

main

October 20, 2021

1 Quality Control for the Caudate Nucleus

```
[1]: suppressMessages({library(repr)
                        library(ggpubr)
                        library(tidyverse)})
```

1.1 Functions

```
[2]: save_img <- function(image, fn, w, h){
      for(ext in c(".svg", ".pdf", ".png")){
        ggsave(file=paste0(fn, ext), plot=image, width=w, height=h)
      }
    }

get_pheno <- function(){
  phenotypes <- data.table::fread(paste0("/ceph/projects/v4_phase3_paper/
    ↪inputs/",
                                         "phenotypes/_m/merged_phenotypes.
    ↪csv")) %>%
    mutate_if(is.character, as.factor)
  levels(phenotypes$Sex) <- c("Female", "Male")
  return(phenotypes)
}

memPHENO <- memoise::memoise(get_pheno)
```

1.2 Covariates examination

1.2.1 Examine if there are significant differences between diagnosis in continous covariates

```
[3]: covarsCont = memPHENO() %>%
      select(-c('RNum', "Race", "Sex", "Dx", 'Region', 'BrNum', "antipsychotics",
                "lifetime_antipsych", "Protocol")) %>% colnames
options(repr.plot.width=12, repr.plot.height=6)
dir.create("covariate_plots")
for(covar in covarsCont){
  set.seed(20211020)
```

```

flush.console()
bxp = memPHENO() %>%
  filter(Age > 17, Race %in% c("AA", "EA"), Dx %in% c("CTL", "SZ")) %>%
  ggboxplot(x='Sex', y=covar, fill='Sex', add='jitter', palette="npg",
            facet.by="Region", legend='bottom', panel.labs.
→font=list(face='bold'),
            add.params=list(alpha=0.75),
→ggtheme=theme_pubr(base_size=20)) +
  stat_compare_means(comparisons=list(c("Female", "Male")),
                    aes(label=..p.signif..), method="wilcox.test") +
  font("title", color="black", face="bold") +
  font("xy.title", face="bold") + #font("xy.text", size=14) +
  font("legend.title", size=16, face="bold") #font("legend.text", size=14)
  save_img(bxp, paste0("covariate_plots/boxplot_sex_filtered_", covar), w=12,
→h=6)
print(bxp)
bxp = memPHENO() %>%
  filter(Age > 17, Race %in% c("AA", "EA"), Dx %in% c("CTL", "SZ")) %>%
  ggboxplot(x='antipsychotics', y=covar, color='Sex', add='jitter',
            palette="npg", facet.by="Region", legend='bottom',
            ggtheme=theme_pubr(base_size=20)) +
  stat_compare_means(comparisons=list(c("TRUE", "FALSE")),
                    aes(label=..p.signif..), method="wilcox.test") +
  font("title", color="black", face="bold") +
  font("xy.title", face="bold") +
  font("legend.title", size=16, face="bold")
  save_img(bxp, paste0("covariate_plots/
→boxplot_sexBYantipsychotics_filtered_", covar),
            w=12, h=6)
print(bxp)
}

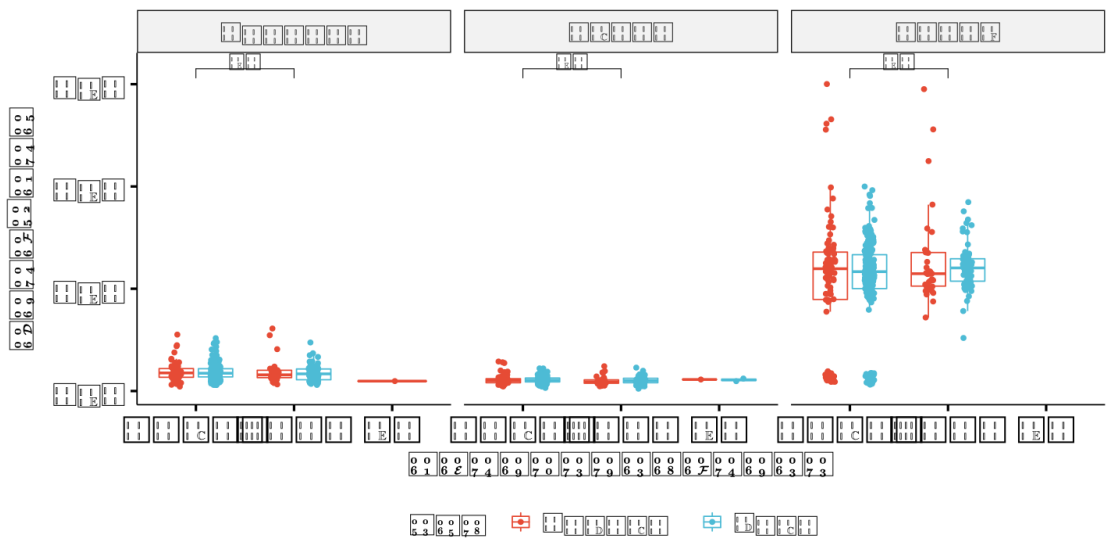
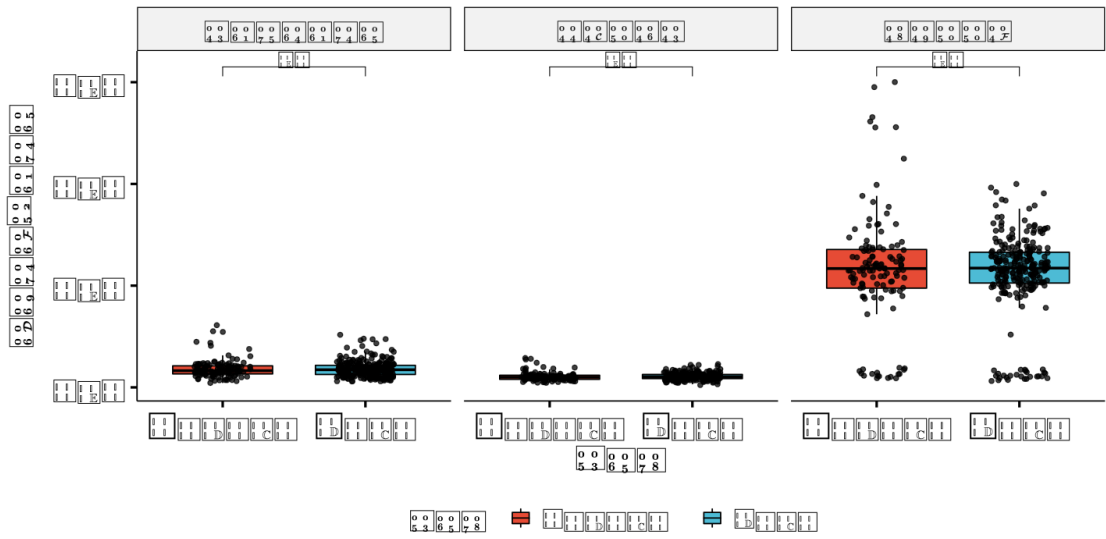
```

