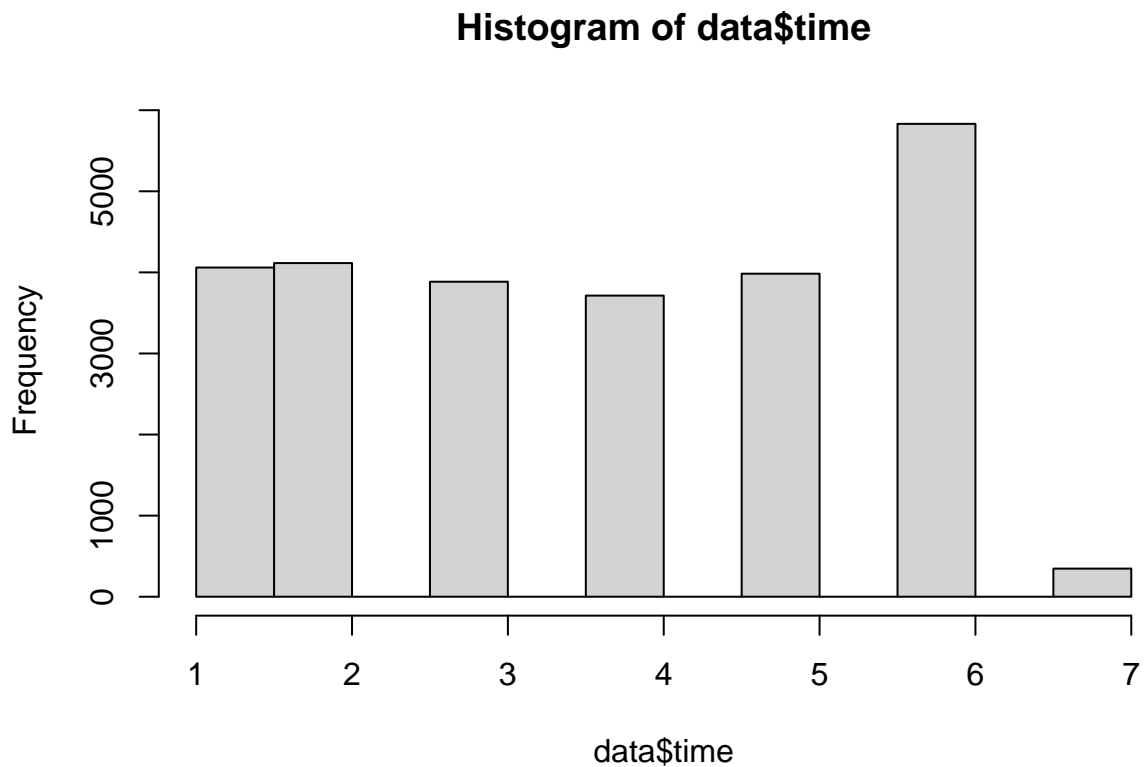# surv_eda

```
library(survival)
```

1. MOF

```
data = read.csv("surv_data_MOF.csv")
data$status = as.integer(as.logical(data$status))
attach(data)

# time histogram
hist(data$time)
```

Histogram of data$time


```
table(data$status)
```

```
##
##     0     1
## 25444   493
```

