Gender discrimination demo

Load libraries

```
library(ggplot2)
library(dplyr)
```

Let's set up the discrimination example to see what a random sample would look like

```
# 48 total employees, 24 Male and 24 Female
employees <- c(rep("Male", 24), rep("Female", 24))

# Let's randomly promote 35 employees (independent of gender)
promoted <- sample(employees, 35)

# Compute the proportion of men and women promoted and not promoted
promoted_men <- sum(promoted == "Male")
promoted_women <- sum(promoted == "Female")
not_promoted_men <- 24 - promoted_men
not_promoted_women <- 24 - promoted_women

# Make a dataframe to show the results
data.frame(gender = c("Male", "Female", "Total"),
           promoted = c(promoted_men, promoted_women,
                        sum(promoted_men, promoted_women)),
           not_promoted = c(not_promoted_men, not_promoted_women,
                            sum(not_promoted_men, not_promoted_women)),
           total = c(24, 24, 48))
```

```
##   gender promoted not_promoted total
## 1   Male       18            6    24
## 2 Female       17            7    24
## 3  Total       35           13    48
```

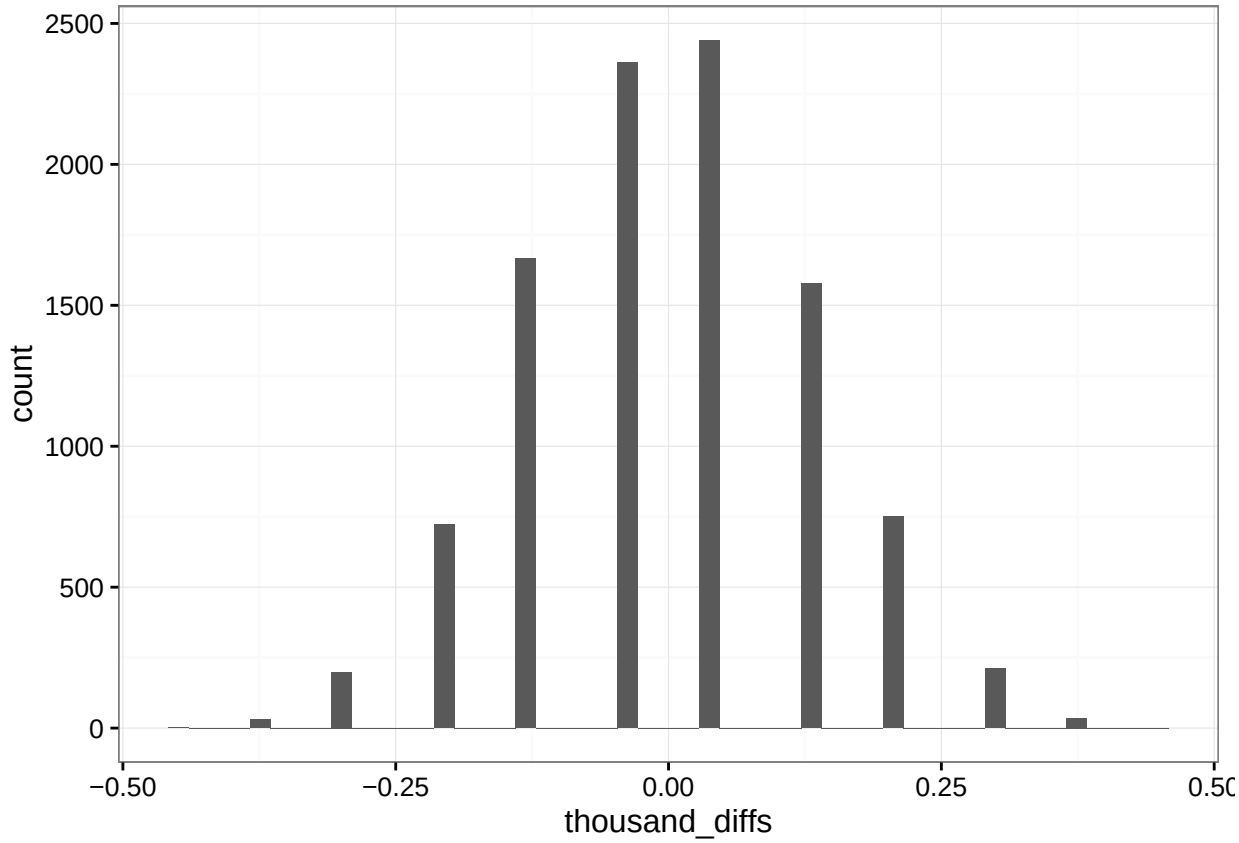
What if we draw a lot of random samples?

```
#Write a function to simulate 1 draw
promoted_diff <- function() {
  # 48 total employees, 24 Male and 24 Female
  employees <- c(rep("Male", 24), rep("Female", 24))

  # Let's randomly promote 35 employees (independent of gender)
  promoted <- sample(employees, 35)

  # Compute the difference in promoted men and women
  promoted_male <- sum(promoted == "Male")/24
  promoted_female <- sum(promoted == "Female")/24

  return(promoted_male - promoted_female)
}

# One function call
promoted_diff()
```

```
## [1] -0.04166667
```

