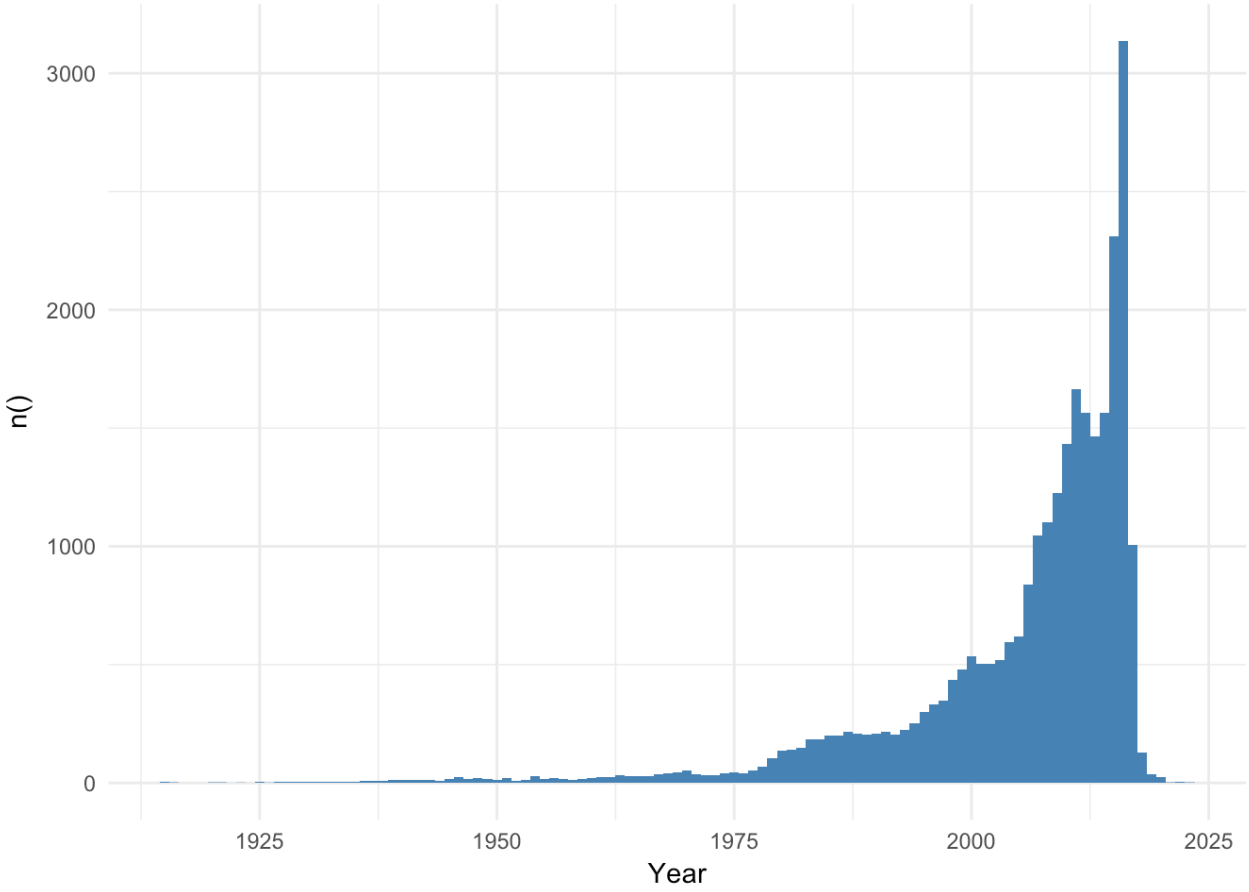
Movie Modeling

Overview

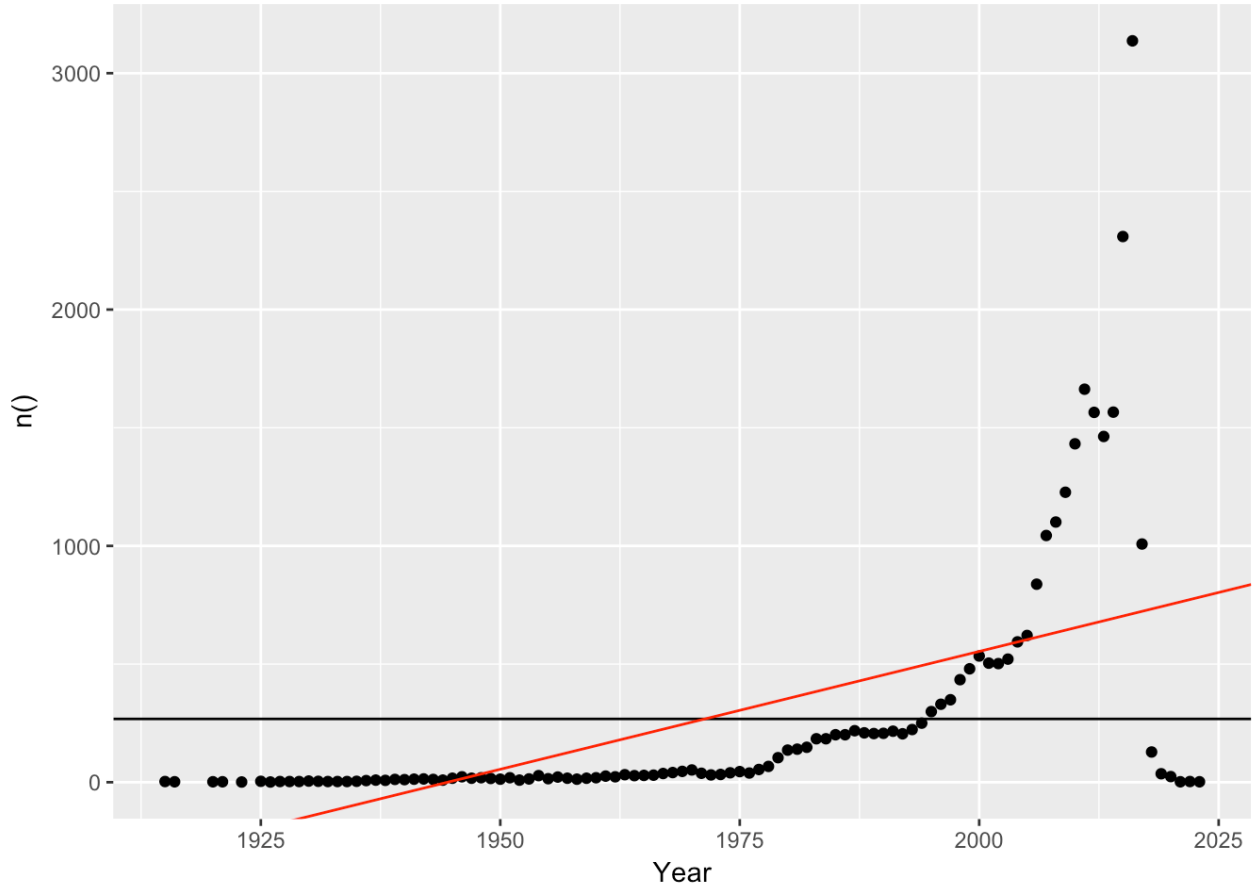
As proposed in initial project submission, this project intends to perform study on the movie industry trends and model
prediction for the future trends.

