

Kickstarter_CaseStudy

Sb Fuller

1. First I installed packages I would need:

```
install.packages("moments")
install.packages("ggplot2")
install.packages("dplyr")
install.packages("plot3D")
library(moments)
library(ggplot2)
library(dplyr)
library(plot3D)
library(readr)
```

2. Next, I loaded the cleaned dataset I could be using:

```
kickstarter_casestudy_DSI_kickstarterscrape_dataset
<- read_csv("kickstarter_casestudy_DSI_kickstarterscrape_dataset.csv",
col_types = cols(goal = col_number(), pledged = col_number(),
funded_percentage_asper = col_number(), funded_percentage_asdec = col_number(),
backers = col_number(), levels = col_number(), updates = col_number(),
comments = col_number(), duration = col_number()))
View(kickstarter_casestudy_DSI_kickstarterscrape_dataset)
```

3. I began my analysis by quickly checking the distribution of the project duration data and found non-normal distribution:

```
ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset, aes(x=duration)) +
  geom_bar(fill = "springgreen4") +
  ggtitle("Distribution of Project Durations") + xlab("Durations") + ylab("Count") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 0),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("duration_dist.jpg")
```

4. The plotted distribution of project backers was so skewed it could barely be plotted, so instead I simply calculated the skew to find that the backers data had a positive skew of approximately 87.32. I also calculated the average pledged amount for all projects which came out to 5035.06:

```
backers = kickstarter_casestudy_DSI_kickstarterscrape_dataset$backers
pledged = kickstarter_casestudy_DSI_kickstarterscrape_dataset$pledged
skewness(backers)
mean(pledged)
```

5. Next, I created subset dataframes of successful and failed projects, and performed general statistical summaries of all columns with numerical value:

```
failed <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
  filter(status == 'failed')
successful <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
```

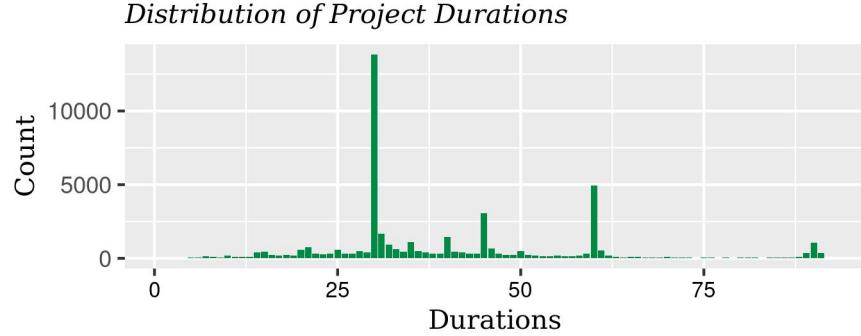


Figure 1: Plot described in Section 3.

```

filter(status == 'successful')
View(successful)
View(failed)
failed %>%
select(goal, pledged, funded_percentage_asper, backers, levels, updates, duration) %>%
summary()
successful %>%
select(goal, pledged, funded_percentage_asper, backers, levels, updates, duration) %>%
summary()

```

6. Addressing the first question this analysis, What is the best length of time to run a campaign?... I started by producing a bar graph to better understand amounts of durations depending on success and failure. These distributions deviated very little from one another and the general duration distribution, with largest numbers of projects around 30, 60, and 90 days:

```

ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset, aes(x=duration)) +
  geom_bar(fill = "springgreen4") + facet_wrap(~status) +
  ggtitle("Projects of Each Duration by Status") + xlab("Duration in Days") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 0),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("duration_status.jpg")

```

7. I then produced the following scatter plots to better understand correlation between duration and funded percentage. This confirmed clustering around 30, 60, and 90 days. The second plot showed that no live project was older than 60 days, giving us a clue about what day of the month this data was collected.

```

ggplot(data=another_most_percentage, aes(x=duration, y=funded_percentage_asdec)) +
  geom_point(alpha = 1/10) + geom_smooth(color = "springgreen4") +
  ggtitle("Correlation of Funded Percentage and Duration") +
  xlab("Duration in Days") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 45),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("duration_perc.jpg")

ggplot(data=another_most_percentage, aes(x=duration, y=funded_percentage_asper)) +
  geom_point(alpha = 1/10) + geom_smooth(color = "springgreen4") + facet_wrap(~status) +

```

Projects of Each Duration by Status

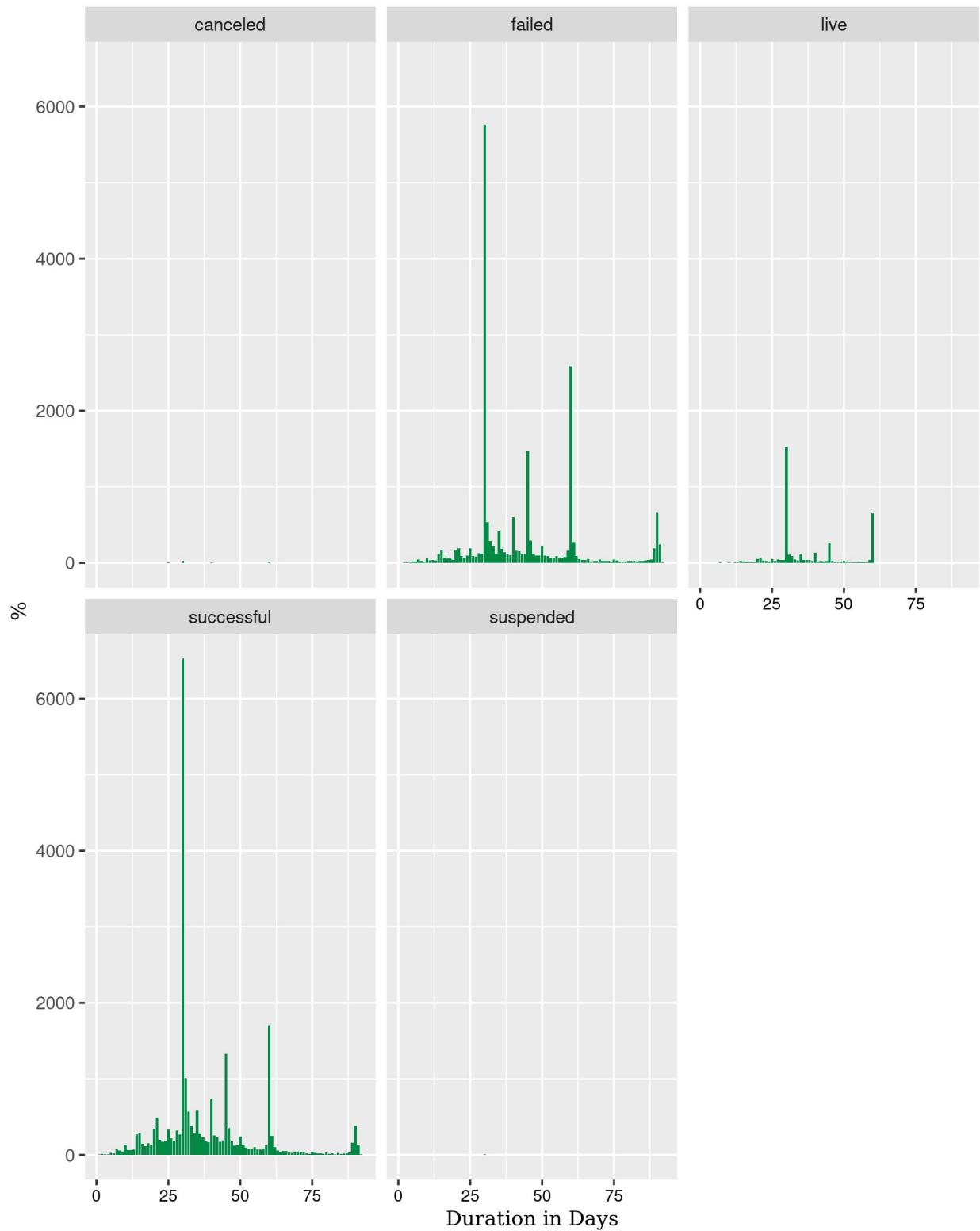


Figure 2: Plot Described in Section 6.

Correlation of Funded Percentage and Duration

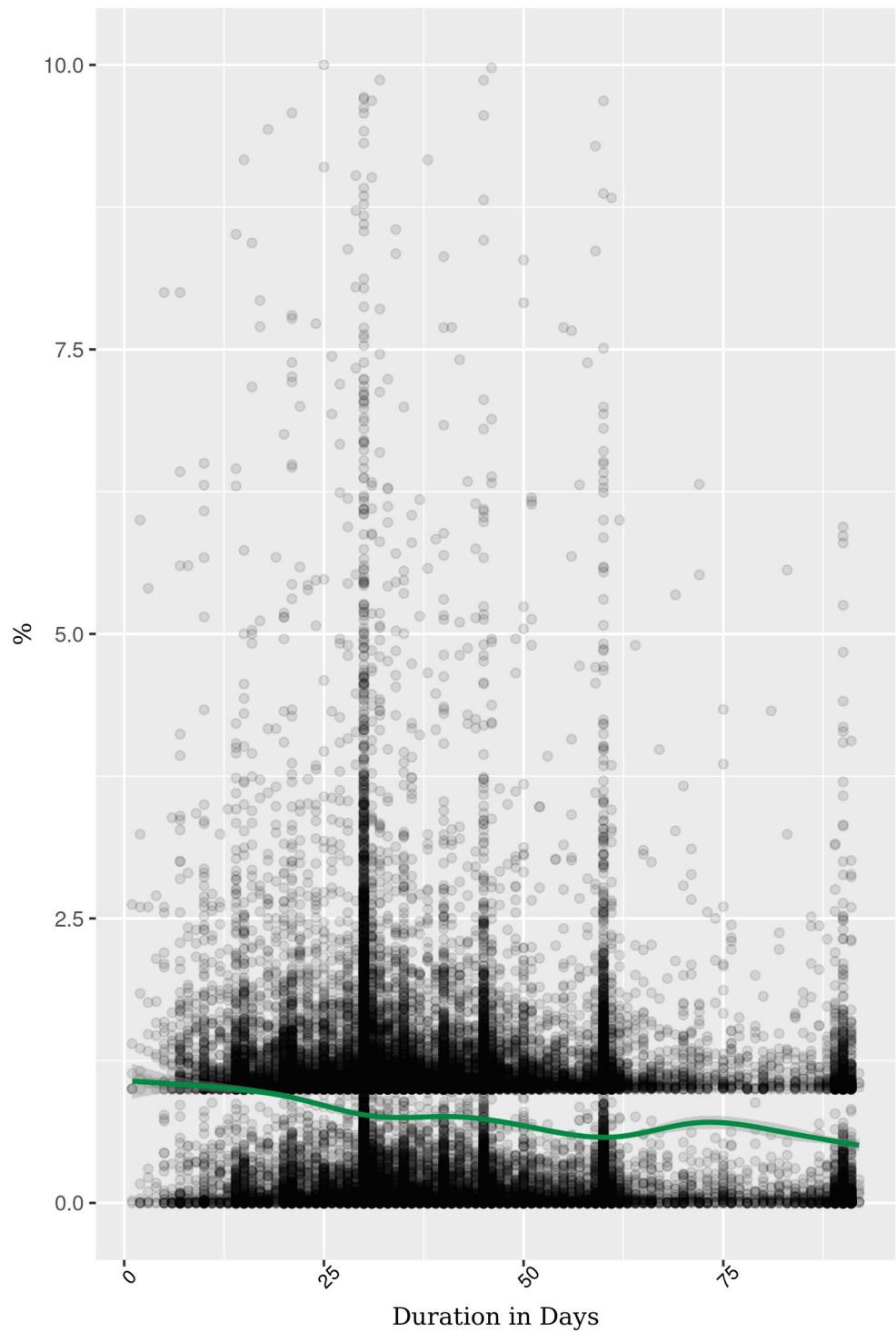


Figure 3: First Plot Described in Section 7.

```

ggtitle("Correlation of Funded Percentage and Duration by Status") +
  xlab("Duration in Days") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 0),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("duration_perc_status.jpg")

