1 Design

Each of the two weeks student received an email with links to the labs. Students within one team had the same version of the lab. They turn in team reports so only one person was working on producing the report (randomly selected "author" for the week) but both the TAs and myself observed that all team members tried most of the R steps on their computers.

Links to labs:

- Week 1 (multi)
- Week 1 (single)
- Week 2 (multi)
- Week 2 (single)

2 Week 1

Survey can be found at:
https://docs.google.com/spreadsheet/viewform?formkey=dEZ0eGNHeDhRd1B0RlltMHJvcXhiZVE6MQ#gid=0

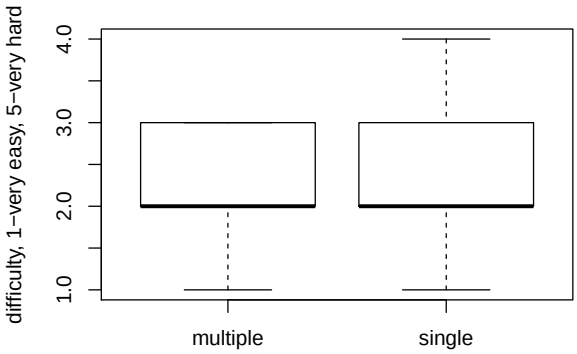
```
d1 = read.csv("Lab survey - Week 1 - Sheet1.csv")
names(d1) = c("timestamp","difficulty","identify_args","no_args",
              "color","identifier")
levels(d1$color) = c("multiple","single")
dim(d1)
```

```
## [1] 85  6
```

2.1 Compare between versions: difficulty

```
boxplot(d1$difficulty ~ d1$color, ylab = "difficulty, 1-very easy, 5-very hard")
t.test(d1$difficulty ~ d1$color)
```

```
##
##  Welch Two Sample t-test
##
## data:  d1$difficulty by d1$color
## t = -1.457, df = 68.69, p-value = 0.1496
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.52407  0.08165
## sample estimates:
## mean in group multiple   mean in group single
##                  2.233                  2.455
##
```

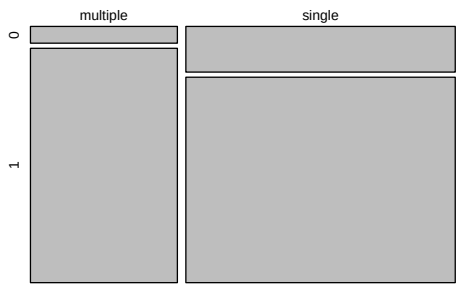


2.2 Compare between versions: % correct on indetifying argument and function

1 means correct, 0 means false.

```
d1$identify_args_score <- NA
d1$identify_args_score[d1$identify_args == "function: names, argument: present"] <- 1
d1$identify_args_score[d1$identify_args == "function: present, argument: names"] <- 0
mosaicplot(table(d1$color, d1$identify_args_score))
chisq.test(table(d1$color, d1$identify_args_score), simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)
##
## data:  table(d1$color, d1$identify_args_score)
## X-squared = 2.123, df = NA, p-value = 0.2089
##
```



2.3 Compare between versions: %correct on number of argument

```
d1$no_args_score <- NA
d1$no_args_score[d1$no_args == 3] <- 1
d1$no_args_score[d1$no_args != 3] <- 0
mosaicplot(table(d1$color, d1$no_args_score))
chisq.test(table(d1$color, d1$no_args_score))
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(d1$color, d1$no_args_score)
## X-squared = 0.0016, df = 1, p-value = 0.9679
##
```

3 Week 2

Survey can be found at:
https://docs.google.com/spreadsheet/viewform?formkey=dHlCMFl2ZmtHT2lDdXR4U0VKWmlicnc6MQ#gid=0

```
d2 = read.csv("Lab survey - Week 2 - Sheet1.csv")
names(d2) = c("timestamp","difficulty","whats_wrong1","whats_wrong2","compare_diff",
              "rewrite", "rewrite_score","typos","color","identifier")
levels(d2$color) = c("multiple","single")
dim(d2)
```

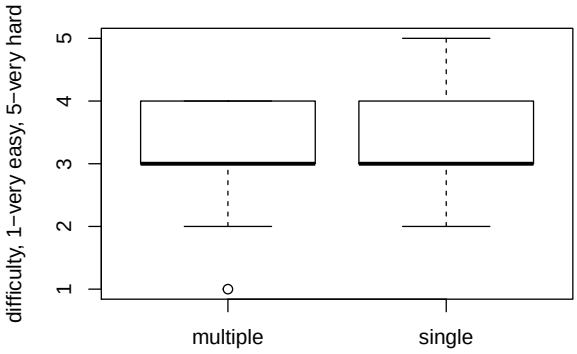
```
## [1] 77 10
```

I assigned scores to the open-ended rewrite the code question manually. There was no partial credit, so those that had only a minor mistake (spell `mosaicplot` as `Mosaicplot`) received the same score as a major mistake or no answer.

