

GPH 2338 Project

Yi Yang

2023-03-13

Data Upload

```
library(haven)
library(psych)
library(caret)
```

```
## Loading required package: ggplot2
```

```
##
```

```
## Attaching package: 'ggplot2'
```

```
## The following objects are masked from 'package:psych':
```

```
##
```

```
##      %+%, alpha
```

```
## Loading required package: lattice
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
```

```
## v tibble  3.1.8      v dplyr   1.0.10
```

```
## v tidyr   1.2.1      v stringr 1.5.0
```

```
## v readr   2.1.3      v forcats 0.5.2
```

```
## v purrr   1.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x ggplot2::%+%( ) masks psych::%+%( )
```

```
## x ggplot2::alpha( ) masks psych::alpha( )
```

```
## x dplyr::filter( ) masks stats::filter( )
```

```
## x dplyr::lag( ) masks stats::lag( )
```

```
## x purrr::lift( ) masks caret::lift( )
```

```
library(ggplot2)
```

```
library(psych)
```

```
library(pander)
```

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
library(pander)
library(readr)
library(r02pro)
library(plyr)
```

```
## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
##
## The following object is masked from 'package:purrr':
##
##   compact
```

```
library(readr)
```

Data Preparation

```
df<-readr::read_tsv("brca_metabric_clinical_data.tsv")
```

```
## Rows: 2509 Columns: 39
## -- Column specification -----
## Delimiter: "\t"
## chr (27): Study ID, Patient ID, Sample ID, Type of Breast Surgery, Cancer Ty...
## dbl (12): Age at Diagnosis, Cohort, Neoplasm Histologic Grade, Lymph nodes e...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
colnames(df)
```

```
## [1] "Study ID"           "Patient ID"
## [3] "Sample ID"          "Age at Diagnosis"
## [5] "Type of Breast Surgery" "Cancer Type"
## [7] "Cancer Type Detailed" "Cellularity"
## [9] "Chemotherapy"       "Pam50 + Claudin-low subtype"
## [11] "Cohort"             "ER status measured by IHC"
## [13] "ER Status"          "Neoplasm Histologic Grade"
## [15] "HER2 status measured by SNP6" "HER2 Status"
## [17] "Tumor Other Histologic Subtype" "Hormone Therapy"
## [19] "Inferred Menopausal State" "Integrative Cluster"
## [21] "Primary Tumor Laterality" "Lymph nodes examined positive"
## [23] "Mutation Count"      "Nottingham prognostic index"
```

```
## [25] "Oncotree Code"           "Overall Survival (Months)"
## [27] "Overall Survival Status"  "PR Status"
## [29] "Radio Therapy"           "Relapse Free Status (Months)"
## [31] "Relapse Free Status"     "Number of Samples Per Patient"
## [33] "Sample Type"             "Sex"
## [35] "3-Gene classifier subtype" "TMB (nonsynonymous)"
## [37] "Tumor Size"              "Tumor Stage"
## [39] "Patient's Vital Status"
```

```
colSums(is.na(df))
```

```
##           Study ID           Patient ID
##           0           0
##           Sample ID       Age at Diagnosis
##           0           11
##           Type of Breast Surgery       Cancer Type
##           554           0
##           Cancer Type Detailed       Cellularity
##           0           592
##           Chemotherapy Pam50 + Claudin-low subtype
##           529           529
##           Cohort       ER status measured by IHC
##           11           83
##           ER Status       Neoplasm Histologic Grade
##           40           121
##           HER2 status measured by SNP6       HER2 Status
##           529           529
##           Tumor Other Histologic Subtype       Hormone Therapy
##           135           529
##           Inferred Menopausal State       Integrative Cluster
##           529           529
##           Primary Tumor Laterality Lymph nodes examined positive
##           639           266
##           Mutation Count       Nottingham prognostic index
##           151           222
##           Oncotree Code       Overall Survival (Months)
##           0           528
##           Overall Survival Status       PR Status
##           528           529
##           Radio Therapy       Relapse Free Status (Months)
##           529           121
##           Relapse Free Status       Number of Samples Per Patient
##           21           0
##           Sample Type       Sex
##           0           0
##           3-Gene classifier subtype       TMB (nonsynonymous)
##           745           0
##           Tumor Size       Tumor Stage
##           149           721
##           Patient's Vital Status
##           529
```

```
print(df)
```

```
## # A tibble: 2,509 x 39
##   'Study ID'   Patie~1 Sampl~2 Age a~3 Type ~4 Cance~5 Cance~6 Cellu~7 Chemo~8
##   <chr>        <chr>   <chr>      <dbl> <chr>   <chr>   <chr>   <chr>   <chr>
## 1 brca_metabric MB-0000 MB-0000    75.6 MASTEC~ Breast~ Breast~ <NA>    NO
## 2 brca_metabric MB-0002 MB-0002    43.2 BREAST~ Breast~ Breast~ High    NO
## 3 brca_metabric MB-0005 MB-0005    48.9 MASTEC~ Breast~ Breast~ High    YES
## 4 brca_metabric MB-0006 MB-0006    47.7 MASTEC~ Breast~ Breast~ Modera~ YES
## 5 brca_metabric MB-0008 MB-0008    77.0 MASTEC~ Breast~ Breast~ High    YES
## 6 brca_metabric MB-0010 MB-0010    78.8 MASTEC~ Breast~ Breast~ Modera~ NO
## 7 brca_metabric MB-0014 MB-0014    56.4 BREAST~ Breast~ Breast~ Modera~ YES
## 8 brca_metabric MB-0020 MB-0020     70  MASTEC~ Breast~ Breast~ High    YES
## 9 brca_metabric MB-0022 MB-0022    89.1 BREAST~ Breast~ Breast~ Modera~ NO
## 10 brca_metabric MB-0025 MB-0025    76.2 <NA>   Breast~ Breast~ <NA>    <NA>
## # ... with 2,499 more rows, 30 more variables:
## #   'Pam50 + Claudin-low subtype' <chr>, Cohort <dbl>,
## #   'ER status measured by IHC' <chr>, 'ER Status' <chr>,
## #   'Neoplasm Histologic Grade' <dbl>, 'HER2 status measured by SNP6' <chr>,
## #   'HER2 Status' <chr>, 'Tumor Other Histologic Subtype' <chr>,
## #   'Hormone Therapy' <chr>, 'Inferred Menopausal State' <chr>,
## #   'Integrative Cluster' <chr>, 'Primary Tumor Laterality' <chr>, ...
```

