

Community-level modeling

Open vocabulary

• Correlate word or LDA topic use with community sentiment

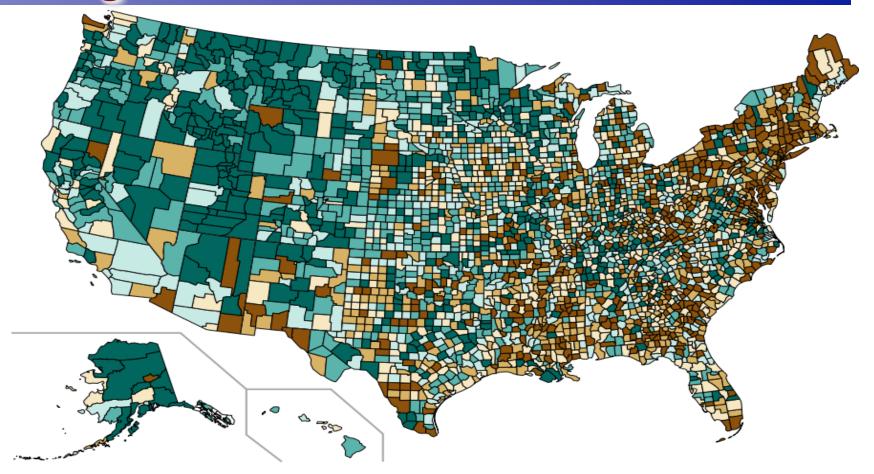
Model-based

- Build language models of sentiment
 - at the tweet or individual level
- Apply them to new tweets or people
 - Group by community
- (Sometimes) correlate with outcome

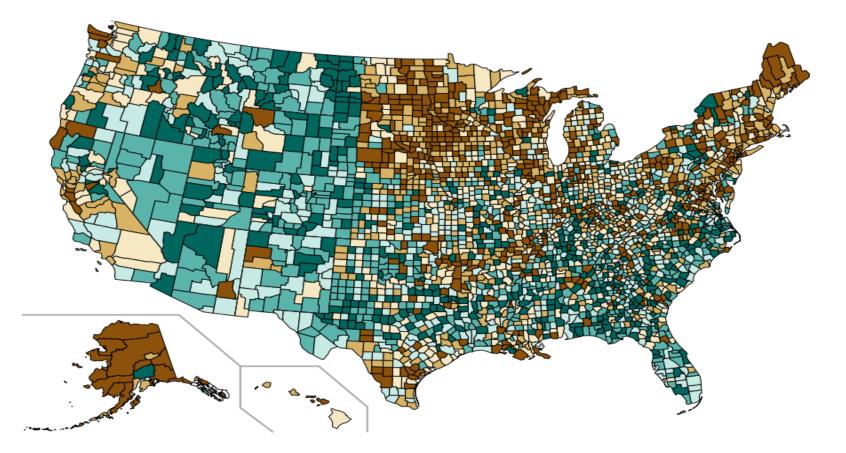
Measuring personality in tweets

- Create regression models for personality
- Use these to estimate scores on tweets collected on the county level
 - Agreeableness
 - Conscientiousness
 - Extroversion
 - Neuroticism
 - Openness to experience

