PS4

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

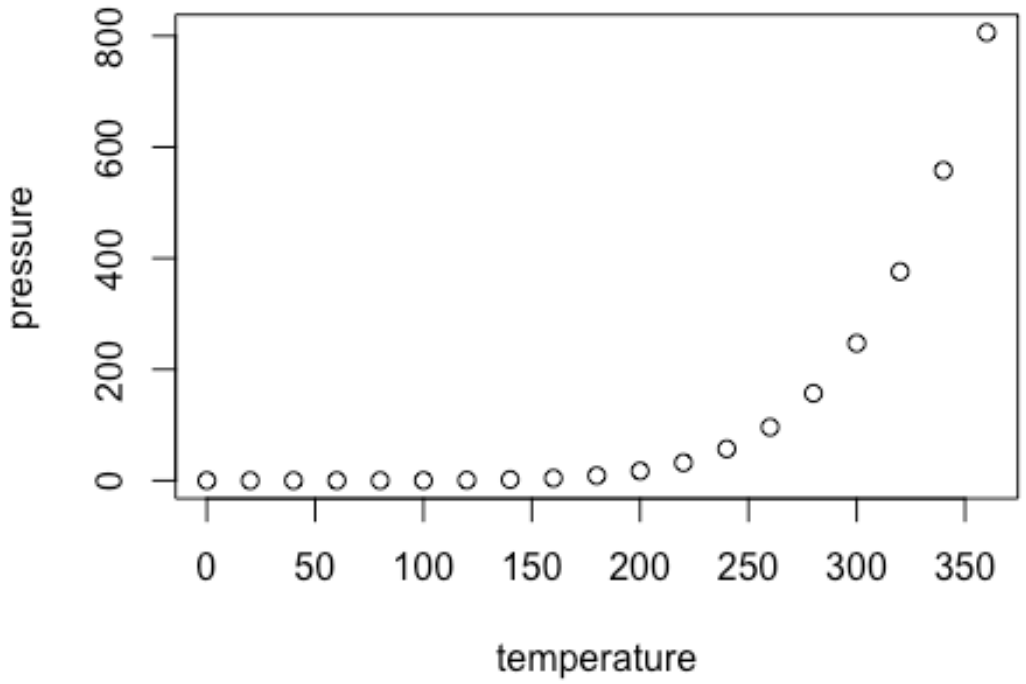
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

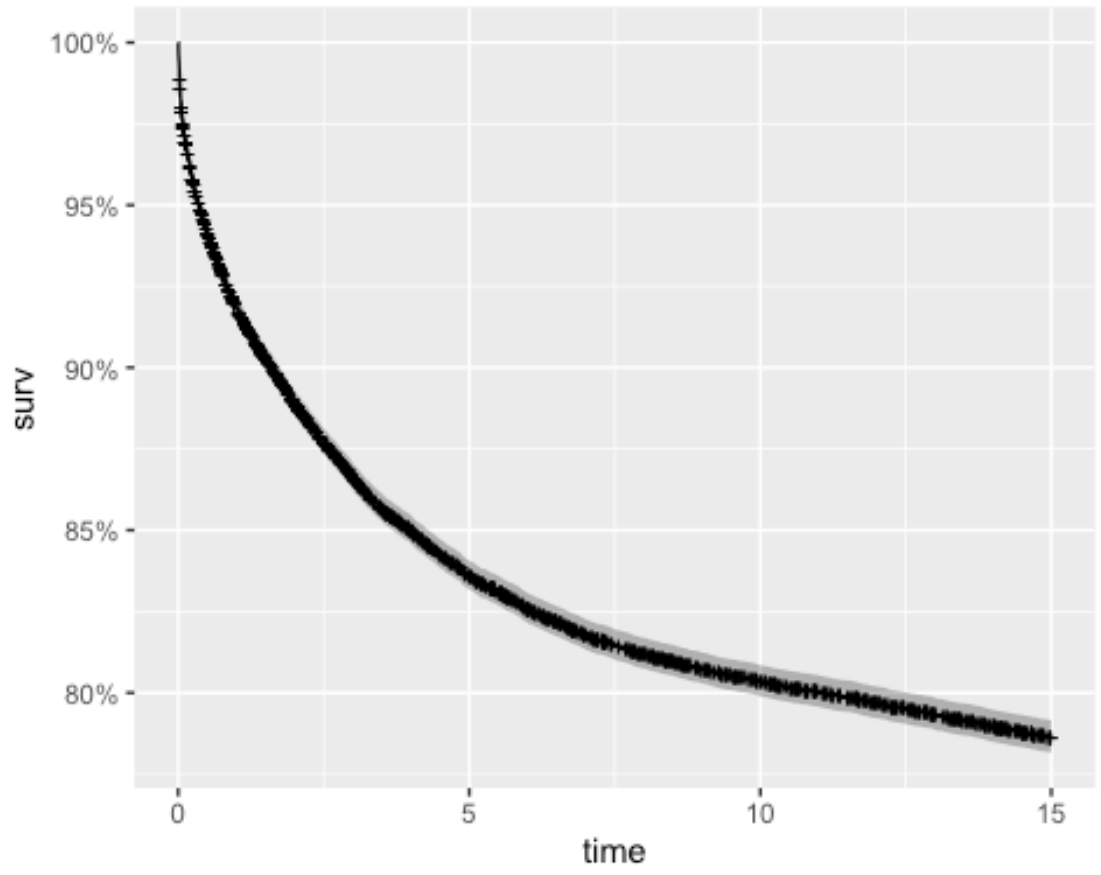
You can also embed plots, for example:



