Movie: Past/Present/Future

Number of Movies/Year

```
library(dplyr)
```

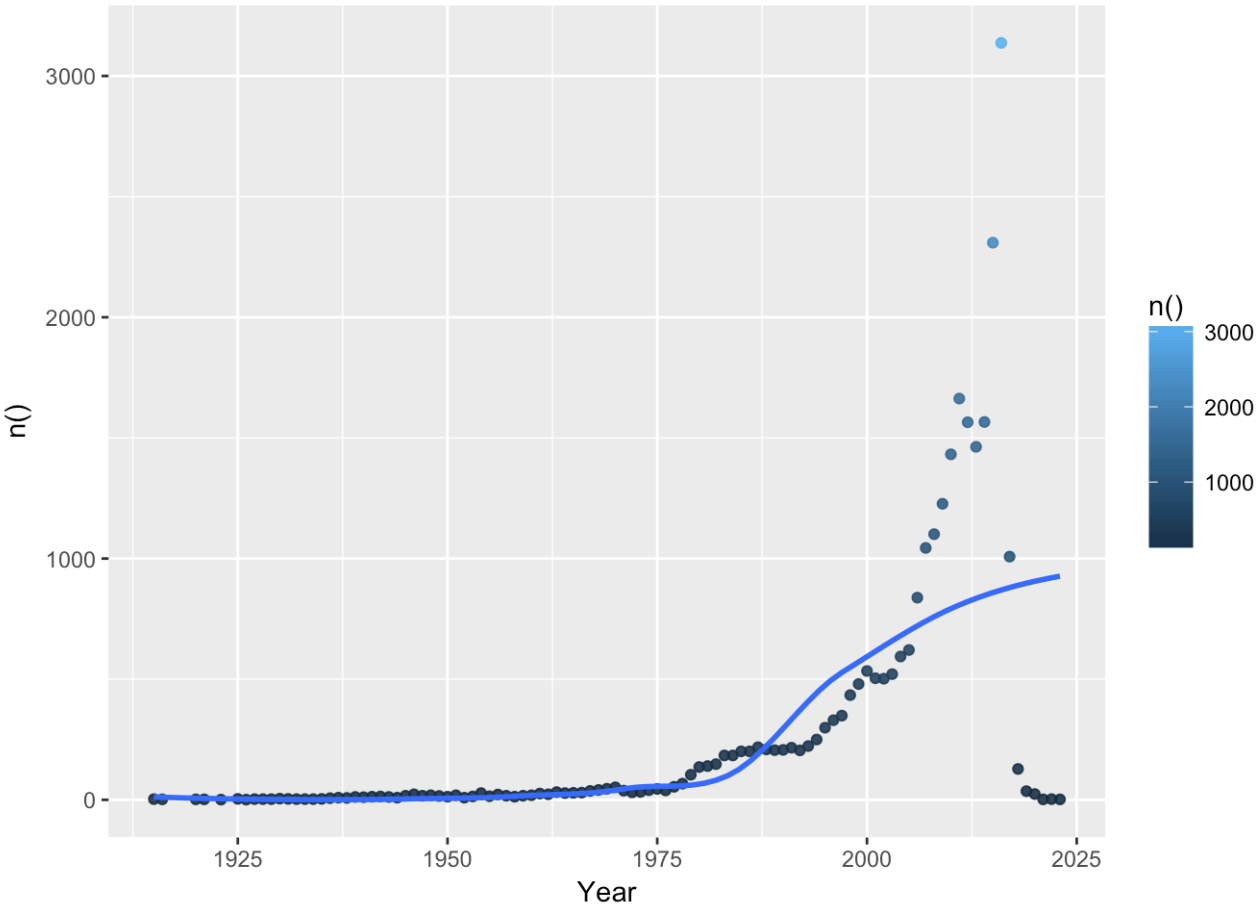
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
DIR<-paste(Sys.getenv("HOME"), "/datascience/springboard/project/", sep="")
setwd(DIR)
df<- read.table(paste(DIR, "MergeDataSet.csv", sep=""), sep=",", header=TRUE)
ldf <- tbl_df(df)
summary(ldf)
```

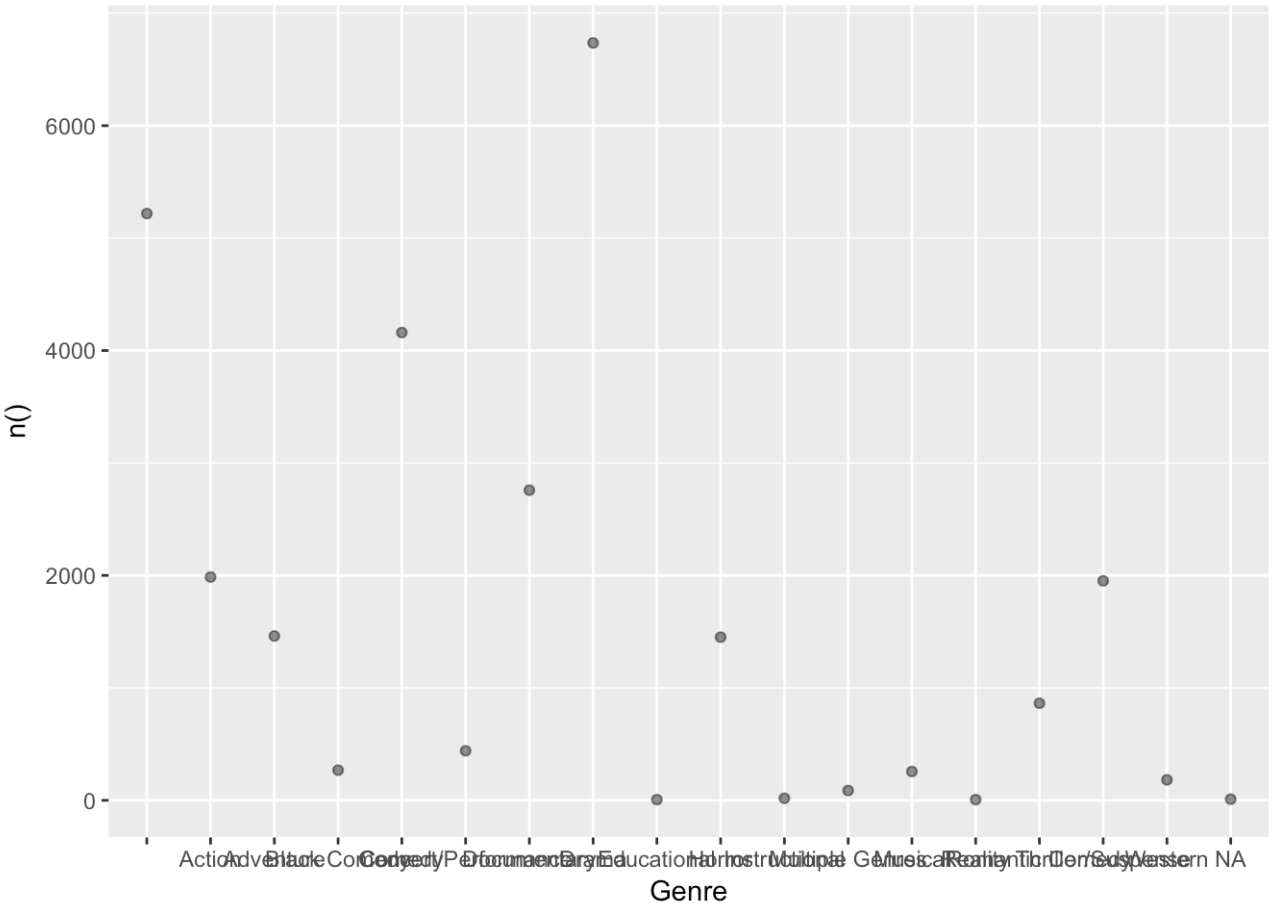
```
##       Release.Date                    Movie                Genre
##              :22561   Hamlet              :    7   Drama      :6736
##  December 31:   94   Home                :    5              :5218
##  January 1  :   76   Pinocchio           :    5   Comedy     :4160
##  December 25:   41   A Star is Born      :    4   Documentary:2758
##  July 1     :   26   Aftermath           :    4   Action     :1986
##  December 21:   25   Beauty and the Beast:    4   (Other)    :6990
##  (Other)    : 5035   (Other)             :27829   NA's       :  10
##     ProductionBudget DomesticBox.Officeto.Date Trailer           Year
##              :22350   $0         :12754             :16917   Min.   :1915
##  $20,000,000:  213   $5,000     :   86         Play: 1448   1st Qu.:2001
##  $10,000,000:  187   $6,000,000 :   33         NA's: 9493   Median :2010
##  $15,000,000:  162   $10,000,000:   30                      Mean   :2005
##  $40,000,000:  159   $3,000,000 :   29                      3rd Qu.:2014
##  (Other)    : 4606   $8,000,000 :   29                      Max.   :2023
##  NA's       :  181   (Other)    :14897
```



Movie Genre

```
category<-ldf %>% group_by(Genre) %>% summarise(n())

p<-ggplot(category, aes(x=Genre, y=`n()`)) + geom_point(alpha=0.5)
print(p)
```


6000
4000
2000
0
n()
Genre
Action
Adventure
Black Comedy
Documentary
Education
Horror
Musical
Reality TV
Thriller/Suspense
Western
NA