# Analysis

2022-09-09

```
library(data.table)
library(rstatix)
```

```
##
## Attaching package: 'rstatix'

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

```
library(jsonlite)

results = data.table(fromJSON("~/CollaborativeVisualization/analysis/HM_responses_08_08.json", flatten=
voltage = data.table(read.csv2("~/CollaborativeVisualization/analysis/voltage_results.csv"))
freq = data.table(read.csv2("~/CollaborativeVisualization/analysis/freq_results.csv", sep=','))
conf = data.table(read.csv2("~/CollaborativeVisualization/data/participant_confidence_data.csv", sep=',
col = c('#386cb0','#fdbf6f')
names(col) = c("Glyphs", "Contours")
confidence=c(0,1,2,3)
names(confidence)=c("No Confidence", "Low Confidence", "Some Confidence", "High Confidence")
```

## Voltage Outliers

### ANOVA

```
analysis = voltage %>% gather(key = "condition", value="score", c, g) %>% convert_as_factor(hash_id, con
get_anova_table(anova_test(data = analysis, dv = score, wid = hash_id, within = condition, between=grou
```

```
## ANOVA Table (type II tests)
##
##            Effect DFn DFd       F        p p<.05      ges
## 1           group   1  28   2.873 1.01e-01       6.10e-02
## 2       condition   1  28 116.475 1.74e-11     * 6.03e-01
## 3 group:condition   1  28   0.002 9.61e-01       3.22e-05
```

### Paired t-test

Show condition effect

```
t.test(voltage[, c], voltage[, g], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  voltage[, c] and voltage[, g]
## t = -10.983, df = 29, p-value = 7.546e-12
## alternative hypothesis: true mean difference is not equal to 0
```

```
## 95 percent confidence interval:
##  -8.580323 -5.886344
## sample estimates:
## mean difference
##       -7.233333
```

```
t.test(voltage[group==1, c], voltage[group==1, g], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  voltage[group == 1, c] and voltage[group == 1, g]
## t = -7.7435, df = 14, p-value = 1.997e-06
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -9.279381 -5.253953
## sample estimates:
## mean difference
##       -7.266667
```

```
t.test(voltage[group==2, c], voltage[group==2, g], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  voltage[group == 2, c] and voltage[group == 2, g]
## t = -7.5221, df = 14, p-value = 2.784e-06
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -9.252943 -5.147057
## sample estimates:
## mean difference
##            -7.2
```

#### t-test

Also shows no interaction between question_order and condition

```
t.test(voltage[group==1, c], voltage[group==2, c])
```

```
##
##  Welch Two Sample t-test
##
## data:  voltage[group == 1, c] and voltage[group == 2, c]
## t = 1.3134, df = 27.993, p-value = 0.1997
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8207817  3.7541151
## sample estimates:
## mean of x mean of y
##  5.733333  4.266667
```

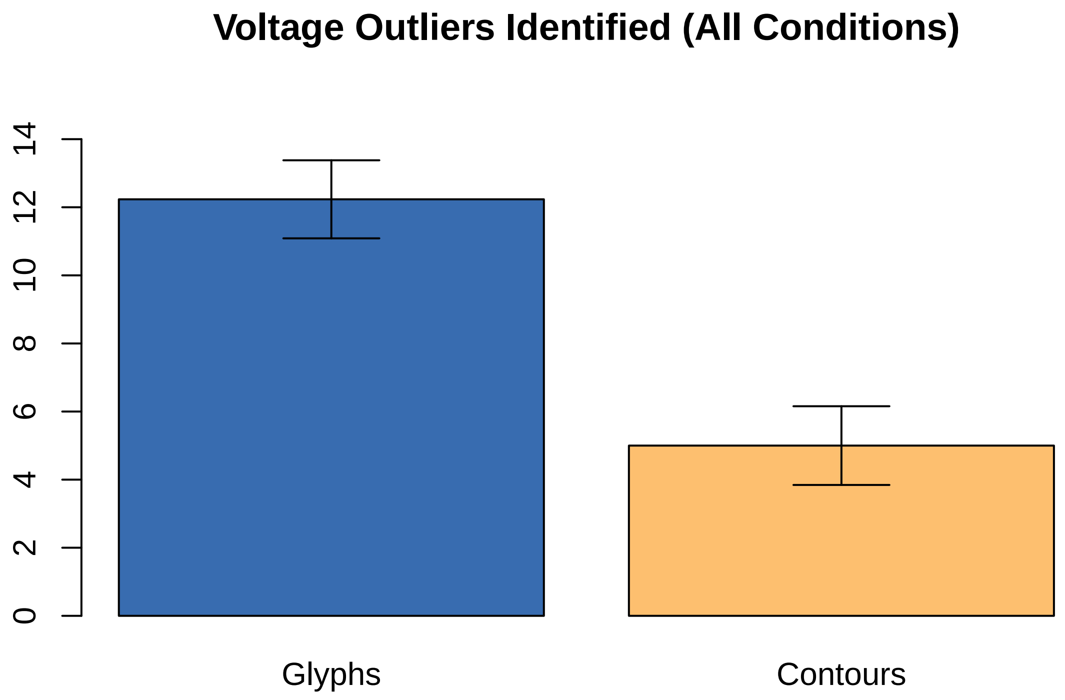
```
t.test(voltage[group==1, g], voltage[group==2, g])
```

