gesis Leibniz Institute for the Social Sciences

Potentials and Pitfalls of Social Media Data

Indira Sen & Katrin Weller

GESIS workshop - December 2022

Leibniz Association

Hands-on Error Analysis

- Python notebook
- Stance vs sentiment
- How accurate is VADER?
- Data collection issues based on different subreddits

How do we measure approval?

O'Connor, B., Balasubramanyan, R., Routledge, B.R. and Smith, N.A., 2010, May. **From tweets to polls: Linking text sentiment to public opinion time series.** In *Fourth international AAAI conference on weblogs and social media*.

Conrad, F.G., Gagnon-Bartsch, J.A., Ferg, R.A., Schober, M.F., Pasek, J. and Hou, E., 2019. **Social media as an alternative to surveys of opinions about the economy.** Social Science Computer Review, p.0894439319875692.

Are we Measuring approval towards Trump?

- Usually approval is defined as ‘sentiment’

Tweet	Untargeted Sentiment	Approval
Trump is the only candidate I fully support	positive	approval

Are we Measuring approval towards Trump?

- This sentiment is untargeted, therefore, if multiple users are mentioned, untargeted sentiment =/= approval

Tweet	Untargeted Sentiment	Targeted Sentiment	Approval
Trump is the only candidate I fully support	positive	positive	approval
What makes me angry is the media unnecessarily attacking Trump	negative	positive	approval

Are we Measuring approval towards Trump?

- Finally, targeted sentiment doesn't work when users are ***not directly*** mentioned

Tweet	Untargeted Sentiment	Targeted Sentiment	Stance	Approval
Trump is the only candidate I fully support	positive	positive	favor	approval
What makes me angry is the media unnecessarily attacking Trump	negative	positive	favor	approval
Jeb Bush is the best choice in the republican lineup	positive	none	against	disapproval

What is 'Stance' Detection?

- "the task of automatically determining from text whether the author of the text is in favor of, against, or neutral towards a proposition or target." [Mohammad et al. 2017]

- **Indirect stance:** "the target is referred to in indirect ways such as through pronouns, epithets, honorifics, and relationships."

Stance as a Theoretical Proxy for Approval

Tweet	Untargeted Sentiment	Targeted Sentiment	Stance	Approval
Trump is the only candidate I fully support	positive	positive	favor	approval
What makes me angry is the media unnecessarily attacking Trump	negative	positive	favor	approval
Jeb Bush is the best choice in the republican lineup	positive	none	against	disapproval

Agenda

- Session 1: Introduction to Research with Social Media Data (SMD)
- Session 2: SM Data Collection
- Session 3: SMD Preprocessing and Analysis
- Session 4: Potential Pitfalls of SMD
- Session 5: Identifying Pitfalls with help from surveys
- Session 6: Identifying Pitfalls in SMD
- **Session 7: Mitigating Pitfalls**
- Session 8: Documenting Pitfalls
- Session 9: Recap and Conclusions

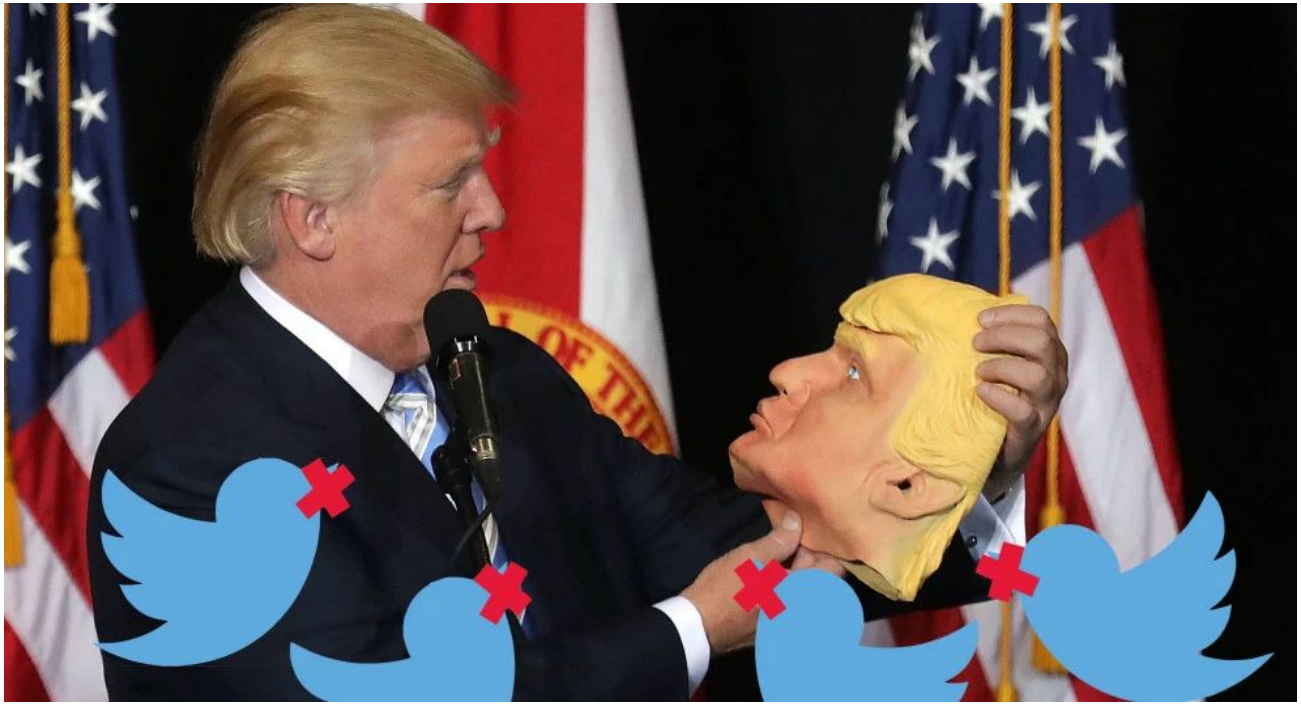
gesis Leibniz Institute for the Social Sciences

5: Mitigating Pitfalls

Photo: Katrin Weller

Case Study

Understanding **Presidential Approval** from tweets

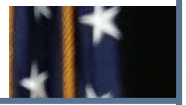


- Understan

Social Science Computer Review
1-18
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0894439318822007
journals.sagepub.com/home/ssc
SAGE

Who's Tweeting About the President? What Big Survey Data Can Tell Us About Digital Traces?

Josh Pasek[1], Colleen A. McClain[1], Frank Newport[2], and Stephanie Marken[2]



Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data.

