Understanding Public Opinion and Demonstrative Support of LGBTIQ+ Population in the United States through Social Media Data

Matthew Kusen[1]

Abstract

BACKGROUND

Understanding support for Lesbian, Gay, Bisexual, Transgender, Intersex, and Queer (LGBTIQ+) populations help to provide a foundation for acceptance, legal protections, and equality for LGBTIQ+ persons in society. Collecting information on attitudes and acceptance can be timely and costly exercise. Fortunately, technological advancements and wide availability of digital trace data have provided new, cost-effective opportunities to collect and analyze data at scale.

OBJECTIVE

This research aims to analyze and understand current levels of LGBTIQ+ support on social media and if women and younger generations are more likely to demonstrated support of people in the United States of America utilizing digital trace data from Facebook and Instagram.

DATA & METHODS

This project utilized Meta's Marketing Ads Manager platform to analyze behavior of Facebook and Instagram users in the United States. Data collection was recorded manually through Facebook's website in September 2021 collecting data on LGBTIQ+ support, age, and gender. Data is provided anonymized, aggregated and rounded to the nearest zero. The total number of users of the sample size is 226.4 million users in the United States.

RESULTS

The results found that young and female persons have higher rates and higher odds of demonstrating LGBTIQ+ support on social media compared to their older male counterparts. Utilizing logistic regression, the odds of a female demonstrating LGBTIQ+ supportive behavior on Instagram and Facebook is 1.73 times higher than those of their male counterparts while Gen Z and Millennials are more likely to demonstrate online support of LGBTIQ+ groups ($p < .0000$) compared to older generations. Overall, 19% of U.S. populations (43.14 million) aged 15 to 64 demonstrated support of LGBTIQ+ persons.

Conclusion

While public opinion is changing and some countries especially in the developed world have instituted progressive LGBTIQ+ support policies and laws, much more is still needed to be done. This research shows us that while we see increasing levels of support of LGBTIQ+ persons, demonstrated support through social and online spaces is still relatively small at less than 20% of Americans.

[1] Master in Demography Student, Chulalongkorn University, College of Population Studies (CPS), Thailand, mrkusen@gmail.com

Table of Contents

1. Introduction

Understanding perspectives and support for the Lesbian, Gay, Bisexual, Transgender, Intersex, and Queer (LGBTIQ+) populations is important. Public opinion and societal opinions towards marginalized and vulnerable populations have widespread impact on public policy, laws, and how members from this community face discrimination and stigma. LGBTIQ+ data collection is severely limited and absent from every national census in the world. In order to collect data on these groups and make sure they are counted, and their voices are heard, public opinion needs to change. Understanding public option has been done in the past using population-based surveys, which can be expensive and a time-consuming process. Through this research paper, I explore new ways to measure and understand LGBTIQ+ support utilizing social media data. While the topic of this has been completed before, these quick and large aggregate level methods have not be used in the past to understand trends and perspectives of LGBTIQ+ support which are useful in the fields of demography and population studies.

1.1. Background and Literature Review

1.1.1. LGBTIQ+ Stigma, Data and Discrimination

According to the United Nations Human Rights Office of the High Commissioner, "the Universal Declaration of Human Rights states that all human beings are born free and equal in dignity and rights, without distinction of any kind. Yet in all regions of the world, there are acts of violence and discrimination committed against individuals because of their sexual orientation or gender identity." (UN OHCHR 2021). Sexual orientation, gender identity and expression, and sex characteristics (SOGIESC) is an umbrella term that is sometimes used interchangeably with Lesbian, Gay, Bisexual, Transgender, Intersex, Queer and other identities (LGBTQ+) to help categorize people who have sexual and gender identities outside of what most societies deem as 'normal'. These terms are sometimes used interchangeably throughout the research proposal along with the term sexual and gender minorities. These terms are all trying to capture the same group and the preference of which phrase to use differs

among various researchers, community members and policy makers. In most societies, the majority of individuals and what society expects is that most people will identify with the gender that they are assigned at birth (e.g., male or female) and that they will be attracted to and marry the opposite sex. For sexual and gender minorities, this is not always the case and this group of individuals face many challenges because of their diversity. According to Human Rights Watch, "People around the world face violence and inequality—and sometimes torture, even execution—because of who they love, how they look, or who they are." (Human Rights Watch 2021) This comes in the form of institutional barriers such as discriminatory laws as well as cultural and societal barriers. Discrimination takes many layers including state violence, criminalization, community stigma, community violence and violence from family members. Statistics and data on accurate population estimates of sexual and gender minorities is extremely limited as most governments do not collect this data through census or civil registration. Most population-based surveys also do not include these demographic questions due to lack of knowledge and stigma among many institutions and researchers.

1.1.2. LGBTIQ+ Data Collection

"Sexuality results in differential outcomes on a number of issues that are fundamental to population study, including migration, fertility, morbidity, and other areas." (Baumle 2013) Demography has been slow to address the heterogeneity of populations. Diverse sexual behavior in literature is rare and most articles that do mention it focus on relating to sexually transmitted infection. (Baumle 2013). Sexual orientation and gender identity correlates with behavioral health burden, and it is important to collect these data. (Cahill et. al 2016). For example, the lack of SOGIESC data not recorded systematically limits the identification of mortality disparities in LGBTIQ+ people (Haas 2015). One study from the United States found that the exclusion of SOGIESC data collection put the burden on patients to disclose their status rather than on the provider and institution. (Allysha et. al 2017) This may result in many people not sharing this information due to stigma, fear or not having the opportunity to share. In order to move

towards data collection of this group, stigma needs to be reduced and increased societal acceptance of this group. In Gilroy's recent 2021 research paper, the authors analyze the expression of sexualities in the contemporary United States using data about disclosure on social media. Through the Facebook advertising platform, the authors review generational differences and disclosure of sexual orientation by age, gender and relationship status to better understand who is disclosing this information through social media platforms and sociological implications. (Gilroy 2021)

Data on numbers of who is supportive of LGBT communities is rare in the academic literature. However, Gallup Polls have been conducted on public opinion related to LGBT for some time now. For example, a recent poll has shown that public support for same sex marriage have increased from only 27% in 1997 to 70% as of 2021.(Gallup 2021) There is very few research looking at LGBT support and internet data. One paper from 2017 analyzed Twitter data to review LGBT support in India. The researchers used deep learning machine learning (ML) methods to collect 580,000 twitters to gauge societal perceptions on LGBT there. Using a classification system, they found that Indian twitter users were more supportive of LGBT than those who opposed with ML accuracy at 85%. (Khatua et. al)

