Survival_Analysis_Final

```
library(mosaic)
library(readr)
library(tidyverse)
library(broom)
library(survival)
library(survminer)
library(praise)
```

```
df <- read_csv("HELPdata.csv", na="*")
```

```
df <- df %>%
  mutate(yrs_education = as.numeric(a9), gender=a1, alcq_30 = as.numeric(alcq_30)) %>%
  select(group, dayslink, linkstatus, yrs_education, gender, age, alcohol, alcq_30)
```

```
## Warning: Problem with `mutate()` input `yrs_education`.
## i NAs introduced by coercion
## i Input `yrs_education` is `as.numeric(a9)`.

## Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion

## Warning: Problem with `mutate()` input `alcq_30`.
## i NAs introduced by coercion
## i Input `alcq_30` is `as.numeric(alcq_30)`.

## Warning in mask$eval_all_mutate(dots[[i]]): NAs introduced by coercion
```

Exploratory Data Analysis

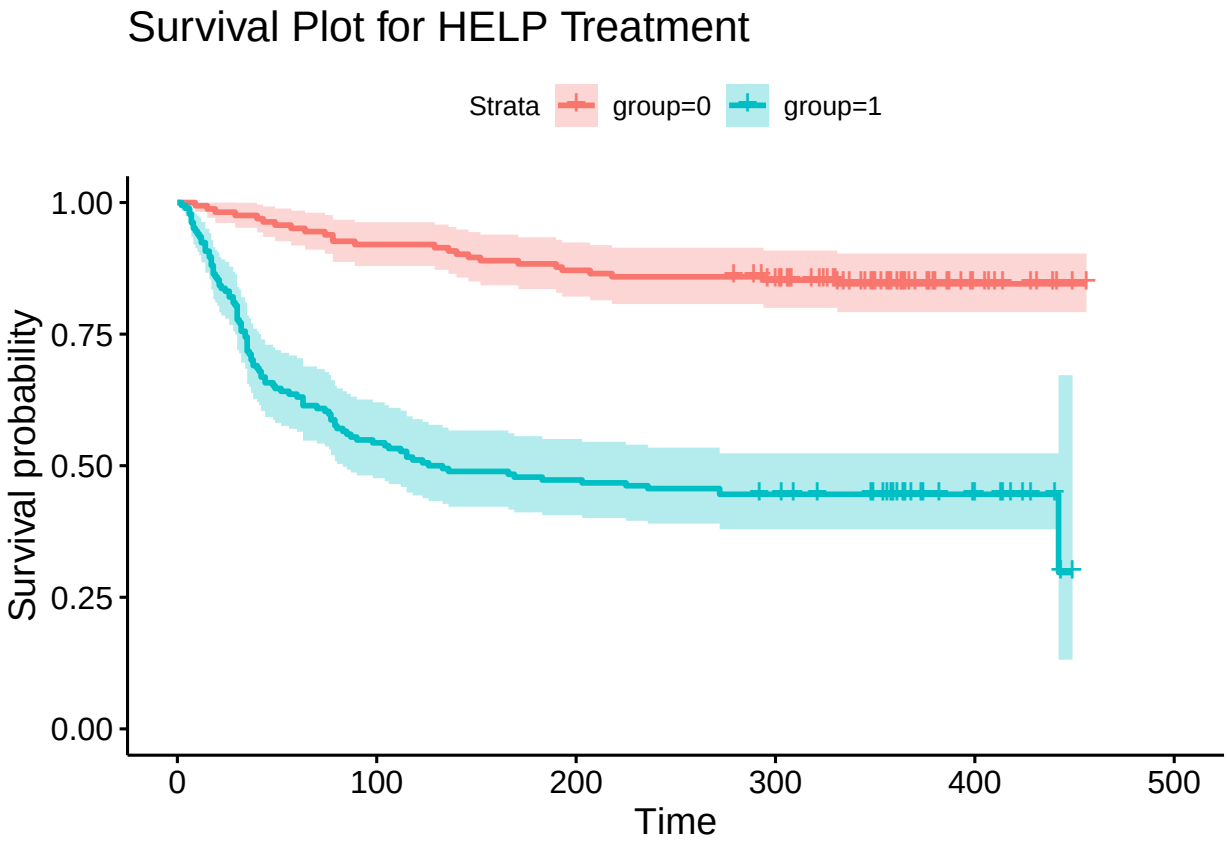
```
summary(df)
```

```
##      group           dayslink       linkstatus     yrs_education
##  Min.   :0.0000   Min.   :  2.0   Min.   :0.0000   Min.   : 3.00
##  1st Qu.:0.0000   1st Qu.: 78.0   1st Qu.:0.0000   1st Qu.:11.00
##  Median :1.0000   Median :364.0   Median :0.0000   Median :12.00
##  Mean   :0.5303   Mean   :257.4   Mean   :0.3689   Mean   :11.94
##  3rd Qu.:1.0000   3rd Qu.:365.0   3rd Qu.:1.0000   3rd Qu.:13.00
##  Max.   :1.0000   Max.   :456.0   Max.   :1.0000   Max.   :24.00
##                                                    NA's   :3
##      gender           age           alcohol          alcq_30
##  Min.   :1.000   Min.   :18.00   Min.   :0.0000   Min.   :  0.0
```

