Netflix im Wandel der Zeit

Márk Reichmann, Simon Keil, Daniel Henke

5. Februar 2022

Tibble modifizieren (1)

unser Datensatz (die interessanten Variablen):

```
## # A tibble: 12,187 x 4
##    Title                       year Genre            country
##    <chr>                      <int> <chr>            <chr>
##  1 Lets Fight Ghost            2008 Crime, Drama, F~ Thailand
##  2 ANNE+                       2016 Drama            Belgium,Netherlands
##  3 Moxie                       2011 Animation, Shor~ Lithuania,Poland,France,Ic~
##  4 Gleboka woda                2011 Drama            Poland
##  5 Instynkt                    2011 Crime            Poland
##  6 Snowroller                  1985 Comedy           Lithuania,Poland,France,It~
##  7 Sunes Summer                2018 Comedy, Family,~ Lithuania,Poland,France,It~
##  8 The Invisible               2007 Crime, Drama, F~ Lithuania,Poland,France,It~
##  9 The Simple Minded Murderer  1985 Drama            Lithuania,Poland,France,It~
## 10 The Stig-Helmer Story       2011 Comedy, Drama    Lithuania,Poland,France,It~
## # ... with 12,177 more rows
```

Tibble modifizieren (2)

`year` und `Title` sind nur zusammen eindeutig

```
## # A tibble: 2 x 3
##   Title                                year `Series or Movie`
##   <chr>                               <int> <chr>
## 1 Cloudy with a Chance of Meatballs  2017 Series
## 2 Cloudy with a Chance of Meatballs  2009 Movie
```

Tibble modifizieren (3)

leider gibt es Duplikate:

```
data %>%
  count(Title, year)%>%
  filter(n>1)
```

```
## # A tibble: 150 x 3
##    Title                                year     n
##    <chr>                               <int> <int>
##  1 A Bride for Rip Van Winkle           2016     2
##  2 A Love So Beautiful                  2017     2
##  3 A Perfect Day                        2016     2
##  4 About Time                           2013     2
##  5 Aftermath                            2017     2
##  6 Aliens                               1986     2
##  7 Alone                                2008     2
##  8 Anastasia                            1997     2
##  9 Anohana: The Flower We Saw That Day  2011     2
## 10 Another                              2012     2
## # ... with 140 more rows
```

Tibble modifizieren (4)

das sind diejenigen Filme, die mehrmals auf Netflix releast wurden

```
## # A tibble: 2 x 4
##   Title          year Genre              country
##   <chr>         <int> <chr>              <chr>
## 1 A Perfect Day  2016 Comedy, Drama, War Italy
## 2 A Perfect Day  2016 Comedy, Drama, War Romania,Poland
```

Lösung:

```
data %>%
  filter(Title == "A Perfect Day") %>%
  select(Title, year, country)%>%
  separate_rows(country, sep = ",")%>%
  distinct(.keep_all = TRUE)
```

```
## # A tibble: 3 x 3
##   Title          year country
##   <chr>         <int> <chr>
## 1 A Perfect Day  2016 Italy
## 2 A Perfect Day  2016 Romania
## 3 A Perfect Day  2016 Poland
```

Tibble modifizieren (5)

selbiges auch mit Genre

```
## # A tibble: 3 x 3
##   Title          year country
##   <chr>         <int> <chr>
## 1 A Perfect Day  2016 Italy
## 2 A Perfect Day  2016 Romania
## 3 A Perfect Day  2016 Poland

## # A tibble: 3 x 3
##   Title          year Genre
##   <chr>         <int> <chr>
## 1 A Perfect Day  2016 Comedy
## 2 A Perfect Day  2016 Drama
## 3 A Perfect Day  2016 War
```

