



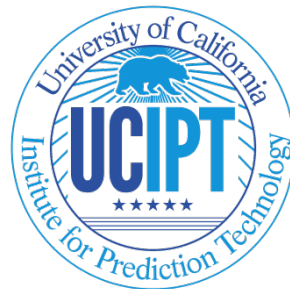
# Reading Between the Tweets:

## Using Social Data to Predict and Change Health Behaviors

Sean D. Young, PhD, MS

Executive Director, Institute for Prediction Technology

Associate Professor, UC Irvine Emergency Medicine and Informatics



# An Indiana town recovering from 190 HIV cases

**Shari Rudavsky**, shari.rudavsky@indystar.com 7:15 a.m. EDT April 11, 2016



The Scott County Health Department Needle Exchange Program was started in 2015 as a response to an HIV outbreak linked to intravenous drug use in Scott County. Austin, IN is the epicenter of the crisis. Kelly Wilkinson/IndyStar



(Photo: Matt Herp / Special to the Courier-Journal)

**AUSTIN, Ind.** — From the start of the HIV outbreak, health officials emphasized that nothing set Scott County apart from many other rural communities where opioid drug use had become an epidemic. This could happen anywhere, people were told.

Many people had viewed HIV as a big-city disease, something that might afflict people in San Francisco or New York. But not in Austin, a small city of about 4,000 people 80 miles south of Indianapolis.

Then, in February 2015, the first 30 cases of HIV were reported. By mid-March, the number had climbed to 55. State health officials, the governor and the federal Centers for Disease Control and Prevention were looking for answers. Few public

**MORE STORIES**

**The Drop for May 26: Jungle Park Speedway, what we're Googling, and Cap's haps**  
May 26, 2016, 5:53 p.m.

**Tully: Despite life's struggles, he's 'a happy-going guy'**  
May 26, 2016, 12:08 a.m.

**Can Hale help Gregg with female voters?**  
May 25, 2016, 10:48 p.m.

- 3176
- 
- 4
- 
- 303





*each day*

500,000,000  
tweets are sent

55,000,000  
status updates on Facebook

Source: internet live stats and KISSmetrics



*each day*

700,000,000  
snaps are sent

70,000,000  
Instagram photos

Source: TheStreet



Given the fact that you didn't update your Facebook status saying you brushed your teeth, I'm a little worried you didn't.



your  cards  
someecards.com



**Jacob Cox-Brown**

2 hours ago · 🌐

Drivin drunk... classic ;) but to whoever's vehicle i hit i am sorry. :P



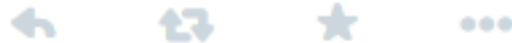
**Jae** @xpink\_homicide · 43m

I'm really tryna have **sex tonight**



**I Throw Ooops** @Last80sbabi · 10m

"@WillPounder: 17. When was the last time u got checked for **std's** ?" Abt Ah week ago



[View conversation](#)



**ambiguous** @YungMeese · 4m

Why **am I addicted to** marihuana?!



**The Pimp Juices** @potatoshitface · 37m

I'm going to **buy cocaine**



**Bon** @MiahMiah · 24h

You made me into a machine with drugs gave me **hiv** made me gay now why don't I **have** anything in my life? Why so desperate for everything?





# What is “*social data*”?

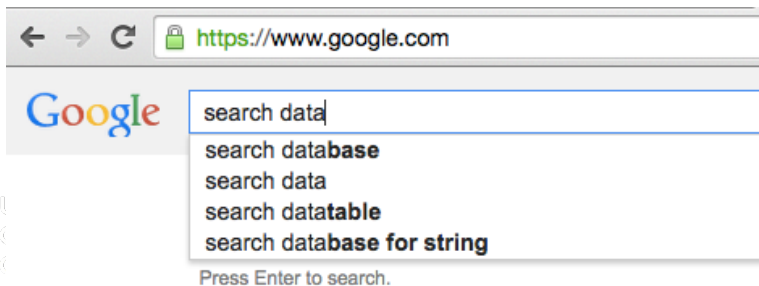
data which provide information  
about user’s behaviors

*examples:* tweets, snaps, wearable  
device, search data, mobile apps





goal: use social data to gain insights into population behaviors and predict events in **medicine**, **politics**, **crime** and **public health**





**Predicting HIV Incidence**

The number of Twitter "tweets" suggestive of HIV risk behaviors predicts the number of HIV cases at the county level.

[Learn More](#)



**HEALTH**



**SECURITY**



**POLITICS**



**FINANCE**



**PRODUCTS**



**POVERTY**

# An Indiana town recovering from 190 HIV cases

Shari Rudavsky, shari.rudavsky@indystar.com 7:15 a.m. EDT April 11, 2016



The Scott County Health Department Needle Exchange Program was started in 2015 as a response to an HIV outbreak linked to intravenous drug use in Scott County. Austin, IN is the epicenter of the crisis. Kelly Wilkinson/IndyStar



(Photo: Matt Herp / Special to the Courier-Journal)

**AUSTIN, Ind.** — From the start of the HIV outbreak, health officials emphasized that nothing set Scott County apart from many other rural communities where opioid drug use had become an epidemic. This could happen anywhere, people were told.

Many people had viewed HIV as a big-city disease, something that might afflict people in San Francisco or New York. But not in Austin, a small city of about 4,000 people 80 miles south of Indianapolis.

Then, in February 2015, the first 30 cases of HIV were reported. By mid-March, the number had climbed to 55. State health officials, the governor and the federal Centers for Disease Control and Prevention were looking for answers. Few public

### MORE STORIES



**The Drop for May 26: Jungle Park Speedway, what we're Googling, and Cap's haps**

May 26, 2016, 5:53 p.m.



**Tully: Despite life's struggles, he's 'a happy-going guy'**

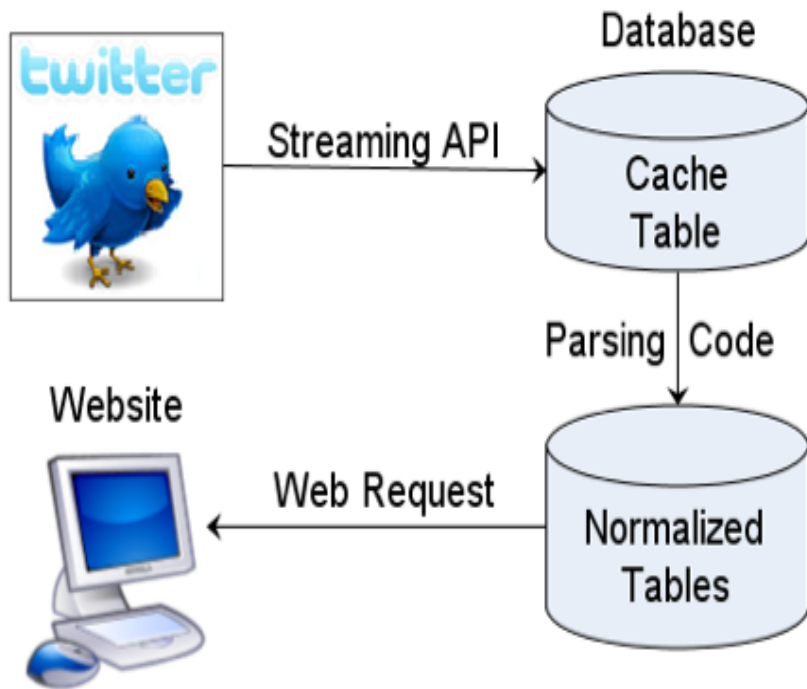
May 26, 2016, 12:08 a.m.



**Can Hale help Gregg with female voters?**

May 25, 2016, 10:48 p.m.