```

8. The following bar graph was produced to better understand average funding percentages of different durations. This shows us that projects lasting 9, 49, 70 days had the greatest average funded percentage:

```

ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset,
aes(x=duration, y=funded_percentage_asper)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean",
           fill = "seagreen3", color = "springgreen4") +
  ggtitle("Average Funded Percentage by Duration") + xlab("Duration in Days") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 45),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("duration_perc_average.jpg")

```

9. Given the apparent randomness of these spikes, subsets were produced with various recombinations of outlier removal for improved plotting... goal amounts less than 20,000, funded percentages less than 10 times the goal, goal amounts less than 2500:

```

most_goals <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>% filter(goal < 20000)
most_percentage <- most_goals %>% filter(funded_percentage_asdec <= 10)
smaller_goals <- most_percentage %>% filter(goal < 2500)
another_most_percentage <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
  filter(funded_percentage_asdec <= 10)

```

10. With outliers removed, the data was plotted again. Here, a more linear negative correlation was shown between duration and funding percentage.

```

ggplot(data=another_most_percentage, aes(x=duration, y=funded_percentage_asper)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean",
           fill = "seagreen3", color = "springgreen4") + geom_smooth(color = "springgreen4") +
  ggtitle("Average Funded Percentage by Duration w/o Outliers") +
  xlab("Duration in Days") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 45),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("duration_percbelow10_average_.jpg")

```

11. Next, regarding the second question of analysis, What's the idea pledge goal?... quick summary statistics were pulled showing that the average goal of all projects was 11,957, the average goal of failed projects was 16,366, and the average goal of successful projects was 5,486. This gives us a clue that perhaps smaller goals are more likely to succeed.

```

kickstarter_casestudy_DSI_kickstarterscrape_dataset %>% select(goal) %>% summary()
successful %>% select(goal) %>% summary()
failed %>% select(goal) %>% summary()

```

12. The following scatter plot was produced to look for correlation between goal amount and funded percentage, with outliers removed. This showed us that funded percentages are highest when the goal is the lowest. This steep correlation exists up to the 1000 goal amount, where the curve flattens a bit

Correlation of Funded Percentage and Duration by Status

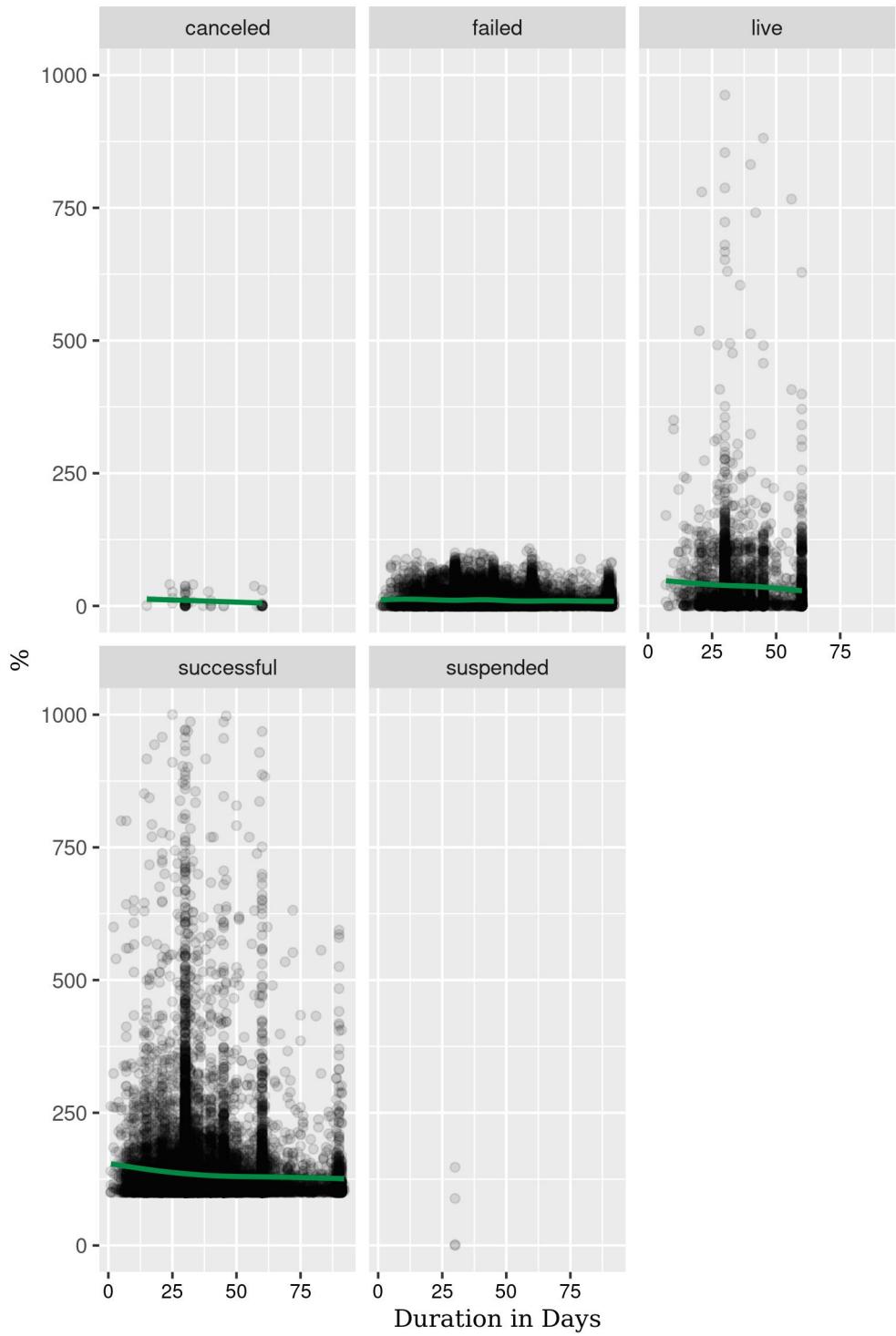


Figure 4: Second Plot Described in Section 7.

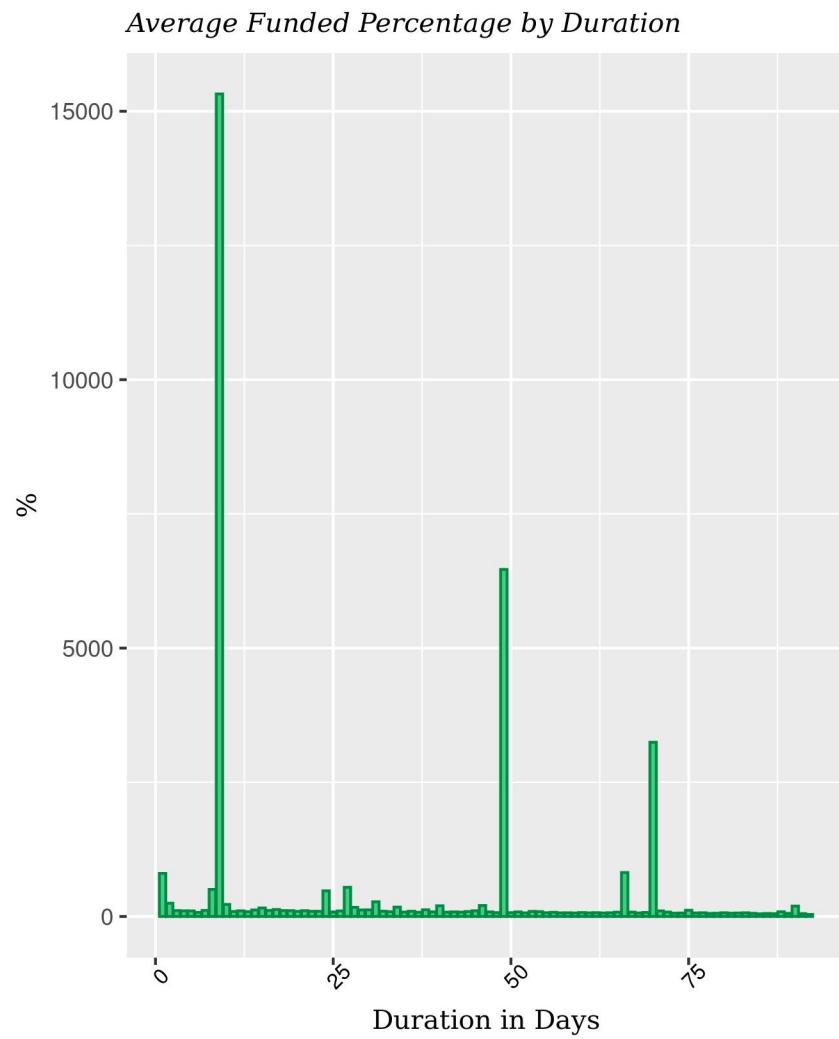


Figure 5: First Plot Described in Section 8.

