5231 pj V3

```r
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
```

```r
library(tidyr)
library(survival)
library(survminer)
```

```
Loading required package: ggplot2
```

```
Loading required package: ggpubr
```

```

Attaching package: 'survminer'

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

    myeloma
```

```r
library(ggplot2)
library(corrplot)
```

```
corrplot 0.95 loaded
```

```r
library(glmnet)
```

```
Loading required package: Matrix
```

```

Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

    expand, pack, unpack

Loaded glmnet 4.1-8
```

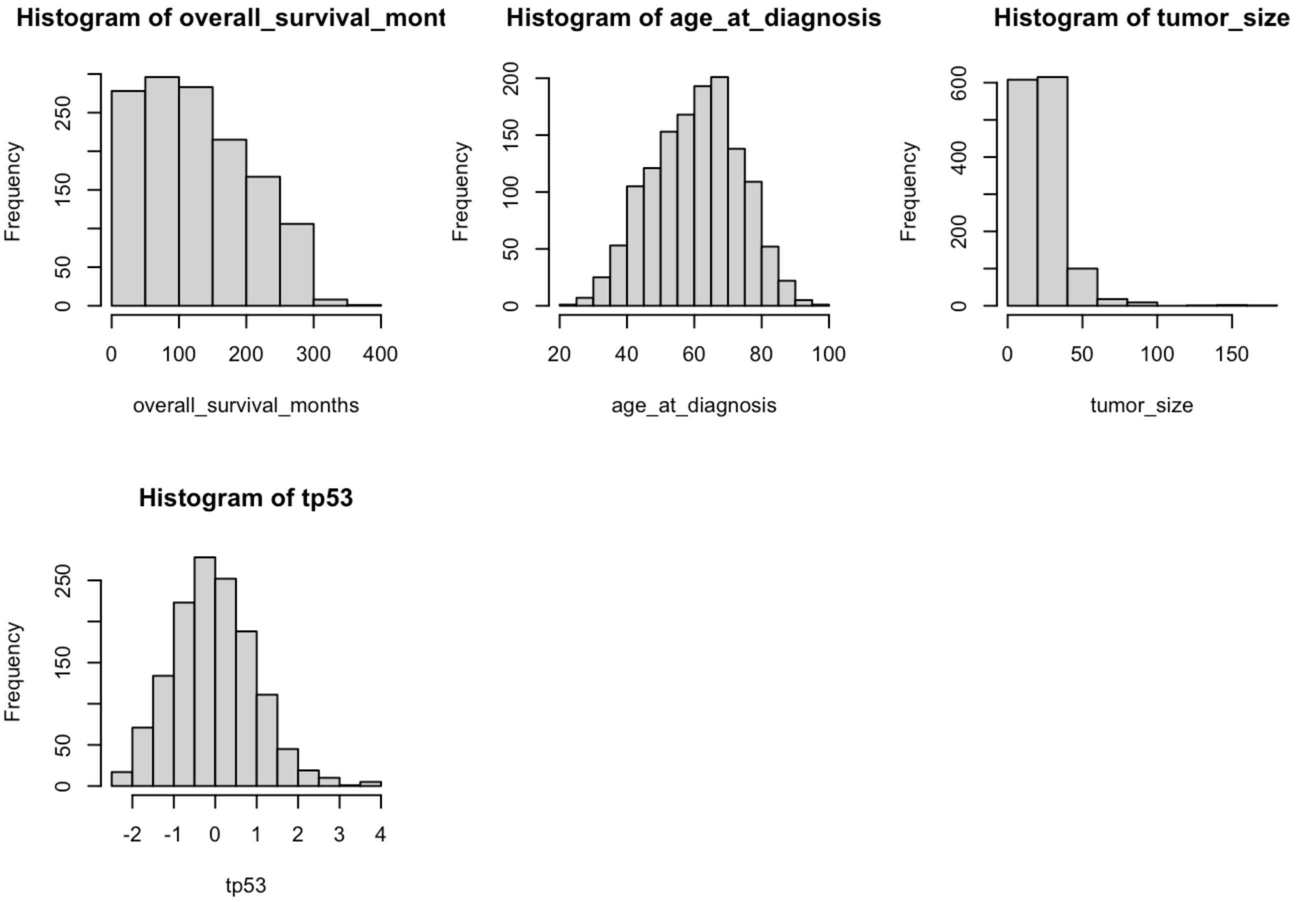
```r
library(flexsurv)
library(patchwork)
```

```r
df <- read.csv('~/Downloads/METABRIC_RNA_Mutation.csv')
```

```r
#data selection: we choose clinical data and the gene we interested:tp53
#tp53: This gene is consider to have a tumor suppressor effect
df$tp53_mut_bin <- ifelse(df$tp53_mut=='0', 0, 1)
df_model<-df %>%
  select(overall_survival_months,
         overall_survival,
         age_at_diagnosis,
         tumor_size,
         tumor_stage,
         neoplasm_histologic_grade,
         chemotherapy,
         hormone_therapy,
         radio_therapy,
         er_status,
         pr_status,
         her2_status,
         tp53_mut_bin,
         type_of_breast_surgery,
         tp53)%>%
  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 = as.factor(tp53_mut_bin))%>%
  drop_na()
```

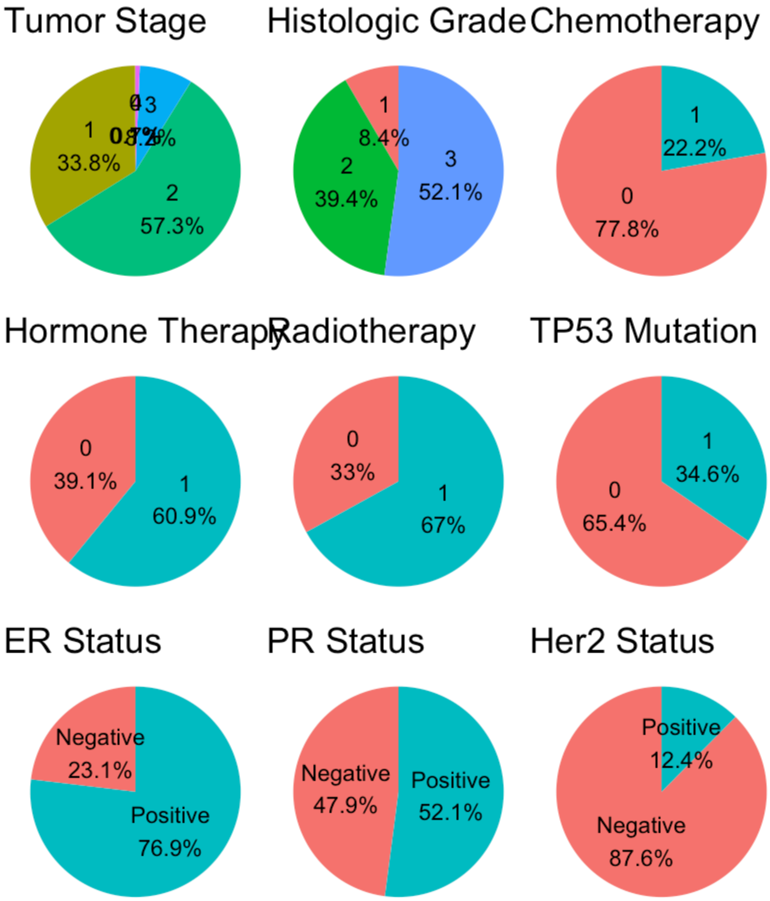
```r
df_model <- df_model %>%
  mutate(tumor_stage_grp = case_when(
    tumor_stage %in% c(0, 1) ~ "Early",
    tumor_stage == 2 ~ "Intermediate",
    tumor_stage %in% c(3, 4) ~ "Late"
  )) %>%
  mutate(tumor_stage_grp = as.factor(tumor_stage_grp))
df_model$age_group <- ifelse(df_model$age_at_diagnosis <= 60, "≤ 60", "> 60")
df_model$size_group <- ifelse(df_model$tumor_size <= 20, "≤ 20mm", "> 20mm")
```

```r
numeric_df <- df_model %>% select(where(is.numeric))
numeric_vars <- names(numeric_df)

par(mfrow = c(2, 3))

for (col in numeric_vars) {
  if (col == "overall_survival") next
  hist(df_model[[col]],
       main = paste("Histogram of", col),
       xlab = col)
}
```

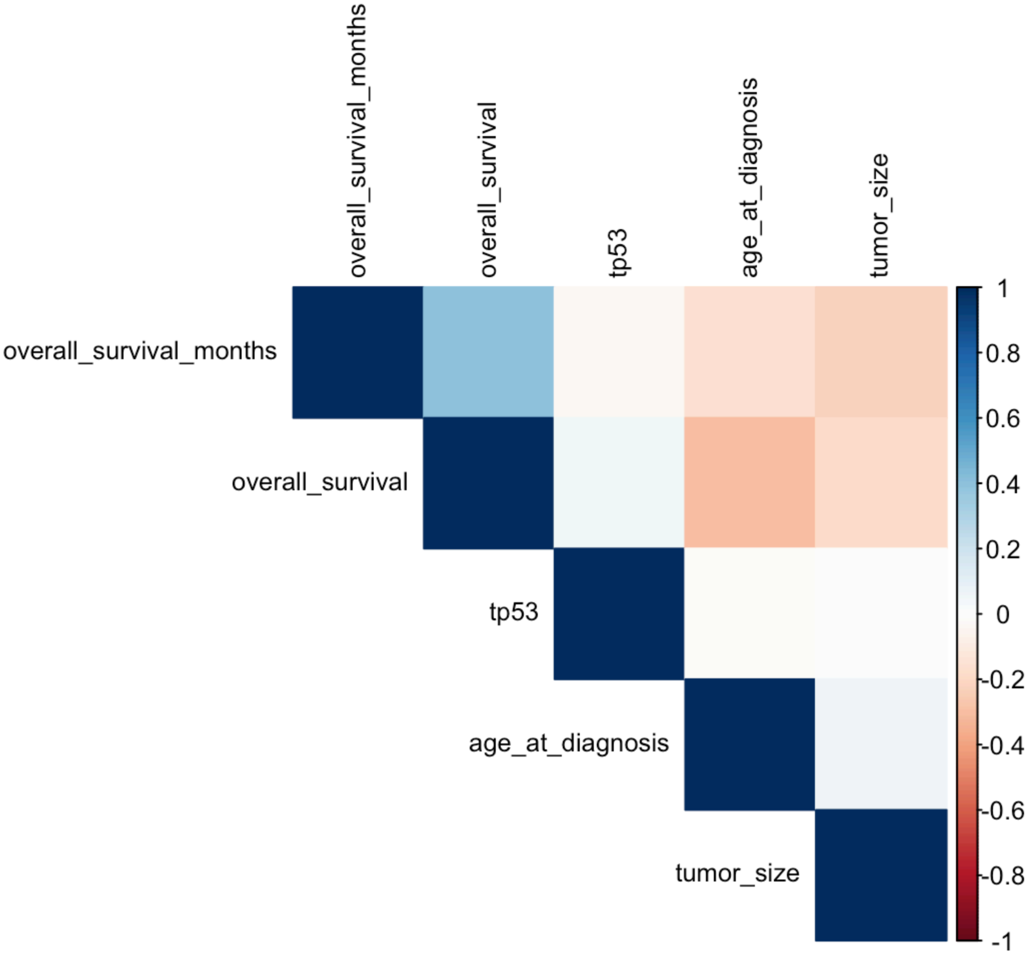


```
make_pie <- function(data1, var, title_text) {
  df <- data1 %>%
    count(!!sym(var)) %>%
    mutate(percent = round(100 * n / sum(n), 1),
           label = paste0(!!sym(var), "\n", percent, "%"))
  ggplot(df, aes(x = "", y = n, fill = !!sym(var))) +
    geom_col(width = 1) +
    coord_polar(theta = "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(df_model, "tumor_stage", "Tumor Stage")
p2 <- make_pie(df_model, "neoplasm_histologic_grade", "Histologic Grade")
p3 <- make_pie(df_model, "chemotherapy", "Chemotherapy")
p4 <- make_pie(df_model, "hormone_therapy", "Hormone Therapy")
```

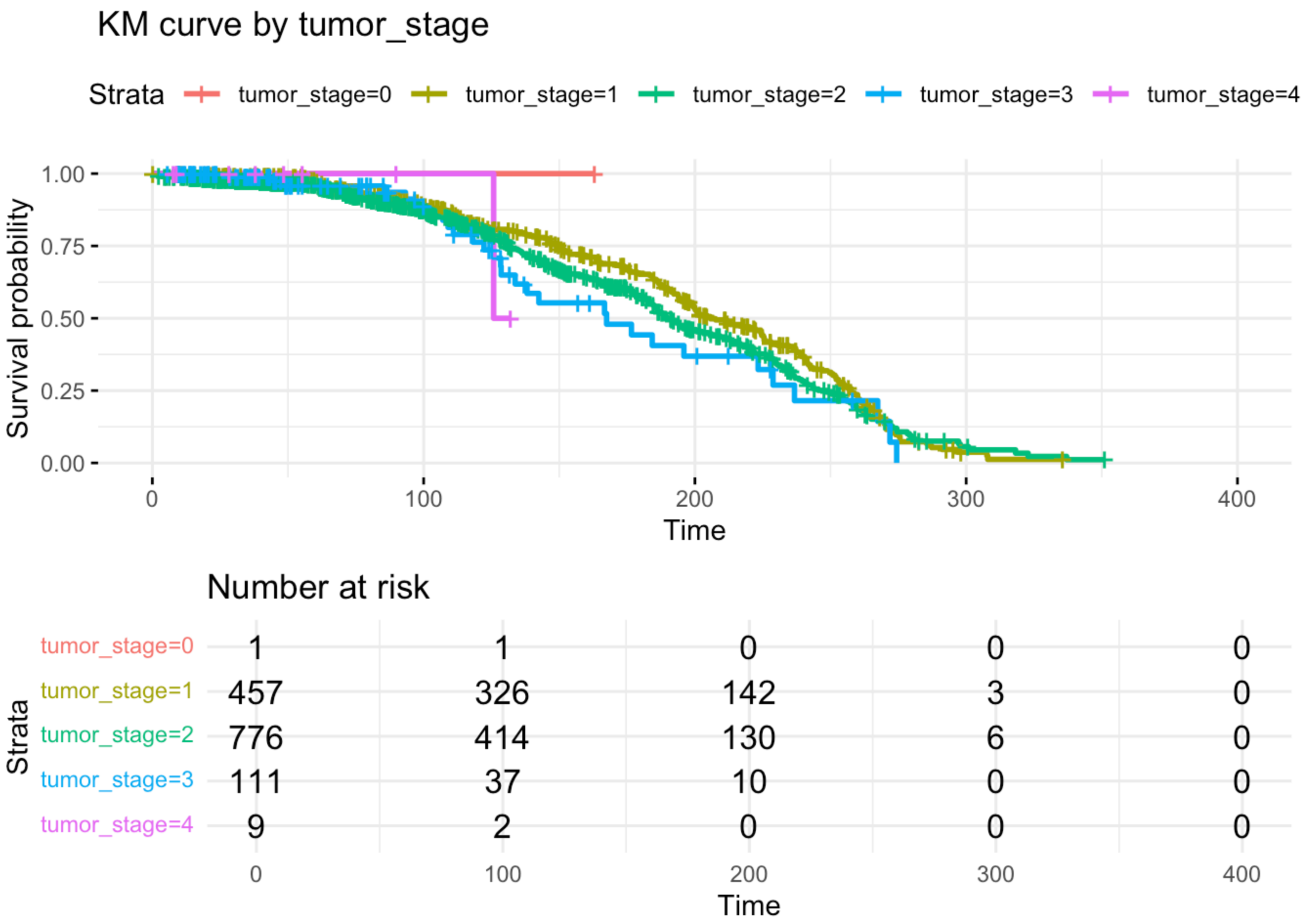
```
p5 <- make_pie(df_model, "radio_therapy", "Radiotherapy")
p6 <- make_pie(df_model, "tp53_mut_bin", "TP53 Mutation")
p7 <- make_pie(df_model, "er_status", "ER Status")
p8 <- make_pie(df_model, "pr_status", "PR Status")
p9 <- make_pie(df_model, "her2_status", "Her2 Status")

(p1 | p2 | p3) /
(p4 | p5 | p6) /
(p7 | p8 | p9)
```



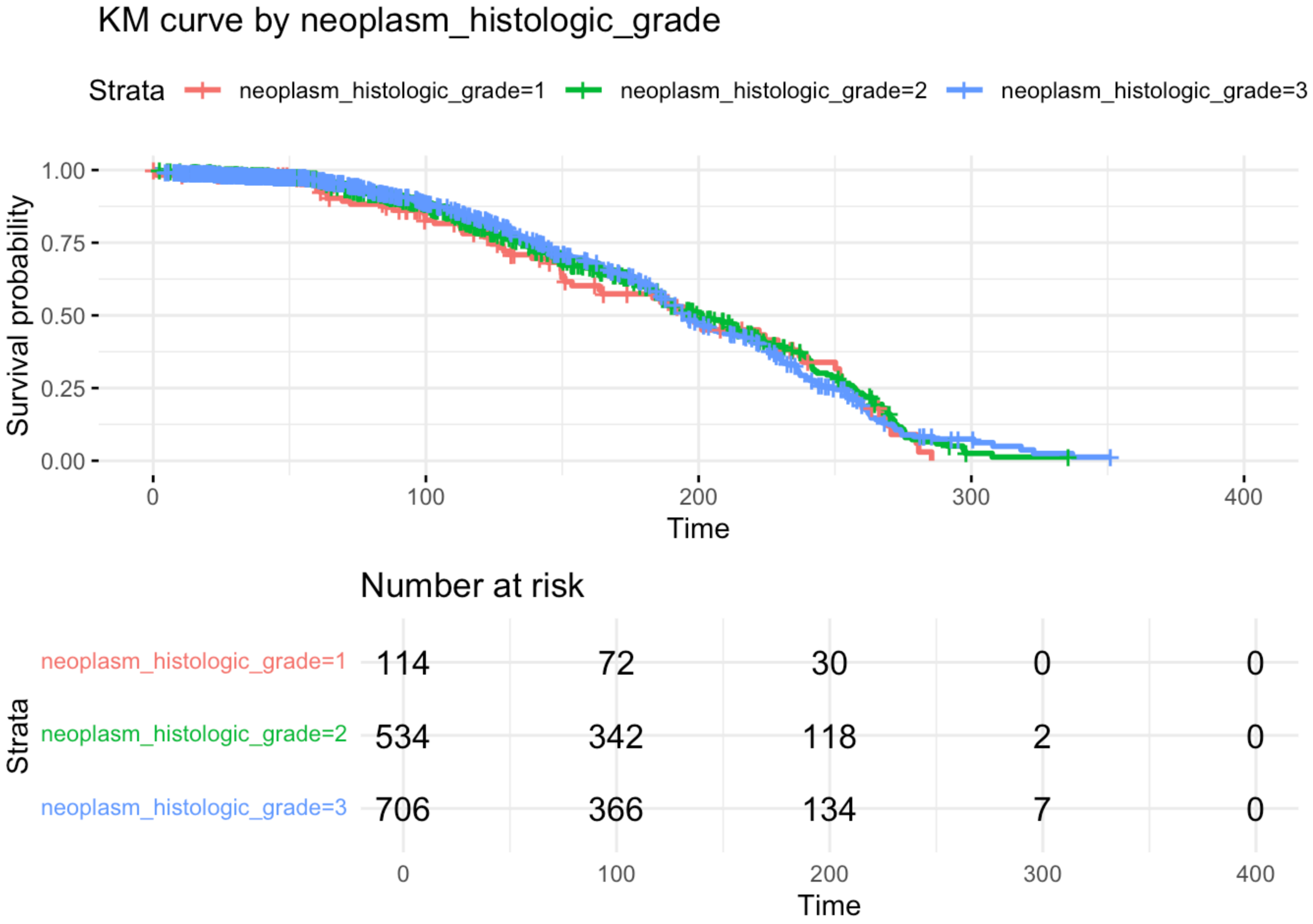
```
cor_mat <- cor(numeric_df, use = "complete.obs")
corrplot(cor_mat, method = "color", type = "upper",
         tl.cex = 0.8, tl.col = "black", order = "hclust")
```



```r
#KM plot
#looking for group feature,I assume the unique length less than 3 is also group not numer
group_vars <- names(df_model)[sapply(df_model, function(x) {
  is.factor(x) || is.character(x) || (is.numeric(x) && length(unique(x)) <= 3)
})]
group_vars <- group_vars[group_vars != "overall_survival"]
for (var in group_vars) {
  formula <- as.formula(paste("Surv(overall_survival_months, overall_survival) ~", var))
  model <- survfit(formula, data = df_model)

  model$call <- list(formula = formula)

  print(ggsurvplot(model, data = df_model,
           risk.table = TRUE,
           risk.table.height = 0.4,
           title = paste("KM curve by", var),
           ggtheme = theme_minimal()))
}
```



