kaplan_meier_analysis

Import library

```
library(survival)
library(survminer)
```

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

```
## Loading required package: ggpubr
```

```
##
## Attaching package: 'survminer'
```

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

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.1     v stringr   1.6.0
## v lubridate 1.9.4     v tibble    3.3.0
## v purrr     1.2.0     v tidyr     1.3.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readr)
library(dplyr)
library(gtsummary)
```

Import the data

```
data <- read.csv("~/Documents/Github/bio_stat_projects/data/seer_breat_cancer_dataset.csv", sep = ",")
```

```
data$Race <- ifelse(data$Race == "Other (American Indian/AK Native, Asian/Pacific Islander)", "Other", d
data$Race <- as.factor(data$Race)

data$Marital.Status <- as.factor(data$Marital.Status)
```

```r
data <- data %>%
  mutate(
    Marital.Status = case_when(
      Marital.Status == "Married (including common law)" ~ "Married",
      Marital.Status == "Single (never married)" ~ "Single",
      TRUE ~ Marital.Status # Keep everything else as it
    )
  )

data$T.Stage <- as.factor(data$T.Stage)
data$N.Stage <- as.factor(data$N.Stage)
data$X6th.Stage <- as.factor(data$X6th.Stage)
data$Grade <- as.factor(data$Grade)
data$A.Stage <- as.factor(data$A.Stage)
data$Estrogen.Status <- as.factor(data$Estrogen.Status)
data$Progesterone.Status <- as.factor(data$Progesterone.Status)

data$Status <- ifelse(data$Status == "Dead", 1, 0)

summary(data)
```

```
##       Age           Race      Marital.Status        X           T.Stage
##  Min.   :30.00   Black: 291   Length:4024        Mode:logical   T1:1603
##  1st Qu.:47.00   Other: 320   Class :character   NA's:4024      T2:1786
##  Median :54.00   White:3413   Mode  :character                  T3: 533
##  Mean   :53.97                                                 T4: 102
##  3rd Qu.:61.00
##  Max.   :69.00
##  N.Stage   X6th.Stage                                    Grade
##  N1:2732   IIA :1305   Moderately differentiated; Grade II   :2351
##  N2: 820   IIB :1130   Poorly differentiated; Grade III      :1111
##  N3: 472   IIIA:1050   Undifferentiated; anaplastic; Grade IV:  19
##            IIIB:  67   Well differentiated; Grade I          : 543
##            IIIC: 472
##
##      A.Stage       Tumor.Size     Estrogen.Status Progesterone.Status
##  Distant :  92   Min.   :  1.00   Negative: 269   Negative: 698
##  Regional:3932   1st Qu.: 16.00   Positive:3755   Positive:3326
##                  Median : 25.00
##                  Mean   : 30.47
##                  3rd Qu.: 38.00
##                  Max.   :140.00
##  Regional.Node.Examined Reginol.Node.Positive Survival.Months     Status
##  Min.   : 1.00          Min.   : 1.000        Min.   :  1.0   Min.   :0.0000
##  1st Qu.: 9.00          1st Qu.: 1.000        1st Qu.: 56.0   1st Qu.:0.0000
##  Median :14.00          Median : 2.000        Median : 73.0   Median :0.0000
##  Mean   :14.36          Mean   : 4.158        Mean   : 71.3   Mean   :0.1531
##  3rd Qu.:19.00          3rd Qu.: 5.000        3rd Qu.: 90.0   3rd Qu.:0.0000
##  Max.   :61.00          Max.   :46.000        Max.   :107.0   Max.   :1.0000
```