3.1 Compare between versions: difficulty

```r
boxplot(d2$difficulty ~ d2$color, ylab = "difficulty, 1-very easy, 5-very hard")
t.test(d2$difficulty ~ d2$color)
```

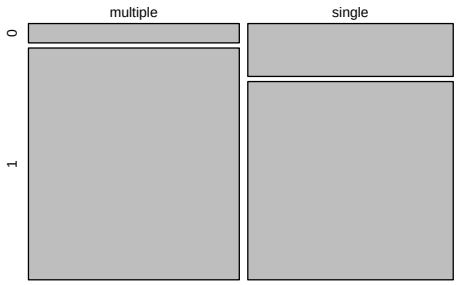
```
##
##  Welch Two Sample t-test
##
## data:  d2$difficulty by d2$color
## t = -1.911, df = 74.27, p-value = 0.05981
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.65323  0.01355
## sample estimates:
## mean in group multiple   mean in group single
##                  3.154                  3.474
##
```



3.2 Compare between versions: what's wrong with the code (1)

```r
d2$whats_wrong1_score <- NA
d2$whats_wrong1_score[d2$whats_wrong1 == "30 should not be in quotation marks."] <- 1
d2$whats_wrong1_score[d2$whats_wrong1 != "30 should not be in quotation marks."] <- 0
mosaicplot(table(d2$color, d2$whats_wrong1_score))
chisq.test(table(d2$color, d2$whats_wrong1_score))
```

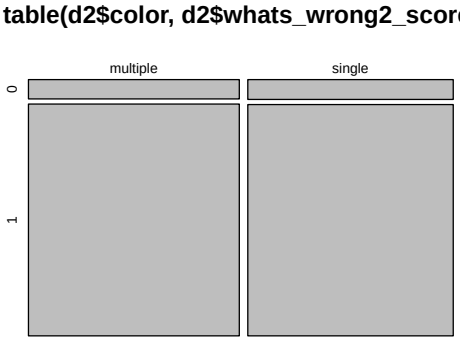
```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(d2$color, d2$whats_wrong1_score)
## X-squared = 1.821, df = 1, p-value = 0.1772
##
```



3.3 Compare between versions: what's wrong with the code (2)

```
d2$whats_wrong2_score <- NA
d2$whats_wrong2_score[d2$whats_wrong2 == " = should instead be ==."] <- 1
d2$whats_wrong2_score[d2$whats_wrong2 != " = should instead be ==."] <- 0
mosaicplot(table(d2$color, d2$whats_wrong2_score))
chisq.test(table(d2$color, d2$whats_wrong2_score), simulate.p.value = TRUE)
```

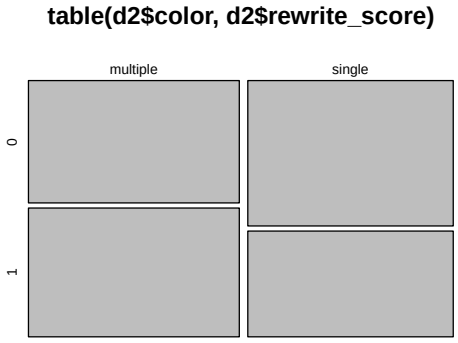
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)
##
## data:  table(d2$color, d2$whats_wrong2_score)
## X-squared = 0.0011, df = NA, p-value = 1
##
```



3.4 Compare between versions: %correct on rewrite

```
mosaicplot(table(d2$color, d2$rewrite_score))
chisq.test(table(d2$color, d2$rewrite_score))
```

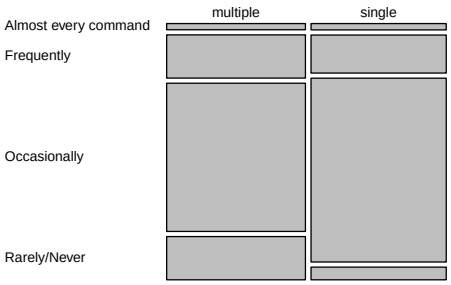
```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(d2$color, d2$rewrite_score)
## X-squared = 0.3346, df = 1, p-value = 0.5629
##
```



3.5 Compare between versions: typos

