Georgia State University
ScholarWorks @ Georgia State University

Business Administration Dissertations | Programs in Business Administration

8-15-2022

Using Consumer-Generated Social Media Posts to Improve Forecasts of Television Premiere Viewership: Extending Diffusion of Innovation Theory

Robert Casey Goodman
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/bus_admin_diss

Recommended Citation

Goodman, Robert Casey, "Using Consumer-Generated Social Media Posts to Improve Forecasts of Television Premiere Viewership: Extending Diffusion of Innovation Theory." Dissertation, Georgia State University, 2022.
doi: https://doi.org/10.57709/30238345

This Dissertation is brought to you for free and open access by the Programs in Business Administration at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Business Administration Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

USING CONSUMER-GENERATED SOCIAL MEDIA POSTS TO IMPROVE FORECASTS OF TELEVISION PREMIERE VIEWERSHIP: EXTENDING DIFFUSION OF INNOVATION THEORY

BY

ROBERT CASEY GOODMAN

A Dissertation

Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Business Administration

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY

ROBINSON COLLEGE OF BUSINESS

2022

Copyright by

Robert Casey Goodman

2022

ACCEPTANCE

This dissertation was prepared under the direction of the ROBERT CASEY GOODMAN Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctorate in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Naveen Donthu, Ph.D., Chair

Denish Shah, Ph.D.

Kai Zhao, Ph.D.

ACKNOWLEDGEMENTS

A number of people helped me on my journey, and I'd like to take this opportunity to recognize them.

The dissertation committee was generous with their valuable time and insights. Naveen Donthu, my advisor, provided guidance in many ways. Denish Shah and Kai Zhao volunteered important ideas that shaped this research.

Lars Mathiassen and Louis Grabowski, two faculty leaders of Georgia State's Executive DBA program, have made this program truly outstanding. They each generously made time to help me along the way many times.

My parents, Bill and Anne Goodman, are proponents of education and lifelong learning. Their unfailing support, encouragement and reassurance made this possible.

Jason Smith was a supportive partner from the program's start to finish. He was always there when I needed him.

My sister, Laura Goodman-Accola, her husband, Pete Accola, and their daughters Emma, Eva and Kate allowed me to get away, have some fun, and rediscover my inner child while playing with my nieces.

My classmates in the DBA program, cohort 2022 – each has unique strengths that I appreciate and find inspiring.

In memory of my grandparents, Mary and Adolph Gallia, and Millie and Gus Goodman. I am grateful to have had time to know each of them.

TABLE OF CONTENTS

ABSTRACT 7
CHAPTER I. INTRODUCTION 8
CHAPTER II. LITERATURE SYNTHESIS 10
CHAPTER III. CONCEPTUAL DEVELOPMENT AND HYPOTHESES 21
CHAPTER IV. DATA SOURCES AND METHODOLOGY 23
CHAPTER V. RESULTS 30
CHAPTER VI. DISCUSSION 41
REFERENCES 46
APPENDIX I: LIST OF 52 SHOWS INCLUDED IN ANALYSIS 51
APPENDIX II: LDA RESULTS FOR EACH SHOW 53
VITA 105

LIST OF TABLES

Table 1: Summary of Literature on Entertainment Forecasting with eWOM 11
Table 2: Other Key Variables Considered 14
Table 3: Operationalization of Constructs in Conditional Process Mediation Model 29
Table 4: Frequency of Series Premieres by Network 30
Table 5: Frequency of Series Premieres by Year 31
Table 6: Twitter Statistics of Consumer Accounts vs. Non-Consumer Accounts 31
Table 7: Means, Standard Deviations, Ranges, and Correlations 32
Table 8: Conditional Process Model Summary 33
Table 9: Comparison of Unscripted Show Prediction Mean Absolute Percentage Error With and Without eWOM in Model 36
Table 10: Zero-Order Correlations of eWOM metrics 38
Table 11: Major Topics Identified with Top 10 Most Frequent Words and their Percentages, Across All 52 Show LDAs 39

Table 12: VADER Sentiment Score Means and Ranges for Total, Show-Related Topics, and Person/Character-Related Topics 40

Table 13: VADER Sentiment Scores for Exemplar Tweets 40

LIST OF FIGURES

Figure 1: Conditional Process Mediation Model 22

Figure 2: Premiere Performance by Genre 32

Figure 3: Interaction of Consumer Online Buzz (M) by Premiere's Genre (Z) 34

Figure 4: Interaction of Media Presence (X) by Lead-In Audience (W) 35

Figure 5: Consumer Buzz vs. Media Presence, 14 Days Prior to Premiere (10 Unscripted Shows) 36

ABSTRACT

Billions of US dollars in transactions occur each year between media companies and advertisers purchasing commercials on television shows to reach target demographics. This study investigates how consumer enthusiasm can be quantified (via social media posts) as an input to improve forecast models of television series premiere viewership beyond inputs that are typically used in the entertainment industry. Results support that Twitter activity (volume of tweets and retweets) is a driver of consumer viewership of unscripted programs (i.e., reality or competition shows). As such, incorporating electronic word of mouth (eWOM) into forecasting models improves accuracy for predictions of unscripted shows. Furthermore, trend analysis suggests it is possible to calculate a forecast as early as 14 days prior to the premiere date. This research also extends the Diffusion of Innovation theory and diffusion modeling by applying them in the television entertainment environment. Evidence was found supporting Rogers's (2003) heterophilous communication, also referred to by Granovetter (1973) as "weak ties." Further, despite a diffusion pattern that differs from other categories, entertainment consumption demonstrates evidence of a mass media (external) channel and an eWOM interpersonal (internal) channel.

CHAPTER I. INTRODUCTION

Nielsen television ratings are the basis for billions of dollars of advertising transactions between media companies and advertisers every year. Forecasts of viewership help establish the cost of advertising in advance of a show's premiere. (Nielsen, 2017) For instance, 30 seconds of advertising during a Super Bowl is more costly than 30 seconds of advertising during a cooking show because of the size and demographics of their respective audiences. As such, networks seek to deliver a large audience within demographic groups (e.g., adults age 18 to 34) that are attractive to advertisers for achieving their marketing objectives. If a show fails to perform to expectations, then "make-goods" are offered by the network – additional commercials in other shows – to ensure an ad gets in front of the number of viewers agreed to. Make-goods, however, often do not fit the advertiser's media plan as well as the shows originally purchased. Reliable forecasts help media companies and advertisers make more accurate, less subjective transactions.

Forecasting viewership can be challenging because it is difficult to incorporate consumer enthusiasm for a new program into a forecasting model as a quantifiable variable. Social media may be a means of assessing this enthusiasm. Social media represents a "town square" of sorts, in which consumers share their opinions. Advances in information technology (e.g., smartphones) and the emergence of online social networking have profoundly changed how some types of information are disseminated. For researchers, electronic word of mouth (eWOM) transcends traditional limitations of conventional word of mouth (WOM) research. Messages are enduring and often visible to the entire world, making them much easier to analyze (Nielsen, 2013a). Harnessing this rich data to facilitate a more accurate marketing strategy would provide a considerable business advantage.

Unlike the consumer-packaged goods industry, which uses a wide range of prelaunch forecasting models at various stages of product development, the television entertainment industry invests millions of dollars into creating (writing, production, casting) and marketing a product with comparatively little data-driven evidence regarding how many consumers will be interested in it. Thus, the television industry is fertile ground for improved forecasting models that provide a decision support system to serve as a

"second opinion." In the German movie, music, and book categories, Hoffmann-Stölting, Clement, Wu, and Albers (2017) found that models can outperform management judgment in most instances. The exception was top-selling products, which usually receive more attention during development and marketing.

The purpose of this study is to investigate how social media can be leveraged as quantitative consumer inputs into forecasting models of television show viewership. We focus on quantifying the volume, sentiment, and topics of consumer-generated social media posts anticipating a particular television premiere. We then use this as an input for modeling viewership of a television series premiere (i.e., episode 1 of season 1). Importantly, the model in this research also includes variables that are already the foundation of viewership forecasting in the entertainment industry. In this way, we will demonstrate whether it is possible to improve accuracy beyond what is already widely used in the entertainment industry.

The research question is: ***How can consumer-generated social media posts improve forecasts of television premiere viewership?*** As such, a quantitative study was conducted combining Nielsen rating data, social media data, and data on media presence to test a new forecasting model for television series premieres (i.e., episode 1 of season 1). Series premieres represent a relatively simple case because there is no need to account for factors such as viewing in prior seasons, and no episodes are yet available via streaming or on-demand. This research strives to make two contributions. First, on the theoretical side, the research contributes to the Diffusion of Innovation theory beyond traditional new product marketing to apply it to the consumption of television entertainment. Second, on the applied side, it adds insight and understanding, as well as more accuracy, to viewership forecasting models used by media companies and advertisers by quantifying consumer enthusiasm.

The literature synthesis that follows draws from several literature streams: forecasting entertainment consumption, the impact of media presence on consumer eWOM, and television network scheduling of lead-in effects. Diffusion of Innovations (Rogers, 2003) and the Bass (1969) diffusion model provide a theoretical lens for this research.

CHAPTER II. LITERATURE SYNTHESIS

This literature synthesis is organized as follows. Section 2.1 presents a review of research on eWOM and forecasting in the entertainment industry. Section 2.2 provides an overview of Diffusion of Innovation Theory (Rogers, 2003) and how the Bass (1969) model will frame this research. Section 2.3 compares the demographics of Twitter users to those of entertainment consumers. Section 2.4 reviews academic research on lead-in effects and audience "flow" from one show to the next. Lastly, section 2.5 discusses media presence, a concept borrowed from policy and public agenda-setting research, to assess consumer awareness and attention.

2.1. Viewership Forecasting and eWOM Literature

The literature on viewership forecasting with eWOM can be divided into three groups. Table 1 summarizes key studies within each group. Group A is most relevant to the current research and describes studies that use eWOM variables to predict Nielsen ratings for television programs. Group B, instead of television ratings, is focused on the movie (i.e., motion picture) box office revenue, which is also a measure of consumer viewership or entertainment consumption, and text analysis of movie reviews (sentiment analysis or topic modeling) to derive eWOM data directly from posts for forecasting inputs. Group C also focuses on predicting movie box office revenue instead of actual text analysis. However, they use *proxy measures* as eWOM data for forecasting inputs. For example, a consumer rating in a movie review is used as a proxy for the sentiment of the text of the review itself. Despite the proxy measures, the studies in group three also demonstrate a link between consumer-generated posts and viewership. Studies in Group B and Group C use post-premiere reviews, which indicate a film's ultimate success. These are, however, not useful for forecasting premiere audiences. As such, the current study pushes beyond this limitation.

This study's contribution to the discussion includes a more comprehensive treatment of eWOM variables, including volume, dispersion, sentiment, and topics. It uses Twitter, which is broadly used by consumers. Its cross-sectional analysis lends itself to real-world forecasting. It also includes variables that are part of typical entertainment industry forecasts.

Table 1: Summary of Literature on Entertainment Forecasting with eWOM

	Reference	Outcome Variable(s)	eWOM Variable(s)	Findings Regarding eWOM
This Research		Nielsen ratings (A18-34)	**Volume** (tweets, unique users), **Dispersion** (retweets), **Sentiment** (positive, negative), **Topics** (via LDA)	eWOM variables Volume (tweets) and Dispersion (retweets), plus Lead-In and Media Presence (typically part of entertainment industry forecasting), demonstrate improved accuracy for Unscripted genre series premieres.
Group A. Television Program Forecasts	Crisci et al. (2018)	Nielsen data (Persons age 2+) reflecting Live viewership	**Volume** (unique users, total tweets), **Dispersion** (retweets), **Sentiment** (positive, negative)	Compared different time-series models for unscripted programs. Models trained on initial episodes of several competitive reality shows and then predicted viewership for the latter episodes. Volume (tweets), Dispersion, and Sentiment were significant.
	Nielsen Media Research (2013b)	Nielsen minute-by-minute ratings of Live viewership	**Volume** of tweets	In 44% of competitive reality episodes measured, Volume caused rating changes; 37% in comedy program episodes; 28% in sports; 18% in drama. Based on 221 broadcasts of primetime episodes.
	Godes & Mayzlin (2004)	Nielsen ratings (Households) Live+SD	Usenet newsgroup **volume** (number of posts) and **dispersion** (number of newsgroups in which program mentioned)	Dispersion is significantly related to television program premiere Nielsen ratings. No consistent support was found for Volume; no measure of sentiment was included.
Group B. Film Forecasts Using Text Analysis	Lehrer & Xie (2017)	Opening weekend box office revenue	**Sentiment** of Twitter posts	Incorporating social media sentiment data significantly improved forecast accuracy. Neither volume nor dispersion was included in the analysis.
	Chiang et al. (2014)	Annual box office revenue	IMDB review **topic** modeling text analysis	The Story Content cluster was most important for explaining box office sales of the four topic clusters.Neither sentiment nor dispersion was included in the analysis.
	Liu (2006)	Aggregate and weekly box office revenue	Yahoo!Movies review **valence** (manually coded)	eWOM valence and volume (percentage of positive reviews) directly impacted aggregate and weekly box office sales. No measure for dispersion was included.
Group C. Film Forecasts Using Proxy Measures for Text	Lee et al. (2019)	Weekly movie box office revenue	IMDB review **rating**, number reading review, or finding it helpful	Adjusted R-square for box office sales was over 50% using reviews from IMDB. Specifically, three variables consistently emerged as significant: sentiment (movie rating), number of people reading the post, and the number indicating the post was helpful (volume). No measure of dispersion was included.
	Holbrook & Addis (2008)	Gross domestic box office revenue	IMDB, Yahoo!Movies, rottentomatoes.com **ratings,** and **volume** (# of reviews)	eWOM volume and valence were significantly related to the box office sales (R-square = 0.369). No measure of dispersion was used.
	Duan, Gu &Whinston (2008)	Daily gross box office revenue	Yahoo!Movies review **volume** (unique YahooIDs) and **valence** (review grade)	eWOM valence generated higher eWOM volume, which in turn significantly impacted box office sales.
	Dellarocas, Zhang & Awad (2007)	Weekly box office revenue	Yahoo!Movies review **volume** (unique IDs), **valence** (review grade), **dispersion** (using gender & age in user profiles)	eWOM volume, valence, and dispersion were each significant predictors of box office sales

Group A: Television Program Forecasts Using eWOM Variables. Group A is comprised of three studies. First, Crisci et al. (2018) used longitudinal analysis to predict audiences for three Italian reality shows (e.g., *X-Factor*) based on tweets, retweets, distinct Twitter accounts, and sentiment (positive or negative) by training a model using social media data during the first ten episodes of the season and then using the model to predict viewership for the final three episodes.

Nielsen Media Research (2013b) published a white paper describing a time series analysis tracking live television ratings and volume of tweets minute-by-minute for 221 primetime episodes. They found that, overall, in 29% of the episodes, tweets influenced ratings. They also found that influence differed by genre. Tweets drive consumers to tune in to a program for 44% of competitive reality shows, 37% of comedies, 28% of sports episodes, and 18% of drama shows.

Godes & Mayzlin (2004) used Usenet newsgroup conversations, rather than Twitter, to predict Nielsen ratings for shows during the 1999-2000 seasons. Usenet is encompasses thousands of newsgroups covering diverse topics (e.g., rec.arts.tv). They found that dispersion (number of newsgroups in which a show is mentioned) was a critical explanatory factor, suggesting the importance of eWOM taking place across heterogeneous communities rather than concentrated within a small set of communities.

Dispersion was included in only two reviewed studies – Godes and Mayzlin (2004) and Dellarocas, Zhang, and Awad (2007). Recall dispersion is the diffusion of information across heterogeneous groups of consumers. Diffusion within a homogeneous group happens quickly; *dispersion* from one group to another takes more time. When dispersion was included, it had a statistically significant relationship with viewership. As such, dispersion merits attention in the current research.

Synthesis of findings. This literature indicates that Volume (number of tweets or number of unique Twitter accounts that were engaged), Sentiment (positive or negative), and Dispersion (the degree to which heterogeneous groups are engaged) are each viable predictors of Nielsen ratings. Furthermore, the genre is related to the degree to which tweets drive ratings (competitive reality shows being most responsive, and dramas least).

Group B: Film Forecasts and Text Analysis of Consumer Posts. Group B is also comprised of three studies. First, Lehrer and Xie (2017) assessed the sentiment of Twitter posts mentioning specific film titles released in North America from 2010 to 2013 (i.e., whether they expressed positive or negative sentiment toward a movie). They found that incorporating this social media metric improved forecast accuracy beyond only movie rating (e.g., PG, PG-13, R), genre (e.g., comedy, drama, thriller), and movie budget (in millions of dollars).

Liu (2006) used Yahoo! Movie reviews in a time-series analysis to predict weekly box office revenue for the opening week and eight weeks after. Three coders assigned 12,136 posts to one of the following five categories: positive, negative, neutral, mixed, and irrelevant. Volume of consumer-generated posts was a key explanatory factor during the early weeks of a release.

Chiang et al. (2014) analyzed consumer-generated reviews posted on IMDb for 29 movies. After generating a keyword frequency distribution, they ran a cluster analysis. Four clusters emerged: content (including "story" and "scene"), promotion (including "potential" and "introduce"), positive WOM (including "pretty," "fantastic," and "beauty"), and negative WOM (including "hate," "horrible" and "terrible"). The promotion keywords significantly affected box office revenue.

Synthesis of findings. This literature in this group underscores the importance of volume of consumer-generated posts. In addition, it indicates a possible link between box office sales and sentiment of reviews on Yahoo! Movies, as well as topic clusters of keywords (e.g., promotion) that were derived across multiple movies.

Group C: Film Forecasts and Proxy Measures for Text Analysis of Consumer Posts. Group C is comprised of four studies. First, Lee et al. (2019) and Holbrook and Addis (2008) considered box office revenue and consumer-generated movie ratings. On IMDb, registered users can rate a movie from one to ten (in addition to writing a review), and ratings are aggregated and summarized on the website. In addition to a significant relationship between box office revenue and rating (considered a proxy for sentiment or valence), these studies found that volume (measured by the number of reviews, number of seeing a review, or number indicating a post was helpful) was also a significant predictor.

Duan, Gu, and Whinston (2008) and Dellarocas, Zhang, and Awad (2007) used consumer-generated Yahoo! Movie review data to explore the relationship with box office revenue. Both teams of researchers found a relationship between box office sales and volume (measured by the number of unique Yahoo user IDs). The latter team also found a significant relationship with dispersion, which they measured by using partial data from user profiles to calculate the heterogeneity of the group of reviewers in terms of age and gender indicated in their profiles.

Synthesis of findings. In addition to movie ratings, this set of publications also found that volume (e.g., number of reviews, number of unique reviewers) was important as well as dispersion (measured in terms of heterogeneity of reviewers).

Factors Other than eWOM. Other explanatory variables were employed by the various studies and are summarized in Table 2 (e.g., genre, production budget). Importantly, some of these variables for films have analogs for television programs and will inform the current research. This research considers genre, eWOM discussion of people/characters involved with the show, the show's lead-in audience, and attention in the mass media.

Table 2: Other Key Variables Considered

Variable	Studies in which it was included
Genre	Lehrer & Xie (2017); Chiang, et al. (2014); Dellarocas, Zhang & Awad (2007); Liu (2006)
Actor "Star Power"/Producer	Duan, Gu & Whinston (2008); Dellarocas, Zhang & Awad (2007); Liu (2006)
Television Network	Godes & Mayzlin (2004)
Film Rating (G, PG, PG-13, R)	Lehrer & Xie (2017); Duan, Gu & Whinston (2008); Liu (2006)
Production Budget	Lehrer & Xie (2017); Holbrook & Addis (2008); Duan, Gu & Whinston (2008)
Number of Opening Screens	Lehrer & Xie (2017); Holbrook & Addis (2008); Duan, Gu & Whinston (2008)
Estimated Marketing Budget	Duan, Gu & Whinston (2008); Dellarocas, Zhang & Awad (2007)

NOTE: Lee et al. (2019) did not include key variables other than eWOM variables.

2.2. Key Literature on Diffusion of Innovations Theory & Diffusion Models

In the fifth edition of his book, *Diffusion of Innovations*, Rogers (2003) describes his theory explaining the proliferation of new ideas and their adoption. Diffusion of Innovations (DOI) is a process involving four components: "[1] an innovation [2] is communicated through certain channels [3] over time [4] among the members of a social system." (Rogers, 2003, p. 5). It is a theoretical frame for researching and understanding awareness and implementing new ideas.

DOI theory has been applied in various disciplines, including new product marketing, such as new computers and automobiles (Mahjan et al., 1990). In this context, potential adopters communicate about a new product (the innovation) to generate awareness and share opinions. This paper argues that DOI can also be used as a lens to interpret the consumption of television entertainment (e.g., television programs). We will consider how communication about a new television program (the innovation) occurs via mass media and eWOM channels over the weeks leading up to the premiere among potential viewers.

Social Media and Diffusion

Rogers (2003) specifically addresses social media in DOI theory, stating that "in addition to mass media and interpersonal communication channels, interactive communication via the Internet has become more important for the diffusion of certain innovations…" (p. 18). Twitter can more readily facilitate the diffusion of ideas compared to other social networking sites like Facebook or LinkedIn because links between individuals can arguably be classified as "weak ties" (Virk, 2011). All content on Twitter is public by default, and users elect to "follow" someone on Twitter without needing to be confirmed by the other as a "friend." Further, decisions on whether to follow another Twitter account are often based on content rather than relationships. That is, users often follow other users because they desire the content of their posts, not necessarily because of a personal or professional relationship. "Weak ties" are more conducive to the diffusion of information (Granovetter, 1973). Weak ties are referred to by Rogers (2005, p. 305) as "heterophilious network links," which can bridge disparate cliques and facilitate the spread of an idea across various groups of consumers.

Nielsen Media Research (2013a) conducted an analysis focusing specifically on television programs and the social media engagement surrounding them. They concluded that, on average, 50 people see the Twitter posts of each author commenting on a television program (via followers, hashtags, and handles). Therefore, if 500 users are tweeting about a program, then 25,000 people see their Tweets. This multiplier decreases, however, as the number of authors for an individual program increases. This is due to the overlap of potential Twitter followers. In other words, a single follower is likely to follow multiple authors, hashtags, etc.

Entertainment Consumption and Diffusion

The diffusion pattern of television entertainment media differs from products in other categories (e.g., durables such as computers or automobiles). It is important to be mindful of five distinctions. First, television entertainment has a short product life cycle. Second, consumers can influence other consumers by simply expressing *intentions* to view entertainment media rather than doing so. Third, early adopters are important for the awareness and persuasion of later adopters. Fourth, this research focuses on the first two stages of Rogers's adoption process: knowledge and persuasion. Fifth, buzz is characterized not only by volume or amount of eWOM but also by its dispersion across populations. Each of these distinctions is discussed briefly below.

One obvious issue is the shorter product life cycle of television entertainment; that is, the time from when a product is introduced until it is "replaced" by a newer version of the product. After the premiere, one week typically elapses until episode two shows; at that point, episode one is no longer the latest. For the television entertainment industry, viewership (i.e., adoption) usually peaks during or shortly after release, except in the rare case of "sleeper hits" that become more successful later (Ainslie, Drèze, and Zufryden; 2005). For other categories, each product life cycle stage can take many months.

A second important distinction, especially relevant for this research, concerns the stages in the adoption process. Rogers (2003) indicates there are five stages: (1) knowledge when a consumer is made aware of the innovation, (2) persuasion when an individual forms her or his attitude toward the innovation, (3) decision when an individual adopts or rejects the innovation, (4) implementation when an

individual uses the innovation and (5) confirmation when individuals encounter reinforcing information or information that leads to reversal of the adoption decision. As this research concerns the time leading up to a series premiere, only the first three stages are relevant (knowledge of a show, attitude formation of a show, and the decision to view). These stages can be influenced by individual demographics such as age and gender (MacVaugh and Schiavone, 2010).

A third key difference is the role of early adopters in the diffusion of television entertainment media before release. Frattini, et al. (2013) conclude that early adopters have two functions in diffusion. First, they disseminate information about a new product. In entertainment, they propagate information about a new show and their opinion. Second, they communicate to later adopters that they intend to view, or have viewed, the new television entertainment show. In this way, early adopters influence later adopters even before the product's release (Hofmann-Stölting et al., 2017).

The fourth distinction is that, for entertainment, the *intention* to view (expressed prior to a release) influences others rather than the *actual adoption* of a product itself (Hui, Elishberg, and George, 2008). This is because television media is experiential, and consumers rely on cues (e.g., star power, critical acclaim) that serve as quality indicators prior to consumption (Caliendo, Clement, and Shehu, 2015). Consequently, marketers attempt to use these cues to generate high anticipation before release (Burmester, Becker, van Heerde, and Clement, 2015).

The fifth and final distinction was investigated by Houston et al. (2014), who found that buzz is a function of not only the *volume* of WOM but also its pervasiveness, which they define as the *dispersion* across populations (p. 514). Their analysis included new product buzz when consumers make adoption decisions before a new product is available (as in the case of deciding whether or not to watch a new series premiere). Their research included two studies. Furthermore, specific to entertainment, they analyzed buzz for 254 movies released in North America between 2010 and 2011 and confirmed volume and dispersion were both predictors of opening-weekend box office revenues.

Diffusion Models for Entertainment Consumption

Several types of diffusion models have been investigated in the context of the entertainment industry. Most work in the area has dealt with movie releases. Two studies are of particular importance for this research.

Dellarocas, Zhang, and Awad (2007) used the Bass (1969) model to demonstrate how online movie reviews generated by consumers during the two to three days following initial release can improve revenue forecasts for later weeks of a film's run. They discuss two parameters of the Bass model in addition to market size and time: external influences (parameter *P*), which included marketing, publicity, and critic reviews prior to the film's release; and internal influences (parameter *Q*), which they measured through online discussion groups, included awareness of the new movie, assessment of the movie's quality, and dispersion of information across various groups of consumers. Both parameters were important additions to the model's accuracy.

Another study by Hui, Eliashberg, and George (2008) modeled DVD preorders and post-release sales. It is similar to the current research on television premiere viewership. When a movie DVD release date is announced, prior to the release date, consumers can typically preorder it to receive it the day of its release. Their model mirrored the typically-observed pattern for DVD sales: an exponentially increasing number of preorder units peaking at release, followed by an exponentially drop post-release.

The Bass (1969) model is a population-level model associated with an S-curve depicting market share (cumulative) as successive groups of consumers (from early adopters to laggards) adopt an innovation over time. The S-curve is an a priori assumption of the Bass model. In contrast, Hui, Eliashberg, and George's model is based on individual-level behavior and is associated with Pareto distributions (highly skewed with a heavy tail). The pattern these authors observed mirrored actual DVD sales data from a major internet DVD provider and was an outcome of the model rather than an a priori assumption.

2.3. Twitter Users Mirror Entertainment Consumers

Twitter social media posts (as opposed to Facebook and Instagram) are publicly accessible. Furthermore, Twitter's demographic is generally an excellent match for the entertainment marketing demographic. In 2019, one in five (22%) U.S. adults used the social media platform Twitter. Pew (2019) surveyed nearly three thousand U.S. adult Twitter users who consented to share their Twitter handles. This allowed their actual Twitter behavior to be linked to their survey responses. Twitter users tend to be considerably younger (age 18-29, 21% US vs 29% US adult Twitter users; age 30-49, 33% vs 44%, respectively). They also had slightly higher household incomes than the general population. In 2014, the Motion Picture Association found that the largest frequent (i.e., monthly or more often) moviegoers were under 40 years old and tended to own more technology products (e.g., smartphones, computers, tablets) than the general population.

2.4 Lead-In Effects and Impact on Ratings

Lead-In Effects, or "Flow," refer to the tendency of a show's audience to stay tuned in to the next show on a network (Webster, 2006). It is the basis for most television program scheduling strategies. If a lead-in show's audience is large, it conveys an advantage to the show immediately following. Conversely, if a lead-in garners a small audience, the next show is handicapped.

For series premieres, networks often precede a premiere with a strong lead-in to give it the best launch possible. As Tiedge and Ksobiech (1986) demonstrate, the lead-in is even more effective when the two programs are of the same genre (e.g., comedy followed by a new comedy). The effect is less effective when the genres do not match (e.g., sports followed by drama).

2.5 Media Presence and Impact on Consumer eWOM

Media presence is most commonly measured by the volume of stories or content dedicated to a topic. This is done by borrowing a key concept from policy and agenda-setting research. Kiousis (2004) found two dimensions within media presence: visibility and valence, with visibility being the dominant dimension (e.g., total story frequency in media or prominence of stories on the front page). This visibility also influences what is salient in consumers' minds. Danner, Hagerer, Pan, and Groh (2020) investigated

agenda setting for organic food in the US and Germany via online news outlets. They demonstrated that topics in online news articles strongly influenced reader comments under those articles. Conway-Silva, Filer, Kenski, and Tsetsi (2018) extended this to social media in an analysis of 2016 presidential campaign tweets and newspaper topics and found a greater influence of newspapers on the campaigns' Twitter feeds.

Barkemeyer et al. (2017) used keyword searches, including the Dow Jones Factiva database, to build a regression model to identify factors that impacted the volume of media coverage of climate change. Similarly, Gurun and Butler (2012) used Factiva to analyze local media mentioning local companies. The present research uses keywords from television show titles and the networks on which they appeared.

2.6 Unscripted Genre and Impact on Ratings

Nielsen Media Research (2011) looked at different genres in primetime and found that reality shows accounted for the least viewership (16%). Dramas were the highest (44%), followed by Sports (22%) and then Comedies (18%). Although unscripted shows generally have lower viewership (except for occasional "hits" such as *American Idol* and *The Voice*), they are also generally less expensive for studios and networks to produce than scripted shows (South University, 2016). Depending on the network and the content of a show, budgets for reality shows can range from $100k to more than $500k per episode, whereas scripted shows can range from $500k to several million dollars per episode.

CHAPTER III. CONCEPTUAL DEVELOPMENT AND HYPOTHESES

Like other social media, Twitter has made a wealth of eWOM readily available for mining in the form of online posts – consumer information through natural language. This unstructured text data can shed light on consumer behavior if appropriately quantified. For example, automated text analysis can help to sift through tens of thousands of Tweets sent every day about a particular subject, allowing researchers to sample from the Twitter "firehose" at any time and assess what consumers are saying in near real-time.

This research will attempt to derive consumer insights that predict entertainment consumption using consumer-generated Twitter posts. This research considers only consumer-generated Twitter posts. Non-consumer social media (e.g., social media marketing, podcasts, entertainment news, etc.) are excluded as much as possible because the intent herein is to use Twitter to glean consumer insights instead of generating PR.

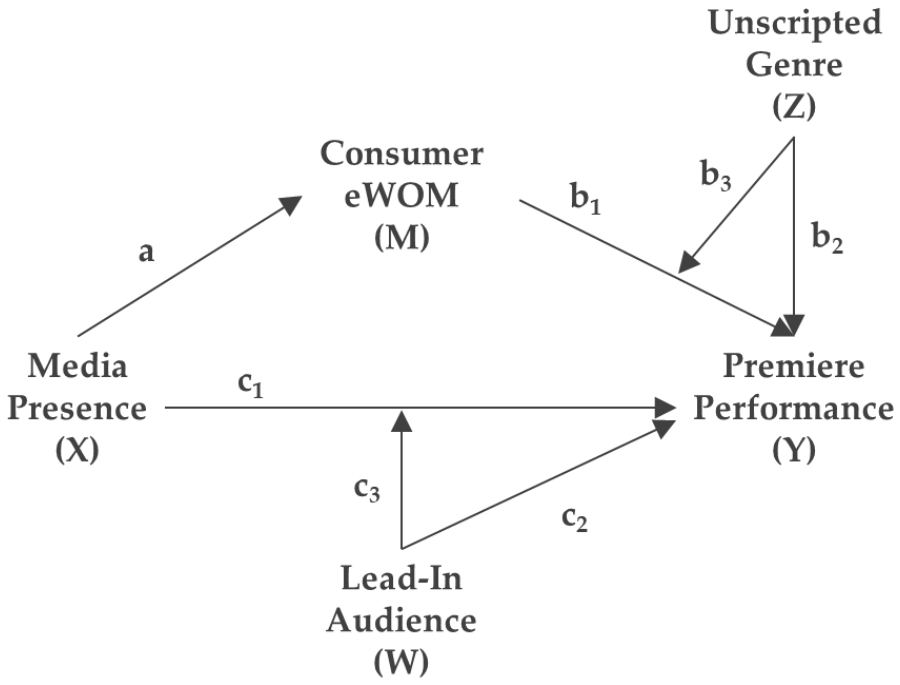
Demand for television premieres is operationalized through Nielsen television program ratings for that premiere. Consumer anticipation of television premieres will be operationalized through Twitter posts (and various aspects of those posts) containing a specific title or hashtag. Tweets about a television program will be analyzed for sentiment (positive or negative valence), number of posts (volume), retweets (dispersion), and likes. Tweets will also be analyzed by key topics and the topics' valence, volume, dispersion, and the number of likes. Mass media is operationalized via the number of mentions a particular show receives in print and television news media.

Hypotheses are based on the literature review, as summarized in Table 1 of the previous section, and the theoretical framework of the Diffusion of Innovation theory. Importantly the results for the television forecast study (group one) dovetail nicely with those for the film forecast studies. All studies in the table will inform the hypotheses for television premiere forecasting. Currently, forecasts are based on historical viewership of the lead-in show (the show airing just prior to the series premiere) and on the amount of attention in the mass media (e.g., interviews of cast members, critical reviews, promotion). The

forecast would then be adjusted up or down based on whether it deviates from the usual genre or content of a network at the time or whether an actor or director has particularly strong "star power."

The statistical model for the hypotheses is below, where e_M and e_Y are error terms in the estimation of M and Y, respectively.

Figure 1: Conditional Process Mediation Model



The equations specifying this model are below, where i_M and i_Y are regression constants.

$$\hat{M} = i_M + aX$$

$$\hat{Y} = i_Y + c_1X + c_2W + c_3XW + b_1M + b_2Z + b_3MZ$$

This research has four hypotheses considering the literature and theory reviewed earlier. Each is operationalized in the next chapter with data sources and methodology.

H_1: Media Presence (X) has a positive indirect effect on Premiere Performance (Y) via generating Consumer Online Buzz (M).

H_2: Media Presence (X) has a positive direct effect on Premiere Performance (Y).

H_3: Lead-In Audience (W) has a positive direct effect on Premiere Performance (Y).

H_4: Unscripted Genre (Z) has a negative direct effect on Premiere Performance (Y).

CHAPTER IV. DATA SOURCES AND METHODOLOGY

This section details the data sources and methodology used to test the regression model research hypotheses., A data set was constructed from multiple data sources to test the regression model. Then, each show in the dataset required topic modeling and sentiment analysis, and these results were also included in the data set. Finally, the regression model was used to test the hypotheses.

4.1 Data Sources

Nielsen Ratings were collected for new show premiere performance and viewership of the lead-in show. Twitter provided consumer social media posts anticipating each new show's premiere. From this unstructured data, volume, dispersion, sentiment, and topic metrics were derived. Dow Jones Factiva was leveraged to collect the media presence metric for each new show premiere. Lastly, each new show was referenced on IMDbPro, used by entertainment industry professionals to identify its genre.

Nielsen Media Research

For viewership data, television program ratings from Nielsen Media Research were used. This data is reported online at ShowbuzzDaily (showbuzzdaily.com). Official Broadcast Nationals are reported for five major networks: ABC, CBS, NBC, FOX, and the CW for primetime (8:00 PM to 11:00 PM). The data is Live Plus Same Day (L+SD) and is reported by three major sales demos: Adults 18 to 34, Adults 18 to 49, and Adults 25 to 54. Live Plus Same Day includes persons who watched a program either while it aired (i.e., "Live") or watched it time-shifted (e.g., via DVR) on the same day the program was broadcast. This measure was used because it captures the increase in time-shifted viewing but focuses on viewing the day of the premiere instead of "binge-watching" or viewing multiple episodes stored on DVR or streaming.

Data are based on a nationally representative panel of over 65 thousand households. These households have meters installed in their homes which continually collect and send to Nielsen data on everything watched on television (Nielsen, 2020). In-home meters reduce potential recall bias that would be a problem with diaries or other recall methods. *Nielsen Ratings* are an audience measurement for programs (Percentage of people in a demographic who tuned into a particular show). For example, a

rating of 3.0 would mean 3% of persons (e.g., persons age 18-34) with access to television were tuned in to a program. Today, of course, access to televisions is ubiquitous among those in the US.

Dow Jones Factiva

Factiva is a tool that aggregates content from various media sources, including newspapers, periodicals, and, importantly for this research, network television news transcripts. The total number of stories will be retrieved from Factiva using a keyword search of the show title and network (e.g., "Murphy Brown" and "CBS"). Duplicates were removed from counts.

Twitter Text Corpus

Python 3 was used to collect the social media data via queries of Twitter's Application Programming Interface (API) v1.1 to return sets of Tweets matching specific criteria from Twitter's historical database. One query was run for each of the 52 television shows. Each query specified the following criteria:

- Time Frame: 14 days leading up to midnight the day of each show's premiere. This was done to assure all downloaded tweets were posted before a premiere and avoided the issue of premieres airing at various times in different time zones.
- Key Words: All posts had to include either that show's handle or its hashtag, which were retrieved from the show's Twitter account.
- Retweet posts were excluded from the query to capture only original tweets (although a count of retweets of original posts was obtained; see below).

This research used only English-language posts generated by U.S.-based accounts to further hone the focus.

The following data fields were downloaded for each post; each represented a column in the text analysis data frame for each show:

1. Date and time the tweet was posted
2. Twitter account/screen name originating the post
3. Text of post itself
4. Count of likes of the post, as of the day the query was run

5. Count of retweets of the post, as of the day the query was run

4.2 Methodology

Text analysis preprocessing, topic modeling, sentiment analysis, and the regression model hypothesis tests are detailed below.

Text Analysis Preprocessing

Text analysis includes sophisticated approaches that quickly and accurately classify and quantify enormous amounts of unstructured textual data. Unlike tabular data, which is generally numeric, text data is unstructured. Natural Language Processing (NLP) is rooted in artificial intelligence research (Ponweiser, 2012) and treats text as data by using dummy variables to represent individual words. We get this data frame by cleaning or "pre-processing" the text. For example, in social media, topics regarding a specific television show premiere (e.g., an aspect of the plot, a popular actor/director) emerged via topic modeling. Then sentiment analysis will score specific topics in terms of the feelings and emotions associated with them. The steps taken in the current analysis are detailed below.

Removing posts, not from consumer accounts. As this analysis was focused on consumer-generated social media, non-consumer accounts were removed. First, a list of Twitter accounts originating tweets for each show was manually reviewed. Tweets originating from the show's marketing account (e.g., "@MaskedSingerFOX") were marked for removal, along with tweets from networks and studios (e.g., "@CBS" or "@CBSTVStudios"), television stations (e.g., "@FOX5Atlanta" or "@CW11Seattle"), entertainment news sources (e.g., "@TVGuideMagazine" or "@TVInsider"), podcasts (e.g., "@MurphyBrownPod"), special topics (e.g., "@BUZZRtv" follows classic game shows) and stars appearing in the shows who posted about it (e.g., Ken Jeong's Twitter account "@kenjeong" posted about the show on which he served as a judge, *I Can See Your Voice*).

Removing extraneous text. The text of Twitter posts contains some information irrelevant for topic modeling. If not removed, these would add unnecessary noise to the analysis. Therefore, the following was done with the text field of each post:

1. Removed text belonging to Twitter handles in each post by searching for strings of text beginning with "@." While removing handles from the text, indicator variables were created for (1) the show's handle and (2) other handles.
2. Removed text belonging to Twitter hashtags in each post by searching for text strings beginning with "#." While removing hashtags from the text, indicator variables were created for (1) the show's hashtag and (2) other hashtags.
3. Remove text belonging to URLs (e.g., to view posted pictures) by searching for strings of text beginning with "http." While URLs were removed from the text, an indicator variable was created for later analysis.
4. Remove punctuation
5. Remove digits

Key words/phrases, stemming and removing stop words (RAKE). Even with the pre-processing done thus far, much noise remains. Rose, Engel, Cramer, and Cowley's (2010) Rapid Automatic Keyword Extraction (RAKE) algorithm, an unsupervised, domain-independent method, was used to remove unwanted variability. RAKE stems words by separating suffixes or prefixes (e.g., "shows," "showing," and "showed" becomes "show"). The function then uses a set of stop words (e.g., "and," "the," "of") and a set of phrase delimiters (e.g., hyphens, quotation marks) to identify candidate keywords (contiguous words often used together in the text). The frequencies of their co-occurrence then identified keywords and phrases. This way, phrases like "Fantasy Island" were retained and analyzed together. RAKE also removes "noise" such as tweets that contain little information that would be useful for topic modeling (e.g., "Are we watching #AMillionLittleThings?"). The use of RAKE before LDA is consistent with the methodology used by Jeong, Yoon, and Lee (2017).

The resulting keyword frequency matrix was cast into a Document Term Matrix for LDA with keywords and document information. In a Document Term Matrix (DTM), each row represents one tweet (i.e., "document"), and each column represents a word or phrase (as identified via RAKE), with as many columns as there are unique words or phrases in the entire corpus of Twitter posts for that show. Values

represent the number of appearances of a word in a post. As expected, this is a sparse matrix comprised mostly of zero values.

Topic Modeling

In this research, the goal of topic modeling is exploratory – to discover the inherent structure of large volumes of Twitter data – rather than to predict an outcome using a training dataset. As such, an unsupervised machine learning process, Latent Dirichlet Allocation (LDA), was used (Silge and Robinson, 2017). LDA assigns the content of the Twitter text corpus for each particular show to topics. However, as LDA is an unsupervised data-driven analysis, steps were taken to reduce "noise" or unwanted variability that would cloud results. A manual review was then needed to confirm the topics were meaningful.

Analogous to cluster analysis of numerical data, topic modeling classifies text such as social media posts into groups. Each post is considered a mixture of topics, and each topic has a mixture of words. The advantage of this method is that, like natural language, topics can have some overlap in their use of words (i.e., they are "soft clusters"). For every word, probabilities (Beta, β) are generated, reflecting how likely that word is to belong to each of the topic "clusters" that emerge.

When running the LDA algorithm for topic modeling on each show's DTM, several topics (k) were manually assigned. Then, a model was run for this number of topics in order to estimate topic and word distributions. The number of topics was determined using a data-driven metric developed by Deveaud, SanJuan, and Bellot (2014). This index indicates a model's optimal number of topics by maximizing dissimilarity between topics. Importantly, the authors conclude that their metric can be used to evaluate latent concepts in short pieces of text, such as tweets.

The resulting topic clusters were manually reviewed and labeled based on the top words associated with each. Recall topics are a mixture of tokens (in this case, keywords or phrases). After arriving at the optimal number of topics, the ten words with the greatest probability (Beta, β) of belonging to each topic were used to help interpret it. Importantly, topic-word density can differ by topic. That is, some topics have fewer words associated with them than others. That is, topics associated with more

words have lower Betas overall than topics with fewer words. Therefore, betas for each topic were assigned a Z-value. In this way, words could be compared within and across topics to standardize this for a more valid cross-topic comparison.

Sentiment Analysis

Sentiment analysis uses text to derive a measure of feelings and emotions. One approach to automated sentiment analysis is to develop general lists of words expressing either positive or negative sentiment (Crossley et al., 2017). (This is referred to as a domain-independent bag-of-words approach.) A computer algorithm then "scores" the text based on the use of these words within the text.

