

# Lung and Breast Cancer

*Thomas Pattara*

*October 9th, 2018*

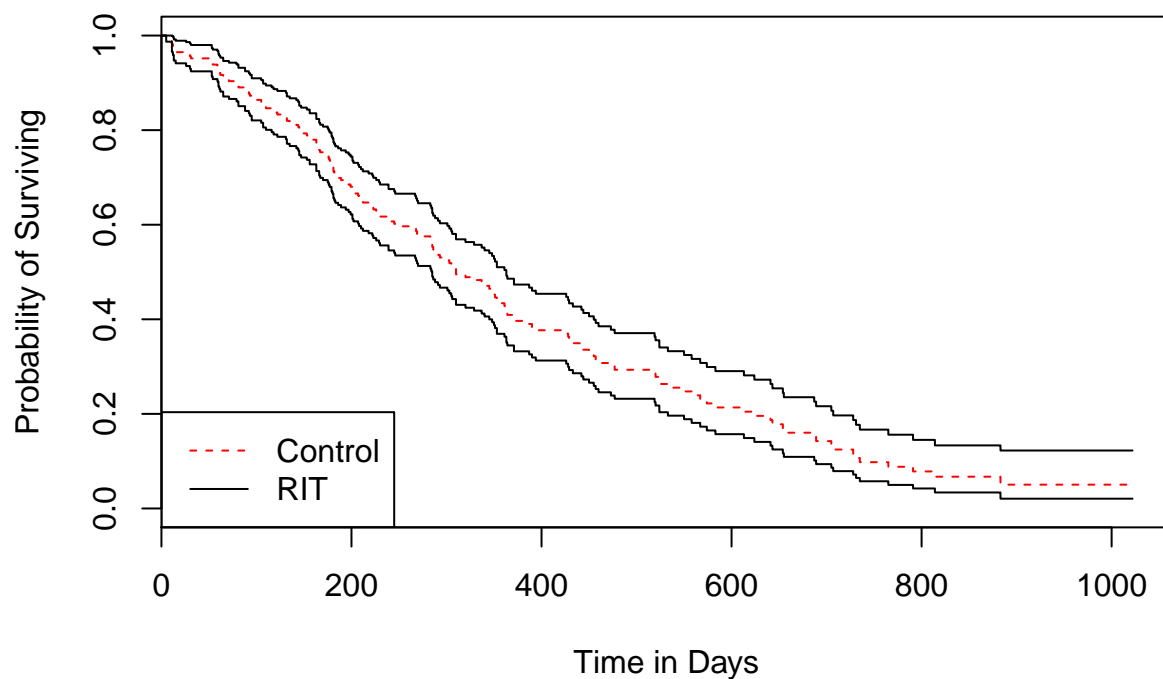
```
library(survival)
library(dplyr)
library(tidyr)
library(coin)
library(ggplot2)
library(HSAUR3)
library(party)
library(survival)
library(survminer)
```

## Question 1

An investigator collected data on survival of patients with lung cancer at Mayo Clinic. The investigator would like you, the statistician, to answer the following questions and provide some graphs. Use the **cancer** data located in the **survival** package.