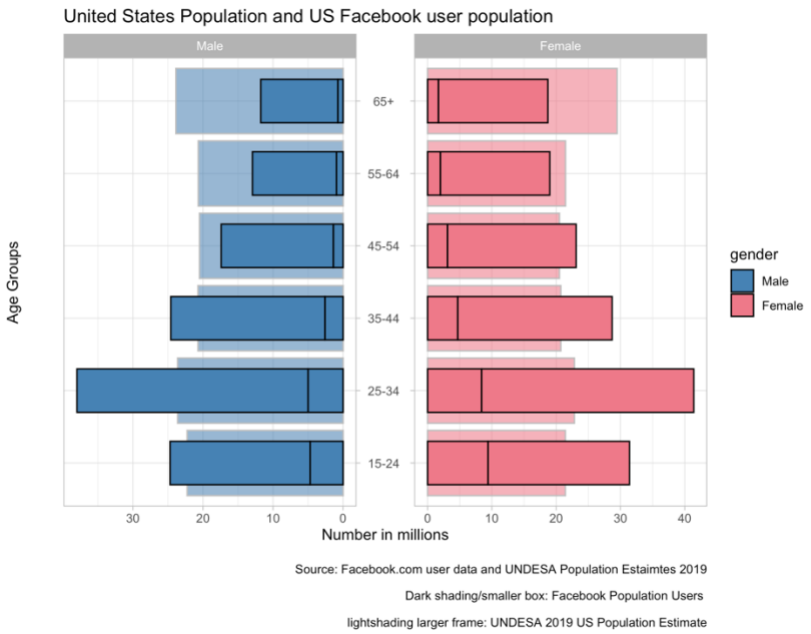
1.1.3. Types of Digital Trace Data

The world is everchanging and so is the field of demography. Technological advancements in the past ten to twenty years has allowed the world to gather, utilize and analyze new data sources and record amount of data than was even imaginable before. The Big Data Revolution and the popularity of social media sites across the globe is a huge opportunity for the demography field to understand macro-level changes. Recent studies have shown that digital trace data can provide us with high precision of large detailed data sets. These digital trace data can be roughly defined as "records of activity (trace data) undertaken through an online information system (thus, digital)" (Howison, Wiggins, & Crowston, 2011). Digital trace data is non-intrusive data that is being collected by service providers (e.g. Facebook, Twitter, Google, etc.) that are stored through Application Programming Interfaces (APIs). The levels of

detail can include many categories along with geotagged data to understand what regions or locations these data are in. (Stier, 2019) According to Data Population Alliance, there are many types of Big data sources ranging from mobile, transportation, financial, social media, and remote/physical sensing electronic data. For the purpose of this research paper, we are focusing in on social media digital trace data and how it has been used for demographic and population research.

1.1.4. Internet Data is non-representative

With opportunities of digital trace data also comes challenges. These include the fact that internet data is non-representative as not all people have access to the internet globally and even in countries where most of the population has access, the number of people who are using particular social sites varies depending on the platform. **Graph 1**: shows the difference between Facebook users in the United States as of 2021 and US Population estimates in the United States. As you can see the real population is larger starting at age 45 and up.



For ages 15-44, there are more Facebook users than there are population. This is most likely due to users having multiple unlinked Facebook accounts.

- In 2015, Zaheni and Weber wrote a paper detailing how to use internet data for demographic studies. They detailed two major scenarios (1) where ground truth data is available and the social media data can be compared to more traditional methods of demographic data and (2) where there is no ground truth data and a difference in differences statistical method would be used to help reduce bias. (Zagheni & Weber, 2015)

- Further demographic research using internet data has mostly quoted and referenced Zagheni's initial paper as a starting point for building these statistical models when using these types of data sets.
- A 2017 study in the United Kingdom using Twitter and Facebook data, compared social media data to a nationally representative probability sample to analyze political attitudes and demographic indicators. The researchers found some apparent differences between the two groups including age, gender and education. "On average social media users are younger and better educated than non-users, and they are more liberal and pay more attention to politics." (Mellon & Prosser, 2017)

1.1.5. Demographic Research using Social Media

Little demographic research exists using social media data due to the fact that this is a new data source and emerging field. The research that does exist tend to focus on Facebook and Twitter data as these sources tend to include high levels of users globally and API marketing platforms to help to extract data that would be useful for demographic research without any cost. Cesare et. al discusses the benefits and challenges of digital trace data to study processes related to fertility, morality and migration. {Cesare, 2018}

Most of the research has focused on migration patterns where researchers are using Facebook or Twitter data to analyze short term mobility trends. Some examples of this work include the following.

- Zagheni et all produced research in 2017 using Facebook data to monitor stock migration and was the first paper at a global scale to do this kind of work and has set the stage for future research in this area.
- The impact of Hurricane Maria on out migration from Puerto Rico (PR) to the United States using Facebook data. They found increase in migration and could locate which states they

found this increase along with return migration back to PR during the timeframe of data collection. (Alexander 2019)

- Spyratos has produced multiple research studies where they utilize Facebook to quantify international human mobility. They found that the data and methods for using Facebook data would correspond with future traditional sources of data that were later reported. They were also successfully able to capture and estimate and increase in Venezuelan migrations in Colombia and Spain in 2018. (Spyratoes et. al 2019, 2020) Zagheni et all produced research in 2017 similar to this framework.

2. Research Questions

Are women and younger generations more likely to be supportive and show support for LGBTIQ+ community on social media?

Hypothesis 1:

Women are more likely than men to show support of LGBTIQ+ community members through social media.

Hypothesis 2

Younger generations are more likely to to demonstrate support of the LGBTIQ+ community on social media compared to older persons.

3. Data and Methods

This project will utilize data from Facebook Marketing Ads Manager platform to understand how LGBT supportive behavior intersects with age and gender. This study utilized aggregated estimates as the outcome of the statistical regression model for three variables including interaction of our variables. The focus will be on Facebook and Instagram Users in the United States aged 15-64. The rationale for the age groups is that 15 is the start of the adolescent group and children under age 13 are not allowed on Facebook. The rationale for the cut off at 64 is that users age 65 and above are not disaggregated so it is

difficult to compare with the age groups below 65. The group of 65+ is also very small as they are not as active on social media due to generational gaps and preferences. Data is provided anonymised, aggregateted and rounded format to the nearest zero. Whenever a data point is less than 1000, "< 1000 users" is provided to ensure confidentiality and privacy is maintained. For this dataset, all entries are above 1000 users.

3.1. **Variables of Interest**

The main variable of interest is to look at LGBTIQ+ support with supplementary variables disaggregating by gender and age group. Further demographic categories are not possible through Facebook such as Race or education level.