Pablo Barberá
Center for Data Science
New York University
www.pablobarbera.com

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series

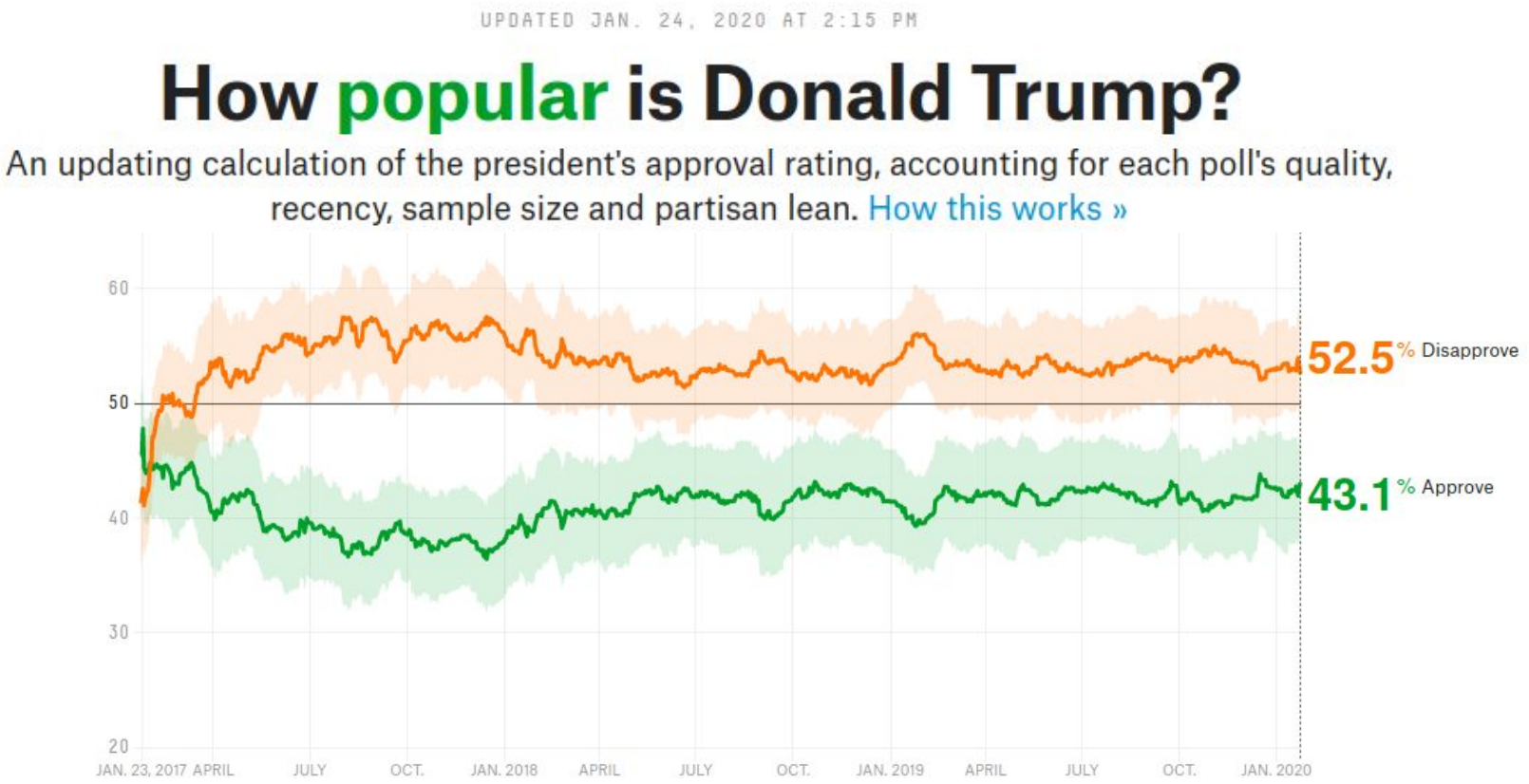
Brendan O'Connor[†] **Ramnath Balasubramanyan**[†] **Bryan R. Routledge**[§] **Noah A. Smith**[†]
brenocon@cs.cmu.edu rbalasub@cs.cmu.edu routledge@cmu.edu nasmith@cs.cmu.edu

[†]School of Computer Science
Carnegie Mellon University

[§]Tepper School of Business
Carnegie Mellon University

Presidential Approval in Surveys

Do you approve or disapprove of the way Donald Trump is handling his job as president?



Data Analysis

Presidential Approval poll aggregate image from https://projects.fivethirtyeight.com/trump-approval-ratings

Construct Definition

Tweets about Trump with negative or positive sentiment

TRACES

[redacted]

Replying to [redacted]

I was one that voted for Trump...BIGGEST mistake. I've always been blue but I thought I was voting for the [redacted] Thank you

Data Collection

Data Preprocessing

Data Analysis

gesis Leibniz Institute for the Social Sciences

Validity

Construct Definition

Tweets about Trump with negative or positive sentiment

TRACES

Replying to [redacted]

I was one that voted for Trump...BIGGEST mistake. I've always been blue but I thought I was voting for the [redacted] Thank you

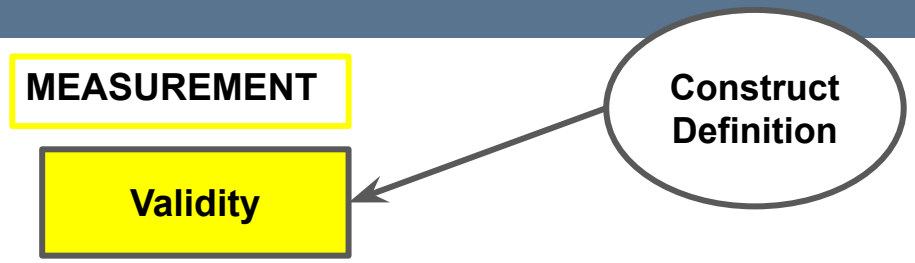
Tweets about Trump maybe related to his personal life

Data Collection

Data Preprocessing

Do you approve or disapprove of the way Donald Trump is handling his job as president?

Data Analysis



Potential remedies for validity issues:

Build on a strong theoretical foundation for the expected traces

- Conduct focus groups with Twitter users and learn how they express approval
- Online panels: combine platform data with offline records (potentially very sensitive)

➔ Joseph, Kenneth, et al. "Polarized, Together: Comparing Partisan Support for Trump's Tweets Using Survey and Platform-Based Measures." Proceedings of the *International AAAI Conference on Web and Social Media*. Vol. 13. No. 01. (2019).

➔ Barberá, Pablo. "Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data." *Work Paper NYU* (2016).

