STAT_306_Project_Group_10

STAT 306 Project

Nicholas Tam, Ivy Cheung, Kaichi Nakajima, Kevin Liu (Group 10)

Introduction

Breast invasive ductal carcinoma (IDC) is the most common type of breast cancer, with about 80% of all forms of breast cancer being IDC, according to the American Cancer Society (DePolo, 2024). There are numerous nonsurgical treatments of IDC, such as radiotherapy, chemotherapy, and hormone therapy (Wright, 2023), but it is unclear how combinations of treatments can interact in a model to predict a patient's survival until death.

For our research project, we have selected a dataset of approximately 1900 primary breast cancer samples, obtained from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) database through cBioPortal for Cancer Genomics.

Our project question is:

"How do radiotherapy, chemotherapy, and hormone therapy influence the length of time a patient with IDC will survive until death, given control variables age, surgery type, and tumor stage?"

Our analysis will involve the inference of covariates within linear models, and thus _____

- Variables from columns 32 to 693 consist of genetic attributes containing m-RNA levels z-score for 331 genes, and mutation for 175 genes; they have been omitted due to being difficult to interpret.

Analysis

Loading relevant libraries

```
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

-- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
v forcats   1.0.0     v readr     2.1.5
v ggplot2   3.5.1     v stringr   1.5.0
v lubridate 1.9.3     v tibble    3.2.1
v purrr     1.0.1     v tidyr     1.3.1
-- 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 becor
Loading required package: carData


Attaching package: 'car'


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

    some


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

    recode


Attaching package: 'gridExtra'
```

```
The following object is masked from 'package:dplyr':

    combine
```

Uploading relevant table and narrowing down to data of interest

- Due to the distribution of "cancer_type_detailed" categories and for ease of computation, we will be filtering the dataset for deceased IDC patients only, as IDC consists of the majority of the dataset, and the "overall_survival" column for living patients would be inconsistent for living patients due to time.

- "tumor_size" has been removed due to already partly contributing to "tumor_stage". (Susan G. Komen®, 2024)

- All rows with "N/A" or " " as their values for any column have been removed to ensure no data is missing.

