmus nem korlátozódott a republikánusokra. Sőt, ami azt illeti, a rasszista tartalmú keresések aránya nem volt magasabb azokon a helyeken, ahol magas volt a republikánusok aránya, mint ott, ahol a demokraták voltak túlsúlyban. Más szóval: a Google-ke-

resések segítettek átrajzolni az Egyesült Államok rasszizmustérképét, és ez a térkép egészen másképp nézett ki, mint amilyennek képzeltük. Lehet, hogy a déli republikánusok nagyobb valószínűséggel vállalják fel a rasszizmusukat. De rengeteg északi demokrata is hasonlóan viszonyul a kérdéshez.

Ez a térkép négy évvel később jelentősen hozzájárult Trump politikai sikerének magyarázatához.

2012-ben arra használtam a Google-keresések alapján kifejlesztett rasszizmustérképet, hogy átértékeljem Obama faji hovatartozásának szerepét. Az adatok egyértelműek voltak: azokban az országrészekben, ahol magas volt a rasszista tartalmú keresések száma, Obama lényegesen rosszabbul szerepelt, mint négy évvel korábban John Kerry, a demokraták fehér elnökjelöltje. A különbségre semmilyen más tényező nem adott magyarázatot: sem az iskolázottság szintje, sem a korosztályi összetétel, sem a templomba járás, sem a fegyverbirtoklás. A rasszista keresések száma nem jelezte előre egyetlen más demokrata jelölt gyenge szereplését sem. Egyedül Obamáét.

És az eredmények óriási hatásra utaltak. Obama országosan durván 4 százalékpontot veszített a nyílt rasszizmus miatt. Ez jóval nagyobb arány, mint amit bármilyen felmérés alapján várhattunk volna. Barack Obamát persze így is elnökké választották, majd újraválasztották, amit néhány, a demokraták számára roppant kedvező körülmény is elősegített. De jócskán több hátrányt kellett leküzdenie, mint azt a hagyományos adatforrásokra támaszkodók – azaz nagyjából mindenki – gondolták volna. Épp elég rasszista volt ahhoz, hogy eldöntsenek egy előválasztást, vagy átbilentsék egy elnökválasztás eredményét egy olyan évben, amikor a körülmények nem kedveztek ennyire a demokratáknak.

A tanulmányomat elsőre öt tudományos folyóirat is elutasította. Ha elnéznek nekem egy kis morgolódást: cikkeim több szakmai bírálója azt mondta, egyszerűen nem hihető, hogy ilyen sok amerikai rejteget ennyire durva rasszizmust a lelkében. Ez egészen egyszerűen nem volt összeegyeztethető mindazzal, amit akkoriban az emberek előszeretettel mondogattak. Ráadásul a Google-keresések igencsak bizarr adathalmaznak tűntek...

Ma azonban, miután tanúi voltunk Donald J. Trump elnöki beiktatásának, az eredményeim már jóval hihetőbbnek tűnnek.

Minél tovább tanulmányoztam a Google-t, annál világosabbá vált, hogy rengeteg olyan információt tartalmaz, amelyekre a közvélemény-kutatások nem derítenek fényt, és amelyek segítségével – sok-sok más téma mellett – egy választást is könnyebben megérthetünk.

Információkat nyerhetünk például arról, hogy ténylegesen ki fog elmenni szavazni. A választás előtti napokban végzett felméréseken a nem szavazó állampolgárok több mint fele azt válaszolja, hogy szándékában áll voksolni, ami torzítja a részvételi arányra vonatkozó becsléseinket. Az viszont, hogy hetekkel a választások előtt hányan keresnek rá arra, hogy „hogyan kell szavazni”, illetve „hol kell szavazni”, képes pontosan előre jelezni, hogy az országnak mely részeiben várható magas részvételi arány.

