

Data Science of Pink Floyd

Renfei Mao*

2022-06-01

*renfeimao@gmail.com

Table of contents

Preface	3
1 Albums	4
1.1 Albums	4
1.2 Inclusion Criteria	5
1.3 Members	6
1.4 Album Lengths	8
1.5 Soundtrack Albums	12
1.6 Collaborative Albums	14
1.7 Productivity	16
2 Songs	20
2.1 Songs	20
2.2 Inclusion Criteria	21
2.3 Song Counts	22
2.4 Song Lengths	23
2.5 Authorship	26
2.6 Instrumentalness	27
2.7 Whose Pink Floyd?	29
3 Popularity	36
3.1 Spotify Dataset	36
3.2 Pink Floyd	38

Preface

This is a work-in-progress book.

The basic idea is to collect and analyze data about Pink Floyd with R data science tools.

The code in the book may seem redundant. This is on purpose so that each code block is self-contained.

1 Albums

```
library(stringr)
library(knitr)
library(dplyr)
library(ggplot2)
library(tidyr)
library(lubridate)

source('data/albums.R')
```

In this article, we will explore the metadata of the albums of Pink Floyd and its members.

1.1 Albums

The sourced data frame `albums` contains the albums we are going to explore:

```
albums |>
  # convert the column names to title case
  rename_with(str_to_title) |>

  # show the data frame in a better format
  kable()
```

Table 1.1: Pink Floyd and Its Members' Albums Based on Some Criteria

Year	Artist	Title	Length
1967	Pink Floyd	The Piper at the Gates of Dawn	41:54
1968	Pink Floyd	A Saucerful of Secrets	39:25
1968	Pink Floyd	The Committee	15:45
1969	Pink Floyd	More	44:57
1969	Pink Floyd	Ummagumma	86:32
1970	Pink Floyd	Zabriskie Point	36:22
1970	Pink Floyd	Atom Heart Mother	52:06

Year	Artist	Title	Length
1971	Pink Floyd	Meddle	46:48
1972	Pink Floyd	Obscured by Clouds	40:08
1973	Pink Floyd	The Dark Side of the Moon	43:09
1975	Pink Floyd	Wish You Were Here	44:11
1977	Pink Floyd	Animals	41:40
1979	Pink Floyd	The Wall	81:08
1983	Pink Floyd	The Final Cut	43:14
1987	Pink Floyd	A Momentary Lapse of Reason	51:09
1994	Pink Floyd	The Division Bell	66:23
2014	Pink Floyd	The Endless River	53:02
1970	Syd Barrett	The Madcap Laughs	37:41
1970	Syd Barrett	Barrett	38:43
1981	Nick Mason	Nick Mason's Fictitious Sports	36:15
1985	Nick Mason	Profiles	44:10
1986	Nick Mason	Life Could Be a Dream	NA
1987	Nick Mason	White of the Eye	43:10
1987	Nick Mason	Body Contact	NA
1989	Nick Mason	Tank Malling	NA
1970	Roger Waters	Music from The Body	41:28
1984	Roger Waters	The Pros and Cons of Hitch Hiking	42:07
1986	Roger Waters	When the Wind Blows	45:36
1987	Roger Waters	Radio K.A.O.S.	41:24
1992	Roger Waters	Amused to Death	72:45
2005	Roger Waters	Ça Ira	108:29
2017	Roger Waters	Is This the Life We Really Want?	54:06
1978	Richard Wright	Wet Dream	43:51
1984	Richard Wright	Identity	46:03
1996	Richard Wright	Broken China	59:34
1978	David Gilmour	David Gilmour	46:19
1984	David Gilmour	About Face	45:18
2006	David Gilmour	On an Island	51:36
2010	David Gilmour	Metallic Spheres	48:53
2015	David Gilmour	Rattle That Lock	51:28

1.2 Inclusion Criteria

Please note that this data frame is by no means exhaustive. Pink Floyd and its members have released much more musical works, but the data frame `albums` includes only **thematically**

unified albums to which they have made significant innovative musical contributions.

Here are some explanations:

- The included cases must be albums. So singles such as *Arnold Layne* are not included.
- Each member's solo albums are also included.
- Each album must have a unified theme. So compilation albums such as *1965: Their First Recordings* are not included.
- Non-musical albums such as Roger Waters' *Igor Stravinsky's The Soldier's Tale* are not included.
- Live albums and box sets are not included, because they do not contain new songs.
- Collaborative albums are included if they contain Pink Floyd or its members' significant contributions. So the soundtrack album *Zabriskie Point* is included, while Michael Mantler's *Something There*, in which Nick Mason played drums, is not.

1.3 Members

Pink Floyd has had five members. The below plot displays the number of albums of each member:

```
albums |>
  # count the number of albums of each member
  group_by(artist) |>
  summarize(count = n()) |>

ggplot() +
  # reorder bars according to the numbers of albums
  geom_col(aes(reorder(artist, count, decreasing = TRUE), count)) +

  labs(x = 'Artist') +
  scale_y_continuous('Number of Albums', 1:20)
```

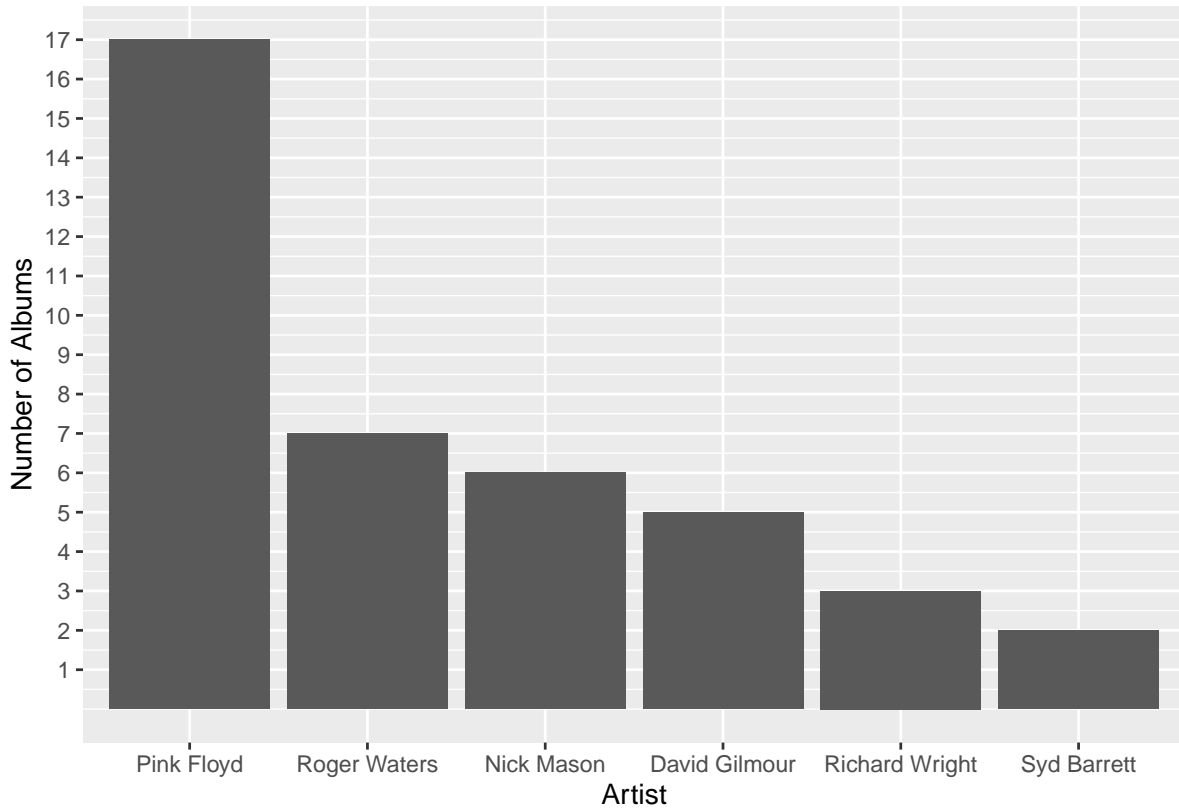


Figure 1.1: Numbers of Albums of Pink Floyd and Its Members

Pink Floyd's members changed from album to album. The below plot displays each member's presence in each Pink Floyd album. The presence information is in the sourced data frame `members`.

```
albums |>
  # merge two data frames
  right_join(members) |>

  # reshape the data frame
  pivot_longer(5:9, 'member', values_to = 'presence') |>
  filter(presence) |>

  # convert x-axis name to title case
  mutate(member = str_to_title(member)) |>

  # append each album's year after it
  mutate(title = paste0(title, ' (', year, ')')) |>
```

```

ggplot() +
  # reorder albums according to their years
  geom_point(aes(member, reorder(title, year))) +

  labs(x = 'Member', y = 'Album')

