

Text Mining

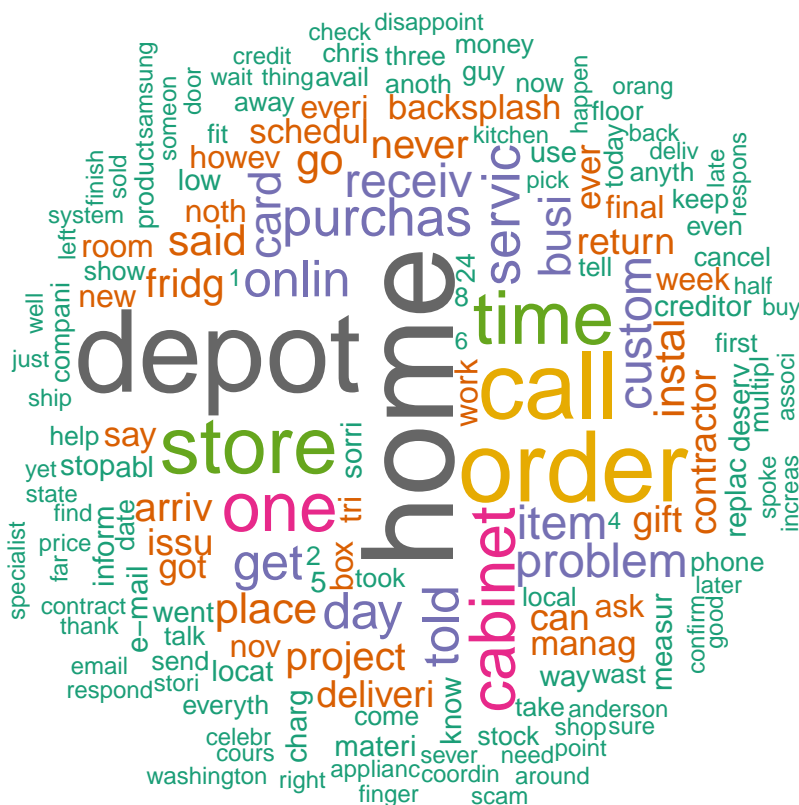
Cody Frisby

12/3/2017

I found customer complaints for Home Depot at https://www.consumeraffairs.com/homeowners/home_depot.htm.

I copied and pasted around 16 of them.

Below is a word cloud of all the combined complaints. The frequency the word appears is represented by the size.



It appears that the most common words are *home* and *depot* which makes sense :). If we remove those two words, the top ten words are in the below table. Most of the people appear to have complaints regarding the service at the store or online. They are also frequently upset about the quality of the products they buy.

words	freq
order	30
call	28
store	23
time	21
one	20
cabinet	18
servic	14
purchas	14
onlin	13
get	13

Thoughts on quanteda.

A (new-ish?) R package for text analysis.

I think it might be crashing my RStudio. Either that or my recent install of `Octave` (I spent all day Saturday studying the Fourier Transform and there was some `matlab` code I wanted to run) and all the packages I updated when downloading it. I'll have to go back and take a look at which binaries might be affected.

Overall, I like the package. It's easy to use, the documentation is well written and the companion website has many in-depth examples. A much more pleasant experience than the last time I attempted text analysis using the `tm` package. There does seem to be one thing that I haven't yet hacked out the details, that is it seems chop letters off some of the words. I'll have to figure out what's going on here.

R Code:

```
library(readtext)
library(quanteda)
library(wordcloud)
#df <- readtext("~/Documents/school/info3130/data/complaints.txt")
df2 <- readLines("~/Documents/school/info3130/data/complaints.txt")
c1 <- corpus(df2)
df <- dfm(c1, remove = stopwords("english"), remove_punct = TRUE,
          stem = TRUE)
#topfeatures(df)
textplot_wordcloud(df, random.order = FALSE,
                   rot.per = 0.25,
                   colors = RColorBrewer::brewer.pal(8, "Dark2"))
# top 10 after removing home and deopt
x <- as.data.frame(topfeatures(df, n = 12))
names(x) <- "freq"
x$words <- row.names(x)
x <- x[3:12, c("words", "freq")]
knitr::kable(x, row.names = FALSE)
```