R Notebook

Functions

```r
true_false_pos_rate <- function(out_studies, pred, n_out_thres = 1) {

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

  # 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))
}
```

```r
roc_rates <- function(contain_outlier, # actual outliers
                      mamba_ppr_val, # vector of mamba ppr values
                      prp_val, # vector of prp values
                      interval = 0.05,
                      out_thres = 1) {

  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,
```

```r
                                               mamba_ppr_pred,
                                               n_out_thres = out_thres))


  prp_rates <- rbind(prp_rates,
                     true_false_pos_rate(contain_outlier,
                                         prp_pred,
                                         n_out_thres = out_thres))
  }

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

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

  mamba_df <- as.data.frame(data)
  x <- seq(from = 0, to = 1, by = 0.05)
  fake_df <-data.frame(x)

  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") +
    stat_function(fun = function(x) {
      x
    })

}
```
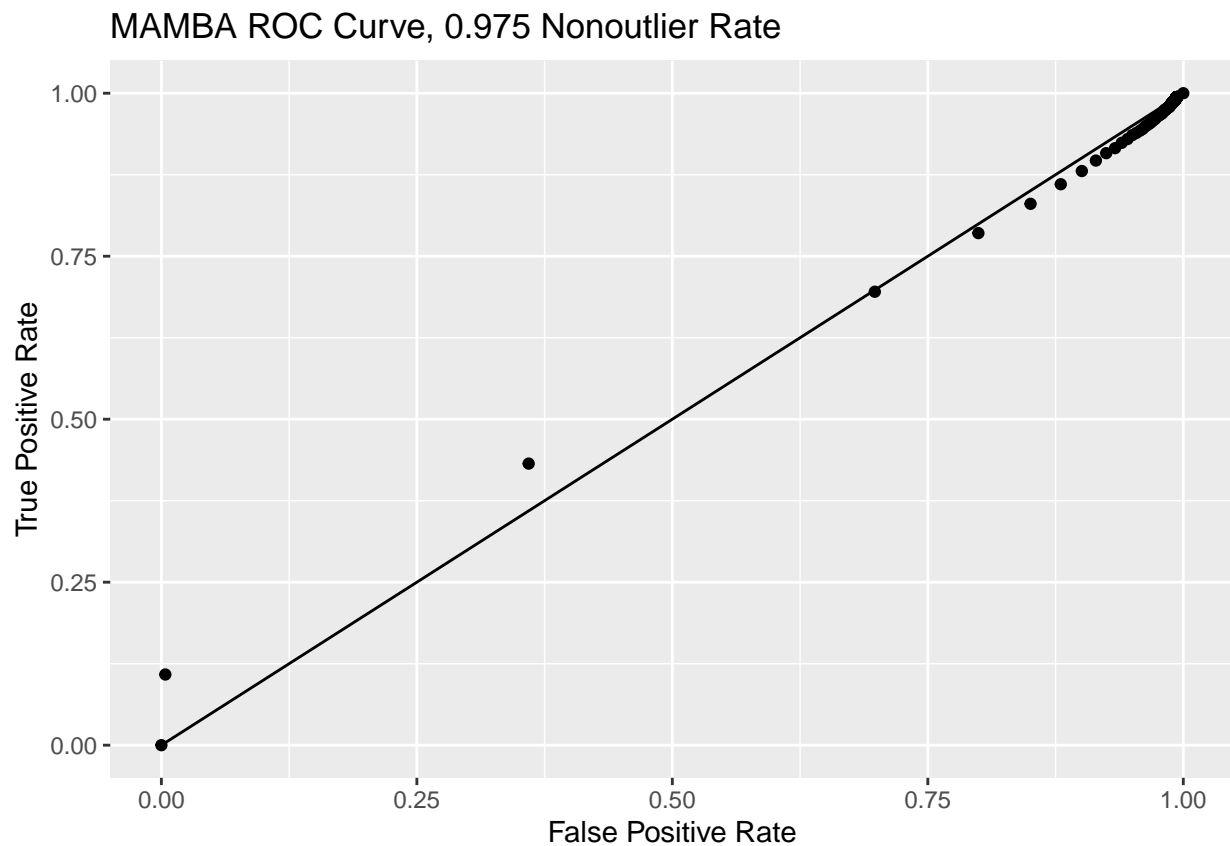
0.975 Nonoutlier Rate

Here, we look at the case where 0.975 is "the proportion of non-replicable SNPs which are well behaved, or nonoutliers" (quoted from the MAMBA documentation).