Sőt, talán még arról is lehetnek ott információk, hogy ki-re fognak szavazni az emberek. De hát hogy jósolhatnánk meg pusztán a Google-kereséseik alapján, hogy melyik jelöltre adják majd voksukat? Nyilván nem elég, ha azt vizsgáljuk, hogy melyik jelöltre keresnek rá a leggyakrabban. Sokan azért keresnek rá valamelyik jelöltre, mert szeretik. Nagyjából ugyanennyien azért keresnek rá valamelyik je-

lőltre, mert utálják. Mindazonáltal Stuart Gabriellel, a Los Angeles-i Kaliforniai Egyetem pénzügyprofesszorával találtunk egy meglepő nyomot arra vonatkozóan, hogyan terveznek szavazni az emberek. A választással kapcsolatos keresések igen nagy százalékában mindkét jelölt neve szerepel. 2016-ban, a Donald Trump és Hillary Clinton közt zajló választási versengés idején voltak, akik arra kerestek rá, hogy „Trump Clinton közvélemény-kutatás”. Másokat a „Clinton–Trump-vita” fontosabb eseményei érdekelték. Összességében a „Trump” szót tartalmazó keresések 12 százaléka tartalmazta a „Clinton” szót is. A „Clinton” szót tartalmazó kereséseknek több mint negyedében szerepelt a „Trump” szó.

Rájöttünk, hogy ezek a látszólag semleges keresések valójában adhatnak némi támpontot arra vonatkozóan, melyik jelöltet támogatja valaki.

Hogy hogyan? Hát a jelöltek nevének sorrendjével. A kutatásunk szerint az ember nagy valószínűséggel az általa támogatott jelölt nevét írja be először egy olyan keresés során, amely mindkét jelölt nevét tartalmazza.

Az előző három elnökválasztáson az a jelölt kapott több szavazatot, akinek neve több Google-keresésben szerepelt elsőként. És ami még érdekesebb: a jelöltek nevének sorrendje azt is előre jelezte, hogy ki szerez többséget egy-egy államban.

A jelöltek nevének sorrendje a jelek szerint más olyan információt is tartalmaz, amely a közvélemény-kutatók előtt rejtve maradhat. A 2012-es választáson, ahol Obama vetélytársa a republikánus Mitt Romney volt, Nate Silver, a zseniális statisztikus és újságíró mind az ötven államban eltalálta a végeredményt. Mi azonban azt is kiderítettük Stuart Gabriellel, hogy azokban az államokban, ahol többször írták be elsőként Romney nevét, mint Obamáét, Rom-

ney jobban teljesített, mint ahogy azt Nate Silver megjósolta. Azokban az államokban pedig, ahol Obama neve került többször az első helyre, Obama ért el jobb eredményt, mint amit Silver várt.

Ez a mutató azért tartalmazhat olyan információt, amelyre a felmérések során nem derül fény, mert a válaszók vagy önmaguknak is hazudnak, vagy nem szívesen fedik fel a közvélemény-kutatók előtt a valódi preferenciáikat. Talán azok, akik a 2012-es felmérésekben bizonytalanul mondták magukat, de következetesen olyanokat írtak be a keresőbe, hogy „Romney–Obama közvélemény-kutatás”, „Romney–Obama vita” és „Romney–Obama-választás”, mindvégig úgy tervezték, hogy Romney-ra adják majd a voksukat.

Szóval a Google megjósolta Trump győzelmét? Nos, még igen sok a tennivalónk – és rengeteg kutatónak kell még csatlakoznia a munkához – ahhoz, hogy kiderüljön, miként használhatjuk fel a Google-adatokat a választási eredmények előrejelzésére. Ez egy teljesen új tudományág, és csak a legutóbbi néhány elnökválasztásra vonatkozóan rendelkezünk ilyen adatokkal. Semmiképpen nem állítom, hogy már ott tartunk – vagy valaha is ott fogunk tartani –, hogy végleg kidobhatjuk az ablakon a közvélemény-kutatást mint a választási eredmények előrejelzésében hasznos eszközt.

Ugyanakkor az internet számos pontján határozottan voltak előjelek, amelyek arra utaltak, hogy Trump jobban szerepelhet, mint azt a közvélemény-kutatások jósolják.

Az elnökválasztás voltak arra utaló jelek, hogy a választók összetétele kedvezhet Trumpnak. A fekete amerikaiak azt mondták a közvélemény-kutatóknak, hogy hatalmas számban fognak Trump ellen szavazni. Ám a túlnyomórészt feketék lakta területeken igen alacsony volt az olyan

Google-keresések száma, amelyek a szavazással kapcsolatos információkra irányultak. A választás napján Hillary Clinton számára súlyos csapást jelentett a feketék alacsony részvétele.

