

mrot diapause - lifespan

Karen M. Kapheim

2025-02-04

This is analysis of the lifespan data for *Megachile rotundata* males & female that either did or did not diapause. These samples were provided as incubating cocoons from Lindsay McCabe (USDA PIRU).

We changed out the cage style after we had some death in the original cage style. That is why there is a 'old_cage' variable. We want to see if the results change when we include bees that spent time in the original cage design.

There were also some bees that had some discrepancies, missing data, or other issues, so made a column that excludes them.

Thorax width was measured by Addie Lloyd in the Leica EZ Suite software.

The original (locked) data sheet and pictures are in the folder 'mrot_diapause_lifespan_immunity' on the KapheimLab google drive.

The original hard copy of the data is in Kapheim's office.

Setup

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(survival)
```

```
## Warning: package 'survival' was built under R version 4.3.3
```

```
library(survminer)
```

```
## Warning: package 'survminer' was built under R version 4.3.3
```

```
## Loading required package: ggpubr
##
## Attaching package: 'survminer'
##
## The following object is masked from 'package:survival':
##
##     myeloma
```

```
library(ggfortify)
```

Data

```
lifespan <- read_csv("MROT_lifespan_thorax_WorkingCopy.csv",
  col_select = c(1:10),
  col_names = TRUE,
  col_types = cols(
    Bee_ID = col_character(),
    Diapause_NonDiapause = col_character(),
    Sex = col_character(),
    Date_emerged = col_date("%m/%d/%y"),
    Date_dead = col_date("%m/%d/%y"),
    Date_frozen = col_date("%m/%d/%y"),
    Thorax_length_mm = col_double(),
    Notes = col_character(),
    use_lifespan = col_character(),
    old_cage_design = col_character()
  ))
```

```
## New names:
## * '' -> '...11'
## * '' -> '...12'
## * '' -> '...13'
## * '' -> '...14'
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
problems(lifespan)
```

Make some new variables We need a censor variable and a survival time.

```
lifespan <- lifespan %>%
  filter(!Bee_ID == "NA") %>%
  mutate(days_surv = as.numeric(Date_dead - Date_emerged),
    status = as.logical("TRUE"),
    combo = paste(Diapause_NonDiapause, Sex, sep = ""))
table(lifespan$combo)
```

```
##
##  DF  DM  NDF  NDM
##  34  24  59   43
```

```
lifespan$Diapause_NonDiapause = relevel(as.factor(lifespan$Diapause_NonDiapause),
                                         ref = "ND")
```

Filter datasets Only those without weird issues

```
lifespan.r <- lifespan %>%
  filter(use_lifespan == "Y")
table(lifespan.r$combo)
```

```
##
##  DF  DM  NDF  NDM
##  34  22  56   39
```

Only those that spent whole life in new cage design

```
lifespan.r2 <- lifespan.r %>%
  filter(old_cage_design == "N")
table(lifespan.r2$combo)
```

```
##
##  DF  DM  NDF  NDM
##  26   4  39   20
```

This eliminates most of the Diapause males. Maybe need to exclude males.

Only females

```
lifespan.r2F <- lifespan.r2 %>%
  filter(Sex == "F")
table(lifespan.r2F$combo)
```

```
##
##  DF  NDF
##  26   39
```

Survival Analysis

```
km <- with(lifespan, Surv(days_surv, status))
km.r <- with(lifespan.r, Surv(days_surv, status))
km.r2 <- with(lifespan.r2, Surv(days_surv, status))
km.r2F <- with(lifespan.r2F, Surv(days_surv, status))
```

Make a survival object

```

km_fit <- survfit(Surv(days_surv, status) ~ combo, data=lifespan)
km_fit.r.combo <- survfit(Surv(days_surv, status) ~ combo, data=lifespan.r)
km_fit.r2.combo <- survfit(Surv(days_surv, status) ~ combo, data=lifespan.r2)
km_fit.r.DN <- survfit(Surv(days_surv, status) ~ Diapause_NonDiapause,
  data=lifespan.r)
km_fit.r2.DN <- survfit(Surv(days_surv, status) ~ Diapause_NonDiapause,
  data=lifespan.r2)
km_fit.r2F.DN <- survfit(Surv(days_surv, status) ~ Diapause_NonDiapause,
  data=lifespan.r2F)
km_fit.r.sex <- survfit(Surv(days_surv, status) ~ Sex, data=lifespan.r)
km_fit.r2.sex <- survfit(Surv(days_surv, status) ~ Sex, data=lifespan.r2)

```

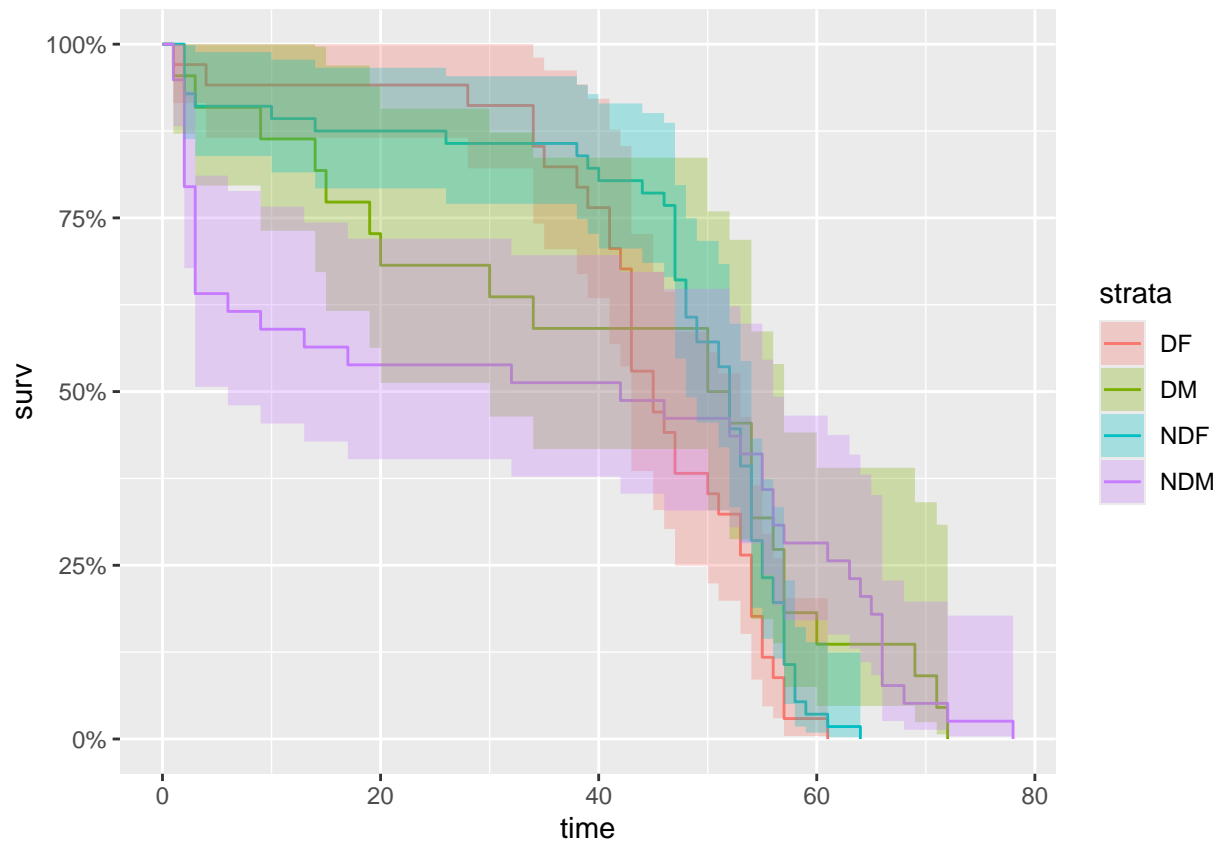
Get the stats

Plots All others

