May 2021

Parenting, Vaccines, and COVID-19: A Machine-Learning Approach

Kathleen Thurson
Clemson University, kthurso@clemson.edu

PARENTING, VACCINES, AND COVID-19: A MACHINE-LEARNING APPROACH

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Parks, Recreation, and Tourism Management

---

by
Kathleen Marie Thurson
May 2021

Accepted by:
Dr. Ryan Gagnon, Committee Chair
Dr. Gwynn Powell
Dr. Barry Garst
Dr. Iryna Sharaievska

ABSTRACT

COVID-19 is currently at the forefront of both out-of-school time program providers' and parents' minds, with additional policies and procedures added existing operating standards to protect the health of participants, staff, and parents (Environmental Health & Engineering, 2020). A failure to adequately prepare and react to different parenting styles may have both operational and financial implications for out-of-school time programs. These implications are only further exacerbated in the additional context of a global pandemic. While the COVID-19 vaccine is a hope to many that the end of the pandemic is near, parental vaccine hesitancy or refusal may pose a significant hurdle to the safe operation of out-of-school time programs. By exploring the topics of vaccine hesitancy, children, and parents in an online environment, this study offers a closer look into a digital leisure space.

In order to better explore the conversations and commentaries occurring on social media about parents, children, vaccines, and COVID-19, web-scraping technologies were employed to aid in a more robust data collection. Due to the nature of web-scraped data as large in size and unruly, a machine learning method was used to analyze the data: Latent Dirichlet Allocation (i.e., LDA), a specific form of topic modelling. After establishing model parameters for the LDA, 25 latent topics were identified from the cleaned dataset (*N* = 31,925). These 25 topics were subsequently sorted into seven categories: Government, Feelings, School, Public Health, Christmas, Risk & Safety, and Parents & Families. Interpretation of the 25 latent topics was aided by a visualization of the top words most relevant to individual topics, in context to the overall dataset.

Representative tweets from each category further identified the range of conversations and commentaries occurring on social media about parents, children, vaccines, and COVID-19. Challenges with research at the cusp of innovation for leisure sciences, as well as implications of practice for out-of-school-time professionals, are also discussed.

ACKNOWLEDGMENTS

No one does anything alone, and I am so grateful to have such a strong network of mentors, friends, family, camp people, and colleagues supporting me. This was not always an easy process, but it was incredibly rewarding and I am so happy with the result. Without my chair, Ryan, I would be down an R rabbit (R-abbit?) hole somewhere on the Internet still. Thank you for pushing me to make things better and different, empowering me to make my own mistakes, and helping me learn from them later. Without Gwynn, I wouldn't be at Clemson, or working with Ryan! Thank you for always cheering me on when I needed it most. Without Barry or Iryna on my committee, I wouldn't have been able to take all the information floating around in my brain and put it on paper (in a coherent fashion, of course). Without my supervisor, Greg Linke, I wouldn't have been able to put all the new things I learned into practice, or get to stay up to date on the camp world. A huge thanks to the Palmetto Computing Cluster team as well, for help with getting set up for a big data project and brining the visualization online!

Good friends tell you what you want to hear, great friends tell you what you need to hear, and grad school friends tell you to take a deep breath, drink some water, and go finish your paper! Without Ali, Akiebia, Olivia, Laura, and Anitra, I would have been a much more difficult person to be around these last few months, and I am so thankful for such a wonderful group of women who care about me so much. Without Madi, Scoobie, Charlie, Hayley, and Roxie, I wouldn't have anyone keeping me current on the latest TikTok trends, or life outside my bubble. Without my committee within Graduate

Student Government, and Jodi with GRAD 360, I wouldn't have been able to stay connected to my community and keep my hands busy when I needed a different task. Without my camp families at Zeke, Happy Acres, Misty Mountain, the Y, ACA Southeast, and Camp 4U, I wouldn't have been interested in graduate school in the first place. And without my family in Jax, I wouldn't have the love and unyielding support I don't always deserve, but always feel, in everything I do.

TABLE OF CONTENTS

Table of Contents (Continued) Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

The COVID-19 pandemic abruptly transformed the summer camp industry, as stay-at-home orders and partial shutdowns limited the operation of out-of-school time programming in summer 2020. While some programs were able to operate (Blaisdell et al., 2020), others were forced or chose to shutdown (Szablewski et al., 2020); due in part to ambiguous or non-existent guidance at federal, state, and organizational levels in combination with a substantial level of labor and resources to effectively and safely provide out-of-school time programs. As out-of-school time providers prepare for another summer of programming amidst a pandemic, the development of COVID-19 vaccines offer potential mitigations to risks associated with the virus. With several vaccines in the beginnings of widespread dispersal (Dooling et al., 2020), health officials offer that widespread vaccination is the key to controlling the COVID-19 pandemic (Gee et al., 2021). However, concerns of vaccine hesitancy and skepticism persist as a threat to mandatory or highly suggested immunization (Quinn et al., 2009). Health and safety are foundational principles to the camp industry, and governing body, the American Camp Association. Given the confluence of deliberate misinformation, inconsistent communication, and increasing "pandemic fatigue" it is unsurprising the COVID-19 pandemic continues to challenge the camp industry's ability to provide safe programming.

In the context of out-of-school time programs, the American Camp Association and the American Academy of Pediatrics offer somewhat succinct guidance on immunizations for summer camps: routine vaccinations should both be mandatory and documented, and non-medical vaccine exemptions are not only inappropriate, but endanger public health (Ambrose & Walton, 2019). Some research has illustrated that vaccine hesitancy may harm the operations of out-of-school time programming (Garst et al., 2021a). The COVID-19 pandemic offers a timely context and event to explore growing parental discontent and attitudes surrounding vaccines, as vaccine hesitancy continues to grow, largely through online discussions via social media (Capurro et al., 2018; Kata, 2012; Sharevski, Jachim & Florek; 2020).

Social media is a rapidly developing environment to conduct research, especially in light of in-person data collection restrictions due to the COVID-19 pandemic. Social media offers a new “leisure space” where parents and other stakeholders engage in frequently unencumbered discussion regarding parental concerns, vaccine hesitancy, and COVID-19. This social media activity is referred to as digital leisure, or the unstructured time spent in digital environments, online, or using digital technologies (Redhead, 2016; Silk et al., 2016). In contextualizing social media as digital leisure, this study uses the social media platform Twitter as a data source to explore concerns related to parents, vaccines, and COVID-19.

Through a machine-learning approach, this study explores two questions: **(1) What are the conversations and commentary occurring on Twitter about parents, vaccines, and COVID-19? and (2) Can machine-learning help us explore this issue of vaccine hesitancy, in the relatively non-traditional context of social media?** In order to address both an exploration of Twitter data as it relates to parents, vaccines, and

COVID-19, and an investigation of machine-learning as a novel method, vaccine hesitancy is presented as an emerging concern for out-of-school time program providers. A primer on collecting data via the Internet, web-scraping, and other foundational concepts to machine-learning are also shared, in presenting machine-learning as the novel method used and then evaluated in this study. The subsequent analysis of web-scraped data through a machine-learning technique is paired with recommendations for the leisure and youth development sciences regarding the use of social media data and machine-learning as an exploratory research context, in combination with recommendations for out-of-school time professionals and researchers in regards to strategies for parent communication during a pandemic.

CHAPTER TWO

LITERATURE REVIEW

Vaccine Hesitancy

Vaccine hesitancy or refusal is an emerging and concerning concept within public health literature and profession (Larson et al., 2014). This concept provides a de-escalation of the pro-vaccine or anti-vaccine (i.e., anti-vaxx) discourse, offering a contextual spectrum to explain hesitancy or skepticism regarding vaccination. Parents who exhibit hesitancy towards vaccinations for their children may reject one or two vaccines, or seek to delay immunization, but nonetheless represent a heterogenous group (Estep & Greenberg, 2020; Opel et al., 2011). Origins (e.g., causes, determinants) of vaccine hesitancy are numerous across the literature, with primary factors including social or cultural differences, contextual issues, and medical or pharmaceutical specific issues (Dubé et al., 2013). Some research suggests up to 40% of medical providers would dismiss families who refuse routine vaccinations (Flanagan-Kylgis, Sharp, & Frader, 2005) which may only further parental anxiety and mistrust associated with vaccines (Leask, Willaby & Kaufman, 2014).

Due to the wide variety of attitudes and groups engaging in and/or influenced by parental vaccine hesitancy or refusal, it is important to understand the range, severity, and propensity of motivations for vaccine hesitancy. Indeed, vaccine hesitancy and/or refusal is referred to as “cultural epidemic” (McIntosh et al., 2015, p. 248) with regard to children’s healthcare, as parents are heavily influenced by sociocultural factors outside of the healthcare setting, including historic discrimination (Quinn et al., 2017), mistrust or worry towards healthcare systems or government agencies (Wiley et al., 2020), and individualism (Estep & Greenberg, 2020). These factors represent a perceived assumption of risk mitigation guided by parental choice rather than a doctor's orders (Sadaf et al., 2013). This presumption of parental expertise exemplifies the other previously mentioned factors, as parents are choosing what's best for their child based on their own research (e.g., individualism), experiences (e.g., discrimination), and fears (e.g., mistrust or worry) rather than adhering to previously-established vaccine schedules. Put simply, some parents are more willing to assume risks related to not vaccinating their child based on their own expertise, rather than their medical providers. Personal belief exemptions from routine vaccinations (e.g., non-medical exemptions) exacerbate the influence of vaccine-hesitancy on public health.

Exemptions from typical vaccinations or a deviation from the traditional vaccination schedule for children fall under three categories: religious, philosophical, or medical (Zier & Bradford, 2020). Non-medical exemptions (e.g., religious or philosophical) have been designated as inappropriate for a childcare setting, and a danger to public health (Ambrose & Walton, 2019). However, vaccination requirements for children differ at the state and local level in the United States, making adherence difficult to track (Estep & Greenberg, 2020; Zier & Bradford, 2020). Non-medical exemptions, along with increased cases of preventable communicable diseases are increasing in the United States (Capurro et al., 2018; Hargreaves et al., 2020). One environment frequently discussed in relation to increasingly concerning vaccine hesitancy is social media.

Social Media in the Context of Vaccine Hesitancy

Social media has become a central context to research regarding vaccine-hesitant parents (Hara & Sanfilippo, 2016; Jenkins & Moreno, 2020), as a mechanism to connect vaccine-hesitant parents with like-minded individuals, and as a way for researchers to explore the interactions between parents online (Gunaratne et al., 2019; Puri et al., 2020; Yuan, Schuchard, & Crooks, 2019). These sites reflect Internet-based communication in a community-setting (Blaszka et al., 2012), where conversations and collaborations can happen quickly and on a global scale (Filo, Lock, & Karg, 2015). Social media is a part of digital leisure, and therefore of interest to recreation and leisure scholars. Digital leisure is defined as non-work or non-required time spent engaged in digital environments (Schultz & McKeown, 2018), ranging a wide variety of activities, including the use of social media, general time spent online, and watching television (Redhead, 2016; Silk et al., 2016). This concept represents an increasingly interdisciplinary field, combining classical leisure research with cultural studies, information communication technology, sociology, and more (Spracklen, 2017).

Within the contexts of digital leisure, social media, and vaccine hesitancy, parents and caregivers with questions or concerns about vaccines often seek out information online. They are then faced with possible outrage (e.g., belittling or berating) from pro-vaccine voices when they are concerned, thus shutting down a possible communication channel to safely educate themselves (Capurro et al., 2018). Or, they are confronted with disinformation that enhances their fears or worries about vaccines (Bonnevie et al., 2019). A common factor in vaccine disinformation is Andrew Wakefield's widely discredited study (Horton, 2004) which linked the Measles, Mumps, and Rubella vaccine (i.e., MMR) to increasing rates of autism (McKeever et al., 2016; Yuan, Schuchard, & Crooks, 2019). The growth of vaccine-hesitant communities, both in-person (Attwell & Smith, 2017) and online (Jenkins & Moreno, 2020; Puri et al., 2020) has also spurred negative reactions from pro-vaccine voices (Capurro et al., 2018). For instance, a measles outbreak traced back to Disneyland in California led to 125 confirmed cases in 2015; 45% of which were not vaccinated individuals (Zipprich et al., 2015). The subsequent media coverage in both the United States and Canada vilified those infected and involved, as not vaccinating your child or yourself was described as intellectual, moral, societal, and ethical parental failure (Capurro et al., 2018; Yuan, Schuchard, & Crooks, 2019).

As noted earlier, social media is a key context for the growing levels of vaccine-hesitancy, as parents and caregivers look to online resources to investigate their concerns regarding their child's health-care needs (Park, Kim & Steinhoff, 2016; Schmidt et al., 2018). These social media sites often act as communities (Jenkins & Moreno, 2020) and are especially important in the face of contention or vilification of vaccine hesitancy from mainstream media sources, as many vaccine-hesitant or vaccine-refusing parents attest to the pressure or isolation they feel from mainstream (e.g., pro-vaccine) culture (Attwell et al., 2018). Digital leisure offers a context in which vaccine hesitancy issue can be further explored, specifically for leisure and recreation researchers and professionals. Emerging techniques such as web-scraping and machine learning, can help capture the often complex and large datasets associated with these vaccine hesitant communities.

Immunization Requirements in Out-Of-School Time Programs

While vaccine-hesitancy literature has primarily originated within public health, motivations and attitudes towards vaccines are also relevant to the out-of-school time industry, specifically in the context of a global pandemic (Ambrose & Walton, 2019; Garst et al., 2021). The American Camp Association (ACA) accredits out-of-school time programs across the United States and includes immunization requirements for both participants and staff members as part of their health and wellness accreditation standards. However, the standards regarding immunization (i.e., HW.1 & HW.15, in the ACA's Accreditation Process Guide v. 2019) do not enforce collection of immunization records. Rather, they require a signed statement from the parent or guardian, attesting that all immunizations are up to date. Both campers and staff members are allowed medical and non-medical exemptions from immunization under these standards, with an additional signed waiver or refusal form.

The Association of Camp Nursing (ACN), an entity that collaborates with both the American Camp Association and with the American Academy of Pediatrics (AAP), offers more strict guidance to camps and out-of-school time programs. The ACN urges camps to publish their immunization policy and collect full immunization history via a health history form (Erceg, 2020). However, the publishing of an immunization policy means that a policy must exist in the first place, and under the current accreditation standards from ACA an immunization policy is not typically required. This caveat reflects a current norm within out-of-school time programs for youth, as immunization policies often fall short or remain difficult to enforce (Garst et al., 2021).

Immunizations and immunization requirements bolster harm prevention policies for out-of-school time programs, from summer camps (Ambrose & Walton, 2019; Garst et al., 2021) to youth sports (Francis & Francis, 2020). However, there is a seemingly lack of formalization of immunization policies among many out-of-school time (i.e., OST) programs, but there are guidelines for other medically important issues facing children and other OST stakeholders. For example, Pop Warner football and cheer programs do not regularly collect immunization information from participants, but they do offer a full program related to other medical emergencies, with special attention to head injuries (Francis & Francis, 2020). As the COVID-19 pandemic continues, review of health and safety policies and procedures for out-of-school time programs are becoming more prevalent (Garst et al., 2021).

Infectious diseases considered to be eradicated in the United States as a result of childhood vaccinations were on the rise pre-COVID-19 pandemic (Opel & Marcuse, 2020). Vaccine hesitancy or refusal is a key factor to be considered as the COVID-19 pandemic continues, with the emergence of several vaccines that hope to stop the spread (Oliver et al., 2021). As vaccine hesitancy and/or refusal illustrates a spectrum of concerns and issues, parental involvement and anxiety also plays a role.

OST professionals already struggle with parent communication in a non-COVID-19 context, so the context of the COVID-19 pandemic may exacerbate existing parental worry and communication struggles between staff and parents. Communication between parents and OST staff persists as a problem which may be invasive in regard to the camp's program and goals (Garst et al., 2020; Garst et al., 2016). The escalating expectations of camper parents of increased and more thorough communication prior to and during their child's camp experience has led to increased strain on camp administrators (ACA Emerging Issues 2017; Garst et al., 2016; Garst et al., 2020).

Given the context of the present study (e.g., social media), a brief overview of how parents communicate online is necessary, particularly in regards to communication regarding health and wellness. With the advent of the Internet and social media, online information is now increasingly more accessible than a visit to the pediatrician's office (Baker, Sanders & Morawska, 2017). However, the veracity of health information online can be a concern, as no credentials are needed to join an online support group, post to a Facebook page, or tweet about your experience. Health information online can be classified into two broad categories: emotional and informational (Pretorius, Choi, Kang, & Mackert, 2020). This distinction assists in determining not only the veracity of the information given, but also in the intention behind it, which is relevant to the extended social network of social media and online communication.

From an information perspective, parenting blogs and social media pages (e.g., Facebook, in this context) offer experiential advice, or may suggest a visit to the pediatrician's office or another credentialed service (Mertan, Croucher, Shafran, & Bennett, 2021). The parents and caregivers using these social networking sites are typically seeking information regarding a concern for their child, whether that is about mental health resources (Mertan, Croucher, Shafran, & Bennett, 2021), Sudden Infant Death Syndrome (Pretorius, Choi, Kang, & Mackert, 2020), or fathering tips in general (Scheibling & Marsiglio, 2020). While a digital divide between parents of varying socioeconomic status was once thought to be present in the search of online health information, that is not the case. Parents of both higher and lower socioeconomic status use online resources to aid in their online health information search (Baker, Sanders, & Morawska, 2017). From an emotional perspective, parents use social media sites to vent, grieve, or otherwise share their emotions with a community of like-minded individuals who may be experiencing similar events (Pretorius, Choi, Kang, & Mackert, 2020). In collecting data from social media sites, this study presents an opportunity to explore both informational and emotional conversations and commentary about vaccine hesitancy during COVID-19.

