

```
> 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

```
> library(tidyr)
```

```
> library(corrplot)
```

corrplot 0.84 loaded

```
> library(ggplot2)
```

```
> library(knitr)
```

```
> #Read in Data Sets ----
```

```
> setwd("C:/Users/sweeneys/Desktop/")
```

```
> dat = read.csv("trumpmin.csv")
```

```
> head(dat)
```

	date	new_id	notweets	retweets	favorites
1	5/4/2009	1	1	500	868
2	5/5/2009	2	1	33	273
3	5/8/2009	3	2	11	18
4	5/12/2009	4	2	27	26
5	5/13/2009	5	1	14	16
6	5/14/2009	6	1	18	25

```
text
```

```
1 Be sure to tune in and watch Donald Trump on Late Night with
David Letterman as he presents the Top Ten List tonight!
2 Donald Trump will be appearing on The View tomorrow morning to discuss Cele
brity Apprentice and his new book Think Like A Champion!
3 Donald Trump reads Top Ten Financia
l Tips on Late Show with David Letterman: - Very funny!
4 Miss USA Tara Conner will not be fired - I've alw
ays been a believer in second chances. says Donald Trump
5 Listen to an interview with Donal
d Trump discussing his new book, Think Like A Champion:
6 Strive for wholeness
and keep your sense of wonder intact. --Donald J. Trump
```

	ss_min	ss_max	ss_compound	presidency	delta_1day	delta_2day
1	0.5255	0.5255	0.52550	1	1	1
2	0.7712	0.7712	0.77120	1	0	1
3	0.0000	0.6468	0.32340	1	1	1
4	0.0000	0.4809	0.24045	1	0	0
5	0.7506	0.7506	0.75060	1	0	0
6	0.2023	0.2023	0.20230	1	1	0

