Data Visualization Using R:(Project Report)

Team members- 1. Aritra Raut

2.Hiranmoy Roy

3.Sourjya Chatterjee

Dataset: https://www.kaggle.com/lava18/google-play-store-apps

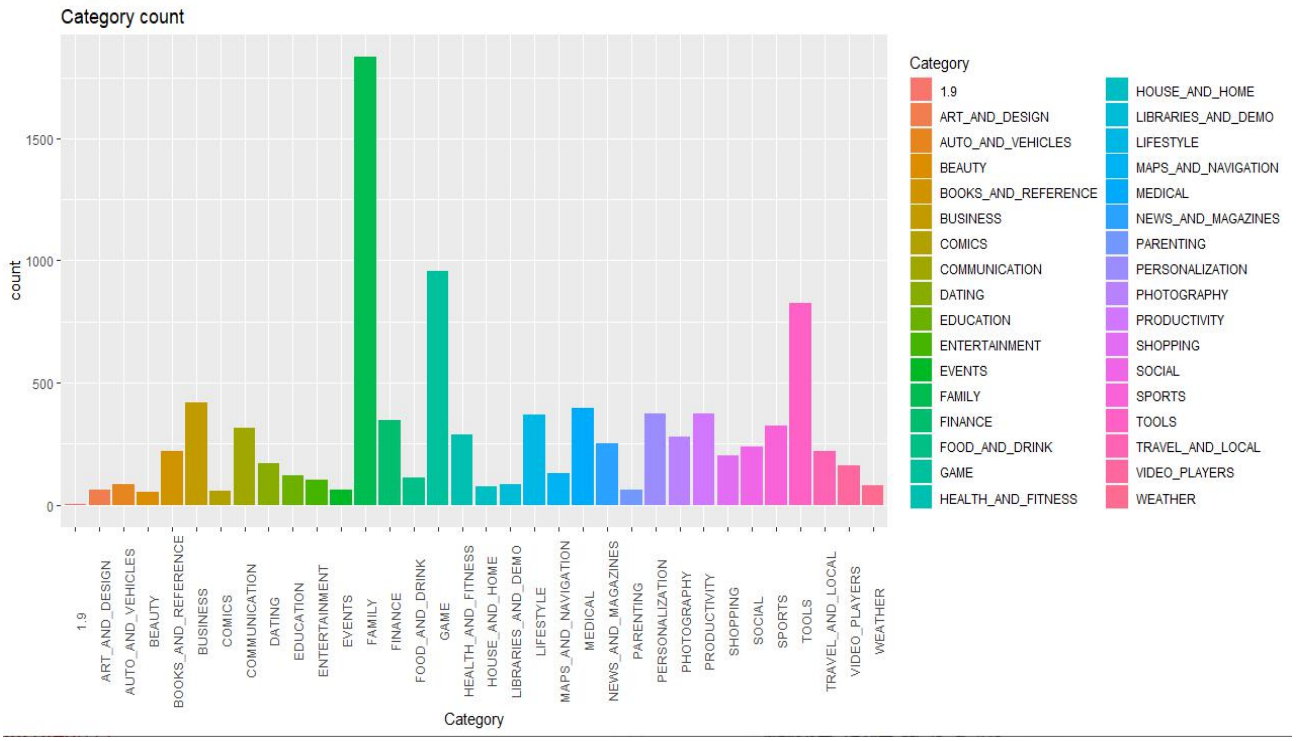
Dataset Description: Set of Playstore Apps with the details of these apps.