```
##
##  Welch Two Sample t-test
##
```

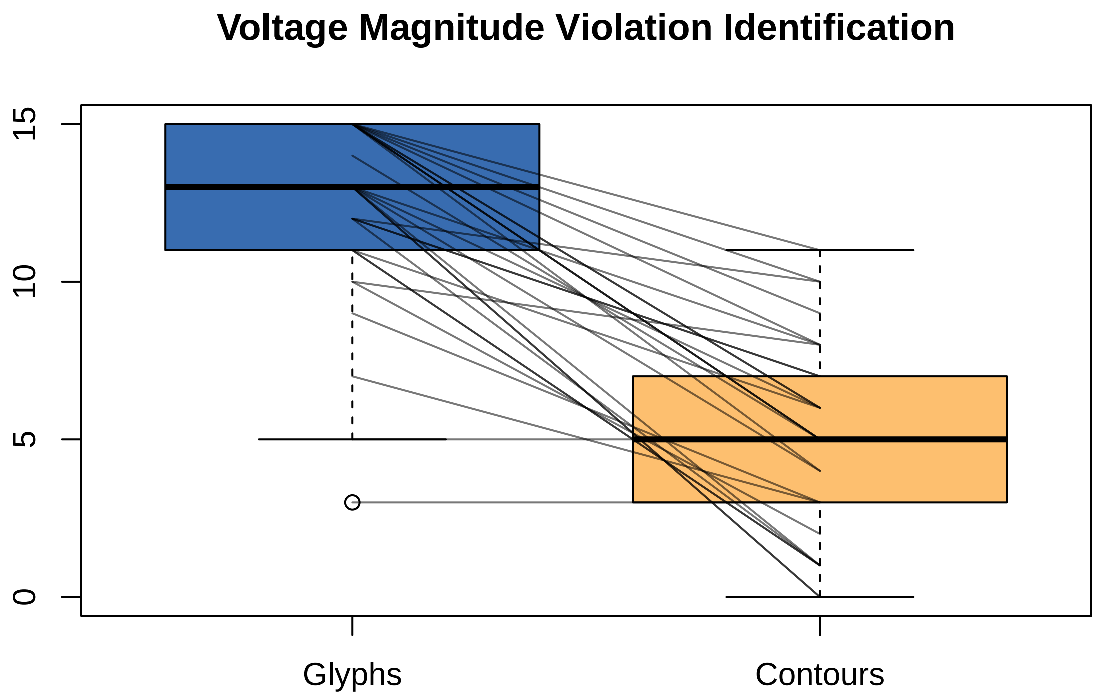
```
## data:  voltage[group == 1, g] and voltage[group == 2, g]
## t = 1.3895, df = 26.551, p-value = 0.1762
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7327139  3.7993806
## sample estimates:
## mean of x mean of y
##  13.00000  11.46667
```

### Plots

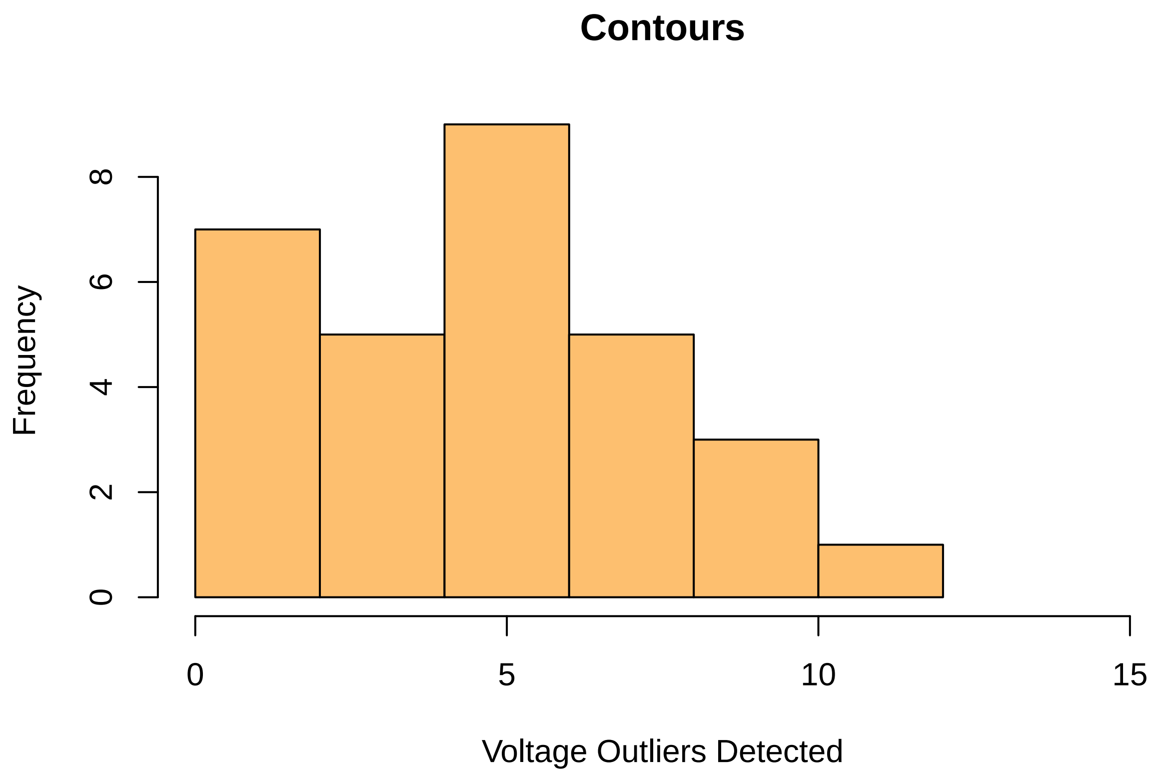
```
bars=barplot(c(voltage[, mean(g)], voltage[, mean(c)]), col=col, ylim=c(0,15), names.arg=names(col))
arrows(bars,
       y0=c(mean(voltage$g)-qt(0.975,df=29)*sd(voltage$g)/sqrt(30), mean(voltage$c)-qt(0.975,df=29)*sd(
       y1=c(mean(voltage$g)+qt(0.975,df=29)*sd(voltage$g)/sqrt(30), mean(voltage$c)+qt(0.975,df=29)*sd(
       angle=90, code=3)
title("Voltage Outliers Identified (All Conditions)")
```



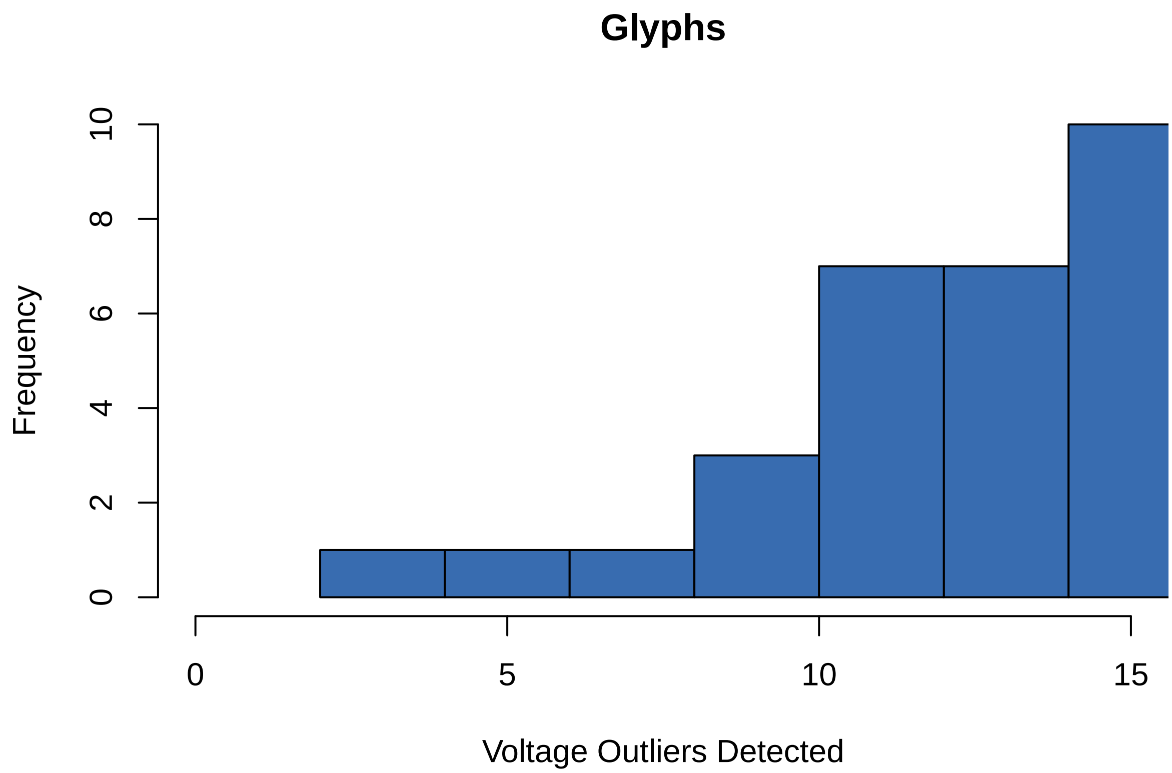

