

Script 2c

Descriptive stats by pain status

Peter Kamerman

08 August 2020

Contents

| | |
|----------------------------------------------------------|---|
| Analysis notes | 1 |
| Import data | 2 |
| First look | 2 |
| Basic clean | 2 |
| Quick tabulation | 2 |
| Analysis data set for the period 0 to 48 weeks | 2 |
| Analysis | 3 |
| Prepare summary data for tabulation | 3 |
| Pain proportion | 3 |
| CD4 | 4 |
| Viral load | 5 |
| General health | 6 |
| Session information | 8 |

Analysis notes

Pain was defined as *pain in the last week* being ‘Yes’, and *pain at its worst* being > 0 .

These two measurements were then the “gatekeeper” measurements, such that the two measurements both had to be positive (‘Yes’ and ‘ > 0 ’, respectively) in order for there to be any entries for *site of pain* and *site of worst pain*. Were the data were incongruent (e.g., when there was no *pain in the last week* and *pain at its worst* = 0, but there were entries for *site of pain* and *site of worst pain*), then the *site of pain* and *site of worst pain* entries were marked as **incongruent**.

Data also were considered **incongruent** when *pain in the last week* = ‘Yes’, but *site of worst pain* = ‘None’.

Lastly, data were considered **incongruent** when *site of worst pain* was not listed as one of the pain locations for a given measurement week.

For analysis purposes, missing data in the *site of pain* columns were changed to ‘**No**’ (pain not present in the site). This approach was conservative, but we believed that the approach would have the least effect on the outcome, while still retaining as many participants as possible.

Import data

```
df <- read_rds('data-cleaned/data-ADVANCE.rds') %>%
  select(ranid, interval_name, pain_in_the_last_week,
         cd4_cells.ul, viral_load_cp.ml, general_health,
         any_missing, interval_numeric)
```

First look

```
head(df)
```

```
## # A tibble: 6 x 8
##   ranid interval_name pain_in_the_last_week cd4_cells.ul viral_load_cp.ml
##   <chr> <ord>         <chr>                <dbl>         <dbl>
## 1 01-0~ 0 weeks      No                642            641
## 2 01-0~ 12 weeks     No                NA              50
## 3 01-0~ 24 weeks     No                525            50
## 4 01-0~ 36 weeks     No                NA              50
## 5 01-0~ 48 weeks     No                668            50
## 6 01-0~ 0 weeks      No                241            3851
## # ... with 3 more variables: general_health <dbl>, any_missing <chr>,
## #   interval_numeric <dbl>
```

```
glimpse(df)
```

```
## Rows: 5,265
## Columns: 8
## $ ranid          <chr> "01-0001", "01-0001", "01-0001", "01-0001", "...
## $ interval_name  <ord> 0 weeks, 12 weeks, 24 weeks, 36 weeks, 48 wee...
## $ pain_in_the_last_week <chr> "No", "No", "No", "No", "No", "No", "Yes", "Y...
## $ cd4_cells.ul    <dbl> 642, NA, 525, NA, 668, 241, NA, 364, NA, 495,...
## $ viral_load_cp.ml <dbl> 641, 50, 50, 50, 50, 3851, 50, 50, 50, 50, 90...
## $ general_health  <dbl> 4, 4, 5, 5, 4, 3, 5, 3, 3, 3, 4, 5, 5, 5, 5, ...
## $ any_missing     <chr> "No", "No", "No", "No", "No", "No", "No", "No...
## $ interval_numeric <dbl> 0, 12, 24, 36, 48, 0, 12, 24, 36, 48, 0, 12, ...
```

Basic clean

```
# Clean and process data
df %<>%
  filter(any_missing == 'No') %>%
  select(-any_missing)
```

Quick tabulation

Analysis data set for the period 0 to 48 weeks

```
# Tabulate data
xtabs(~interval_name, data = df)
```

```
## interval_name
## 0 weeks 12 weeks 24 weeks 36 weeks 48 weeks
##      787      787      787      787      787
```

