Movie Ratings

Abhinav Kumar

2025-09-19

```
getwd()
```

```
## [1] "/Users/abhinav/Downloads/essentials/projects/R prog/Movie Ratings and Revenue"
```

```
setwd("/Users/abhinav/Downloads/essentials/projects/R prog/Movie Ratings and Revenue")
getwd()
```

```
## [1] "/Users/abhinav/Downloads/essentials/projects/R prog/Movie Ratings and Revenue"
```

```
ratings1 <- read.csv("S6-Movie-Ratings.csv")
```

```
str(ratings1)
```

```
## 'data.frame':    562 obs. of  6 variables:
##  $ Film                     : chr  "(500) Days of Summer " "10,000 B.C." "12 Rounds " "127 Hours" ..
##  $ Genre                    : chr  "Comedy" "Adventure" "Action" "Adventure" ...
##  $ Rotten.Tomatoes.Ratings..: int  87 9 30 93 55 39 40 50 43 93 ...
##  $ Audience.Ratings..       : int  81 44 52 84 70 63 71 57 48 93 ...
##  $ Budget..million...       : int  8 105 20 18 20 200 30 32 28 8 ...
##  $ Year.of.release          : int  2009 2008 2009 2010 2009 2009 2008 2007 2011 2011 ...
```

```
ratings1$Genre <- factor(ratings1$Genre)
ratings1$Year.of.release <- factor(ratings1$Year.of.release)
ratings1$Film <- factor(ratings1$Film)
str(ratings1)
```

```
## 'data.frame':    562 obs. of  6 variables:
##  $ Film                     : Factor w/ 562 levels "(500) Days of Summer ",..: 1 2 3 4 5 6 7 8 9 10
##  $ Genre                    : Factor w/ 7 levels "Action","Adventure",..: 3 2 1 2 3 1 3 5 3 3 ...
##  $ Rotten.Tomatoes.Ratings..: int  87 9 30 93 55 39 40 50 43 93 ...
##  $ Audience.Ratings..       : int  81 44 52 84 70 63 71 57 48 93 ...
##  $ Budget..million...       : int  8 105 20 18 20 200 30 32 28 8 ...
##  $ Year.of.release          : Factor w/ 5 levels "2007","2008",..: 3 2 3 4 3 3 2 1 5 5 ...
```

```
head(ratings1)
```

```
##                    Film     Genre Rotten.Tomatoes.Ratings.. Audience.Ratings..
## 1 (500) Days of Summer     Comedy                        87                 81
## 2           10,000 B.C. Adventure                         9                 44
## 3            12 Rounds     Action                        30                 52
## 4            127 Hours Adventure                        93                 84
## 5            17 Again     Comedy                        55                 70
## 6                 2012    Action                        39                 63
##   Budget..million... Year.of.release
## 1                  8            2009
## 2                105            2008
```

```
## 3                  20            2009
## 4                  18            2010
## 5                  20            2009
## 6                 200            2009
```

```
colnames(ratings1) <- c("Movie", "Genre", "CriticRating", "AudienceRating", "BudgetMill", "Year")
head(ratings1)
```

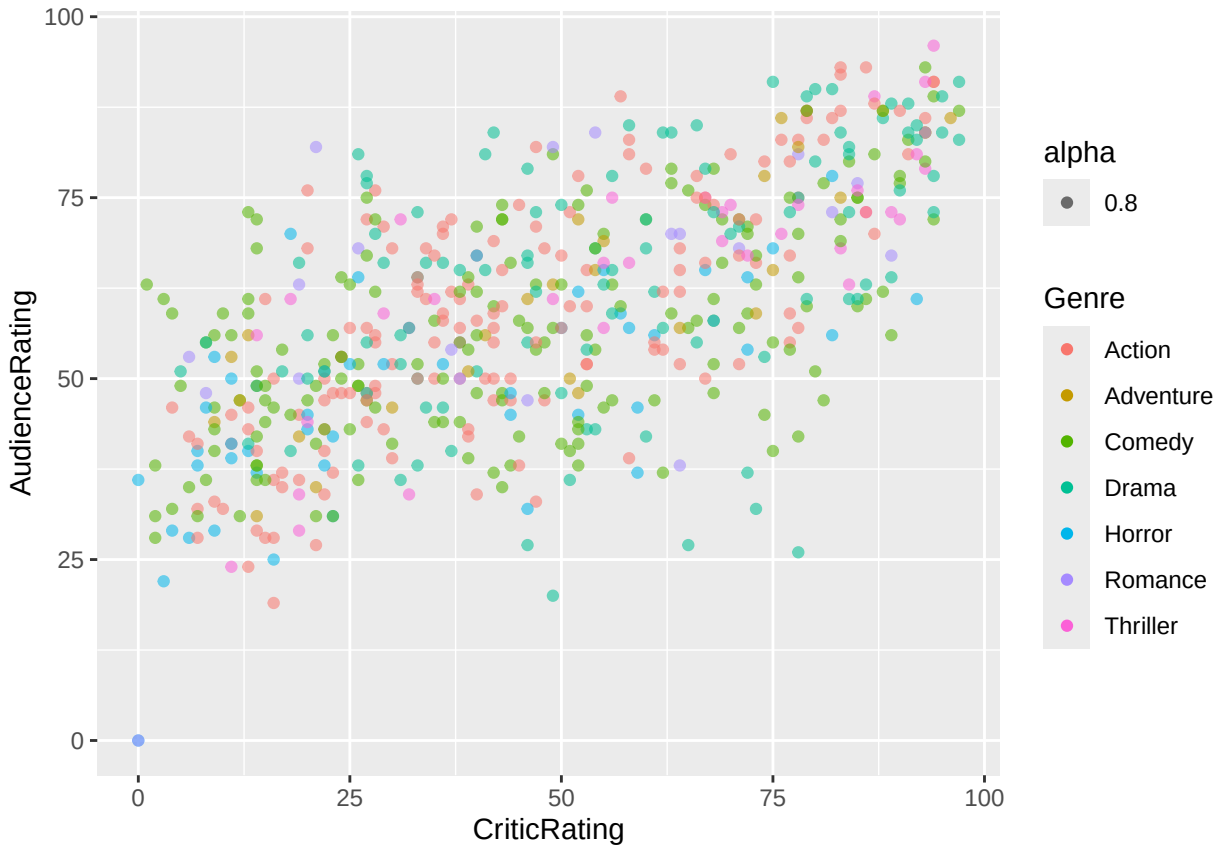
```
##                   Movie     Genre CriticRating AudienceRating BudgetMill Year
## 1 (500) Days of Summer     Comedy           87             81          8 2009
## 2           10,000 B.C. Adventure            9             44        105 2008
## 3            12 Rounds     Action           30             52         20 2009
## 4             127 Hours Adventure           93             84         18 2010
## 5             17 Again     Comedy           55             70         20 2009
## 6                 2012     Action           39             63        200 2009
```

```
summary(ratings1)
```

