

1st homework prep

```
library("dplyr")

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library("ggplot2")
library("tidyverse")

## -- Attaching packages -----
## v tibble  2.1.3      v purrr   0.3.3
## v tidyr   1.0.0      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Hypothesis 1: Relation between positive and negative emotions

```
video1 <- read.csv("NeuroDesignVid01.csv")
video2 <- read.csv("NeuroDesignVid02.csv")

# remove all the frames where the face presumably couldn't be captured well enough
video1 %>%
  select(frame, timestamp, confidence, AU01_c, AU01_r, AU04_c, AU04_r, AU05_c, AU05_r, AU06_c, AU06_r, A
  filter(confidence > 0.7) -> video1

video2 %>%
  select(frame, timestamp, confidence, AU01_c, AU01_r, AU04_c, AU04_r, AU05_c, AU05_r, AU06_c, AU06_r, A
  filter(confidence > 0.7) -> video2

# Some data exploration

plot_aus <- function(dataset, title) {
  df1 <- dataset %>%
    select(-timestamp, - confidence, -AU01_c, -AU04_c, -AU05_c, -AU06_c, -AU12_c) %>%
    gather(key = "variable", value = "value", -frame)

  ggplot(df1, aes(x = frame, y = value)) +
    geom_line(aes(color = variable)) +
```

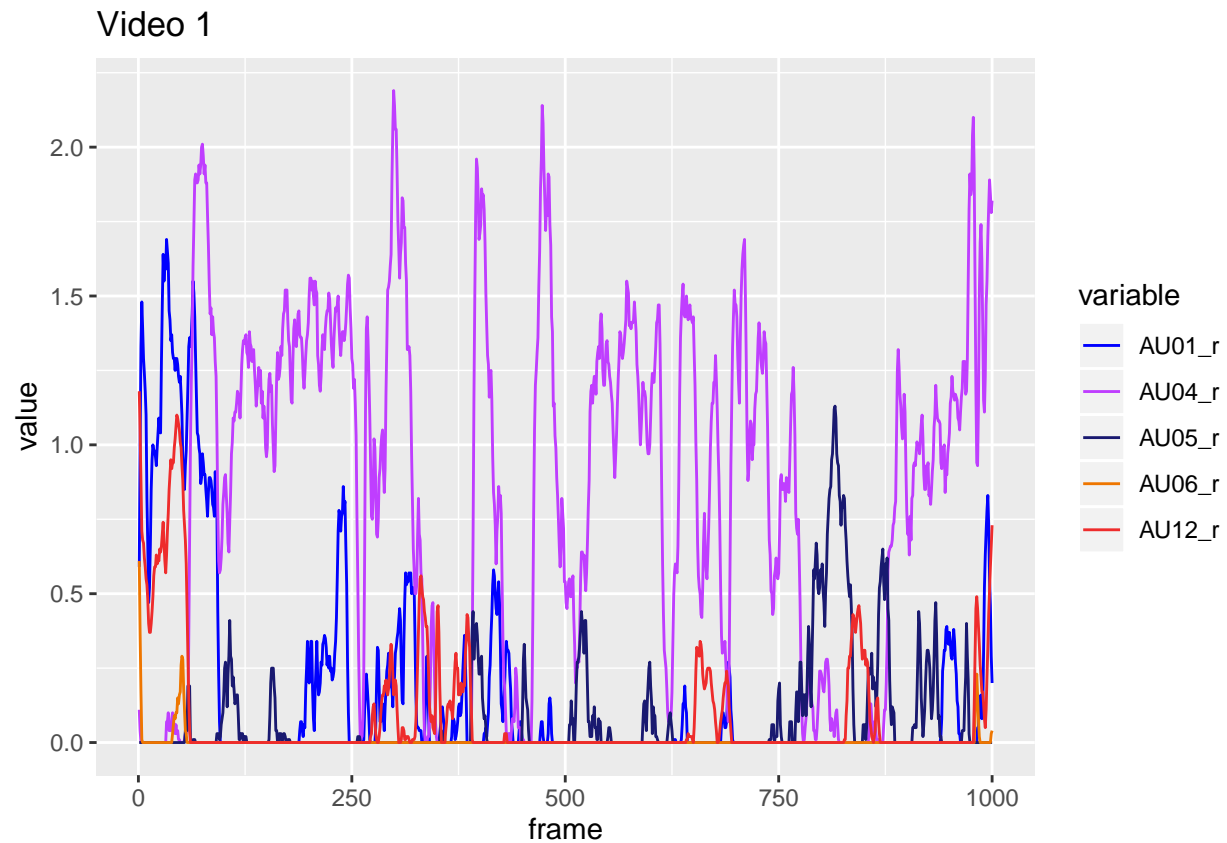
```

scale_color_manual(values = c("blue1", "darkorchid1", "midnightblue", "darkorange2", "firebrick2"))
ggtitle(title)
}

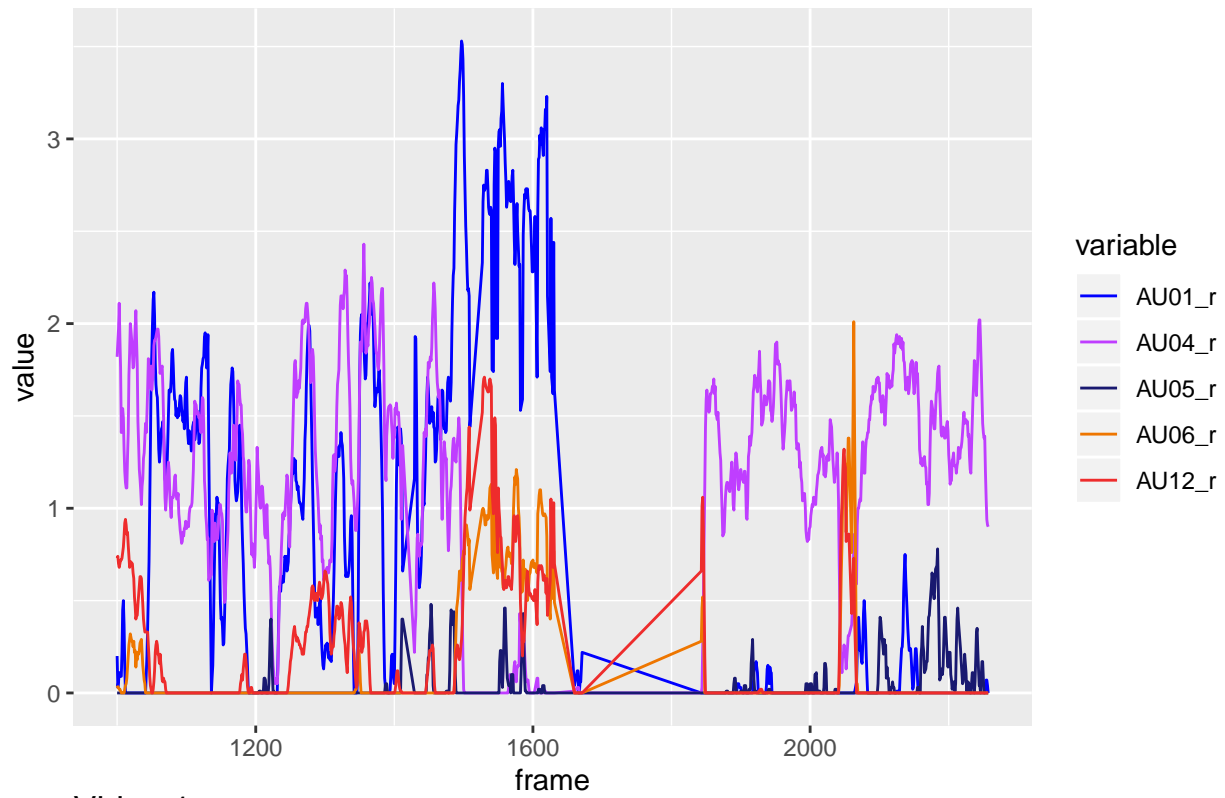
plot_in_steps <- function(dataset, title) {
  for (i in 1:as.integer(nrow(dataset) / 1000)) {
    start = (i - 1) * 1000
    end = i * 1000
    print(plot_au(dataset[start:end,], title))
  }
}