```

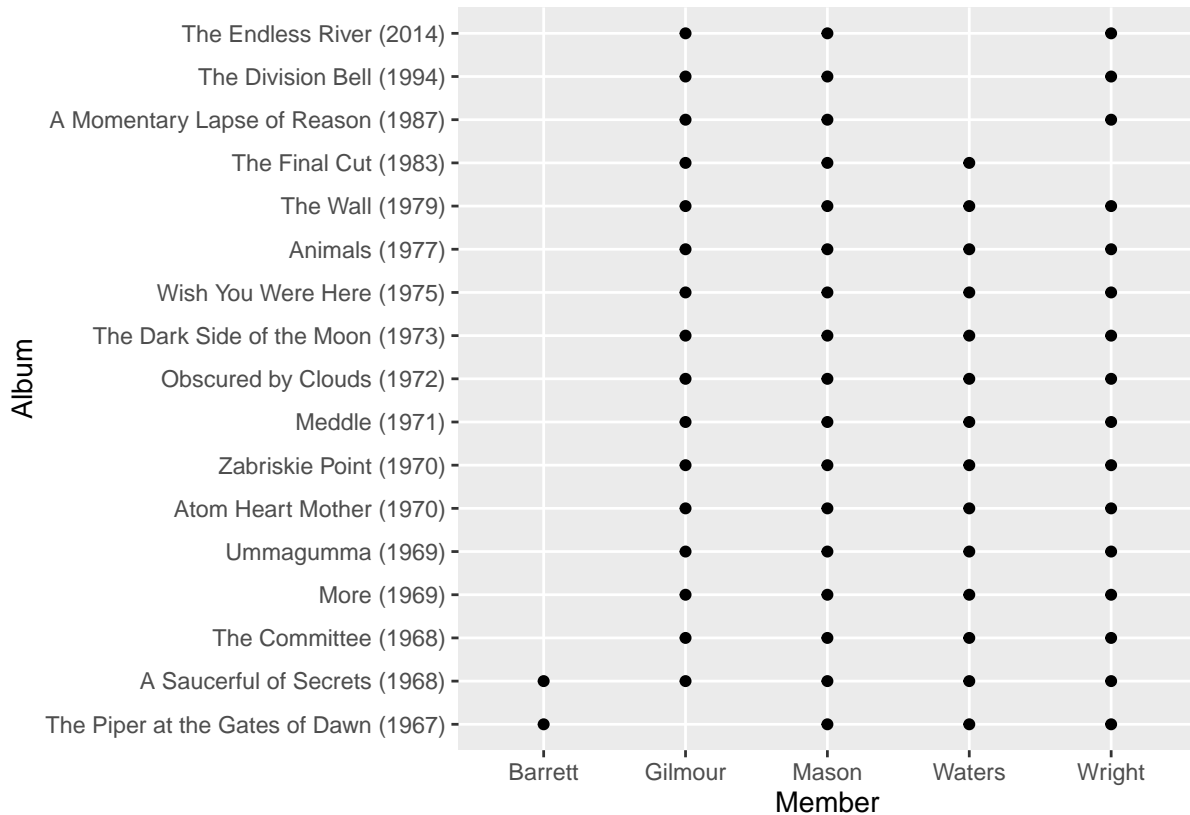


Figure 1.2: Members' Presence in Pink Floyd Albums

1.4 Album Lengths

The below plot displays the album lengths of Pink Floyd and its members:

```

albums |>
  # remove cases with missing values
  filter(!is.na(length)) |>

```



```

# convert character album lengths to seconds
mutate(second = as.double(ms(length))) |>

# indicate quantiles
mutate(q = cut(second, quantile(second), include.lowest = TRUE)) |>

ggplot(aes(fill = q)) +
  # reorder albums according to their lengths
  geom_col(aes(second, reorder(title, second)), width = 0.3) +

  # append album lengths after bars
  geom_text(aes(second, title, label = length), nudge_x = 700, size = 3) +

  # modify the legend
  scale_fill_manual(
    name = 'Length Quantile',
    values = c('#4E84C4', '#52854C', '#293352', '#D16103'),
    labels = c('[0, 25%]', '(25%, 50%]', '(50%, 75%]', '(75%, 100%]' )
  ) +

  # reverse the legend labels
  guides(fill = guide_legend(reverse = TRUE)) +

  labs(x = 'Length', y = 'Album') +
  xlim(0, 7500) +

  # remove x-axis labels and ticks
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank())

```

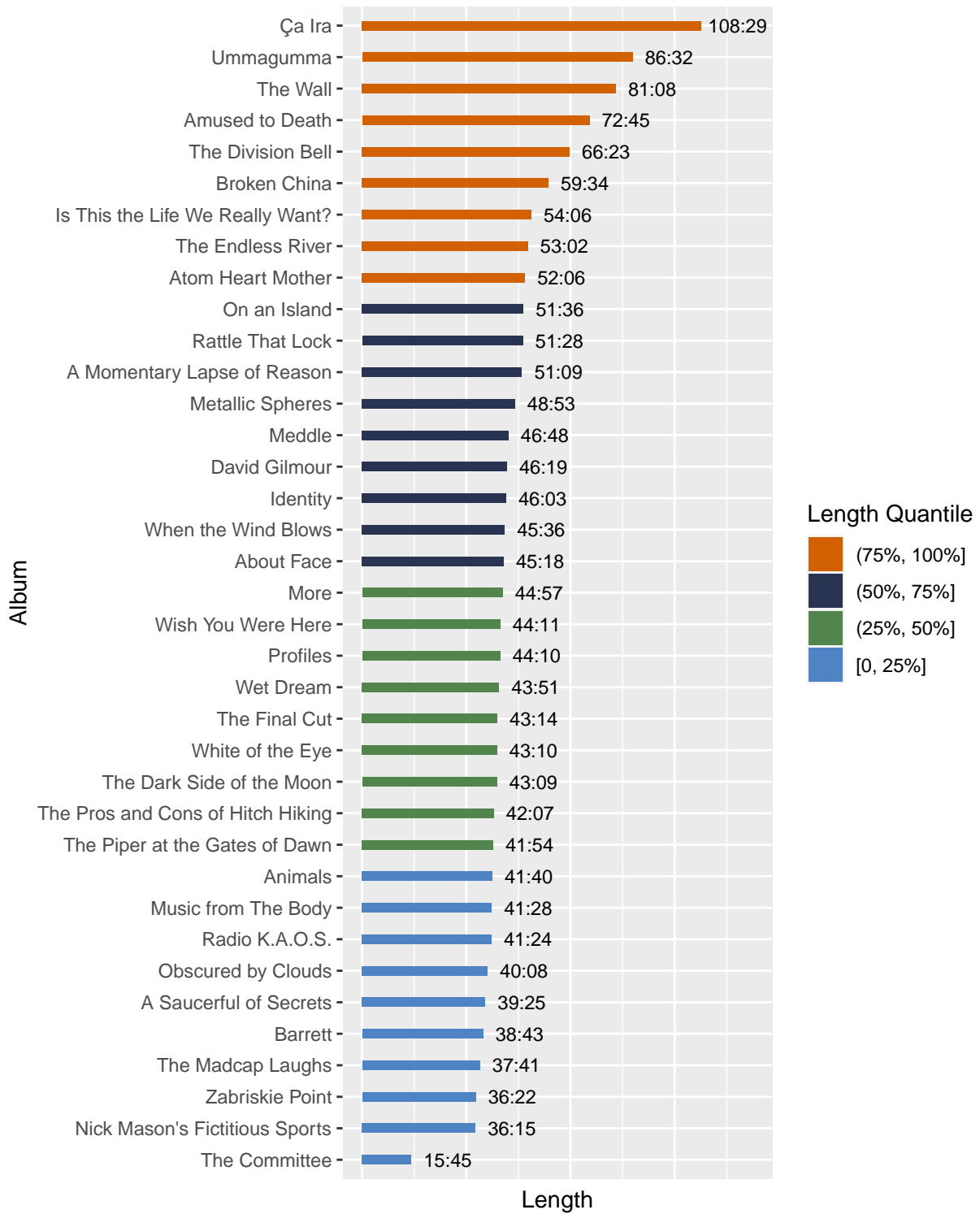


Figure 1.3: Album Lengths of Pink Floyd and Members

The below plot displays only Pink Floyd studio album lengths:

```
albums |>
  # keep only Pink Floyd's studio albums
  filter(artist == 'Pink Floyd') |>
  filter(!(title %in% c('The Committee', 'Zabriskie Point'))) |>

  # convert character album lengths to seconds
  mutate(second = as.double(ms(length))) |>

  # indicate quantiles
  mutate(q = cut(second, quantile(second), include.lowest = TRUE)) |>

  ggplot(aes(fill = q)) +
  # reorder albums according to their lengths
  geom_col(aes(second, reorder(title, second)), width = 0.3) +

  # append album lengths after bars
  geom_text(aes(second, title, label = length), nudge_x = 700, size = 3) +

  # modify the legend
  scale_fill_manual(
    name = 'Length Quantile',
    values = c('#4E84C4', '#52854C', '#293352', '#D16103'),
    labels = c('[0, 25%', '(25%, 50%', '(50%, 75%', '(75%, 100%]'')
  ) +

  # reverse the legend labels
  guides(fill = guide_legend(reverse = TRUE)) +

  labs(x = 'Length', y = 'Album') +
  xlim(0, 7500) +

  # remove x-axis labels and ticks
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank())
```

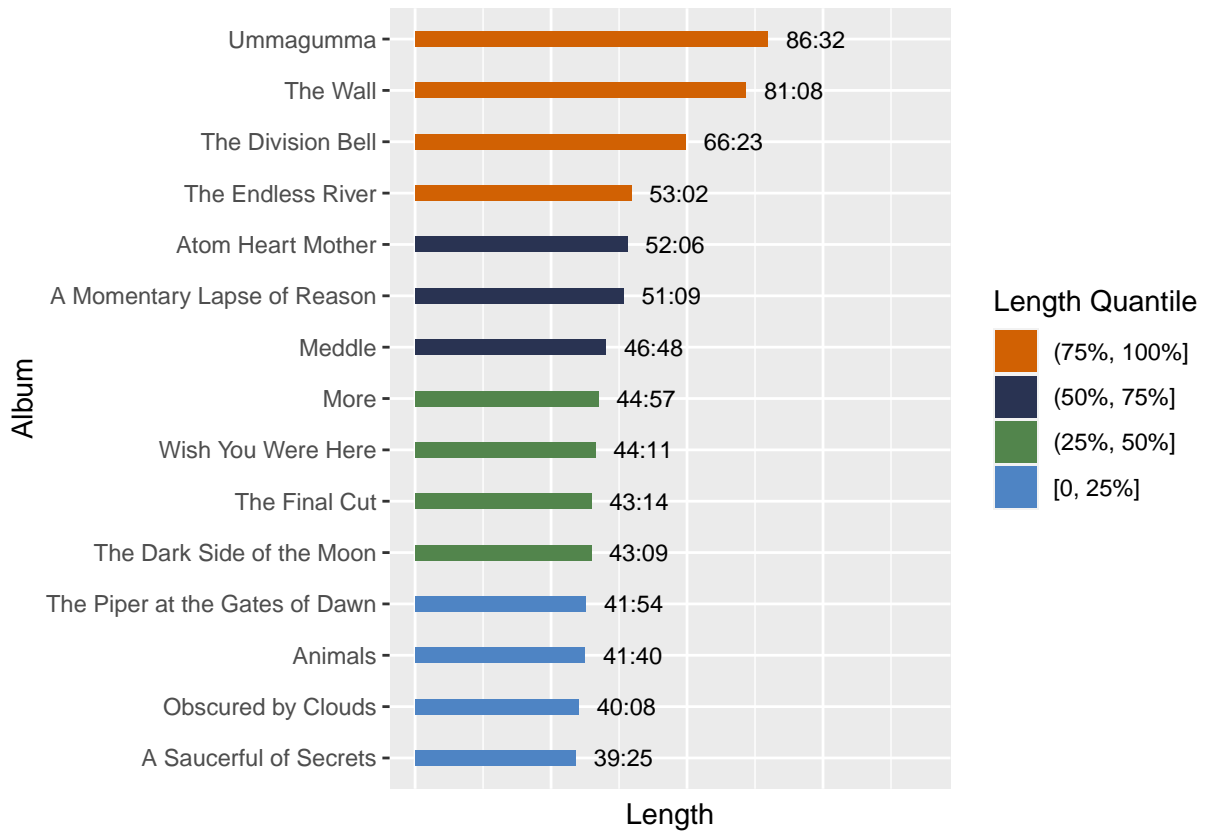


Figure 1.4: Pink Floyd Studio Album Lengths

1.5 Soundtrack Albums

Pink Floyd and members have had several soundtrack albums. The sourced vector `soundtrack` contains titles of their soundtrack albums.

```
albums |>
  filter(title %in% soundtrack) |>
  rename_with(str_to_title) |>
  kable()
```