diese beiden Tibbles werden jetzt an `Title`, `year` gejoint

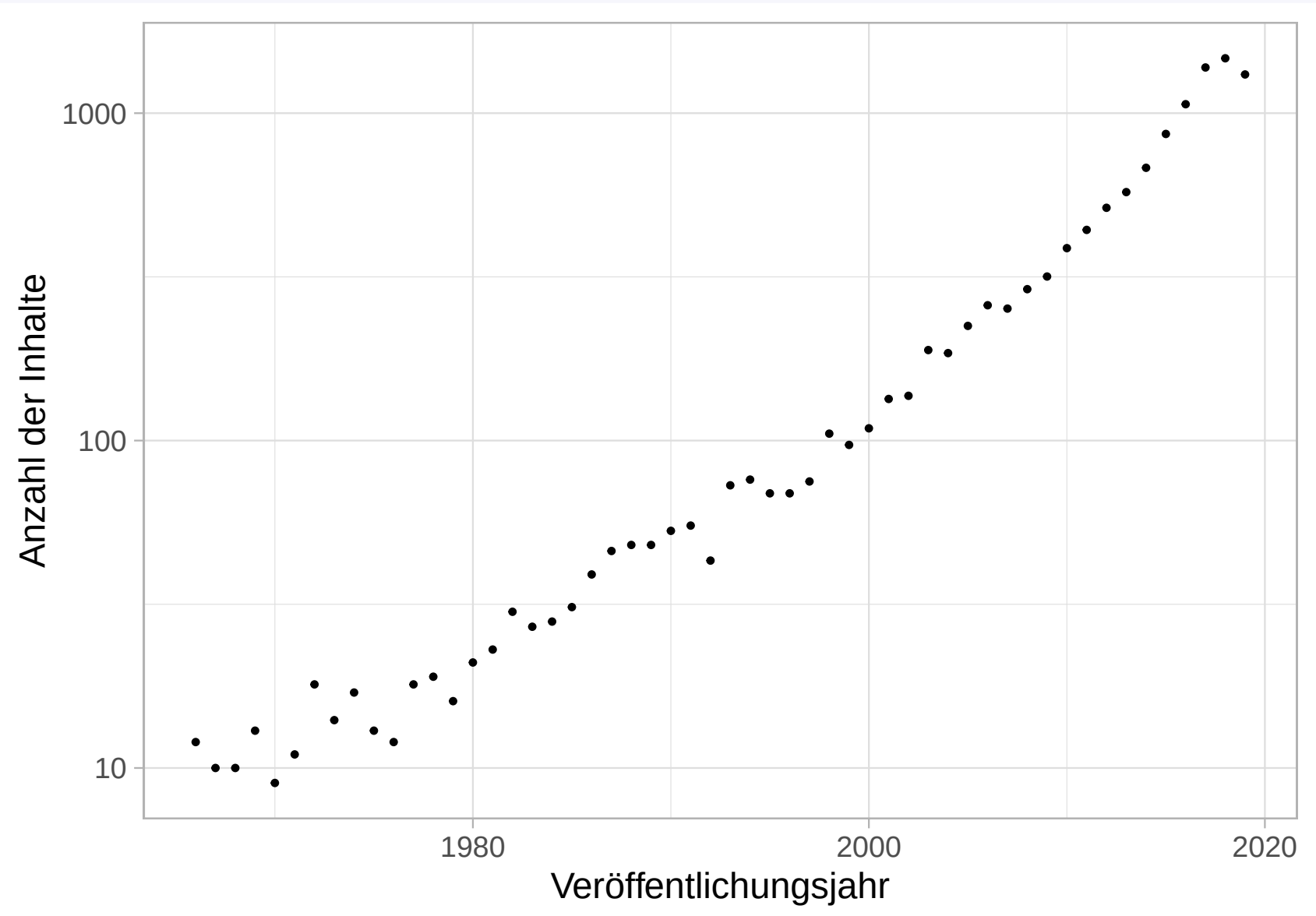
Tibble modifizieren (6)

