

## Longitudinal data exploratory analysis

Antonia Chroni [achroni@stjude.org](mailto:achroni@stjude.org) for St. Jude Children's Research Hospital BioHackathon Team 1

### 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 . . . . .	4
<b>5</b>	<b>Number of samples with assay information</b>	<b>4</b>
<b>6</b>	<b>Number of samples per brain cancer type and assay</b>	<b>5</b>
6.1	Overall assays . . . . .	5
6.2	Per assay . . . . .	7
6.3	Per disease stage . . . . .	10
<b>7</b>	<b>Number of samples per brain cancer type, longitudinal, assay, and SJUID</b>	<b>11</b>
7.1	Plots . . . . .	19
<b>8</b>	<b>Future directions</b>	<b>20</b>
<b>9</b>	<b>Session Info</b>	<b>21</b>
##	The following object is masked _by_ .GlobalEnv:	
##		
##	root_dir	

**Project: Comprehensive Omics Catalogue for Hartwell**

**St. Jude Children's Research Hospital BioHackathon Team 1**

Date started: 09/04/2024 Date completed: 9/06/2024 Report generated: 18:28:35 CDT  
09/11/2024

## 1 Information about this notebook

This is an exploratory analysis of the data availability in terms of assays in the Comprehensive Omics Catalogue for Hartwell. This is critical for mitigating duplicate sequencing requests and efforts through Hartwell. This notebook aims to showcase: (1) which samples have already been sequenced by Hartwell, and (2) what omics data are available per sample.

For demo purposes, we use dummy data cohort and subset by human brain tumor samples. Finally, we investigate the number of samples per `cancer_type_brain`, Assay and longitudinal.

## 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
```

```
analysis_dir <- file.path(root_dir, "analyses", "data-exploratory-analysis")
```

```
input_dir <- file.path(analysis_dir, "input")
```

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

```
metadata_file <- file.path(input_dir, input_file) # metadata input file
```

```
palette_file <- file.path(root_dir, "figures", "palettes", "assay_color_palette.tsv")
```

```
tumor_palette_file <- file.path(root_dir, "figures", "palettes", "tumor_color_palette.tsv")
```

```
longitudinal_palette_file <- file.path(root_dir, "figures", "palettes", "longitudinal_color_palette.tsv")
```

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

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

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

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

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

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

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

## 4 Read metadata file

We will subset by human brain tumor samples.

```
# Read metadata
```

```
df <- read.csv(metadata_file, stringsAsFactors=FALSE)
```

```
# Number of samples per cancer_type_brain
```

```
assays_number <- length(df$SJUID)
```

```
samples_number <- length(unique(df$SJUID))
```

```

cancer_type_brain_order <- cancer_type_order
longitudinal_order <- disease_stage_order

# Re-order df
f <- assay_order
df <- df %>%
  dplyr::mutate(Assay = factor(Assay),
                Assay = fct_relevel(Assay, f),
                longitudinal = factor(longitudinal),
                longitudinal = fct_relevel(longitudinal, longitudinal_order)) %>%
  arrange(cancer_type_brain, longitudinal, Assay)

```

## 4.1 Color palette for plotting

```

# Read color palette
palette_df <- readr::read_tsv(palette_file, guess_max = 100000, show_col_types = FALSE)

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

# Read color palette for tumor type
tumor_palette_df <- readr::read_tsv(tumor_palette_file, guess_max = 100000, show_col_types = FALSE)

# Define and order palette
tumor_palette <- tumor_palette_df$hex_codes
names(tumor_palette) <- tumor_palette_df$color_names

# Read color palette for longitudinal
longitudinal_palette_df <- readr::read_tsv(longitudinal_palette_file, guess_max = 100000, show_col_types = FALSE)

