

Twitter Corpus of the #BlackLivesMatter Movement And Counter Protests: 2013 to 2021

Salvatore Giorgi,^{1,2} Sharath Chandra Guntuku,¹ McKenzie Himelein-Wachowiak,² Amy Kwarteng,²

Sy Hwang,¹ Muhammad Rahman,² Brenda Curtis²

¹University of Pennsylvania, ²National Institute on Drug Abuse

Introduction

- Black Lives Matter (BLM) is a decentralized social movement protesting violence against Black individuals and communities, with a focus on police brutality
- Movement gained significant attention following the murders of Ahmaud Arbery, Breonna Taylor, and George Floyd in 2020

- Police violence toward Black individuals perpetuates negative mental and physical health

Data Collection

- Data pulled continuously from June 2016 to Dec. 2021 using official Twitter API

- Pre-June 2016 data was pulled using the GetOldTweets Python package

- All tweets contain some variation of one of three keywords: BlackLivesMatter, AllLivesMatter, BlueLivesMatter

Ethics Statement

- Only public tweets IDs are made available, no raw or inferred data is shared

- While all tweets are publicly available, users did not consent to share data

- Hydrated data may contain identifying information

- Authors do not intend to draw equivalences between counter protests and BLM

	Tweets	Users	Retweets	Replies	Geotagged	User Location	Top Languages
<i>All</i>	63,884,799	13,061,316	47,083,420	3,266,120	86,641	10,820,854	en, fr, es, pt, ja
<i>BlackLivesMatter</i>	56,693,715	12,322,212	42,693,046	2,590,724	77,257	9,552,502	en, fr, es, pt, ja
<i>AllLivesMatter</i>	4,343,704	1,845,937	2,287,247	564,714	9,928	698,252	en, es, nl, ja, fr
<i>BlueLivesMatter</i>	5,075,833	1,224,933	3,494,159	306,711	2,329	938,335	en, fr, es, ja, de

Table: Descriptive counts for the entire data set and each keyword

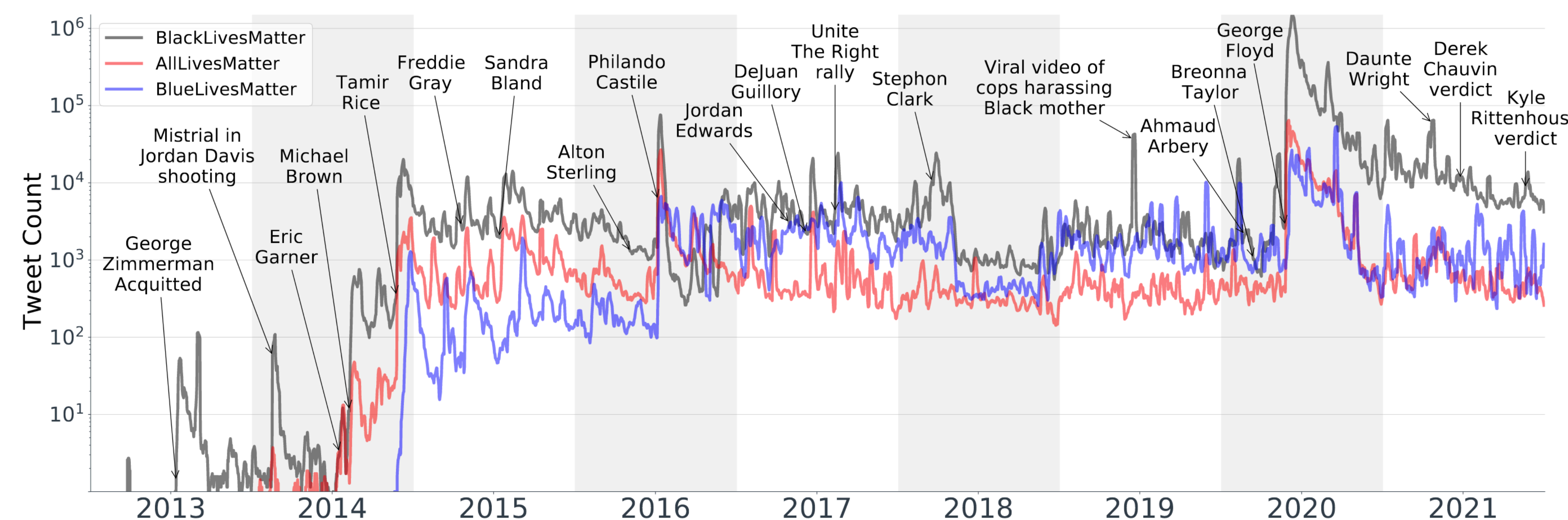


Figure 1: Seven day moving average of monthly tweet count from 2013 to 2021

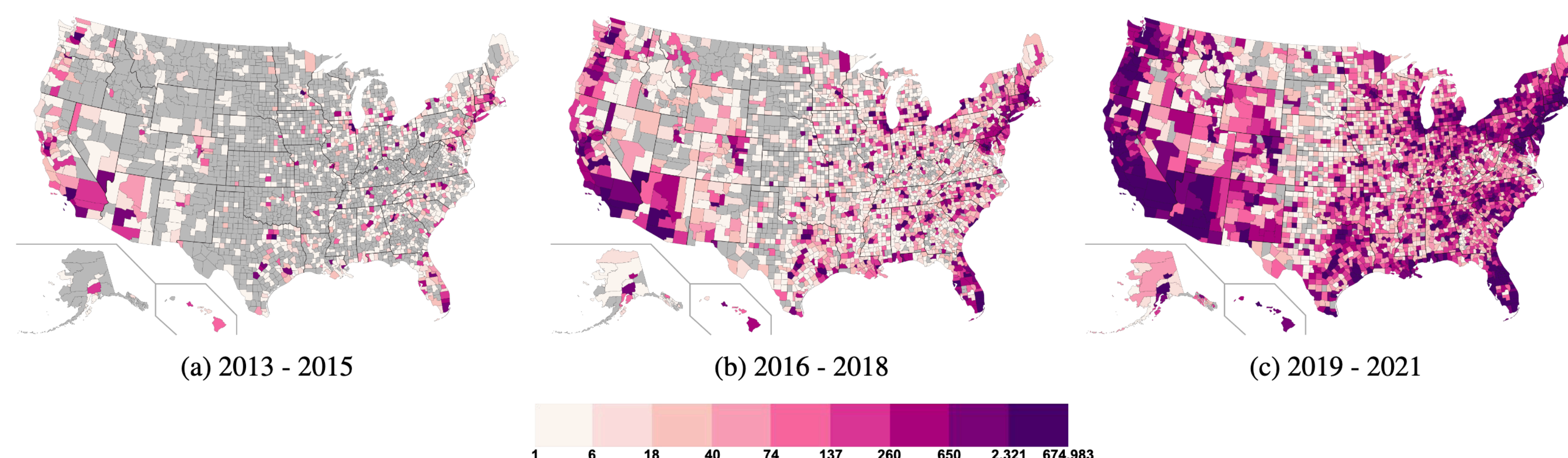


Figure 2: Distribution of BlackLivesMatter tweets across the United States for three 3-year periods: 2013 to 2015, 2016 to 2018, and 2019 to 2021

LDA Topics



Figure 3: Most frequently used topic for each keyword

Data and Software

- Data is available on Zenodo:
zenodo.org/record/5835260
- Software for hydrating tweets:
github.com/sjgiorgi/blm_twitter_corpus

