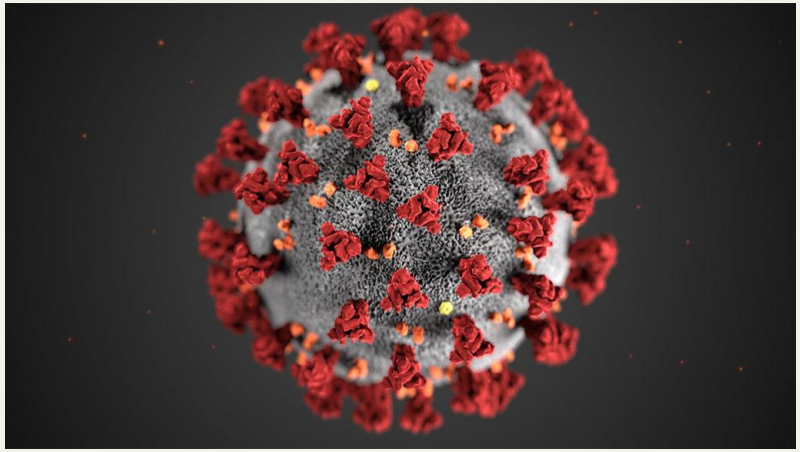
Pigs, Bats & Microchips

Using Survey Data as a Predictor of Pandemic Vaccination

Mark Patterson

March 2021

Covid-19



In the USA as of March 4, 2021:

29,130,219

Confirmed cases

525,560

Deaths

> 2.4 million deaths globally

Vaccination is Key

Began: Dec. 20th

Progress:

- 15.5% one dose
- **7.8%** fully vaccinated

Need: 75% to 85% to be effective

Vaccination is Key

Began: Dec. 20th

Progress:

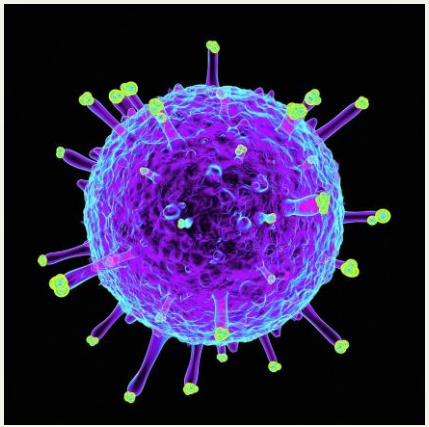
- 15.5% one dose
- **7.8%** fully vaccinated

Need: 75% to 85% to be effective

*Only 49% of <65yo said they would…**

* According to a Dec. 2020 poll of adults in USA.

H1N1:
(swine flu)



April 2009 – viral pandemic

~ 573,000 deaths globally

In the USA 2009/2010*:

