# Project 1

Clare Flynn

In this analysis, I use data from a study testing the success of different HIV medication by Hammer et al. (1997). The response variable in the study is *time*, which indicates in days the time to death, AIDs diagnosis, or the end of the study. There is also a *censor* variable, which indicates whether each patient made it to the end of the study without a diagnosis or death, or not. The explanatory variable of interest is tx, or treatment group, so whether each patient recieved the treatment with the IDV, or the control treatment without the IDV. There are also many other potentially explanatory variables recorded, such as sex, race, age, Karnofsky score, and IV drug use history.

There are 782 uncensored data points in this study, meaning that 782 people survived and were not diagnosed with AIDs by the end of the study period. There are 69 censored data points, meaning 69 people either died or were diagnosed with AIDs before the end of the study period. Of those censored data, 20 of them died during the study.

The following histogram displays the times to diagnosis, death, or the end of the study.
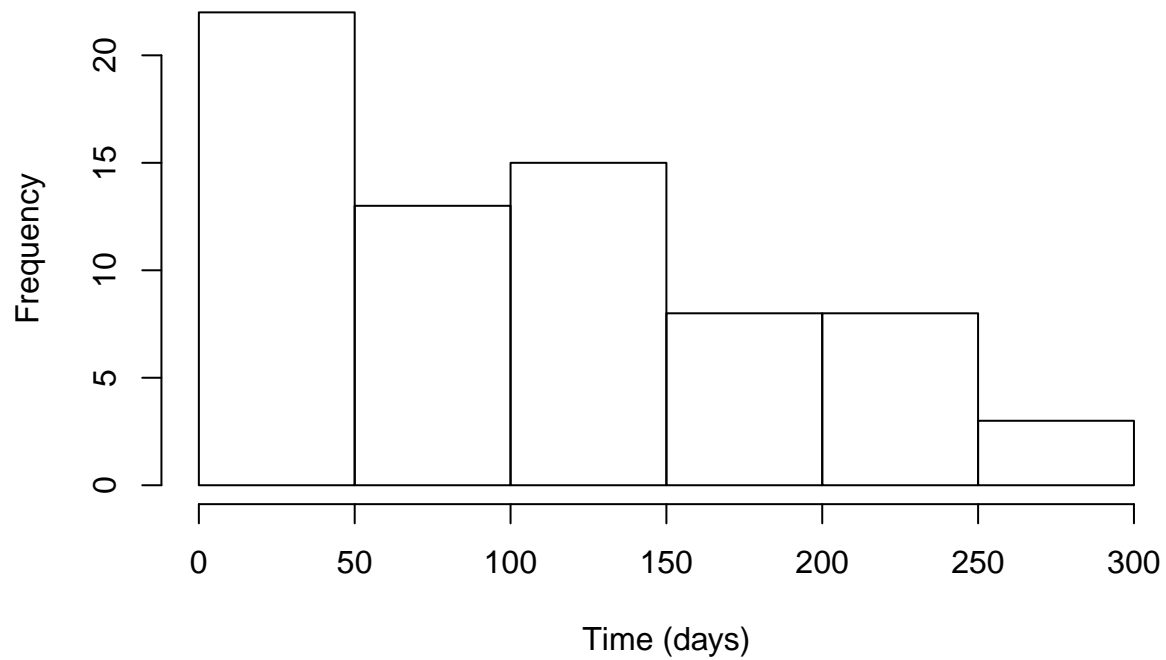
```
hist(aids$time, xlab="Time (days)", main="Time to Diagnosis, Death, or End of Study")
```
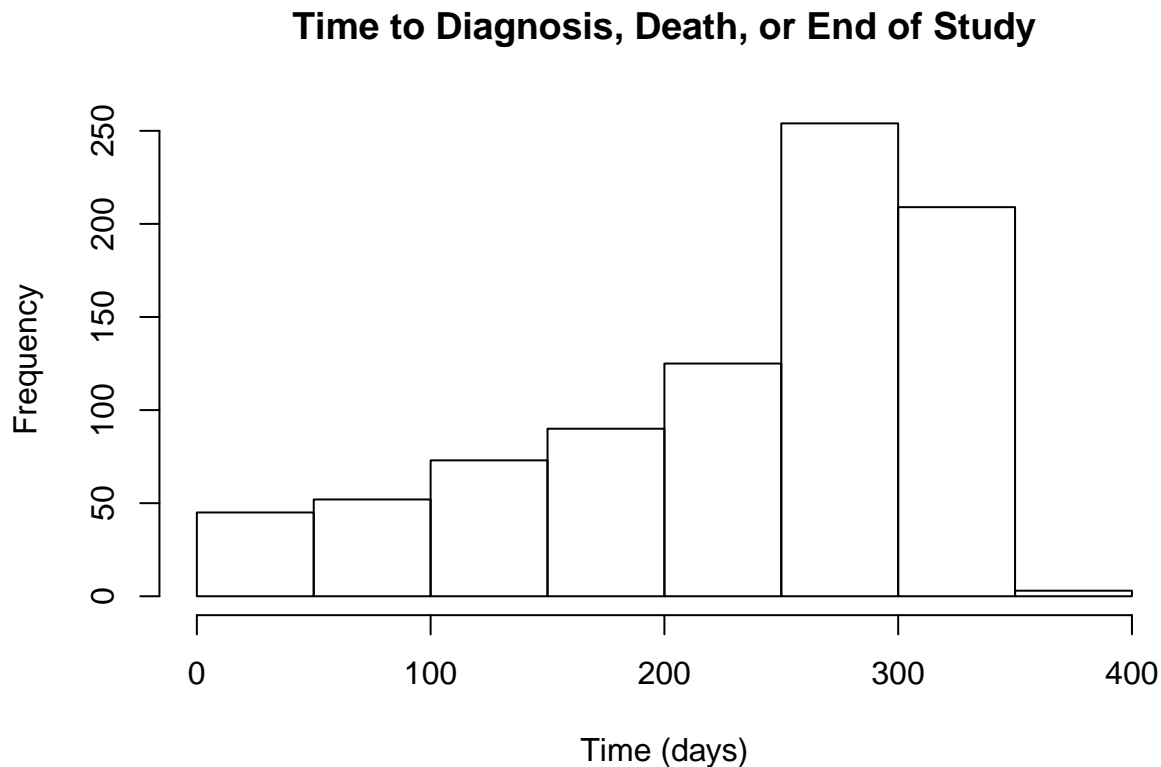