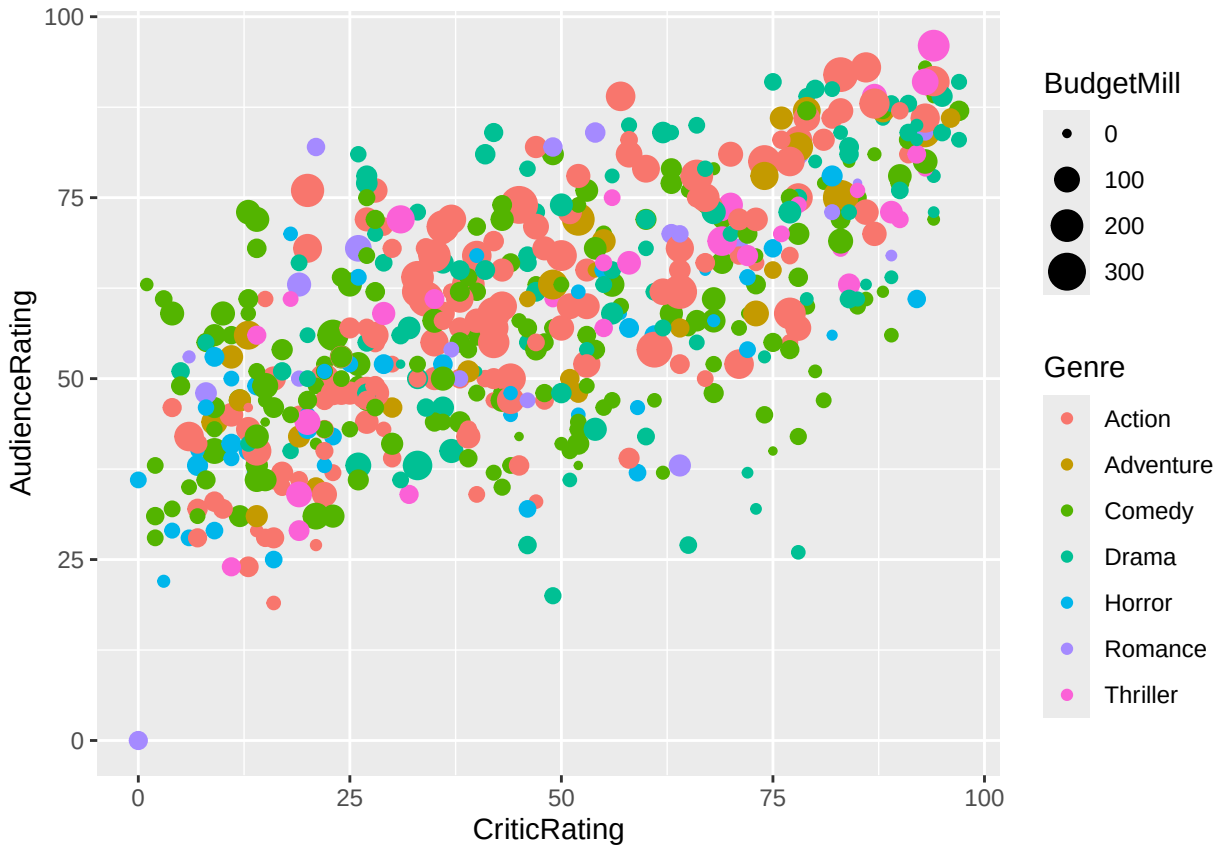
```
##                   Movie           Genre      CriticRating  AudienceRating
##  (500) Days of Summer :  1   Action   :154   Min.   : 0.0   Min.   : 0.00
##  10,000 B.C.          :  1   Adventure: 29   1st Qu.:25.0   1st Qu.:47.00
##  12 Rounds            :  1   Comedy   :172   Median :46.0   Median :58.00
##  127 Hours            :  1   Drama    :101   Mean   :47.4   Mean   :58.83
##  17 Again             :  1   Horror   : 49   3rd Qu.:70.0   3rd Qu.:72.00
##  2012                 :  1   Romance  : 21   Max.   :97.0   Max.   :96.00
##  (Other)              :556   Thriller : 36
##    BudgetMill      Year
##  Min.   :  0.0   2007: 79
##  1st Qu.: 20.0   2008:125
##  Median : 35.0   2009:116
##  Mean   : 50.1   2010:119
##  3rd Qu.: 65.0   2011:123
##  Max.   :300.0
##
```

```
library(ggplot2)
```

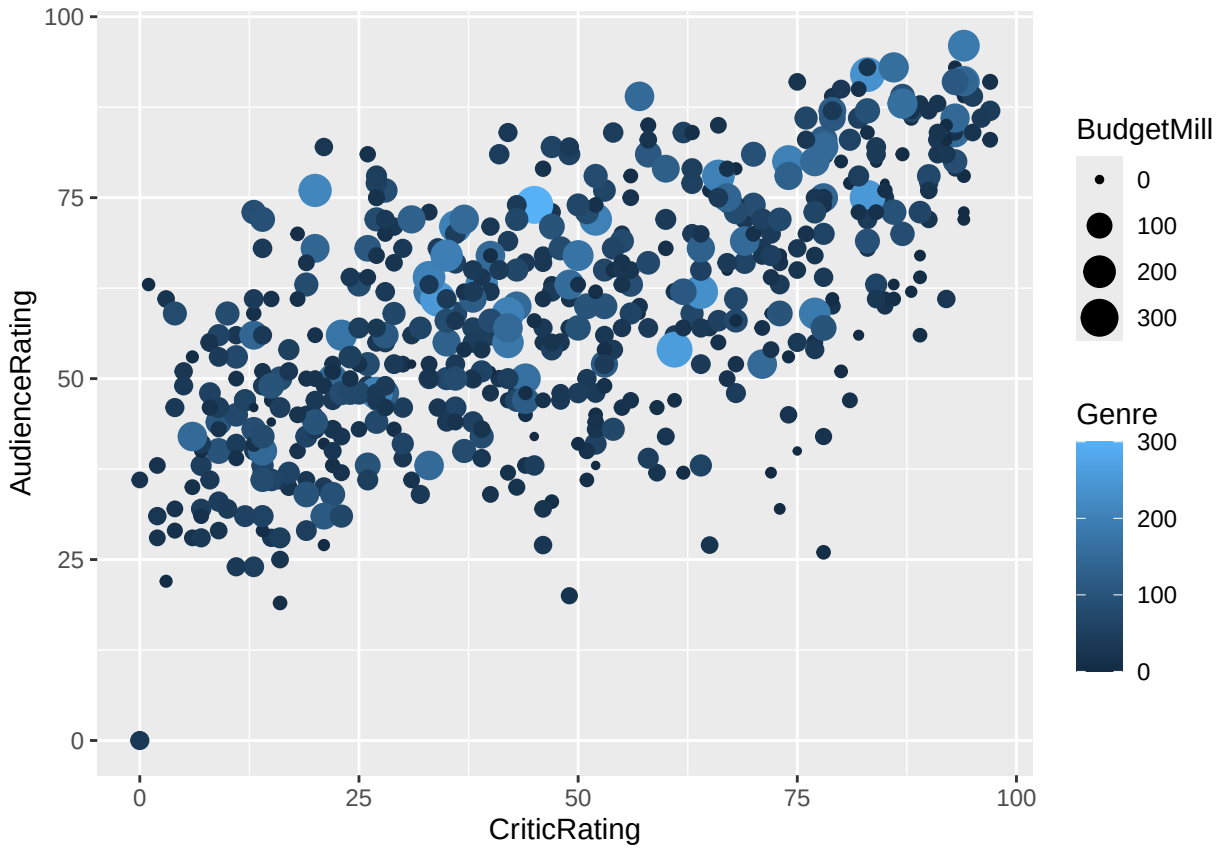
```
ggplot(data = ratings1, aes(x = CriticRating, y = AudienceRating,
                            colour = Genre, alpha = 0.8)) +
  geom_point()
```



