R Notebook

```
library(stringr)
library(gender)
```

```
## PLEASE NOTE: The method provided by this package must be used cautiously
## and responsibly. Please be sure to see the guidelines and warnings about
## usage in the README or the package documentation.
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

attempt to get a story out of the combined pay gap data for years ending 2017,2018,2019

```
data = readRDS("PayGap_Combined.rds")
```

Separate the name and position and gender of the responsible person

```
data$respFirstName <- word(data$ResponsiblePerson, 1, sep=" ")
data$respLastName <- word(data$ResponsiblePerson, 2, sep=" ")
data$respPos <- str_replace(word(data$ResponsiblePerson, 2, sep="\\("), "\\)", "")

genders <- distinct(gender(data$respFirstName))
data$name <- data$respFirstName
dataWithGender <- data %>% inner_join(genders)
```

```
## Joining, by = "name"
```

What is the ratio of male/female within the responsible person's population

```
gcounts = dataWithGender %>% count(gender)
gsum <- gcounts %>% summarise(sum = min(n)) %>% merge(gcounts)
gsum$prop = gsum$n / gsum$sum
gsum
```

```
##    sum gender     n     prop
## 1 8488 female  8488 1.000000
## 2 8488   male 13516 1.592366
```

Turns out that for years ending on 1017-2019, there were 60% more males responsible for pay gap policy enforcements and reporting than women.