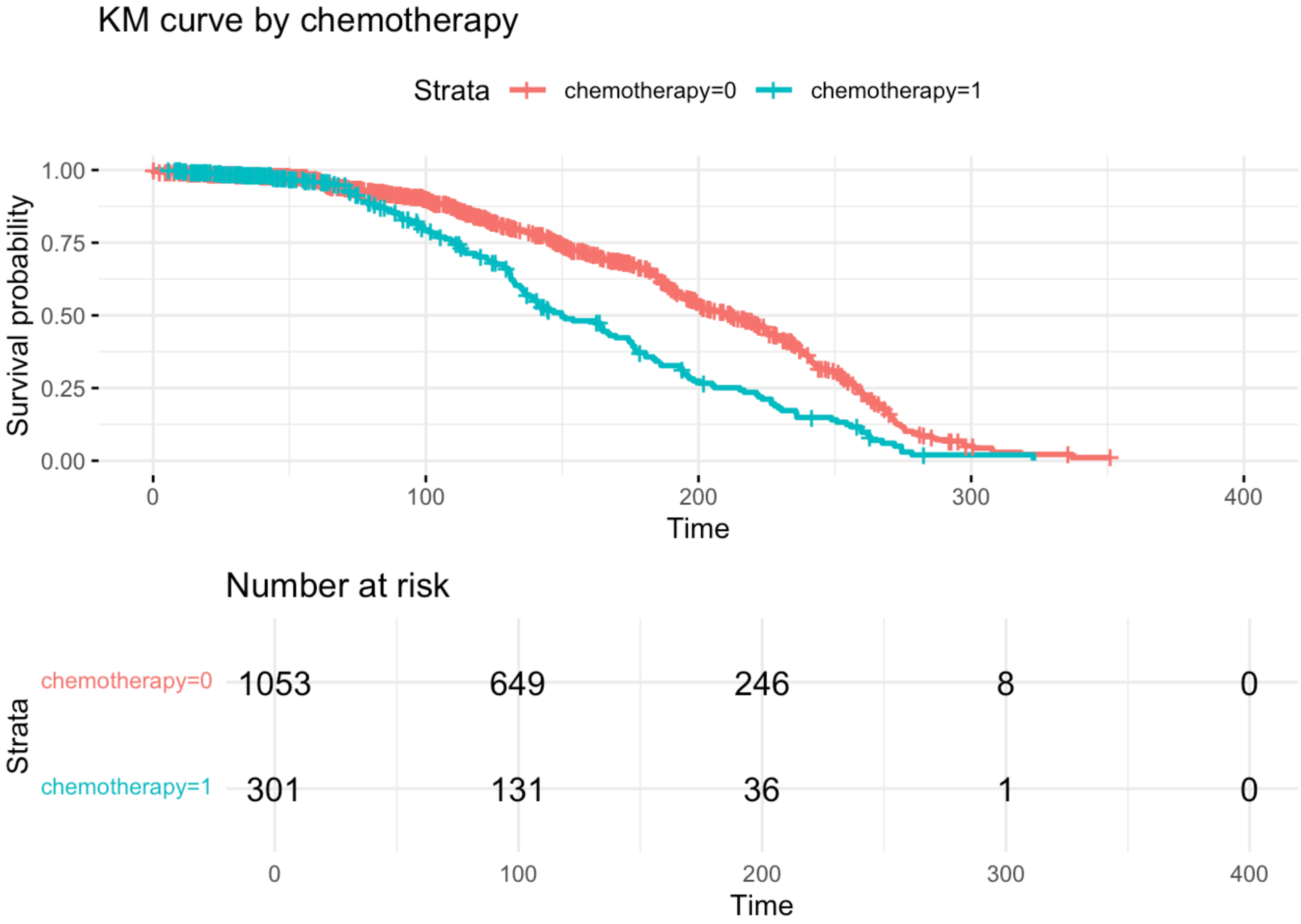
Strata	0	100	200	300	400
tumor_stage=0	1	1	0	0	0
tumor_stage=1	457	326	142	3	0
tumor_stage=2	776	414	130	6	0
tumor_stage=3	111	37	10	0	0
tumor_stage=4	9	2	0	0	0



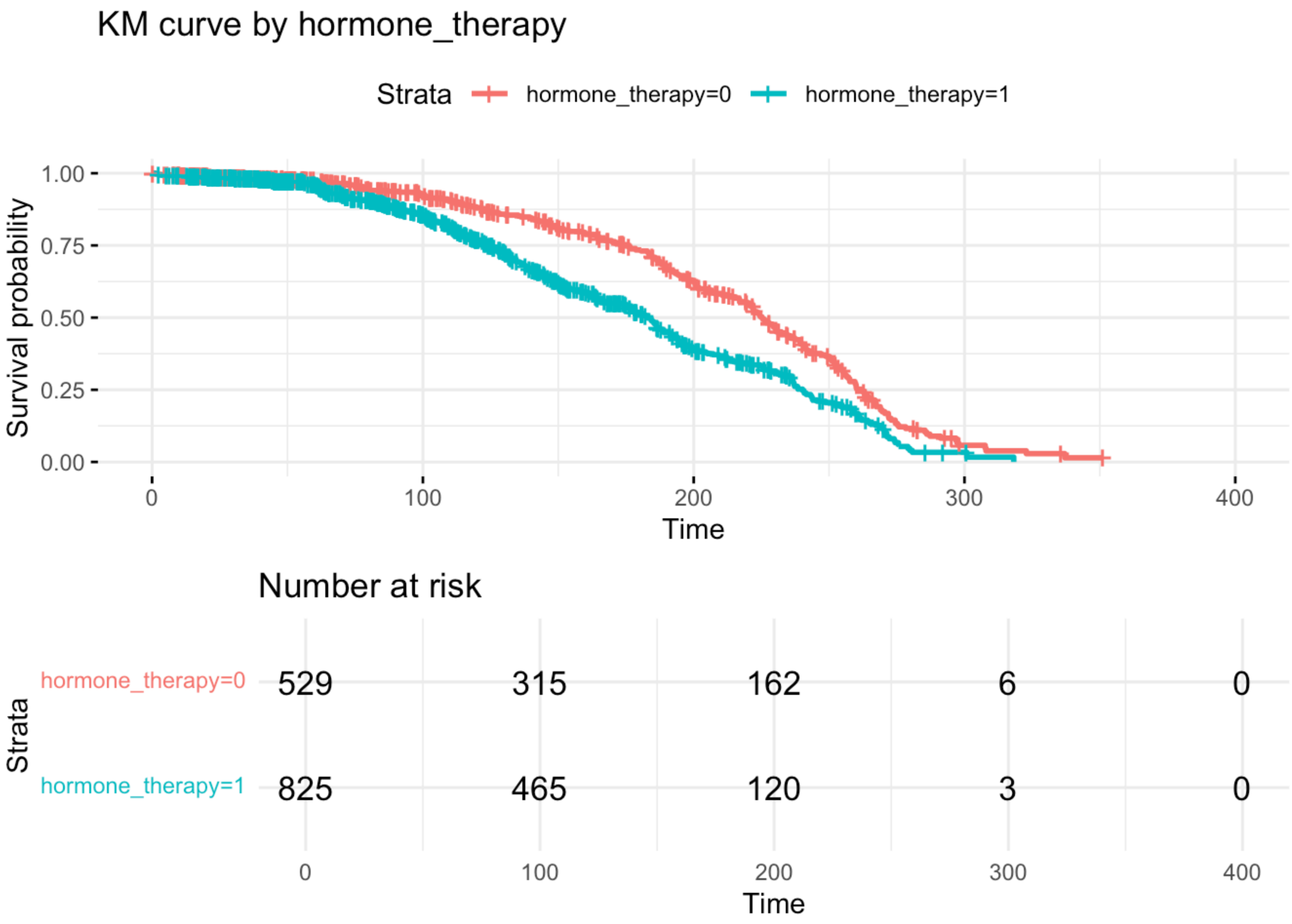
KM curve by neoplasm_histologic_grade

Number at risk

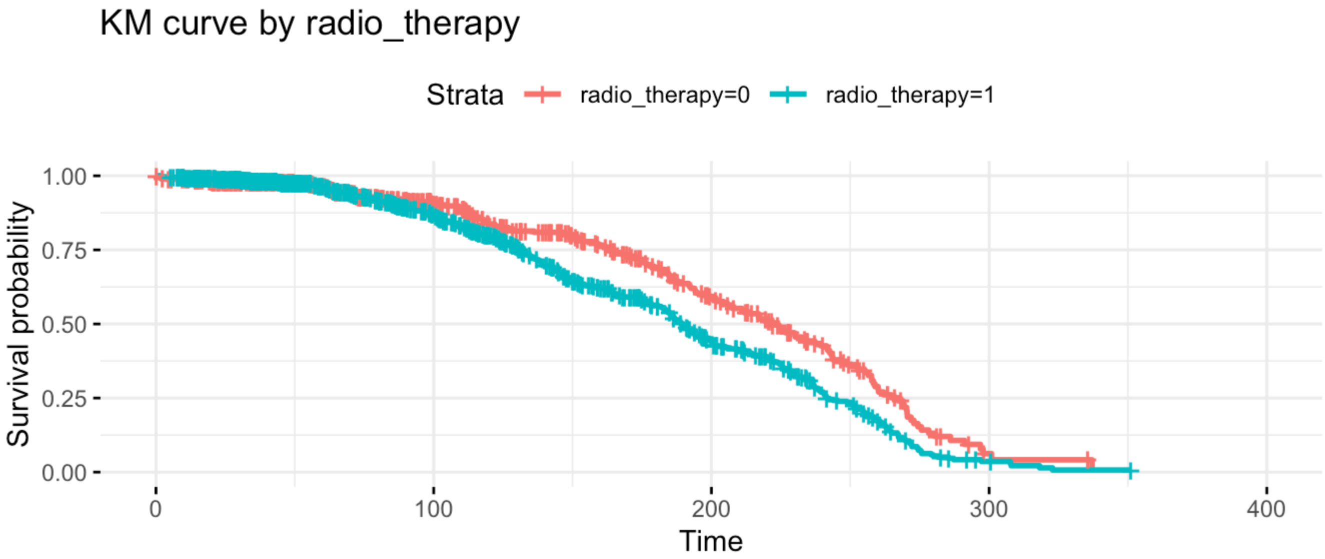
Strata	0	100	200	300	400
neoplasm_histologic_grade=1	114	72	30	0	0
neoplasm_histologic_grade=2	534	342	118	2	0
neoplasm_histologic_grade=3	706	366	134	7	0



Strata	0	100	200	300	400
chemotherapy=0	1053	649	246	8	0
chemotherapy=1	301	131	36	1	0



Strata	0	100	200	300	400
hormone_therapy=0	529	315	162	6	0
hormone_therapy=1	825	465	120	3	0

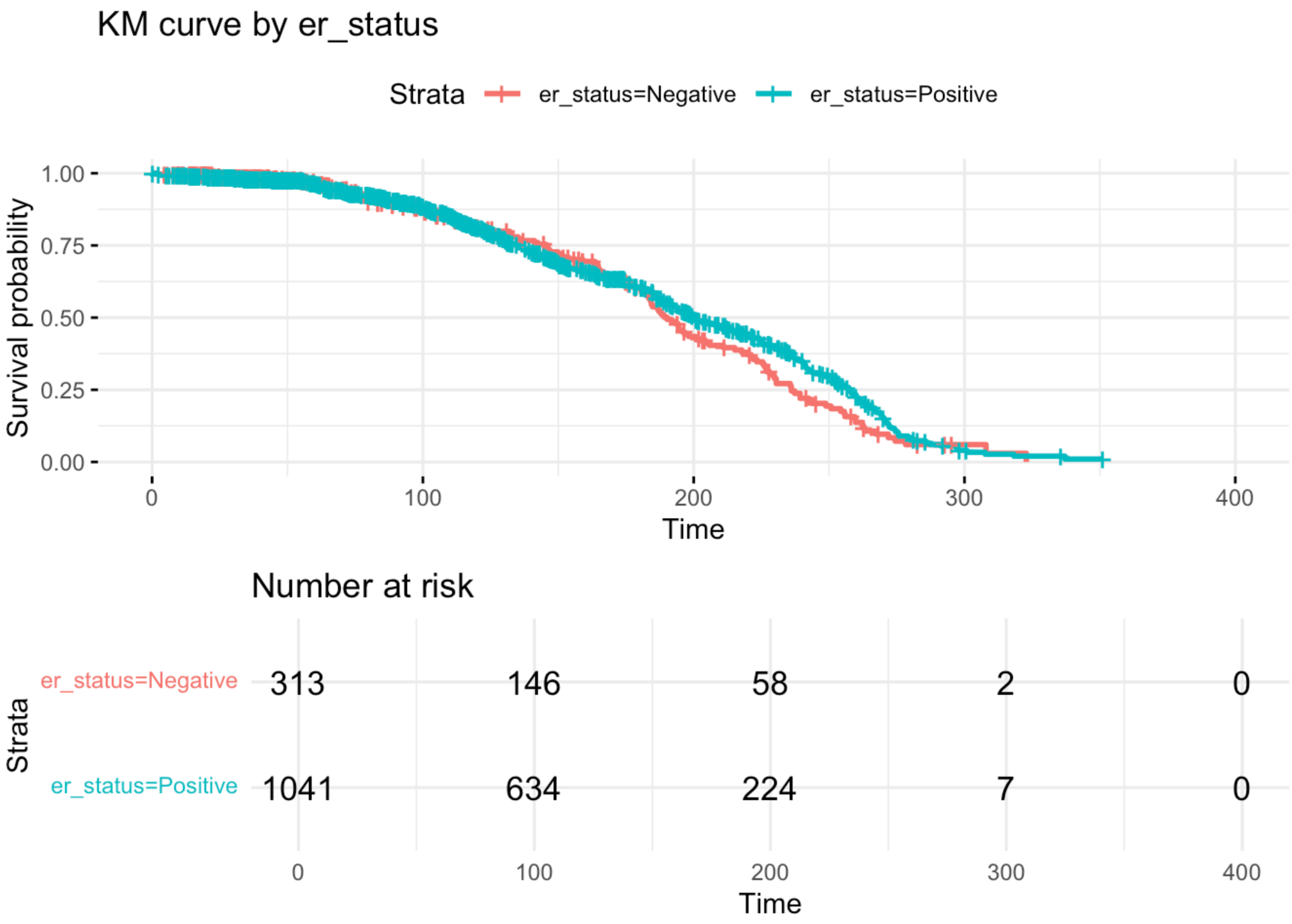


Number at risk

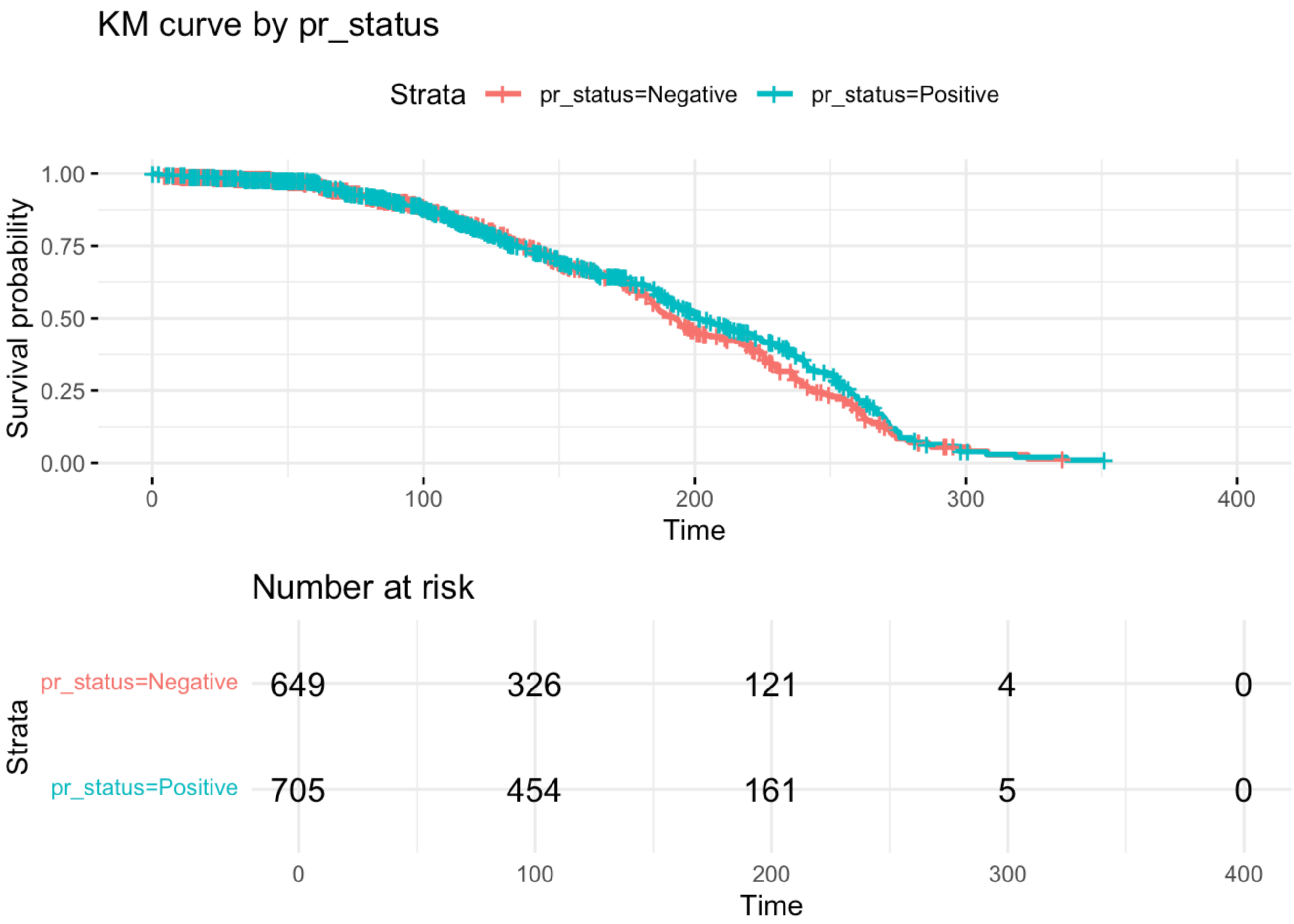
Strata	0	100	200	300	400
radio_therapy=0	447	259	105	3	0
radio_therapy=1	907	521	177	6	0