# Data Science to Predict HIV?



**All collected tweets  
N = 553,186,061 (100%)**

**USA geolocated tweets  
N = 2,157,260 (0.4%)**

**Includes keyword  
N = 9,880 (0.5%)**

**Drug keyword  
N = 1,342  
(14%)**

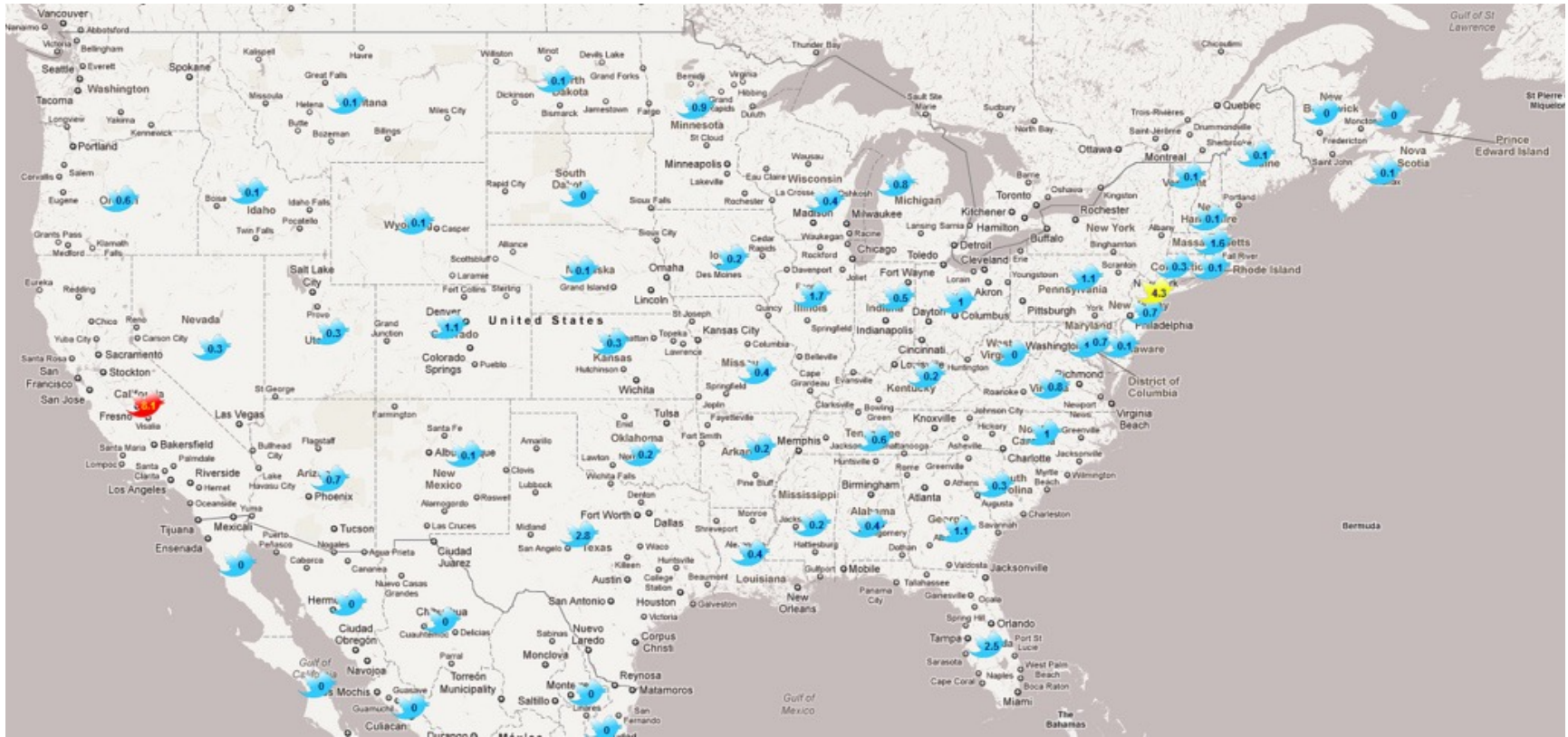
**Sex keyword  
N = 8,538  
(86%)**

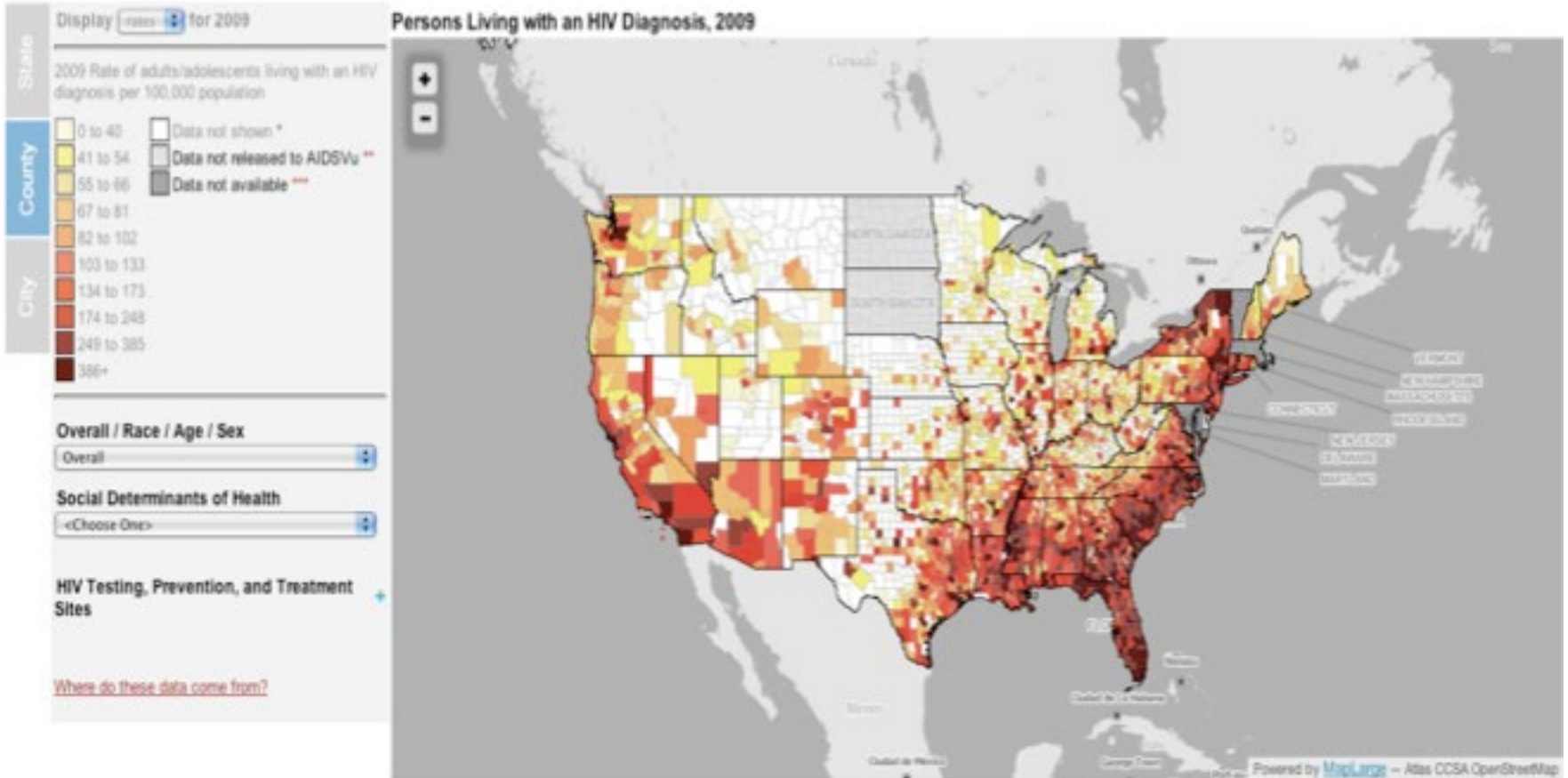
**From county  
with HIV data  
N = 1,233  
(92%)**

**From county  
with HIV data  
N = 7,811  
(92%)**

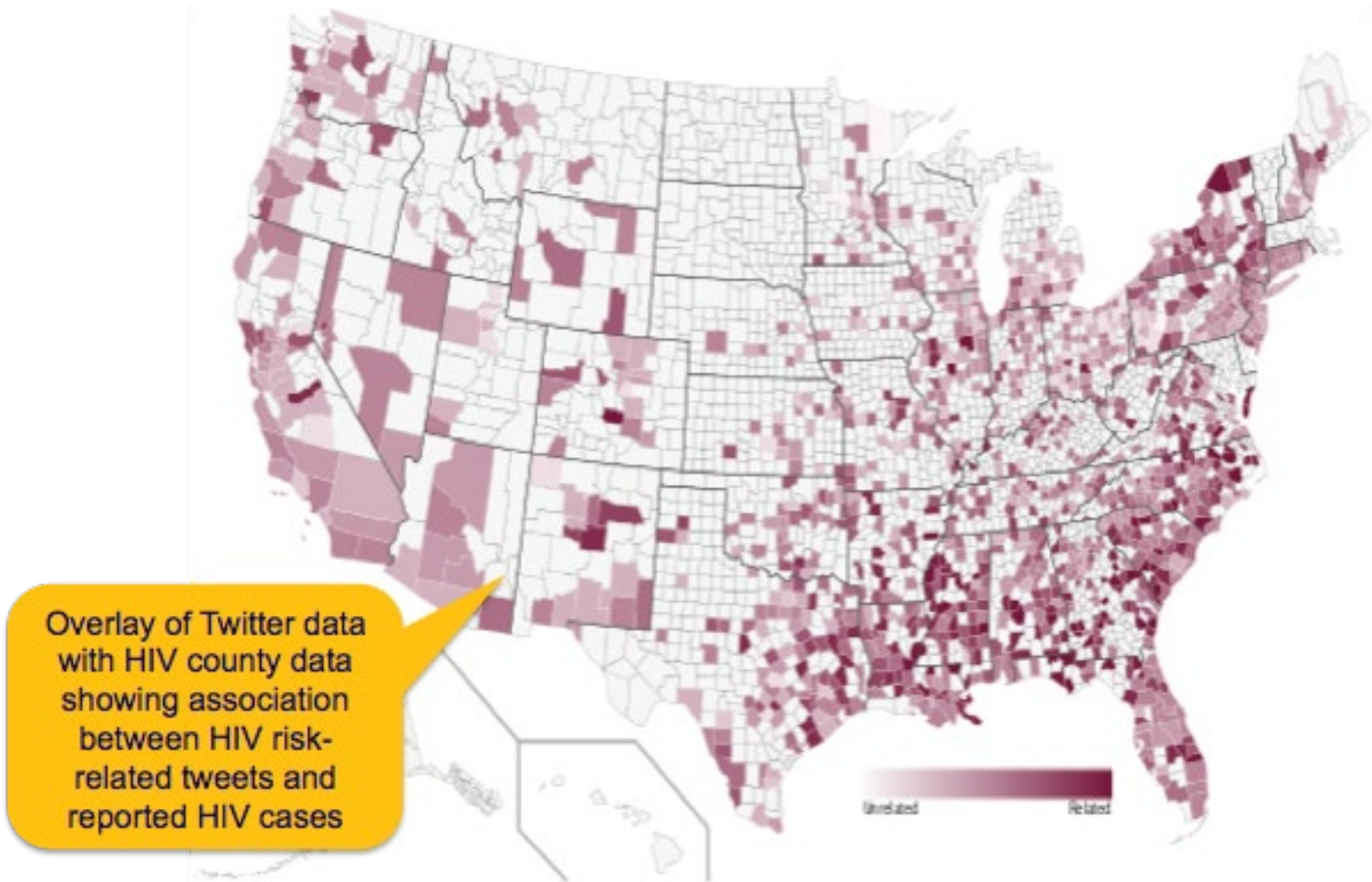


<i>Category</i>	<i>Sample 1</i>	<i>Sample 2</i>	<i>Sample 3</i>
<b><i>Sex risk-related</i></b>	I neeeeeeed sex now!	Sex ain't better than love	[@ another user] all about the sex
<b><i>Substance use-related</i></b>	Anybody wanna get high?	who got some drugs for tonight?	need an easier way to get drugs
<b><i>General</i></b>	This weather is crazy!	Watching the Celtics game.	headed to take my exam, wish me luck







