

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Leveraging multiple digital footprint datasets to predict racial, sex-based, and sexual-orientation bias across US states

Raphael Derecki¹, Brian O'Shea¹, and James Goulding¹

¹University of Nottingham

Introduction & Background

Food retailers' transaction data are increasingly used for research. Unlike many other digital footprints, the representativeness of automatically accumulating food purchase data as such is less biased as food is consumed by all individuals. However, the process of obtaining individual/household level data requires consents from the consumers and, thus, may create selection bias.

Objectives & Approach

We explored biases using geospatial analyses throughout multiple datasets based on US participants: Project Implicit, American National Election Studies (ANES), Google Trends, and Twitter/X. We included state-based variables to compare between states and assess the most important environmental-level predictors of biases. We built generalised linear and linear mixed-effect models and general linear models. Within Project implicit ($n > 3,000,000$) and ANES datasets ($n > 30,000$), we assessed racial and sexual-orientation biases via explicit and implicit measures. For Google Trends and Twitter/X datasets, we assessed racial and sex-based biases via search and tweet-per-state scores. To analyse the biases, we included environmental-level variables, e.g., infectious disease rates (developed by Thornhill and Fincher in 2014), and individual-level variables, e.g., political orientation.

Relevance to Digital Footprints

These preliminary findings analyse everyday people's online behaviour including volunteered surveys, searches and posts. We attempt to address the pressing societal issue of bias by leveraging modern datasets. Our primary goal is to aid policy makers by recommending cost-effective solutions that can improve several factors of the population's quality of life.

Results

We find that the most consistently significant predictor of racial bias is infectious disease rates. When leveraging Google Trends data including anti-women terminology, infectious disease rates and population density are consistent predictors of bias. Finally, we find preliminary results suggesting that increased levels of infectious diseases increases homophobic bias.

Conclusions & Implications

Overall, we find that as infectious disease rates increase in a state, the level of racial and sexist bias significantly increases. Consistent with parasite-stress theory, we argue that focusing on reducing infectious disease rates in an area can have a plethora of benefits including improving physical and mental health and reducing biases that damage society.