Time

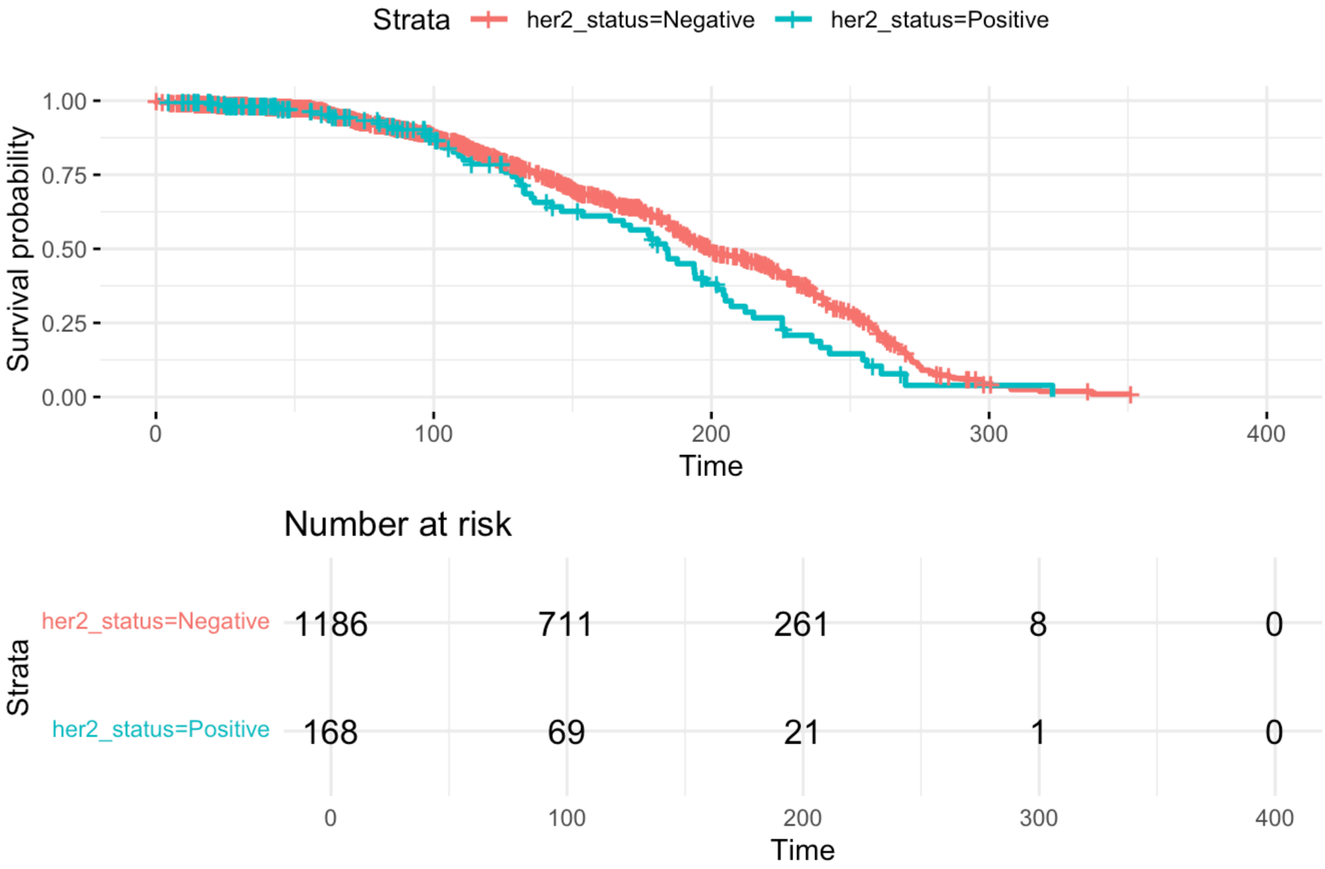
KM curve by er_status



Strata	0	100	200	300	400
er_status=Negative	313	146	58	2	0
er_status=Positive	1041	634	224	7	0

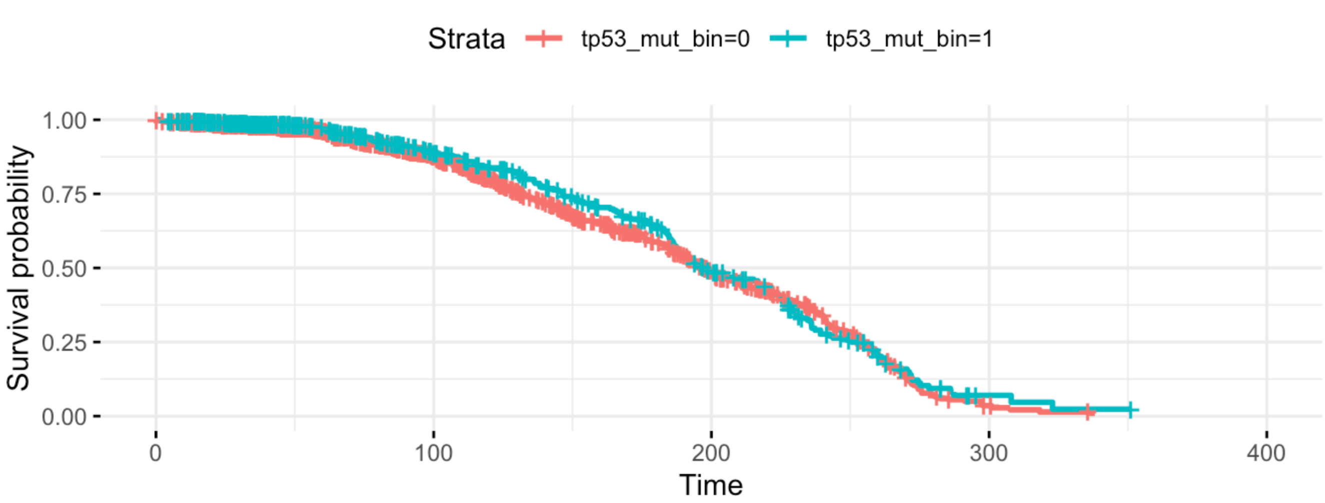

KM curve by pr_status
Strata
pr_status=Negative
pr_status=Positive
Survival probability
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
Time
Number at risk
Strata
pr_status=Negative
649
326
121
4
0
pr_status=Positive
705
454
161
5
0
0
100
200
300
400
Time

KM curve by her2_status



Number at risk

Strata	0	100	200	300	400
her2_status=Negative	1186	711	261	8	0
her2_status=Positive	168	69	21	1	0

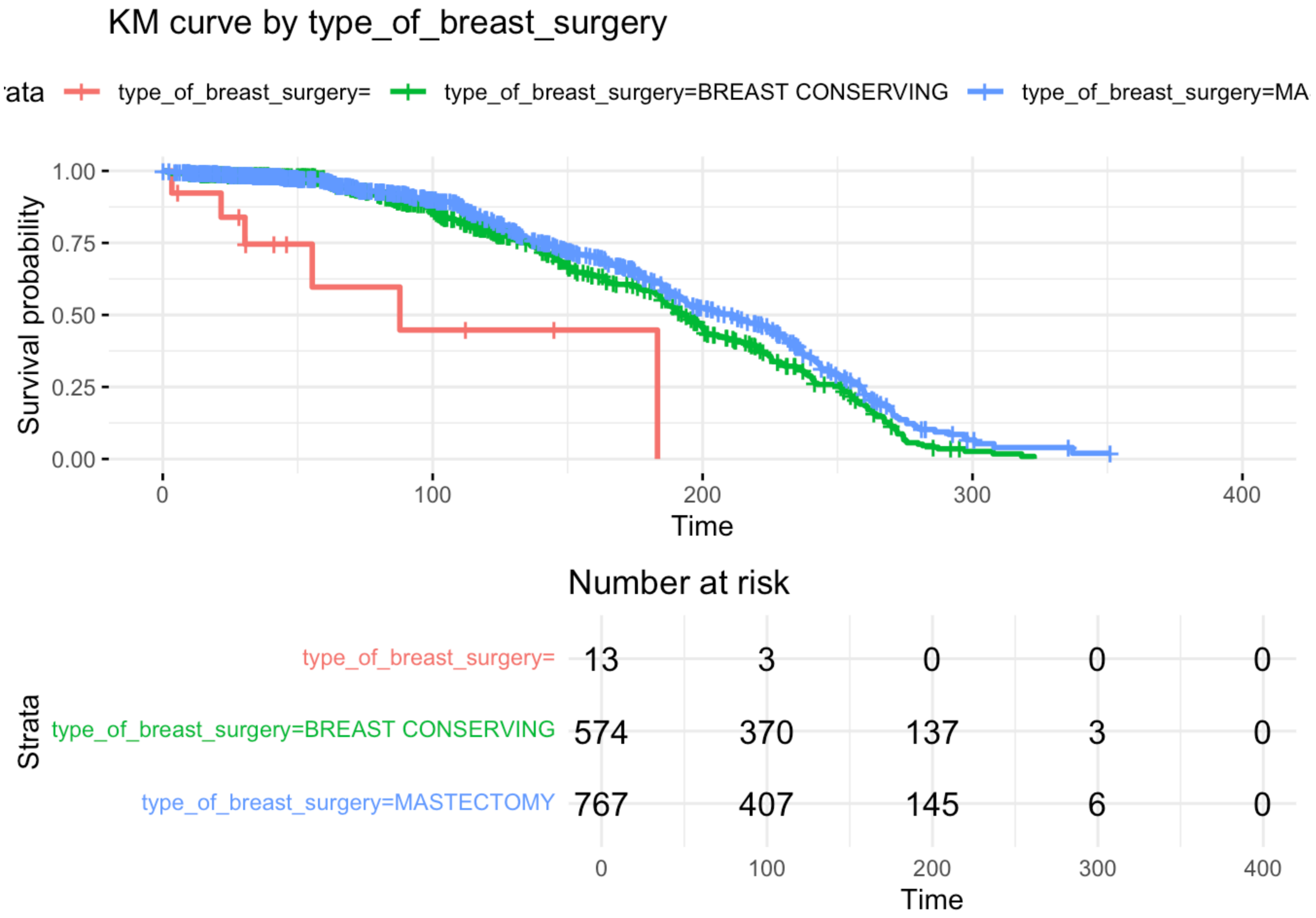


KM curve by tp53_mut_bin

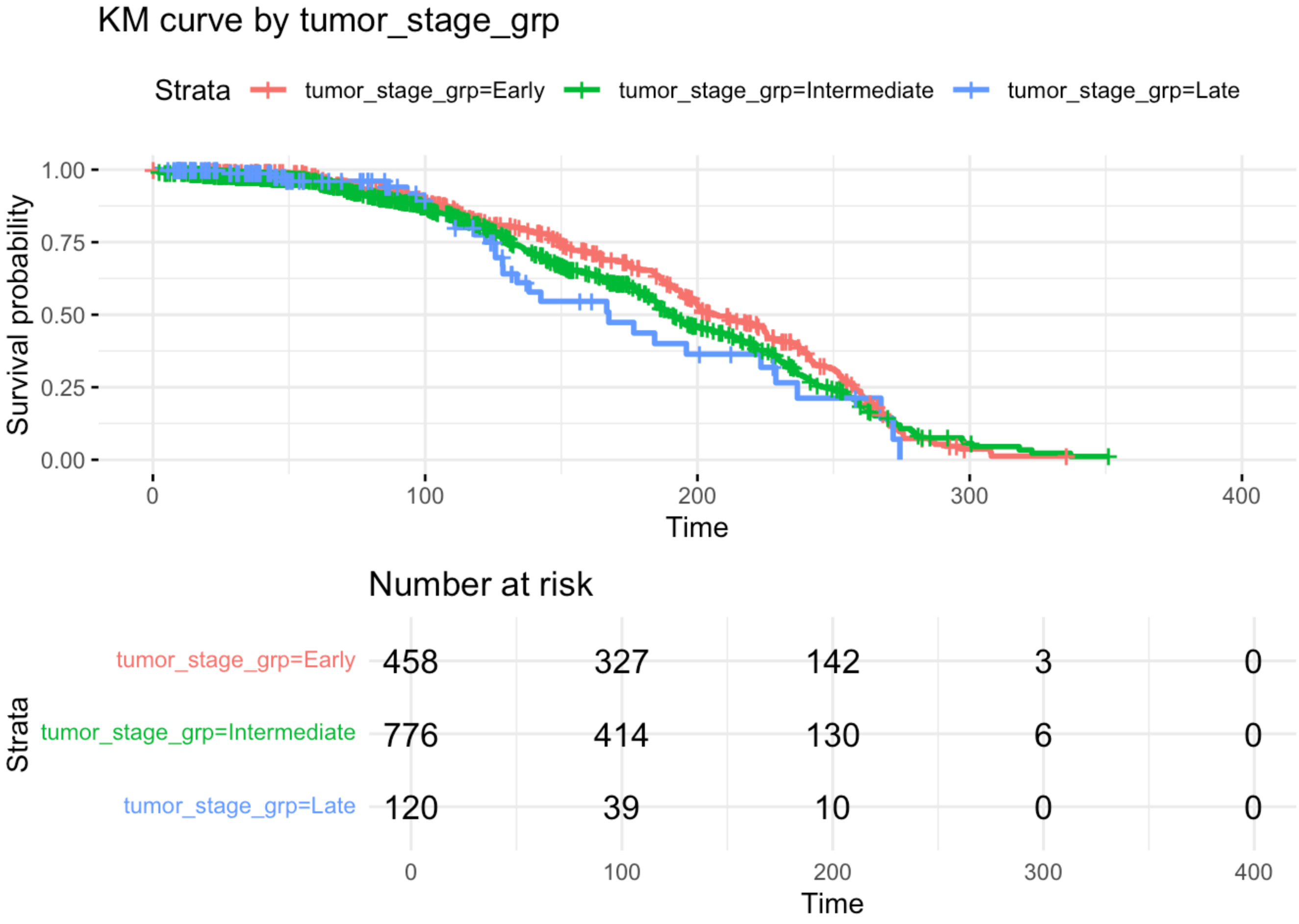
Number at risk

Strata	0	100	200	300	400
tp53_mut_bin=0	886	552	191	6	0
tp53_mut_bin=1	468	228	91	3	0

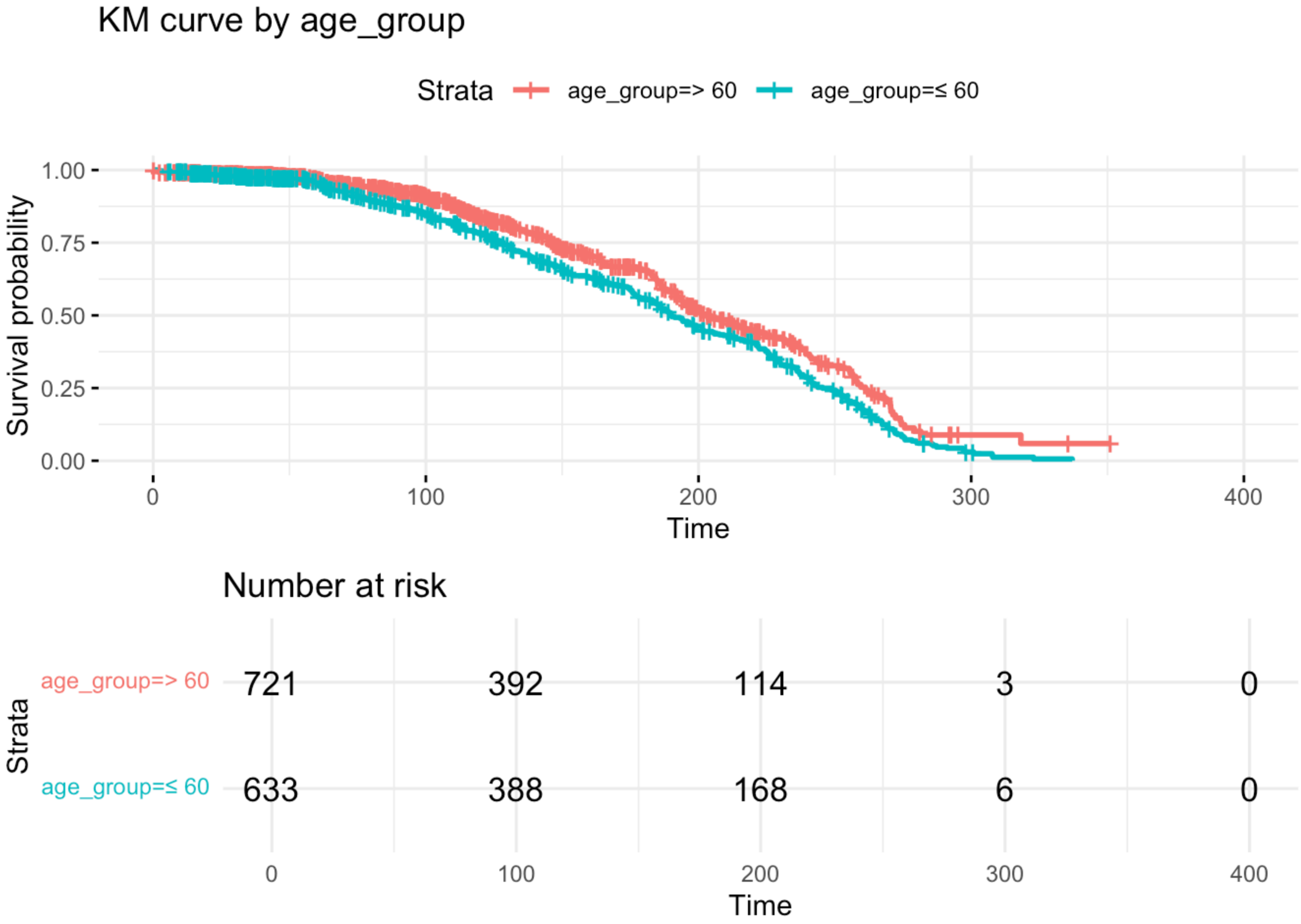
Time



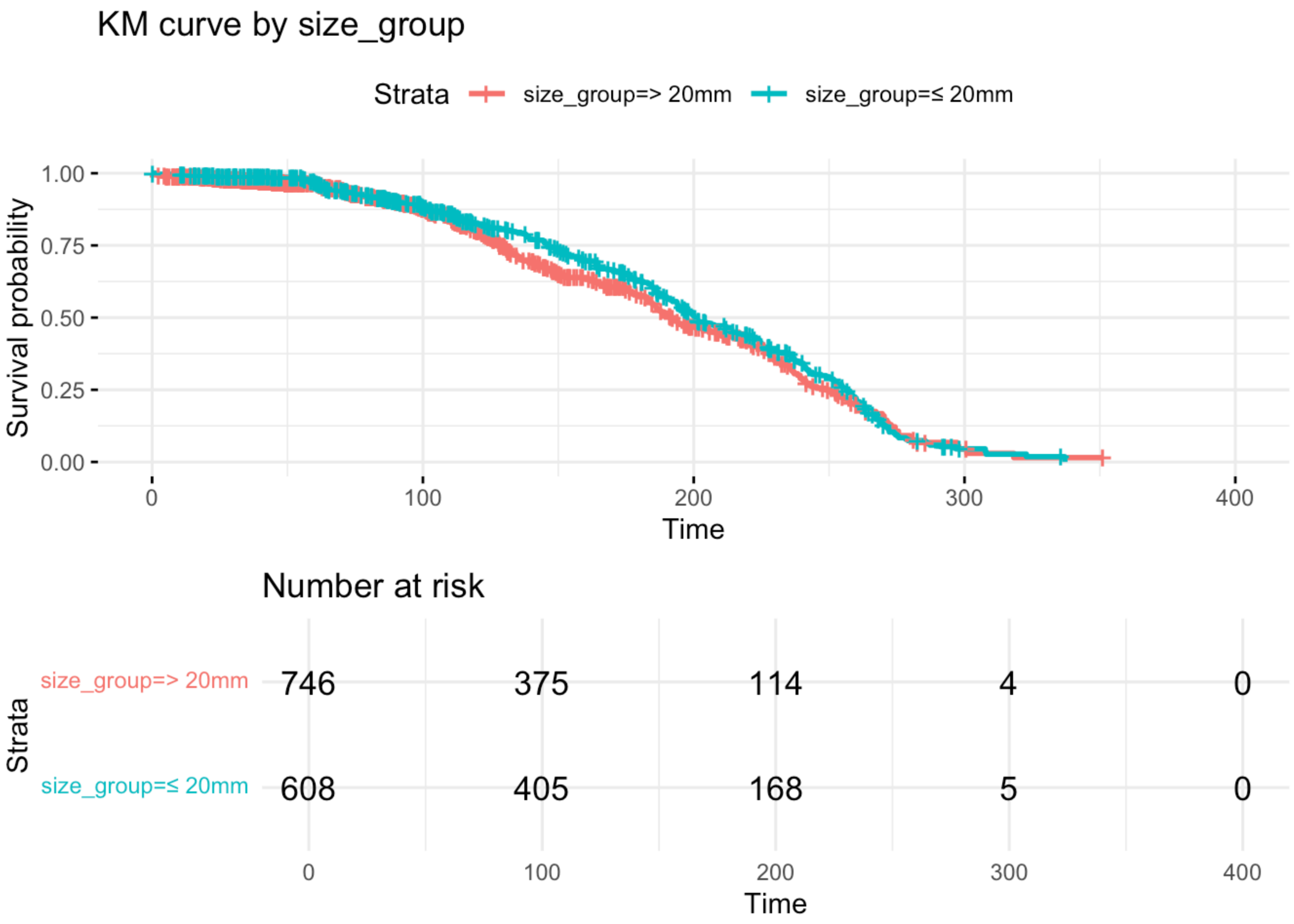
Strata	0	100	200	300	400
type_of_breast_surgery=	13	3	0	0	0
type_of_breast_surgery=BREAST CONSERVING	574	370	137	3	0
type_of_breast_surgery=MASTECTOMY	767	407	145	6	0