Még arra utaló jelek is voltak, hogy a bizonytalan szavazók Trump felé mozdulnak. Stuart Gabriellel felfedeztük, hogy a kulcsfontosságú középnyugati államokban, ahol Clinton győzelemre számított, többen kerestek arra, hogy „Trump–Clinton”, mint arra, hogy „Clinton–Trump”. És Trump éppen annak köszönhetette megválasztását, hogy ezekben az államokban jócskán túlteljesítette a közvélemény-kutatási eredményeit.

A sikere legfőbb előjele azonban – már az előválasztás idején – szerintem az a gondosan titkolt rasszizmus volt, amelyre az Obama-tanulmányom fényt derített. A Google-keresések olyan, az amerikaiak jelentős hányadában jelen lévő sötétségről és gyűlöletről tanúskodtak, amely felett a véleményformálók hosszú éveken keresztül elsiklottak. A keresési adatok elárulták, hogy a társadalom, amelyben élünk, erősen különbözik attól, amit a tudósok és az újságírók a közvélemény-kutatási eredmények alapján gondolnak róla. Rusnya, ijesztő, széles körben elharapózott indulatokról árulkodtak az adatok, és ez a felgyülemlett indulat csak egy jelöltre várt, aki majd hangot ad neki.

Az emberek gyakran hazudnak – önmaguknak és másoknak is. Az amerikaiak 2008-ban azt állították a felméréseken, hogy már nem érdekli őket a faji hovatartozás. Nyolc évvel később pedig elnökké választották Donald J. Trumpot, egy olyan embert, aki azt a hamis állítást osztotta meg a Twitteren, hogy a feketék felelősek a fehér amerikaiak ellen elkövetett gyilkosságok többségéért, majd egyik nagygyűlésén védelmébe vette azon támogatóit, akik ellátták a baját a Black Lives Matter Mozgalom egyik

tiltakozó aktivistájának, aztán pedig csak hosszas habozás után vonta meg támogatását a Ku-Klux-Klan egyik egykori vezetőjétől. Ugyanaz a rejtett rasszizmus, amely Barack Obamának hátrányt jelentett, Donald Trumpot segítette.

Még az előválasztási időszak kezdetén hangzott el Nate Silver elhíresült kijelentése, mely szerint Trump gyakorlatilag esélytelen. Aztán egyre nyilvánvalóbbá vált, hogy Trump széles körű támogatottságot élvez, és Silver úgy döntött, megnézi az adatokat. Meg akarta érteni, mi folyik itt. Hogy szerepelhet Trump ilyen jól?

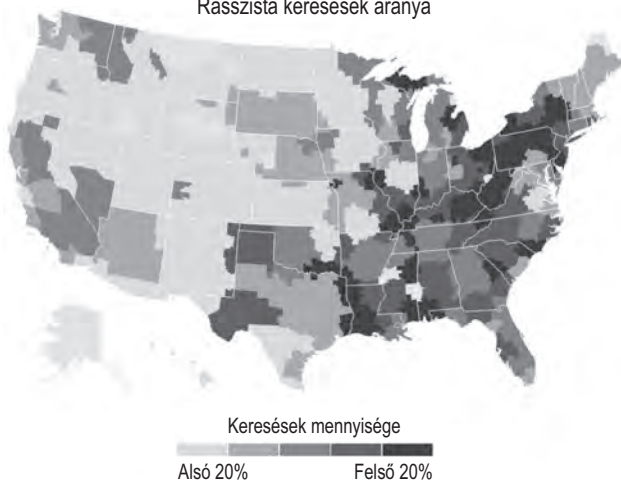
Silver észrevette, hogy azok a területek, ahol Trump a legjobb eredményeket érte el, furcsa térképet rajzolnak ki. Trump jól szerepelt az északkeleti országrészben, a Közép-Nyugat iparvidékein, valamint Délen is. Az ország nyugati részén viszont lényegesen rosszabb eredményeket ért el. Silver olyan változókat keresett, amelyek magyarázatot adhatnak erre a térképre. Mi lehetett Trump sikerének titka? A munkanélküliség? A vallásosság? A fegyverbirtoklás? A bevándorlás mértéke? Az Obama-ellenesség?