Average Funded Percentage by Duration w/o Outliers

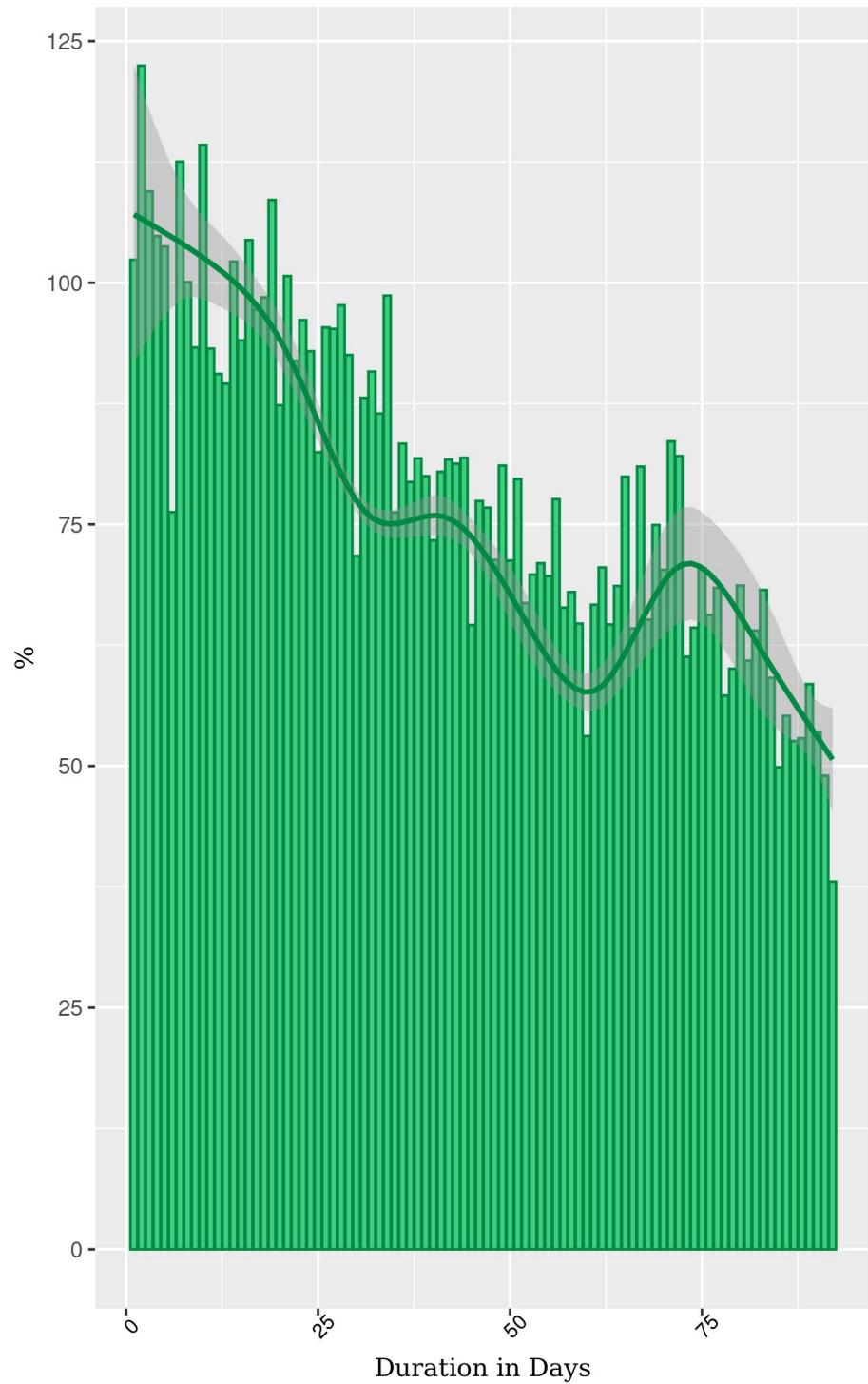


Figure 6: First Plot Described in Section 10.

until the 5000 goal amount and so on. Vertical spikes along 500 dollar increments likely have to do with the frequency of projects structuring their goals in 500 dollar increments and less so on correlation between funded percentage and goal amount.

```
ggplot(data=most_percentage, aes(x=goal, y=funded_percentage_asdec)) +
  geom_point(alpha = 1/10) + geom_smooth(color = "springgreen2") +
  ggtitle("Correlation of Funded Percentage to Goal") + xlab("Funding Goal") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 45),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("goal_percbelow10and20000.jpg")
```

13. Zoomed in further using tighter subsets, the line of correlation dips below one (100% funded project) near the goal amount of 1000 dollars:

```
ggplot(data=smaller_goals, aes(x=goal, y=funded_percentage_asdec)) +
  geom_point(alpha = 1/10) + geom_smooth(color = "springgreen2") +
  ggtitle("Correlation of Funded Percentage to Goal") + xlab("Funding Goal") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 45),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("goal_percbelow10and2500.jpg")
```

14. To better understand the correlation between goal amount, duration, and funded percentage, the following variables were defined for quicker analysis and correlation calculation were run. These both showed similarly weak correlation between funded percentage and both duration and goal amount:

```
x <- most_percentage$goal
y <- most_percentage$duration
z <- most_percentage$funded_percentage_asdec
sx <- smaller_goals$goal
sy <- smaller_goals$duration
sz <- smaller_goals$funded_percentage_asdec
sw <- smaller_goals$status
cor(y, z)
cor(x, z)
```

15. Furthermore, the following 3D plots were produced from the two filtered subsets to visualize two degrees of zoom on the core cluster of data points showing two potential axes of correlation on funded percentage:

```
scatter3D(x, y, z, bty = "b2", pch=1, col = gg.col(n=10, alpha=.5),
          theta = 0, phi = 20,
          xlab = "Goal", ylab ="Duration", zlab = "%", ticktype = "detailed",
          type = "h", nticks=12, cex.lab = 1, cex.axis = .5)
ggsave("gdp_3D_01.jpg")
scatter3D(x, y, z, bty = "b2", pch=1, col = gg.col(n=10, alpha=.5),
          theta = 23.5, phi = 20,
          xlab = "Goal", ylab ="Duration", zlab = "%", ticktype = "detailed",
          type = "h", nticks=12, cex.lab = 1, cex.axis = .5)
ggsave("gdp_3D_02.jpg")
scatter3D(x, y, z, bty = "b2", pch=1, col = gg.col(n=10, alpha=.5),
          theta = 45, phi = 20,
          xlab = "Goal", ylab ="Duration", zlab = "%", ticktype = "detailed",
          type = "h", nticks=12, cex.lab = 1, cex.axis = .5)
ggsave("gdp_3D_03.jpg")
```

Correlation of Funded Percentage to Goal

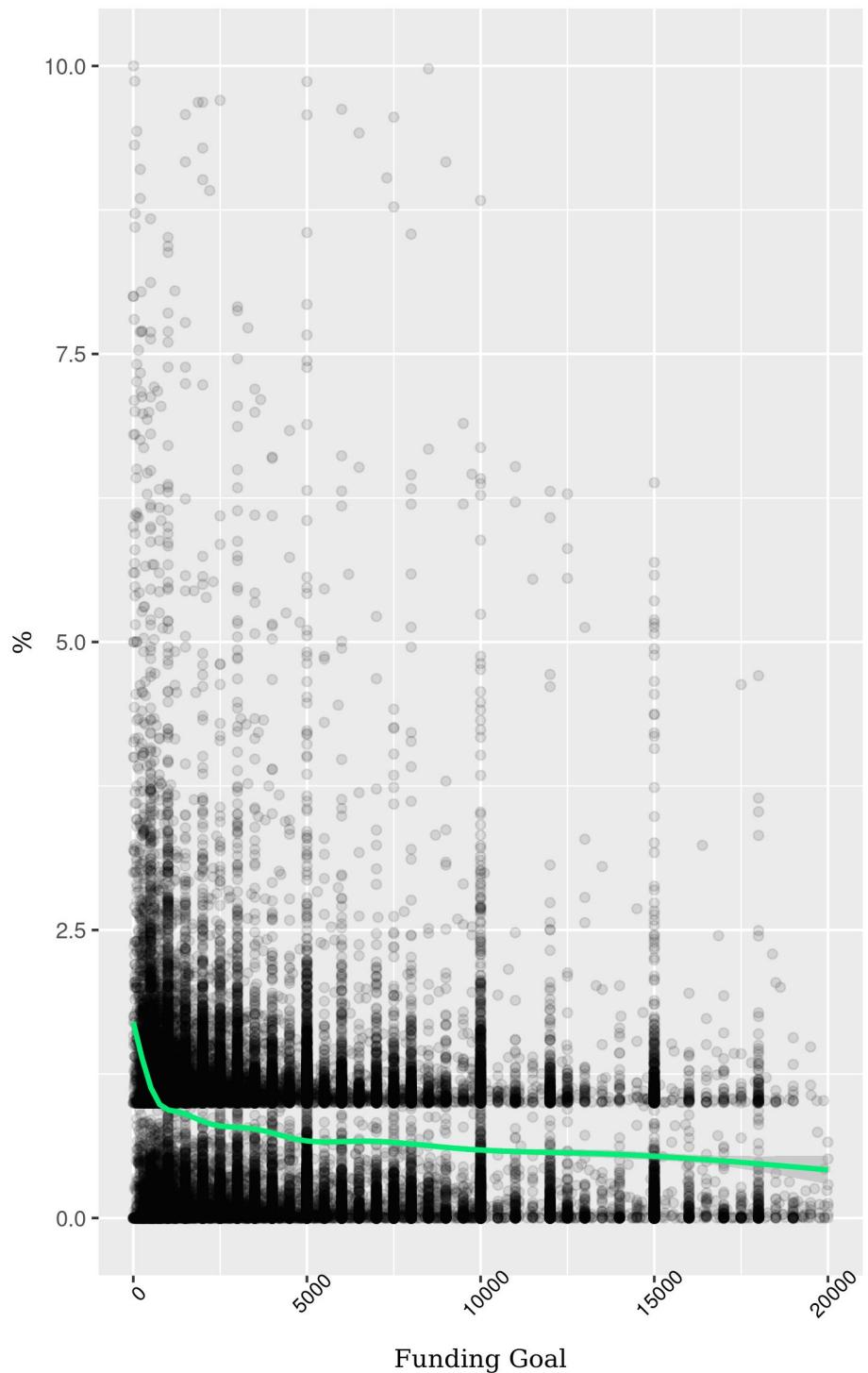


Figure 7: First Plot Described in Section 12.