Number of Movies/Year

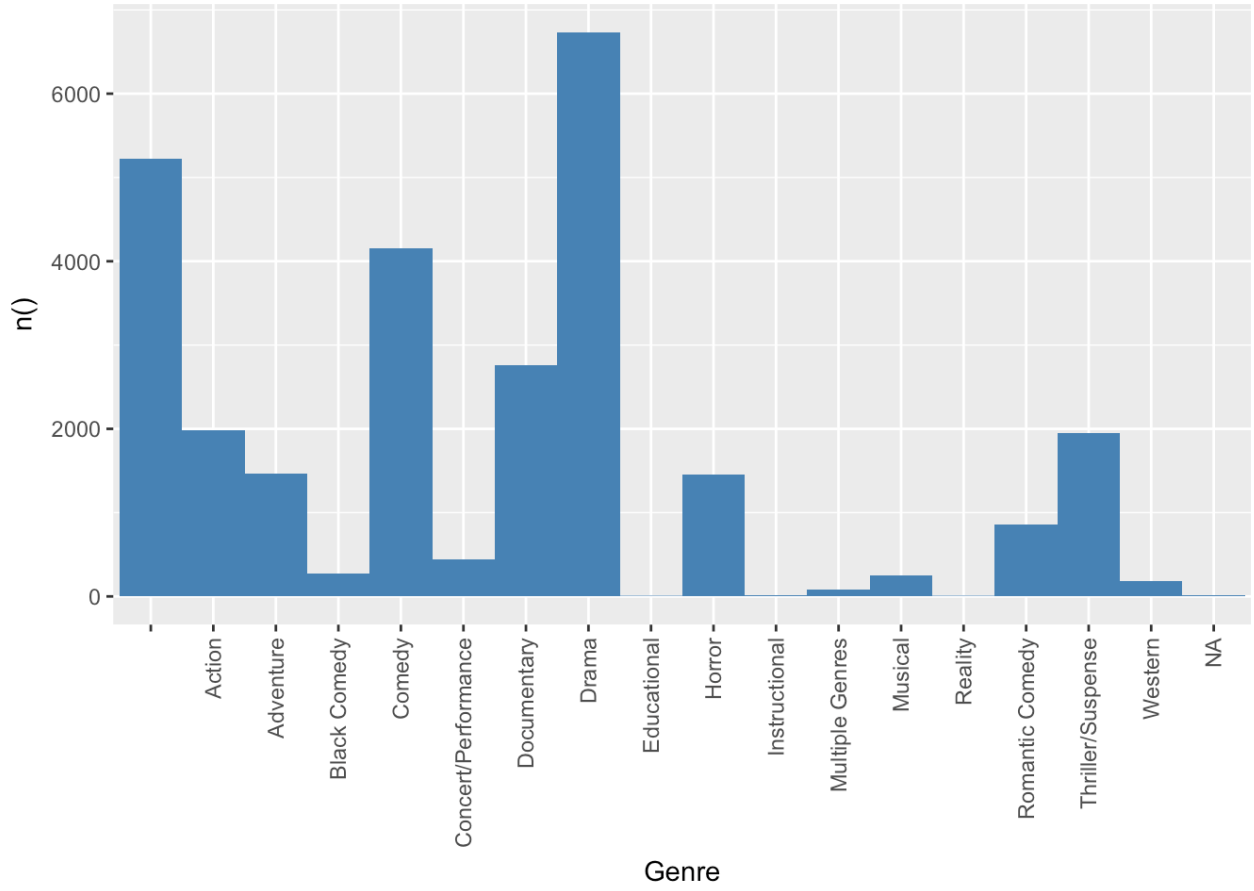


Model Linear Regression

```
#train_data <- subset(ldf, Year > 1975)
train_data <- ldf
yrsum<-train_data %>% group_by(Year) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)
p<-ggplot(yrsum, aes(x=Year, y=`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")

print(p)
```

