

Banking Dataset

2019 HACKATHON

TEAM MEMBER: HOWARD, STEVEN, AILING, VIVIAN, ISSAC

MISSION

- Find out which variable is most important in determining the term deposit.
- Find out correlation between important variables and gain insight from it.

Codes – Training Data

```
```{r}
library(randomForest)
set.seed(1234)
rf = randomForest(y~., data = train)
print(rf)
```
```

randomForest 4.6-14
Type rfNews() to see new features/changes/bug fixes.

Call:
randomForest(formula = y ~ ., data = train)
Type of random forest: classification
Number of trees: 500
No. of variables tried at each split: 4

OOB estimate of error rate: 9.47%

Confusion matrix:

| | no | yes | class.error |
|-----|-------|------|-------------|
| no | 30730 | 1227 | 0.03839534 |
| yes | 2201 | 2048 | 0.51800424 |

Codes – Testing Data

```
```{r}
p2 = predict(rf,test)
confusionMatrix(p2, test$y)
```
```

Confusion Matrix and Statistics

| | | Reference | |
|------------|-----|------------|--------|
| | | Prediction | no yes |
| Prediction | no | 7723 | 513 |
| | yes | 242 | 527 |

Accuracy : 0.9162
95% CI : (0.9102, 0.9218)
No Information Rate : 0.8845
P-Value [Acc > NIR] : < 2.2e-16

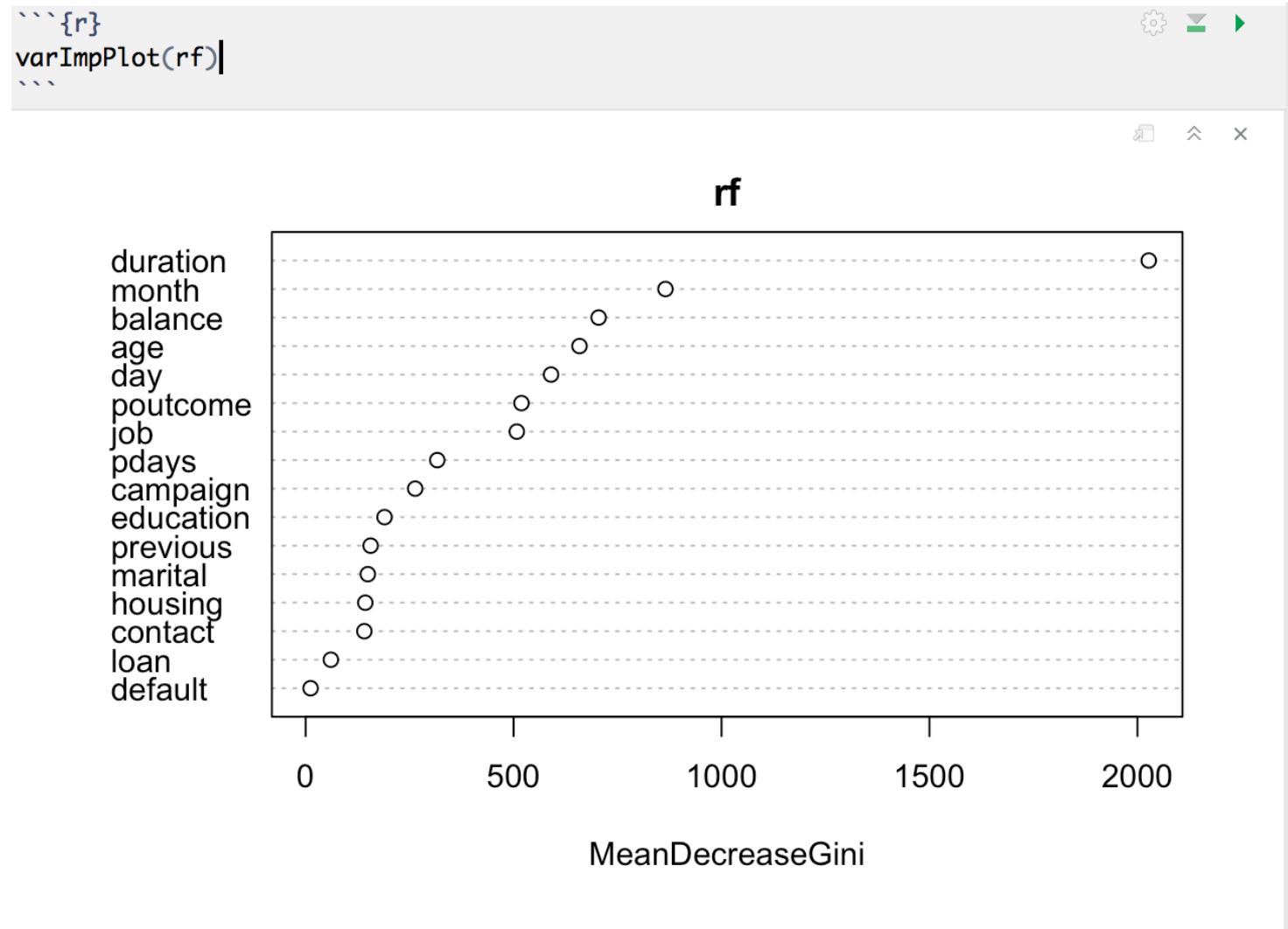
Kappa : 0.5372

Mcnemar's Test P-Value : < 2.2e-16

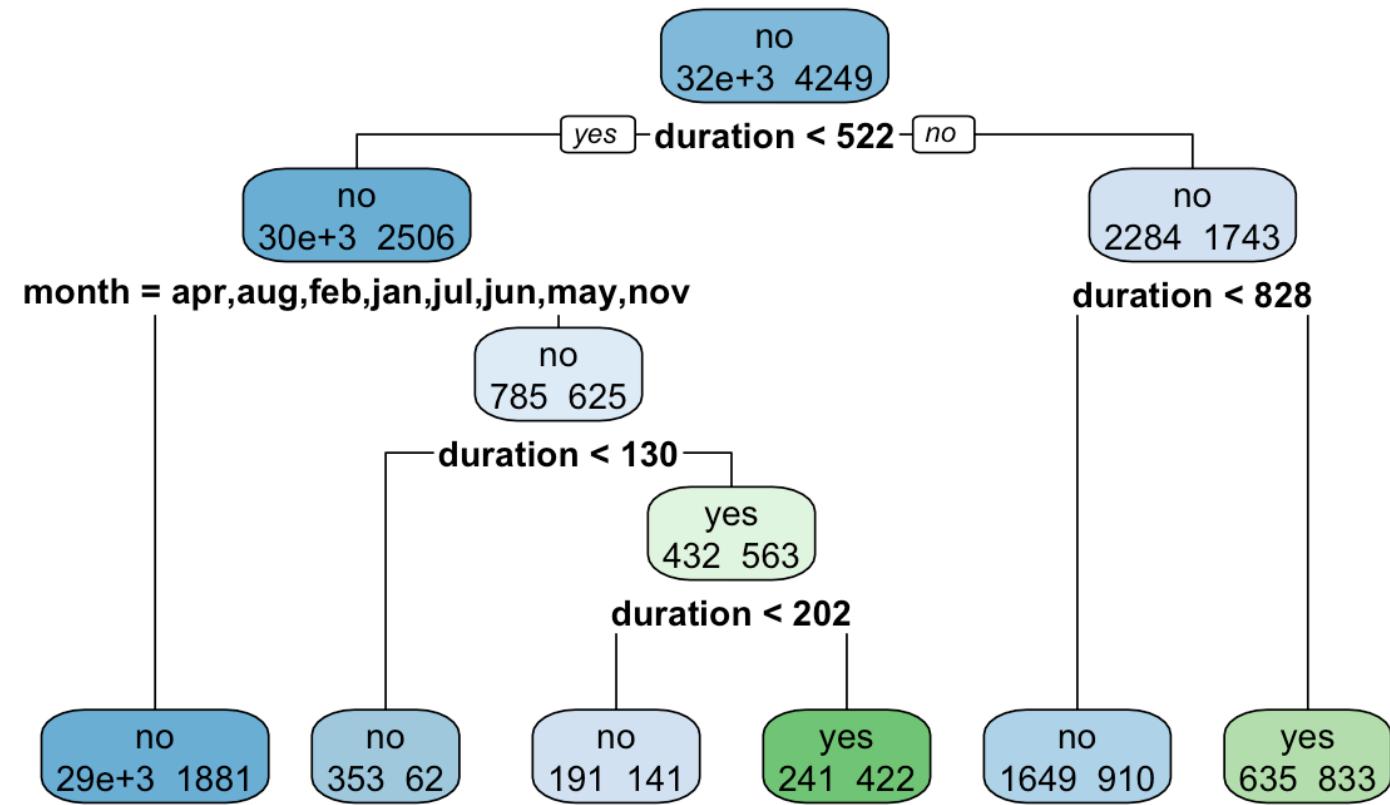
Sensitivity : 0.9696
Specificity : 0.5067
Pos Pred Value : 0.9377
Neg Pred Value : 0.6853
Prevalence : 0.8845
Detection Rate : 0.8576
Detection Prevalence : 0.9146
Balanced Accuracy : 0.7382

'Positive' Class : no

Codes – Random Forest

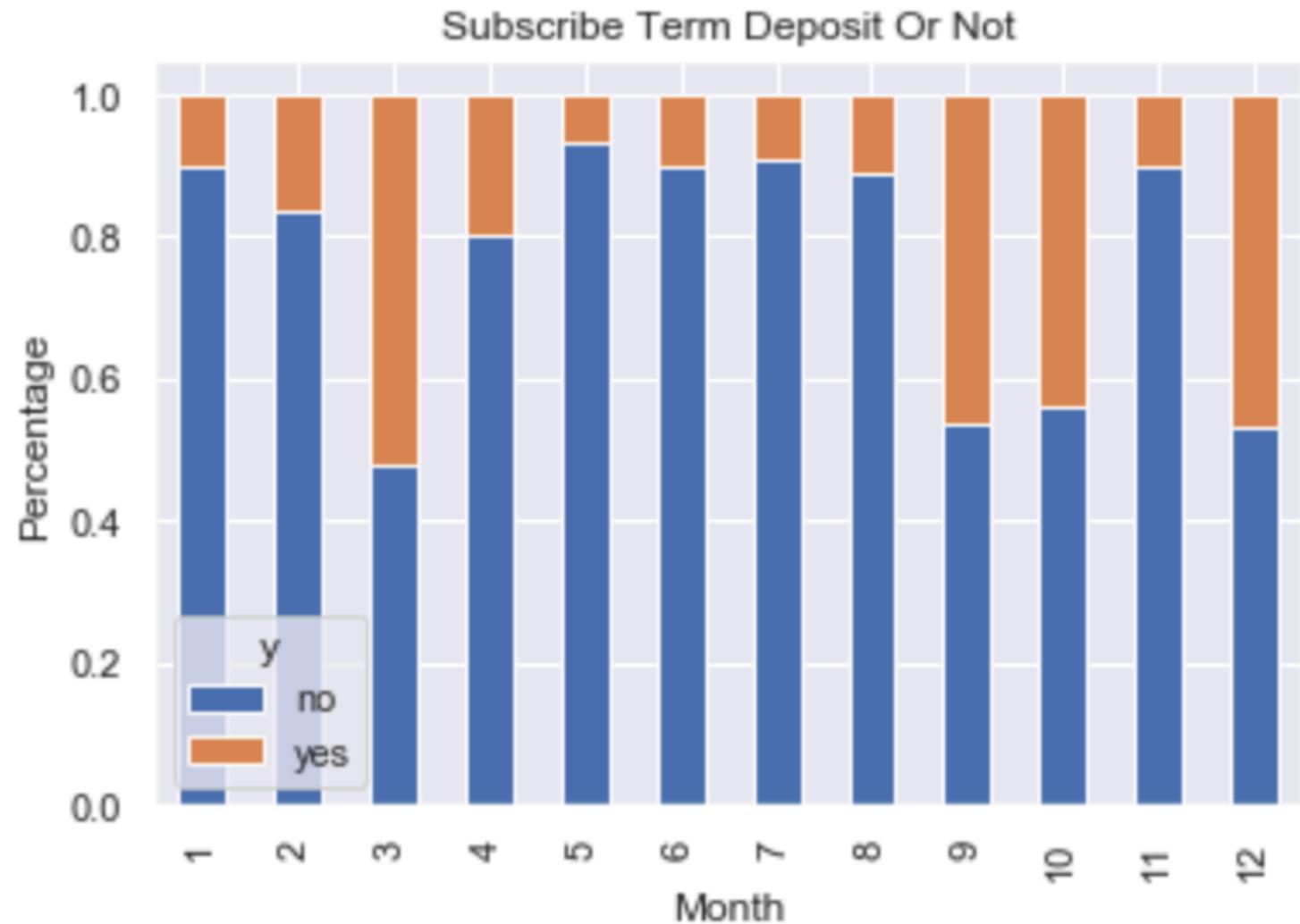


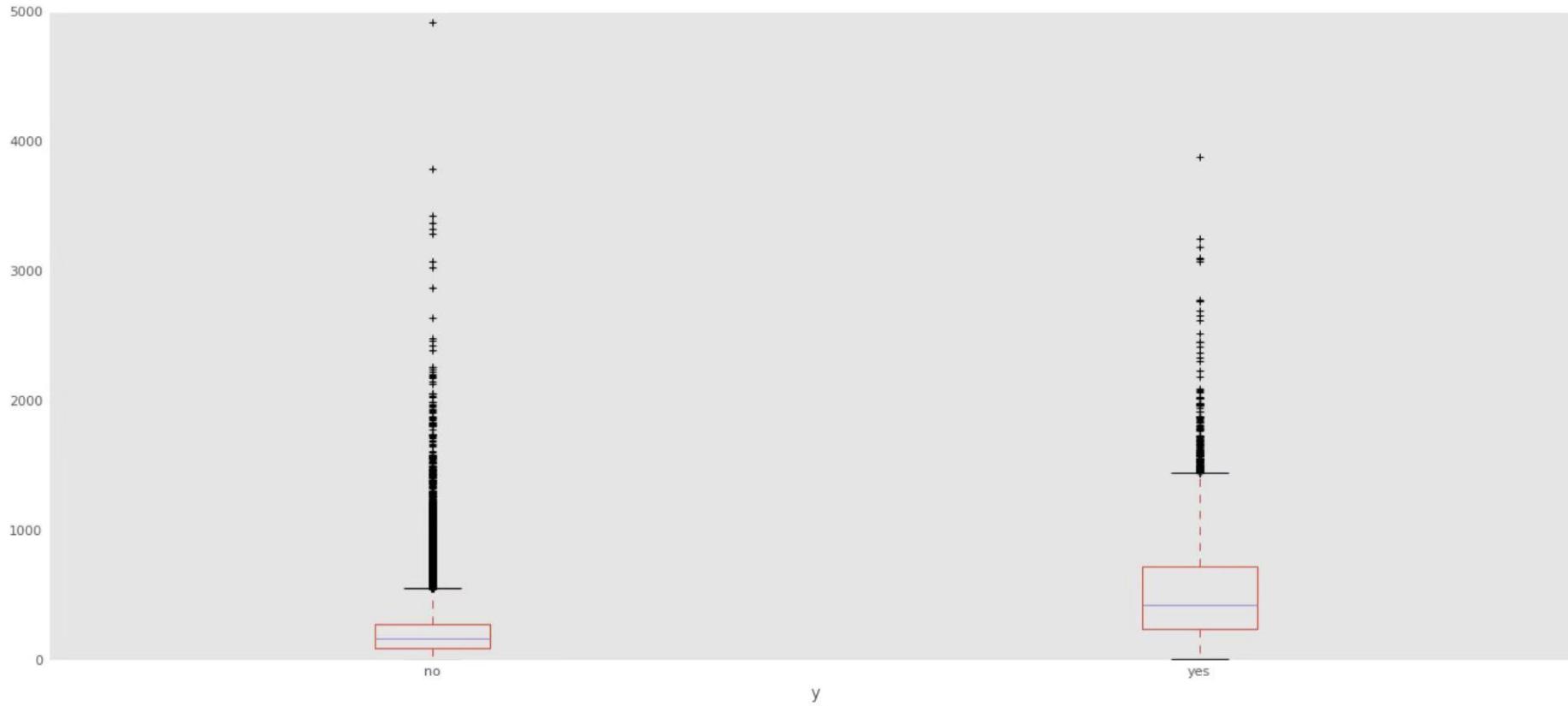
Codes – Decision Tree



Yes if duration > 522 and > 828
if duration < 522 , within when month = march,sep,oct,dec, duration > 202

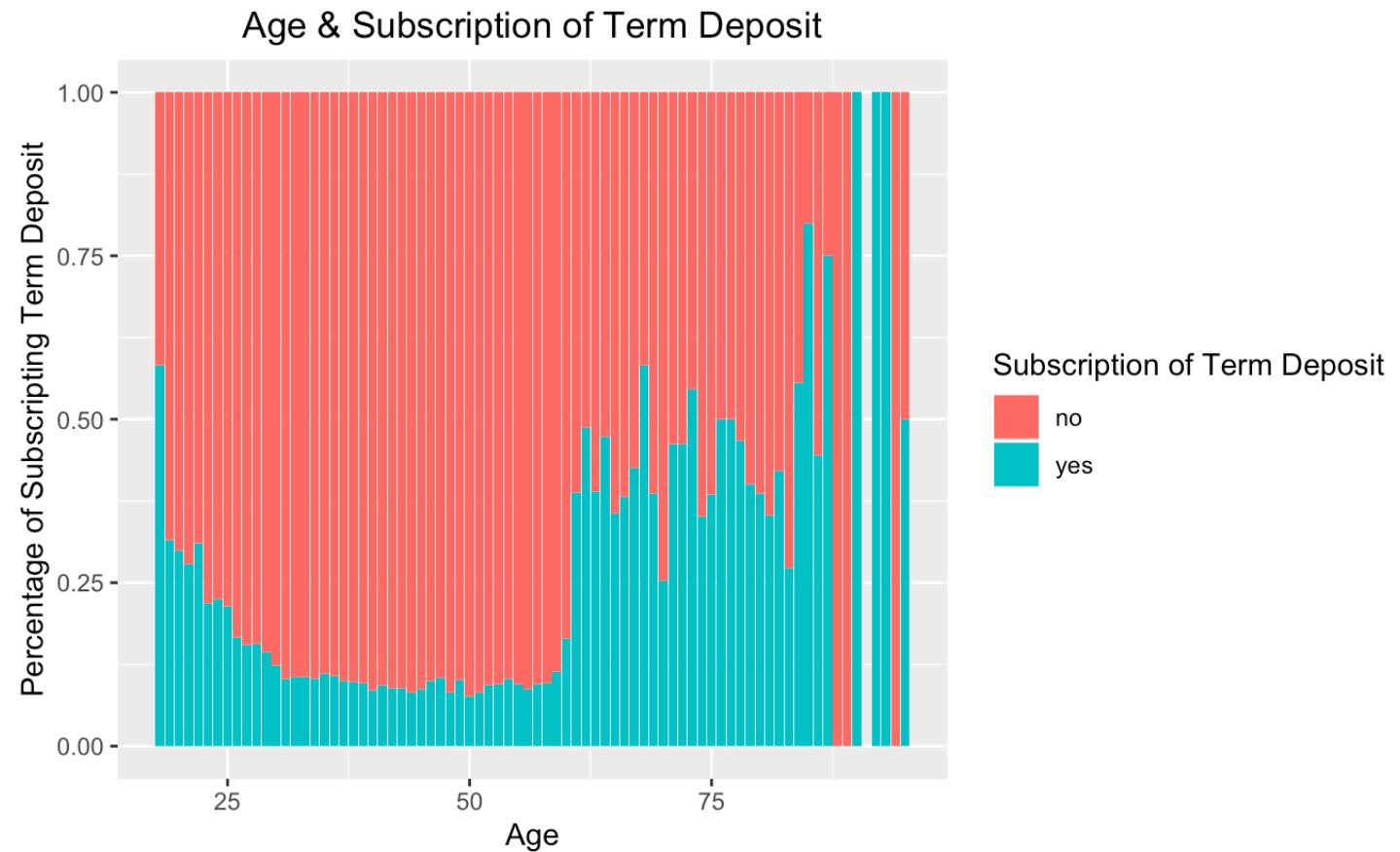
Visualization – Acceptance of term deposit by month



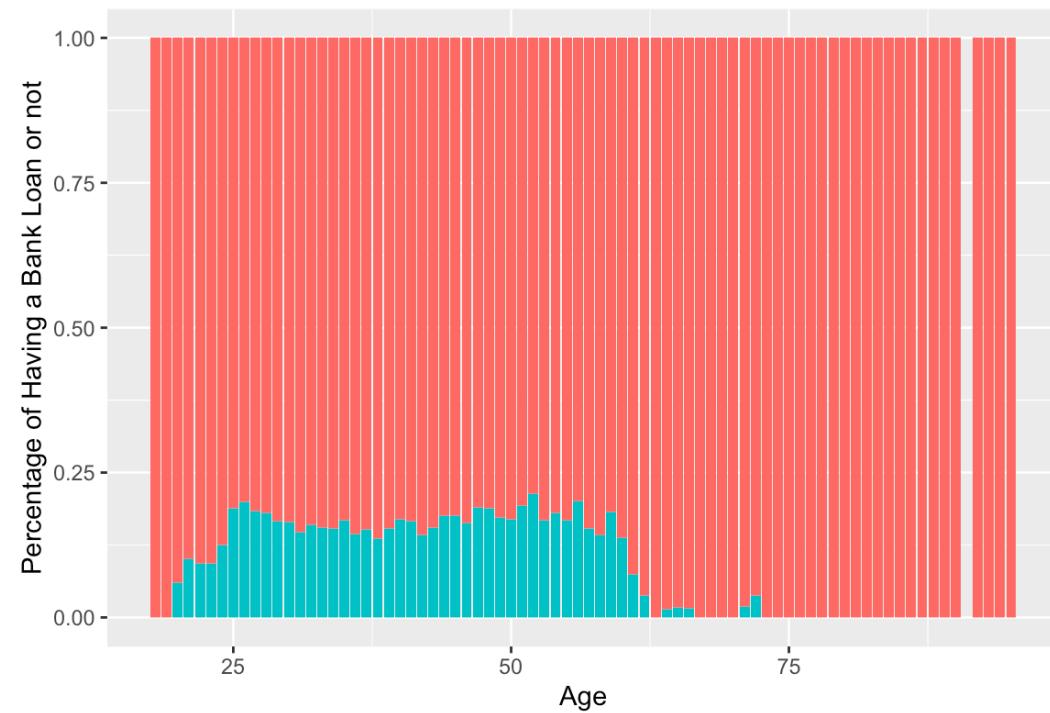


Visualization - Boxplot of Duration and term deposit

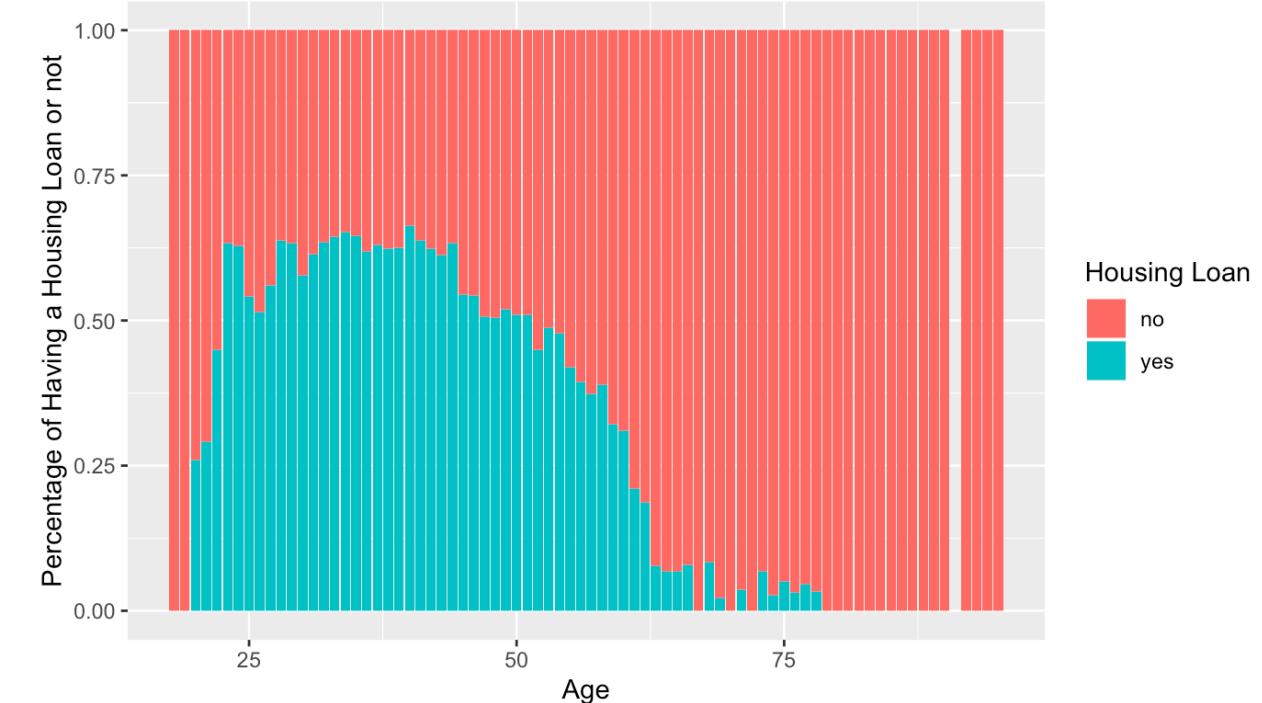
Visualization – Acceptance of term deposit by age



Age & Bank Loan



Age & Housing Loan



Visualization – Acceptance of
term deposit by loan

```

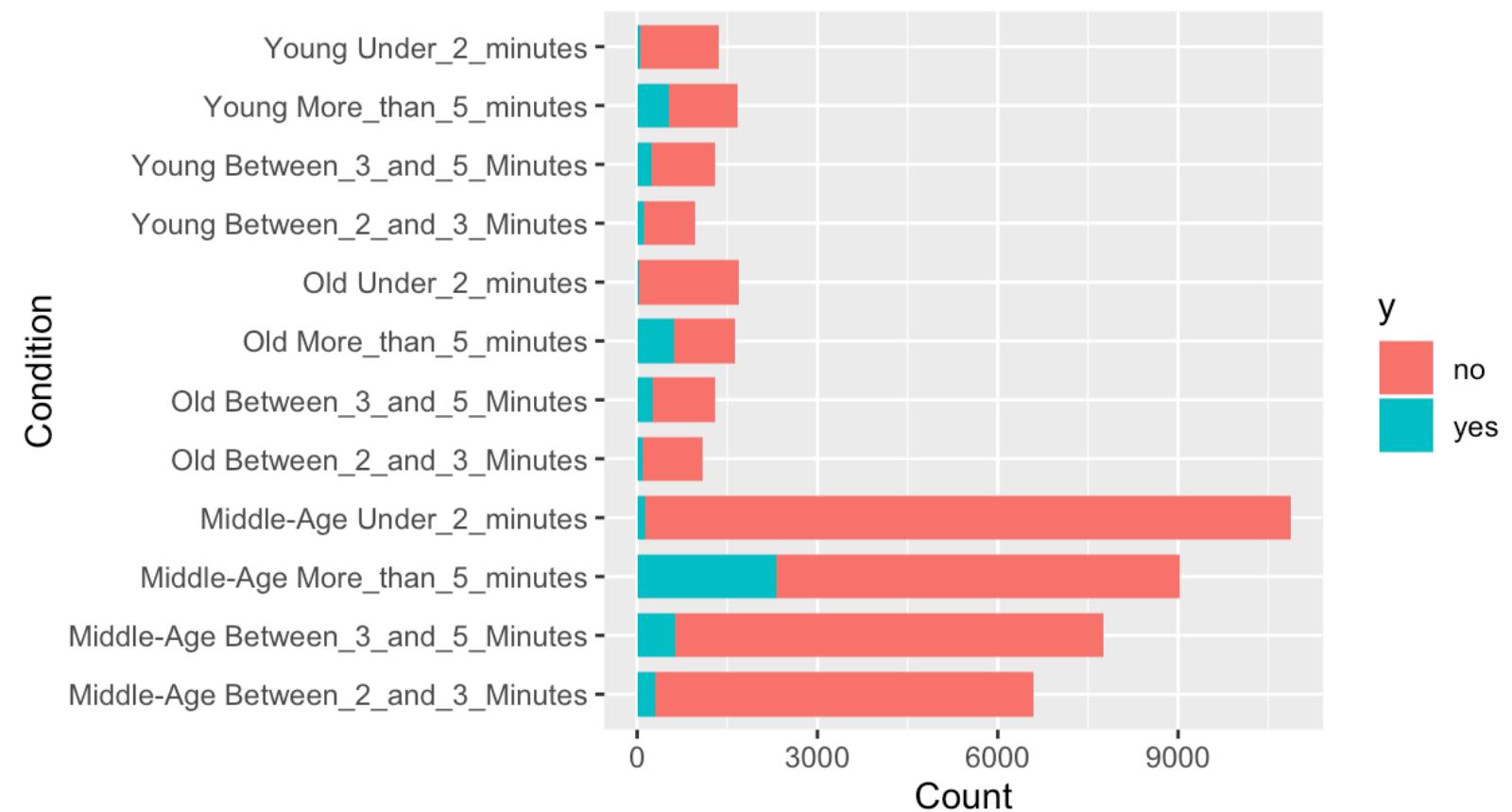
```{r}
library(dplyr)
bank = mutate(bank,
 age_cat = ifelse(age < 30, "Young",
 ifelse(age >= 30 & age < 55, "Middle-Age",
 ifelse(age >= 55, "Old", "N/A"))))

```
```{r}
bank = mutate(bank,
 duration_cat = ifelse(duration < 120, "Under_2_minutes",
 ifelse(duration >= 120 & duration < 180, "Between_2_and_3_Minutes",
 ifelse(duration >= 180 & duration < 300, "Between_3_and_5_Minutes",
 ifelse(duration >= 300, "More_than_5_minutes", "N/A"))))

```
```{r}
bank = mutate(bank, combine = paste(bank$age_cat, bank$duration_cat, sep=" "))
bank
```

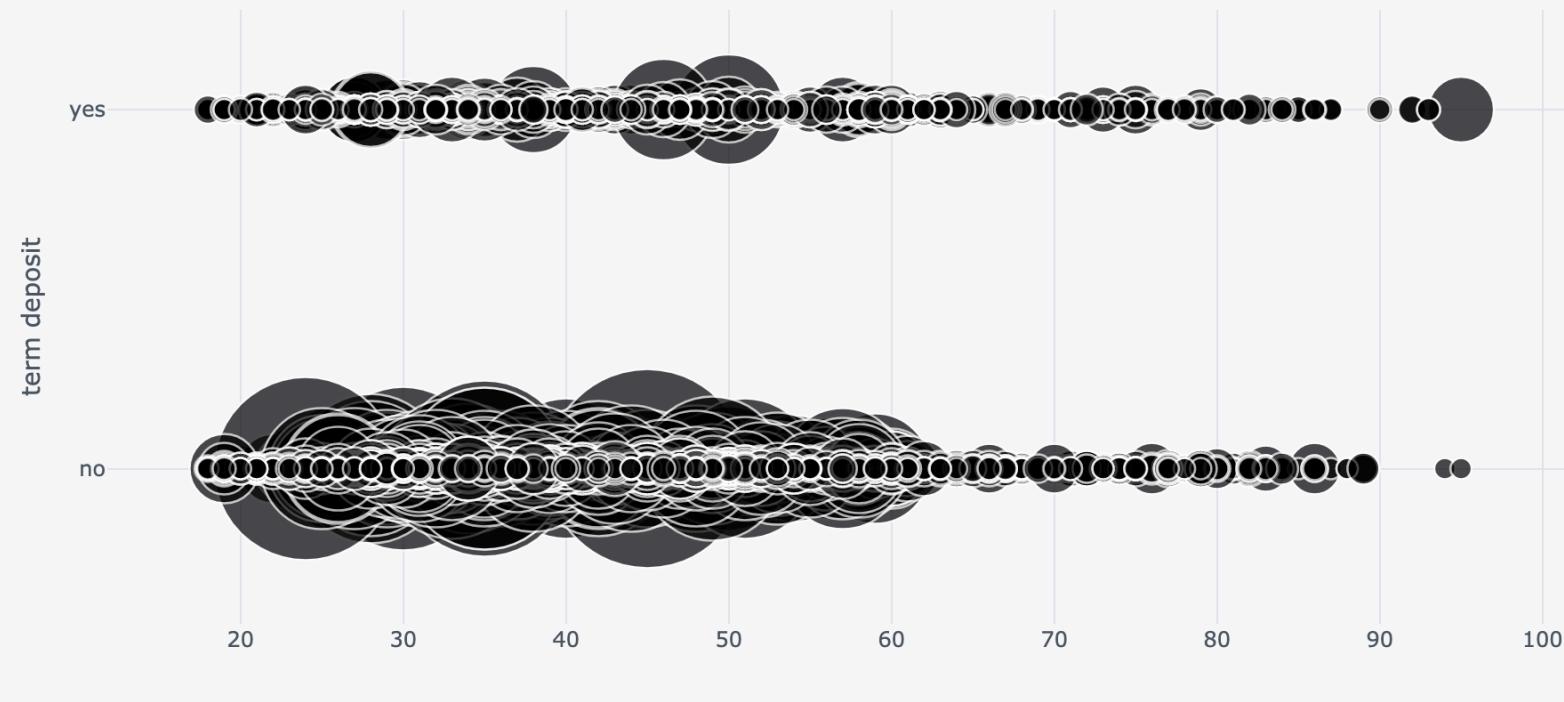
combine
<chr>
Old Between_3_and_5_Minutes
Middle-Age Between_2_and_3_Minutes
Middle-Age Under_2_minutes
Middle-Age Under_2_minutes
Middle-Age Between_3_and_5_Minutes

```



Visualization – Acceptance of Term Deposit grouped by Age Range and Duration

The relationship among age,term deposit and campaign



[Export to plot.ly »](#)

Visualization – Acceptance of
Term Deposit grouped by Age
and Campaign

Recommendations For the Next Campaign

- Age : target under 25 and older than 60
- Month of Marketing Activity : March, September, October and December
- Duration : Make the call more interesting