# AU 01, 04, 05 correspond to happiness
# AU 06, 12 correspond to negative emotions
plot_in_steps(video1, "Video 1")

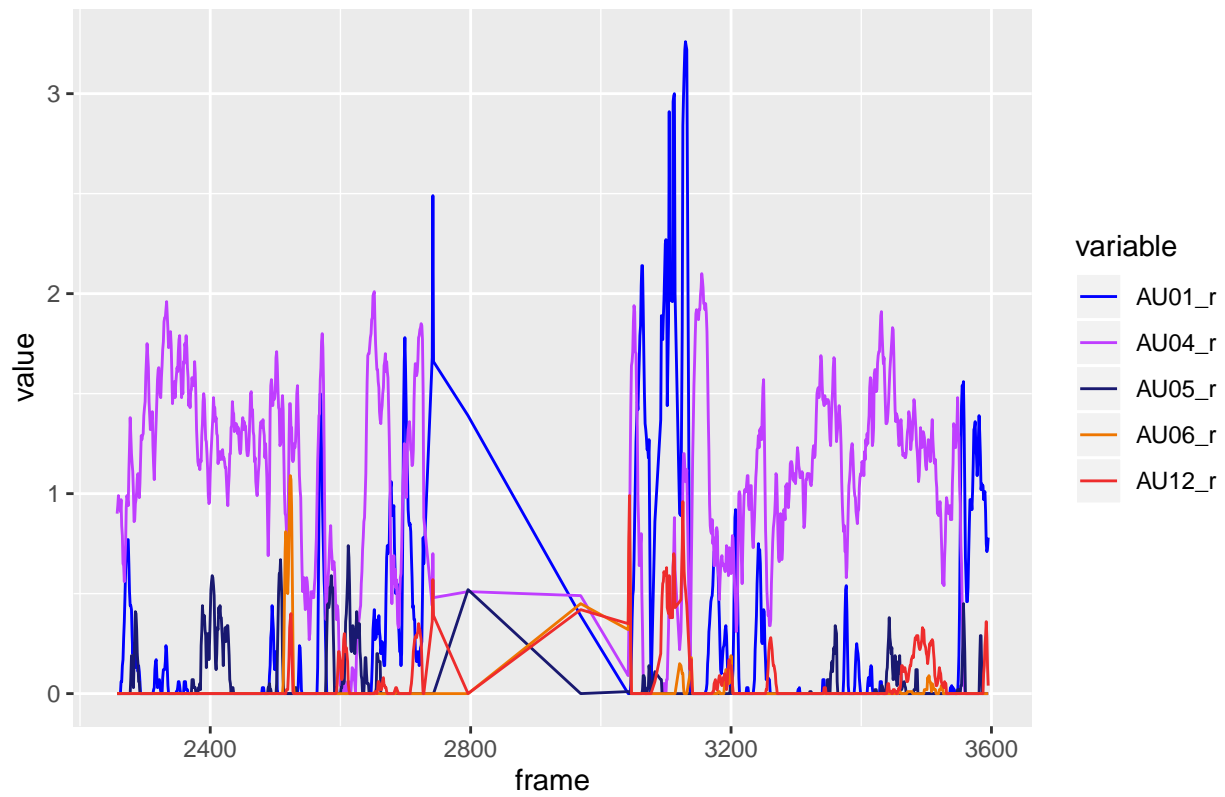
