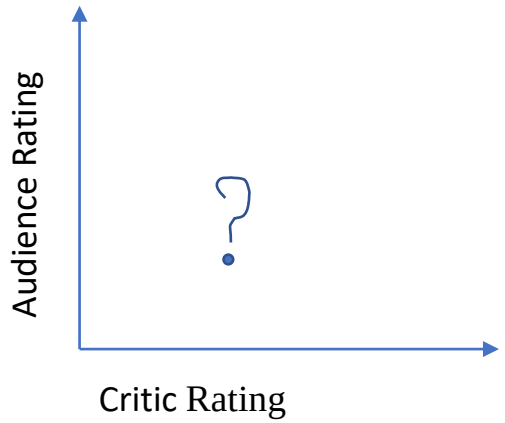
Problem Statement

- We have been approached by an analytics consultant from a movie review website. They are writing up an article analyzing movie rating by critics and audience as well as movie budgets for the years 2007 to 2011
- They have asked us to look at the data and provide them with 5 graphs. Given an excel sheet has information on movies, its genre, year in which they were released, Critic Rating(Rotten Tomatoes Rating) and Audience Rating.
- The CEO has requested a specific graph that has to be included with the other 5 graphs. The diagram is supposed to look like this –



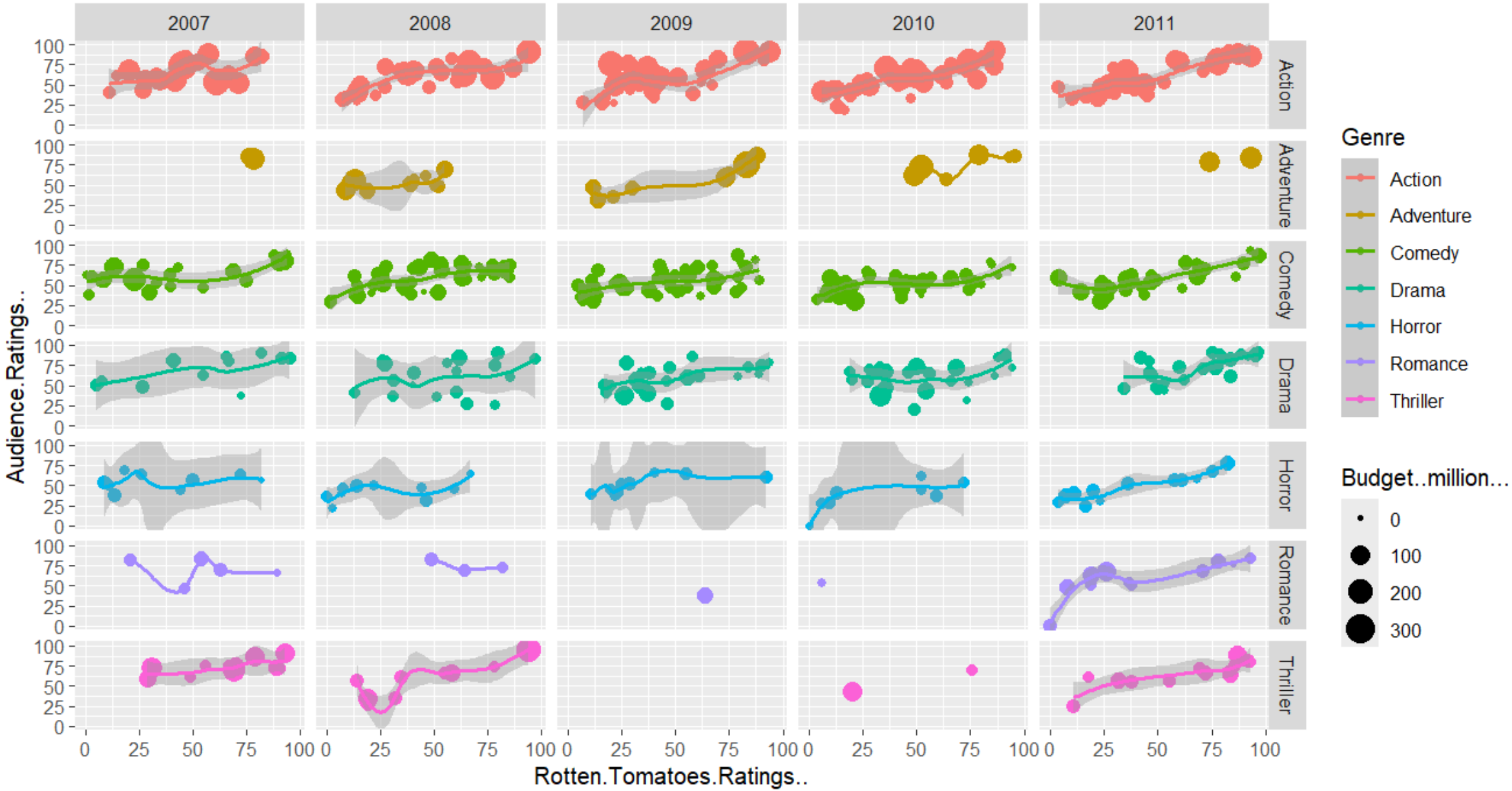
To visualize the data- R programming used