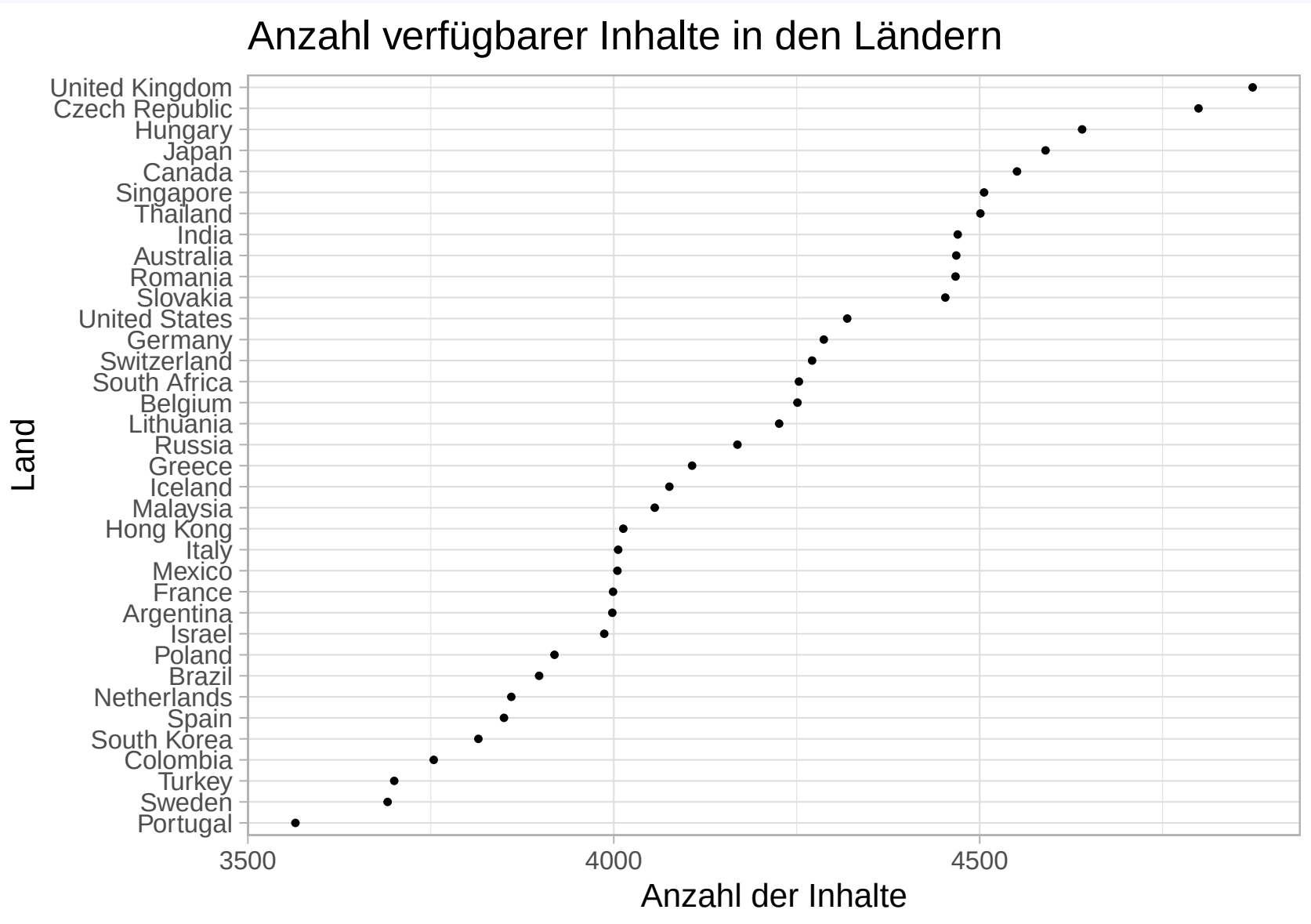
```
data_tidy <- full_join(countries, genres, by = c("year", "Title"))
data_tidy
```

