

1 Interactions

2 Logistic Regression

1 Interactions

2 Logistic Regression

thesis

Code ▼

Kamil Zukowski (S2732947)

Generated on: Mon Jun 18 17:00:56 2018

Hide

Hide

```
mydf <- read.csv("final.csv") # load data
names(mydf)
```

```
## [1] "X" "FOMO"
## [3] "FUD" "HODL"
## [5] "MOON" "URL"
## [7] "Unnamed..0" "Unnamed..0.1"
## [9] "archived" "author"
## [11] "author_flair_css_class" "author_flair_text"
## [13] "body" "controversiality"
## [15] "created_utc" "distinguished"
## [17] "downs" "gilded"
## [19] "id" "link_id"
## [21] "name" "parent_id"
## [23] "pump" "question"
## [25] "removal_reason" "retrieved_on"
## [27] "score" "score_hidden"
## [29] "sentiment" "smiley_neg"
## [31] "smiley_pos" "subreddit"
## [33] "subreddit_id" "ups"
```

Hide

Hide

```
head(mydf)
```

```
## X FOMO FUD HODL MOON URL Unnamed..0 Unnamed..0.1 archived
## 1 0 0 0 0 0 0 0 1
## 2 1 0 0 0 0 0 1 1
## 3 2 0 0 0 0 0 2 1
## 4 3 0 0 0 0 0 3 1
## 5 4 0 0 0 0 0 4 1
```

```
## 6 5 0 0 0 0 5 1
## author author_flair_css_class author_flair_text
## 1 The_Evil_Within
## 2 dylanbob5
## 3 __Cyber_Dildonics__
## 4 randomhuman99
## 5 coinoleum
## 6 dudetalking
##
```

body

```
## 1
```

If it's inconvenient enough, and the practical amounts are limited by that inconvenience... it's no longer much of an investment vehicle and it has zero worth as a daily practical currency.\n\nI'm not saying a government could wipe out bitcoin entirely, but they could easily make it so inconvenient that it would be effectively dead for the average person, and useless to those who already have means. They could drive it back to a black market currency of limited circulation.

```
## 2
```

He paid in bitcoin

```
## 3
```

That is all interesting but I think if anything we are to the 1990 stage of bitcoin, it is just hard to tell since because of the internet we know about every small step. \n\n\nI think the biggest aspect though is that people can believe and understand that you can be connected on a fast computer network to the rest of the world, but people still don't really believe that bitcoin works, let alone even understand it. But that is the real breakthrough, it does work, because it is working even with billions at stake. And the price has gone up and down like an internet stock, but adoption is steadily gaining all the time. Most people in the US have heard of it, they just have no idea what it is, how it works, and if it is safe.\n\n\nCommunication and media are definitely sexier and more interesting, and no one can make many promises with bitcoin right now because the price volatility and security track record mean that telling the average person about it implies informing them of a gamble. \n\n\nIt also kicks a lot of banks and credit card services to the curb, so they have no desire for people to adopt it. So basically I think adoption will go much faster because of the internet, and the fact that there is money to be made, but the fact that talking about money is not right now seen as something people want to deal with. \n\n\nThere may come a time very soon though, when advertisements on TV start talking about everyone becoming an entrepreneur and easily selling some sort of service to people. It is still very early, we don't have ATMs, ATM/credit cards, hardware wallets, physical locations or a lot of the things needed to make it easy.

```
## 4
```

Crypto believer. I believe in them all because they all exist in reality. I have only bought real world items with bitcoin because bitpay, and slowly starting to use GoCoin for LTC\\Doge. I have bought groceries and pay a few bills for sure. Localbitcoins is good for that since it promotes crypto locally, and you can always mention DOGE or LTC & GoCoin to the retailer to promote it. \n\n\nIf people are attracted to the community in general, they will find their niche be it charitable causes or paying the bills in an open, transparent, trustless way.\n\n\nHolding any coin is a good bet IMO, but the marketcaps are still small. Until we can show the huge whales of the world that it's beneficial to business to take these thin

gs into account, I'd say the whole community is at risk. \n\nThis is where self-regulation comes in until the big whales come in and try to fork the protocol to something more centrally regulated to their benefit.

5

China did ban bitcoin.\n\nBut this is the last time, seriously. No really. Last time banning bitcoin, they swear it. It's really banned this time. \n\n**It's triple-dog banned this time.**

6

The scarcity argument i believe works against bitcoin in a long time frame as a currency.\n\nBitcoin will become a store of value in that model, while another currency functions as the transfer medium.\n\nHowever the artificial scarcity of bitcoin will not determine its value.\n\nThere are plenty of "scarce" things that have no value.\n\n

```
##      controversiality created_utc distinguished downs gilded      id
## 1              0 1397152628              0      0 cgp472v
## 2              0 1397153908              0      0 cgp4t6s
## 3              0 1397139819              0      0 cgoyajb
## 4              0 1397094949              0      0 cgom05s
## 5              0 1397149200              0      0 cgp2iou
## 6              0 1397167623              0      0 cgpbh3r
```

```
##      link_id      name parent_id pump question removal_reason
## 1 t3_22oayg t1_cgp472v t1_cgp3ylj 1      0      NA
## 2 t3_22ois9 t1_cgp4t6s t1_cgoutbe 1      0      NA
## 3 t3_22n39j t1_cgoyajb t1_cgotmro 1      0      NA
## 4 t3_22niq9 t1_cgom05s t1_cgollpe 1      0      NA
## 5 t3_22ols2 t1_cgp2iou t1_cgov6bo 1      0      NA
## 6 t3_22p5un t1_cgpbh3r t1_cgp78rn 1      0      NA
```

```
##      retrieved_on score score_hidden sentiment smiley_neg smiley_pos
## 1 1433296793      1      0 -0.8873      0      0
## 2 1433297083      1      0 0.0000      0      0
## 3 1433293945      1      0 0.9851      0      0
## 4 1433288087      1      0 0.9393      0      0
## 5 1433296003      8      0 -0.9462      0      0
## 6 1433300291      1      0 0.3612      0      0
```

```
##      subreddit subreddit_id ups
## 1 Bitcoin t5_2s3qj 1
## 2 AskReddit t5_2qh1i 1
## 3 BitcoinMarkets t5_2wwh3 1
## 4 dogecoin t5_2zcp2 1
## 5 Bitcoin t5_2s3qj 8
## 6 BitcoinMarkets t5_2wwh3 1
```

