

# Discovering People's Needs with Data Science


---

**Harvey Alférez, Ph.D.**

Director, Institute of Data Science  
School of Engineering and Technology  
Montemorelos University, Mexico

<https://icd.um.edu.mx/en/>





“the Savior mingled with men as one who desired their good. He showed His sympathy for them, ***ministered to their needs***, and won their confidence. Then He bade them, ‘Follow Me.’” [1]

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


**Data Science**

analyzed with

**Big Data**

**Discover the  
Needs of  
People**

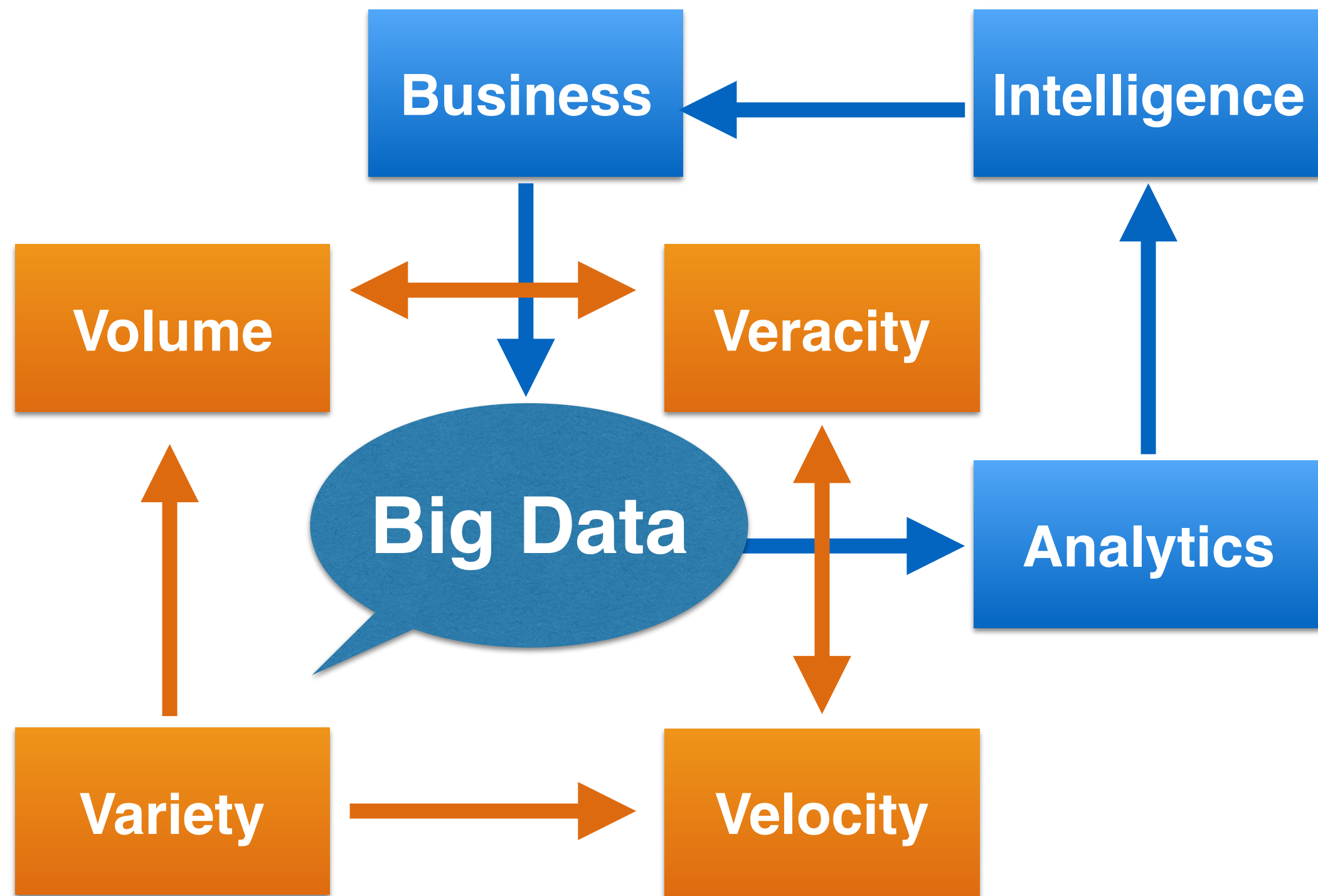


- 
- **Big data** is a term that can be used to describe data sets so **large** and **complex** that they become difficult to work with using standard techniques [1].
  - **Big data is the next big thing. The new oil** [2].

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/>

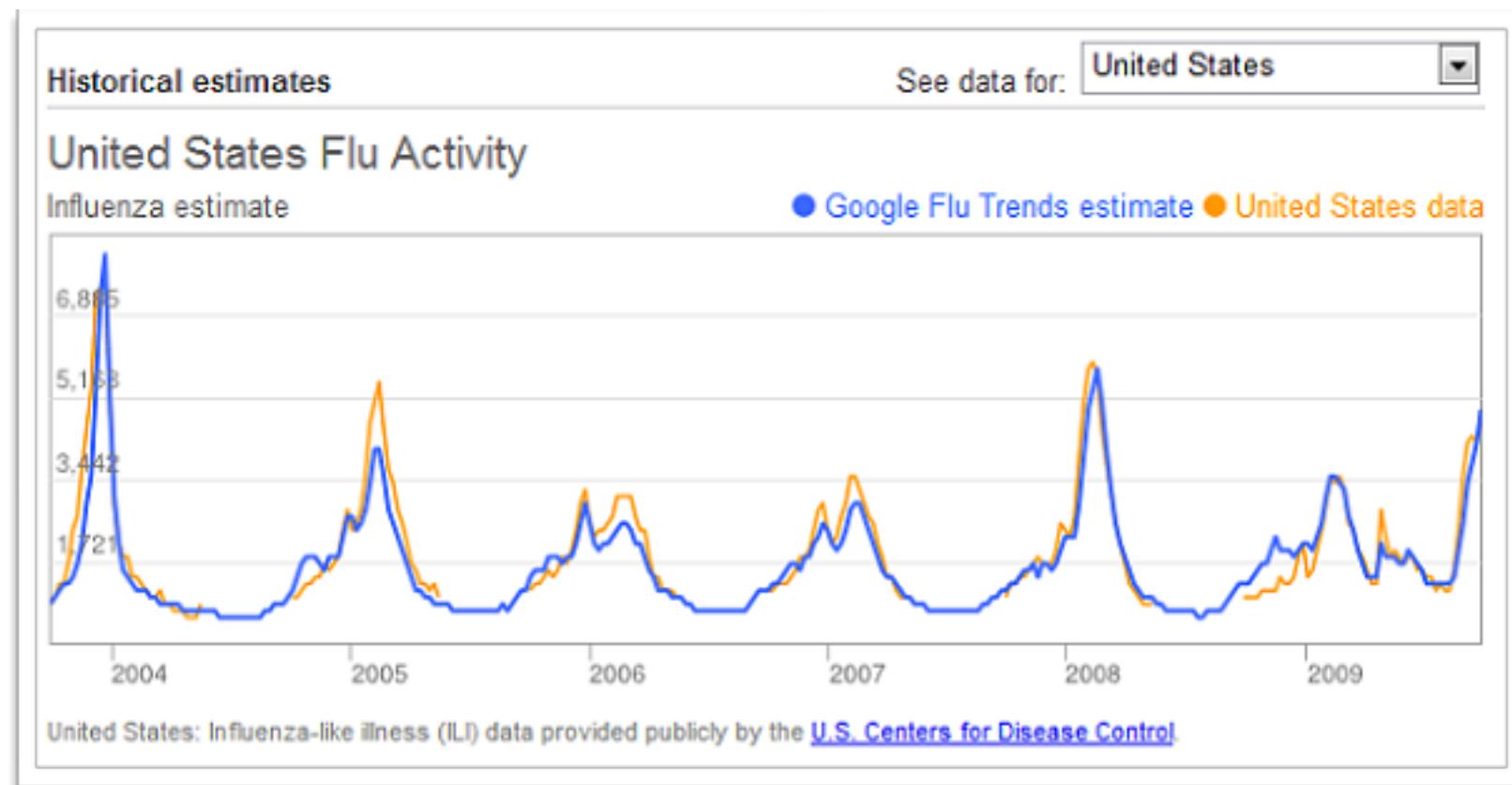




3. Maheshwari, A. (2016). Big data essentials. Amazon Digital Services LLC.

# 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)

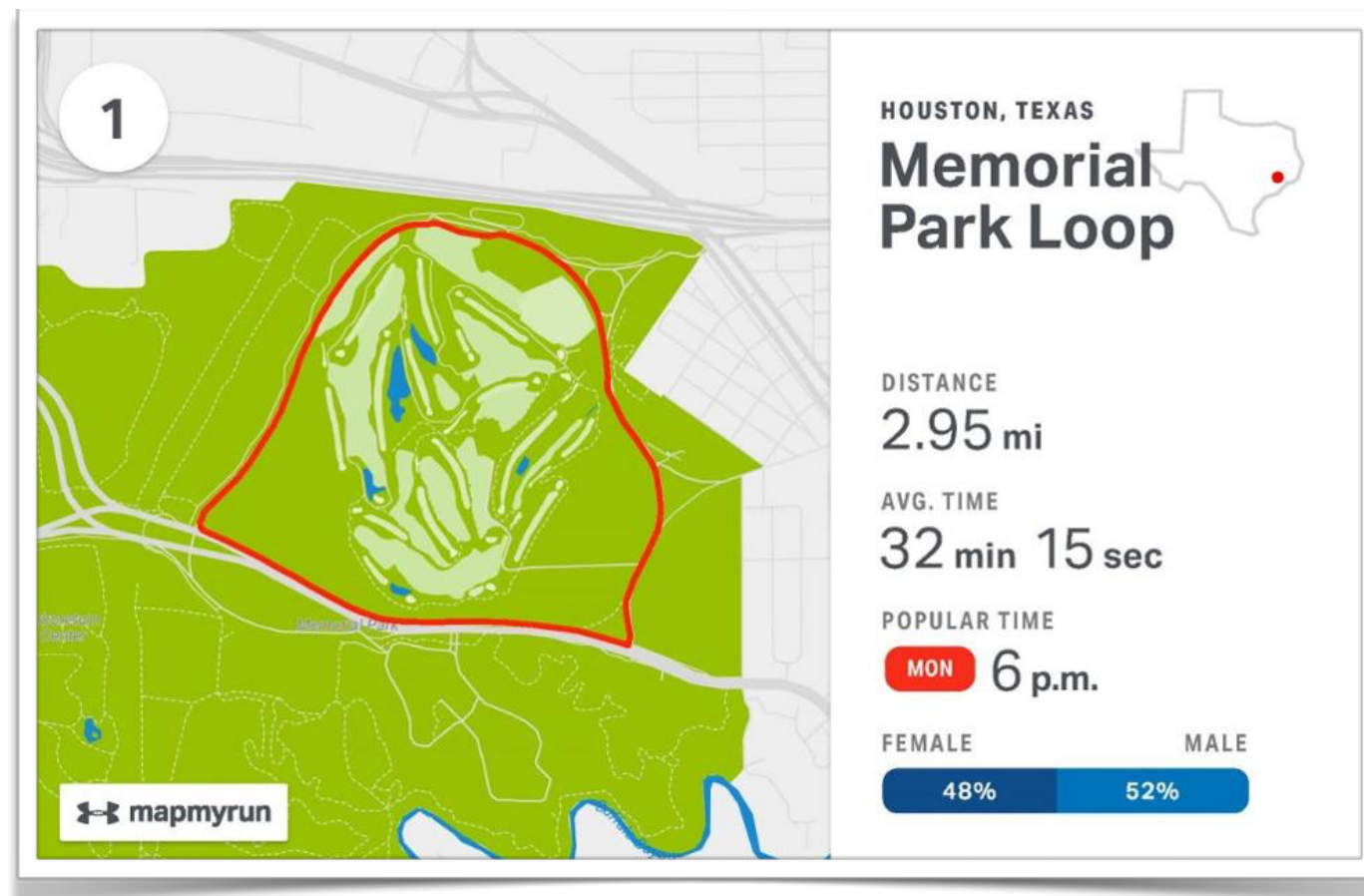
# Big Data Applications

- **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** can be defined as the study of the generalizable extraction of knowledge from data [4].

