SimulationStudySummary

Xiang Ji

9/7/2017

This R markdown file summarizes Simulation Study results.

```
rm(list=ls())  # clean up workspace
setwd("/Users/xji3/GitFolders/YeastIGCTract/SimulationStudy/")

Tract.list <- c(3.0, 10.0, 50.0, 100.0, 200.0, 300.0, 400.0, 500.0)
# First read in HMM results
# from summary file
for(tract in Tract.list){
  hmm.tract.summary <- NULL
  for(sim in 1:100){
    hmm.summary <- paste("./summary/Tract_", toString(tract), '.0/sim_',
                         toString(sim), '/HMM_YDR418W_YEL054C_MG94_nonclock_sim_',
                         toString(sim), '_1D_summary.txt', sep = "")
    if (file.exists(hmm.summary)){
      all <- readLines(hmm.summary, n = -1)
      col.names <- paste("sim_", toString(sim), sep = "")
      row.names <- strsplit(all[length(all)], ' ')[[1]][-1]
      summary_mat <- as.matrix(read.table(hmm.summary,
                                          row.names = row.names,
                                          col.names = col.names))
      hmm.tract.summary <- cbind(hmm.tract.summary, summary_mat)
    }

  }
  assign(paste("HMM_Tract_", toString(tract), "_summary", sep = ""), hmm.tract.summary)
}

# from plots
for(tract in Tract.list){
  hmm.tract.plots <- NULL
  for(sim in 1:100){
    hmm.plot <- paste("./plot/Tract_", toString(tract), '.0/sim_',
                      toString(sim), '/HMM_YDR418W_YEL054C_lnL_sim_',