One such word list, VADER (Valence Aware Dictionary and sEntiment Reasoner; Hutto and Gilbert, 2014), was specifically developed for sentiments expressed in social media and will thus be used here. For instance, it is sensitive to sentiment expressions in social media such as "LOL" (a popular initialism for "Laughing Out Loud"), emoticons (e.g., ":-)"), and slang such as "meh" (used to express a lack of enthusiasm or boredom). It yields polarity scores in both semantic dimensions (positive and negative). It represents a gold standard based on combining qualitative and quantitative methods. It includes grammar and syntax conventions (e.g., capitalization or modifiers like "very" or "extremely") for expressing sentiment and its intensity. Further, VADER accounts for negation (e.g., "The situation is *not* good.") to avoid misclassifying such text as a positive statement.

The overall sentiment score for each Twitter post can be aggregated to arrive at a sentiment score for each topic identified in the topic analysis. In this way, it is possible to understand what is driving overall sentiment.

Hypothesis Tests of Conditional Process Model with Moderation

All data for each program (ratings, media presence, Twitter-related data) were assembled into a single database. Regression analyses were conducted in line with current standards of mediation analysis to test the hypotheses (Hayes, 2013). Specifically, multiple regression was conducted in which the indirect effects of the predictor variable (media presence) via a proposed mediator (online consumer buzz) and moderators (lead-in audience and genre) are simultaneously tested using bootstrapping, which is

especially appropriate given the limited sample size for this study (Preacher and Hayes, 2004). The SPSS macro for this analysis is available at www.processmacro.org.

Table 3: Operationalization of Constructs in Conditional Process Mediation Model

Consequent Variable (Y) and Antecedents

Y Viewership of show premiere (A18-34 Live+SD)

W Viewership of lead-in show (A18-34)

Z Genre (Unscripted reality or Comedy/Drama)

X Media Presence (# of times show appears in the media in the 14 days leading up to premiere)

M eWOM Volume (# of consumer-generated posts in the 14 days leading up to premiere)

eWOM Dispersion (# of retweets of consumer-generated posts in the 14 days leading up to premiere)

CHAPTER V. RESULTS

A total of 52 series premieres (i.e., shows designated at episode 1 of season 1) were included in this analysis. Two shows were removed prior to analysis because they represented extraordinary cases. The first show, *The Equalizer* (CBS, 2021), had the *Super Bowl LV* as its lead-in. Because of this, lead-in and premiere ratings were extremely inflated (Lead-in A18-34 = 11.90; premiere A18-34 = 3.41) and not representative of other *The Equalizer* episodes that aired later. The second show, *Evil* (CBS, 2019), used the hashtag "#Evil," which was used as a Twitter hashtag in many other contexts beyond the discussion of the television show. This skewed sentiment analysis and Twitter statistics (e.g., retweets). In addition, a third show, *The Masked Singer* (FOX, 2019), had high premiere ratings that artificially inflated regression results. Given the limited number of cases, instead of excluding this program due to a single outlying datapoint, the rating was Winsorized, replaced with the next largest rating for an unscripted program, to limit the impact of the outlier (A18-34 = 1.70; recoded to 0.80).

5.1. Key Data Distributions

The 52 shows aired on the following networks (see Table 4). Proportions for the major networks were similar, with the CW network contributing the fewest cases.

Table 4: Frequency of Series Premieres by Network

Network	Count	Proportion
ABC	15	29%
CBS	11	21%
FOX	11	21%
NBC	10	19%
CW	5	10%
Total	**52**	**100%**

Below (Table 5) is a breakdown of the shows by the year they premiere. Proportions were similar for each year except for 2020. In 2020, the COVID-19 pandemic shut down or delayed the production of television programs, especially those scheduled to air in the latter half of that year (White, 2020). Consequently, fewer premieres from 2020 were available to include in the sample.

Table 5: Frequency of Series Premieres by Year

Premiere Year	Count	Percentage
2018	18	35%
2019	19	37%
2020	4	8%
2021	11	21%
Total	**52**	**100%**

5.2 Data Manipulation Check

A total of 71,543 tweets were downloaded across all 52 shows. Non-consumer accounts were removed manually. As displayed in Table 6, although these accounts represented only 2% of total Unique Accounts, they accounted for 50% of total Twitter posts. Non-consumer accounts averaged 63.2 tweets per account (median = 6) versus 1.5 tweets per account (median = 1) for consumer accounts.

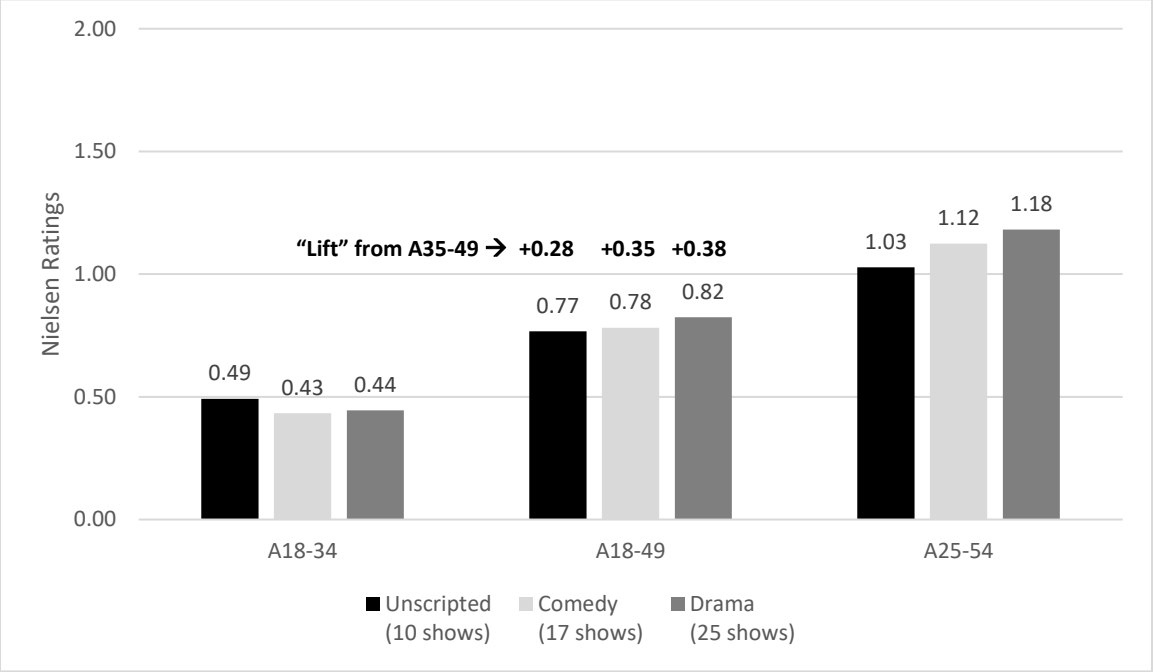
Table 6: Twitter Statistics of Consumer Accounts vs. Non-Consumer Accounts

	Total Tweets		Unique Accounts		Average # Tweets	Median # Tweets
Consumer Accounts	35,732	50%	23,927	98%	1.5	1
Non-Consumer Accounts	35,811	50%	566	2%	63.2	6
Total	71,543		24,493			

5.3 Nielsen Ratings

The 52 shows comprised three major genres: Unscripted (e.g., competition or reality shows), Comedy, and Drama. Ratings by the three major sales demos (Adults age 18 to 34, Adults age 18 to 49, and Adults age 25 to 54) are displayed in the figure below (Figure 2). Note the "Lift" provided by including older adults ages 35 to 49 is smaller for Unscripted than for the other genres. Therefore, unscripted audiences tend to skew younger.

Figure 2: Premiere Performance by Genre



5.4 Hypothesis Tests via Conditional Process Regression Analysis

Each of the features of the regression model is described in Table 7, with means, standard deviations, and ranges; zero-order correlations are also presented. In addition, a log transformation was used on Measure 3, Total # of Consumer Tweets + Retweets, to reduce the variability of the data.

Table 7: Means, Standard Deviations, Ranges, and Correlations

	Mean	Std. Deviation	Range	1	2	3	4
1. Premiere Performance (A18-34)	0.45	0.23	0.04 - 1.20	-			
2. Lead-In Audience (A18-34)	0.71	0.43	0.10 - 2.10	0.74**	-		
3. LOG Total # Consumer Tweets + Retweets	2.84	0.61	1.08 - 3.99	0.26	0.13	-	
4. Media Presence (# mentions)	39.79	27.83	0 – 164	0.23	0.11	0.43**	-
5. Genre (1=Unscripted / 0=Comedy or Drama)	0.19	0.40	0 - 1	0.09	0.13	-0.31*	-0.45**

N=52
* Significant ($p < 0.05$)
** Highly significant ($p < 0.01$)

Table 8 summarizes the regression model, with overall model statistics toward the bottom of the panel.

Table 8: Conditional Process Model Summary

		Consequent									
		Consumer eWOM (M)					Premiere Performance (Y)				
		Coeff	**SE**	**t**	**p-val**			**Coeff**	**SE**	**t**	**p-val**	
Media Presence (X)	*a*	0.010	0.003	3.407	0.001	*	c_1	0.002	0.002	1.228	0.226	
Consumer eWOM (M)		–	–	–	–		b_1	0.011	0.041	0.268	0.790	
Lead-In Audience (W)		–	–	–	–		c_2	0.117	0.112	1.040	0.304	
Unscripted Genre (Z)		–	–	–	–		b_2	-0.476	0.225	-2.116	0.040	*
Interaction (X by W)		–	–	–	–		c_3	0.006	0.002	2.538	0.015	*
Interaction (M by Z)		–	–	–	–		b_3	0.232	0.088	2.642	0.011	*
		R^2 = 0.19						**R^2 = 0.67**				
		F(1, 50) = 11.61, p < 0.01**						**F(6, 45) = 15.45, p < 0.01****				

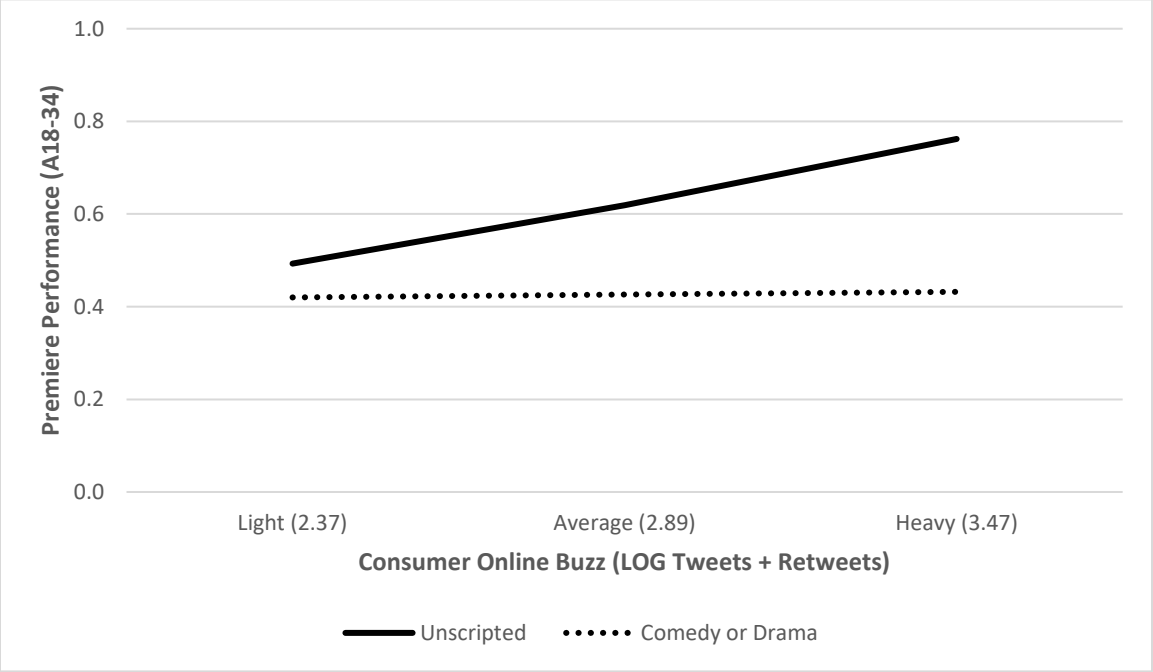
N = 52
** Significant (p < 0.05)*
*** Highly significant (p < 0.01)*

H_1: Media Presence (X) has a positive indirect effect on Premiere Performance (Y) via generating Consumer eWOM (M).

Conditional support was found for Hypothesis 1. As shown in Table 8, there was a significant positive relationship between Media Presence (X) and Consumer eWOM (M), $t(50) = 3.407$, $p < 0.01$. Although the relationship between Consumer eWOM (M) and Premiere Performance (Y) did not achieve significance, $t(45) = 0.268$, $p > 0.10$, the effect of Consumer eWOM on Premiere Performance was contingent on Unscripted Genre, $t(45) = 2.642$, $p < .05$.

Figure 3 (below) shows that, for Unscripted shows, Consumer eWOM positively affected Premiere Performance. However, this relationship did not hold for the other genres (comedies or dramas). This interaction accounts for 5.1 percent of variance in the model ($\Delta R^2 = 0.051$), $F(1, 45) = 6.98$, $p = 0.01$.

Figure 3: Interaction of Consumer Online Buzz (M) by Premiere's Genre (Z)



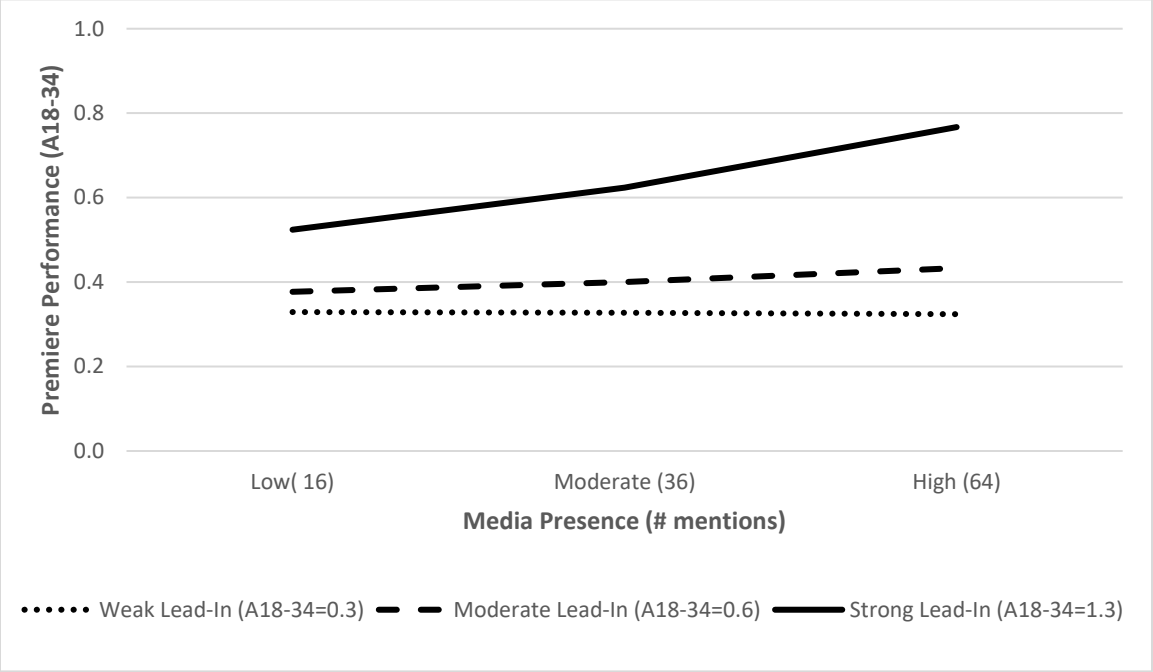
H_2: Media Presence (X) has a positive direct effect on Premiere Performance (Y).

H_3: Lead-In Audience (W) has a positive direct effect on Premiere Performance (Y).

Hypothesis 2 and Hypothesis 3 are considered together because conditional support was found for both. The direct effect of Media Presence (X) on Premiere Performance (Y) was not significant, $t(45) = 1.228$, $p > 0.10$. Similarly, the direct effect of Lead-In Audience (W) on Premiere Performance (Y) did not achieve significance, $t(45) = 1.040$, $p > 0.10$.

However, a significant interaction between Media Presence (X) and Lead-In Audience (W) was found, $t(45) = 2.538$, $p < 0.05$. Figure 4 (below) shows that, for strong Lead-Ins (with A18-34 ratings around 1.3), Media Presence positively affected Premiere Performance. However, this relationship disappeared for moderate and weak Lead-Ins (with A18-34 around 0.6 and below). The interaction accounts for 4.7 percent of variance in the model ($\Delta R^2 = 0.047$), $F(1, 45) = 6.44$, $p = 0.02$.

Figure 4: Interaction of Media Presence (X) by Lead-In Audience (W)



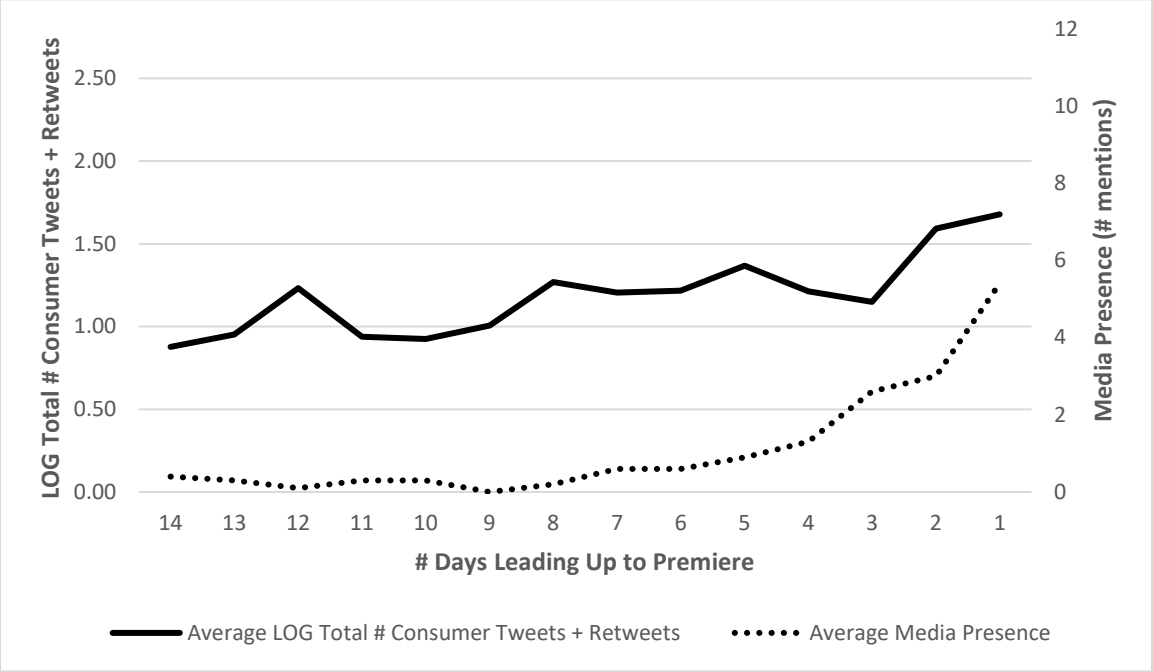
H_4: Unscripted Genre (Z) has a negative direct effect on Premiere Performance (Y).

Hypothesis 4 was supported by the data. Genre (Z) had a significant direct effect on Premiere Performance (Y), $t(45) = -2.116$, $p < 0.05$. Unscripted shows tended to have lower Premiere Performance than other genres of shows, such as comedies or dramas.

5.5 Longitudinal Analysis

Unscripted Program totals were broken out by the 14 days leading up to each premiere to further explore the temporal relationship between Consumer Online Buzz and Media Presence. There was a significant lagged correlation between Media Presence and Consumer Online Buzz, indicating a causal relationship between the two, $r(10) = 0.81$, $p < 0.01$. Further, one-way ANOVA contrasts show a highly significant linear trend for Consumer Online Buzz, $F(1, 12) = 10.32$, $p < 0.01$ and a highly significant quadratic trend for Media Presence, $F(1, 11) = 28.03$, $p < 0.01$. See Figure 5 below.

Figure 5: Consumer Buzz vs. Media Presence, 14 Days Prior to Premiere (10 Unscripted Shows)



5.6 Predictive Validation

The conditional process regression model in Table 8 was then used to predict performance for three unscripted shows premiering in 2022. Results are reported in Table 9, which shows the actual A18-34 rating and compares the model *with* eWOM included and *without* eWOM. Mean absolute percentage error (MAPE), summarized at the bottom, with eWOM was 26% compared to 80% MAPE without eWOM.

Table 9: Comparison of Unscripted Show Prediction Mean Absolute Percentage Error With and Without eWOM in Model

				Absolute Percentage Error	
Unscripted Show Title	**Network**	**Premiere Date**	**A18-34 Rating**	**With eWOM in Model**	**Without eWOM in Model**
To Tell The Truth	ABC	2/22/2022	0.17	25%	78%
Who Do You Believe?	ABC	5/3/2022	0.10	12%	80%
That's My Jam	NBC	1/3/2022	0.20	40%	82%
Mean				**26%**	**80%**

Although even 26% MAPE for A18-34 ratings between 0.10 and 0.20 requires additional refinement to be useful in practical applications, the improvement with eWOM (26% vs. 80%) is compelling. Furthermore, recall significant longitudinal trends were observed for Social Media Buzz and Media Presence (see Figure 5). Finally, paired with data available to networks on the historical performance of lead-in shows, this research demonstrates it is possible to calculate a prediction up to 14 days in advance. This would allow marketing managers to use consumer eWOM for better decision support to fulfill audience delivery targets.

5.7 Social Media Measure Correlations

All social media metrics considered in this study, plus the outcome measure (Premiere Performance), are shown in the table below (Table 10), with zero-order correlations. The metrics associated with volume and dispersion (numbers 2 through 6) are all strongly intercorrelated. The same can be said for metrics of sentiment (numbers 7 through 10). These seem to represent independent dimensions within the data. Importantly, none have significant zero-order correlations with the outcome measure. Measure 6 (LOG of Total Consumer Tweets + Retweets) is most strongly associated with it. Measure 11 (Average # of Likes per Tweet) is most strongly associated with Measure 4 (Total # of Consumer Retweets).

Table 10: Zero-Order Correlations of eWOM metrics

	1	2	3	4	5	6	7	8	9	10
1. Premiere Performance (A18-34)	–									
2. Total Unique Accounts	0.15	–								
3. Total # Consumer Tweets	0.13	.98**	–							
4. Total # Consumer Retweets	0.07	.63**	.67**	–						
5. Total # Consumer Tweets + Retweets	0.10	.82**	.85**	.96**	–					
6. LOG Total # Consumer TW + RT	0.26	.77**	.76**	.65**	.75**	–				
7. VADER Composite Valence Score	0.06	0.05	0.08	0.13	0.12	0.12	–			
8. VADER Positive Valence Score	0.12	0.23	0.24	0.16	0.21	0.24	.85**	–		
9. VADER Negative Valence Score	-0.03	0.14	0.12	0.01	0.06	0.13	-.44**	0.00	–	
10. VADER Neutral Valence Score	-0.09	-0.27	-0.27	-0.15	-0.21	-0.27	-.53**	-0.87	-.49**	–
11. Average # of Likes per Tweet	0.15	0.12	0.13	.48**	.38**	.34*	0.16	-0.03	-0.26	0.15

5.8 Topic Analysis Results

Of the 52 LDAs run, two topics emerged as optimal for 49 shows, and three topics emerged as optimal for three shows. Topics were coded into two categories: show-related and person/character-related. Table 11 reports the keywords for each topic.

Show Topics

Consumers typically posted that they can't "wait" or have been "waiting" a while or are looking "forward" to the "show." The "premier" "episode" is this "week" or "[weekday of premiere]" on "[ABC, CBS, CW, FOX, NBC]."

Person/Character Topics

Premiere ratings (A18-34) did not differ by whether a Person/Character topic emerged, $F(1, 49) = 0.050$, $p = 0.83$. Consumers typically posted using the name of a cast member (Alec Baldwin, *The Alec Baldwin Show*; Nathan Fillion, *The Rookie*; Candice Bergen, *Murphy Brown*) who is "back" on television or posted about a character depicted in the show (Higgins, *Mangum P.I.*; Hannibal Lecter, *Clarice*), or someone otherwise related to the show (Mariah Carey wrote the theme song for *Mixed-ish*) or notable in some way (Azita Ghanizada an Afghan American actor in the *United States of Al*). They then went on to

mention in their posts the “show” “premiere” this “week” on “[ABC, CBS, CW, FOX, NBC]” and how they can’t “wait.”

Table 11: Major Topics Identified with Top 10 Most Frequent Words and their Percentages, Across All 52 Show LDAs

Show-Related Topics (52 total)			Person-Related Topics (25 total)		
Words (stemmed)	**# of Shows**	**% of top Words**	**Words (stemmed)**	**# of Shows**	**% of top Words**
Show	41	22%	[star name]	34	33%
Week	19	10%	show	20	19%
[network name]	33	17%	wait	8	8%
forward	17	9%	week	8	8%
Episode	14	7%	[network name]	8	8%
wait	14	7%	premier	6	6%
premier	14	7%	back	5	5%
[weekday of premiere]	15	8%	episode	5	5%
day	12	6%	forward	5	5%
time	10	5%	day	4	4%

6.4 Sentiment Analysis

VADER Sentiment Scores are shown in Table 12 for total, show-related, and person/character-related topics. Overall, posts tended to be only slightly positive, and this held for both types of topics as well (see Table 12).

Total Sentiment Scores were not significantly correlated with Premiere Performance (A18-34 ratings; see Table 10). In addition, it did not differ by whether a person/character topic emerged (see Table 10).

Table 12: VADER Sentiment Score Means and Ranges for Total, Show-Related Topics, and Person/Character-Related Topics

	Positive (0 to 1)	**Negative (0 to 1)**	**Neutral (0 to 1)**	**Composite (-1.0 to +1.0)**
Total	0.16	0.05	0.79	0.06
(52 shows)	0.00 – 1.00	0.00 – 1.00	0 – 1.00	-0.90 – 0.96
Show-Related Topic	0.15	0.05	0.80	0.05
(52 shows)	0.00 – 1.00	0. .00 – 1.00	0. .00 – 1.00	-0.89 – 0.96
Person/Character-Related Topic	0.17	0.04	0.79	0.07
(25 shows)	0. 00 – 1.00	0. 00 – 1.00	0. 00 – 1.00	-0.90 – 0.96

Exemplar tweets are below (Table 13) for illustration.

Table 13: VADER Sentiment Scores for Exemplar Tweets

	Positive	**Negative**	**Neutral**	**Composite**
Tweets Assigned to Show Topic				
"@DatingGameABC @ZooeyDeschanel @mbsings @OnTheRedCarpet This show is gonna be super awesome!! Super excited to watch!!"	0.69	0	0.31	0.95
"@AMillionABC This show looks horrible. Suicide is a sad reality in some lives, but no where should it be the basis of a tv show. #notwatching"	0	0.41	0.59	-0.92
Tweets Assigned to Person/Character Topic				
"@TheRookie @NathanFillion I love it, Nathan is so funny, charming & gorgeous #TheRookie #SuperExcited"	0.76	0	0.24	0.95
"@AlecBaldwinShow I'm a die hard progressive but Alec Baldwin is not who we need as an ambassador of the left. His horrific treatment of his exwife and homophobic comments that got him fired from @MSNBC are disqualifying in my mind. We can do better."	0.07	0.26	0.68	-0.90

CHAPTER VI. DISCUSSION

The relationship between social media and entertainment consumption is indisputable, but researchers have long been working to understand its nature. This study bridges theory and practice around this issue with the research question: *How can consumer-generated social media posts improve forecasts of television premiere viewership?* Guided by this research question, a quantitative data frame was constructed, composed of a total of 52 television programs across genres, with data from Nielsen, Twitter, Dow Jones Factiva, and IMDbPro to test a statistical model that included Consumer eWOM (Tweets and Retweets).

6.1 Viewership Forecasting and eWOM

This study demonstrates the complexity of the relationship between social media and entertainment consumption in that conditional support was found for key hypotheses. Conditional support was found for *Hypothesis 1: Media Presence (X) has a positive indirect effect on Premiere Performance (Y) via generating Consumer Online Buzz (M).* Twitter activity (volume of tweets and retweets) helps to drive ratings for unscripted programs (i.e., reality or competition shows). This is consistent with prior research on eWOM and forecasts of television programs. Nielsen (2013b) found in a minute-by-minute analysis that tweets had the greatest impact on ratings for competitive reality programs. Similarly, Crisci et al. (2018) were able to train a model on the initial episodes of several competitive reality shows and then predict viewership for the latter episodes based on tweets and retweets.

The mechanism for this interaction effect is unclear. It may be that younger audiences are more likely to use social media for the diffusion of ideas, and unscripted shows generally skew younger than other genres (e.g., dramas). Another explanation involves the genre itself and audience engagement. For example, Crisci et al. (2018) selected talent competition shows because of the high level of audience engagement in voting for acts to push them ahead and keep them from elimination. While this study includes several types of competitions (talent, cooking, dating, obstacle courses), judges rather than the audience determine who advances, and other shows in the data are reality show documentaries (e.g., Emergency Call focuses on 911 call takers and the dramatic events leading up to the arrival of help).

Furthermore, this study investigates the 14 days leading up to the first episode of season one. As such, there was little opportunity for competition shows to develop the level of engagement found in Crisci et al.

It is also noteworthy that neither Twitter sentiment nor Twitter topic emerged as significant predictors in the model. Tweets and Retweets indicate support that, prior to a premiere, social media plays a role in generating awareness of a show. Still, their role is not persuasive in that positive sentiment improves viewership. Perhaps personal opinion of a show largely depends on personal preference, and the sentiment is disregarded. This might change later in a season because Twitter posts would be based on actual experience with plot, character, etc. If so, spin-off series (i.e., new series containing characters originating in a previous series) might also be more likely to be impacted by sentiment because of experience with the previous series.

Hypothesis 2, Media Presence (X), has a positive direct effect on Premiere Performance (Y), and *Hypothesis 3, Lead-In Audience (W), has a positive direct effect on Premiere Performance (Y),* are considered together because their relationship is conditional. Findings indicate that strong lead-ins (with ratings around 1.3 for A18-34) work together with media presence to give premieres an advantage. Audience "delivered" from a lead-in show will stay tuned for the premiere if they are primed for it by attention in the media in the two weeks preceding the premiere. However, the effect does not hold for moderate or weak lead-ins (with ratings around 0.6 or below for A18-34).

Hypothesis 4, Unscripted Genre (Z) has a negative direct effect on Premiere Performance (Y), was confirmed, which is consistent with general trends in the entertainment industry. Unscripted shows, apart from breakout hits (e.g., American Idol), tend to deliver somewhat lower audiences. However, networks include them in their season line-ups because they are generally less expensive to produce than their scripted counterparts (comedies and dramas). In addition, they are attractive to advertisers because they tend to attract younger audiences.

6.2 Contributions & Implications for Theory

This research provides support for the important role of dispersion across groups, referred to by Granovetter (1973) as "weak ties" and by Rogers (2003) as heterophilous communication. Although tweets, retweets, and unique accounts were all strongly intercorrelated, only Consumer Tweets plus Consumer Retweets were significant in the model. Therefore, retweets are a mechanism for information diffusion on Twitter, considering the content-driven, low-reciprocity relationships in the Twitter environment. This is consistent with Ahn and Park's (2015) network analysis using Twitter data, which found that weak links and sharing information via retweets play a more important role in information diffusion. In addition, the Diffusion of Innovations theory predicted the importance of volume and dispersion (Houston et al., 2014) and the literature on entertainment forecasting with eWOM (Table 1).

Diffusion of Innovations theory also provides a theoretical framework to understand why neither Sentiment nor Topic was significant in the series premiere forecast model. Pre-premiere, during the early stages of DOI adoption, Twitter posts serve primarily as an awareness function for innovators and early adopters. Their role is not generally persuasion (second stage of DOI adoption process) because post sentiment as early as 14 days prior to premiere was not associated with greater viewership. At this stage, with such limited information (television promos and news media coverage), consumers may recognize that personal opinion of a show does not reflect the show itself; therefore, sentiment is disregarded. Awareness rather than persuasion (attitude formation via sentiment valence) as a mechanism for early diffusion is consistent with extant literature on DOI (Frattini et al., 2013) and entertainment forecasting (Table 1). This relationship might change later in the season when Twitter posts would be based on experience with the show's plot, characters, etc.

This research also provides support for the Bass (1969) diffusion model's assumption that product adoption (in this case, entertainment consumption) is a function of two channels: mass media (external) and interpersonal (internal) channels, in addition to market size and time. In this research, media presence is an external influence in that, together with a strong lead-in, media presence drives

premiere ratings. Consumer eWOM is an interpersonal (internal) influence in that it was shown to promote awareness of new unscripted series and drives premiere ratings.

6.3 Contributions & Implications for Practice

Above and beyond variables typically included in television-network entertainment show forecasts (i.e., lead-in and media presence), consumer-generated social media can be incorporated into forecasts to further hone forecast accuracy for unscripted programs. Significant trends were found in a longitudinal analysis of the 14 days prior to the premiere: eWOM increased linearly, and media attention increased in a quadratic fashion. Therefore, it is possible to use these inputs to predict performance before the premiere day, making a forecast even more useful to practitioners.

This study also provided evidence that consumer-generated social media posts have a role in predicting and *driving* viewership for unscripted shows. Temporal sequencing provides strong evidence for causality. Marketing implications for Unscripted shows include making sharable content available to encourage this.

The influence of lead-ins was also found to be contingent on media attention. This has implications for the network television program schedule. An effect was found only for strong lead-in increases with more media presence, but no effect was found for moderate or weak lead-ins. The strategic implication is to save strong lead-ins and the greatest mass media efforts for high-priority "flagship" shows that are key for a network. Shows that are lower priority can be scheduled to follow less-strong lead-ins, using shows of the same type (e.g., following a drama with another drama). Resources for mass media efforts can be reserved for flagship shows because this data suggest the impact would be negligible on actual viewership.

6.4 Study Limitations and Next Steps

There were a few unavoidable limitations of the study imposed by the data sources available. The showbuzzdaily.com site is an exceptionally reliable source of information for broadcast network ratings in primetime. However, it limited the current analysis to four broadcast networks (ABC, CBS, NBC, FOX, and the CW). Including series premieres on cable networks would increase the amount of data available

for analysis and improve the generalizability of the current model. Additionally, although media salience (number of mentions in media during the days leading up to premiere) was an informative variable, a separate but related construct would be the level of promotion or advertising for new series (e.g., commercials promoting the new series during the days leading up to premiere). With this additional data, the model could predict when additional promotion (which has monetary cost or opportunity cost for a studio or network) would be helpful.

The next steps for additional exploration are also clear. Recall that non-consumer Twitter accounts were the source of 50 percent of total tweet volume. It would be interesting to look at the impact of these social media posts on ratings to add to the current model and our understanding of social media marketing and series premiere ratings.

This analytic approach can be applied to a series of digital streaming services (e.g., Netflix, Amazon Prime, Disney +, HBO Max). These tend to have younger audiences who are more engaged with social media. Specifically, it would be interesting to look at the relationship between eWOM and viewership of premieres to increase subscriptions to streaming services. Also, although streaming services are on-demand, they use algorithms to suggest a new show at the conclusion of shows consumers choose to view. The current model can be used to assess the effectiveness of the algorithms in creating audience flow from established series to new series premieres, which makes the streaming service “sticky” and can prevent subscription cancelation.

REFERENCES

Ahn, H., & Park, J.-H. (2015). The structural effects of sharing function on Twitter networks: Focusing on the retweet function. *Journal of Information Science, 41*(3), 354–365.

Ainslie, A., Drèze, X., & Zufryden, F. (2005). Modeling movie life cycles and market share. *Marketing Science, 24*(3), 508-517.

Barkemeyer, R., Figge, F., Hoepner, A., Holt, D., Kraak, J. M., & Yu, P. S. (2017). Media coverage of climate change: An international comparison. *Environment and Planning C: Politics and Space, 35*(6), 1029-1054.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science, 15*(5), 215-227.

Burmester, A. B., Becker, J. U., van Heerde, H. J., & Clement, M. (2015). The impact of pre-and post-launch publicity and advertising on new product sales. *International Journal of Research in Marketing, 32*(4), 408-417.

Caliendo, M., Clement, M., & Shehu, E. (2015). The effect of individual professional critics on books' sales: Capturing selection biases from observable and unobservable factors. *Marketing Letters, 26*(4), 423-436.

Chiang, I., Wen, Y., Luo, Y, Li, M., & Hsu, C. (2014). Using text mining technique to analyze how movie forums affect the box office. *International Journal of Electronic Commerce Studies, 5*(1), 91-96.

Conway-Silva, B. A., Filer, C. R., Kenski, K., & Tsetsi, E. (2018). Reassessing Twitter's agenda-building power: An analysis of intermedia agenda-setting effects during the 2016 presidential primary season. *Social Science Computer Review, 36*(4), 469-483.

Crisci, A., Grasso, V., Nesi, P., Pantaleo, G., Paoli, I., & Zaza, I. (2018). Predicting TV program audience by using twitter-based metrics. *Multimedia Tools and Applications, 77*(10), 12203-12232.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2017). Sentiment analysis and social cognition engine (I): An automatic tool for sentiment, social cognition, and social order analysis. *Behavior Research Methods 49*(3), pp. 803-821. Doi:10.3758/s13428-016-0743-z.

Danner, H., Hagerer, G., Pan, Y., & Groh, G. (2022). The news media and its audience: Agenda setting on organic food in the United States and Germany. *Journal of Cleaner Production, 354*, 131503.

Dellarocas, C., Zhang, X. and Awad, N. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing 21*(4), 23-45.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique, 17*(1), 61-84.

Duan, W., Gu, B. and Whinston, A. (2008). The dynamics of online word-of-mouth and product sales – And empirical investigation of the movie industry. *Journal of Retailing, 82*(2), 233-242.

Frattini, F., Bianchi, M., De Massis, A., & Sikimic, U. (2014). The role of early adopters in the diffusion of new products: Differences between platform and non-platform innovations. *Journal of Product Innovation Management, 31*(3), 466-488.

Hofmann-Stölting, C., Clement, M., Wu, S., & Albers, S. (2017). Sales forecasting of new entertainment media products. *Journal of Media Economics, 30*(3), 143-171.

Godes, D and Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23*(4), 545-560.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380.

Gurun, U. G., & Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance, 67*(2), 561-598.

Hayes, Andrew F. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis*. Guilford Press, New York. ISBN: 978-1-60918-230-4.

Holbrook, M. B. & Addis, M. (2008). Art versus commerce in the movie industry: A two-path model of motion-picture success. *Journal of Cultural Economics, 32*, 87-107.

Houston, M. B., Knapp, A. K., Hennig-Thurau, T., & Spann, M. (2014). Conceptualizing new product buzz. *ACR North American Advances*.

Hui, S. K., Eliashberg, J., & George, E. I. (2008). Modeling DVD preorder and sales: An optimal stopping approach. *Marketing Science, 27*(6), 1097-1110.

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1, pp. 216-225).

Jeong, B., Yoon, J., & Lee, J. M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management, 48*, 280-290.

Kiousis, S. (2004). Explicating media salience: A factor analysis of New York Times issue coverage during the 2000 US presidential election. *Journal of Communication, 54*(1), 71-87.

Lee, C., Xu, X. and Lin, C (2019). Using online user-generated reviews to predict offline box-office sales and online DVD store sales in the O2O era. *Journal of Theoretical and Applied Electronic Commerce Research, 14*(1), 68-83.

Lehrer, S., and Xie, T. (2017). Box office Buzz: Does social media data steal the show from model uncertainty when forecasting for Hollywood? *The Review of Economics and Statistics, 99*(5), 749-755.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing 70*, 74-89.

MacVaugh, J., & Schiavone, F. (2010). Limits to the diffusion of innovation: A literature review and integrative model. *European Journal of Innovation Management*.

Mahajan, V., Muller, E., & Bass, F.M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing 54*, 1-26.

MPAA (2014), Theatrical market statistics 2014. Retrieved from www.mpaa.org/wp-content/uploads/2015/03/MPAA- Theatrical-Market-Statistics-2014.pdf

Nielsen Media Research (2011). The relationship between social media buzz and TV ratings. Retrieved from https://www.nielsen.com/us/en/insights/article/2011/the-relationship-between-social-media-buzz-and-tv-ratings.

Nielsen Media Research (2013a). Nielsen launches 'Nielsen Twitter ratings' Retrieved from www.nielsen.com/us/en/press-room/2013/ nielsen-launches-nielsen-twitter-tv-ratings.html.

Nielsen Media Research (2013b). The follow-back: Understanding the two-way causal influence between Twitter activity and TV viewership. Retrieved from https://www.nielsen.com/us/en/insights/article/2013/the-follow-back-understanding-the-two-way-causal-influence-betw/.

Nielsen Media Research (2017). Using machine learning to predict future TV ratings. Retrieved from https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/using-machine-learning-to-predict-future-tv-ratings.pdf.

Nielsen Media Research (2020). How we do it. Retrieved from https://www.nielsen.com/bh/en/solutions/measurement/television/.

Pew Research Center (2019). Sizing up Twitter Users. Retrieved from https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

Ponweiser, M. (2012). *Latent Dirichlet allocation in R*. Thesis, Institute for Statistics and Mathematics, Vienna University of Economics and Business, Vienna.

Preacher, K.J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments and Computers, 36*, 717-731.

Richards, Tia (2021). Predicting the future of the entertainment industry after COVID. Retrieved from https://news.usc.edu/183870/future-of-entertainment-after-covid-movies-tv-streaming-usc-experts/.

Rogers, E. M. (2003). *Diffusion of innovations. 5th Edition.* New York: Free Press.

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010) Automatic keyword extraction from individual documents. In *Text Mining: Applications and Theory* edited by Michael W. Berry and Jacob Kogan (John Wiley & Sons).

Silge, J, and Robinson, D (2017). *Text Mining with R: A tidy approach.* Sebastopol, CA: O'Reilly Media, Inc.

South University (2016). Reality TV – Low-cost programming that produces high ratings. Retrieved from https://www.southuniversity.edu/news-and-blogs/2016/08/reality-tv-low-cost-programming-that-produces-high-ratings-119585.

Tiedge, J. T., & Ksobiech, K. J. (1986). The "lead-in" strategy for prime-time: Does it increase the audience? *Journal of Communication, 36*(3), 64-76.

Virk, A. (2011). Twitter: The strength of weak ties. *University of Auckland Business Review, 13*(1), 19-21.

Webster, J. G. (2006). Audience flow past and present: Television inheritance effects reconsidered. *Journal of Broadcasting & Electronic Media, 50*(2), 323-337.

