# Survival

Zander Bonnet

2024-05-29

```
echocardiogram <- read.csv("~/Desktop/GCU/DSC_520/Data/echocardiogram/echocardiogram.dat
a", header=FALSE)
colnames(echocardiogram) <- c('survival','still-alive','age-at-heart-attack',
                        "pericardial-effusion","fractional-shortening",
                        "epss","lvdd","wall-motion-score",
                        "wall-motion-index","mult","name",
                        "group","alive-at-1")
echocardiogram[,(c('name', 'group'))] <- list(NULL)

suppressWarnings(echocardiogram$survival <- as.numeric(echocardiogram$survival))

echocardiogram$`still-alive` <- factor(echocardiogram$`still-alive`, levels = c(0,1))

suppressWarnings(echocardiogram$`age-at-heart-attack` <- as.numeric(echocardiogram$`age-
at-heart-attack`))

echocardiogram$`pericardial-effusion` <- factor(echocardiogram$`pericardial-effusion`, l
evels = c(0,1))

suppressWarnings(echocardiogram$`fractional-shortening` <- as.numeric(echocardiogram$`fr
actional-shortening`))

suppressWarnings(echocardiogram$epss <- as.numeric(echocardiogram$epss))

suppressWarnings(echocardiogram$lvdd <- as.numeric(echocardiogram$lvdd))

suppressWarnings(echocardiogram$`wall-motion-score` <- as.numeric(echocardiogram$`wall-m
otion-score`))

suppressWarnings(echocardiogram$`wall-motion-index` <- as.numeric(echocardiogram$`wall-m
otion-index`))

suppressWarnings(echocardiogram$mult <- as.numeric(echocardiogram$mult))

echocardiogram <- na.omit(echocardiogram)

echocardiogram$`alive-at-1` <- ifelse((echocardiogram$survival >= 12 | echocardiogram$`s
till-alive` == 1),1,0)

echocardiogram$`alive-at-1` <- factor(echocardiogram$`alive-at-1`, levels = c(0,1))
echocardiogram <- na.omit(echocardiogram)
head(echocardiogram)
```

```
##   survival still-alive age-at-heart-attack pericardial-effusion
## 1       11           0                  71                    0
## 2       19           0                  72                    0
## 3       16           0                  55                    0
## 4       57           0                  60                    0
## 5       19           1                  57                    0
## 6       26           0                  68                    0
##   fractional-shortening   epss  lvdd wall-motion-score wall-motion-index  mult
## 1                 0.260  9.000 4.600                14              1.00 1.000
## 2                 0.380  6.000 4.100                14              1.70 0.588
## 3                 0.260  4.000 3.420                14              1.00 1.000
## 4                 0.253 12.062 4.603                16              1.45 0.788
## 5                 0.160 22.000 5.750                18              2.25 0.571
## 6                 0.260  5.000 4.310                12              1.00 0.857
##   alive-at-1
## 1          0
## 2          1
## 3          1
## 4          1
## 5          1
## 6          1
```

First I must clean the dataset, so that it can be used. I assigned every factor its desired data type, and fixed the alive at 1 year factor as it can be easily manually filled. I then removed the data with missing values since they tended to have multiple missing data points if they were missing one.

```
# Load in libraries
library(knitr)
library(devtools)
```

```
## Loading required package: usethis
```

```
library(survival)
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.1
## ✔ purrr     1.0.2
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
to become errors
```

```
library(survminer)
```

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

```
library(ggplot2)
library(ggfortify)
```

```
kable(echocardiogram)
```

| | survival | still-alive | age-at-heart-attack | pericardial-effusion | fractional-shortening | epss | lvdd | wall-motion-score | wall-motion-index | mult | alive-at-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11.00 | 0 | 71.000 | 0 | 0.260 | 9.000 | 4.600 | 14.00 | 1.000 | 1.000 | 0 |
| 2 | 19.00 | 0 | 72.000 | 0 | 0.380 | 6.000 | 4.100 | 14.00 | 1.700 | 0.588 | 1 |
| 3 | 16.00 | 0 | 55.000 | 0 | 0.260 | 4.000 | 3.420 | 14.00 | 1.000 | 1.000 | 1 |
| 4 | 57.00 | 0 | 60.000 | 0 | 0.253 | 12.062 | 4.603 | 16.00 | 1.450 | 0.788 | 1 |
| 5 | 19.00 | 1 | 57.000 | 0 | 0.160 | 22.000 | 5.750 | 18.00 | 2.250 | 0.571 | 1 |
| 6 | 26.00 | 0 | 68.000 | 0 | 0.260 | 5.000 | 4.310 | 12.00 | 1.000 | 0.857 | 1 |
| 7 | 13.00 | 0 | 62.000 | 0 | 0.230 | 31.000 | 5.430 | 22.50 | 1.875 | 0.857 | 1 |
| 8 | 50.00 | 0 | 60.000 | 0 | 0.330 | 8.000 | 5.250 | 14.00 | 1.000 | 1.000 | 1 |
| 9 | 19.00 | 0 | 46.000 | 0 | 0.340 | 0.000 | 5.090 | 16.00 | 1.140 | 1.003 | 1 |
| 10 | 25.00 | 0 | 54.000 | 0 | 0.140 | 13.000 | 4.490 | 15.50 | 1.190 | 0.930 | 1 |
| 11 | 10.00 | 1 | 77.000 | 0 | 0.130 | 16.000 | 4.230 | 18.00 | 1.800 | 0.714 | 1 |
| 12 | 52.00 | 0 | 62.000 | 1 | 0.450 | 9.000 | 3.600 | 16.00 | 1.140 | 1.003 | 1 |
| 13 | 52.00 | 0 | 73.000 | 0 | 0.330 | 6.000 | 4.000 | 14.00 | 1.000 | 1.000 | 1 |
| 14 | 44.00 | 0 | 60.000 | 0 | 0.150 | 10.000 | 3.730 | 14.00 | 1.000 | 1.000 | 1 |
| 15 | 0.50 | 1 | 62.000 | 0 | 0.120 | 23.000 | 5.800 | 11.67 | 2.330 | 0.358 | 1 |
| 16 | 24.00 | 0 | 55.000 | 1 | 0.250 | 12.063 | 4.290 | 14.00 | 1.000 | 1.000 | 1 |
| 17 | 0.50 | 1 | 69.000 | 1 | 0.260 | 11.000 | 4.650 | 18.00 | 1.640 | 0.784 | 1 |
| 18 | 0.50 | 1 | 62.529 | 1 | 0.070 | 20.000 | 5.200 | 24.00 | 2.000 | 0.857 | 1 |