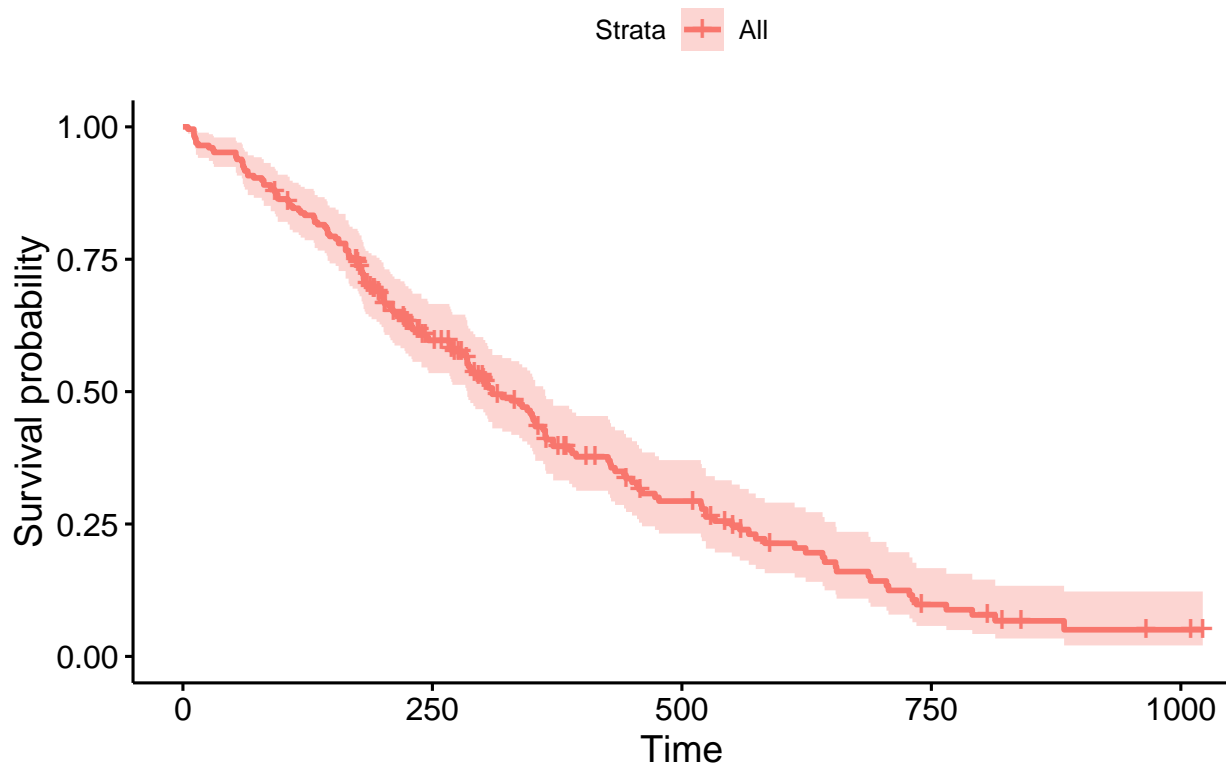
a. Provide a graph, including 95% confidence limits, of the Kaplan-Meier estimate of the entire study.

### Kaplan-Meier Estimate



## GGPLOT

### Kaplan–Meir Estimate



b. What is the probability that someone will survive past 300 days?

```
## $surv
## [1] 0.5306081
```

### Answer

The probability that someone will surviving lung cancer past 300 days is approximately 0.53061.

c. Is there a difference in the survival rates between males and females? Provide a formal statistical test with a p-value and visual evidence.

```
##
## Asymptotic Two-Sample Logrank Test
##
## data: Surv(time, status == 2) by sex (1, 2)
## Z = -3.2779, p-value = 0.001046
## alternative hypothesis: true theta is not equal to 1
```

### Probability of Men Surviving Past 300 Days

The probability of men surviving past 300 days is approximately 0.441 which is lower than for women.

