DM-Project1

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.7
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## -- Conflicts ------------------------------------------------- tidyverse_conflicts() --
## x dplyr::combine()  masks randomForest::combine()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()      masks stats::lag()
## x ggplot2::margin() masks randomForest::margin()
```

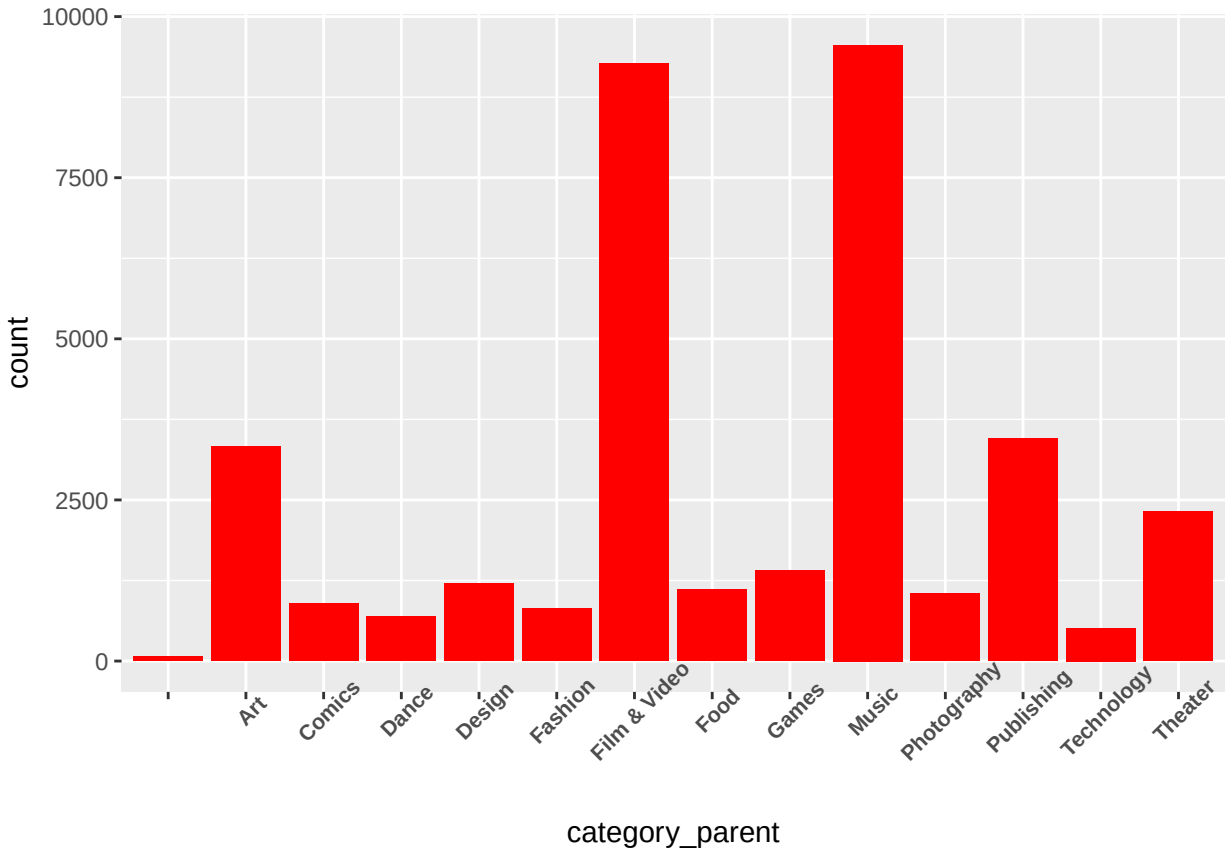
```
crowd = read.csv(file = '/Users/DwayneChen/Desktop/Project1_Crowdfunding/KickStarterData.csv')
head(crowd)
```

```
##   id
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
##                                                                                 url
## 1            http://www.kickstarter.com/projects/darkpony/drawing-for-dollars
## 2   http://www.kickstarter.com/projects/dphiffer/offline-wikipedia-iphone-app
## 3             http://www.kickstarter.com/projects/nymab/new-york-makes-a-book
## 4 http://www.kickstarter.com/projects/260428906/help-me-write-my-second-novel
## 5                  http://www.kickstarter.com/projects/kicey/kicey-to-iceland
## 6            http://www.kickstarter.com/projects/ericberlin/crossword-puzzles
##   name description backers      category category_parent  duration
## 1                        3  Illustration             Art  8.463843
## 2                       25 Open Software      Technology 79.641412
## 3                      110    Journalism      Publishing 17.918970
## 4                       18       Fiction      Publishing 29.965394
## 5                       31   Photography     Photography 48.969630
## 6                      163      Art Book      Publishing 61.902743
##           endDate goal likes latitude longitude     location
## 1  2009/5/2 23:59   20     0        0         0
## 2 2009/7/13 23:59   99     0        0         0
## 3  2009/5/16 2:59 3000     0        0         0 New York, NY
## 4 2009/5/28 17:09  500     0        0         0
## 5 2009/6/16 21:00  350     0        0         0
## 6  2009/6/30 2:32 1500     0        0         0
##   combined_location   State       Country location.population owner_name
## 1                 0 #VALUE!       #VALUE!                #N/A
```