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





ARTWORK: TAMAR COHEN, ANDREW J BUBOLTZ, 2011, SILK SCREEN ON A PAGE FROM A HIGH SCHOOL YEARBOOK, 8.5" X 12"

DATA

# Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

WHAT TO READ NEXT



**Big Data: The Management Revolution**

5. Davenport, T.H. y Patil D.J. (2012). Data scientist: the sexiest job of the 21st century. URL: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>



# Data Analysis Results for Washington Conference

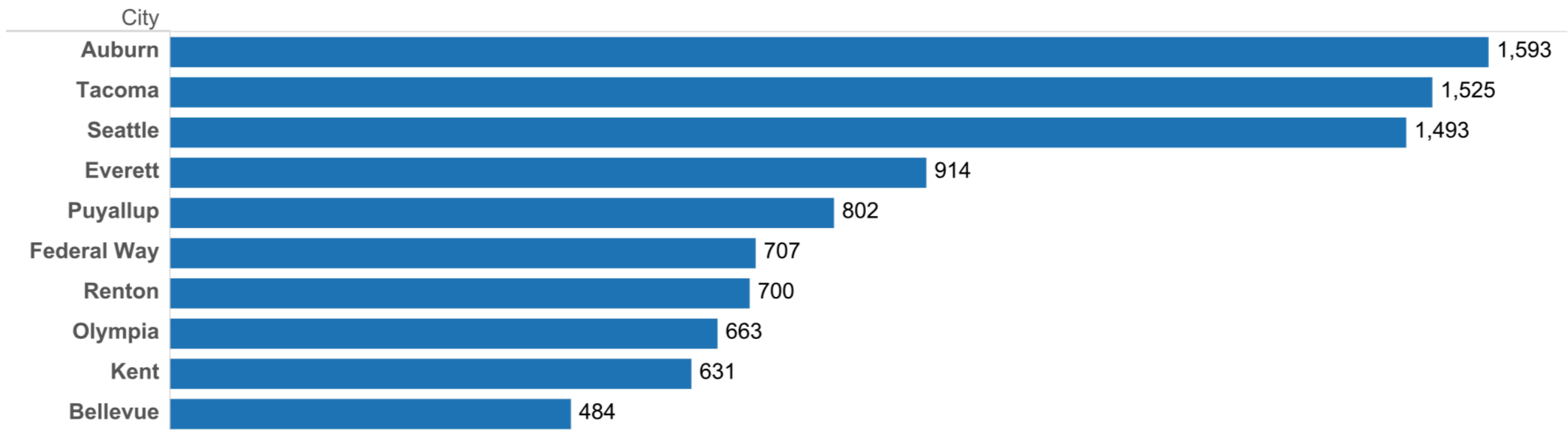
- Data related to church members, churches, schools, and social media activity in Washington Conference - **close data**.
- June 5-30, 2017.



# Datasets

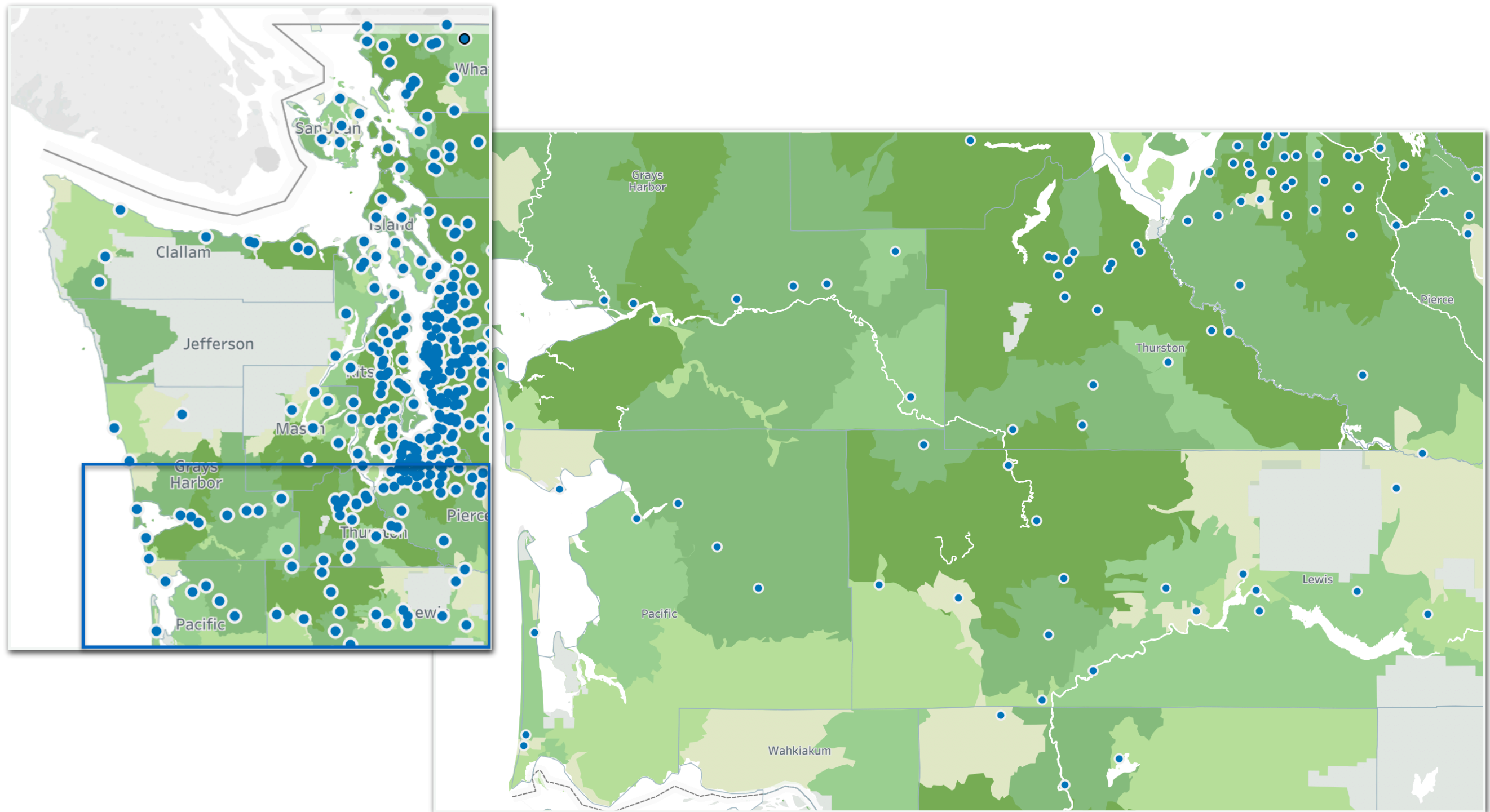
- TowerData: 31,013 records
- Directory (geolocation) of Seventh-day Adventist churches and schools in Washington Conference
- Data related to the activity in Washington Conference's Facebook page

# Top Ten Cities with the Highest Concentration of Church Members

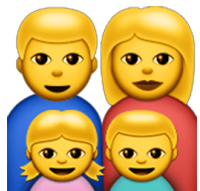
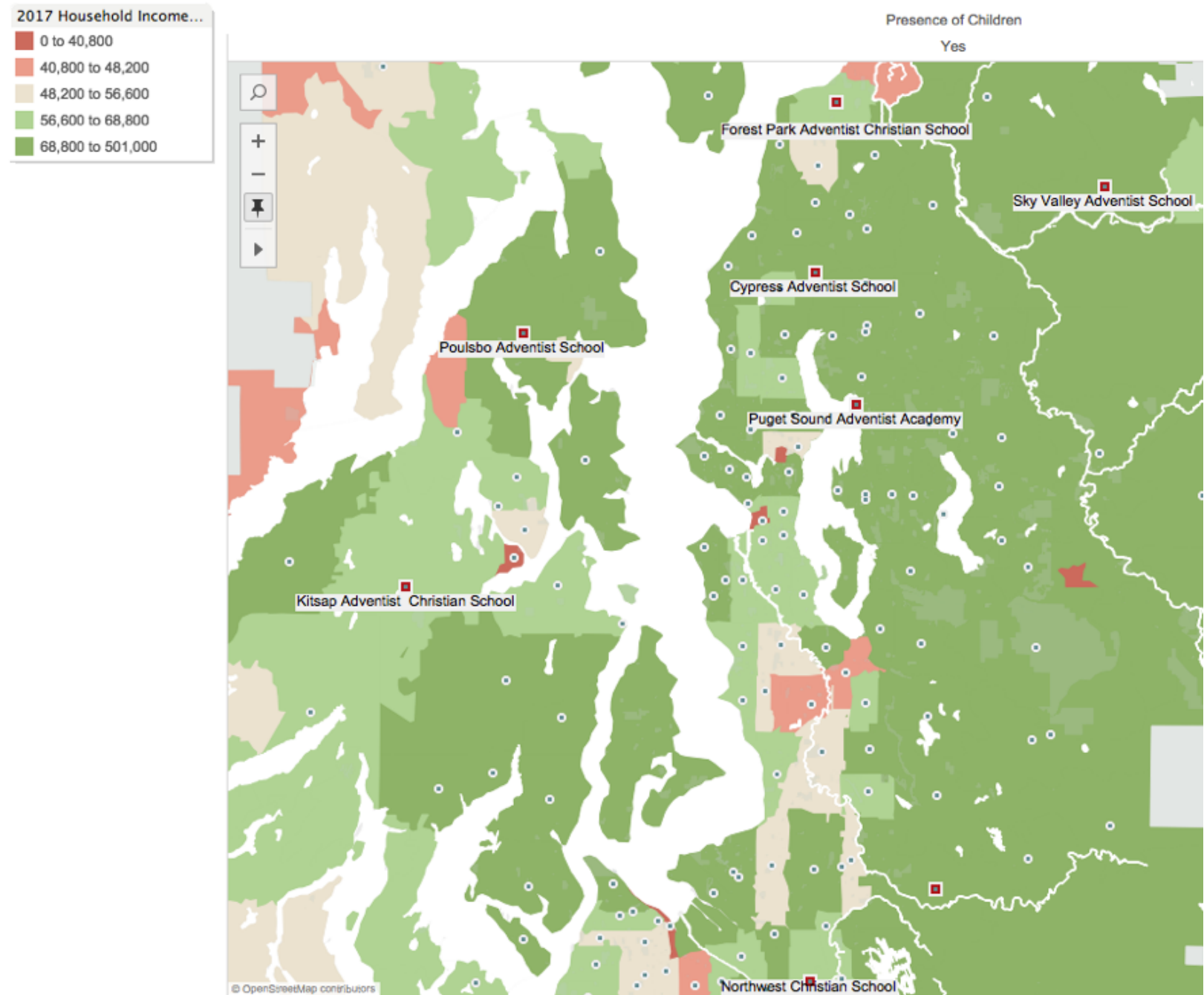