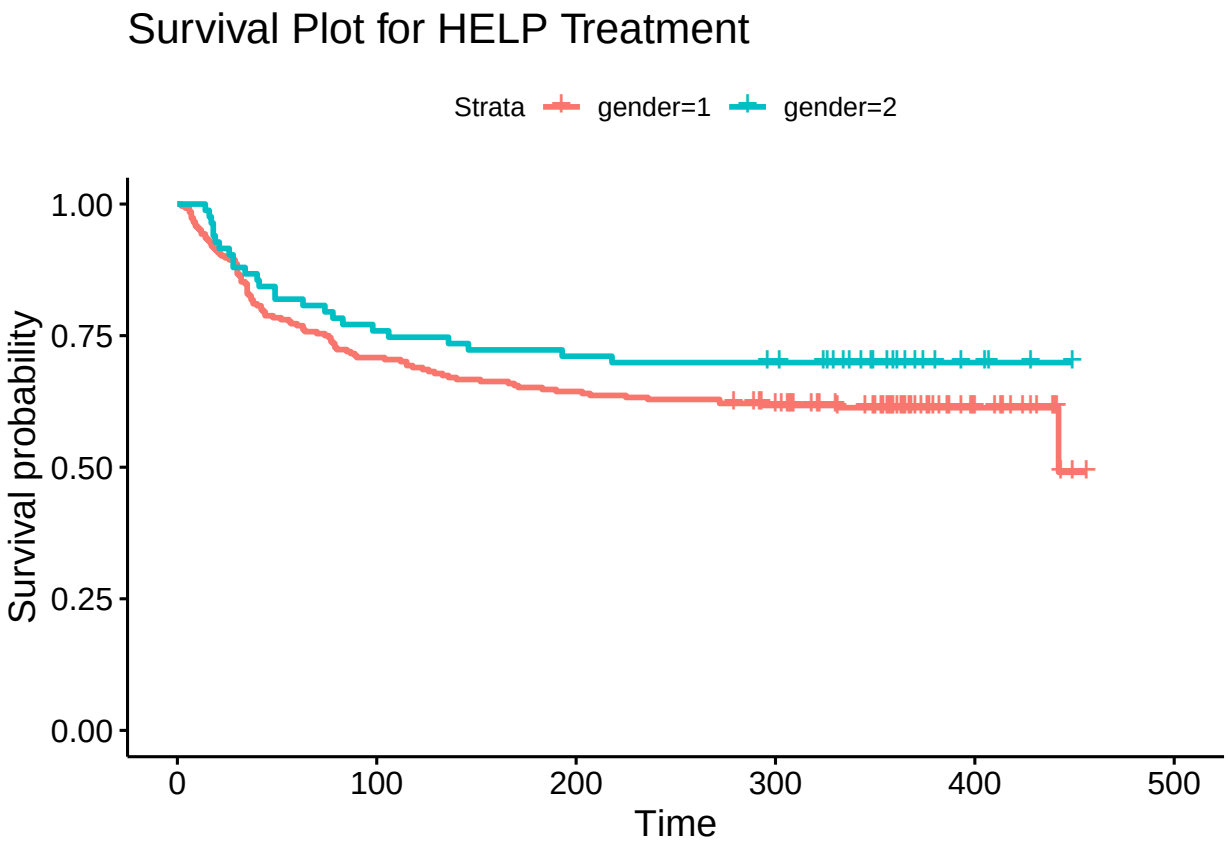
```
##  1st Qu.:1.000   1st Qu.:31.00   1st Qu.:0.0000   1st Qu.:  15.0
##  Median :1.000   Median :35.00   Median :1.0000   Median : 165.0
##  Mean   :1.239   Mean   :35.91   Mean   :0.6398   Mean   : 393.9
##  3rd Qu.:1.000   3rd Qu.:41.00   3rd Qu.:1.0000   3rd Qu.: 571.5
##  Max.   :2.000   Max.   :59.00   Max.   :1.0000   Max.   :4020.0
##                                                   NA's   :1
```

```
# df %>%
#   group_by(group) %>%
#   ggplot() + geom_line(aes(x=dayslink, y=linkstatus, color=group))
```