Machine Learning for Leisure and Recreation Scientists

In the context of this study, web-scraping is the data collection or extraction tool, and machine-learning is the method used for data analyses. Before presenting the current study, the following sections offer some basics of web-scraping technologies. This is done in order to facilitate an answer to the study's second research question (Can machine-learning help us explore this issue, in the relatively non-traditional context of social media?) and demonstrate the technique's potential usefulness to researchers interested in out-of-school time programs, youth, and parenting.

Therefore, after exploring the basics of web-scraping as a data collection process, machine-learning will be discussed in a similar fashion: some basics of machine-learning and its potential usefulness to out-of-school time researchers. While these two concepts of web-scraping and machine-learning are relatively novel for out-of-school time researchers, research utilizing machine learning has rapidly expanded across the social sciences including: public health (Allem et al., 2018; Luo, Zimet & Shah, 2019; Yuan, Schuchard & Crooks, 2019), environmental science (Dahal, Kumar & Li, 2019), and communication (Linvill & Warren, 2020).

Web-Scraping

When sharing their parenting styles and techniques with researchers, parents may choose to discuss what they think is appropriate and rational, rather than their typical behaviors (Huber et al., 2018; Morsbach & Prinz, 2006; Napolitano et al., 2018). These views shared with a researcher may not reflect reality, but rather a more socially desirable response. Social desirability illustrates a typical challenge to survey-centered, interview-, and focus group-based research (Nederhof, 1985; Grimm, 2010). Survey research also comes with a potentially low return on investment in relation to funding and client outreach, as well as a large time commitments even when using previously validated measures (Landers, Brusso, Cavanaugh & Collums, 2016). One approach to mitigate these limitations is the use of social media content, collected through the use of web-scraping technology.

Mechanics of Web-Scraping

Web-scraping (i.e., web or content mining) is the (semi)automated collection or extraction of content from webpages (Cooley, Marbasher & Srivastava, 1997). In a "scrape", a researcher might search or pull information from a specific website or collection of websites, such as the American Camp Association, such as blog titles or authors. A web-scrape focused on individual or group use patterns and networks, would search or pull data from users of a specific site, like Twitter, Facebook, or Instagram. The content collected in an individual or group use focused web-scrape could be tweets, posts, or replies, while in a content-based web-scrape the information would be larger sections of text, headings, and other information available on a website (Landers, Brusso, Cavanaugh & Collums, 2016).

Information Structure on the Internet

In order to understand how a web-scrape is generally conducted, a brief introduction to the structure of information on the internet is useful. The primary language of the Internet is HTML (i.e., Hypertext Markup Language), and by using HTML the users of the front-facing or visual website, can interact with information easily, without reading through lines of code (Antoniou & Van Harmelen, 2011). By creating HTML objects or categories, website developers create web pages that users can easily interact with, while also adhering to best practices in website creation and development.. The common language of HTML and its affiliates facilitate web-scraping, as the structure of the data housed in webpages is similar across platforms, sites, and content (Antoniou & Van Harmelen, 2011). At the risk of making an overly broad generalization, a scrape can treat websites like a series of spreadsheets.

In terms of actual data collection or extraction from the Internet, web-scraping technologies vary, and often depend on the context and purpose of the scrape. In the present study, web-scraping via APIs (i.e., Application Programming Interfaces) are the primary tool to gather data. However, without a public API, researcher-designed web-scrapes (i.e., algorithms written in computer code by the researcher or research team), can be also be implemented (Freelon, 2018). For the present study, an API designed by

Twitter for research was used (Twitter, 2021). APIs are the connectors or communication-facilitators between computer programs, allowing these different websites and web-based content programs or tools the ability to interact with each other through an endpoint (Twitter, 2021). Researchers must apply for a "Twitter Developer" account in order to access the Twitter API, and then use a computing or statistical software of their choice to manage the search or pulls from the API (Twitter, 2021). For context, a tweet is a message of 280 characters or less sent via the Twitter website or mobile app (Twitter, 2021), are the main unit of analysis for this study.

Figure 1. Example of a Tweet & Reply

After the data is scraped using whatever selected tool(s), it is generally transferred to another software package (i.e., R, Python) to be analyzed. While unrelated to the focus of the present study, it may be clear that identifiable and personal data can be easily collected by APIs. To mitigate this concern within the context of Twitter, the connection between the API and the computing or statistical software, in this context the open-access software R (R Core Team, 2021; version 4.0.4) must also be authenticated in order to protect user information (Kearney, 2019). Twitter uses a process called Open Authentication (i.e., OAuth), in which the researcher is given unique credentials via their Twitter Developer account, which ties their search or use of the Twitter API to their account, regardless of the statistical or computing software used (Kumar, Morstatter & Liu, 2014). A more detailed account and the subsequent code used to scrape Twitter using the Twitter API is in the proceeding methods section, but this section offers a macro-level view of web-scraping technologies and Twitter. The following section builds on this foundation of the technological aspect of web-scraping and explains why web-scraping is useful to out-of-school time researchers.

Connection to Digital Leisure Studies

As previously discussed, social media intersects with digital leisure studies and communication research. The use of social media content as data is not necessarily new to the leisure studies field (see Lopez, Muldon & McKeown, 2020, Outley, Pinckney & Brown, 2020; Pinckney et al., 2018), but the collection of social media content with web-scraping technologies is less apparent. Web-scraping allows for the extraction of social media content in a manner that facilitates more robust and replicable data collection, (i.e., more data and search parameters). Web-scraping also facilitates replicability, as the code/syntax/script used to collect data, is be published with the study and reproducible by other researchers as needed (Jacobi et al., 2016; Welbers, Van Atteveldt & Benoit, 2017)].

Data Management

One challenge researchers using web-scraping technologies face is data management post- collection, as web-scraped data is both visually and structurally different than more typical data in out-of-school-time research (i.e., questionnaires, interview transcripts). Within the context of web scrapping data management is often described using the four Vs: velocity, variety, volume (Laney, 2001, Fan & Bifet, 2013) and veracity (Lukoianova & Rubin, 2013). First, velocity refers to the speed at which web-based data, in this case social media, is generated (Laney, 2001; Russom, 2011; Rodriguez & Storer, 2020). In 2020, 500 million tweets were sent per day, (Omnicore, 2021),amounting to approximately one hundred eighty-two billion five hundred million potential data points. Variety refers to the increasingly diverse range of content available on web-based platforms (Fan & Bifet, 2013; Rodriguez & Storer, 2020). On Twitter, users can share messages containing text, pictures, videos, and links to other websites. Volume refers to the large amount of content available on web-based platforms (Fan & Bifet, 2013; Rodriguez & Storer, 2020). In continuing with the previous example given regarding the 500 million tweets sent per day, multiplying that number by 365 days exhibits the volume of data just within the Twitter platform.

However, the entire volume of tweets may not be usable for a research study, leading to the fourth V; Veracity, refers to information quality (Lukoianova & Rubin, 2013), as a reoccurring issue in both web-scraping and machine learning studies is the large amount of unusable data, characterized most often as non-unique data points (e.g., retweets; a tweet that has been forwarded from a different user) or uninterpretable words and characters (Allem et al., 2018; Dahal, Kumar & Li, 2019; Twitter, 2021). A contextual example of veracity with the present study's data is presented below. These considerations of velocity, variety, volume, and veracity are important in understanding the technical complexity and value of web-scraped data.

In a review of data science innovations' applicability to the organizational science field, Tonindandel, King, and Cortina (2018) offer the following points to illustrate the potential web-scraped data provides social scientists: opportunities to investigate old questions in new ways and opportunities to address emerging practice needs. Web-scraped content in a social media context offers the opportunity investigate new questions, with emerging technologies and understanding of digital spaces (e.g., Bonnevie et al., 2020; Yuan et al., 2019). Social media data is publicly available and easily accessed in exceedingly large quantities, with content creation happening constantly (Allem et al., 2018; Sinneberg et al., 2017).

The relative ease of accessibility with social media data, along with the associated volume, has facilitated methodological improvements in the social sciences, most notably in the use of machine-learning as a method to assist in the analysis of datasets deemed too large or unruly for more traditional quantitative analysis (Lucas; 2020). After data collection, in this case web-scraping, data analysis begins. The following section discusses the analytic methods used in this study: machine-learning.

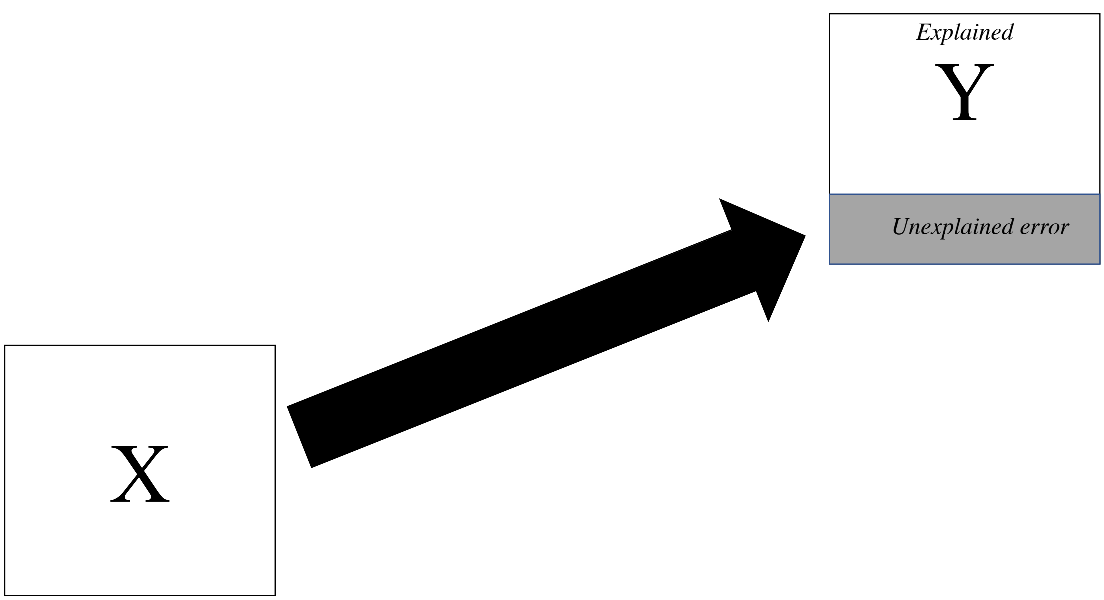
Machine Learning

Machine-learning is an intersection between computational science, statistics, and communication, defined as an automation of learning process algorithms (Mitchell, 1997). Put simply, machine-learning allows computers to learn and be taught, and then generate predictions based on the prior and incoming data (Burger, 2018; Lantz 2019; Landers, Brusso, Cavanaugh & Collums, 2016). For instance, a search engine's autocomplete feature can be eerily correct or humorously off the mark, but both instances are examples of machine learning (Goldberg et al., 2020). Machine-learning allows for the automation of tasks that would take an extraordinary amount of time/resources if attempted by a human. A human may be able to reasonably analyze the content of 200 tweets, but 20,000 could be untenable. This is the primary reason for machine learning: we have simply too much data to analyze using the techniques of the 19th and 20th century.

Linear Regression

The simplification of the algorithms that make up machine learning do not only describe the model this study uses—topic modelling—but also a model more familiar to the social sciences: linear regression. Put simply, linear regression predicts one variable or outcome from a single independent variable (Field, 2012): more of this (x), leads to more of that (y) (see Figure 1). While machine-learning models can get increasingly more complex, linear regression informs the overarching science of machine-learning (Burger, 2018; Lantz, 2019). Like machine learning, linear regression also results in unexplained variance or error. Regression models are predictive, as the independent variable (x) predicts the dependent variable (y) with a degree of mismatch, (i.e., unexplained/unsystematic error). Algorithm and model are sometimes used interchangeably, but in the context of this study these two terms are distinct (See Table 2 for guiding machine-learning definitions). An algorithm is a set list of instructions, to be followed rigidly and in a prescribed order (Burger, 2018). An algorithm(s) is then passed into a model, as a model requires some input to then calculate an output (Burger, 2018). There are several different machine learning model types, and this study focuses on classification models designed for text data: Latent Dirichlet Allocation (Blei, Ng & Jordan, 2003). Latent Dirichlet Allocation (i.e., LDA) is form of topic modelling that uses natural language processing (i.e., NLP) techniques to offer generative classifications of data (Hagen, 2020). While the boundaries between natural language processing and machine learning have blurred with the advent of advanced computing technology, a working understanding of both fields offers a better foundation for the current study.

Figure 2. Linear Regression



Natural Language Processing

Natural language processing (i.e., NLP) rose out of the linguistics and artificial intelligence fields (Nadkarni, Ohno-Machado & Chapman, 2011). NLP is often described as the manipulation of natural human language by computing technology (Bird, Loper & Klein, 2009); as a process used to aid computers in understanding natural languages (e.g., English, Spanish, French). Virtual voice assistants, such as Siri or Alexa, are popular applications of NLP in daily life (Hagen, 2020). Siri or Alexa are able to process input (e.g., human voice commands; "Where's the closest gas station?"), and then produce output in the form of verbal communication, often paired with web-based location services (Burbach et al., 2019). The relation between NLP and machine learning is relatively complex compared to linear regression but can be understood through the model implemented in the present study: Latent Dirichlet Allocation (i.e., LDA), a form of topic-modeling. LDA is machine-learning applied to natural language processing (Hldaka & Holub, 2015). Keeping in mind the definition of a model (see Table 2), the following sections offer a conceptual explanation of what Latent Dirilecht Allocation (LDA) is, and then a discussion on why LDA is useful to social scientists studying out-of-school time.

Table 1. Guiding Definitions: Machine Learning

Term	Definition
Algorithm	set list of instructions, to be followed rigidly and in a prescribed order[1]
Model	"a function with predictive power"; requiring input and output[1]
Machine learning	A process or set of instructions that allows computers to learn from data, and then generate predictions[1]

Supervised machine learning	Deductive processes reliant on researcher-involvement (coding, annotations, themes)[2]
Unsupervised machine learning	Inductive processes without coding roles or research annotations in which a model is determined by the data[2]
Boolean operators	AND, NOT, or OR; used to refine searches[3]
[1] Burger, 2018; [2] Lucas, 2020; [3] Dinet et al., 2004	

Latent Dirilecht Allocation

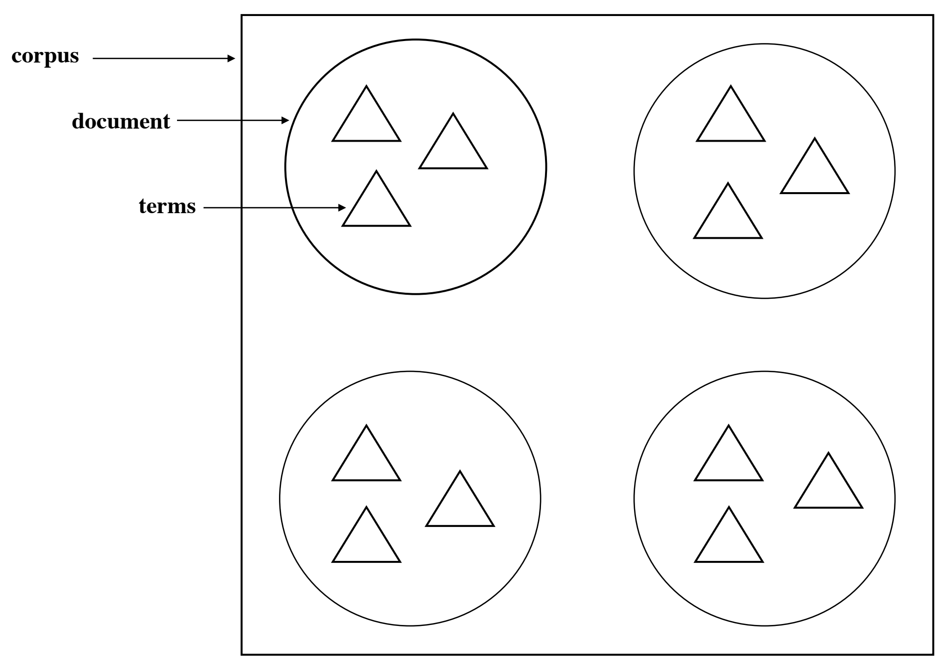
Introduced by Blei, Ng & Jordan (2003) Latent Dirilecht Allocation (LDA) is a generative, probabilistic Bayesian model which identifies topics across a collection of data (Ostrowski, 2015). In the context of LDA, generative refers to the input-output nature of the model where there is generation of content or output after the model is run. Similarly, probabilistic refers to the structure of the algorithm employed by an LDA model; this is best explained using the "bag of words" analogy (Blei, 2012; Ostrowski, 2015; Rodriguez & Storer, 2020; Silge & Robinson, 2020). A bag of words assumption on a basic level assumes that the position of the words in a sentence do not matter (Blei, 2012). LDA uses a hierarchical structure (see Figure 3), beginning with the corpus (e.g., the entire dataset), then the documents (e.g., each tweet is a document in this study), and the terms (e.g., words within each tweet) (Blei, Ng, & Jordan, 2003; Jacobi et al., 2016).