```
mosaicplot(table(d2$color, d2$typos), las = 1)
chisq.test(table(d2$color, d2$typos), simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)
##
## data:  table(d2$color, d2$typos)
## X-squared = 3.314, df = NA, p-value = 0.3833
##
```



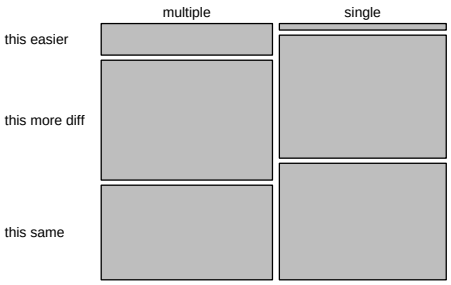
3.6 Compare difficulty between this and previous

```
levels(d2$compare_diff) <- c("this easier", "this more diff", "this same")
mosaicplot(table(d2$color, d2$compare_diff), las = 1)
chisq.test(table(d2$color, d2$compare_diff), simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)
##
## data:  table(d2$color, d2$compare_diff)
## X-squared = 2.927, df = NA, p-value = 0.2894
##
```

```
d2$this_easier <- NA
d2$this_easier[d2$compare_diff == "this easier"] <- 1
d2$this_easier[d2$compare_diff != "this easier"] <- 0
chisq.test(table(d2$color, d2$this_easier), simulate.p.value = TRUE)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)
##
## data:  table(d2$color, d2$this_easier)
## X-squared = 2.781, df = NA, p-value = 0.1944
##
```



4 Compare across weeks

Only 51 students were matched.

```
d <- merge(d1, d2, by = c("identifier", "identifier"))
d <- d[d$identifier != "123", ]  # 4 people chose 123!
dim(d)
```

```
## [1] 51 20
```

4.1 Sanity check

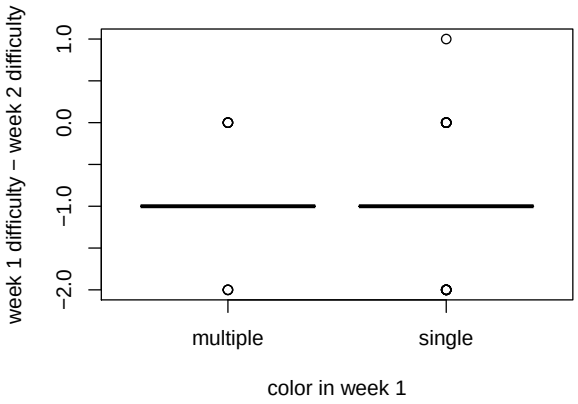
There should have been nobody who had multiple or single colors both weeks, but there are. I'm confident that they got the correct links, so I think they don't quite understand what the question is asking. I don't know what this means about the reliability of the rest of the results.

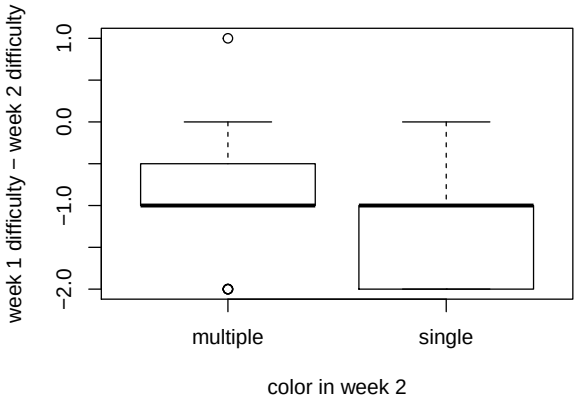
```
table(d$color.x, d$color.y)
```

```
##
##            multiple single
##   multiple       11      7
##   single         17     16
```

4.2 Change in perceived in difficulty

```
d$diff_diff <- d$difficulty.x - d$difficulty.y
boxplot(d$diff_diff ~ d$color.x, xlab = "color in week 1", ylab = "week 1 difficulty -
week 2 difficulty")
boxplot(d$diff_diff ~ d$color.y, xlab = "color in week 2", ylab = "week 1 difficulty -
week 2 difficulty")
```


week 1 difficulty – week 2 difficulty
1.0
0.0
−1.0
−2.0
multiple
single
color in week 1


week 1 difficulty – week 2 difficulty
1.0
0.0
−1.0
−2.0
multiple
single
color in week 2