60,800,000

Confirmed cases

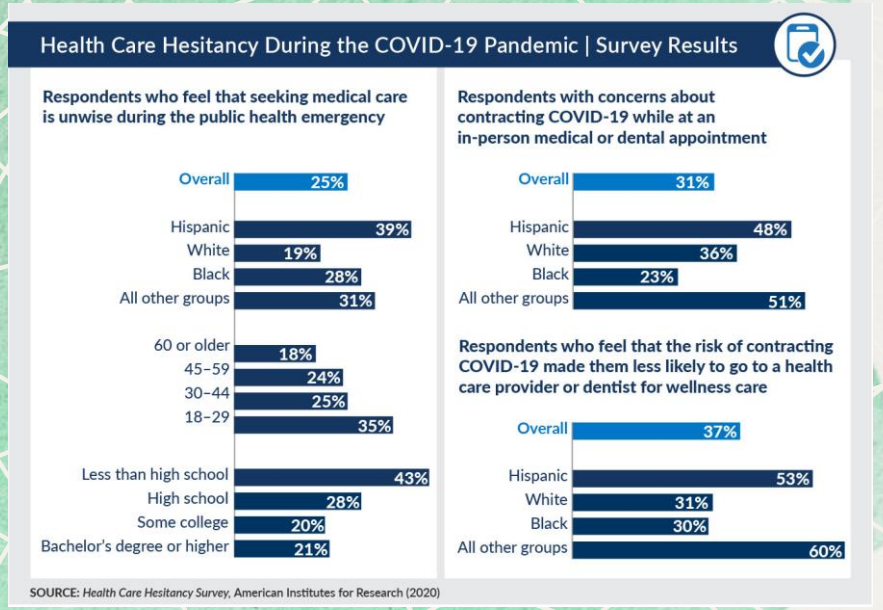
12,469

Deaths

* Based on estimates from the CDC

Analyis Questions

1. Can we use a **survey** to predict if people got vaccinated (h1n1)?
2. What are the key **factors**?
3. Are there other patterns or groupings to help ID **who** got vaccinated?
4. Can we ID current **concerns** to improve chances of getting vaccinated?

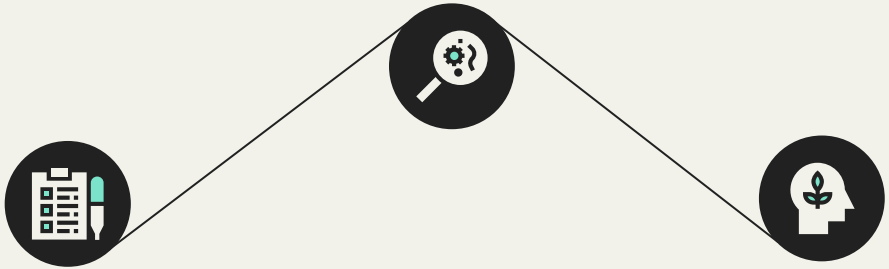


A Muti-Method Approach

2

Clustering

Do any patterns emerge to help tell who got vaccinated?



1

Classification

Data from 2009 H1N1 Survey to predict vaccination status

3

NLP (LDA)

Examine current Tweets to ID why not get vaccinated

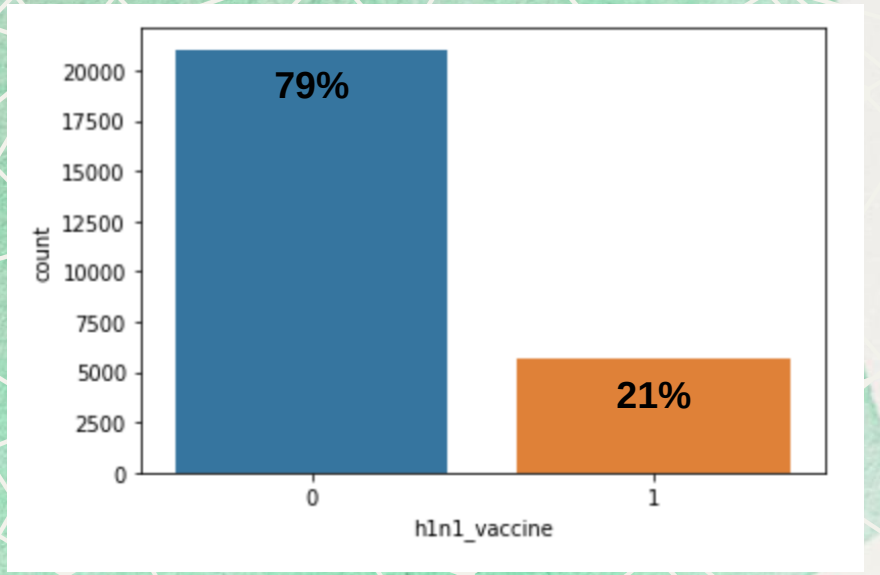
1

Classification

Can we predict who got vaccinated and who didn't? What key features help us predict?

