Project-3

Dongyu Wang

2023-10-25

Project Deliverable 3 Q3

Outcome Page.total.likes, main predictor Type

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
library(RNOmni)
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```
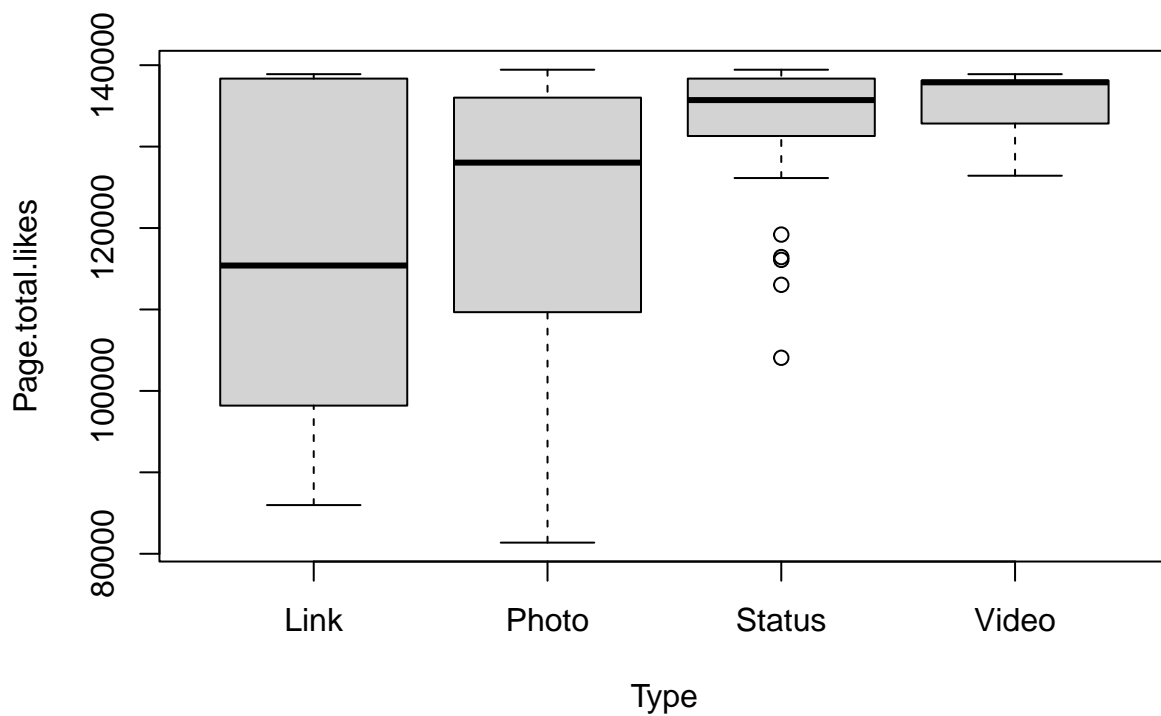
```
setwd("C:/Users/wdy24/Dropbox/BU_PHD/Coursework/BU MA series/MA575/Lab/Project/Code_dataset")
fb <- read.csv(file="dataset_Facebook(3).csv",header = T,sep = ";")
```

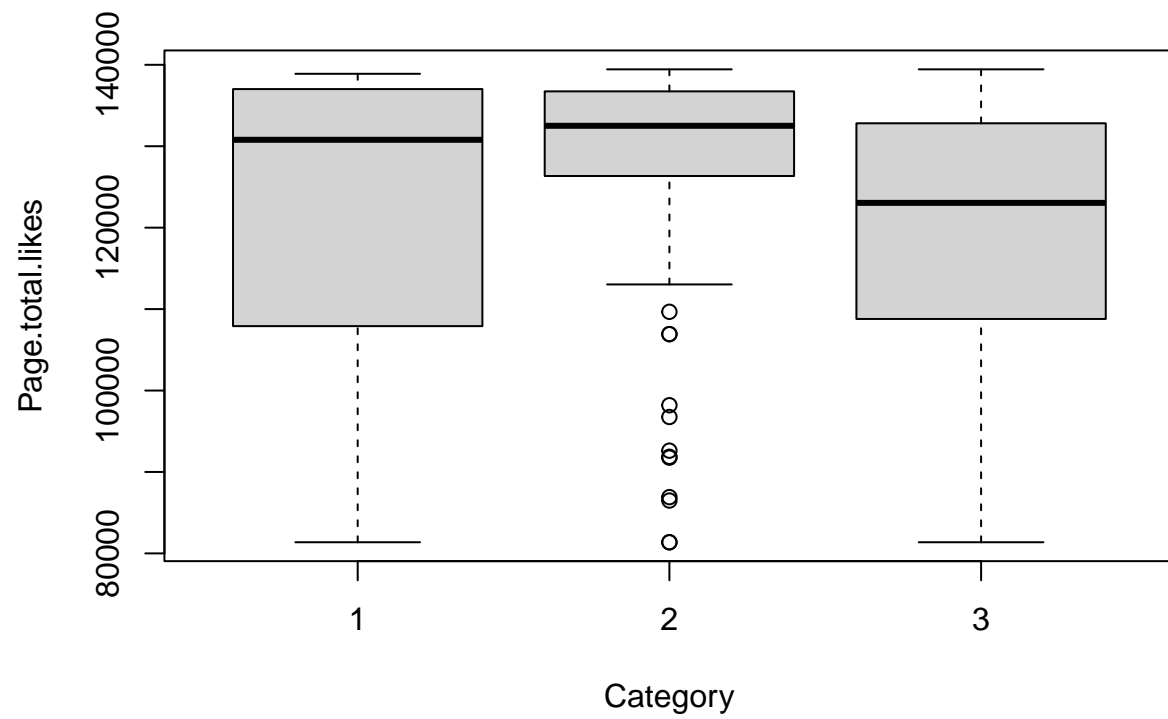## Descriptive analysis

```
attach(fb)
table(fb$Type, useNA = 'always')
```

```
##
##   Link  Photo Status  Video   <NA>
##     22    426     45      7      0
```
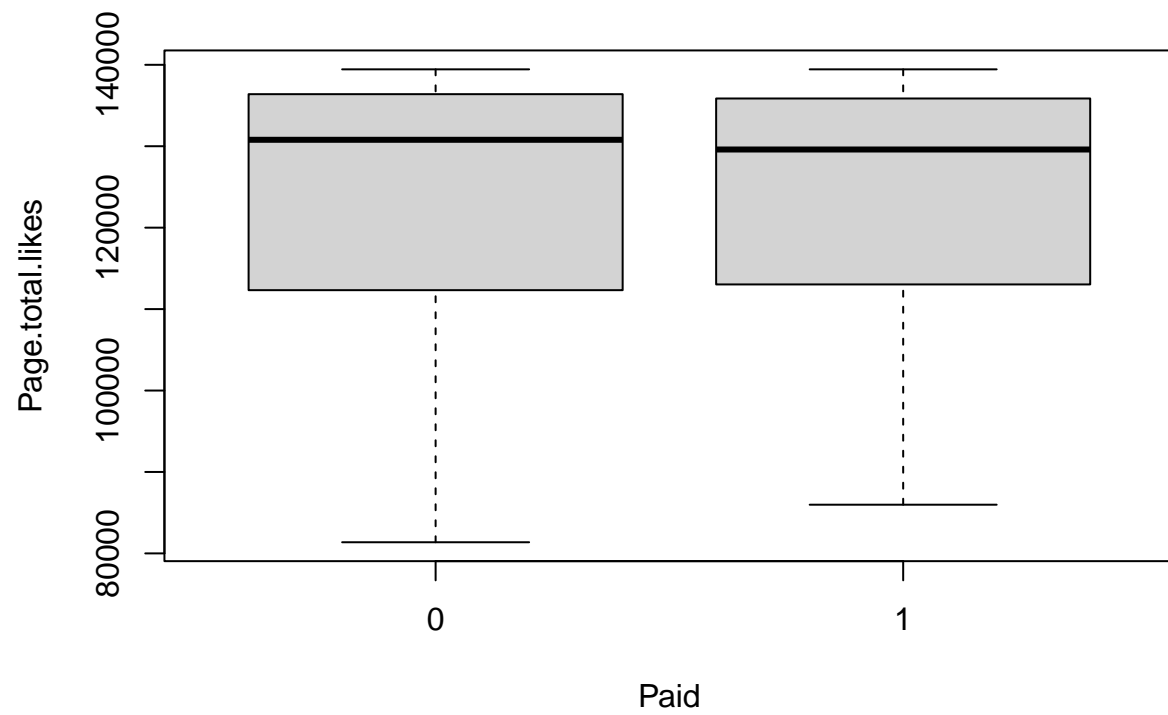
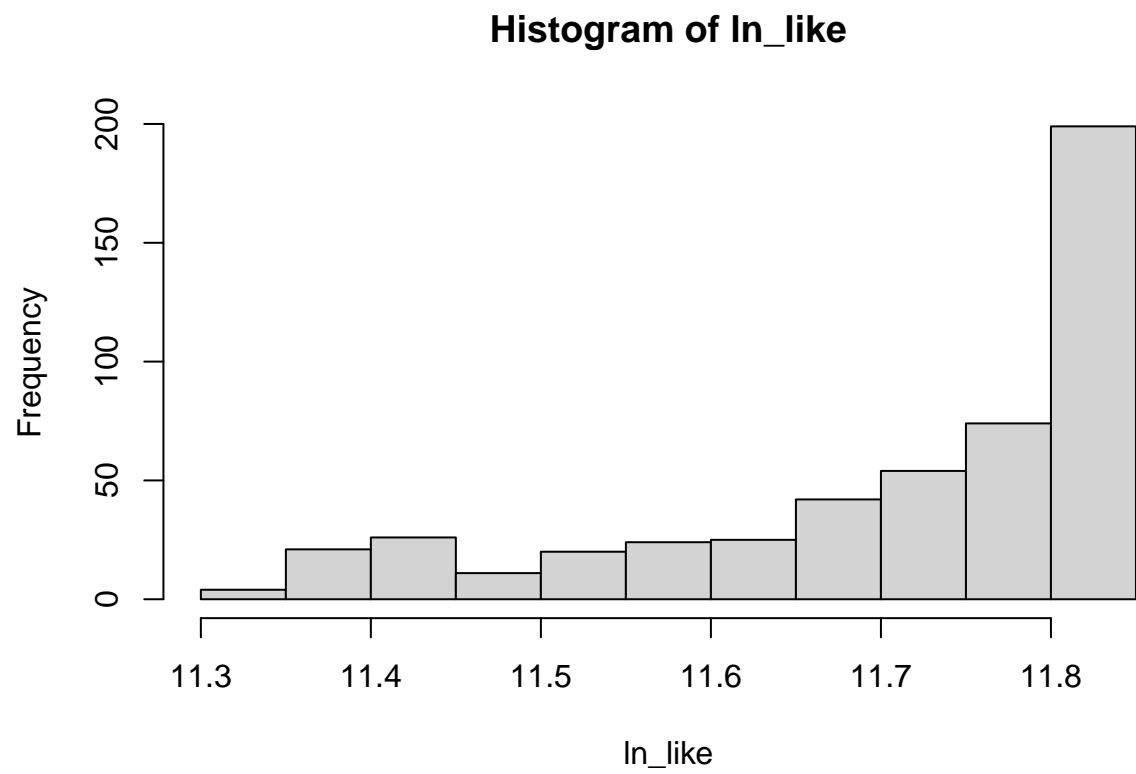```
boxplot(Page.total.likes ~ Type)
```