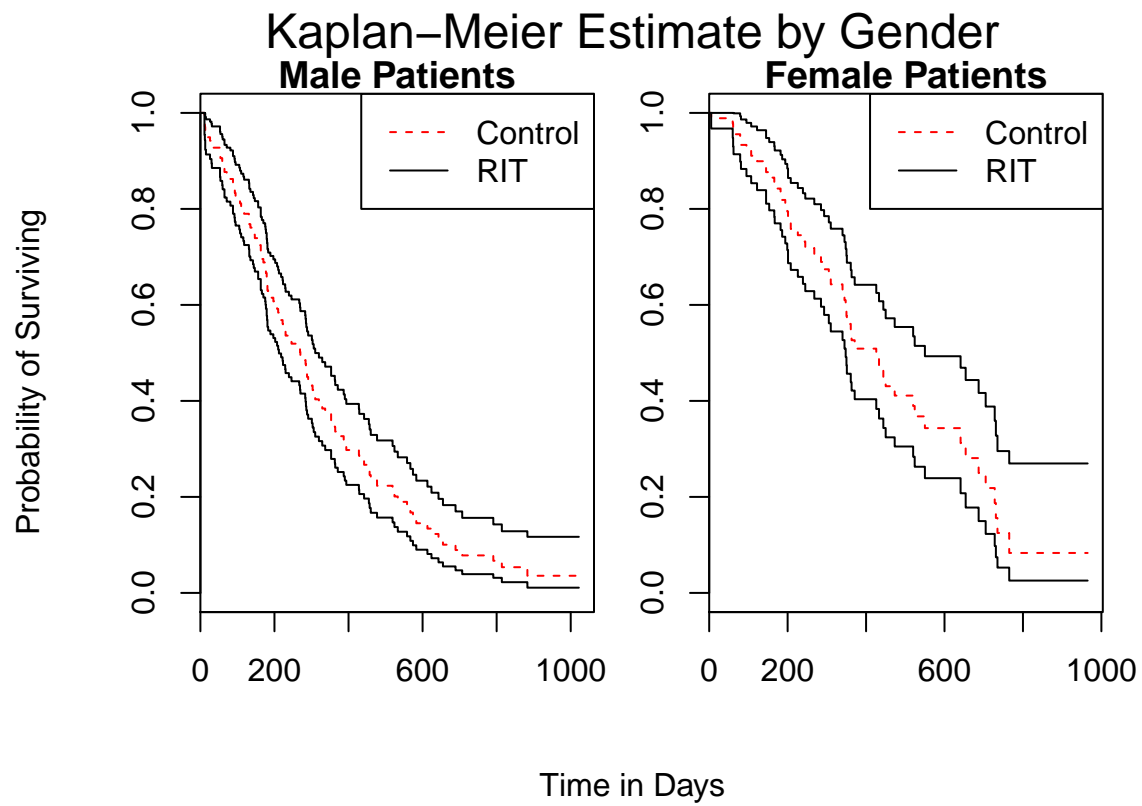
```
## $surv
## [1] 0.4410889
```

## Probability of Women Surviving Past 300 Days

The probability of men surviving past 300 days is approximately 0.674 which is higher than for men.