Graph 1 – scatter plot with faceting

```
x<-ggplot(data=movie,aes(x=Rotten.Tomatoes.Ratings..,y=Audience.Ratings..,colour=Genre))
x+geom_point(aes(size=Budget..million...))+facet_grid(Genre~Year)+geom_smooth()
```

Inference

- Over the years (2011),audience has developed liking in Romance and Thriller movies but in years 2009-2010 the audience did not like those genres
- Action, Drama and Comedy movies have been liked by the audience over the years

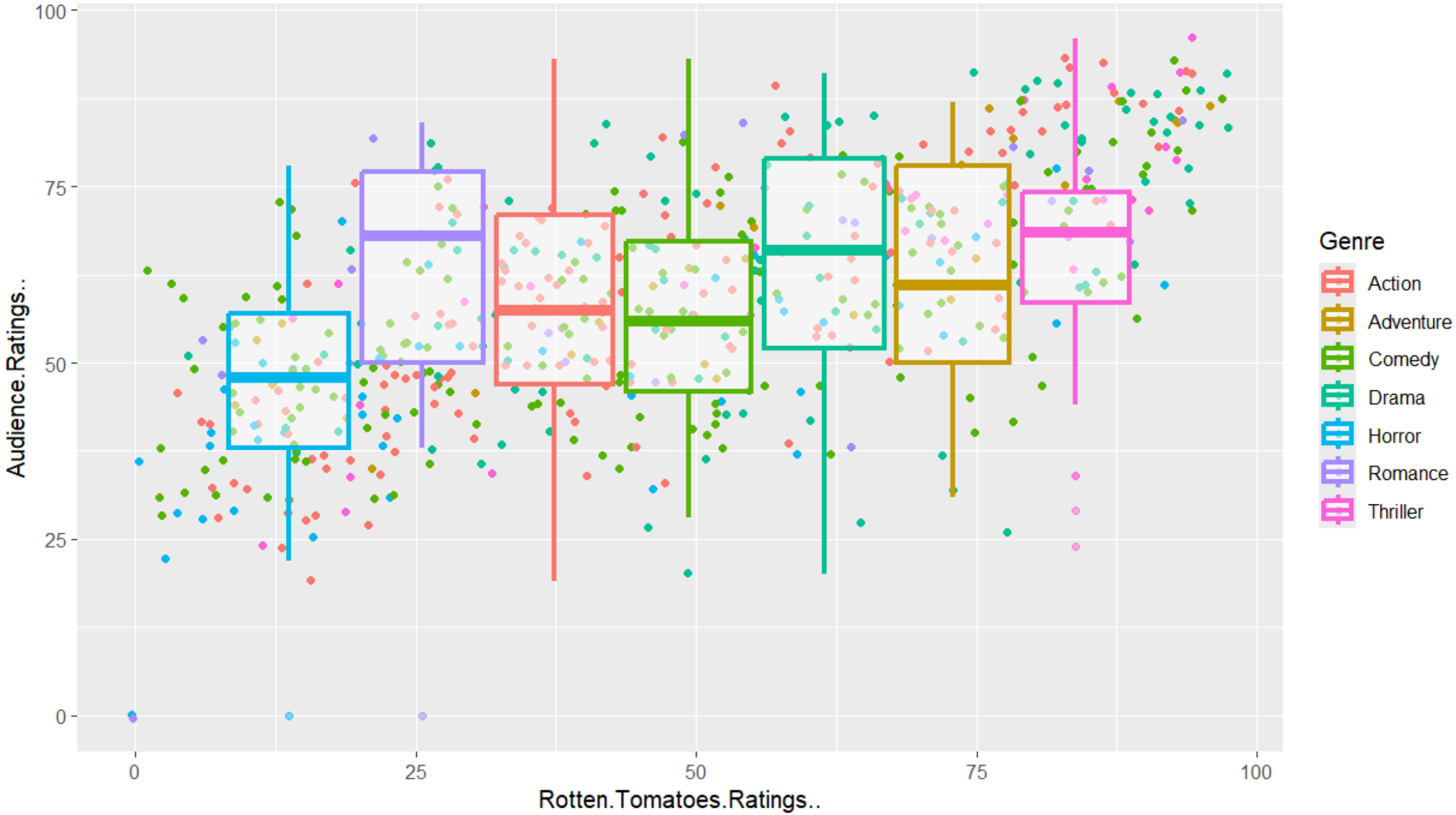

2007
2008
2009
2010
2011
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller
Audience.Ratings..
Rotten.Tomatoes.Ratings..
0
25
50
75
100
Genre
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller
Budget..million...
0
100
200
300

Graph 2 –box plot with scatter plot

```
v<-ggplot(data = movie,aes(x=Genre,y=Audience.Ratings..,colour=Genre))
v+geom_boxplot(size=1.2) #increases border size
v+geom_boxplot(size=1.2) + geom_point()
#tip- to make it look better
 geom_jitterv+geom_boxplot(size=1.2) +
 geom_jitter()
# to change the layering you change the order of layering
u+geom_point()+geom_boxplot(size=1.2)
#looks shabby
u+geom_jitter()+geom_boxplot(size=1.2,alpha=0.5) #aplha adds transparency
```