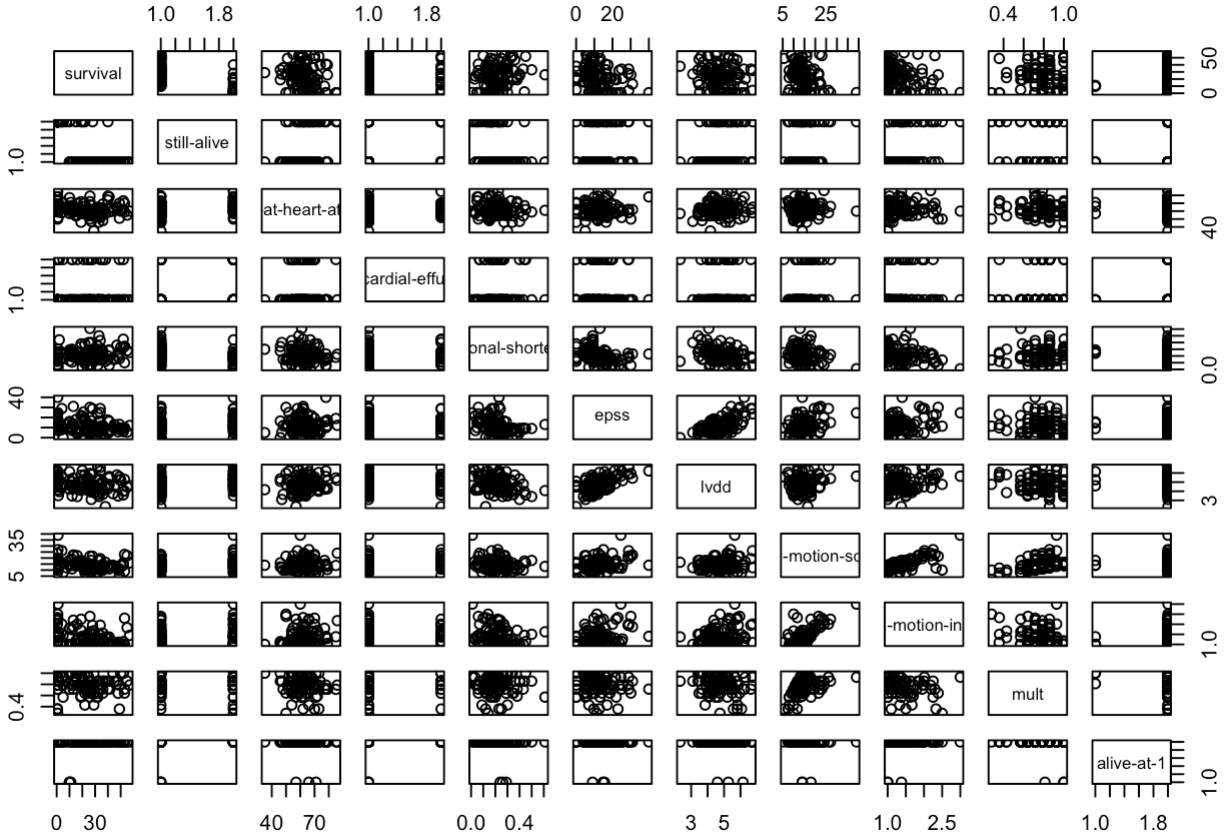
| | survival | still-alive | age-at-heart-attack | pericardial-effusion | fractional-shortening | epss | lvdd | wall-motion-score | wall-motion-index | mult | alive-at-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | 22.00 | 1 | 66.000 | 0 | 0.090 | 17.000 | 5.819 | 8.00 | 1.333 | 0.429 | 1 |
| 20 | 1.00 | 1 | 66.000 | 1 | 0.220 | 15.000 | 5.400 | 27.00 | 2.250 | 0.857 | 1 |
| 21 | 0.75 | 1 | 69.000 | 0 | 0.150 | 12.000 | 5.390 | 19.50 | 1.625 | 0.857 | 1 |
| 22 | 0.75 | 1 | 85.000 | 1 | 0.180 | 19.000 | 5.460 | 13.83 | 1.380 | 0.710 | 1 |
| 23 | 0.50 | 1 | 73.000 | 0 | 0.230 | 12.733 | 6.060 | 7.50 | 1.500 | 0.360 | 1 |
| 24 | 5.00 | 1 | 71.000 | 0 | 0.170 | 0.000 | 4.650 | 8.00 | 1.000 | 0.570 | 1 |
| 25 | 48.00 | 0 | 64.000 | 0 | 0.190 | 5.900 | 3.480 | 10.00 | 1.110 | 0.640 | 1 |
| 26 | 29.00 | 0 | 54.000 | 0 | 0.300 | 7.000 | 3.850 | 10.00 | 1.667 | 0.430 | 1 |
| 27 | 29.00 | 0 | 35.000 | 0 | 0.300 | 5.000 | 4.170 | 14.00 | 1.000 | 1.000 | 1 |
| 30 | 36.00 | 0 | 55.000 | 1 | 0.210 | 4.200 | 4.160 | 14.00 | 1.560 | 0.640 | 1 |
| 32 | 1.00 | 1 | 52.000 | 1 | 0.170 | 17.200 | 5.320 | 14.00 | 1.170 | 0.857 | 1 |
| 34 | 27.00 | 0 | 47.000 | 0 | 0.400 | 5.120 | 3.100 | 12.00 | 1.000 | 0.857 | 1 |
| 36 | 26.00 | 0 | 61.000 | 0 | 0.610 | 13.100 | 4.070 | 13.00 | 1.625 | 0.571 | 1 |
| 41 | 32.00 | 0 | 54.000 | 0 | 0.350 | 9.300 | 3.630 | 11.00 | 1.222 | 0.640 | 1 |
| 42 | 16.00 | 0 | 70.000 | 1 | 0.270 | 4.700 | 4.490 | 22.00 | 2.000 | 0.786 | 1 |
| 43 | 40.00 | 0 | 79.000 | 0 | 0.150 | 17.500 | 4.270 | 13.00 | 1.300 | 0.714 | 1 |
| 45 | 2.00 | 1 | 67.000 | 1 | 0.440 | 9.000 | 3.960 | 17.50 | 1.450 | 0.857 | 1 |
| 48 | 20.00 | 1 | 59.000 | 0 | 0.030 | 21.300 | 6.290 | 17.00 | 1.310 | 0.928 | 1 |
| 54 | 10.00 | 0 | 57.000 | 0 | 0.240 | 14.800 | 5.260 | 18.00 | 1.380 | 0.812 | 0 |
| 55 | 12.00 | 0 | 58.000 | 0 | 0.300 | 9.400 | 3.490 | 14.00 | 1.000 | 1.000 | 1 |
| 56 | 1.00 | 1 | 60.000 | 0 | 0.010 | 24.600 | 5.650 | 39.00 | 3.000 | 0.928 | 1 |
| 57 | 10.00 | 0 | 66.000 | 0 | 0.290 | 15.600 | 6.150 | 14.00 | 1.000 | 1.000 | 0 |
| 58 | 45.00 | 0 | 63.000 | 0 | 0.150 | 13.000 | 4.570 | 13.00 | 1.080 | 0.857 | 1 |
| 59 | 22.00 | 0 | 57.000 | 0 | 0.130 | 18.600 | 4.370 | 12.33 | 1.370 | 0.642 | 1 |
| 60 | 53.00 | 0 | 70.000 | 0 | 0.100 | 9.800 | 5.300 | 23.00 | 2.300 | 0.714 | 1 |
| 62 | 26.00 | 0 | 79.000 | 0 | 0.170 | 11.900 | 5.150 | 10.50 | 1.050 | 0.714 | 1 |
| 64 | 26.00 | 0 | 72.000 | 0 | 0.187 | 12.000 | 5.020 | 13.00 | 1.180 | 0.785 | 1 |
| 65 | 0.50 | 1 | 59.000 | 0 | 0.130 | 16.400 | 4.960 | 17.83 | 1.370 | 0.928 | 1 |
| 66 | 12.00 | 0 | 67.000 | 1 | 0.110 | 10.300 | 4.680 | 11.00 | 1.000 | 0.785 | 1 |

