Toward Insight

- Growth of social media
  => unprecedented amounts of personal discourse
- Leveraged in computational linguistics mostly for prediction:
  - sentiment analysis
  - authorship attribution
  - personality prediction
    (Argamon 2005, 2009; Mairesse et al. 2006, 2007; Golbeck et al. 2011; Iocabelli et al. 2011)

Toward Insight

- Growth of social media
  => unprecedented amounts of personal discourse
- Leveraged in computational linguistics mostly for prediction:
  - sentiment analysis
  - authorship attribution
  - personality prediction
    (Argamon 2005, 2009; Mairesse et al. 2006, 2007; Golbeck et al. 2011; Iocabelli et al. 2011)
- Analysis of language in social media can also be used for the purpose of gaining psychological insights.

This Work

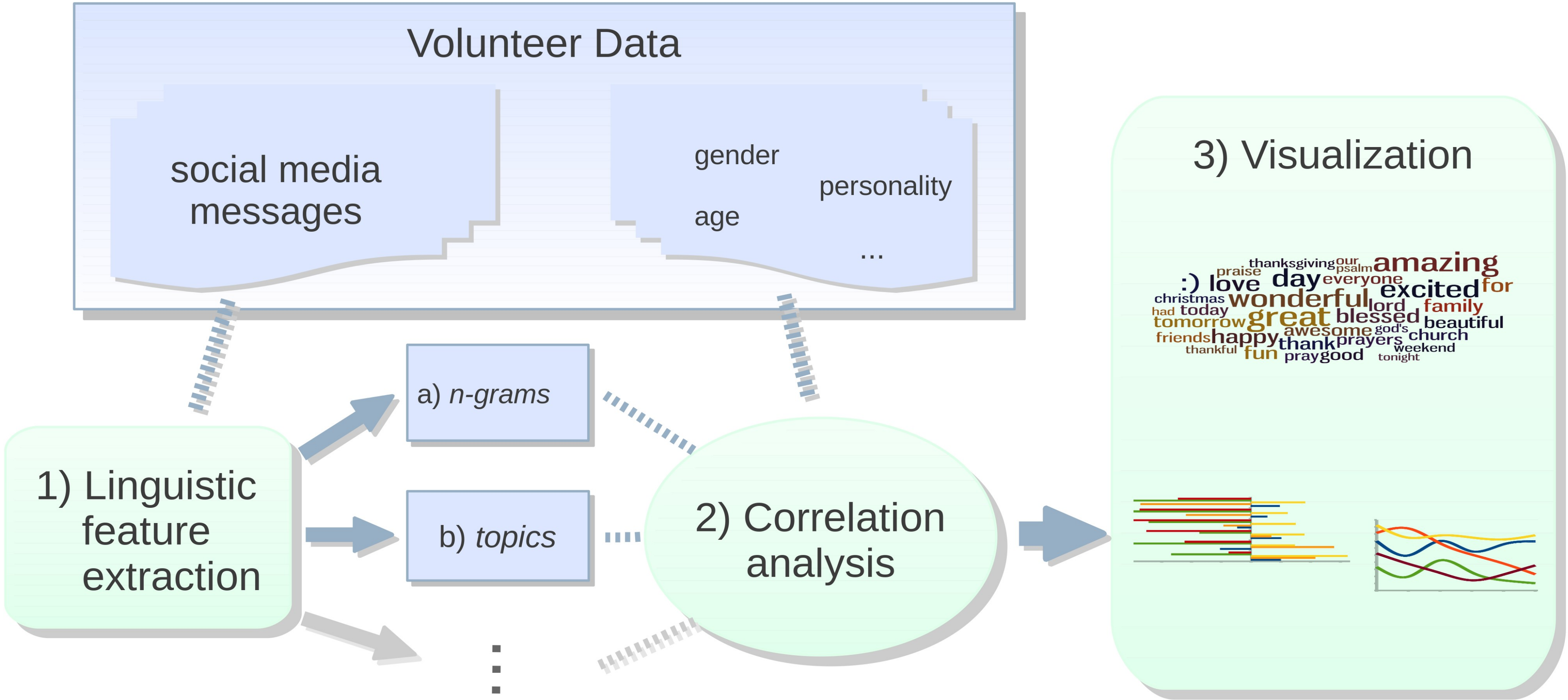
- Explores language in social media as a function of gender, age, and personality.

