1. Movies Per Original Language

I used original\_language column as the only variable for point #1. Also, I found that original\_language column is considered categorical column in R.

1. plot

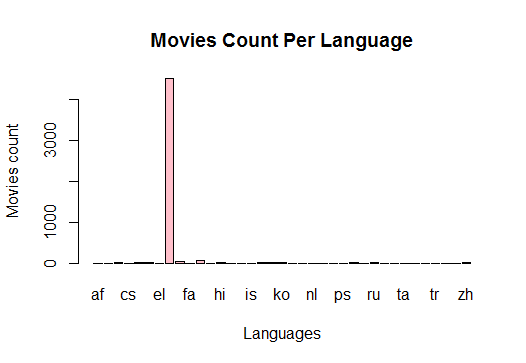
plot(x = data$original\_language,

main = "Movies Count Per Language",

col = "pink",

xlab = "Languages",

ylab = "Movies count")

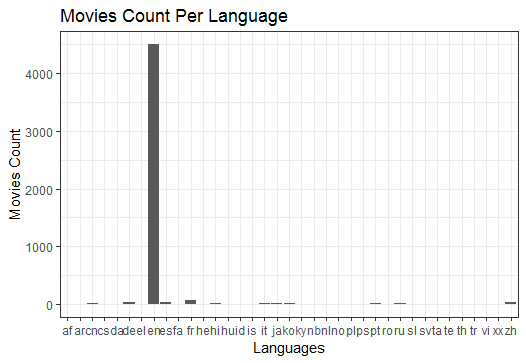


1. ggplot

ggplot(data, aes(x = data$original\_language))+theme\_bw()+geom\_bar()+

labs(y = "Movies Count", x="Languages",

title = "Movies Count Per Language")

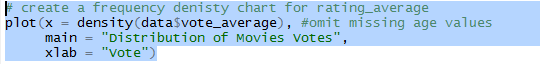


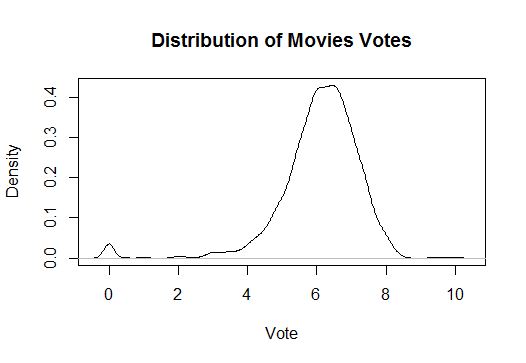
**Findings:** most of the movies are English language movies

2. Voting average

I used voting average as the quantitavive column.