Correlation of Funded Percentage to Goal

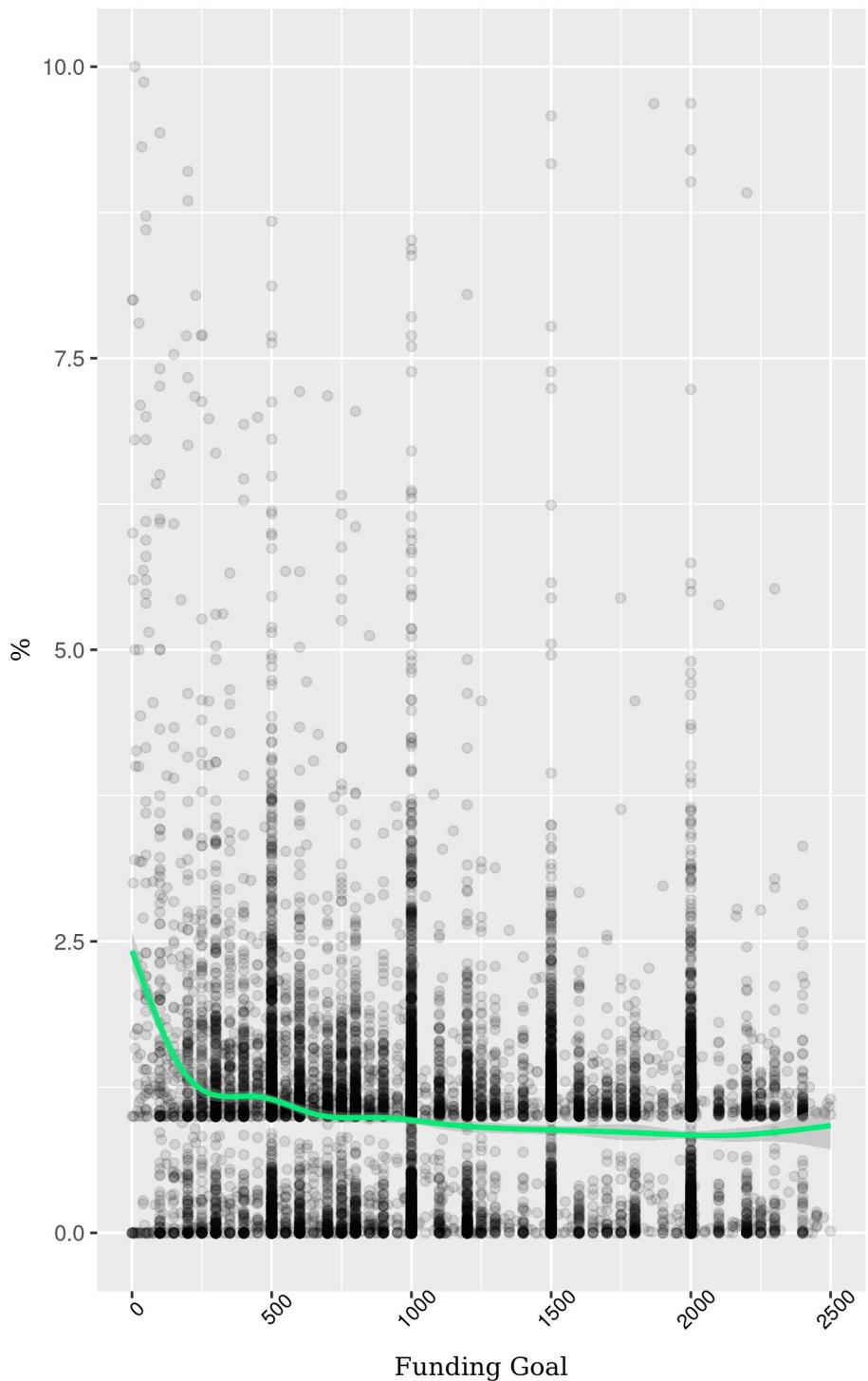


Figure 8: First Plot Described in Section 13.

```

scatter3D(x, y, z, bty = "b2", pch=1, col = gg.col(n=10, alpha=.5),
theta = 57.5, phi = 20,
xlab = "Goal", ylab ="Duration", zlab = "%", ticktype = "detailed",
type = "h", nticks=12, cex.lab = 1, cex.axis = .5)
ggsave("gdp_3D_04.jpg")
scatter3D(x, y, z, bty = "b2", pch=1, col = gg.col(n=10, alpha=.5),
theta = 90, phi = 20,
xlab = "Goal", ylab ="Duration", zlab = "%", ticktype = "detailed",
type = "h", nticks=12, cex.lab = 1, cex.axis = .5)
ggsave("gdp_3D_05.jpg")

```

16. Next, attending to the question of what type of project would be most successful at getting funded, the following bar graph was produced to better understand amounts and proportions of failed and successful project by category. This showed us that Design, Fashion, Games, Photography, Publishing, and Technology had more failed projects than successful projects and Art, Film & Video, and Music appeared to be, not only the most popular, but most likely to have more successes than failure. Music appears to show the most extreeme difference of success over failure:

```

ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset, aes(x=status)) +
  geom_bar(colour = "black", fill = "seagreen3") + facet_wrap(~category) +
  ggtitle("Number of Failed and Successful Projects by Category") +
  xlab("Status") + ylab("Count") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 90),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("status_category_count.jpg")

```

17. To confirm this, the following ratios of failur to success were calculated and input into a new dataframe:

```

#General Ratio of success to failure (0.826934):
sum(kickstarter_casestudy_DSI_kickstarterscrape_dataset$status == "failed", na.rm=TRUE) /
sum(kickstarter_casestudy_DSI_kickstarterscrape_dataset$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Art (0.7524988):
art <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Art')
sum(art$status == "failed", na.rm=TRUE) / sum(art$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Film & Video (0.9610755):
film_video <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Film & Video')
sum(film_video$status == "failed", na.rm=TRUE) / sum(film_video$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Music (0.4805847) highest ratio for success:
music <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Music')
sum(music$status == "failed", na.rm=TRUE) / sum(music$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Design (1.150138):
design <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Design')
sum(design$status == "failed", na.rm=TRUE) / sum(design$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Fashion (2.038806) Highest ratio for failure:
fashion <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Fashion')
sum(fashion$status == "failed", na.rm=TRUE) / sum(fashion$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Games ( 1.313788):
games <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Games')

```

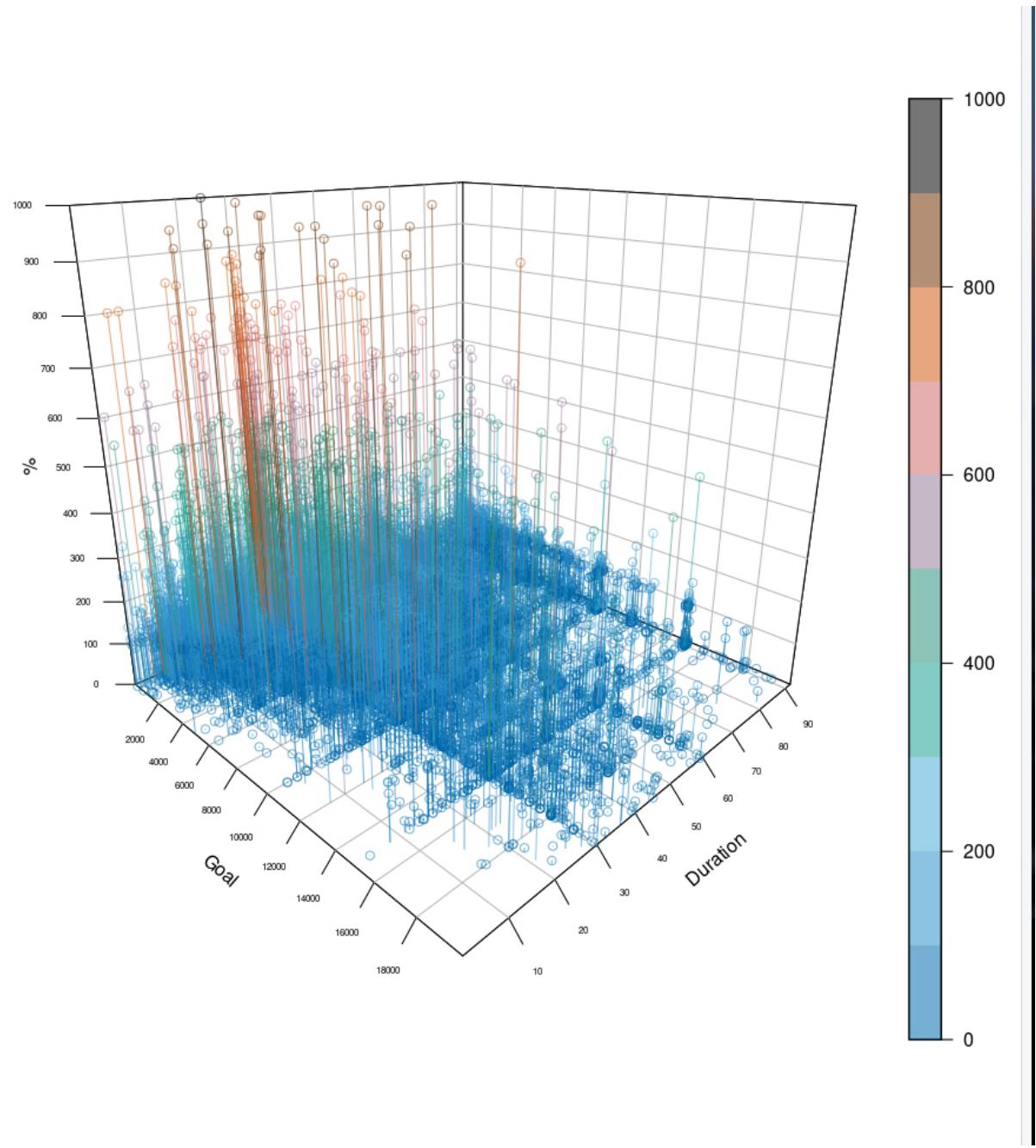


Figure 9: Plot Described in Section 15.