                      toString(sim), '_1D_surface.txt', sep = "")
    if (file.exists(hmm.plot)){
      lnL.surface <- read.table(hmm.plot)
      max.idx <- which.max(lnL.surface[, 2])
      new.summary <- matrix(c(3.0*exp(-lnL.surface[max.idx, 1]), lnL.surface[max.idx, 2]), 2, 1)
      rownames(new.summary) <- c("tract in nt", "lnL")
      colnames(new.summary) <- paste("sim_", toString(sim), sep = "")
      hmm.tract.plots <- cbind(hmm.tract.plots, new.summary)
    }
  }
  assign(paste("HMM_Tract_", toString(tract), "_plot", sep = ""), hmm.tract.plots)
}
```

```
# Now read in PSJS summary results
for(tract in Tract.list){
  PSJS.tract.summary <- NULL
  for(sim in 1:100){
    PSJS.summary <- paste("./summary/Tract_", toString(tract), '.0/sim_',
                          toString(sim), '/PSJS_HKY_rv_sim_',
                          toString(sim), "_Tract_", toString(tract), '.0_summary.txt', sep = "")
    if (file.exists(PSJS.summary)){
      all <- readLines(PSJS.summary, n = -1)
      col.names <- paste("sim_", toString(sim), sep = "")
      row.names <- strsplit(all[length(all)], ' ')[[1]][-1]
      summary_mat <- as.matrix(read.table(PSJS.summary,
                                          row.names = row.names,
                                          col.names = col.names))
      PSJS.tract.summary <- cbind(PSJS.tract.summary, summary_mat)
    }

  }
  assign(paste("PSJS_Tract_", toString(tract), "_summary", sep = ""), PSJS.tract.summary)