Variable	Definition	Response Options	Type of Variable
LGBT support	Demonstrated support of LGBTIQ+ movements on Facebook and Instagram	1) Supportive 2) No support identified	Binary / Categorical
Age Group	Age group based on Generation categories	1) Gen Z (age 15-24) 2) Millenials (age 25-34) 3) Gen X - younger (age 35-44) 4) Gen X – older (age 45-54) 5) Baby Boomers (age 55-64)	Categorical 10 year age groups
Gender	Gender	1) Male 2) Female	Binary / Categorical
"Weights"	Estimated population size for each observation	Discrete value (>0)	

3.1.1. Dependent Variable

LGBTIQ+ Support is our main variable of interest and for the purposes of this research, we have coded this as a binary categorical variable due to the limitations of Facebook categorization. No previous studies have covered this topic using Facebook marketing data. The following categories were determined:

1. “Supportive” are Facebook and Instagram users who have demonstrated support towards the LGBTIQ+ community trough Facebook’s interest section under the marketing tool.

We have used the following three interest classification phrases, which best captures LGBTIQ+ support: "LGBT community", "LGBT culture", "LGBT social movements". According to Facebook, detailed targeting including interests and behaviors are based on “pages [people] engage with, activities people engage with” (Facebook 2021)

2. “No support identified” are all users who have no record of these behavior classifications.

The “No support identified” category maybe include people who are neutral as well as people who are anti-LGBTIQ+ support. While support may be possible to classify on a scale if this data is collected through surveys. It is not possible to do so through routine Facebook marketing data as this is passed on through algorithms and Facebook users’ history on the platform such as interactions and liked pages.

3.1.2. Independent Variables

Gender and age are the two main independent variables we are interested in this study. Additional indicators would have been included such as race and education if the platform allowed for this level of detail.

Gender (X_1) is collected for male and female Facebook users. It is not possible to include 3rd gender options as Facebook does not allow for Facebook marketing towards non-binary or gender non-conforming individuals.

Age (X_2) was collected for Facebook groups by 5 year age groups from 15-64 during data collection. For the purposes of this research, these were transformed to 10 year age-groups based on generational groups to help differentiate and simplify the analysis. The following age groups were determined:

- Generation Z ages 15-24
- Millennials ages 25-34
- Generation X – Younger ages 35-44
- Generation X – Older ages 45-54
- Baby Boomers ages 55- 64

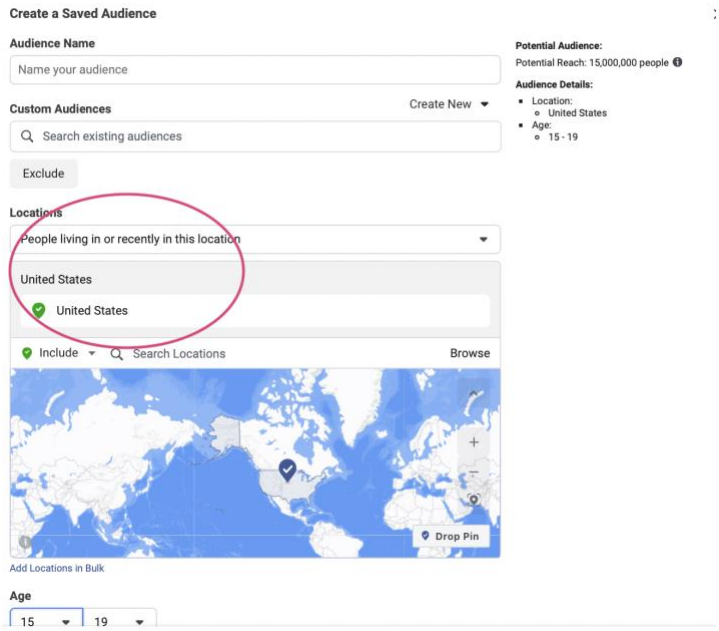
The age variable categories were created based on different generation groups that have been defined and used through sociology and media that helps to categorize groups based on common similiarities and differences. Using the Gallup poll definition and ensuring that the groups fit into 10-year-age group to ensure standardized groups. (Gallup 2021)

Age and Gender Interaction (X_3) is the third variable of interest in order to understand how gender and age interreact with one another to influence likelihood of showing support of LGBTIQ+ support on social media.

3.2. Timeframe and Data Collection

The data used is up-to-date estimates from 2021 as of time of data extraction on September 19, 2021. Data was extracted on September 19, 2021 from Facebook Ads Manager through Facebook Marketing website. Anyone can extract and utilize Facebook Ads Manger as long as you sign up for Facebook business account. We utilized the saved audience and potential targeting option (see Figure 1). As our audience is in the United States, we have chosen that as the geographic location. Creating potential audience targeting is free of charge. Data extracted from Facebook was then inputted into tidy format as a CSV file and uploaded into R for exploratory data analysis followed by statistical analysis.

Figure 1: Example of creating potential audience through Facebook marketing Ads Manger



4. Analysis and Results

4.1. Statistical Model

We propose to use descriptive univariate statistics, followed by descriptive and chi-squared test for bivariate statistical analysis and finally logistic regression model for multivariate analysis. Logistic regression model is used for a binary outcome variable. Logistic regression uses a binomial distribution to predict the outcome of Y based on various dependent variables. Y_i is a the realization of a random variable of Yi, that can take the form of 0 or 1 with probabilities of π_i and $1 - \pi_i$. This results in the binomial probability function and then takes the logit transformation. Logistic regression uses the Maximum Likelihood estimator to find the odds of binary outcome. (Rodriguez 2017) The basic logit model will take the log of odds for the various beta coefficients of our dependent variables.

$$ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$$

4.2. Proposed Model:

Our proposed model looks at gender and age along with gender and age interactions to understand the intersection of these two variables and their impact on our outcome variable of LGBT support. Our systematic and random components of our model include:

$$Y_i \sim B(n_i, \pi_i)$$

$$\text{Log(odds)} = \beta_0 + \beta_1 x_1 + \beta_2 x_{2a} + \beta_3 x_{2b} + \beta_4 x_{2c} + \beta_4 x_{2e} + \beta_6 x_{2f} + \beta_7 x_{2g} + \beta_8 x_{2h} + \beta_9 x_{2i} + \beta_{10} x_3 + \beta_{11} x_{3a} + \beta_{12} x_{3b} + \beta_{12} x_{3c} + \beta_{14} x_{3e} + \beta_{15} x_{3f} + \beta_{16} x_{3g} + \beta_{17} x_{3h} + \beta_{18} x_{3i}$$

*the interactions of variables will include all of categories within the age variable.

4.3. Analysis

4.3.1. Univariate and Bivariate Statistical Analysis

In order to better understand the data, a number of exploratory data analysis was conducted in R including data visualization, frequency tables and a review of univariate and bivariate analysis. Table 1 provides the detail of the descriptive statistics including our LGBT supportive population (19%) along with distribution of gender and age collected. The gender distribution is close to an even although women are more represented compared to men (women closer to 53% and men close to 47%). The age distribution is largest around 20-29 accounting for 30% of the sample. As age increased, the size reduces and is most likely attributed to smaller number of older persons, Gen X and Baby Boomers, using internet and social media compared with Gen Z and Millennials. Within the Gen Z age group and in particular, 15-19 we see smaller counts compared to older Gen Z and millennials most likely due to this group using other social media platforms (most notably Tik Tok) along with restrictions or parents not allowing them to use Facebook.