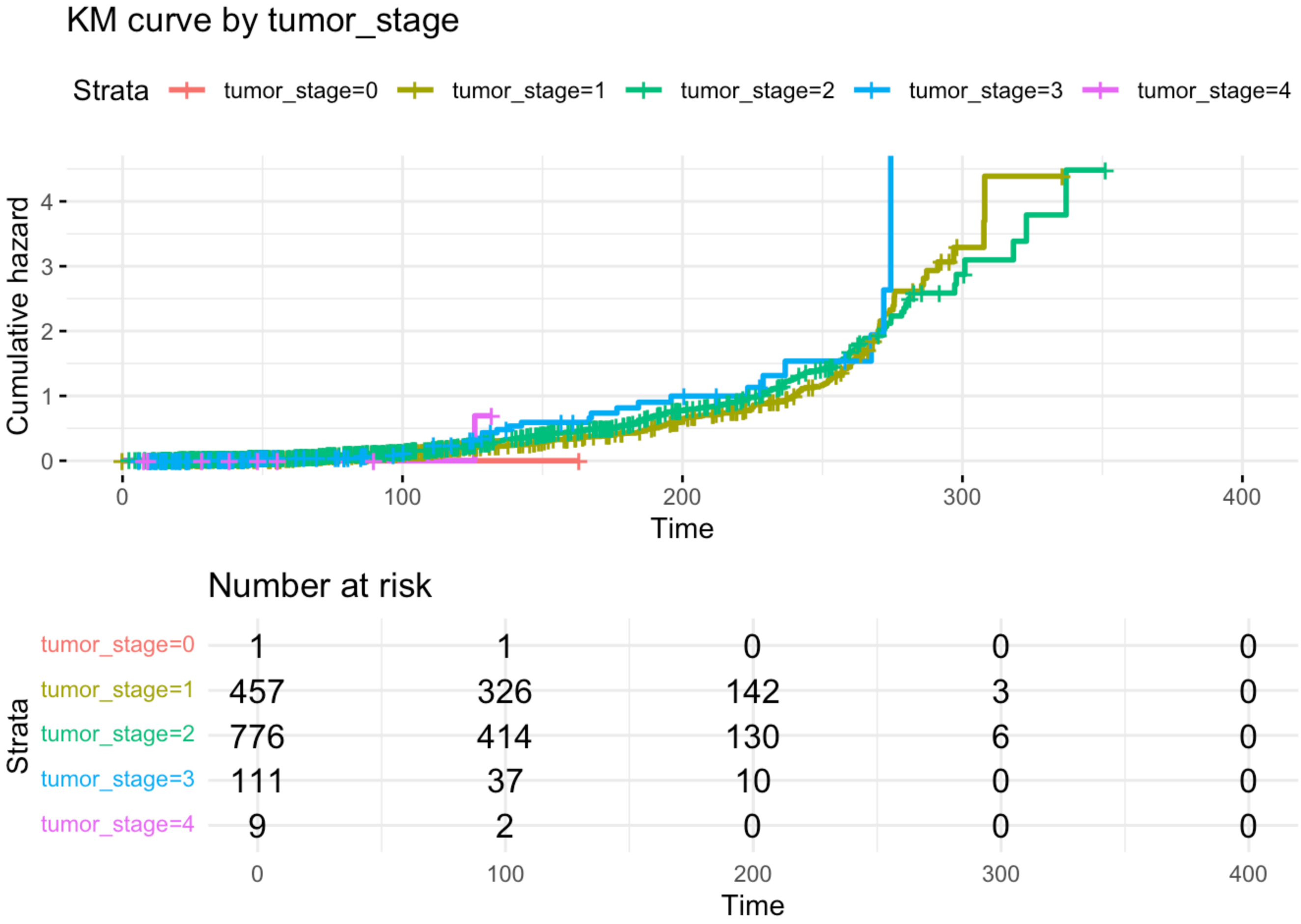
Strata	0	100	200	300	400
tumor_stage_grp=Early	458	327	142	3	0
tumor_stage_grp=Intermediate	776	414	130	6	0
tumor_stage_grp=Late	120	39	10	0	0



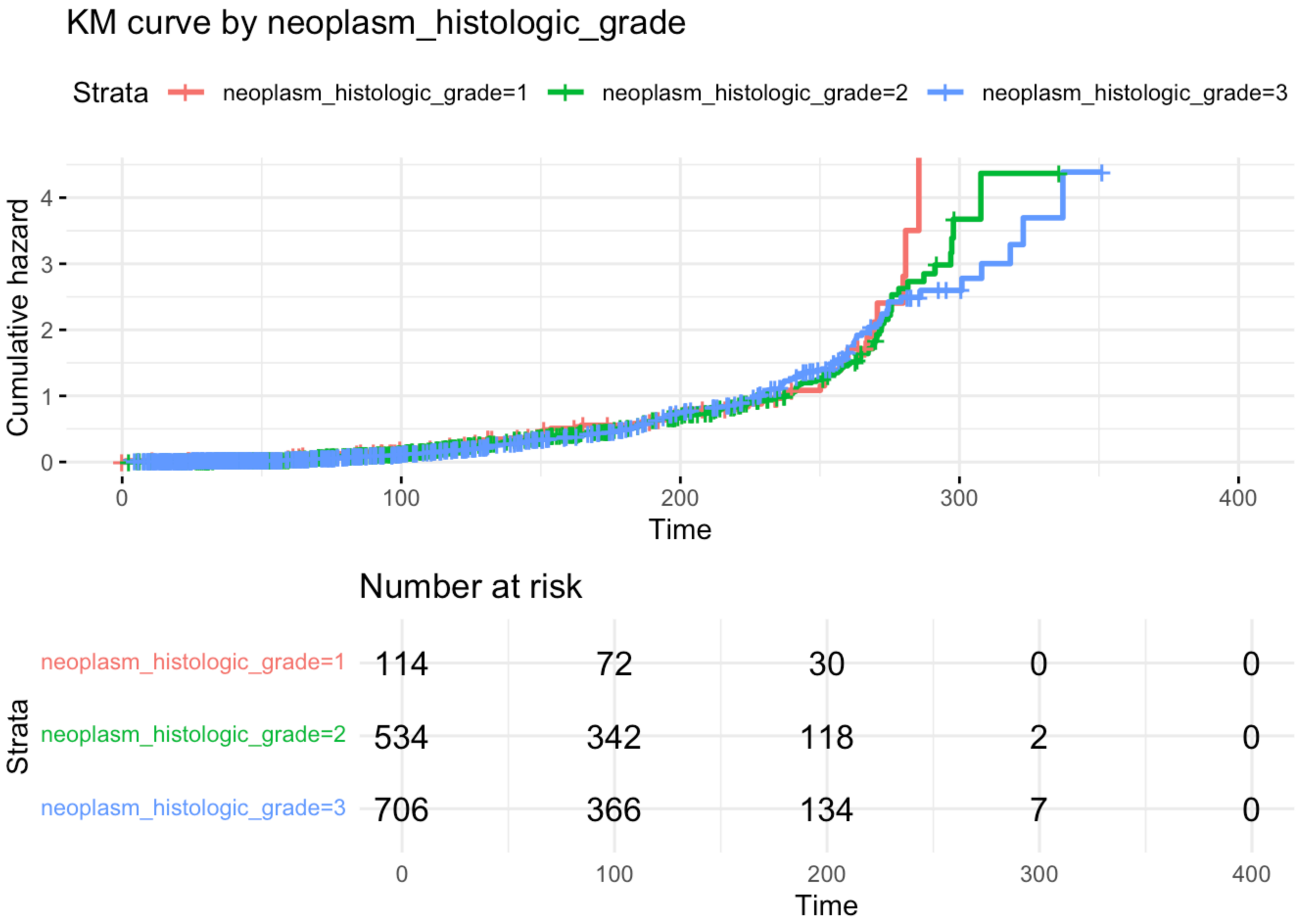
Strata	0	100	200	300	400
age_group=> 60	721	392	114	3	0
age_group=≤ 60	633	388	168	6	0



```
#Cumulative Hazard Plot
for (var in group_vars) {
  formula <- as.formula(paste("Surv(overall_survival_months, overall_survival) ~", var))
  model <- survfit(formula, data = df_model)

  model$call <- list(formula = formula)

  print(ggsurvplot(model,fun ="cumhaz", data = df_model,
             risk.table = TRUE,
             risk.table.height = 0.4,
             title = paste("KM curve by", var),
             ggtheme = theme_minimal()))
}
```

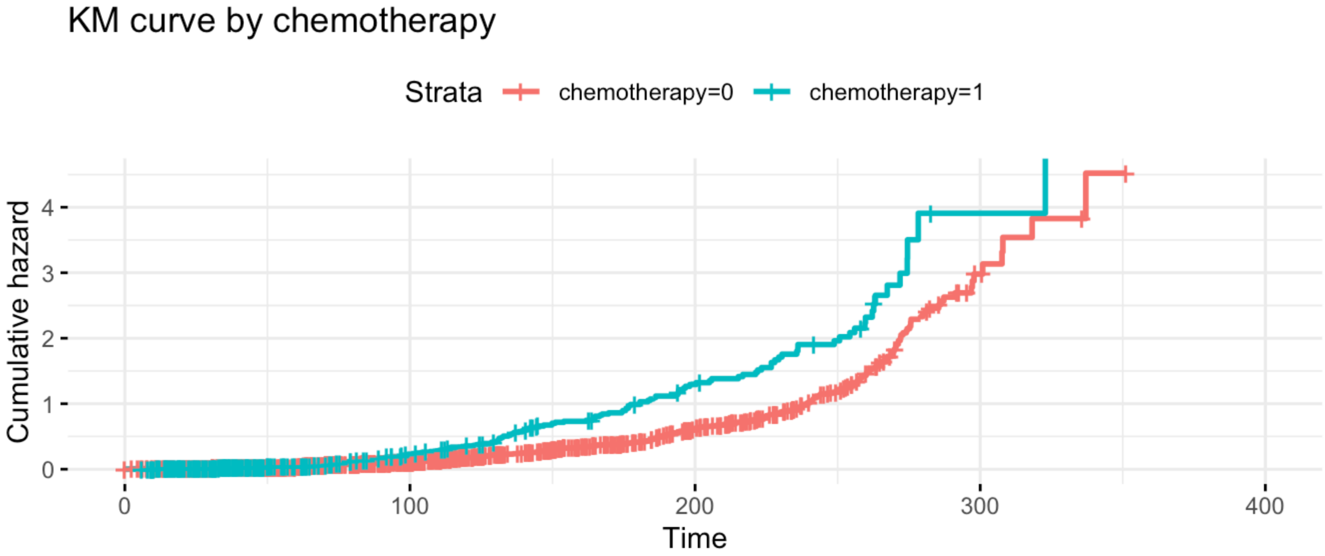


Strata	0	100	200	300	400
tumor_stage=0	1	1	0	0	0
tumor_stage=1	457	326	142	3	0
tumor_stage=2	776	414	130	6	0
tumor_stage=3	111	37	10	0	0
tumor_stage=4	9	2	0	0	0



Number at risk

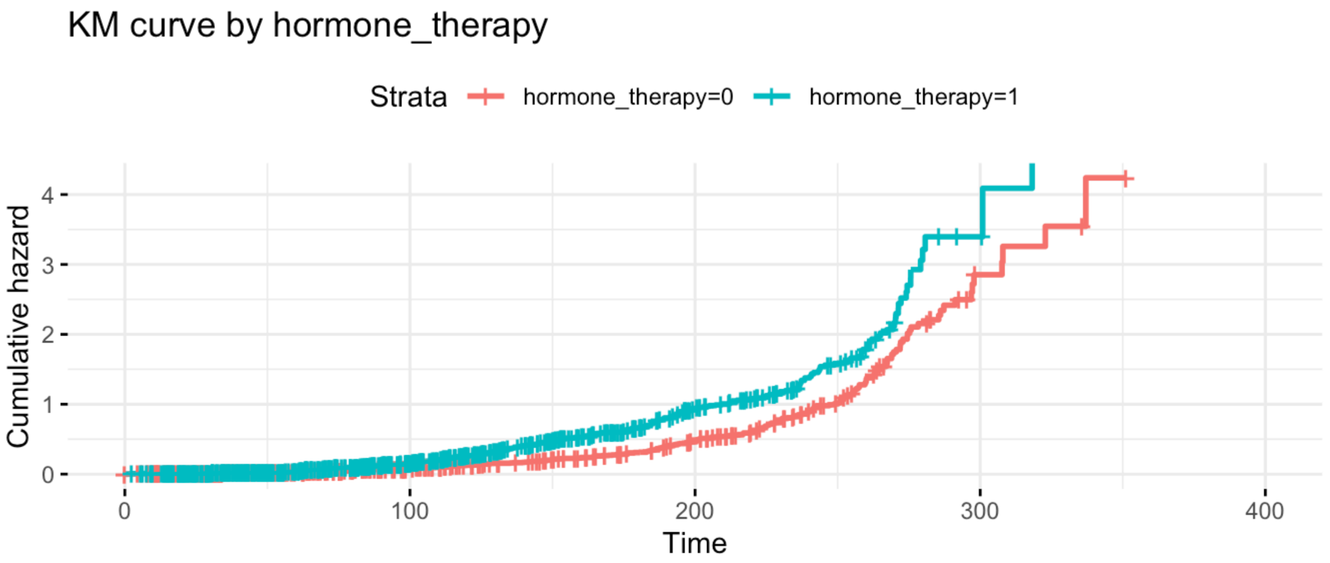
Strata	0	100	200	300	400
neoplasm_histologic_grade=1	114	72	30	0	0
neoplasm_histologic_grade=2	534	342	118	2	0
neoplasm_histologic_grade=3	706	366	134	7	0

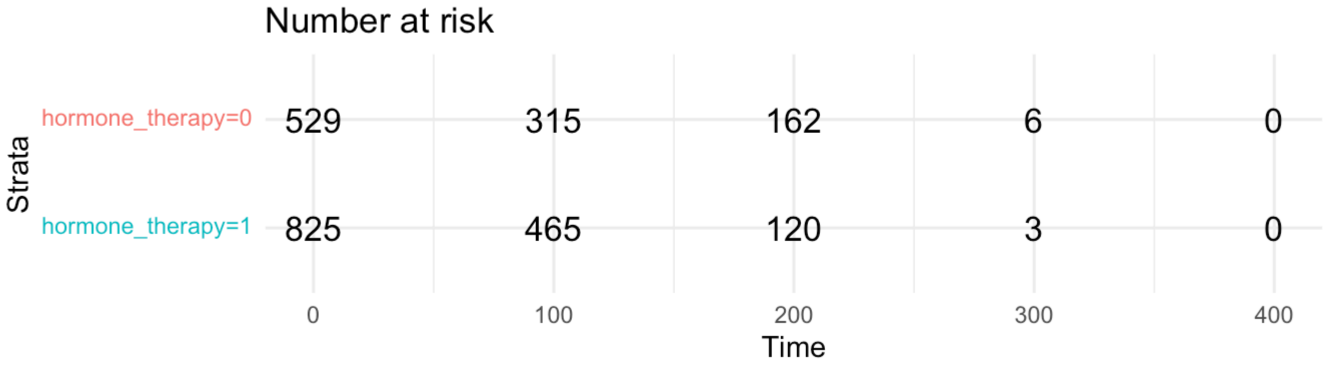

KM curve by chemotherapy

Number at risk

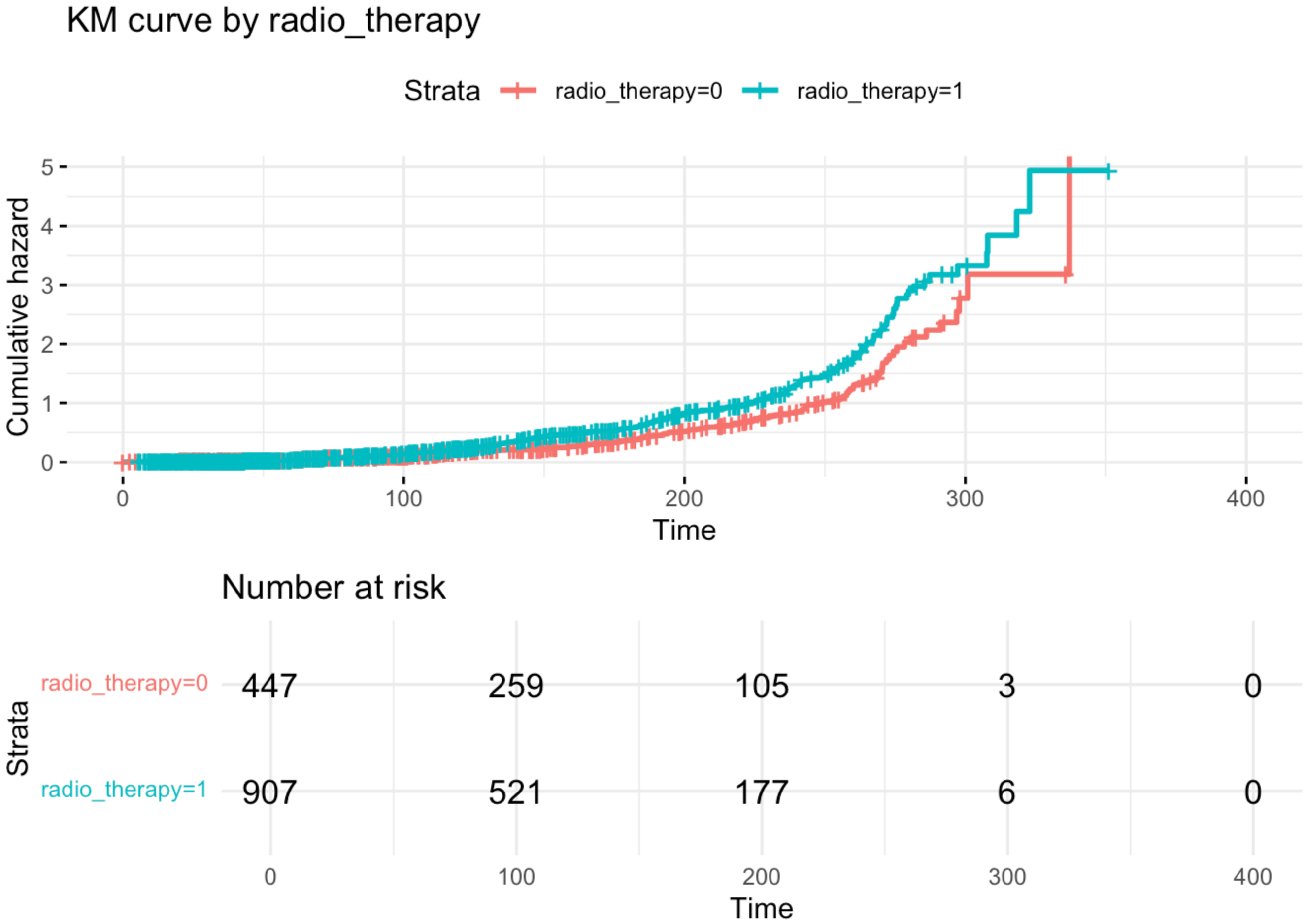
Strata	0	100	200	300	400
chemotherapy=0	1053	649	246	8	0
chemotherapy=1	301	131	36	1	0

Time




Number at risk
Strata
Time

Strata	0	100	200	300	400
hormone_therapy=0	529	315	162	6	0
hormone_therapy=1	825	465	120	3	0

