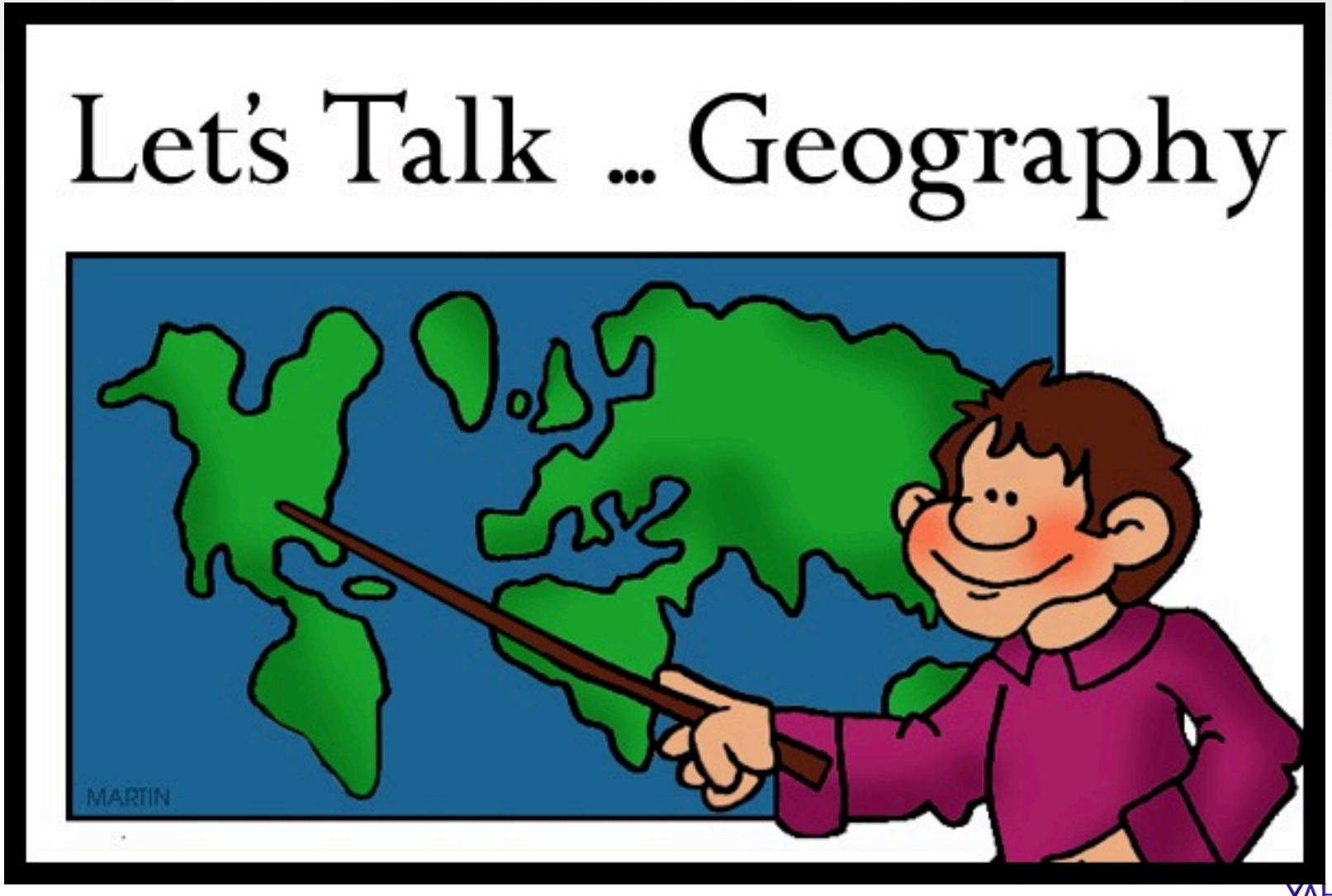
The Geography of Online News Engagement

Martin Saveski, MIT Media Lab, Cambridge, USA
Daniele Quercia, Yahoo Labs, Barcelona, Spain
Amin Mantrach, Yahoo Labs, Barcelona, Spain



YAHOO!