- All categorical variables are transformed to be treated as categorical by the table.

```
   age_at_diagnosis type_of_breast_surgery tumor_stage chemotherapy
1            75.65             MASTECTOMY           2            0
2            43.19      BREAST CONSERVING           1            0
3            48.87             MASTECTOMY           2            1
4            56.45      BREAST CONSERVING           2            1
5            86.41      BREAST CONSERVING           2            0
6            70.91      BREAST CONSERVING           1            0
7            45.27             MASTECTOMY           2            1
8            83.02             MASTECTOMY           2            0
9            51.46      BREAST CONSERVING           2            1
10           70.02      BREAST CONSERVING           2            0
   hormone_therapy radio_therapy      death_from_cancer overall_survival_months
1                1             1                 Living               140.50000
2                1             1                 Living                84.63333
3                1             0        Died of Disease               163.70000
4                1             1                 Living               164.33333
5                1             1 Died of Other Causes                 36.56667
6                1             1                 Living               163.53333
7                0             1                 Living               164.90000
8                1             1 Died of Other Causes                 14.13333
9                1             1                 Living               103.83333
10               1             1                 Living               161.06667
```

```
age_at_diagnosis       type_of_breast_surgery tumor_stage chemotherapy
Min.   :26.36    BREAST CONSERVING:462          1:353       0:808
1st Qu.:50.35    MASTECTOMY       :610          2:632       1:264
Median :60.94                                   3: 87
Mean   :60.17
3rd Qu.:69.87
Max.   :96.29
hormone_therapy radio_therapy          death_from_cancer
0:433           0:345        Living               :468
1:639           1:727        Died of Disease      :368
                             Died of Other Causes:236



overall_survival_months
Min.   :  0.10
1st Qu.: 61.27
Median :117.62
Mean   :126.69
3rd Qu.:187.03
Max.   :351.00
```
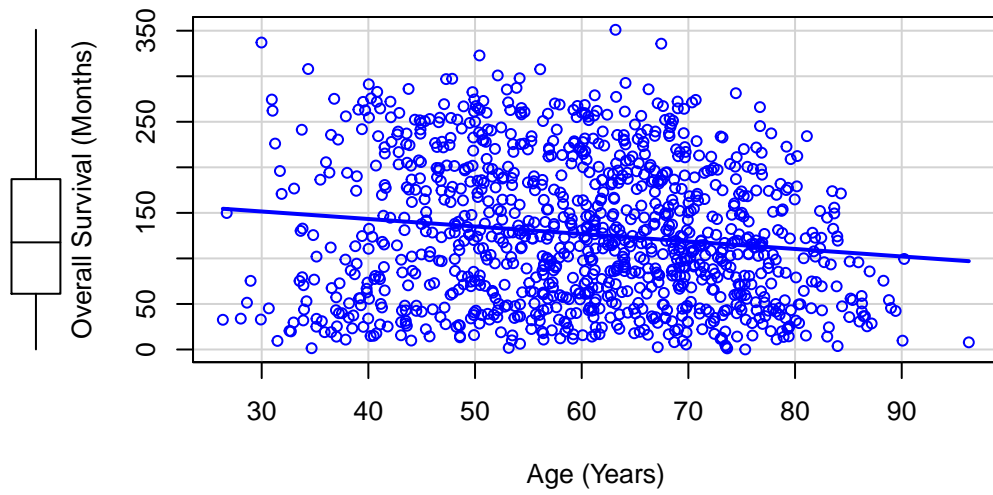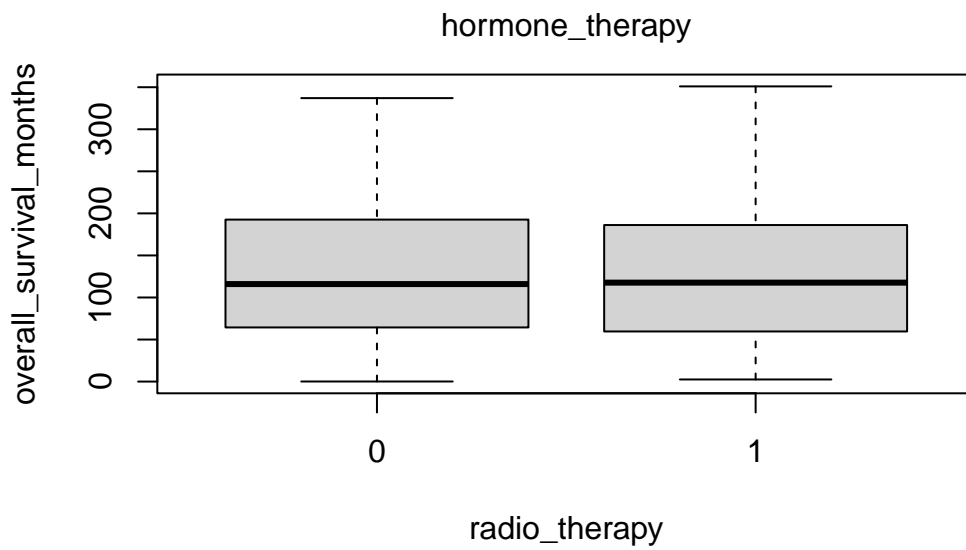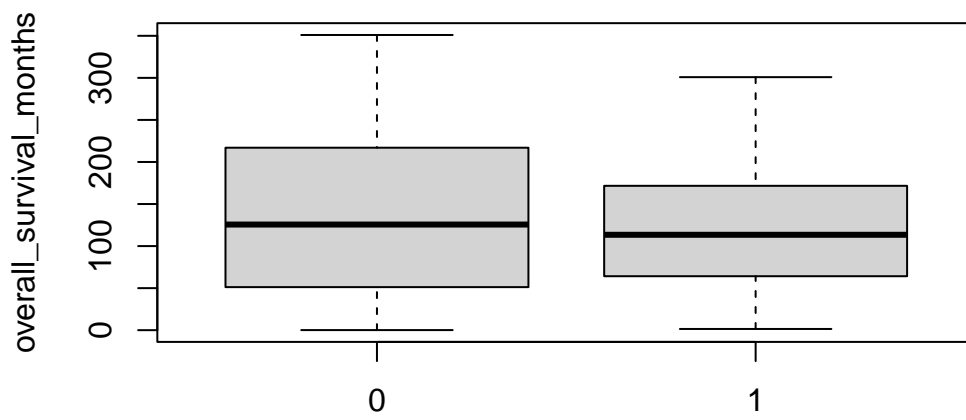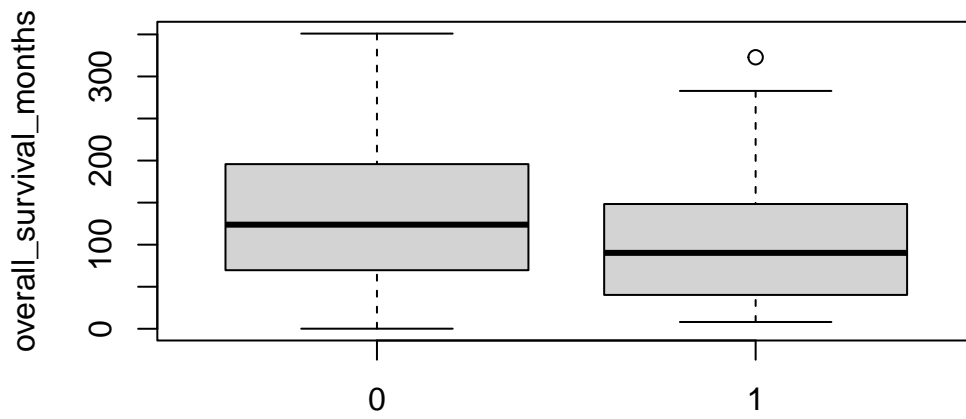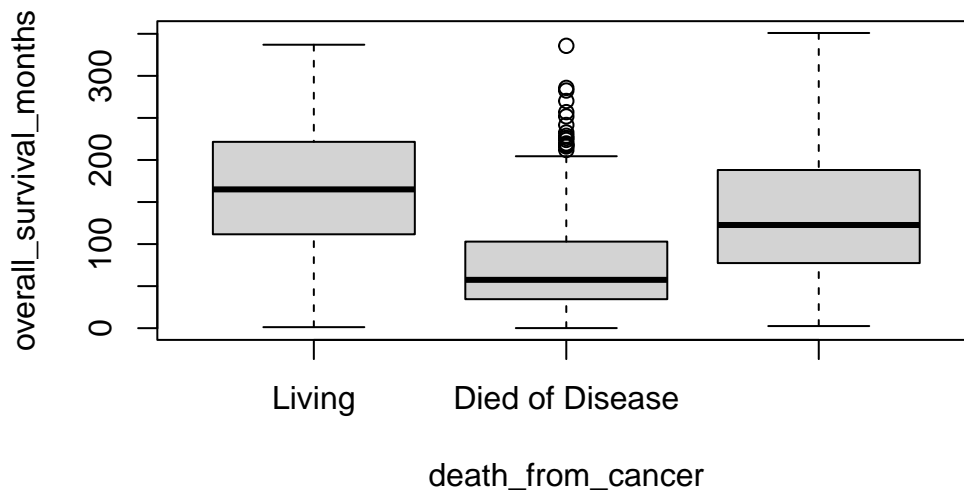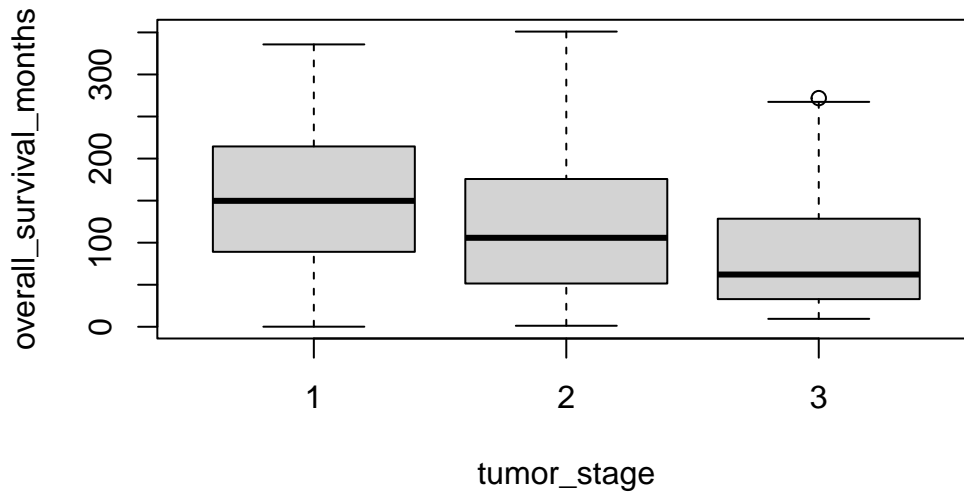
Exploratory data analysis