Silver arra jutott, hogy az az egy tényező, amely a leginkább korrelált Donald Trump támogatottságával a republikánus előválasztáson, az volt, amelyre én bukkantam rá négy évvel korábban. Azokon a területeken támogatták a legtöbben Trumpot, ahol a legtöbbször kerestek rá a „nigger” szóra.

Az utóbbi négy év szinte minden napját Google-adatok elemzésével töltöttem. Egy ideig adattudósként is dolgoztam a Google-nál, ahová azért vettek fel, mert tudomást szereztek a rasszizmussal kapcsolatos kutatásomról. Ezeket az adatokat azóta is folyamatosan vizsgálom a *The New York Times* publicistájaként és adatújságírójaként. És egymást követik a megvilágosodások. Mentális betegségek, emberi szexualitás, gyerekbántalmazás, abortusz,



### Rasszista keresések aránya



### Donald Trump támogatottsága a republikánus előválasztáson



vallás, egészségügy... Nem éppen kis horderejű témák, de a Google-adathalmaz, amely pár évtizede még sehol sem volt, meglepő, új távlatokat nyitott a feltérképezésükben. A közgazdászok és más társadalomtudósok egyfolytában új adatforrásokra vadásznak, úgyhogy engedjék meg, hogy ne kerteljek: meggyőződésem, hogy a Google-keresések képezik az emberi pszichéről valaha összegyűjtött legfontosabb adathalmazt.

Ráadásul ez az adathalmaz nem is az egyetlen eszköz, amit az internetről kaptunk, hogy megérthessük a világot. Hamar rájöttem, hogy vannak más digitális aranybányák is. Letöltöttem az egész Wikipédiát, Facebook-profilokat böngésztem, és a Stormfrontot is megkapirgáltam. Mindezekon felül az egyik legnagyobb pornográf weboldal, a Pornhub is átadta nekem a világ legkülönbözőbb pontjain élő, anonim emberek kereséseire, videómegtekintéseire vonatkozó teljes adatállományát. Fejest ugorhattam abba, amit manapság Big Datának, azaz „nagy adatnak” hívunk. Többtucatnyi emberrel – tudósokkal, szakújságírókkal, vállalkozókkal – készítettem továbbá interjút, akik szintén ennek az új keletű birodalomnak a feltárásán dolgoznak. Számos tanulmányukat fogom tárgyalni ebben a könyvben.

De mindenekelőtt egy vallomás: nem fogom precízen definiálni, hogy mi az a Big Data. Hogy miért? Azért, mert ez eredendően homályos fogalom. Hogyan döntsük el, milyen nagy az, ami nagy? Hol húzzuk meg a határt? Talán ott, hogy 18 462 megfigyelés még Small Data, azaz kis adat, 18 463 pedig már Big Data? A magam részéről inkább minél több mindent lefedő módon tekintek arra, hogy mi minősül ilyennek: az adatok többsége, amelyekkel babrálok, az internetről származik, de más forrásokról is szót ejtek. Napjainkban az elérhető információk mennyiségi és minőségi robbanásának korát éljük, és az új adatok zöme

a Google, illetve a közösségi média révén áramlik be. Az adathalmaz egy része a korábban iratszekrényekben, aktákban rejtőző információk digitalizálásának eredménye. Más része piackutatásokhoz felhasznált, egyre bővülő forrásokból származik. A könyvben tárgyalt tanulmányok némelyike pedig egyáltalán nem dolgozik óriás adathalmazokkal, viszont újszerű, kreatív módon közelíti meg az adatokat – márpedig a mi információktól túltengő körünkben létfontosságú az efféle kreativitás.

Pontosan miért is bír tehát ekkora hatalommal a Big Data? Gondoljunk csak bele, mennyi információ szóródik szét online egy adott napon – ami azt illeti, van is pontos számunk arra vonatkozóan, hogy mennyi. Az emberek a 21. század elején, egy átlagos napon 2,5 millió trillió bájtnyi adatot generálnak.

És ezek a bájtok nyomok.