White, P. (2020). Coronavirus shutdown will delay 60% Of U.S. & global drama as scripted hangover set to extend well in 2021, data shows" *Deadline Hollywood.* Retrieved from https://deadline.com/2020/05/covid-19-shutdown-delay-international-drama-1202938005/

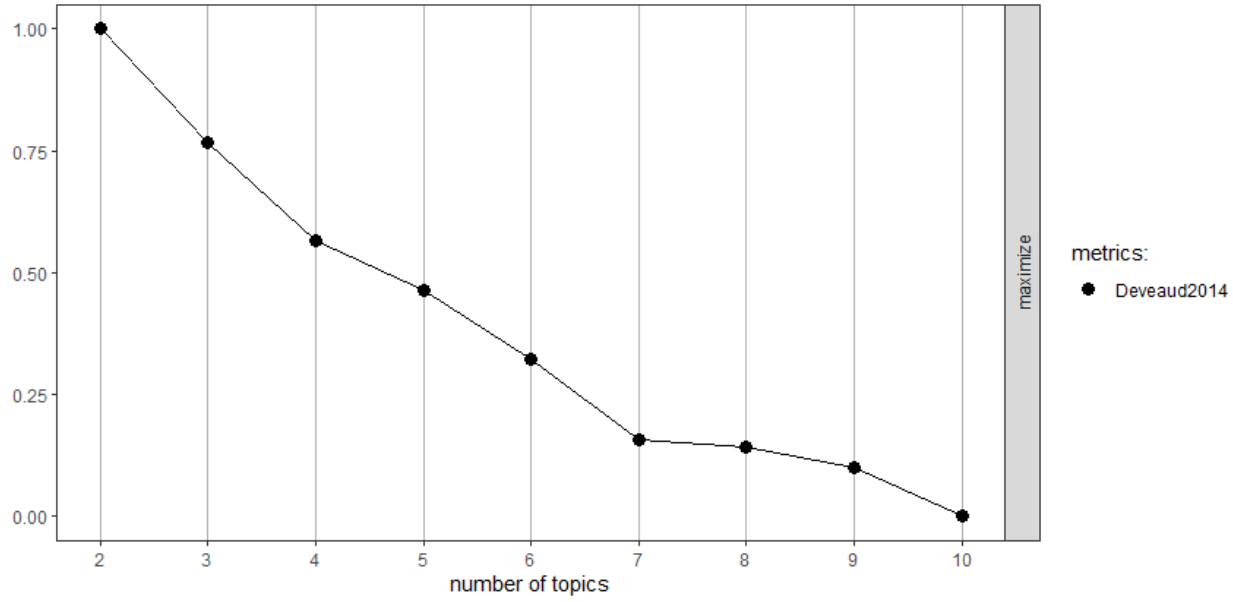
APPENDIX I: LIST OF 52 SHOWS INCLUDED IN ANALYSIS

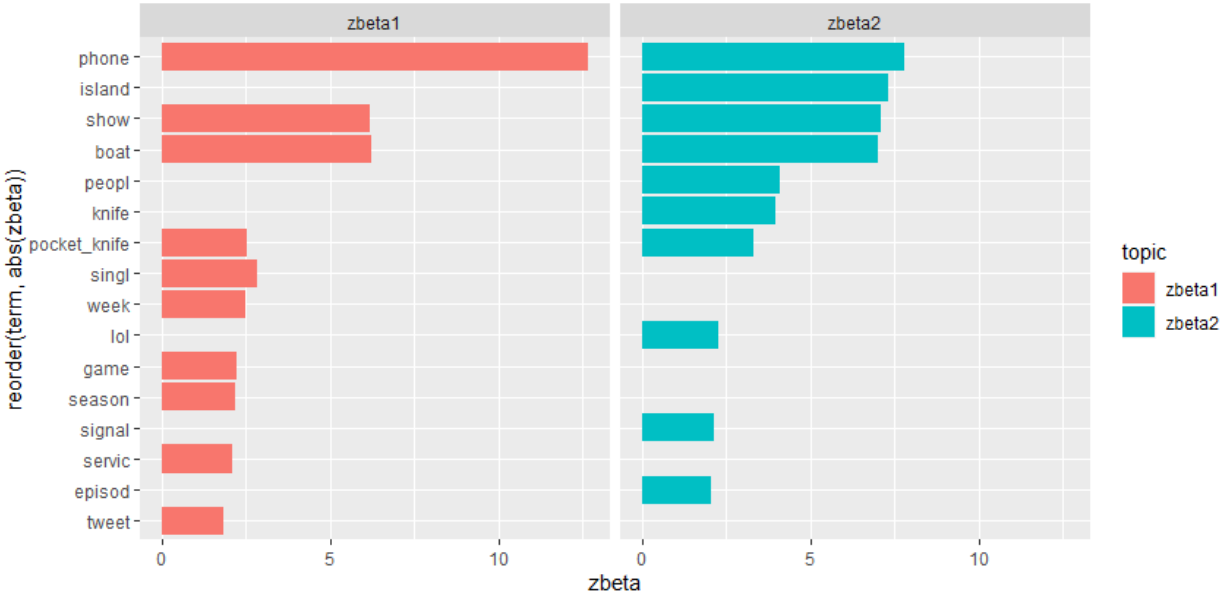
	Program Title	Genre	Premiere Date	Network	Lead-In
1	Castaways	Unscripted	8/7/2018	ABC	Bachelor in Paradise
2	Emergency Call	Unscripted	9/28/2020	ABC	Dancing with the Stars
3	Family Game Fight!	Unscripted	8/8/2021	NBC	Summer Olympics
4	Gordon Ramsay's 24 Hours to Hell & Back	Unscripted	13-Jun-18	FOX	MasterChef
5	I Can See Your Voice	Unscripted	9/23/2020	FOX	The Masked Singer
6	Press Your Luck	Unscripted	6/11/2019	ABC	The Bachelorette
7	Red Bull Peaking	Unscripted	9/13/2019	The CW	Madden NFL
8	The Celebrity Dating Game	Unscripted	6/14/2021	ABC	The Bachelorette
9	The Masked Singer	Unscripted	1/2/2019	FOX	24 Hrs to Hell & Back
10	TKO: Total Knock Out	Unscripted	7/11/2018	CBS	Big Brother
11	B Positive	Comedy	4/1/2021	CBS	Mom
12	Bless the Harts	Comedy	9/29/2019	FOX	The Simposons
13	Bob Hearts Abishola	Comedy	9/23/2019	CBS	The Neighborhood
14	Call Me Kat	Comedy	1/3/2021	FOX	The OT
15	Carol’s Second Act	Comedy	9/26/2019	CBS	Mom
16	Home Economics	Comedy	4/7/2021	ABC	The Goldbergs
17	I Feel Bad	Comedy	10/4/2018	NBC	Will & Grace
18	Kenan	Comedy	2/16/2021	NBC	Young Rock
19	Making It	Comedy	7/31/2018	NBC	America's Got Talent
20	Mixed-ish	Comedy	9/24/2019	ABC	Bless This Mess
21	Murphy Brown	Comedy	9/27/2018	CBS	Mom
22	Perfect Harmony	Comedy	9/26/2019	NBC	Superstore
23	Rel	Comedy	9/30/2018	FOX	Family Guy
24	Sunnyside	Comedy	9/26/2019	NBC	The Good Place
25	The Alec Baldwin Show	Comedy	10/14/2018	ABC	Shark Tank
26	The Cool Kids	Comedy	9/28/2018	FOX	Last Man Standing
27	United States of Al	Comedy	4/1/2021	CBS	Young Sheldon
28	A Million Little Things	Drama	9/26/2018	ABC	Single Parents
29	All Rise	Drama	9/23/2019	CBS	Bob Hearts Abishola
30	Almost Family	Drama	10/2/2019	FOX	The Masked Singer
31	BH90210	Drama	8/7/2019	FOX	MasterChef
32	Big Sky	Drama	11/17/2020	ABC	The Bachelorette
33	Bluff City Law	Drama	9/23/2019	NBC	The Voice
34	Charmed	Drama	10/14/2018	The CW	Supergirl
35	Clarice	Drama	2/11/2021	CBS	The Unicorn
36	Debris	Drama	3/1/2021	NBC	The Voice
37	Emergence	Drama	9/24/2019	ABC	Black-ish
38	Fantasy Island	Drama	8/10/2021	FOX	Lego Masters
39	God Friended Me	Drama	9/30/2018	CBS	60 Minutes

40	Grand Hotel	Drama	6/17/2019	ABC	The Bachelorette
41	Legacies	Drama	10/25/2018	The CW	Supernatural
42	Magnum P.I.	Drama	9/24/2018	CBS	Young Sheldon
43	Nancy Drew	Drama	10/9/2019	The CW	Riverdale
44	New Amsterdam	Drama	9/25/2018	NBC	This is Us
45	Nurses	Drama	12/7/2020	NBC	The Voice
46	Prodigal Son	Drama	9/23/2019	FOX	911
47	Single Parents	Drama	9/26/2018	ABC	Modern Family
48	Stumptown	Drama	9/25/2019	ABC	Single Parents
49	The Kids Are Alright	Drama	10/16/2018	ABC	The Conners
50	The Republic of Sarah	Drama	6/14/2021	The CW	All American
51	The Rookie	Drama	10/16/2018	ABC	Splitting Up Together
52	The Unicorn	Drama	9/26/2019	CBS	Young Sheldon

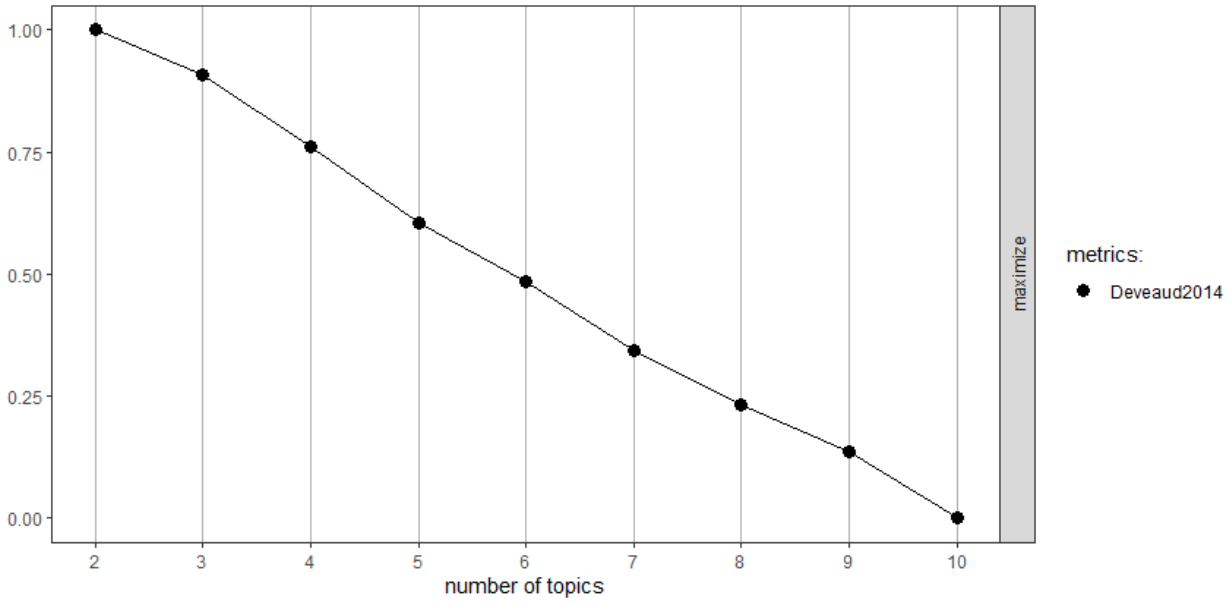
APPENDIX II: LDA RESULTS FOR EACH SHOW

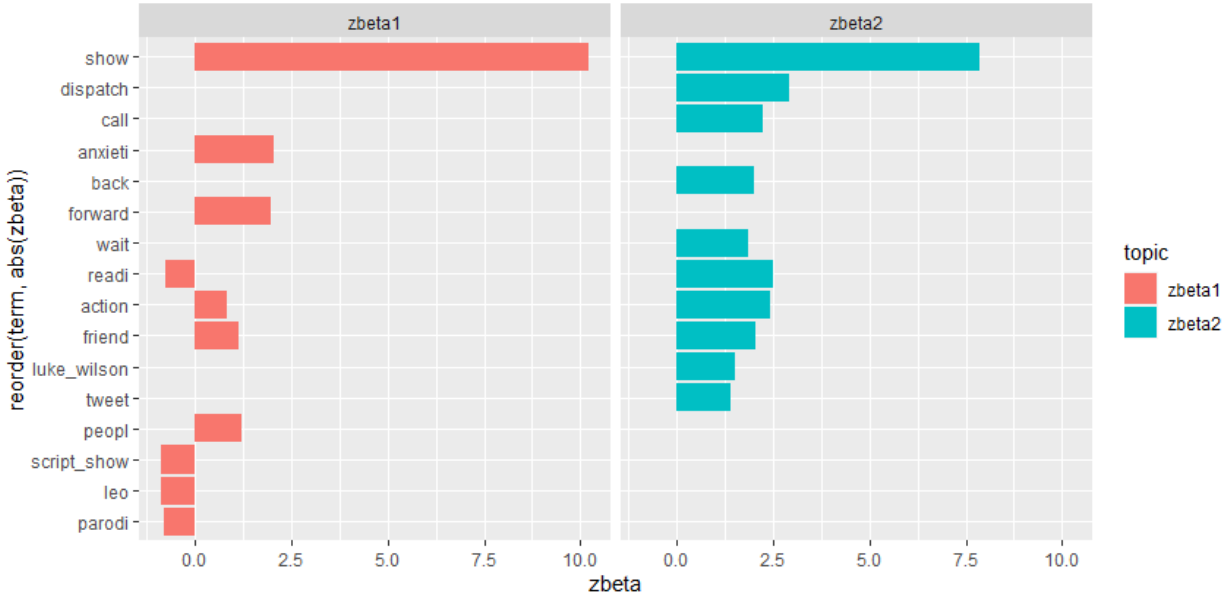
Castaways



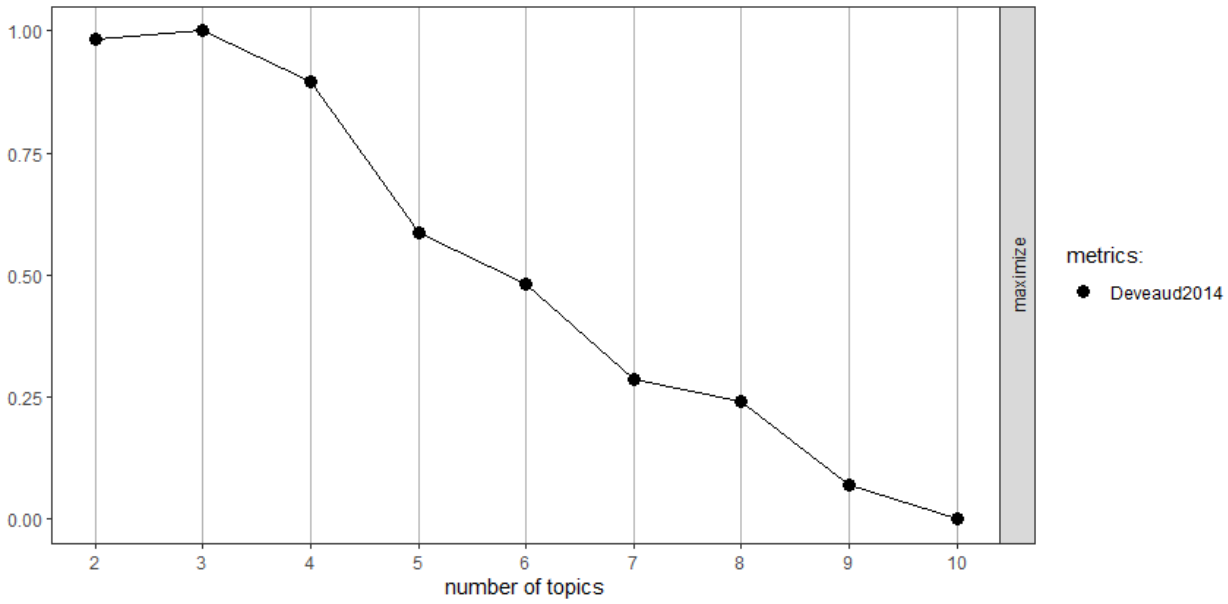


Emergency Call

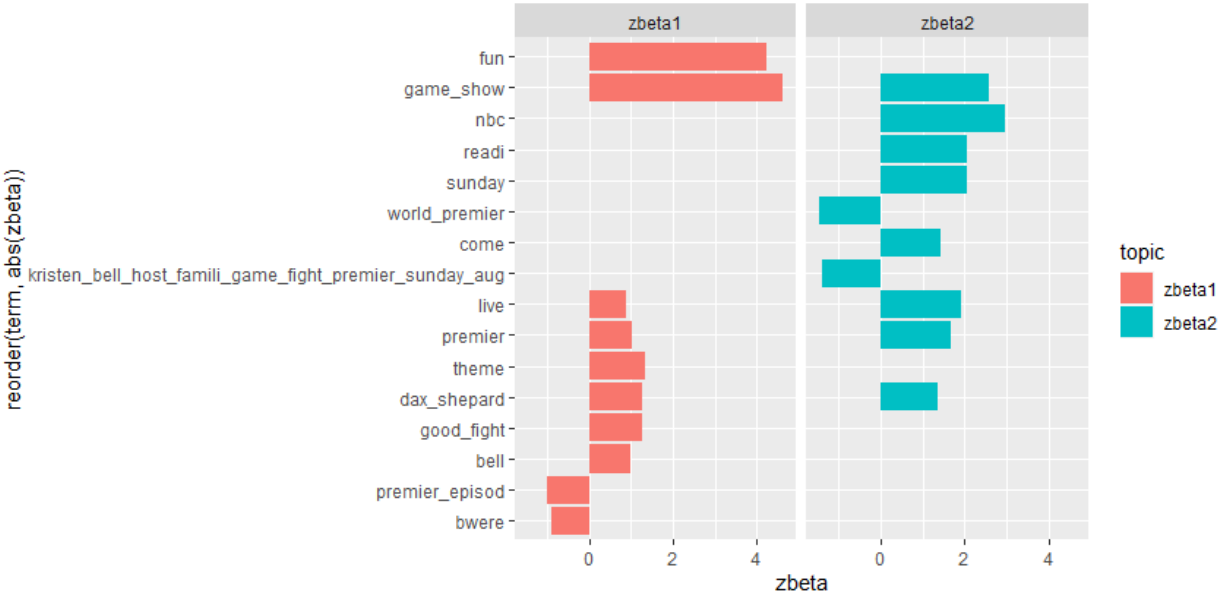




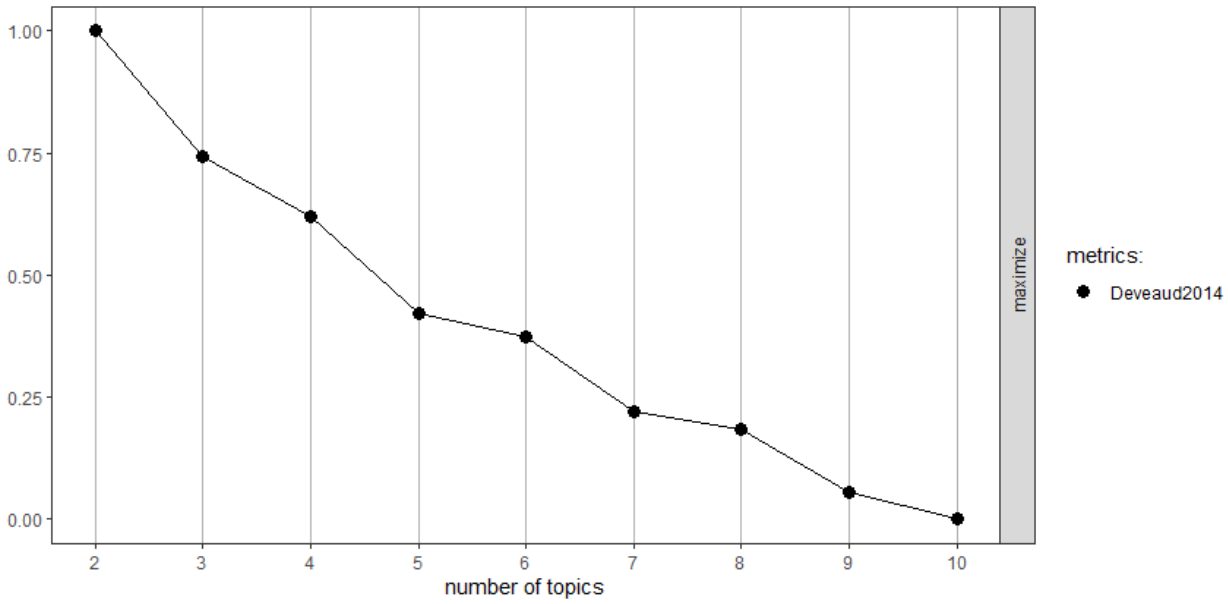
Family Game Fight!

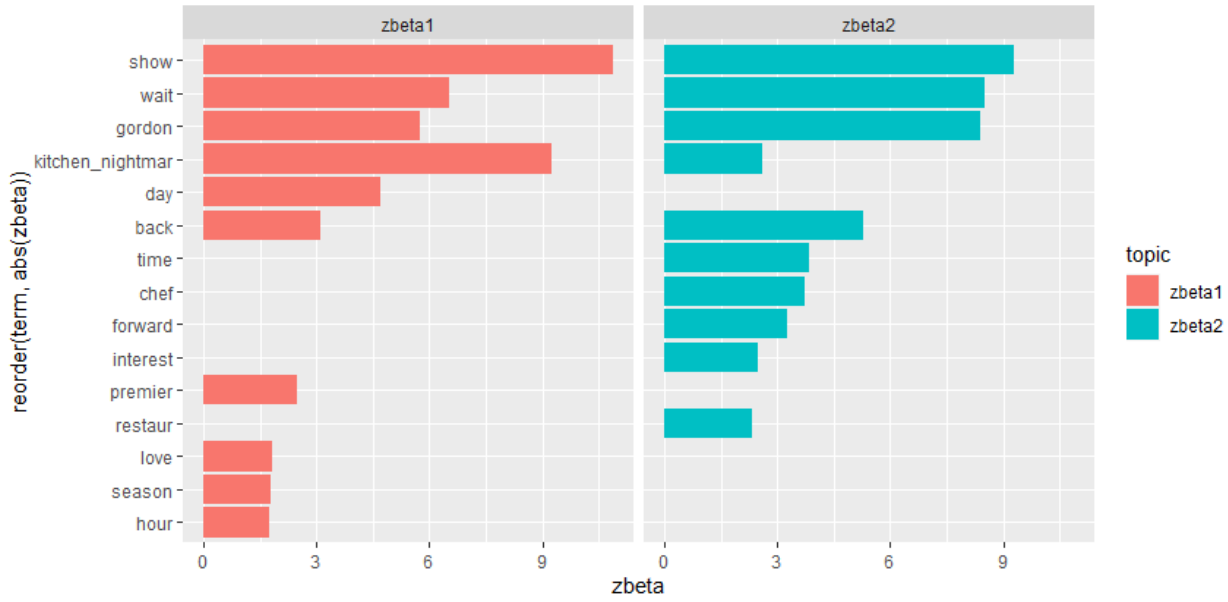


NOTE Sample size was too small for 3 topics (under 50)

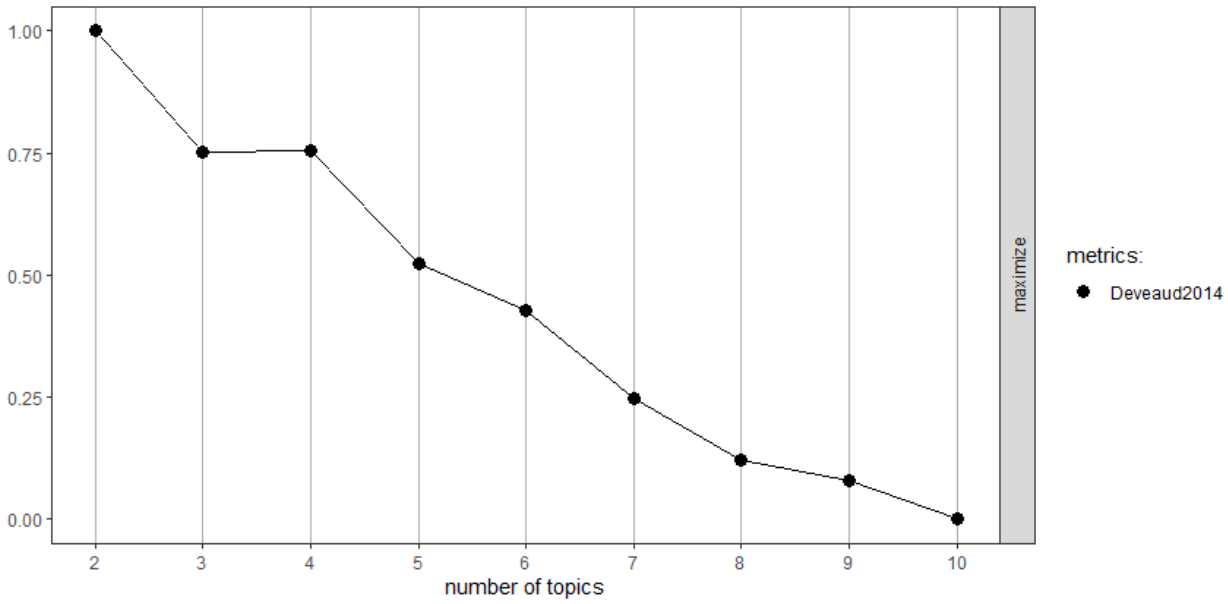


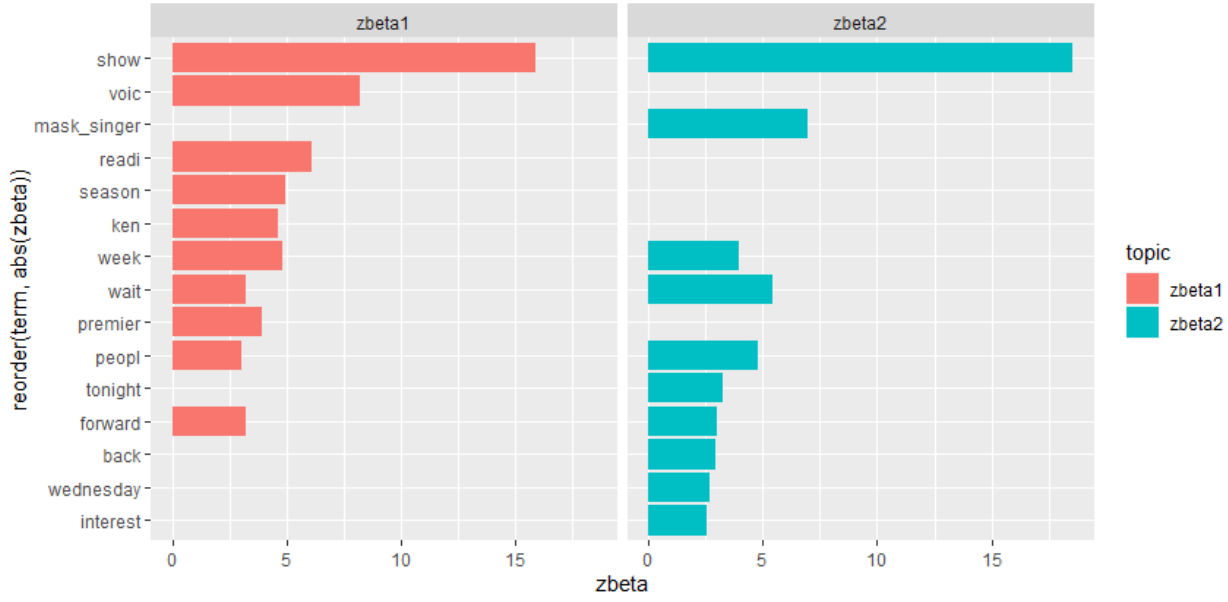
Gordon Ramsay's 24 Hours to Hell & Back



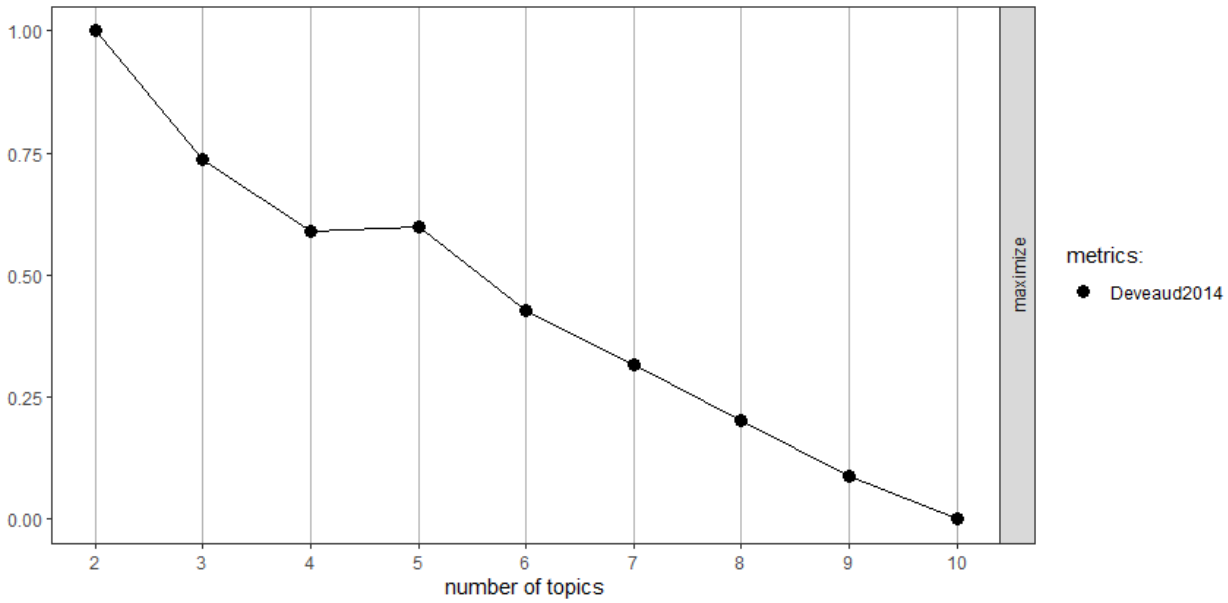


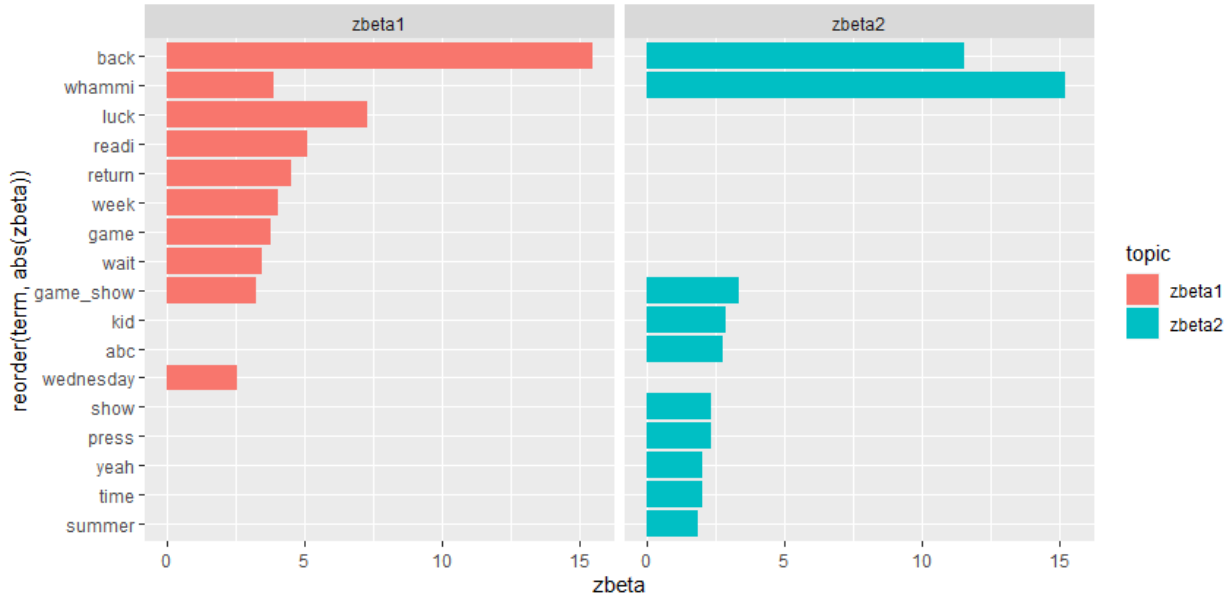
I Can See Your Voice



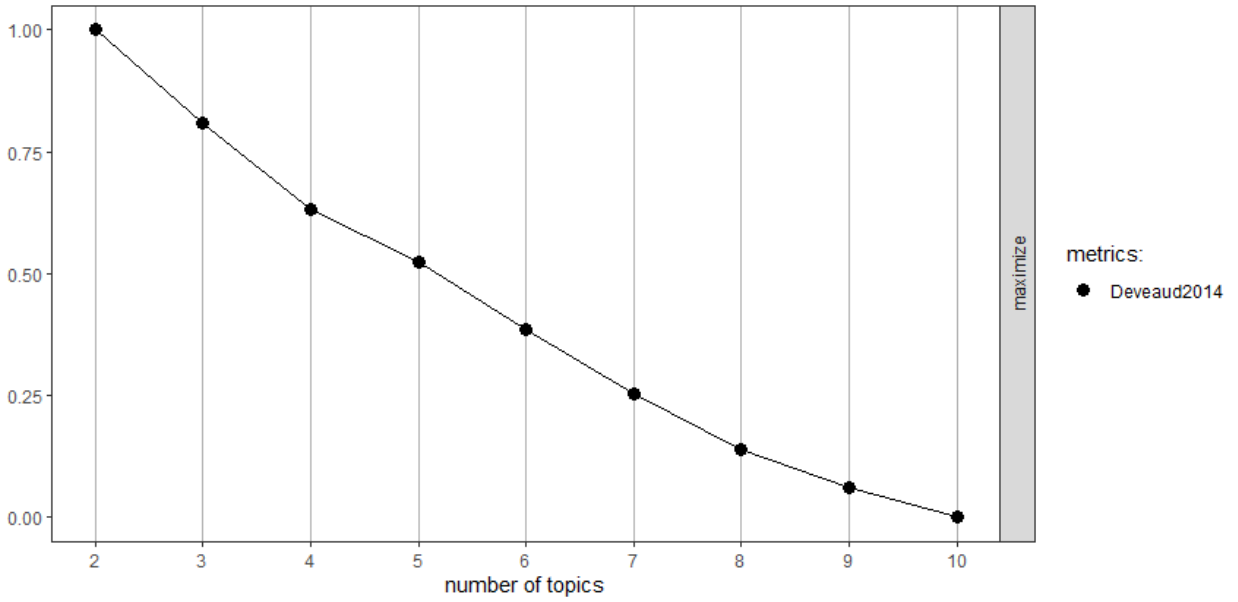


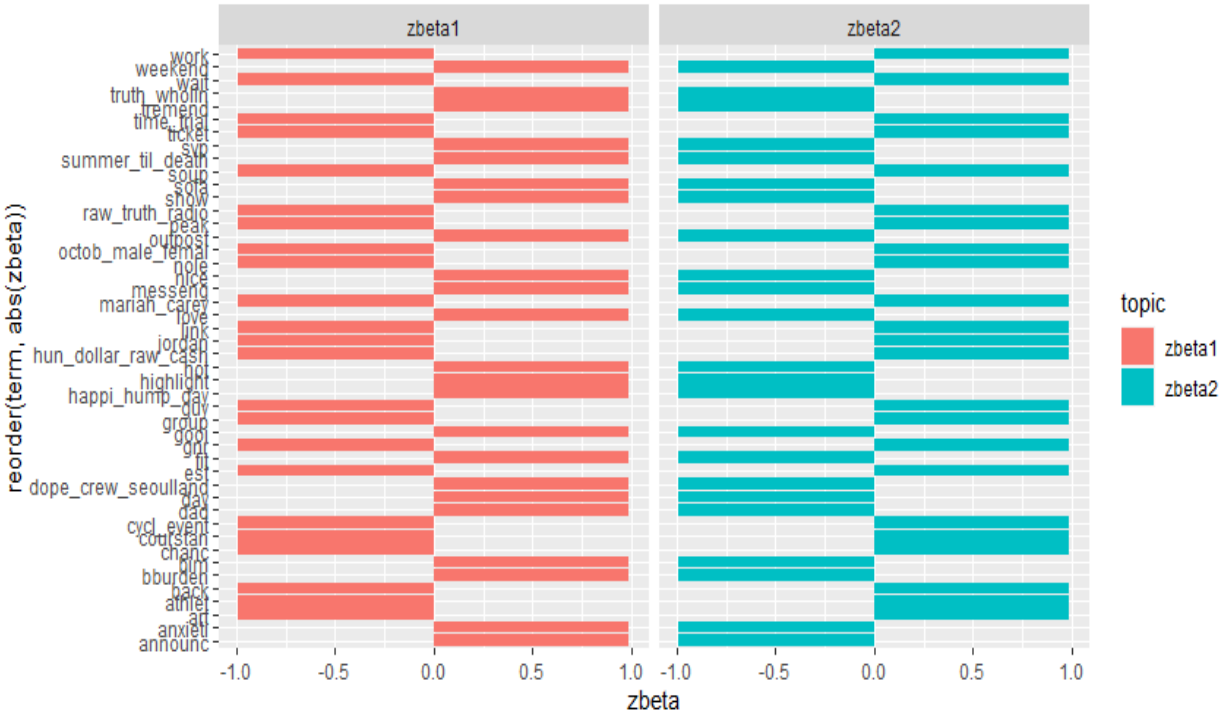
Press Your Luck





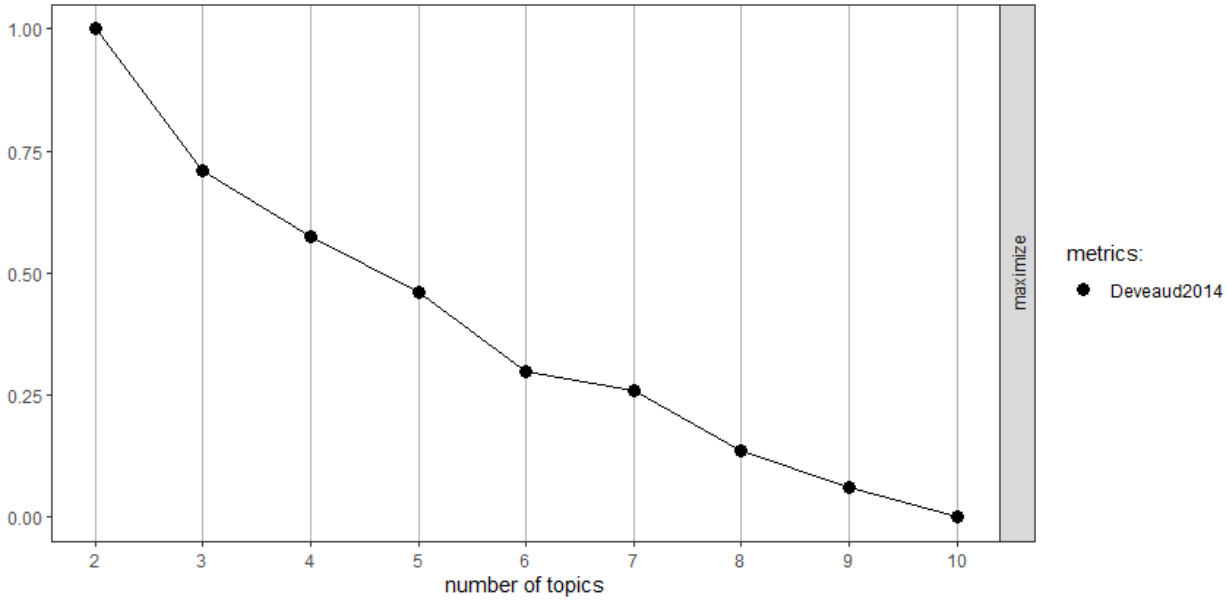
Red Bull Peaking

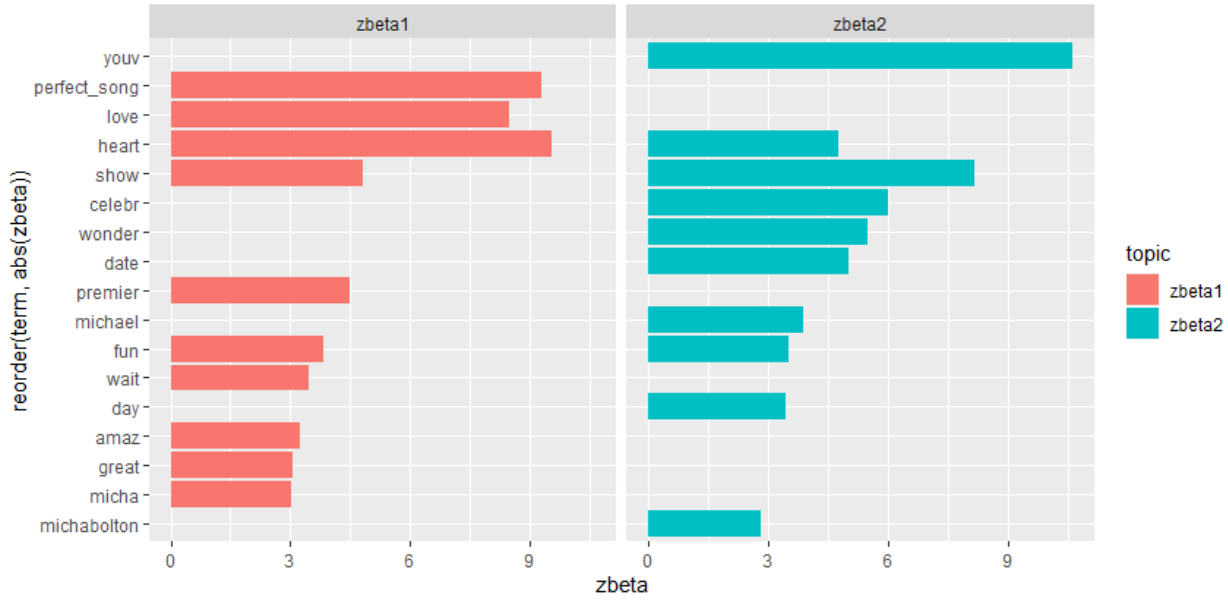




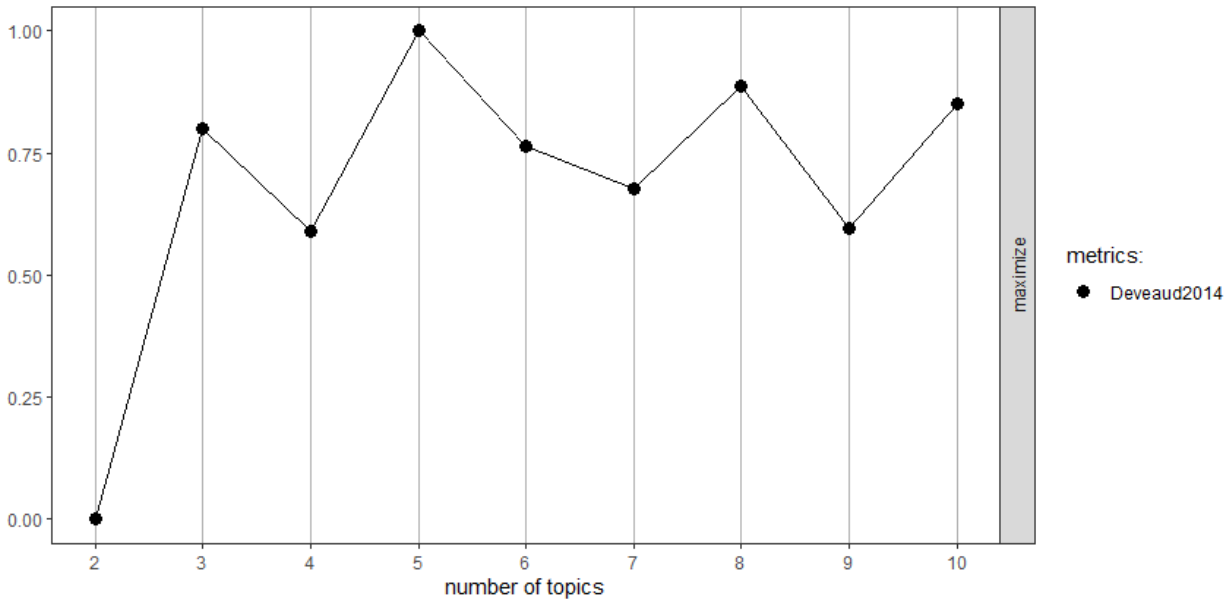
NOTE: Small Sample Size for 2 topics (under 50)