1.1 KM-model

```
km.model = survfit(Surv(time, status) ~ 1, type = "kaplan-meier")
summary(km.model)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, type = "kaplan-meier")
##
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1  25937     137    0.995 0.000450        0.994        0.996
##     2  21877     107    0.990 0.000649        0.989        0.991
##     3  17762      89    0.985 0.000832        0.983        0.987
##     4  13877      75    0.980 0.001029        0.978        0.982
##     5  10163      50    0.975 0.001229        0.972        0.977
##     6   6179      34    0.969 0.001529        0.966        0.972
##     7    347       1    0.967 0.003179        0.960        0.973
```

```r
plot(km.model, confi.inf = F, xlab = "Time(year)", ylab = "S(t)", main = "KM-model", ylim = c(0.95, 1))
```

```
## Warning in plot.window(...): "confi.inf" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "confi.inf" is not a graphical parameter
```
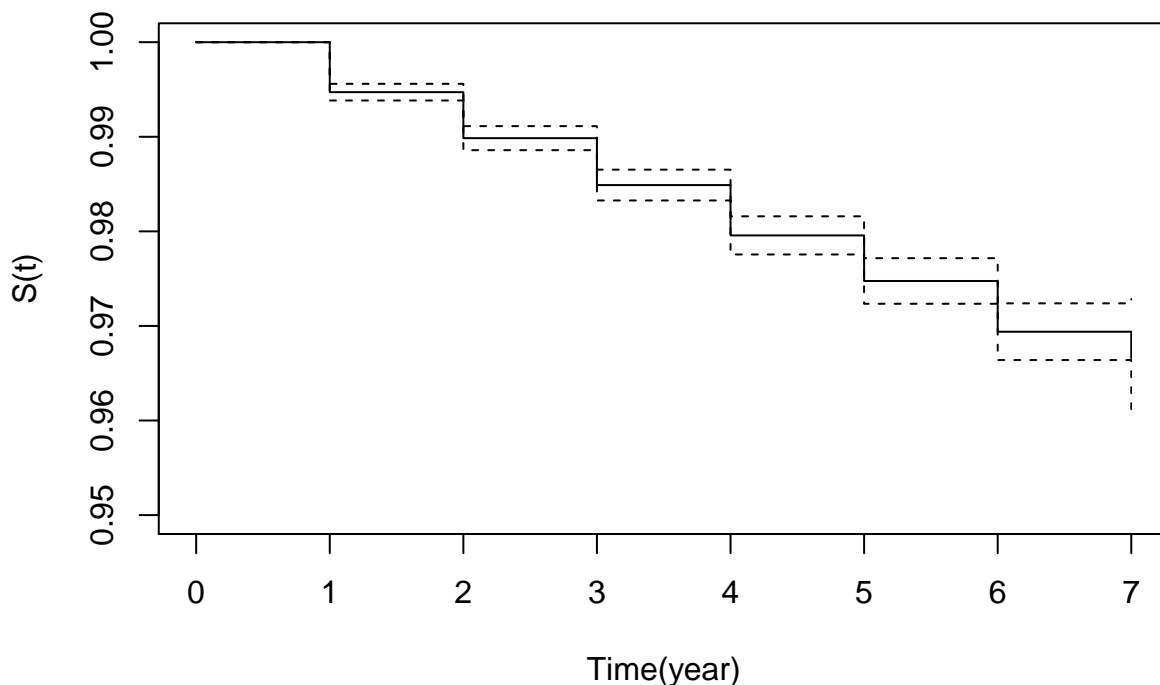
```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```

```
## Warning in box(...): "confi.inf" is not a graphical parameter
```

```
## Warning in title(...): "confi.inf" is not a graphical parameter
```

KM−model

1.2 KM-model ~ gender

```r
table(male)
```

```
## male
```

```
##     0     1
## 13968 11969
```

```
km.model.g = survfit(Surv(time, status) ~ male, type = "kaplan-meier")
summary(km.model.g)
```

```
## Call: survfit(formula = Surv(time, status) ~ male, type = "kaplan-meier")
##
##                male=0
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1  13968      95    0.993 0.000695        0.992        0.995
##     2  11773      63    0.988 0.000961        0.986        0.990
##     3   9591      49    0.983 0.001197        0.980        0.985
##     4   7443      45    0.977 0.001481        0.974        0.980
##     5   5400      30    0.971 0.001774        0.968        0.975
##     6   3283      22    0.965 0.002240        0.961        0.969
##     7    228       1    0.961 0.004776        0.951        0.970
##
##                male=1
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1  11969      42    0.996 0.000541        0.995        0.998
##     2  10104      44    0.992 0.000846        0.990        0.994
##     3   8171      40    0.987 0.001138        0.985        0.990
##     4   6434      30    0.983 0.001409        0.980        0.985
##     5   4763      20    0.979 0.001679        0.975        0.982
##     6   2896      12    0.975 0.002039        0.971        0.979
```

