

# Big Data and Social Scientific Applications

Stony Brook University  
CSE545, Fall 2016

# Why Social Scientific Applications?

Applications that make a difference in the world.

Often public data available.

Experience working toward an objective and/or using data to answer questions.





# SUSTAINABLE DEVELOPMENT GOALS

**1** NO POVERTY

**2** ZERO HUNGER

**3** GOOD HEALTH AND WELL-BEING

**4** QUALITY EDUCATION

**5** GENDER EQUALITY

**6** CLEAN WATER AND SANITATION

**7** AFFORDABLE AND CLEAN ENERGY

**8** DECENT WORK AND ECONOMIC GROWTH

**9** INDUSTRY, INNOVATION AND INFRASTRUCTURE

**10** REDUCED INEQUALITIES

**11** SUSTAINABLE CITIES AND COMMUNITIES

**12** RESPONSIBLE CONSUMPTION AND PRODUCTION

**13** CLIMATE ACTION

**14** LIFE BELOW WATER

**15** LIFE ON LAND

**16** PEACE, JUSTICE AND STRONG INSTITUTIONS

**17** PARTNERSHIPS FOR THE GOALS

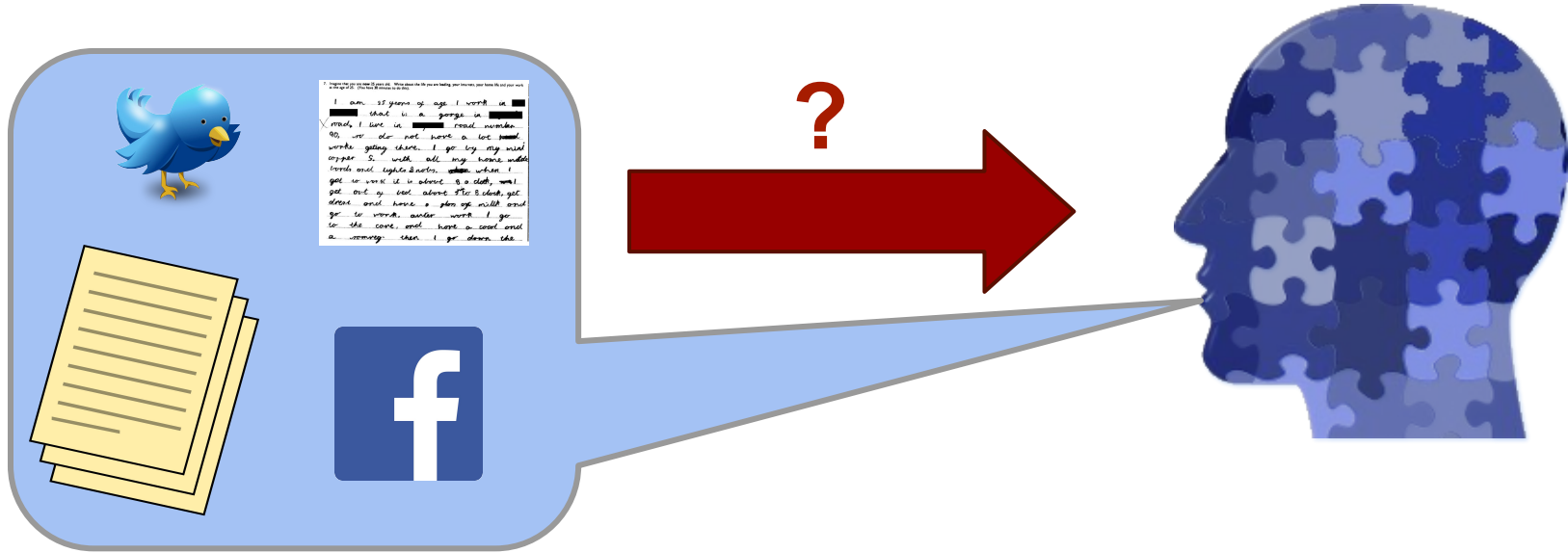
**SUSTAINABLE DEVELOPMENT GOALS**

# SDG Related Data Sets

| Name               | Desc.  | Metadata                          | Link   | SDG Goal |
|--------------------|--|-----------------------------------|--|----------|
| Air Pollution Data | Per country air pollution  | 2977 Towns                        | Multiple Sources: <a href="https://ourworldindata.org/air-pollution">https://ourworldindata.org/air-pollution</a><br><a href="http://www.who.int/phe/health_topics/outdoorair/databases/AAP_database_summary_results_2016_v02.pdf">http://www.who.int/phe/health_topics/outdoorair/databases/AAP_database_summary_results_2016_v02.pdf</a> | SDG 3    |
| GHO Data           | Health indicators including, mortality, burden of disease, life expectancy     |                                   | <a href="https://www.who.int/gho/mortality_burden_disease/en/">https://www.who.int/gho/mortality_burden_disease/en/</a>  | SDG 3    |
| Child Mortality    | Child mortality rate   | 210 countries                     | Multiple Sources: <a href="https://ourworldindata.org/child-mortality#data-quality-definition">https://ourworldindata.org/child-mortality#data-quality-definition</a>  | SDG 3    |
| Poverty and Equity | Helps in understanding the evolution of poverty in various countries           | 5394 rows, 31 poverty indicators  | <a href="http://povertydata.worldbank.org/poverty/home/">http://povertydata.worldbank.org/poverty/home/</a>  | SDG 1    |
| Water Quality      | Ensure availability and sustainable management of water and sanitation for all | 550246 rows, 5 quality parameters | <a href="https://data.gov.in/catalog/water-quality-affected-habitations">https://data.gov.in/catalog/water-quality-affected-habitations</a>  | SDG 6    |
| Access to water    |  | 181 countries                     | <a href="https://data.unicef.org/topic/water-and-sanitation/drinking-water/">https://data.unicef.org/topic/water-and-sanitation/drinking-water/</a>  | SDG 6    |
| WDI data           | World development indicators   | 84000 rows, 300 features          | <a href="https://data.world/worldbank/sustainable-development-goals/workspace/file?filename=SDG_csv_en%2FSDG_Data.csv">https://data.world/worldbank/sustainable-development-goals/workspace/file?filename=SDG_csv_en%2FSDG_Data.csv</a>  | SDG 1    |
| Quality Education  | Quality of education in countries  | 291 countries 1600+ features      | <a href="http://data.uis.unesco.org/">http://data.uis.unesco.org/</a>  | SDG 4    |
| Depression Data    | Various mental health factors in USA   | 1986-2016, 5486 features          | <a href="https://qssdataexplorer.norc.org/variables/vfilter">https://qssdataexplorer.norc.org/variables/vfilter</a>  | SDG 3    |

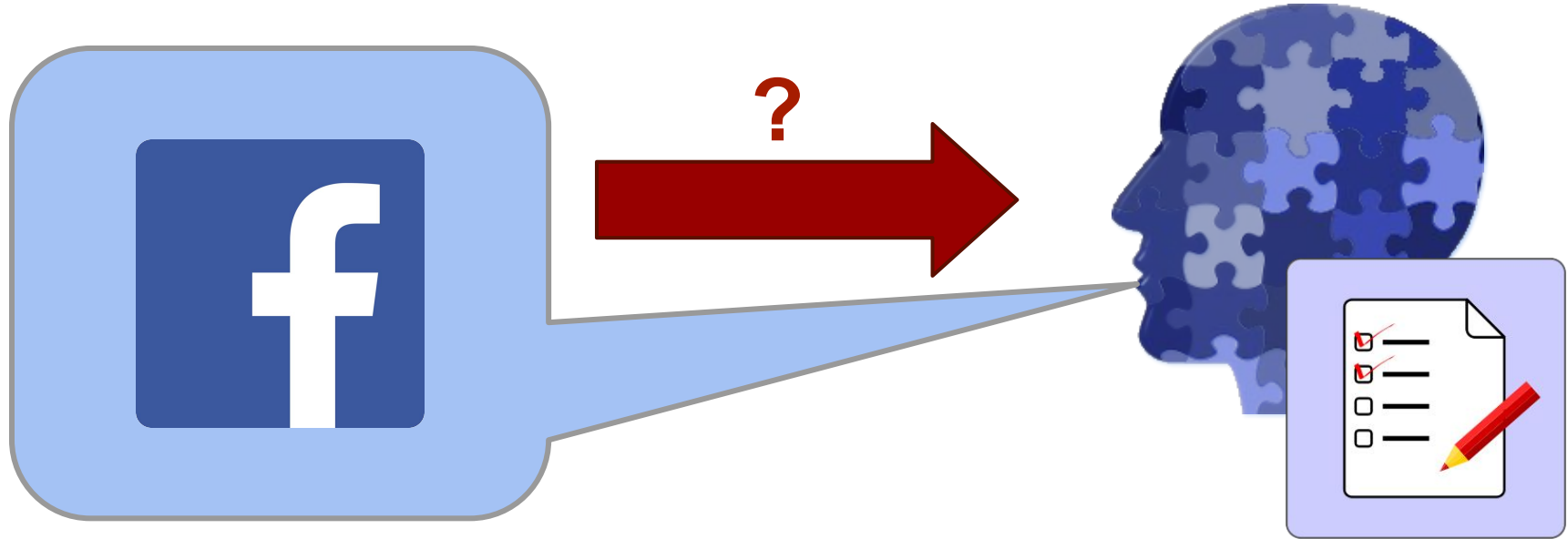
# A (Biased) Sample of Computational Psychology

# Language Says A Lot About People



**Does language use reflect who we are?**

# Language Says A Lot About People



**19M Facebook posts**

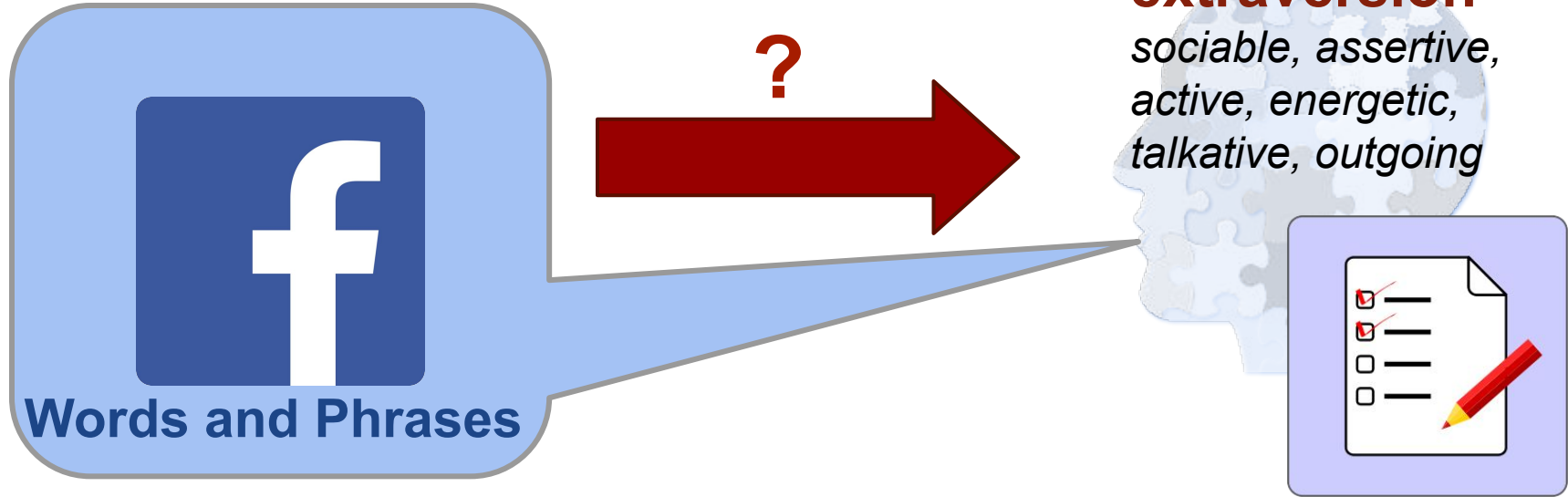
**75,000 personality surveys**

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). **Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach.** *In PLOS ONE 8(9).*

**Does language use reflect who we are?**



# Language Says A Lot About People



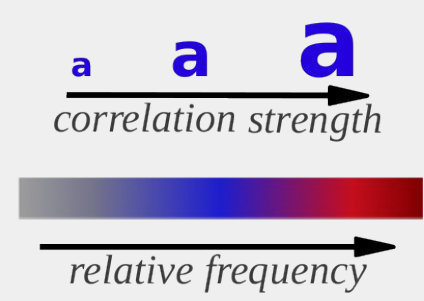
**19M Facebook posts**

**75,000 personality surveys**

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). **Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach.** *In PLOS ONE 8(9).*

**Does language use reflect who we are?**

# extraversion







# Language Says A Lot About People

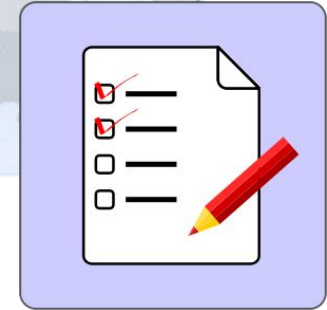
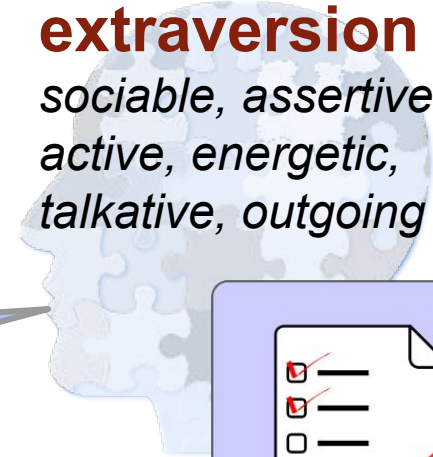


19M Facebook posts

**insights**



**extraversion** --  
*sociable, assertive,  
active, energetic,  
talkative, outgoing*



75,000 personality surveys

# Language Says A Lot About People

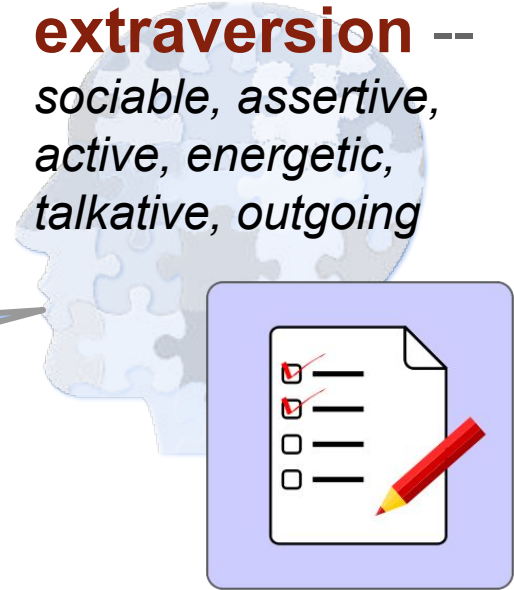


19M Facebook posts

Predict?



**extraversion** --  
*sociable, assertive,  
active, energetic,  
talkative, outgoing*



75,000 personality surveys

***“Language-based Assessments”***

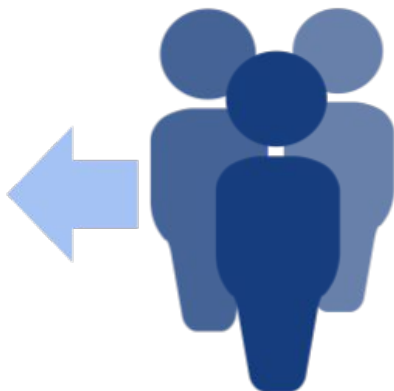
## Language use patterns

*I am **blessed** to spend so much **time** with my **family**.*

*Need **some help**!*

...

Research  
Participants



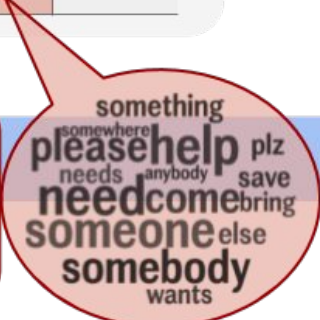
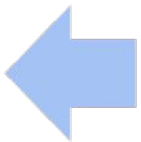
Language use patterns

*I am **blessed** to spend so much **time** with my **family**.*

*Need **some help**!*

...

Research  
Participants





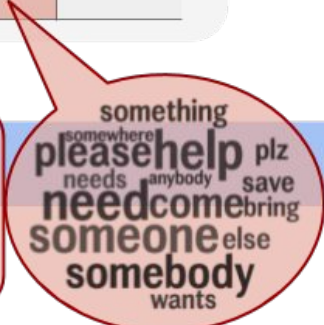
Language use patterns

*I am **blessed** to spend so much **time** with my **family**.*

*Need **some help**!*

...

Research  
Participants



Language-based  
Assessments

- regression
- classification
- deep learning

Language use patterns

*I am **blessed** to spend so much **time** with my **family**.*

*Need **some help**!*

...

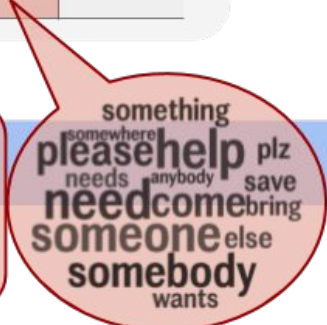
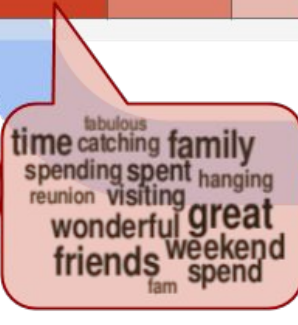
Research  
Participants



States and Traits

*affective valence*  
*depression*  
*anxiety*      *personality*  
*mood*

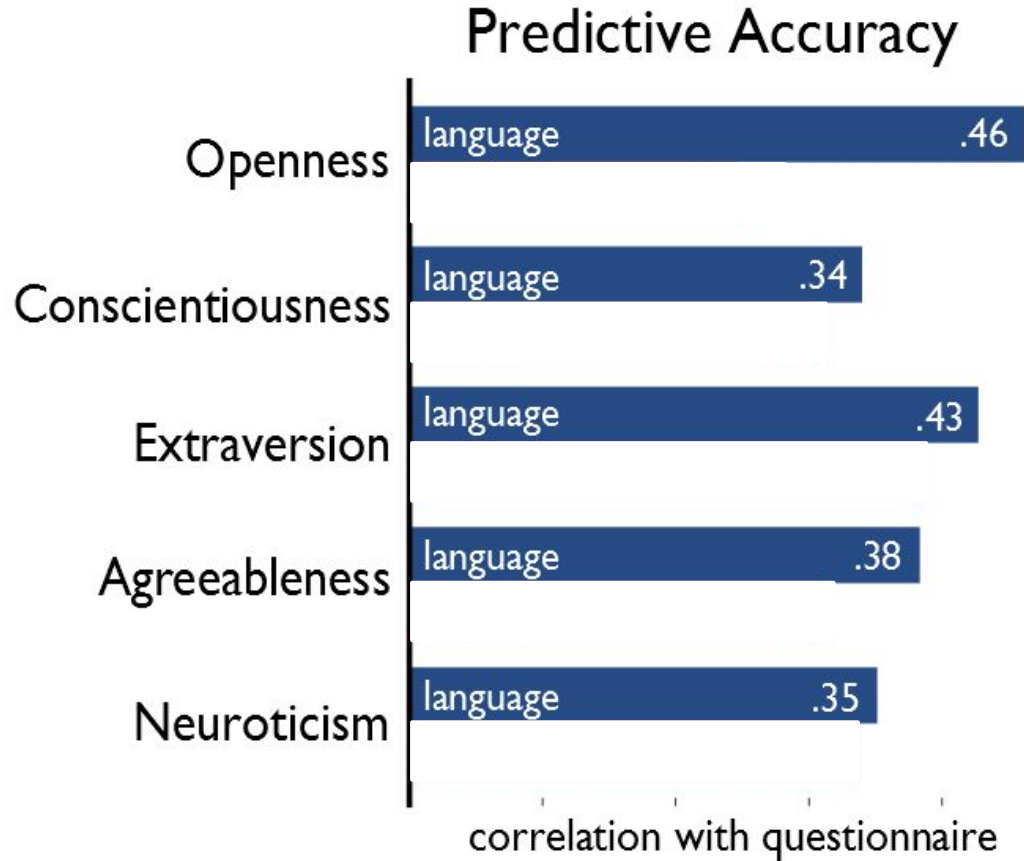
...