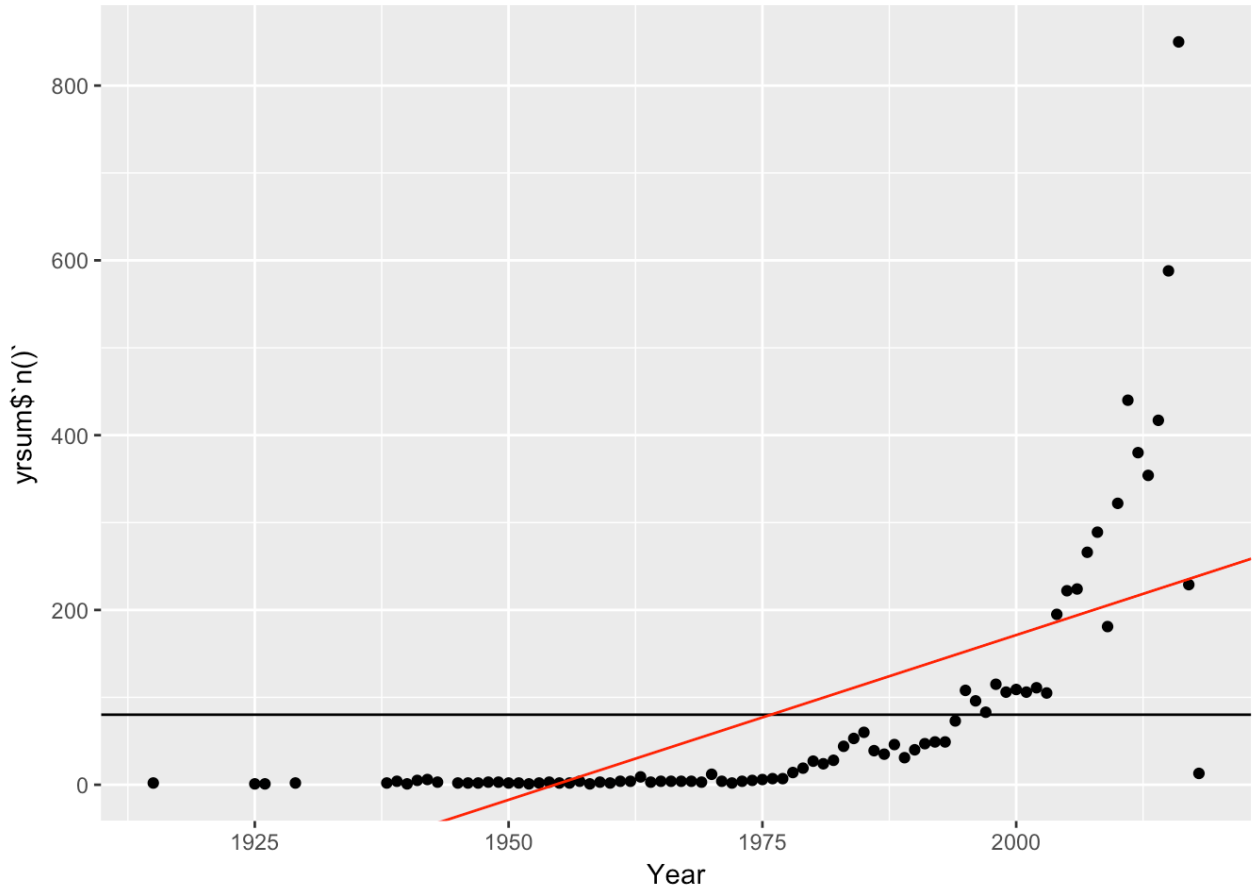


Movies Genre



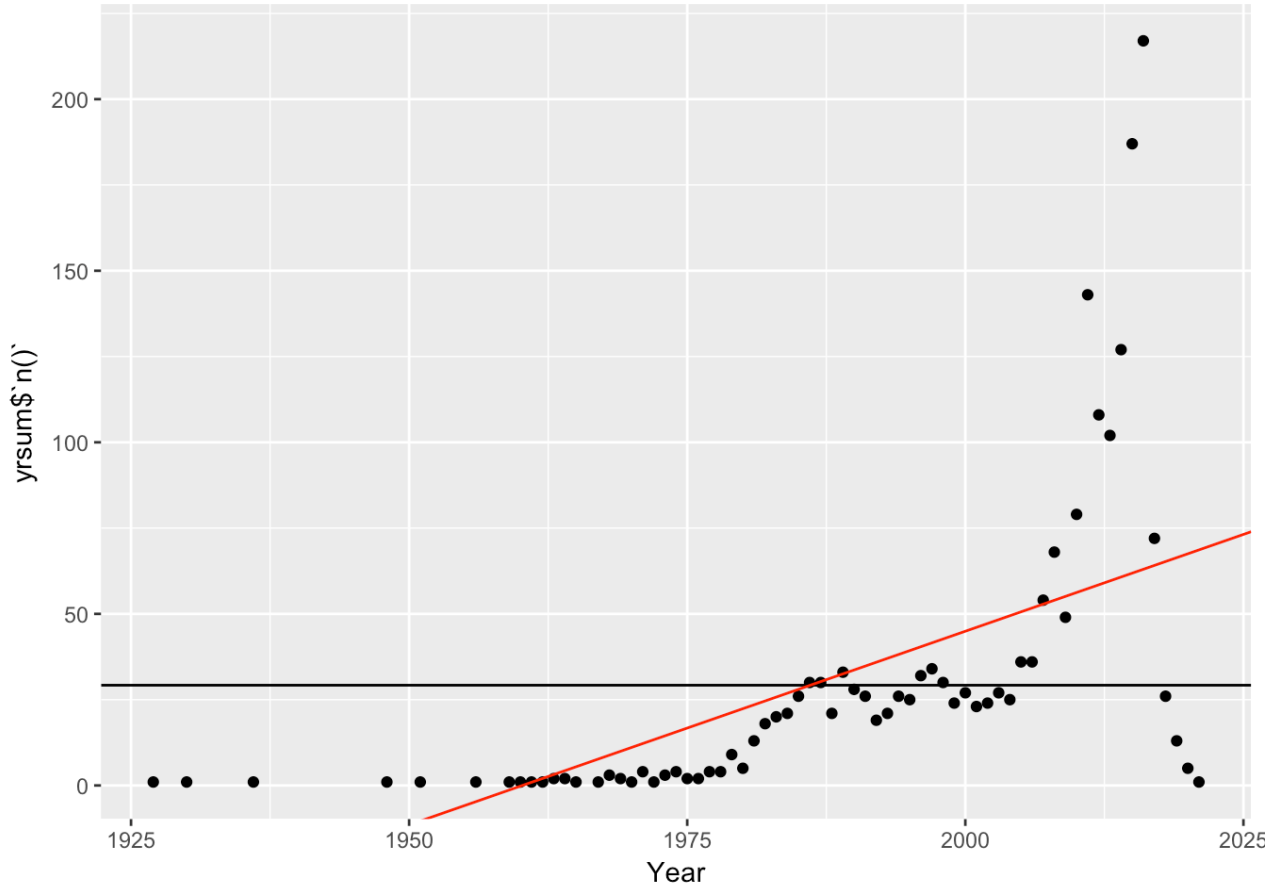
Genre = “Drama”" Linear Regression

```
train_data <- ldf
train_data <- train_data %>% filter(train_data$Genre == "Drama")
yrsum<-train_data %>% group_by(Year, Genre) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)

p<-ggplot(yrsum, aes(x=Year, y=yrsum$`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")
print(p)
```



