```r
# R course for beginners
# HW 10 - Analysis
# assignment by Amir Mano, id 205779788

# import libraries
library(tidyverse)
library(pROC)

#### loading and organizing the data ----
data <- read.csv('HW_10/Titanic.csv')
df <- data.frame(gender = factor(data$Sex), is_first = factor(data$PClass=="1st"), survived = factor(data$Survived))

contrasts(df$gender) <- contr.treatment(2, base = ifelse(df$gender[1]=="male",1,2))
contrasts(df$is_first) <- contr.treatment(2, base = ifelse(df$is_first[1]==TRUE,1,2))

#### loading and organizing the data ----
summary(df)

print(stats<- df |>
  count(is_first, gender) |>
  pivot_wider(names_from = gender, values_from = n))

#### building logistic regression models ----
model_1 <- glm(df$survived ~ 1, family = binomial())
model_2 <- glm(df$survived ~ df$gender, family = binomial())
model_3 <- glm(df$survived ~ df$gender * df$is_first, family = binomial())

summary(model_1)
summary(model_2)
summary(model_3)
#exp(coef(model_1))
#exp(coef(model_2))
#exp(coef(model_3))

#### ROC ----
df$predict_model_1 <- predict(model_1, type = "response")
df$predict_model_2 <- predict(model_2, type = "response")
df$predict_model_3 <- predict(model_3, type = "response")

roc_model_1 <- roc(df$survived, df$predict_model_1)
roc_model_2 <- roc(df$survived, df$predict_model_2)
roc_model_3 <- roc(df$survived, df$predict_model_3)

auc(roc_model_1)
auc(roc_model_2)
auc(roc_model_3)

#### Plotting ----

# Plot ROC curves
plot(roc_model_1, col = "blue2", main = "ROC Curve Comparison")
plot(roc_model_2, add = TRUE, col = "red2")
plot(roc_model_3, add = TRUE, col = "green3")
legend("right", legend = c("Intercept", "+ gender", "+ gender + class"), col = c("blue2","red2", "green3"), lwd = 3)
```