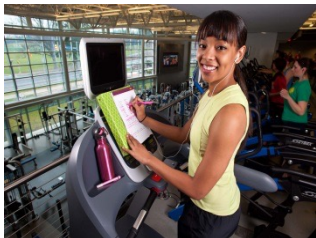


# *Student Health*





# Partnerships – UCLA Chancellor and Intel



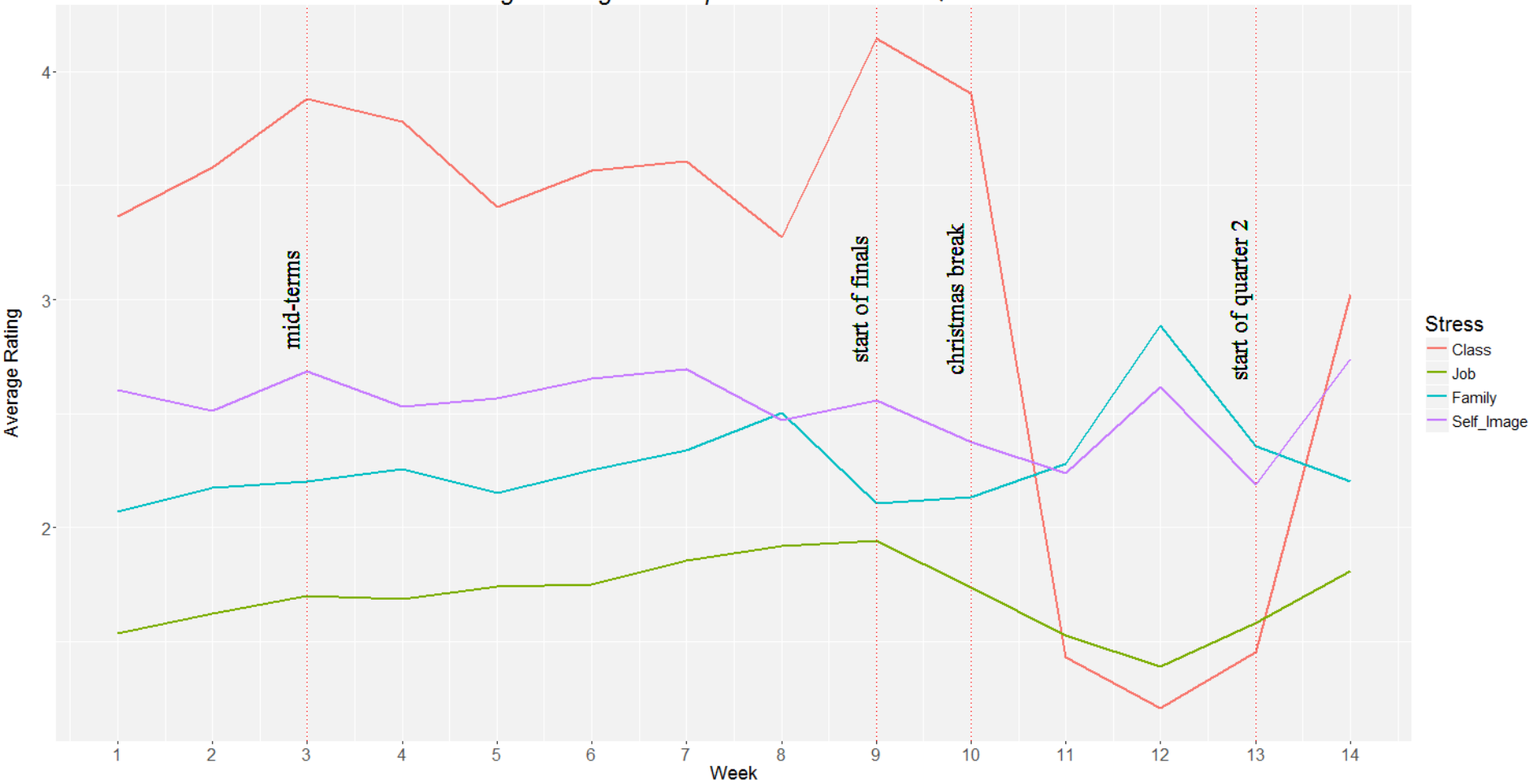
# Data and Approach: Freshmen 200

- 200 UCLA freshmen took weekly surveys on emotions, stress-levels, activities
  - Followed them on Twitter
  - Wore fitness/sleep trackers
- 
- We seek to use these social data to predict student health
  - Can be used by administration and campus services