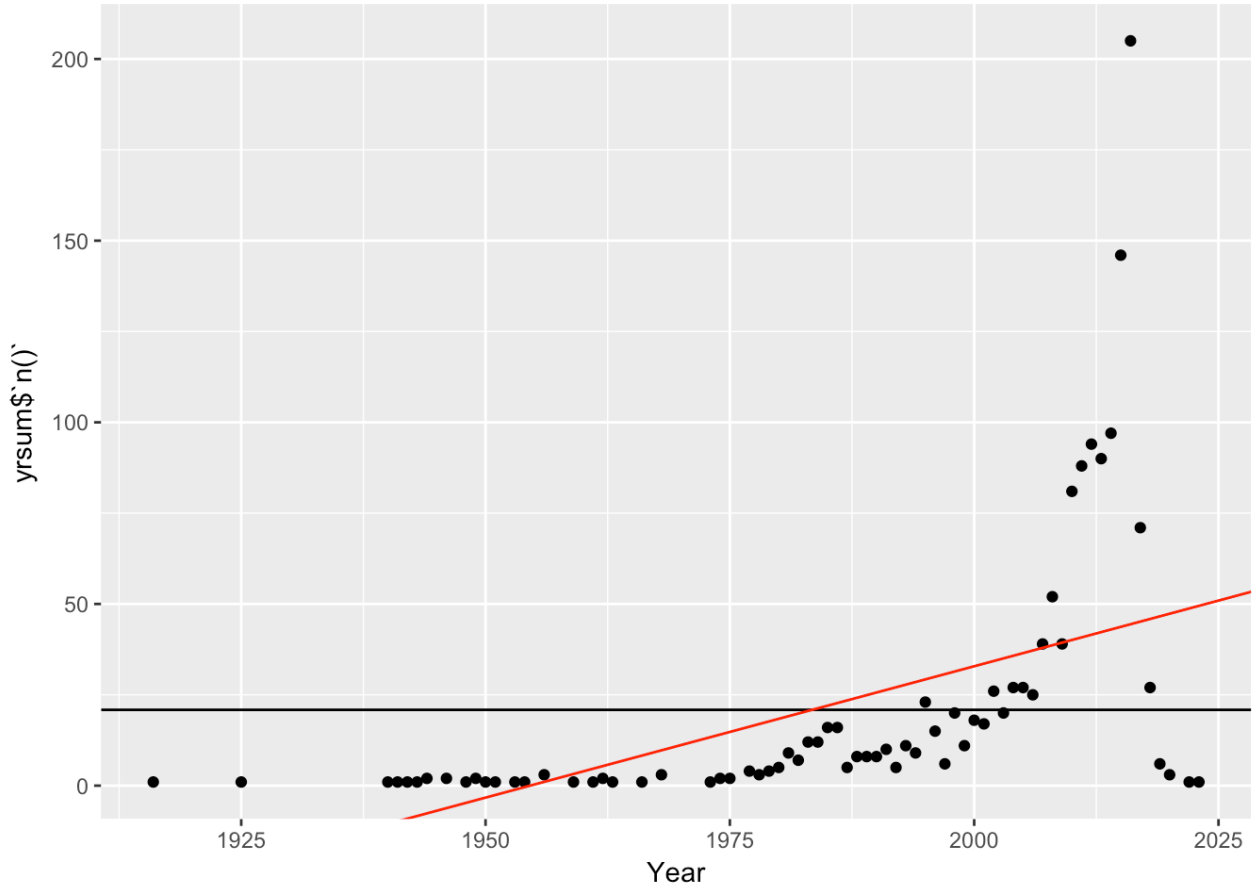
Genre = “Action” Linear Regression

```
train_data <- ldf
train_data <- train_data %>% filter(train_data$Genre == "Action")
yrsum<-train_data %>% group_by(Year, Genre) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)

p<-ggplot(yrsum, aes(x=Year, y=yrsum$`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")
print(p)
```