```
boxplot(Page.total.likes ~ Category)
```
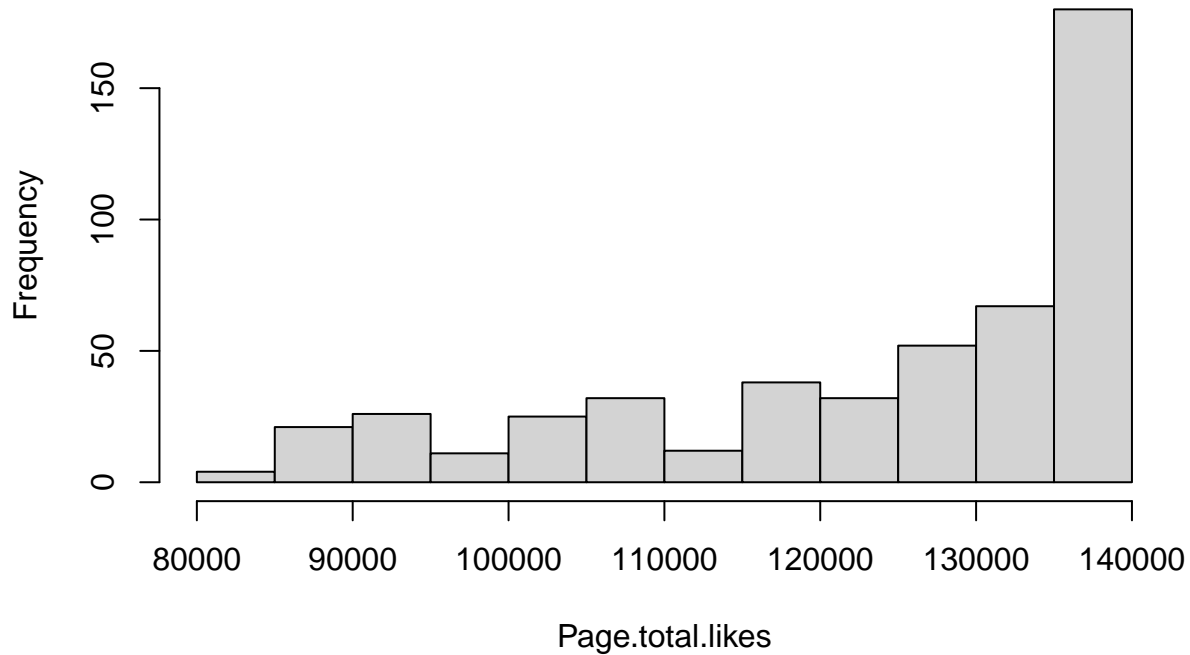
```
boxplot(Page.total.likes ~ Paid)
```

```r
ln_like <- log(Page.total.likes)
hist(ln_like)
```

Histogram of ln_like

```
hist(Page.total.likes)
```

Histogram of Page.total.likes


```r
tapply(Page.total.likes,Type, summary)
```

```
## $Link
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   85979   98829  115396  116363  138059  138895
##
## $Photo
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   81370  109670  128032  122354  136013  139441
##
## $Status
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  104070  131300  135713  132647  138353  139441
##
## $Video
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  126424  132835  137893  135015  138111  138895
```
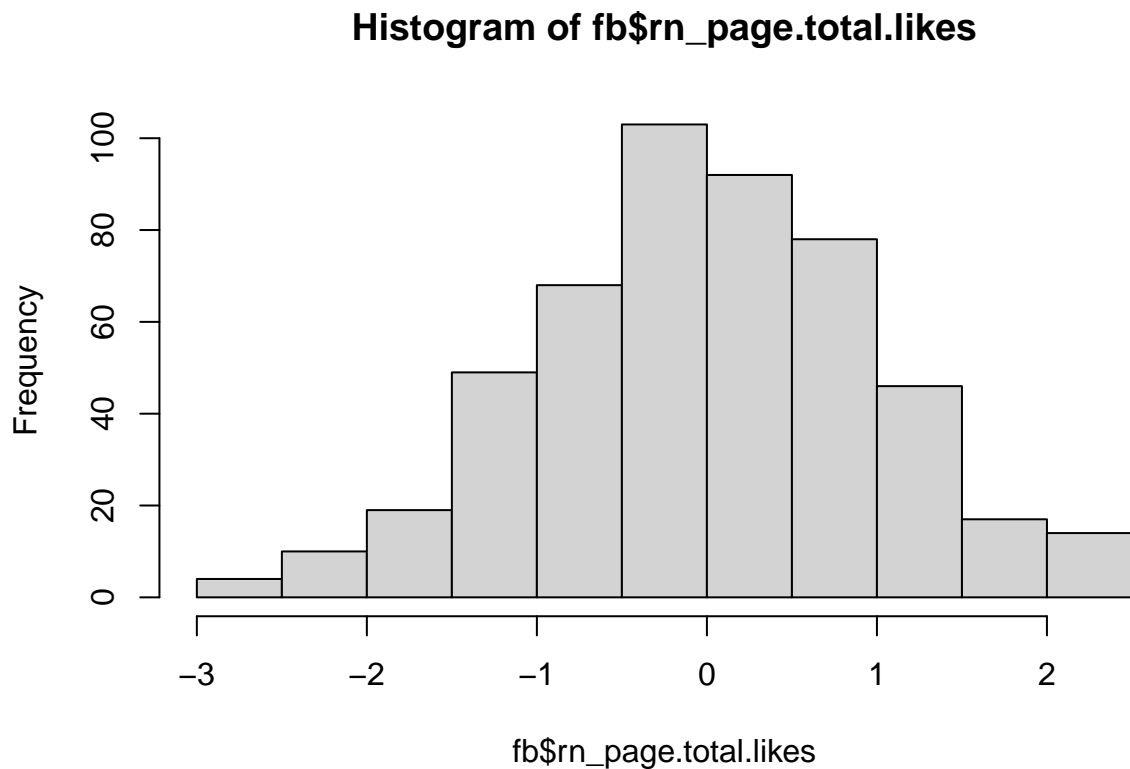
```r
fb$rn_page.total.likes <- RankNorm(Page.total.likes,ties.method = "average")
summary(fb$rn_page.total.likes)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.631705 -0.664782 -0.022550 -0.002038  0.655348  2.190531
```

```
sd(fb$rn_page.total.likes)
```

```
## [1] 0.9878953
```

```
hist(fb$rn_page.total.likes)
```

Histogram of fb$rn_page.total.likes


```
attach(fb)
```

```
## The following objects are masked from fb (pos = 3):
##
##     Category, comment, Lifetime.Engaged.Users,
##     Lifetime.People.who.have.liked.your.Page.and.engaged.with.your.post,
##     Lifetime.Post.Consumers, Lifetime.Post.Consumptions,
##     Lifetime.Post.Impressions.by.people.who.have.liked.your.Page,
##     Lifetime.Post.reach.by.people.who.like.your.Page,
##     Lifetime.Post.Total.Impressions, Lifetime.Post.Total.Reach, like,
##     Page.total.likes, Paid, Post.Hour, Post.Month, Post.Weekday, share,
##     Total.Interactions, Type
```
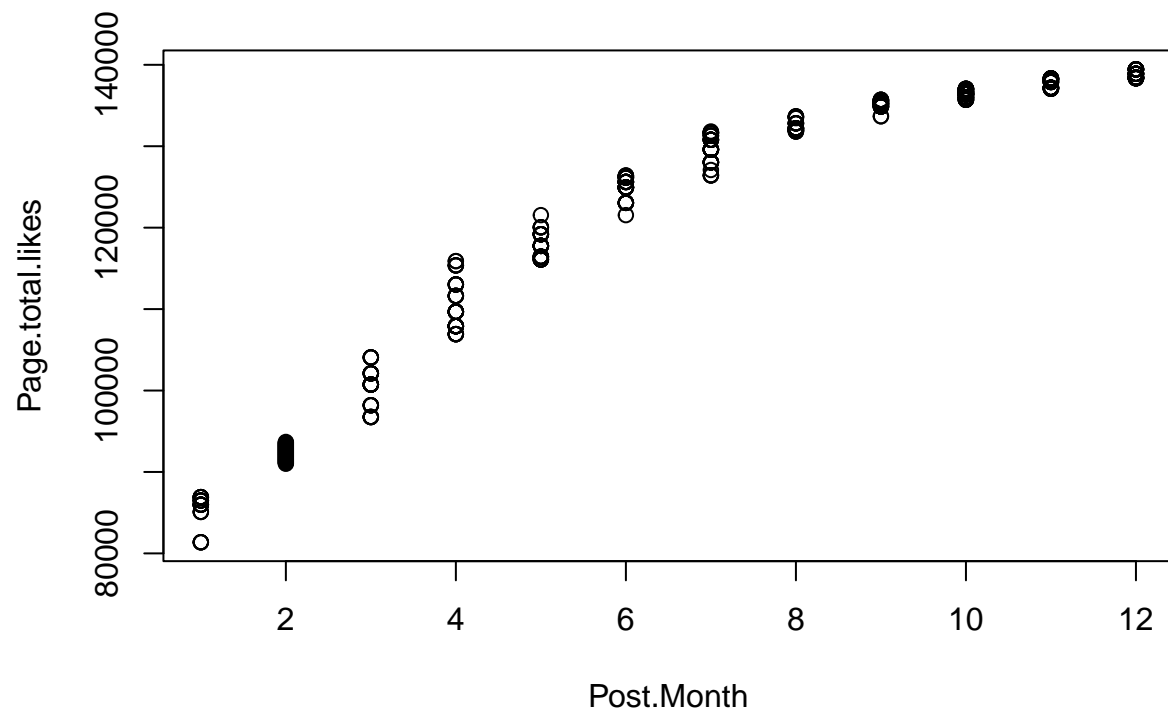
```
## Page.total.likes highly skewed toward the left - need to transform but first proceed with the MV mode
## Inital variable selection from scatter plot + heatmap
## Type Post.Month Post.Weekday Post.Hour Total.Interactions Lifetime.Post.Consumers
```
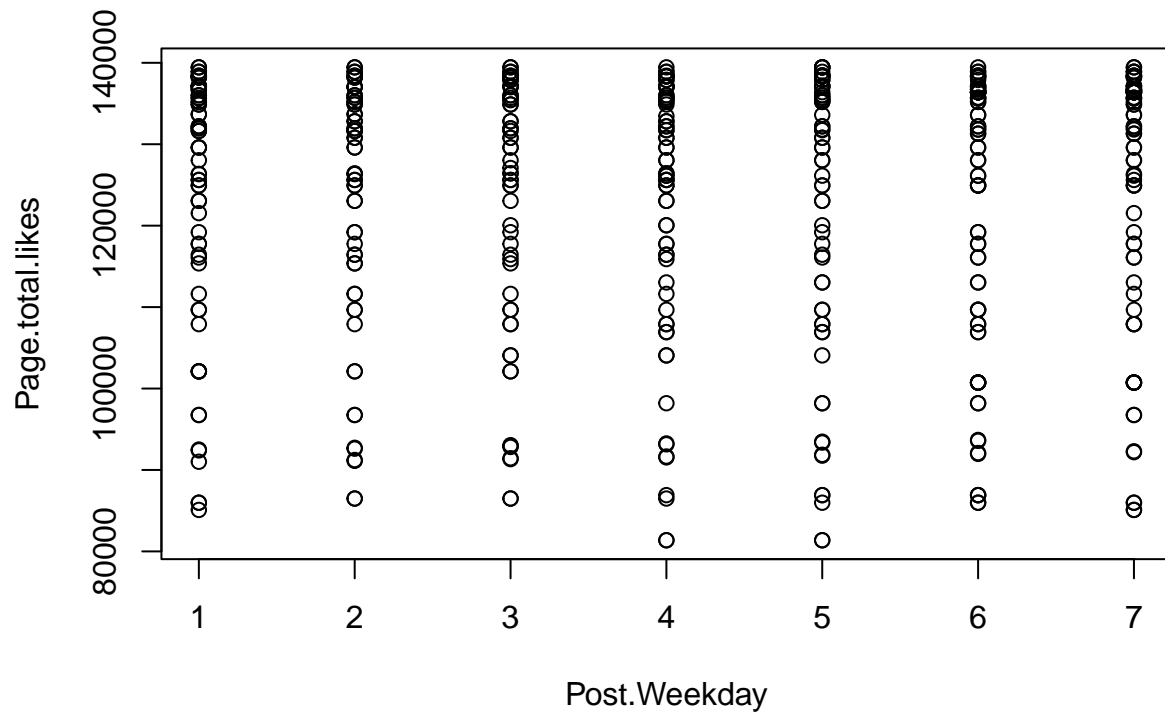
```
plot(Page.total.likes ~ Post.Month)
```
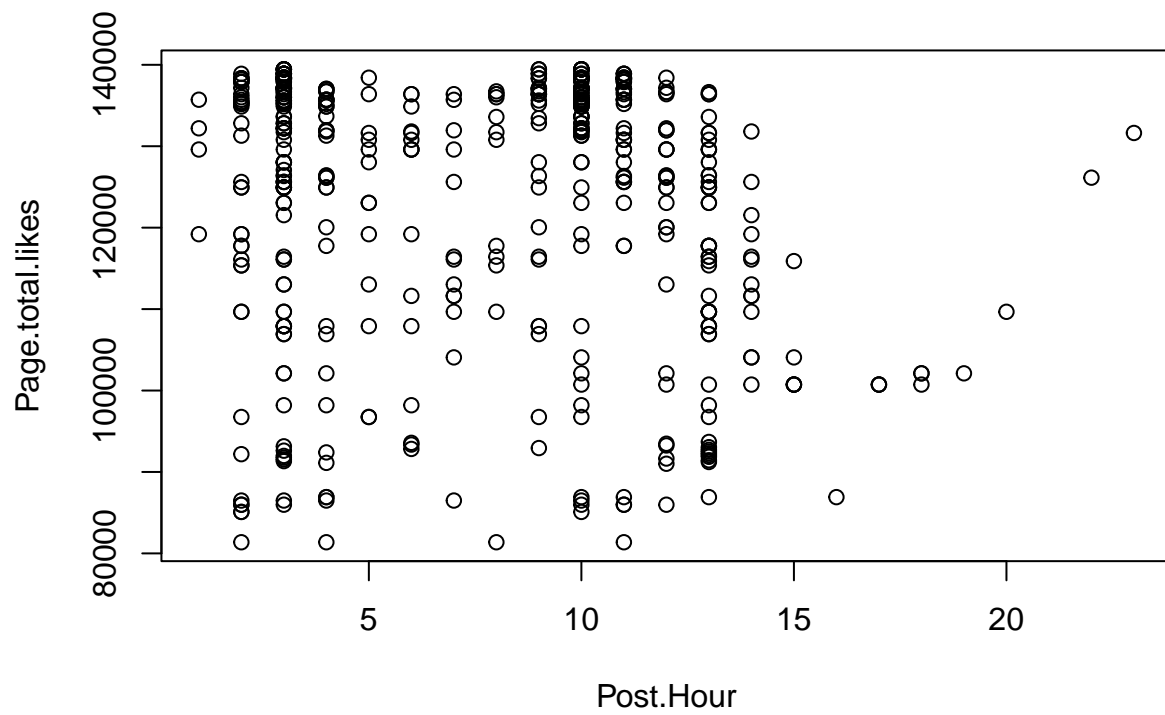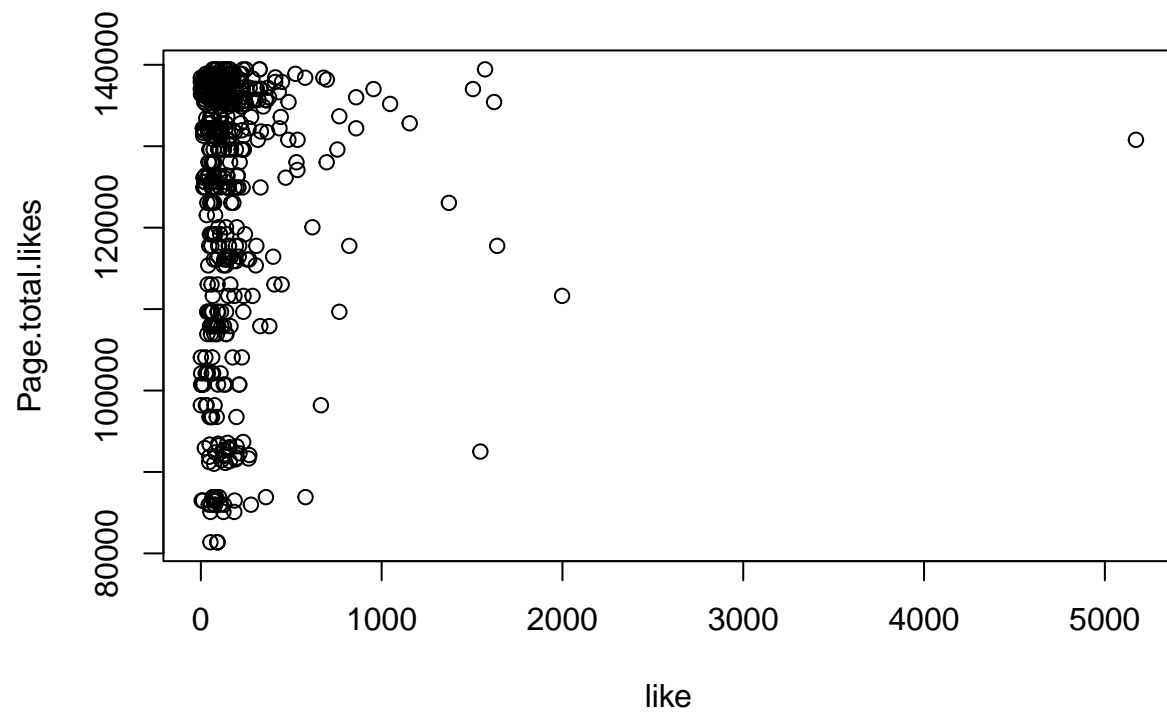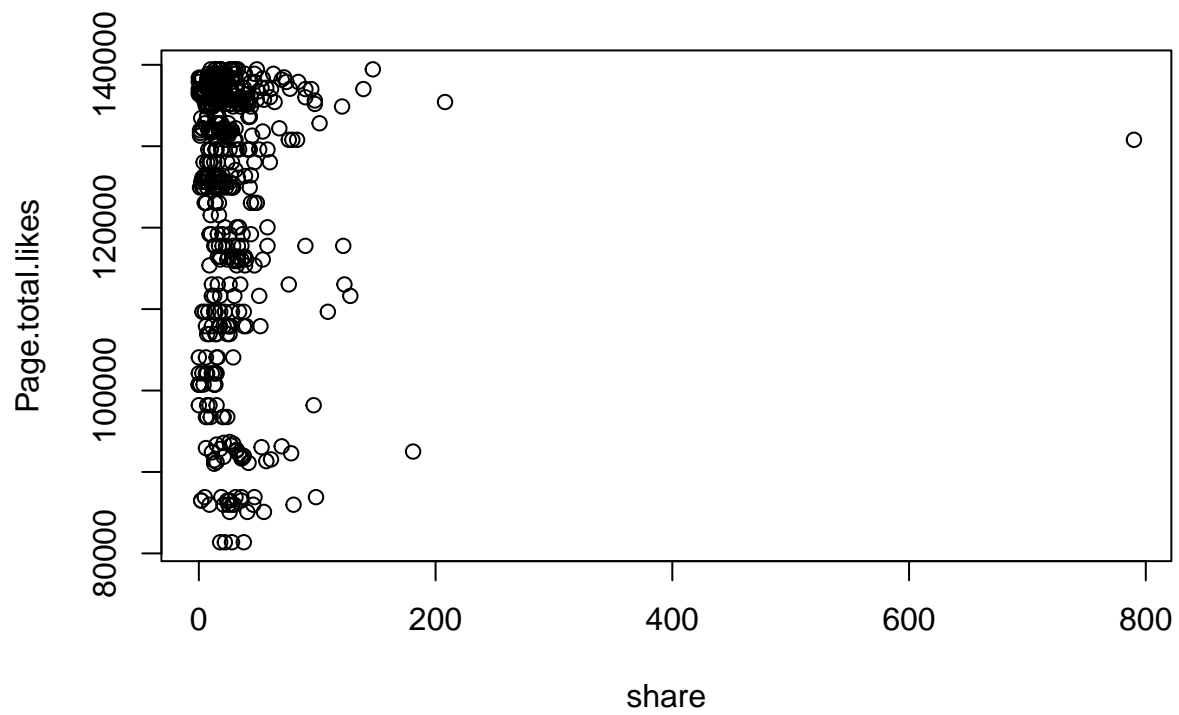
```
plot(Page.total.likes ~ Post.Weekday)
```