*Egy nő unatkozik csütörtök délután. Rákeres még néhány „nem disznó vicc”-re. Megnézi az e-mailjét. Bejelentkezik a Twitterre. Rákeres a „nigger viccek”-re.*

*Egy férfi lehangolt. Rákeres arra, hogy „depresszió tünetei”, és arra, hogy „depressziós történetek”. Pasziánszozik egyet.*

*Egy nő meglátja, hogy egy barátnője bejelentette az eljegyzését a Facebookon. A nő, aki szingli, letiltja a barátnőjét.*

*Egy férfi kis időre félbehagyja az NFL-lel meg a rappel kapcsolatos böngészést, és feltesz a keresőmotornak egy kérdést: „Normális, ha valaki álmában férfiakkal csókolózik?”*

*Egy nő rákattint egy BuzzFeed-sztorira, amely bemutatja „a 15 legcukibb cicát”.*

*Egy férfi is látja ugyanezt a cicás sztorit, de az ő képernyőjén az a felirat, hogy „a 15 legimádnivalóbb cica”. Ő nem kattint rá.*

*Egy nő rákeres arra, hogy „zseni a fiam?”.*

*Egy férfi rákeres arra, hogy „hogyan vehetném rá a lányomat, hogy fogyjon le?”.*

*Egy nő a hat legjobb barátnőjével üdül. A barátnői egyfolytában azt mondogatják, mennyire jól érzik magukat. Ő pedig titokban rákeres arra, hogy „magány a férjtől távol”.*

*Egy férfi, az utóbbi nő férje, a hat legjobb barátjával vakációzik. Ő ezt gépeli be titokban a Google-ba: „arra utaló jelek, hogy a feleséged megcsal”.*

A fenti adatok némelyike olyan információkat tartalmaz, amit az emberek más körülmények között senkinek sem vallanának be. Ha mindezt összesítjük – természetesen nevek nélkül, hogy soha ne szerezzünk tudomást egyetlen konkrét személy félelmeiről, vágyairól, viselkedéséről sem –, és beleviszünk a dologba némi adattudományt, máris kezdjük más szemmel látni az embereket: a viselkedésüket, a vágyaikat, a természetüket. Ami azt illeti, megkockáztatva akár a fellengzősség vádját is, arra a meggyőződésre jutottam, hogy a digitalizált korunkban egyre könnyebben hozzáférhető, újfajta adatok radikálisan ki fogják bővíteni az emberiségre vonatkozó tudásunkat. A mikroszkóp megmutatta nekünk, hogy a tó vizének egyetlen cseppjében jóval több van, mint amennyit szabad szemmel látni vélünk benne. A teleszkóp megmutatta, hogy az éjjeli égbolton több a néznivaló, mint amennyit szabad szemmel látni vélünk. Az újszerű, digitális adatok pedig most azt mutatják meg, hogy az emberi társadalomban is több van, mint amennyit látni vélünk benne. A Big Data lehet korunk mikroszkópja vagy teleszkópja, amely fontos, sőt forradalmi jelentőségű felismeréseket tesz lehetővé.

Egy ilyen állításban egy további kockázat is rejlik: az, hogy nem pusztán fellengzősen hangzik, hanem még trendi is. Számos nagyotmondó kinyilatkoztatás hangzott már

el a Big Data hatalmára vonatkozóan. De bizonyítékokkal egyik sem nagyon szolgált.

Ez arra inspirálta a „Big Data-szkeptikusokat”, akikből szintén jó néhány akad, hogy elutasítsák a nagyobb volumenű adathalmazok kutatását. Nassim Taleb esszéista és statisztikus például így vélekedik: „Nem mondom, hogy a Big Data nem tartalmaz információt. Van ott információ bőven. Csak az a gond – és ez a középponti kérdés –, hogy egyre nagyobb szénakazalban kell keresnünk a tűt.”

Ennek a könyvnek az az egyik fő célja, hogy felmutassa a hiányzó bizonyítékokat arra vonatkozóan, mit lehet kezdeni a Big Datával – vagyis hogyan találhatjuk meg a tűt az egyre nagyobb szénakazalban. Remélem, sikerül elegendő példával szolgálnom arra, hogy a Big Data újszerű felismerésekkel szolgál az emberi pszichológia és viselkedés terén, hogy végül önök előtt elkezdjenek kirajzolódni valami egészen újszerű, forradalmi dolog körvonalai.

