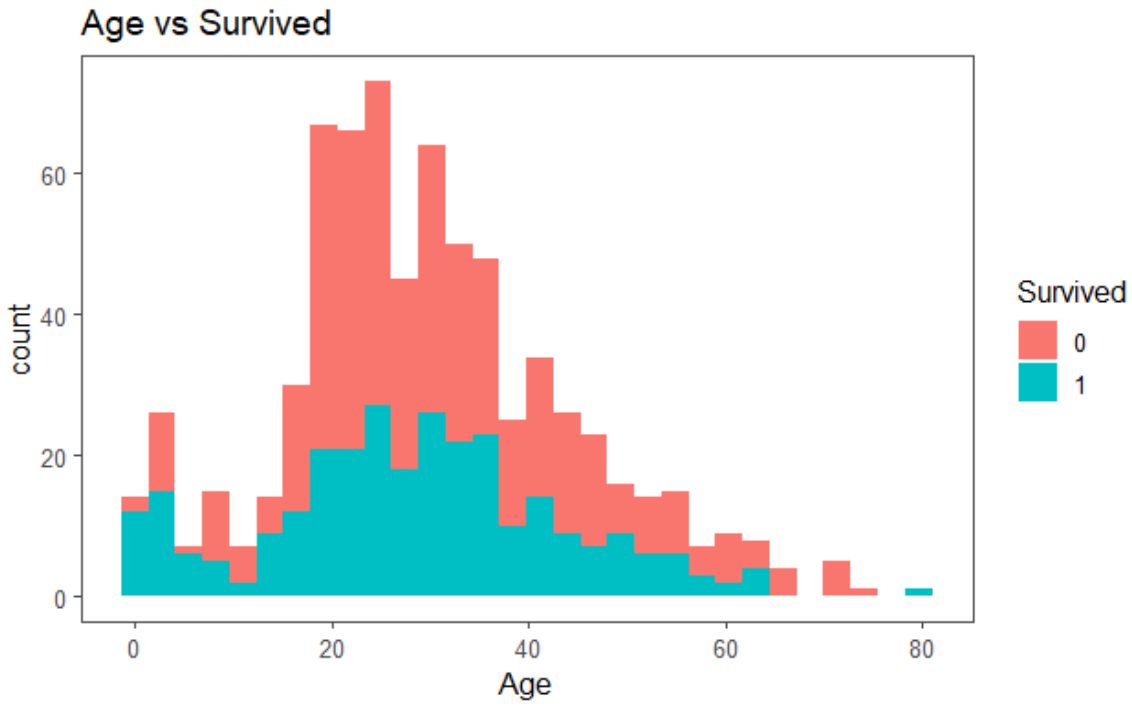
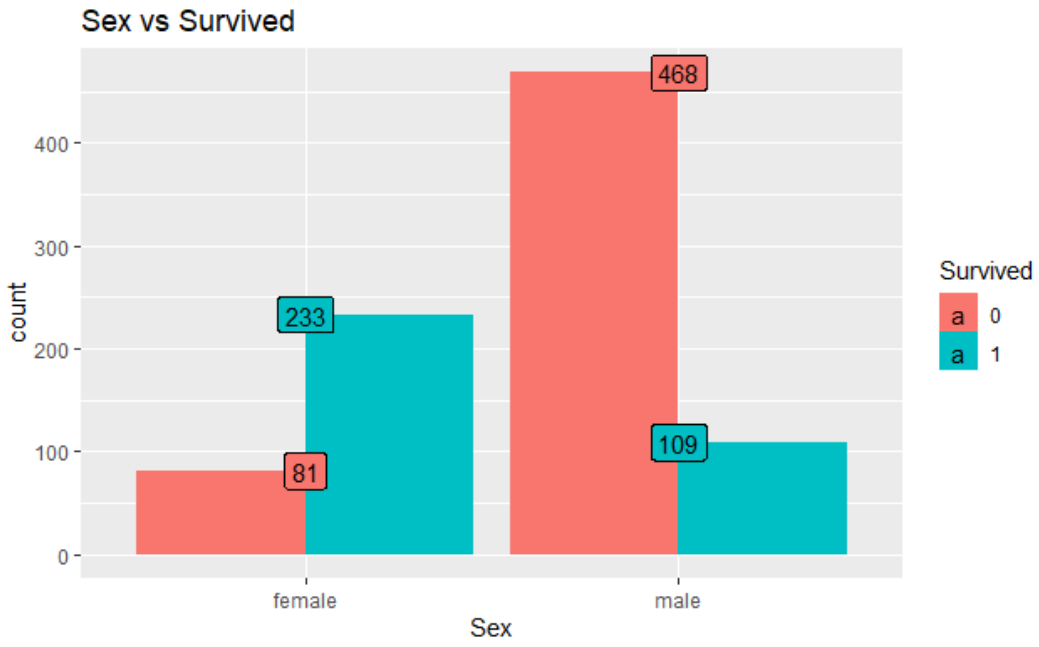
- [vedio] How to Get Started with Kaggle's Titanic Competition | Kaggle, [How to Get Started with Kaggle's Titanic Competition | Kaggle]
    - EDA
        - How to improve