```
plot(km.model.g, confi.inf = F, xlab = "Time(year)", ylab = "S(t)", main = "KM-model", ylim = c(0.96, 1)
```
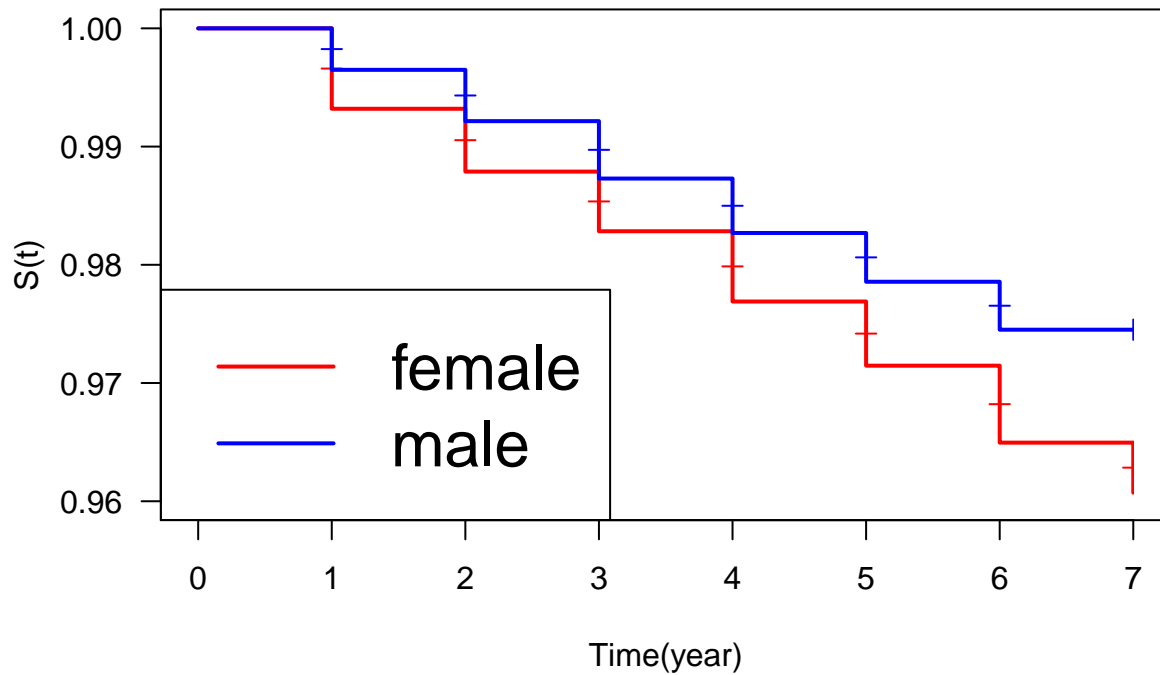
```
## Warning in plot.window(...): "confi.inf" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "confi.inf" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```

```
## Warning in box(...): "confi.inf" is not a graphical parameter
```

```
## Warning in title(...): "confi.inf" is not a graphical parameter
```

```
legend("bottomleft", legend = c("female", "male"), col = c("red", "blue"), lty=1, lwd=2, cex = 2)
```

KM–model


```
survdiff(Surv(time, status) ~ male)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ male)
##
##            N Observed Expected (O-E)^2/E (O-E)^2/V
## male=0 13968      305      265      6.06      13.2
## male=1 11969      188      228      7.04      13.2
##
##  Chisq= 13.2  on 1 degrees of freedom, p= 3e-04
```