```
bars = boxplot(voltage$g, voltage$c, names=names(col), ylim=c(0,15), col=col, main="Voltage Magnitude Vi
segments(rep(1,30), voltage$g, rep(2,30), voltage$c, col="#00000088")
```



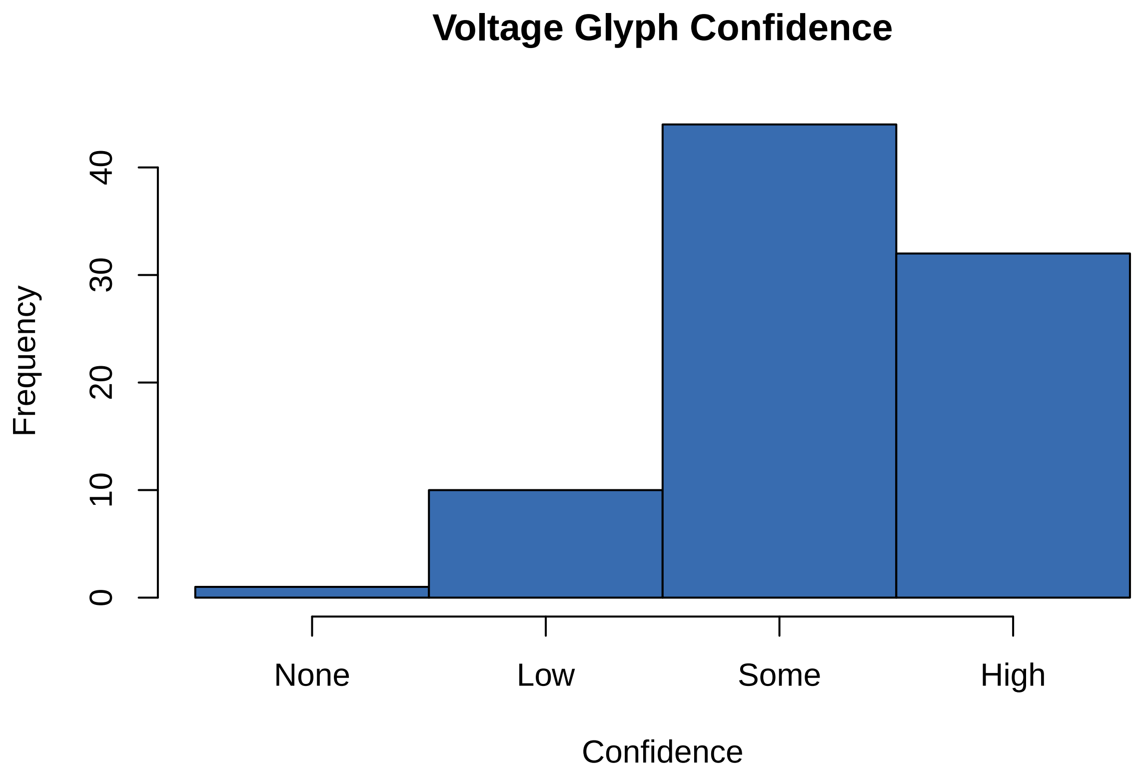

```
hist(voltage$c, xlim=c(0,15), col=col["Contours"], xlab="Voltage Outliers Detected", main="Contours")
```



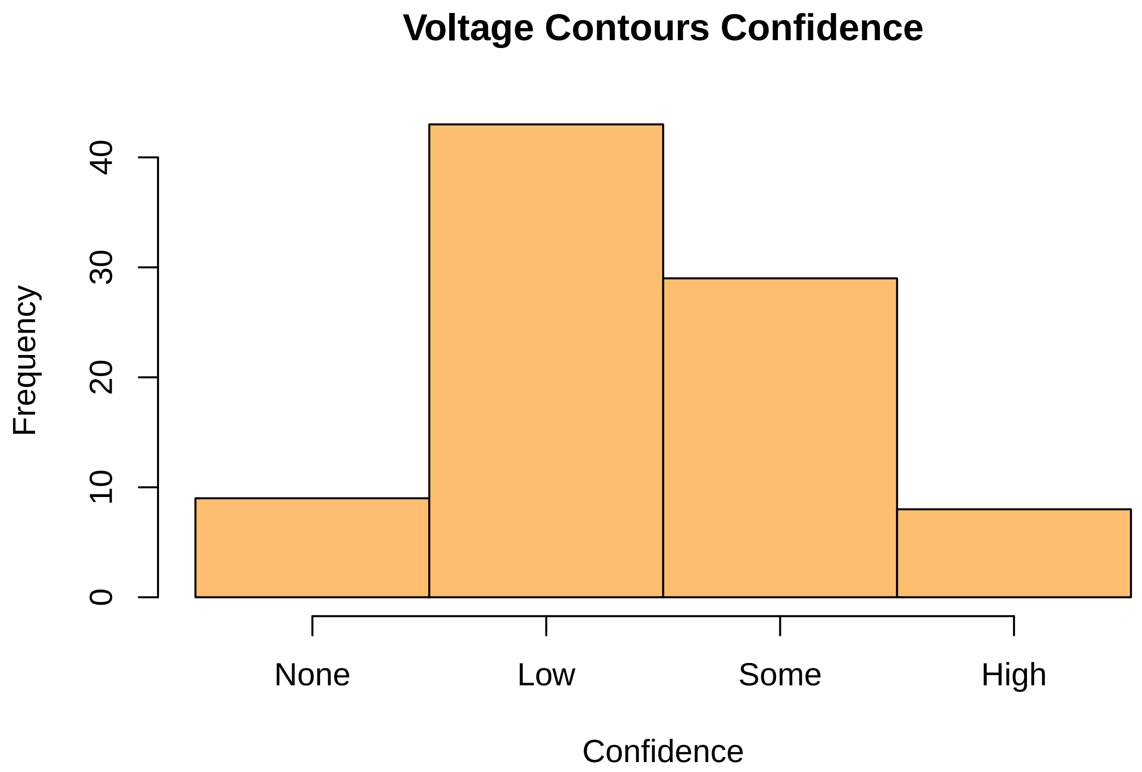

```r
hist(voltage$g, xlim=c(0,15), col=col["Glyphs"], xlab="Voltage Outliers Detected", main="Glyphs")
```



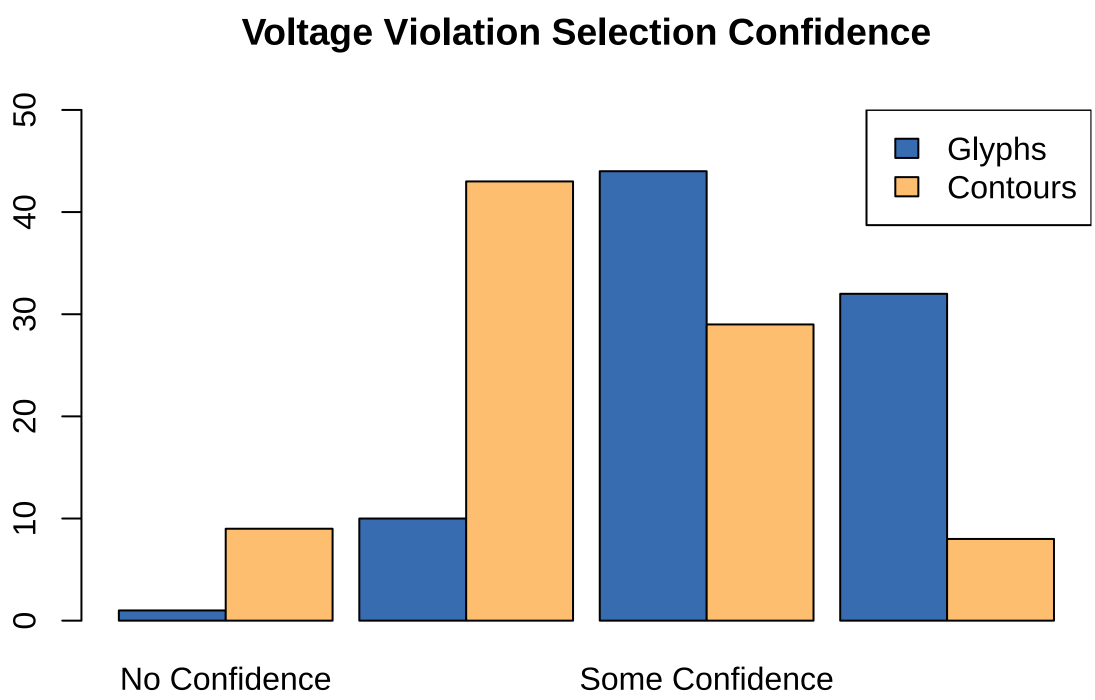