| | survival | still-alive | age-at-heart-attack | pericardial-effusion | fractional-shortening | epss | lvdd | wall-motion-score | wall-motion-index | mult | alive-at-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 67 | 49.00 | 0 | 51.000 | 0 | 0.160 | 13.200 | 5.260 | 11.00 | 1.000 | 0.786 | 1 |
| 68 | 0.75 | 1 | 50.000 | 0 | 0.140 | 11.400 | 4.750 | 10.00 | 2.500 | 0.280 | 1 |
| 69 | 49.00 | 0 | 70.000 | 1 | 0.250 | 9.700 | 5.570 | 5.50 | 1.100 | 0.357 | 1 |
| 70 | 47.00 | 0 | 65.000 | 0 | 0.360 | 8.800 | 5.780 | 12.00 | 1.000 | 0.857 | 1 |
| 71 | 41.00 | 0 | 78.000 | 0 | 0.060 | 16.100 | 5.620 | 13.67 | 1.367 | 0.714 | 1 |
| 72 | 0.25 | 1 | 86.000 | 0 | 0.225 | 12.200 | 5.200 | 24.00 | 2.180 | 0.786 | 1 |
| 73 | 33.00 | 0 | 56.000 | 0 | 0.250 | 11.000 | 4.720 | 11.00 | 1.000 | 0.785 | 1 |
| 74 | 29.00 | 0 | 60.000 | 0 | 0.120 | 10.200 | 4.310 | 15.00 | 1.670 | 0.640 | 1 |
| 75 | 41.00 | 0 | 59.000 | 0 | 0.290 | 7.500 | 4.750 | 13.00 | 1.080 | 0.857 | 1 |
| 76 | 26.00 | 0 | 50.000 | 0 | 0.060 | 30.100 | 5.950 | 21.50 | 2.390 | 0.643 | 1 |
| 77 | 15.00 | 0 | 54.000 | 0 | 0.217 | 17.900 | 4.540 | 16.50 | 1.180 | 1.000 | 1 |
| 78 | 0.25 | 1 | 68.000 | 0 | 0.220 | 21.700 | 4.850 | 15.00 | 1.150 | 0.928 | 1 |
| 80 | 12.00 | 0 | 64.000 | 0 | 0.200 | 7.100 | 4.580 | 14.00 | 1.000 | 1.000 | 1 |
| 81 | 32.00 | 0 | 63.000 | 0 | 0.200 | 5.000 | 5.200 | 8.00 | 1.000 | 0.570 | 1 |
| 82 | 32.00 | 0 | 65.000 | 0 | 0.060 | 23.600 | 6.740 | 12.00 | 1.090 | 0.785 | 1 |
| 83 | 27.00 | 0 | 54.000 | 1 | 0.070 | 16.800 | 4.160 | 18.00 | 1.500 | 0.857 | 1 |
| 84 | 23.00 | 0 | 62.000 | 0 | 0.250 | 6.000 | 4.480 | 11.00 | 1.000 | 0.786 | 1 |
| 85 | 0.75 | 1 | 78.000 | 0 | 0.050 | 10.000 | 4.440 | 15.00 | 1.360 | 0.786 | 1 |
| 87 | 34.00 | 0 | 52.000 | 0 | 0.140 | 25.000 | 6.210 | 11.50 | 1.150 | 0.714 | 1 |
| 88 | 1.00 | 1 | 73.000 | 0 | 0.050 | 14.800 | 4.140 | 15.50 | 1.410 | 0.786 | 1 |
| 89 | 21.00 | 1 | 70.000 | 1 | 0.160 | 19.200 | 5.250 | 11.00 | 1.000 | 0.786 | 1 |
| 90 | 55.00 | 0 | 55.000 | 0 | 0.280 | 5.500 | 4.480 | 22.00 | 1.830 | 0.857 | 1 |
| 91 | 15.00 | 1 | 60.000 | 0 | 0.180 | 8.700 | 4.560 | 13.50 | 1.040 | 0.928 | 1 |
| 92 | 0.50 | 1 | 67.000 | 0 | 0.155 | 11.300 | 5.160 | 13.00 | 1.000 | 0.928 | 1 |
| 93 | 35.00 | 0 | 64.000 | 0 | 0.300 | 6.600 | 4.360 | 14.00 | 1.270 | 0.786 | 1 |
| 94 | 53.00 | 0 | 59.000 | 0 | 0.344 | 9.100 | 4.040 | 9.00 | 1.000 | 0.643 | 1 |
| 95 | 33.00 | 0 | 46.000 | 0 | 0.272 | 16.500 | 5.360 | 12.67 | 1.060 | 0.857 | 1 |
| 97 | 33.00 | 0 | 63.000 | 0 | 0.250 | 5.600 | 3.870 | 18.00 | 1.500 | 0.857 | 1 |
| 98 | 40.00 | 1 | 74.000 | 0 | 0.200 | 4.800 | 4.560 | 12.50 | 1.040 | 0.857 | 1 |

| | survival | still-alive | age-at-heart-attack | pericardial-effusion | fractional-shortening | epss | lvdd | wall-motion-score | wall-motion-index | mult | alive-at-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 99 | 33.00 | 0 | 59.000 | 0 | 0.500 | 9.100 | 3.420 | 18.00 | 1.500 | 0.857 | 1 |
| 100 | 5.00 | 1 | 65.000 | 1 | 0.160 | 8.500 | 5.470 | 16.00 | 1.450 | 0.786 | 1 |
| 101 | 4.00 | 1 | 58.000 | 0 | 0.170 | 28.900 | 6.730 | 26.08 | 2.010 | 0.928 | 1 |
| 104 | 22.00 | 0 | 70.000 | 0 | 0.380 | 0.000 | 4.550 | 10.00 | 1.000 | 0.714 | 1 |
| 105 | 25.00 | 0 | 62.000 | 0 | 0.258 | 11.800 | 4.870 | 11.00 | 1.000 | 0.786 | 1 |
| 106 | 1.25 | 1 | 63.000 | 0 | 0.300 | 6.900 | 3.520 | 18.16 | 1.510 | 0.857 | 1 |
| 107 | 24.00 | 0 | 59.000 | 0 | 0.170 | 14.300 | 5.490 | 13.50 | 1.500 | 0.643 | 1 |
| 108 | 25.00 | 0 | 57.000 | 0 | 0.228 | 9.700 | 4.290 | 11.00 | 1.000 | 0.786 | 1 |
| 109 | 24.00 | 0 | 57.000 | 0 | 0.036 | 7.000 | 4.120 | 13.50 | 1.230 | 0.786 | 1 |
| 110 | 0.75 | 1 | 78.000 | 0 | 0.230 | 40.000 | 6.230 | 14.00 | 1.400 | 0.714 | 1 |
| 111 | 3.00 | 1 | 62.000 | 0 | 0.260 | 7.600 | 4.420 | 14.00 | 1.000 | 1.000 | 1 |
| 112 | 27.00 | 0 | 62.000 | 0 | 0.220 | 12.100 | 3.920 | 11.00 | 1.000 | 0.785 | 1 |
| 113 | 13.00 | 0 | 66.000 | 0 | 0.240 | 13.600 | 4.380 | 22.00 | 2.200 | 0.714 | 1 |
| 114 | 36.00 | 0 | 61.000 | 0 | 0.270 | 9.000 | 4.060 | 12.00 | 1.000 | 0.857 | 1 |
| 115 | 25.00 | 0 | 59.000 | 1 | 0.400 | 9.200 | 5.360 | 12.00 | 1.000 | 0.857 | 1 |
| 116 | 27.00 | 0 | 57.000 | 0 | 0.290 | 9.400 | 4.770 | 9.00 | 1.000 | 0.640 | 1 |
| 117 | 34.00 | 0 | 62.000 | 1 | 0.190 | 28.900 | 6.630 | 19.50 | 1.950 | 0.714 | 1 |
| 119 | 34.00 | 0 | 54.000 | 0 | 0.430 | 9.300 | 4.790 | 10.00 | 1.000 | 0.714 | 1 |
| 120 | 28.00 | 1 | 62.000 | 1 | 0.240 | 28.600 | 5.860 | 21.50 | 1.950 | 0.786 | 1 |
| 122 | 17.00 | 0 | 64.000 | 0 | 0.150 | 6.600 | 4.170 | 14.00 | 1.270 | 0.786 | 1 |
| 123 | 38.00 | 0 | 57.000 | 1 | 0.120 | 0.000 | 2.320 | 16.50 | 1.375 | 0.857 | 1 |
| 124 | 31.00 | 0 | 61.000 | 0 | 0.180 | 0.000 | 4.480 | 11.00 | 1.375 | 0.570 | 1 |
| 125 | 12.00 | 0 | 61.000 | 1 | 0.190 | 13.200 | 5.040 | 19.00 | 1.730 | 0.786 | 1 |
| 126 | 36.00 | 0 | 48.000 | 0 | 0.150 | 12.000 | 3.660 | 10.00 | 1.000 | 0.714 | 1 |
| 128 | 21.00 | 0 | 61.000 | 0 | 0.140 | 25.500 | 5.160 | 14.00 | 1.270 | 0.786 | 1 |
| 129 | 7.50 | 1 | 64.000 | 0 | 0.240 | 12.900 | 4.720 | 12.00 | 1.000 | 0.857 | 1 |
| 130 | 41.00 | 0 | 64.000 | 0 | 0.280 | 5.400 | 5.470 | 11.00 | 1.100 | 0.714 | 1 |
| 131 | 36.00 | 0 | 69.000 | 0 | 0.200 | 7.000 | 5.050 | 14.50 | 1.210 | 0.857 | 1 |
| 132 | 22.00 | 0 | 57.000 | 0 | 0.140 | 16.100 | 4.360 | 15.00 | 1.360 | 0.786 | 1 |

| | survival | still-alive | age-at-heart-attack | pericardial-effusion | fractional-shortening | epss | lvdd | wall-motion-score | wall-motion-index | mult | alive-at-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 133 | 20.00 | 0 | 62.000 | 0 | 0.150 | 0.000 | 4.510 | 15.50 | 1.409 | 0.786 | 1 |

```
plot(echocardiogram)
```




We can see that there are a handful of possibly correlated factors, and there are no visual extreme outliers.