Number of Failed and Successful Projects by Category

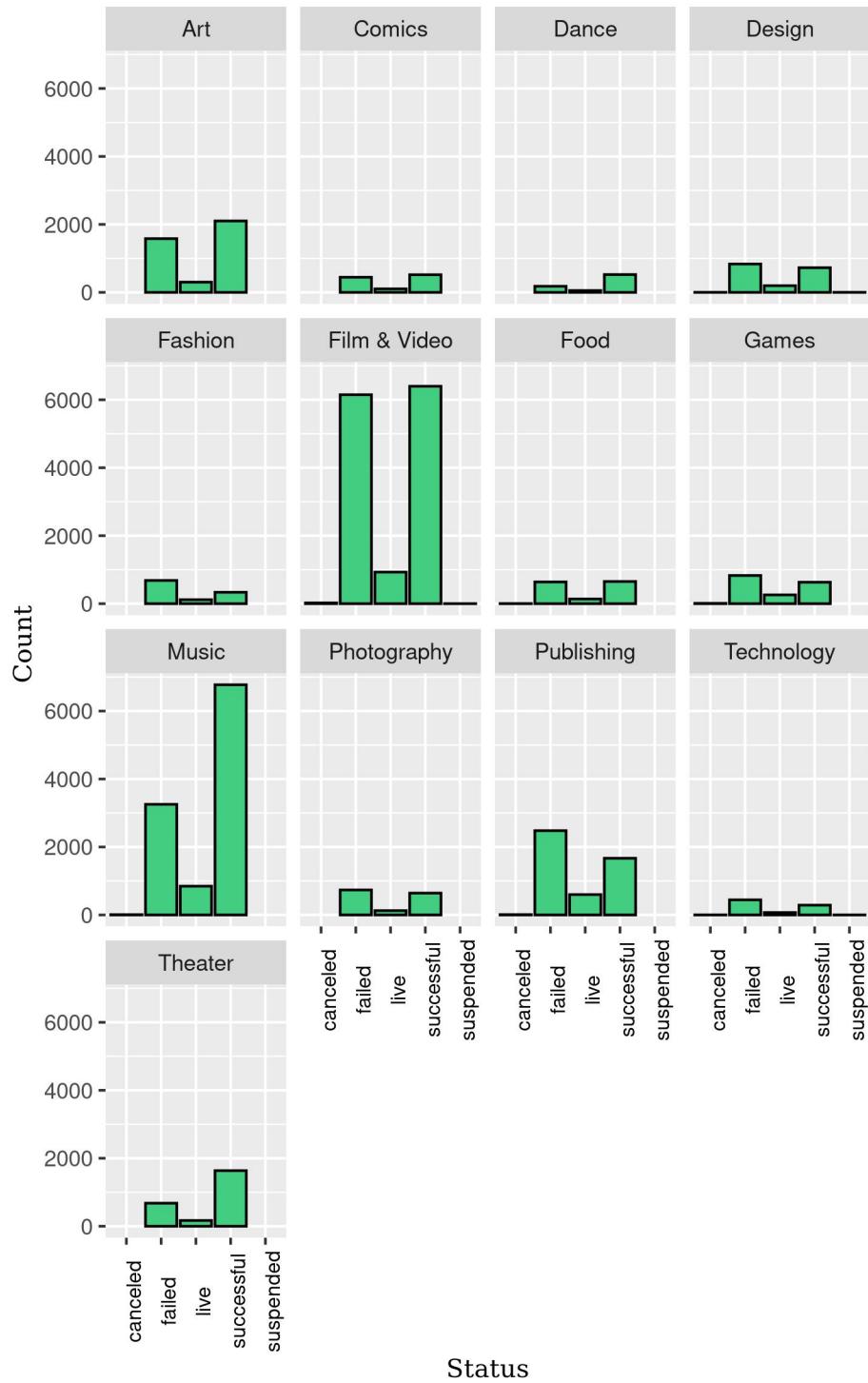


Figure 10: Plot Described in Section 16.

```

sum(games$status == "failed", na.rm=TRUE) / sum(games$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Photography (1.143079):
photography <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Photography')
sum(photography$status == "failed", na.rm=TRUE) / sum(photography$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Publishing ( 1.486811):
publishing <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Publishing')
sum(publishing$status == "failed", na.rm=TRUE) / sum(publishing$status == "successful", na.rm=TRUE)
#Ratio of success to failure for Technology (1.541667):
tech <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>%
filter(category == 'Technology')
sum(tech$status == "failed", na.rm=TRUE) / sum(tech$status == "successful", na.rm=TRUE)
#Creating a data.frame of failure ratios by category:
success_failure_ratios <- data.frame (Category = c("Art", "Film & Video", "Music", "General", "Design",
Ratio = c(sum(art$status == "failed", na.rm=TRUE) / sum(art$status == "successful", na.rm=TRUE),
sum(film_video$status == "failed", na.rm=TRUE) / sum(film_video$status == "successful", na.rm=TRUE),
sum(music$status == "failed", na.rm=TRUE) / sum(music$status == "successful", na.rm=TRUE),
sum(kickstarter_casestudy_DSI_kickstarterscrape_dataset$status == "failed", na.rm=TRUE) /
sum(design$status == "failed", na.rm=TRUE) / sum(design$status == "successful", na.rm=TRUE),
sum(fashion$status == "failed", na.rm=TRUE) / sum(fashion$status == "successful", na.rm=TRUE),
sum(games$status == "failed", na.rm=TRUE) / sum(games$status == "successful", na.rm=TRUE),
sum(photography$status == "failed", na.rm=TRUE) / sum(photography$status == "successful", na.rm=TRUE),
sum(publishing$status == "failed", na.rm=TRUE) / sum(publishing$status == "successful", na.rm=TRUE),
sum(tech$status == "failed", na.rm=TRUE) / sum(tech$status == "successful", na.rm=TRUE)))
View(success_failure_ratios)

```

18. This new data frame was then plotted to show us liklihoods of failure in the top and bottom performing categories. This confirmed Music to be the category most likely to bring about success, and fashion the least likely to bring about success:

```

ggplot(data=success_failure_ratios, aes(x=Category, y=Ratio)) +
  ggtitle("Failure Likelihood") + xlab("Category") + ylab("Failure Ratio") +
  geom_point(shape = 21, colour = "black", fill = "seagreen3", size = 5, stroke = 5) +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 45),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("failure.jpg")

```

19. As the distinction of success/failure alone isnt the entirity of information at our disposal to understand the streghths or weaknesses of different categories in being funded, the following bar graph was produces to show average funding percentage by category. This showed that the top funded percentages by category are Art, Design, and Film & Video, and Comics:

```

ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset,
aes(x=category, y=funded_percentage_asper)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean",
fill = "seagreen3", color = "springgreen4") +
  ggtitle("Number of Failed and Successful Projects by Subcategory") +
  xlab("Category") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 90),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=10, family="Serif"))
ggsave("cat_dec_avg.jpg")

```

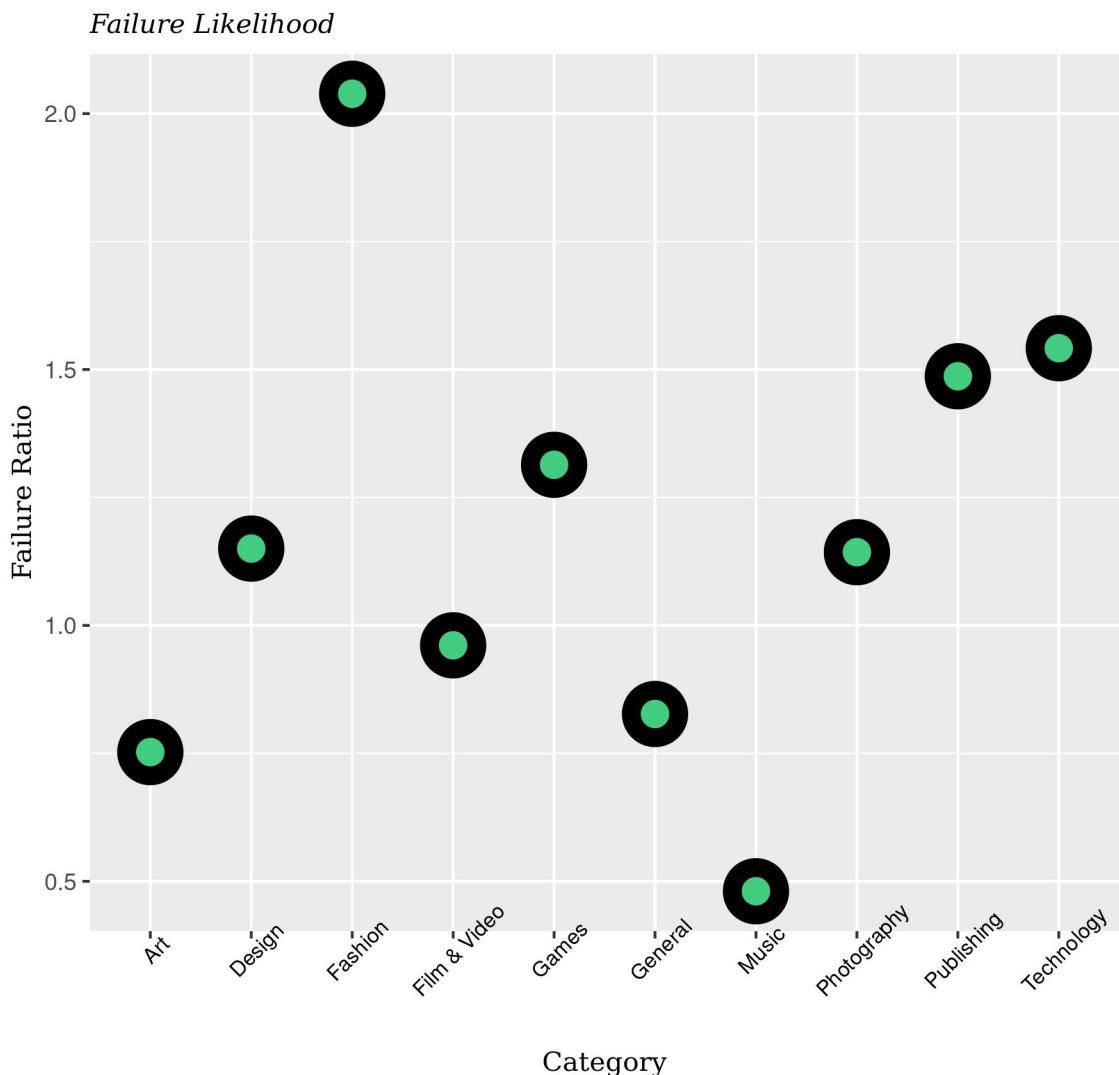


Figure 11: Plot Described in Section 18.