Hide

Hide

summary(mydf)

```
##      X      FOMO      FUD      HODL
## Min. : 0 Min. :0.000000 Min. :0.00000 Min. :0.00000
## 1st Qu.:13941 1st Qu.:0.000000 1st Qu.:0.00000 1st Qu.:0.00000
## Median :27883 Median :0.000000 Median :0.00000 Median :0.00000
## Mean :28598 Mean :0.001739 Mean :0.00772 Mean :0.01132
```

```

## 3rd Qu.:41825    3rd Qu.:0.000000    3rd Qu.:0.00000    3rd Qu.:0.00000
## Max.      :64693    Max.      :1.000000    Max.      :1.00000    Max.      :1.00000
##
##          MOON          URL          Unnamed..0          Unnamed..0.1
## Min.      :0.000000    Min.      :0.0000    Min.      :    0    Min.      :    0
## 1st Qu.:0.000000    1st Qu.:0.0000    1st Qu.: 1386    1st Qu.:  542
## Median :0.000000    Median :0.0000    Median : 9594    Median : 1287
## Mean     :0.006877    Mean     :0.2534    Mean     :19620    Mean     : 1953
## 3rd Qu.:0.000000    3rd Qu.:1.0000    3rd Qu.:36810    3rd Qu.: 2760
## Max.     :1.000000    Max.     :1.0000    Max.     :64693    Max.     :10260
##
##          archived          author          author_flair_css_class
## Min.      :0.00    BitcoinAllBot          : 5815          :94764
## 1st Qu.:0.00    tippr          : 2757    noflair          : 3501
## Median :1.00    [deleted]          : 1904    noob          : 3207
## Mean     :0.71    censorship_notifier:  540    bulllongterm:  543
## 3rd Qu.:1.00    AutoModerator          :  431    Training          :  525
## Max.     :1.00    bitcointip          :  404    training          :  360
## NA's     :84947    (Other)          :99683    (Other)          : 8634
##          author_flair_text
##                      :96986
## redditor for 3 months: 1226
## Long-term Holder      :  543
## Bullish                :  357
## redditor for 2 months:  317
## redditor for 2 weeks  :  310
## (Other)                :11795
##
body
## **Bitcoin (BTC) Basic Info:** [Website](https://bitcoin.org) - r/Bitcoin - [Abstract](https://www.reddit.com/r/CryptoWikis/wiki/Bitcoin_basicinfo#wiki_short_description) - [History](https://www.reddit.com/r/CryptoWikis/wiki/Bitcoin_basicinfo#wiki_history) - [Exchanges](https://www.reddit.com/r/CryptoWikis/wiki/Bitcoin_basicinfo#wiki_exchanges) - [Wallets](https://www.reddit.com/r/CryptoWikis/wiki/Bitcoin_basicinfo#wiki_wallets) \n\n**Biases:** [Arguments For](https://www.reddit.com/r/CryptoWikis/wiki/Bitcoin_pros) & [Arguments Against](https://www.reddit.com/r/CryptoWikis/wiki/Bitcoin_cons) | **CryptoWikis:** [Policy](https://www.reddit.com/r/CryptoCurrency/wiki/policy#wiki_cryptowikis_program_-_28work_in_progress.29) - [Contribute Content](https://www.reddit.com/r/CryptoWikis/wiki/contributing)\n\n---\n\n\n*I am a bot, and this action was performed automatically. Please [contact the moderators of this subreddit](/message/compose/?to=/r/CryptoCurrency) if you have any questions or concerns.*: 71
## It's called **Bitcoin Cash** or **BCC**/**XBC** abbreviated. [Why is this important?](https://www.reddit.com/r/btc/comments/6r4no6/its_called_bitcoin_cash_the_term_bcash_is_a/)\n\n**BCC** and **XBC** are the recommended ticker symbols on the [official website](https://www.bitcoincash.org) and used by all 4 current client implementations.\n
: 70
## This is good for bitcoin.
: 61
## +1 Bitcoin. Verified.\n\n\nYou will be getting a bitcoin tip address shortly.\n\n\nThanks, have a merry Christmas
: 57