„Várj csak, Seth!” – mondhatják nagyjából most. „Forradalmat ígérsz, és költőien áradozol ezekről a hatalmas, újszerű adathalmazokról. De mindeddig alapvetően két dolgot tudtam meg tőled ezeknek a rendkívüli, elképesztő, lélegzetelállító, úttörő jelentőségű adataidnak a révén: azt, hogy Amerikában sok a rasszista, meg azt, hogy az emberek, különösen a férfiak, tódítanak arról, hogy mennyit szexelnek.”

Beismerem, hogy ezek az újfajta adatok néha csak megerősítik a nyilvánvalót. De aki úgy gondolja, hogy eddig triviális dolgokról esett szó, várjon csak, amíg elér a 4. fejezetig, ahol a Google-keresések alapján egyértelműen, kétségbevonhatatlanul bebizonyítom, hogy a férfiaknál rengeteg aggodalom és bizonytalanság okozója a – dobpergés! – péniszméretük!

Azt állítom, hogy értékes dolog olyasmit bizonyítani, amit eleve gyanítottunk, de nemigen volt rá bizonyítékunk. A gyanú egy dolog. Igazolni már egészen más. De ha a Big Data csak arra lenne jó, hogy beigazolja a sejtéseinket, nem lenne forradalmi jelentőségű. A Big Data azonban, hála az égnek, ennél sokkal többre képes. Az adatok egyre-másra arról árulkodnak nekem, hogy a világ pont fordítva működik, mint gondoltam. Íme, néhány példa, amelyekkel talán sikerül meglepnem önöket.

Talán azt gondolják, hogy a rasszizmus fő oka a gazdasági bizonytalanság és kiszolgáltatottság. Ezért természetes módon arra gyanakszunk, hogy ha az emberek elveszítik a munkahelyüket, a rasszizmus erősödni fog. Pedig valójában sem a rasszista keresések száma, sem a Stormfront taglétszáma nem lesz magasabb, amikor nő a munkanélküliség.

Talán azt gondolják, hogy a szorongás mértéke a túlképzett nagyvárosokban a legmagasabb. A nagyvárosi neurózis közismert sztereotípiája. A szorongással kapcsolatos Google-keresések – úgymint „szorongás tünetei”, illetve „szorongás segítség” – rendszerint mégis gyakoribbak az alacsonyabb iskolázottsági szintű és átlagjövedelmű helyeken, illetve ott, ahol a népesség nagyobb része él vidékies területeken. A szorongás keresési rátája magasabb New York állam vidékies, északi részein, mint New York Cityben.

Talán azt gondolják, hogy a többtucatnyi vagy akár több száz ember életét követelő terrortámadásokat automatikusan nagyarányú, széles körben elharapózó szorongáshullám követi. A terrorizmusnak definíció szerint az a lényege, hogy rettegést váltson ki. Megnéztem a szorongást tükröző Google-kereséseket. Azt vizsgáltam, milyen mértékben nőtt a számuk a 2004 óta végrehajtott nagy európai, illetve amerikai terrorakciók utáni napokban, hetekben, hónapokban. Szóval átlagosan mennyivel is nőtt ezekben

az időszakokban a szorongással kapcsolatos keresések száma? Semennyivel. Egyáltalán semennyivel.

Talán azt gondolják, hogy az emberek olyankor keresnek gyakrabban vicceket, amikor szomorúak. A történelem legnagyobb gondolkodói közül sokan állították, hogy a fájdalomunk enyhítése céljából fordulunk a humor felé. Régóta úgy gondoljuk, hogy a humor a frusztrációval, a fájdalommal, az élet óhatatlan csalódásaival való megbirkózás eszköze. Charlie Chaplin szavaival élve: „A nevetés felüdít, könnyebbséget hoz, és enyhíti a fájdalmat.”