```



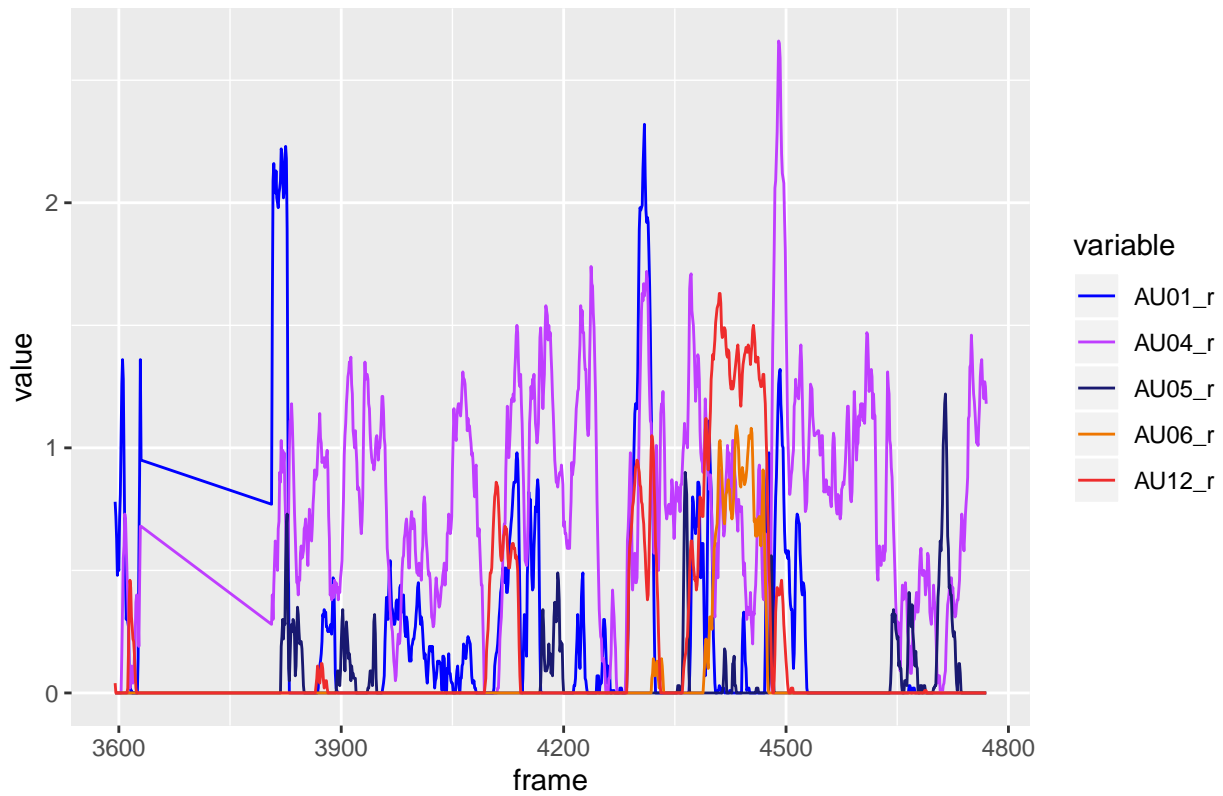
Video 1



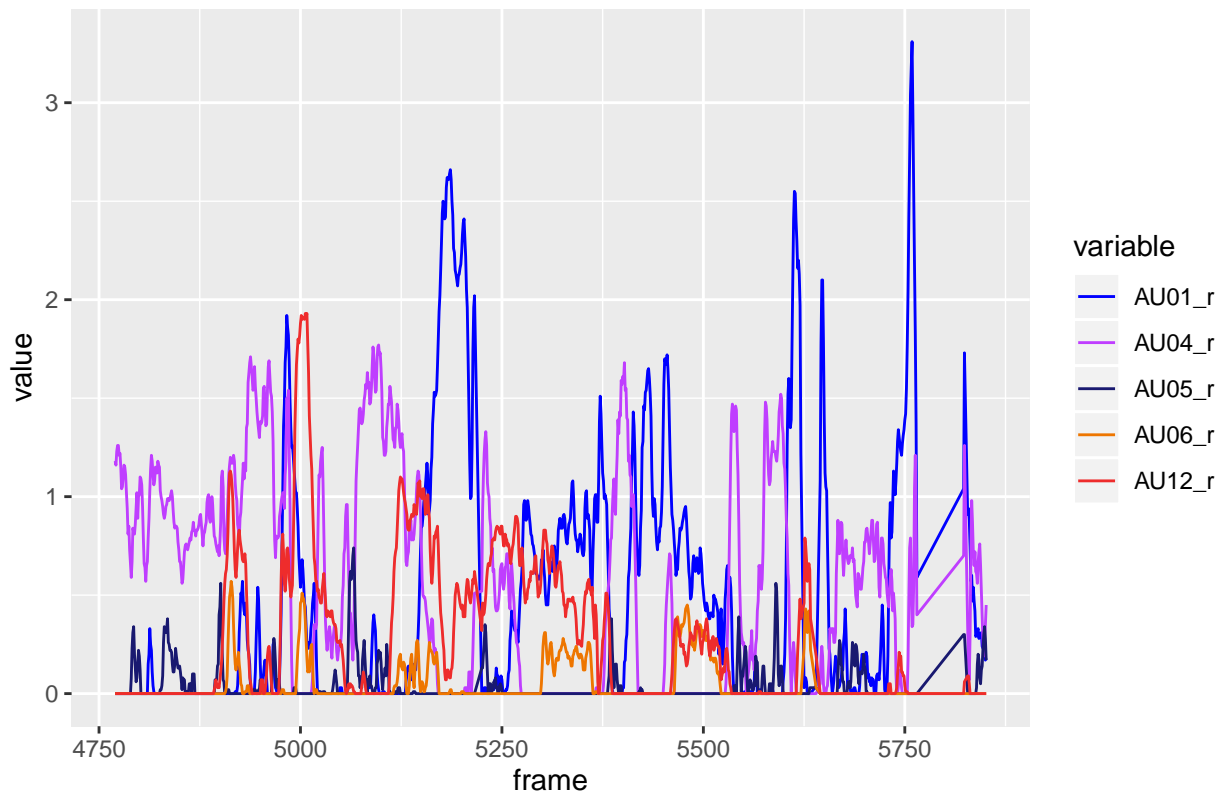
Video 1