➔ Pew Research on how people tweet: https://pewrsr.ch/2SgbujQ

Construct Definition

Platform Selection

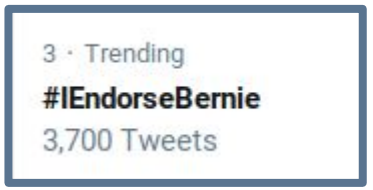
Data Collection

Data Preprocessing

Data Analysis

3 · Trending
#IEndorseBernie
3,700 Tweets

Platform Affordances Error



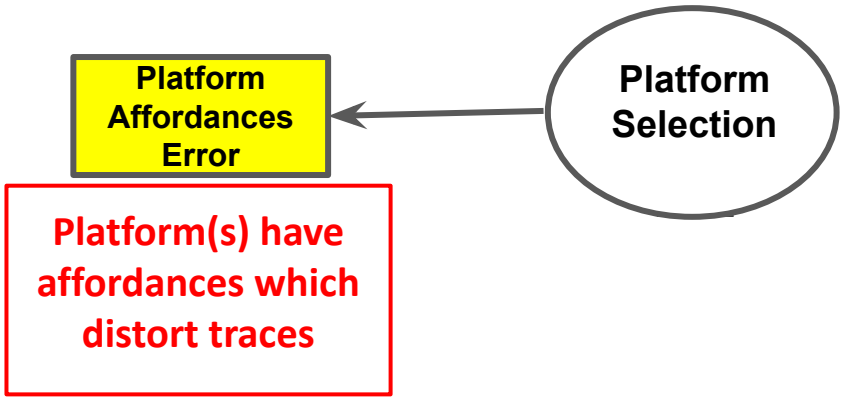
Platform affordances:

- Terse tweets or a thread consisting of multiple tweets to express their opinion
- Less likely to expose an opinion that they assume to be unpopular on the platform
- Twitter's "trends" feature, showing popular (including political) hashtags might nudge users

MEASUREMENT

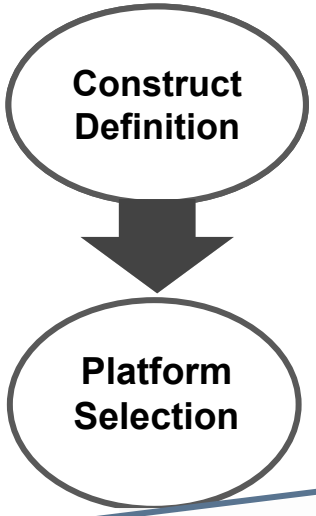
Potential remedies for Platform Affordances Error:

- **Quantify the effect of a feature vs. its non-existence**
- **Technical solutions around trace sharing restrictions**

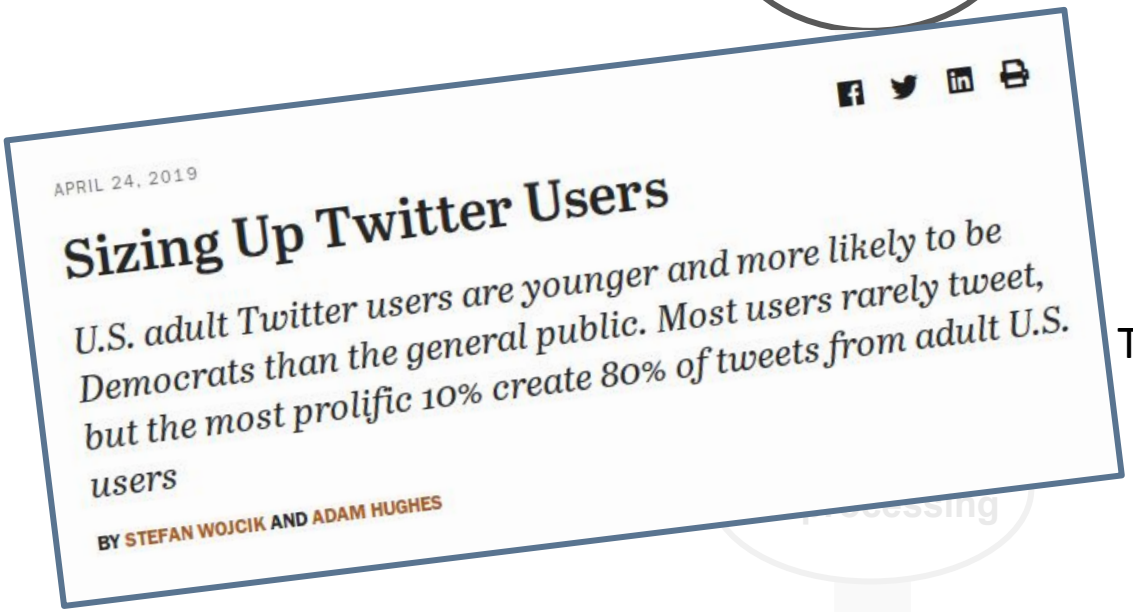


➔ Malik, Momin M., and Jürgen Pfeffer. "Identifying platform effects in social media data." *Tenth International AAAI Conference on Web and Social Media* (2016).

➔ Gligorić, Kristina, Ashton Anderson, and Robert West. "How constraints affect content: The case of Twitter's switch from 140 to 280 characters." *Twelfth International AAAI Conference on Web and Social Media* (2018).



Platform Coverage Error



APRIL 24, 2019

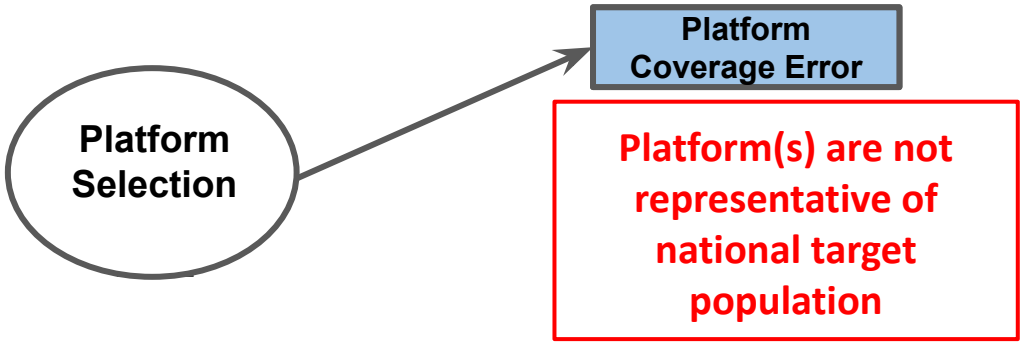
Sizing Up Twitter Users

U.S. adult Twitter users are younger and more likely to be Democrats than the general public. Most users rarely tweet, but the most prolific 10% create 80% of tweets from adult U.S. users

