data_prep

```
#info 370 final project
library(readr)
library(dplyr)
library(ggplot2)
df_update<-read_csv("./facebook_with_reactions.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
df<-df_update %>% select(Category,Page,`Date Published`,`Post Type`,Rating,
                         share_count,reaction_count,comment_count,
                         status_message,link_name,status_type,status_published,
                         num_reactions,num_comments,num_shares,num_loves,
                         num_likes,num_wows,num_hahas,num_sads,num_angrys)

#how do we deal with NA values? found in many cells, can't just omit them
#checking: sum(is.na(df_numeric$share_count))

#select only the numerical cols for now
df_numeric<-df %>% select(share_count,reaction_count,comment_count,
                          num_reactions,num_comments,num_shares,num_loves,
                          num_likes,num_wows,num_hahas,num_sads,num_angrys)

reaction_numeric<-matrix(0,nrow=nrow(df))
for(i in 1:nrow(df)){
  if(df$Rating[i]=="no factual content"){
    reaction_numeric[i]=1
  }
  if(df$Rating[i]=="mostly false"){
    reaction_numeric[i]=0
  }
  if(df$Rating[i]=="mixture of true and false"){
    reaction_numeric[i]=2
  }
  if(df$Rating[i]=="mostly true"){
    reaction_numeric[i]=3
  }
}

df_numeric=cbind(df_numeric,reaction_numeric)
##############################done with df
```

Tables

```
#see how many in each group
df_numeric %>% group_by(reaction_numeric) %>% summarize(n())
```

```
## # A tibble: 4 x 2
##   reaction_numeric `n()`
##              <dbl> <int>
```

```
## 1                0    78
## 2                1   257
## 3                2   219
## 4                3  1612
```

```
df %>% group_by(`Post Type`) %>% summarize(n())
```

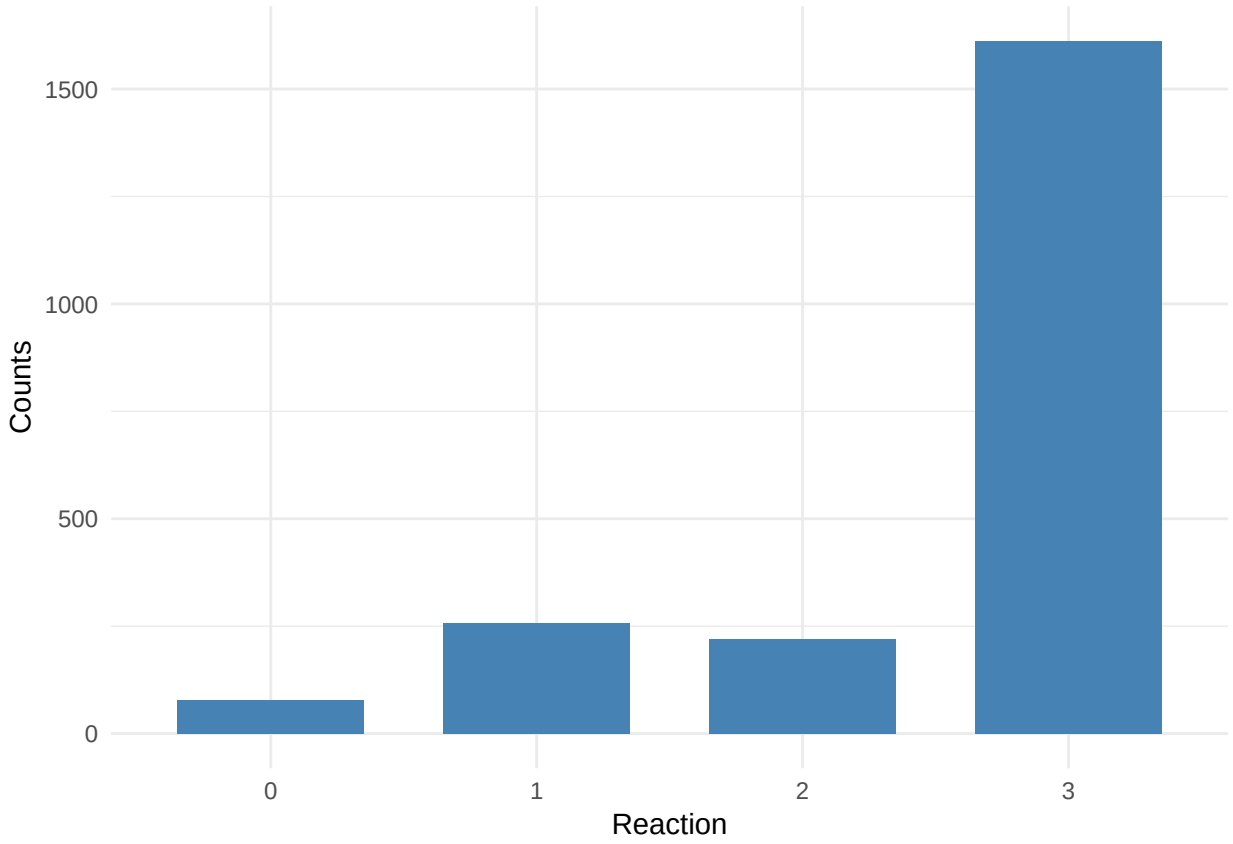
```