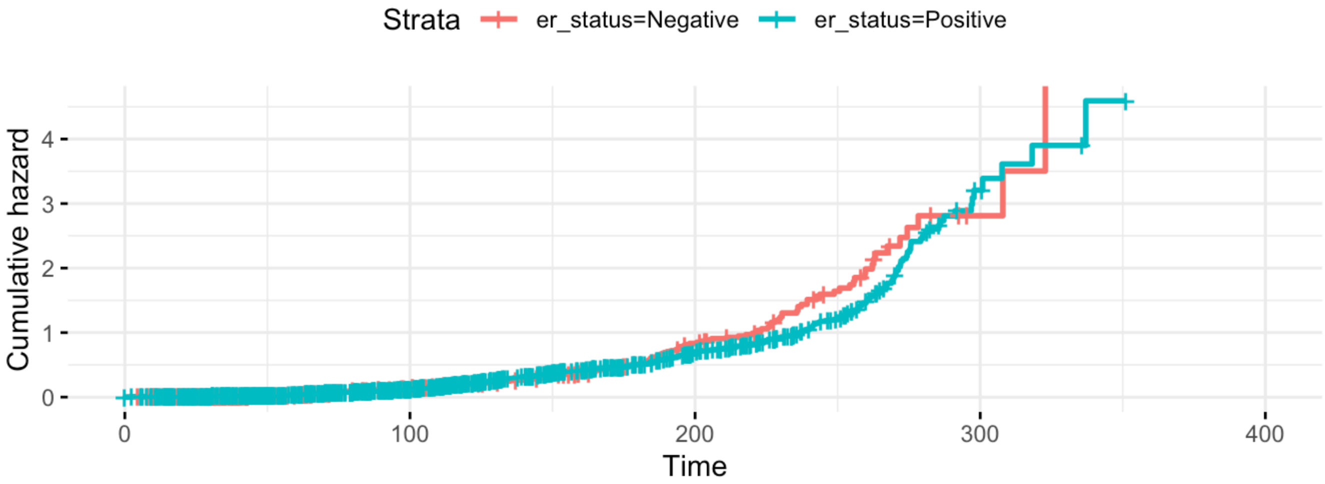


Number at risk

Strata	0	100	200	300	400
radio_therapy=0	447	259	105	3	0
radio_therapy=1	907	521	177	6	0

Time

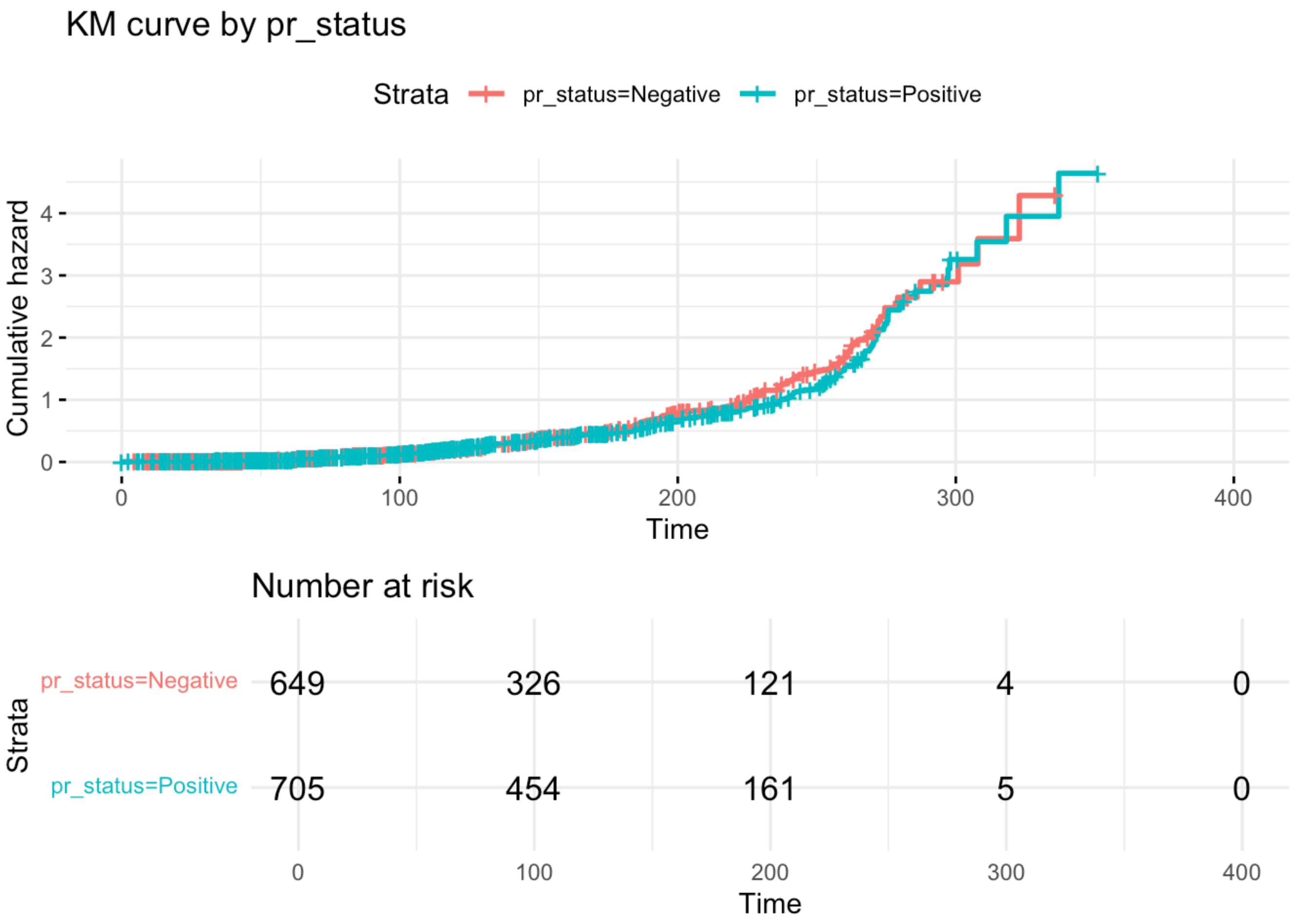
KM curve by er_status



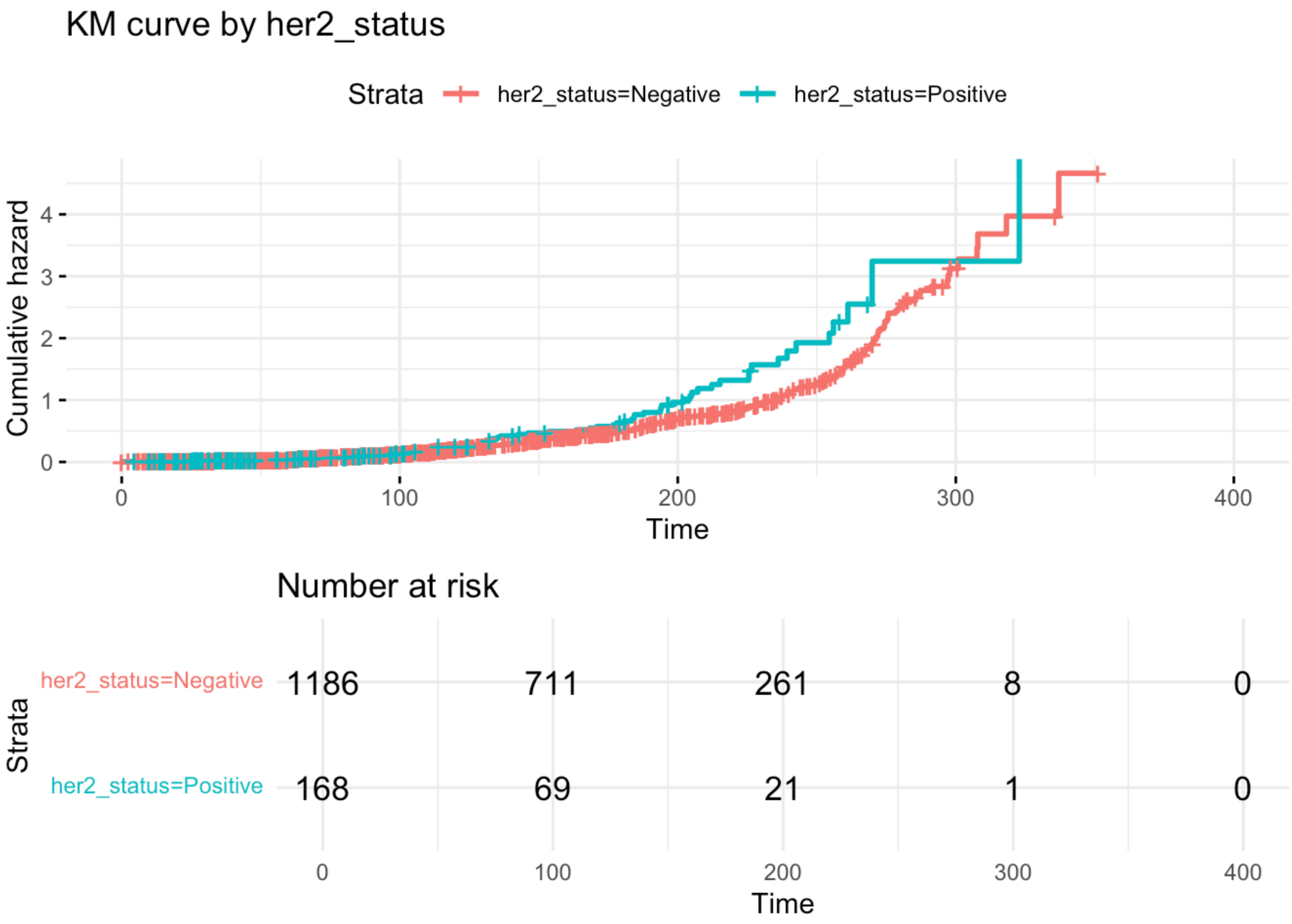
Number at risk

Strata	0	100	200	300	400
er_status=Negative	313	146	58	2	0
er_status=Positive	1041	634	224	7	0

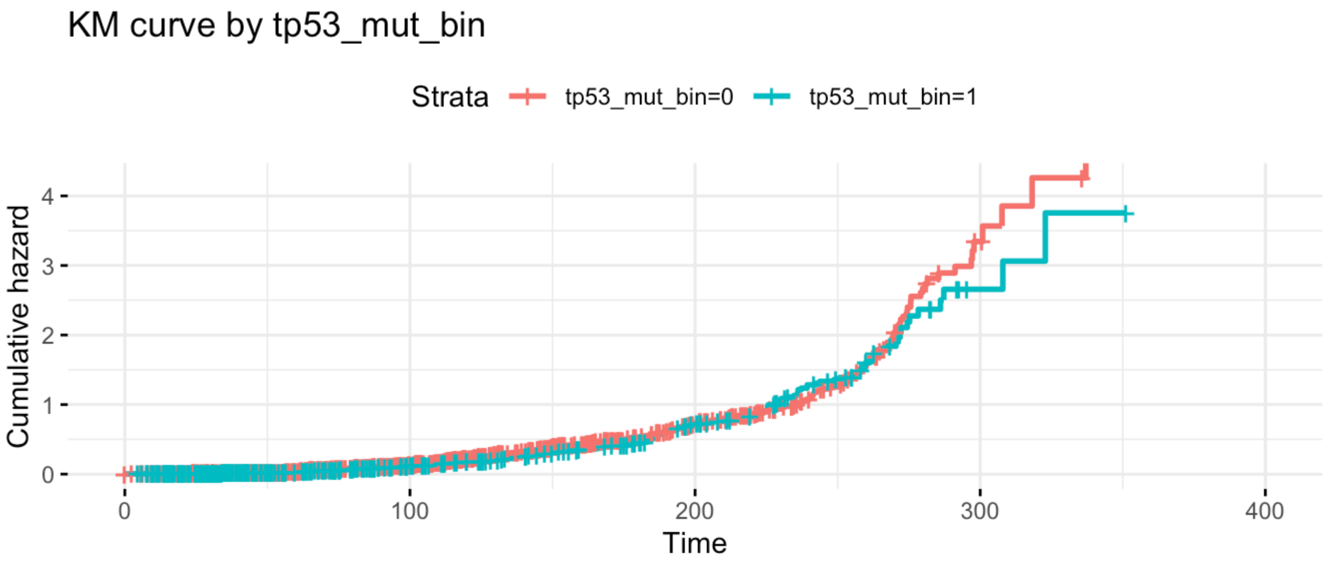
Time



Strata	0	100	200	300	400
pr_status=Negative	649	326	121	4	0
pr_status=Positive	705	454	161	5	0



Strata	0	100	200	300	400
her2_status=Negative	1186	711	261	8	0
her2_status=Positive	168	69	21	1	0

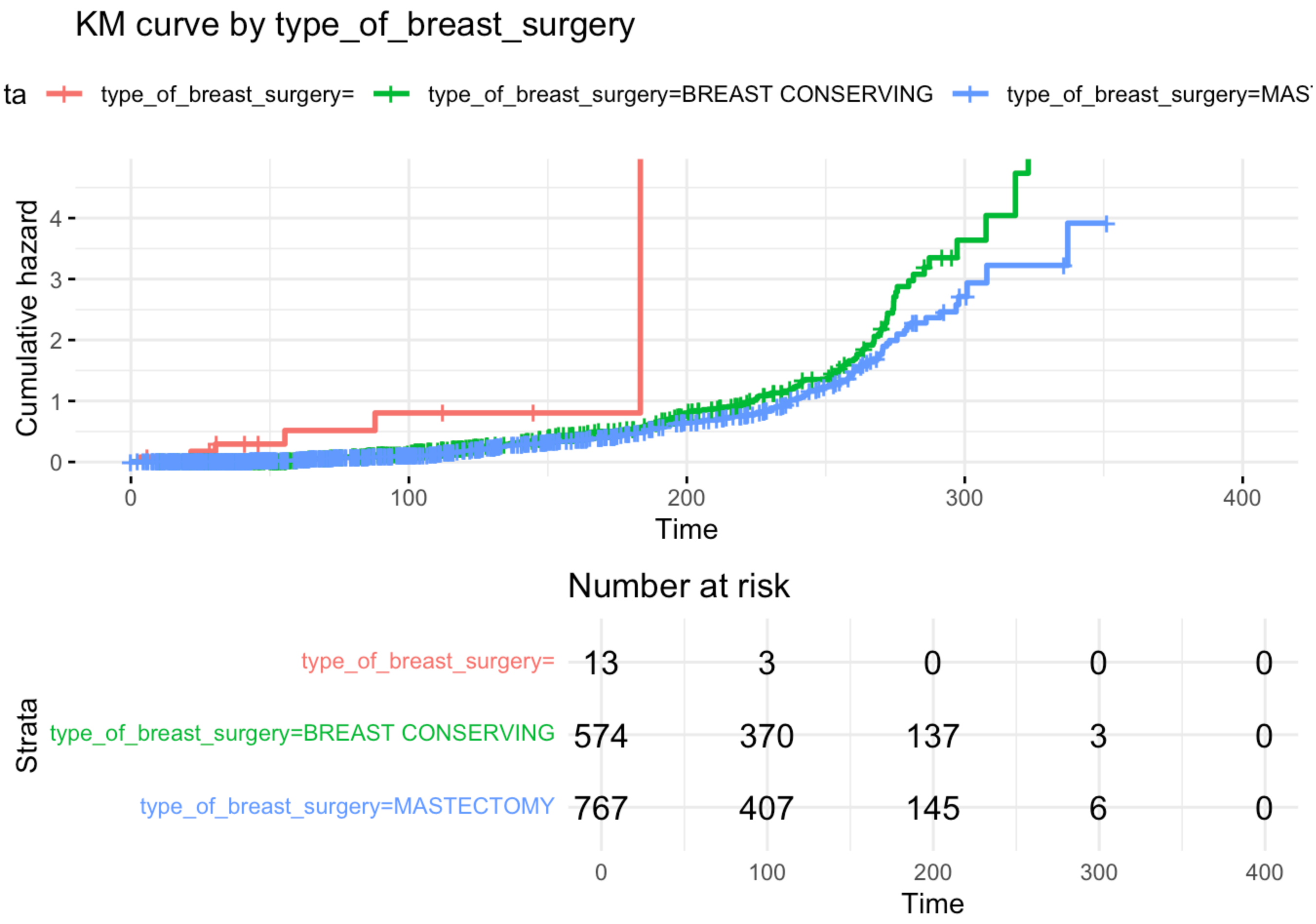


KM curve by tp53_mut_bin

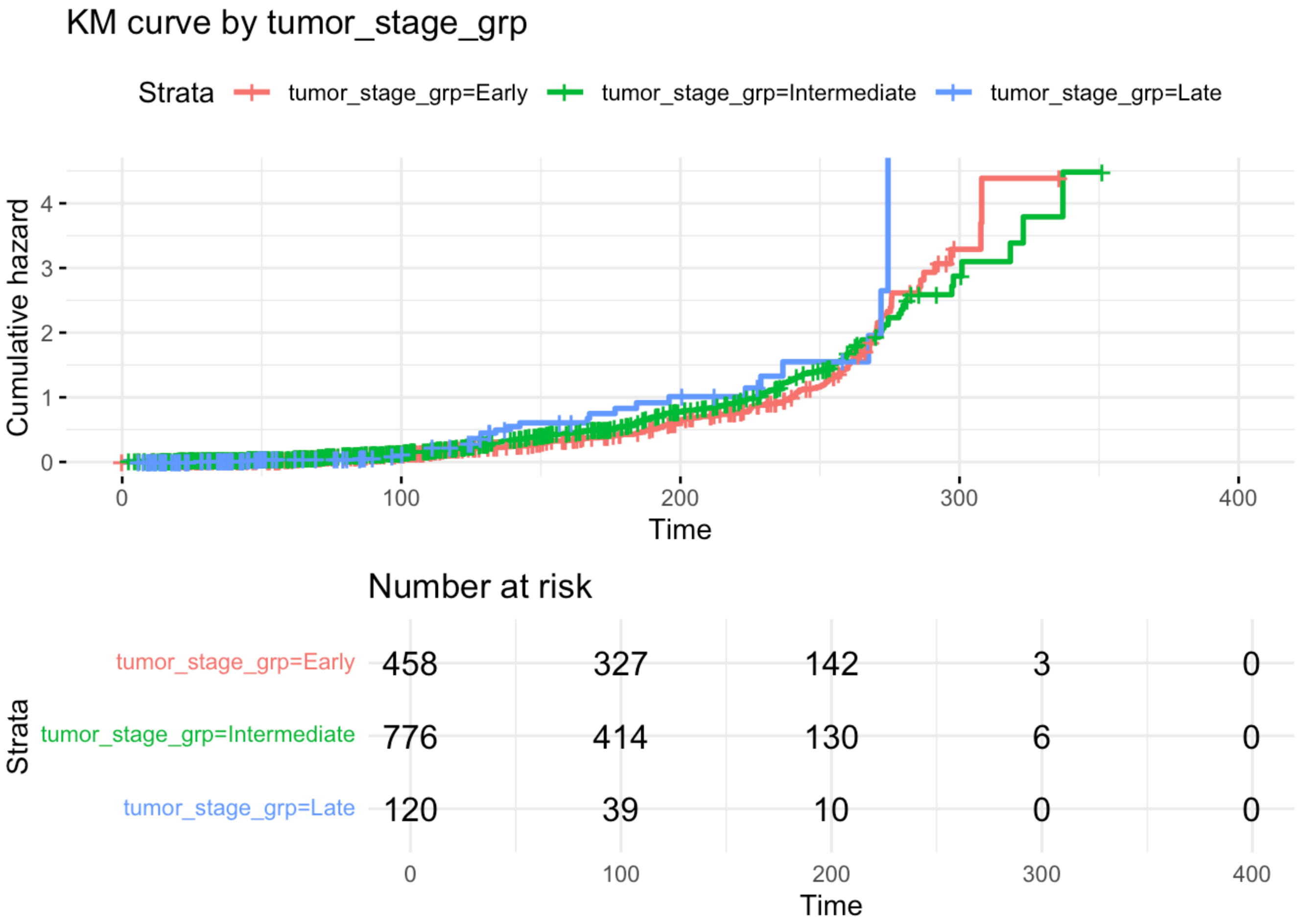
Number at risk

Strata	0	100	200	300	400
tp53_mut_bin=0	886	552	191	6	0
tp53_mut_bin=1	468	228	91	3	0

