# ResultsFigures

Alex Kale

2/11/2020

In this document, we create the visualizations that drive our reporting of results, and we save the svg images used to produce the figures in the paper. Although, I do some touching up in Adobe Illustrator, this is basically how the figures in the paper are generated programatically.

# Load in Data

This is the data we used to fit the models.

```
# read in data
model_df <- read_csv("model-data.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   worker_id = col_character(),
##   condition = col_character(),
##   start_means = col_logical(),
##   gender = col_character(),
##   age = col_character(),
##   education = col_character(),
##   chart_use = col_character(),
##   strategy_with_means = col_character(),
##   strategy_without_means = col_character(),
##   outcome = col_logical(),
##   means = col_logical(),
##   exclude = col_logical()
## )
```

```
## See spec(...) for full column specifications.
```

```
# preprocessing
model_df <- model_df %>%
  mutate(
    # factors for modeling
    means = as.factor(means),
    start_means = as.factor(start_means),
    sd_diff = as.factor(sd_diff),
    # evidence scale for decision model
    p_diff = p_award_with - (p_award_without + (1 / award_value)),
    evidence = qlogis(p_award_with) - qlogis(p_award_without + (1 / award_value))
  )
```

# Probability of Superiority

We load in the model of probability of superiority judgments that we arrived at through a process of model expansion described in our preregistration[https://osf.io/9kpmb (https://osf.io/9kpmb)]. This is basically a hierachical linear model of probability of superiority judgments where both judgments and the ground truth have been transformed onto a log odds scale. We call this a linear log odds model. See the paper and **experiment/analysis/PSuperiority.Rmd** in the supplemental materials for details.

```
# hierarchical linear log odds model
m.p_sup <- brm(data = model_df, family = "gaussian",
             formula = bf(lo_p_sup ~  (1 + lo_ground_truth*trial + means*sd_diff|worker_
id) + lo_ground_truth*means*sd_diff*condition*start_means + lo_ground_truth*condition*tr
ial,
                          sigma ~ (1 + lo_ground_truth + trial|worker_id) + lo_ground_tr
uth*condition*trial + means*start_means),
             prior = c(prior(normal(1, 0.5), class = b),
                       prior(normal(1.3, 1), class = Intercept),
                       prior(normal(0, 0.15), class = sd, group = worker_id),
                       prior(normal(0, 0.3), class = b, dpar = sigma),
                       prior(normal(0, 0.15), class = sd, dpar = sigma),
                       prior(lkj(4), class = cor)),
             iter = 12000, warmup = 2000, chains = 2, cores = 2, thin = 2,
             control = list(adapt_delta = 0.99, max_treedepth = 12),
             file = "model-fits/llo_mdl-min-r_means_sd_trial_block_sigma_gt_trial_means_
block")
```

```
summary(m.p_sup)
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = log
## Formula: lo_p_sup ~ (1 + lo_ground_truth * trial + means * sd_diff | worker_id) + lo_
ground_truth * means * sd_diff * condition * start_means + lo_ground_truth * condition *
trial
##          sigma ~ (1 + lo_ground_truth + trial | worker_id) + lo_ground_truth * condit
ion * trial + means * start_means
##    Data: model_df (Number of observations: 19924)
## Samples: 2 chains, each with iter = 12000; warmup = 2000; thin = 2;
##          total post-warmup samples = 10000
##
## Group-Level Effects:
## ~worker_id (Number of levels: 623)
##                                                Estimate Est.Error l-95% CI
## sd(Intercept)                                      0.06      0.01     0.05
## sd(lo_ground_truth)                                0.39      0.01     0.37
## sd(trial)                                          0.03      0.01     0.00
## sd(meansTRUE)                                      0.03      0.01     0.02
## sd(sd_diff15)                                      0.08      0.01     0.07
## sd(lo_ground_truth:trial)                          0.24      0.01     0.21
## sd(meansTRUE:sd_diff15)                            0.06      0.01     0.04
## sd(sigma_Intercept)                                1.18      0.03     1.12
## sd(sigma_lo_ground_truth)                          0.41      0.01     0.38
## sd(sigma_trial)                                    1.19      0.04     1.12
## cor(Intercept,lo_ground_truth)                    -0.47      0.09    -0.64
## cor(Intercept,trial)                               0.20      0.23    -0.28
## cor(lo_ground_truth,trial)                        -0.25      0.23    -0.64
## cor(Intercept,meansTRUE)                           0.04      0.18    -0.29
## cor(lo_ground_truth,meansTRUE)                    -0.60      0.13    -0.81
## cor(trial,meansTRUE)                               0.21      0.25    -0.31
## cor(Intercept,sd_diff15)                          -0.02      0.11    -0.23
## cor(lo_ground_truth,sd_diff15)                     0.03      0.09    -0.14
## cor(trial,sd_diff15)                               0.02      0.21    -0.40
## cor(meansTRUE,sd_diff15)                          -0.00      0.16    -0.34
## cor(Intercept,lo_ground_truth:trial)              -0.27      0.10    -0.45
## cor(lo_ground_truth,lo_ground_truth:trial)         0.40      0.06     0.28
## cor(trial,lo_ground_truth:trial)                  -0.36      0.23    -0.72
## cor(meansTRUE,lo_ground_truth:trial)              -0.13      0.16    -0.43
## cor(sd_diff15,lo_ground_truth:trial)               0.06      0.09    -0.10
## cor(Intercept,meansTRUE:sd_diff15)                -0.33      0.14    -0.58
## cor(lo_ground_truth,meansTRUE:sd_diff15)           0.23      0.13    -0.04
## cor(trial,meansTRUE:sd_diff15)                     0.17      0.22    -0.28
## cor(meansTRUE,meansTRUE:sd_diff15)                 0.03      0.18    -0.33
## cor(sd_diff15,meansTRUE:sd_diff15)                -0.30      0.12    -0.51
## cor(lo_ground_truth:trial,meansTRUE:sd_diff15)    -0.12      0.12    -0.36
## cor(sigma_Intercept,sigma_lo_ground_truth)        -0.71      0.02    -0.75
## cor(sigma_Intercept,sigma_trial)                   0.10      0.04     0.02
## cor(sigma_lo_ground_truth,sigma_trial)            -0.05      0.04    -0.14
##                                                u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)                                      0.07 1.00     3205     6294
## sd(lo_ground_truth)                                0.42 1.00     3332     6907
## sd(trial)                                          0.06 1.00     1235     2547
## sd(meansTRUE)                                      0.05 1.00     1379     2192
```

```
## sd(sd_diff15)                                      0.09 1.00     4032     7833
## sd(lo_ground_truth:trial)                          0.27 1.00     1677     5455
## sd(meansTRUE:sd_diff15)                            0.07 1.00     3520     7180
## sd(sigma_Intercept)                                1.24 1.00     2740     4467
## sd(sigma_lo_ground_truth)                          0.43 1.00     3982     6582
## sd(sigma_trial)                                    1.27 1.00     5467     7404
## cor(Intercept,lo_ground_truth)                    -0.29 1.00      587     1241
## cor(Intercept,trial)                               0.61 1.00     6065     7549
## cor(lo_ground_truth,trial)                         0.26 1.00     4894     5531
## cor(Intercept,meansTRUE)                           0.41 1.00     2402     5259
## cor(lo_ground_truth,meansTRUE)                    -0.30 1.00     2874     5664
## cor(trial,meansTRUE)                               0.63 1.00     1796     3322
## cor(Intercept,sd_diff15)                           0.19 1.00     2076     4506
## cor(lo_ground_truth,sd_diff15)                     0.20 1.00     3829     7001
## cor(trial,sd_diff15)                               0.44 1.00      345      816
## cor(meansTRUE,sd_diff15)                           0.31 1.00      627     1393
## cor(Intercept,lo_ground_truth:trial)              -0.07 1.00     1131     2593
## cor(lo_ground_truth,lo_ground_truth:trial)         0.52 1.00     6278     8237
## cor(trial,lo_ground_truth:trial)                   0.18 1.01      286      427
## cor(meansTRUE,lo_ground_truth:trial)               0.18 1.00      544     1427
## cor(sd_diff15,lo_ground_truth:trial)               0.23 1.00     2636     5672
## cor(Intercept,meansTRUE:sd_diff15)                -0.05 1.00     3043     6243
## cor(lo_ground_truth,meansTRUE:sd_diff15)           0.47 1.00     4341     8138
## cor(trial,meansTRUE:sd_diff15)                     0.57 1.00     1029     2030
## cor(meansTRUE,meansTRUE:sd_diff15)                 0.39 1.00     2070     4493
## cor(sd_diff15,meansTRUE:sd_diff15)                -0.04 1.00     3507     7193
## cor(lo_ground_truth:trial,meansTRUE:sd_diff15)     0.12 1.00     3241     6173
## cor(sigma_Intercept,sigma_lo_ground_truth)        -0.67 1.00     3969     6400
## cor(sigma_Intercept,sigma_trial)                   0.18 1.00     4842     6926
## cor(sigma_lo_ground_truth,sigma_trial)             0.03 1.00     4098     6587
##
## Population-Level Effects:
##                                                                    Estimate
## Intercept                                                             -0.02
## sigma_Intercept                                                       -1.71
## lo_ground_truth                                                        0.45
## meansTRUE                                                             -0.00
## sd_diff15                                                              0.04
## conditionHOPs                                                         -0.09
## conditionintervals                                                    -0.01
## conditionQDPs                                                          0.02
## start_meansTRUE                                                        0.01
## trial                                                                 -0.06
## lo_ground_truth:meansTRUE                                             -0.05
## lo_ground_truth:sd_diff15                                              0.08
## meansTRUE:sd_diff15                                                    0.02
## lo_ground_truth:conditionHOPs                                         -0.01
## lo_ground_truth:conditionintervals                                    -0.10
## lo_ground_truth:conditionQDPs                                          0.07
## meansTRUE:conditionHOPs                                                0.08
## meansTRUE:conditionintervals                                           0.01
## meansTRUE:conditionQDPs                                               -0.02
## sd_diff15:conditionHOPs                                                0.03
## sd_diff15:conditionintervals                                           0.02
```

```
## sd_diff15:conditionQDPs                                                   -0.01
## lo_ground_truth:start_meansTRUE                                           -0.14
## meansTRUE:start_meansTRUE                                                 -0.02
## sd_diff15:start_meansTRUE                                                  0.00
## conditionHOPs:start_meansTRUE                                              0.08
## conditionintervals:start_meansTRUE                                         0.00
## conditionQDPs:start_meansTRUE                                             -0.01
## lo_ground_truth:trial                                                      0.13
## conditionHOPs:trial                                                        0.01
## conditionintervals:trial                                                   0.04
## conditionQDPs:trial                                                        0.05
## lo_ground_truth:meansTRUE:sd_diff15                                        0.05
## lo_ground_truth:meansTRUE:conditionHOPs                                   -0.08
## lo_ground_truth:meansTRUE:conditionintervals                              -0.01
## lo_ground_truth:meansTRUE:conditionQDPs                                   -0.00
## lo_ground_truth:sd_diff15:conditionHOPs                                    0.06
## lo_ground_truth:sd_diff15:conditionintervals                              -0.01
## lo_ground_truth:sd_diff15:conditionQDPs                                    0.03
## meansTRUE:sd_diff15:conditionHOPs                                         -0.00
## meansTRUE:sd_diff15:conditionintervals                                    -0.02
## meansTRUE:sd_diff15:conditionQDPs                                          0.00
## lo_ground_truth:meansTRUE:start_meansTRUE                                  0.04
## lo_ground_truth:sd_diff15:start_meansTRUE                                  0.03
## meansTRUE:sd_diff15:start_meansTRUE                                       -0.01
## lo_ground_truth:conditionHOPs:start_meansTRUE                             -0.07
## lo_ground_truth:conditionintervals:start_meansTRUE                         0.03
## lo_ground_truth:conditionQDPs:start_meansTRUE                              0.14
## meansTRUE:conditionHOPs:start_meansTRUE                                   -0.09
## meansTRUE:conditionintervals:start_meansTRUE                               0.01
## meansTRUE:conditionQDPs:start_meansTRUE                                    0.02
## sd_diff15:conditionHOPs:start_meansTRUE                                   -0.02
## sd_diff15:conditionintervals:start_meansTRUE                              -0.01
## sd_diff15:conditionQDPs:start_meansTRUE                                   -0.02
## lo_ground_truth:conditionHOPs:trial                                       -0.03
## lo_ground_truth:conditionintervals:trial                                   0.00
## lo_ground_truth:conditionQDPs:trial                                       -0.00
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs                         -0.02
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals                     0.02
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs                         -0.04
## lo_ground_truth:meansTRUE:sd_diff15:start_meansTRUE                        0.04
## lo_ground_truth:meansTRUE:conditionHOPs:start_meansTRUE                    0.12
## lo_ground_truth:meansTRUE:conditionintervals:start_meansTRUE               0.03
## lo_ground_truth:meansTRUE:conditionQDPs:start_meansTRUE                   -0.01
## lo_ground_truth:sd_diff15:conditionHOPs:start_meansTRUE                    0.01
## lo_ground_truth:sd_diff15:conditionintervals:start_meansTRUE               0.00
## lo_ground_truth:sd_diff15:conditionQDPs:start_meansTRUE                   -0.02
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                          0.04
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE                     0.02
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                          0.01
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE         -0.07
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE    -0.04
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE         -0.00
## sigma_lo_ground_truth                                                      0.45
## sigma_conditionHOPs                                                        0.58
```

```
## sigma_conditionintervals                                         0.16
## sigma_conditionQDPs                                             -0.05
## sigma_trial                                                     -0.45
## sigma_meansTRUE                                                  0.00
## sigma_start_meansTRUE                                           -0.04
## sigma_lo_ground_truth:conditionHOPs                             -0.17
## sigma_lo_ground_truth:conditionintervals                        -0.10
## sigma_lo_ground_truth:conditionQDPs                             -0.03
## sigma_lo_ground_truth:trial                                      0.02
## sigma_conditionHOPs:trial                                        0.06
## sigma_conditionintervals:trial                                   0.12
## sigma_conditionQDPs:trial                                       -0.06
## sigma_meansTRUE:start_meansTRUE                                 -0.23
## sigma_lo_ground_truth:conditionHOPs:trial                        0.05
## sigma_lo_ground_truth:conditionintervals:trial                   0.06
## sigma_lo_ground_truth:conditionQDPs:trial                       -0.02
##                                                            Est.Error
## Intercept                                                       0.02
## sigma_Intercept                                                 0.09
## lo_ground_truth                                                 0.04
## meansTRUE                                                       0.02
## sd_diff15                                                       0.02
## conditionHOPs                                                   0.03
## conditionintervals                                              0.02
## conditionQDPs                                                   0.02
## start_meansTRUE                                                 0.02
## trial                                                           0.02
## lo_ground_truth:meansTRUE                                       0.02
## lo_ground_truth:sd_diff15                                       0.02
## meansTRUE:sd_diff15                                             0.02
## lo_ground_truth:conditionHOPs                                   0.07
## lo_ground_truth:conditionintervals                              0.06
## lo_ground_truth:conditionQDPs                                   0.06
## meansTRUE:conditionHOPs                                         0.03
## meansTRUE:conditionintervals                                    0.02
## meansTRUE:conditionQDPs                                         0.03
## sd_diff15:conditionHOPs                                         0.04
## sd_diff15:conditionintervals                                    0.03
## sd_diff15:conditionQDPs                                         0.03
## lo_ground_truth:start_meansTRUE                                 0.06
## meansTRUE:start_meansTRUE                                       0.03
## sd_diff15:start_meansTRUE                                       0.03
## conditionHOPs:start_meansTRUE                                   0.04
## conditionintervals:start_meansTRUE                              0.03
## conditionQDPs:start_meansTRUE                                   0.03
## lo_ground_truth:trial                                           0.03
## conditionHOPs:trial                                             0.04
## conditionintervals:trial                                        0.03
## conditionQDPs:trial                                             0.03
## lo_ground_truth:meansTRUE:sd_diff15                             0.02
## lo_ground_truth:meansTRUE:conditionHOPs                         0.04
## lo_ground_truth:meansTRUE:conditionintervals                    0.03
## lo_ground_truth:meansTRUE:conditionQDPs                         0.03
## lo_ground_truth:sd_diff15:conditionHOPs                         0.03
```

```
## lo_ground_truth:sd_diff15:conditionintervals                                0.02
## lo_ground_truth:sd_diff15:conditionQDPs                                     0.03
## meansTRUE:sd_diff15:conditionHOPs                                           0.04
## meansTRUE:sd_diff15:conditionintervals                                      0.03
## meansTRUE:sd_diff15:conditionQDPs                                           0.03
## lo_ground_truth:meansTRUE:start_meansTRUE                                   0.03
## lo_ground_truth:sd_diff15:start_meansTRUE                                   0.02
## meansTRUE:sd_diff15:start_meansTRUE                                         0.03
## lo_ground_truth:conditionHOPs:start_meansTRUE                               0.09
## lo_ground_truth:conditionintervals:start_meansTRUE                          0.09
## lo_ground_truth:conditionQDPs:start_meansTRUE                               0.09
## meansTRUE:conditionHOPs:start_meansTRUE                                     0.05
## meansTRUE:conditionintervals:start_meansTRUE                                0.04
## meansTRUE:conditionQDPs:start_meansTRUE                                     0.04
## sd_diff15:conditionHOPs:start_meansTRUE                                     0.05
## sd_diff15:conditionintervals:start_meansTRUE                                0.04
## sd_diff15:conditionQDPs:start_meansTRUE                                     0.04
## lo_ground_truth:conditionHOPs:trial                                         0.05
## lo_ground_truth:conditionintervals:trial                                    0.04
## lo_ground_truth:conditionQDPs:trial                                         0.05
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs                           0.04
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals                      0.03
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs                           0.03
## lo_ground_truth:meansTRUE:sd_diff15:start_meansTRUE                         0.03
## lo_ground_truth:meansTRUE:conditionHOPs:start_meansTRUE                     0.05
## lo_ground_truth:meansTRUE:conditionintervals:start_meansTRUE                0.04
## lo_ground_truth:meansTRUE:conditionQDPs:start_meansTRUE                     0.04
## lo_ground_truth:sd_diff15:conditionHOPs:start_meansTRUE                     0.04
## lo_ground_truth:sd_diff15:conditionintervals:start_meansTRUE                0.03
## lo_ground_truth:sd_diff15:conditionQDPs:start_meansTRUE                     0.03
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                           0.05
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE                      0.04
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                           0.04
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE           0.05
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE      0.04
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE           0.04
## sigma_lo_ground_truth                                                       0.03
## sigma_conditionHOPs                                                         0.12
## sigma_conditionintervals                                                    0.12
## sigma_conditionQDPs                                                         0.12
## sigma_trial                                                                 0.10
## sigma_meansTRUE                                                             0.03
## sigma_start_meansTRUE                                                       0.07
## sigma_lo_ground_truth:conditionHOPs                                         0.05
## sigma_lo_ground_truth:conditionintervals                                    0.05
## sigma_lo_ground_truth:conditionQDPs                                         0.05
## sigma_lo_ground_truth:trial                                                 0.05
## sigma_conditionHOPs:trial                                                   0.14
## sigma_conditionintervals:trial                                              0.14
## sigma_conditionQDPs:trial                                                   0.14
## sigma_meansTRUE:start_meansTRUE                                             0.05
## sigma_lo_ground_truth:conditionHOPs:trial                                   0.07
## sigma_lo_ground_truth:conditionintervals:trial                              0.07
## sigma_lo_ground_truth:conditionQDPs:trial                                   0.07
```

```
##                                                    l-95% CI
## Intercept                                            -0.05
## sigma_Intercept                                      -1.89
## lo_ground_truth                                       0.36
## meansTRUE                                            -0.04
## sd_diff15                                             0.00
## conditionHOPs                                        -0.14
## conditionintervals                                   -0.05
## conditionQDPs                                        -0.02
## start_meansTRUE                                      -0.03
## trial                                                -0.10
## lo_ground_truth:meansTRUE                            -0.09
## lo_ground_truth:sd_diff15                             0.04
## meansTRUE:sd_diff15                                  -0.03
## lo_ground_truth:conditionHOPs                        -0.14
## lo_ground_truth:conditionintervals                   -0.22
## lo_ground_truth:conditionQDPs                        -0.05
## meansTRUE:conditionHOPs                               0.02
## meansTRUE:conditionintervals                         -0.04
## meansTRUE:conditionQDPs                              -0.07
## sd_diff15:conditionHOPs                              -0.04
## sd_diff15:conditionintervals                         -0.04
## sd_diff15:conditionQDPs                              -0.07
## lo_ground_truth:start_meansTRUE                      -0.26
## meansTRUE:start_meansTRUE                            -0.07
## sd_diff15:start_meansTRUE                            -0.05
## conditionHOPs:start_meansTRUE                         0.01
## conditionintervals:start_meansTRUE                   -0.05
## conditionQDPs:start_meansTRUE                        -0.07
## lo_ground_truth:trial                                 0.06
## conditionHOPs:trial                                  -0.06
## conditionintervals:trial                             -0.02
## conditionQDPs:trial                                  -0.01
## lo_ground_truth:meansTRUE:sd_diff15                   0.00
## lo_ground_truth:meansTRUE:conditionHOPs              -0.15
## lo_ground_truth:meansTRUE:conditionintervals         -0.06
## lo_ground_truth:meansTRUE:conditionQDPs              -0.06
## lo_ground_truth:sd_diff15:conditionHOPs               0.00
## lo_ground_truth:sd_diff15:conditionintervals         -0.05
## lo_ground_truth:sd_diff15:conditionQDPs              -0.03
## meansTRUE:sd_diff15:conditionHOPs                    -0.09
## meansTRUE:sd_diff15:conditionintervals               -0.08
## meansTRUE:sd_diff15:conditionQDPs                    -0.06
## lo_ground_truth:meansTRUE:start_meansTRUE            -0.02
## lo_ground_truth:sd_diff15:start_meansTRUE            -0.02
## meansTRUE:sd_diff15:start_meansTRUE                  -0.07
## lo_ground_truth:conditionHOPs:start_meansTRUE        -0.25
## lo_ground_truth:conditionintervals:start_meansTRUE   -0.14
## lo_ground_truth:conditionQDPs:start_meansTRUE        -0.04
## meansTRUE:conditionHOPs:start_meansTRUE              -0.19
## meansTRUE:conditionintervals:start_meansTRUE         -0.06
## meansTRUE:conditionQDPs:start_meansTRUE              -0.05
## sd_diff15:conditionHOPs:start_meansTRUE              -0.11
## sd_diff15:conditionintervals:start_meansTRUE         -0.08
```

```
## sd_diff15:conditionQDPs:start_meansTRUE                                    -0.09
## lo_ground_truth:conditionHOPs:trial                                        -0.13
## lo_ground_truth:conditionintervals:trial                                   -0.09
## lo_ground_truth:conditionQDPs:trial                                        -0.09
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs                          -0.10
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals                     -0.04
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs                          -0.10
## lo_ground_truth:meansTRUE:sd_diff15:start_meansTRUE                        -0.02
## lo_ground_truth:meansTRUE:conditionHOPs:start_meansTRUE                     0.02
## lo_ground_truth:meansTRUE:conditionintervals:start_meansTRUE               -0.05
## lo_ground_truth:meansTRUE:conditionQDPs:start_meansTRUE                    -0.09
## lo_ground_truth:sd_diff15:conditionHOPs:start_meansTRUE                    -0.06
## lo_ground_truth:sd_diff15:conditionintervals:start_meansTRUE               -0.05
## lo_ground_truth:sd_diff15:conditionQDPs:start_meansTRUE                    -0.08
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                          -0.06
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE                     -0.05
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                          -0.07
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE          -0.16
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE     -0.11
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE          -0.08
## sigma_lo_ground_truth                                                       0.39
## sigma_conditionHOPs                                                         0.36
## sigma_conditionintervals                                                   -0.07
## sigma_conditionQDPs                                                        -0.29
## sigma_trial                                                                -0.65
## sigma_meansTRUE                                                            -0.06
## sigma_start_meansTRUE                                                      -0.18
## sigma_lo_ground_truth:conditionHOPs                                        -0.27
## sigma_lo_ground_truth:conditionintervals                                   -0.19
## sigma_lo_ground_truth:conditionQDPs                                        -0.12
## sigma_lo_ground_truth:trial                                                -0.08
## sigma_conditionHOPs:trial                                                  -0.22
## sigma_conditionintervals:trial                                             -0.15
## sigma_conditionQDPs:trial                                                  -0.33
## sigma_meansTRUE:start_meansTRUE                                            -0.32
## sigma_lo_ground_truth:conditionHOPs:trial                                  -0.09
## sigma_lo_ground_truth:conditionintervals:trial                             -0.07
## sigma_lo_ground_truth:conditionQDPs:trial                                  -0.15
##                                                                          u-95% CI
## Intercept                                                                   0.01
## sigma_Intercept                                                            -1.53
## lo_ground_truth                                                             0.54
## meansTRUE                                                                   0.03
## sd_diff15                                                                   0.08
## conditionHOPs                                                              -0.03
## conditionintervals                                                          0.03
## conditionQDPs                                                               0.06
## start_meansTRUE                                                             0.06
## trial                                                                      -0.01
## lo_ground_truth:meansTRUE                                                  -0.00
## lo_ground_truth:sd_diff15                                                   0.11
## meansTRUE:sd_diff15                                                         0.06
## lo_ground_truth:conditionHOPs                                               0.12
## lo_ground_truth:conditionintervals                                          0.03
```

```
## lo_ground_truth:conditionQDPs                                  0.19
## meansTRUE:conditionHOPs                                        0.15
## meansTRUE:conditionintervals                                   0.06
## meansTRUE:conditionQDPs                                        0.03
## sd_diff15:conditionHOPs                                        0.10
## sd_diff15:conditionintervals                                   0.07
## sd_diff15:conditionQDPs                                        0.05
## lo_ground_truth:start_meansTRUE                               -0.02
## meansTRUE:start_meansTRUE                                      0.04
## sd_diff15:start_meansTRUE                                      0.05
## conditionHOPs:start_meansTRUE                                  0.16
## conditionintervals:start_meansTRUE                             0.06
## conditionQDPs:start_meansTRUE                                  0.04
## lo_ground_truth:trial                                          0.19
## conditionHOPs:trial                                            0.09
## conditionintervals:trial                                       0.10
## conditionQDPs:trial                                            0.11
## lo_ground_truth:meansTRUE:sd_diff15                            0.10
## lo_ground_truth:meansTRUE:conditionHOPs                       -0.01
## lo_ground_truth:meansTRUE:conditionintervals                   0.04
## lo_ground_truth:meansTRUE:conditionQDPs                        0.05
## lo_ground_truth:sd_diff15:conditionHOPs                        0.12
## lo_ground_truth:sd_diff15:conditionintervals                   0.04
## lo_ground_truth:sd_diff15:conditionQDPs                        0.08
## meansTRUE:sd_diff15:conditionHOPs                              0.08
## meansTRUE:sd_diff15:conditionintervals                         0.04
## meansTRUE:sd_diff15:conditionQDPs                              0.07
## lo_ground_truth:meansTRUE:start_meansTRUE                      0.10
## lo_ground_truth:sd_diff15:start_meansTRUE                      0.07
## meansTRUE:sd_diff15:start_meansTRUE                            0.04
## lo_ground_truth:conditionHOPs:start_meansTRUE                  0.11
## lo_ground_truth:conditionintervals:start_meansTRUE             0.20
## lo_ground_truth:conditionQDPs:start_meansTRUE                  0.31
## meansTRUE:conditionHOPs:start_meansTRUE                        0.02
## meansTRUE:conditionintervals:start_meansTRUE                   0.09
## meansTRUE:conditionQDPs:start_meansTRUE                        0.10
## sd_diff15:conditionHOPs:start_meansTRUE                        0.07
## sd_diff15:conditionintervals:start_meansTRUE                   0.06
## sd_diff15:conditionQDPs:start_meansTRUE                        0.05
## lo_ground_truth:conditionHOPs:trial                            0.08
## lo_ground_truth:conditionintervals:trial                       0.09
## lo_ground_truth:conditionQDPs:trial                            0.09
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs              0.05
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals         0.08
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs              0.03
## lo_ground_truth:meansTRUE:sd_diff15:start_meansTRUE            0.09
## lo_ground_truth:meansTRUE:conditionHOPs:start_meansTRUE        0.21
## lo_ground_truth:meansTRUE:conditionintervals:start_meansTRUE   0.10
## lo_ground_truth:meansTRUE:conditionQDPs:start_meansTRUE        0.07
## lo_ground_truth:sd_diff15:conditionHOPs:start_meansTRUE        0.08
## lo_ground_truth:sd_diff15:conditionintervals:start_meansTRUE   0.06
## lo_ground_truth:sd_diff15:conditionQDPs:start_meansTRUE        0.04
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE              0.15
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE         0.10
```

```
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                          0.09
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE          0.02
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE     0.03
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE          0.07
## sigma_lo_ground_truth                                                       0.52
## sigma_conditionHOPs                                                         0.81
## sigma_conditionintervals                                                    0.40
## sigma_conditionQDPs                                                         0.18
## sigma_trial                                                                -0.25
## sigma_meansTRUE                                                             0.06
## sigma_start_meansTRUE                                                       0.10
## sigma_lo_ground_truth:conditionHOPs                                        -0.08
## sigma_lo_ground_truth:conditionintervals                                   -0.01
## sigma_lo_ground_truth:conditionQDPs                                         0.07
## sigma_lo_ground_truth:trial                                                 0.11
## sigma_conditionHOPs:trial                                                   0.34
## sigma_conditionintervals:trial                                              0.39
## sigma_conditionQDPs:trial                                                   0.22
## sigma_meansTRUE:start_meansTRUE                                            -0.13
## sigma_lo_ground_truth:conditionHOPs:trial                                   0.18
## sigma_lo_ground_truth:conditionintervals:trial                              0.20
## sigma_lo_ground_truth:conditionQDPs:trial                                   0.12
##                                                                         Rhat
## Intercept                                                               1.00
## sigma_Intercept                                                         1.00
## lo_ground_truth                                                         1.00
## meansTRUE                                                               1.00
## sd_diff15                                                               1.00
## conditionHOPs                                                           1.00
## conditionintervals                                                      1.00
## conditionQDPs                                                           1.00
## start_meansTRUE                                                         1.00
## trial                                                                   1.00
## lo_ground_truth:meansTRUE                                               1.00
## lo_ground_truth:sd_diff15                                               1.00
## meansTRUE:sd_diff15                                                     1.00
## lo_ground_truth:conditionHOPs                                           1.00
## lo_ground_truth:conditionintervals                                      1.00
## lo_ground_truth:conditionQDPs                                           1.00
## meansTRUE:conditionHOPs                                                 1.00
## meansTRUE:conditionintervals                                            1.00
## meansTRUE:conditionQDPs                                                 1.00
## sd_diff15:conditionHOPs                                                 1.00
## sd_diff15:conditionintervals                                            1.00
## sd_diff15:conditionQDPs                                                 1.00
## lo_ground_truth:start_meansTRUE                                         1.00
## meansTRUE:start_meansTRUE                                               1.00
## sd_diff15:start_meansTRUE                                               1.00
## conditionHOPs:start_meansTRUE                                           1.00
## conditionintervals:start_meansTRUE                                      1.00
## conditionQDPs:start_meansTRUE                                           1.00
## lo_ground_truth:trial                                                   1.00
## conditionHOPs:trial                                                     1.00
## conditionintervals:trial                                                1.00
```

```
## conditionQDPs:trial                                                      1.00
## lo_ground_truth:meansTRUE:sd_diff15                                      1.00
## lo_ground_truth:meansTRUE:conditionHOPs                                  1.00
## lo_ground_truth:meansTRUE:conditionintervals                             1.00
## lo_ground_truth:meansTRUE:conditionQDPs                                  1.00
## lo_ground_truth:sd_diff15:conditionHOPs                                  1.00
## lo_ground_truth:sd_diff15:conditionintervals                             1.00
## lo_ground_truth:sd_diff15:conditionQDPs                                  1.00
## meansTRUE:sd_diff15:conditionHOPs                                        1.00
## meansTRUE:sd_diff15:conditionintervals                                   1.00
## meansTRUE:sd_diff15:conditionQDPs                                        1.00
## lo_ground_truth:meansTRUE:start_meansTRUE                                1.00
## lo_ground_truth:sd_diff15:start_meansTRUE                                1.00
## meansTRUE:sd_diff15:start_meansTRUE                                      1.00
## lo_ground_truth:conditionHOPs:start_meansTRUE                            1.00
## lo_ground_truth:conditionintervals:start_meansTRUE                       1.00
## lo_ground_truth:conditionQDPs:start_meansTRUE                            1.00
## meansTRUE:conditionHOPs:start_meansTRUE                                  1.00
## meansTRUE:conditionintervals:start_meansTRUE                             1.00
## meansTRUE:conditionQDPs:start_meansTRUE                                  1.00
## sd_diff15:conditionHOPs:start_meansTRUE                                  1.00
## sd_diff15:conditionintervals:start_meansTRUE                             1.00
## sd_diff15:conditionQDPs:start_meansTRUE                                  1.00
## lo_ground_truth:conditionHOPs:trial                                      1.00
## lo_ground_truth:conditionintervals:trial                                 1.00
## lo_ground_truth:conditionQDPs:trial                                      1.00
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs                        1.00
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals                   1.00
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs                        1.00
## lo_ground_truth:meansTRUE:sd_diff15:start_meansTRUE                      1.00
## lo_ground_truth:meansTRUE:conditionHOPs:start_meansTRUE                  1.00
## lo_ground_truth:meansTRUE:conditionintervals:start_meansTRUE             1.00
## lo_ground_truth:meansTRUE:conditionQDPs:start_meansTRUE                  1.00
## lo_ground_truth:sd_diff15:conditionHOPs:start_meansTRUE                  1.00
## lo_ground_truth:sd_diff15:conditionintervals:start_meansTRUE             1.00
## lo_ground_truth:sd_diff15:conditionQDPs:start_meansTRUE                  1.00
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                        1.00
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE                   1.00
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                        1.00
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE        1.00
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE 1.00
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE        1.00
## sigma_lo_ground_truth                                                    1.00
## sigma_conditionHOPs                                                      1.00
## sigma_conditionintervals                                                 1.00
## sigma_conditionQDPs                                                      1.00
## sigma_trial                                                              1.00
## sigma_meansTRUE                                                          1.00
## sigma_start_meansTRUE                                                    1.00
## sigma_lo_ground_truth:conditionHOPs                                      1.00
## sigma_lo_ground_truth:conditionintervals                                 1.00
## sigma_lo_ground_truth:conditionQDPs                                      1.00
## sigma_lo_ground_truth:trial                                              1.00
## sigma_conditionHOPs:trial                                                1.00
```

```
## sigma_conditionintervals:trial                                    1.00
## sigma_conditionQDPs:trial                                         1.00
## sigma_meansTRUE:start_meansTRUE                                   1.00
## sigma_lo_ground_truth:conditionHOPs:trial                         1.00
## sigma_lo_ground_truth:conditionintervals:trial                    1.00
## sigma_lo_ground_truth:conditionQDPs:trial                         1.00
##                                                                   Bulk_ESS
## Intercept                                                             2580
## sigma_Intercept                                                       1655
## lo_ground_truth                                                       3504
## meansTRUE                                                             2453
## sd_diff15                                                             2639
## conditionHOPs                                                         3514
## conditionintervals                                                    2964
## conditionQDPs                                                         2871
## start_meansTRUE                                                       2462
## trial                                                                 3463
## lo_ground_truth:meansTRUE                                             2696
## lo_ground_truth:sd_diff15                                             2505
## meansTRUE:sd_diff15                                                   2658
## lo_ground_truth:conditionHOPs                                         4181
## lo_ground_truth:conditionintervals                                    3571
## lo_ground_truth:conditionQDPs                                         3670
## meansTRUE:conditionHOPs                                               3465
## meansTRUE:conditionintervals                                          2593
## meansTRUE:conditionQDPs                                               2767
## sd_diff15:conditionHOPs                                               3785
## sd_diff15:conditionintervals                                          3151
## sd_diff15:conditionQDPs                                               3202
## lo_ground_truth:start_meansTRUE                                       3467
## meansTRUE:start_meansTRUE                                             2400
## sd_diff15:start_meansTRUE                                             2584
## conditionHOPs:start_meansTRUE                                         3524
## conditionintervals:start_meansTRUE                                    2934
## conditionQDPs:start_meansTRUE                                         2567
## lo_ground_truth:trial                                                 4230
## conditionHOPs:trial                                                   4865
## conditionintervals:trial                                              4120
## conditionQDPs:trial                                                   3855
## lo_ground_truth:meansTRUE:sd_diff15                                   2540
## lo_ground_truth:meansTRUE:conditionHOPs                               3416
## lo_ground_truth:meansTRUE:conditionintervals                          2894
## lo_ground_truth:meansTRUE:conditionQDPs                               2967
## lo_ground_truth:sd_diff15:conditionHOPs                               3141
## lo_ground_truth:sd_diff15:conditionintervals                          2823
## lo_ground_truth:sd_diff15:conditionQDPs                               2842
## meansTRUE:sd_diff15:conditionHOPs                                     3533
## meansTRUE:sd_diff15:conditionintervals                                3032
## meansTRUE:sd_diff15:conditionQDPs                                     3287
## lo_ground_truth:meansTRUE:start_meansTRUE                             2635
## lo_ground_truth:sd_diff15:start_meansTRUE                             2445
## meansTRUE:sd_diff15:start_meansTRUE                                   2625
## lo_ground_truth:conditionHOPs:start_meansTRUE                         4090
## lo_ground_truth:conditionintervals:start_meansTRUE                    3643
```

```
## lo_ground_truth:conditionQDPs:start_meansTRUE                           3720
## meansTRUE:conditionHOPs:start_meansTRUE                                 3535
## meansTRUE:conditionintervals:start_meansTRUE                            2619
## meansTRUE:conditionQDPs:start_meansTRUE                                 2646
## sd_diff15:conditionHOPs:start_meansTRUE                                 3806
## sd_diff15:conditionintervals:start_meansTRUE                            3192
## sd_diff15:conditionQDPs:start_meansTRUE                                 3112
## lo_ground_truth:conditionHOPs:trial                                     5206
## lo_ground_truth:conditionintervals:trial                                4917
## lo_ground_truth:conditionQDPs:trial                                     5060
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs                       3055
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals                  2877
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs                       2801
## lo_ground_truth:meansTRUE:sd_diff15:start_meansTRUE                     2424
## lo_ground_truth:meansTRUE:conditionHOPs:start_meansTRUE                 3423
## lo_ground_truth:meansTRUE:conditionintervals:start_meansTRUE            2877
## lo_ground_truth:meansTRUE:conditionQDPs:start_meansTRUE                 2890
## lo_ground_truth:sd_diff15:conditionHOPs:start_meansTRUE                 3444
## lo_ground_truth:sd_diff15:conditionintervals:start_meansTRUE            2946
## lo_ground_truth:sd_diff15:conditionQDPs:start_meansTRUE                 2849
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                       3464
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE                  3157
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                       3243
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE       3116
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE  2893
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE       2780
## sigma_lo_ground_truth                                                   2449
## sigma_conditionHOPs                                                     1577
## sigma_conditionintervals                                                1646
## sigma_conditionQDPs                                                     1924
## sigma_trial                                                             5088
## sigma_meansTRUE                                                         8196
## sigma_start_meansTRUE                                                   2687
## sigma_lo_ground_truth:conditionHOPs                                     2485
## sigma_lo_ground_truth:conditionintervals                                2546
## sigma_lo_ground_truth:conditionQDPs                                     2642
## sigma_lo_ground_truth:trial                                             7275
## sigma_conditionHOPs:trial                                               5451
## sigma_conditionintervals:trial                                          4927
## sigma_conditionQDPs:trial                                               5451
## sigma_meansTRUE:start_meansTRUE                                         8703
## sigma_lo_ground_truth:conditionHOPs:trial                               7402
## sigma_lo_ground_truth:conditionintervals:trial                          7054
## sigma_lo_ground_truth:conditionQDPs:trial                               7836
##                                                                     Tail_ESS
## Intercept                                                               4762
## sigma_Intercept                                                         3213
## lo_ground_truth                                                         5086
## meansTRUE                                                               4637
## sd_diff15                                                               5257
## conditionHOPs                                                           5745
## conditionintervals                                                      5420
## conditionQDPs                                                           4919
## start_meansTRUE                                                         4345
```