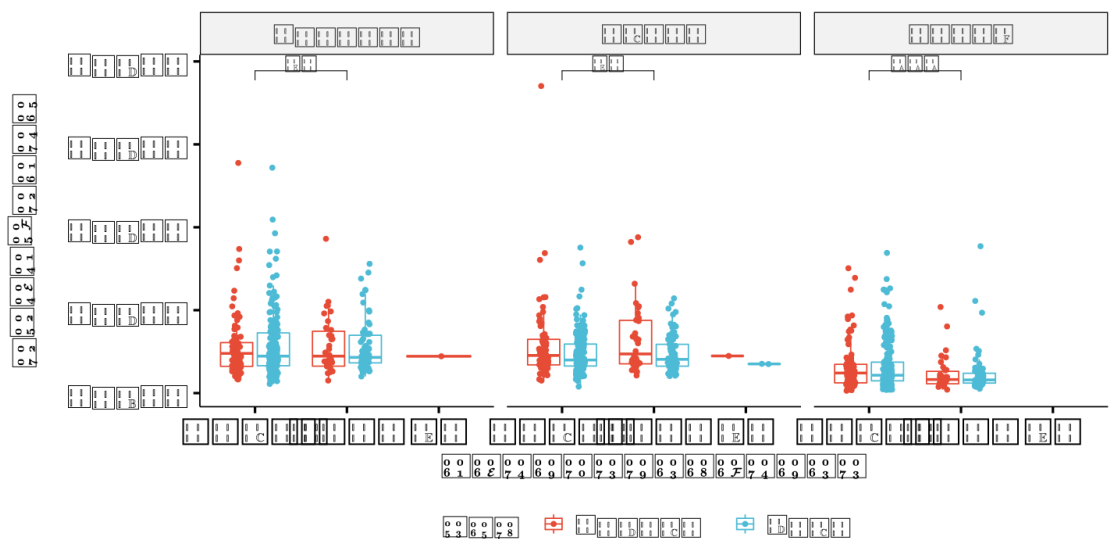
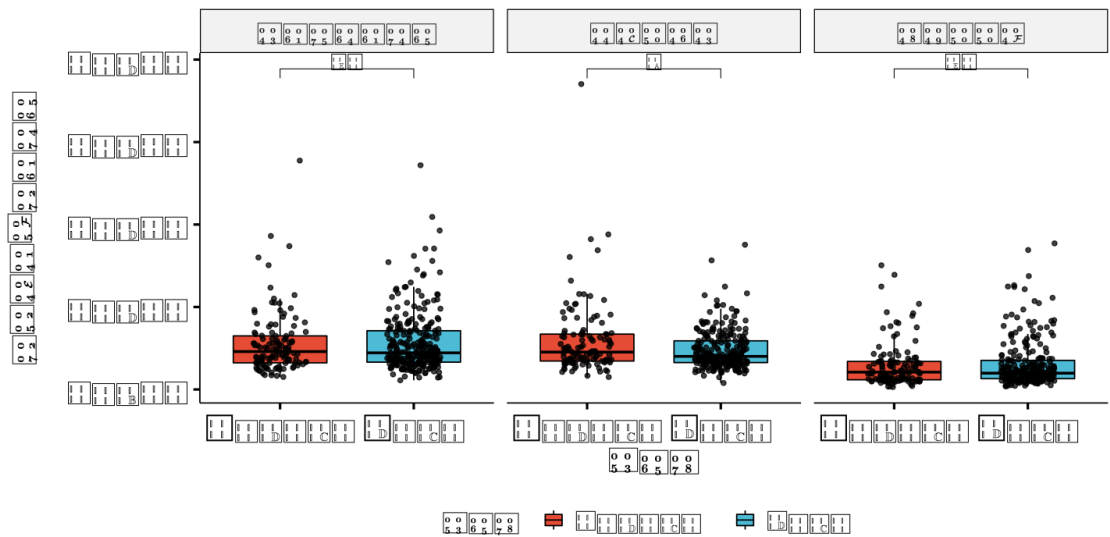
Warning message:

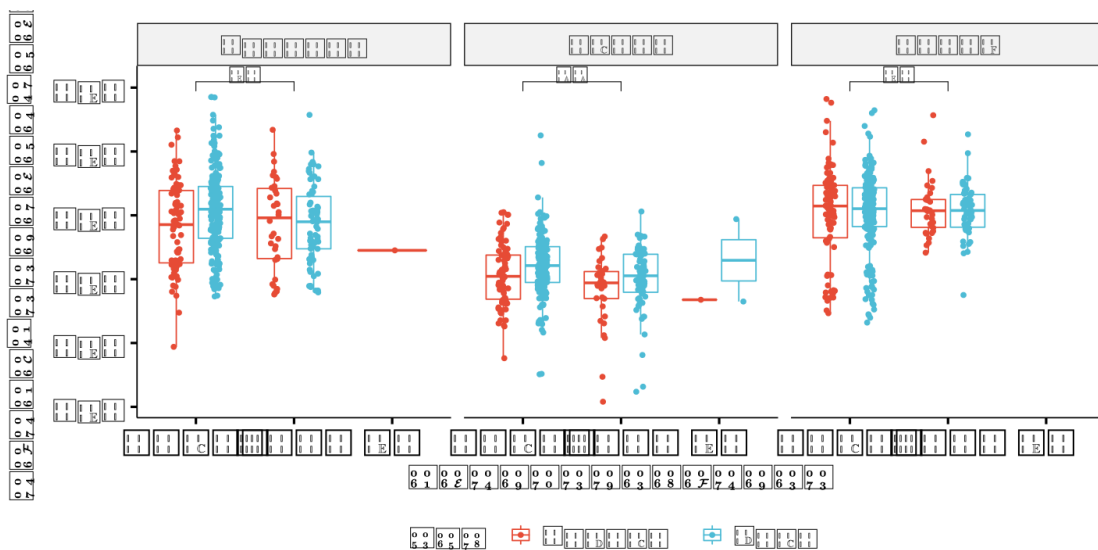
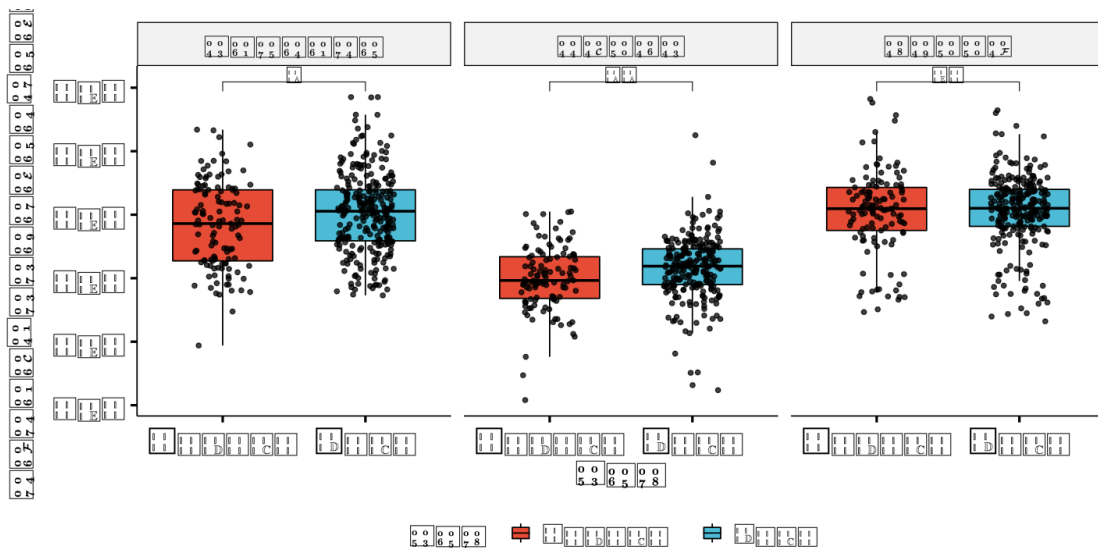
"Using `as.character()` on a quosure is deprecated as of rlang 0.3.0.