Table 1.2: Pink Floyd and Members' Soundtrack Albums

Year	Artist	Title	Length
1968	Pink Floyd	The Committee	15:45
1969	Pink Floyd	More	44:57

Year	Artist	Title	Length
1970	Pink Floyd	Zabriskie Point	36:22
1972	Pink Floyd	Obscured by Clouds	40:08
1986	Nick Mason	Life Could Be a Dream	NA
1987	Nick Mason	White of the Eye	43:10
1987	Nick Mason	Body Contact	NA
1989	Nick Mason	Tank Malling	NA
1970	Roger Waters	Music from The Body	41:28
1986	Roger Waters	When the Wind Blows	45:36

The below plot displays the numbers of soundtrack albums of Pink Floyd and members:

```
albums |>
  # indicate soundtrack albums
  mutate(soundtrack = title %in% soundtrack) |>

  ggplot() +
    geom_bar(
      aes(artist, fill = soundtrack),
      position = position_dodge2(preserve = 'single')
    ) +

    labs(x = 'Artist') +
    scale_y_continuous('Number of Albums', 1:15) +

    # modify legend
    scale_fill_manual(
      name = 'Soundtrack',
      values = c('white', '#595959'),
      labels = c('No', 'Yes')
    )
```

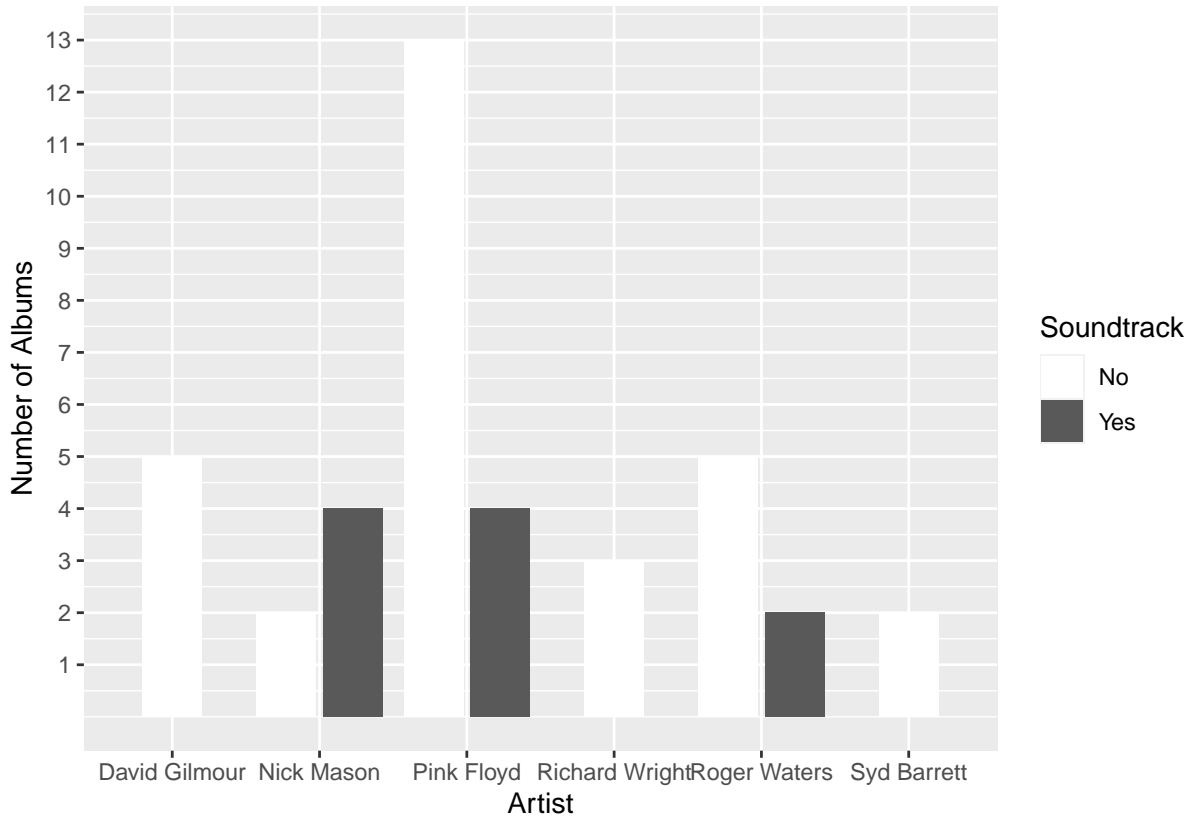


Figure 1.5: Pink Floyd and Members' Soundtrack Albums

1.6 Collaborative Albums

Pink Floyd and members have had collaboration with other artists. The sourced vector `collaborative` contains titles of their collaborative albums.

```
albums |>
  filter(title %in% collaborative) |>
  rename_with(str_to_title) |>
  kable()
```

Table 1.3: Pink Floyd and Members' Collaborative Albums

Year	Artist	Title	Length
1970	Pink Floyd	Zabriskie Point	36:22
1981	Nick Mason	Nick Mason's Fictitious Sports	36:15

Year	Artist	Title	Length
1985	Nick Mason	Profiles	44:10
1986	Nick Mason	Life Could Be a Dream	NA
1987	Nick Mason	White of the Eye	43:10
1987	Nick Mason	Body Contact	NA
1989	Nick Mason	Tank Malling	NA
1970	Roger Waters	Music from The Body	41:28
1984	Richard Wright	Identity	46:03
2010	David Gilmour	Metallic Spheres	48:53

The below plot displays the numbers of collaborative albums of Pink Floyd and members:

```
albums |>
# indicate collaborative albums
mutate(collaborative = title %in% collaborative) |>

ggplot() +
  geom_bar(
    aes(artist, fill = collaborative),
    position = position_dodge2(preserve = 'single')
  ) +

  labs(x = 'Artist') +
  scale_y_continuous('Number of Albums', 1:20) +

  # modify legend
  scale_fill_manual(
    name = 'Collaborative',
    values = c('white', '#595959'),
    labels = c('No', 'Yes')
  )
```

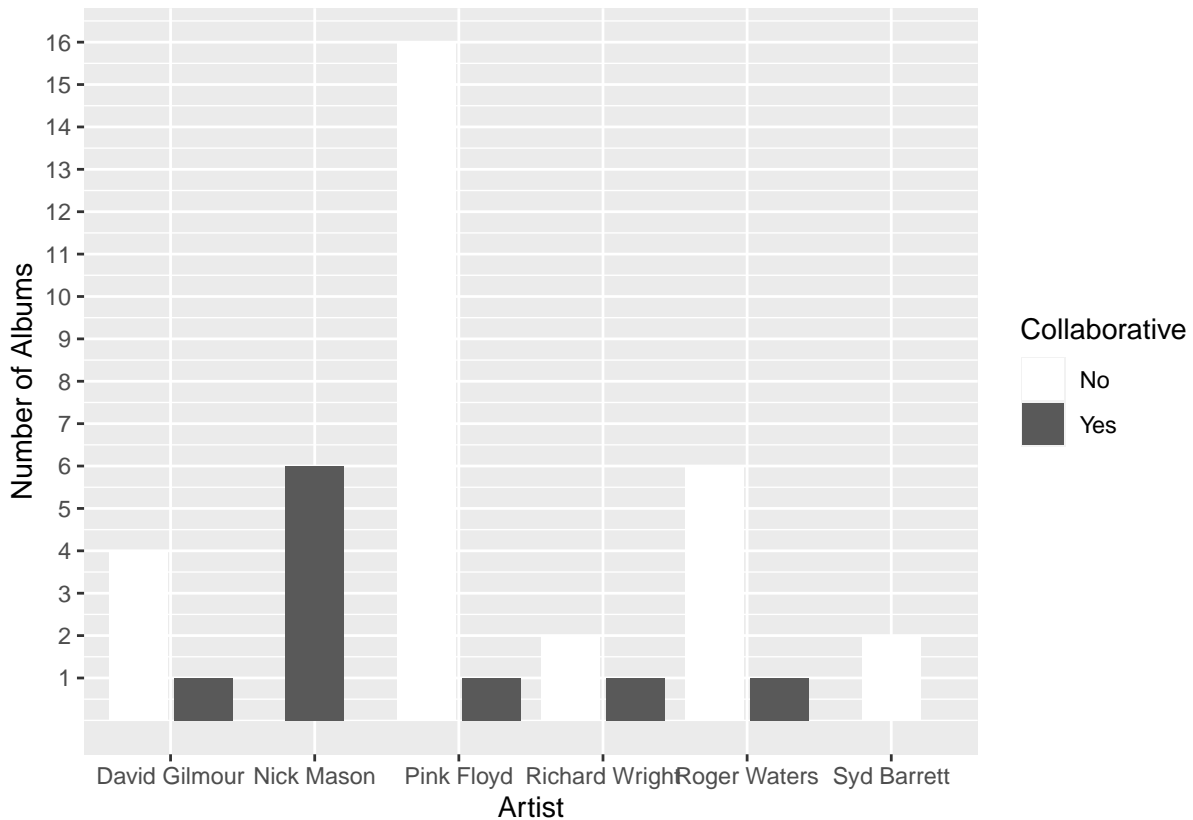


Figure 1.6: Pink Floyd and Members' Collaborative Albums

1.7 Productivity

How has Pink Floyd and its members' productivity changed over time?

