

```

data=read.csv('METABRIC_RNA_Mutation.csv')

library(ggplot2)
library(survival)
library(survminer)

## Loading required package: ggpubr

##
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##      myeloma

library(corrplot)

## corrplot 0.92 loaded

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union

library(tidyr)
library(patchwork)

data1= data %>%
  select(overall_survival_months,
         overall_survival,
         age_at_diagnosis,
         tumor_size,
         tumor_stage,
         neoplasm_histologic_grade,
         chemotherapy,
         hormone_therapy,
         radio_therapy,
         tp53,
         tp53_mut)%>%
  mutate(tumor_stage = as.factor(tumor_stage),
         neoplasm_histologic_grade = as.factor(neoplasm_histologic_grade),
         chemotherapy = as.factor(chemotherapy),
         hormone_therapy = as.factor(hormone_therapy),
         radio_therapy = as.factor(radio_therapy),
         tp53_mut_bin= if_else(tp53_mut == "0", 0, 1),
         tp53_mut_bin = as.factor(tp53_mut_bin))%>%
  drop_na()

str(data1);dim(data1)

```

```

## 'data.frame': 1354 obs. of 12 variables:
## $ overall_survival_months : num 140.5 84.6 163.7 164.9 41.4 ...
## $ overall_survival : int 1 1 0 1 0 0 1 0 0 0 ...
## $ age_at_diagnosis : num 75.7 43.2 48.9 47.7 77 ...
## $ tumor_size : num 22 10 15 25 40 31 10 29 16 28 ...
## $ tumor_stage : Factor w/ 5 levels "0","1","2","3",...: 3 2 3 3 3 5 3 3 3
3 ...
## $ neoplasm_histologic_grade: Factor w/ 3 levels "1","2","3": 3 3 2 2 3 3 2 2 3 2 ...
## $ chemotherapy : Factor w/ 2 levels "0","1": 1 1 2 2 2 1 2 1 1 1 ...
## $ hormone_therapy : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 1 ...
## $ radio_therapy : Factor w/ 2 levels "0","1": 2 2 1 2 2 2 2 2 2 1 ...
## $ tp53 : num 0.3504 -0.0136 0.5141 1.6708 0.3484 ...
## $ tp53_mut : chr "0" "H178P" "0" "0" ...
## $ tp53_mut_bin : Factor w/ 2 levels "0","1": 1 2 1 1 2 2 1 1 2 1 ...

## [1] 1354 12

p1=ggplot(data1, aes(x = age_at_diagnosis)) +
  geom_histogram(aes(y = ..density..),color="darkblue", fill="lightblue", bins = 30) +
  geom_density(color = "darkblue")+
  labs(title = "Age at Diagnosis", x = "Age")

p2=ggplot(data1, aes(x = tumor_size)) +
  geom_histogram(aes(y = ..density..),colour="black", fill="white", bins = 30) +
  geom_density(color = "black")+
  labs(title = "Tumor Size", x = "Size (mm)")

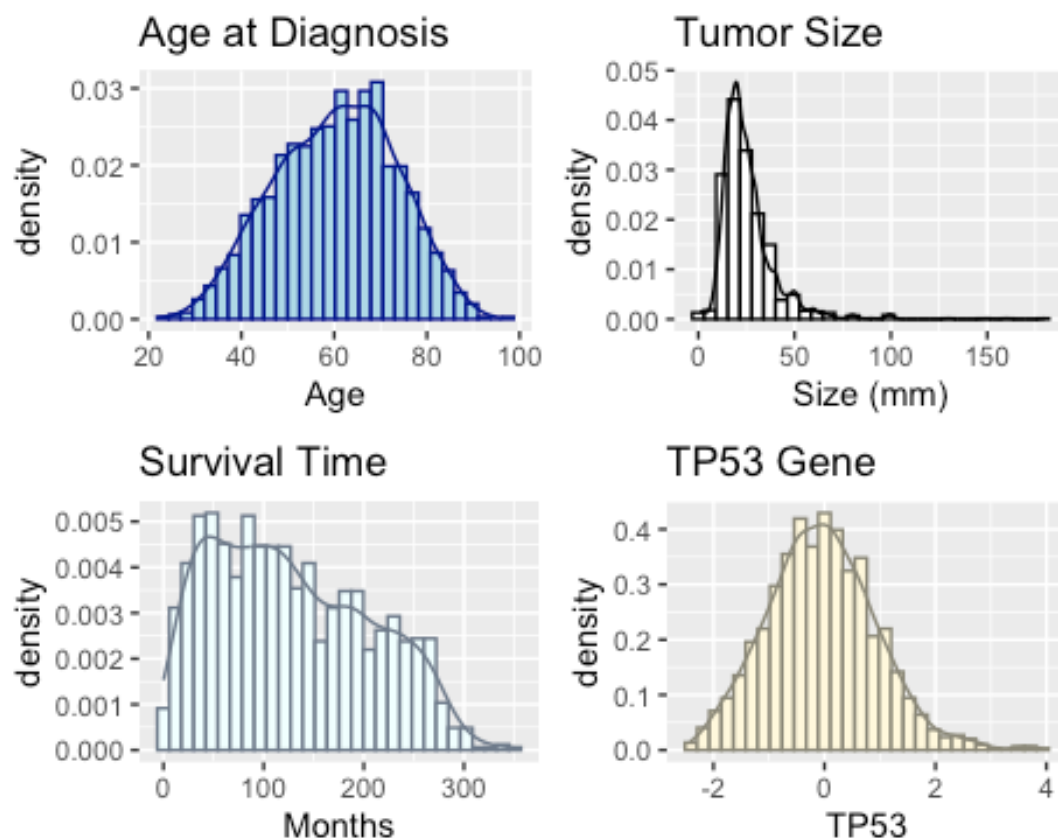
p3=ggplot(data1, aes(x = overall_survival_months)) +
  geom_histogram(aes(y = ..density..),color="lightsteelblue4", fill="azure", bins = 30)
+
  geom_density(color = "lightsteelblue4")+
  labs(title = "Survival Time", x = "Months")

p4=ggplot(data1, aes(x = tp53)) +
  geom_histogram(aes(y = ..density..),color="#8B8878",fill = "#FFF8DC", bins = 30) +
  geom_density(color = "#8B8878")+
  labs(title = "TP53 Gene", x = "TP53")

(p1 | p2 ) /
(p3 | p4 )

## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
##  Please use `after_stat(density)` instead.