Language-based Assessments

- regression
- classification
- deep learning

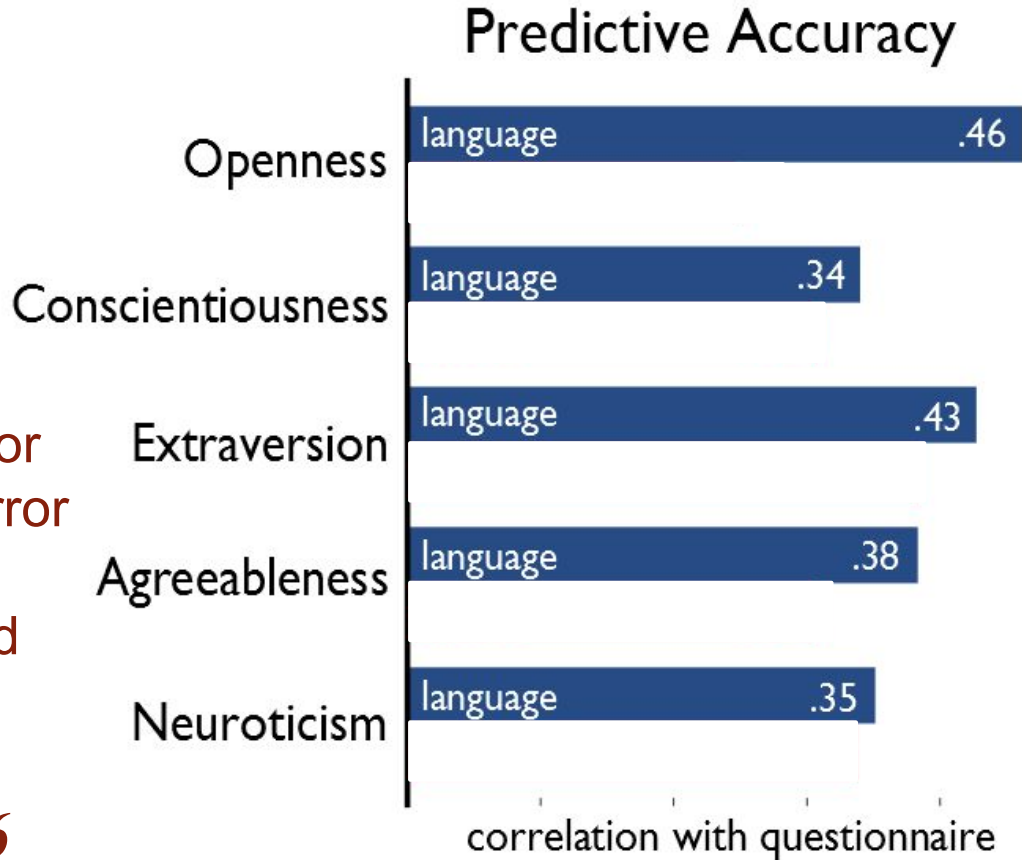
# Language-Based Assessment Evaluation



# Language-Based Assessment Evaluation

disattenuated for  
measurement error  
and  
shared method  
variance:

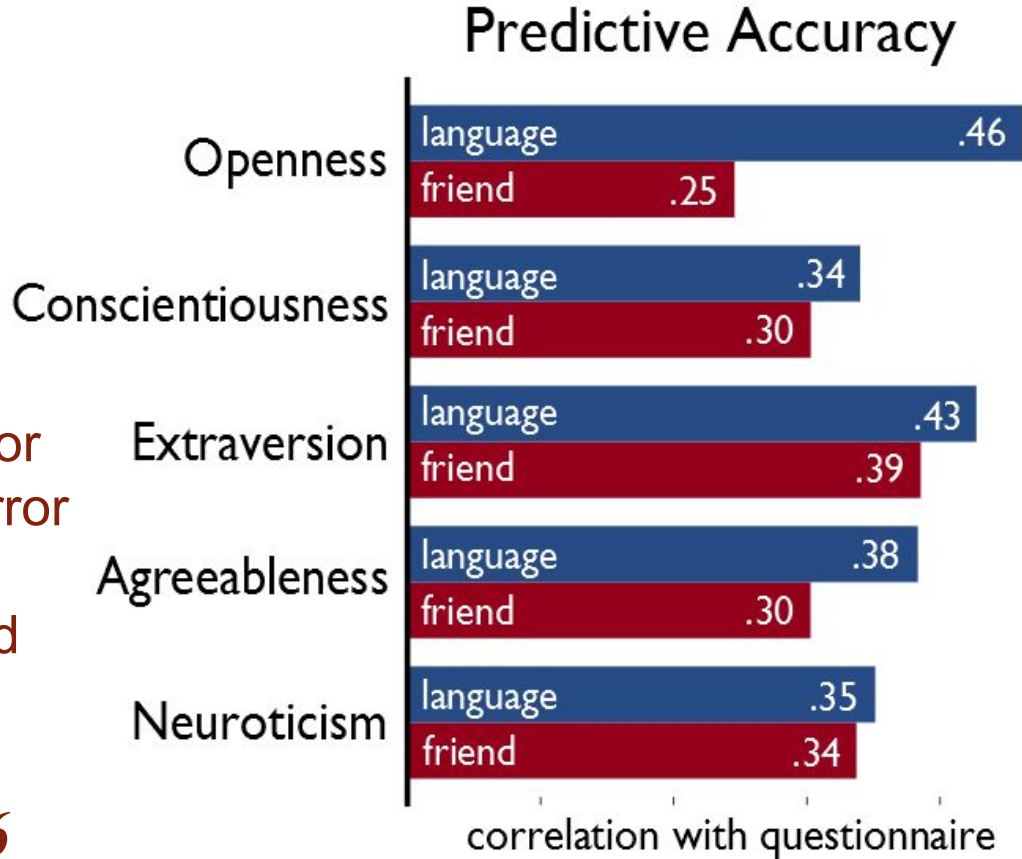
*r = .49 to .66*



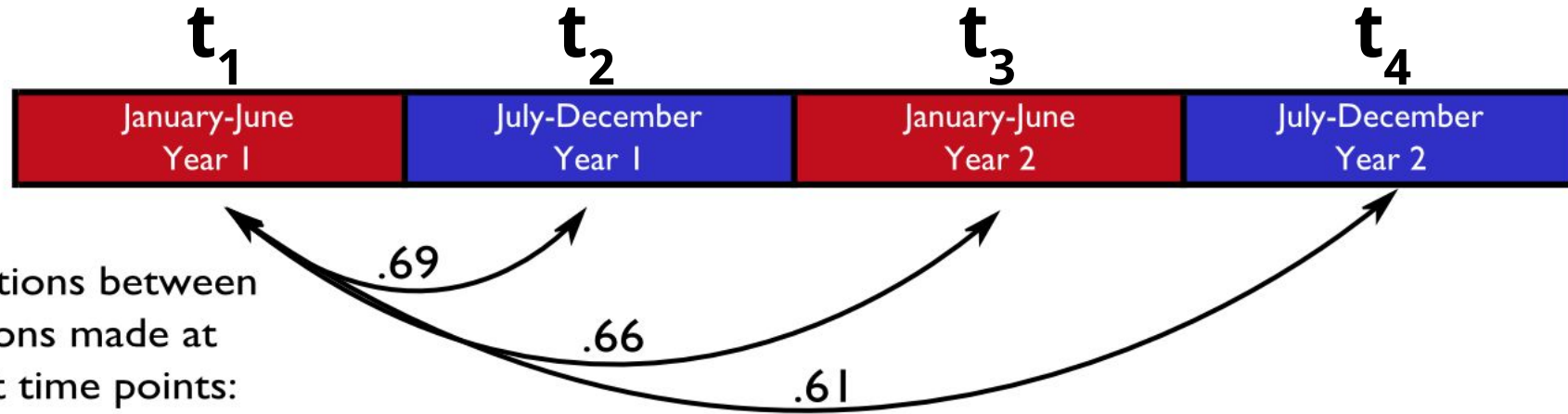
# Language-Based Assessment Evaluation

disattenuated for  
measurement error  
and  
shared method  
variance:

*r = .49 to .66*



# Is it reliable and stable over time?

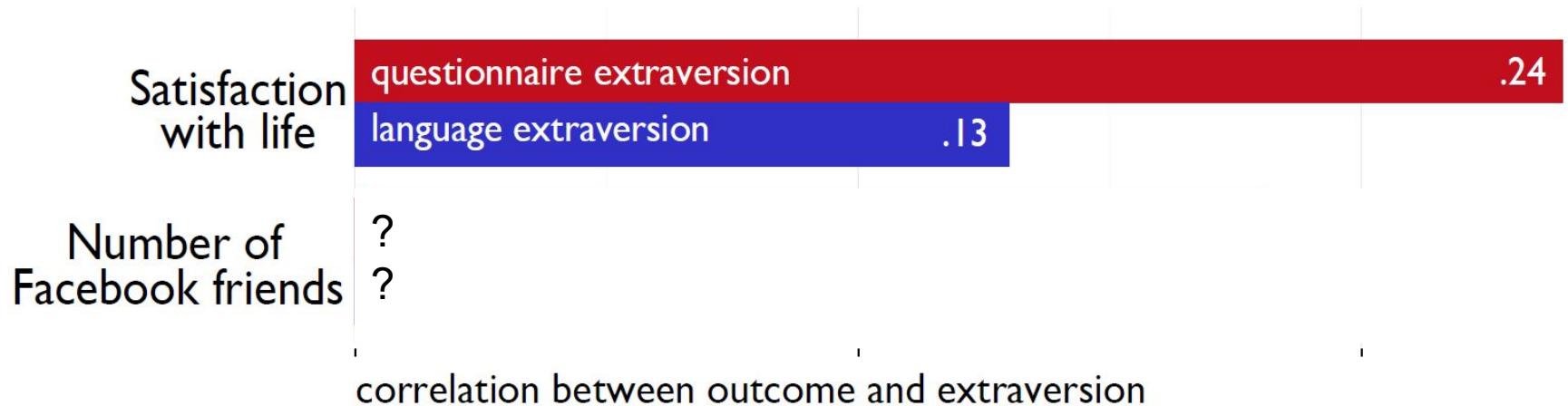


Online 6-month test-retest of 20-item questionnaire:

$$r = .71$$

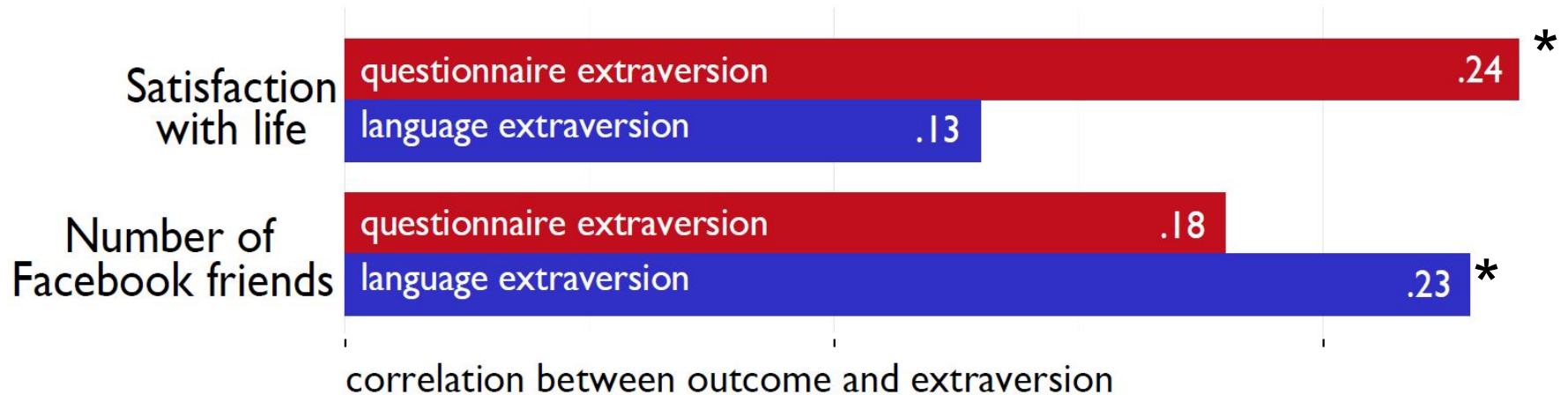
# Does it generalize to other outcomes?

## Example: extraversion



# Does it generalize to other outcomes?

## Example: extraversion





# Other Outcomes?

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)



# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

## Temporal Orientation

(Schwartz et al., 2015)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Meaning in Life

(Schwartz et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

## Temporal Orientation

(Schwartz et al., 2015)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Meaning in Life

(Schwartz et al., 2016)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

## Temporal Orientation

(Schwartz et al., 2015)

## Trustfulness

(Buffone et al., 2018)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Meaning in Life

(Schwartz et al., 2016)

## Control

(Rouhizadeh et al., 2018)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

## Temporal Orientation

(Schwartz et al., 2015)

## Trustfulness

(Buffone et al., 2018)

# Other Outcomes?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

Eichstaedt, ..., & Schwartz  
*PNAS, October, 2018.*

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Meaning in Life

(Schwartz et al., 2016)

## Control

(Rouhizadeh et al., 2018)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

## Temporal Orientation

(Schwartz et al., 2015)

## Trustfulness

(Buffone et al., 2018)

# Other Outcome

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

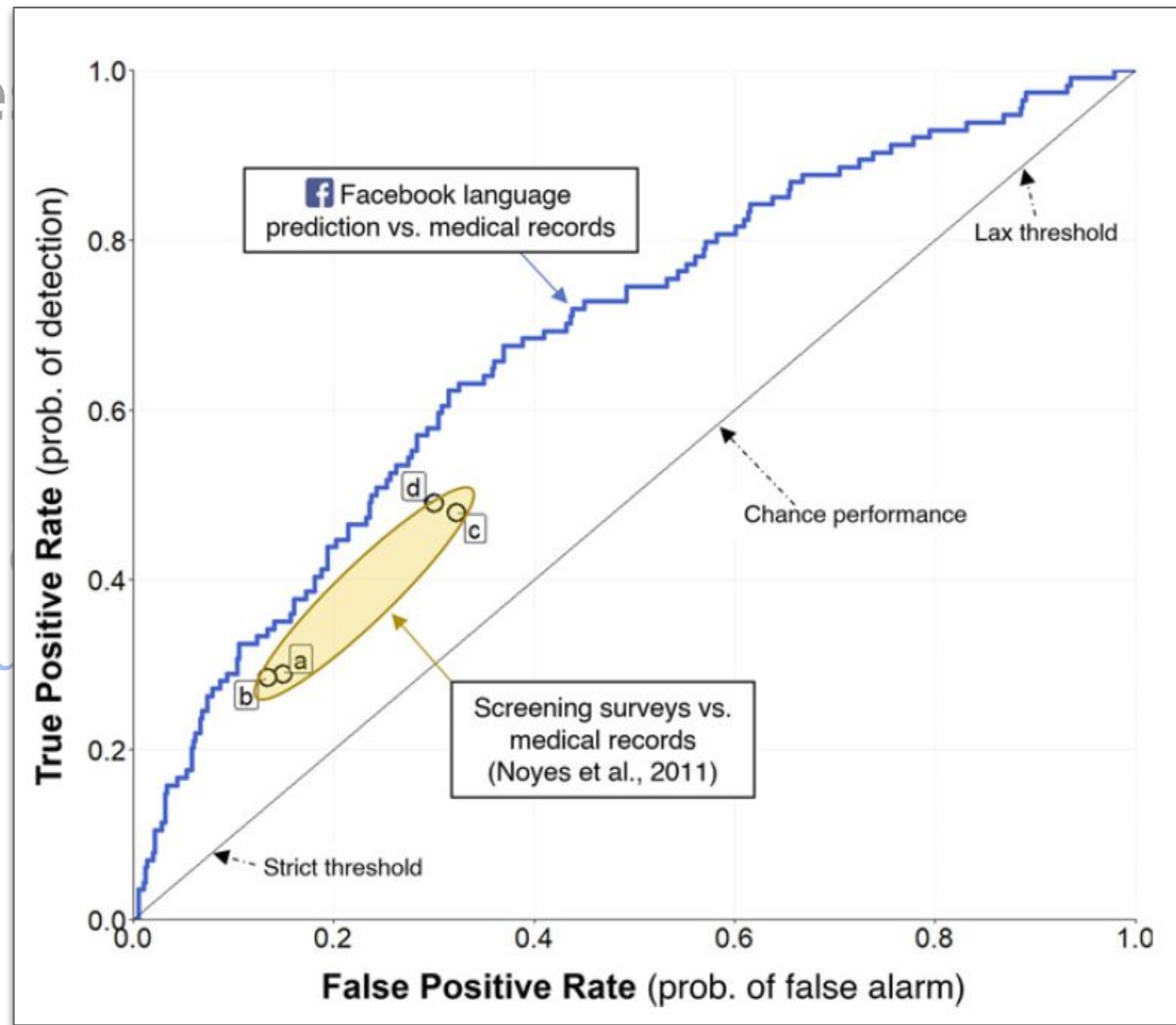
## Spiritual/Religious Outcome

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

Eichstaedt, ..., & Schwartz  
*PNAS, October, 2018.*



(Rouhizadeh et al., 2018)

Other

Life Satisfaction

(Schwartz et al., 2018)

Mental Health

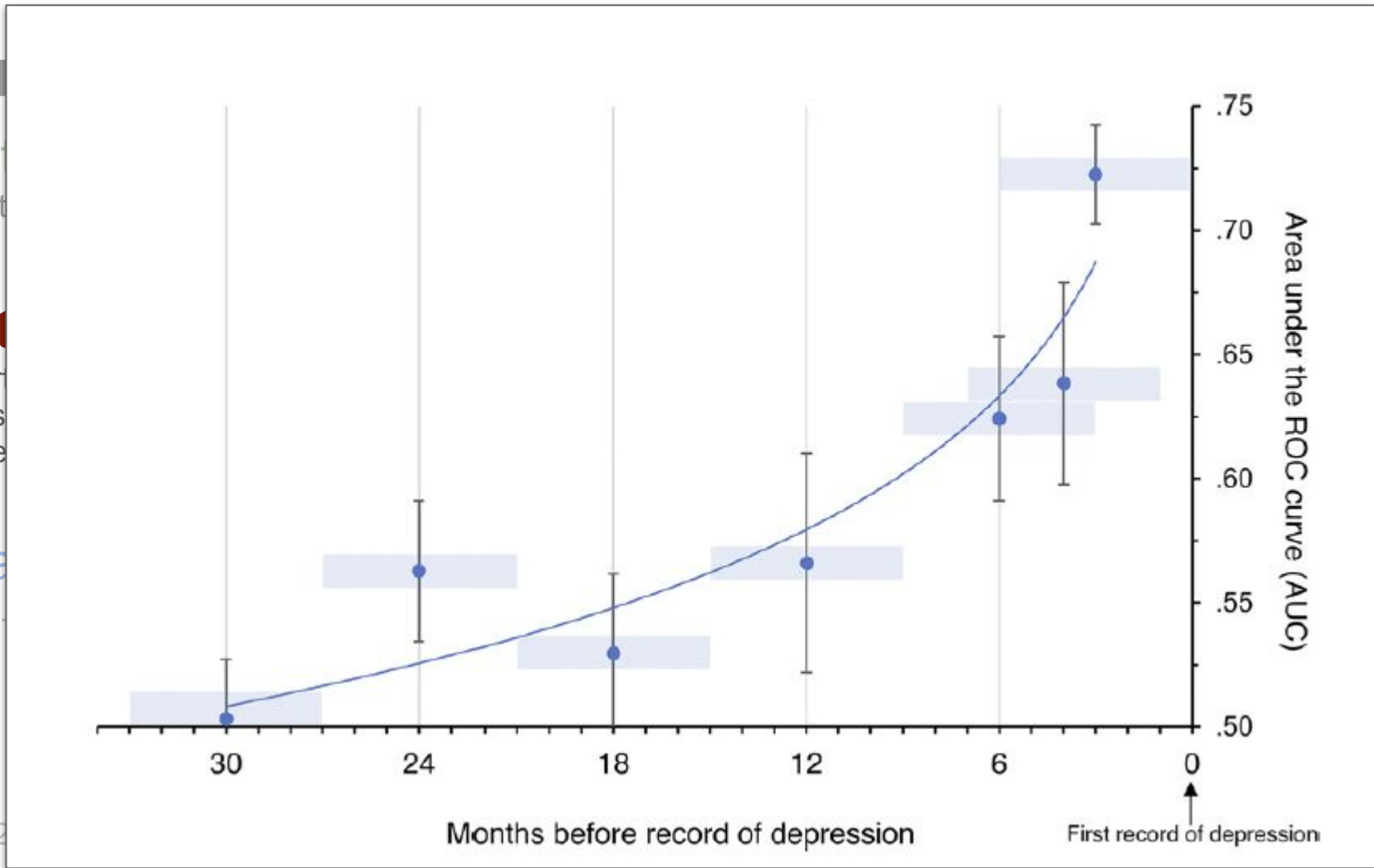
(Schwartz et al., 2018; Coppersmith et al., 2018; Eichstaedt et al., 2018)