Main objective: finding the apps which have the most passionate developers

Finding the best category →

q.1>*which category has the most no. of applications?*

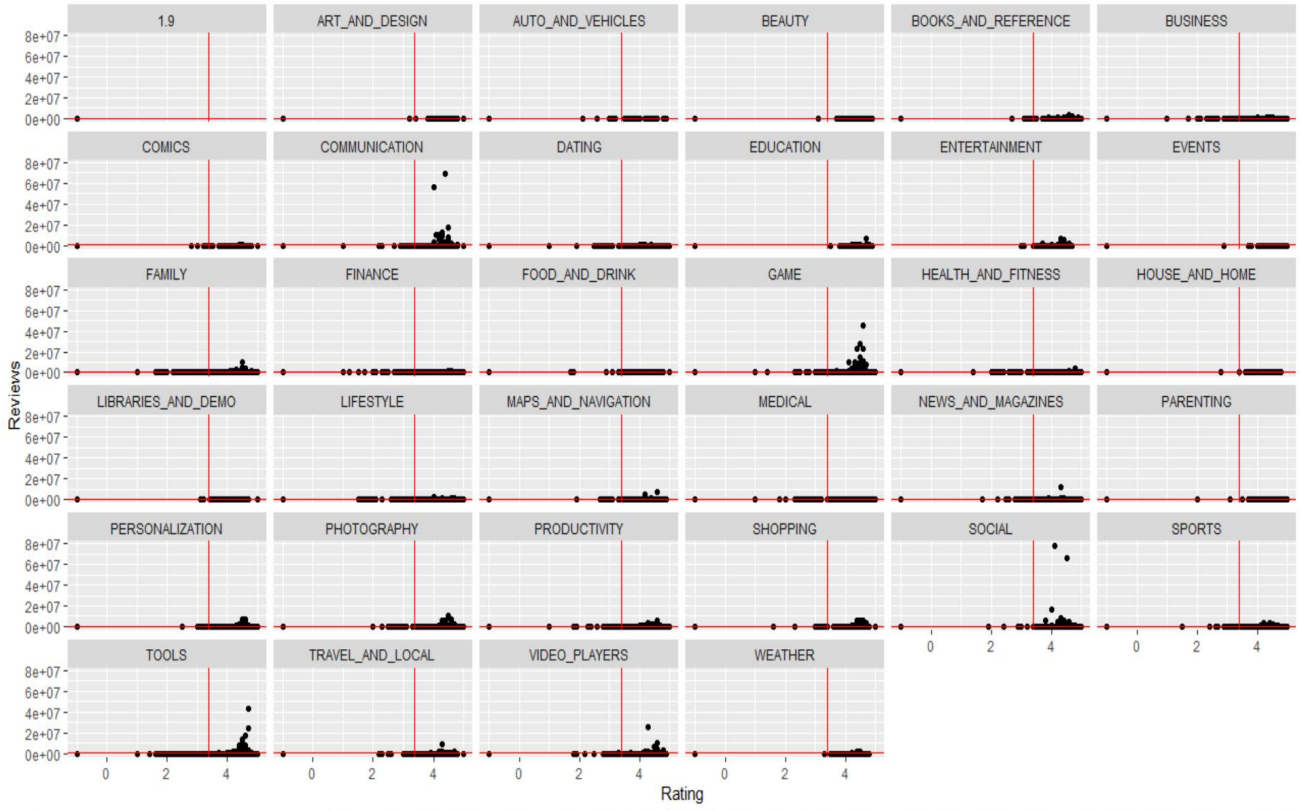
plot—>



Plot Analysis ->We can clearly see that the category “family” has the most no. of applications here,followed by “games”,”tools” and others. But we cannot say that this categories with higher no. of apps under them are the most popular categories.