Let's first check Pink Floyd's studio album productivity:

```
albums |>
  # keep only Pink Floyd's studio albums
  filter(artist == 'Pink Floyd') |>
  filter(!(title %in% c('The Committee', 'Zabriskie Point'))) |>

  ggplot(aes(year)) +
    geom_histogram(breaks = seq(1960, 2025, 5), color = 'black') +
    scale_x_continuous('Year', seq(1960, 2025, 10)) +
    scale_y_continuous('Number of Albums', 1:15)
```

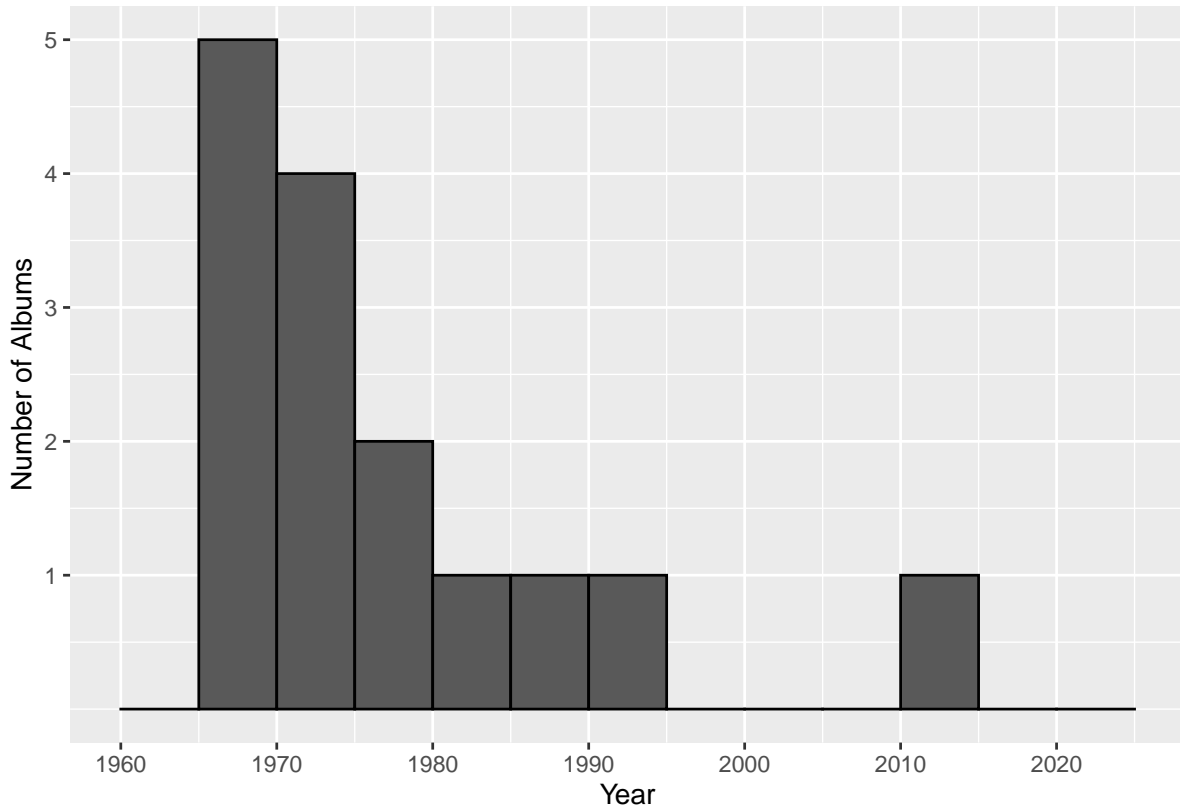



Figure 1.7: Pink Floyd's Studio Album Productivity over Time

How about Pink Floyd and its members' total productivity?

```
ggplot(albums, aes(year)) +  
  geom_histogram(breaks = seq(1960, 2025, 5), color = 'black') +  
  scale_x_continuous('Year', seq(1960, 2025, 10)) +  
  scale_y_continuous('Number of Albums', 1:15)
```

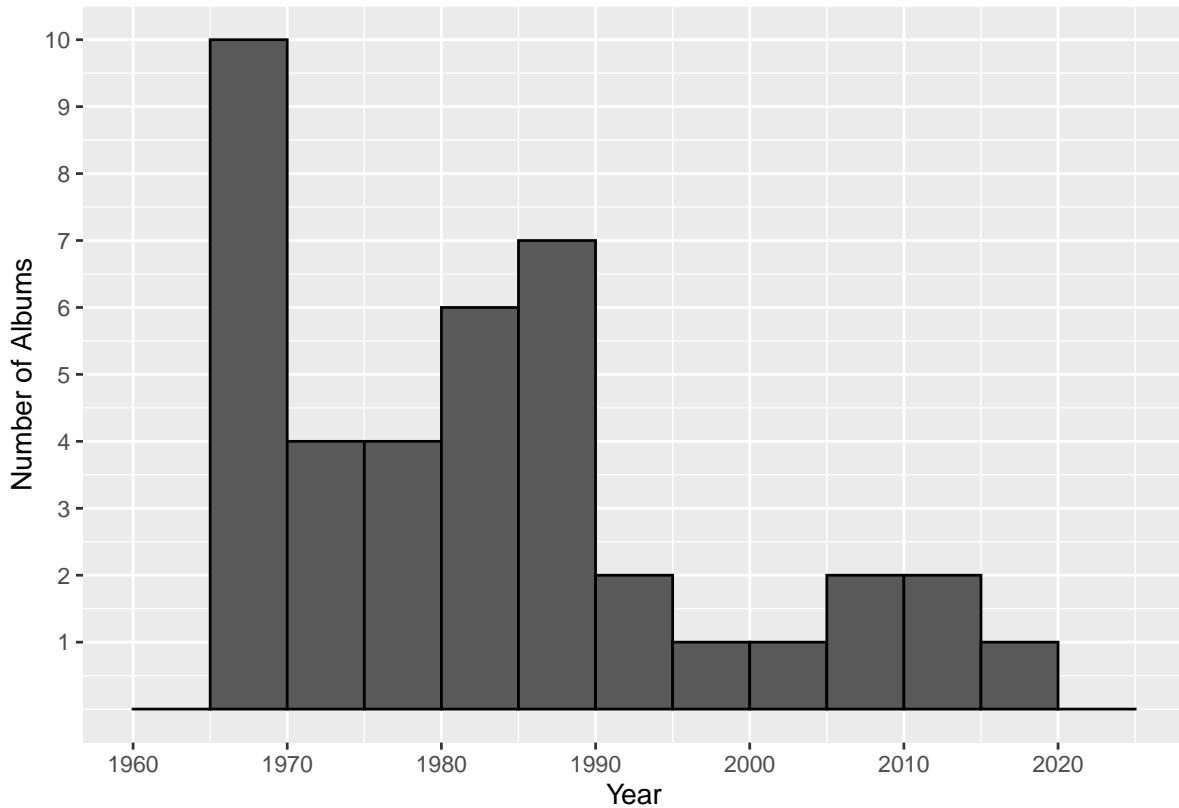


Figure 1.8: Pink Floyd and Members' Total Productivity over Time

Each member's productivity:

```
ggplot(albums, aes(year)) +
  # add total productivity in the background
  geom_histogram(
    data = transform(albums, artist = NULL),
    fill = 'grey85',
    breaks = seq(1960, 2025, 5)
  ) +

  geom_histogram(breaks = seq(1960, 2025, 5), color = 'black') +
  scale_x_continuous('Year', seq(1960, 2025, 10)) +
  scale_y_continuous('Number of Albums', 1:15) +
  facet_wrap(~ artist)
```

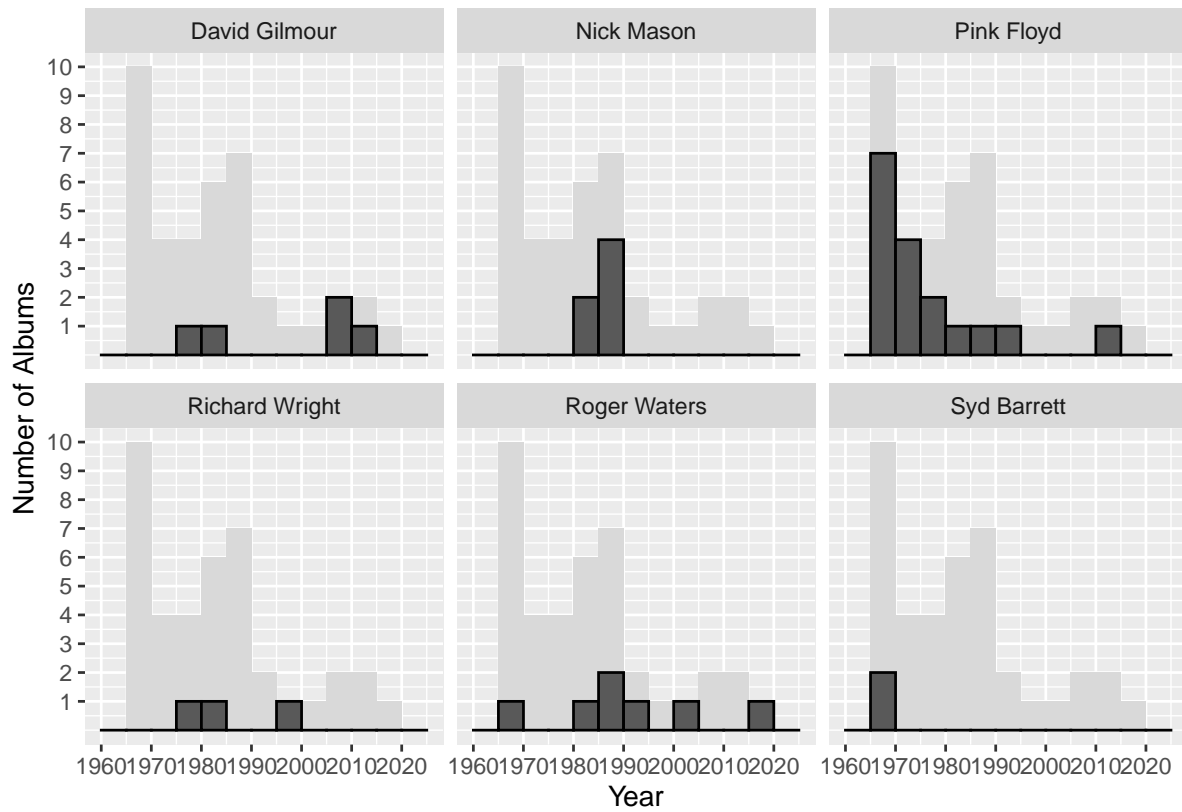


Figure 1.9: Pink Floyd and Members' Productivity over Time

2 Songs

```
library(stringr)
library(knitr)
library(dplyr)
library(ggplot2)
library(tidyr)
library(lubridate)

source('data/songs.R')
source('data/albums.R')
source('utils.R')
```

In this article, we will explore the metadata of Pink Floyd songs.