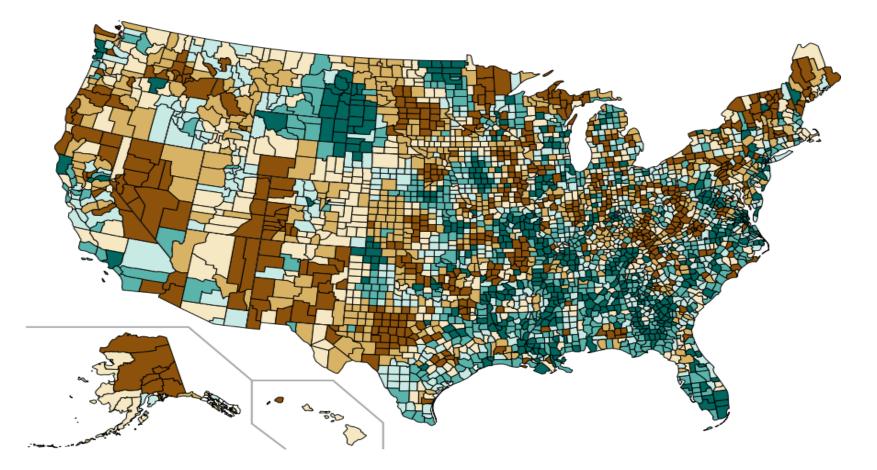
Agreeableness



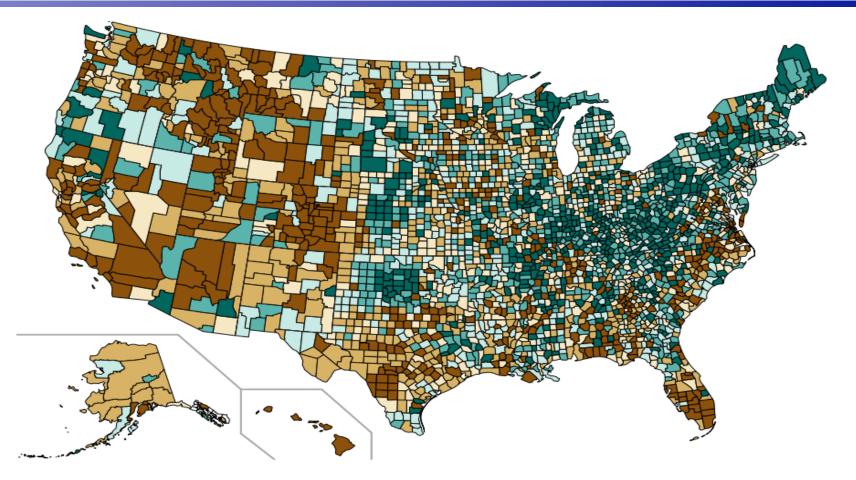
Conscientiousness



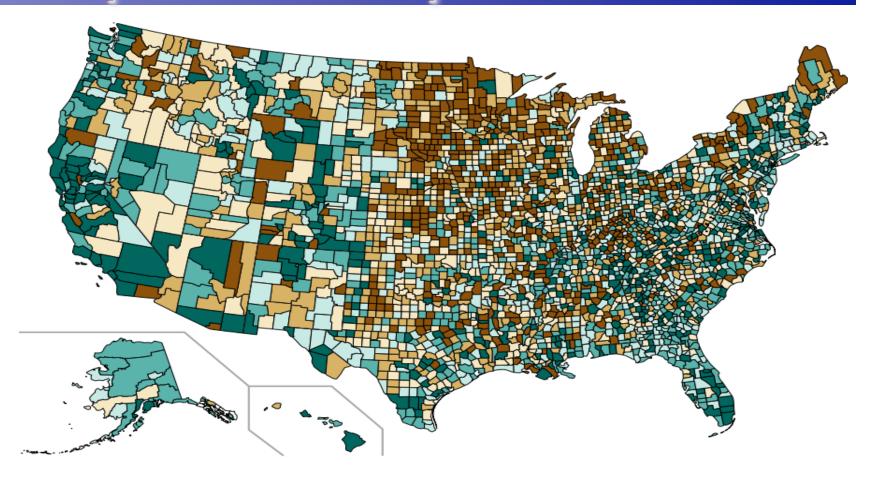


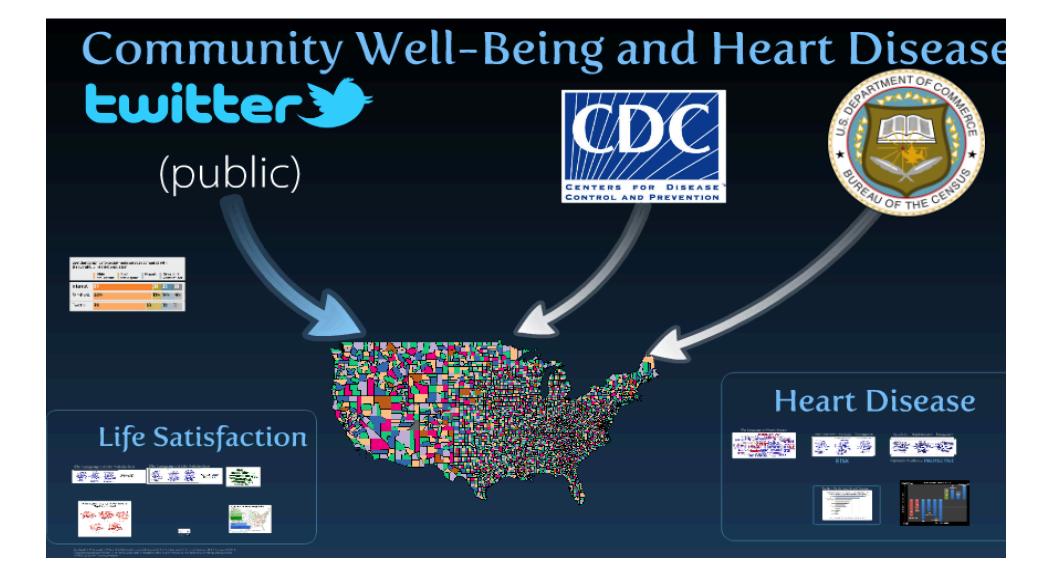


Neuroticism



Openness to experience





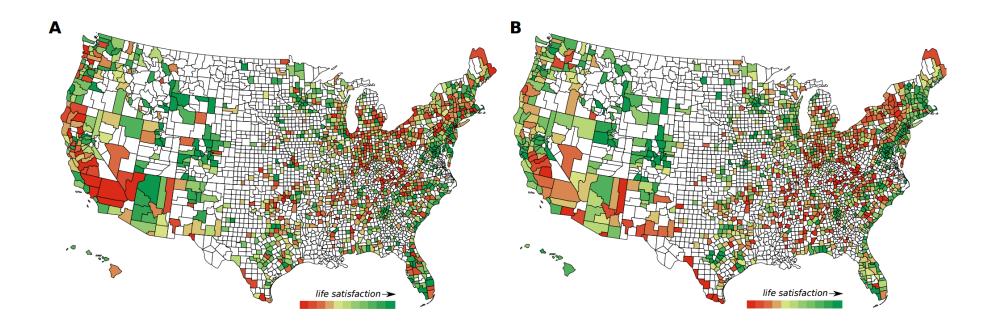
Satisfaction with Life Prediction

data	r
Lexica	0.264
Topics	0.307
Topics & Lexica	0.307
Controls	0.435
Controls, Topics & Lexica	0.535

county-level predictions

<u>Controls</u> median age sex (percentage female) minorities (percentage black and Hispanic). median household income (log-transformed) educational attainment (% high school grads or higher; % BS or higher).