```
g=hist(confidence[c(unlist(results[question_order==2]$response_data.6.fields),unlist(results[question_o
axis(side=1,at=c(0,1,2,3),labels=c("None","Low","Some","High"))
```




```
c=hist(confidence[c(unlist(results[question_order==1]$response_data.6.fields),unlist(results[question_o
axis(side=1,at=c(0,1,2,3),labels=c("None","Low","Some","High"))
```




```
cdf = data.frame(None=c(glyph=g$counts[1], contour=c$counts[1]), Low=c(glyph=g$counts[2], contour=c$cou
barplot(height=as.matrix(cdf), beside=TRUE, col=c(col["Glyphs"], col["Contours"]), main="Voltage Violat
legend("topright", c("Glyphs", "Contours"), fill=c(col["Glyphs"], col["Contours"]))
```




**ANOVA Voltage False Positives**

```
analysis = voltage %>% gather(key = "condition", value="score", cf, gf) %>% convert_as_factor(hash_id, c
get_anova_table(anova_test(data = analysis, dv = score, wid = hash_id, within = condition, between=grou
```

```
## ANOVA Table (type II tests)
##
##              Effect DFn DFd     F     p p<.05      ges
## 1             group   1  28 0.005 0.946       9.74e-05
## 2         condition   1  28 1.509 0.230       2.10e-02
## 3 group:condition   1  28 4.888 0.035     * 6.60e-02
```

t-tests

```
t.test(voltage[, cf], voltage[, gf], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  voltage[, cf] and voltage[, gf]
## t = 1.1534, df = 29, p-value = 0.2582
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.3866278  1.3866278
## sample estimates:
## mean difference
##             0.5
```

```r
t.test(voltage[group==1, cf], voltage[group==2, cf])
```

```
##
##  Welch Two Sample t-test
##
## data:  voltage[group == 1, cf] and voltage[group == 2, cf]
## t = 1.7863, df = 19.496, p-value = 0.08962
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.158412  2.025079
## sample estimates:
## mean of x mean of y
## 1.4000000 0.4666667
```

```r
t.test(voltage[group==1, gf], voltage[group==2, gf])
```

```
##
##  Welch Two Sample t-test
##
## data:  voltage[group == 1, gf] and voltage[group == 2, gf]
## t = -1.1777, df = 14, p-value = 0.2586
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.4450713  0.7117379
## sample estimates:
## mean of x mean of y
## 0.0000000 0.8666667
```

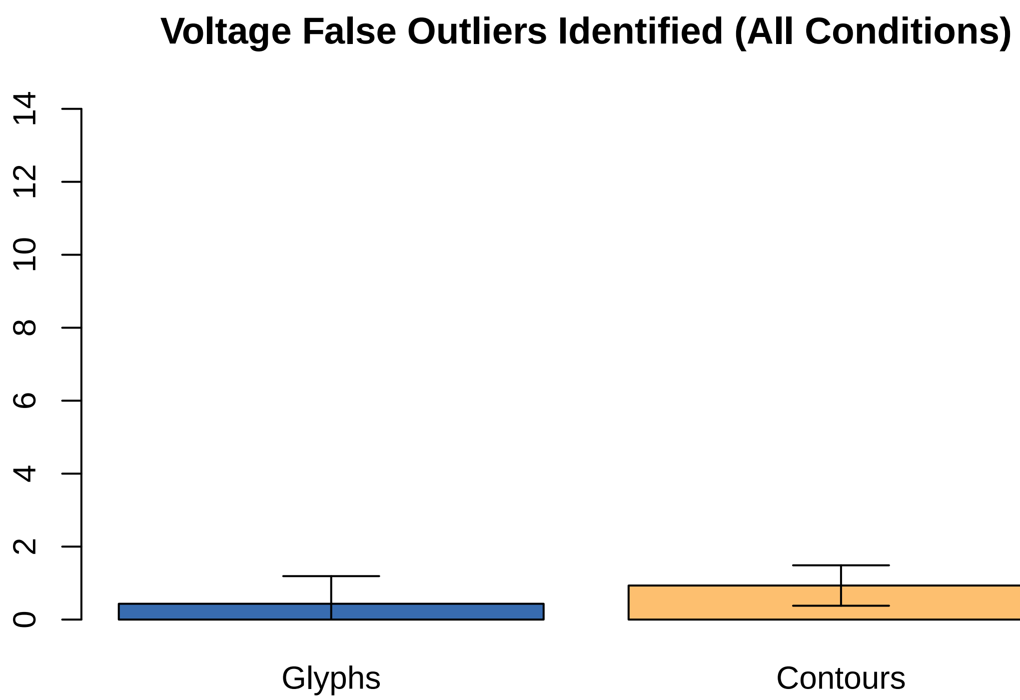
```r
t.test(voltage[group==1, cf], voltage[group==1, gf], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  voltage[group == 1, cf] and voltage[group == 1, gf]
## t = 2.9406, df = 14, p-value = 0.01074
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.3788773 2.4211227
## sample estimates:
## mean difference
##             1.4
```

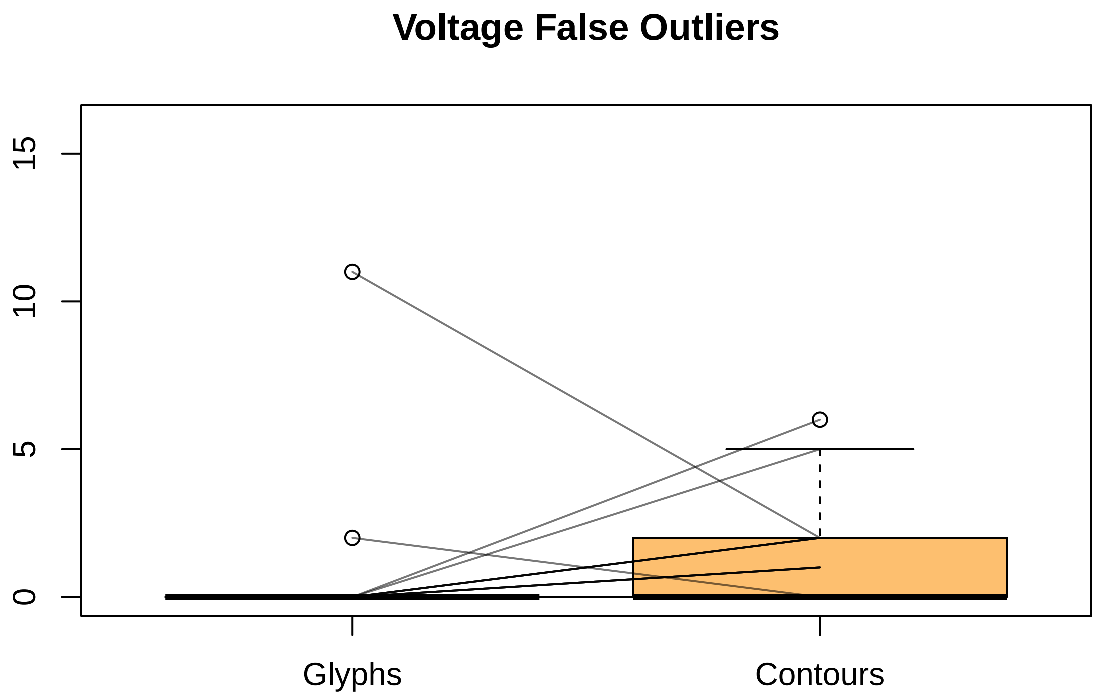
```r
t.test(voltage[group==2, cf], voltage[group==2, gf], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  voltage[group == 2, cf] and voltage[group == 2, gf]
## t = -0.60565, df = 14, p-value = 0.5544
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -1.816518  1.016518
## sample estimates:
## mean difference
##            -0.4
```

