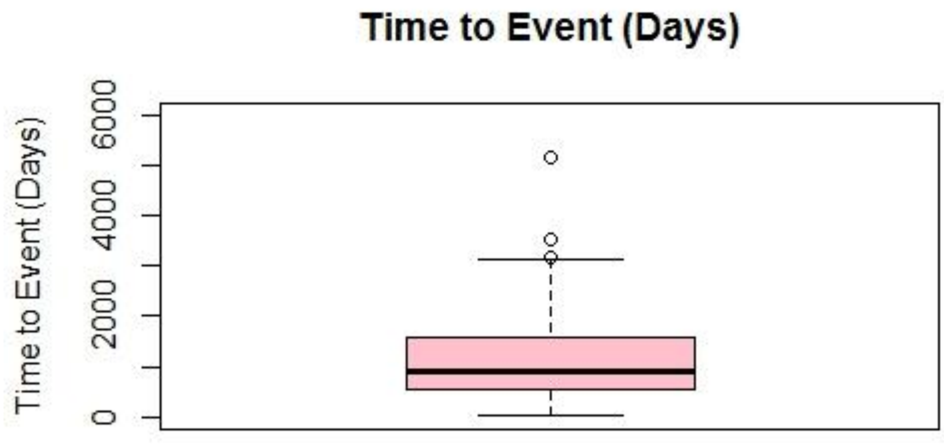
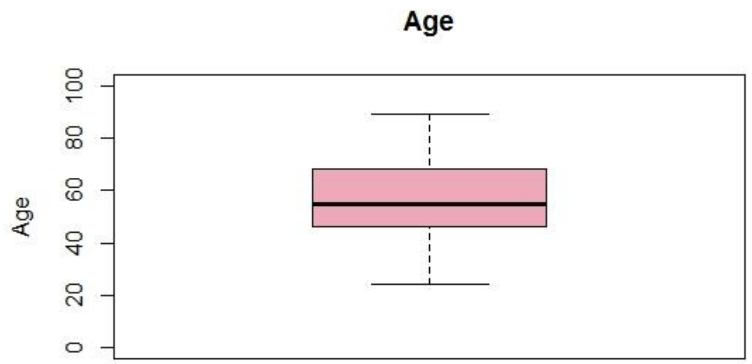
```
##Data cleaning.
```

```
##Time to event boxplot.
> summary(UNC4_assignment3$t.os)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
     30     540     885    1104    1590    5190 
> boxplot(UNC4_assignment3$t.os)
> quantile(UNC4_assignment3$t.os, probs=c(0, 0.25, 0.5, 0.75, 1))
  0%  25%  50%  75% 100% 
  30  540  885 1590 5190 
```

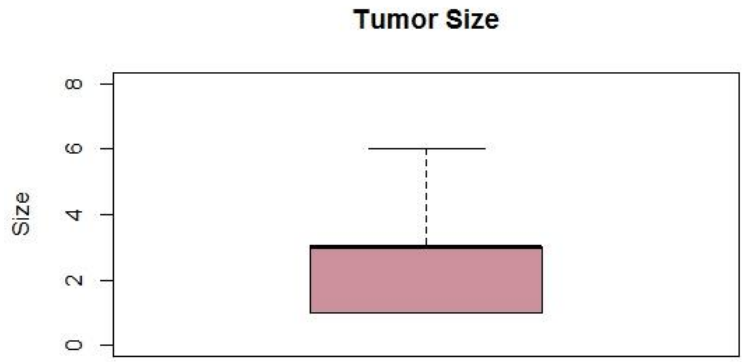


##Creating histogram for time to event.

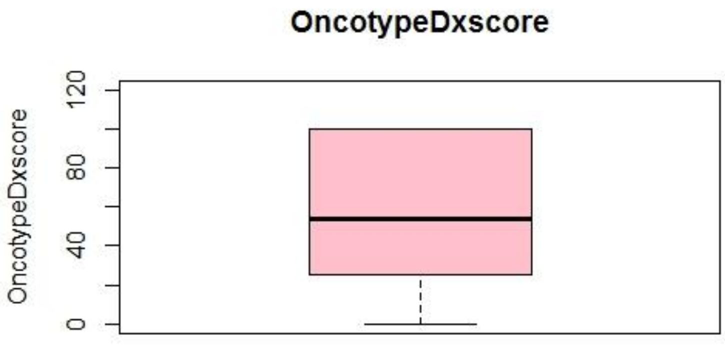
```
##Boxplot for age as a continuous variable.
> summary(UNC4_assignment3$age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  24.00   46.00   55.00   56.96   68.00   89.00      15 
> boxplot(UNC4_assignment3$age)
> quantile(UNC4_assignment3$t.os, probs=c(0, 0.25, 0.5, 0.75, 1))
  0%  25%  50%  75% 100% 
  30  540  885 1590 5190 
> boxplot(UNC4_assignment3$age, col=(c("pink2")), main="Age", ylab="Age", 
ylim=c(0,100))
```



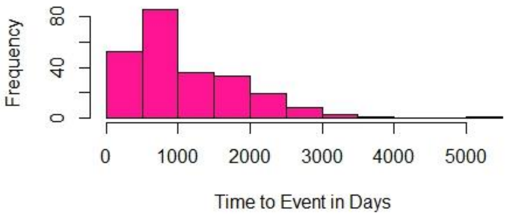
```
##Boxplot for Size as a Continuous variable.

> summary(UNC4_assignment3$size)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  1.000   1.000   3.000   3.072   3.000   6.000      26 
> summary(UNC4_assignment3$size)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  1.000   1.000   3.000   3.072   3.000   6.000      26 
> boxplot(UNC4_assignment3$size)
> boxplot(UNC4_assignment3$size, col=(c("pink3")), main="Tumor Size",
ylab="Size", ylim=c(0,8))
```



```
##OncotypeDxscore Boxplot.

> summary(UNC4_assignment3$OncotypeDxscore)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   25.63   53.90   55.98  100.00  100.00 
> boxplot(UNC4_assignment3$OncotypeDxscore)
> boxplot(UNC4_assignment3$OncotypeDxscore, col=(c("pink")),
main="OncotypeDxscore", ylab="OncotypeDxscore", ylim=c(0,120))
```



```
##Histogram for Time to event.
> hist(UNC4_assignment3$t.os,  col="deeppink1", main="", xlab="Time to Event
in Days", ylab="Frequency")
```



```
## Time to event and age are already continuous.
## Categorizing Event variable.
> UNC4_assignment3["UNC4e.os"] <- NA

> UNC4_assignment3$UNC4e.os <-UNC4_assignment3$e.os
> UNC4_assignment3$UNC4e.os <-cut(UNC4_assignment3$UNC4e.os, br=c(-1,0,1))
> summary(UNC4_assignment3$UNC4e.os)
(-1,0]  (0,1]   NA's
   189     51     65
> levels(UNC4_assignment3$UNC4e.os)<- c("No death", "Death")
> summary(UNC4_assignment3$UNC4e.os)
   0     1 NA's
 189    51   65
```

```
## Cleaning up pgr variable.

> is.factor(UNC4_assignment3$pgr)
[1] FALSE
> UNC4_assignment3$pgr <-cut(UNC4_assignment3$pgr, br=c(-1,0,1))
> levels(UNC4_assignment3$pgr)<- c("Not expressing", "Expressing")
> summary(UNC4_assignment3$pgr)
Not expressing     Expressing           NA's
```

```
           126            109             70
```

Cleaning up er variable.

```
> is.factor(UNC4_assignment3$er)
[1] FALSE
> UNC4_assignment3$er <-cut(UNC4_assignment3$er, br=c(-1,0,1))
> levels(UNC4_assignment3$er)<- c("Not expressing", "Expressing")
> summary(UNC4_assignment3$er)
Not expressing     Expressing           NA's
            99            154             52
```