q.2> *what is the relation between* ***no. of Reviews and Rating*** *?*

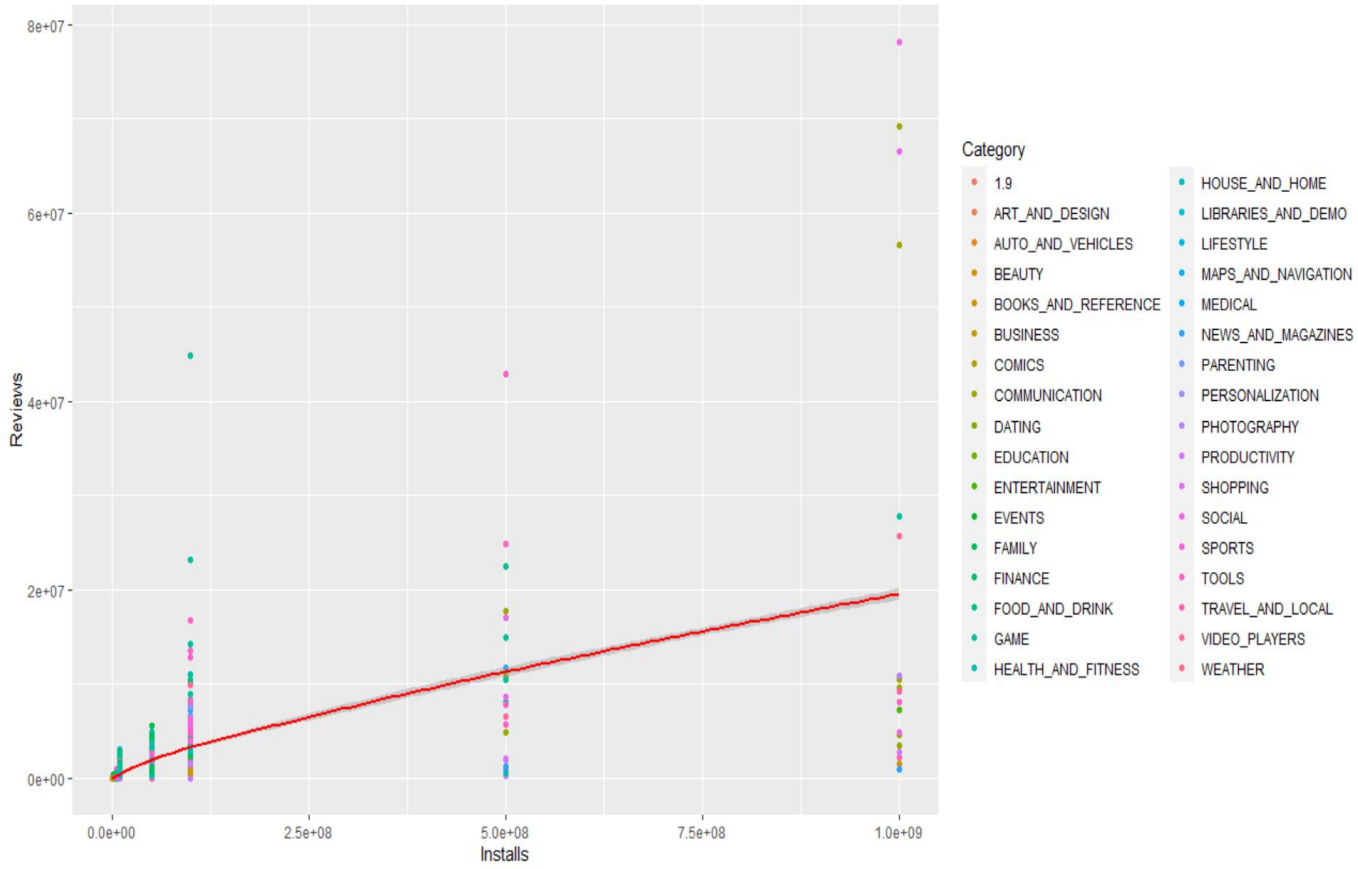
plot—>



Plot Analysis ->We can see that few categories like ,”game”,”communication”,”entertainment”,”photography”,”social”,”tools” , ”video player” etc. has a few no. of apps with high ratings and a large no. of reviews and we can conclude that if the no. of reviews are high then the rating of the apps are higher automatically, i.e, higher no. of reviews implies high ratings

q.3>*Is there any relation between* ***installs*** *and* ***no. of reviews.***

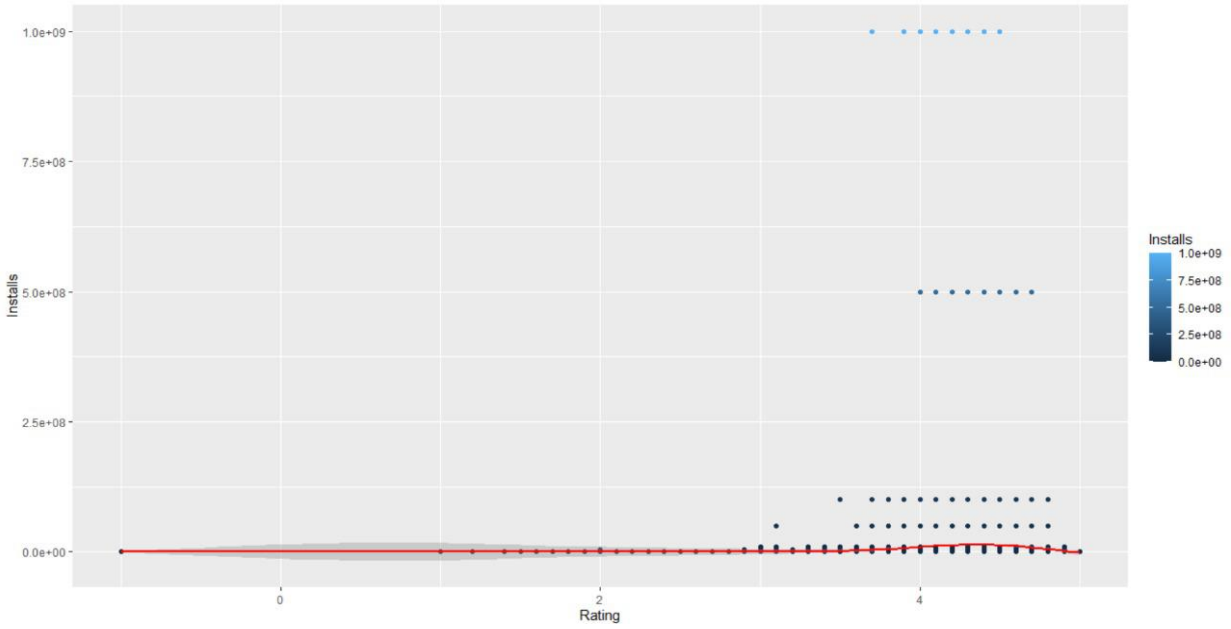
plot—>



Plot Analysis ->There is not a very high correlation between this two parameters. So we cannot say anything surely from here.