BY STEFAN WOJCIK AND ADAM HUGHES

Platform population = people with Twitter accounts != general American population

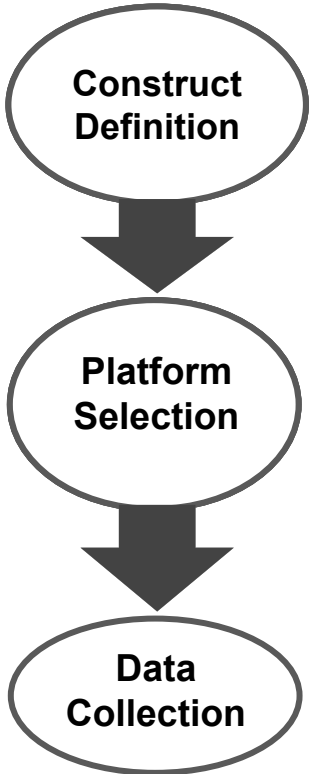
REPRESENTATION



Potential remedies for Platform Coverage Error:

- **Reweighting (adjustment) of platform population**
- **Using multiple platforms**

➔ Zagheni, Emilio, and Ingmar Weber. "Demographic Research with Non-Representative Internet Data." (2014).



Trace Selection Error

Tweets with keywords related to Trump:

- 'Trump' phrases unrelated to Donald Trump, which e.g. refer to Melania Trump or Donald Trump Jr.
- Likewise, relevant queries might be excluded that simply refer to "the president", "Donald" or acronyms.

MEASUREMENT

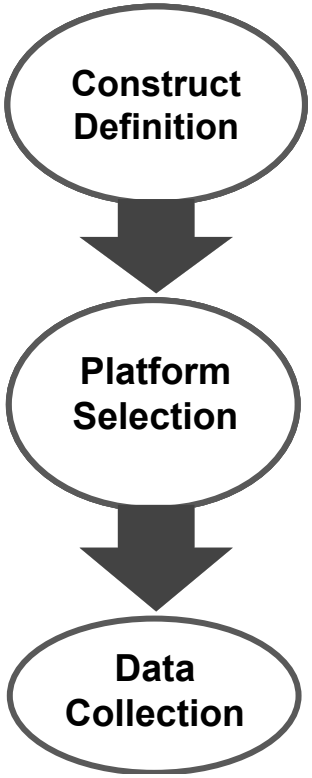
Potential remedies for Trace Selection Error:

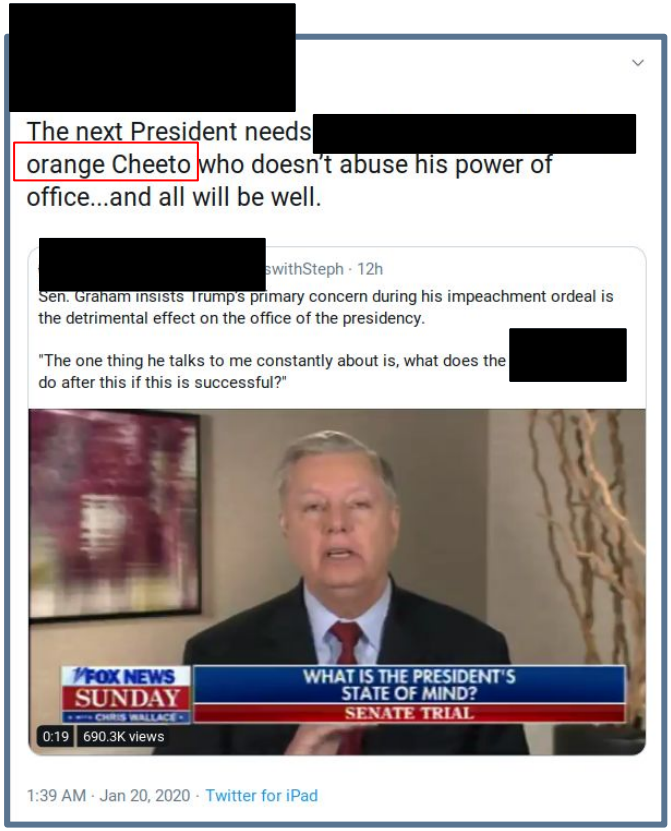
- **Error analysis of queries, specifically understanding precision and recall**
- **Improving data collection strategies with query expansion**

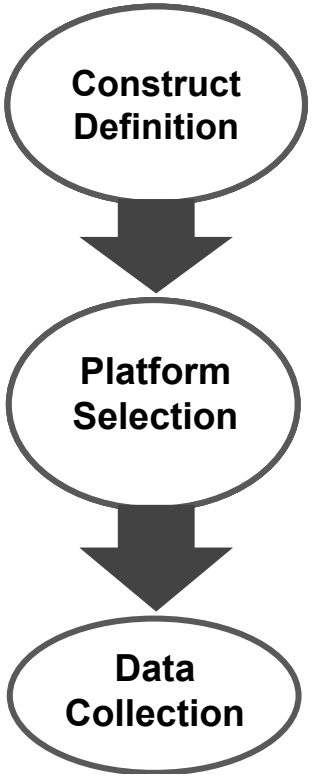
➔ Ruiz, Eduardo J., Vagelis Hristidis, and Panagiotis G. Ipeirotis. "Efficient filtering on hidden document streams." *Eighth International AAAI Conference on Weblogs and Social Media* (2014).

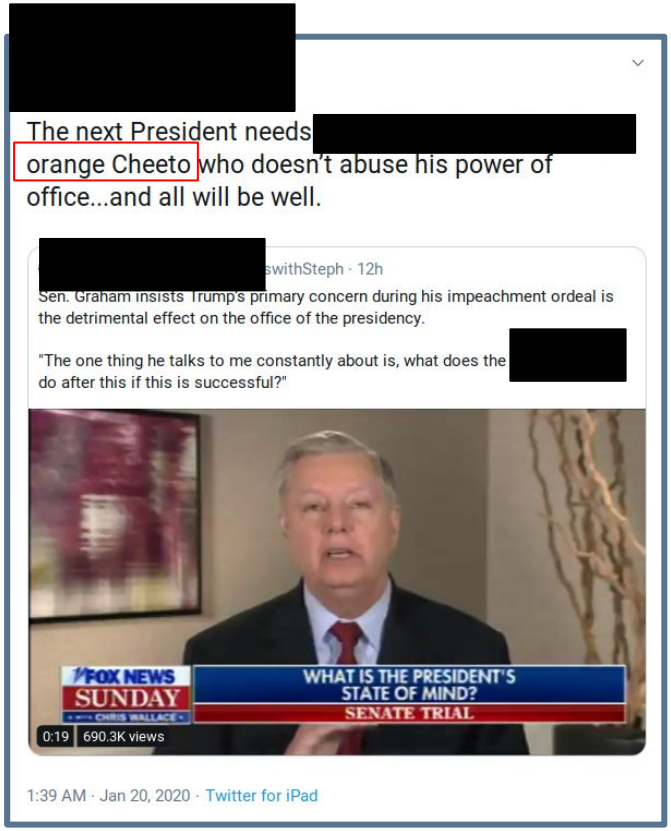
➔ Linder, Fridolin. "Improved data collection from online sources using query expansion and active learning." Available at SSRN 3026393 (2017).