Time



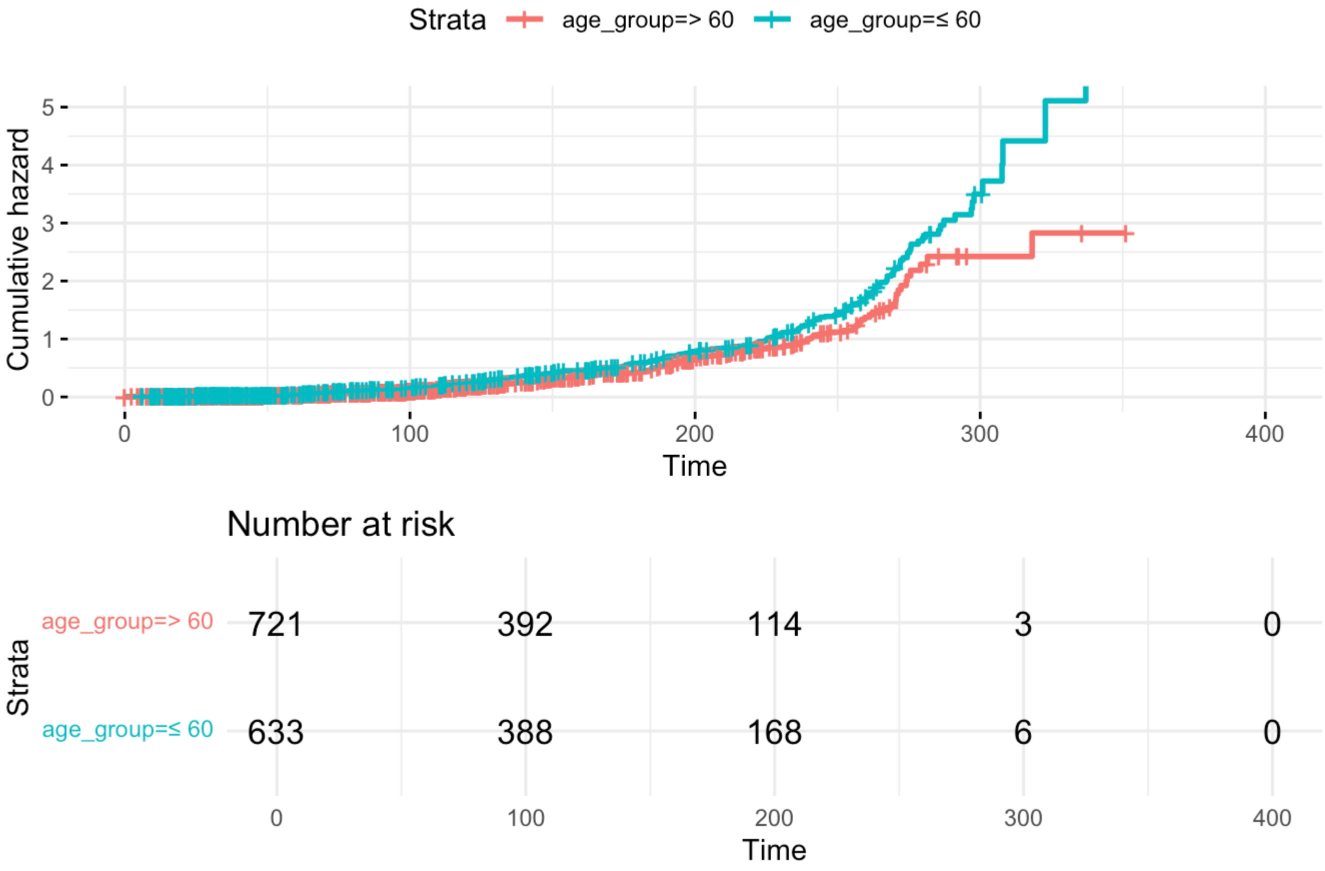
Strata	0	100	200	300	400
type_of_breast_surgery=	13	3	0	0	0
type_of_breast_surgery=BREAST CONSERVING	574	370	137	3	0
type_of_breast_surgery=MASTECTOMY	767	407	145	6	0



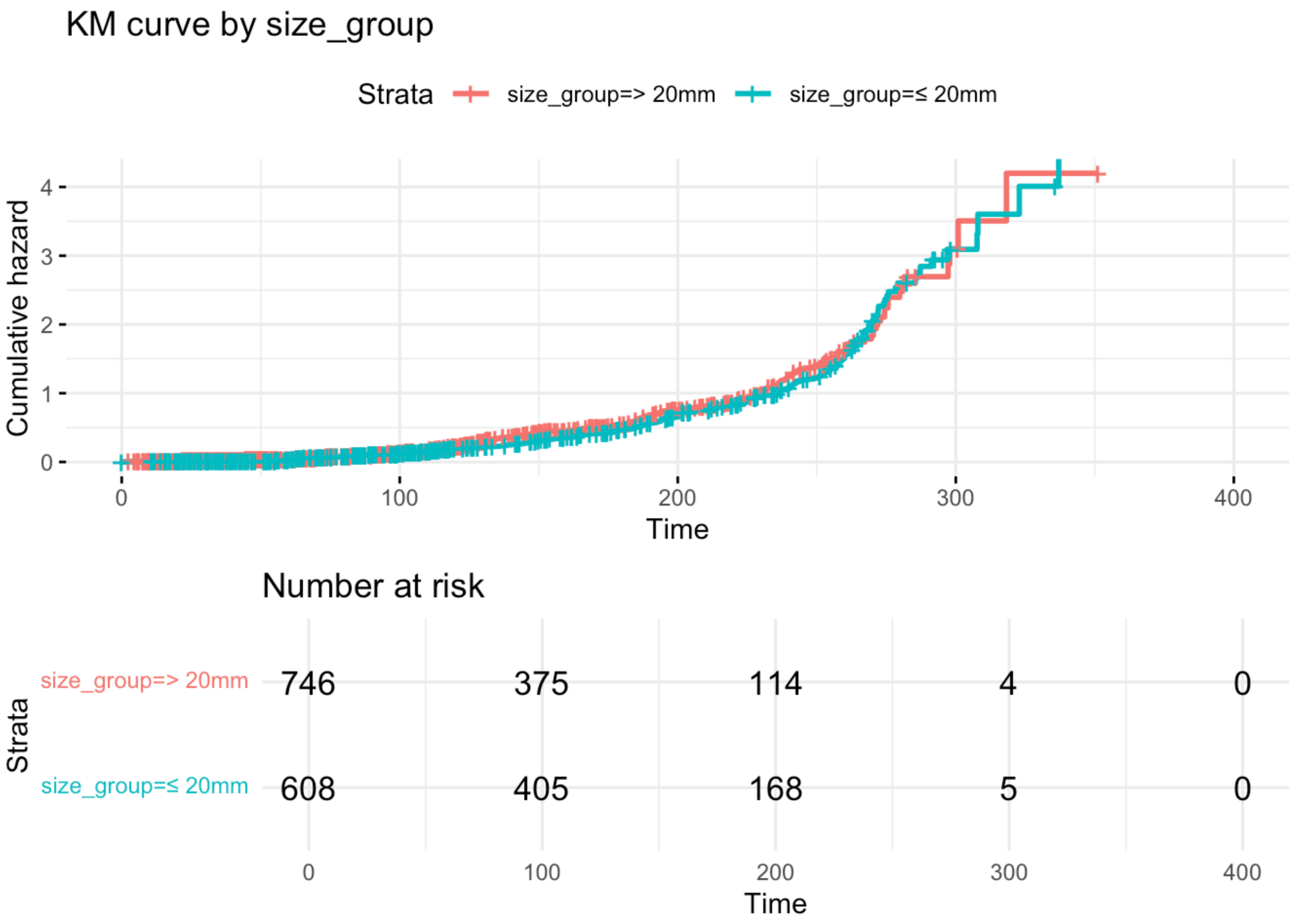
KM curve by tumor_stage_grp

Strata	0	100	200	300	400
tumor_stage_grp=Early	458	327	142	3	0
tumor_stage_grp=Intermediate	776	414	130	6	0
tumor_stage_grp=Late	120	39	10	0	0

KM curve by age_group



Strata	0	100	200	300	400
age_group=> 60	721	392	114	3	0
age_group=≤ 60	633	388	168	6	0



```
get_cox <- function(var, df){
  formula <- as.formula(paste("Surv(overall_survival_months, overall_survival) ~", var))
  model <- coxph(formula, data = df_model)
  var1 <- rownames(summary(model)$coefficients)
  p_value <- summary(model)$coefficients[,"Pr(>|z|)"]
  hr <- summary(model)$coefficients[, "exp(coef)"]
  return(list(variable = var1,
              p_value = p_value,
              harzed_ratio = hr))
}

cox_var <- c(
  "age_at_diagnosis", "tumor_size", "tumor_stage_grp", "neoplasm_histologic_grade",
  "er_status", "pr_status", "her2_status","type_of_breast_surgery", "chemotherapy",
  "hormone_therapy", "radio_therapy", "tp53", "tp53_mut_bin","age_group","size_group"
)
cox_result <- lapply(cox_var, get_cox, df = df_model)
cox_result_df <- do.call(rbind, lapply(cox_result, as.data.frame))
```

```
cox_result_df
```

```
                                                                       variable
1                                                              age_at_diagnosis
2                                                                    tumor_size
tumor_stage_grpIntermediate                         tumor_stage_grpIntermediate
tumor_stage_grpLate                                         tumor_stage_grpLate
neoplasm_histologic_grade2                           neoplasm_histologic_grade2
neoplasm_histologic_grade3                           neoplasm_histologic_grade3
11                                                            er_statusPositive
12                                                            pr_statusPositive
13                                                          her2_statusPositive
type_of_breast_surgeryBREAST CONSERVING type_of_breast_surgeryBREAST CONSERVING
type_of_breast_surgeryMASTECTOMY               type_of_breast_surgeryMASTECTOMY
14                                                                chemotherapy1
15                                                             hormone_therapy1
16                                                               radio_therapy1
17                                                                         tp53
18                                                                tp53_mut_bin1
19                                                                age_group≤ 60
110                                                            size_group≤ 20mm
                                             p_value harzed_ratio
1                                       1.603294e-02    0.9914464
2                                       6.605071e-01    1.0015440
tumor_stage_grpIntermediate             1.169882e-01    1.1426042
tumor_stage_grpLate                     1.128365e-01    1.3599931
neoplasm_histologic_grade2              5.205099e-01    0.9137327
neoplasm_histologic_grade3              5.289992e-01    0.9163688
11                                      1.536674e-01    0.8688951
12                                      2.656370e-01    0.9123345
13                                      1.915166e-02    1.3680895
type_of_breast_surgeryBREAST CONSERVING 4.538662e-04    0.2338315
type_of_breast_surgeryMASTECTOMY        5.837651e-05    0.1888122
14                                      2.499391e-10    1.8317210
15                                      1.155988e-08    1.6316368
16                                      7.050152e-05    1.4387202
17                                      1.933130e-04    1.1674635
18                                      3.865868e-01    0.9255697
19                                      1.558315e-03    1.3101685
110                                     2.446481e-01    0.9082791
```

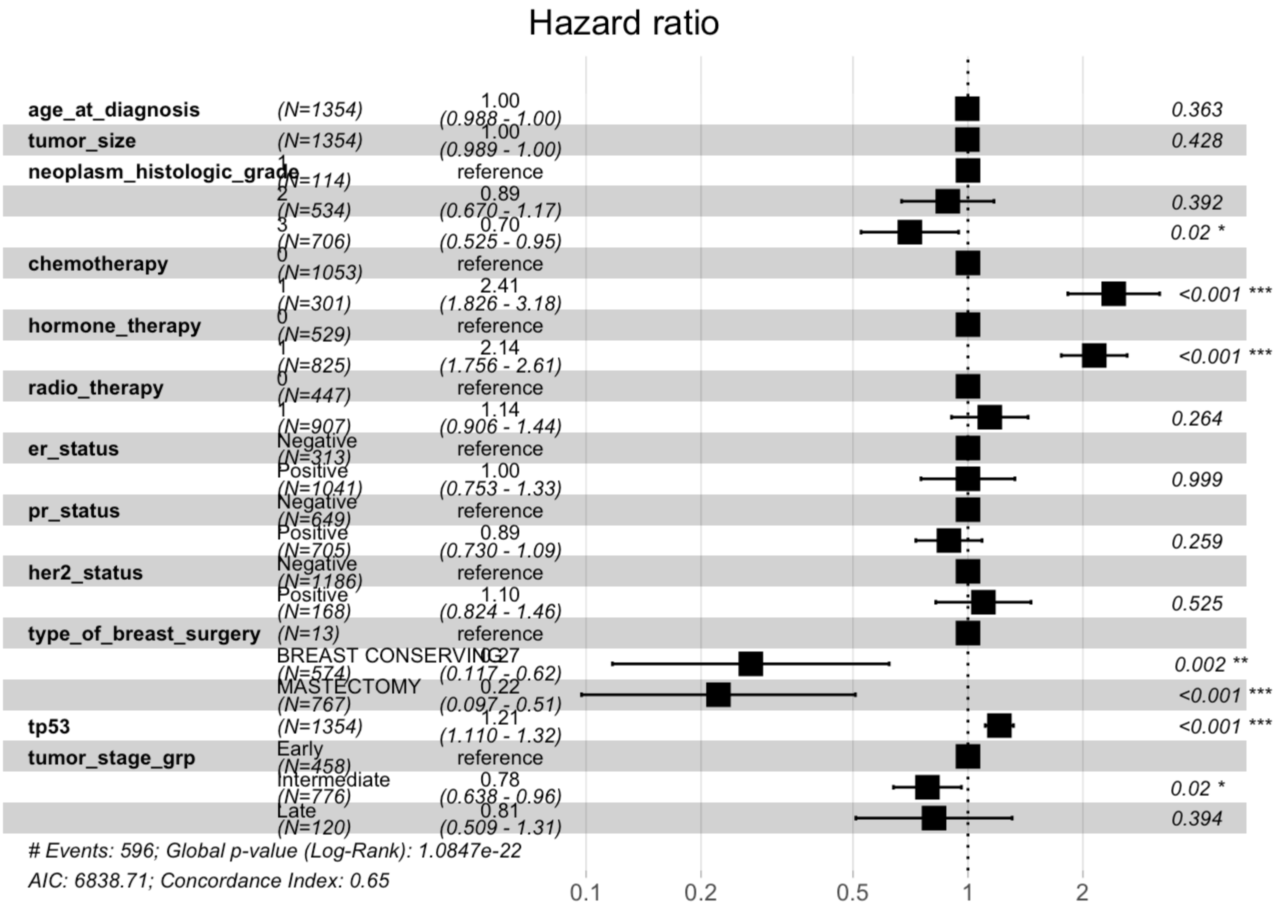
```r
df_cox <- df_model %>% select(-age_group,-size_group,-tp53_mut_bin,-tumor_stage)
cox_fit_total <- coxph(Surv(overall_survival_months,overall_survival) ~ ., data = df_cox)
summary(cox_fit_total)
```

```
Call:
coxph(formula = Surv(overall_survival_months, overall_survival) ~ 
    ., data = df_cox)

  n= 1354, number of events= 596 

                                              coef  exp(coef)   se(coef)      z
```

```
age_at_diagnosis                        -0.0038330  0.9961743  0.0042129 -0.910
tumor_size                              -0.0032810  0.9967244  0.0041428 -0.792
neoplasm_histologic_grade2              -0.1216521  0.8854564  0.1421567 -0.856
neoplasm_histologic_grade3              -0.3504253  0.7043884  0.1500841 -2.335
chemotherapy1                            0.8790028  2.4084967  0.1412044  6.225
hormone_therapy1                         0.7612930  2.1410429  0.1012790  7.517
radio_therapy1                           0.1317201  1.1407890  0.1178195  1.118
er_statusPositive                       -0.0002493  0.9997508  0.1444976 -0.002
pr_statusPositive                       -0.1148094  0.8915361  0.1017957 -1.128
her2_statusPositive                      0.0930050  1.0974672  0.1464554  0.635
type_of_breast_surgeryBREAST CONSERVING -1.3092310  0.2700276  0.4256437 -3.076
type_of_breast_surgeryMASTECTOMY        -1.5042121  0.2221923  0.4212916 -3.570
tp53                                     0.1894068  1.2085324  0.0433754  4.367
tumor_stage_grpIntermediate             -0.2442978  0.7832543  0.1045932 -2.336
tumor_stage_grpLate                     -0.2046972  0.8148940  0.2403548 -0.852
                                        Pr(>|z|)    
age_at_diagnosis                        0.362905    
tumor_size                              0.428386    
neoplasm_histologic_grade2              0.392130    
neoplasm_histologic_grade3              0.019551 *  
chemotherapy1                           4.81e-10 ***
hormone_therapy1                        5.61e-14 ***
radio_therapy1                          0.263575    
er_statusPositive                       0.998624    
pr_statusPositive                       0.259387    
her2_statusPositive                     0.525402    
type_of_breast_surgeryBREAST CONSERVING 0.002099 ** 
type_of_breast_surgeryMASTECTOMY        0.000356 ***
tp53                                    1.26e-05 ***
tumor_stage_grpIntermediate             0.019507 *  
tumor_stage_grpLate                     0.394411    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                                        exp(coef) exp(-coef) lower .95
age_at_diagnosis                           0.9962     1.0038    0.9880
tumor_size                                 0.9967     1.0033    0.9887
neoplasm_histologic_grade2                 0.8855     1.1294    0.6701
neoplasm_histologic_grade3                 0.7044     1.4197    0.5249
chemotherapy1                              2.4085     0.4152    1.8262
hormone_therapy1                           2.1410     0.4671    1.7556
radio_therapy1                             1.1408     0.8766    0.9056
er_statusPositive                          0.9998     1.0002    0.7532
pr_statusPositive                          0.8915     1.1217    0.7303
her2_statusPositive                        1.0975     0.9112    0.8236
type_of_breast_surgeryBREAST CONSERVING    0.2700     3.7033    0.1172
type_of_breast_surgeryMASTECTOMY           0.2222     4.5006    0.0973
tp53                                       1.2085     0.8274    1.1100
tumor_stage_grpIntermediate                0.7833     1.2767    0.6381
tumor_stage_grpLate                        0.8149     1.2272    0.5088
                                        upper .95
```

```
age_at_diagnosis                          1.0044
tumor_size                                1.0049
neoplasm_histologic_grade2                1.1700
neoplasm_histologic_grade3                0.9453
chemotherapy1                             3.1764
hormone_therapy1                          2.6112
radio_therapy1                            1.4371
er_statusPositive                         1.3271
pr_statusPositive                         1.0884
her2_statusPositive                       1.4624
type_of_breast_surgeryBREAST CONSERVING   0.6219
type_of_breast_surgeryMASTECTOMY          0.5074
tp53                                      1.3158
tumor_stage_grpIntermediate               0.9615
tumor_stage_grpLate                       1.3052

Concordance= 0.651  (se = 0.013 )
Likelihood ratio test= 141.7  on 15 df,   p=<2e-16
Wald test            = 149.5  on 15 df,   p=<2e-16
Score (logrank) test = 155.1  on 15 df,   p=<2e-16
```

```
ggforest(cox_fit_total, data = df_model)
```



```
#这个我感觉没啥用可以删了
df_lasso <- df_model[df_model$overall_survival_months > 0, ]
df_lasso <- na.omit(df_lasso)
X <- model.matrix(Surv(overall_survival_months,overall_survival) ~ ., data = df_lasso)[,
y <- Surv(df_lasso$overall_survival_months,df_lasso$overall_survival)
cox_lasso <- cv.glmnet(X, y, family = "cox", alpha = 1)
selected_vars <- rownames(coef(cox_lasso))[
  as.vector(coef(cox_lasso, s = "lambda.min")) != 0
]
selected_vars
```

```
 [1] "tumor_size"                        "tumor_stage1"
 [3] "neoplasm_histologic_grade3"        "chemotherapy1"
 [5] "hormone_therapy1"                  "radio_therapy1"
 [7] "pr_statusPositive"                 "her2_statusPositive"
 [9] "tp53_mut_bin1"                     "type_of_breast_surgeryMASTECTOMY"
[11] "tp53"                              "age_group≤ 60"
```

```
reg_feature <- c("overall_survival_months", "overall_survival","tumor_stage","neoplasm_hi
df_data <- df_model %>%
  select(all_of(reg_feature)) %>%
  filter(overall_survival_months > 0) %>%
  na.omit()
```

```
aft_formula <- as.formula(paste("Surv(overall_survival_months, overall_survival) ~",
                                paste(reg_feature[-c(1,2)], collapse = " + ")))
```

```
model_lognormal <- survreg(aft_formula, data = df_data, dist = "lognormal")
model_weibull <- survreg(aft_formula, data = df_data, dist = "weibull")
model_exponential <- survreg(aft_formula, data = df_data, dist = "exponential")
model_loglogistic <- survreg(aft_formula, data = df_data, dist = "loglogistic")
```

```
AIC(model_lognormal, model_weibull, model_exponential, model_loglogistic)
```