Satisfaction with Life



survey

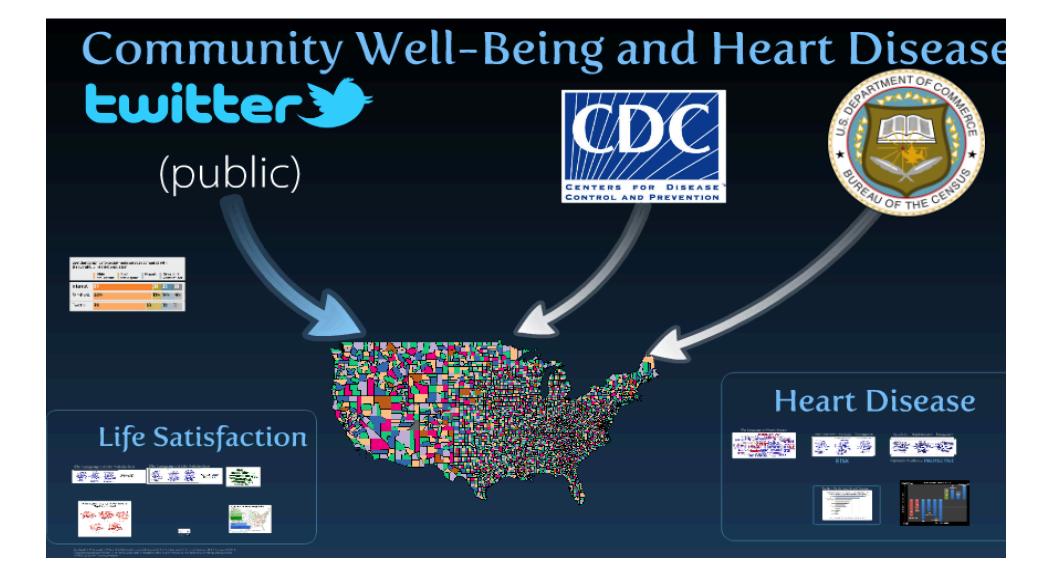
predicted

Life Satisfaction Words

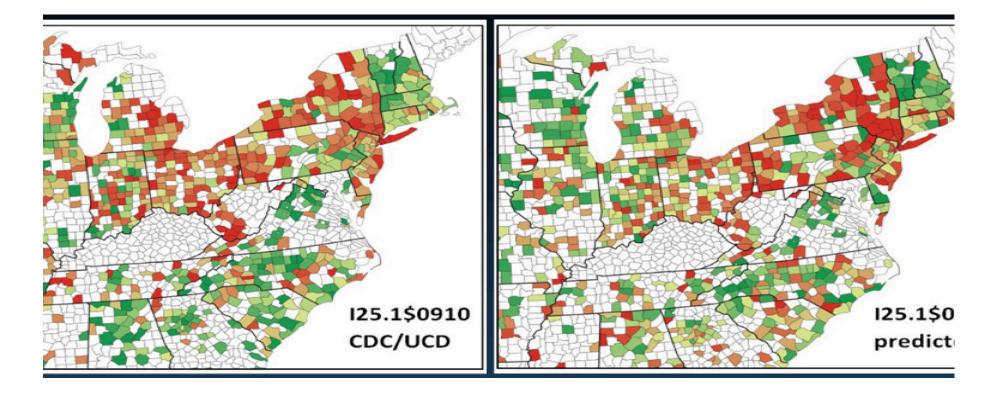
zumba basic training personalvista personalcertified intense potty class session trainer fitness sessionsgym workout	figure ideas idea suggestion helpful suggestions appreciated greatly advice decide tips	haiti.benefit donationraise_donated moneydonate charity Support cancer donations raised reliefhelping fund	blockstakes building foundation builtconstruction builtconstruction ground brick noah	physical universe compassion human experience nature reality spiritual sense individual humans divine
meetingstudent convention leadership meetings centerCouncil staff students boardattend conference group	houseplace apartment_interested bedroom_signed_area monthmoving_apt move lease	technology businesslearning process information communicationeducation marketing management engineering analysis skills development design research	Customer company Customer entertainment announcement customers community gravity servicessuggestions public	judgementpleasant experienced.judgment exciting experiences experience changing journey .wonderfulshare learning painful bound esjevitte
sailing wave waves ocean drowningdeepboat swim ship sail sinking water waters sea hill view tent swimmingblast weekend hikinglake beachcamping trip fishing cabin		1	stressed boredfreakin _{bore} tired.tire ughh ughhh IM b soooosooo sooooughSOO effin	boredom extremelyentertained hmu entertainsooooooo oring entertainment bore bored textincredibly insanely

z, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Lucas, R. E., Agrawal, M., Park, G. J., Lakshmikanth, S. K., Jha, S., Seligman, M. E. gar, L. H. (2013). Characterizing Geographic Variation in Well-Being using Tweets. In *Proceedings of the Seventh International AAAI* nce on Weblogs and Social Media (ICWSM). Boston, MA.

