SURVIVAL ANALYSIS - TCGA PRAD CANCER

Kelvin Ofori-Minta   University of Texas at El Paso (UTEP)

June 27, 2022

Contents

1 Loading and Cleaning Data

```r
data <- read.csv("thyroidcancer.csv", header = T, na.strings = "NA")

data[data==""]<-NA #replace all empty cells with na

# write.csv(data,"C://Users//Kelvin//Desktop//Spring 2022//research with Dr. Leung//survival//
```

1.1 Inspecting dataframe for missing values

```r
require(inspectdf)
show_plot(inspect_na(data))
```

Prevalence of NAs in df::data

df::data has 37 columns, of which 14 have missing values
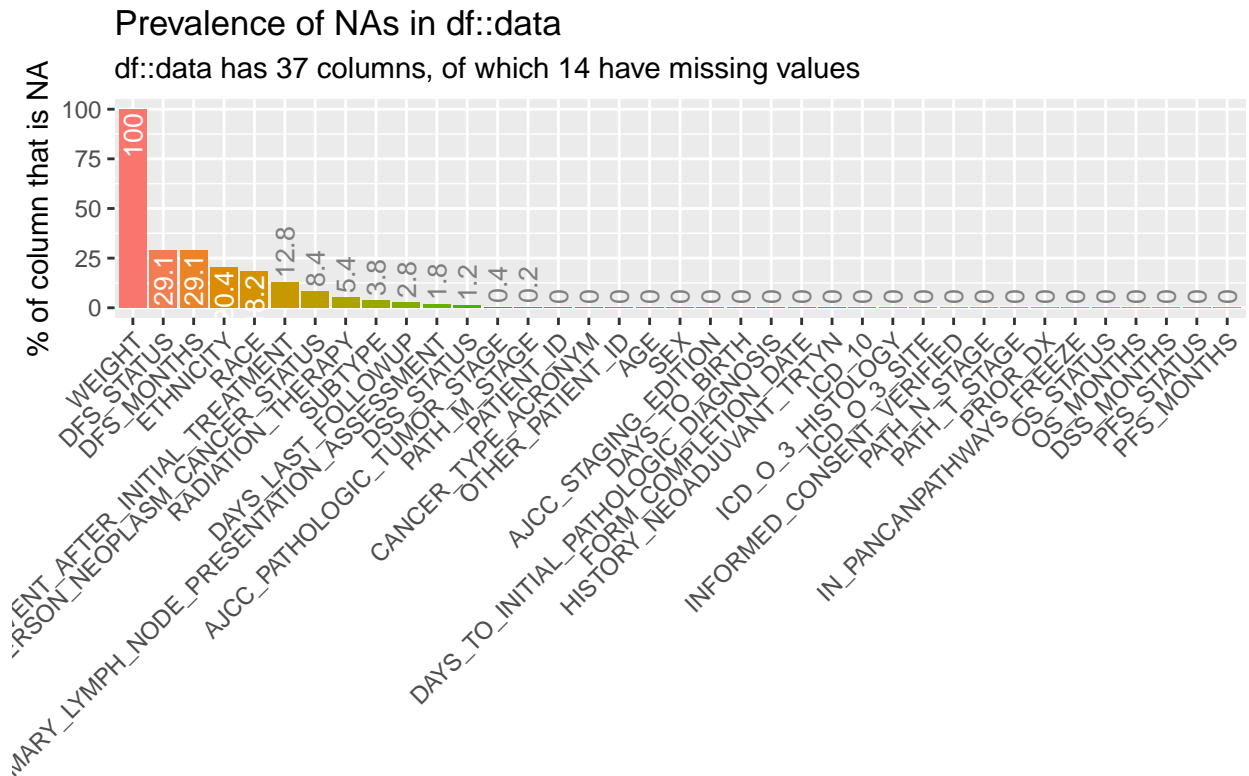

```
missing = inspect_na(data)
missing[ , 3] = round(missing[ ,3], 2)
names(missing) = c("variable", "count", "proportion")
require(kableExtra)
kable(missing)
```