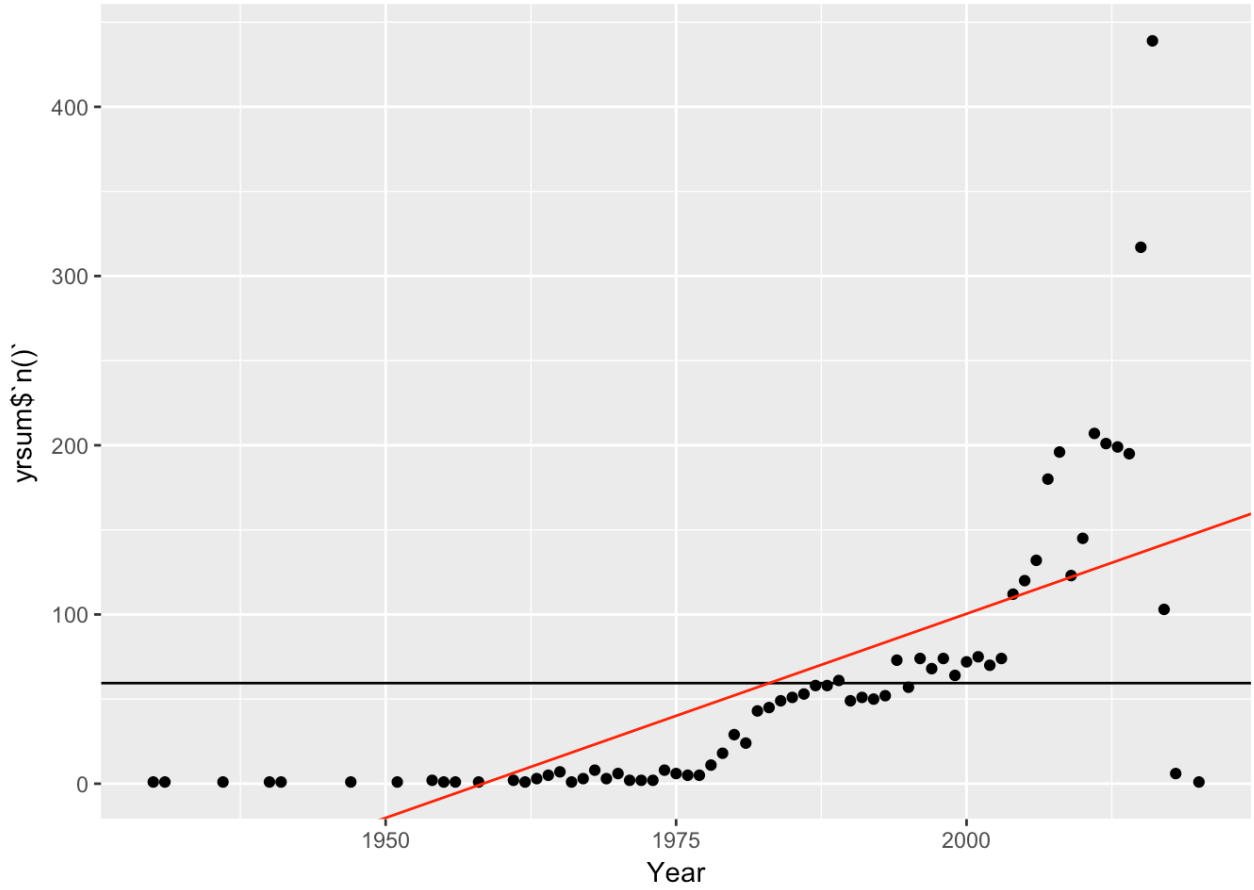
Genre = “Adventure” Linear Regression

```
train_data <- ldf
train_data <- train_data %>% filter(train_data$Genre == "Adventure")
yrsum<-train_data %>% group_by(Year, Genre) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)

p<-ggplot(yrsum, aes(x=Year, y=yrsum$`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")
print(p)
```