```
> #Data subsets ----
```

```
> datpre <- dat[which(dat$presidency=='1'),]
```

```
> datcand <- dat[which(dat$presidency=='2'),]
```

```
> datpres <- dat[which(dat$presidency=='3'),]
```

```
> datred <- dat[c(-2,-6)]
```

```
> datred =datred[complete.cases(datred),]
```

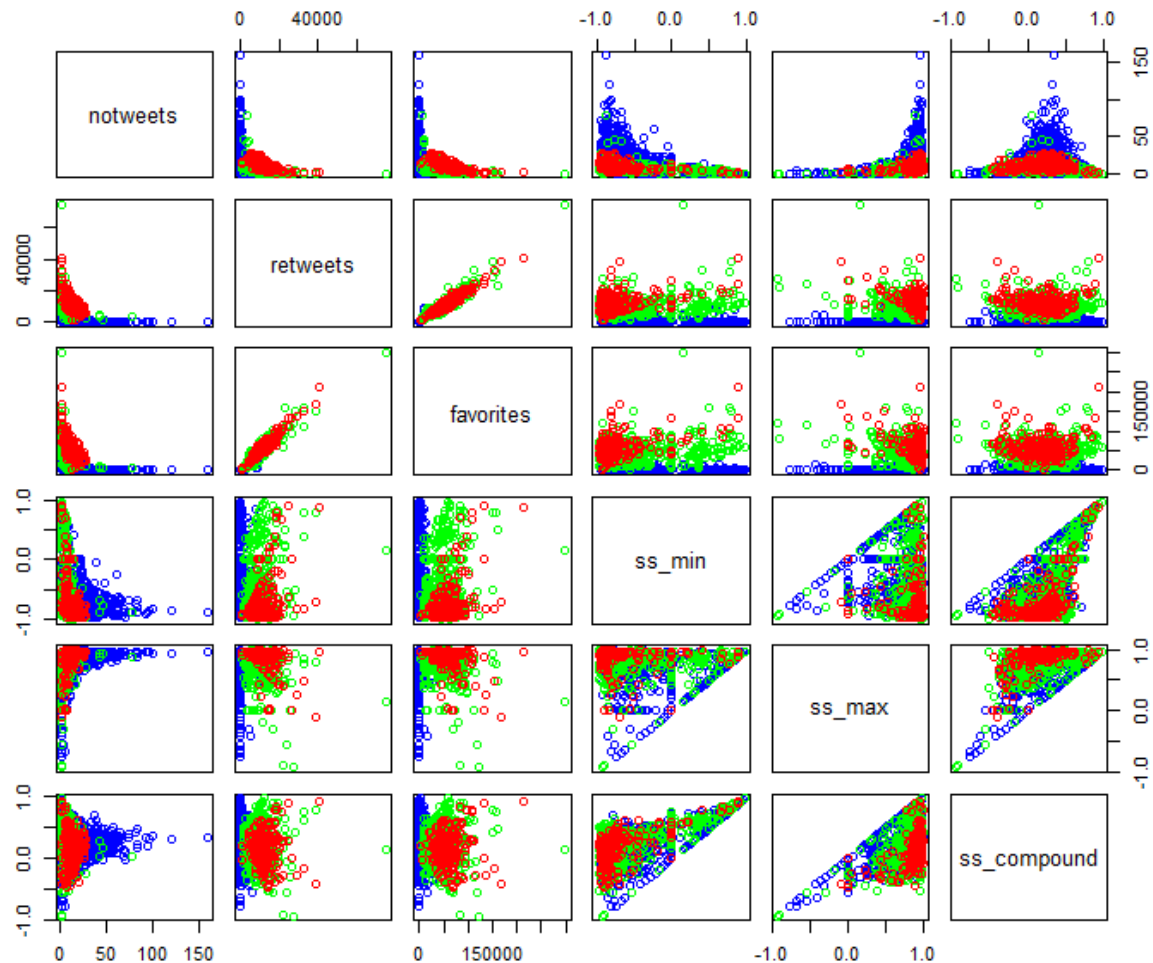
```
> head(datred)
```

	date	notweets	retweets	favorites	ss_min	ss_max	ss_compound	presidency
1	5/4/2009	1	500	868	0.5255	0.5255	0.52550	1
1		1						
2	5/5/2009	1	33	273	0.7712	0.7712	0.77120	1
0		1						
3	5/8/2009	2	11	18	0.0000	0.6468	0.32340	1
1		1						