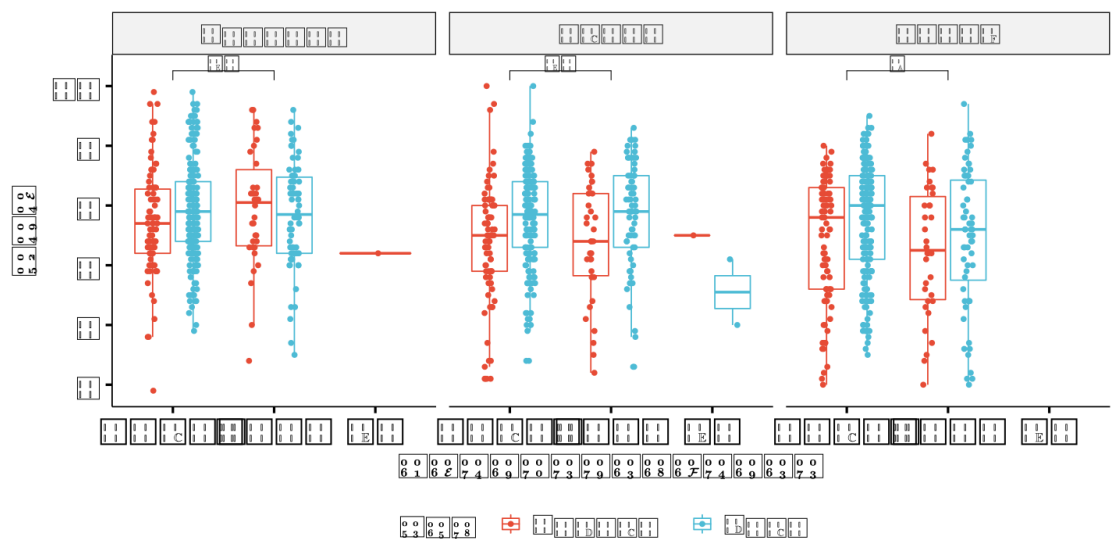
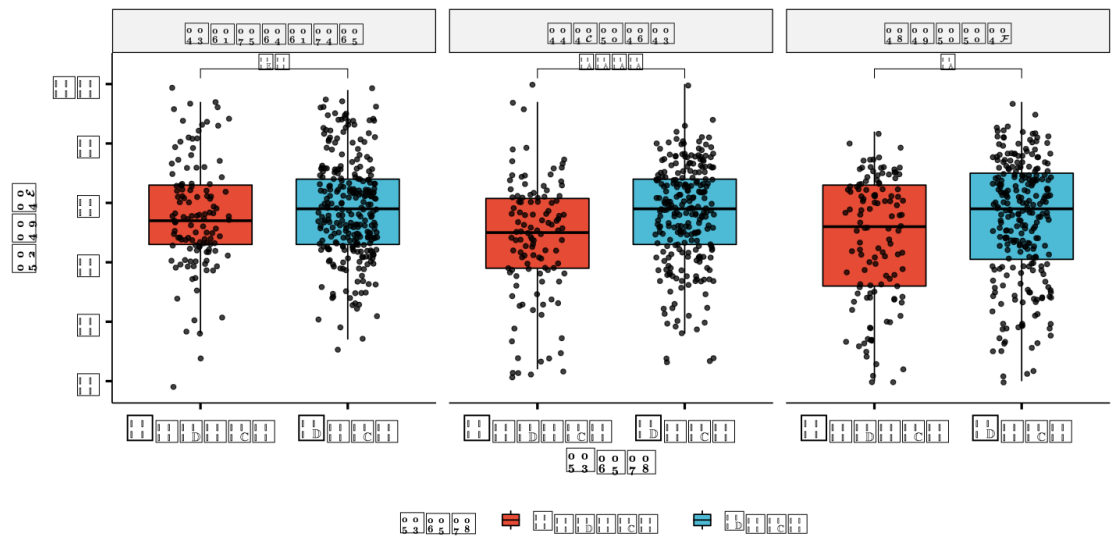
Please use `as_label()` or `as_name()` instead.