In this we will aim to answer if there is a difference in survival rates between individuals with percardial effusion and without pericaridal effusion.

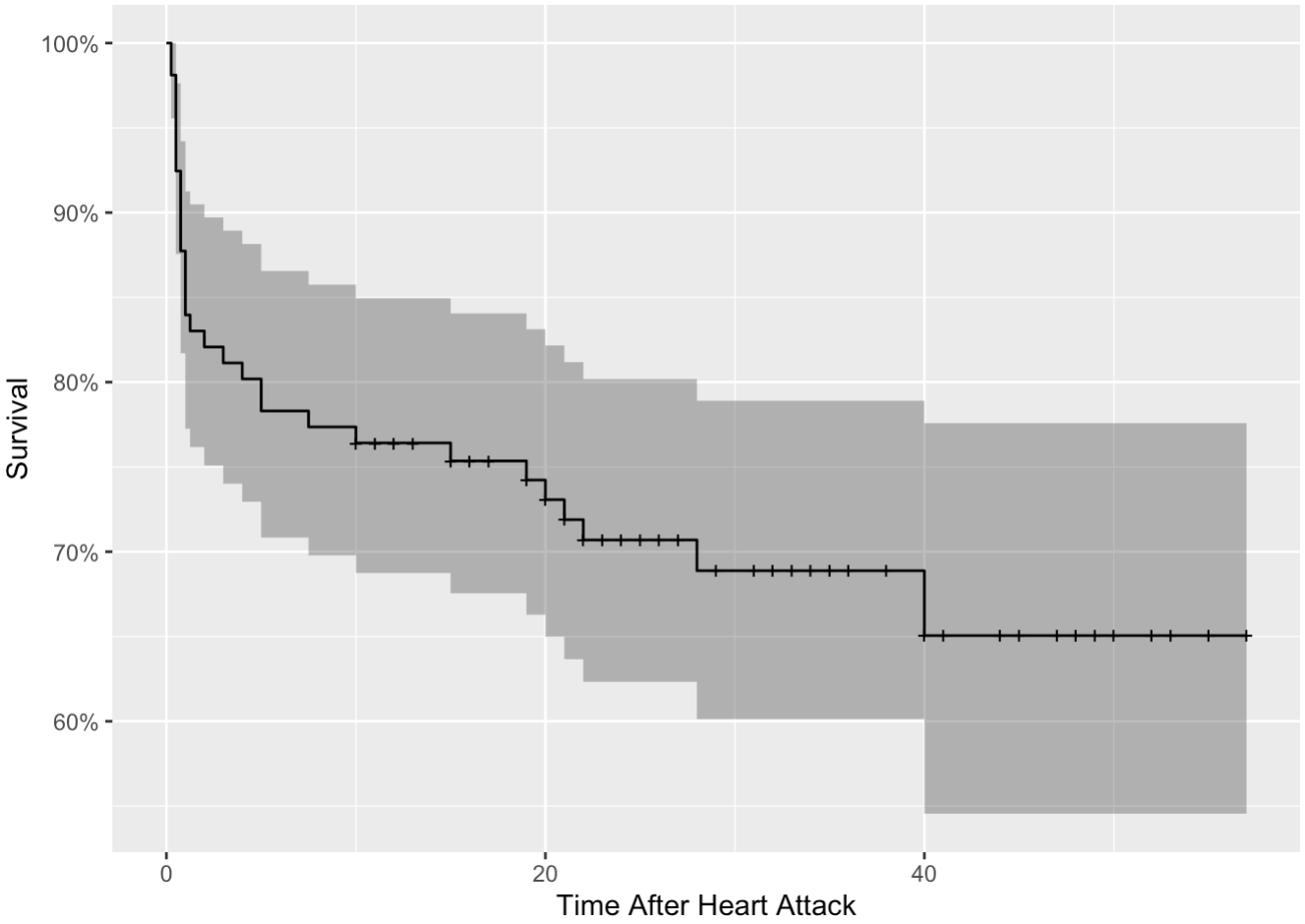
```
sur <- Surv(echocardiogram$survival, as.numeric(as.character(echocardiogram$`still-alive`)))
print(sur)
```

```
##   [1] 11.00+ 19.00+ 16.00+ 57.00+ 19.00  26.00+ 13.00+ 50.00+ 19.00+ 25.00+
##  [11] 10.00  52.00+ 52.00+ 44.00+  0.50  24.00+  0.50   0.50  22.00   1.00
##  [21]  0.75   0.75   0.50   5.00  48.00+ 29.00+ 29.00+ 36.00+  1.00  27.00+
##  [31] 26.00+ 32.00+ 16.00+ 40.00+  2.00  20.00  10.00+ 12.00+  1.00  10.00+
##  [41] 45.00+ 22.00+ 53.00+ 26.00+ 26.00+  0.50  12.00+ 49.00+  0.75  49.00+
##  [51] 47.00+ 41.00+  0.25  33.00+ 29.00+ 41.00+ 26.00+ 15.00+  0.25  12.00+
##  [61] 32.00+ 32.00+ 27.00+ 23.00+  0.75  34.00+  1.00  21.00  55.00+ 15.00
##  [71]  0.50  35.00+ 53.00+ 33.00+ 33.00+ 40.00  33.00+  5.00   4.00  22.00+
##  [81] 25.00+  1.25  24.00+ 25.00+ 24.00+  0.75   3.00  27.00+ 13.00+ 36.00+
##  [91] 25.00+ 27.00+ 34.00+ 34.00+ 28.00  17.00+ 38.00+ 31.00+ 12.00+ 36.00+
## [101] 21.00+  7.50  41.00+ 36.00+ 22.00+ 20.00+
```

```
fit <- survfit(sur ~ 1 , data = echocardiogram )
summary(fit)
```