```

```

## This is good for bitcoin
: 48
## bitcoin
: 47
## (Other)
:111180
## controversiality created_utc distinguished downs
## Min. :0.00000 Min. :1.368e+09 :110948 Min. :0
## 1st Qu.:0.00000 1st Qu.:1.446e+09 admin : 3 1st Qu.:0
## Median :0.00000 Median :1.507e+09 moderator: 583 Median :0
## Mean :0.02252 Mean :1.479e+09 Mean :0
## 3rd Qu.:0.00000 3rd Qu.:1.512e+09 3rd Qu.:0
## Max. :1.00000 Max. :1.522e+09 Max. :0
## NA's :8494
7
## gilded id link_id name
## Min. :0.000000 dn02b2z: 2 t3_7g9cd3: 772 :849
47
## 1st Qu.:0.000000 dn02bcq: 2 t3_7hhpxs: 593 t1_c9rc1bc:
1
## Median :0.000000 dn02bpl: 2 t3_7lybxy: 357 t1_c9rc1px:
1
## Mean :0.001013 dn02c7s: 2 t3_7rguaj: 357 t1_c9rc1tg:
1
## 3rd Qu.:0.000000 dn02c9p: 2 t3_1tod7a: 329 t1_c9rc2rl:
1
## Max. :5.000000 dn02dsn: 2 t3_702o6b: 224 t1_c9rc5kw:
1
## (Other):111522 (Other) :108902 (Other) :265
82
## parent_id pump question removal_reason
## t3_7hhpxs: 492 Min. :0.00 Min. :0.0000 Mode:logical
## t3_1tod7a: 238 1st Qu.:0.00 1st Qu.:0.0000 NA's:111534
## t3_7g9cd3: 193 Median :1.00 Median :0.0000
## t3_7hmjdw: 103 Mean :0.58 Mean :0.3133
## t3_7gb4ou: 89 3rd Qu.:1.00 3rd Qu.:1.0000
## t3_1sbeq8: 70 Max. :1.00 Max. :1.0000
## (Other) :110349
## retrieved_on score score_hidden sentiment
## Min. :1.425e+09 Min. : -236.000 Min. :0 Min. : -0.99
96
## 1st Qu.:1.450e+09 1st Qu.: 1.000 1st Qu.:0 1st Qu.: -0.10
27
## Median :1.507e+09 Median : 1.000 Median :0 Median : 0.25
00
## Mean :1.488e+09 Mean : 4.577 Mean :0 Mean : 0.19
33
## 3rd Qu.:1.514e+09 3rd Qu.: 3.000 3rd Qu.:0 3rd Qu.: 0.64
74
## Max. :1.526e+09 Max. :17424.000 Max. :1 Max. : 0.99
97
## NA's :84947
## smiley_neg smiley_pos subreddit
## Min. :0.000000 Min. :0.00000 Bitcoin :36900

```

```
## 1st Qu.:0.000000 1st Qu.:0.000000 btc :12656
## Median :0.000000 Median :0.000000 BitcoinAll : 6352
## Mean :0.002753 Mean :0.01752 Cryptocurrency: 4741
## 3rd Qu.:0.000000 3rd Qu.:0.000000 BitcoinMarkets: 3965
## Max. :1.000000 Max. :1.000000 ethtrader : 2726
## (Other) :44194
## subreddit_id ups
## t5_2s3qj:36900 Min. : -155.00
## t5_2si5v:12656 1st Qu.: 1.00
## t5_39m7n: 6352 Median : 1.00
## t5_2wlj3: 4741 Mean : 3.65
## t5_2wwh3: 3965 3rd Qu.: 3.00
## t5_37jgj: 2726 Max. :3546.00
## (Other) :44194 NA's :76685
```

Hide

Hide

```
nrow(mydf)
```

```
## [1] 111534
```

Hide

Hide

```
ncol(mydf)
```

```
## [1] 34
```

Hide

Hide

```
#data <- subset(mydf,select=c(7,8,9,10,11,13,14,15,29,30,31,32,33))
```

Hide

Hide

```
#library(Amelia)
#library(mlbench)
#missmap(data, col=c("blue", "red"), legend=FALSE)

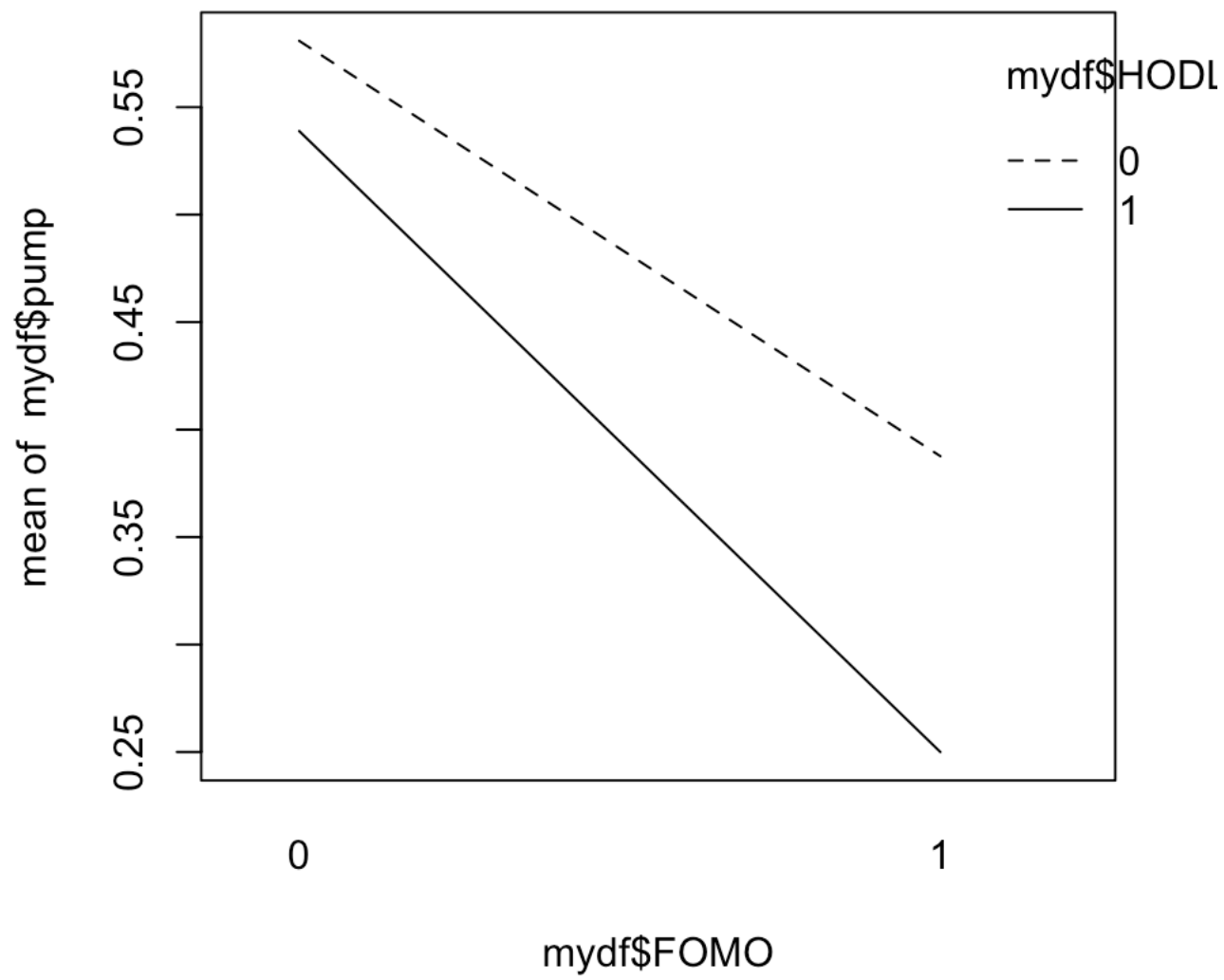
#library(corrplot)
#correlations <- cor(data[,1:13])
#corrplot(correlations, method="circle")
```