1.2 KM-model ~ race

```
table(race)
```

```
## race
##     0     1     2     3
## 12758  7844  3587  1748
```

```
km.model.r = survfit(Surv(time, status) ~ race, type = "kaplan-meier")
summary(km.model.r)
```

```
## Call: survfit(formula = Surv(time, status) ~ race, type = "kaplan-meier")
##
##                  race=0
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1  12758      78    0.994 0.00069        0.993        0.995
##     2  10779      73    0.987 0.00104        0.985        0.989
##     3   8812      50    0.982 0.00130        0.979        0.984
##     4   6816      46    0.975 0.00162        0.972        0.978
##     5   4920      29    0.969 0.00193        0.965        0.973
```

```
##     6  3015     16   0.964 0.00231          0.960          0.969
##
##                  race=1
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1  7844     32   0.996 0.000720          0.995          0.997
##     2  6740     15   0.994 0.000918          0.992          0.996
##     3  5617     19   0.990 0.001196          0.988          0.993
##     4  4429     20   0.986 0.001553          0.983          0.989
##     5  3280     10   0.983 0.001816          0.979          0.986
##     6  1969     13   0.976 0.002544          0.971          0.981
##     7   241      1   0.972 0.004771          0.963          0.982
##
##                  race=2
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1  3587     21   0.994 0.00127          0.992          0.997
##     2  3018     12   0.990 0.00170          0.987          0.994
##     3  2315     15   0.984 0.00237          0.979          0.988
##     4  1856      7   0.980 0.00274          0.975          0.985
##     5  1387      9   0.974 0.00345          0.967          0.980
##     6   835      4   0.969 0.00414          0.961          0.977
##
##                  race=3
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1  1748      6   0.997 0.00140          0.994          0.999
##     2  1340      7   0.991 0.00241          0.987          0.996
##     3  1018      5   0.986 0.00323          0.980          0.993
##     4   776      2   0.984 0.00369          0.977          0.991
##     5   576      2   0.981 0.00440          0.972          0.989
##     6   360      1   0.978 0.00516          0.968          0.988
```

```r
plot(km.model.r, confi.inf = F, xlab = "Time(year)", ylab = "S(t)", main = "KM-model", ylim = c(0.96, 1)
```

```
## Warning in plot.window(...): "confi.inf" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "confi.inf" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```
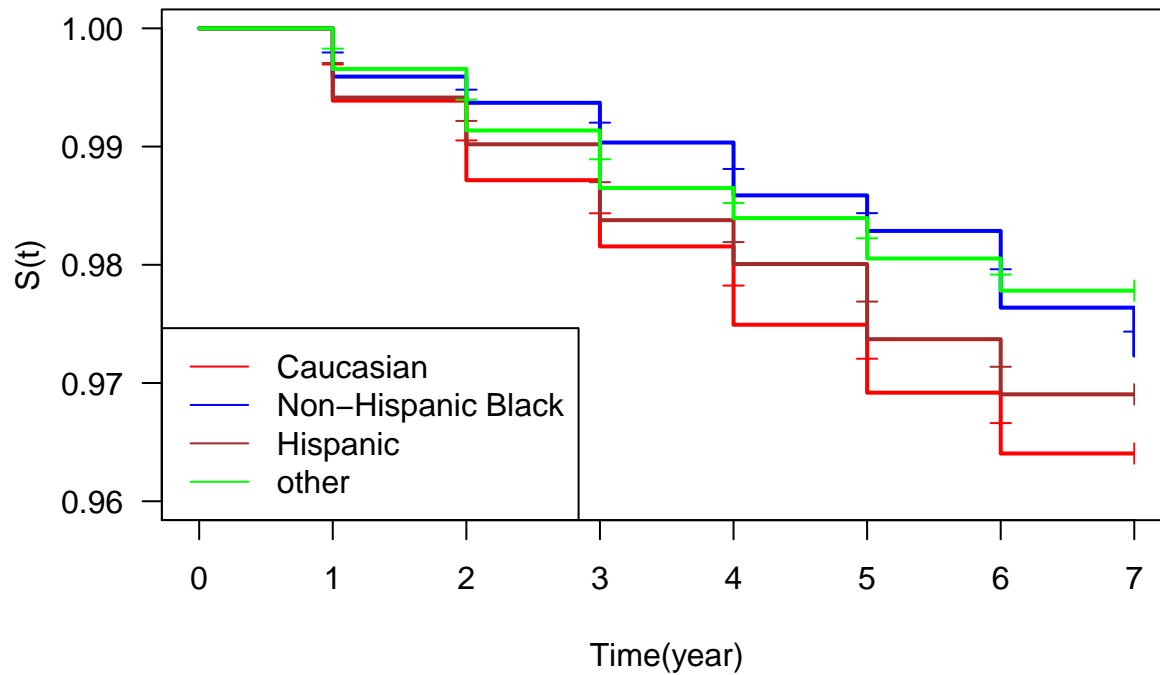
```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```

```
## Warning in box(...): "confi.inf" is not a graphical parameter
```

```
## Warning in title(...): "confi.inf" is not a graphical parameter
```

```r
legend("bottomleft", legend = c("Caucasian", "Non-Hispanic Black", "Hispanic", "other"), col = c("red",
```