```
df1 <- df %>% na.omit()
colSums(is.na(df1))
```

```
##           Study ID           Patient ID
##           0              0
##      Sample ID      Age at Diagnosis
##           0              0
##   Type of Breast Surgery      Cancer Type
##           0              0
##      Cancer Type Detailed      Cellularity
##           0              0
##      Chemotherapy      Pam50 + Claudin-low subtype
##           0              0
##           Cohort      ER status measured by IHC
##           0              0
##      ER Status      Neoplasm Histologic Grade
##           0              0
##   HER2 status measured by SNP6      HER2 Status
##           0              0
##   Tumor Other Histologic Subtype      Hormone Therapy
##           0              0
##   Inferred Menopausal State      Integrative Cluster
##           0              0
##   Primary Tumor Laterality      Lymph nodes examined positive
##           0              0
##      Mutation Count      Nottingham prognostic index
##           0              0
##      Oncotree Code      Overall Survival (Months)
##           0              0
##   Overall Survival Status      PR Status
```

```
##          0          0
##          Radio Therapy   Relapse Free Status (Months)
##          0          0
##          Relapse Free Status   Number of Samples Per Patient
##          0          0
##          Sample Type          Sex
##          0          0
##          3-Gene classifier subtype          TMB (nonsynonymous)
##          0          0
##          Tumor Size          Tumor Stage
##          0          0
##          Patient's Vital Status
##          0
```

```
df1<- df1[df1$`Cancer Type Detailed` == "Breast Invasive Ductal Carcinoma",]
df2<-df1[,-c(1, 2, 3, 6)] # remove StudyID, Patient ID, Sample ID, Cancer Type
print(df2)
```

```
## # A tibble: 859 x 35
##   Age at Diagn~1 Type ~2 Cance~3 Cellu~4 Chemo~5 Pam50~6 Cohort ER st~7 ER St~8
##   <dbl> <chr>   <chr>   <chr>   <chr>   <chr>   <dbl> <chr>   <chr>
## 1      43.2 BREAST~ Breast~ High    NO      LumA      1 Positive Positi~
## 2      78.8 MASTEC~ Breast~ Modera~ NO      LumB      1 Positive Positi~
## 3      86.4 BREAST~ Breast~ Modera~ NO      LumB      1 Positive Positi~
## 4      85.5 MASTEC~ Breast~ Modera~ NO      LumA      1 Positive Positi~
## 5      45.4 BREAST~ Breast~ High    YES     LumB      1 Positive Positi~
## 6      61.5 BREAST~ Breast~ High    NO      LumB      1 Positive Positi~
## 7      68.7 MASTEC~ Breast~ Low     YES     Basal     1 Negati~ Negati~
## 8      49.9 MASTEC~ Breast~ Modera~ YES     LumA      1 Positive Positi~
## 9      54.2 MASTEC~ Breast~ High    NO      LumA      1 Positive Positi~
## 10     48.6 MASTEC~ Breast~ Low     NO      LumA      1 Positive Positi~
## # ... with 849 more rows, 26 more variables: 'Neoplasm Histologic Grade' <dbl>,
## # 'HER2 status measured by SNP6' <chr>, 'HER2 Status' <chr>,
## # 'Tumor Other Histologic Subtype' <chr>, 'Hormone Therapy' <chr>,
## # 'Inferred Menopausal State' <chr>, 'Integrative Cluster' <chr>,
## # 'Primary Tumor Laterality' <chr>, 'Lymph nodes examined positive' <dbl>,
## # 'Mutation Count' <dbl>, 'Nottingham prognostic index' <dbl>,
## # 'Oncotree Code' <chr>, 'Overall Survival (Months)' <dbl>, ...
```

```
dim(df2)
```

```
## [1] 859 35
```

```
is_categorical <- sapply(df2, is.character)
is_categorical
```

```
##          Age at Diagnosis          Type of Breast Surgery
##          FALSE                      TRUE
##          Cancer Type Detailed          Cellularity
##          TRUE                      TRUE
##          Chemotherapy          Pam50 + Claudin-low subtype
```

```

##                TRUE                TRUE
##                Cohort                ER status measured by IHC
##                FALSE                TRUE
##                ER Status                Neoplasm Histologic Grade
##                TRUE                FALSE
##    HER2 status measured by SNP6                HER2 Status
##                TRUE                TRUE
## Tumor Other Histologic Subtype                Hormone Therapy
##                TRUE                TRUE
##    Inferred Menopausal State                Integrative Cluster
##                TRUE                TRUE
##    Primary Tumor Laterality    Lymph nodes examined positive
##                TRUE                FALSE
##                Mutation Count    Nottingham prognostic index
##                FALSE                FALSE
##                Oncotree Code    Overall Survival (Months)
##                TRUE                FALSE
##    Overall Survival Status                PR Status
##                TRUE                TRUE
##                Radio Therapy    Relapse Free Status (Months)
##                TRUE                FALSE
##    Relapse Free Status    Number of Samples Per Patient
##                TRUE                FALSE
##                Sample Type                Sex
##                TRUE                TRUE
##    3-Gene classifier subtype                TMB (nonsynonymous)
##                TRUE                FALSE
##                Tumor Size                Tumor Stage
##                FALSE                FALSE
##    Patient's Vital Status
##                TRUE

```

Create Dummy Variable for Categorical Variable