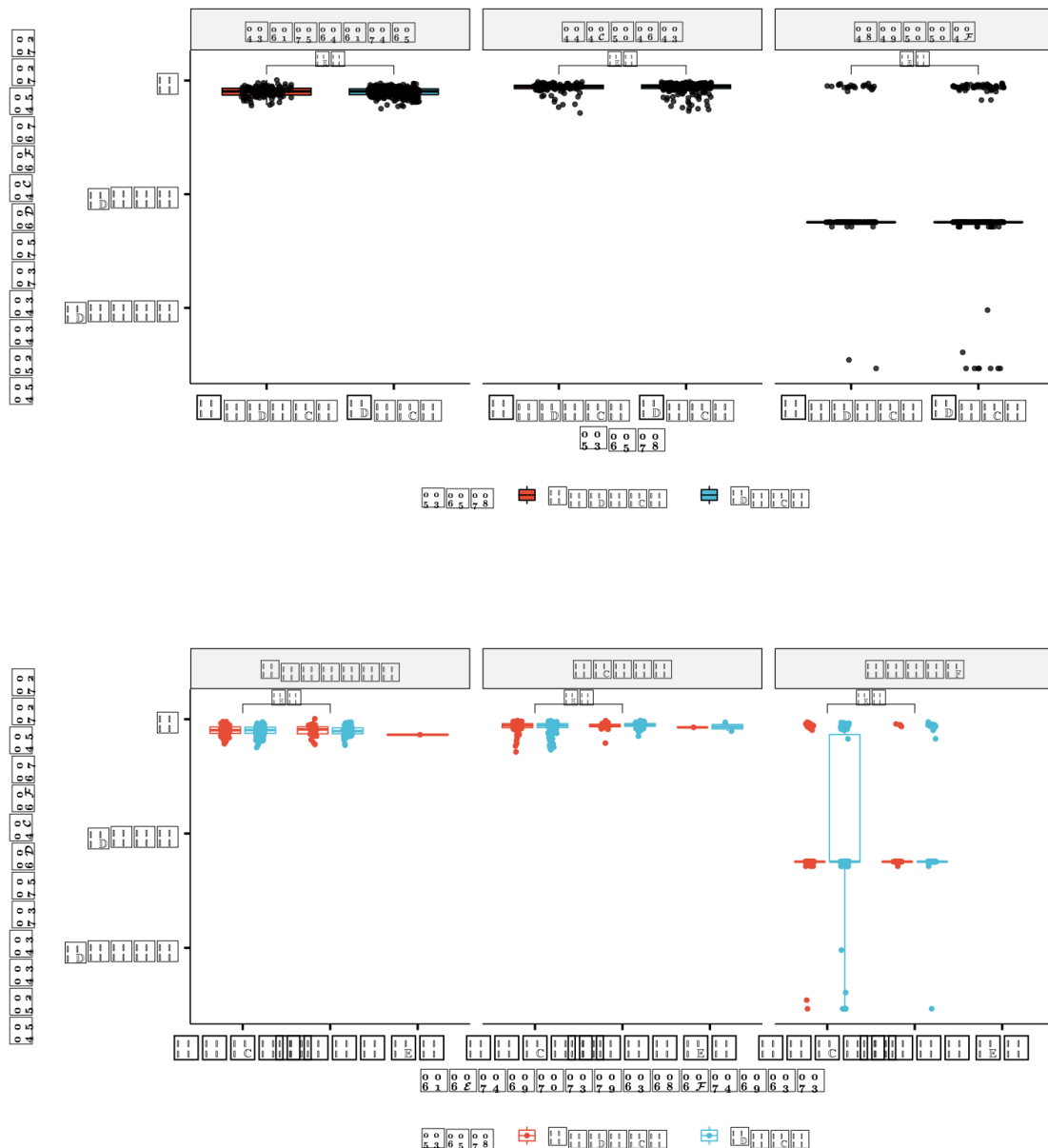
This warning is displayed once per session."



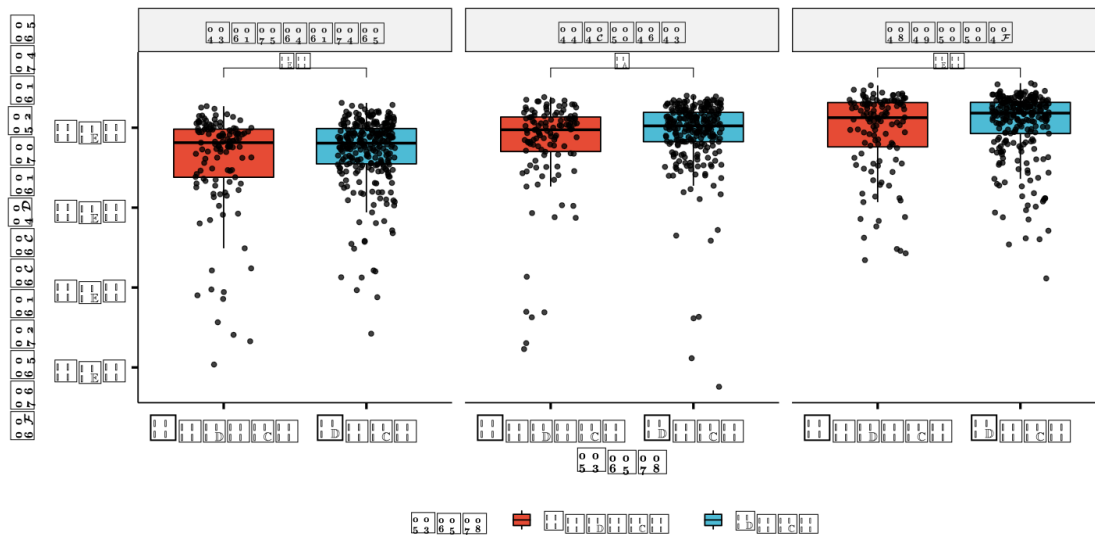






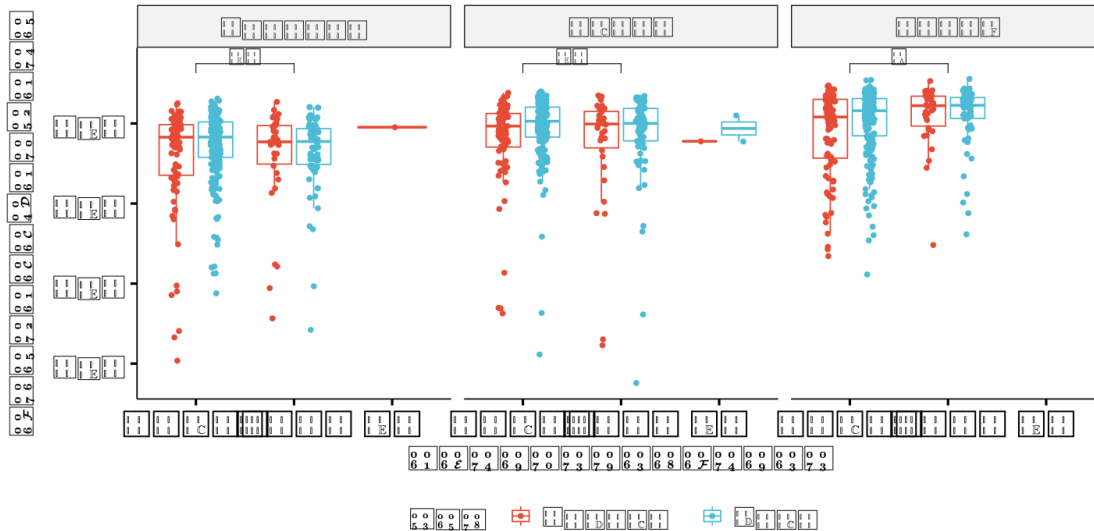



```
"Removed 3 rows containing missing values (geom_point)."
```