Number of Failed and Successful Projects by Subcategory

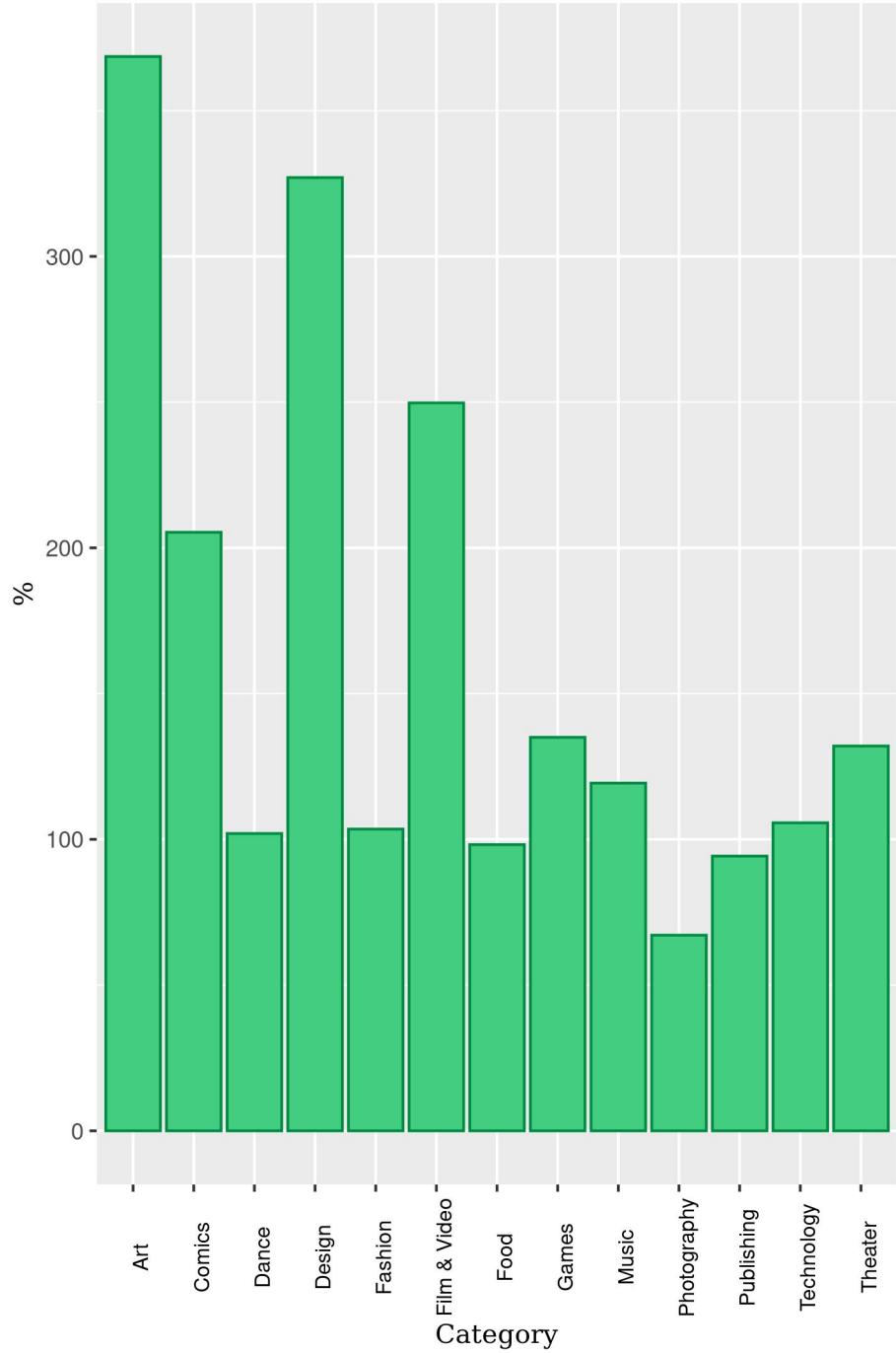


Figure 12: Plot Described in Section 19.

20. A bar chart showing numbers of successes and failure by subcategory:

```
ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset, aes(x=status)) +  
  geom_bar(fill = "seagreen3", color = "springgreen4") + facet_wrap(~subcategory) +  
  ggtitle("Number of Failed and Successful Projects by Subcategory") +  
  xlab("Subcategory") + ylab("Count") +  
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),  
        axis.text.x = element_text(color = "black", size = 8, angle = 90),  
        axis.title.x = element_text(color="black", size=10, family="Serif"),  
        axis.title.y = element_text(color="black", size=10, family="Serif"))  
ggsave(stat_subcat.jpg)
```

21. Another bar graph was produced showing average funded percentages by subcategory. This showed Conceptual Art as, by far, the highest average funded percentage. Other notably high sub categories are Graphic Design and Narrative Film:

```
ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset,  
       aes(x=subcategory, y=funded_percentage_asper)) +  
  ggtitle("Average Funded Percentage by Subcategory") +  
  xlab("Subcategory") + ylab("%") +  
  geom_bar(position = "dodge", stat = "summary", fun = "mean",  
           fill = "seagreen3", color = "springgreen4") +  
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),  
        axis.text.x = element_text(color = "black", size = 8, angle = 70),  
        axis.title.x = element_text(color="black", size=10, family="Serif"),  
        axis.title.y = element_text(color="black", size=10, family="Serif"))  
ggsave("subcat_perc_average.jpg")
```

22. Finally, I approached the question of ideal Months, Days, Times to launch a campaign by producing a series of bar graphs showing the numbers of success and failures depending on day of the week. This showed that projects are most likely to be launched on a week day, and least likely to be launched on a weekend, regardless of eventual success or failure:

```
ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset, aes(x=status)) +  
  geom_bar(fill = "seagreen3", color = "springgreen4") + facet_wrap(~day) +  
  ggtitle("Number Failed and Successful Projects by Day of the Week") +  
  xlab("Status") + ylab("Count") +  
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),  
        axis.text.x = element_text(color = "black", size = 8, angle = 90),  
        axis.title.x = element_text(color="black", size=10, family="Serif"),  
        axis.title.y = element_text(color="black", size=10, family="Serif"))  
ggsave("stat_day.jpg")  
  
ggplot(data=successful, aes(x=day)) + geom_bar()  
ggplot(data=failed, aes(x=day)) + geom_bar()
```

23. Another series of bar graphs were then produced showing the average funded percentage by day of the week. Projects launched on Fridays and Wednesdays, even with failures filtered out, had the highest funded percentage by far:

```
ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset,  
       aes(x=day, y=funded_percentage_asdec)) +  
  geom_bar(position = "dodge", stat = "summary", fun = "mean",  
           fill = "seagreen3", color = "springgreen4") +  
  ggtitle("Average Funded % of Projects by Day of the Week") +  
  xlab("Day") + ylab("%") +  
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
```

Number of Failed and Successful Projects by Subcategory



Figure 13: Plot Described in Section 20.

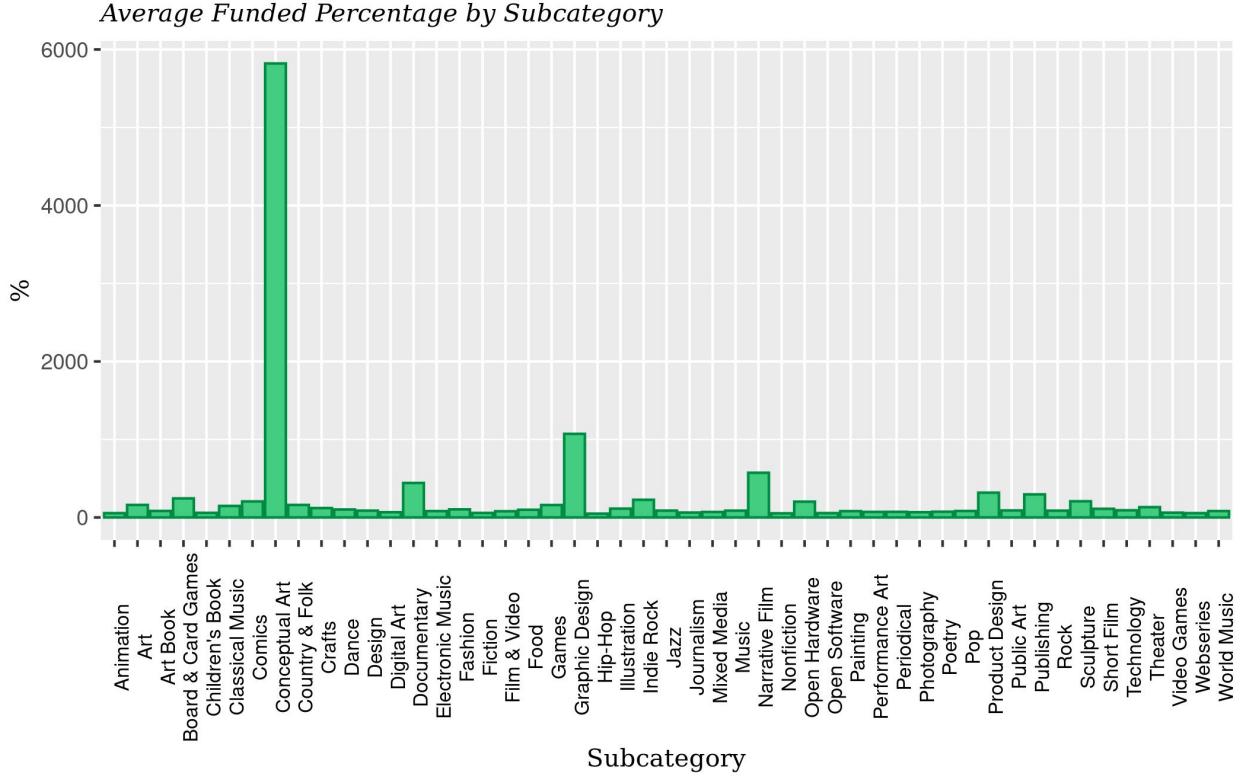


Figure 14: Plot Described in Section 21.

```

axis.title.x = element_text(color="black", size=10, family="Serif"),
axis.title.y = element_text(color="black", size=16, face="bold", family="Serif"))
ggsave("day_dec_avg.jpg")

ggplot(data=successful, aes(x=day, y=funded_percentage_asper)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean",
  fill = "seagreen3", color = "springgreen4") +
  ggtitle("Average Funded % of Successful Projects by Day of the Week") +
  xlab("Day") + ylab("%") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
  axis.title.x = element_text(color="black", size=10, family="Serif"),
  axis.title.y = element_text(color="black", size=16, face="bold", family="Serif"))
ggsave("day_dec_avg_success.jpg")

```

24. Average funded percentage of all projects is 184.8%, average funded percentage of projects launched on Wednesdays is 283.15%, and average funded percentage of projects launched on Fridays is 334%. Average funded percentage of all successful projects is 353.2%, average funded percentage of successful projects launched on Wednesdays is 540.4%, and average funded percentage of successful projects launched on Fridays is 672.5%. These are remarkable deviations from the average:

```

wednesday <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>% filter(day == "Wed")
friday <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>% filter(day == "Fri")
mean(kickstarter_casestudy_DSI_kickstarterscrape_dataset$funded_percentage_asper)
mean(wednesday$funded_percentage_asper)
mean(friday$funded_percentage_asper)
wednesday_nofail <- successful %>% filter(day == "Wed")
friday_nofail <- successful %>% filter(day == "Fri")

