# gender_distribution

## Maisy Hallam

## 2023-02-10

Import the simulation data. It has three columns: language number (1-1000), gender symmetry score, and set (which simulation run it comes from.)

```r
gender_sims <- list.files(path="data/gender/new/",pattern="*.csv",full.names=TRUE)
data <- do.call(rbind, lapply(gender_sims, function(x) read.csv(x)))
```

Import the real language data - layout same as above.

```r
real_data <- read_csv('data/kinbank_gender.csv')
```

```
## Rows: 1080 Columns: 2
## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr (2): LANGUAGE, GENDER
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```r
data$GENDER <- as.factor(data$GENDER)
```

```r
sym_proportion <- group_by(data,SET) %>%
  summarise(
    mean_symmetrical = mean(GENDER == '[1, 1]')
  )
```

```r
sym_proportion
```

```
## # A tibble: 1,000 x 2
##       SET mean_symmetrical
##     <int>          <dbl>
## 1     1           0.847
## 2     2           0.847
## 3     3           0.846
## 4     4           0.829
## 5     5           0.841
## 6     6           0.833
## 7     7           0.849
## 8     8           0.837
## 9     9           0.855
## 10    10          0.831
## # ... with 990 more rows
```

```
kinbank <- real_data %>%
  summarise(
    mean_symmetrical = mean(GENDER == '[1, 1]')
)
```

```
sym_proportion %>%
  ggplot(aes(mean_symmetrical)) +
  geom_density(fill = 'lightblue', alpha = 1) +
  geom_vline(aes(xintercept = 0.7684), color="navy", linetype="dashed", size=0.5) +
  labs(
    title = 'Proportion of 1,1 gender symmetry scores',
    color = legend) +
  xlab('Proportion of 1,1 symmetry') +
  xlim(0.75,0.9)
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
```
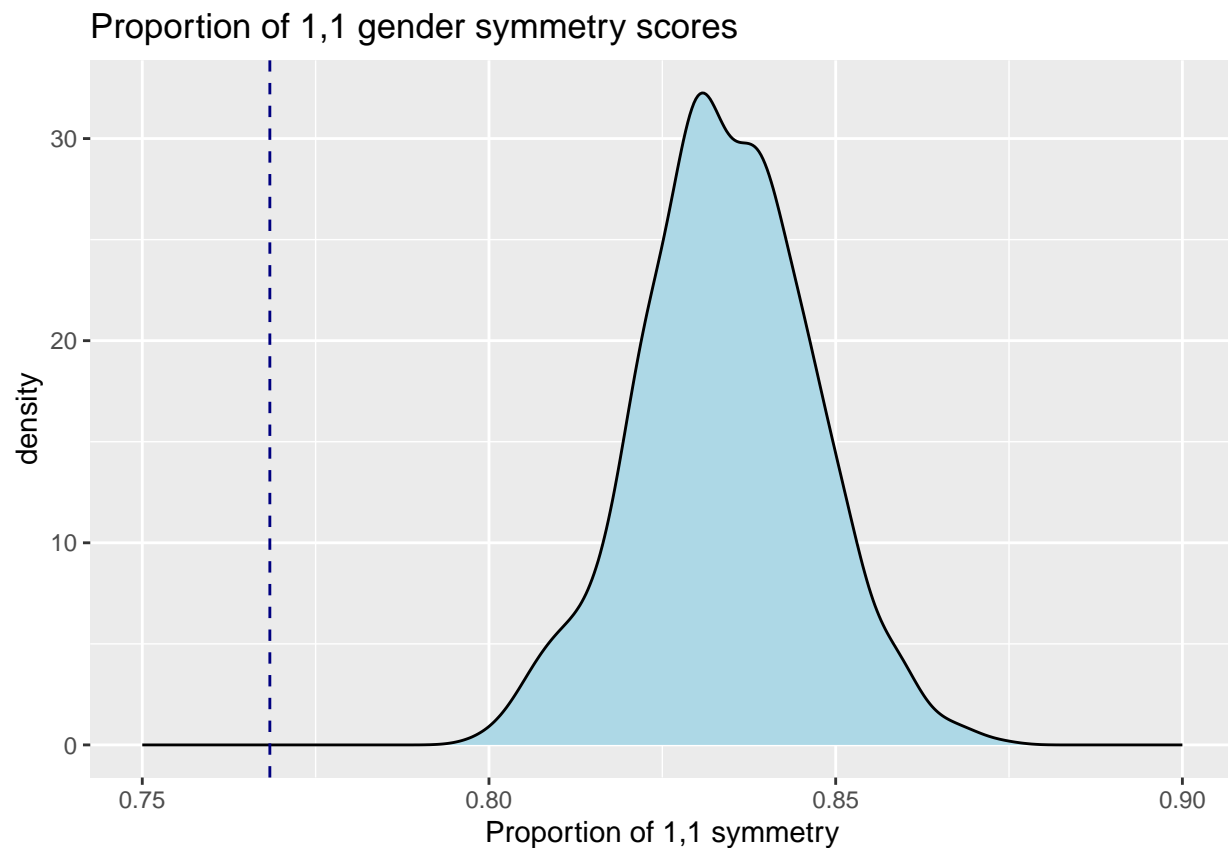


```
ggsave('plots/gender_1,1_density.png')
```

```
## Saving 6.5 x 4.5 in image
```

```
proportion_10 <- group_by(data,SET) %>%
  summarise(
    mean = mean(GENDER == '[1, 0]')
  )

proportion_01 <- group_by(data,SET) %>%
  summarise(
    mean = mean(GENDER == '[0, 1]')
  )

proportion_00 <- group_by(data,SET) %>%
  summarise(
    mean = mean(GENDER == '[0, 0]')
  )

proportion_10
```

```