```

df2_1 <- df2 %>%
  mutate(Type_of_Breast_Surgery_BREAST_CONSERVING =
    as.integer(df2$`Type of Breast Surgery` == "BREAST CONSERVING"))
df2_2 <- df2_1 %>%
  mutate(Type_of_Breast_Surgery_MASTECTOMY =
    as.integer(df2$`Type of Breast Surgery` == "MASTECTOMY"))

df2_3 <- df2_2 %>%
  mutate(Cellularity_High = as.integer(df2_2$Cellularity == "High"))
df2_4 <- df2_3 %>%
  mutate(Cellularity_Moderate = as.integer(df2_3$Cellularity == "Moderate"))
df2_5 <- df2_4 %>%
  mutate(Cellularity_Low = as.integer(df2_4$Cellularity == "Low"))

df2_6 <- df2_5 %>%
  mutate(Chemotherapy_No = as.integer(df2_5$Chemotherapy == "NO"))
df2_7 <- df2_6 %>%
  mutate(Chemotherapy_Yes = as.integer(df2_6$Chemotherapy == "Yes"))

df2_8 <- df2_7 %>%

```

```

mutate(Pam50_Claudin_low_subtype_LumA =
  as.integer(df2_7$`Pam50 + Claudin-low subtype` == "LumA"))
df2_9 <- df2_8 %>%
  mutate(Pam50_Claudin_low_subtype_LumB =
    as.integer(df2_8$`Pam50 + Claudin-low subtype` == "LumB"))
df2_10 <- df2_9 %>%
  mutate(Pam50_Claudin_low_subtype_Her2 =
    as.integer(df2_9$`Pam50 + Claudin-low subtype` == "Her2"))
df2_11 <- df2_10 %>%
  mutate(Pam50_Claudin_low_subtype_Basal =
    as.integer(df2_10$`Pam50 + Claudin-low subtype` == "Basal"))
df2_12 <- df2_11 %>%
  mutate(Pam50_Claudin_low_subtype_Normal =
    as.integer(df2_11$`Pam50 + Claudin-low subtype` == "Normal"))
df2_13 <- df2_12 %>%
  mutate(Pam50_Claudin_low_subtype_claudin_low =
    as.integer(df2_12$`Pam50 + Claudin-low subtype` == "claudin-low"))
df2_14 <- df2_13 %>%
  mutate(Pam50_Claudin_low_subtype_NC =
    as.integer(df2_13$`Pam50 + Claudin-low subtype` == "NC"))

df2_15 <- df2_14 %>%
  mutate(ER_status_measured_by_IHC_Positive =
    as.integer(df2_14$`ER status measured by IHC` == "Positive"))
df2_16 <- df2_15 %>%
  mutate(ER_status_measured_by_IHC_Negative =
    as.integer(df2_15$`ER status measured by IHC` == "Negative"))
df2_17 <- df2_16 %>%
  mutate(ER_Status_Positive = as.integer(df2_16$`ER Status` == "Positive"))
df2_18 <- df2_17 %>%
  mutate(ER_Status_Negative = as.integer(df2_17$`ER Status` == "Negative"))

df2_19 <- df2_18 %>%
  mutate(HER2_status_measured_by_SNP6_Neutral =
    as.integer(df2_18$`HER2 status measured by SNP6` == "NEUTRAL"))
df2_20 <- df2_19 %>%
  mutate(HER2_status_measured_by_SNP6_GAIN =
    as.integer(df2_19$`HER2 status measured by SNP6` == "GAIN"))
df2_21 <- df2_20 %>%
  mutate(HER2_status_measured_by_SNP6_LOSS =
    as.integer(df2_20$`HER2 status measured by SNP6` == "LOSS"))
df2_22 <- df2_21 %>%
  mutate(HER2_status_measured_by_SNP6_UNDEF =
    as.integer(df2_21$`HER2 status measured by SNP6` == "UNDEF"))

df2_23 <- df2_22 %>%
  mutate(HER2_Status_Positive = as.integer(df2_22$`HER2 Status` == "Positive"))
df2_24 <- df2_23 %>%
  mutate(HER2_Status_Negative = as.integer(df2_23$`HER2 Status` == "Negative"))

df2_25 <- df2_24 %>%

```

```

mutate(Tumor_Other_Histologic_Subtype_DN=
  as.integer(df2_24$`Tumor Other Histologic Subtype` == "Ductal/NST"))
df2_26 <- df2_25 %>%
  mutate(Tumor_Other_Histologic_Subtype_Mixed =
    as.integer(df2_25$`Tumor Other Histologic Subtype` == "Mixed"))
df2_27 <- df2_26 %>%
  mutate(Tumor_Other_Histologic_Subtype_Lobular =
    as.integer(df2_26$`Tumor Other Histologic Subtype` == "Lobular"))
df2_28 <- df2_27 %>%
  mutate(Tumor_Other_Histologic_Subtype_TC =
    as.integer(df2_27$`Tumor Other Histologic Subtype` == "Tubular/ cribriform"))
df2_29 <- df2_28 %>%
  mutate(Tumor_Other_Histologic_Subtype_Mucinous =
    as.integer(df2_28$`Tumor Other Histologic Subtype` == "Mucinous"))
df2_30 <- df2_29 %>%
  mutate(Tumor_Other_Histologic_Subtype_Medullary =
    as.integer(df2_29$`Tumor Other Histologic Subtype` == "Medullary"))
df2_31 <- df2_30 %>%
  mutate(Tumor_Other_Histologic_Subtype_Other =
    as.integer(df2_30$`Tumor Other Histologic Subtype` == "Other"))

df2_32 <- df2_31 %>%
  mutate(Hormone_Therapy_No = as.integer(df2_31$`Hormone Therapy` == "NO"))
df2_33 <- df2_32 %>%
  mutate(Hormone_Therapy_Yes = as.integer(df2_32$`Hormone Therapy` == "Yes"))

df2_34 <- df2_33 %>%
  mutate(Inferred_Menopausal_State_Pre =
    as.integer(df2_33$`Inferred Menopausal State` == "Pre"))
df2_35 <- df2_34 %>%
  mutate(Inferred_Menopausal_State_Post =
    as.integer(df2_34$`Inferred Menopausal State` == "Post"))

df2_36 <- df2_35 %>%
  mutate(Primary_Tumor_Laterality_Right =
    as.integer(df2_35$`Primary Tumor Laterality` == "Right"))
df2_37 <- df2_36 %>%
  mutate(Primary_Tumor_Laterality_left =
    as.integer(df2_36$`Primary Tumor Laterality` == "Left"))

