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.

1

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 becon
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
```

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             56.45      BREAST CONSERVING           2            1
4             70.91      BREAST CONSERVING           1            0
5             45.27             MASTECTOMY           2            1
6             51.46      BREAST CONSERVING           2            1
7             70.02      BREAST CONSERVING           2            0
8             66.91             MASTECTOMY           2            0
9             45.43      BREAST CONSERVING           2            1
10            52.14             MASTECTOMY           1            1
   hormone_therapy radio_therapy overall_survival_months
1                1             1               140.50000
2                1             1                84.63333
3                1             1               164.33333
4                1             1               163.53333
5                0             1               164.90000
6                1             1               103.83333
7                1             1               161.06667
8                1             1               160.30000
9                1             1               140.86667
10               0             1               153.96667
```

```
 age_at_diagnosis       type_of_breast_surgery tumor_stage chemotherapy
 Min.   :26.72    BREAST CONSERVING:246        1:190       0:339
 1st Qu.:47.71    MASTECTOMY       :222        2:254       1:129
 Median :55.52                                 3: 24
 Mean   :55.82
```

3

```
3rd Qu.:64.03
Max.   :85.21
hormone_therapy radio_therapy overall_survival_months
0:191           0:126         Min.   :  1.233
1:277           1:342         1st Qu.:111.742
                              Median :165.033
                              Mean   :164.249
                              3rd Qu.:221.400
                              Max.   :337.033
```
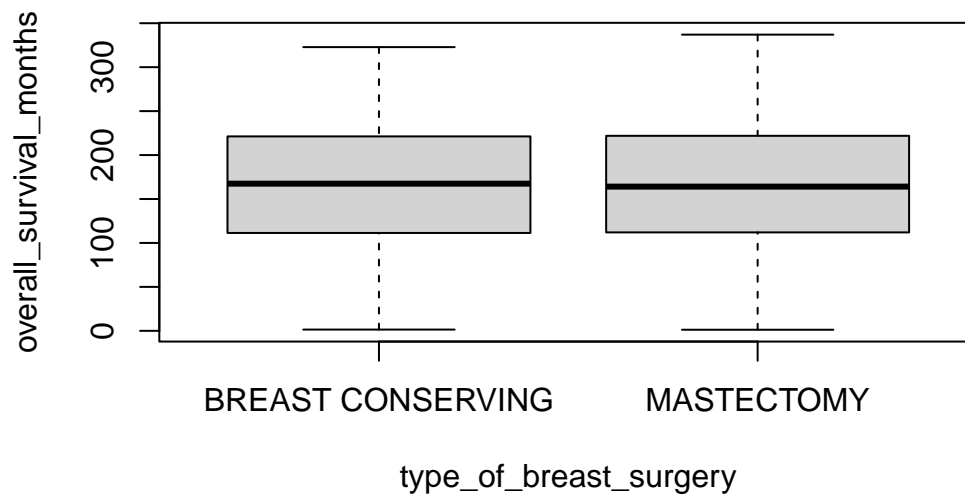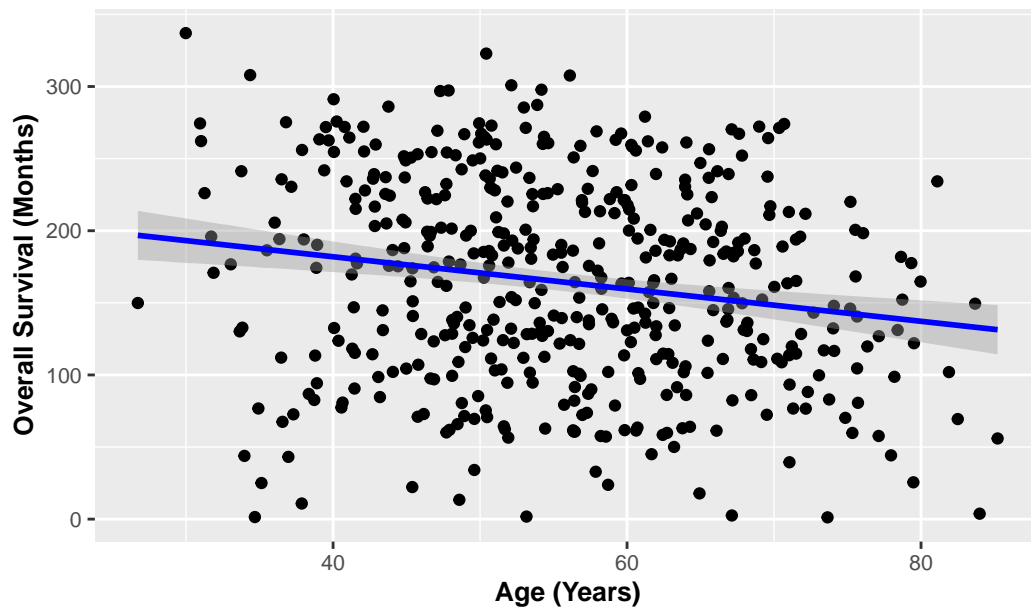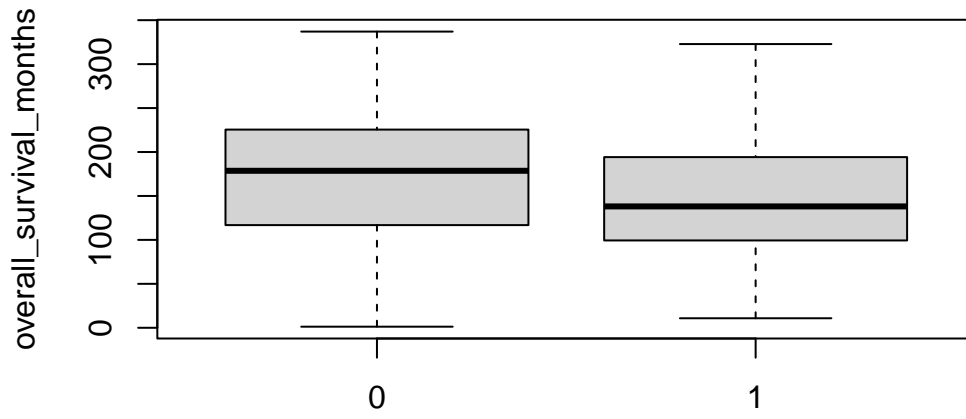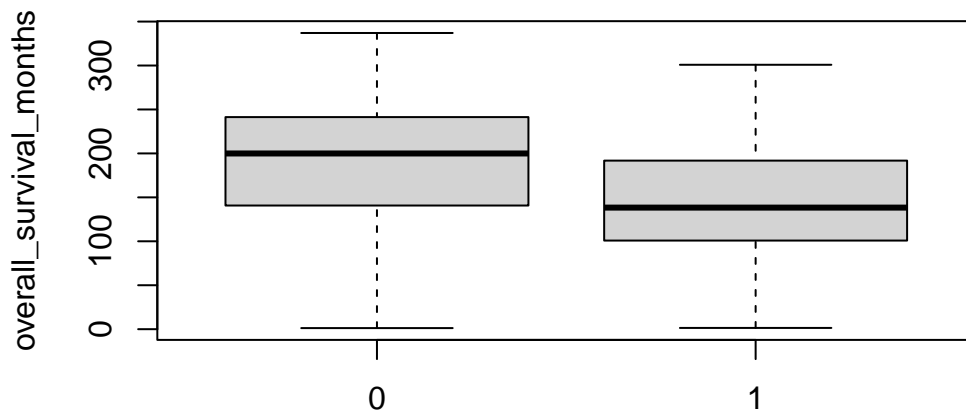
Exploratory data analysis