```

# combo
autoplot(km_fit.r.combo)

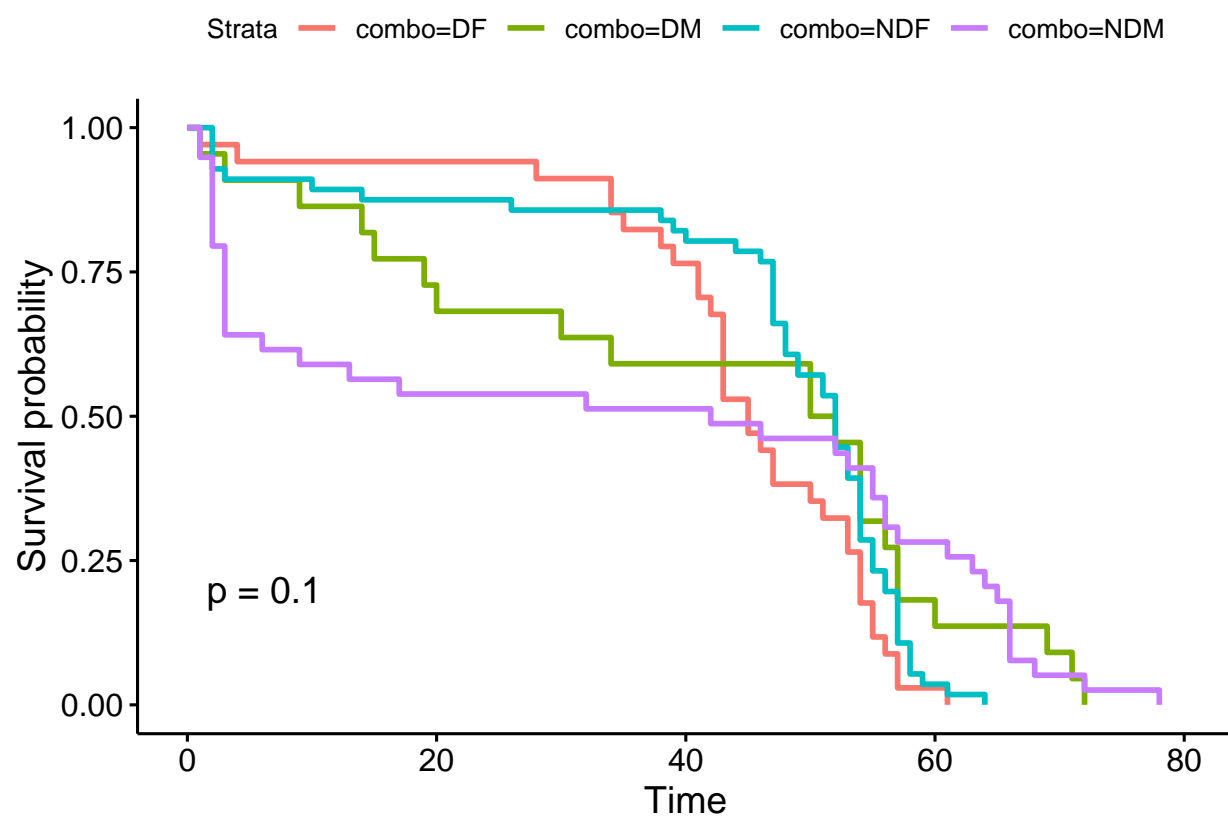
```



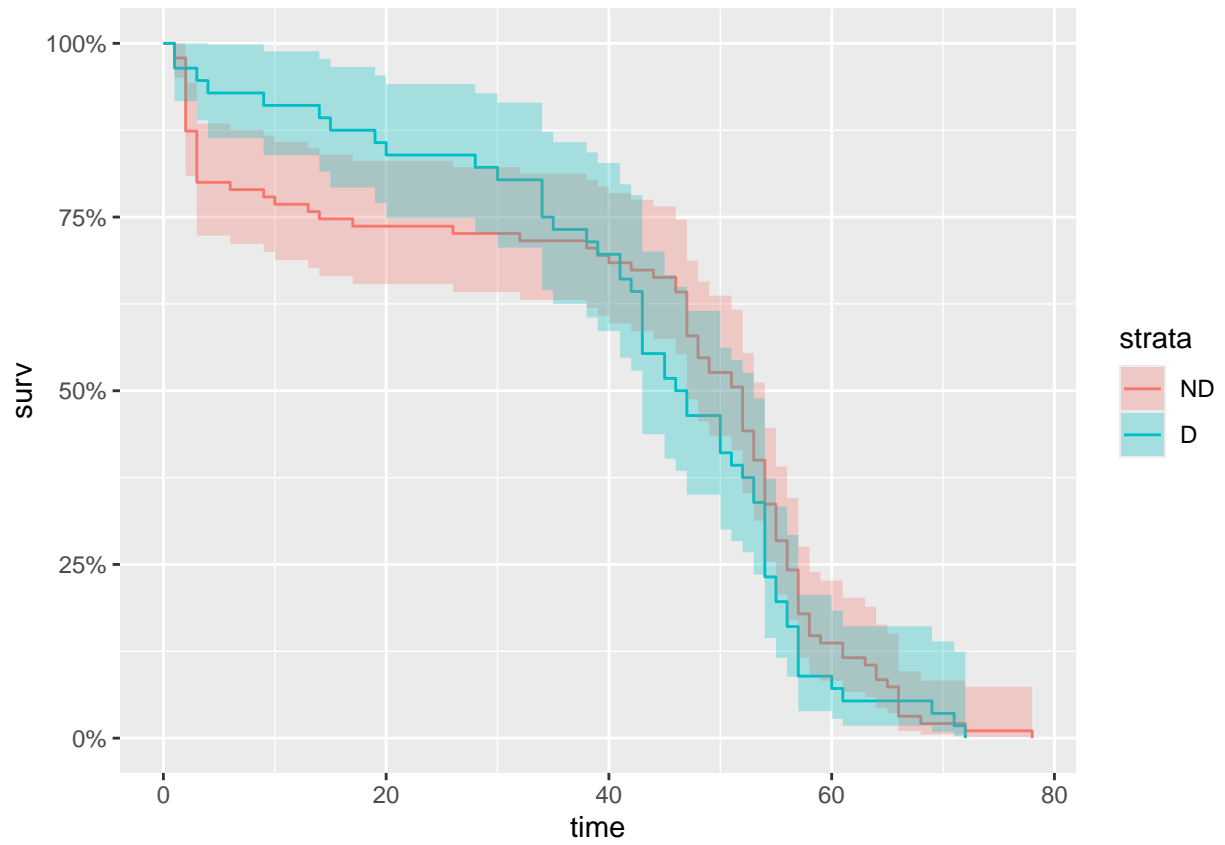
```

ggsurvplot(km_fit.r.combo, data = lifespan.r, pval = TRUE)