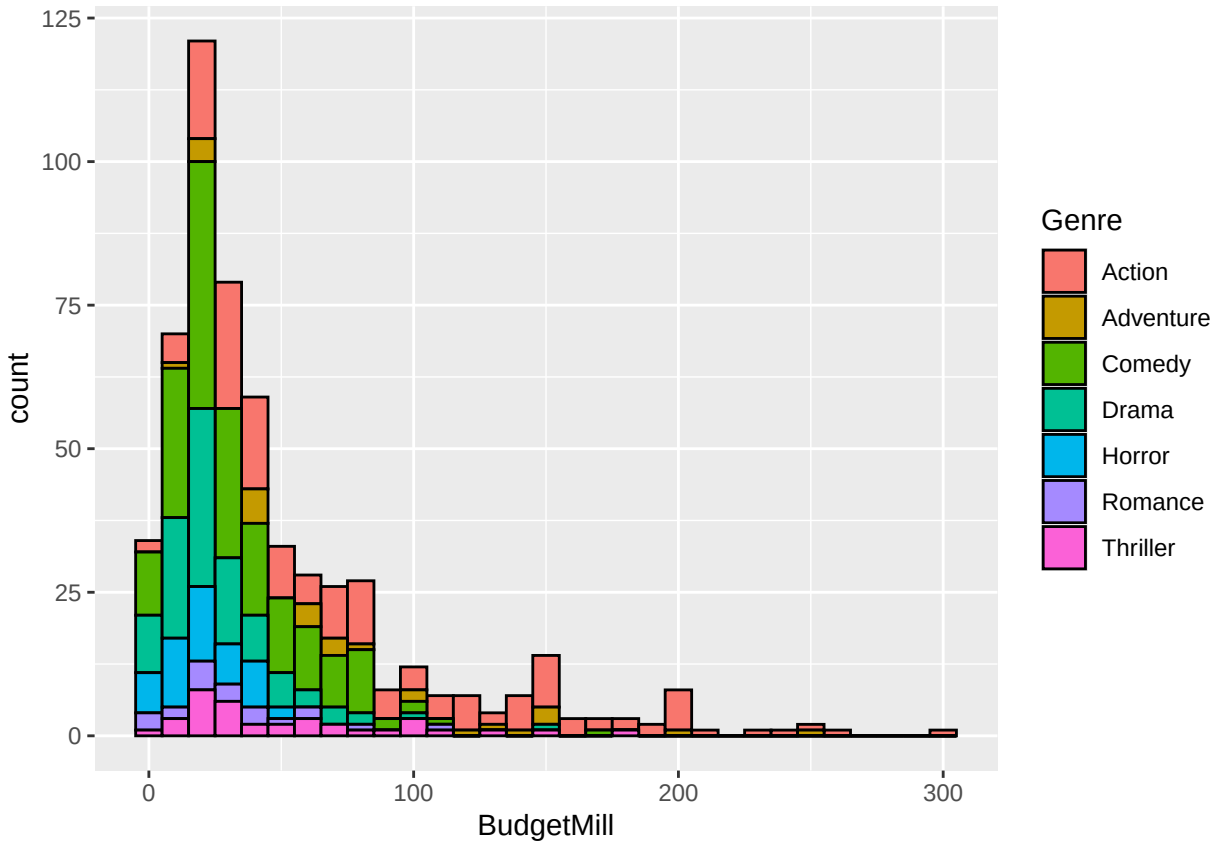

```
(p <- ggplot(data = ratings1, aes(x = CriticRating, y = AudienceRating,
                                  colour = Genre, size = BudgetMill))) +
  geom_point()
```



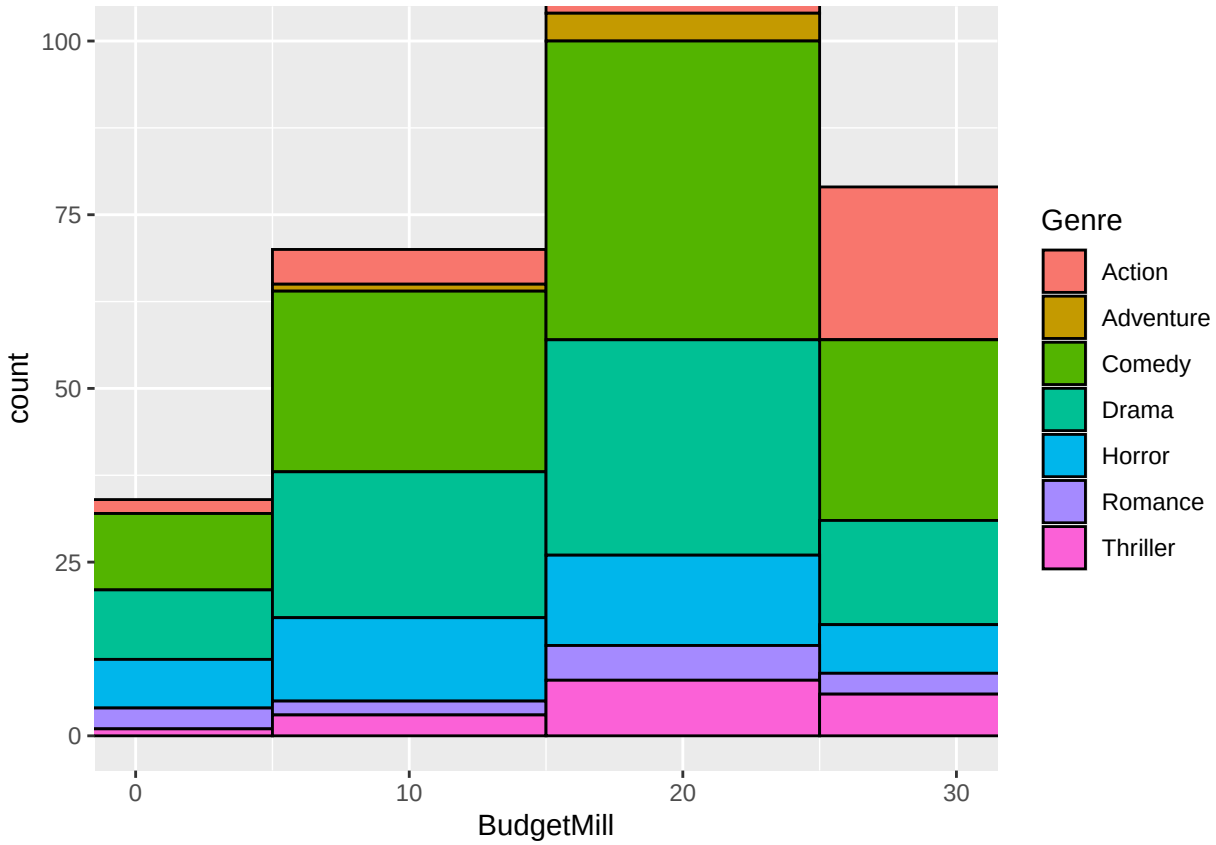

```
p + geom_point(aes(colour = BudgetMill))
```