```r
plot(Page.total.likes ~ Post.Hour)
```
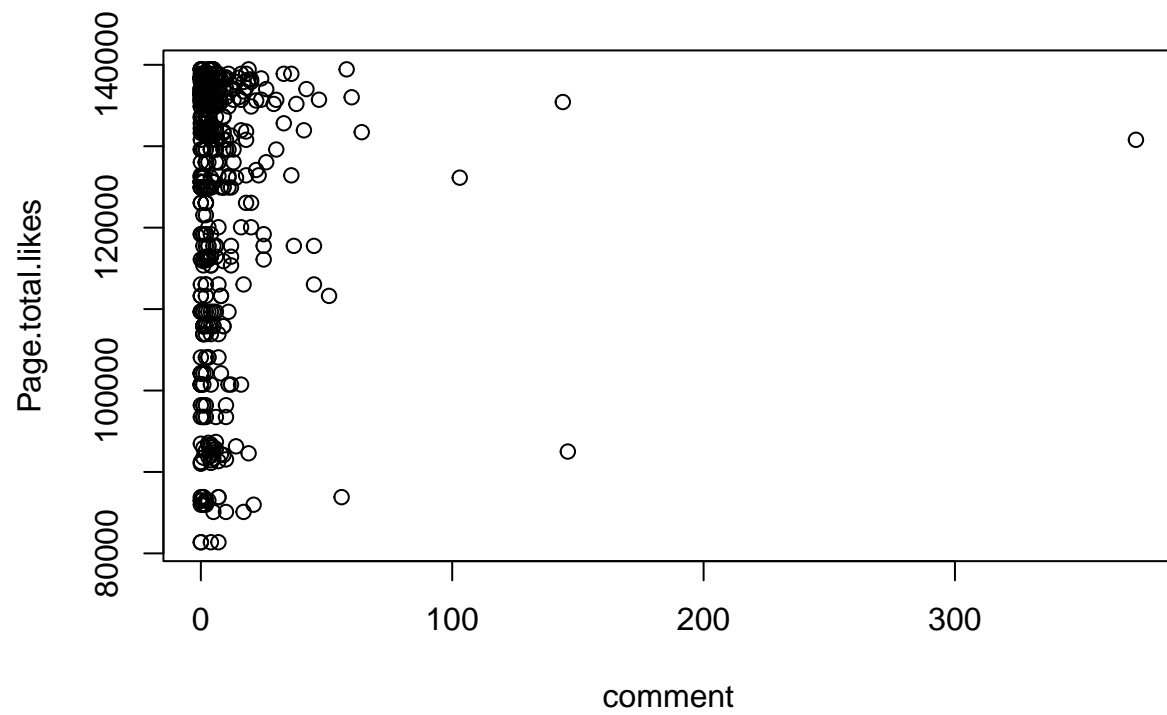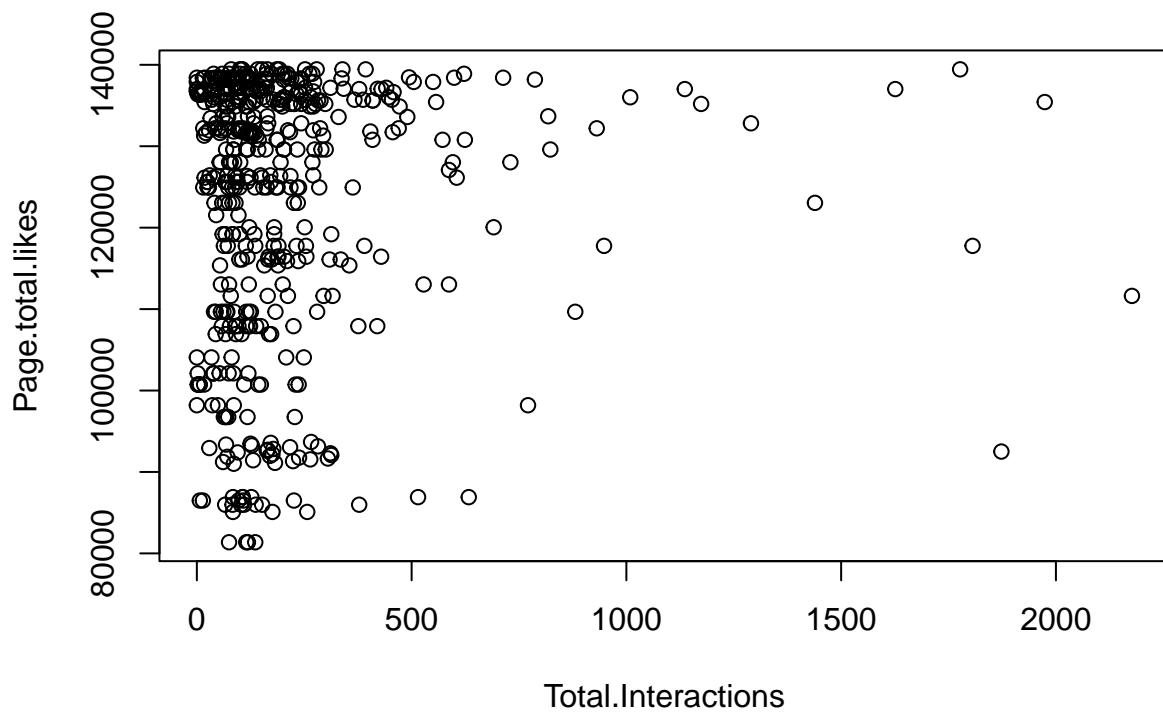
```
plot(Page.total.likes ~ like)
```