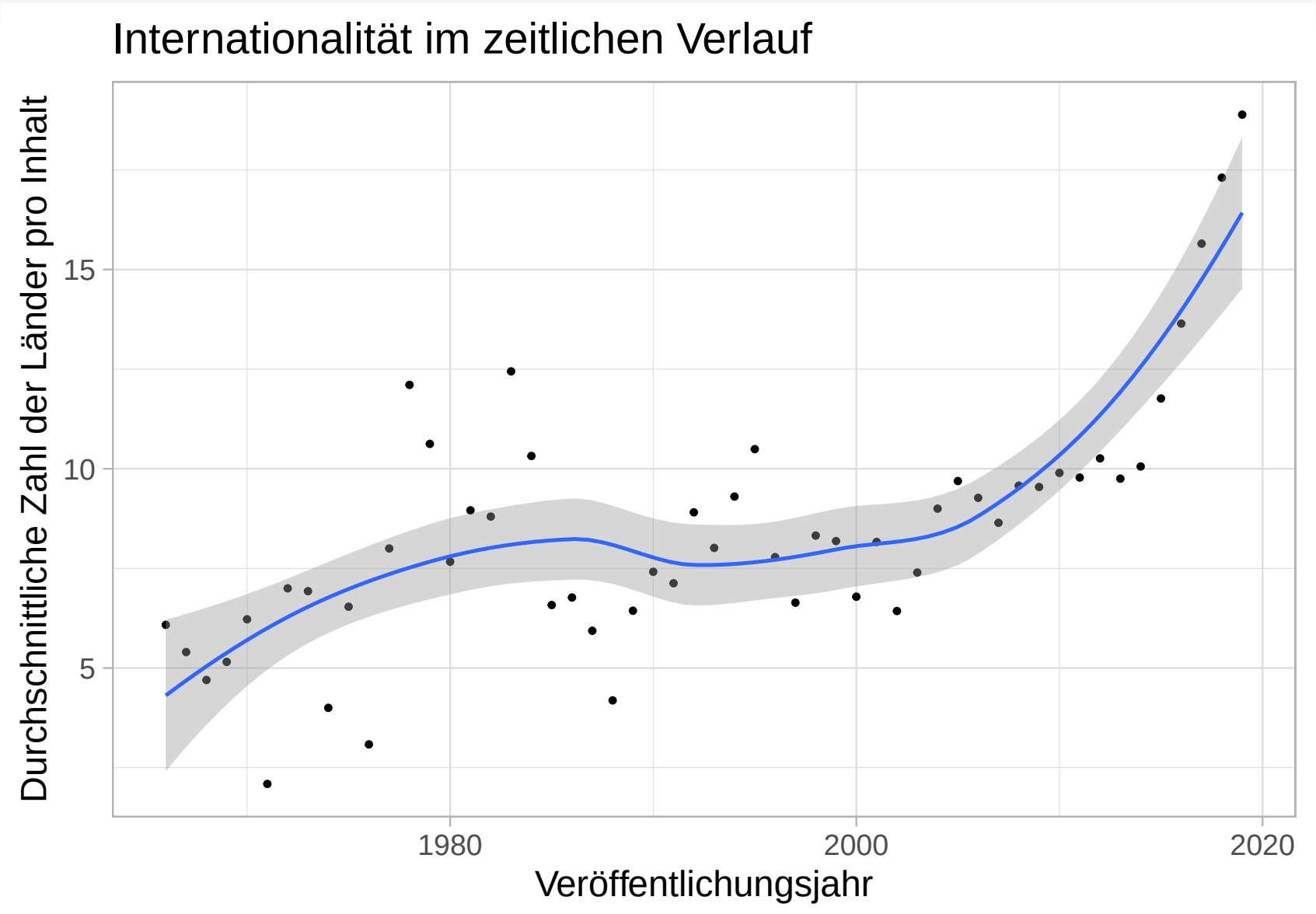
```
## # A tibble: 9 x 4
##   Title          year country Genre
##   <chr>         <int> <chr>   <chr>
## 1 A Perfect Day  2016 Italy   Comedy
## 2 A Perfect Day  2016 Italy   Drama
## 3 A Perfect Day  2016 Italy   War
## 4 A Perfect Day  2016 Romania Comedy
## 5 A Perfect Day  2016 Romania Drama
## 6 A Perfect Day  2016 Romania War
## 7 A Perfect Day  2016 Poland  Comedy
## 8 A Perfect Day  2016 Poland  Drama
## 9 A Perfect Day  2016 Poland  War
```