```
load(file = "../data/post_prp_data_pval.rda")
load(file = "../data/sim_mamba_mod.rda")
load(file = "../data/mamba_data.rda")

p98_outliers <- mamba_data$Ojk
p98_mamba_ppr <- sim_mod$ppr
p98_prp <- post_prp_data_pval
```
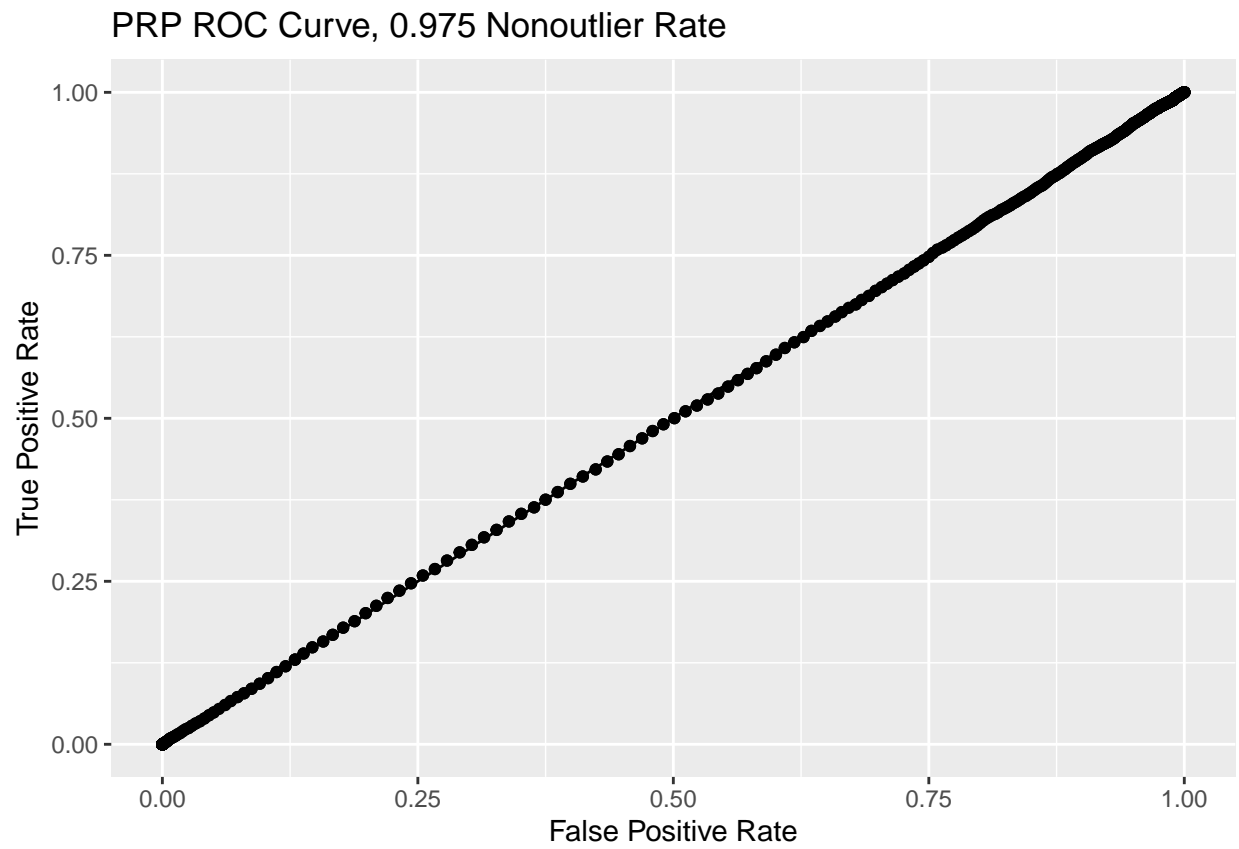
```
p98_rates <- roc_rates(contain_outlier = p98_outliers,
                       mamba_ppr_val = p98_mamba_ppr,
                       prp_val = p98_prp,
                       interval = 0.0005,