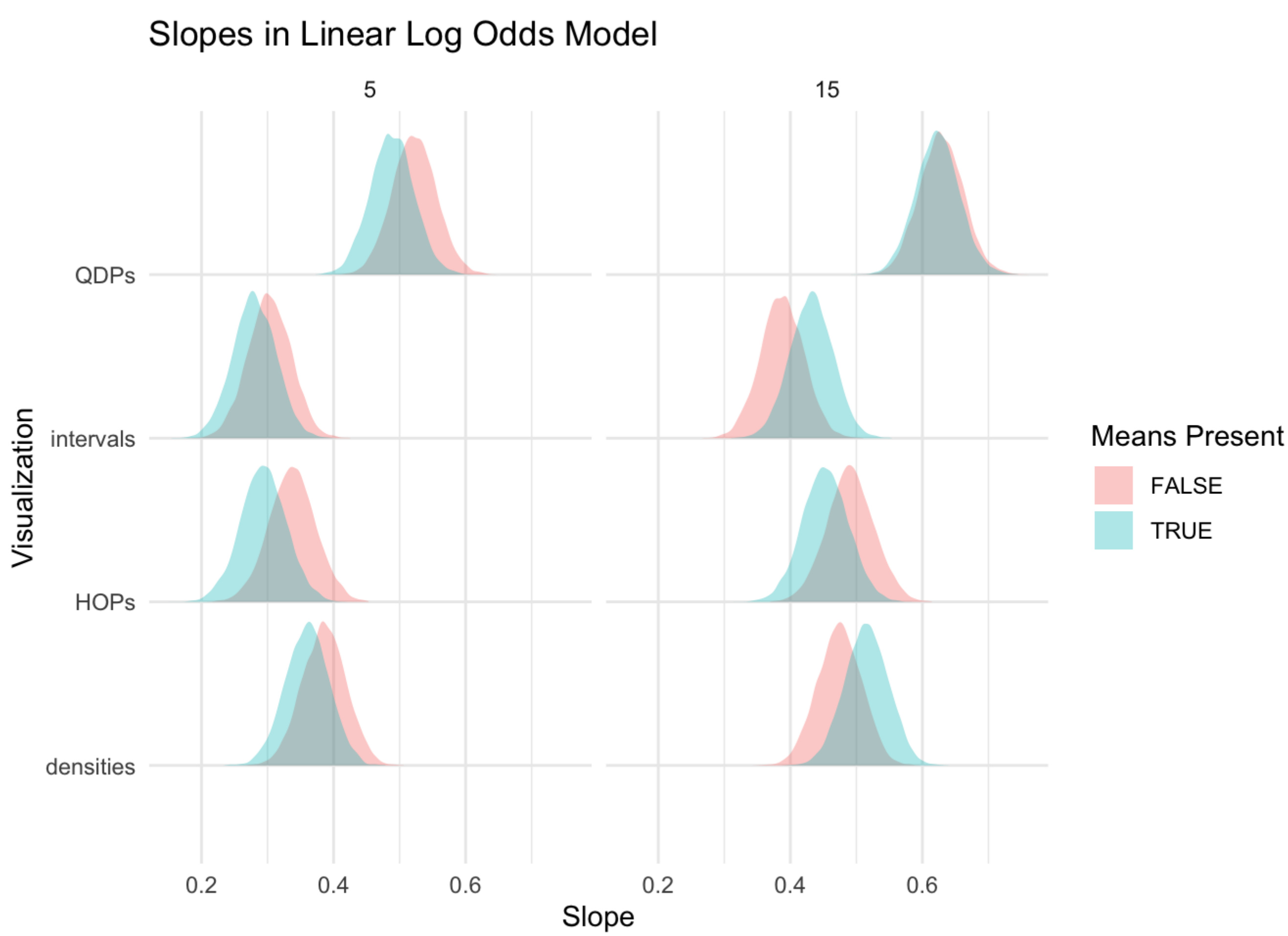
```
## trial                                                           5812
## lo_ground_truth:meansTRUE                                       5245
## lo_ground_truth:sd_diff15                                       5313
## meansTRUE:sd_diff15                                             5779
## lo_ground_truth:conditionHOPs                                   6754
## lo_ground_truth:conditionintervals                              5562
## lo_ground_truth:conditionQDPs                                   6501
## meansTRUE:conditionHOPs                                         6165
## meansTRUE:conditionintervals                                    4680
## meansTRUE:conditionQDPs                                         5192
## sd_diff15:conditionHOPs                                         6343
## sd_diff15:conditionintervals                                    5559
## sd_diff15:conditionQDPs                                         6196
## lo_ground_truth:start_meansTRUE                                 5426
## meansTRUE:start_meansTRUE                                       4675
## sd_diff15:start_meansTRUE                                       5065
## conditionHOPs:start_meansTRUE                                   6332
## conditionintervals:start_meansTRUE                              4873
## conditionQDPs:start_meansTRUE                                   4070
## lo_ground_truth:trial                                           6496
## conditionHOPs:trial                                             7227
## conditionintervals:trial                                        6041
## conditionQDPs:trial                                             6534
## lo_ground_truth:meansTRUE:sd_diff15                             4524
## lo_ground_truth:meansTRUE:conditionHOPs                         6114
## lo_ground_truth:meansTRUE:conditionintervals                    5694
## lo_ground_truth:meansTRUE:conditionQDPs                         5268
## lo_ground_truth:sd_diff15:conditionHOPs                         6623
## lo_ground_truth:sd_diff15:conditionintervals                    5344
## lo_ground_truth:sd_diff15:conditionQDPs                         5599
## meansTRUE:sd_diff15:conditionHOPs                               6233
## meansTRUE:sd_diff15:conditionintervals                          5547
## meansTRUE:sd_diff15:conditionQDPs                               5799
## lo_ground_truth:meansTRUE:start_meansTRUE                       4775
## lo_ground_truth:sd_diff15:start_meansTRUE                       5190
## meansTRUE:sd_diff15:start_meansTRUE                             5272
## lo_ground_truth:conditionHOPs:start_meansTRUE                   6134
## lo_ground_truth:conditionintervals:start_meansTRUE              6001
## lo_ground_truth:conditionQDPs:start_meansTRUE                   6013
## meansTRUE:conditionHOPs:start_meansTRUE                         5655
## meansTRUE:conditionintervals:start_meansTRUE                    4972
## meansTRUE:conditionQDPs:start_meansTRUE                         5144
## sd_diff15:conditionHOPs:start_meansTRUE                         7108
## sd_diff15:conditionintervals:start_meansTRUE                    5579
## sd_diff15:conditionQDPs:start_meansTRUE                         5919
## lo_ground_truth:conditionHOPs:trial                             7002
## lo_ground_truth:conditionintervals:trial                        7467
## lo_ground_truth:conditionQDPs:trial                             7159
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs               6430
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals          5071
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs               4493
## lo_ground_truth:meansTRUE:sd_diff15:start_meansTRUE             4851
## lo_ground_truth:meansTRUE:conditionHOPs:start_meansTRUE         6432
## lo_ground_truth:meansTRUE:conditionintervals:start_meansTRUE    5980
```

```
## lo_ground_truth:meansTRUE:conditionQDPs:start_meansTRUE                     5374
## lo_ground_truth:sd_diff15:conditionHOPs:start_meansTRUE                     6616
## lo_ground_truth:sd_diff15:conditionintervals:start_meansTRUE                5255
## lo_ground_truth:sd_diff15:conditionQDPs:start_meansTRUE                     5729
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                           6080
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE                      5534
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                           5861
## lo_ground_truth:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE           6537
## lo_ground_truth:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE      5046
## lo_ground_truth:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE           4737
## sigma_lo_ground_truth                                                       4855
## sigma_conditionHOPs                                                         3284
## sigma_conditionintervals                                                    3342
## sigma_conditionQDPs                                                         3827
## sigma_trial                                                                 7328
## sigma_meansTRUE                                                             8609
## sigma_start_meansTRUE                                                       4777
## sigma_lo_ground_truth:conditionHOPs                                         4456
## sigma_lo_ground_truth:conditionintervals                                    5287
## sigma_lo_ground_truth:conditionQDPs                                         4578
## sigma_lo_ground_truth:trial                                                 8503
## sigma_conditionHOPs:trial                                                   7735
## sigma_conditionintervals:trial                                              7564
## sigma_conditionQDPs:trial                                                   7428
## sigma_meansTRUE:start_meansTRUE                                             8607
## sigma_lo_ground_truth:conditionHOPs:trial                                   8543
## sigma_lo_ground_truth:conditionintervals:trial                              8990
## sigma_lo_ground_truth:conditionQDPs:trial                                   8508
##
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
## is a crude measure of effective sample size, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

# Interaction Effects

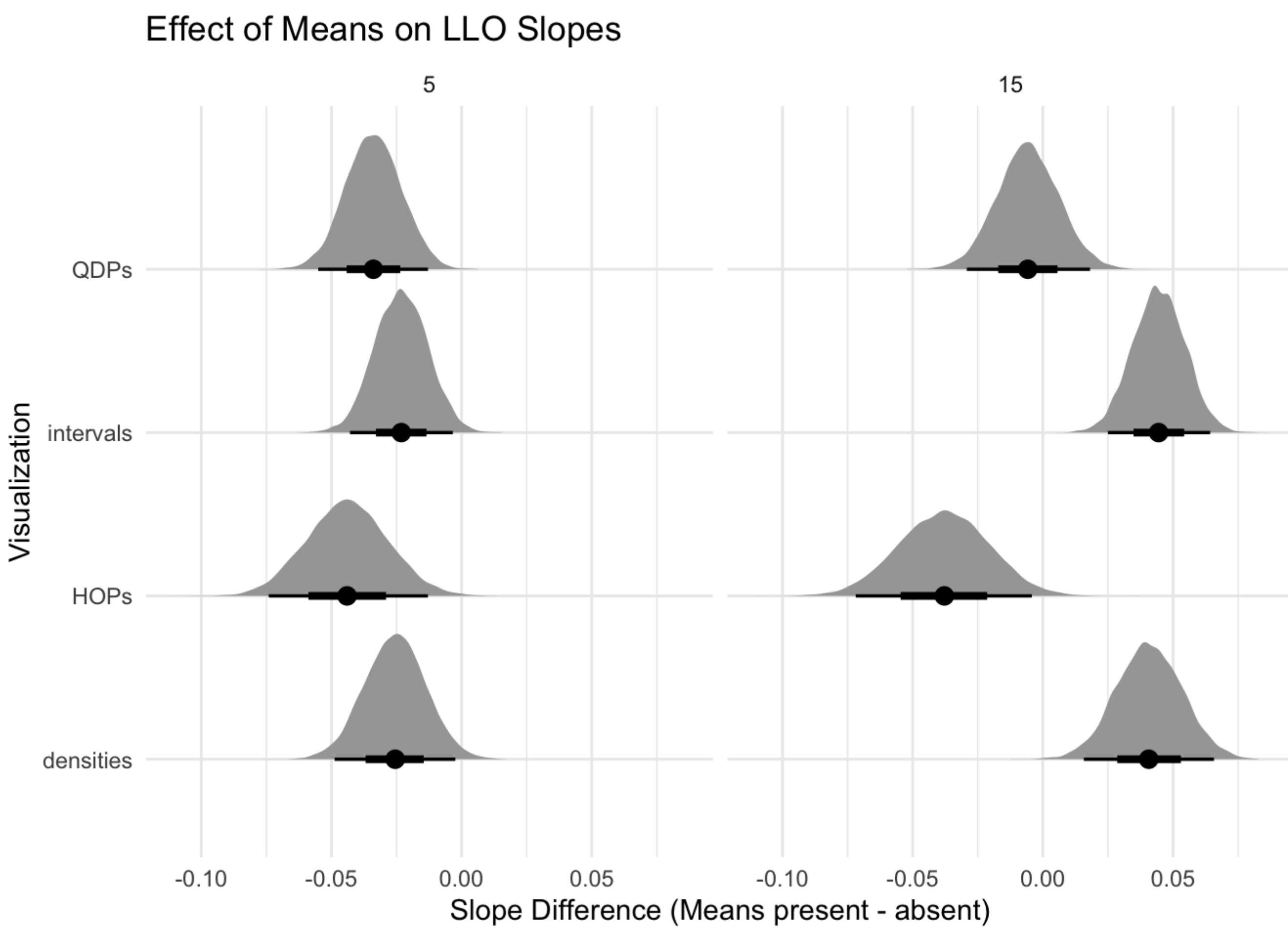
The primary results about probability of superiority that we present in the paper concern the three way interaction between the ground truth probability of superiority, the presence or absence of extrinsic means, and the level of uncertainty shown `lo_ground_truth*means*sd_diff` for each uncertainty visualization format we tested. In order to show this effect, we want to show how the slope of the linear log odds (LLO) model, changes as a function of extrinsic means, uncertainty show, and visualization format. The charts below highlight this effect.