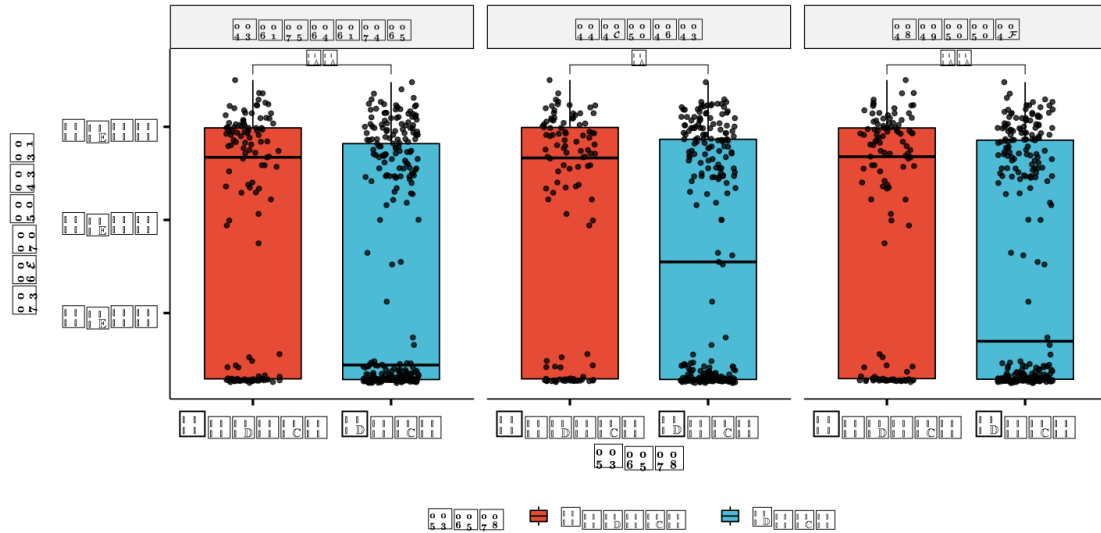


```
"Removed 3 rows containing non-finite values (stat_boxplot)."
```

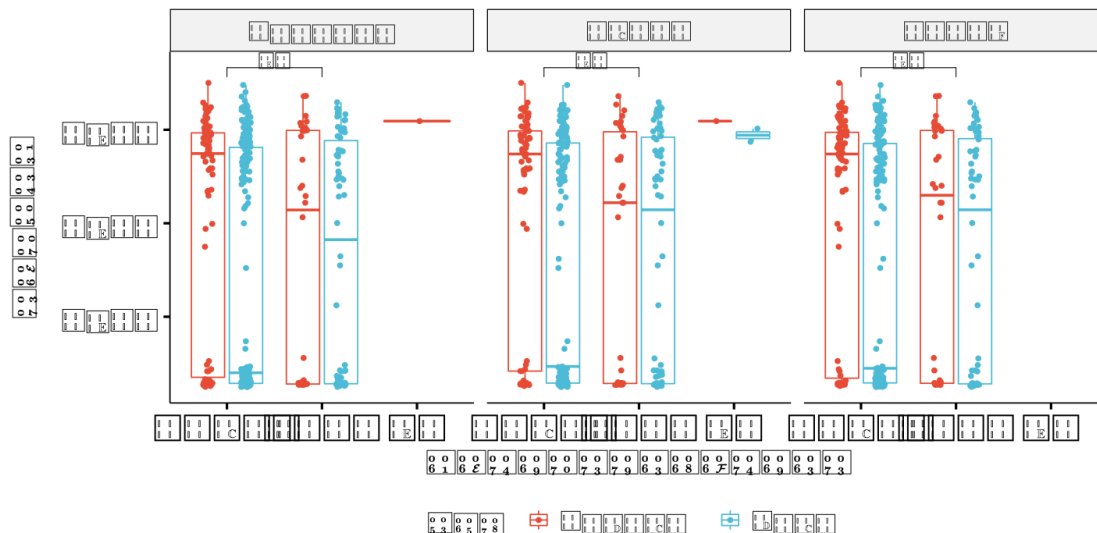
Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



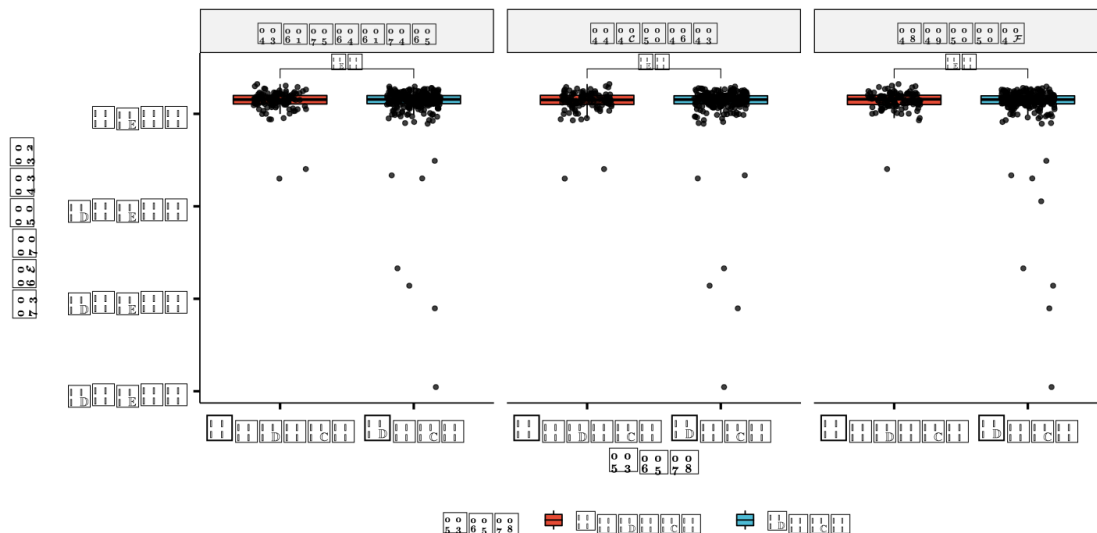
Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



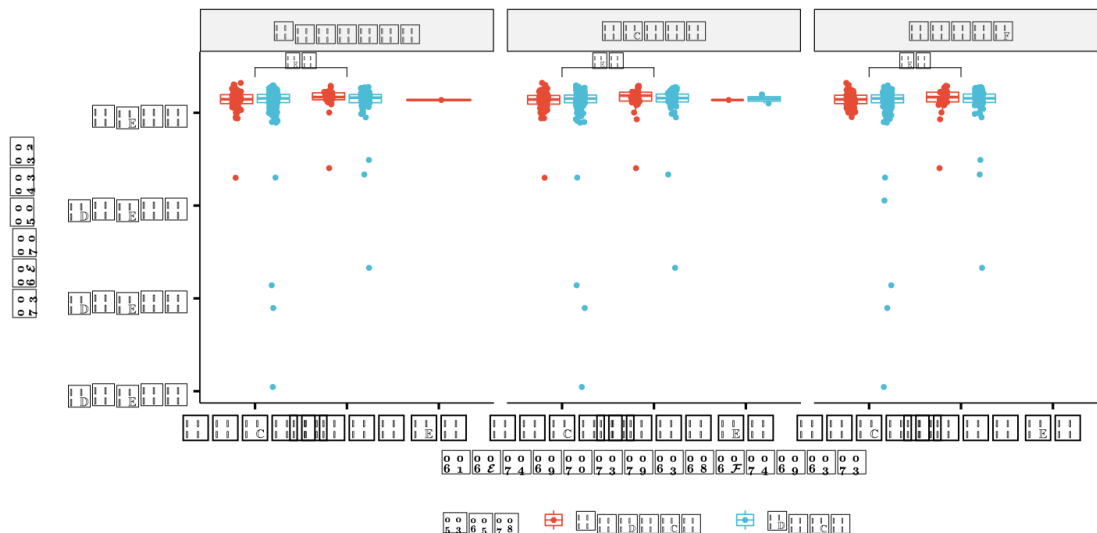
Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."