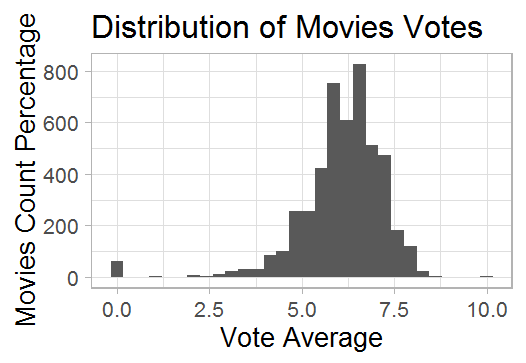
1. plot





1. ggplot



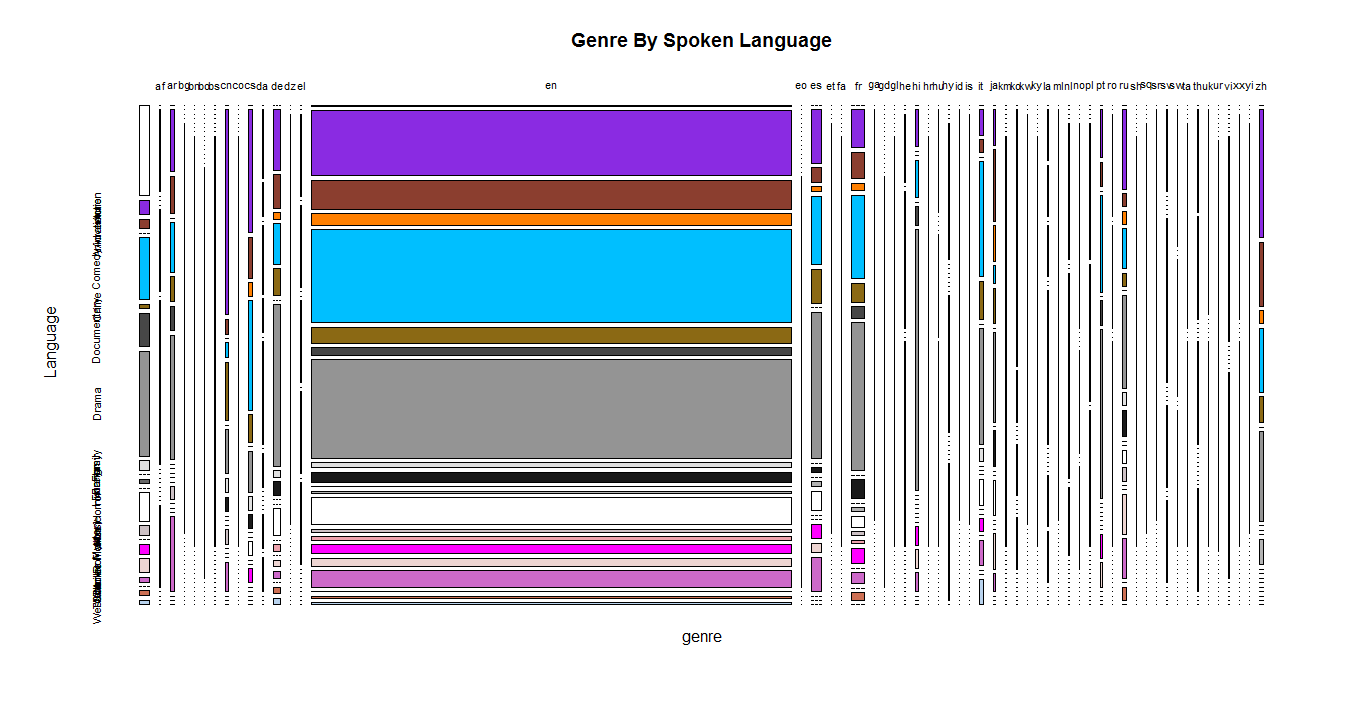


**Findings:** Most of the user’s ratings are in between 5 & 8. So it seems those movies are of a good quality.