Spirituality

(Yaden et al., 2018)

Causal Inference

(Son et al., 2018)



Eichstaedt, ..., & Schwartz  
*PNAS, October, 2018.*

(Rouhizadeh et al., 2018)

(Buttone et al., 2018)





# Depth?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Meaning in Life

(Schwartz et al., 2016)

## Control

(Rouhizadeh et al., 2018)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)

## Temporal Orientation

(Schwartz et al., 2015)

## Trustfulness

(Buffone et al., 2018)

# Depth?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Meaning in Life

(Schwartz et al., 2016)

## Control

(Rouhizadeh et al., 2018)

## Demographics

(Sap et al., 2014)

## Temporal Orientation

(Schwartz et al., 2015)

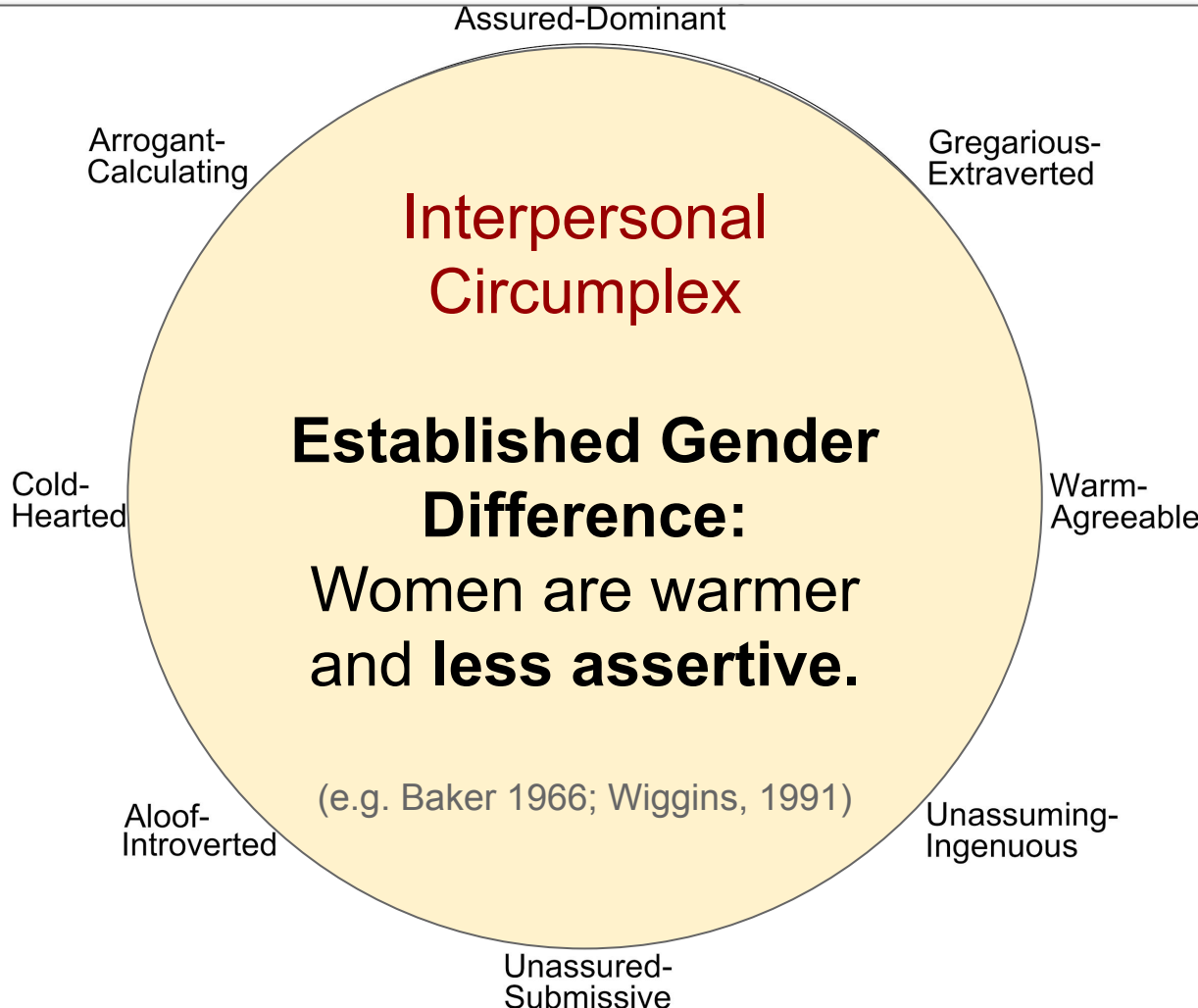
## Trustfulness

(Buffone et al., 2018)

# Depth?

## Personality

## Demographics

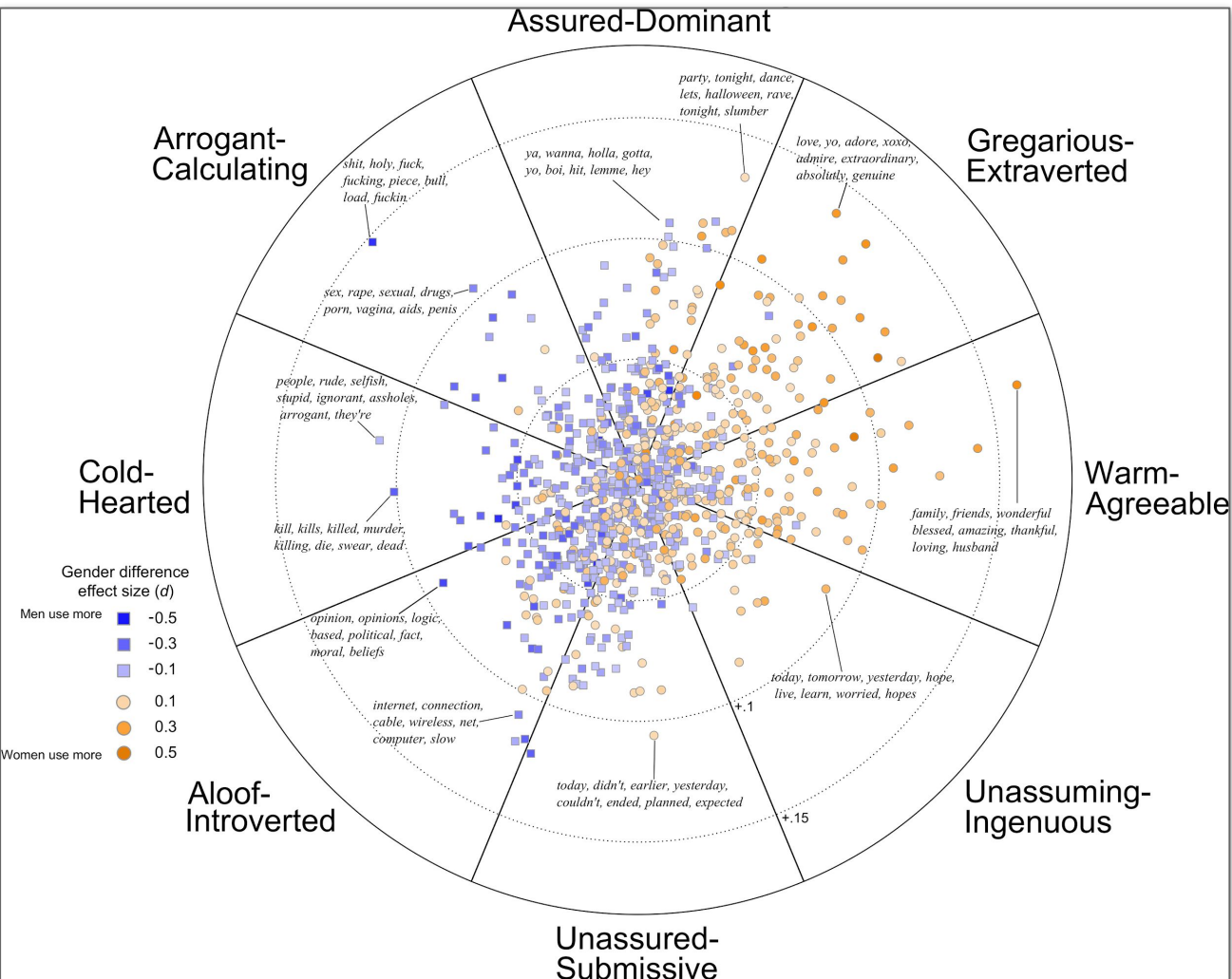


(Park et al., 2017)

# Depth?

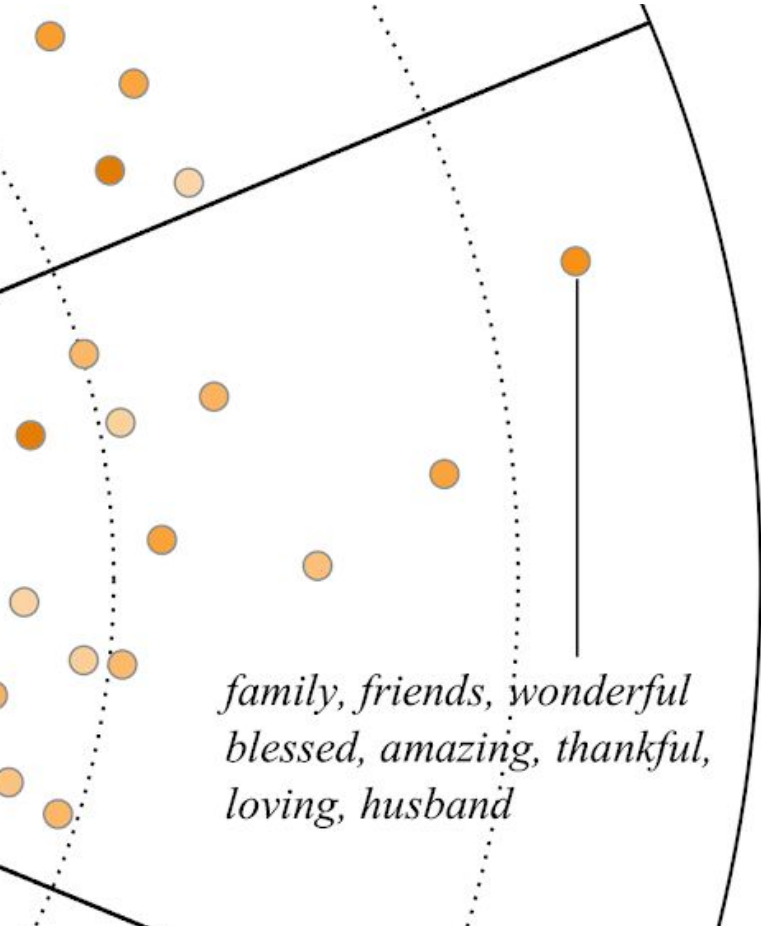
# Personality

# Demographics



(Park et al., 2017)

# Depth?



## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

# Warm- Agreeable

## Demographics

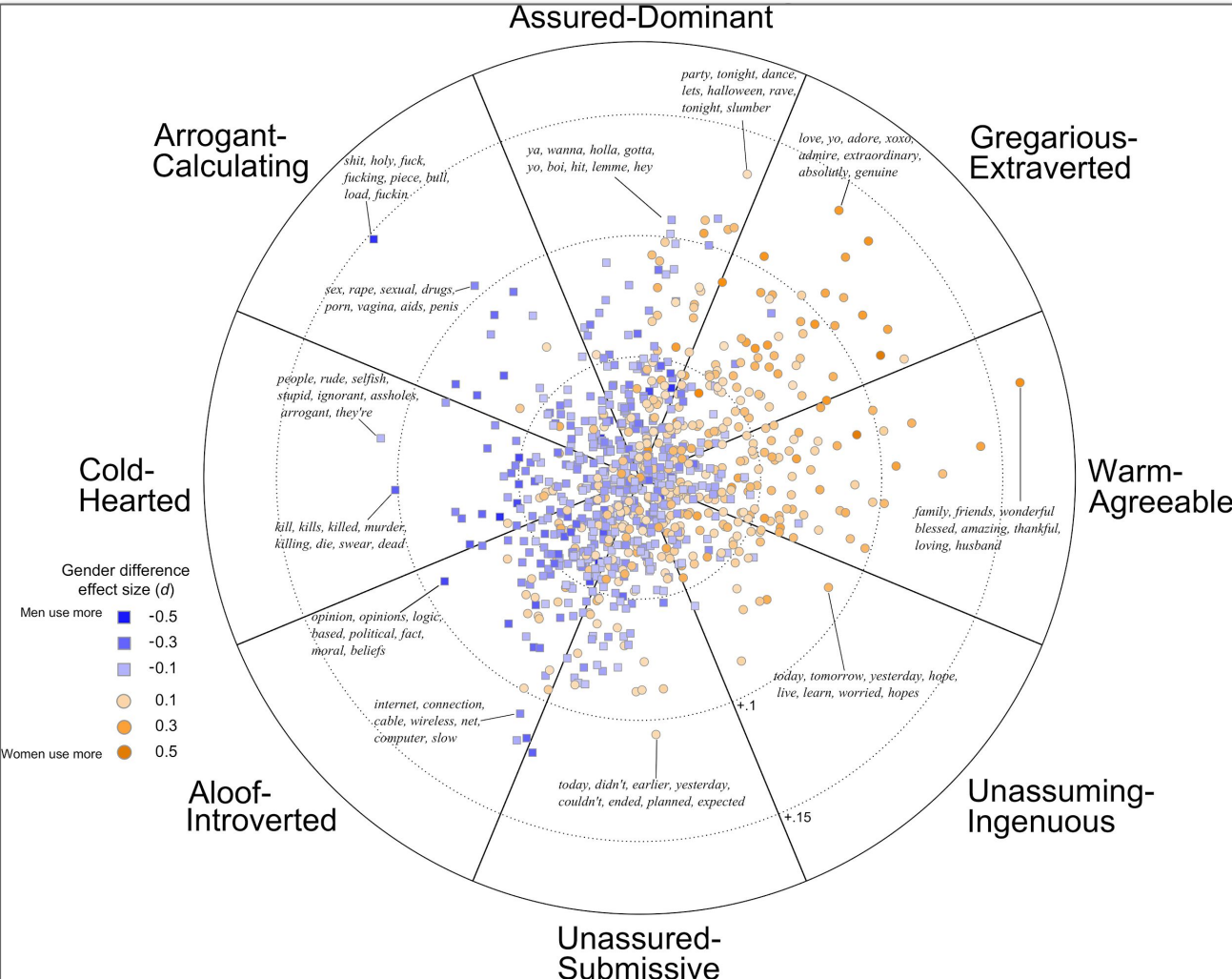
(Sap et al., 2014)

(Park et al., 2017)

# Depth?

# Personality

# Demographics



(Park et al., 2017)

# Depth?

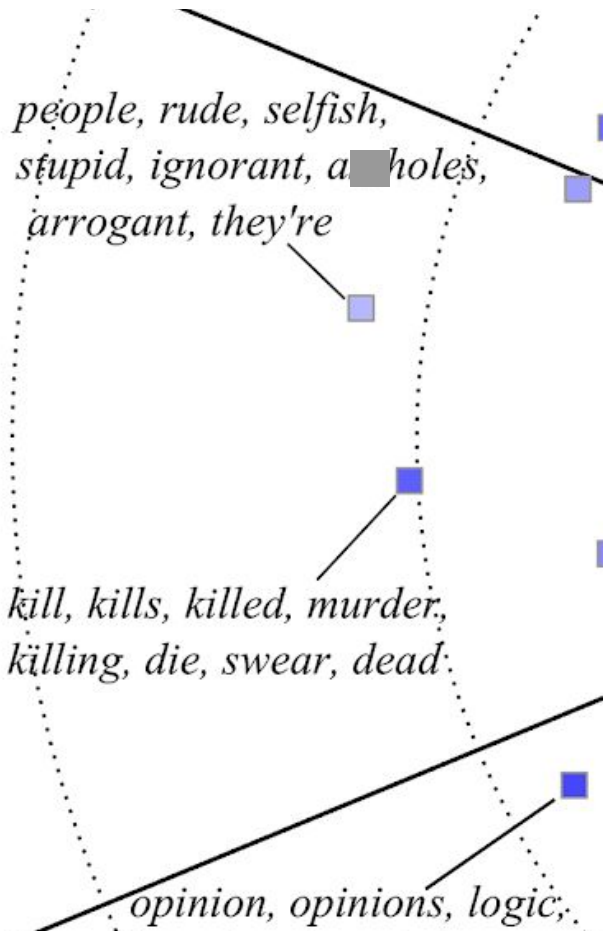
## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

Cold-  
Hearted

Gender difference  
effect size ( $d$ )

more  -0.5



**Demographics**

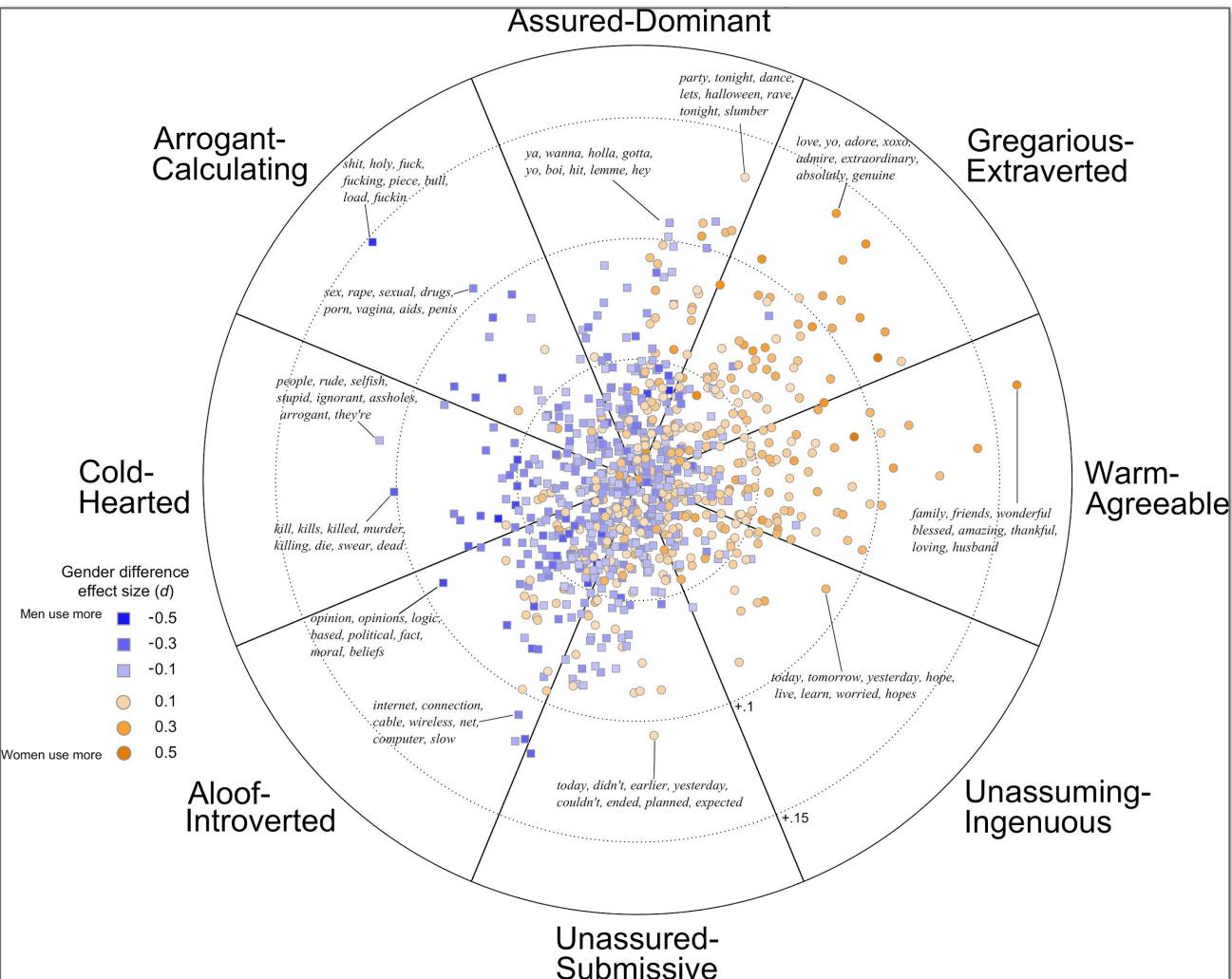
(Park et al., 2017)



# Depth?

# Personality

# Demographics



(Park et al., 2017)

# Depth?

# Personality

# Demographics

## Interpersonal Circumplex

Questionnaire-based  
Gender Differences:  
Women are warmer  
but less assertive.

