Venn Diagram and Table 1

Aditya Dube

2024-10-30

```r
library(VennDiagram)
```

```
## Loading required package: grid
```

```
## Loading required package: futile.logger
```

```r
library(readxl)
library(kableExtra)
library(dplyr)
```

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

```
## The following object is masked from 'package:kableExtra':
##
##     group_rows
```

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

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

```r
library(gt)
```

```r
# Load the dataset (ignoring metadata rows starting with '!')
file_path <- "Venn Diagram_AF1.xlsx"
f_vs_m <- read_excel(file_path, sheet = "FvsM")
y_vs_o <- read_excel(file_path, sheet = "YvsO")

f_vs_m_probes <- f_vs_m[[1]]   # Probe IDs are in the first column
y_vs_o_probes <- y_vs_o[[1]]   # Probe IDs are in the first column

f_vs_m_set <- unique(f_vs_m_probes)
y_vs_o_set <- unique(y_vs_o_probes)

intersection_size <- length(intersect(f_vs_m_set, y_vs_o_set))
f_vs_m_size <- length(f_vs_m_set)
y_vs_o_size <- length(y_vs_o_set)

venn.plot <- draw.pairwise.venn(
  area1 = f_vs_m_size,
  area2 = y_vs_o_size,
  cross.area = intersection_size,
  category = c("Male vs Female", "Young vs Old"),
```
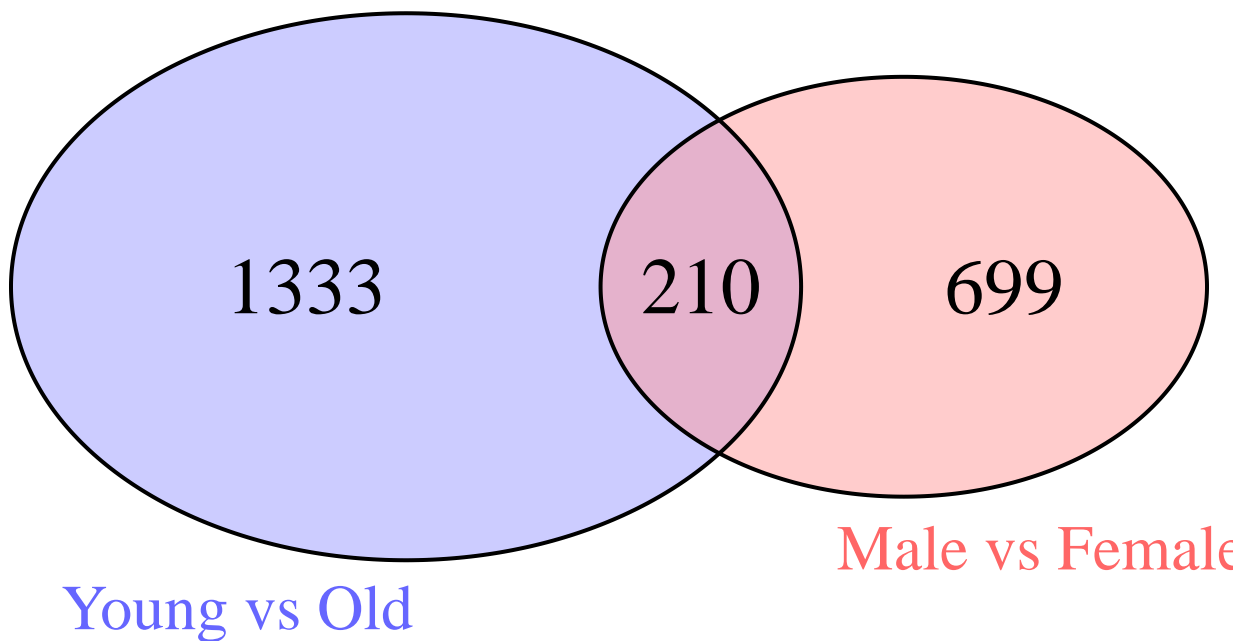
```r
  fill = c("#FF9999", "#9999FF"),   # Softer colors
  alpha = 0.5,                      # Transparency
  lty = "solid",                    # Solid border
  lwd = 2,                          # Thicker border
  cex = 2.5,                        # Font size for numbers
  cat.cex = 2,                      # Font size for labels
  cat.col = c("#FF6666", "#6666FF"), # Custom label colors
  cat.pos = c(-20, 20),             # Label positions
  cat.dist = c(0.05, 0.05),         # Distance between label and circle
  scaled = TRUE                     # Scale the circles better
)
```



```r
png(filename = "venn_diagram.png", width = 800, height = 800)
grid.draw(venn.plot)
dev.off()

## pdf
##   2
# Read the Excel file
file_path <- "Table1.xlsx"
f_vs_m_data <- read_excel(file_path, sheet = "Female vs Male", skip = 16)  # Start from row 17
y_vs_o_data <- read_excel(file_path, sheet = "Young vs Old", skip = 20) # Start from row 21

# List of GO-IDs to match the expected table
selected_go_ids <- c(5575, 8150, 9987, 30154, 10468, 7399, 9653,
                     22008, 48699, 1525, 7420)
```

```r
# Filter for significant GO-IDs (p < 0.05)
female_vs_male_significant <- f_vs_m_data %>%
  filter(`p-value` < 0.05, `GO-ID` %in% selected_go_ids)

young_vs_old_significant <- y_vs_o_data %>%
  filter(`p-value` < 0.05, `GO-ID` %in% selected_go_ids)

female_filtered <- female_vs_male_significant %>%
  select(`GO-ID`, Description, x, X) %>%
  mutate(E = paste0(round((x / X) * 100), "% male"))

young_filtered <- young_vs_old_significant %>%
  select(`GO-ID`, Description, x, X) %>%
  mutate(E = paste0(round((x / X) * 100), "% young"))

female_filtered <- female_filtered %>% distinct(`GO-ID`, .keep_all = TRUE)