# Define and order palette
longitudinal_palette <- longitudinal_palette_df$hex_codes
names(longitudinal_palette) <- longitudinal_palette_df$color_names

```

## 5 Number of samples with assay information

Table 1: Summary of samples per assay

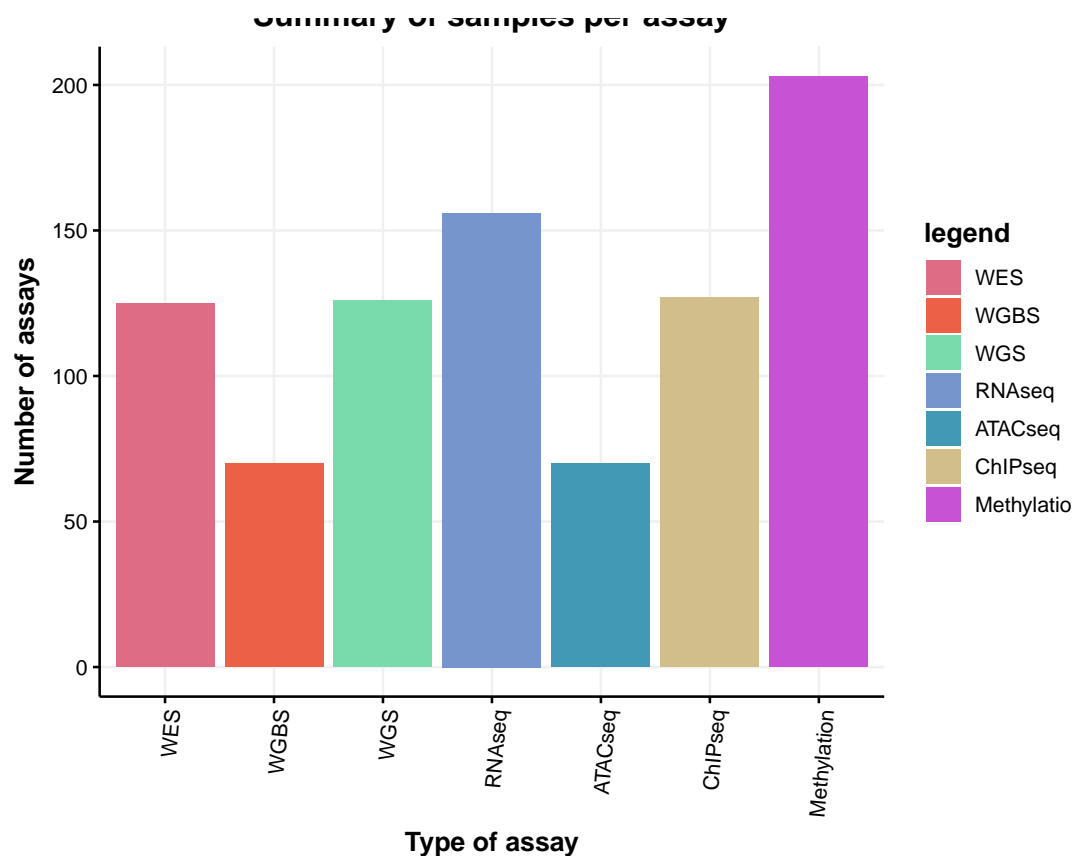
Assay	n
WES	125
WGBS	70
WGS	126
RNAseq	156
ATACseq	70
ChIPseq	127
Methylation	203

```

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

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

```



```

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

```

```

## pdf
## 2

```

## 6 Number of samples per brain cancer type and assay

### 6.1 Overall assays

There are 59 brain tumor samples with 877 assays in total.

Table 2: Summary of samples and assays per brain cancer type

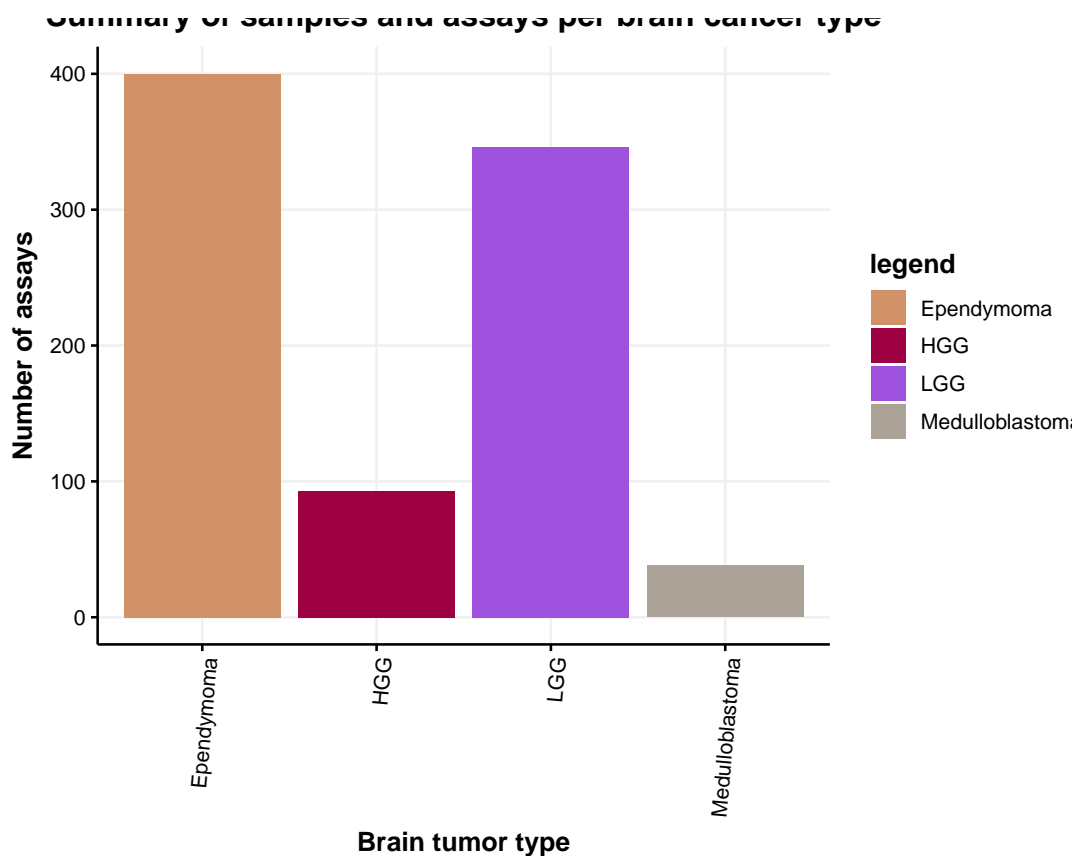
cancer_type_brain	n
Ependymoma	400
HGG	93
LGG	346
Medulloblastoma	38

```

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

# Run function
fname <- paste0(figures_plots_dir, "/", "cancer-type-brain-assay-overall-longitudinal.pdf")
p <- create_barplot(plot_df = tables1,
                    ylim = ylim,
                    x_value = tables1$cancer_type_brain,
                    use_palette = tumor_palette,
                    xtitle = "Brain tumor type",
                    title_value = caption_value)

```