2.1 Songs

The sourced data frame `songs` contains all the songs we are going to explore. The below table displays the first 15 songs:

```
songs |>
  # show only the first several entries
  head(n = 15) |>

  # convert the column names to title case
  rename_with(str_to_title) |>

  # show the data frame in a better format
  kable()
```

Table 2.1: Some Songs from Pink Floyd Studio Albums

Album	Side	Number	Title	Length
The Piper at the Gates of Dawn	1	1	Astronomy Domine	4:12
The Piper at the Gates of Dawn	1	2	Lucifer Sam	3:07
The Piper at the Gates of Dawn	1	3	Matilda Mother	3:08
The Piper at the Gates of Dawn	1	4	Flaming	2:46
The Piper at the Gates of Dawn	1	5	Pow R. Toc H.	4:26
The Piper at the Gates of Dawn	1	6	Take Up Thy Stethoscope and Walk	3:05
The Piper at the Gates of Dawn	2	1	Interstellar Overdrive	9:41
The Piper at the Gates of Dawn	2	2	The Gnome	2:13
The Piper at the Gates of Dawn	2	3	Chapter 24	3:42
The Piper at the Gates of Dawn	2	4	The Scarecrow	2:11
The Piper at the Gates of Dawn	2	5	Bike	3:21
A Saucerful of Secrets	1	1	Let There Be More Light	5:38
A Saucerful of Secrets	1	2	Remember a Day	4:33
A Saucerful of Secrets	1	3	Set the Controls for the Heart of the Sun	5:28
A Saucerful of Secrets	1	4	Corporal Clegg	4:13

2.2 Inclusion Criteria

Only the songs from Pink Floyd studio albums are included. So singles, soundtrack albums like *Zabriskie Point*, and Pink Floyd members' albums are not included.

There are several versions or releases of Pink Floyd studio albums. Only the original releases are considered. For example, the extra songs from *The Endless River* deluxe edition are not included.

2.3 Song Counts

The total number of songs from Pink Floyd studio albums:

```
nrow(songs)
```

```
[1] 159
```

The number of songs from each album:

```
songs |>
  # count the number of songs for each album
  group_by(album) |>
  summarize(count = n()) |>

  # use the `year` column of `albums`
  left_join(albums, by = c('album' = 'title')) |>

  # append each album's year after it
  mutate(album = paste0(album, ' (', year, ')')) |>

  ggplot() +
  # reorder albums according to their years
  geom_col(aes(count, reorder(album, year))) +

  # append song counts after bars
  geom_text(aes(count, album, label = count), nudge_x = 1, size = 3) +

  labs(x = 'Number of Songs', y = 'Album') +

  # remove x-axis labels and ticks
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank())
```

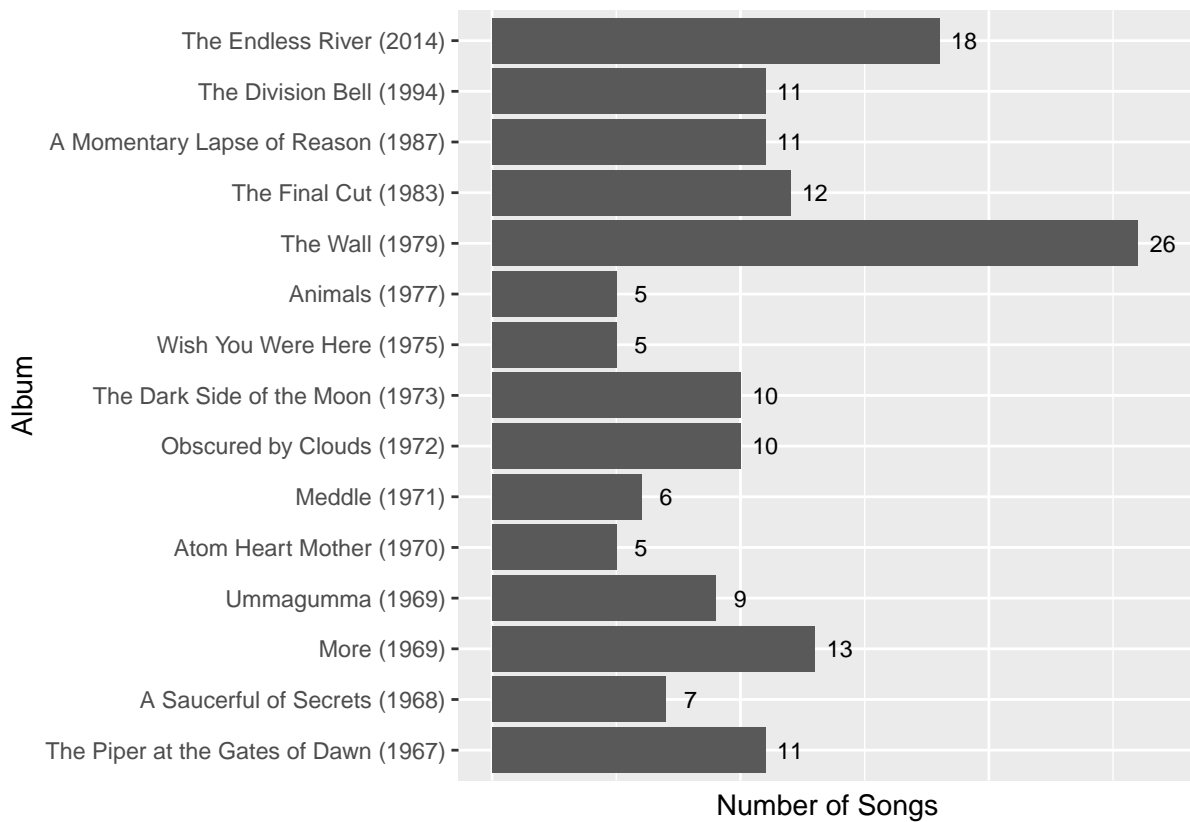


Figure 2.1: Number of Songs of Each Pink Floyd Studio Album

2.4 Song Lengths

The distribution of Pink Floyd song lengths:

```
# make a copy of `songs`
d <- songs |>
  # convert song lengths to seconds
  mutate(second = as.double(ms(length)))

# quantiles of song lengths
qs <- quantile(d$second)
# corresponding labels
labels <- paste(to_length(qs), names(qs), sep = '\n')

ggplot(d, aes(second)) +
```

```
geom_histogram(binwidth = 30) +
geom_density(aes(y = 30 * ..count..)) +
scale_y_continuous('Number of Songs', 0:20) +
scale_x_continuous('Length\nQuantile', qs, labels = labels)
```

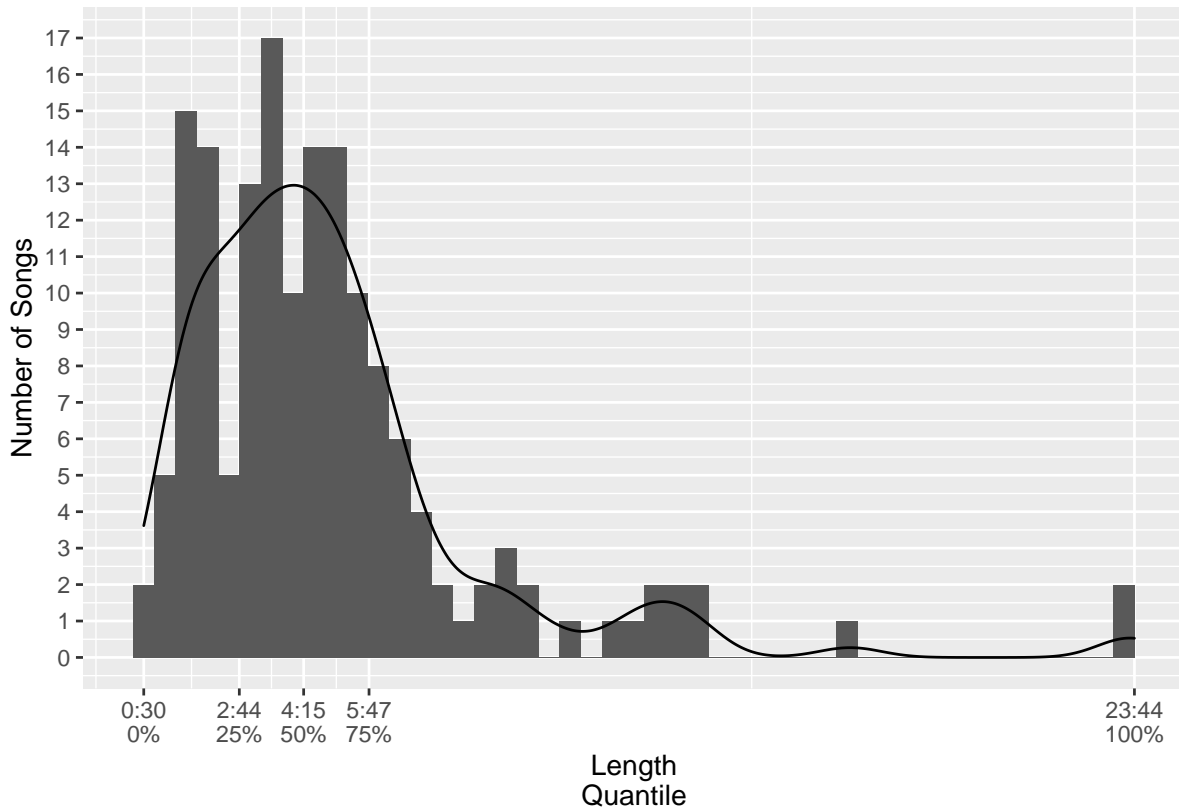


Figure 2.2: Distribution of Pink Floyd Song Lengths

Song lengths of each album:

```
songs |>
  # convert song lengths to seconds
  mutate(second = as.double(ms(length))) |>

  # use the `year` column of `albums`
  left_join(albums, by = c('album' = 'title')) |>

  # append each album's year after it
  mutate(album = paste0(album, ' (', year, ')')) |>
```



```

ggplot() +
  # reorder albums according to their years
  geom_boxplot(aes(second, reorder(album, year))) +

  labs(x = 'Length', y = 'Album') +
  scale_x_continuous(
    breaks = seq(0, 1500, 120),
    labels = to_length(seq(0, 1500, 120))
  )