```

Number Failed and Successful Projects by Day of the Week

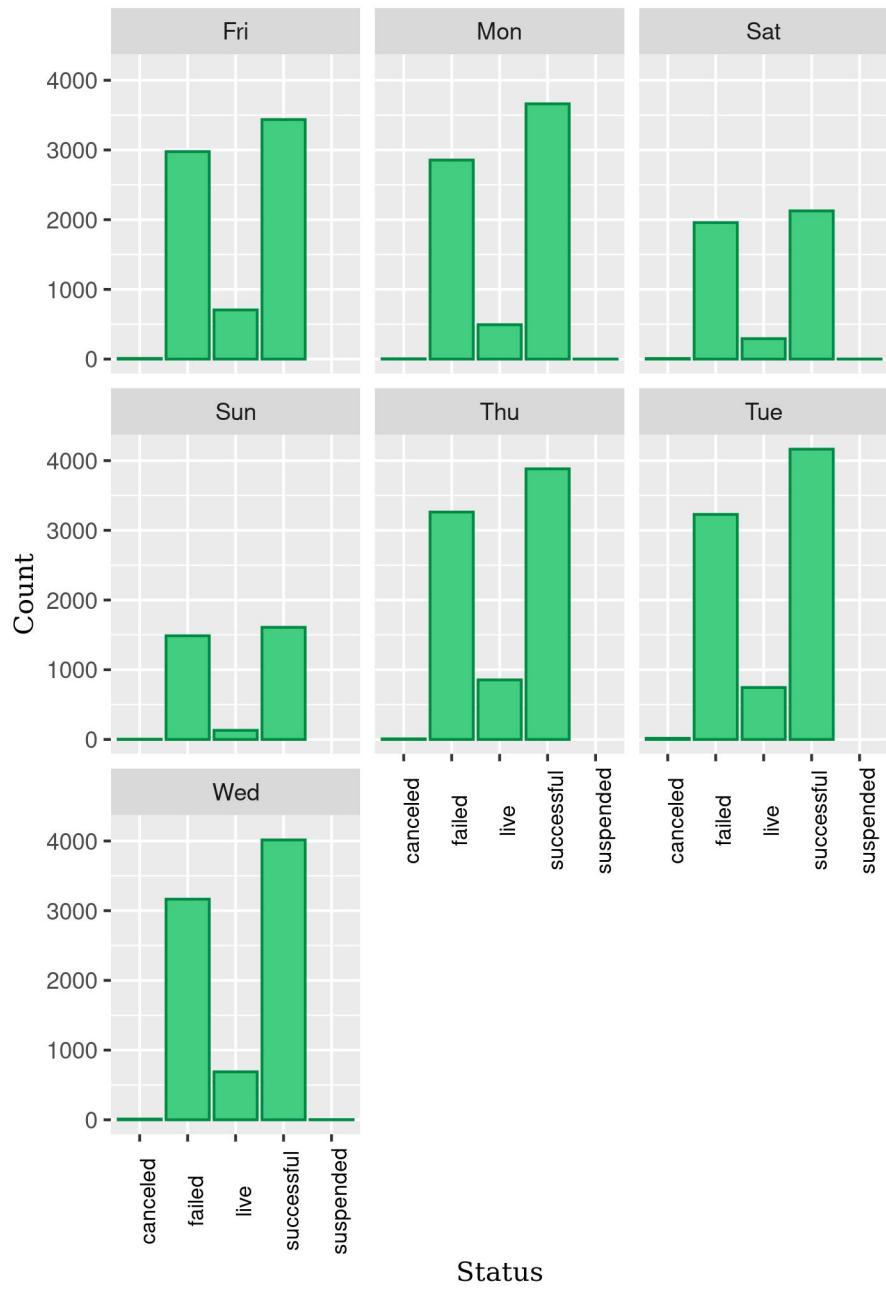


Figure 15: Plot Described in Section 22.

Average Funded % of Projects by Day of the Week

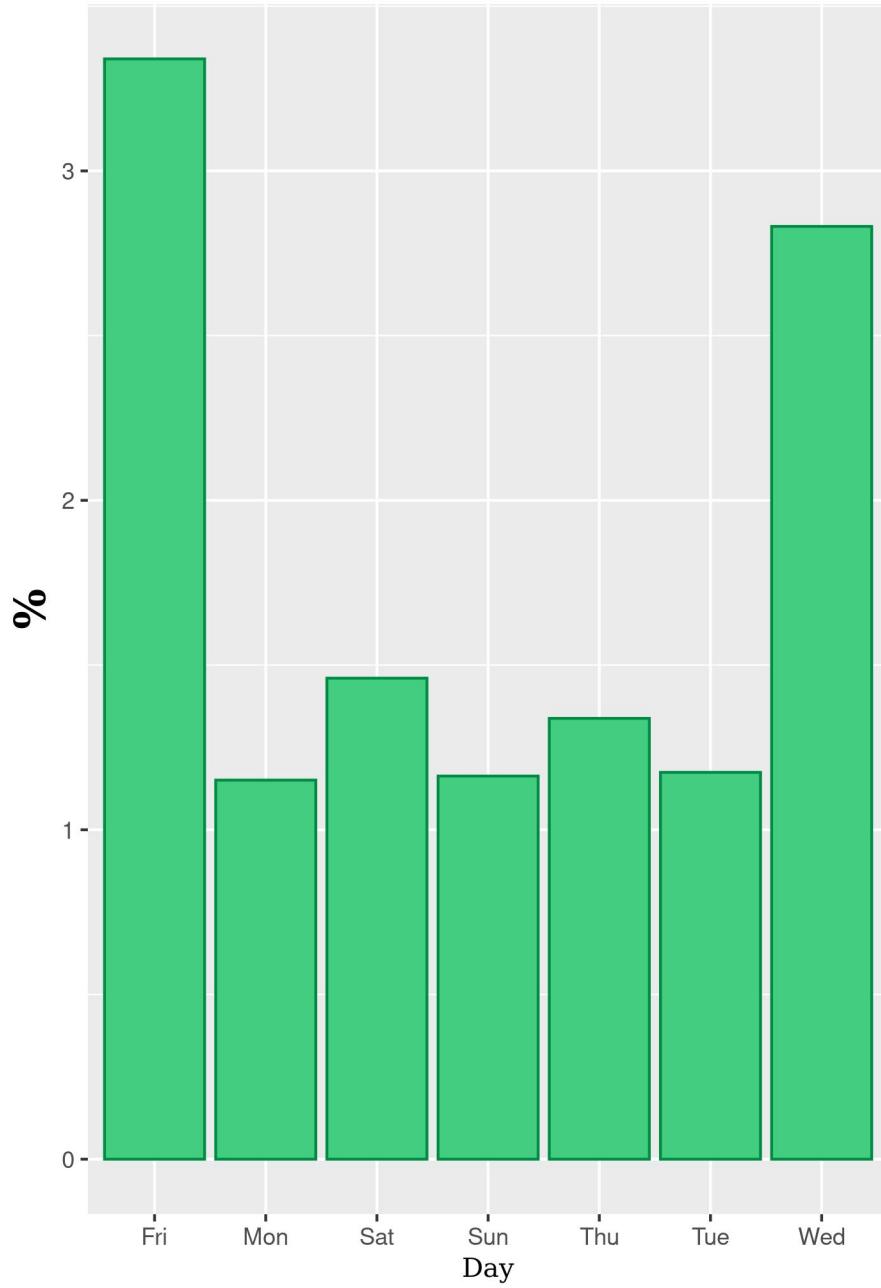


Figure 16: First Plot Described in Section 23.

Average Funded % of Successful Projects by Day of the Week

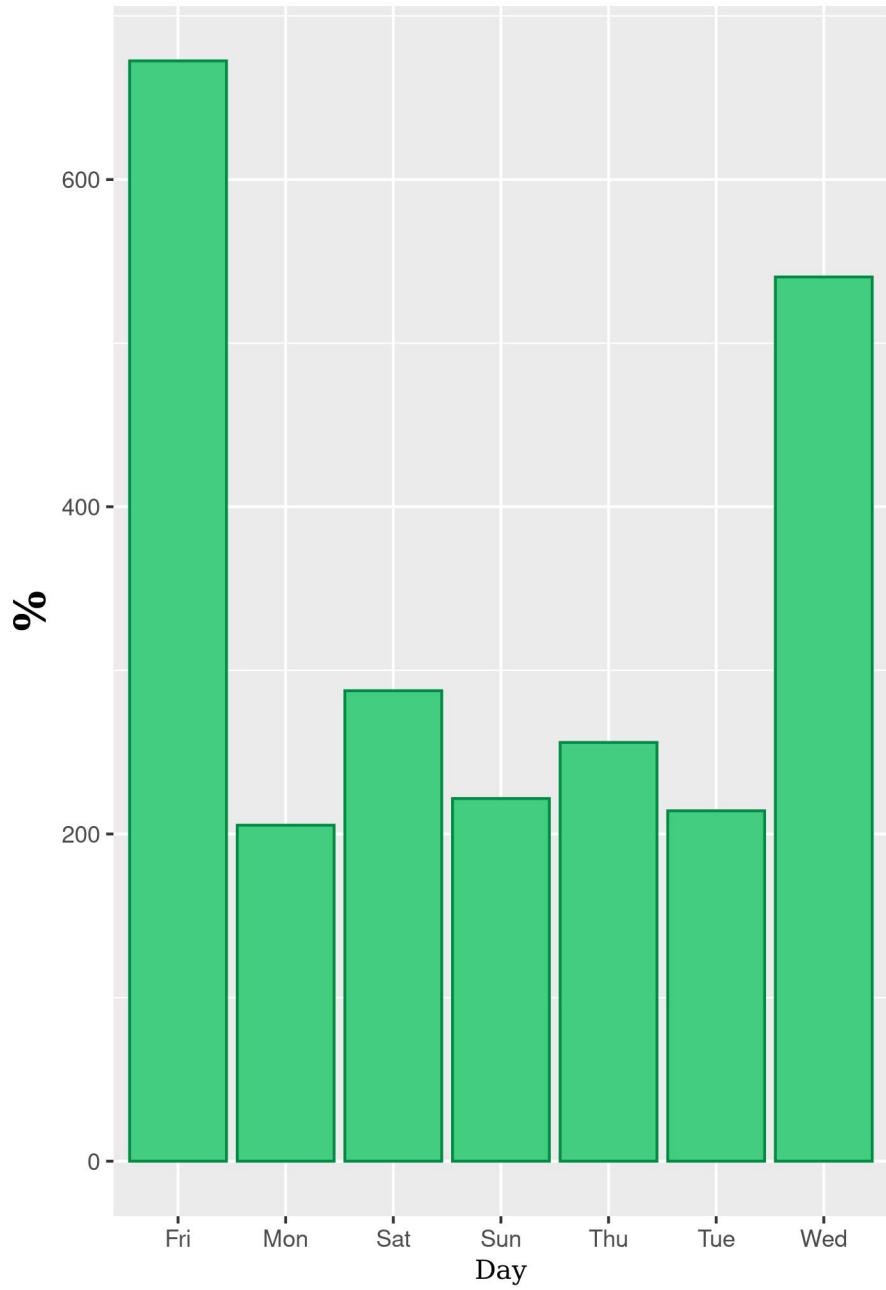


Figure 17: Second Plot Described in Section 23.

```

mean(successful$funded_percentage_asper)
mean(wednesday_nofail$funded_percentage_asper)
mean(friday_nofail$funded_percentage_asper)