```

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

```

```

## pdf
## 2

```

## 6.2 Per assay

Table 3: Summary of samples and assays per brain cancer type and per assay

cancer_type_brain	WES	WGBS	WGS	RNAseq	ChIPseq	Methylation	ATACseq
Ependymoma	64	70	70	64	68	64	0

cancer_type_brain	WES	WGBS	WGS	RNAseq	ChIPseq	Methylation	ATACseq
HGG	0	0	0	51	0	42	0
LGG	61	0	56	41	59	59	70
Medulloblastoma	0	0	0	0	0	38	0



```

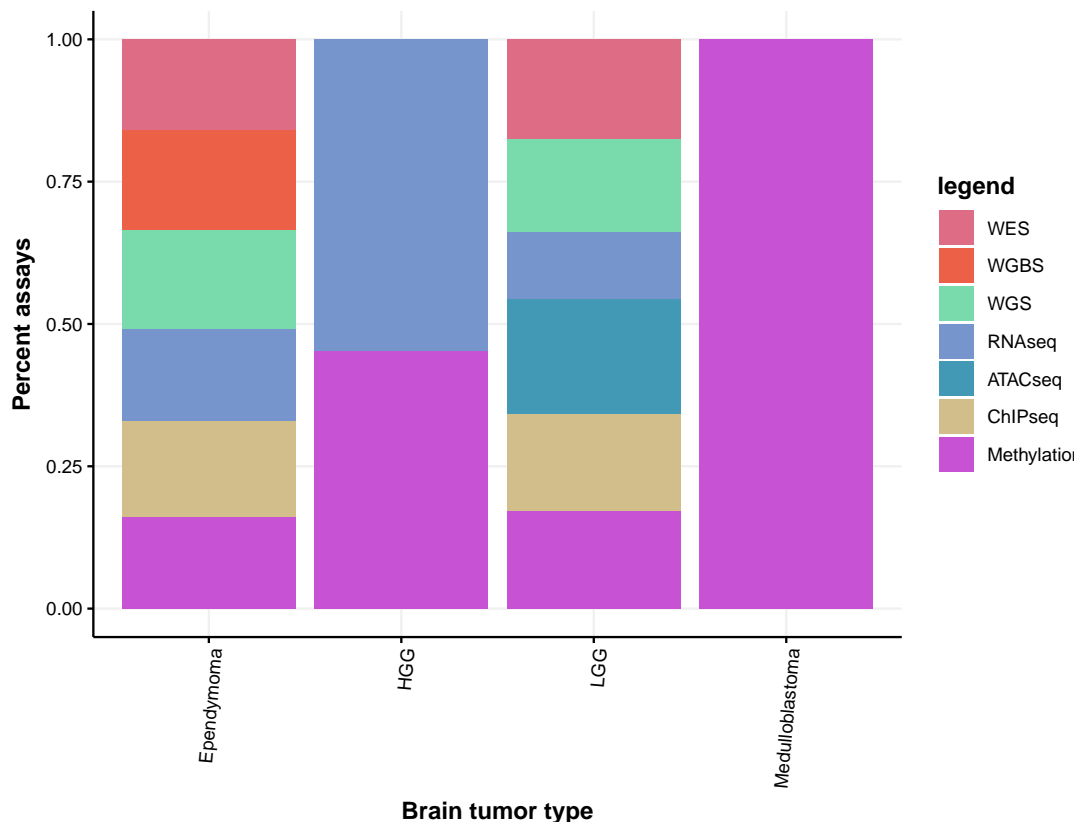
tables1 <- df %>% count(cancer_type_brain, Assay) %>%
  as.data.frame() %>%
  mutate_all(funs(replace_na(.,0)))

## Warning: `funs()` was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with `tibble::lst()`: tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.

# Plot stacked barplot
fname <- paste0(figures_plots_dir, "/", "cancer-type-brain-per-assay-longitudinal.pdf")
p <- create_stacked_barplot(plot_df = tables1,
  x_value = tables1$cancer_type_brain,
  use_palette = palette,
  xtitle = "Brain tumor type",
  legend = tables1$Assay,
  title_value = caption_value)

```

Summary of samples and assays per brain cancer type and per assay