```
model_df %>%
  group_by(means, sd_diff, condition, trial, start_means) %>%
  data_grid(lo_ground_truth = c(0, 1)) %>%                     # get fitted draws (in l
og odds units) only for ground truth of 0 and 1
  add_fitted_draws(m.p_sup, re_formula = NA) %>%
  compare_levels(.value, by = lo_ground_truth) %>%             # calculate the differen
ce between fits at 1 and 0 (i.e., slope)
  rename(slope = .value) %>%
  group_by(means, sd_diff, condition, .draw) %>%   # group by predictors to keep
  summarise(slope = weighted.mean(slope)) %>%      # marginalize out other predictors by
taking a weighted average
  ggplot(aes(x = slope, y = condition, group = means, fill = means)) +
  stat_slabh(alpha = 0.35) +
  labs(
    title = "Slopes in Linear Log Odds Model",
    x = "Slope",
    y = "Visualization",
    fill = "Means Present"
  ) +
  theme_minimal() +
  facet_grid(. ~ sd_diff)
```

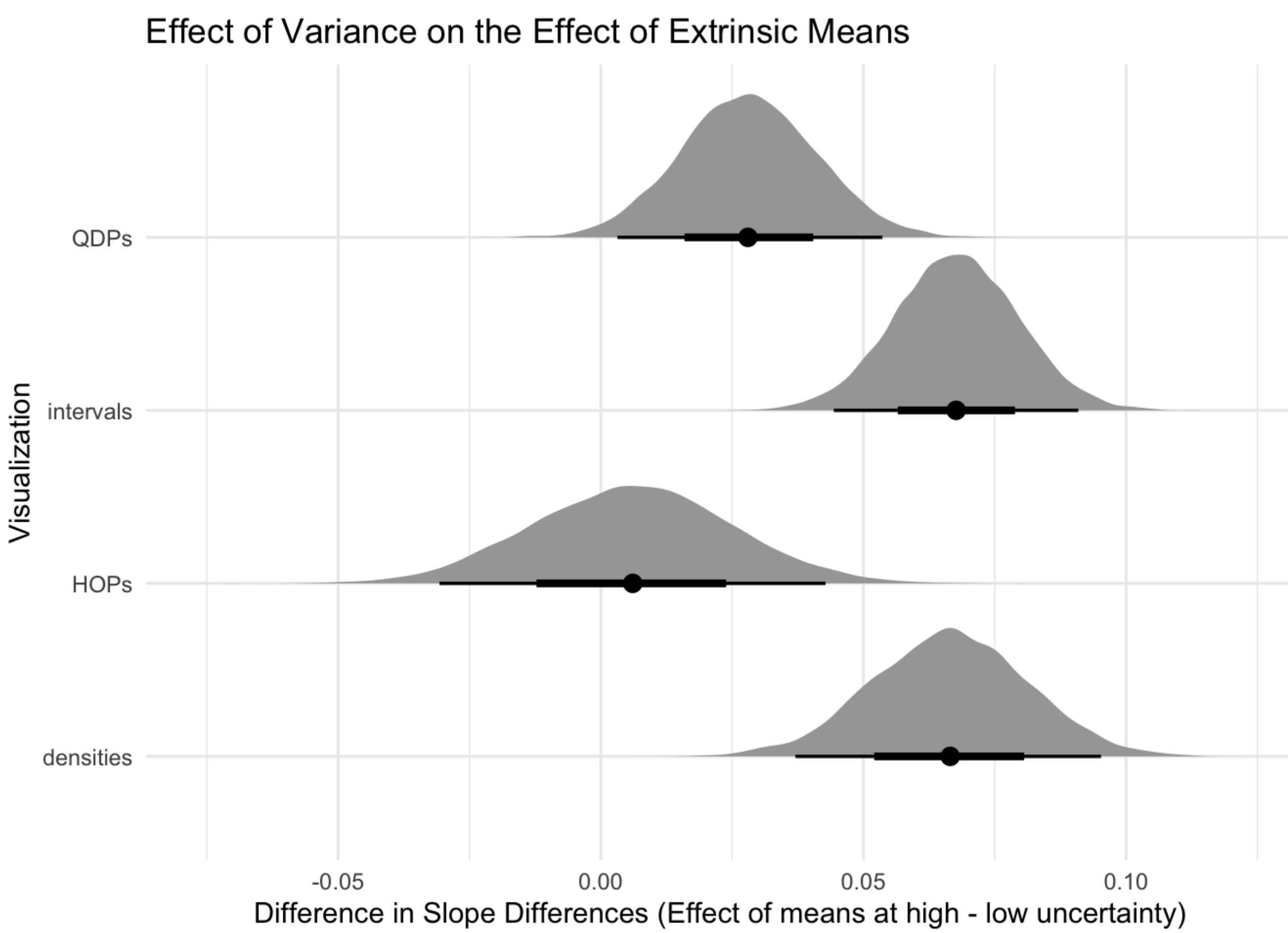



We’ll break this chart down into contrasts and contrasts of contrasts to do some visual reliability testing.

```
model_df %>%
  group_by(means, sd_diff, condition, trial, start_means) %>%
  data_grid(lo_ground_truth = c(0, 1)) %>%                    # get fitted draws (in l
og odds units) only for ground truth of 0 and 1
  add_fitted_draws(m.p_sup, re_formula = NA) %>%
  compare_levels(.value, by = lo_ground_truth) %>%            # calculate the differen
ce between fits at 1 and 0 (i.e., slope)
  rename(slope = .value) %>%
  group_by(means, sd_diff, condition, .draw) %>%   # group by predictors to keep
  summarise(slope = weighted.mean(slope)) %>%      # marginalize out other predictors by
taking a weighted average
  compare_levels(slope, by = means) %>%            # contrast mean present - absent
  ggplot(aes(x = slope, y = condition)) +
  stat_halfeyeh() +
  labs(
    title = "Effect of Means on LLO Slopes",
    x = "Slope Difference (Means present - absent)",
    y = "Visualization"
  ) +
  theme_minimal() +
  facet_grid(. ~ sd_diff)
```



```
model_df %>%
  group_by(means, sd_diff, condition, trial, start_means) %>%
  data_grid(lo_ground_truth = c(0, 1)) %>%                    # get fitted draws (in l
og odds units) only for ground truth of 0 and 1
  add_fitted_draws(m.p_sup, re_formula = NA) %>%
  compare_levels(.value, by = lo_ground_truth) %>%            # calculate the differen
ce between fits at 1 and 0 (i.e., slope)
  rename(slope = .value) %>%
  group_by(means, sd_diff, condition, .draw) %>%   # group by predictors to keep
  summarise(slope = weighted.mean(slope)) %>%      # marginalize out other predictors by
taking a weighted average
  compare_levels(slope, by = means) %>%            # contrast mean present - absent
  compare_levels(slope, by = sd_diff) %>%            # contrast sd_diff high - low (I th
ink)
  ggplot(aes(x = slope, y = condition)) +
  stat_halfeyeh() +
  labs(
    title = "Effect of Variance on the Effect of Extrinsic Means",
    x = "Difference in Slope Differences (Effect of means at high - low uncertainty)",
    y = "Visualization"
  ) +
  theme_minimal()
```

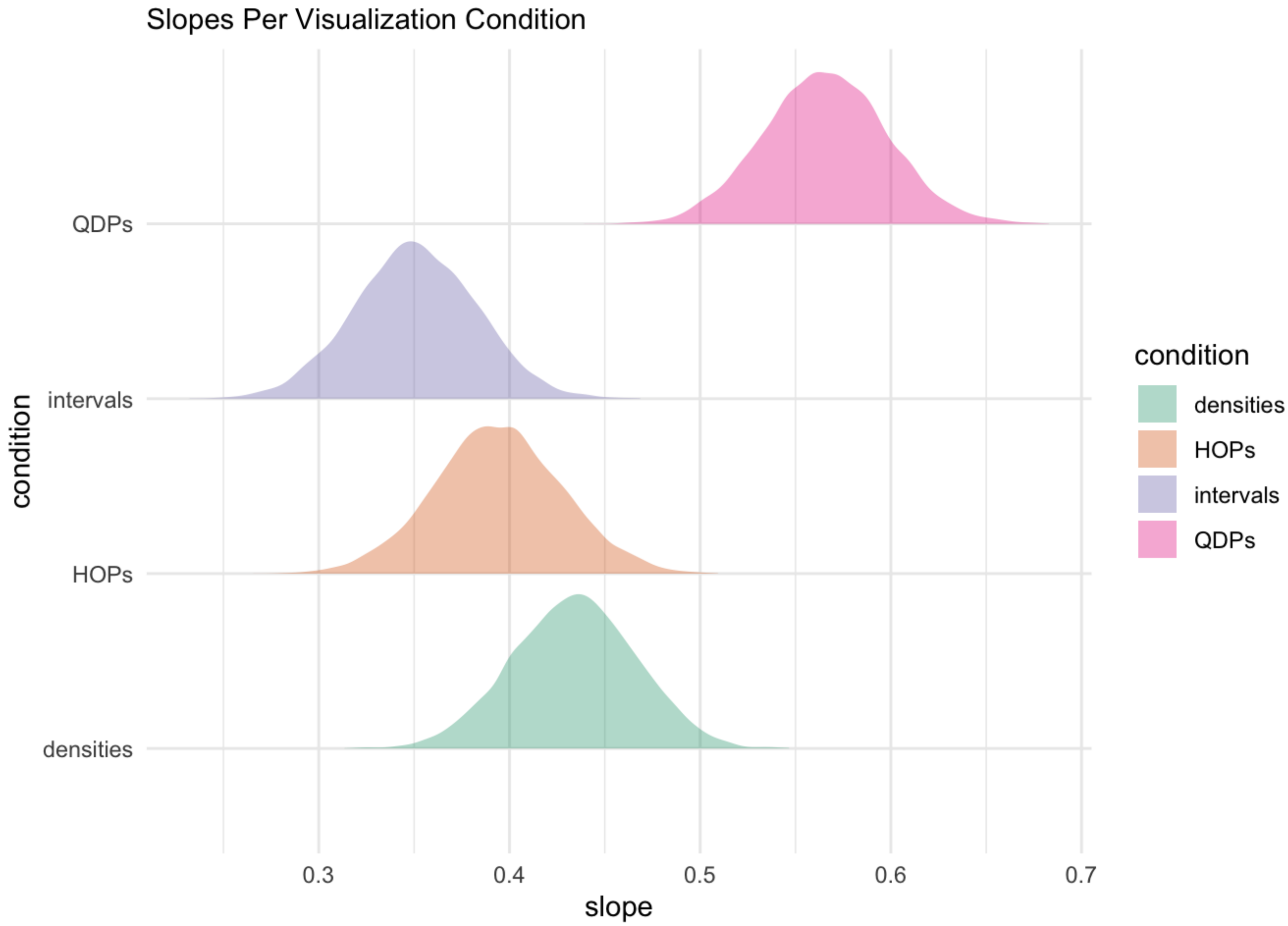



It looks like extrinsic means lead to greater underestimation of probability of superiority (lower LLO slopes) when uncertainty is low, regardless of visualization condition. This is the effect we expected to see but which eluded us until we controlled for order effects. Surprisingly, the impact of extrinsic means does not seem to depend on the intinsic salience of the mean in the uncertainty visualization conditions. At high levels of uncertainty, extrinsic means improve slopes for intervals and densities but still reduce slopes for HOPs. *These results suggest that adding extrinsic means is not a good design choice for HOPs or when the distributions visualized on a common axis differ in their variance*.

# Visualization Effects

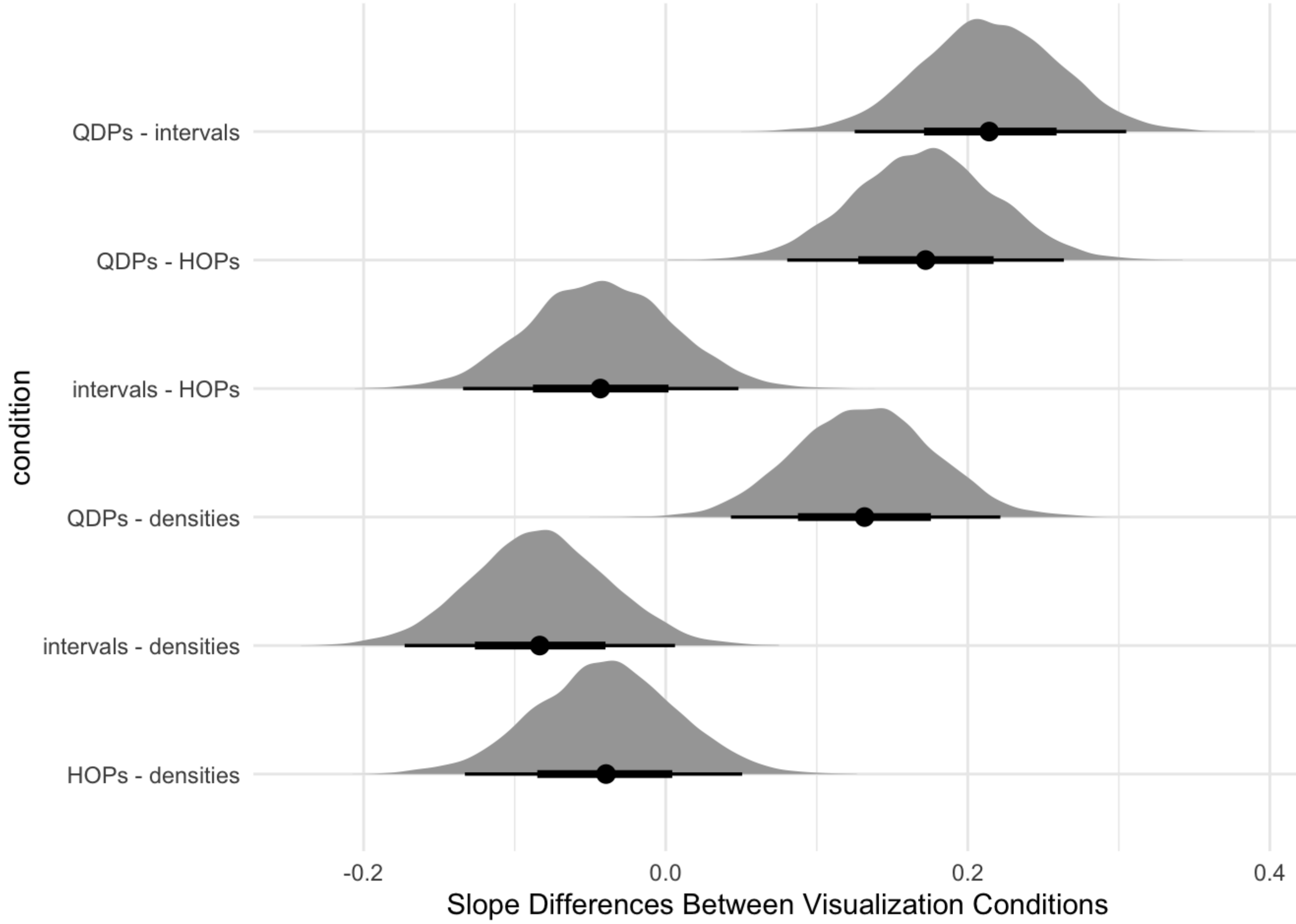
We also want to look at the LLO slopes in each uncertainty visualization condition after adjusting for other predictors.

