# Gender_Model.R

*JingbinXu*

*Thu Dec 22 16:26:38 2016*

```
# Jingbin - 10 Jan 2016
# Titanic

# Set working directory and import datafiles
train <- read.csv("train.csv")
test <- read.csv("test.csv")

# Install and load required packages for fancy decision tree plotting
library(rpart)
library(rattle)
```

```
## Warning: Failed to load RGtk2 dynamic library, attempting to install it.
```

```
## Please install GTK+ from http://r.research.att.com/libs/GTK_2.24.17-X11.pkg
```

```
## If the package still does not load, please ensure that GTK+ is installed and that
it is on your PATH environment variable
```

```
## IN ANY CASE, RESTART R BEFORE TRYING TO LOAD THE PACKAGE AGAIN
```

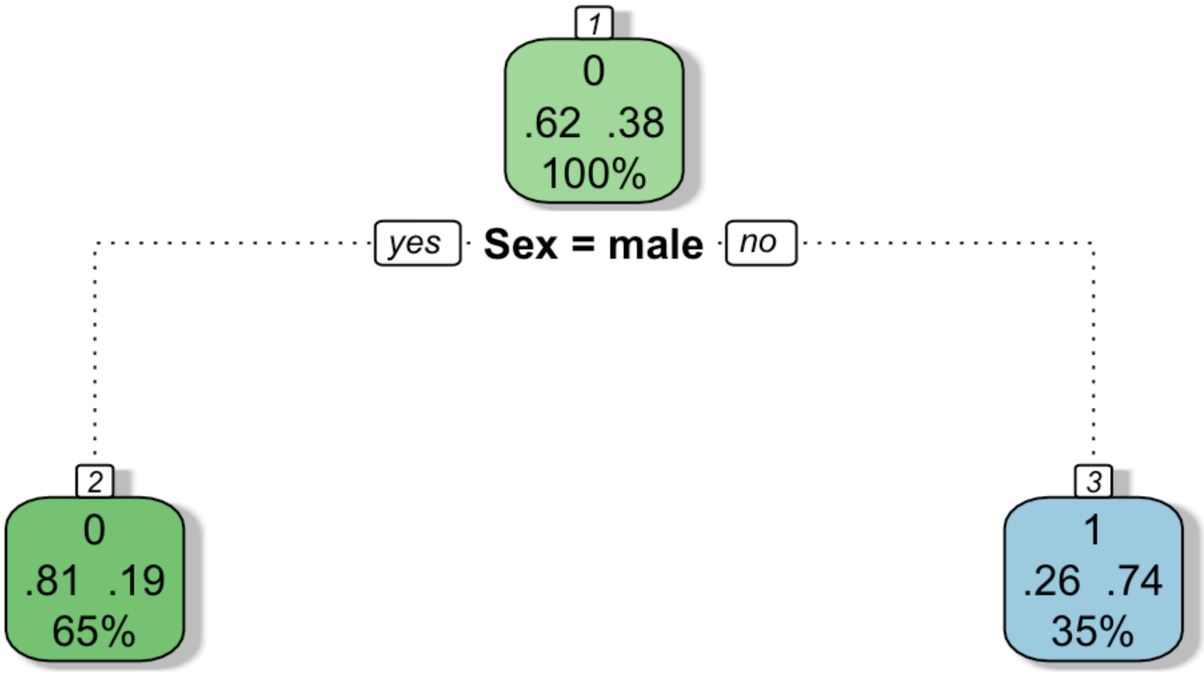
```
## Rattle: A free graphical interface for data mining with R.
## Version 4.1.0 Copyright (c) 2006-2015 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

```
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 3.2.5
```

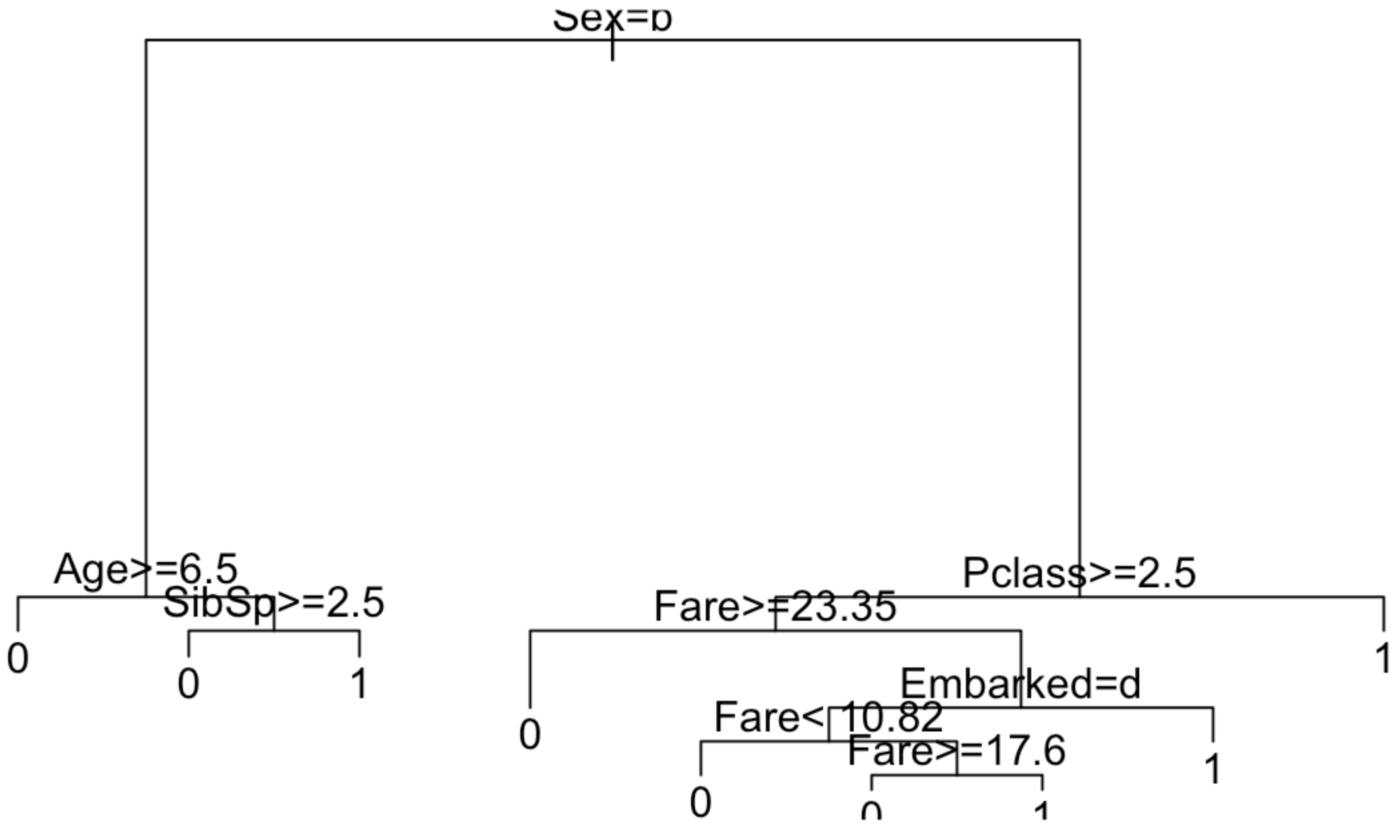
```
library(RColorBrewer)

# Recreate the gender model
fit <- rpart(Survived ~ Sex, data=train, method="class")
fancyRpartPlot(fit)
```

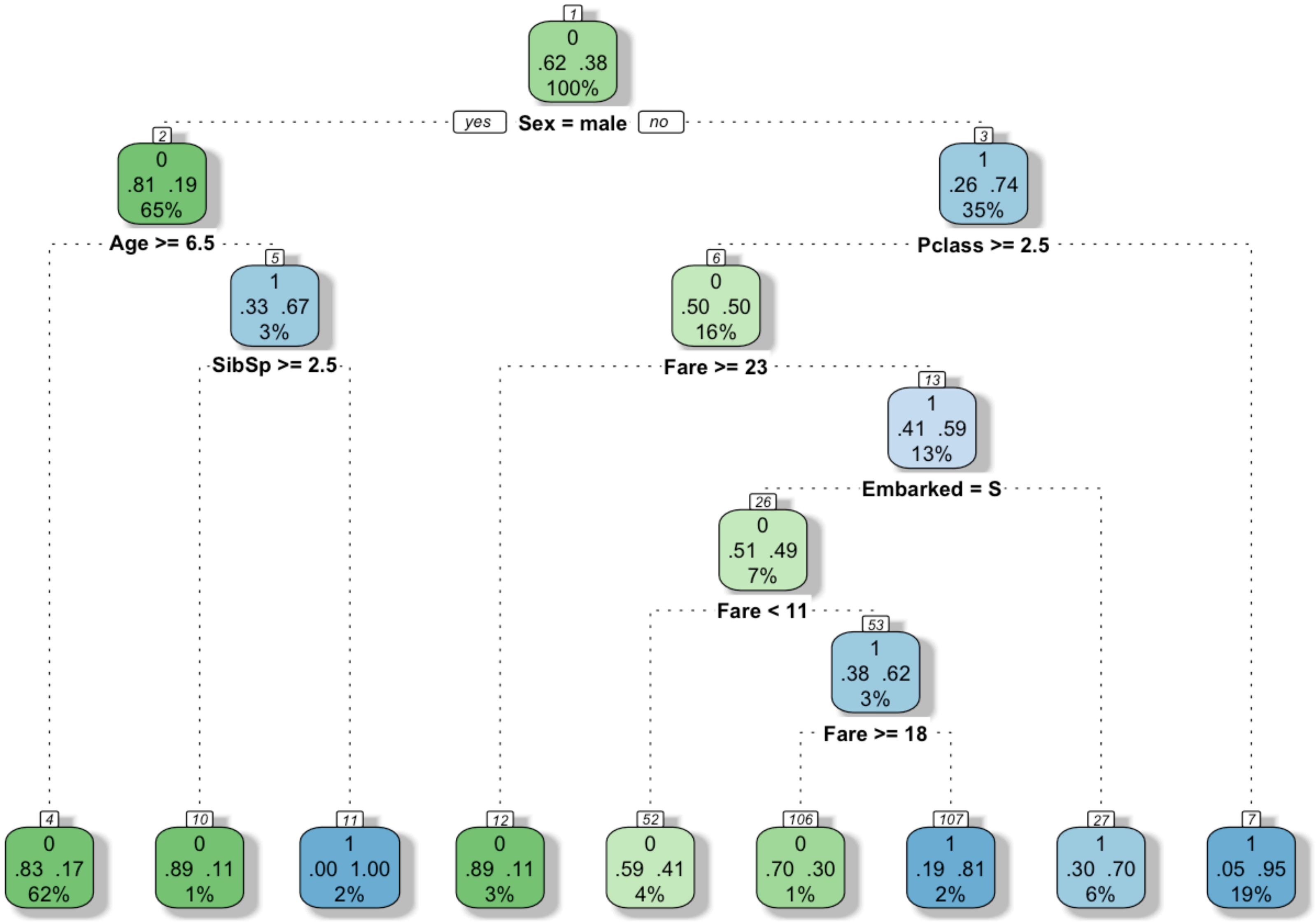


Rattle 2016-Dec-22 16:26:39 JingbinXu

```
# Build a deeper tree
fit <- rpart(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked, data=train, method="class")
# Plot it with base-R
plot(fit)
text(fit)
```



```
# And then make it look better with fancyRpartPlot!
fancyRpartPlot(fit)
```

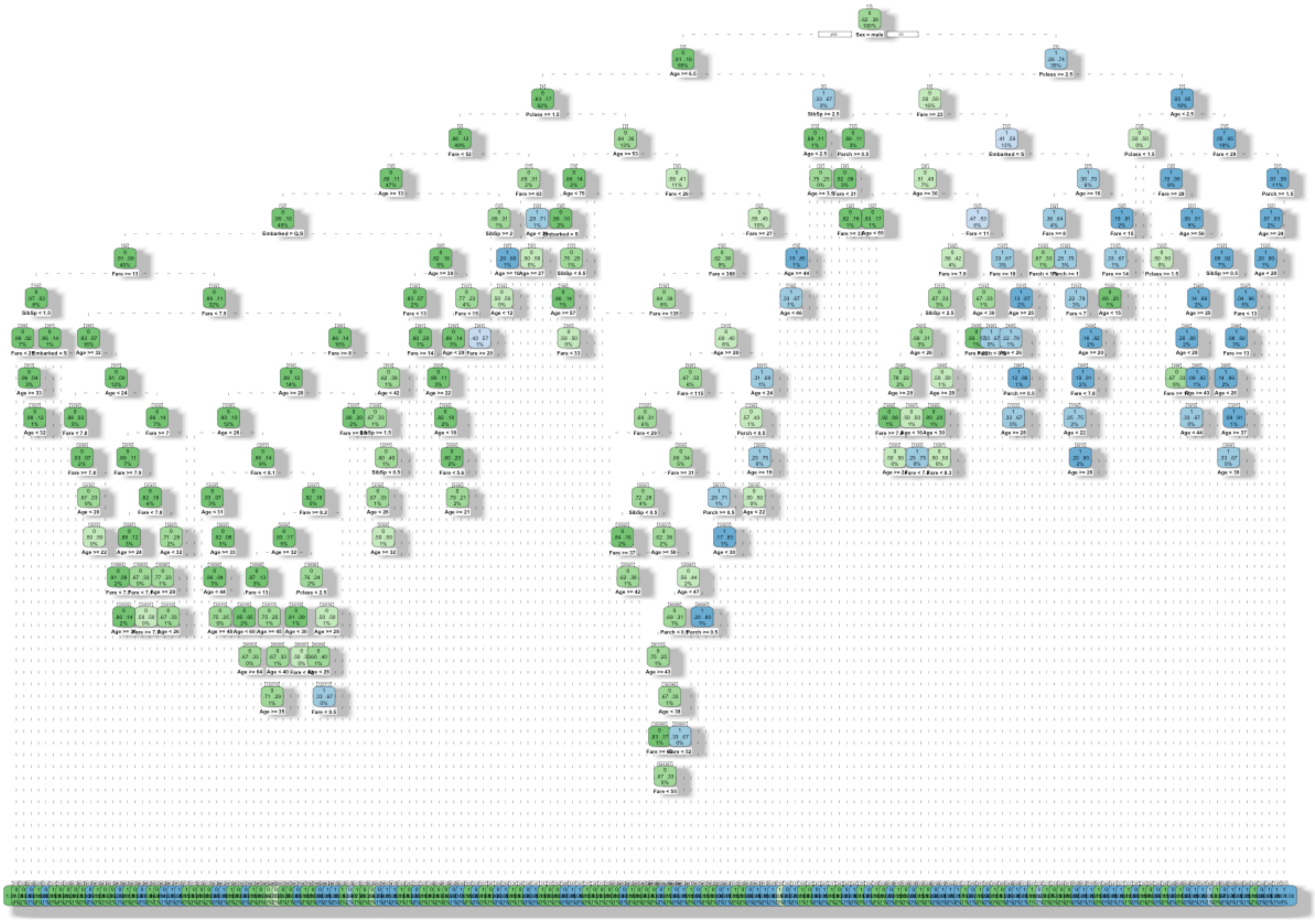


Rattle 2016-Dec-22 16:26:39 JingbinXu

```
# Now let's make a prediction and write a submission file
Prediction <- predict(fit, test, type = "class")
submit <- data.frame(PassengerId = test$PassengerId, Survived = Prediction)
write.csv(submit, file = "myfirstdtree.csv", row.names = FALSE)

# Let's unleash the decision tree and let it grow to the max
fit <- rpart(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked, data=tr
ain,
             method="class", control=rpart.control(minsplit=2, cp=0))
fancyRpartPlot(fit)
```

```
## Warning: labs do not fit even at cex 0.15, there may be some overplotting
```

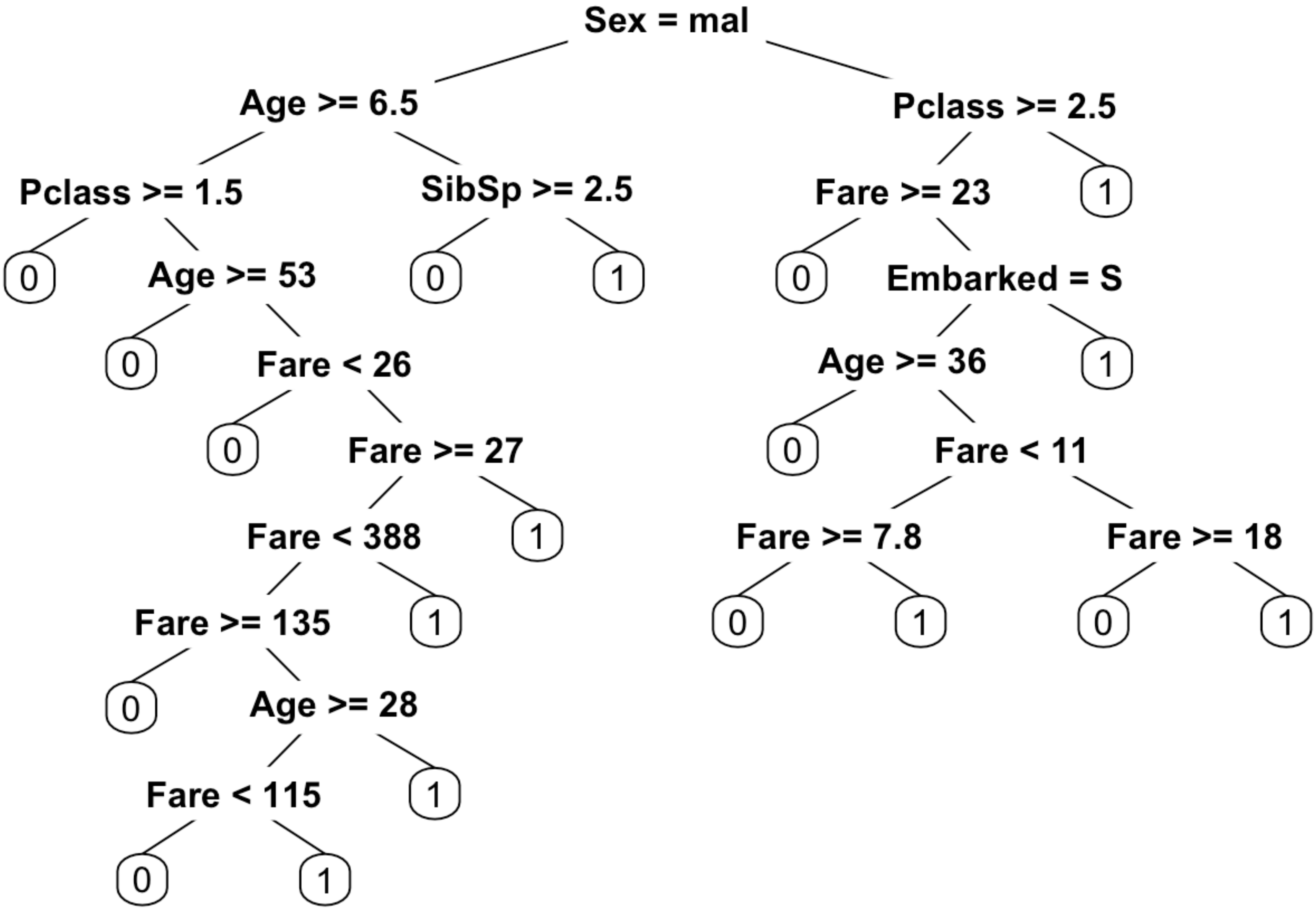


Rattle 2016-Dec-22 16:26:40 JingbinXu

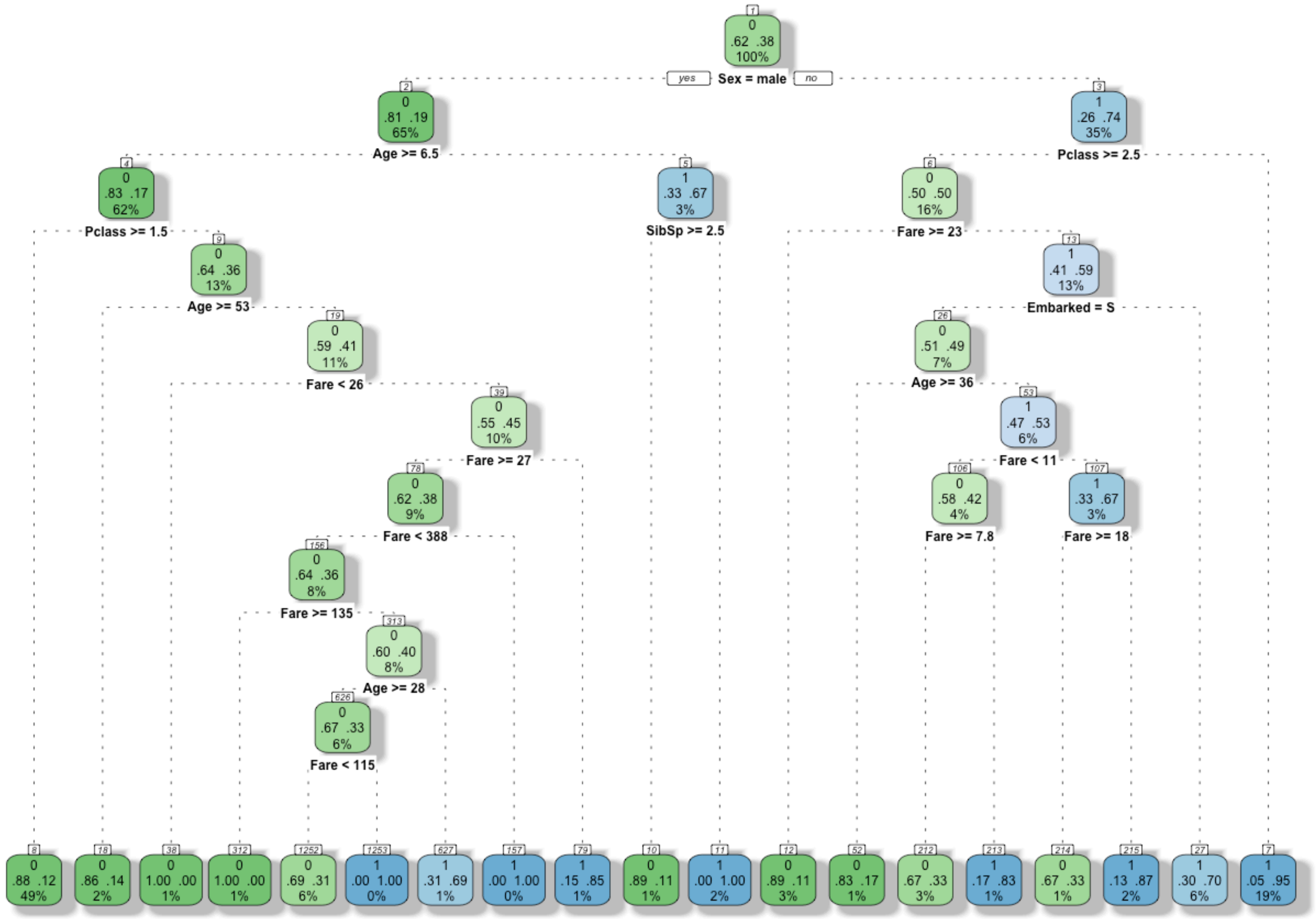
```
# Now let's make a prediction and write a submission file
Prediction <- predict(fit, test, type = "class")
submit <- data.frame(PassengerId = test$PassengerId, Survived = Prediction)
write.csv(submit, file = "myfullgrowntree.csv", row.names = FALSE)

# Manually trim a decision tree
fit <- rpart(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare + Embarked, data=tr
ain,
             method="class", control=rpart.control(minsplit=2, cp=0.005))
new.fit <- prp(fit,snip=TRUE)$obj
```

```
## Warning: ignoring snip=TRUE for quartz_off_screen device
```



```
fancyRpartPlot(new.fit)
```



Rattle 2016-Dec-22 16:26:53 JingbinXu