-

Overall Survival against Age
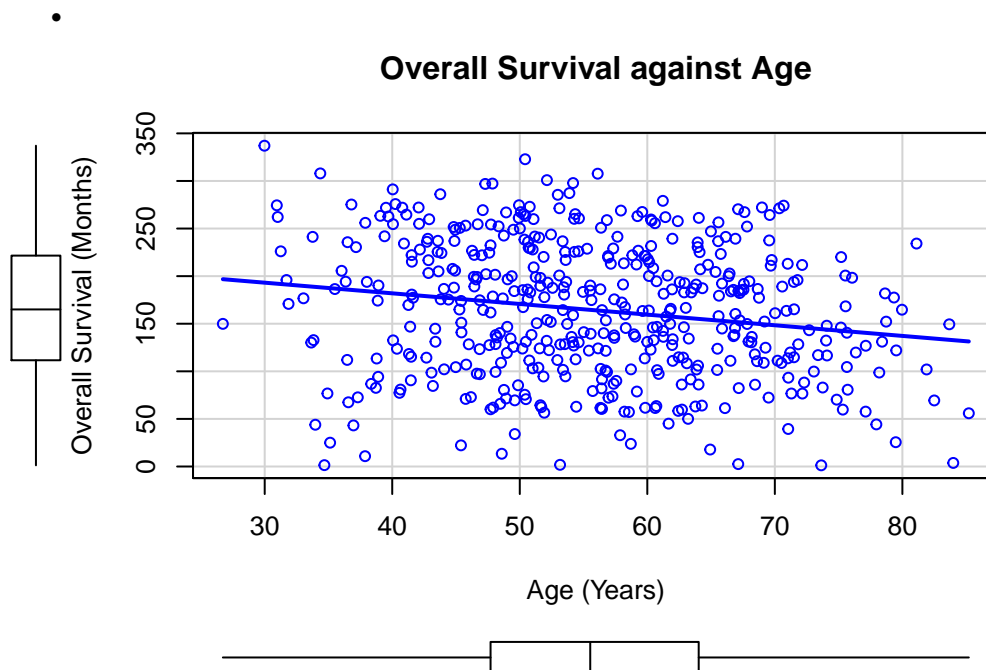
```
`geom_smooth()` using formula = 'y ~ x'
```

