GPH 2338 Project

Yi Yang

2023-03-13

Data Upload

```
library(haven)
library(psych)
library(caret)
library(tidyverse)
library(ggplot2)
library(psych)
library(pander)
library(corrplot)
library(pander)
library(readr)
library(r02pro)
library(plyr)
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
```

```r
#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: 1,092 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 Positve Positi~
##  2          77.0 MASTEC~ Breast~ High    YES     LumB         1 Positve Positi~
##  3          78.8 MASTEC~ Breast~ Modera~ NO      LumB         1 Positve Positi~
##  4          86.4 BREAST~ Breast~ Modera~ NO      LumB         1 Positve Positi~
##  5          84.2 MASTEC~ Breast~ High    NO      Her2         1 Negati~ Positi~
##  6          85.5 MASTEC~ Breast~ Modera~ NO      LumA         1 Positve Positi~
##  7          45.4 BREAST~ Breast~ High    YES     LumB         1 Positve Positi~
##  8          61.5 BREAST~ Breast~ High    NO      LumB         1 Positve Positi~
##  9          68.7 MASTEC~ Breast~ Low     YES     Basal        1 Negati~ Negati~
## 10          46.9 MASTEC~ Breast~ Modera~ NO      Normal       1 Positve Positi~
## # ... with 1,082 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>, ...
```

```r
dim(df2)
```

```
## [1] 1092    35
```

```r
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 %>%
```

```r
  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$`ER status measured by IHC` == "Positve"))
df2_16 <- df2_15 %>%
  mutate(ER_status_measured_by_IHC_Negative =
           as.integer(df2_15$`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"))
```

```r
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 %>%
```

```r
  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_DO =
           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_68 <- df2_67 %>%
  mutate(Integrative_Cluster_4m =
           as.integer(df2_67$`Integrative Cluster` == "4ER-"))


df2_69 <- df2_68 %>%
  mutate(Cancer_Type_Detailed=
           df2$`Cancer Type Detailed`)
```