The Celebrity Dating Game

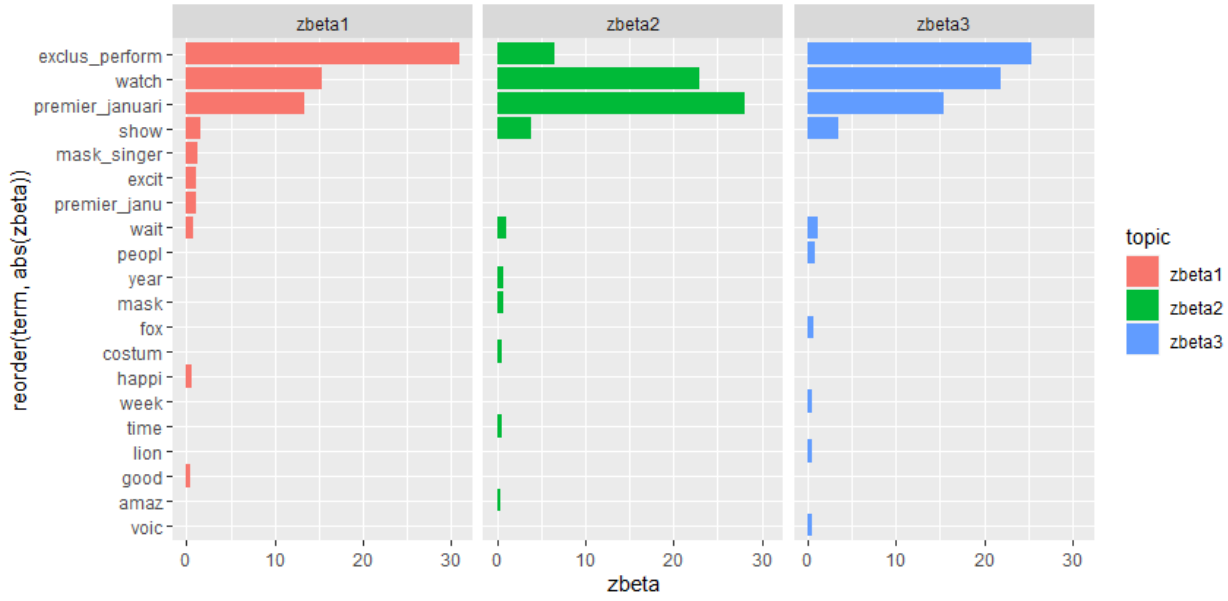




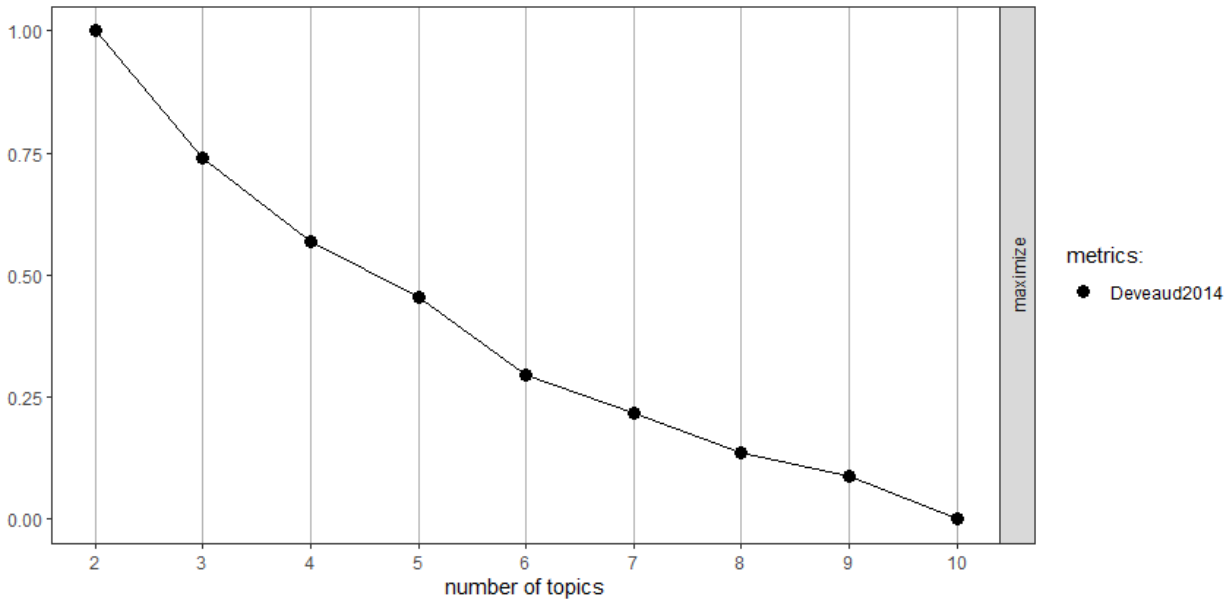
The Masked Singer

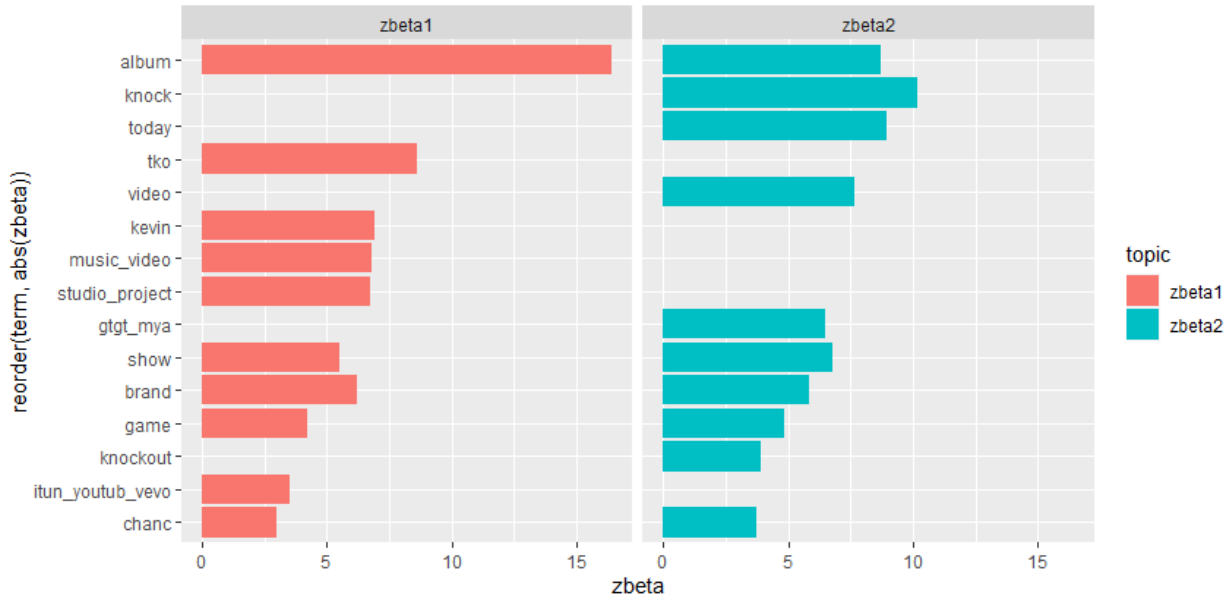


NOTE: Small Sample Size for 5 topics or more (under 50)

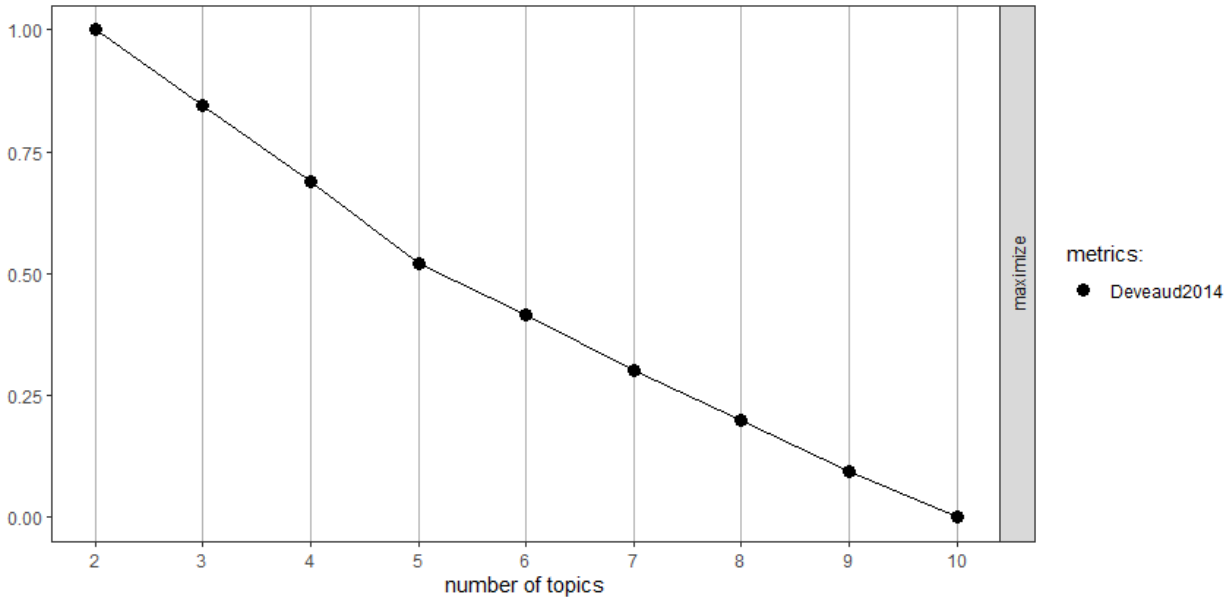


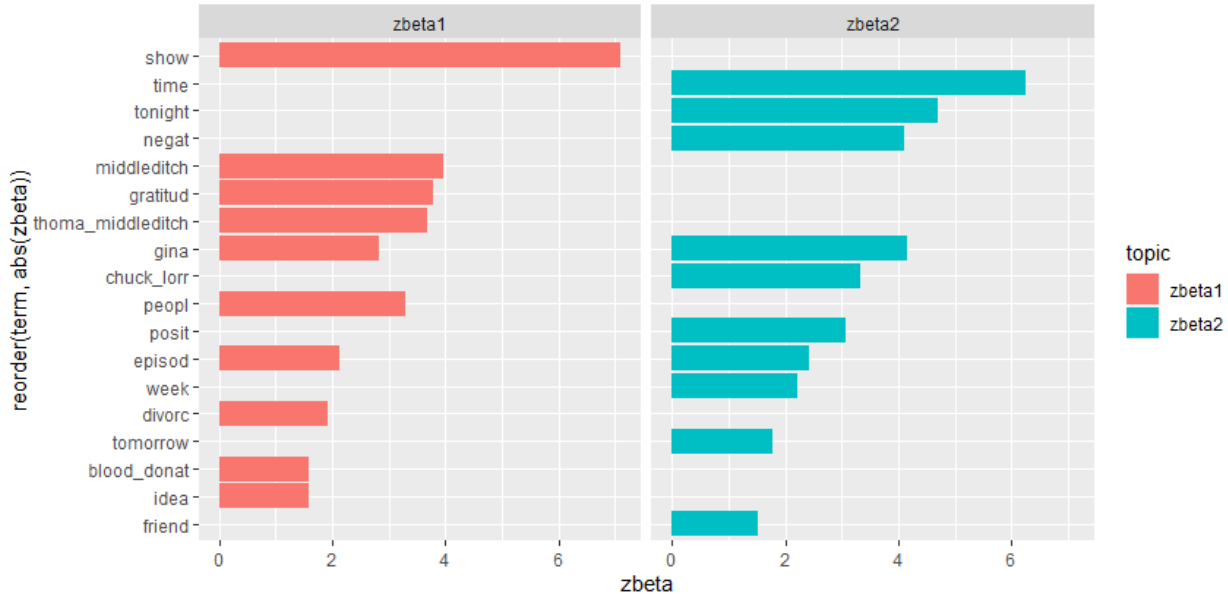
TKO: Total Knock Out



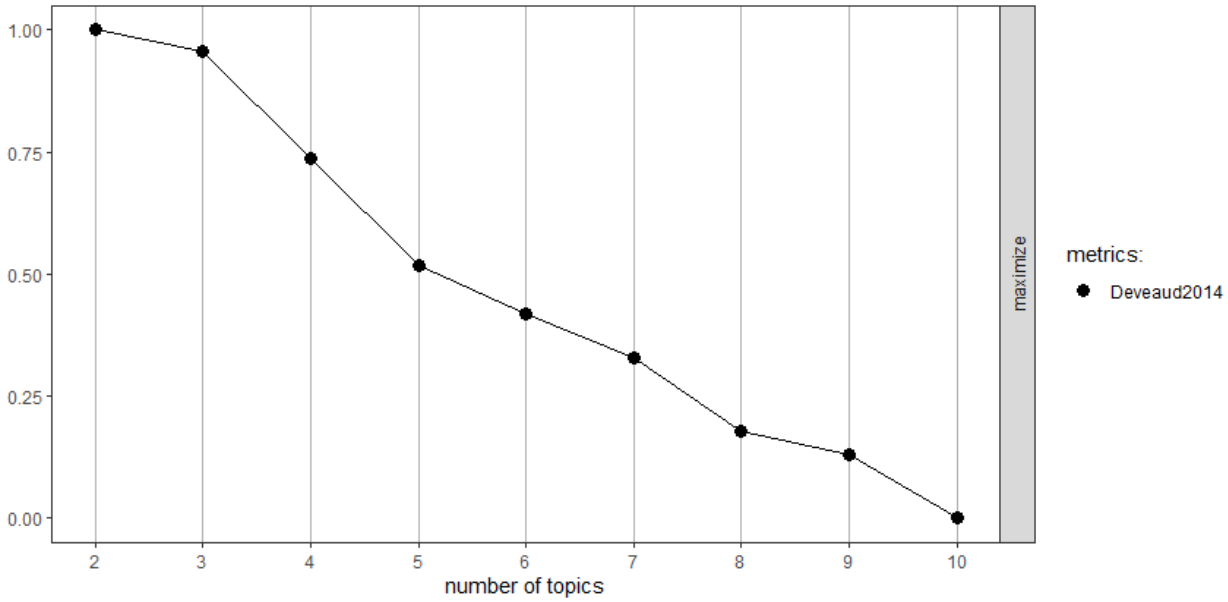


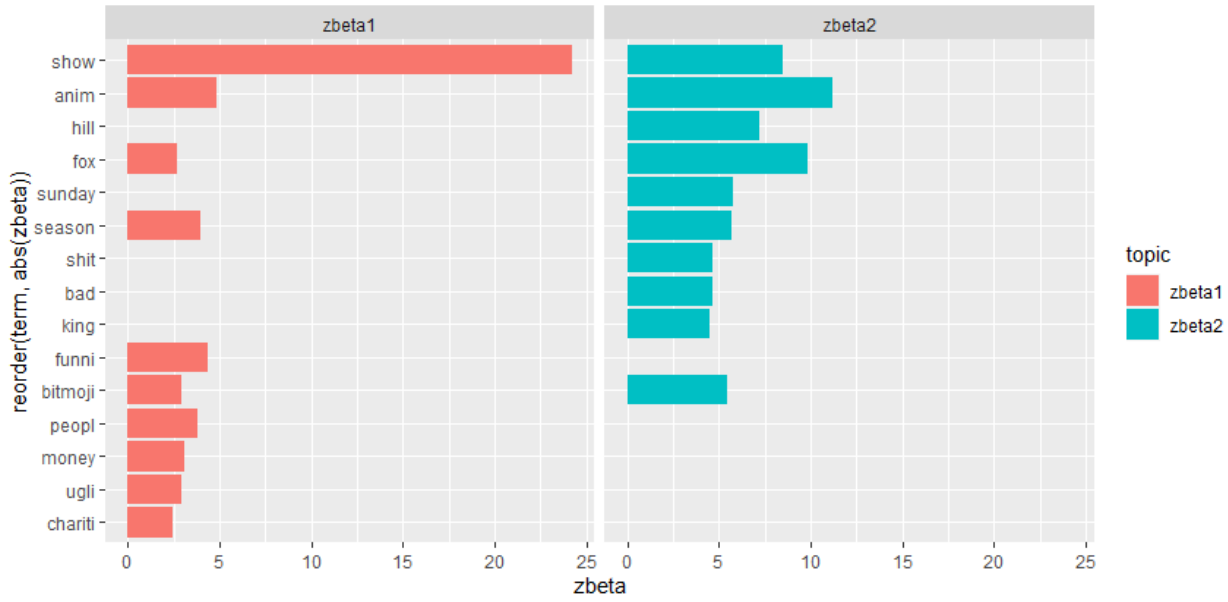
B Positive


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

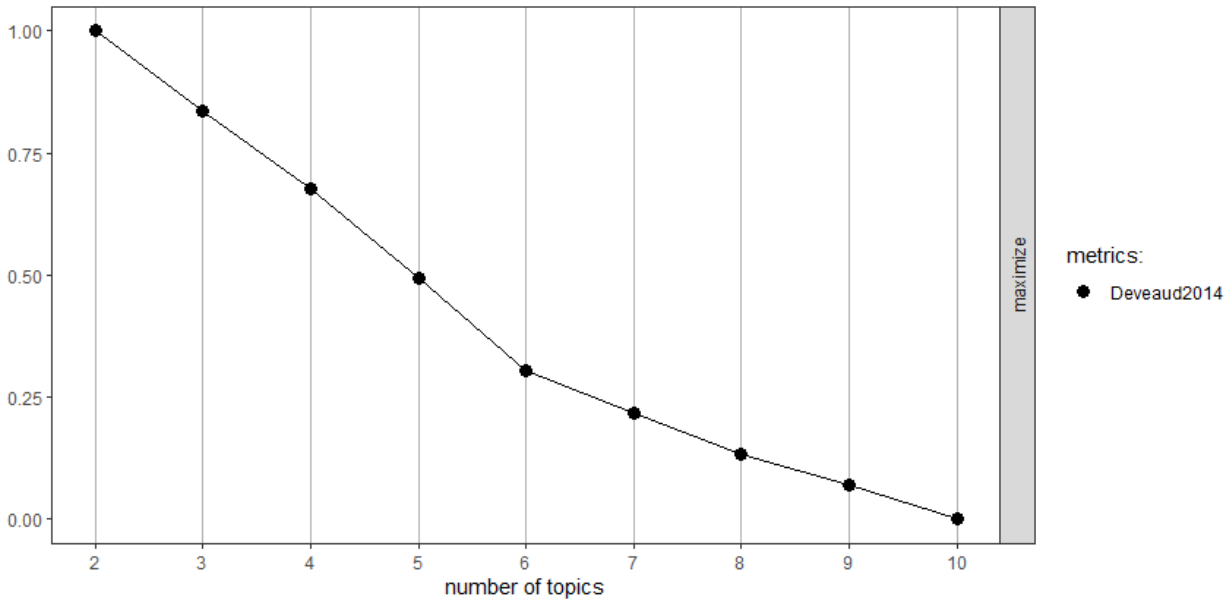

zbeta1
zbeta2
show
time
tonight
negat
middleditch
gratitud
thoma_middleditch
gina
chuck_lorr
peopl
posit
episod
week
divorc
tomorrow
blood_donat
idea
friend
reorder(term, abs(zbeta))
0
2
4
6
zbeta
topic
zbeta1
zbeta2

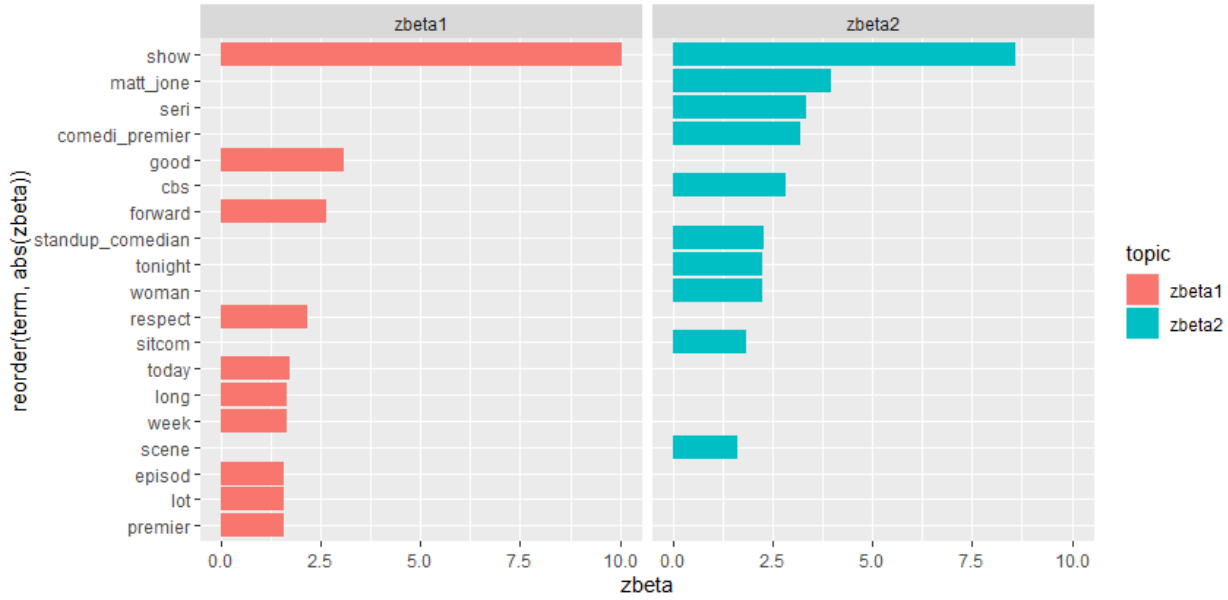
Bless the Harts



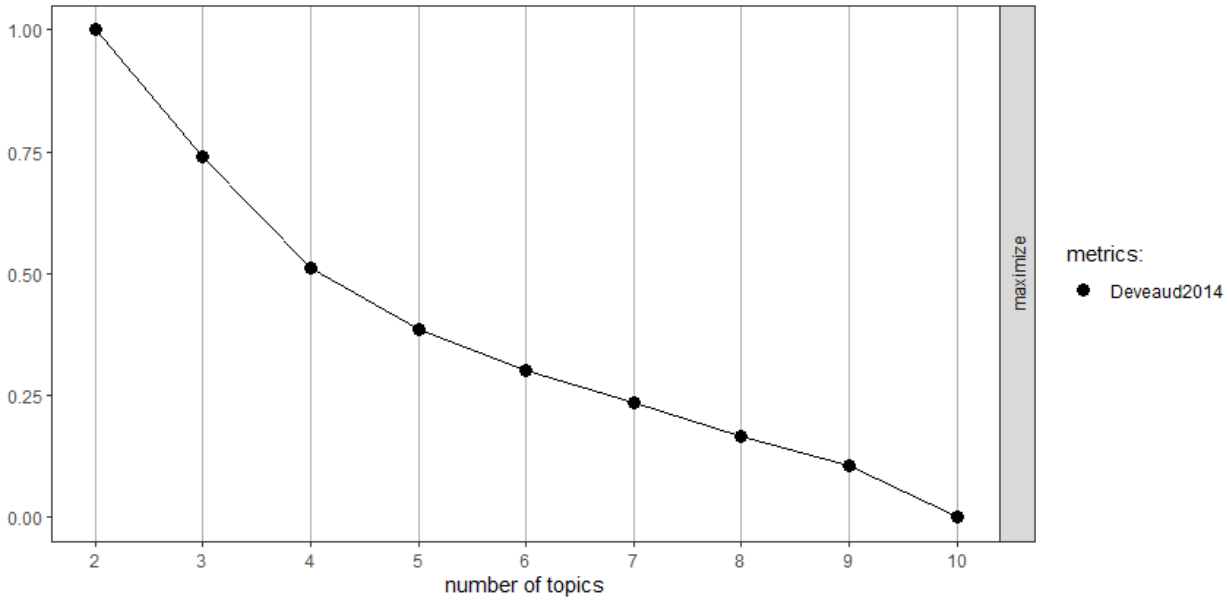


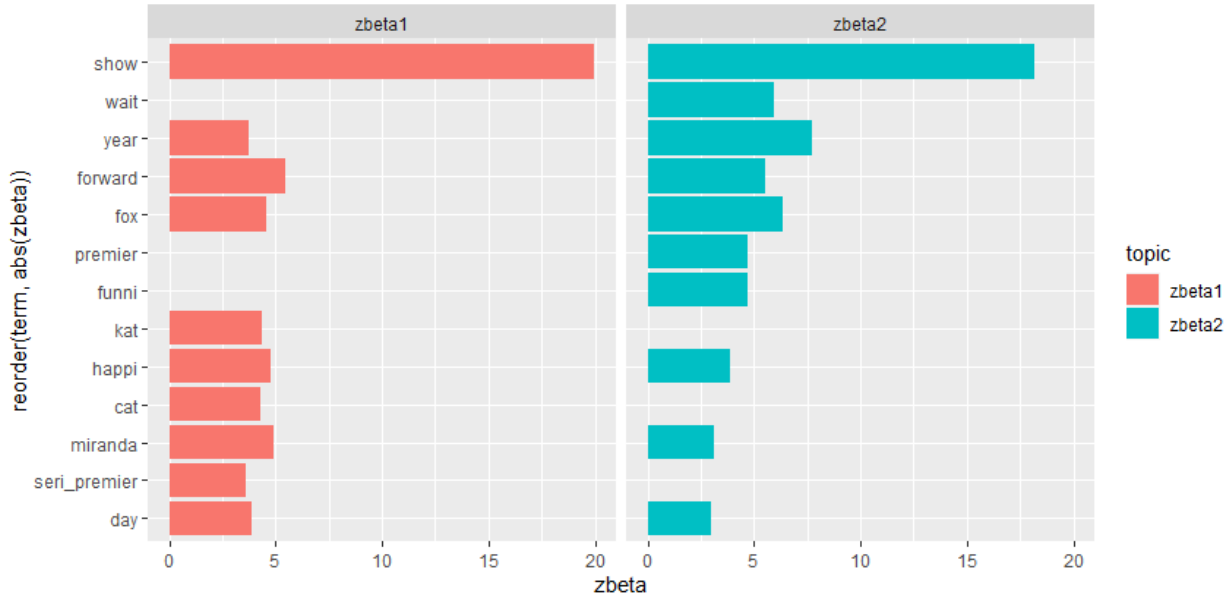
Bob Hearts Abishola



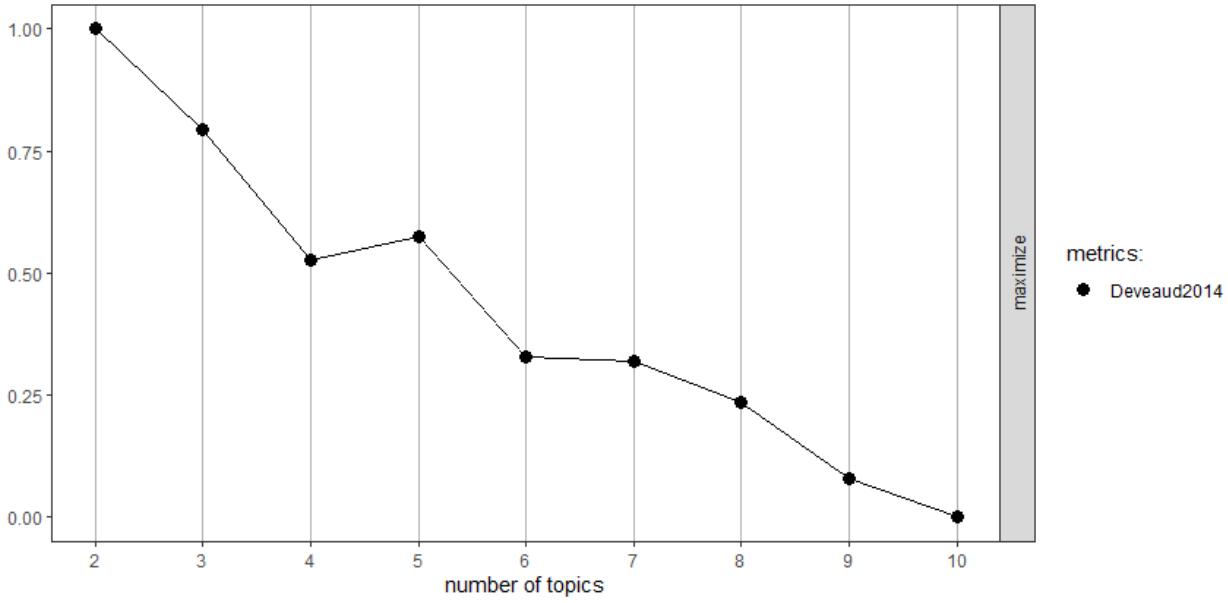


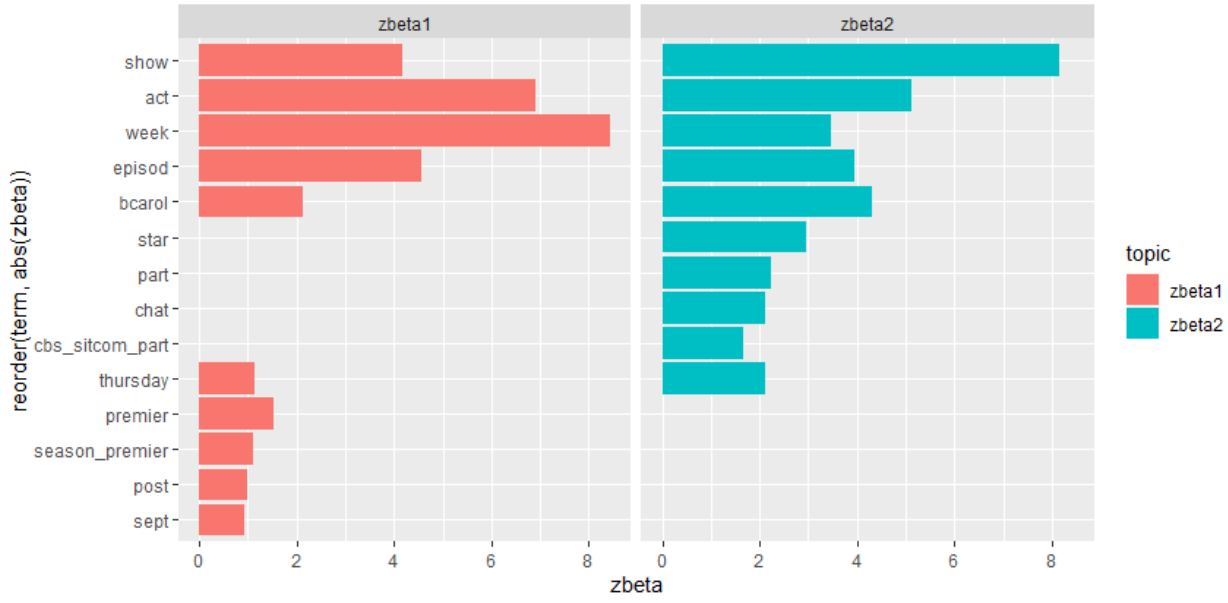
Call Me Kat


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

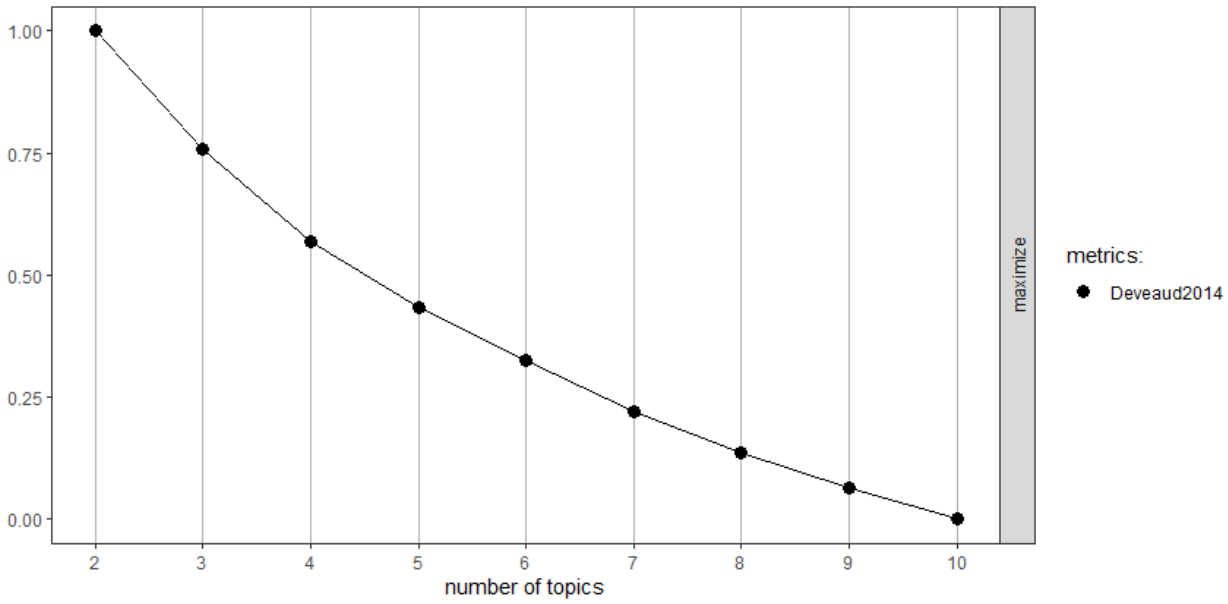

zbeta1
zbeta2
show
wait
year
forward
fox
premier
funni
kat
happi
cat
miranda
seri_premier
day
reorder(term, abs(zbeta))
0
5
10
15
20
zbeta
topic
zbeta1
zbeta2

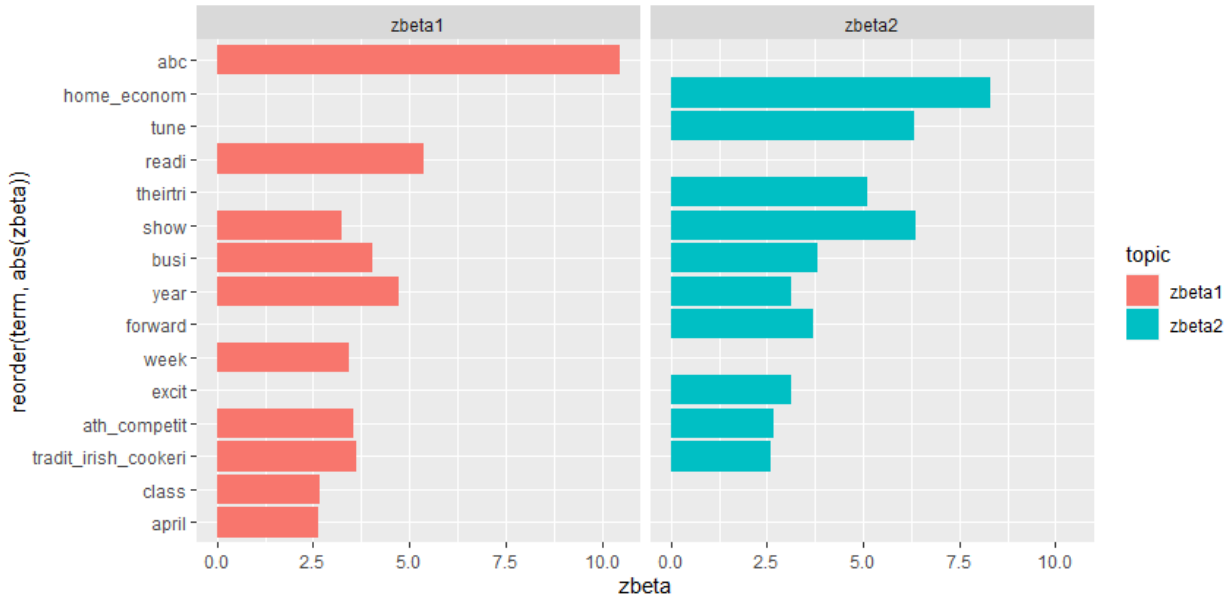
Carol's Second Act



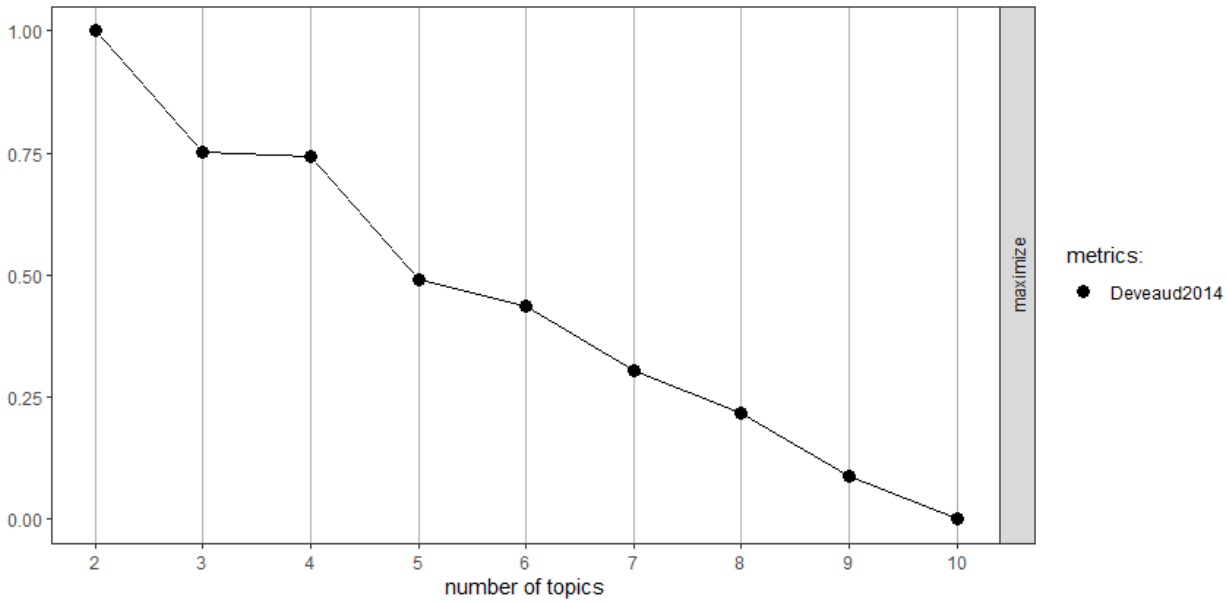


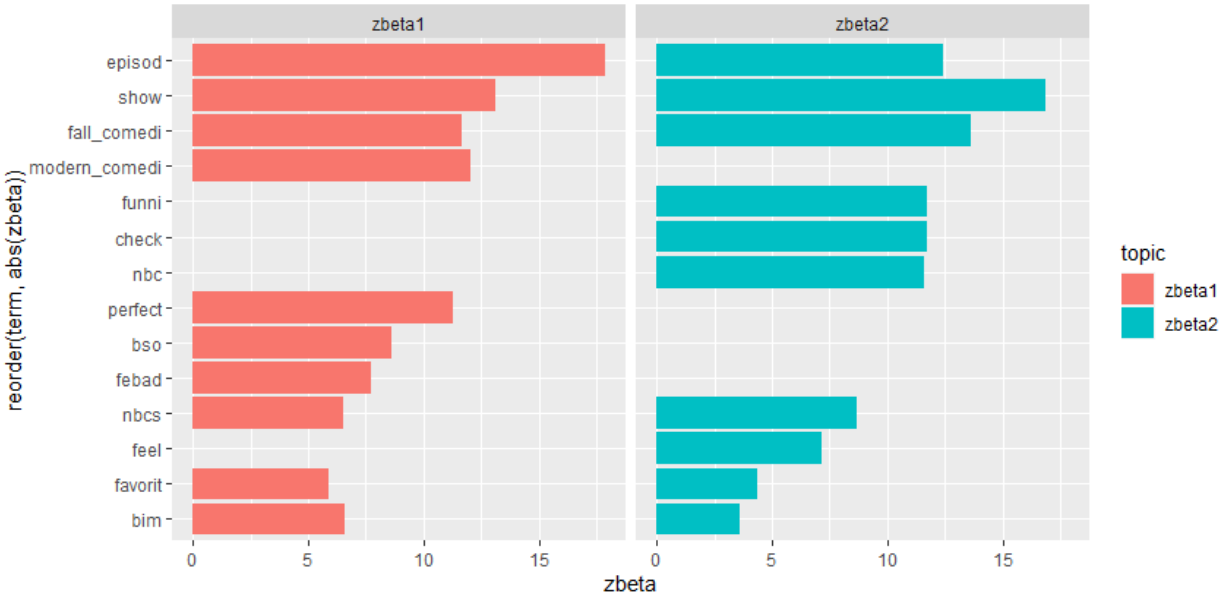
Home Economics



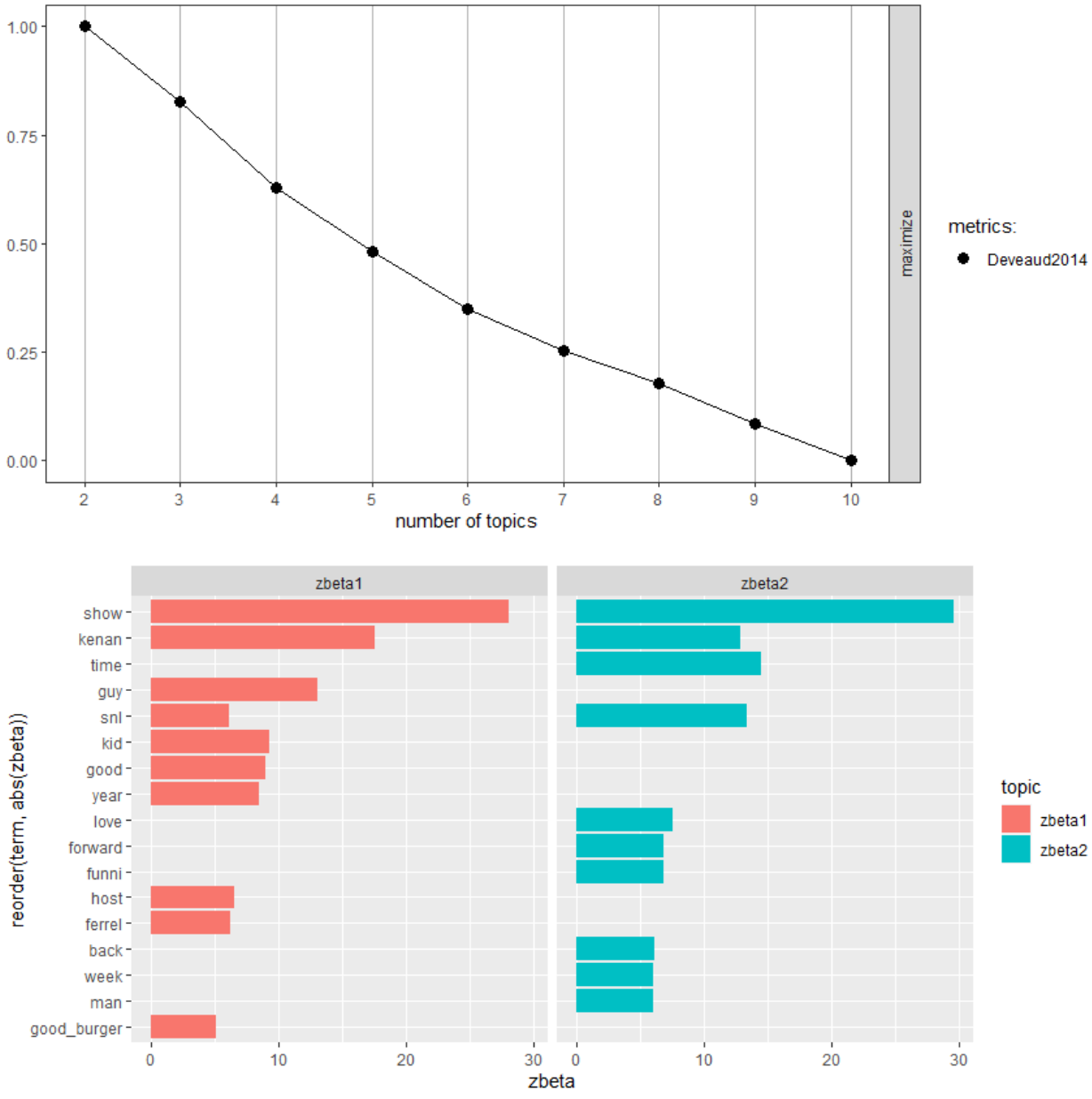


I Feel Bad

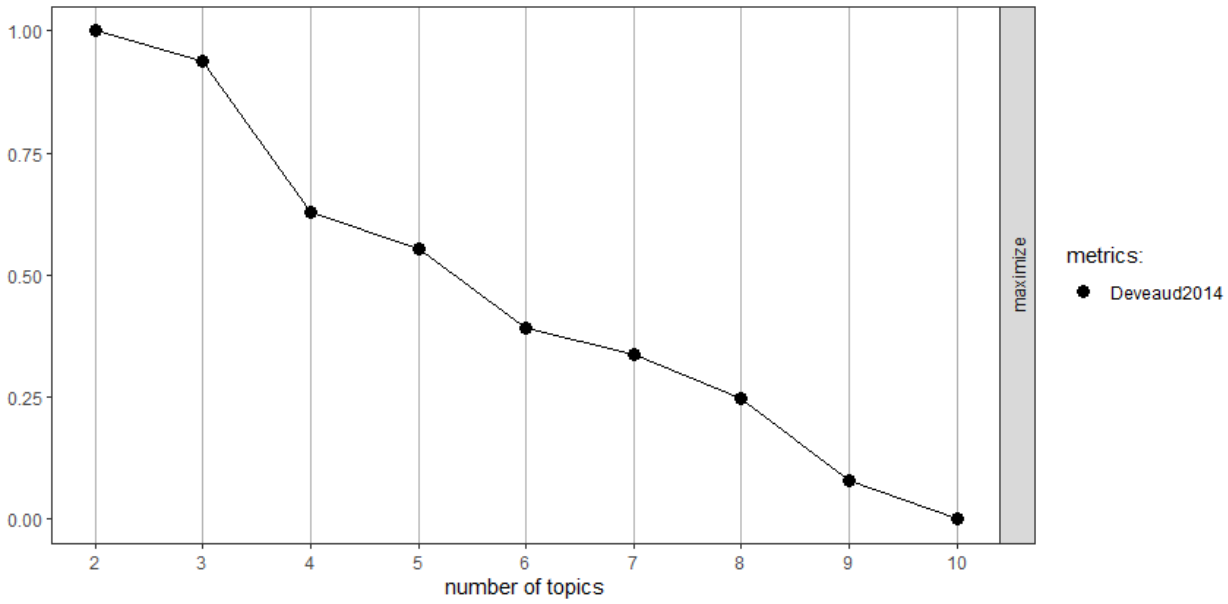

1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014


zbeta1
zbeta2
episod
show
fall_comedi
modern_comedi
funni
check
nbc
perfect
bso
febad
nbcs
feel
favorit
bim
reorder(term, abs(zbeta))
0
5
10
15
zbeta
topic
zbeta1
zbeta2