```
model_df %>%
  group_by(means, sd_diff, condition, trial, start_means) %>%
  data_grid(lo_ground_truth = c(0, 1)) %>%          # get fitted draws (in log odds unit
s) only for ground truth of 0 and 1
  add_fitted_draws(m.p_sup, re_formula = NA) %>%
  compare_levels(.value, by = lo_ground_truth) %>%  # calculate the difference between f
its at 1 and 0 (i.e., slope)
  rename(slope = .value) %>%
  group_by(condition, .draw) %>%                    # group by predictors to keep
  summarise(slope = weighted.mean(slope)) %>%       # marginalize out means present/abse
nt by taking a weighted average
  ggplot(aes(x = slope, y = condition, fill = condition)) +
  stat_slabh(alpha = 0.35) +
  scale_fill_brewer(type = "qual", palette = 2) +
  labs(subtitle = "Slopes Per Visualization Condition") +
  theme_minimal()
```




Let’s look at contrasts between visualization conditions to get a sense of which differences are reliable.

```
model_df %>%
  group_by(means, sd_diff, condition, trial, start_means) %>%
  data_grid(lo_ground_truth = c(0, 1)) %>%          # get fitted draws (in log odds unit
s) only for ground truth of 0 and 1
  add_fitted_draws(m.p_sup, re_formula = NA) %>%
  compare_levels(.value, by = lo_ground_truth) %>%  # calculate the difference between f
its at 1 and 0 (i.e., slope)
  rename(slope = .value) %>%
  group_by(condition, .draw) %>%                    # group by predictors to keep
  summarise(slope = weighted.mean(slope)) %>%       # marginalize out means present/abse
nt by taking a weighted average
  compare_levels(slope, by = condition) %>%
  # compare_levels(slope, by = condition, comparison = list(c("QDPs", "intervals"), c("Q
DPs", "HOPs"), c("QDPs", "densities"), c("densities", "intervals"))) %>%
# show only reliable contrasts
  ggplot(aes(x = slope, y = condition)) +
  stat_halfeyeh() +
  labs(x = "Slope Differences Between Visualization Conditions") +
  theme_minimal()
```




The chart above shows only the contrasts between visualization conditions that seem reliable.

# Intervention Decisions

Next, we load in the model of intervention decisions that we arrived at through a process of model expansion described in our preregistration[https://osf.io/9kpmb (https://osf.io/9kpmb)]. This is a hierachical logistic regression modeling the probability that chart users choose to pay for an intervention based on its effect size compared to status quo if they do not pay. See the paper and **experiment/analysis/InterventionDecisions.Rmd** in the supplemental materials for details.

```
m.decisions <- brm(
  data = model_df, family = bernoulli(link = "logit"),
  formula = bf(intervene ~ (1 + evidence*means*sd_diff + evidence*trial|worker_id) + evi
dence*means*sd_diff*condition*start_means + evidence*trial),
  prior = c(prior(normal(0, 1), class = Intercept),
            prior(normal(1, 1), class = b, coef = evidence),
            prior(normal(0, 0.5), class = b),
            prior(normal(0, 0.5), class = sd),
            prior(lkj(4), class = cor)),
  iter = 8000, warmup = 2000, chains = 2, cores = 2, thin = 2,
  file = "model-fits/logistic_mdl-min_order-r_means_sd_trial-long_chains")
```

```
summary(m.decisions)
```

```
##  Family: bernoulli
##   Links: mu = logit
## Formula: intervene ~ (1 + evidence * means * sd_diff + evidence * trial | worker_id)
+ evidence * means * sd_diff * condition * start_means + evidence * trial
##    Data: model_df (Number of observations: 19924)
## Samples: 2 chains, each with iter = 8000; warmup = 2000; thin = 2;
##          total post-warmup samples = 6000
##
## Group-Level Effects:
## ~worker_id (Number of levels: 623)
##                                                       Estimate Est.Error
## sd(Intercept)                                             1.86      0.09
## sd(evidence)                                              1.24      0.08
## sd(meansTRUE)                                             1.32      0.12
## sd(sd_diff15)                                             1.16      0.09
## sd(trial)                                                 2.50      0.16
## sd(evidence:meansTRUE)                                    0.77      0.11
## sd(evidence:sd_diff15)                                    0.75      0.10
## sd(meansTRUE:sd_diff15)                                   0.73      0.16
## sd(evidence:trial)                                        1.52      0.16
## sd(evidence:meansTRUE:sd_diff15)                          0.67      0.19
## cor(Intercept,evidence)                                   0.54      0.05
## cor(Intercept,meansTRUE)                                 -0.11      0.09
## cor(evidence,meansTRUE)                                   0.05      0.09
## cor(Intercept,sd_diff15)                                 -0.37      0.08
## cor(evidence,sd_diff15)                                  -0.04      0.09
## cor(meansTRUE,sd_diff15)                                  0.24      0.10
## cor(Intercept,trial)                                      0.36      0.07
## cor(evidence,trial)                                       0.13      0.08
## cor(meansTRUE,trial)                                      0.22      0.08
## cor(sd_diff15,trial)                                     -0.05      0.09
## cor(Intercept,evidence:meansTRUE)                        -0.17      0.11
## cor(evidence,evidence:meansTRUE)                         -0.04      0.13
## cor(meansTRUE,evidence:meansTRUE)                         0.44      0.12
## cor(sd_diff15,evidence:meansTRUE)                         0.30      0.12
## cor(trial,evidence:meansTRUE)                             0.17      0.12
## cor(Intercept,evidence:sd_diff15)                        -0.36      0.11
## cor(evidence,evidence:sd_diff15)                         -0.07      0.12
## cor(meansTRUE,evidence:sd_diff15)                        -0.09      0.13
## cor(sd_diff15,evidence:sd_diff15)                         0.64      0.09
## cor(trial,evidence:sd_diff15)                            -0.10      0.12
## cor(evidence:meansTRUE,evidence:sd_diff15)                0.17      0.15
## cor(Intercept,meansTRUE:sd_diff15)                       -0.10      0.15
## cor(evidence,meansTRUE:sd_diff15)                         0.31      0.14
## cor(meansTRUE,meansTRUE:sd_diff15)                       -0.17      0.17
## cor(sd_diff15,meansTRUE:sd_diff15)                        0.05      0.17
## cor(trial,meansTRUE:sd_diff15)                           -0.08      0.16
## cor(evidence:meansTRUE,meansTRUE:sd_diff15)               0.15      0.18
## cor(evidence:sd_diff15,meansTRUE:sd_diff15)               0.23      0.18
## cor(Intercept,evidence:trial)                             0.25      0.09
## cor(evidence,evidence:trial)                              0.38      0.10
## cor(meansTRUE,evidence:trial)                             0.05      0.12
## cor(sd_diff15,evidence:trial)                            -0.17      0.11
```

```
## cor(trial,evidence:trial)                                  0.51      0.09
## cor(evidence:meansTRUE,evidence:trial)                     0.26      0.13
## cor(evidence:sd_diff15,evidence:trial)                    -0.10      0.14
## cor(meansTRUE:sd_diff15,evidence:trial)                    0.13      0.17
## cor(Intercept,evidence:meansTRUE:sd_diff15)               -0.21      0.15
## cor(evidence,evidence:meansTRUE:sd_diff15)                 0.11      0.16
## cor(meansTRUE,evidence:meansTRUE:sd_diff15)               -0.01      0.18
## cor(sd_diff15,evidence:meansTRUE:sd_diff15)                0.05      0.18
## cor(trial,evidence:meansTRUE:sd_diff15)                   -0.08      0.16
## cor(evidence:meansTRUE,evidence:meansTRUE:sd_diff15)      -0.07      0.20
## cor(evidence:sd_diff15,evidence:meansTRUE:sd_diff15)       0.08      0.19
## cor(meansTRUE:sd_diff15,evidence:meansTRUE:sd_diff15)      0.48      0.18
## cor(evidence:trial,evidence:meansTRUE:sd_diff15)          -0.10      0.18
##                                                        l-95% CI u-95% CI Rhat
## sd(Intercept)                                              1.70      2.04 1.00
## sd(evidence)                                               1.09      1.40 1.00
## sd(meansTRUE)                                              1.09      1.54 1.00
## sd(sd_diff15)                                              0.99      1.34 1.00
## sd(trial)                                                  2.18      2.83 1.00
## sd(evidence:meansTRUE)                                     0.55      0.99 1.00
## sd(evidence:sd_diff15)                                     0.56      0.95 1.00
## sd(meansTRUE:sd_diff15)                                    0.40      1.05 1.00
## sd(evidence:trial)                                         1.21      1.83 1.00
## sd(evidence:meansTRUE:sd_diff15)                           0.26      1.01 1.00
## cor(Intercept,evidence)                                    0.44      0.64 1.00
## cor(Intercept,meansTRUE)                                  -0.28      0.07 1.00
## cor(evidence,meansTRUE)                                   -0.12      0.23 1.00
## cor(Intercept,sd_diff15)                                  -0.52     -0.21 1.00
## cor(evidence,sd_diff15)                                   -0.22      0.14 1.00
## cor(meansTRUE,sd_diff15)                                   0.04      0.42 1.00
## cor(Intercept,trial)                                       0.23      0.49 1.00
## cor(evidence,trial)                                       -0.02      0.28 1.00
## cor(meansTRUE,trial)                                       0.06      0.38 1.00
## cor(sd_diff15,trial)                                      -0.23      0.13 1.00
## cor(Intercept,evidence:meansTRUE)                         -0.39      0.05 1.00
## cor(evidence,evidence:meansTRUE)                          -0.29      0.23 1.00
## cor(meansTRUE,evidence:meansTRUE)                          0.19      0.66 1.00
## cor(sd_diff15,evidence:meansTRUE)                          0.06      0.54 1.00
## cor(trial,evidence:meansTRUE)                             -0.07      0.40 1.00
## cor(Intercept,evidence:sd_diff15)                         -0.56     -0.14 1.00
## cor(evidence,evidence:sd_diff15)                          -0.31      0.18 1.00
## cor(meansTRUE,evidence:sd_diff15)                         -0.33      0.17 1.00
## cor(sd_diff15,evidence:sd_diff15)                          0.44      0.81 1.00
## cor(trial,evidence:sd_diff15)                             -0.34      0.12 1.00
## cor(evidence:meansTRUE,evidence:sd_diff15)                -0.13      0.45 1.00
## cor(Intercept,meansTRUE:sd_diff15)                        -0.40      0.20 1.00
## cor(evidence,meansTRUE:sd_diff15)                          0.02      0.58 1.00
## cor(meansTRUE,meansTRUE:sd_diff15)                        -0.47      0.18 1.00
## cor(sd_diff15,meansTRUE:sd_diff15)                        -0.27      0.40 1.00
## cor(trial,meansTRUE:sd_diff15)                            -0.38      0.24 1.00
## cor(evidence:meansTRUE,meansTRUE:sd_diff15)               -0.20      0.50 1.00
## cor(evidence:sd_diff15,meansTRUE:sd_diff15)               -0.13      0.57 1.00
## cor(Intercept,evidence:trial)                              0.08      0.43 1.00
## cor(evidence,evidence:trial)                               0.19      0.57 1.00
```

```
## cor(meansTRUE,evidence:trial)                          -0.18     0.27 1.00
## cor(sd_diff15,evidence:trial)                          -0.38     0.04 1.00
## cor(trial,evidence:trial)                               0.34     0.68 1.00
## cor(evidence:meansTRUE,evidence:trial)                 -0.00     0.50 1.00
## cor(evidence:sd_diff15,evidence:trial)                 -0.37     0.17 1.00
## cor(meansTRUE:sd_diff15,evidence:trial)                -0.21     0.46 1.00
## cor(Intercept,evidence:meansTRUE:sd_diff15)            -0.49     0.08 1.00
## cor(evidence,evidence:meansTRUE:sd_diff15)             -0.22     0.40 1.00
## cor(meansTRUE,evidence:meansTRUE:sd_diff15)            -0.37     0.34 1.00
## cor(sd_diff15,evidence:meansTRUE:sd_diff15)            -0.29     0.40 1.00
## cor(trial,evidence:meansTRUE:sd_diff15)                -0.39     0.24 1.00
## cor(evidence:meansTRUE,evidence:meansTRUE:sd_diff15)   -0.42     0.33 1.00
## cor(evidence:sd_diff15,evidence:meansTRUE:sd_diff15)   -0.27     0.46 1.00
## cor(meansTRUE:sd_diff15,evidence:meansTRUE:sd_diff15)   0.06     0.77 1.00
## cor(evidence:trial,evidence:meansTRUE:sd_diff15)       -0.44     0.26 1.00
##                                                      Bulk_ESS Tail_ESS
## sd(Intercept)                                            3677     4519
## sd(evidence)                                             3813     4963
## sd(meansTRUE)                                            2027     2972
## sd(sd_diff15)                                            3313     4691
## sd(trial)                                                3220     4741
## sd(evidence:meansTRUE)                                   1557     2593
## sd(evidence:sd_diff15)                                   2365     4201
## sd(meansTRUE:sd_diff15)                                  1402     2249
## sd(evidence:trial)                                       2571     4219
## sd(evidence:meansTRUE:sd_diff15)                          988     1582
## cor(Intercept,evidence)                                  3175     4410
## cor(Intercept,meansTRUE)                                 2976     3738
## cor(evidence,meansTRUE)                                  2907     4196
## cor(Intercept,sd_diff15)                                 3697     4683
## cor(evidence,sd_diff15)                                  2639     4702
## cor(meansTRUE,sd_diff15)                                 1992     3185
## cor(Intercept,trial)                                     3168     4552
## cor(evidence,trial)                                      2504     4105
## cor(meansTRUE,trial)                                     2340     3770
## cor(sd_diff15,trial)                                     2242     3444
## cor(Intercept,evidence:meansTRUE)                        3296     4512
## cor(evidence,evidence:meansTRUE)                         2895     3501
## cor(meansTRUE,evidence:meansTRUE)                        1887     3124
## cor(sd_diff15,evidence:meansTRUE)                        1770     3369
## cor(trial,evidence:meansTRUE)                            1866     3366
## cor(Intercept,evidence:sd_diff15)                        3858     4853
## cor(evidence,evidence:sd_diff15)                         3829     4342
## cor(meansTRUE,evidence:sd_diff15)                        2743     4225
## cor(sd_diff15,evidence:sd_diff15)                        2805     4317
## cor(trial,evidence:sd_diff15)                            3512     4423
## cor(evidence:meansTRUE,evidence:sd_diff15)               2280     4067
## cor(Intercept,meansTRUE:sd_diff15)                       4086     4917
## cor(evidence,meansTRUE:sd_diff15)                        3809     4755
## cor(meansTRUE,meansTRUE:sd_diff15)                       3245     4435
## cor(sd_diff15,meansTRUE:sd_diff15)                       2556     3875
## cor(trial,meansTRUE:sd_diff15)                           2874     4386
## cor(evidence:meansTRUE,meansTRUE:sd_diff15)              2569     4096
## cor(evidence:sd_diff15,meansTRUE:sd_diff15)              2359     3571
```

```
## cor(Intercept,evidence:trial)                                    3465     4797
## cor(evidence,evidence:trial)                                     3305     4903
## cor(meansTRUE,evidence:trial)                                    2206     4040
## cor(sd_diff15,evidence:trial)                                    2922     4698
## cor(trial,evidence:trial)                                        2977     4499
## cor(evidence:meansTRUE,evidence:trial)                           2314     3769
## cor(evidence:sd_diff15,evidence:trial)                           2123     3253
## cor(meansTRUE:sd_diff15,evidence:trial)                          1755     3306
## cor(Intercept,evidence:meansTRUE:sd_diff15)                      4834     5025
## cor(evidence,evidence:meansTRUE:sd_diff15)                       4528     5298
## cor(meansTRUE,evidence:meansTRUE:sd_diff15)                      3660     4700
## cor(sd_diff15,evidence:meansTRUE:sd_diff15)                      3595     4547
## cor(trial,evidence:meansTRUE:sd_diff15)                          4215     4865
## cor(evidence:meansTRUE,evidence:meansTRUE:sd_diff15)             2967     4274
## cor(evidence:sd_diff15,evidence:meansTRUE:sd_diff15)             2599     3793
## cor(meansTRUE:sd_diff15,evidence:meansTRUE:sd_diff15)            1792     2821
## cor(evidence:trial,evidence:meansTRUE:sd_diff15)                 2907     4528
##
## Population-Level Effects:
##                                                                       Estimate
## Intercept                                                                 0.32
## evidence                                                                  2.14
## meansTRUE                                                                -0.39
## sd_diff15                                                                 1.07
## conditionHOPs                                                            -0.22
## conditionintervals                                                       -0.41
## conditionQDPs                                                             0.29
## start_meansTRUE                                                          -0.48
## trial                                                                     1.45
## evidence:meansTRUE                                                       -0.13
## evidence:sd_diff15                                                        0.60
## meansTRUE:sd_diff15                                                       0.63
## evidence:conditionHOPs                                                   -0.10
## evidence:conditionintervals                                              -0.30
## evidence:conditionQDPs                                                    0.27
## meansTRUE:conditionHOPs                                                  -0.01
## meansTRUE:conditionintervals                                              0.03
## meansTRUE:conditionQDPs                                                  -0.34
## sd_diff15:conditionHOPs                                                   0.47
## sd_diff15:conditionintervals                                              0.36
## sd_diff15:conditionQDPs                                                   0.10
## evidence:start_meansTRUE                                                 -0.48
## meansTRUE:start_meansTRUE                                                 0.49
## sd_diff15:start_meansTRUE                                                 0.50
## conditionHOPs:start_meansTRUE                                            -0.38
## conditionintervals:start_meansTRUE                                       -0.25
## conditionQDPs:start_meansTRUE                                             0.20
## evidence:trial                                                            1.81
## evidence:meansTRUE:sd_diff15                                              0.02
## evidence:meansTRUE:conditionHOPs                                         -0.38
## evidence:meansTRUE:conditionintervals                                     0.28
## evidence:meansTRUE:conditionQDPs                                          0.11
## evidence:sd_diff15:conditionHOPs                                          0.13
## evidence:sd_diff15:conditionintervals                                     0.35
```

```
## evidence:sd_diff15:conditionQDPs                                      0.23
## meansTRUE:sd_diff15:conditionHOPs                                    -0.53
## meansTRUE:sd_diff15:conditionintervals                                0.55
## meansTRUE:sd_diff15:conditionQDPs                                     0.25
## evidence:meansTRUE:start_meansTRUE                                    0.40
## evidence:sd_diff15:start_meansTRUE                                    0.22
## meansTRUE:sd_diff15:start_meansTRUE                                  -0.05
## evidence:conditionHOPs:start_meansTRUE                               -0.49
## evidence:conditionintervals:start_meansTRUE                           0.24
## evidence:conditionQDPs:start_meansTRUE                               -0.06
## meansTRUE:conditionHOPs:start_meansTRUE                               0.19
## meansTRUE:conditionintervals:start_meansTRUE                         -0.03
## meansTRUE:conditionQDPs:start_meansTRUE                              -0.03
## sd_diff15:conditionHOPs:start_meansTRUE                               0.13
## sd_diff15:conditionintervals:start_meansTRUE                         -0.31
## sd_diff15:conditionQDPs:start_meansTRUE                              -0.23
## evidence:meansTRUE:sd_diff15:conditionHOPs                           -0.50
## evidence:meansTRUE:sd_diff15:conditionintervals                       0.28
## evidence:meansTRUE:sd_diff15:conditionQDPs                           -0.04
## evidence:meansTRUE:sd_diff15:start_meansTRUE                          0.31
## evidence:meansTRUE:conditionHOPs:start_meansTRUE                      0.83
## evidence:meansTRUE:conditionintervals:start_meansTRUE                -0.31
## evidence:meansTRUE:conditionQDPs:start_meansTRUE                      0.00
## evidence:sd_diff15:conditionHOPs:start_meansTRUE                      0.07
## evidence:sd_diff15:conditionintervals:start_meansTRUE                -0.37
## evidence:sd_diff15:conditionQDPs:start_meansTRUE                      0.00
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                     0.34
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE               -0.27
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                    -0.18
## evidence:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE            0.32
## evidence:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE       0.25
## evidence:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE           -0.09
##                                                                 Est.Error
## Intercept                                                            0.17
## evidence                                                             0.15
## meansTRUE                                                            0.18
## sd_diff15                                                            0.16
## conditionHOPs                                                        0.23
## conditionintervals                                                   0.23
## conditionQDPs                                                        0.23
## start_meansTRUE                                                      0.22
## trial                                                                0.18
## evidence:meansTRUE                                                   0.18
## evidence:sd_diff15                                                   0.16
## meansTRUE:sd_diff15                                                  0.19
## evidence:conditionHOPs                                               0.19
## evidence:conditionintervals                                          0.19
## evidence:conditionQDPs                                               0.20
## meansTRUE:conditionHOPs                                              0.25
## meansTRUE:conditionintervals                                         0.25
## meansTRUE:conditionQDPs                                              0.26
## sd_diff15:conditionHOPs                                              0.22
## sd_diff15:conditionintervals                                         0.22
## sd_diff15:conditionQDPs                                              0.22
```