-

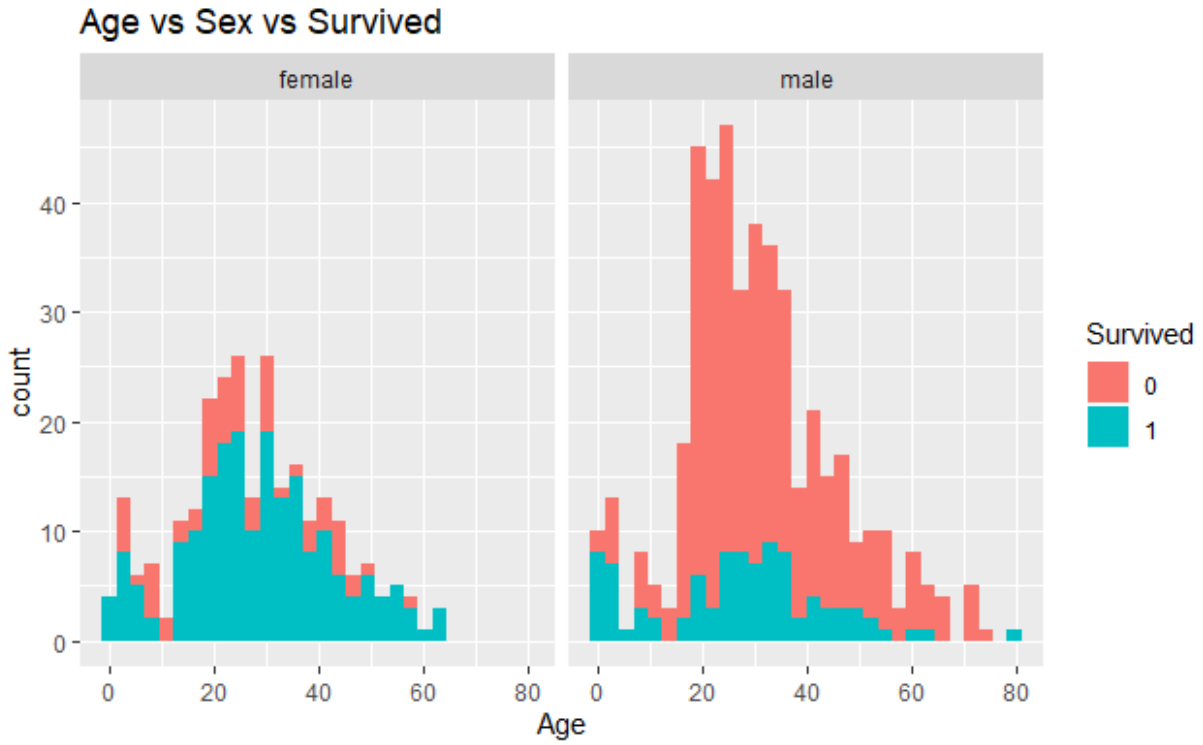
2.1 Age



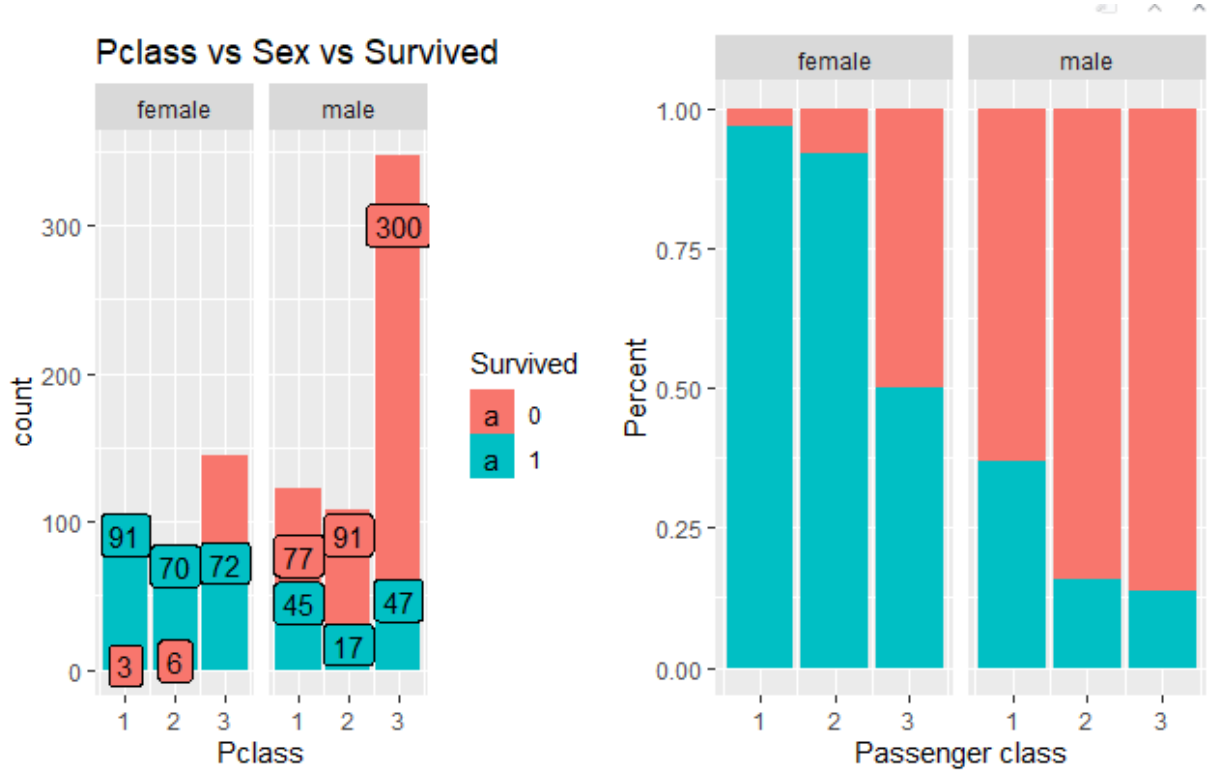
2.2 Sex Vs Survive



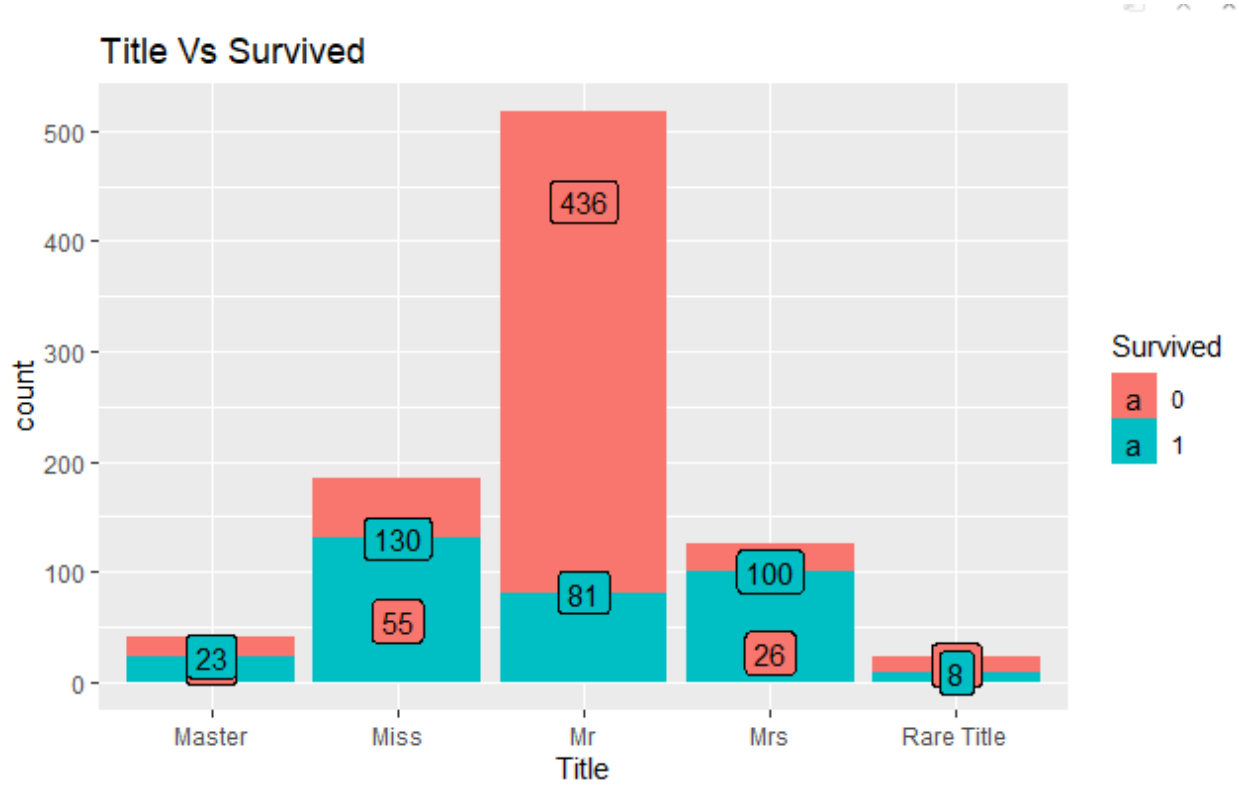
2.3 Age Vs Sex Vs Survived



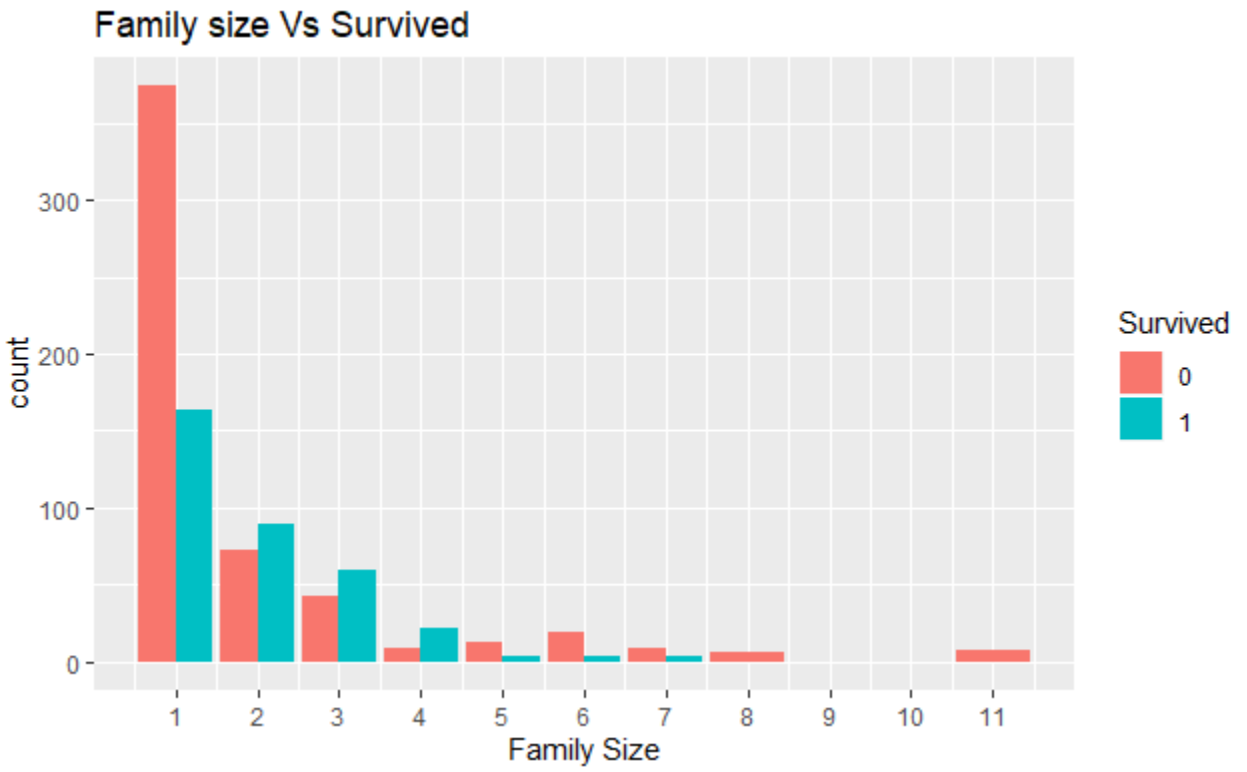
2.4. Pclass vs Sex



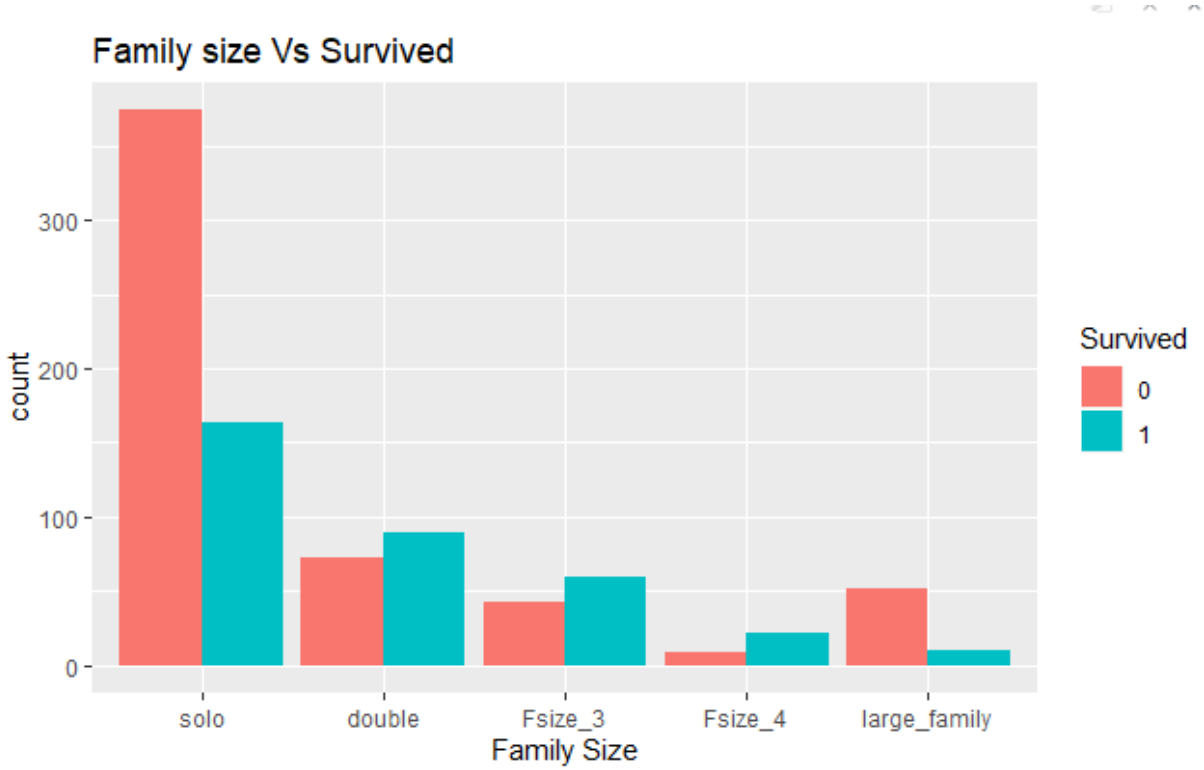
3.2 Title Vs Survived



3.3 Family size Vs Survived



4.1 Redefined the Fsize to factor:



4.2 SLLR on Survived~Age

```
Call:
glm(formula = train$Survived ~ train$Age, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1488  -1.0361  -0.9544   1.3159   1.5908

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.05672    0.17358  -0.327   0.7438
train$Age   -0.01096    0.00533  -2.057   0.0397 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 964.52  on 713  degrees of freedom