q.4>*Is there any relation between* ***installs*** *and* ***rating.***

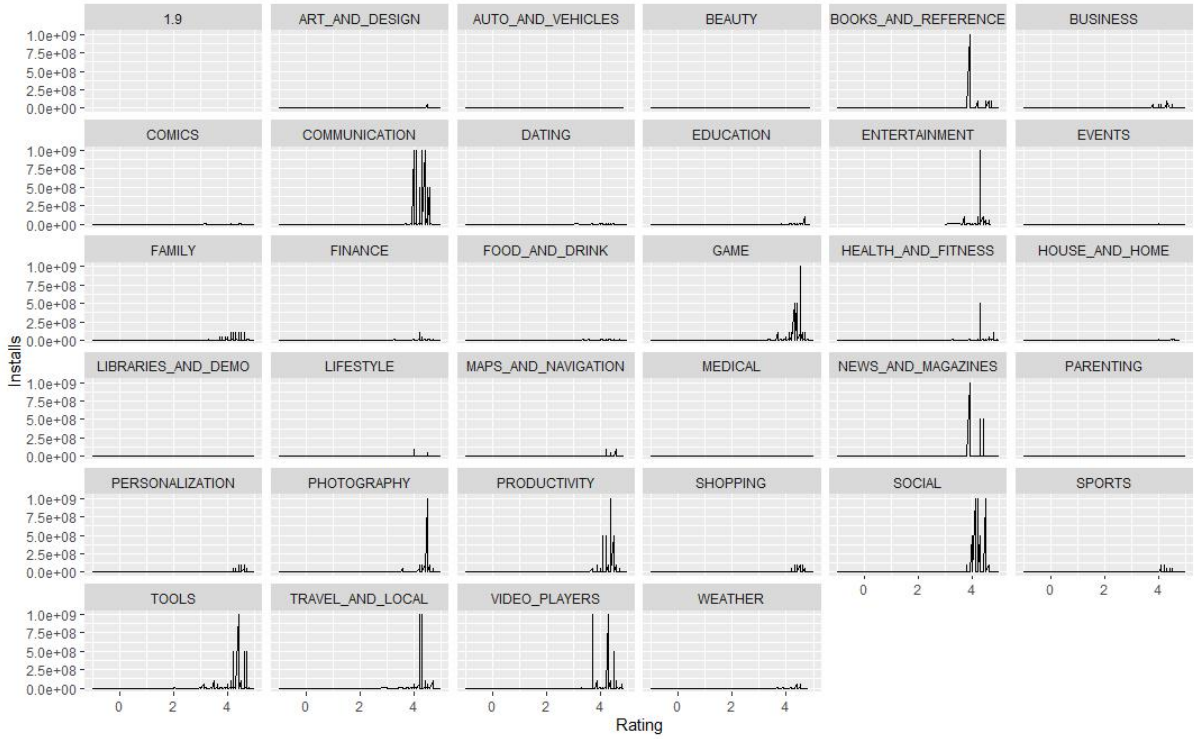
Plot->



Plot Analysis ->By This plotting again we don't get any Correlation between Rating and Installs.

Again we plot this ,this time category wise,

Plot->



Plot Analysis ->By this we can say that there are very few categories which have higher rated apps with higher no. of installs.

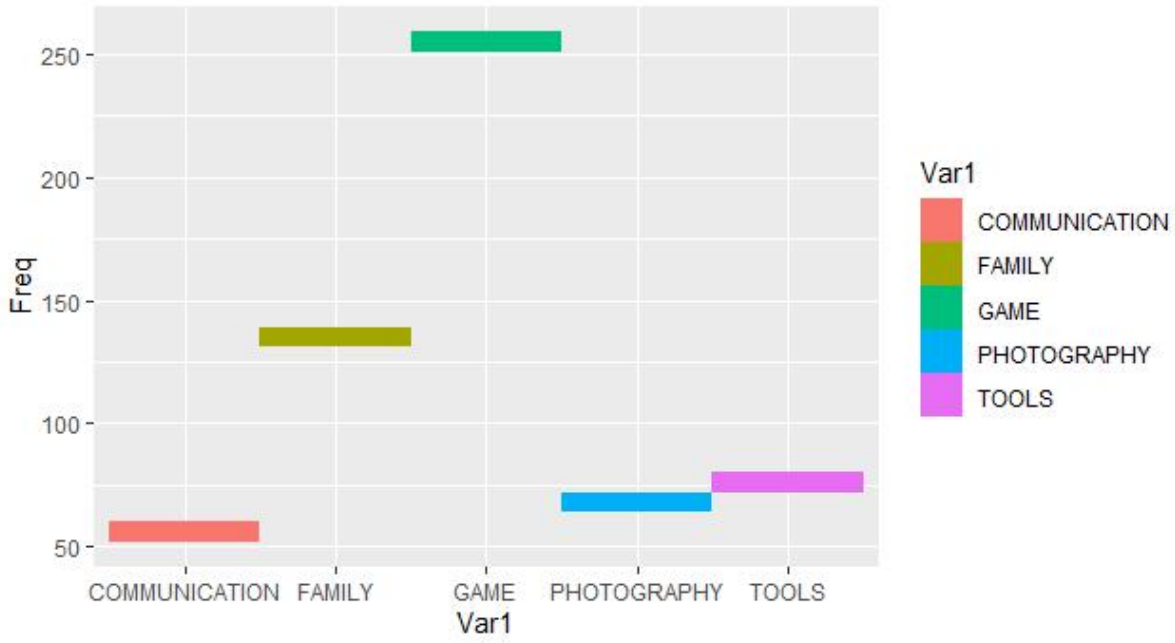
And further we can say that there is an one way relation between the no. of installs and the rating,i.e,higher no. of installs implies higher ratings,but it is not true for the other way round.