KM-model


```
survdiff(Surv(time, status) ~ race)
```
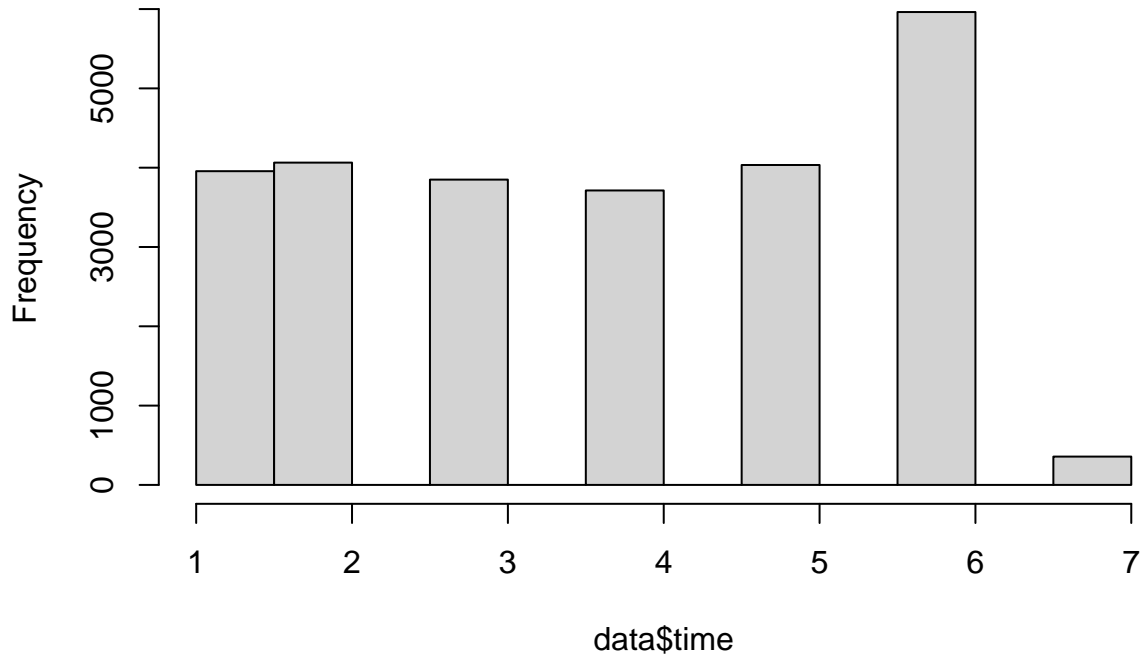
```
## Call:
## survdiff(formula = Surv(time, status) ~ race)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## race=0 12758      292    242.0   10.3164    20.391
## race=1  7844      110    154.1   12.6428    18.518
## race=2  3587       68     66.9    0.0189     0.022
## race=3  1748       23     29.9    1.6121     1.726
##
##  Chisq= 24.8  on 3 degrees of freedom, p= 2e-05
```

```
detach(data)
```

2. Hip

```
data = read.csv("surv_data_hip.csv")
data$status = as.integer(as.logical(data$status))
attach(data)

## plot distribution
hist(data$time)
```

Histogram of data$time


```
table(data$status)
```

```
##
##     0     1
## 25870    67
```