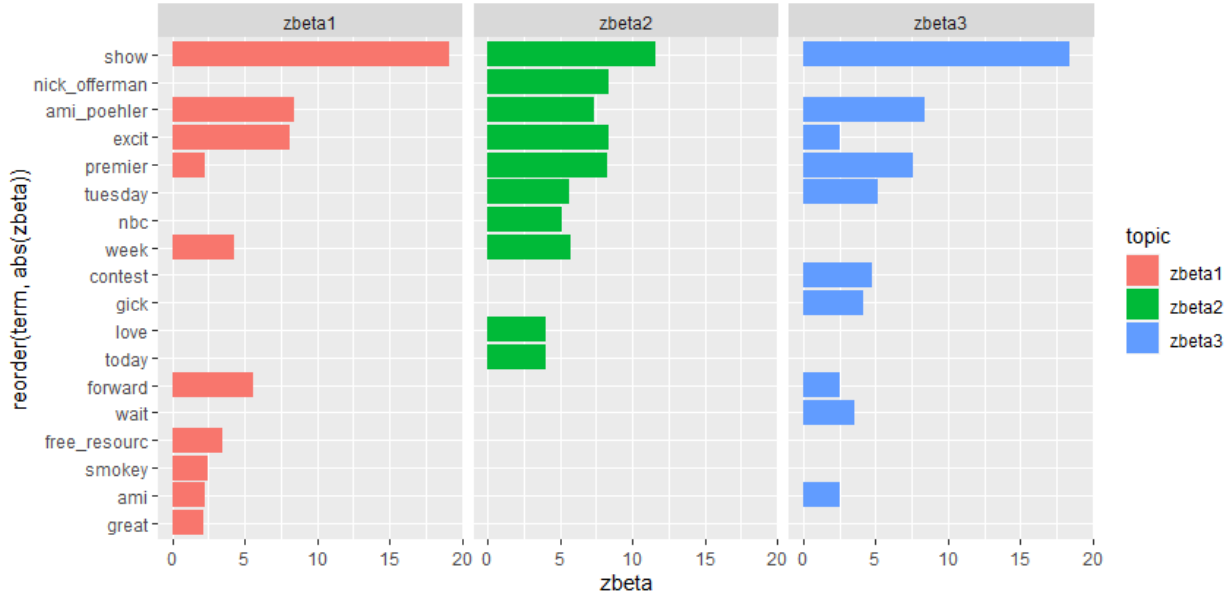
Kenan


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014
zbeta1
zbeta2
show
kenan
time
guy
snl
kid
good
year
love
forward
funni
host
ferrel
back
week
man
good_burger
reorder(term, abs(zbeta))
0
10
20
30
zbeta
topic
zbeta1
zbeta2

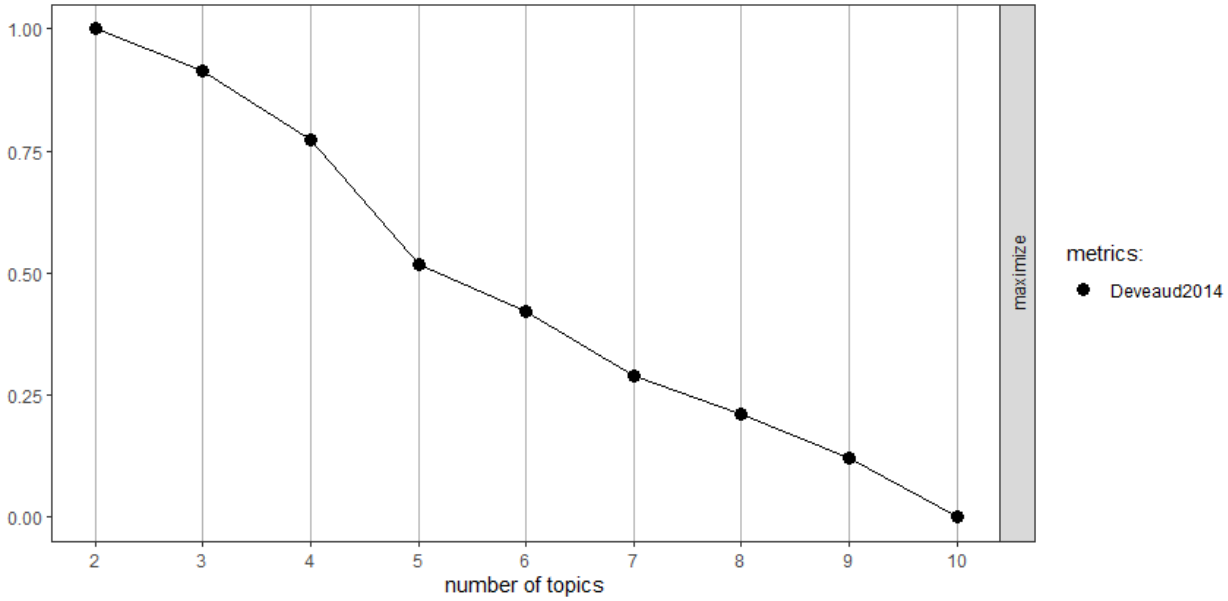
Making It

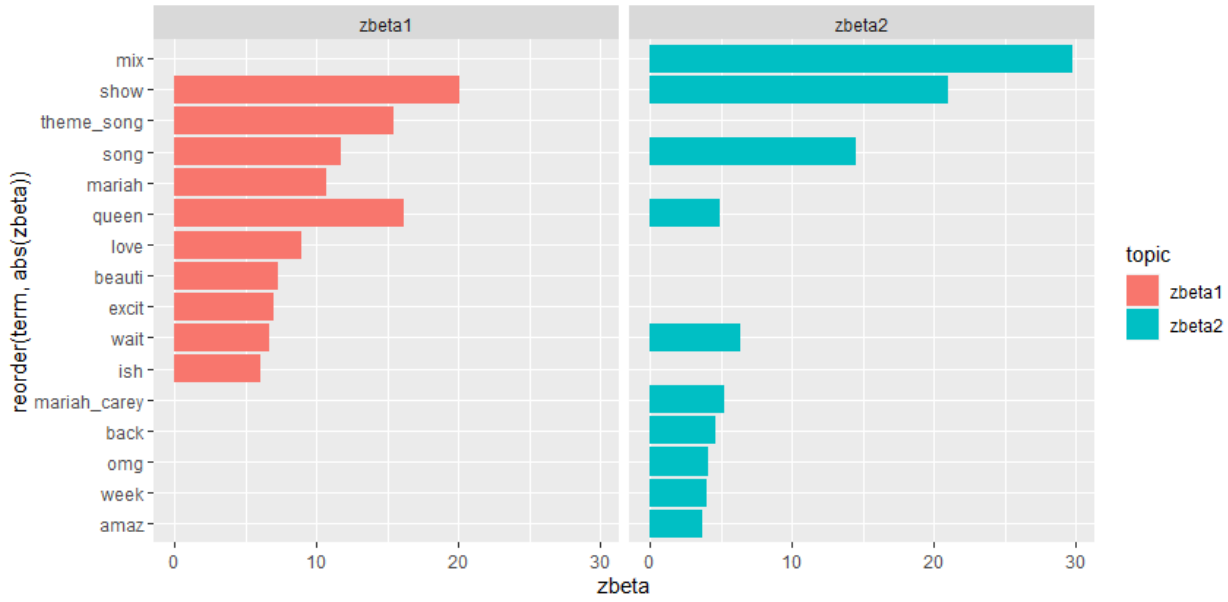


NOTE: 3-topic solution was more interpretable and sample size was sufficient.

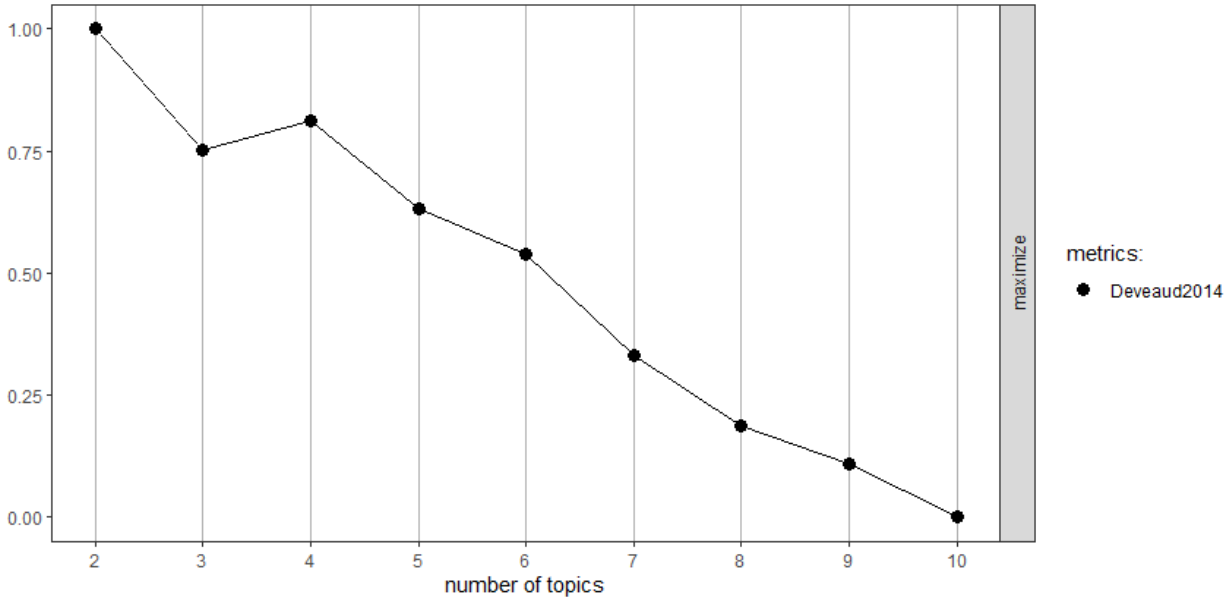


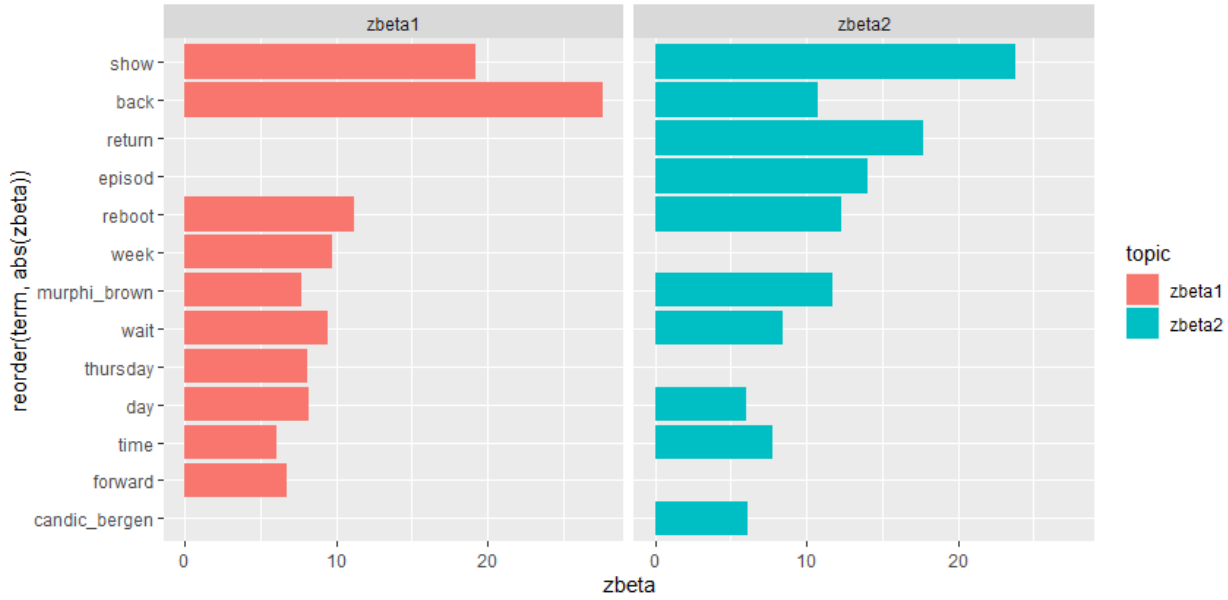
Mixed-ish


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

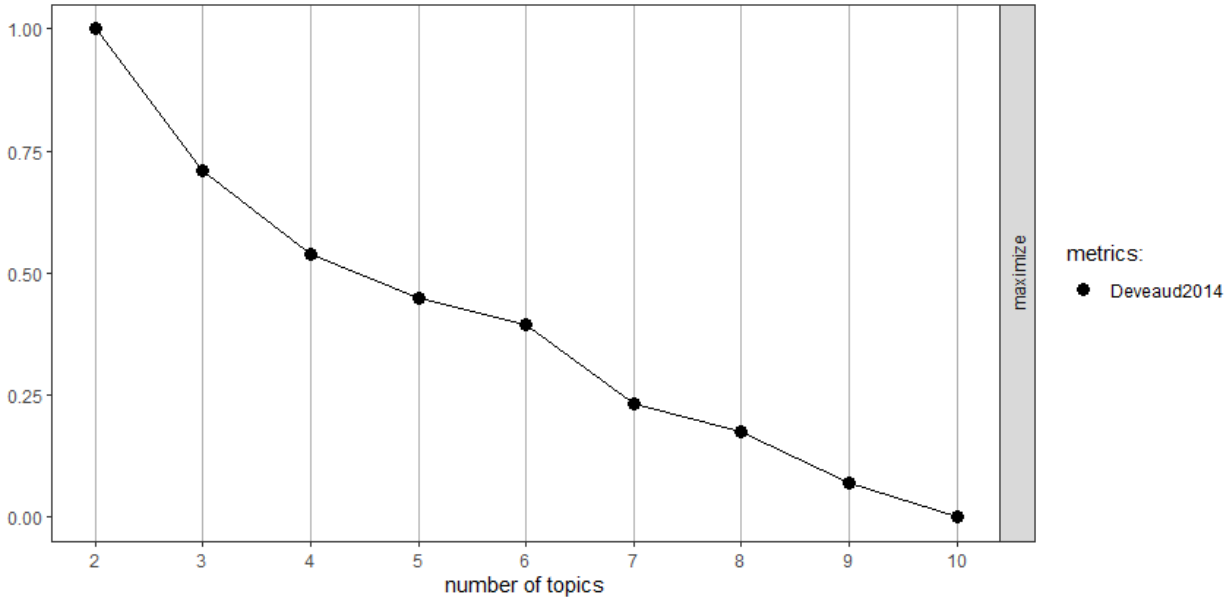

zbeta1
zbeta2
mix
show
theme_song
song
mariah
queen
love
beauti
excit
wait
ish
mariah_carey
back
omg
week
amaz
reorder(term, abs(zbeta))
0
10
20
30
zbeta
topic
zbeta1
zbeta2

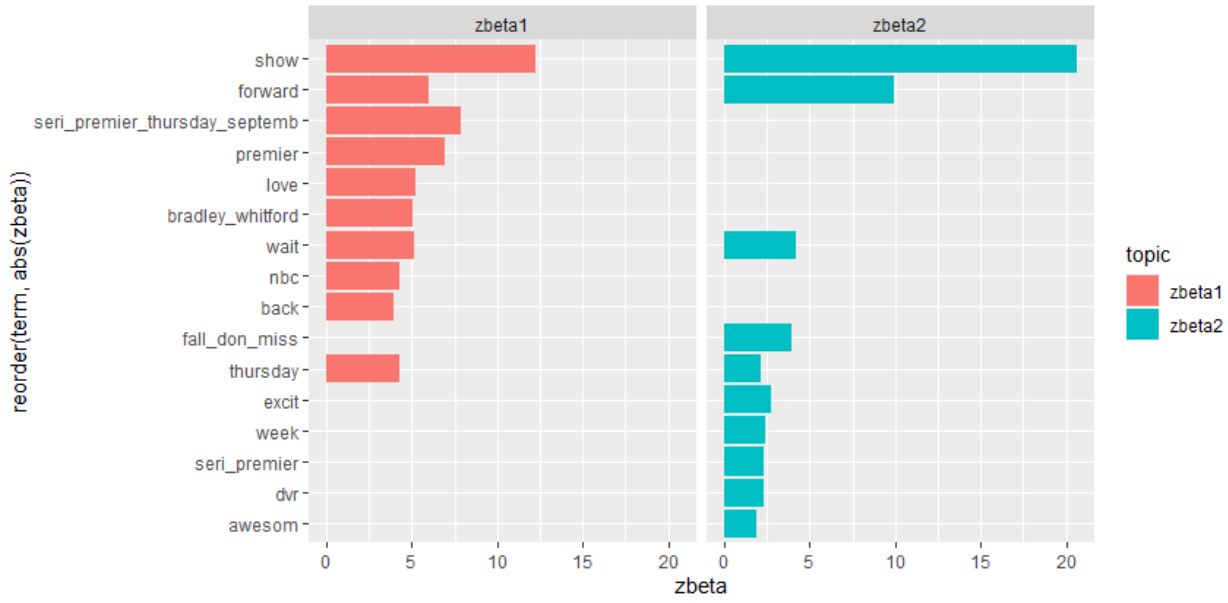
Murphy Brown



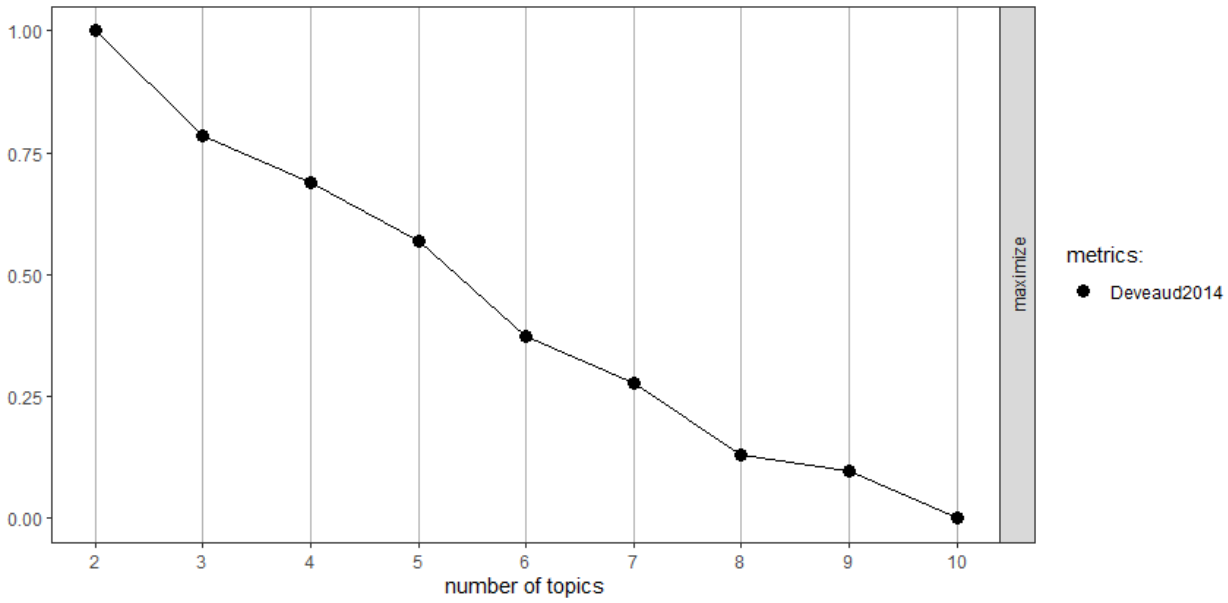


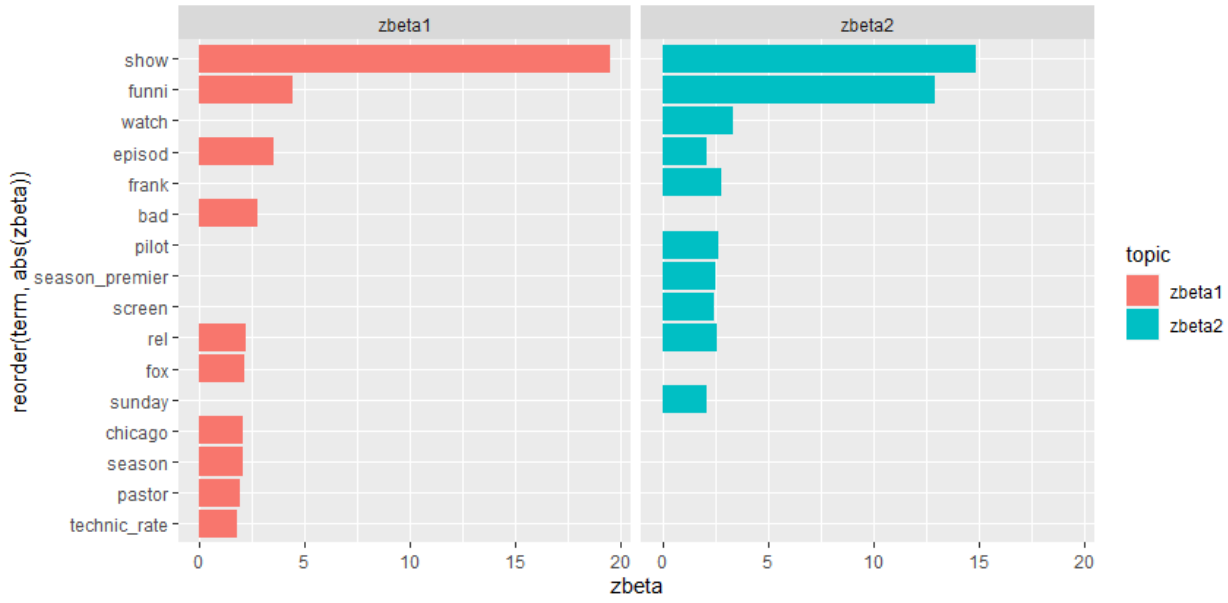
Perfect Harmony



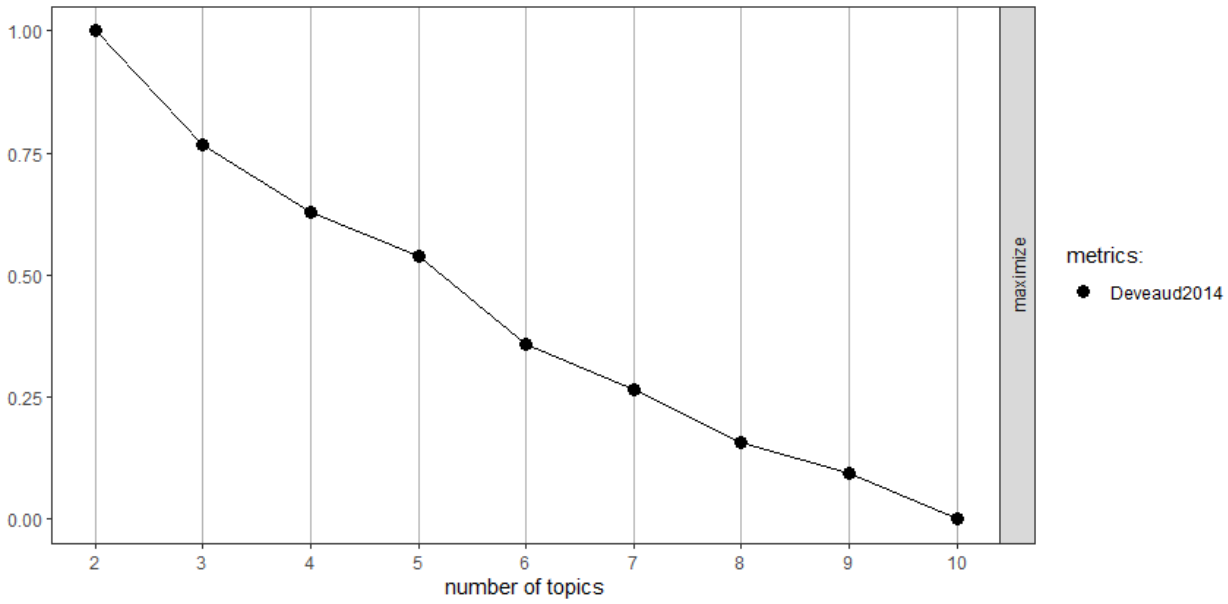


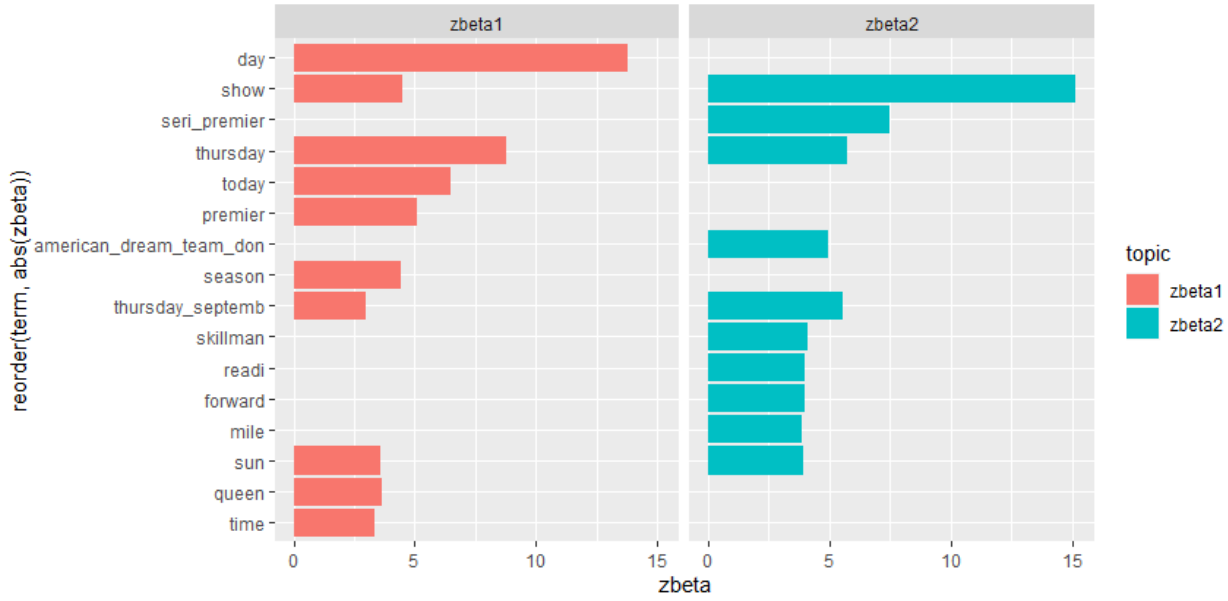
Rel


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

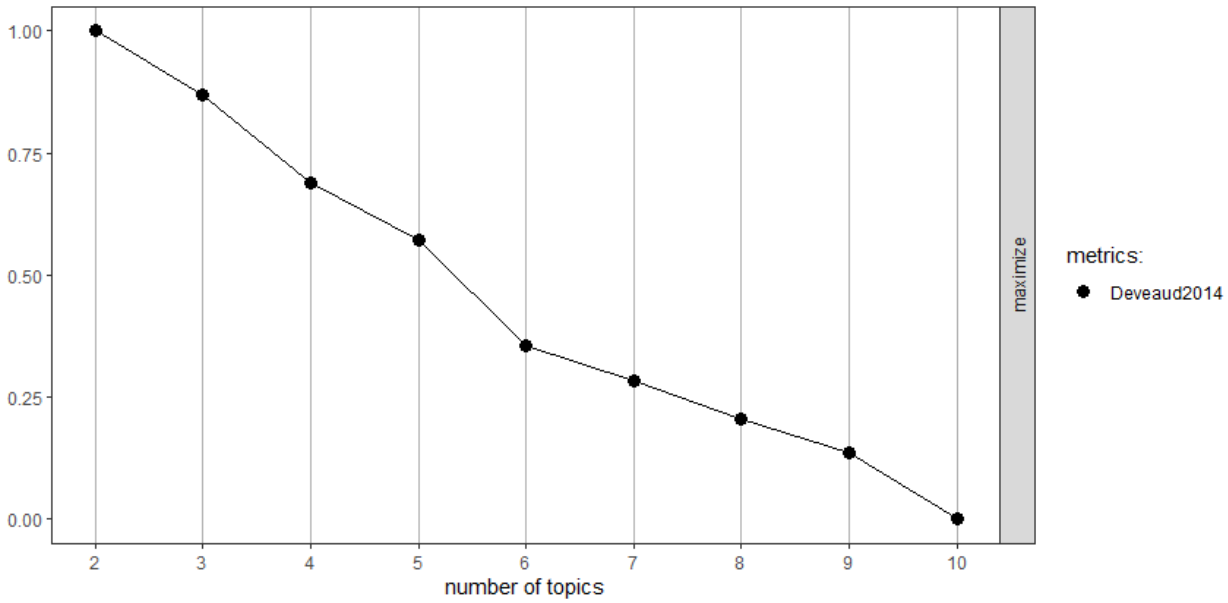

zbeta1
zbeta2
show
funni
watch
episod
frank
bad
pilot
season_premier
screen
rel
fox
sunday
chicago
season
pastor
technic_rate
reorder(term, abs(zbeta))
0
5
10
15
20
zbeta
topic
zbeta1
zbeta2

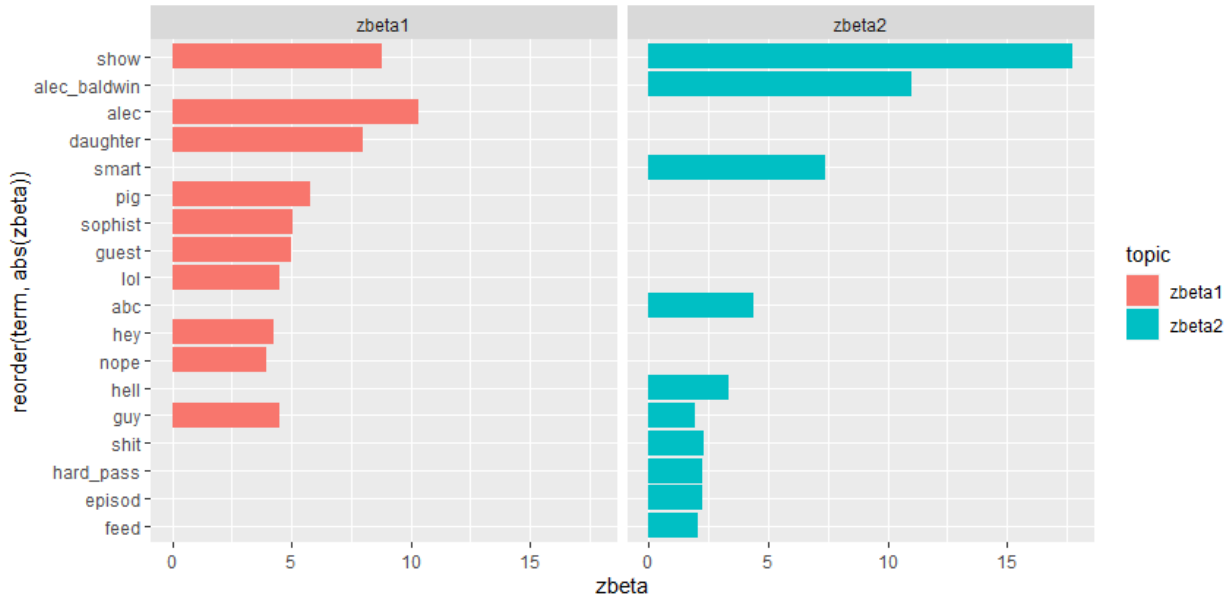
Sunnyside


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

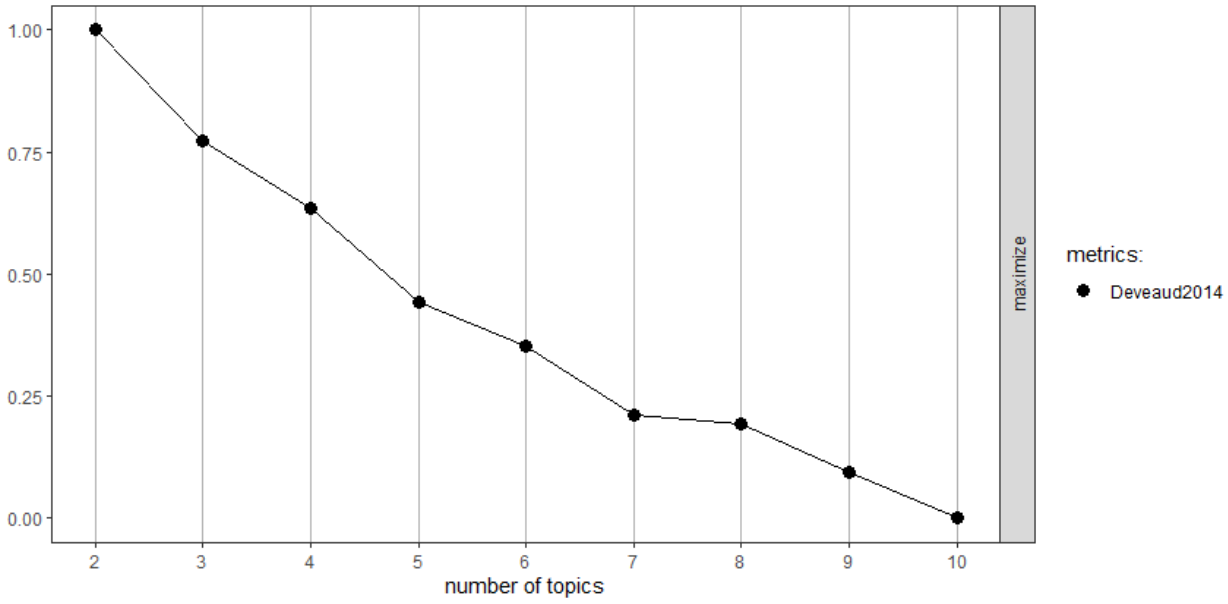

zbeta1
zbeta2
day
show
seri_premier
thursday
today
premier
american_dream_team_don
season
thursday_septemb
skillman
readi
forward
mile
sun
queen
time
reorder(term, abs(zbeta))
0
5
10
15
zbeta
topic
zbeta1
zbeta2

