Exercise 1: Titanic

Part A

Titanic data summary

Known age distribution
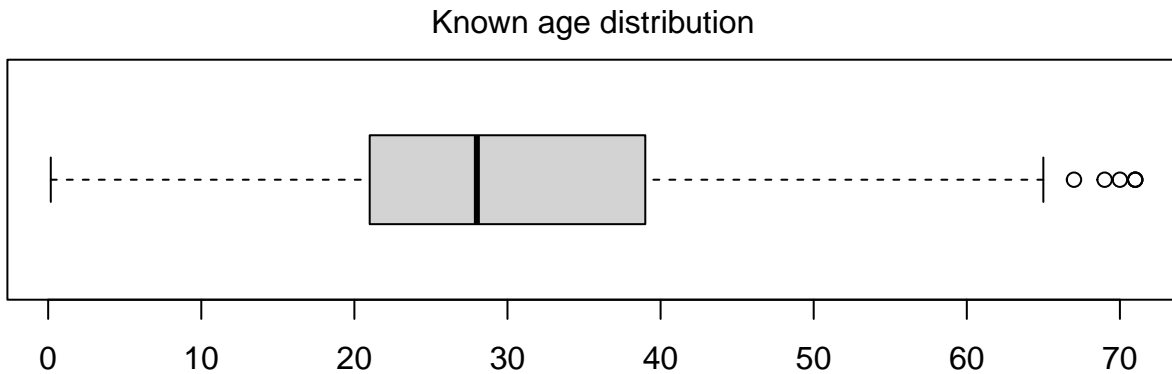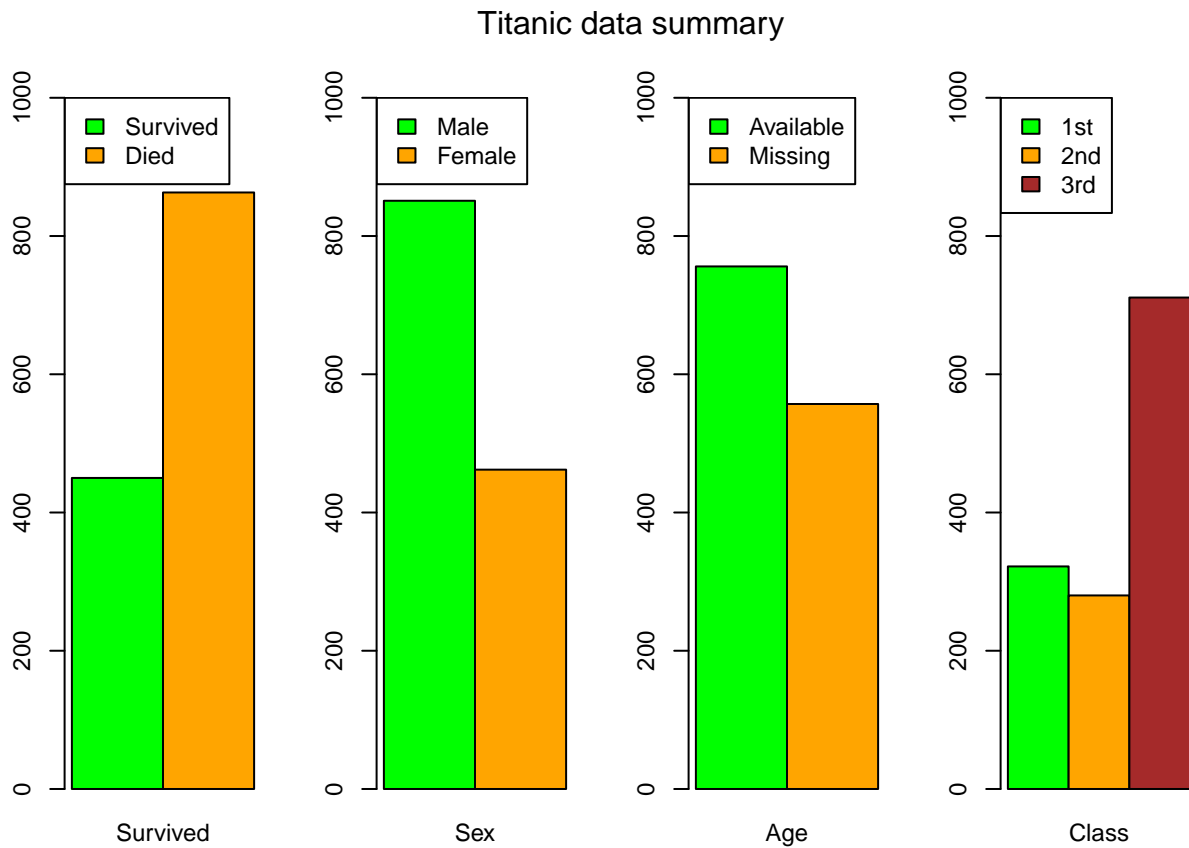
The data for this exercise is a part of the passenger manifest for the Titanic. The dataset covers the names, passenger class, survival status and, when available, the age of 1313 passengers. A summary of the data is given above. Because the age of some passengers is unavailable, we will use a which does not include these

passengers when their age is required.

```
age_glm <- glm(Survived ~ Age, data = data_not_na, family=binomial)
summary(age_glm)
```

```
##
## Call:
## glm(formula = Survived ~ Age, family = binomial, data = data_not_na)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.081428   0.173862  -0.468   0.6395
## Age         -0.008795   0.005232  -1.681   0.0928 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1025.6  on 755  degrees of freedom
## Residual deviance: 1022.7  on 754  degrees of freedom
## AIC: 1026.7
##
## Number of Fisher Scoring iterations: 4
```

```
sex_glm <- glm(Survived ~ Sex, data = data, family=binomial)
summary(sex_glm)
```

```
##
## Call:
## glm(formula = Survived ~ Sex, family = binomial, data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.69315    0.09869   7.023 2.17e-12 ***
## Sexmale     -2.30118    0.13488 -17.061  < 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: 1688.1  on 1312  degrees of freedom
## Residual deviance: 1355.5  on 1311  degrees of freedom
## AIC: 1359.5
##
## Number of Fisher Scoring iterations: 4
```

```
exp(0.69315-2.30118)
```

```
## [1] 0.2002818
```

```
cls_glm <- glm(Survived ~ PClass, data = data, family=binomial)
summary(cls_glm)
```

```
##
## Call:
## glm(formula = Survived ~ PClass, family = binomial, data = data)
##
```

```
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.4029     0.1137   3.543 0.000396 ***
## PClass2nd    -0.7052     0.1660  -4.249 2.15e-05 ***
## PClass3rd    -1.8265     0.1481 -12.335  < 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: 1688.1  on 1312  degrees of freedom
## Residual deviance: 1515.2  on 1310  degrees of freedom
## AIC: 1521.2
##
## Number of Fisher Scoring iterations: 4
```

```r
exp(0.4029-0.7052)
```

```
## [1] 0.7391163
```

```r
exp(0.4029-1.8265)
```

```
## [1] 0.2408454
```