df2_38 <- df2_37 %>%
  mutate(Oncotree_Code_IDC = as.integer(df2_37$`Oncotree Code` == "IDC"))
df2_39 <- df2_38 %>%
  mutate(Oncotree_Code_MDLC = as.integer(df2_38$`Oncotree Code` == "MDLC"))
df2_40 <- df2_39 %>%
  mutate(Oncotree_Code_ILC = as.integer(df2_39$`Oncotree Code` == "ILC"))
df2_41 <- df2_40 %>%
  mutate(Oncotree_Code_IMMC = as.integer(df2_40$`Oncotree Code` == "IMMC"))
df2_42 <- df2_41 %>%
  mutate(Oncotree_Code_BREAST = as.integer(df2_41$`Oncotree Code` == "BREAST"))

df2_43 <- df2_42 %>%
  mutate(Overall_Survival_Status_0 =

```



```

      as.integer(df2_42$`Overall Survival Status` == "0:LIVING"))
df2_44 <- df2_43 %>%
  mutate(Overall_Survival_Status_1 =
    as.integer(df2_43$`Overall Survival Status` == "1:DECEASED"))

df2_45 <- df2_44 %>%
  mutate(PR_Status_Positive = as.integer(df2_44$`PR Status` == "Positive"))
df2_46 <- df2_45 %>%
  mutate(PR_Status_Negative = as.integer(df2_45$`PR Status` == "Negative"))

df2_47 <- df2_46 %>%
  mutate(Radio_Therapy_No = as.integer(df2_46$`Radio Therapy` == "NO"))
df2_48 <- df2_47 %>%
  mutate(Radio_Therapy_Yes = as.integer(df2_47$`Radio Therapy` == "Yes"))

df2_49 <- df2_48 %>%
  mutate(Relapse_Free_Status_0 =
    as.integer(df2_48$`Relapse Free Status` == "0:Not Recurred"))
df2_50 <- df2_49 %>%
  mutate(Relapse_Free_Status_1 =
    as.integer(df2_49$`Relapse Free Status` == "1:Recurred"))

df2_51 <- df2_50 %>%
  mutate(Gene_classifier_subtype_ERH =
    as.integer(df2_50$`3-Gene classifier subtype` == "ER+/HER2- High Prolif"))
df2_52 <- df2_51 %>%
  mutate(Gene_classifier_subtype_ERL =
    as.integer(df2_51$`3-Gene classifier subtype` == "ER+/HER2- Low Prolif"))
df2_53 <- df2_52 %>%
  mutate(Gene_classifier_subtype_ERM =
    as.integer(df2_52$`3-Gene classifier subtype` == "ER-/HER2-"))
df2_54 <- df2_53 %>%
  mutate(Gene_classifier_subtype_HER =
    as.integer(df2_53$`3-Gene classifier subtype` == "HER2+"))

df2_55 <- df2_54 %>%
  mutate(Patients_Vital_Status_L =
    as.integer(df2_54$`Patient's Vital Status` == "Living"))
df2_56 <- df2_55 %>%
  mutate(Patients_Vital_Status_D =
    as.integer(df2_55$`Patient's Vital Status` == "Died of Disease"))
df2_57 <- df2_56 %>%
  mutate(Patients_Vital_Status_D0 =
    as.integer(df2_56$`Patient's Vital Status` == "Died of Other Causes"))

df2_58 <- df2_57 %>%
  mutate(Integrative_Cluster_4p =
    as.integer(df2_57$`Integrative Cluster` == "4ER+"))
df2_59 <- df2_58 %>%
  mutate(Integrative_Cluster_9 = as.integer(df2_58$`Integrative Cluster` == "9"))
df2_60 <- df2_59 %>%
  mutate(Integrative_Cluster_7 = as.integer(df2_59$`Integrative Cluster` == "7"))

```

```

df2_61 <- df2_60 %>%
  mutate(Integrative_Cluster_3 = as.integer(df2_60$`Integrative Cluster` == "3"))
df2_62 <- df2_61 %>%
  mutate(Integrative_Cluster_10 = as.integer(df2_61$`Integrative Cluster` == "10"))
df2_63 <- df2_62 %>%
  mutate(Integrative_Cluster_8 = as.integer(df2_62$`Integrative Cluster` == "8"))
df2_64 <- df2_63 %>%
  mutate(Integrative_Cluster_6 = as.integer(df2_63$`Integrative Cluster` == "6"))
df2_65 <- df2_64 %>%
  mutate(Integrative_Cluster_1 = as.integer(df2_64$`Integrative Cluster` == "1"))
df2_66 <- df2_65 %>%
  mutate(Integrative_Cluster_2 = as.integer(df2_65$`Integrative Cluster` == "2"))
df2_67 <- df2_66 %>%
  mutate(Integrative_Cluster_5 = as.integer(df2_66$`Integrative Cluster` == "5"))
df2_69 <- df2_67 %>%
  mutate(Integrative_Cluster_4m =
    as.integer(df2_67$`Integrative Cluster` == "4ER-"))

```

Create Dummy Variable for Numeric Variable

```

df2_70 <- df2_69 %>%
  mutate(Age_at_Diagnosis = df2_69$`Age at Diagnosis`)
df2_71 <- df2_70 %>%
  mutate(Neo_Grade = df2_70$`Neoplasm Histologic Grade`)
df2_72 <- df2_71 %>%
  mutate(Lymph = df2_71$`Lymph nodes examined positive`)
df2_73 <- df2_72 %>%
  mutate(Mutation = df2_72$`Mutation Count`)
df2_74 <- df2_73 %>%
  mutate(Nottingham = df2_73$`Nottingham prognostic index`)
df2_75 <- df2_74 %>%
  mutate(Overall_Month = df2_74$`Overall Survival (Months)`)
df2_76 <- df2_75 %>%
  mutate(Relapse_Month = df2_75$`Relapse Free Status (Months)`)
df2_77 <- df2_76 %>%
  mutate(TMB = df2_76$`TMB (nonsynonymous)`)
df2_78 <- df2_77 %>%
  mutate(Tumor_Size = df2_77$`Tumor Size`)
df2_79 <- df2_78 %>%
  mutate(Tumor_Stage = df2_78$`Tumor Stage`)

```

```
df3 <- df2_79[, -c(1:35)]
```

```
head(df3)
```

```
## # A tibble: 6 x 78
##   Type_of_Brea~1 Type_~2 Cellu~3 Cellu~4 Cellu~5 Chemo~6 Chemo~7 Pam50~8 Pam50~9
##         <int>   <int>   <int>   <int>   <int>   <int>   <int>   <int>   <int>
## 1             1       0       1       0       0       1       0       1       0
## 2             0       1       0       1       0       1       0       0       1
## 3             1       0       0       1       0       1       0       0       1
## 4             0       1       0       1       0       1       0       1       0

```