```



```
c1=ggplot(data1, aes(x = tumor_stage)) +
  geom_bar(fill='steelblue') +
  labs(title = "Tumor Stage", x = "Stage")

c2=ggplot(data1, aes(x = neoplasm_histologic_grade)) +
  geom_bar(fill='steelblue') +
  labs(title = "Histologic Grade", x = "Grade")

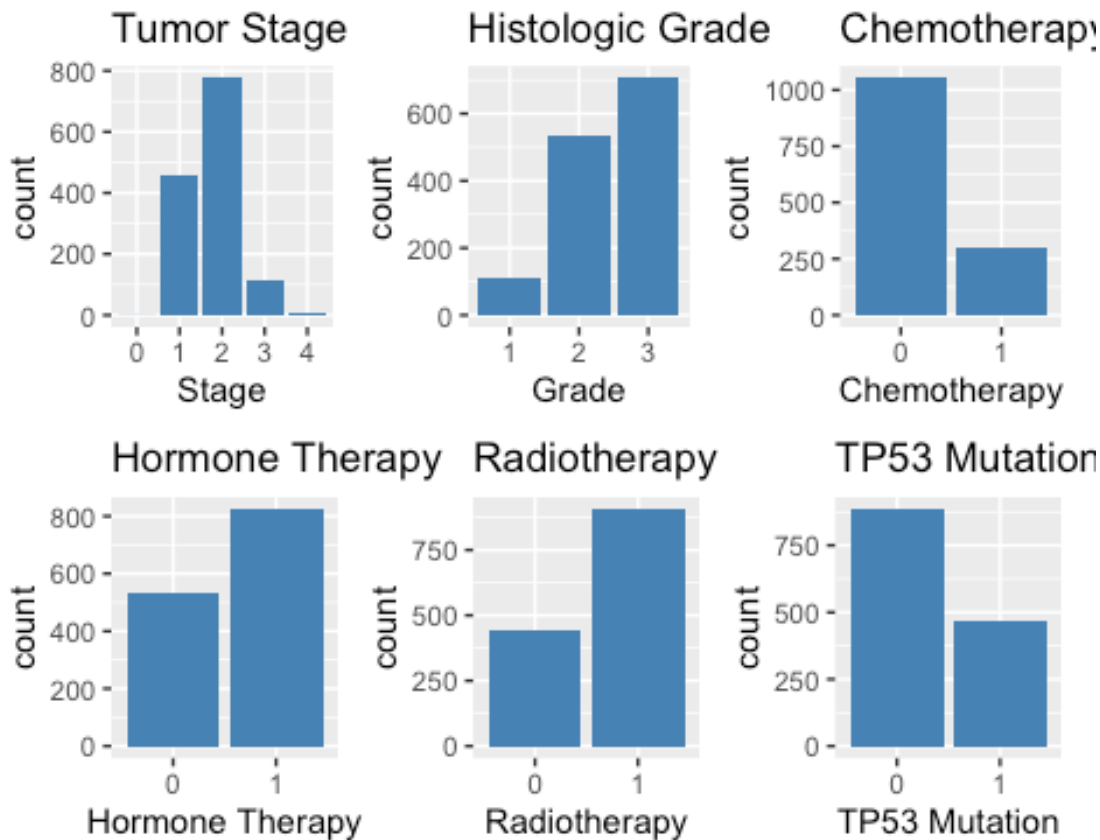
c3=ggplot(data1, aes(x = chemotherapy)) +
  geom_bar(fill='steelblue') +
  labs(title = "Chemotherapy", x = "Chemotherapy")

c4=ggplot(data1, aes(x = hormone_therapy)) +
  geom_bar(fill='steelblue') +
  labs(title = "Hormone Therapy", x = "Hormone Therapy")

c5=ggplot(data1, aes(x = radio_therapy)) +
  geom_bar(fill='steelblue') +
  labs(title = "Radiotherapy", x = "Radiotherapy")

c6=ggplot(data1, aes(x = tp53_mut_bin)) +
  geom_bar(fill='steelblue') +
  labs(title = "TP53 Mutation", x = "TP53 Mutation")

(c1 | c2 | c3) /
(c4 | c5 | c6)
```