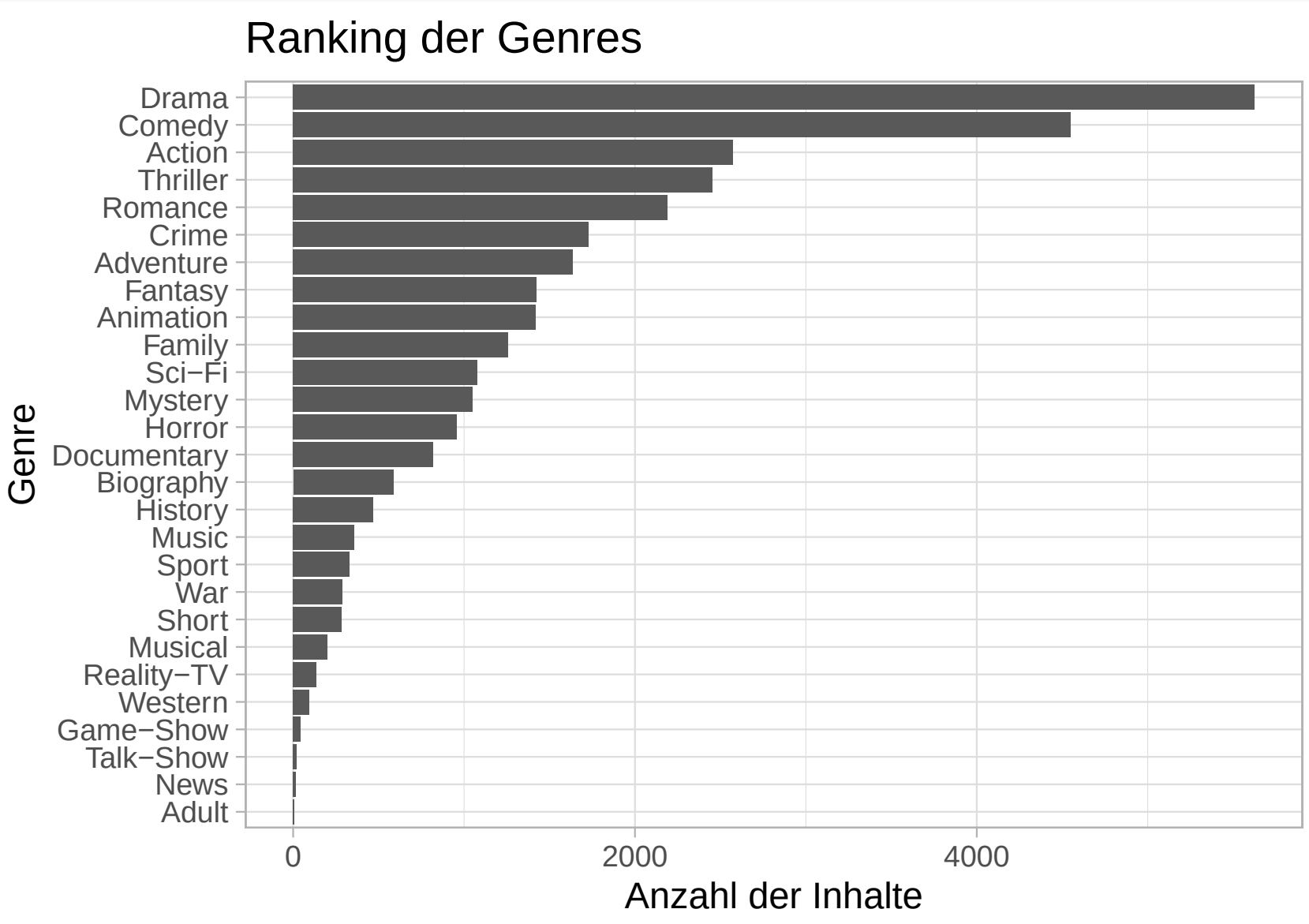
```
data_tidy %>%
  group_by(Title, year, country) %>%
  summarise(Genre = paste(Genre, collapse = ", ")) %>%
  group_by(Title, year, Genre) %>%
  summarise(country = paste(country, collapse = ", ")) %>%
  ungroup()
```

```
## # A tibble: 1 x 4
##   Title          year Genre              country
##   <chr>         <int> <chr>              <chr>
## 1 A Perfect Day  2016 Comedy, Drama, War Italy, Poland, Romania
```

