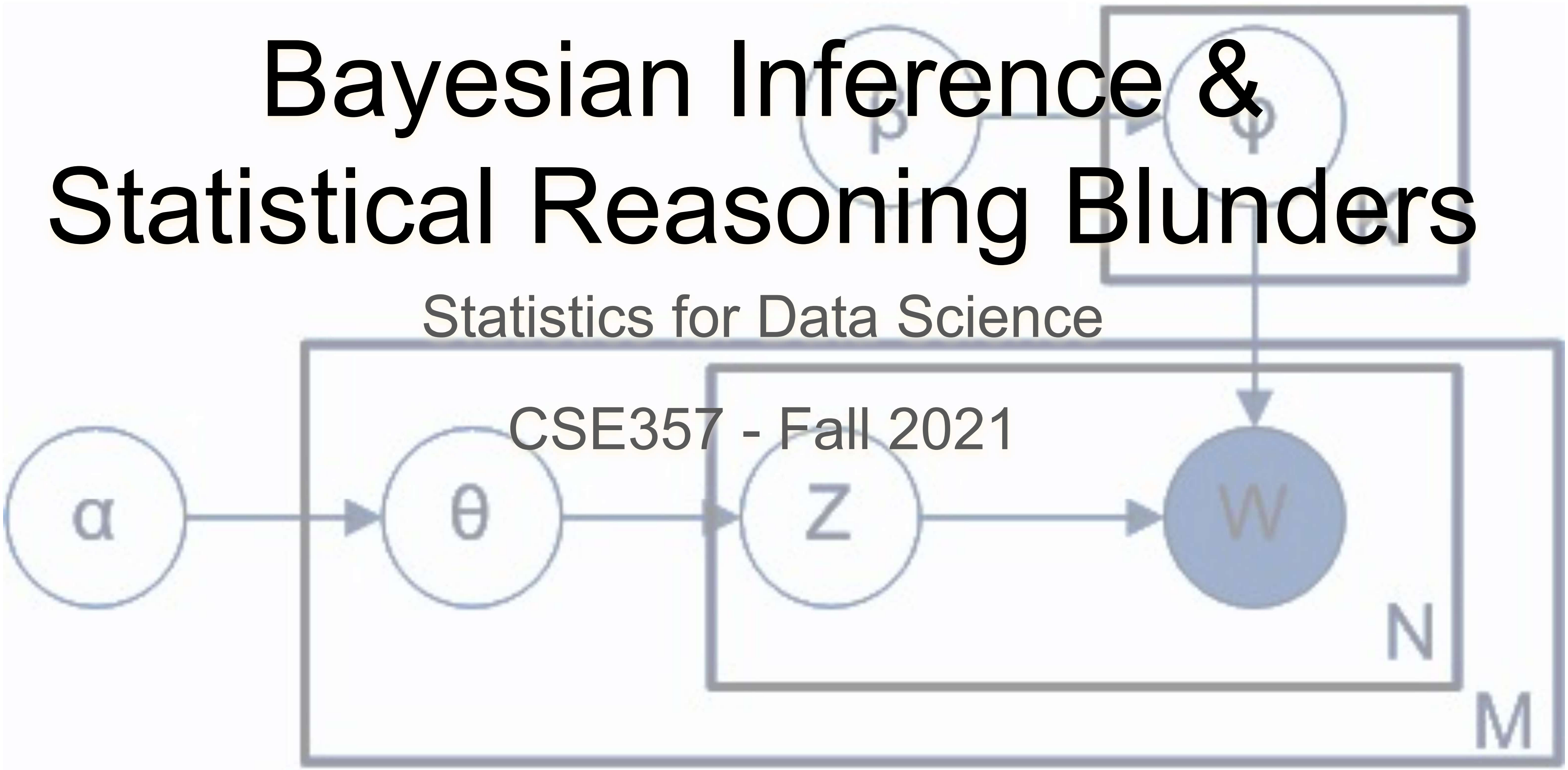
Bayesian Inference & Statistical Reasoning Blunders

Statistics for Data Science

CSE357 - Fall 2021



Bayes Rule

$P(A|B) = (P(B|A)P(A)) / P(B)$

GOAL: Relate $P(A_i|B)$ to $P(B|A_i)$,

for all i = 1 ... k, where $A_1 \dots A_k$ **partition Ω**

Law of Total Probability

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

Let's try:

(1) $P(A_i|B) = P(A_i,B) / P(B)$

(2) $P(A_i,B) / P(B) = P(B|A_i)\, P(A_i) / P(B)$, by multiplication rule

but in practice, we might not know P(B)

(3) $P(B|A_i)\, P(A_i) / P(B) = P(B|A_i)\, P(A_i) / \left(\sum_{i=1}^{k} P(B|A_i)P(A_i) \right)$, by law of total probability

Bayes Rule

$P(A|B) = (P(B|A)P(A)) / P(B)$

GOAL: Relate $P(A_i|B)$ to $P(B|A_i)$,

for all i = 1 … k, where $A_1 \ldots A_k$ **partition** Ω

Law of Total Probability

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

Let's try:

(1) $P(A_i|B) = P(A_i,B) / P(B)$

(2) $P(A_i,B) / P(B) = P(B|A_i)\, P(A_i) / P(B)$, by multiplication rule

but in practice, we might not know P(B)

(3) $P(B|A_i)\, P(A_i) / P(B) = P(B|A_i)\, P(A_i) / \left(\sum_{i=1}^{k} P(B|A_i)P(A_i)\right)$, by law of total probability

Bayes Rule (in practice):

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

Bayes Inference

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(B|A_i)\mathrm{P}(A_i)}{\sum_{i=1}^{k} \mathrm{P}(B|A_i)\mathrm{P}(A_i)}$$

Inference, in general, comes in 3 forms:

1. **Point Estimation** -- finding a "best guess" for the value of a particular parameter. (e.g. mean)

2. **Confidence Sets** -- describing some range (interval) in which a parameter is likely to fall. (e.g. 95% CI of mean; "set": multivariate version of "interval")

3. **Hypothesis Testing** -- Given null hypothesis, does data provide enough evidence to reject? (e.g. t-test for difference in means)

Bayes Rule

likelihood prior

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(B|A_i)\mathrm{P}(A_i)}{\sum_{i=1}^{k} \mathrm{P}(B|A_i)\mathrm{P}(A_i)}$$

posterior marginalization

Bayes Inference

Inference, in general, comes in 3 forms:

1. **Point Estimation** -- finding a "best guess" for the value of a particular parameter. (e.g. mean)

2. **Confidence Sets** -- describing some range (interval) in which a parameter is likely to fall. (e.g. 95% CI of mean; "set": multivariate version of "interval")

3. **Hypothesis Testing** -- Given null hypothesis, does data provide enough evidence to reject? (e.g. t-test for difference in means)

likelihood prior

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

posterior marginalization

Bayes Inference

Inference, in general, comes in 3 forms:

1. **Point Estimation** -- finding a "best guess" for the value of a particular parameter. (e.g. mean)

2. **Confidence Sets** -- describing some range (interval) in which a parameter is likely to fall. (e.g. 95% CI of mean; "set": multivariate version of "interval")

3. **Hypothesis Testing** -- Given null hypothesis, does data provide enough evidence to reject? (e.g. t-test for difference in means)

Consider that A is a parameter for a distribution. Bayesian Inference allows us to have a ***prior*** belief about what A it.

likelihood prior

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

posterior marginalization

Bayes Inference

Point-Estimation:

Point = A discrete variable (i.e. a class)

likelihood prior

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

posterior marginalization

Bayes Inference

Point-Estimation:

Point = A discrete variable (i.e. a class)

likelihood prior

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

posterior marginalization

Bayes classifier: choose the class most likely according to $P(y|X)$. (y is a class label)

Naive Bayes classifier: Assumes all predictors are independent given y.

$$P(Y = y|A = a, B = b, C = c) = p(y|a)p(y|b)p(y|c)$$

Bayes Inference

Point-Estimation:

Point = A discrete variable (i.e. a class)

likelihood prior

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

posterior marginalization

Bayes classifier: choose the class most likely according to $P(y|X)$. (y is a class label)

Naive Bayes classifier: Assumes all predictors are independent given y.

$$P(Y = y|A = a, B = b, C = c) = p(y|a)p(y|b)p(y|c)$$

$$P(y|X) = \prod_{i=1}^{m} P(y|X_i)$$

Bayes Inference

$$P(y|X) = \prod_{i=1}^{m} P(y|X_i)$$

$$P(y|X) \propto P(y, X_1, ..., X_m) \propto P(y) \prod_{i=1}^{m} P(X_i|y)$$

likelihood prior

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

posterior marginalization

Maximum a Posteriori (MAP): Pick the class with the maximum posterior probability:

$$\hat{y} = arg\max_{y} P(y) \prod_{i=1}^{m} P(X_i|y)$$

unnormalized posterior

$P(y \mid X)$

Bayes Inference

$$P(y|X) = \prod_{i=1}^{m} P(y|X_i)$$

$$P(y|X) \propto P(y, X_1, ..., X_m) \propto P(y) \prod_{i=1}^{m} P(X_i|y)$$

Will be constant, so can ignore.

likelihood prior

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)}$$

posterior marginalization

Maximum a Posteriori (MAP): Pick the class with the maximum posterior probability:

$$\hat{y} = arg \max_{y} P(y) \prod_{i=1}^{m} P(X_i|y)$$

unnormalized posterior

$P(y \mid X)$

Gaussian Naive Bayes

Maximum a Posteriori (MAP): Pick the class with the maximum posterior probability:

$$\hat{y} = arg\max_{y} \boxed{\mathrm{P}(y) \prod_{i=1}^{m} \mathrm{P}(X_i|y)}$$

unnormalized posterior

$\mathrm{P}(y \mid X)$

Assume $\mathrm{P}(\mathrm{X}|\mathrm{Y})$ is *Normal* (*Y* is still discrete)

Gaussian Naive Bayes

Maximum a Posteriori (MAP): Pick the class with the maximum posterior probability:

$$\hat{y} = arg\max_y \mathrm{P}(y) \prod_{i=1}^{m} \mathrm{P}(X_i|y)$$

unnormalized posterior

$P(y \mid X)$

Assume $P(X|Y)$ is *Normal* (*Y* is still discrete)

Then, to "train" the model, need $P(y)$, $P(X|Y)$ (i.e. find values for the Gaussian X, given y):

1. Estimate the prior: $P(Y = k)$;

   π_k = count(Y = k) / Count(Y = *)

Gaussian Naive Bayes

Maximum a Posteriori (MAP): Pick the class with the maximum posterior probability:

$$\hat{y} = arg\max_{y} P(y) \prod_{i=1}^{m} P(X_i|y)$$

unnormalized posterior

$P(y \mid X)$

Assume $P(X|Y)$ is *Normal* (*Y* is still discrete)

Then, to "train" the model, need $P(y)$, $P(X|Y)$ (i.e. find values for the Gaussian X, given y):

1. Estimate the prior: $P(Y = k)$;

   $\pi_k = count(Y = k) / Count(Y = *)$
2. Use MLE to find parameters parameters of likelihood $P(X|Y)$:

   μ, σ, for each class of Y. (the "class conditional distribution")

Gaussian Naive Bayes

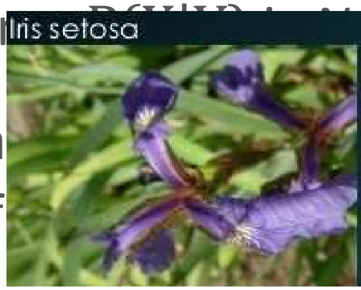
a … AP): Pick the class … axi … probability:

$$\hat{y} = arg \max_y \mathrm{P}(y) \prod_{i=1}^{m} \mathrm{P}(X_i|y)$$

unnormalized posterior

$P(y \mid X)$

Assu … *ormal* (*Y* is still disc…



Then … nodel, need P …
(i.e. f … e Gaussian X, g …).

1. Estimate the prior: P(Y = k);

   π_k = count(Y = k) / Count(Y = *)

2. Use MLE to find parameters parameters of likelihood P(X|Y):

   μ, σ, for each class of Y. (the “class conditional distribution”)

https://colab.research.google.com/drive/1nXAl95cGL7PKzrNVxA96EPV91IlIPzyr

Case Study: COVID-19

Statistical Reasoning Blunders

Case Study: COVID-19

- Base rate fallacy: effect on false-positive and false-negatives
- Availability heuristic: Effect on evaluation of people/groups
- Causal Confounds: Understanding case/death rates
- Measurement Bias/Confounds: Relation of Twitter to Real World

Base Rate Fallacy

Effect on false-positive and false-negatives

Sensitivity: |True Positives| / |Actual Positives| :   P(test=True | disease=True)

Specificity: |True Negatives| / |Actual Negatives|: P(test=False | disease=False)

(plot source: NYT U.S. Coronavirus Data)

Base Rate Fallacy

Effect on false-positive and false-negatives

The test has a true positive

(plot source: NYT U.S. Coronavirus Data)

Base Rate Fallacy

Effect on false-positive and false-negatives

Sensitivity: |True Positives| / |Actual Positives| :   P(test=True | disease=True)

Specificity: |True Negatives| / |Actual Negatives|: P(test=False | disease=False)

P(disease = True | test = True)

(plot source: NYT U.S. Coronavirus Data)

Base Rate Fallacy

Effect on false-positive and false-negatives

Sensitivity: |True Positives| / |Actual Positives| :   P(test=True | disease=True)

Specificity: |True Negatives| / |Actual Negatives|: P(test=False | disease=False)

