Titanic: Machine Learning from Disaster
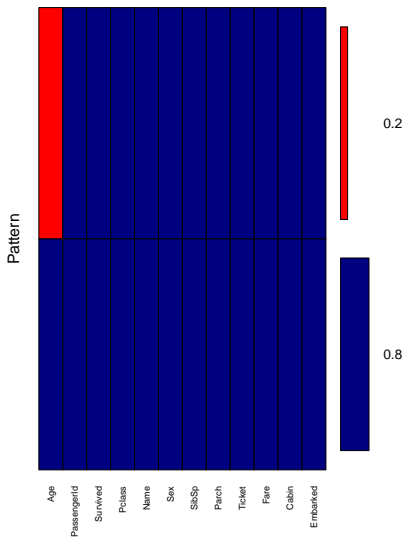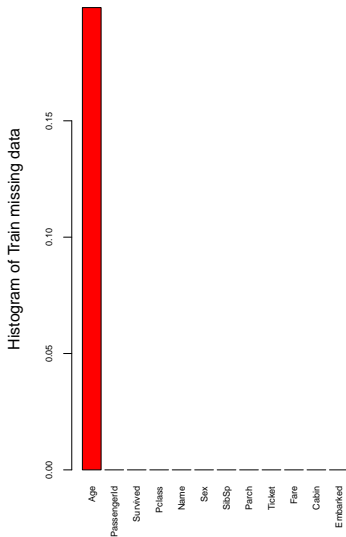
Casey Quinn

February 10, 2016

Kaggle Data

- train.csv
- 891 observations
- 11 potential predictor variables
- Survivor Information
- test.csv
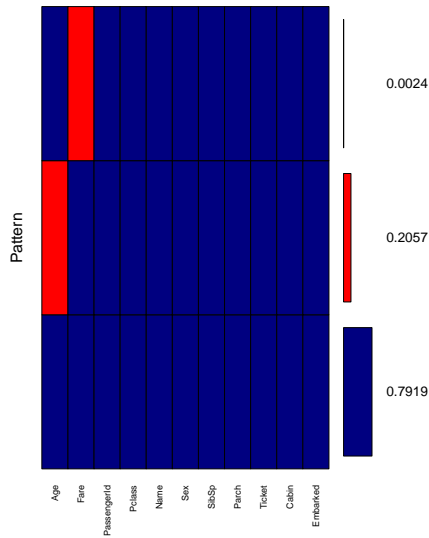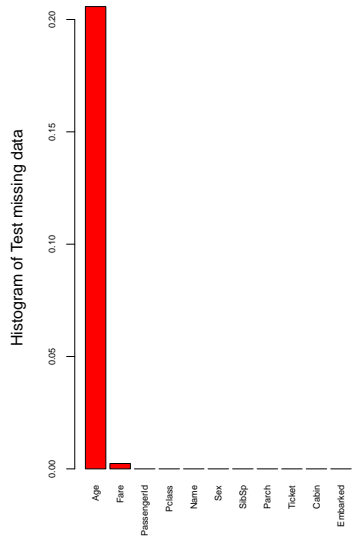- 418 observations
- 11 potential predictor variables

Train Data


```
##
## Variables sorted by number of missings:
```
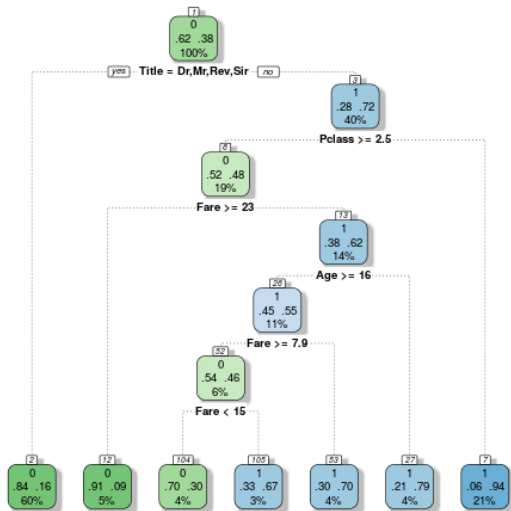
Test Data


```
##
## Variables sorted by number of missings:
```

Results

Model	Package/function	Description	Training Accuracy	Test Accuracy
KNN	e1071/knn	Fare K=5	0.776	0.660
KNN	e1071/knn	Sex&Fare K = 5	0.767	0.766
Tree	rpart	All Variables + Derived Values	0.840	0.804
Tree	rpart	All Varibles	0.840	0.785
Tree	tree	All	0.616	0.780
Random Forest	rPart/randomForest	All	0.912	0.799
Bagging	rPart/randomForest	All	0.984	0.737
Logistic Regression	glm	Sex&Fare	0.764	0.761
LDA	MASS/lda	Sex&Fare	0.765	0.761
Naive Bayes	e1071/naiveBayes	Pclass	0.679	0.655
Naive Bayes	e1071/naiveBayes	Pclass + Sex + Age + Fare	0.777	0.727
SVM	e1071/svm poly2 c=1	Pclass Sex Sib Par	0.810	0.770
SVM	e1071/svm tuned	Pclass Sex Sib Par	0.807	0.775
SVM	e1071/svm gamma = 0.5 c=1	Pclass Sex Sib Par	0.816	0.804

Winning Tree with Title

Rattle 2016-Feb-11 22:10:32 repl