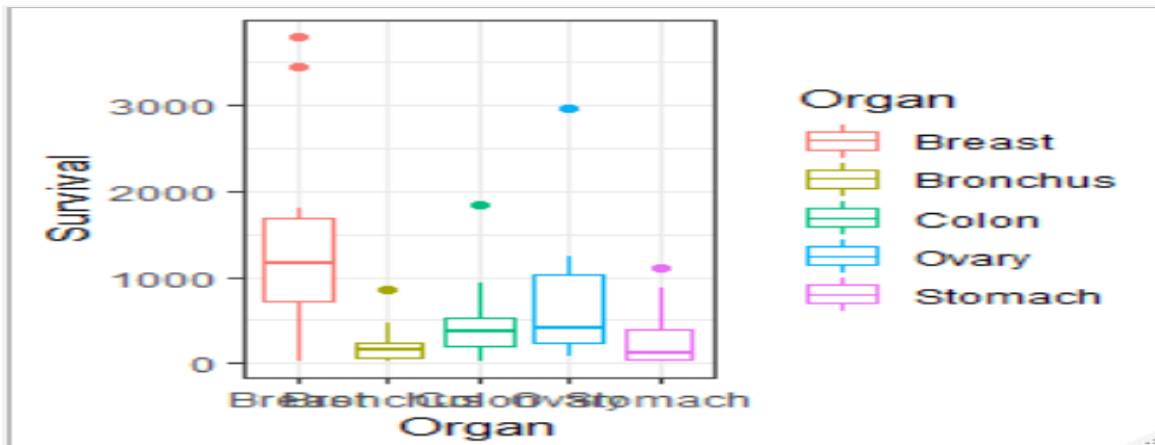
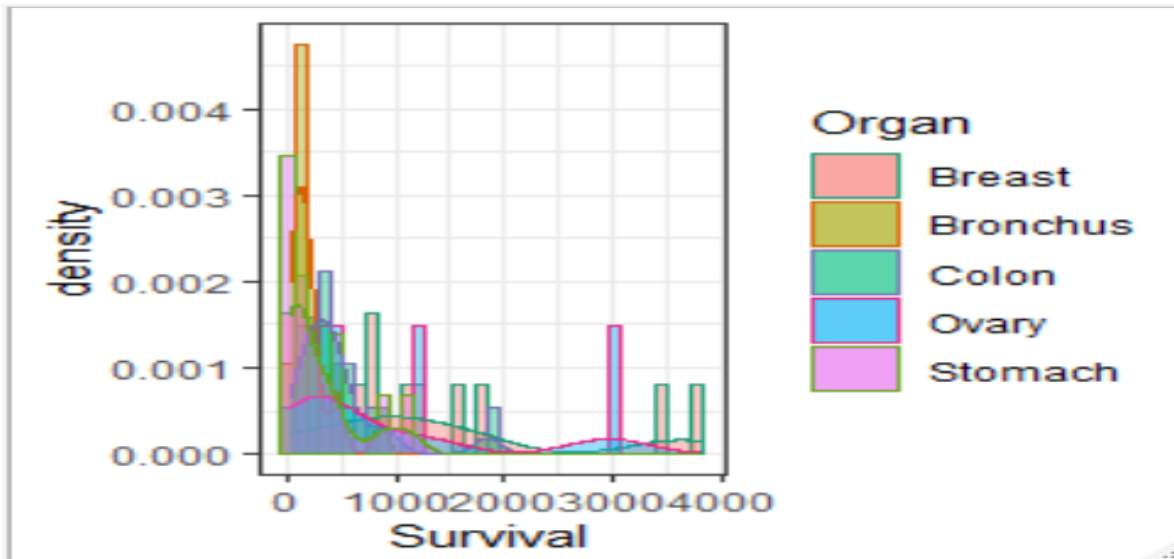