```

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

```

```
## pdf
```

### 6.3 Per disease stage

Table 4: Summary of samples and assays per brain cancer type, per disease stage and per assay

can- cer_type_brain	longitudinal	WES	WGBS	WGS	RNAseq	ChIPseq	Methylation	ATAC- seq
Ependymoma	Diagnosis	15	15	15	15	15	13	0
Ependymoma	Remission	15	17	17	15	17	17	0
Ependymoma	Relapse	16	18	18	16	17	16	0
Ependymoma	Deceased	18	20	20	18	19	18	0
HGG	Diagnosis	0	0	0	13	0	11	0
HGG	Remission	0	0	0	13	0	11	0
HGG	Relapse	0	0	0	14	0	12	0
HGG	Deceased	0	0	0	11	0	8	0
LGG	Diagnosis	16	0	14	11	15	15	17
LGG	Remission	14	0	13	10	14	14	17
LGG	Relapse	17	0	15	11	16	16	19
LGG	Deceased	14	0	14	9	14	14	17
Medulloblastoma	Diagnosis	0	0	0	0	0	9	0
Medulloblastoma	Remission	0	0	0	0	0	10	0
Medulloblastoma	Relapse	0	0	0	0	0	10	0
Medulloblastoma	Deceased	0	0	0	0	0	9	0

## 7 Number of samples per brain cancer type, longitudinal, assay, and SJUID

Table 5: Summary of cohort per brain cancer type, SJUID, disease stage, and assay

can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
Ependymoma	SJH0H5WYRHR	Remission	1	1	1	1	1	0	0
Ependymoma	SJH0H5WYREIR	Relapse	1	1	1	1	1	0	0
Ependymoma	SJH0H5WYREID	Deceased	1	1	1	1	1	0	0
Ependymoma	SJH2HPPEEKMD	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJH2HPPEEKMR	Remission	1	1	1	1	1	1	0
Ependymoma	SJH2HPPEEKMR	Relapse	1	1	1	1	1	1	0
Ependymoma	SJH2HPPEEKMD	Deceased	1	1	1	1	1	1	0
Ependymoma	SJH51B396IW	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJH51B396IW	Remission	1	1	1	1	1	1	0
Ependymoma	SJH51B396IW	Relapse	1	1	1	1	1	1	0
Ependymoma	SJH51B396IW	Deceased	1	1	1	1	1	1	0
Ependymoma	SJH5HCKPC97R	Remission	0	1	1	1	1	1	0
Ependymoma	SJH5HCKPC97R	Relapse	0	1	1	1	1	1	0
Ependymoma	SJH5HCKPC97R	Deceased	0	1	1	1	1	1	0
Ependymoma	SJH98QNKIUU	Remission	0	1	1	1	1	1	0
Ependymoma	SJH98QNKIUU	Relapse	0	1	1	1	1	1	0
Ependymoma	SJH98QNKIUU	Deceased	0	1	1	1	1	1	0
Ependymoma	SJH9YML- FOTI	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJH9YML- FOTI	Remission	1	1	1	1	1	1	0
Ependymoma	SJH9YML- FOTI	Relapse	1	1	1	1	1	1	0
Ependymoma	SJH9YML- FOTI	Deceased	1	1	1	1	1	1	0
Ependymoma	SJHA6KC56J6	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHA6KC56J6	Remission	1	1	1	1	1	1	0
Ependymoma	SJHA6KC56J6	Relapse	1	1	1	1	1	1	0
Ependymoma	SJHA6KC56J6	Deceased	1	1	1	1	1	1	0
Ependymoma	SJH- BKS7QSFO	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJH- BKS7QSFO	Remission	1	1	1	1	1	1	0
Ependymoma	SJH- BKS7QSFO	Relapse	1	1	1	1	1	1	0
Ependymoma	SJH- BKS7QSFO	Deceased	1	1	1	1	1	1	0
Ependymoma	SJHC70DZRJS	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHC70DZRJS	Remission	1	1	1	1	1	1	0
Ependymoma	SJHC70DZRJS	Relapse	1	1	1	1	1	1	0
Ependymoma	SJHC70DZRJS	Deceased	1	1	1	1	1	1	0
Ependymoma	SJHCN03RCVD	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHCN03RCVD	Remission	1	1	1	1	1	1	0
Ependymoma	SJHCN03RCVD	Deceased	1	1	1	1	1	1	0

can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
Ependymoma	SJHD1KI- UMM6	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHD1KI- UMM6	Remission	1	1	1	1	1	1	0
Ependymoma	SJHD1KI- UMM6	Relapse	1	1	1	1	1	1	0
Ependymoma	SJHD1KI- UMM6	Deceased	1	1	1	1	1	1	0
Ependymoma	SJHFG- GJRHYO	Remission	1	1	1	1	1	0	0
Ependymoma	SJHFG- GJRHYO	Relapse	1	1	1	1	1	0	0