Analysis

Prepare summary data for tabulation

```
# Pain in the last week
df_pain <- as.data.frame(xtabs(~interval_name + pain_in_the_last_week,
                             data = df)) %>%
  rename(count = Freq) %>%
  group_by(interval_name) %>%
  mutate(total = sum(count)) %>%
  mutate(proportion = round(count / total, 3)) %>%
  filter(pain_in_the_last_week == 'Yes')

# CD4 (by pain status)
df_cd4 <- df %>%
  group_by(interval_name) %>%
  summarise(missing = sum(is.na(cd4_cells.ul)),
            median = median(cd4_cells.ul, na.rm = TRUE),
            Q25 = quantile(cd4_cells.ul, probs = 0.25, na.rm = TRUE),
            Q75 = quantile(cd4_cells.ul, probs = 0.75, na.rm = TRUE),
            min = min(cd4_cells.ul, na.rm = TRUE),
            max = max(cd4_cells.ul, na.rm = TRUE))

df_cd4 <- df_cd4[c(1, 3, 5), ] # weeks 0, 12, 48

# Viral load (by pain status)
df_vl <- df %>%
  group_by(interval_name, pain_in_the_last_week) %>%
  mutate(viral_load_cp.ml = ifelse(viral_load_cp.ml < 50,
                                   yes = 50,
                                   no = viral_load_cp.ml)) %>%
  summarise(missing = sum(is.na(viral_load_cp.ml)),
            median = median(viral_load_cp.ml, na.rm = TRUE),
            Q25 = quantile(viral_load_cp.ml, probs = 0.25, na.rm = TRUE),
            Q75 = quantile(viral_load_cp.ml, probs = 0.75, na.rm = TRUE),
            min = min(viral_load_cp.ml, na.rm = TRUE),
            max = max(viral_load_cp.ml, na.rm = TRUE))

# general health (by pain status)
df_gh <- df %>%
  group_by(interval_name, pain_in_the_last_week) %>%
  summarise(missing = sum(is.na(general_health)),
            median = median(general_health, na.rm = TRUE),
            Q25 = quantile(general_health, probs = 0.25, na.rm = TRUE),
            Q75 = quantile(general_health, probs = 0.75, na.rm = TRUE),
            min = min(general_health, na.rm = TRUE),
            max = max(general_health, na.rm = TRUE))
```

Pain proportion

```
# Pain
df_pain %>%
  kable(caption = 'Proportion with pain in the last week')
```

Table 1: Proportion with pain in the last week

| interval_name | pain_in_the_last_week | count | total | proportion |
|---------------|-----------------------|-------|-------|------------|
| 0 weeks | Yes | 151 | 787 | 0.192 |
| 12 weeks | Yes | 143 | 787 | 0.182 |
| 24 weeks | Yes | 127 | 787 | 0.161 |
| 36 weeks | Yes | 114 | 787 | 0.145 |
| 48 weeks | Yes | 96 | 787 | 0.122 |

CD4

```
# Tabulate CD4 (by pain status)
df_cd4 %>%
  kable(caption = '5-number summary of CD4 T-cell count (cells/ul) by pain status')
```