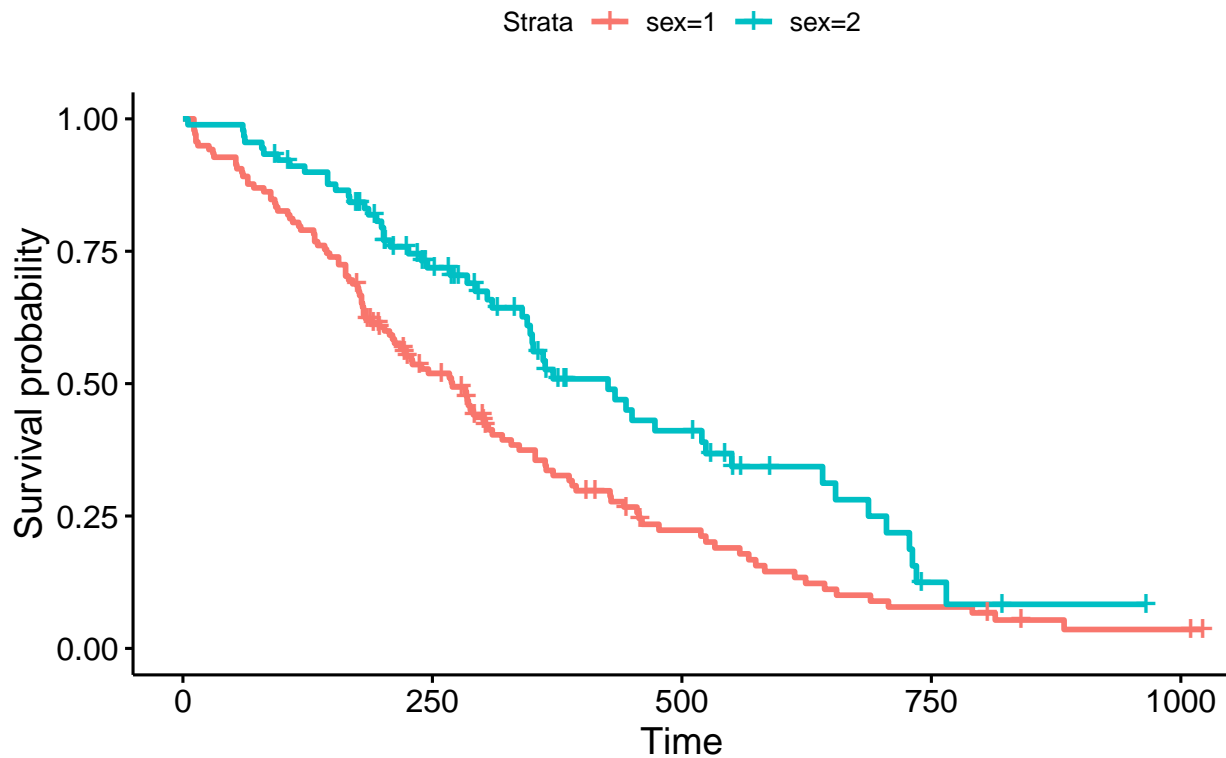
```
## $surv
## [1] 0.6742026
```

## Plot of Kaplan-Meier Estimates For Men and Women



GGPLOT

Male(1) vs Female(2)