(e.g. Baker 1966; Wiggins, 1991)

Assured-Dominant

Arrogant-  
Calculating

Gregarious-  
Extraverted

Cold-  
Hearted

Warm-  
Agreeable

Aloof-  
Introverted

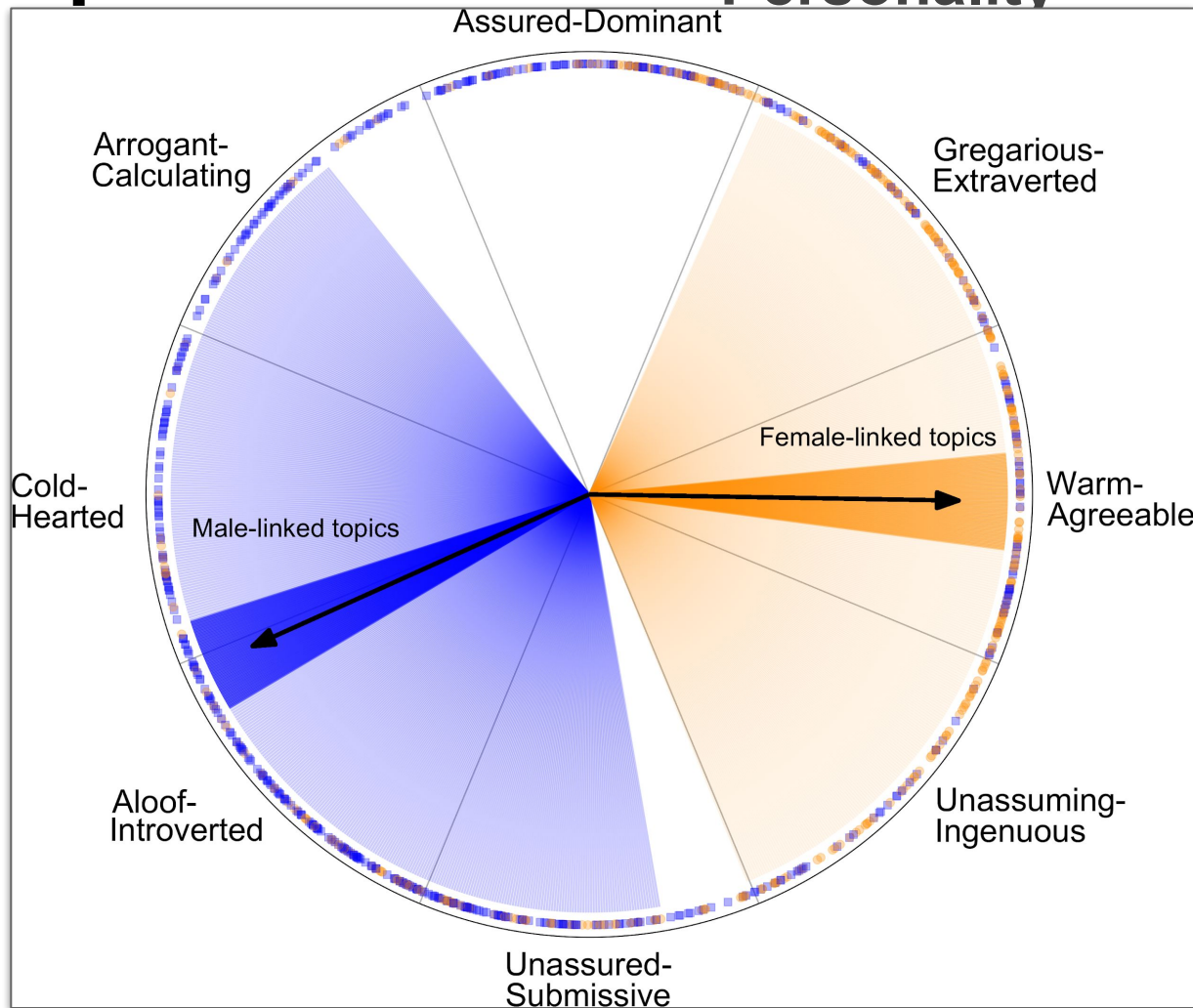
Unassuming-  
Ingenuous

Unassured-  
Submissive

(Park et al., 2017)

# Depth?

# Personality



# Demographics

(Park et al., 2017)

# Depth?

## Life Satisfaction

(Schwartz et al., 2013; 2016)

## Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

## Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

## Causal Explanations

(Son et al., 2018)

## Personality

(Schwartz et al., 2013;  
Park et al., 2015)

## Emotion / Affect

(Preotiuc-Pietro et al., 2016)

## Dark Triad

(Preotiuc-Pietro et al., 2016)

## Meaning in Life

(Schwartz et al., 2016)

## Control

(Rouhizadeh et al., 2018)

## Characterizing Gratitude

(Carpenter et al., 2016)

## Demographics

(Sap et al., 2014)  
(Park et al., 2017)

## Temporal Orientation

(Schwartz et al., 2015)

## Trustfulness

(Buffone et al., 2018)

# Depth?

Life Satisfaction

(Schwartz et al., 2013; 2016)

Mental Health

(Schwartz et al., 2013;  
Coppersmith et al., 2014;  
Eichstaedt et al., 2018)

Spiritual/Religious Outcomes

(Yaden et al., 2016, 2017)

Causal Explanations

(Son et al., 2018)

Personality

(Schwartz et al., 2013;  
Park et al., 2015)

Emotion / Affect

(Preotiuc-Pietro et al., 2016)

Dark Triad

(Preotiuc-Pietro et al., 2016)

Meaning in Life

(Schwartz et al., 2016)

Control

(Rouhizadeh et al., 2018)

Characterizing Gratitude

(Carpenter et al., 2016)

Demographics

(Sap et al., 2014)

Temporal Orientation

(Schwartz et al., 2015)

Trustfulness

(Buffone et al., 2018)

# Depth?

Trust from  
Language

Questionnaire:

---

*Dis att.  $\beta$*

Trust

.54

**facets**

**of**

**agreeableness**

Morality

Altruism

Cooperation

Modesty

Sympathy

---

*Age-gender adjusted beta*

**Trustfulness**

(Buffone, Zamani et al., 2018)

# Depth?

Trust from  
Language

Questionnaire:

*Dis att.  $\beta$*

**facets  
of  
agreeableness**

Trust

.54

Morality

.15

Altruism

Cooperation

Modesty

-.03

Sympathy

.08

*Age-gender adjusted beta*

**Trustfulness**

(Buffone, Zamani et al., 2018)

# Depth?

Trust from  
Language

Questionnaire:

*Dis att.  $\beta$*

**facets  
of  
agreeableness**

Trust

.54

Morality

**.15**

Altruism

.23

Cooperation

.35

Modesty

**-.03**

Sympathy

**.08**

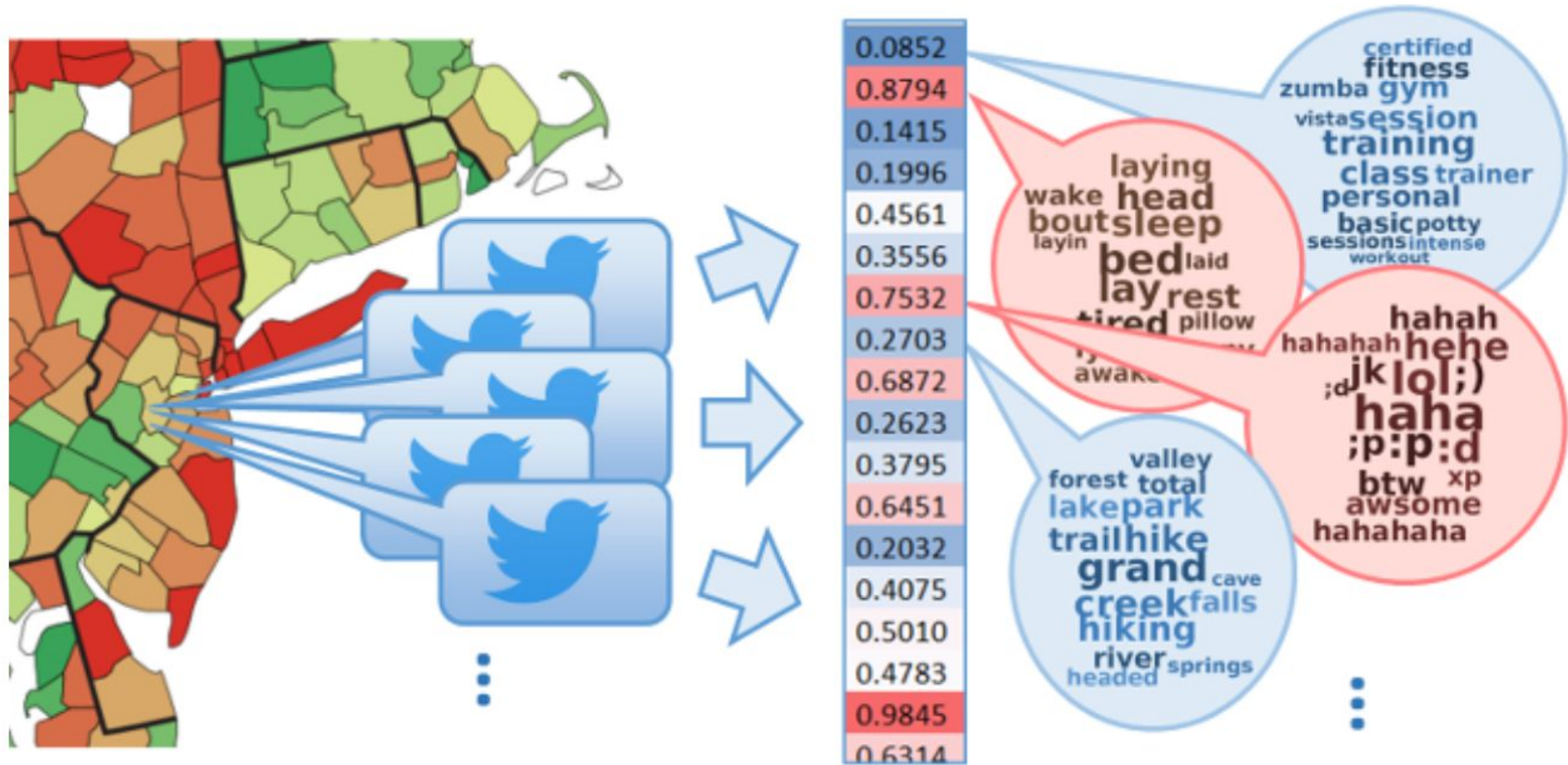
*Age-gender adjusted beta*

**Trustfulness**

(Buffone, Zamani et al., 2018)



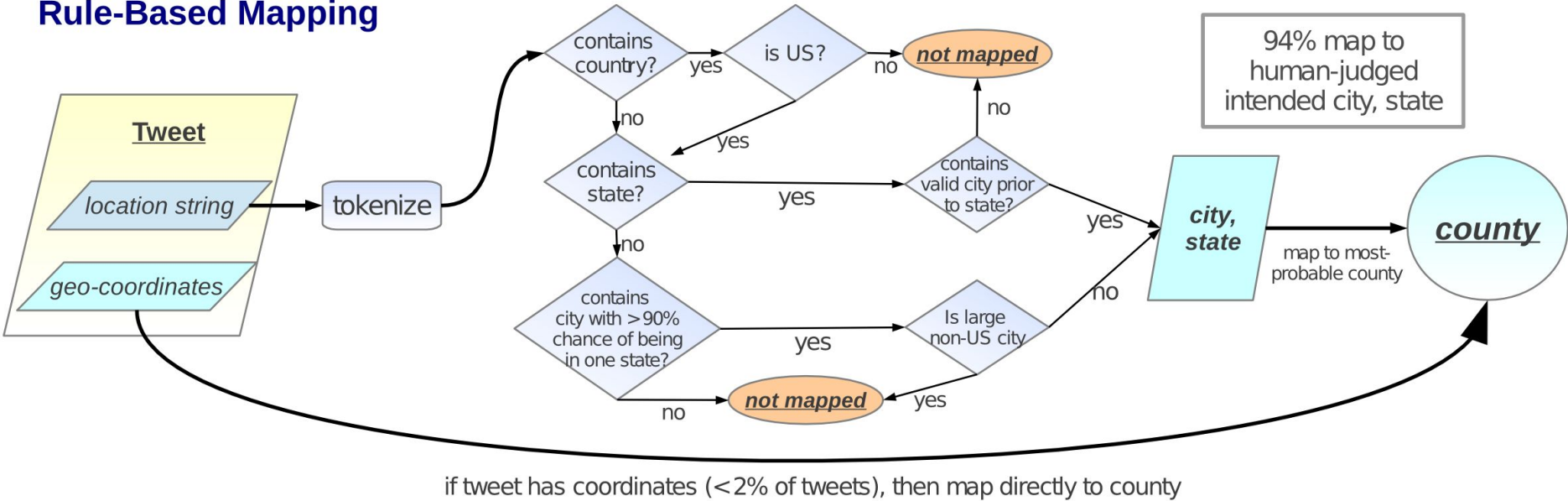
# A Community's Digital Footprint as an Encoding of Who They Are



(Schwartz, 2018)

# 1. Text-based Geolocation

## Rule-Based Mapping

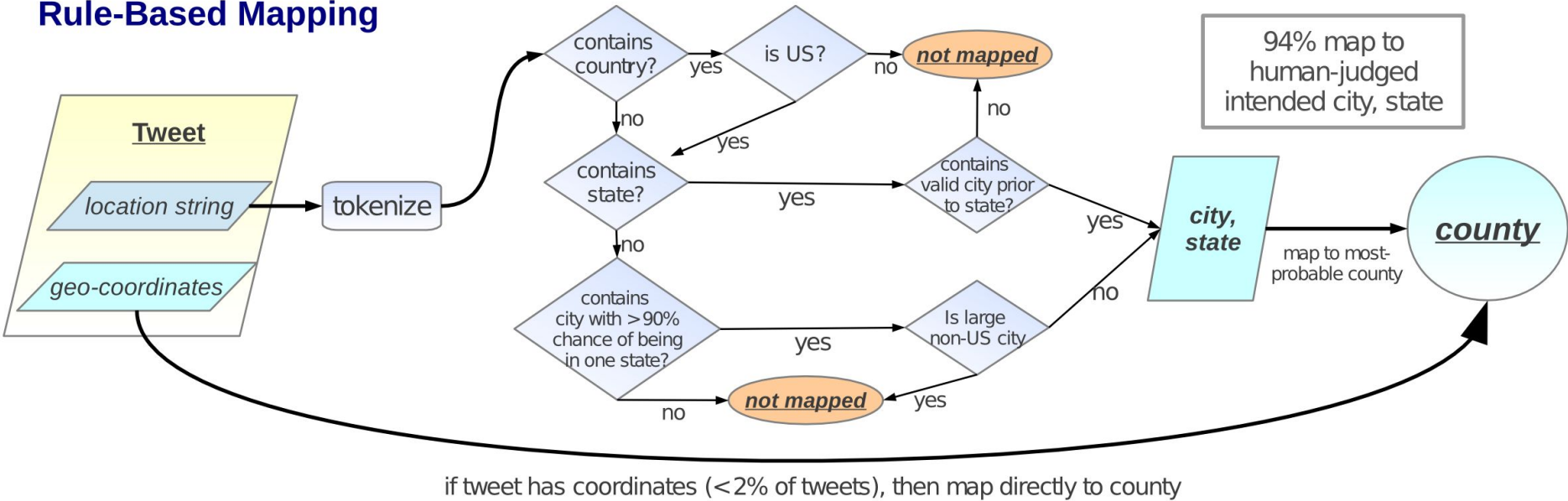


## versions

1. Language-behavior based (e.g. status updates, tweets)
2. Profile-based (e.g. free-response location field)

# 1. Text-based Geolocation

## Rule-Based Mapping



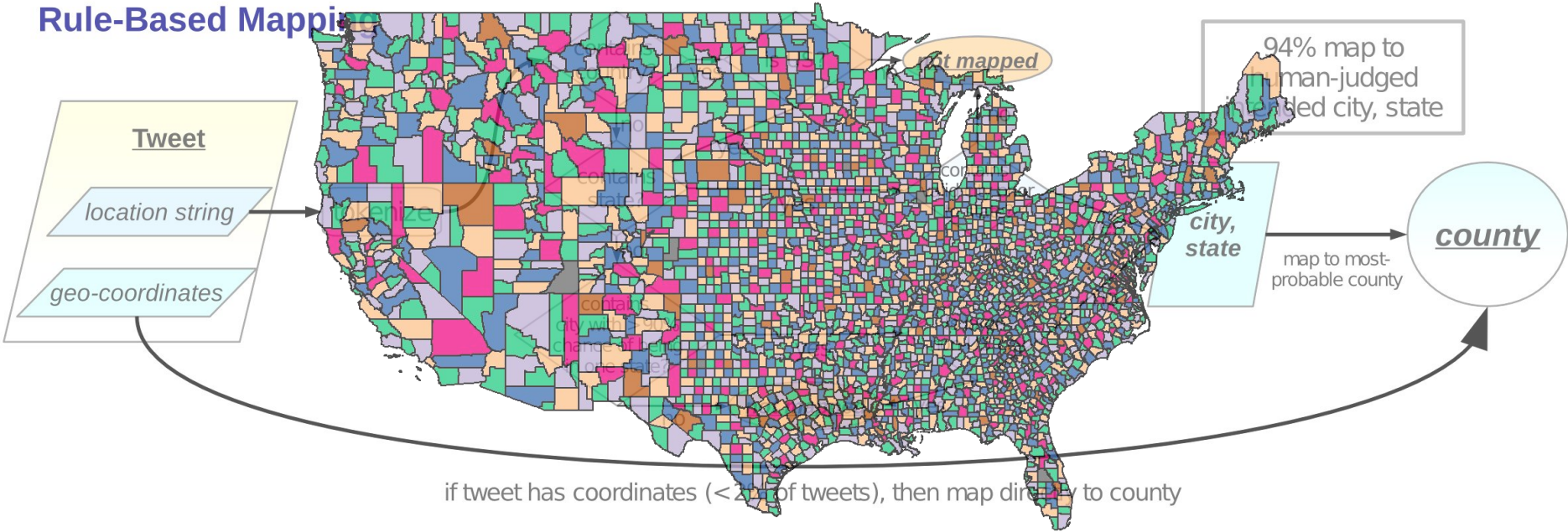
## versions

94% accuracy at mapping to U.S. counties

1. Language-behavior based (e.g. status updates, tweets)
2. Profile-based (e.g. free-response location field)

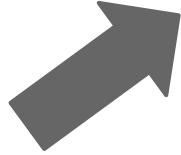
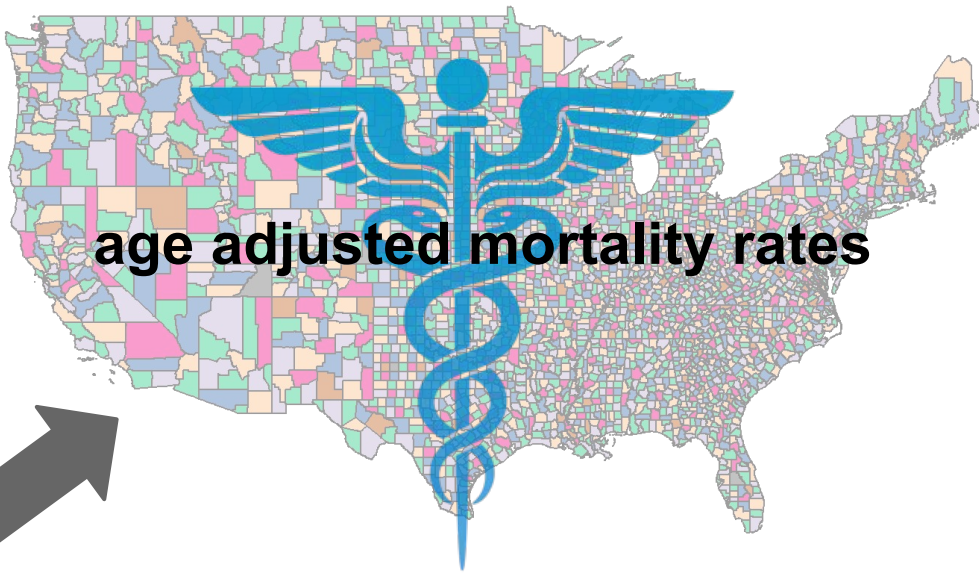
# 1. Text-based Geolocation