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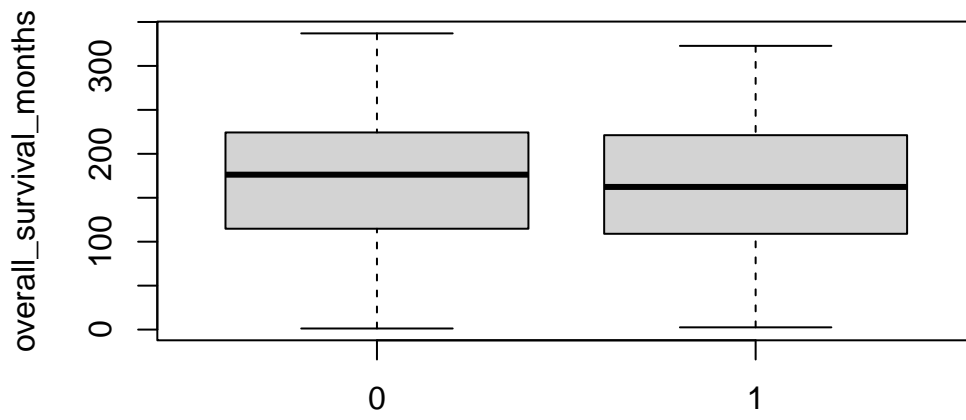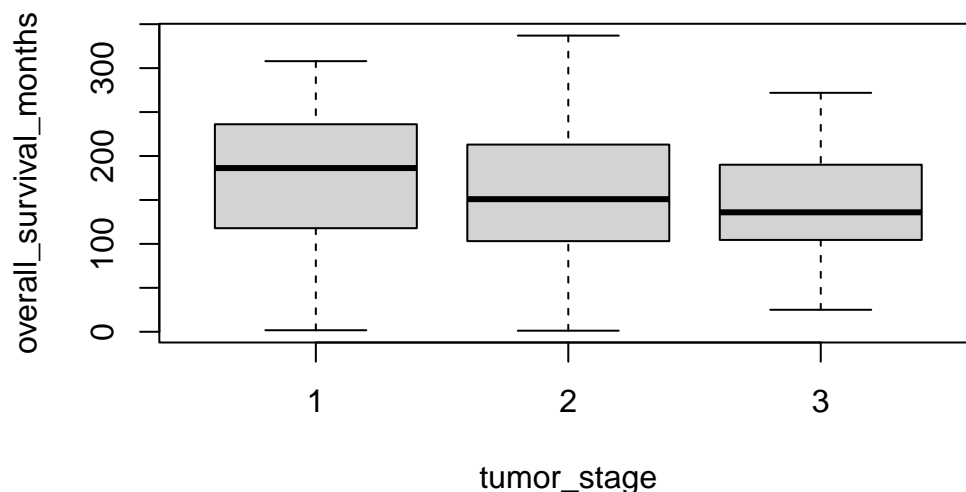
```
Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * type_of_breast_surgery,
    data = cancer_new)

Residuals:
     Min       1Q   Median       3Q      Max
-179.139  -50.337    2.962   53.683  153.534

Coefficients:
                                                      Estimate Std. Error t value
(Intercept)                                           196.6102    23.8003   8.261
age_at_diagnosis                                       -0.5417     0.4235  -1.279
type_of_breast_surgeryMASTECTOMY                       51.2802    31.7724   1.614
age_at_diagnosis:type_of_breast_surgeryMASTECTOMY      -0.9877     0.5582  -1.769
                                                      Pr(>|t|)
(Intercept)                                           1.53e-15 ***
age_at_diagnosis                                        0.2015
type_of_breast_surgeryMASTECTOMY                        0.1072
age_at_diagnosis:type_of_breast_surgeryMASTECTOMY      0.0775 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 68.86 on 464 degrees of freedom
Multiple R-squared:  0.04126,   Adjusted R-squared:  0.03506
F-statistic: 6.656 on 3 and 464 DF,  p-value: 0.0002082
```
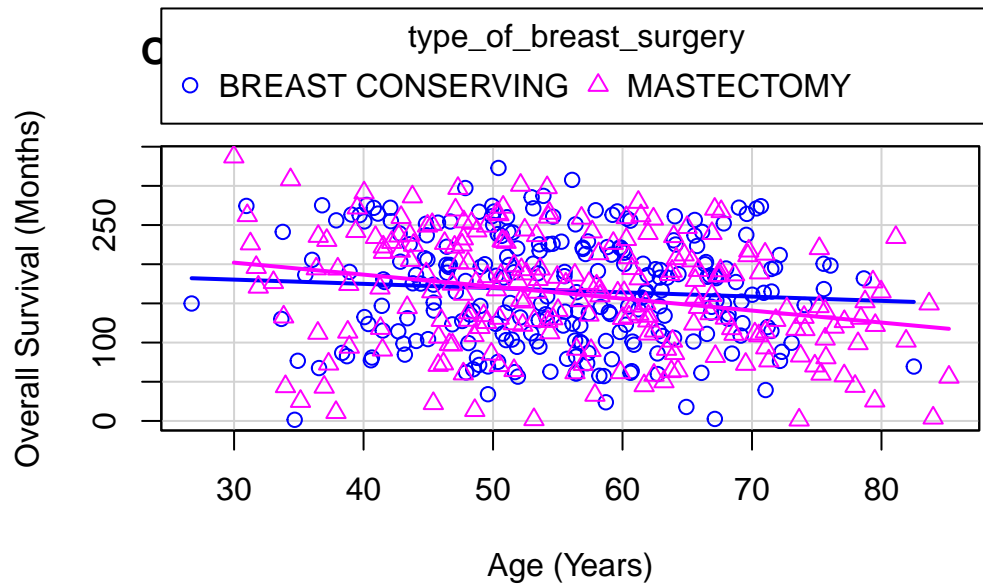
```
Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * chemotherapy,
    data = cancer_new)

Residuals:
     Min        1Q    Median        3Q       Max
-218.438   -50.470     0.868    49.029   172.544

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     292.3982    19.4735  15.015  < 2e-16 ***
age_at_diagnosis                 -2.0913     0.3256  -6.424  3.3e-10 ***
chemotherapy1                  -138.8200    38.6703  -3.590 0.000366 ***
age_at_diagnosis:chemotherapy1    2.0261     0.7576   2.674 0.007752 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 66.9 on 464 degrees of freedom
Multiple R-squared:  0.09527,   Adjusted R-squared:  0.08942
F-statistic: 16.29 on 3 and 464 DF,  p-value: 4.432e-10
```

Overall Survival against Age and Chemotherapy

(plot axes: Overall Survival (Months), Age (Years); legend: chemotherapy, 0, 1)
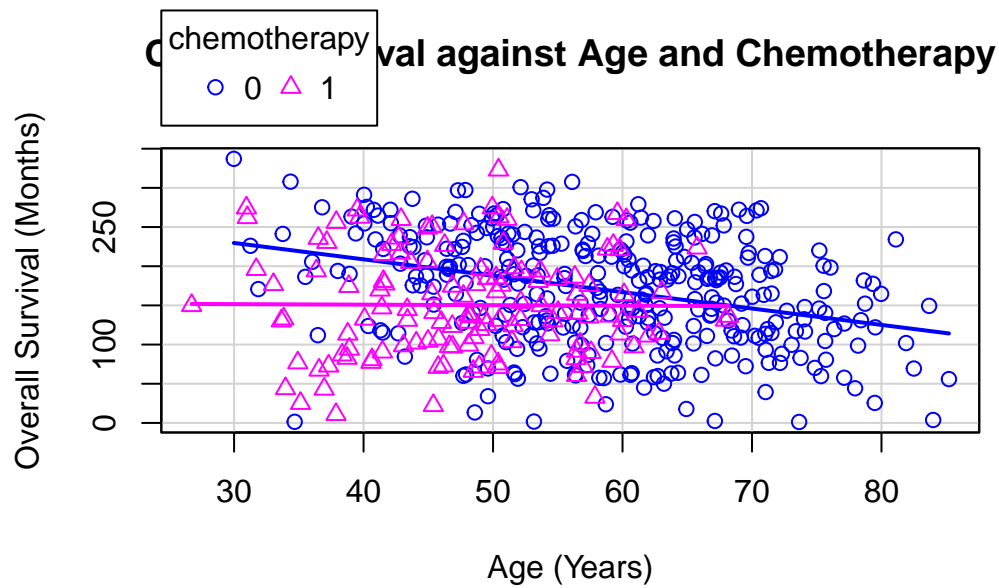
```
Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * hormone_therapy,
    data = cancer_new)

Residuals:
     Min       1Q   Median       3Q      Max
-194.326  -46.759    1.527   48.502  151.428

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     247.4224    22.9894  10.762  < 2e-16 ***
age_at_diagnosis                 -1.1160     0.4314  -2.587  0.00998 **
hormone_therapy1                -77.6135    31.4090  -2.471  0.01383 *
age_at_diagnosis:hormone_therapy1  0.7252     0.5619   1.291  0.19750
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 66.6 on 464 degrees of freedom
Multiple R-squared:  0.1032,    Adjusted R-squared:  0.09741
F-statistic:  17.8 on 3 and 464 DF,  p-value: 5.961e-11
```