Table 1: Descriptive Statistics of the LGBTIQ Supportive Statistics (source: Facebook & Instagram)

Variable	Count (in millions)	Percentage (%)
LGBT Support		
- Supportive	43.1	19.1
- No Support Identified	183	80.9
Gender		
- Men	107	47.4
- Women	119	52.6
Age		
- 15-24 (Gen Z)	48.8	21.5
- 25-34 (Millennials)	64	28.2
- 35-44 (Gen X – group 1)	48	21.2
- 45-54 (Gen X – group 2)	36.2	15.9
- 55-64 (Baby Boomers)	39.4	12.9
TOTAL	226.4	

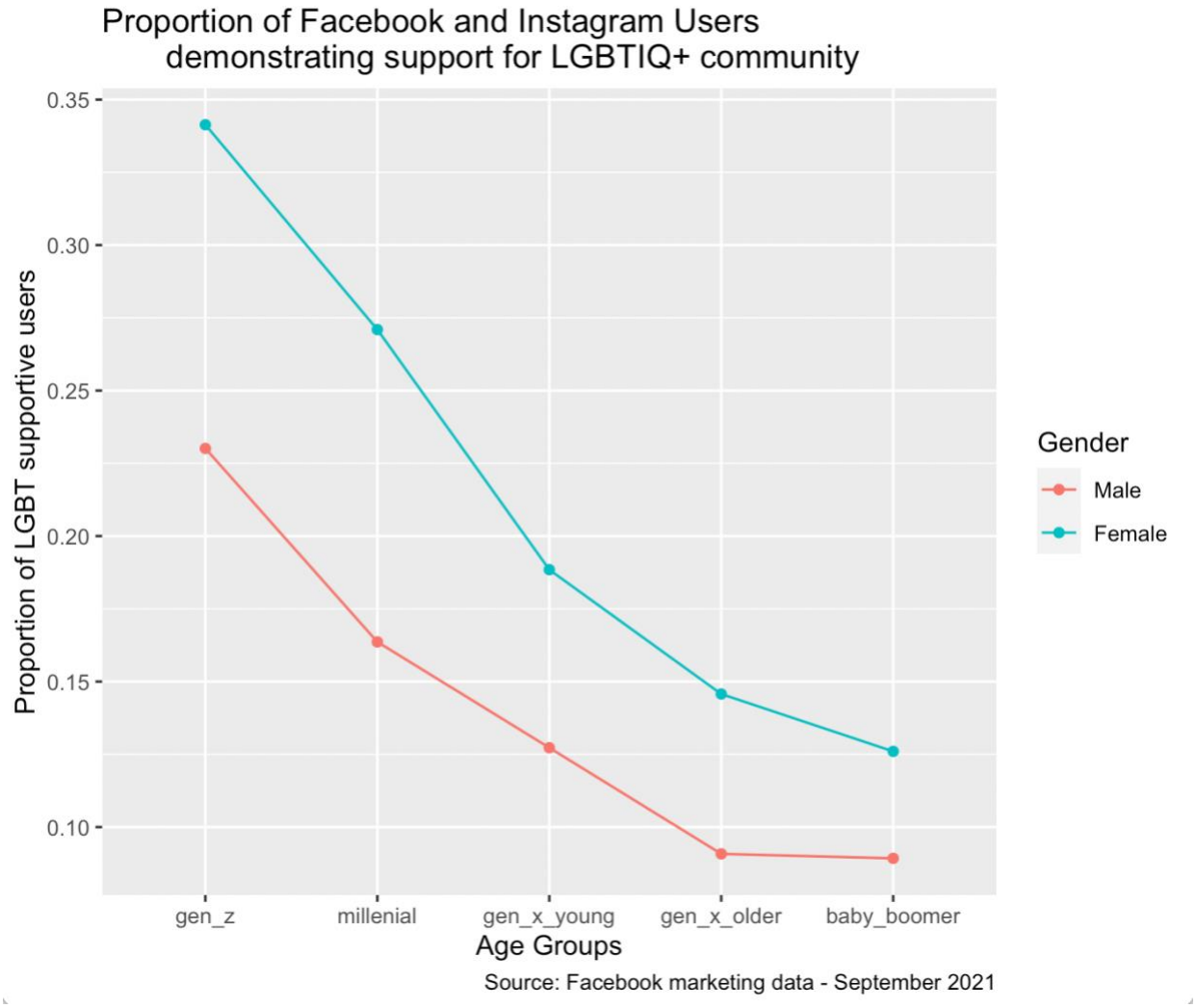
After conducting the initial univariate analysis of indicators, we reviewed the bivariate analysis comparing our outcome variable (LGBT support) against gender and age. The counts are provided (in millions) in Table 2 along with the Chi Squared test. All tests are considered to be statistically significant given the large sample size of 226.4 million. The proportion of women being LGBT supportive is greater than men at 22.6% compared with 15%. Additionally, the age groups are more supportive at younger ages and the proportion reduces as age group increases. This is shown in Graph 1, which calculated the proportion of supportive LGBTIQ+ Facebook and Instagram by age group and by gender. Female proportions are always higher for more supportive than their male counterparts in the age group. It is also interesting to note 4 age groups (women 15 up to 34) have higher LGBT supportive proportions compared to the highest male group (age 15-19). This is most likely attributed to societal stigma against effeminate boys and men resulting in high rejection among males compared to females for

LGBT persons. A bivariate analysis of the chi-squared test was also conducted to understand the expected vs. observed frequencies for these group.

Table 2: Bivariate Statistics of Facebook data collected comparing LGBT support (Y) with X1 (gender) and X2 (age).

Variable	**Count (in millions)**	**Count (in millions)**	**Chi Squared Test (in millions)**
	LGBT Supportive	No Evidence of Support	
Gender			2 ***
- Men	16.3 (15%)	91 (85%)	
- Women	26.9 (22.6%	92.2 (77.4%)	
Age			5.82***
- 15-24 (Gen Z)	14	34.8	
- 25-34 (Millennials)	13.8	50.2	
- 35-44 (Gen X – group 1)	7.7	40.3	
- 45-54 (Gen X – group 2)	4.38	31.82	
- 55-64 (Baby Boomers)	3.26	26.14	
TOTAL	43.14 (19%)	138.26 (81%)	
*** p-value < 0.00000000000000022			

Graph 1: Proportion of social media users demonstrating LGBT support by gender & generation



4.3.2. Multivariate Statistical Analysis

Multivariate analysis was conducted using logistic regression including interaction terms for age and gender. The results are included in Table 4 with the estimated coefficients, standard errors, and probability levels for each Beta. Table 5 takes the exponential form of the log (estimated beta) to come up with the odds ratio for each variable along with the upper and lower confidence interval and the change in odds. Females had 1.58 higher odds of showing support for LGBT support on Instagram or Facebook compared to men. For each generation, the reference group is the youngest group, Gen Z (age 15-24). All age groups have a decreased odd of being LGBT supportive compared to Gen Z. For each age

group, the odds decrease by more and more with the baby boomers having the lowest odds. Looking at the interaction of women age-groups as compared to the reference of Gen Z males, you will see that the odds of being LGBT supportive are higher only for millennial females. For Gen X and baby boomer females, they were less like likely than Gen Z males to show support. It is important to note that all of these findings are at the p value of greater than 0.0000000000000002. This is very significant findings and attributed to the large sample size of the study at over 226 million.