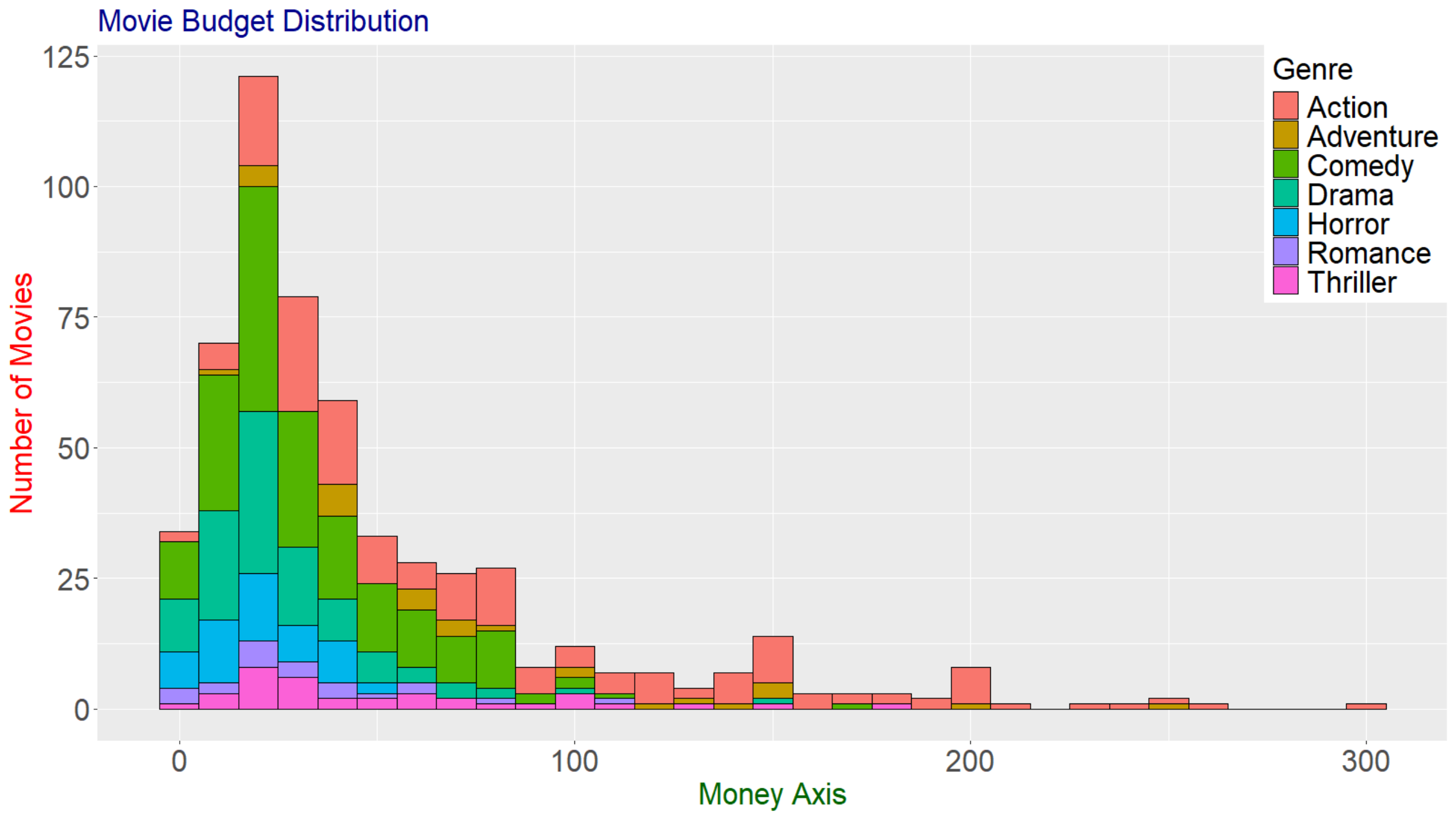
Inference

- Thriller, Drama and Adventure have high Rotten Tomatoes Ratings and is liked by the audience as well
- Romance has low critic rating but high audience ratings
- Horror has comparatively low audience rating and critic rating


100
75
50
25
0
Audience.Ratings..
0
25
50
75
100
Rotten.Tomatoes.Ratings..
Genre
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller

Graph 3:Histogram (Number of Movies VS Money Axis)

```
o<-ggplot(data=movie,aes(x=Budget..million...,))o+geom_histogram(binwidth=10,aes(fill=Genre))
i<-o+geom_histogram(binwidth=10,aes(fill=Genre),colour="Black")
i +xlab("Money Axis")+ylab("Number of Movies”)+
ggtitle("Movie Budget Distribution")+
 theme(axis.title.x = element_text(colour="Dark Green",size = 20),
axis.title.y = element_text(colour="Red",size=20),
axis.text.x = element_text(size=20),
axis.text.y=element_text(size=20),
legend.title=element_text(size=20),
legend.text=element_text(size=20),
legend.position=c(1,1),
legend.justification=c(1,1),
plot.title=element_text(colour="Dark Blue", size=20,family="Courier"))
```