```
## 5          1          0          1          0          0          0          0          0          1
## 6          1          0          1          0          0          1          0          0          1
## # ... with 69 more variables: Pam50_Claudin_low_subtype_Her2 <int>,
## #   Pam50_Claudin_low_subtype_Basal <int>,
## #   Pam50_Claudin_low_subtype_Normal <int>,
## #   Pam50_Claudin_low_subtype_claudin_low <int>,
## #   Pam50_Claudin_low_subtype_NC <int>,
## #   ER_status_measured_by_IHC_Positive <int>,
## #   ER_status_measured_by_IHC_Negative <int>, ER_Status_Positive <int>, ...
```

```
summary(df3)
```

```
## Type_of_Breast_Surgery_BREAST_CONSERVING Type_of_Breast_Surgery_MASTECTOMY
## Min.      :0.0000                      Min.      :0.0000
## 1st Qu.:0.0000                      1st Qu.:0.0000
## Median :0.0000                      Median :1.0000
## Mean    :0.4319                      Mean     :0.5681
## 3rd Qu.:1.0000                      3rd Qu.:1.0000
## Max.    :1.0000                      Max.     :1.0000
## Cellularity_High Cellularity_Moderate Cellularity_Low Chemotherapy_No
## Min.      :0.0000 Min.      :0.0000 Min.      :0.0000 Min.      :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :1.0000 Median :0.0000 Median :0.0000 Median :1.0000
## Mean    :0.5308 Mean    :0.3597 Mean    :0.1094 Mean    :0.7462
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max.    :1.0000 Max.    :1.0000 Max.    :1.0000 Max.    :1.0000
## Chemotherapy_Yes Pam50_Claudin_low_subtype_LumA Pam50_Claudin_low_subtype_LumB
## Min.      :0 Min.      :0.0000 Min.      :0.0000
## 1st Qu.:0 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0 Median :0.0000 Median :0.0000
## Mean    :0 Mean    :0.3667 Mean    :0.2608
## 3rd Qu.:0 3rd Qu.:1.0000 3rd Qu.:1.0000
## Max.    :0 Max.    :1.0000 Max.    :1.0000
## Pam50_Claudin_low_subtype_Her2 Pam50_Claudin_low_subtype_Basal
## Min.      :0.0000 Min.      :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000
## Mean    :0.1036 Mean    :0.1141
## 3rd Qu.:0.0000 3rd Qu.:0.0000
## Max.    :1.0000 Max.    :1.0000
## Pam50_Claudin_low_subtype_Normal Pam50_Claudin_low_subtype_claudin_low
## Min.      :0.00000 Min.      :0.0000
## 1st Qu.:0.00000 1st Qu.:0.0000
## Median :0.00000 Median :0.0000
## Mean    :0.05122 Mean    :0.1013
## 3rd Qu.:0.00000 3rd Qu.:0.0000
## Max.    :1.00000 Max.    :1.0000
## Pam50_Claudin_low_subtype_NC ER_status_measured_by_IHC_Positive
## Min.      :0.000000 Min.      :0
## 1st Qu.:0.000000 1st Qu.:0
## Median :0.000000 Median :0
## Mean    :0.002328 Mean    :0
## 3rd Qu.:0.000000 3rd Qu.:0
## Max.    :1.000000 Max.    :0
```

ER_status_measured_by_IHC_Negative		ER_Status_Positive	ER_Status_Negative
##	Min. :0.0000	Min. :0.0000	Min. :0.0000
##	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
##	Median :0.0000	Median :1.0000	Median :0.0000
##	Mean :0.2596	Mean :0.7404	Mean :0.2596
##	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
##	Max. :1.0000	Max. :1.0000	Max. :1.0000
HER2_status_measured_by_SNP6_Neutral		HER2_status_measured_by_SNP6_GAIN	
##	Min. :0.0000	Min. :0.0000	
##	1st Qu.:0.0000	1st Qu.:0.0000	
##	Median :1.0000	Median :0.0000	
##	Mean :0.6903	Mean :0.2515	
##	3rd Qu.:1.0000	3rd Qu.:1.0000	
##	Max. :1.0000	Max. :1.0000	
HER2_status_measured_by_SNP6_LOSS		HER2_status_measured_by_SNP6_UNDEF	
##	Min. :0.00000	Min. :0.000000	
##	1st Qu.:0.00000	1st Qu.:0.000000	
##	Median :0.00000	Median :0.000000	
##	Mean :0.05704	Mean :0.001164	
##	3rd Qu.:0.00000	3rd Qu.:0.000000	
##	Max. :1.00000	Max. :1.000000	
HER2_Status_Positive	HER2_Status_Negative	Tumor_Other_Histologic_Subtype_DN	
##	Min. :0.0000	Min. :0.0000	Min. :0.0000
##	1st Qu.:0.0000	1st Qu.:1.0000	1st Qu.:1.0000
##	Median :0.0000	Median :1.0000	Median :1.0000
##	Mean :0.1444	Mean :0.8556	Mean :0.9674
##	3rd Qu.:0.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
##	Max. :1.0000	Max. :1.0000	Max. :1.0000
Tumor_Other_Histologic_Subtype_Mixed		Tumor_Other_Histologic_Subtype_Lobular	
##	Min. :0	Min. :0	
##	1st Qu.:0	1st Qu.:0	
##	Median :0	Median :0	
##	Mean :0	Mean :0	
##	3rd Qu.:0	3rd Qu.:0	
##	Max. :0	Max. :0	
Tumor_Other_Histologic_Subtype_TC		Tumor_Other_Histologic_Subtype_Mucinous	
##	Min. :0.00000	Min. :0	
##	1st Qu.:0.00000	1st Qu.:0	
##	Median :0.00000	Median :0	
##	Mean :0.01746	Mean :0	
##	3rd Qu.:0.00000	3rd Qu.:0	
##	Max. :1.00000	Max. :0	
Tumor_Other_Histologic_Subtype_Medullary		Tumor_Other_Histologic_Subtype_Other	
##	Min. :0.00000	Min. :0	
##	1st Qu.:0.00000	1st Qu.:0	
##	Median :0.00000	Median :0	
##	Mean :0.01513	Mean :0	
##	3rd Qu.:0.00000	3rd Qu.:0	
##	Max. :1.00000	Max. :0	
Hormone_Therapy_No	Hormone_Therapy_Yes	Inferred_Menopausal_State_Pre	
##	Min. :0.0000	Min. :0	Min. :0.0000
##	1st Qu.:0.0000	1st Qu.:0	1st Qu.:0.0000
##	Median :0.0000	Median :0	Median :0.0000
##	Mean :0.3958	Mean :0	Mean :0.2503

