

# Longitudinal Analysis of TV Series Data

Siyue Gao, Yuyan Liu & Miten Shah

Department of Statistics, Rice University, Houston TX

## 1 Introduction

TV Series (a.k.a. TV Shows) are a widely known and consumed form of entertainment. We wish to analyse the constantly changing aspects of TV Series over time, be it genre, recurring themes, or impact of streaming-networks.

A TV series is basically divided in to multiple seasons and each season consists of a finite number of episodes. In general, there is a new season each year. A TV series consists of a lead cast, supporting cast members and recurring cast members. The TV series usually airs on a weekly basis on a particular network (such as NBC, Netflix and BBC) on a particular timeslot. The genre of the TV series can span from drama, mystery and thriller to light-hearted comedies.

We have a huge dataset with TV series data ranging from early 1900s to the one aired last week. We wish to hence, analyse the data longitudinally, i.e., through the years. There are many questions at hand that we think we can answer with this project using R[1].

We have divided this report into two main sections:

- Over the Years: This section will be a broad comparison of TV series in general and how they have developed across the years.
- TV Networks in the Past Decade: This section will compare the prominent TV Networks based on their type.

## 2 The Data

The source for our database has been [TheTVDB.com](http://TheTVDB.com)[2]. We used the JSON API to download the datasets and stored them as csv files on the local drive. The data was cleaned to remove any NULL/(empty) placeholder strings from the files and to remove unidentified unicode segments using RegEx. We then imported them into PostgreSQL[3]. For Emmy data, we scraped the data from Wikipedia and did some cleaning in R using RegEx to be able to easily extract actor/network/series from the data.

### 2.1 The Series

We have 54498 TV series data.

Variable	Definition
seriesName	Name of the series
status	Whether the series is running or ended
firstAired	Date of first airing
network	On which network is / was the series running
runtime	Approximate Length in minutes of each episode of the series
genre	Genre which the series deals with
overview	A brief summary of the series
airDayOfWeek	Day of the week on which the series airs.
airsTime	Time of the day on which the series airs.
rating	Viewers advisory rating
siteRating	Average user rating of the series on TVDB
siteRatingCount	Count of users who rated the series on TVDB

## 2.2 The Episodes

We have 2017171 episodes data.

Variable	Definition
seriesID	ID of the series of which the episode was part of
airedEpisodeNumber	The $n^{th}$ episode of the season
airedSeason	The season number in which the episode takes place
episodeName	Name of the episode
firstAired	Date on which the episode was first aired
overview	A brief summary of the episode

## 2.3 The Actors

We have 156807 actors data.

Variable	Definition
seriesID	ID of the series of which the actor was part of
name	Name of the actor
role	Name of the role played by the actor in the series
sortOrder	The importance of the actor on the show (0 = lead role, 1 = main cast, 2 = recurring character ...)

## 2.4 The Emmys

We have Primetime Emmy Awards data[4] spanning from 1949 to 2017. It consists of the following categories:

- Programs
  1. Outstanding Comedy Series
  2. Outstanding Drama Series
  3. Outstanding Variety Series
- Acting
  1. Outstanding Lead Actor in a Comedy Series
  2. Outstanding Lead Actor in a Drama Series
  3. Outstanding Lead Actress in a Comedy Series
  4. Outstanding Lead Actress in a Drama Series