2.1 KM-model

```
km.model = survfit(Surv(time, status) ~ 1, type = "kaplan-meier")
summary(km.model)
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, type = "kaplan-meier")
##
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1  25937      19    0.999 0.000168        0.999        1.000
##     2  21981      18    0.998 0.000256        0.998        0.999
##     3  17917      11    0.998 0.000315        0.997        0.998
##     4  14067       8    0.997 0.000374        0.997        0.998
##     5  10354       7    0.997 0.000452        0.996        0.997
##     6   6320       3    0.996 0.000528        0.995        0.997
##     7    357       1    0.993 0.002836        0.988        0.999
```

```
plot(km.model, confi.inf = F, xlab = "Time(year)", ylab = "S(t)", main = "KM-model", ylim = c(0.99, 1))
```

```
## Warning in plot.window(...): "confi.inf" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "confi.inf" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
```
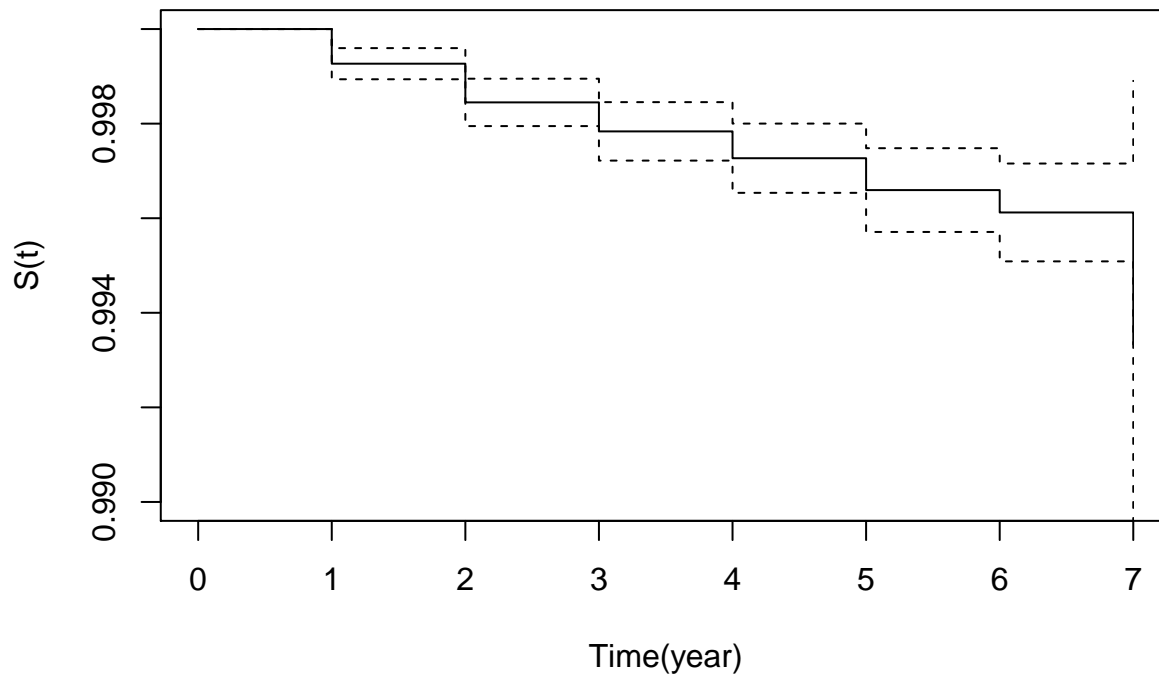
```
## a graphical parameter
## Warning in box(...): "confi.inf" is not a graphical parameter
## Warning in title(...): "confi.inf" is not a graphical parameter
```