```
plot(Page.total.likes ~ share)
```

```r
plot(Page.total.likes ~ comment)
```

```r
plot(Page.total.likes ~ Total.Interactions,subset = Total.Interactions < 6000)
```

```r
plot(Page.total.likes ~ Lifetime.Post.Consumers)
```
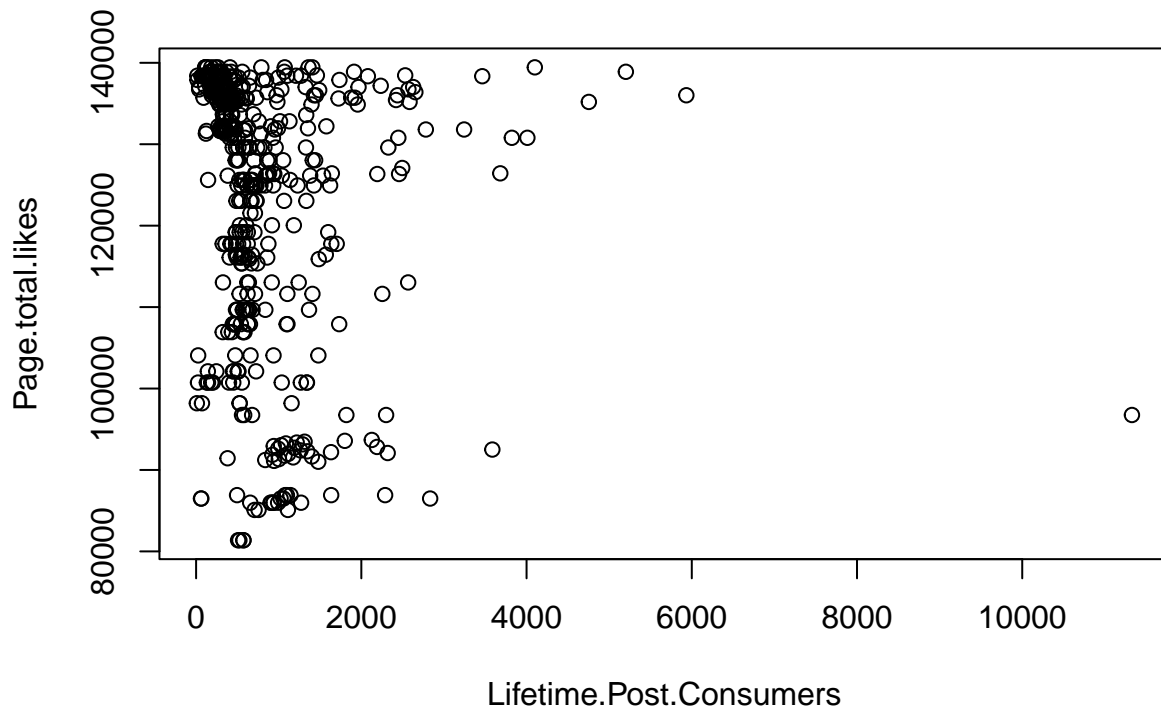
```
# temp_data <- fb %>% select("Page.total.likes",starts_with("Lifetime"))
# pair1<- ggpairs(temp_data,
#        lower=list(continuous=wrap("cor", alpha=0.5), combo="box"),
#        upper=list(continuous=wrap("points", alpha=0.3, size=0.1)), labeller = label_wrap_gen(5, mult
# pair1<- pair1 + theme(axis.text.x = element_text(angle = 40, hjust = 1, size = 7), axis.text.y = elem
# pair1
```

Fit the MV model

```
## dummy variable for type
fb$type.photo <- ifelse(Type == c("Photo"),1,0)
fb$type.status <- ifelse(Type == c("Status"),1,0)
fb$type.video <- ifelse(Type == c("Video"),1,0)

## simple linear model

m00 <- lm(Page.total.likes~type.photo+type.status+type.video, data=fb)
summary(m00)
```

```
##
## Call:
## lm(formula = Page.total.likes ~ type.photo + type.status + type.video,
##     data = fb)
```

```
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -40984 -10734   4070  13359  22532
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   116363       3396  34.270  < 2e-16 ***
## type.photo      5991       3482   1.721   0.0860 .
## type.status    16284       4143   3.930 9.69e-05 ***
## type.video     18652       6911   2.699   0.0072 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15930 on 496 degrees of freedom
## Multiple R-squared:  0.04788,    Adjusted R-squared:  0.04212
## F-statistic: 8.314 on 3 and 496 DF,  p-value: 2.102e-05
```

```
plot(m00)
```


Residuals vs Fitted

Fitted values
lm(Page.total.likes ~ type.photo + type.status + type.video)

16

Q–Q Residuals

Theoretical Quantiles
lm(Page.total.likes ~ type.photo + type.status + type.video)

Scale−Location

Fitted values
lm(Page.total.likes ~ type.photo + type.status + type.video)

Residuals vs Leverage

lm(Page.total.likes ~ type.photo + type.status + type.video)