9

Ovhormone_therapy**gainst Age and Hormone Therapy**
○ 0 △ 1

Overall Survival (Months)

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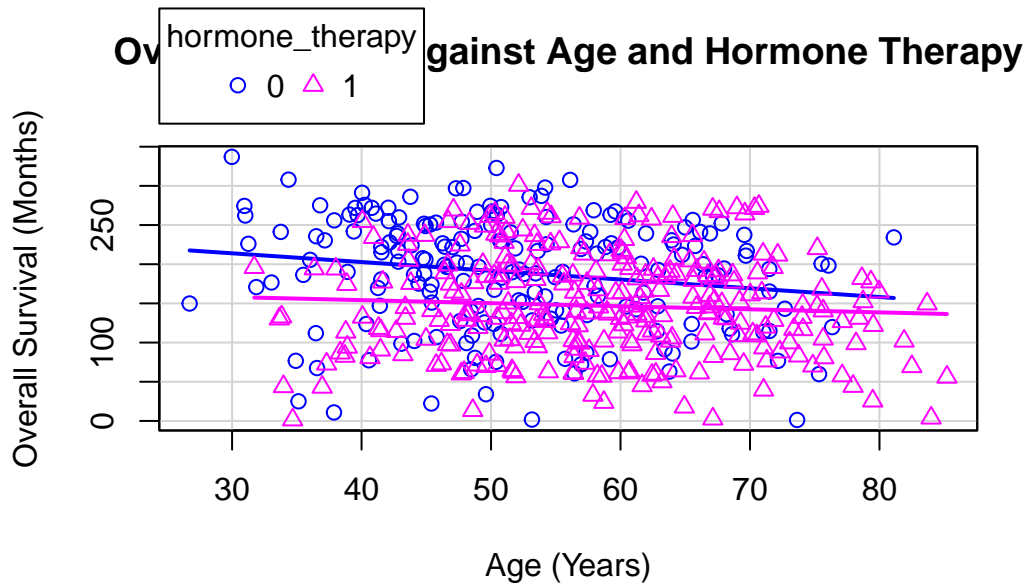
```
Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * radio_therapy,
    data = cancer_new)

Residuals:
     Min       1Q   Median       3Q      Max
-193.595  -49.230    1.223   53.764  154.621

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   237.0139    29.9138   7.923 1.73e-14 ***
age_at_diagnosis               -1.2106     0.5226  -2.316    0.021 *
radio_therapy1                -13.9996    35.1846  -0.398    0.691
age_at_diagnosis:radio_therapy1   0.1237     0.6157   0.201    0.841
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 69.05 on 464 degrees of freedom
Multiple R-squared:  0.03615,   Adjusted R-squared:  0.02992
F-statistic: 5.802 on 3 and 464 DF,  p-value: 0.0006725
```

Overall Survival against Age and Radio Therapy
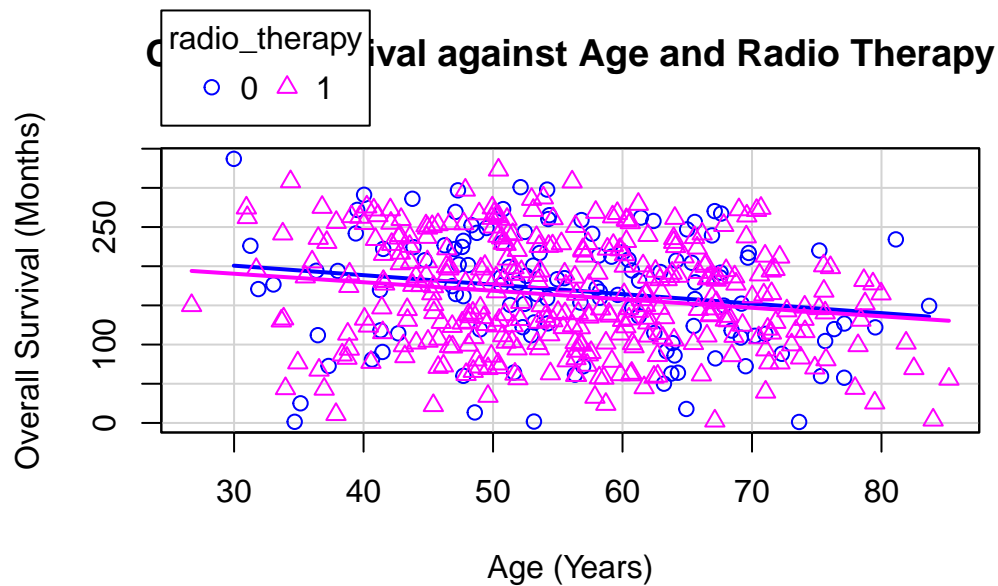
Call:
lm(formula = overall_survival_months ~ age_at_diagnosis * tumor_stage,
    data = cancer_new)

Residuals:
     Min   1Q   Median   3Q   Max
-177.133 -48.121   0.087   54.851 160.598

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)   257.5701   26.5445   9.703 < 2e-16 ***
age_at_diagnosis   -1.4799   0.4727 -3.130 0.00186 **
tumor_stage2   -49.8562   33.4707 -1.490 0.13703
tumor_stage3   -73.0883   69.3013 -1.055 0.29214
age_at_diagnosis:tumor_stage2   0.5779   0.5908   0.978 0.32852
age_at_diagnosis:tumor_stage3   0.8441   1.2009   0.703 0.48249