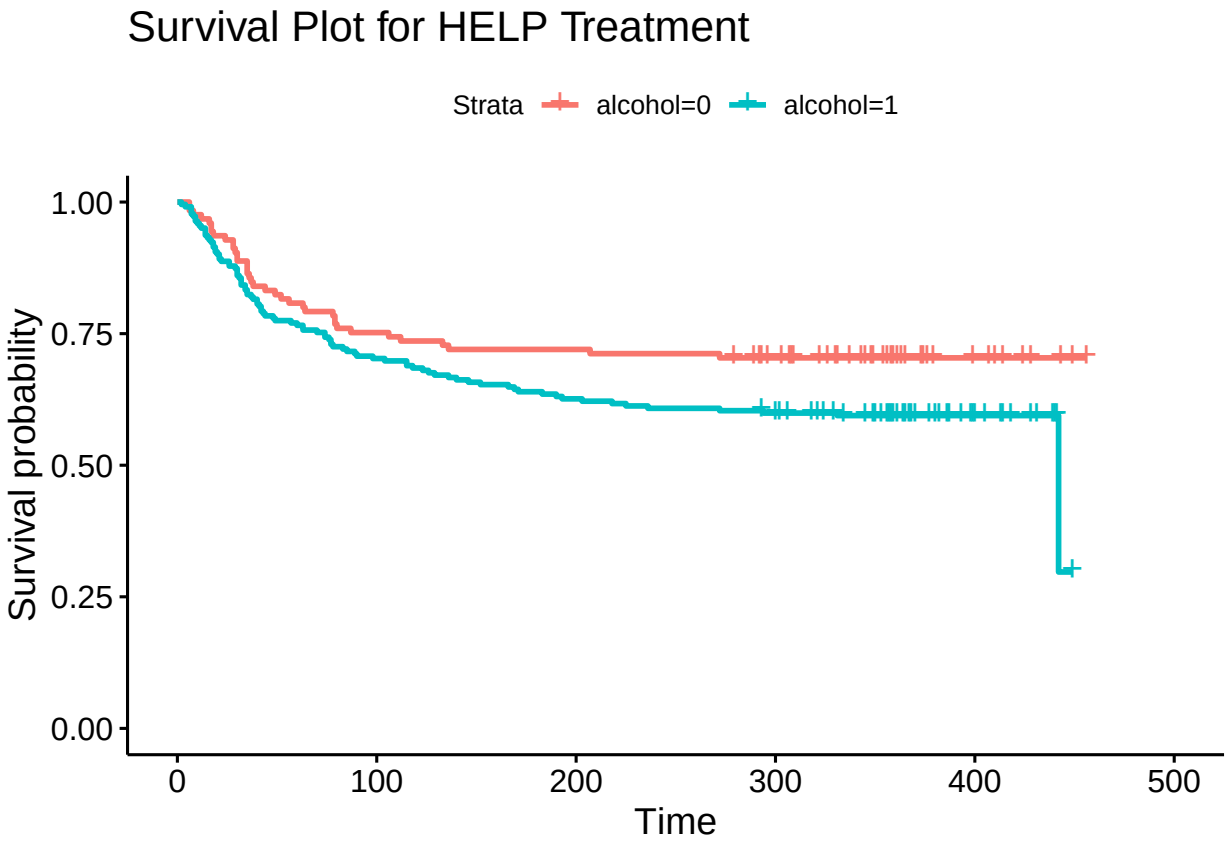
```
care_fit <- survfit(Surv(dayslink, linkstatus) ~ group, data=df)
ggsurvplot(care_fit, conf.int = T) +
  ggtitle("Survival Plot for HELP Treatment")
```




```
care_fit <- survfit(Surv(dayslink, linkstatus) ~ gender, data=df)
ggsurvplot(care_fit) +
  ggtitle("Survival Plot for HELP Treatment")
```




```
care_fit <- survfit(Surv(dayslink, linkstatus) ~ alcohol, data=df)
ggsurvplot(care_fit) +
  ggtitle("Survival Plot for HELP Treatment")
```


Survival Plot for HELP Treatment
Strata
alcohol=0
alcohol=1
Survival probability
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
500
Time