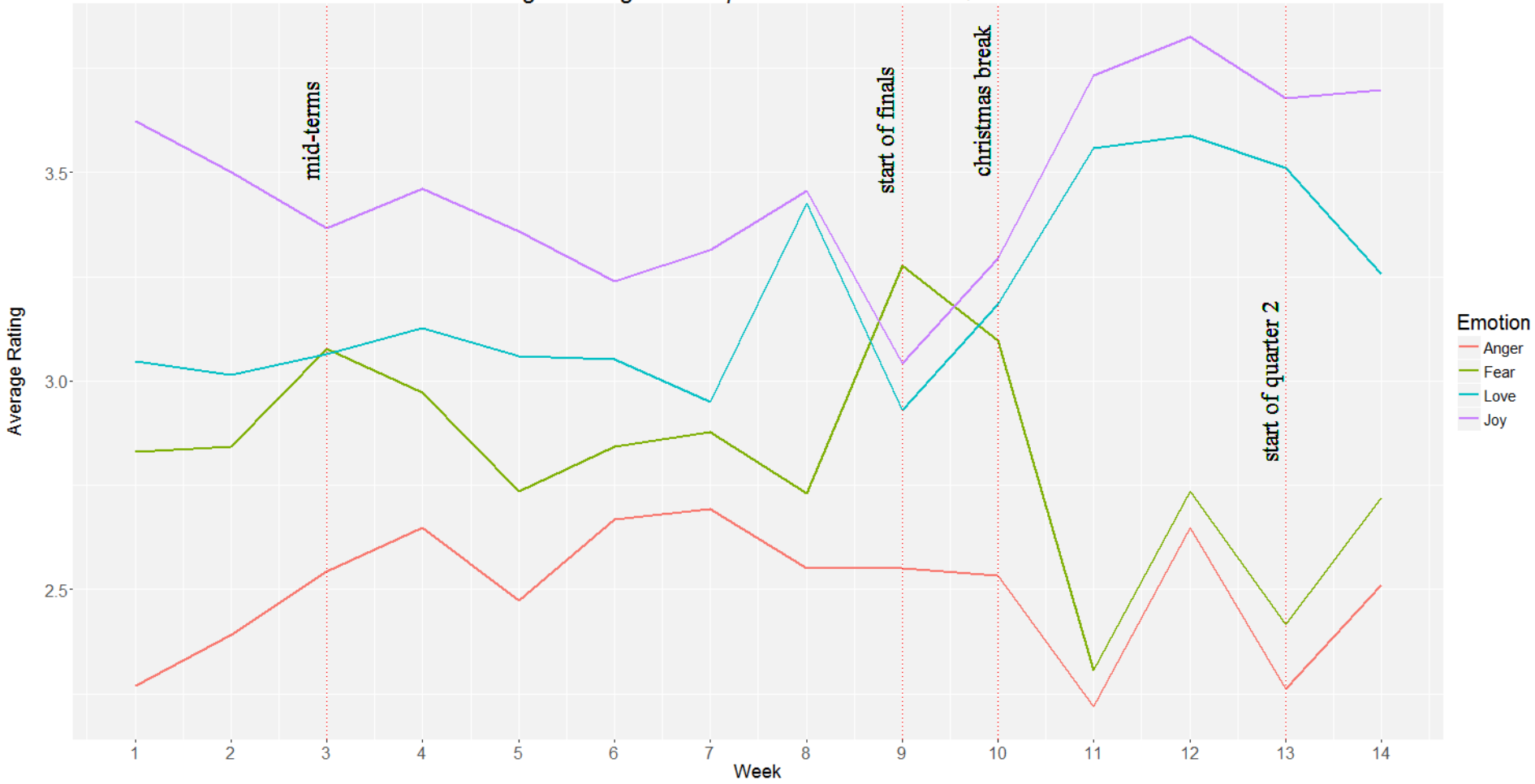
# Study 1: Results

Average Rating of Multiple Stressors over Quarter 1



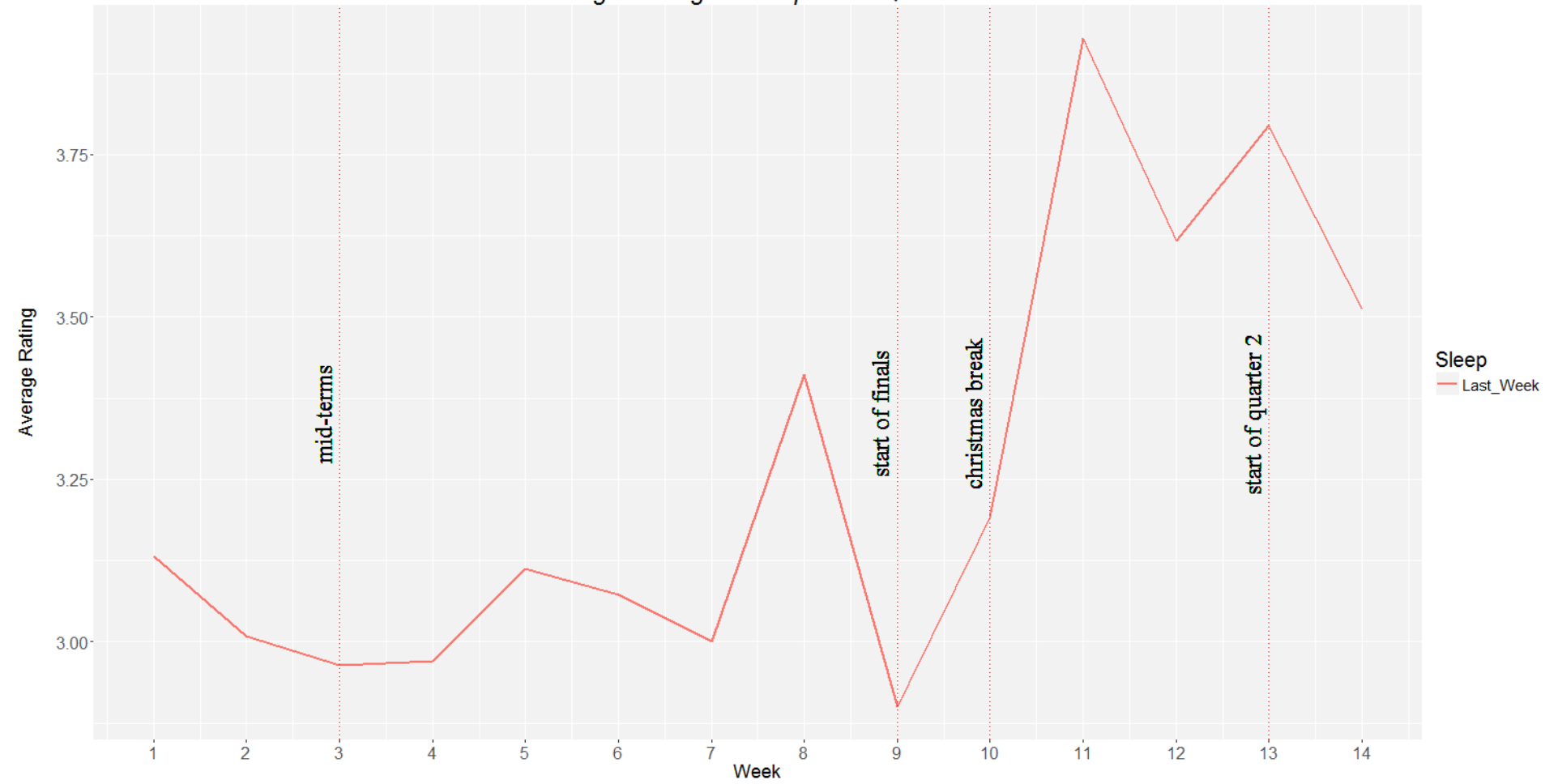


Average Rating of Multiple Emotions Over Quarter 1





Average Rating of Sleep Over Quarter 1





**Tweets that expressed emotions of love and kindness (Positive)**

- Blessed At The End Of Every Day.
- Lord do I have such amazing and supportive friends
- I'm Telling You, There Is A Silver Lining To Everything :-')

**Tweets that expressed Happiness/excitement (Positive)**

- Finally done wit midterms
- Looks like a successful night ..
- I just ate next to Zaza Pachulia at In N Out's westwood !!!

**Tweets that expressed emotions of anger (Negative)**

- I swear these people just stare at me it annoying
- I hate essays with a passion....
- I hate math now

**Tweets that expressed emotions of fear (Negative)**

- College is just too stressful
- I chose Netflix and sleep over studying and now 1 hour before my test I hate myself
- Oh how I missed eating breakfast at my house

**Neutral tweets**

- I want tacos.
- In history waiting for my caffeine to kick in
- McDonalds run





# Study 1: Twitter Results

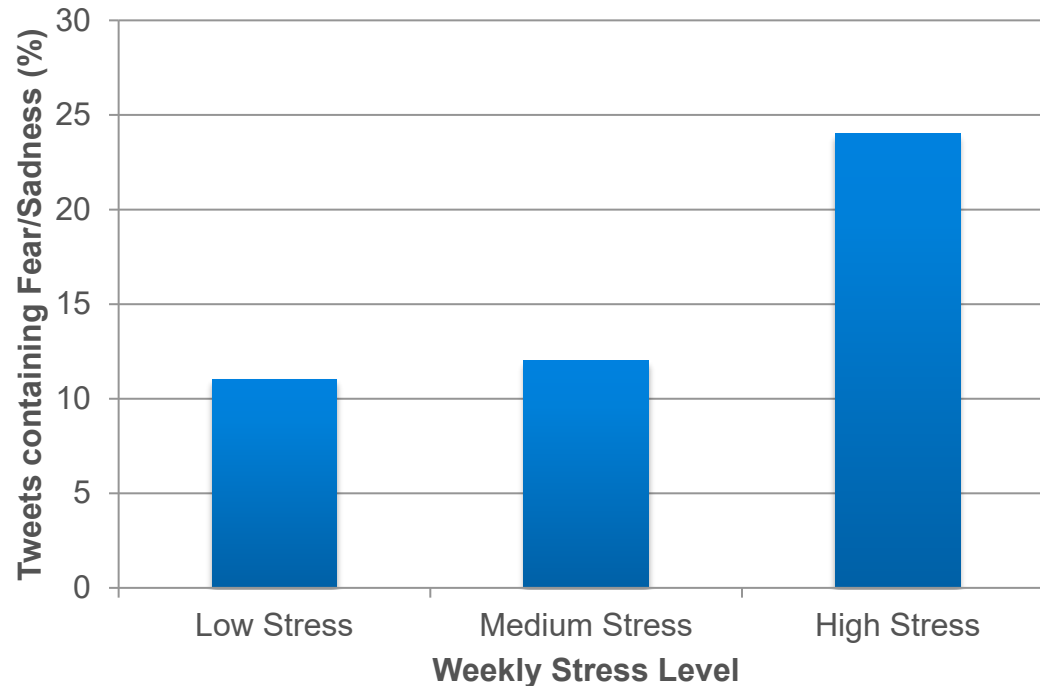
- Weekly stress was significantly associated with
  - a greater percentage of negative sentiment tweets
  - tweets containing sadness
  - tweets containing love and hope ( $\beta \pm SE: 3.6 \pm 1.4; P=0.01$ ).
  
- Anger was negatively associated with
  - percent of positive sentiment ( $\beta \pm SE: -1.6 \pm 0.8; P=0.05$ )
  - tweets related to joy ( $\beta \pm SE: -2.2 \pm 0.9; P=0.02$ ).
  
- Fear was positively associated with
  - negative sentiment ( $\beta \pm SE: 1.67 \pm 0.7; P=0.012$ )
  - a greater proportion of sadness tweets ( $\beta \pm SE: 2.4 \pm 0.8; P=0.01$ ).



# Students with ↑ weekly stress level

↑ Tweets containing Fear/sadness

*“College is just too stressful..”  
“I chose Netflix and sleep over studying and now 1 hour before my test I hate myself”*





