Data Science: How to Turn Data Into Actionable Mission-Oriented Decisions

Harvey Alférez, Ph.D.
Global Software Lab,
School of Engineering and Technology,
Universidad de Montemorelos, Mexico
www.harveyalferez.com

At the Church, We're Drowning in a Sea of Data

Video & Audio

Textual Documents

Social Media

Web Access & Search Logs

Databases

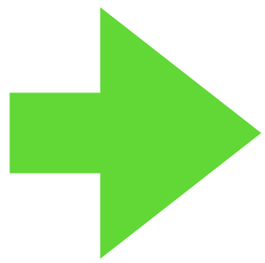
Geolocation

- **How to discover the current and future needs of people at your conference?**
- **How to understand your members with data?**
- **Are SDA churches and schools near to families with children?**
- **Where to open a new school or church?**
- **Where to launch evangelistic campaigns?**
- **Which places don’t have Adventist presence?**
- **How to know the apostasy profile from data?**

Data Science

Big Data

Open Data



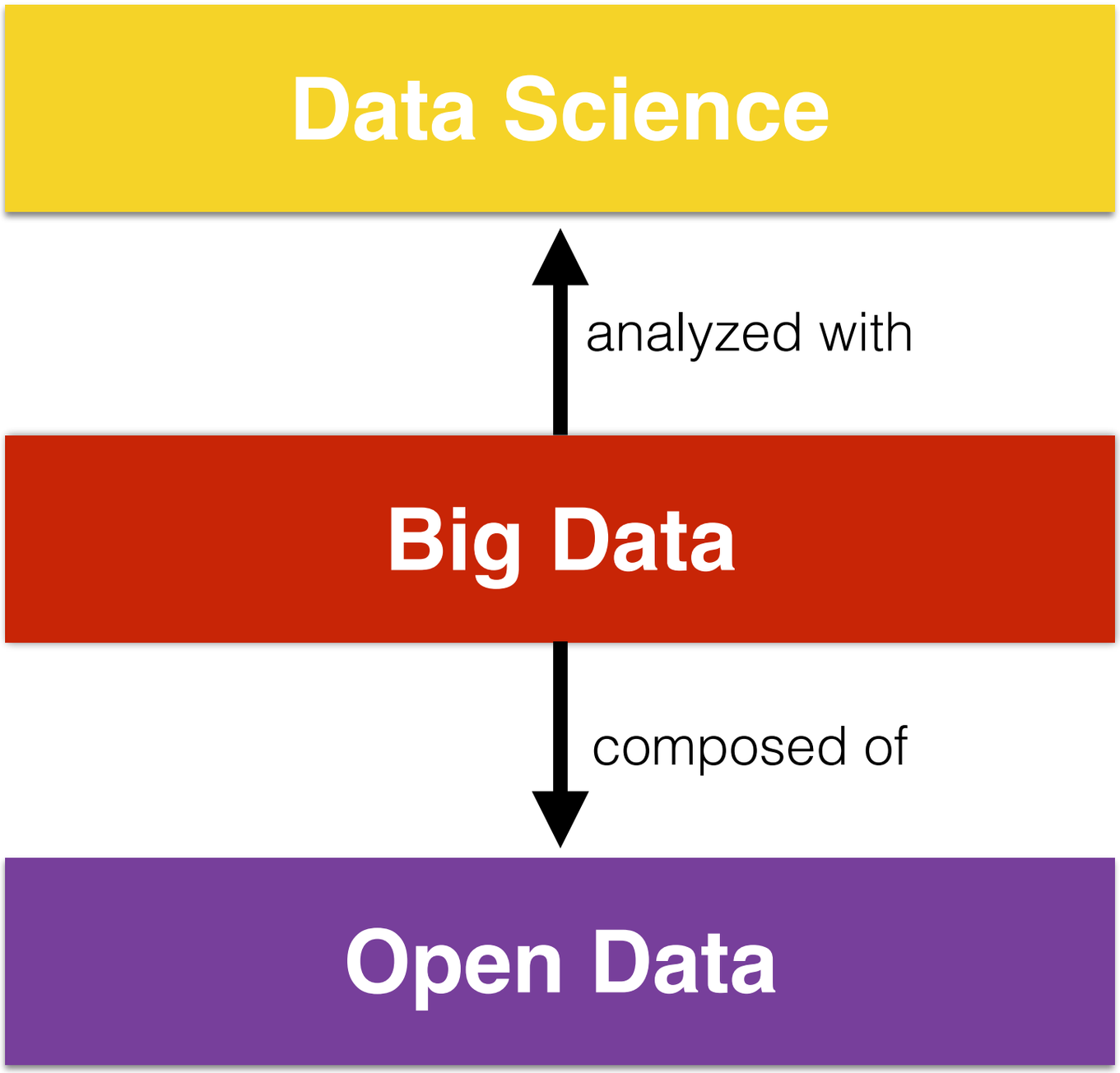
Our Mission

Agenda

1. What is big data, data science, and open data?
2. Successful case studies
3. How to use data science at your conference
4. Conclusions

Agenda

1. **What is big data, data science, and open data?**
2. Successful case studies
3. How to use data science at your conference
4. Conclusions


Data Science
analyzed with
Big Data
composed of
Open Data

1. Snijders, C., Matzat, U., and Reips, U.-D. (2012). “Big data”: Big gaps of knowledge in the field of Internet science. International Journal of Internet Science, 1(1):1–5.

2. Rotella, P. (2012). Is data the new oil? URL: http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/

IN 60 SECONDS...

1 NEW DEFINITION IS ADDED ON urban

1,600+ READS ON Scribd.

13,000+ HOURS MUSIC STREAMING ON PANDORA

12,000+ NEW ADS POSTED ON craigslist

370,000+ MINUTES VOICE CALLS ON skype

98,000+ TWEETS

20,000+ NEW POSTS ON tumblr.

THE LARGEST SOCIAL READING PUBLISHING COMPANY!!

320+ NEW twitter ACCOUNTS

13,000+ iPhone APPLICATIONS DOWNLOADED

100+ NEW Linked in ACCOUNTS

1 associatedcontent NEW ARTICLE IS PUBLISHED

THE WORLD'S LARGEST COMMUNITY CREATED CONTENT!!

QUESTIONS ASKED ON THE INTERNET...

100+ Answers.com

40+ YAHOO! ANSWERS

6,600+ NEW PICTURES ARE UPLOADED ON flickr

600+ NEW VIDEOS

50+ WordPress DOWNLOADS

25+ HOURS TOTAL DURATION

70+ DOMAINS REGISTERED

60+ NEW BLOGS

168 MILLION EMAILS ARE SENT

694,445 SEARCH QUERIES

695,000+ facebook STATUS UPDATES

1,700+ Firefox DOWNLOADS

125+ PLUGIN DOWNLOADS

1,500+ BLOG POSTS

79,364 WALL POSTS

510,040 COMMENTS

Google

Google Search

GO-Globe.com web technologies

THE SOURCE OF IT ALL

BIG DATA

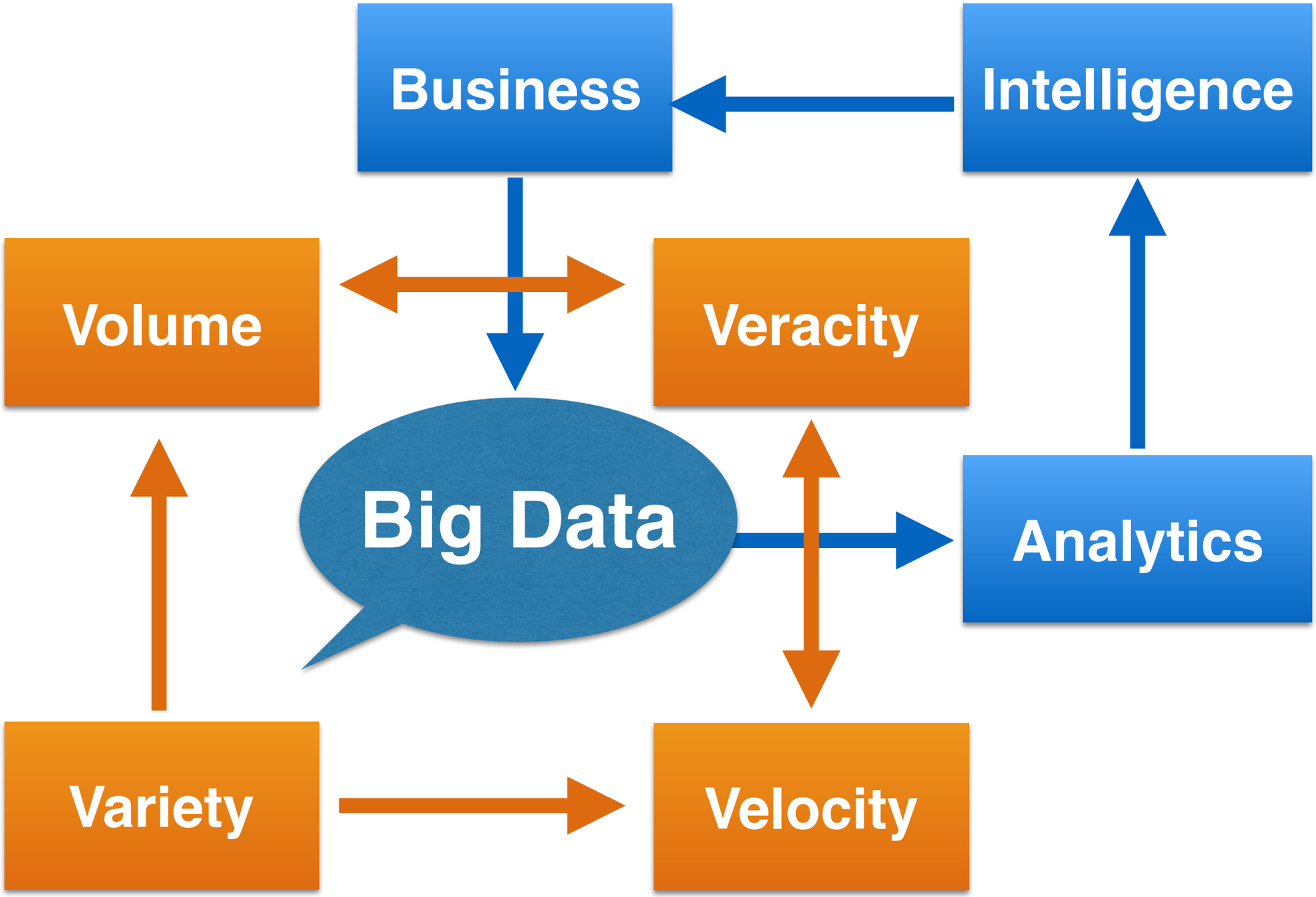
Big data is used to describe data sets that are too large and complex to be processed and analyzed using traditional technologies.

THE DIGITAL UNIVERSE

50x GROWTH FROM 2010 TO 2020

1 exabyte (EB) = 1,000,000,000,000,000,000 bytes

Source: IDC's Digital Universe Study, sponsored by EMC, December 2012



Big Data Context [3]

3. Maheshwari, A. (2016). Big data essentials. Amazon Digital Services LLC.

Volume of Data

The quantity of data generated in the world has been relentlessly <u>doubling</u> every <u>12-18 months</u>!

Traditional Data	Big Data
Measured in **Gigabytes** (GB) and **Terabytes** (TB)	Measured in **Petabytes** (PB) and **Exabytes** (1 Exabyte = 1 Million TB)

The **primary reason** for the **growth of data** is the **dramatic reduction** in the **cost of storing data (30-40% every year)**.

Velocity of Data

Traditional Data

Big Data

Velocity of Data

- **Reasons for the increased velocity of data:**

1. Increase in internet speed.

- 10MB/sec to 1 GB/sec (100 times faster)!

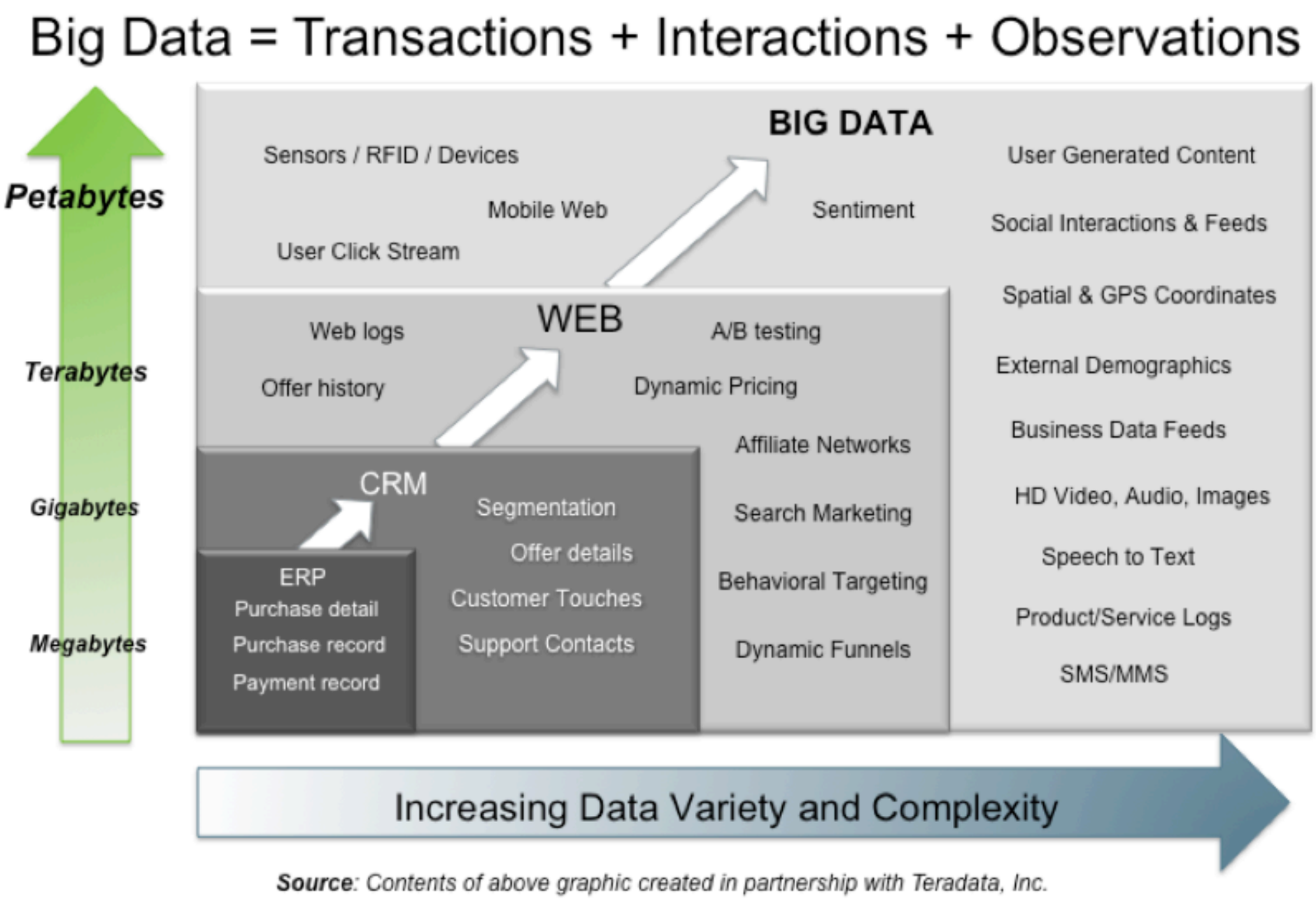
2. Increased variety of sources, such as **mobile devices**.