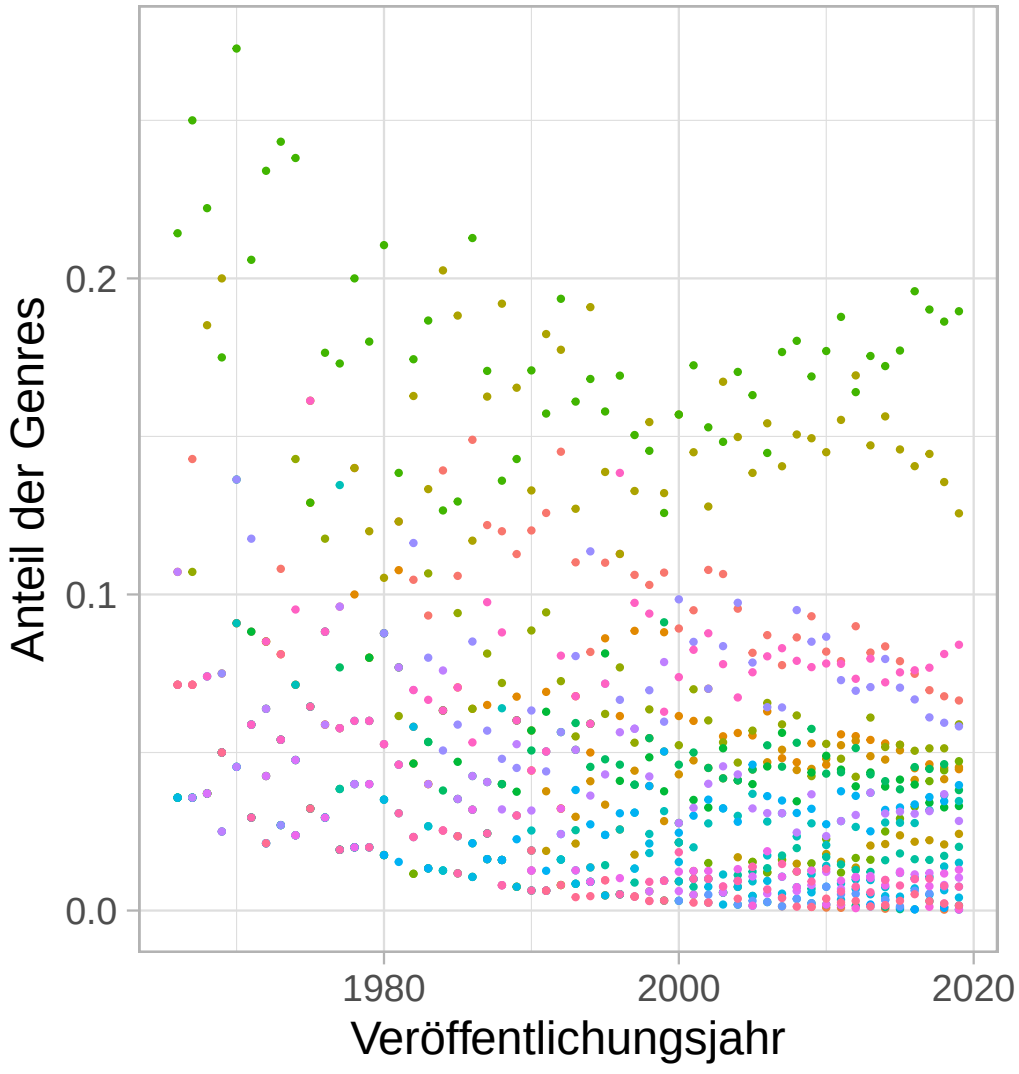

1000
100
10
Anzahl der Inhalte
1980
2000
2020
Veröffentlichungsjahr


Anzahl verfügbarer Inhalte in den Ländern
Land
United Kingdom
Czech Republic
Hungary
Japan
Canada
Singapore
Thailand
India
Australia
Romania
Slovakia
United States
Germany
Switzerland
South Africa
Belgium
Lithuania
Russia
Greece
Iceland
Malaysia
Hong Kong
Italy
Mexico
France
Argentina
Israel
Poland
Brazil
Netherlands
Spain
South Korea
Colombia
Turkey
Sweden
Portugal
3500
4000
4500
Anzahl der Inhalte


Internationalität im zeitlichen Verlauf
Durchschnittliche Zahl der Länder pro Inhalt
15
10
5
1980
2000
2020
Veröffentlichungsjahr