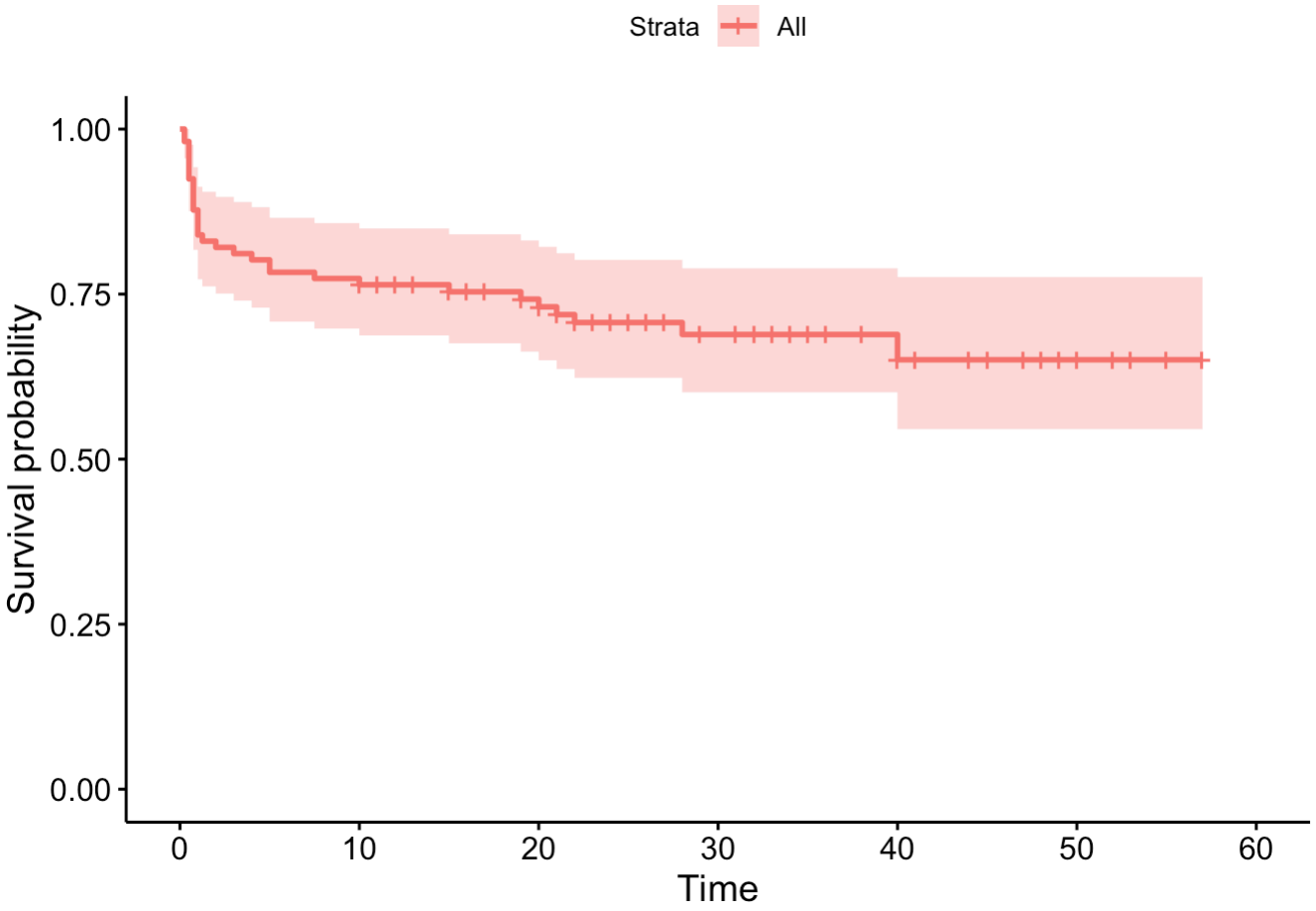
```
## Call: survfit(formula = sur ~ 1, data = echocardiogram)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   0.25    106       2    0.981  0.0132        0.956        1.000
##   0.50    104       6    0.925  0.0257        0.876        0.976
##   0.75     98       5    0.877  0.0319        0.817        0.942
##   1.00     93       4    0.840  0.0356        0.773        0.912
##   1.25     89       1    0.830  0.0365        0.762        0.905
##   2.00     88       1    0.821  0.0373        0.751        0.897
##   3.00     87       1    0.811  0.0380        0.740        0.889
##   4.00     86       1    0.802  0.0387        0.729        0.881
##   5.00     85       2    0.783  0.0400        0.708        0.866
##   7.50     83       1    0.774  0.0406        0.698        0.858
##  10.00     82       1    0.764  0.0412        0.687        0.849
##  15.00     72       1    0.754  0.0420        0.676        0.841
##  19.00     67       1    0.742  0.0429        0.663        0.831
##  20.00     64       1    0.731  0.0437        0.650        0.822
##  21.00     62       1    0.719  0.0446        0.637        0.812
##  22.00     60       1    0.707  0.0454        0.623        0.802
##  28.00     39       1    0.689  0.0477        0.601        0.789
##  40.00     18       1    0.651  0.0584        0.546        0.776
```

```
autoplot(fit, ylab = 'Survival',  xlab = 'Time After Heart Attack', legLabs = c('dea
d','alive'))
```




When analyzing survival over time we can see that the more time after the heart attack the less likely the surviuval is.

```
ggsurvplot(fit = fit, data = echocardiogram)
```