Cleaning up her2 variable.

```
> is.factor(UNC4_assignment3$her2)
[1] FALSE
> UNC4_assignment3$her2 <-cut(UNC4_assignment3$her2, br=c(-1,0,1))
> levels(UNC4_assignment3$her2)<- c("Not expressing", "Expressing")
> summary(UNC4_assignment3$her2)
Not expressing     Expressing           NA's
           203             58             44
```

```
##Cleaning up tumor grade variable.
> is.factor(UNC4_assignment3$grade)
[1] FALSE
> UNC4_assignment3$grade <-cut(UNC4_assignment3$grade, br=c(0,1,2,3))
> levels(UNC4_assignment3$grade) <-c("Low Grade","Intermediate Grade","High
Grade")
> summary(UNC4_assignment3$grade)
         Low Grade Intermediate Grade         High Grade
                25                 80                137
              NA's
                63

##Cleaning up node status.

> is.factor(UNC4_assignment3$node)
[1] FALSE
> UNC4_assignment3$node <-cut(UNC4_assignment3$node, br=c(-1,0,1))
> levels(UNC4_assignment3$node) <-c("It has not spread to the lymph
nodes","It has spread to the lymph nodes")
> summary(UNC4_assignment3$node)
It has not spread to the lymph nodes
                                 126
    It has spread to the lymph nodes
                                 134
                                NA's
                                  45


##Cleaning up OncotypeDxrisk.

> is.factor(UNC4_assignment3$OncotypeDxrisk)
[1] FALSE
> UNC4_assignment3$OncotypeDxrisk <-cut(UNC4_assignment3$OncotypeDxrisk,
br=c(-1,0,0.5,1))
> levels(UNC4_assignment3$OncotypeDxrisk) <-c("low risk","intermediate
risk","high risk")
> summary(UNC4_assignment3$OncotypeDxrisk)
         low risk intermediate risk         high risk
```

```
               54                41               210


##Categorizing age.

> UNC4_assignment3["UNC4age"] <- NA
> UNC4_assignment3["UNC4age"] <-UNC4_assignment3$age
> UNC4_assignment3$UNC4age <-cut(UNC4_assignment3$UNC4age,
br=c(20,39,49,59,89))
> summary(UNC4_assignment3$UNC4age)
(20,39] (39,49] (49,59] (59,89]    NA's
     29      61      61      95      59
> levels(UNC4_assignment3$UNC4age) <-c("24-39","40-49","50-59","60-89")
> summary(UNC4_assignment3$UNC4age)
24-39 40-49 50-59 60-89  NA's
   29    61    61    95    59

##Categorizing size.
> UNC4_assignment3$size <-cut(UNC4_assignment3$size, br=c(0,2,4,6))
> levels(UNC4_assignment3$size) <-c("1 cm or 2 cm","3 cm or 4 cm","5 cm or 6
cm")
> summary(UNC4_assignment3$size)
1 cm or 2 cm 3 cm or 4 cm 5 cm or 6 cm         NA's
          61          129           43           72



##Categorizing time to event


> UNC4_assignment3["UNC4t.os"] <- NA
> UNC4_assignment3["UNC4t.os"] <- UNC4_assignment3$t.os
> UNC4t.os <-cut(UNC4t.os, br=c(-1,1500,3000,4500,6000))
> levels(UNC4_assignment3$UNC4t.os) <-c("0-50 months","51-100 months","101-
150 months","151-200 months")
> summary(UNC4_assignment3$UNC4t.os)
   0-50 months  51-100 months 101-150 months 151-200 months
           175             60              4              1
          NA's
            65


##Creating table for question #1.

library(plyr)

##Event for question 1.

> eventq.1 <-cbind("N"= count(UNC4_assignment3$UNC4e.os), "Percent" =
(count(UNC4_assignment3$UNC4e.os)[,2]/sum(count(UNC4_assignment3$UNC4e.os)[2]
))*100)
> eventq.1
   N.x N.freq  Percent
1    0    189 61.96721
2    1     51 16.72131
3 <NA>     65 21.31148

##Time to event question 1.

> summary(UNC4_assignment3$UNC4t.os)
   0-50 months  51-100 months 101-150 months  151-200 monts
           175             60              4              1
          NA's
            65
```

```
> timetoevq.1 <-cbind("N"= count(UNC4_assignment3$UNC4t.os), "Percent" =
(count(UNC4_assignment3$UNC4t.os)[,2]/sum(count(UNC4_assignment3$UNC4t.os)[2]
))*100)
```

```
> timetoevq.1
               N.x N.freq    Percent
1     0-50 months    175 57.3770492
2   51-100 months     60 19.6721311
3 101-150 months      4  1.3114754
4   151-200 monts      1  0.3278689
5            <NA>     65 21.3114754
```

Age for question 1.

```
> ageq.1 <-cbind("N"= count(UNC4_assignment3$UNC4age), "Percent" =
(count(UNC4_assignment3$UNC4age)[,2]/sum(count(UNC4_assignment3$UNC4age)[2]))
*100)
> ageq.1
    N.x N.freq   Percent
1 24-39     29  9.508197
2 40-49     61 20.000000
3 50-59     61 20.000000
4 60-89     95 31.147541
5  <NA>     59 19.344262
```

pgr for question 1.

```
> pgr.1 <-cbind("N"= count(UNC4_assignment3$pgr), "Percent" =
(count(UNC4_assignment3$pgr)[,2]/sum(count(UNC4_assignment3$pgr)[2]))*100)
> pgr.1
             N.x N.freq  Percent
1 Not expressing    126 41.31148
2     Expressing    109 35.73770
3           <NA>     70 22.95082
```

##er for question 1.

```
> er.1 <-cbind("N"= count(UNC4_assignment3$er), "Percent" =
(count(UNC4_assignment3$er)[,2]/sum(count(UNC4_assignment3$er)[2]))*100)
> er.1
             N.x N.freq  Percent
1 Not expressing     99 32.45902
2     Expressing    154 50.49180
3           <NA>     52 17.04918
```

##her2 for question 1.

```
> her2q.1 <-cbind("N"= count(UNC4_assignment3$her2), "Percent" =
(count(UNC4_assignment3$her2)[,2]/sum(count(UNC4_assignment3$her2)[2]))*100)
> her2q.1
             N.x N.freq  Percent
1 Not expressing    203 66.55738
2     Expressing     58 19.01639
3           <NA>     44 14.42623
```

##tumor grade for question 1.

```
> gradeq.1 <-cbind("N"= count(UNC4_assignment3$grade), "Percent" =
(count(UNC4_assignment3$grade)[,2]/sum(count(UNC4_assignment3$grade)[2]))*100
)
> gradeq.1
                 N.x N.freq   Percent
1          Low Grade     25  8.196721
2 Intermediate Grade     80 26.229508
3         High Grade    137 44.918033
4               <NA>     63 20.655738
```

Tumor size for question 1.

```
> sizeq.1 <-cbind("N"= count(UNC4_assignment3$size), "Percent" =
(count(UNC4_assignment3$size)[,2]/sum(count(UNC4_assignment3$size)[2]))*100)
> sizeq.1
           N.x N.freq  Percent
1 1 cm or 2 cm     61 20.00000
2 3 cm or 4 cm    129 42.29508
3 5 cm or 6 cm     43 14.09836
4         <NA>     72 23.60656
```