Table 4: Logistic Regression Results

	Log of Odds	***Odds Ratio***
Dependent Variable: LGBT Support	LGBT Support	LGBT Support
Gender - Female	0.550***	1.734***
	(0.001)	(0.001)
age_group - millenial	-0.424***	0.655***
	(0.001)	(0.001)
age_group - gen_x_young	-0.718***	0.488***
	(0.001)	(0.001)
age_group - gen_x_older	-1.096***	0.334***
	(0.001)	(0.001)
age_group - baby_boomer	-1.115***	0.328***
	(0.001)	(0.001)
Gender – Female : age_group - millenial	0.091***	1.096***
	(0.001)	(0.001)
Gender – Female : age_group - gen_x_young	-0.085***	0.918***
	(0.001)	(0.001)
Gender – Female : age_group – Gen X - Older	-0.015***	0.985***
	(0.001)	(0.001)
Gender – Female : age_group – Baby Boomer	-0.164***	0.848***
	(0.001)	(0.001)
Constant	-1.208***	0.299***
	(0.0005)	(0.0005)
Observations	10	
Log Likelihood	-83.563	
Akaike Inf. Crit.	187.126	
Note:	*p<0.1; **p<0.05; ***p<0. .0000000000000002	

5. Discussion and Recommendations

5.1. Results

5.1.1. Result 1: The odds of a female demonstrating LGBTIQ+ supportive behavior on Instagram and Facebook is 1.73 times higher than those of their male counterparts.

This finding is in line with our original hypothesis of females being more supportive. In the United States and in many countries across the globe, there is greater stigma of being LGBTIQ+ among men due to societal pressures and sociological norms of masculinity. Members of the LGBTIQ+ community, especially those who identify as male are less likely to see by society as conforming to what is expected of the male. “Demonstrations of heterosexuality, through both words and actions, are used by males to express their masculinity.” (Kelley, 2014) Kelley explored violence against LGBTIQ+ through the eyes of masculinity. Similarly, Herek has found that attitudes towards gay men were more negative compared with lesbians that further reflects societal gendered differences among men and women. His study also found significant differences among heterosexual men and women, where men held more negative values. (Herek, 2002) While this study is over 20 years old, we can see that this remains true today across all age cohorts included in this study. This finding is in line with the literature and expected outcomes of the research.

5.1.2. Result 2: Gen Z and Millennials are more likely to demonstrate online support of LGBTIQ+ groups. All age groups are less likely to demonstrate LGBTIQ+ support through behavior on Instagram and Facebook compared to the youngest cohort, Gen Z.

Our finding related to age and likelihood of support is in line with our expected hypothesis that the younger age groups are more likely to show support due to generational and societal differences among the cohorts. Gen Z tend to hold more progressive views than older generations and thanks to the impact of the media, LGBTIQ+ characters are more and more visible as compared to the past. (Hacknel 2016)

5.1.3. Result 3: Gen Z males have a higher odds compared with Females from Gen X and Baby Boomer generations to support for LGBTIQ+ on social media Baby boomer females were .848 times the odds of a Gen Z to show support for Gen Z.

Our third finding relates to our interaction variables of interest to understand how gender and age intersect to help explain behaviors and support for LGBTIQ+ movements. Age seems to be more of a factor than gender where females from Gen X and Baby boomer groups (35+) had a lower odd of showing support compared to their Gen Z male counterparts. This points to the fact that societal differences and generational impact may be a more important factor than gender when looking at the interaction. Future generations may include a more supportive generation and help to change attitudes to be more inclusive of LGBTIQ+ groups. Attitudes toward LGBT marriage and legal protections in the United States have increased over the past 40 years. A 2020 published study by Kaufman and Compton found that 60% of respondents favor LGBT marriage and 76% of them favor laws to protect LGBT persons. Although a sizeable minority of respondents, 40%, were in favor of small business to deny products or services to LGBT persons. They also found that in certain more conservative regions and those with higher approval of Trump were less likely to be support of LGBT persons. (Kaufman 2021)

5.2. Data feasibility and Limitations

Facebook Marketing API allows extraction of this type of data. Extracting data from Facebook Marketing through API requires programming with APIs. Research practitioners who are trying to replicate this study or utilizing this kind of data and research methods may find difficulty in extracting the data via API as there is a steep learning curve with utilizing API since it is a programming framework and language. Utilizing Python programming package may help for those who are proficient in Python coding language. Manual data collection is possible by utilizing the Facebook Ads Manager Audience interface in case of difficulty with the API although this process can be time consuming depending on the level of disaggregation you are hoping to extract. Additional limitations is that at any time, Facebook may

change what data is available free of charge from the Facebook Marketing APIs. For example, in the past research has looks as sexuality disclosure through Facebook, which is no longer available from Facebook. (Gilroy 2021) Social media research is not representative of the population (Alexander et al., 2019) and it generally tends to be younger than the overall population. "A majority of Americans say they use YouTube and Facebook, while use of Instagram, Snapchat and TikTok is especially common among adults under 30." (Pew Research 2021)

Another limitation of the study is the restrictions in some of the independent variables such as Race, Education or other variables that may be of interest. While these variables are not available to the public, Facebook has this data internally and future research collaborations to access some of the internal data could be beneficial for future studies to improve the methodology.

5.3. Strengths of the Study

This study has a number of strengths that are useful for researchers and practitioners in population studies, LGBTIQ+ rights movements and in public opinion. The ability to quickly collect data from Facebook and Instagram, in real-time, and at scale is one of the most significant ones. The aggregate level data is provided in the millions, which make it easier to have statistically significant findings. The ability to choose from a variety of topics and locations is another strength of this research methodology and data collection method using social media data. The research methodology also allows for agility in terms of more complex data collection and statistical analyses such a longitudinal and time series analysis by collecting this data over time and comparing changes to public opinion overtime. Social media data and big data has many promising opportunities in the field of social science and is under utilized within the field.