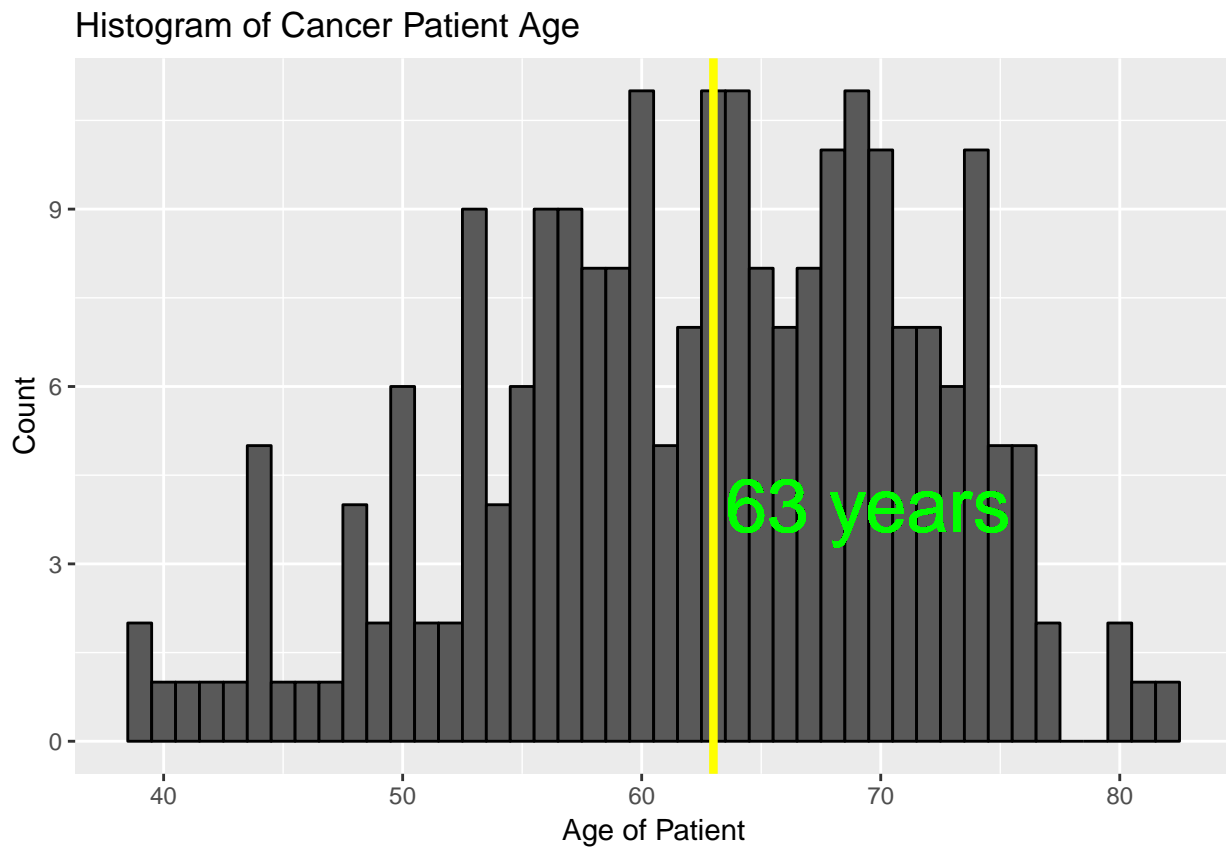


Answer

Yes, women have a higher probability of surviving lung cancer than men. This can be verified by the Kaplan-Meier estimate plots and by the probability of each surviving past 300 days done above.

d. Is there a difference in the survival rates for the older half of the group versus the younger half? Provide a formal statistical test with a p-value and visual evidence.

## Dividing the Data into Two Groups GGPLOT



## Calculate the p-value

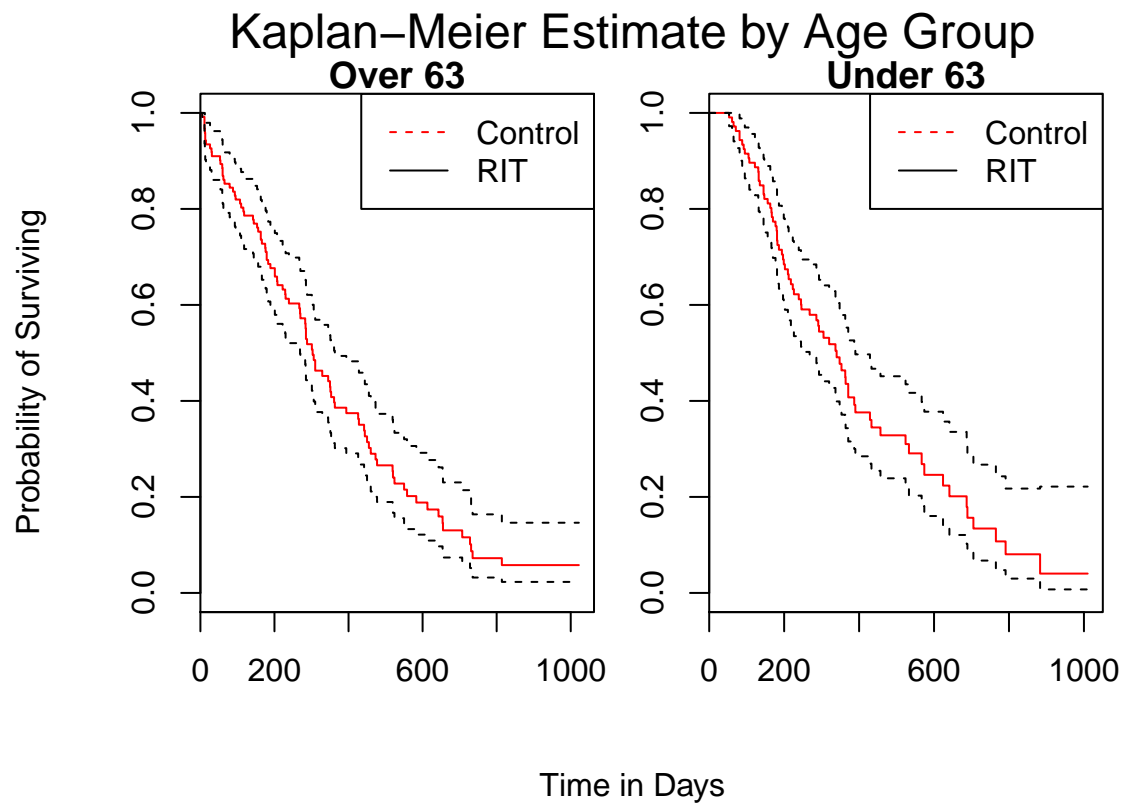
```
##
## Asymptotic Two-Sample Logrank Test
##
## data: Surv(time, status == 2) by agegroup (Old, Young)
## Z = -0.8637, p-value = 0.3878
## alternative hypothesis: true theta is not equal to 1
```