LDA is a series of probability distributions which use the Dirilecht family of distributions, commonly used in Bayesian statistics (Maier et al., 2018). There are two distributions within an LDA model: 1) the latent topics' distribution over words, and 2) the collection of documents' distribution over the topics (Blei, Ng, & Jordan, 2003). An LDA model develops latent categories based on repeated word occurrence in documents.

A major assumption of LDA is that documents are a mixture of the latent topics, so words from one tweet may show up in multiple topics (Silge & Robinson, 2020). Much like in multiple regression, where multiple independent variables may co-vary/interact, and thereby better explain a dependent variable.

As noted earlier, machine-learning is a method of analysis which uses computers to assist researchers in developing algorithms and models resulting in the generation of predictions (Burger, 2018; Lantz, 2013). While machine-learning is a relatively numbers driven approach, several machine learning models lend themselves to textual analysis. This study uses a machine-learning approach, applied to text data (e.g., tweets) as a natural language processing technique to demonstrate the possibilities machine-learning methods offer out-of-school time researchers.

Figure 3. Latent Dirichlet Allocation[1]



[1] based on Maier et al., 2018

The present study

This study explored two research questions: (1) What are the conversations and commentary occurring on Twitter about parents, vaccines, and COVID-19? and (2) Can machine-learning help us explore this issue, in the relatively non-traditional context of social media? In order to develop meaningful recommendations for researchers and for practitioners, two guiding questions were used to aid interpretation of results: (1) How can out-of-school time professionals better equip themselves and their staff to address parent concerns related to health and safety in OST? and (2) How can leisure and recreation scientists use machine-learning in their own research?

CHAPTER THREE

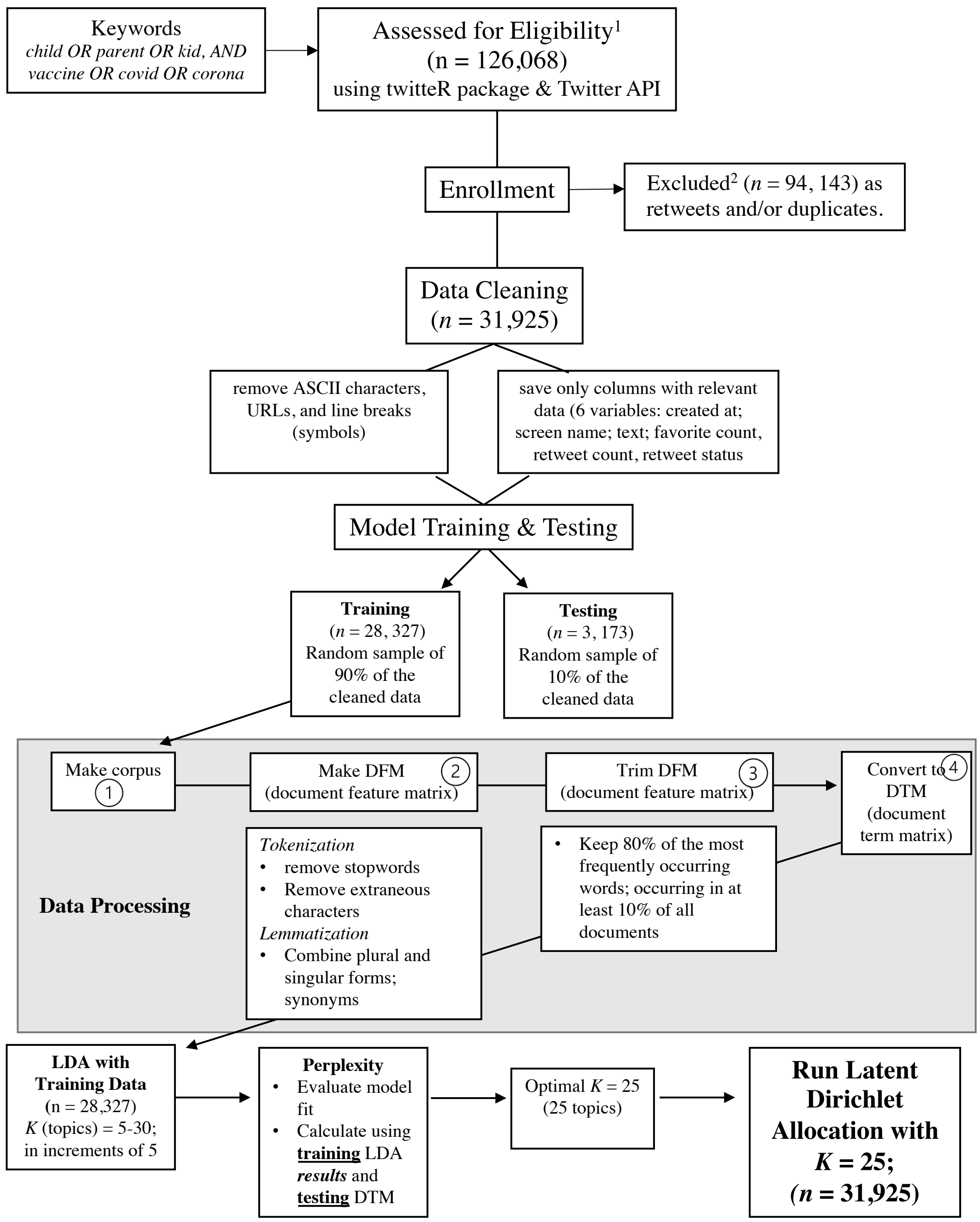
METHOD

Data Collection

Data were collected using the Twitter Application Programming Interface (API) between December 14, 2020 to December 21, 2020 (See Figure 4). Consisting of 126,068 tweets, the data were collected through the use of the rtweet package v0.4.0 (Kearney, 2017) in RStudio v.1.3.1056. Any web-scrape of Twitter using the Twitter API requires user authentication; user-authentication is an approval process through Twitter that works to ensure privacy standards and data protection (Kearney, 2019; Twitter; 2021). Twitter was selected as a data source due to the established evidence of conversations on Twitter leading to "real-world" behaviors and authentic discourse regarding vaccines (Bonnevie et al., 2020; Sinneberg et al., 2017).

Collection through the API was filtered in two ways: (1) date, as only tweets sent within the previous seven days are available to the API and (2) keywords with Boolean operators. The keywords utilized within this study were child OR parent OR kid, AND vaccine OR covid OR corona, notated in R script as child OR parent OR kid (vaccine OR covid OR corona). As a study focused on children and parents within the context of vaccines and the COVID-19 pandemic, it was important to have both sets of Boolean operators, to ensure that tweets collected mentioned both topic areas (Allem et al., 2018; Dahal, Kumar & Li; 2019). Additional information was included in the raw dataset related to user engagement such as likes, replies and retweets.

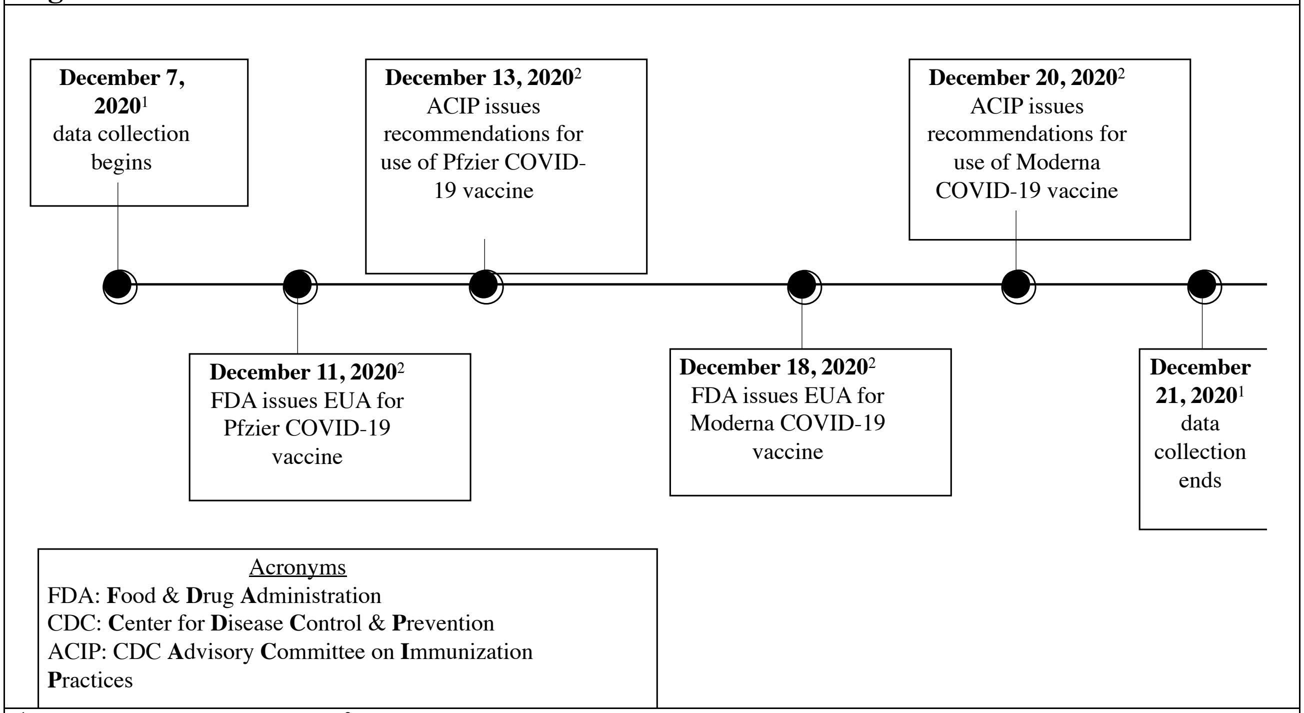
Figure 4. Study Method (Data Flow)



The Twitter application programming interface (i.e., API) streams a random sample of 1% of all public tweets from the last seven days, at the time of data collection

(Dahal, Kumar & Li, 2019; Karami et al., 2018; Twitter, 2021). Researchers are impacted by rate limits which restricted pulls (e.g., collected tweets from the web-scrape itself) to 18,000 tweets per 15 minute window (Kearney, 2017). Therefore, the tweets available for data collection and subsequent analysis are not only limited by keywords and Boolean operators, but also by the Twitter API process. Keywords refer to the words used as focus points of the API, to narrow down the data available for collection (e.g., parent, child, kid, vaccine, covid, corona). Boolean operators work as the connection points between keywords, similar to how conjunctions work in the context of grammar. For example, a Google search with "kids AND vaccines" would only show results that both the words "kids" and "vaccines" were included in. Tweets that were not publicly available (e.g., Twitter users with private accounts) or beyond the seven day window at the time of data collection were unable for use in the present study, as the API is unable to collect data from outside the seven day window or private accounts.

Figure 5. Data Collection & Immunization Authorization Timeline



[1] see Data Collection Section; [2] CDC, 2021

As noted earlier this study's parallel purpose is to introduce machine-learning as an emergent method within the out-of-school-time research (Behrend & Landers, 2019; Karamshuk et al., 2017). Machine-learning, and many quantitative analyses at large, are often categorized as fully empirical or free from researcher bias (Birhane, 2021; Mehrabi et al., 2019). Machine learning analysis is a series of both computer and researcher led decisions, which ultimately shape the output. However, the decision to include the keywords used in this study, to remove certain aspects of the data (e.g., retweets, links to other websites, line breaks), and the multiple processing stages are all examples of researcher-driven decisions which subsequently may affect the outcome/interpretation of analyses (Jacobi et al., 2016).

Data Cleaning

Prior to the analyses, the dataset went through both cleaning and processing stages in RStudio to facilitate both usability and study reproduction, following the recommendations of Jacobi et al. (2016) and Maier et al. (2018) (Figure 4). Complete study data and code are available on the author's website (https://katiethurson.github.io/LDAvis/#topic=0&lambda=0.6&term=). Data cleaning involved removing retweets (i.e., non-unique tweets, similar to a copy and paste or email forward), as well as ensuring the remaining tweets were interpretable to human researchers. ASCII (computer encoded symbols), URLs or external links, and line breaks were also removed at this stage (see example Table 3). Due to the size of the initial dataset (*N* = 126,068 tweets), data cleaning was done using RStudio on the Palmetto Computing Cluster, to facilitate more efficient computation. All subsequent stages (i.e., processing and analysis) were conducted using RStudio (v.4.0.4) run on a local server and using the Palmetto computing cluster.

Table 2. Raw and Cleaned Tweet Example

Raw Tweet	Cleaned Tweet
I just hope that all parents who decide to vaccinate their kids .. also decide to give their kid the COVID vaccine. Why pick & choose which vaccines to take now ? Ô£ø√º¬ß√Æ why not give your self & your kid the flu vaccine too ? Ô£ø√º√¥√á why not those ?! Lol	I just hope that all parents who decide to vaccinate their kids.. also decide to give their kid the COVID vaccine. Why pick and choose which vaccines to take now? why not give yourself and your kid the flu vaccine too? why not those ?! Lol

Data Processing

After data cleaning, data processing (See Figure 3) prepares the dataset for analysis in converting the cleaned file into the different R data-storage objects used for

LDA. This data processing and object conversion is fundamental to analysis, as converting from a data frame, to a corpus, to a document-feature matrix, to a document-term matrix facilitates analysis using the quanteda (v.2.1.2; Benoit et al., 2018) and topicmodels (v.0.2.12) packages in R, as well as subsequent visualization using LDAvis (v.0.3.2; Sievert & Shirley, 2015) packages. Imported data from the aforementioned Twitter API search, is converted to a data frame, with the full text of all tweets still intact. These tweets are categorized as string variables, meaning the entire phrase of each tweet is a single unit.

There are several storage methods or objects for data within R, one of the most common and useful being a data frame (Landers, 2018). The tweets collected using the process described above are initially stored in a list (i.e., combination of data types in one structure) (Landers, 2018), which is not always usable for analytic procedures involving machine- learning. Conversion to a data frame, which is a special type of list, allows for easy conversion to the wide range of other data storage options in RStudio, including .csv files and document-feature matrices which will be discussed in the subsequent sections. A random sample of the dataset was visually inspected at this stage to verify correct structure and collection procedures, using the head() and str() functions available in the pre-loaded utilities package within RStudio. Head() gives the first six lines of the selected data object, and str() details the structure of the selected document, in regards to data class, data type, and breakdown of individual variables.

Data Conversion

Data conversion is an essential part of text analysis (Silge & Robinson, 2020), especially when using multiple packages in RStudio. Different packages require different data objects (i.e., structure or format), so understanding how the different formats can work together in crucial. For this study, data processing was a five-step conversion, including: a corpus, a document-feature matrix, a trimmed document-feature matrix, and a document-term matrix (see Figure 3).

A corpus is a collection of texts, represented as character or string variables (Welbers, Van Atteveldt, & Benoit, 2017). Within this study, our corpus is made up of cleaned individual tweets (*N* = 31,925) and the relevant metadata saved during the cleaning process: screen name, favorite count, retweet count, and retweet status (i.e.,retweet or not). Each tweet was also given an identification number, from 1 to 31,925 represented in the corpus as text_1, text_2….text 31,925. This unique identification number helped ensure that the tweets could be accounted for at each stage of data processing. After the corpus was created from the cleaned data frame (see Figure 4) using the corpus() function in the quanteda package (v.2.1.2), the corpus was converted to a document-feature matrix, which converts the string variables (e.g., full tweet as a sentence) to individual words (e.g., tokens), in a process called tokenization (Watanabe & Müller, 2020). Tokenization in combination with another processing technique, lemmatization, are crucial for ensuring a more interpretable model (Jacobi et al., 2016). Lemmatization groups similar words together, usually the singular and plural forms, different tenses of a verb, or synonyms. An example of the lemmatization process used in this study is in Table 4.

Table 3. Lemmatization Example	
Words Prior to Lemmatization	Lemmatized Form
baby, babys, babies, infant, infants	baby
government, govt	government
kid, kids, child, childs, children	child

A document-feature matrix uses a corpus object (from step 1; see Figure 2) to create a sparse matrix in which rows are documents (e.g., tweets); columns are terms (e.g., individual words), and cells represent how many times each term appeared within each document (Benoit et al., 2018; Welbers, Van Atteveldt, & Benoit, 2017). A sparse matrix refers to a matrix that is mostly composed of zeros (Maechler, 2008). As it is unlikely for all tweets to share the same or even most of the same words, the document-feature matrix created from our corpus in step 1 is sparse.

A trimmed document-feature matrix limits the amount of features (e.g., words), using minimum and maximum term and document frequencies calculated with the dfm_trim() function in quanteda (v.2.1.2). The limits for trimmed DFM vary, and this study set a minimum term frequency of 80% and a maximum document frequent of 10%; keeping terms that occurring in at least 80% of the entire corpus, in less than 10% of all the documents. This allows the subsequent analysis to focus on representative, but distinct features (Watanabe & Müller, 2020). A document-term matrix (i.e., DTM) uses the same structure of DFM, but terms can only be one word, whereas in a DFM a feature could be set to more than one word (e.g., first and last names). Converting the trimmed

DFM to the DTM was necessary in order to run the LDA model in the topicmodels (v.0.212) using the LDA() function.