```
## evidence:start_meansTRUE                                              0.19
## meansTRUE:start_meansTRUE                                             0.23
## sd_diff15:start_meansTRUE                                             0.20
## conditionHOPs:start_meansTRUE                                         0.30
## conditionintervals:start_meansTRUE                                    0.30
## conditionQDPs:start_meansTRUE                                         0.30
## evidence:trial                                                        0.18
## evidence:meansTRUE:sd_diff15                                          0.21
## evidence:meansTRUE:conditionHOPs                                      0.21
## evidence:meansTRUE:conditionintervals                                 0.22
## evidence:meansTRUE:conditionQDPs                                      0.24
## evidence:sd_diff15:conditionHOPs                                      0.21
## evidence:sd_diff15:conditionintervals                                 0.21
## evidence:sd_diff15:conditionQDPs                                      0.22
## meansTRUE:sd_diff15:conditionHOPs                                     0.25
## meansTRUE:sd_diff15:conditionintervals                                0.26
## meansTRUE:sd_diff15:conditionQDPs                                     0.26
## evidence:meansTRUE:start_meansTRUE                                    0.22
## evidence:sd_diff15:start_meansTRUE                                    0.20
## meansTRUE:sd_diff15:start_meansTRUE                                   0.23
## evidence:conditionHOPs:start_meansTRUE                                0.26
## evidence:conditionintervals:start_meansTRUE                           0.27
## evidence:conditionQDPs:start_meansTRUE                                0.27
## meansTRUE:conditionHOPs:start_meansTRUE                               0.31
## meansTRUE:conditionintervals:start_meansTRUE                          0.30
## meansTRUE:conditionQDPs:start_meansTRUE                               0.31
## sd_diff15:conditionHOPs:start_meansTRUE                               0.28
## sd_diff15:conditionintervals:start_meansTRUE                          0.28
## sd_diff15:conditionQDPs:start_meansTRUE                               0.28
## evidence:meansTRUE:sd_diff15:conditionHOPs                            0.25
## evidence:meansTRUE:sd_diff15:conditionintervals                       0.26
## evidence:meansTRUE:sd_diff15:conditionQDPs                            0.27
## evidence:meansTRUE:sd_diff15:start_meansTRUE                          0.24
## evidence:meansTRUE:conditionHOPs:start_meansTRUE                      0.26
## evidence:meansTRUE:conditionintervals:start_meansTRUE                 0.27
## evidence:meansTRUE:conditionQDPs:start_meansTRUE                      0.28
## evidence:sd_diff15:conditionHOPs:start_meansTRUE                      0.27
## evidence:sd_diff15:conditionintervals:start_meansTRUE                 0.28
## evidence:sd_diff15:conditionQDPs:start_meansTRUE                      0.28
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                     0.31
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE                0.31
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                     0.32
## evidence:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE            0.31
## evidence:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE       0.33
## evidence:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE            0.33
##                                                                    l-95% CI
## Intercept                                                             -0.02
## evidence                                                               1.85
## meansTRUE                                                             -0.73
## sd_diff15                                                              0.76
## conditionHOPs                                                         -0.68
## conditionintervals                                                    -0.85
## conditionQDPs                                                         -0.17
## start_meansTRUE                                                       -0.91
```

```
## trial                                                       1.09
## evidence:meansTRUE                                         -0.47
## evidence:sd_diff15                                          0.30
## meansTRUE:sd_diff15                                         0.27
## evidence:conditionHOPs                                     -0.47
## evidence:conditionintervals                                -0.67
## evidence:conditionQDPs                                     -0.11
## meansTRUE:conditionHOPs                                    -0.49
## meansTRUE:conditionintervals                               -0.45
## meansTRUE:conditionQDPs                                    -0.83
## sd_diff15:conditionHOPs                                     0.04
## sd_diff15:conditionintervals                               -0.06
## sd_diff15:conditionQDPs                                    -0.32
## evidence:start_meansTRUE                                   -0.84
## meansTRUE:start_meansTRUE                                   0.03
## sd_diff15:start_meansTRUE                                   0.10
## conditionHOPs:start_meansTRUE                              -0.95
## conditionintervals:start_meansTRUE                         -0.85
## conditionQDPs:start_meansTRUE                              -0.38
## evidence:trial                                              1.45
## evidence:meansTRUE:sd_diff15                               -0.38
## evidence:meansTRUE:conditionHOPs                           -0.80
## evidence:meansTRUE:conditionintervals                      -0.16
## evidence:meansTRUE:conditionQDPs                           -0.36
## evidence:sd_diff15:conditionHOPs                           -0.28
## evidence:sd_diff15:conditionintervals                      -0.06
## evidence:sd_diff15:conditionQDPs                           -0.20
## meansTRUE:sd_diff15:conditionHOPs                          -1.01
## meansTRUE:sd_diff15:conditionintervals                      0.04
## meansTRUE:sd_diff15:conditionQDPs                          -0.25
## evidence:meansTRUE:start_meansTRUE                         -0.03
## evidence:sd_diff15:start_meansTRUE                         -0.17
## meansTRUE:sd_diff15:start_meansTRUE                        -0.50
## evidence:conditionHOPs:start_meansTRUE                     -1.00
## evidence:conditionintervals:start_meansTRUE                -0.28
## evidence:conditionQDPs:start_meansTRUE                     -0.58
## meansTRUE:conditionHOPs:start_meansTRUE                    -0.41
## meansTRUE:conditionintervals:start_meansTRUE               -0.60
## meansTRUE:conditionQDPs:start_meansTRUE                    -0.63
## sd_diff15:conditionHOPs:start_meansTRUE                    -0.42
## sd_diff15:conditionintervals:start_meansTRUE               -0.85
## sd_diff15:conditionQDPs:start_meansTRUE                    -0.79
## evidence:meansTRUE:sd_diff15:conditionHOPs                 -0.98
## evidence:meansTRUE:sd_diff15:conditionintervals            -0.24
## evidence:meansTRUE:sd_diff15:conditionQDPs                 -0.58
## evidence:meansTRUE:sd_diff15:start_meansTRUE               -0.15
## evidence:meansTRUE:conditionHOPs:start_meansTRUE            0.31
## evidence:meansTRUE:conditionintervals:start_meansTRUE      -0.85
## evidence:meansTRUE:conditionQDPs:start_meansTRUE           -0.56
## evidence:sd_diff15:conditionHOPs:start_meansTRUE           -0.46
## evidence:sd_diff15:conditionintervals:start_meansTRUE      -0.91
## evidence:sd_diff15:conditionQDPs:start_meansTRUE           -0.54
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE          -0.27
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE     -0.89
```

```
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                 -0.80
## evidence:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE        -0.30
## evidence:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE   -0.38
## evidence:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE        -0.73
##                                                                 u-95% CI Rhat
## Intercept                                                            0.66 1.00
## evidence                                                             2.44 1.00
## meansTRUE                                                           -0.03 1.00
## sd_diff15                                                            1.38 1.00
## conditionHOPs                                                        0.23 1.00
## conditionintervals                                                   0.04 1.00
## conditionQDPs                                                        0.73 1.00
## start_meansTRUE                                                     -0.05 1.00
## trial                                                                1.80 1.00
## evidence:meansTRUE                                                   0.23 1.00
## evidence:sd_diff15                                                   0.92 1.00
## meansTRUE:sd_diff15                                                  1.00 1.00
## evidence:conditionHOPs                                               0.28 1.00
## evidence:conditionintervals                                          0.07 1.00
## evidence:conditionQDPs                                               0.67 1.00
## meansTRUE:conditionHOPs                                              0.49 1.00
## meansTRUE:conditionintervals                                         0.51 1.00
## meansTRUE:conditionQDPs                                              0.16 1.00
## sd_diff15:conditionHOPs                                              0.91 1.00
## sd_diff15:conditionintervals                                         0.78 1.00
## sd_diff15:conditionQDPs                                              0.52 1.00
## evidence:start_meansTRUE                                            -0.10 1.00
## meansTRUE:start_meansTRUE                                            0.94 1.00
## sd_diff15:start_meansTRUE                                            0.89 1.00
## conditionHOPs:start_meansTRUE                                        0.21 1.00
## conditionintervals:start_meansTRUE                                   0.34 1.00
## conditionQDPs:start_meansTRUE                                        0.80 1.00
## evidence:trial                                                       2.16 1.00
## evidence:meansTRUE:sd_diff15                                         0.43 1.00
## evidence:meansTRUE:conditionHOPs                                     0.04 1.00
## evidence:meansTRUE:conditionintervals                                0.72 1.00
## evidence:meansTRUE:conditionQDPs                                     0.57 1.00
## evidence:sd_diff15:conditionHOPs                                     0.54 1.00
## evidence:sd_diff15:conditionintervals                                0.76 1.00
## evidence:sd_diff15:conditionQDPs                                     0.65 1.00
## meansTRUE:sd_diff15:conditionHOPs                                   -0.04 1.00
## meansTRUE:sd_diff15:conditionintervals                               1.05 1.00
## meansTRUE:sd_diff15:conditionQDPs                                    0.76 1.00
## evidence:meansTRUE:start_meansTRUE                                   0.83 1.00
## evidence:sd_diff15:start_meansTRUE                                   0.60 1.00
## meansTRUE:sd_diff15:start_meansTRUE                                  0.40 1.00
## evidence:conditionHOPs:start_meansTRUE                               0.00 1.00
## evidence:conditionintervals:start_meansTRUE                          0.76 1.00
## evidence:conditionQDPs:start_meansTRUE                               0.45 1.00
## meansTRUE:conditionHOPs:start_meansTRUE                              0.79 1.00
## meansTRUE:conditionintervals:start_meansTRUE                         0.57 1.00
## meansTRUE:conditionQDPs:start_meansTRUE                              0.58 1.00
## sd_diff15:conditionHOPs:start_meansTRUE                              0.67 1.00
## sd_diff15:conditionintervals:start_meansTRUE                         0.24 1.00
```

```
## sd_diff15:conditionQDPs:start_meansTRUE                          0.32 1.00
## evidence:meansTRUE:sd_diff15:conditionHOPs                      -0.03 1.00
## evidence:meansTRUE:sd_diff15:conditionintervals                  0.80 1.00
## evidence:meansTRUE:sd_diff15:conditionQDPs                       0.49 1.00
## evidence:meansTRUE:sd_diff15:start_meansTRUE                     0.76 1.00
## evidence:meansTRUE:conditionHOPs:start_meansTRUE                 1.35 1.00
## evidence:meansTRUE:conditionintervals:start_meansTRUE            0.22 1.00
## evidence:meansTRUE:conditionQDPs:start_meansTRUE                 0.56 1.00
## evidence:sd_diff15:conditionHOPs:start_meansTRUE                 0.59 1.00
## evidence:sd_diff15:conditionintervals:start_meansTRUE            0.18 1.00
## evidence:sd_diff15:conditionQDPs:start_meansTRUE                 0.56 1.00
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                0.94 1.00
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE           0.34 1.00
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                0.43 1.00
## evidence:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE       0.94 1.00
## evidence:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE  0.90 1.00
## evidence:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE       0.56 1.00
##                                                               Bulk_ESS
## Intercept                                                         3144
## evidence                                                          3259
## meansTRUE                                                         4044
## sd_diff15                                                         3946
## conditionHOPs                                                     3577
## conditionintervals                                                3063
## conditionQDPs                                                     3703
## start_meansTRUE                                                   3582
## trial                                                             5084
## evidence:meansTRUE                                                4395
## evidence:sd_diff15                                                3987
## meansTRUE:sd_diff15                                               4977
## evidence:conditionHOPs                                            3972
## evidence:conditionintervals                                       4378
## evidence:conditionQDPs                                            3867
## meansTRUE:conditionHOPs                                           4605
## meansTRUE:conditionintervals                                      4909
## meansTRUE:conditionQDPs                                           4686
## sd_diff15:conditionHOPs                                           4623
## sd_diff15:conditionintervals                                      4627
## sd_diff15:conditionQDPs                                           4715
## evidence:start_meansTRUE                                          3680
## meansTRUE:start_meansTRUE                                         5253
## sd_diff15:start_meansTRUE                                         4374
## conditionHOPs:start_meansTRUE                                     4371
## conditionintervals:start_meansTRUE                                3795
## conditionQDPs:start_meansTRUE                                     4441
## evidence:trial                                                    4917
## evidence:meansTRUE:sd_diff15                                      4175
## evidence:meansTRUE:conditionHOPs                                  4760
## evidence:meansTRUE:conditionintervals                             5467
## evidence:meansTRUE:conditionQDPs                                  5038
## evidence:sd_diff15:conditionHOPs                                  4688
## evidence:sd_diff15:conditionintervals                             4639
## evidence:sd_diff15:conditionQDPs                                  4952
## meansTRUE:sd_diff15:conditionHOPs                                 4407
```

```
## meansTRUE:sd_diff15:conditionintervals                              5249
## meansTRUE:sd_diff15:conditionQDPs                                   5235
## evidence:meansTRUE:start_meansTRUE                                  4687
## evidence:sd_diff15:start_meansTRUE                                  3596
## meansTRUE:sd_diff15:start_meansTRUE                                 5194
## evidence:conditionHOPs:start_meansTRUE                              4528
## evidence:conditionintervals:start_meansTRUE                         3984
## evidence:conditionQDPs:start_meansTRUE                              4528
## meansTRUE:conditionHOPs:start_meansTRUE                             4565
## meansTRUE:conditionintervals:start_meansTRUE                        4881
## meansTRUE:conditionQDPs:start_meansTRUE                             5011
## sd_diff15:conditionHOPs:start_meansTRUE                             5219
## sd_diff15:conditionintervals:start_meansTRUE                        4626
## sd_diff15:conditionQDPs:start_meansTRUE                             4719
## evidence:meansTRUE:sd_diff15:conditionHOPs                          4954
## evidence:meansTRUE:sd_diff15:conditionintervals                     5285
## evidence:meansTRUE:sd_diff15:conditionQDPs                          5394
## evidence:meansTRUE:sd_diff15:start_meansTRUE                        4622
## evidence:meansTRUE:conditionHOPs:start_meansTRUE                    5116
## evidence:meansTRUE:conditionintervals:start_meansTRUE               5363
## evidence:meansTRUE:conditionQDPs:start_meansTRUE                    5288
## evidence:sd_diff15:conditionHOPs:start_meansTRUE                    5135
## evidence:sd_diff15:conditionintervals:start_meansTRUE               4917
## evidence:sd_diff15:conditionQDPs:start_meansTRUE                    4534
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                   4830
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE              5510
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                   5223
## evidence:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE          5348
## evidence:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE     5265
## evidence:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE          5189
##                                                                 Tail_ESS
## Intercept                                                           4700
## evidence                                                            4190
## meansTRUE                                                           5249
## sd_diff15                                                           4969
## conditionHOPs                                                       4789
## conditionintervals                                                  4229
## conditionQDPs                                                       4772
## start_meansTRUE                                                     4895
## trial                                                               5439
## evidence:meansTRUE                                                  5246
## evidence:sd_diff15                                                  4826
## meansTRUE:sd_diff15                                                 5473
## evidence:conditionHOPs                                              4693
## evidence:conditionintervals                                         4956
## evidence:conditionQDPs                                              4668
## meansTRUE:conditionHOPs                                             5503
## meansTRUE:conditionintervals                                        5534
## meansTRUE:conditionQDPs                                             5546
## sd_diff15:conditionHOPs                                             5025
## sd_diff15:conditionintervals                                        4870
## sd_diff15:conditionQDPs                                             5075
## evidence:start_meansTRUE                                            4769
## meansTRUE:start_meansTRUE                                           4736
```

```
## sd_diff15:start_meansTRUE                                          5484
## conditionHOPs:start_meansTRUE                                      5470
## conditionintervals:start_meansTRUE                                 4689
## conditionQDPs:start_meansTRUE                                      5206
## evidence:trial                                                     5347
## evidence:meansTRUE:sd_diff15                                       4860
## evidence:meansTRUE:conditionHOPs                                   5181
## evidence:meansTRUE:conditionintervals                              5226
## evidence:meansTRUE:conditionQDPs                                   5501
## evidence:sd_diff15:conditionHOPs                                   5209
## evidence:sd_diff15:conditionintervals                              4754
## evidence:sd_diff15:conditionQDPs                                   5232
## meansTRUE:sd_diff15:conditionHOPs                                  5068
## meansTRUE:sd_diff15:conditionintervals                             5537
## meansTRUE:sd_diff15:conditionQDPs                                  5546
## evidence:meansTRUE:start_meansTRUE                                 4928
## evidence:sd_diff15:start_meansTRUE                                 5488
## meansTRUE:sd_diff15:start_meansTRUE                                5463
## evidence:conditionHOPs:start_meansTRUE                             5299
## evidence:conditionintervals:start_meansTRUE                        5140
## evidence:conditionQDPs:start_meansTRUE                             5048
## meansTRUE:conditionHOPs:start_meansTRUE                            5422
## meansTRUE:conditionintervals:start_meansTRUE                       5005
## meansTRUE:conditionQDPs:start_meansTRUE                            4671
## sd_diff15:conditionHOPs:start_meansTRUE                            5305
## sd_diff15:conditionintervals:start_meansTRUE                       5179
## sd_diff15:conditionQDPs:start_meansTRUE                            5255
## evidence:meansTRUE:sd_diff15:conditionHOPs                         4731
## evidence:meansTRUE:sd_diff15:conditionintervals                    5648
## evidence:meansTRUE:sd_diff15:conditionQDPs                         5469
## evidence:meansTRUE:sd_diff15:start_meansTRUE                       5180
## evidence:meansTRUE:conditionHOPs:start_meansTRUE                   5432
## evidence:meansTRUE:conditionintervals:start_meansTRUE              5261
## evidence:meansTRUE:conditionQDPs:start_meansTRUE                   5400
## evidence:sd_diff15:conditionHOPs:start_meansTRUE                   5351
## evidence:sd_diff15:conditionintervals:start_meansTRUE              4861
## evidence:sd_diff15:conditionQDPs:start_meansTRUE                   5084
## meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE                  5278
## meansTRUE:sd_diff15:conditionintervals:start_meansTRUE             5433
## meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE                  5612
## evidence:meansTRUE:sd_diff15:conditionHOPs:start_meansTRUE         5599
## evidence:meansTRUE:sd_diff15:conditionintervals:start_meansTRUE    5537
## evidence:meansTRUE:sd_diff15:conditionQDPs:start_meansTRUE         4671
## 
## Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
## is a crude measure of effective sample size, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

Our research questions are about the just-noticable differences (JND) and points of subjective equality (PSE) for this logistic regression model. We derive estimates of these two statistics from the model’s posterior distribution.

```
# get slopes from transformed linear model
slopes_df <- model_df %>%
  group_by(means, sd_diff, condition, trial, start_means) %>%
  data_grid(evidence = c(0, 1)) %>%
  add_fitted_draws(m.decisions, re_formula = NA, scale = "linear") %>%
  compare_levels(.value, by = evidence) %>%
  rename(slope = .value)

# get intercepts from linear model
intercepts_df <- model_df %>%
  group_by(means, sd_diff, condition, trial, start_means) %>%
  data_grid(evidence = 0) %>%
  add_fitted_draws(m.decisions, re_formula = NA, scale = "linear") %>%
  rename(intercept = .value)

# join dataframes for slopes and intercepts, calculate PSE and JND
stats_df <- slopes_df %>%
  full_join(intercepts_df, by = c("means", "sd_diff", "condition", "trial", "start_mean
s", ".draw")) %>%
  mutate(
    pse = -intercept / slope,
    jnd = qlogis(0.75) / slope
  )
```

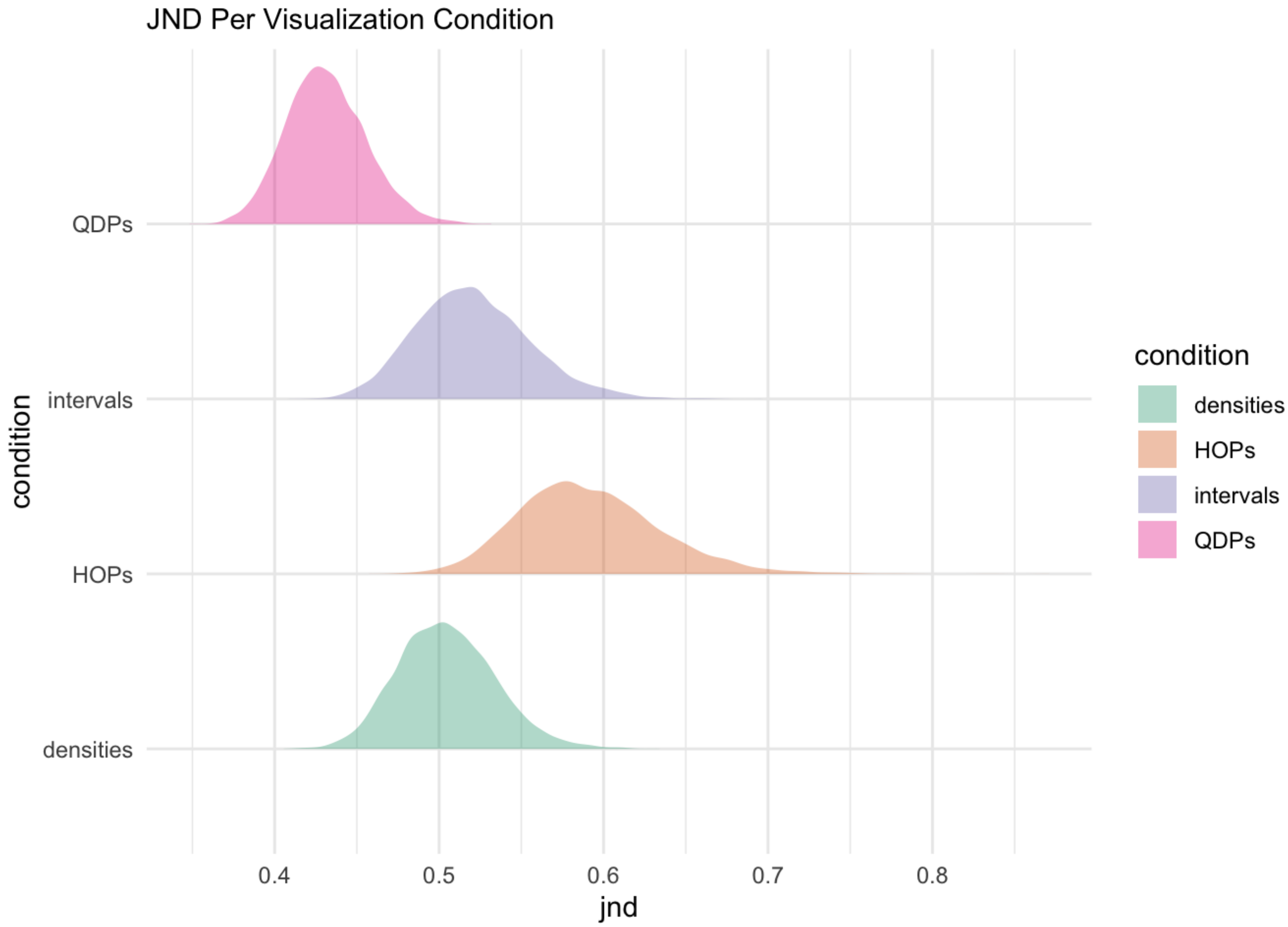
# Just-Noticeable Differences (JNDs)

JNDs describe a chart user's sensitivity to effect size information (i.e., evidence) for the purpose of making decisions.