# Create an Automated Tool



**Sentiment Analysis:** Classify each tweet into Positive, Neutral, or Negative



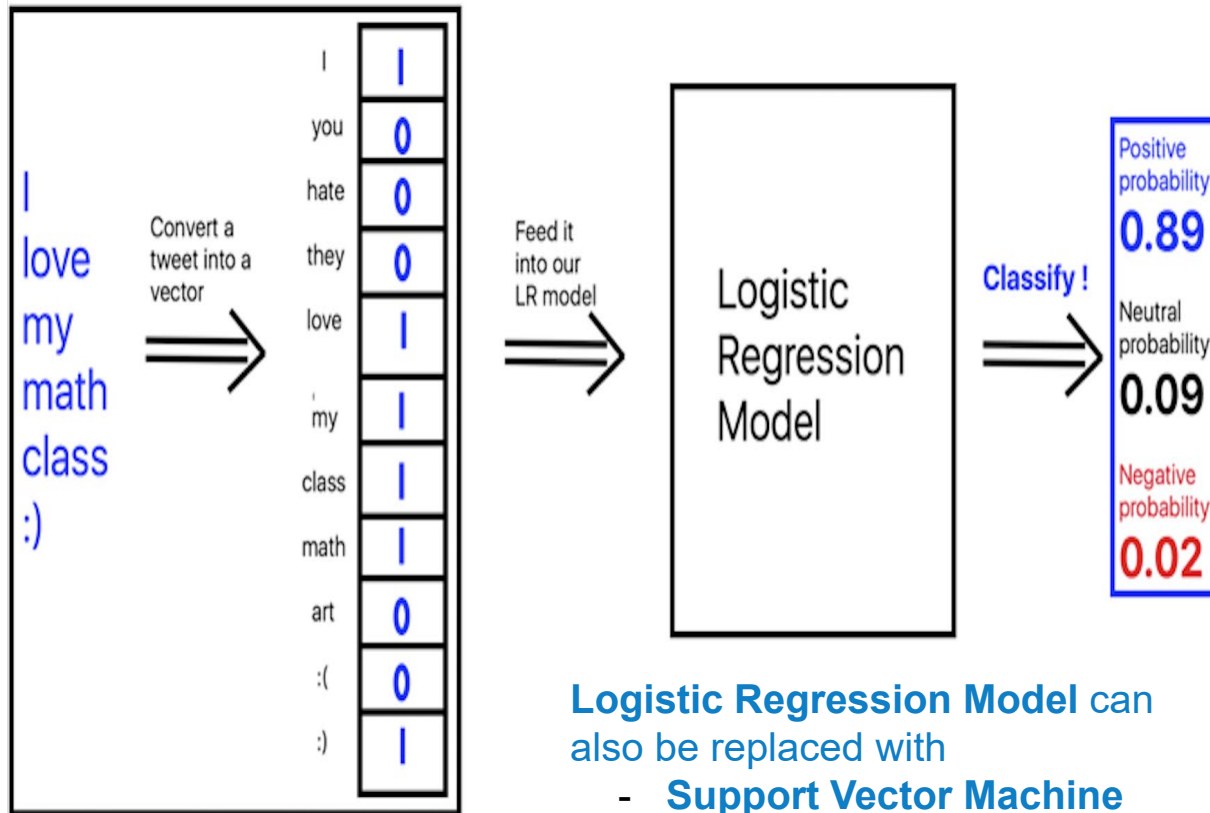
# AI: Bag of Words Model

**Bag of Words Model** - relies on the frequency of words that are highly correlated with each sentiment

We first explored various machine learning models based on bag of words

1. **Logistic Regression** - can capture linear relationships between predictors and dependent variables (targets) well in general
2. **Support Vector Machine** - can capture more complex, high-dimensional relationships better
3. **Random Forest** - can learn generalized, universal patterns from input dataset

# Bag of Words Model



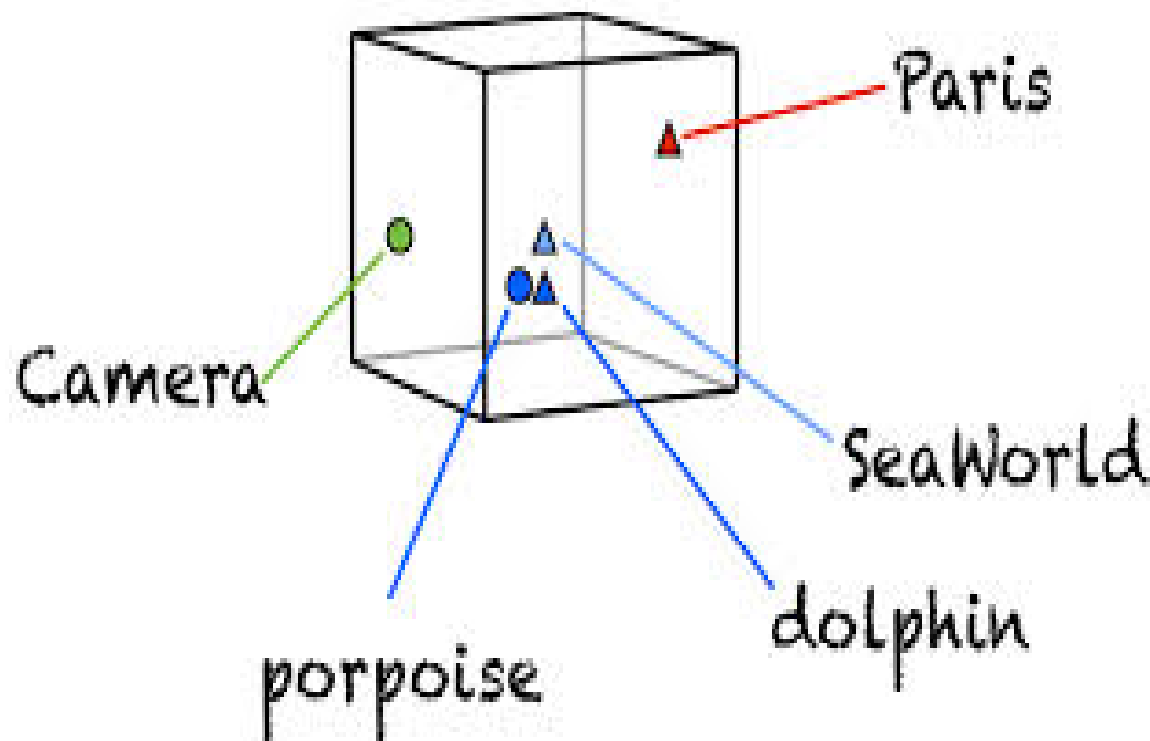
1. An input tweet is converted into a vector, each of whose elements corresponds to the frequency of a particular word in the tweet

2. The vector is fed into LR model, and the model predicts the probability for each sentiment

**Logistic Regression Model** can also be replaced with

- **Support Vector Machine**
- **Random Forest**

# AI: Deep Neural Network



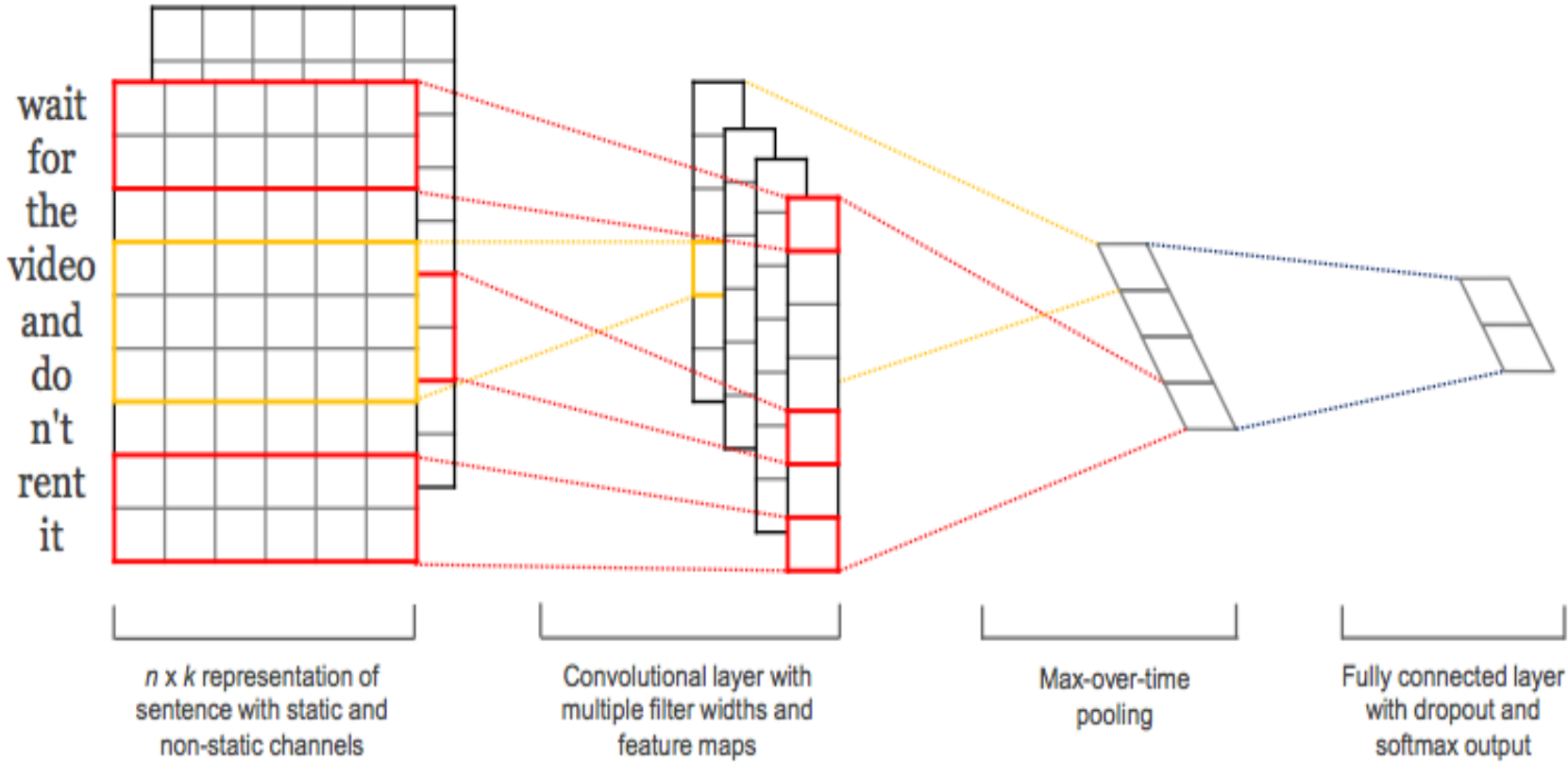
Deep Neural Network uses **word embeddings**, which are the vectorized representations of words.