Analyses

Model Parameters

LDA requires parameters to be set prior to analysis, namely α, *β*, and *K.* In LDA, α deals primarily with the distribution of topics within documents; limiting the number of topics a document can contain (Jacobi, van Atteveldt, & Welbers, 2016). *K* refers to the number of topics a model contains, is set apriori (Maier et al., 2018), and then evaluated. The α is typically estimated at 50/*K*, and defaults to this estimation in the topicmodels (v.0.2.12) package (Grün & Hornik, 2021). The *β* is the topic distribution over each word (Maier et al., 2018), and defaults to an estimation of 1/*K* in the topicmodels (v.0.2.12) package (Grün & Hornik, 2021). The method to be used to fit the subsequent model is also specified within the the topicmodels (v.0.2.12) package (Grün & Hornik, 2021), as either variational expectation-maximization (i.e., VEM) or Gibbs sampling technique (Griffiths & Steyvers, 2004) for a Bayesian estimation (Grün & Hornik, 2011).

Model Training and Testing

Machine-learning, and Latent Dirichlet Allocation more specifically, are Bayesian approaches, using a process of training and testing models in order to reach better conclusions (Blei, Ng & Jordan, 2003). Logistically, this requires splitting the dataset into a training sample and testing sample. The two sample groups were randomly assigned to reduce potential biases and misinterpretation. Training the model (e.g., the LDA) on a sample of 90% percent of the data (Maier et al., 2018) allows us to optimize the output on a large portion of the data. Testing the output of the LDA performed on training data (e.g., 90% sample) with data that has been reserved for model testing (e.g., testing data; 10% of overall cleaned sample) allows us to evaluate model fit. In LDA, model fit is evaluated using a measure called perplexity.

Perplexity

Perplexity is a measure of goodness of fit (Blei, Ng & Jordan, 2003); comparable to an R^2 in linear regression (Jacobi, van Atteveldt & Welbers, 2016). More specifically, an R^2 is a metric that determines the percent of variance explained by the predictors variables in a dependent variable [e.g., $R^2 = .85$ indicates 85% of variance is explained by the independent variable(s) with 15% unexplained]. Held-out likelihood refers to the Bayesian foundations of the perplexity measure, as the trained (i.e., fitted) model is used in comparison with data that has been "held-out" (e.g., 10% sample versus 90% sample). Generally, the lower the perplexity score, the better the goodness of fit (Jacobi, van Atteveldt & Welbers, 2016). The lower perplexity score corresponds to a specific K (e.g., number of topics), indicating the optimal number of topics for the model. The LDA is then run again on the full dataset (e.g., testing and training data together), with the value of K set to determine the optimal number of topics.

Model Interpretation

While all of these parameters do involve researcher decisions, model interpretation is place in which the researcher becomes more involved in the process. Topics can be named and further categorized based on the researcher's interpretation of the top terms occurring in each topic, based on the β (e.g., topic probability distribution per word) (Blei, Ng, & Jordan, 2003; Jacobi, van Atteveldt & Welbers, 2016). However, this frequency based approach can make interpretation difficult, as terms can appear across multiple topics (Sievert & Shirley, 2014).

Relevance

To address the limitation of a purely frequency based approach, the relevance metric, reorders the top terms for each topic based on overall corpus frequency (Maier et al., 2018; Sievert & Shirley, 2014). For instance, the keywords used in this study meant that all tweets must include one word in at least each of the two categories: Category (1) child, kid, parent and Category(2) vaccine, covid, corona. By the nature of the sampling strategy, these words occur frequently throughout the entire corpus. Interpreting the topics based only on the top words specified without incorporating overall frequency within the entire corpus, may make interpretation difficult as dissimilar topics appear similar.

Relevance is set using λ as a weighting parameter set between 0 and 1, and optimized at 0.6 (Sievert & Shirley, 2014). When λ is set to 1, the top words reflect the standard probability, while when $\lambda = 0$, the top words are the most specific words to that topic (e.g., occurring less frequently in the rest of the corpus) (Maier et al., 2018; Sievert & Shirley, 2014). The use of the visualization package LDAvis (v.0.3.2) aids in interpretation, not only in the use of the relevance metric to identify top words more specific to each topic, but also in visualizing the distribution of top terms across the entire corpus.

[This page is blank]

CHAPTER FOUR

RESULTS & DISCUSSION

In order to develop recommendations for researchers and practitioners in a coherent fashion, the results and discussion of this study are presented together in an integrated fashion. As an exploratory study with two research questions: (1) What are the conversations and commentary occurring on Twitter about parents, vaccines and COVID-19 and (2) Can machine-learning help us explore this issue of vaccine hesitancy, in the relatively non-traditional context of social media? Thus, an integrated results and discussion section aids in study interpretation and implications. Specifically, the results of the LDA model (perplexity and most relevant terms) are presented, followed by a discussion separated into two parts: recommendations for practice and recommendations for research. Study limitations and overall challenges are discussed as part of challenges with research at the frontier.

This study explored the commentaries and conversations occurring on Twitter about parents, vaccines, and COVID-19 using a method somewhat novel within the leisure and OST sciences: machine-learning. Machine-learning, specifically Latent Dirichlet Allocation, was used to explore a large dataset (n = 31,925 tweets) collected during a key point in the COVID-19 pandemic: federal emergency authorizations of two major vaccines (See Figure 4). 25 latent topics identified by the model were further sorted into seven categories: Government, Feelings, School, Public health, Christmas, Risk & Safety, and Parents & Families for additional interpretability (Table 5). The main challenge with unsupervised machine-learning is interpretability, as the theoretical success of a machine-learning model is not necessarily indicative of applied knowledge or understanding. Put simply, the model can "work" but mean very little when attempting to distinguish interpretable topics.

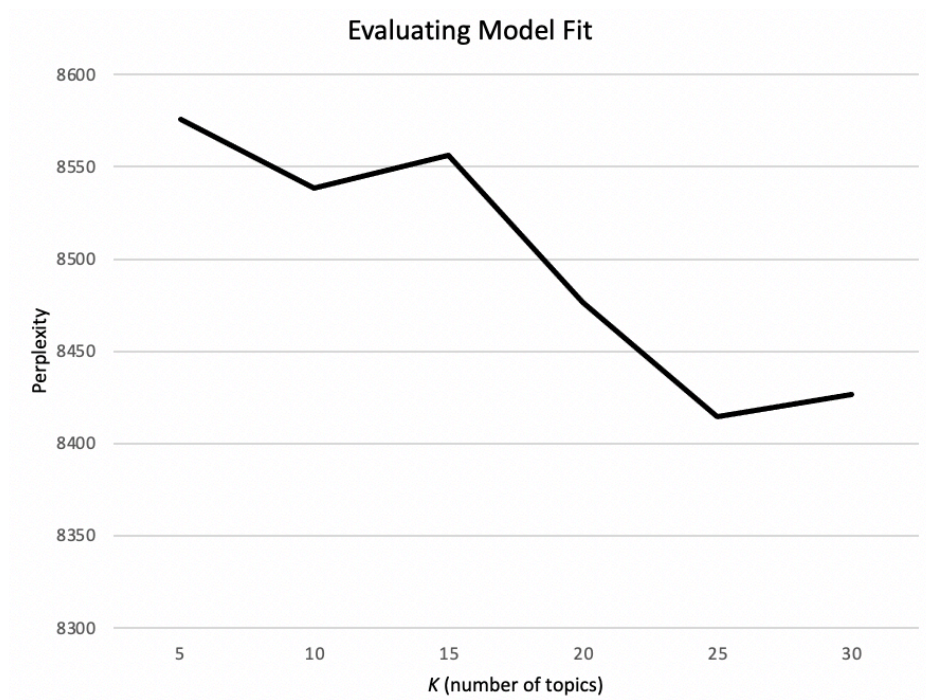
This study was able to explore the commentaries and conversations occurring on Twitter about parenting, vaccines, and COVID-19 by using web-scraped data. The use of a web-scrape as the method of data collection led to a large dataset, which led researchers to incorporate methods of analysis most appropriate for this kind of exploratory research: machine-learning. The specific machine-learning model used, LDA, is a Bayesian approach, using randomly sampled training data to optimize the model, followed by a test of model fit using data randomly reserved for that purpose. Interpretation of a Bayesian model leads to more Bayesian inferences; the more we learned in order to explore results, the more we realized how different a machine-learning model different from other forms of textual analysis.

The use of keywords during data collection meant that every tweet available for use towards further analysis included at least one of each set of keywords (e.g., parent, child, kid; covid, corona, vaccine). While this helped to ensure coherence across the entire dataset, it also resulted in words which were closely aligned to the topic (e.g., parent, child, kid; covid, corona, vaccine) that occurred so frequently throughout the dataset they were not relevant to specific topics for model interpretation. Logistically, these words occurred so frequently across the entire dataset that topics characterized by any of the keywords would not have resulted in interpretable findings. Therefore, the latent topics identified through the LDA may be interpreted as topics which exist under these larger categories assumed by the use of the keywords. More interpretation research regarding LDA using social media data in OST and leisure sciences is needed, as this study served as the beginning, not the end, of machine-learning in these fields.

Perplexity Results

Perplexity was calculated using the perplexity() function in the topicmodels package (v. 0.2-12). Models at $K = 5$, $K = 10$, $K = 15$, $K = 20$, $K = 25$, and $K = 30$ were evaluated, using the test data ($n = 3,193$) that had been randomly assigned and reserved for comparison. Both datasets underwent the same cleaning procedures prior to tokenization and lemmatization, and were subject to the same control methods when creating the data objects needed to perform LDA. In using the testing data ($n = 3,193$) to calculate the perplexity of the three fitted models, we were able to evaluate how well the fitted model is able to generate predictions using new or held-out data (Maier et al., 2018).The lowest perplexity score was at 25 topics (see Figure 5), so $K = 25$ was selected for further analysis using the full dataset (Blei, Ng & Jordan, 2003).

Figure 6. Perplexity Graph



Most Relevant Terms

As anticipated, the LDA model with *K* = 25 resulted in 25 topics. For interpretability, topics were manually grouped together, taking the top relevant terms into account (see Table 4). The LDAvis package, specifically the interactive visualization, assisted in this process as topics were able to be explored beyond a table of the top 5 words (Appendix A). A range of topics were identified, and further explored using the kwic() function(i.e., keyword in context), from the quanteda package (Benoit et al., 2018), and sorted into 7 categories. This approach resulted in identifying representative tweets from the cleaned dataset, containing top relevant terms from the topics identified with the selected LDA model.

Table 4. Topics with most relevant terms

Category	Topic	Most relevant words
Government	1: Relief Needed	care, relief, families, workers, food
	2: Trump	trump, realdonaldtrump, white, man, god
	3: Support Seeking	support, hope, rise, share, economy
	7: Economic Impact	deal, closed, big, time, small
	8: Jobs	work, due, job, time, single
	13: Poverty	government, public, poverty, years, lives
Feelings	4: Mixed Emotions	good, day, feel, bad, make
	19: Positive	great, play, making, real, left
	20: Negative Communication	put, lot, things, talking, poor
	23: Upset	fuck, sick, give, gonna, won
School	6: Teachers & Students	teacher, student, learning, person, part
	24: Masks	year, mask, wear, masks, primary
	9: Abuse	die, abuse, rate, community, number
Public Health	5: Symptoms	positive, tested, case, symptoms, case
	12: Pregnancy	woman, baby, age, pregnant, pfzier
	17: Vaccine History	polio, anti, remember, doctor, disease
	15: Patient Care	live, medical, line, heart, patients
	18: Health Issues	health, life, issues, early, immune
Christmas	14: Christmas Cheer	home, christmas, safe, stay, love
	25: Santa Worries	worry, santa, are, worried, restrictions
Risk & Safety	11: Safety Concerns	safety, important, learn, call, visit
	16: Risk of Spread	risk, young, spread, stop, virus
	21: Long-Term Effects	long, world, social, effects, term
Parents & Family	10: Fathers & Sons	back, dad, son, lost, friend
	22: Mothers	family, mom, flu, court, test

This process is where a social science perspective becomes more valuable than computational technique, in order to take initial model output (e.g., list of topics with top 5 words) and interpret it to answer our research questions. From exploring the visualizations of each topic in this category, and looking at representative tweets, recommendations were developed to address the issues raised from the topics, in two parts: recommendations for practice (e.g., evaluation of conversations and commentary occurring on Twitter about parents, vaccines, and COVID-19) and recommendations for research (e.g., evaluation of machine-learning).

Some users spoke to their concerns about overall health and wellness, from both a maternal health and pediatric perspective (Public Health category; Table 6). Vaccine safety and parenting concerns, specifically thoughts regarding vaccine safety for children, as well as the risks associated with in-person education were evident across topics, concentrated in the Risk & Safety category (Table 7). The influence of the period during which data were collected was also evident, as users expressed concerns related to holiday celebrations, from Santa Claus' visits during a pandemic to lamentation regarding the loss of previous tradition (Christmas category; Table 11). Other concerns related to the difficulties COVID-19 caused families were also present (Parents & Families category; Table 8). The Parent & Families category is an excellent example of one of the important considerations to keep in mind when using an unsupervised machine-learning model like LDA: identical words may not be used the same way across the dataset (Table 8; second tweet).

Recommendations for Practice

As instant communication has become more normative, camp directors and administrators have reported increasing struggles to maintain a balance between customer service and program presence (Henderson, 2007; Kingery et al., 2014). A brief phone call or email from the parent of a first time camper could be expected by an administrator, but daily messages followed by comments on the camp's Facebook photo album may be excessive (Garst, Gagnon & Bennett, 2016). The idea of no news is good news has been phased out, and camp administrators report the consequences of constant contact as communication channels frequently overwhelm their time. Therefore, the interpretation of the School, Public Health, Risk & Safety, and Parent & Families categories was structured around communication recommendations for OST professionals, particularly camp directors gearing up for summer programming during the COVID-19 pandemic,

School

The *School* category includes three of the latent topics identified in the LDA (K = 25), characterized initially by the top relevant words in Table 5. Terms associated with this category ranged from teacher and student, to abuse and masks, reflecting concerns related to those most involved in education (e.g., teachers and students) and the concerns associated with education during a pandemic, specifically mask usage and lack of abuse prevention due to the lack of in-person education.

Table 5. School Category Representative Tweet
@GovInslee Zero kids in OR and WA have died of covid. Death by suicide is 120x more likely to happen to a kid than death by seasonal influenzas. Zero educators in WA have died of covid. Average age of teacher is 40. No one will die! #openourschools
Why would the parents of my mother's student - who felt sick last week - wait FOUR DAYS to tell her (and the school) the kid tested POSITIVE for Covid It feels like the scene in every zombie movie when the bitten person goes "I'm fine, I'm totally not bitten" #StayHomeSaveLives

Education during the pandemic received varied responses, as some focused on the lack of training educators received in the transition to online education (ElSaheli-Elhage, 2021), while others were concerned about students' minimal access to social services and the associated consequences (Lancker & Parolin, 2020). Mental health, suicide rates in particular, was also a continued concern (Reger, Stanley, & Joiner, 2020), as healthcare workers worried about the convergence of conditions all typically associated with higher suicide rates (e.g., growth of unemployment, political turmoil, health crises). Some campers may have been fully in-person all school year, while some were fully online. How will their needs differ at camp in this context, and what can camps do to prepare? OST professionals should allocate time, energy, staff, and funding to additional mental health resources. The American Camp Association, along with the American Academy of Pediatrics, has developed resources for this issue, available on the ACA's COVID-19 resource website.

Public Health

The *Public* category includes five of the latent topics identified in the LDA, characterized initially by the top relevant words in Table 5. Terms associated with this category ranged from "pregnancy" and "women" to "polio" and "doctor," reflective of the diverse range of concerns from parents regarding the COVID-19 pandemic.

Table 6. Public Health Category Representative Tweet
COVID-19: **Pregnant women** allowed partner at birth under new coronavirus rules. This is how sheep like we have become. 'Allowed'? Fuck off! You'd need to fight me to stop me being at the birth of my child!
@savagebabs1 @KareemFoster79 @DanRather do you realize that it usually takes a bit of time for babies to show symptoms of autism after being born? Stop acting like a vaccine causes autism. go talk to people who lived through smallpox or polio. all of these diseases are vaccinated for a reason. protect your child.

As noted previously, the COVID-19 pandemic exacerbated existing health disparities, felt not only by those contracting COVID-19 but by others suffering from a clinic closures and difficulties of telehealth, including pregnant women (Bruno, Shalowitz, & Arora, 2021). Support during labor and delivery (e.g., partner in the room) is associated with better perinatal outcomes (Bruno, Shalowitz, & Arora, 2021), and lack of support due to COVID-19 procedures (e.g., partner not allowed in the delivery room) was not well-received by the public. The initial COVID-19 vaccine trials did not include pregnant women, and fueled concerns that the vaccines were not safe for this population (Farrell, Michie, & Pope, 2020). Health concerns related to COVID-19 and vaccines vary greatly, and it is crucial that OST professionals are equipped with a variety of responses to these concerns. OST professionals should make a plan on how they are going to communicate their new COVID policies and procedures to parents, and then develop responses for their staff to use when talking with parents.

Risk & Safety

The Risk & Safety category includes three of the latent topics identified in the LDA, characterized initially by the top relevant words in Table 5. Terms associated with this category ranged from "safety" and "risk" to "worry" and "virus."