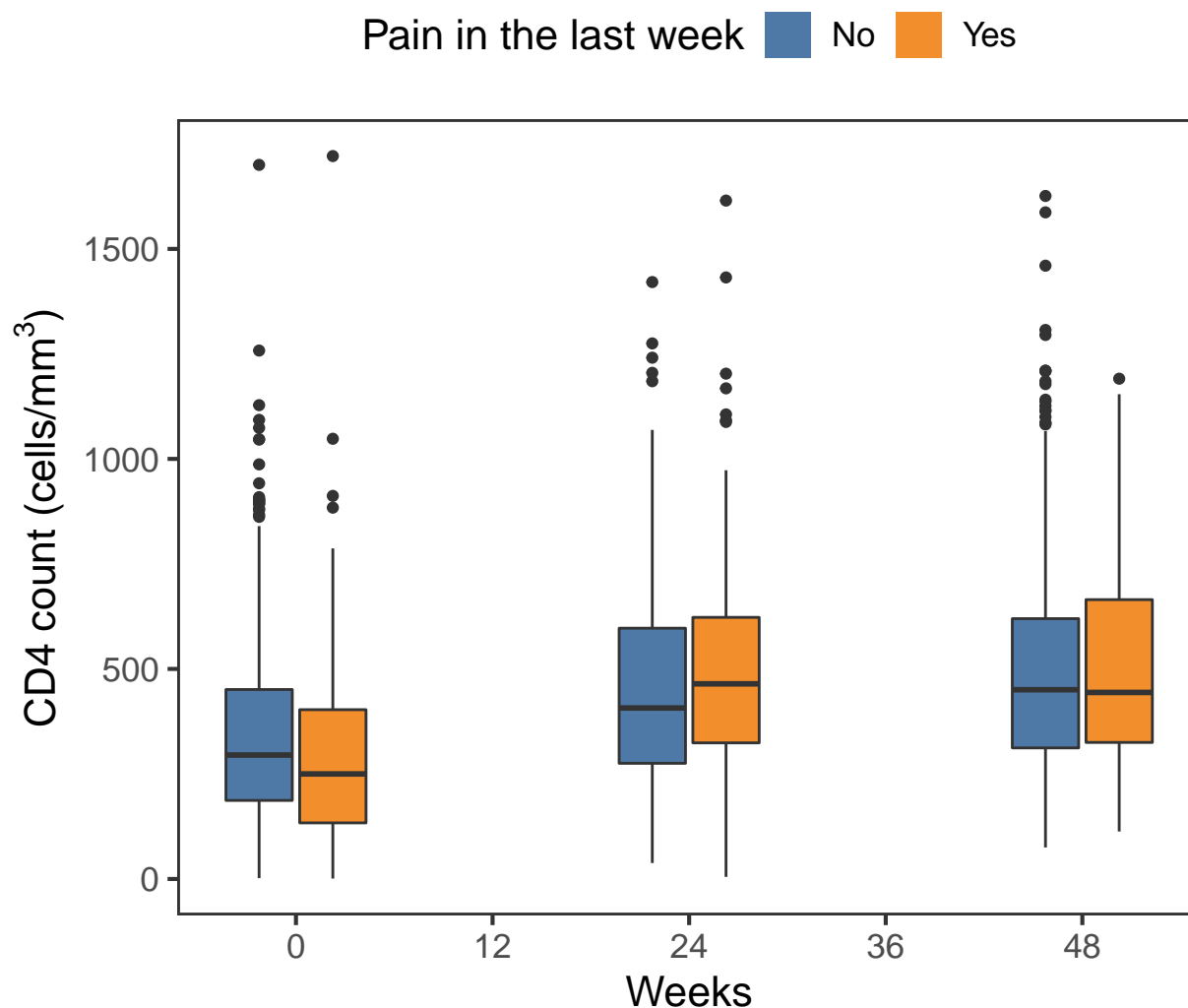
Table 2: 5-number summary of CD4 T-cell count (cells/ul) by pain status

| interval_name | missing | median | Q25 | Q75 | min | max |
|---------------|---------|--------|--------|--------|-----|------|
| 0 weeks | 0 | 290 | 173.50 | 441.50 | 1 | 1721 |
| 24 weeks | 11 | 411 | 277.75 | 604.25 | 5 | 1615 |
| 48 weeks | 16 | 450 | 313.50 | 629.00 | 75 | 1626 |

```
# Plot CD4 (by pain status)
p_cd4 <- df %>%
  select(interval_name, interval_numeric,
         pain_in_the_last_week, cd4_cells.ul) %>%
  mutate(cd4_cells.ul = ifelse(interval_numeric == 12 | interval_numeric == 36,
                              yes = NA,
                              no = cd4_cells.ul)) %>%

  ggplot(data = .) +
  aes(y = cd4_cells.ul,
      x = interval_name,
      fill = pain_in_the_last_week) +
  geom_boxplot(key_glyph = "rect") +
  labs(title = 'CD4 T-cell count',
       y = expression('CD4 count (cells/mm'^3)'),
       x = 'Weeks') +
  scale_x_discrete(labels = c('0', '12', '24', '36', '48')) +
  scale_fill_tableau(name = 'Pain in the last week') +
  theme(legend.position = 'top',
        panel.grid = element_blank()); p_cd4
```