## # A tibble: 1,000 x 2
##      SET    mean
##    <int>   <dbl>
## 1      1  0.0523
## 2      2  0.0543
## 3      3  0.0568
## 4      4  0.0687
## 5      5  0.0556
## 6      6  0.0624
## 7      7  0.0517
## 8      8  0.0692
## 9      9  0.0620
## 10    10  0.0612
## # ... with 990 more rows
```

```
kinbank_prop_10 <- real_data %>%
  summarise(
    mean = mean(GENDER == '[1, 0]')
)

kinbank_prop_01 <- real_data %>%
  summarise(
    mean = mean(GENDER == '[0, 1]')
)

kinbank_prop_00 <- real_data %>%
  summarise(
    mean = mean(GENDER == '[0, 0]')
)

kinbank_prop_10
```

```
## # A tibble: 1 x 1
##      mean
##     <dbl>
## 1  0.0370
```
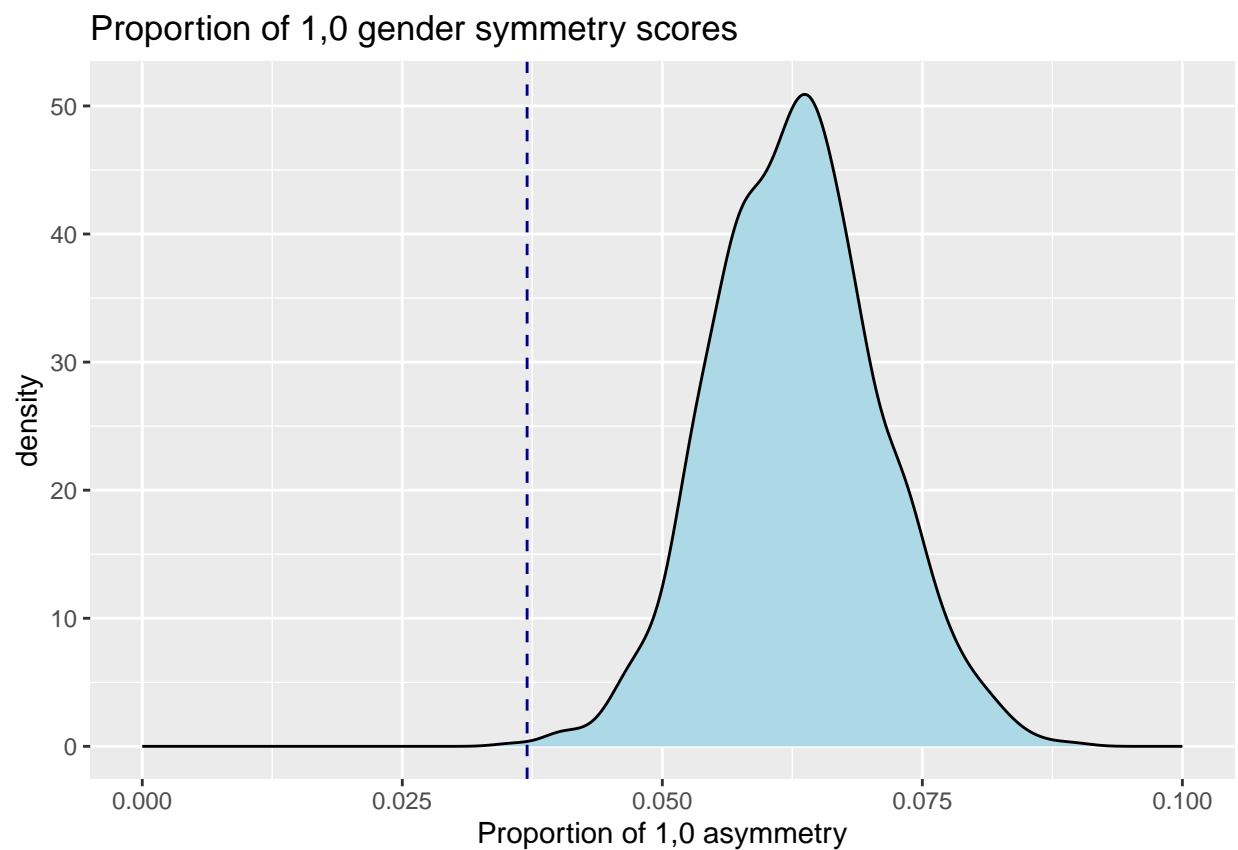
```
kinbank_prop_01
```

```
## # A tibble: 1 x 1
##    mean
##   <dbl>
## 1 0.170
```

```
kinbank_prop_00
```

```
## # A tibble: 1 x 1
##     mean
##    <dbl>
## 1 0.0278
```
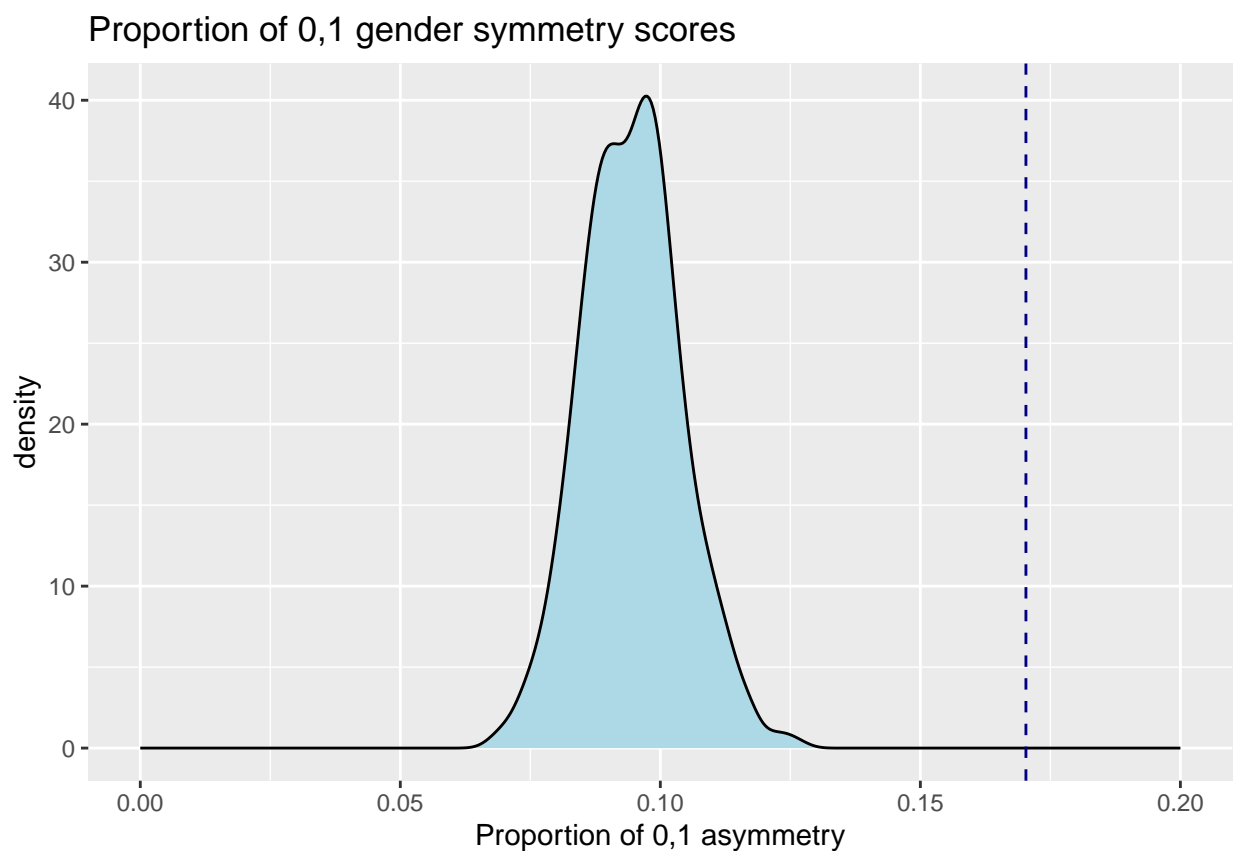
```
proportion_10 %>%
  ggplot(aes(mean)) +
  geom_density(fill = 'lightblue', alpha = 1) +
  geom_vline(aes(xintercept = 0.037), color="navy", linetype="dashed", size=0.5) +
  labs(
    title = 'Proportion of 1,0 gender symmetry scores',
    color = legend) +
  xlab('Proportion of 1,0 asymmetry') +
  xlim(0,0.10)
```