Precision (PPV): |True Positives| / |Predicted Positive|: P(disease = True | test = True)

(plot source: NYT U.S. Coronavirus Data)

Base Rate Fallacy

Effect on false-positive and false-negatives

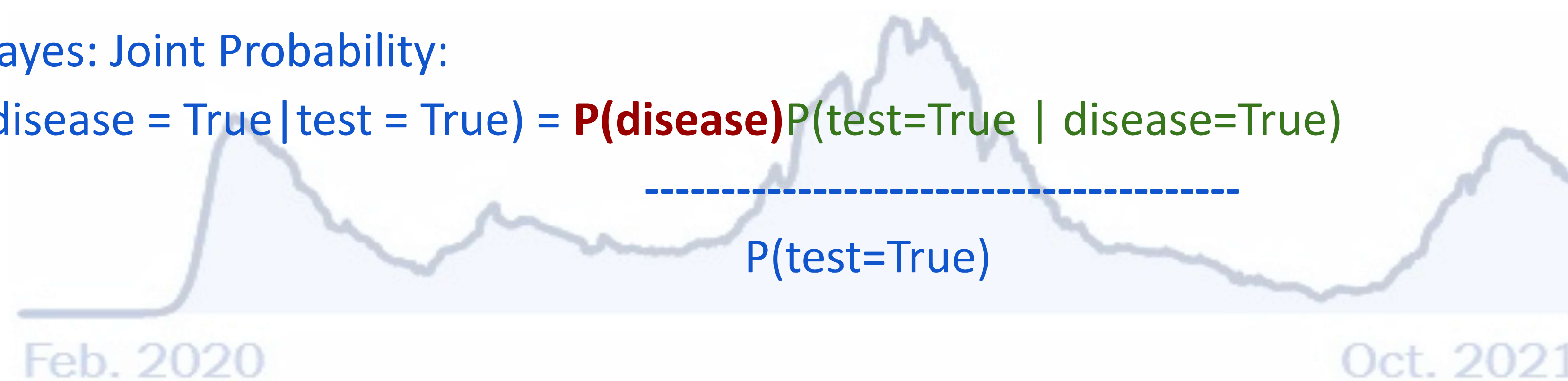
Sensitivity: |True Positives| / |Actual Positives| : P(test=True | disease=True)

Specificity: |True Negatives| / |Actual Negatives|: P(test=False | disease=False)

Precision (PPV): |True Positives| / |Predicted Positive|: P(disease = True | test = True)

Using Bayes: Joint Probability:

$$P(\text{disease} = \text{True} \mid \text{test} = \text{True}) = \frac{\mathbf{P(disease)}\,P(\text{test=True} \mid \text{disease=True})}{P(\text{test=True})}$$



(plot source: NYT U.S. Coronavirus Data)

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

1. Linda is a bank teller.
2. Linda is a bank teller and vaccinated.

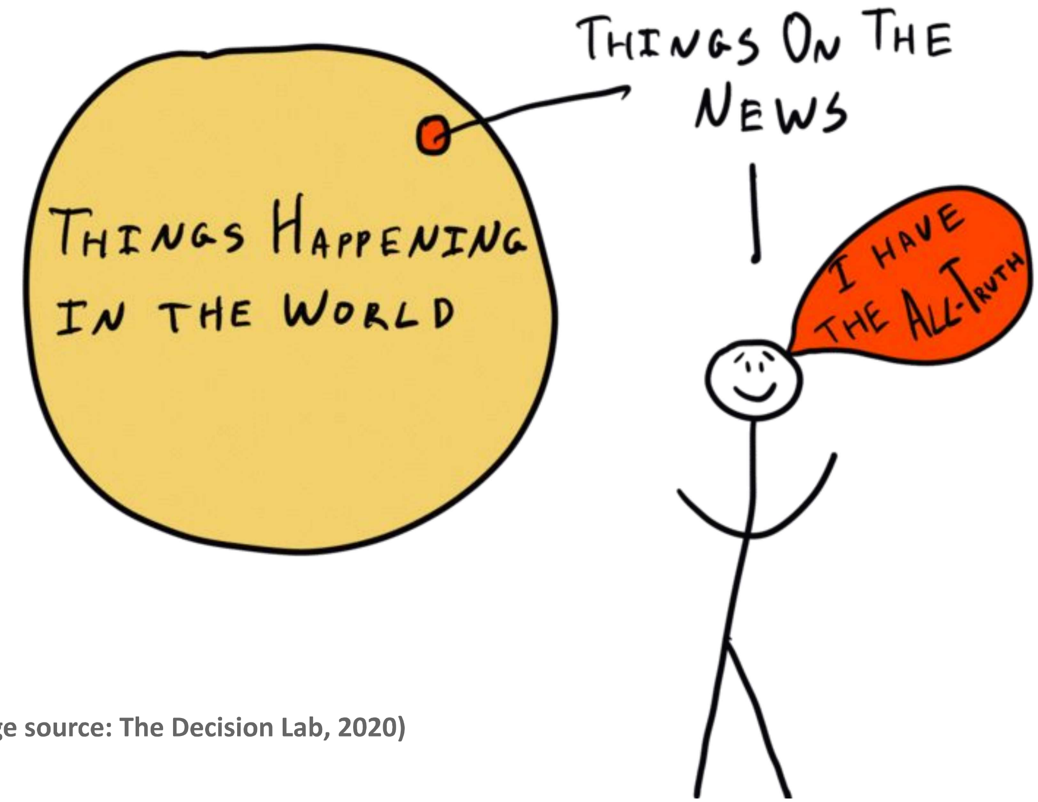
Availability heuristic

Effect on evaluation of people/groups

(image source: The Decision Lab, 2020)

Availability heuristic

Effect on evaluation of people/groups



(image source: The Decision Lab, 2020)