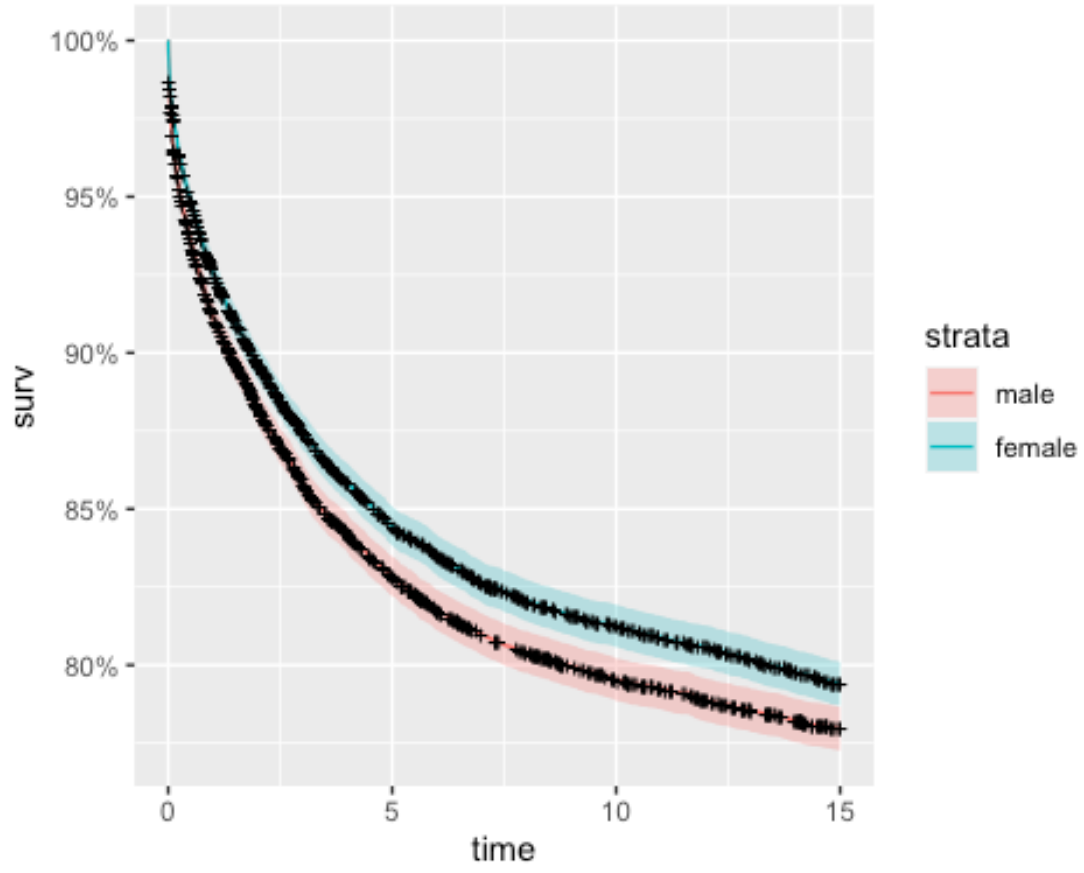
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
library('eha')
library('survival')
library('tidyverse')

## ── Attaching packages ─────────────────────────────────────── tidyverse
1.3.2 ──
## ✓ ggplot2 3.4.1     ✓ purrr   1.0.1
## ✓ tibble  3.1.8     ✓ dplyr   1.1.0
## ✓ tidyr   1.3.0     ✓ stringr 1.5.0
## ✓ readr   2.1.2     ✓ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()

library('ggfortify')
library('stargazer')