## Probability of Adults Above 63 Years Surviving After 300 Days

The probability of an adult age 63 and older surviving lung cancer is approximately 0.518.

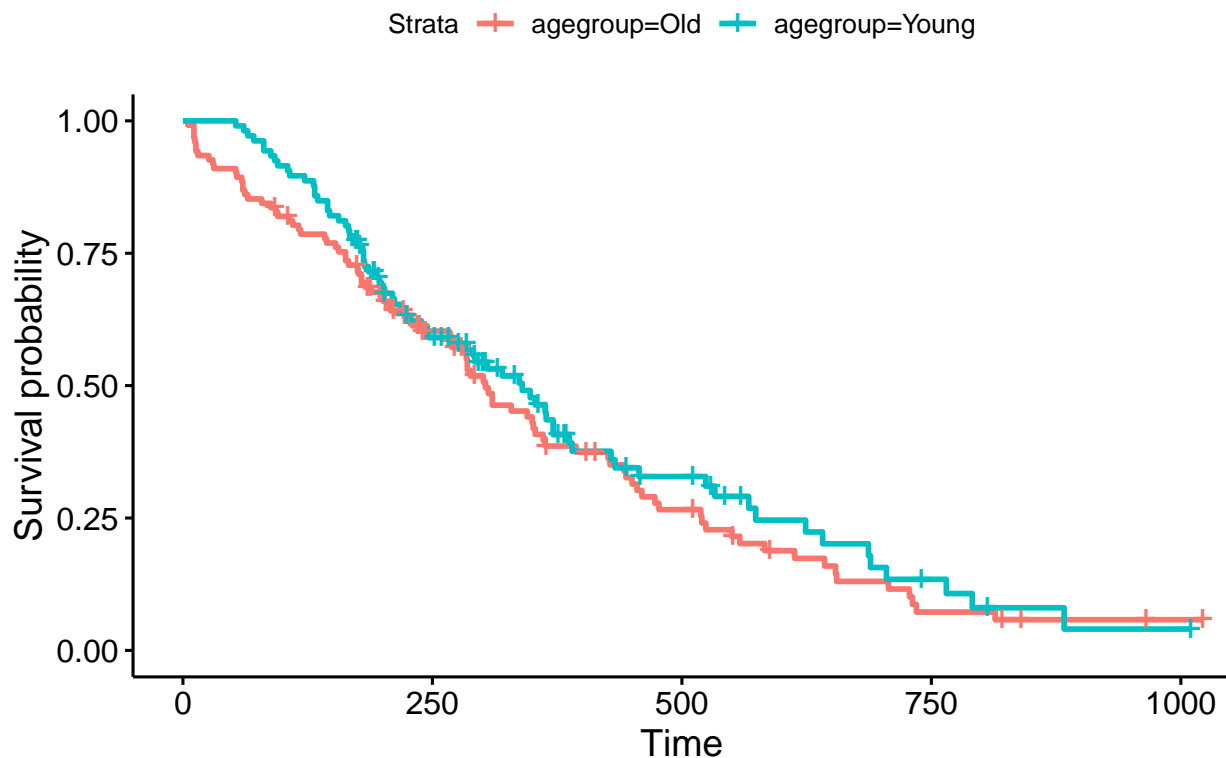
```
## $surv
## [1] 0.5181564
```

Plot of Kaplan-Meier Estimates For Older and Younger than 63 Years of Age



## GGPLOT

### Over 63 vs Under 63



## Answer

The median being 63 years, there is a very small difference in the probability of death in the two age groups. The older group does not have that much probability of surviving as the younger group as days increase.