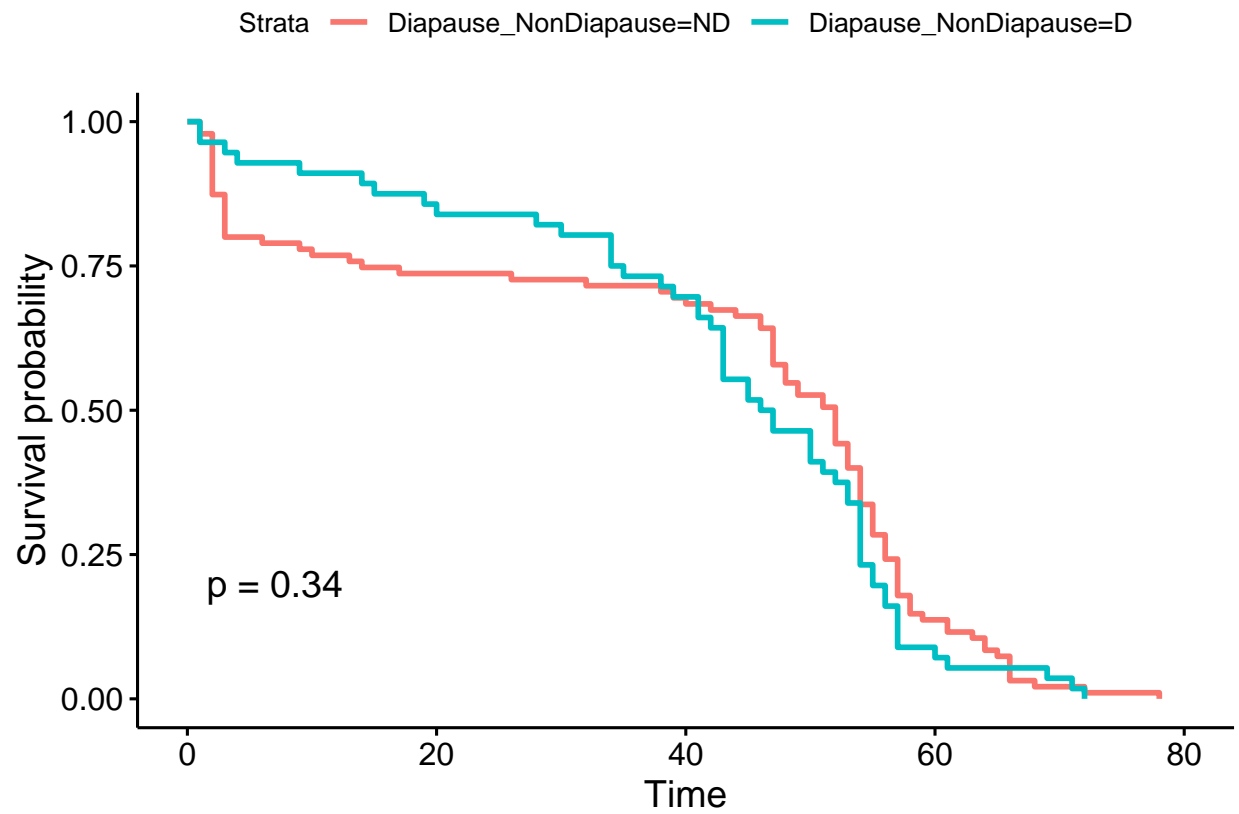
```



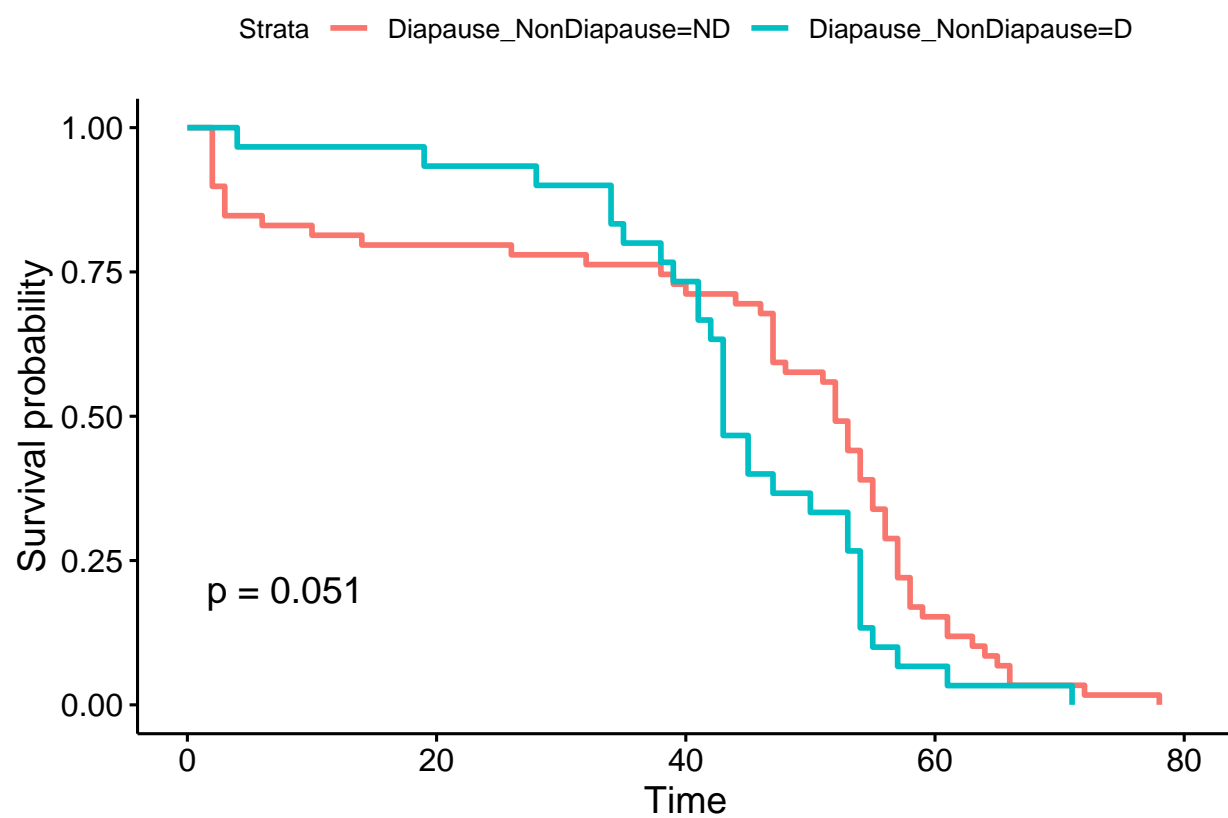
```
# diapause  
autoplot(km_fit.r.DN)
```



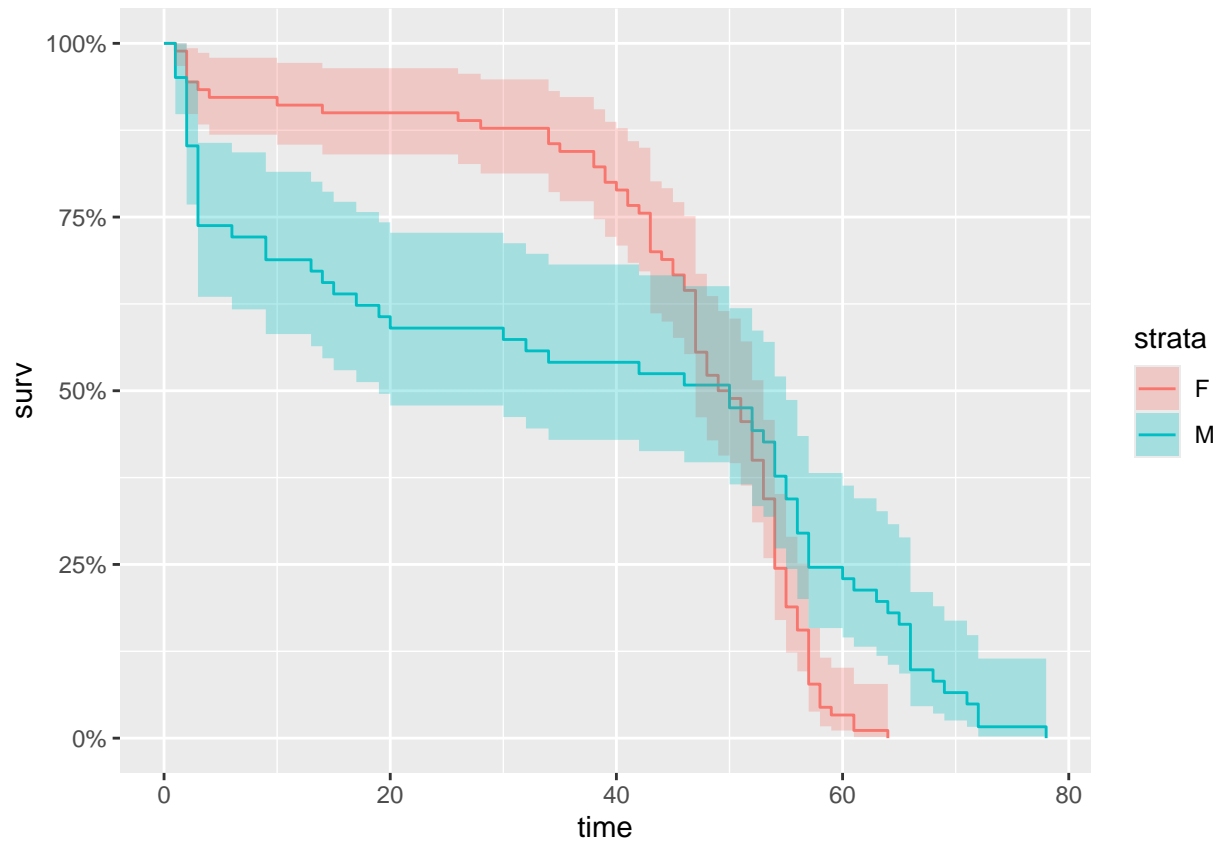
```
ggsurvplot(km_fit.r.DN, data = lifespan.r, pval = TRUE)