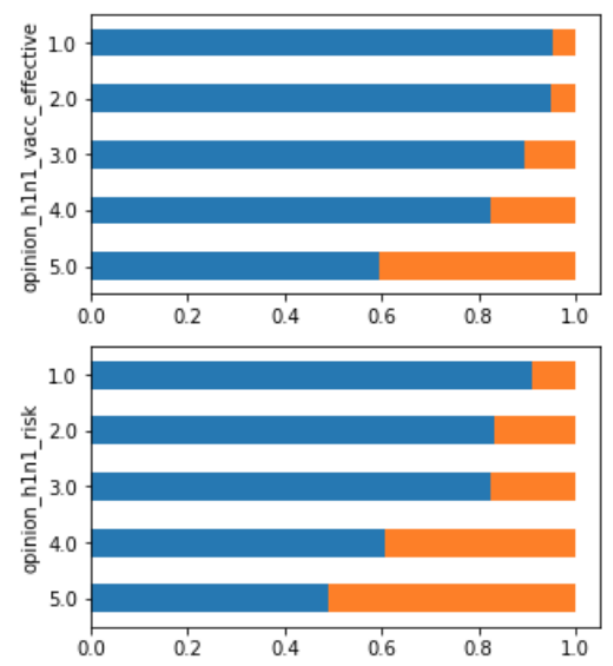
Data & Processing

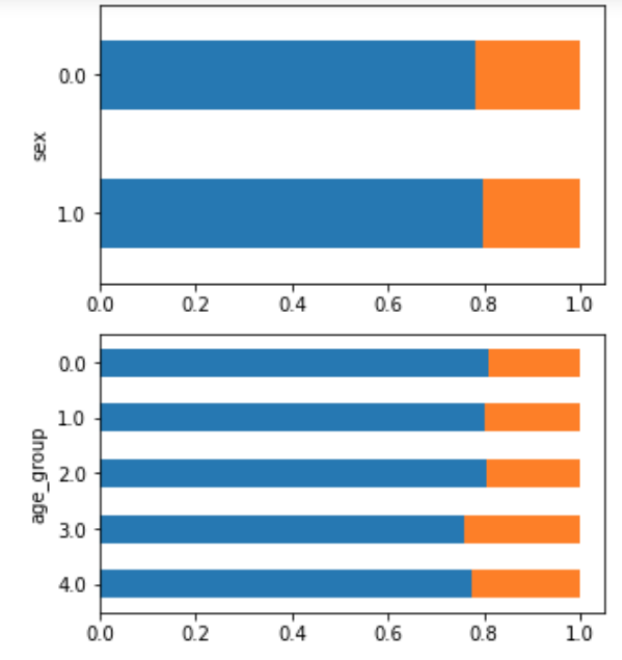
- Phone survey: 2009 to 2010
- 26,700 responses
- 36 questions
- Class imbalance (79% not vax)
- Data Preparation:
  (Ordinal Encoder and One-Hot; KNN Imputer; StandardScaler; SMOTE)



Opinions appear important; demographics less so

Orange bars represent the proportion of respondents that got the H1N1 vaccination





Classification Modeling

6 models:

XGBoost,
RandomForest,
SVC,
LogisticRegeression,
KNN,
DecisionTrees

Preprocessing & optimization

Model simplification:

19 features
10 features

XGBoost: best results*

.84
Accuracy

.71
Precission (1)

Implication: Ok ability to predict who got vaccinated, but mistakenly claims 29% got vaccined who didn't.

* This model included:
72 features (lots of one-hot encoding); no SMOTE

Key Factors

- Looked at feature importance from Random Forest models
- Consistent across various data preparations and # features.
  - Doctor recommended is top
  - Opinions (h1n1; seasonal flu) also important
  - Demographics less important

Feature Importance	
doctor reccomends h1n1 vaccine	0.12
doctor reccomends seasonal vaccine	0.04
o_h1n1 feel at risk	0.09
o_h1n1 vaccine is effective	0.08
o_h1n1 sick from vaccine	0.06
concerned about h1n1	0.06
o_seasonal flu feel at risk	0.06
o_seasonal flu can get sick from vaccine	0.06
o_seasonal flu vaccine is effective	0.05
age group	0.07
education	0.06