Csakhogy a viccek keresési rátája valójában hétfőnként a legalacsonyabb, pedig az emberek saját bevallásuk szerint olyankor a legboldogtalanabbak. A viccek keresési rátája felhős, esős napokon a legalacsonyabb. Egy-egy komolyabb tragédiát követően pedig a mélybe zuhan, mint például a 2013-as bostoni maraton után, ahol két bombabarobbanás során meghalt három ember, és több százan megsebesültek. Az emberek nagyobb valószínűséggel keresnek vicceket a neten, ha szép az élet, és jól mennek a dolgok, mint ha nem.

Egy új adathalmaz néha olyan viselkedésmódokat, vágyakat, aggályokat tár fel, amelyeket eszembe sem jutott volna vizsgálni. Akad jó néhány szexuális hajlam, amely ebbe a kategóriába tartozik. Tudták például, hogy Indiában a leggyakoribb, „a férjem azt akarja, hogy...” kezdetű keresés ez: „a férjem azt akarja, hogy szoptassam”? Ez a komment sokkal gyakoribb Indiában, mint más országokban. Sőt mit több: a férfit szoptató nőt ábrázoló pornóra való keresés négyszer olyan gyakori Indiában és Bangladesben, mint a világ bármely más országában. Ezt egészen biztosan nem gyanítottam, amíg meg nem láttam az adatokat.

Az, hogy a férfiakat rögeszmésen foglalkoztatja a péniszük mérete, nem hangzik túl meglepően, de a nők legna-

gyobb, a testükkel összefüggő és a Google-nak megvallott bizonytalansági tényezője már annál inkább. Ezen új adatok szerint a péniszméret miatti aggodás női megfelelője – rövid hatásszünet – az azon való görcsölés, hogy nem büdös-e a hüvelyük. A nőknél majdnem ugyanannyi a nemi szervük miatti aggályokat tükröző keresés, mint a férfiaknál, csak nekik az a legfőbb gondjuk, hogy milyen a szóban forgó nemi szerv szaga, és miként javíthatnának rajta. Ezt egészen biztosan nem tudtam, amíg meg nem láttam az adatokat.

Az új adatok időnként olyan kulturális különbségekről árulkodnak, amelyek eddig meg sem fordultak a fejemben. Egy példa: az emberek a világ különböző tájain nagyon másként reagálnak arra, hogy a feleségük állapotos. Mexikóban a terhes feleséggel kapcsolatos leggyakoribb keresések között szerepel az, hogy „*frases de amor para mi esposa embarazada*” (szerelmes szavak várandós feleségemnek), valamint az, hogy „*poemas para mi esposa embarazada*” (versek várandós feleségemnek). Az Egyesült Államokban viszont olyanokat találunk a leggyakoribb keresések között, mint hogy „terhes a feleségem, most mi van”, és hogy „terhes a feleségem, mit csináljak”.

A könyvem azonban több mint furcsaságok és egyedi esettanulmányok gyűjteménye, bár lesznek még ilyenek szép számmal. Mivel az alkalmazott módszerek nagyon újak és egyre erőteljesebbek, kifejték néhány gondolatot a működésükről, illetve arról, hogy mitől úttörő jelentőségűek. De a Big Data korlátait is el fogom ismerni.

Az adatforradalomban rejlő lehetőségek iránti lelkesedés időnként félrevisz. A Big Data szerelmeseinek többsége arról ömleng, milyen iradatlan nagyok tudnak lenni ezek az új adathalmazok. Ez az adathalmaz méretével kapcsolatos rögeszme nem újdonság. Dallasban már akkor rendeztek



egy *Nagy és komplex adathalmazok* című konferenciát, amikor a Google, az Amazon és a Facebook még nem volt sehol, sőt még a „Big Data” kifejezés sem létezett. Jerry Friedman, a Stanford statisztikaprofesszora, aki a kollégám volt a Google-nál, felidézte ezt az 1977-es konferenciát. Egy jeles statisztikus emelkedett szólásra. Elmondta, hogy döbbenetes, elképesztő adattömeget halmozott fel: öt gigabájtot! Aztán a következő jeles statisztikus emelkedett szólásra. Így kezdte: „Elöttem szóló kollégámnak gigabájtjai vannak. Az semmi. Nekem terabájtjaim vannak.” Más szóval azon volt a hangsúly, hogy mennyi információt tudunk felhalmozni, nem pedig azon, hogy reményeink szerint mit tudunk majd kezdeni vele, illetve milyen kérdéseket tervezünk megválaszolni a segítségükkel. „Akkor mulatságosnak találtam” – mondja Friedman –, „hogyan az adatbázisuk méretével akartak lenyűgözni bennünket. Ilyesmi még mindig előfordul.”