```



Figure 2.3: Song Lengths of Each Pink Floyd Studio Album

The longest five songs:

```

songs |>
  mutate(second = as.double(ms(length))) |>
  arrange(desc(second)) |>

```

```

select(title, length, album) |>
head(n = 5) |>
rename_with(str_to_title) |>
kable()

```

Table 2.2: Longest Five Pink Floyd Songs

Title	Length	Album
Atom Heart Mother	23:44	Atom Heart Mother
Echoes	23:31	Meddle
Dogs	17:04	Animals
Shine On You Crazy Diamond (Parts 1-5)	13:32	Wish You Were Here
Sisyphus	13:28	Ummagumma

The shortest five songs:

```

songs |>
mutate(second = as.double(ms(length))) |>
arrange(second) |>
select(title, length, album) |>
head(n = 5) |>
rename_with(str_to_title) |>
kable()

```

Table 2.3: Shortest Five Pink Floyd Songs

Title	Length	Album
Stop	0:30	The Wall
A New Machine Part 2	0:38	A Momentary Lapse of Reason
A Spanish Piece	1:05	More
Speak to Me	1:05	The Dark Side of the Moon
Party Sequence	1:07	More

2.5 Authorship

The sourced data frame `authorship` contains the lyricists, composers, and vocalists of Pink Floyd songs. The below table displays the first several cases:

```

authorship |>
  # show only the first several entries
  head() |>

  # convert the column names to title case
  rename_with(str_to_title) |>

  # show the data frame in a better format
  kable()

```

Table 2.4: Authorship of Some Pink Floyd Songs

Song	Lyrics	Music	Vocal
Astronomy Domine	Barrett	Barrett	Barrett, Wright
Lucifer Sam	Barrett	Barrett	Barrett
Matilda Mother	Barrett	Barrett	Wright , Barrett
Flaming	Barrett	Barrett	Barrett
Pow R. Toc H.	NA	Barrett, Waters , Wright , Mason	Barrett, Waters , Wright
Take Up Thy Stethoscope and Walk	Waters	Waters	Waters

2.6 Instrumentalness

Any song with a missing value in its `lyrics` column is instrumental.

The percentage of instrumental songs:

```

(sum(is.na(authorship$lyrics)) / nrow(authorship) * 100) |>
  round(digits = 2) |>
  paste0('%')

```

```
[1] "28.85%"
```

The percentage of instrumental songs for each album:

```

authorship |>
  # indicate instrumentalness
  mutate(instr = sapply(lyrics, anyNA)) |>

```

```

# use the 'album' column of `songs`
left_join(songs, by = c('song' = 'title')) |>

# use the `year` column of `albums`
left_join(albums, by = c('album' = 'title')) |>

# append each album's year after it
mutate(album = paste0(album, ' (', year, ')')) |>

# reorder albums according to their years
ggplot(aes(y = reorder(album, year), fill = instr)) +

# show percentages
geom_bar(position = 'fill') +

labs(x = 'Percentage of Instrumental Songs', y = 'Album') +
scale_x_continuous(labels = paste0(seq(0, 100, 25), '%')) +

# modify legend
scale_fill_manual(
  name = 'Instrumental',
  values = c('white', '#595959'),
  labels = c('No', 'Yes')
)

```

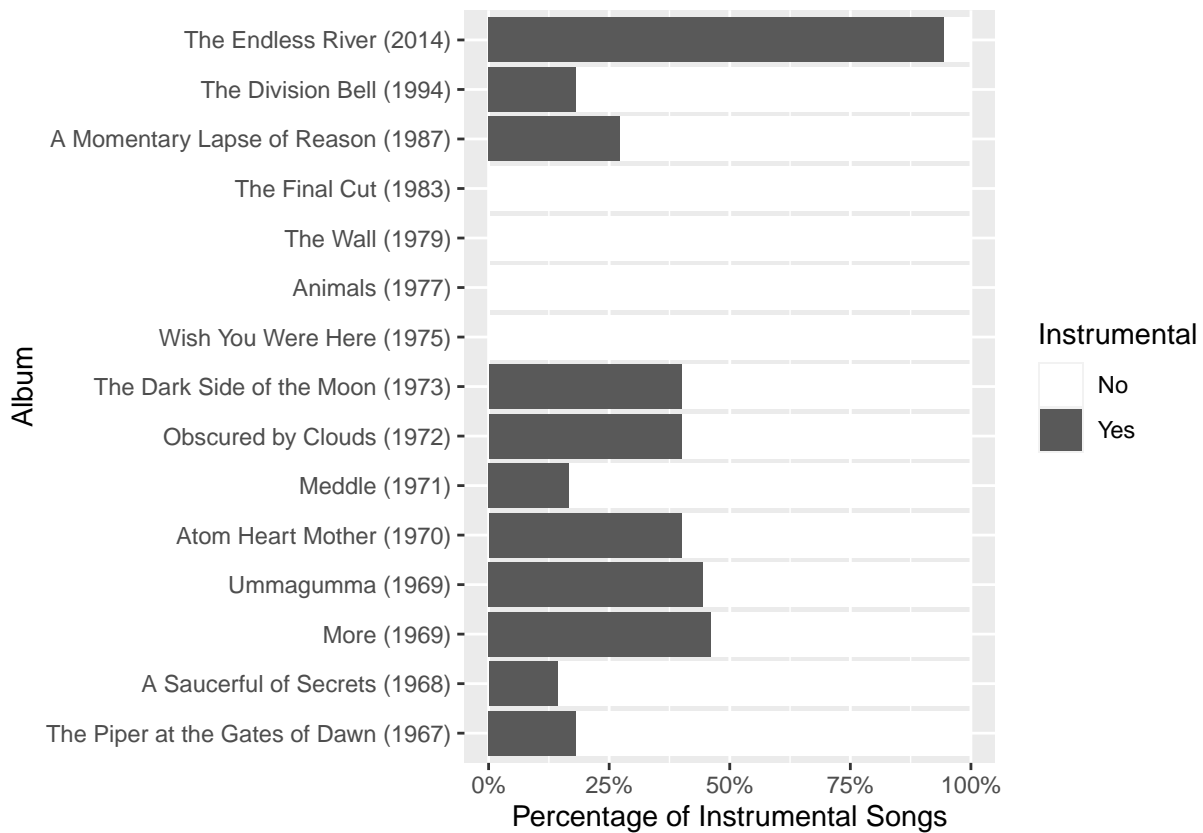


Figure 2.4: Percentage of Instrumental Songs for Each Album

2.7 Whose Pink Floyd?

Pink Floyd had been led by different members in different eras. Let's see if this can be reflected by the data.

The column `music` in the sourced data frame `authorship` contains the composers of each Pink Floyd song. The below bar chart displays the percentage of each member's contributions to the music of each album:

```
authorship |>
  # add each song's album
  left_join(songs, by = c('song' = 'title')) |>

  # add each album's year
  left_join(albums, by = c('album' = 'title')) |>
```

```

# append each album's year after its title
mutate(album = paste0(album, '(', year, ')')) |>

# remove unnecessary columns
select(song, music, album, year) |>

# expand the list column `music`
unnest_longer('music') |>

# indicate composers who are not Pink Floyd members
mutate(music = ifelse(
  music %in% c('Barrett', 'Waters', 'Wright', 'Mason', 'Gilmour'),
  music, 'Other'
)) |>

# to reorder plot legend labels
mutate(music = factor(
  music,
  c('Waters', 'Gilmour', 'Wright', 'Mason', 'Barrett', 'Other')
)) |>

# reorder albums according to their years
ggplot(aes(y = reorder(album, year), fill = music)) +

# show percentages
geom_bar(position = position_fill(reverse = TRUE)) +

labs(x = 'Percentage of Music Contributions', y = 'Album') +
scale_x_continuous(labels = paste0(seq(0, 100, 25), '%')) +

# modify legend
scale_fill_manual(
  name = 'Composer',
  values = c('#ED5564', '#4FC1E8', '#AC92EB',
    '#A0D568', '#FFCE54', 'white')
)