##
## Please cite as:
```

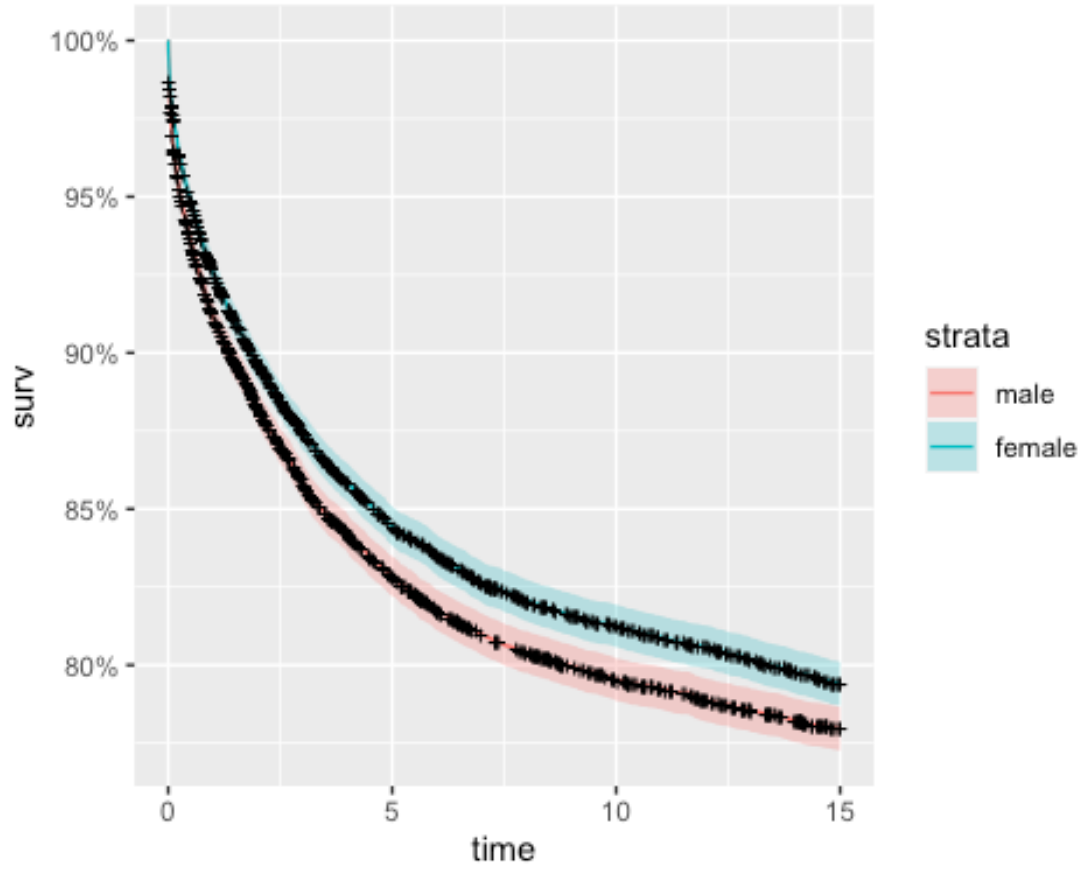
```
##
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary
Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

data(child)

child$m.agegroup <- cut(child$m.age, breaks = seq(10, 60, 10), right = FALSE)

child_surv <- with(child, Surv(enter, exit, event))

# Overall survival
km_overall <- survfit(child_surv ~ 1, data = child)
summary(km_overall, times = seq(0, 15, 1))

## Call: survfit(formula = child_surv ~ 1, data = child)
##
##  time n.risk n.event censored survival std.err lower 95% CI upper 95% CI
##     0  26574       0        0    1.000 0.00000        1.000        1.000
##     1  24319    2161       94    0.919 0.00168        0.915        0.922
##     2  23450     778       91    0.889 0.00193        0.885        0.893
##     3  22766     596       88    0.867 0.00209        0.862        0.871
##     4  22269     430       68    0.850 0.00220        0.846        0.854
##     5  21859     365       44    0.836 0.00228        0.832        0.841
##     6  21533     261       65    0.826 0.00233        0.822        0.831
##     7  21266     214       53    0.818 0.00238        0.813        0.823
##     8  21077     151       38    0.812 0.00241        0.807        0.817
##     9  20915     117       45    0.808 0.00243        0.803        0.812
##    10  20777     103       35    0.804 0.00245        0.799        0.808
##    11  20655      81       41    0.801 0.00246        0.796        0.805
##    12  20531      91       33    0.797 0.00248        0.792        0.802
##    13  20404      89       38    0.794 0.00250        0.789        0.798
##    14  20277      95       32    0.790 0.00251        0.785        0.795
##    15  20141      84    20193    0.787 0.00253        0.782        0.792

