



## Data exploratory analysis

Antonia Chroni [achroni@stjude.org](mailto:achroni@stjude.org) for St. Jude Children's Research Hospital

### Contents

<b>1</b>	<b>Information about this notebook</b>	<b>3</b>
<b>2</b>	<b>Set up</b>	<b>3</b>
<b>3</b>	<b>Directories and paths to file Inputs/Outputs</b>	<b>3</b>
<b>4</b>	<b>Read metadata file</b>	<b>3</b>
4.1	Color palette for plotting . . . . .	3
<b>5</b>	<b>Number of samples by Sex</b>	<b>4</b>
<b>6</b>	<b>Session Info</b>	<b>6</b>
##	The following object is masked _by_ .GlobalEnv:	
##		
##	root_dir	

**PI: Stanislav Zakharenko**

**Project: snRNASeq\_of\_DS\_in\_Df1\_mice**

Task: Training 1 - Demo

Project Lead(s): NA

Department: Developmental Neurobiology

DNB Bioinformatics Core Analysis Team:

**Lead Analyst(s): Antonia Chroni, PhD**

Group Lead: Cody A. Ramirez, PhD

**Contact E-mail:** [achroni@stjude.org](mailto:achroni@stjude.org)

**DNB Bioinformatics Core Pipeline:** Training 1 - Demo

Date started: 10/10/2024

Date completed: 10/10/2024

Report generated: 13:43:37 CDT 10/10/2024

Reviewed by: \_\_\_\_\_ Date: \_\_\_\_\_

## 1 Information about this notebook

This is an exploratory analysis of the data in the project. We are investigating the number of samples per Sex. This is a demo designed for the purposes of the **Best practices** training in the context of the **Automation and Reproducibility in Computational Biology** course.

## 2 Set up

```
suppressPackageStartupMessages({
  library(tidyverse)
})
```

### 3 Directories and paths to file Inputs/Outputs

```
attach(params)
```

```
## The following object is masked _by_ .GlobalEnv:
```

##

```
##      root_dir
```

```
demo_dir <- file.path(root_dir, "courses", "Automation-Reproducibility-compbio", "demo")
analysis_dir <- file.path(demo_dir, "analyses", "data-exploratory-analysis")
```

```
# We will first read in metadata file as we need to define sample_name
```

```
metadata_file <- file.path(metadata_dir, "project_metadata.tsv") # metadata input file
```

```
palette_file <- file.path(demo_dir, "figures", "palettes", "binary_color_palette.tsv")
```

```
# File path to `plots` directory
```

```
plots_dir <- file.path(analysis_dir, "plots")
```

```
if (!dir.exists(plots_dir)) {
```

```
dir.create(plots_dir)}
```

```
figures_plots_dir <- file.path(plots_dir, "figures")
```

```
if (!dir.exists(figures_plots_dir)) {
```

```
dir.create(figures_plots_dir)}
```

```
source(paste0(analysis_dir, "/util/function-create-barplot.R"))
```

```
source(paste0(demo_dir, "/figures/scripts/theme_plot.R"))
```

#### 4 Read metadata file

```
# Read metadata
```

```
df <- read.csv(metadata_file, sep = '\\t', header = TRUE)
```

### 4.1 Color palette for plotting

## # Read color palette

[illegible]

```
# Define and order palette
palette <- palette_df$hex_codes
names(palette) <- palette_df$color_names
```