}

# Now read in actual mean tract length in each simulated dataset
for (tract in Tract.list){
  sim.tract <- NULL
  for(sim in 1:100){
    sim_log <- paste("./Tract_", toString(tract), ".0_HKY/sim_", toString(sim),
                     "/YDR418W_YEL054C_sim_", toString(sim), "_IGC.log", sep = "")
    # now read in log file
    log_info <- read.table(sim_log, header = TRUE)
    performed.tract.length <- log_info[, "stop_pos"] - log_info[, "start_pos"] + 1
    proposed.tract.length <- log_info[, "tract_length"]
    new.info <- matrix(c(dim(log_info)[1], mean(proposed.tract.length), sd(proposed.tract.length),
                         mean(performed.tract.length), sd(performed.tract.length)), 5, 1)
    rownames(new.info) <- c("num IGC", "mean proposed tract length", "sd proposed tract length",
                            "mean performed tract length", "sd performed tract length")
    colnames(new.info) <- paste("sim_", toString(sim), sep = "")
    sim.tract <- cbind(sim.tract, new.info)
  }
  assign(paste("sim.tract.", toString(tract), sep = ""), sim.tract)
}

# read in simulated tract info of simulated datasets with shrinked tau
for (tract in Tract.list){
  HKY.shrinked.tract <- NULL
  for(sim in 1:100){
    sim_log <- paste("./Shrinked_Tau_HKY_Simulation/Tract_", toString(tract), ".0_HKY/sim_", toString(si
                     "/YDR418W_YEL054C_sim_", toString(sim), "_IGC.log", sep = "")