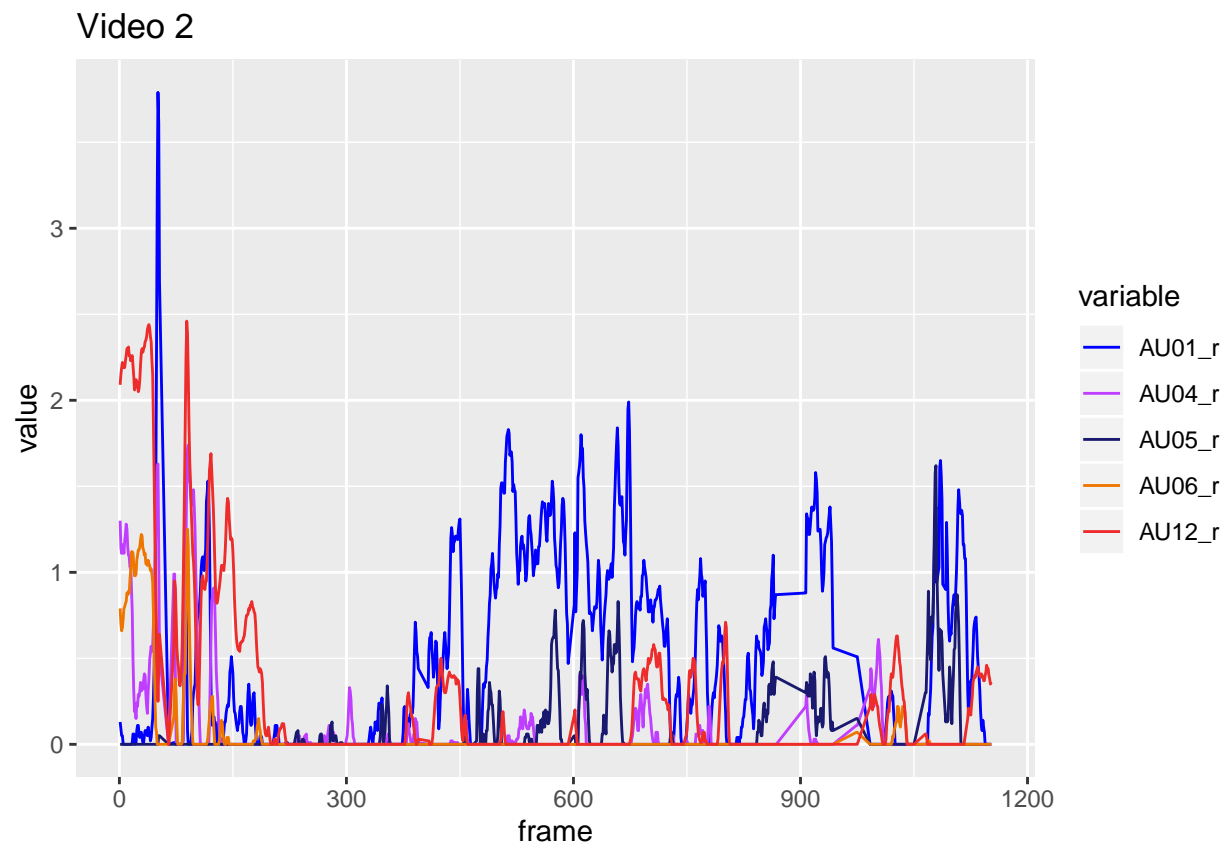
Video 1

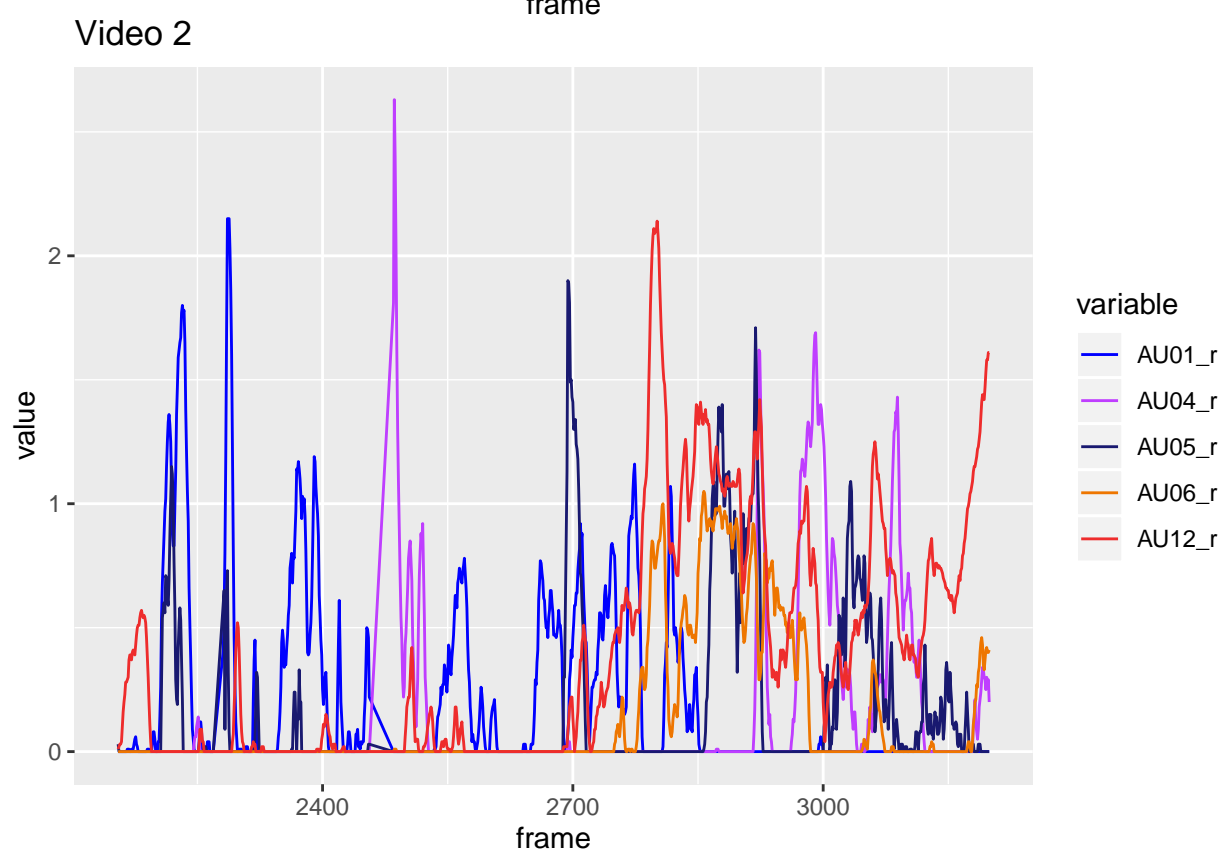
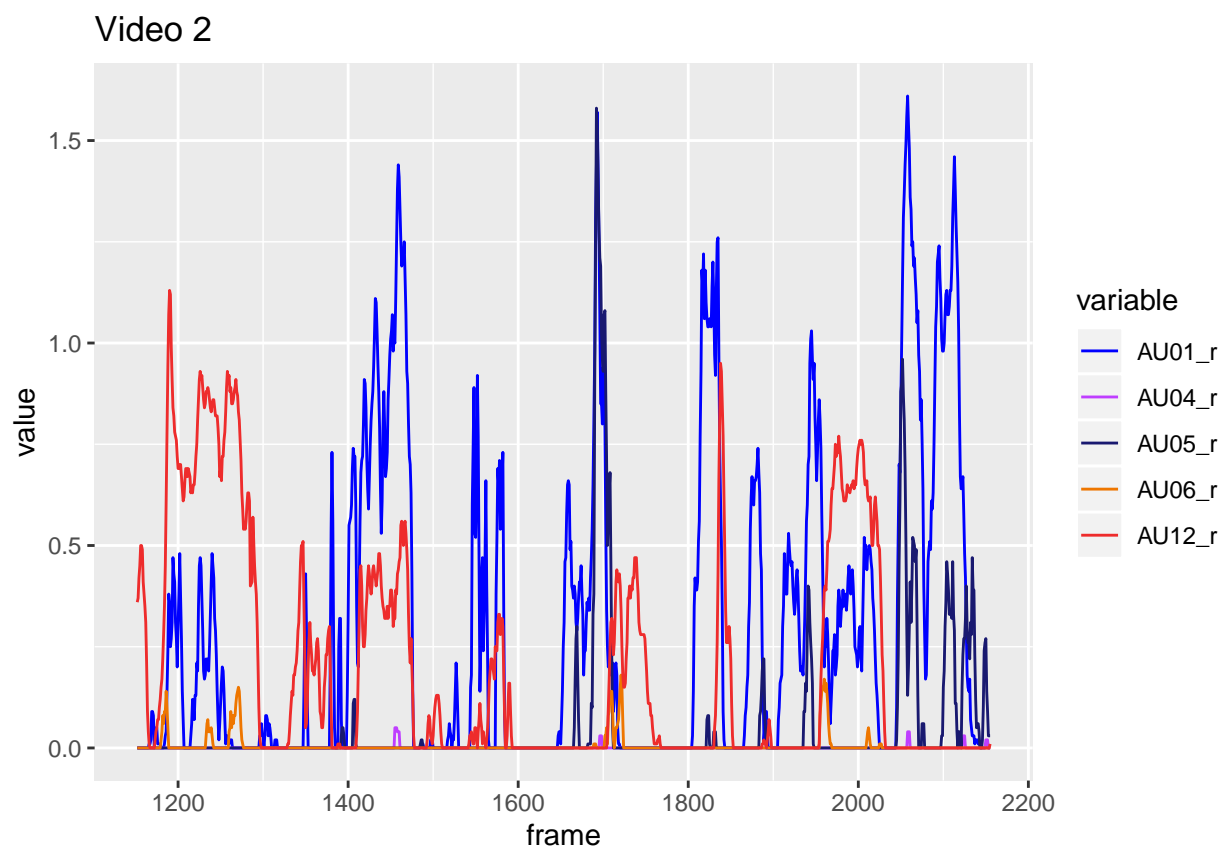


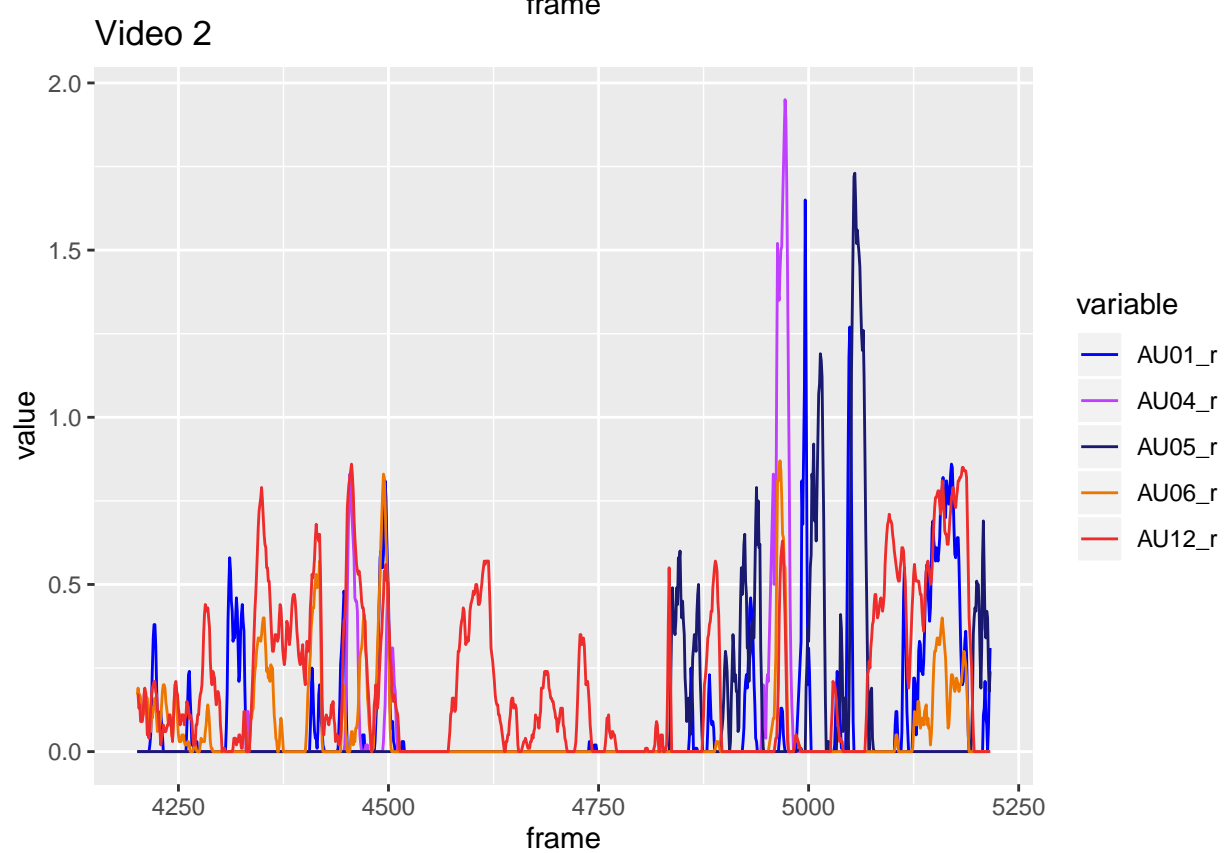