-

Overall Survival against Age
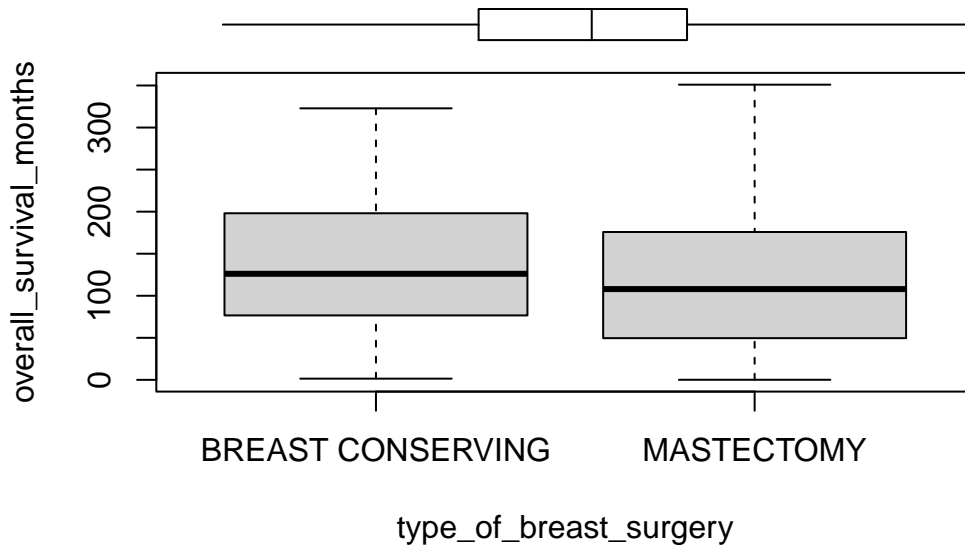
Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * type_of_breast_surgery,
    data = cancer_new)

Residuals:
```
     Min       1Q   Median       3Q      Max
-146.840  -63.418   -8.192   59.119  233.599
```