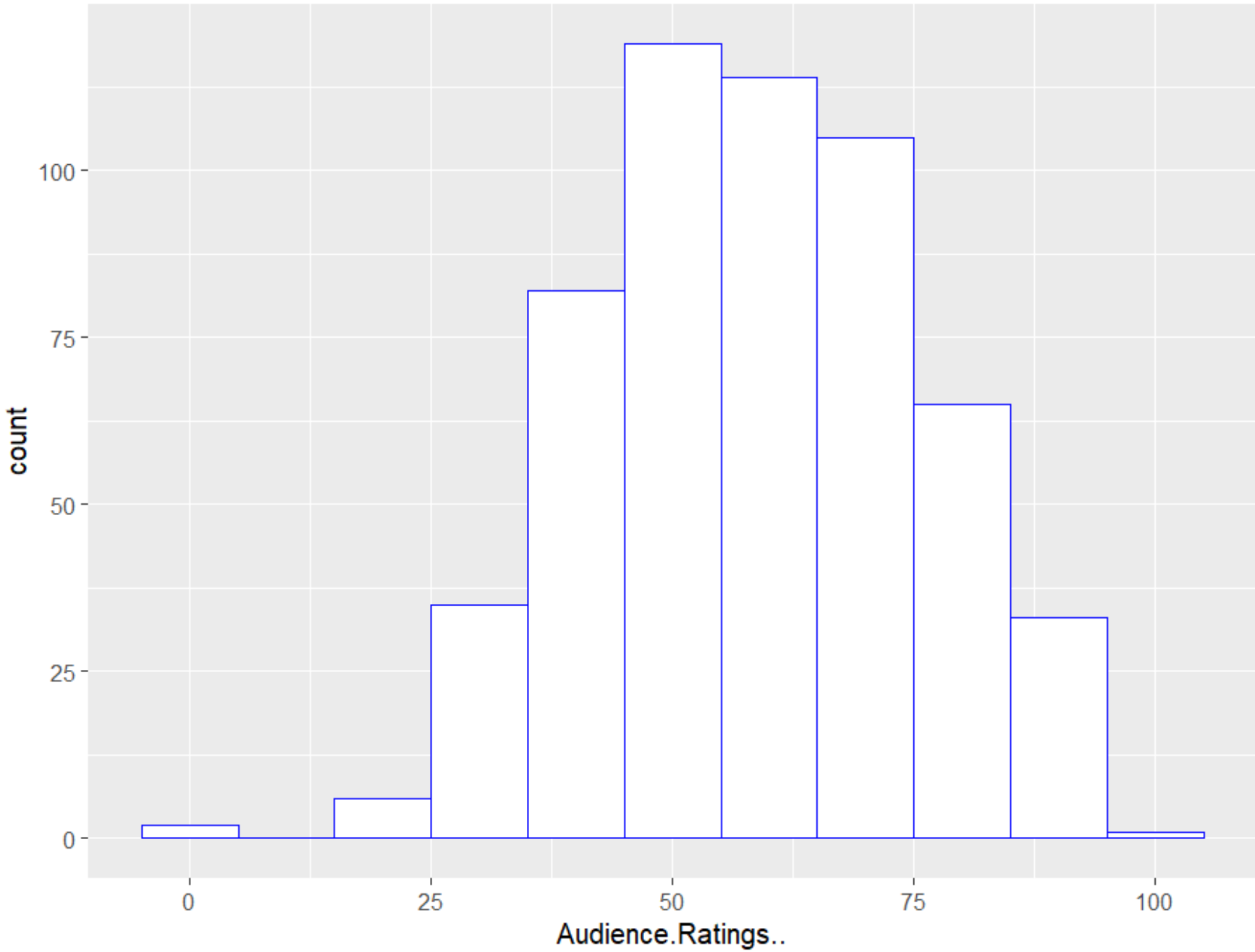

Movie Budget Distribution
Number of Movies
125
100
75
50
25
0
0
100
200
300
Money Axis
Genre
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller

Graph 4:Histogram- Audience Rating

- t<-ggplot(data=movie,aes(x=Audience.Ratings..))
- t+geom_histogram(binwidth=10,fill="White",colour="Blue")

Inference

- Maximum Rating- Over 100 movies have been rated by audience to be 50%
- Minimum Rating- Very less movies have 100% rating