                       out_thres = 1)
```

```
plot_point(p98_rates$mamba_ppr_rates,
           title = "MAMBA ROC Curve, 0.975 Nonoutlier Rate")
```



```
plot_point(p98_rates$prp_rates,
           title = "PRP ROC Curve, 0.975 Nonoutlier Rate")
```

PRP ROC Curve, 0.975 Nonoutlier Rate
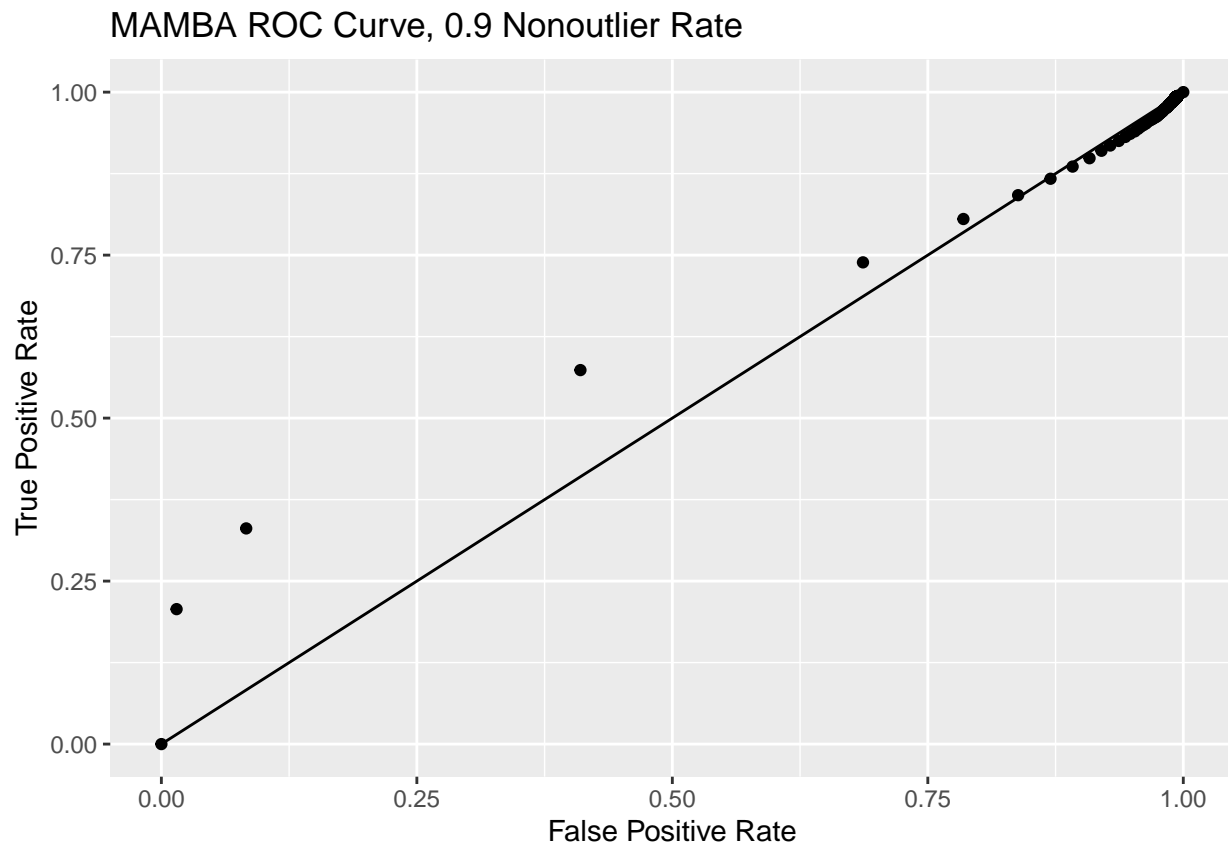
0.9 Nonoutlier Rate

Here, we look at the case where 0.9 is "the proportion of non-replicable SNPs which are well behaved, or nonoutliers."