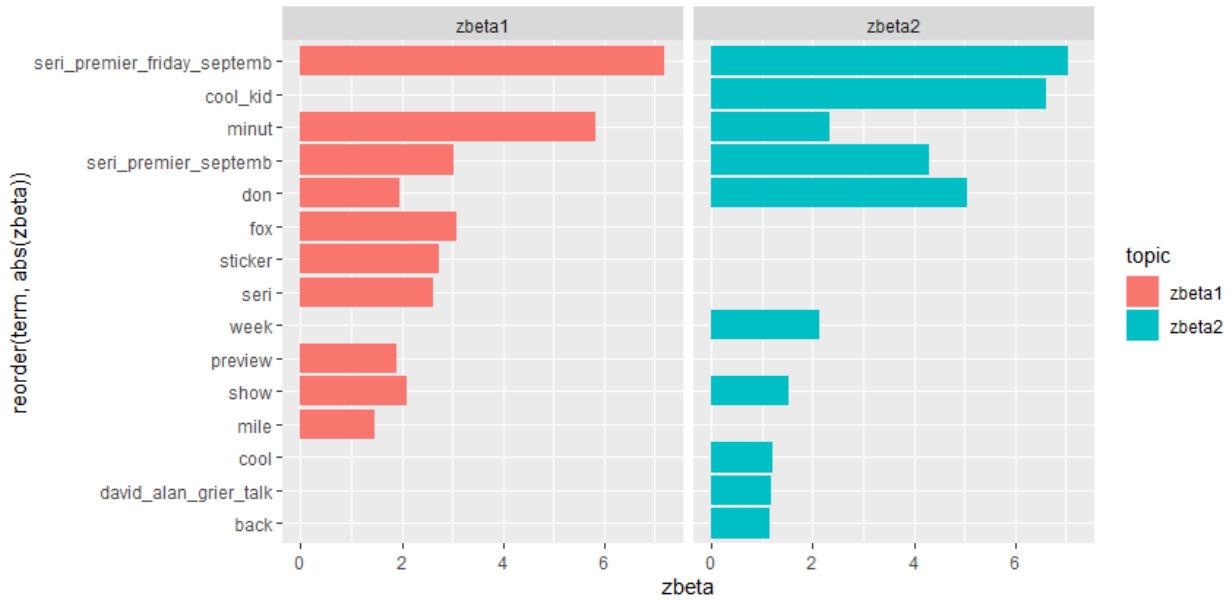
The Alec Baldwin Show



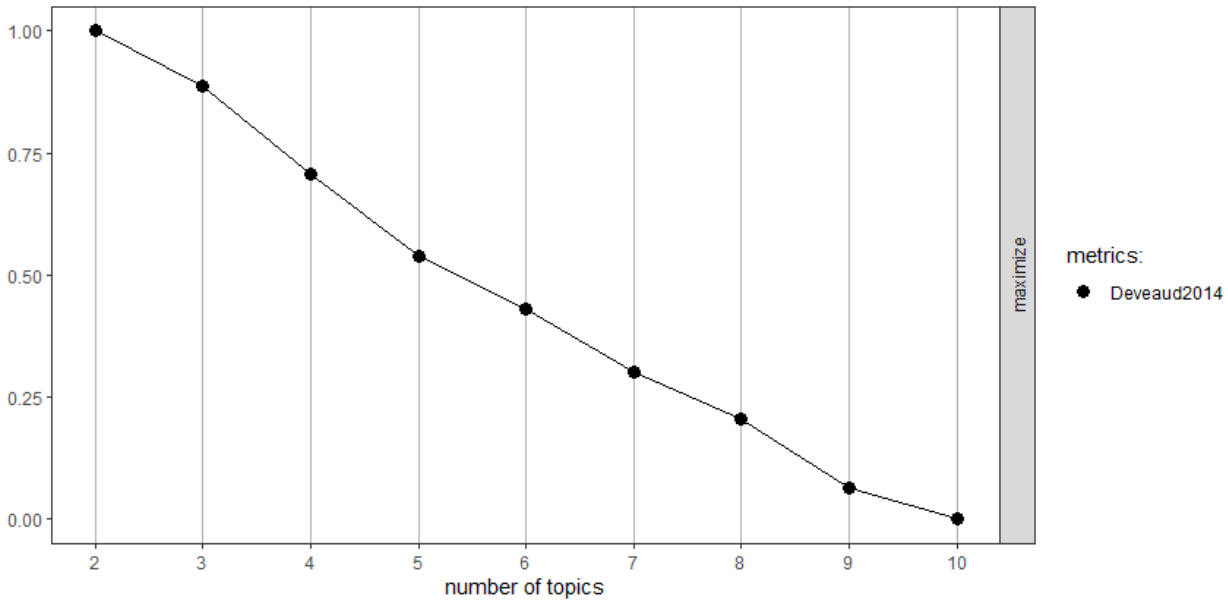


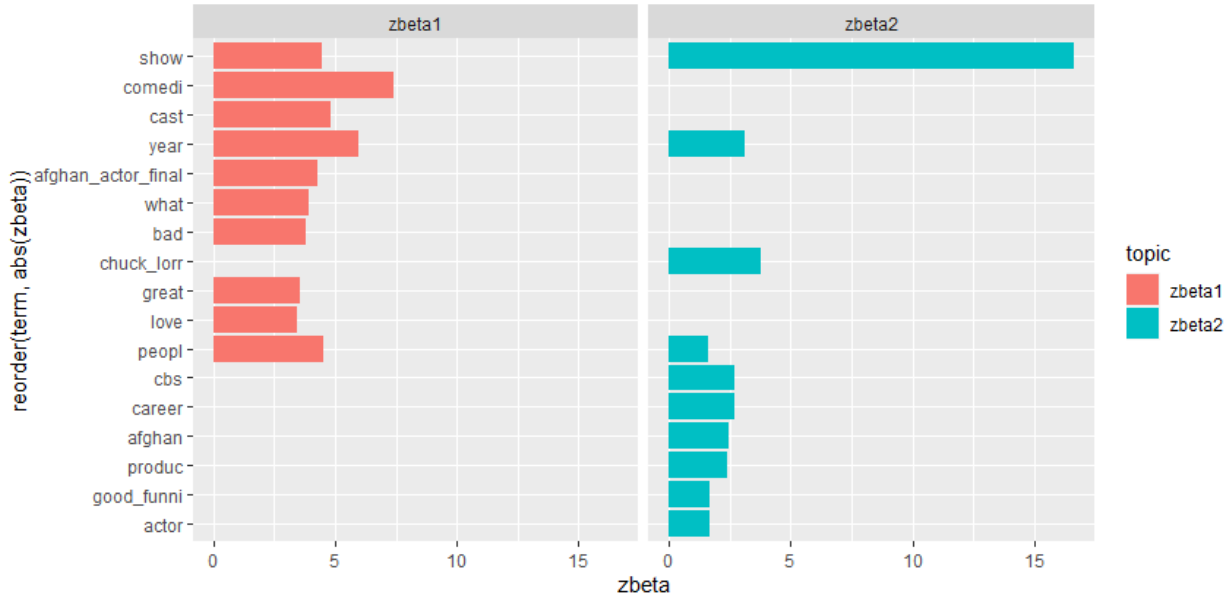
The Cool Kids



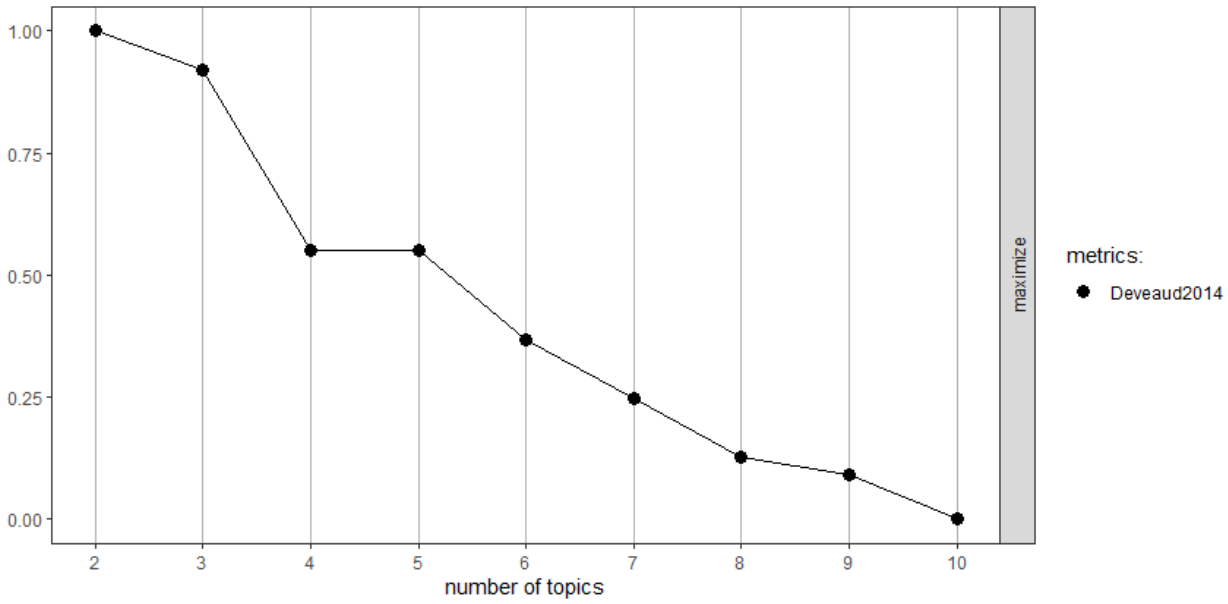


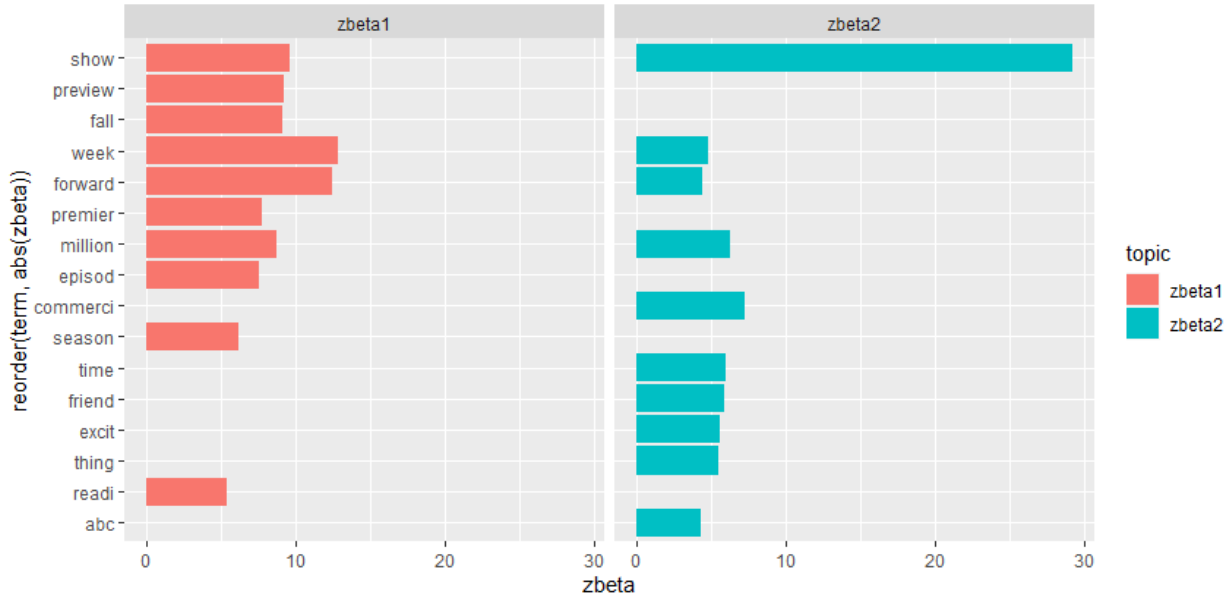
The United States of Al



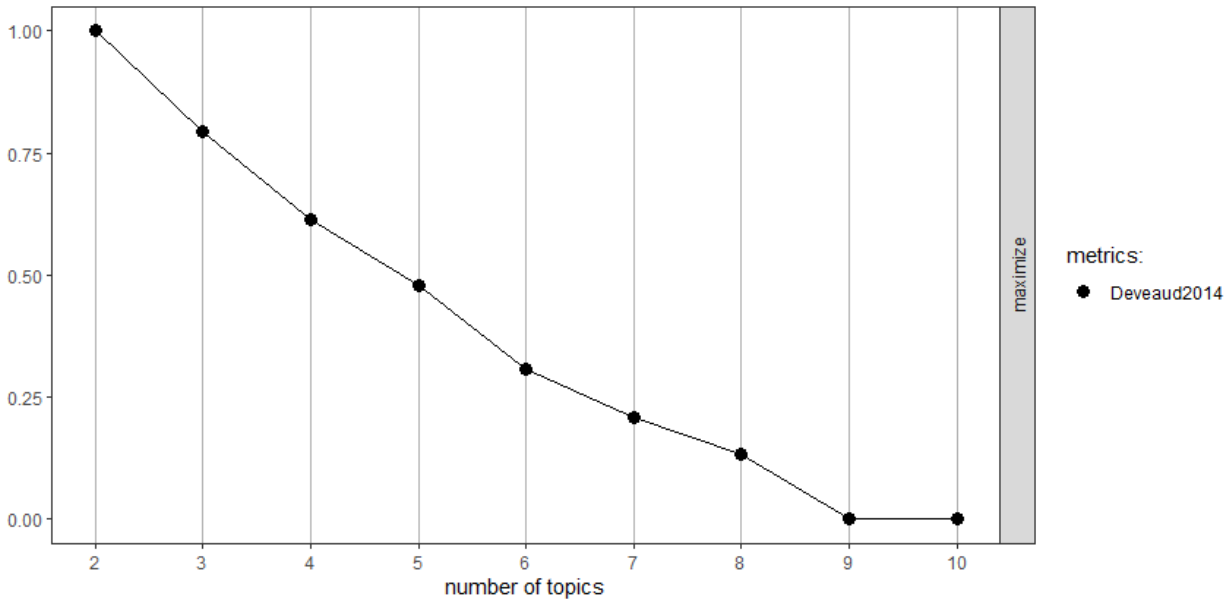


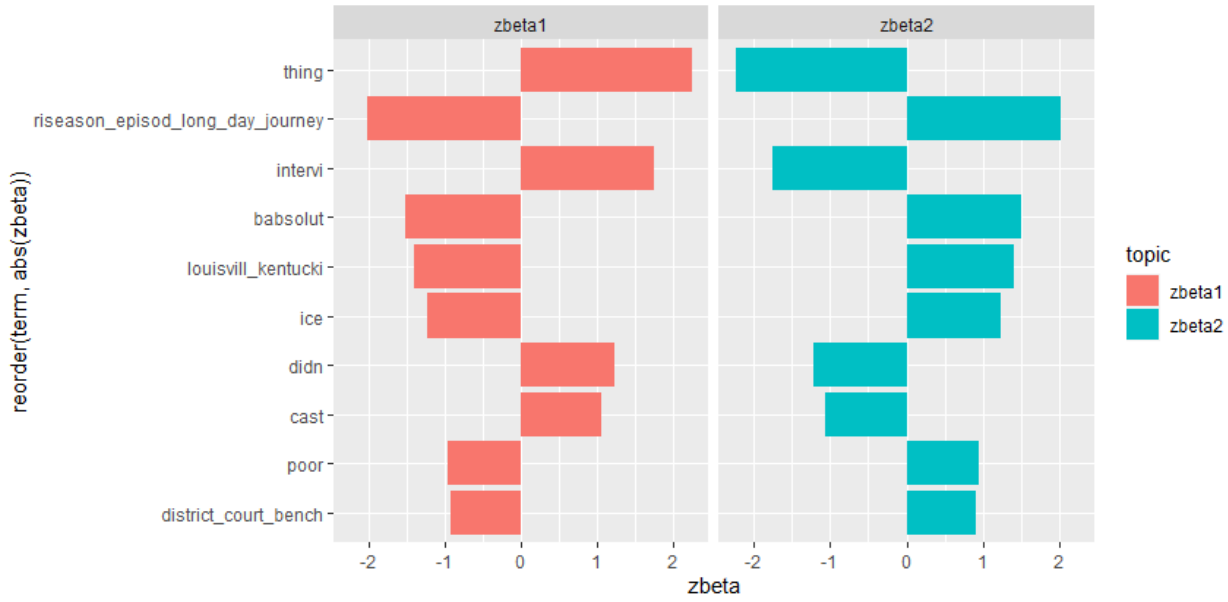
A Million Little Things



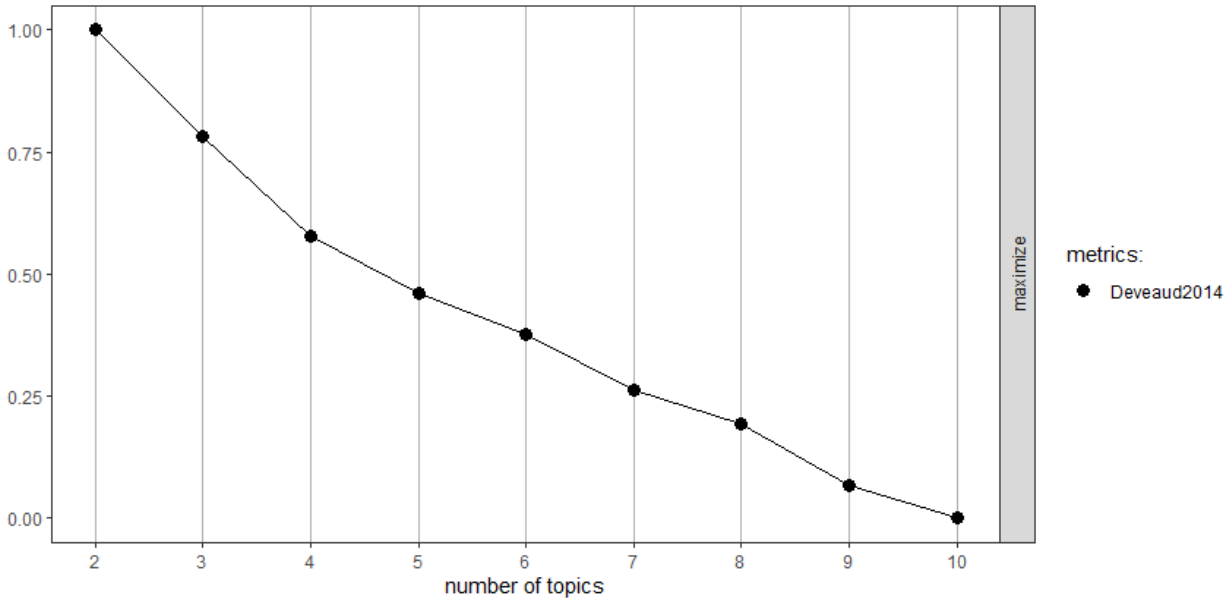


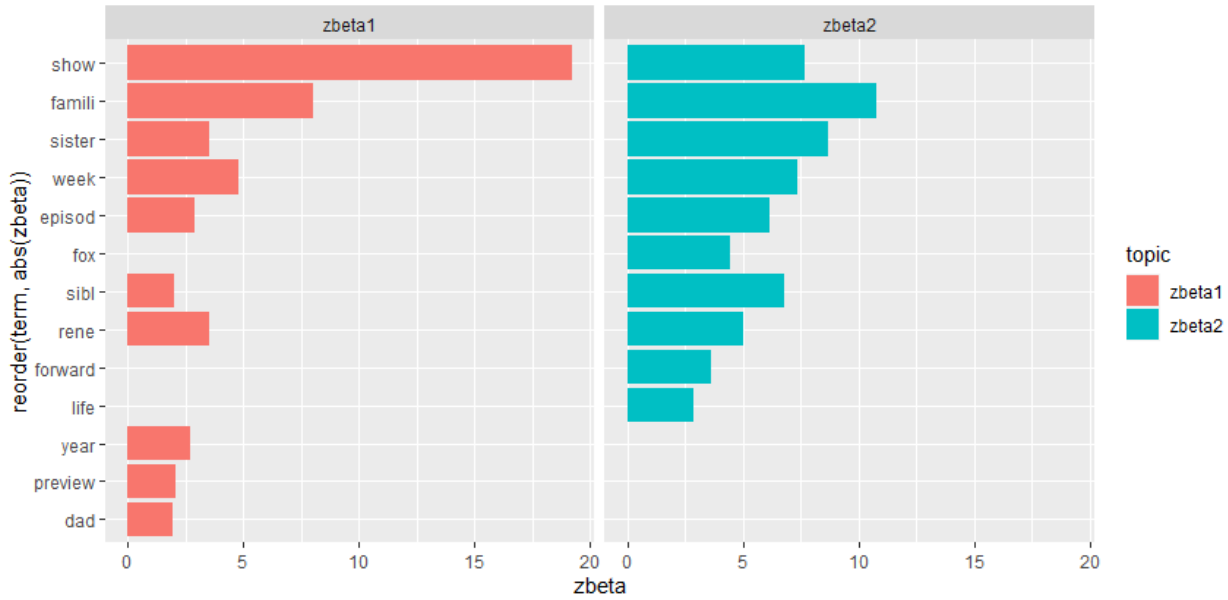
All Rise


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

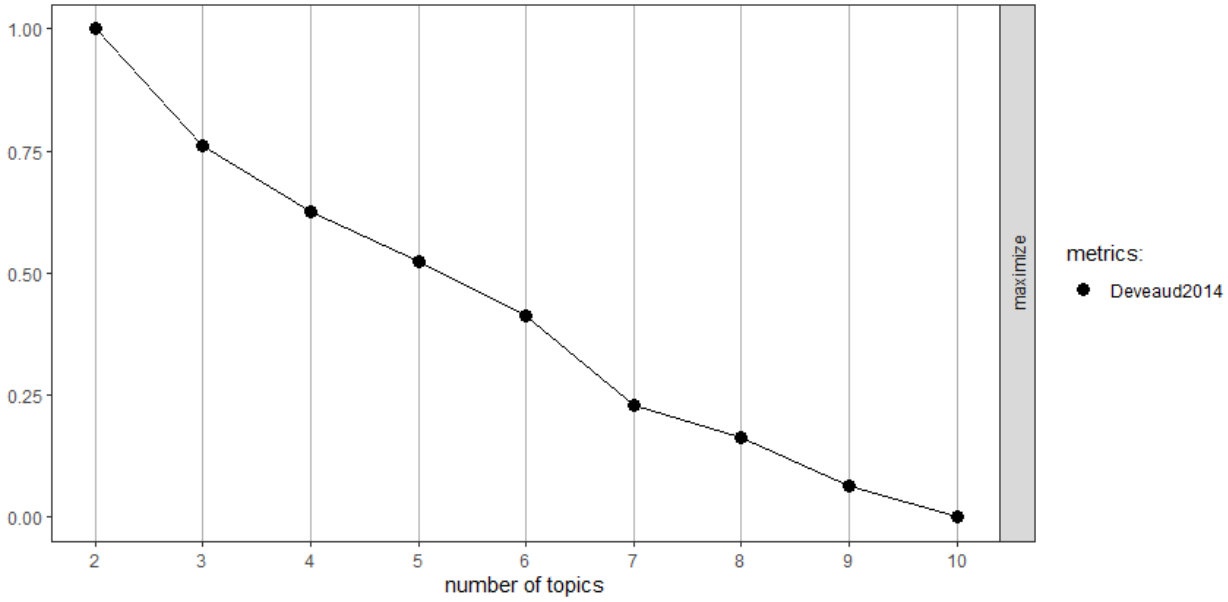

zbeta1
zbeta2
thing
riseason_episod_long_day_journey
intervi
babsolut
louisvill_kentucki
ice
didn
cast
poor
district_court_bench
reorder(term, abs(zbeta))
-2
-1
0
1
2
zbeta
topic
zbeta1
zbeta2

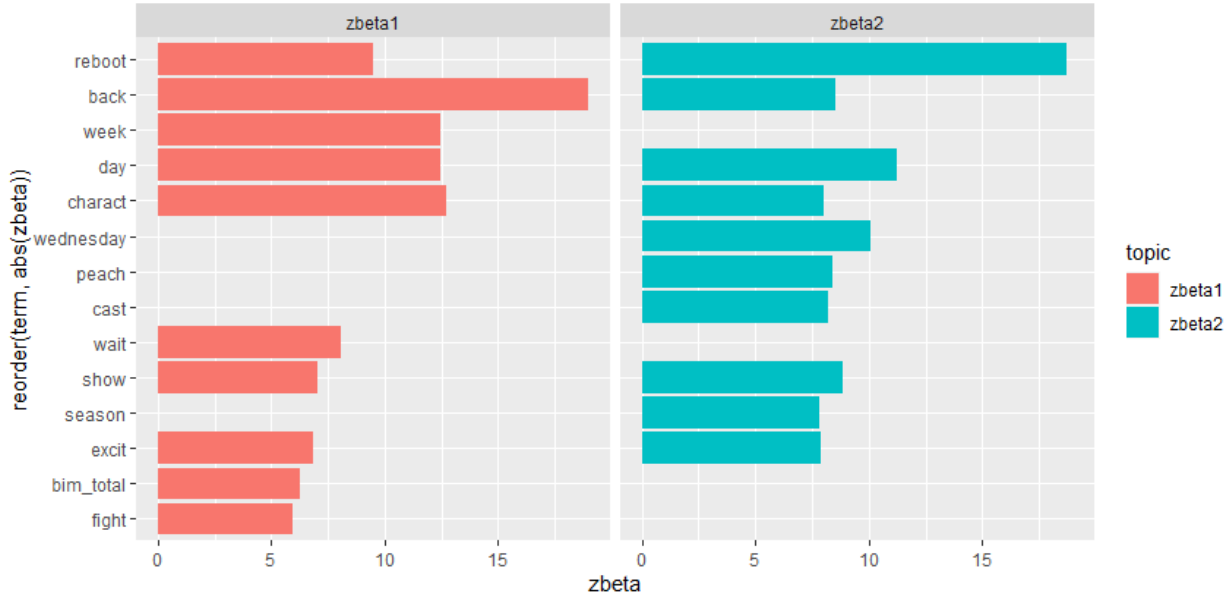
Almost Family



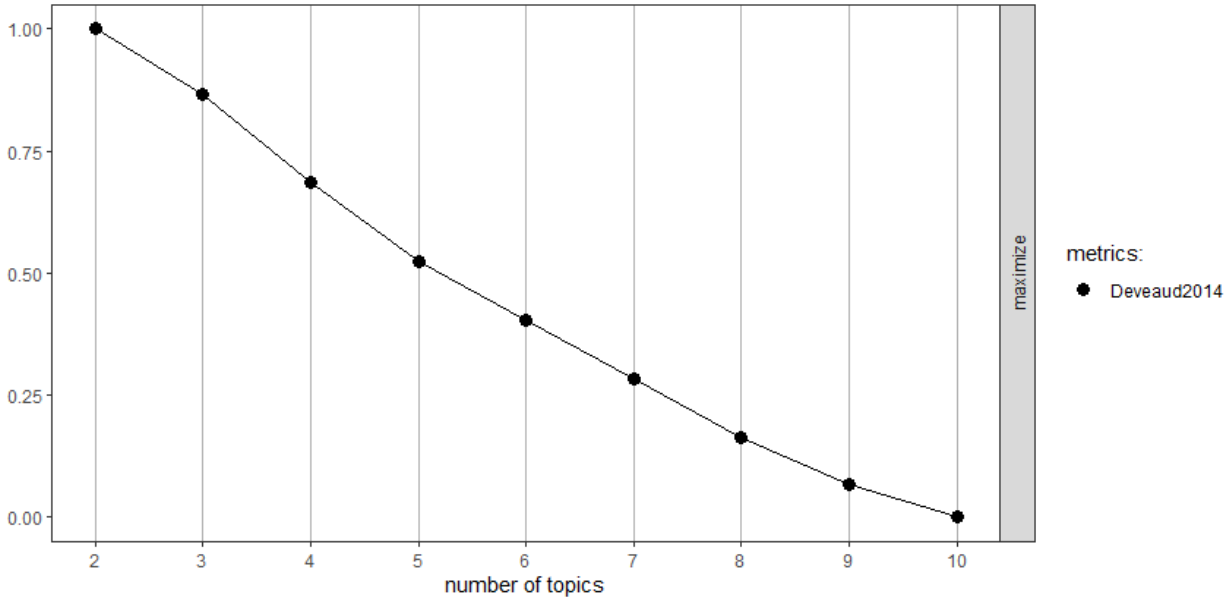


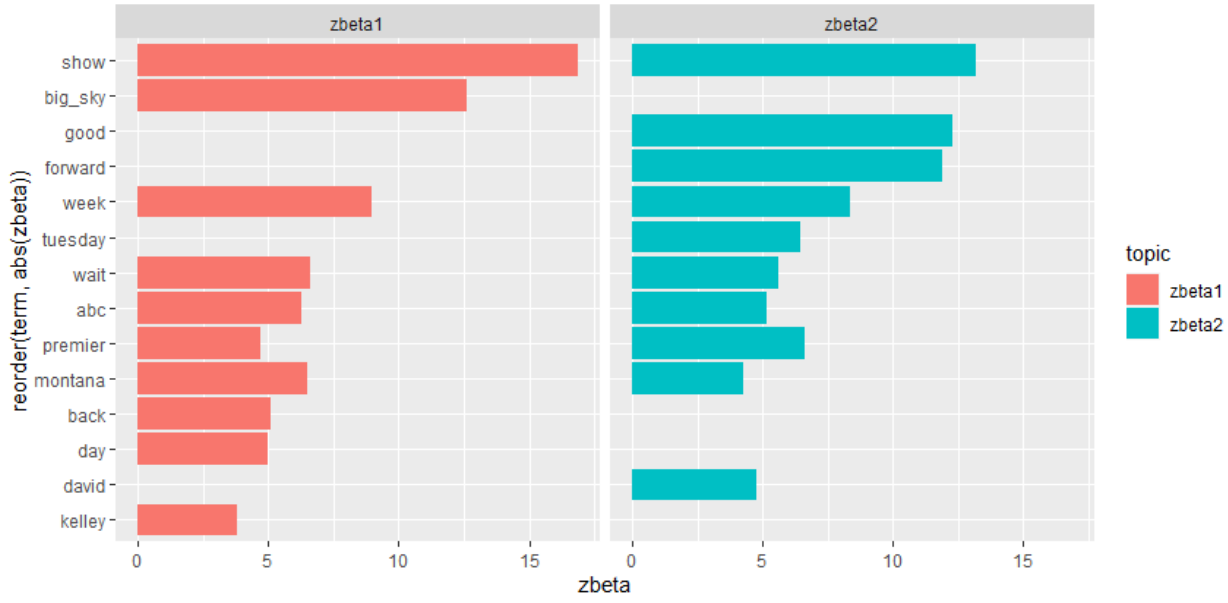
BH90210


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

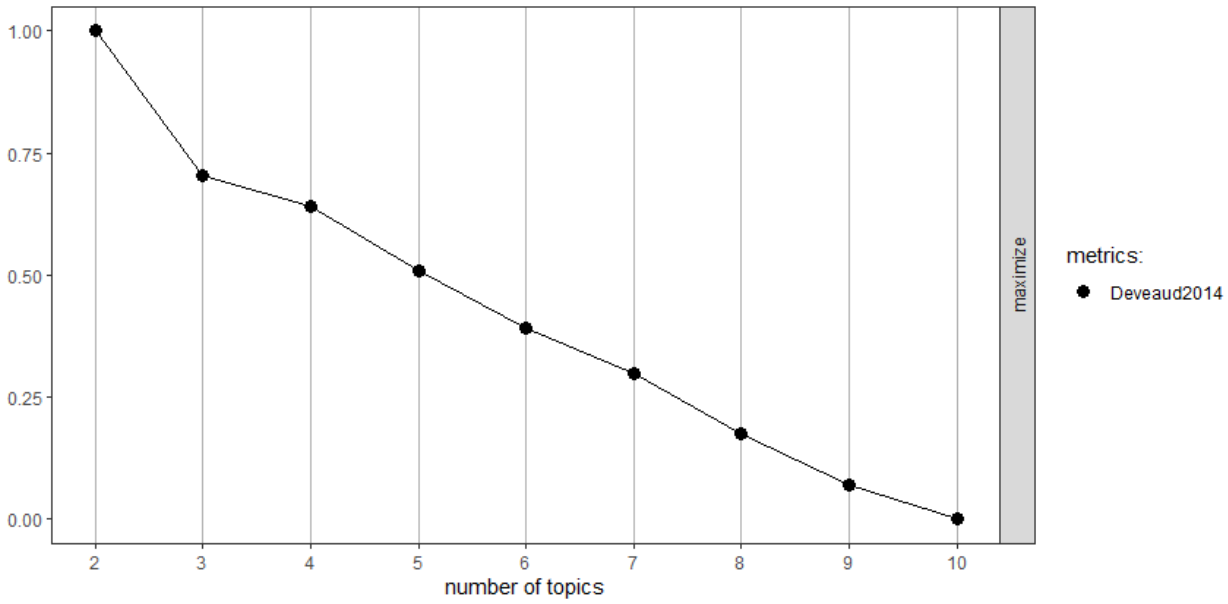

zbeta1
zbeta2
reboot
back
week
day
charact
wednesday
peach
cast
wait
show
season
excit
bim_total
fight
reorder(term, abs(zbeta))
0
5
10
15
zbeta
topic
zbeta1
zbeta2

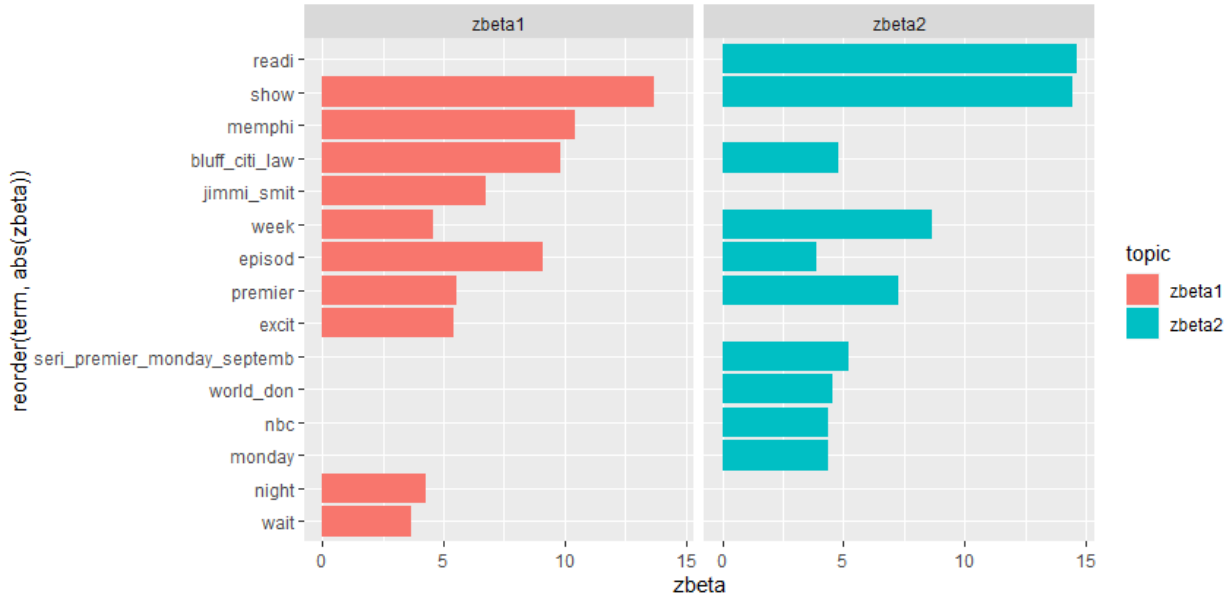
Big Sky


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

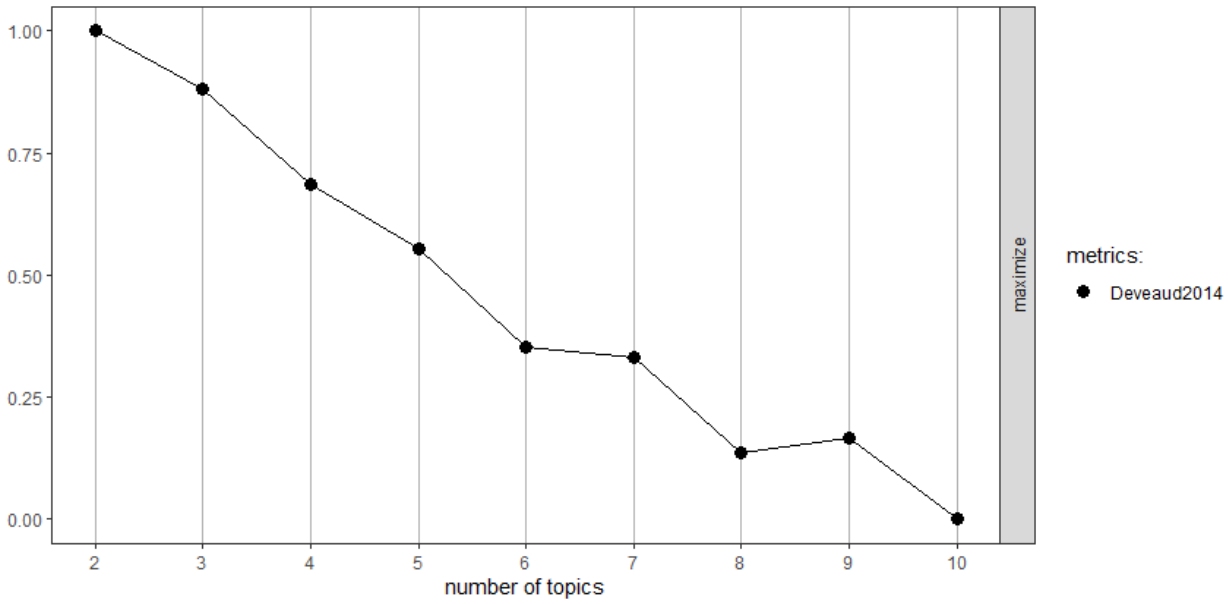

zbeta1
zbeta2
show
big_sky
good
forward
week
tuesday
wait
abc
premier
montana
back
day
david
kelley
reorder(term, abs(zbeta))
0
5
10
15
zbeta
topic
zbeta1
zbeta2

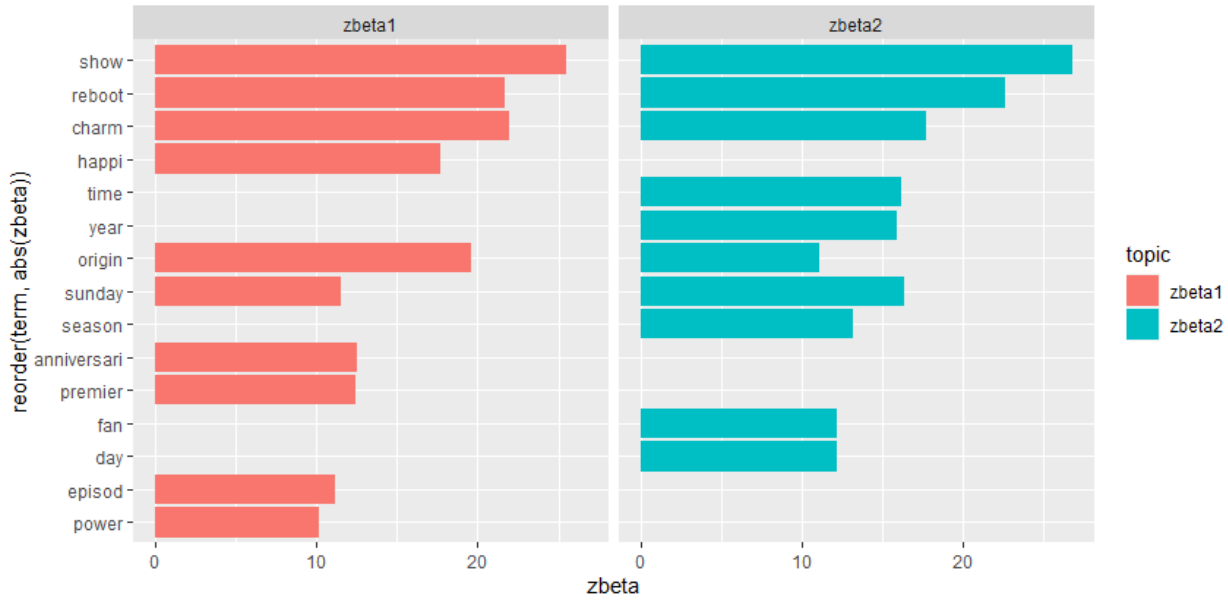
Bluff City Law



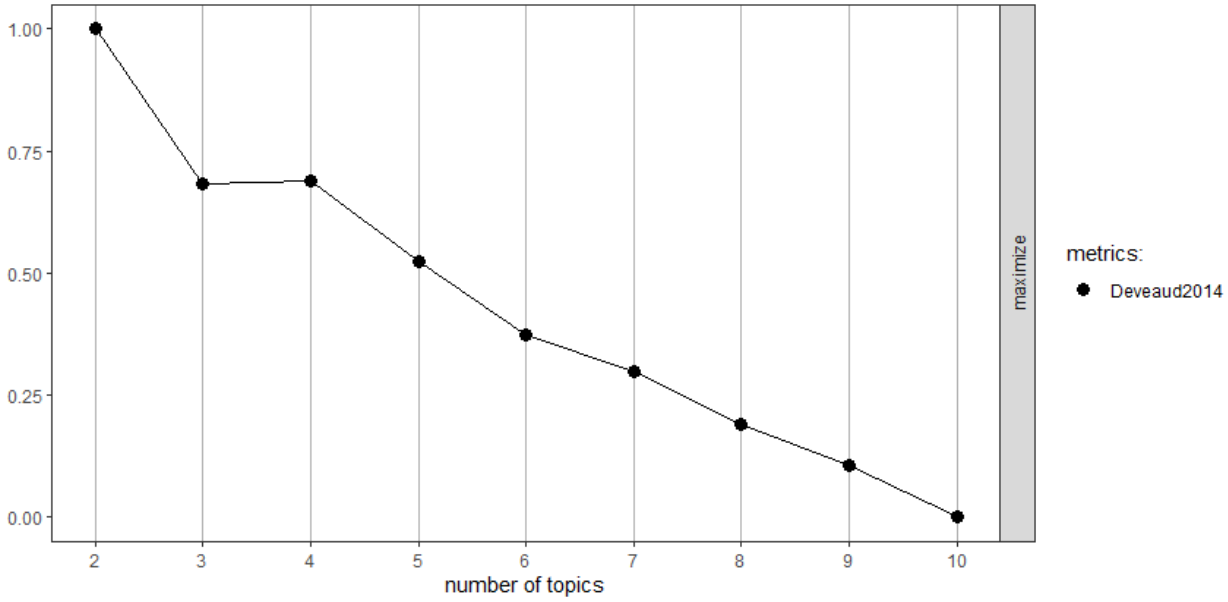


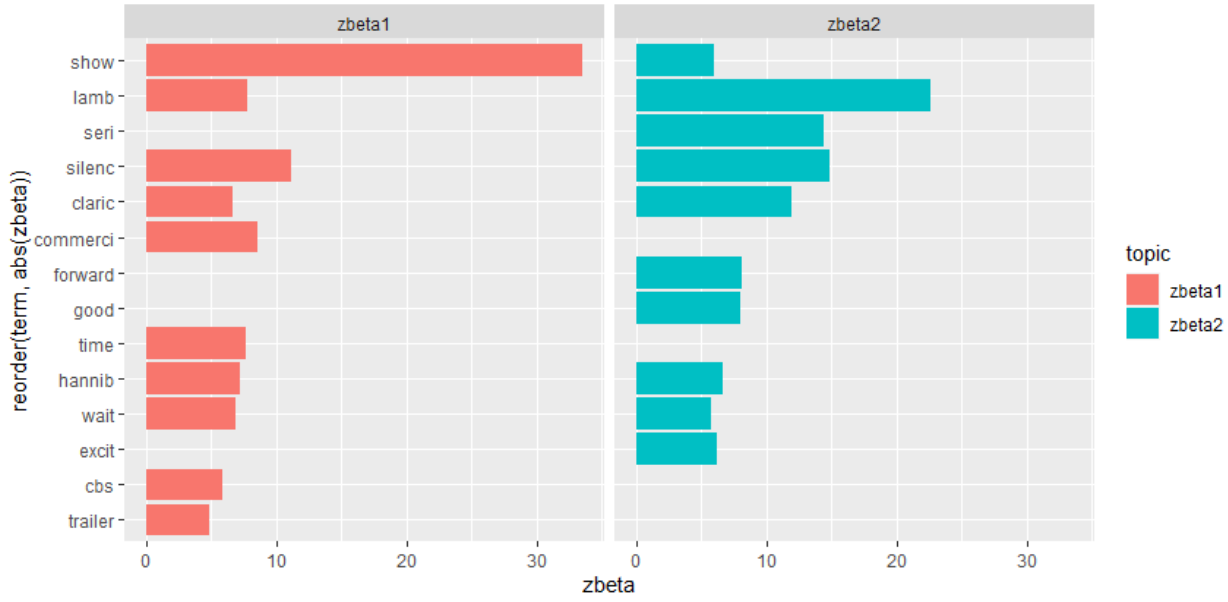
Charmed


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

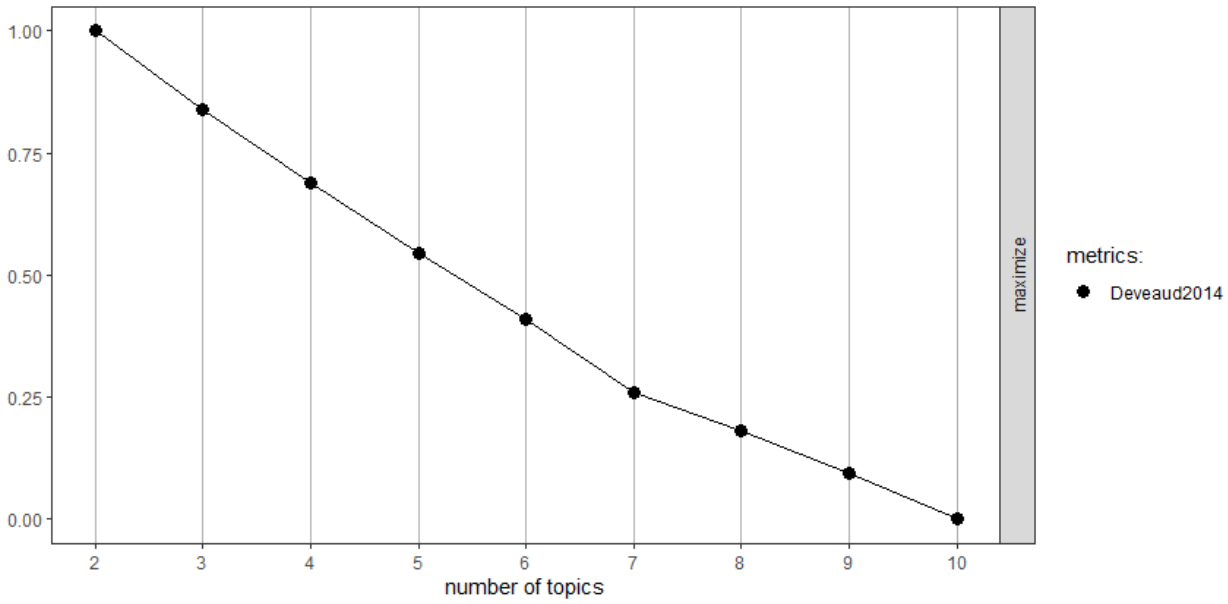

zbeta1
zbeta2
show
reboot
charm
happi
time
year
origin
sunday
season
anniversari
premier
fan
day
episod
power
reorder(term, abs(zbeta))
0
10
20
zbeta
topic
zbeta1
zbeta2

