Midlevel analysis (χ^2)

Marcin

```
data <- read.csv('data_matrix_items.csv')
data$dim <- as.factor(str_replace(data$dim, '_1', ''))
```

Number of stimuli per subject and category:

```
all(xtabs(~ dim + id, data) == 12)
```

```
## [1] TRUE
```

Number of subjects:

```
length(unique(data$id))
```

```
## [1] 99
```

Number of unique stimuli:

```
length(unique(data$item))
```

```
## [1] 47
```

The above should be 48 - is there a duplicate?

```
data %>%
  group_by(id, item) %>%
  summarize(reps_per_subj = n()) %>%
  filter(reps_per_subj > 1) %>%
  group_by(item, reps_per_subj) %>%
  count()
```

```
## `summarise()` has grouped output by 'id'. You can override using the `.groups` argument.
```

```
## # A tibble: 1 x 3
## # Groups:   item, reps_per_subj [1]
##   item                    reps_per_subj     n
##   <chr>                           <int> <int>
## 1 dim10_1_low_utep05_2.wav            2    99
```

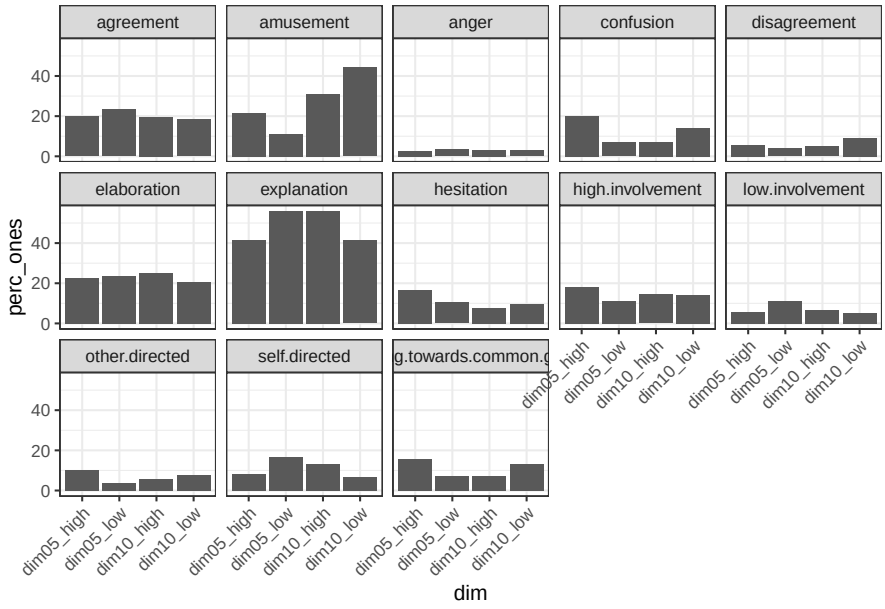
Remove the second occurrence of the duplicate stimulus:

```
data <- data %>%
  group_by(id, item) %>%
  arrange(X) %>%
  slice(1)
```

Overall plot

```
data %>%
  pivot_longer(!c('id', 'dim', 'item', 'X')) %>%
  group_by(dim, name) %>%
  summarize(perc_ones = 100 * sum(value) / n()) %>%
  ggplot(aes(x = dim, y = perc_ones)) +
  geom_col() +
  facet_wrap(~ name, ncol = 5) +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust= 1))
```

```
## `summarise()` has grouped output by 'dim'. You can override using the `.groups` argument.
```



χ^2 tests

Function to run multiple comparisons and correct p-values:

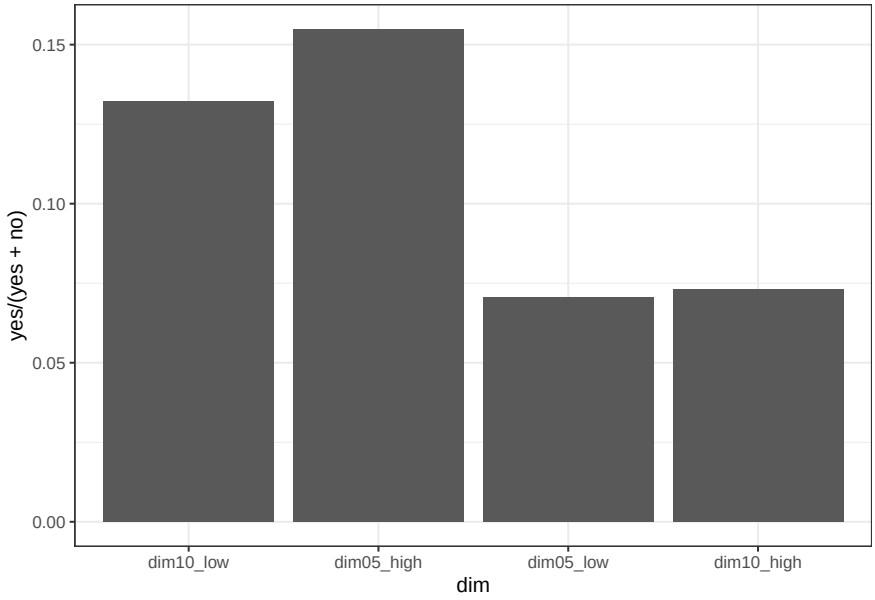
```
chisq_multicomp <- function(df) {
  df <- as.data.frame(df)
  res <- matrix(NA, nrow(df) - 1, 3)
  for (i in 2:nrow(df)) {
    chisq <- chisq.test(df[c(1, i), -1])
    res[i-1, 1] <- chisq$statistic
    res[i-1, 2] <- chisq$p.value
  }
  res[,3] <- res[,2] * (nrow(df) - 1)
  res[res[,3] > 1, 3] <- 1
  rownames(res) <- df[-1, 1]
  colnames(res) <- c('chisq', 'p_value', 'p_value_cor')
  return(res)
}
```