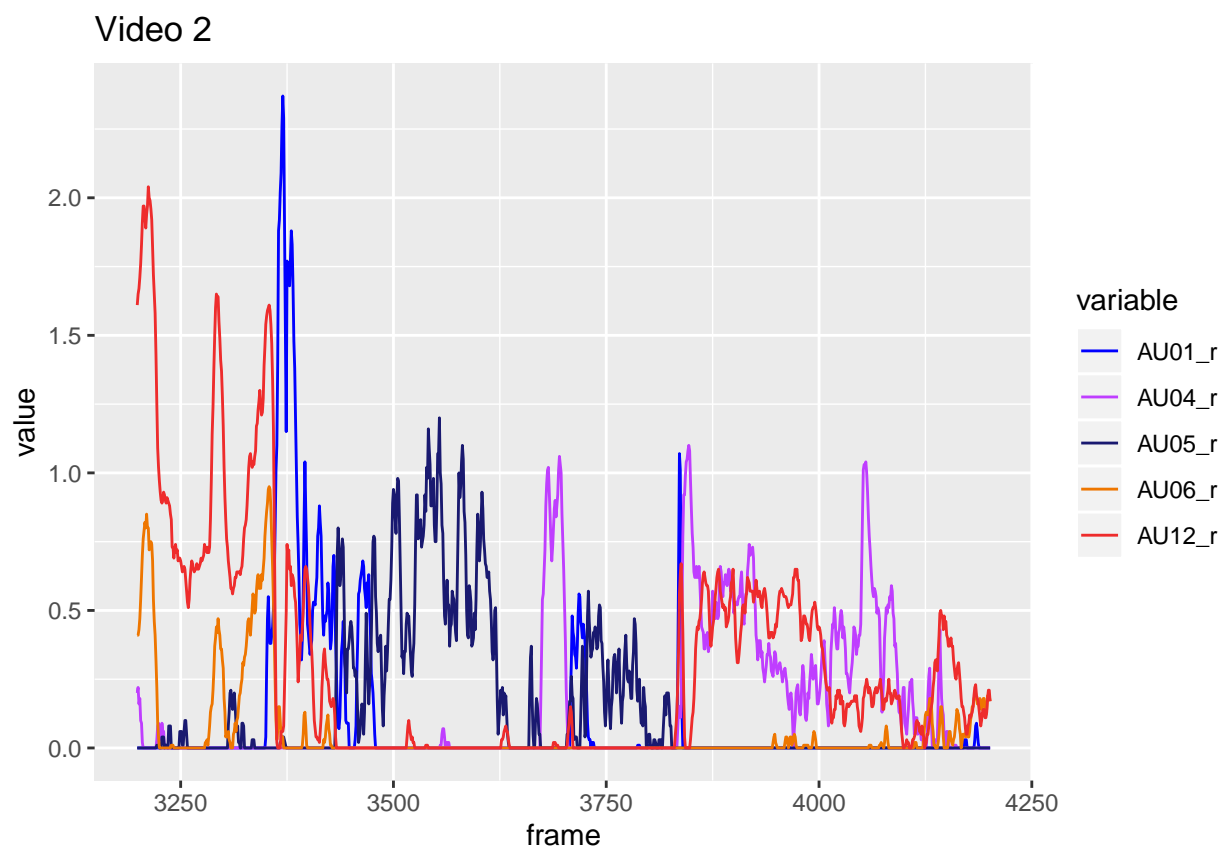
Video 1

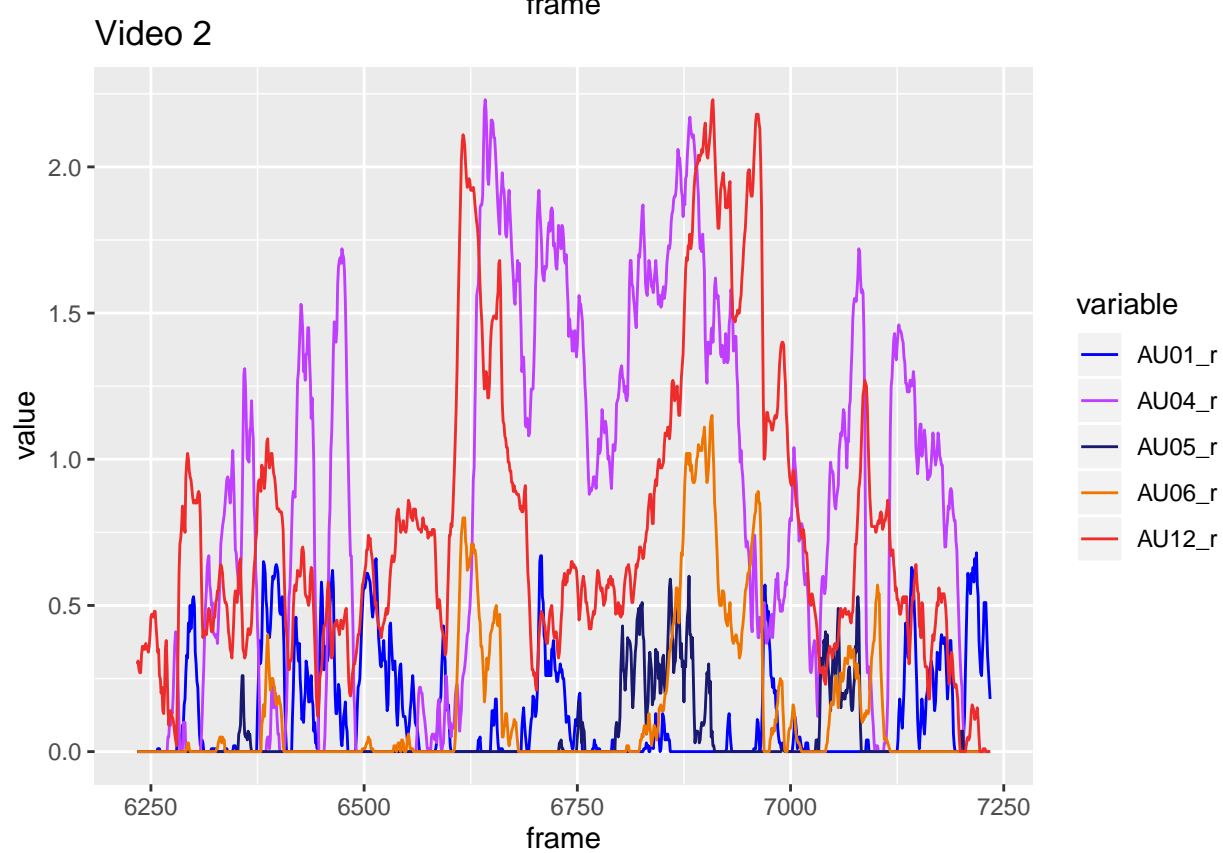
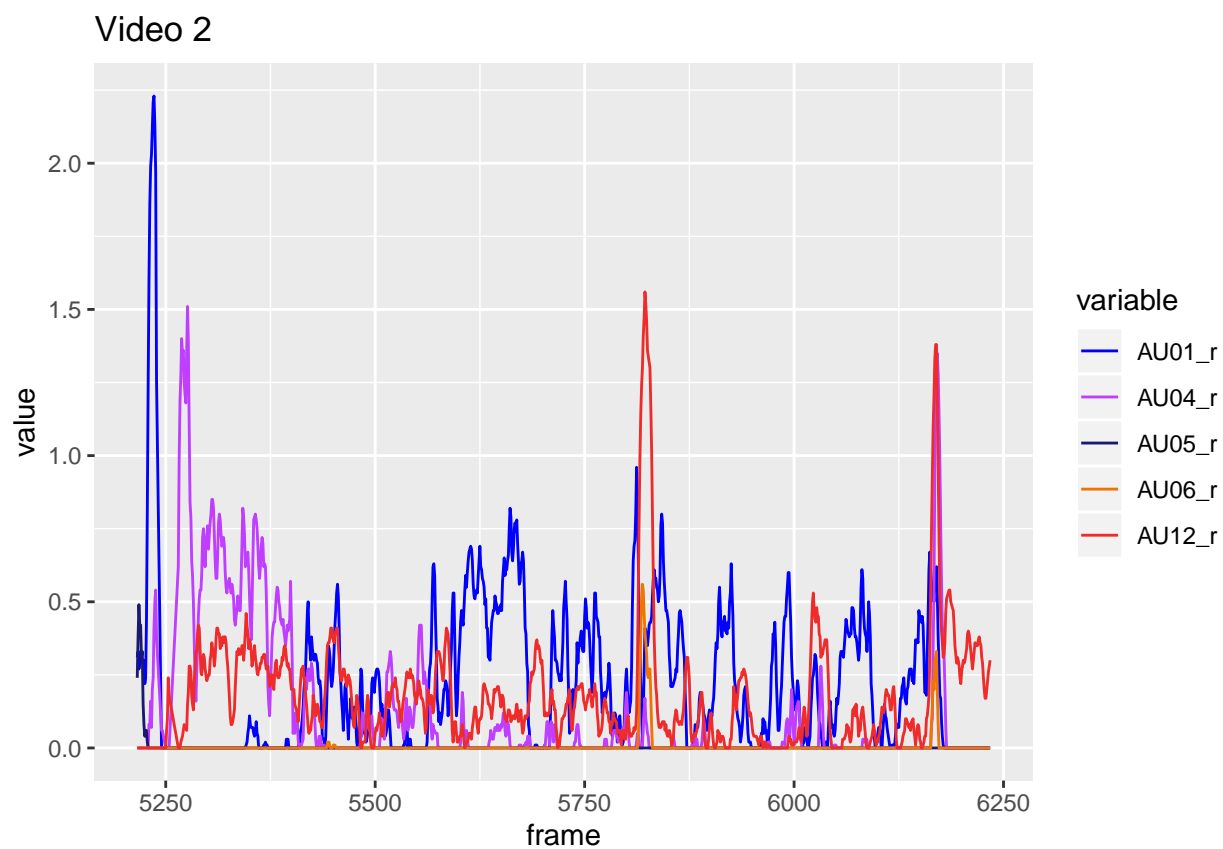