```
## 2                 0 #VALUE!       #VALUE!               #N/A
## 3      New York, NY      NY United States             8391881
## 4                 0 #VALUE!       #VALUE!               #N/A
## 5                 0 #VALUE!       #VALUE!               #N/A
## 6                 0 #VALUE!       #VALUE!               #N/A
##   owner_profile owner_backing_count owner_friends pledged X..of.Goal
## 1                                 1           539      35   1.750000
## 2                                 4             0     145   1.464646
## 3                                 0             0    3329   1.109667
## 4                                 1           574     563   1.126000
## 5                                 3           326    1630   4.657143
## 6                                 2             0    2265   1.510000
##       update_time reward_count Finished Successful
## 1 2012/5/22 20:49            1     TRUE       TRUE
## 2 2012/5/22 20:44            1     TRUE       TRUE
## 3 2012/5/22 19:44            2     TRUE       TRUE
## 4 2012/5/22 22:39            7     TRUE       TRUE
## 5 2012/5/22 20:09            5     TRUE       TRUE
## 6 2012/5/22 20:39            3     TRUE       TRUE
```

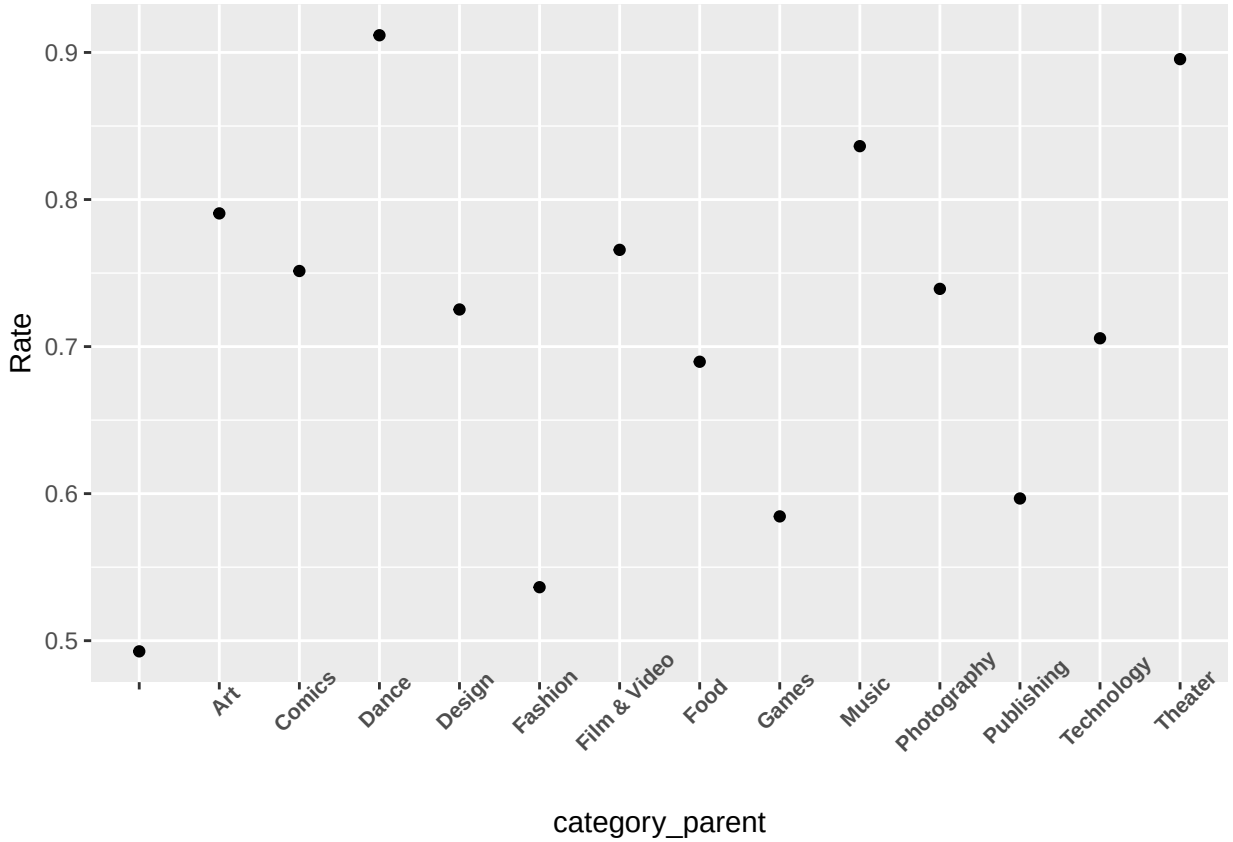
Data Exploration

```
crowd_1 <- select(crowd, "category_parent")
ggplot(data = crowd_1) + geom_bar(mapping = aes(x = category_parent), fill = "red") + theme(axis.text.x
```



We can see that *Film & Video* and *Music* have a overwhelmingly exceeding quantity than other categories.

```
crowd_2 <- select(crowd , 'category_parent', 'Successful') %>% table() %>% as.data.frame() %>% spread(.
names(crowd_2) <- c('category_parent', 'False', 'True')
succ_rate <- mutate(crowd_2, Rate = True / (False + True))
ggplot(data = succ_rate) + geom_point(mapping = aes(x = category_parent, y = Rate)) + theme(axis.text.x
```



We can see that though *Film & Video* is large in quantity of examples, it's successful rate is not as high as *Music*. Besides, the successful rate of *Dance* and *Theater* is fairly high with 0.9's successful rate. It was pretty out of my expectation, for I've thought that most of the crowd-funding programs are producing some advanced technology. Through these two plots, we can easily say that if you cast an crowd-funding program in *Dance* and *Theater* you are just a step from success.