```



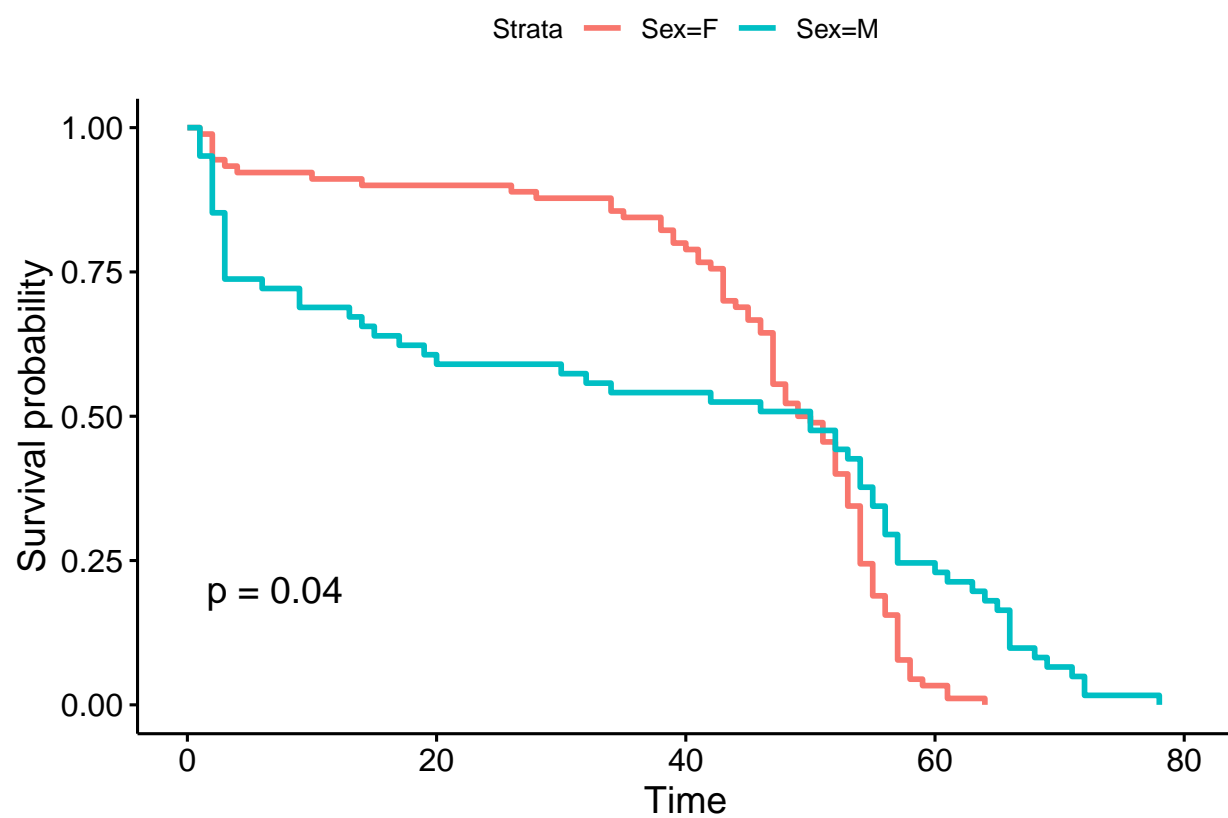
```
ggsurvplot(km_fit.r2.DN, data = lifespan.r2, pval = TRUE)
```



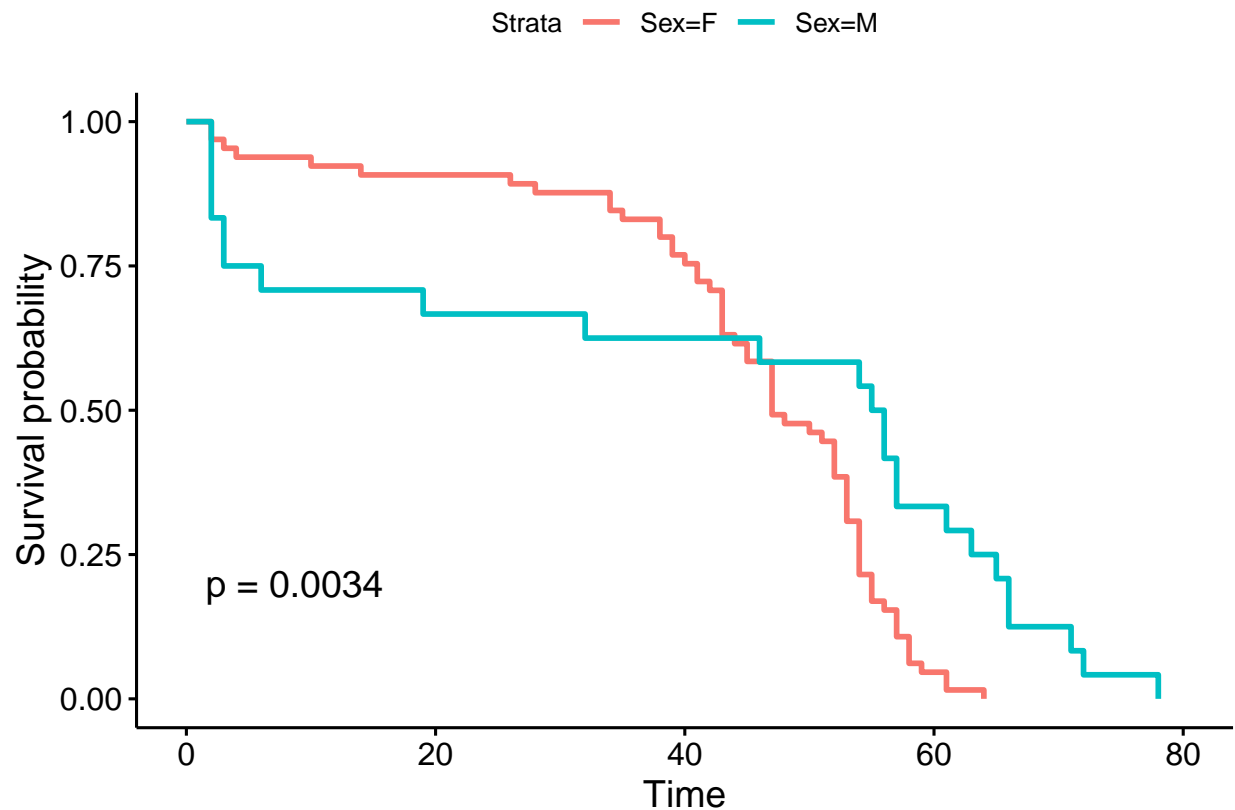
```
# sex  
autoplot(km_fit.r.sex)
```