```
bars=barplot(c(voltage[, mean(gf)], voltage[, mean(cf)]), col=col, ylim=c(0,14), names.arg=names(col))
arrows(bars,
       y0=c(mean(voltage$gf)-qt(0.975,df=29)*sd(voltage$gf)/sqrt(30), mean(voltage$cf)-qt(0.975,df=29)*s
       y1=c(mean(voltage$gf)+qt(0.975,df=29)*sd(voltage$gf)/sqrt(30), mean(voltage$cf)+qt(0.975,df=29)*s
       angle=90, code=3)
title("Voltage False Outliers Identified (All Conditions)")
```




```
bars = boxplot(voltage$gf, voltage$cf, names=names(col), ylim=c(0,16), col=col, main="Voltage False Outl
segments(rep(1,30), voltage$gf, rep(2,30), voltage$cf, col="#00000088")
```




## Freq Outliers

```
freq[, id:=.I]
analysis = freq %>% gather(key = "condition", value="score", mags_gl, mags_hm) %>% convert_as_factor(id
get_anova_table(anova_test(data = analysis, dv = score, wid = id, within = condition, between=question_
```

```
## ANOVA Table (type II tests)
##
##                    Effect DFn DFd     F     p p<.05   ges
## 1          question_order   1  28 0.053 0.819       0.002
## 2               condition   1  28 7.254 0.012     * 0.019
## 3 question_order:condition   1  28 0.653 0.426       0.002
```

### Paired t-test

```
t.test(freq$mags_gl, freq$mags_hm, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  freq$mags_gl and freq$mags_hm
## t = 2.7096, df = 29, p-value = 0.01119
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.1634554 1.1698779
```

```
## sample estimates:
## mean difference
##       0.6666667
```

```
t.test(freq[question_order==1, mags_hm], freq[question_order==1, mags_gl], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  freq[question_order == 1, mags_hm] and freq[question_order == 1, mags_gl]
## t = -1.9441, df = 14, p-value = 0.07226
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -1.82279458  0.08946124
## sample estimates:
## mean difference
##      -0.8666667
```

```
t.test(freq[question_order==2, mags_hm], freq[question_order==2, mags_gl], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  freq[question_order == 2, mags_hm] and freq[question_order == 2, mags_gl]
## t = -2.1676, df = 14, p-value = 0.04792
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.928414915 -0.004918419
## sample estimates:
## mean difference
##      -0.4666667
```

**t-test**

```
t.test(freq[question_order==1, mags_hm], freq[question_order==2, mags_gl])
```

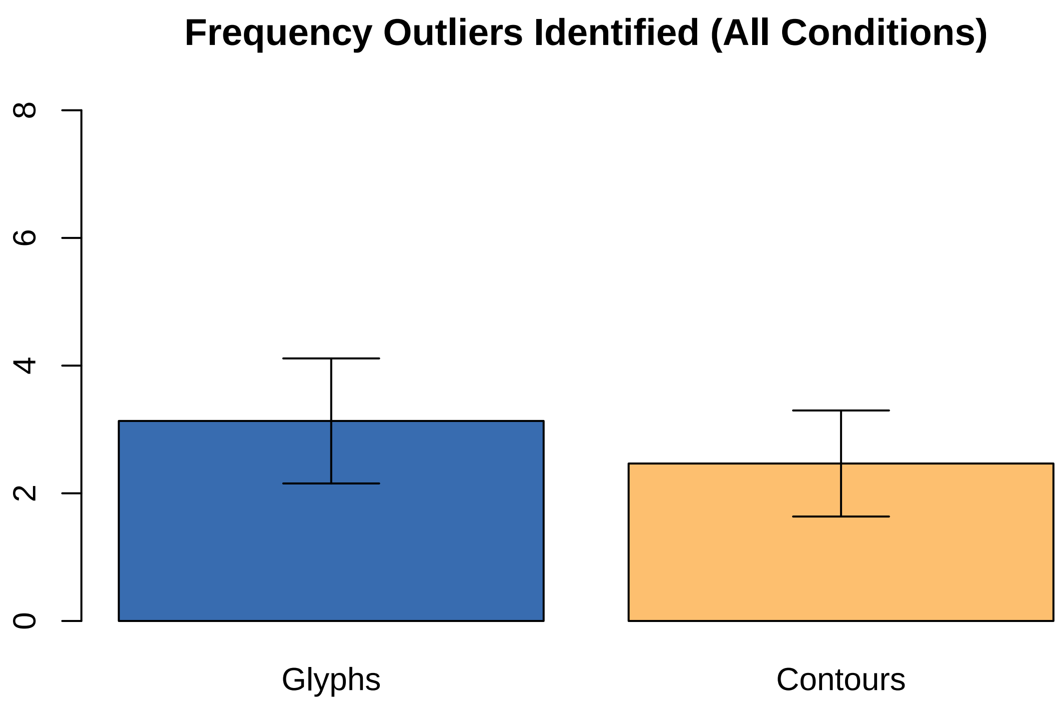
```
##
##  Welch Two Sample t-test
##
## data:  freq[question_order == 1, mags_hm] and freq[question_order == 2, mags_gl]
## t = -0.50879, df = 27.845, p-value = 0.6149
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.345954  1.412621
## sample estimates:
## mean of x mean of y
##  2.466667  2.933333
```