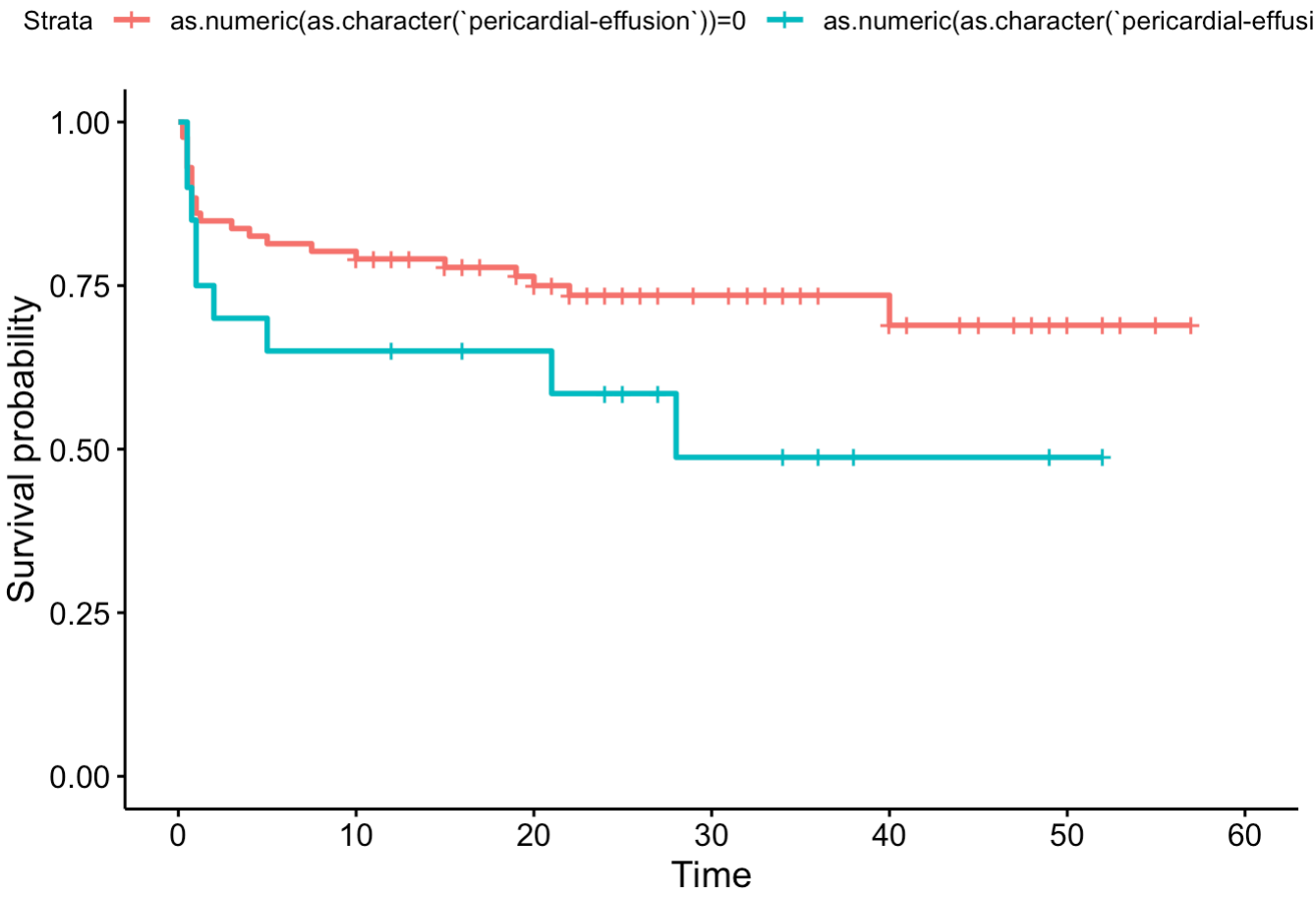



This shows a similar story with the error bar included.

```
fit2 <- survfit(sur ~ as.numeric(as.character(`pericardial-effusion`)) , data = echocard
iogram )
summary(fit2)
```

```
## Call: survfit(formula = sur ~ as.numeric(as.character(`pericardial-effusion`)),
##     data = echocardiogram)
##
##                as.numeric(as.character(`pericardial-effusion`))=0
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   0.25     86       2    0.977  0.0163        0.945        1.000
##   0.50     84       4    0.930  0.0275        0.878        0.986
##   0.75     80       4    0.884  0.0346        0.819        0.954
##   1.00     76       2    0.860  0.0374        0.790        0.937
##   1.25     74       1    0.849  0.0386        0.776        0.928
##   3.00     73       1    0.837  0.0398        0.763        0.919
##   4.00     72       1    0.826  0.0409        0.749        0.910
##   5.00     71       1    0.814  0.0420        0.736        0.900
##   7.50     70       1    0.802  0.0429        0.722        0.891
##  10.00     69       1    0.791  0.0439        0.709        0.882
##  15.00     61       1    0.778  0.0450        0.694        0.871
##  19.00     57       1    0.764  0.0463        0.679        0.860
##  20.00     54       1    0.750  0.0475        0.662        0.849
##  22.00     51       1    0.735  0.0488        0.646        0.837
##  40.00     16       1    0.689  0.0638        0.575        0.826
##
##                as.numeric(as.character(`pericardial-effusion`))=1
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   0.50     20       2    0.900  0.0671        0.778        1.000
##   0.75     18       1    0.850  0.0798        0.707        1.000
##   1.00     17       2    0.750  0.0968        0.582        0.966
##   2.00     15       1    0.700  0.1025        0.525        0.933
##   5.00     14       1    0.650  0.1067        0.471        0.897
##  21.00     10       1    0.585  0.1141        0.399        0.857
##  28.00      6       1    0.488  0.1302        0.289        0.823
```

```
ggsurvplot(fit = fit2, data = echocardiogram)
```

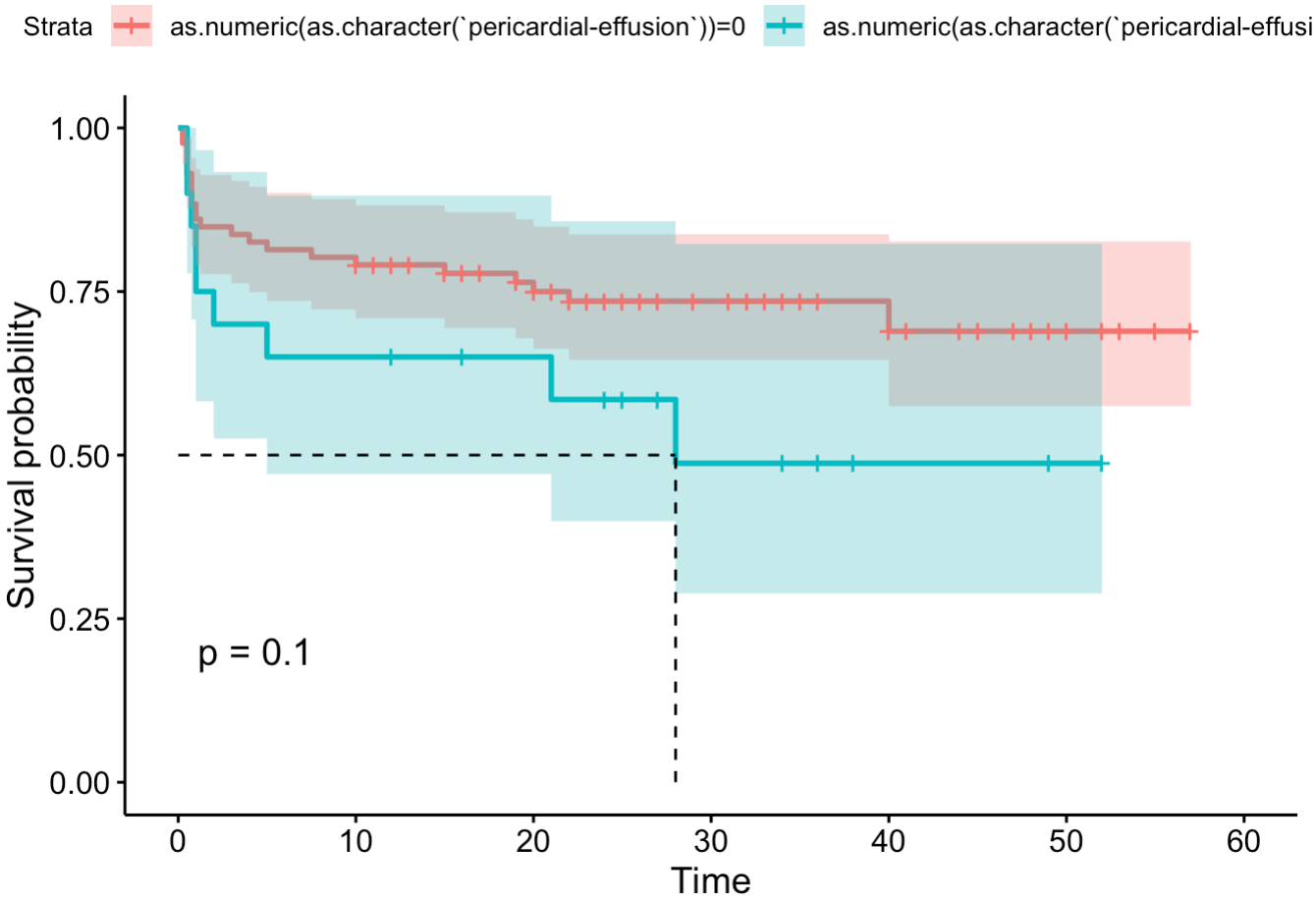



When splitting it between paricardial effusion we can see that having pericardial effusion tends to leave to a lower survival chance over time.

```
dif <- survdiff(sur ~ as.numeric(as.character(`pericardial-effusion`)) , data = echocard
iogram)
dif
```

```
## Call:
## survdiff(formula = sur ~ as.numeric(as.character(`pericardial-effusion`)),
##     data = echocardiogram)
##
##                                                  N Observed Expected
## as.numeric(as.character(`pericardial-effusion`))=0 86       23    26.48
## as.numeric(as.character(`pericardial-effusion`))=1 20        9     5.52
##                                                  (O-E)^2/E (O-E)^2/V
## as.numeric(as.character(`pericardial-effusion`))=0     0.457      2.71
## as.numeric(as.character(`pericardial-effusion`))=1     2.190      2.71
##
##  Chisq= 2.7  on 1 degrees of freedom, p= 0.1
```

```
ggsurvplot(fit = fit2, data = echocardiogram, surv.median.line = 'hv',pval = TRUE, conf.
int = TRUE)
```




When including the error bars we can see that there is a lot of overlap so, there may not be a significant difference in the two groups. By then looking at the log rank test we can see that there is p-value of .1, so there is not significant evidence to prove that the two survival curves are different, at a critical value of .05.

```
write.csv(echocardiogram, 'echo.csv')
```