1 Interactions

Hide

Hide

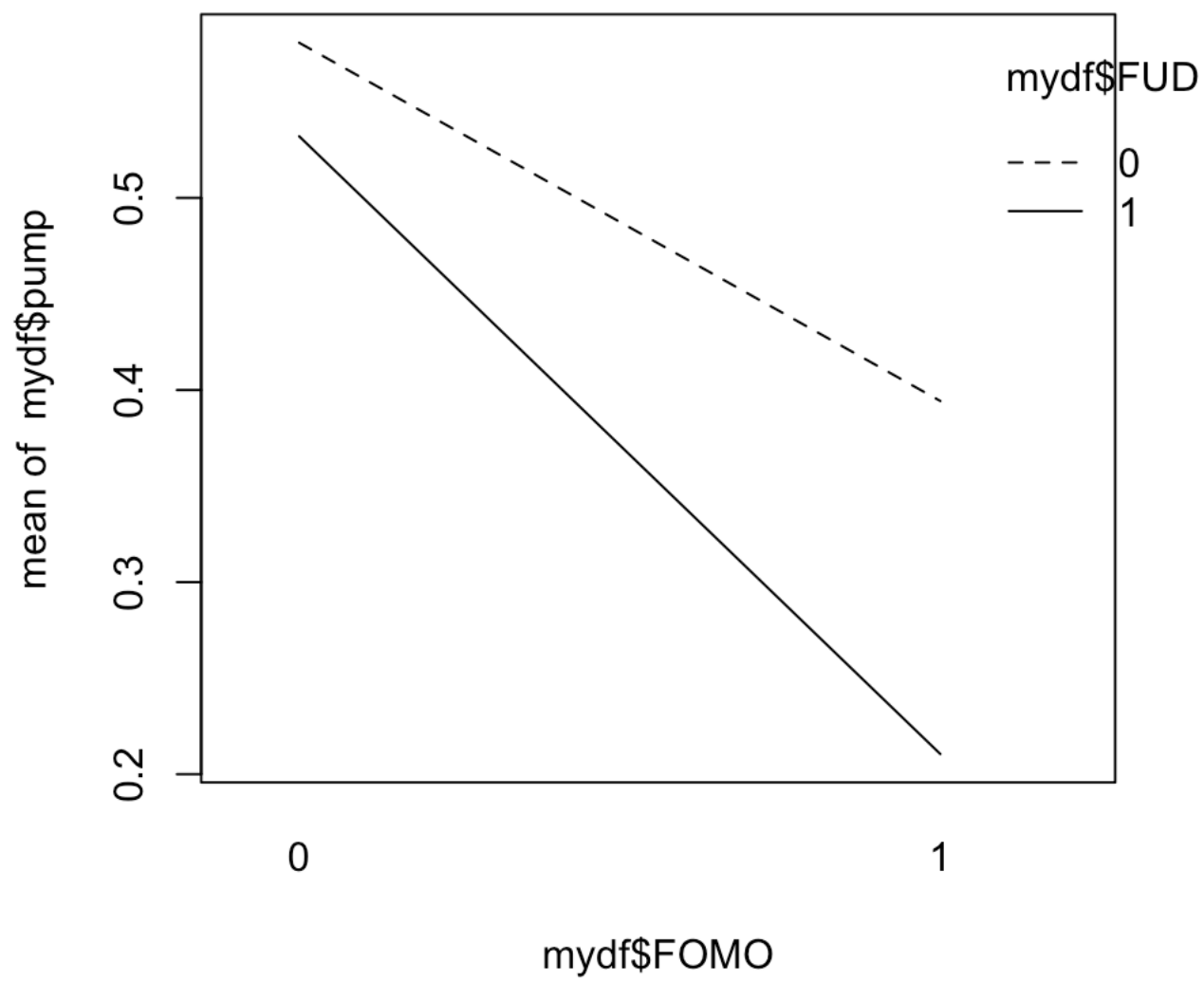
```
interaction.plot(mydf$FOMO, mydf$HODL, mydf$pump)
```