Variety of Data

Traditional Data

Big Data

Variety of Data



***Source**: Contents of above graphic created in partnership with Teradata, Inc.*

4. Connoly, S. (2012). 7 key drivers for the big data market. URL: http://hortonworks.com/blog/7-key-drivers-for-the-big-data-market/.

Veracity of Data

- **Veracity** relates to the **truthfulness**, **believability** and **quality of data**.

- **Big Data is messy.**

Veracity of Data

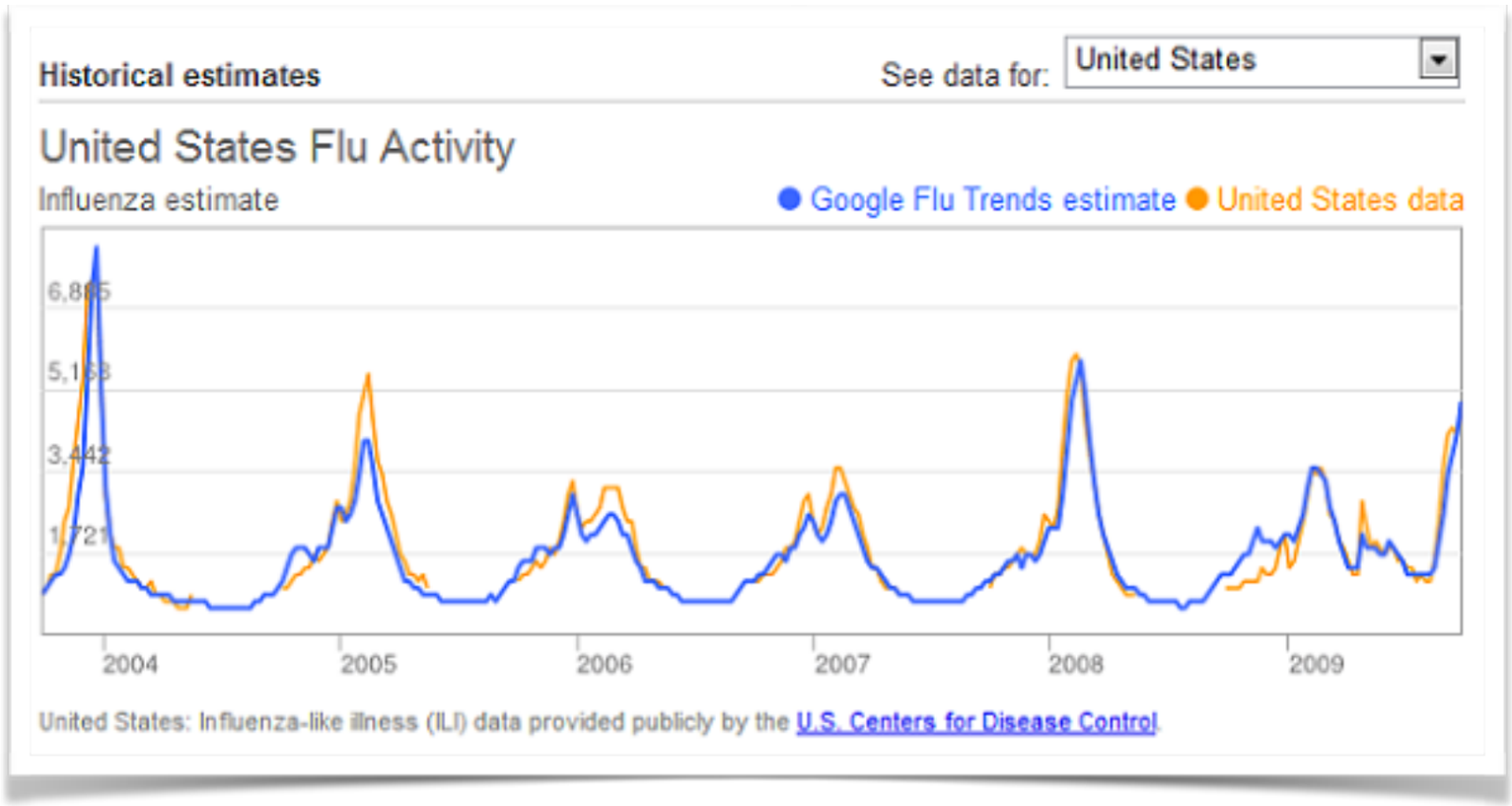
The **source of information** may **not be authoritative**.

The New York Times

WIKIPEDIA
The Free Encyclopedia

Big Data Applications

- **Public Health Monitoring**



Google Flu Trends better predicted flu outbreaks than U.S. Centers for Disease Control and Prevention (CDC) (2004-2012)

Big Data Applications

- **Predictive Policing**

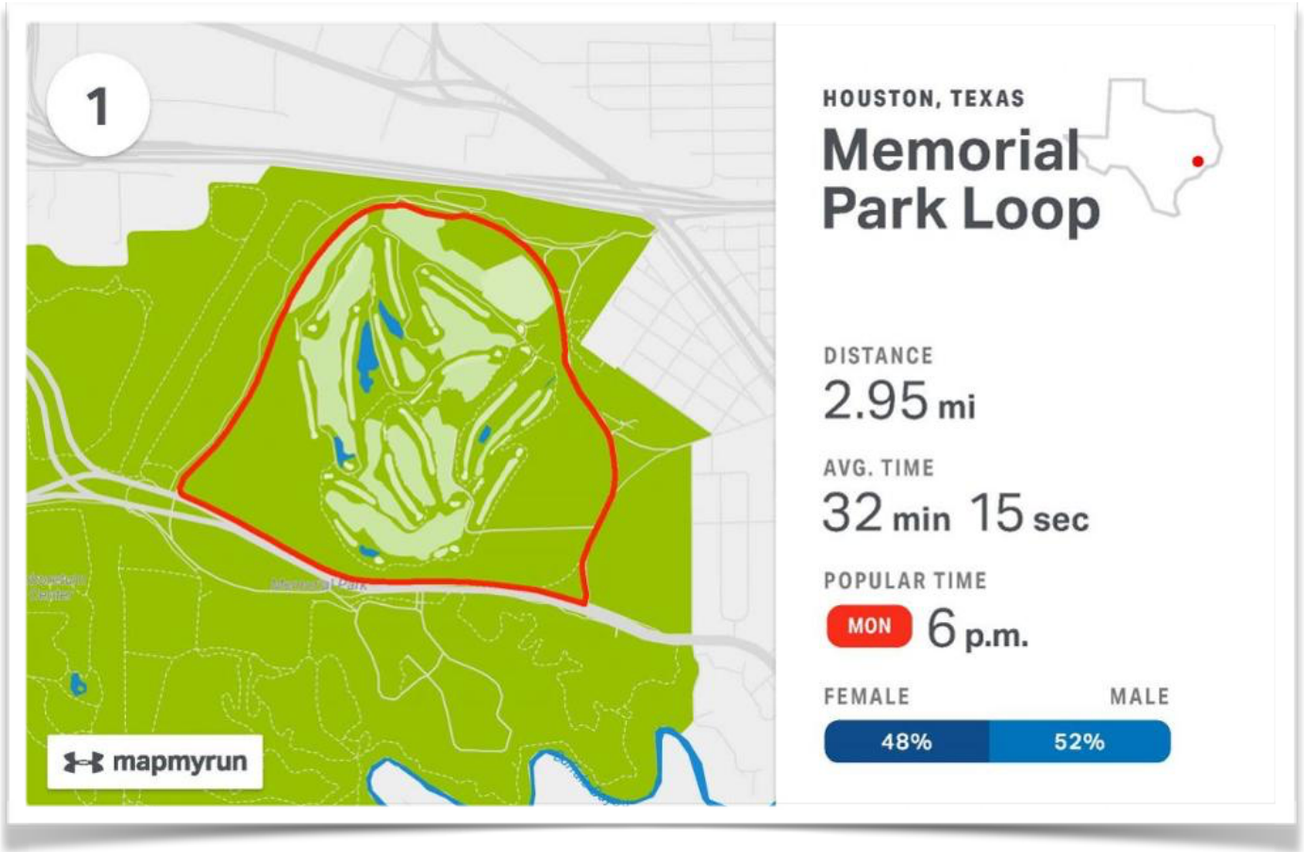
  - **Los Angeles Police Department + UC Berkeley**

  - Analyzed **13 million crimes** recorded over **80 years**

  - The identified **hotspots of crimes** were likely to happen in the **future!**

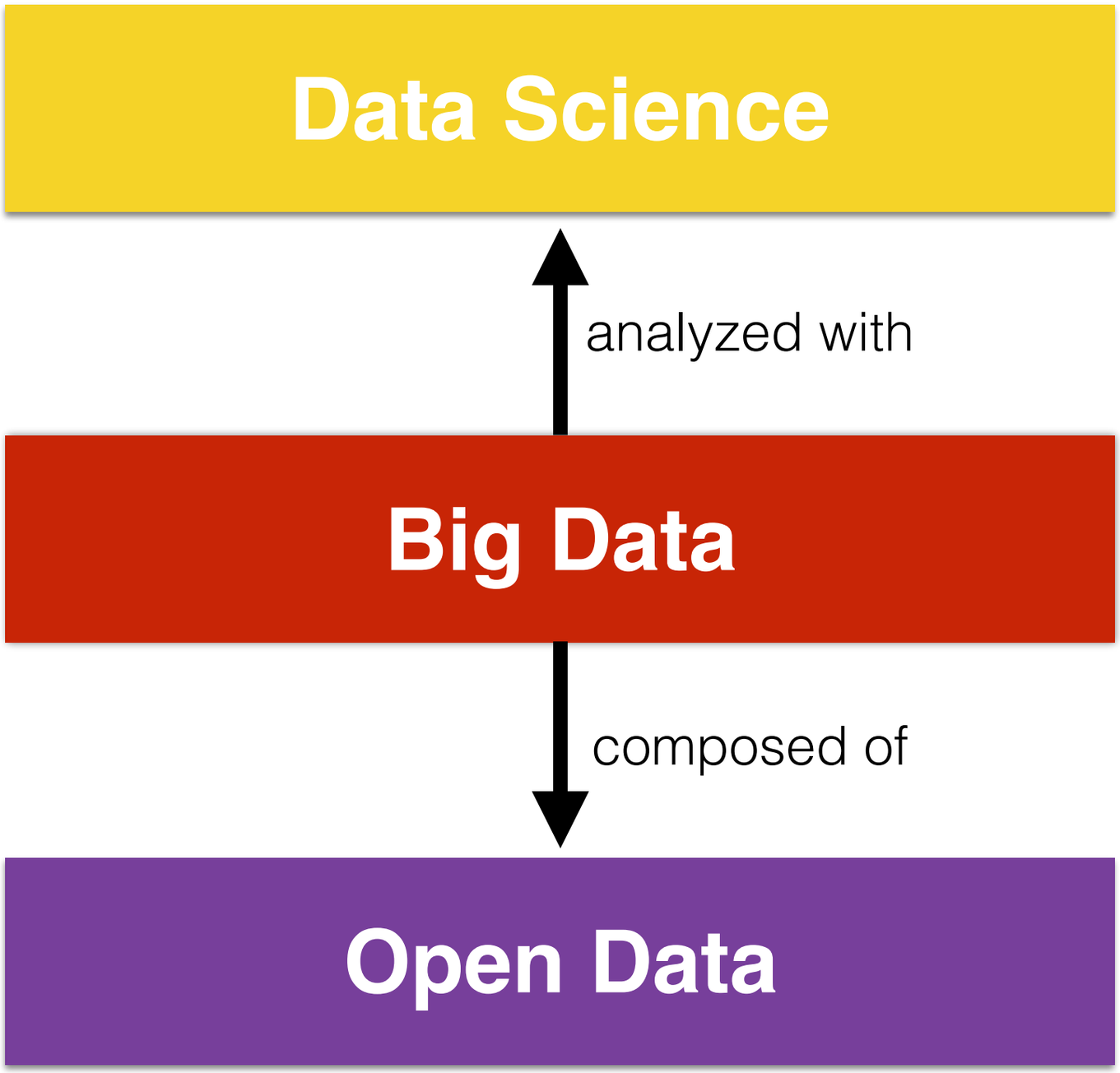
  - **Reduce crime by 12% to 26%** in different categories of crime

Big Data Applications



Under Armour knows:

- The most popular running route in the country.
- The most popular time for this route to be run is Monday at 6pm.
- Among the 20 most popular running routes in the country, 14 of them are next to or around a body of water (river or lake).


Data Science
analyzed with
Big Data
composed of
Open Data

Data Science can be defined as the study of the generalizable extraction of knowledge from data [5].

5. V. Dhar, “Data science and prediction,” *Commun. ACM , 56* (12, 2013), pp. 64-73.

- **Data science:**
  - A **multi-disciplinary approach**.
    - Programming skills + math + statistics knowledge + machine learning.
  - ***Not*** only restricted to big data.
  - Data science is the **new statistics** [6].

6. Miller, T. (2015). Marketing data science. Pearson.

Why do we need a new term like **data science** when we have had statistics for centuries?

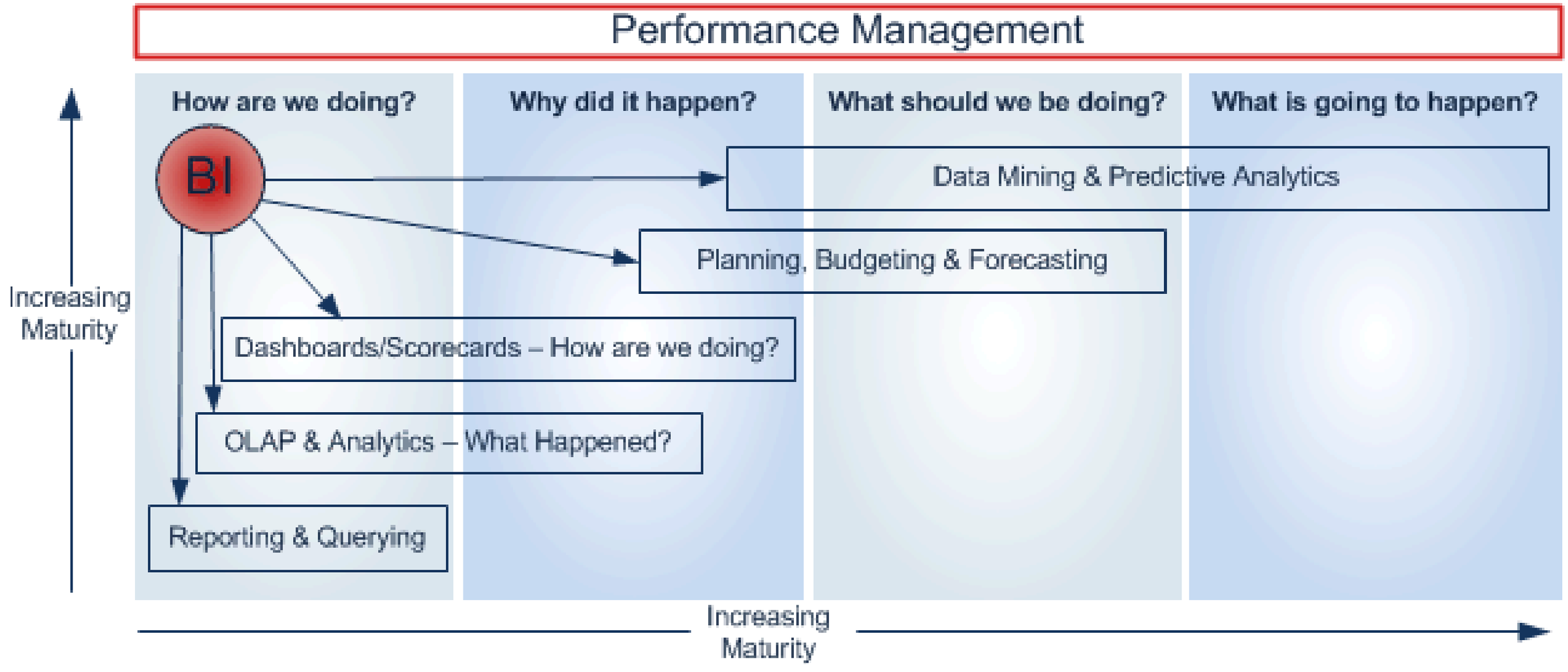
1. The raw material, the "**data**" part of data science, is increasingly **heterogeneous** and **unstructured**.