## Rule-Based Mapping



94% accuracy at mapping to U.S. cities

1. Language-behavior based (e.g. status updates, tweets)
2. Profile-based (e.g. free-response location field)



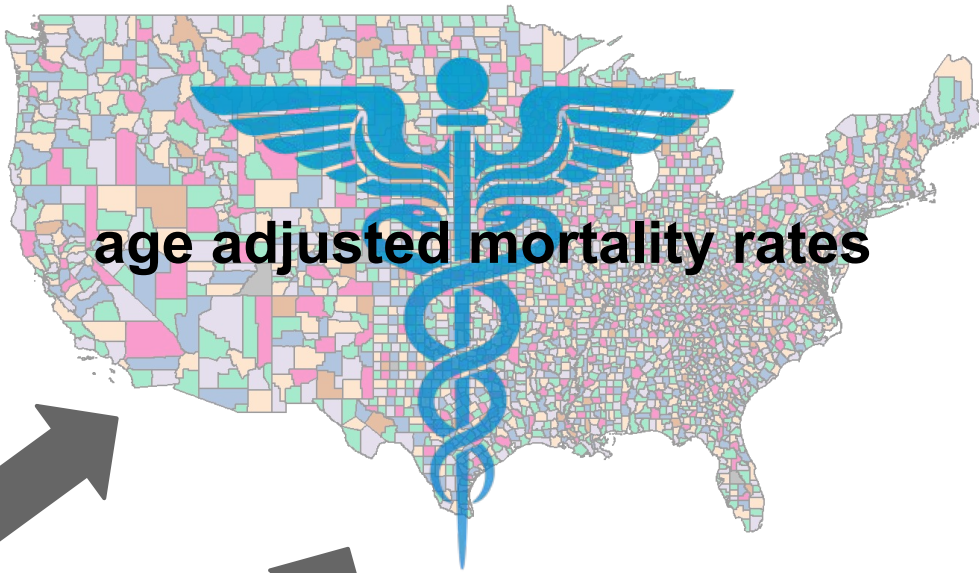
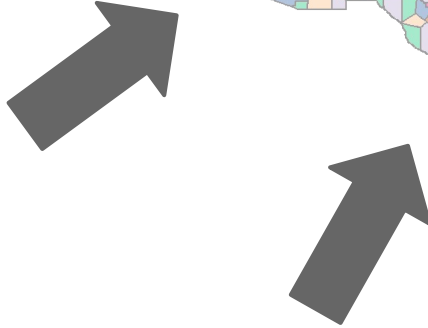
age  
gender  
race

**age adjusted mortality rates**



**age  
gender  
race**

**income  
unemployment**



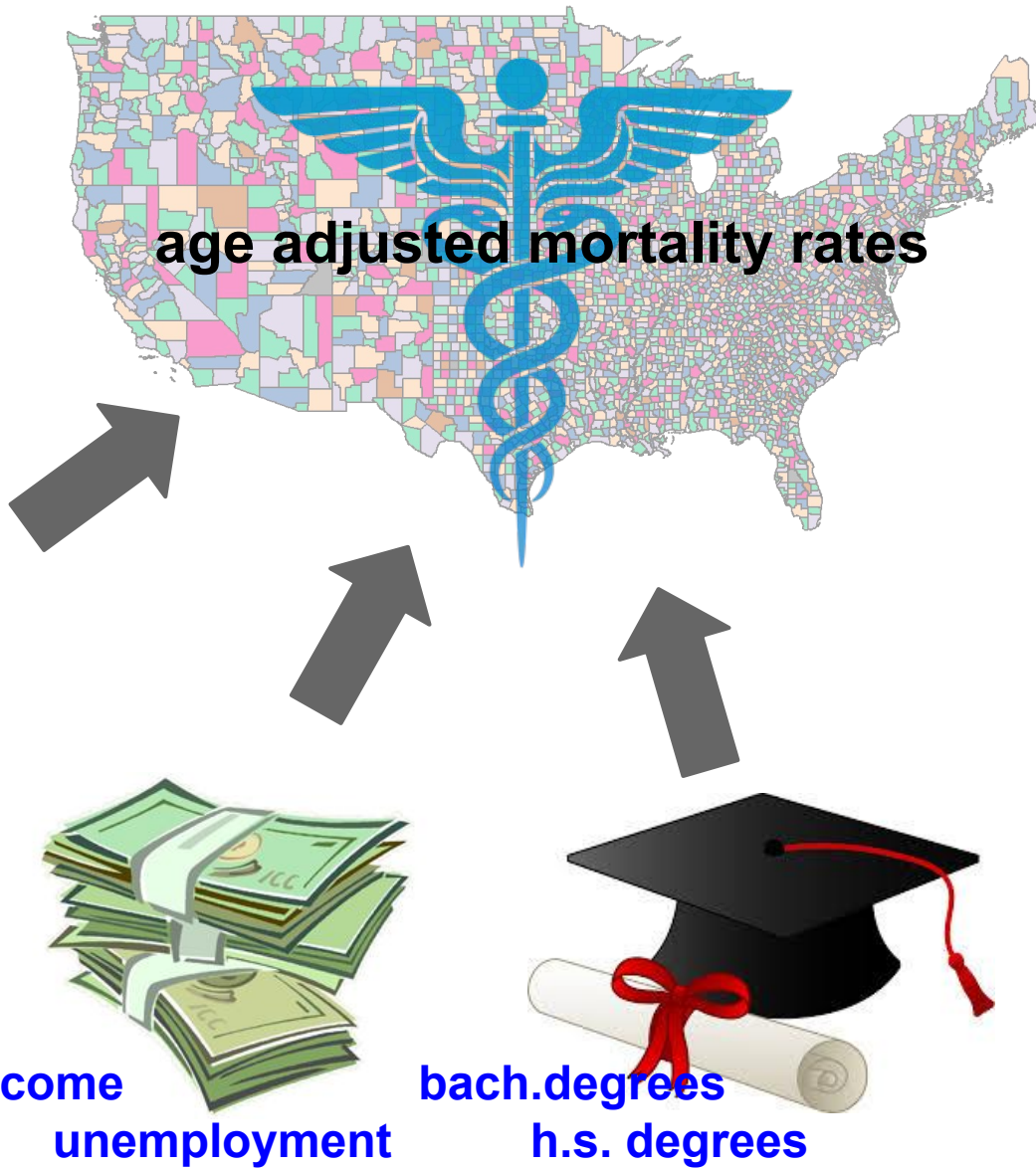
**age adjusted mortality rates**



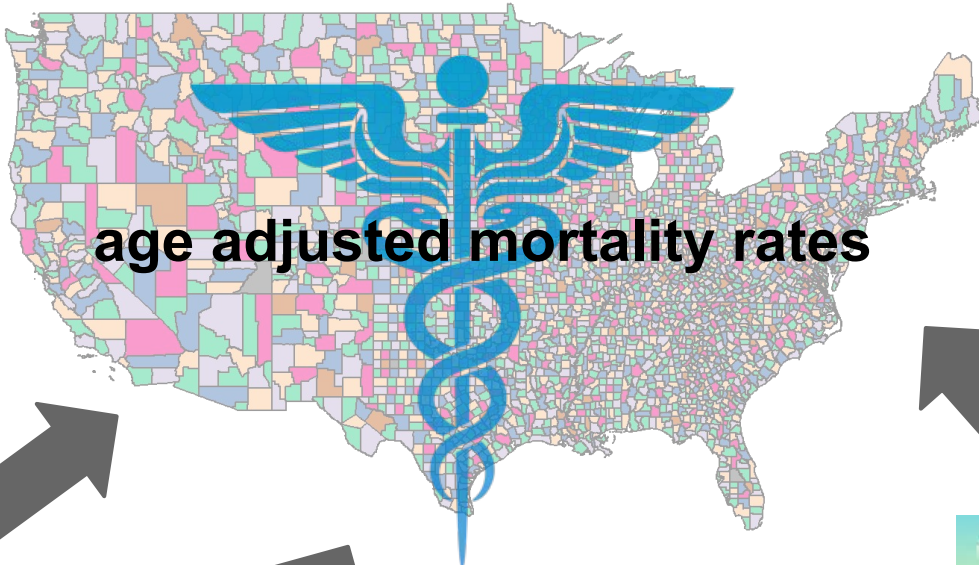
**age  
gender  
race**

**income  
unemployment**

**bach.degrees  
h.s. degrees**



# age adjusted mortality rates



age  
gender  
race

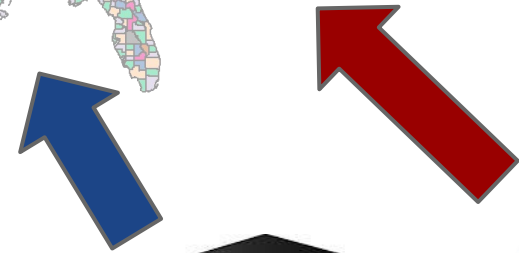
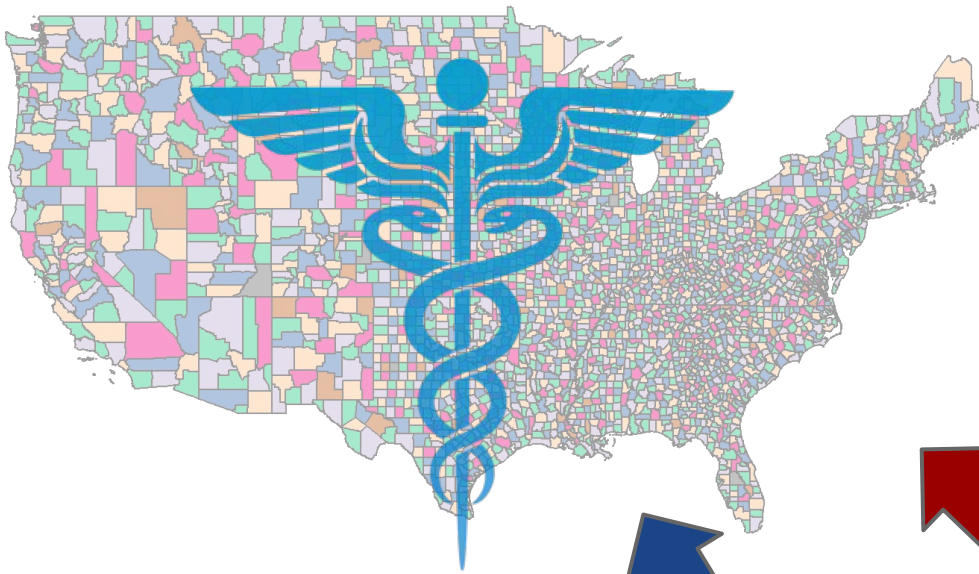
income  
unemployment

bach.degrees  
h.s. degrees

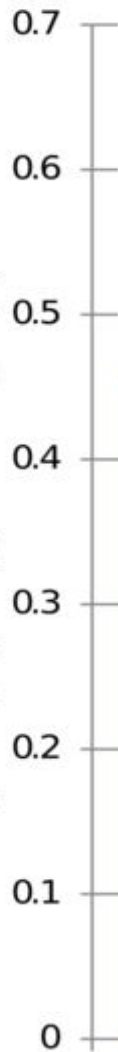
pop. density  
% married





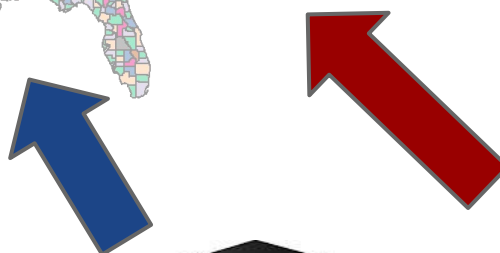
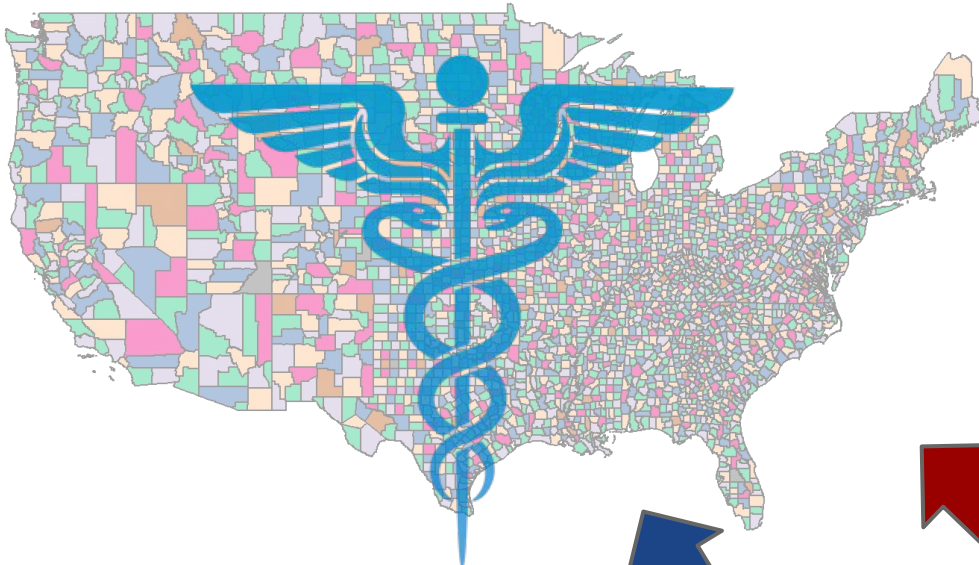