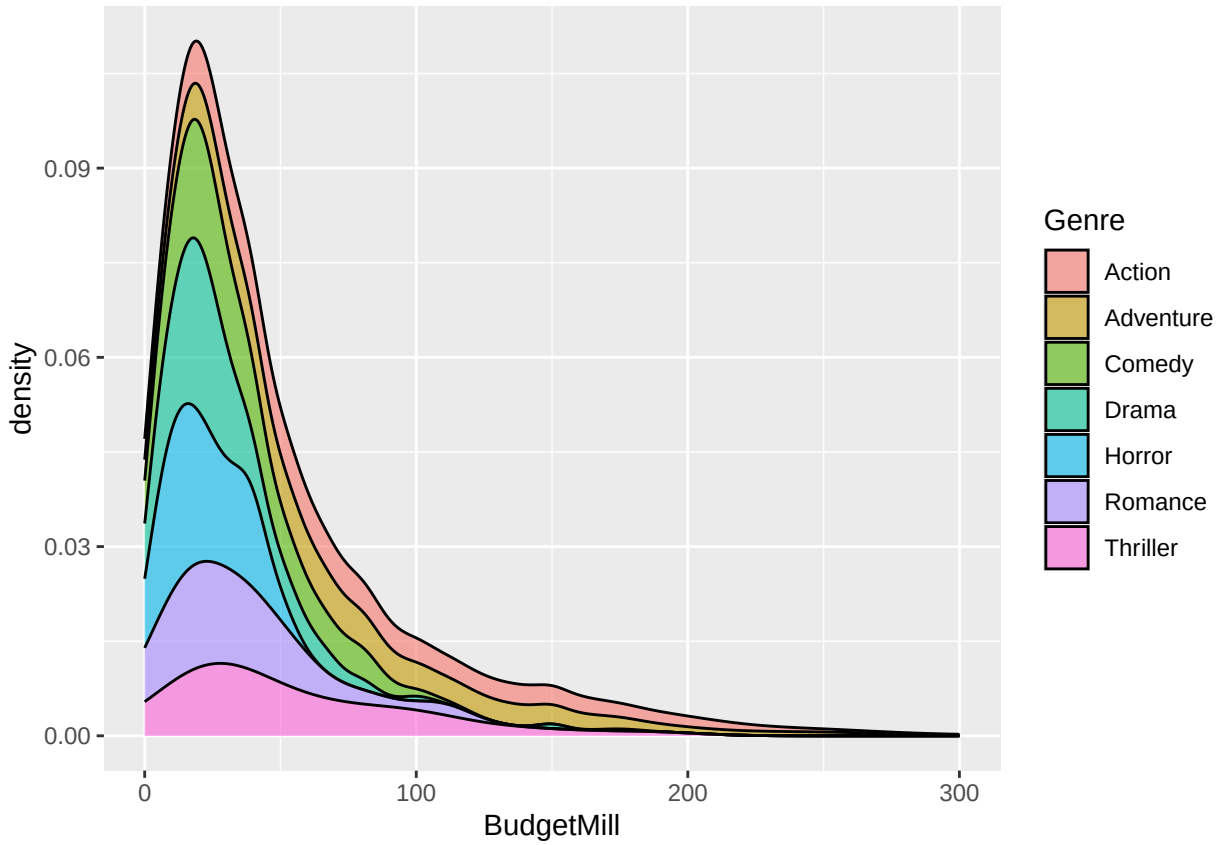

```
q <- ggplot(data = ratings1, aes(x = BudgetMill))
q + geom_histogram(binwidth = 10, aes(fill = Genre), colour = "black")
```



