SURVIVAL ANALYSIS - TCGA PRAD CANCER

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

June 29, 2022

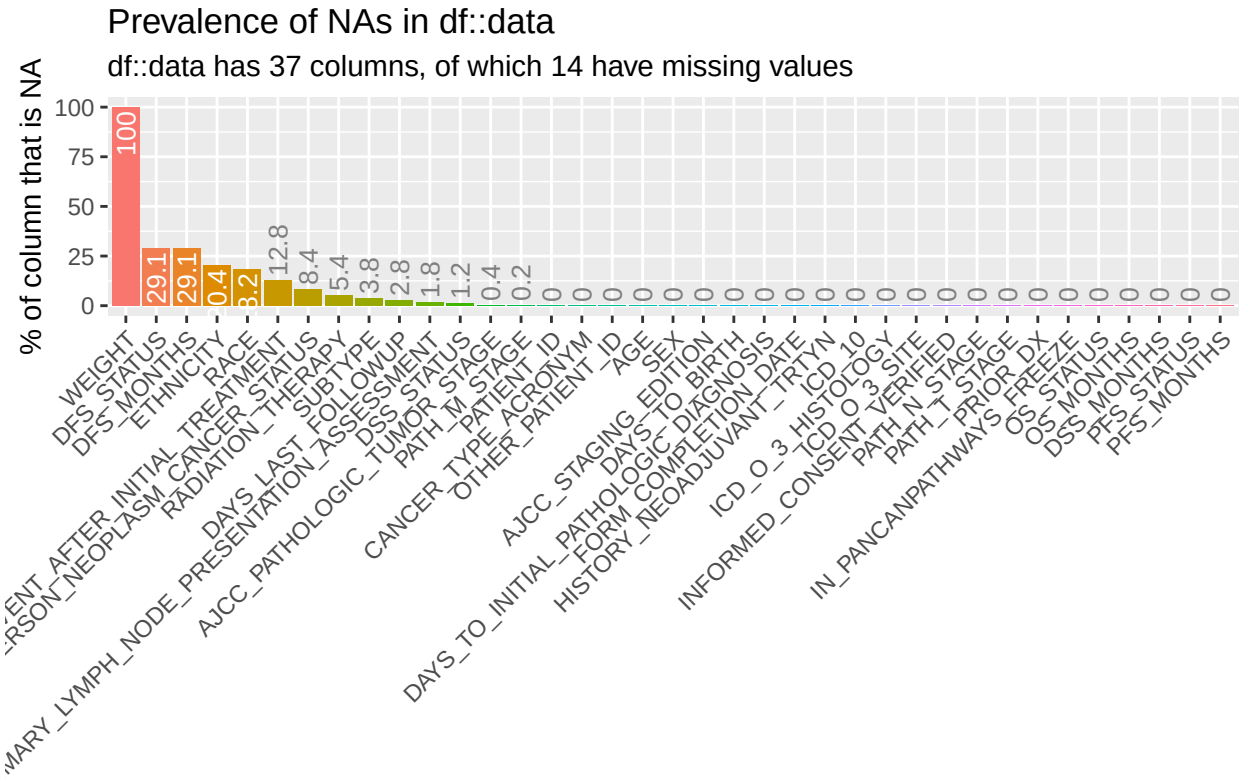
Contents

1 Loading and Cleaning Data

```
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

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

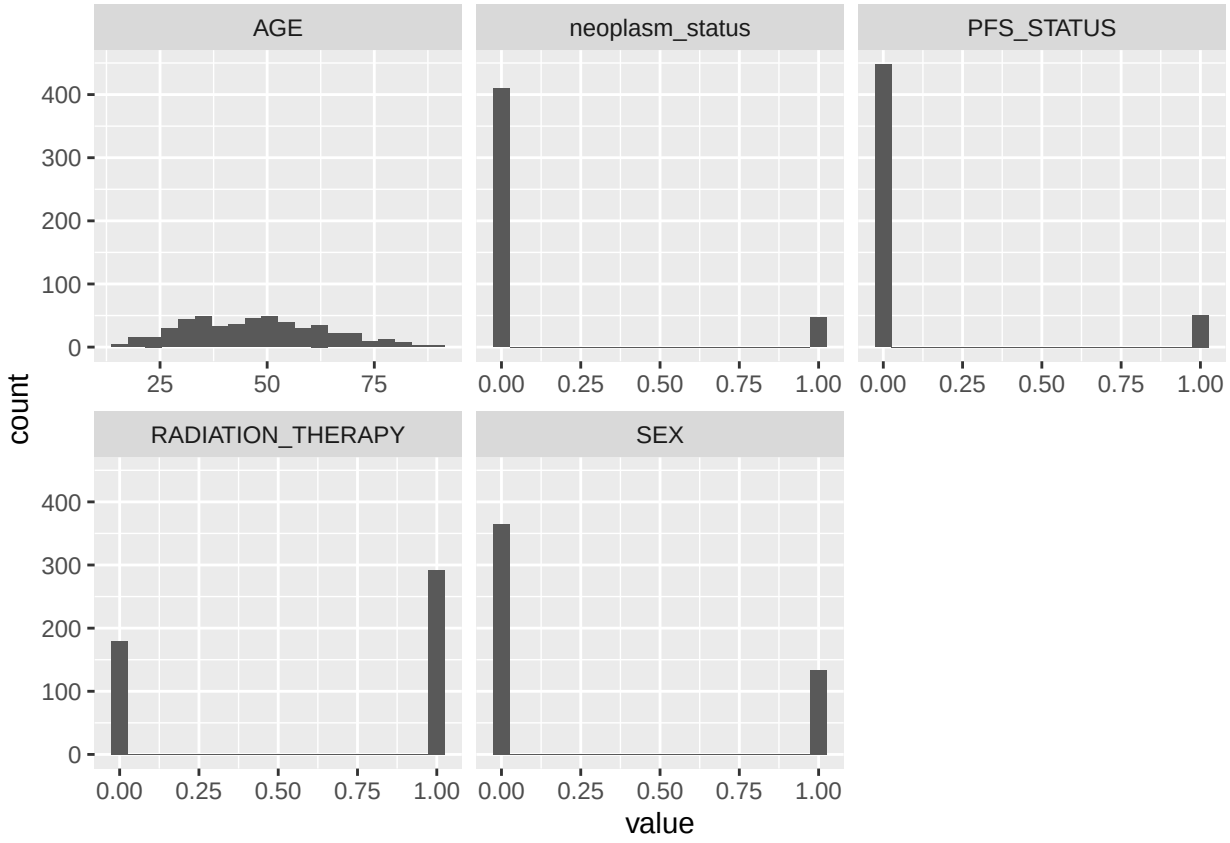



```
missing = inspect_na(data)
missing[ , 3] = round(missing[ ,3], 2)
names(missing) = c("variable", "count", "proportion")
require(kableExtra)
kable(missing)
```

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