5.4. Policy Implications

While the original reason for data collection is for marketing purposes, these marketing tools can be adapted and use by those working on public policy, government and research. This research helps to

understand what more needs to be done to improve public opinion on LGBTIQ+ groups. More outreach and behavior change communication (BCC) is needed to improve the well-being and support for this marginalized group as the demonstrated support on social media is still quite low compared to the overall number of users.

5.4.1. **Multisectoral partnerships need to be established and maintained between Research Institutes, Governments and Private Sector companies, and in particular large social media and technology companies such as Facebook, Tik Tok, Google and others**.

Data that is available to the public is ever changing and some private sector companies restrict the level of information available due to privacy concerns and the ability for people to use the open-source data for bad. For example, in 2018 Facebook eliminated its Pages API invalidating many computational research methods from being replicated and eliminated a magnitude of publicly available information to become only accessible internally at Facebook. (Freelon, 2018) In 2021, Facebook is currently under investigation by a number of congressional committees in the United States and some European countries due to recent whistleblowers who have published internal documents and have blamed their algorithms for promoting hate and polarization in the name of profits over public safety. This may lead to social media companies to become more conservative with their data and additional legislative over technology laws and guidelines. This ever-changing field could greatly benefit from new partnerships to ensure that the data is used for good in line with public safety and adequate transparency.

5.5. Recommendations for Future Studies

As this is a relatively under-utilized methodology for collecting data in demography, there are many recommendations for future studies to utilize and build on this type of social media data.

1. Utilize time series methods to understand trends over time of social media data

Future studies should also use social media data to review changes of trends over time to better understand how perceptive and online behaviors change over time.

2. Collect data in multiple contexts and locations to understand comparative differences among internet users in different countries.

While this study looked at the United States context, it is possible to replicate this exact study in other locations and countries given Facebook's global use of almost 3 billion users. A comparative perspective will help rank and understand LGBTIQ+ support in different contexts and countries to support public policy makers and activists understand trends.

3. Sub-national studies to understand where more investment in BCC projects to improve LGBTIQ+ support.

Lastly, sub-national studies in the United States context for future studies could be useful to better understand where the lowest levels of support are and more investment might be needed on community activism, LGBTIQ+ programs and utilize behavior, change communication (BCC) strategies.

5.6. Conclusion

Support for Lesbian, Gay, Bisexual, Trans, Intersex and Queer plus (LGBTIQ+) communities have made significant progress and strides over the last 50 years in the United States and in many other countries globally. While public opinion is changing and some countries especially in the developed world have instituted progressive LGBTIQ+ support policies and laws, much more is still needed to be done. This research shows us that while we see increasing levels of support of LGBTIQ+ persons, demonstrated support through social and online spaces is still relatively small at only around 20% of Americans. Analyzing these social media and online trends will help understand public perspectives and online behavior towards this marginalized group. This new method of quickly gathering large scale aggregated information helps leverage the power of big-data and the 21st century move to an online and global data driven world.

6. Acknowledgments

I would like to thank my professor, Dr. Sorapop Kiatpongsan, PhD, for his invaluable feedback and oversight of the study design, data collection, analysis and manuscript writing. I would also like to thank my classmate Myo Thandar and Thao Nguyen for their feedback and support on the research and manuscript. Lastly, I would like to thank all of the persons who have demonstrated support of the LGBTIQ+ population as this helps to reduce stigma, promote equal rights and protection of LGBTIQ+ persons in the United States and globally.

References

Alexander, M., Polimis, K., & Zagheni, E. (2019). The Impact of Hurricane Maria on Out-migration from Puerto Rico: Evidence from Facebook Data [https://doi.org/10.1111/padr.12289]. *Population and Development Review*, *45*(3), 617-630. https://doi.org/https://doi.org/10.1111/padr.12289

Allysha C. Maragh-Bass, M. T., Rachel Adler, Eric Schneider, Anju Ranjit, Lisa M. Kodadek, Ryan Shields, Danielle German, Claire Snyder, Susan Peterson, Jeremiah Schuur, Brandyn Lau, and Adil H. Haider. (2017). Risks, Benefits, and Importance of Collecting Sexual Orientation and Gender Identity Data in Healthcare Settings: A Multi-Method Analysis of Patient and Provider Perspectives. *LGBT Health*, *4*(2), 141-152. https://doi.org/10.1089/lgbt.2016.0107

Baumle, A. K. (2013). *International handbook on the demography of sexuality*. Springer. https://doi.org/10.1007/978-94-007-5512-3_1

Cahill, S., Singal, R., Grasso, C., King, D., Mayer, K., Baker, K., & Makadon, H. (2014). Do Ask, Do Tell: High Levels of Acceptability by Patients of Routine Collection of Sexual Orientation and Gender Identity Data in Four Diverse American Community Health Centers. *PLOS ONE*, *9*(9), e107104. https://doi.org/10.1371/journal.pone.0107104

Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). Promises and Pitfalls of Using Digital Traces for Demographic Research. *Demography*, *55*(5), 1979-1999. https://doi.org/10.1007/s13524-018-0715-2

Facebook. (2021). https://www.facebook.com/business/help/182371508761821?id=176276233019487

Feeney, G., Keilman, N., Schmertmann, C., & Bijak, J. (2019). Editorial: The past, present, and future of Demographic Research. *Demographic Research*, *41*(41), 1197-1204. https://www.demographic-research.org/volumes/vol41/41/

https://www.demographic-research.org/volumes/vol41/41/41-41.pdf

Feldman, J. L., Luhur, W. E., Herman, J. L., Poteat, T., & Meyer, I. H. (2021). Health and health care access in the US transgender population health (TransPop) survey [https://doi.org/10.1111/andr.13052]. *Andrology*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/andr.13052

Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, *35*(4), 665-668. https://doi.org/10.1080/10584609.2018.1477506

Gallup. (2021). https://news.gallup.com/poll/1651/gay-lesbian-rights.aspx

Gilroy, C., & Kashyap, R. (2021). Digital Traces of Sexualities: Understanding the Salience of Sexual Identity through Disclosure on Social Media. *Socius*, *7*, 23780231211029499. https://doi.org/10.1177/23780231211029499

Haas, A. P., Lane, A., & Working Group for Postmortem Identification of, S. G. (2015). Collecting Sexual Orientation and Gender Identity Data in Suicide and Other Violent Deaths: A Step Towards Identifying and Addressing LGBT Mortality Disparities. *LGBT Health*, *2*(1), 84-87. https://doi.org/10.1089/lgbt.2014.0083

HEREK, G. M. (2002). Gender Gaps in Public Opinion about Lesbians and Gay Men*. *Public Opinion Quarterly*, *66*(1), 40-66. https://doi.org/10.1086/338409

Kaufman, G., & Compton, D. L. (2021). Attitudes Toward LGBT Marriage and Legal Protections Post-Obergefell. *Sexuality Research and Social Policy*, *18*(2), 321-330. https://doi.org/10.1007/s13178-020-00460-y