This next histogram shows the time to event only for those who died or were diagnosed during the study period.

```
hist(aids$time[which(aids$censor==1)], xlab="Time (days)", main="Time to Diagnosis or Death")
```

Time to Diagnosis or Death

The following is a Kaplan-Meier Curve for all participants in the study. The curve is not very steep, since very few participants died or were diagnosed with AIDs during the study period.

```
aid.surv <- survfit(Surv(aids$time,  aids$censor)~1, data=aids)
survminer::ggsurvplot(aid.surv, conf.int=TRUE, censor=F) + ggtitle("Overall")
```

Overall
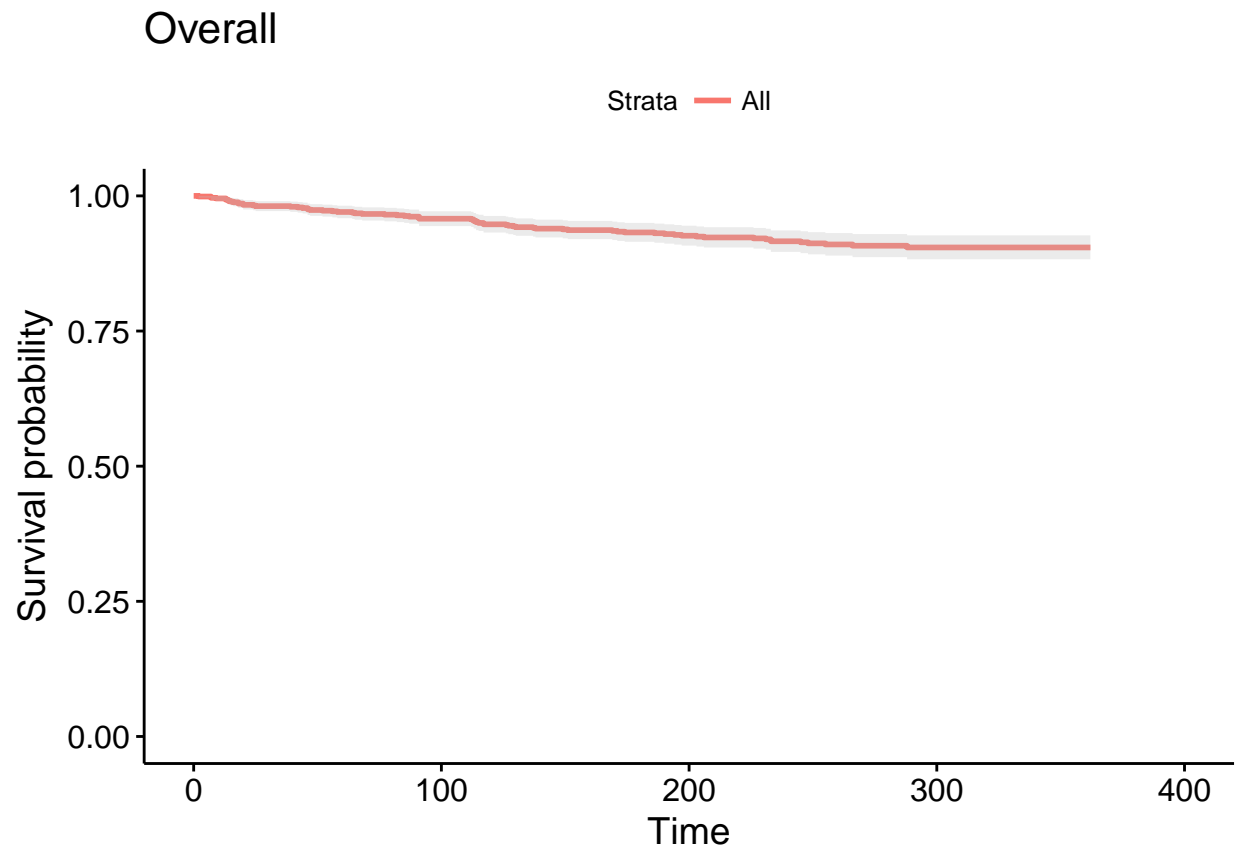
The following are Kaplan-Meier Curves for the groups control and high Karnofsky score (red), control and low Karnofsky score (green), treatment and high Karnofsky score (blue), and treatment and low Karnofsky score (purple).

```r
aids.surv <- survfit(Surv(aids$time,  aids$censor)~tx + k, data=aids)
survminer::ggsurvplot(aids.surv, conf.int=TRUE, censor=F) + ggtitle("Overall")
```