# Communities with No SDA Members



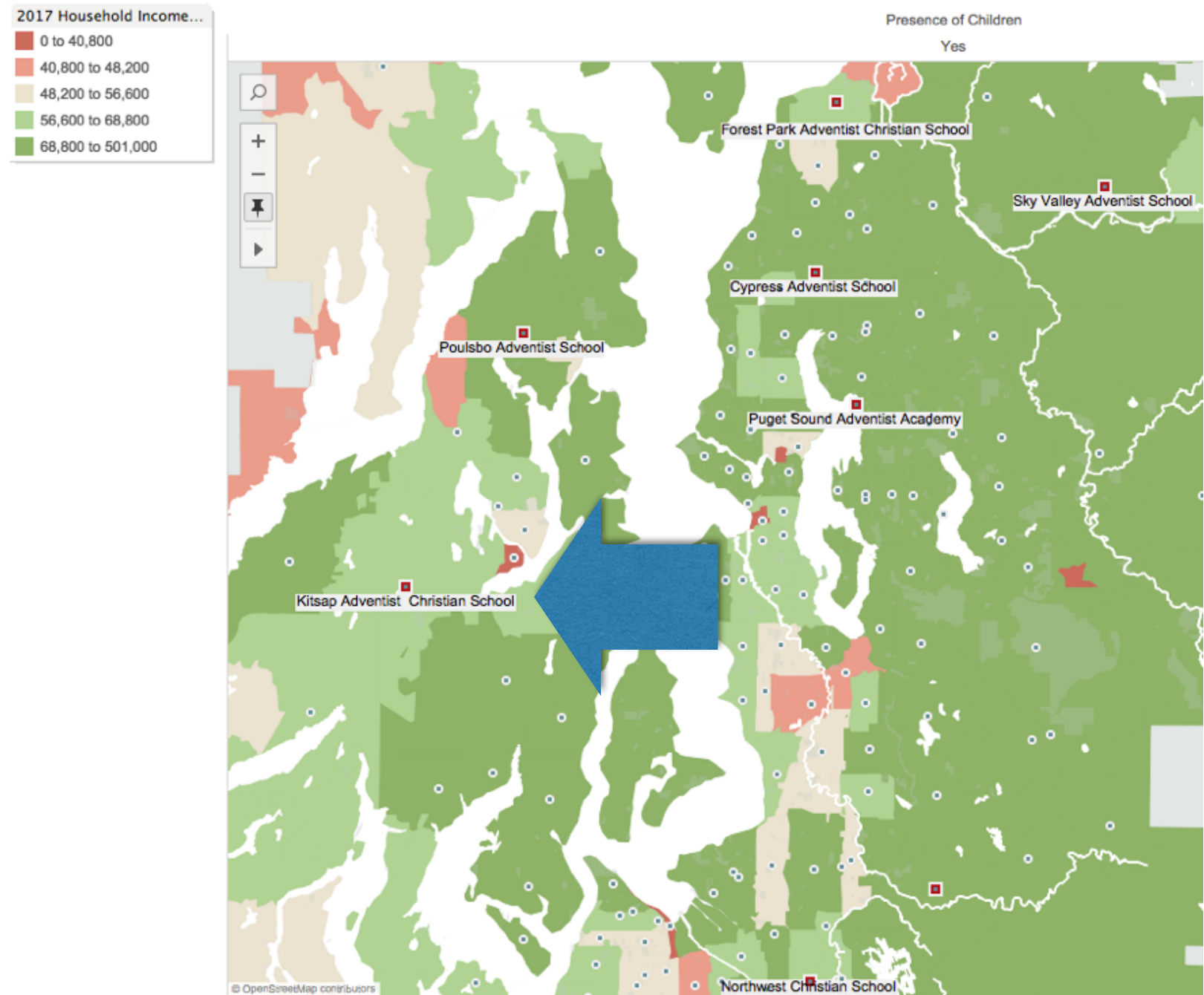
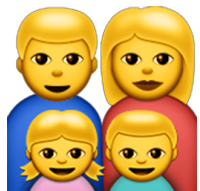
# Members with Children Present at Home and Adventist Schools



# Members with Children Present at Home and Adventist Schools

**Kitsap Adventist Christian School** is located in an area with the following household income: **\$56,600 to \$68,800**

- 18 members in an area of \$0 to \$48,200
- 18 members in an area of \$48,200 to \$56,600
- **109 members in an area of \$56,600 to \$68,800**
- 36 members in an area of \$68,800 to \$501,000

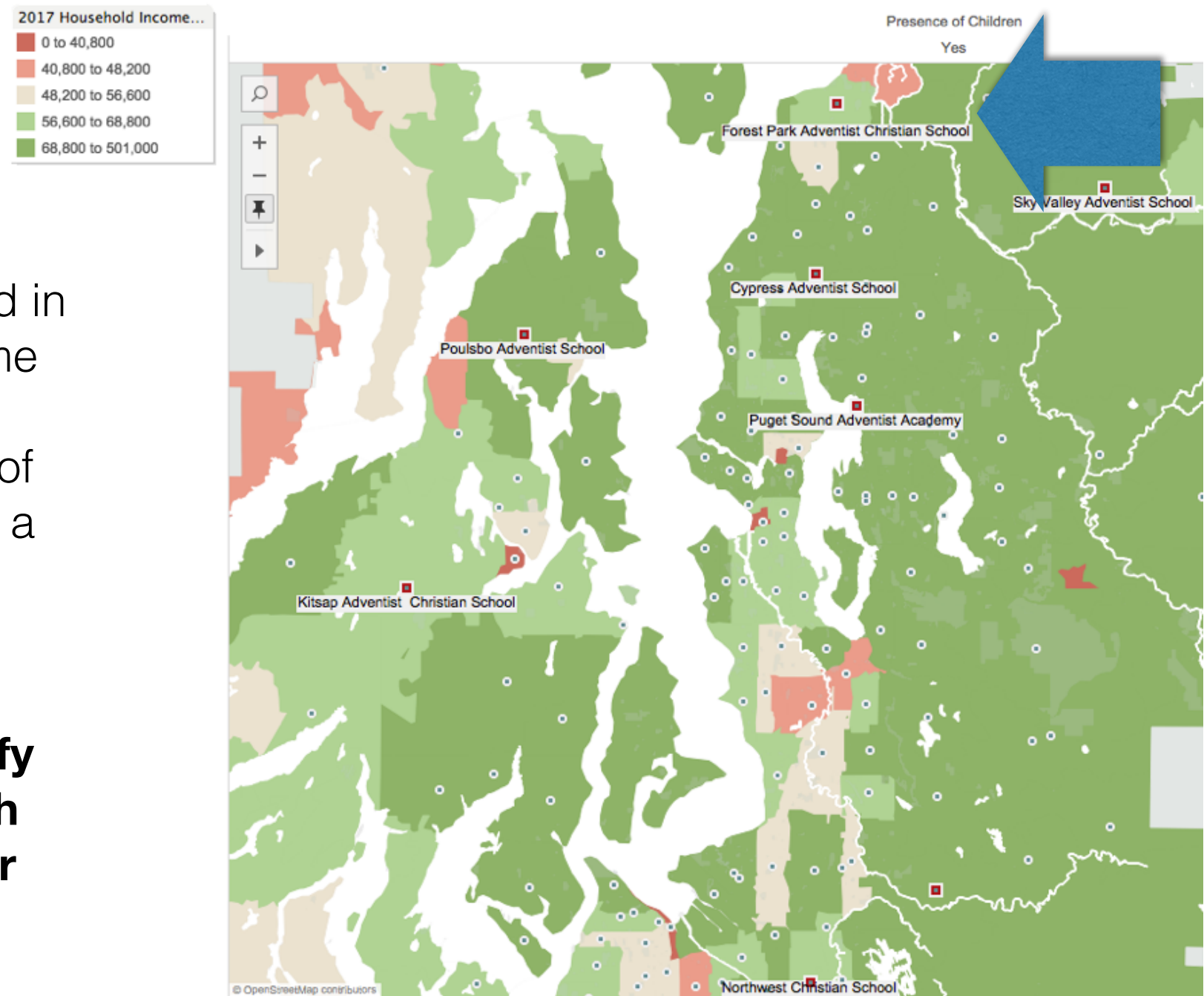
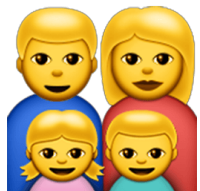




# Members with Children Present at Home and Adventist Schools

**Forest Park Adventist Christian School** is located in an area of household income values **lower than** the household income of most of the church members within a radius of 10 miles from this school

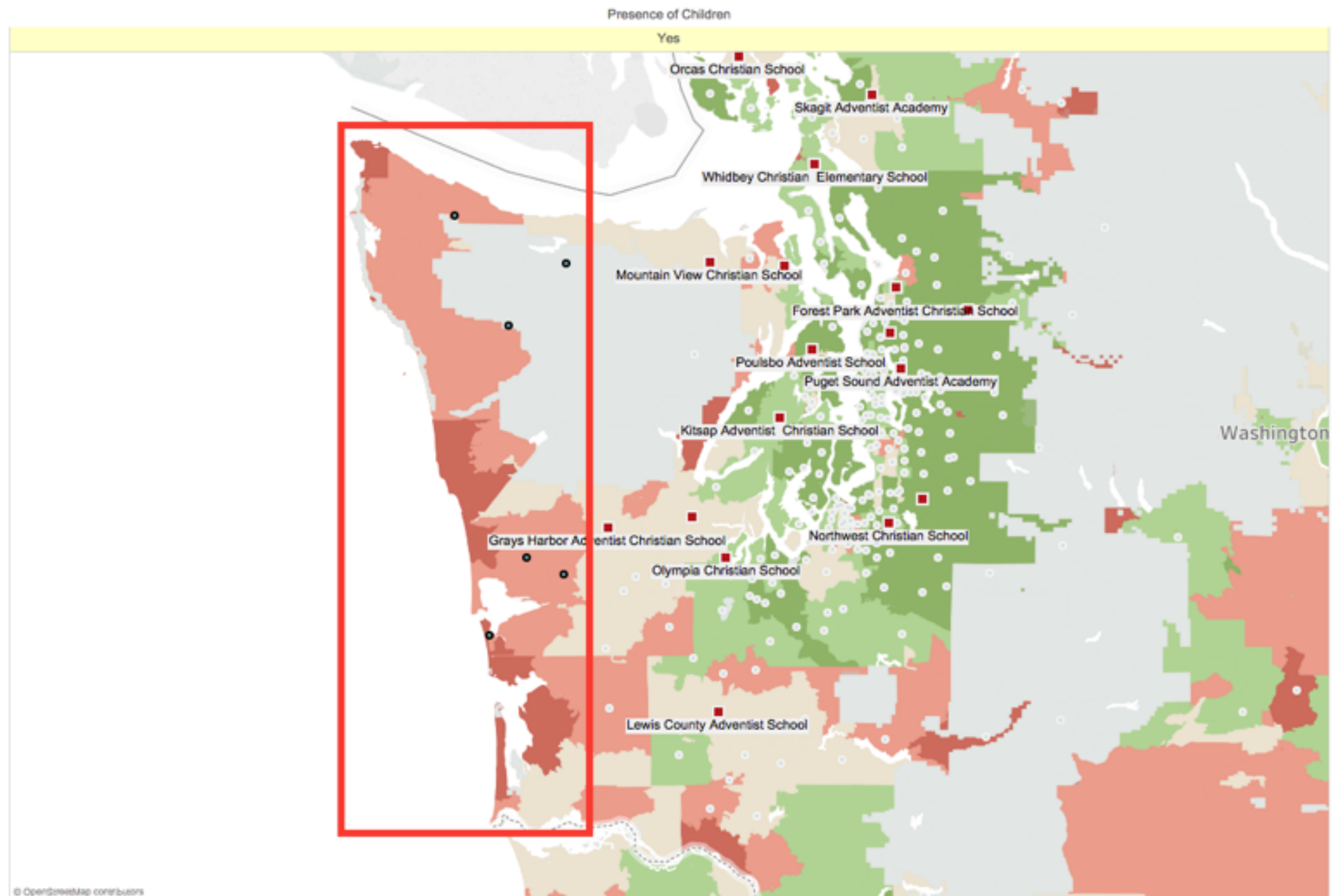
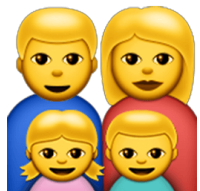
**Does this school satisfy the demands of church members with a higher household income?**