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)]
```

```
pander(summary(df3),caption='Descriptive Statistics of The Data')
```

Table 1: Descriptive Statistics of The Data (continued below)

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 :0.4231	Mean :0.5769
3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000

Table 2: Table continues below

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.:1.0000
Median :1.0000	Median :0.0000	Median :0.0000	Median :1.0000
Mean :0.5092	Mean :0.3773	Mean :0.1136	Mean :0.7793
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

Table 3: Table continues below

Chemotherapy_Yes	Pam50_Claudin_low_subtype_LumA
Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000
Median :0.0000	Median :0.0000
Mean :0.2207	Mean :0.4057
3rd Qu.:0.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000

Table 4: Table continues below

Pam50_Claudin_low_subtype_LumB	Pam50_Claudin_low_subtype_Her2
Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.00000
Mean :0.2491	Mean :0.09249
3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.00000

Table 5: Table continues below

Pam50_Claudin_low_subtype_Basal	Pam50_Claudin_low_subtype_Normal
Min. :0.00000	Min. :0.00000
1st Qu.:0.00000	1st Qu.:0.00000
Median :0.00000	Median :0.00000
Mean :0.09341	Mean :0.06227

Pam50_Claudin_low_subtype_Basal	Pam50_Claudin_low_subtype_Normal
3rd Qu.:0.00000	3rd Qu.:0.00000
Max. :1.00000	Max. :1.00000

Table 6: Table continues below

Pam50_Claudin_low_subtype_claudin_low	Pam50_Claudin_low_subtype_NC
Min. :0.00000	Min. :0.000000
1st Qu.:0.00000	1st Qu.:0.000000
Median :0.00000	Median :0.000000
Mean :0.09524	Mean :0.001832
3rd Qu.:0.00000	3rd Qu.:0.000000
Max. :1.00000	Max. :1.000000

Table 7: Table continues below

ER_status_measured_by_IHC_Positive	ER_status_measured_by_IHC_Negative
Min. :0.0000	Min. :0.0000
1st Qu.:1.0000	1st Qu.:0.0000
Median :1.0000	Median :0.0000
Mean :0.7747	Mean :0.2253
3rd Qu.:1.0000	3rd Qu.:0.0000
Max. :1.0000	Max. :1.0000

Table 8: Table continues below

ER_Status_Positive	ER_Status_Negative	HER2_status_measured_by_SNP6_Neutral
Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:1.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :1.0000	Median :0.0000	Median :1.0000
Mean :0.7747	Mean :0.2253	Mean :0.7262
3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000

Table 9: Table continues below

HER2_status_measured_by_SNP6_GAIN	HER2_status_measured_by_SNP6_LOSS
Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.00000
Mean :0.2207	Mean :0.05128
3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.00000

Table 10: Table continues below

HER2_status_measured_by_SNP6_UNDEF	HER2_Status_Positive
Min. :0.000000	Min. :0.0000
1st Qu.:0.000000	1st Qu.:0.0000
Median :0.000000	Median :0.0000
Mean :0.001832	Mean :0.1227
3rd Qu.:0.000000	3rd Qu.:0.0000
Max. :1.000000	Max. :1.0000

Table 11: Table continues below

HER2_Status_Negative	Tumor_Other_Histologic_Subtype_DN
Min. :0.0000	Min. :0.000
1st Qu.:1.0000	1st Qu.:1.000
Median :1.0000	Median :1.000
Mean :0.8773	Mean :0.761
3rd Qu.:1.0000	3rd Qu.:1.000
Max. :1.0000	Max. :1.000

Table 12: Table continues below

Tumor_Other_Histologic_Subtype_Mixed	Tumor_Other_Histologic_Subtype_Lobular
Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.00000
Mean :0.1264	Mean :0.07051
3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.00000

Table 13: Table continues below

Tumor_Other_Histologic_Subtype_TC	Tumor_Other_Histologic_Subtype_Mucinous
Min. :0.00000	Min. :0.0000
1st Qu.:0.00000	1st Qu.:0.0000
Median :0.00000	Median :0.0000
Mean :0.01374	Mean :0.0119
3rd Qu.:0.00000	3rd Qu.:0.0000
Max. :1.00000	Max. :1.0000

Table 14: Table continues below

Tumor_Other_Histologic_Subtype_Medullary
Min. :0.0000
1st Qu.:0.0000
Median :0.0000
Mean :0.0119

Tumor__Other__Histologic__Subtype__Medullary
3rd Qu.:0.0000
Max. :1.0000

Table 15: Table continues below

Tumor__Other__Histologic__Subtype__Other	Hormone__Therapy__No
Min. :0.000000	Min. :0.0000
1st Qu.:0.000000	1st Qu.:0.0000
Median :0.000000	Median :0.0000
Mean :0.004579	Mean :0.3864
3rd Qu.:0.000000	3rd Qu.:1.0000
Max. :1.000000	Max. :1.0000

Table 16: Table continues below

Hormone__Therapy__Yes	Inferred__Menopausal__State__Pre
Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000
Median :1.0000	Median :0.0000
Mean :0.6136	Mean :0.2299
3rd Qu.:1.0000	3rd Qu.:0.0000
Max. :1.0000	Max. :1.0000

Table 17: Table continues below

Inferred__Menopausal__State__Post	Primary__Tumor__Laterality__Right
Min. :0.0000	Min. :0.0000
1st Qu.:1.0000	1st Qu.:0.0000
Median :1.0000	Median :0.0000
Mean :0.7701	Mean :0.4817