Motivations

- Despite its importance,
- the **geographic processes of online engagement** on news platforms have not been widely studied;

- → **Yahoo News** site for more than **two years**: articles and user comments.

Contributions

In Online news platforms,

Users engage with each other depending on where they live

Contributions: Socio economic factors and user engagement

Users in states with:

*high levels of **education** and **well-being** comment articles about **research** & **technology**, but not **politics**, **gossip** or **sport**;*

*high levels of **crime** and **unemployment** comment on articles about **sports**, but not those about **economy** or **research** and **technology**;*

YAHOO!

Contributions: Personality Traits and user engagement

Users from

*states **low** in **neuroticism** (emotionally stable) comment on articles about **music** ;*

***open and extravert** states comment on articles about **sport**;*

***conscientious** states on articles about **economics**.*

YAHOO!

Related work

- **Influence of time** on our actions online:
  - "***emotion words*** *by Twitter users influenced by* ***times zones***"

- **News** in tweets and **geographics spread** on Twitter:
  - "**Reciprocal relations** between people who live **no more than 3 times zones away**"
  - **Physical distance** constrained the spread of **hashtags**

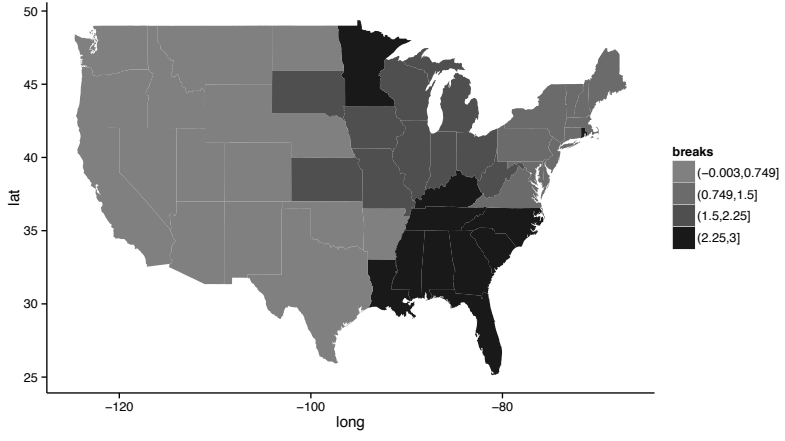
YAHOO!

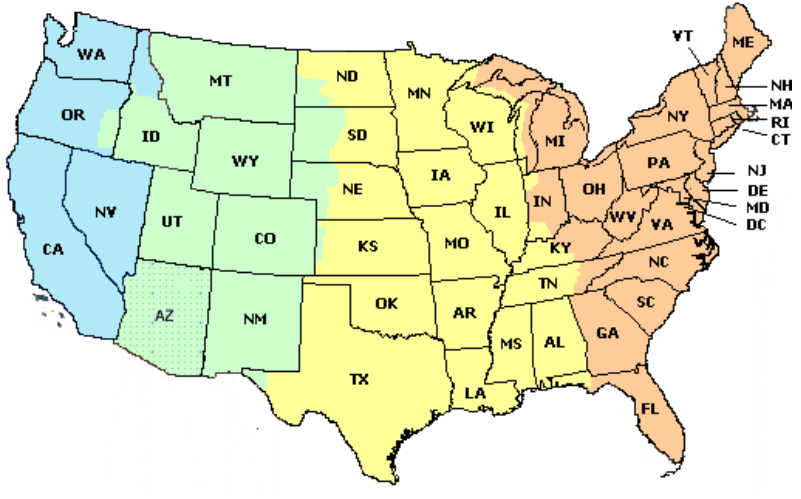
Data description

- Random sample of **200K news articles** and **41M associated comments**;
- Published from August 2010 to February 2013;
- On Yahoo! News US;
- English articles;
- Sources: Reuters, ABC News, AP, etc.;
- From anonymous user we have **IP address → state**(Yahoo places Web service).