## # A tibble: 4 x 2
##   `Post Type` `n()`
##   <chr>       <int>
## 1 link         1669
## 2 photo         206
## 3 text            3
## 4 video         288
```

```
df %>% group_by(status_type) %>% summarize(n())
```

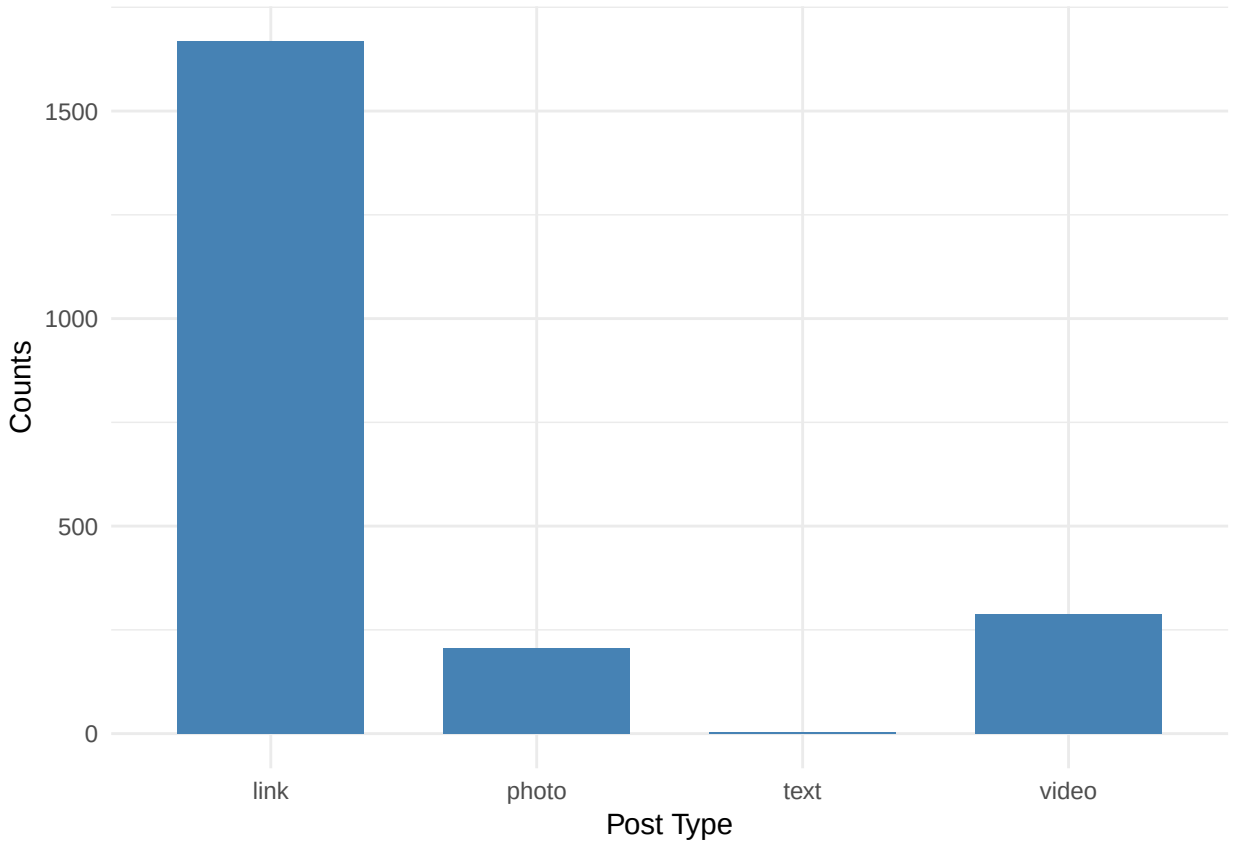
```
## # A tibble: 4 x 2
##   status_type `n()`
##   <chr>       <int>
## 1 link         1645
## 2 photo         212
## 3 status         12
## 4 video         297
```

Plots

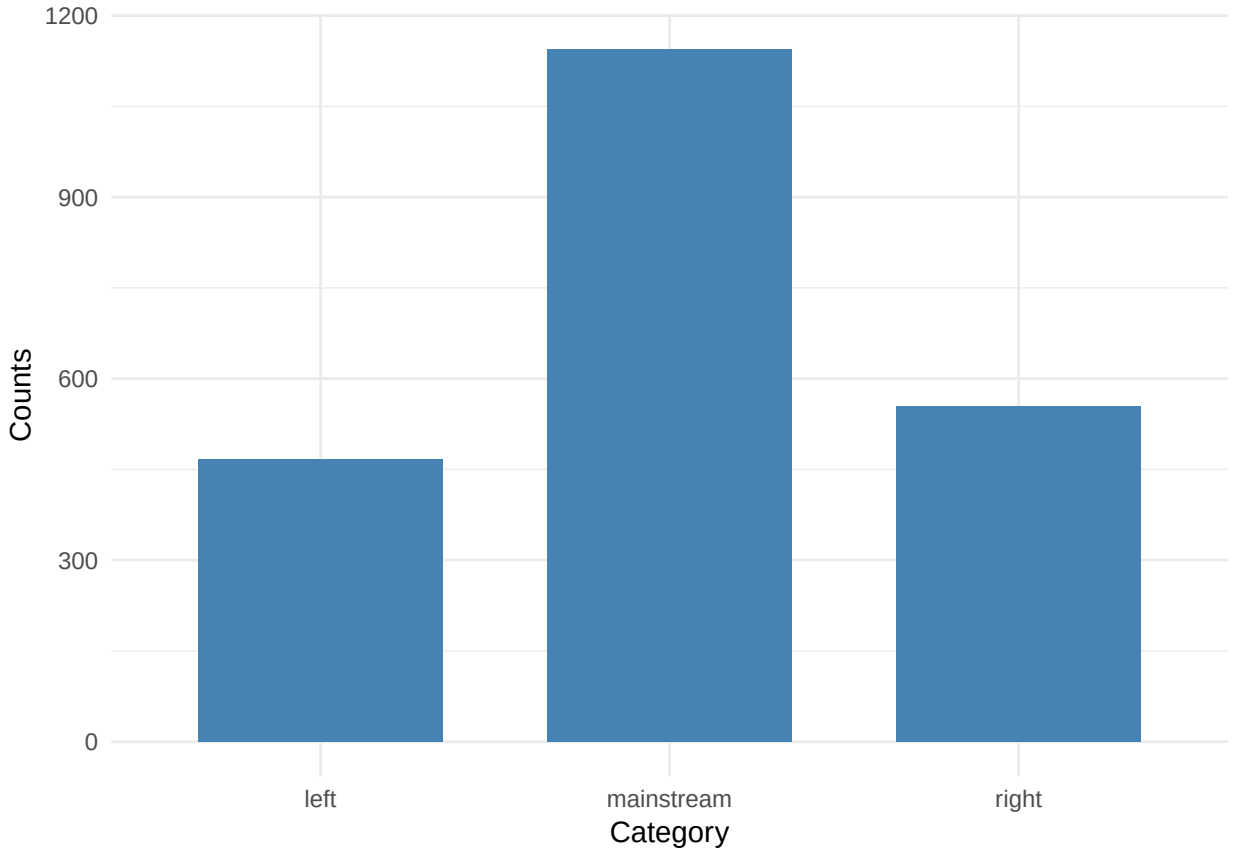
```
p<-ggplot(df_numeric, aes(x=factor(reaction_numeric)))+
  labs(x="Reaction",y="Counts")+
  geom_bar(stat="count", width=0.7, fill="steelblue")+
  theme_minimal() #barplot of reaction
p
```