### 3 Over the Years

#### 3.1 Contemporary Themes

##### 3.1.1 From TV Series Overviews

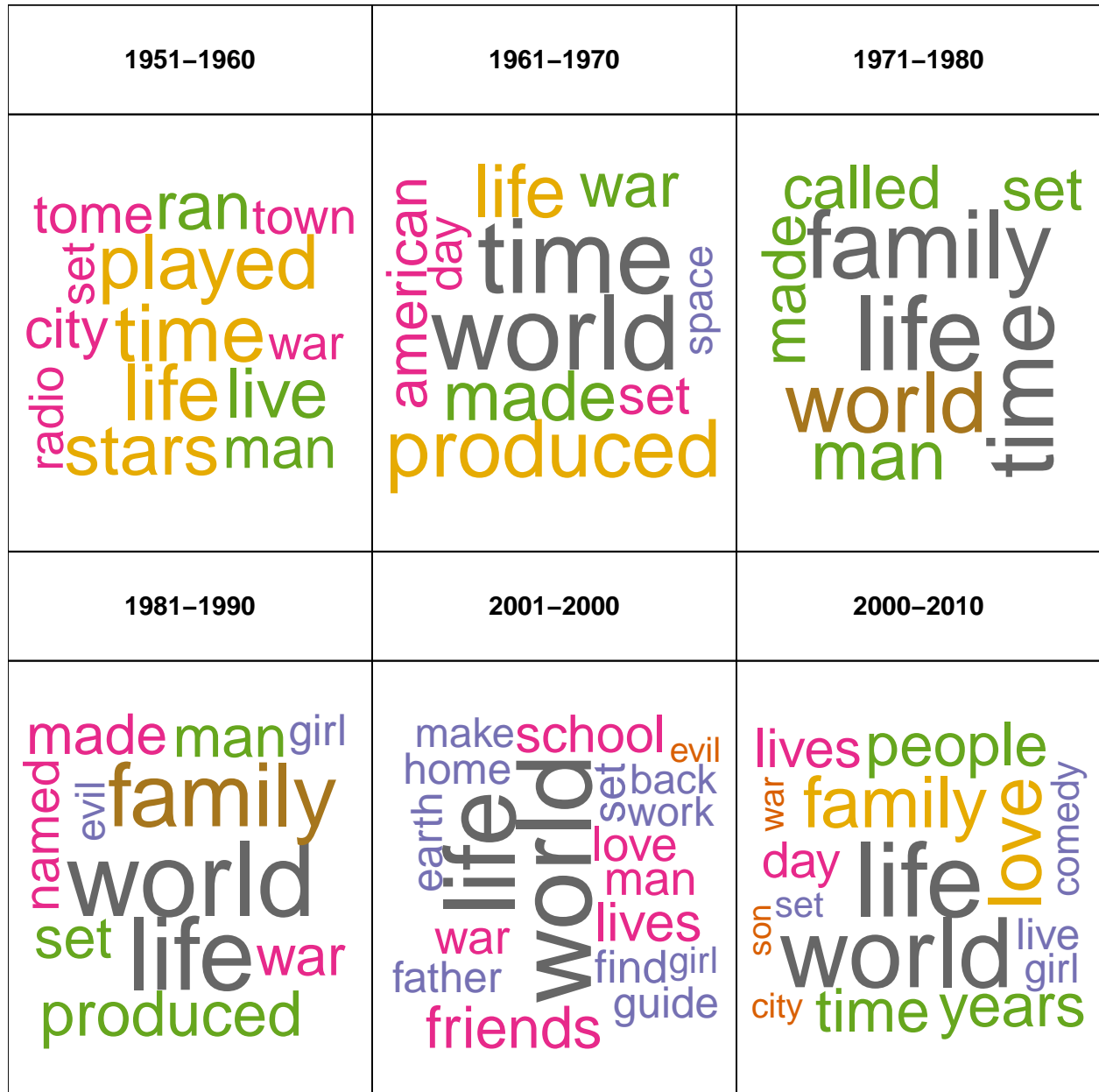


Figure 1: Prominent Themes Across the Years

From Fig. 1, "War" and "Life" are the common theme across all the decades. "World" and "Family" are also quite prominent since the 60s.

### 3.1.2 From TV Episodes Overviews

TABLE 1 Top Words Occurring in Episodes Overviews

1951-1960	Fred - Barney - Wilma - Man - Freds - Betty - Home - Flintstones - Time - Back - Twilight - Zone - Pebbles - Wife - Rubbles - Bedrock - Finds - Day - Serling - Slate
1961-1970	Batman - Max - Hogan - Robin - Kaos - Chief - Plan - Control - Colonel - Gotham - Enterprise - Agent - City - Kirk - Find - German - Back - Gang - Men - Klink
1971-1980	Sketches - Include - Live - York - Thomas - Hawkeye - Performs - Charles - Man - Magnum - Performed - Laura - Higgins - Find - Back - Time - Love - Muppet - Home - Weekend
1981-1990	Jerry - George - Segment - Movie - Garfield - Elaine - Kramer - Sam - Joel - Man - Film - Made - Back - Find - Finds - Murder - Plot - Woman - Sol - Original
1991-2000	Discusses - Back - Find - Fran - Time - Takes - Carrie - Finds - Frasier - Harm - Jon - Man - Decides - Ray - Doug - Team - Carter - House - Hes - Game
2001-2010	Team - Find - Teams - Stephen - Case - Michael - Family - Back - Finds - Time - Life - Murder - Make - Takes - Man - Jim - Home - Guests - Work - Mike