2. Traditional database methods are *not* suited for **knowledge discovery**.

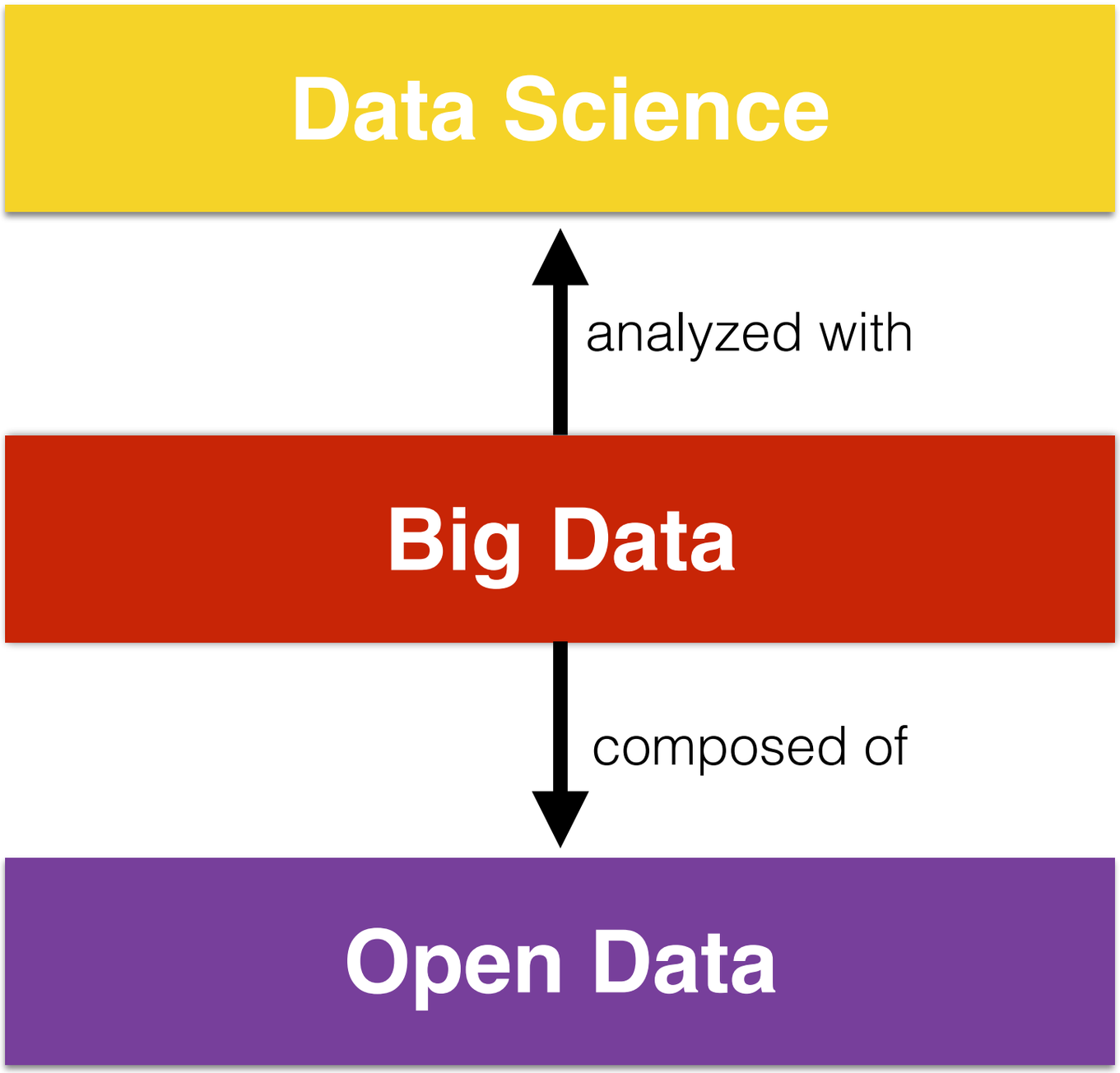
Unlike **database querying**, which asks "What data satisfies this pattern (query)?"

discovery asks "What **interesting** and **robust patterns satisfy** this **data**?"

Data Science in the Context of Business Intelligence

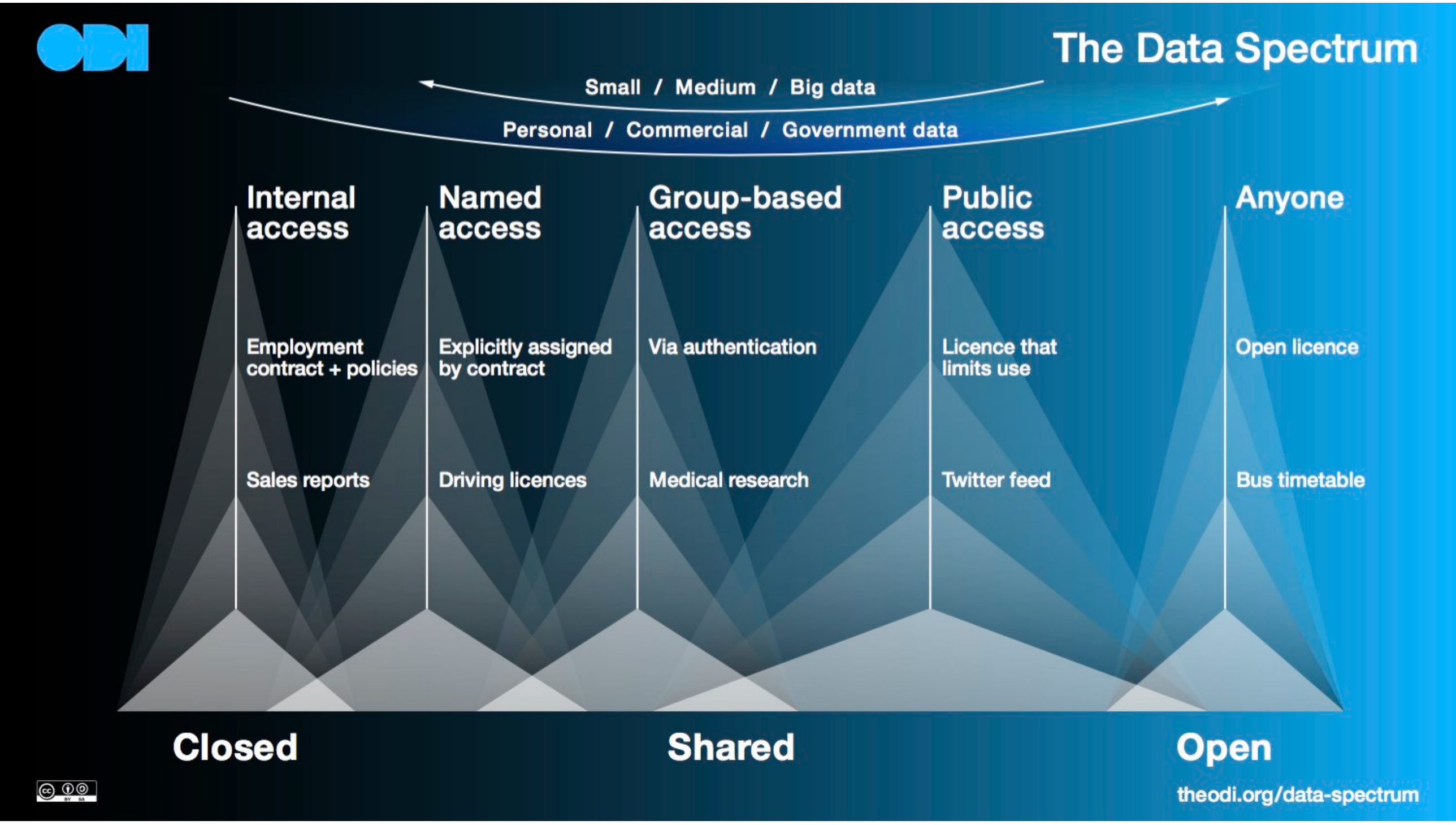


7. University of Glsgow (2015). Business intelligence strategy. URL: http://www.gla.ac.uk/media/media_434476_en.pdf


Data Science
analyzed with
Big Data
composed of
Open Data

Open data is **data** that **anyone** can **access**, **use** or **share**. Simple as that.

The Data Spectrum



8. Open Data Institute (n.d.). The data spectrum. URL: http://theodi.org/data-spectrum.

Agenda

1. What is big data, data science, and open data?
2. **Successful case studies**
3. How to use data science at your conference
4. Conclusions

Big Data for Reaching a Big World

Andrews Research Conference

9. Alférez, G.H. (2015). Big Data for Reaching a Big World. Adventist Review, 192(11), 47-51

Use **big data analysis** to try to understand how **culture** perceives our **fundamental beliefs.**

Big Data and Our Church

- In this study, the computational data analysis was based on **culturomics**.
- The application of high-throughput data collection and analysis to the study of human culture [10].

10. Michel, J.-B. et al. (2011). Quantitative analysis of culture using millions of digitized books. Science, 331(6014):176–182.

Big Data and Our Church

- The full **data set** used in the experiments is available for download at:

  https://books.google.com/ngrams

- This data set is composed of digitized texts containing about **4% of all books ever printed** between **1800** and **2008** (5,195,769 books).

- **Books** in **English** (361 billion words) and in **Spanish** (45 billion words)

Big Data and Our Church

- The corpus **cannot be read by a human** [11]:
  - If you try to read only English-language entries from the year 2000 alone, at the reasonable pace of 200 words/min, without interruptions for food or sleep, it would take **80 years**.

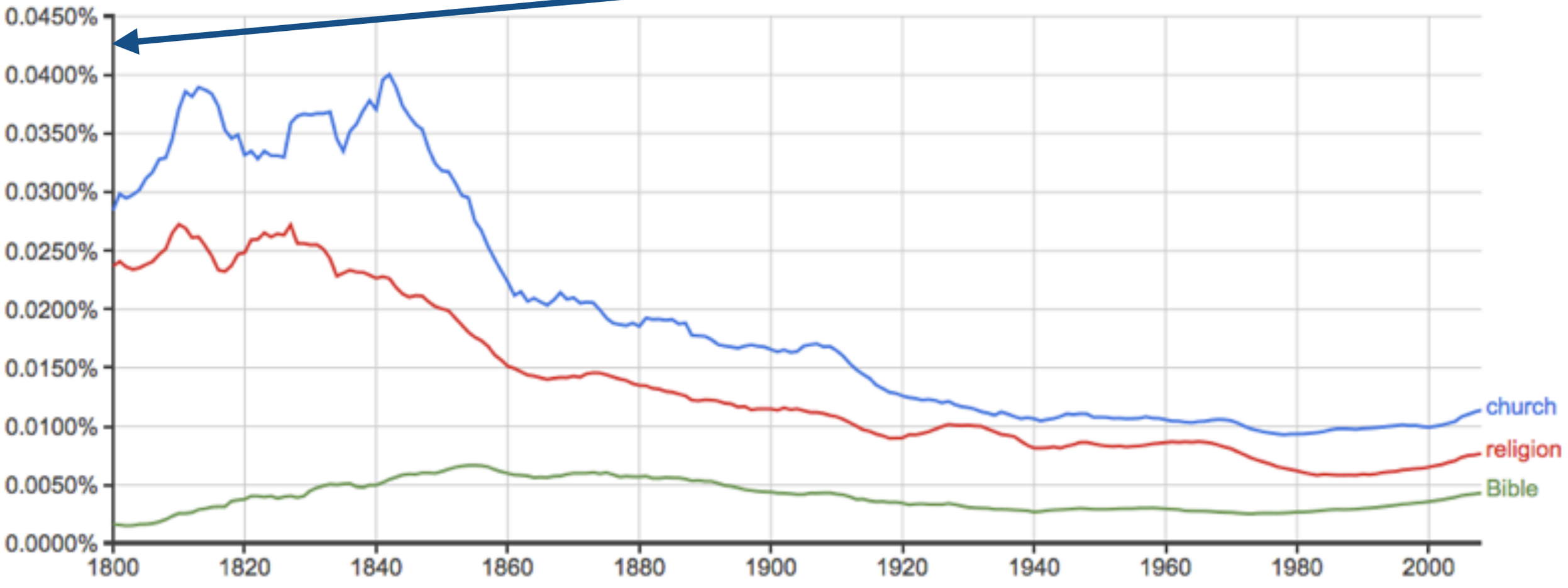
11. Michel, J.-B. et al. (2011). Quantitative analysis of culture using millions of digitized books. Science, 331(6014):176–182.

Big Data and Our Church

- The **Google Ngram Viewer** was used to visualize the results.
  - A **1-gram** is a string of characters uninterrupted by a space. This includes words (“car”, “MICHIGAN”) but also numbers (“3.14”) and typos (“excesss”).
  - An **n-gram** is a sequence of 1-grams, such as the phrases “stock market” (a 2-gram) and “the United States of America” (a 5-gram).