Therefore as we can see that any App for most of all Category if their no. of installs and no. of reviews is high then ratings of these apps will also be high.

q.5>So now,what are the categories that are carrying apps with higher no. of installs and reviews?

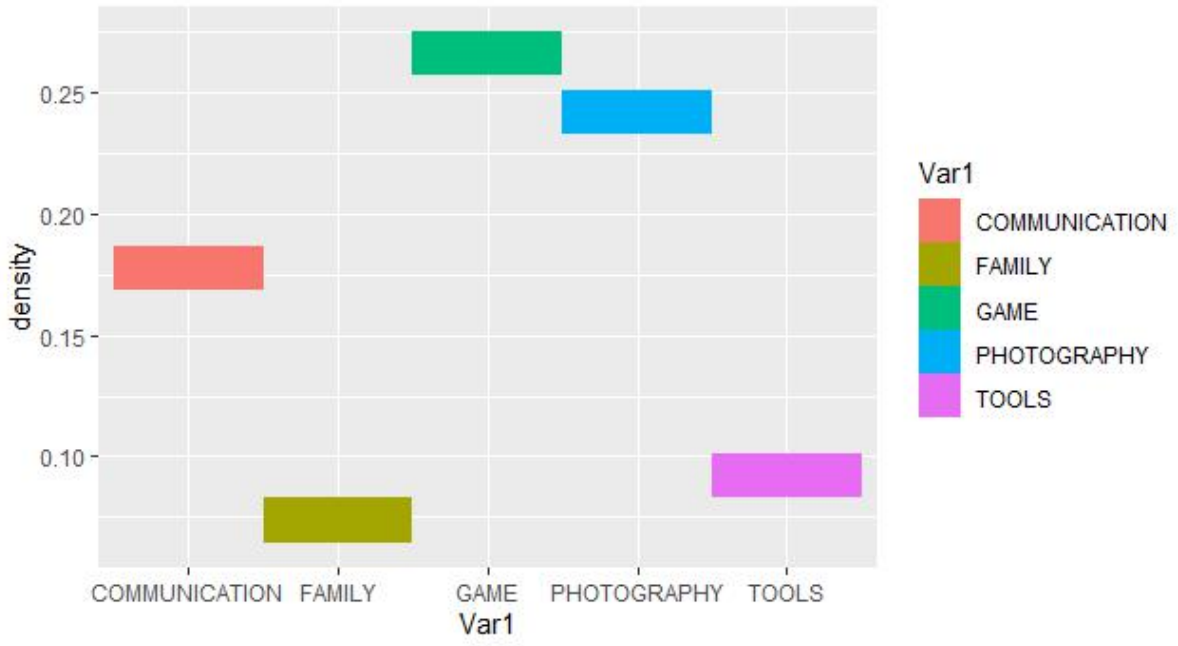
For finding Category Set

plot:-



Plot Analysis->But we cannot say that developers of gameing categories are more passionate by looking at this, as there are a huge difference in the no. of total apps of under categories. So, we will look into there density,