We narrowed down the episodes data to remove shows with rating less than 5 and rating count less than 25. From Table 1, we can see the following:

- In the 50s, the most common words come from the characters and themes in "The Flintstones" and "The Twilight Zone".
- In the 60s, the most common words come from the characters and themes in "Batman" and "Star Trek".
- In the 70s, the most common words suggest inception of "Saturday Night Live" from words like "sketches", "York", "perform", "Higgins", "weekend".
- In the 80s, the most common words come from the characters and themes in "Seinfeld".
- Since the 90s, the common words do not suggest any particular TV shows.

## 3.2 Genres

Rank	1950-60	1960-70	1970-80	1980-90	1990-00	2000-10	2010-20
1	Drama	Comedy	Drama	Drama	Comedy	Drama	Documentary
2	Adventure	Animation	Comedy	Comedy	Drama	Documentary	Drama
3	Action	Drama	Children	Animation	Animation	Comedy	Comedy
4	Comedy	Children	Animation	Children	Documentary	Animation	Reality
5	Western	Adventure	Adventure	Action	Children	Reality	Animation
6	Children	Action	Action	Adventure	Adventure	Action	Romance
7	Documentary	Documentary	Mini-Series	Mini-Series	Action	Adventure	Special Interest
8	Animation	Science-Fiction	Science-Fiction	Documentary	Science-Fiction	Special Interest	Action
9	Science-Fiction	Family	Documentary	Science-Fiction	Mini-Series	Children	Mini-Series
10	Family	Crime	Family	Fantasy	Fantasy	Mini-Series	Crime

Table 2: Top genres over the decades

Table 2 suggests Drama and Comedy have stood the strongest across time along with Animation & Childrens' TV series. We can observe the peak of the Western genre in the 1950s. Documentary and Reality TV series

have risen to the top in the new millennium. Surprisingly, not many Romance TV series were produced until the current decade.

### 3.3 Actors

#### 3.3.1 Prolific Actors

Actor	Number of series
Frank Welker	129
Takahiro Sakurai	87
Kana Hanazawa	78
Yui Horie	71
Mamiko Noto	69
Takehito Koyasu	68
Daisuke Namikawa	67
Miyuki Sawashiro	66
Jun Fukuyama	66
Don Messick	62

Table 3: Industrious Actors

Table 3 shows that most prolific actors are technically voice-actors either from America or Japan.

#### 3.3.2 Actors working in Highly-Rated Series

Actor	Number of series	Average Rating
Leo Laporte	12	9.66
Shaun Micallef	11	9.10
Laura Bailey	13	8.82
Tooru Ookawa	16	8.71
Ogura, Yui	11	8.67

Table 4: Actors working in highly-rated shows ()

For Table 4, we only look at actors here have worked in at least 5 TV series and the average siteRatingcount is at least greater than 10. We again observe dominance of Japanese voice-actors.

#### 3.3.3 Diverse Actors

Actor	Diversity
Takehito Koyasu	17
Takahiro Sakurai	17
David Tennant	17

Table 5: Most diverse actors

For Table 5 we observe that David Tennant, of "Doctor Who" and "Broadchurch" fame, is the among the most diverse actors along with two other Japanese voice actors, with 17 different genres.

### 3.4 Awards

#### 3.4.1 Primetime Emmy for Lead Actors

Actor	Number of Series	Average Series Rating	Average Series Rating Count	Emmys
Julia Louis-Dreyfus	8	5.25	51.88	7
Mary Tyler Moore	8	2.85	3.00	6
Candice Bergen	3	5.53	31.00	5
Jane Wyatt	1	0.00	0.00	5
Lucille Ball	6	4.97	4.17	5

Table 6: Actors with most lead Primetime Emmy Awards

Table 6 shows the actors with most number of Lead Emmys along with the number of shows they have done, average rating of those TV shows and the average rating count for each show. Notice all are females.