```
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

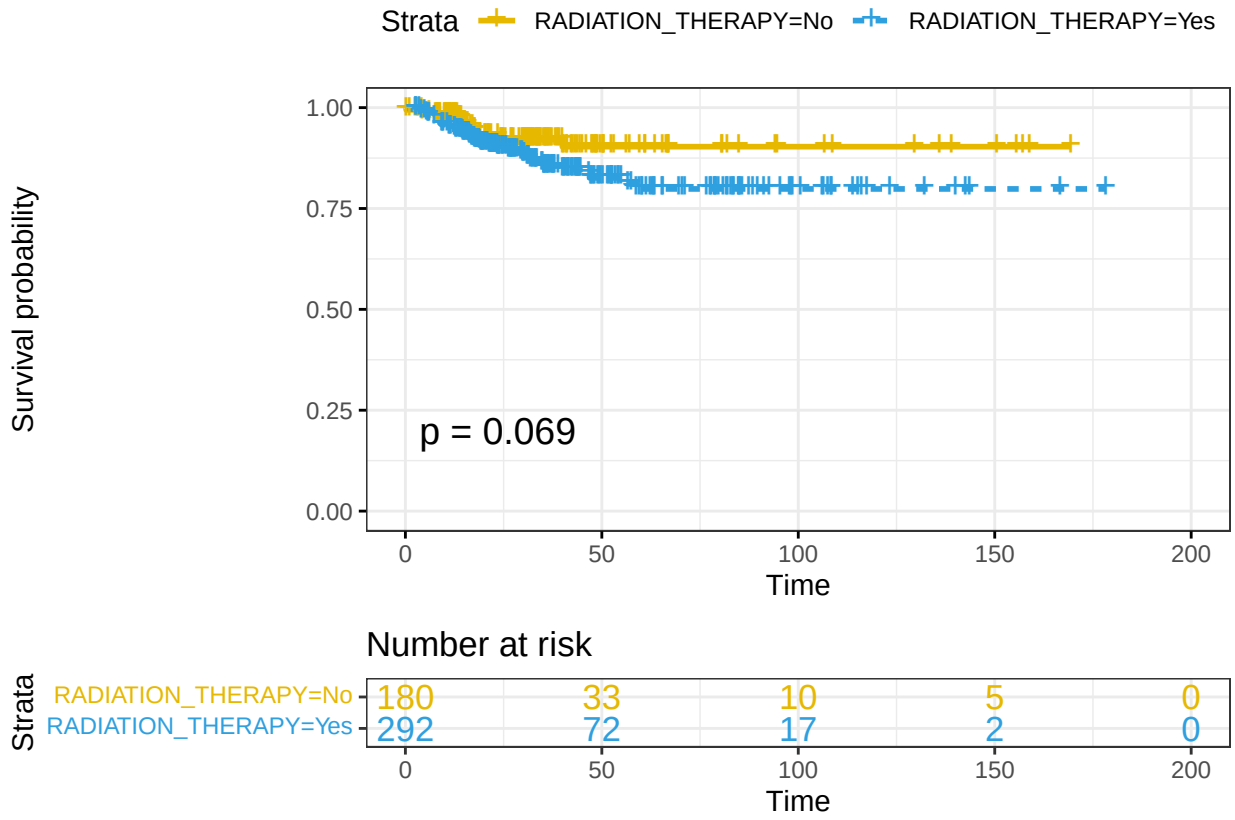
```
library("survival")
library("survminer")
ndata<-data
fit1<-survfit(Surv(ndata$PFS_MONTHS, ndata$PFS_STATUS==1)~ndata$RADIATION_THERAPY
              ,data=ndata)
print(fit1)
```

```