## 5 Number of samples by Sex

Table 1: Summary of samples per Sex

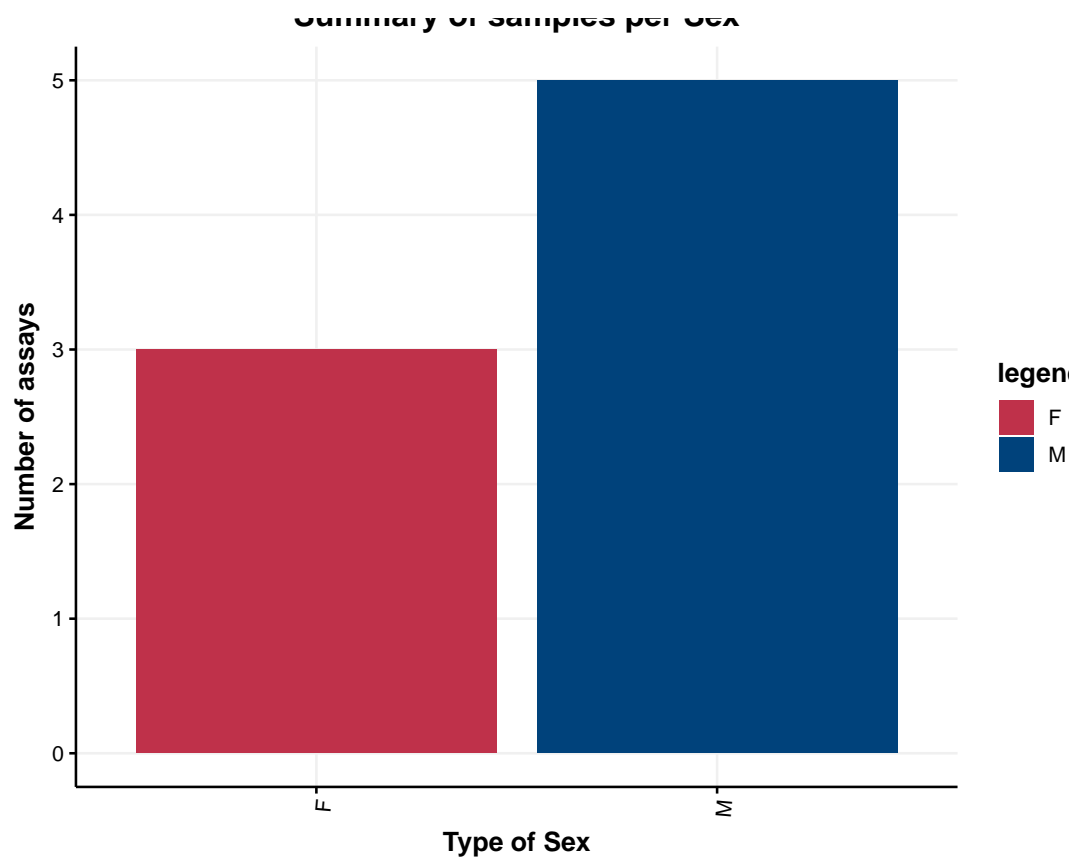
Sex	n
F	3
M	5

```

# Define parameters for function
ylim <- max(tables1$n)

# Run function
fname <- paste0(figures_plots_dir, "/", "samples-per-sex.pdf")
p <- create_barplot(plot_df = tables1,
                    ylim = ylim,
                    x_value = tables1$Sex,
                    use_palette = palette,
                    xtitle = "Type of Sex",
                    title_value = caption_value)

```



```

pdf(file = fname, width = 6, height = 5)
print(p)
dev.off()

```

```

## pdf
## 2

```

## 6 Session Info

```
## R version 4.4.0 (2024-04-24)
## Platform: x86_64-pc-linux-gnu
## Running under: Red Hat Enterprise Linux 8.8 (Ootpa)
##
## Matrix products: default
## BLAS: /usr/lib64/libblas.so.3.8.0
## LAPACK: /usr/lib64/liblapack.so.3.8.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8      LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/Chicago
## tzcode source: system (glibc)
##
## attached base packages:
## [1] grid      stats      graphics  grDevices utils      datasets  methods
## [8] base
##
## other attached packages:
## [1] ggthemes_5.1.0  lubridate_1.9.3 forcats_1.0.0  stringr_1.5.1
## [5] dplyr_1.1.4     purrr_1.0.2     readr_2.1.5    tidyr_1.3.1
## [9] tibble_3.2.1    ggplot2_3.5.1   tidyverse_2.0.0 yaml_2.3.10
##
## loaded via a namespace (and not attached):
## [1] sass_0.4.9      utf8_1.2.4      generics_0.1.3  stringi_1.8.4
## [5] hms_1.1.3       digest_0.6.37   magrittr_2.0.3  evaluate_0.24.0
## [9] timechange_0.3.0 fastmap_1.2.0   jsonlite_1.8.8  tinytex_0.52
## [13] fansi_1.0.6     scales_1.3.0    jquerylib_0.1.4 cli_3.6.3
## [17] rlang_1.1.4     crayon_1.5.3    bit64_4.0.5     munsell_0.5.1
## [21] withr_3.0.1     cachem_1.1.0    tools_4.4.0     parallel_4.4.0
## [25] tzdb_0.4.0      colorspace_2.1-1 vctrs_0.6.5     R6_2.5.1
## [29] mime_0.12       lifecycle_1.0.4 bit_4.0.5        vroom_1.6.5
## [33] pkgconfig_2.0.3 pillar_1.9.0    bslib_0.8.0     gtable_0.3.5
## [37] glue_1.7.0      xfun_0.47       tidyselect_1.2.1 highr_0.11
## [41] knitr_1.48      farver_2.1.2    htmltools_0.5.8.1 rmarkdown_2.28
## [45] labeling_0.4.3  compiler_4.4.0
```