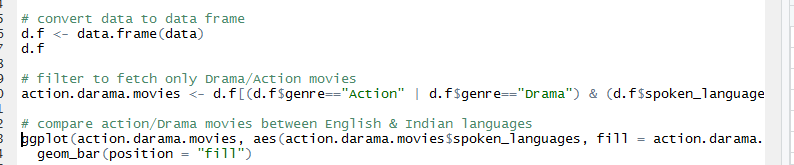
3. Language VS. Genre

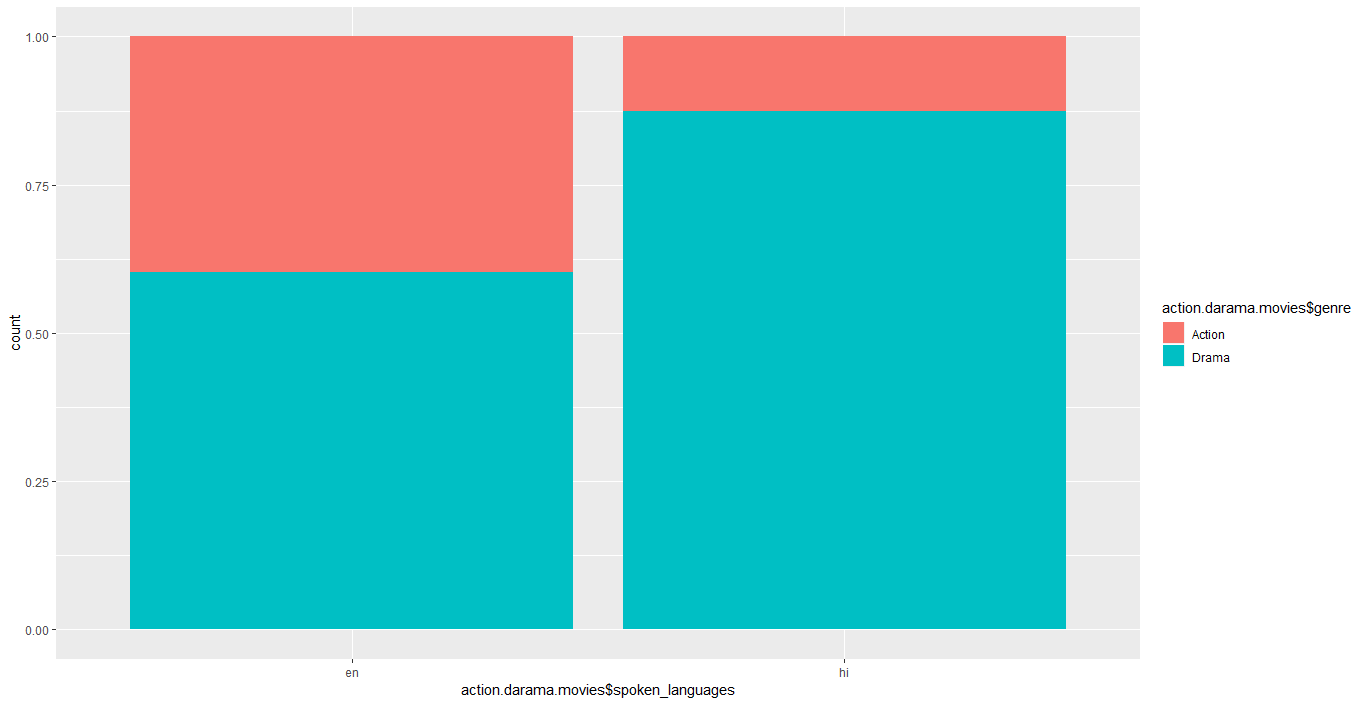
I picked spoken language & Genre for this bivariate analysis. Also, I selected Mosaic plot here.