## 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"))
```



3 PF Survival of Neoplasm Tumor Patients Exposed to Radiation Therapy

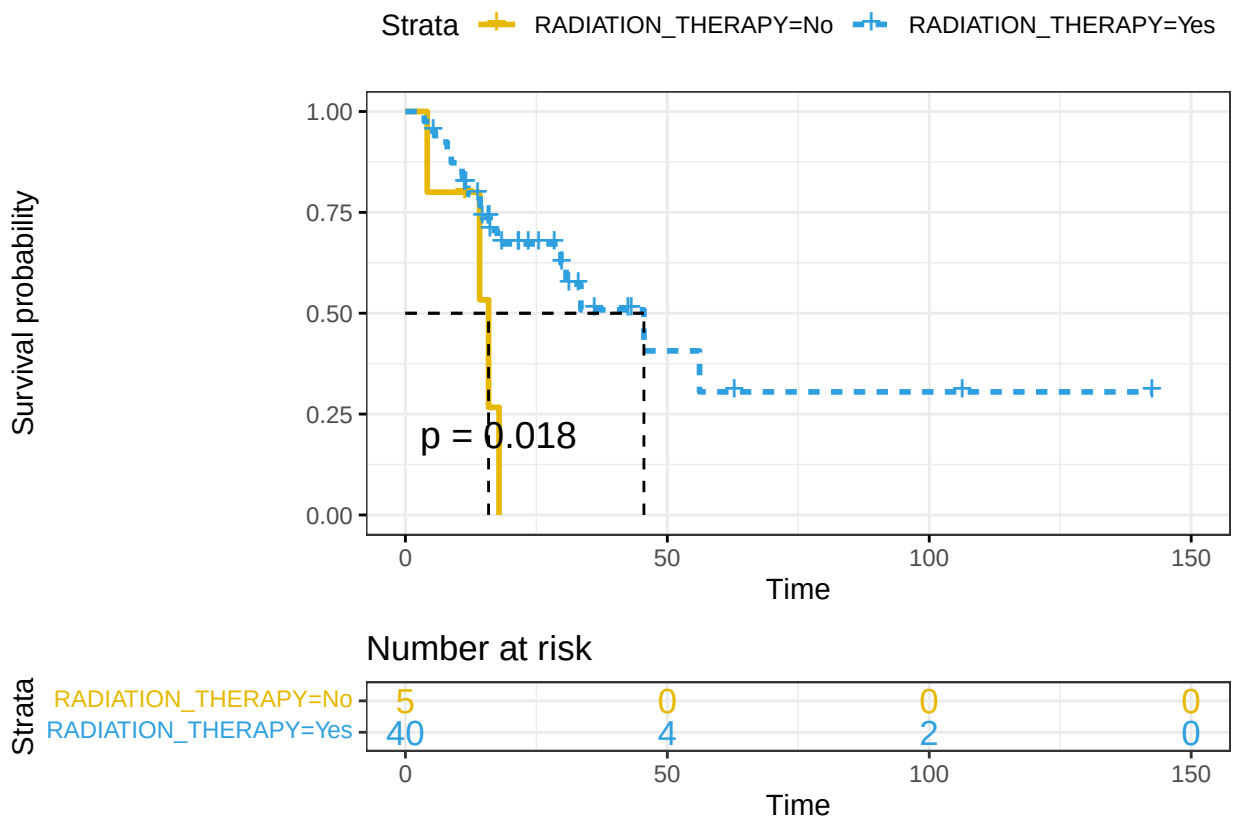
```
tumor=ndata[ndata$PERSON_NEOPLASM_CANCER_STATUS=="With Tumor",]

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)
```