```
plot_in_steps(video2, "Video 2")
```










```

# returns value greater 1 if more positive than negative emotions, a value smaller 1 otherwise
get_relation_emotional_frequencies <- function(dataset) {
  frequencies <- colMeans(select(dataset, AU01_c, AU04_c, AU05_c, AU06_c, AU12_c))

  f <- matrix(frequencies)
  pos_mean <- mean(f[0:3])
  neg_mean <- mean(f[3:5])

  return(pos_mean / neg_mean)
}

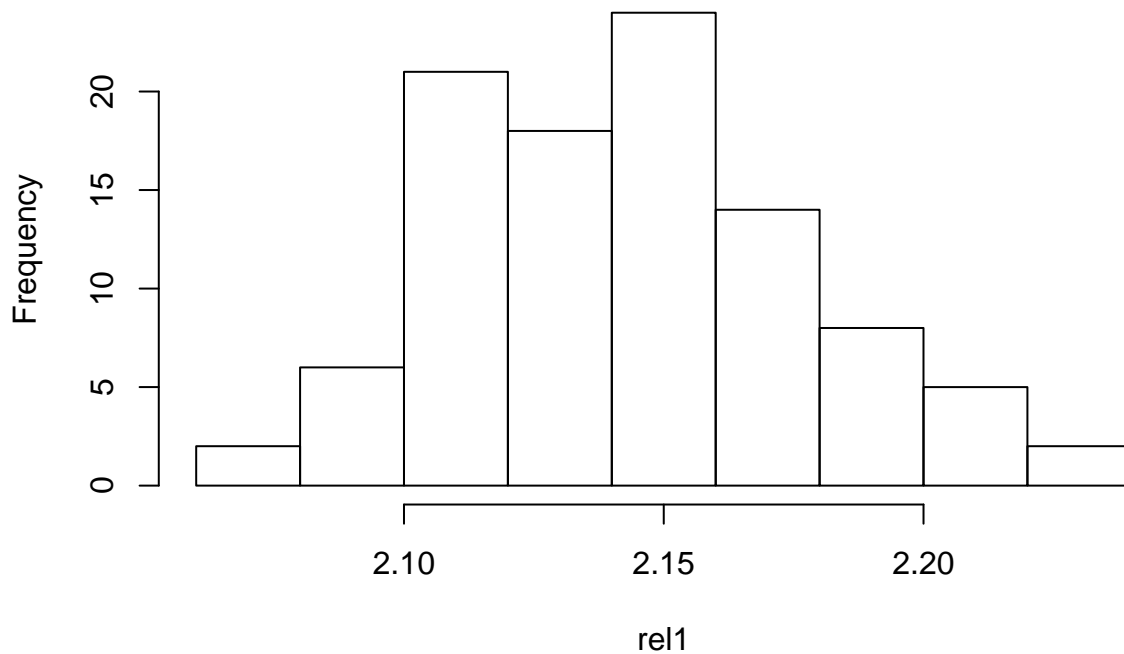
# Creating artificial datasets because we only have two samples
inflate_and_calc_rel_emotions <- function(df) {
  rel = c()
  for (i in 1:100) {
    new_size <- as.integer(nrow(video1) / 5)
    sample = sample_n(video1, new_size)
    rel = c(rel, get_relation_emotional_frequencies(sample))
  }
  return(rel)
}

rel1 <- inflate_and_calc_rel_emotions(video1)
rel2 <- inflate_and_calc_rel_emotions(video2)

hist(rel1)