## Visualization Effects

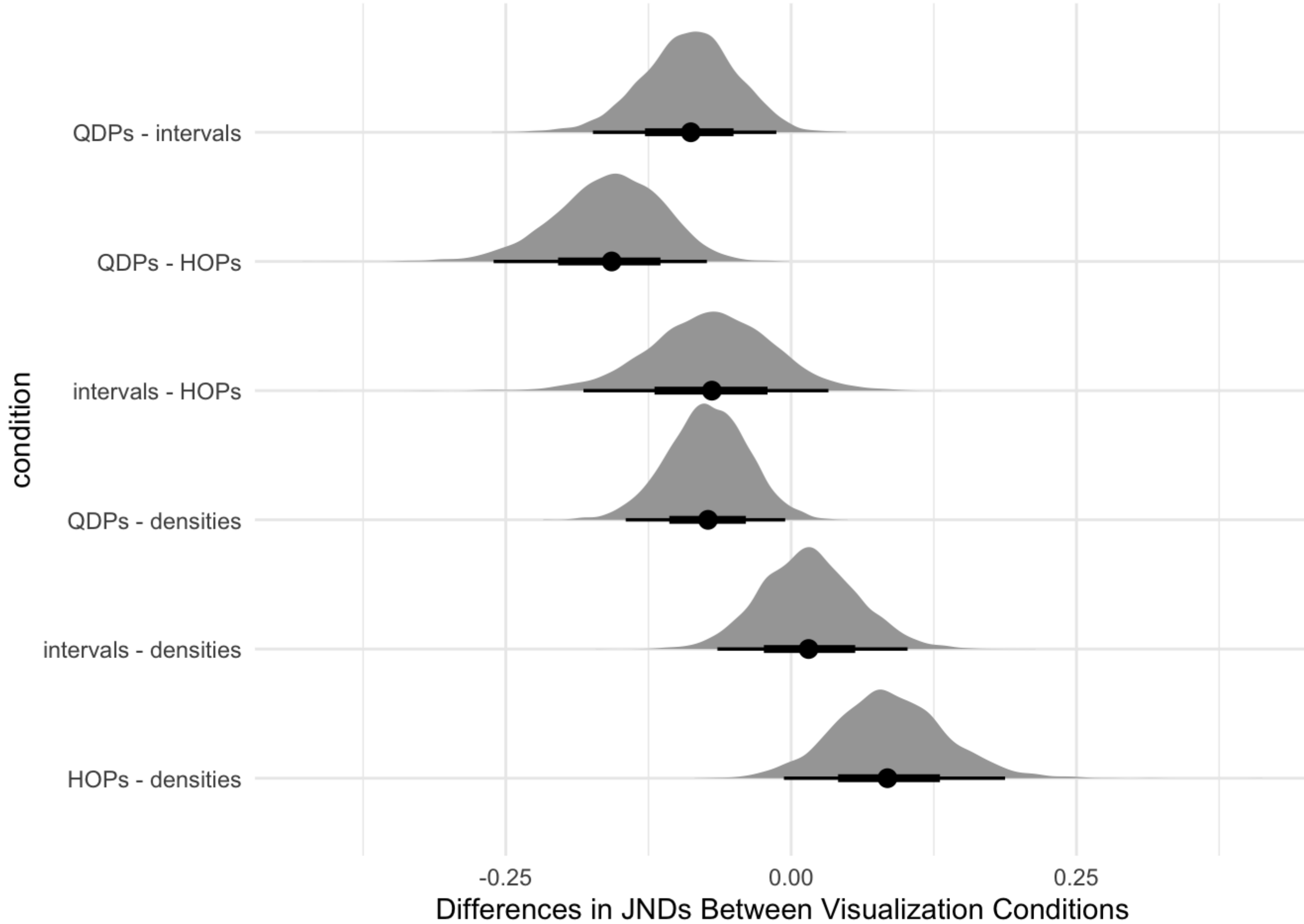
We are interested in estimates of JNDs per visualization, marginalizing across other manipulations.

```
stats_df %>%
  group_by(condition, .draw) %>%          # maginalize out other manipulations
  summarise(jnd = weighted.mean(jnd)) %>%
  ggplot(aes(x = jnd, y = condition, fill = condition)) +
  stat_slabh(alpha = 0.35) +
  scale_fill_brewer(type = "qual", palette = 2) +
  labs(subtitle = "JND Per Visualization Condition") +
  theme_minimal()
```




Let’s look at contrasts between visualization conditions for visual reliability tests.

```
stats_df %>%
  group_by(condition, .draw) %>%          # maginalize out other manipulations
  summarise(jnd = weighted.mean(jnd)) %>%
  compare_levels(jnd, by = condition) %>%
  # compare_levels(jnd, by = condition, comparison = list(c("QDPs", "intervals"), c("QDP
s", "HOPs"), c("QDPs", "densities"), c("densities", "HOPs"), c("intervals", "HOPs"))) %
>%
  ggplot(aes(x = jnd, y = condition)) +
  stat_halfeyeh() +
  labs(x = "Differences in JNDs Between Visualization Conditions") +
  theme_minimal()
```

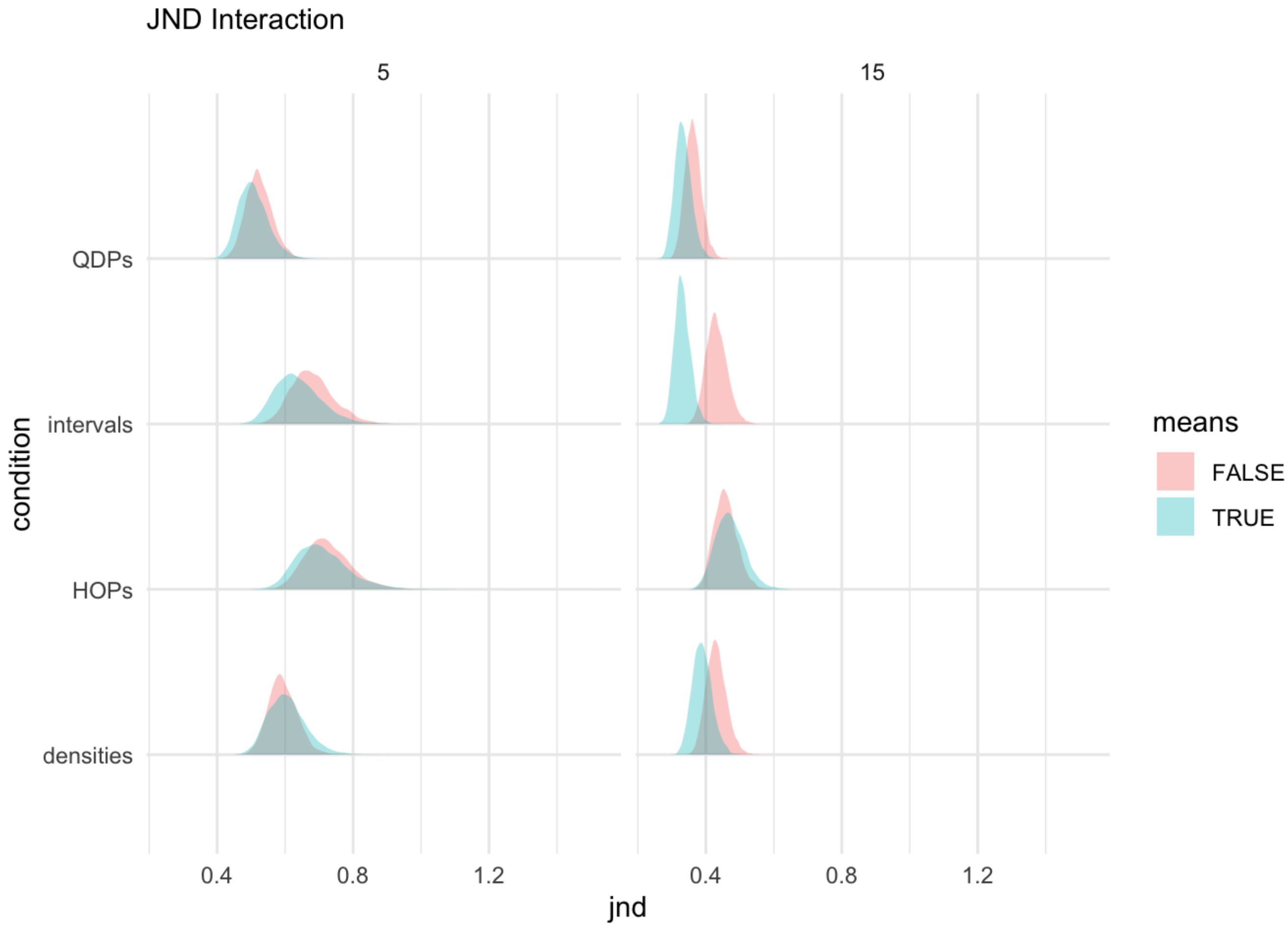



It looks like users are most sensitive to evidence (i.e., JNDs are smaller) in the quantile dotplots condition and are least sensitive with HOPs.

## Interaction Effects

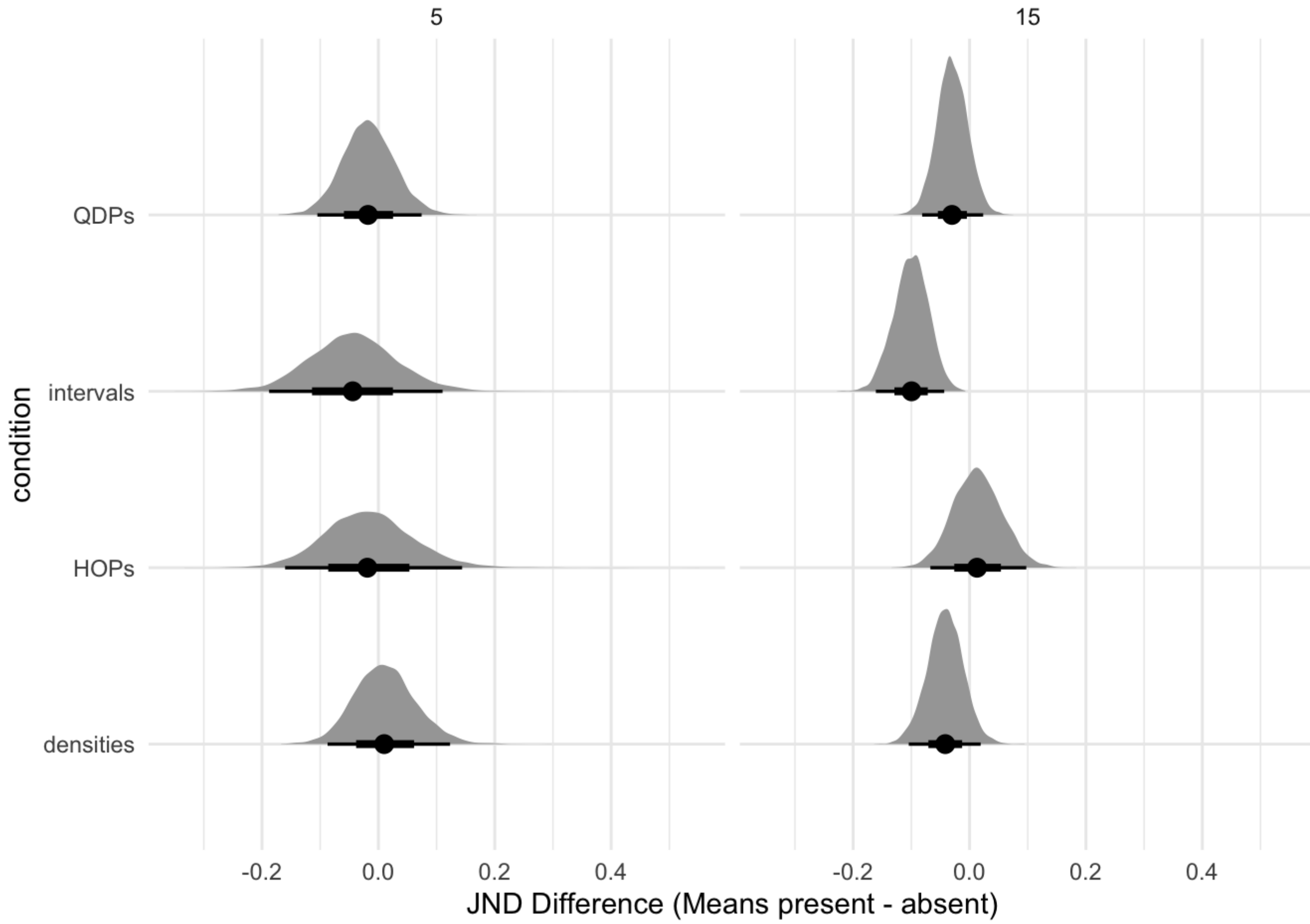
Since we are interested in the way that extinsic means impact the perception of effect size at difference levels of uncertainty, we also look at how this effect manifests in JNDs.

```
stats_df %>%
  group_by(means, sd_diff, condition, .draw) %>%          # maginalize out other manipul
ations
  summarise(jnd = weighted.mean(jnd)) %>%
  ggplot(aes(x = jnd, y = condition, group = means, fill = means)) +
  stat_slabh(alpha = 0.35) +
  labs(subtitle = "JND Interaction") +
  theme_minimal() +
  facet_grid(. ~ sd_diff)
```




Let's look at contrasts for the impact of the mean.

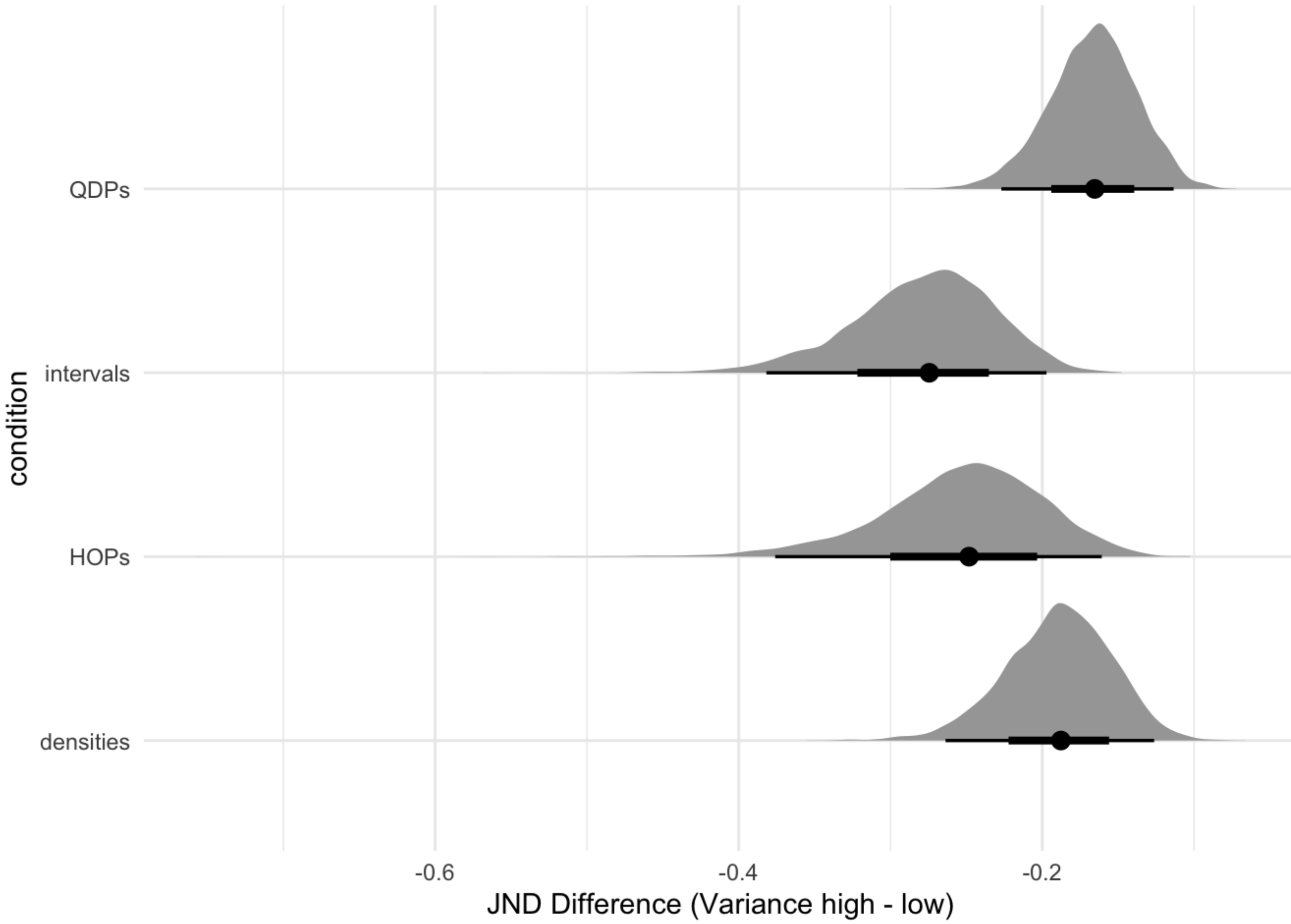
```
stats_df %>%
  group_by(means, sd_diff, condition, .draw) %>%          # maginalize out other manipul
ations
  summarise(jnd = weighted.mean(jnd)) %>%
  compare_levels(jnd, by = means) %>%
  ggplot(aes(x = jnd, y = condition)) +
  stat_halfeyeh() +
  labs(x = "JND Difference (Means present - absent)") +
  theme_minimal() +
  facet_grid(. ~ sd_diff)
```




Extrinsic means seem to improve sensitivity for intervals at high uncertainty.

Now, let’s look at constrasts for the impact of the level of uncertainty per condition.

```
stats_df %>%
  group_by(sd_diff, condition, .draw) %>%           # maginalize out other manipulations
 (including means present/absent)
  summarise(jnd = weighted.mean(jnd)) %>%
  compare_levels(jnd, by = sd_diff) %>%
  ggplot(aes(x = jnd, y = condition)) +
  stat_halfeyeh() +
  labs(x = "JND Difference (Variance high - low)") +
  theme_minimal()
```




Users seem to be consistently more sensitive to evidence (smaller JNDs) when uncertainty is high. This might be because charts in the high uncertainty condition use more of the space on a chart to convey effect size compared to the low uncertainty charts which have a lot of white space such that smaller visual differences convey the same effect size.

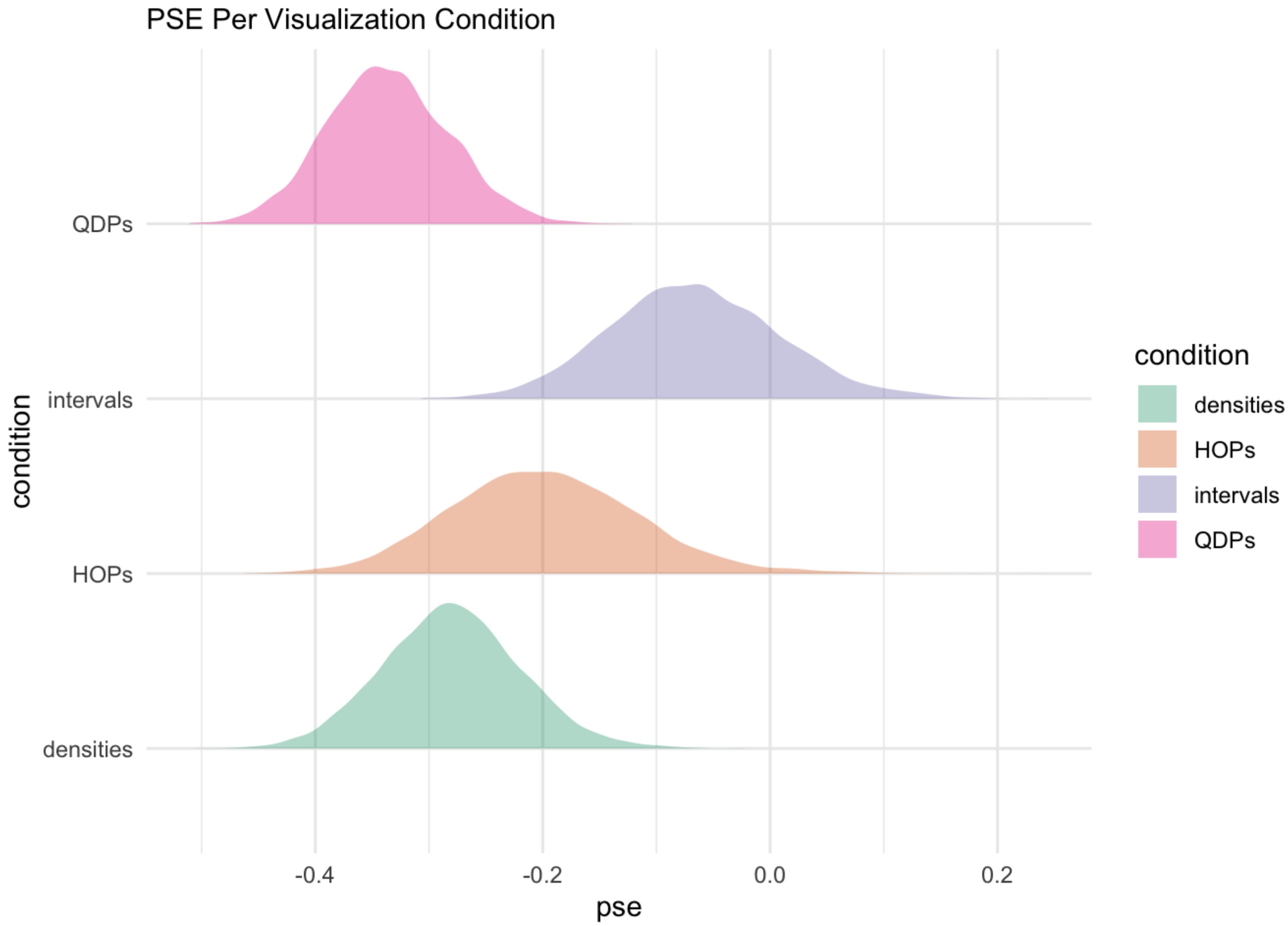
# Points of Subjective Equality (PSE)

PSE describe a chart user's bias toward or against intervening compared to utility optimal decision criterion on the evidence scale (i.e., a proxy for effect size).