```r
tbl_summary(data)
```

```r
surv_object <- Surv(time = data$Survival.Months, event=data$Status)
```

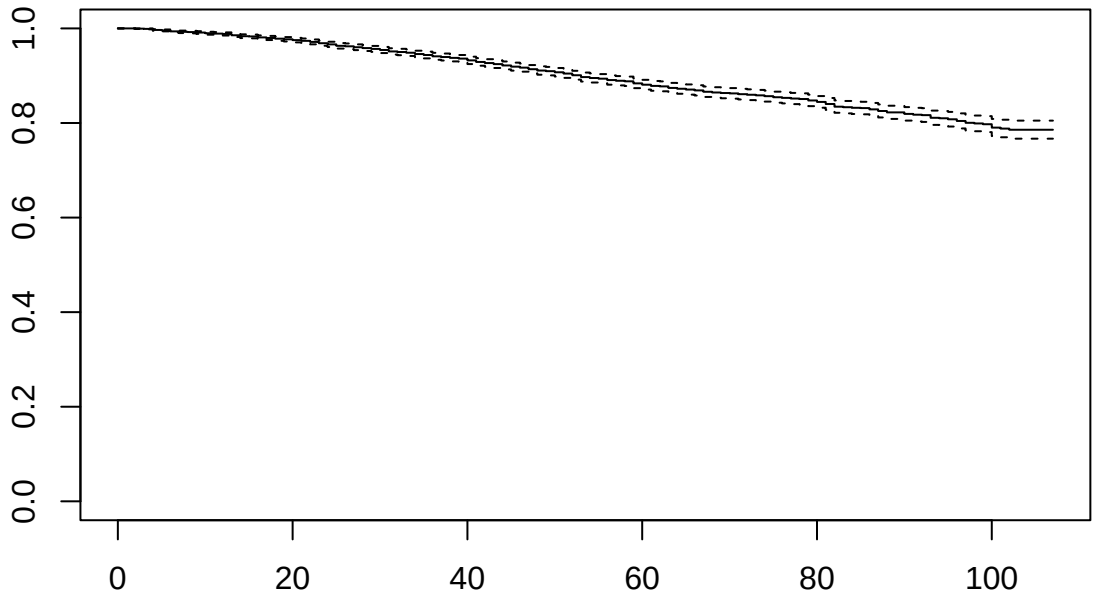
```r
fit <- survfit(surv_object ~ 1)
summary(fit)
```

```
## Call: survfit(formula = surv_object ~ 1)
##
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     2   4023       2    1.000 0.000351        0.999        1.000
##     3   4020       2    0.999 0.000497        0.998        1.000
##     4   4016       9    0.997 0.000895        0.995        0.999
##     5   4006       4    0.996 0.001023        0.994        0.998
##     6   4000       8    0.994 0.001240        0.991        0.996
##     7   3992       4    0.993 0.001335        0.990        0.995
##     8   3987       2    0.992 0.001380        0.990        0.995
##     9   3980       4    0.991 0.001466        0.988        0.994
##    10   3971       6    0.990 0.001586        0.987        0.993
##    11   3963       3    0.989 0.001643        0.986        0.992
##    12   3959       3    0.988 0.001698        0.985        0.992
##    13   3953      11    0.986 0.001885        0.982        0.989
##    14   3941       8    0.984 0.002009        0.980        0.987
##    15   3931       2    0.983 0.002039        0.979        0.987
##    16   3925       8    0.981 0.002154        0.977        0.985
##    17   3917       5    0.980 0.002223        0.975        0.984
##    18   3911       6    0.978 0.002303        0.974        0.983
##    19   3900       5    0.977 0.002367        0.972        0.982
##    20   3895       7    0.975 0.002454        0.970        0.980
##    21   3885       6    0.974 0.002526        0.969        0.979
##    22   3878       9    0.972 0.002630        0.966        0.977
##    23   3867      11    0.969 0.002752        0.963        0.974
##    24   3851      10    0.966 0.002857        0.961        0.972
##    25   3839      12    0.963 0.002978        0.957        0.969
##    26   3827       4    0.962 0.003018        0.956        0.968
##    27   3822       8    0.960 0.003094        0.954        0.966
##    28   3809       7    0.958 0.003160        0.952        0.965
##    29   3801       7    0.957 0.003223        0.950        0.963
##    30   3794       9    0.954 0.003303        0.948        0.961
##    31   3784      12    0.951 0.003406        0.945        0.958
##    32   3769       5    0.950 0.003448        0.943        0.957
##    33   3764       6    0.949 0.003498        0.942        0.955
##    34   3755      11    0.946 0.003587        0.939        0.953
##    35   3744       8    0.944 0.003649        0.937        0.951
##    36   3733      11    0.941 0.003734        0.934        0.948
##    37   3722       6    0.939 0.003779        0.932        0.947
##    38   3711       7    0.938 0.003830        0.930        0.945
##    39   3698       6    0.936 0.003874        0.929        0.944
##    40   3689      15    0.932 0.003981        0.925        0.940
##    41   3671      13    0.929 0.004071        0.921        0.937
##    42   3647       9    0.927 0.004132        0.919        0.935
##    43   3633       9    0.924 0.004192        0.916        0.933
##    44   3623      11    0.922 0.004264        0.913        0.930
##    45   3607      10    0.919 0.004328        0.911        0.928
##    46   3591       8    0.917 0.004379        0.909        0.926
```