```
ggsurvplot(km_fit.r.sex, data = lifespan.r, pval = TRUE)
```



```
ggsurvplot(km_fit.r2.sex, data = lifespan.r2, pval = TRUE)
```



Only females without issues

Summary data Min and max lifespan

```
lifespan.r2F %>%
  group_by(Diapause_NonDiapause) %>%
  summarise(min = min(days_surv),
            max = max(days_surv),
            mean = mean(days_surv),
            median = median(days_surv),
            sd = sd(days_surv),
            se = sd(days_surv)/sqrt(n()),
            n = n())
```

```
## # A tibble: 2 x 8
##   Diapause_NonDiapause   min   max  mean median    sd    se    n
##   <fct>             <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <int>
## 1 ND                 2     64  45.5    52  16.9  2.71    39
## 2 D                  4     61  43.2    43  11.2  2.19    26
```

```
lifespan.r2F %>%
  summarise(min = min(days_surv),
            max = max(days_surv),
            mean = mean(days_surv),
```

```

median = median(days_surv),
sd = sd(days_surv),
se = sd(days_surv)/sqrt(n()),
n = n())

```

```

## # A tibble: 1 x 7
##   min    max  mean median    sd    se     n
##   <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <int>
## 1     2    64  44.6    47  14.8  1.84    65

```

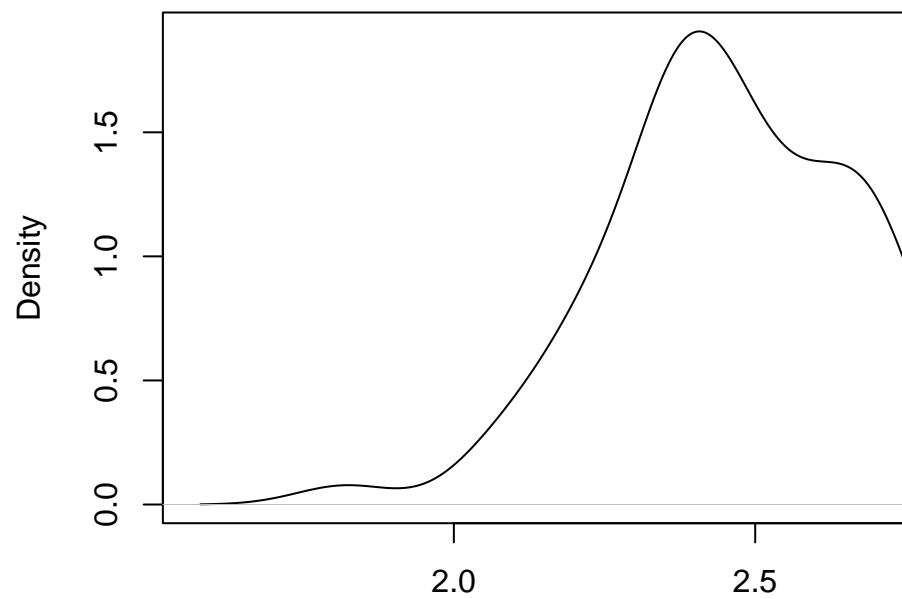
Min, Max, and Mean are similar, but median lifespan is 9 d longer (21%) for ND.

```

plot(density(lifespan.r2F$Thorax_length_mm, na.rm = TRUE))

```

density(x = lifespan.r2F\$Thorax_length_mm



N = 65 Bandwidth = 0.08025

Body size difference between D and ND

```

shapiro.test(lifespan.r2F$Thorax_length_mm)

```

```

##
##  Shapiro-Wilk normality test
##
## data:  lifespan.r2F$Thorax_length_mm
## W = 0.98126, p-value = 0.429

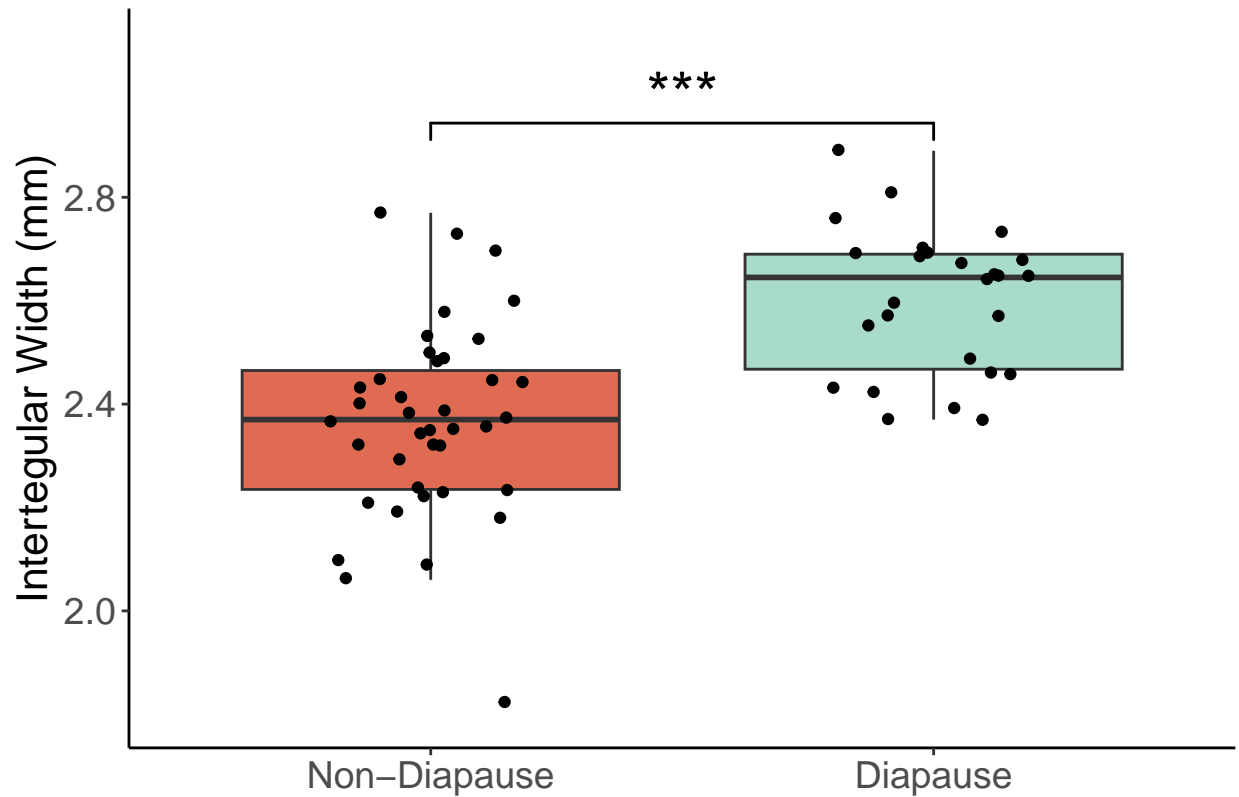
```

```
lifespan.r2F %>%
  group_by(Diapause_NonDiapause) %>%
  summarise(min = min(Thorax_length_mm),
            max = max(Thorax_length_mm),
            mean = mean(Thorax_length_mm),
            median = median(Thorax_length_mm),
            sd = sd(Thorax_length_mm),
            se = sd(Thorax_length_mm)/sqrt(n()),
            n = n())
```

```
## # A tibble: 2 x 8
##   Diapause_NonDiapause   min    max  mean median    sd    se    n
##   <fct>                <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl> <int>
## 1 ND                  1.82  2.77  2.36   2.37 0.189 0.0302   39
## 2 D                   2.37  2.89  2.60   2.64 0.140 0.0275   26
```

```
#Plot body size D vs ND
```

```
ggplot(lifespan.r2F, aes(x = Diapause_NonDiapause, y = Thorax_length_mm, fill = Diapause_NonDiapause, ))
  geom_boxplot(alpha = 0.8, outlier.color=NA) +
  geom_jitter(position = position_jitter(width = 0.20)) +
  geom_signif(data = lifespan.r2F, stat = "signif", position = "identity",
              comparisons = list(c("D", "ND")), map_signif_level = TRUE,
              annotations = "***", textsize = 8) +
  coord_cartesian(ylim = c(1.8, 3.1)) +
  theme_classic() +
  scale_fill_manual(values = c("#D84727", "#94D2BD")) +
  theme(legend.position = "none",
        axis.title=element_text(size=16), axis.text=element_text(size=14)) +
  xlab("") +
  scale_x_discrete(labels = c("Non-Diapause", "Diapause")) +
  ylab("Intertegular Width (mm)")
```