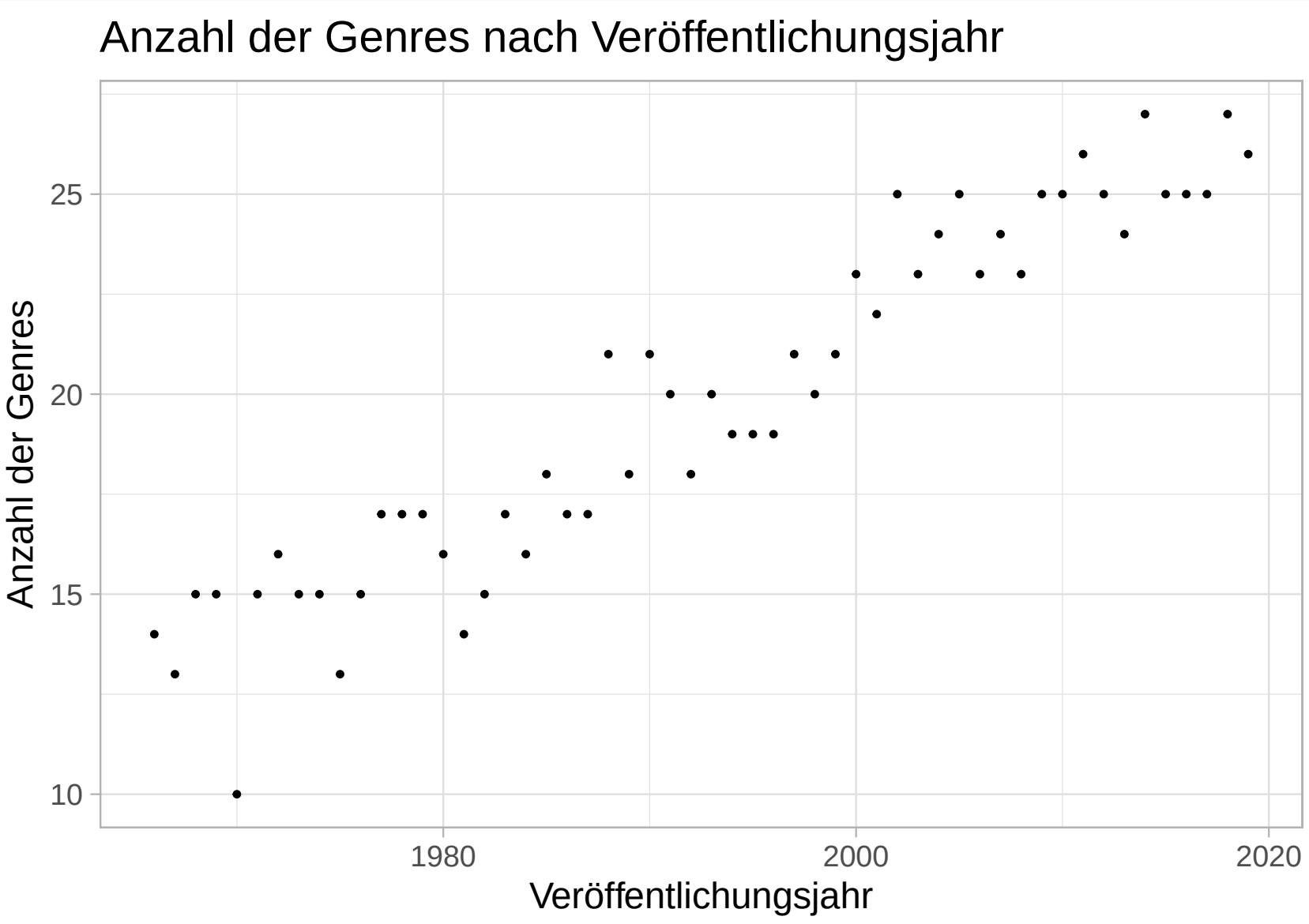

Ranking der Genres
Drama
Comedy
Action
Thriller
Romance
Crime
Adventure
Fantasy
Animation
Family
Sci−Fi
Mystery
Horror
Documentary
Biography
History
Music
Sport
War
Short
Musical
Reality−TV
Western
Game−Show
Talk−Show
News
Adult
Genre
0
2000
4000
Anzahl der Inhalte