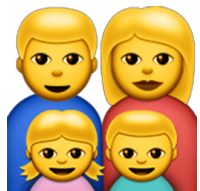
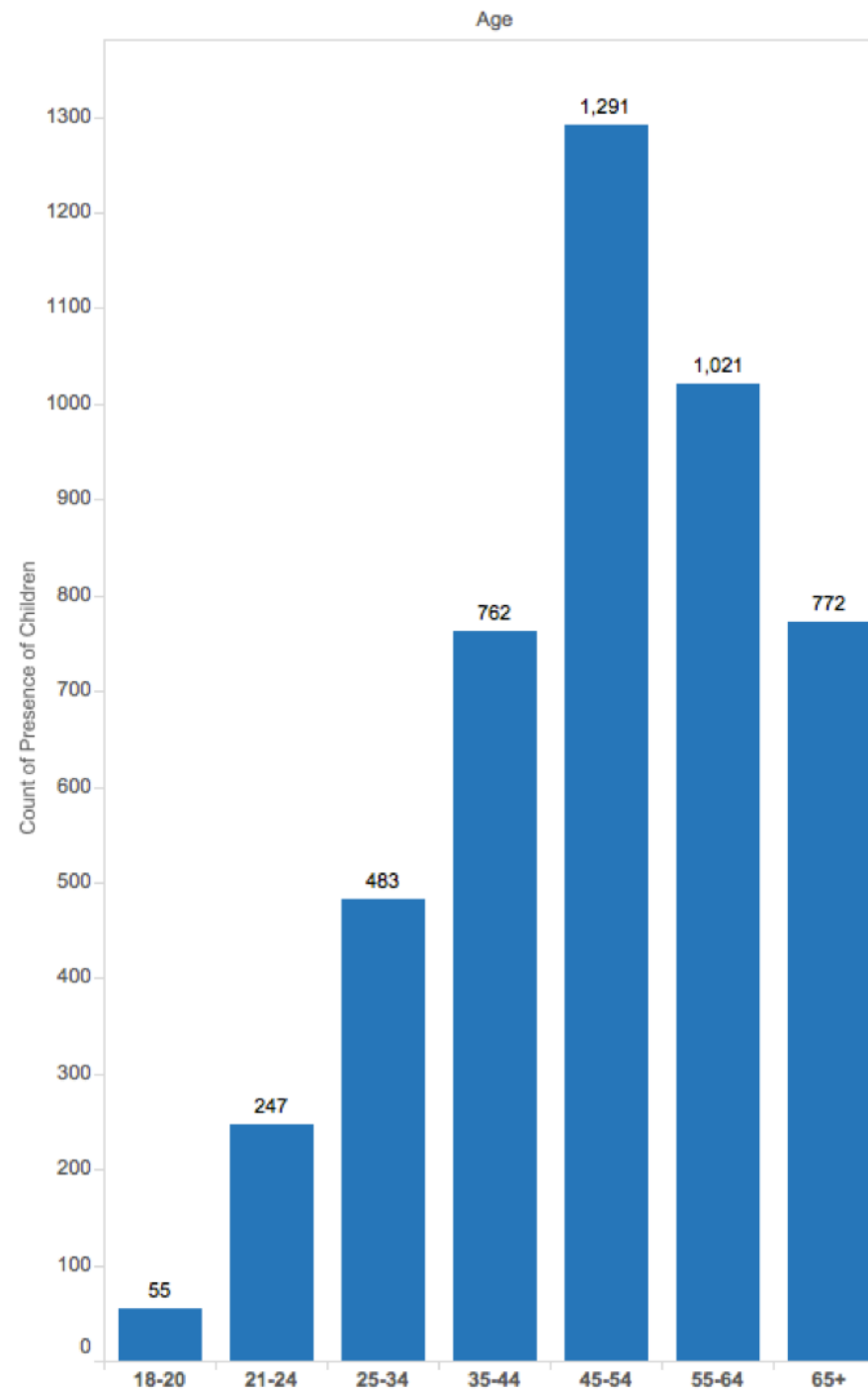
# Members with Children Present at Home and Adventist Schools



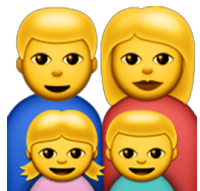
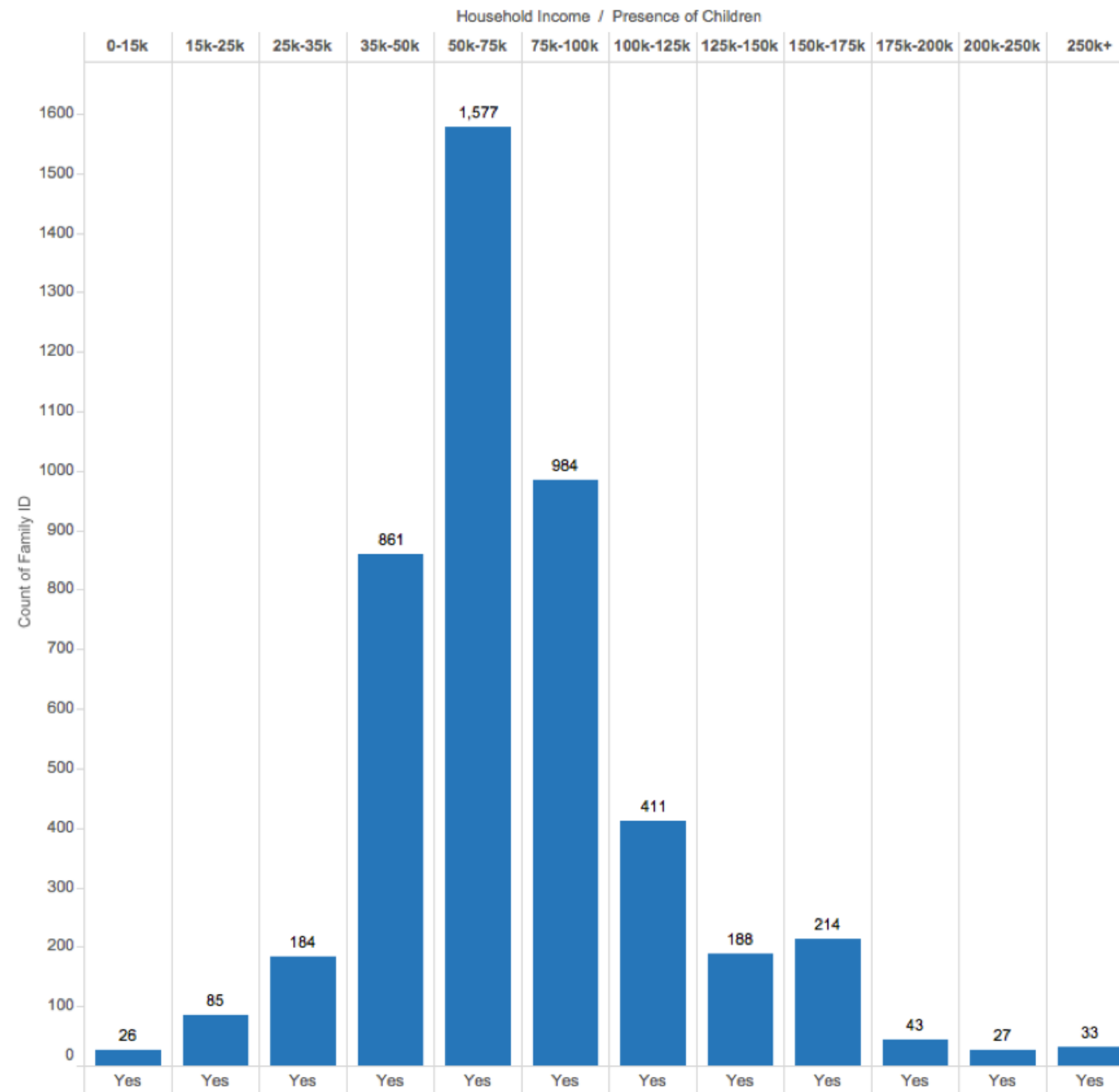
**108** church members with children at home at the west **without any Adventist School** in an area of approximately 4,800 square miles