State commenting graph

- Nodes: US states
- Egde weight: number of times two users in state i and j comment on the same article

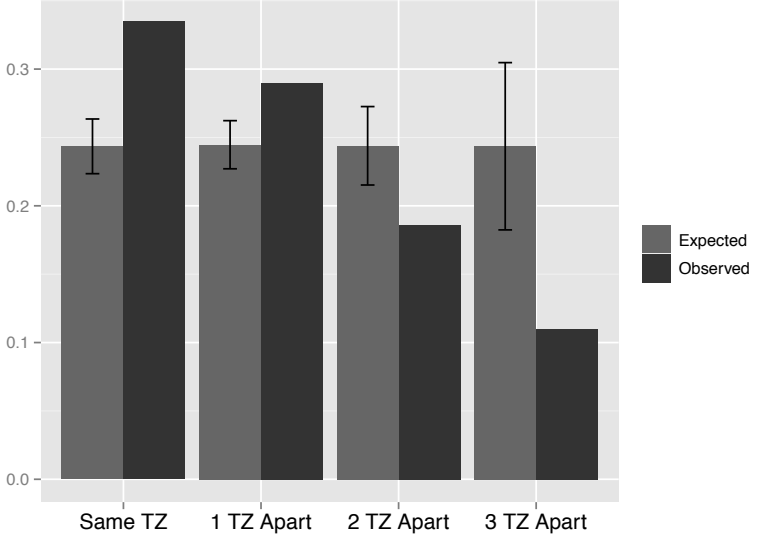




YAHOO!

The time zone effect

- Users in the same ***TZ*** preferentially engage with the same articles, while users in different time zones engage with different articles;
- **Engagement in k-time zone apart**: count the number of times users from k-time zone apart engage in the same articles;
- **Null model**: We shuffle user's time zone assignment.



YAHOO!

