wide intervals

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
library(ggplot2)
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::combine()       masks randomForest::combine()
## x dplyr::filter()        masks stats::filter()
## x dplyr::lag()           masks stats::lag()
## x purrr::lift()          masks caret::lift()
## x randomForest::margin() masks ggplot2::margin()
```

```
setwd("~/CSP571ProjectGroup")
df <- read_csv("df_with_class_wide.csv")
```

```
## New names:
## * `` -> ...1
```

```
## Rows: 114660 Columns: 17
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (10): number, incident_state, sys_updated_by, contact_type, category, su...
## dbl  (6): ...1, reassignment_count, reopen_count, sys_mod_count, problem_id,...
## lgl  (1): made_sla
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df %>% group_by(class) %>% summarise(n = n())
```

```
## # A tibble: 8 x 2
##   class       n
##   <chr>   <int>
## 1 >14days 12232
## 2 0mins   27618
## 3 14days  11837
## 4 1day     9859
## 5 1hr     15481
## 6 3days   13533
## 7 6hr      9244
## 8 7days   14856
```

```
df = subset(df, select=-c(sys_updated_by, number, subcategory, resolved_updated_diff))
df = subset(df, select=-1)
df$class = as.factor(df$class)
```

```
# try on the full data set (df_small from previous version of the code)
df_small = df
train_idx = createDataPartition(df_small$class, p = 0.8, list=F)
df_train = df_small[train_idx,]
df_test = df_small[-train_idx,]
# try a randomforest model
rf_mod = randomForest(class~., data=df_train, ntree=1000)
df_test_X = subset(df_test, select=-c(class))
df_test_y = df_test$class
pred = predict(rf_mod, df_test)


columns = c("pred", "actual")
acc_df = data.frame(matrix(nrow=nrow(df_test), ncol = length(columns)))
colnames(acc_df) = columns
acc_df$pred = c(pred)
acc_df$actual = c(df_test_y)
#acc_df
acc = nrow(acc_df[acc_df$pred == acc_df$actual,])
acc/nrow(df_test)
```

```
## [1] 0.3918789
```

```
table(pred, df_test$class)
```

```
##
## pred      >14days 0mins 14days 1day  1hr 3days  6hr 7days
##   >14days     373    72    155  135  125   149  100   146
##   0mins        20  4670     22   16   31    23   15    25
##   14days       48     9     67   46   52    58   32    64
##   1day         12     4     12   52    7    19   22    11
##   1hr         820   387    785  652 1971   901  738   935
##   3days       122    21    152  132   86   230  123   175
##   6hr           3     2      8    8    5     4   10     3
##   7days      1048   358   1166  930  819  1322  808  1612
```

```
summary(rf_mod)
```

```
##                 Length Class  Mode
## call                 4 -none- call
## type                 1 -none- character
## predicted        91732 factor numeric
## err.rate          9000 -none- numeric
## confusion           72 -none- numeric
## votes           733856 matrix numeric
## oob.times        91732 -none- numeric
## classes              8 -none- character
## importance          11 -none- numeric
## importanceSD         0 -none- NULL
## localImportance      0 -none- NULL
## proximity            0 -none- NULL
## ntree                1 -none- numeric
## mtry                 1 -none- numeric
## forest              14 -none- list
## y                91732 factor numeric
## test                 0 -none- NULL
## inbag                0 -none- NULL
## terms                3 terms  call
```