# Nowacek HW 2-4

```r
library(tidyverse)
library(survival)
library(survminer)
library(tidyr)
library(kableExtra)
library(broom)
```

Problem 4

```r
prison <- read_csv("https://www2.stat.duke.edu/courses/Fall24/sta490.01/data/hw2.csv")
```

```r
prison_fit <- survfit(Surv(week, arrest) ~ mar, data = prison)

ggsurvplot(survfit(Surv(week, arrest) ~ mar, data = prison),
           main = "Kaplan-Meier Estimates by Marriage Status",
           xlab = "Weeks",
           ylab = "Estimated Survival Probability",
           ylim = c(0, 1), conf.int = T, censor = F,
           legend.labs = c("Married", "Not Married"))
```
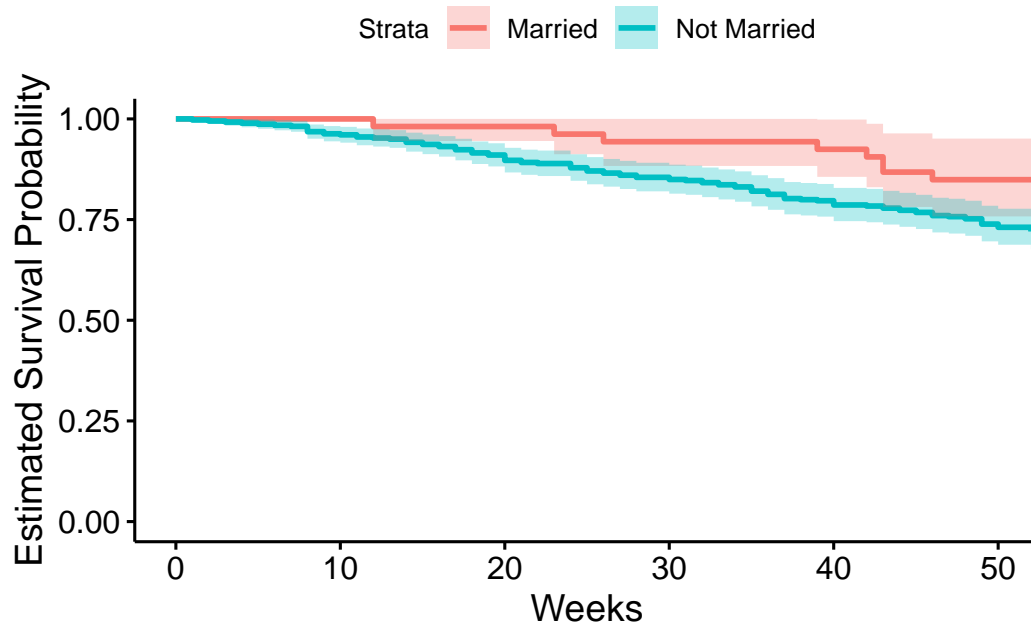
Define what "survival" means here. What are the estimated "survival" times at the 6 month mark for the two groups?

Survival here means making it through a year post-release from prison without recidivism.

```r
summary(prison_fit, time = 24)
```

```
Call: survfit(formula = Surv(week, arrest) ~ mar, data = prison)

                mar=married
        time        n.risk       n.event       survival        std.err lower 95% CI
     24.0000       51.0000        2.0000         0.9623         0.0262         0.9123
upper 95% CI
      1.0000


                mar=not married
        time        n.risk       n.event       survival        std.err lower 95% CI
     24.0000      337.0000       46.0000         0.8786         0.0168         0.8464
upper 95% CI
      0.9121
```

The survival estimate for the married subjects is 0.9623 and 0.8786 for not married subjects.

Describe what you see in the relationship here (no need to do any hypothesis test or anything, we haven't learned any yet!).

The married group is more likely to survive the full year without recidivism, especially towards the beginning, after about 38 weeks, the married group drops in survival probability to come closer to the non-married group which decreases in survival probability fairly steadily over the year.

Why do you think one of the curves seems much "blockier" than the other? (you may attach any files as necessary).

```
prison |>
  group_by(mar) |>
  count() |>
  kable()
```

mar	n
married	53
not married	379

The married group's curve seems blockier because the sample size is smaller, leading to each individual case being more apparent in its influence on the plot.