KM−model

2.2 KM-model ~ gender

```
table(male)
```

```
## male
##     0     1
## 13968 11969
```

```
km.model.g = survfit(Surv(time, status) ~ male, type = "kaplan-meier")
summary(km.model.g)
```

```
## Call: survfit(formula = Surv(time, status) ~ male, type = "kaplan-meier")
##
##              male=0
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1  13968      14    0.999 0.000268        0.998        1.000
##     2  11844      11    0.998 0.000387        0.997        0.999
##     3   9690       5    0.998 0.000450        0.997        0.998
##     4   7558       3    0.997 0.000505        0.996        0.998
##     5   5519       3    0.997 0.000594        0.995        0.998
##     6   3373       1    0.996 0.000663        0.995        0.998
##     7    236       1    0.992 0.004264        0.984        1.000
##
##              male=1
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
```

```
##     1  11969       5    1.000 0.000187          0.999            1.000
##     2  10137       7    0.999 0.000321          0.998            1.000
##     3   8227       6    0.998 0.000437          0.997            0.999
##     4   6509       5    0.997 0.000555          0.996            0.998
##     5   4835       4    0.997 0.000691          0.995            0.998
##     6   2947       2    0.996 0.000840          0.994            0.998
```

```r
plot(km.model.g, confi.inf = F, xlab = "Time(year)", ylab = "S(t)", main = "KM-model", ylim = c(0.99, 1)
```

```
## Warning in plot.window(...): "confi.inf" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "confi.inf" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter

## Warning in box(...): "confi.inf" is not a graphical parameter

## Warning in title(...): "confi.inf" is not a graphical parameter
```
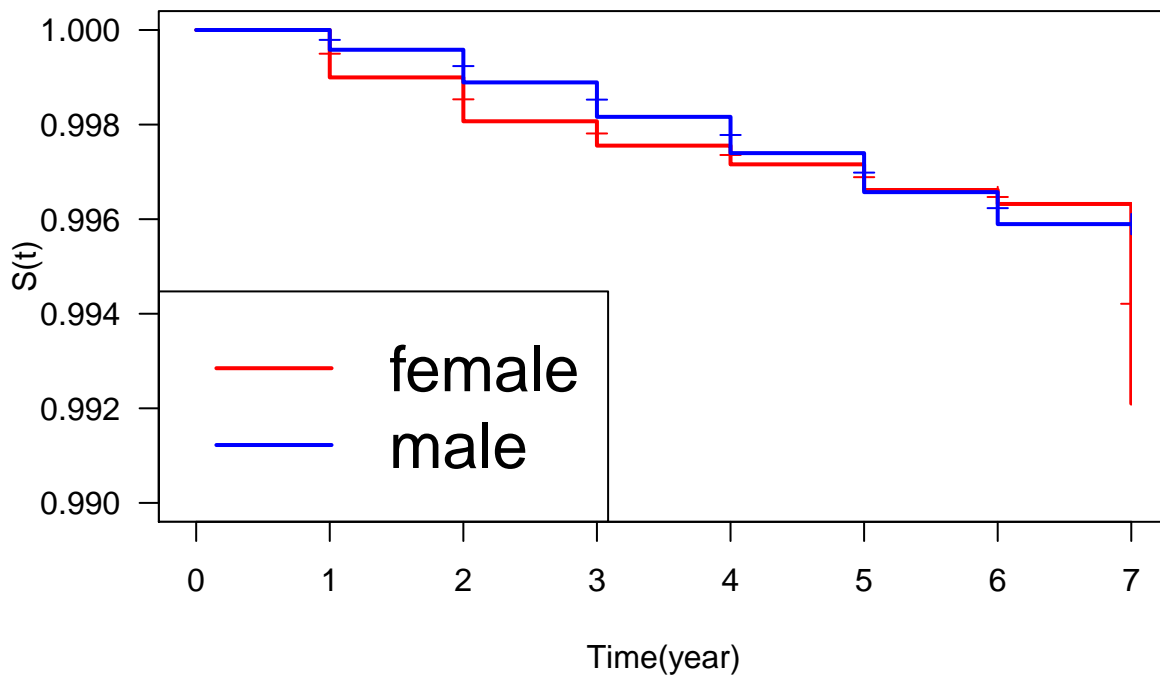
```r
legend("bottomleft", legend = c("female", "male"), col = c("red", "blue"), lty=1, lwd=2, cex = 2)
```



```r
survdiff(Surv(time, status) ~ male)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ male)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## male=0 13968       38     36.2    0.0924     0.201
## male=1 11969       29     30.8    0.1084     0.201
```

```
##
##  Chisq= 0.2  on 1 degrees of freedom, p= 0.7
```