Ependymoma	SJHFG- GJRHYO	Deceased	1	1	1	1	1	0	0
Ependymoma	SJH- HZH67WMF	Relapse	1	1	1	0	1	1	0
Ependymoma	SJH- HZH67WMF	Deceased	1	1	1	0	1	1	0
Ependymoma	SJHI8CC7QT8	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHI8CC7QT8	Remission	1	1	1	1	1	1	0
Ependymoma	SJHI8CC7QT8	Relapse	1	1	1	1	1	1	0
Ependymoma	SJHI8CC7QT8	Deceased	1	1	1	1	1	1	0
Ependymoma	SJHIQT- NAYIF	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHIQT- NAYIF	Remission	1	1	1	1	1	1	0
Ependymoma	SJHIQT- NAYIF	Relapse	1	1	1	1	1	1	0
Ependymoma	SJHIQT- NAYIF	Deceased	1	1	1	1	1	1	0
Ependymoma	SJHKVQE- BOCP	Diagnosis	1	1	1	1	0	1	0
Ependymoma	SJHKVQE- BOCP	Relapse	1	1	1	1	0	1	0
Ependymoma	SJHKVQE- BOCP	Deceased	1	1	1	1	0	1	0
Ependymoma	SJHKYUIYAM	Diagnosis	1	1	1	1	0	1	0
Ependymoma	SJHKYUIYAM	Relapse	1	1	1	1	0	1	0
Ependymoma	SJHKYUIYAM	Deceased	1	1	1	1	0	1	0
Ependymoma	SJH- PVXLA- CLM	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJH- PVXLA- CLM	Remission	1	1	1	1	1	1	0
Ependymoma	SJH- PVXLA- CLM	Relapse	1	1	1	1	1	1	0
Ependymoma	SJH- PVXLA- CLM	Deceased	1	1	1	1	1	1	0

can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
Ependymoma	SJHUMKP2L6M	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHUMKP2L6N	Remission	1	1	1	1	1	1	0
Ependymoma	SJHUMKP2L6V	Deceased	1	1	1	1	1	1	0
Ependymoma	SJHXTBTQ5ZD	Diagnosis	1	1	1	1	1	1	0
Ependymoma	SJHXTBTQ5ZH	Remission	1	1	1	1	1	1	0
Ependymoma	SJHXTBTQ5ZT	Relapse	1	1	1	1	1	1	0
Ependymoma	SJHXTBTQ5ZD	Deceased	1	1	1	1	1	1	0
HGG	SJH5QHZM4U8	Diagnosis	0	0	0	0	1	1	0
HGG	SJH5QHZM4U8	Remission	0	0	0	0	1	1	0
HGG	SJH5QHZM4U8	Relapse	0	0	0	0	1	1	0
HGG	SJH5QHZM4U8	Deceased	0	0	0	0	1	1	0
HGG	SJHADL5OJME	Diagnosis	0	0	0	0	1	1	0
HGG	SJHADL5OJME	Remission	0	0	0	0	1	1	0
HGG	SJHADL5OJME	Relapse	0	0	0	0	1	1	0
HGG	SJHADL5OJME	Deceased	0	0	0	0	1	1	0
HGG	SJH- BISK5KBU	Diagnosis	0	0	0	0	1	1	0
HGG	SJH- BISK5KBU	Remission	0	0	0	0	1	1	0
HGG	SJH- BISK5KBU	Relapse	0	0	0	0	1	1	0
HGG	SJHBL7CDYRN	Diagnosis	0	0	0	0	1	1	0
HGG	SJHBL7CDYRN	Remission	0	0	0	0	1	1	0
HGG	SJHBL7CDYRN	Relapse	0	0	0	0	1	1	0
HGG	SJHBL7CDYRN	Deceased	0	0	0	0	1	1	0
HGG	SJHD- DTE0SYL	Diagnosis	0	0	0	0	0	1	0
HGG	SJHD- DTE0SYL	Relapse	0	0	0	0	0	1	0
HGG	SJHEOVURBMD	Diagnosis	0	0	0	0	1	1	0
HGG	SJHEOVURBMD	Remission	0	0	0	0	1	1	0
HGG	SJHEOVURBMD	Relapse	0	0	0	0	1	1	0
HGG	SJHEOVURBMD	Deceased	0	0	0	0	1	1	0
HGG	SJHEQG3P4FK	Diagnosis	0	0	0	0	1	0	0
HGG	SJHEQG3P4FK	Relapse	0	0	0	0	1	0	0
HGG	SJHHPN- QERSQ	Diagnosis	0	0	0	0	1	1	0
HGG	SJHHPN- QERSQ	Remission	0	0	0	0	1	1	0
HGG	SJHHPN- QERSQ	Relapse	0	0	0	0	1	1	0
HGG	SJHHPN- QERSQ	Deceased	0	0	0	0	1	1	0
HGG	SJHHW23UJ9R	Remission	0	0	0	0	1	1	0
HGG	SJHHW23UJ9P	Relapse	0	0	0	0	1	1	0
HGG	SJHJ59RLHSU	Diagnosis	0	0	0	0	1	1	0
HGG	SJHJ59RLHSU	Remission	0	0	0	0	1	1	0
HGG	SJHJ59RLHSU	Relapse	0	0	0	0	1	1	0
HGG	SJHJ59RLHSU	Deceased	0	0	0	0	1	1	0
HGG	SJHKKD- DXOYH	Diagnosis	0	0	0	0	1	1	0

can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
HGG	SJHKKD-DXOYH	Remission	0	0	0	0	1	1	0
HGG	SJHKKD-DXOYH	Relapse	0	0	0	0	1	1	0
HGG	SJHKKD-DXOYH	Deceased	0	0	0	0	1	1	0
HGG	SJHMI643DMID	Diagnosis	0	0	0	0	1	1	0
HGG	SJHMI643DMID	Remission	0	0	0	0	1	1	0
HGG	SJHMI643DMID	Relapse	0	0	0	0	1	1	0
HGG	SJHMI643DMID	Deceased	0	0	0	0	1	1	0
HGG	SJHQBOPS9FID	Diagnosis	0	0	0	0	0	1	0
HGG	SJHQBOPS9FID	Remission	0	0	0	0	0	1	0