Variance Explained

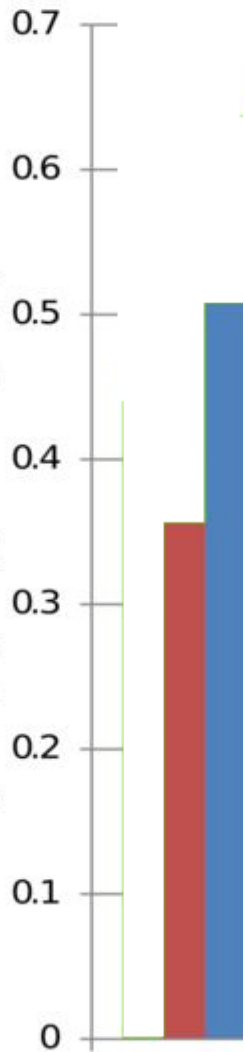


Total Mortality Rates

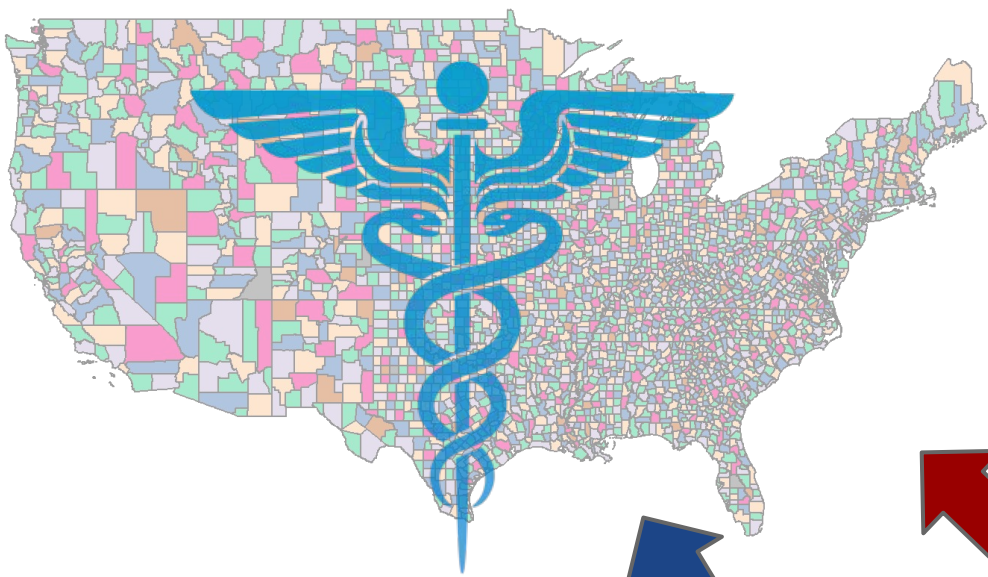
Tested Out-of-Sample



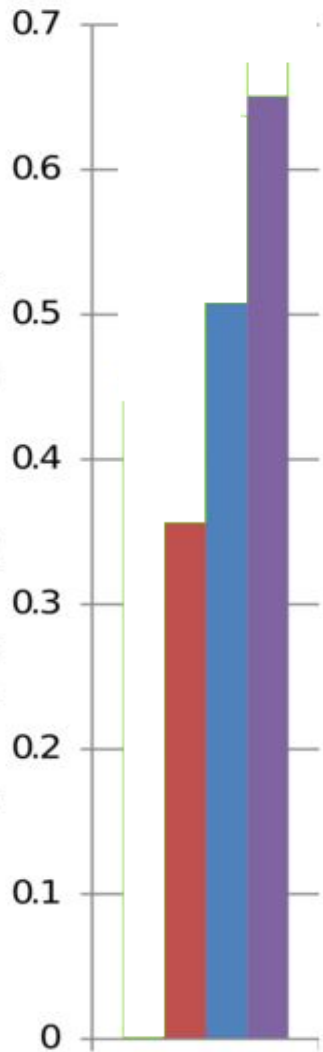
Variance Explained



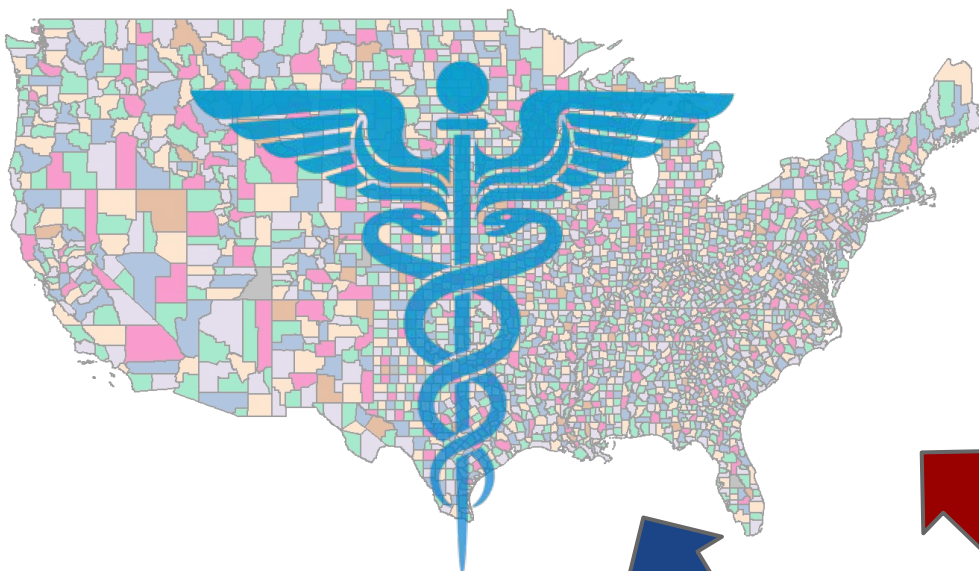
Total Mortality Rates



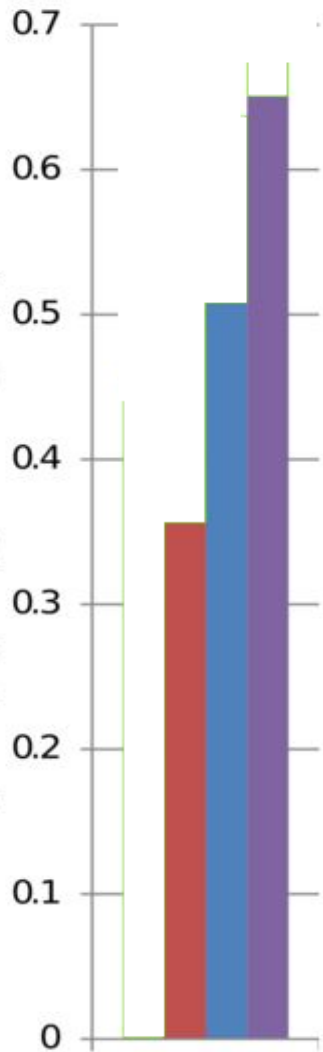
Variance Explained



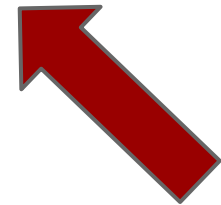
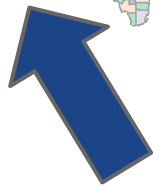
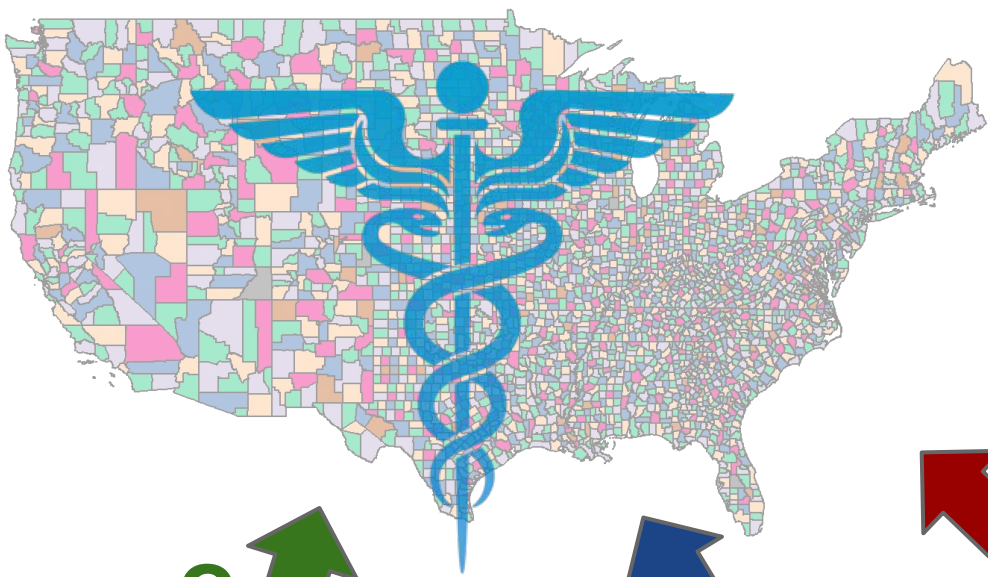
Total Mortality Rates



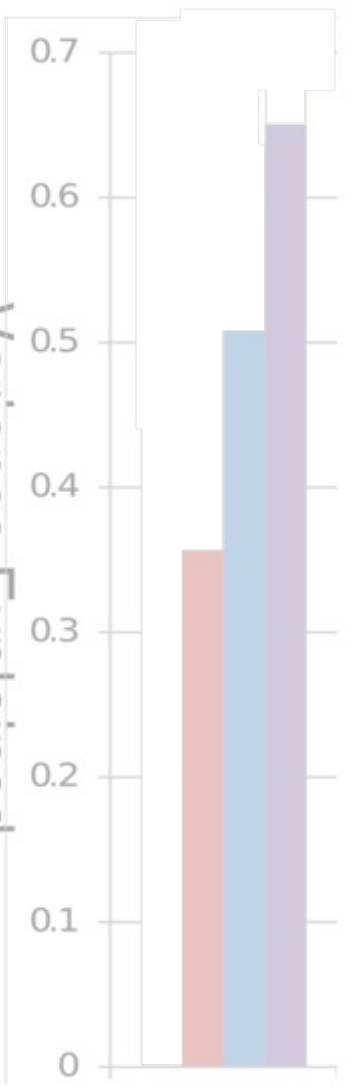
Variance Explained



Total Mortality Rates



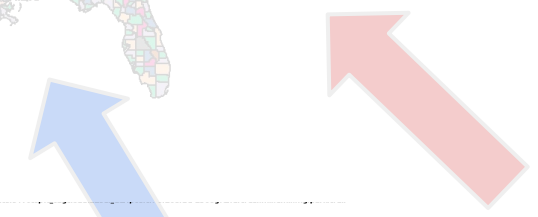
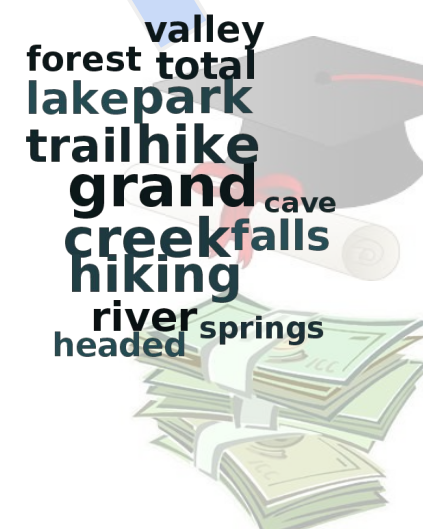
Variance Explained



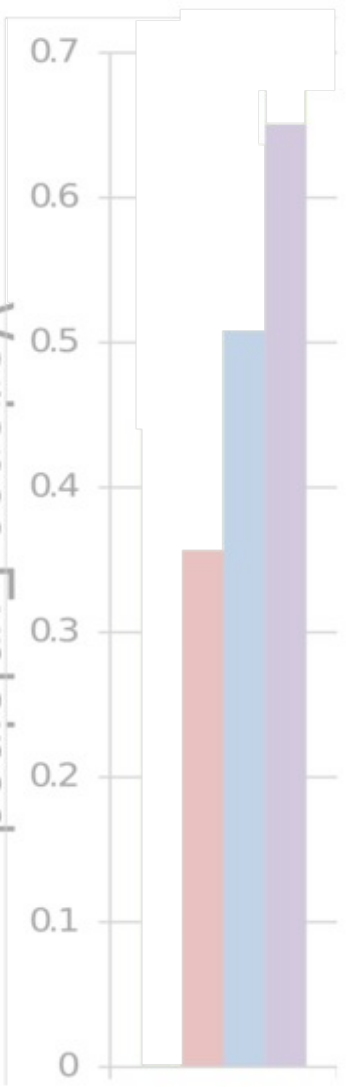
Total Mortality Rates



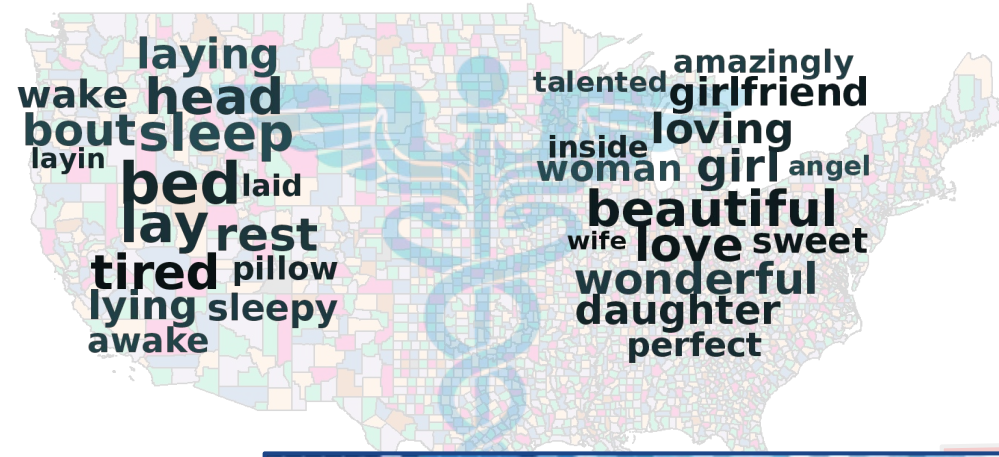
homecoming  
basketball  
hockey play  
football  
game  
twins  
tonight intense  
watch  
boys season  
volleyball pumped  
baseball



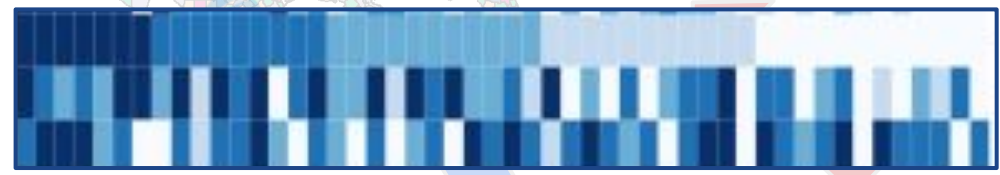
Variance Explained



Total Mortality Rates



homecoming  
basketball  
hockey play  
football  
game  
twins  
tonight intense  
watch  
boys season  
volleyball pumped  
baseball



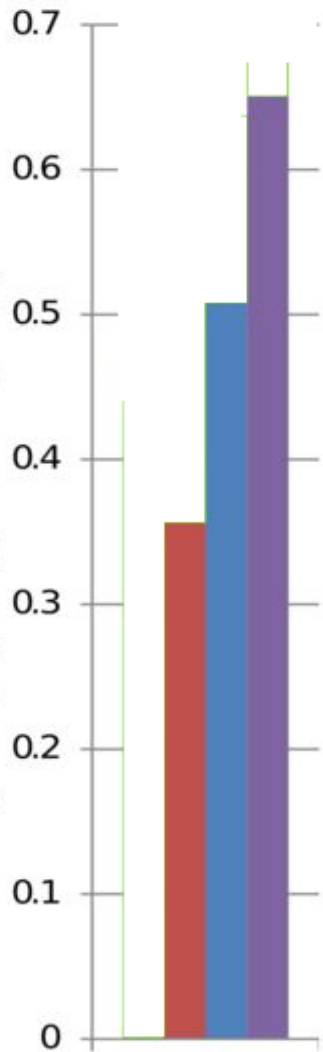
valley  
forest total  
lake park  
trail hike  
grand  
creek falls  
hiking  
river springs  
headed



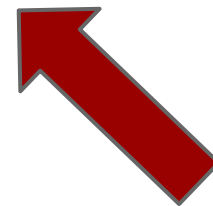
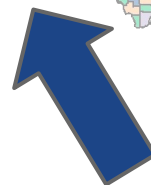
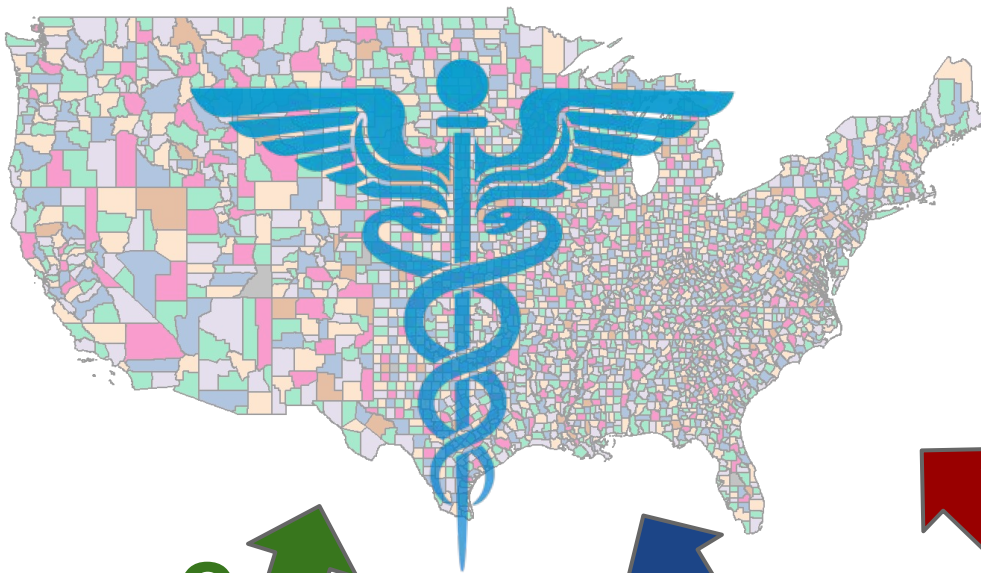
honestly  
arent alot  
confused  
idc dont ive  
isnt im idk  
anymore  
whats mad  
wont upset



Variance Explained

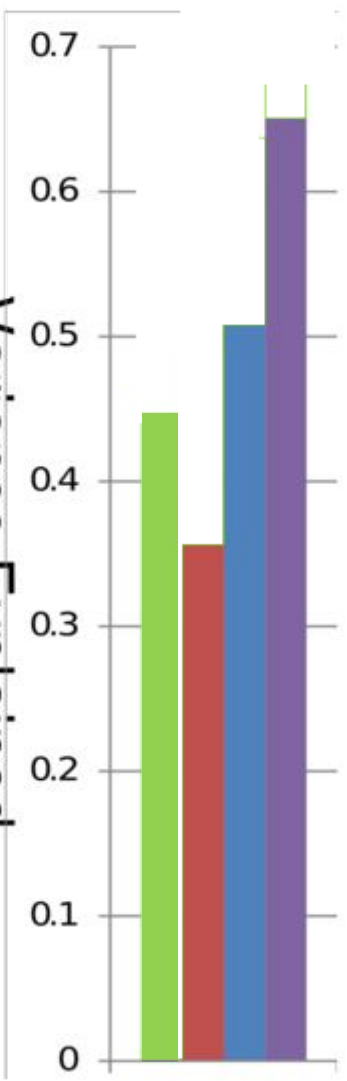


Total Mortality Rates

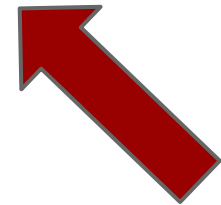
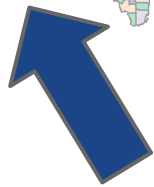
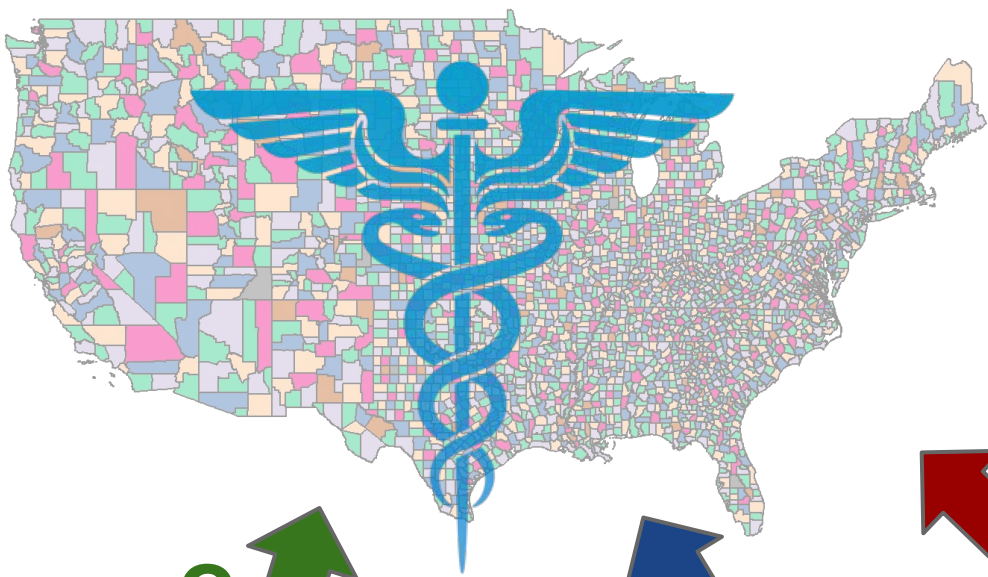


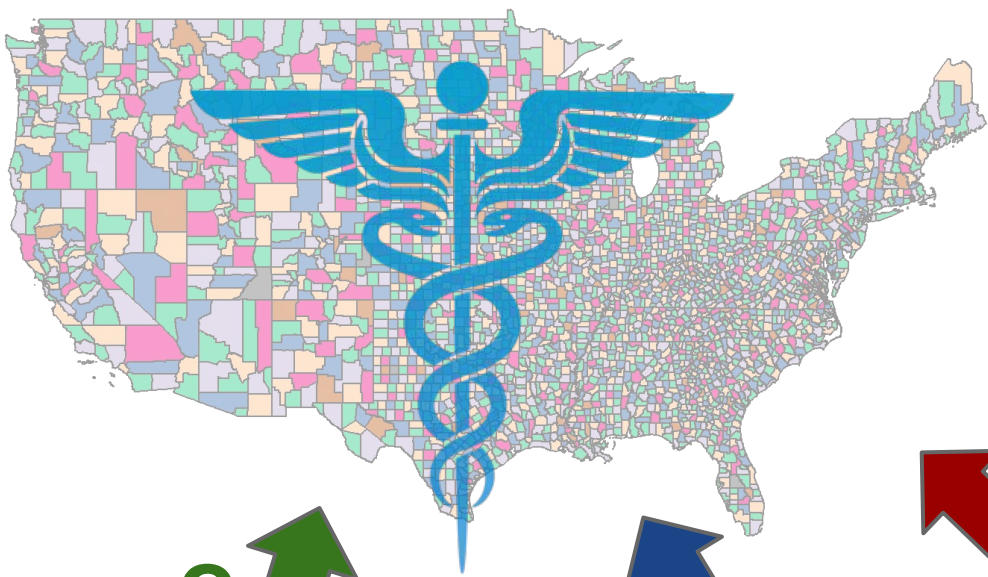
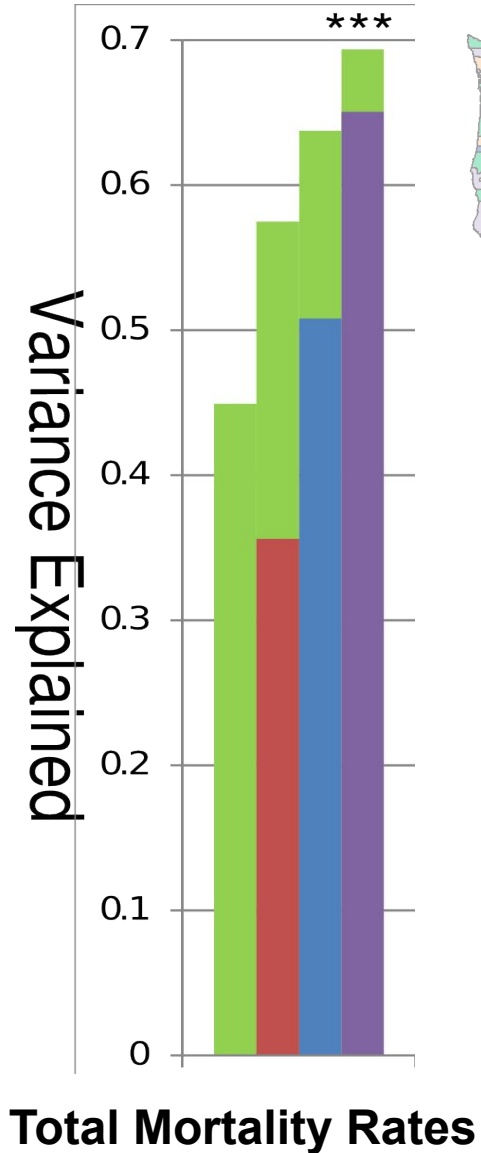