Overall
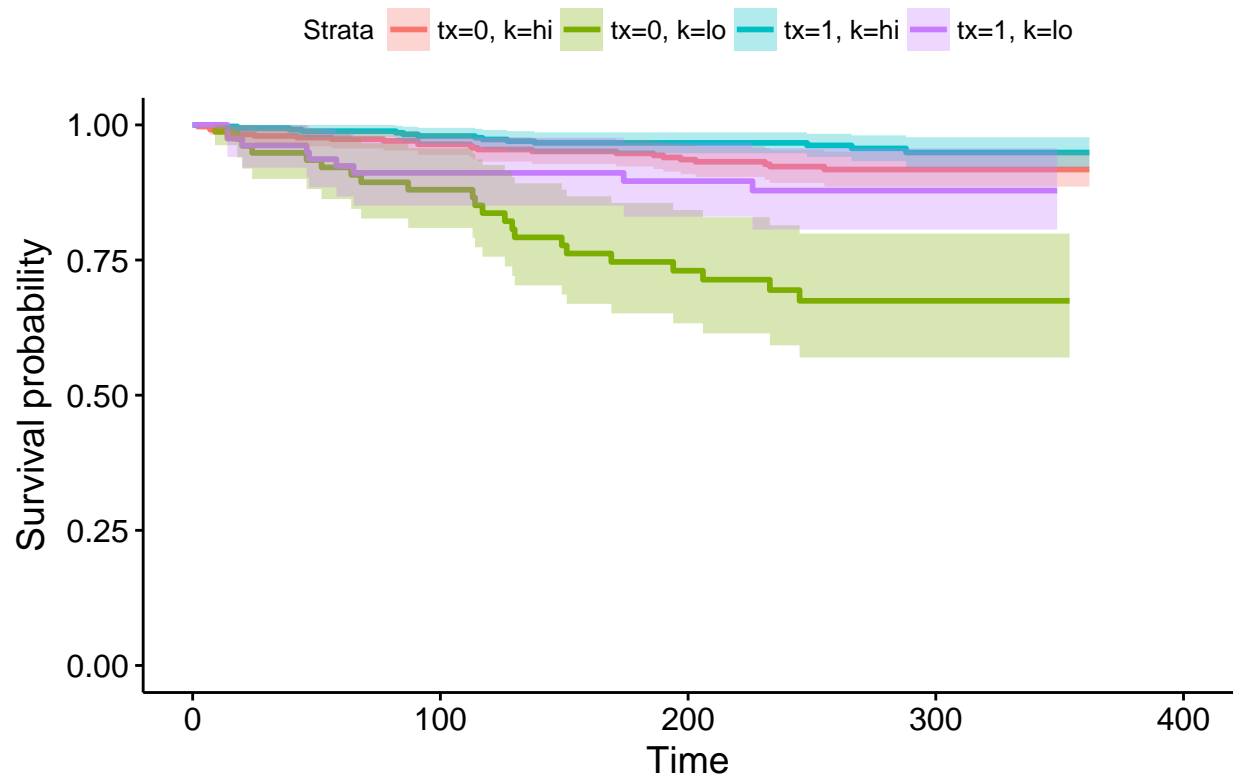
There is a significant difference in survival time between the four groups based on the Log-Rank test (p = 3.21e-12) and Wilcoxon test (p = 4.26e-12). There is an especially noticeable difference in survival times for those with low Karnofsky scores- those on the treatment appeared to have much higher survival probabilities than those on the control.