Church, Religion and Bible

N-gram Frequency (Corpus of English Books)



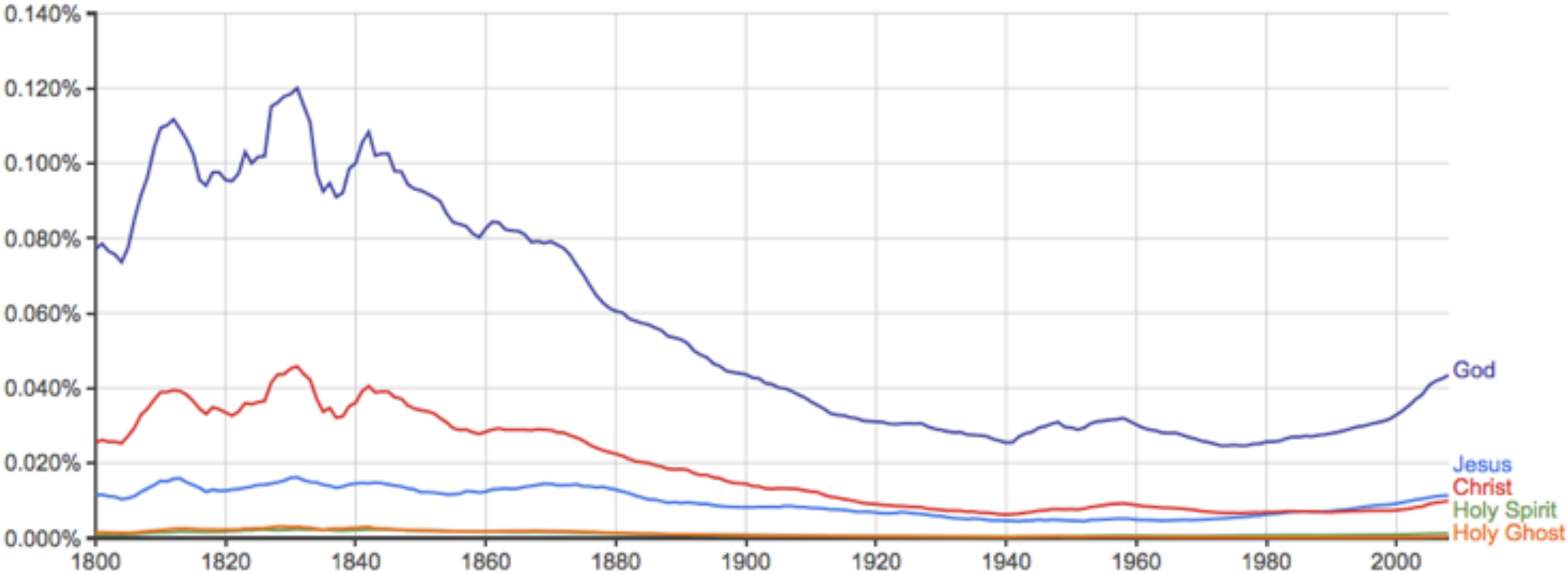
Church, Religion, and Bible Trends

Years

Secularization

Secularization Trend

God



God-Related Trend

Creation

Creationism vs. Theory of Evolution (Case Insensitive Search)

The Sabbath

Sabbath Trend

Nature of Man

Sin and Salvation Trends

The Second Coming of Christ

Second Coming of Christ, English Vs. Spanish Trends
(Case Insensitive Search)

Healthy Living

Increasing Interest in Healthy Living and Vegetarianism

Understanding the Needs of People in Big Cities through Data Science

12. Alférez, G.H. (2016). Tweeting in New York City - Data Science Can Teach Us to Sympathize. Adventist Review, 193(2), 47-49

Cities are growing fast

- **66%** of the world's population will live in urban areas by **2050** [13].

- There are more than **500** cities with a population of **1 million or more people**. However, these cities have an average of **1 Adventist congregation** for every **89,000 people**! [14].

13. Department of Economic and Social Affairs, United Nations, "World's Population Increasingly Urban with More than Half Living in Urban Areas," *United Nations* (July 10, 2014) https://www.un.org/development/desa/en/news/population/world-urbanization-prospects.html; retrieved November 10, 2015.

14. Oliver, A. "Adventist Church Implements Assessment Plan for Urban Mission," *Adventist News Network* (October 25, 2013) http://news.adventist.org/en/all-news/news/go/2013-10-25/adventist-church-implements-assessment-plan-for-urban-mission/; retrieved November 11, 2015.

“**The work in the cities is the essential work for this time.** When the cities are worked as God would have them, the result will be the setting in operation of **a mighty movement such as we have not yet witnessed**” [15].

15. White, E. G., Medical Ministry (Pacific Press Pub, 1963), p. 304.

“the Savior mingled with men as one who desired their good. He showed His sympathy for them, ***ministered to their <u>needs</u>***, and won their confidence. Then He bade them, ‘Follow Me.’” [16]

16. White, E. G., The Ministry of Healing (Review & Herald, 1905), p. 143.

Use **data science** to understand the **needs** of people in **New York City**.

Which ***data*** to use to understand the needs of people in big cities?

17. Simonite, T., "Twitter Boasts of What It Can Do with Your Data," *MIT Technology Review* (October 21, 2015) http://www.technologyreview.com/news/542711/twitter-boasts-of-what-it-can-do-with-your-data/; retrieved November 10, 2015.

Reaching People's Tweets

Sentiment analysis was used to discover the **needs** of people from tweets.

> The **computational study** of **opinions**, **sentiments**, and **emotions** expressed in text [18].

Sentiment analysis has been **satisfactory** used to classify users' sentiments in tweets [19].

18. B. Ling, "Sentiment Analysis and Subjectivity," in N. Indurkhya, & F. J. Damerau, *Handbook of Natural Language Processing*, 2nd ed., (Boca Raton, Fl: Chapman & Hall, 2010), pp. 627-665.

19. A. Tumasjan, T. O. Sprenger, & P. G., Sa. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.* AAAI, (2010), pp. 178-185.

Reaching People's Tweets

- **Tweets** are **classified**
  - as ***positive*** when they communicate a positive sentiment, such as happiness;
  - as ***negative*** when a negative sentiment is attached to them (e.g. sadness);
  - and as ***neutral*** when no emotions are implied.

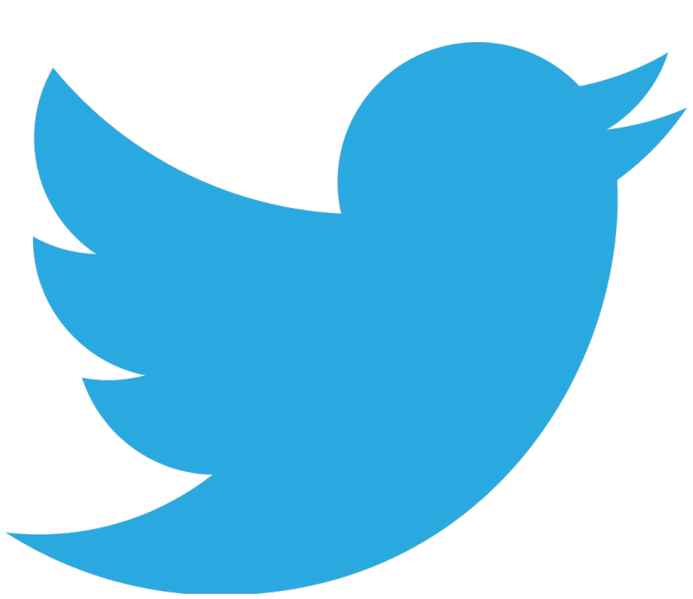
Reaching People's Tweets

Machine learning [20] was used as a tool to differentiate tweets with *positive*, *negative*, and *neutral* sentiments.

Machine learning explores the study and construction of **algorithms** that can **learn from** and **make predictions on data**.

20. A. Go, R. Bhayani, & L. Huang, *Twitter Sentiment Classification using Distant Supervision* (Stanford University, 2009)

Listening Closely to the Birds



Over a period of six weeks (September 22 to November 3, 2015), we collected 2,084 tweets from New York City, 1,633 of them bearing positive sentiments and 451 expressing negative sentiments. Tweets with neutral sentiments were not collected.

Listening Closely to the Birds

30 specified keywords:

Adventist, addiction, Bible, children, Christ, church, contamination, divorce, education, elderly, exercise, family, God, health, Jesus, obesity, peace, poverty, religion, rest, safety, salvation, Savior, stress, teenagers, teens, terrorism, vegetarian, violence, youth

1. Collects

2. Classifies

+ and -

3. Stores

```
                str(j["coordinates"]).split(",")[1].split("[")[1]+"\\*"+
                str(j["coordinates"]).split(",")[2].split("]")[0] +"\n")
            fd.close()

        elif int(j["polarity"]) == 4:
            positive_tweets += 1
            #print "Positive^ " + str(j["username"]) +"^ "+str(j["screen_name"]) +"^ "+str(j["date"])+"^ "+ j["text"] +"^ "+
(str(j["coordinates"]).split(",")[1])+"^ "+ str(j["coordinates"]).split(",")[2]
            fd = open('DataSets/'+user_city+'_'+EXPRESSION+'.csv','a')
            fd.write("Positive" + "\\*"+
                str(j["postId"])+"\\*"+
                str(j["username"]) +"\\*"+
                str(j["screen_name"])[2:len(str(j["screen_name"]))-1] +"\\*"+
                str(j["date"])+"\\*"+
                j["text"].replace("\n", "") +"\\*"+
                str(j["coordinates"]).split(",")[1].split("[")[1]+"\\*"+
                str(j["coordinates"]).split(",")[2].split("]")[0] +"\n")
            fd.close()

    return negative_tweets, positive_tweets

def main(argv):

    global user_city
    global EXPRESSION
    global radio

    try:
        opts, args = getopt.getopt(argv,"he:c:r:l:",["expr=","city=","ra
    except getopt.GetoptError:
        print os.path.basename(__file__) + ' -e <expression> -c <city> -
        pass
    for opt, arg in opts:
        if opt == '-h':
            print os.path.basename(__file__) + ' -e <expression> -c <cit
            sys.exit()
        elif opt in ("-e", "--expr"):
```

python

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Positive*65	friends	new family and... https://t.co/H23IGJNwc0*-73.99278645* 40.75052792																
2	Positive*65	haters and family near and far and my Jetblue family@too lets... https://t.co/IDsQLyobRl*-73.77555571* 40.64579177																	
3	Positive*651003442270351360*248009151*AliceLaurissa*2015-10-05 11:58:10*Hanging with my nieces for my family bday dinner. Love these guys. #alicelaurissa #birthdaygirl... https://t.co/i1GL5dBqIc*-73.74631052* 40.70295328																		
4	Positive*650915559132000256*2472332402*cityscapesny*2015-10-05 06:08:57*TOMORROW NIGHT! #JADAKISS @therealkiss #THOTLINEBLING HNY Family invades #sooperstarmondays with The... https://t.co/nuQiyLen4g*-73.9118195* 40.725																		
5	Positive*65	my man	my family. I love you SO much Mama Nancy- you are a wonderful mom- we celebrate YOU... https://t.co/zqCXfCvFNr*-73.8831482* 40.7555008																
6	Positive*650860212287590400*50217891*ALovelyToy*2015-10-05 02:29:01*Love my family!!!! HAPPY BIRTHDAY TO MY BABY SISTER!!!!! LOVE U2 til DEATH DO US AND BEYOND!!!! LIBRA... https://t.co/FJHGxEOemc*-73.983223* 40.7449989																		
7	Positive*650848490151653376*67786749*MannyJR99*2015-10-05 01:42:26*This is my beautiful family God bless my Beautiful family I love you... https://t.co/YT4PLHkQks*-74.0064* 40.7142																		
8	Positive*650847671297536001*110518597*WJBLAl*2015-10-05 01:39:11*Suite life w/ the family! Thanks for the great memories! Mets won to put the cherry on top... https://t.co/Slf4EGjG45*-73.84604275* 40.75685645																		
9	Positive*650826880317128704*91659013*fritziedotnyc*2015-10-05 00:16:34*I always love spending time with my family. @ Tamashi Ramen https://t.co/owJBqIEDZF*-73.92671723* 40.76262602																		
10	Positive*650820329963655168*2658278567*Fortheloveto*2015-10-04 23:50:32*My family loves this crest #sugarshield it is the best!! With cleaner teethes and fresher breath for... https://t.co/w79APDdQ25*-74.0064* 40.7142																		
11	Positive*650818621313302529*3777775125*_wavydayv*2015-10-04 23:43:45*Happy Birthday Cineya #Family @ Sweet Chick https://t.co/1qOD2gFqcT*-73.9573975* 40.718399																		
12	Positive*650795826457870336*1571019223*IntoTheRowBlog*2015-10-04 22:13:10*Happy Birthday to october_babi! Enjoying another round of drinks just for you #family #instacool... https://t.co/4yfIzAKf9B*-74.2596588* 40.7449684																		
13	Positive*650785188331413504*16436343*the_captain66*2015-10-04 21:30:54*What a perfect day to watch the PLAYOFF BOUND NY METS with my favorite mets loving family glenbing... https://t.co/5xh0nerOi9*-73.84604275* 40.75685645																		
14	Positive*650760792846036993*753249602*Cinderella_miu*2015-10-04 19:53:58*Sunday tapping with my New York family @ Champion Dance Studios https://t.co/Z1GHb57PKS*-73.9903564* 40.7552299																		
15	Positive*650745263175401472*851804210*wellingtonmk6*2015-10-04 18:52:15*Happy Sunday!! #family #sunday #funday #sunny #intrepid #nyc gabygarcia20... https://t.co/RmoPivldW8*-73.99992301* 40.76453124																		
16	Positive*650745035844100099*14768046*baristanet*2015-10-04 18:51:21*Love this family that turned out to honor Yogi @ Yogi Berra Museum & Learning Center https://t.co/Eb5nQ14gA3*-74.19464988* 40.86826073																		
17	Positive*650736517833433089*14301804*travelingheidi*2015-10-04 18:17:30*#Hamilton with the family! So excited!!! (@ Richard Rodgers Theatre - @nederlanderbway for Hamilton (NY)) https://t.co/wXxBDxC72A*-73.98670752* 40.75930975																		
18	Positive*650725601553305600*71097356*BRONXGURL12*2015-10-04 17:34:07*#aboutlastnight Family Fun with My Brother My Daughter And My Godkids... https://t.co/sl1swKyY6h*-73.98872539* 40.73639096																		
19	Positive*65	family & finally signing a lease! @ MP Taverna Astoria https://t.co/AHn3JXwdfH*-73.90966687* 40.77549002																	
20	Positive*650699463791505409*255800296*drjordanmetzl*2015-10-04 15:50:16*Best morning ever! Hangin with my niece in Central Park! #family #NYC #newyork #outdoors @ Balto... https://t.co/TyrCGju6P8*-73.9701224* 40.77005404																		
21	Positive*650684013724631040*374570973*KingsleyGil*2015-10-04 14:48:52*Blessed that I can worship along side my family watching this wonderful... https://t.co/X5dKfIUvan*-73.83593728* 40.6957819																		
22	Positive*650674928929386496*302211068*JohnTripodi59*2015-10-04 14:12:46*Sunday morning NEDA walk with the family #TeamBec @ Foley Square https://t.co/U2qOCk5lt*-74.00297031* 40.7146602																		
23	Positive*650585442094460928*2904596325*dlopez580*2015-10-04 08:17:11*Happy birthday to me! #birthday #family #sister #brother ##nyc @ Therapy NYC https://t.co/z1lIGnPAKL*-73.9869537* 40.7640419																		
24	Positive*650574034401423360*105627214*rblang3*2015-10-04 07:31:51*Great to have role models of family and relationship like President and Mrs Obama!!!! Happy Anniversary!!!*-73.60379227* 40.71022649																		
25	Positive*650560138496081920*248939856*DANTHEMANBABY*2015-10-04 06:36:38*SQUAD UP FAMILY RIGHT HERE HAPPY BDAY RICK I LOVE YOU KID #UnionDjsInc #UnionDjsIncManagement... https://t.co/lUboNhTNnL*-73.91515992* 40.760																		
26	Positive*650548266300534784*31569621*jeffbeacher*2015-10-04 05:49:27*We are live from New York !!! On Saturday night !!!! #Snl #family the queen hosting @mileycyrus @... https://t.co/kqpPAMaVuV*-73.97891732* 40.75882475																		
27	Positive*65	let's give it up for @_sotor24 's... https://t.co/4s3shuytyH*-73.9839783* 40.7439308																	
28	Positive*650519060069916672*103855993*irunbk718diva*2015-10-04 03:53:24*I swear ain't nothing like family!!! I LOVE MY COUSINS!!! #familyfirst... https://t.co/QZnmVChlME*-73.96116732* 40.71674563																		
29	Positive*650503477429411840*2867179032*FLIP_NYC*2015-10-04 02:51:29*Family affair happy 50th Simon and Bsnk$ happy 200th old man ... https://t.co/4cT29pQFxU*-74.0064468* 40.7406578																		
30	Positive*65	NY w/ @justinbreton @alliecrozier) https://t.co/iNMFw9JtbH http://t.co/5cDGho3y3u*-73.9807189* 40.7301443																	
31	Positive*650493057322852352*428652528*7ames4be*2015-10-04 02:10:05*Happy wedding anniversary my man president B Obama. You're a great role model. May the Lord continue 2 strengthen you and bless your family*-73.8413014* 40.87																		
32	Negative*6	and his own	died in the (2/3)*-74.13895387* 40.77477477																
33	Positive*650481904605597696*16570056*kathykattenburg*2015-10-04 01:25:46*@DelilahNoBSzone agree family is nice. (2/2)*-74.13895387* 40.77477477																		
34	Positive*65	New York https://t.co/eCclgrpSRa*-73.8644* 40.9415																	
35	Positive*650458289566781440*15042927*GeorgeGSmithJr*2015-10-03 23:51:55*The most important part of today is Friends and Family...That's sweet....wait...You're talking... https://t.co/BuMyd9Sfl1*-73.96762272* 40.762187																		
36	Positive*650458069240033280*1043211744*mastodon_1*2015-10-03 23:51:03*Drinks with my BFF! #family #loveher #saturdaynight @ Maja's Tapas Bar https://t.co/RS1patMCKV*-73.9275208* 40.8659515																		
37	Positive*650452529713557504*24227930*AntonioPulgarin*2015-10-03 23:29:02*What a great night with the http://t.co/F9e9EaRSn5 #GUILD Family. Special thanks to everyone who came out to... https://t.co/LVKPz5Q0Sy*-74.0022125* 40.727																		
38	Positive*65	NY) https://t.co/8zW2SM0yTe*-73.99330097* 40.75739368																	
39	Positive*65	today was a great... https://t.co/gOecid6Bcz*-73.9786606* 40.7677116																	
40	Positive*650407771691728896*46807995*CookEatForOne*2015-10-03 20:31:11*A well deserved glass of wine after surviving the Bloomingdales friends and family sale. But I made a lot of... https://t.co/pz69CyH4Tu*-74.009575* 40.706304																		