##Lymph node status for question 1.

```
> nodeq.1 <-cbind("N"= count(UNC4_assignment3$node), "Percent" =
(count(UNC4_assignment3$node)[,2]/sum(count(UNC4_assignment3$node)[2]))*100)
>
> nodeq.1
                                 N.x N.freq  Percent
1 It has not spread to the lymph nodes    126 41.31148
2     It has spread to the lymph nodes    134 43.93443
3                                 <NA>     45 14.75410
```

```
##Oncotype risk for question 1.
> oncoriskq.1 <-cbind("N"= count(UNC4_assignment3$OncotypeDxrisk), "Percent"
=
(count(UNC4_assignment3$OncotypeDxrisk)[,2]/sum(count(UNC4_assignment3$Oncoty
peDxrisk)[2]))*100)
> oncoriskq.1
                N.x N.freq  Percent
1          low risk     54 17.70492
2 intermediate risk     41 13.44262
3         high risk    210 68.85246
```

```
##Creating the entire frequencies table.

> tableq.1 <- rbind(eventq.1, timetoevq.1, ageq.1, pgr.1, er.1, her2q.1,
gradeq.1, sizeq.1, nodeq.1, oncoriskq.1)
> tableq.1
                                     N.x N.freq    Percent
1                               No death    189 61.9672131
2                                  Death     51 16.7213115
3                                   <NA>     65 21.3114754
4                            0-50 months    175 57.3770492
5                          51-100 months     60 19.6721311
6                         101-150 months      4  1.3114754
7                          151-200 monts      1  0.3278689
8                                   <NA>     65 21.3114754
9                                  24-29      5  1.6393443
10                                 30-39     24  7.8688525
11                                 40-49     61 20.0000000
12                                 50-59     61 20.0000000
```

```
13                                60-69     39 12.7868852
14                                70-79     35 11.4754098
15                                80-89     21  6.8852459
16                                 <NA>     59 19.3442623
17                       Not expressing    126 41.3114754
18                           Expressing    109 35.7377049
19                                 <NA>     70 22.9508197
20                       Not expressing     99 32.4590164
21                           Expressing    154 50.4918033
22                                 <NA>     52 17.0491803
23                       Not expressing    203 66.5573770
24                           Expressing     58 19.0163934
25                                 <NA>     44 14.4262295
26                            Low Grade     25  8.1967213
27                   Intermediate Grade     80 26.2295082
28                           High Grade    137 44.9180328
29                                 <NA>     63 20.6557377
30                                 1 cm     60 19.6721311
31                                 2 cm      1  0.3278689
32                                 3 cm    129 42.2950820
33                                 6 cm     43 14.0983607
34                                 <NA>     72 23.6065574
35 It has not spread to the lymph nodes    126 41.3114754
36     It has spread to the lymph nodes    134 43.9344262
37                                 <NA>     45 14.7540984
38                             low risk     54 17.7049180
39                    intermediate risk     41 13.4426230
40                            high risk    210 68.8524590
> write.csv(tableq.1, "C:/Users/James/Desktop/Tablequestion1.csv")

##2x2 Tables

##pgr x er
> pgrxer <- table(UNC4_assignment3$pgr, UNC4_assignment3$er)
> pgrxer

                 Not expressing Expressing
  Not expressing             65         35
  Expressing                  6         98

> chisq.test(pgrxer)

        Pearson's Chi-squared test with Yates' continuity correction

data:  pgrxer
X-squared = 73.7448, df = 1, p-value < 2.2e-16

##pgr x her2

> pgrxher2 <- table(UNC4_assignment3$pgr, UNC4_assignment3$her2)
> pgrxher2

                  Not expressing Expressing
  Not expressing             94         30
  Expressing                 86         19

> chisq.test(pgrxher2)

        Pearson's Chi-squared test with Yates' continuity correction

data:  pgrxher2
X-squared = 0.888, df = 1, p-value = 0.346
```

```
##pgr x grade

> pgrxgrade <- table(UNC4_assignment3$pgr, UNC4_assignment3$grade)
> pgrxgrade

                   Low Grade Intermediate Grade High Grade
  Not expressing         5                 24         64
  Expressing            16                 46         40

> chisq.test(pgrxgrade)

        Pearson's Chi-squared test

data:  pgrxgrade
X-squared = 16.6775, df = 2, p-value = 0.0002391


##pgr x node

> pgrxnode <- table(UNC4_assignment3$pgr, UNC4_assignment3$node)
> pgrxnode

                 It has not spread to the lymph nodes
  Not expressing                                   53
  Expressing                                       52

                 It has spread to the lymph nodes
  Not expressing                               47
  Expressing                                   51


> chisq.test(pgrxnode)

        Pearson's Chi-squared test with Yates' continuity correction

data:  pgrxnode
X-squared = 0.0496, df = 1, p-value = 0.8238


##pgr x oncorisk

> pgrxoncorisk <- table(UNC4_assignment3$pgr,
UNC4_assignment3$OncotypeDxrisk)
> pgrxoncorisk

                 low risk intermediate risk high risk
  Not expressing       18                17        91
  Expressing           29                18        62

> chisq.test(pgrxoncorisk)

        Pearson's Chi-squared test

data:  pgrxoncorisk
X-squared = 6.332, df = 2, p-value = 0.04217


##pgr x age

> pgrxage <- table(UNC4_assignment3$pgr, UNC4_assignment3$age)
> pgrxage

                 24-39 40-59 60-89
```

```
  Not expressing    11    43    42
  Expressing         8    55    38

> chisq.test(pgrxage)

        Pearson's Chi-squared test

data:  pgrxage
X-squared = 1.7386, df = 3, p-value = 0.6284


##pgr x size

> pgrxsize <- table(UNC4_assignment3$pgr, UNC4_assignment3$size)
> pgrxsize

                 1 cm or 2 cm 3 cm or 4 cm 5 cm or 6 cm
  Not expressing           17           53           20
  Expressing               28           56           13
        Pearson's Chi-squared test

data:  pgrxsize
X-squared = 4.0174, df = 2, p-value = 0.1342


## er x her2

> erxher2 <- table(UNC4_assignment3$er, UNC4_assignment3$her2)
> erxher2

                 Not expressing Expressing
  Not expressing             59         26
  Expressing                118         21

> chisq.test(erxher2)

        Pearson's Chi-squared test with Yates' continuity correction

data:  erxher2
X-squared = 6.1923, df = 1, p-value = 0.01283


##er x grade

> erxgrade <- table(UNC4_assignment3$er, UNC4_assignment3$grade)
> erxgrade

                 Low Grade Intermediate Grade High Grade
  Not expressing         1                 19         73
  Expressing            23                 60         63


> chisq.test(erxgrade)

        Pearson's Chi-squared test

data:  erxgrade
X-squared = 28.9319, df = 2, p-value = 5.218e-07


## er x node

> erxnode <- table(UNC4_assignment3$er, UNC4_assignment3$node)
```

```
> erxnode

                 It has not spread to the lymph nodes
  Not expressing                                   49
  Expressing                                       73

                 It has spread to the lymph nodes
  Not expressing                               49
  Expressing                                   80

> chisq.test(erxnode)

        Pearson's Chi-squared test with Yates' continuity correction

data:  erxnode
X-squared = 0.042, df = 1, p-value = 0.8377


## er x OncotypeDxrisk


> erxoncorisk <- table(UNC4_assignment3$er,
UNC4_assignment3$OncotypeDxrisk)
> erxoncorisk

                 low risk intermediate risk high risk
  Not expressing        6                 6        87
  Expressing           39                29        86

> chisq.test(erxoncorisk)

        Pearson's Chi-squared test

data:  erxoncorisk
X-squared = 25.6721, df = 2, p-value = 2.663e-06

## er x age

> erxage <- table(UNC4_assignment3$er, UNC4_assignment3$age)
> erxage

                 24-39 40-59 60-89
  Not expressing    14    45    31
  Expressing        14    73    60


> chisq.test(erxage)

        Pearson's Chi-squared test

data:  erxage
X-squared = 2.2908, df = 3, p-value = 0.5143


## er x size

> erxsize <- table(UNC4_assignment3$er, UNC4_assignment3$size)
> erxsize

                 1 cm or 2 cm 3 cm or 4 cm 5 cm or 6 cm
  Not expressing           19           46           19
  Expressing               34           82           24
> chisq.test(erxsize)
```

```
        Pearson's Chi-squared test

data:  erxsize
X-squared = 1.3584, df = 2, p-value = 0.507
```

her2 x grade

```
> her2xgrade <- table(UNC4_assignment3$her2, UNC4_assignment3$grade)
> her2xgrade

                 Low Grade Intermediate Grade High Grade
  Not expressing        22                 59         89
  Expressing             0                 18         26

> chisq.test(her2xgrade)

        Pearson's Chi-squared test

data:  her2xgrade
X-squared = 5.7064, df = 2, p-value = 0.05766


## her2 x node

> her2xnode <- table(UNC4_assignment3$her2, UNC4_assignment3$node)
> her2xnode

                 It has not spread to the lymph nodes
  Not expressing                                   92
  Expressing                                       17

                 It has spread to the lymph nodes
  Not expressing                               86
  Expressing                                   29

> chisq.test(her2xnode)

        Pearson's Chi-squared test with Yates' continuity correction

data:  her2xnode
X-squared = 0.9334, df = 1, p-value = 0.334


##her2 x oncotypedxrisk

> her2xoncorisk <- table(UNC4_assignment3$her2,
UNC4_assignment3$OncotypeDxrisk)
> her2xoncorisk

                 low risk intermediate risk high risk
  Not expressing       39                30       134
  Expressing            9                10        39


> chisq.test(her2xoncorisk)

        Pearson's Chi-squared test

data:  her2xoncorisk
X-squared = 0.7228, df = 2, p-value = 0.6967
```

```
##her2 x age

> her2xage <- table(UNC4_assignment3$her2, UNC4_assignment3$age)
> her2xage

                 24-39 40-59 60-89
  Not expressing    15    84    69
  Expressing         7    24    14

> chisq.test(her2xage)

        Pearson's Chi-squared test

data:  her2xage
X-squared = 2.3381, df = 3, p-value = 0.5053


##her2 x size

> her2xsize <- table(UNC4_assignment3$her2, UNC4_assignment3$size)
> her2xsize

                 1 cm or 2 cm 3 cm or 4 cm 5 cm or 6 cm
  Not expressing           40           96           25
  Expressing                8           20           12
> chisq.test(her2xsize)

        Pearson's Chi-squared test

data:  her2xsize
X-squared = 5.7864, df = 2, p-value = 0.0554
```

```
## grade x node.

> gradexnode <- table(UNC4_assignment3$grade, UNC4_assignment3$node)
> gradexnode

                     It has not spread to the lymph nodes
  Low Grade                                            15
  Intermediate Grade                                   41
  High Grade                                           66

                     It has spread to the lymph nodes
  Low Grade                                        10
  Intermediate Grade                               39
  High Grade                                       70

> chisq.test(gradexnode)

        Pearson's Chi-squared test

data:  gradexnode
X-squared = 1.0403, df = 2, p-value = 0.5944


## grade x OncotypeDxrisk

> gradexoncorisk <- table(UNC4_assignment3$grade,
UNC4_assignment3$OncotypeDxrisk)
```

```
> gradexoncorisk

                     low risk intermediate risk high risk
  Low Grade                 6                 5        14
  Intermediate Grade       19                15        46
  High Grade               18                13       106


> chisq.test(gradexoncorisk)

        Pearson's Chi-squared test

data:  gradexoncorisk
X-squared = 9.9483, df = 4, p-value = 0.04131

## grade x age.

> gradexage <- table(UNC4_assignment3$grade, UNC4_assignment3$age)
> gradexage

                     24-39 40-59 60-89
  Low Grade              2    13    10
  Intermediate Grade     6    40    30
  High Grade            20    59    49

> chisq.test(gradexage)

        Pearson's Chi-squared test

data:  gradexage
X-squared = 2.6837, df = 6, p-value = 0.8474



## grade x size

> gradexsize <- table(UNC4_assignment3$grade, UNC4_assignment3$size)
> gradexsize

                     1 cm or 2 cm 3 cm or 4 cm 5 cm or 6 cm
  Low Grade                     8           11            4
  Intermediate Grade           23           39           13
  High Grade                   22           72           24

> chisq.test(gradexsize)

        Pearson's Chi-squared test

data:  gradexsize
X-squared = 4.1981, df = 4, p-value = 0.3799


## node x oncotypedxrisk

> nodexoncorisk <- table(UNC4_assignment3$node,
UNC4_assignment3$OncotypeDxrisk)
> nodexoncorisk

                                       low risk intermediate risk
  It has not spread to the lymph nodes       21                22
  It has spread to the lymph nodes           25                13
```

```
                                       high risk
  It has not spread to the lymph nodes        83
  It has spread to the lymph nodes            96

> chisq.test(nodexoncorisk)

        Pearson's Chi-squared test

data:  nodexoncorisk
X-squared = 4.7241, df = 2, p-value = 0.09423

##node x age

> nodexage <- table(UNC4_assignment3$node, UNC4_assignment3$age)
> nodexage

                                       24-39 40-59 60-89
  It has not spread to the lymph nodes    10    48    61
  It has spread to the lymph nodes        19    73    33


> chisq.test(nodexage)

        Pearson's Chi-squared test

data:  nodexage
X-squared = 17.1828, df = 3, p-value = 0.0006481

##node x size

> nodexsize <- table(UNC4_assignment3$node, UNC4_assignment3$size)
> nodexsize

                                       1 cm or 2 cm 3 cm or 4 cm
  It has not spread to the lymph nodes           44           64
  It has spread to the lymph nodes               17           65

                                       5 cm or 6 cm
  It has not spread to the lymph nodes           13
  It has spread to the lymph nodes               30
> chisq.test(nodexsize)

        Pearson's Chi-squared test

data:  nodexsize
X-squared = 19.0848, df = 2, p-value = 7.174e-05


##OncotypeDxrisk x age

> oncoriskxage <- table(UNC4_assignment3$OncotypeDxrisk,
UNC4_assignment3$age)
> oncoriskxage

                    24-39 40-59 60-89
  low risk              4    24    15
  intermediate risk     4    16    15
  high risk            21    82    65

> chisq.test(oncoriskxage)

        Pearson's Chi-squared test
```

```
data:  oncoriskxage
X-squared = 1.312, df = 6, p-value = 0.971


## OncotypeDxrisk x size

> oncoriskxsize <- table(UNC4_assignment3$OncotypeDxrisk,
UNC4_assignment3$size)
> oncoriskxsize

                    1 cm or 2 cm 3 cm or 4 cm 5 cm or 6 cm
  low risk                    12           22            7
  intermediate risk           14           16            3
  high risk                   35           91           33

> chisq.test(oncoriskxsize)

        Pearson's Chi-squared test

data:  oncoriskxsize
X-squared = 7.9292, df = 4, p-value = 0.0942

## age x size

> agexsize <- table(UNC4_assignment3$age, UNC4_assignment3$size)
> agexsize

        1 cm or 2 cm 3 cm or 4 cm 5 cm or 6 cm
  24-39            8           12            5
  40-59           24           66           19
  60-89           25           48           14
> chisq.test(agexsize)

        Pearson's Chi-squared test

data:  agexsize
X-squared = 6.8761, df = 6, p-value = 0.3325
```