# Python

```
import lifelines as lf
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
```

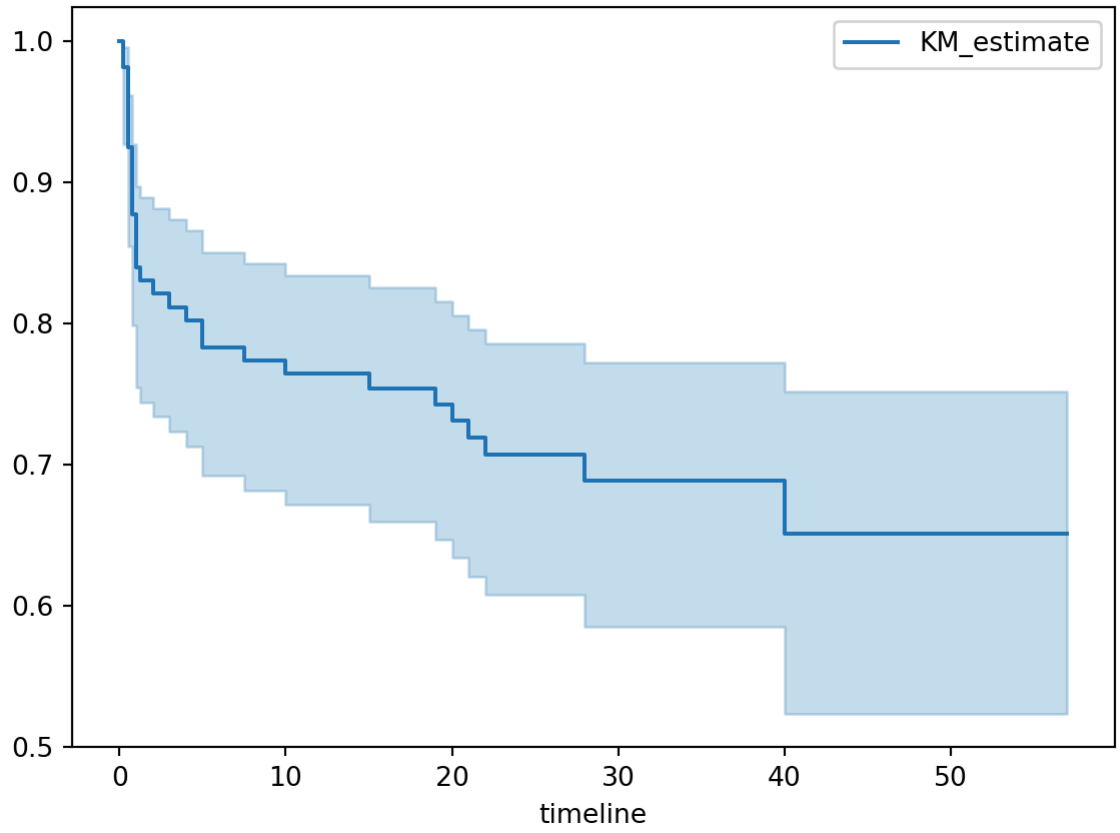
```
echo = pd.read_csv('echo.csv', index_col = 0)
echo.head()
```

```
##    survival  still-alive  ...   mult  alive-at-1
## 1      11.0            0  ...  1.000           0
## 2      19.0            0  ...  0.588           1
## 3      16.0            0  ...  1.000           1
## 4      57.0            0  ...  0.788           1
## 5      19.0            1  ...  0.571           1
##
## [5 rows x 11 columns]
```

```
kmf = lf.KaplanMeierFitter()
kmf.fit(echo['survival'],echo['still-alive'])
```

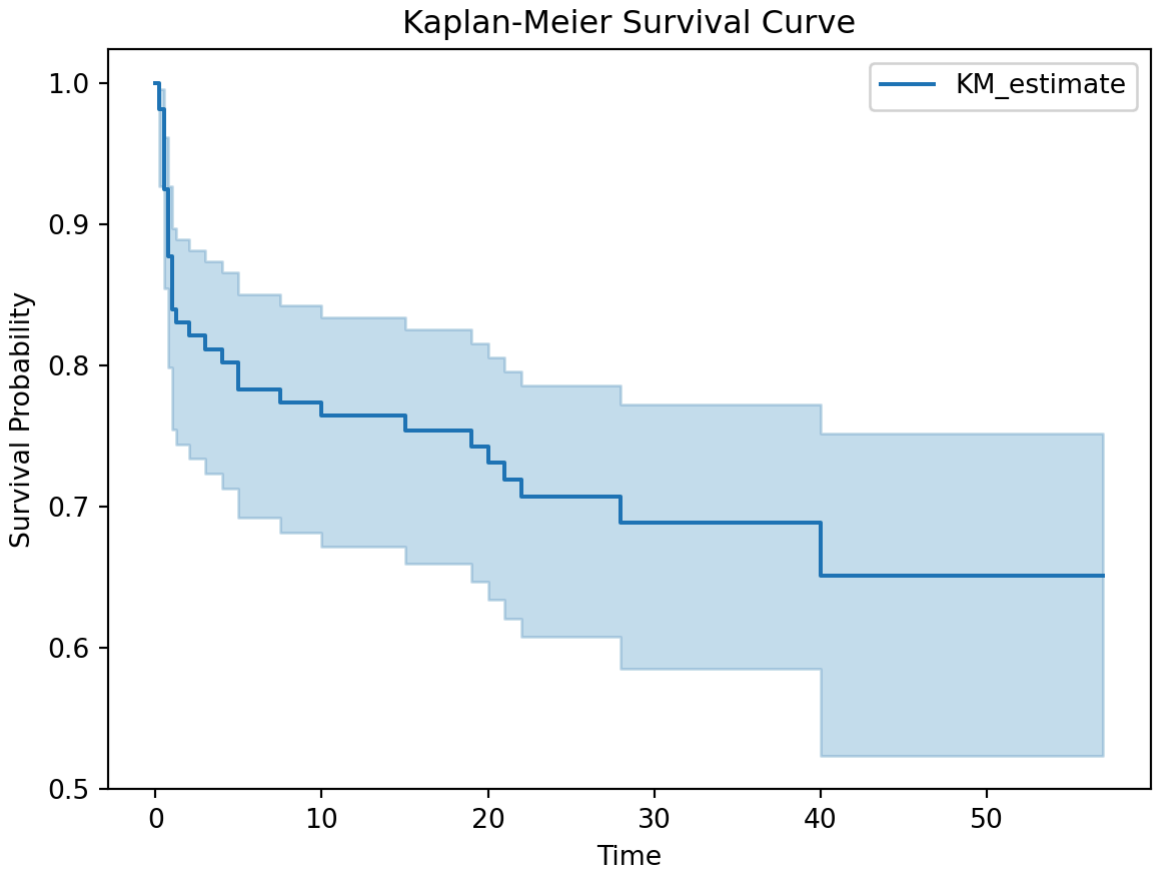
```
## <lifelines.KaplanMeierFitter:"KM_estimate", fitted with 106 total observations, 74 ri
ght-censored observations>
```

```
kmf.plot_survival_function()
plt.show()
```




```
plt.close()
```

```
kmf.plot_survival_function(title= 'Kaplan-Meier Survival Curve', xlabel='Time', ylabel
='Survival Probability')
plt.show()
```




```
plt.close()
```