## Visualization Effects

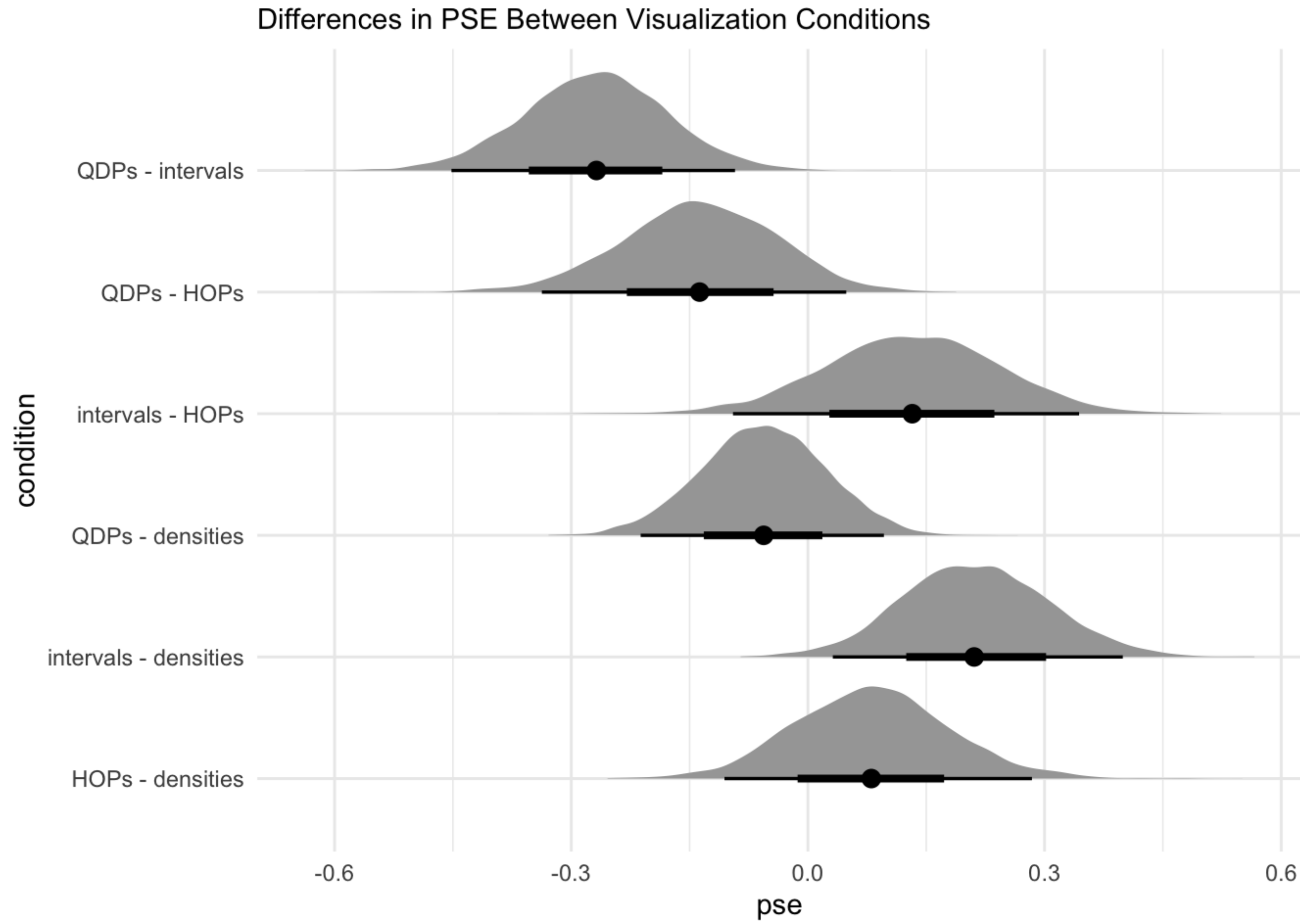
We are interested in estimates of PSE per visualization, marginalizing across other manipulations.

```
stats_df %>%
  group_by(condition, .draw) %>%          # maginalize out other manipulations
  summarise(pse = weighted.mean(pse)) %>%
  ggplot(aes(x = pse, y = condition, fill = condition)) +
  stat_slabh(alpha = 0.35) +
  scale_fill_brewer(type = "qual", palette = 2) +
  labs(subtitle = "PSE Per Visualization Condition") +
  theme_minimal()
```




Let’s look at contrasts between visualization conditions for visual reliability tests.

```
stats_df %>%
  group_by(condition, .draw) %>%          # maginalize out other manipulations
  summarise(pse = weighted.mean(pse)) %>%
  compare_levels(pse, by = condition) %>%
  ggplot(aes(x = pse, y = condition)) +
  stat_halfeyeh() +
  labs(subtitle = "Differences in PSE Between Visualization Conditions") +
  theme_minimal()
```

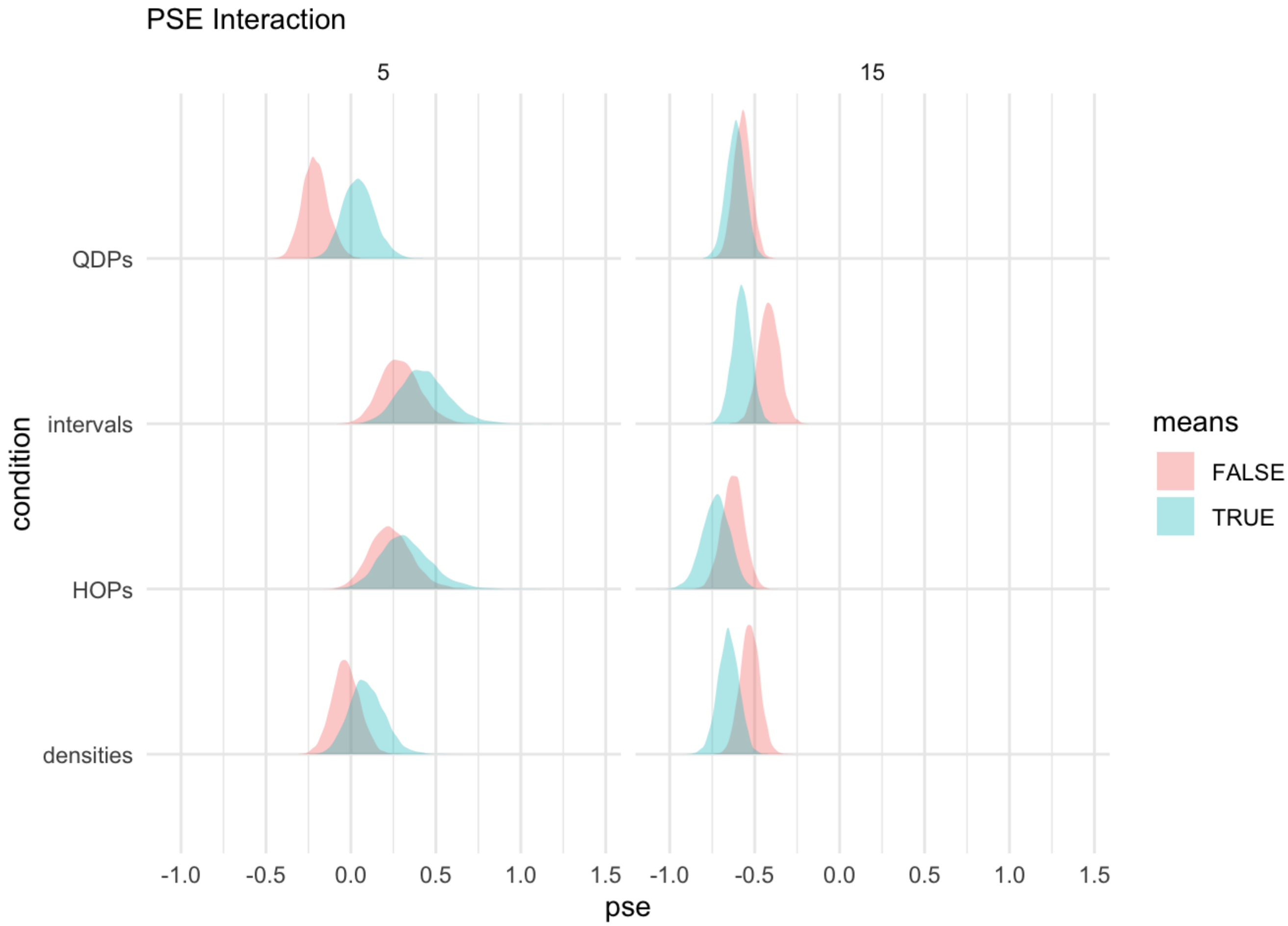



It looks like the point of subjective equality is least biased with intervals, with increasing bias toward intervening (i.e., negative PSE) with HOPs, densities, and quantile dotplots, respectively.

## Interaction Effects

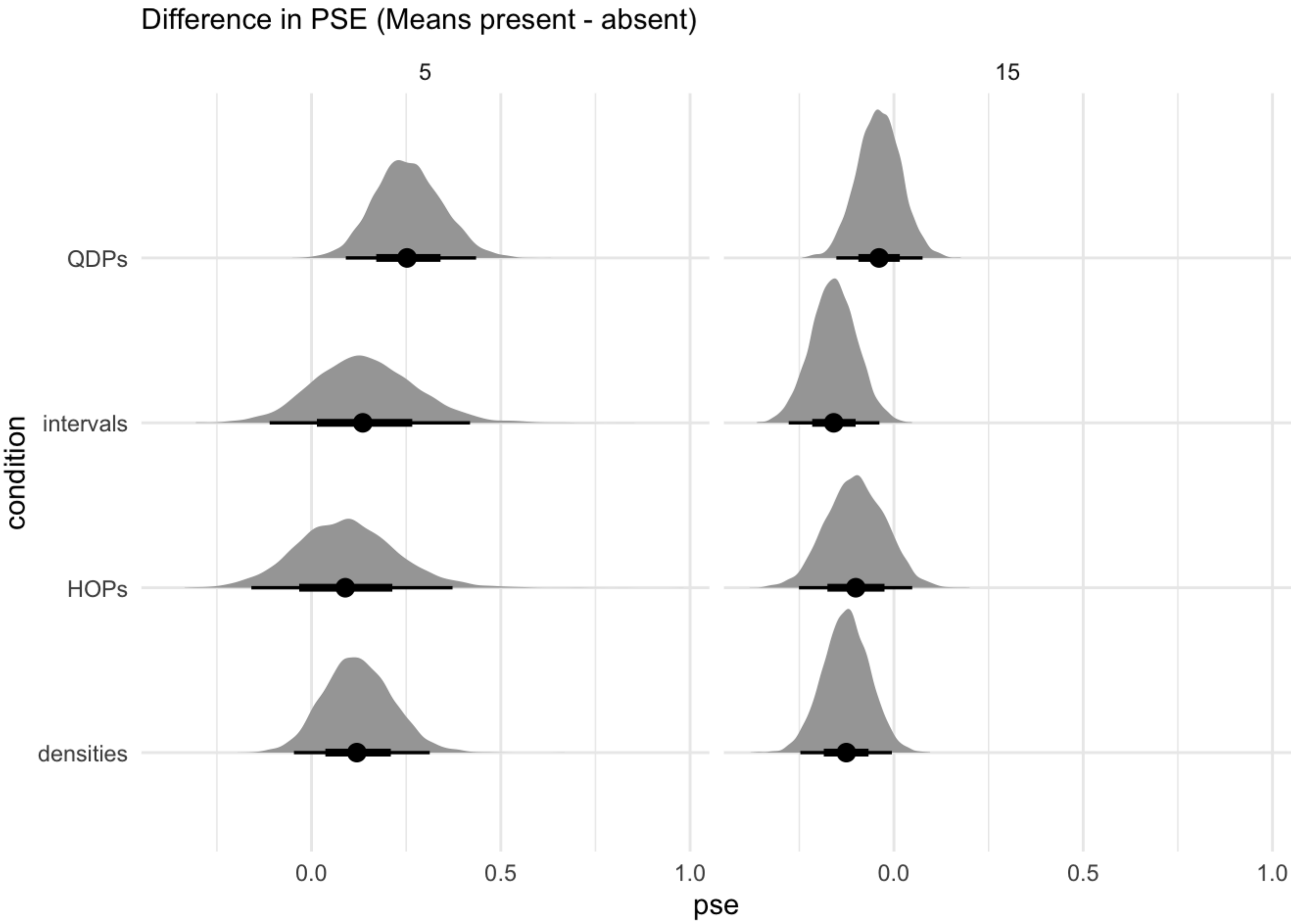
Let's also take a look at the interaction effect of extinsic means at difference levels of uncertainty on PSE.

```
stats_df %>%
  group_by(means, sd_diff, condition, .draw) %>%          # maginalize out other manipul
ations
  summarise(pse = weighted.mean(pse)) %>%
  ggplot(aes(x = pse, y = condition, group = means, fill = means)) +
  stat_slabh(alpha = 0.35) +
  labs(subtitle = "PSE Interaction") +
  theme_minimal() +
  facet_grid(. ~ sd_diff)
```




Let's look at contrasts for the impact of the mean.

```
stats_df %>%
  group_by(means, sd_diff, condition, .draw) %>%          # maginalize out other manipul
ations
  summarise(pse = weighted.mean(pse)) %>%
  compare_levels(pse, by = means) %>%
  ggplot(aes(x = pse, y = condition)) +
  stat_halfeyeh() +
  labs(subtitle = "Difference in PSE (Means present - absent)") +
  theme_minimal() +
  facet_grid(. ~ sd_diff)
```

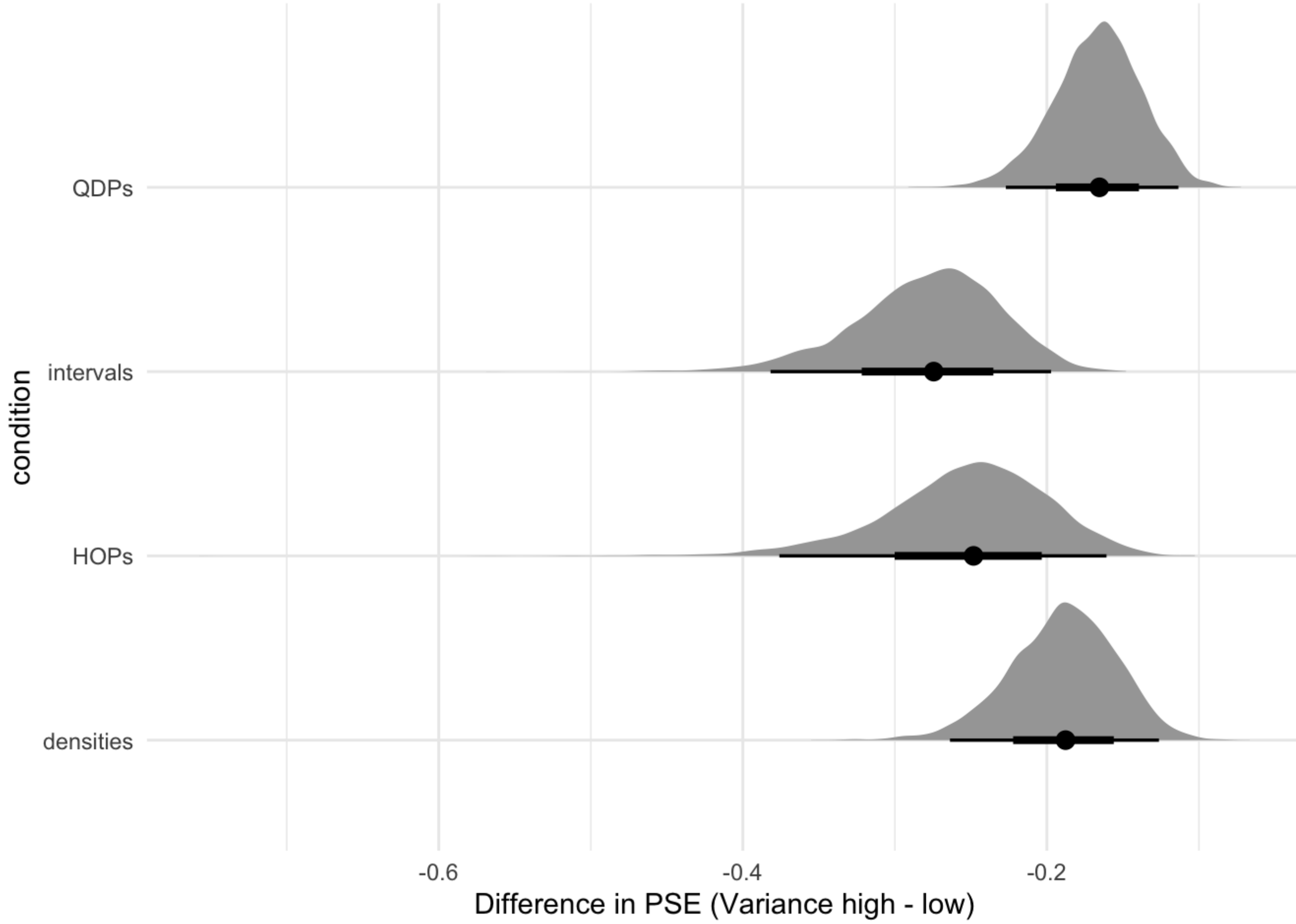



In terms of the direction of effect, extrinsic means seem to consistently bias PSE toward intervention at high variance and away from intervention at low variance. This has the impact of exacerbating biases in decisions compared to when means are absent (with the exception of quantile dotplots at low variance). However, this effect only appears to be reliable for intervals and densities at high variance. We suspect that more data would shrink the uncertainty in these estimates revealing this to be persistent trend, but we can only speculate about whether extrinsic means might lead to less utility optimal decisions overall.

Quantile dotplots are slightly different than other charts in that they are the only uncertainty encoding that consistently biases users toward intervention, regardless of the level of uncertainty. This means that the positive impact on PSE induced by adding extrinsic means at low uncertainty is debiasing for quantile doplots, which is the only case where we can say that adding means is reliably helpful for decision-making.

Now, let’s look at constrasts for the impact of the level of uncertainty per condition.

```
stats_df %>%
  group_by(sd_diff, condition, .draw) %>%          # maginalize out other manipulations
 (including means present/absent)
  summarise(pse = weighted.mean(jnd)) %>%
  compare_levels(pse, by = sd_diff) %>%
  ggplot(aes(x = pse, y = condition)) +
  stat_halfeyeh() +
  labs(x = "Difference in PSE (Variance high - low)") +
  theme_minimal()
```



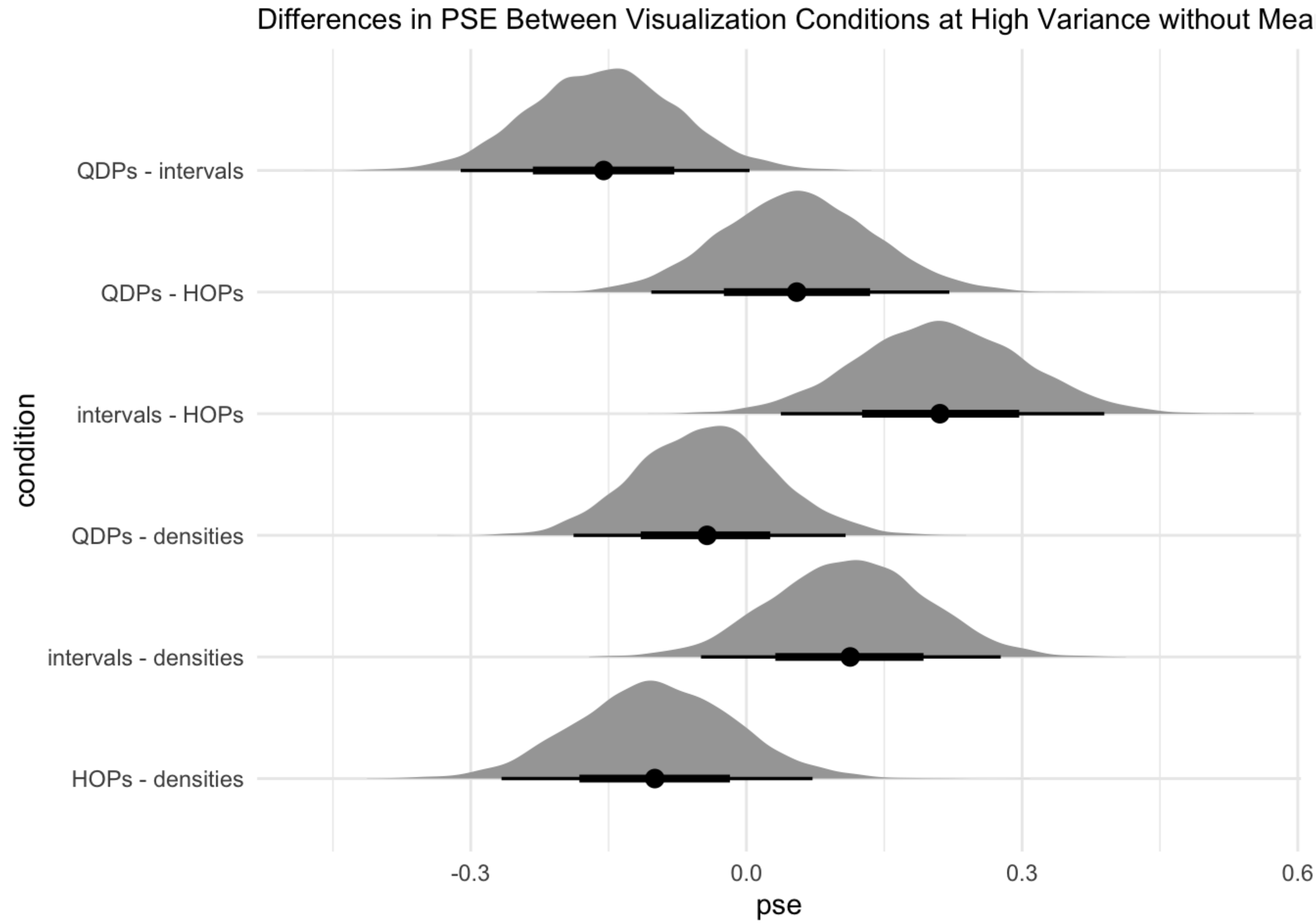

People seem to intervene more than they should when uncertainty is high in all visualization conditions.