## Proportion of 1,0 gender symmetry scores

```
ggsave('plots/gender_1,0_density.png')
```

```
## Saving 6.5 x 4.5 in image
```

```
proportion_01 %>%
  ggplot(aes(mean)) +
  geom_density(fill = 'lightblue', alpha = 1) +
  geom_vline(aes(xintercept = 0.1703), color="navy", linetype="dashed", size=0.5) +
  labs(
    title = 'Proportion of 0,1 gender symmetry scores',
    color = legend) +
  xlab('Proportion of 0,1 asymmetry') +
  xlim(0,0.2)
```



Proportion of 0,1 gender symmetry scores

```
ggsave('plots/gender_0,1_density.png')
```

```
## Saving 6.5 x 4.5 in image
```

```
proportion_00 %>%
  ggplot(aes(mean)) +
  geom_density(fill = 'lightblue', alpha = 1) +
  geom_vline(aes(xintercept = 0.0278), color="navy", linetype="dashed", size=0.5) +
  labs(
```

```
    title = 'Proportion of 0,0 gender symmetry scores',
    color = legend) +
xlab('Proportion of 0,0 symmetry') +
xlim(0,0.1)
```

## Proportion of 0,0 gender symmetry scores
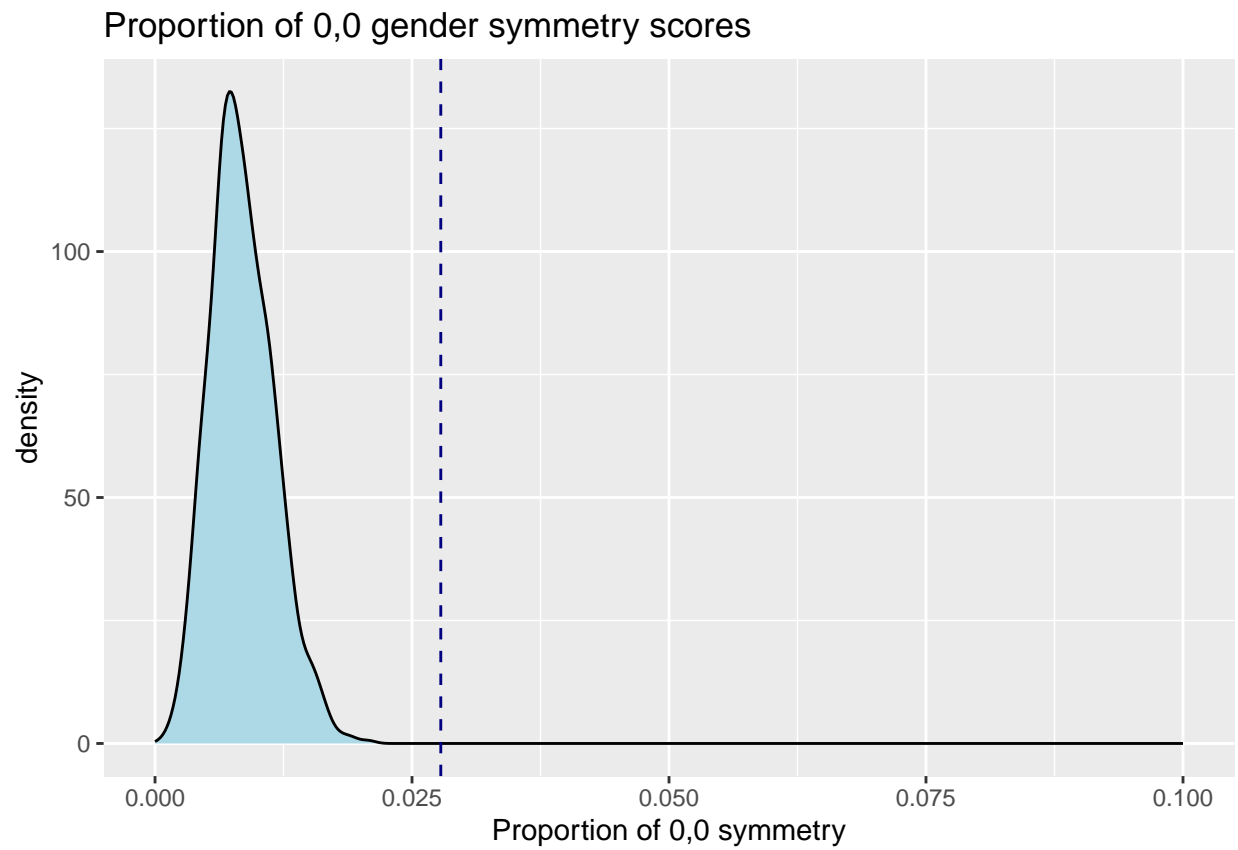


```
ggsave('plots/gender_0,0_density.png')
```

```
## Saving 6.5 x 4.5 in image
```