Availability heuristic

Effect on evaluation of people/groups

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

Causal Confounds

Understanding case/death rates

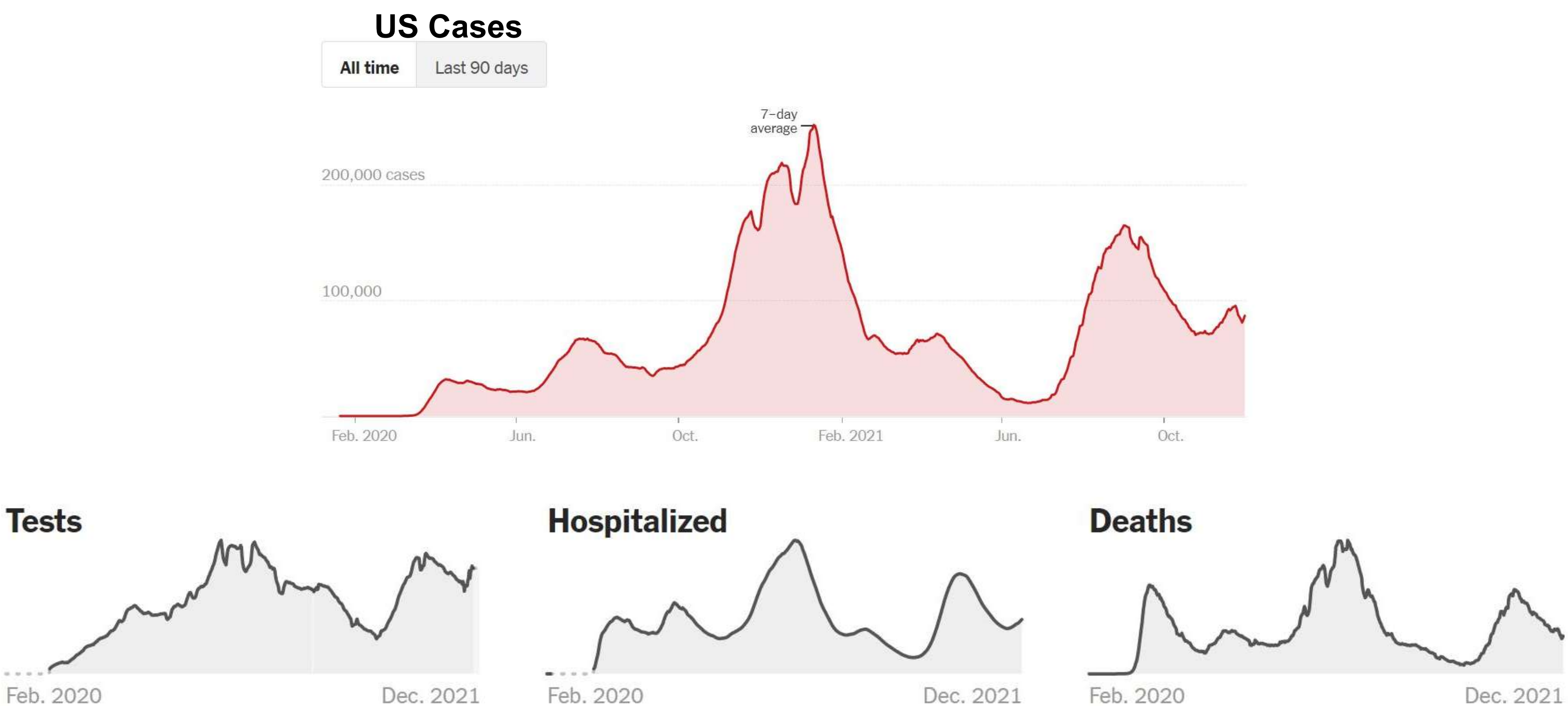
Case Rates:

https://www.nytimes.com/interactive/2021/us/covid-cases.html

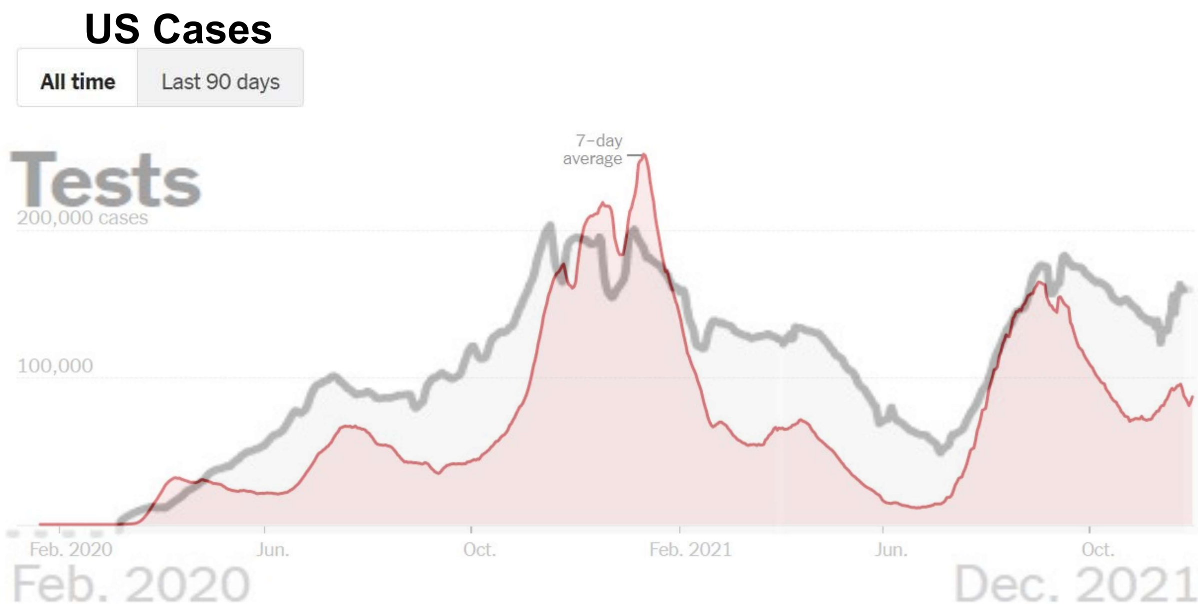
Role of Excess Deaths:

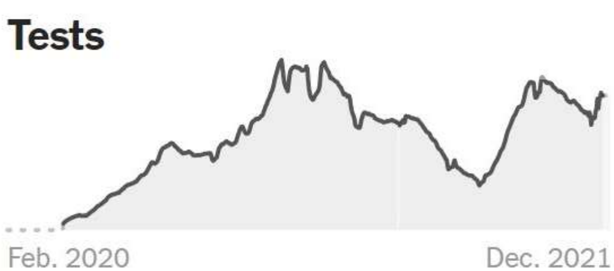
https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker

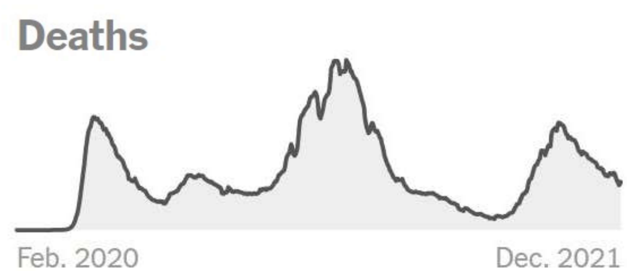
Causal Confounds



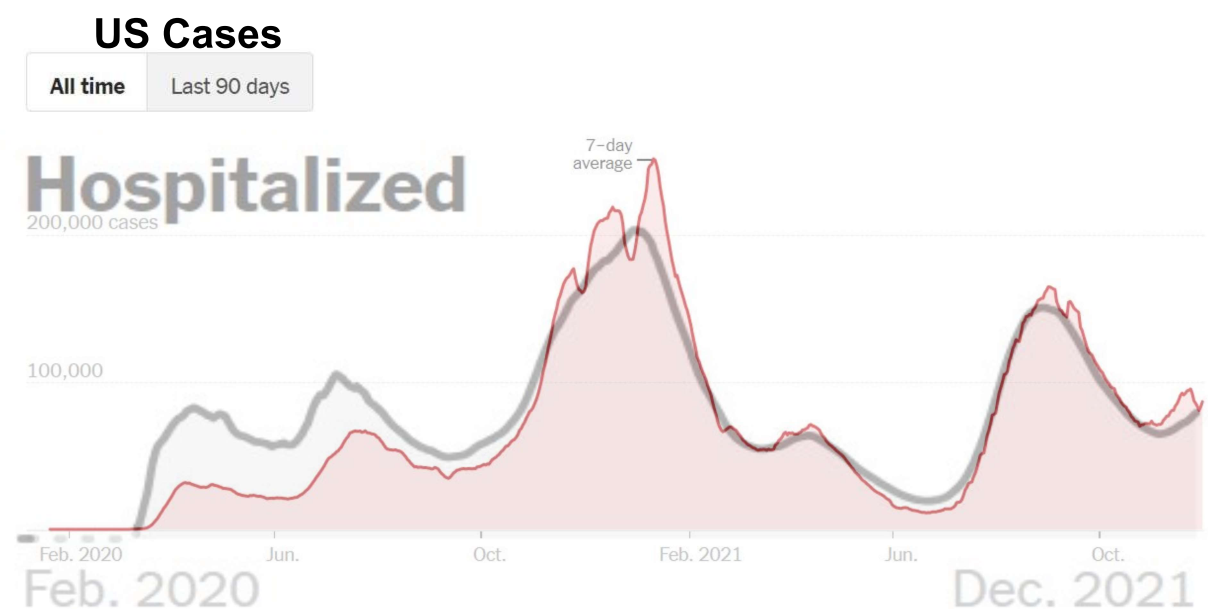
Causal Confounds

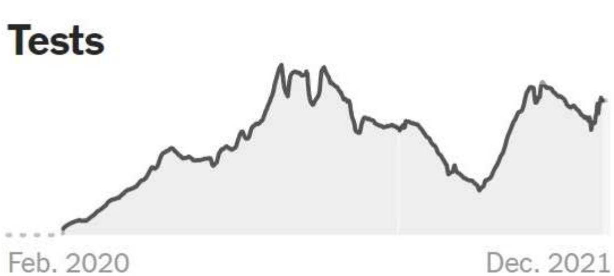


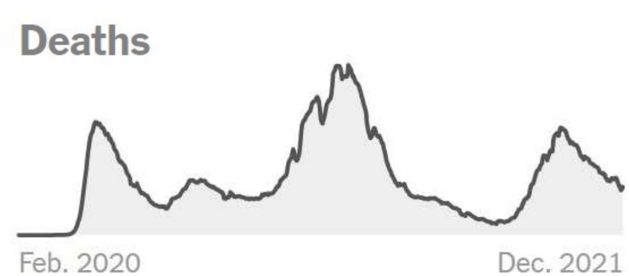




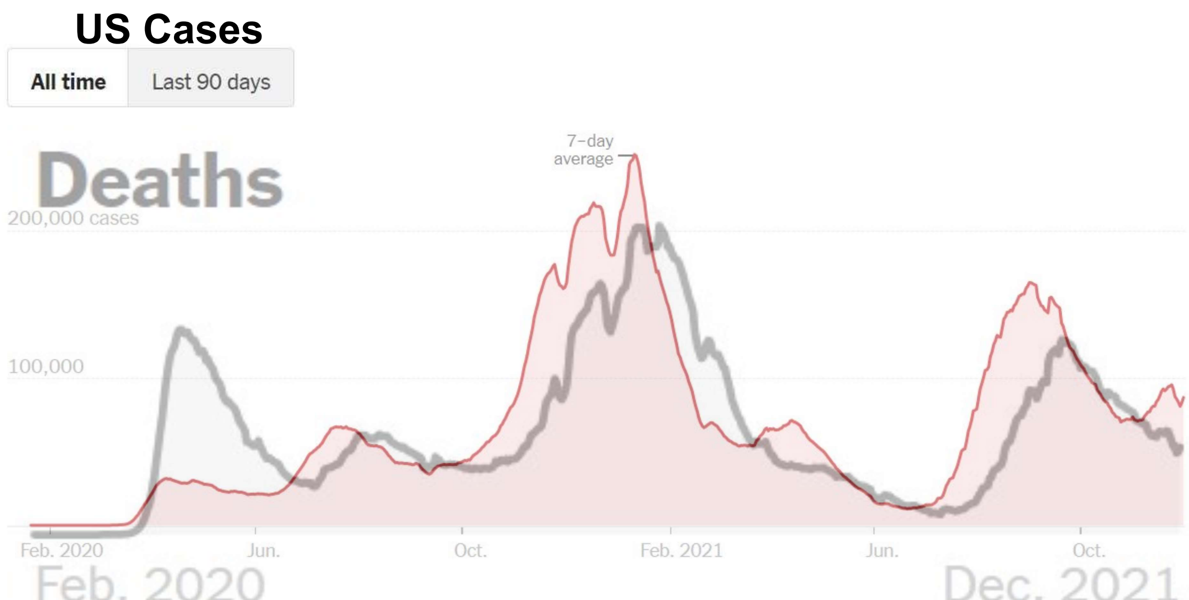
Causal Confounds

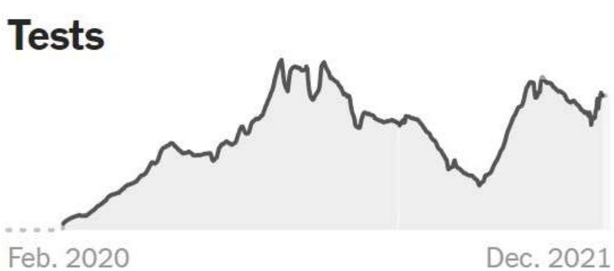


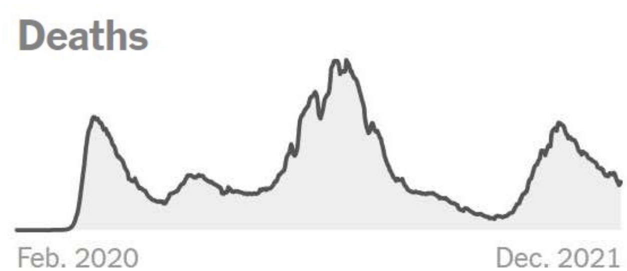




