Assignment #3: Data Manipulation with base R

Ryan Wynns

September 26, 2018

# Questions #1, 2, 4

sum(is.na(Titanic))

## [1] 177

Titanic[is.na(Titanic)] = mean(Titanic$Age, na.rm = TRUE)  
Titanic$SibSp = NULL  
Titanic$Parch = NULL  
Titanic$Ticket = NULL  
Titanic$Cabin = NULL  
Titanic$Name = NULL  
Titanic$PassengerID = NULL

# Questions #5-11

mean(Titanic$Age)

## [1] 29.69912

TitanicFemale <- subset(Titanic, Titanic$Sex == "female")  
 mean(TitanicFemale$Age)

## [1] 28.21673

Class1 <- subset(Titanic, Titanic$Pclass == "1")  
 median(Class1$Fare)

## [1] 60.2875

NonClassFemales <- subset(Titanic, Titanic$Pclass != "1" & Titanic$Sex == "female")  
median(NonClassFemales$Fare)

## [1] 14.45625

NonClassFemalesAlive <- subset(Titanic, Titanic$Pclass != "3" & Titanic$Sex == "female" & Titanic$Survived == "1")  
 median(NonClassFemalesAlive$Age)

## [1] 30

FemaleTeensAlive<- subset(Titanic, Titanic$Survived == "1" & Titanic$Sex == "female" & Titanic$Age >= "13" & Titanic$Age <= "19")  
 mean(FemaleTeensAlive$Fare)

## [1] 49.17966

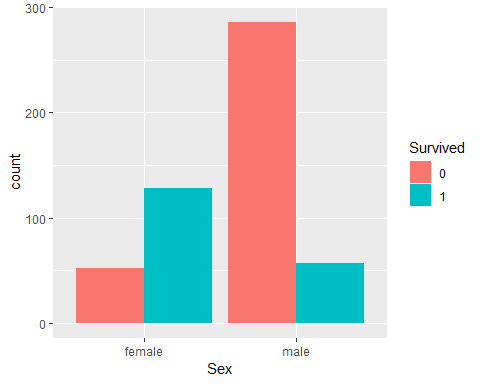
FemaleTeensAlive1<- subset(Titanic, Titanic$Survived == "1" & Titanic$Sex == "female" & Titanic$Age >= "13" & Titanic$Age <= "19" & Titanic$Pclass == "1")  
FemaleTeensAlive2 <- subset(Titanic, Titanic$Survived == "1" & Titanic$Sex == "female" & Titanic$Age >= "13" & Titanic$Age <= "19" & Titanic$Pclass == "2")  
 FemaleTeensAlive3<- subset(Titanic, Titanic$Survived == "1" & Titanic$Sex == "female" & Titanic$Age >= "13" & Titanic$Age <= "19" & Titanic$Pclass == "3")

# Questions #12-13

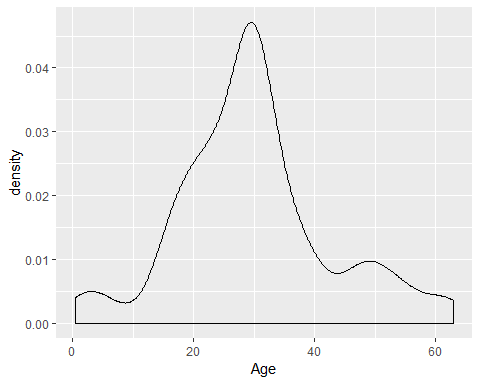
AboveAvgFare <-subset(Titanic, Titanic$Fare >= "32.2")  
SurvivedFare <- sum(AboveAvgFare$Survived == "1")  
DeathFare <- sum(AboveAvgFare$Survived == "0")  
Ratio <- SurvivedFare/DeathFare  
  
Titanic$sFare <- (Titanic$Fare - mean(Titanic$Fare)) / sd(Titanic$Fare)

# Questions #14-20

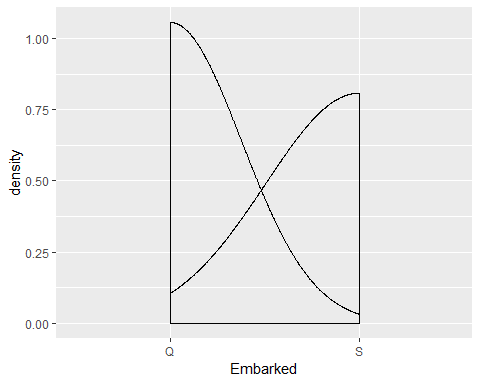
YoungPassengers <- subset(Titanic, Titanic$Age <= "29.70")  
ggplot(data = YoungPassengers) + geom\_bar(mapping = aes(x = Sex, fill = Survived), position = "dodge")



SurviveHighFare <- subset(Titanic, Titanic$Survived == "1" & Titanic$Fare >= "32.2")  
ggplot(data = SurviveHighFare) + geom\_density(mapping = aes(x = Age))



EmbSHandQT<- subset(Titanic, Titanic$Embarked =="Q" | Titanic$Embarked == "S")  
ggplot(data = EmbSHandQT) + geom\_density(mapping = aes(x = Embarked))



Titanic$cfare <- ifelse(Titanic$Fare > 32.2, "expensive", ifelse(Titanic$Fare < 32.2, "cheap",""))  
  
Titanic$cage = Titanic$Age  
Titanic$cage[(Titanic$cage >0) & (Titanic$cage <=10)] = 0  
Titanic$cage[(Titanic$cage >10) & (Titanic$cage <=20)] = 1  
Titanic$cage[(Titanic$cage >20) & (Titanic$cage <=30)] = 2  
Titanic$cage[(Titanic$cage >30) & (Titanic$cage <=40)] = 3  
Titanic$cage[(Titanic$cage >40) & (Titanic$cage <=50)] = 4  
Titanic$cage[(Titanic$cage >50) & (Titanic$cage <=60)] = 5  
Titanic$cage[(Titanic$cage >60) & (Titanic$cage <=70)] = 6  
Titanic$cage[(Titanic$cage >70) & (Titanic$cage <=80)] = 7  
Titanic$cage[(Titanic$cage >80) & (Titanic$cage <=90)] = 8

### 

ggplot(data = AboveAvgFare) + geom\_bar(mapping = aes(x = cage, fill = Survived), position = “dodge”)

ggplot(data = AboveAvgFare) + geom\_bar(mapping = aes(x = cage, fill = Survived), position = “dodge”) + facet\_wrap(~Pclass, nrow =1)

```