##Univariate Cox proportional hazard regression.

```
> library(splines)

> library(survival)
> coxpgr <- coxph(Surv(t.os, e.os) ~ pgr, data=UNC4_assignment3)
> summary(coxpgr)
Call:
coxph(formula = Surv(t.os, e.os) ~ pgr, data = UNC4_assignment3)

  n= 196, number of events= 40
   (109 observations deleted due to missingness)

                 coef exp(coef) se(coef)      z Pr(>|z|)
pgrExpressing -1.5340    0.2157   0.3805 -4.032 5.54e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

              exp(coef) exp(-coef) lower .95 upper .95
pgrExpressing    0.2157      4.637    0.1023    0.4546

Concordance= 0.697  (se = 0.044 )
Rsquare= 0.097   (max possible= 0.842 )
Likelihood ratio test= 19.97  on 1 df,   p=7.848e-06
Wald test            = 16.25  on 1 df,   p=5.542e-05
Score (logrank) test = 19.58  on 1 df,   p=9.624e-06
```

```
> coxpgrtable <- cbind(OR=exp(coef(coxpgr)), exp(confint(coxpgr))
+ )
> coxpgrtable
                    OR    2.5 %    97.5 %
pgrExpressing 0.215666 0.102304 0.4546431


## Cox Proportional Hazard Model, er.
> coxer <- coxph(Surv(t.os, e.os) ~ er, data=UNC4_assignment3)
> summary(coxer)
Call:
coxph(formula = Surv(t.os, e.os) ~ er, data = UNC4_assignment3)

  n= 236, number of events= 51 
   (69 observations deleted due to missingness)

                coef exp(coef) se(coef)      z Pr(>|z|)    
erExpressing -1.2037    0.3001   0.2904 -4.145  3.4e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

             exp(coef) exp(-coef) lower .95 upper .95
erExpressing    0.3001      3.332    0.1698    0.5302

Concordance= 0.652  (se = 0.038 )
Rsquare= 0.072   (max possible= 0.87 )
Likelihood ratio test= 17.69  on 1 df,   p=2.596e-05
Wald test            = 17.18  on 1 df,   p=3.397e-05
Score (logrank) test = 19.22  on 1 df,   p=1.165e-05
> coxertable <- cbind(OR=exp(coef(coxer)), exp(confint(coxer)))
> coxertable
                    OR     2.5 %    97.5 %
erExpressing 0.3000796 0.1698442 0.5301786


## Cox Proportional Hazard Model, her2.
> coxher2 <- coxph(Surv(t.os, e.os) ~ her2, data=UNC4_assignment3)
> summary(coxher2)
Call:
coxph(formula = Surv(t.os, e.os) ~ her2, data = UNC4_assignment3)

  n= 208, number of events= 43 
   (97 observations deleted due to missingness)

                 coef exp(coef) se(coef)     z Pr(>|z|)  
her2Expressing 0.7039    2.0216   0.3435 2.049   0.0405 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

               exp(coef) exp(-coef) lower .95 upper .95
her2Expressing     2.022     0.4947     1.031     3.964

Concordance= 0.545  (se = 0.032 )
Rsquare= 0.018   (max possible= 0.859 )
Likelihood ratio test= 3.76  on 1 df,   p=0.05247
Wald test            = 4.2  on 1 df,   p=0.04046
Score (logrank) test = 4.37  on 1 df,   p=0.03656
> coxher2 <- coxph(Surv(t.os, e.os) ~ her2, data=UNC4_assignment3)
> coxher2table <- cbind(OR=exp(coef(coxher2)), exp(confint(coxher2)))
> coxher2table
                    OR    2.5 %   97.5 %
her2Expressing 2.02156 1.031057 3.963606
```

```
## Cox Proportional Hazard Model, grade.
> coxgrade <- coxph(Surv(t.os, e.os) ~ grade, data=UNC4_assignment3)
> summary(coxgrade)
Call:
coxph(formula = Surv(t.os, e.os) ~ grade, data = UNC4_assignment3)

  n= 224, number of events= 46 
   (81 observations deleted due to missingness)

                          coef exp(coef) se(coef)     z Pr(>|z|)  
gradeIntermediate Grade 0.6268    1.8716   0.7715 0.812   0.4165  
gradeHigh Grade         1.3655    3.9179   0.7313 1.867   0.0619 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                        exp(coef) exp(-coef) lower .95 upper .95
gradeIntermediate Grade     1.872     0.5343    0.4126     8.491
gradeHigh Grade             3.918     0.2552    0.9344    16.427

Concordance= 0.623  (se = 0.042 )
Rsquare= 0.037   (max possible= 0.855 )
Likelihood ratio test= 8.39  on 2 df,   p=0.01509
Wald test            = 7.13  on 2 df,   p=0.02835
Score (logrank) test = 7.75  on 2 df,   p=0.02072
> coxgrade <- coxph(Surv(t.os, e.os) ~ grade, data=UNC4_assignment3)
> coxgradetable <- cbind(OR=exp(coef(coxgrade)), exp(confint(coxgrade)))
> coxgradetable
                              OR     2.5 %    97.5 %
gradeIntermediate Grade 1.871636 0.4125805  8.490515
gradeHigh Grade         3.917870 0.9344042 16.427267
```

```
## Cox Proportional Hazard Model, node.
> coxnode <- coxph(Surv(t.os, e.os) ~ node, data=UNC4_assignment3)
> summary(coxnode)
Call:
coxph(formula = Surv(t.os, e.os) ~ node, data = UNC4_assignment3)

  n= 238, number of events= 49 
   (67 observations deleted due to missingness)

                                     coef exp(coef) se(coef)     z
nodeIt has spread to the lymph nodes 1.1614    3.1943   0.3332 3.486
                                     Pr(>|z|)    
nodeIt has spread to the lymph nodes 0.000491 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                                     exp(coef) exp(-coef) lower .95
nodeIt has spread to the lymph nodes     3.194     0.3131     1.663
                                     upper .95
nodeIt has spread to the lymph nodes     6.137

Concordance= 0.627  (se = 0.04 )
Rsquare= 0.058   (max possible= 0.858 )
Likelihood ratio test= 14.16  on 1 df,   p=0.0001679
Wald test            = 12.15  on 1 df,   p=0.0004908
Score (logrank) test = 13.55  on 1 df,   p=0.0002322
> coxnode <- coxph(Surv(t.os, e.os) ~ node, data=UNC4_assignment3)
> coxnodetable <- cbind(OR=exp(coef(coxnode)), exp(confint(coxnode)))
> coxnodetable
                                           OR    2.5 %   97.5 %
nodeIt has spread to the lymph nodes 3.194271 1.662535 6.137237
```

```
## Cox Proportional Hazard Model, OncotypeDxrisk.
> coxoncorisk <- coxph(Surv(t.os, e.os) ~ OncotypeDxrisk, 
data=UNC4_assignment3)
> summary(coxoncorisk)
Call:
coxph(formula = Surv(t.os, e.os) ~ OncotypeDxrisk, data = UNC4_assignment3)

  n= 240, number of events= 51 
   (65 observations deleted due to missingness)

                                   coef exp(coef) se(coef)      z
OncotypeDxriskintermediate risk -0.4542    0.6350   0.5399 -0.841
OncotypeDxriskhigh risk         -0.0365    0.9642   0.3486 -0.105
                                Pr(>|z|)
OncotypeDxriskintermediate risk    0.400
OncotypeDxriskhigh risk            0.917

                                exp(coef) exp(-coef) lower .95
OncotypeDxriskintermediate risk    0.6350      1.575    0.2204
OncotypeDxriskhigh risk            0.9642      1.037    0.4868
                                upper .95
OncotypeDxriskintermediate risk     1.829
OncotypeDxriskhigh risk             1.909

Concordance= 0.517  (se = 0.037 )
Rsquare= 0.004   (max possible= 0.867 )
Likelihood ratio test= 0.93  on 2 df,   p=0.6287
Wald test            = 0.83  on 2 df,   p=0.6603
Score (logrank) test = 0.84  on 2 df,   p=0.6562
> coxoncorisk <- coxph(Surv(t.os, e.os) ~ OncotypeDxrisk, 
data=UNC4_assignment3)
> coxoncorisktable <- cbind(OR=exp(coef(coxoncorisk)), 
exp(confint(coxoncorisk)))
> coxoncorisktable
                                       OR     2.5 %   97.5 %
OncotypeDxriskintermediate risk 0.6349612 0.2204058 1.829243
OncotypeDxriskhigh risk         0.9641616 0.4868428 1.909462
```

```
## Cox Proportional Hazard Model, Age.

> coxage <- coxph(Surv(t.os, e.os) ~ age, data=UNC4_assignment3)
> summary(coxage)
Call:
coxph(formula = Surv(t.os, e.os) ~ age, data = UNC4_assignment3)

  n= 225, number of events= 46 
   (80 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
age40-59 -0.1965    0.8216   0.4984 -0.394    0.693
age60-89 -0.1059    0.8995   0.5011 -0.211    0.833

         exp(coef) exp(-coef) lower .95 upper .95
age40-59    0.8216      1.217    0.3093     2.182
age60-89    0.8995      1.112    0.3369     2.402

Concordance= 0.53  (se = 0.043 )
Rsquare= 0.001   (max possible= 0.852 )
Likelihood ratio test= 0.18  on 2 df,   p=0.912
Wald test            = 0.19  on 2 df,   p=0.9108
```

```
Score (logrank) test = 0.19  on 2 df,   p=0.9106


## Changing reference groups for certain variables.

> UNC4_assignment3$pgr <-cut(UNC4_assignment3$pgr, br=c(-1,0,1))
> levels(UNC4_assignment3$pgr)<- c("Not expressing", "Expressing")
> UNC4_assignment3$er <-cut(UNC4_assignment3$er, br=c(-1,0,1))
> levels(UNC4_assignment3$er)<- c("Not expressing", "Expressing")
> UNC4_assignment3$her2 <-cut(UNC4_assignment3$her2, br=c(-1,0,1))
> levels(UNC4_assignment3$her2)<- c("Not expressing", "Expressing")
> pgrxer <- table(UNC4_assignment3$pgr, UNC4_assignment3$er)

> library(MASS)

##Chi-square test for pgr x er
> chisq.test(pgrxer)

        Pearson's Chi-squared test with Yates' continuity correction

data:  pgrxer
X-squared = 76.2334, df = 1, p-value < 2.2e-16

> pgrxer

                 Not expressing Expressing
  Not expressing             65         35
  Expressing                  6         98

##Chi-square test for pgr x her2


> pgrxher2 <- table(UNC4_assignment3$pgr, UNC4_assignment3$her2)
> chisq.test(pgrxher2)

        Pearson's Chi-squared test with Yates' continuity correction

data:  pgrxher2
X-squared = 0.9207, df = 1, p-value = 0.3373

##Chi-square test for pgr x er.

> pgrxer <- table(UNC4_assignment3$pgr, UNC4_assignment3$er)
> pgrxer

                 Not expressing Expressing
  Not expressing             65         35
  Expressing                  6         98
> chisq.test(pgrxer)

        Pearson's Chi-squared test with Yates' continuity correction

data:  pgrxer
X-squared = 76.2334, df = 1, p-value < 2.2e-16

##Chi-square test for er x her2

> erxher2 <- table(UNC4_assignment3$er, UNC4_assignment3$her2)
> chisq.test(erxher2)

        Pearson's Chi-squared test with Yates' continuity correction

data:  erxher2
```

```
X-squared = 6.7187, df = 1, p-value = 0.009541


## Changing some variables to categorical.

> UNC4_assignment3$grade <-cut(UNC4_assignment3$grade, br=c(0,1,2,3))
> levels(UNC4_assignment3$grade) <-c("Low Grade","Intermediate Grade","High
Grade")
> UNC4_assignment3$node <-cut(UNC4_assignment3$node, br=c(-1,0,1))
> levels(UNC4_assignment3$node) <-c("It has not spread to the lymph
nodes","It has spread to the lymph nodes")
> UNC4_assignment3$OncotypeDxrisk <-cut(UNC4_assignment3$OncotypeDxrisk,
br=c(-1,0,0.5,1))
> levels(UNC4_assignment3$OncotypeDxrisk) <-c("low risk","intermediate
risk","high risk")

> library(splines)
> library(survival)
## Putting the univariate Cox regression (with hazard ratios, and 95% CIs)
table together.

##Pgr, univariate Cox model
> coxpgr <- coxph(Surv(t.os, e.os) ~ pgr, data=UNC4_assignment3)
> summary(coxpgr)
Call:
coxph(formula = Surv(t.os, e.os) ~ pgr, data = UNC4_assignment3)

  n= 196, number of events= 40
   (109 observations deleted due to missingness)

                 coef exp(coef) se(coef)      z Pr(>|z|)
pgrExpressing -1.5340    0.2157   0.3805 -4.032 5.54e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

              exp(coef) exp(-coef) lower .95 upper .95
pgrExpressing    0.2157      4.637    0.1023    0.4546

Concordance= 0.697  (se = 0.044 )
Rsquare= 0.097   (max possible= 0.842 )
Likelihood ratio test= 19.97  on 1 df,   p=7.848e-06
Wald test            = 16.25  on 1 df,   p=5.542e-05
Score (logrank) test = 19.58  on 1 df,   p=9.624e-06

> coxpgrtable <- cbind(HR=exp(coef(coxpgr)), exp(confint(coxpgr))
+ )
> coxpgrtable
                    HR    2.5 %    97.5 %
pgrExpressing 0.215666 0.102304 0.4546431

## er, univariate Cox model.
> coxer <- coxph(Surv(t.os, e.os) ~ er, data=UNC4_assignment3)
> summary(coxer)
Call:
coxph(formula = Surv(t.os, e.os) ~ er, data = UNC4_assignment3)

  n= 236, number of events= 51
   (69 observations deleted due to missingness)

                coef exp(coef) se(coef)      z Pr(>|z|)
erExpressing -1.2037    0.3001   0.2904 -4.145  3.4e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
             exp(coef) exp(-coef) lower .95 upper .95
erExpressing    0.3001      3.332    0.1698    0.5302

Concordance= 0.652  (se = 0.038 )
Rsquare= 0.072   (max possible= 0.87 )
Likelihood ratio test= 17.69  on 1 df,   p=2.596e-05
Wald test            = 17.18  on 1 df,   p=3.397e-05
Score (logrank) test = 19.22  on 1 df,   p=1.165e-05

> coxpgrtable <- cbind(HR=exp(coef(coxer)), exp(confint(coxer)))
> coxpgrtable <- cbind(HR=exp(coef(coxpgr)), exp(confint(coxpgr))
+ )
> coxpgrtable
                    HR    2.5 %    97.5 %
pgrExpressing 0.215666 0.102304 0.4546431
> coxertable <- cbind(HR=exp(coef(coxer)), exp(confint(coxer)))
> coxertable
                    HR     2.5 %    97.5 %
erExpressing 0.3000796 0.1698442 0.5301786


##Her2, univariate Cox regression.
> coxher2 <- coxph(Surv(t.os, e.os) ~ her2, data=UNC4_assignment3)
> coxher2table <- cbind(HR=exp(coef(coxher2)), exp(confint(coxher2)))
> coxher2table
                    HR    2.5 %   97.5 %
her2Expressing 2.02156 1.031057 3.963606


##Grade, univariate Cox regression.
> coxgrade <- coxph(Surv(t.os, e.os) ~ grade, data=UNC4_assignment3)
> coxgradetable <- cbind(HR=exp(coef(coxgrade)), exp(confint(coxgrade)))
> coxgradetable
                              HR     2.5 %    97.5 %
gradeIntermediate Grade 1.871636 0.4125805  8.490515
gradeHigh Grade         3.917870 0.9344042 16.427267

##Node, univariate Cox regression.
> coxnode <- coxph(Surv(t.os, e.os) ~ node, data=UNC4_assignment3)
> coxnodetable <- cbind(HR=exp(coef(coxnode)), exp(confint(coxnode)))
> coxnodetable
                                           HR    2.5 %   97.5 %
nodeIt has spread to the lymph nodes 3.194271 1.662535 6.137237


##OncotypeDxrisk, univariate Cox regression.
> coxoncorisk <- coxph(Surv(t.os, e.os) ~ OncotypeDxrisk,
data=UNC4_assignment3)
> coxoncorisktable <- cbind(HR=exp(coef(coxoncorisk)),
exp(confint(coxoncorisk)))
> coxoncorisktable
                                       HR     2.5 %   97.5 %
OncotypeDxriskintermediate risk 0.6349612 0.2204058 1.829243
OncotypeDxriskhigh risk         0.9641616 0.4868428 1.909462

##Age, univariate Cox regression.
> coxage <- coxph(Surv(t.os, e.os) ~ age, data=UNC4_assignment3)
> summary(coxage)
Call:
coxph(formula = Surv(t.os, e.os) ~ age, data = UNC4_assignment3)

  n= 225, number of events= 46 
   (80 observations deleted due to missingness)
```