plot→

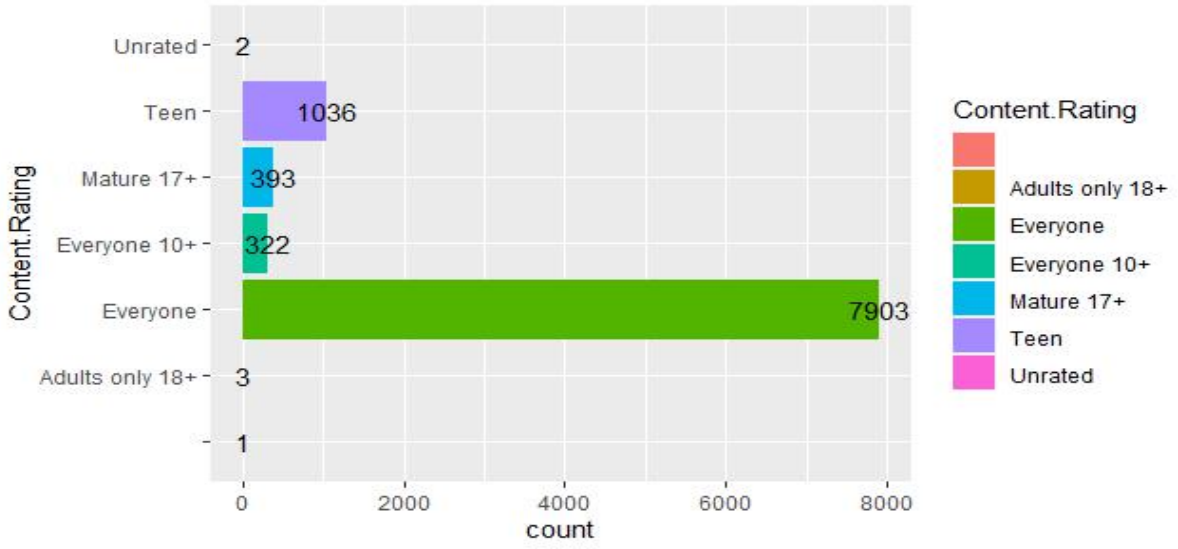


Plot Analysis->From this we can say that the category GAME and PHOTOGRAPHY has more passionate developers.

Finding the most popular age group →

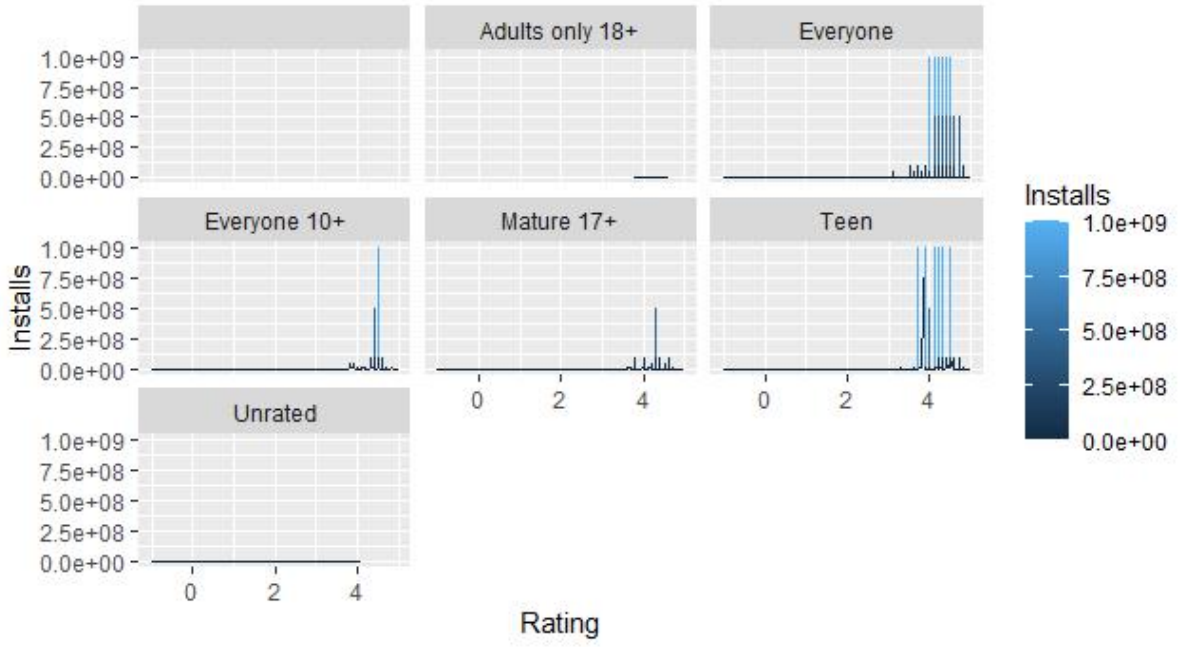
q.6> which Content.Rating has the highest no. of apps?