Table 7. Risk & Safety Category Representative Tweet
Nothing like sacrificing your precious child to a vaccine with NO **safety** data for pregnant, breastfeeding mothers or for rapidly growing children. Sure hope he's not harmed
@SkyNews @RachelReevesMP I'm a single parent dad and I would rather b at home with my kids then put them at risk in school which every week u here a new case of Covid. Only parents that seem 2 want to put the kids in school are the 1s that don't want to stop working or don't want 2 b stuck at home wiv them.

This category reiterated concerns in both the Schools and the Parents & Families category, from language reflecting vaccine-hesitancy (Estep & Greenberg, 2020) as well as associated risks in returning to in-person education (ElSaheli-Elhage, 2021). This category is an excellent example of the connectivity between topics in an LDA model, as terms are not mutually exclusive to individual topics. From a risk mitigation perspective, explored in our risk & safety category, OST professionals should consider: What is your program's immunization policy management strategy? Who is checking forms, or attestations? Or do you have a policy to begin with? Policies are not the same as procedures, and the logistics of public health at camp can be very complicated. Policy management is key to public health and safety in OST programs.

Parents & Families

The Parent & Families category includes two of the latent topics identified in the LDA, characterized initially by the top relevant words in Table 5. Terms associated with this category ranged from "family" and "mom" to "son" and "court."

Table 8. Parents & Families Category Representative Tweet
My abuser owes nearly $16k to my children. Stopped paying 1.5 yrs ago...but state retirement he receives still sends his money, but has refused to cooperate with the child support office & **court** order & take CS out of his retirement. Covid cancelled our **court** date in March.
@Canadabuster @JustinTrudeau Yeah cause Justin time-travelled back to August and renegotiated the vaccine deals because Erin criticized him on Twitter three days ago. Did your **mom** drop you on your head as a kid?

The Parent & Families category reflected concerns shared with other studies focused on the effects of the COVID-19 pandemic on families; exacerbating issues associated with single parent homes and the difficulties in work-life balance (Fisher et al., 2020). Spouses planning on divorcing were unable to do so, leaving their families in a holding pattern (Lebow, 2020). Even when court proceedings were able to be held in an online format, the resources required to do so were often lacking and further disrupted the process (Baldwin, Eassey, & Brooke, 2020). Humor was also present, in keeping with studies associating humor as a coping mechanism during the COVID-19 (Bischetti, Canal, & Bambini, 2020), as people sought stimulation from online spaces for their daily interactions (Barnes et al., 2021). From the parents & families category, OST programs should look into how their scholarship funds are currently allocated, in addition to planning a communication strategy when faced with "dark humor", and how staff should address it.

Recommendations for Research

The use of Twitter as a data source complicated the process of training the LDA model, as tweets are not edited by a publishing company like a book or article would be, or accompanied with additional clarifications regarding their meaning. Twitter data is messy, both in its raw form as incomplete sentences with grammatical errors and misspelling, as well as the use of slang and other characteristics specific to social media (e.g., the @ symbol noting a reply to another user, or # followed by words which may or may not relate to the tweet's overall message). While this messiness did result in several stages of data cleaning and data processing (Figure 4), it also indicates the authenticity of the data. Opinions, jokes, complaints, and debates regarding parents, children, vaccines, and COVID-19 all indicate how multidimensional these issues are. Some users focused on the actions or inactions of politicians to curb the pandemic (Government category; Table 6), while others detailed the difficulty of holidays amidst a pandemic (Christmas category; Table 8). While the three categories below (Government, Feelings, and Christmas) did not aid in an exploration of conversations and commentary about parents, vaccines, and COVID-19, they do help illustrate the complexity of a machine-learning approach, specifically data management and interdisciplinary challenges.

Government

The *Government* category includes six of the latent topics identified in the LDA characterized initially by the top relevant words in Table 5. Terms associated with this category ranged from "realdonaldtrump" (i.e., former US president Donald Trump's personal Twitter username) to concerns related to economic relief and job security.

Table 9. Government Category Representative Tweets
It isnt the dems who want YOU to be free an in charge of your own life no that is Trump. IT WASNT THE DEMS WHO SIGNED AN E.O. TO STOP CHILD TRAFFIKING IT WAS TRUMP. It wasnt the dems who wanted to give you a check for covid cuz they held it up but Trump wanted to. Its not the *(tweet ends)*
@FLOTUS @ToysForTotsUSA @USMC @JBABdc Your husband pulled food, housing subsidies. Let COVID run rampant, costing millions their jobs & lets McConnell delay any relief. GOP is the reason there are so many needy children. Just go away.

In the United States, the COVID-19 pandemic continued to strengthen the political, partisan divide (Druckman et al., 2020) as Republican (e.g., GOP) and former President Trump's approval ratings and election support suffered (Warshaw, Vavreck, & Baxter-King, 2020). Regions with more deaths from COVID-19 were less likely to support Republicans in upcoming elections (Warshaw, Vavreck, & Baxter-King), though it is important to note that larger, urban cities are typically more left-leaning. COVID-19 relief and unemployment was also an intensely politicized conversation in the United States, demonstrating the divide between those able to work from home and those unable to do so, which typically reflected higher versus lower education and overall income, respectively (Blustein et al., 2020).

As recreation and leisure scientists exploring a novel method through a vaccine hesitancy lens, this category did not aid in an understanding of conversations and commentary occurring on Twitter about parents, vaccines, and COVID-19. However, the Government category did present issues relevant to the study from a methods perspective. Namely, how do recreation and leisure scientists integrate the politicization of leisure into their study design? More work is needed in order to understand the role of politics and governmental agencies role in digital leisure spaces, and how that may change the nature of the online space.

Feelings

The *Feelings* category includes four of the latent topics identified in the LDA characterized initially by the top relevant words in Table 5. Terms associated with this category ranged from “good” and “great”, to “fuck” and “bad”, indicating the range of emotions associated the cleaned dataset of tweets related to the COVID-19 pandemic, children, and parents.

Table 10. Feelings Category Representative Tweet
I’m not taking no vaccine and neither is my child. **Fuck**[1] these pharmaceutical companies.
@FortyYoung @MarkChangizi My kid had tumor surgery postponed 6+ months — it’s unlikely to be cancer and we’re **good**, but others haven’t been that lucky. Happened all around the world to millions. The people who decided to deny care due to Covid restrictions are genocidal sociopaths.
[1] The term “fuck” was not modified for presentation in text in order to preserve the tweet in it’s original form.

In other studies exploring mental and emotional health during the pandemic, anger was associated with increased dissemination of misinformation (Han, Cha, & Lee, 2020), as individuals faced frustration and resentment towards the long-term effects of

COVID-19 on social and political environments. As such, medical practitioners were urged to monitor mental health of routine patients, both during and post-pandemic (Pffefferbaum & North, 2020). Routine care can suffer during other health crises (e.g., a global pandemic), as hospitals and clinics reallocate resources (Chudasama et al., 2020) to combat the crisis, causing some patients long periods of rescheduling. The challenge of veracity in regard to data management is evident here, as curse words and other incendiary language is particularly visible in this category. The dataset has already been cleaned, trimmed, and processed (see Figure 4), and removing curse words may diminish the authenticity of the cleaned dataset. More work is needed regarding the logistics of cleaning and processing social media data.

Christmas

The Christmas category includes two of the latent topics identified in the LDA, characterized initially by the top relevant words in Table 5. Terms associated with this category ranged from "Christmas" and "santa" to "worry" and "restrictions." The appearance of two latent topics related to Christmas is not surprising when you consider the time of data collection (see Figure 4). This category speaks the most to the concerns of children, evident in the following tweet that contains relevant keywords from the topics in this category. This user speaks about the end of the term (e.g., academic semester) and the concerns of children related to holiday celebrations during the COVID-19 pandemic.

As illustrated in Table 10, the concerns raised by Twitter users in our dataset mirrored both broader research concerns (i.e., Boccia, 2020) regarding restrictions on travel and family gatherings, concerns regarding tourism and travel (Grydehoj et al., 2020), and the need to persist as a family despite the fatigue associated with lockdowns and social isolation (Reicher & Drury, 2021).

Table 11. Christmas Category Representative Tweet
In case you're wondering how the end of term is going, I'm in bits listening to the kids speaking to **Santa** with @NickyAACampbell and @rachelburden on R5. Their questions for **Santa**: When will coronavirus end, and can you give an extra present to children who lost a parent to Covid?
Please God get us to Friday so I can get my kid out of school. The covid anxiety is too much. We were supposed to be going on a massive two week sunny vacation this Christmas. Now just looking forward to staying in and getting to know our new games and puzzles.

Methods-wise, social media data is messy. Even with keywords, we got a lot of other "stuff," from curse words to Christmas wishes. This is why model interpretation is the beginning, not conclusion to LDA (Blei, Ng, & Jordan, 2003), particularly for social scientists. In many studies utilizing LDA for data analysis, model interpretation includes perplexity evaluation and top-terms, concluding in an conversation regarding whether or not the model was able to identify interpretable topics (see Allem et al., 2018; Dahal et al., 2019; Jacobi et al., 2016; Maier et al., 2018). While this is not terribly surprising given the exploratory nature of LDA, it does leave social scientists somewhat unfulfilled. Prior to data collection, we investigated the topic areas surrounding the selected keywords: parenting styles, vaccine-hesitancy, and COVID-19, following a similar process used in traditional experimental design. This prior investigation did inform our interpretation of the LDA, but assigning topics into categories a priori did not serve the data or research question well. More work is needed regarding LDA interpretation, and the implications of such interpretation, within the social sciences.

Research at the frontier: Limitations, challenges, and future directions

As noted previously, interpretation can prove difficult after running an LDA model even after calculating perplexity and establishing an optimal value for *K* (see Figure 5). The other model parameters (α and β) also influence output of the model in terms of probability distribution across topics and terms, and while usually estimated using the default estimations in the topicmodels package (Grün & Hornik, 2011), both parameters can be fixed prior to analysis (Jacobi et al., 2015). While an optimal value for the number of topics (*K*) was established for this study, the calculation of perplexity still involved decision-making by the researchers, to run training models with a range of *K* values to use for the perplexity evaluation. λ, used to as a weighting parameter to aid in interpreting the most relevant terms for each topic, is also a scalable parameter (e.g., 0 to 1 scale). Sievert and Shirely (2015) and others (Maier et al., 2018) optimized λ at 0.6, but interpretation is still possible with a different λ value, with additional language regarding why the value was set lower (e.g., to identify more unique, relevant words) or higher (e.g., to identify words more likely to be shared across the entire dataset).

In addition to parameter estimation challenges, data cleaning and processing resulted in several interesting situations, in which the researchers were the mechanism used to decide what to keep or what to remove. For example, during the model training phase (See Figure 3) several words continued to show up within the top 30 most relevant terms for a topic, but were seemingly nonsense (e.g., “goibibo” and “ik4ea9l4kr”). Instead of taking a more conservative approach and removing the terms from the cleaned dataset, we were able to use both R and the original data saved as a spreadsheet, to trace where these terms came from. All tweets within the dataset were publicly available through the Twitter API, and using the full tweet containing “goibibo” and “ik4ea9l4kr,” we were able to make sense of what appeared to be a misspelling. Goibibo is an Indian airline and hotel reservation website, and “ik4ea9l4kr” corresponds to a specific reservation identification code. Customers used Twitter to communicate with the travel company after COVID-19 cancelled their travel plans. To aid in model interpretation, “ik4ea9l4kr” was removed but “goibibo” was kept and was one of the top 30 terms for Topic 9 (See Appendix A).

In a similar manner, “bong” and “jae” were identified as top terms during model training and were lemmatized to “jaehyun” after deeper investigation. Bong Jae-Hyun is a Korean musician (e.g., K-pop) star who tested positive for COVID-19 in December, which led to an outpouring of support across social media of fans offering wishes for a speedy recovery. Jaehyun’s positive COVID-19 test resulted in his entire music group’s quarantine, and fueled concerns about a COVID-19 cluster in the K-pop industry. While it may seem like a specific situation unable to be applied to a general audience, the reoccurrence of Jaehyun’s name offers a poignant example of how different audiences contextualized the COVID-19 pandemic. While other users expressed worries about school closures or Christmas plans, others exhibited concern after K-pop star they liked contract COVID-19. These concerns about Jaehyun’s wellbeing exemplify famous actors and musicians who tested positive for COVID-19 led to conversations about the pandemic on social media, as the effects of COVID-19 were fault across different sectors and audiences.

Survey research about parenting is also fairly homogenous, with samples generally comprised of white, college-aged female students and/or their white, college-educated mothers, in addition to issues regarding sub-optimal (e.g., small) sample size (Cui et al., 2019a; 2019b). In 2019, Twitter had over 31 million monetizable daily active users in the United States (Twitter Annual Report, 2019). Monetizable Daily Active Users (mDAU) is a metric used by Twitter to more accurately reflect their active users; it represents users who are active daily on the platform that can be shown advertisements. While the dataset used in this study did not total 31 million users, it included over 120,000 tweets (prior to data cleaning procedures; see Figure 3) exhibiting characteristics and content of issues beyond a program or specific location (e.g., an individual program, or state and region).

The COVID-19 pandemic has exacerbated health disparities in the United States, as Black and Hispanic individuals are more than four times as likely to require hospitalizations because of COVID-19 (Callaghan et al., 2020; Wortham et al., 2020). Racial discrimination, governmental distrust, and lack of culturally appropriate resources and medical providers are all factors of vaccine hesitancy (Ford & Airhihenbuwa, 2010; Quinn et al., 2017). Hinman & McKinley (2015) admonished that immunizations could be fundamental in establishing health equity, but vaccine hesitancy continues to grow (Estep & Greenberg, 2020; Lu et al., 2015). In using social media data, we have the opportunity to incorporate communities and contexts typically under-represented in out-of-school time research.

Conclusion

Addressing twenty-first century issues requires twenty-first century skills. Research on emergent issues needs not only topical expertise, but additional competencies in communication, computer science, digital technologies, and cultural studies. Emergent issues may span multiple disciplines, lived experiences, and environments, and a machine-learning approach helps to continue providing research that serves our communities best in a changing landscape. Transdisciplinary research, or research that combines knowledge from multiple sources, sectors, and experiences (Wada et al., 2021) embodies both the successes and shortcomings of this study. Machine-learning offers social scientists a critical capacity to explore concerns and commentaries occurring on social media, web-based platforms, large datasets, and more. A machine-learning approach affects not only data analysis but study design and development, as researchers utilize testing and training data to better infer results indicative of the problem in its entirety. As with any worthwhile research study, we are left with more questions than answers, and we look forward to exploring these questions further through a machine-learning approach to transdisciplinary research.

APPENDICES

Appendix A

Selected LDA Visualizations

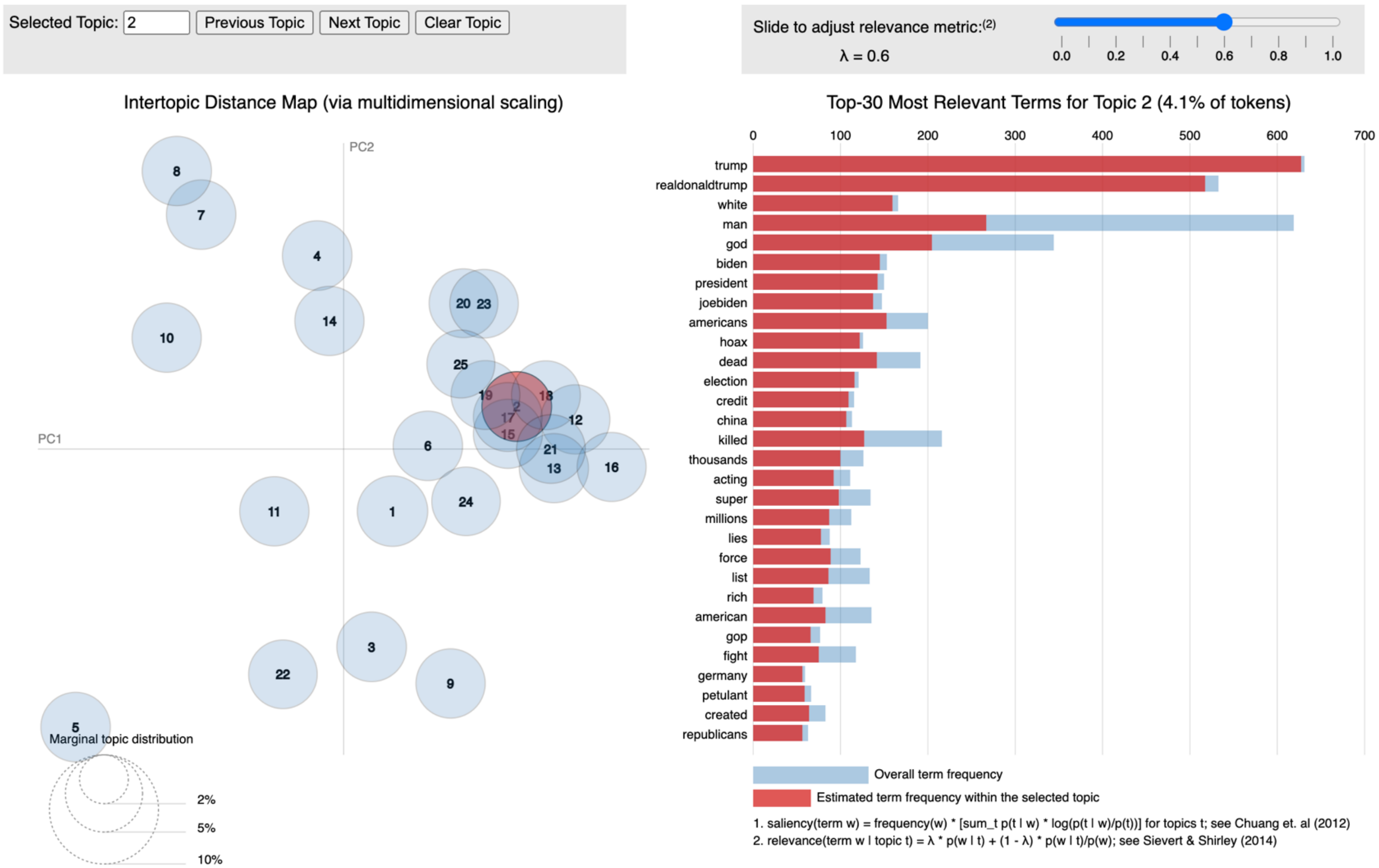


Figure A-1: LDA model visualization with topic 2 (in Government Category) selected