CD4 T-cell count



Viral load

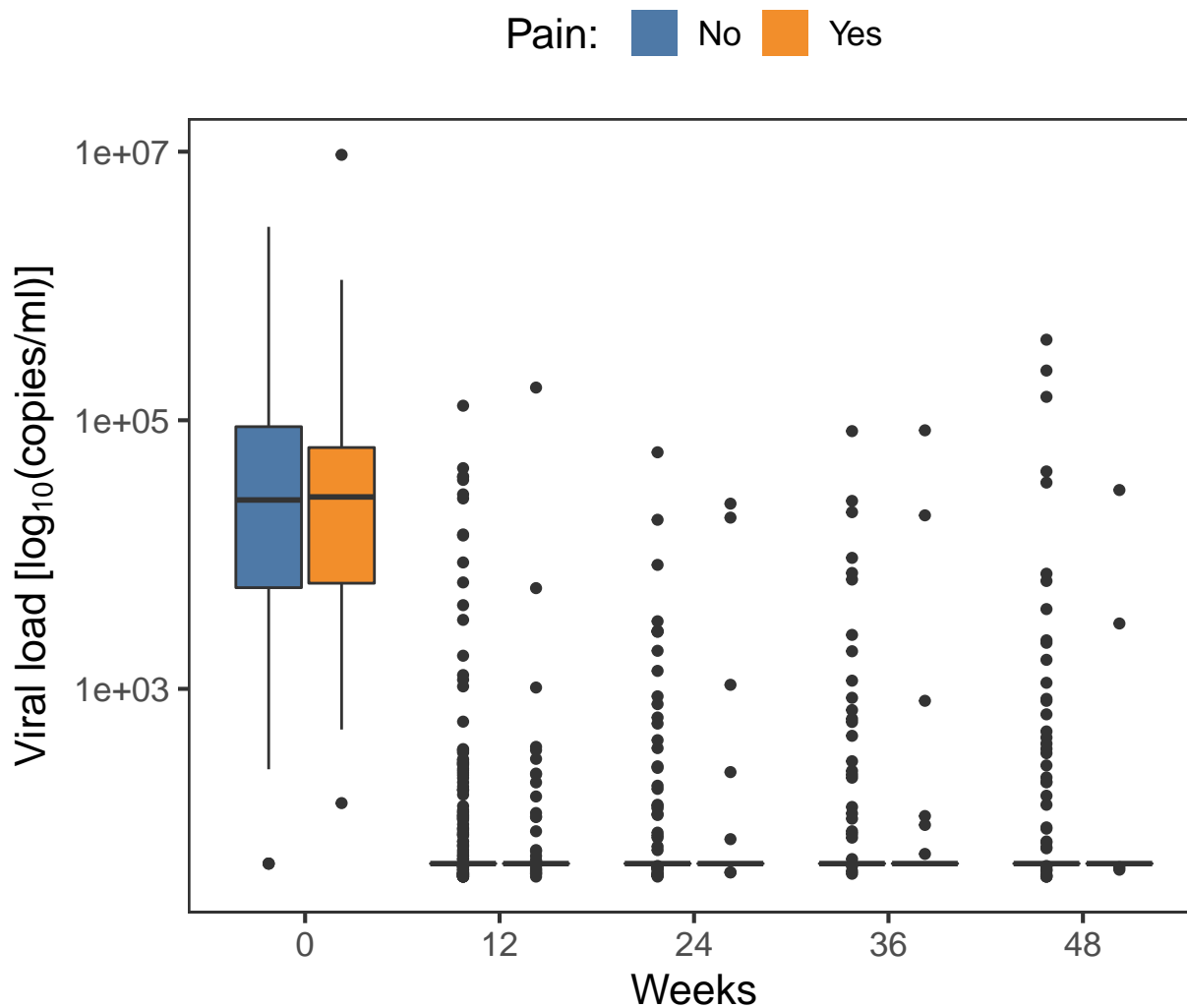
```
# Tabulate viral load (by pain status)
df_vl %>%
  kable(caption = '5-number summary of viral load (copies/ml) by pain status')
```

Table 3: 5-number summary of viral load (copies/ml) by pain status

| interval_name | pain_in_the_last_week | missing | median | Q25 | Q75 | min | max |
|---------------|-----------------------|---------|---------|---------|---------|-----|---------|
| 0 weeks | No | 0 | 25495.5 | 5667.25 | 89600.5 | 50 | 2757298 |
| 0 weeks | Yes | 0 | 26859.0 | 6123.00 | 62706.5 | 141 | 9475772 |
| 12 weeks | No | 2 | 50.0 | 50.00 | 50.0 | 50 | 128334 |
| 12 weeks | Yes | 1 | 50.0 | 50.00 | 50.0 | 50 | 175168 |
| 24 weeks | No | 2 | 50.0 | 50.00 | 50.0 | 50 | 57754 |
| 24 weeks | Yes | 0 | 50.0 | 50.00 | 50.0 | 50 | 23938 |
| 36 weeks | No | 8 | 50.0 | 50.00 | 50.0 | 50 | 83020 |
| 36 weeks | Yes | 1 | 50.0 | 50.00 | 50.0 | 50 | 84167 |
| 48 weeks | No | 6 | 50.0 | 50.00 | 50.0 | 50 | 397926 |
| 48 weeks | Yes | 3 | 50.0 | 50.00 | 50.0 | 50 | 30237 |

```
# Plot viral load (by pain status)
p_vl <- df %>%
  select(interval_name, pain_in_the_last_week, viral_load_cp.ml) %>%
  ggplot(data = .) +
  aes(y = viral_load_cp.ml,
      x = interval_name,
      fill = pain_in_the_last_week) +
  geom_boxplot(key_glyph = "rect") +
  labs(title = 'HIV-1 RNA viral load ',
       y = expression('Viral load [log'[10]*'(copies/ml)']),
       x = 'Weeks') +
  scale_x_discrete(labels = c('0', '12', '24', '36', '48')) +
  scale_y_log10() +
  scale_fill_tableau(name = 'Pain: ') +
  theme(legend.position = 'top',
        panel.grid = element_blank()); p_vl
```