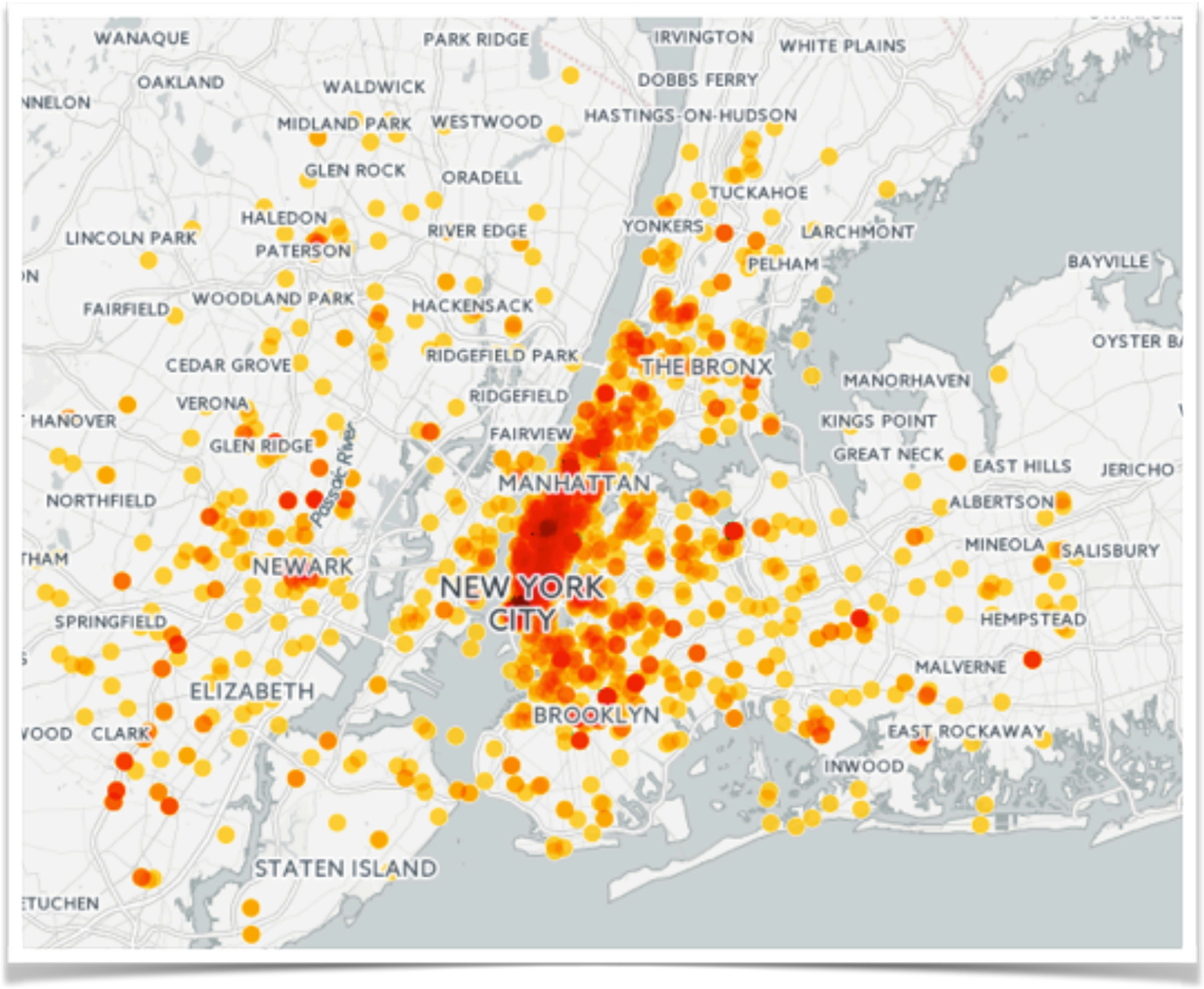
Positive Tweet about Vegetarian Food

- Positive
- her*
- 2015/10/02 02:08:16
- I want to be vegetarian. I really do. @arrogantswine @ East Williamsburg Brooklyn https://t.co/rpatPGyhXw\
- -73.939 (longitude)
- 40.714 (latitude)

Negative Tweet about Family

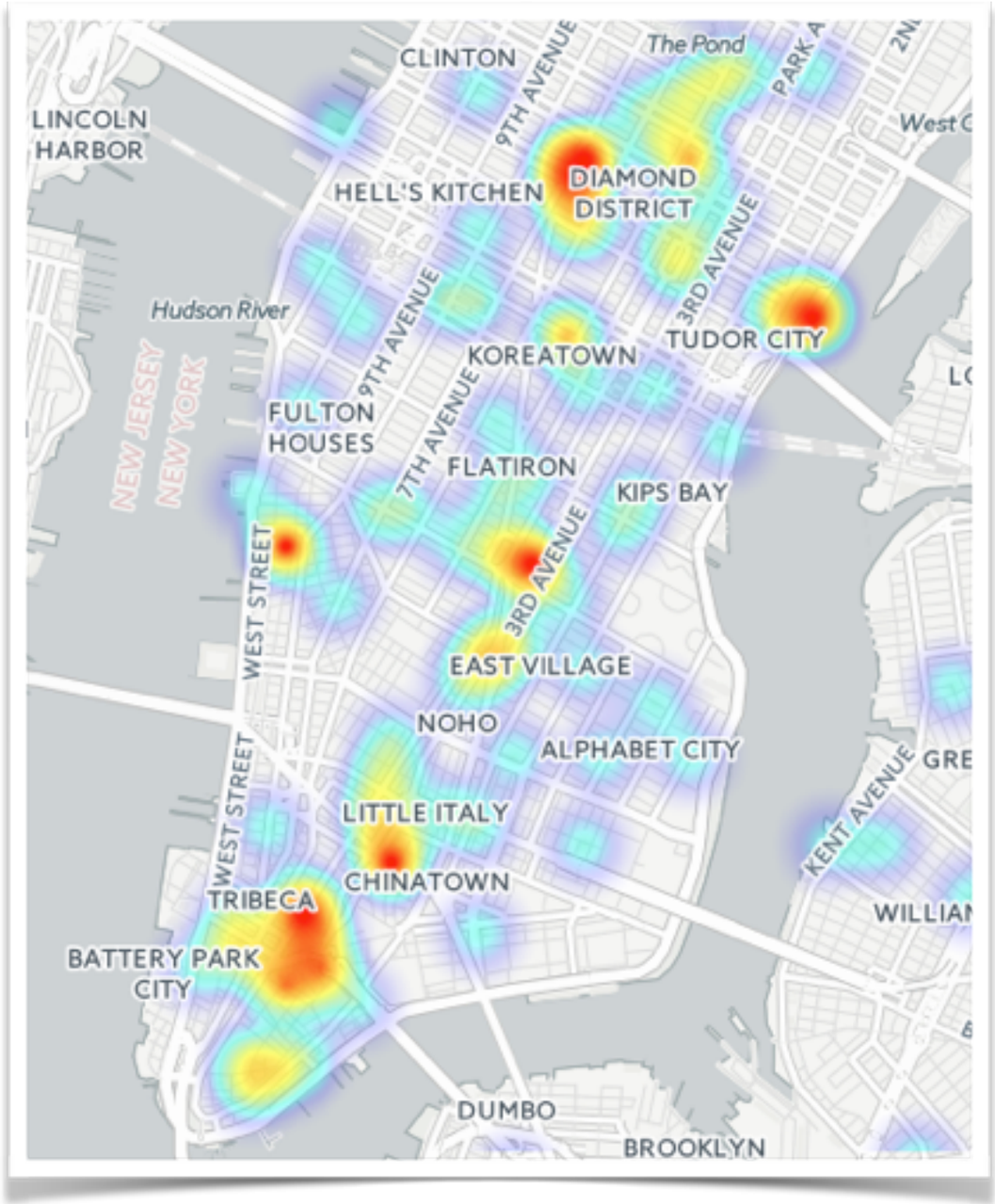
- Negative
- And*
- 11/10/15 18:48
- My ex has made them hate me, but I still see the children in my dreams.
- -73.74663446 (longitude)
- 40.69729011 (latitude)