Kelley, K., & Gruenewald, J. (2014). Accomplishing Masculinity through Anti-Lesbian, Gay, Bisexual, and Transgender Homicide: A Comparative Case Study Approach. *Men and Masculinities*, *18*(1), 3-29. https://doi.org/10.1177/1097184X14551204

Khatua, A., Cambria, E., Ghosh, K., Chaki, N. (2019). *Tweeting in Support of LGBT? A Deep Learning Approach* Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, Kolkata, India. https://doi.org/10.1145/3297001.3297057

Lewis, D. C., Flores, A. R., Haider-Markel, D. P., Miller, P. R., Tadlock, B. L., & Taylor, J. K. (2017). Degrees of Acceptance: Variation in Public Attitudes toward Segments of the LGBT Community. *Political Research Quarterly*, *70*(4), 861-875. https://doi.org/10.1177/1065912917717352

LGBT Rights. (2021). Human Rights Watch. https://www.hrw.org/topic/lgbt-rights

McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E. S. (2015). Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing. *Sociological Methods & Research*, *46*(3), 390-421. https://doi.org/10.1177/0049124115605339

Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, *4*(3), 2053168017720008. https://doi.org/10.1177/2053168017720008

Singh, A., Halgamuge, M. N., & Moses, B. (2019). An Analysis of Demographic and Behavior Trends Using Social Media: Facebook, Twitter, and Instagram. *Social Network Analytics*, 87-108. https://doi.org/10.1016/B978-0-12-815458-8.00005-0

Spyratos, S., Vespe, M., Natale, F., Iacus, S. M., & Santamaria, C. (2020). Explaining the travelling behaviour of migrants using Facebook audience estimates. *PLOS ONE*, *15*(9), e0238947. https://doi.org/10.1371/journal.pone.0238947

Spyratos, S., Vespe, M., Natale, F., Weber, I., Zagheni, E., & Rango, M. (2019). Quantifying international human mobility patterns using Facebook Network data. *PLOS ONE*, *14*(10), e0224134. https://doi.org/10.1371/journal.pone.0224134

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2019). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, *38*(5), 503-516. https://doi.org/10.1177/0894439319843669

Independent Expert on sexual orientation and gender Identity. (2021) United Nations Office of the High Commissioner on Human Rights (UN OHCHR). https://www.ohchr.org/en/issues/sexualorientationgender/pages/index.aspx

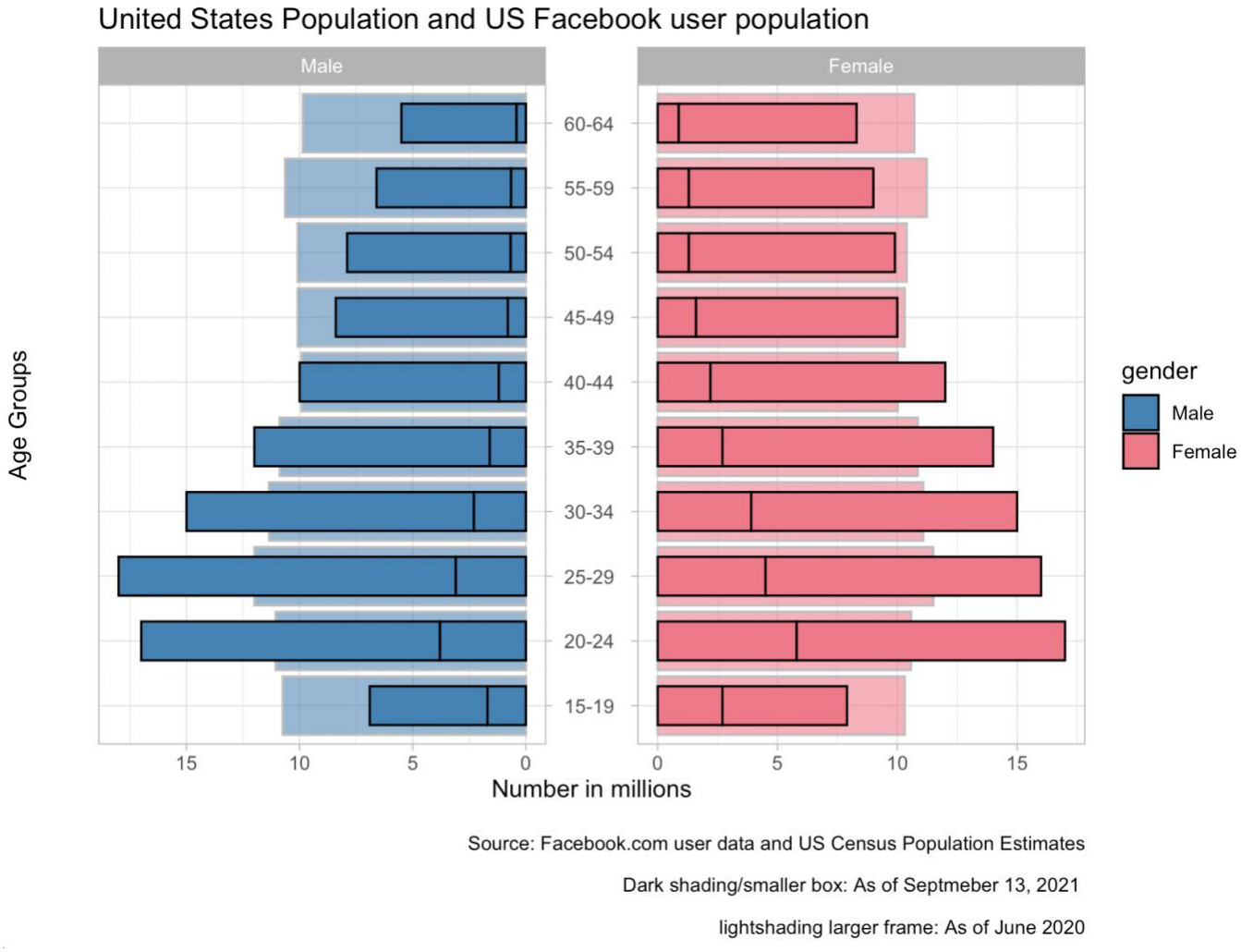
Yildiz, D., Munson, J., Vitali, A., Tinati, R., & Holland, J. A. (2017). Using Twitter data for demographic research. *Demographic Research*, *37*, 1477-1514. http://www.jstor.org/stable/26332233

Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, *36*(1), 13-25. https://doi.org/10.1108/IJM-12-2014-0261