These word vectors contain semantic as well as syntactic information of words.

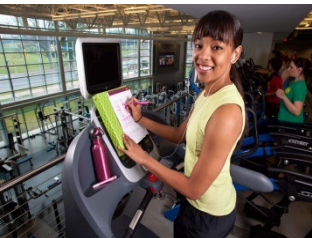
Notice the word vectors of “dolphin,” “porpoise,” “sea world” are close to each other in the above 3-D word space. (We can go beyond 3-D space to represent words with vectors)



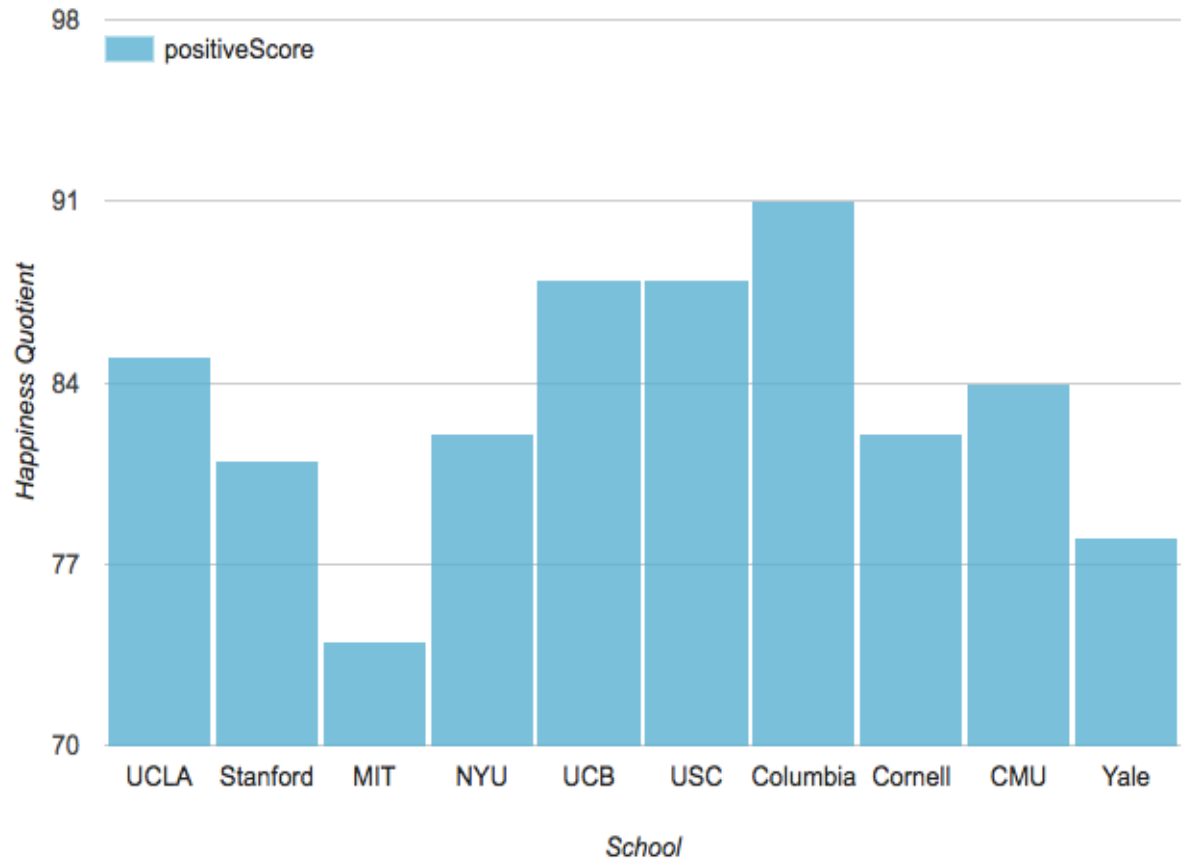
# Deep Convolutional Neural Network



# Partnerships – UCLA Chancellor and Intel



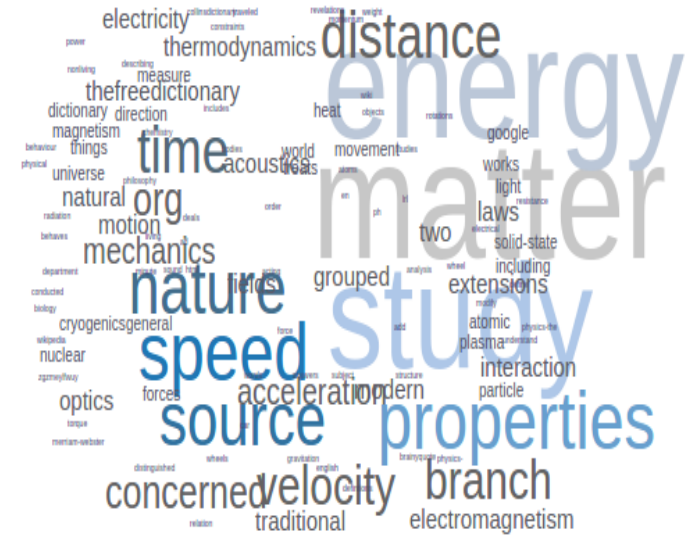
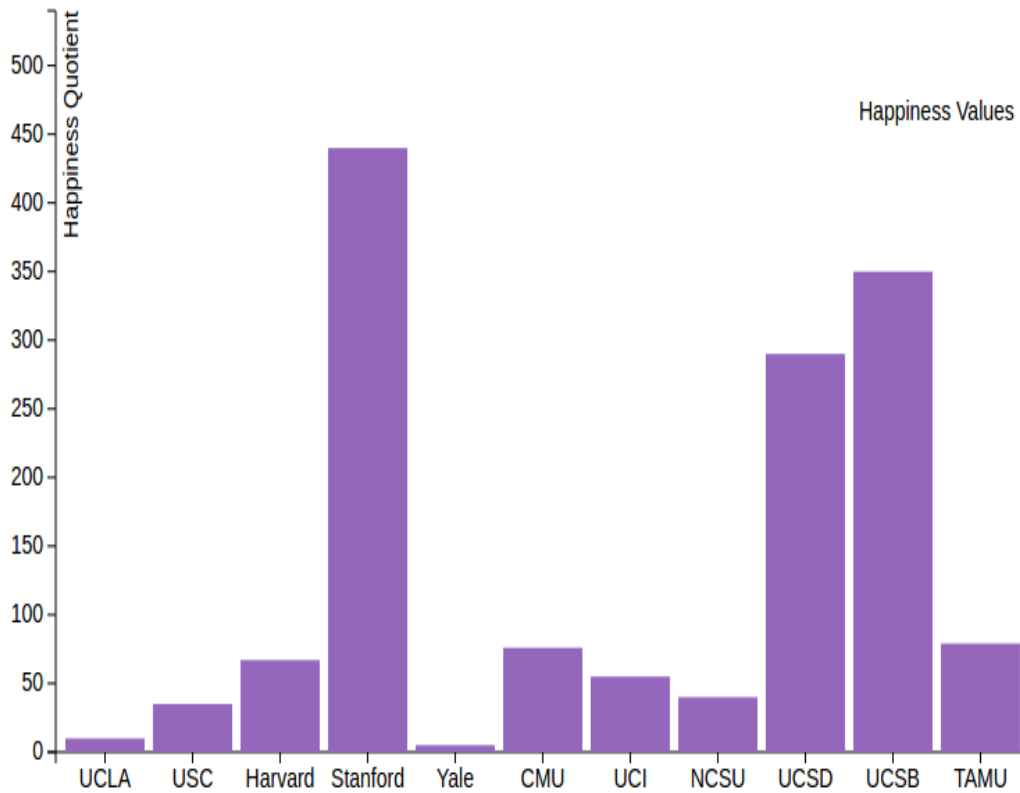
## Happiness by School





# Stress Quotient

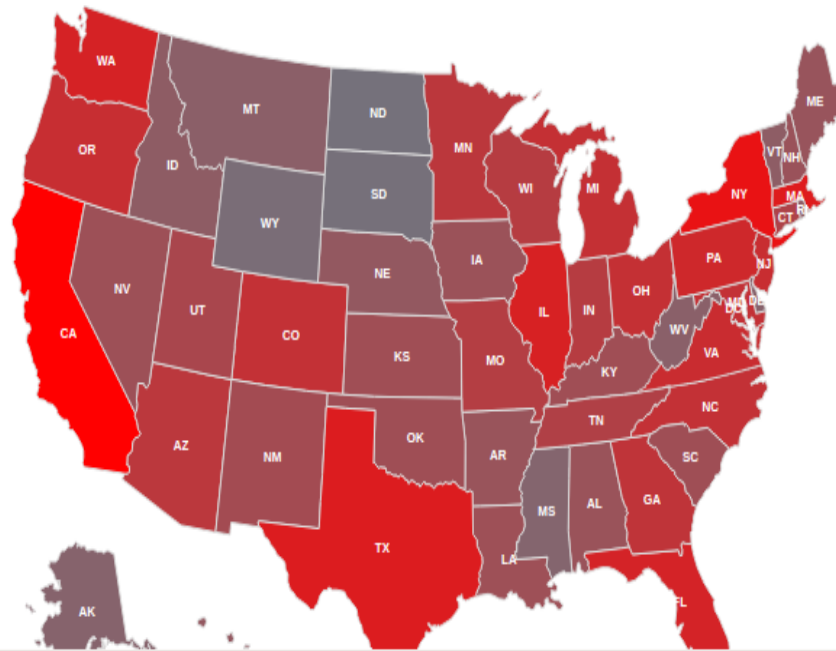
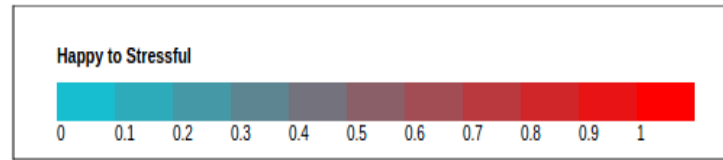
[Home](#) [Stress Map](#) [School Ranking](#) [Contact us](#)