#### 3.4.2 Primetime Emmy for TV Series

Series	Series Rating	Series Rating Count	Emmys
Late Show with David Letterman	7.90	17	6
Frasier	9.20	85	5
Modern Family	8.80	310	5
All in the Family	8.40	8	4
Cheers	8.30	48	4

Table 7: Shows with most Primetime Emmy Awards

Table 7 shows the series with most number of Lead Emmys along with their average rating and the average rating count. Most of them are family-based comedies.

## 4 TV Networks in the Past Decade

We are looking specifically into the US TV Networks in the last decade (2007-2017) for the fact that Streaming Networks only arrived to the scene in this time and Cable Networks have created some of the top shows to compete to the ever-present Broadcast Networks. We are particularly looking into the following network representation based on popularity, Nielsen ratings and their propensity of winning Emmys:

- Broadcast Networks
  1. ABC
  2. CBS
  3. NBC
- Cable Networks
  1. AMC
  2. Comedy Central
  3. HBO
- Streaming Networks
  1. Amazon
  2. Hulu
  3. Netflix

## 4.1 Production

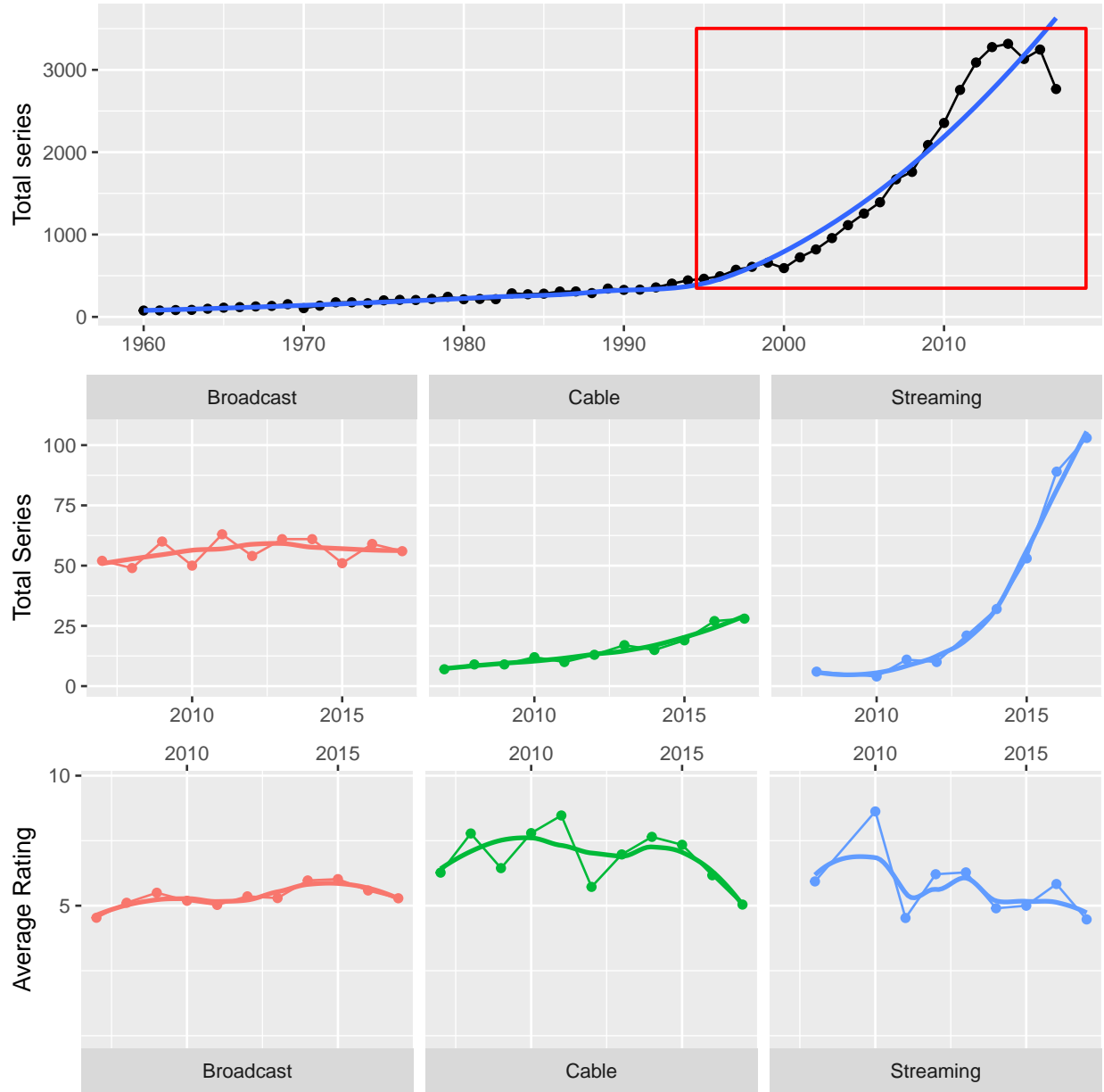


Figure 2: TV Networks (Production)

We can see in Fig. 2 (top), the highlighted region is many times referred to as the New Golden Age of Television[5]. From Fig. 2 (bottom-right) point out that there is a streaming boom in the last decade. More shows seem to be produced by the major streaming networks compared to their conventional counterparts. The conventional networks have been producing new shows on an almost constant levels. From Fig. 2 (bottom-left) we can establish that the streaming networks may have started out with high ratings but are converging to the similar ratings as the conventional networks.