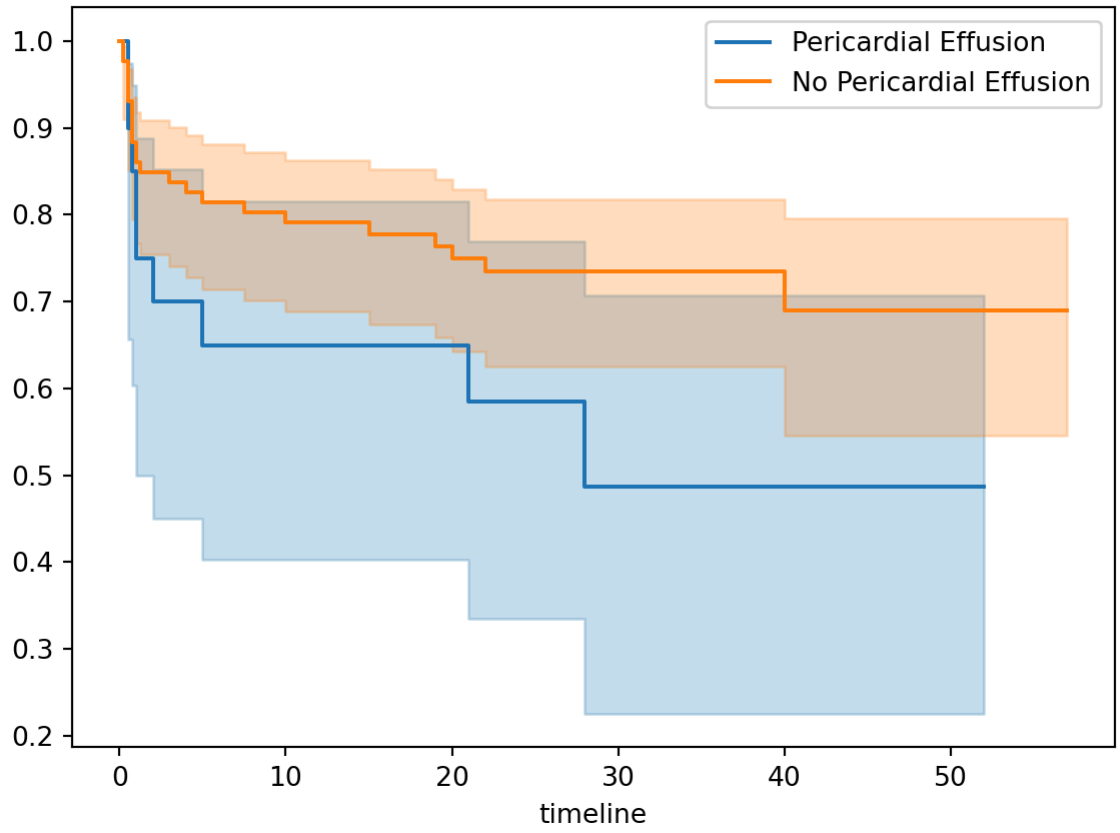
```
ax = plt.subplot(111)

kmf2 = lf.KaplanMeierFitter()
kmf2.fit(echo.loc[echo['pericardial-effusion'] == 1]['survival'],echo.loc[echo['pericard
ial-effusion'] == 1]['still-alive'], label = 'Pericardial Effusion')
```

```
## <lifelines.KaplanMeierFitter:"Pericardial Effusion", fitted with 20 total observation
s, 11 right-censored observations>
```

```
kmf2.plot_survival_function(ax=ax)

kmf2.fit(echo.loc[echo['pericardial-effusion'] == 0]['survival'],echo.loc[echo['pericard
ial-effusion'] == 0]['still-alive'], label = 'No Pericardial Effusion')
```

```
## <lifelines.KaplanMeierFitter:"No Pericardial Effusion", fitted with 86 total observat
ions, 63 right-censored observations>
```

```
kmf2.plot_survival_function(ax=ax)

plt.show()
```




```
plt.close()
```

```
results = lf.statistics.logrank_test(echo.loc[echo['pericardial-effusion'] == 1]['survival'], echo.loc[echo['pericardial-effusion'] == 0]['survival'], event_observed_A=echo.loc[echo['pericardial-effusion'] == 1]['still-alive'], event_observed_B=echo.loc[echo['pericardial-effusion'] == 0]['still-alive'])

results
```

| | |
|---|---|
| **t_0** | -1 |
| **null_distribution** | chi squared |
| **degrees_of_freedom** | 1 |
| **test_name** | logrank_test |

| | test_statistic | p | -log2(p) |
|---|---|---|---|
| **0** | 2.71 | 0.10 | 3.33 |

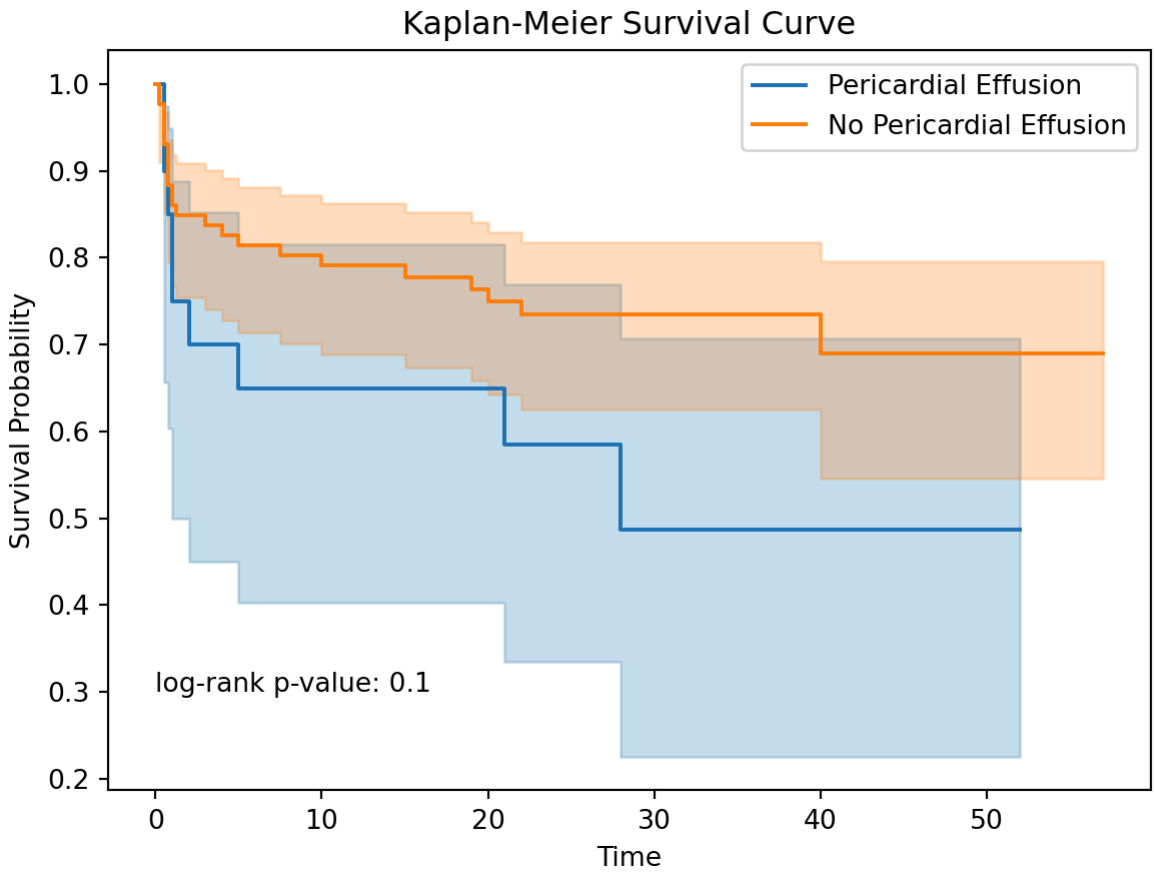
```
ax = plt.subplot(111)

kmf2 = lf.KaplanMeierFitter()
kmf2.fit(echo.loc[echo['pericardial-effusion'] == 1]['survival'],echo.loc[echo['pericard
ial-effusion'] == 1]['still-alive'], label = 'Pericardial Effusion')
```

```
## <lifelines.KaplanMeierFitter:"Pericardial Effusion", fitted with 20 total observation
s, 11 right-censored observations>
```

```
kmf2.plot_survival_function(ax=ax)

kmf2.fit(echo.loc[echo['pericardial-effusion'] == 0]['survival'],echo.loc[echo['pericard
ial-effusion'] == 0]['still-alive'], label = 'No Pericardial Effusion')
```

```
## <lifelines.KaplanMeierFitter:"No Pericardial Effusion", fitted with 86 total observat
ions, 63 right-censored observations>
```

```
kmf2.plot_survival_function(ax=ax,title= 'Kaplan-Meier Survival Curve', xlabel='Time', y
label='Survival Probability')

plt.text(s = "log-rank p-value: {:.2}".format(results.p_value), x = 0, y= .3 )

plt.show()
```




```
plt.close()
```

We came to the same conclusion as in R where there is not a significant difference in the two groups. There is a p-value of .1 which is not significant at the .05 level.

In the end we found that there is no significant difference in survival rates in the groups with pericardial effusion and without pericardial effusion. The dataset is fairly small, so we might be able to make a more definitive conclusion if we were able to increase the sample size.

# References

Echocardiogram. (1989). UCI Machine Learning Repository. https://doi.org/10.24432/C5QW24 (https://doi.org/10.24432/C5QW24).