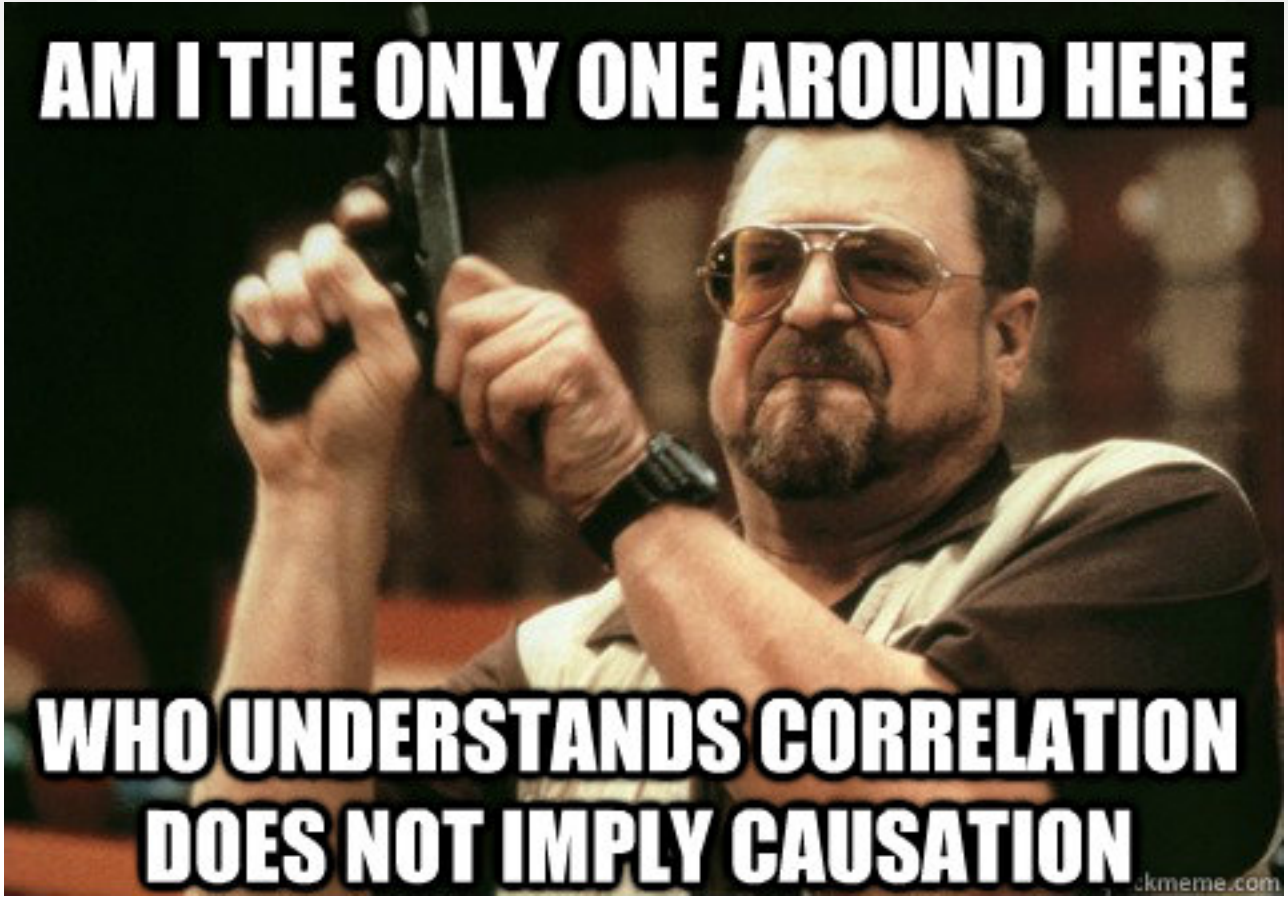
The Geography of News Engagement

- The gravity model

$$F_{i,j}^{est} = g\frac{m_i m_j}{d_{i,j}^2}$$

- **F** measures the estimated engagement between two states;
- **g** is a scaling constant;
