

00-An-R-Example

Frances Lin

March 2022

```
# Load packages
library(randomForestSRC)

##
## randomForestSRC 3.0.2
##
## Type rfsrc.news() to see new features, changes, and bug fixes.
##

#library(ggRandomForests) # Require older version of R
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::partial() masks randomForestSRC::partial()

library(pander)
library(ggplot2)

# Load data
data(peakV02, package = "randomForestSRC")

peakV02 %>% head(3) %>% pander
```

Table 1: Table continues below

age	betablok	dilver	nifed	acei	angioten.II	anti.arrhy	anti.coag
74	1	0	0	0	1	1	1
77	1	0	0	1	0	0	0

age	betablok	dilver	nifed	acei	angioten.II	anti.arrhy	anti.coag
79	0	0	0	0	0	0	0

Table 2: Table continues below

aspirin	digoxin	nitrates	vasodilator	diuretic.loop	diuretic.thiazide
1	0	0	0	1	0
1	1	0	0	1	0
1	1	0	0	1	0

Table 3: Table continues below

diuretic.potassium.spar	lipidrx.statin	insulin	surgery.pacemaker
0	1	1	0
0	0	0	0
0	0	0	0

Table 4: Table continues below

surgery.cabg	surgery.pci	surgery.aicd.implant	resting.systolic.bp
0	0	1	90
0	0	0	134
0	0	0	108

Table 5: Table continues below

resting.hr	smknow	q.wave.mi	bmi	niddm	lvef.metabl	peak.rer
74	0	0	26.08	0	15	1.12
53	0	0	20.84	0	20	0.98
104	0	0	20.73	0	40	1.15

Table 6: Table continues below

peak.vo2	interval	cad	died	ttodead	bun	sodium	hgb	glucose
11.9	374	1	0	1.207	25	141	14.6	89
24	755	0	0	3.611	20	138	13.1	90
11.2	225	0	1	0.2382	29.69	139.8	12.88	98.74

male	black	crcl
1	0	49.93
0	0	60.35
1	0	33.61

```
# Fit RSF
dta <- peakVO2
obj <- rfsrc(Surv(ttodead, died)~., dta,
             ntree = 1000, nodesize = 5, nsplit = 50, importance = TRUE) # This takes a while
```

```
# Print the outcome
print(obj)
```

```
##                      Sample size: 2231
##                      Number of deaths: 726
##                      Number of trees: 1000
##          Forest terminal node size: 5
##          Average no. of terminal nodes: 260.322
## No. of variables tried at each split: 7
##                      Total no. of variables: 39
##          Resampling used to grow trees: swor
##          Resample size used to grow trees: 1410
##                      Analysis: RSF
##                      Family: surv
##                      Splitting rule: logrank *random*
##          Number of random split points: 50
##                      (OOB) CRPS: 0.15486118
##          (OOB) Requested performance error: 0.29810376
```

```
# Get the C-index
get.cindex(obj$yvar[,1], obj$yvar[,2], obj$predicted.oob)
```

```
## [1] 0.2981038
```

```
df <- obj$xvar
peak_vo2 <- df[,30]
quantile(peak_vo2)
```

```
##    0%   25%   50%   75%  100%
##  4.2 12.8 15.7 19.3 43.8
```

Patient(s) info:

```
# Plot the estimated survival functions
newdata <- data.frame(lapply(1:ncol(obj$xvar),function(i){median(obj$xvar[,i])}))
colnames(newdata) <- obj$xvar.names
newdata1 <- newdata2 <- newdata
newdata1[,which(obj$xvar.names == "peak_vo2")] <- 12.8 # NA
newdata2[,which(obj$xvar.names == "peak_vo2")] <- 19.3 # NA
newdata <- rbind(newdata1,newdata2)
newdata[1,30] = 12.8 # Now it works
newdata[2,30] = 19.3
newdata %>% pander
```

Table 8: Table continues below

age	betablok	dilver	nifed	acei	angioten.II	anti.arrhy	anti.coag
55	1	0	0	1	0	0	0
55	1	0	0	1	0	0	0

Table 9: Table continues below

aspirin	digoxin	nitrates	vasodilator	diuretic.loop	diuretic.thiazide
0	1	0	0	1	0
0	1	0	0	1	0

Table 10: Table continues below

diuretic.potassium.spar	lipidrx.statin	insulin	surgery.pacemaker
0	0	0	0
0	0	0	0

Table 11: Table continues below

surgery.cabg	surgery.pci	surgery.aicd.implant	resting.systolic.bp
0	0	0	110
0	0	0	110