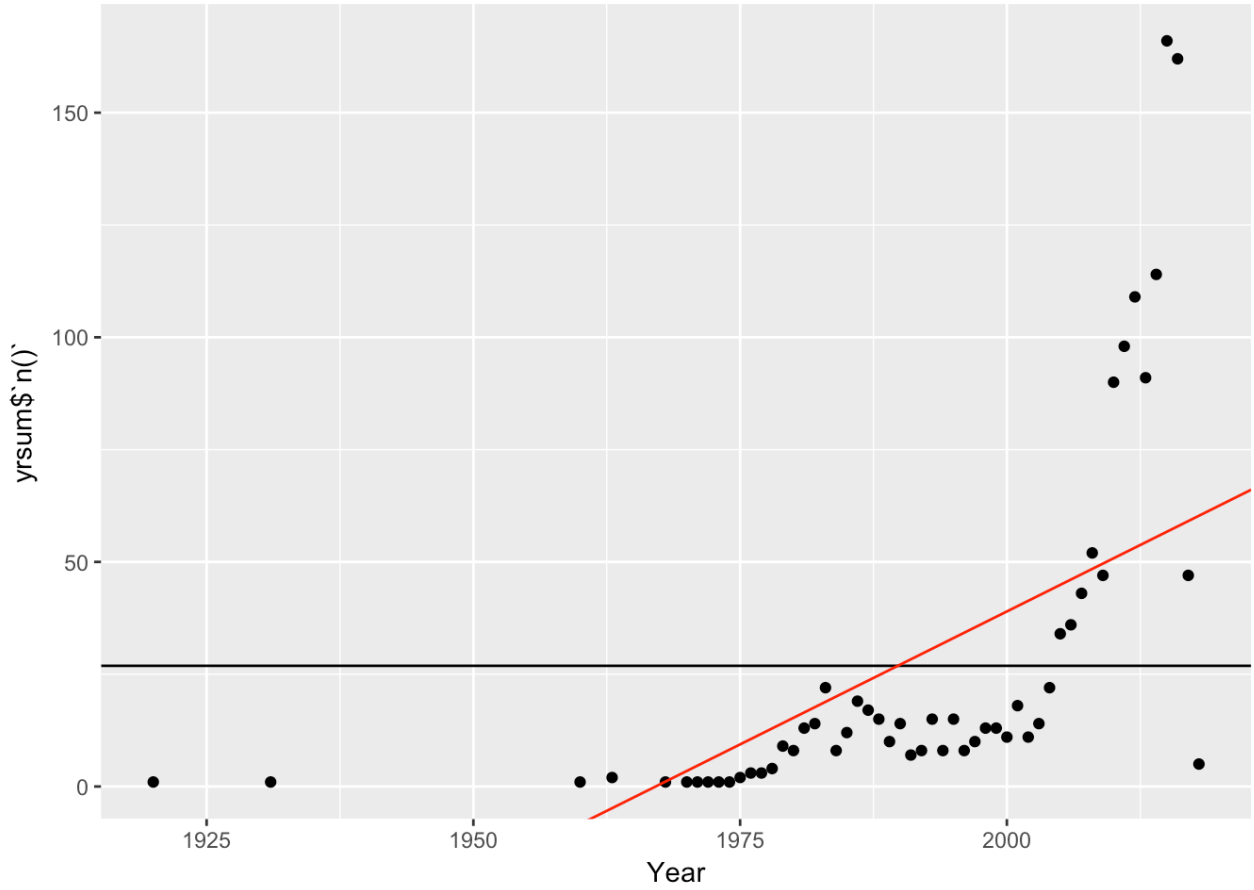
Genre = “Comedy” Linear Regression

```
train_data <- ldf
train_data <- train_data %>% filter(train_data$Genre == "Comedy")
yrsum<-train_data %>% group_by(Year, Genre) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)

p<-ggplot(yrsum, aes(x=Year, y=yrsum$`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")
print(p)
```