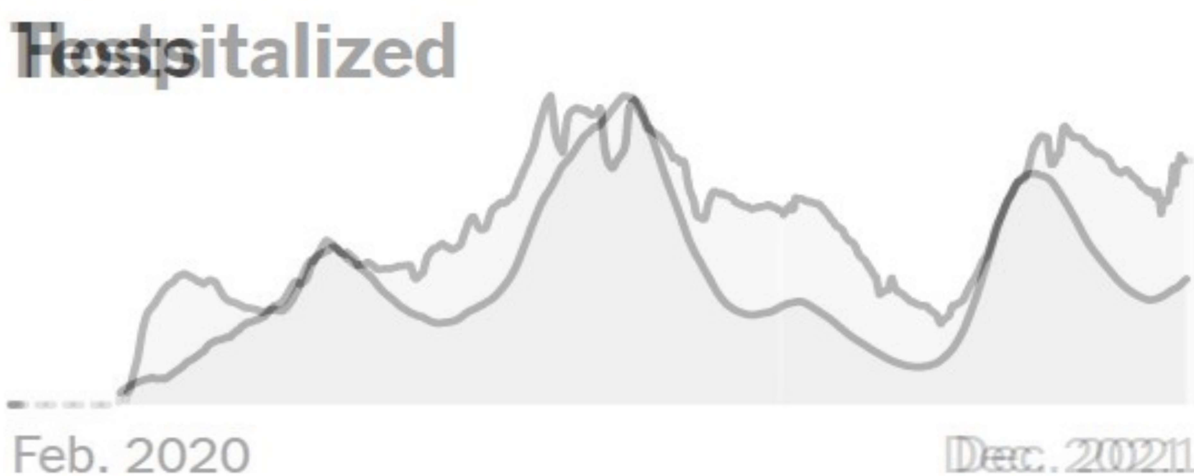
Causal Confounds

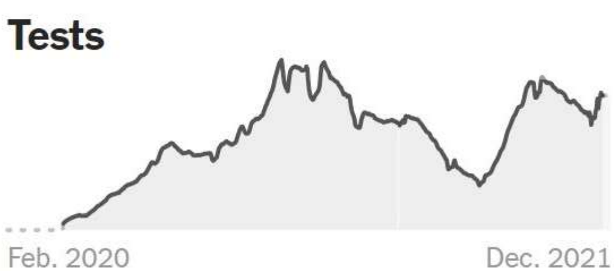


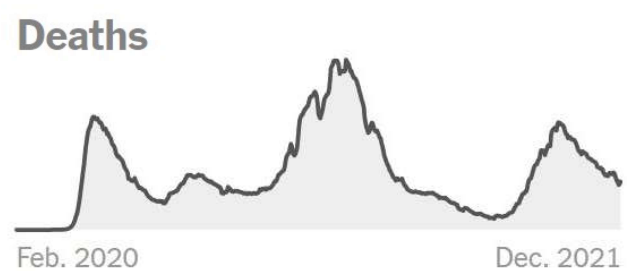




Causal Confounds







Causal Confounds

Understanding case/death rates

Case Rates:

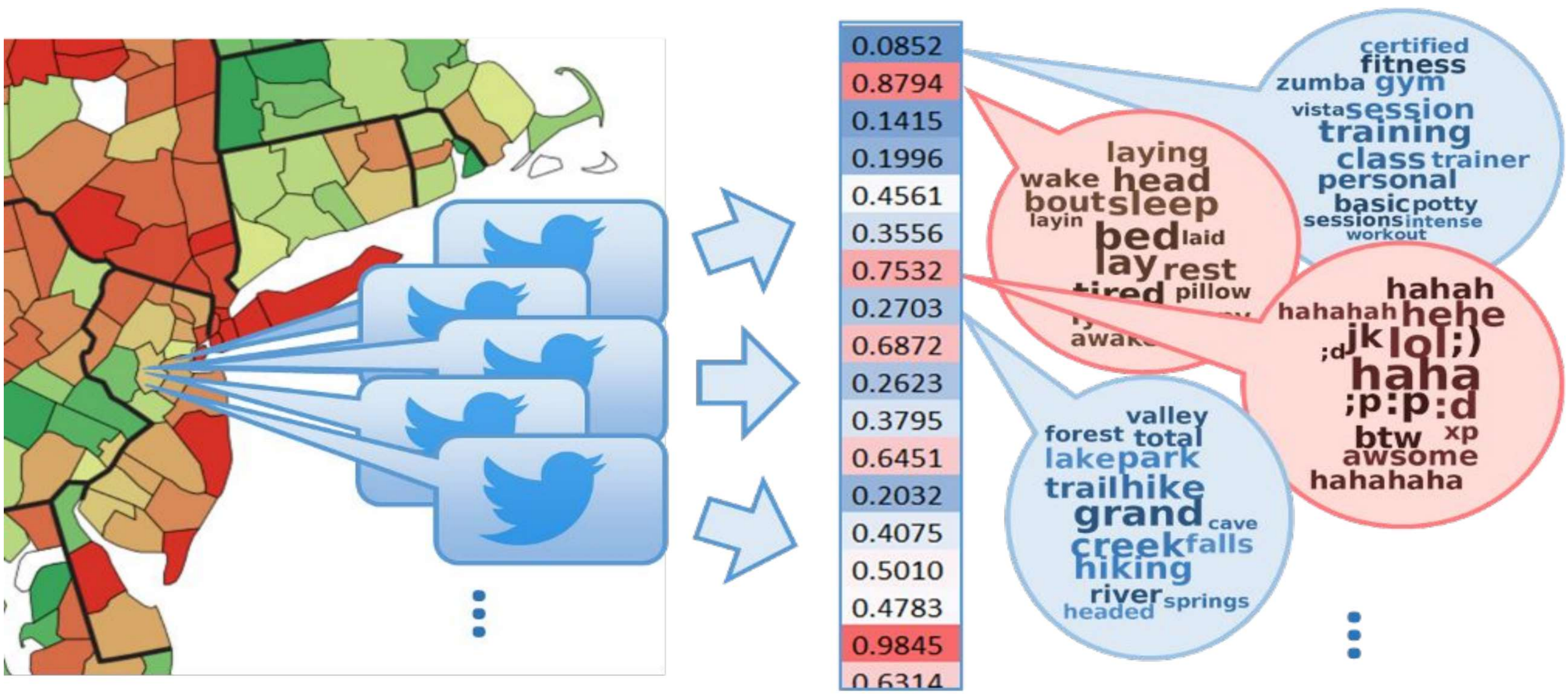
https://www.nytimes.com/interactive/2021/us/covid-cases.html