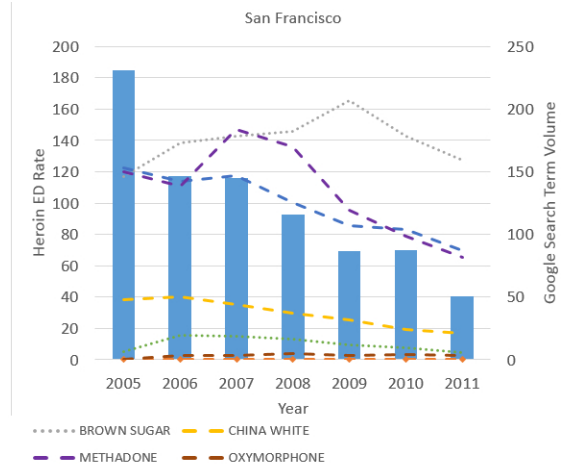
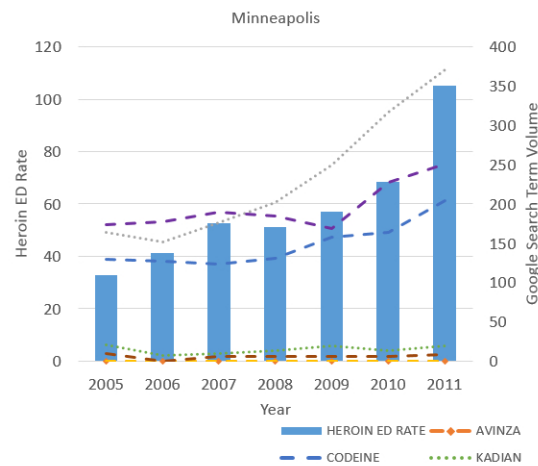


# Stress Quotient

- Home
- Stress Map
- School Ranking
- Contact us

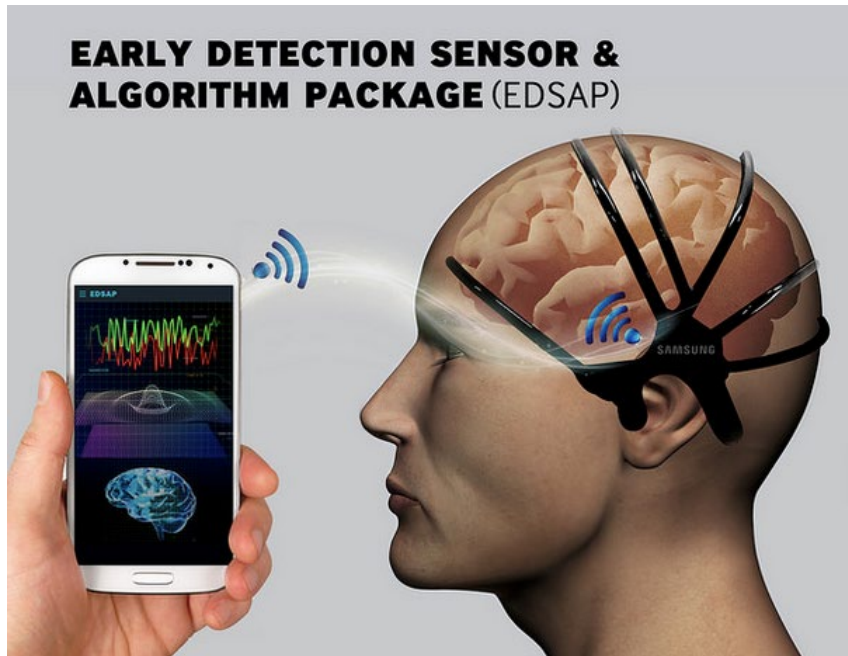


# Partnerships – Opioids



Note: Google Search Term Volume is calculated by Google as the probability of the search based on geography and time period which is then multiplied by 10 million to be human readable.

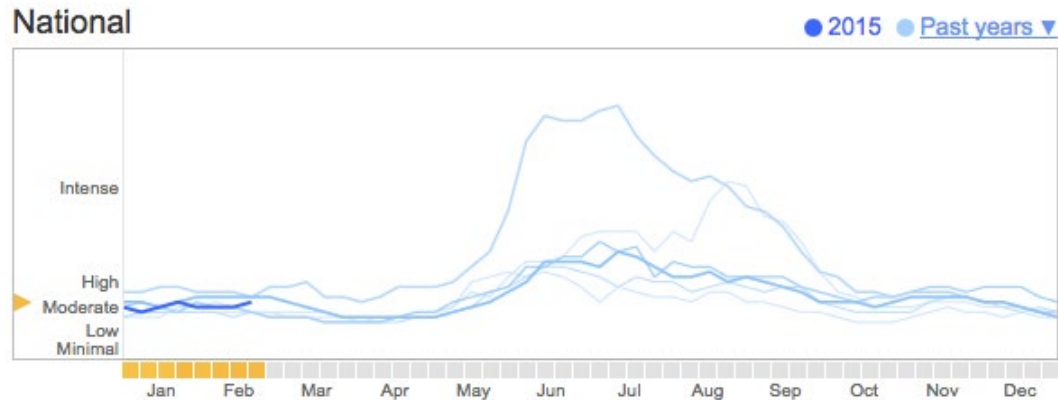
# Wearable Data and Health Events



Source: Samsung, Apple



# Search Query Trends



Top 5 most searched celebrity selfies in 2014:

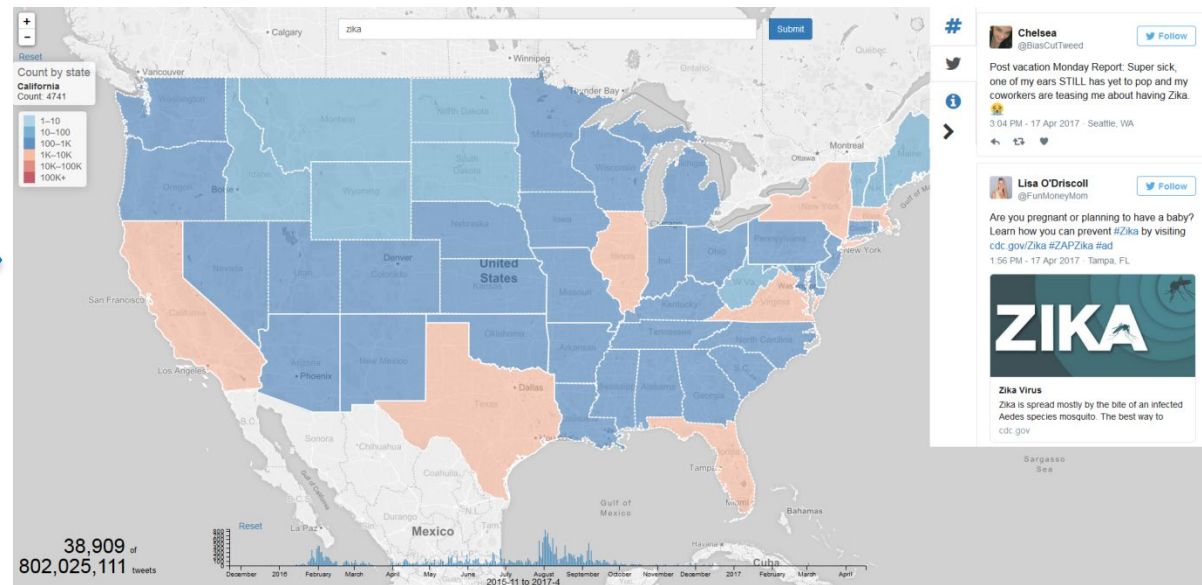
1. Ellen DeGeneres
2. Emma Stone
3. President Obama
4. Kim Kardashian
5. Danica Patrick

Top 5 most searched global news events in 2014:

1. Ebola
2. ISIS
3. Malaysia Airlines
4. Crimea / Ukraine
5. Ferguson

Source: Google

# Partnerships – Others



# Implementation Challenges



## *Digital Health Behavior Change*

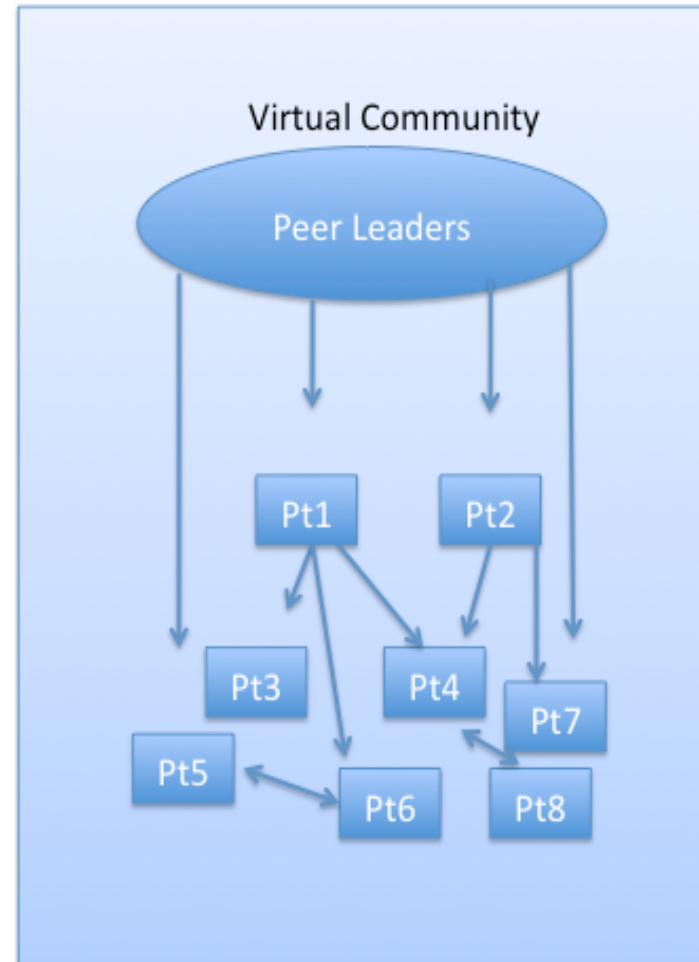


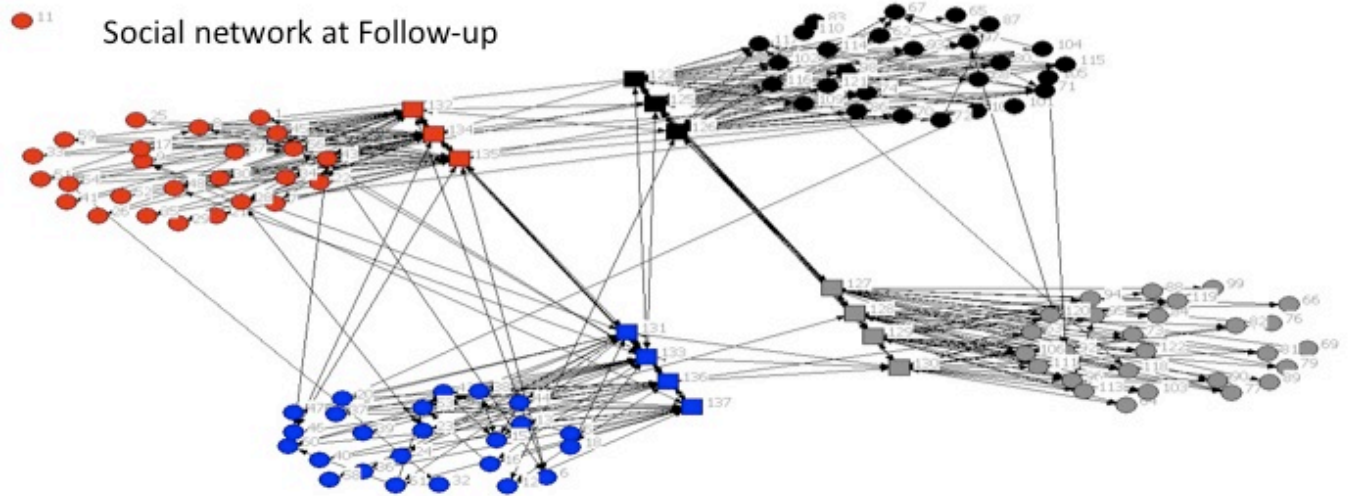
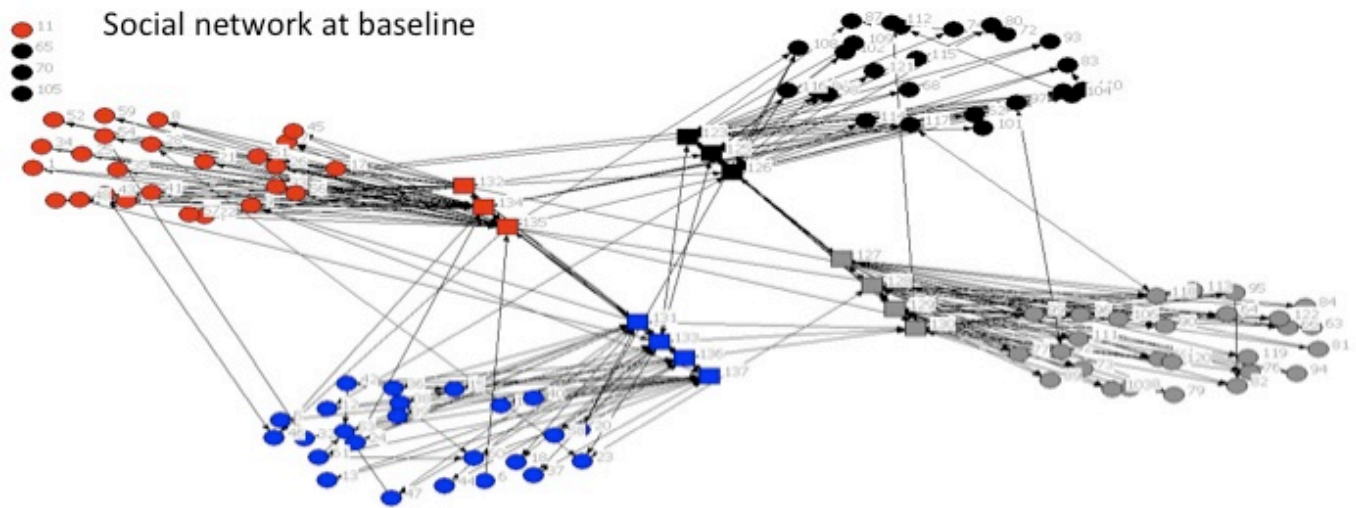
**Step 1: Establish community norms**

Peer leaders reach out at community and individual level to promote HIV prevention norms

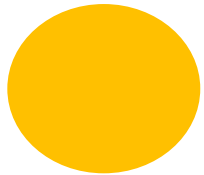
**Step 2: Community norm diffusion**

Participants communicate with each other about the importance of HIV prevention and testing. Vocal community members help to spread norms to other community members, and ultimately, to people outside of the virtual community.

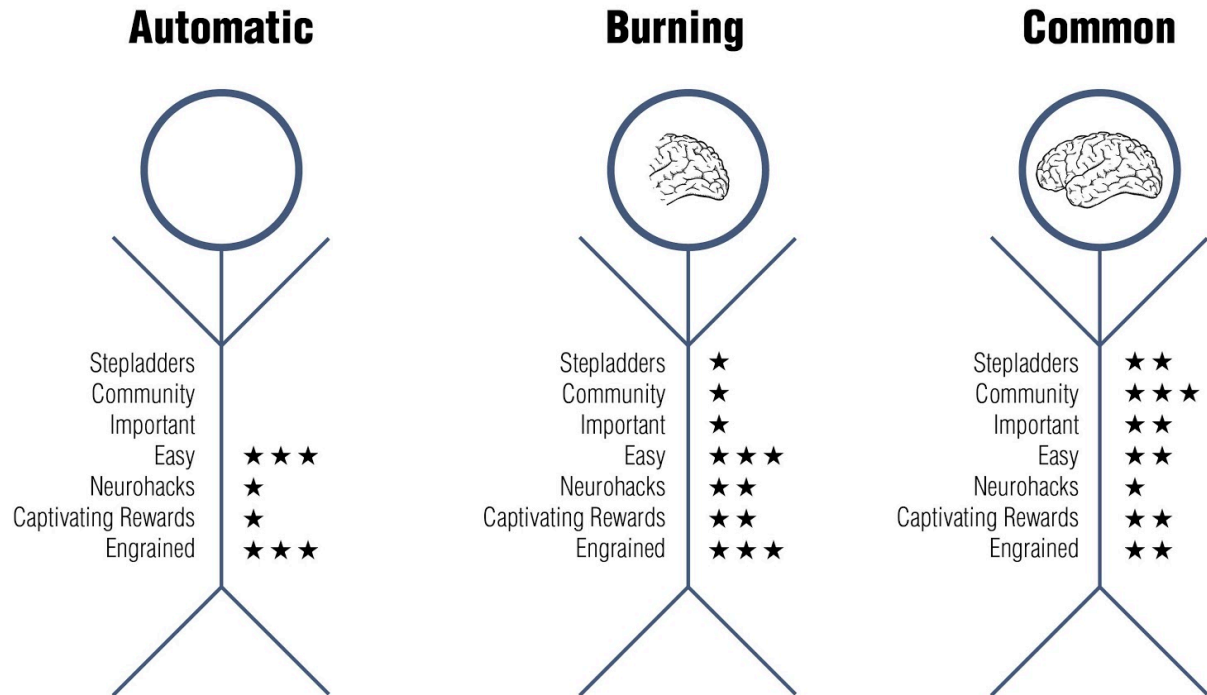




# A Guiding Theory Is Needed



Stick with It  
At a Glance



Young, S.D. *Stick With It* (2017). Harper Collins. New York



*research • partnerships • sponsorships*

***get involved***



predictiontechnology.ucla.edu  
[hello@predictiontechnology.ucla.edu](mailto:hello@predictiontechnology.ucla.edu)  
*@seanyoungphd*