variable	count	proportion
WEIGHT	499	100.00
DFS_STATUS	145	29.06
DFS_MONTHS	145	29.06
ETHNICITY	102	20.44
RACE	91	18.24
NEW_TUMOR_EVENT_AFTER_INITIAL_TREATMENT	64	12.83
PERSON_NEOPLASM_CANCER_STATUS	42	8.42
RADIATION_THERAPY	27	5.41
SUBTYPE	19	3.81
DAYS_LAST_FOLLOWUP	14	2.81
PRIMARY_LYMPH_NODE_PRESENTATION_ASSESSMENT	9	1.80
DSS_STATUS	6	1.20
AJCC_PATHOLOGIC_TUMOR_STAGE	2	0.40
PATH_M_STAGE	1	0.20
PATIENT_ID	0	0.00
CANCER_TYPE_ACRONYM	0	0.00
OTHER_PATIENT_ID	0	0.00
AGE	0	0.00
SEX	0	0.00
AJCC_STAGING_EDITION	0	0.00
DAYS_TO_BIRTH	0	0.00
DAYS_TO_INITIAL_PATHOLOGIC_DIAGNOSIS	0	0.00
FORM_COMPLETION_DATE	0	0.00
HISTORY_NEOADJUVANT_TRTYN	0	0.00
ICD_10	0	0.00
ICD_O_3_HISTOLOGY	0	0.00
ICD_O_3_SITE	0	0.00
INFORMED_CONSENT_VERIFIED	0	0.00
PATH_N_STAGE	0	0.00
PATH_T_STAGE	0	0.00
PRIOR_DX	0	0.00
IN_PANCANPATHWAYS_FREEZE	0	0.00
OS_STATUS	0	0.00
OS_MONTHS	0	0.00
DSS_MONTHS	0	0.00
PFS_STATUS	0	0.00
PFS_MONTHS	0	0.00