```
## simple linear model - rank normalized
m0 <- lm(rn_page.total.likes~type.photo+type.status+type.video, data=fb)
summary(m0)
```
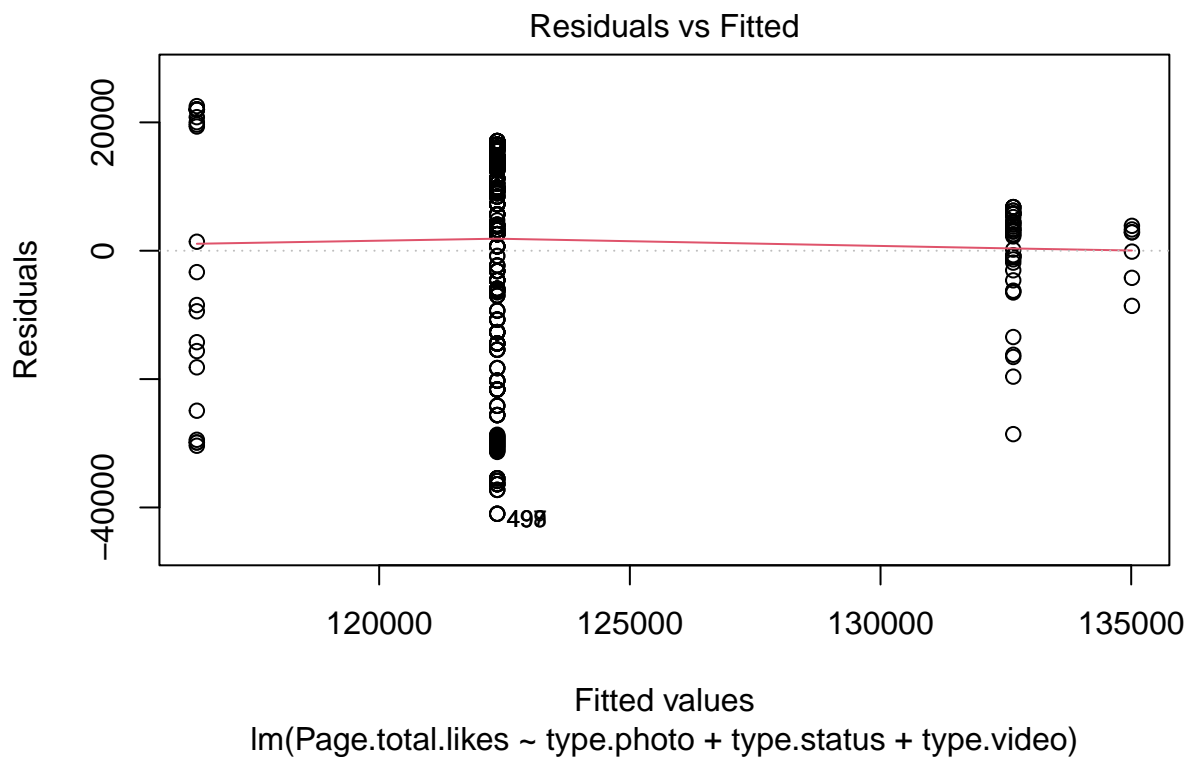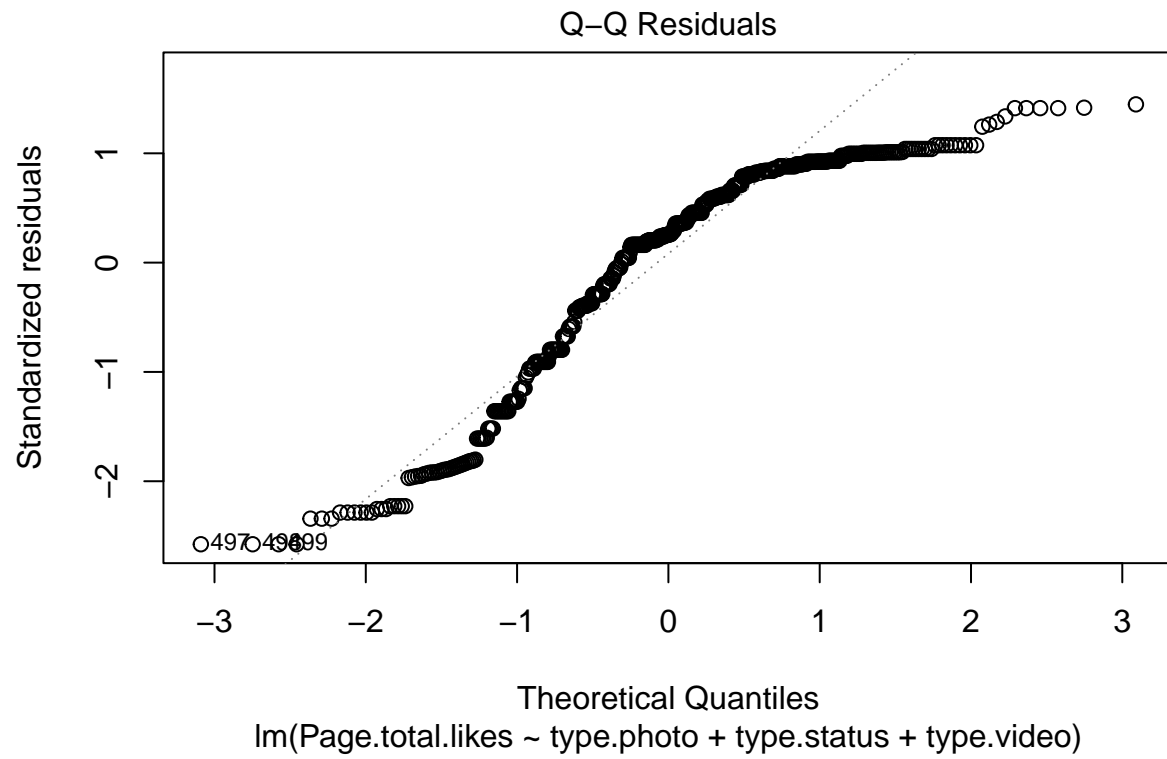
```
##
## Call:
## lm(formula = rn_page.total.likes ~ type.photo + type.status +
##     type.video, data = fb)
##
## Residuals:
##      Min      1Q   Median       3Q      Max
## -2.56038 -0.68363 -0.01396  0.72667  2.26186
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.1810     0.2059  -0.879  0.37977
## type.photo    0.1097     0.2111   0.519  0.60368
## type.status   0.8045     0.2512   3.202  0.00145 **
## type.video    0.9371     0.4191   2.236  0.02579 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9658 on 496 degrees of freedom
## Multiple R-squared:  0.05006,    Adjusted R-squared:  0.04432
## F-statistic: 8.713 on 3 and 496 DF,  p-value: 1.214e-05
```
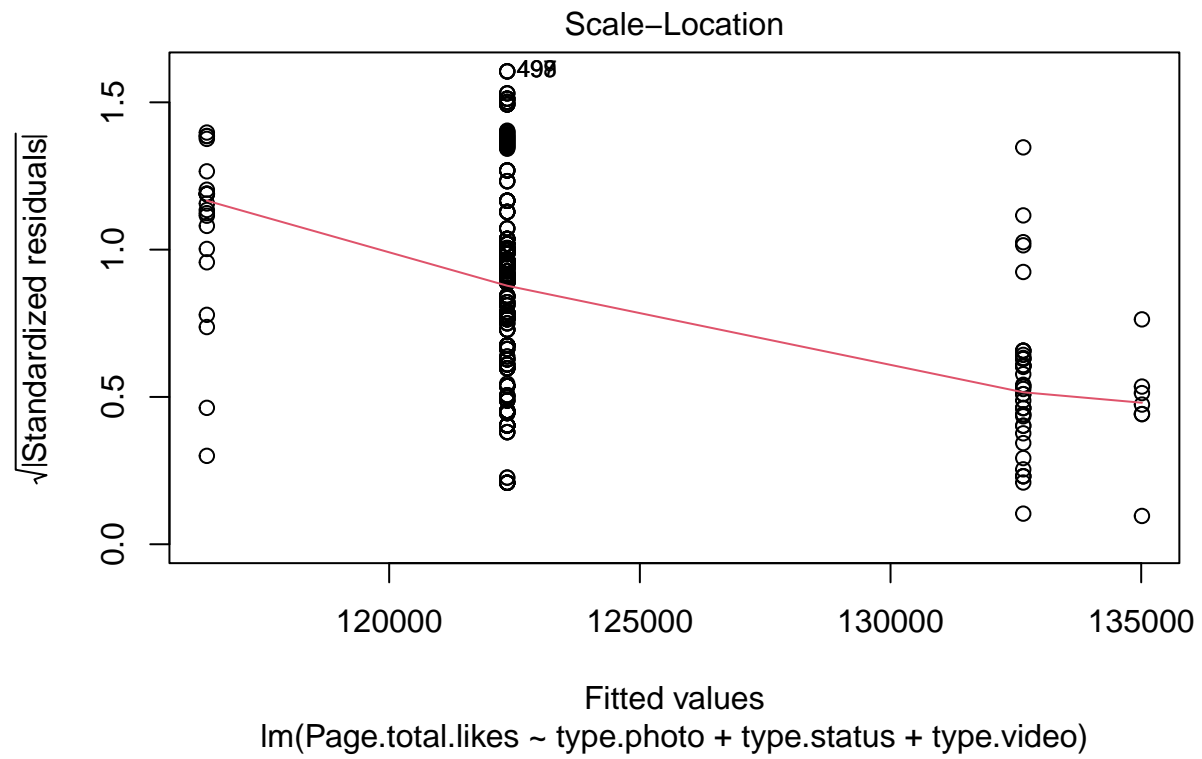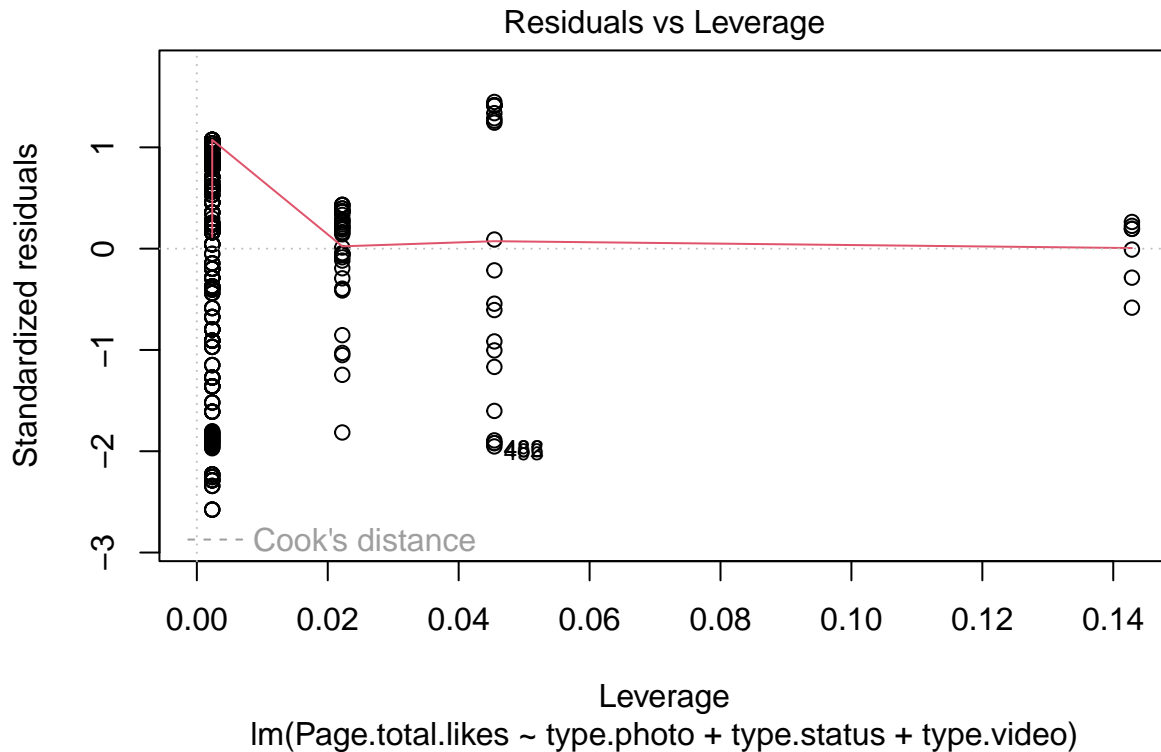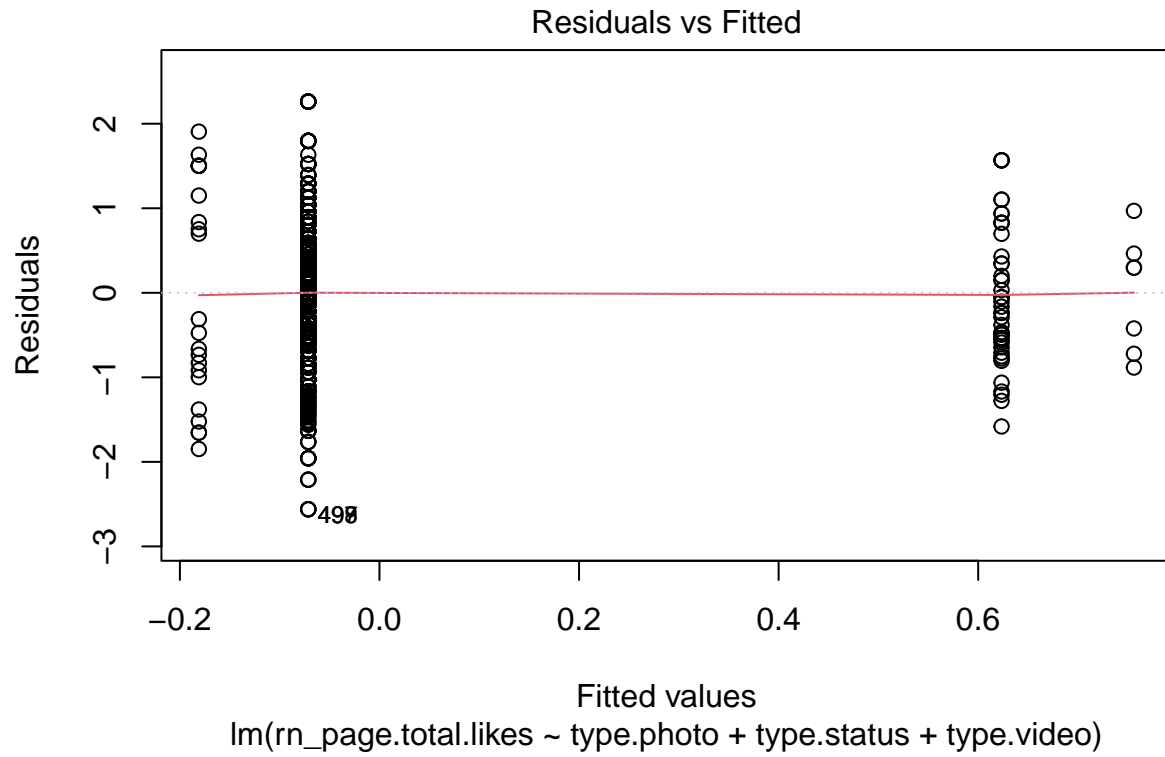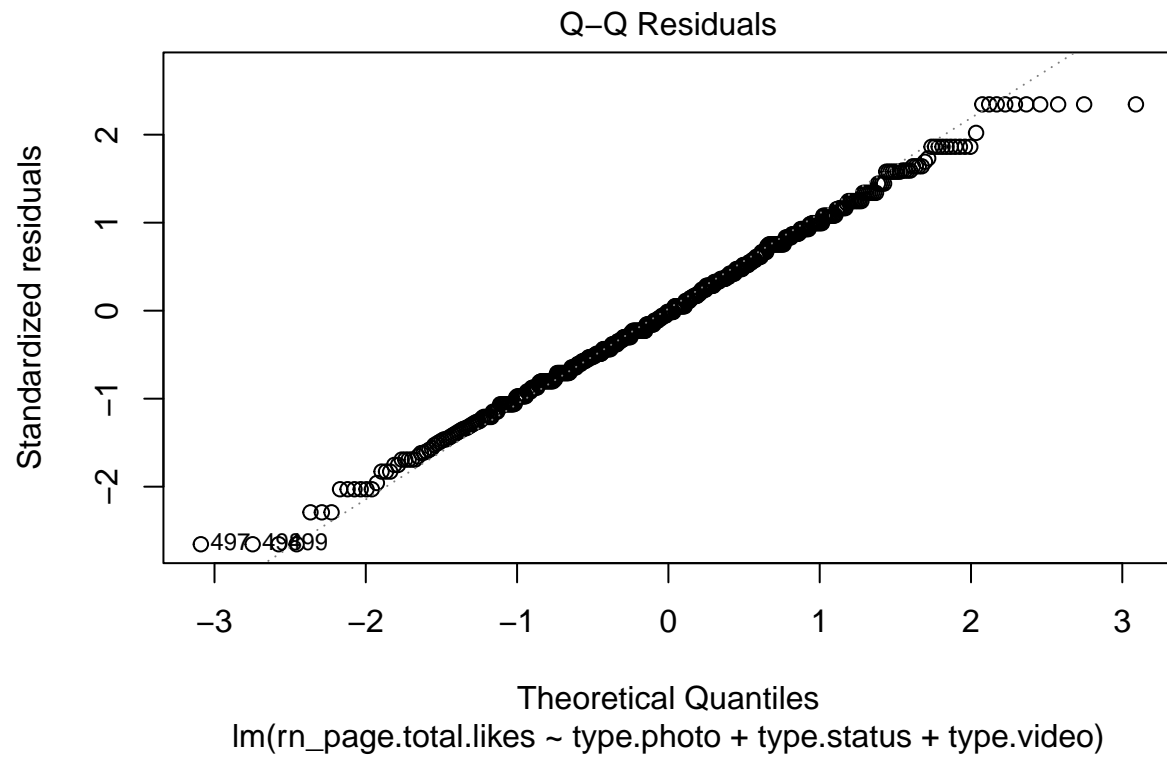
```
plot(m0)
```

Residuals vs Fitted

Fitted values
lm(rn_page.total.likes ~ type.photo + type.status + type.video)

Q−Q Residuals

Theoretical Quantiles
lm(rn_page.total.likes ~ type.photo + type.status + type.video)

Scale−Location

Fitted values
lm(rn_page.total.likes ~ type.photo + type.status + type.video)

Residuals vs Leverage

Leverage
lm(rn_page.total.likes ~ type.photo + type.status + type.video)
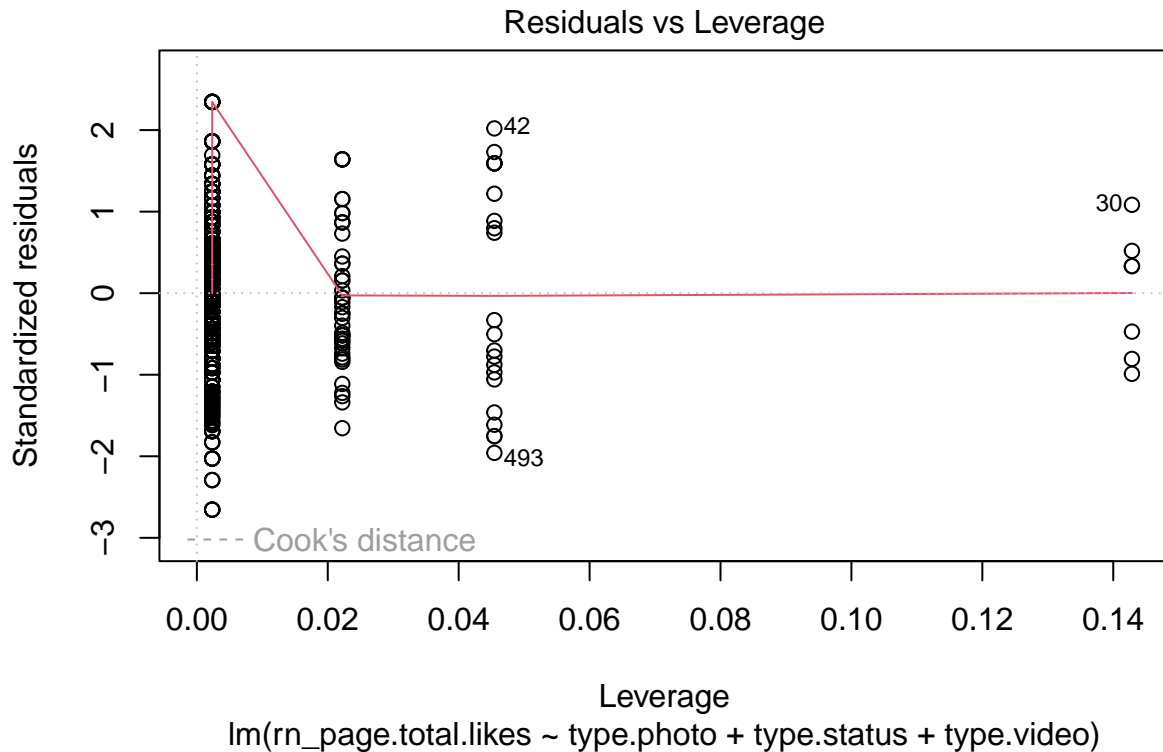
```
## first model - initial try
m1 <- lm(rn_page.total.likes~type.photo+type.status+type.video+Total.Interactions+Lifetime.Post.Consumer
summary(m1)
```