## Looking at High vs Low Uncertainty Separately

It seems like the patterns of results for PSE at high vs low uncertainty are different enough that we might want to make different design recommendations depending on the level of uncertainty shown in charts.

Let's start by looking at contrasts between visualization conditions at high uncertainty. Since means seem to bias performance toward intervention at high uncertainty (if they have a reliably effect), we'll look specifically at the effectiveness of visualization conditions **without means**.

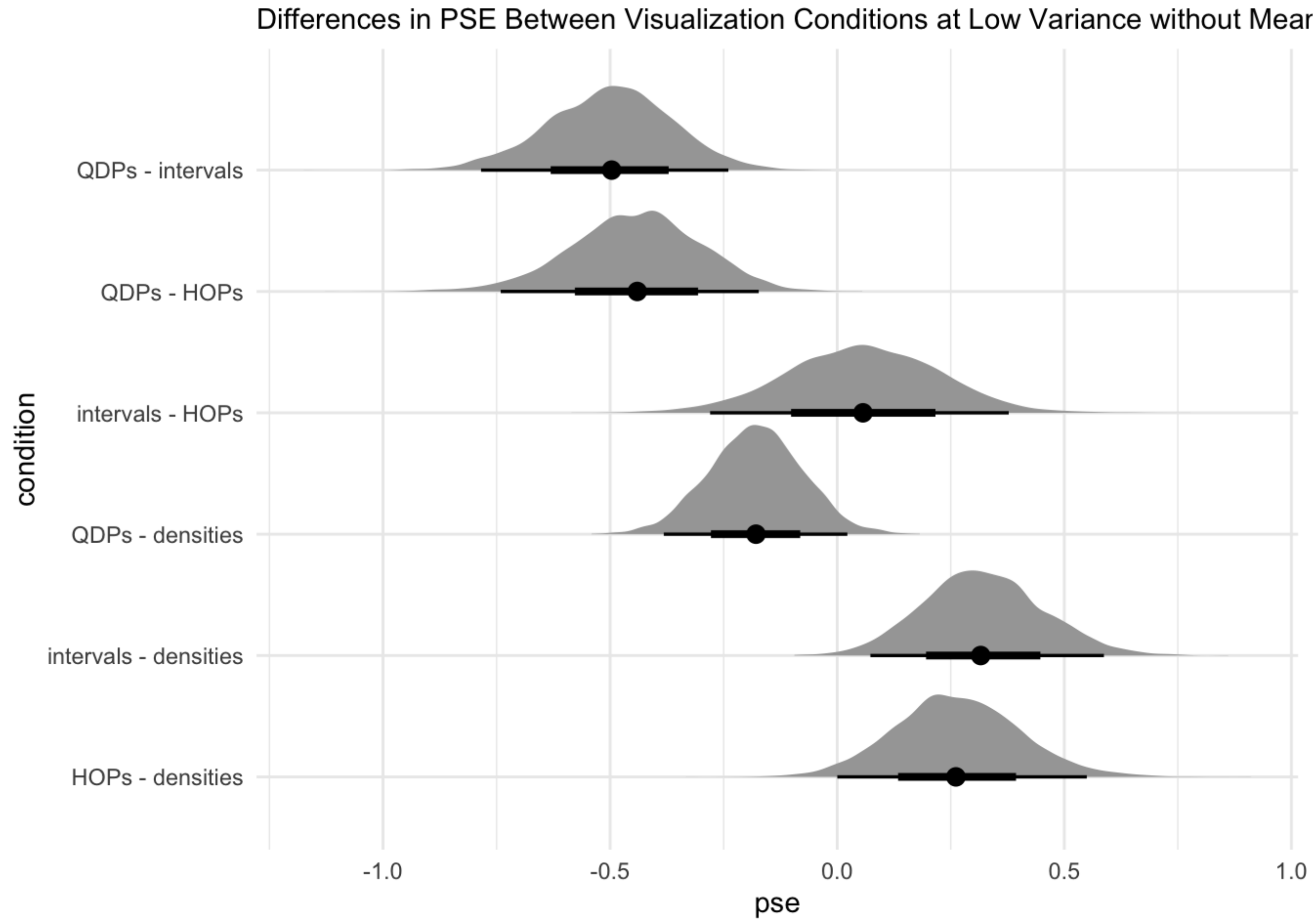
```
stats_df %>%
  group_by(means, sd_diff, condition, .draw) %>%          # maginalize out other manipul
ations
  summarise(pse = weighted.mean(pse)) %>%
  filter(sd_diff == 15 & !as.logical(means)) %>%
  compare_levels(pse, by = condition) %>%
  ggplot(aes(x = pse, y = condition)) +
  stat_halfeyeh() +
  labs(subtitle = "Differences in PSE Between Visualization Conditions at High Variance
 without Means") +
  theme_minimal()
```




When we compare visualization conditions without means, looking for the least biased distributional encoding at high variance, intervals are less biased than quantile dotplots and HOPs, but are not reliably less biased than densities.

Now, we'll consider contrasts between visualization conditions without means at low uncertainty.

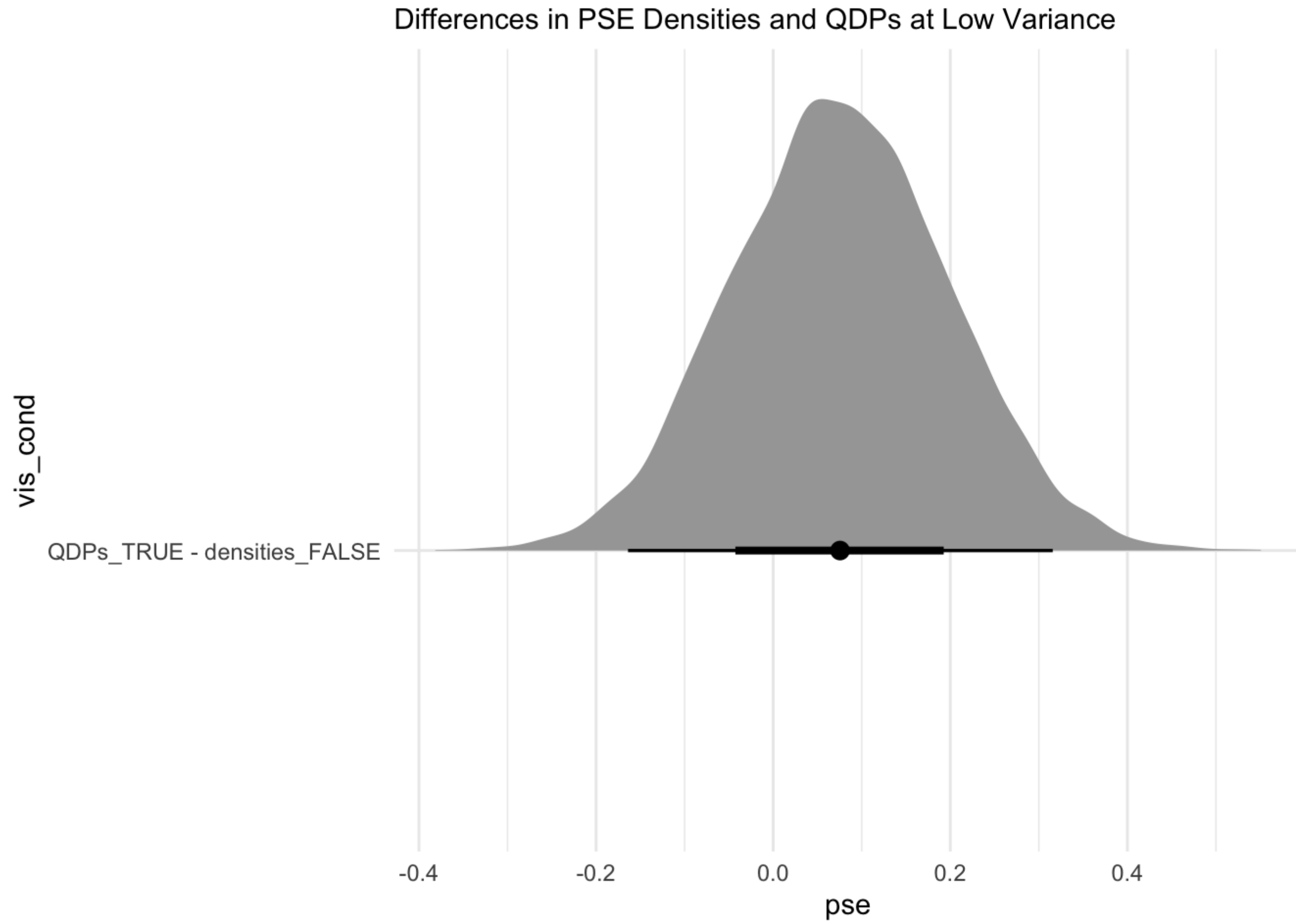
```
stats_df %>%
  group_by(means, sd_diff, condition, .draw) %>%          # maginalize out other manipul
ations
  summarise(pse = weighted.mean(pse)) %>%
  filter(sd_diff == 5 & !as.logical(means)) %>%
  compare_levels(pse, by = condition) %>%
  ggplot(aes(x = pse, y = condition)) +
  stat_halfeyeh() +
  labs(subtitle = "Differences in PSE Between Visualization Conditions at Low Variance w
ithout Means") +
  theme_minimal()
```




Densities without means are the best bet at low uncertainty. They are reliably closer to an unbiased PSE at low uncertainty than intervals, HOPs, and maybe quantile dotplots, suggesting that densities may be preferred when decision aids show distributions with different levels of uncertainty on a common axis.

To clarify, densities without means are no better than quantile dotplots with means in this respect.

```
stats_df %>%
  group_by(means, sd_diff, condition, .draw) %>%          # maginalize out other manipul
ations
  summarise(pse = weighted.mean(pse)) %>%
  filter(sd_diff == 5) %>%
  unite(vis_cond, condition, means) %>%
  filter(vis_cond %in% c("densities_FALSE", "QDPs_TRUE")) %>%
  compare_levels(pse, by = vis_cond) %>%
  ggplot(aes(x = pse, y = vis_cond)) +
  stat_halfeyeh() +
  labs(subtitle = "Differences in PSE Densities and QDPs at Low Variance") +
  theme_minimal()
```




# Does Perceptual Accuracy Lead to Better Decision-Making?

We want to explore how perceptual bias as measured by LLO slopes impacts decision quality as measured by JND and PSE. To do this, we derive point estimates of estimates LLO slope, JND, and PSE for each worker in our data set and combine these statistics into one dataframe.

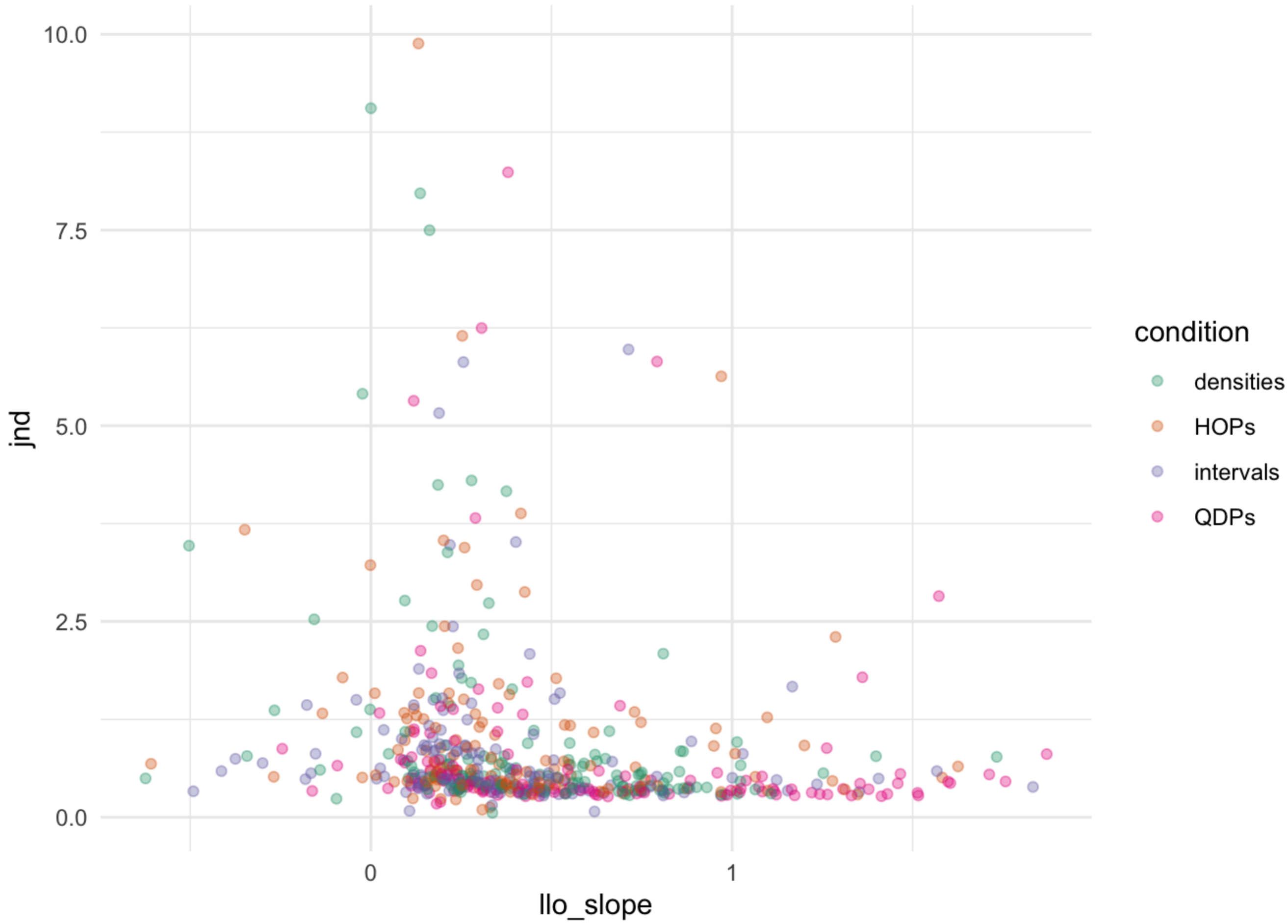
```
# get linear log odds (LLO) slopes per worker
wrkr_llo_slopes_df <- model_df %>%
  group_by(worker_id, means, sd_diff, condition, trial, start_means) %>%
  data_grid(lo_ground_truth = c(0, 1)) %>%          # get fitted draws (in log odds unit
s) only for ground truth of 0 and 1
  add_fitted_draws(m.p_sup, n = 500) %>%
  compare_levels(.value, by = lo_ground_truth) %>%  # calculate the difference between f
its at 1 and 0 (i.e., slope)
  rename(llo_slope = .value) %>%
  group_by(worker_id, condition) %>%                # calculate point estimate of margin
al LLO slope per worker
  summarise(llo_slope = weighted.mean(llo_slope))
```

```
# get logistic regression slopes per worker
wrkr_logistic_slopes_df <- model_df %>%
  group_by(worker_id, means, sd_diff, condition, trial, start_means) %>%
  data_grid(evidence = c(0, 1)) %>%
  add_fitted_draws(m.decisions, scale = "linear", n = 500, seed = 1234) %>%
  compare_levels(.value, by = evidence) %>%
  rename(slope = .value)

# get logistic regression intercepts per worker
wrkr_logistic_intercepts_df <- model_df %>%
  group_by(worker_id ,means, sd_diff, condition, trial, start_means) %>%
  data_grid(evidence = 0) %>%
  add_fitted_draws(m.decisions, scale = "linear", n = 500, seed = 1234) %>%
  rename(intercept = .value)

# join dataframes for logistic slopes and intercepts, calculate PSE and JND
wrkr_logistic_stats_df <- wrkr_logistic_slopes_df %>%
  full_join(wrkr_logistic_intercepts_df, by = c("worker_id", "means", "sd_diff", "condit
ion", "trial", "start_means", ".draw")) %>%
  mutate(
    pse = -intercept / slope,
    jnd = qlogis(0.75) / slope
  ) %>%
  group_by(worker_id, condition) %>%  # calculate point estimate of marginal JND and PSE
per worker
  summarise(
    pse = weighted.mean(pse),
    jnd = weighted.mean(jnd)
  )
```

```
# join the dataframes of summary statistics per worker
wrkr_stats_df <- wrkr_llo_slopes_df %>%
  full_join(wrkr_logistic_stats_df, by = c("worker_id", "condition"))
```

Now let's look at the relationship between LLO slopes and JNDs. This should give a rough indication of how much perceptual accuracy for effect size judgments translates into sensitivity to effect size information for the purpose of decision-making. We've had to filter some workers with extreme JNDs out of this view to get a chart we can read. These are the subset of workers with JND estimates in a reasonable range.

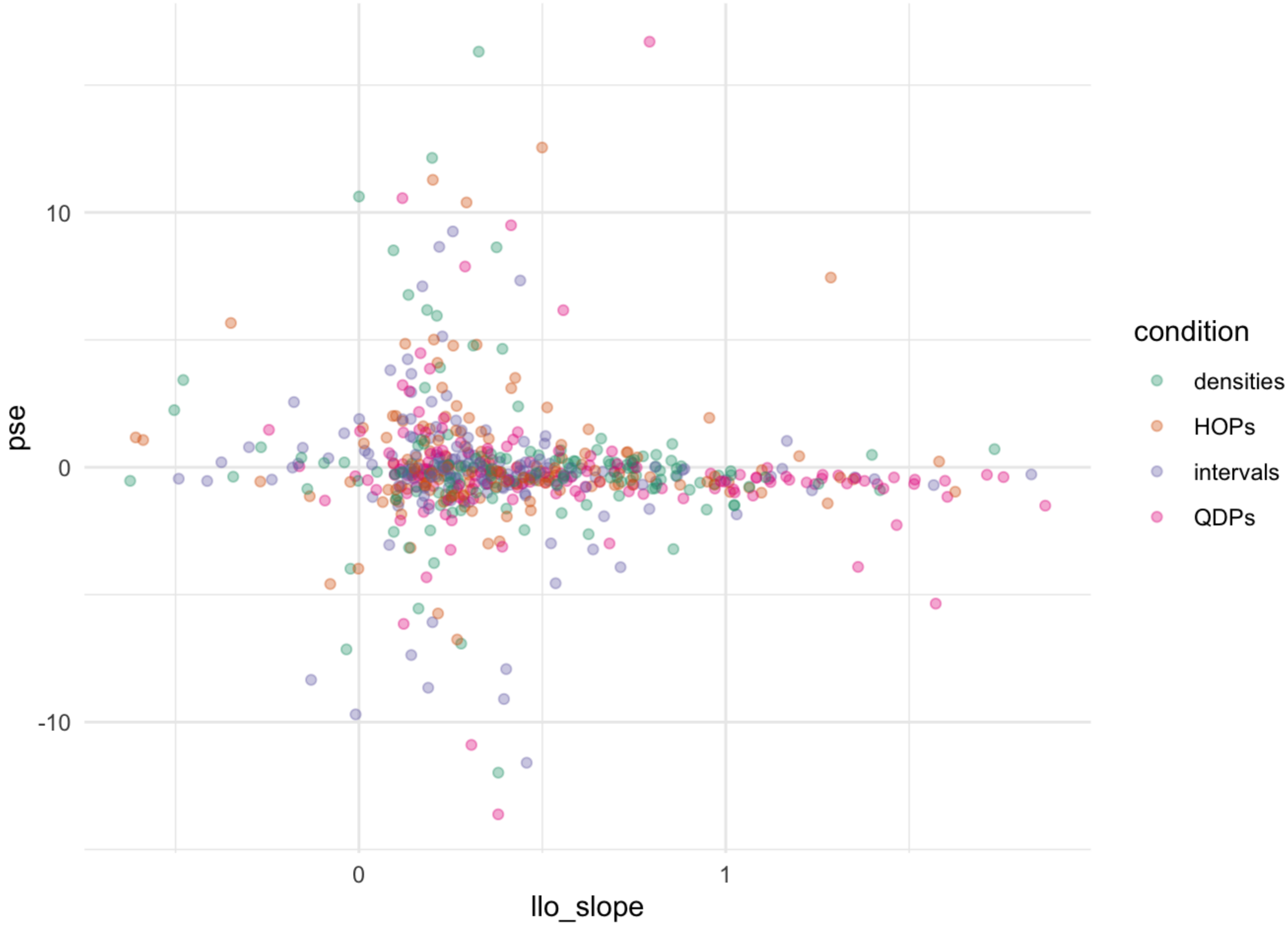
```
wrkr_stats_df %>%
  filter(jnd < 10 & jnd > 0) %>%
  ggplot(aes(x = llo_slope, y = jnd, color = condition)) +
  geom_point(alpha = 0.35) +
  scale_color_brewer(type = "qual", palette = 2) +
  theme_minimal()
```




We can see that while more of the high JNDs (indicating insensitivity) are for workers with low LLO slopes (indicating a tendency to underestimate effect size). However, most workers have relatively small JNDs across the full range of observed LLO slopes, suggesting that perceptual accuracy and sensitivity are only loosely linked with additional factors probably impacting decision-making.

What about the relationship between LLO slopes and PSE. This should give a rough sense of how much perceptual bias translates into bias in decision-making. Again, we've had to filter some workers with extreme PSE out of this view to get a chart we can read.

```
wrkr_stats_df %>%
  filter(pse < 20 & pse > -20) %>%
  ggplot(aes(x = llo_slope, y = pse, color = condition)) +
  geom_point(alpha = 0.35) +
  scale_color_brewer(type = "qual", palette = 2) +
  theme_minimal()
```




Here again we see that the most extreme biases in decision-making (PSE far from 0) tend to correspond with the most extreme tendency to underestimate effect size (slopes less than 1). While biases in decision-making are less common among users with more accurate effect size judgments, the opposite is not the case: There are many users with poor perceptual accuracy who have close to utility optimal decisions. This suggests that perceptual accuracy does not determine a user's ability to make a decision. The implication for the visualization community is that we ought to study and design for magnitude estimation and decisions separately, although not independently.

Part of this mismatch between perceptual performance and decision-making performance may be explained by the fact that out magnitude estimation task was more difficult than the decision task. Some users struggled with the more granular response scale of probability of superiority in pilot testing. By comparison, a binary decision is rather straightforward. We also incentivized the decision task and not the magnitude estimation task. Although we told participants that the best way to maximize their bonus was to answer both questions to the best of their ability, some participants may have sped through the probability of superiority judgments and focused on the decision task. This might explain some of the mismatch between performance on the two tasks.