gesis Leibniz Institute for the Social Sciences









User Selection Error

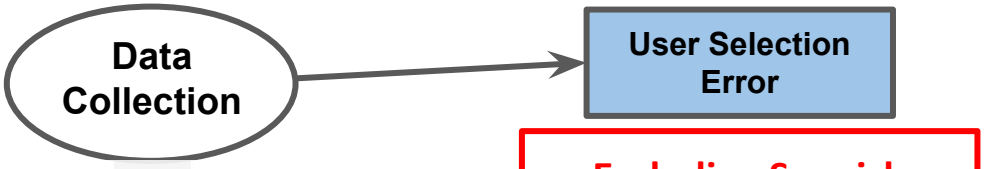
Certain groups of users may be underrepresented if they refer to the President with a certain nickname or other name: 'orange cheeto' for detractors

REPRESENTATION

Construct Definition

Potential remedies for User Selection Error:

- **Assess the characteristics of users included in the data with changing queries**
- **Reweighting (adjustment)**



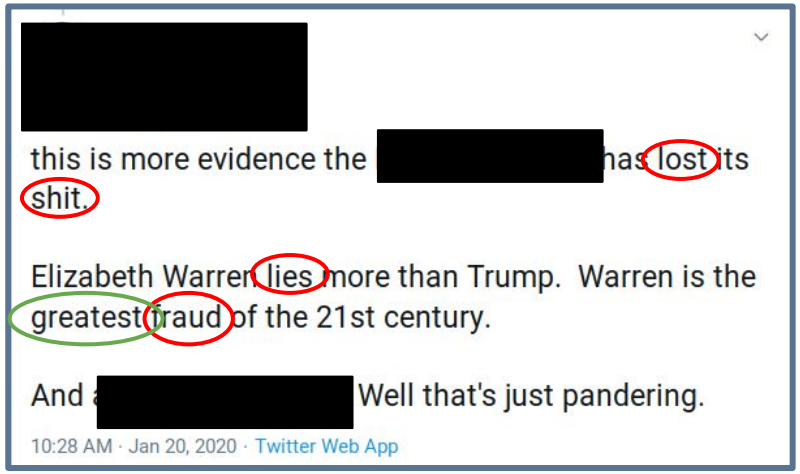
Excluding Spanish tweets excludes Hispanic population of the US. Alternately, includes English speakers from other countries

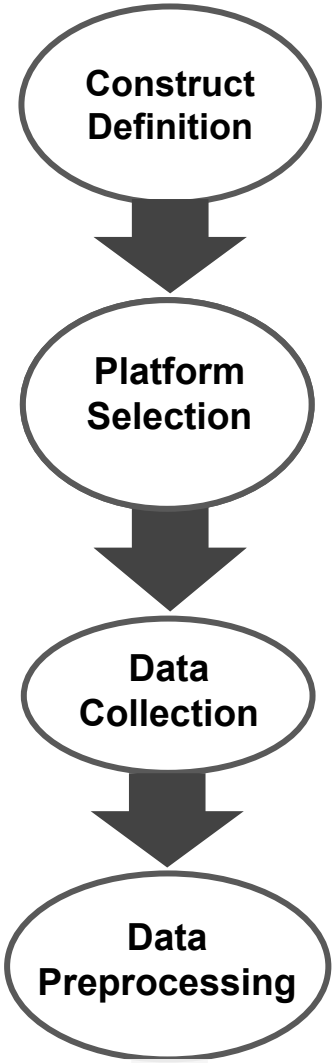
Data Preprocessing

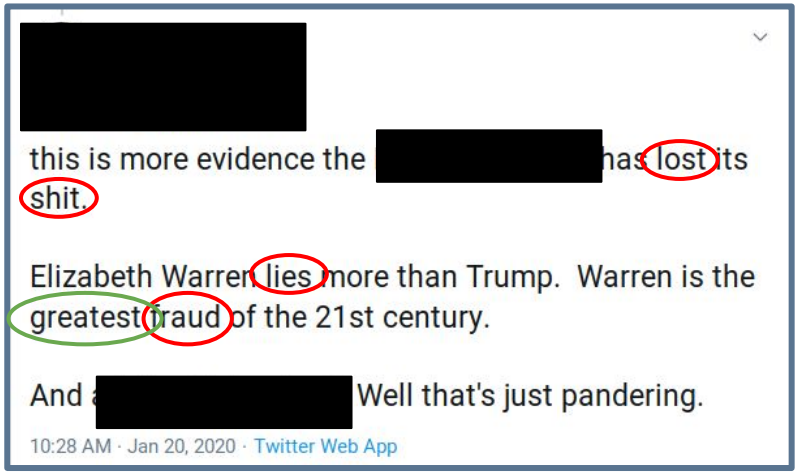
Data Analysis

➔ Barberá, Pablo. "Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data." *Work Paper NYU* (2016).

Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis


this is more evidence the has lost its shit.
Elizabeth Warren lies more than Trump. Warren is the greatest fraud of the 21st century.
And Well that's just pandering.
10:28 AM · Jan 20, 2020 · Twitter Web App



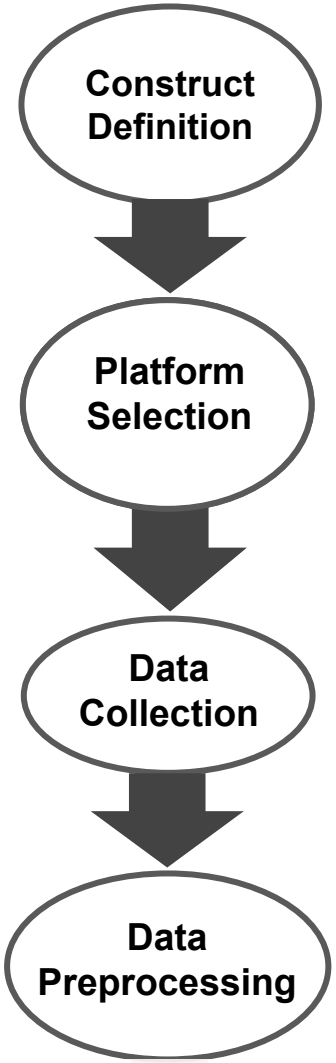


Trace Augmentation Error

Augmentation: sentiment analysis for annotating approval --- social media vocabulary mismatch, target-independent lexicon

Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

[redacted]
@DiamondandSilk @Its_A_GOP_COUP
@realDonaldTrump @LindseyGrahamSC @VP
@senatemajldr @IvankaTrump @SenateGOP
@HouseGOP @GOP @JudiciaryGOP @DevinNunes
@LouDobbs @marcorubio @SenRickScott
@MittRomney @SenatorCollins @lisamurkowski
@BenSasse @CoryGardner @SenMikeLee
@SenatorBurr
DO I LOOK LIKE I CARE ABOUT YOUR PROBLEMS ?
BEAUTIFUL CLEAN COAL
TRUMP NOBEL PEACE PRIZE
BEAUTIFUL CLEAN COAL



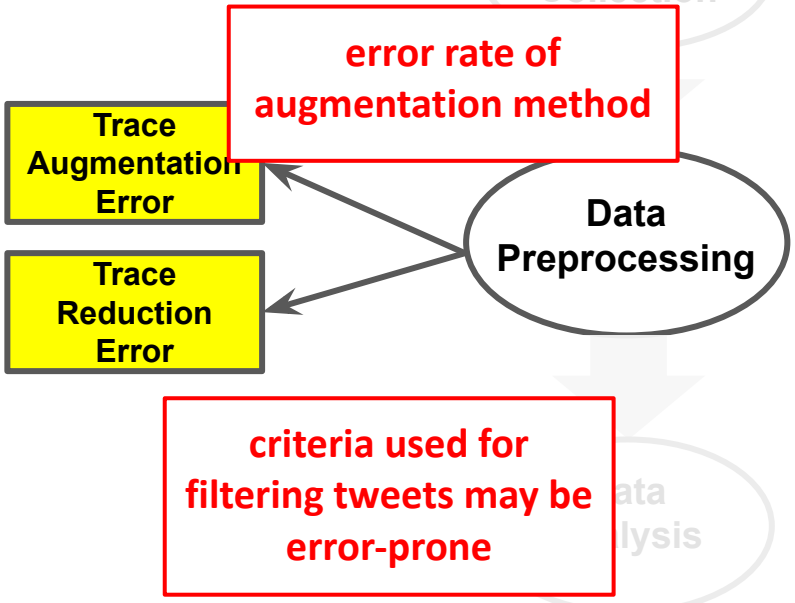
Reduction: remove spam tweets --- may inadvertently remove useful traces or discard tweets that do not contain any textual content, thereby ignoring relevant statements made through hyperlinks or embedded pictures/videos

MEASUREMENT

- Denny, Matthew J., and Arthur Spirling. "Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it." *Political Analysis* 26.2 (2018): 168-189.
- González-Bailón, Sandra, and Georgios Paltoglou. "Traces of public opinion in online communication: A comparison of methods and data sources." *The ANNALS of the American Academy of Political and Social Science* 659.1 (2015): 95-107.

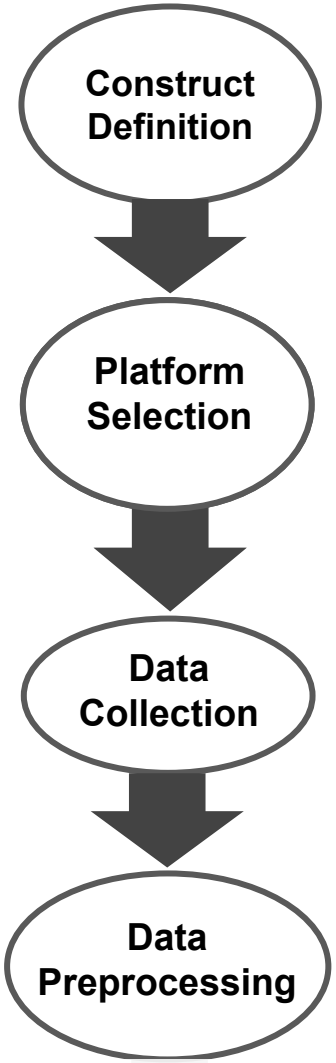
Potential remedies for Trace Preprocessing Error:

- **Assess which augmentation and reduction methods are suitable**



Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

OLIVER MCGEE
AEROSPACE, MECHANICAL AND CIVIL ENGINEER
BATTLE
Follow
Oliver McGee PhD MBA
@OliverMcGee
FMR Sr Advisor @WhiteHouse, FMR US Dep Asst Secretary #MAGA|Guest @SkyNews @BBCWorld @LBC |@FoxNews|@MSNBC|@CNN|@CGTNOfficial |BookMe ogmcgee@yahoo.com
Washington, DC, USA
OliverMcGee.org
Joined October 2010
144.5K Following
201.9K Followers

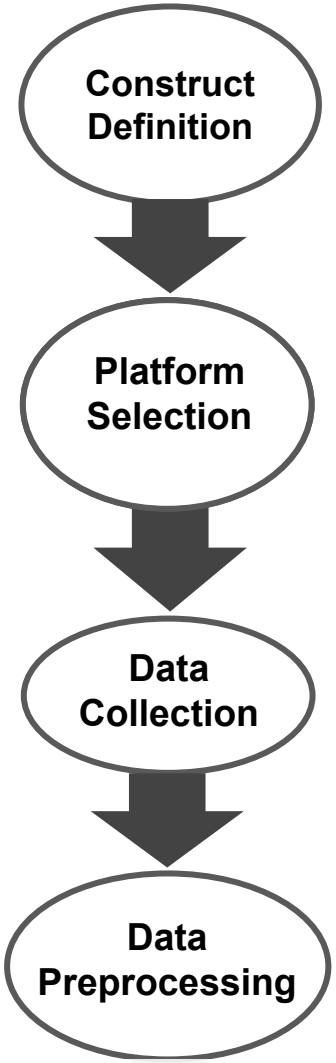


User Augmentation Error

Augmentation: Twitter profiles' age, gender, ethnicity, or location may be inferred to understand how presidential approval differs across demographics. But current methods have limitations

Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

[redacted]
Joined January 2020
208 Following
6 Followers



User Reduction Error

Reduction: Remove users who are not representative of human users, e.g. bots and organizations

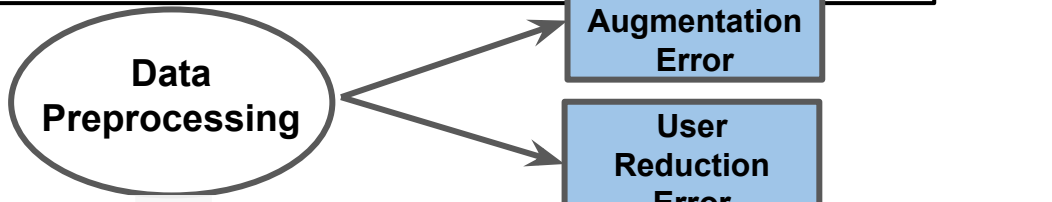
REPRESENTATION

Construct Definition
Platform Selection
Data Analysis

➔ Karimi, Fariba, et al. "Inferring gender from names on the web: A comparative evaluation of gender detection methods." Proceedings of the *25th International Conference Companion on World Wide Web* (2016).