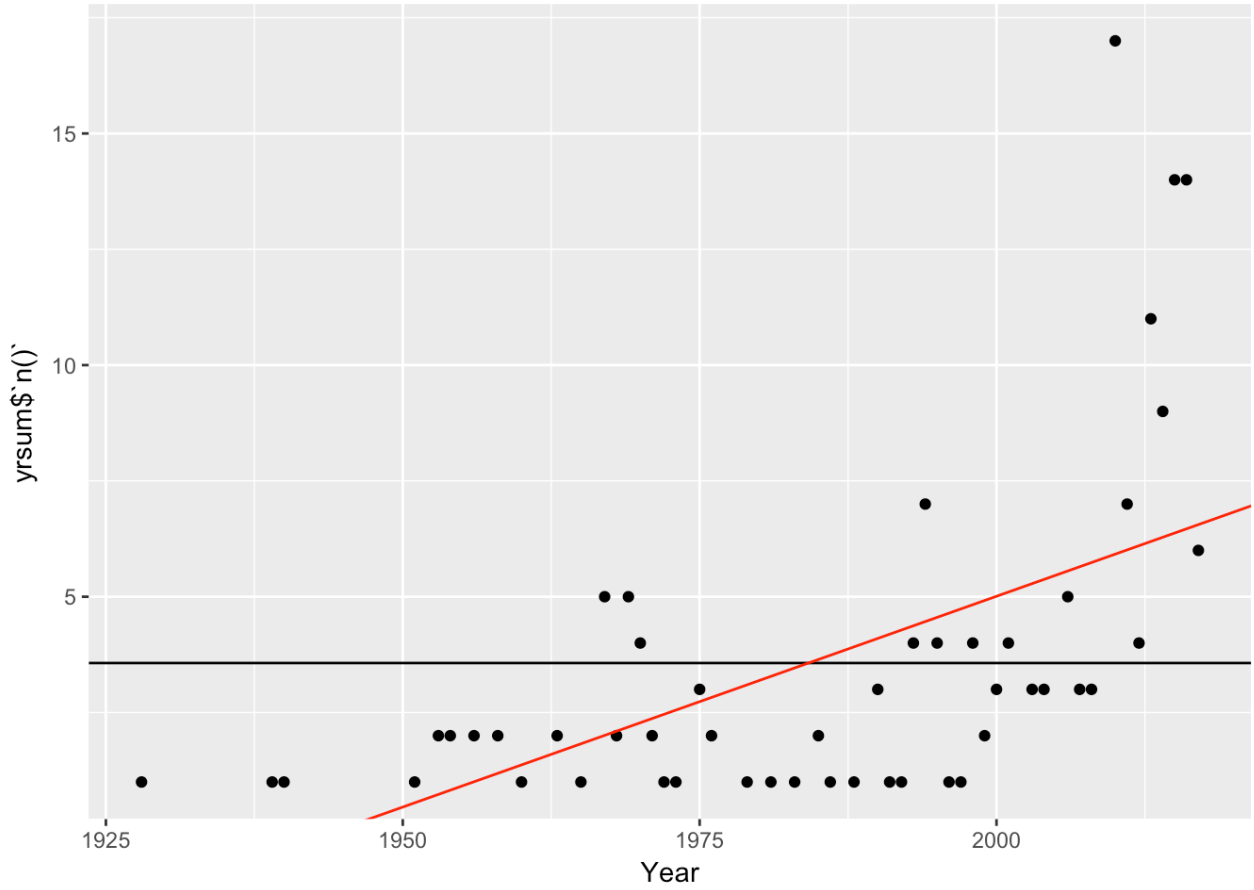
Genre = “Horror” Linear Regression

```
train_data <- ldf
train_data <- train_data %>% filter(train_data$Genre == "Horror")
yrsum<-train_data %>% group_by(Year, Genre) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)

p<-ggplot(yrsum, aes(x=Year, y=yrsum$`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")
print(p)
```