## Exploratory Data Analysis:

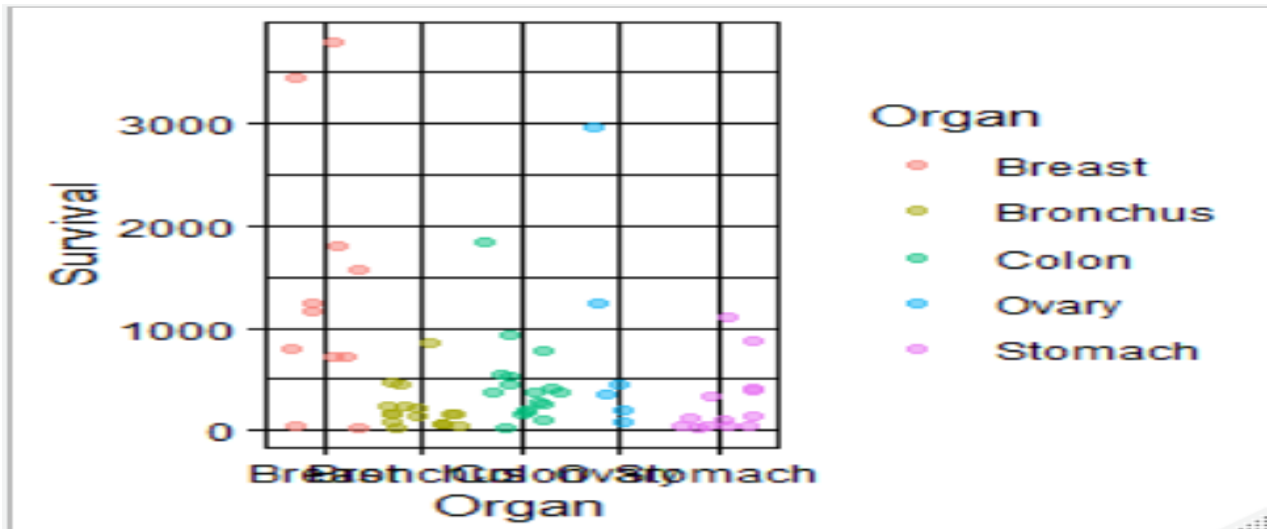
Box Plot shows that there are some extreme outliers for the breast cancer, also the median of the breast cancer is higher than other organ cancer. Among all the organ cancers from the box plot it is estimated that the Bronchus cancer seems more consistence in term of its variance and Breast cancer and Ovary cancer seem to have wider variance and inconsistency contrasting others.



Density and Histogram plot shows that the data is highly positive skewed and it may need some transformation. The most frequency survival seen data points is Bronchus cancer and Breast cancer.



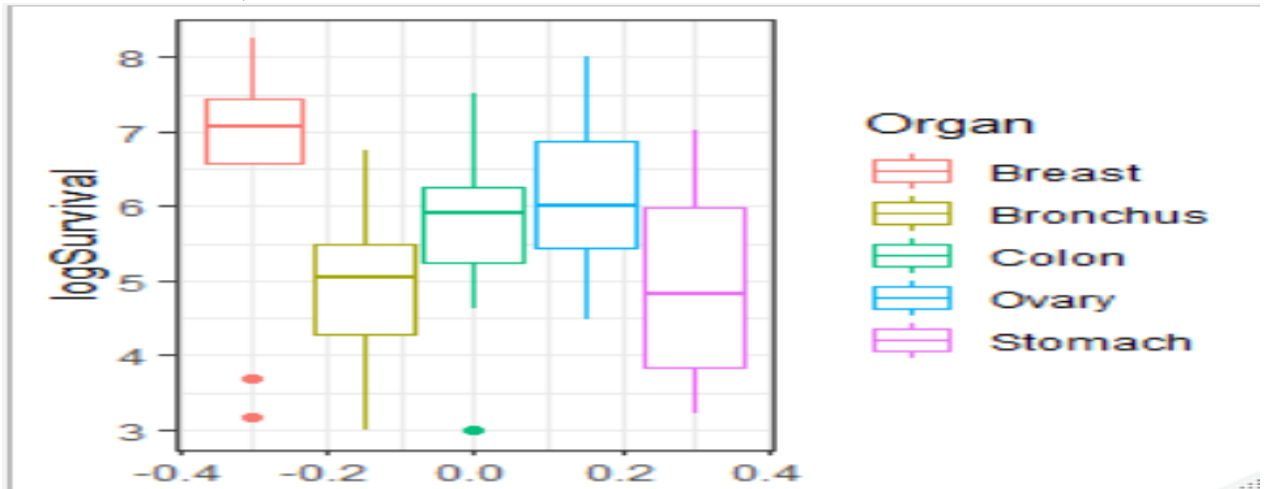
Scatter Plot confirms that the consistency of Breast in term of its Survival is lesser than other groups and also the three cancer groups of Bronchus and Colon and Stomach are seemed same behavior. Also, the extreme outliers in Breast cancer and Ovary cancer is targeted.