HIV-1 RNA viral load



General health

```
# Tabulate general health (by pain status)
df_gh %>%
```

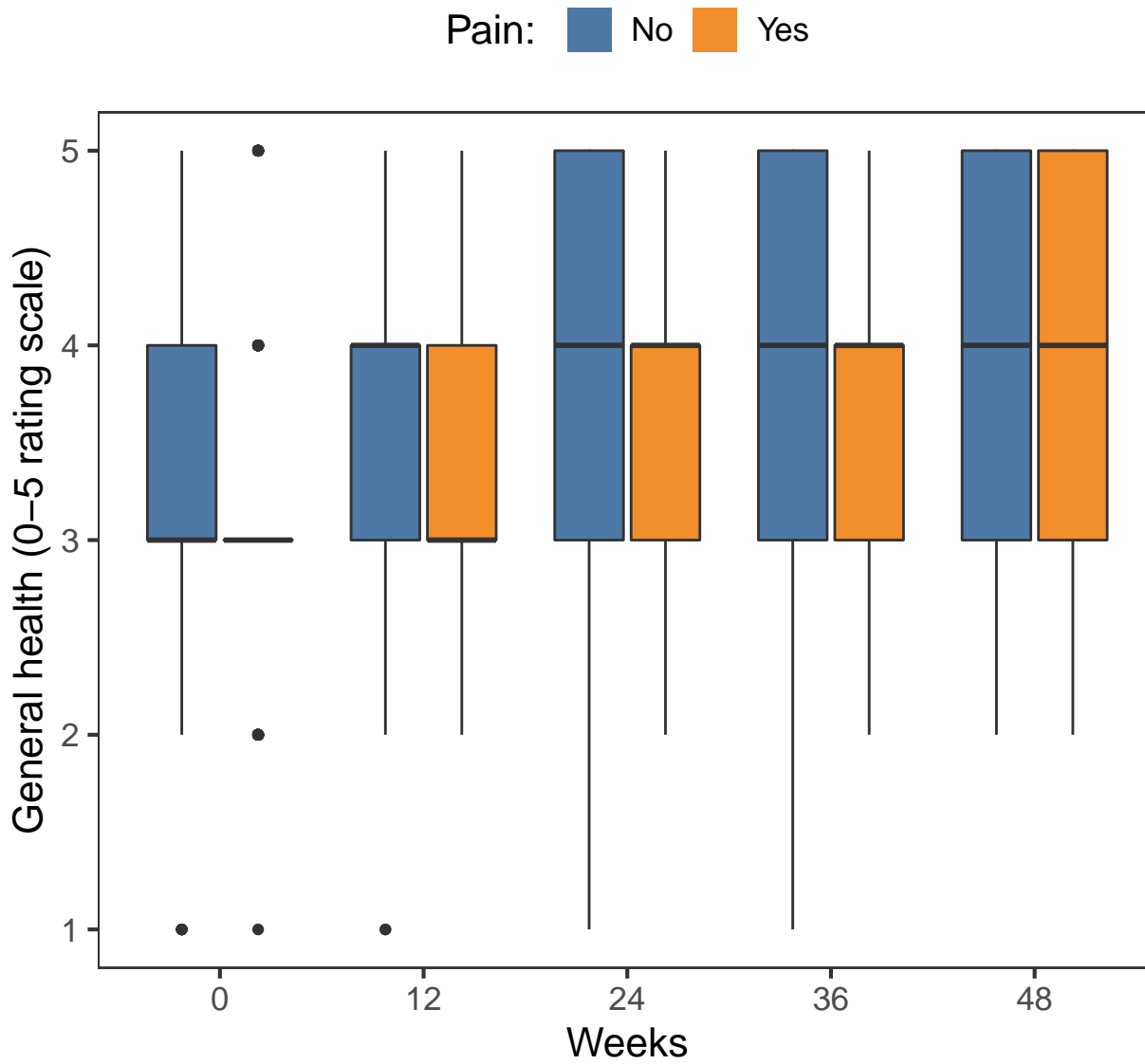
```
kable(caption = '5-number summary of general health (0-5 rating scale) by pain status')
```