    # now read in log file
    log_info <- read.table(sim_log, header = TRUE)
    proposed.tract.length <- log_info[, "tract_length"]
    performed.tract.length <- log_info[, "stop_pos"] - log_info[, "start_pos"] + 1
    new.info <- matrix(c(mean(proposed.tract.length), sd(proposed.tract.length),
                         mean(performed.tract.length), sd(performed.tract.length)), 4, 1)
```

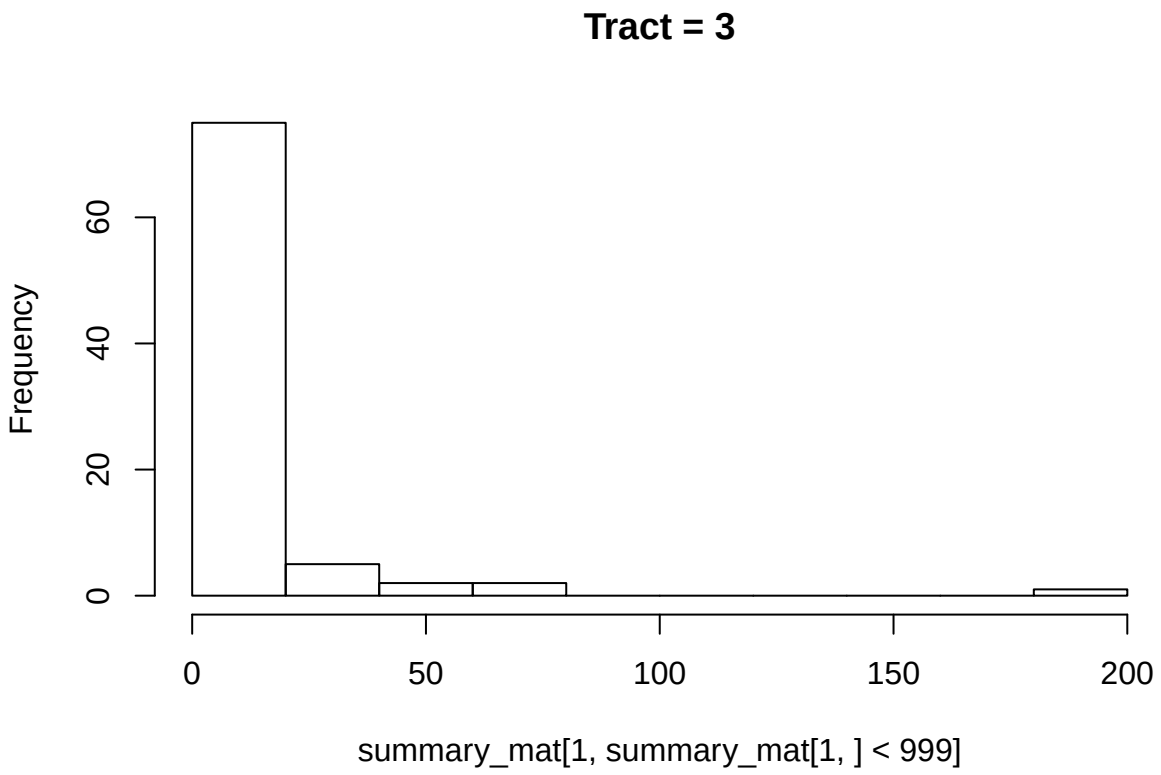
```
    rownames(new.info) <- c("mean proposed tract length", "sd proposed tract length",
                            "mean performed tract length", "sd performed tract length")
    colnames(new.info) <- paste("sim_", toString(sim), sep = "")
    HKY.shrinked.tract <- cbind(HKY.shrinked.tract, new.info)
  }
  assign(paste("HKY.shrinked.tract.", toString(tract), sep = ""), HKY.shrinked.tract)
}

guess.list <- c(50.0, 100.0, 250.0, 500.0)
# Now read in PSJS summary results of shrinked tau datasets
for(tract in Tract.list){
  for(guess in guess.list){
    PSJS.tract.summary <- NULL
    for(sim in 1:100){

      PSJS.summary <- paste("./summary/Tract_", toString(tract), '.0_HKY/sim_',
                            toString(sim), '/PSJS_HKY_rv_sim_',
                            toString(sim), "_Tract_", toString(tract), '.0_guess_',
                            toString(guess),'.0_nt_summary.txt', sep = "")
      if (file.exists(PSJS.summary)){