```

4 5/12/2009      2      27      26 0.0000 0.4809      0.24045      1
0              0
5 5/13/2009      1      14      16 0.7506 0.7506      0.75060      1
0
6 5/14/2009      1      18      25 0.2023 0.2023      0.20230      1
1
> datless <- datred[c(-1,-8,-9,-10)]
> datless = datless[complete.cases(datless),]
> head(datless)
  notweets retweets favorites ss_min ss_max ss_compound
1         1        500      868 0.5255 0.5255      0.52550
2         1         33       273 0.7712 0.7712      0.77120
3         2         11        18 0.0000 0.6468      0.32340
4         2         27        26 0.0000 0.4809      0.24045
5         1         14        16 0.7506 0.7506      0.75060
6         1         18        25 0.2023 0.2023      0.20230
> #Plots ----
> pairs(datless,col=c("blue","green","red")[dat$presidency])

```



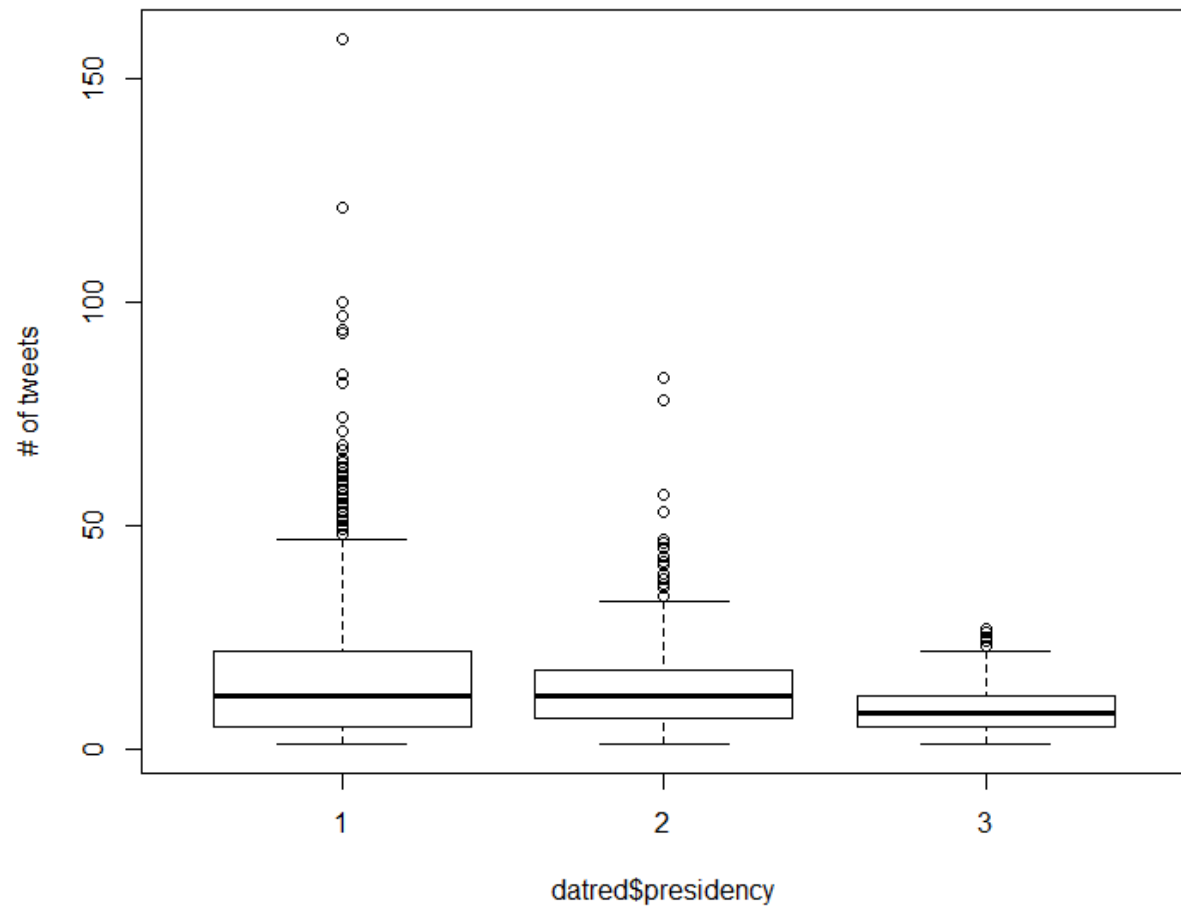
```

> boxplot(datred$notweets ~ datred$presidency,main="Number of Daily Tweets by
Presidency status",ylab="# of tweets")
> boxplot(datred$ss_min ~ datred$presidency,main="Minimum Sentiment Score by
Presidency status",ylab="Sentiment score")