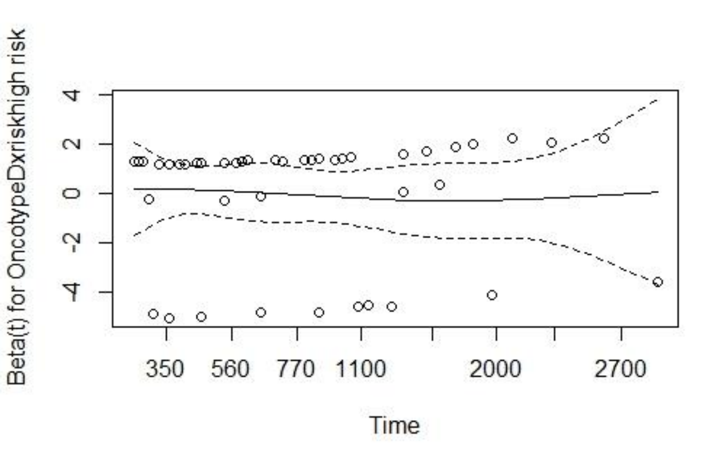
```
         coef exp(coef)  se(coef)      z Pr(>|z|)
age -0.002953  0.997052  0.009906 -0.298    0.766

    exp(coef) exp(-coef) lower .95 upper .95
age    0.9971      1.003    0.9779     1.017

Concordance= 0.498  (se = 0.048 )
Rsquare= 0   (max possible= 0.852 )
Likelihood ratio test= 0.09  on 1 df,   p=0.7653
Wald test            = 0.09  on 1 df,   p=0.7657
Score (logrank) test = 0.09  on 1 df,   p=0.7656

> coxagetable <- cbind(HR=exp(coef(coxage)), exp(confint(coxage)))
> coxagetable
           HR     2.5 % 97.5 %
age 0.9970517 0.9778795 1.0166



## Cox Proportional Hazard Model, Size.

> coxsize <- coxph(Surv(t.os, e.os) ~ size, data=UNC4_assignment3)
> summary(coxsize)
Call:
coxph(formula = Surv(t.os, e.os) ~ size, data = UNC4_assignment3)

  n= 214, number of events= 37 
   (91 observations deleted due to missingness)

                    coef exp(coef) se(coef)    z Pr(>|z|)   
size3 cm or 4 cm  1.7449    5.7254   0.7394 2.36  0.01829 * 
size5 cm or 6 cm  2.3925   10.9412   0.7620 3.14  0.00169 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                 exp(coef) exp(-coef) lower .95 upper .95
size3 cm or 4 cm     5.725     0.1747     1.344     24.39
size5 cm or 6 cm    10.941     0.0914     2.457     48.71

Concordance= 0.718  (se = 0.049 )
Rsquare= 0.072   (max possible= 0.793 )
Likelihood ratio test= 15.88  on 2 df,   p=0.0003556
Wald test            = 10.79  on 2 df,   p=0.004548
Score (logrank) test = 14.12  on 2 df,   p=0.0008582
> coxsizetable <- cbind(HR=exp(coef(coxsize)), exp(confint(coxsize)))
> coxsizetable
           HR    2.5 %   97.5 %
size 1.402578 1.169536 1.682056
```