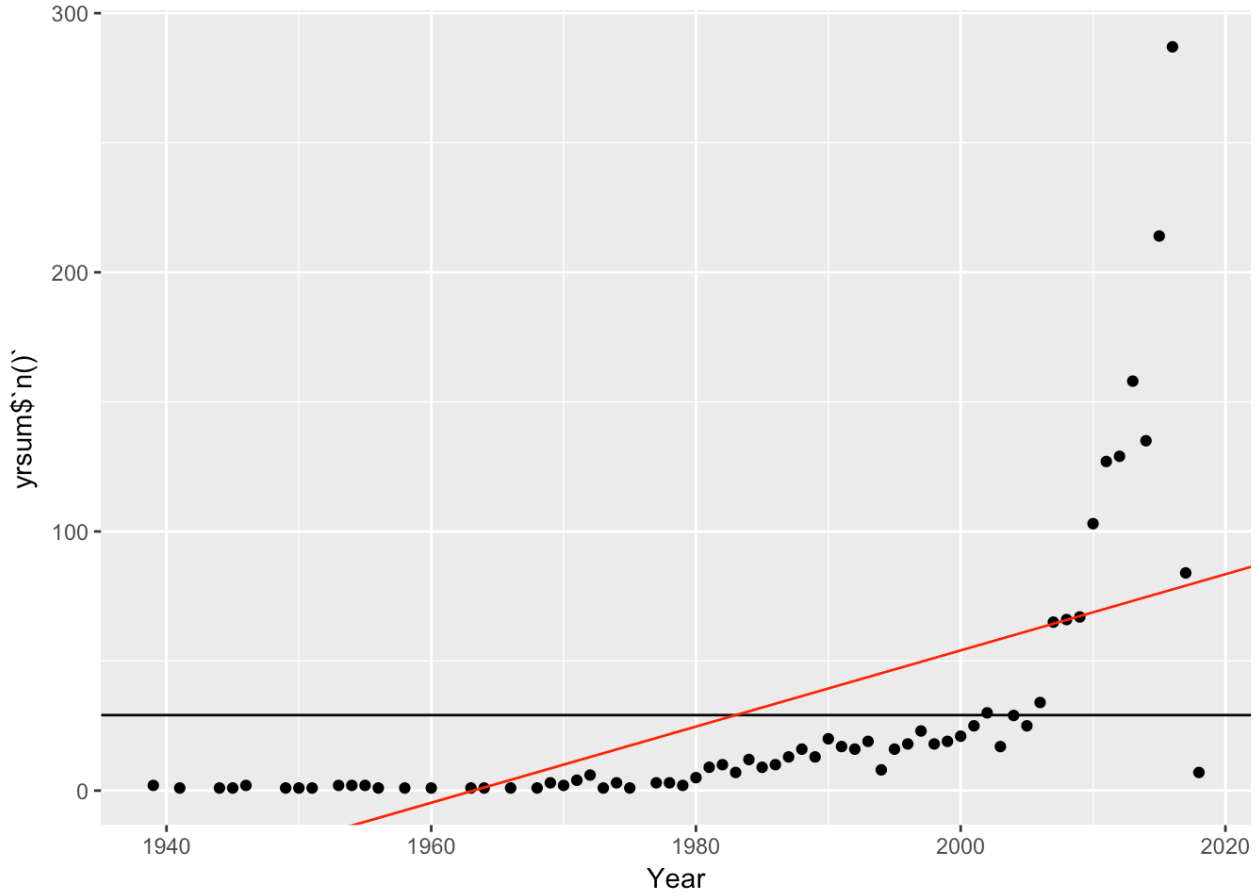
Genre = “Western” Linear Regression

```
train_data <- ldf
train_data <- train_data %>% filter(train_data$Genre == "Western")
yrsum<-train_data %>% group_by(Year, Genre) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)

p<-ggplot(yrsum, aes(x=Year, y=yrsum$`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")
print(p)
```



Genre = “Thriller/Suspense” Linear Regression

```
train_data <- ldf
train_data <- train_data %>% filter(train_data$Genre == "Thriller/Suspense")
yrsum<-train_data %>% group_by(Year, Genre) %>% summarise(n())
mean.num_of_movie=mean(yrsum$`n()`, na.rm=T)
model1 <- lm(yrsum$`n()`~Year, data=yrsum)

p<-ggplot(yrsum, aes(x=Year, y=yrsum$`n()`)) + geom_point() +
  geom_hline(yintercept=mean.num_of_movie) +
  geom_abline(intercept=model1$coefficients[1], slope=model1$coefficients[2],
color="red")
print(p)
```


300
200
100
0
yrsum$`n()`
1940
1960
1980
2000
2020
Year