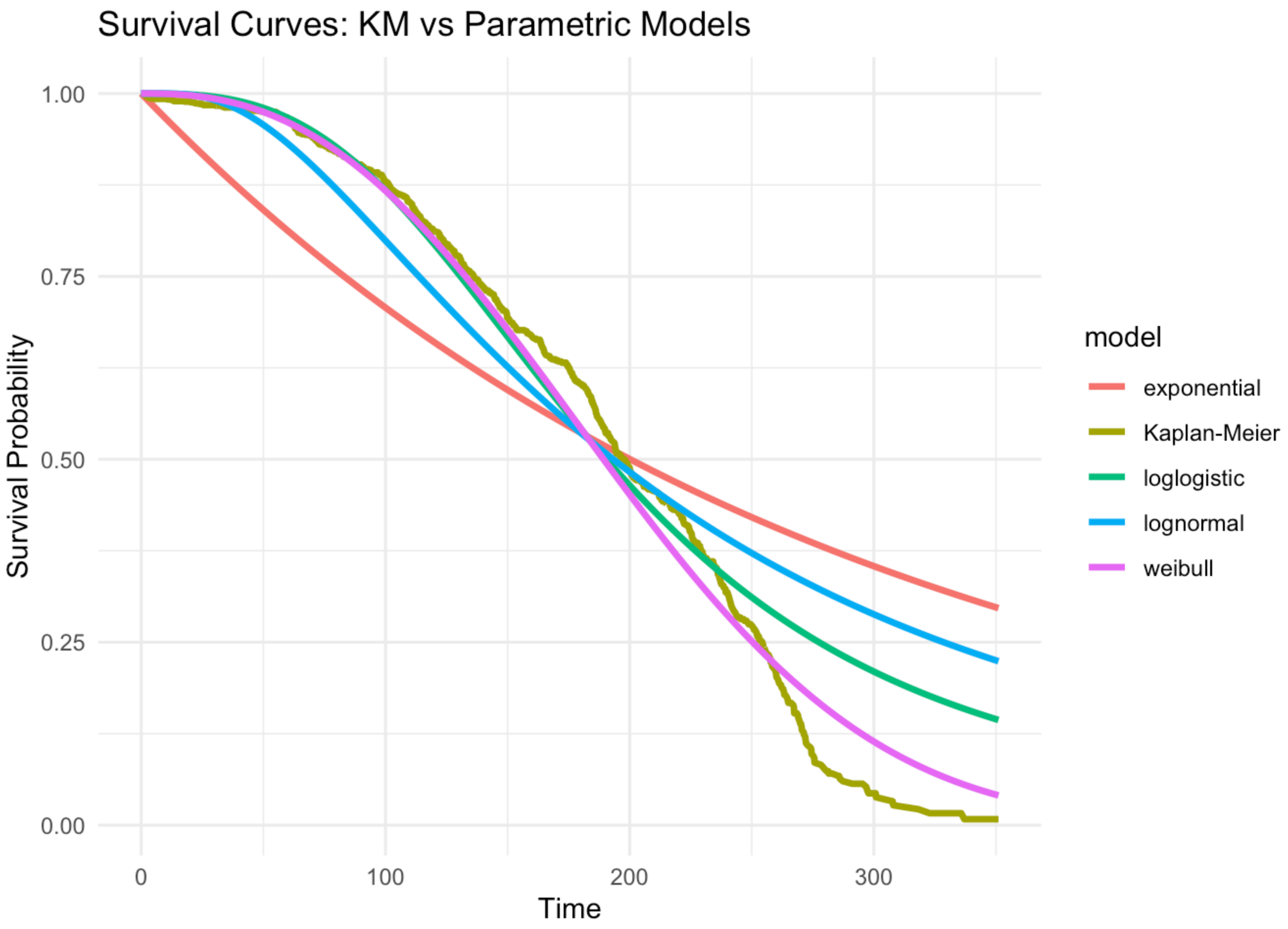
```
                  df      AIC
model_lognormal   13 7724.989
model_weibull     13 7283.647
model_exponential 12 7916.010
model_loglogistic 13 7420.526
```

```
sapply(list(lognormal = model_lognormal,
            weibull = model_weibull,
            exponential = model_exponential,
            loglogistic = model_loglogistic), logLik)
```

```
  lognormal     weibull exponential loglogistic 
  -3849.494   -3628.823   -3946.005   -3697.263 
```

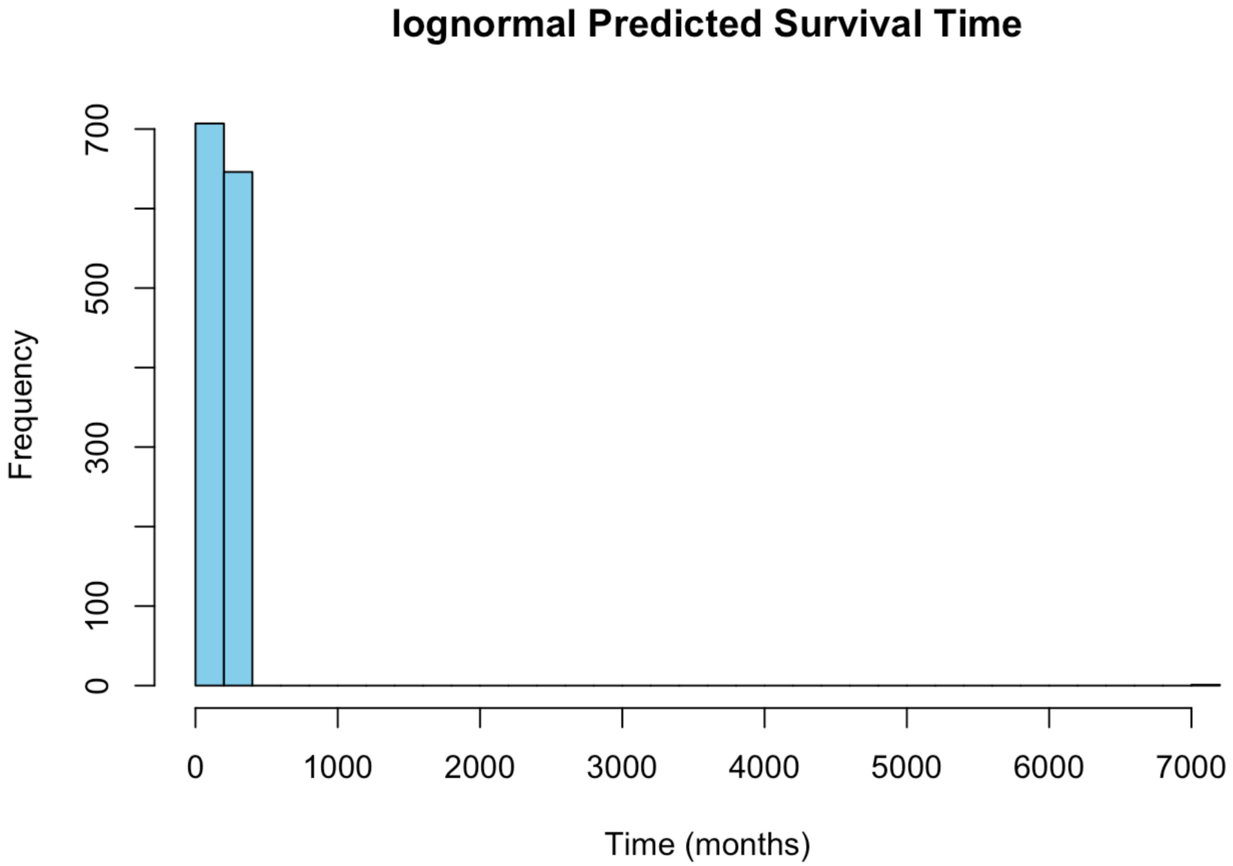
```
km_fit <- survfit(Surv(overall_survival_months, overall_survival) ~ 1, data = df_data)
fit_exp  <- flexsurvreg(Surv(overall_survival_months, overall_survival) ~ 1, data = df_da
fit_weib <- flexsurvreg(Surv(overall_survival_months, overall_survival) ~ 1, data = df_da
fit_ln   <- flexsurvreg(Surv(overall_survival_months, overall_survival) ~ 1, data = df_da
fit_ll   <- flexsurvreg(Surv(overall_survival_months, overall_survival) ~ 1, data = df_da

t_grid <- seq(0, max(df_data$overall_survival_months), length.out = 200)

df_pred <- data.frame(
  time = t_grid,
  exponential = summary(fit_exp, t = t_grid)[[1]]$est,
  weibull     = summary(fit_weib, t = t_grid)[[1]]$est,
  lognormal   = summary(fit_ln, t = t_grid)[[1]]$est,
  loglogistic = summary(fit_ll, t = t_grid)[[1]]$est
)

df_long <- tidyr::pivot_longer(df_pred, -time, names_to = "model", values_to = "surv")


km_df <- data.frame(time = km_fit$time,
                    surv = km_fit$surv,
                    model = "Kaplan-Meier")



plot_df <- rbind(df_long, km_df)

ggplot(plot_df, aes(x = time, y = surv, color = model)) +
  geom_line(size = 1.2) +
  labs(title = "Survival Curves: KM vs Parametric Models",
       x = "Time", y = "Survival Probability") +
  theme_minimal()
```

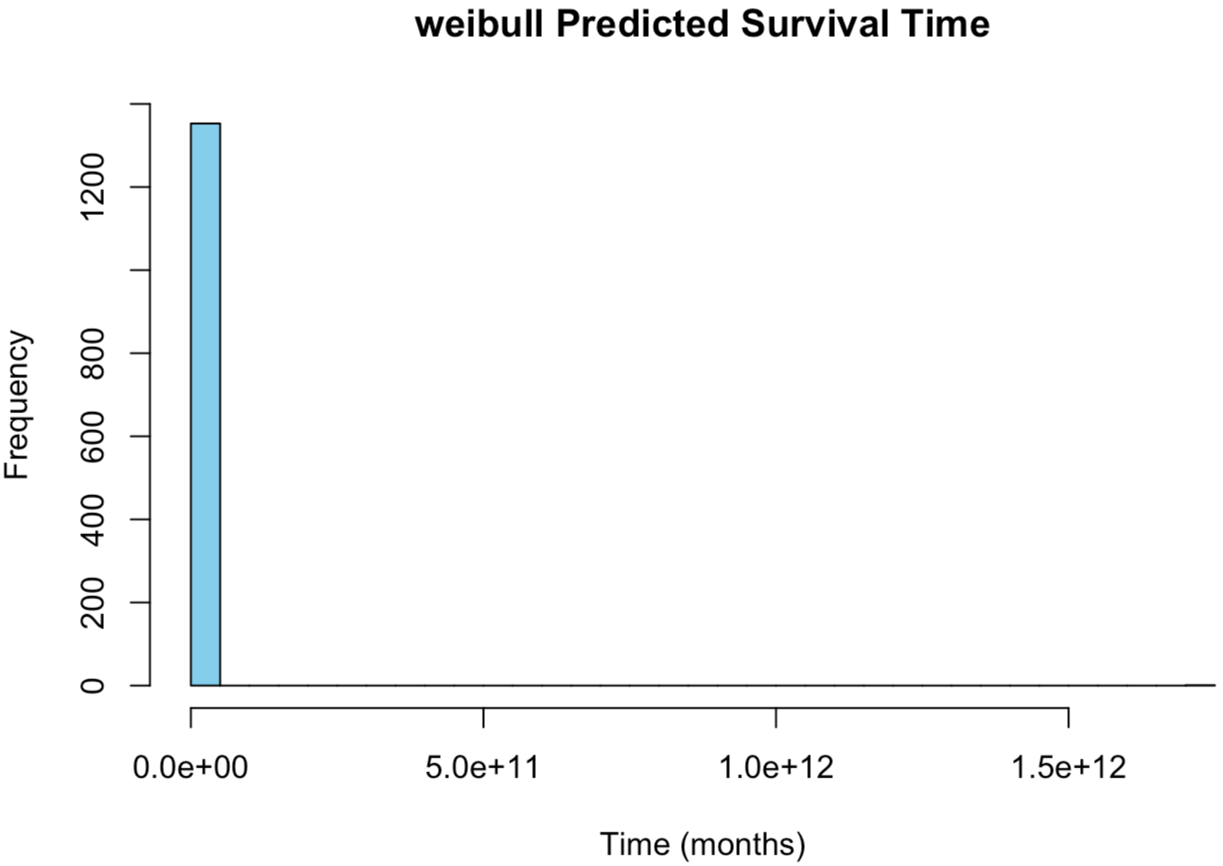
```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.
```

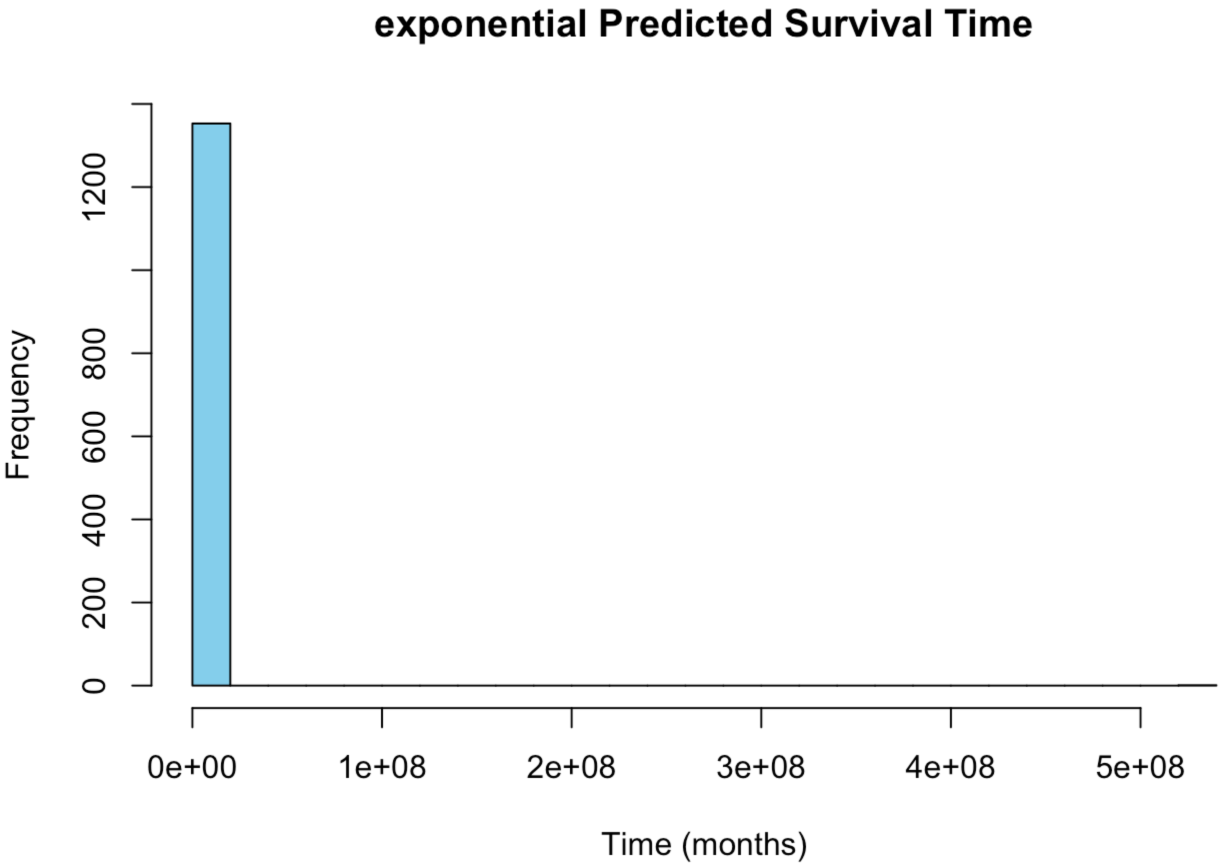


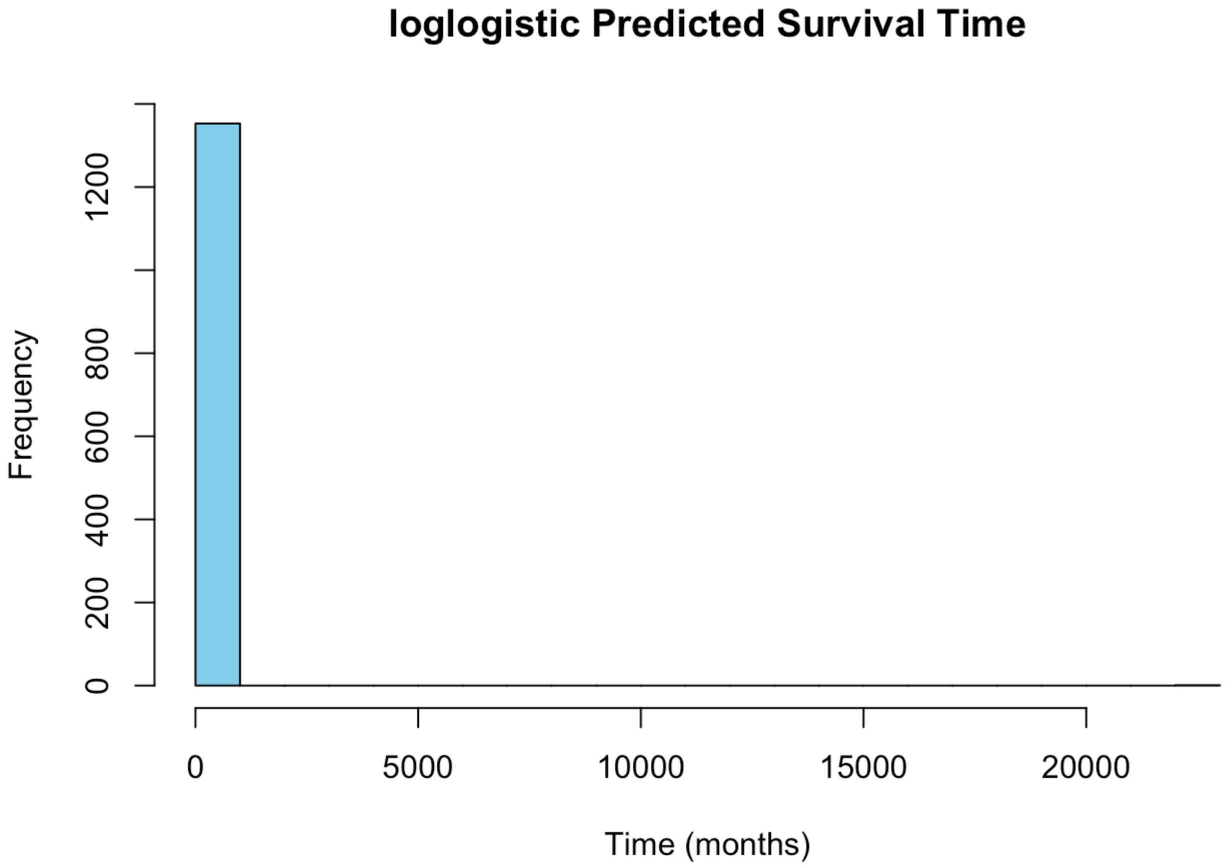
```r
pred_result <- list(lognormal = 0, weibull = 0, exponential = 0, loglogistic = 0)
for (name in names(pred_result)) {
  model_obj <- get(paste0("model_", name))
  model_pred <- predict(model_obj, type = "response")
  pred_result[[name]] <- model_pred
}
```

```r
for(name in names(pred_result)){
  hist(pred_result[[name]], main = paste(name,"Predicted Survival Time"),
     xlab = "Time (months)",col = "skyblue" ,breaks = 30)
}
```


lognormal Predicted Survival Time
Frequency
0
100
300
500
700
0
1000
2000
3000
4000
5000
6000
7000
Time (months)