```
t.test(freq[question_order==1, mags_gl], freq[question_order==2, mags_gl])
```

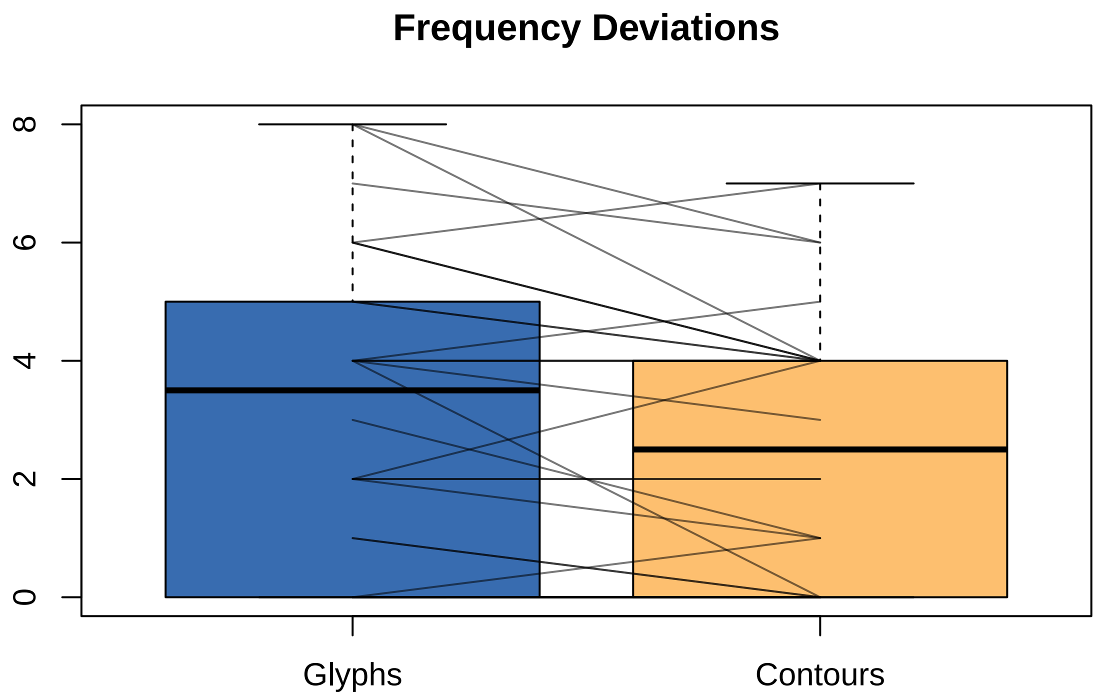
```
##
##  Welch Two Sample t-test
##
## data:  freq[question_order == 1, mags_gl] and freq[question_order == 2, mags_gl]
## t = 0.41167, df = 27.95, p-value = 0.6837
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -1.59052  2.39052
## sample estimates:
## mean of x mean of y
##  3.333333  2.933333
```

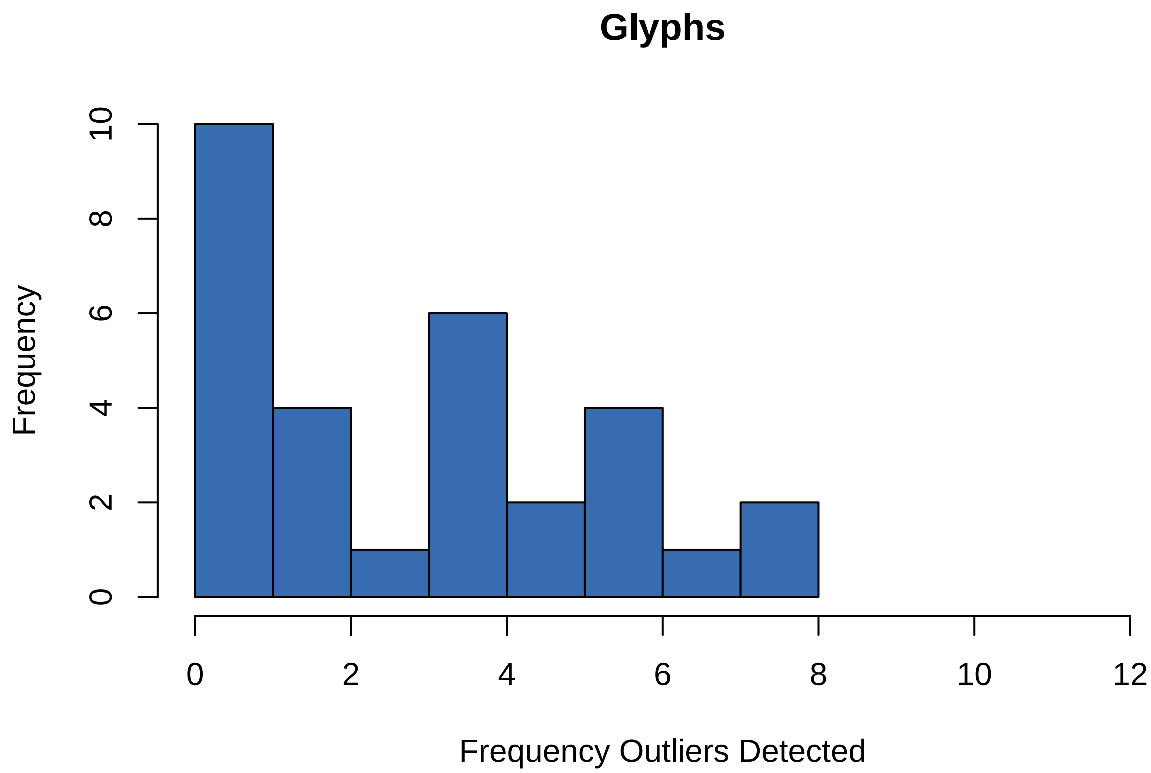
```
bars=barplot(c(freq[, mean(mags_gl)], freq[, mean(mags_hm)]), col=col, ylim=c(0,8), names.arg=names(col
arrows(bars,
       y0=c(mean(freq$mags_gl)-qt(0.975,df=29)*sd(freq$mags_gl)/sqrt(30), mean(freq$mags_hm)-qt(0.975,d
       y1=c(mean(freq$mags_gl)+qt(0.975,df=29)*sd(freq$mags_gl)/sqrt(30), mean(freq$mags_hm)+qt(0.975,d
       angle=90, code=3)
title("Frequency Outliers Identified (All Conditions)")
```



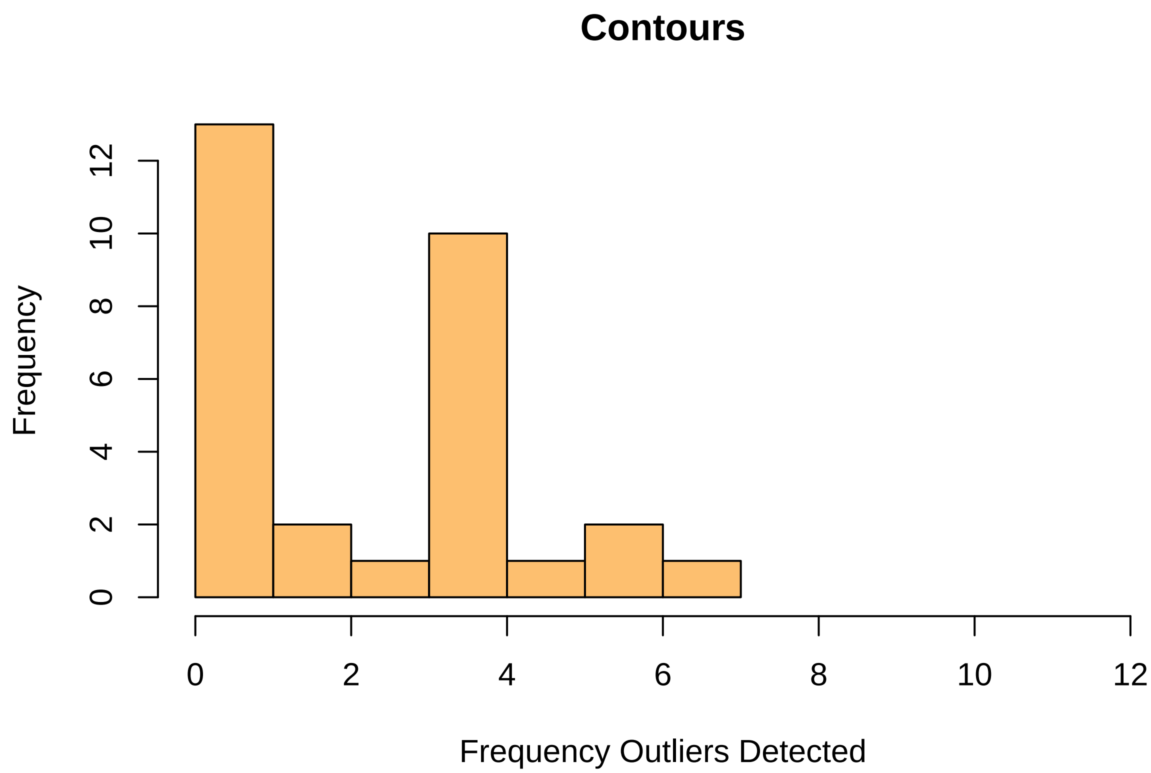