```
ggsave("body_size_thorax_lifespan.png", width = 7, height = 5)
```

```
model <- lm(Thorax_length_mm ~ Diapause_NonDiapause, data = lifespan.r2F)
summary(model)
```

```
##
## Call:
## lm(formula = Thorax_length_mm ~ Diapause_NonDiapause, data = lifespan.r2F)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.54462 -0.13462  0.00538  0.09077  0.40538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.36462    0.02739   86.320 < 2e-16 ***
## Diapause_NonDiapauseD  0.23462    0.04331    5.417 1.01e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1711 on 63 degrees of freedom
## Multiple R-squared:  0.3177, Adjusted R-squared:  0.3069
## F-statistic: 29.34 on 1 and 63 DF, p-value: 1.013e-06
```

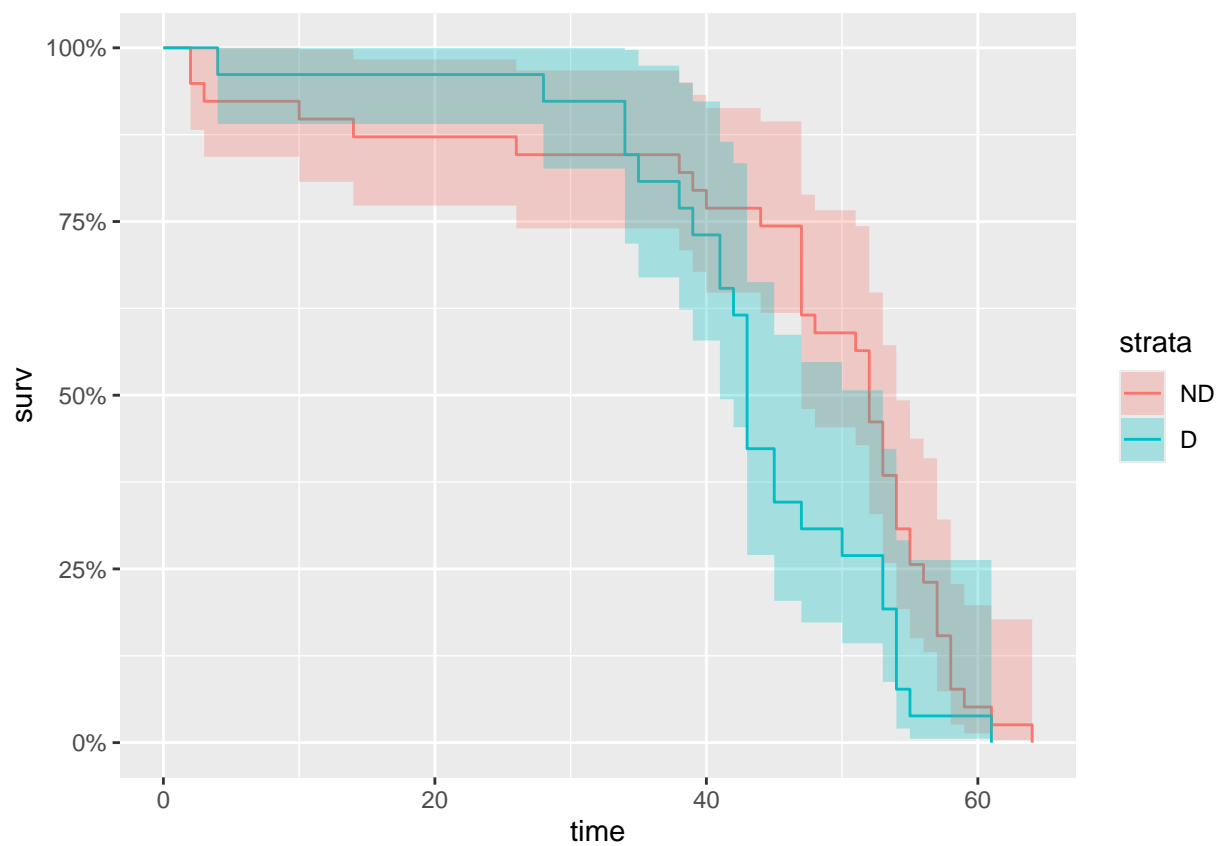
```
print(km_fit.r2F.DN)
```

Summary of the survival curve

```
## Call: survfit(formula = Surv(days_surv, status) ~ Diapause_NonDiapause,  
##      data = lifespan.r2F)  
##  
##              n events median 0.95LCL 0.95UCL  
## Diapause_NonDiapause=ND 39      39      52      47      54  
## Diapause_NonDiapause=D  26      26      43      41      53
```

Plot

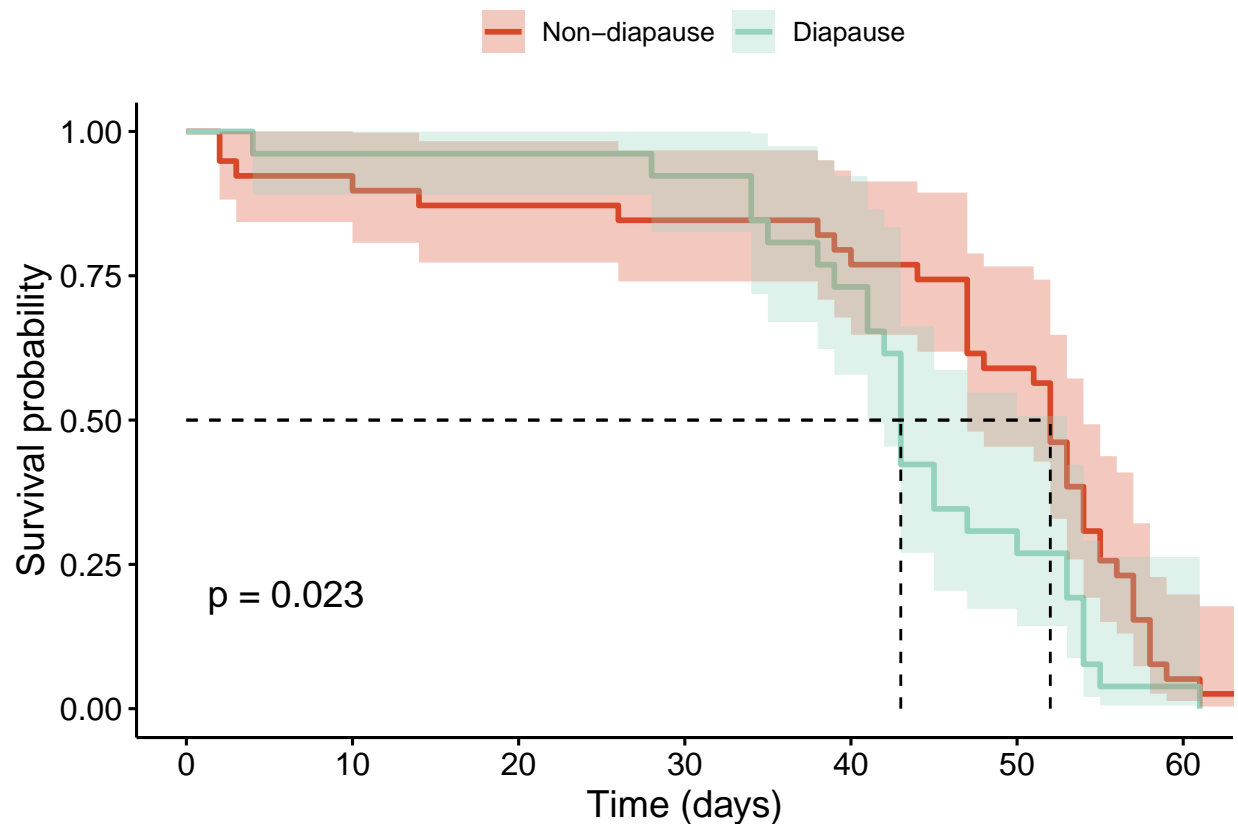
```
autoplot(km_fit.r2F.DN)
```