- **$d_{i,j}$** is the euclian distance between two states centers;
- **m_k** is number of users of state k.

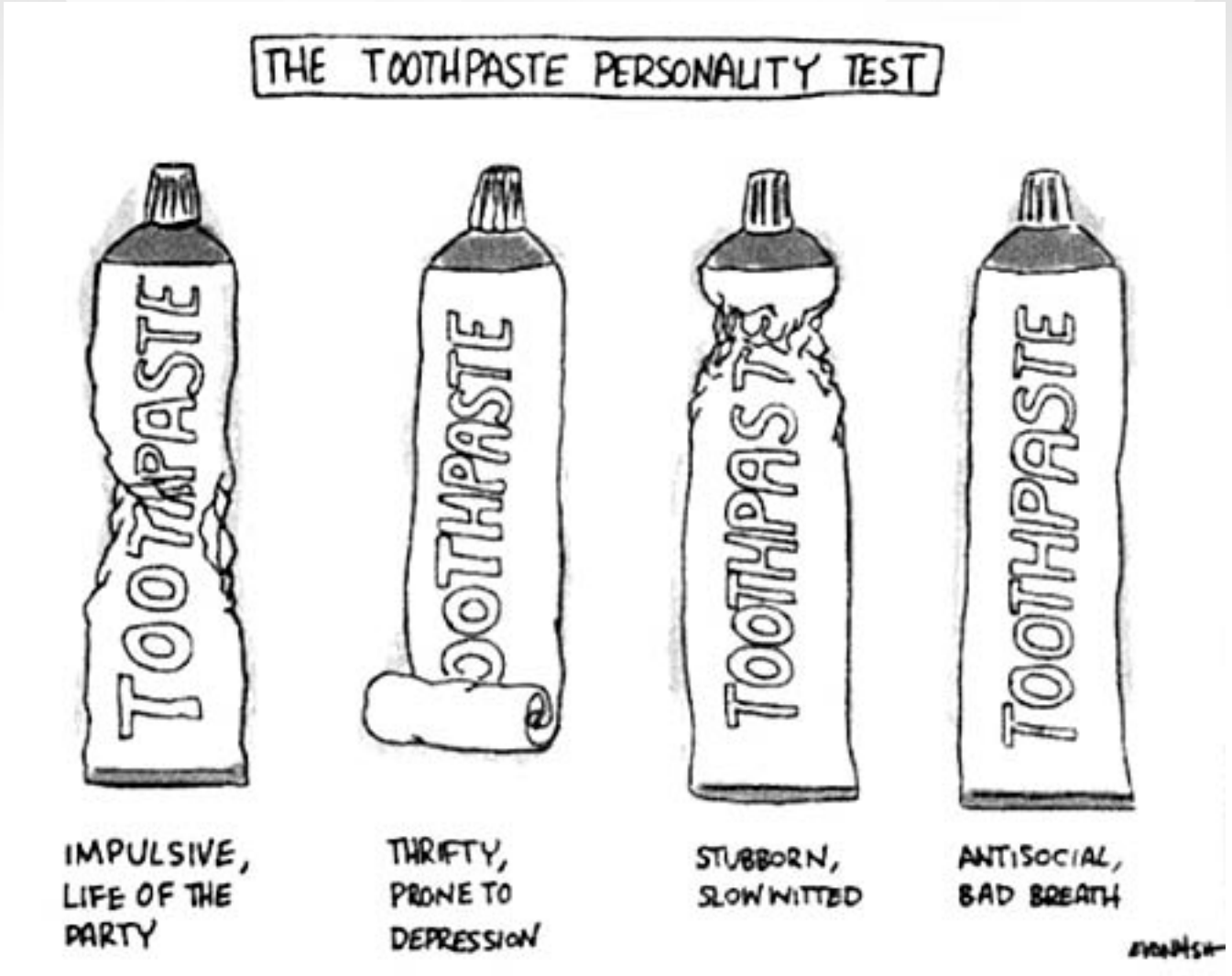
- F correlates with number of the observed number of comments with a **pearson correlation of 0.70**