```
bars = boxplot(freq$mags_gl, freq$mags_hm, names=names(col), ylim=c(0,8), col=col, main="Frequency Devi
segments(rep(1,30), freq$mags_gl, rep(2,30), freq$mags_hm, col="#00000088")
```




```
hist(freq$mags_gl, xlim=c(0,12), col=col["Glyphs"], xlab="Frequency Outliers Detected", main="Glyphs")
```




```
hist(freq$mags_hm, xlim=c(0,12), col=col["Contours"], xlab="Frequency Outliers Detected", main="Contours
```




## Frequency Spreading

### ANOVA

```
analysis = freq %>% gather(key = "condition", value="score", spread_gl, spread_hm) %>% convert_as_facto
get_anova_table(anova_test(data = analysis, dv = score, wid = id, within = condition, between=question_
```

```
## ANOVA Table (type II tests)
##
##                   Effect DFn DFd     F     p p<.05   ges
## 1         question_order   1  28 0.686 0.415       0.017
## 2              condition   1  28 2.383 0.134       0.024
## 3 question_order:condition   1  28 2.383 0.134       0.024
```

```
t.test(freq$spread_gl, freq$spread_hm, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  freq$spread_gl and freq$spread_hm
## t = 1.5081, df = 29, p-value = 0.1423
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.1424481  0.9424481
## sample estimates:
## mean difference
```

```
##             0.4
```

```r
t.test(freq[question_order==1, spread_hm], freq[question_order==1, spread_gl], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  freq[question_order == 1, spread_hm] and freq[question_order == 1, spread_gl]
## t = -2.347, df = 14, p-value = 0.03416
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -1.53108752 -0.06891248
## sample estimates:
## mean difference
##            -0.8
```

```r
t.test(freq[question_order==2, spread_hm], freq[question_order==2, spread_gl], paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  freq[question_order == 2, spread_hm] and freq[question_order == 2, spread_gl]
## t = 0, df = 14, p-value = 1
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.837239  0.837239
## sample estimates:
## mean difference
##               0
```

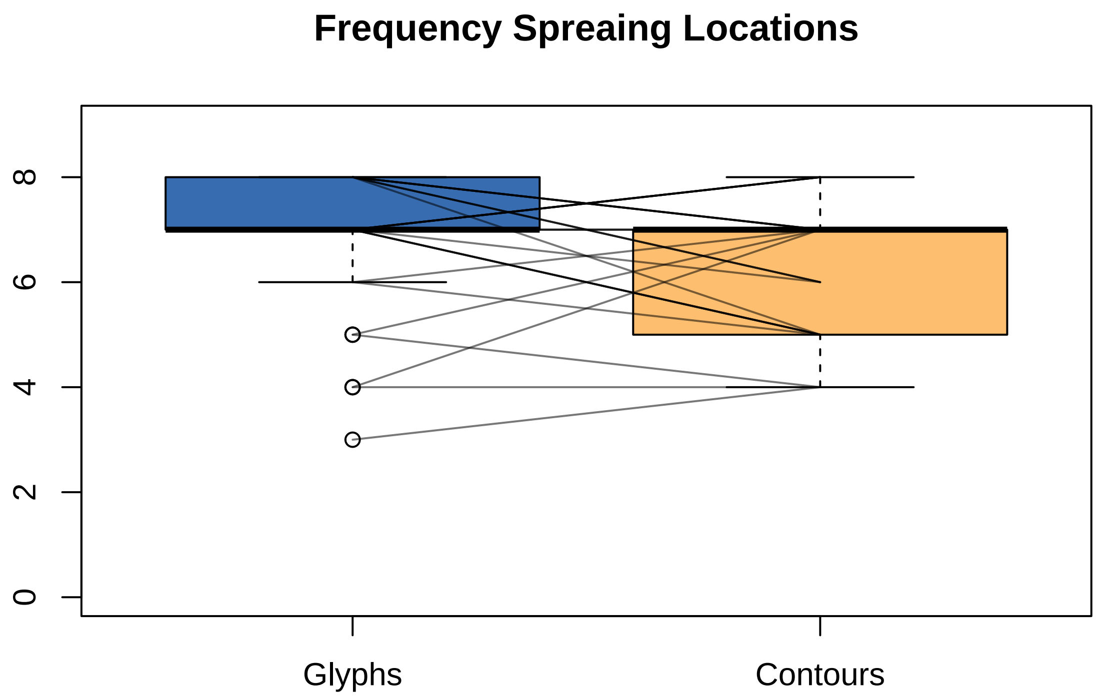
```r
t.test(freq[question_order==1, spread_hm], freq[question_order==2, spread_gl])
```

```
##
##  Welch Two Sample t-test
##
## data:  freq[question_order == 1, spread_hm] and freq[question_order == 2, spread_gl]
## t = -1.5192, df = 27.99, p-value = 0.1399
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.7221509  0.2554842
## sample estimates:
## mean of x mean of y
##  6.000000  6.733333
```

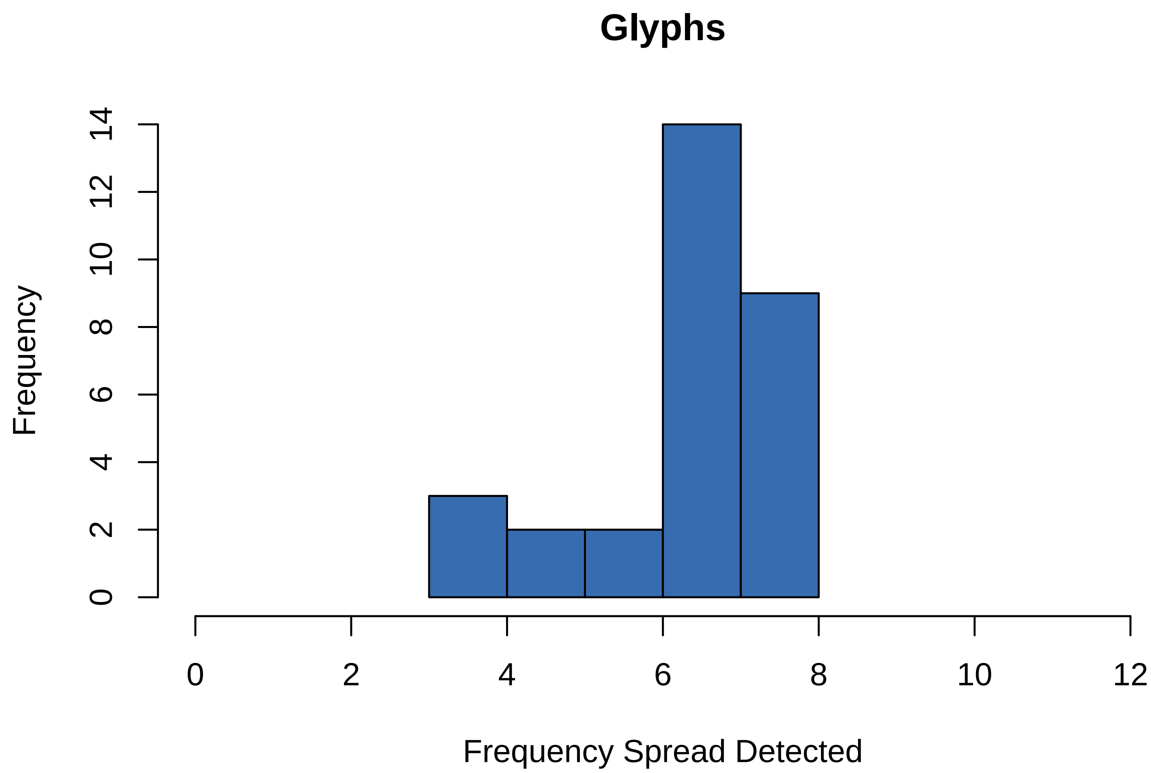
```r
t.test(freq[question_order==1, spread_gl], freq[question_order==2, spread_gl])
```