Potential remedies for User Preprocessing Error:

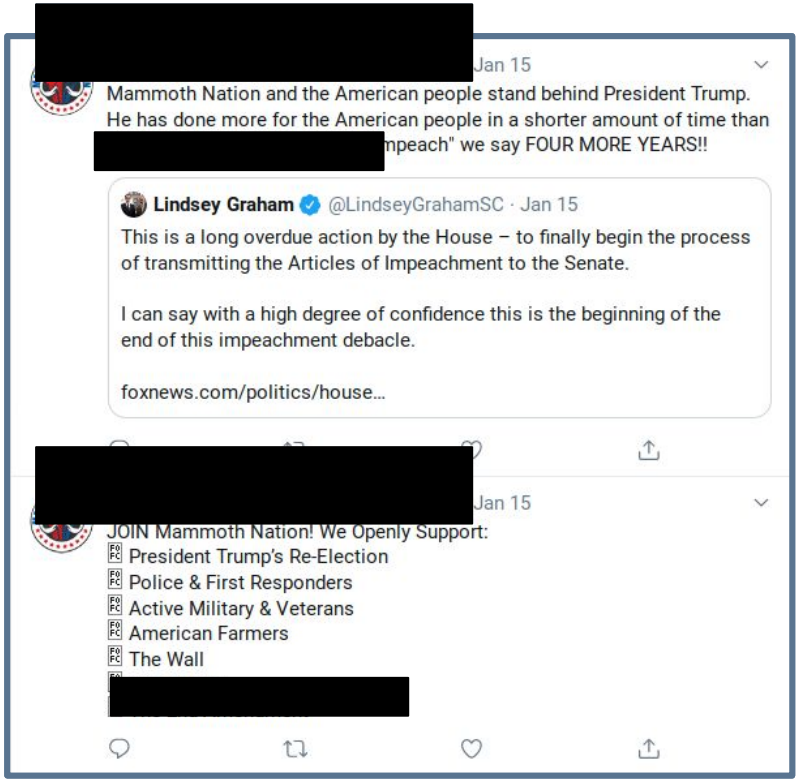
- **Assess the errors of user augmentation and reduction methods, especially for minorities**
  - **Ensemble methods**
- **(Only using self-reports)**



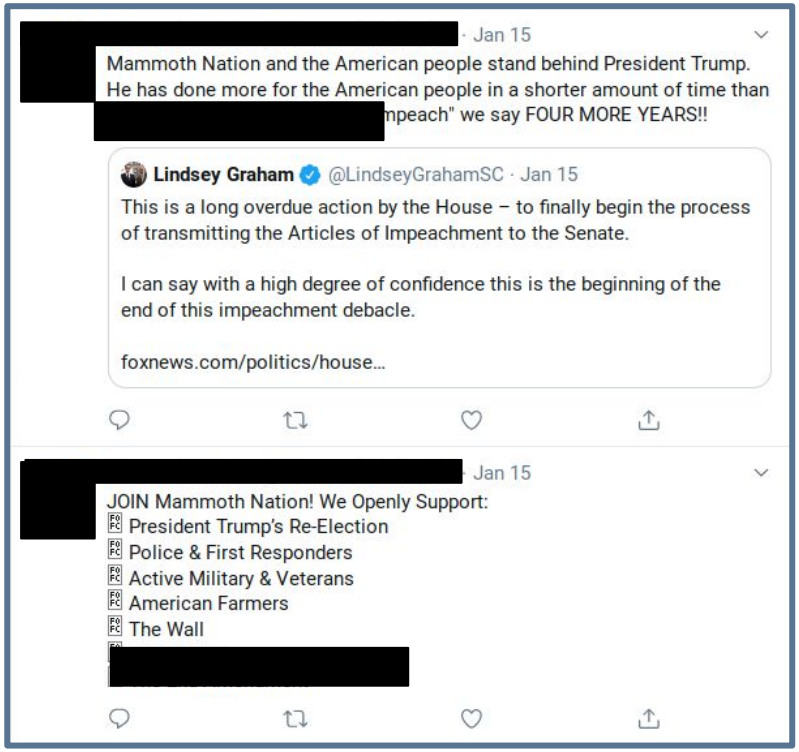
error rate of augmentation method

criteria used for filtering users may be error-prone

Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis


[redacted] Jan 15
Mammoth Nation and the American people stand behind President Trump. He has done more for the American people in a shorter amount of time than [redacted] mpeach" we say FOUR MORE YEARS!!
Lindsey Graham @LindseyGrahamSC · Jan 15
This is a long overdue action by the House – to finally begin the process of transmitting the Articles of Impeachment to the Senate.
I can say with a high degree of confidence this is the beginning of the end of this impeachment debacle.
foxnews.com/politics/house...
[redacted] Jan 15
JOIN Mammoth Nation! We Openly Support:
President Trump's Re-Election
Police & First Responders
Active Military & Veterans
American Farmers
The Wall
[redacted]

Trace Measurement Error



- Count the normalized positive words in a day
- The ratio of positive and negative words per tweet
- Add all the ratio of all tweets in a day

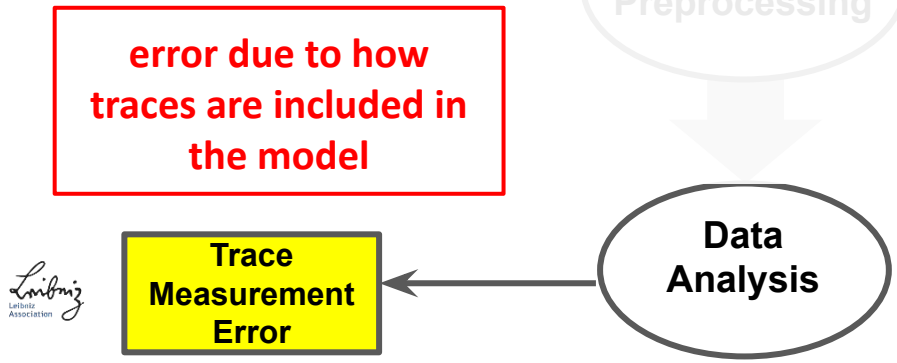
Different aggregation methods lead to different measures of approval

MEASUREMENT

➔ Ahmed, Saifuddin, Kokil Jaidka, and Marko M. Skoric. "Tweets and votes: A four-country comparison of volumetric and sentiment analysis approaches." *Tenth International AAAI Conference on Web and Social Media* (2016).