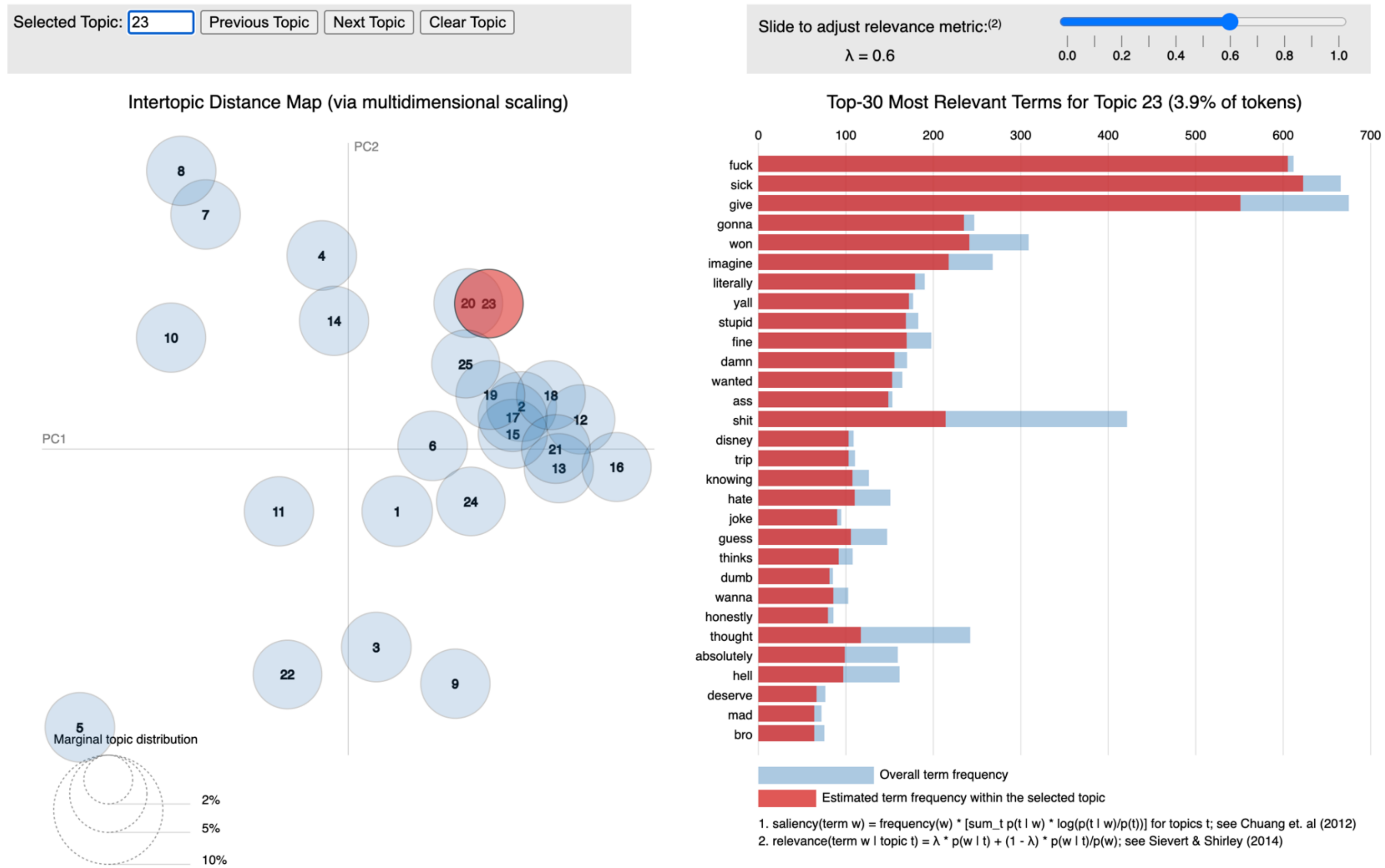


Figure A-2: LDA model visualization with topic 23 (in Feelings Category) selected

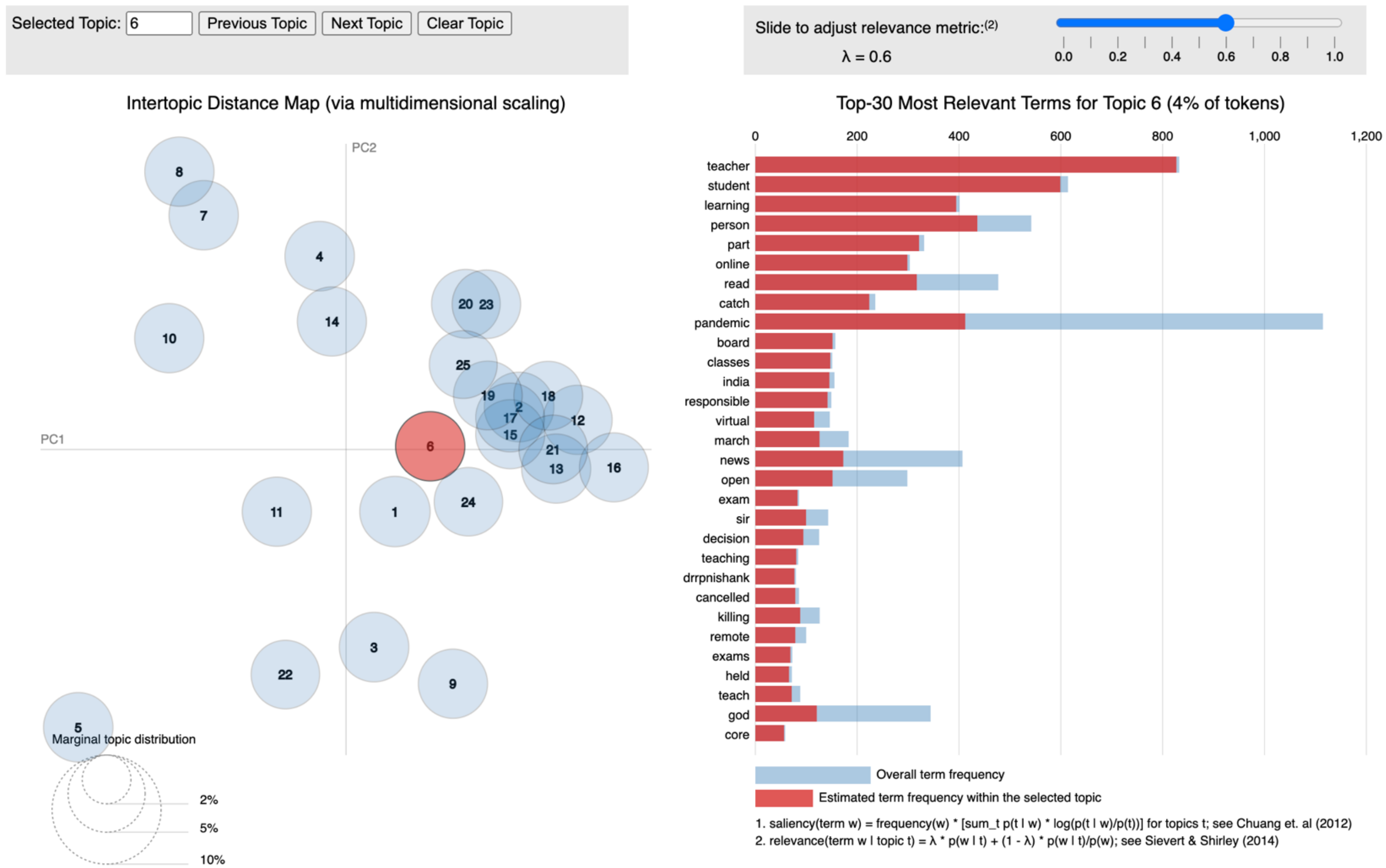


Figure A-3: LDA model visualization with topic 6 (in School Category) selected

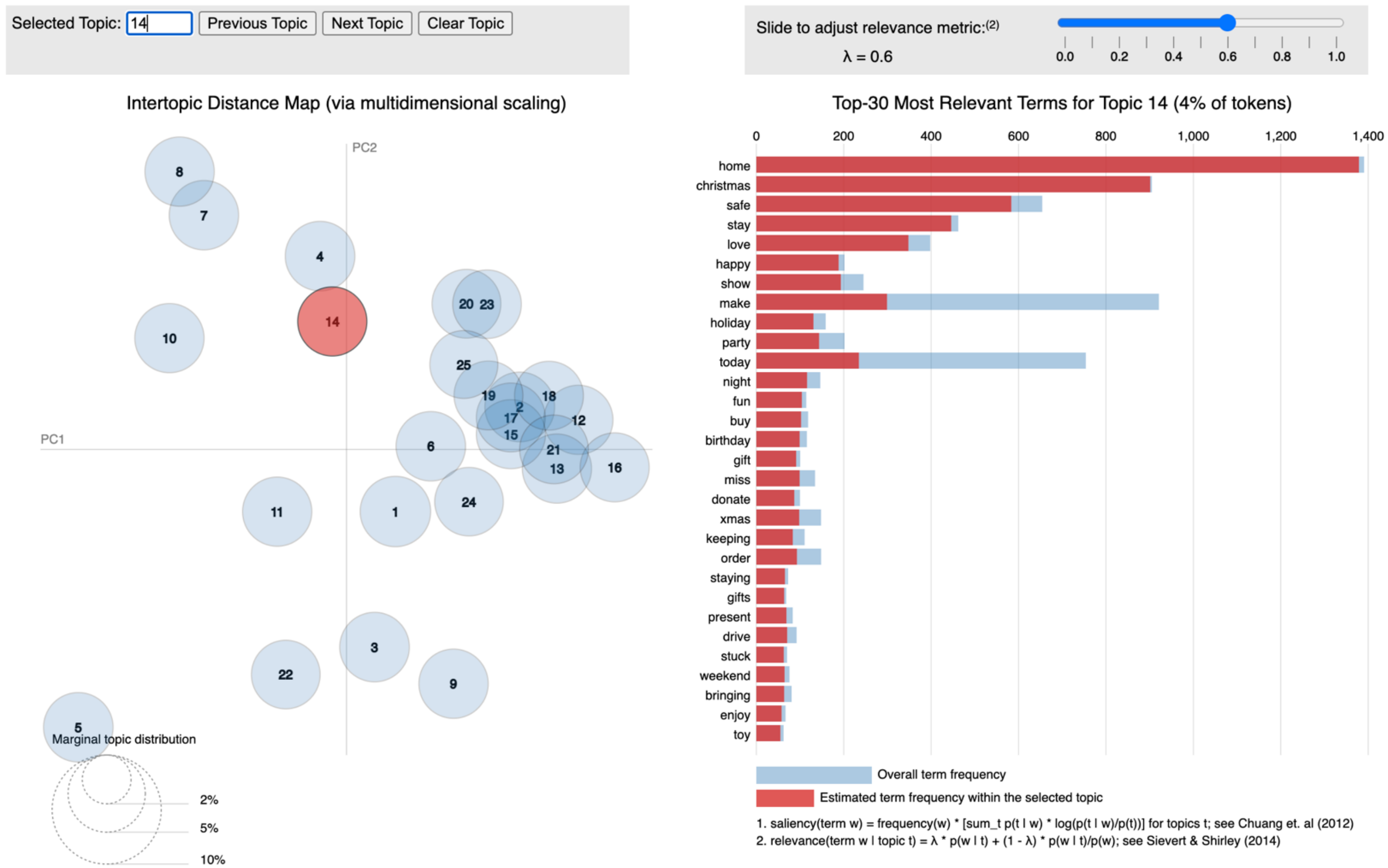


Figure A-4: LDA model visualization with topic 14 (in Christmas Category) selected

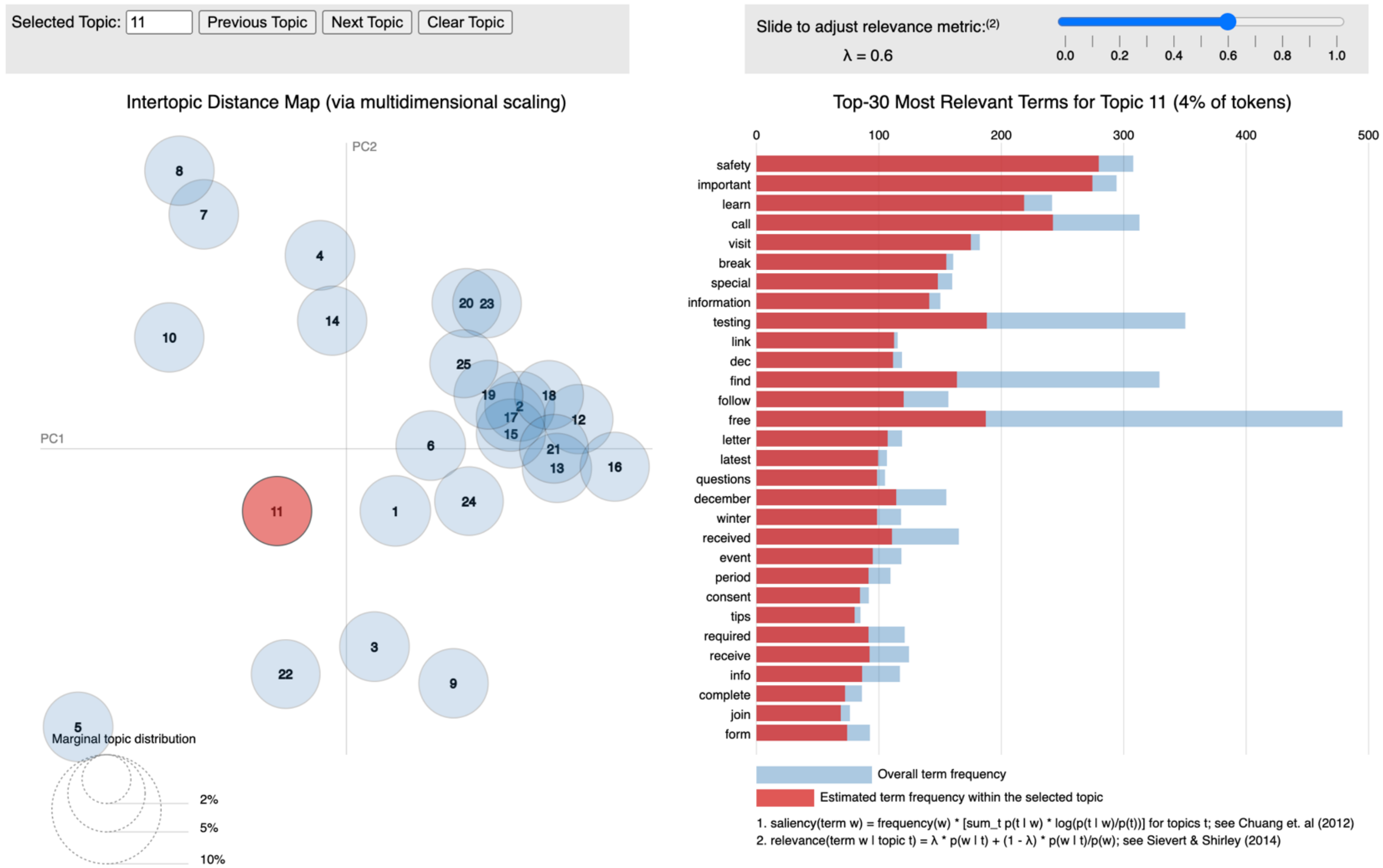


Figure A-5: LDA model visualization with topic 11 (in Risk & Safety Category) selected