```
##
## Call:
## lm(formula = rn_page.total.likes ~ type.photo + type.status +
##     type.video + Total.Interactions + Lifetime.Post.Consumers +
##     Lifetime.Post.Consumptions, data = fb)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.59221 -0.65462  0.02478  0.61289  3.03808
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -9.984e-02  1.975e-01  -0.506 0.613347
## type.photo                   2.217e-01  2.031e-01   1.091 0.275623
## type.status                  1.410e+00  2.562e-01   5.503 6.02e-08 ***
## type.video                   1.372e+00  4.062e-01   3.377 0.000791 ***
## Total.Interactions           3.785e-04  1.180e-04   3.206 0.001433 **
## Lifetime.Post.Consumers     -3.796e-04  7.535e-05  -5.038 6.60e-07 ***
## Lifetime.Post.Consumptions  -1.003e-05  2.939e-05  -0.341 0.732936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.9231 on 493 degrees of freedom
## Multiple R-squared:  0.1375, Adjusted R-squared:  0.127
## F-statistic: 13.09 on 6 and 493 DF,  p-value: 9.045e-14
```
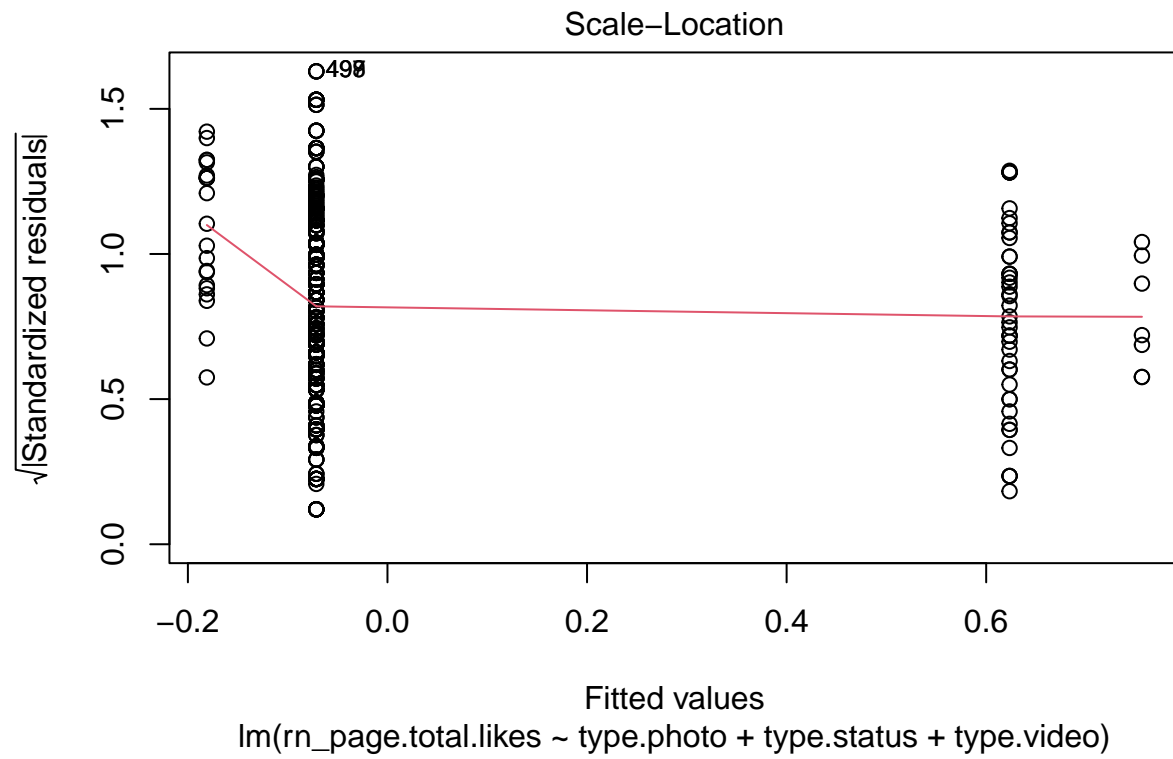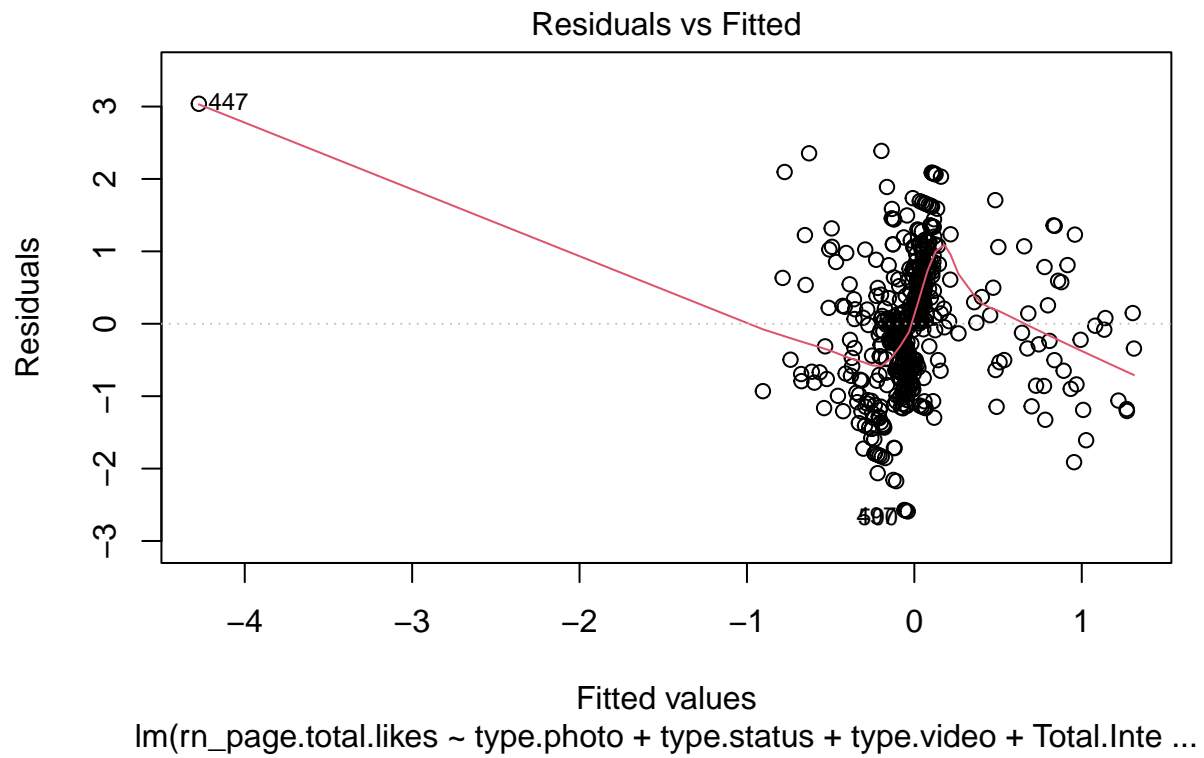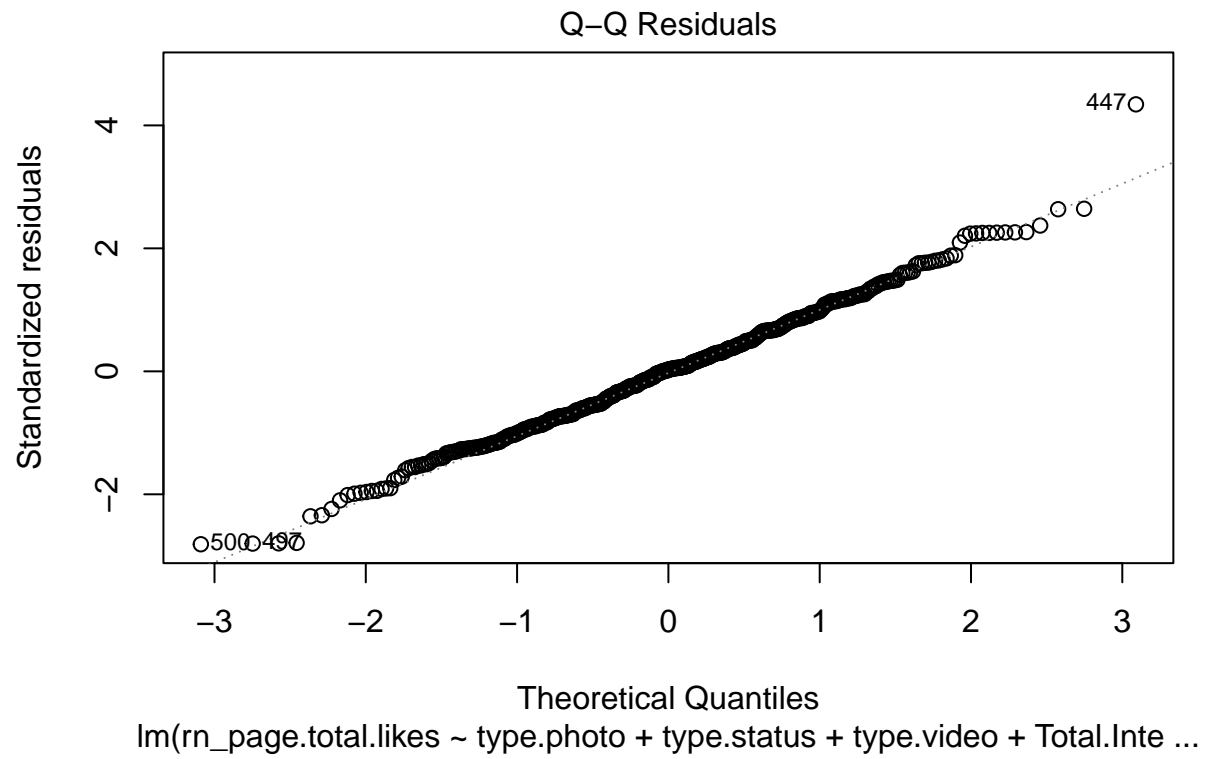
```
plot(m1)
```

Residuals vs Fitted

lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Q–Q Residuals

Standardized residuals

Theoretical Quantiles
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Scale−Location

Fitted values
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Residuals vs Leverage

Leverage
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

```r
AIC(m1)
```

```
## [1] 1347.825
```

```r
## outlines diagnostics are way off... will need to: 1) remove bad outliers
##                                                     2) add in a squared term
##                                                     3) check collinearity


## better but still not good - remove bad outliers 447 245
fb2 <- fb[-c(447,245),]
m2 <- lm(rn_page.total.likes~type.photo+type.status+type.video+Total.Interactions+Lifetime.Post.Consumer
summary(m2)
```