### 1.1.1 Inspect distribution of variables

```
explorecolumns_thyroid=c("AGE","SEX","PERSON_NEOPLASM_CANCER_STATUS","ETHNICITY",
                         "RACE","RADIATION_THERAPY", "PFS_STATUS")
dat=data[,explorecolumns_thyroid]

colnames(dat)[colnames(dat)=="PERSON_NEOPLASM_CANCER_STATUS"]<-"neoplasm_status"

dat$PFS_STATUS<-as.integer(ifelse(dat$PFS_STATUS=="0:CENSORED",0,1))
dat$RADIATION_THERAPY = as.integer(ifelse(dat$RADIATION_THERAPY=="No",0,1))
dat$SEX = as.integer(ifelse(dat$SEX=="Female",0,1))
dat$neoplasm_status=as.integer(ifelse(dat$neoplasm_status=="Tumor Free",0,1))

suppressPackageStartupMessages(library(tidyverse))
dat%>%
  pivot_longer(cols=c(PFS_STATUS,RADIATION_THERAPY,AGE, SEX,neoplasm_status),
               names_to ="key", values_to = "value", drop_na(dat)) %>%
  ggplot(aes(value)) +
  geom_histogram(bins = 20) +
  facet_wrap(~key, scales='free_x')
```
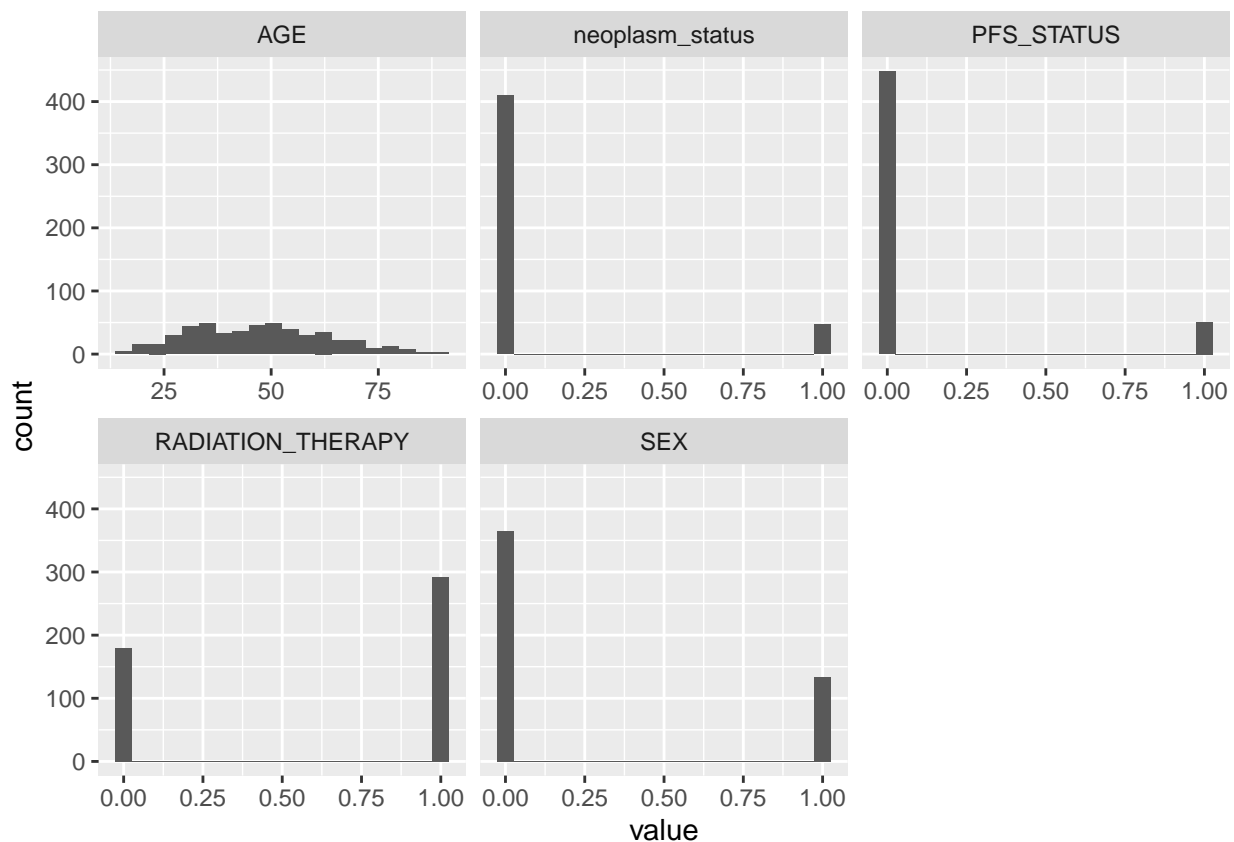
1.1.2 Re-coding variables

```r
data=data[ ,-c(28)] #remove weight,it has empty cells



data$RADIATION_THERAPY = factor(data$RADIATION_THERAPY, levels = c("No","Yes"),
                                labels =c("No", "Yes"))

data$SEX = factor(data$SEX, levels=c("Female","Male"), labels=c("Female","Male"))

data$AJCC_PATHOLOGIC_TUMOR_STAGE=factor(data$AJCC_PATHOLOGIC_TUMOR_STAGE,levels = c("STAGE I","

data$AJCC_STAGING_EDITION = factor(data$AJCC_STAGING_EDITION,
                                   levels = c("4TH","5TH","6TH","7TH"),
                                   labels = c("4TH","5TH","6TH","7TH"))
data$ETHNICITY=factor(data$ETHNICITY,
                      levels=c("Hispanic Or Latino", "Not Hispanic Or Latino"),
                      labels =c("Hispanic Or Latino", "Not Hispanic Or Latino"))

data$PFS_STATUS<-as.integer(ifelse(data$PFS_STATUS=="0:CENSORED",0,1))
```