## 4.2 Themes

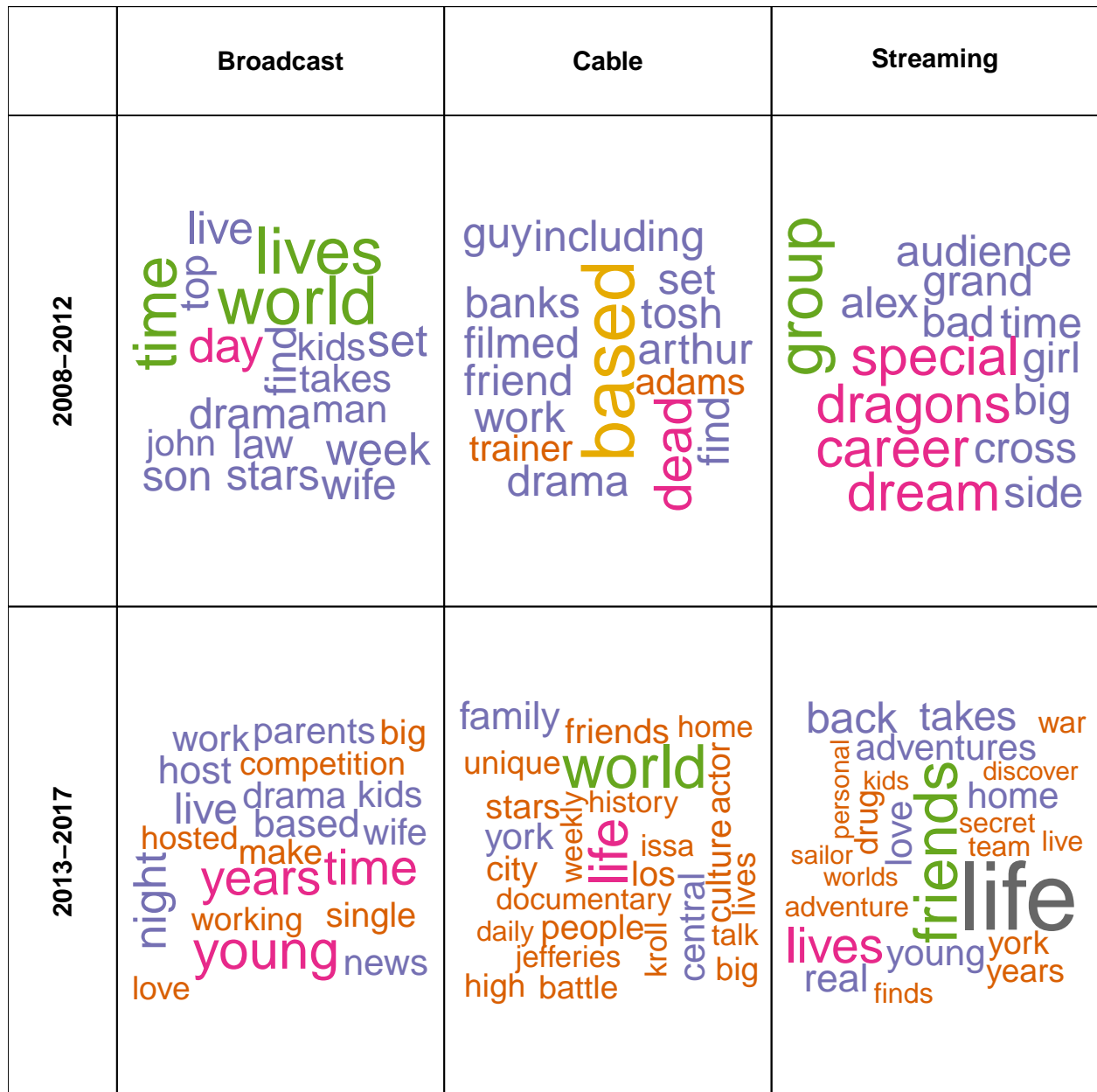


Figure 3: TV Networks (Themes)

From Fig. 3, as we move from left to right, we observe how the themes change from family-friendly themes in Broadcast Networks to themes like "sex", "death", "dark" and "evil" in Streaming Networks. There are many factors including the target age-group and