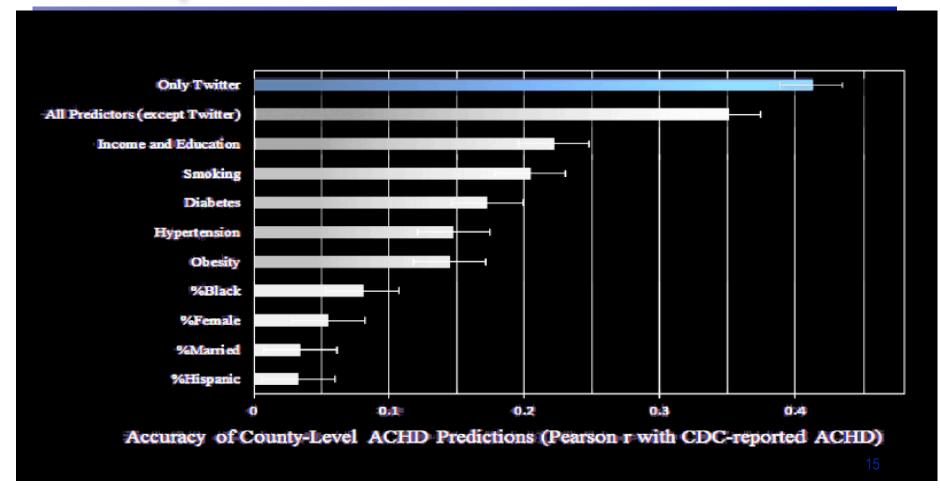
12



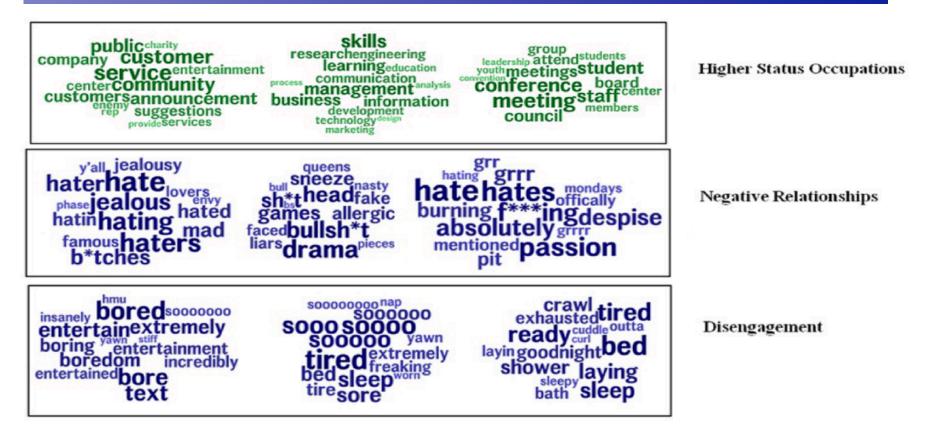
Twitter predicts cardiovascular disease



Twitter predicts cardiovascular disease



Cardiovascular Disease Words



Person & Community Conclusions

Language reveals demographics, personality

• Also psychopathy, emotion, SES, political orientation, happiness, depression, health, disease

Language models generalize

- Across individuals
- From Facebook to Twitter
- From Individuals to Counties

Language allows data-driven hypothesis generation

• As well as hypothesis testing