plot→



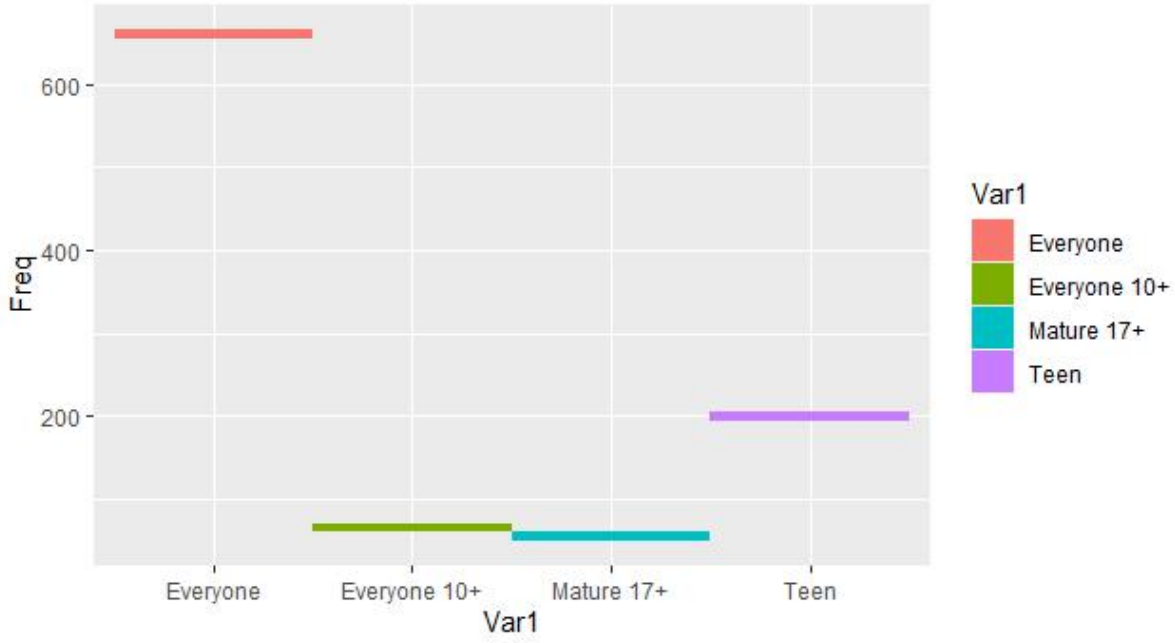
. Study of Install vs Rating, Content Rating wise—

plot →



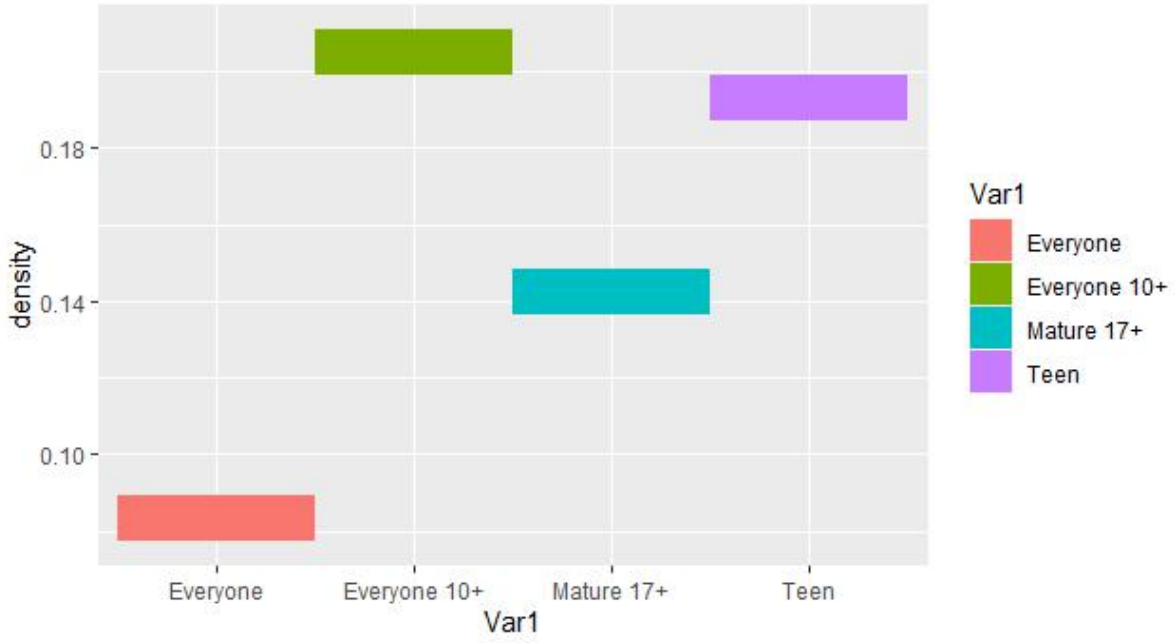
q.7>higher rated apps belongs to which age groups?

Plot→



But for the same reason we cannot conclude here, So we plot another way.

Plot →



Plot Analysis->So,we can now conclude that content rating “Everyone 10+” and “Teen” has the more passionate developers.

Now , we can say the apps which are of these content ratings and categories have the most passionate developers

The most passionate developers are→

Code—

```
app1<-subset(app,Installs>=mean(Installs),Reviews>=mean(Reviews))

w<-subset(app1,Content.Rating %in% c("Everyone 10+","Teen"))

p_developers<-subset(w,Category %in% c("GAME","PHOTOGRAPHY"))
```