## Question 2

A healthcare group has asked you to analyse the **mastectomy** data from the **HSAUR3** package, which is the survival times (in months) after a mastectomy of women with breast cancer. The cancers are classified as having metastasized or not based on a histochemical marker. The healthcare group requests that your report should not be longer than one page, and must only consist of one plot, one table, and one paragraph. Do the following:

a. Plot the survivor functions of each group only using GGPlot, estimated using the Kaplan-Meier estimate.

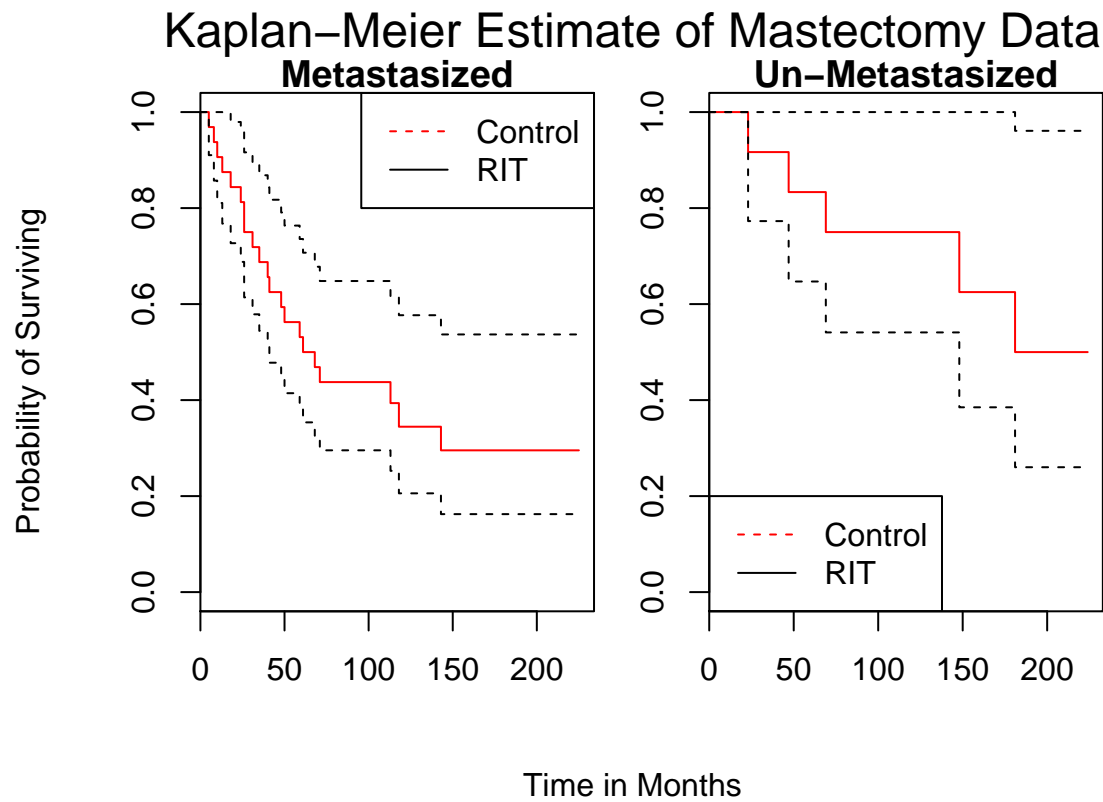
### Probability of Surviving After 50 Months for Metastasized Patients

```
## $surv
## [1] 0.5625
```