2   KM Curve - PF Survival of patients with Radiation Therapy

```
library("survival")
library("survminer")
ndata<-data
fit1<-survfit(Surv(ndata$PFS_MONTHS, ndata$PFS_STATUS==1)~ndata$RADIATION_THERAPY
              ,data=ndata)
print(fit1)
```

```
## Call: survfit(formula = Surv(ndata$PFS_MONTHS, ndata$PFS_STATUS ==
##      1) ~ ndata$RADIATION_THERAPY, data = ndata)
##
##     27 observations deleted due to missingness
##                                n events median 0.95LCL 0.95UCL
## ndata$RADIATION_THERAPY=No   180     12     NA      NA      NA
## ndata$RADIATION_THERAPY=Yes  292     38     NA      NA      NA
```

```
summary(fit1)$table
```

```
##                            records n.max n.start events     rmean se(rmean)
## ndata$RADIATION_THERAPY=No      180   180     180     12  162.8716  4.441883
## ndata$RADIATION_THERAPY=Yes     292   292     292     38  147.8427  4.827914
##                            median 0.95LCL 0.95UCL
## ndata$RADIATION_THERAPY=No     NA      NA      NA
## ndata$RADIATION_THERAPY=Yes    NA      NA      NA
```

```
ggsurvplot(fit1,
           #legend.labs=c("tumor_free", "with_tumor"),