1. Plot



1. Ggplot2 -> I filtered data to fetch only Action/Drama movies for 2 languages (English, Indian)

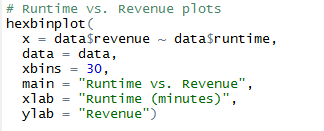


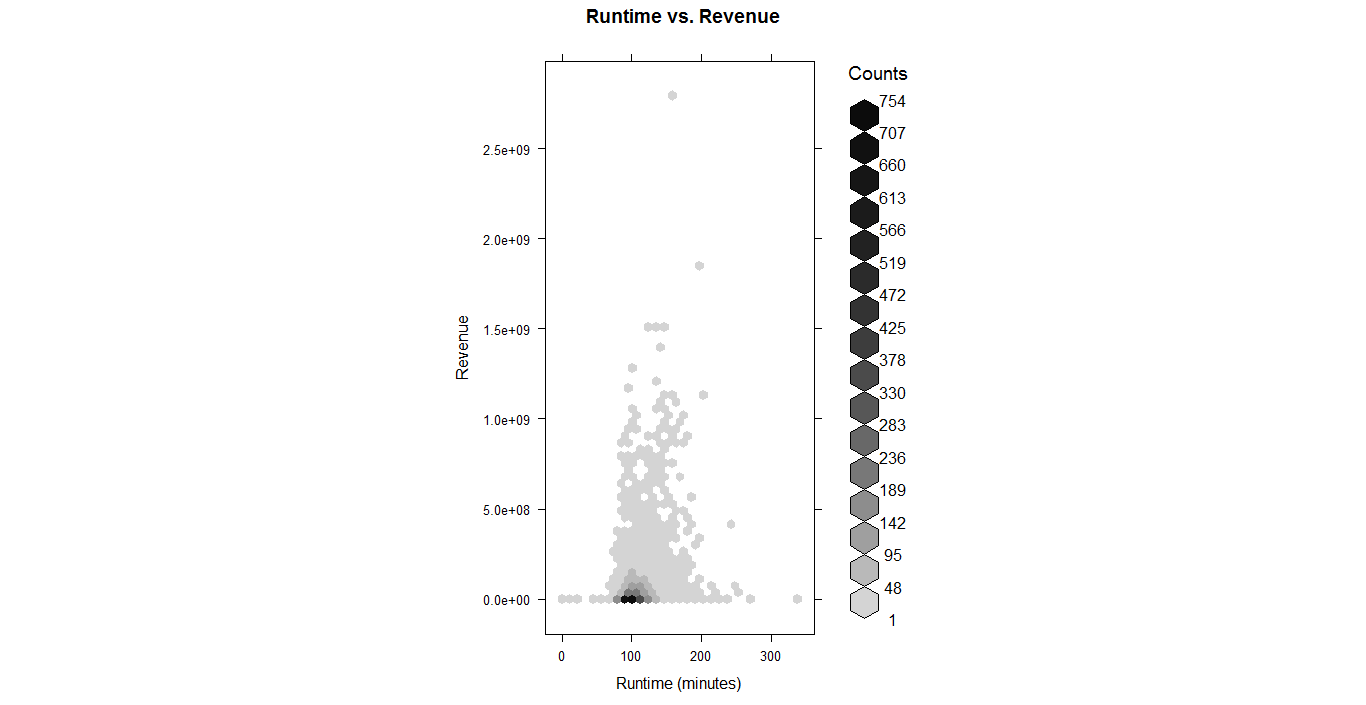


**Findings: apparently there are more interests in drama like movies in the Indian culture. In the English culture it’s more of a 50/50.**