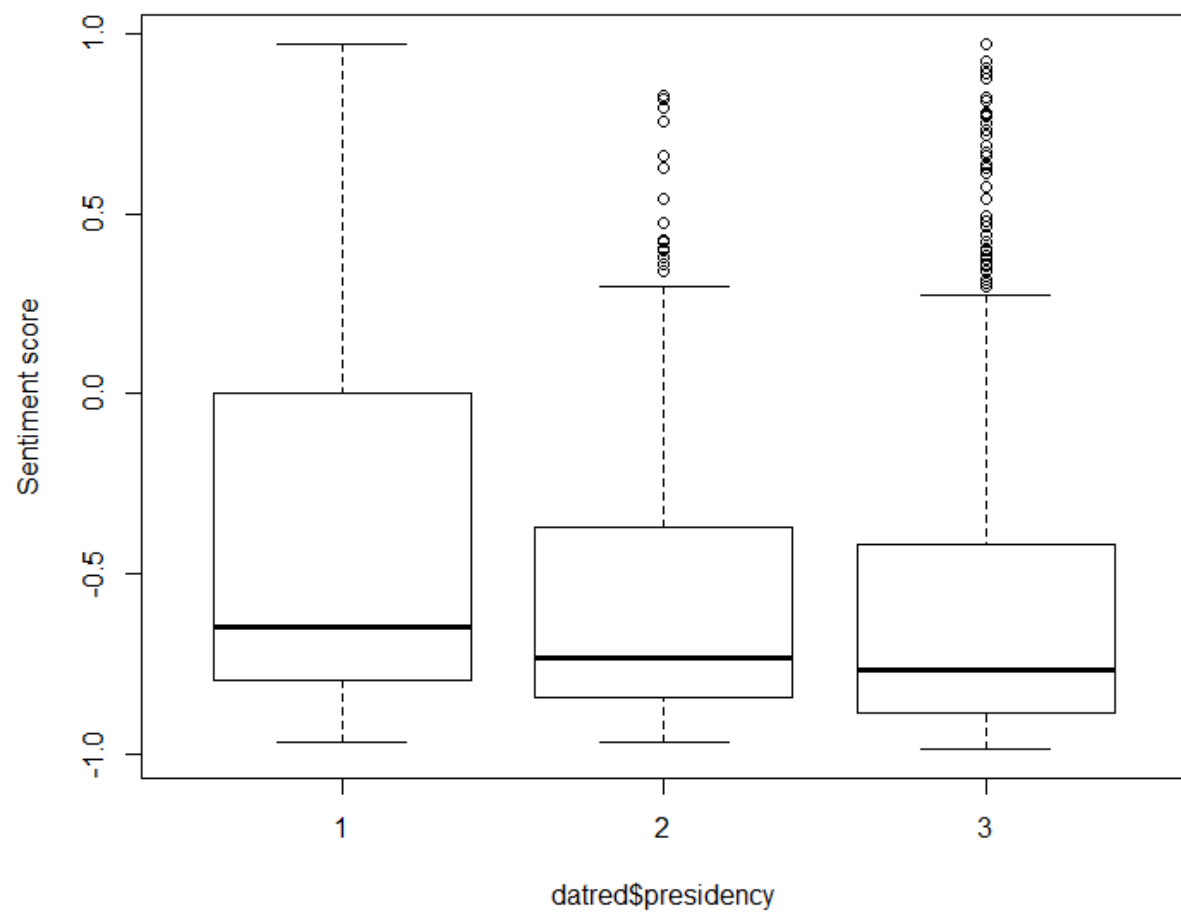
```

```
> boxplot(datred$ss_max ~ datred$presidency,main="Maximum Sentiment Score by  
Presidency status",ylab="Sentiment score")  
> boxplot(datred$ss_compound ~ datred$presidency,main="Compound Sentiment Sco  
re by Presidency status",ylab="Sentiment score")