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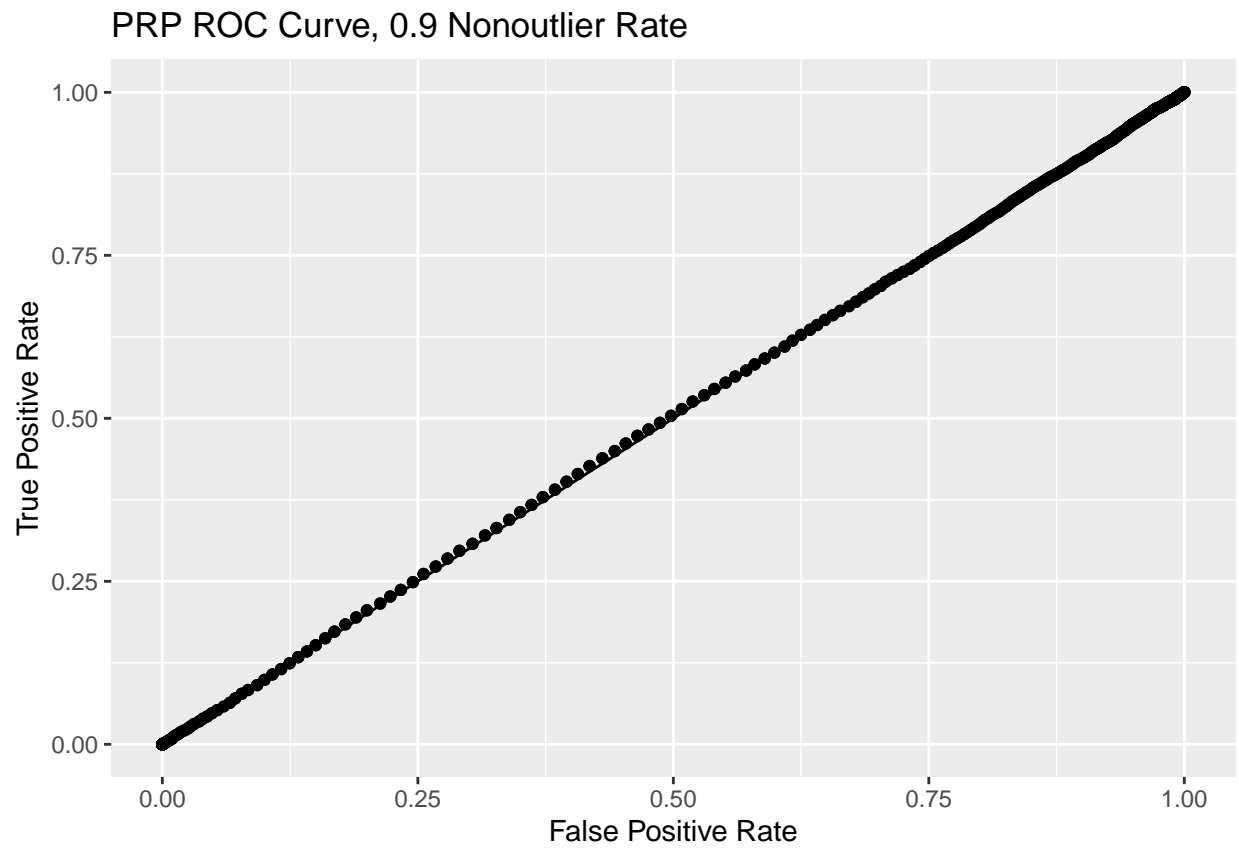
```
plot_point(p90_rates$mamba_ppr_rates,
           title = "MAMBA ROC Curve, 0.9 Nonoutlier Rate")
```



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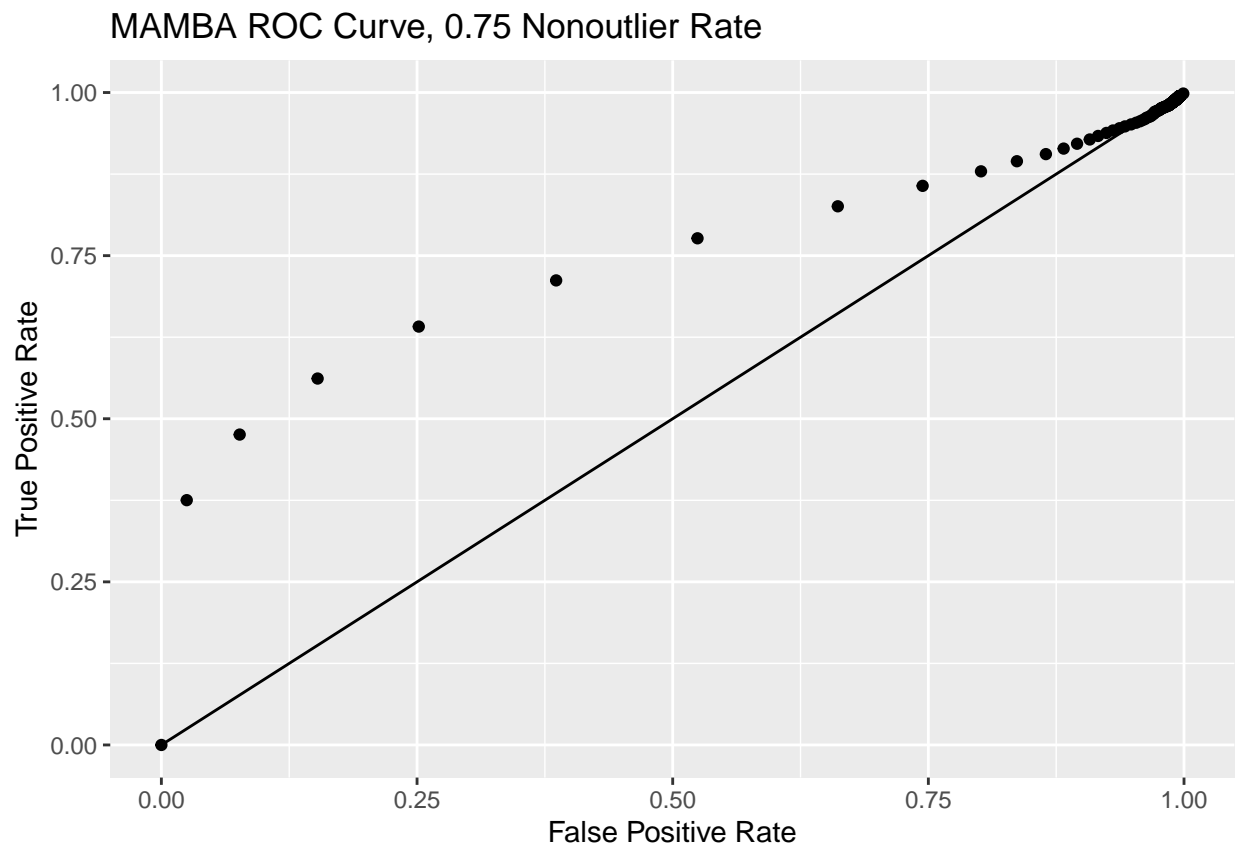