        all <- readLines(PSJS.summary, n = -1)
        col.names <- paste("sim_", toString(sim), sep = "")
        row.names <- strsplit(all[length(all)], ' ')[[1]][-1]
        summary_mat <- as.matrix(read.table(PSJS.summary,
                                            row.names = row.names,
                                            col.names = col.names))
        PSJS.tract.summary <- cbind(PSJS.tract.summary, summary_mat)
      }
    }
    assign(paste("PSJS_HKY_Tract_", toString(tract), "_guess_",
                 toString(guess), "_summary", sep = ""), PSJS.tract.summary)
  }
}

# Now combine all initial guess results
for(tract in Tract.list){
  combined.PSJS.tract.summary <- NULL
  col.list <- NULL
  for ( sim_num in 1:100){
    sim_col <- paste("sim_", toString(sim_num), sep = "")
    best.lnL <- -Inf
    best.guess <- NULL
    for(guess in guess.list){
      target_summary <- get(paste("PSJS_HKY_Tract_", toString(tract), "_guess_", toString(guess), "_sum
      if(sim_col %in% colnames(target_summary) ){
        if (target_summary["ll", sim_col] > best.lnL){
          best.lnL <- target_summary["ll", sim_col]
          best.guess <- guess
        }
      }
    }
    if(! is.null(best.guess)){
      combined.PSJS.tract.summary <- cbind(combined.PSJS.tract.summary,
```

```
                                        get(paste("PSJS_HKY_Tract_", toString(tract), "_guess_", toS
      col.list <- c(col.list, sim_col)