Variance Explained



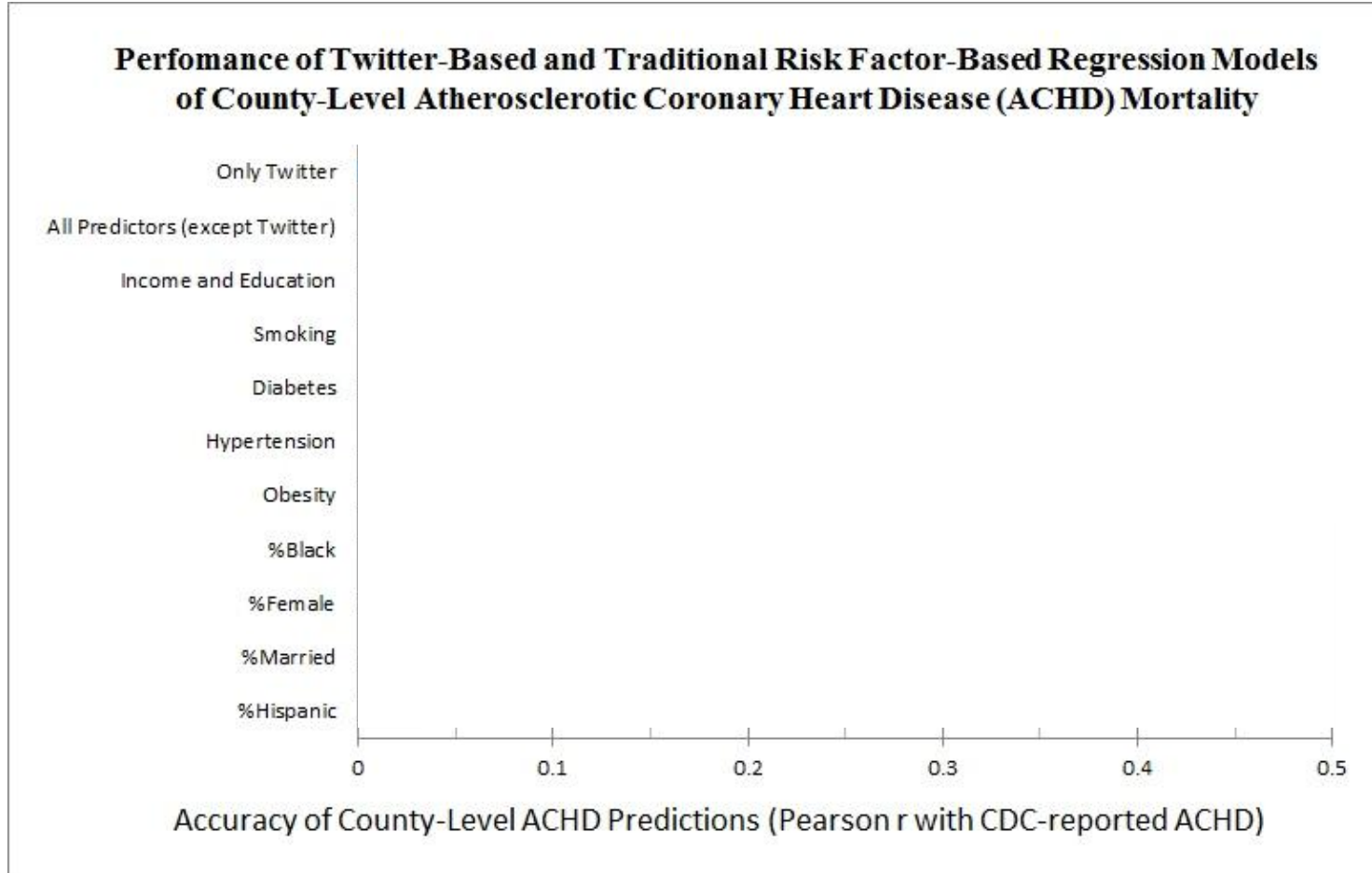
Total Mortality Rates





Total Mortality Rates

# Twitter Predicts Heart Disease





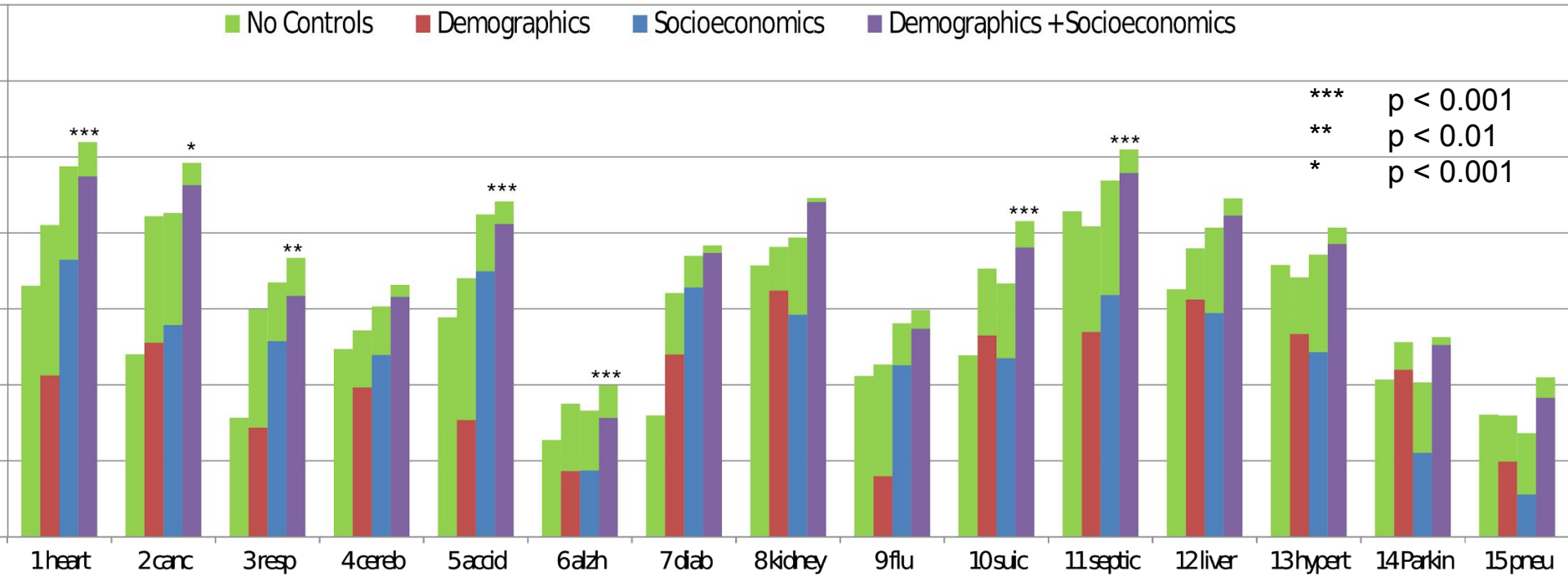
1. Diseases of heart
2. Malignant neoplasms (cancers)
3. Chronic lower respiratory
4. Cerebrovascular diseases  
(strokes)
5. Accidents, unintentional
6. Alzheimer's disease
7. Diabetes melitus
8. Kidney Diseases
9. Influenza & Pneumonia
10. Intentional self-harm  
(suicide)
11. Septicemia
12. Liver Disease
13. Hypertension
14. Parkinson's
15. Pneumonitus

## TOP 15 Causes of Death, 2013

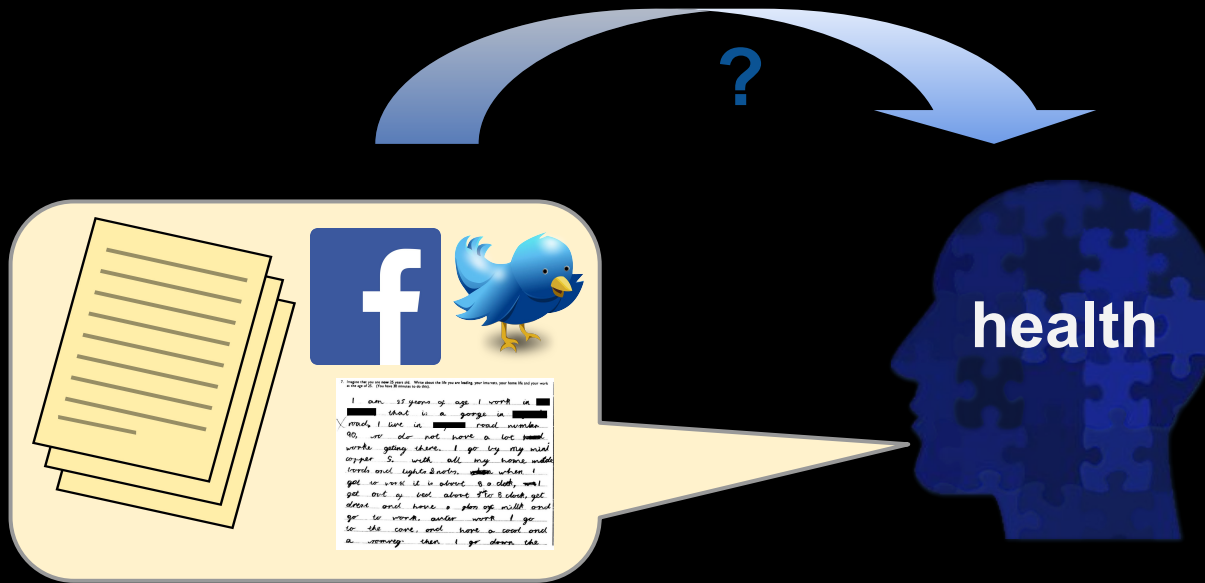
1. Diseases of heart
2. Malignant neoplasms (cancers)
3. Chronic lower respiratory
4. Cerebrovascular diseases (strokes)
5. Accidents, unintentional

6. Alzheimer's disease
7. Diabetes melitus
8. Kidney Diseases
9. Influenza & Pneumonia
10. Intentional self-harm (suicide)

11. Septicemia
12. Liver Disease
13. Hypertension
14. Parkinson's
15. Pneumonitus

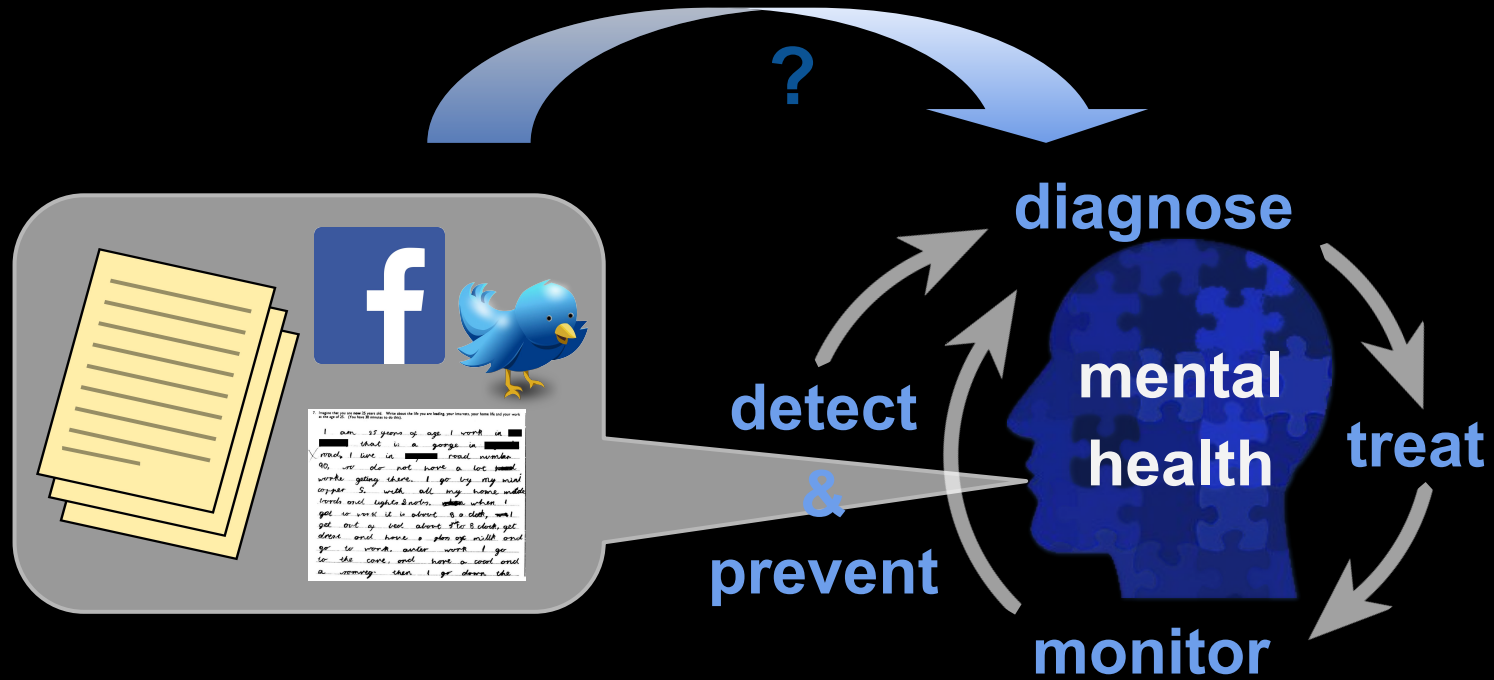


# Language-based Assessment



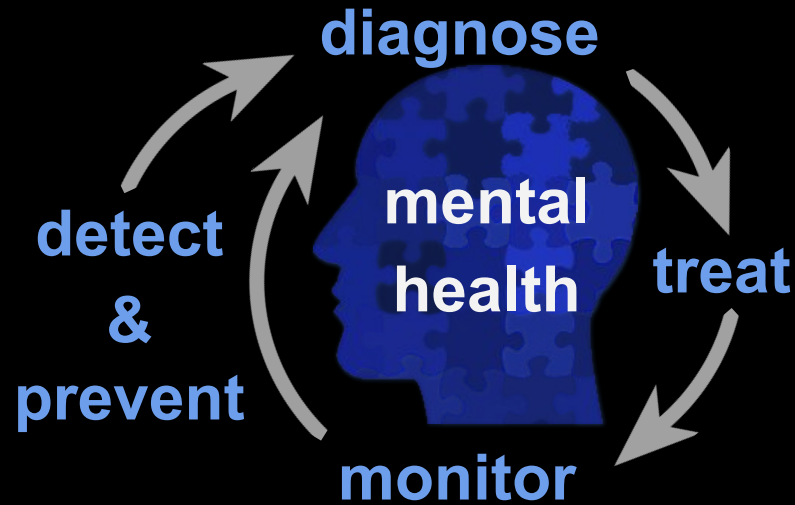
- inexpensive, behavior-based (different biases than questionnaires)
- reliable: test-rest; external validity
- signals many different conditions / behaviors / psychological traits and states
- captures nuance -- nearly infinite conceptual resolution

# Toward Forecasting in Mental Health

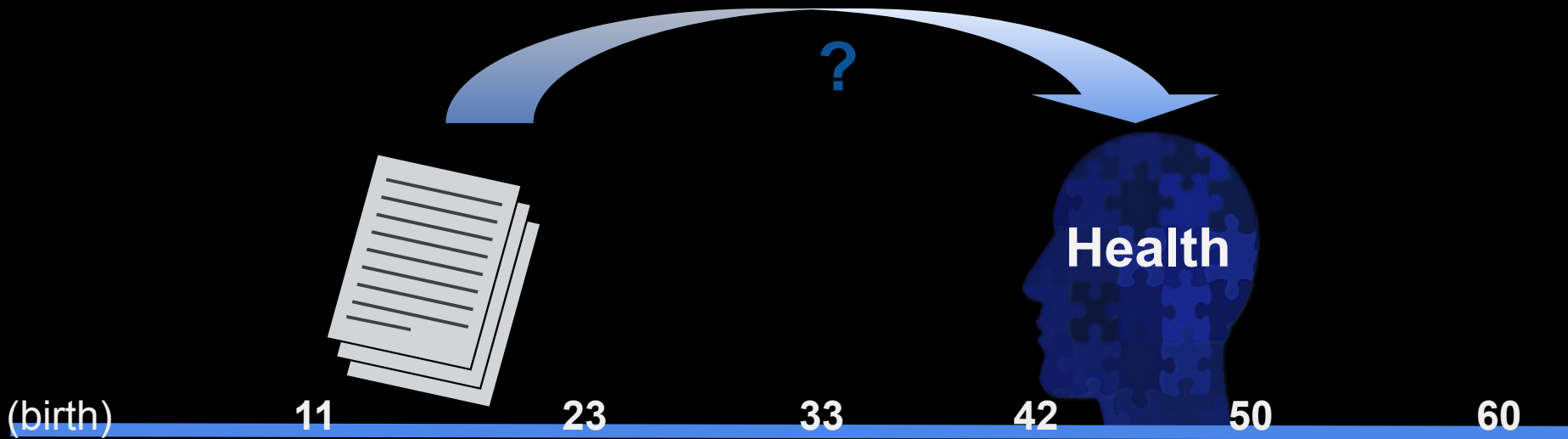




# Toward Forecasting in Mental Health



# Long-term Health Forecasting



- Early assessment of lifetime risk
- Targeted interventions for prevention and treatment

Joint Work with Veronica Lynn (SBU), Alissa Goodman (University College London), Margaret Kern (Univ. of Melbourne)

# National Child Development Study (NCDS)

Imagine that you are now 22 years old. Write about the life you are leading, your interests, your home life and your work at the age of 22. (You have 30 minutes to do this).

I am 22 years of age I work in [redacted] that is a garage in [redacted] road number 90, we do not have a lot of work getting there. I go by my mini copper 5. with all my home made boards and lights & notes. when I get to work it is about 8 o'clock, I get out of bed about 7:30, get dressed and have a glass of milk and go to work. after work I go to the care, and have a salad and a washing. then I go down the



(birth) 11 23 33 42 50 60

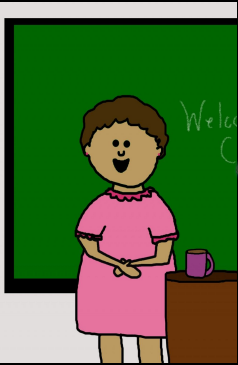
# National Child Development Study (NCDS)

Imagine that you were now 22 years old. Write about the life you are leading, your interests, your home life and your work at the age of 22. (You have 30 minutes to do this).

I am 22 years of age I work in [redacted] that is a garage in [redacted] road. I live in [redacted] road number 90, we do not have a lot of work getting there. I go by my mini copper. I with all my home made breads and light & notes. when I get to work it is about 8 o'clock, I get out of bed about 7:30 o'clock, get dressed and have a glass of milk and go to work. after work I go to the care, and have a salad and a washing. then I go down the



(birth) 11 23 33 42 50 60



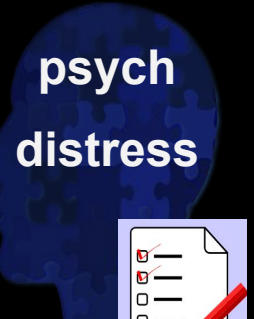
# National Child Development Study (NCDS)

Imagine that you are now 22 years old. Write down the 90 you are feeling, your interests, your best life and your work at the age of 22. (You have 30 minutes to do this).