```
ggsurvplot(km_fit.r2F.DN, data = lifespan.r2F,  
  pval = TRUE,  
  palette = c("#D84727", "#94D2BD"),  
  conf.int = TRUE,  
  # risk.table = "percentage",  
  # risk.table.col = "strata",  
  legend.title = "",  
  legend.labs = c("Non-diapause", "Diapause"),
```

```
#      risk.table.height = 0.3,
      break.time.by = 10,
      xlab = "Time (days)",
      surv.median.line = "hv",
      font.x = c(14),
      font.y = c(14)
    )
```

```
## Warning in geom_segment(aes(x = 0, y = max(y2), xend = max(x1), yend = max(y2)), : All aesthetics ha
## i Please consider using 'annotate()' or provide this layer with data containing
##   a single row.
## All aesthetics have length 1, but the data has 2 rows.
## i Please consider using 'annotate()' or provide this layer with data containing
##   a single row.
```



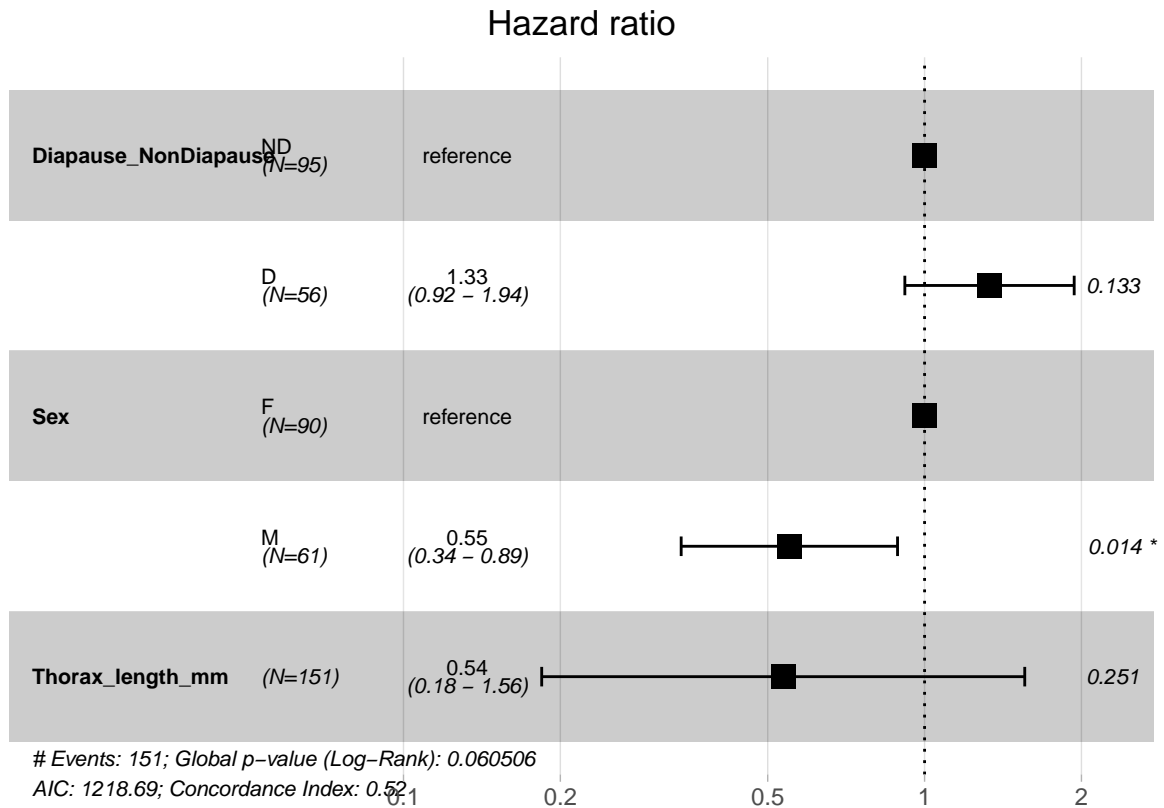
```
ggsave("survival.png", height = 4, width = 6)
```

```
## Warning in geom_segment(aes(x = 0, y = max(y2), xend = max(x1), yend = max(y2)), : All aesthetics ha
## i Please consider using 'annotate()' or provide this layer with data containing
##   a single row.
```

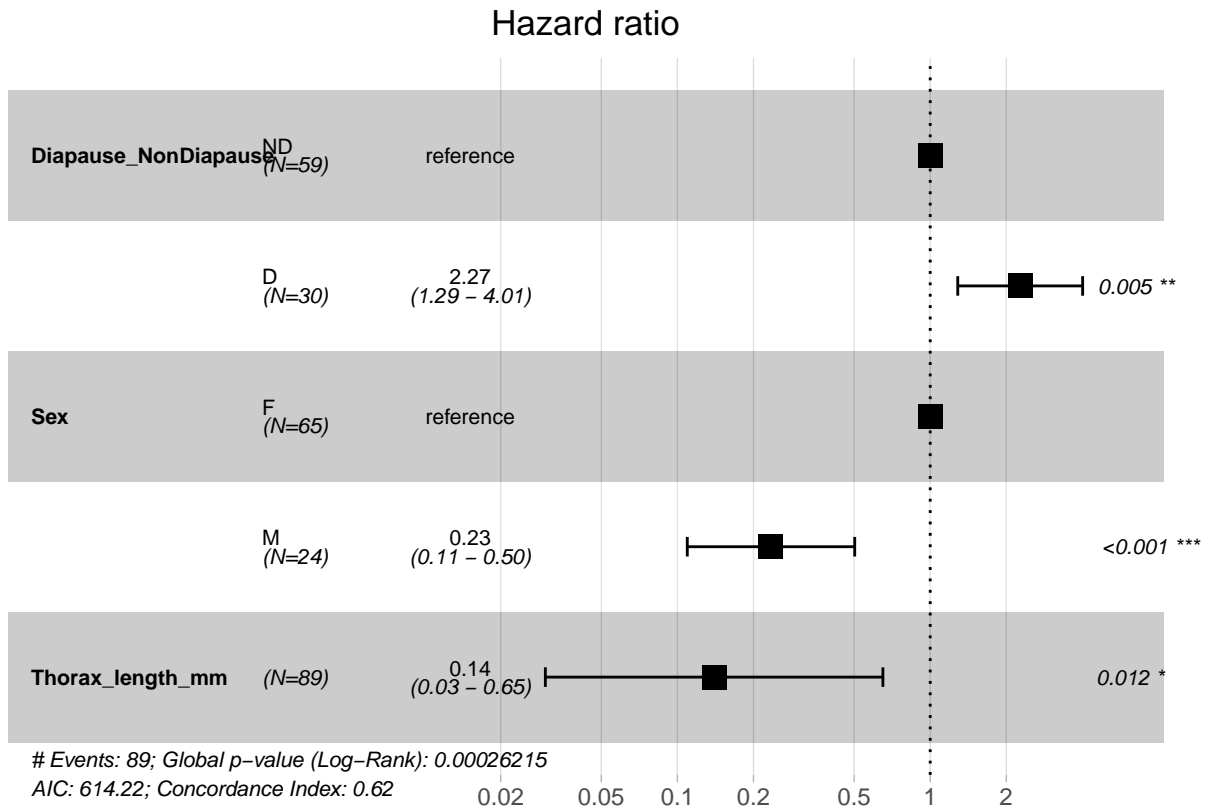

Cox-Hazard models

All others