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	164.9777	17.7215	9.309
age_at_diagnosis	-0.4817	0.2961	-1.627
type_of_breast_surgeryMASTECTOMY	10.5483	22.5335	0.468

```
age_at_diagnosis:type_of_breast_surgeryMASTECTOMY  -0.4385     0.3696  -1.186
                                                    Pr(>|t|)
(Intercept)                                         <2e-16 ***
age_at_diagnosis                                     0.104
type_of_breast_surgeryMASTECTOMY                     0.640
age_at_diagnosis:type_of_breast_surgeryMASTECTOMY    0.236
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 76.3 on 1068 degrees of freedom
Multiple R-squared:  0.03087,   Adjusted R-squared:  0.02815
F-statistic: 11.34 on 3 and 1068 DF,  p-value: 2.524e-07




Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * chemotherapy,
    data = cancer_new)

Residuals:
     Min       1Q   Median       3Q      Max
-185.048  -59.736   -6.789   58.059  218.435

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   249.6778    13.8963  17.967  < 2e-16 ***
age_at_diagnosis               -1.8223     0.2148  -8.484  < 2e-16 ***
chemotherapy1                -117.0181    25.4821  -4.592 4.91e-06 ***
age_at_diagnosis:chemotherapy1  1.2618     0.4707   2.681  0.00746 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73.97 on 1068 degrees of freedom
Multiple R-squared:  0.08924,   Adjusted R-squared:  0.08668
F-statistic: 34.88 on 3 and 1068 DF,  p-value: < 2.2e-16




Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * hormone_therapy,
    data = cancer_new)

Residuals:
     Min       1Q   Median       3Q      Max
```

```
-151.456  -64.283   -5.846   61.586  216.895


Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                       148.4095    15.5851   9.523   <2e-16 ***
age_at_diagnosis                   -0.2264     0.2748  -0.824   0.4101
hormone_therapy1                   43.4052    22.5901   1.921   0.0549 .
age_at_diagnosis:hormone_therapy1  -0.8960     0.3733  -2.400   0.0166 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 76.4 on 1068 degrees of freedom
Multiple R-squared:  0.02823,   Adjusted R-squared:  0.0255
F-statistic: 10.34 on 3 and 1068 DF,  p-value: 1.035e-06




Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * radio_therapy,
    data = cancer_new)

Residuals:
     Min       1Q   Median       3Q      Max
-160.961  -64.542   -6.219   60.830  228.620

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     203.1604    19.6689  10.329  < 2e-16 ***
age_at_diagnosis                 -1.1755     0.3090  -3.804  0.00015 ***
radio_therapy1                  -37.2283    23.6896  -1.572  0.11636
age_at_diagnosis:radio_therapy1   0.4861     0.3781   1.286  0.19885
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 76.6 on 1068 degrees of freedom
Multiple R-squared:  0.02326,   Adjusted R-squared:  0.02052
F-statistic: 8.479 on 3 and 1068 DF,  p-value: 1.436e-05




Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * tumor_stage,
    data = cancer_new)
```

```
Residuals:
     Min      1Q   Median      3Q      Max
-154.935  -60.562   -6.494   58.906  234.345


Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   206.9199    20.1807  10.253  < 2e-16 ***
age_at_diagnosis               -0.9447     0.3392  -2.785  0.00545 **
tumor_stage2                  -44.6092    24.2427  -1.840  0.06603 .
tumor_stage3                 -128.9948    40.9834  -3.147  0.00169 **
age_at_diagnosis:tumor_stage2   0.2219     0.4013   0.553  0.58041
age_at_diagnosis:tumor_stage3   1.0893     0.6635   1.642  0.10092
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 74.41 on 1066 degrees of freedom
Multiple R-squared:  0.08009,   Adjusted R-squared:  0.07577
F-statistic: 18.56 on 5 and 1066 DF,  p-value: < 2.2e-16




Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * death_from_cancer,
    data = cancer_new)

Residuals:
     Min      1Q   Median      3Q      Max
-194.458  -44.887   -8.238   46.802  257.828

Coefficients:
                                                      Estimate Std. Error
(Intercept)                                           226.6910    14.8952
age_at_diagnosis                                       -1.1187     0.2613
death_from_cancerDied of Disease                     -168.5633    21.2342
death_from_cancerDied of Other Causes                  88.0976    34.2708
age_at_diagnosis:death_from_cancerDied of Disease       1.4119     0.3610
age_at_diagnosis:death_from_cancerDied of Other Causes -1.4804     0.5073
                                                      t value Pr(>|t|)
(Intercept)                                            15.219  < 2e-16 ***
age_at_diagnosis                                       -4.281 2.03e-05 ***
death_from_cancerDied of Disease                       -7.938 5.17e-15 ***
death_from_cancerDied of Other Causes                   2.571  0.01029 *
age_at_diagnosis:death_from_cancerDied of Disease       3.911 9.76e-05 ***
```

```
age_at_diagnosis:death_from_cancerDied of Other Causes  -2.918  0.00359 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 65.31 on 1066 degrees of freedom
Multiple R-squared:  0.2912,    Adjusted R-squared:  0.2879
F-statistic:  87.6 on 5 and 1066 DF,  p-value: < 2.2e-16


`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```
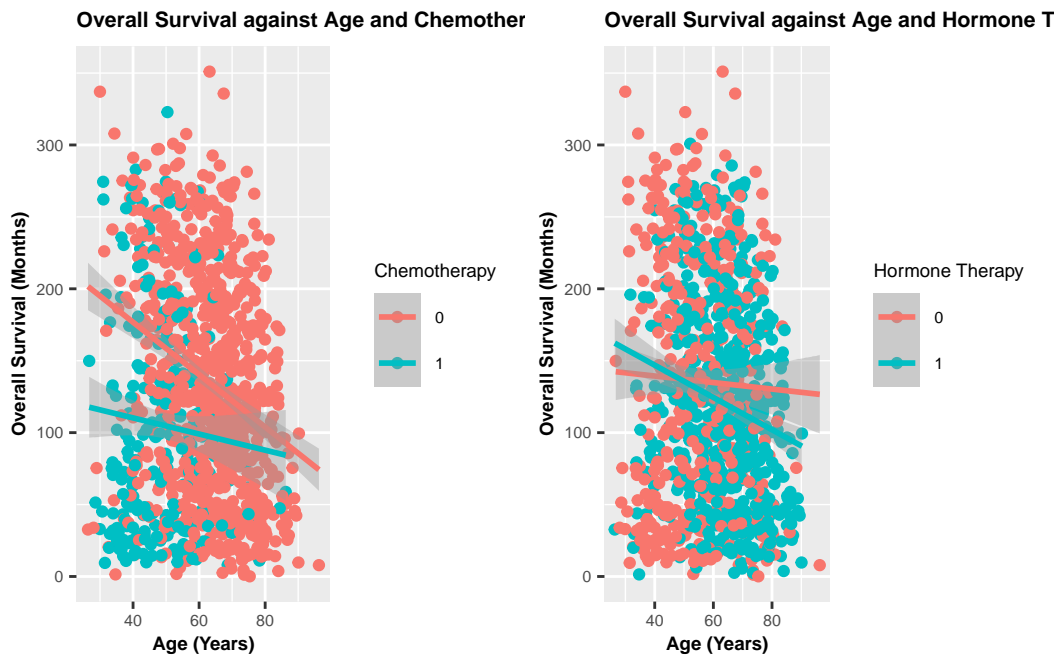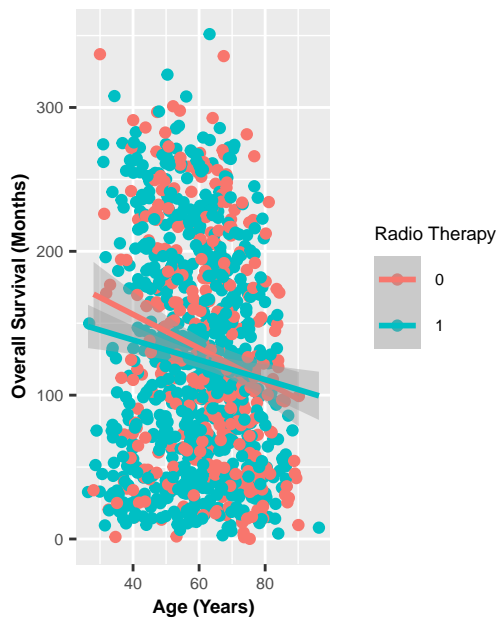