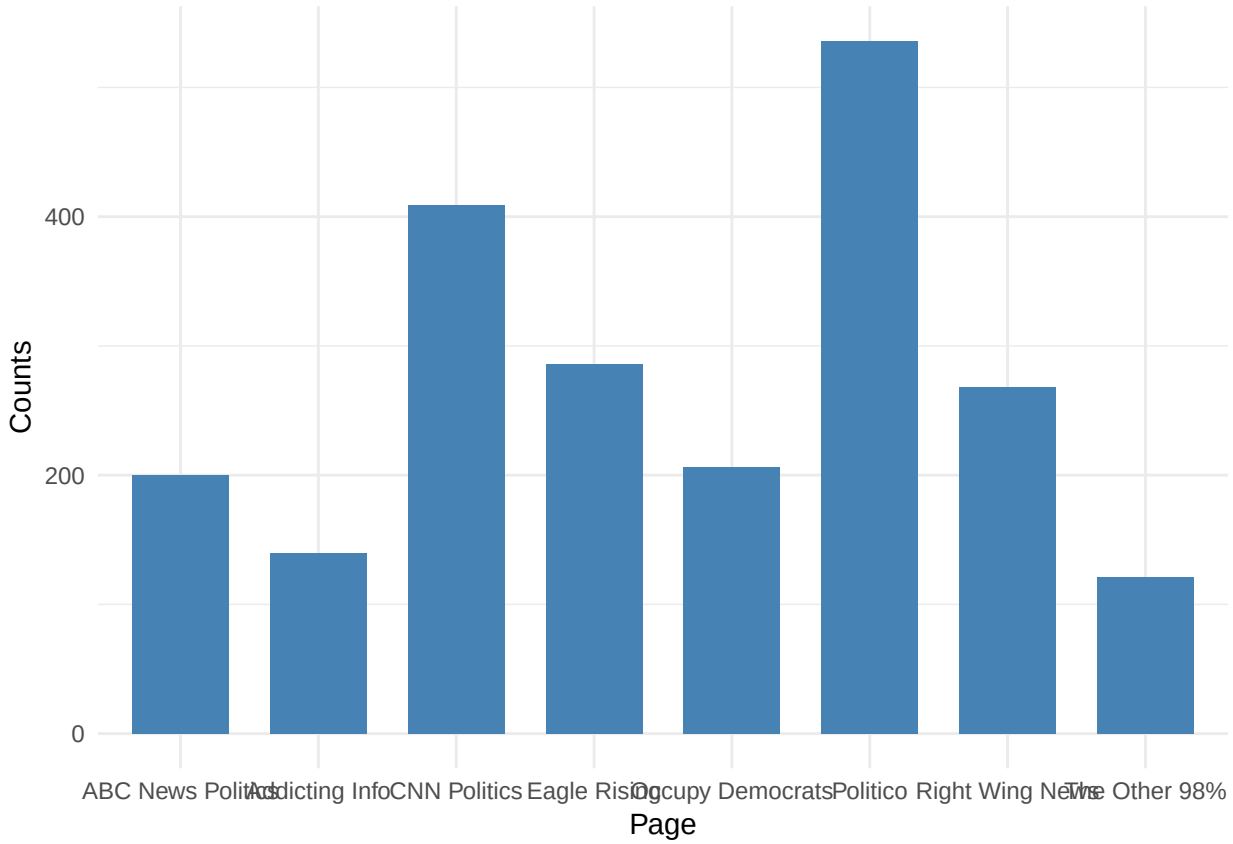

```
p<-ggplot(df_update, aes(x=factor(`Post Type`)))+
  theme(text=element_text(size=0.1))+
  labs(x="Post Type",y="Counts")+
  geom_bar(stat="count", width=0.7, fill="steelblue")+
  theme_minimal() #barplot of page
p
```



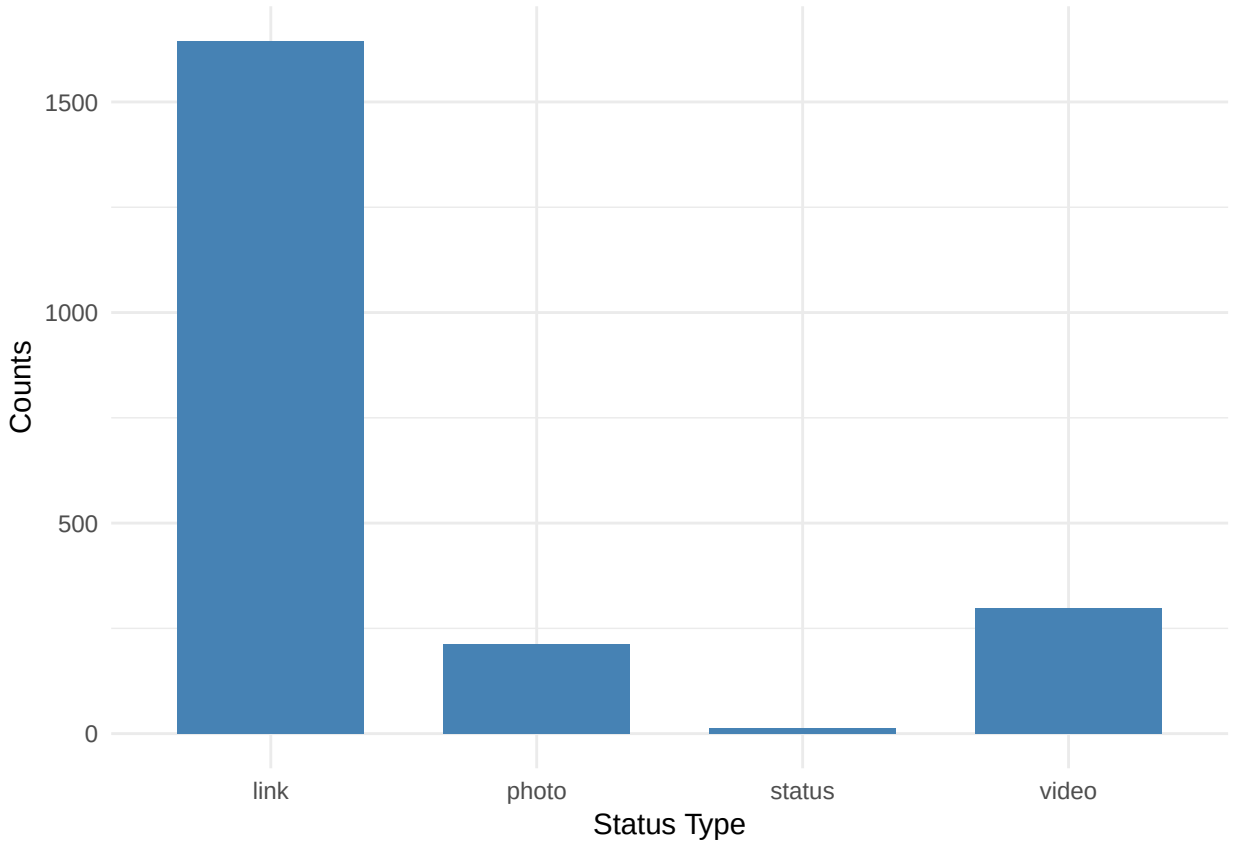

```
p<-ggplot(df_update, aes(x=factor(Category)))+
  theme(text=element_text(size=0.1))+
  labs(x="Category",y="Counts")+
  geom_bar(stat="count", width=0.7, fill="steelblue")+
  theme_minimal() #barplot of status type
p
```




```
p<-ggplot(df_update, aes(x=factor(Page)))+
  theme(text=element_text(size=0.1))+