2.3 KM-model ~ race

```
table(race)
```

```
## race
##     0     1     2     3
## 12758  7844  3587  1748
```

```
km.model.r = survfit(Surv(time, status) ~ race, type = "kaplan-meier")
summary(km.model.r)
```

```
## Call: survfit(formula = Surv(time, status) ~ race, type = "kaplan-meier")
##
##               race=0
##   time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##      1  12758      13    0.999 0.000282        0.998        1.000
##      2  10835      13    0.998 0.000436        0.997        0.999
##      3   8904       7    0.997 0.000527        0.996        0.998
##      4   6924       6    0.996 0.000633        0.995        0.997
##      5   5018       2    0.996 0.000693        0.994        0.997
##      6   3091       2    0.995 0.000829        0.993        0.997
##
##               race=1
##   time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##      1   7844       4    0.999 0.000255        0.999        1.000
##      2   6768       4    0.999 0.000390        0.998        1.000
##      3   5652       1    0.999 0.000428        0.998        1.000
##      4   4473       1    0.998 0.000483        0.998        0.999
##      5   3333       4    0.997 0.000769        0.996        0.999
##      6   2005       1    0.997 0.000915        0.995        0.999
##      7    245       1    0.993 0.004161        0.985        1.000
##
##               race=2
##   time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##      1   3587       1    1.000 0.000279        0.999        1.000
##      2   3036       1    0.999 0.000431        0.999        1.000
##      3   2338       3    0.998 0.000856        0.996        1.000
##      4   1886       1    0.998 0.001006        0.996        1.000
##      5   1420       1    0.997 0.001226        0.994        0.999
##
##               race=3
##          time      n.risk     n.event     survival      std.err lower 95% CI
##      1.00e+00    1.75e+03    1.00e+00     9.99e-01     5.72e-04     9.98e-01
## upper 95% CI
##      1.00e+00
```

```
plot(km.model.r, confi.inf = F, xlab = "Time(year)", ylab = "S(t)", main = "KM-model", ylim = c(0.99, 1]
```
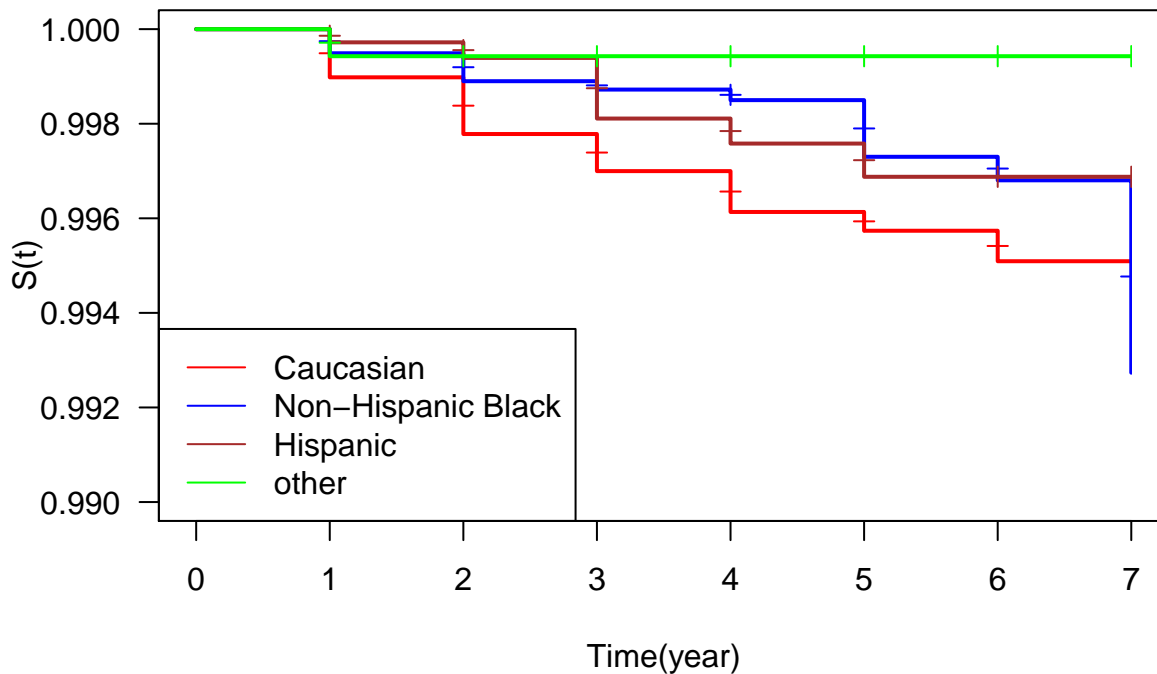
```
## Warning in plot.window(...): "confi.inf" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "confi.inf" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "confi.inf" is not
## a graphical parameter
```

```
## Warning in box(...): "confi.inf" is not a graphical parameter
```

```
## Warning in title(...): "confi.inf" is not a graphical parameter
```

```
legend("bottomleft", legend = c("Caucasian", "Non-Hispanic Black", "Hispanic", "other"), col = c("red",
```

KM−model

```
survdiff(Surv(time, status) ~ race)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ race)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## race=0 12758       43    32.62      3.30      6.485
## race=1  7844       16    21.19      1.27      1.881
## race=2  3587        7     9.11      0.49      0.567
## race=3  1748        1     4.07      2.32      2.469
##
##   Chisq= 7.4  on 3 degrees of freedom, p= 0.06
```

```
detach(data)
```