I am 22 years of age I work in [redacted] that is a garage in [redacted] road number 90, we do not have a lot of work getting there. I go by my mini copper 5. with all my home made boards and lights & notes. when I get to work it is about 8 o'clock, I get out of bed about 4:30, get dressed and have a glass of milk and go to work. after work I go to the care, and have a meal and a washing. then I go down the



(birth)                      11                      23                      33                      42                      50                      60



# National Child Development Study (NCDS)

Imagine that you are now 22 years old. Write about the life you are leading, your interests, your best life and your worst at the age of 22. (This has 30 minutes to do this.)

I am 22 years of age I work in [redacted] that is a garage in [redacted] road. I live in [redacted] road number 90, we do not have a lot of work getting there. I go by my mini copper. I with all my home made breads and light notes. when I get to work it is about 8 o'clock, I get out of bed about 7:30, get dressed and have a glass of milk and go to work. after work I go to the care, and have a salad and a washing. then I go down the



Predict?

(birth) 11 23 33 42 50 60

BSAG

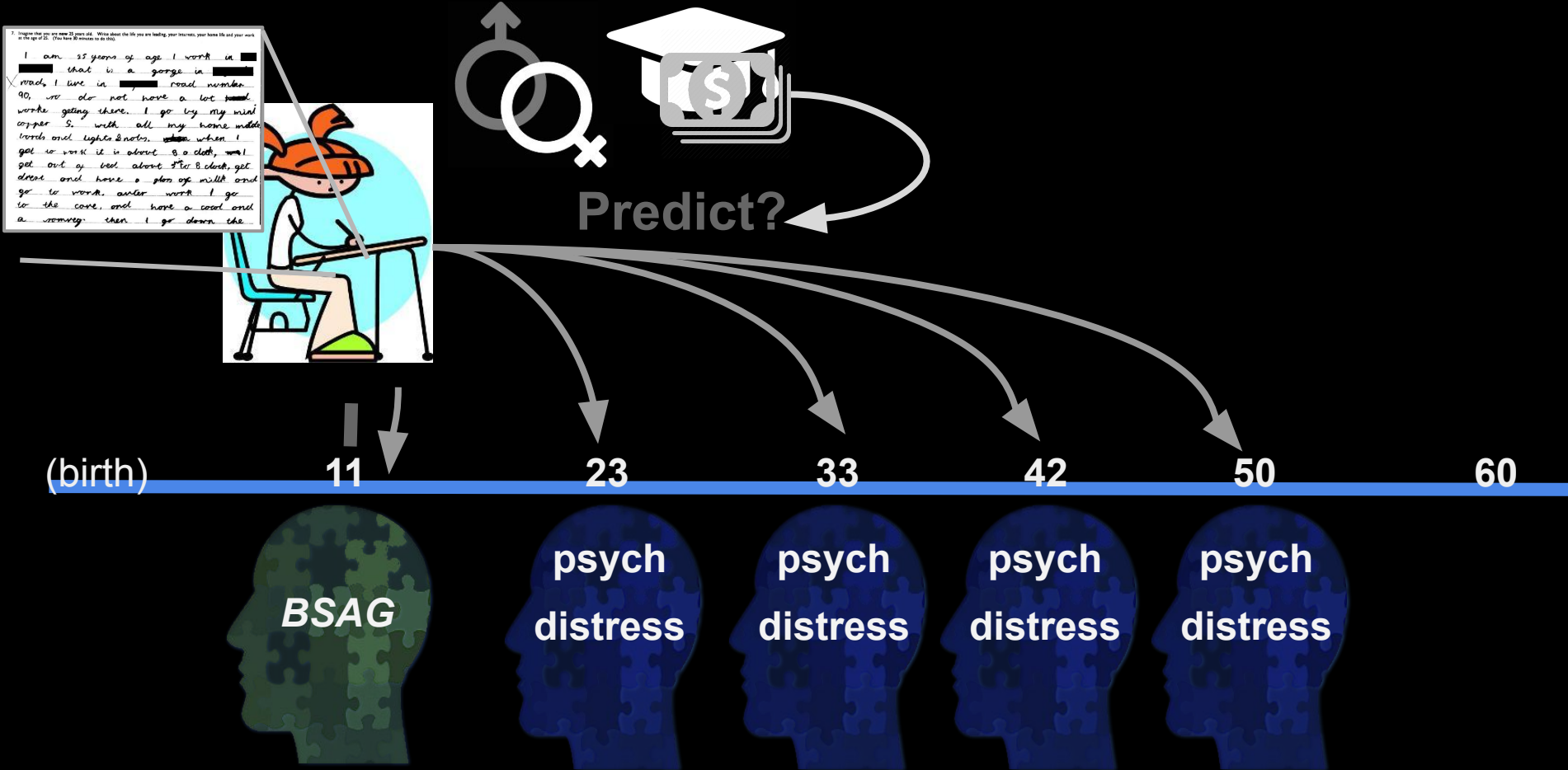
psych  
distress

psych  
distress

psych  
distress

psych  
distress

# National Child Development Study (NCDS)



# CLPSYCH 2018

Computational Linguistics and Clinical Psychology Workshop

[Workshop Description](#)

[Past Proceedings](#)

[Call for Papers 2018](#)

[Shared Task 2018](#) ▾

[Shared Task 2017](#)

## SHARED TASK 2018

# CLPsych-2018 Shared Task: Call for Participation

## Predicting Current and Future Psychological Health from Childhood Essays

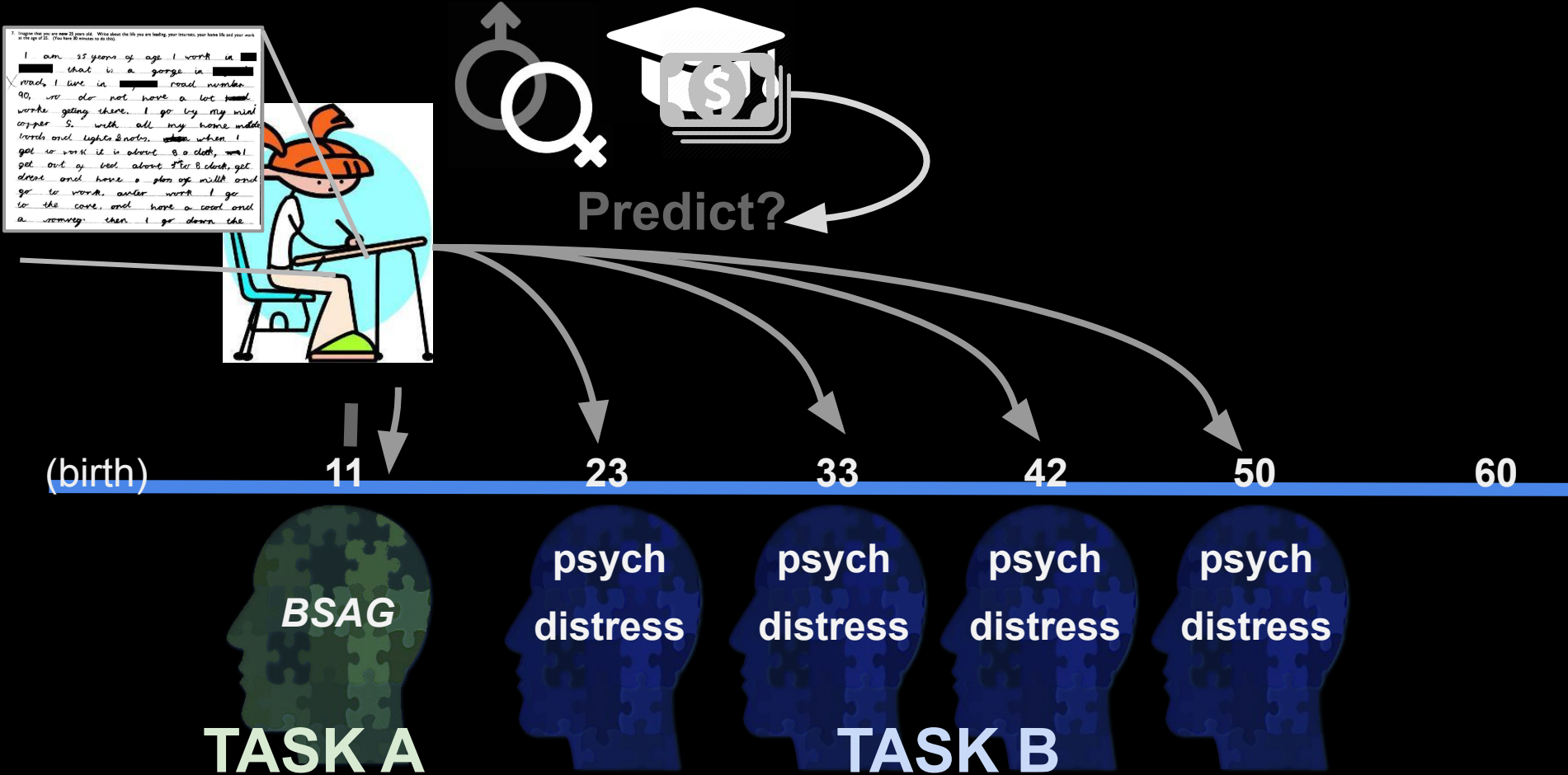
We invite participants for the 2018 CLPsych Shared Task.

### Motivation and Background

This shared task seeks to encourage new methods not only for analyzing current language use as a signal for mental health, as in previous CL Psych shared tasks, but also for understanding childhood



# National Child Development Study (NCDS)



# National Child Development Study (NCDS)

Imagine that you are now 22 years old. Write down the life you are leading, your interests, your home life and your work at the age of 22. (This has to be done in 60 mins.)

I am 22 years of age I work in [redacted] that is a garage in [redacted] road number 90, we do not have a lot [redacted] work getting there. I go by my [redacted] copper S. with all my home [redacted] [redacted] and lights & notes. [redacted] when I get to work it is about 8 o'clock, [redacted] I get out of bed about 7:30, I get dressed and have a glass of milk and go to work. After work I go to the [redacted] and have a [redacted] and a [redacted] then I go down the



Predict?

(birth) 11 23 33 42 50 60

BSAG

.56

**TASK A**

psych  
distress  
.43

psych  
distress  
.32

psych  
distress  
.21

psych  
distress  
.30\*

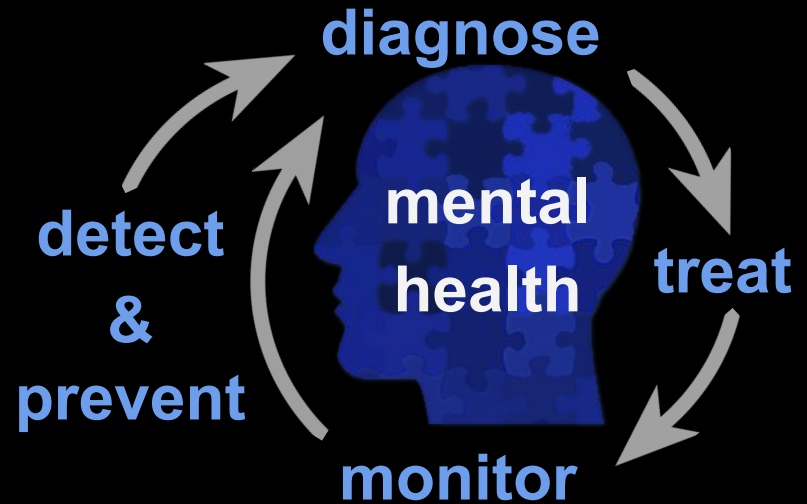
**TASK B**

Accuracy:

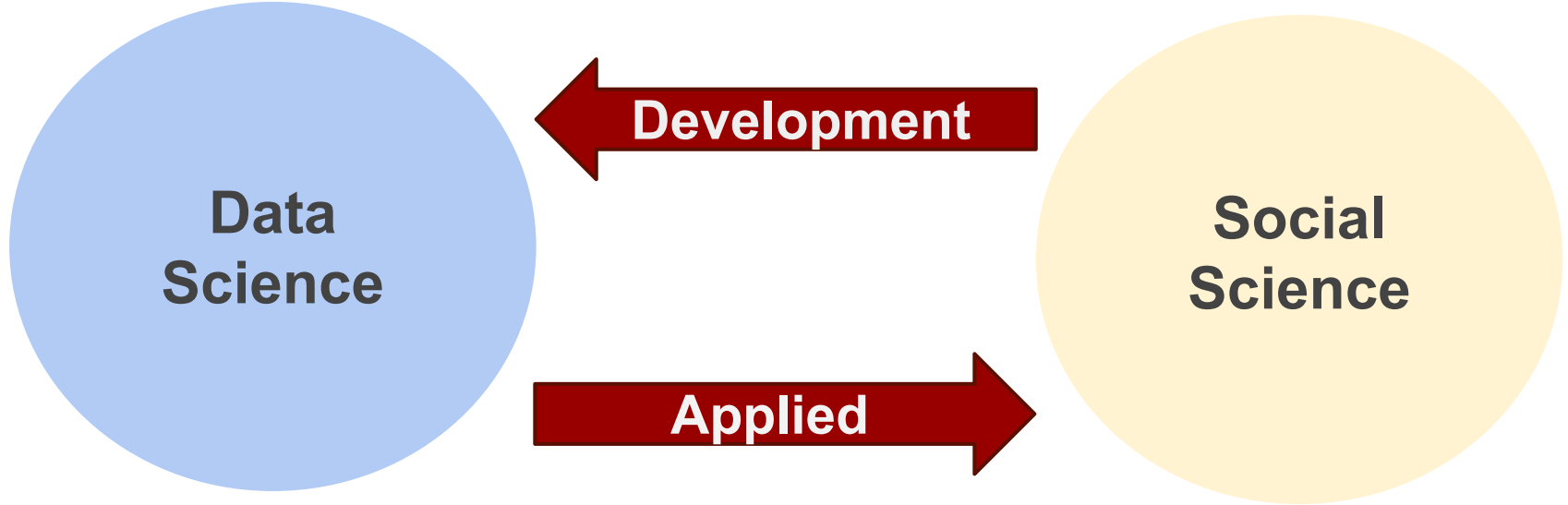
# Toward Forecasting in Mental Health

Language, a *human behavior*, often more predictive than traditional assessments alone.

- + study in people's "own words".
- + nearly infinite conceptual resolution.
- + longitudinal context.



# *How can your project make an impact?*



# *How can your project make an impact?*

General Big Picture Idea: <https://www.youtube.com/watch?v=v-zGHqMyd7o>

## Other Areas for Projects

- **Satellite Images for Crop Yield:**
  - <https://blogs.worldbank.org/developmenttalk/can-satellites-deliver-accurate-measures-crop-yields-smallholder-farming-systems>
  - <https://www.youtube.com/watch?v=NwRFTccODP4>
  - <https://www.mdpi.com/2313-433X/4/4/52/pdf>
- **Mobile Device Usage for Migration Patterns**
  - <https://migrationdataportal.org/themes/big-data>
- **Calls and Texts for Tracking Poverty**
  - <https://www.futurity.org/data-poverty-maps-1628812/>
  - <https://www.kdnuggets.com/2015/03/how-big-data-can-improve-lives-poor.html>
  - <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.408.9095&rep=rep1&type=pdf>
- **Social and Information Networks for Disaster Response**
  - <https://www.technologyreview.com/s/523711/how-information-flows-during-emergencies/>

UN's Inventory of Big Data and SDG Projects:

<https://unstats.un.org/bigdata/inventory>

# *How can your project make an impact?*

- Teams of 3 to 4 students.
- Towards solving or measuring a sustainable development goal.
- 2 teams per goal
- The projects must utilize:
  - 2 data pipeline (MapReduce, Spark, Tensorflow)
  - 2 other concepts from the course
- Components:
  - (1) a brief proposal presentation,
  - (2) analysis code
  - (3) a report, and
  - (4) a final presentation (Poster or Oral)

# *How can your project make an impact?*



# SUSTAINABLE DEVELOPMENT GOALS

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| <b>1</b> NO POVERTY<br>                  | <b>2</b> ZERO HUNGER<br>                     | <b>3</b> GOOD HEALTH AND WELL-BEING<br>              | <b>4</b> QUALITY EDUCATION<br>                         | <b>5</b> GENDER EQUALITY<br>                     | <b>6</b> CLEAN WATER AND SANITATION<br>              |
| <b>7</b> AFFORDABLE AND CLEAN ENERGY<br> | <b>8</b> DECENT WORK AND ECONOMIC GROWTH<br> | <b>9</b> INDUSTRY, INNOVATION AND INFRASTRUCTURE<br> | <b>10</b> REDUCED INEQUALITIES<br>                     | <b>11</b> SUSTAINABLE CITIES AND COMMUNITIES<br> | <b>12</b> RESPONSIBLE CONSUMPTION AND PRODUCTION<br> |
| <b>13</b> CLIMATE ACTION<br>           | <b>14</b> LIFE BELOW WATER<br>             | <b>15</b> LIFE ON LAND<br>                         | <b>16</b> PEACE, JUSTICE AND STRONG INSTITUTIONS<br> | <b>17</b> PARTNERSHIPS FOR THE GOALS<br>       | <br><b>SUSTAINABLE DEVELOPMENT GOALS</b>            |

# SDG Related Data Sets

| Name               | Desc.  | Metadata                          | Link   | SDG Goal |
|--------------------|--|-----------------------------------|--|----------|
| Air Pollution Data | Per country air pollution  | 2977 Towns                        | Multiple Sources: <a href="https://ourworldindata.org/air-pollution">https://ourworldindata.org/air-pollution</a><br><a href="http://www.who.int/phe/health_topics/outdoorair/databases/AAP_database_summary_results_2016_v02.pdf">http://www.who.int/phe/health_topics/outdoorair/databases/AAP_database_summary_results_2016_v02.pdf</a> | SDG 3    |
| GHO Data           | Health indicators including, mortality, burden of disease, life expectancy     |                                   | <a href="https://www.who.int/gho/mortality_burden_disease/en/">https://www.who.int/gho/mortality_burden_disease/en/</a>  | SDG 3    |
| Child Mortality    | Child mortality rate   | 210 countries                     | Multiple Sources: <a href="https://ourworldindata.org/child-mortality#data-quality-definition">https://ourworldindata.org/child-mortality#data-quality-definition</a>  | SDG 3    |
| Poverty and Equity | Helps in understanding the evolution of poverty in various countries           | 5394 rows, 31 poverty indicators  | <a href="http://povertydata.worldbank.org/poverty/home/">http://povertydata.worldbank.org/poverty/home/</a>  | SDG 1    |
| Water Quality      | Ensure availability and sustainable management of water and sanitation for all | 550246 rows, 5 quality parameters | <a href="https://data.gov.in/catalog/water-quality-affected-habitations">https://data.gov.in/catalog/water-quality-affected-habitations</a>  | SDG 6    |
| Access to water    |  | 181 countries                     | <a href="https://data.unicef.org/topic/water-and-sanitation/drinking-water/">https://data.unicef.org/topic/water-and-sanitation/drinking-water/</a>  | SDG 6    |
| WDI data           | World development indicators   | 84000 rows, 300 features          | <a href="https://data.world/worldbank/sustainable-development-goals/workspace/file?filename=SDG_csv_en%2FSDG_Data.csv">https://data.world/worldbank/sustainable-development-goals/workspace/file?filename=SDG_csv_en%2FSDG_Data.csv</a>  | SDG 1    |
| Quality Education  | Quality of education in countries  | 291 countries 1600+ features      | <a href="http://data.uis.unesco.org/">http://data.uis.unesco.org/</a>  | SDG 4    |
| Depression Data    | Various mental health factors in USA   | 1986-2016, 5486 features          | <a href="https://qssdataexplorer.norc.org/variables/vfilter">https://qssdataexplorer.norc.org/variables/vfilter</a>  | SDG 3    |