AM I THE ONLY ONE AROUND HERE
WHO UNDERSTANDS CORRELATION
DOES NOT IMPLY CAUSATION
ckmeme.com

Assigning topic distributions to states

- 13.8% of articles are editorially labeled with IPTC categories
- IPTC consists of 1400 topics organized in a taxonomy;
- Average category per articles is 5;
- By using user engagement (on comments) we aggregated articles topics to form states topics:
  - Each time a user from state comment on an article, the tags contribute to the state topical distribution;
  - Tags contribution is normalized by the number of times they are used (to discount popular tags, similar to idf to discount stop words).

YAHOO!


THE TOOTHPASTE PERSONALITY TEST
TOOTHPASTE
TOOTHPASTE
TOOTHPASTE
TOOTHPASTE
IMPULSIVE, LIFE OF THE PARTY
THRIFTY, PRONE TO DEPRESSION
STUBBORN, SLOW WITTED
ANTISOCIAL, BAD BREATH

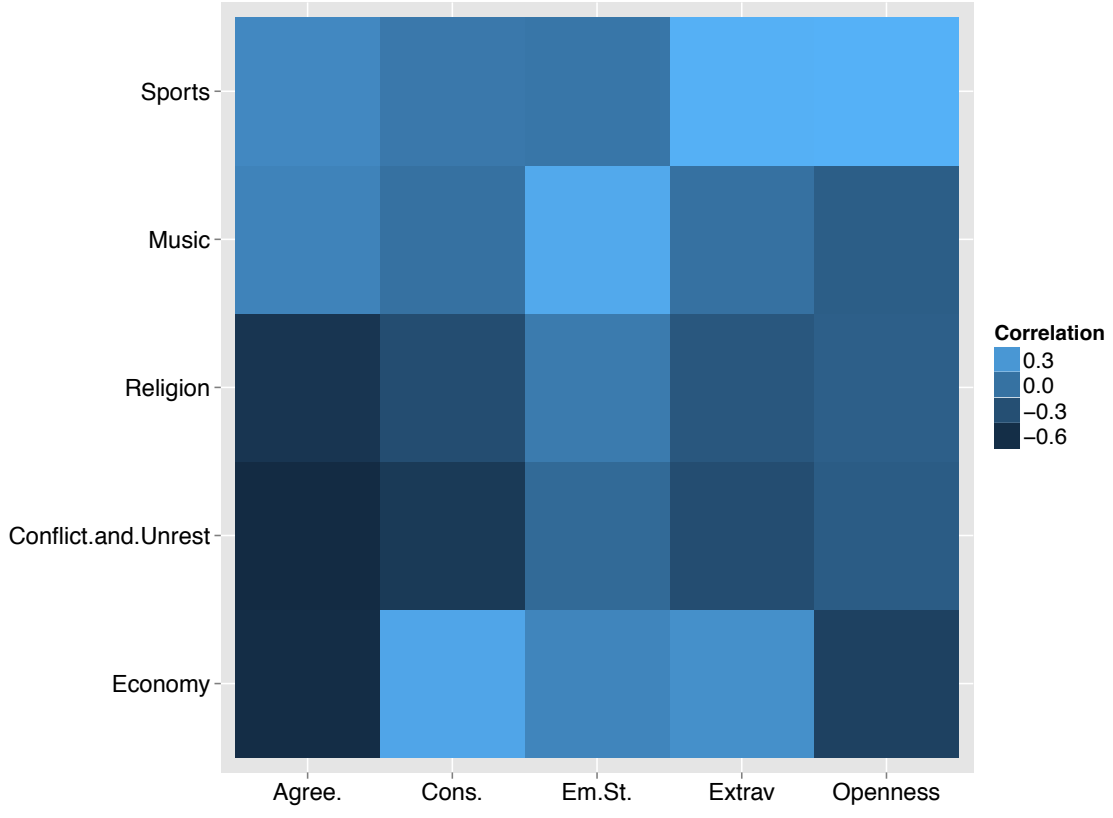
YAHOO!

The Big Five Personnality Traits

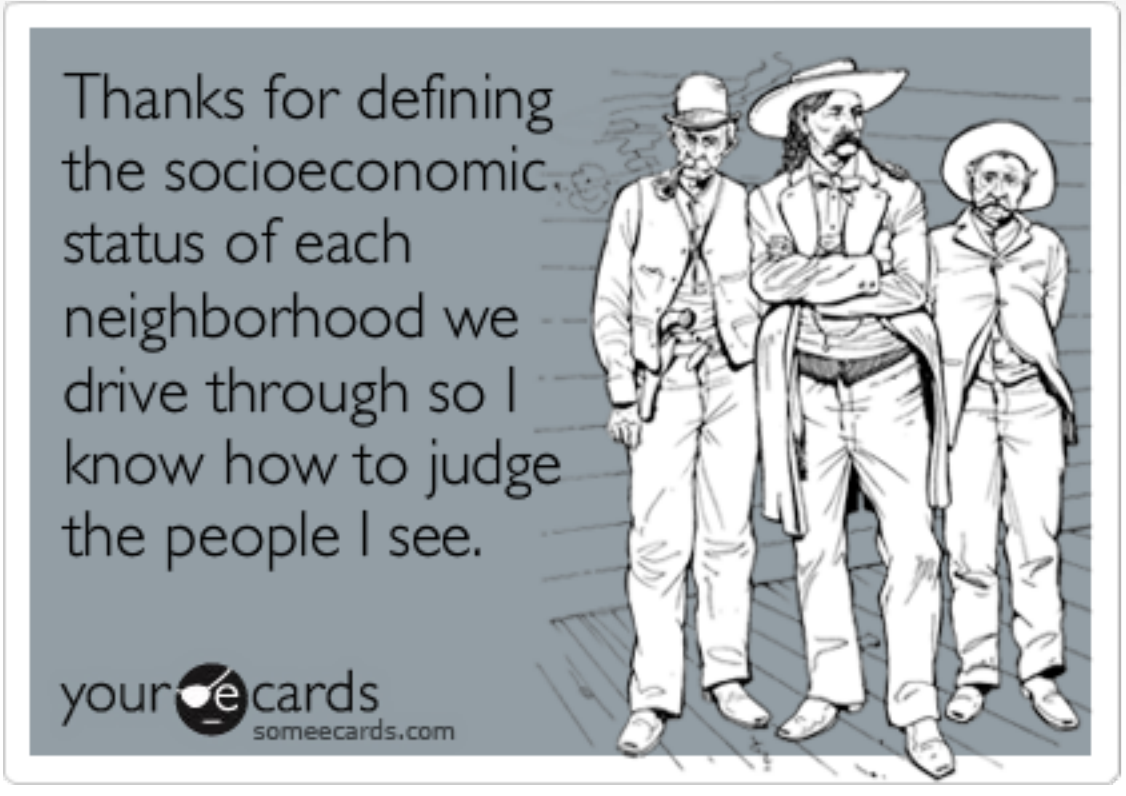
1. **Openess:** *Imaginative*, *spontaneous*, and *adventurous* individuals;
2. **Conscientiousness:** *Ambitious*, *resourceful* and *persistent* individuals;
3. **Extraversion:** Individuals who are *sociable* and tend to seek *excite-ment;*
4. **Agreeableness:** are *trusting*, *altruistic*, *tender-minded*, and are motivated to *maintain positive relationships* with others
5. **Neuroticism:** Finally, *emotionally liable* and *impulsive* individuals.