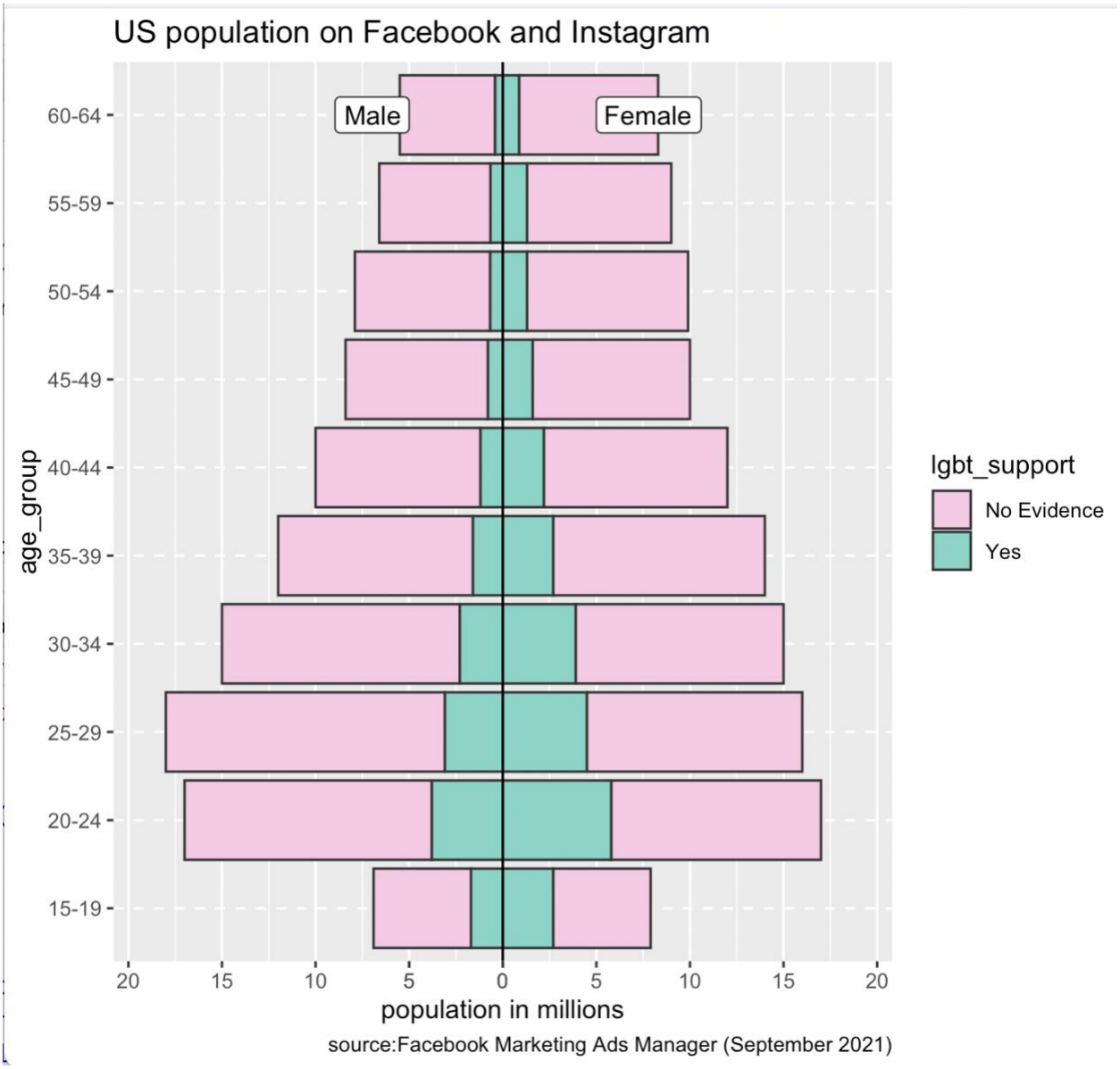
Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants. *Population and Development Review*, *43*. https://doi.org/10.1111/padr.12102

Appendix 1: Graphs

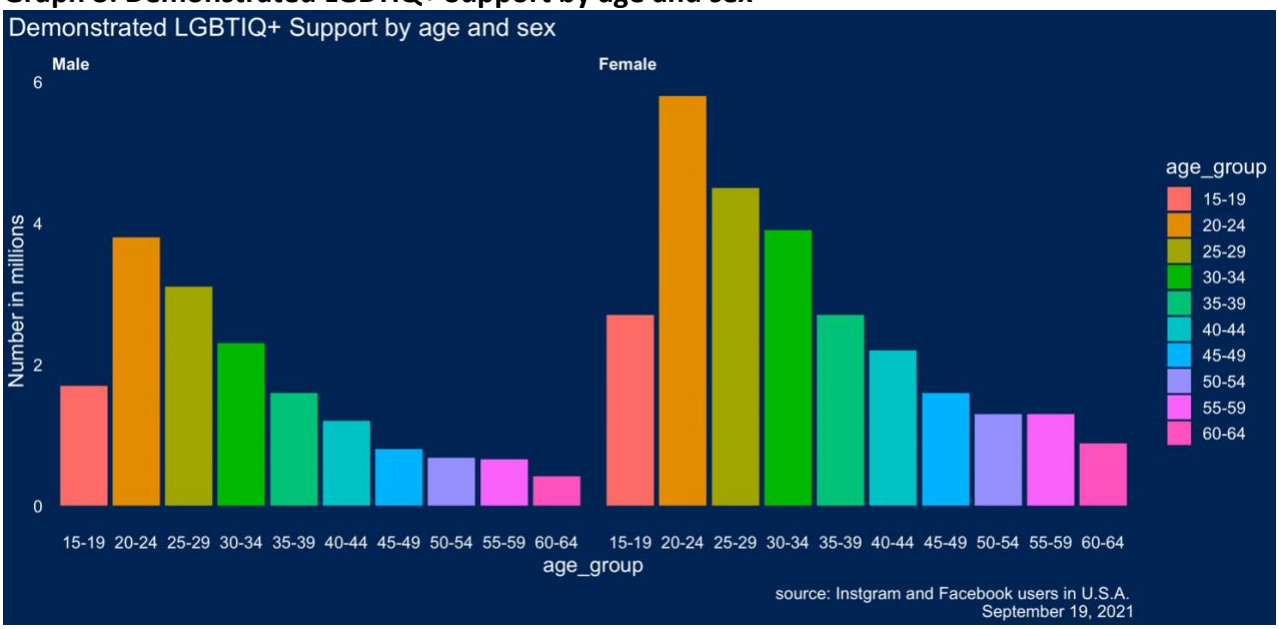
Graph 1: Population Pyramid of U.S. Population against US Facebook and Instagram Users



Graph 2: Population Pyramid of US Facebook and Instagram Users with demonstrated LGBTIQ+ support and no evidence of support



Graph 3: Demonstrated LGBTIQ+ Support by age and sex



Number in millions of those who have demonstrated their LGBTIQ+ support on Facebook and Instagram by age-group and gender.

Graph 4: Proportion of Facebook and Instagram Users demonstrated support online for LGBTIQ+ community