censorship issues that are involved in such stark contrasts amongst the networks. Notice, even in the last decade, how the themes have evolved within each network. ]

## 4.3 Emmy Awards

We are looking into Emmy Awards since they are considered the Oscars of TV Shows.



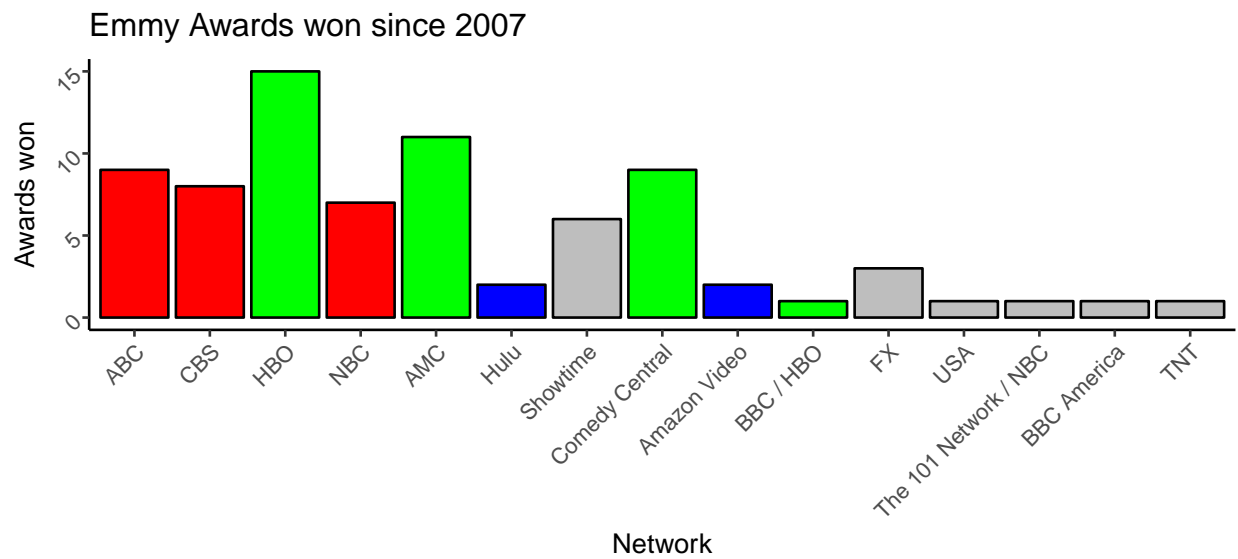
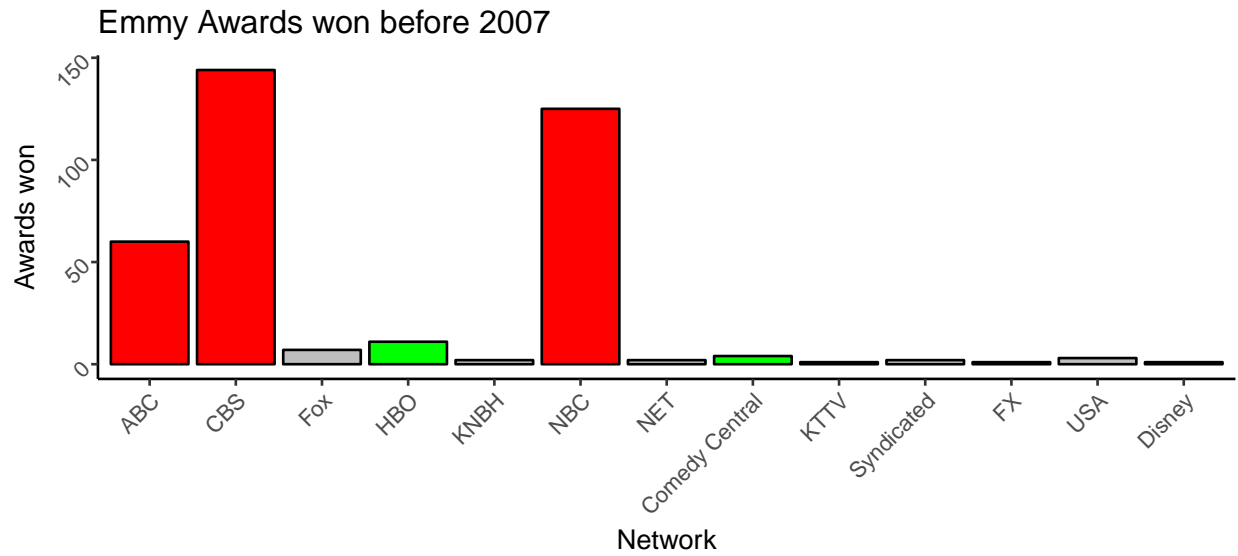