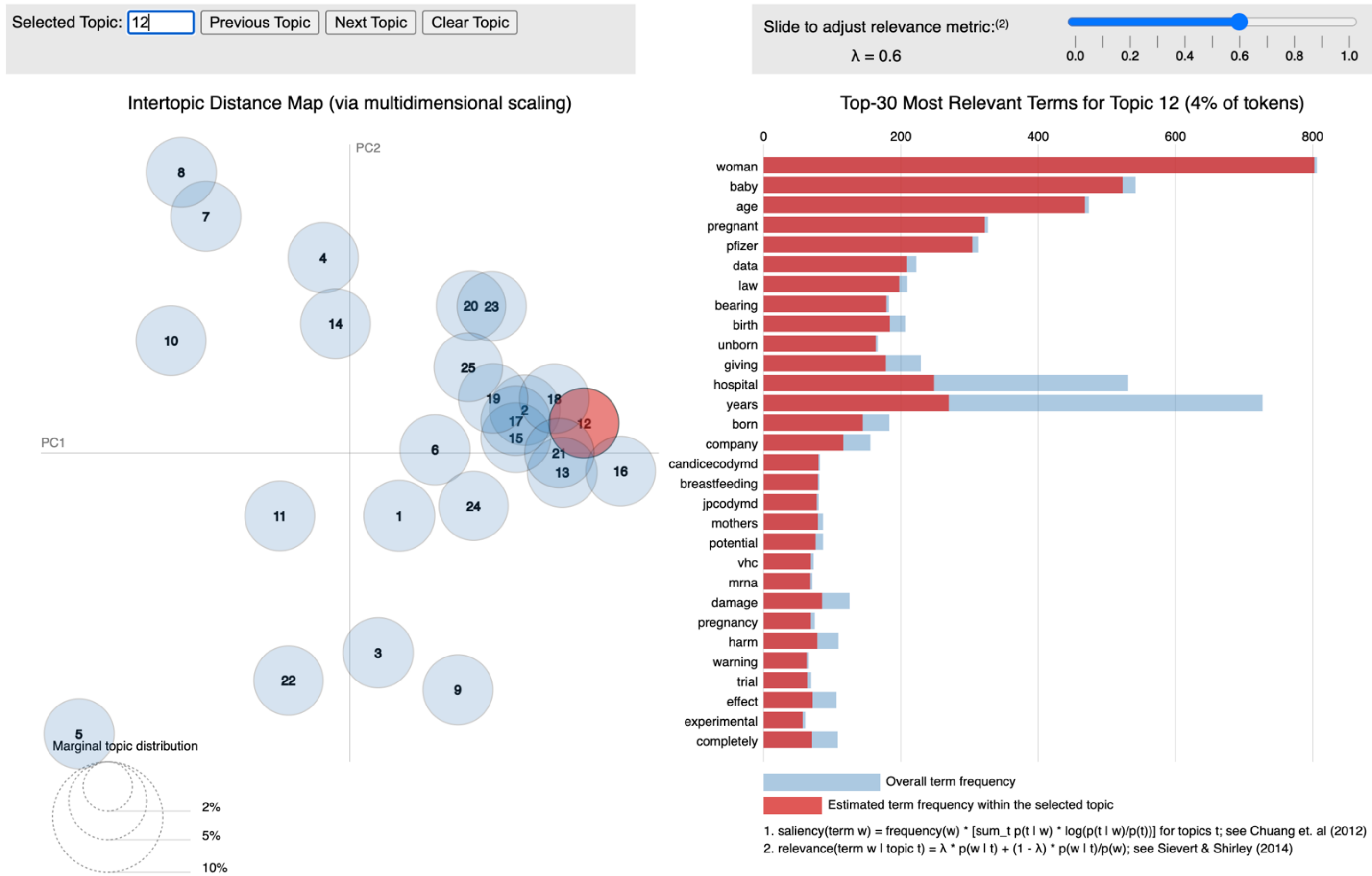


Figure A-6: LDA model visualization with topic 12 (in Public Health Category) selected

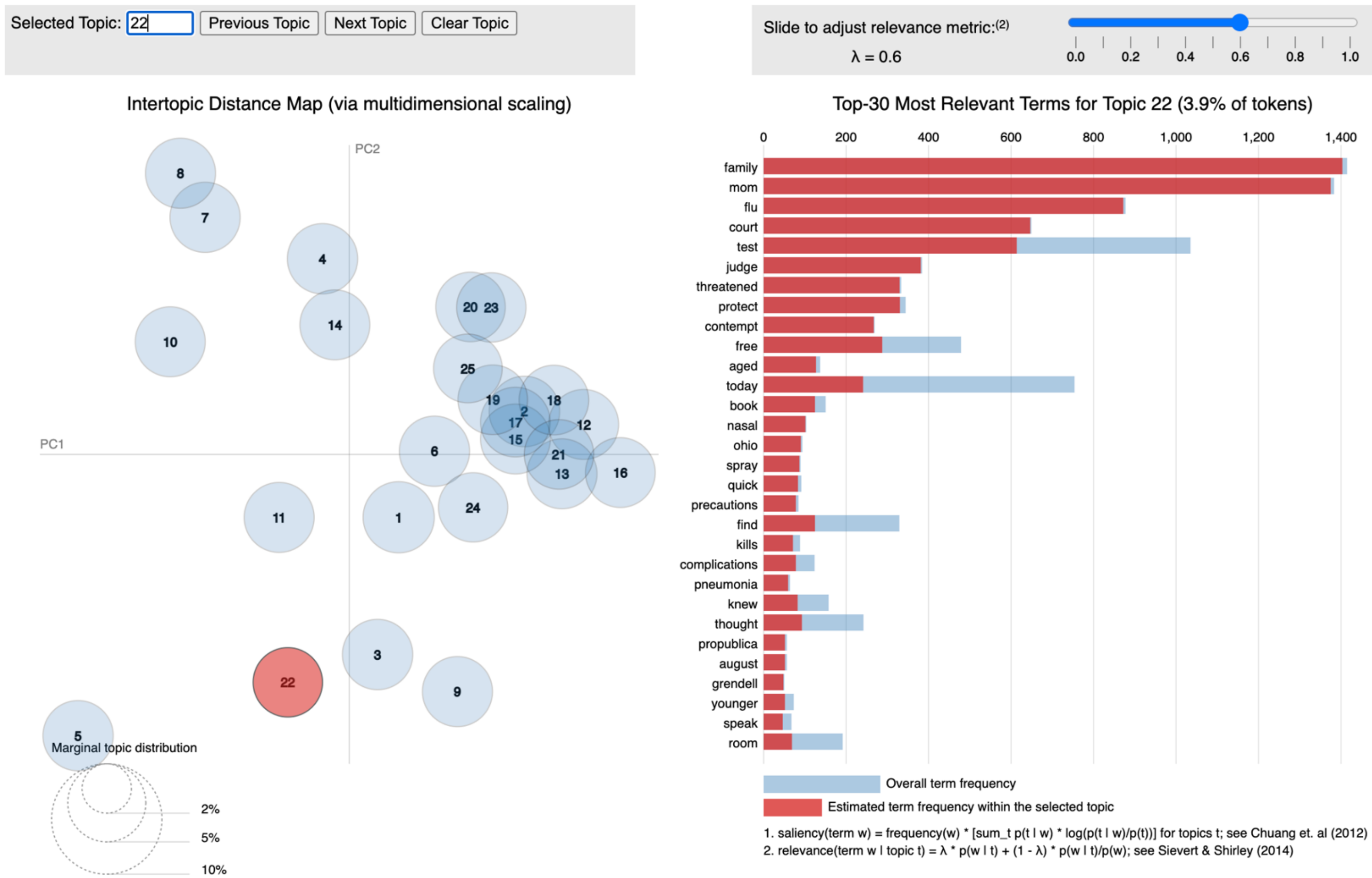


Figure A-7: LDA model visualization with topic 22 (in Parents & Families Category) selected

REFERENCES

Ambrose, M. J., Walton, E. A., & Health, C. on S. (2019). Improving Health and Safety at Camp. *Pediatrics*. https://doi.org/10.1542/peds.2019-1355

American Camp Association (2018). ACA Facts and Trends. Retrieved from https://www.acacamps.org/press-room/aca-facts-trends

American Camp Association. (2017). Emerging Issues Report. Retrieved from https://www.acacamps.org/resource-library/research/emerging-issues

Amin, A. B., Bednarczyk, R. A., Ray, C. E., Melchiori, K. J., Graham, J., Huntsinger, J. R., & Omer, S. B. (2017). Association of moral values with vaccine hesitancy. *Nature Human Behaviour*, *1*(12), 873–880. https://doi.org/10.1038/s41562-017-0256-5

Attwell, K., & Smith, D. T. (2017). Parenting as politics: Social identity theory and vaccine hesitant communities. *International Journal of Health Governance*, *22*(3), 183–198. https://doi.org/10.1108/IJHG-03-2017-0008

Attwell, K., Smith, D. T., & Ward, P. R. (2018). 'The Unhealthy Other': How vaccine rejecting parents construct the vaccinating mainstream. *Vaccine*, *36*(12), 1621–1626. https://doi.org/10.1016/j.vaccine.2018.01.076

Baldwin, J. M., Eassey, J. M., & Brooke, E. J. (2020). Court Operations during the COVID-19 Pandemic. *American Journal of Criminal Justice*, *45*(4), 743–758. https://doi.org/10.1007/s12103-020-09553-1

Barnes, K., Riesenmy, T., Trinh, M. D., Lleshi, E., Balogh, N., & Molontay, R. (2021). Dank or not? Analyzing and predicting the popularity of memes on Reddit. *Applied Network Science*, *6*(1), 1–24. https://doi.org/10.1007/s41109-021-00358-7

Baumrind, D. (1966). Effects of Authoritative Parental Control on Child Behavior. *Child Development*, *37*(4), 887–907. https://doi.org/10.2307/1126611

Behrend, T., & Landers, R. (2019). Introduction to the Special Issue on Advanced Technologies in Assessment: A Science-Practice Concern. *Personnel Assessment and Decisions*, *5*. https://doi.org/10.25035/pad.2019.03.001

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A (2018). "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software*, **3**(30), 774. doi: 10.21105/joss.00774, https://quanteda.io.

Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. *Patterns*, *2*(2), 100205. https://doi.org/10.1016/j.patter.2021.100205

Bischetti, L., Canal, P., & Bambini, V. (2021). Funny but aversive: A large-scale survey of the emotional response to Covid-19 humor in the Italian population during the lockdown. *Lingua*, *249*, 102963. https://doi.org/10.1016/j.lingua.2020.102963

Blaisdell L., Cohn W., Pavell J., Rubin, D., Vergales, J. (2020). Preventing and mitigating SARS-CoV-2 transmission — four overnight camps, Maine, June–August 2020. *Morbidity and Mortality Weekly Report, 69*(35), 1216–1220. http://dx.doi.org/10.15585/mmwr.mm6935e1ex

Blustein, D. L., Duffy, R., Ferreira, J. A., Cohen-Scali, V., Cinamon, R. G., & Allan, B. A. (2020). Unemployment in the time of COVID-19: A research agenda. *Journal of Vocational Behavior*, *119*, 103436. https://doi.org/10.1016/j.jvb.2020.103436

Boccia, S. (2021). Christmas Festivities and COVID-19: A Foreseeable Risk to Anticipate. *Frontiers in Public Health*, *8*. https://doi.org/10.3389/fpubh.2020.639647

Broomhead, K. E. (2019). The influence of the nature of children's disabilities on societal reactions experienced by their parents. *Journal of Research in Special Educational Needs, 19*(2), 126–134. https://doi.org/10.1111/1471-3802.12435

Bruno, B., Shalowitz, D. I., & Arora, K. S. (2021). Ethical challenges for women's healthcare highlighted by the COVID-19 pandemic. *Journal of Medical Ethics*, *47*(2), 69–72. https://doi.org/10.1136/medethics-2020-106646

Burger, S. V. (2018). *Introduction to Machine Learning with R: Rigorous Mathematical Analysis*. O'Reilly Media, Inc.

Burke, T. J., Segrin, C., & Farris, K. L. (2018). Young adult and parent perceptions of facilitation: Associations with overparenting, family functioning, and student adjustment. *Journal of Family Communication, 18*(3), 233–247. https://doi.org/10.1080/15267431.2018.1467913

Capurro, G., Greenberg, J., Dubé, E., & Driedger, M. (2018). Measles, Moral Regulation and the Social Construction of Risk: Media Narratives of "Anti-Vaxxers" and the 2015 Disneyland Outbreak. *Canadian Journal of Sociology*, *43*(1), 25–48. https://doi.org/10.29173/cjs29301

Carney-Hall, K.C. (2008). Understanding current trends in family involvement. *New Directions for Student Services,* 2008(122):3-14. https://doi.org/10.1002/ss.271

CDC. (2021, March 3). *How CDC Is Making COVID-19 Vaccine Recommendations*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations-process.html

Chandra L. Ford, Collins O. Airhihenbuwa (2010). Critical Race Theory, Race Equity, and Public Health: Toward Antiracism Praxis. *American Journal of Public Health* 100, no. S1 pp. S30-S35.

Chudasama, Y. V., Gillies, C. L., Zaccardi, F., Coles, B., Davies, M. J., Seidu, S., &

Clarke, K., Cooper, P., & Creswell, C. (2013). The parental overprotection scale: Associations with child and parental anxiety. *Journal of Affective Disorders, 151*(2), 618–624. https://doi.org/10.1016/j.jad.2013.07.007

Commissioner, Office of Food and Drug Administration. (2021a). Moderna COVID-19 Vaccine. *FDA*. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/moderna-covid-19-vaccine

Commissioner, Office of Food and Drug Administration. (2021b). Pfizer-BioNTech COVID-19 Vaccine. *FDA*. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/pfizer-biontech-covid-19-vaccine

Creswell, J. W., & Poth, C. N. (2016). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). Sage.

Cui, M., Allen, J. W., Fincham, F. D., May, R. W., & Love, H. (2019a). Helicopter parenting, self-regulatory processes, and alcohol use among female college students. *Journal of Adult Development, 26*(2), 97–104. https://doi.org/10.1007/s10804-018-9301-5
Cui, M., Darling, C. A., Coccia, C., Fincham, F. D., & May, R. W. (2019b). Indulgent parenting, helicopter parenting, and well-being of parents and emerging adults. *Journal of Child and Family Studies, 28*(3), 860–871. https://doi.org/10.1007/s10826-018-01314-3
Davies, H. D., Jackson, M. A., Rice, S. G., Diseases, C. on I., & Fitness, C. on S. M. A. (2017). Infectious Diseases Associated With Organized Sports and Outbreak Control. *Pediatrics*, *140*(4). https://doi.org/10.1542/peds.2017-2477
DeCarlo, Matthew (2018). Scientific inquiry in social work. Open Social Work Education. https://scientificinquiryinsocialwork.pressbooks.com/
Dinet, J., Favart, M., & Passerault, J.-M. (2004). Searching for information in an online public access catalogue (OPAC): The impacts of information search expertise on the use of Boolean operators. *Journal of Computer Assisted Learning*, *20*(5), 338–346. https://doi.org/10.1111/j.1365-2729.2004.00093.x
Dooling, K. (2021). The Advisory Committee on Immunization Practices' Updated Interim Recommendation for Allocation of COVID-19 Vaccine—United States, December 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*. https://doi.org/10.15585/mmwr.mm695152e2
Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (undefined/ed). How Affective Polarization Shapes Americans' Political Beliefs: A Study of Response to the COVID-19 Pandemic. *Journal of Experimental Political Science*, 1–12. https://doi.org/10.1017/XPS.2020.28
Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. A. (2013). Vaccine hesitancy. *Human Vaccines & Immunotherapeutics*, *9*(8), 1763–1773. https://doi.org/10.4161/hv.24657
Environmental Health & Engineering, Inc. (2020). *Field guide for camps on implementation of CDC guidance*. Newtown, MA: USA (EH&E Project #240206). Retrieved from: https://www.acacamps.org/resource-library/coronavirus/camp-business/field-guide-camps
ElSaheli-Elhage, R. (2021). Access to Students and Parents and Levels of Preparedness of Educators during the COVID-19 Emergency Transition to e-Learning. *International Journal on Studies in Education*, *3*(2), 61–69. https://doi.org/10.46328/ijonse.35
Farrell, R., Michie, M., & Pope, R. (2020). Pregnant Women in Trials of Covid-19: A Critical Time to Consider Ethical Frameworks of Inclusion in Clinical Trials. *Ethics & Human Research*, *42*(4), 17–23. https://doi.org/10.1002/eahr.500060
Filipo Sharevski, Peter Jachim, and Kevin Florek. 2020. To Tweet or Not to Tweet: Covertly Manipulating a Twitter Debate on Vaccines Using Malware Induced Misperceptions. In The 15th International Conference on Availability, Reliability and Security (ARES 2020), August 25–28, 2020, Virtual Event, Ireland. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3407023. 340702

Fisher, J., Languilaire, J.-C., Lawthom, R., Nieuwenhuis, R., Petts, R. J., Runswick-Cole, K., & Yerkes, M. A. (2020). Community, work, and family in times of COVID-19. *Community, Work & Family*, *23*(3), 247–252. https://doi.org/10.1080/13668803.2020.1756568

Flynn, R. M., Ricker, A. A., Dolezal, C., Kunin, M., & Mellins, C. A. (2019). Residential summer camp for youth with special needs: A longitudinal approach to investigating differences in social skills. *Children and Youth Services Review, 96*, 354–363. https://doi.org/10.1016/j.childyouth.2018.10.036

Ford, C. L., & Airhihenbuwa, C. O. (2010). Critical Race Theory, Race Equity, and Public Health: Toward Antiracism Praxis. *American Journal of Public Health*, *100*(Suppl 1), S30–S35. https://doi.org/10.2105/AJPH.2009.171058

Francis, J., & Francis, L. (2020). Immunization and participation in amateur youth sports. *Journal of the Philosophy of Sport*, *47*(2), 151–167. https://doi.org/10.1080/00948705.2020.1750960

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

Gagnon, R. J., & Garst, B. A. (2019). Exploring overparenting in summer camp: Adapting, developing, and implementing a measure. *Annals of Leisure Research, 22*(2), 161–182. https://doi.org/10.1080/11745398.2018.1452619

Gagnon, R. J., Garst, B. A., Kouros, C. D., Schiffrin, H. H., & Cui, M. (2020). When overparenting is normal parenting: Examining child disability and overparenting in early adolescence. *Journal of Child and Family Studies, 29*(2), 413–425. https://doi.org/10.1007/s10826-019-01623-1

Garst, B., Dubin, A., Bunke, C., Schellpfeffer, N., Gaslin, T., Ambrose, M., & Hashikawa, A. (2021a). Barriers impacting organizational immunization policy implementation in U.S. and Canadian summer camps. *Children's Health Care*, *0*(0), 1–13. https://doi.org/10.1080/02739615.2020.1870118

Garst, B., Gagnon, R., & Bennett, T. (2016). Parent anxiety causes and consequences: perspectives from camp program providers. *The Cyber Journal of Applied Leisure and Recreation Research, 18*(1). https://tigerprints.clemson.edu/parksrec_pubs/34

Gau, S. S.-F., Chiu, Y.-N., Soong, W.-T., & Lee, M.-B. (2008). Parental characteristics, parenting style, and behavioral problems among Chinese children with down syndrome, their siblings and controls in Taiwan. *Journal of the Formosan Medical Association, 107*(9), 693–703. https://doi.org/10.1016/S0929-6646(08)60114-X

Gee, J. (2021). First Month of COVID-19 Vaccine Safety Monitoring—United States, December 14, 2020–January 13, 2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*. https://doi.org/10.15585/mmwr.mm7008e3

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*(2), 105–112. https://doi.org/10.1016/j.nedt.2003.10.001

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(Supplement 1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Grün B, Hornik K (2011). “topicmodels: An R Package for Fitting Topic Models.” *Journal of Statistical Software*, **40**(13), 1–30. doi: 10.18637/jss.v040.i13.