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 68.56 on 462 degrees of freedom
Multiple R-squared: 0.05377,   Adjusted R-squared: 0.04353
F-statistic: 5.251 on 5 and 462 DF, p-value: 0.0001058
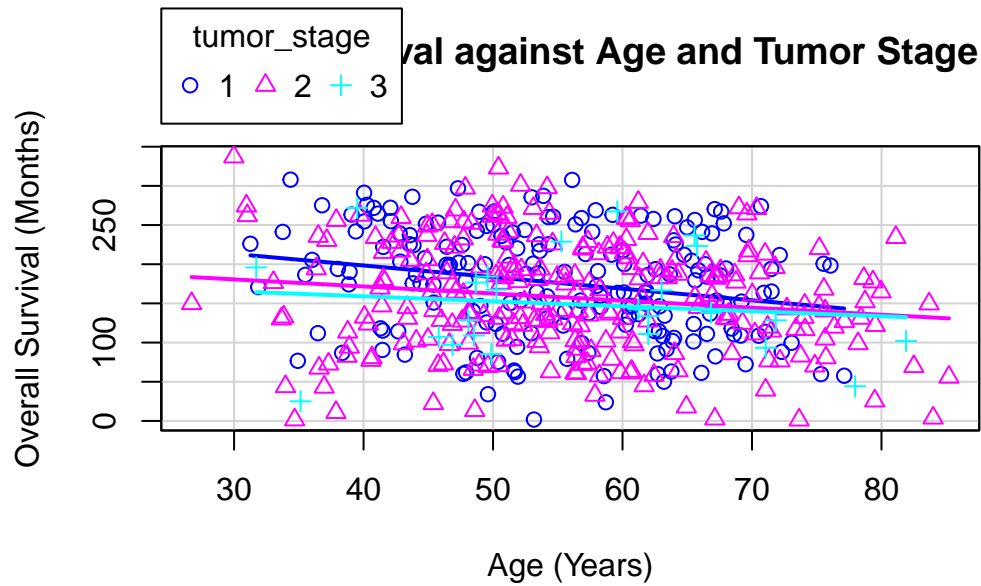
Model selection

-

```
Start:  AIC=3928.4
overall_survival_months ~ (age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy) *
    (age_at_diagnosis + type_of_breast_surgery + tumor_stage +
        chemotherapy + hormone_therapy + radio_therapy)
```