  labs(x="Page",y="Counts")+
  geom_bar(stat="count", width=0.7, fill="steelblue")+
  theme_minimal() #barplot of page
p
```

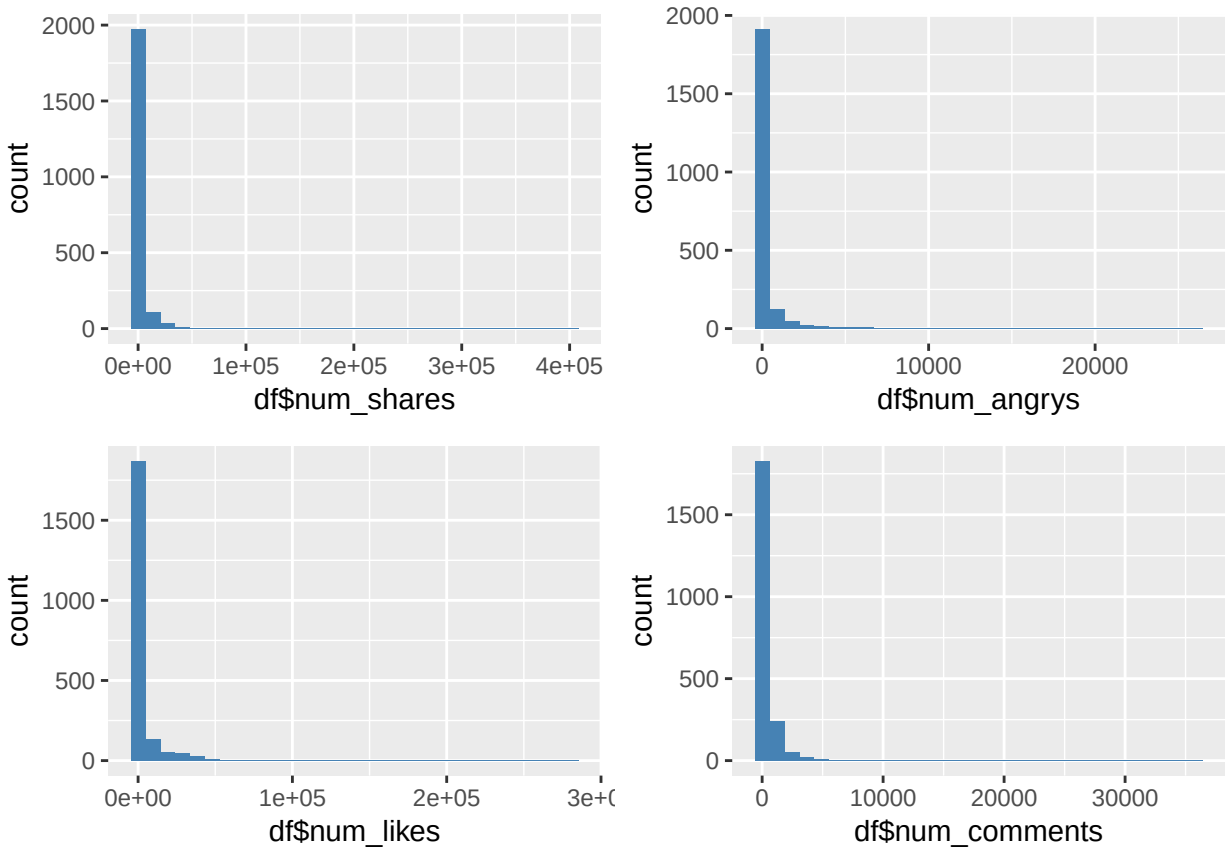



```
p<-ggplot(df_update, aes(x=factor(status_type)))+
  theme(text=element_text(size=0.1))+
  labs(x="Status Type",y="Counts")+
  geom_bar(stat="count", width=0.7, fill="steelblue")+
  theme_minimal() #barplot of status type
p
```



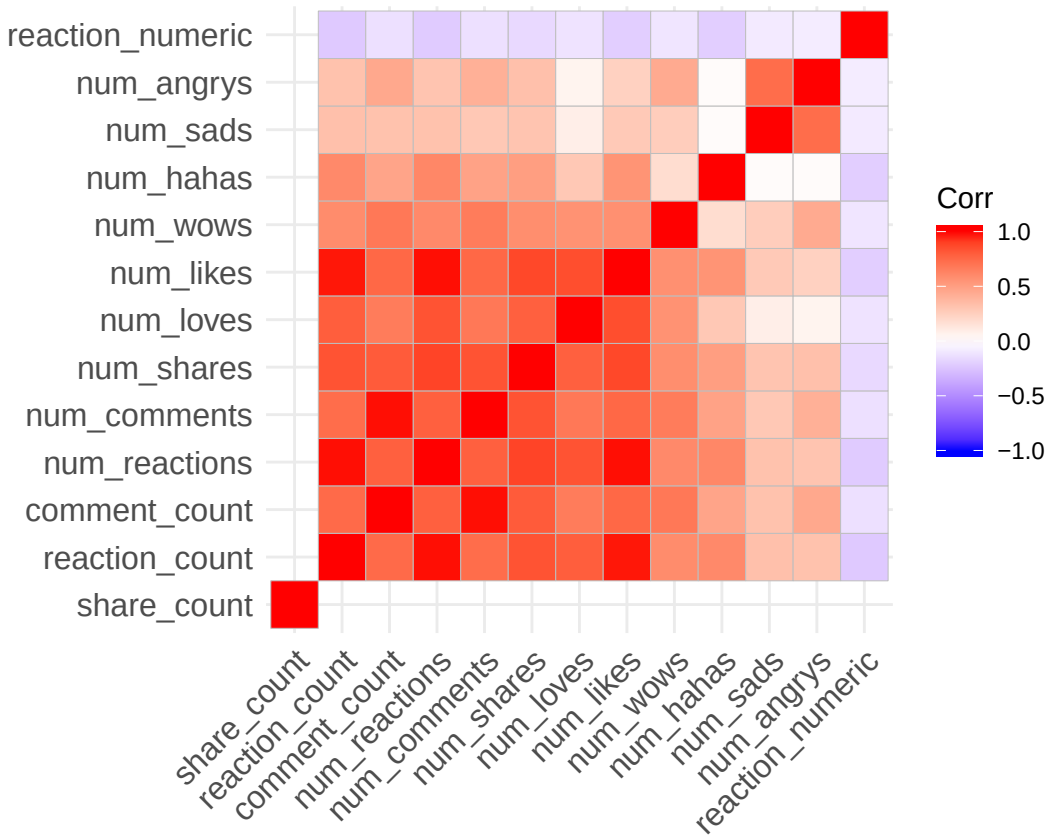
Histogram

```
library(gridExtra)
p1<-ggplot(data=df,aes(df$num_shares)) + geom_histogram(fill="steelblue")
p2<-ggplot(data=df,aes(df$num_angrys)) + geom_histogram(fill="steelblue")
p3<-ggplot(data=df,aes(df$num_likes)) + geom_histogram(fill="steelblue")
p4<-ggplot(data=df,aes(df$num_comments)) + geom_histogram(fill="steelblue")
grid.arrange(p1,p2,p3,p4,nrow=2)
```



Correlation Heat Graph

```
library(ggcorrplot)
cormat<-round(cor(df_numeric),2)
ggcorrplot(cormat)
```


reaction_numeric
num_angrys
num_sads
num_hahas
num_wows
num_likes
num_loves
num_shares
num_comments
num_reactions
comment_count
reaction_count
share_count
share_count
reaction_count
comment_count
num_reactions
num_comments
num_shares
num_loves
num_likes
num_wows
num_hahas
num_sads
num_angrys
reaction_numeric
Corr
1.0
0.5
0.0
−0.5
−1.0