Warning message:

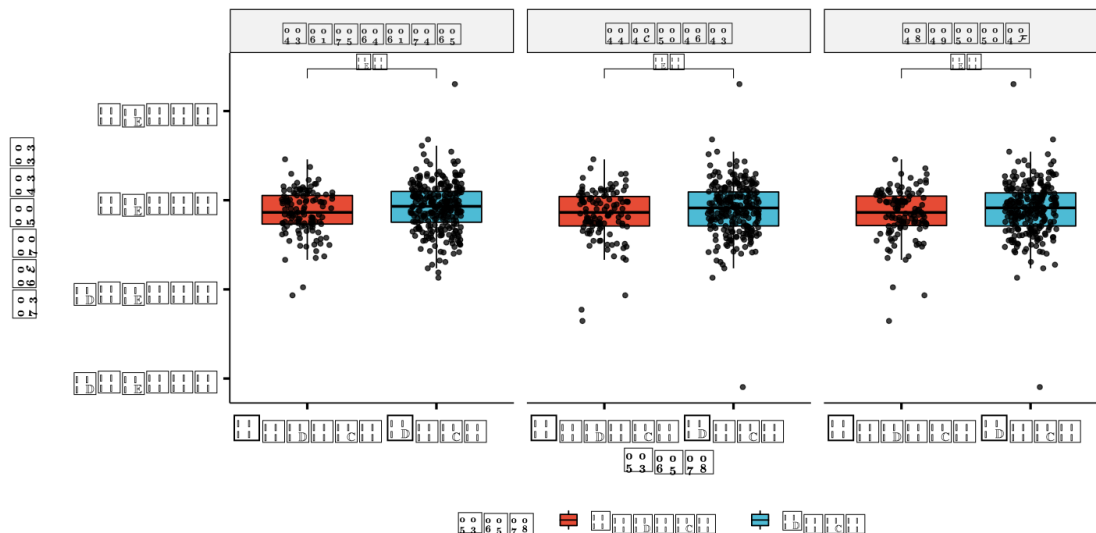
"Removed 3 rows containing non-finite values (stat_boxplot)."

Warning message:

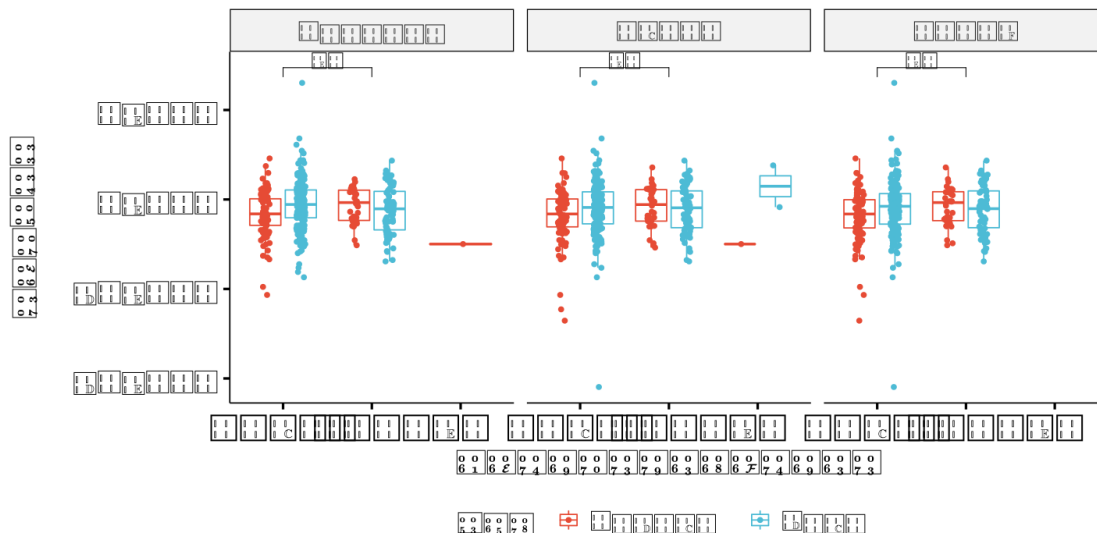
"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

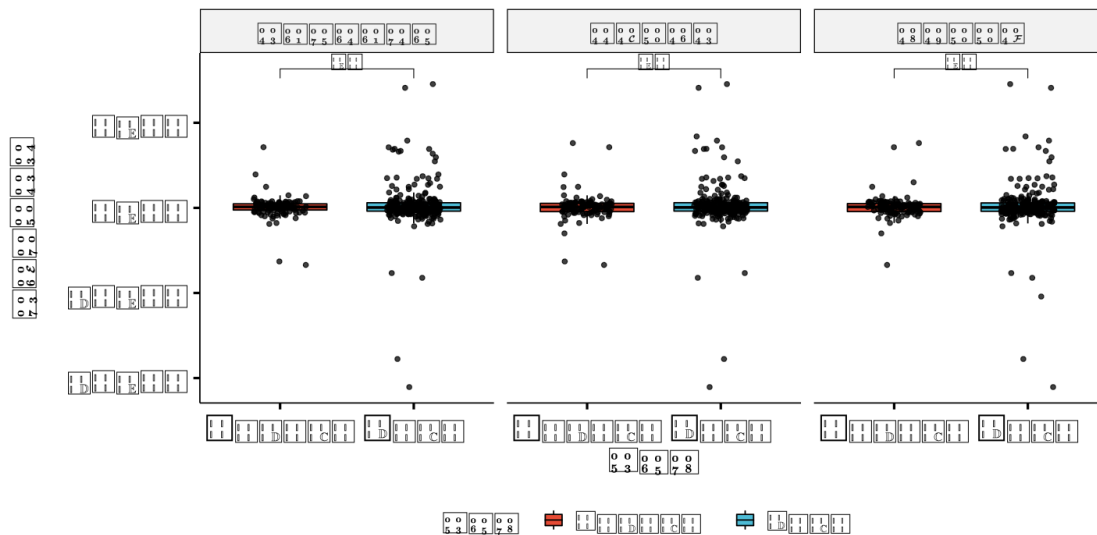
"Removed 3 rows containing missing values (geom_point)."