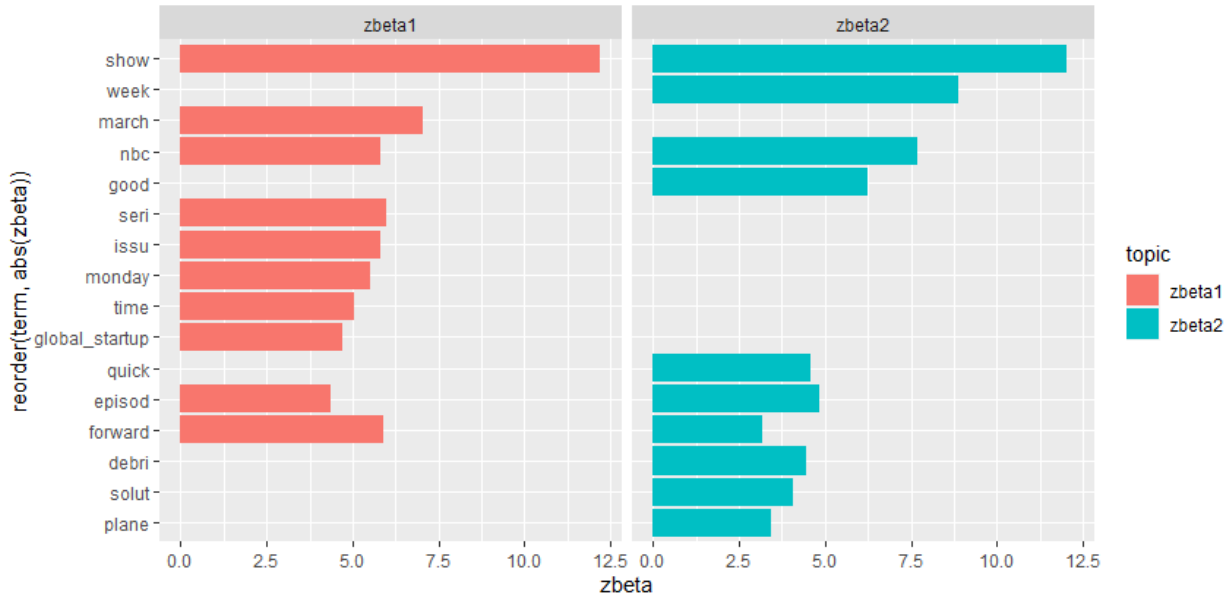
Clarice


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

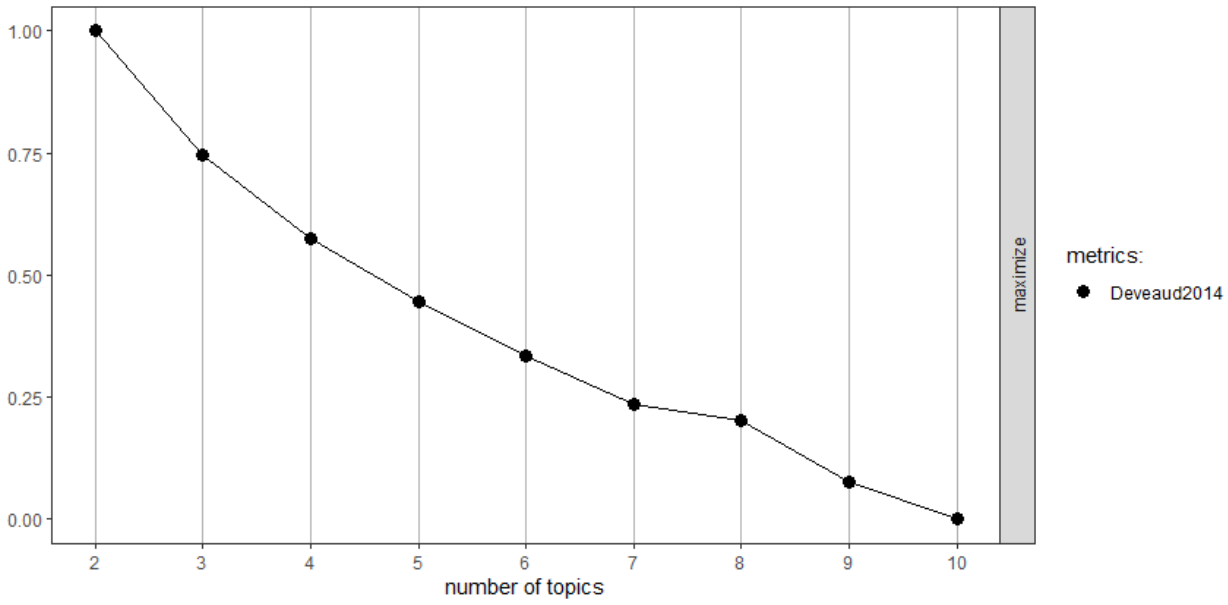

zbeta1
zbeta2
show
lamb
seri
silenc
claric
commerci
forward
good
time
hannib
wait
excit
cbs
trailer
reorder(term, abs(zbeta))
0
10
20
30
zbeta
topic
zbeta1
zbeta2

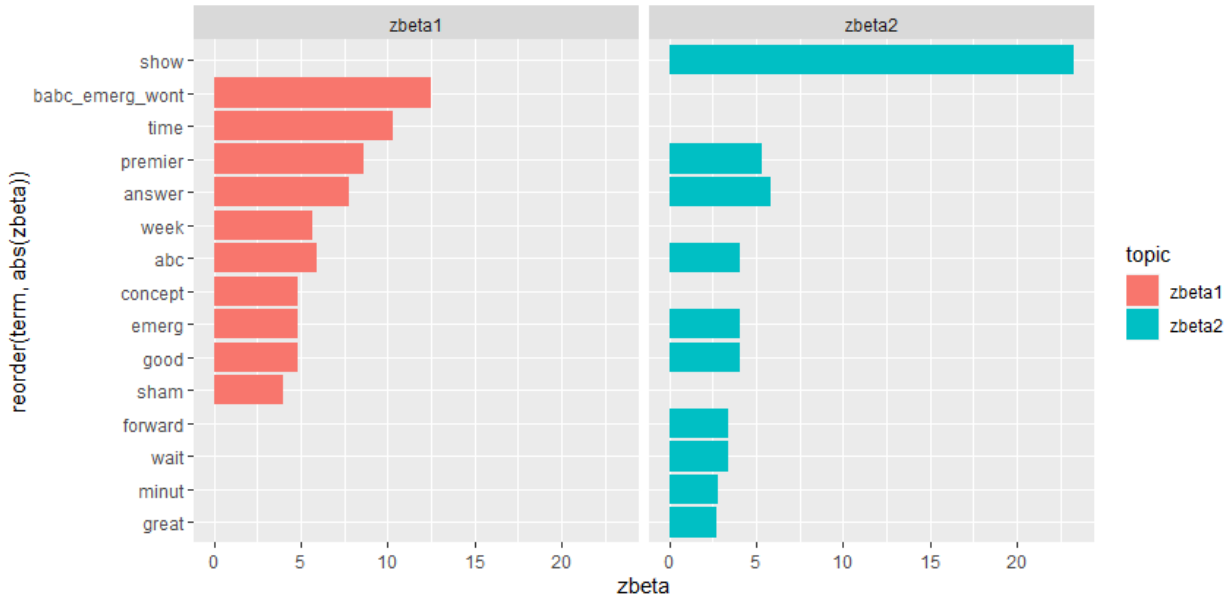
Debris


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

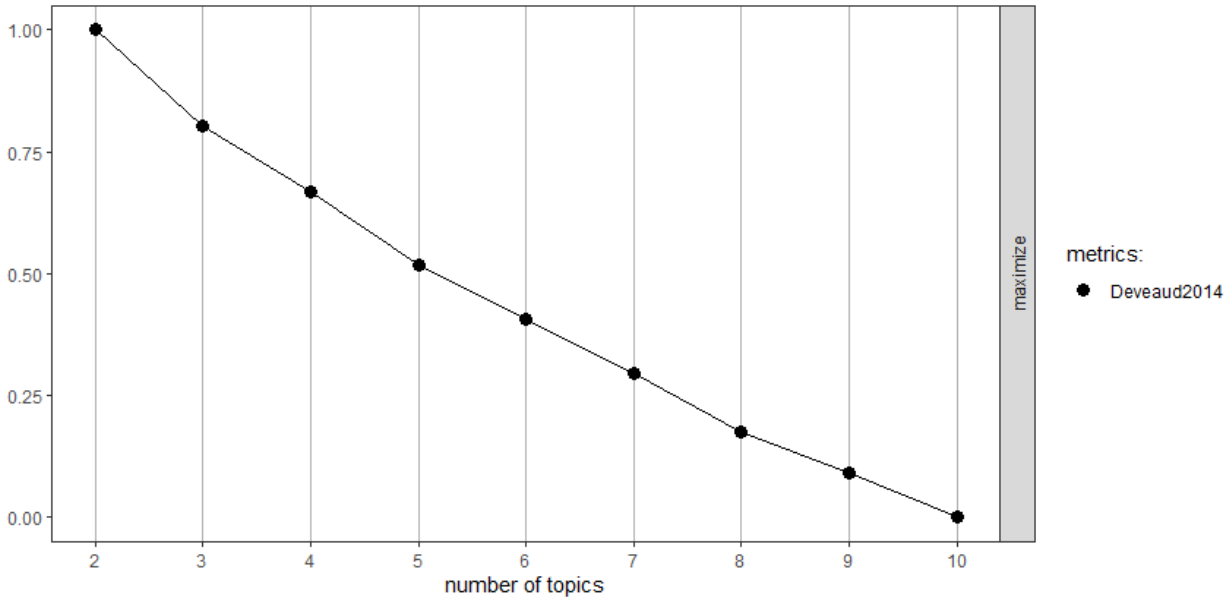

zbeta1
zbeta2
show
week
march
nbc
good
seri
issu
monday
time
global_startup
quick
episod
forward
debri
solut
plane
reorder(term, abs(zbeta))
0.0
2.5
5.0
7.5
10.0
12.5
zbeta
topic
zbeta1
zbeta2

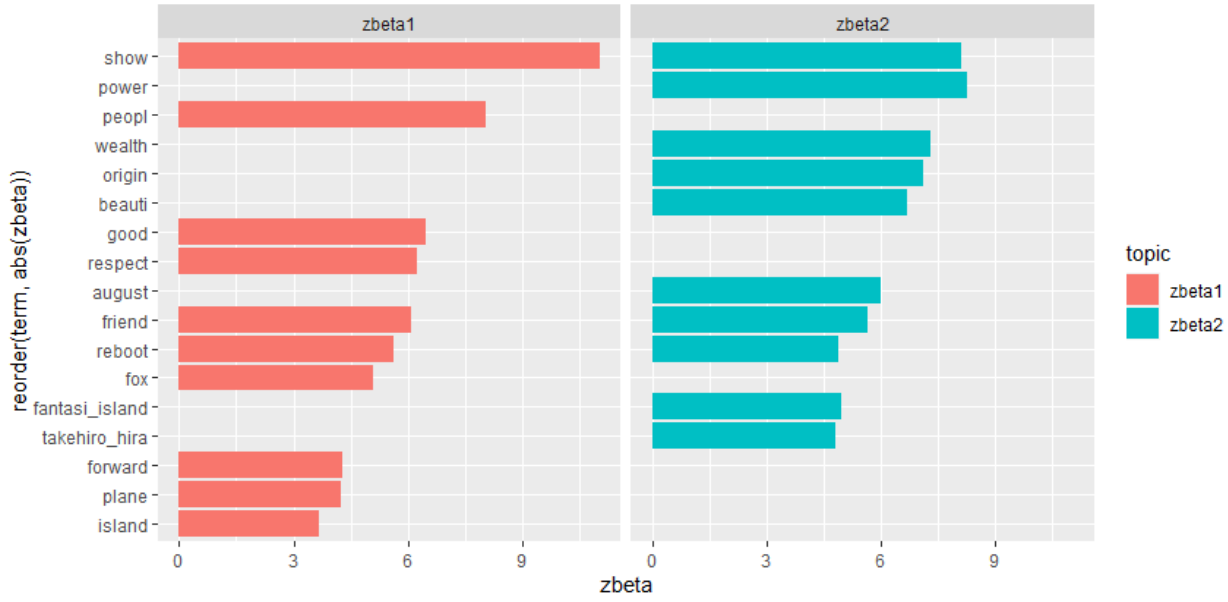
Emergence


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

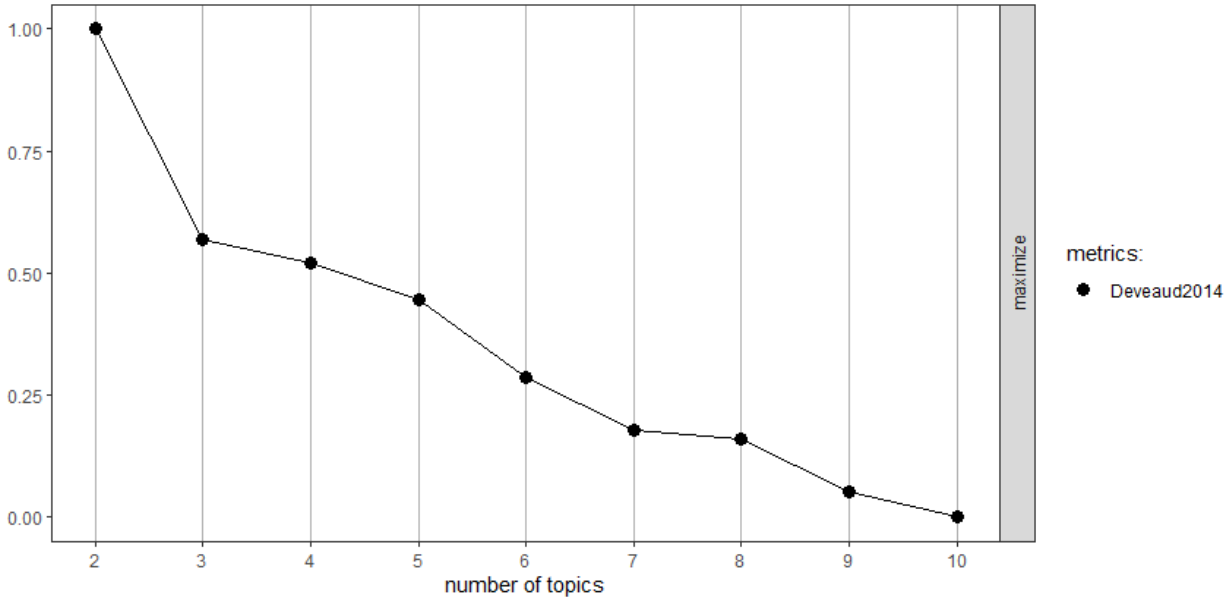

zbeta1
zbeta2
show
babc_emerg_wont
time
premier
answer
week
abc
concept
emerg
good
sham
forward
wait
minut
great
reorder(term, abs(zbeta))
0
5
10
15
20
zbeta
topic
zbeta1
zbeta2

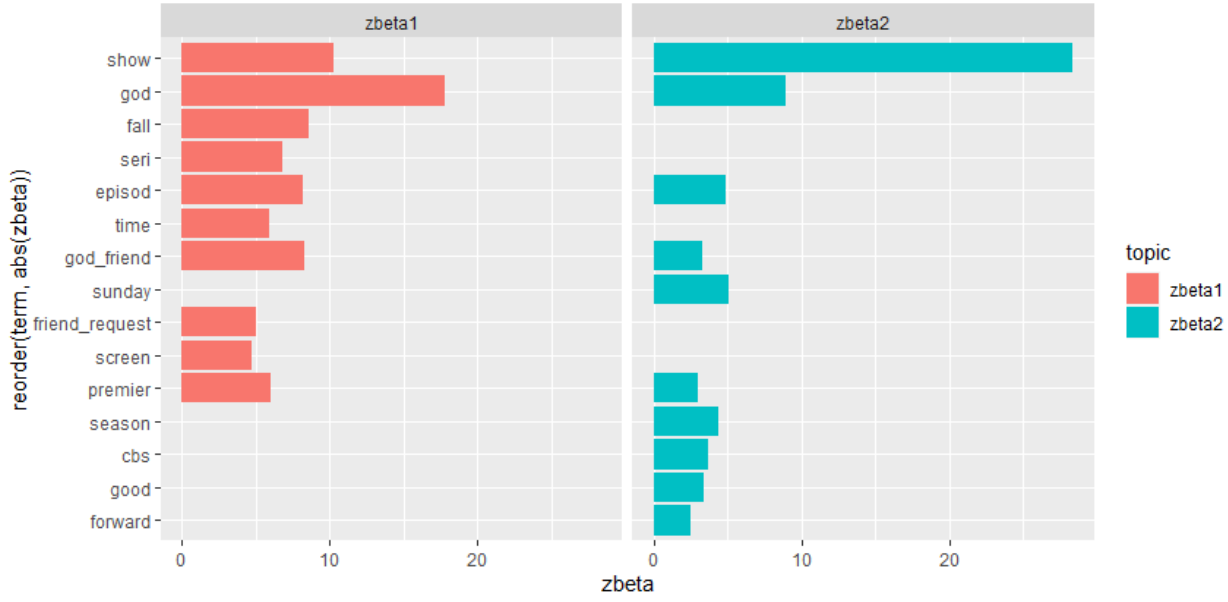
Fantasy Island



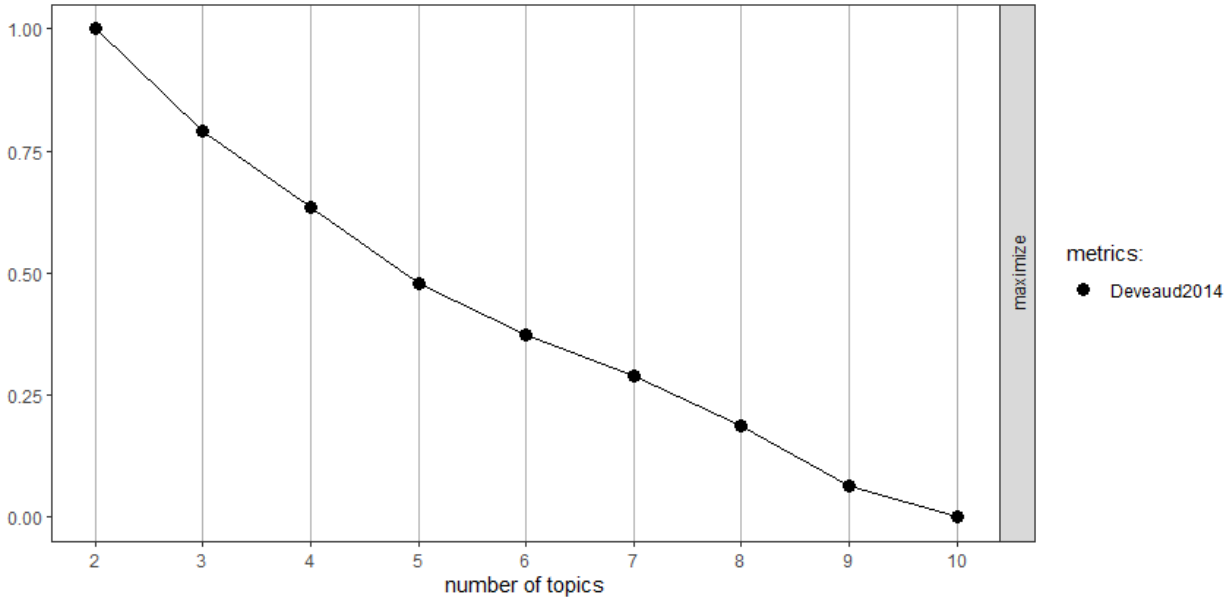


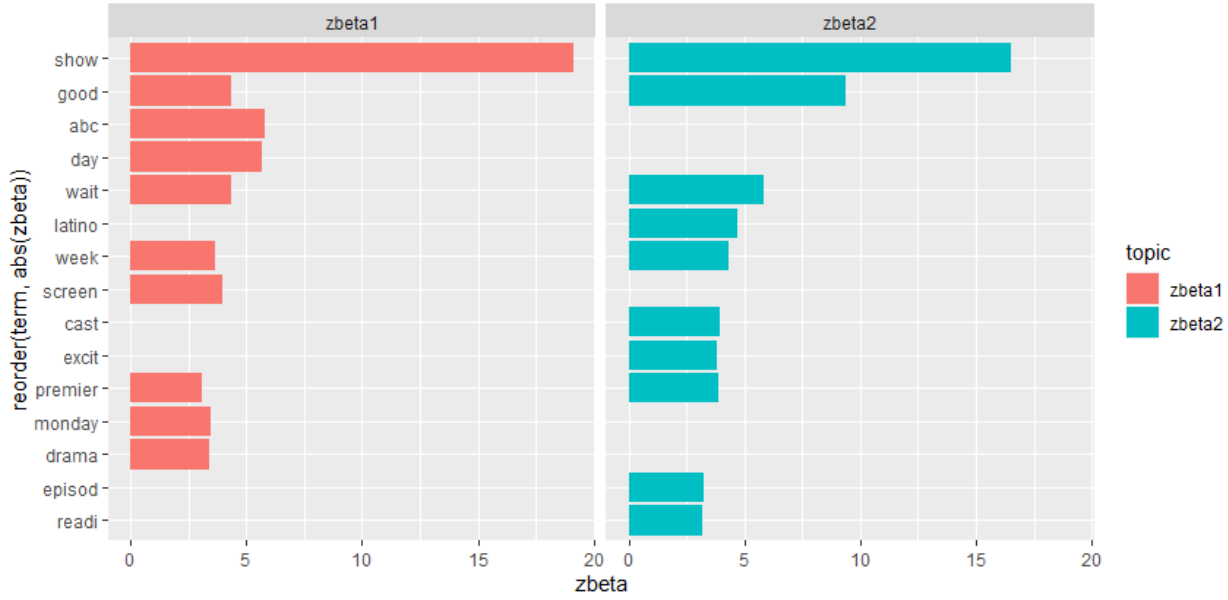
God Friended Me



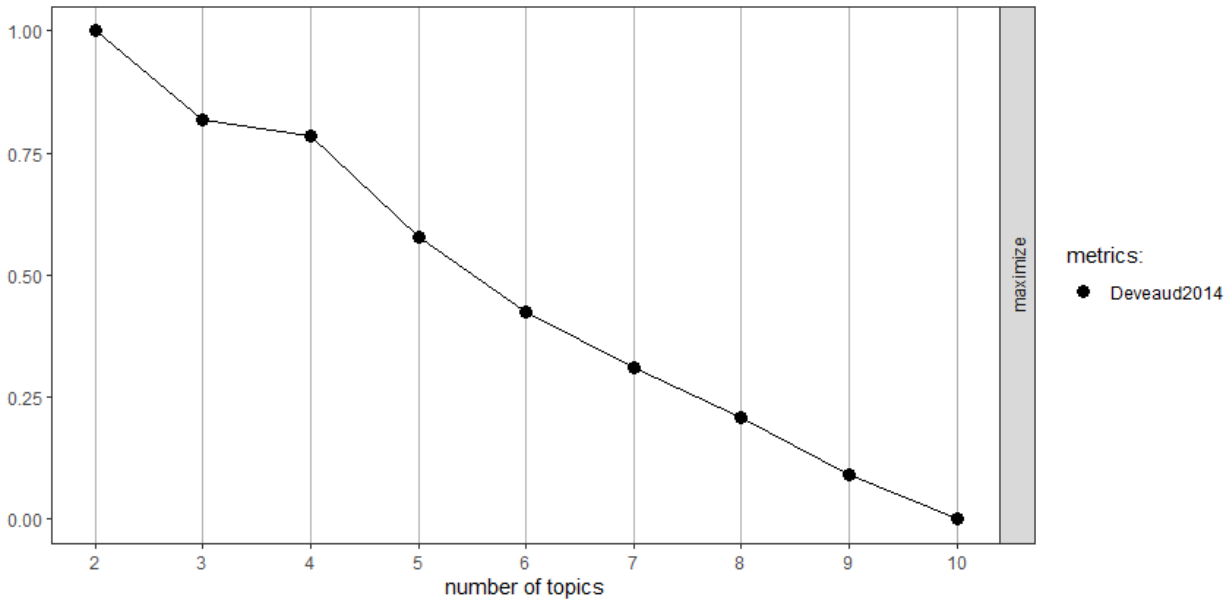


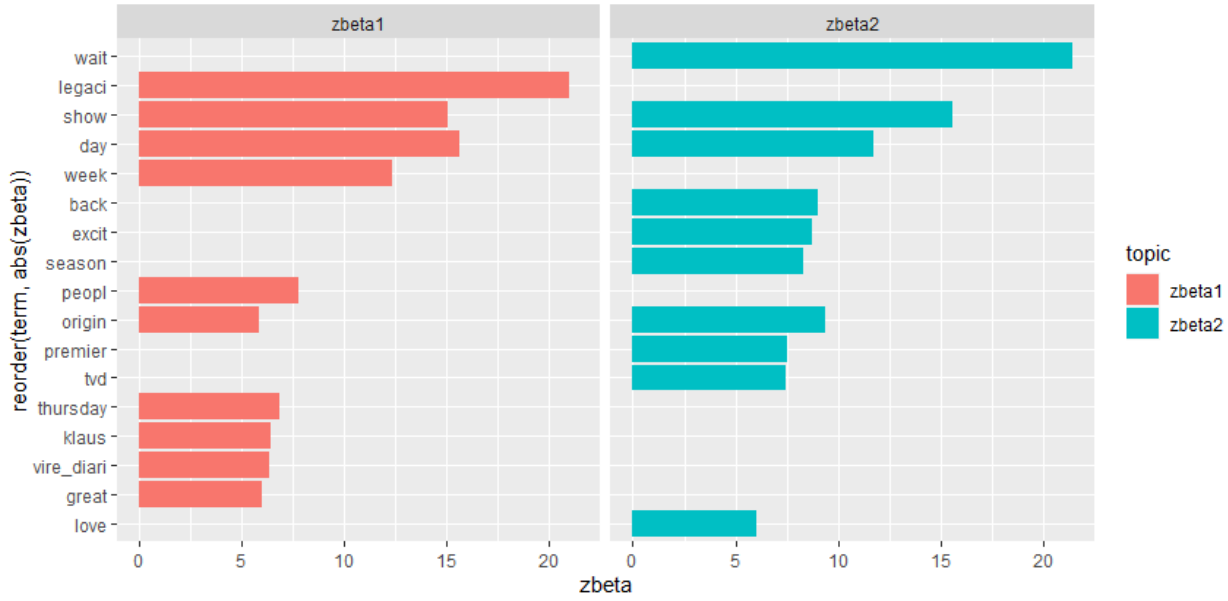
Grand Hotel


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

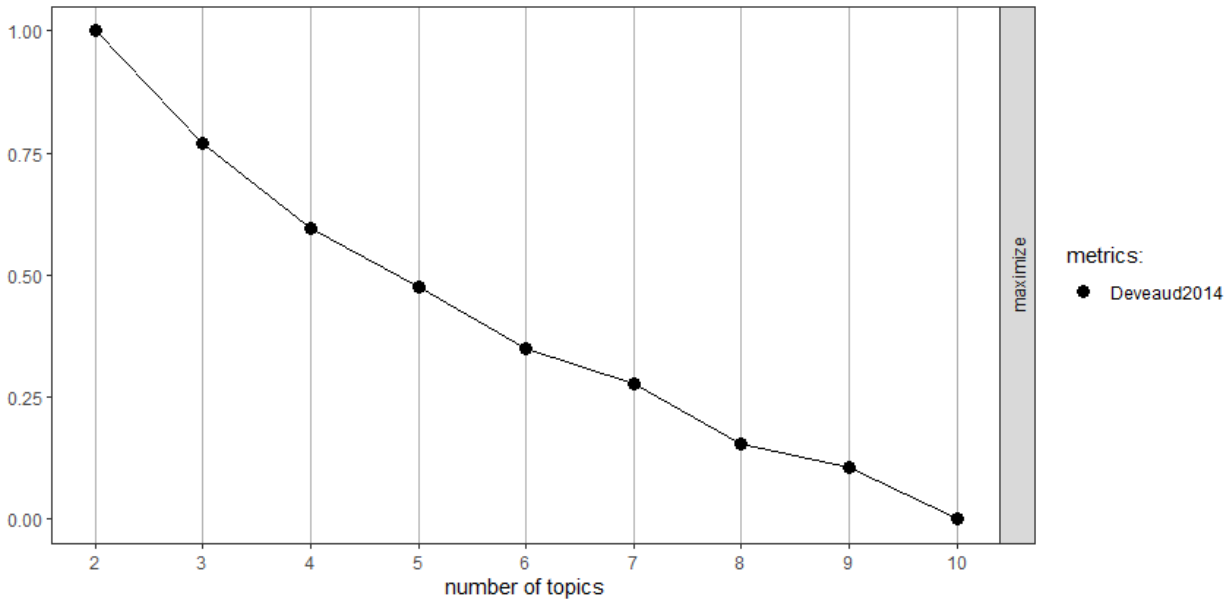

zbeta1
zbeta2
show
good
abc
day
wait
latino
week
screen
cast
excit
premier
monday
drama
episod
readi
reorder(term, abs(zbeta))
0
5
10
15
20
zbeta
topic
zbeta1
zbeta2

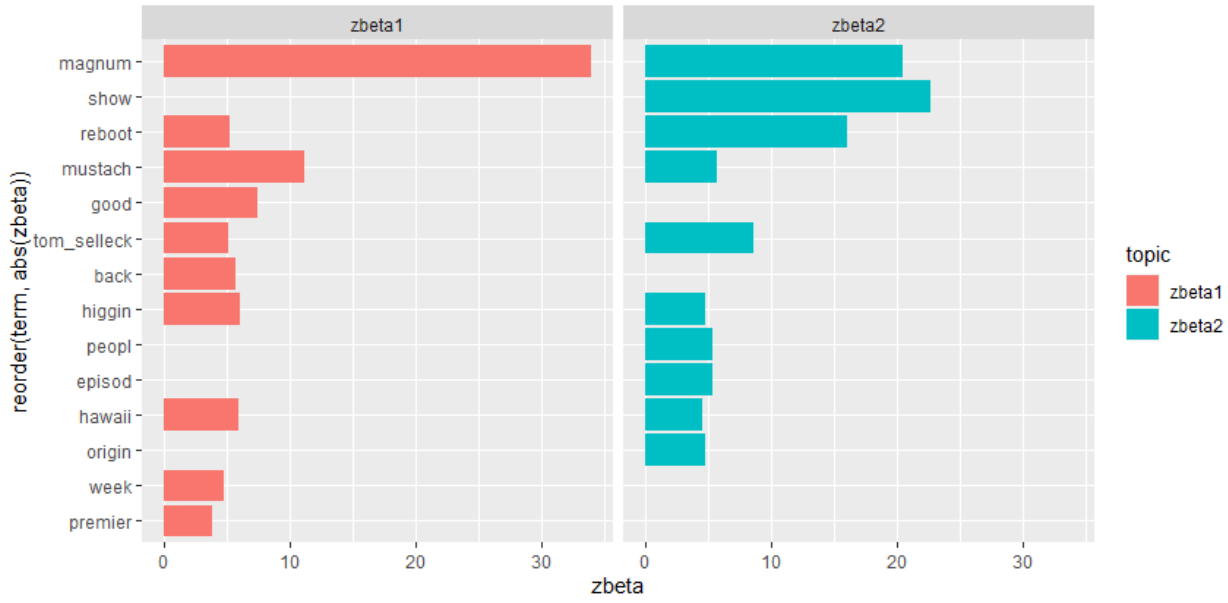
Legacies


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

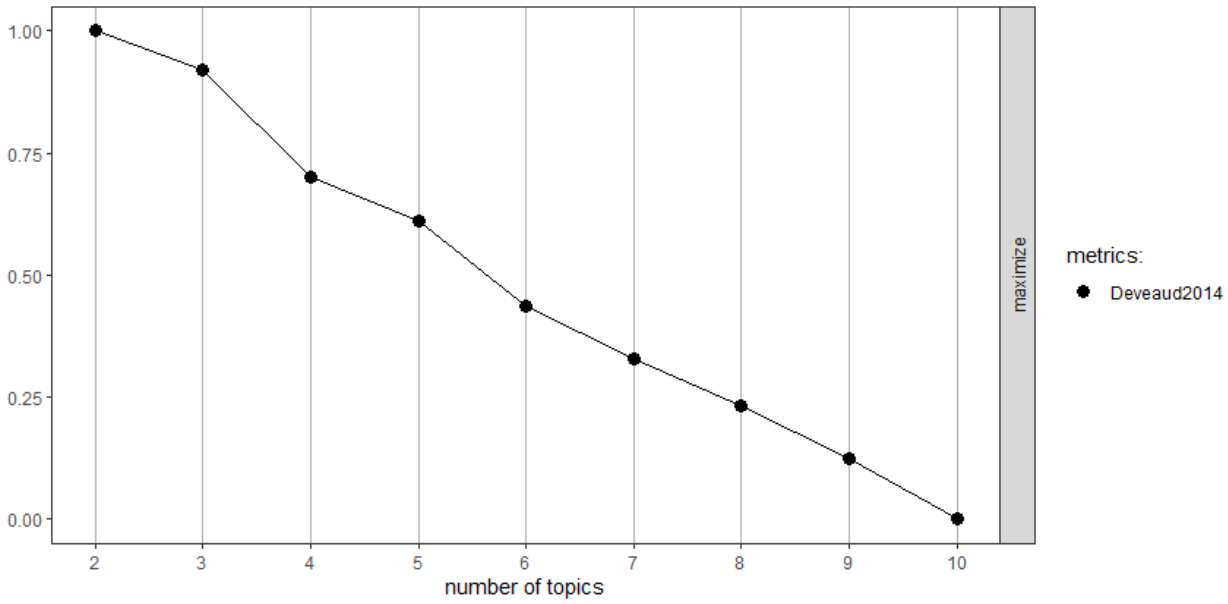

zbeta1
zbeta2
wait
legaci
show
day
week
back
excit
season
peopl
origin
premier
tvd
thursday
klaus
vire_diari
great
love
reorder(term, abs(zbeta))
0
5
10
15
20
zbeta
topic
zbeta1
zbeta2

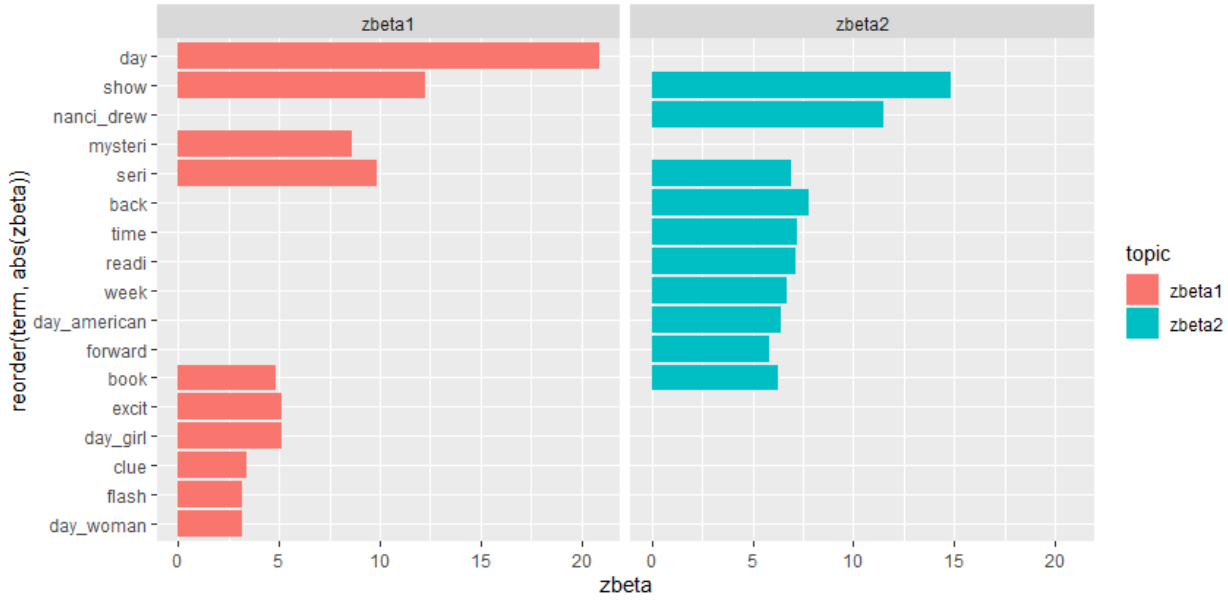
Magnum P.I.


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

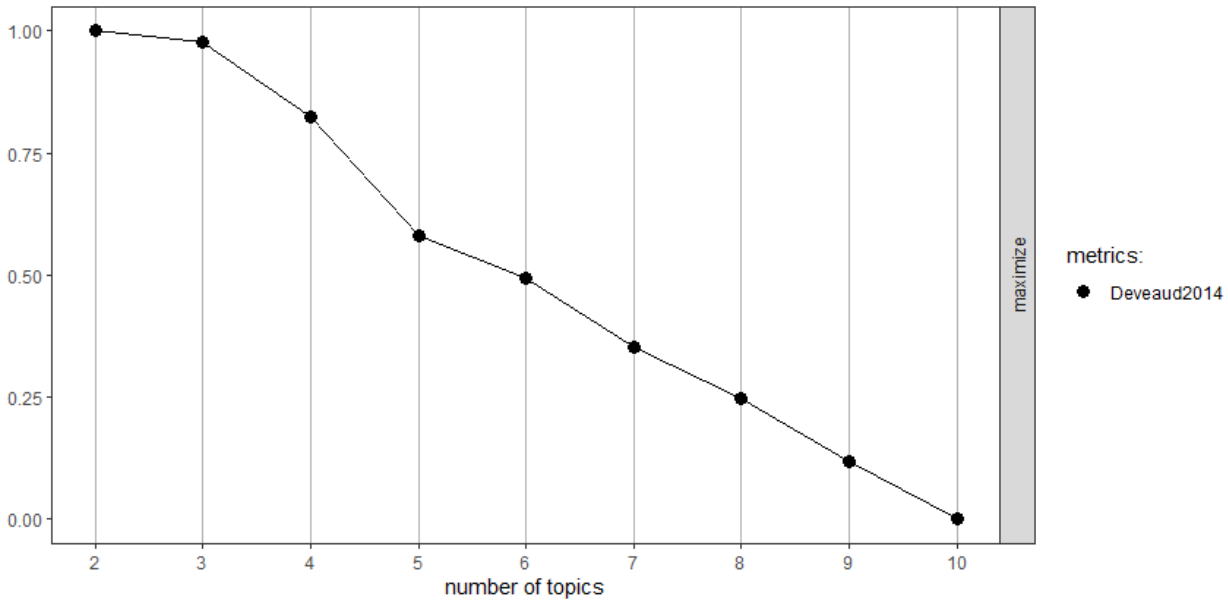

zbeta1
zbeta2
magnum
show
reboot
mustach
good
tom_selleck
back
higgin
peopl
episod
hawaii
origin
week
premier
reorder(term, abs(zbeta))
0
10
20
30
zbeta
topic
zbeta1
zbeta2

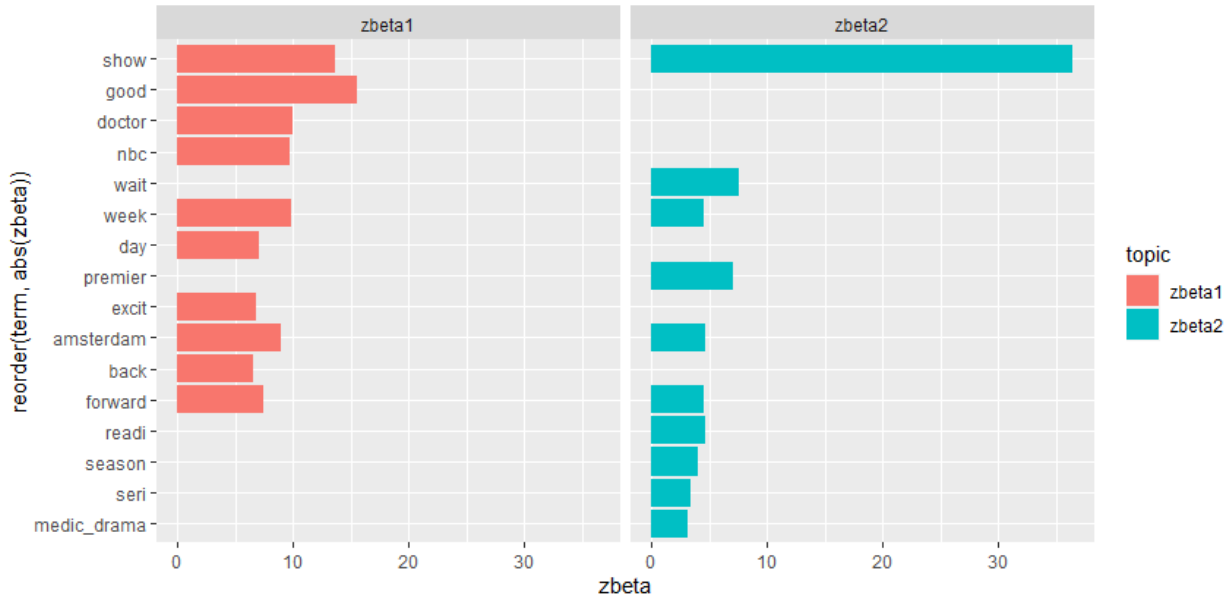
Nancy Drew


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

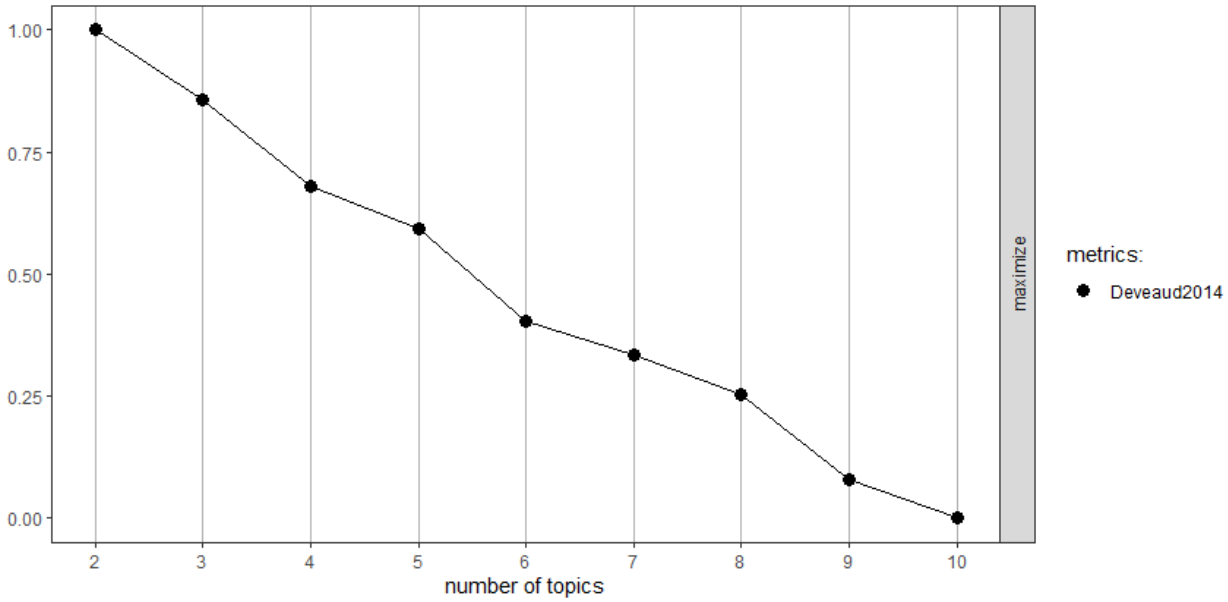

zbeta1
zbeta2
day
show
nanci_drew
mysteri
seri
back
time
readi
week
day_american
forward
book
excit
day_girl
clue
flash
day_woman
reorder(term, abs(zbeta))
0
5
10
15
20
zbeta
topic
zbeta1
zbeta2