HGG	SJHQBOPS9FID	Relapse	0	0	0	0	0	1	0
HGG	SJHQBOPS9FID	Deceased	0	0	0	0	0	1	0
HGG	SJHRO-JUQZAP	Diagnosis	0	0	0	0	1	0	0
HGG	SJHRO-JUQZAP	Remission	0	0	0	0	1	0	0
HGG	SJHRO-JUQZAP	Relapse	0	0	0	0	1	0	0
HGG	SJHUT-PISXFQ	Diagnosis	0	0	0	0	0	1	0
HGG	SJHUT-PISXFQ	Remission	0	0	0	0	0	1	0
HGG	SJHUT-PISXFQ	Relapse	0	0	0	0	0	1	0
HGG	SJHUT-PISXFQ	Deceased	0	0	0	0	0	1	0
HGG	SJHVIS5Q8HXD	Diagnosis	0	0	0	0	0	1	0
HGG	SJHVIS5Q8HXD	Remission	0	0	0	0	0	1	0
HGG	SJHVIS5Q8HXD	Relapse	0	0	0	0	0	1	0
HGG	SJHVIS5Q8HXD	Deceased	0	0	0	0	0	1	0
LGG	SJH2W47P7DGD	Diagnosis	1	0	1	1	1	1	1
LGG	SJH2W47P7DGD	Remission	1	0	1	1	1	1	1
LGG	SJH2W47P7DGD	Relapse	1	0	1	1	1	1	1
LGG	SJH2W47P7DGD	Deceased	1	0	1	1	1	1	1
LGG	SJH5HCKPC97D	Remission	0	0	0	0	0	0	1
LGG	SJH5HCKPC97D	Relapse	0	0	0	0	0	0	1
LGG	SJH5HCKPC97D	Deceased	0	0	0	0	0	0	1
LGG	SJHAY4GX4AND	Diagnosis	1	0	1	1	1	1	1
LGG	SJHAY4GX4AND	Remission	1	0	1	1	1	1	1
LGG	SJHAY4GX4AND	Relapse	1	0	1	1	1	1	1
LGG	SJHAY4GX4AND	Deceased	1	0	1	1	1	1	1
LGG	SJHBN-JSZHW6	Remission	1	0	1	1	1	0	1
LGG	SJHBN-JSZHW6	Relapse	1	0	1	1	1	0	1
LGG	SJHBN-JSZHW6	Deceased	1	0	1	1	1	0	1
LGG	SJHBV3Q6UVB	Diagnosis	1	0	1	1	1	1	1
LGG	SJHBV3Q6UVB	Remission	1	0	1	1	1	1	1

can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
LGG	SJHBV3Q6UVB	Relapse	1	0	1	1	1	1	1
LGG	SJHBV3Q6UVB	Deceased	1	0	1	1	1	1	1
LGG	SJHC70DZRJS	Diagnosis	0	0	0	0	0	0	1
LGG	SJHC70DZRJS	Remission	0	0	0	0	0	0	1
LGG	SJHC70DZRJS	Relapse	0	0	0	0	0	0	1
LGG	SJHC70DZRJS	Deceased	0	0	0	0	0	0	1
LGG	SJH- CLGFJTIG	Diagnosis	1	0	1	1	1	1	1
LGG	SJH- CLGFJTIG	Remission	1	0	1	1	1	1	1
LGG	SJH- CLGFJTIG	Relapse	1	0	1	1	1	1	1
LGG	SJH- CLGFJTIG	Deceased	1	0	1	1	1	1	1
LGG	SJHD- DTE0SYL	Diagnosis	1	0	1	1	0	0	0
LGG	SJHD- DTE0SYL	Relapse	1	0	1	1	0	0	0
LGG	SJHI52BLNWK	Diagnosis	1	0	1	1	1	1	1
LGG	SJHI52BLNWK	Remission	1	0	1	1	1	1	1
LGG	SJHI52BLNWK	Relapse	1	0	1	1	1	1	1
LGG	SJHI52BLNWK	Deceased	1	0	1	1	1	1	1
LGG	SJHIQT- NAYIF	Diagnosis	0	0	0	0	0	0	1
LGG	SJHIQT- NAYIF	Remission	0	0	0	0	0	0	1
LGG	SJHIQT- NAYIF	Relapse	0	0	0	0	0	0	1
LGG	SJHIQT- NAYIF	Deceased	0	0	0	0	0	0	1
LGG	SJHN- PJTQHIT	Diagnosis	1	0	1	1	1	0	1
LGG	SJHN- PJTQHIT	Relapse	1	0	1	1	1	0	1
LGG	SJHN- PJTQHIT	Deceased	1	0	1	1	1	0	1
LGG	SJHO- FORRR7C	Diagnosis	1	0	1	1	1	1	1
LGG	SJHO- FORRR7C	Remission	1	0	1	1	1	1	1
LGG	SJHO- FORRR7C	Relapse	1	0	1	1	1	1	1
LGG	SJHO- FORRR7C	Deceased	1	0	1	1	1	1	1
LGG	SJHOY05OJJN	Diagnosis	1	0	0	0	1	1	1
LGG	SJHOY05OJJN	Relapse	1	0	0	0	1	1	1
LGG	SJHQBOPS9FD	Diagnosis	1	0	1	1	1	0	1
LGG	SJHQBOPS9FD	Remission	1	0	1	1	1	0	1
LGG	SJHQBOPS9FD	Relapse	1	0	1	1	1	0	1
LGG	SJHQBOPS9FD	Deceased	1	0	1	1	1	0	1
LGG	SJHUT- PISXFQ	Diagnosis	1	0	1	1	1	0	1

can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
LGG	SJHUT- PISXFQ	Remission	1	0	1	1	1	0	1
LGG	SJHUT- PISXFQ	Relapse	1	0	1	1	1	0	1
LGG	SJHUT- PISXFQ	Deceased	1	0	1	1	1	0	1
LGG	SJHVIS5Q8HX	Diagnosis	1	0	1	1	1	0	1
LGG	SJHVIS5Q8HX	Remission	1	0	1	1	1	0	1
LGG	SJHVIS5Q8HX	Relapse	1	0	1	1	1	0	1
LGG	SJHVIS5Q8HX	Deceased	1	0	1	1	1	0	1
LGG	SJHWS0NRZV	Diagnosis	1	0	0	1	1	1	1
LGG	SJHWS0NRZV	Remission	1	0	0	1	1	1	1
LGG	SJHWS0NRZV	Relapse	1	0	0	1	1	1	1
LGG	SJHXIKCWNK	Diagnosis	1	0	1	1	1	1	1
LGG	SJHXIKCWNK	Remission	1	0	1	1	1	1	1
LGG	SJHXIKCWNK	Relapse	1	0	1	1	1	1	1
LGG	SJHXIKCWNK	Deceased	1	0	1	1	1	1	1
LGG	SJHYP- KTG3P5	Diagnosis	1	0	1	1	1	1	1