```

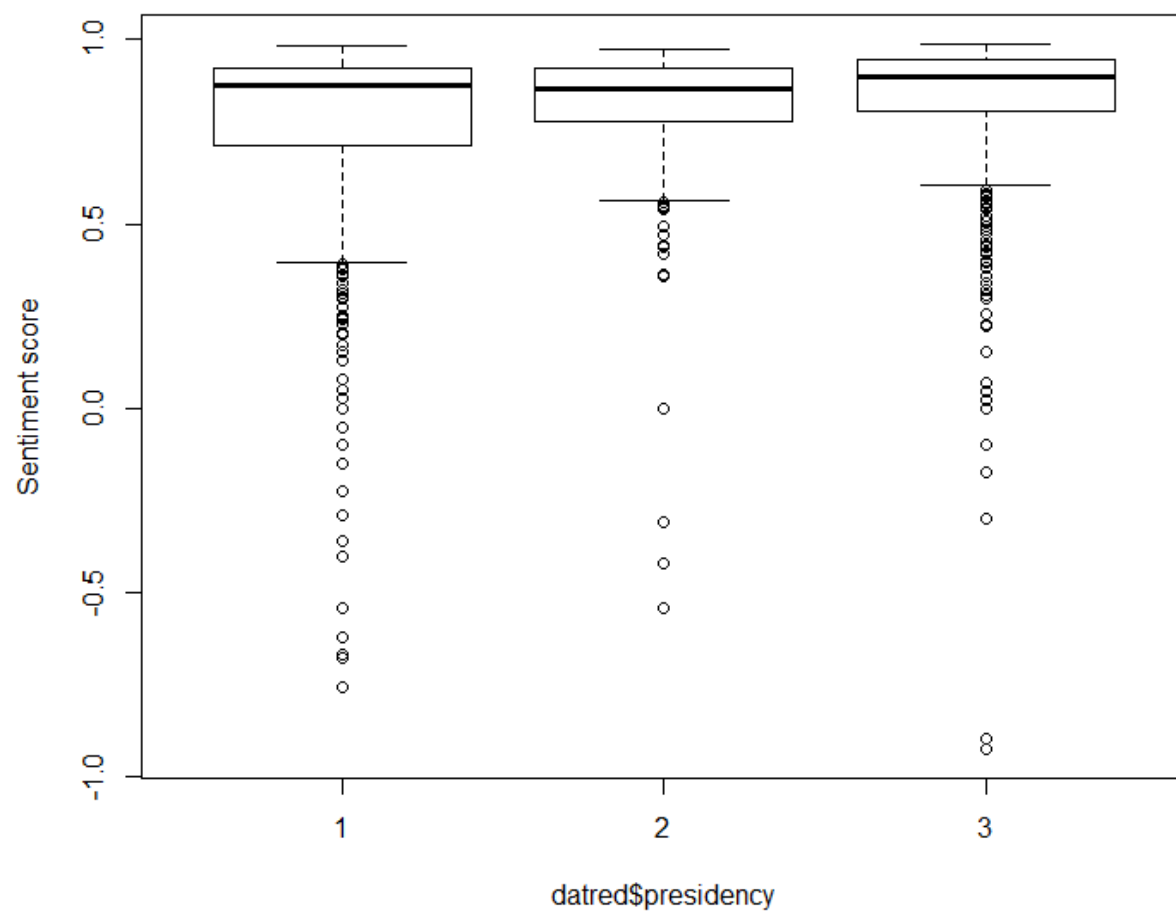
Number of Daily Tweets by Presidency status



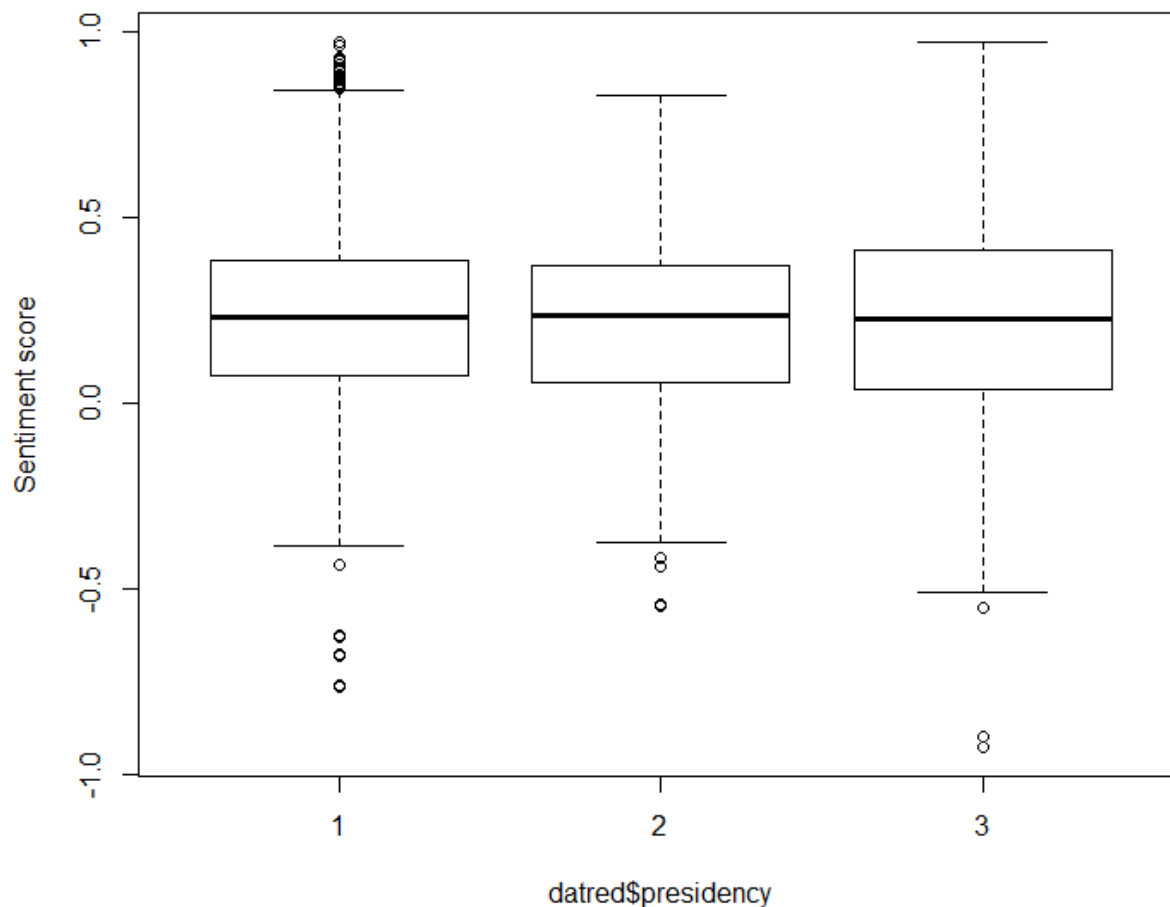
Minimum Sentiment Score by Presidency status



Maximum Sentiment Score by Presidency status



Compound Sentiment Score by Presidency status