	Df	Sum of Sq	RSS	AIC
- tumor_stage:radio_therapy	2	92	1835603	3924.4
- age_at_diagnosis:tumor_stage	2	1925	1837436	3924.9
- tumor_stage:chemotherapy	2	4239	1839751	3925.5
- tumor_stage:hormone_therapy	2	6116	1841627	3926.0
- type_of_breast_surgery:radio_therapy	1	0	1835512	3926.4
- chemotherapy:hormone_therapy	1	64	1835576	3926.4
- type_of_breast_surgery:hormone_therapy	1	320	1835831	3926.5
- chemotherapy:radio_therapy	1	2460	1837971	3927.0
- hormone_therapy:radio_therapy	1	3017	1838528	3927.2
- type_of_breast_surgery:chemotherapy	1	5589	1841101	3927.8
<none>			1835511	3928.4
- age_at_diagnosis:radio_therapy	1	8689	1844201	3928.6
- age_at_diagnosis:hormone_therapy	1	8950	1844462	3928.7

```
- type_of_breast_surgery:tumor_stage          2      21217 1856729 3929.8
- age_at_diagnosis:chemotherapy               1      27278 1862790 3933.3
- age_at_diagnosis:type_of_breast_surgery     1      32905 1868417 3934.7

Step:  AIC=3924.43
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:tumor_stage +
    age_at_diagnosis:chemotherapy + age_at_diagnosis:hormone_therapy +
    age_at_diagnosis:radio_therapy + type_of_breast_surgery:tumor_stage +
    type_of_breast_surgery:chemotherapy + type_of_breast_surgery:hormone_therapy +
    type_of_breast_surgery:radio_therapy + tumor_stage:chemotherapy +
    tumor_stage:hormone_therapy + chemotherapy:hormone_therapy +
    chemotherapy:radio_therapy + hormone_therapy:radio_therapy

                                            Df Sum of Sq      RSS    AIC
- age_at_diagnosis:tumor_stage               2      2382 1837985 3921.0
- tumor_stage:chemotherapy                   2      4729 1840332 3921.6
- tumor_stage:hormone_therapy                2      6505 1842108 3922.1
- type_of_breast_surgery:radio_therapy       1         0 1835603 3922.4
- chemotherapy:hormone_therapy               1        59 1835662 3922.4
- type_of_breast_surgery:hormone_therapy     1       350 1835953 3922.5
- chemotherapy:radio_therapy                 1      2777 1838380 3923.1
- hormone_therapy:radio_therapy              1      2931 1838534 3923.2
- type_of_breast_surgery:chemotherapy        1      5597 1841200 3923.9
<none>                                                   1835603 3924.4
- age_at_diagnosis:radio_therapy             1      8611 1844214 3924.6
- age_at_diagnosis:hormone_therapy           1      9349 1844952 3924.8
- type_of_breast_surgery:tumor_stage         2     21447 1857050 3925.9
+ tumor_stage:radio_therapy                  2        92 1835511 3928.4
- age_at_diagnosis:chemotherapy              1     27442 1863045 3929.4
- age_at_diagnosis:type_of_breast_surgery    1     32840 1868443 3930.7

Step:  AIC=3921.03
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + type_of_breast_surgery:chemotherapy +
    type_of_breast_surgery:hormone_therapy + type_of_breast_surgery:radio_therapy +
    tumor_stage:chemotherapy + tumor_stage:hormone_therapy +
    chemotherapy:hormone_therapy + chemotherapy:radio_therapy +
    hormone_therapy:radio_therapy
```

```
                                        Df Sum of Sq      RSS    AIC
- tumor_stage:hormone_therapy            2      6680 1844665 3918.7
- type_of_breast_surgery:radio_therapy   1         1 1837986 3919.0
- chemotherapy:hormone_therapy           1         8 1837993 3919.0
- type_of_breast_surgery:hormone_therapy 1       352 1838337 3919.1
- tumor_stage:chemotherapy               2      8833 1846818 3919.3
- hormone_therapy:radio_therapy          1      2837 1840822 3919.8
- chemotherapy:radio_therapy             1      3040 1841025 3919.8
- type_of_breast_surgery:chemotherapy    1      5394 1843379 3920.4
<none>                                             1837985 3921.0
- age_at_diagnosis:hormone_therapy       1      7933 1845918 3921.0
- age_at_diagnosis:radio_therapy         1      9873 1847858 3921.5
- type_of_breast_surgery:tumor_stage     2     19861 1857846 3922.1
+ age_at_diagnosis:tumor_stage           2      2382 1835603 3924.4
+ tumor_stage:radio_therapy              2       549 1837436 3924.9
- age_at_diagnosis:chemotherapy          1     25514 1863499 3925.5
- age_at_diagnosis:type_of_breast_surgery 1    38234 1876219 3928.7

Step:  AIC=3918.73
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + type_of_breast_surgery:chemotherapy +
    type_of_breast_surgery:hormone_therapy + type_of_breast_surgery:radio_therapy +
    tumor_stage:chemotherapy + chemotherapy:hormone_therapy +
    chemotherapy:radio_therapy + hormone_therapy:radio_therapy

                                        Df Sum of Sq      RSS    AIC
- tumor_stage:chemotherapy               2      7719 1852384 3916.7
- chemotherapy:hormone_therapy           1         1 1844666 3916.7
- type_of_breast_surgery:radio_therapy   1         3 1844668 3916.7
- type_of_breast_surgery:hormone_therapy 1       379 1845045 3916.8
- chemotherapy:radio_therapy             1      2246 1846911 3917.3
- hormone_therapy:radio_therapy          1      2728 1847394 3917.4
- type_of_breast_surgery:chemotherapy    1      5033 1849699 3918.0
- age_at_diagnosis:hormone_therapy       1      7492 1852157 3918.6
<none>                                             1844665 3918.7
- type_of_breast_surgery:tumor_stage     2     16708 1861373 3919.0
- age_at_diagnosis:radio_therapy         1     11197 1855863 3919.6
+ tumor_stage:hormone_therapy            2      6680 1837985 3921.0
+ age_at_diagnosis:tumor_stage           2      2558 1842108 3922.1
```

```
+ tumor_stage:radio_therapy                    2        58 1844607 3922.7
- age_at_diagnosis:chemotherapy                1     25140 1869805 3923.1
- age_at_diagnosis:type_of_breast_surgery  1     39516 1884181 3926.7

Step:  AIC=3916.69
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + type_of_breast_surgery:chemotherapy +
    type_of_breast_surgery:hormone_therapy + type_of_breast_surgery:radio_therapy +
    chemotherapy:hormone_therapy + chemotherapy:radio_therapy +
    hormone_therapy:radio_therapy

                                              Df Sum of Sq      RSS     AIC
- type_of_breast_surgery:radio_therapy     1         4 1852389 3914.7
- chemotherapy:hormone_therapy             1        50 1852435 3914.7