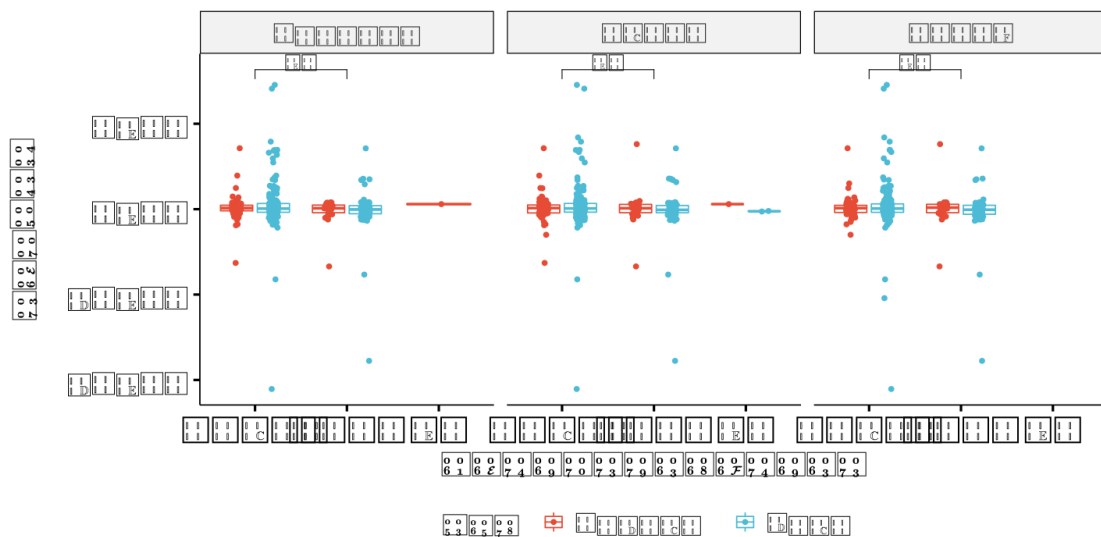
Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_boxplot)."
 Warning message:
 "Removed 3 rows containing non-finite values (stat_signif)."
 Warning message:
 "Removed 3 rows containing missing values (geom_point)."



Warning message:

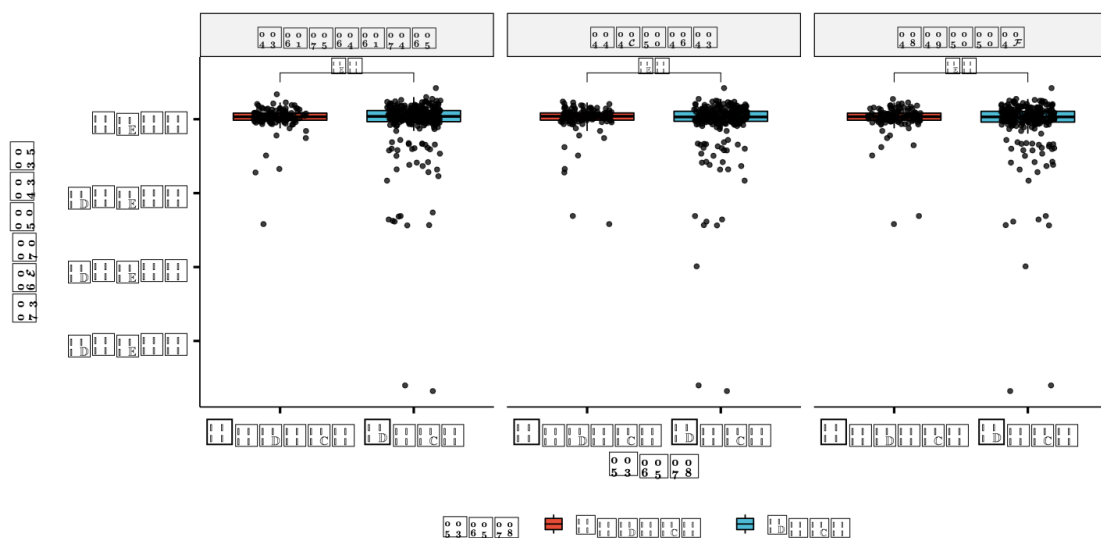
"Removed 3 rows containing non-finite values (stat_boxplot)."

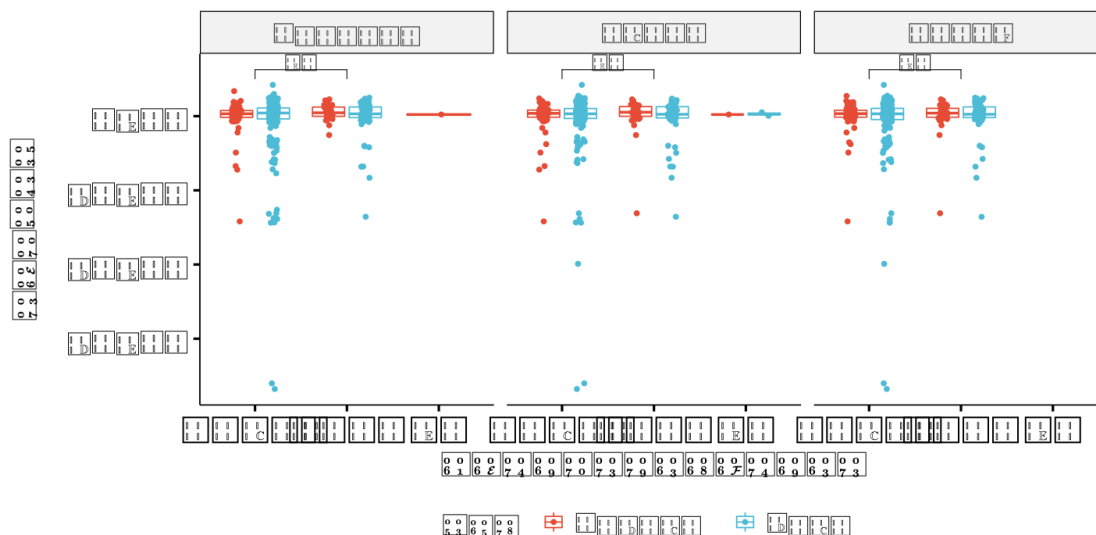
Warning message:

"Removed 3 rows containing non-finite values (stat_signif)."

Warning message:

"Removed 3 rows containing missing values (geom_point)."





1.3 Reproducibility Information

```
[4]: Sys.time()
proc.time()
options(width = 120)
sessioninfo::session_info()
```

```
[1] "2021-10-20 15:29:16 EDT"
```

```
user system elapsed
222.507 1.318 223.426
```

```
Session info
```

```
setting value
```

```
version R version 4.1.0 (2021-05-18)
```

```
os Arch Linux
```

```
system x86_64, linux-gnu
```

```
ui X11
```

```
language (EN)
```

```
collate en_US.UTF-8
```

```
ctype en_US.UTF-8
```

```
tz America/New_York
```

```
date 2021-10-20
```

```
Packages
```

package	* version	date	lib	source
abind	1.4-5	2016-07-21	[1]	CRAN (R 4.1.0)
assertthat	0.2.1	2019-03-21	[1]	CRAN (R 4.1.0)
backports	1.2.1	2020-12-09	[1]	CRAN (R 4.1.0)
base64enc	0.1-3	2015-07-28	[1]	CRAN (R 4.1.0)