```

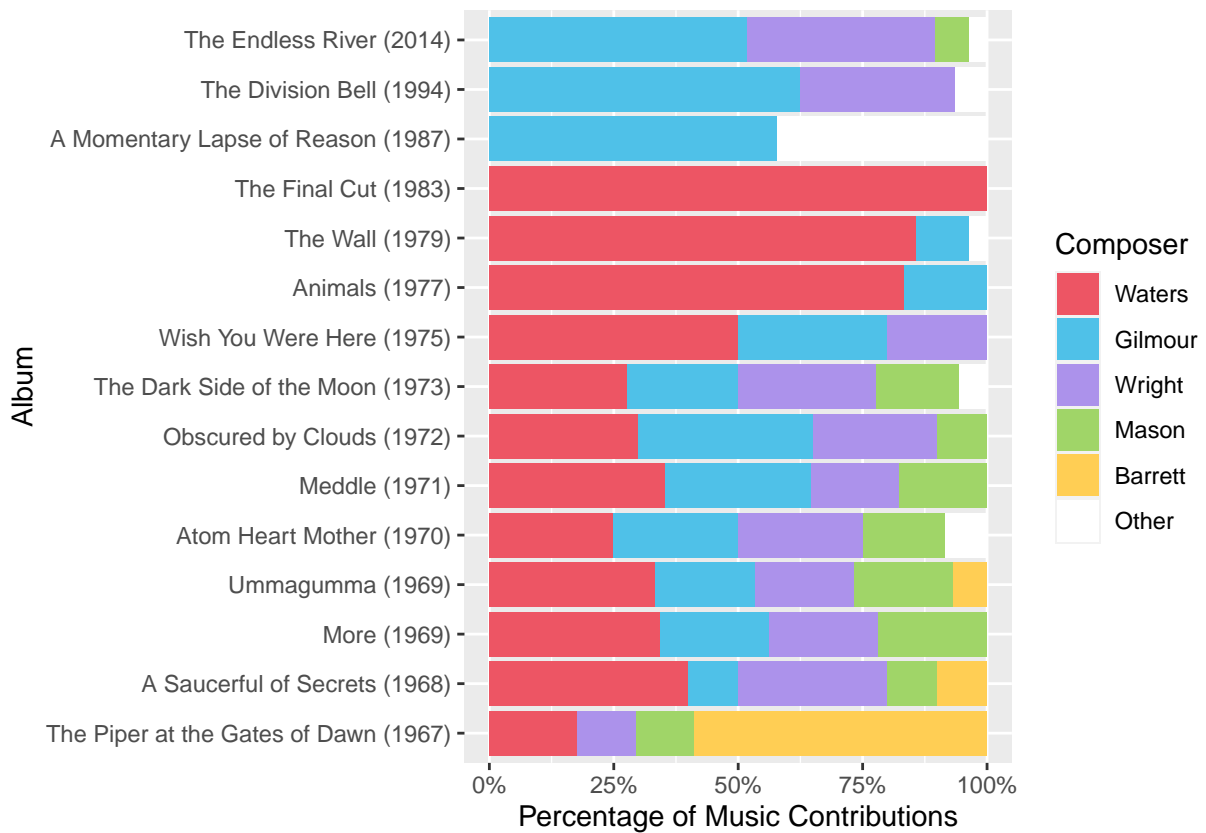


Figure 2.5: Percentage of Each Member's Music Contributions

The column `lyrics` in the sourced data frame `authorship` contains the lyricists of each Pink Floyd song. The below bar chart displays the percentage of each member's contributions to the lyrics of each album:

```
authorship |>
  # add each song's album
  left_join(songs, by = c('song' = 'title')) |>

  # add each album's year
  left_join(albums, by = c('album' = 'title')) |>

  # append each album's year after its title
  mutate(album = paste0(album, '(', year, ')')) |>

  # remove unnecessary columns
  select(song, lyrics, album, year) |>
```

```

# expand the list column `lyrics`
unnest_longer('lyrics') |>

# indicate instrumental songs
mutate(lyrics = ifelse(is.na(lyrics), 'Instrumental', lyrics)) |>

# indicate lyricists who are not Pink Floyd members
mutate(lyrics = ifelse(
  lyrics %in% c('Barrett', 'Waters', 'Wright', 'Mason',
    'Gilmour', 'Instrumental'),
  lyrics, 'Other'
)) |>

# to reorder plot legend labels
mutate(lyrics = factor(
  lyrics,
  c('Waters', 'Gilmour', 'Wright', 'Mason',
    'Barrett', 'Other', 'Instrumental')
)) |>

# reorder albums according to their years
ggplot(aes(y = reorder(album, year), fill = lyrics)) +

# show percentages
geom_bar(position = position_fill(reverse = TRUE)) +

labs(x = 'Percentage of Lyrics Contributions', y = 'Album') +
scale_x_continuous(labels = paste0(seq(0, 100, 25), '%')) +

# modify legend
scale_fill_manual(
  name = 'Lyricist',
  values = c('#ED5564', '#4FC1E8', '#AC92EB',
    '#A0D568', '#FFCE54', 'white', 'grey85')
)

```

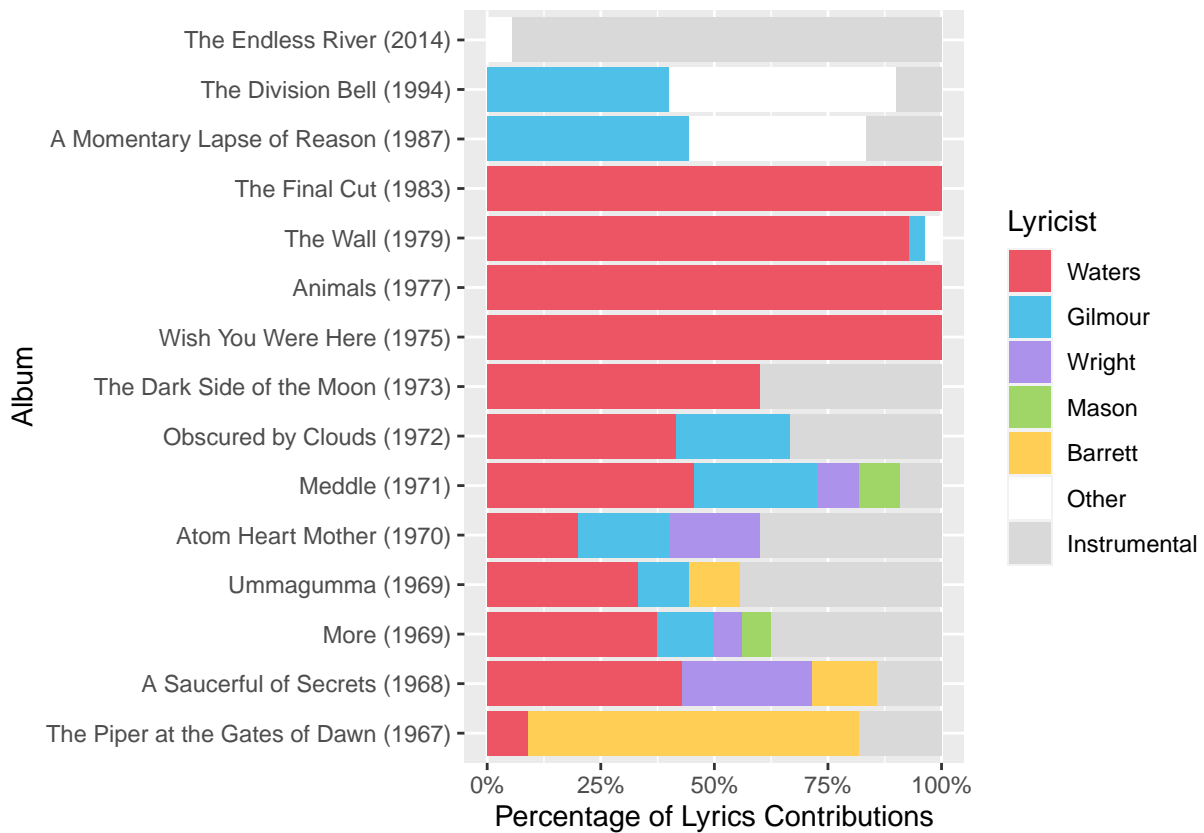



Figure 2.6: Percentage of Each Member's Lyrics Contributions

Let's combine these two plots and see each member's total contributions to each album:

```

authorship |>
  # add each song's album
  left_join(songs, by = c('song' = 'title')) |>

  # add each album's year
  left_join(albums, by = c('album' = 'title')) |>

  # append each album's year after its title
  mutate(album = paste0(album, ' (', year, ')')) |>

  # merge `music` and `lyrics` columns
  mutate(author = mapply(c, music, lyrics, SIMPLIFY = FALSE)) |>

  # remove unnecessary columns

```

```

select(song, author, album, year) |>

# expand the list column `author`
unnest_longer('author') |>
filter(!is.na(author)) |>

# indicate writers who are not Pink Floyd members
mutate(author = ifelse(
  author %in% c('Barrett', 'Waters', 'Wright', 'Mason', 'Gilmour'),
  author, 'Other'
)) |>

# to reorder plot legend labels
mutate(author = factor(
  author,
  c('Waters', 'Gilmour', 'Wright', 'Mason', 'Barrett', 'Other')
)) |>

# reorder albums according to their years
ggplot(aes(y = reorder(album, year), fill = author)) +

# show percentages
geom_bar(position = position_fill(reverse = TRUE)) +

labs(x = 'Percentage of Contributions', y = 'Album') +
scale_x_continuous(labels = paste0(seq(0, 100, 25), '%')) +

# modify legend
scale_fill_manual(
  name = 'Writer',
  values = c('#ED5564', '#4FC1E8', '#AC92EB',
    '#A0D568', '#FFCE54', 'white')
)

```

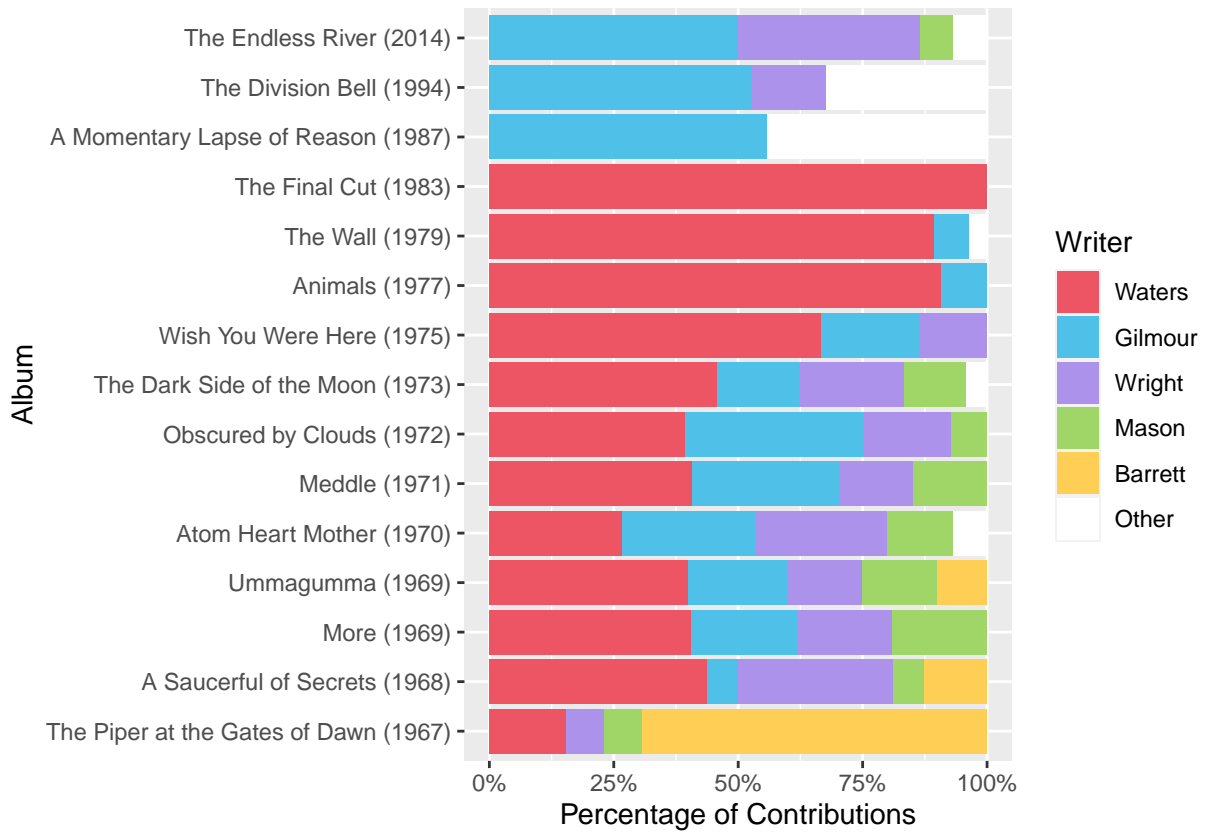


Figure 2.7: Percentage of Each Member's Contributions

Based on the above plot, we might divide the history of Pink Floyd into four eras:

1. Barrett-led era (1967)
2. collective leadership era (1968-1975)
3. Waters-led era (1977-1983)
4. Gilmour-led era (1987-2014)

3 Popularity

```
library(stringr)
library(knitr)
library(dplyr)
library(ggplot2)
library(tidyr)
library(lubridate)

load('data/popularity.RData')
```

In this article, we will explore the data about Pink Floyd’s popularity.