Table 12: Table continues below

resting.hr	smknow	q.wave.mi	bmi	niddm	lvef.metabl	peak.rer
75	0	0	27.78	0	20	1.1
75	0	0	27.78	0	20	1.1

peak.vo2	interval	cad	bun	sodium	hgb	glucose	male	black	crcl
12.8	480	0	23	139.8	13.58	96.93	1	0	83.13
19.3	480	0	23	139.8	13.58	96.93	1	0	83.13

```
y.pred <- predict(obj,newdata = rbind(newdata,obj$xvar)[1:2,])
y.pred
```

```
## Sample size of test (predict) data: 2
##           Number of grow trees: 1000
## Average no. of grow terminal nodes: 260.322
##           Total no. of grow variables: 39
##           Resampling used to grow trees: swor
## Resample size used to grow trees: 1
##           Analysis: RSF
##           Family: surv
```

```
class(y.pred) # Cannot plot in ggplot. Need to convert to df.
```

```
## [1] "rfsrc"    "predict" "surv"
```

```
# Convert to a df
y_pred <- tibble(
  time.interest = y.pred$time.interest,
  # Change legend label here instead...
  PeakVO2_12.8 = y.pred$survival[1,],
  PeakVO2_19.3 = y.pred$survival[2,] # Sth is off. These two rows are the same.
)
```

```
# Convert to long table for plotting purpose
y_pred_long <- pivot_longer(y_pred, cols = 2:3, names_to = "patient", values_to = "time")
```

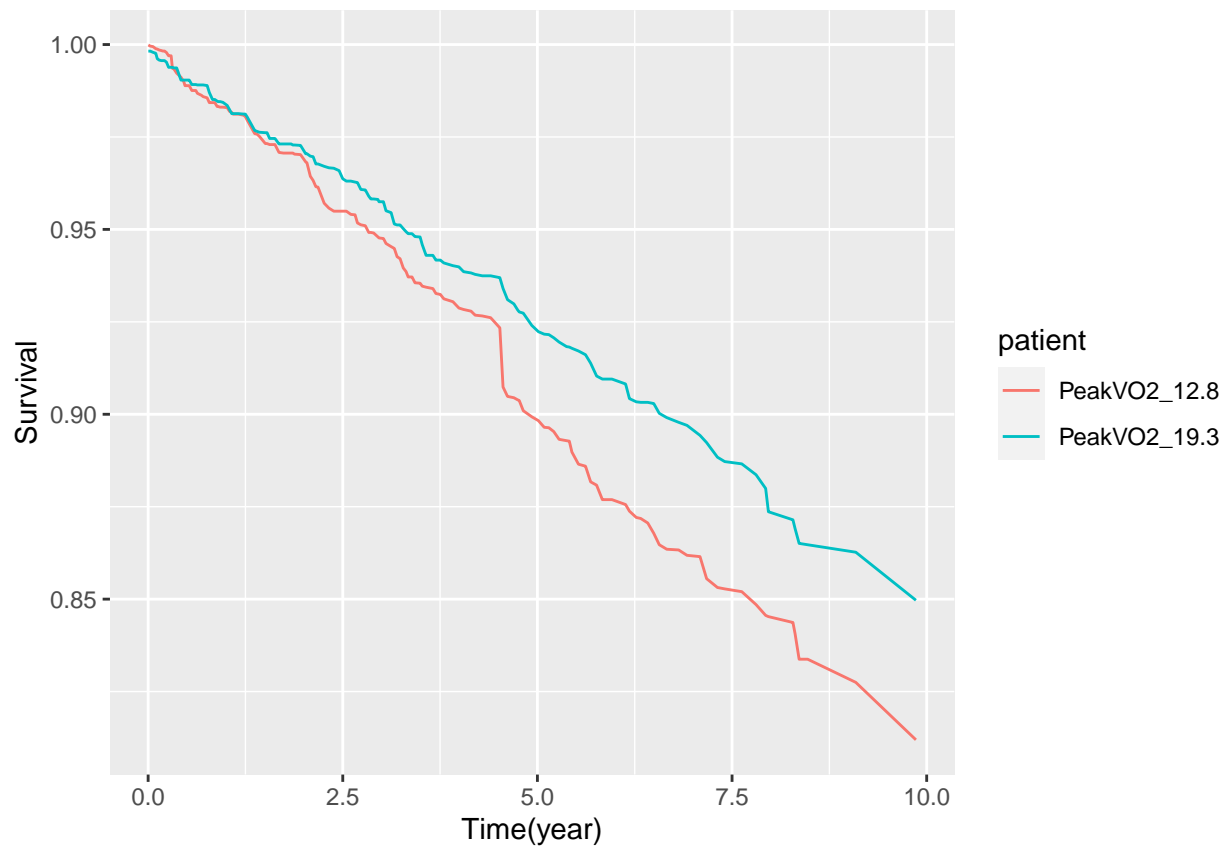
```
# Change patient to factor
y_pred_long$patient = factor(y_pred_long$patient)
```

```
#which(y_pred_long$patient == "survival")
```

```
#str(y_pred_long)
```