- Data Set over Facebook
  - ~75,000 volunteers, writing 14m status updates
    - wrote at least 1000 words
    - reported gender and age
    - took standard personality survey:
      “the big five” / five factor model.

      (MyPersonality App.)

Method

Differential Language Analysis



Volunteer Data

social media messages

gender

personality

age

...

1) Linguistic feature extraction

a) *n-grams*

b) *topics*

⋮

2) Correlation analysis

3) Visualization

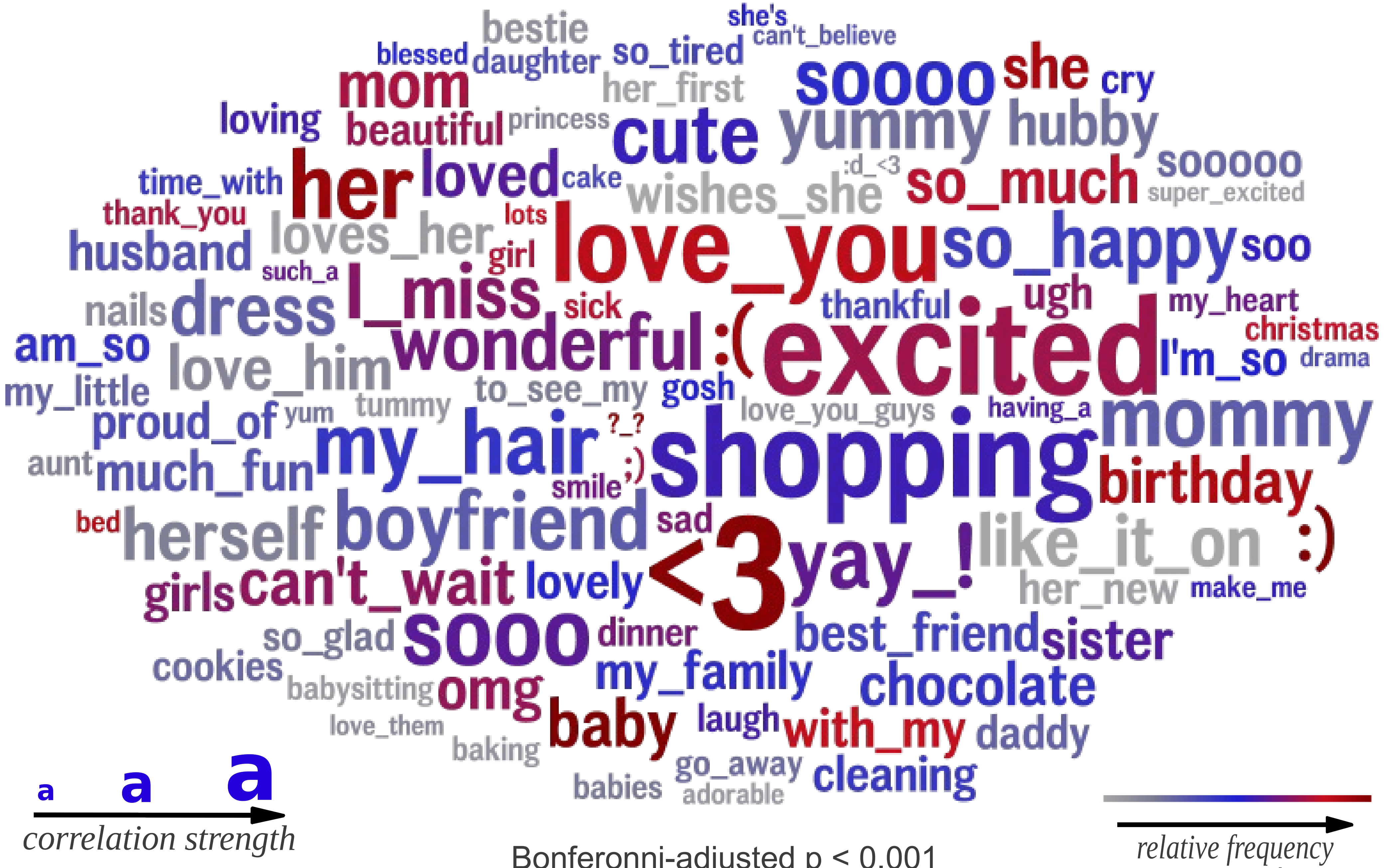
Features

- N-grams
  - emoticon-aware tokenization
  - collocation filter based on point-wise mutual information; used by at least 1% of volunteers
- LDA topics
  - Latent Dirichlet Allocation
  - 2000 topics from larger 20m status updates
  - usage defined as:

$$p(topic, person) = \sum_{tok \in topic} p(topic|tok) * p(tok|person)$$

Explicit Language Warning

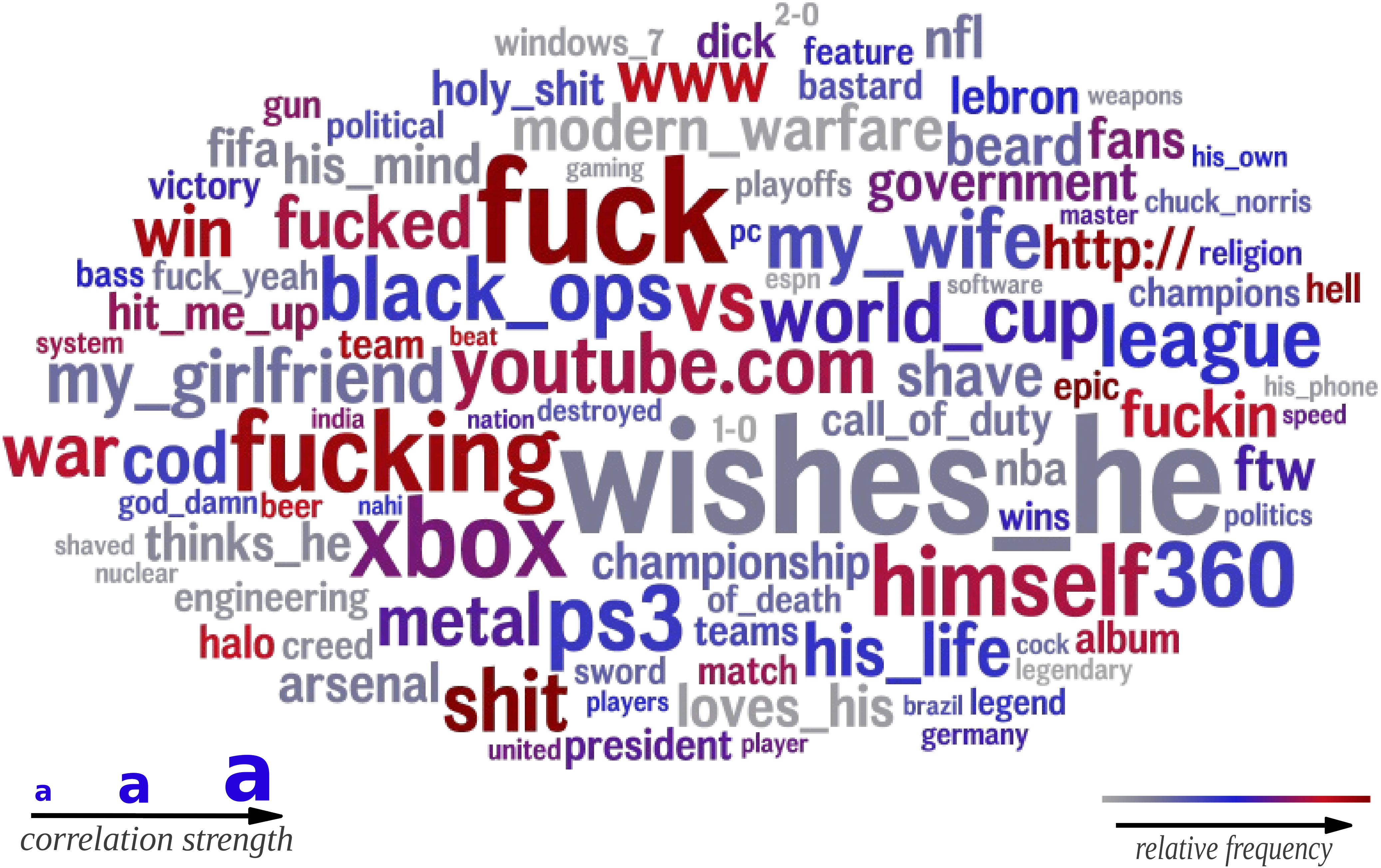
Results: Gender



Bonferonni-adjusted p < 0.001

wwbp.org

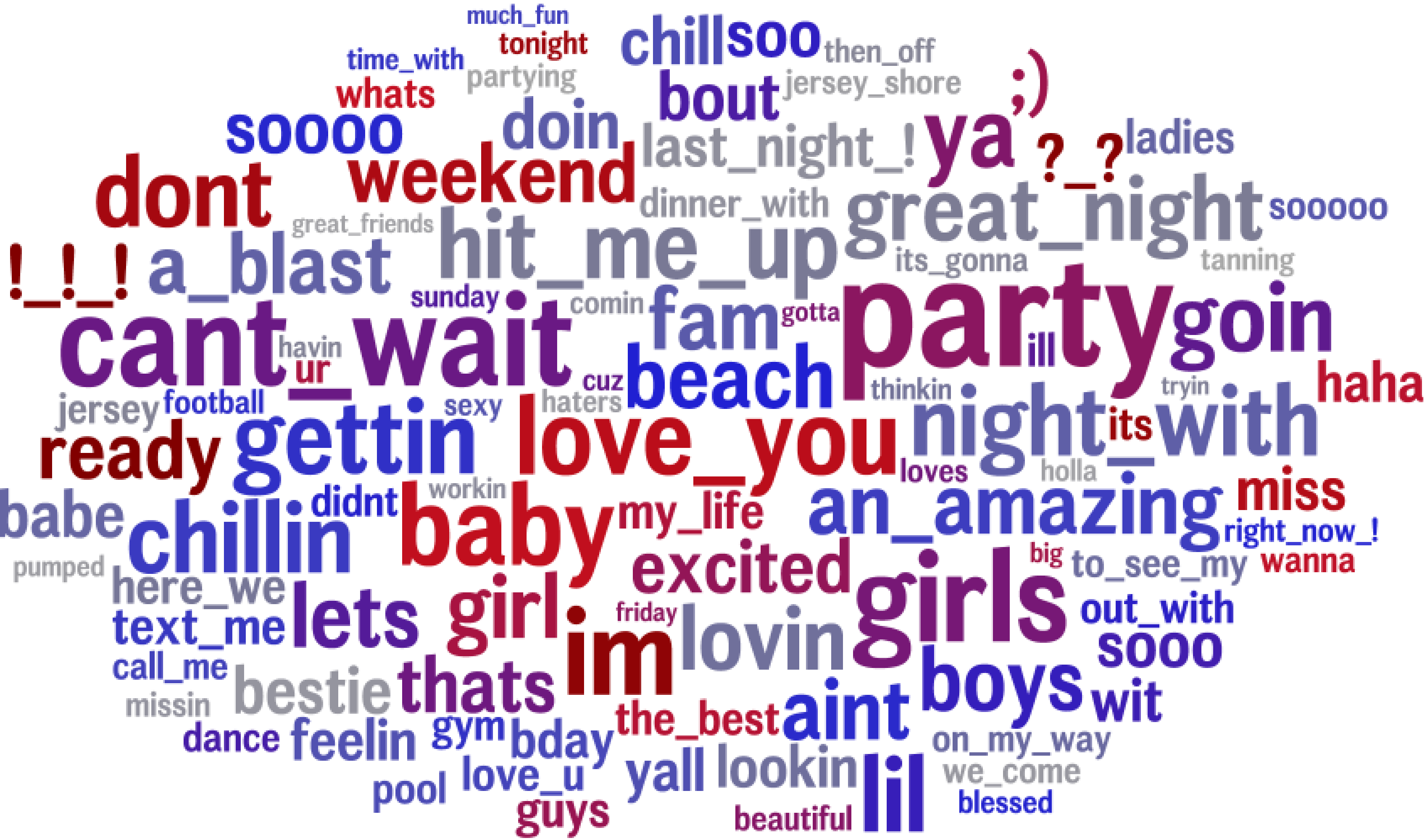
Results: Gender



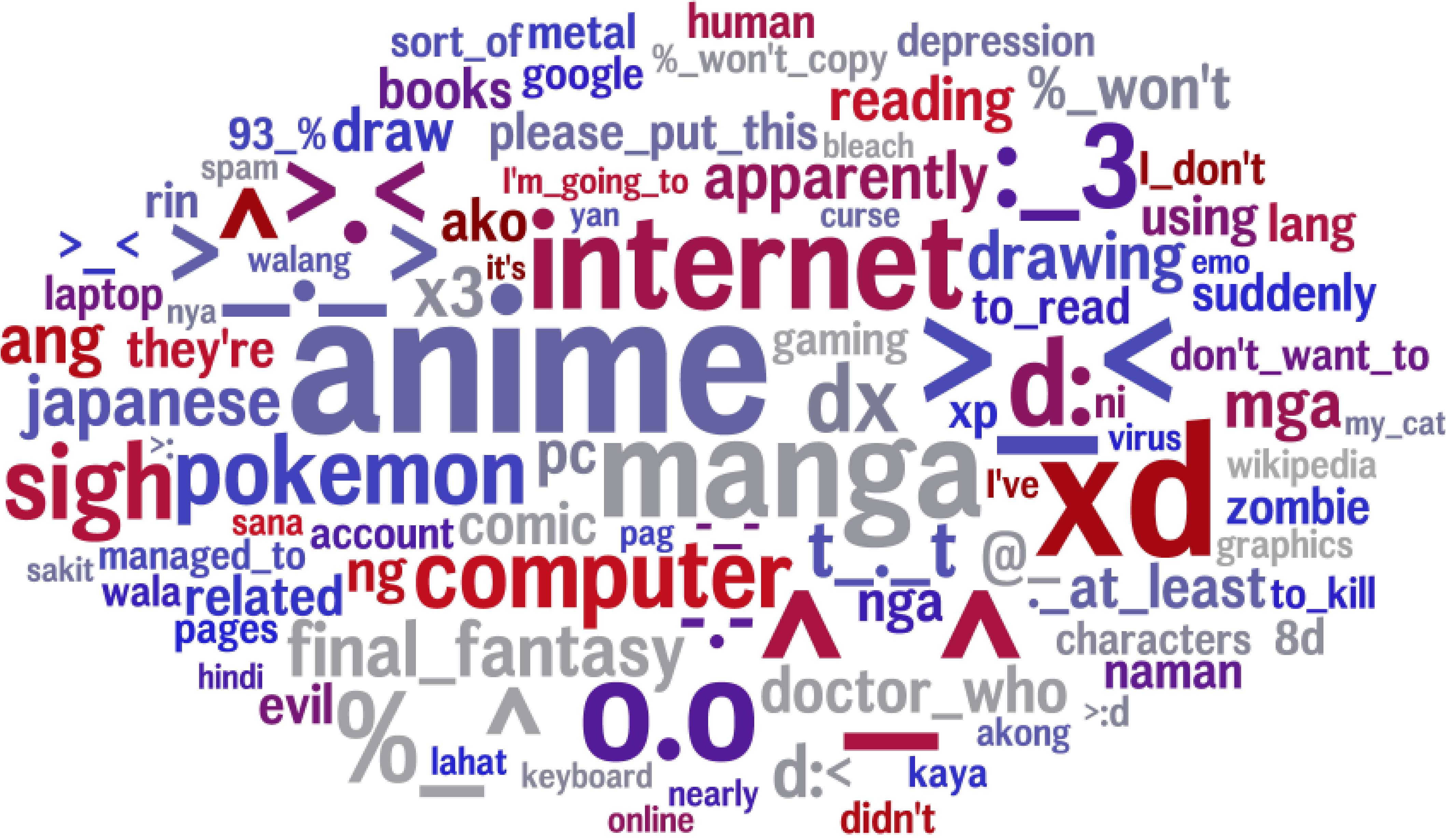
correlation strength

relative frequency

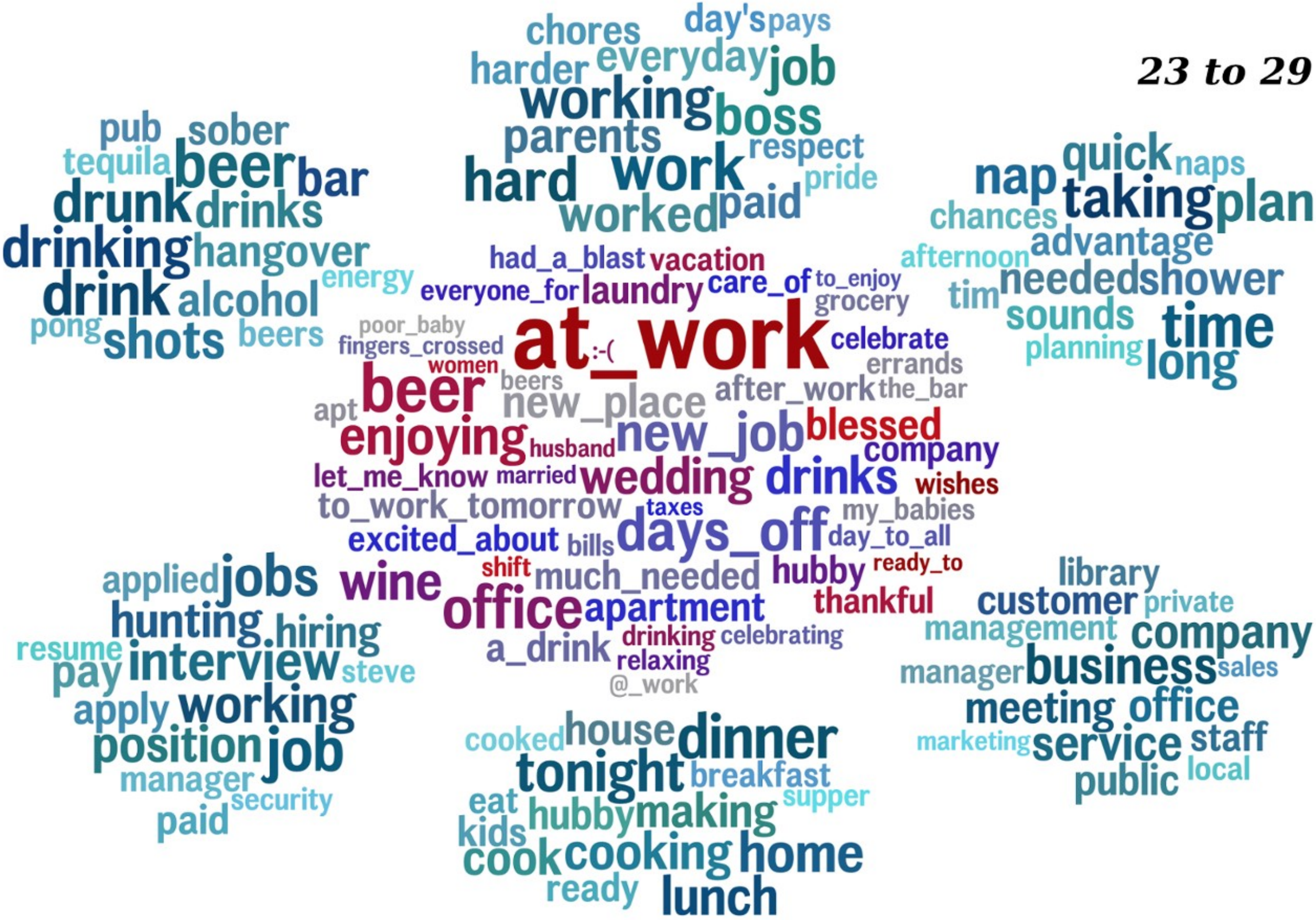
Results: Personality - *Extraversion*



Results: Personality - *Introversion*



Results: Age


23 to 29
chores
day's
pays
harder