```

## 3rd Qu.:1.0000      3rd Qu.:0      3rd Qu.:0.5000
## Max. :1.0000      Max. :0      Max. :1.0000
## Inferred_Menopausal_State_Post Primary_Tumor_Laterality_Right
## Min. :0.0000      Min. :0.000
## 1st Qu.:0.5000      1st Qu.:0.000
## Median :1.0000      Median :0.000
## Mean :0.7497      Mean :0.475
## 3rd Qu.:1.0000      3rd Qu.:1.000
## Max. :1.0000      Max. :1.000
## Primary_Tumor_Laterality_left Oncotree_Code_IDC Oncotree_Code_MDLC
## Min. :0.000      Min. :1      Min. :0
## 1st Qu.:0.000      1st Qu.:1      1st Qu.:0
## Median :1.000      Median :1      Median :0
## Mean :0.525      Mean :1      Mean :0
## 3rd Qu.:1.000      3rd Qu.:1      3rd Qu.:0
## Max. :1.000      Max. :1      Max. :0
## Oncotree_Code_ILC Oncotree_Code_IMMC Oncotree_Code_BREAST
## Min. :0      Min. :0      Min. :0
## 1st Qu.:0      1st Qu.:0      1st Qu.:0
## Median :0      Median :0      Median :0
## Mean :0      Mean :0      Mean :0
## 3rd Qu.:0      3rd Qu.:0      3rd Qu.:0
## Max. :0      Max. :0      Max. :0
## Overall_Survival_Status_0 Overall_Survival_Status_1 PR_Status_Positive
## Min. :0.0000      Min. :0.0000      Min. :0.0000
## 1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:0.0000
## Median :0.0000      Median :1.0000      Median :0.0000
## Mean :0.4435      Mean :0.5565      Mean :0.4994
## 3rd Qu.:1.0000      3rd Qu.:1.0000      3rd Qu.:1.0000
## Max. :1.0000      Max. :1.0000      Max. :1.0000
## PR_Status_Negative Radio_Therapy_No Radio_Therapy_Yes Relapse_Free_Status_0
## Min. :0.0000      Min. :0.0000      Min. :0      Min. :0.0000
## 1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:0      1st Qu.:0.0000
## Median :1.0000      Median :0.0000      Median :0      Median :1.0000
## Mean :0.5006      Mean :0.3201      Mean :0      Mean :0.5786
## 3rd Qu.:1.0000      3rd Qu.:1.0000      3rd Qu.:0      3rd Qu.:1.0000
## Max. :1.0000      Max. :1.0000      Max. :0      Max. :1.0000
## Relapse_Free_Status_1 Gene_classifier_subtype_ERH Gene_classifier_subtype_ERL
## Min. :0.0000      Min. :0.000      Min. :0.0000
## 1st Qu.:0.0000      1st Qu.:0.000      1st Qu.:0.0000
## Median :0.0000      Median :0.000      Median :0.0000
## Mean :0.4214      Mean :0.362      Mean :0.3132
## 3rd Qu.:1.0000      3rd Qu.:1.000      3rd Qu.:1.0000
## Max. :1.0000      Max. :1.000      Max. :1.0000
## Gene_classifier_subtype_ERM Gene_classifier_subtype_HER
## Min. :0.0000      Min. :0.0000
## 1st Qu.:0.0000      1st Qu.:0.0000
## Median :0.0000      Median :0.0000
## Mean :0.1921      Mean :0.1327
## 3rd Qu.:0.0000      3rd Qu.:0.0000
## Max. :1.0000      Max. :1.0000
## Patients_Vital_Status_L Patients_Vital_Status_D Patients_Vital_Status_DO
## Min. :0.0000      Min. :0.0000      Min. :0.0000
## 1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:0.0000

```