```

25. A bar graph showing the count of success, live projects, and failures depending on month was produced and this mostly suggested that information was web scraped in early to late June as April, May, and June, show the most live and cancelled projects.

```

ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset, aes(x=status)) +
  geom_bar(fill = "seagreen3", color = "springgreen4") + facet_wrap(~month) +
  ggtitle("Number of Failed and Successful Projects by Month") +
  xlab("Month") + ylab("Count") +
  theme(plot.title = element_text(color="black", size=10, face="italic", family="Serif"),
        axis.text.x = element_text(color = "black", size = 8, angle = 90),
        axis.title.x = element_text(color="black", size=10, family="Serif"),
        axis.title.y = element_text(color="black", size=16, family="Serif"))
  ggsave("stat_month.jpg")

```

26. Another series of bar graphs showed averages of funded percentage by Month and this revealed September, August, May, and July to have the lowest funding percentages, while November and October show the highest funding percentages by far, even with failed projects filtered out:

```

ggplot(data=kickstarter_casestudy_DSI_kickstarterscrape_dataset,
       aes(x=month, y=funded_percentage_asper)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean",
           fill = "seagreen3", color = "springgreen4") +
  ggtitle("Average Funded % of Projects by Month") +
  xlab("Month") + ylab("%") +
  theme(
    plot.title = element_text(color="springgreen4", size=14, face="bold.italic", family="Serif"),
    axis.title.x = element_text(color="black", size=12, family="Serif"),
    axis.title.y = element_text(color="black", size=16, face="bold", family="Serif"))
  ggsave("month_dec_avg.jpg")

ggplot(data=successful, aes(x=month, y=funded_percentage_asper)) +
  geom_bar(position = "dodge", stat = "summary", fun = "mean",
           fill = "seagreen3", color = "springgreen4") +
  ggtitle("Average Funded % of Successful Projects by Month") +
  xlab("Month") + ylab("%") +
  theme(
    plot.title = element_text(color="springgreen4", size=14, face="bold.italic", family="Serif"),
    axis.title.x = element_text(color="black", size=12, family="Serif"),
    axis.title.y = element_text(color="black", size=16, face="bold", family="Serif"))
  ggsave("month_dec_avg_success.jpg")

```

27. The average funded percentage of all projects is 184.8%, the average funded percentage of all successful projects is 353.2%, and the average percentage of all October projects, successful or failed, is 596.7%. This too is a remarkably outstanding deviation from the average:

```

october <- kickstarter_casestudy_DSI_kickstarterscrape_dataset %>% filter(month == "Oct")
mean(kickstarter_casestudy_DSI_kickstarterscrape_dataset$funded_percentage_asper)
mean(successful$funded_percentage_asper)
mean(october$funded_percentage_asper)

```

Number of Failed and Successful Projects by Month



Figure 18: Plot Described in Section 25.

Average Funded % of Projects by Month

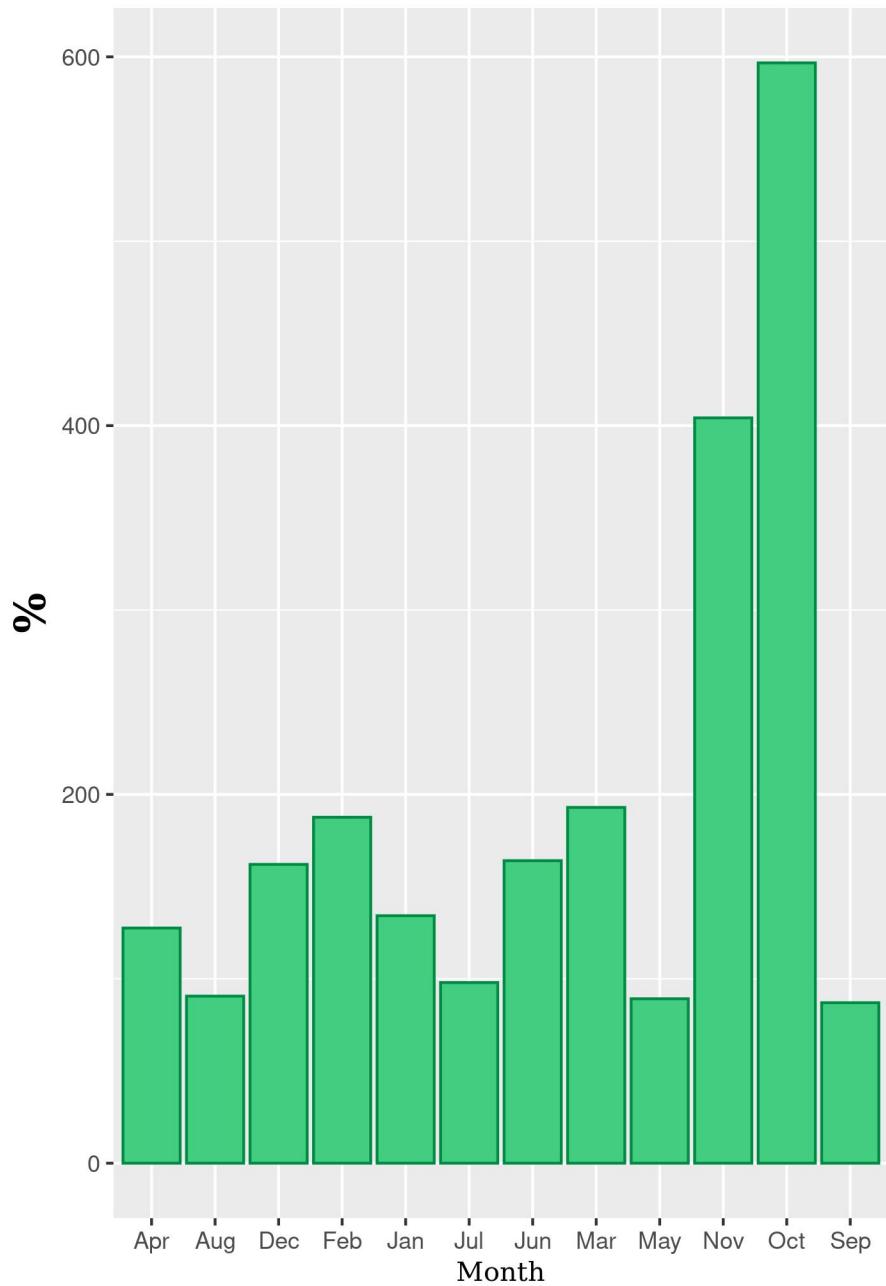


Figure 19: First Plot Described in Section 26.

Average Funded % of Successful Projects by Month

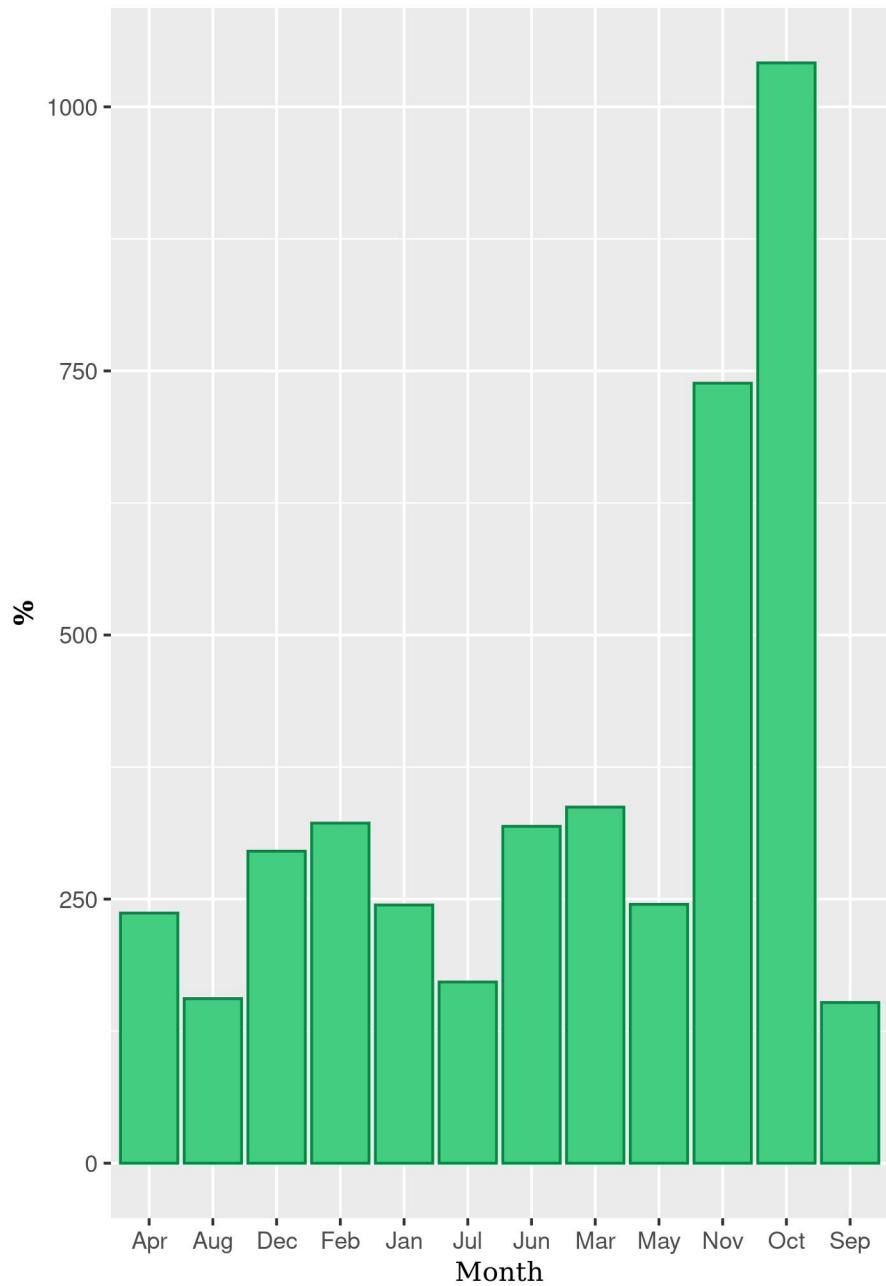


Figure 20: Second Plot Described in Section 26.