1. Import the Titanic Dataset from the link Titanic Data Set.

Perform the following:

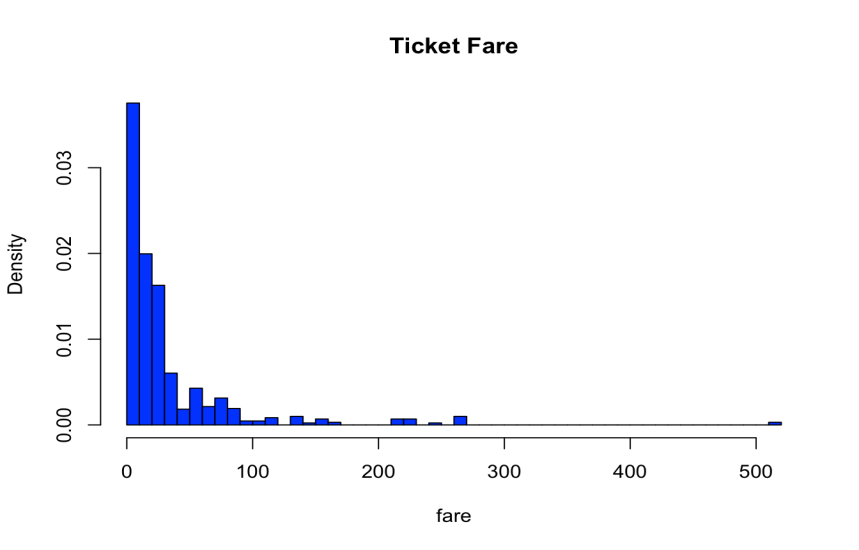
a. Is there any difference in fares by different class of tickets?

Note - Show a boxplot displaying the distribution of fares by class

fare <- titanic$fare

length(fare); index<-which(!is.na(fare)); fare<-fare[index]; length(fare)

hist(fare,50,col="blue",freq=FALSE, main="Ticket Fare")



summary(fare)

qqnorm(fare)

qqline(fare)

b. Is there any association with Passenger class and gender?

Note – Show a stacked bar chart

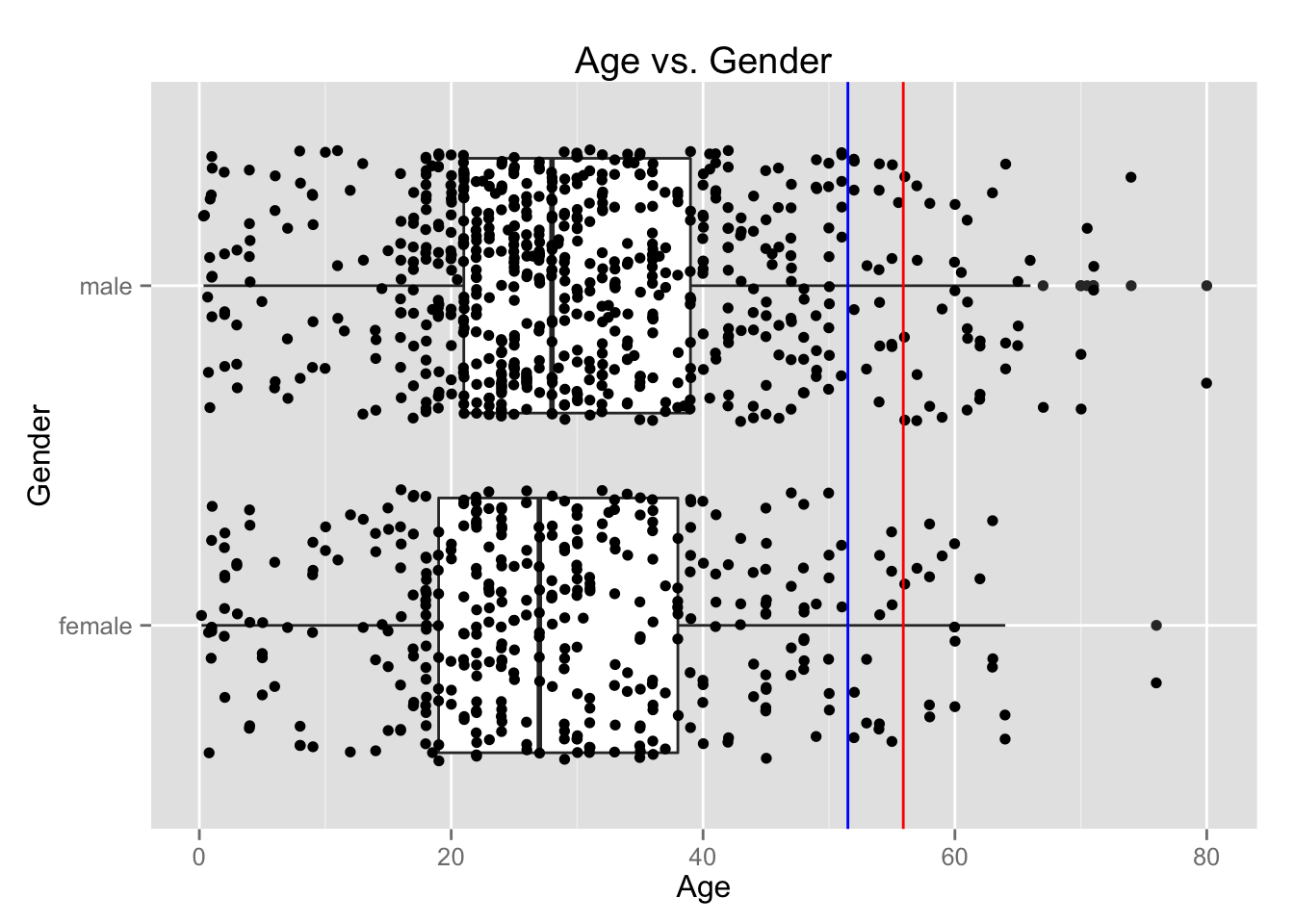
p <- qplot(x=age.nna$sex, y=age.nna$age, data=age.nna,

geom=c("boxplot", "jitter"), main="Age vs. Gender", xlab="Gender",

ylab="Age") + coord\_flip()

p + geom\_hline(yintercept = 51.5, color="blue", label="Life Exp. (M)") +

geom\_hline(yintercept = 55.9, colour="red", label="Life Exp. (F)")



length(which(titanic$age < 18))

titanic$age\_group <- "adult"

titanic$age\_group[titanic$age < 18] <- "child"

age\_group.survived <- table(titanic$age\_group, titanic$survived); age\_group.survived

age\_group.survived.prop <- prop.table(age\_group.survived, 1); age\_group.survived.prop

adults.survived <- age\_group.survived.prop[3]

children.survived <- age\_group.survived.prop[4]

observed <- children.survived / adults.survived; observed