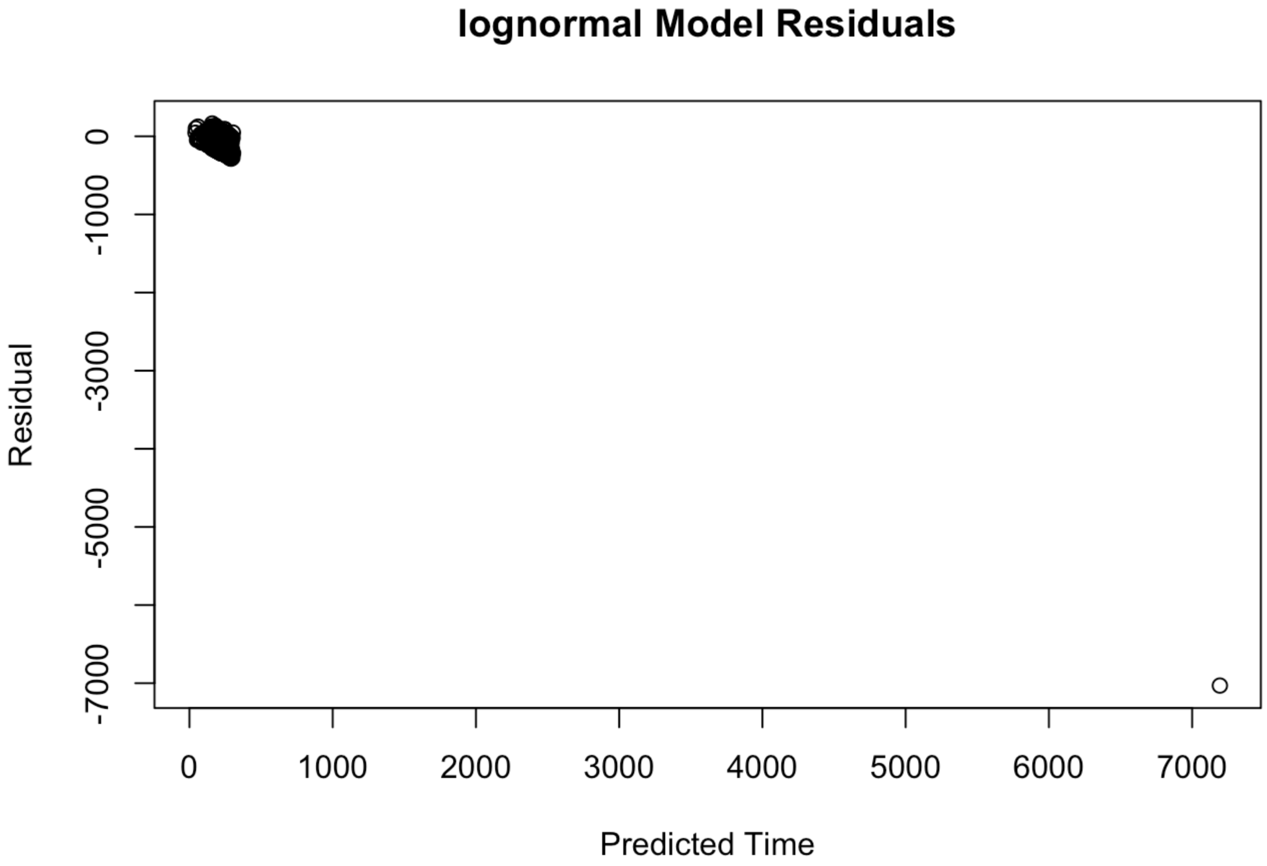

weibull Predicted Survival Time
Frequency
0
200
400
600
800
1200
0.0e+00
5.0e+11
1.0e+12
1.5e+12
Time (months)

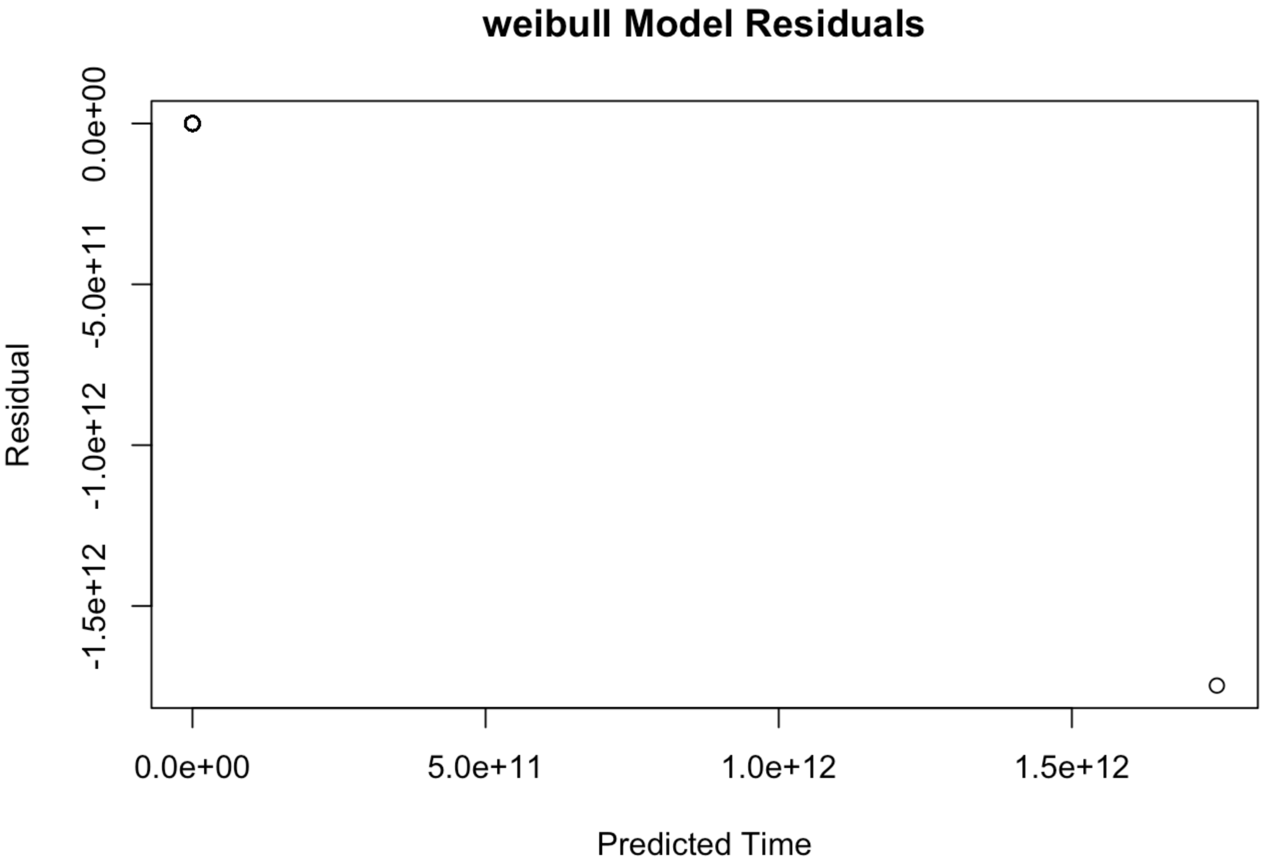

exponential Predicted Survival Time
Frequency
0
200
400
600
800
1200
0e+00
1e+08
2e+08
3e+08
4e+08
5e+08
Time (months)

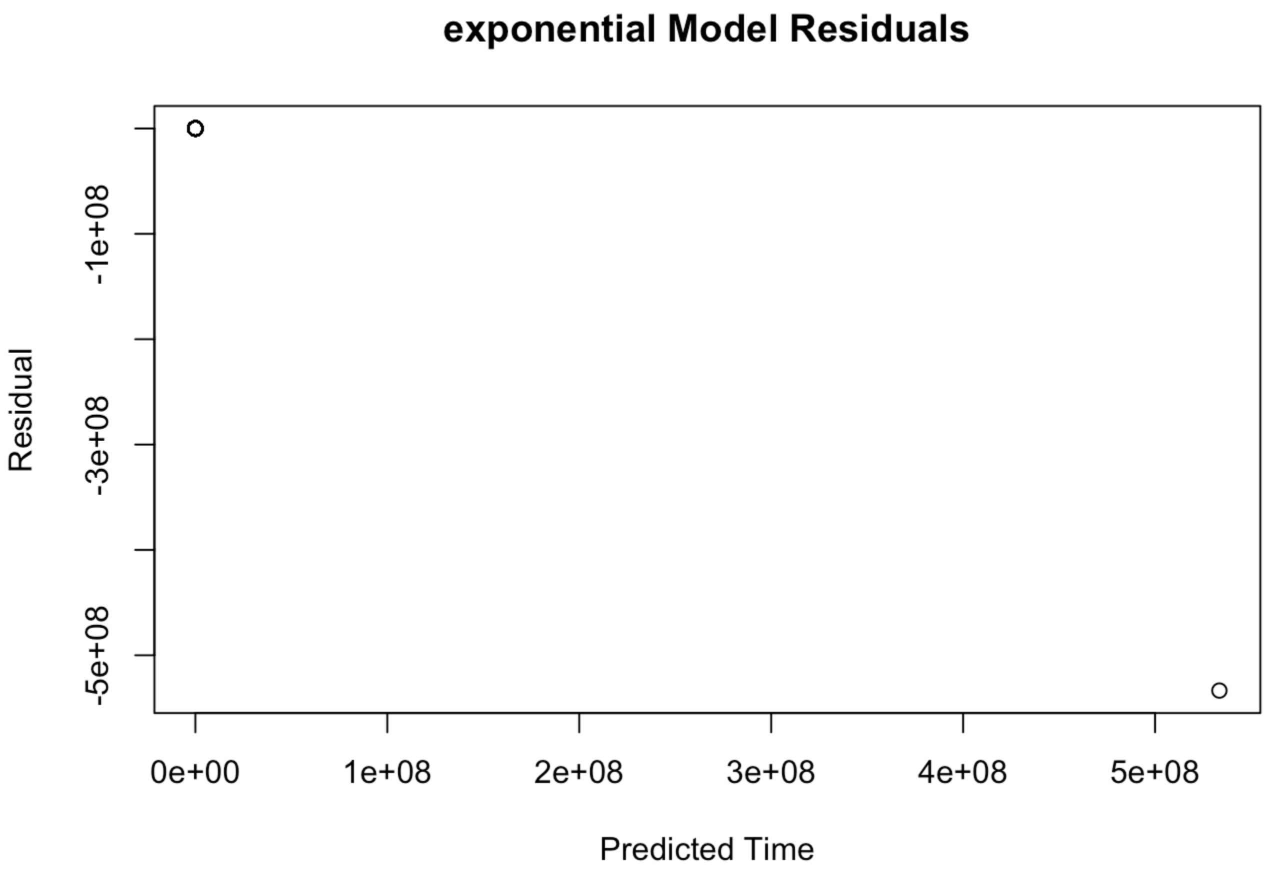


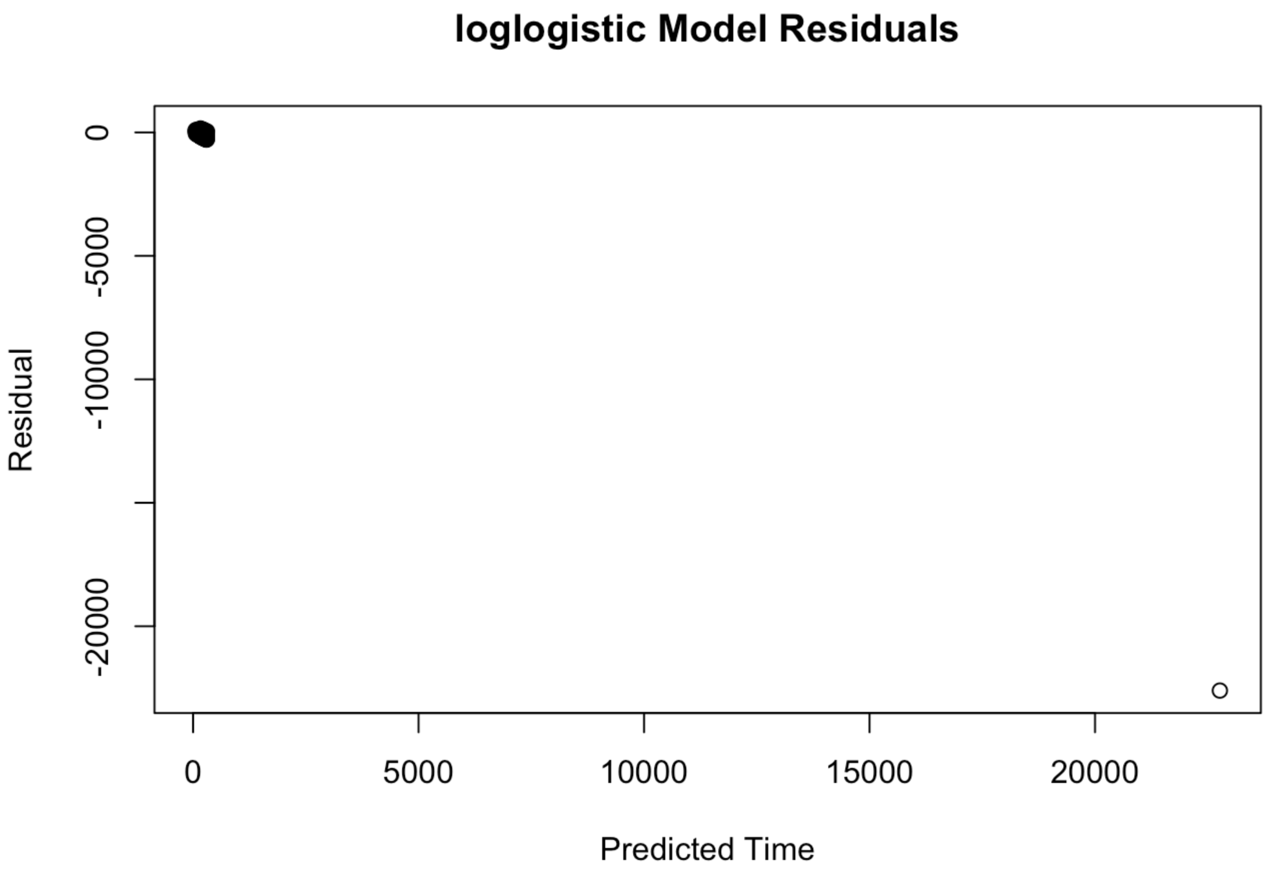
```r
resid_result <- list(lognormal = 0, weibull = 0, exponential = 0, loglogistic = 0)
for (name in names(resid_result)) {
  model_obj <- get(paste0("model_", name))
  model_resid <- resid(model_obj, type = "response")
  resid_result[[name]] <- model_resid
}
```

```r
for(name in names(resid_result)){
  plot(pred_result[[name]],resid_result[[name]],
       xlab = "Predicted Time", ylab = "Residual",
       main = paste(name,"Model Residuals"))
}
```

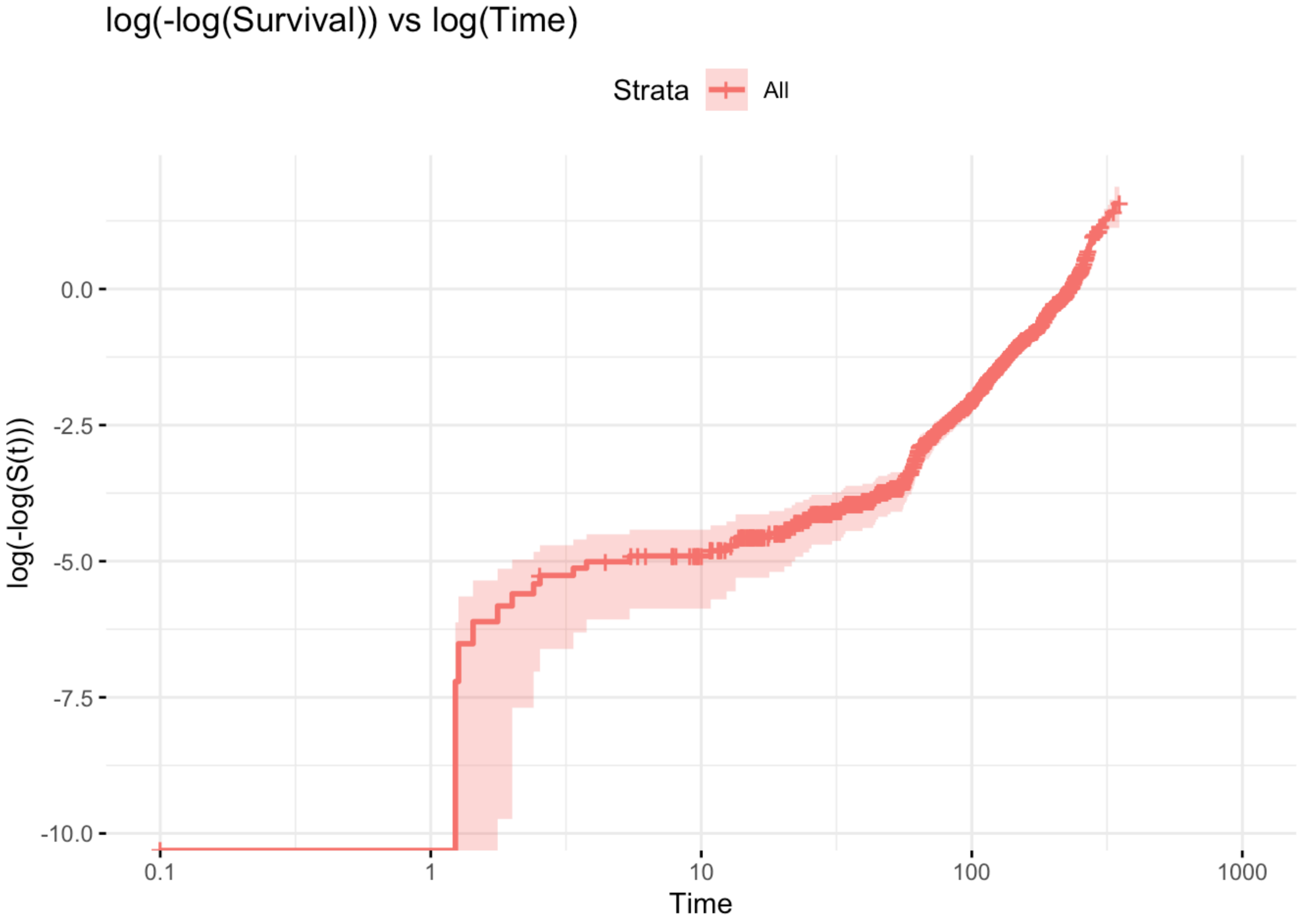

lognormal Model Residuals
Residual
0
-1000
-3000
-5000
-7000
Predicted Time
0
1000
2000
3000
4000
5000
6000
7000


weibull Model Residuals
Residual
0.0e+00
-5.0e+11
-1.0e+12
-1.5e+12
0.0e+00
5.0e+11
1.0e+12
1.5e+12
Predicted Time


exponential Model Residuals
Residual
-1e+08
-3e+08
-5e+08
0e+00
1e+08
2e+08
3e+08
4e+08
5e+08
Predicted Time



```
km_fit <- survfit(Surv(overall_survival_months, overall_survival) ~ 1,
                  data = df_data)
ggsurvplot(km_fit, fun = "cloglog",
           title = "log(-log(Survival)) vs log(Time)",
           ggtheme = theme_minimal())
```



```
survdiff(Surv(overall_survival_months, overall_survival) ~ tp53_mut_bin,
         data = df_model)
```

```
Call:
survdiff(formula = Surv(overall_survival_months, overall_survival) ~ 
    tp53_mut_bin, data = df_model)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
tp53_mut_bin=0 886      416      406     0.238     0.751
tp53_mut_bin=1 468      180      190     0.510     0.751

 Chisq= 0.8  on 1 degrees of freedom, p= 0.4 
```