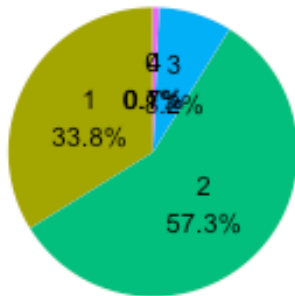


```
make_pie <- function(data1, var, title_text) {
  df <- data1 %>%
    count(!!sym(var)) %>%
    mutate(percent = round(100 * n / sum(n), 1),
           label = paste0(!!sym(var), "\n", percent, "%"))
  ggplot(df, aes(x = "", y = n, fill = !!sym(var))) +
    geom_col(width = 1) +
    coord_polar(theta = "y") +
    geom_text(aes(label = label), position = position_stack(vjust = 0.5), size = 3) +
    labs(title = title_text, x = NULL, y = NULL) +
    theme_void() +
    theme(legend.position = "none")
}

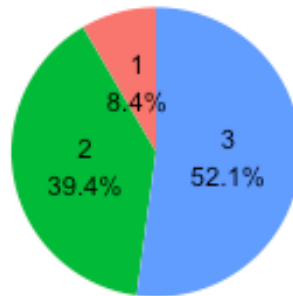
p1 <- make_pie(data1, "tumor_stage", "Tumor Stage")
p2 <- make_pie(data1, "neoplasm_histologic_grade", "Histologic Grade")
p3 <- make_pie(data1, "chemotherapy", "Chemotherapy")
p4 <- make_pie(data1, "hormone_therapy", "Hormone Therapy")
p5 <- make_pie(data1, "radio_therapy", "Radiotherapy")
p6 <- make_pie(data1, "tp53_mut_bin", "TP53 Mutation")

(p1 | p2 | p3) /
(p4 | p5 | p6)
```