```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```

Overall Survival against Age and Radio Ther
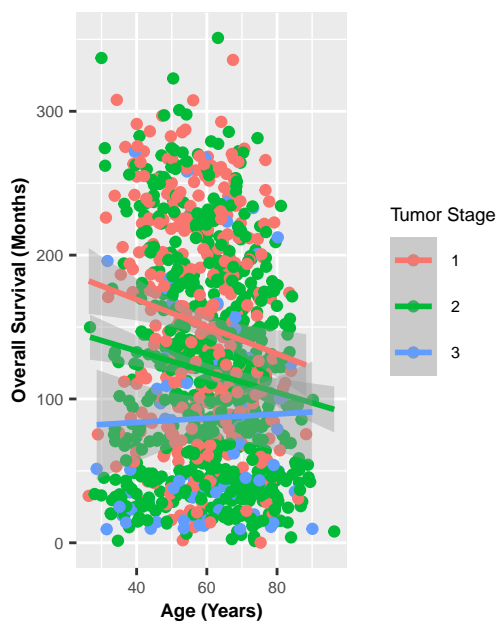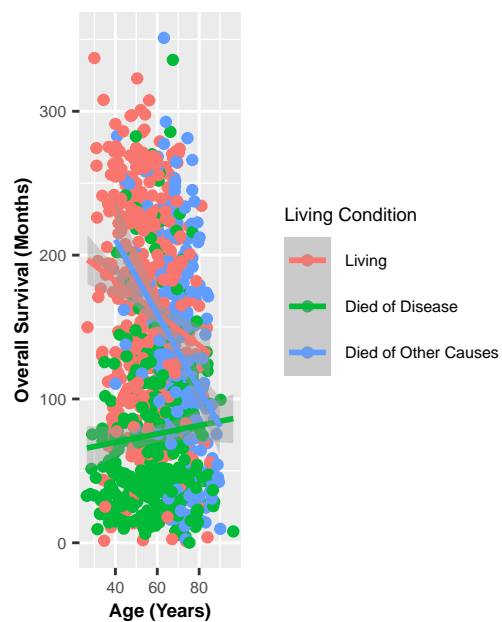
Overall Survival against Age and Breast Surg

```
`geom_smooth()` using formula = 'y ~ x'
`geom_smooth()` using formula = 'y ~ x'
```

Overall Survival against Age and Tumor Stag

Overall Survival against Age and Living Con

Model selection

- The interaction terms for all therapies

```
Start:  AIC=8939.01
overall_survival_months ~ age_at_diagnosis + chemotherapy * hormone_therapy *
    radio_therapy + age_at_diagnosis:chemotherapy + age_at_diagnosis:hormone_therapy +
    age_at_diagnosis:radio_therapy + tumor_stage + tumor_stage:chemotherapy +
    tumor_stage:hormone_therapy + tumor_stage:radio_therapy +
    type_of_breast_surgery + death_from_cancer
```

	Df	Sum of Sq	RSS	AIC
- chemotherapy:hormone_therapy:radio_therapy	1	170	4295339	8937.1
- radio_therapy:tumor_stage	2	8551	4303720	8937.1
- age_at_diagnosis:radio_therapy	1	1750	4296918	8937.4
- age_at_diagnosis:hormone_therapy	1	1836	4297005	8937.5
- type_of_breast_surgery	1	2863	4298032	8937.7
<none>			4295169	8939.0
- chemotherapy:tumor_stage	2	22615	4317783	8940.6
- age_at_diagnosis:chemotherapy	1	30237	4325405	8944.5
- hormone_therapy:tumor_stage	2	40117	4335286	8945.0
- death_from_cancer	2	1264608	5559776	9211.7

```
Step:  AIC=8937.05
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    radio_therapy + tumor_stage + type_of_breast_surgery + death_from_cancer +
    chemotherapy:hormone_therapy + chemotherapy:radio_therapy +
    hormone_therapy:radio_therapy + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    chemotherapy:tumor_stage + hormone_therapy:tumor_stage +
    radio_therapy:tumor_stage
```