PRP ROC Curve, 0.9 Nonoutlier Rate

0.75 Nonoutlier Rate Here, we look at the case where 0.75 is "the proportion of non-replicable SNPs which are well behaved, or nonoutliers."

```
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 Nonoutlier Rate")
```
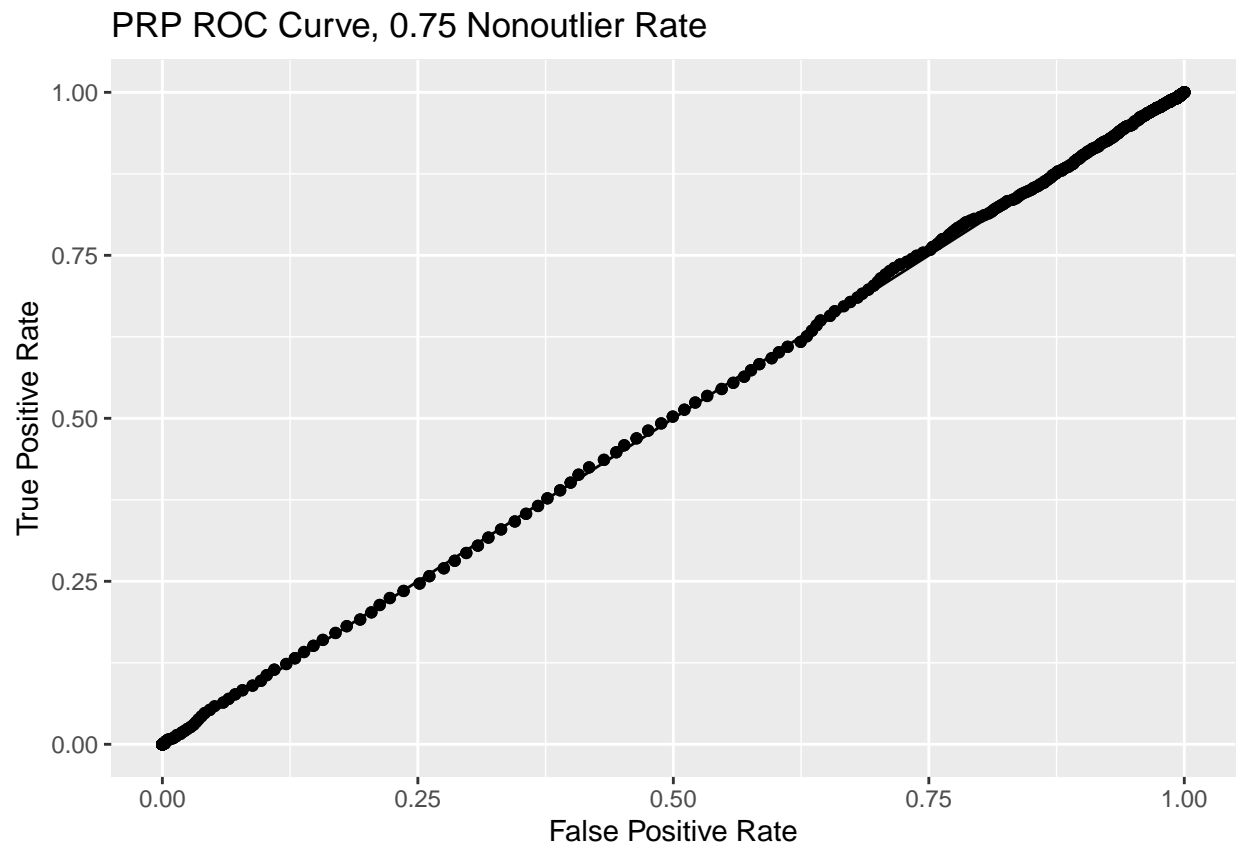