- type_of_breast_surgery:hormone_therapy   1       296 1852680 3914.8
- chemotherapy:radio_therapy               1      2298 1854683 3915.3
- hormone_therapy:radio_therapy            1      2789 1855173 3915.4
- type_of_breast_surgery:chemotherapy      1      5333 1857718 3916.0
<none>                                                   1852384 3916.7
- age_at_diagnosis:hormone_therapy         1      8293 1860678 3916.8
- type_of_breast_surgery:tumor_stage       2     16939 1869324 3916.9
- age_at_diagnosis:radio_therapy           1     12508 1864892 3917.8
+ tumor_stage:chemotherapy                 2      7719 1844665 3918.7
+ tumor_stage:hormone_therapy              2      5567 1846818 3919.3
+ age_at_diagnosis:tumor_stage             2      5354 1847031 3919.3
+ tumor_stage:radio_therapy                2       394 1851991 3920.6
- age_at_diagnosis:chemotherapy            1     27438 1879822 3921.6
- age_at_diagnosis:type_of_breast_surgery  1     42618 1895002 3925.3

Step:  AIC=3914.69
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + type_of_breast_surgery:chemotherapy +
    type_of_breast_surgery:hormone_therapy + chemotherapy:hormone_therapy +
    chemotherapy:radio_therapy + hormone_therapy:radio_therapy

                                              Df Sum of Sq      RSS     AIC
- chemotherapy:hormone_therapy             1        49 1852437 3912.7
```

```
- type_of_breast_surgery:hormone_therapy    1          296 1852685 3912.8
- chemotherapy:radio_therapy                1         2568 1854957 3913.3
- hormone_therapy:radio_therapy             1         2799 1855188 3913.4
- type_of_breast_surgery:chemotherapy       1         5873 1858262 3914.2
<none>                                                     1852389 3914.7
- age_at_diagnosis:hormone_therapy          1         8289 1860678 3914.8
- type_of_breast_surgery:tumor_stage        2        16952 1869341 3915.0
- age_at_diagnosis:radio_therapy            1        13316 1865705 3916.0
+ type_of_breast_surgery:radio_therapy      1            4 1852384 3916.7
+ tumor_stage:chemotherapy                  2         7721 1844668 3916.7
+ tumor_stage:hormone_therapy               2         5571 1846818 3917.3
+ age_at_diagnosis:tumor_stage              2         5357 1847032 3917.3
+ tumor_stage:radio_therapy                 2          382 1852007 3918.6
- age_at_diagnosis:chemotherapy             1        27765 1880154 3919.6
- age_at_diagnosis:type_of_breast_surgery   1        44775 1897164 3923.9

Step:  AIC=3912.7
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + type_of_breast_surgery:chemotherapy +
    type_of_breast_surgery:hormone_therapy + chemotherapy:radio_therapy +
    hormone_therapy:radio_therapy

                                           Df Sum of Sq      RSS    AIC
- type_of_breast_surgery:hormone_therapy    1          460 1852897 3910.8
- chemotherapy:radio_therapy                1         2539 1854976 3911.3
- hormone_therapy:radio_therapy             1         3038 1855476 3911.5
- type_of_breast_surgery:chemotherapy       1         5826 1858263 3912.2
<none>                                                     1852437 3912.7
- type_of_breast_surgery:tumor_stage        2        16905 1869342 3913.0
- age_at_diagnosis:hormone_therapy          1         8995 1861432 3913.0
- age_at_diagnosis:radio_therapy            1        13356 1865794 3914.1
+ chemotherapy:hormone_therapy              1           49 1852389 3914.7
+ type_of_breast_surgery:radio_therapy      1            3 1852435 3914.7
+ tumor_stage:chemotherapy                  2         7769 1844669 3914.7
+ tumor_stage:hormone_therapy               2         5230 1847207 3915.4
+ age_at_diagnosis:tumor_stage              2         5135 1847303 3915.4
+ tumor_stage:radio_therapy                 2          377 1852060 3916.6
- age_at_diagnosis:chemotherapy             1        28283 1880721 3917.8
- age_at_diagnosis:type_of_breast_surgery   1        44760 1897197 3921.9
```

```
Step:  AIC=3910.82
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + type_of_breast_surgery:chemotherapy +
    chemotherapy:radio_therapy + hormone_therapy:radio_therapy

                                           Df Sum of Sq      RSS    AIC
- chemotherapy:radio_therapy                1      2720  1855617 3909.5
- hormone_therapy:radio_therapy             1      6093  1858991 3910.4
- type_of_breast_surgery:chemotherapy       1      6563  1859460 3910.5
<none>                                                   1852897 3910.8
- age_at_diagnosis:hormone_therapy          1      8672  1861569 3911.0
- type_of_breast_surgery:tumor_stage        2     16994  1869891 3911.1
- age_at_diagnosis:radio_therapy            1     13090  1865988 3912.1
+ type_of_breast_surgery:hormone_therapy    1       460  1852437 3912.7
+ chemotherapy:hormone_therapy              1       213  1852685 3912.8
+ type_of_breast_surgery:radio_therapy      1         1  1852896 3912.8
+ tumor_stage:chemotherapy                  2      7801  1845096 3912.8
+ age_at_diagnosis:tumor_stage              2      4955  1847942 3913.6
+ tumor_stage:hormone_therapy               2      4941  1847956 3913.6
+ tumor_stage:radio_therapy                 2       417  1852480 3914.7
- age_at_diagnosis:chemotherapy             1     28411  1881308 3915.9
- age_at_diagnosis:type_of_breast_surgery   1     44438  1897335 3919.9

Step:  AIC=3909.5
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + type_of_breast_surgery:chemotherapy +
    hormone_therapy:radio_therapy

                                           Df Sum of Sq      RSS    AIC
- type_of_breast_surgery:chemotherapy       1      4398  1860015 3908.6
- hormone_therapy:radio_therapy             1      6019  1861637 3909.0
<none>                                                   1855617 3909.5
- age_at_diagnosis:hormone_therapy          1      8279  1863896 3909.6
- type_of_breast_surgery:tumor_stage        2     17989  1873606 3910.0
- age_at_diagnosis:radio_therapy            1     10396  1866013 3910.1
+ chemotherapy:radio_therapy                1      2720  1852897 3910.8
+ type_of_breast_surgery:hormone_therapy    1       641  1854976 3911.3