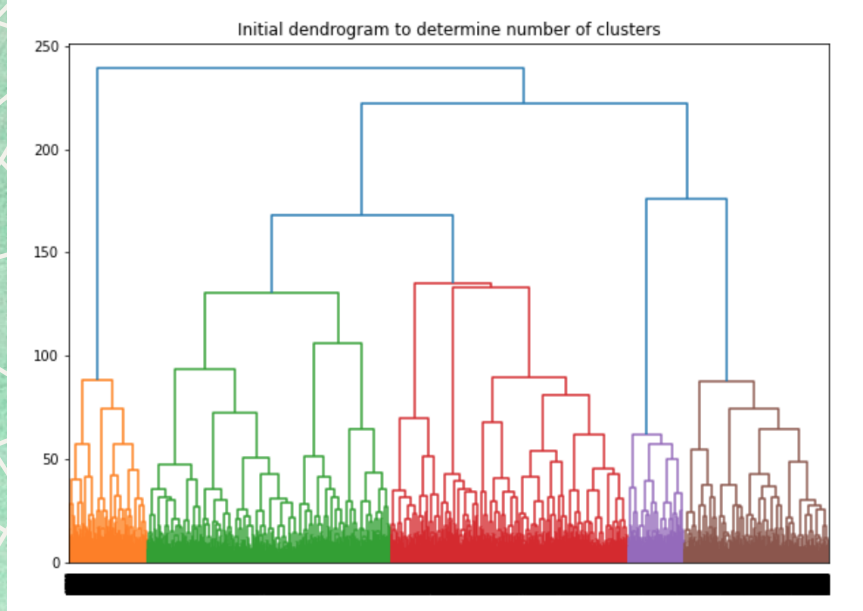
2

Clustering

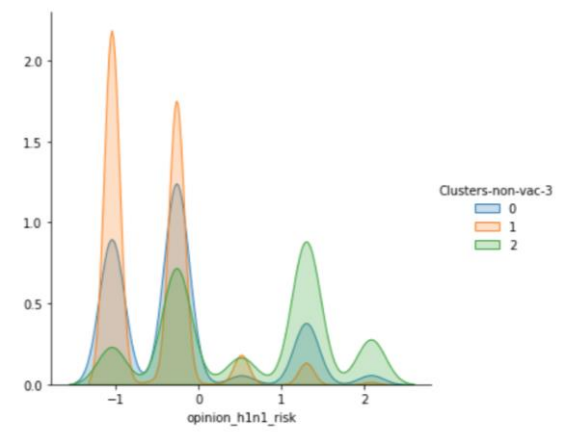
Are there any underlying patterns to tell us WHO got vaccinated or didn't?

KMeans Clustering

- Non-vaccinated set (21,000)
- 19 variables
- Data Imputed and Scaled
  (KNN Imputer; StandardScaler)
- KMeans with 3 groups
  (3 to 5 suggested)

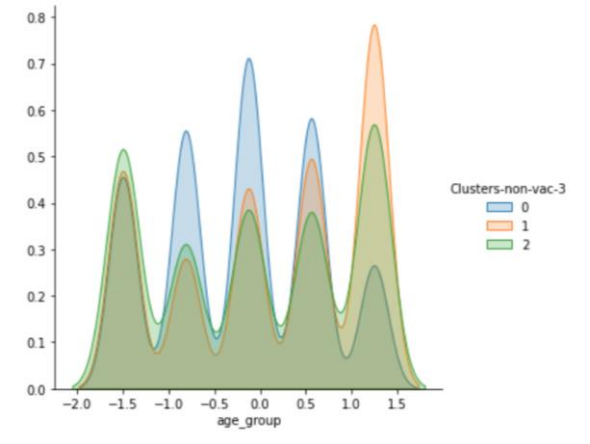


Age, education, opinions, contact



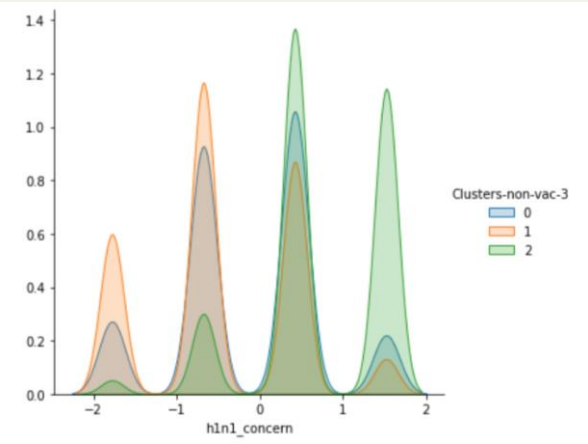
Group 1

Older, less contact, **feel less at risk**, lower doct. recc. vax



Group 0

Younger, more contact, mid-level concern / at risk



Group 2

Higher knowledge, education, income, **higher concern**

NOTE: Similar groupings from those who were vaccinated (5,700)