MAMBA ROC Curve, 0.75 Nonoutlier Rate


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

PRP ROC Curve, 0.75 Nonoutlier Rate
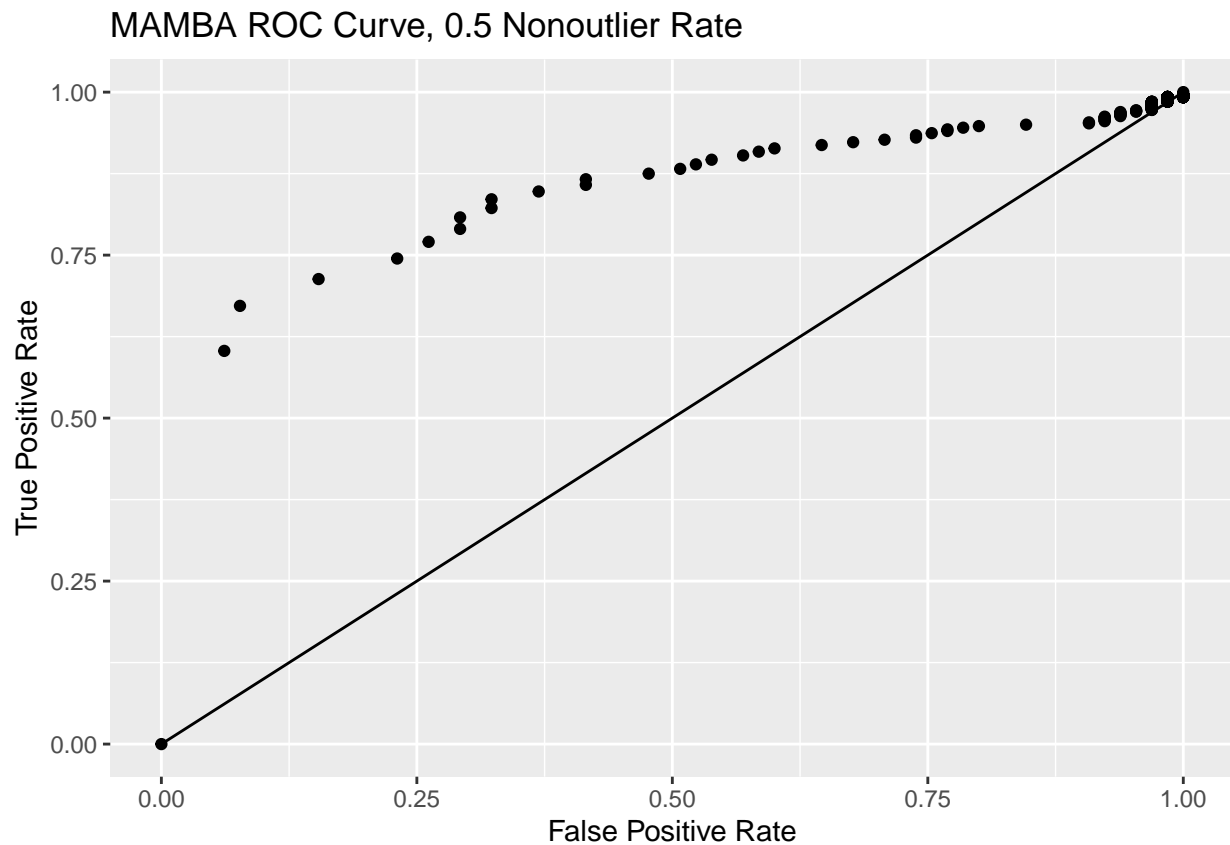
0.5 Nonoutlier Rate

We look at the case where 0.5 is "the proportion of non-replicable SNPs which are well behaved, or nonoutliers."