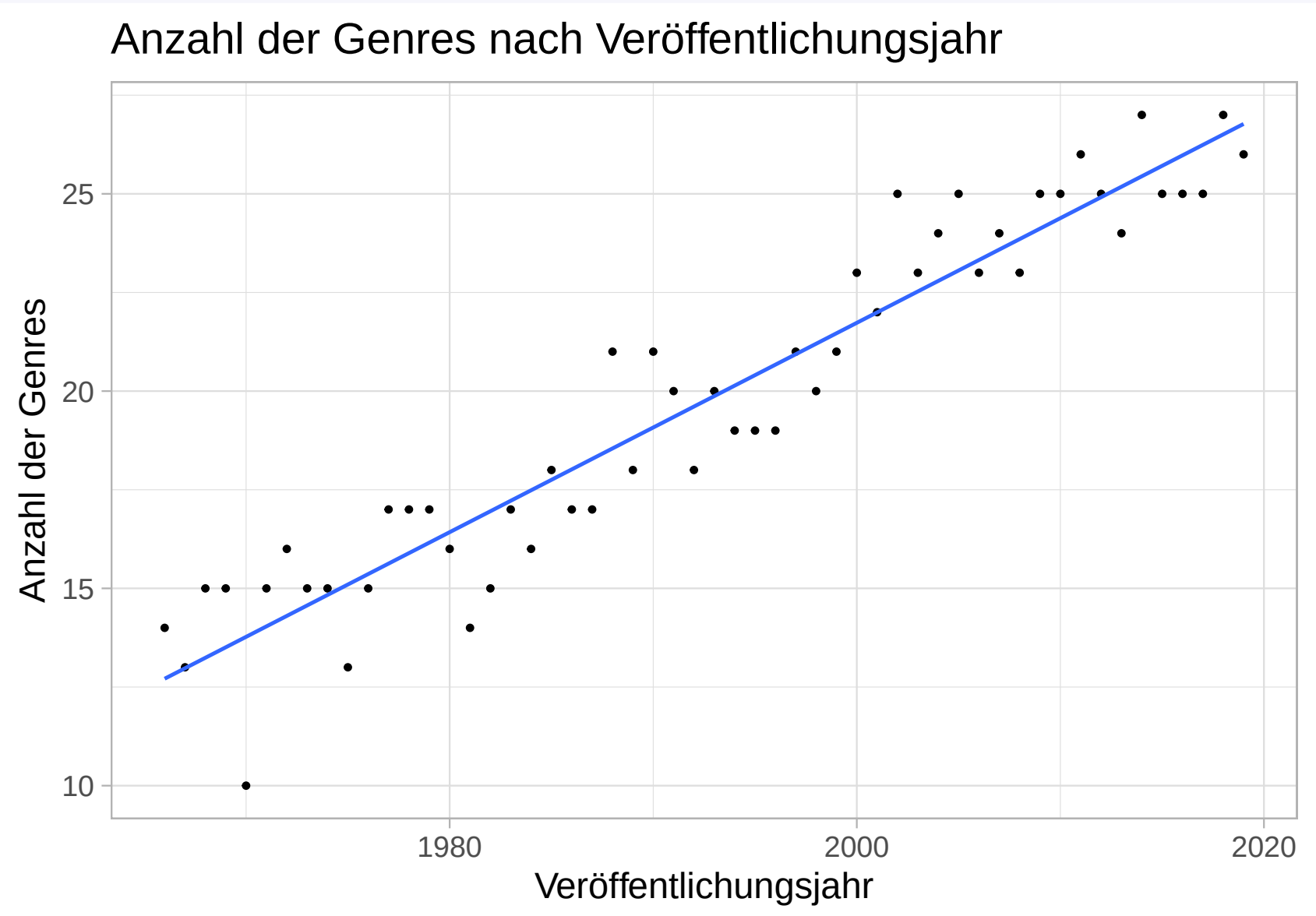
Anteil der verfügbaren Genres nach Veröffentlichungsjahr



Genre

- Action
- Adult
- Adventure
- Animation
- Biography
- Comedy
- Crime
- Documentary
- Drama
- Family
- Fantasy
- Game−Show
- History
- Horror
- Music
- Musical
- Mystery
- News
- Reality−TV
- Romance
- Sci−Fi
- Short
- Sport
- Talk−Show
- Thriller
- War
- Western


Anzahl der Genres nach Veröffentlichungsjahr
Anzahl der Genres
25
20
15
10
1980
2000
2020
Veröffentlichungsjahr


Anzahl der Genres nach Veröffentlichungsjahr
Anzahl der Genres
25
20
15
10
1980
2000
2020
Veröffentlichungsjahr