```
##
## Call:
## lm(formula = rn_page.total.likes ~ type.photo + type.status +
##     type.video + Total.Interactions + Lifetime.Post.Consumers +
##     Lifetime.Post.Consumptions, data = fb2)
##
## Residuals:
##     Min      1Q   Median      3Q      Max
## -2.58208 -0.61760  0.01333  0.60762  3.11487
##
## Coefficients:
```
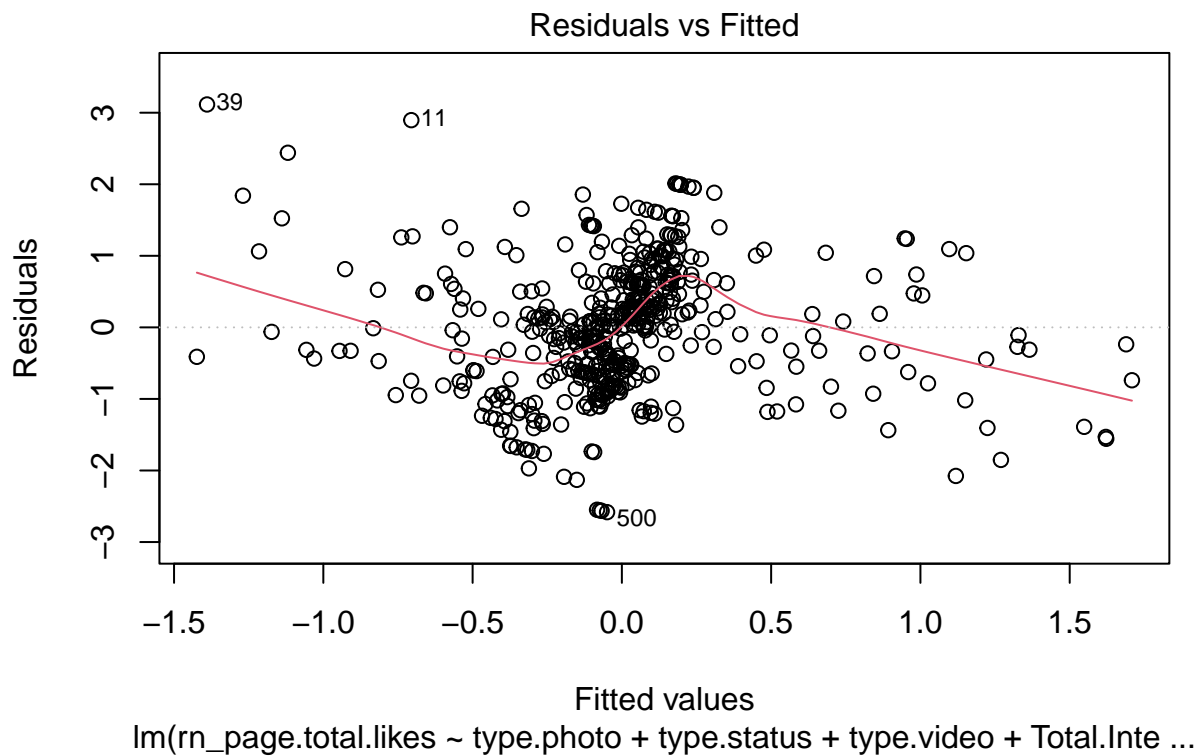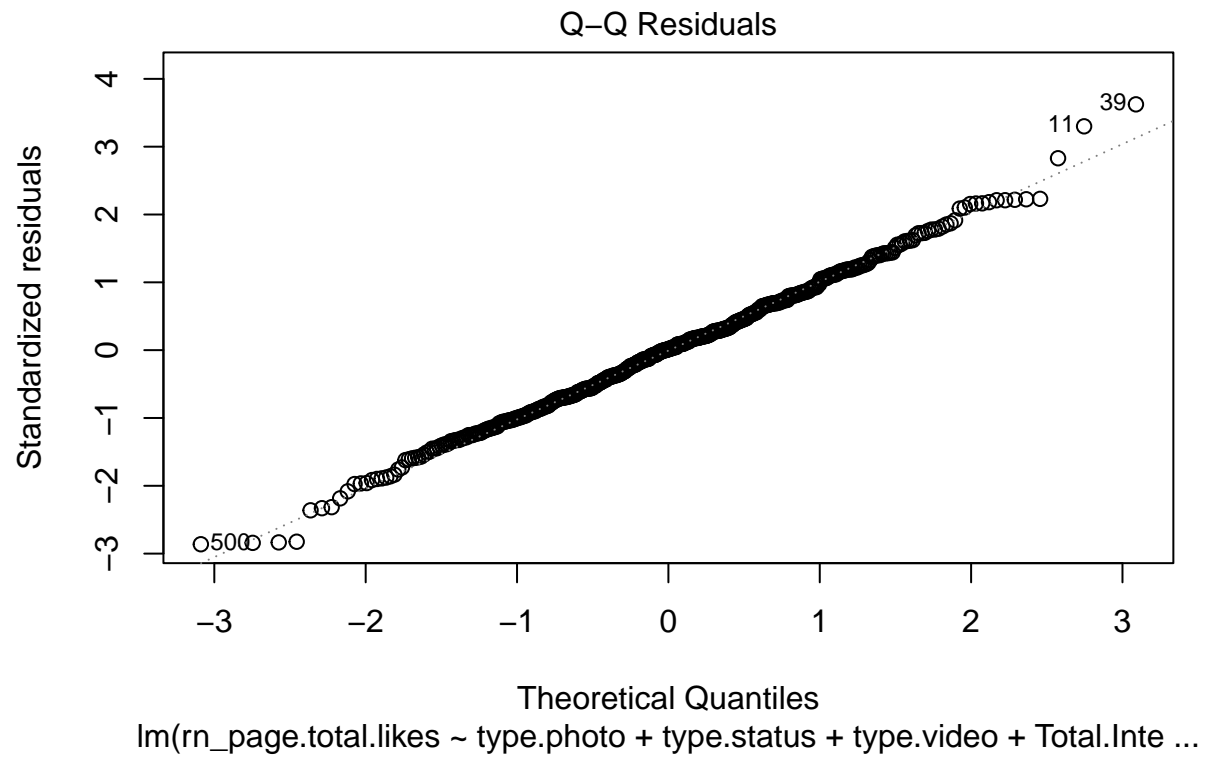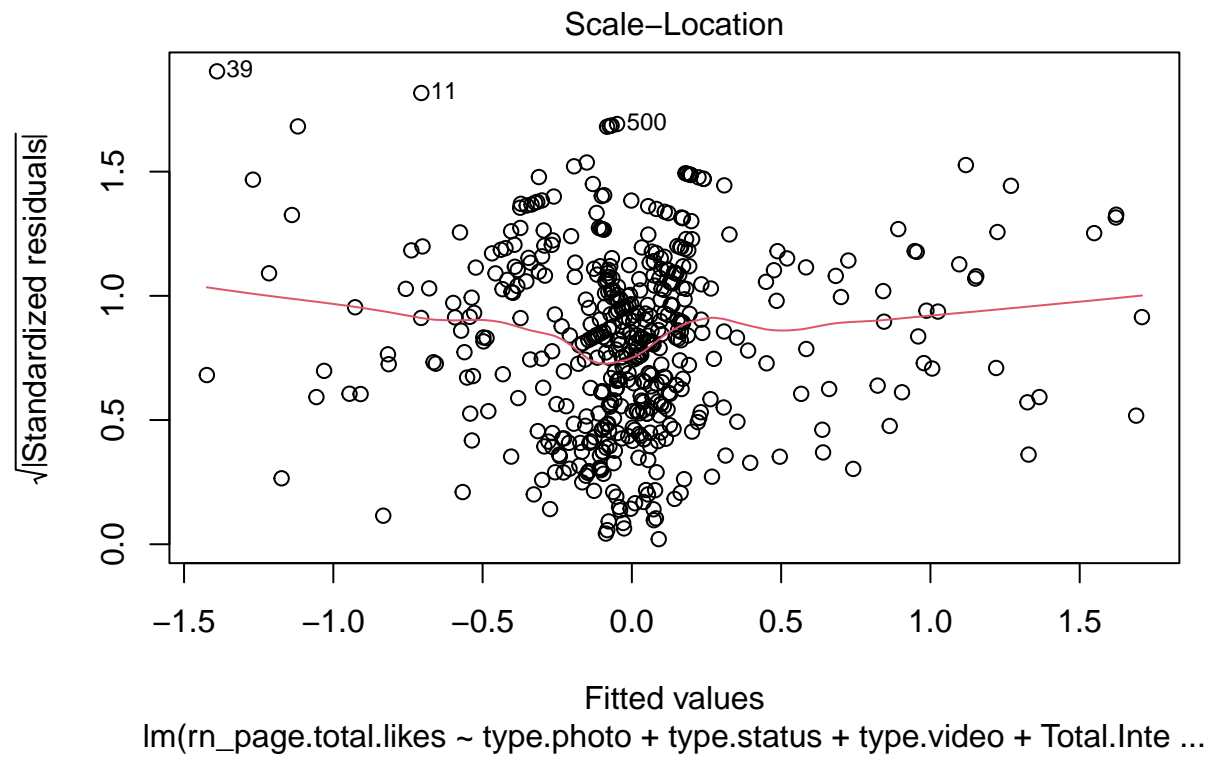
```
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -6.666e-02  1.937e-01  -0.344    0.731
## type.photo                 2.540e-01  1.993e-01   1.274    0.203
## type.status                1.751e+00  2.615e-01   6.694 5.94e-11 ***
## type.video                 1.591e+00  4.012e-01   3.965 8.43e-05 ***
## Total.Interactions         7.927e-04  1.738e-04   4.560 6.47e-06 ***
## Lifetime.Post.Consumers   -6.290e-04  9.016e-05  -6.977 9.82e-12 ***
## Lifetime.Post.Consumptions -2.219e-06  2.881e-05  -0.077    0.939
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9035 on 491 degrees of freedom
## Multiple R-squared:  0.1744, Adjusted R-squared:  0.1643
## F-statistic: 17.29 on 6 and 491 DF,  p-value: < 2.2e-16
```

```
plot(m2)
```


Residuals vs Fitted

Fitted values
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Q–Q Residuals

Standardized residuals

Theoretical Quantiles
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Scale−Location

√|Standardized residuals|

Fitted values
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Residuals vs Leverage

lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

```r
AIC(m2)
```

```
## [1] 1321.111
```

```r
##remove consumption
m3 <- lm(rn_page.total.likes~type.photo+type.status+type.video+Total.Interactions+Lifetime.Post.Consume
summary(m3)
```