Potential remedies for Trace Measurement Error:

- **Assess what different aggregation strategies *mean***
- **Consult domain experts**



Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

Construct Definition

Platform Selection

Data Collection

Data Preprocessing

Data Analysis

Who's Tweeting About the President? What Big Survey Data Can Tell Us About Digital Traces?

Social Science Computer Review
1-18
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0894439318822007
journals.sagepub.com/home/ssc
SAGE

Josh Pasek[1], Colleen A. McClain[1], Frank Newport[2], and Stephanie Marken[2]

Adjustment Error

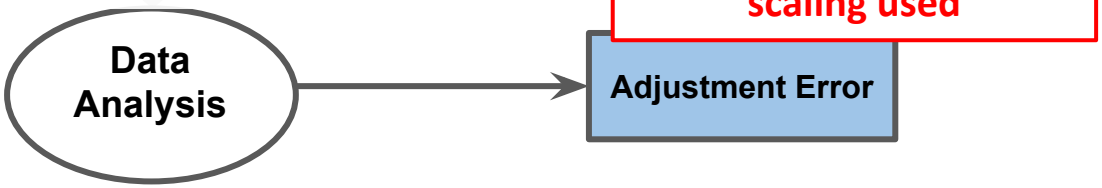
- Reweighting survey estimates by Twitter demographics to find an alignment may not work since **political Twitterers != random Twitterers**

REPRESENTATION

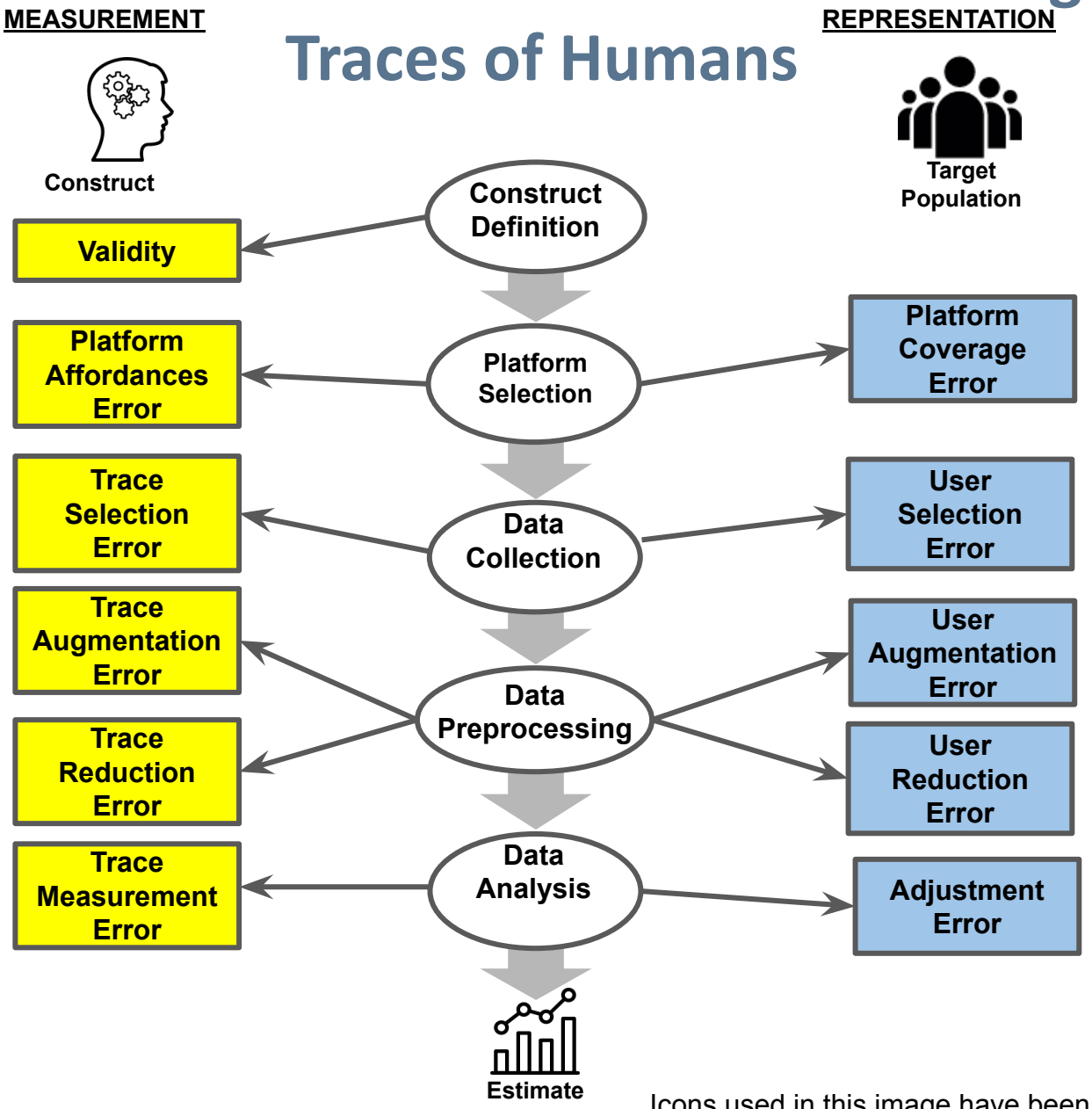
➔ Li, Lingling, et al. "On weighting approaches for missing data." Statistical methods in medical research 22.1 (2013): 14-30.

Potential remedies for Adjustment Error:

- **Carefully select variables for reweighting**
- **Assess different reweighting methods and see which is best for the problem**



TED-On: A Total Error Framework for Digital Traces of Humans



Icons used in this image have been designed by Becris, EliasBikbulatov and Pixel perfect from www.flaticon.com

To Sum Up...

- TED-On: A Total error framework inspired by survey methodology to decompose errors in DTD-based study
- Acts as guidelines for researchers designing their studies
- A way to systematically illuminate errors - known and unknown, similar to Datasheets [Gebru et al. 2018] and Model Cards [Mitchell et al. 2018]