Manapság valóban igen sok az olyan adattudós, aki iradatlan méretű adathalmazokat dolgoz fel, amelyekből rendkívül érdektelen dolgokat hoz ki – például hogy a New York Knicks népszerű New Yorkban. Túl sok cég fulladozik az adatokban. Jó sok terabájtjuk van, és annál kevesebb érdemi megfigyelésük. Úgy vélem, az adatbázis méretét sokszor túlértékelik. Van erre egy összetett, de igen fontos magyarázat: minél nagyobb valaminek a hatása, annál kisebb számú megfigyelés kell ahhoz, hogy megértsük. A forró kályhához például elég egyszer hozzányúlunk, rögtön tudni fogjuk, hogy veszélyes. Kávézni viszont több ezerszer kell ahhoz, hogy megállapítsuk, hajlamos-e megfájdulni tőle a fejünk, vagy sem. Vajon melyik lecke a fontosabb? Nyilván a forró kályha leckéje, amely a hatás intenzitása miatt villámgyorsan, minimális adatmennyiség mellett is megmutatkozik.

Az igazság az, hogy a legokosabb Big Data-cégek sokszor tudatosan csökkentik az adatmennyiséget. A Google-nál a nagy horderejű döntéseket az összes adatból vett parányi minta alapján hozzák meg. Ahhoz, hogy az ember fontos dolgokra jöjjön rá, nincs mindig szükség egy tonna adatra. A megfelelő adatokra van szükség. A Google-keresések hatalmas értékét elsősorban nem az adja, hogy olyan sok van belőlük – hanem az, hogy az emberek annyira őszintén nyilvánulnak meg általuk. Hazudnak a barátaiknak, a szeretőjüknek, az orvosuknak, a közvélemény-kutatóknak és önmaguknak. A Google-lal azonban sokszor kínos információkat is megosztanak, többek között olyan témákban, mint a szexmentes házasságuk, a mentális egészségük, a bizonytalanságaik, vagy éppen a fekete bőrű emberek iránti ellenérzéseik.

És ami a legfontosabb: annak érdekében, hogy jelentős felismeréseket sajtolhassunk ki a Big Datából, a megfelelő kérdéseket kell feltennünk. Ha csak véletlenszerűen pásztázzuk az éjjeli égboltot a teleszkóppal, nem várhatjuk, hogy felfedezze nekünk a Plutót. Ugyanígy nem várhatjuk el egy nagy rakás adattól sem, hogy feltárja előttünk az emberi természet titkait. A kutakodást ígéretes helyeken kell megkezdenünk: például az „a férjem azt akarja, hogy...” kezdetű indiai Google-keresések között.

Ez a könyv azt fogja szemléltetni, hogyan lehet a legjobban használni a Big Datát, és részletesen elmagyarázza, miért is lehet ez hatalmas jelentőségű eszköz. Menet közben az is ki fog derülni, mi mindent felfedeztünk már, én is, mások is, például a következő kérdések kapcsán:

- Hány férfi meleg?
- Működnek-e a hirdetések?
- Miért volt nagy versenyló American Pharoah?
- Elfogult-e a média?

- Létezik-e a freudi elszólás?
- Ki csal adót?
- Számít-e, hogy az ember hova jár egyetemre?
- Kijátszhatjuk a részvényt piacot?
- Hol a legjobb gyereket nevelni?
- Mitől lesz virális egy sztori?
- Miről beszéljünk az első randin, ha azt szeretnénk, hogy második is legyen?

...és még rengeteg, rengeteg további kérdés kapcsán.

De mielőtt mindebbe belevágnánk, meg kell vitatnunk a legalapvetőbb kérdést: azt, hogy miért is van egyáltalán szükségünk adatokra. Ennek érdekében pedig mindenekelőtt bemutatom a nagymamámat.