6 Formating Cox Regression Results

```
broom::tidy(fitph ,
            exp=TRUE) %>%
  kable()
```

term	estimate	std.error	statistic	p.v
RADIATION_THERAPYYes	1.489350e-01	0.6160075	-3.0912701	0.0019
AGE	1.079070e-01	0.0162990	-136.6023018	0.0000
SEXMale	6.634732e+00	0.6387226	2.9626605	0.0030
RACEAsian	0.000000e+00	4512.3164391	-0.0081625	0.9934
RACEBlack or African American	0.000000e+00	0.7510061	-24.6836042	0.0000
RACEWhite	0.000000e+00	0.6828295	-28.6648894	0.0000
AJCC_PATHOLOGIC_TUMOR_STAGESTAGE II	4.021354e+07	1.0561633	16.5786051	0.0000
AJCC_PATHOLOGIC_TUMOR_STAGESTAGE III	2.375540e+07	0.5364540	31.6584817	0.0000
AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IV	1.000000e+00	0.0000000	NaN	I
AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVA	4.537365e+06	1.0438310	14.6842319	0.0000
AJCC_PATHOLOGIC_TUMOR_STAGESTAGE IVC	1.211187e+08	1.2335399	15.0885125	0.0000
DAYS_TO_BIRTH	9.937521e-01	0.0000447	-140.2243367	0.0000
DAYS_LAST_FOLLOWUP	9.994395e-01	0.0002623	-2.1372922	0.0325
AJCC_STAGING_EDITION5TH	3.184256e-01	0.8234036	-1.3898001	0.1645
AJCC_STAGING_EDITION6TH	2.081824e+08	0.6830696	28.0409577	0.0000
AJCC_STAGING_EDITION7TH	1.000000e+00	0.6109654	0.0000000	1.0000

```
fitph %>%
  gtsummary::tbl_regression(exp=TRUE)
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**HR**	**95% CI**	**p-value**
RADIATION_THERAPY			
No			
Yes	0.15	0.04, 0.50	0.002
AGE	0.11	0.10, 0.11	<0.001
SEX			
Female			
Male	6.63	1.90, 23.2	0.003
RACE			
American Indian or Alaska Native			
Asian	0.00	0.00, Inf	>0.9
Black or African American	0.00	0.00, 0.00	<0.001
White	0.00	0.00, 0.00	<0.001
AJCC_PATHOLOGIC_TUMOR_STAGE			
STAGE I			
STAGE II	40,213,538	5,074,163, 318,698,566	<0.001
STAGE III	23,755,403	8,301,028, 67,981,841	<0.001
STAGE IV	1.00	1.00, 1.00	
STAGE IVA	4,537,365	586,534, 35,100,584	<0.001
STAGE IVC	121,118,746	10,794,942, 1,358,946,680	<0.001
DAYS_TO_BIRTH	0.99	0.99, 0.99	<0.001
DAYS_LAST_FOLLOWUP	1.00	1.00, 1.00	0.033
AJCC_STAGING_EDITION			
4TH			
5TH	0.32	0.06, 1.60	0.2
6TH	208,182,433	54,577,570, 794,097,750	<0.001
7TH	1.00	0.30, 3.31	>0.9