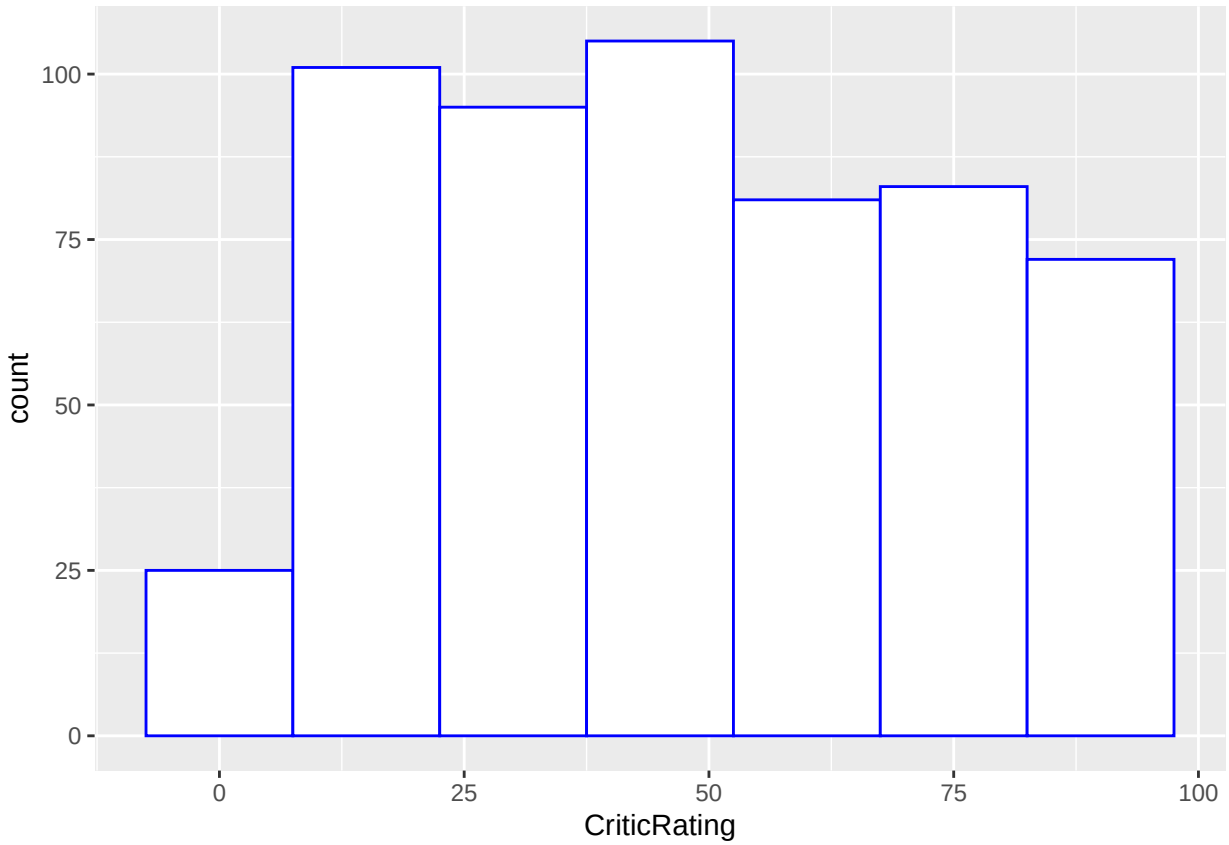

```
q + geom_histogram(binwidth = 10, aes(fill = Genre), colour = "black") + coord_cartesian(ylim = c(0, 10
```



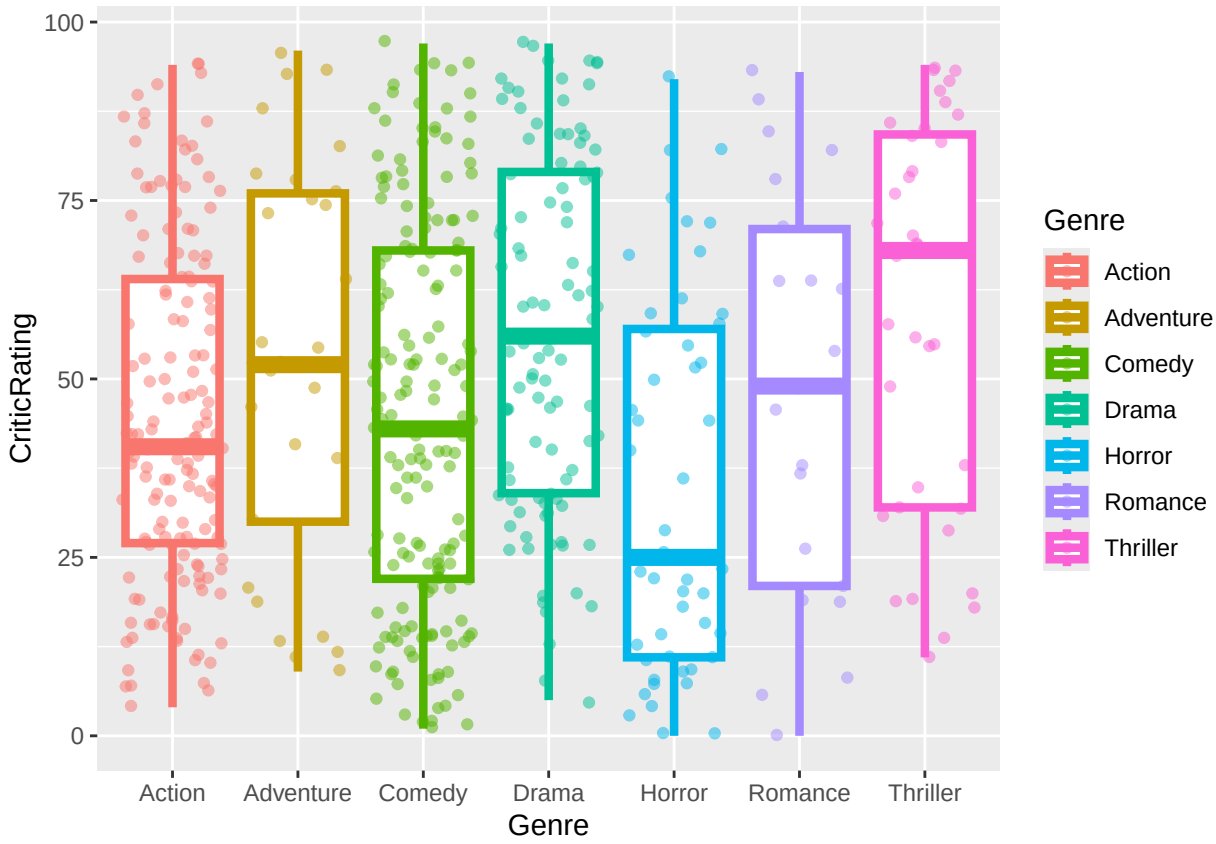

```
q + geom_density(aes(fill = Genre), colour = "black", position = "stack", alpha = 0.6)
```