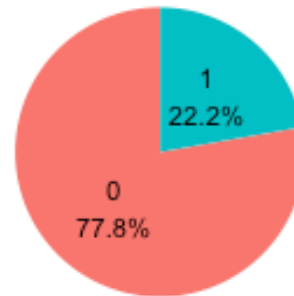
Tumor Stage



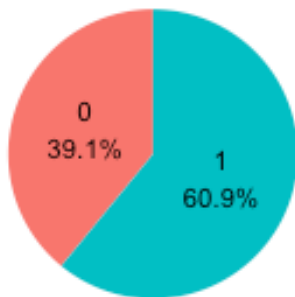
Histologic Grade



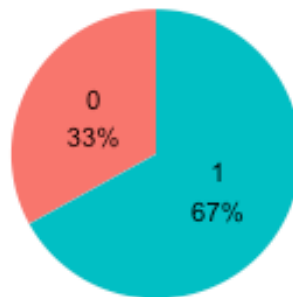
Chemotherapy



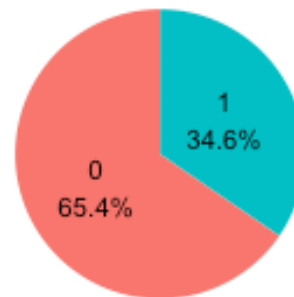
Hormone Therapy



Radiotherapy



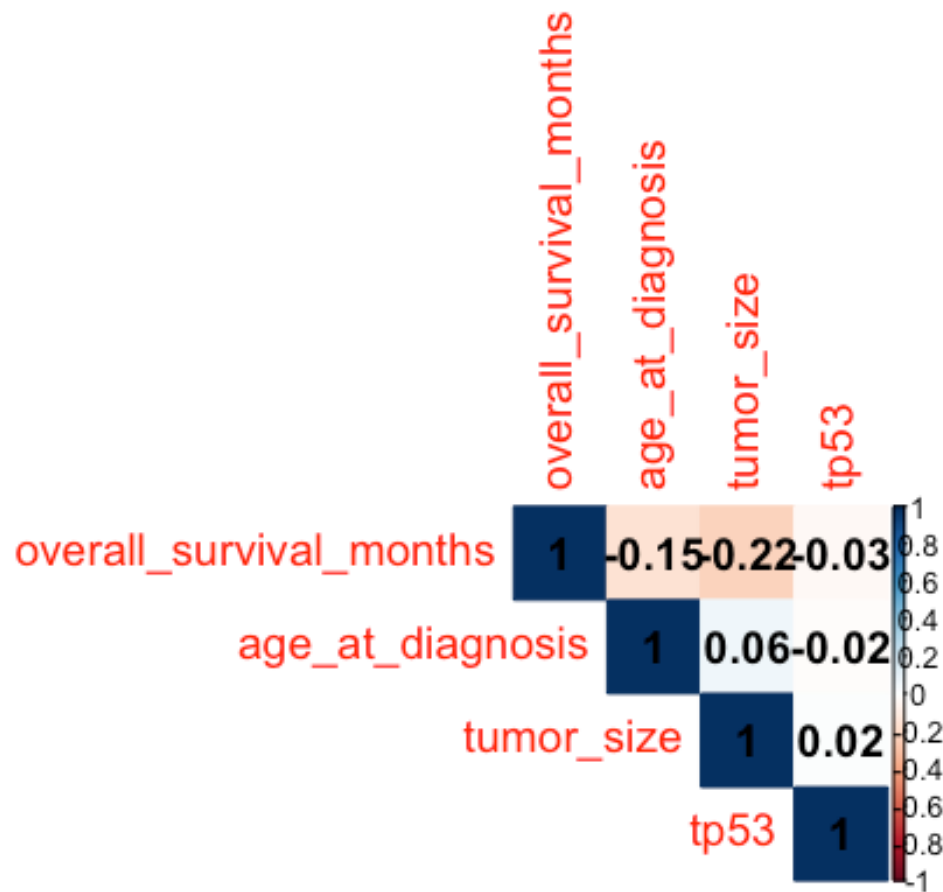
TP53 Mutation



```
cor_data <- data1 %>%
  select(overall_survival_months, age_at_diagnosis, tumor_size, tp53)
cor_matrix <- cor(cor_data, method = "pearson") # 也可以用 method = "spearman"
print(cor_matrix)

##               overall_survival_months age_at_diagnosis tumor_size
## overall_survival_months      1.00000000      -0.15365888 -0.22111171
## age_at_diagnosis             -0.15365888       1.00000000  0.06432339
## tumor_size                   -0.22111171       0.06432339  1.00000000
## tp53                        -0.03405508      -0.01715542  0.01930368
##
##               tp53
## overall_survival_months -0.03405508
## age_at_diagnosis        -0.01715542
## tumor_size              0.01930368
## tp53                    1.00000000

corrplot(cor_matrix, method = "color", type = "upper",
  addCoef.col = "black", tl.cex = 1.2, number.cex = 1.2)
```



```
surv_obj <- Surv(data1$overall_survival_months, data1$overall_survival)
fit <- survfit(surv_obj ~ tp53_mut_bin, data = data1)
ggsurvplot(fit, data = data1, pval = TRUE, conf.int = TRUE)
```