Listening Closely to the Birds



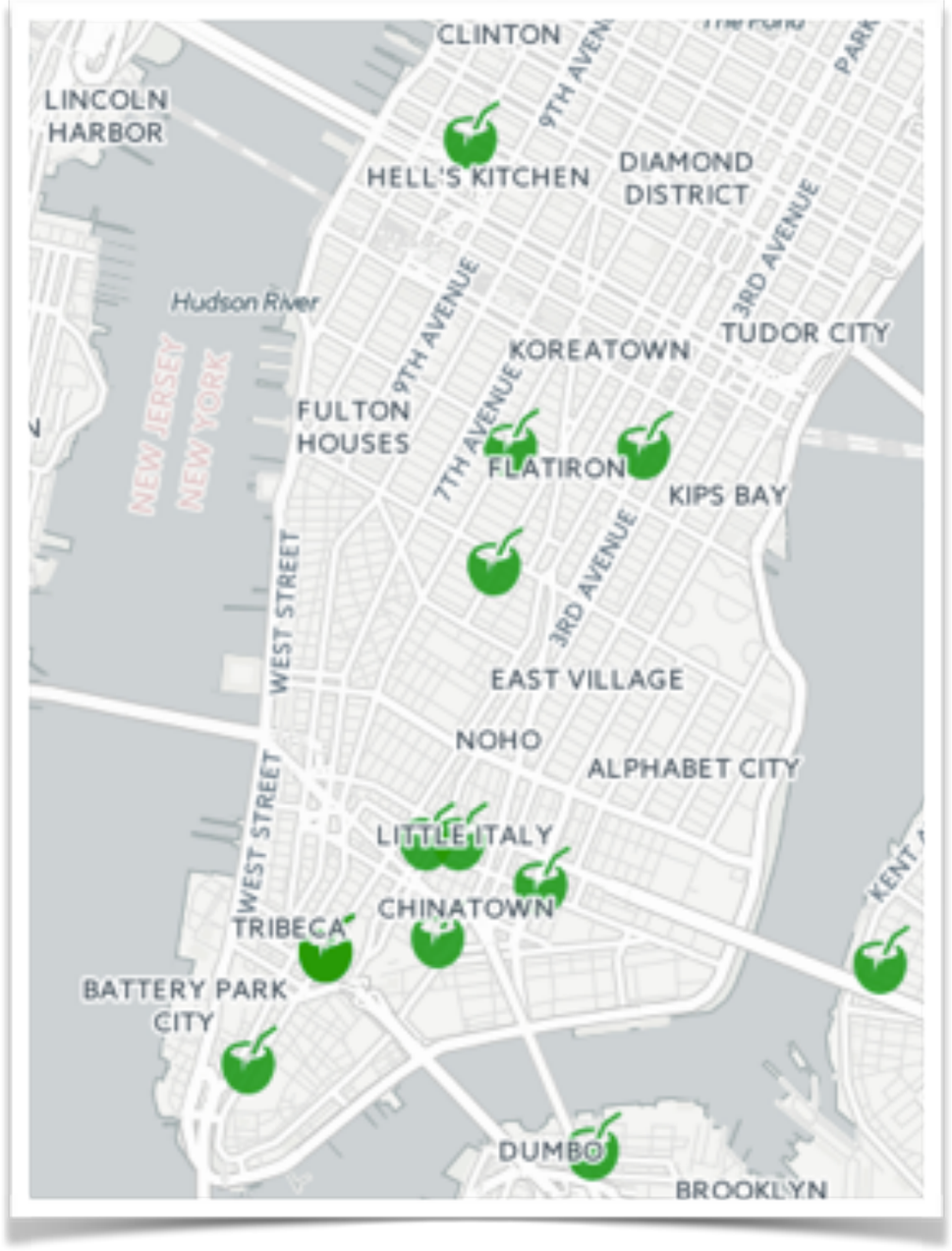
Intensity of tweets in New York City

Listening Closely to the Birds



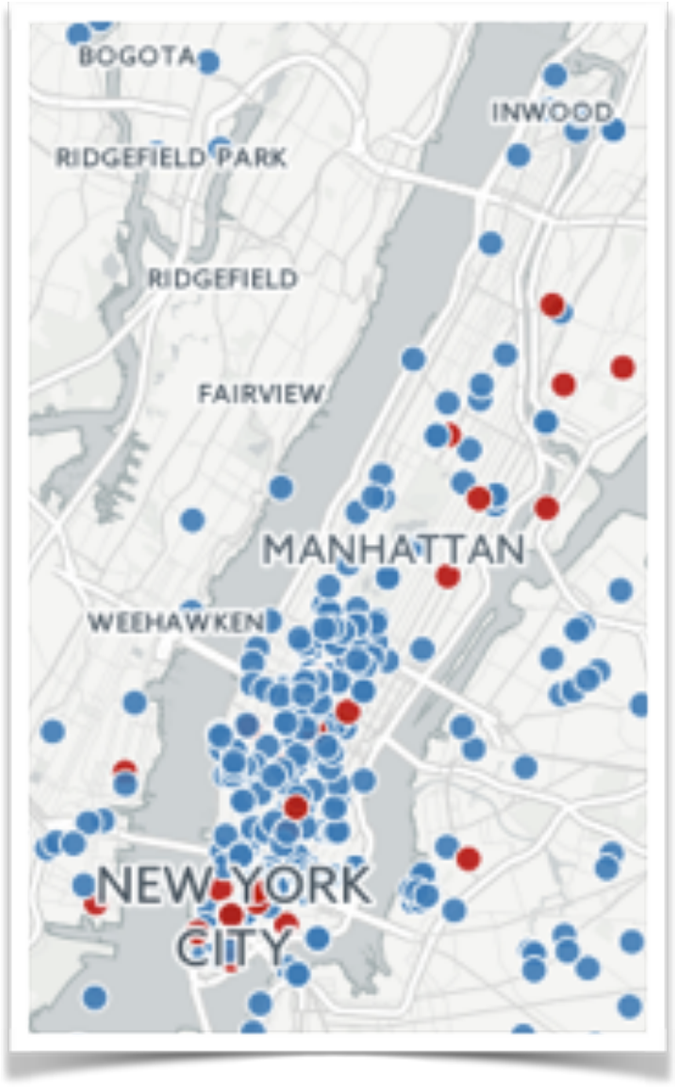
Areas with negative tweets in Manhattan

Upbeat and Downbeat



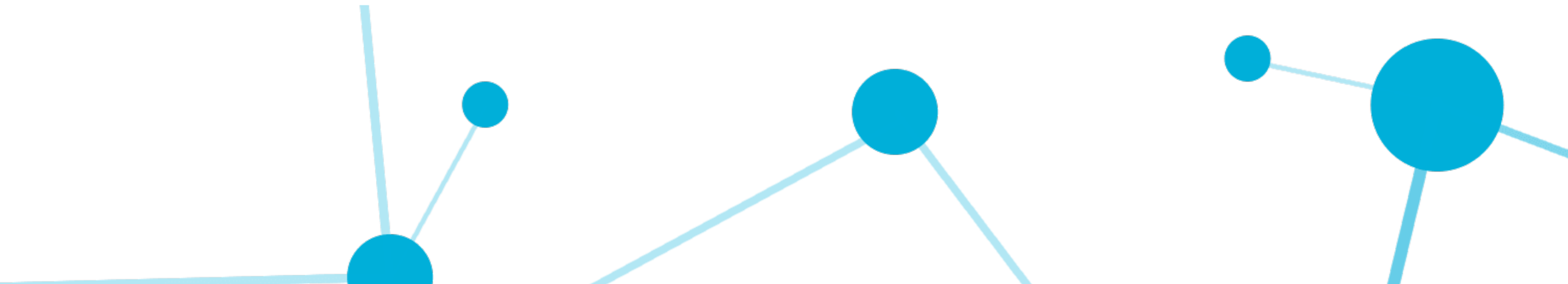
Positive tweets about vegetarian food in Manhattan

Upbeat and Downbeat



Positive [blue] and negative [red] tweets about family in Manhattan

Other Data Science Mission-Oriented Case Studies

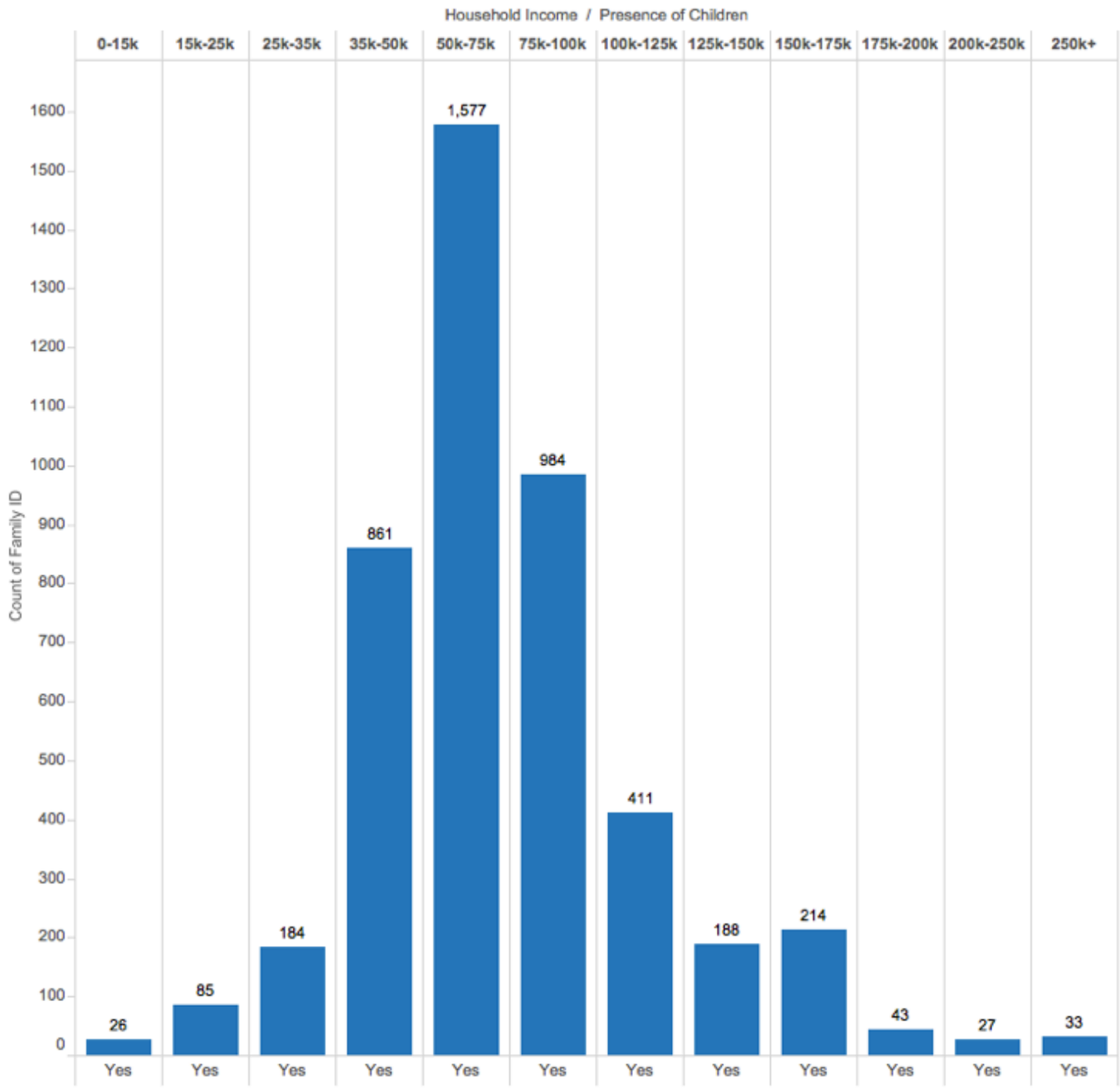


Data Analysis for Washington Conference

- What communities have the highest concentrations of members?
- What are the needs of the communities according to open data?
- What communities have no SDA presence?
- What age range do the parents fall into?
- What is the general income of families?
- What are key membership interests based on age and income?
- Geo-localization of all the Adventist churches and schools at the Washington Conference compared to the information in 2017 US Census
- Social media analysis
- Cluster analysis for marketing

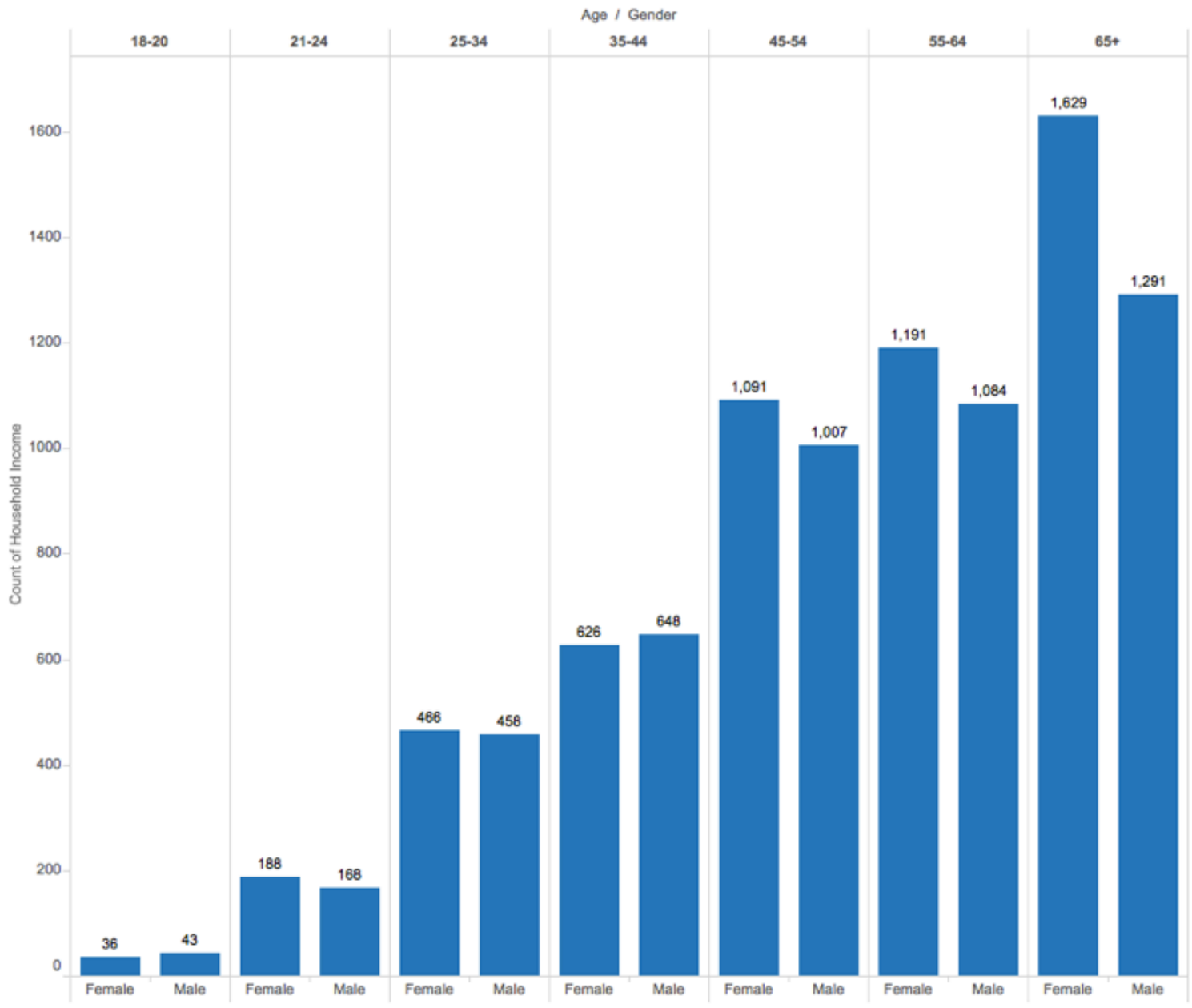
SEVENTH-DAY ADVENTIST CHURCH
NORTH AMERICAN DIVISION

Data Analysis for Washington Conference



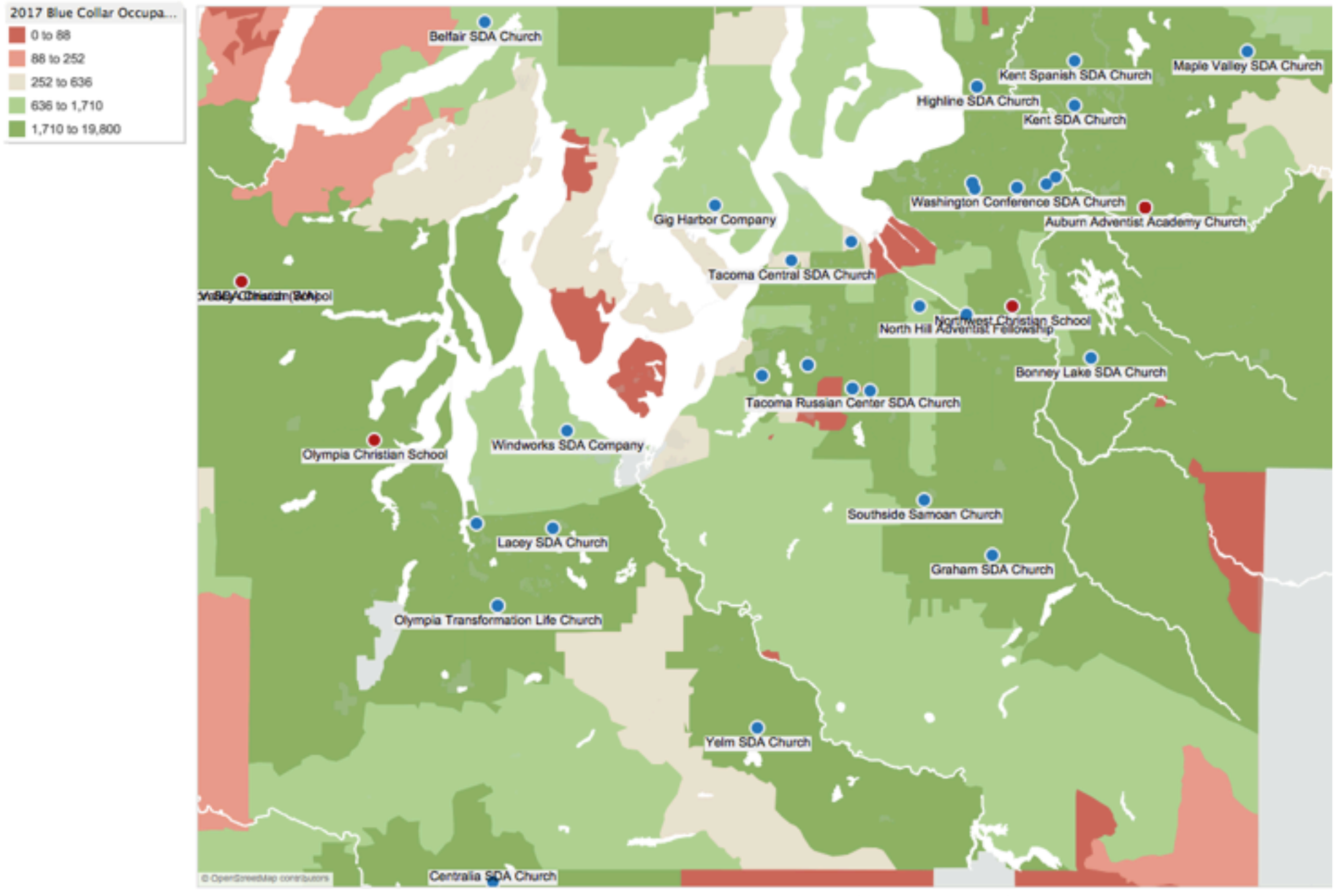
Household income of families with children at home