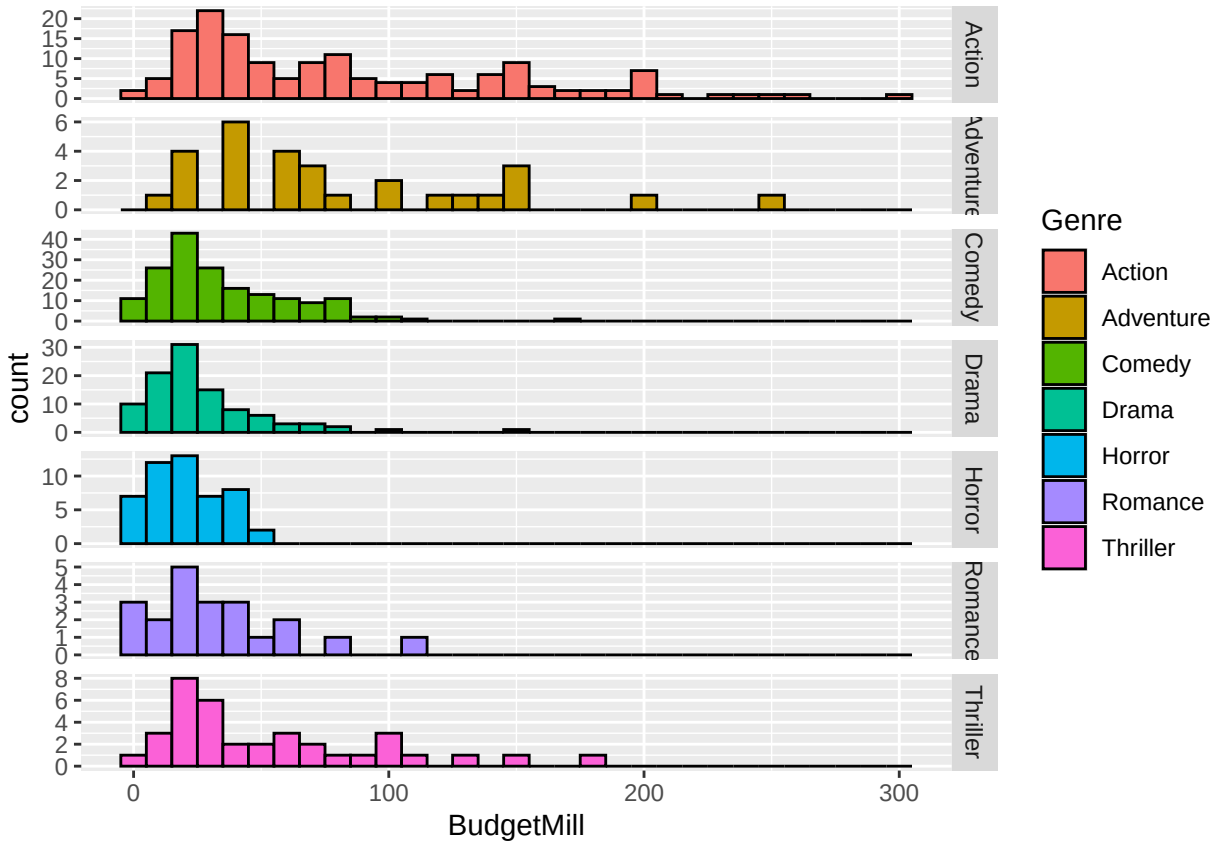

```
r <- ggplot(data = ratings1)
r + geom_histogram(binwidth = 15, aes(x = CriticRating), fill = "white", colour = "blue")
```




```
s <- ggplot(data = ratings1, aes(x = Genre, y = CriticRating, colour = Genre))
s + geom_boxplot(size = 1.5)+ geom_jitter(alpha = 0.5)
```

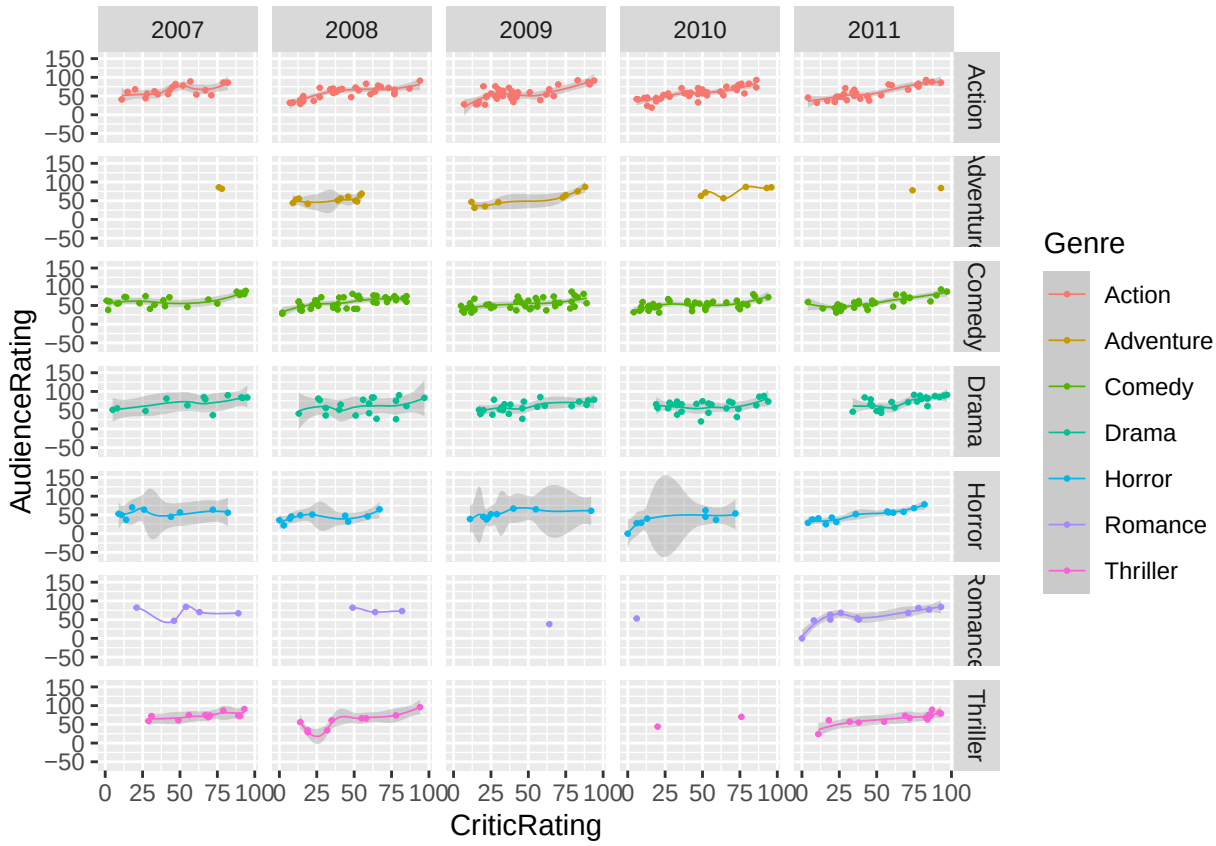



```
t <- ggplot(data = ratings1, aes(x = BudgetMill))
t + geom_histogram(binwidth = 10, aes(fill = Genre), colour = "black") + facet_grid(Genre~., scales = "
```




```
u <- ggplot(data = ratings1, aes(x = CriticRating, y = AudienceRating, colour = Genre))
u + geom_smooth(size = 0.3) + geom_point(size = 0.5) + facet_grid(Genre~Year)
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ratings2 <- read.csv("S6-Homework-Data.csv")
```

```
head(ratings2)
```

```
##   Day.of.Week               Director  Genre       Movie.Title Release.Date
## 1      Friday              Brad Bird action      Tomorrowland   22/05/2015
## 2      Friday            Scott Waugh action    Need for Speed   14/03/2014
## 3      Friday         Patrick Hughes action The Expendables 3   15/08/2014
## 4      Friday Phil Lord, Chris Miller comedy    21 Jump Street   16/03/2012
## 5      Friday        Roland Emmerich action  White House Down   28/06/2013
## 6      Friday             David Ayer action               Fury   17/10/2014
##                Studio Adjusted.Gross...mill. Budget...mill. Gross...mill.
## 1 Buena Vista Studios                202.1            170         202.1
## 2 Buena Vista Studios                204.2             66         203.3
## 3           Lionsgate                207.1            100         206.2
## 4                Sony                208.8             42         201.6
## 5                Sony                209.7            150         205.4
## 6                Sony                212.8             80         211.8
##   IMDb.Rating MovieLens.Rating Overseas...mill. Overseas. Profit...mill.
## 1         6.7             3.26            111.9      55.4           32.1
## 2         6.6             2.97            159.7      78.6          137.3
## 3         6.1             2.93            166.9      80.9          106.2
## 4         7.2             3.62             63.1      31.3          159.6
## 5         8.0             3.65            132.3      64.4           55.4
## 6         5.8             2.85              126      59.5          131.8
##   Profit. Runtime..min. US...mill. Gross...US
## 1    18.9           130       90.2       44.6
```

```
## 2   208.0           132       43.6       21.4
## 3   106.2           126       39.3       19.1
## 4   380.0           109      138.4       68.7
## 5    36.9           131       73.1       35.6
## 6   164.8           134       85.8       40.5
```

```
colnames(ratings2) <- c("Day", "Director", "Genre", "Movie", "LaunchDate", "Studio", "AdjGrossMill", "Bu
```

```
str(ratings2)
```

```
## 'data.frame':    608 obs. of  18 variables:
##  $ Day         : chr  "Friday" "Friday" "Friday" "Friday" ...
##  $ Director    : chr  "Brad Bird" "Scott Waugh" "Patrick Hughes" "Phil Lord, Chris Miller" ...
##  $ Genre       : chr  "action" "action" "action" "comedy" ...
##  $ Movie       : chr  "Tomorrowland" "Need for Speed" "The Expendables 3" "21 Jump Street" ...
##  $ LaunchDate  : chr  "22/05/2015" "14/03/2014" "15/08/2014" "16/03/2012" ...
##  $ Studio      : chr  "Buena Vista Studios" "Buena Vista Studios" "Lionsgate" "Sony" ...
##  $ AdjGrossMill: chr  "202.1" "204.2" "207.1" "208.8" ...
##  $ BudgetMill  : num  170 66 100 42 150 80 50 85 70 5 ...
##  $ GrossMill   : chr  "202.1" "203.3" "206.2" "201.6" ...
##  $ IMDBRating  : num  6.7 6.6 6.1 7.2 8 5.8 6 6.8 6.3 5.9 ...
##  $ LensRating  : num  3.26 2.97 2.93 3.62 3.65 2.85 3.16 3.45 2.92 2.9 ...
##  $ OverMill    : chr  "111.9" "159.7" "166.9" "63.1" ...
##  $ Over%       : num  55.4 78.6 80.9 31.3 64.4 59.5 39.9 39.3 73.9 49.8 ...
##  $ ProfMill    : chr  "32.1" "137.3" "106.2" "159.6" ...
##  $ OverProf    : num  18.9 208 106.2 380 36.9 ...
##  $ Runtime     : int  130 132 126 109 131 134 125 115 92 84 ...
##  $ USMill      : num  90.2 43.6 39.3 138.4 73.1 ...
##  $ USGross     : num  44.6 21.4 19.1 68.7 35.6 40.5 60.1 60.7 26.1 50.2 ...
```

```
ratings2$Day <- factor(ratings2$Day)
ratings2$Genre <- factor(ratings2$Genre)
ratings2$Director <- factor(ratings2$Director)
ratings2$Studio <- factor(ratings2$Studio)
ratings2$AdjGrossMill <- as.numeric(ratings2$AdjGrossMill)
ratings2$GrossMill <- as.numeric(ratings2$GrossMill)
ratings2$OverMill <- as.numeric(ratings2$OverMill)
ratings2$ProfMill <- as.numeric(ratings2$ProfMill)
ratings2$Runtime <- as.numeric(ratings2$Runtime)
str(ratings2)
```