3.1 Spotify Dataset

The popularity data is from [a Spotify dataset on Kaggle](#). This dataset contains information of about 600 thousand tracks. The popularity data is extracted and saved in `data/popularity.RData` as data frame `popularity`.

The first several cases of `popularity`:

```
popularity |>
  # show only the first several entries
  head() |>

  # convert the column names to title case
  rename_with(str_to_title) |>

  # show the data frame in a better format
  kable()
```

Table 3.1: Popularity of Some Spotify Tracks

Name	Artists	Popularity
Carve	['Uli']	6

Name	Artists	Popularity
Capítulo 2.16 - Banquero Anarquista	['Fernando Pessoa']	0
Vivo para Quererte - Remasterizado	['Ignacio Corsini']	0
El Prisionero - Remasterizado	['Ignacio Corsini']	0
Lady of the Evening	['Dick Haymes']	0
Ave Maria	['Dick Haymes']	0

The distribution of Spotify tracks' popularity:

```
ggplot(popularity) +
  geom_histogram(aes(popularity), binwidth = 1) +
  labs(x = 'Popularity', y = 'Number of Songs')
```

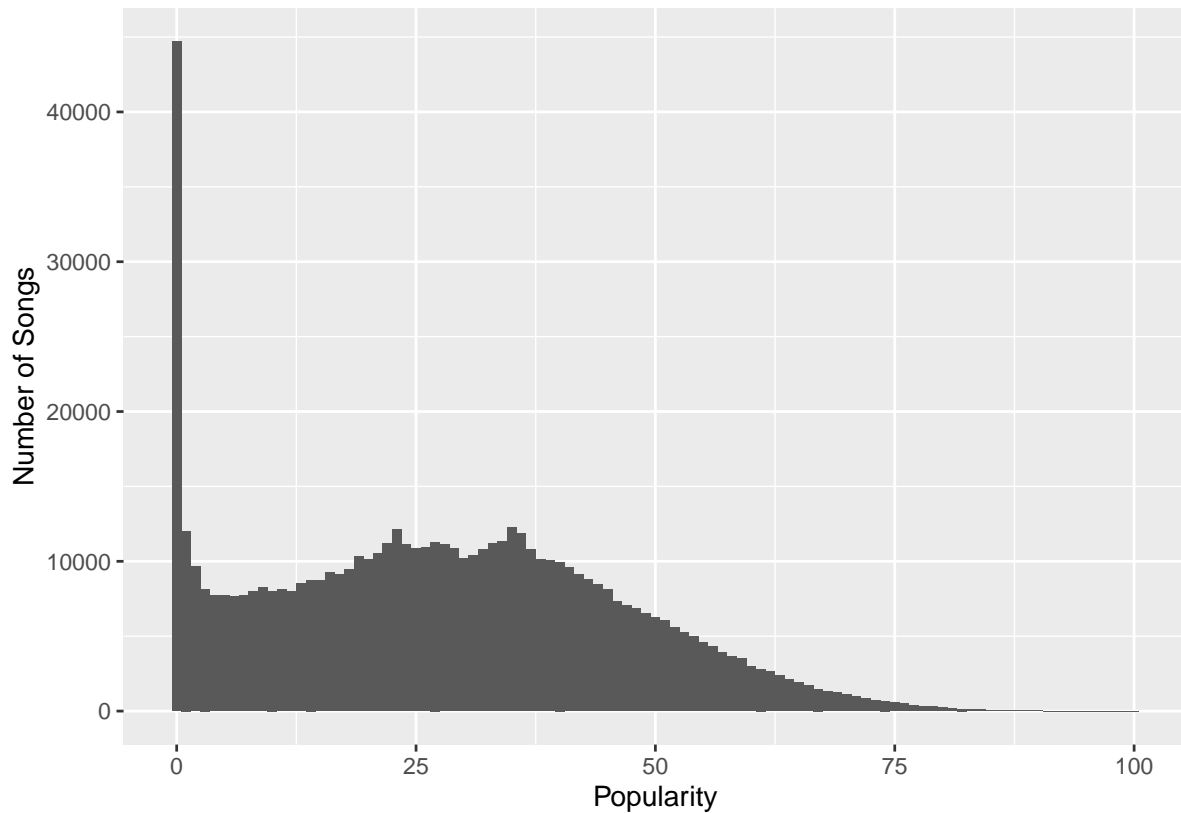


Figure 3.1: Distribution of Spotify Tracks' Popularity

3.2 Pink Floyd

The below boxplot displays the song popularity of Pink Floyd and its members. Wright and Mason are not found in the Spotify dataset.

```
popularity |>
  mutate(artist = ifelse(
    str_detect(artists, 'Pink Floyd'), 'Pink Floyd', 'Other')) |>

  mutate(artist = ifelse(
    str_detect(artists, 'Roger Waters'), 'Waters', artist)) |>

  mutate(artist = ifelse(
    str_detect(artists, 'David Gilmour'), 'Gilmour', artist)) |>

  mutate(artist = ifelse(
    str_detect(artists, 'Syd Barrett'), 'Barrett', artist)) |>

  ggplot(aes(artist, popularity)) +
    geom_boxplot() +
    labs(x = 'Artist', y = 'Popularity') +
    scale_y_continuous(breaks = seq(0, 100, 10))
```

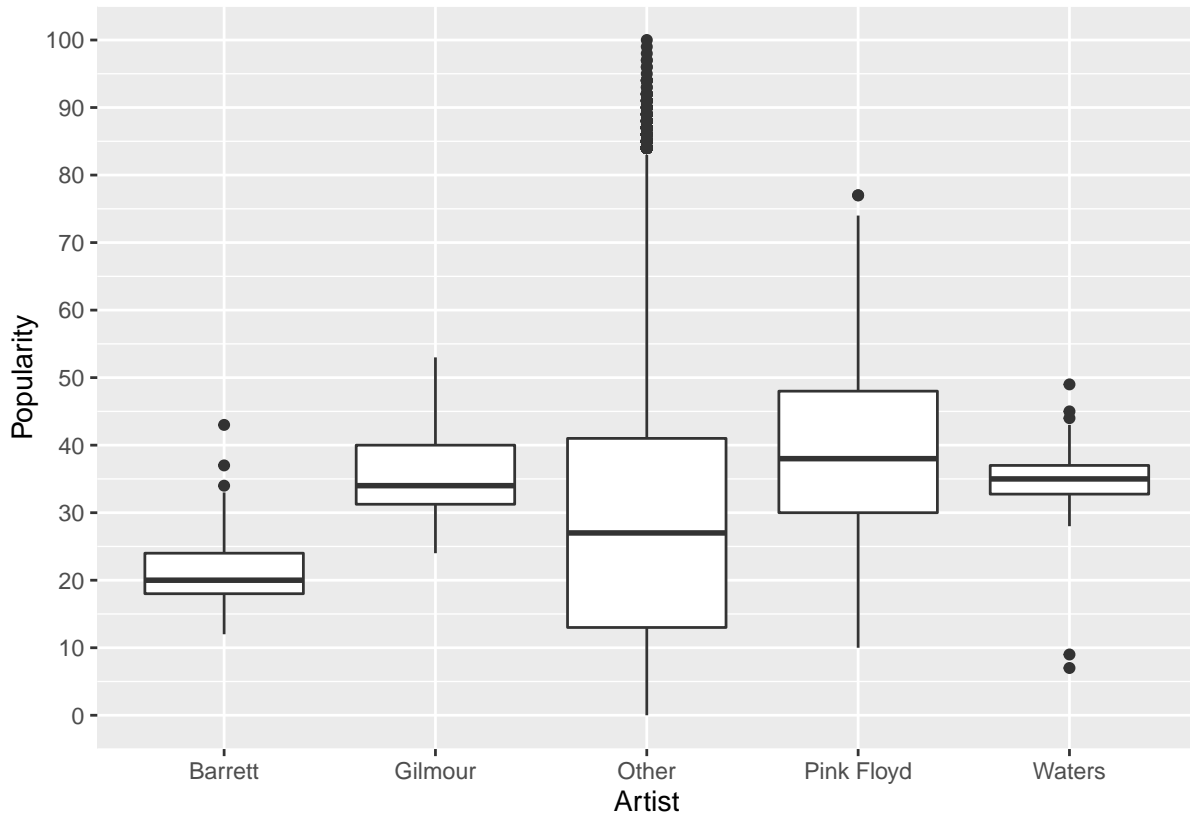


Figure 3.2: Song Popularity of Pink Floyd and Members

The five most popular Pink Floyd songs:

```
popularity |>
  filter(str_detect(artists, 'Pink Floyd')) |>
  arrange(desc(popularity)) |>
  select(name, popularity) |>
  head(5) |>
  rename_with(str_to_title) |>
  kable()
```

Table 3.2: Five Most Popular Pink Floyd Songs

Name	Popularity
Wish You Were Here	77
Another Brick in the Wall, Pt. 2	77
Comfortably Numb	74

Name	Popularity
Money	73
Another Brick In The Wall, Pt. 2 - 2011 Remastered Version	72

The five least popular Pink Floyd songs:

```
popularity |>
  filter(str_detect(artists, 'Pink Floyd')) |>
  arrange(popularity) |>
  select(name, popularity) |>
  head(5) |>
  rename_with(str_to_title) |>
  kable()
```

Table 3.3: Five Least Popular Pink Floyd Songs

Name	Popularity
Song 1 - Capitol Studio Session, 22 August 1968, 2016 Mix	10
Song 2 - Roger's Boogie, Capitol Studio Session, 22 August 1968, 2016 Mix	10
Careful With That Axe, Eugene - Single Version, 2016 Remastered Version	13
Absolutely Curtains - 2016 Remix	13
On the Turning Away	13