```

```
+ type_of_breast_surgery:radio_therapy       1       329 1855289 3911.4
+ chemotherapy:hormone_therapy               1       186 1855431 3911.5
+ tumor_stage:chemotherapy                   2      7809 1847808 3911.5
+ age_at_diagnosis:tumor_stage               2      5113 1850504 3912.2
+ tumor_stage:hormone_therapy                2      4259 1851359 3912.4
+ tumor_stage:radio_therapy                  2        60 1855557 3913.5
- age_at_diagnosis:chemotherapy              1     25717 1881334 3913.9
- age_at_diagnosis:type_of_breast_surgery    1     41739 1897356 3917.9

Step:  AIC=3908.61
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage + hormone_therapy:radio_therapy

                                            Df Sum of Sq      RSS    AIC
- hormone_therapy:radio_therapy              1      5809 1865824 3908.1
<none>                                                    1860015 3908.6
- age_at_diagnosis:hormone_therapy           1      8597 1868612 3908.8
- age_at_diagnosis:radio_therapy             1      8618 1868633 3908.8
+ type_of_breast_surgery:chemotherapy        1      4398 1855617 3909.5
- type_of_breast_surgery:tumor_stage         2     19912 1879927 3909.6
+ type_of_breast_surgery:hormone_therapy     1      1219 1858796 3910.3
+ type_of_breast_surgery:radio_therapy       1       977 1859038 3910.4
+ chemotherapy:radio_therapy                 1       555 1859460 3910.5
+ chemotherapy:hormone_therapy               1       133 1859882 3910.6
+ tumor_stage:chemotherapy                   2      7993 1852022 3910.6
+ age_at_diagnosis:tumor_stage               2      5048 1854967 3911.3
+ tumor_stage:hormone_therapy                2      4357 1855658 3911.5
+ tumor_stage:radio_therapy                  2        40 1859975 3912.6
- age_at_diagnosis:chemotherapy              1     24919 1884934 3912.8
- age_at_diagnosis:type_of_breast_surgery    1     39045 1899060 3916.3

Step:  AIC=3908.07
overall_survival_months ~ age_at_diagnosis + type_of_breast_surgery +
    tumor_stage + chemotherapy + hormone_therapy + radio_therapy +
    age_at_diagnosis:type_of_breast_surgery + age_at_diagnosis:chemotherapy +
    age_at_diagnosis:hormone_therapy + age_at_diagnosis:radio_therapy +
    type_of_breast_surgery:tumor_stage

                                            Df Sum of Sq      RSS    AIC
<none>                                                    1865824 3908.1
```

```
- age_at_diagnosis:hormone_therapy          1     10061 1875885 3908.6
+ hormone_therapy:radio_therapy             1      5809 1860015 3908.6
+ type_of_breast_surgery:hormone_therapy    1      4810 1861014 3908.9
- age_at_diagnosis:radio_therapy            1     11407 1877231 3908.9
+ type_of_breast_surgery:chemotherapy       1      4187 1861637 3909.0
- type_of_breast_surgery:tumor_stage        2     21163 1886987 3909.3
+ type_of_breast_surgery:radio_therapy      1       699 1865125 3909.9
+ chemotherapy:radio_therapy                1       552 1865272 3909.9
+ chemotherapy:hormone_therapy              1        78 1865746 3910.0
+ tumor_stage:chemotherapy                  2      7507 1858317 3910.2
+ age_at_diagnosis:tumor_stage              2      5614 1860210 3910.7
+ tumor_stage:hormone_therapy               2      4643 1861181 3910.9
+ tumor_stage:radio_therapy                 2       181 1865643 3912.0
- age_at_diagnosis:chemotherapy             1     25231 1891055 3912.4
- age_at_diagnosis:type_of_breast_surgery   1     38408 1904233 3915.6


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

Residuals:
     Min       1Q   Median       3Q      Max
-197.554  -43.883    4.063   47.466  163.426

Coefficients:
                                                    Estimate Std. Error t value
(Intercept)                                         218.4268    41.4618   5.268
age_at_diagnosis                                     -0.5964     0.7247  -0.823
type_of_breast_surgeryMASTECTOMY                    107.6380    34.8111   3.092
tumor_stage2                                          3.8998     9.1429   0.427
tumor_stage3                                         49.2187    26.0629   1.888
chemotherapy1                                      -142.0485    40.2583  -3.528
hormone_therapy1                                    -84.8574    31.4836  -2.695
radio_therapy1                                       67.8692    37.9016   1.791
age_at_diagnosis:type_of_breast_surgeryMASTECTOMY    -1.7878     0.5848  -3.057
age_at_diagnosis:chemotherapy1                        1.9085     0.7702   2.478
age_at_diagnosis:hormone_therapy1                     0.8895     0.5685   1.565
age_at_diagnosis:radio_therapy1                      -1.0887     0.6535  -1.666
```

```
type_of_breast_surgeryMASTECTOMY:tumor_stage2        5.8723    12.9730    0.453
type_of_breast_surgeryMASTECTOMY:tumor_stage3      -63.7049    31.4736   -2.024
                                                  Pr(>|t|)

(Intercept)                                       2.13e-07 ***
age_at_diagnosis                                  0.410940
type_of_breast_surgeryMASTECTOMY                  0.002110 **
tumor_stage2                                      0.669916
tumor_stage3                                      0.059602 .
chemotherapy1                                     0.000461 ***
hormone_therapy1                                  0.007294 **
radio_therapy1                                    0.074013 .
age_at_diagnosis:type_of_breast_surgeryMASTECTOMY 0.002367 **
age_at_diagnosis:chemotherapy1                    0.013584 *
age_at_diagnosis:hormone_therapy1                 0.118371
age_at_diagnosis:radio_therapy1                   0.096397 .
type_of_breast_surgeryMASTECTOMY:tumor_stage2     0.651013
type_of_breast_surgeryMASTECTOMY:tumor_stage3     0.043547 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 64.11 on 454 degrees of freedom
Multiple R-squared:  0.1871,    Adjusted R-squared:  0.1638
F-statistic: 8.035 on 13 and 454 DF,  p-value: 1.445e-14
```
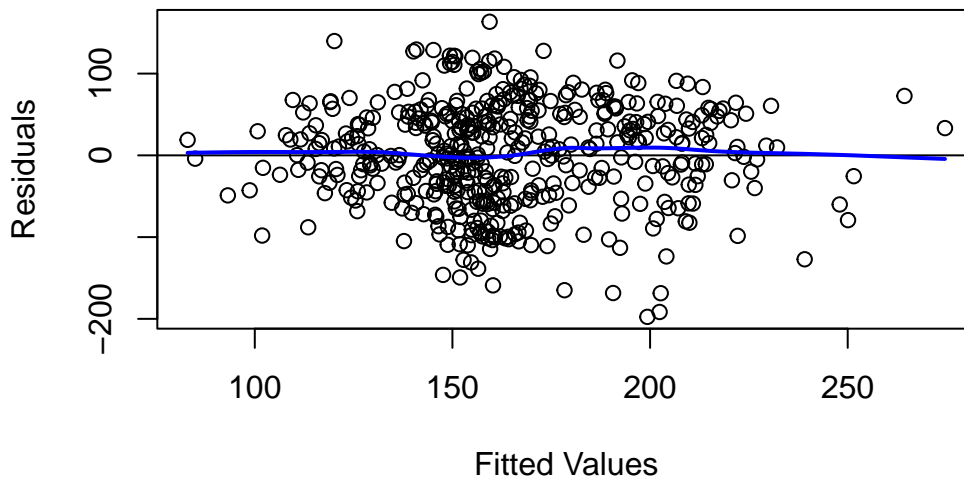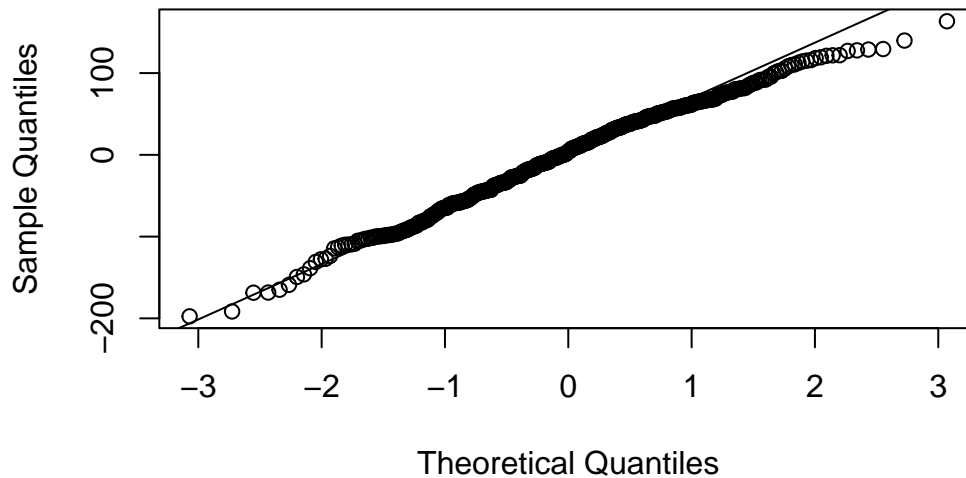
Normal Q–Q Plot

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.