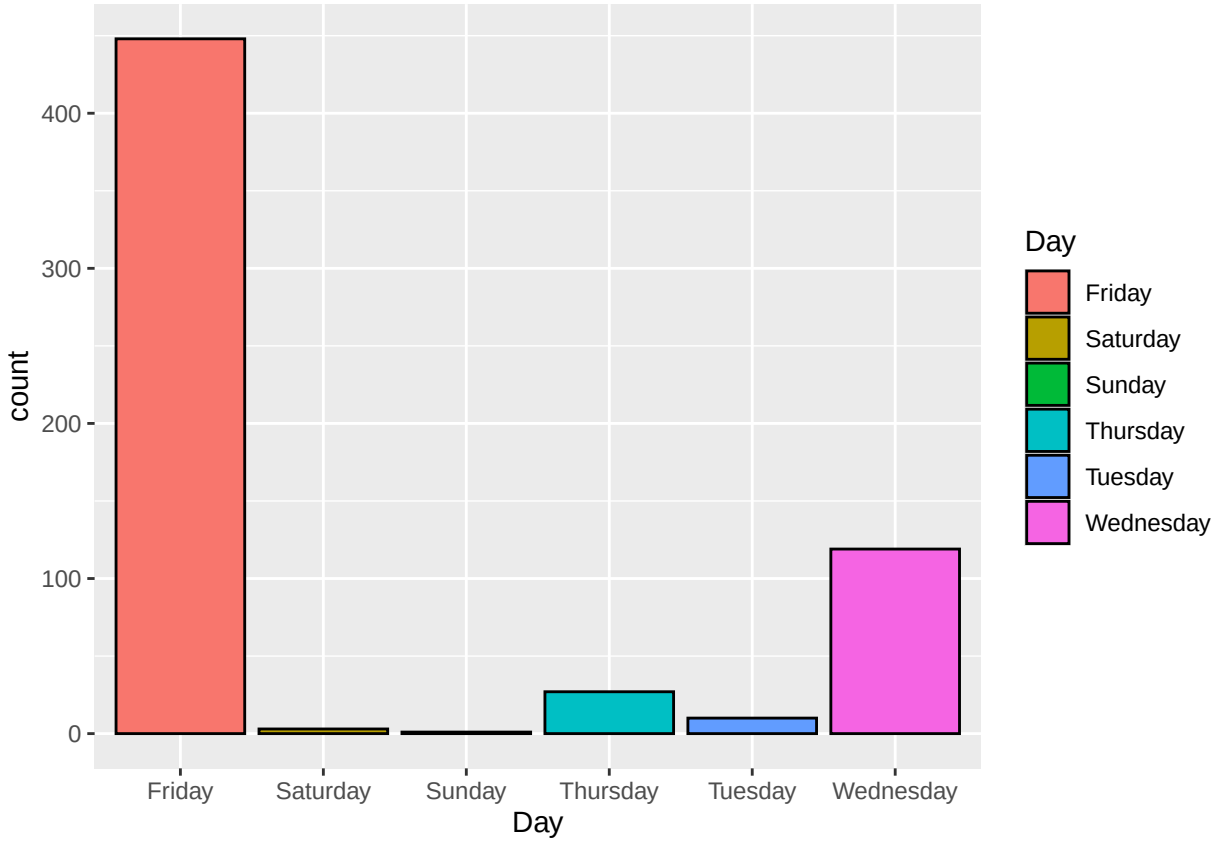
```
## 'data.frame':    608 obs. of  18 variables:
##  $ Day         : Factor w/ 6 levels "Friday","Saturday",..: 1 1 1 1 1 1 4 1 1 1 ...
##  $ Director    : Factor w/ 337 levels "Aaron Blaise, Robert A. Walker",..: 31 297 233 256 287 76 276
##  $ Genre       : Factor w/ 15 levels "action","adventure",..: 1 1 1 5 1 1 2 1 1 10 ...
##  $ Movie       : chr  "Tomorrowland" "Need for Speed" "The Expendables 3" "21 Jump Street" ...
##  $ LaunchDate  : chr  "22/05/2015" "14/03/2014" "15/08/2014" "16/03/2012" ...
##  $ Studio      : Factor w/ 36 levels "Art House Studios",..: 2 2 11 25 25 25 2 31 31 20 ...
##  $ AdjGrossMill: num  202 204 207 209 210 ...
##  $ BudgetMill  : num  170 66 100 42 150 80 50 85 70 5 ...
##  $ GrossMill   : num  202 203 206 202 205 ...
##  $ IMDBRating  : num  6.7 6.6 6.1 7.2 8 5.8 6 6.8 6.3 5.9 ...
##  $ LensRating  : num  3.26 2.97 2.93 3.62 3.65 2.85 3.16 3.45 2.92 2.9 ...
##  $ OverMill    : num  111.9 159.7 166.9 63.1 132.3 ...
##  $ Over%       : num  55.4 78.6 80.9 31.3 64.4 59.5 39.9 39.3 73.9 49.8 ...
##  $ ProfMill    : num  32.1 137.3 106.2 159.6 55.4 ...
```

```
##  $ OverProf    : num  18.9 208 106.2 380 36.9 ...
##  $ Runtime     : num  130 132 126 109 131 134 125 115 92 84 ...
##  $ USMill      : num  90.2 43.6 39.3 138.4 73.1 ...
##  $ USGross     : num  44.6 21.4 19.1 68.7 35.6 40.5 60.1 60.7 26.1 50.2 ...
```