	Df	Sum of Sq	RSS	AIC
- radio_therapy:tumor_stage	2	8426	4303764	8935.2
- chemotherapy:radio_therapy	1	1115	4296454	8935.3
- age_at_diagnosis:radio_therapy	1	1719	4297058	8935.5
- age_at_diagnosis:hormone_therapy	1	1871	4297210	8935.5
- type_of_breast_surgery	1	2903	4298241	8935.8
- chemotherapy:hormone_therapy	1	5113	4300451	8936.3
- hormone_therapy:radio_therapy	1	5670	4301009	8936.5
<none>			4295339	8937.1
- chemotherapy:tumor_stage	2	22460	4317799	8938.6

```
+ chemotherapy:hormone_therapy:radio_therapy  1       170 4295169 8939.0
- age_at_diagnosis:chemotherapy               1     30104 4325442 8942.5
- hormone_therapy:tumor_stage                 2     39983 4335321 8943.0
- death_from_cancer                           2   1264495 5559834 9209.7

Step:  AIC=8935.16
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    radio_therapy + tumor_stage + type_of_breast_surgery + death_from_cancer +
    chemotherapy:hormone_therapy + chemotherapy:radio_therapy +
    hormone_therapy:radio_therapy + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    chemotherapy:tumor_stage + hormone_therapy:tumor_stage
```

	Df	Sum of Sq	RSS	AIC
- chemotherapy:radio_therapy	1	334	4304099	8933.2
- age_at_diagnosis:radio_therapy	1	832	4304597	8933.4
- age_at_diagnosis:hormone_therapy	1	1987	4305751	8933.6
- type_of_breast_surgery	1	2736	4306501	8933.8
- hormone_therapy:radio_therapy	1	3506	4307270	8934.0
- chemotherapy:hormone_therapy	1	7072	4310836	8934.9
<none>			4303764	8935.2
+ radio_therapy:tumor_stage	2	8426	4295339	8937.1
+ chemotherapy:hormone_therapy:radio_therapy	1	45	4303720	8937.1
- chemotherapy:tumor_stage	2	24662	4328426	8937.3
- age_at_diagnosis:chemotherapy	1	27301	4331065	8939.9
- hormone_therapy:tumor_stage	2	41669	4345434	8941.5
- death_from_cancer	2	1281516	5585281	9210.6

```
Step:  AIC=8933.24
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    radio_therapy + tumor_stage + type_of_breast_surgery + death_from_cancer +
    chemotherapy:hormone_therapy + hormone_therapy:radio_therapy +
    age_at_diagnosis:chemotherapy + age_at_diagnosis:hormone_therapy +
    age_at_diagnosis:radio_therapy + chemotherapy:tumor_stage +
    hormone_therapy:tumor_stage
```