Residual deviance: 960.23  on 712  degrees of freedom
  (177 observations deleted due to missingness)
AIC: 964.23

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: train$Survived

Terms added sequentially (first to last)


          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
NULL                        713     964.52
train$Age  1   4.2876       712     960.23  0.03839 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4.3 SLLR on Survived~Sex

```
Call:
glm(formula = train$Survived ~ train$Sex, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6462  -0.6471  -0.6471   0.7725   1.8256

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     1.0566     0.1290   8.191 2.58e-16 ***
train$Sexmale  -2.5137     0.1672 -15.036  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1186.7  on 890  degrees of freedom
Residual deviance:  917.8  on 889  degrees of freedom
AIC: 921.8

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: train$Survived

Terms added sequentially (first to last)


          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                        890     1186.7
train$Sex  1   268.85        889      917.8 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.4 SLLR on Survived~Pclass

```
Call:
glm(formula = train$Survived ~ train$Pclass, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4094  -0.7450  -0.7450   0.9619   1.6836

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)           0.5306     0.1409   3.766 0.000166 ***
train$PclassClass_2  -0.6394     0.2041  -3.133 0.001731 **
train$PclassClass_3  -1.6704     0.1759  -9.496  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1186.7  on 890  degrees of freedom
Residual deviance: 1083.1  on 888  degrees of freedom
AIC: 1089.1

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: train$Survived

Terms added sequentially (first to last)


             Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                           890     1186.7
train$Pclass  2   103.55       888     1083.1 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4.5 SLLR on Survived~Fsize

```
Call:
glm(formula = train$Survived ~ train$Fsize, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6049  -0.8506  -0.8506   1.0888   1.9103

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)              0.2120     0.1585   1.337   0.1811
train$FsizeFsize_3       0.1044     0.2556   0.408   0.6830
train$FsizeFsize_4       0.7531     0.4447   1.694   0.0903 .
train$Fsizelarge_family -1.8606     0.3799  -4.897 9.72e-07 ***
train$Fsizesolo         -1.0425     0.1842  -5.659 1.52e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1186.7  on 890  degrees of freedom
Residual deviance: 1108.5  on 886  degrees of freedom
AIC: 1118.5

Number of Fisher Scoring iterations: 4

Analysis of Deviance Table

Model: binomial, link: logit

Response: train$Survived

Terms added sequentially (first to last)


            Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                          890     1186.7
train$Fsize  4   78.176       886     1108.5 4.238e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4.6 SLLR on Survived~Age + Sex + Pclass + Fsize

```
Call:
glm(formula = train$Survived ~ train$Age + train$Sex + train$Pclass +
    train$Fsize, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0603  -0.6347  -0.3936   0.5997   2.5288

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               3.948604   0.453484   8.707  < 2e-16 ***
train$Age                -0.040119   0.008448  -4.749 2.04e-06 ***
train$Sexmale            -2.680324   0.223901 -11.971  < 2e-16 ***
train$PclassClass_2      -1.466435   0.291928  -5.023 5.08e-07 ***
train$PclassClass_3      -2.494861   0.293589  -8.498  < 2e-16 ***
train$FsizeFsize_3        0.480501   0.348034   1.381 0.167398
train$FsizeFsize_4        0.805076   0.604634   1.332 0.183021
train$Fsizelarge_family -1.808759   0.491987  -3.676 0.000237 ***
train$Fsizesolo           0.102978   0.264714   0.389 0.697264
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 964.52  on 713  degrees of freedom
Residual deviance: 621.02  on 705  degrees of freedom
  (177 observations deleted due to missingness)
AIC: 639.02

Number of Fisher Scoring iterations: 5

Analysis of Deviance Table

Model: binomial, link: logit

Response: train$Survived

Terms added sequentially (first to last)


             Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                          713     964.52
train$Age     1    4.288       712     960.23   0.03839 *
train$Sex     1  210.271       711     749.96 < 2.2e-16 ***
train$Pclass  2  102.674       709     647.28 < 2.2e-16 ***
train$Fsize   4   26.267       705     621.02 2.795e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

5. SLLP with interaction terms

```
glm.fit: fitted probabilities numerically 0 or 1 occurredAnalysis of Deviance
Table

Model: binomial, link: logit

Response: train$Survived

Terms added sequentially (first to last)


                                          Df Deviance Resid. Df
NULL                                                      713
train$Age                                  1    4.288       712
train$Sex                                  1  210.271       711
train$Pclass                               2  102.674       709
train$Fsize                                4   26.267       705
train$Age:train$Sex                        1   12.176       704
train$Age:train$Pclass                     2    4.029       702
train$Sex:train$Pclass                     2   24.456       700
train$Age:train$Fsize                      4    7.680       696
train$Sex:train$Fsize                      4    2.089       692
train$Pclass:train$Fsize                   8   10.967       684
train$Age:train$Sex:train$Pclass           2    1.497       682
train$Age:train$Sex:train$Fsize            4   10.050       678
train$Age:train$Pclass:train$Fsize         8   25.485       670
train$Sex:train$Pclass:train$Fsize         7   12.943       663
train$Age:train$Sex:train$Pclass:train$Fsize 7   1.634       656
                                          Resid. Dev  Pr(>Chi)
NULL                                          964.52
train$Age                                     960.23 0.0383922 *
train$Sex                                     749.96 < 2.2e-16 ***
train$Pclass                                  647.28 < 2.2e-16 ***
train$Fsize                                   621.02 2.795e-05 ***
train$Age:train$Sex                           608.84 0.0004841 ***
train$Age:train$Pclass                        604.81 0.1334124
train$Sex:train$Pclass                        580.36 4.891e-06 ***
train$Age:train$Fsize                         572.68 0.1040205
train$Sex:train$Fsize                         570.59 0.7194818
train$Pclass:train$Fsize                      559.62 0.2035988
train$Age:train$Sex:train$Pclass              558.12 0.4731723
train$Age:train$Sex:train$Fsize               548.07 0.0395971 *
train$Age:train$Pclass:train$Fsize            522.59 0.0012859 **
train$Sex:train$Pclass:train$Fsize            509.65 0.0735103 .
train$Age:train$Sex:train$Pclass:train$Fsize   508.01 0.9772951
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```