autoplot(km_overall)
```




```r
# Impact of child gender
km_sex <- survfit(child_surv ~ sex, data = child)
autoplot(km_sex)
```



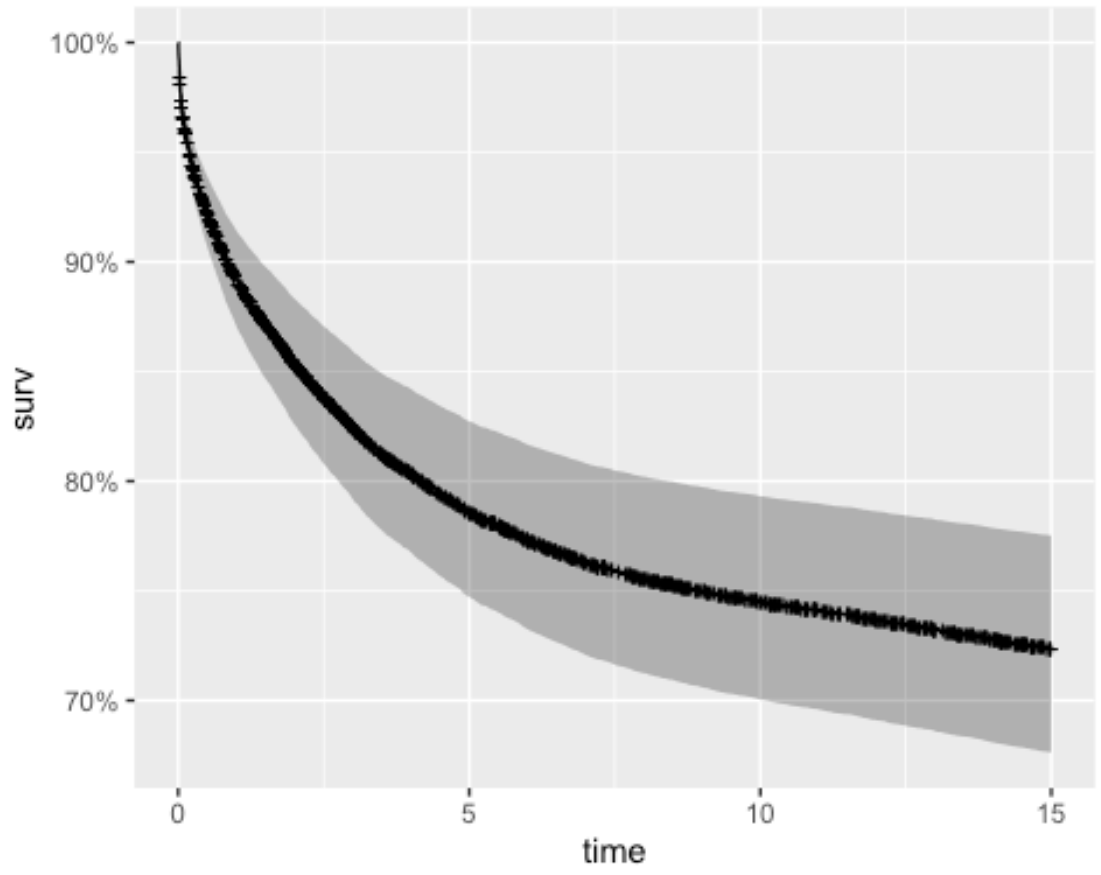