Hide

Hide

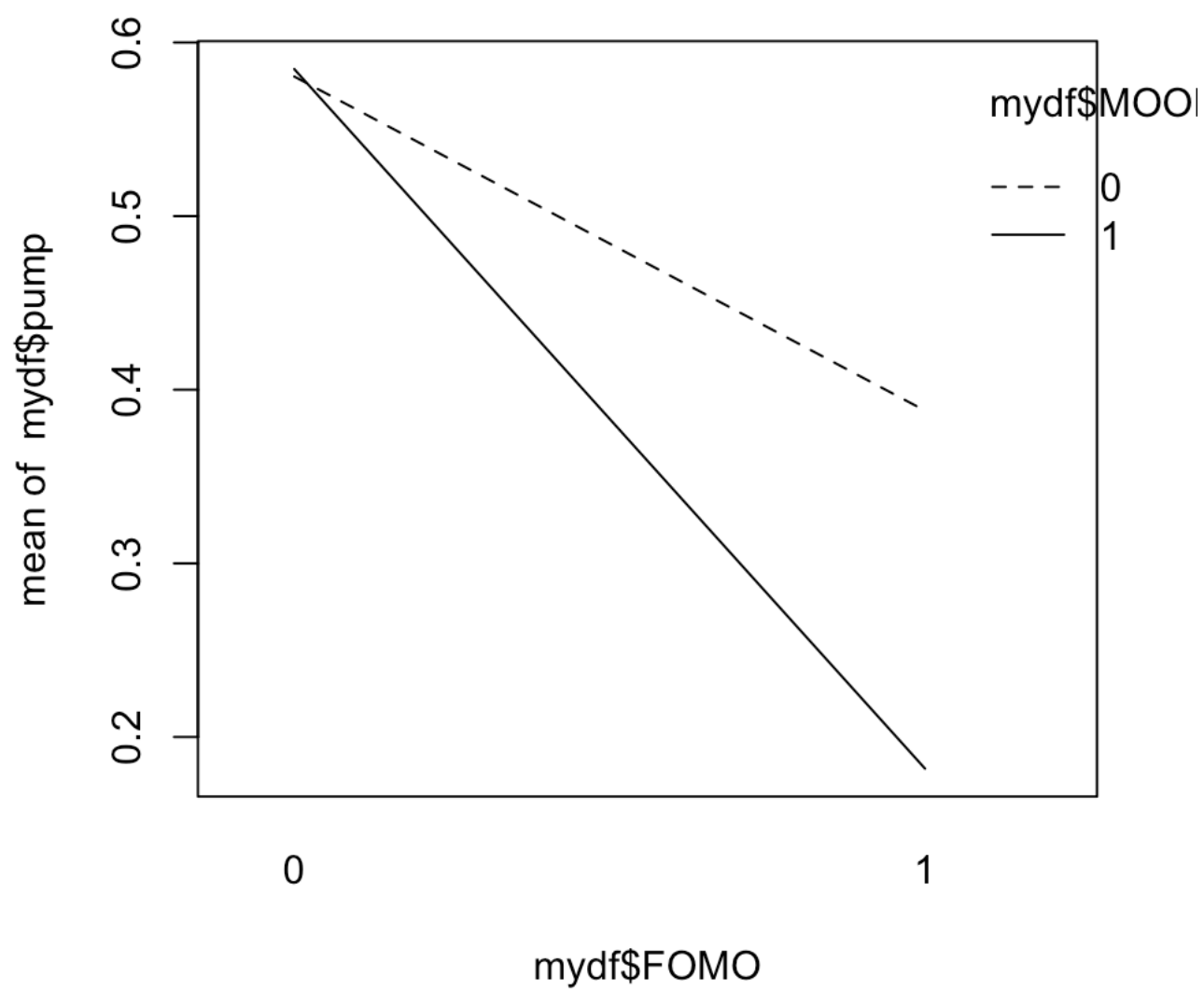
```
interaction.plot(mydf$FOMO, mydf$FUD, mydf$pump)
```



Hide

Hide

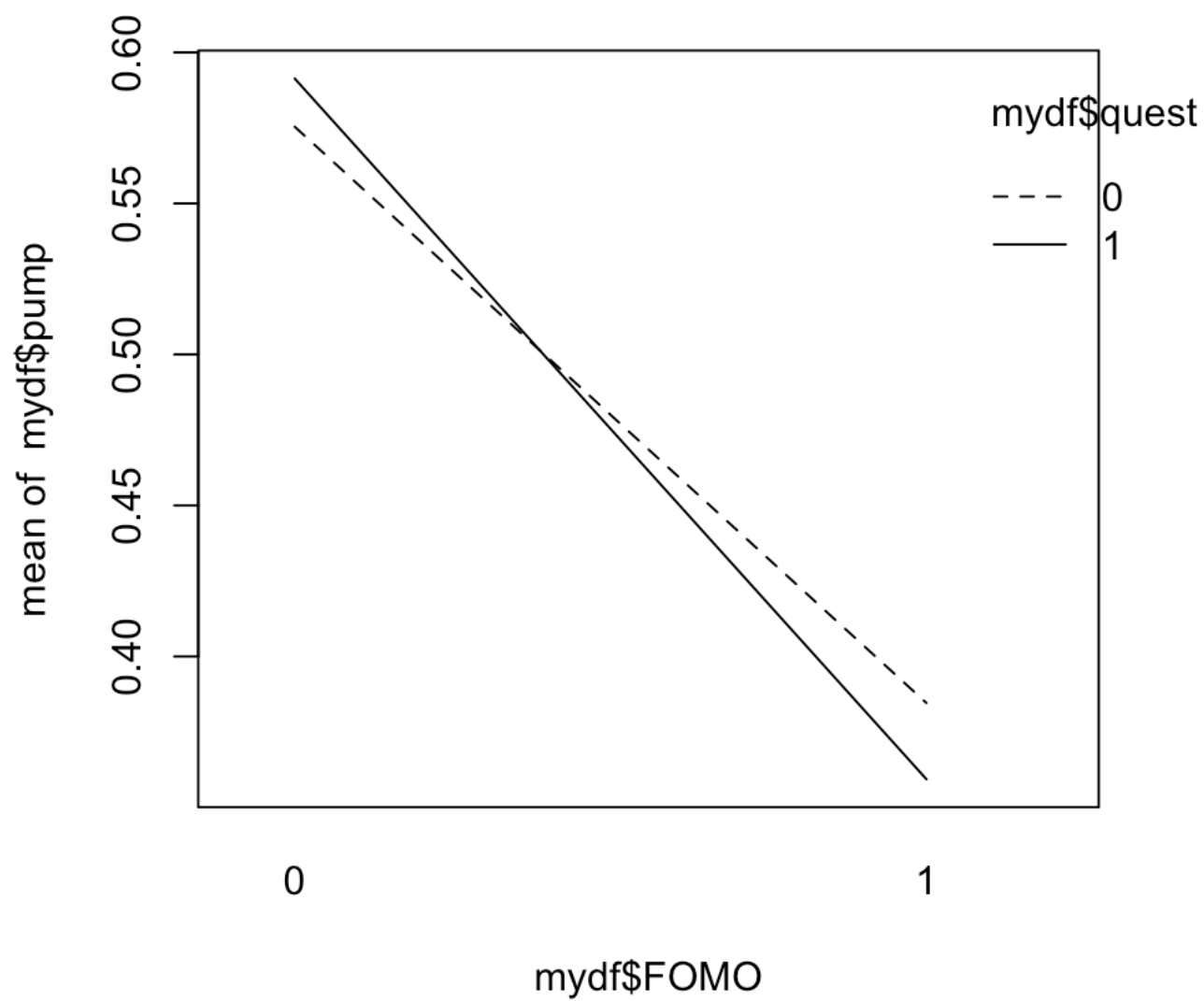
```
interaction.plot(mydf$FOMO, mydf$MOON, mydf$pump)
```