##Question 1d.

Cox Proportional Hazards Model for OncotypeDxscore.

```
> summary(UNC4_assignment3$OncotypeDxscore)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   25.71   56.08   57.41  100.00  100.00 
> coxoncotypescore <- coxph(Surv(t.os, e.os) ~ OncotypeDxscore, 
data=UNC4_assignment3)
> summary(coxoncotypescore)
Call:
coxph(formula = Surv(t.os, e.os) ~ OncotypeDxscore, data = UNC4_assignment3)
```

```
  n= 240, number of events= 51 
   (65 observations deleted due to missingness)

                    coef exp(coef) se(coef)     z Pr(>|z|)
OncotypeDxscore 0.004023  1.004031 0.004028 0.999    0.318

                exp(coef) exp(-coef) lower .95 upper .95
OncotypeDxscore     1.004      0.996    0.9961     1.012

Concordance= 0.537  (se = 0.045 )
Rsquare= 0.004   (max possible= 0.867 )
Likelihood ratio test= 1  on 1 df,   p=0.317
Wald test            = 1  on 1 df,   p=0.3179
Score (logrank) test = 1  on 1 df,   p=0.3168

> schoenfeldoncorisk <- cox.zph(coxoncorisk)
> plot(schoenfeldoncorisk)
```



```
##OncotypeDxscore, univariate Cox model.
> coxoncoscoretable <- cbind(HR=exp(coef(coxoncoscore)), 
exp(confint(coxoncoscore)))
> coxoncoscoretable
                      HR     2.5 %   97.5 %
OncotypeDxscore 1.004031 0.9961358 1.011989


## Creating entire table.

> coxunivtable <- rbind(coxpgrtable, coxertable, coxher2table, coxgradetable, 
coxnodetable, coxoncorisktable, coxagetable, coxsizetable, coxoncoscoretable)
> coxunivtable
                                         HR     2.5 %     97.5 %
pgrExpressing                     0.2156660 0.1023040  0.4546431
erExpressing                      0.3000796 0.1698442  0.5301786
her2Expressing                    2.0215601 1.0310574  3.9636057
gradeIntermediate Grade           1.8716359 0.4125805  8.4905155
gradeHigh Grade                   3.9178703 0.9344042 16.4272666
nodeIt has spread to the lymph nodes 3.1942712 1.6625346  6.1372367
OncotypeDxriskintermediate risk   0.6349612 0.2204058  1.8292431
OncotypeDxriskhigh risk           0.9641616 0.4868428  1.9094615
age                               0.9970517 0.9778795  1.0165998
size                              1.4025779 1.1695356  1.6820564
OncotypeDxscore                   1.004031  0.9961358  1.011989

> write.csv(coxunivtable, "C:/Users/James/Desktop/coxunivtable.csv")
```

```
##Part 2.
##Subsetting data to only include node-negative patients.
> summary(UNC4_assignment3$node)
It has not spread to the lymph nodes
                                 126
    It has spread to the lymph nodes
                                 134
                                NA's
                                  45
> levels(UNC4_assignment3$node) <-c("0","1")
> summary(UNC4_assignment3$node)
   0    1 NA's
 126  134   45

> UNC4_assignment3$node[UNC4_assignment3$node==1] <- NA
> final <- UNC4_assignment3[complete.cases(UNC4_assignment3[c("node")]),]
> write.csv(final, "C:/Users/James/Desktop/Question2.csv")
```

```
Part 2:

##Changing referent categories.

> levels(Question2$pgr)
[1] "Expressing"     "Not expressing"
> Question2$pgr2=factor(Question2$pgr, c("Not
expressing","Expressing"))
> levels(Question2$pgr2)
[1] "Not expressing" "Expressing"
> Question2$er2=factor(Question2$er, c("Not expressing","Expressing"))
> levels(Question2$er2)
[1] "Not expressing" "Expressing"
> Question2$her22=factor(Question2$her2, c("Not expressing","Expressing"))
> levels(Question2$her22)
[1] "Not expressing" "Expressing"
```

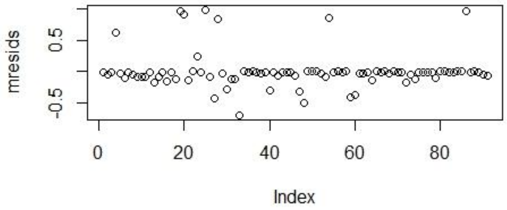
```
##Initial Cox model with all predictors.
> allmodel <- coxph(Surv(t.os, e.os) ~ pgr2+er2+her22+grade+size+age,
data=Question2)
> summary(allmodel)
Call:
coxph(formula = Surv(t.os, e.os) ~ pgr2 + er2 + her22 + grade +
    size + age, data = Question2)

  n= 89, number of events= 8
   (28 observations deleted due to missingness)

                            coef exp(coef) se(coef)      z Pr(>|z|)
pgr2Expressing          -1.86098   0.15552  1.23933 -1.502    0.133
er2Expressing           -0.75210   0.47137  1.01297 -0.742    0.458
her22Expressing         -1.61586   0.19872  1.46360 -1.104    0.270
gradeIntermediate Grade -0.85601   0.42486  1.14592 -0.747    0.455
gradeLow Grade           0.61399   1.84780  1.28466  0.478    0.633
size                     0.41973   1.52156  0.28531  1.471    0.141
age                      0.04721   1.04835  0.02785  1.695    0.090

pgr2Expressing
er2Expressing
her22Expressing
```

```
gradeIntermediate Grade
gradeLow Grade
size
age                     .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                        exp(coef) exp(-coef) lower .95 upper .95
pgr2Expressing             0.1555     6.4300   0.01370     1.765
er2Expressing              0.4714     2.1215   0.06473     3.432
her22Expressing            0.1987     5.0322   0.01128     3.500
gradeIntermediate Grade    0.4249     2.3537   0.04496     4.015
gradeLow Grade             1.8478     0.5412   0.14899    22.917
size                       1.5216     0.6572   0.86982     2.662
age                        1.0483     0.9539   0.99266     1.107

Concordance= 0.834  (se = 0.113 )
Rsquare= 0.129   (max possible= 0.487 )
Likelihood ratio test= 12.25  on 7 df,   p=0.09254
Wald test            = 8.36  on 7 df,   p=0.3016
Score (logrank) test = 10.73  on 7 df,   p=0.1506

> AIC(allmodel)
[1] 61.19693


##Backward Step Selection.
## Taking out er decreases AIC.
> allmodelwithouter <- coxph(Surv(t.os, e.os) ~ pgr2+her22+grade+size+age, 
data=Question2)
> AIC(allmodelwithouter)
[1] 59.79456
##Toss this out.


##Taking out pgr increases AIC.
> allmodelwithoutpgr <- coxph(Surv(t.os, e.os) ~ er2+her22+grade+size+age, 
data=Question2)
> AIC(allmodelwithoutpgr)
[1] 66.24968
##Keep this in.


##Taking out her2 increases AIC.
> allmodelwithouther2 <- coxph(Surv(t.os, e.os) ~ pgr2+er2+grade+size+age, 
data=Question2)
> AIC(allmodelwithouther2)
[1] 63.22979
##Keep this in.

##Taking out grade decreases AIC.
> allmodelwithoutergrade <- coxph(Surv(t.os, e.os) ~ pgr2+her22+size+age, 
data=Question2)
> AIC(allmodelwithoutergrade)
[1] 57.92907
##Toss this out.

##Taking out size increases AIC.
> allmodelwithoutergradesize <- coxph(Surv(t.os, e.os) ~ pgr2+her22+age, 
data=Question2)
> AIC(allmodelwithoutergradesize)
[1] 73.48523
## Keep this in.
```