everyday
job
working
boss
parents
respect
hard
work
pride
worked
paid
pub
sober
tequila
beer
bar
drunk
drinks
drinking
hangover
energy
drink
alcohol
pong
shots
beers
quick
naps
nap
taking
plan
chances
advantage
afternoon
tim
needed
shower
sounds
time
planning
long
had_a_blast
vacation
everyone_for
laundry
care_of
to_enjoy
grocery
poor_baby
fingers_crossed
at_work
:-(
celebrate
women
errands
beer
beers
new_place
after_work
the_bar
apt
enjoying
husband
new_job
blessed
company
let_me_know
married
wedding
drinks
wishes
to_work_tomorrow
taxes
my_babies
excited_about
bills
days_off
day_to_all
shift
much_needed
hubby
ready_to
wine
office
apartment
thankful
drinking
celebrating
a_drink
relaxing
@_work
applied
jobs
hunting
hiring
resume
pay
interview
steve
apply
working
position
job
manager
security
paid
library
customer
private
management
company
manager
business
sales
meeting
office
marketing
service
staff
public
local
cooked
house
dinner
tonight
breakfast
eat
supper
kids
hubby
making
cook
cooking
home
ready
lunch

Conclusions

- This was a case study in gaining personality insights
  - expected results
  - insightful results
- Moving Forward
  - more feature types (semantic relations, named entities)
  - more analyses (capture interactions)

Conclusions

- This was a case study in gaining personality insights
  - expected results
  - insightful results
- Moving Forward
  - more feature types (semantic relations, named entities)
  - more analyses (capture interactions)

- Other Works
  - predicting gender *(91.9%)*, age *(r = 0.84)*, personality *(r = .31 to .42)*
  - characterizing geographic variance in well-being and health

Personality

- "The Big Five" / Five Factor Model
  (McCrae and John 1992)

  - *extraversion:* active, assertive, energetic, enthusiastic, outgoing
  - *agreeableness:* appreciative, forgiving, generous, kind
  - *conscientiousness:* efficient, organized, planful, reliable
  - *neuroticism:* anxious, self-pittying, tense, touchy, unstable
  - *openness:* artistic, curious, imaginative, insightful, original