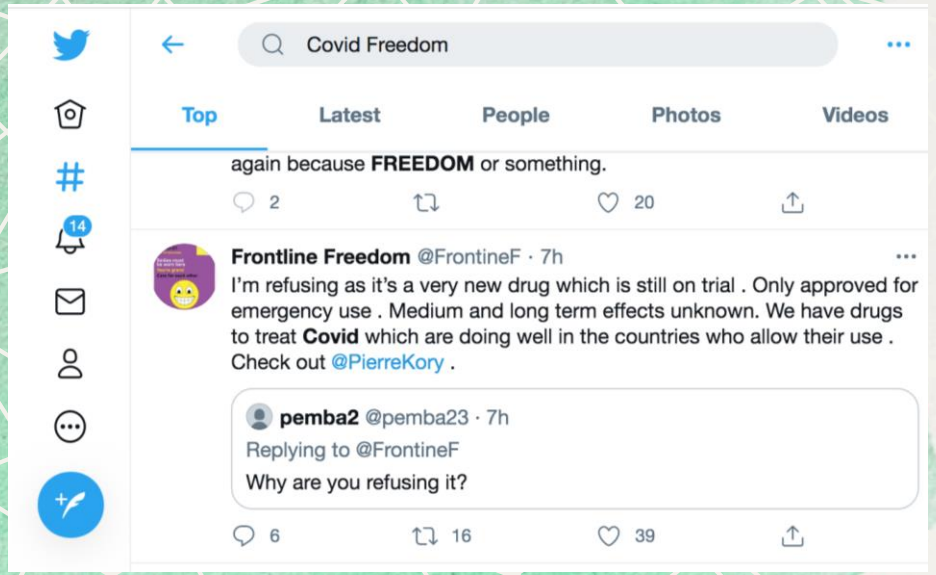
3

NLP

Using sentiment analysis and soft clustering what CONCERNS need to be communicated?

Covid-19 Tweets

- via Twint
  (Feb. 2020 to present; EN; at least 5 likes)
- 2 sets: neutral; negative
  - “covid vaccination” (26K)
  - “covid death”, “vaccine choice”, “vaccine freedom”, etc. (38K)
- TextBlob used to id negative sentiment tweets



Uncovering Topics and Meaning

1. List Top Words
2. View Bi-grams & Tri-grams
3. Topic clustering
4. Read a sample of tweets

1
List Top Words

freedom
refuse
prevent
petition
restrictions
serious
bad
long
term

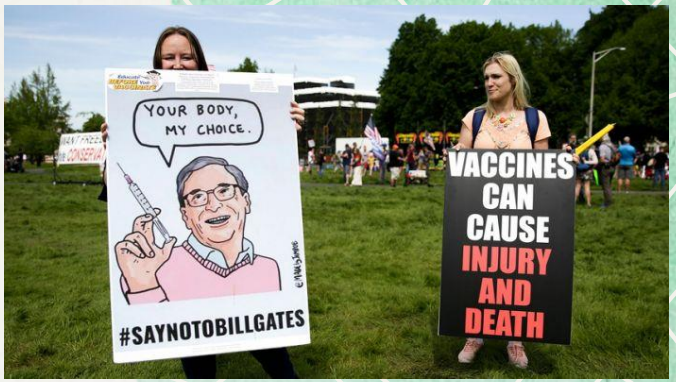
don’t
worst
sick
dangerous
stop
lockdown
trump
pfizer

2
View Bi-grams & Tri-grams

rushed, vaccine
vaccine, trial
cause, death
unnecessary, death
bill, gates
freedom, choice
medical, freedom
vaccine, passport
don’t, want

serious, side, effect
long, term, side
petition, prevent, restrictions
unknown, long, term
within, 28, days

Note: positive set similar but also – getting, received, want; good, better, great; safe, zero side effects, 100% effective, works



3, 4

Topic clustering & reading

Vaccine:

ineffective, rushed, not tested, side effects, deaths

Process:

forced, status tracked, denial of services, lack of choice / freedom

Actions:

sign petition to prevent restrictions; protest – close vacc. site

Political / News:

Bill Gates, WHO, Democrats, government, big pharma, church; microchip, medical exp.