Figure 4: TV Networks (Emmy Awards)

We can see from Fig. 4 an obvious rise in the quality of cable networks especially in the past decades winning majority of the Emmys.

## 4.4 Genre

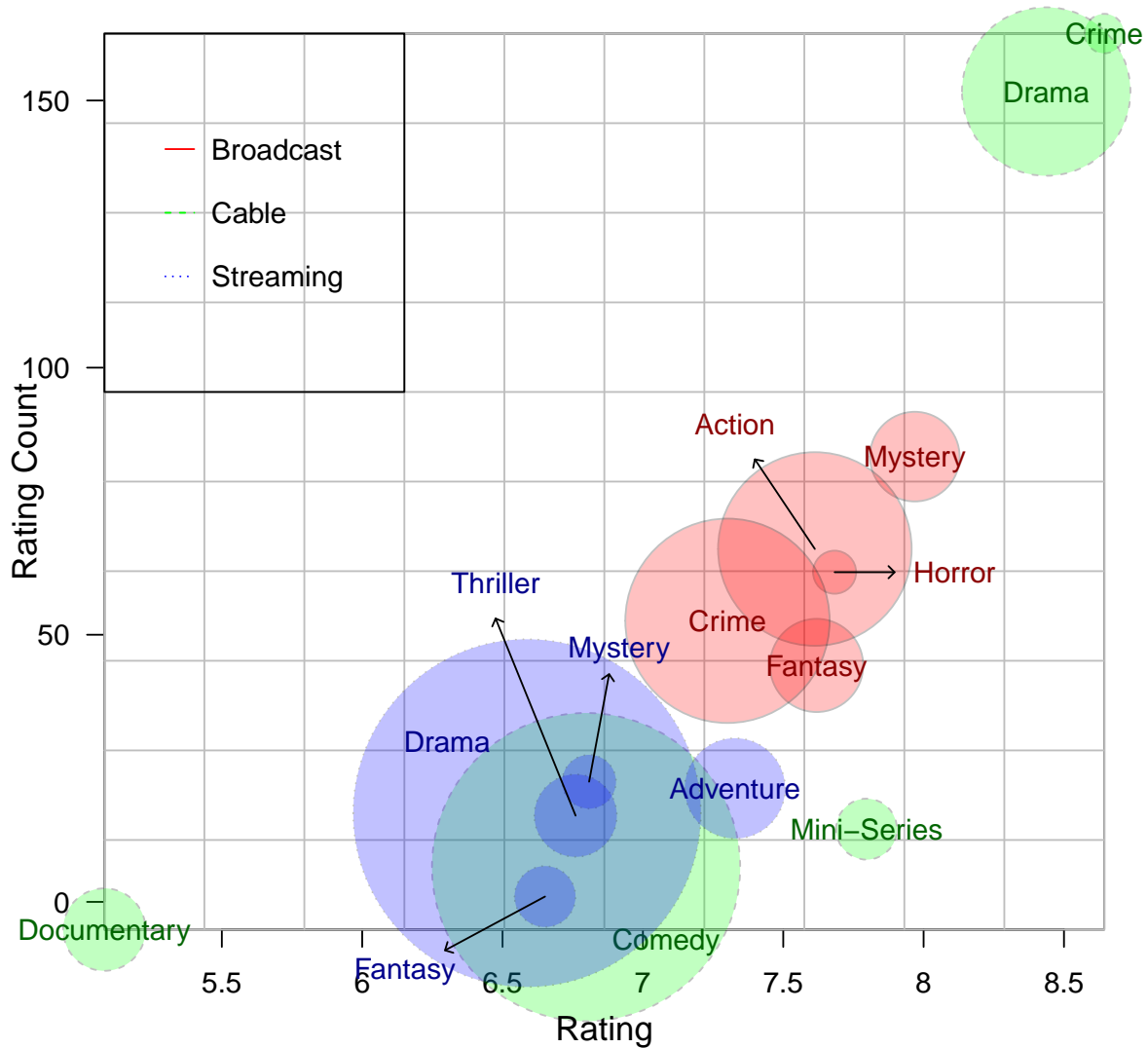


Figure 5: TV Networks (Genres)

We have subsetting the data to have top 5 genres based on the rating with at least 5 series produced for that genre. We have plotted the genres as circles across rating and rating count as in Fig. 5. The radii of the circles signify the genre count of the series. Broadcast Networks produce few shows in genres other than Drama and Comedy, but those shows are very highly rated. Crime and Drama shows produced by Cable Networks are the most well-regarded subset of TV series. There's not much difference in the rating across genres among Streaming Networks.

## 5 Remarks

Based on the dataset we possess, we thought it would be best to adopt visual story telling. Variables like *airsTime* and *rating* had sparse and inconsistent data and were not used. We are taking the variables like *siteRating* and *siteRatingCount* with a bit of skepticism since it's not as popular and reliable as IMDB, RottenTomatoes or Metacritic. For the popular shows at least, the ratings are quite similar to those sites.

In many cases when we had to join the Emmy Data (which is in CSV format) and SQL data, there might be not have been complete merging due to differences in the names of TV Series or Actors.