```
> # Logistic Regression 1 day ----
> mod_log = glm(datred$delta_1day ~ datred$notweets + datred$retweets + datred$favorites + datred$ss_min + datred$ss_max + datred$ss_compound + datred$presidency, data = datred, family = binomial)
> summary(mod_log)
```

```
Call:
glm(formula = datred$delta_1day ~ datred$notweets + datred$retweets + datred$favorites + datred$ss_min + datred$ss_max + datred$ss_compound + datred$presidency, family = binomial, data = datred)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.619	-1.233	1.053	1.121	1.302

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.932e-01	1.632e-01	1.184	0.2363
datred\$notweets	-1.182e-03	4.063e-03	-0.291	0.7712
datred\$retweets	-5.465e-05	3.848e-05	-1.420	0.1555
datred\$favorites	1.437e-05	8.348e-06	1.722	0.0851
datred\$ss_min	1.601e-02	1.829e-01	0.088	0.9302
datred\$ss_max	1.256e-01	2.686e-01	0.468	0.6400

```

datred$ss_compound -2.364e-01  3.299e-01  -0.716  0.4737
datred$presidency -6.365e-02  8.942e-02  -0.712  0.4766
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 3132.6  on 2268  degrees of freedom
Residual deviance: 3127.8  on 2261  degrees of freedom
AIC: 3143.8

```

Number of Fisher Scoring iterations: 3

```

> mod_log2 = glm(datred$delta_2day ~ datred$notweets + datred$retweets + datred$favorites + datred$ss_min + datred$ss_max + datred$ss_compound + datred$presidency, data = datred, family = binomial)
> summary(mod_log2)