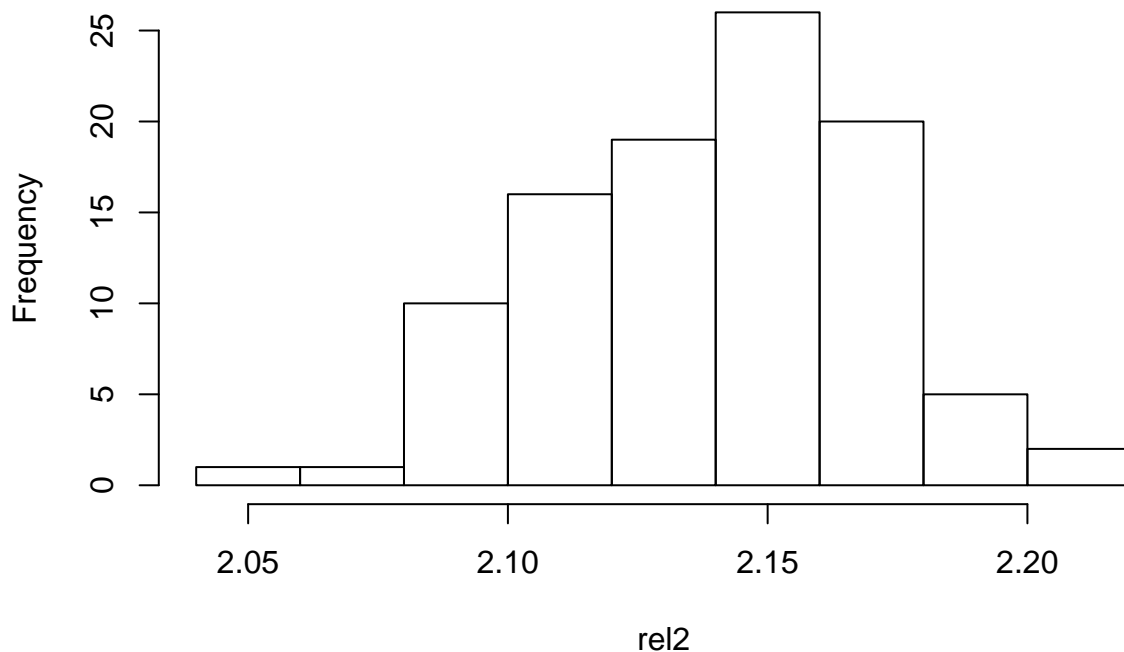
```

Histogram of rel1



```
hist(rel2)
```

Histogram of rel2



```
t.test(rel1, rel2)
```

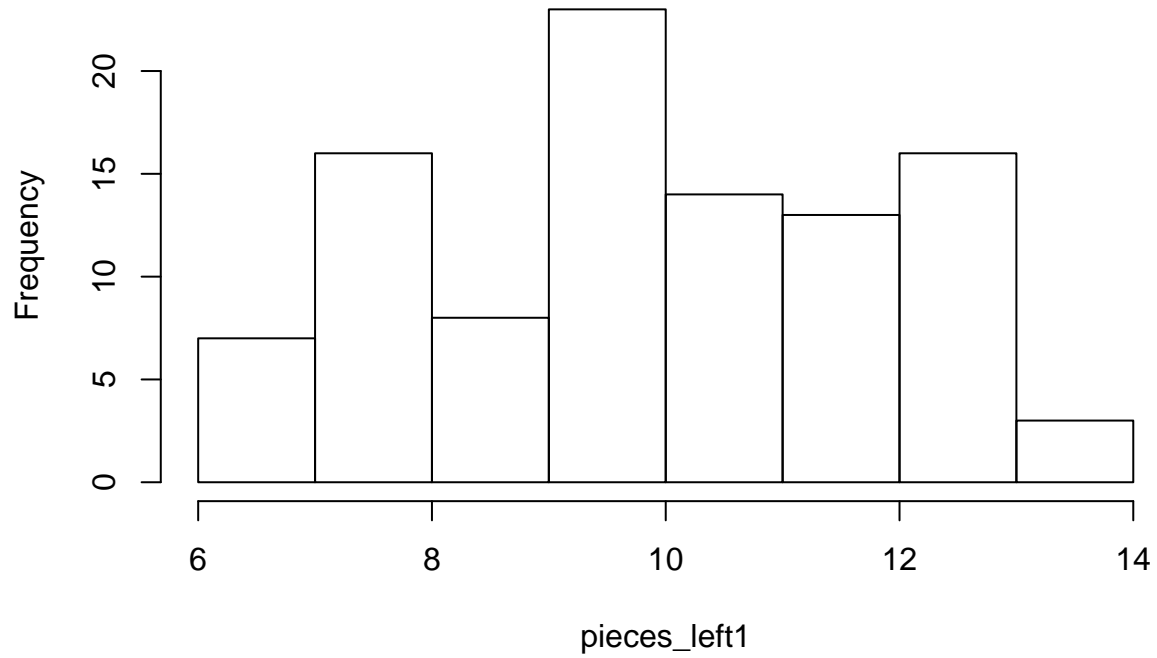
```
##
##  Welch Two Sample t-test
##
## data:  rel1 and rel2
## t = 0.91105, df = 195.49, p-value = 0.3634
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.004942423  0.013429276
## sample estimates:
## mean of x mean of y
##  2.143489  2.139245
```

Hypothesis 2: Number of puzzle pieces left

```
# Creating artificial dataset with number of puzzle pieces left, because the amount of pieces left in t
pieces_left1 <- rnorm(100, 10, 2)
pieces_left2 <- rnorm(100, 8, 2)
```

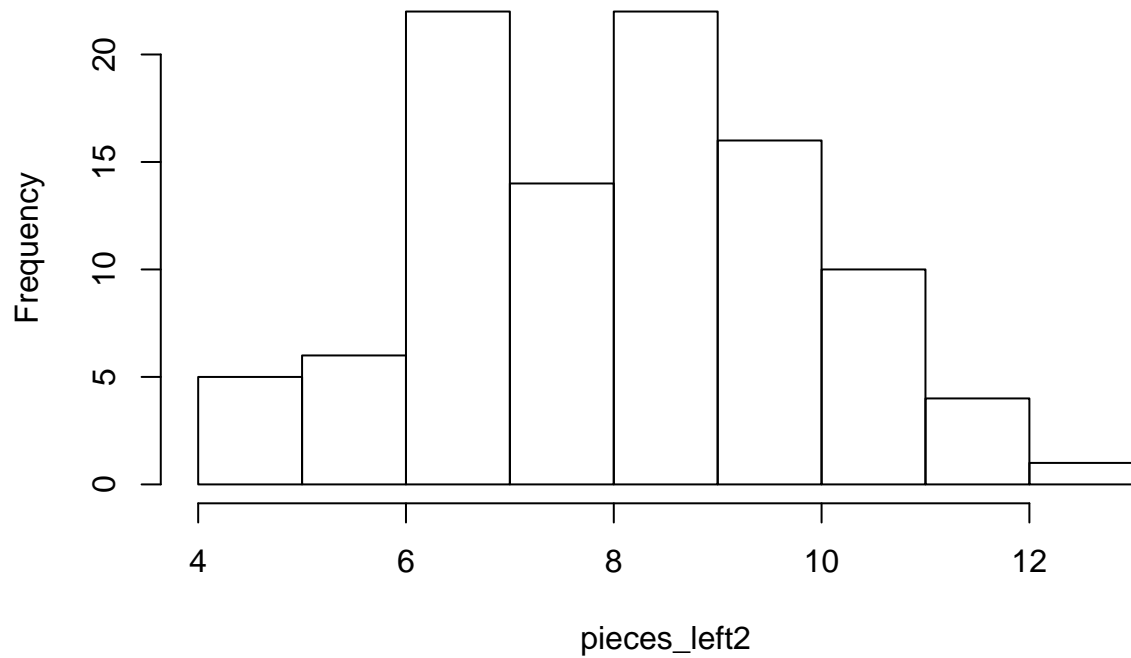
```
hist(pieces_left1)
```

Histogram of pieces_left1



