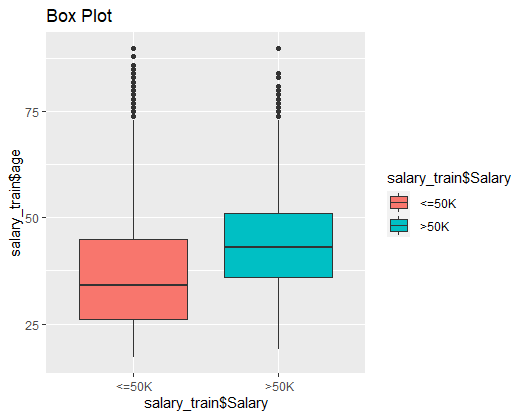
**Naïve Bayes ML Algo**

* **Salary Dataset:**
* We are Range of Salary column as Target column to predict on and classifying it.
* Below is the boxplot of salary range of class1 <=50k and class 2 >50k.



* X= salary\_train$Salary y =salary$train
* The higher the age (blue boxplot ) the higher the salary snce its experienced professional, where as the younger generation seems to have less salary it falls majoritvely under (<=50k).

hoursperweek

Y [,1] [,2]

<=50K 39.34856 11.95104

>50K 45.70658 10.73699

* NB model will predict the salary baased on various categories inside each columns [ hours / week ] has data 1 and 2 .
* > mean(y\_pred==salary\_test$Salary)
* [1] 0.8193227
* The model prediction accuracy was 81 %.
* y
* x <=50K >50K
* <=50K 10550 810
* >50K 1911 1789
* Confusion matrix of Salary dataset.
* **Ham Spam Detection:**
* data frame with 0 columns and 4812 rows in ham.
* data frame with 0 columns and 747 rows in spam
* 86% ham and 13% spam in messages data
* > test\_acc
* [1] 0.9755396
* Test Accuracy ::
* 91 % accuracy in terms of identifying given test data was ham or spam based on test data.
* Train Accuracy::
* 0.9846486
* Train accuracy is more than test accuracy.
* Prediction for this specific term “accept”
* accept
* sms\_raw\_train$type No Yes
* ham 0.998613037 0.001386963
* spam 1.000000000 0.000000000
* No and yes column are made ia custome function based on if a single message or doc has repeated words that means it’s a spam