Grydehøj, A., Kelman, I., & Su, P. (2020). Island Geographies of Separation and Cohesion: The Coronavirus (COVID-19) Pandemic and the Geopolitics of Kalaallit Nunaat (Greenland). *Tijdschrift Voor Economische En Sociale Geografie*, *111*(3), 288–301. https://doi.org/10.1111/tesg.12423

Gunaratne, K., Coomes, E. A., & Haghbayan, H. (2019). Temporal trends in anti-vaccine discourse on Twitter. *Vaccine*, *37*(35), 4867–4871. https://doi.org/10.1016/j.vaccine.2019.06.086

Han, J., Cha, M., & Lee, W. (2020). Anger contributes to the spread of COVID-19 misinformation. *Harvard Kennedy School Misinformation Review*, *1*(3). https://doi.org/10.37016/mr-2020-39

Henderson, K. A., Whitaker, L. S., Bialeschki, M. D., Scanlin, M. M., & Thurber, C. (2007). Summer camp experiences: Parental perceptions of youth development outcomes. *Journal of Family Issues*, *28*(8), 987–1007. https://doi.org/10.1177/0192513X07301428

Hinman, A. R., & McKinlay, M. A. (2015). Immunization Equity. *American Journal of Preventive Medicine*, *49*(6), S399–S405. https://doi.org/10.1016/j.amepre.2015.04.018

Horton, R. (2004). The lessons of MMR. *The Lancet*, *363*(9411), 747–749. https://doi.org/10.1016/S0140-6736(04)15714-0

Huber, B., Highfield, K., & Kaufman, J. (2018). Detailing the digital experience: Parent reports of children’s media use in the home learning environment. *British Journal of Educational Technology*, *49*(5), 821–833. https://doi.org/10.1111/bjet.12667

Jacobi, C., Atteveldt, W. H. van, & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *4*(1), 89–106. https://doi.org/10.1080/21670811.2015.1093271

Jenkins, M. C., & Moreno, M. A. (2020). Vaccination Discussion among Parents on Social Media: A Content Analysis of Comments on Parenting Blogs. *Journal of Health Communication*, *25*(3), 232–242. https://doi.org/10.1080/10810730.2020.1737761

Karamshuk, D., Shaw, F., Brownlie, J., & Sastry, N. (2017). Bridging big data and qualitative methods in the social sciences: A case study of Twitter responses to high profile deaths by suicide. *Online Social Networks and Media*, *1*, 33–43. https://doi.org/10.1016/j.osnem.2017.01.002

Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm—An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, *30*(25), 3778–3789. https://doi.org/10.1016/j.vaccine.2011.11.112

Kearney, M. W. (2019). rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software*, *4*(42), 1829. https://doi.org/10.21105/joss.01829

Keller, J. (2008). On the development of regulatory focus: The role of parenting styles. *European Journal of Social Psychology*, *38*(2), 354–364. https://doi.org/10.1002/ejsp.460

Khunti, K. (2020). Impact of COVID-19 on routine care for chronic diseases: A global survey of views from healthcare professionals. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(5), 965–967. https://doi.org/10.1016/j.dsx.2020.06.042

Kingery, J., Peneston, K., Rice, S., & Wormuth, B. (2012a). Parental anxious expectations and child anxiety predicting homesickness during overnight summer camp. *Journal of Outdoor Recreation, Education, and Leadership*, *4*(3), 172–184. https://doi.org/10.7768/1948-5123.1116

Kolodziej, S. P., Aznaveh, M., Bullock, M., David, J., Davis, T. A., Henderson, M., Hu, Y., & Sandstrom, R. (2019). The SuiteSparse Matrix Collection Website Interface. *Journal of Open Source Software*, *4*(35), 1244. https://doi.org/10.21105/joss.01244

Kondracki, N. L., Wellman, N. S., & Amundson, D. R. (2002). Content Analysis: Review of Methods and Their Applications in Nutrition Education. *Journal of Nutrition Education and Behavior, 34*(4), 224–230. https://doi.org/10.1016/S1499-4046(06)60097-3

Lancker, W. V., & Parolin, Z. (2020). COVID-19, school closures, and child poverty: A social crisis in the making. *The Lancet Public Health*, *5*(5), e243–e244. https://doi.org/10.1016/S2468-2667(20)30084-0

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods, 21*(4), 475–492. https://doi.org/10.1037/met0000081

Landers, R. (2018, January 3). Data Types and Basic Variable Manipulation [PowerPoint slides]. http://datascience.tntlab.org/wp-content/uploads/2017/12/Module-2-Data-Types-and-Basic-Variable-Manipulation.pdf

Lantz, B. (2019). *Machine Learning with R: Expert techniques for predictive modeling, 3rd Edition*. Packt Publishing Ltd.

Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M. D., & Paterson, P. (2014). Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012. *Vaccine*, *32*(19), 2150–2159. https://doi.org/10.1016/j.vaccine.2014.01.081

Leask, J., Willaby, H. W., & Kaufman, J. (2014). The big picture in addressing vaccine hesitancy. *Human Vaccines & Immunotherapeutics*, *10*(9), 2600–2602. https://doi.org/10.4161/hv.29725

Lebow, J.L. (2020), The Challenges of COVID-19 for Divorcing and Post-divorce Families. *Family Process*, 59: 967-973. https://doi.org/10.1111/famp.12574

Locke, J., Campbell, M., & Kavanagh, D. (2012). Can a Parent Do Too Much for Their Child? An Examination By Parenting Professionals of the Concept of Overparenting. Australian Journal of Guidance and Counselling, 22(2), 249-265. doi:10.1017/jgc.2012.29

Locke, Judith, Kavanagh, David, & Campbell, Marilyn (2016) Overparenting and homework: The student's task, but everyone's responsibility. *Journal of Psychologists and Counsellors in Schools*, *26*(1), pp. 1-15.

Lombard, M., Snyder-Duch, J., & Bracken, C.C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587-604.

Love, H., Cui, M., Hong, P. & McWey, L. (2020). Parent and child perceptions of indulgent parenting and female emerging adults' depressive symptoms. *Journal of Family Studies*, DOI: 10.1080/13229400.2020.1794932

Luebbe, A., Mancini, K, Kiel, E., Spangler, B., Semlak, J., & Fussner, L. (2018). Dimensionality of helicopter parenting and relations to emotional, decision-making, and academic functioning in emerging adults. *Assessment, 25*(7), 841–857. https://doi.org/10.1177/1073191116665907

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

McKeever, B. W., McKeever, R., Holton, A. E., & Li, J.-Y. (2016a). Silent Majority: Childhood Vaccinations and Antecedents to Communicative Action. *Mass Communication and Society*, *19*(4), 476–498. https://doi.org/10.1080/15205436.2016.1148172

Measles Outbreak—California, December 2014–February 2015. (n.d.). Retrieved March 20, 2021, from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6406a5.htm

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A Survey on Bias and Fairness in Machine Learning. *ArXiv:1908.09635 [Cs]*. http://arxiv.org/abs/1908.09635

Mertan, E., Croucher, L., Shafran, R., & Bennett, S. D. (2021). An investigation of the information provided to the parents of young people with mental health needs on an internet forum. *Internet Interventions*, *23*, 100353. https://doi.org/10.1016/j.invent.2020.100353

Mitchell, T. M. (1997). Does Machine Learning Really Work?. AI Magazine, 18(3), 11. https://doi.org/10.1609/aimag.v18i3.1303

Morsbach, S. K., & Prinz, R. J. (2006). Understanding and improving the validity of self-report of parenting. *Clinical Child and Family Psychology Review*, *9*(1), 1–21. https://doi.org/10.1007/s10567-006-0001-5

Mowatt, R. A., & Schmalz, D. L. (2014). The Conspicuous Nature of Power: Conclusion to the Special Issue. *Journal of Leisure Research*, *46*(3), 353–358. http://dx.doi.org/10.1080/00222216.2014.11950331

Napolitano, F., D'Alessandro, A., & Angelillo, I. F. (2018). Investigating Italian parents' vaccine hesitancy: A cross-sectional survey. *Human Vaccines & Immunotherapeutics*, *14*(7), 1558–1565. https://doi.org/10.1080/21645515.2018.1463943

Oliver, S. E. (2021). The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Janssen COVID-19 Vaccine—United States, February 2021. *MMWR. Morbidity and Mortality Weekly Report*, *70*. https://doi.org/10.15585/mmwr.mm7009e4

Park, E., Kim, H., & Steinhoff, A. (2016). Health-Related Internet Use by Informal Caregivers of Children and Adolescents: An Integrative Literature Review. *Journal of Medical Internet Research*, *18*(3), e4124. https://doi.org/10.2196/jmir.4124

Pfefferbaum, B., & North, C. S. (2020). Mental Health and the Covid-19 Pandemic. *New England Journal of Medicine*, *383*(6), 510–512. https://doi.org/10.1056/NEJMp2008017

Pretorius, K., Choi, E., Kang, S., & Mackert, M. (2020). Sudden Infant Death Syndrome on Facebook: Qualitative Descriptive Content Analysis to Guide Prevention Efforts. *Journal of Medical Internet Research*, *22*(7), e18474. https://doi.org/10.2196/18474

Puri, N., Coomes, E. A., Haghbayan, H., & Gunaratne, K. (2020). Social media and vaccine hesitancy: New updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics*, *16*(11), 2586–2593. https://doi.org/10.1080/21645515.2020.1780846

Quinn, S. C., Jamison, A. M., An, J., Hancock, G. R., & Freimuth, V. S. (2019). Measuring vaccine hesitancy, confidence, trust and flu vaccine uptake: Results of a national survey of White and African American adults. *Vaccine*, *37*(9), 1168–1173. https://doi.org/10.1016/j.vaccine.2019.01.033

R Core Team (2021). R: A language and environment for statistical computing (Version 4.0.4) [Software]. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Reger MA, Stanley IH, Joiner TE. Suicide Mortality and Coronavirus Disease 2019—A Perfect Storm? *JAMA Psychiatry.* 2020;77(11):1093–1094. doi:10.1001/jamapsychiatry.2020.1060

Reicher, S., & Drury, J. (2021). Pandemic fatigue? How adherence to covid-19 regulations has been misrepresented and why it matters. *BMJ*, *372*, n137. https://doi.org/10.1136/bmj.n137

Rodriguez, M. Y., & Storer, H. (2020). A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data*. *Journal of Technology in Human Services*, *38*(1), 54–86. https://doi.org/10.1080/15228835.2019.1616350

Rozbroj, T., Lyons, A., & Lucke, J. (2020). Vaccine-Hesitant and Vaccine-Refusing Parents' Reflections on the Way Parenthood Changed Their Attitudes to Vaccination. *Journal of Community Health*, *45*(1), 63–72. https://doi.org/10.1007/s10900-019-00723-9

Scheibling, C., & Marsiglio, W. (2020). #HealthyDads: "Fit Fathering" Discourse and Digital Health Promotion in Dad Blogs. *Journal of Marriage and Family*. https://doi.org/10.1111/jomf.12743

Schmidt, A. L., Zollo, F., Scala, A., Betsch, C., & Quattrociocchi, W. (2018). Polarization of the vaccination debate on Facebook. *Vaccine*, *36*(25), 3606–3612. https://doi.org/10.1016/j.vaccine.2018.05.040

Segrin, C., Woszidlo, A., Givertz, M., & Montgomery, N. (2013). Parent and Child Traits Associated with Overparenting. Journal of Social and Clinical Psychology, 32(6), 569–595. https://doi.org/10.1521/jscp.2013.32.6.569

Segrin, C., Woszidlo, A., Givertz, M., Bauer, A., & Murphy, M. T. (2012). The Association Between Overparenting, Parent-Child Communication, and Entitlement and Adaptive Traits in Adult Children. *Family Relations*, *61*(2), 237–252. https://doi.org/10.1111/j.1741-3729.2011.00689.x

Sharevski, F., Jachim, P., & Florek, K. (2020). To tweet or not to tweet: Covertly manipulating a Twitter debate on vaccines using malware-induced misperceptions. *Proceedings of the 15th International Conference on Availability, Reliability and Security*, 1–12. https://doi.org/10.1145/3407023.3407025

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70. https://doi.org/10.3115/v1/W14-3110

Silge, J., & Robinson, D. (2020). *Text Mining with R: A Tidy Approach*. O'Reilly Media, Inc.

Silk, M., Millington, B., Rich, E., & Bush, A. (2016). (Re-)thinking digital leisure. *Leisure Studies*, *35*(6), 712–723. https://doi.org/10.1080/02614367.2016.1240223

Szablewski, C., Chang K., Brown M., Chu, V., Yousaf, A., Anyalechi, N., Aryee, P., Kirking, H., Lumsden, M., Mayweather, E., McDaniel,C., Montierth, R., Mohammed, A., Schwartz, N., Shah, J., Tate, J., Dirlikov, E., Drenzek, C., Lanzieri, T., Stewart, R. SARS-CoV-2 transmission and infection among attendees of an overnight camp — Georgia, June 2020. Morbidity and Mortality Weekly Report, 69(31), 1023–1025. http://dx.doi.org/10.15585/mmwr.mm6931e1ex

Uhls, Y. T., Michikyan, M., Morris, J., Garcia, D., Small, G. W., Zgourou, E., & Greenfield, P. M. (2014). Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues. *Computers in Human Behavior*, *39*, 387–392. https://doi.org/10.1016/j.chb.2014.05.036

Wada, M., Grigorovich, A., Kontos, P., Fang, M. L., & Sixsmith, J. (2021). Addressing Real-World Problems Through Transdisciplinary Working. In A. Sixsmith, J. Sixsmith, A. Mihailidis, & M. L. Fang (Eds.), *Knowledge, Innovation, and Impact: A Guide for the Engaged Health Researcher* (pp. 121–129). Springer International Publishing. https://doi.org/10.1007/978-3-030-34390-3_17

Warshaw, C., Vavreck, L., & Baxter-King, R. (2020). Fatalities from COVID-19 are reducing Americans' support for Republicans at every level of federal office. *Science Advances*, *6*(44), eabd8564. https://doi.org/10.1126/sciadv.abd8564

Wiley, K. E., Leask, J., Attwell, K., Helps, C., Degeling, C., Ward, P., & Carter, S. M. (2020). Parenting and the vaccine refusal process: A new explanation of the relationship between lifestyle and vaccination trajectories. *Social Science & Medicine*, *263*, 113259. https://doi.org/10.1016/j.socscimed.2020.113259

Williams, J. T. B., O'Leary, S. T., & Nussbaum, A. M. (2020). Caring for the Vaccine-Hesitant Family: Evidence-Based Alternatives to Dismissal. *The Journal of Pediatrics*, *224*, 137–140. https://doi.org/10.1016/j.jpeds.2020.05.029

Winner, N. A., & Nicholson, B. C. (2018). Overparenting and Narcissism in Young Adults: The Mediating Role of Psychological Control. *Journal of Child and Family Studies*, *27*(11), 3650–3657. https://doi.org/10.1007/s10826-018-1176-3

Yuan, X., Schuchard, R. J., & Crooks, A. T. (2019). Examining Emergent Communities and Social Bots Within the Polarized Online Vaccination Debate in Twitter. *Social Media + Society*, *5*(3), 2056305119865465. https://doi.org/10.1177/2056305119865465