Data Analysis for Washington Conference



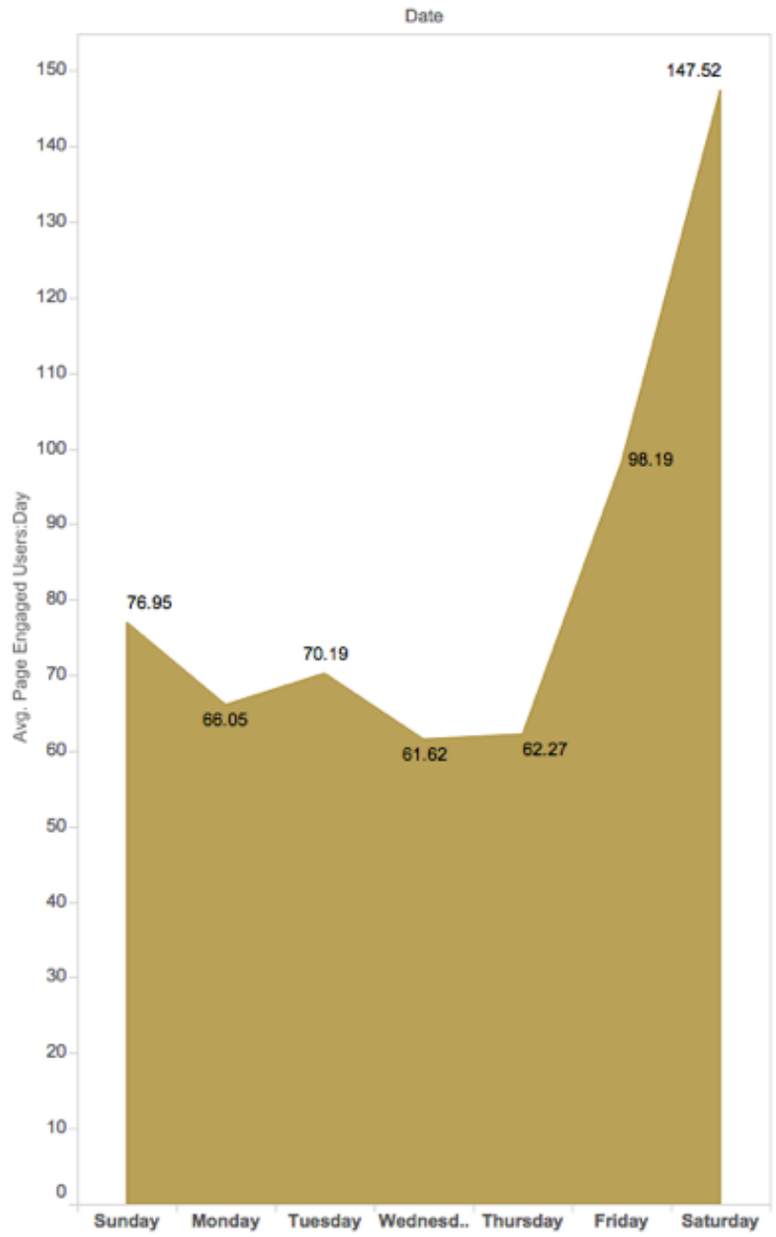
Household income by gender and age

Data Analysis for Washington Conference



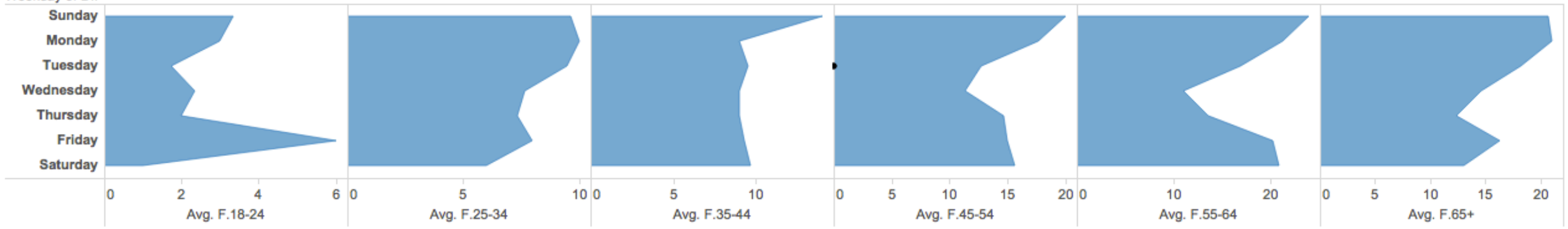
Sample map with the geo-localization of churches (in blue) and schools (in red) and census data related to blue-collar occupation (background colors)

Data Analysis for Washington Conference



Average page engaged users per day (January 12 to June 6, 2017)

Weekday of D..



Average page female visitors per day

SEVENTH-DAY ADVENTIST CHURCH
NORTH AMERICAN DIVISION

Exploring Open Demographic Data

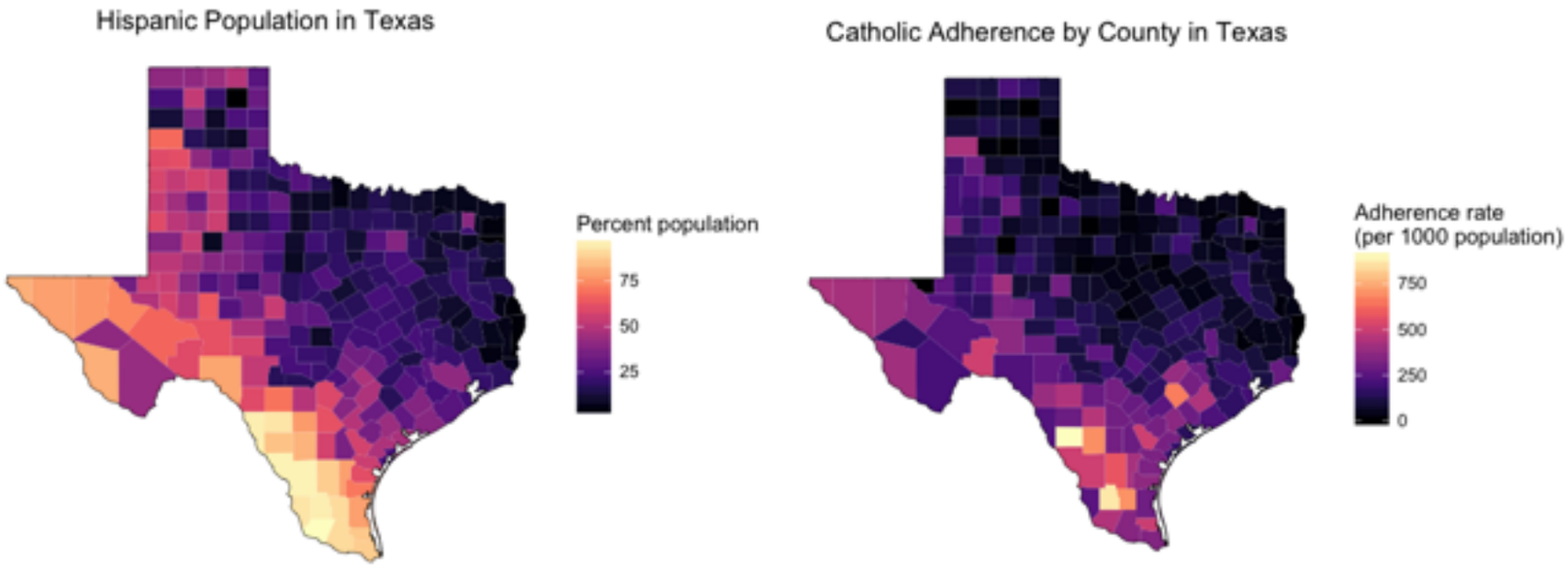
Data from the latest religion census that was published in 2010 by the Association of Statisticians of American Religious Bodies (ASARB)

+

Demographic data from the 2013 American Community Survey (ACS) county demographic data

Exploring Open Demographic Data

The correlation coefficient between the Catholic adherence rate and the percentage of the population that is Hispanic in Texas counties is high 0.7521401, with a 95% confidence interval.



Hispanic Population in Texas Vs. Catholic Adherence by County in Texas

Using Data Science to Understand Segments of Individuals Who Have been Removed from Membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013

Dr. Germán H. Alférez, *Universidad de Montemorelos,* Erón Zebadúa, *Inter-Oceanic Mexican Union Conference,* and Enoc Cruz, *Universidad Linda Vista*

Technical Report June 23, 2016. Global Software Lab, School of Engineering and Technology, Universidad de Montemorelos

Abstract—Removing individuals from membership in the Seventh-day Adventist Church is the ultimate discipline that the church can administer. Our contribution is to present how we have applied state-of-the-art data science techniques to identify the segments of individuals who have been baptized from 2005 to 2013 and also been removed from membership in the same period of time at the Inter-Oceanic Mexican Union Conference. The dataset that was analyzed is composed of 14,388 records of members who have been removed. The results can guide further church decisions to prevent membership lost, specially among youth and among people who are baptized after evangelistic campaigns. Our data-science approach could be easily extrapolated to other divisions and conferences.

21. Alférez, G.H., Zebadúa, E., & Cruz, E. (2016). Using Data Science to Understand Segments of Individuals Who Have been Removed from Membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013. Technical Report June 23, 2016. Global Software Lab, School of Engineering and Technology, Universidad de Montemorelos. Retrieved June 23, 2016, from http://www.harveyalferez.com/publications/TechnicalReport_June_23_2016_GSL_UM.pdf

1. Most individuals who are removed from membership are young and last around 3 years at church.

2. The percentage of retention of youth (15 to 33 years old) with a SDA background is very similar to the segment of youth without any religious background.

3. People who enter the church after an evangelistic campaign tend to leave the church in a higher percentage than people who take Bible courses or who are invited by friends.

21. Alférez, G.H., Zebadúa, E., & Cruz, E. (2016). Using Data Science to Understand Segments of Individuals Who Have been Removed from Membership in the Inter-Oceanic Mexican Union Conference from 2005 to 2013. Technical Report June 23, 2016. Global Software Lab, School of Engineering and Technology, Universidad de Montemorelos. Retrieved June 23, 2016, from http://www.harveyalferez.com/publications/TechnicalReport_June_23_2016_GSL_UM.pdf

100 languages

Software to Guide the Application of Data Science in the 10/40 Window. Case Study: Data Analysis in the Middle East and North Africa Union

AfricaNews, Agence France Presse, Associated Press, Associated Press Online, Associated Press Worldstream, BBC Monitoring, Christian Science Monitor, Facts on File, Foreign Broadcast Information Service, The New York Times, United Press International and The Washington Post

Discovering Mission-Oriented Patterns with Open Data in New York City

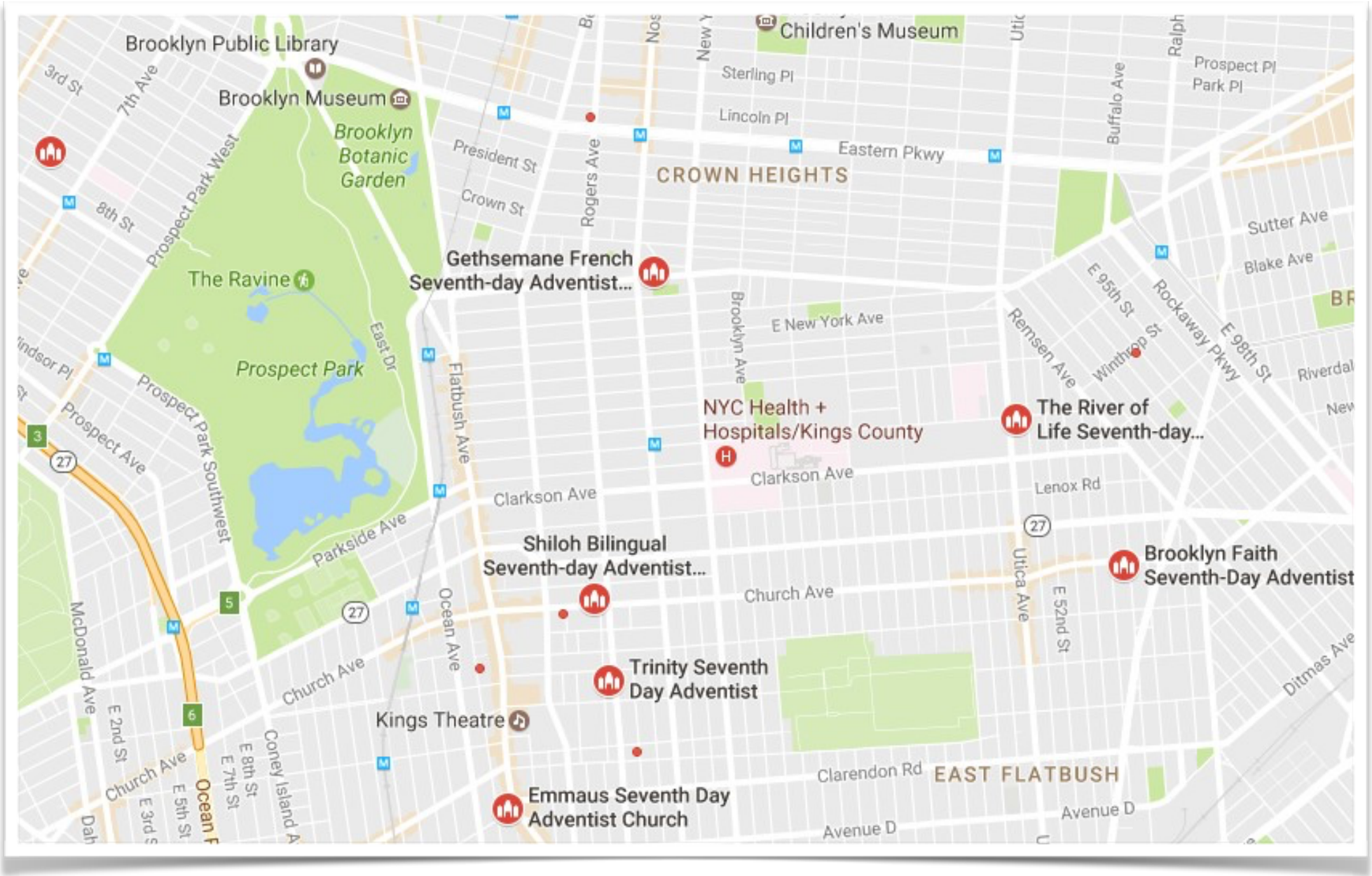
- We analyzed an open dataset of motor vehicle collisions in NYC (2014-2016), which is freely provided by NYPD. 932,904 registered incidents with 30 variables. We applied the K-Means algorithm to this dataset.

- On Thursdays, Fridays, and Saturdays, drivers tend to drive aggressively. This situation increases the number of accidents during those days.

- On Fridays, around Prospect Park, Brooklyn, there were over 77,000 pedicab accidents.

22. Alférez, G.H. (2017). Discovering Mission-Oriented Patterns with Open Data in New York City. Social Media and Big Data Services, North American Division of the Seventh-day Adventist Church. Retrieved January 31, 2017, from http://www.sdadata.org/blog/discovering-mission-oriented-patterns-with-open-data-in-new-york-city

Discovering Mission-Oriented Patterns with Open Data in New York City



22. Alférez, G.H. (2017). Discovering Mission-Oriented Patterns with Open Data in New York City. Social Media and Big Data Services, North American Division of the Seventh-day Adventist Church. Retrieved January 31, 2017, from http://www.sdadata.org/blog/discovering-mission-oriented-patterns-with-open-data-in-new-york-city

Application of Data Science to Classify Causes of Maternal Mortality in Mexico

- In Mexico, the maternal mortality rate is very high: 38 deaths per 100,000 live births in 2015 (WHO).