broom	0.7.9	2021-07-27	[1]	CRAN	(R 4.1.0)
cachem	1.0.6	2021-08-19	[1]	CRAN	(R 4.1.0)
car	3.0-11	2021-06-27	[1]	CRAN	(R 4.1.0)
carData	3.0-4	2020-05-22	[1]	CRAN	(R 4.1.0)
cellranger	1.1.0	2016-07-27	[1]	CRAN	(R 4.1.0)
cli	3.0.1	2021-07-17	[1]	CRAN	(R 4.1.0)
colorspace	2.0-2	2021-06-24	[1]	CRAN	(R 4.1.0)
crayon	1.4.1	2021-02-08	[1]	CRAN	(R 4.1.0)
curl	4.3.2	2021-06-23	[1]	CRAN	(R 4.1.0)
data.table	1.14.2	2021-09-27	[1]	CRAN	(R 4.1.0)
DBI	1.1.1	2021-01-15	[1]	CRAN	(R 4.1.0)
dbplyr	2.1.1	2021-04-06	[1]	CRAN	(R 4.1.0)
digest	0.6.28	2021-09-23	[1]	CRAN	(R 4.1.0)
dplyr	* 1.0.7	2021-06-18	[1]	CRAN	(R 4.1.0)
ellipsis	0.3.2	2021-04-29	[1]	CRAN	(R 4.1.0)
evaluate	0.14	2019-05-28	[1]	CRAN	(R 4.1.0)
fansi	0.5.0	2021-05-25	[1]	CRAN	(R 4.1.0)
farver	2.1.0	2021-02-28	[1]	CRAN	(R 4.1.0)
fastmap	1.1.0	2021-01-25	[1]	CRAN	(R 4.1.0)
forcats	* 0.5.1	2021-01-27	[1]	CRAN	(R 4.1.0)
foreign	0.8-81	2020-12-22	[2]	CRAN	(R 4.1.0)
fs	1.5.0	2020-07-31	[1]	CRAN	(R 4.1.0)
generics	0.1.0	2020-10-31	[1]	CRAN	(R 4.1.0)
ggplot2	* 3.3.5	2021-06-25	[1]	CRAN	(R 4.1.0)
ggpubr	* 0.4.0	2020-06-27	[1]	CRAN	(R 4.1.0)
ggsci	2.9	2018-05-14	[1]	CRAN	(R 4.1.0)
ggsignif	0.6.3	2021-09-09	[1]	CRAN	(R 4.1.0)
glue	1.4.2	2020-08-27	[1]	CRAN	(R 4.1.0)
gtable	0.3.0	2019-03-25	[1]	CRAN	(R 4.1.0)
haven	2.4.3	2021-08-04	[1]	CRAN	(R 4.1.0)
hms	1.1.1	2021-09-26	[1]	CRAN	(R 4.1.0)
htmltools	0.5.2	2021-08-25	[1]	CRAN	(R 4.1.0)
httr	1.4.2	2020-07-20	[1]	CRAN	(R 4.1.0)
IRdisplay	1.0	2021-01-20	[1]	CRAN	(R 4.1.0)
IRkernel	1.2	2021-05-11	[1]	CRAN	(R 4.1.0)
jsonlite	1.7.2	2020-12-09	[1]	CRAN	(R 4.1.0)
labeling	0.4.2	2020-10-20	[1]	CRAN	(R 4.1.0)
lifecycle	1.0.1	2021-09-24	[1]	CRAN	(R 4.1.0)
lubridate	1.7.10	2021-02-26	[1]	CRAN	(R 4.1.0)
magrittr	2.0.1	2020-11-17	[1]	CRAN	(R 4.1.0)
memoise	2.0.0	2021-01-26	[1]	CRAN	(R 4.1.0)
modelr	0.1.8	2020-05-19	[1]	CRAN	(R 4.1.0)
munsell	0.5.0	2018-06-12	[1]	CRAN	(R 4.1.0)
openxlsx	4.2.4	2021-06-16	[1]	CRAN	(R 4.1.0)
pbdZMQ	0.3-5	2021-02-10	[1]	CRAN	(R 4.1.0)
pillar	1.6.3	2021-09-26	[1]	CRAN	(R 4.1.0)
pkgconfig	2.0.3	2019-09-22	[1]	CRAN	(R 4.1.0)
purrr	* 0.3.4	2020-04-17	[1]	CRAN	(R 4.1.0)

R6	2.5.1	2021-08-19	[1]	CRAN	(R 4.1.0)
Rcpp	1.0.7	2021-07-07	[1]	CRAN	(R 4.1.0)
readr	* 2.0.2	2021-09-27	[1]	CRAN	(R 4.1.0)
readxl	1.3.1	2019-03-13	[1]	CRAN	(R 4.1.0)
repr	* 1.1.3	2021-01-21	[1]	CRAN	(R 4.1.0)
reprex	2.0.1	2021-08-05	[1]	CRAN	(R 4.1.0)
rio	0.5.27	2021-06-21	[1]	CRAN	(R 4.1.0)
rlang	0.4.11	2021-04-30	[1]	CRAN	(R 4.1.0)
rstatix	0.7.0	2021-02-13	[1]	CRAN	(R 4.1.0)
rstudioapi	0.13	2020-11-12	[1]	CRAN	(R 4.1.0)
rvest	1.0.1	2021-07-26	[1]	CRAN	(R 4.1.0)
scales	1.1.1	2020-05-11	[1]	CRAN	(R 4.1.0)
sessioninfo	1.1.1	2018-11-05	[1]	CRAN	(R 4.1.0)
stringi	1.7.5	2021-10-04	[1]	CRAN	(R 4.1.0)
stringr	* 1.4.0	2019-02-10	[1]	CRAN	(R 4.1.0)
svglite	2.0.0	2021-02-20	[1]	CRAN	(R 4.1.0)
systemfonts	1.0.2	2021-05-11	[1]	CRAN	(R 4.1.0)
tibble	* 3.1.5	2021-09-30	[1]	CRAN	(R 4.1.0)
tidyr	* 1.1.4	2021-09-27	[1]	CRAN	(R 4.1.0)
tidyselect	1.1.1	2021-04-30	[1]	CRAN	(R 4.1.0)
tidyverse	* 1.3.1	2021-04-15	[1]	CRAN	(R 4.1.0)
tzdb	0.1.2	2021-07-20	[1]	CRAN	(R 4.1.0)
utf8	1.2.2	2021-07-24	[1]	CRAN	(R 4.1.0)
uuid	0.1-4	2020-02-26	[1]	CRAN	(R 4.1.0)
vctrs	0.3.8	2021-04-29	[1]	CRAN	(R 4.1.0)
withr	2.4.2	2021-04-18	[1]	CRAN	(R 4.1.0)
xml2	1.3.2	2020-04-23	[1]	CRAN	(R 4.1.0)
zip	2.2.0	2021-05-31	[1]	CRAN	(R 4.1.0)

[1] /home/jbenja13/R/x86_64-pc-linux-gnu-library/4.1

[2] /usr/lib/R/library