Table 4: 5-number summary of general health (0-5 rating scale) by pain status

| interval_name | pain_in_the_last_week | missing | median | Q25 | Q75 | min | max |
|---------------|-----------------------|---------|--------|-----|-----|-----|-----|
| 0 weeks | No | 2 | 3 | 3 | 4 | 1 | 5 |
| 0 weeks | Yes | 2 | 3 | 3 | 3 | 1 | 5 |
| 12 weeks | No | 0 | 4 | 3 | 4 | 1 | 5 |
| 12 weeks | Yes | 1 | 3 | 3 | 4 | 2 | 5 |
| 24 weeks | No | 0 | 4 | 3 | 5 | 1 | 5 |
| 24 weeks | Yes | 0 | 4 | 3 | 4 | 2 | 5 |
| 36 weeks | No | 1 | 4 | 3 | 5 | 1 | 5 |
| 36 weeks | Yes | 0 | 4 | 3 | 4 | 2 | 5 |
| 48 weeks | No | 2 | 4 | 3 | 5 | 2 | 5 |
| 48 weeks | Yes | 0 | 4 | 3 | 5 | 2 | 5 |

```
# Plot general health (by pain status)
p_gh <- df %>%
  select(interval_name, pain_in_the_last_week, general_health) %>%
  ggplot(data = .) +
  aes(y = general_health,
       x = interval_name,
       fill = pain_in_the_last_week) +
  geom_boxplot(key_glyph = "rect") +
  labs(title = 'General health',
       y = 'General health (0-5 rating scale)',
       x = 'Weeks') +
  scale_x_discrete(labels = c('0', '12', '24', '36', '48')) +
  scale_fill_tableau(name = 'Pain: ') +
  theme(legend.position = 'top',
        panel.grid = element_blank()); p_gh
```

General health



Session information

```
sessionInfo()
```

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.5
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
```



```
##
## other attached packages:
## [1] knitr_1.29      patchwork_1.0.1 ggthemes_4.2.0  magrittr_1.5
## [5] forcats_0.5.0   stringr_1.4.0   dplyr_1.0.0     purrr_0.3.4
## [9] readr_1.3.1     tidyr_1.1.0     tibble_3.0.1    ggplot2_3.3.2
## [13] tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.1.0 xfun_0.15      haven_2.3.1    lattice_0.20-41
## [5] colorspace_1.4-1 vctrs_0.3.1    generics_0.0.2 htmltools_0.5.0
## [9] yaml_2.2.1       utf8_1.1.4     blob_1.2.1     rlang_0.4.6
## [13] pillar_1.4.4     glue_1.4.1     withr_2.2.0    DBI_1.1.0
## [17] dbplyr_1.4.4     modelr_0.1.8   readxl_1.3.1   lifecycle_0.2.0
## [21] munsell_0.5.0    gtable_0.3.0   cellranger_1.1.0 rvest_0.3.5
## [25] evaluate_0.14    labeling_0.3    fansi_0.4.1    highr_0.8
## [29] broom_0.5.6      Rcpp_1.0.4.6   scales_1.1.1   backports_1.1.8
## [33] jsonlite_1.6.1   farver_2.0.3   fs_1.4.1       hms_0.5.3
## [37] digest_0.6.25    stringi_1.4.6  grid_4.0.2     cli_2.0.2
## [41] tools_4.0.2      crayon_1.3.4   pkgconfig_2.0.3 ellipsis_0.3.1
## [45] xml2_1.3.2       reprex_0.3.0   lubridate_1.7.9 assertthat_0.2.1
## [49] rmarkdown_2.3    httr_1.4.1     rstudioapi_0.11 R6_2.4.1
## [53] nlme_3.1-148     compiler_4.0.2
```