```
##
##  Welch Two Sample t-test
##
## data:  freq[question_order == 1, spread_gl] and freq[question_order == 2, spread_gl]
## t = 0.13484, df = 27.977, p-value = 0.8937
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.9461302  1.0794635
## sample estimates:
## mean of x mean of y
##  6.800000  6.733333
```

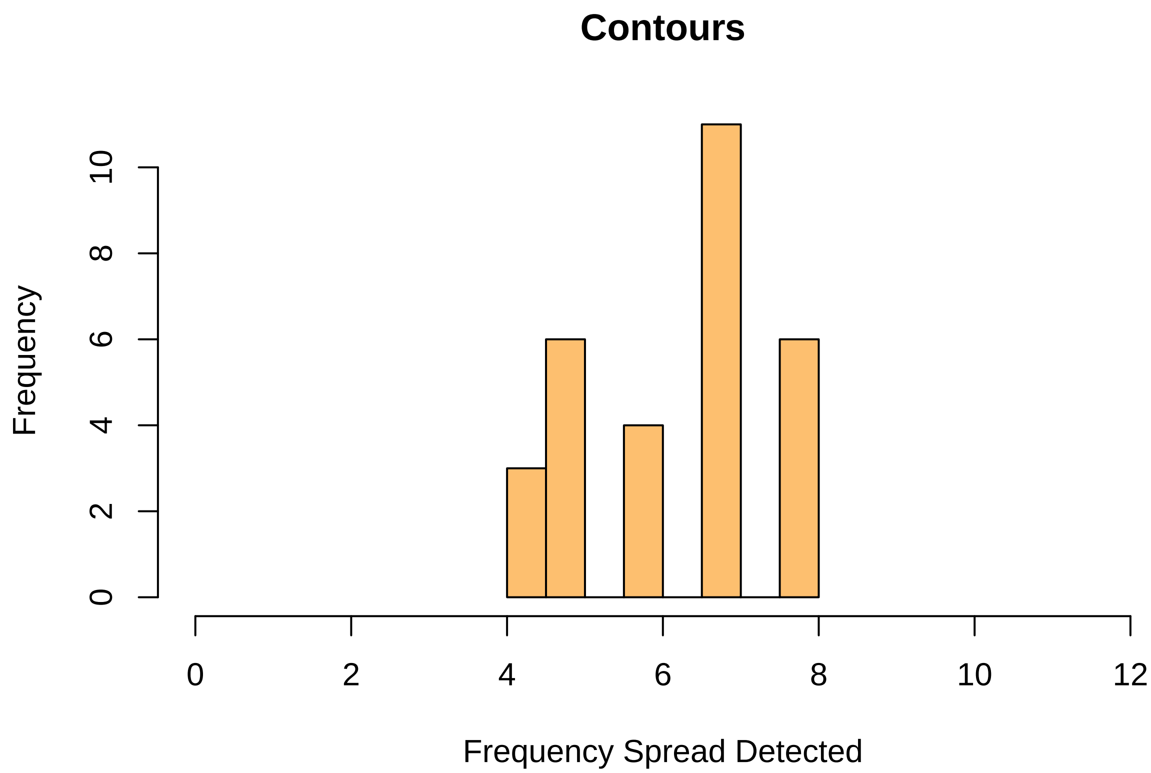
```
bars = boxplot(freq$spread_gl, freq$spread_hm, names=names(col), ylim=c(0,9), col=col, main="Frequency S
segments(rep(1,30), freq$spread_gl, rep(2,30), freq$spread_hm, col="#00000088")
```



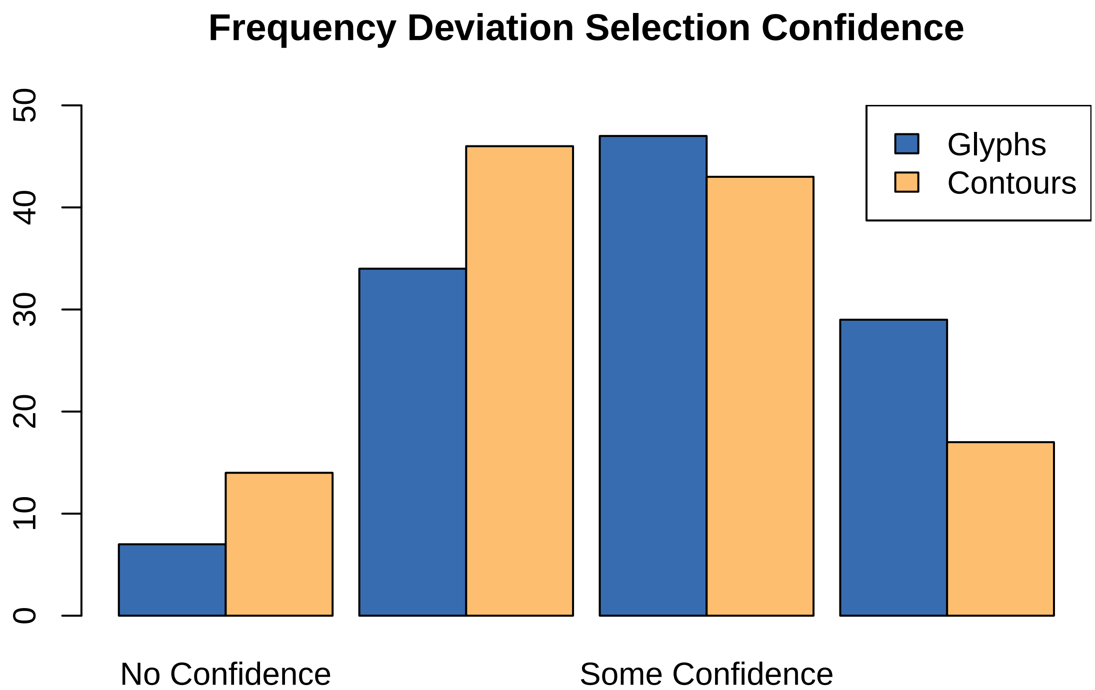

```
hist(freq$spread_gl, xlim=c(0,12), col=col["Glyphs"], xlab="Frequency Spread Detected", main="Glyphs")
```




```
hist(freq$spread_hm, xlim=c(0,12), col=col["Contours"], xlab="Frequency Spread Detected", main="Contours
```




```
cdf = data.frame("No Confidence"=c(glyph=conf[qtype=="gl", sum(mag_conf=="No Confidence")], contour=con
barplot(height=as.matrix(cdf), beside=TRUE, col=c(col["Glyphs"], col["Contours"]), main="Frequency Devia
legend("topright", c("Glyphs", "Contours"), fill=c(col["Glyphs"], col["Contours"]))
```


Frequency Deviation Selection Confidence
50
40
30
20
10
0
Glyphs
Contours
No Confidence
Some Confidence