           pval = TRUE, conf.int = F,
           risk.table = TRUE, # Add risk table
           risk.table.col = "strata", # Change risk table color by groups
           linetype = "strata", # Change line type by groups
           surv.median.line = "hv", # Specify median survival
           ggtheme = theme_bw(), # Change ggplot2 theme
           palette = c("#E7B800", "#2E9FDF"))
```

```
## Warning in .add_surv_median(p, fit, type = surv.median.line, fun = fun, : Median
## survival not reached.
```
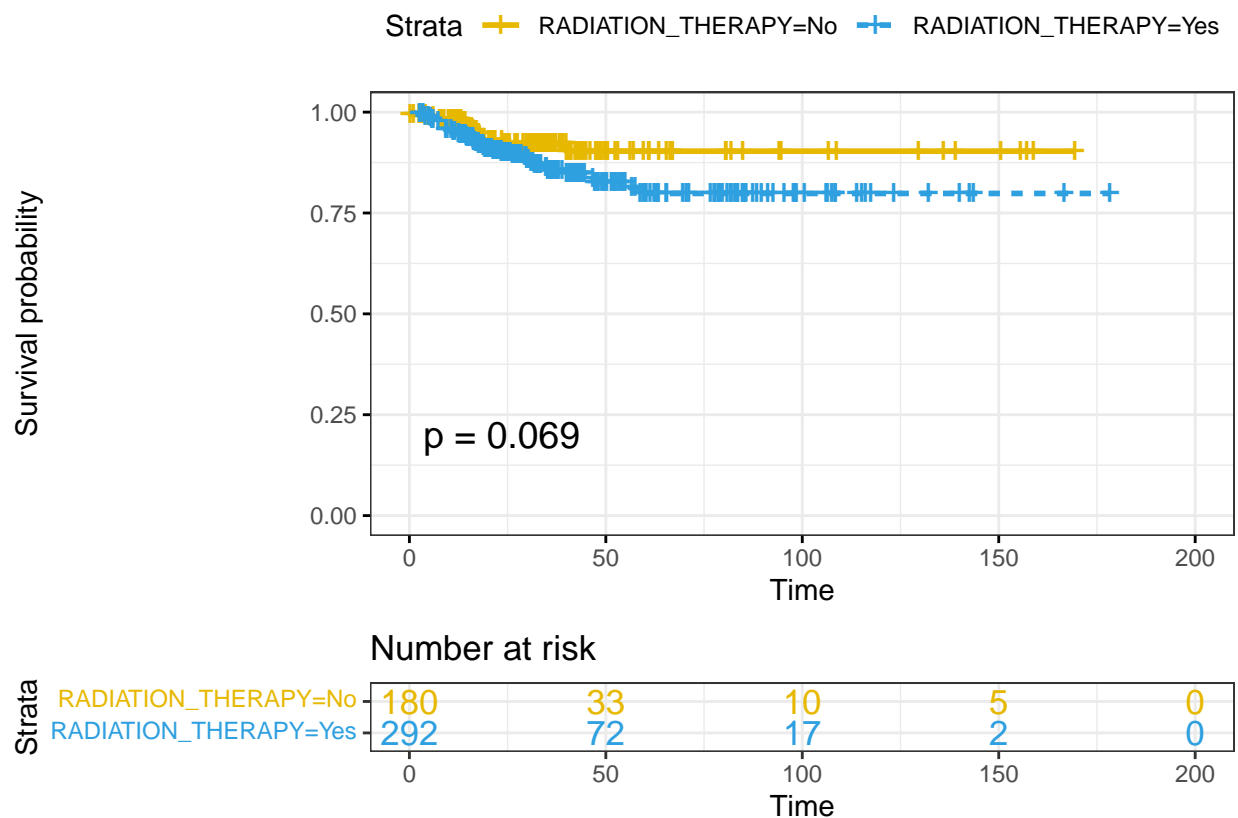
3 PF Survival of Neoplasm Tumor Patients Exposed to Radiation Therapy

```
tumor=ndata[ndata$PERSON_NEOPLASM_CANCER_STATUS=="With Tumor",]

fit2<-survfit(Surv(tumor$PFS_MONTHS, tumor$PFS_STATUS==1)~tumor$RADIATION_THERAPY
             ,data=tumor)

print(fit2)
```
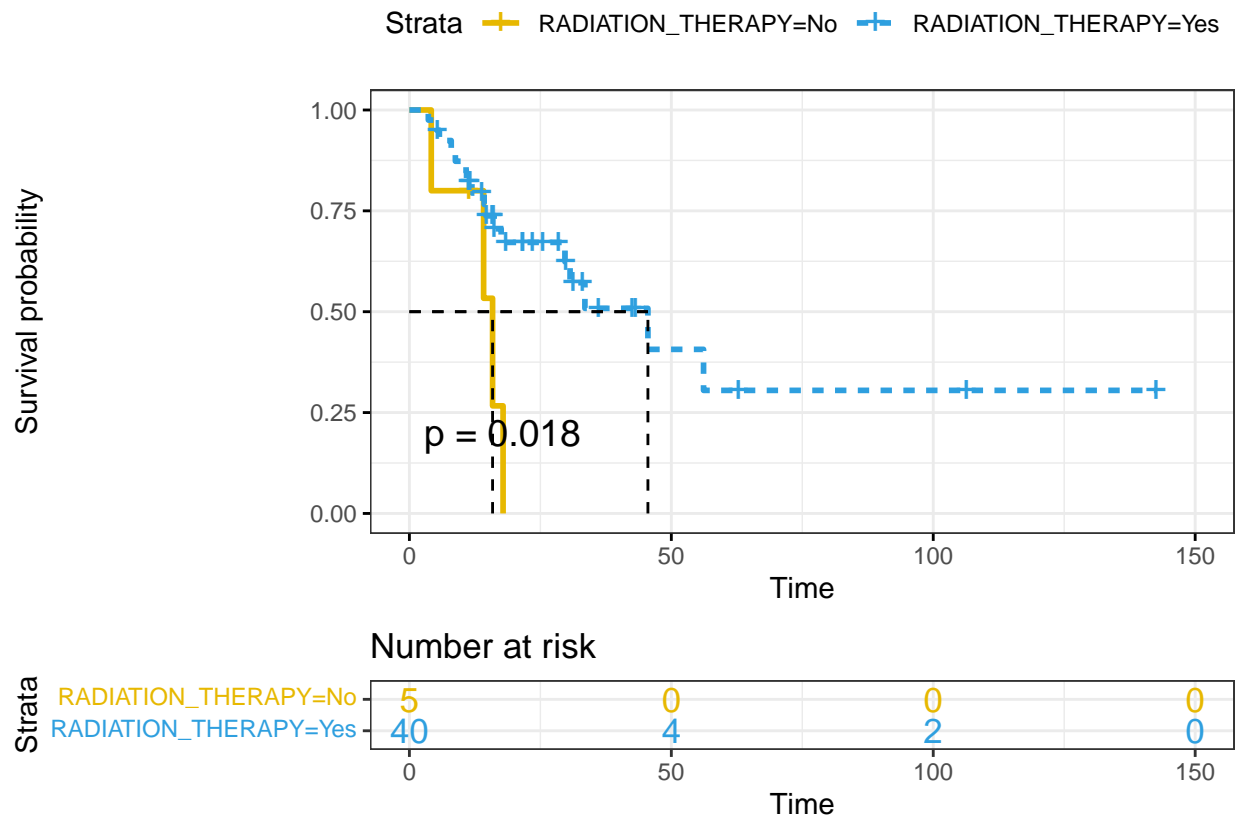
```
## Call: survfit(formula = Surv(tumor$PFS_MONTHS, tumor$PFS_STATUS ==
##      1) ~ tumor$RADIATION_THERAPY, data = tumor)
##
##     44 observations deleted due to missingness
##                              n events median 0.95LCL 0.95UCL
## tumor$RADIATION_THERAPY=No   5      4   15.9    14.2      NA
## tumor$RADIATION_THERAPY=Yes 40     17   45.5    29.7      NA
```

```
summary(fit2)$table
```

```
##                             records n.max n.start events    rmean se(rmean)
## tumor$RADIATION_THERAPY=No        5     5       5      4 13.61739  2.224726
## tumor$RADIATION_THERAPY=Yes      40    40      40     17 62.56331 13.398065
##                               median  0.95LCL 0.95UCL
## tumor$RADIATION_THERAPY=No   15.87928 14.16971      NA
## tumor$RADIATION_THERAPY=Yes 45.53375 29.68735      NA
```

```
ggsurvplot(fit2,
          pval = TRUE, conf.int = F,
          risk.table = TRUE, # Add risk table
          risk.table.col = "strata", # Change risk table color by groups
          linetype = "strata", # Change line type by groups
          surv.median.line = "hv", # Specify median survival
          ggtheme = theme_bw(), # Change ggplot2 theme
          palette = c("#E7B800", "#2E9FDF"))
```

```
table(data$PERSON_NEOPLASM_CANCER_STATUS)
```

```
##
## Tumor Free With Tumor
##        410         47
```

```
table(tumor$PFS_STATUS)
```

```
##
##  0  1
## 25 22
```

```
table(tumor$RADIATION_THERAPY)
```

```
##
##  No Yes
##   5  40
```