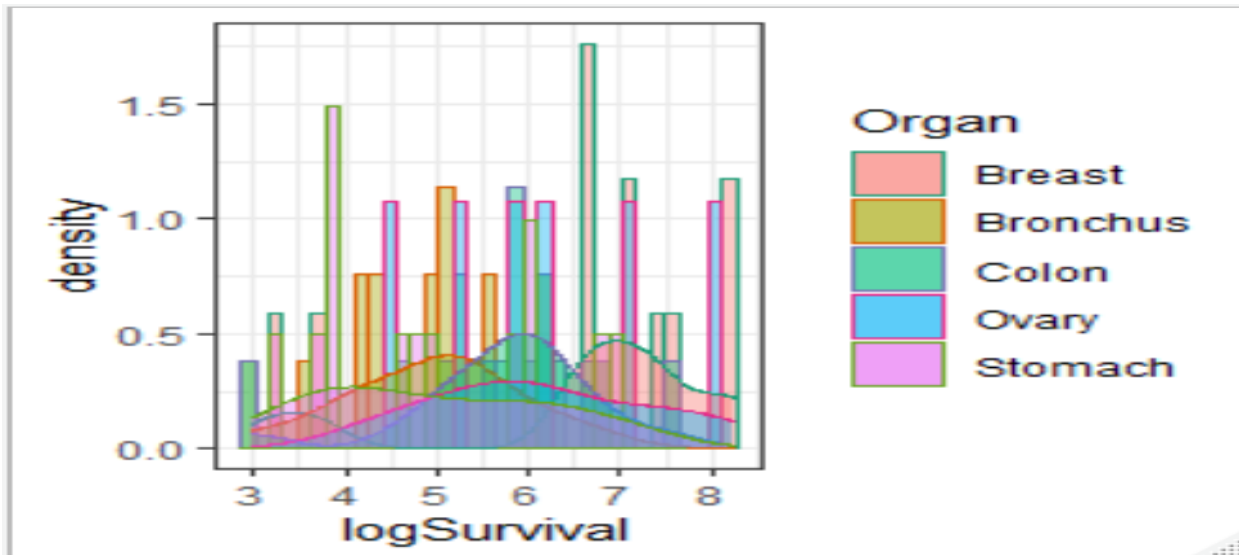
Transforming the cancer data to get a log transformation to Survival variable.

Exploratory Data Analysis:

Box Plot shows that the extreme outliers have eliminated as well as skewness of the data seems to be more normal than before transformation. Also, among the two Ovary and Colon cancer, they seem to be more closer their median and also Stomach and Bronchus's medians are more close to each other, but breast cancer.



Density and Histogram after log transformation precisely suggest that the cancer dataset is now almost has a normal distribution and without outliers. After log transformation on Survival variables, the plot shows that frequency data points in all cancer groups approximately the same, but there seems that Breast cancer is a bit more frequent than others.



Examination of the Data by fitting the Cancer dataset into a linear model:

RESULTS OF SUMMARY:

```
Call:
lm(formula = survival ~ organ, data = cancer)

Residuals:
    Min       1Q   Median       3Q      Max
-1371.91  -241.75  -111.50    87.19   2412.09

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1395.9     201.9    6.915 3.77e-09 ***
OrganBronchus -1184.3     259.1   -4.571 2.53e-05 ***
OrganColon    -938.5     259.1   -3.622 0.000608 ***
OrganOvary    -511.6     339.8   -1.506 0.137526
OrganStomach  -1109.9     274.3   -4.046 0.000153 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 669.5 on 59 degrees of freedom
Multiple R-squared:  0.3037, Adjusted R-squared:  0.2565
F-statistic: 6.433 on 4 and 59 DF, p-value: 0.0002295
```

Examination of the Data by fitting the Cancer dataset AFTER LOG TRANSFORMATION into a linear model:

RESULTS OF SUMMARY:

```
Call:
lm(formula = logSurvival ~ organ, data = logTransformedSurvavial)
```

```

Residuals:
    Min       1Q   Median       3Q      Max
-3.3805 -0.6607  0.1025  0.8207  2.0460

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.5586    0.3603  18.201 < 2e-16 ***
OrganBronchus -1.6054    0.4625  -3.472 0.000975 ***
OrganColon    -0.8095    0.4625  -1.750 0.085247 .
OrganOvary    -0.4080    0.6065  -0.673 0.503801 .
OrganStomach  -1.5907    0.4896  -3.249 0.001915 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.195 on 59 degrees of freedom
Multiple R-squared:  0.2252,    Adjusted R-squared:  0.1726
F-statistic: 4.286 on 4 and 59 DF,  p-value: 0.004122

```

From fitting both Cancer dataset and Log Transformed Cancer data set and comparing the p values, it seems sticking to Cancer dataset with itself and further tests on original Cancer data set can draw approximately the same results from log transformed data. Therefore, I will stick with the original Cancer dataset, but further tests on it to see the comparing and contrasting between and within the groups.