```
hist(pieces_left2)
```

Histogram of pieces_left2



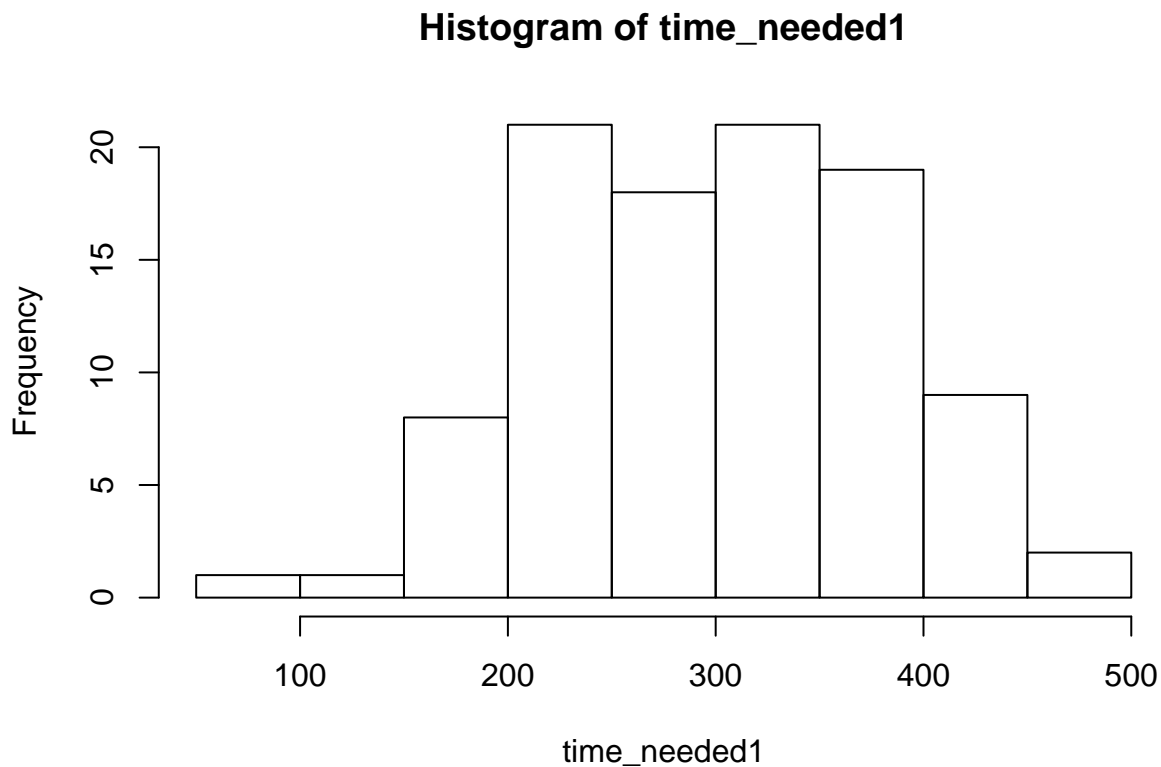
```
t.test(pieces_left1, pieces_left2)
```

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

```
##  
## data: pieces_left1 and pieces_left2  
## t = 6.9794, df = 197.09, p-value = 4.412e-11  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.303803 2.330785  
## sample estimates:  
## mean of x mean of y  
## 9.874802 8.057508
```

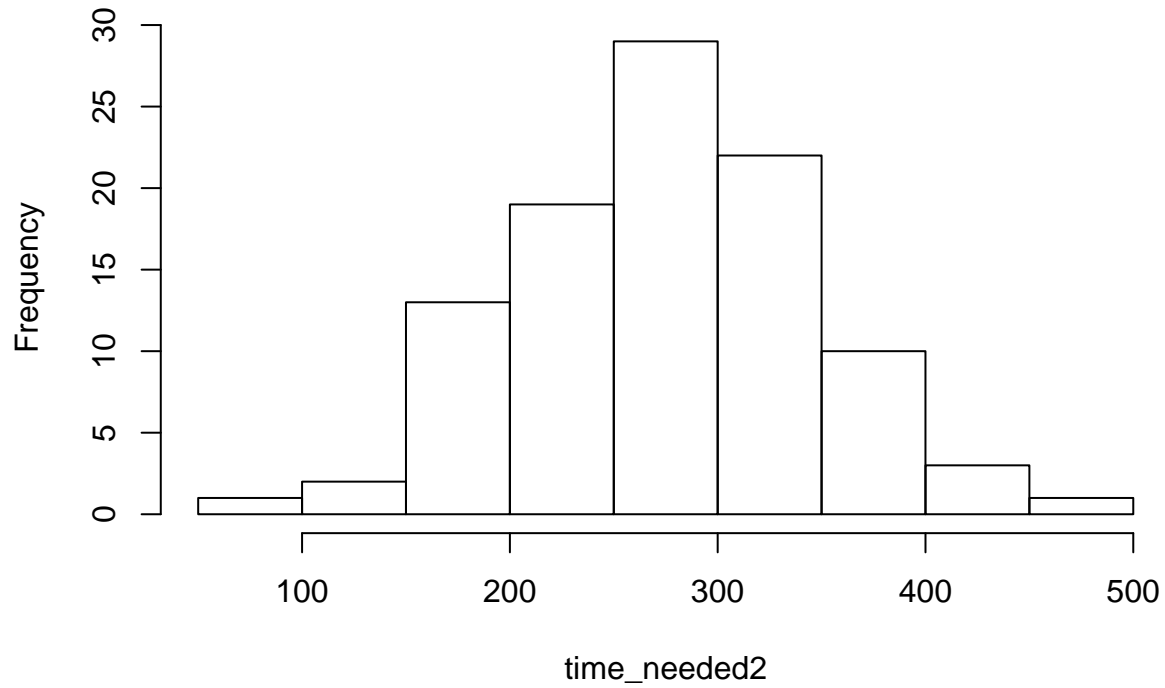
Hypothesis 3: time needed for successful completion

```
# creating artificial dataset with time spent in seconds  
time_needed1 <- rnorm(100, 300, 70)  
time_needed2 <- rnorm(100, 280, 70)  
  
hist(time_needed1)
```



```
hist(time_needed2)
```

Histogram of time_needed2



```
t.test(time_needed1, time_needed2)
```

```
##
##  Welch Two Sample t-test
##
## data:  time_needed1 and time_needed2
## t = 2.4871, df = 196.67, p-value = 0.01371
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.561299 48.156295
## sample estimates:
## mean of x mean of y
##  300.2968  273.4380
```