```
# Impact of mother's age
km_magegroup <- survfit(child_surv ~ m.agegroup, data = child)
autoplot(km_sex)
```



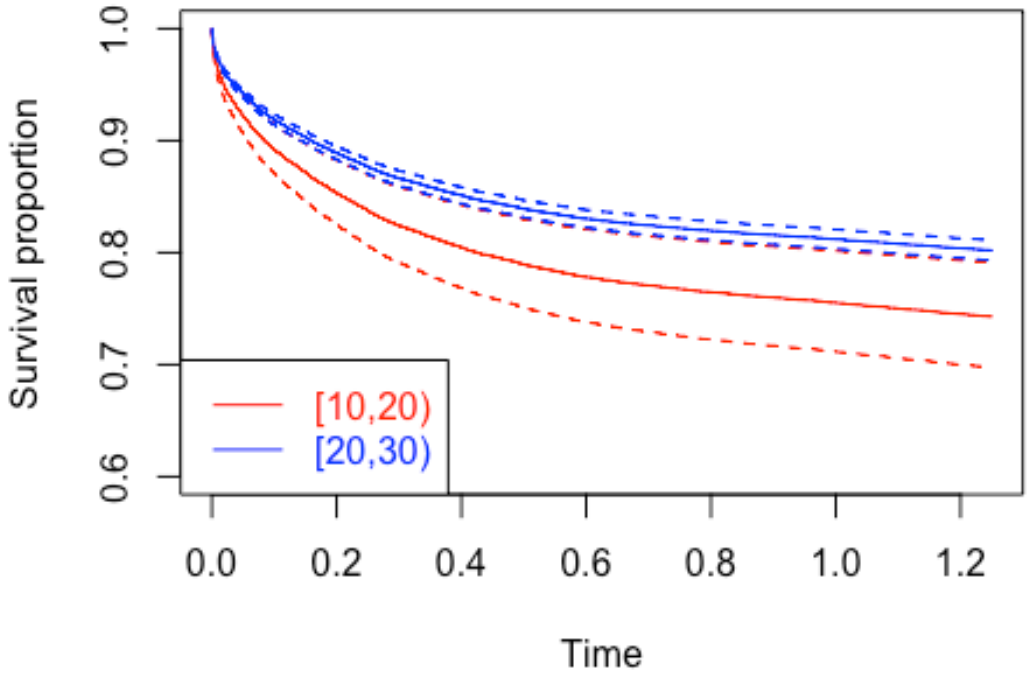

```
### MODEL

cox <- coxph(child_surv ~ m.agegroup + sex, data = child)

## Warning in agreg.fit(X, Y, istrat, offset, init, control, weights =
weights, :
## Loglik converged before variable 4 ; beta may be infinite.

summary(cox)

## Call:
## coxph(formula = child_surv ~ m.agegroup + sex, data = child)
##
##   n= 26574, number of events= 5616
##
##                          coef  exp(coef)   se(coef)      z Pr(>|z|)
## m.agegroup[20,30) -2.985e-01  7.420e-01  1.094e-01 -2.728  0.00638 **
## m.agegroup[30,40) -2.618e-01  7.697e-01  1.089e-01 -2.403  0.01627 *
## m.agegroup[40,50) -1.391e-01  8.702e-01  1.132e-01 -1.228  0.21936
## m.agegroup[50,60) -1.022e+01  3.656e-05  2.145e+02 -0.048  0.96201
## sexfemale         -8.250e-02  9.208e-01  2.674e-02 -3.085  0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##                   exp(coef) exp(-coef)  lower .95  upper .95
## m.agegroup[20,30) 7.420e-01      1.348  5.988e-01  9.194e-01
## m.agegroup[30,40) 7.697e-01      1.299  6.217e-01  9.529e-01
## m.agegroup[40,50) 8.702e-01      1.149  6.970e-01  1.086e+00
## m.agegroup[50,60) 3.656e-05  27350.930 1.014e-187 1.319e+178
## sexfemale         9.208e-01      1.086  8.738e-01  9.704e-01
##
## Concordance= 0.52  (se = 0.004 )
## Likelihood ratio test= 29.76  on 5 df,   p=2e-05
## Wald test            = 29.58  on 5 df,   p=2e-05
## Score (logrank) test = 30.13  on 5 df,   p=1e-05
```

```
# the model for age group shows us that mothers in their teens have higher child mortality rates than the other age brackets
# the model also highlights that child mortality increases with mothers over the age of 40
# the model shows us that there are up to 26% fewer child deaths in mothers 20 and older versus teenage mothers


### FIT MODEL

cox_fit <- survfit(cox)
autoplot(cox_fit)
```




```
newdat <- with(child,
               data.frame(
                   sex = "female", m.agegroup=c("[10,20)", "[20,30)")
               ))

plot(survfit(cox, newdata = newdat), xscale = 12,
     conf.int = T,
     ylim = c(0.6, 1),
     col = c("red", "blue"),
     xlab = "Time",
     ylab = "Survival proportion",
     main = "")
legend("bottomleft",
       legend=c("[10,20)", "[20,30)"),
       lty = 1,
       col = c("red", "blue"),
       text.col = c("red", "blue"))
```


Survival proportion
1.0
0.9
0.8
0.7
0.6
[10,20)
[20,30)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Time