4 Logrank Test

```
logrank <- survdiff(Surv(tumor$PFS_MONTHS, tumor$PFS_STATUS==1)~tumor$RADIATION_THERAPY
              ,data=tumor)
logrank
```

```
## Call:
## survdiff(formula = Surv(tumor$PFS_MONTHS, tumor$PFS_STATUS ==
##      1) ~ tumor$RADIATION_THERAPY, data = tumor)
##
## n=45, 44 observations deleted due to missingness.
##
##                                N Observed Expected (O-E)^2/E (O-E)^2/V
## tumor$RADIATION_THERAPY=No     5        4     1.37     5.077      5.59
## tumor$RADIATION_THERAPY=Yes 40       17    19.63     0.353      5.59
##
##   Chisq= 5.6  on 1 degrees of freedom, p= 0.02
```

5 Cox Proportional Hazard Model with Neoplasm Tumor Data

```
fitph<-coxph(Surv(PFS_MONTHS,PFS_STATUS==1) ~ RADIATION_THERAPY +
                AGE + SEX + RACE + AJCC_PATHOLOGIC_TUMOR_STAGE + DAYS_TO_BIRTH
                + DAYS_LAST_FOLLOWUP + AJCC_PATHOLOGIC_TUMOR_STAGE +
                AJCC_STAGING_EDITION,
            data=tumor)
summary(fitph)
```

```
## Call:
## coxph(formula = Surv(PFS_MONTHS, PFS_STATUS == 1) ~ RADIATION_THERAPY +
##      AGE + SEX + RACE + AJCC_PATHOLOGIC_TUMOR_STAGE + DAYS_TO_BIRTH +
##      DAYS_LAST_FOLLOWUP + AJCC_PATHOLOGIC_TUMOR_STAGE + AJCC_STAGING_EDITION,
##      data = tumor)
##
##   n= 33, number of events= 15
##    (56 observations deleted due to missingness)
##
##                                              coef  exp(coef)  se(coef)        z
## RADIATION_THERAPYYes                    -1.904e+00  1.489e-01  6.160e-01   -3.091
## AGE                                     -2.226e+00  1.079e-01  1.630e-02 -136.602
## SEXMale                                  1.892e+00  6.635e+00  6.387e-01    2.963
## RACEAsian                               -3.683e+01  1.010e-16  4.512e+03   -0.008
## RACEBlack or African American           -1.854e+01  8.897e-09  7.510e-01  -24.684
## RACEWhite                               -1.957e+01  3.158e-09  6.828e-01  -28.665
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE II      1.751e+01  4.021e+07  1.056e+00   16.579
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE III     1.698e+01  2.376e+07  5.365e-01   31.658
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IV      0.000e+00  1.000e+00  0.000e+00       NA
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVA     1.533e+01  4.537e+06  1.044e+00   14.684
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVC     1.861e+01  1.211e+08  1.234e+00   15.089
## DAYS_TO_BIRTH                           -6.267e-03  9.938e-01  4.470e-05 -140.224
## DAYS_LAST_FOLLOWUP                       -5.606e-04  9.994e-01  2.623e-04   -2.137
## AJCC_STAGING_EDITION5TH                  -1.144e+00  3.184e-01  8.234e-01   -1.390
## AJCC_STAGING_EDITION6TH                   1.915e+01  2.082e+08  6.831e-01   28.041
## AJCC_STAGING_EDITION7TH                   0.000e+00  1.000e+00  6.110e-01    0.000
##                                          Pr(>|z|)
## RADIATION_THERAPYYes                      0.00199 **
## AGE                                       < 2e-16 ***
## SEXMale                                   0.00305 **
## RACEAsian                                 0.99349
## RACEBlack or African American             < 2e-16 ***
## RACEWhite                                 < 2e-16 ***
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE II       < 2e-16 ***
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE III      < 2e-16 ***
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IV            NA
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVA      < 2e-16 ***
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVC      < 2e-16 ***
```

```
## DAYS_TO_BIRTH                          < 2e-16 ***
## DAYS_LAST_FOLLOWUP                      0.03257 *
## AJCC_STAGING_EDITION5TH                 0.16459
## AJCC_STAGING_EDITION6TH                 < 2e-16 ***
## AJCC_STAGING_EDITION7TH                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                       exp(coef) exp(-coef) lower .95 upper .95
## RADIATION_THERAPYYes                  1.489e-01  6.714e+00 4.453e-02 4.981e-01
## AGE                                   1.079e-01  9.267e+00 1.045e-01 1.114e-01
## SEXMale                               6.635e+00  1.507e-01 1.897e+00 2.320e+01
## RACEAsian                             1.010e-16  9.906e+15 0.000e+00       Inf
## RACEBlack or African American         8.897e-09  1.124e+08 2.042e-09 3.877e-08
## RACEWhite                             3.158e-09  3.166e+08 8.284e-10 1.204e-08
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE II   4.021e+07  2.487e-08 5.074e+06 3.187e+08
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE III  2.376e+07  4.210e-08 8.301e+06 6.798e+07
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IV   1.000e+00  1.000e+00 1.000e+00 1.000e+00
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVA  4.537e+06  2.204e-07 5.865e+05 3.510e+07
## AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVC  1.211e+08  8.256e-09 1.079e+07 1.359e+09
## DAYS_TO_BIRTH                         9.938e-01  1.006e+00 9.937e-01 9.938e-01
## DAYS_LAST_FOLLOWUP                    9.994e-01  1.001e+00 9.989e-01 1.000e+00
## AJCC_STAGING_EDITION5TH               3.184e-01  3.140e+00 6.341e-02 1.599e+00
## AJCC_STAGING_EDITION6TH               2.082e+08  4.803e-09 5.458e+07 7.941e+08
## AJCC_STAGING_EDITION7TH               1.000e+00  1.000e+00 3.020e-01 3.312e+00
##
## Concordance= 0.869  (se = 0.049 )
## Likelihood ratio test= 30.75  on 16 df,   p=0.01
## Wald test            = 22618  on 16 df,   p=<2e-16
## Score (logrank) test = 32.25  on 16 df,   p=0.009
```

6