Common ground

```
data$dim <- relevel(data$dim, which(levels(data$dim) == 'dim10_low'))
common_ground <- group_by(data, dim) %>%
  summarize(yes = sum(working.towards.common.ground),
            no = n() - yes)
common_ground
```

```
## # A tibble: 4 x 3
##   dim          yes    no
##   <fct>      <int> <int>
## 1 dim10_low    144   945
## 2 dim05_high   184  1004
## 3 dim05_low     84  1104
## 4 dim10_high    87  1101
```

```
ggplot(common_ground, aes(x = dim, y = yes / (yes + no))) + geom_col()
```




```
chisq_multicomp(common_ground)
```

```
##                chisq      p_value  p_value_cor
## dim05_high  2.184077 1.394446e-01 4.183339e-01
## dim05_low  23.190539 1.467174e-06 4.401522e-06
## dim10_high 21.053633 4.466053e-06 1.339816e-05
```

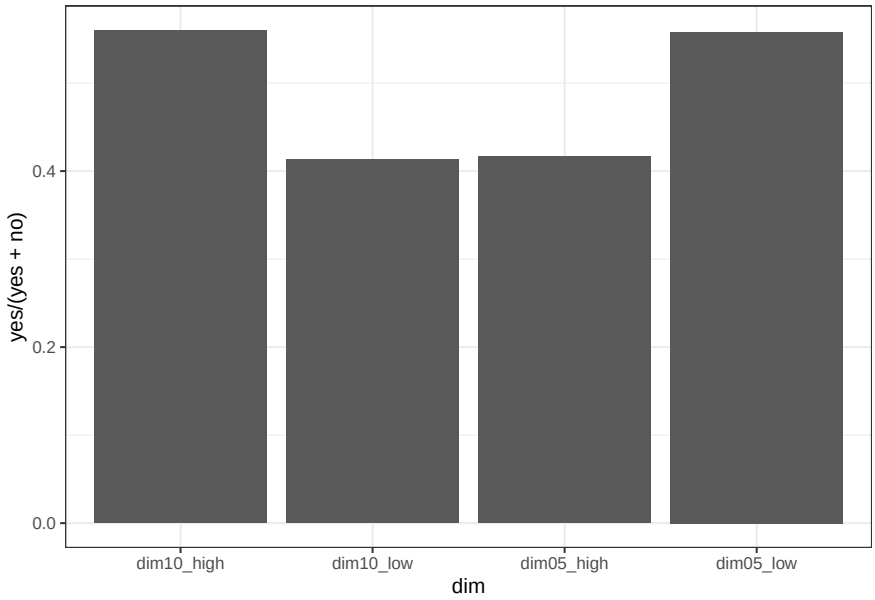
Explanation

```
data$dim <- relevel(data$dim, which(levels(data$dim) == 'dim10_high'))
explanation <- group_by(data, dim) %>%
  summarize(yes = sum(explanation),
            no = n() - yes)
explanation
```

```
## # A tibble: 4 x 3
##   dim          yes    no
```

```
##   <fct>      <int> <int>
## 1 dim10_high   665   523
## 2 dim10_low    450   639
## 3 dim05_high   495   693
## 4 dim05_low    663   525
```

```
ggplot(explanation, aes(x = dim, y = yes / (yes + no))) + geom_col()
```




```
chisq_multicomp(explanation)
```