Hide

Hide

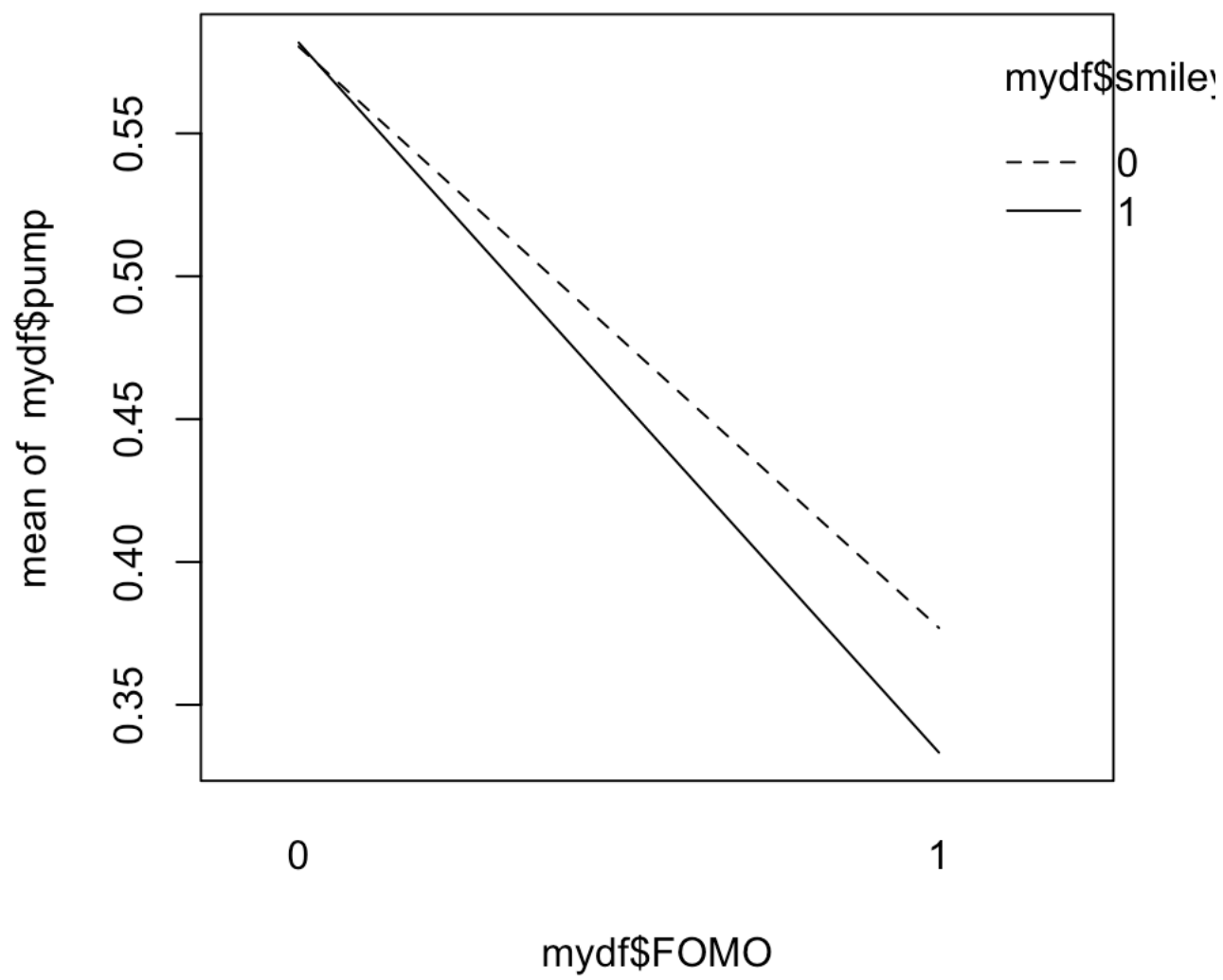
```
interaction.plot(mydf$FOMO, mydf$question, mydf$pump)
```