Role of Excess Deaths:

https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker

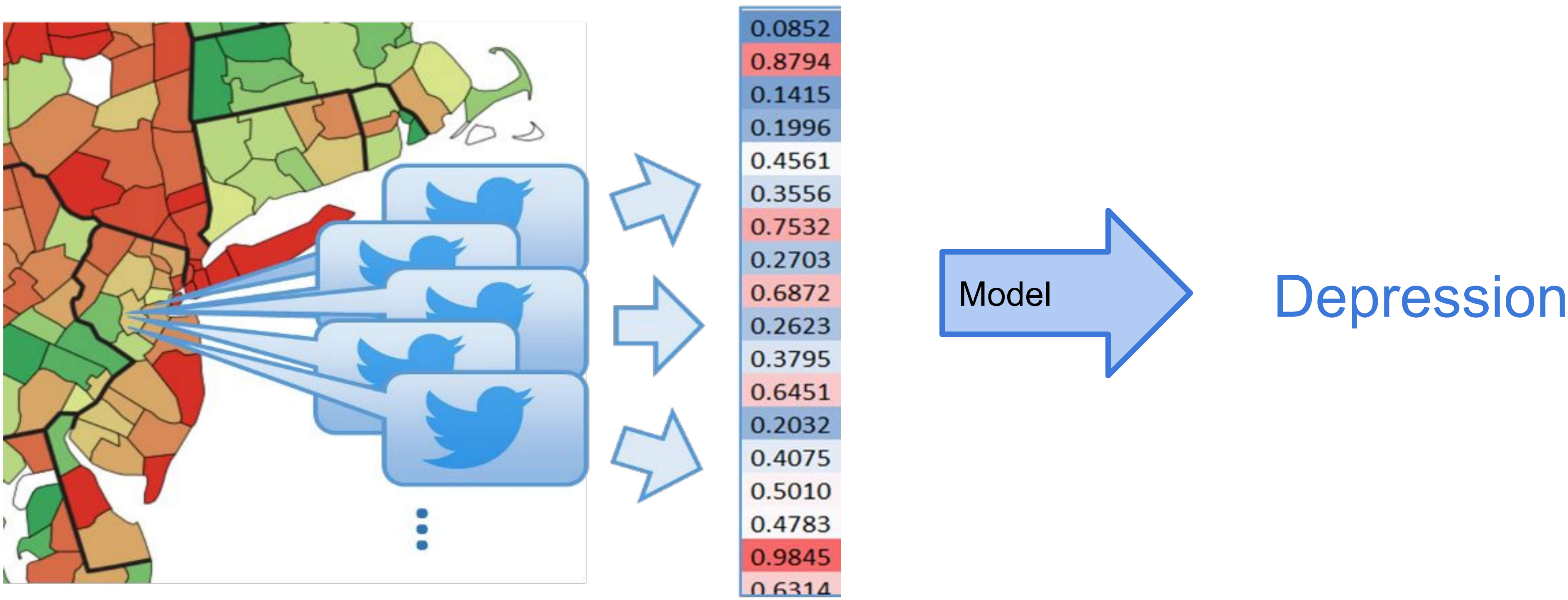
Measurement Bias/Confounds

Twitter assessment of community well-being



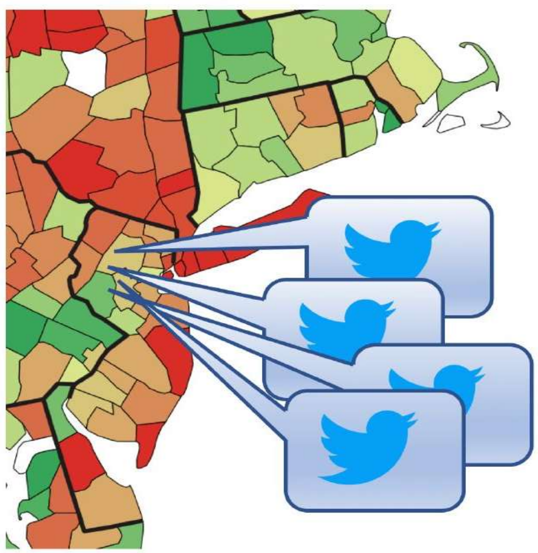
Measurement Bias/Confounds

Twitter assessment of community well-being



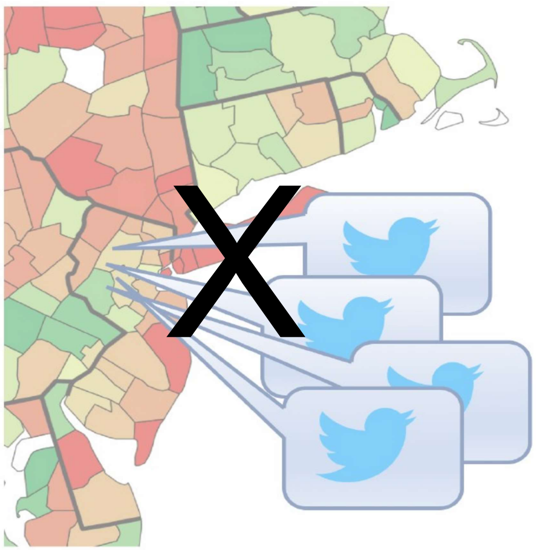
Measurement Bias/Confounds

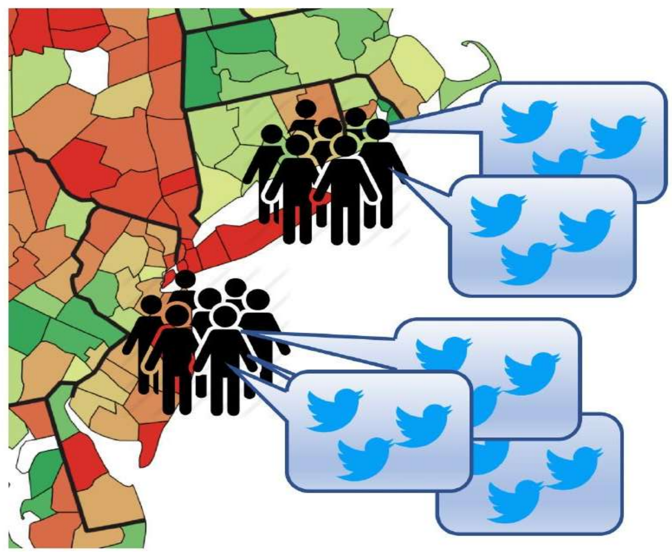
(Giorgi, Preotiuc-Petro, Buffone, Reiman, Ungar, Schwartz, 2018)