```
##
## Call:
## lm(formula = rn_page.total.likes ~ type.photo + type.status +
##     type.video + Total.Interactions + Lifetime.Post.Consumers,
##     data = fb2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.58144 -0.62227  0.01378  0.60747  3.11957
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -6.638e-02  1.935e-01  -0.343    0.732
## type.photo            2.534e-01  1.989e-01   1.274    0.203
## type.status           1.752e+00  2.609e-01   6.714 5.23e-11 ***
## type.video            1.591e+00  4.008e-01   3.969 8.28e-05 ***
## Total.Interactions    7.934e-04  1.735e-04   4.574 6.07e-06 ***
```
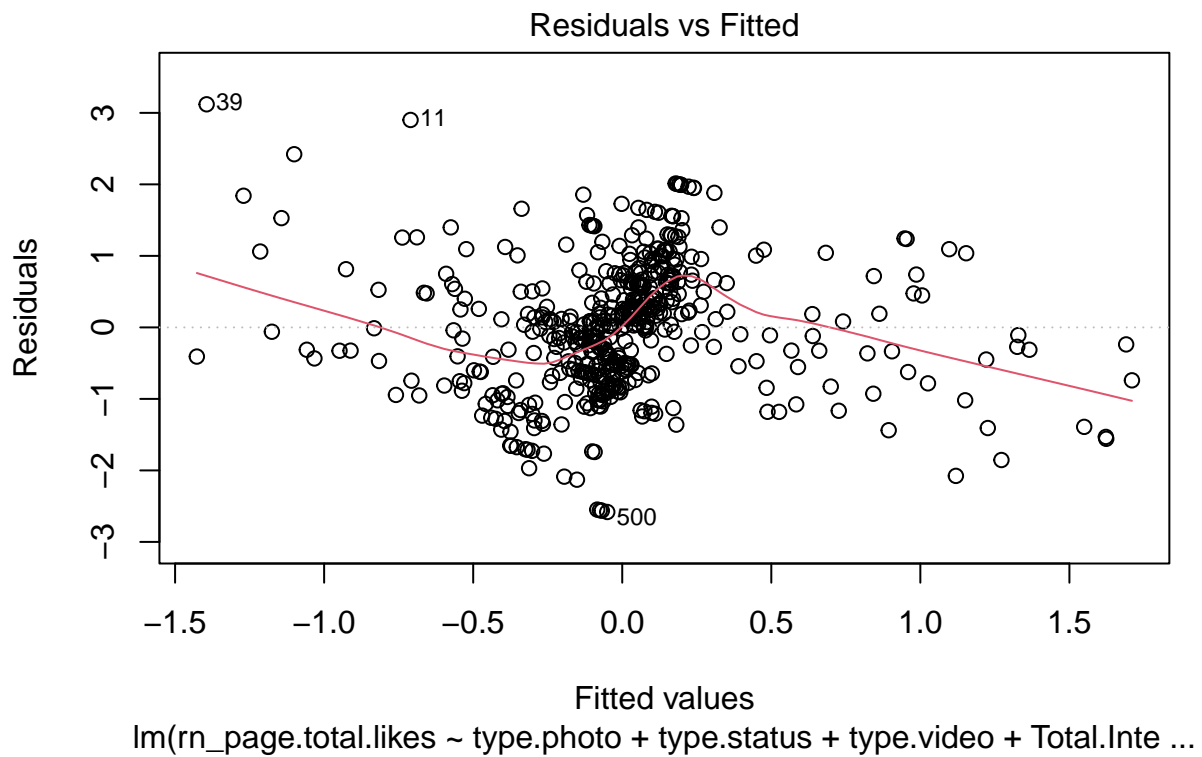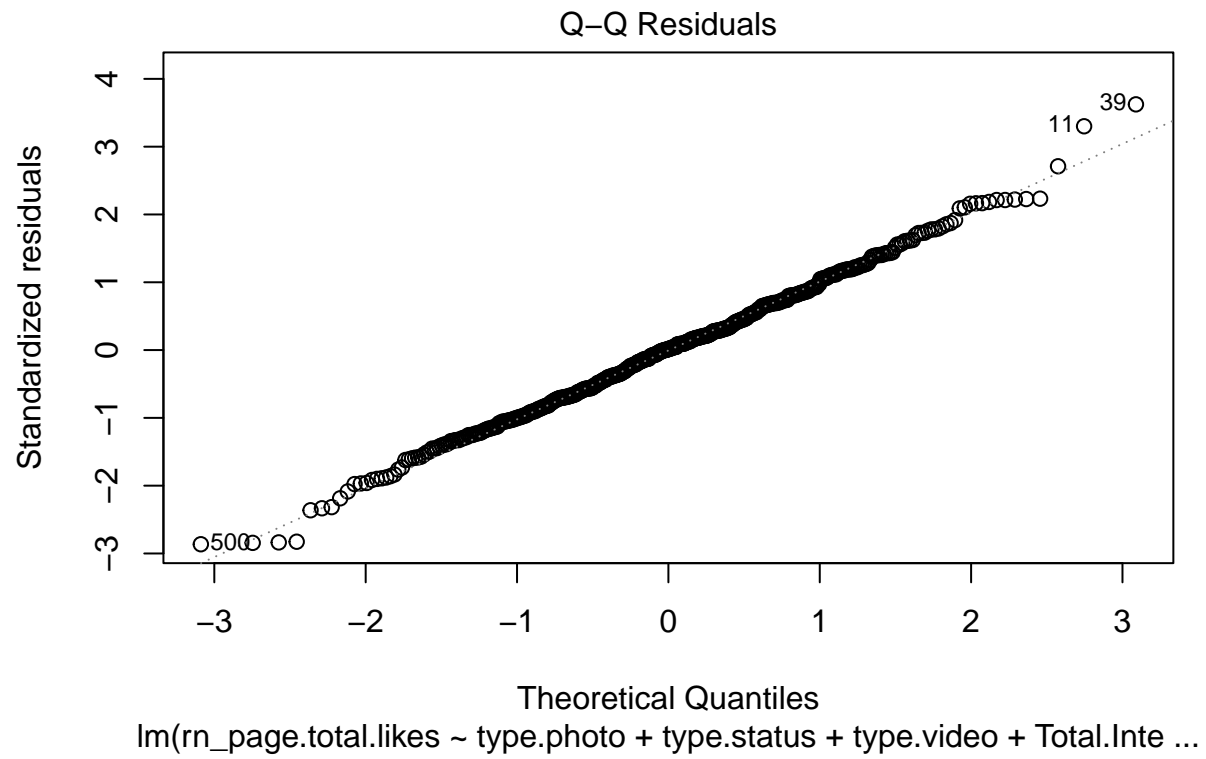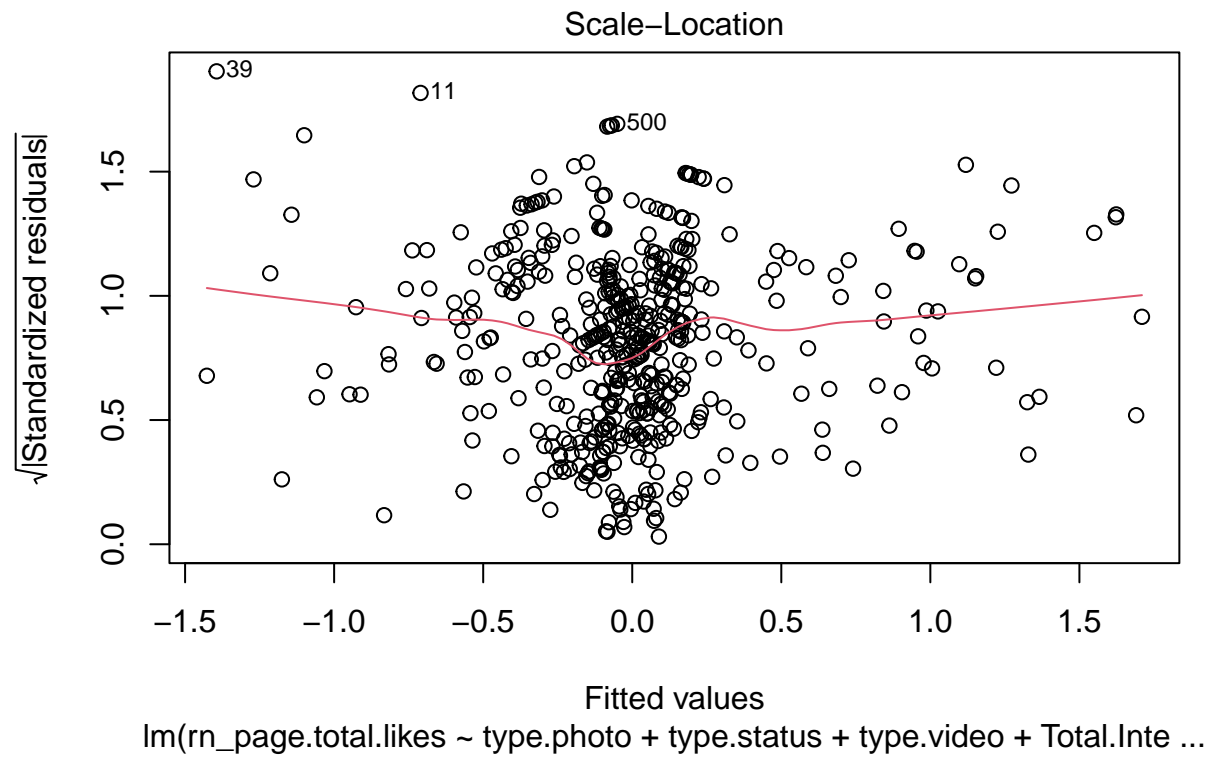
31

```
## Lifetime.Post.Consumers -6.330e-04  7.375e-05  -8.584  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9026 on 492 degrees of freedom
## Multiple R-squared:  0.1744, Adjusted R-squared:  0.166
## F-statistic: 20.79 on 5 and 492 DF,  p-value: < 2.2e-16
```

```
plot(m3)
```


Residuals vs Fitted

lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Q–Q Residuals

Standardized residuals

Theoretical Quantiles
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Scale−Location

Fitted values
lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

Residuals vs Leverage

lm(rn_page.total.likes ~ type.photo + type.status + type.video + Total.Inte ...

```
##partial F test
anova(m2,m3)
```

```
## Analysis of Variance Table
##
## Model 1: rn_page.total.likes ~ type.photo + type.status + type.video +
##     Total.Interactions + Lifetime.Post.Consumers + Lifetime.Post.Consumptions
## Model 2: rn_page.total.likes ~ type.photo + type.status + type.video +
##     Total.Interactions + Lifetime.Post.Consumers
##   Res.Df    RSS Df  Sum of Sq      F Pr(>F)
## 1    491 400.79
## 2    492 400.79 -1 -0.0048417 0.0059 0.9386
```

```
## check collinearity

vif(m3)
```

```
##           type.photo           type.status            type.video
##             3.060808              3.420518              1.360975
##    Total.Interactions Lifetime.Post.Consumers
##             1.277676              1.783045
```

```
## final model

summary(m3)
```

```
## 
## Call:
## lm(formula = rn_page.total.likes ~ type.photo + type.status +
##     type.video + Total.Interactions + Lifetime.Post.Consumers,
##     data = fb2)
## 
## Residuals:
##      Min      1Q   Median      3Q      Max
## -2.58144 -0.62227  0.01378  0.60747  3.11957
## 
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -6.638e-02  1.935e-01  -0.343    0.732
## type.photo               2.534e-01  1.989e-01   1.274    0.203
## type.status              1.752e+00  2.609e-01   6.714 5.23e-11 ***
## type.video               1.591e+00  4.008e-01   3.969 8.28e-05 ***
## Total.Interactions       7.934e-04  1.735e-04   4.574 6.07e-06 ***
## Lifetime.Post.Consumers -6.330e-04  7.375e-05  -8.584  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.9026 on 492 degrees of freedom
## Multiple R-squared:  0.1744, Adjusted R-squared:  0.166
## F-statistic: 20.79 on 5 and 492 DF,  p-value: < 2.2e-16
```

Visualization

```
## m3
par(mfrow=c(2,2))
plot(m3)
```
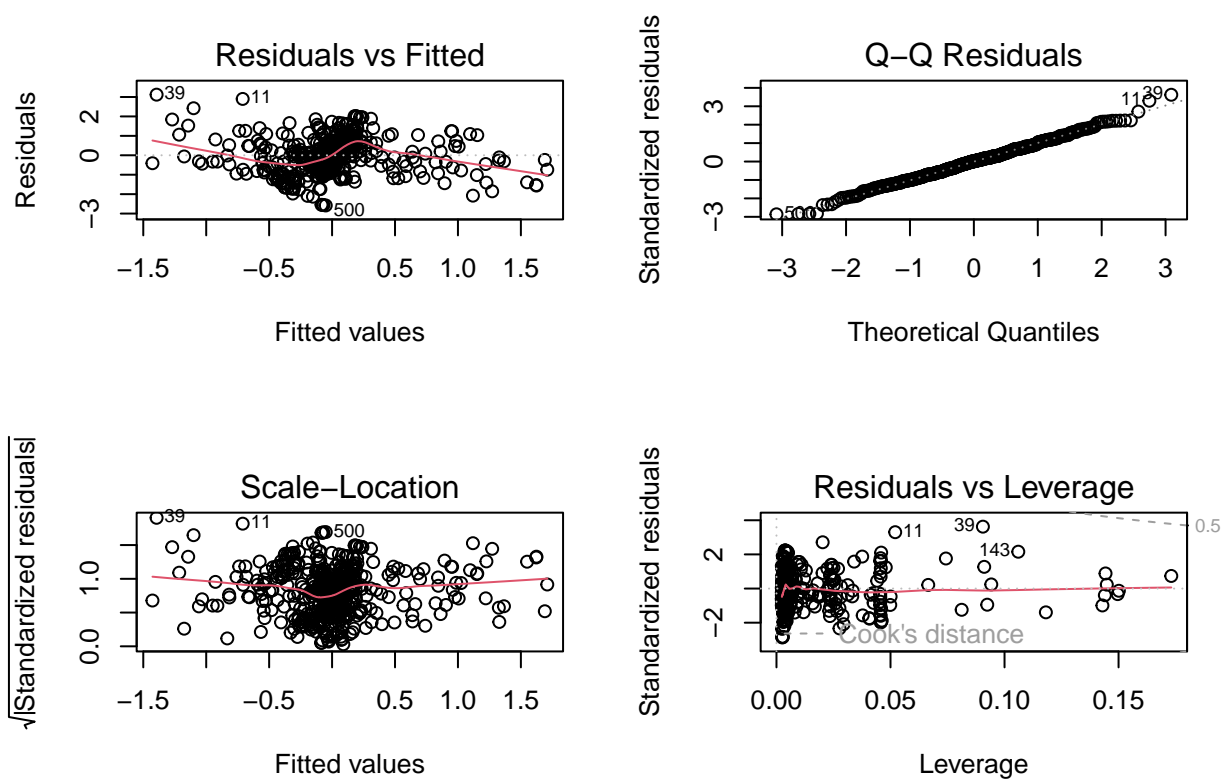
```
avPlots(m3)
```

Added−Variable Plots