**Exploratory Analysis**[¶](https://www.kaggleusercontent.com/kf/1538771/eyJhbGciOiJkaXIiLCJlbmMiOiJBMTI4Q0JDLUhTMjU2In0..6xk13l_v0ghUh_6TFc3TVQ.KZlZ973vHHhmKRojoPQIao1pppo7Qw9hezq6rVIAspLo8t6yf3s6l4Uam6VTnDofgEQpBrQ1FYkWf3A3HbjgtOlL3YrrwwUxu4IpH75DiFi6q_2eWaBiAywH5W_3MOqNI7GmXTCXfFkIFtLmbsV3rQ.akUQueixawCbvNnCmGs7Iw/__results__.html#2.-Exploratory-Analysis)

* Missing value : (part 3)
  + Age : 263
  + Fare : 1
  + Embarked : 2
  + Cabin : too many

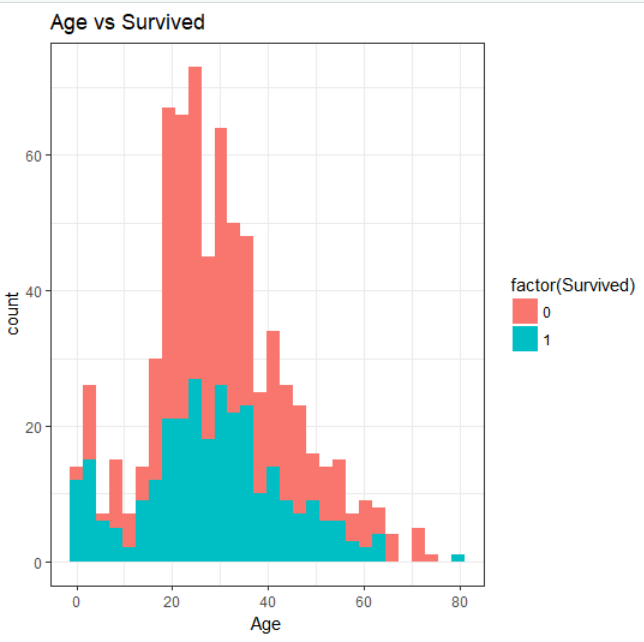
1. **Age**

ggplot(train[1:891,], aes(Age, fill = factor(Survived))) +

geom\_histogram(bins=30) +

xlab("Age") +

ggtitle("Age vs Survived")+theme\_bw()

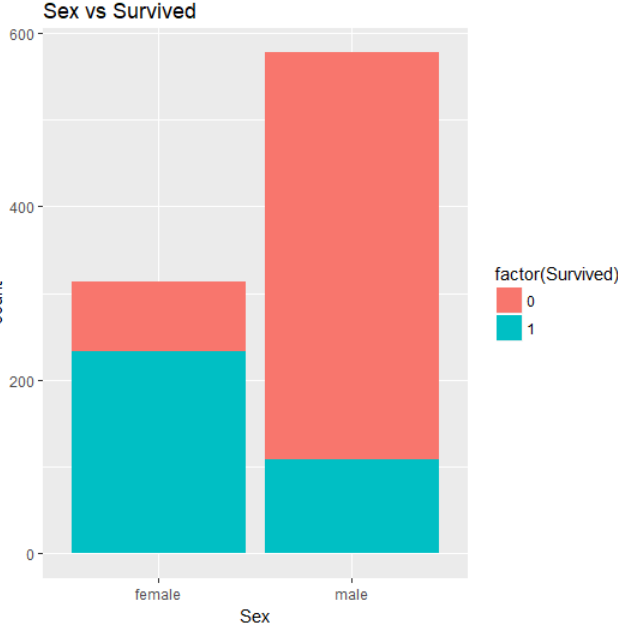


1. Sex

#2.Sex vs survived

ggplot(train[1:891,],mapping=aes(Sex,fill=factor(Survived)))+geom\_bar(stat='count')+ggtitle("Sex vs Survived")

tapply(train[1:891,]$Survived,train[1:891,]$Sex,mean)# apply the function mean into data train$survived,index is &Sex

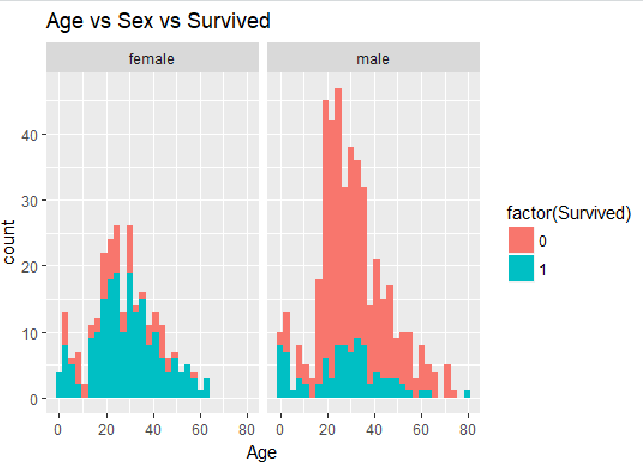


1. Age vs Sex

ggplot(train[1:891,],mapping=aes(Age,fill=factor(Survived)))+geom\_histogram(bins=30)+

ggtitle("Age vs Sex vs Survived")+

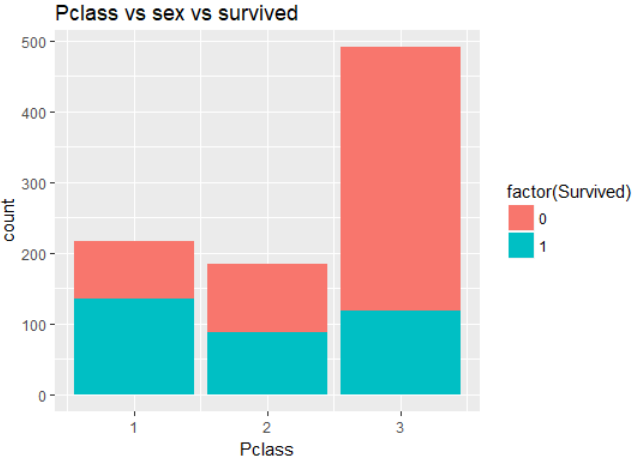
xlab('Age')+ylab('count')+facet\_grid(.~Sex)# horizonal to show the graph(Sex~. ->vertical)



1. Plcass vs Survived

ggplot(train[1:891,],aes(Pclass,fill=factor(Survived)))+geom\_bar(stat='count')+# for categorical

xlab('Pclass')+ylab('count')+ggtitle('Pclass vs sex vs survived')



1. Pclass vs sex vs age vs survived

#5. Pclass vs sex vs age vs survival

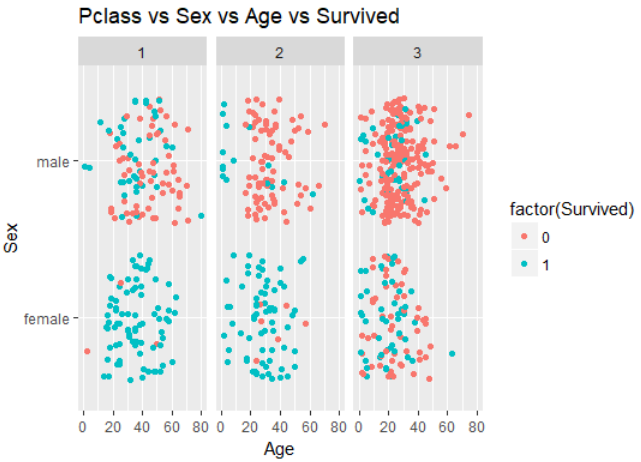
# four variable, x,y to age,sex. fill to survived,facet\_wrap to pclass

ggplot(train[1:891,], aes(x = Age, y = Sex)) +

geom\_jitter(aes(colour = factor(Survived))) +

facet\_wrap(~Pclass) +

labs(x = "Age", y = "Sex", title = "Pclass vs Sex vs Age vs Survived")



1. Fare vs Pclass

ggplot(train[1:891,],aes(x=Fare, y=Pclass))+geom\_jitter(aes(colour=factor(Survived)))+

theme(element\_blank())+labs(x='Age',y='Pclass',title='Fare vs pclass')+

scale\_fill\_discrete(name = "survived")