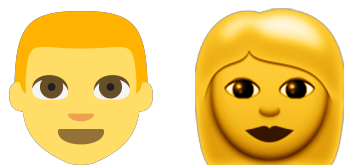
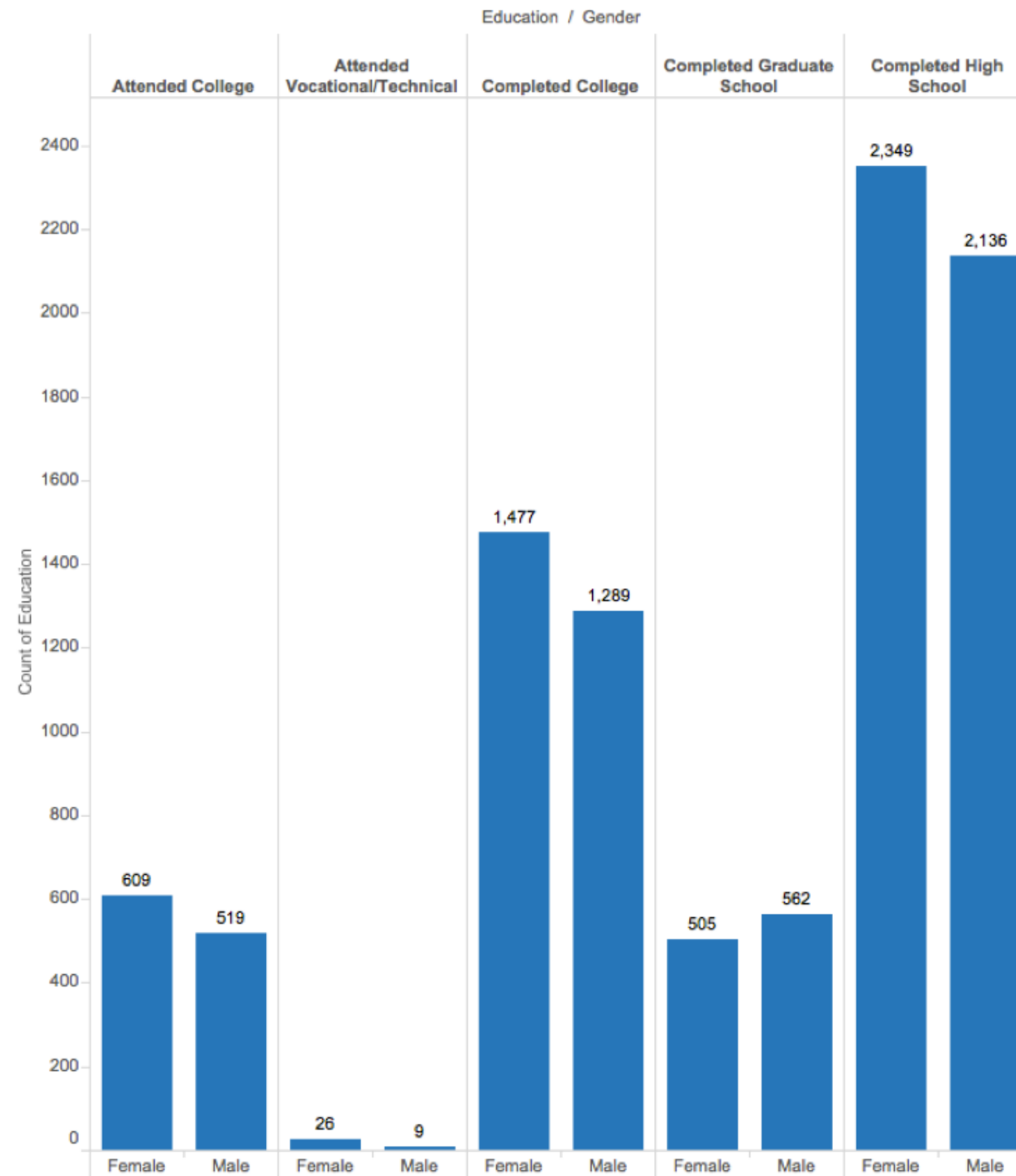
# Parents' Age-Range



# General Income of Families

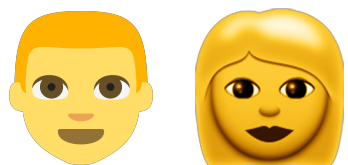
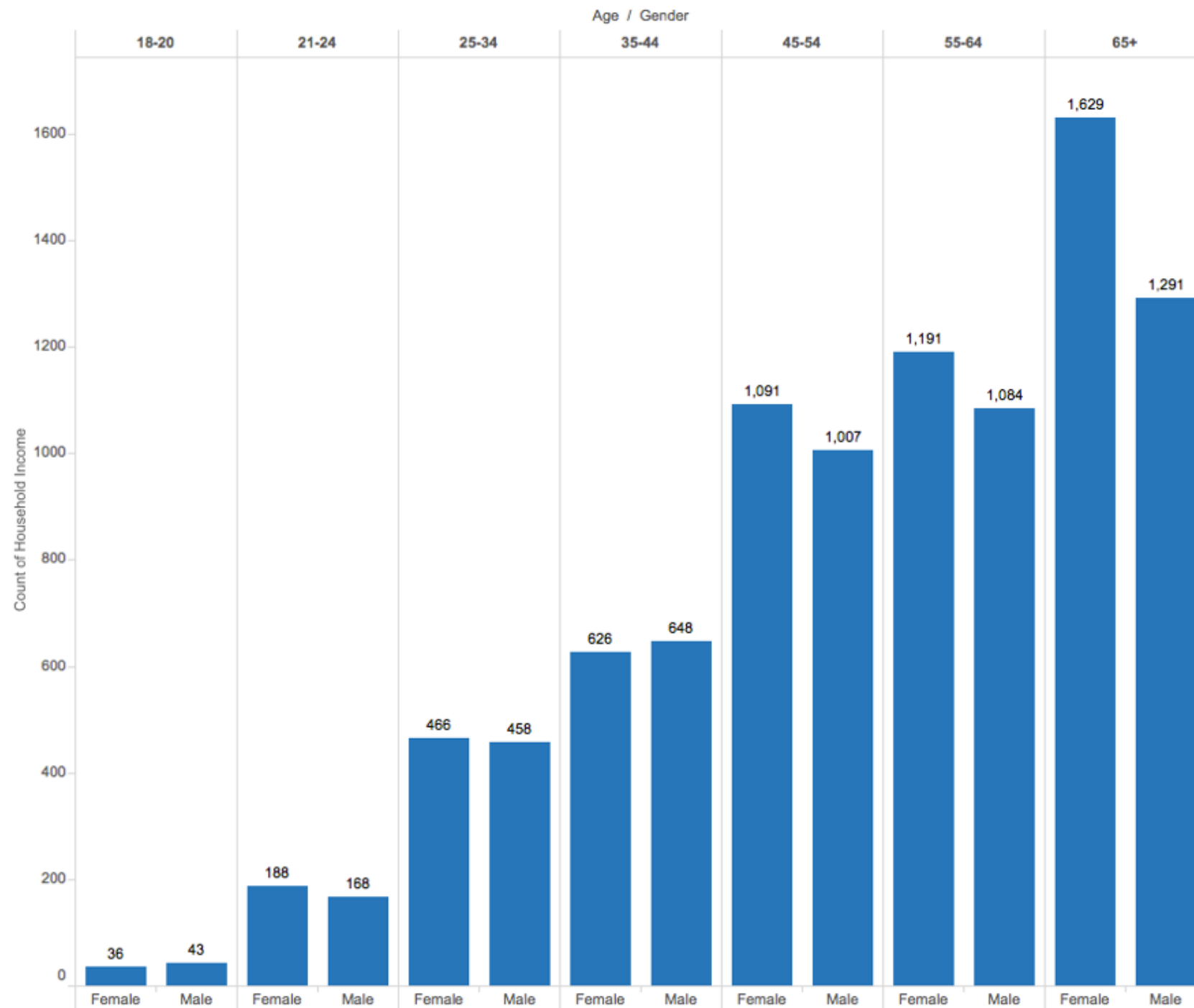


# Socio-Economic Status

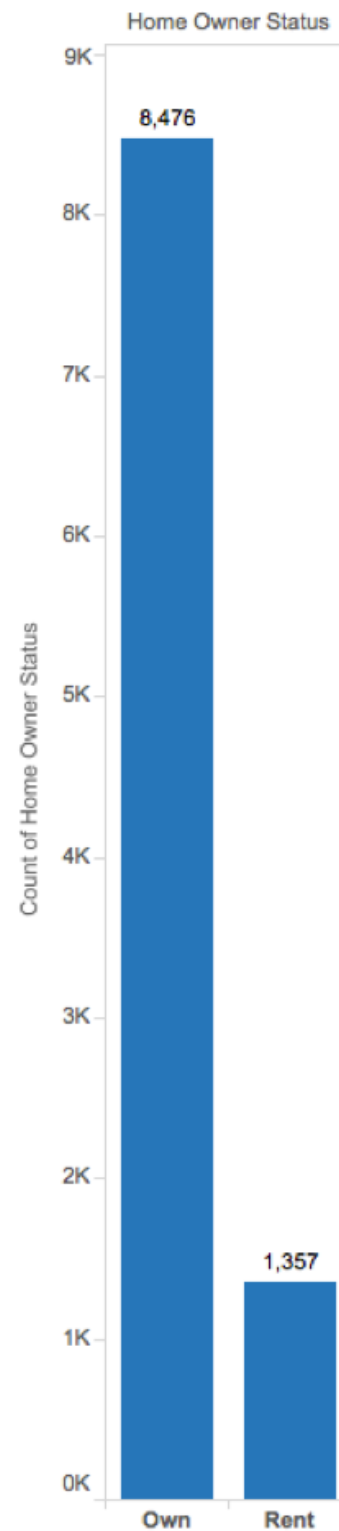




# Socio-Economic Status



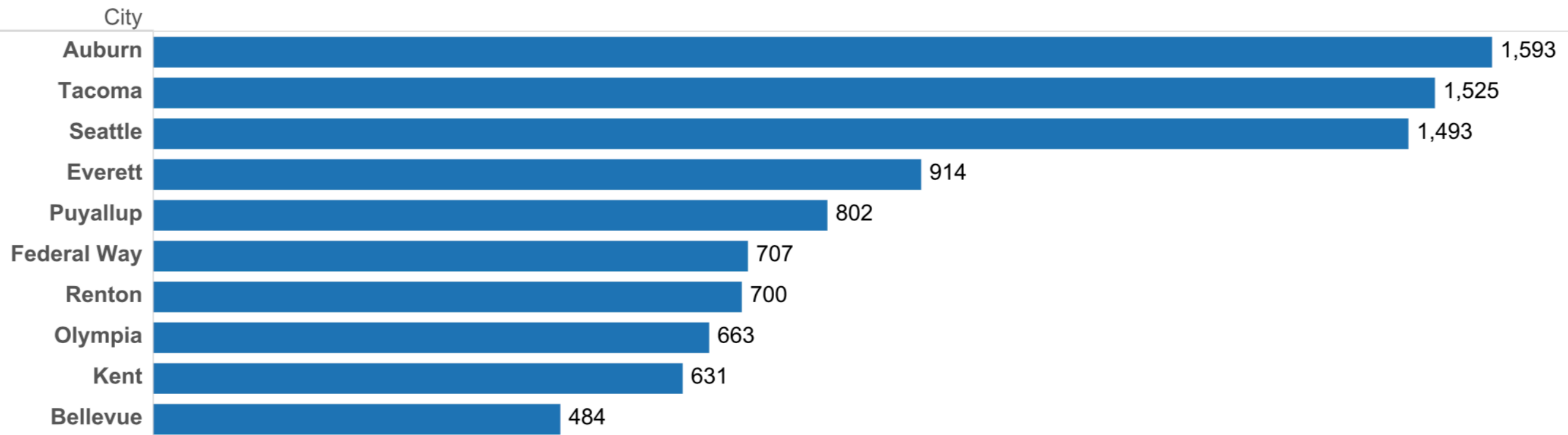
# Socio-Economic Status



# Socio-Economic Status

- Church members who have completed high school have a **similar home market value** and **household income value** than church members who have not finished their college degree.
- The **spending score increases** according to the **level of education**.
- Having terminated graduate school ***does not*** imply a higher income or a higher spending among church members.