```
## Median :0.0000      Median :0.0000      Median :0.0000
## Mean   :0.4435      Mean    :0.3481      Mean    :0.2084
## 3rd Qu.:1.0000      3rd Qu.:1.0000      3rd Qu.:0.0000
## Max.   :1.0000      Max.    :1.0000      Max.    :1.0000
## Integrative_Cluster_4p Integrative_Cluster_9 Integrative_Cluster_7
## Min.   :0.0000      Min.    :0.00000      Min.    :0.0000
## 1st Qu.:0.0000      1st Qu.:0.00000      1st Qu.:0.0000
## Median :0.0000      Median :0.00000      Median :0.0000
## Mean   :0.1153      Mean    :0.08033      Mean    :0.1001
## 3rd Qu.:0.0000      3rd Qu.:0.00000      3rd Qu.:0.0000
## Max.   :1.0000      Max.    :1.00000      Max.    :1.0000
## Integrative_Cluster_3 Integrative_Cluster_10 Integrative_Cluster_8
## Min.   :0.0000      Min.    :0.0000      Min.    :0.0000
## 1st Qu.:0.0000      1st Qu.:0.0000      1st Qu.:0.0000
## Median :0.0000      Median :0.0000      Median :0.0000
## Mean   :0.1246      Mean    :0.1385      Mean    :0.1257
## 3rd Qu.:0.0000      3rd Qu.:0.0000      3rd Qu.:0.0000
## Max.   :1.0000      Max.    :1.0000      Max.    :1.0000
## Integrative_Cluster_6 Integrative_Cluster_1 Integrative_Cluster_2
## Min.   :0.00000      Min.    :0.00000      Min.    :0.00000
## 1st Qu.:0.00000      1st Qu.:0.00000      1st Qu.:0.00000
## Median :0.00000      Median :0.00000      Median :0.00000
## Mean   :0.04773      Mean    :0.07334      Mean    :0.03958
## 3rd Qu.:0.00000      3rd Qu.:0.00000      3rd Qu.:0.00000
## Max.   :1.00000      Max.    :1.00000      Max.    :1.00000
## Integrative_Cluster_5 Integrative_Cluster_4m Age_at_Diagnosis Neo_Grade
## Min.   :0.0000      Min.    :0.00000      Min.    :26.36      Min.    :1.000
## 1st Qu.:0.0000      1st Qu.:0.00000      1st Qu.:49.99      1st Qu.:2.000
## Median :0.0000      Median :0.00000      Median :60.62      Median :3.000
## Mean   :0.1153      Mean    :0.03958      Mean    :60.05      Mean    :2.517
## 3rd Qu.:0.0000      3rd Qu.:0.00000      3rd Qu.:69.75      3rd Qu.:3.000
## Max.   :1.0000      Max.    :1.00000      Max.    :96.29      Max.    :3.000
## Lymph      Mutation      Nottingham      Overall_Month
## Min.   : 0.000      Min.    : 1.000      Min.    :2.018      Min.    : 0.10
## 1st Qu.: 0.000      1st Qu.: 3.000      1st Qu.:3.080      1st Qu.: 58.05
## Median : 0.000      Median : 5.000      Median :4.050      Median :115.30
## Mean   : 1.916      Mean    : 5.423      Mean    :4.216      Mean    :124.05
## 3rd Qu.: 2.000      3rd Qu.: 7.000      3rd Qu.:5.050      3rd Qu.:186.32
## Max.   :41.000      Max.    :46.000      Max.    :6.360      Max.    :337.03
## Relapse_Month      TMB      Tumor_Size      Tumor_Stage
## Min.   : 0.10      Min.    : 1.308      Min.    : 1.0      Min.    :1.000
## 1st Qu.: 40.10      1st Qu.: 3.923      1st Qu.: 17.0      1st Qu.:1.000
## Median : 98.42      Median : 6.538      Median : 22.0      Median :2.000
## Mean   :109.35      Mean    : 7.072      Mean    : 25.7      Mean    :1.767
## 3rd Qu.:172.12      3rd Qu.: 9.153      3rd Qu.: 30.0      3rd Qu.:2.000
## Max.   :296.91      Max.    :60.146      Max.    :180.0      Max.    :4.000
```