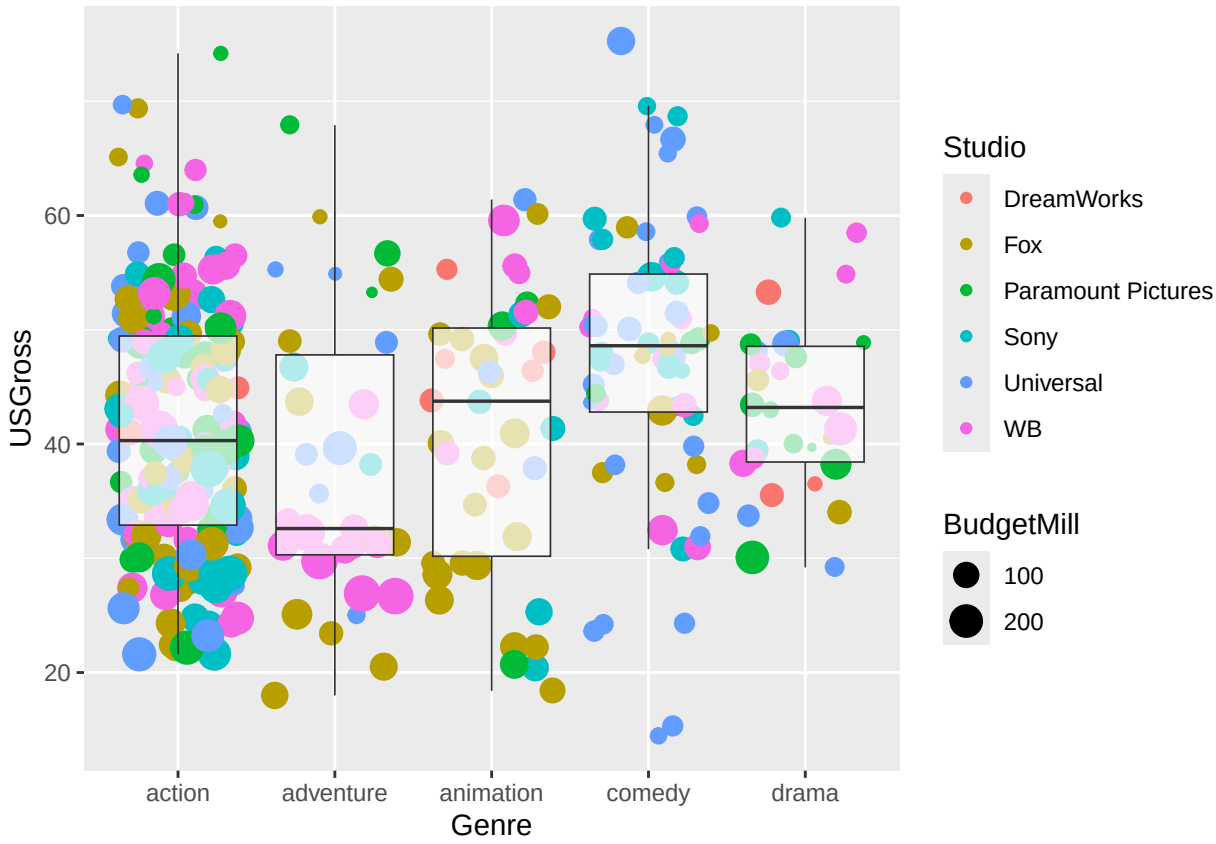
```
filt1 <- ratings2$Genre %in% c("action", "adventure", "animation", "comedy", "drama")
filt2 <- ratings2$Studio %in% c("WB", "Fox", "Paramount Pictures", "Sony", "Universal", "DreamWorks")

ratings3 <- ratings2[(filt1 & filt2),]
```

```
a <- ggplot(data = ratings2, aes(x = Day))
a + geom_bar(aes(fill = Day), colour = "black")
```




```
b <- ggplot(data = ratings3, aes(x = Genre, y = USGross))
b + geom_jitter(aes(size =BudgetMill, colour = Studio)) + geom_boxplot(size = 0.3, alpha = 0.7, outlier
```


USGross
60
40
20
action
adventure
animation
comedy
drama
Genre
Studio
DreamWorks
Fox
Paramount Pictures
Sony
Universal
WB
BudgetMill
100
200