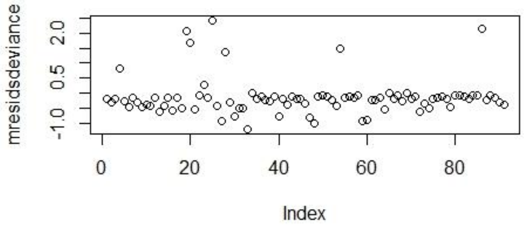
```
##Taking out age will increase AIC.
> allmodelwithoutergradeage <- coxph(Surv(t.os, e.os) ~ pgr2+her22+size,
data=Question2)
> AIC(allmodelwithoutergradeage)
[1] 58.17043
##Keep this in.

##Final Model.
> allmodelwithoutergrade <- coxph(Surv(t.os, e.os) ~ pgr2+her22+size+age,
data=Question2)
> AIC(allmodelwithoutergrade)
[1] 57.92907
```

```
##Final Model:

> allmodelwithoutergrade <- coxph(Surv(t.os, e.os) ~ pgr2+her22+size+age,
data=Question2)
> summary(allmodelwithoutergrade)
Call:
coxph(formula = Surv(t.os, e.os) ~ pgr2 + her22 + size + age, 
    data = Question2)

  n= 91, number of events= 8 
   (26 observations deleted due to missingness)

                   coef exp(coef) se(coef)      z Pr(>|z|)  
pgr2Expressing  -2.1199    0.1201   1.1451 -1.851   0.0641 .
her22Expressing -1.4561    0.2331   1.5404 -0.945   0.3445  
size             0.3228    1.3810   0.2354  1.371   0.1703  
age              0.0398    1.0406   0.0283  1.407   0.1596  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                exp(coef) exp(-coef) lower .95 upper .95
pgr2Expressing     0.1200     8.3301   0.01272     1.133
her22Expressing    0.2331     4.2892   0.01139     4.774
size               1.3810     0.7241   0.87056     2.191
age                1.0406     0.9610   0.98446     1.100

Concordance= 0.813  (se = 0.113 )
Rsquare= 0.104   (max possible= 0.482 )
Likelihood ratio test= 9.97  on 4 df,   p=0.04102
Wald test            = 7.07  on 4 df,   p=0.1323
Score (logrank) test = 9.21  on 4 df,   p=0.05616

> AIC(allmodelwithoutergrade)
[1] 57.92907
```

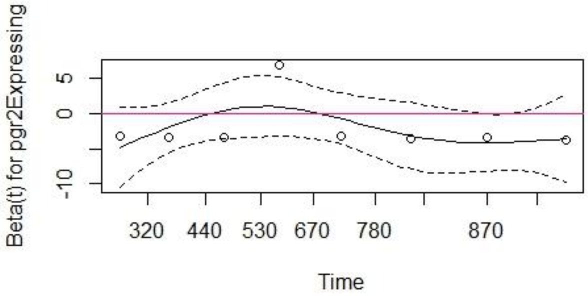
```
##Residuals Analysis:

##Martingale:

> mresids <- residuals(allmodelwithoutergrade,type="martingale")
> plot(mresids)
```



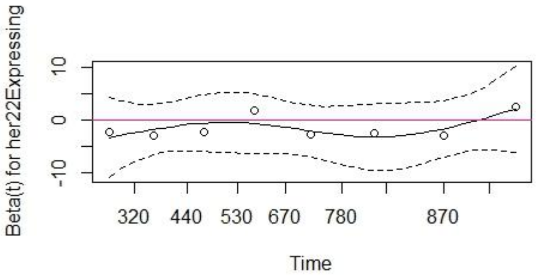
```
## Deviance residuals.
> mresidsdeviance <- residuals(allmodelwithouternode,type="deviance")
> plot(mresidsdeviance)
```



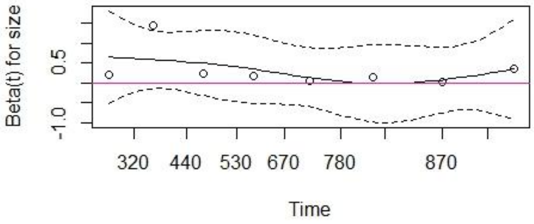
```
##Schoenfeld Residuals for final model.

> schoenfeldfit <- cox.zph(allmodelwithoutergrade)


## Schoenfeld residuals plot for final model’s pgr.
> plot(schoenfeldfit[1])
>abline(h=0, col="deeppink1")
```



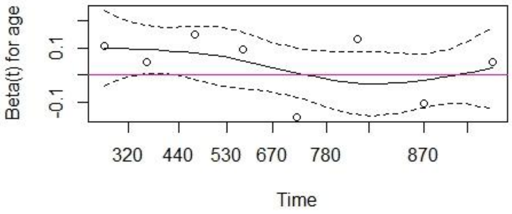
```
## Schoenfeld residuals plot for final model’s er.
> plot(schoenfeldfit[2])
> abline(h=0, col="deeppink1")
```



```
## Schoenfeld residuals plot for final model’s size.
> plot(schoenfeldfit[3])
> abline(h=0, col="deeppink1")
```



```
## Schoenfeld residuals plot for final model’s age.
> plot(schoenfeldfit[4])
> abline(h=0, col="deeppink1")
```