```r
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)
```
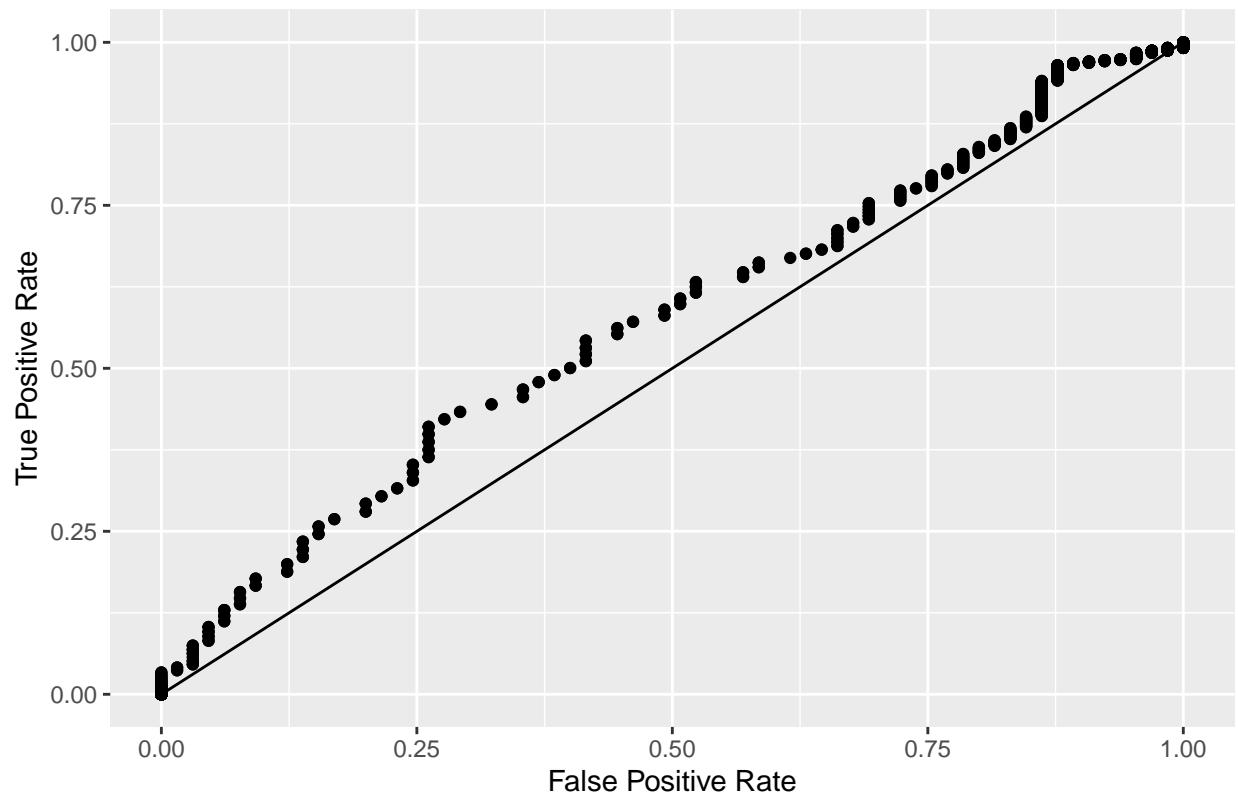
```r
plot_point(p50_rates$mamba_ppr_rates,
           title = "MAMBA ROC Curve, 0.5 Nonoutlier Rate")
```



```r
plot_point(p50_rates$prp_rates,
           title = "PRP ROC Curve, 0.5 Nonoutlier Rate")
```

PRP ROC Curve, 0.5 Nonoutlier Rate
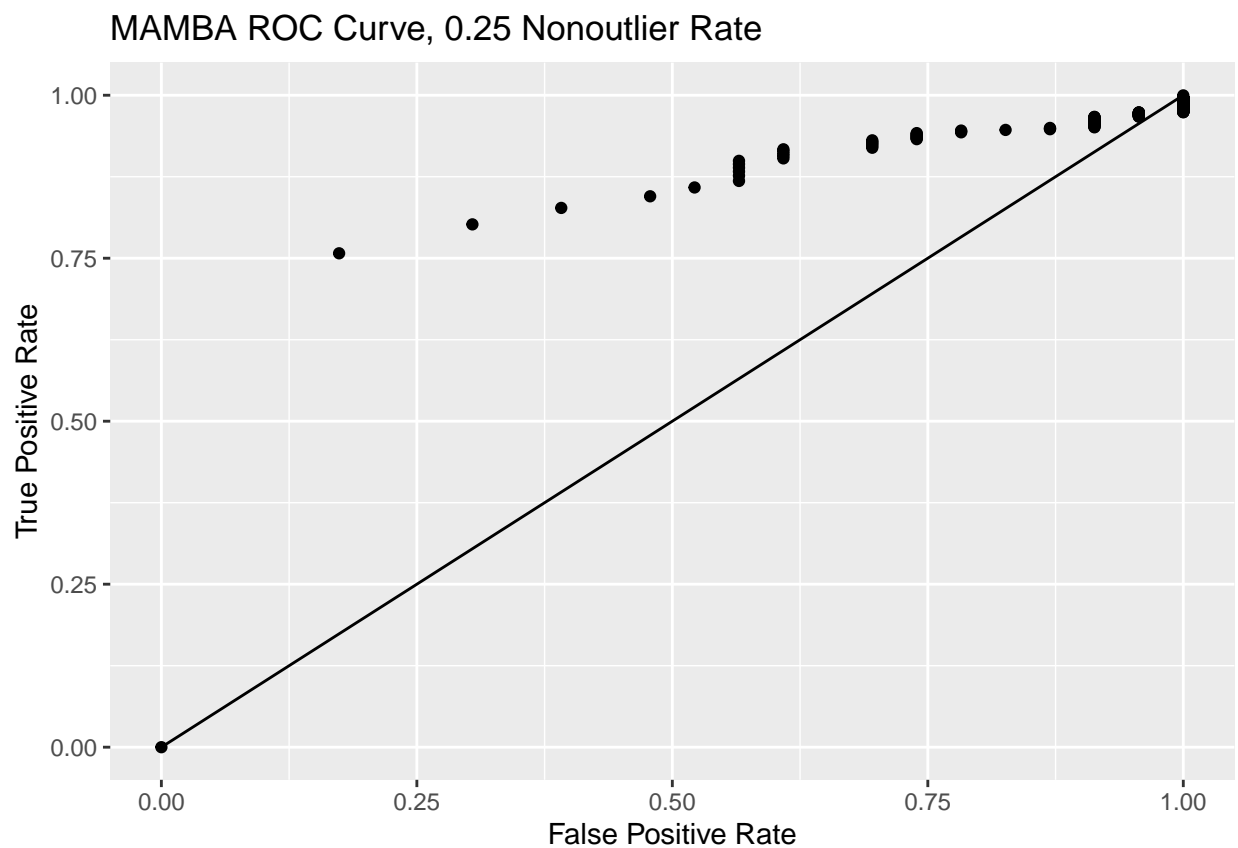
0.25 Nonoutlier Rate

We look at the case where 0.25 is "the proportion of non-replicable SNPs which are well behaved, or nonoutliers."

```
load(file = "../data/post_prp_data_pval_p25.rda")
load(file = "../data/sim_mamba_mod_p25.rda")
load(file = "../data/mamba_data_p25.rda")

p25_outliers <- mamba_data_p25$Ojk
p25_mamba_ppr <- sim_mod_p25$ppr
p25_prp <- post_prp_data_pval_p25


p25_rates <- roc_rates(contain_outlier = p25_outliers,
                       mamba_ppr_val = p25_mamba_ppr,
                       prp_val = p25_prp,
                       interval = 0.0005,
                       out_thres = 3)
```

```
plot_point(p25_rates$mamba_ppr_rates,
           title = "MAMBA ROC Curve, 0.25 Nonoutlier Rate")
```



```
plot_point(p25_rates$prp_rates,
           title = "PRP ROC Curve, 0.25 Nonoutlier Rate")
```

11

PRP ROC Curve, 0.25 Nonoutlier Rate