R code Hypothesis Test

```
#read in csv file
getwd()
```

```
## [1] "/Users/joeathas/Documents/GitHub/ds710fall2018finalproject"
```

```
library(readr)
df <- read_csv("/Users/joeathas/Documents/GitHub/ds710fall2018finalproject/sentiments (2).csv")
```

```
## Parsed with column specification:
## cols(
##   Hashtag = col_character(),
##   Positive = col_integer(),
##   Negative = col_integer(),
##   Neutral = col_integer(),
##   `n Tweets` = col_integer()
## )
```

```
#make first column row names
result <- df[-1]
row.names(result) <- df$Hashtag
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
dfp<-result
```

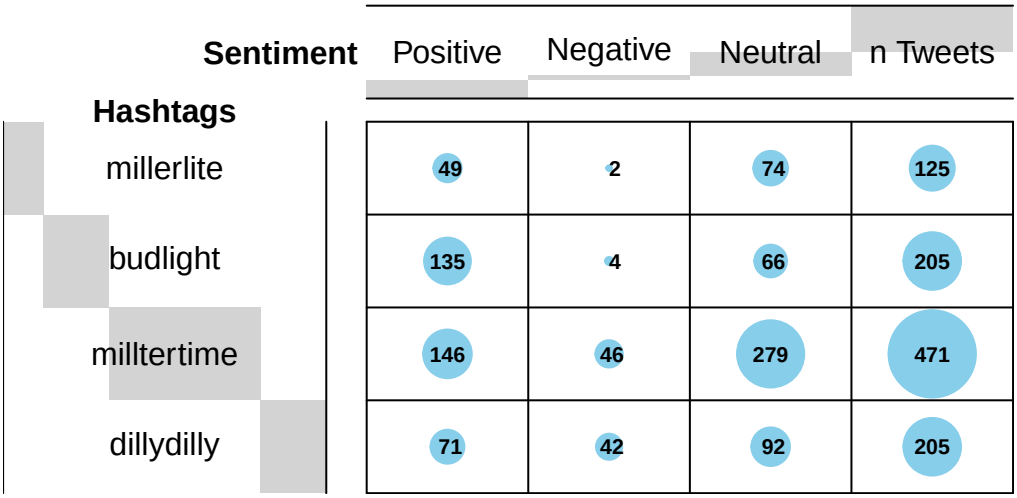
```
#Balloon Plot to visualize
library("gplots")
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
# 1. convert the data as a table
dt <- as.table(as.matrix(dfp))
# 2. Graph
balloonplot(t(dt), main ="Hashtag Sentiment Proportions", xlab ="Sentiment", ylab="Hashtags",  label.si
            label = T, show.margins = FALSE)
```

Hashtag Sentiment Proportions




```
#remove "n tweets" row in order to do analysis
dt<-dt[,1:3]
```

```
#Do Chi-Square
chisq <- chisq.test(dt)
chisq
```

```
##
##  Pearson's Chi-squared test
##
## data:  dt
## X-squared = 116.52, df = 6, p-value < 2.2e-16
```

```
#Do Post-Hoc Tests
library(fifer)
```

```
## Loading required package: xtable
```

```
## Loading required package: MASS
```

```
x<-chisq.post.hoc(dt)
```

```
## Adjusted p-values used the fdr method.
```

```
print(x)
```

```
##                   comparison  raw.p  adj.p
## 1    millerlite vs. budlight 0.0000 0.0000
## 2 millerlite vs. milltertime 0.0023 0.0023
## 3  millerlite vs. dillydilly 0.0000 0.0000
## 4   budlight vs. milltertime 0.0000 0.0000
## 5    budlight vs. dillydilly 0.0000 0.0000
## 6 milltertime vs. dillydilly 0.0001 0.0001
```

```
library(kableExtra)
#Import and make table of Observed proportions of hashtags
Proportions <- read_csv("/Users/joeathas/Documents/GitHub/ds710fall2018finalproject/Proportions.csv")
```

```
## Parsed with column specification:
## cols(
##   Hashtag = col_character(),
##   `Positive Tweet Proportion` = col_character(),
##   `Negative Tweet Proportion` = col_character(),
##   `Neutral Tweet Proportion` = col_character(),
##   `n Tweets` = col_integer()
## )
```

```
kable(Proportions,align = 'c')
```

Hashtag	Positive Tweet Proportion	Negative Tweet Proportion	Neutral Tweet Proportion	n Tweets
millerlite	39.20%	1.60%	59.20%	125
budlight	65.85%	1.95%	32.20%	205
millertime	30.99%	9.77%	59.24%	471
dillydilly	34.64%	20.49%	44.87%	205