young_filtered <- young_filtered %>% distinct(`GO-ID`, .keep_all = TRUE)

final_table <- merge(
  female_filtered, young_filtered, by = c("GO-ID", "Description"),
  suffixes = c(" (F vs. M)", " (Y vs. O)")
)
# Select and reorder columns to match desired output
final_table <- final_table %>%
  select(`GO-ID`, Description, `x (F vs. M)`, `E (F vs. M)`, `x (Y vs. O)`, `E (Y vs. O)`)

# Print the final table
print(final_table)
```

```
##      GO-ID                  Description x (F vs. M) E (F vs. M) x (Y vs. O)
## 1   10468   regulation of gene expression          31     6% male          49
## 2    1525                    angiogenesis          10     2% male          12
## 3   22008                    neurogenesis          18     4% male          30
## 4   30154            cell differentiation          35     7% male          54
## 5   48699             generation of neurons          17     3% male          28
## 6    5575              cellular_component         371    74% male         624
## 7    7399       nervous system development          28     6% male          45
## 8    7420               brain development          13     3% male          18
## 9    8150              biological_process         337    67% male         560
## 10   9653 anatomical structure morphogenesis     37     7% male          59
## 11   9987                  cellular process         176    35% male         336
##     E (Y vs. O)
## 1       6% young
## 2       1% young
## 3       4% young
## 4       6% young
## 5       3% young
## 6      73% young
## 7       5% young
## 8       2% young
## 9      66% young
## 10      7% young
## 11     40% young
```

```r
# Create a styled gt table
gt_table <- final_table %>%
  gt() %>%
  tab_header(
    title = md("**Gene Ontology Comparisons by Gender and Age**")
  ) %>%
  cols_label(
    `GO-ID` = "GO ID",
    Description = "Description",
    `x (F vs. M)` = "x (F vs. M)",
    `E (F vs. M)` = "E (F vs. M)",
    `x (Y vs. O)` = "x (Y vs. O)",
    `E (Y vs. O)` = "E (Y vs. O)"
  ) %>%
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_column_labels(everything())
  ) %>%
  tab_options(
    table.font.size = "small",
    table.width = pct(100)
  ) %>%
  opt_table_font(
    font = "Times New Roman"
  )

# Render the gt table
gt_table
```

Gene Ontology Comparisons by Gender and Age

GO ID	Description	x (F vs. M)	E (F vs. M)	x (Y vs. O)	E (Y vs. O)
10468	regulation of gene expression	31	6% male	49	6% young
1525	angiogenesis	10	2% male	12	1% young
22008	neurogenesis	18	4% male	30	4% young
30154	cell differentiation	35	7% male	54	6% young
48699	generation of neurons	17	3% male	28	3% young
5575	cellular_component	371	74% male	624	73% young
7399	nervous system development	28	6% male	45	5% young
7420	brain development	13	3% male	18	2% young
8150	biological_process	337	67% male	560	66% young
9653	anatomical structure morphogenesis	37	7% male	59	7% young
9987	cellular process	176	35% male	336	40% young