```
# Fit a Cox proportional hazards model
fit.coxph.r <- coxph(km.r ~ Diapause_NonDiapause + Sex + Thorax_length_mm,
                    data = as.data.frame(lifespan.r))
ggforest(fit.coxph.r, data = as.data.frame(lifespan.r))
```

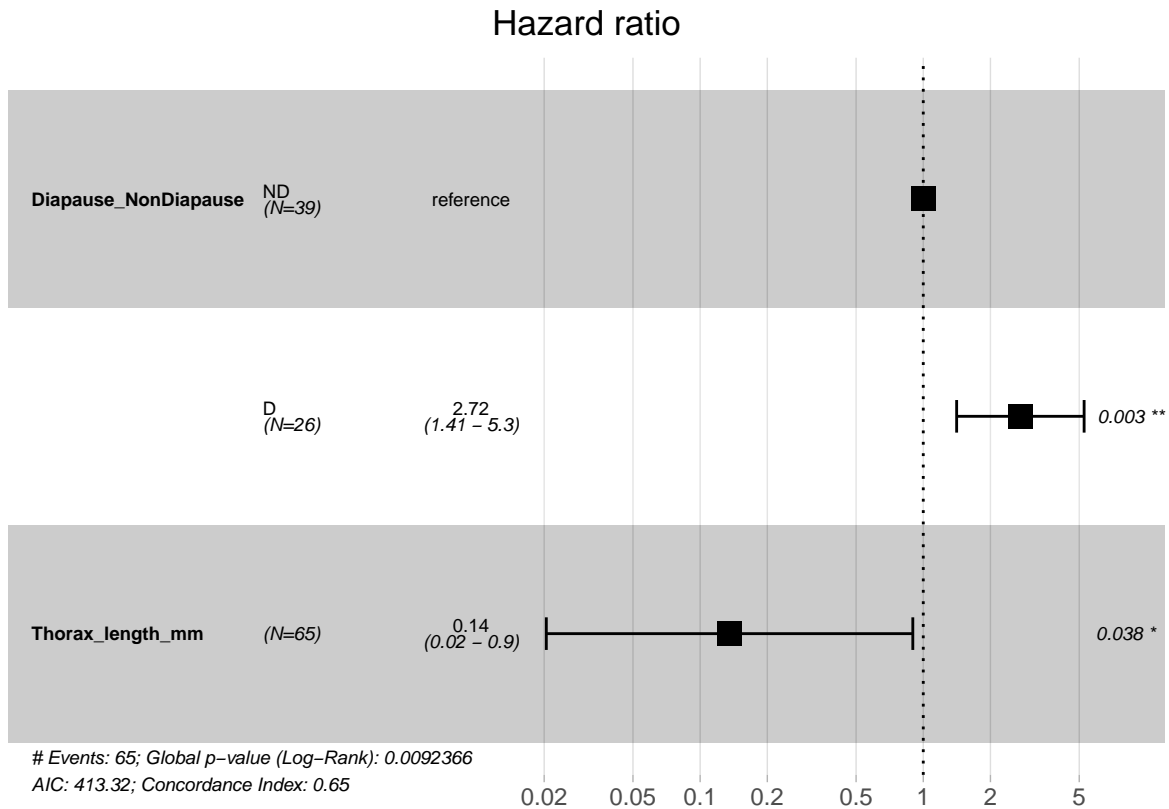


```
# Fit a Cox proportional hazards model
fit.coxph.r2 <- coxph(km.r2 ~ Diapause_NonDiapause + Sex + Thorax_length_mm,
                    data = as.data.frame(lifespan.r2))
ggforest(fit.coxph.r2, data = as.data.frame(lifespan.r2))
```



Females only

```
# Fit a Cox proportional hazards model
fit.coxph.r2F <- coxph(km.r2F ~ Diapause_NonDiapause + Thorax_length_mm,
  data = as.data.frame(lifespan.r2F))
ggforest(fit.coxph.r2F, data = as.data.frame(lifespan.r2F), fontsize = 0.65)
```



```
ggsave("hazards.svg", height = 4.5, width = 6.5)
```

book-keeping

```
sessionInfo()
```

```
## R version 4.3.2 (2023-10-31)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Sonoma 14.2.1
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib; LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/Denver
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
```

```
## other attached packages:
## [1] ggfortify_0.4.17 survminer_0.5.0 ggpubr_0.6.0 survival_3.7-0
## [5] lubridate_1.9.3 forcats_1.0.0 stringr_1.5.1 dplyr_1.1.4
## [9] purrr_1.0.2 readr_2.1.5 tidyr_1.3.1 tibble_3.2.1
## [13] ggplot2_3.5.1 tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] gtable_0.3.5 xfun_0.46 rstatix_0.7.2 lattice_0.22-6
## [5] tzdb_0.4.0 vctrs_0.6.5 tools_4.3.2 generics_0.1.3
## [9] parallel_4.3.2 fansi_1.0.6 highr_0.11 pkgconfig_2.0.3
## [13] Matrix_1.6-5 data.table_1.15.4 lifecycle_1.0.4 farver_2.1.2
## [17] compiler_4.3.2 textshaping_0.4.0 munsell_0.5.1 carData_3.0-5
## [21] htmltools_0.5.8.1 yaml_2.3.10 crayon_1.5.3 pillar_1.9.0
## [25] car_3.1-2 abind_1.4-5 km.ci_0.5-6 tidyselect_1.2.1
## [29] digest_0.6.36 stringi_1.8.4 labeling_0.4.3 splines_4.3.2
## [33] cowplot_1.1.3 fastmap_1.2.0 grid_4.3.2 colorspace_2.1-1
## [37] cli_3.6.3 magrittr_2.0.3 utf8_1.2.4 broom_1.0.6
## [41] withr_3.0.0 scales_1.3.0 backports_1.5.0 bit64_4.0.5
## [45] timechange_0.3.0 rmarkdown_2.27 bit_4.0.5 gridExtra_2.3
## [49] ggsignif_0.6.4 ragg_1.3.2 zoo_1.8-12 hms_1.1.3
## [53] evaluate_0.24.0 knitr_1.48 KMSurv_0.1-5 survMisc_0.5.6
## [57] rlang_1.1.4 xtable_1.8-4 glue_1.7.0 svglite_2.1.3
## [61] rstudioapi_0.16.0 vroom_1.6.5 R6_2.5.1 systemfonts_1.1.0
```

```
citation("tidyverse")
```

```
## To cite package 'tidyverse' in publications use:
```

```
##
```

```
## Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R,
## Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller
## E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V,
## Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to
## the tidyverse." Journal of Open Source Software, 4*(43), 1686.
## doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>.
```

```
##
```

```
## A BibTeX entry for LaTeX users is
```

```
##
```

```
## @Article{,
##   title = {Welcome to the {tidyverse}},
##   author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostini
##   year = {2019},
##   journal = {Journal of Open Source Software},
##   volume = {4},
##   number = {43},
##   pages = {1686},
##   doi = {10.21105/joss.01686},
## }
```

```
citation("survival")
```

```
## To cite package 'survival' in publications use:
```

```
##
```

```
## Therneau T (2024). A Package for Survival Analysis in R. R package
```

```
## version 3.7-0, <https://CRAN.R-project.org/package=survival>.
##
## Terry M. Therneau, Patricia M. Grambsch (2000). _Modeling Survival
## Data: Extending the Cox Model_. Springer, New York. ISBN
## 0-387-98784-3.
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

```
citation("survminer")
```

```
## To cite package 'survminer' in publications use:
##
## Kassambara A, Kosinski M, Biecek P (2024). _survminer: Drawing
## Survival Curves using 'ggplot2'_. R package version 0.5.0,
## <https://CRAN.R-project.org/package=survminer>.
##
## A BibTeX entry for LaTeX users is
##
## @Manual{,
##   title = {survminer: Drawing Survival Curves using 'ggplot2'},
##   author = {Alboukadel Kassambara and Marcin Kosinski and Przemyslaw Biecek},
##   year = {2024},
##   note = {R package version 0.5.0},
##   url = {https://CRAN.R-project.org/package=survminer},
## }
```

```
citation("ggfortify")
```

```
## To cite ggfortify in publications, please use:
##
## Yuan Tang, Masaaki Horikoshi, and Wenxuan Li. "ggfortify: Unified
## Interface to Visualize Statistical Result of Popular R Packages." The
## R Journal 8.2 (2016): 478-489.
##
## Masaaki Horikoshi and Yuan Tang (2016). ggfortify: Data Visualization
## Tools for Statistical Analysis Results.
## https://CRAN.R-project.org/package=ggfortify
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```