```
##    47   3577      13    0.914 0.004459        0.905        0.923
##    48   3551      11    0.911 0.004526        0.902        0.920
##    49   3497       5    0.910 0.004557        0.901        0.919
##    50   3436       9    0.907 0.004614        0.898        0.916
##    51   3369      10    0.905 0.004678        0.895        0.914
##    52   3314      13    0.901 0.004762        0.892        0.910
##    53   3252      14    0.897 0.004853        0.888        0.907
##    54   3191       7    0.895 0.004899        0.886        0.905
##    55   3132       5    0.894 0.004933        0.884        0.903
##    56   3069      10    0.891 0.005002        0.881        0.901
##    57   2996       5    0.889 0.005038        0.880        0.899
##    58   2943       4    0.888 0.005067        0.878        0.898
##    59   2882      14    0.884 0.005172        0.874        0.894
##    60   2821       9    0.881 0.005240        0.871        0.891
##    61   2758       8    0.878 0.005302        0.868        0.889
##    62   2698       3    0.877 0.005326        0.867        0.888
##    63   2646      10    0.874 0.005408        0.864        0.885
##    64   2573       5    0.872 0.005451        0.862        0.883
##    65   2511       5    0.871 0.005495        0.860        0.882
##    66   2456       5    0.869 0.005541        0.858        0.880
##    67   2385       8    0.866 0.005617        0.855        0.877
##    68   2324       4    0.865 0.005657        0.854        0.876
##    69   2261       3    0.863 0.005688        0.852        0.875
##    70   2180       2    0.863 0.005710        0.851        0.874
##    71   2131       4    0.861 0.005757        0.850        0.872
##    72   2090       3    0.860 0.005792        0.848        0.871
##    73   2033       3    0.858 0.005830        0.847        0.870
##    74   1978       4    0.857 0.005882        0.845        0.868
##    75   1922       5    0.854 0.005951        0.843        0.866
##    76   1857       3    0.853 0.005994        0.841        0.865
##    77   1798       3    0.852 0.006041        0.840        0.864
##    78   1734       2    0.851 0.006073        0.839        0.863
##    79   1678       6    0.848 0.006177        0.836        0.860
##    80   1623       6    0.845 0.006286        0.832        0.857
##    81   1555       8    0.840 0.006438        0.828        0.853
##    82   1487      10    0.835 0.006638        0.822        0.848
##    83   1426       2    0.833 0.006680        0.820        0.847
##    84   1360       2    0.832 0.006726        0.819        0.845
##    85   1302       1    0.831 0.006752        0.818        0.845
##    86   1250       4    0.829 0.006860        0.815        0.842
##    87   1193       5    0.825 0.007005        0.812        0.839
##    88   1140       4    0.822 0.007128        0.809        0.837
##    90   1011       4    0.819 0.007283        0.805        0.834
##    91    968       2    0.818 0.007366        0.803        0.832
##    92    917       1    0.817 0.007412        0.802        0.831
##    93    868       6    0.811 0.007710        0.796        0.826
##    94    806       1    0.810 0.007766        0.795        0.825
##    95    749       2    0.808 0.007895        0.792        0.823
##    96    682       3    0.804 0.008122        0.788        0.820
##    97    628       3    0.800 0.008381        0.784        0.817
##    98    571       1    0.799 0.008483        0.783        0.816
##    99    499       1    0.797 0.008615        0.781        0.814
##   100    444       4    0.790 0.009256        0.772        0.809
##   101    378       1    0.788 0.009465        0.770        0.807
```

```
##    102     324       1     0.786 0.009743        0.767        0.805
```

```
plot(fit)
```



Characteristic	**N = 4,024**[1]
Age	54 (47, 61)
Race	
Black	291 (7.2%)
Other	320 (8.0%)
White	3,413 (85%)
Marital.Status	
Divorced	486 (12%)
Married	2,643 (66%)
Separated	45 (1.1%)
Single	615 (15%)
Widowed	235 (5.8%)
X	0 (NA%)
Unknown	4,024
T.Stage	
T1	1,603 (40%)
T2	1,786 (44%)
T3	533 (13%)
T4	102 (2.5%)
N.Stage	
N1	2,732 (68%)
N2	820 (20%)
N3	472 (12%)
X6th.Stage	
IIA	1,305 (32%)
IIB	1,130 (28%)
IIIA	1,050 (26%)
IIIB	67 (1.7%)
IIIC	472 (12%)
Grade	
Moderately differentiated; Grade II	2,351 (58%)
Poorly differentiated; Grade III	1,111 (28%)
Undifferentiated; anaplastic; Grade IV	19 (0.5%)
Well differentiated; Grade I	543 (13%)
A.Stage	
Distant	92 (2.3%)
Regional	3,932 (98%)
Tumor.Size	25 (16, 38)
Estrogen.Status	
Negative	269 (6.7%)
Positive	3,755 (93%)
Progesterone.Status	
Negative	698 (17%)
Positive	3,326 (83%)
Regional.Node.Examined	14 (9, 19)
Reginol.Node.Positive	2 (1, 5)
Survival.Months	73 (56, 90)
Status	616 (15%)

[1] Median (Q1, Q3); n (%)