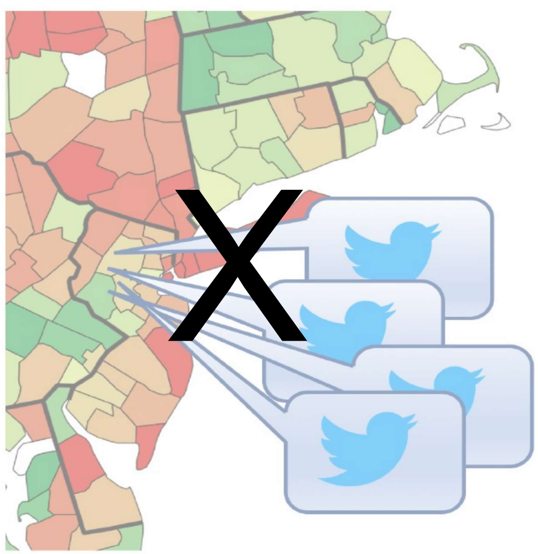
Measurement Bias/Confounds

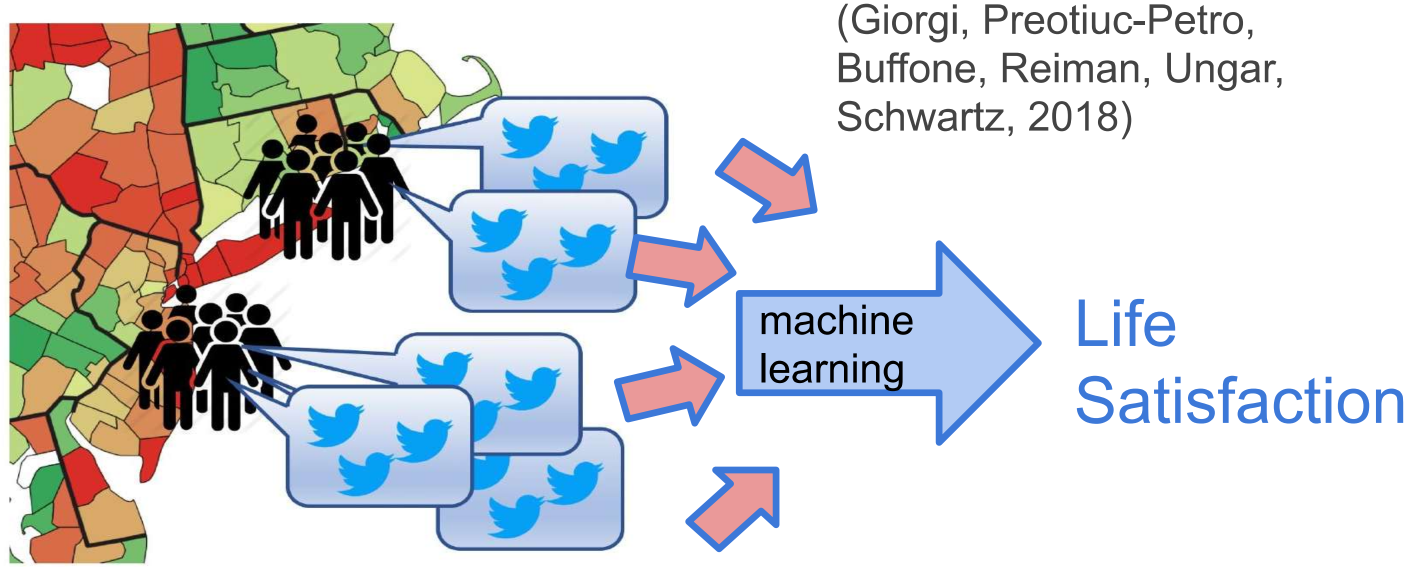
(Giorgi, Preotiuc-Petro, Buffone, Reiman, Ungar, Schwartz, 2018)





Measurement Bias/Confounds

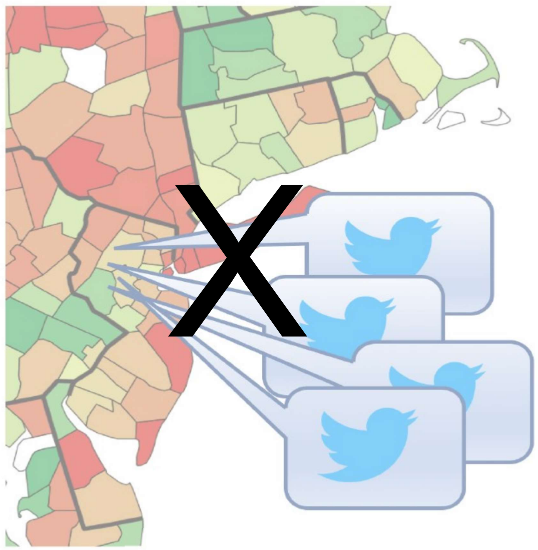


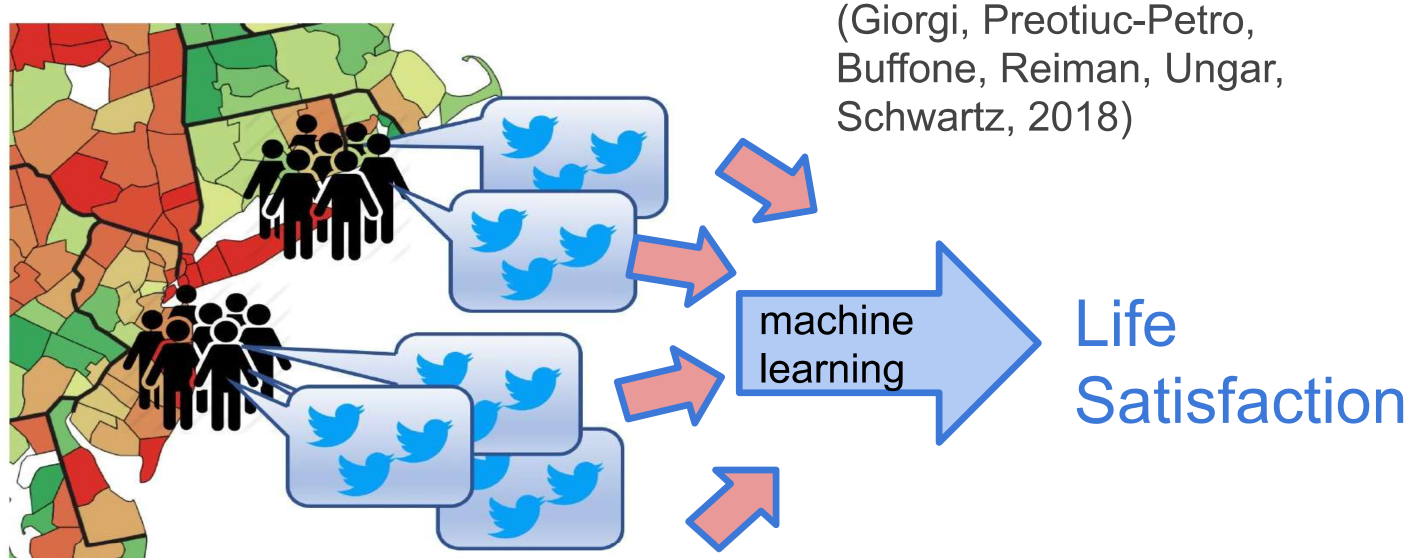


(Giorgi, Preotiuc-Petro, Buffone, Reiman, Ungar, Schwartz, 2018)

***

Measurement Bias/Confounds





(Giorgi, Preotiuc-Petro, Buffone, Reiman, Ungar, Schwartz, 2018)

N = *2040* US counties