# Around 17% of Members are Blue-Collar Workers

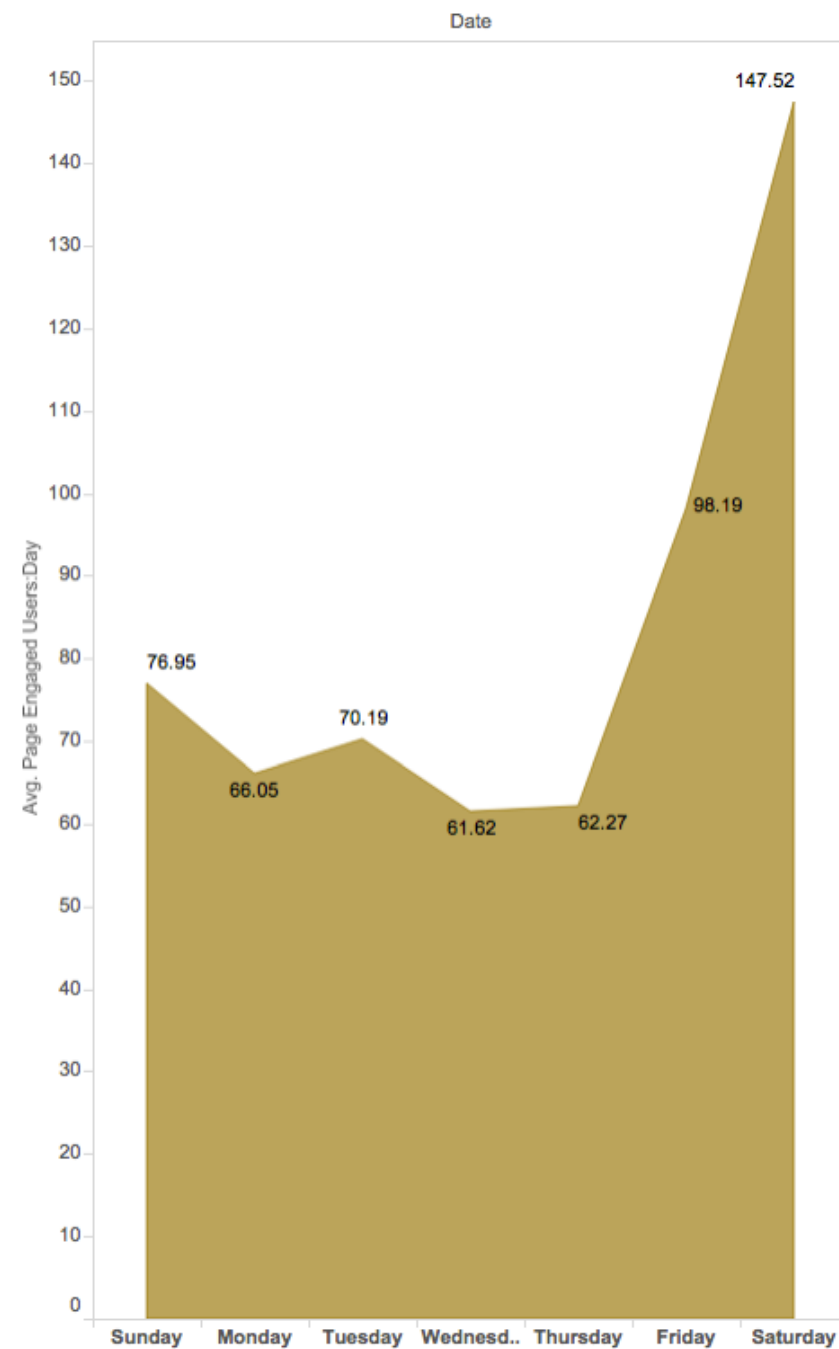


# Key Members' Interests Based on Age

- **Example: 65+ members** have the following **interests** (which can be satisfied by the church):
  - **news & current events** (with news from Adventist News, Adventist Review or other publications)
  - **kitchen & dining, garden, health & wellness, and cooking** (with online courses on these areas)
  - **traveling** (with the organization of mission trips)
  - **books and magazines** (this is a key fact for ABC stores)
  - **charitable donations** (this is an interesting fact for fund raising campaigns)



# Social Media Analysis



Facebook Page Activity of Washington Conference

# Other 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:** Demographic information, etc.

# Big Data for Reaching a Big World



7. 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

- 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)

# Creation



Creationism vs. Theory of Evolution (Case Insensitive Search)



# Healthy Living



Increasing Interest in Healthy Living and Vegetarianism

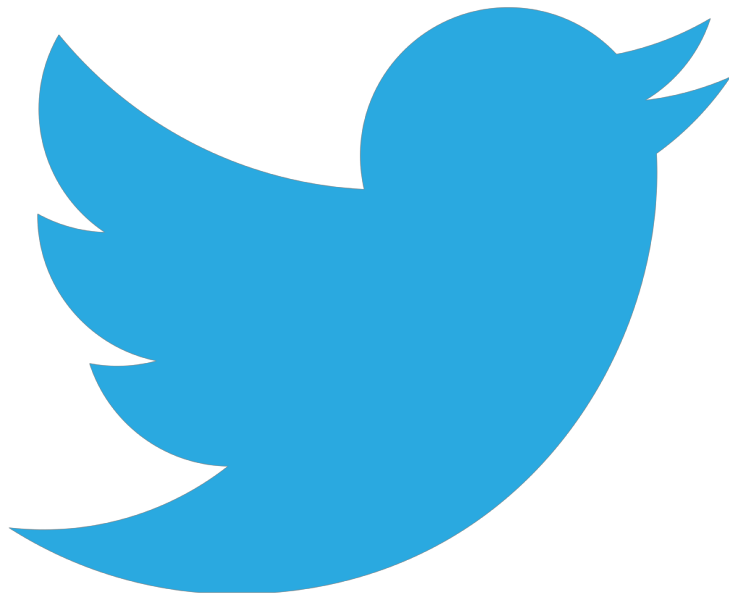
# Understanding the Needs of People in Big Cities through Data Science



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

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

# 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.

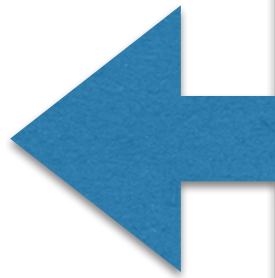
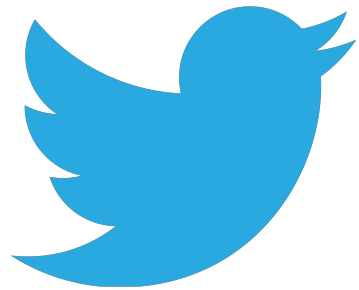
# Listening Closely to the Birds

## **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



```
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> -r'
        sys.exit()

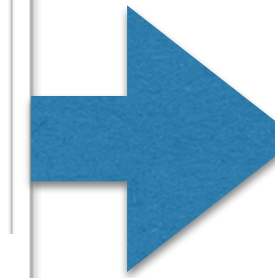
    for opt, arg in opts:
        if opt == '-h':
            print os.path.basename(__file__) + ' -e <expression> -c <city> -r'
            sys.exit()
        elif opt in ("-e", "--expr"):
            EXPRESSION = arg
        elif opt in ("-c", "--city"):
            user_city = arg
        elif opt in ("-r", "--radio"):
            radio = arg

    tweets = collect_tweets(user_city, EXPRESSION, radio)

    return negative_tweets, positive_tweets
```

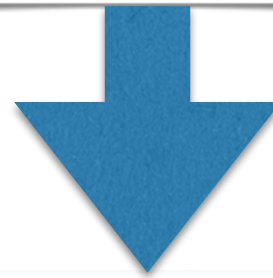


# 2. Classifies



+ and -

# 3. Stores



	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/H23IGJNw0	*73.99278645	*40.75052792												
2	Positive	*65	haters and family near and far and my Jetblue family@too lets...	https://t.co/IDsqLyobR	*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. #alichelaurissa #birthdaygirl...	https://t.co/1GL5dBqC	*73.74631052	*40.70295328										
4	Positive	*650915559132000256	*2472332402	*cityscapesny	*2015-10-05 06:08:57	*TOMORROW NIGHT! #ADAKISS @thereakiss #THOTLINEBLING HNY Family invades #sooperstarmondays with The...	https://t.co/nuQilyLen4g	*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/zqCXfCvFn	*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	*WJBLAI	*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/Sif4EGJG45	*73.84604275	*40.75685645										
9	Positive	*650826880317128704	*91659013	*fritzedotnyc	*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	*377775125	*_wavydayv	*2015-10-04 23:43:45	*Happy Birthday Cineya #Family @ Sweet Chick https://t.co/1qOD2gFqC	*73.9573975	*40.718399											
12	Positive	*65079582645780336	*1571019223	*IntoTheRowBlog	*2015-10-04 22:13:10	*Happy Birthday to october_ bab! Enjoying another round of drinks just for you #family #instacool...	https://t.co/4yflzAKF9B	*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/RmoPividW8	*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	*travelingheidl	*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/s1swKyY6h	*73.98872539	*40.73639096										
19	Positive	*65	family & finally signing a lease!	@ MP Taverna Astoria https://t.co/AHn3JKwdfH	*73.90966687	*40.77549002													
20	Positive	*650699463791505409	*255800296	*drjordanmetz	*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/X5dKfUvan	*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/I2qOck5lt	*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/zLIIGnPAKL	*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/UboNhtNnL	*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/QZnmVChIME	*73.96116732	*40.71674563											
29	Positive	*650503477429411840	*2867179032	*FLIP_NYC	*2015-10-04 02:51:29	*Family affair happy 50th Simon and Bsnk's happy 200th old man ... https://t.co/4cT29pQFxFU	*74.0064468	*40.7406578											
30	Positive	*65	NY w/ @justinbreton @allicrozier	https://t.co/NMFw9ltbH http://t.co/ScDgho3y3u	*73.9807189	*40.7301443													
31	Positive	*650493057322852352	*428652528	*James4be	*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	*40.8											
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	*@DeiliahNoB5zone 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/BuMyd95f1	*73.96762272	*40.762187											
36	Positive	*650458069240033280	*104321174	*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/LVKPz5Q05y	*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										