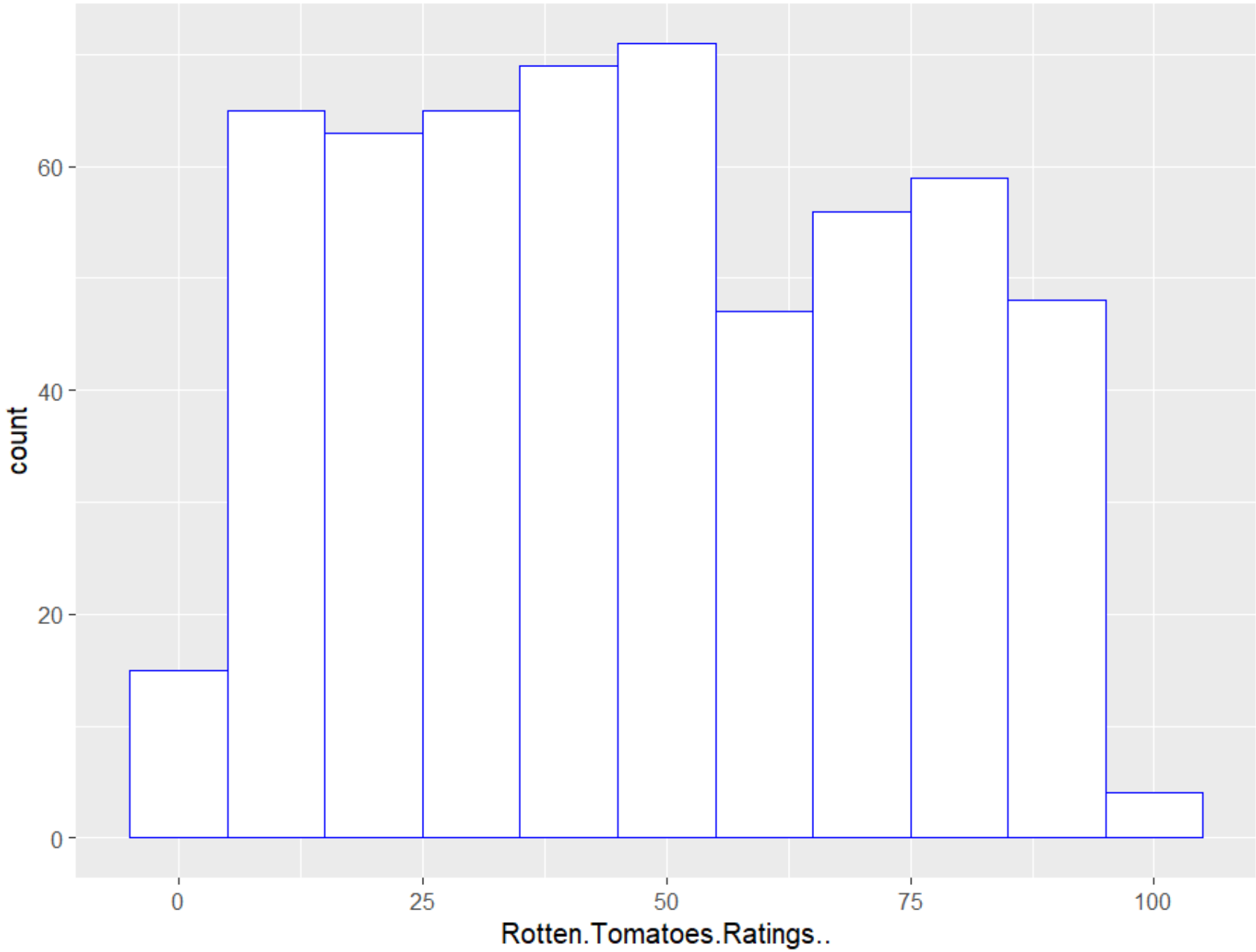

count
100
75
50
25
0
0
25
50
75
100
Audience.Ratings..

Graph 5:Histogram- Rotten tomatoes rating

- t<-ggplot(data=movie,aes(x=Rotten.Tomatoes.Ratings..))
- t+geom_histogram(binwidth=10,fill="White",colour="Blue")