## 6 Conclusion

Across the time, the most common themes were "life" and "war", which is quite poetic. Japanese voice-actors dominate in terms of diversity, prolificacy and quality of the shows they are part of. The production for Broadcast Networks can be considered constant, Cable Networks can be considered linear and Streaming Networks exponential. Genres such as Crime, Mystery and Fantasy have been more well-received compared to other genres. Broadcast Networks serve all age-groups and hence often portray family-friendly themes however shows on Streaming networks having fewer restrictions can delve into wider variety of themes.

## References

- [1] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017. URL: [\url{https://www.R-project.org/}](https://www.R-project.org/).
- [2] TheTVDB.com. *TheTVDB.com: An open database for television fans*. <https://api.thetvdb.com/swagger/>. 2017.
- [3] PostgreSQL Global Development Group. *PostgreSQL*. <https://www.postgresql.org/>. 1996.
- [4] Wikipedia. *List of Primetime Emmy Award winners — Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/wiki/List\\_of\\_Primetime\\_Emy\\_Award\\_winners/](https://en.wikipedia.org/wiki/List_of_Primetime_Emy_Award_winners/). 2017.
- [5] Todd Leopold. *The new, new TV golden age*. Ed. by CNN. [Online; posted 6-May-2013]. 2013. URL: [\url{http://www.cnn.com/2013/05/06/showbiz/golden-age-of-tv/}](http://www.cnn.com/2013/05/06/showbiz/golden-age-of-tv/).