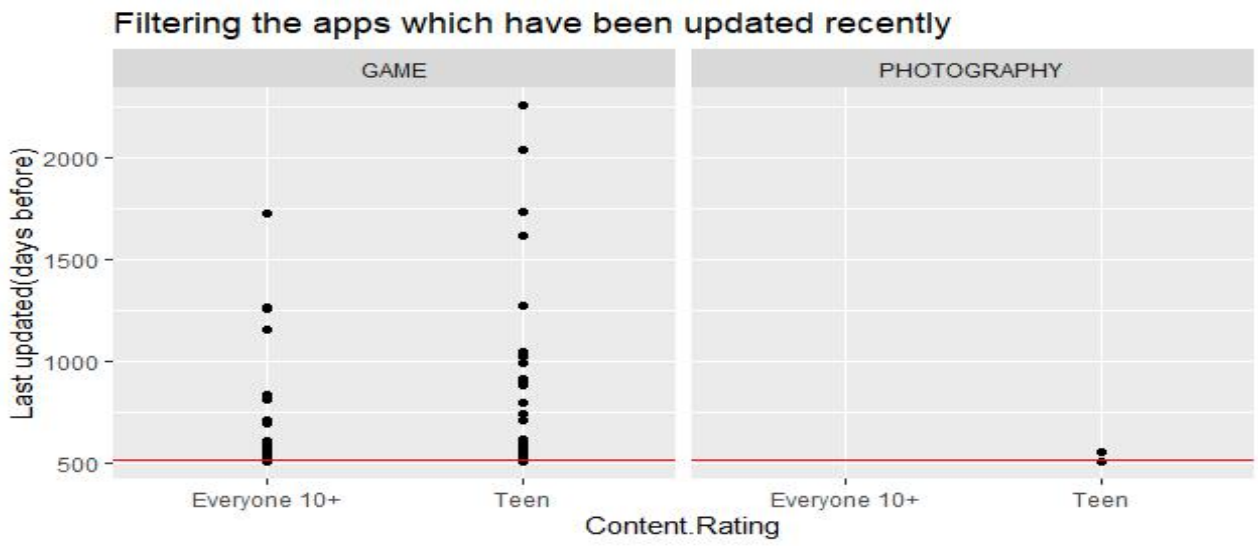
So in the dataframe p_developers we have the apps with the passionate developers compared to the others.

q.8> which of these apps have been updated recently?

#we first introduce an extra column "Time.Diffrence" denoting the difference between last updating date and 31-12-2019—

#now we plot time difference vs category vs content rating—

Plot→



We now filter these apps which have been updated recently(whose time difference is >= mode(time difference)

Code→

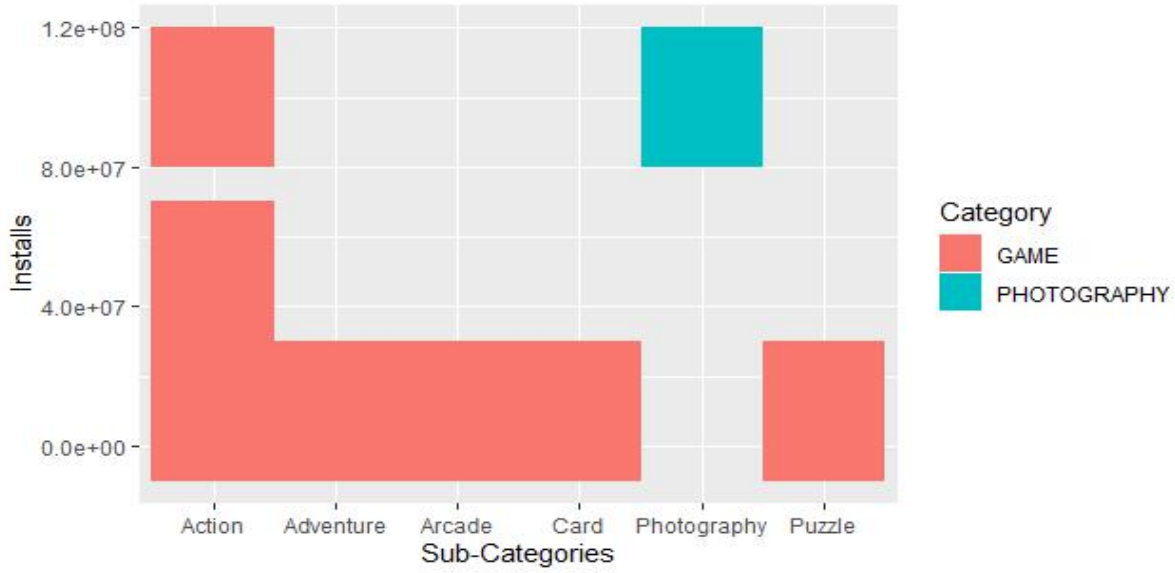
```
v<-
subset(p_developers,p_developers$Time.Difference<=getmode(p_developers
$Time.

Difference))
```

So now we have 19 applications which we can say has the most passionate developers,who really worked hard and working hard to serve the best item.

##Further observations(on these applications)→

1. What are the sub-categories which have the highest downloads(Installs) in them?



2. **Avg cost of these apps?**

   table(v$Type)

   free
   19

   The most passionate developers made there apps free