Inference

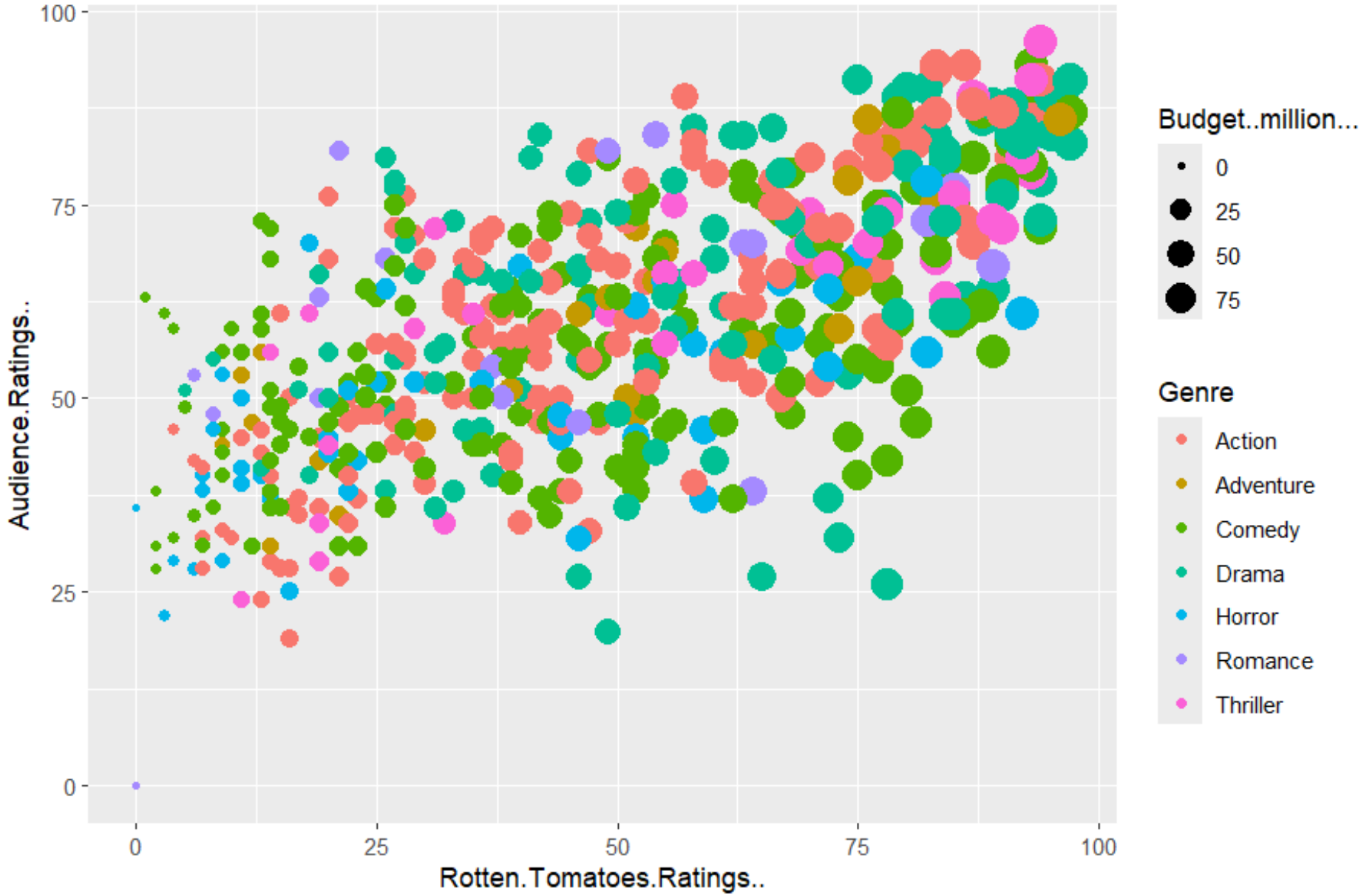
- Maximum Rating- Over 70 movies have 50% Rotten Tomatoes Rating
- Minimum Rating- About 5-10 movies have 100% Rating


count
60
40
20
0
0
25
50
75
100
Rotten.Tomatoes.Ratings..

Graph 6:Scatter plot with respect to Genre and Budget

```
q<_ggplot(data=movie,aes(x=Rotten.Tomatoes.Ratings..,y=Audience.Ratings..,colour=Genre,size=Budget..million...))

q+geom_point(aes(size=Rotten.Tomatoes.Ratings..))
```


Audience.Ratings..
100
75
50
25
0
0
25
50
75
100
Rotten.Tomatoes.Ratings..
Budget..million...
0
25
50
75
Genre
Action
Adventure
Comedy
Drama
Horror
Romance
Thriller