- Eclampsia during labor and postpartum hemorrhage.

- The model generated with Naïve Bayes was chosen to carry out classifications within the software (accuracy = 0.72, precision = 0.75, recall = 0.74). The model was trained with 1,018 instances.

23. Domínguez, R. (2017). Aplicación de Ciencia de Datos para la Creación de Software Predictivo de Morbimortalidad Materna en México. Maestría en Ciencias Computacionales. Universidad de Montemorelos.

Analyzing the CHSI

Community Health Status Indicators (CHSI) open dataset provided by the Centers for Disease Control and Prevention (CDC) on Data.Gov. The CHSI dataset provides key health indicators for local communities. It contains over 200 measures for each of the 3,141 United States counties.

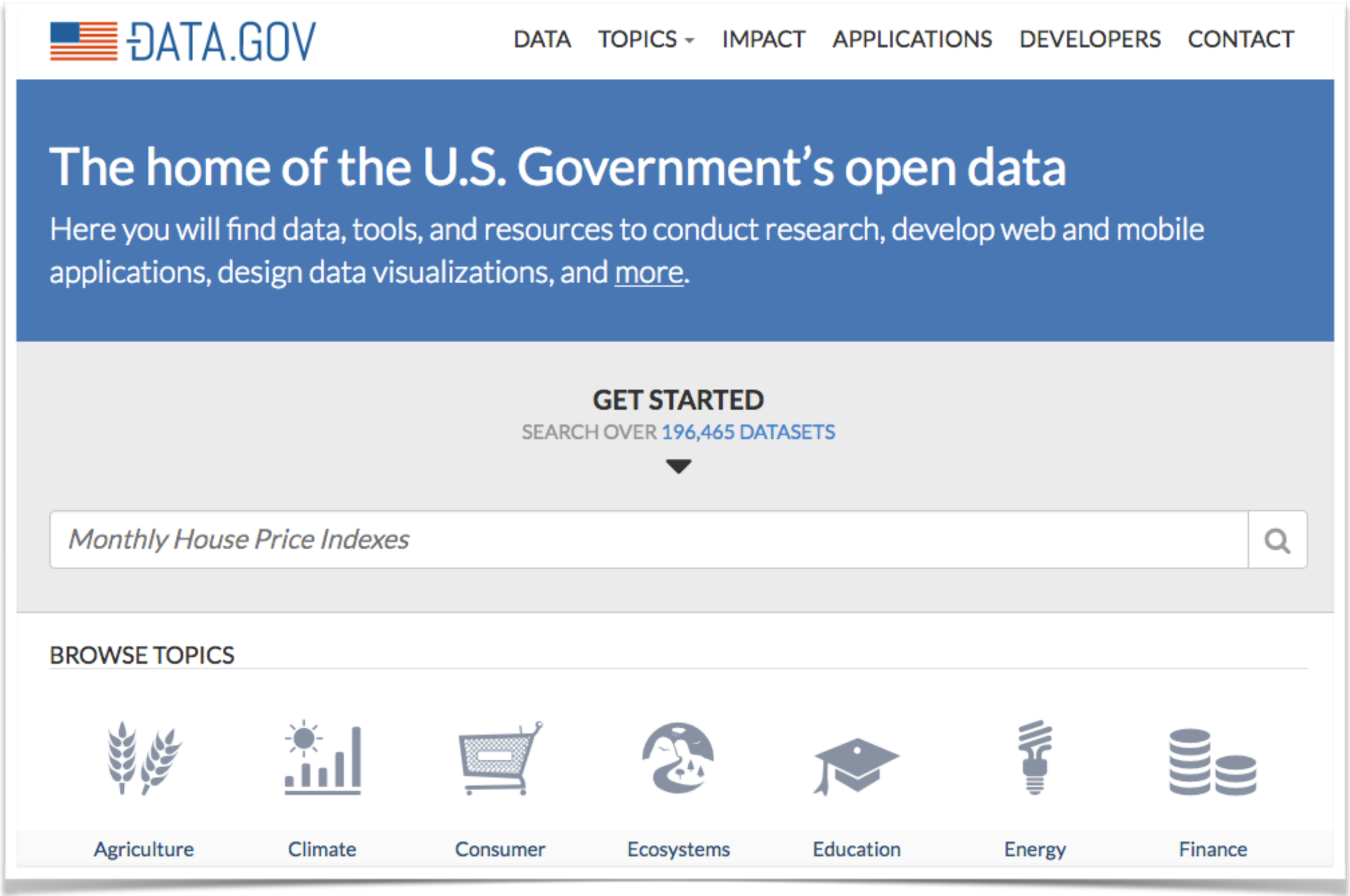
An increasing number of community/migrant health centers correlates with a decreasing number of people with diabetes per county.

24. Alférez, G.H. (2016). Discovering Hidden Patterns in US Health-Related Open Data with Machine Learning. Social Media and Big Data Services, North American Division of the Seventh-day Adventist Church. Retrieved November 14, 2016, from http://www.sdadata.org/blog/discovering-hidden-patterns-in-us-health-related-open-data-with-machine-learning

Agenda

1. What is big data, data science, and open data?
2. Successful case studies
3. **How to use data science at your conference**
4. Conclusions and recommendations

Use Open Data Sources



Use Open Data Sources

HealthData.gov

search

About Datasets Developers Agencies Feedback Topics

Home / Datasets / Food Affordability, 2006-2010

View published Add Feedback

State of California

License

Open Data Commons Open Database License (ODbL)

OPEN DATA

Food Affordability, 2006-2010

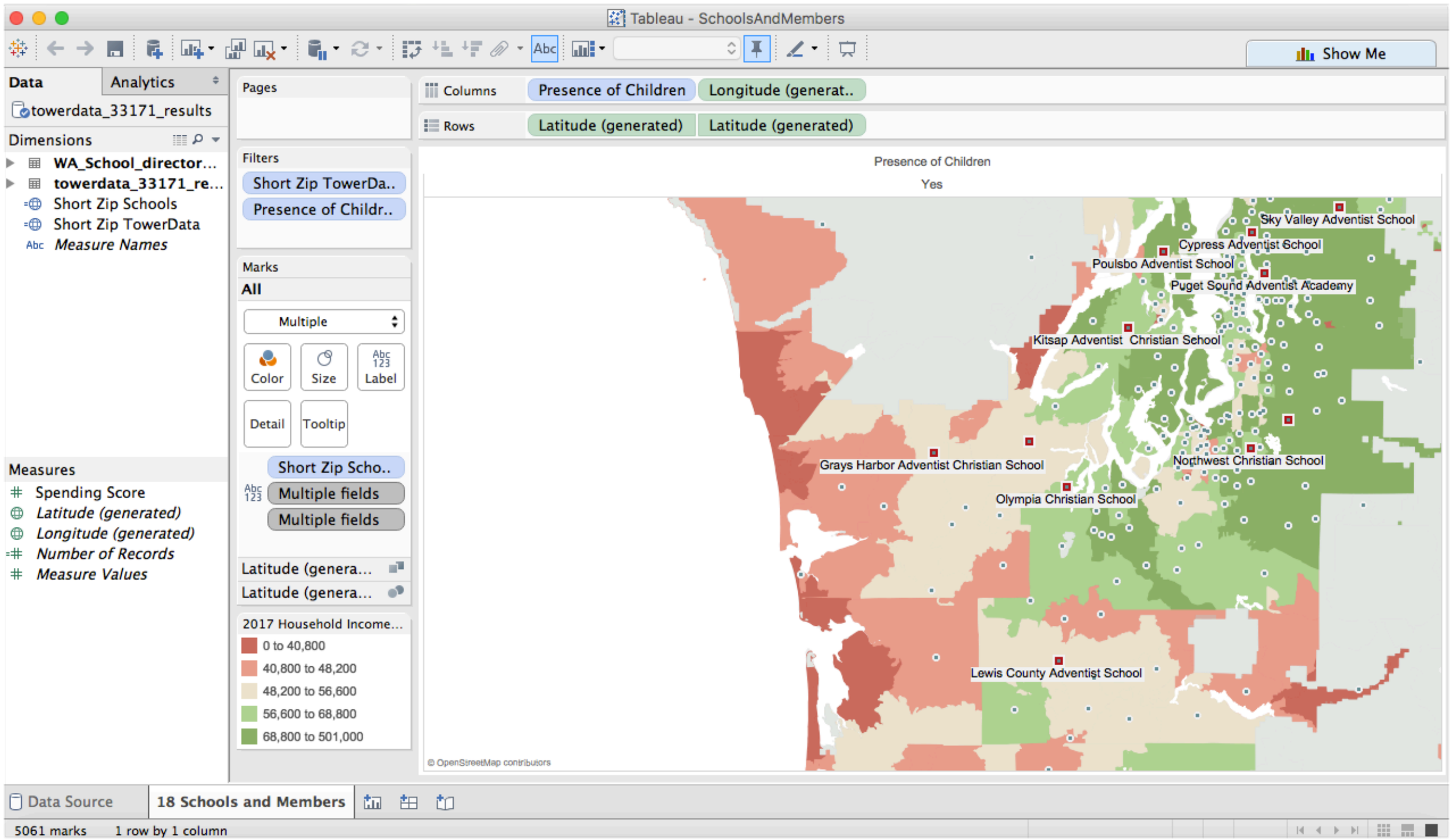
This table contains data on the average cost of a market basket of nutritious food items relative to income for female-headed households with children, for California, its regions, counties, and cities/towns. The ratio uses data from the U.S. Department of Agriculture and the U.S. Census Bureau. The table is part of a series of indicators in the Healthy Communities Data and Indicators Project of the Office of Health Equity.

An adequate, nutritious diet is a necessity at all stages of life. Inadequate diets can impair intellectual performance and have been linked to more frequent school absence and poorer educational achievement in children. Nutrition also plays a significant role in causing or preventing a number of illnesses, such as cardiovascular disease, some cancers, obesity, type 2 diabetes, and anemia.

At least two factors influence the affordability of food and the dietary choices of families â the cost of food and

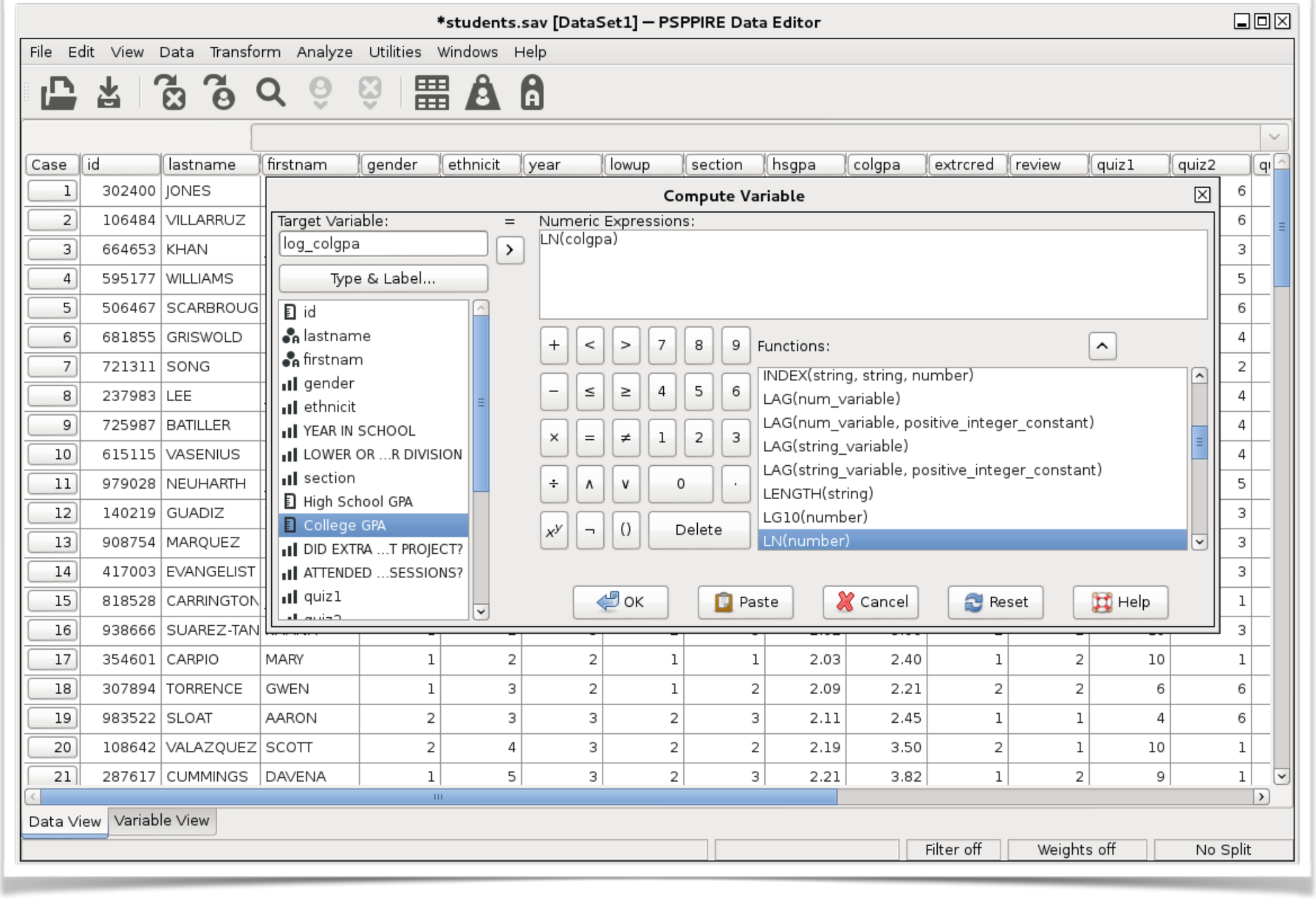
Use Simple and Powerful Software

Tableau



Use Free Software

GNU PSPP



Contact Us

SOCIAL MEDIA + BIG DATA SERVICES

Connecting members & mission through technology

Agenda

1. What is big data, data science, and open data?
2. Successful case studies
3. How to use data science at your conference
4. **Conclusions**

The Role of Big Data in Our Church

- Our **Church** can do something **valuable** with **big data**.
    - For instance, big data can help us to **make our beliefs relevant in a postmodern culture**.

- **Computational approaches** can be used to understand large pools of data, discover patterns, and make **“data-driven” decisions**.

Data science has the potential to help us understand the **needs of people in big cities** in an **unprecedented way**.

Areas of Interest

- **Evangelism:** Understand the community, etc.
- **Church Treasury**: Analysis of tithes and offering, donors profiles, offering distribution, etc.
- **Education:** Student profiles, retention forecasting, etc.
- **Media:** Create profiles of TV/radio/Internet viewers or listeners (i.e., personas), etc.
- **Geolocation Analysis:** Create maps related to schools and students, church and members, small groups, demographic information, etc.

Data Science: How to Turn Data Into Actionable Mission-Oriented Decisions

Harvey Alférez, Ph.D.
Global Software Lab,
School of Engineering and Technology,
Universidad de Montemorelos, Mexico
www.harveyalferez.com