LGG	SJHYP- KTG3P5	Remission	1	0	1	1	1	1	1
LGG	SJHYP- KTG3P5	Relapse	1	0	1	1	1	1	1
LGG	SJHYP- KTG3P5	Deceased	1	0	1	1	1	1	1
LGG	SJHZW7GYEF	Diagnosis	1	0	1	1	1	1	1
LGG	SJHZW7GYEF	Remission	1	0	1	1	1	1	1
LGG	SJHZW7GYEF	Relapse	1	0	1	1	1	1	1
LGG	SJHZW7GYEF	Deceased	1	0	1	1	1	1	1
Medulloblas- toma	SJH77NRD- WUX	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJH77NRD- WUX	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJH77NRD- WUX	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJH77NRD- WUX	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHAF- TIOMPQ	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHAF- TIOMPQ	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHAF- TIOMPQ	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHAF- TIOMPQ	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHC1QST5GR	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHC1QST5GR	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHC1QST5GR	Deceased	0	0	0	0	1	0	0



can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
Medulloblas- toma	SJHF- CYGKSDY	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHF- CYGKSDY	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHF- CYGKSDY	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHF- CYGKSDY	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHJAC- CGA3S	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHJAC- CGA3S	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHJAC- CGA3S	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHJAC- CGA3S	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHQS0D51KH	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHQS0D51KH	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHQS0D51KH	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHQS0D51KH	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHXVMEU21I	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHXVMEU21I	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHXVMEU21I	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHXVMEU21I	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHY4W1ZWCY	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHY4W1ZWCY	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHY4W1ZWCY	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHY4W1ZWCY	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHZ40PT- CYH	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHZ40PT- CYH	Remission	0	0	0	0	1	0	0
Medulloblas- toma	SJHZ40PT- CYH	Relapse	0	0	0	0	1	0	0
Medulloblas- toma	SJHZ40PT- CYH	Deceased	0	0	0	0	1	0	0
Medulloblas- toma	SJHZZLR- CGJ6	Diagnosis	0	0	0	0	1	0	0
Medulloblas- toma	SJHZZLR- CGJ6	Remission	0	0	0	0	1	0	0

can- cer_type_brain	SJUID	longitudi- nal	WES	WGBS	WGS	ChIPseq	Methyla- tion	RNAseq	ATAC- seq
Medulloblas- toma	SJHZZLR- CGJ6	Relapse	0	0	0	0	1	0	0

## 7.1 Plots

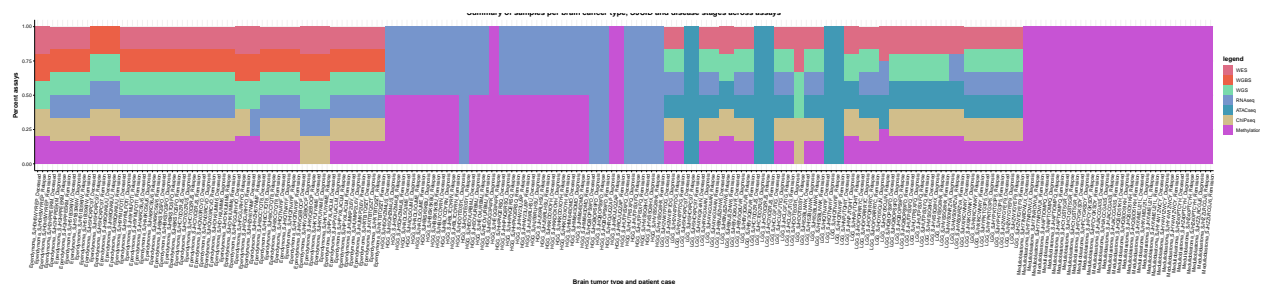
```
tables1 <- df %>% count(cancer_type_brain, SJUID, longitudinal, Assay) %>%
  as.data.frame() %>%
  arrange(cancer_type_brain, longitudinal, Assay) %>%
  mutate(match_id_brain = paste(cancer_type_brain, SJUID, sep = "_"),
         match_id_brain_longitudinal = paste(cancer_type_brain, SJUID, longitudinal, sep = "_")) %>%
  mutate_all(funs(replace_na(.,0)))
```