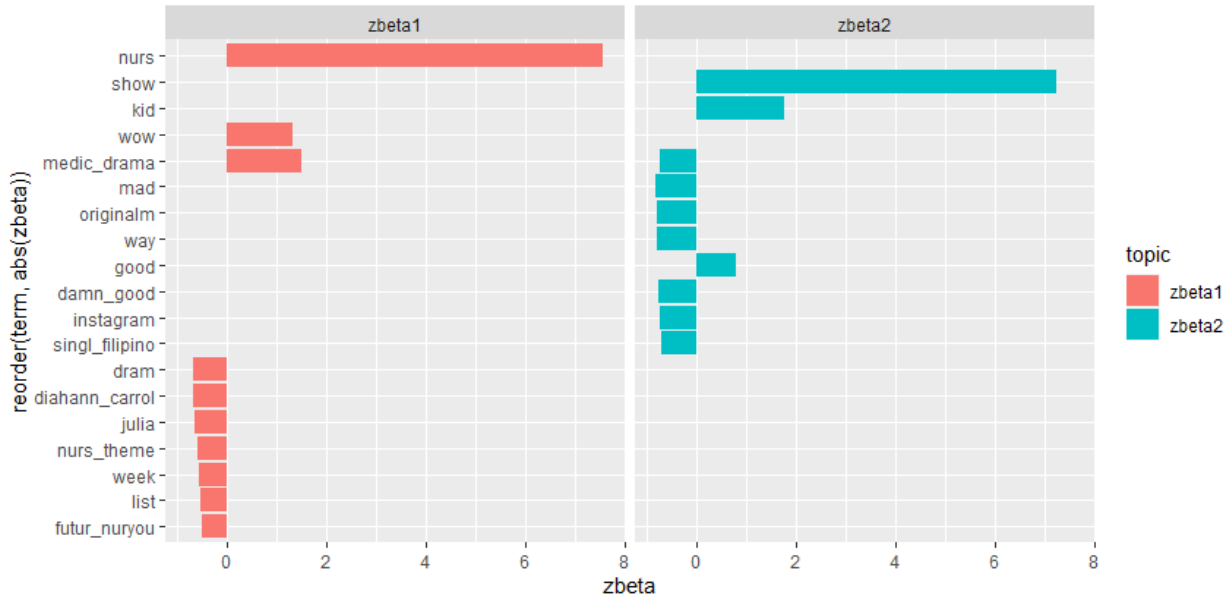
New Amsterdam



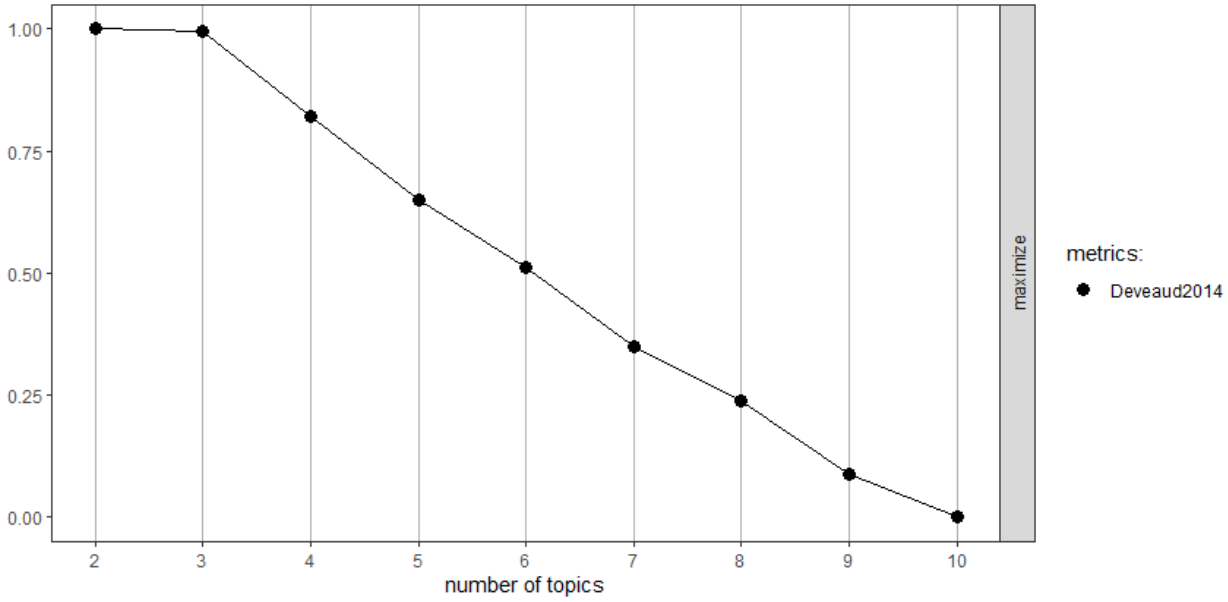


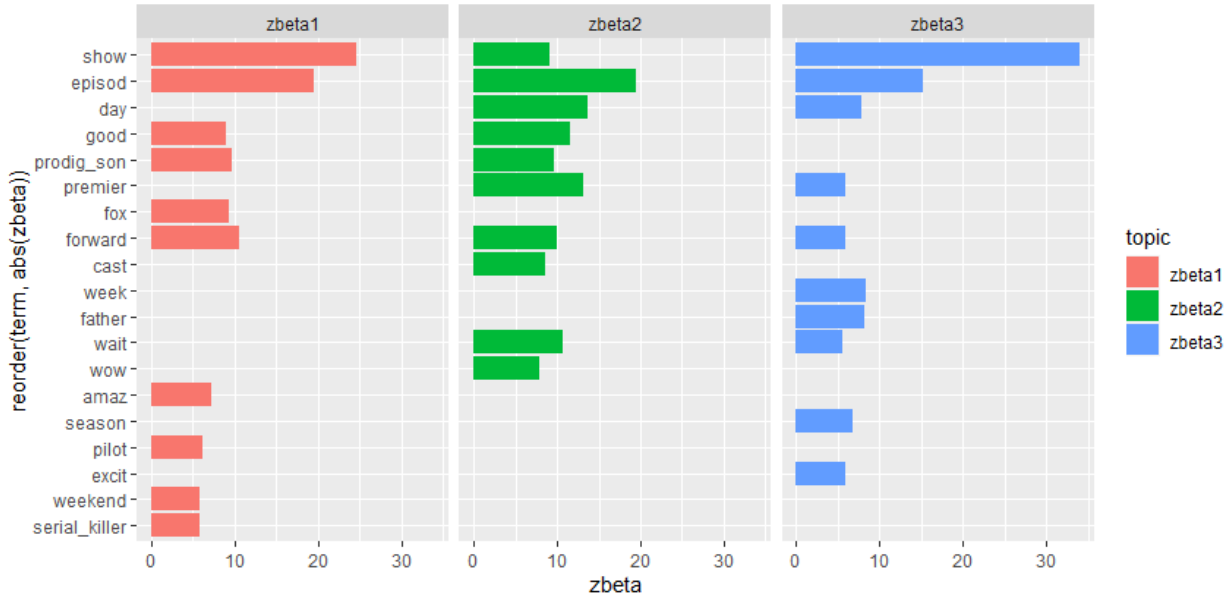
Nurses


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

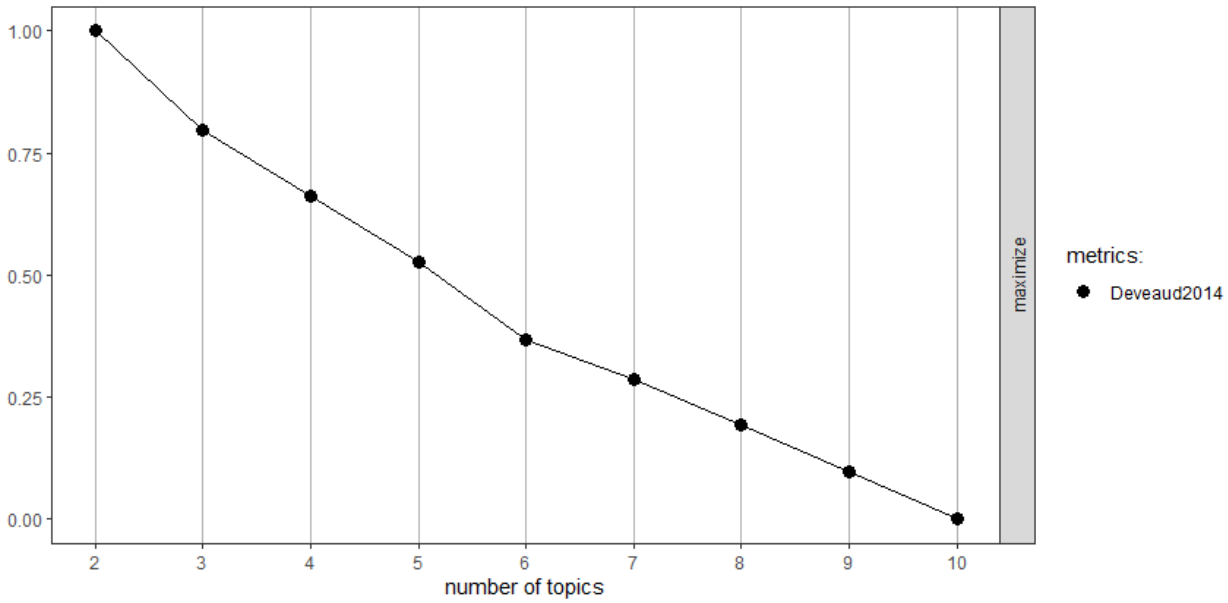

zbeta1
zbeta2
nurs
show
kid
wow
medic_drama
mad
originalm
way
good
damn_good
instagram
singl_filipino
dram
diahann_carrol
julia
nurs_theme
week
list
futur_nuryou
reorder(term, abs(zbeta))
0
2
4
6
8
zbeta
topic
zbeta1
zbeta2

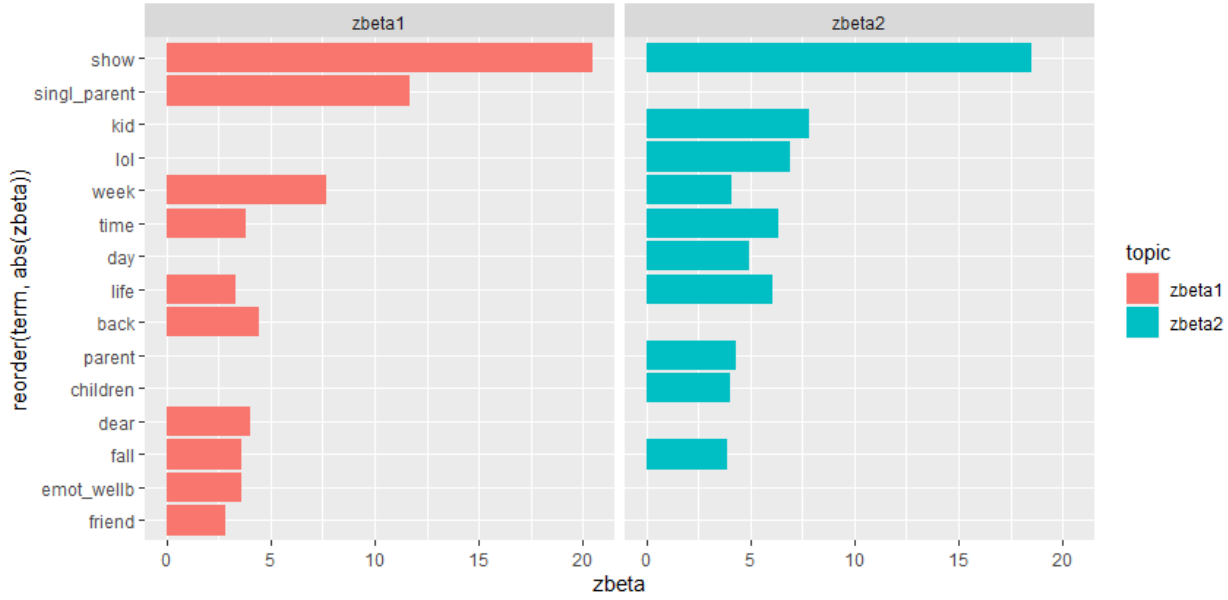
Prodigal Son



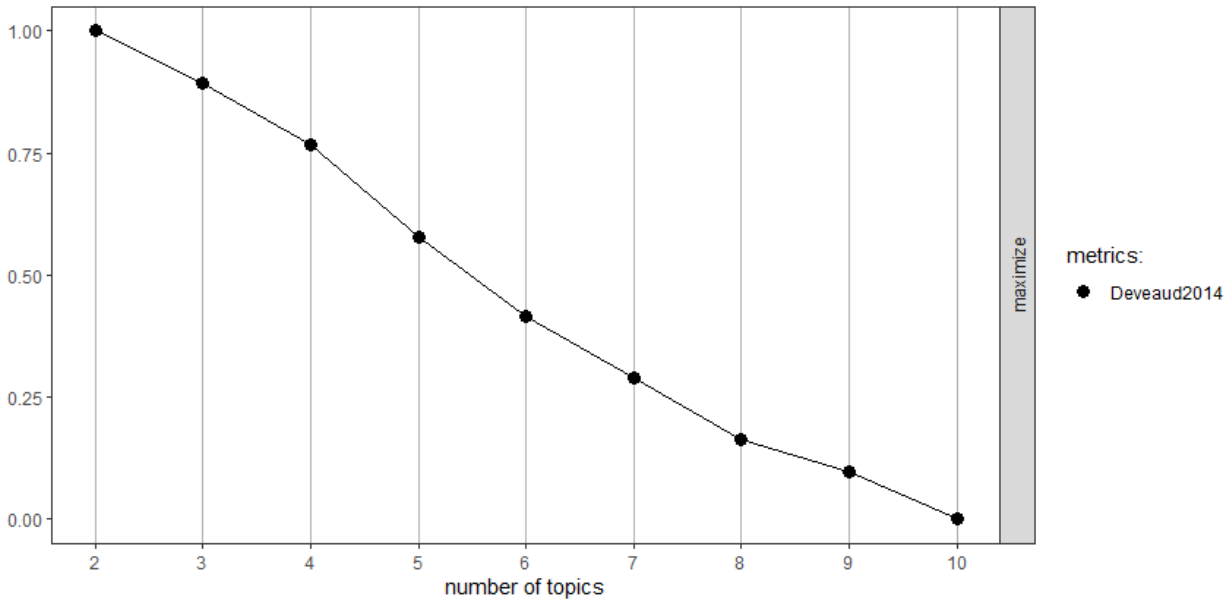


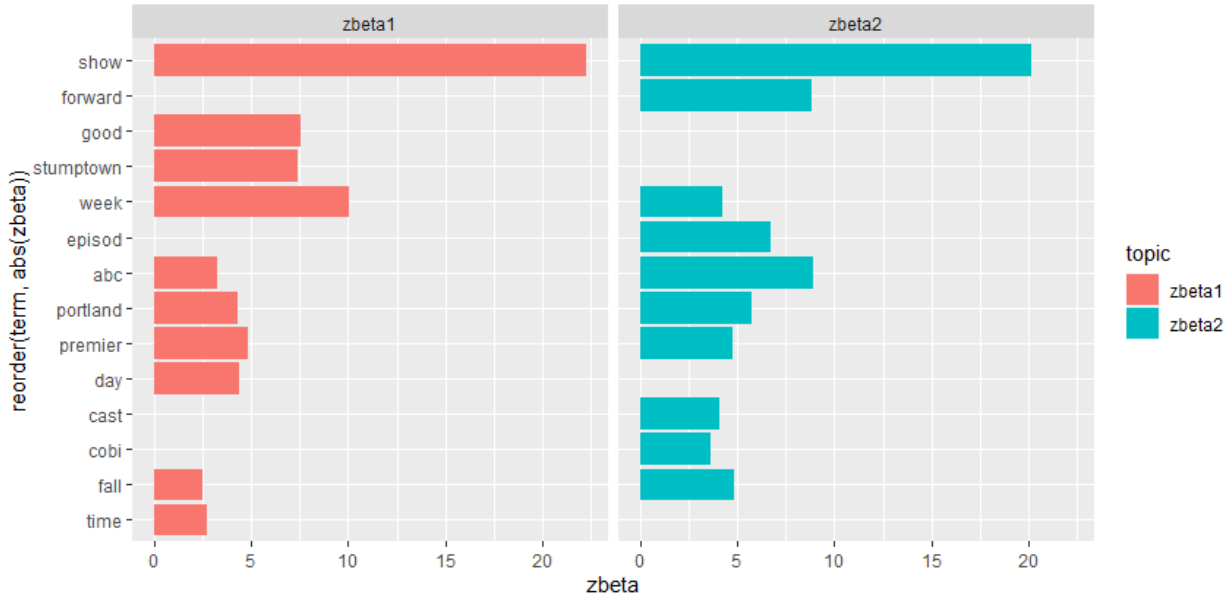
Single Parents



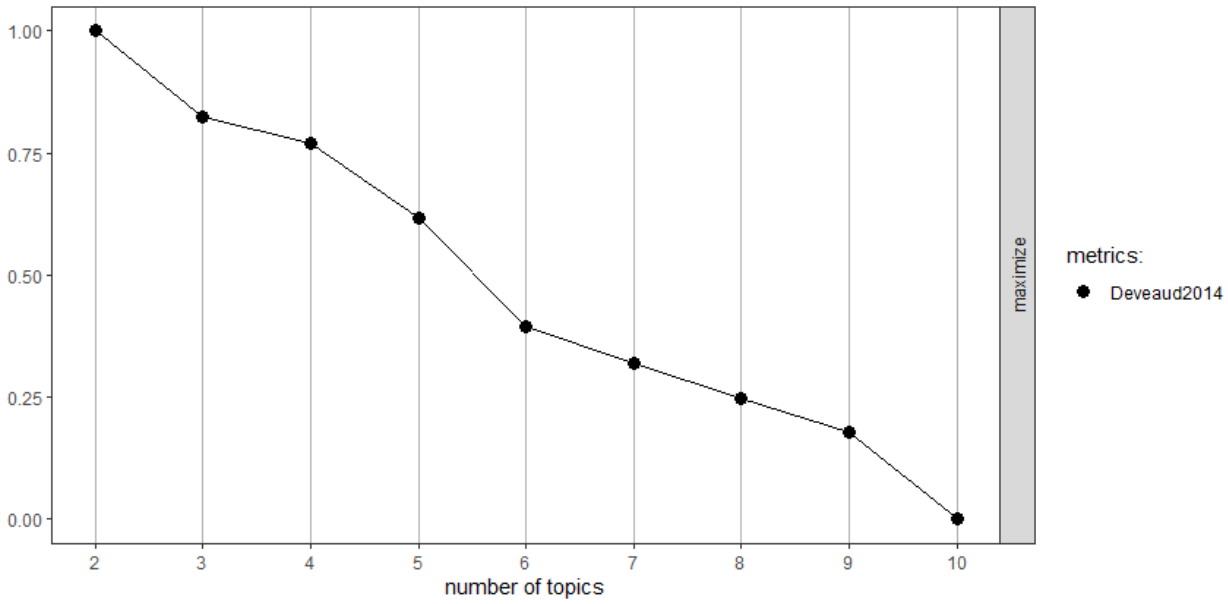


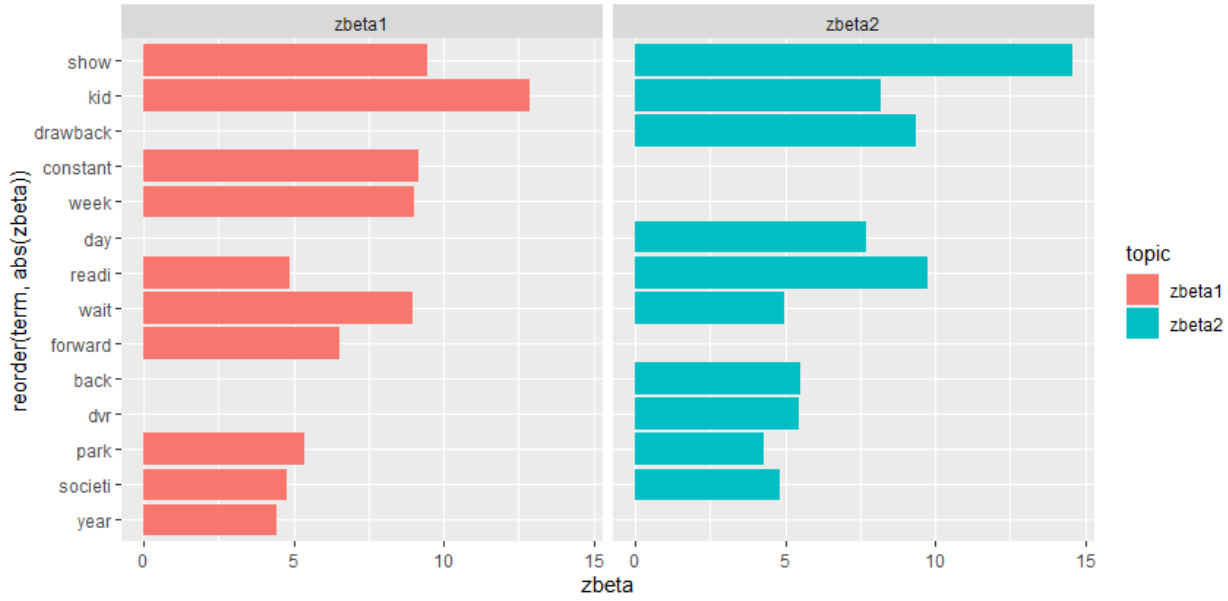
Stumptown


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014

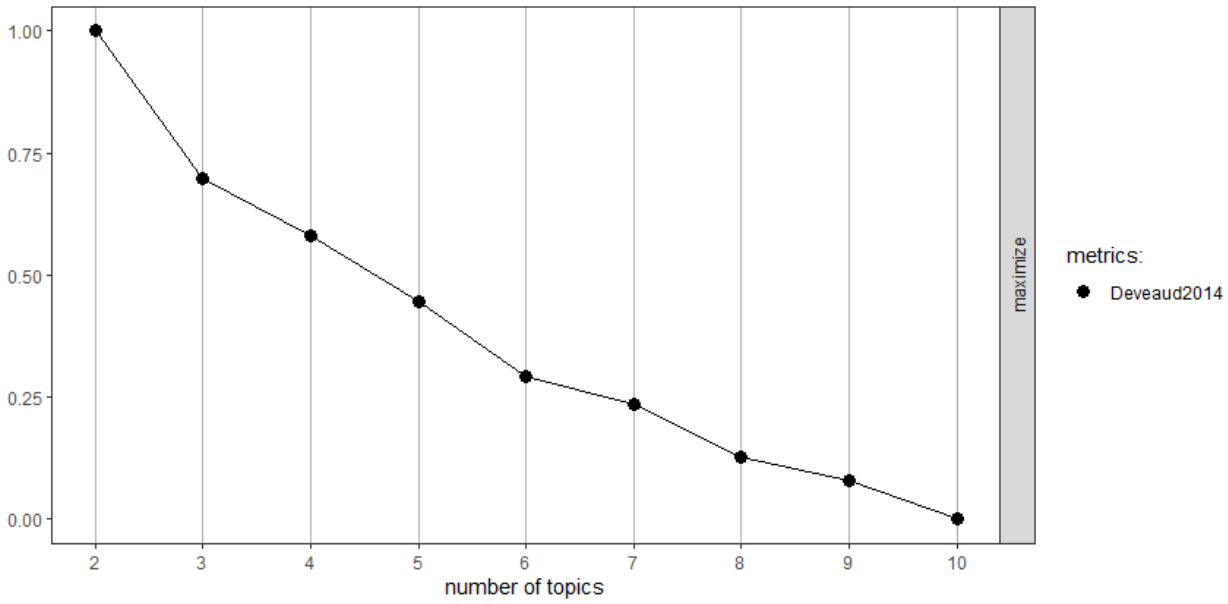

zbeta1
zbeta2
show
forward
good
stumptown
week
episod
abc
portland
premier
day
cast
cobi
fall
time
reorder(term, abs(zbeta))
0
5
10
15
20
zbeta
topic
zbeta1
zbeta2

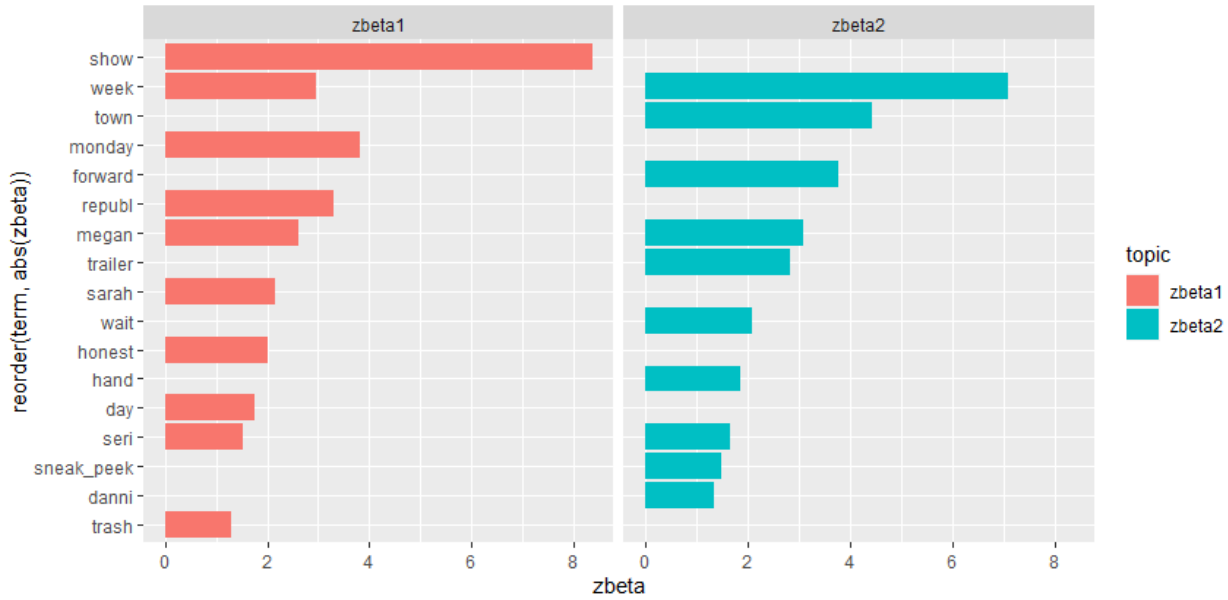
The Kids Are Alright



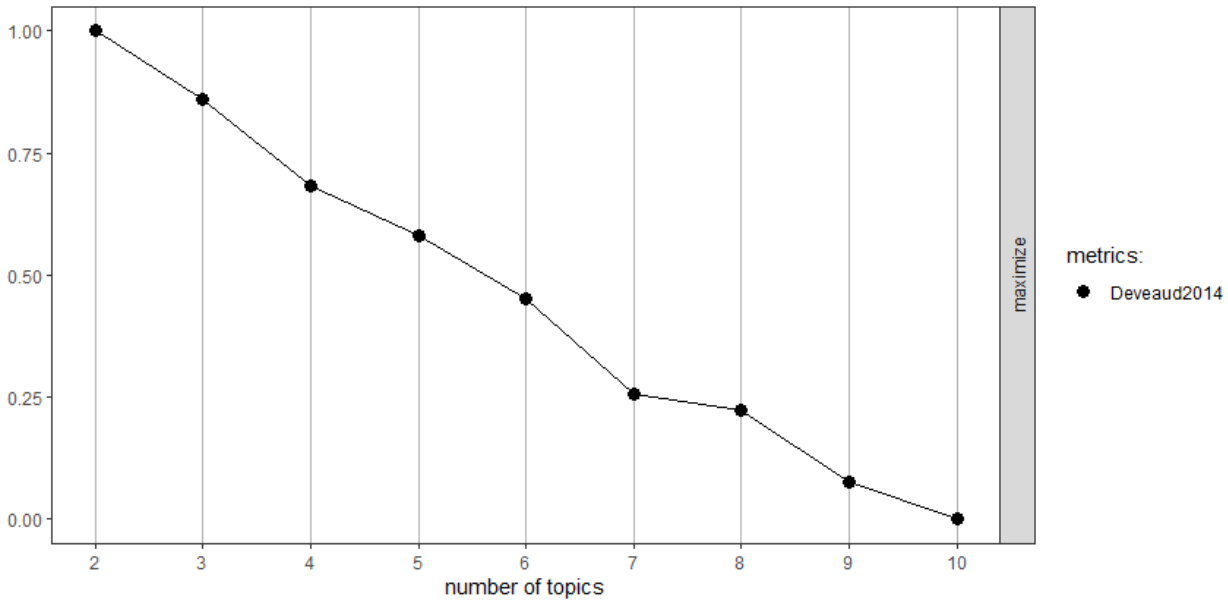


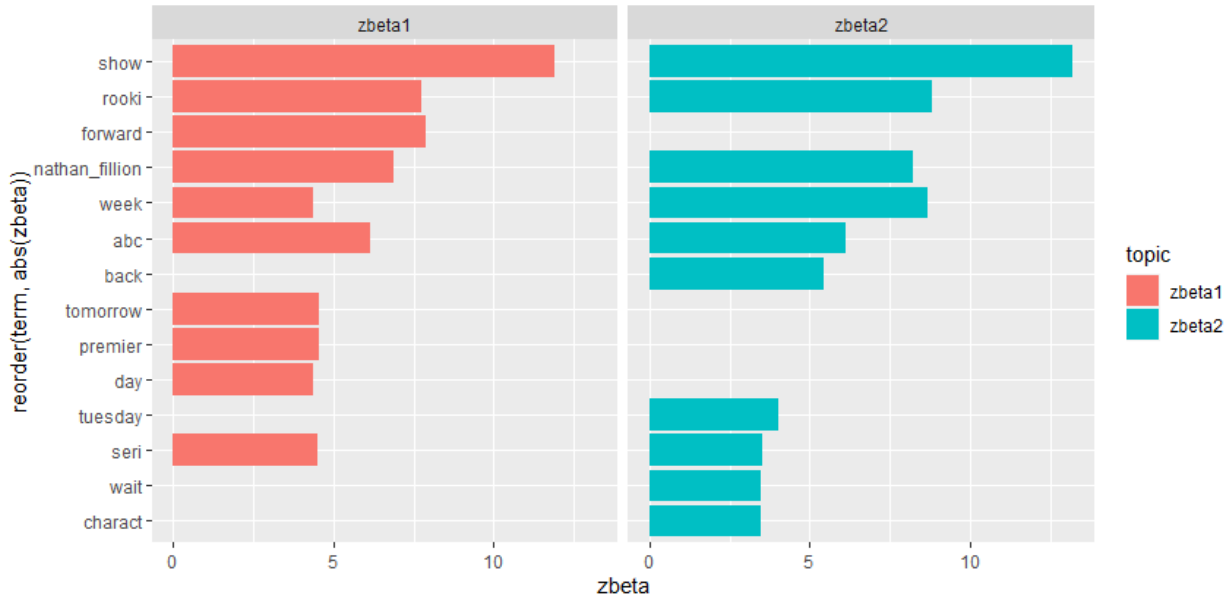
The Republic of Sarah



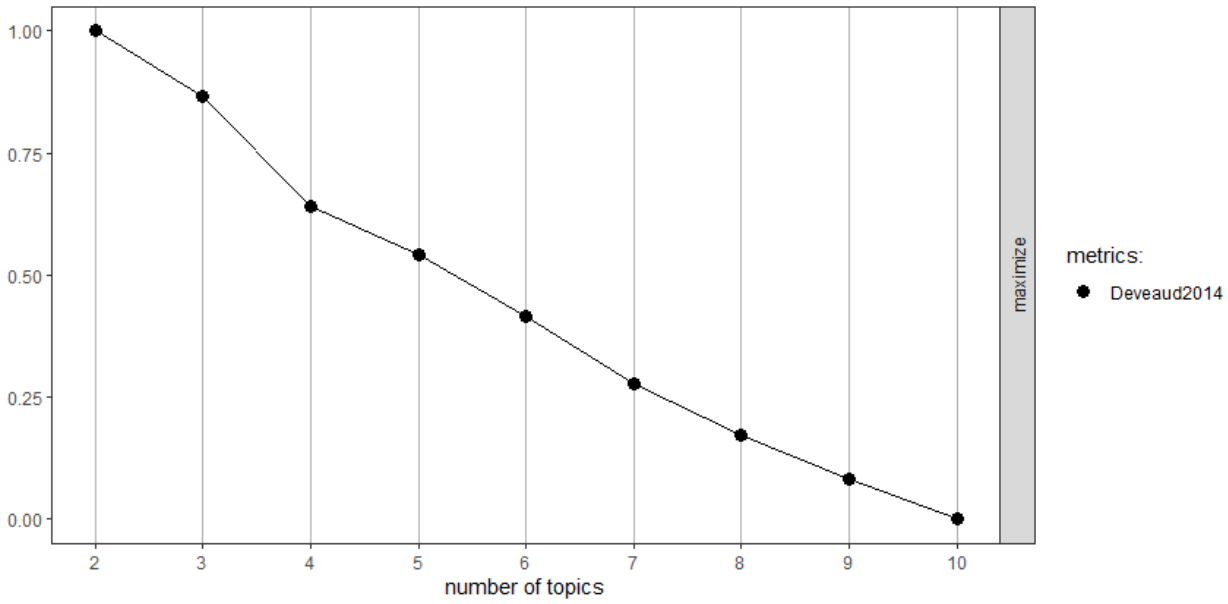


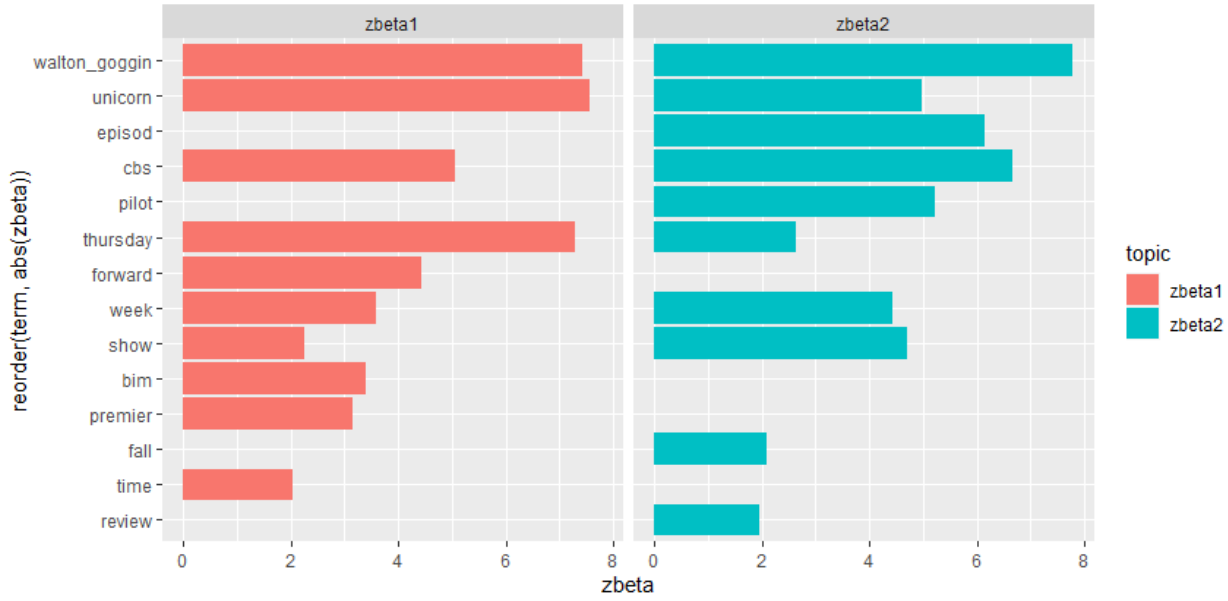
The Rookie


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014


zbeta1
zbeta2
show
rooki
forward
nathan_fillion
week
abc
back
tomorrow
premier
day
tuesday
seri
wait
charact
reorder(term, abs(zbeta))
0
5
10
zbeta
topic
zbeta1
zbeta2

The Unicorn


1.00
0.75
0.50
0.25
0.00
2
3
4
5
6
7
8
9
10
number of topics
maximize
metrics:
Deveaud2014


zbeta1
zbeta2
walton_goggin
unicorn
episod
cbs
pilot
thursday
forward
week
show
bim
premier
fall
time
review
reorder(term, abs(zbeta))
0
2
4
6
8
zbeta
topic
zbeta1
zbeta2

VITA

R. CASEY GOODMAN

SCIENTIST-PRACTITIONER
& BUSINESS ANALYTICS EXECUTIVE

casey.goodman@mindmeldinsights.com
(404) 405-6722
www.linkedin.com/in/rcaseygoodman
Atlanta, GA

Academic and professional background combining a Doctorate in Business Administration (DBA), with MA and BA focusing on Experimental Social Psychology, plus a 20-year career in Market Research, Analytics & Data Visualization. New analytic capabilities and best practices are developed and then deployed throughout business organizations. Natural mentor/coach driven to inspire students and colleagues to pursue academic and personal excellence via a challenging yet engaging environment.

PEER-REVIEWED RESEARCH AND APPLIED RESEARCH FOR CAPABILITIES DEVELOPMENT

Peer-Reviewed Publications and Presentations

- "Effect of media coverage, social media, advertising & prevention spending on e-cigarette use among US youth" by Arbogast, M.D., Goodman, R.C., Hernandez, A.L., Donthu, N. Presented at *2020 International Conference on Engaged Management Scholarship*.
- *Basic and Applied Social Psychology*: "Social Identity Orientation & Schema Maintenance" by Seta, Catherine E; Seta, John J; Goodman, R. Casey; 10/1998, Volume 20, Issue 4.

Applied Research for Capabilities Development

MindMeld Insights, LLC

- Developing capability to identify product innovation opportunities via mining social media or web scrape data (in Python) and applying keyword extraction, topic modeling and sentiment analysis (in R).
- Supported development and validation of *Chooseology*, a new consumer online shopping simulation for StandPoint, an agency specializing in product innovation; *Chooseology* used with traditional concept testing and pricing analytics.

Turner Broadcasting

- Spearheaded development and application of new proprietary tool, *TOPCAT (Turner Optimal Promotion Campaign Analysis Tool)*, to analyze promo ad campaigns across TV, digital, social, radio and cinema, to drive viewership of entertainment networks. *TOPCAT* was then deployed across Turner's other News & Kids networks as a best practice.

IPSOS (formerly Synovate)

- Presented at *2008 Global Conference* (Kuala Lumpur, Malaysia) "Unlocking the Power of Unilever's Product Testing Database."
- Presented at *2007 Global Conference* (Prague, Czech Republic) "Package Testing for Coca-Cola: A Case Study."
- Developed template for visualizing product test results, plus a mining procedure for global product testing databases. Both deployed globally across IPSOS as best practices.
- Worked with Product Design & Development Practice to improve analytic & text mining capabilities. This enhanced their value to clients globally.

Polaris Marketing Research

- Spearheaded text mining and structural equation modeling at Polaris. These were deployed across Polaris account teams.

Georgians for Children (non-profit)

- Wrote and published *Georgia KIDS COUNT Factbook* (state profile of child & family well-being). The publication was used as a reference for state legislators and lobbyists and cited by Georgia media.

Education and Academic Credentials

Doctoral Candidate, Executive Doctorate in Business Administration (DBA; May 2022)

Georgia State University, Robinson College of Business, Atlanta, GA

- GPA: 4.13 out of 4.0
- Dissertation (April 2022): "Consumer-Generated Social Media Posts to Improve Forecasting Models of Television Premiere Viewership: Extending Diffusion of Innovation Theory" Naveen Donthu, Advisor. Using Python for Twitter API data service; R for text analytics (topic modeling, sentiment analysis, keyword extraction), and statistical analysis.

Master of Arts in Psychology

Wake Forest University, Winston-Salem, NC

- GPA: 2.9 out of 3.0
- Sigma Xi – Scientific Research Honor Society
- Instructor: Research Methods & Statistics Laboratory
- Thesis: "The Moderating Effects of Self-Esteem on Self-Presentation: Interpersonal and Intrapersonal Processes" Mark R. Leary, Advisor.

Bachelor of Arts in Psychology, minor in Politics

Wake Forest University, Winston-Salem, NC

- Summa Cum Laude
- GPA: 3.8 out of 4.0
- President, Psi Chi Psychology Honor Society

Certifications

- R, Python, Web Scraping, Tableau, Network Analysis. Georgia State University Research Data Services. Atlanta, GA. Fall 2020
- Text Classification in R with quanteda. Coursera. June 2021.
- Teaching at the University Level. Georgia State University. Atlanta, GA. Fall 2021.

Practitioner Experience

MindMeld Insights, LLC (Research Consulting), Atlanta, GA — 2015 – Present

Principal & Chief Strategist

Established MindMeld Insights (mindmeldinsights.com), a market analytics consulting practice that optimizes intelligence gained from marketing research & analytics to drive business growth. Clients include research suppliers, consulting agencies, corporate marketing insights teams, and university-based programs.

The Goodman Organization, LLC (Fitness Consulting), Atlanta, GA — 2017 – Present

Owner

Established fitness and nutrition consulting practice to guide and inspire clients to achieve their best results, while addressing their individual goals and fitness challenges. National Academy of Sports Medicine (NASM) Certified Personal Trainer (considered a premier fitness industry certification).

Turner Broadcasting, Atlanta, GA **2011 – 2014**
Senior Director, Strategic Entertainment Research

Built a new Strategic Research function at TNT & TBS to support Turner's Business Strategy Team. The research budget grew to nearly $1 million. Routinely worked with Chief Marketing Officer, Chief Strategy Officer, and their teams. Leveraged custom research & statistics, plus Nielsen & other syndicated media measurement, to inform Marketing, Strategy, and Programming for Turner Entertainment Networks, including digital, mobile, social media, VOD, and streaming extensions.

- Conducted team training on Quantitative & Qualitative Research Methods & Statistics

Marketing & Planning Systems, Boston, MA **2010 – 2011**
Principal

Spearheaded key accounts Wal-Mart, and Sam's Club via a suite of programs tracking retail brand equity and brought a new account, CIBA Vision. Used consultative approach when working with clients to frame business issues and design research. Presented business implications derived from statistical analysis of results.

IPSOS *(formerly Synovate)*, New York, NY **2004 – 2010**
Vice President (2008 – 2010)
Account Group Manager (2005 – 2008)
Account Executive (2004 – 2006)

Led a diverse team of project managers and analysts (several worked in remote offices) in providing high-quality service, reports, and presentations on time and within budget. Functioned as key US-based relationship manager for Unilever, The Coca-Cola Company, Florida Power & Light, Philips Consumer Lifestyle, Georgia Pacific, CIBA Vision, and Turner Studios.

- Invited to present Lunch-and-Learns for Unilever and Colgate. *Male Beauty: A US and Global Perspective.*

Polaris Marketing Research, Inc, Atlanta, GA **1999 – 2004**
Account Director (2003 – 2004)
Senior Project Manager (2001 – 2003)
Project Manager (1999 – 2001)

Led team of managers, programmers, analysts, and interns to optimize client service & efficiency. Presented business implications from Consumer and B2B satisfaction and brand studies. Lead for key accounts in healthcare, financial services, manufacturing, technology, and media/entertainment.

- Conducted internal training on Quantitative Research Methods & Statistics
- Worked with relational databases (e.g., SQL, dBase, Paradox, FoxPro)

Georgians for Children (non-profit), Atlanta, GA **2011 – 2014**
Research Director

Headed statewide project to track health, education, and social statistics to measure the status of children in Georgia. Worked with vendors to analyze data and publish findings. Generated research cited by the media and used in testimony before legislative committees.

- Conducted workshops throughout the state on the use of secondary data

Analytics & Research Techniques

Machine Learning, Descriptive & Predictive Modeling

Decision Trees and Random Forests, Latent Dirichlet Allocation (LDA), Ensemble Modeling.

Statistical Analysis, Optimization & Simulation

Sentiment Analysis, Structural Equation Modeling, Data Fusion, Meta-Analysis, Psychometric Assessments (e.g., reliability, validity), Discrete Choice and Conjoint, Max-Diff, Segmentation, Perceptual Mapping, Penalty Analysis, TURF, Kano, Forecasting.

Quantitative & Qualitative Research Techniques

Ad Hoc & Tracking studies. Advertising Effectiveness, Brand Image/Positioning, Satisfaction, Attitude & Usage, Concept/Product, User Experience, Diversity, Emerging Markets, Emotional/Motivational, Shopper Insights (including Path-to-Purchase). Home Use, Central Location, and Online Community methodologies. Omnibus, Telephone, IVR, and Panel data collection. Focus Groups, in-Depth Interviews, Ethnographies, Online Qualitative.

Syndicated Data Tools

Nielsen Media Research television ratings (including on-demand and DVR), Nielsen Audio, Nielsen Catalina, Nielsen Buyer Insights, Nielsen VideoCensus, Rentrak, MRI, Keller Fay, Omniture, comScore.

Volunteer & Service

International Conference on Engaged Management Scholarship. Peer reviewer of manuscript submissions for 2020 conference.

Nic's Animal Sanctuary (Shiner, TX)

- Vice President and board member.
- Promoted sanctuary's free spay and neuter program to rural residents of the county via advertising and PR outreach efforts.