Hide

Hide

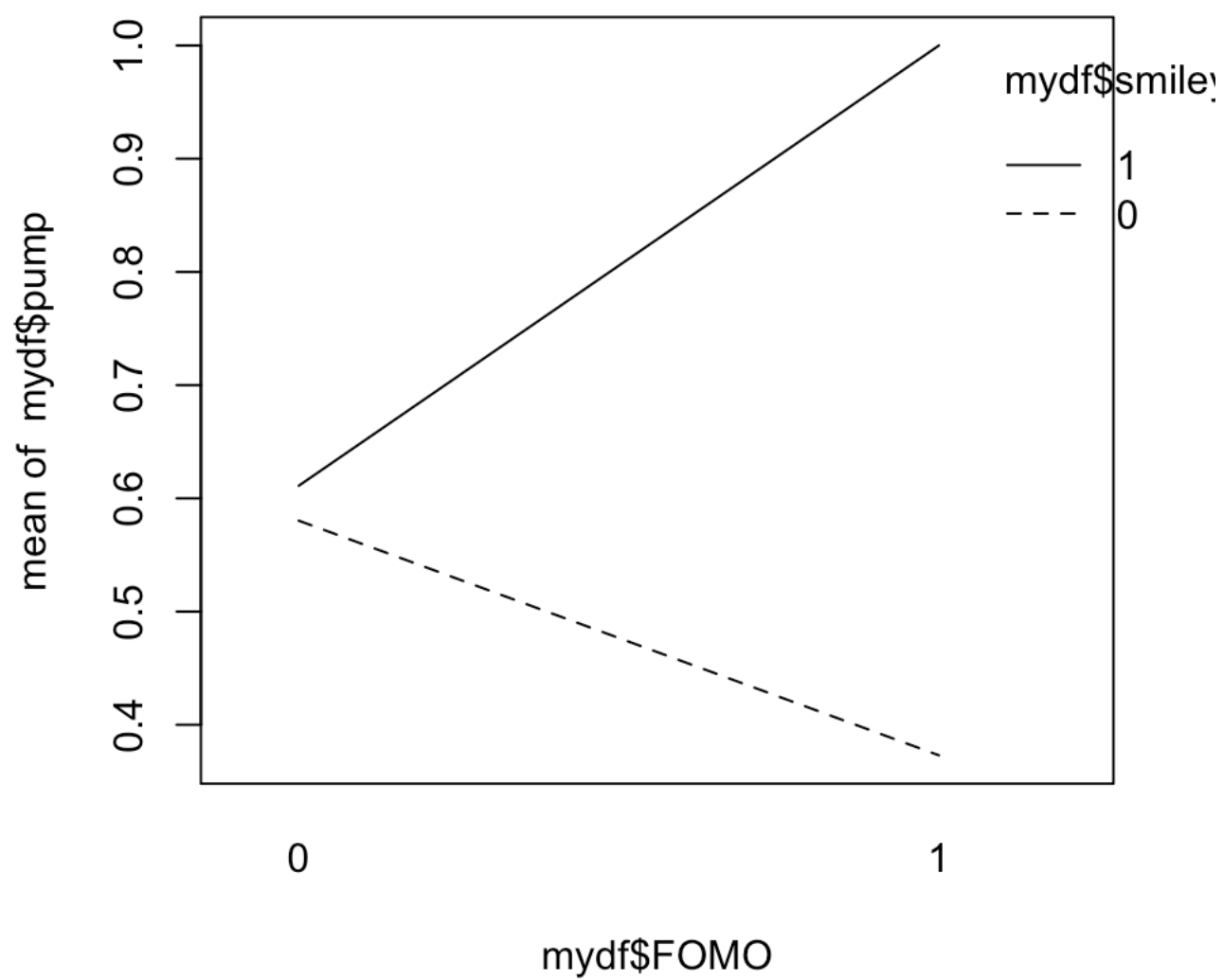
```
interaction.plot(mydf$FOMO, mydf$smiley_pos, mydf$pump)
```