```

```

Call:
glm(formula = datred$delta_2day ~ datred$notweets + datred$retweets + datred$favorites + datred$ss_min + datred$ss_max + datred$ss_compound + datred$presidency, family = binomial, data = datred)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.647  -1.298   1.008   1.053   1.217

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.571e-01	1.652e-01	2.162	0.0306 *
datred\$notweets	-6.293e-04	4.106e-03	-0.153	0.8782
datred\$retweets	-4.806e-05	3.971e-05	-1.210	0.2262
datred\$favorites	1.757e-05	8.575e-06	2.049	0.0405 *
datred\$ss_min	-8.722e-02	1.849e-01	-0.472	0.6372
datred\$ss_max	1.042e-01	2.705e-01	0.385	0.7002
datred\$ss_compound	8.505e-02	3.331e-01	0.255	0.7985
datred\$presidency	-1.826e-01	9.234e-02	-1.977	0.0480 *

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 3094.7  on 2268  degrees of freedom
Residual deviance: 3084.6  on 2261  degrees of freedom
AIC: 3100.6

```

Number of Fisher Scoring iterations: 4

```

> mod_log4 = glm(datred$delta_2day ~ datred$retweets + datred$favorites + datred$ss_min + datred$ss_max + datred$presidency, data = datred, family = binomial)
> summary(mod_log4)

```

```

Call:
glm(formula = datred$delta_2day ~ datred$retweets + datred$favorites + datred$ss_min + datred$ss_max + datred$presidency, family = binomial, data = datred)

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.649  -1.298   1.009   1.052   1.217

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
--	----------	------------	---------	----------

(Intercept)	3.514e-01	1.638e-01	2.146	0.0319 *
datred\$retweets	-4.881e-05	3.956e-05	-1.234	0.2173
datred\$favorites	1.771e-05	8.557e-06	2.070	0.0385 *
datred\$ss_min	-4.382e-02	9.628e-02	-0.455	0.6490
datred\$ss_max	1.469e-01	1.737e-01	0.846	0.3976
datred\$presidency	-1.799e-01	9.039e-02	-1.990	0.0466 *

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3094.7 on 2268 degrees of freedom
 Residual deviance: 3084.7 on 2263 degrees of freedom
 AIC: 3096.7

Number of Fisher Scoring iterations: 4

```
> mod_log5 = glm(datred$delta_2day ~ datred$presidency ,data = datred, family
= binomial)
> summary(mod_log5)
```

Call:
 glm(formula = datred\$delta_2day ~ datred\$presidency, family = binomial,
 data = datred)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.325	-1.297	1.037	1.062	1.062

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.24646	0.09533	2.586	0.00972 **
datred\$presidency	0.03100	0.04851	0.639	0.52275

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3094.7 on 2268 degrees of freedom
 Residual deviance: 3094.3 on 2267 degrees of freedom
 AIC: 3098.3

Number of Fisher Scoring iterations: 4

```
> mod_log6 = glm(datred$delta_2day ~ datred$retweets + datred$favorites + dat
red$ss_max +datred$presidency ,data = datred, family = binomial)
> summary(mod_log6)
```

Call:
 glm(formula = datred\$delta_2day ~ datred\$retweets + datred\$favorites +
 datred\$ss_max + datred\$presidency, family = binomial, data = datred)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.670	-1.303	1.011	1.050	1.212

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.523e-01	1.637e-01	2.152	0.0314 *
datred\$retweets	-4.773e-05	3.950e-05	-1.208	0.2269
datred\$favorites	1.732e-05	8.515e-06	2.034	0.0420 *
datred\$ss_max	1.623e-01	1.704e-01	0.952	0.3409
datred\$presidency	-1.731e-01	8.917e-02	-1.941	0.0522 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3094.7 on 2268 degrees of freedom
Residual deviance: 3084.9 on 2264 degrees of freedom
AIC: 3094.9

Number of Fisher Scoring iterations: 4