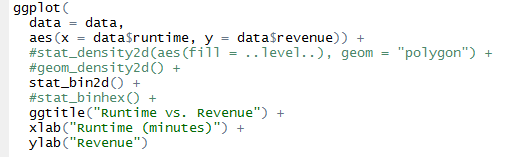
**4. Revenue VS. Runtime**

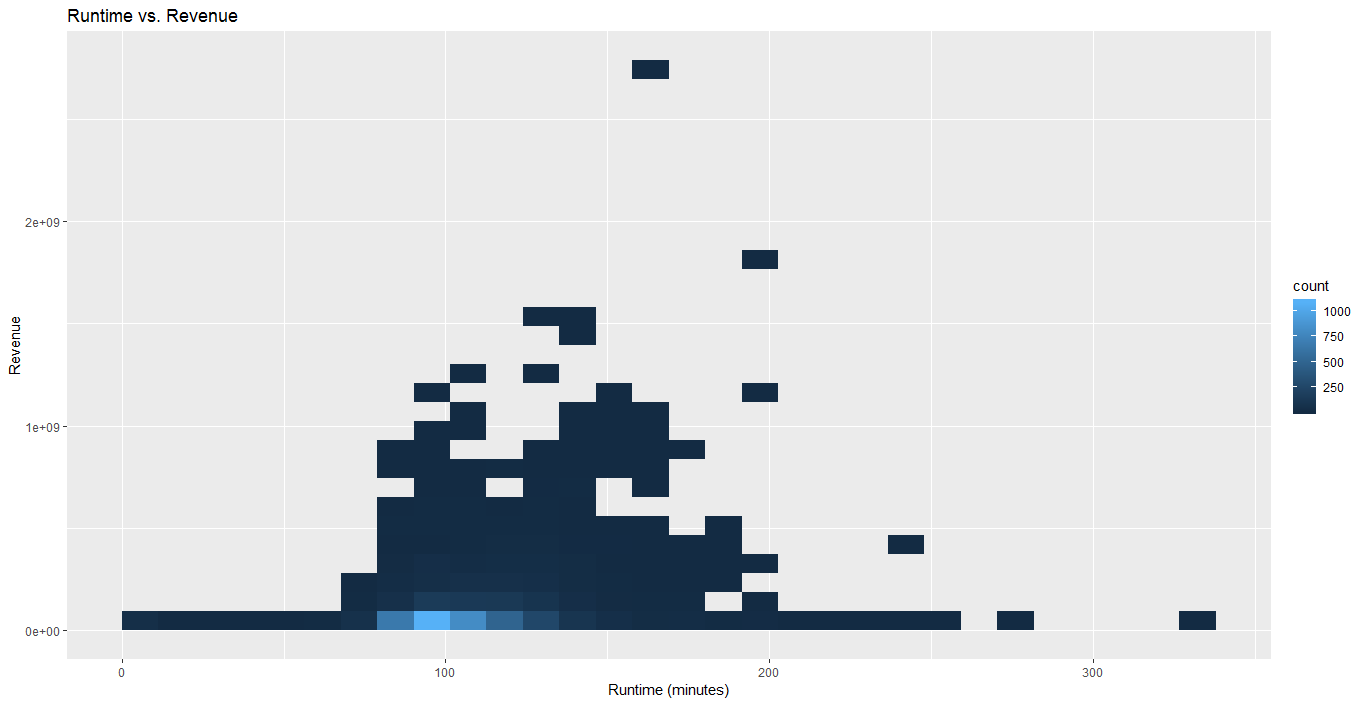
**a- plot**

****

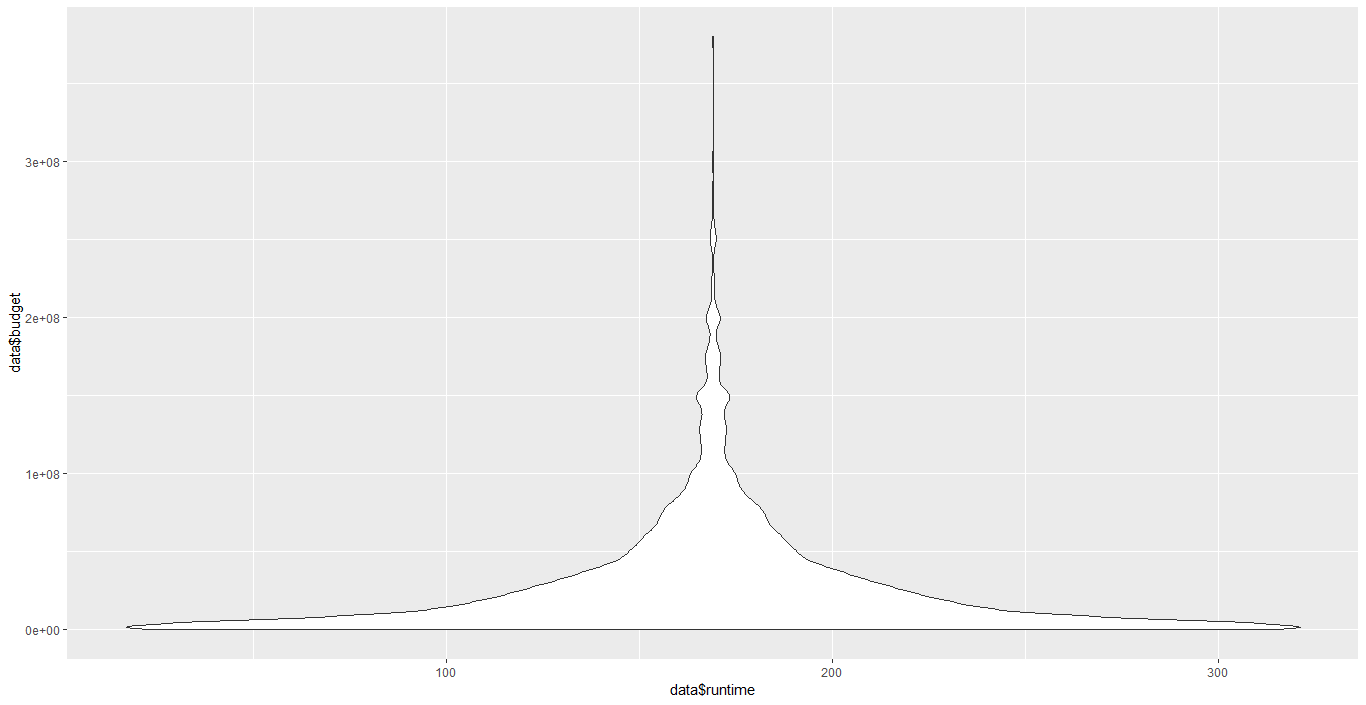


**b- ggplot**

****



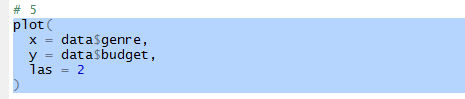