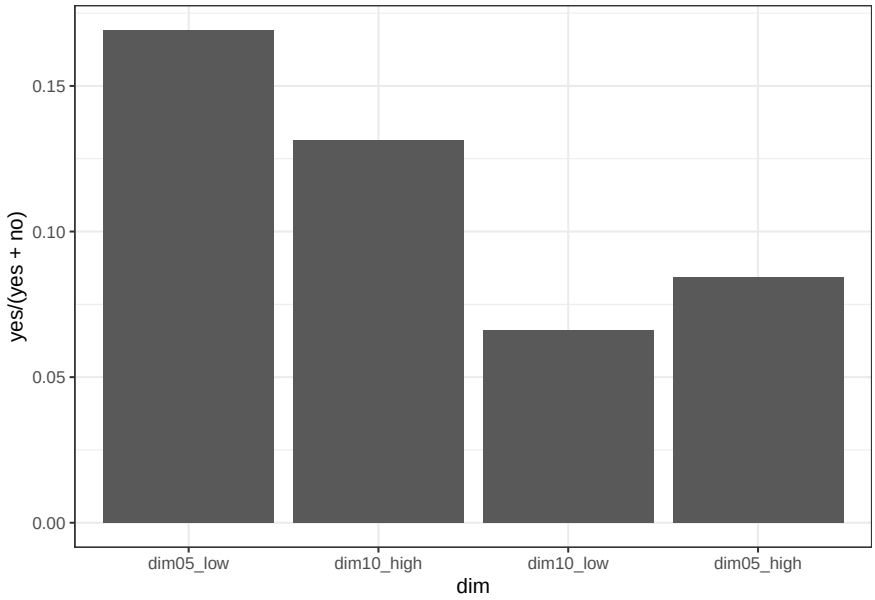
```
##                   chisq      p_value  p_value_cor
## dim10_low  48.240729462 3.769762e-12 1.130929e-11
## dim05_high 48.109216198 4.031272e-12 1.209382e-11
## dim05_low   0.001707211 9.670421e-01 1.000000e+00
```

Self-directed

```
data$dim <- relevel(data$dim, which(levels(data$dim) == 'dim05_low'))
self_dir <- group_by(data, dim) %>%
  summarize(yes = sum(self.directed),
            no = n() - yes)
self_dir
```

```
## # A tibble: 4 x 3
##   dim          yes    no
##   <fct>      <int> <int>
## 1 dim05_low    201   987
## 2 dim10_high   156  1032
## 3 dim10_low     72  1017
## 4 dim05_high   100  1088
```

```
ggplot(self_dir, aes(x = dim, y = yes / (yes + no))) + geom_col()
```




```
chisq_multicomp(self_dir)
```

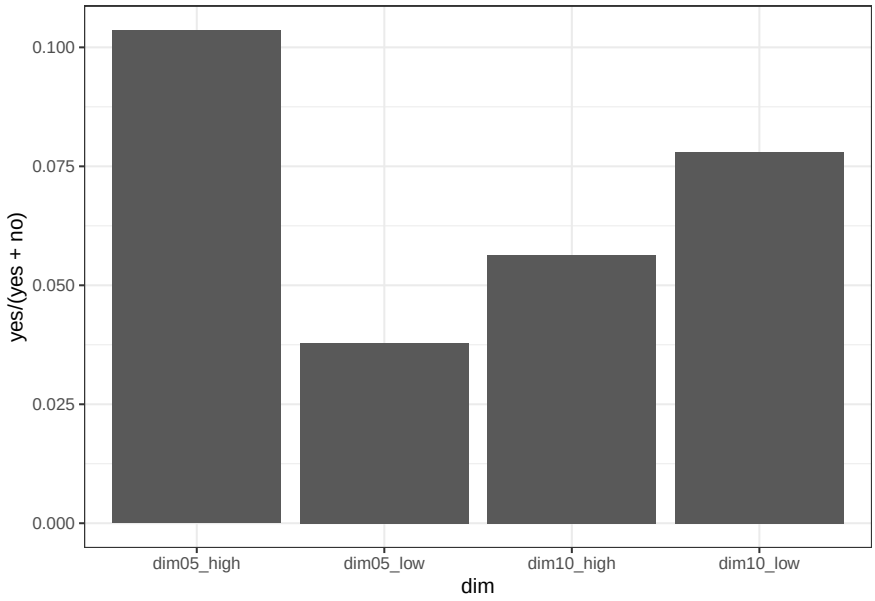
```
##                chisq      p_value  p_value_cor
## dim10_high  6.38186 1.152926e-02 3.458777e-02
## dim10_low  56.23627 6.426498e-14 1.927950e-13
## dim05_high 38.04187 6.924269e-10 2.077281e-09
```

Other-directed

```
data$dim <- relevel(data$dim, which(levels(data$dim) == 'dim05_high'))
other_dir <- group_by(data, dim) %>%
  summarize(yes = sum(other.directed),
            no = n() - yes)
other_dir
```

```
## # A tibble: 4 x 3
##   dim          yes    no
##   <fct>      <int> <int>
## 1 dim05_high   123  1065
## 2 dim05_low     45  1143
## 3 dim10_high    67  1121
## 4 dim10_low     85  1004
```

```
ggplot(other_dir, aes(x = dim, y = yes / (yes + no))) + geom_col()
```




```
chisq_multicomp(other_dir)
```

```
##                chisq      p_value  p_value_cor
## dim05_low  37.976902 7.158713e-10 2.147614e-09
## dim10_high 17.304859 3.183718e-05 9.551153e-05
## dim10_low   4.143117 4.180412e-02 1.254124e-01
```