```
## Warning: `funs()` was deprecated in dplyr 0.8.0.
## i Please use a list of either functions or lambdas:
##
## # Simple named list: list(mean = mean, median = median)
##
## # Auto named with `tibble::lst()`: tibble::lst(mean, median)
##
## # Using lambdas list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

*# Run function*

```
fname <- paste0(figures_plots_dir, "/", "cancer-type-brain-per-assay-patient.pdf")
```

```
p <- create_stacked_barplot(plot_df = tables1,
                           x_value = tables1$match_id_brain_longitudinal,
                           use_palette = palette,
                           xtitle = "Brain tumor type and patient case",
                           legend = tables1$Assay,
                           title_value = "Summary of samples per brain cancer type, SJUID and disease state")
```



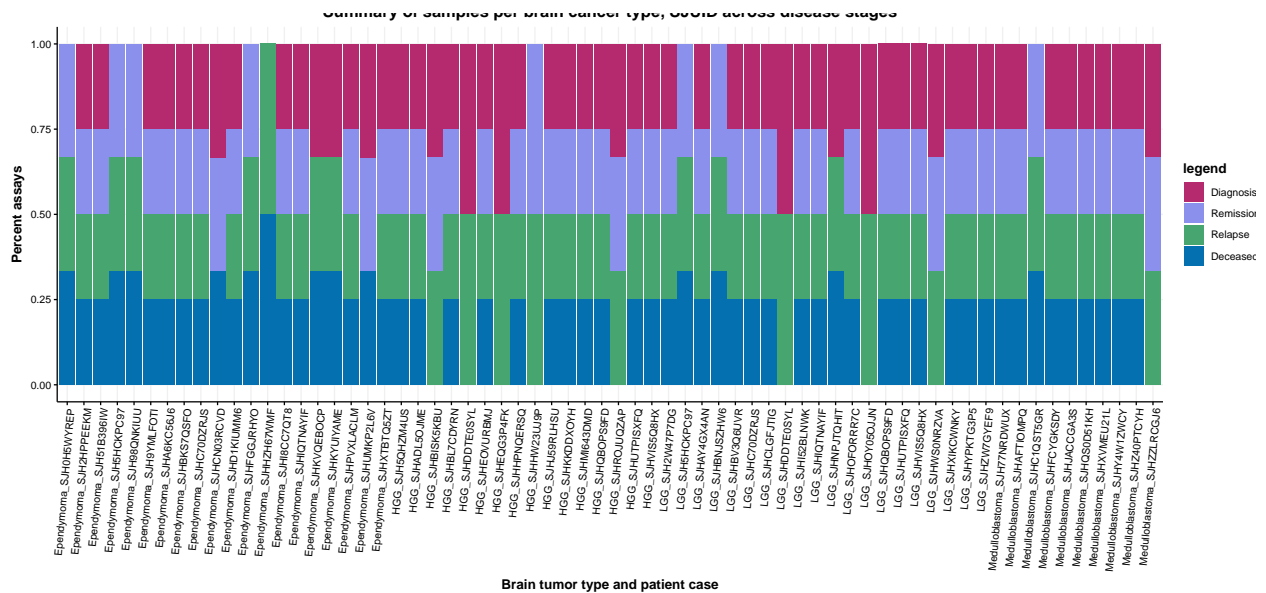
```
pdf(file = fname, width = 25, height = 6)
print(p)
dev.off()
```

```
## pdf
## 2
```

*# Run function*

```
fname <- paste0(figures_plots_dir, "/", "cancer-type-brain-patient-longitudinal.pdf")
```

```
p <- create_stacked_barplot(plot_df = tables1,
                           x_value = tables1$match_id_brain,
                           use_palette = longitudinal_palette,
                           xtitle = "Brain tumor type and patient case",
                           legend = tables1$longitudinal,
                           title_value = "Summary of samples per brain cancer type, SJUID across disease state")
```



```
pdf(file = fname, width = 12, height = 6)
print(p)
dev.off()
```

```
## pdf
## 2
```

## 8 Future directions

The current exploratory data analysis module can be expanded by investigating samples with paired assays. Moreover, if other metadata are available, e.g., disease\_stage, treatment, this will build large, longitudinal cohorts with multi-omic sequencing data. Such an analysis permits consideration of samples according to the condition(s) of the experiment and research aims. In addition, it can be used to refine research questions and/or generate new ones.

This will facilitate collaboration across departments at St. Jude, expedite discoveries, and find cures for children with cancer and other catastrophic diseases.

## 9 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:   /research/rgs01/applications/hpcf/authorized_apps/rhel8_apps/lapack/3.10.1/install/lib64/libblas.so.3
## LAPACK: /research/rgs01/applications/hpcf/authorized_apps/rhel8_apps/lapack/3.10.1/install/lib64/liblapack.so.3
##
## 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
##
## 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    yaml_2.3.10     tools_4.4.0
## [25] parallel_4.4.0  tzdb_0.4.0      colorspace_2.1-1 vctrs_0.6.5
## [29] R6_2.5.1        mime_0.12        lifecycle_1.0.4 bit_4.0.5
## [33] vroom_1.6.5     pkgconfig_2.0.3 pillar_1.9.0    bslib_0.8.0
## [37] gtable_0.3.5    glue_1.7.0      xfun_0.47       tidysselect_1.2.1
## [41] highr_0.11      knitr_1.48       farver_2.1.2    htmltools_0.5.8.1
## [45] rmarkdown_2.28  labeling_0.4.3   compiler_4.4.0
```