	Df	Sum of Sq	RSS	AIC
- age_at_diagnosis:radio_therapy	1	552	4304651	8931.4
- age_at_diagnosis:hormone_therapy	1	2061	4306160	8931.8
- type_of_breast_surgery	1	2512	4306611	8931.9
- hormone_therapy:radio_therapy	1	3450	4307548	8932.1
- chemotherapy:hormone_therapy	1	7255	4311353	8933.0
<none>			4304099	8933.2

```
+ chemotherapy:radio_therapy            1          334 4303764 8935.2
- chemotherapy:tumor_stage              2        24481 4328580 8935.3
+ radio_therapy:tumor_stage             2         7645 4296454 8935.3
- age_at_diagnosis:chemotherapy         1        27079 4331177 8938.0
- hormone_therapy:tumor_stage           2        41983 4346082 8939.6
- death_from_cancer                     2      1281655 5585754 9208.7

Step:  AIC=8931.38
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    radio_therapy + tumor_stage + type_of_breast_surgery + death_from_cancer +
    chemotherapy:hormone_therapy + hormone_therapy:radio_therapy +
    age_at_diagnosis:chemotherapy + age_at_diagnosis:hormone_therapy +
    chemotherapy:tumor_stage + hormone_therapy:tumor_stage

                                  Df Sum of Sq      RSS    AIC
- age_at_diagnosis:hormone_therapy  1      2175 4306825 8929.9
- type_of_breast_surgery            1      2404 4307055 8930.0
- hormone_therapy:radio_therapy     1      4687 4309337 8930.5
- chemotherapy:hormone_therapy      1      7199 4311849 8931.2
<none>                                          4304651 8931.4
+ age_at_diagnosis:radio_therapy    1       552 4304099 8933.2
+ chemotherapy:radio_therapy        1        54 4304597 8933.4
- chemotherapy:tumor_stage          2     24731 4329382 8933.5
+ radio_therapy:tumor_stage         2      7315 4297336 8933.6
- age_at_diagnosis:chemotherapy     1     26862 4331512 8936.0
- hormone_therapy:tumor_stage       2     42677 4347327 8938.0
- death_from_cancer                 2   1281676 5586326 9206.8

Step:  AIC=8929.92
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    radio_therapy + tumor_stage + type_of_breast_surgery + death_from_cancer +
    chemotherapy:hormone_therapy + hormone_therapy:radio_therapy +
    age_at_diagnosis:chemotherapy + chemotherapy:tumor_stage +
    hormone_therapy:tumor_stage

                                  Df Sum of Sq      RSS    AIC
- type_of_breast_surgery            1      2173 4308999 8928.5
- hormone_therapy:radio_therapy     1      4529 4311355 8929.0
- chemotherapy:hormone_therapy      1      5192 4312018 8929.2
<none>                                          4306825 8929.9
+ age_at_diagnosis:hormone_therapy  1      2175 4304651 8931.4
+ age_at_diagnosis:radio_therapy    1       666 4306160 8931.8
- chemotherapy:tumor_stage          2     24070 4330895 8931.9
```

```
+ chemotherapy:radio_therapy          1         68 4306757 8931.9
+ radio_therapy:tumor_stage           2       7238 4299587 8932.1
- age_at_diagnosis:chemotherapy       1      27146 4333971 8934.7
- hormone_therapy:tumor_stage         2      40542 4347367 8936.0
- death_from_cancer                   2    1295194 5602020 9207.8

Step:  AIC=8928.46
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    radio_therapy + tumor_stage + death_from_cancer + chemotherapy:hormone_therapy +
    hormone_therapy:radio_therapy + age_at_diagnosis:chemotherapy +
    chemotherapy:tumor_stage + hormone_therapy:tumor_stage

                                     Df Sum of Sq     RSS    AIC
- hormone_therapy:radio_therapy       1       3837 4312835 8927.4
- chemotherapy:hormone_therapy        1       5390 4314389 8927.8
<none>                                             4308999 8928.5
+ type_of_breast_surgery              1       2173 4306825 8929.9
+ age_at_diagnosis:hormone_therapy    1       1944 4307055 8930.0
+ age_at_diagnosis:radio_therapy      1        546 4308453 8930.3
+ chemotherapy:radio_therapy          1          7 4308992 8930.5
- chemotherapy:tumor_stage            2      24580 4333579 8930.6
+ radio_therapy:tumor_stage           2       7281 4301718 8930.6
- age_at_diagnosis:chemotherapy       1      26590 4335589 8933.1
- hormone_therapy:tumor_stage         2      41496 4350495 8934.7
- death_from_cancer                   2    1295324 5604323 9206.2

Step:  AIC=8927.41
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    radio_therapy + tumor_stage + death_from_cancer + chemotherapy:hormone_therapy +
    age_at_diagnosis:chemotherapy + chemotherapy:tumor_stage +
    hormone_therapy:tumor_stage

                                     Df Sum of Sq     RSS    AIC
- radio_therapy                       1       2245 4315080 8926.0
<none>                                             4312835 8927.4
- chemotherapy:hormone_therapy        1       8547 4321382 8927.5
+ hormone_therapy:radio_therapy       1       3837 4308999 8928.5
+ age_at_diagnosis:hormone_therapy    1       1842 4310993 8929.0
+ age_at_diagnosis:radio_therapy      1       1650 4311186 8929.0
+ type_of_breast_surgery              1       1481 4311355 8929.0
+ chemotherapy:radio_therapy          1         31 4312804 8929.4
- chemotherapy:tumor_stage            2      25706 4338541 8929.8
+ radio_therapy:tumor_stage           2       5066 4307769 8930.2
```

```
- age_at_diagnosis:chemotherapy      1      26066 4338901 8931.9
- hormone_therapy:tumor_stage        2      42656 4355491 8934.0
- death_from_cancer                  2    1296638 5609473 9205.2
```

Step: AIC=8925.97
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    tumor_stage + death_from_cancer + chemotherapy:hormone_therapy +
    age_at_diagnosis:chemotherapy + chemotherapy:tumor_stage +
    hormone_therapy:tumor_stage

```
                                    Df Sum of Sq      RSS    AIC
- chemotherapy:hormone_therapy       1       7884 4322964 8925.9
<none>                                             4315080 8926.0
+ type_of_breast_surgery             1       3216 4311864 8927.2
+ radio_therapy                      1       2245 4312835 8927.4
+ age_at_diagnosis:hormone_therapy   1       1723 4313357 8927.5
- chemotherapy:tumor_stage           2      25400 4340479 8928.3
- age_at_diagnosis:chemotherapy      1      25596 4340676 8930.3
- hormone_therapy:tumor_stage        2      41255 4356334 8932.2
- death_from_cancer                  2    1295104 5610184 9203.3
```

Step: AIC=8925.93
overall_survival_months ~ age_at_diagnosis + chemotherapy + hormone_therapy +
    tumor_stage + death_from_cancer + age_at_diagnosis:chemotherapy +
    chemotherapy:tumor_stage + hormone_therapy:tumor_stage

```
                                    Df Sum of Sq      RSS    AIC
<none>                                             4322964 8925.9
+ chemotherapy:hormone_therapy       1       7884 4315080 8926.0
+ type_of_breast_surgery             1       2777 4320187 8927.2
- chemotherapy:tumor_stage           2      21980 4344944 8927.4
+ radio_therapy                      1       1582 4321382 8927.5
+ age_at_diagnosis:hormone_therapy   1          0 4322964 8927.9
- age_at_diagnosis:chemotherapy      1      20901 4343865 8929.1
- hormone_therapy:tumor_stage        2      33533 4356497 8930.2
- death_from_cancer                  2    1288783 5611747 9201.6
```