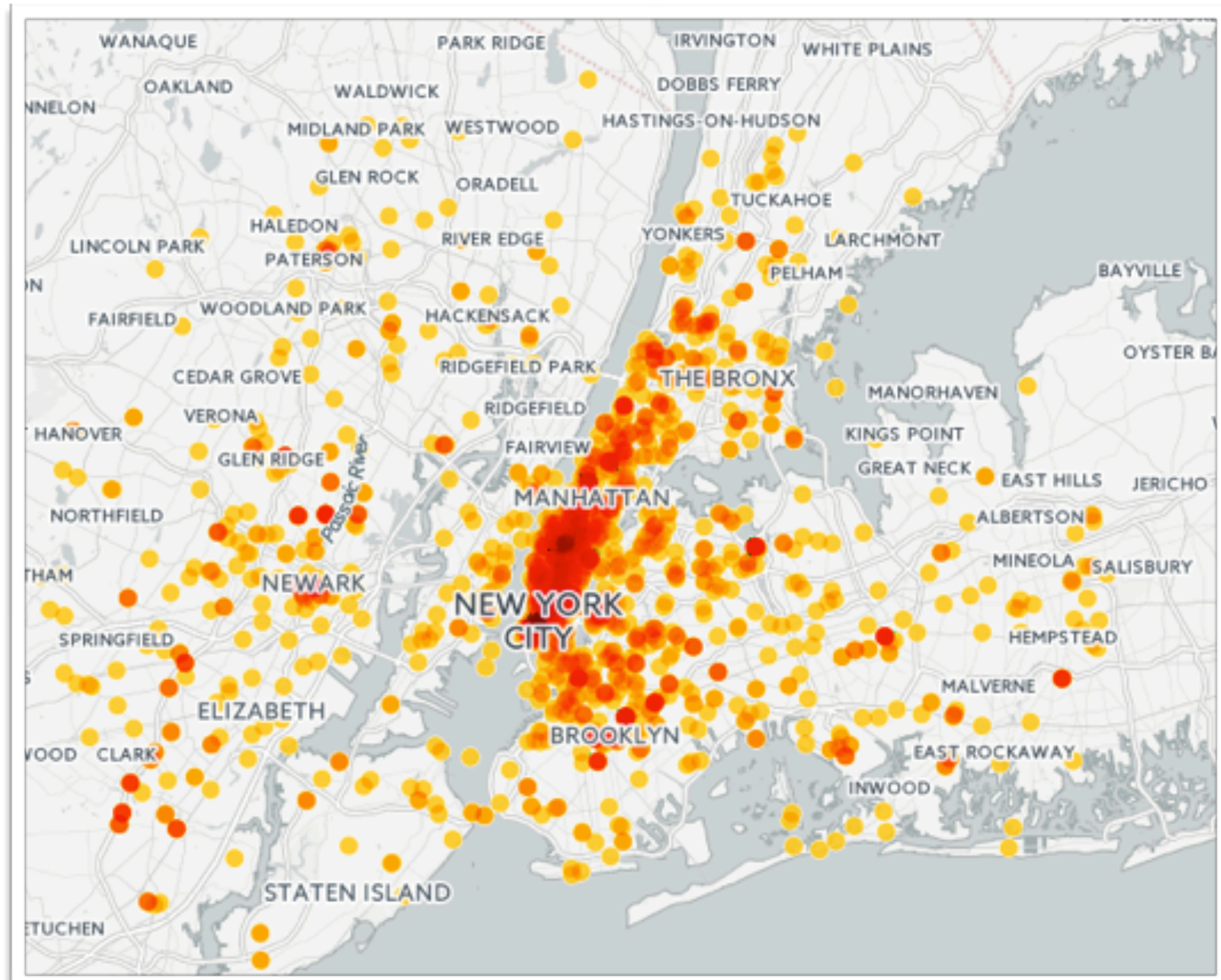
# Positive Tweet about Vegetarian Food

- Positive
- her\*
- 2015/10/02 02:08:16
- I want to be vegetarian. I really do. @arrogantwine @ 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)