3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000

Table 18: Table continues below

Primary__Tumor__Laterality__left	Oncotree__Code__IDC	Oncotree__Code__MDLC
Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:1.0000	1st Qu.:0.0000
Median :1.0000	Median :1.0000	Median :0.0000
Mean :0.5183	Mean :0.7866	Mean :0.1264
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:0.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000

Table 19: Table continues below

Oncotree_Code_ILC	Oncotree_Code_IMMC	Oncotree_Code_BREAST
Min. :0.00000	Min. :0.0000	Min. :0.000000
1st Qu.:0.00000	1st Qu.:0.0000	1st Qu.:0.000000
Median :0.00000	Median :0.0000	Median :0.000000
Mean :0.07051	Mean :0.0119	Mean :0.004579
3rd Qu.:0.00000	3rd Qu.:0.0000	3rd Qu.:0.000000
Max. :1.00000	Max. :1.0000	Max. :1.000000

Table 20: Table continues below

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 :1.0000
Mean :0.4432	Mean :0.5568	Mean :0.5229
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000

Table 21: Table continues below

PR_Status_Negative	Radio_Therapy_No	Radio_Therapy_Yes
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 :1.0000
Mean :0.4771	Mean :0.3361	Mean :0.6639
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000

Table 22: Table continues below

Relapse_Free_Status_0	Relapse_Free_Status_1	Gene_classifier_subtype_ERH
Min. :0.0000	Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.0000
Median :1.0000	Median :0.0000	Median :0.0000
Mean :0.5833	Mean :0.4167	Mean :0.3535
3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000

Table 23: Table continues below

Gene_classifier_subtype_ERL	Gene_classifier_subtype_ERM
Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000
Median :0.0000	Median :0.0000
Mean :0.3663	Mean :0.1694

Gene_classifier_subtype_ERL	Gene_classifier_subtype_ERM
3rd Qu.:1.0000	3rd Qu.:0.0000
Max. :1.0000	Max. :1.0000

Table 24: Table continues below

Gene_classifier_subtype_HER	Patients_Vital_Status_L
Min. :0.0000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.0000
Median :0.0000	Median :0.0000
Mean :0.1108	Mean :0.4432
3rd Qu.:0.0000	3rd Qu.:1.0000
Max. :1.0000	Max. :1.0000

Table 25: Table continues below

Patients_Vital_Status_D	Patients_Vital_Status_DO	Integrative_Cluster_4p
Min. :0.0000	Min. :0.0000	Min. :0.00
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00
Median :0.0000	Median :0.0000	Median :0.00
Mean :0.3388	Mean :0.2179	Mean :0.13
3rd Qu.:1.0000	3rd Qu.:0.0000	3rd Qu.:0.00
Max. :1.0000	Max. :1.0000	Max. :1.00

Table 26: Table continues below

Integrative_Cluster_9	Integrative_Cluster_7	Integrative_Cluster_3
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 :0.0696	Mean :0.1007	Mean :0.1566
3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.0000
Max. :1.0000	Max. :1.0000	Max. :1.0000

Table 27: Table continues below

Integrative_Cluster_10	Integrative_Cluster_8	Integrative_Cluster_6
Min. :0.0000	Min. :0.0000	Min. :0.00000
1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000
Median :0.0000	Median :0.0000	Median :0.00000
Mean :0.1126	Mean :0.1346	Mean :0.04853
3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :1.0000	Max. :1.0000	Max. :1.00000

Table 28: Table continues below

Integrative_Cluster_1	Integrative_Cluster_2	Integrative_Cluster_5
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 :0.07051	Mean :0.04212	Mean :0.09799
3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000
Max. :1.00000	Max. :1.00000	Max. :1.00000

Table 29: Table continues below

Integrative_Cluster_4m	Cancer_Type_Detailed	Age_at_Diagnosis
Min. :0.00000	Length:1092	Min. :21.93
1st Qu.:0.00000	Class :character	1st Qu.:50.79
Median :0.00000	Mode :character	Median :60.99
Mean :0.03663	NA	Mean :60.39
3rd Qu.:0.00000	NA	3rd Qu.:69.70
Max. :1.00000	NA	Max. :96.29

Table 30: Table continues below

Neo_Grade	Lymph	Mutation	Nottingham
Min. :1.000	Min. : 0.000	Min. : 1.000	Min. :2.002
1st Qu.:2.000	1st Qu.: 0.000	1st Qu.: 3.000	1st Qu.:3.052
Median :3.000	Median : 0.000	Median : 5.000	Median :4.046
Mean :2.446	Mean : 1.889	Mean : 5.484	Mean :4.129
3rd Qu.:3.000	3rd Qu.: 2.000	3rd Qu.: 7.000	3rd Qu.:5.046
Max. :3.000	Max. :41.000	Max. :46.000	Max. :6.360

Overall_Month	Relapse_Month	TMB	Tumor_Size	Tumor_Stage
Min. : 0.10	Min. : 0.00	Min. : 1.308	Min. : 1.00	Min. :1.000
1st Qu.: 60.83	1st Qu.: 40.88	1st Qu.: 3.923	1st Qu.: 17.00	1st Qu.:1.000
Median :116.98	Median : 99.82	Median : 6.538	Median : 22.00	Median :2.000
Mean :126.55	Mean :111.23	Mean : 7.153	Mean : 26.03	Mean :1.759
3rd Qu.:188.43	3rd Qu.:173.43	3rd Qu.: 9.153	3rd Qu.: 30.00	3rd Qu.:2.000
Max. :337.03	Max. :298.88	Max. :60.146	Max. :180.00	Max. :4.000