Call:
lm(formula = overall_survival_months ~ age_at_diagnosis + chemotherapy +
    hormone_therapy + tumor_stage + death_from_cancer + age_at_diagnosis:chemotherapy +
    chemotherapy:tumor_stage + hormone_therapy:tumor_stage, data = cancer_new)

```
Residuals:
     Min       1Q   Median       3Q      Max
-194.714  -42.322   -7.887   44.326  236.983

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                             261.3739    12.7534  20.494  < 2e-16 ***
age_at_diagnosis                         -1.2207     0.2155  -5.666 1.88e-08 ***
chemotherapy1                          -114.6748    26.9279  -4.259 2.24e-05 ***
hormone_therapy1                        -24.2402     6.9092  -3.508  0.00047 ***
tumor_stage2                            -20.8586     7.2140  -2.891  0.00391 **
tumor_stage3                            -59.1460    22.7620  -2.598  0.00949 **
death_from_cancerDied of Disease        -80.2741     4.5958 -17.467  < 2e-16 ***
death_from_cancerDied of Other Causes   -22.3403     5.6899  -3.926 9.18e-05 ***
age_at_diagnosis:chemotherapy1            0.9526     0.4210   2.263  0.02385 *
chemotherapy1:tumor_stage2               36.9777    16.9455   2.182  0.02932 *
chemotherapy1:tumor_stage3               48.4794    24.6965   1.963  0.04991 *
hormone_therapy1:tumor_stage2            19.1196     8.9760   2.130  0.03339 *
hormone_therapy1:tumor_stage3            50.1839    20.7534   2.418  0.01577 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 63.89 on 1059 degrees of freedom
Multiple R-squared:  0.3262,    Adjusted R-squared:  0.3185
F-statistic: 42.72 on 12 and 1059 DF,  p-value: < 2.2e-16
```
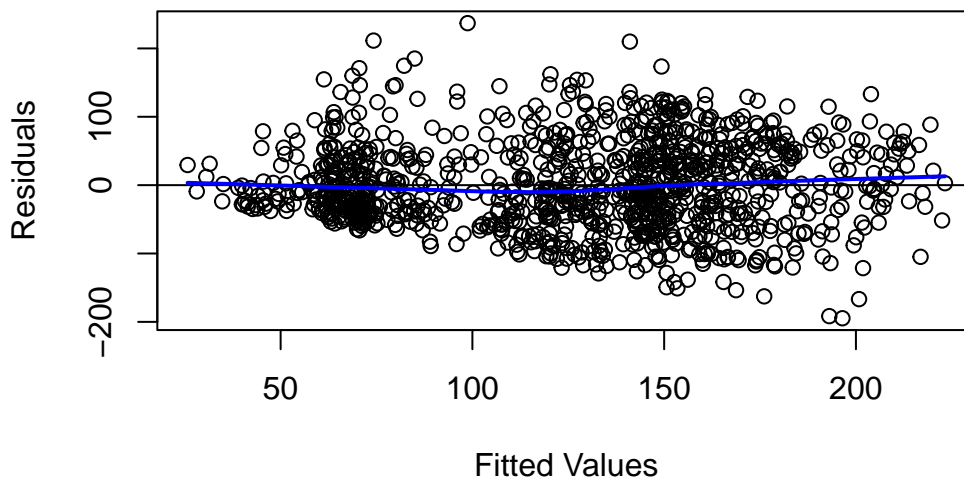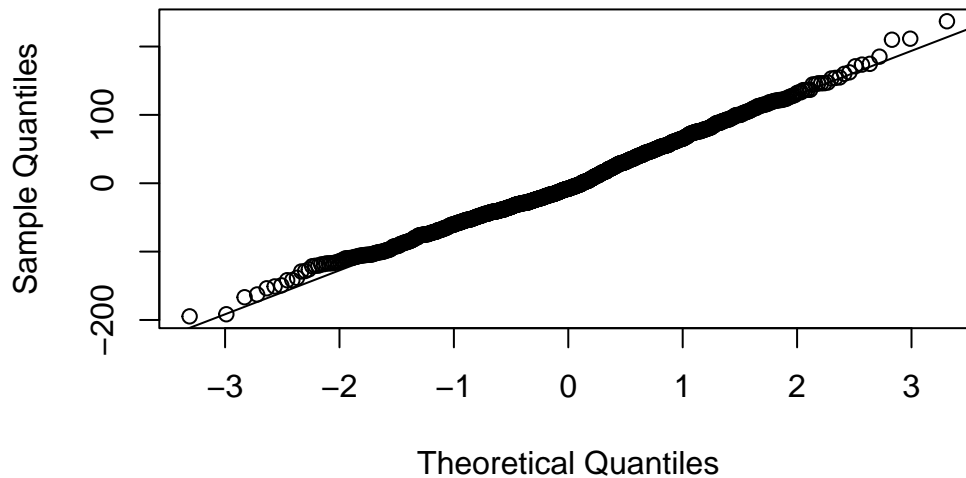
Normal Q–Q Plot

Conclusion

Appendix

- Breast cancer gene expression profiles (METABRIC). Kaggle. (2016, May 10). https://www.kaggle.com/datasets/raghadalharbi/breast-cancer-gene-expression-profiles-metabric

- CBioPortal for Cancer Genomics. (n.d.). https://www.cbioportal.org/study/summary?id=brca_metabric

- DePolo, J. (2024, October 2). Invasive ductal carcinoma (IDC). Breastcancer.org - Breast Cancer Information and Support. https://www.breastcancer.org/types/invasive-ductal-carcinoma

- Wright , P. (2023, March 21). Invasive ductal carcinoma (IDC). Johns Hopkins Medicine. https://www.hopkinsmedicine.org/health/conditions-and-diseases/breast-cancer/invasive-ductal-carcinoma-idc#:~:text=Radiation%20therapy%20might%20be%20part,lymph%20nodes%2C%E2%80%9D%20Wright%20says.

- Tumor size and staging. Susan G. Komen®. (2024, May 2). https://www.komen.org/breast-cancer/diagnosis/stages-staging/tumor-size/#:~:text=Tumor%20size%20is%20related%20to,the%20size%20of%20the%20tumor.