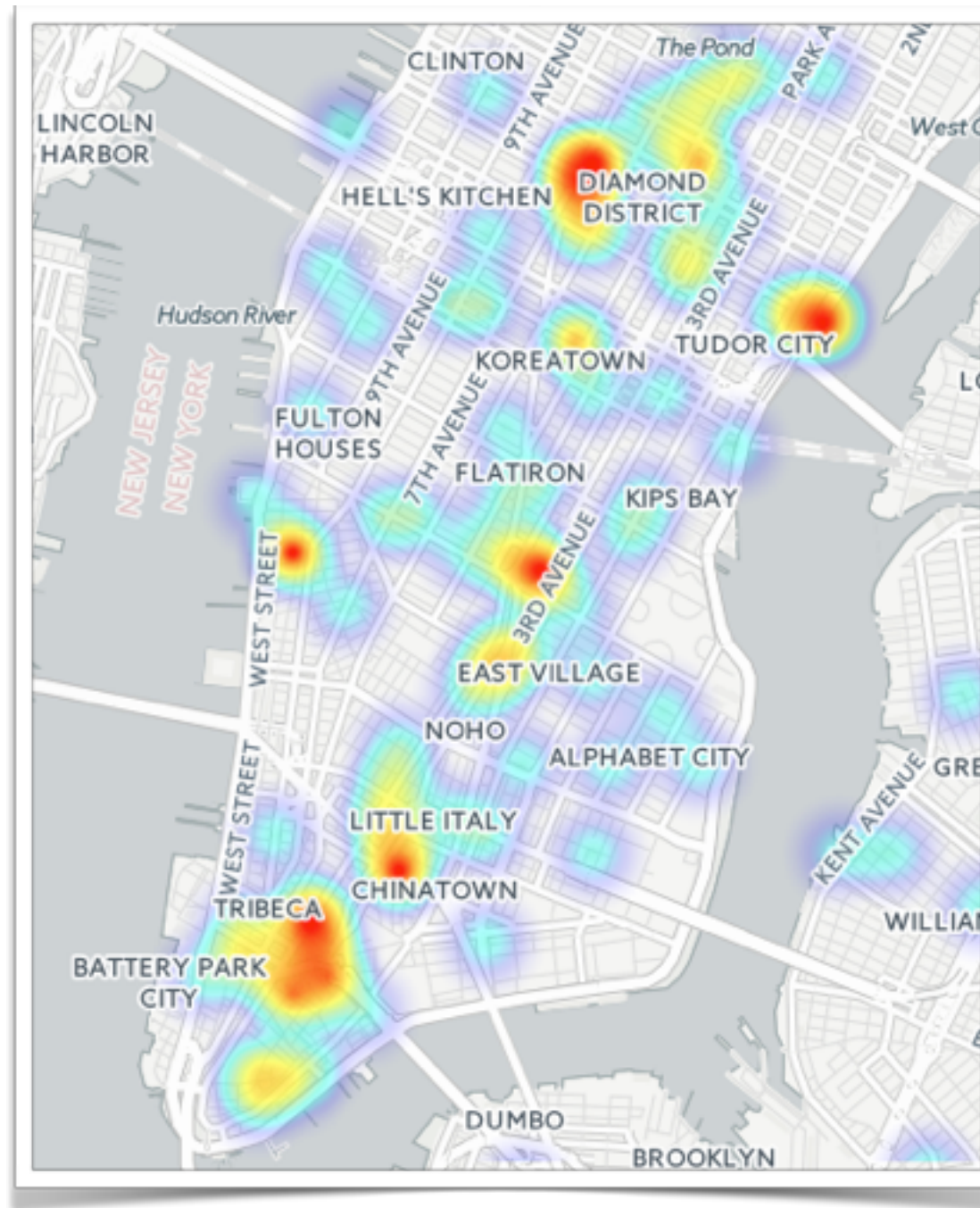
# Listening Closely to the Birds



Intensity of tweets in New York City



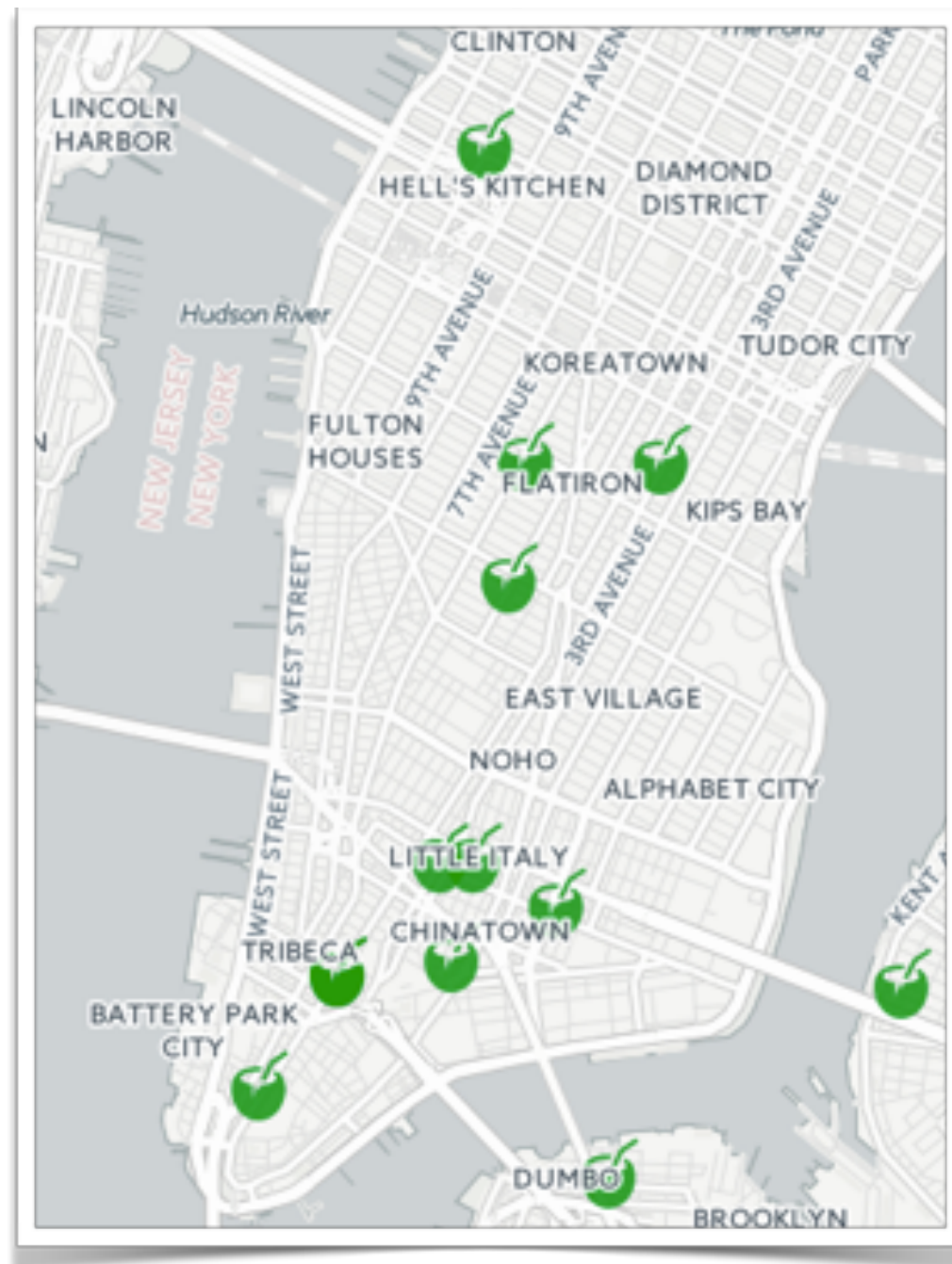
# Listening Closely to the Birds



Areas with negative tweets in Manhattan



# Upbeat and Downbeat



Positive tweets about vegetarian food in Manhattan

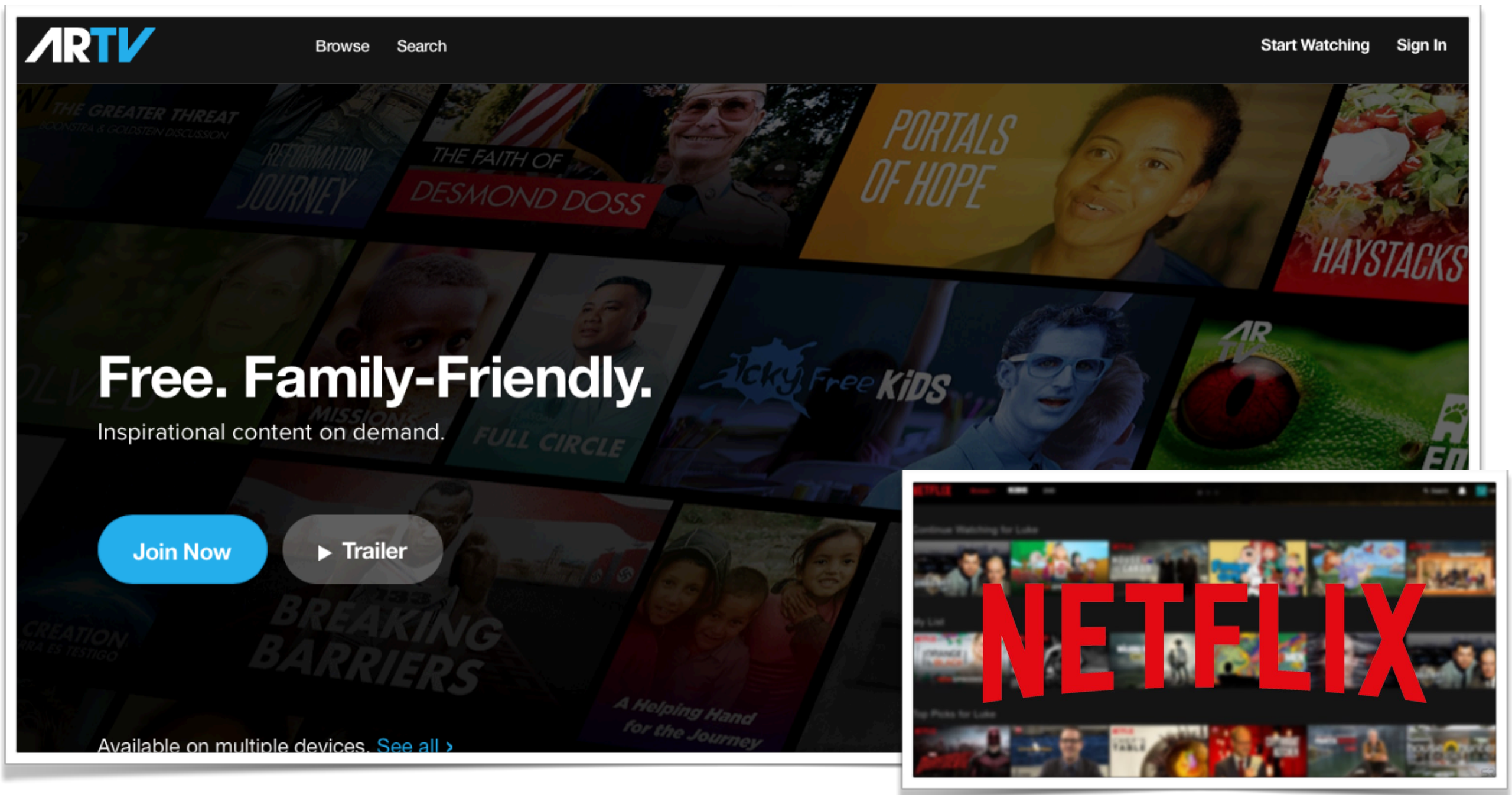
# Living in the Age of Anxiety as Revealed in Google's Big Data



Searches for “uncertainty” in the US in the past 5 years

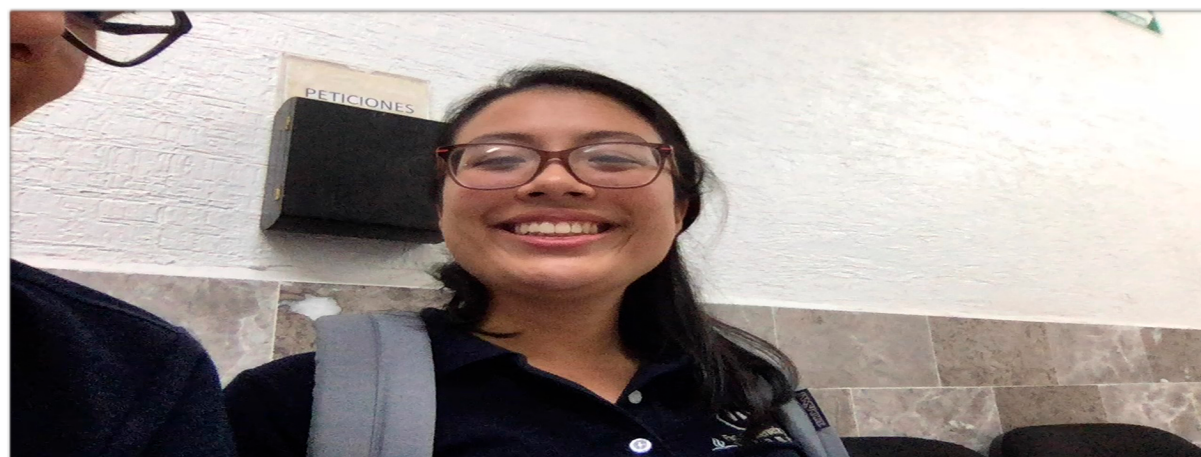
9. Alférez, G.H. (2017). Living in the Age of Anxiety as Revealed in Google's Big Data. ADvindicate. URL: <http://advindicate.com/articles/2017/11/23/living-in-the-age-of-anxiety-as-revealed-in-googles-big-data>

# Media Recommender Based on Big Data



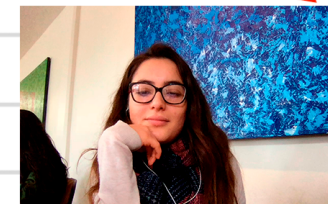


# Using Deep Learning to Discover the Emotions of People when Watching Videos



```
The TensorFlow library wasn't compiled to use SSE4.2 instructions, but these are
available on your machine and could speed up CPU computations.
2017-11-01 18:27:02.708320: W tensorflow/core/platform/cpu_feature_guard.cc:45]
The TensorFlow library wasn't compiled to use AVX instructions, but these are av
ailable on your machine and could speed up CPU computations.
2017-11-01 18:27:02.708324: W tensorflow/core/platform/cpu_feature_guard.cc:45]
The TensorFlow library wasn't compiled to use AVX2 instructions, but these are a
vailable on your machine and could speed up CPU computations.
2017-11-01 18:27:02.708328: W tensorflow/core/platform/cpu_feature_guard.cc:45]
The TensorFlow library wasn't compiled to use FMA instructions, but these are av
ailable on your machine and could speed up CPU computations.
2017-11-01 18:27:03.809365: W tensorflow/core/framework/op_def_util.cc:333] Op B
atchNormWithGlobalNormalization is deprecated. It will cease to work in GraphDef
version 9. Use tf.nn.batch_normalization().
Feliz (score = 0.81066)
Neutro (score = 0.17384)
```

1



00:28 00:42 00:56 01:10 01:24 01:39 01:53 02:07 02:21 02:35 02:49 03:03 03:17 03:31 03:45 03:59 04:13 04:27 04:41 04:55

# The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and [more](#).

## GET STARTED

SEARCH OVER **301,193 DATASETS**



## BROWSE TOPICS



Agriculture



Climate



Consumer



Ecosystems



Education



Energy



Finance



Health



Local  
Government



Manufacturing



Maritime



Ocean



Public Safety



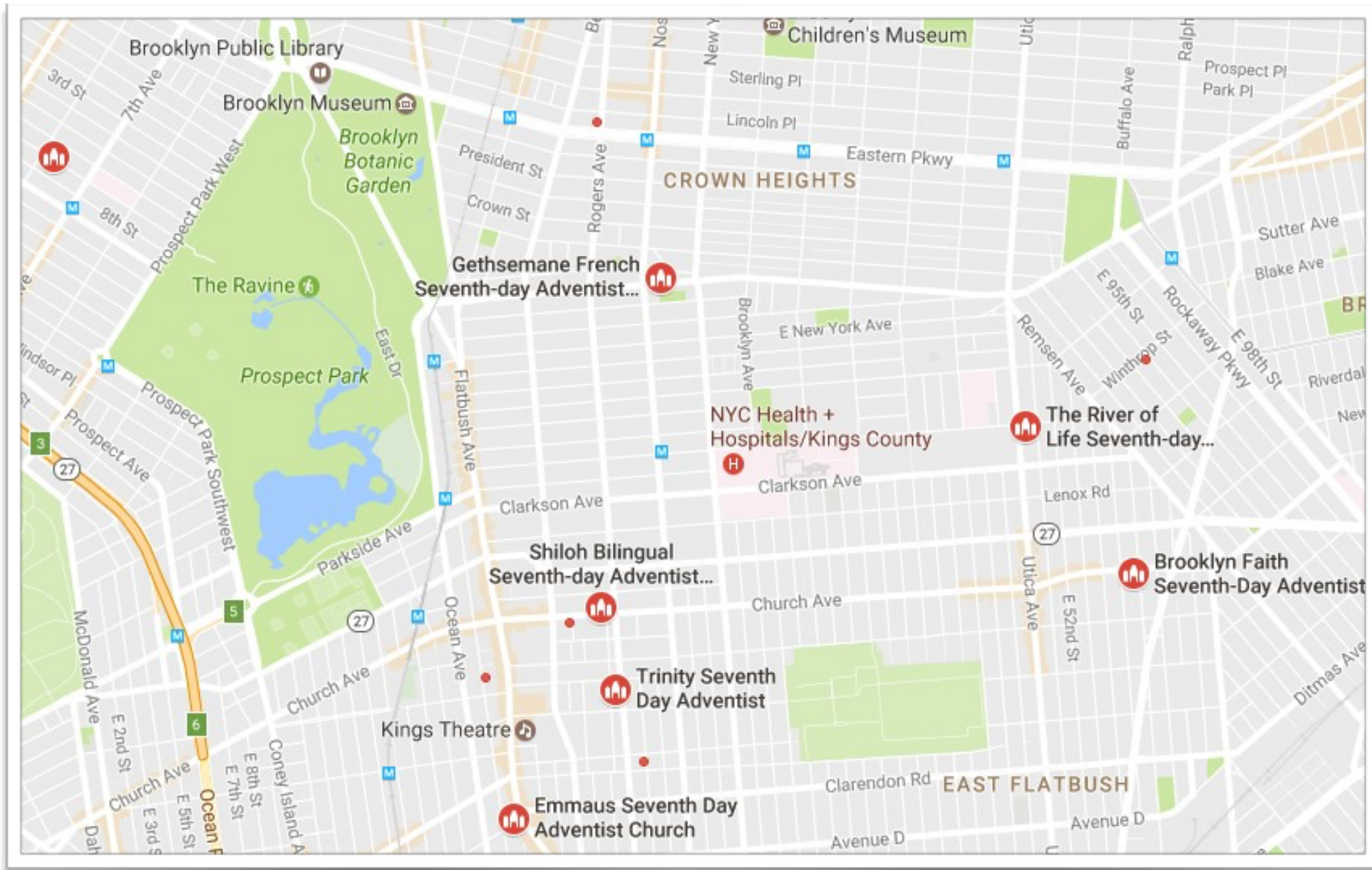
Science &  
Research



# 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 are several pedicab accidents.

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



10. 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. URL: <http://www.sdadata.org/blog/discovering-mission-oriented-patterns-with-open-data-in-new-york-city>



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



# Discovering People's Needs with Data Science

---

**Harvey Alférez, Ph.D.**

Director, Institute of Data Science  
School of Engineering and Technology  
Montemorelos University, Mexico

<https://icd.um.edu.mx/en/>

[harveyalferez@um.edu.mx](mailto:harveyalferez@um.edu.mx)