```
model11 <- glm(df3$Overall_Month ~., data = df3)
summary(model11)
```

```
##
## Call:
## glm(formula = df3$Overall_Month ~ ., data = df3)
##
```

```

## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -90.281   -8.692   -1.188    6.222   141.390
##
## Coefficients: (29 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    134.70073    31.20199    4.317 1.78e-05
## Type_of_Breast_Surgery_BREAST_CONSERVING    0.11047    1.95883    0.056 0.95504
## Type_of_Breast_Surgery_MASTECTOMY            NA         NA         NA      NA
## Cellularity_High    -2.65880    2.72377   -0.976 0.32928
## Cellularity_Moderate -2.94965    2.74693   -1.074 0.28323
## Cellularity_Low      NA         NA         NA      NA
## Chemotherapy_No      5.29901    2.51119    2.110 0.03515
## Chemotherapy_Yes      NA         NA         NA      NA
## Pam50_Claudin_low_subtype_LumA    -23.65547   15.87227   -1.490 0.13652
## Pam50_Claudin_low_subtype_LumB    -24.54846   15.81912   -1.552 0.12110
## Pam50_Claudin_low_subtype_Her2    -24.18549   16.16497   -1.496 0.13500
## Pam50_Claudin_low_subtype_Basal    -25.94962   16.29142   -1.593 0.11159
## Pam50_Claudin_low_subtype_Normal    -28.92091   16.15994   -1.790 0.07388
## Pam50_Claudin_low_subtype_claudin_low    -27.28122   16.15691   -1.689 0.09170
## Pam50_Claudin_low_subtype_NC      NA         NA         NA      NA
## ER_status_measured_by_IHC_Positive      NA         NA         NA      NA
## ER_status_measured_by_IHC_Negative    -0.08680    4.05265   -0.021 0.98292
## ER_Status_Positive      4.62436    4.14155    1.117 0.26451
## ER_Status_Negative      NA         NA         NA      NA
## HER2_status_measured_by_SNP6_Neutral    1.41869   22.20971    0.064 0.94908
## HER2_status_measured_by_SNP6_GAIN      3.19252   22.29363    0.143 0.88617
## HER2_status_measured_by_SNP6_LOSS    -2.11852   22.41497   -0.095 0.92472
## HER2_status_measured_by_SNP6_UNDEF      NA         NA         NA      NA
## HER2_Status_Positive    -4.80564    4.77797   -1.006 0.31482
## HER2_Status_Negative      NA         NA         NA      NA
## Tumor_Other_Histologic_Subtype_DN    -5.47118    6.31821   -0.866 0.38678
## Tumor_Other_Histologic_Subtype_Mixed      NA         NA         NA      NA
## Tumor_Other_Histologic_Subtype_Lobular      NA         NA         NA      NA
## Tumor_Other_Histologic_Subtype_TC    -10.24649    8.71807   -1.175 0.24021
## Tumor_Other_Histologic_Subtype_Mucinous      NA         NA         NA      NA
## Tumor_Other_Histologic_Subtype_Medullary      NA         NA         NA      NA
## Tumor_Other_Histologic_Subtype_Other      NA         NA         NA      NA
## Hormone_Therapy_No      5.31230    1.99512    2.663 0.00791
## Hormone_Therapy_Yes      NA         NA         NA      NA
## Inferred_Menopausal_State_Pre    -1.39328    2.76611   -0.504 0.61461
## Inferred_Menopausal_State_Post      NA         NA         NA      NA
## Primary_Tumor_Laterality_Right      0.50774    1.53036    0.332 0.74014
## Primary_Tumor_Laterality_left      NA         NA         NA      NA
## Oncotree_Code_IDC      NA         NA         NA      NA
## Oncotree_Code_MDLC      NA         NA         NA      NA
## Oncotree_Code_ILC      NA         NA         NA      NA
## Oncotree_Code_IMMC      NA         NA         NA      NA
## Oncotree_Code_BREAST      NA         NA         NA      NA
## Overall_Survival_Status_0      4.13879    2.27208    1.822 0.06889
## Overall_Survival_Status_1      NA         NA         NA      NA
## PR_Status_Positive      1.00374    1.93534    0.519 0.60416
## PR_Status_Negative      NA         NA         NA      NA
## Radio_Therapy_No      3.28757    2.00578    1.639 0.10159

```

## Radio_Therapy_Yes	NA	NA	NA	NA
## Relapse_Free_Status_0	-84.14957	3.11108	-27.048	< 2e-16
## Relapse_Free_Status_1	NA	NA	NA	NA
## Gene_classifier_subtype_ERH	-1.20919	5.20485	-0.232	0.81635
## Gene_classifier_subtype_ERL	-5.08175	5.55668	-0.915	0.36071
## Gene_classifier_subtype_ERM	-0.93733	5.65201	-0.166	0.86832
## Gene_classifier_subtype_HER	NA	NA	NA	NA
## Patients_Vital_Status_L	NA	NA	NA	NA
## Patients_Vital_Status_D	-66.81109	3.57280	-18.700	< 2e-16
## Patients_Vital_Status_DO	NA	NA	NA	NA
## Integrative_Cluster_4p	3.40217	5.34627	0.636	0.52472
## Integrative_Cluster_9	0.09066	5.33512	0.017	0.98645
## Integrative_Cluster_7	1.60445	5.55187	0.289	0.77266
## Integrative_Cluster_3	-0.04801	5.51899	-0.009	0.99306
## Integrative_Cluster_10	2.30212	4.69278	0.491	0.62387
## Integrative_Cluster_8	3.16357	5.45142	0.580	0.56186
## Integrative_Cluster_6	4.38502	6.02859	0.727	0.46721
## Integrative_Cluster_1	2.12743	5.47486	0.389	0.69769
## Integrative_Cluster_2	-1.57031	6.17571	-0.254	0.79935
## Integrative_Cluster_5	-0.77025	6.04703	-0.127	0.89867
## Integrative_Cluster_4m	NA	NA	NA	NA
## Age_at_Diagnosis	-0.30591	0.10125	-3.021	0.00260
## Neo_Grade	1.86862	2.36510	0.790	0.42971
## Lymph	0.13387	0.28983	0.462	0.64429
## Mutation	-0.77145	5.29892	-0.146	0.88428
## Nottingham	-3.43446	1.87887	-1.828	0.06793
## Relapse_Month	0.95635	0.01270	75.317	< 2e-16
## TMB	0.85798	4.04852	0.212	0.83222
## Tumor_Size	0.08219	0.06496	1.265	0.20612
## Tumor_Stage	0.17160	1.85285	0.093	0.92623
##				
## (Intercept)	***			
## Type_of_Breast_Surgery_BREAST_CONSERVING				
## Type_of_Breast_Surgery_MASTECTOMY				
## Cellularity_High				
## Cellularity_Moderate				
## Cellularity_Low				
## Chemotherapy_No	*			
## Chemotherapy_Yes				
## Pam50_Claudin_low_subtype_LumA				
## Pam50_Claudin_low_subtype_LumB				
## Pam50_Claudin_low_subtype_Her2				
## Pam50_Claudin_low_subtype_Basal				
## Pam50_Claudin_low_subtype_Normal	.			
## Pam50_Claudin_low_subtype_claudin_low	.			
## Pam50_Claudin_low_subtype_NC				
## ER_status_measured_by_IHC_Positive				
## ER_status_measured_by_IHC_Negative				
## ER_Status_Positive				
## ER_Status_Negative				
## HER2_status_measured_by_SNP6_Neutral				
## HER2_status_measured_by_SNP6_GAIN				
## HER2_status_measured_by_SNP6_LOSS				
## HER2_status_measured_by_SNP6_UNDEF				


```

## HER2_Status_Positive
## HER2_Status_Negative
## Tumor_Other_Histologic_Subtype_DN
## Tumor_Other_Histologic_Subtype_Mixed
## Tumor_Other_Histologic_Subtype_Lobular
## Tumor_Other_Histologic_Subtype_TC
## Tumor_Other_Histologic_Subtype_Mucinous
## Tumor_Other_Histologic_Subtype_Medullary
## Tumor_Other_Histologic_Subtype_Other
## Hormone_Therapy_No                **
## Hormone_Therapy_Yes
## Inferred_Menopausal_State_Pre
## Inferred_Menopausal_State_Post
## Primary_Tumor_Laterality_Right
## Primary_Tumor_Laterality_left
## Oncotree_Code_IDC
## Oncotree_Code_MDLC
## Oncotree_Code_ILC
## Oncotree_Code_IMMC
## Oncotree_Code_BREAST
## Overall_Survival_Status_0         .
## Overall_Survival_Status_1
## PR_Status_Positive
## PR_Status_Negative
## Radio_Therapy_No
## Radio_Therapy_Yes
## Relapse_Free_Status_0             ***
## Relapse_Free_Status_1
## Gene_classifier_subtype_ERH
## Gene_classifier_subtype_ERL
## Gene_classifier_subtype_ERM
## Gene_classifier_subtype_HER
## Patients_Vital_Status_L
## Patients_Vital_Status_D           ***
## Patients_Vital_Status_DO
## Integrative_Cluster_4p
## Integrative_Cluster_9
## Integrative_Cluster_7
## Integrative_Cluster_3
## Integrative_Cluster_10
## Integrative_Cluster_8
## Integrative_Cluster_6
## Integrative_Cluster_1
## Integrative_Cluster_2
## Integrative_Cluster_5
## Integrative_Cluster_4m
## Age_at_Diagnosis                  **
## Neo_Grade
## Lymph
## Mutation
## Nottingham                        .
## Relapse_Month                     ***
## TMB
## Tumor_Size

```

```
## Tumor_Stage
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 473.9313)
##
##      Null deviance: 5026495  on 858  degrees of freedom
## Residual deviance:  383884  on 810  degrees of freedom
## AIC: 7779.6
##
## Number of Fisher Scoring iterations: 2
```

```
with(summary(model1), 1 - deviance/null.deviance)
```

```
## [1] 0.9236278
```