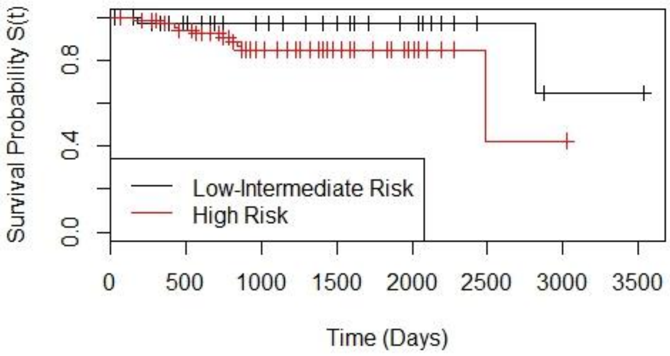
```
##Adding OncotypeDxscore to the model yields a lower.
> allmodelwithoutergradewithoncoscore <- coxph(Surv(t.os, e.os) ~
pgr2+her22+size+age+OncotypeDxscore, data=Question2)
> AIC(allmodelwithoutergradewithoncoscore)
[1] 55.36747

> summary(allmodelwithoutergradewithoncoscore)
Call:
coxph(formula = Surv(t.os, e.os) ~ pgr2 + her22 + size + age +
    OncotypeDxscore, data = Question2)

  n= 91, number of events= 8
   (26 observations deleted due to missingness)

                    coef exp(coef) se(coef)      z Pr(>|z|)
pgr2Expressing  -2.57698   0.07600  1.29478 -1.990   0.0466 *
her22Expressing -2.79576   0.06107  2.90257 -0.963   0.3354
```

```
size                0.57354   1.77453  0.32402  1.770   0.0767 .
age                 0.04818   1.04936  0.02672  1.803   0.0714 .
OncotypeDxscore     0.02565   1.02599  0.01307  1.963   0.0496 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                exp(coef) exp(-coef) lower .95 upper .95
pgr2Expressing    0.07600    13.1573 0.0060078    0.9615
her22Expressing   0.06107    16.3751 0.0002066   18.0506
size              1.77453     0.5635 0.9403151    3.3488
age               1.04936     0.9530 0.9958187    1.1058
OncotypeDxscore   1.02599     0.9747 1.0000429    1.0526

Concordance= 0.857  (se = 0.113 )
Rsquare= 0.148   (max possible= 0.482 )
Likelihood ratio test= 14.53  on 5 df,   p=0.01259
Wald test            = 7.93  on 5 df,   p=0.1601
Score (logrank) test = 11.43  on 5 df,   p=0.04346


##Creating the table for adding Oncotypedxscore to the final Cox model

> oncoscoretable <- cbind(HR=exp(coef(allmodelwithoutergradewithoncoscore)),
exp(confint(allmodelwithoutergradewithoncoscore)),
AIC(allmodelwithoutergradewithoncoscore))
> oncoscoretable
                        HR        2.5 %     97.5 %
pgr2Expressing  0.07600330 0.0060077614  0.9615066 55.36747
her22Expressing 0.06106837 0.0002066052 18.0505923 55.36747
size            1.77453250 0.9403150757  3.3488409 55.36747
age             1.04936026 0.9958187315  1.1057805 55.36747
OncotypeDxscore 1.02598633 1.0000428603  1.0526028 55.36747
> write.csv(oncoscoretable, "C:/Users/James/Desktop/Oncoscoretable.csv")


## Adding OncotypeDxrisk to final Cox model.
## Need to fix the OncotypeDxrisk variable. Variable 5,6 are the categories
for OncotypeDxrisk. OncotypeDxrisk variable is the problematic one.
> allmodelwithoutergradewithoncorisk <- coxph(Surv(t.os, e.os) ~
pgr2+her22+age+size+OncotypeDxrisk2, data=Question2)
Warning message:
In fitter(X, Y, strats, offset, init, control, weights = weights,  :
  Loglik converged before variable  5,6 ; beta may be infinite.




## Combined low and intermediate categories because there were too few
individuals in the low risk category. So, I decided to combine the
intermediate category with the low category.
##Adding OncotypeDxrisk to the model yields a lower AIC.
> Question2$OncotypeDxrisk2=factor(Question2$OncotypeDxrisk, c("low
risk","intermediate risk","high risk"))
> levels(Question2$OncotypeDxrisk2)
[1] "low risk"          "intermediate risk" "high risk"
> levels(Question2$OncotypeDxrisk2) <- c("low-intermediate risk", "low-
intermediate risk", "high risk")
> allmodelwithoutergradewithoncorisk <- coxph(Surv(t.os, e.os) ~
pgr2+her22+age+size+OncotypeDxrisk2, data=Question2)
> summary(allmodelwithoutergradewithoncorisk)
Call:
coxph(formula = Surv(t.os, e.os) ~ pgr2 + her22 + age + size +
    OncotypeDxrisk2, data = Question2)
```

```
  n= 91, number of events= 8 
   (26 observations deleted due to missingness)

                             coef exp(coef) se(coef)      z
pgr2Expressing           -2.24892   0.10551  1.17175 -1.919
her22Expressing          -2.14751   0.11677  2.38916 -0.899
age                       0.04074   1.04158  0.02607  1.563
size                      0.50902   1.66366  0.29902  1.702
OncotypeDxrisk2high risk  1.68180   5.37522  1.17806  1.428
                         Pr(>|z|)  
pgr2Expressing             0.0549 .
her22Expressing            0.3687  
age                        0.1180  
size                       0.0887 .
OncotypeDxrisk2high risk   0.1534  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                         exp(coef) exp(-coef) lower .95 upper .95
pgr2Expressing              0.1055     9.4775  0.010615     1.049
her22Expressing             0.1168     8.5635  0.001081    12.618
age                         1.0416     0.9601  0.989706     1.096
size                        1.6637     0.6011  0.925844     2.989
OncotypeDxrisk2high risk    5.3752     0.1860  0.534109    54.096

Concordance= 0.807  (se = 0.113 )
Rsquare= 0.131   (max possible= 0.482 )
Likelihood ratio test= 12.77  on 5 df,   p=0.02561
Wald test            = 7.45  on 5 df,   p=0.1893
Score (logrank) test = 10.15  on 5 df,   p=0.07115

> AIC(allmodelwithoutergradewithoncorisk)
[1] 57.12249

##Creating oncotypedxrisk table (oncotypedxrisk added to Cox model).
> oncorisktable <- cbind(HR=exp(coef(allmodelwithoutergradewithoncorisk)),
exp(confint(allmodelwithoutergradewithoncorisk)),
AIC(allmodelwithoutergradewithoncorisk))
> oncorisktable
                                HR       2.5 %    97.5 %         
pgr2Expressing           0.1055131 0.010614770  1.048822 57.12249
her22Expressing          0.1167748 0.001080664 12.618491 57.12249
age                      1.0415841 0.989706420  1.096181 57.12249
size                     1.6636580 0.925844328  2.989442 57.12249
OncotypeDxrisk2high risk 5.3752166 0.534109137 54.095599 57.12249
> write.csv(oncorisktable, "C:/Users/James/Desktop/Oncorisktable.csv")


##Cox OncotypeDxrisk Unadjusted for other variables.

> unadjoncorisk <- coxph(Surv(t.os, e.os) ~ OncotypeDxrisk2, data=Question2)
> summary(unadjoncorisk)
Call:
coxph(formula = Surv(t.os, e.os) ~ OncotypeDxrisk2, data = Question2)

  n= 117, number of events= 12 

                           coef exp(coef) se(coef)    z Pr(>|z|)
OncotypeDxrisk2high risk 1.2336    3.4336   0.8062 1.53    0.126

                         exp(coef) exp(-coef) lower .95 upper .95
OncotypeDxrisk2high risk     3.434     0.2912    0.7072     16.67

Concordance= 0.602  (se = 0.076 )
```

```
Rsquare= 0.025   (max possible= 0.558 )
Likelihood ratio test= 2.92  on 1 df,   p=0.08724
Wald test            = 2.34  on 1 df,   p=0.126
Score (logrank) test = 2.58  on 1 df,   p=0.1084
```

```
##Unadjusted KM plot.

> kmfitunadjusted <- survfit(Surv(t.os, e.os)~OncotypeDxrisk2,
data=Question2)

> plot(kmfitunadjusted, col=c(1:2), xlab="Time (Days)", ylab="Survival
Probability S(t)")

> legend("bottomleft", legend=c("Low-Intermediate Risk","High Risk"),
col=(1:2), lwd=0.5)
```

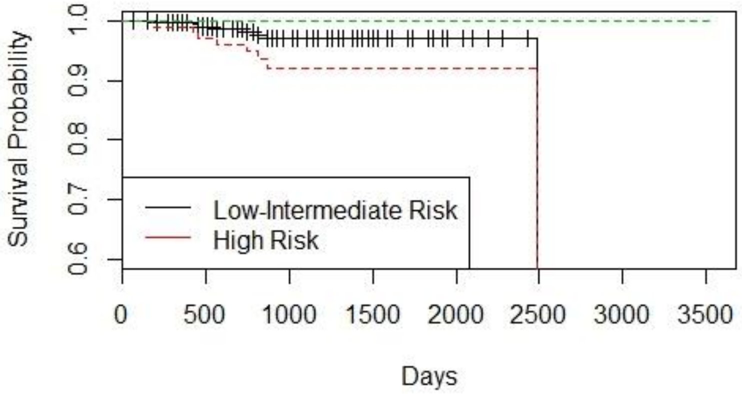


```
##Log rank Test for OncotypeDxrisk(unadjusted).

> survdiff(Surv(t.os, e.os) ~ OncotypeDxrisk2, data=Question2)
Call:
survdiff(formula = Surv(t.os, e.os) ~ OncotypeDxrisk2, data = Question2)

                                     N Observed Expected
OncotypeDxrisk2=low-intermediate risk 41        2      4.6
OncotypeDxrisk2=high risk             76       10      7.4
                                    (O-E)^2/E (O-E)^2/V
OncotypeDxrisk2=low-intermediate risk     1.468      2.58
OncotypeDxrisk2=high risk                 0.912      2.58

 Chisq= 2.6  on 1 degrees of freedom, p= 0.108
```

```
##Wilcoxon test (Alternative to log rank test.)
> survdiff(Surv(Question2$t.os, Question2$e.os) ~ Question2$OncotypeDxrisk2,
rho=1)
Call:
survdiff(formula = Surv(Question2$t.os, Question2$e.os) ~
Question2$OncotypeDxrisk2,
    rho = 1)

                                               N Observed Expected
Question2$OncotypeDxrisk2=low-intermediate risk 41     1.70     4.17
```

```
Question2$OncotypeDxrisk2=high risk             76      9.45      6.98
                                                (O-E)^2/E (O-E)^2/V
Question2$OncotypeDxrisk2=low-intermediate risk     1.462      2.66
Question2$OncotypeDxrisk2=high risk                 0.873      2.66

 Chisq= 2.7  on 1 degrees of freedom, p= 0.103

## Tarone-ware test (Also another alternative).
> survdiff(Surv(Question2$t.os, Question2$e.os) ~ Question2$OncotypeDxrisk2,
rho=1.5)
Call:
survdiff(formula = Surv(Question2$t.os, Question2$e.os) ~ 
Question2$OncotypeDxrisk2, 
    rho = 1.5)

                                                 N Observed Expected
Question2$OncotypeDxrisk2=low-intermediate risk 41      1.59      3.99
Question2$OncotypeDxrisk2=high risk             76      9.20      6.79
                                                (O-E)^2/E (O-E)^2/V
Question2$OncotypeDxrisk2=low-intermediate risk      1.45      2.68
Question2$OncotypeDxrisk2=high risk                  0.85      2.68

 Chisq= 2.7  on 1 degrees of freedom, p= 0.102
```

```
##For survival analysis information.

> summary(kmfitunadjusted)
Call: survfit(formula = Surv(t.os, e.os) ~ OncotypeDxrisk2, data = Question2)

                OncotypeDxrisk2=low-intermediate risk 
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  180     38       1    0.974   0.026        0.924            1
 2820      3       1    0.649   0.266        0.291            1

                OncotypeDxrisk2=high risk 
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  150     75       1    0.987  0.0132        0.961        1.000
  330     71       1    0.973  0.0190        0.936        1.000
  420     67       1    0.958  0.0236        0.913        1.000
  450     66       1    0.944  0.0274        0.892        0.999
  570     60       1    0.928  0.0311        0.869        0.991
  750     53       1    0.910  0.0351        0.844        0.982
  810     47       1    0.891  0.0393        0.817        0.972
  840     45       1    0.871  0.0432        0.791        0.960
  870     44       1    0.852  0.0465        0.765        0.948
 2490      2       1    0.426  0.3020        0.106        1.000

##Adjusted KM plot.

> adjsurvivalcurve <- plot(survfit(allmodelwithoutergradewithoncorisk),
col=c(1:2), ylim=c(.6, 1), xlab='Days', ylab='Survival Probability')
> legend("bottomleft", legend=c("Low-Intermediate Risk","High Risk"),
col=(1:2), lwd=0.5)
```


Survival Probability
1.0
0.9
0.8
0.7
0.6
0
500
1000
1500
2000
2500
3000
3500
Days
Low-Intermediate Risk
High Risk