****

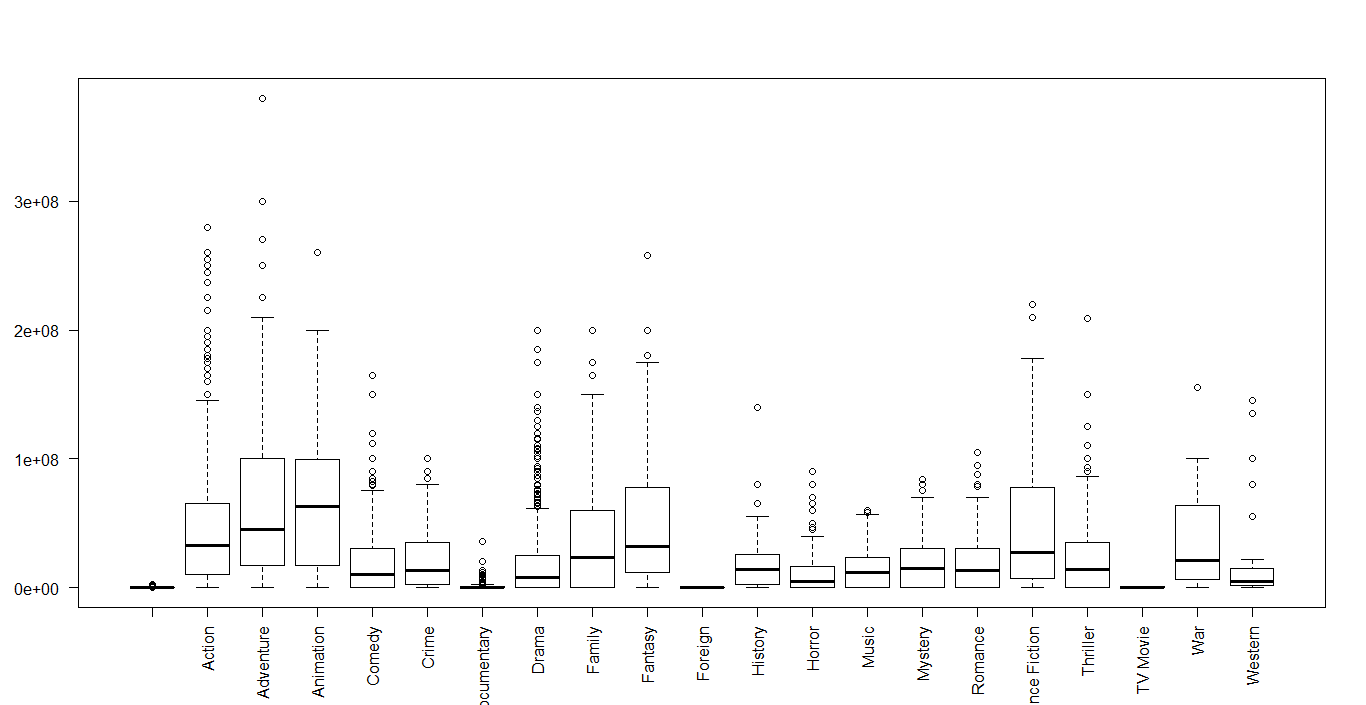


**Findings: Movies are most likely to generate more revenue if length is between 80 & 180 minutes.**

**5. Genre VS. Budget**

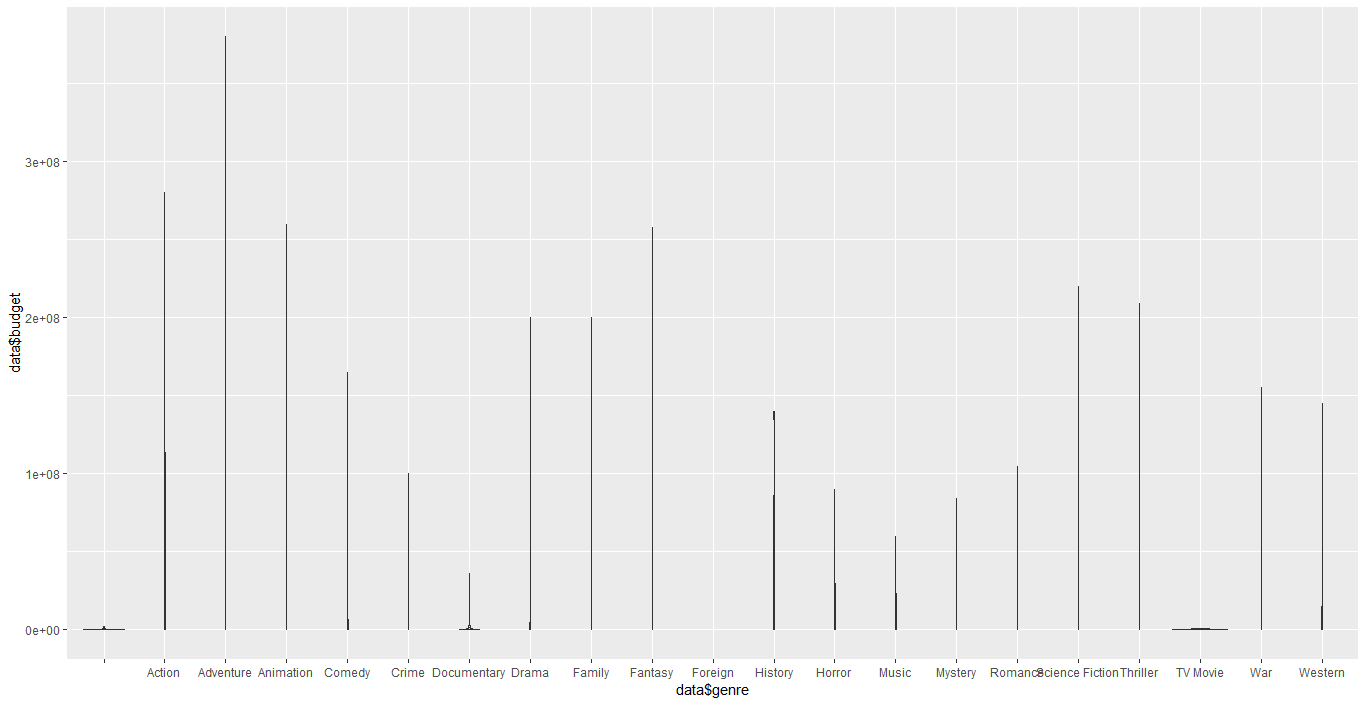
1. **Plot**

****



1. **Ggplot2**

****



**Findings: lowest budget are assigned to Music/Documentary movies. And highest budget are dedicated to Animation, Adventure, and science fiction movies.**