Plot of Predicted survival curves of two hypothetical individuals:

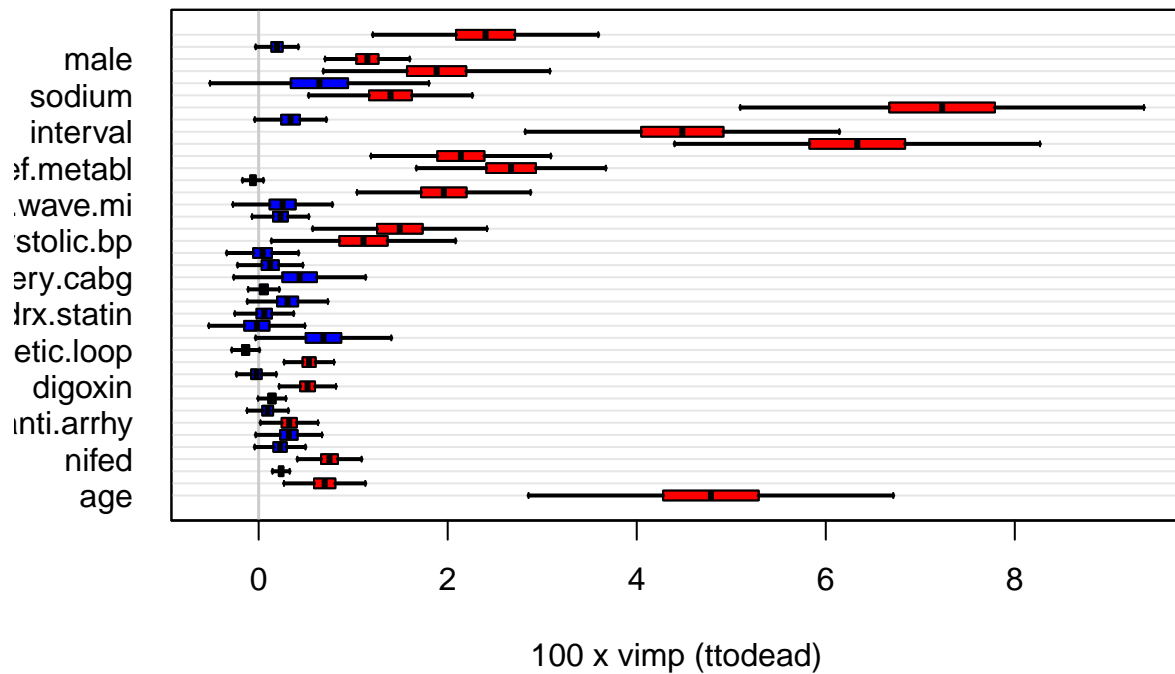
```
ggplot(data = y_pred_long, aes(x = time.interest, y = time, color = patient)) +
  geom_line() +
  labs(x = "Time(year)", y = "Survival") -> p1 #+
  #scale_fill_discrete(labels = c("A", "B")) # Does not work
p1
```



```
# Calculate VIMP (Variable Importance)
jk.obj <- subsample(obj)
```

```
## |
```

```
# Plot it
p2 <- plot(jk.obj, xlab = "Variable Importance (x 100)")
```



p2

```
## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
## [26] 26 27 28 29 30 31 32 33 34 35 36 37 38 39
```

Ran out of time to convert it to ggplot.

```
# jk.obj
```

```
#as.data.frame(jk.obj)
```

```
#ggplot(jk.obj)
```

```
# Write results
library(here)
```

```
## here() starts at /Users/franceslinyc/Random-Survival-Forests-with-Liver-data-2022
```

```
write_rds(p1, here("results", "p1.rds"))
write_rds(p2, here("results", "p2.rds"))
```

Reference

H. Ishwaran, M. S. Lauer, E. H. Blackstone, M. Lu, and U. B. Kogalur. 2021. “randomForestSRC: random survival forests vignette.” <https://luminwin.github.io/randomForestSRC/articles/survival.html>.