These factors have been **also studied at the state level** (strongly correlated with socio-economic indicators) [Rentfrow et al.]

Correlation of news user engagement factors with state personality scores



YAHOO!

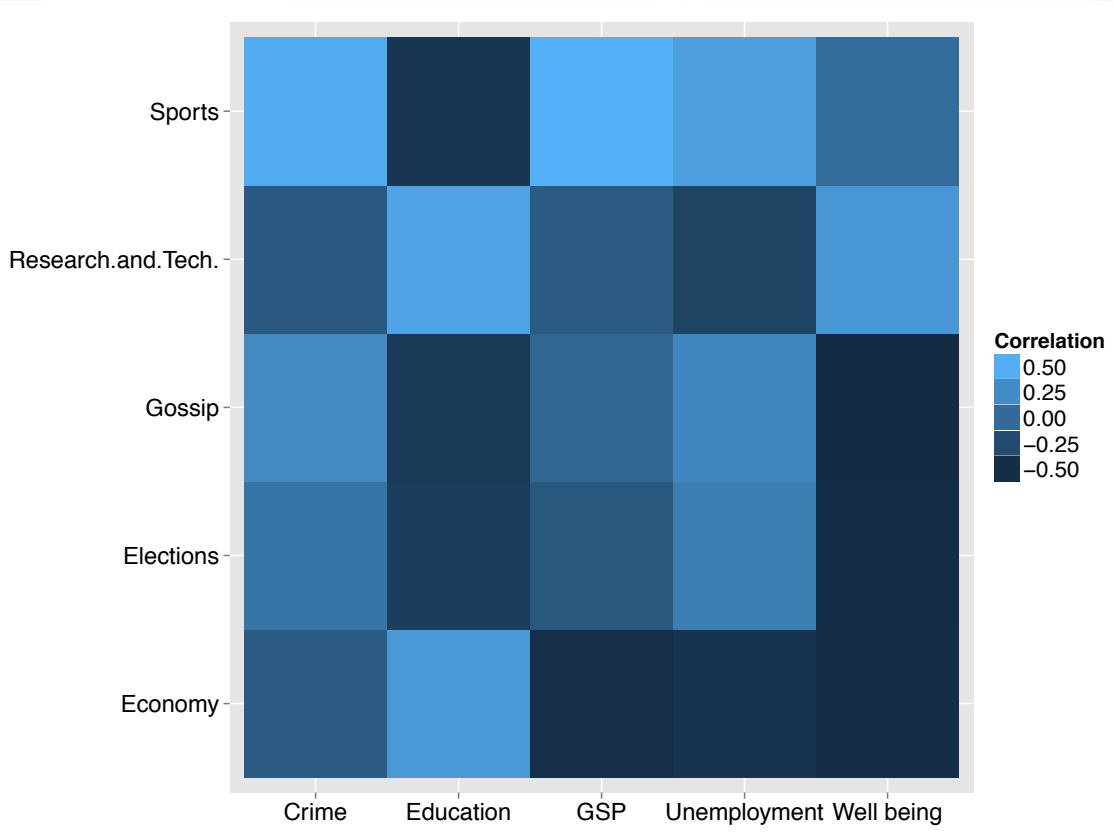

Thanks for defining the socioeconomic status of each neighborhood we drive through so I know how to judge the people I see.
yourecards
someecards.com