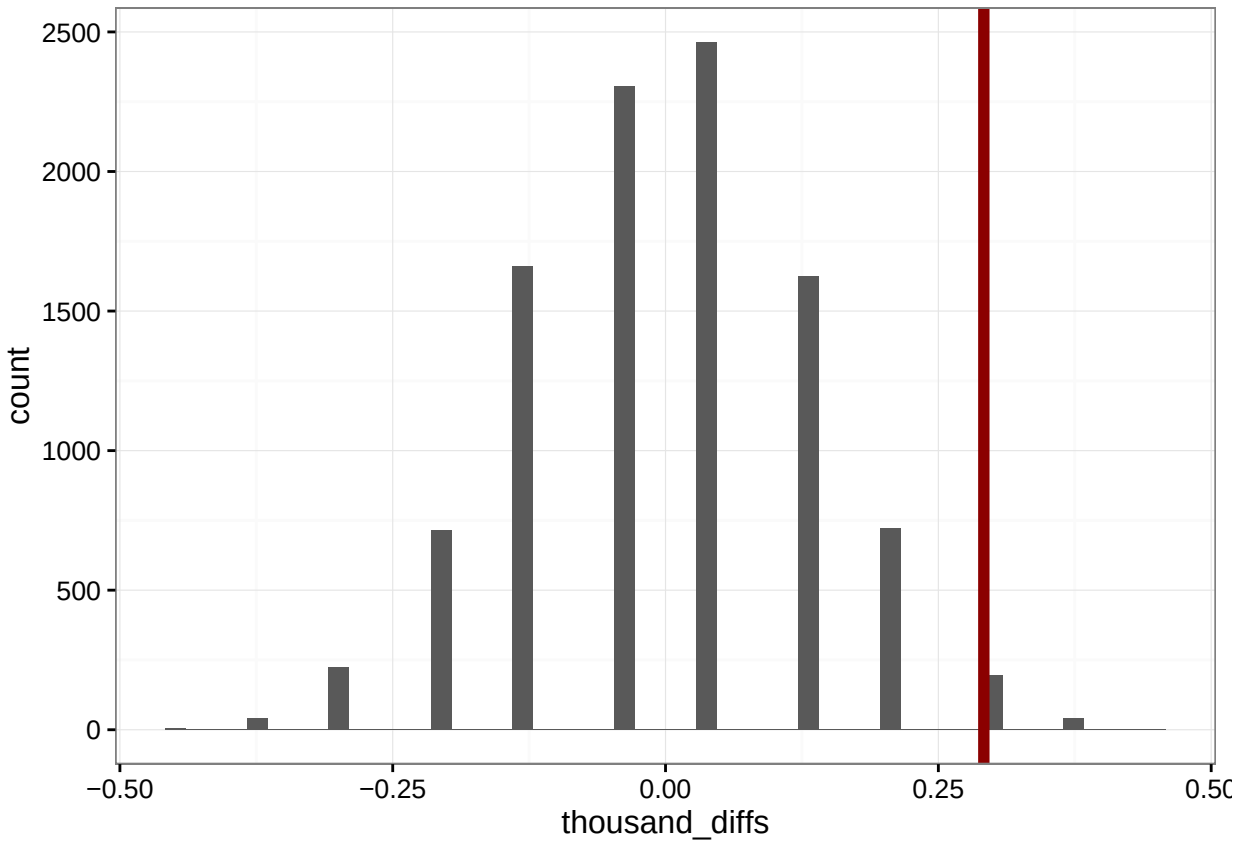
Sample and plot

```
thousand_diffs <- replicate(10000, promoted_diff())

qplot(thousand_diffs, bins = 50) +
  theme_bw()
```



Sample and plot along with our original result

```
#find difference in the actual data
actual_diff <- (21/24) - (14/24)

thousand_diffs <- replicate(10000, promoted_diff())

#add a vertical line to the random samples to show the empirical data
qplot(thousand_diffs, bins = 50) +
  theme_bw() +
  geom_vline(xintercept = actual_diff, color = "darkred", size = 2)
```



Where does our difference lie?

```
# Compute the proportion of samples that the actual difference was greater than
sum(actual_diff > thousand_diffs) / 10000
```

```
## [1] 0.976
```