### Least Significant Means Difference Test on Fitted Linear Model

Results:

```

Organ    lsmmean    SE df lower.CL upper.CL
Breast    1396    202 59    992.0    1800
Bronchus   212    162 59   -113.3     537
Colon     457    162 59    132.5     782
Ovary     884    273 59    337.4    1431
Stomach    286    186 59   -85.6     658

Confidence level used: 0.95

```

### Contrasting the groups Least Significant Difference

Results:

```

contrast      estimate    SE df t.ratio p.value
Breast - Bronchus    1184.3    259 59   4.571 <.0001
Breast - Colon       938.5    259 59   3.622 0.0006
Breast - Ovary       511.6    340 59   1.506 0.1375

```

Breast - Stomach	1109.9	274	59	4.046	0.0002
Bronchus - Colon	-245.8	230	59	-1.070	0.2888
Bronchus - Ovary	-672.7	318	59	-2.116	0.0386
Bronchus - Stomach	-74.4	247	59	-0.302	0.7640
Colon - Ovary	-426.9	318	59	-1.343	0.1845
Colon - Stomach	171.4	247	59	0.695	0.4899
Ovary - Stomach	598.3	330	59	1.811	0.0753

### Confidence Intervals 95% Summary:

contrast	estimate	SE	df	lower.CL	upper.CL
Breast - Bronchus	1184.3	259	59	665.9	1702.7
Breast - Colon	938.5	259	59	420.1	1456.9
Breast - Ovary	511.6	340	59	-168.4	1191.5
Breast - Stomach	1109.9	274	59	561.1	1658.8
Bronchus - Colon	-245.8	230	59	-705.3	213.7
Bronchus - Ovary	-672.7	318	59	-1308.9	-36.6
Bronchus - Stomach	-74.4	247	59	-568.0	419.2
Colon - Ovary	-426.9	318	59	-1063.1	209.3
Colon - Stomach	171.4	247	59	-322.2	665.0
Ovary - Stomach	598.3	330	59	-62.9	1259.6

Outcomes of research for the Cancer dataset, determining if the survival times differ with respect to the organ affected by the cancer study, extracted from <https://math.tntech.edu/e-stat/DASL/page24.html>

- Contrasting the Survival time and Organ affected by cancer of two groups, Breast and Bronchus cancers shows that there is a strong convincing evidence that increase was estimated to 1184 days from 666 to 1702 days with the 95% confidence. The p-value < .0001.
- Contrasting the Survival time and Organ affected by cancer of two groups, Breast and Colon shows that there is strong convincing evidence that increase was estimated to 938.5 days from 420 to 1456 days with 95% confidence. The p-value is equal to 0.0006.
- Contrasting the Survival time and Organ affected by cancer among the two groups of Breast and Ovary shows that there is a suggestive but not inclusive evidence that increase was estimated to 511 days from -168 to 1191 days with 95% confidence interval. The p-value is equal to 0.1.
- Contrasting the Survival time and Organ affected by cancer of two groups, Breast and Stomach shows that there is a strong convincing evidence that the increase was estimated to 1109 days from 561 to 1659 days with 95% confidence. The p-value equal to 0.0002.
- Contrasting the Survival time and Organ affected by cancer of two groups, Bronchus and colon shows that there is NO evidence that decrease estimated to 245 days from -705 to 214 days of 95 % confidence. P-value equals to 0.3.
- Contrasting the Survival time and Organ affected by cancer of two groups, Bronchus and Ovary shows that there is a moderate evidence that increase was estimated to 672 days from 36.6 to 1308 days of 95 % confidence interval. The p-value equals to 0.04.

- Contrasting the Survival time and Organ affected by cancer of two groups, Bronchus and Stomach shows that there is NO evidence that increase was estimated 74 days from -568 to 419 days with 95% confidence. P-value equals to 0.8.
- Contrasting the Survival time and Organ affected by cancer of two groups, Colon and Ovary shows that there is No evidence that increase estimated to 426 days from -1063 to 209 with 95% confidence. The p-value equals to 0.2.
- Contrasting the Survival time and Organ affected by cancer of two groups, Colon and stomach shows that there is NO evidence that the increase was estimated to 171 days from -322 to 665 days with 95% confidence. P-value equals to 0.5.
- Contrasting the Survival time and Organ affected by cancer of two groups, Ovary and Stomach shows that there is a suggestive but inconclusive evidence that the increase was estimated to 598 days from -63 to 1259.6 days with 95 % confidence-value equals to 0.07.