Socioeconomic Indicators

- 1. **Well-being** index;
- 2. **Crime** level
- 3. Rate of **unemployment**;
- 4. **Gross** State **Product**;
- 5. **Education** level.

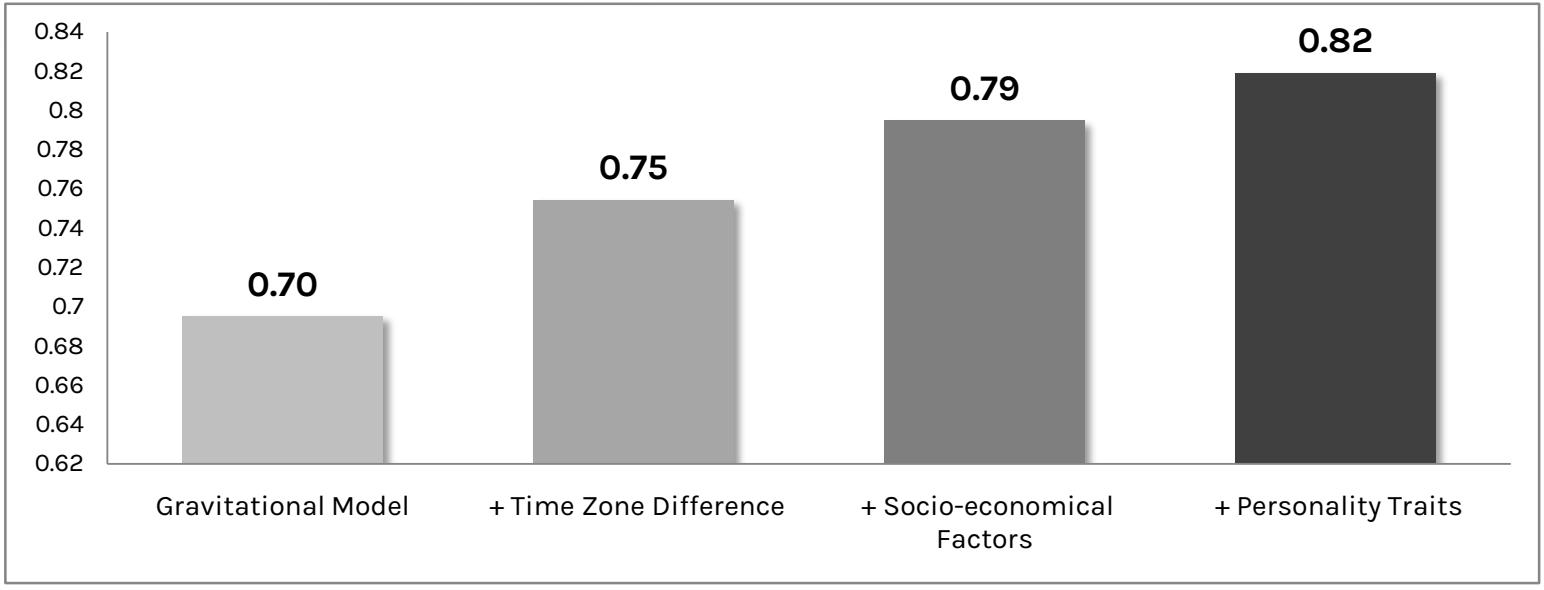
YAHOO!

Correlation of news user engagement factors with socio economic indicators



YAHOO!

Conclusions: Putting all together



YAHOO!

Questions