## Probability of Surviving After 50 Months for Un-Metastasized Patients

```
## $surv  
## [1] 0.8333333
```

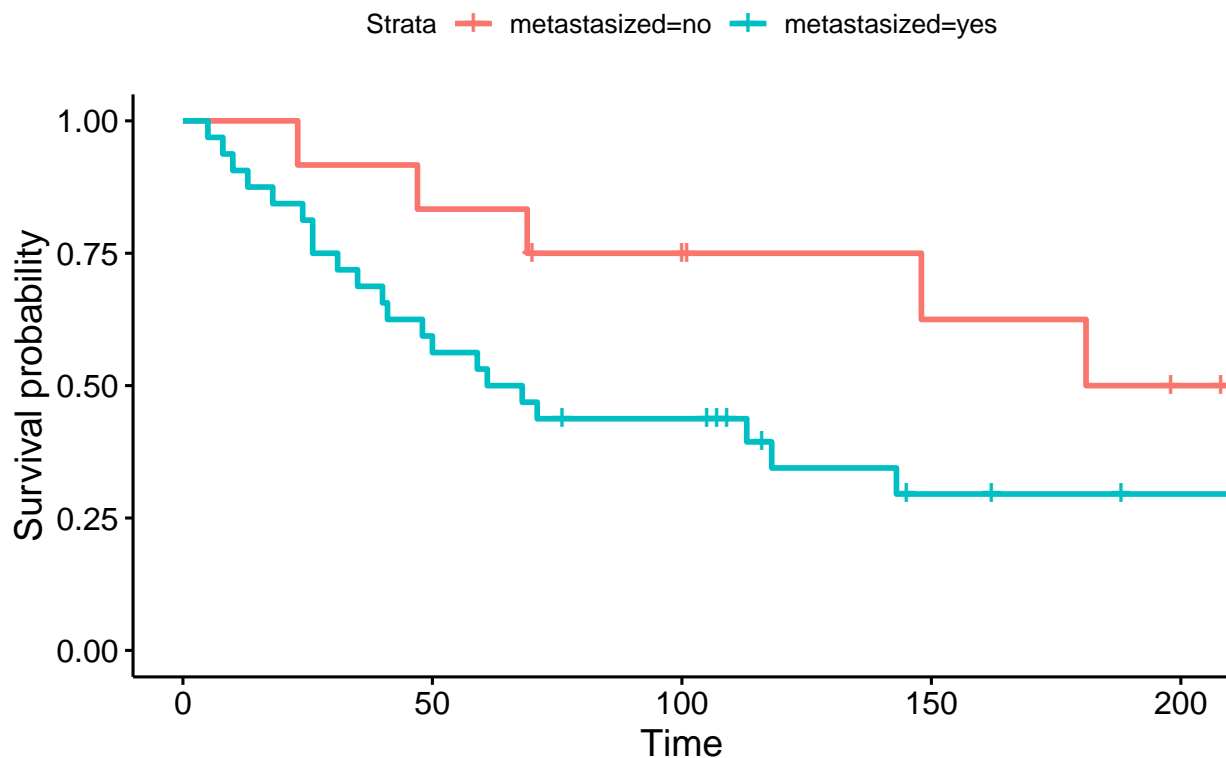
## Plot of Kaplan-Meier Estimates For Mastasized and Un-Mastasized Patients





## GGPLOT

### Metastasized vs Un-Metastasized



b. Use a log-rank test to compare the survival experience of each group more formally. Only present a formal table of your results.

```
##
## Asymptotic Two-Sample Logrank Test
##
## data: Surv(time, event == 2) by metastasized (no, yes)
## Z = 1.8667, p-value = 0.06194
## alternative hypothesis: true theta is not equal to 1
```

c. Write one paragraph summarizing your findings and conclusions.

There is a noticeable difference in the probability of death for metastasized and un-metastasized patients. There is a much higher probability of patients surviving if the cancer is un-metastasized. The probability of surviving cancer for an un-metastasized patient is 0.83 while for a metastasized patient is only 0.56. There is significant visual confirmation from the diagrams given above to this finding.