Hide

Hide

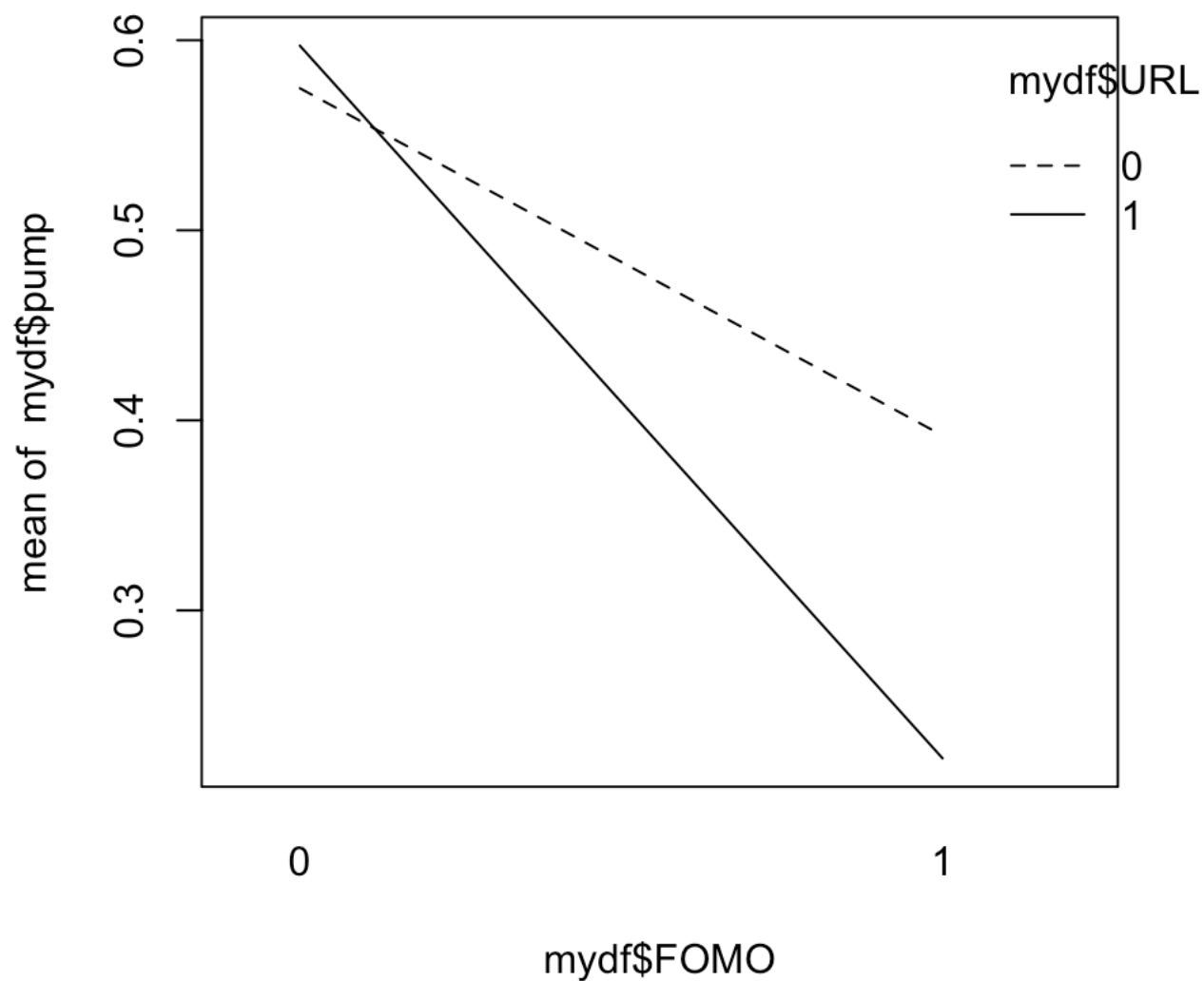
```
interaction.plot(mydf$FOMO, mydf$smiley_neg, mydf$pump)
```



Hide

Hide

```
interaction.plot(mydf$FOMO, mydf$URL, mydf$pump)
```



Hide

Hide

```
#interaction.plot(mydf$FOMO, mydf$HODL, mydf$URL, mydf$pump)

#interaction.plot(mydf$FOMO, mydf$HODL, mydf$FUD, mydf$pump)
```

2 Logistic Regression

Hide

Hide

```
# Logistics Regression
glm.fit <- glm(pump ~ sentiment + smiley_pos + smiley_neg + FOMO + MOON
+ HODL + FUD + question + URL, data = mydf, family = binomial)

summary(glm.fit)
```

```
##
## Call:
## glm(formula = pump ~ sentiment + smiley_pos + smiley_neg + FOMO +
##      MOON + HODL + FUD + question + URL, family = binomial, data = mydf
## )
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.438   -1.306    1.019    1.053    1.518
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.285698   0.008194  34.868 < 2e-16 ***
## sentiment    0.022496   0.011507   1.955  0.05058 .
## smiley_pos   -0.010804   0.046567  -0.232  0.81653
## smiley_neg    0.150325   0.117439   1.280  0.20054
## FOMO         -0.790225   0.148666  -5.315 1.06e-07 ***
## MOON          0.023845   0.073698   0.324  0.74628
## HODL         -0.169222   0.057013  -2.968  0.00300 **
## FUD          -0.199821   0.068895  -2.900  0.00373 **
## question      0.054216   0.013262   4.088 4.35e-05 ***
## URL           0.081281   0.014201   5.724 1.04e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 151749  on 111533  degrees of freedom
## Residual deviance: 151633  on 111524  degrees of freedom
## AIC: 151653
##
## Number of Fisher Scoring iterations: 4
```

[Hide](#)
[Hide](#)

```
glm.probs <- predict(glm.fit,type = "response")
glm.probs[1:5]
```

```
##           1           2           3           4           5
## 0.5660461 0.5709426 0.5763625 0.5761109 0.5657206
```

[Hide](#)
[Hide](#)

```
glm.pred <- ifelse(glm.probs > 0.5, "1", "0")
attach(mydf)
table(glm.pred,pump)
```

```
##          pump
## glm.pred    0    1
##          0   134   95
##          1 46706 64599
```

[Hide](#)[Hide](#)

```
mean(glm.pred == pump)
```

```
## [1] 0.580388
```

[Hide](#)[Hide](#)

```
anova(glm.fit, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: pump
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                111533    151749
## sentiment    1     6.206    111532    151742  0.012734 *
## smiley_pos   1     0.021    111531    151742  0.883536
## smiley_neg   1     1.518    111530    151741  0.217960
## FOMO         1    32.626    111529    151708 1.117e-08 ***
## MOON         1     0.002    111528    151708  0.963317
## HODL         1     9.868    111527    151698  0.001681 **
## FUD          1     7.684    111526    151691  0.005573 **
## question     1    24.892    111525    151666 6.065e-07 ***
## URL          1    32.840    111524    151633 1.001e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

[Hide](#)[Hide](#)

```
#train <- mydf[1:55766,]
#test  <- mydf[55767:111534,]

#model <- glm(pump ~ sentiment + smiley_pos + smiley_neg + FOMO + MOON +
HODL + FUD + question + URL, family = binomial(link='logit'),data=train)
#summary(model)
```