Example Tweets

"...Rushed, experimental, untested vaccine with a tracking chip..."

"I am a healthy 53 yr old and I am not going near that vaccine. You watch, there will be some profound side effects to this thing they rushed through."

"This is AMERICA. We have FREEDOM. Where your fear begins, my rights don't end."

"...Either fight for your freedom or remain passive to the Gov'ts will. I will not take a vaccine that Bill Gates has a hand in and I'm damn sure not going to submit to being microchipped."

"If you have a death wish, take the vaccine."

Take-aways

1

Survey:

Can use to predict if get vaccinated

2

Opinions are key:

A driver in determining if get vaccinated

3

Tweets:

Help us identify concerns and misinformation

4

Feedback Loop:

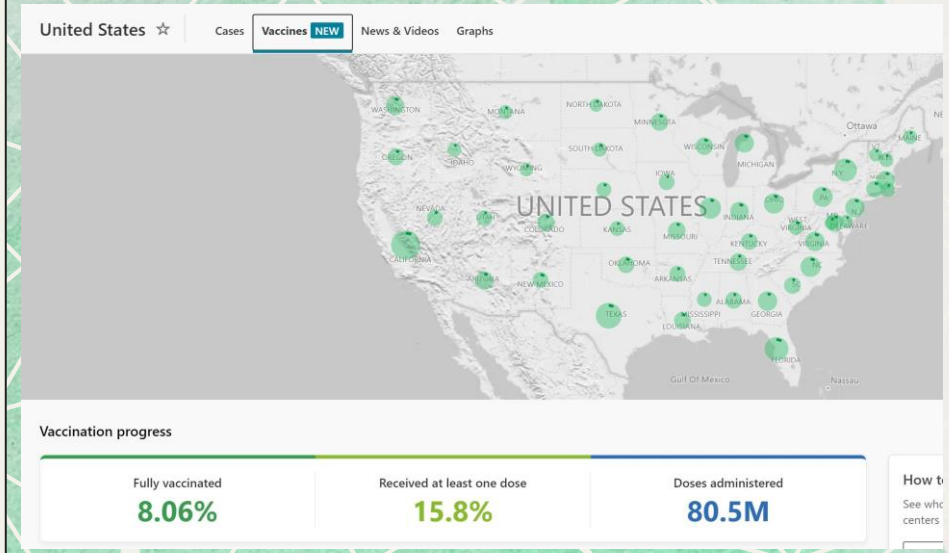
To refine and create future surveys

Recommendations

1. Gain support of local **doctors**; community leaders
2. Communicate the facts; address **misperceptions**
3. Be equitable and target most needed **people / places**
4. Keep alert to **trending topics** of concern (ongoing data collection)
5. **Bridge** federal and state schisms

Next Steps

1. More social media **sources**; better targeted
2. **Recent** survey data; examine change in opinions over time
3. **Geographic** dimension to identify places of need
4. Infographic and **Dashboard** improvements



Dashboard... in progress

Classification Model Results

Select data preperation to see results of 6 models

A-Basic-preprocessing

Classification_model	Accuracy	Precision1
XGBoost	0.85	0.68
Random Forest	0.84	0.68
SVC	0.84	0.67
Logistic Regression	0.84	0.66
KNN	0.81	0.54
Decision Trees	0.75	0.4

Thanks

For details:

Email: markpatterson8@hotmail.com
Github: https://github.com/markp-rankin/H1N1-vaccination-predictions

Thanks to:

Yish, Dara, and Flatiron School
Classmates / Friends

Credits

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution