R Notebook

```
library(PRP)
library(rstudioapi)
library(mamba)
```

```
## Loading required package: TMB
```

```
## Loading required package: RcppEigen
```

```
## Warning: package 'RcppEigen' was built under R version 4.0.5
```

```
## Loading required package: data.table
```

```
## Warning: package 'data.table' was built under R version 4.0.5
```

```
library(data.table)
library(parallel)
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.0.5
```

```
library(pander)
```

```
## Warning: package 'pander' was built under R version 4.0.5
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

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

```
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1
## v purrr   0.3.4
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'readr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## Warning: package 'forcats' was built under R version 4.0.5

## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()
```

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

set.seed(2021)
```

Functions

```
true_false_pos_rate <- function(out_studies, pred) {

  # truth, SNP that actually contains outliers
  actual <- rowSums(out_studies == 0) > 0

  # true and false positives
  tpos <- (actual == pred) & (actual == TRUE)
  fpos <- ((actual == FALSE) == pred) & (actual == FALSE)

  # false positive rate
  fpos_count <- length(fpos[fpos == TRUE])
  act_neg_count <- length(actual[actual == FALSE])
  fpr <- fpos_count/act_neg_count


  # true positive rate
  tpos_count <- length(tpos[tpos == TRUE])
  act_pos_count <- length(actual[actual == TRUE])
  tpr <- tpos_count/act_pos_count
  return(c(fpr, tpr))
}
```

```
roc_rates <- function(contain_outlier, # actual outliers
                      mamba_ppr_val, # vector of mamba ppr values
                      prp_val, # vector of prp values
                      interval = 0.05) {
```

```
  mamba_ppr_rates <- list()
  prp_rates <- list()

  # iterate through different cutoffs
  for (cutoff in seq(from = 0, to = 1, by = interval)) {

    # outlier or not based on cutoff
    mamba_ppr_pred <- mamba_ppr_val < cutoff
    prp_pred <- prp_val < cutoff



    mamba_ppr_rates <- rbind(mamba_ppr_rates,
                             true_false_pos_rate(contain_outlier,
                                                 mamba_ppr_pred))


    prp_rates <- rbind(prp_rates,
                       true_false_pos_rate(contain_outlier,
                                           prp_pred))
  }

  rates_list <- list("mamba_ppr_rates" = mamba_ppr_rates, "prp_rates" = prp_rates)
  return(rates_list)
}
```

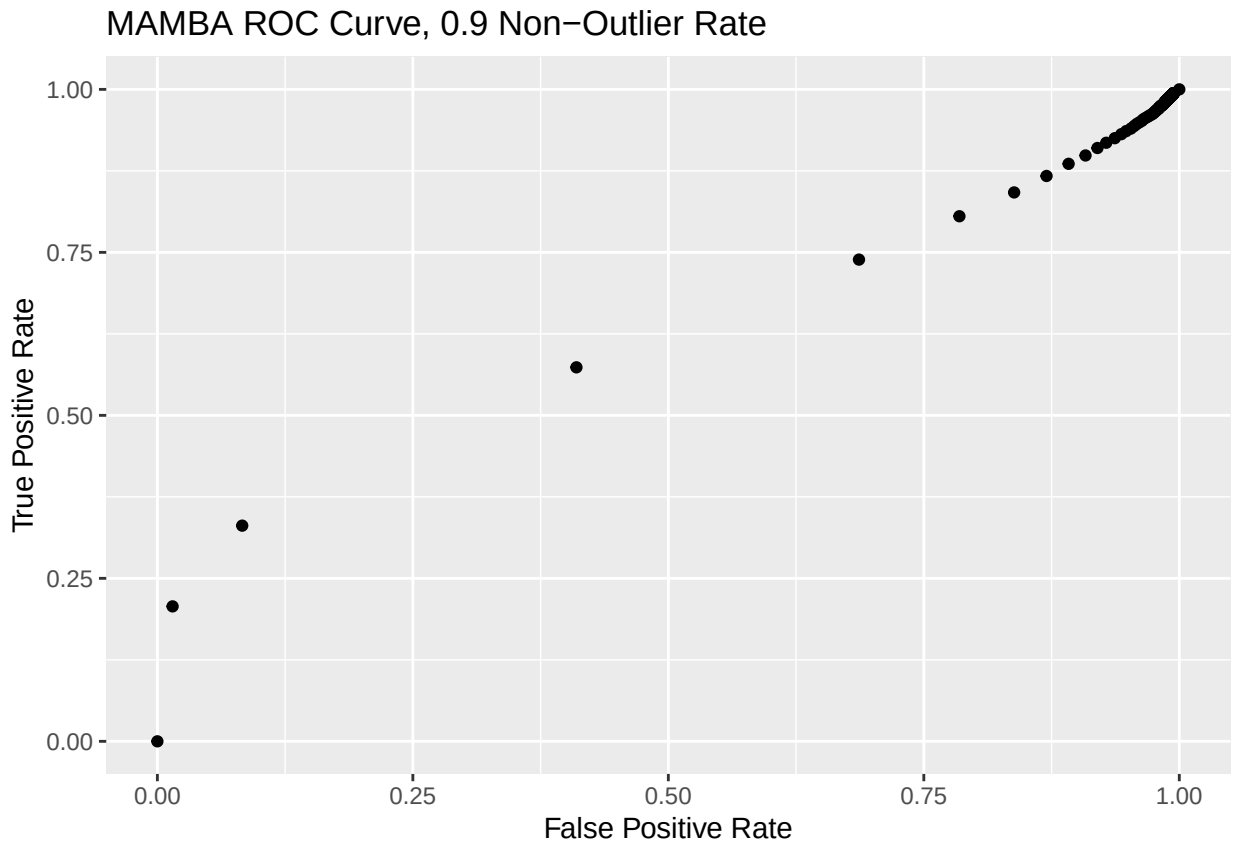
```
# function for graphing ROC
plot_point <- function(data,
                       title = "") {

  mamba_df <- as.data.frame(data)

  ggplot(data = mamba_df,
         aes(x = as.numeric(mamba_df[, 1]),
             y = as.numeric(mamba_df[, 2]))) +
    geom_point() +
    ggtitle(title) +
    xlab("False Positive Rate") +
    ylab("True Positive Rate")

}
```

0.9 nonoutlier rate

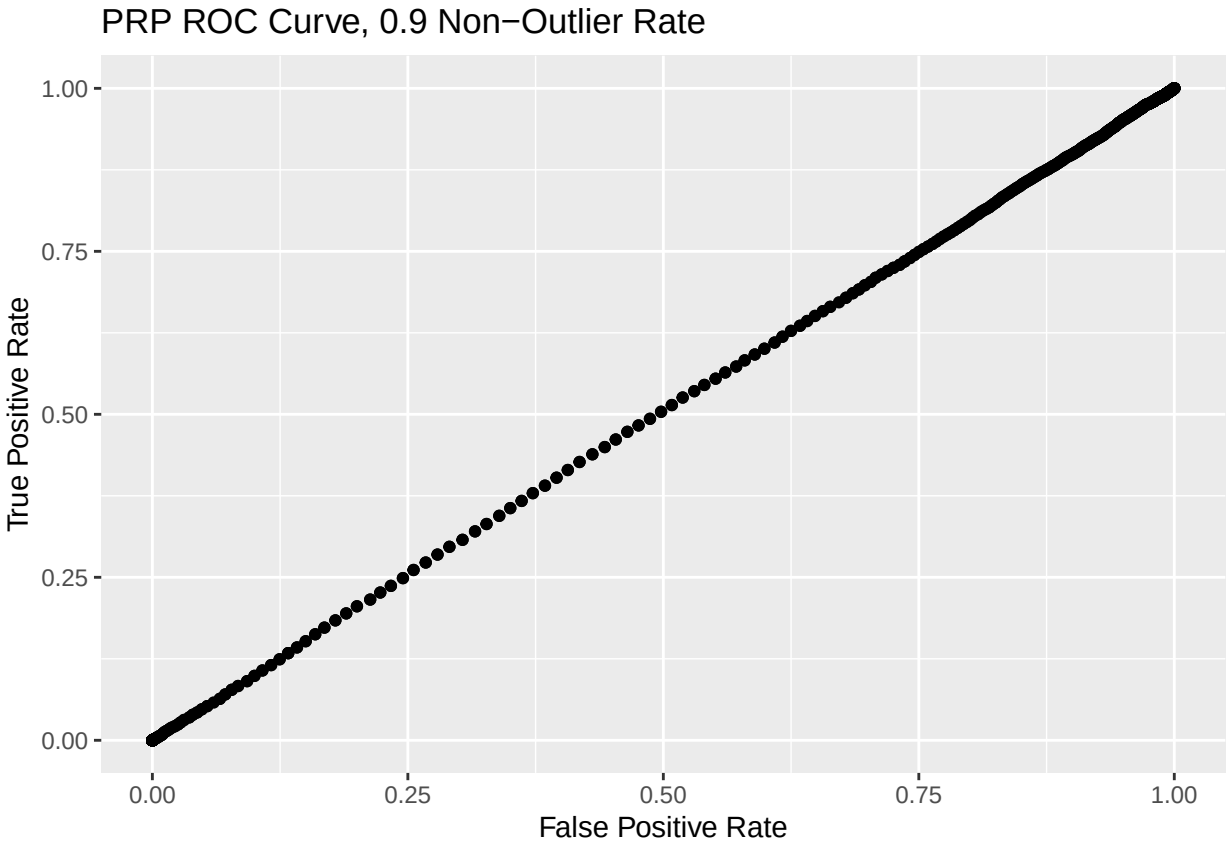
```
load(file = "../data/post_prp_data_pval_p90.rda")
load(file = "../data/sim_mamba_mod_p90.rda")
load(file = "../data/mamba_data_p90.rda")

p90_outliers <- mamba_data_p90$Ojk
p90_mamba_ppr <- sim_mod_p90$ppr
p90_prp <- post_prp_data_pval_p90
```

```
p90_rates <- roc_rates(contain_outlier = p90_outliers,
                       mamba_ppr_val = p90_mamba_ppr,
                       prp_val = p90_prp,
                       interval = 0.0005)
```

```
plot_point(p90_rates$mamba_ppr_rates,
           title = "MAMBA ROC Curve, 0.9 Non-Outlier Rate")
```




```
plot_point(p90_rates$prp_rates,
           title = "PRP ROC Curve, 0.9 Non-Outlier Rate")
```

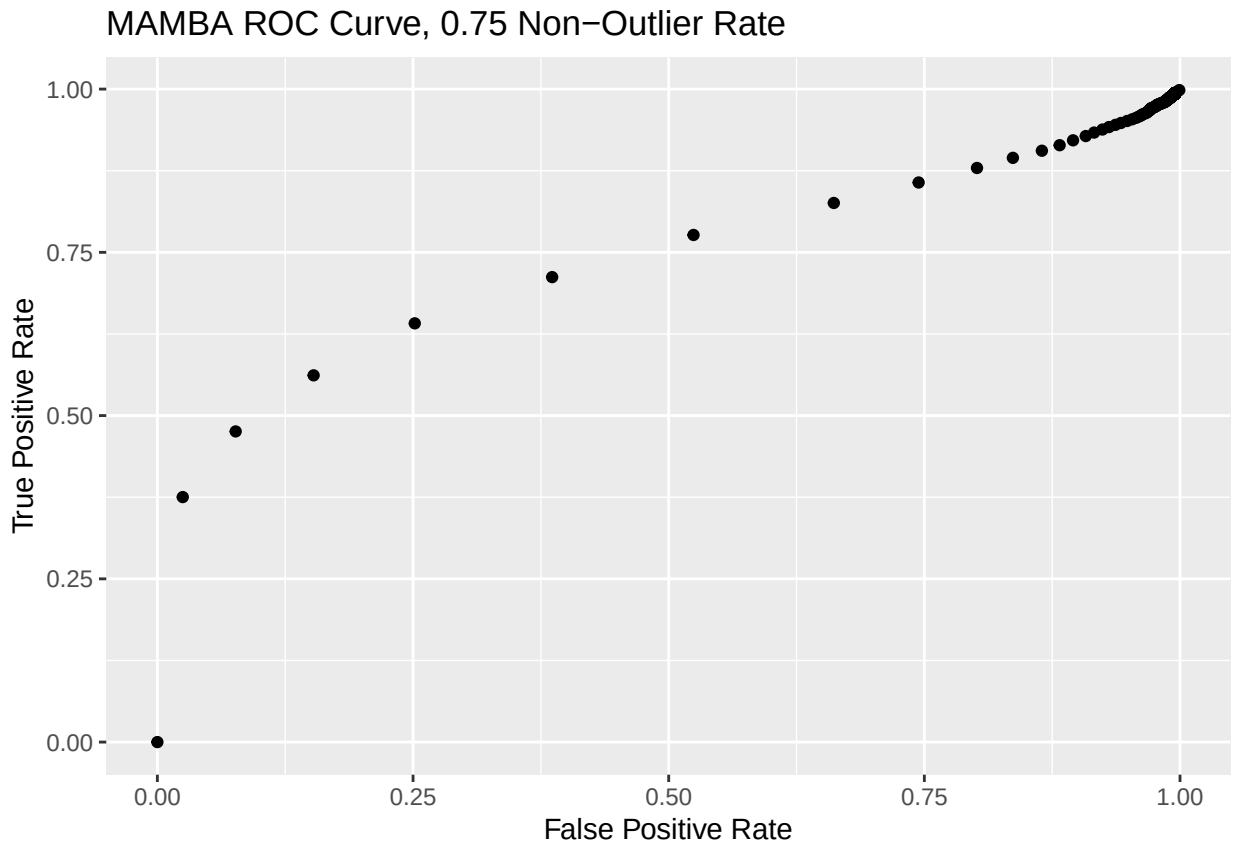

PRP ROC Curve, 0.9 Non−Outlier Rate
True Positive Rate
False Positive Rate
0.00
0.25
0.50
0.75
1.00

0.75 nonoutlier rate

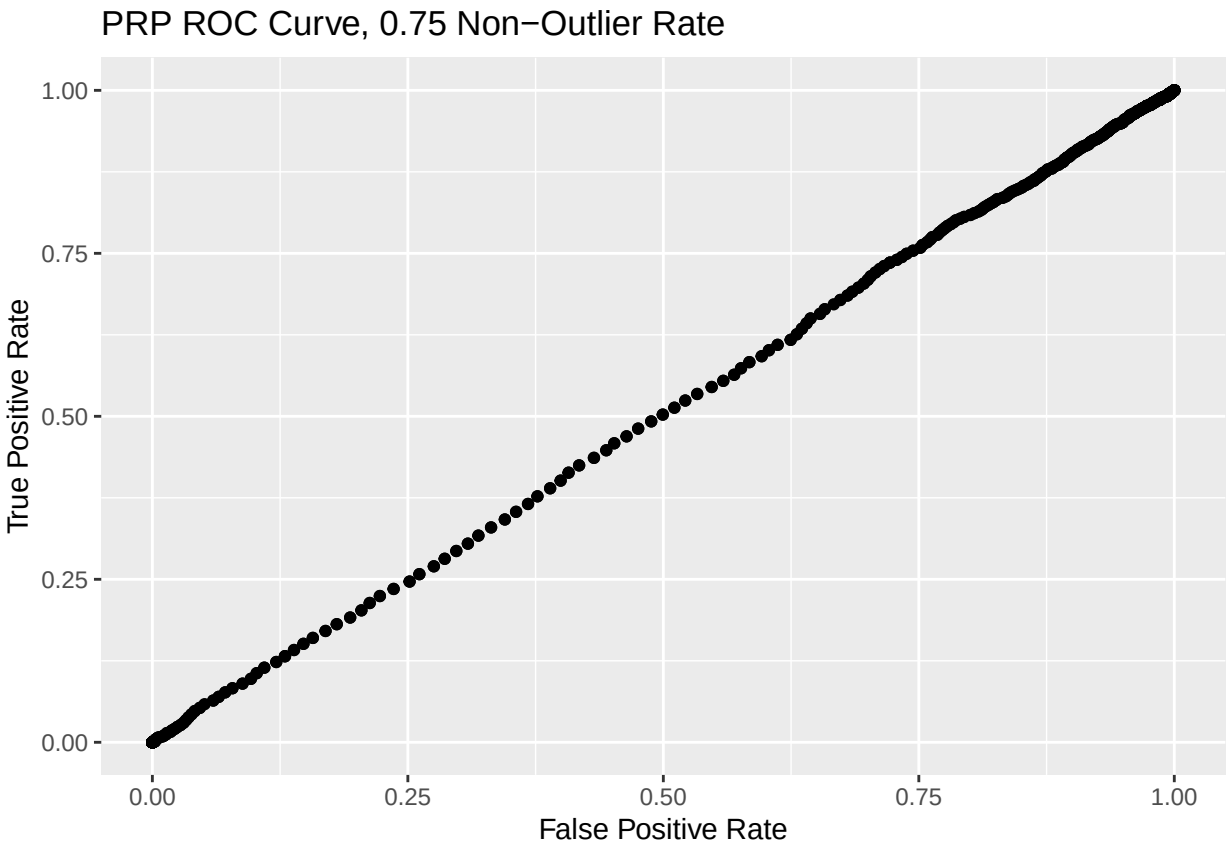
```
load(file = "../data/post_prp_data_pval_p75.rda")
load(file = "../data/sim_mamba_mod_p75.rda")
load(file = "../data/mamba_data_p75.rda")

p75_outliers <- mamba_data_p75$Ojk
p75_mamba_ppr <- sim_mod_p75$ppr
p75_prp <- post_prp_data_pval_p75
```

```
p75_rates <- roc_rates(contain_outlier = p75_outliers,
                       mamba_ppr_val = p75_mamba_ppr,
                       prp_val = p75_prp,
                       interval = 0.0005)
```

```
plot_point(p75_rates$mamba_ppr_rates,
           title = "MAMBA ROC Curve, 0.75 Non-Outlier Rate")
```




```
plot_point(p75_rates$prp_rates,
           title = "PRP ROC Curve, 0.75 Non-Outlier Rate")
```

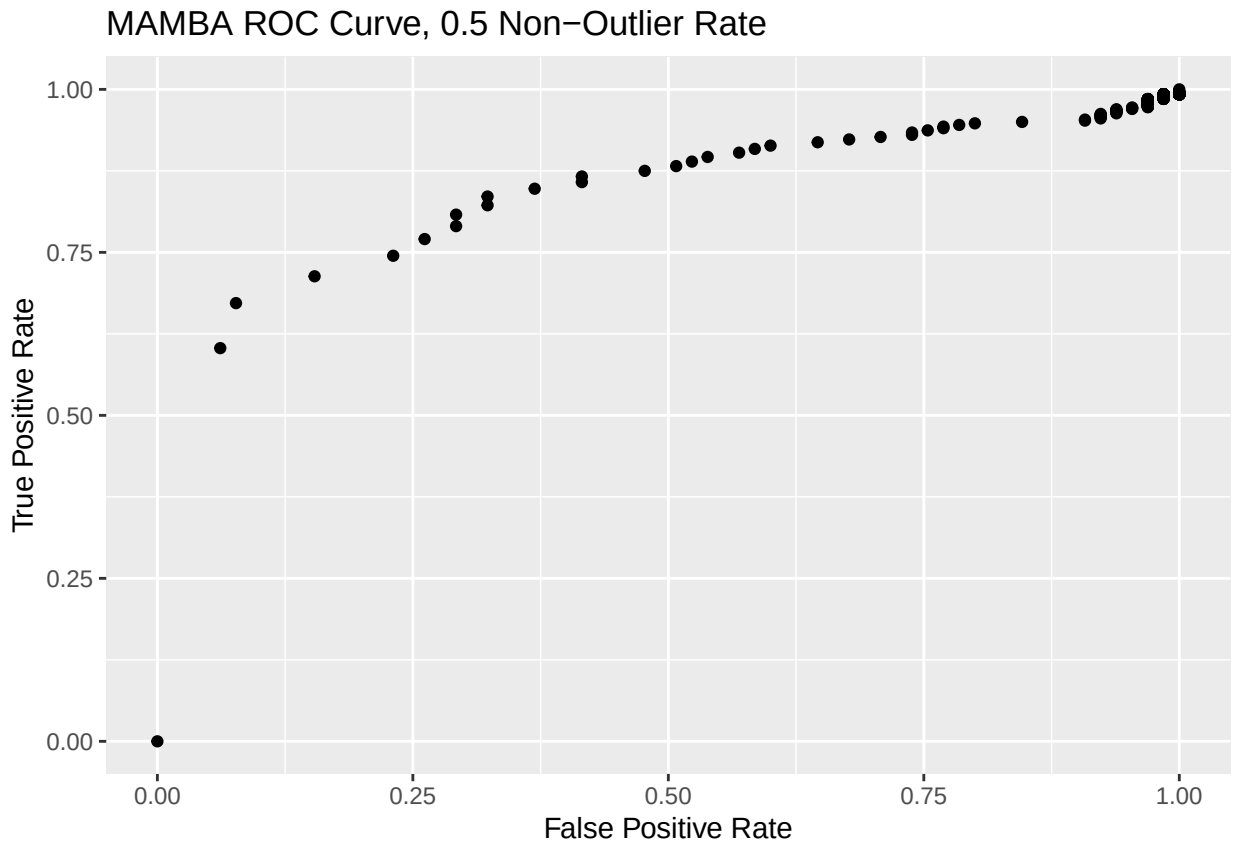

PRP ROC Curve, 0.75 Non−Outlier Rate
True Positive Rate
False Positive Rate
0.00
0.25
0.50
0.75
1.00

0.5 nonoutlier rate

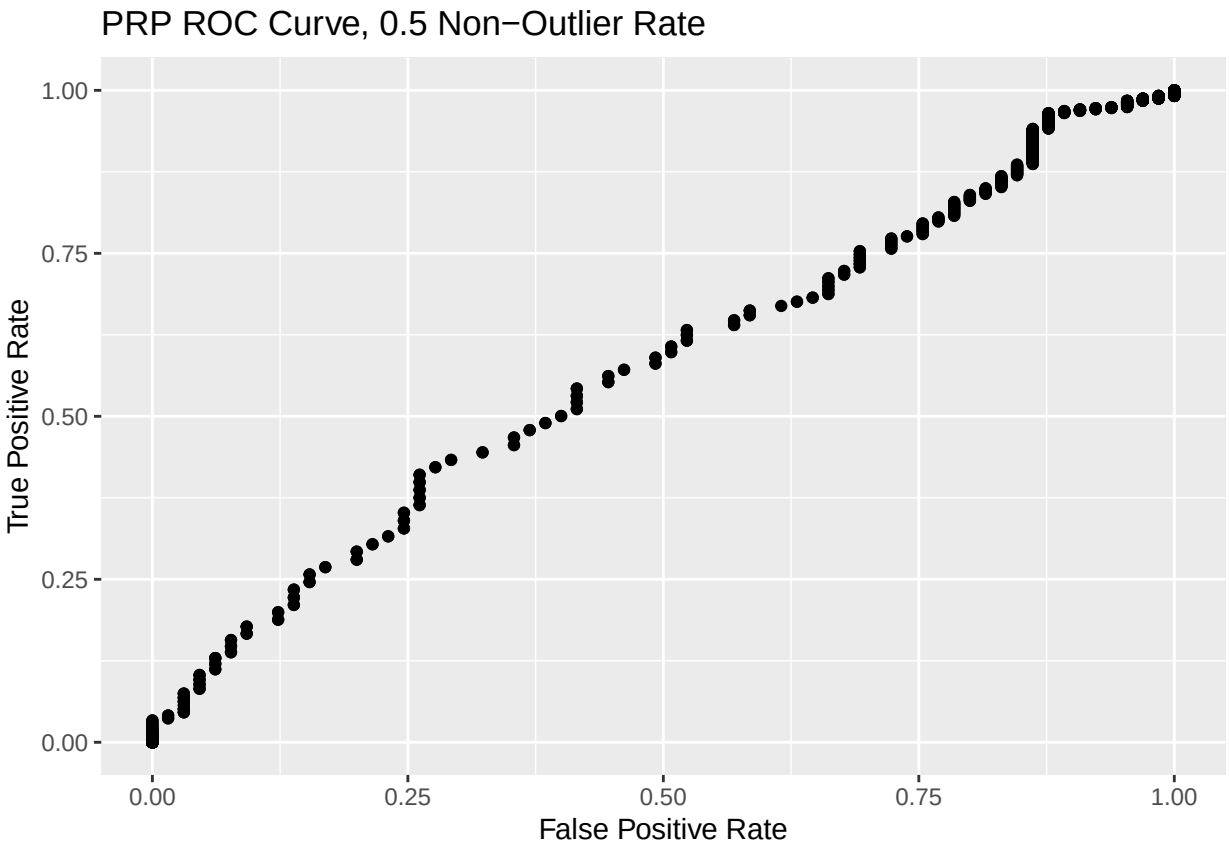
```
load(file = "../data/post_prp_data_pval_p50.rda")
load(file = "../data/sim_mamba_mod_p50.rda")
load(file = "../data/mamba_data_p50.rda")

p50_outliers <- mamba_data_p50$Ojk
p50_mamba_ppr <- sim_mod_p50$ppr
p50_prp <- post_prp_data_pval_p50
```

```
p50_rates <- roc_rates(contain_outlier = p50_outliers,
                       mamba_ppr_val = p50_mamba_ppr,
                       prp_val = p50_prp,
                       interval = 0.0005)
```

```
plot_point(p50_rates$mamba_ppr_rates,
           title = "MAMBA ROC Curve, 0.5 Non-Outlier Rate")
```




```
plot_point(p50_rates$prp_rates,
           title = "PRP ROC Curve, 0.5 Non-Outlier Rate")
```


PRP ROC Curve, 0.5 Non−Outlier Rate
True Positive Rate
False Positive Rate
0.00
0.25
0.50
0.75
1.00