    }

  }
  colnames(combined.PSJS.tract.summary) <- col.list
  assign(paste("PSJS_HKY_Tract_", toString(tract), "_combined_summary", sep = ""), combined.PSJS.tract.s
}
```

OK, Now show the performance summary

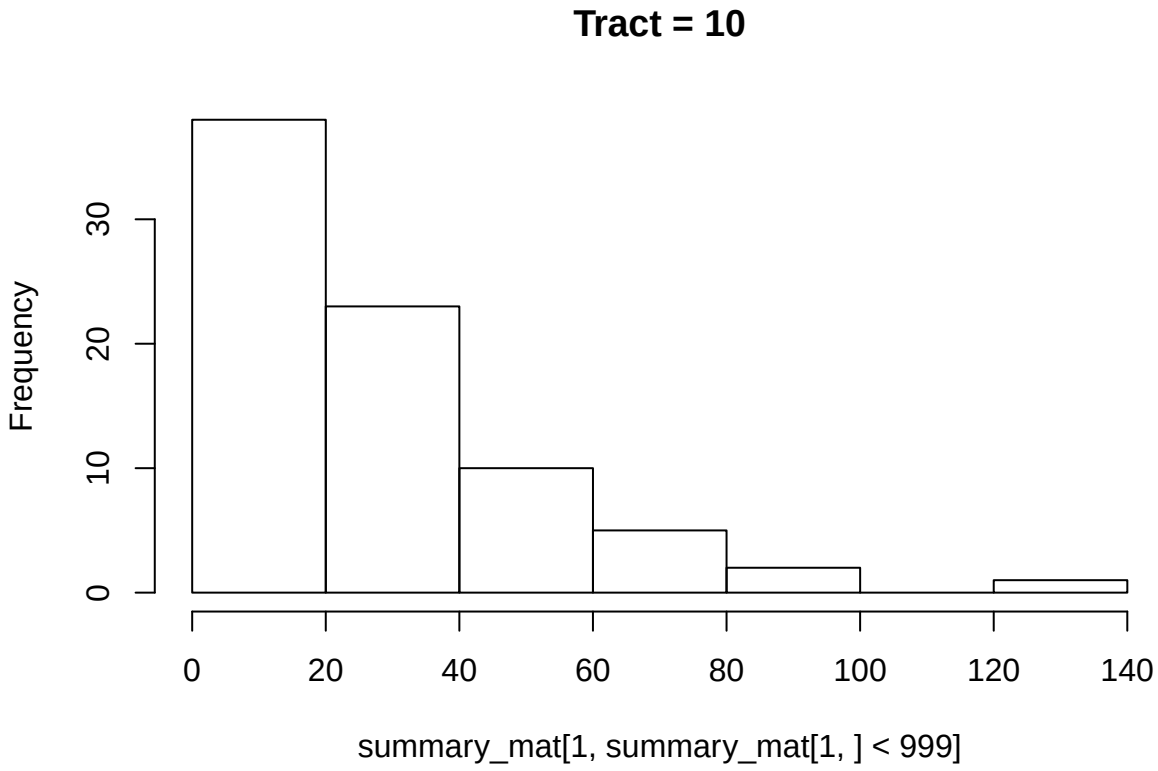
```
# HMM results
for (tract in Tract.list){
  # show how many stuck at boundary 1000 nt first
  print(paste("Tract = ", toString(tract), sep = ""))
  summary_mat <- get(paste("HMM_Tract_", toString(tract), "_plot", sep = ""))

  # histogram of inferred tract length
  hist(summary_mat[1, summary_mat[1, ] < 999.], main = paste("Tract = ", toString(tract), sep = ""))
  print(paste("Among total 100 simulated data sets, ", toString(sum(summary_mat[1, ] > 999)),
              " datasets stuck at 1000", sep = ""))
  print(matrix(c("mean", mean(summary_mat[1, summary_mat[1, ] < 999.]),
          "sd", sd(summary_mat[1, summary_mat[1, ] < 999.])), 2, 2))
}
```

```
## [1] "Tract = 3"
```

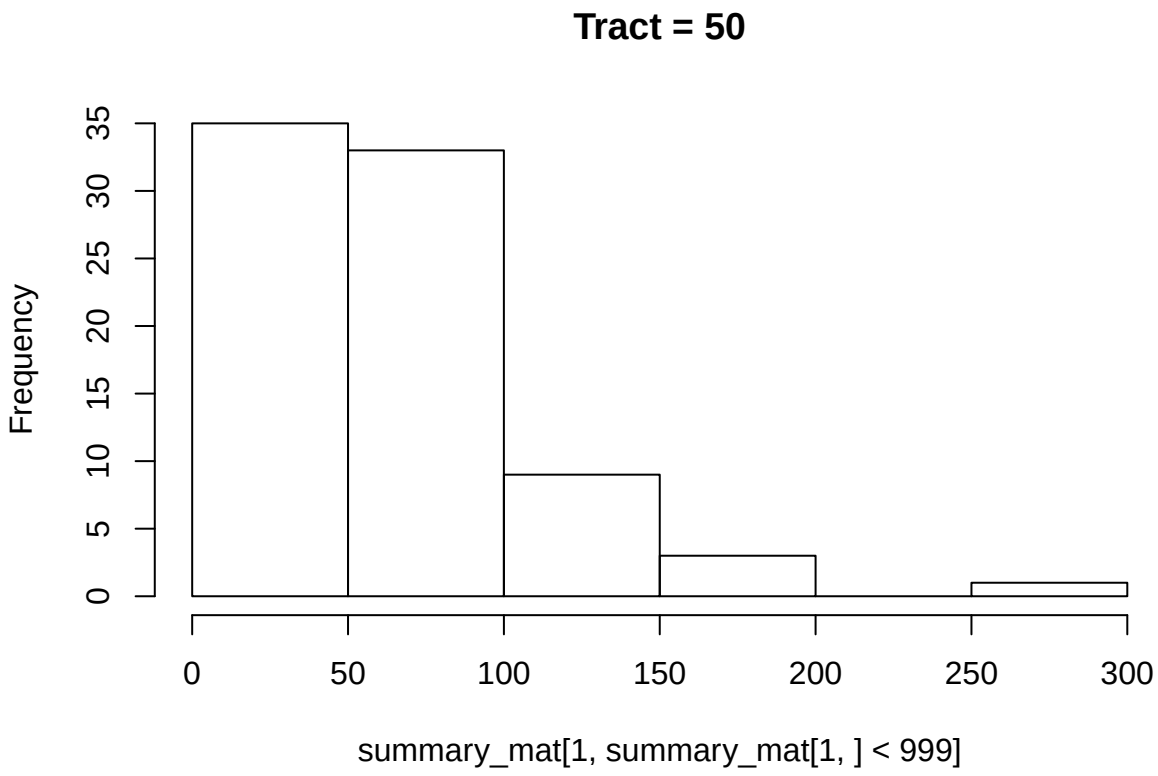



```
## [1] "Among total 100 simulated data sets, 15 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "11.6941176470561" "23.409920327626"
```

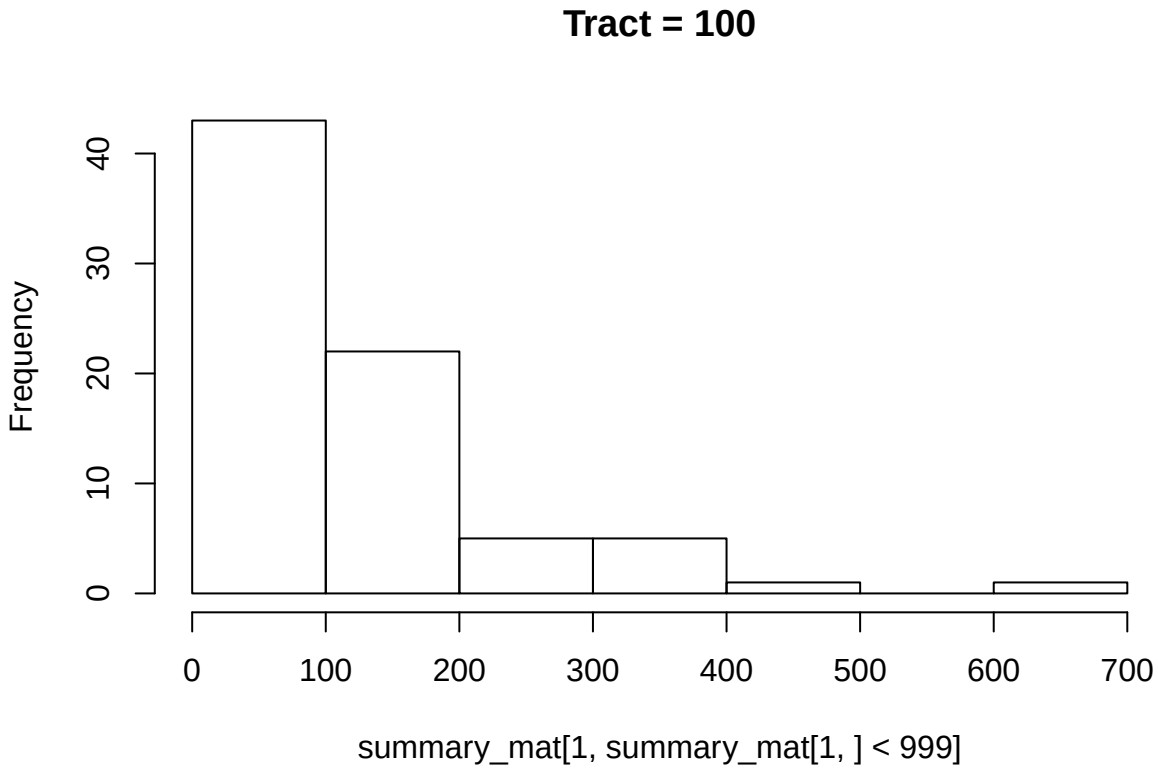
```
## [1] "Tract = 10"
```



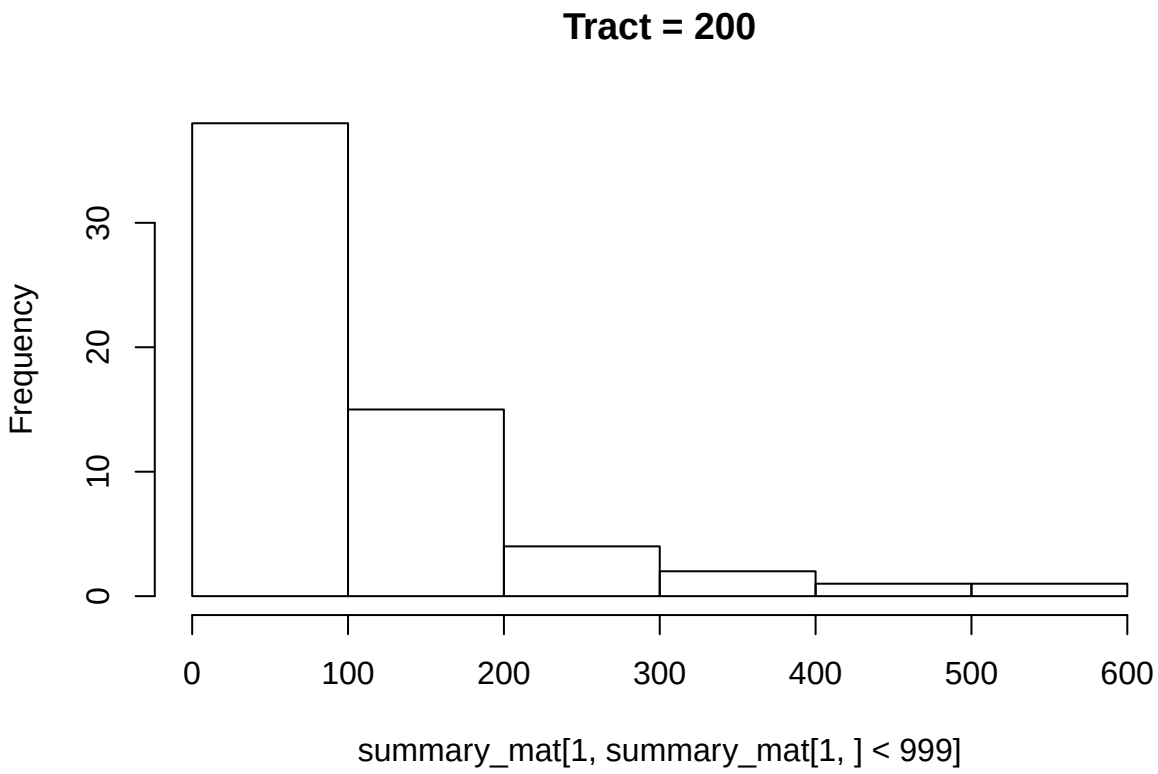

```
## [1] "Among total 100 simulated data sets, 21 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "28.5316455696352" "24.4245753935622"
## [1] "Tract = 50"
```



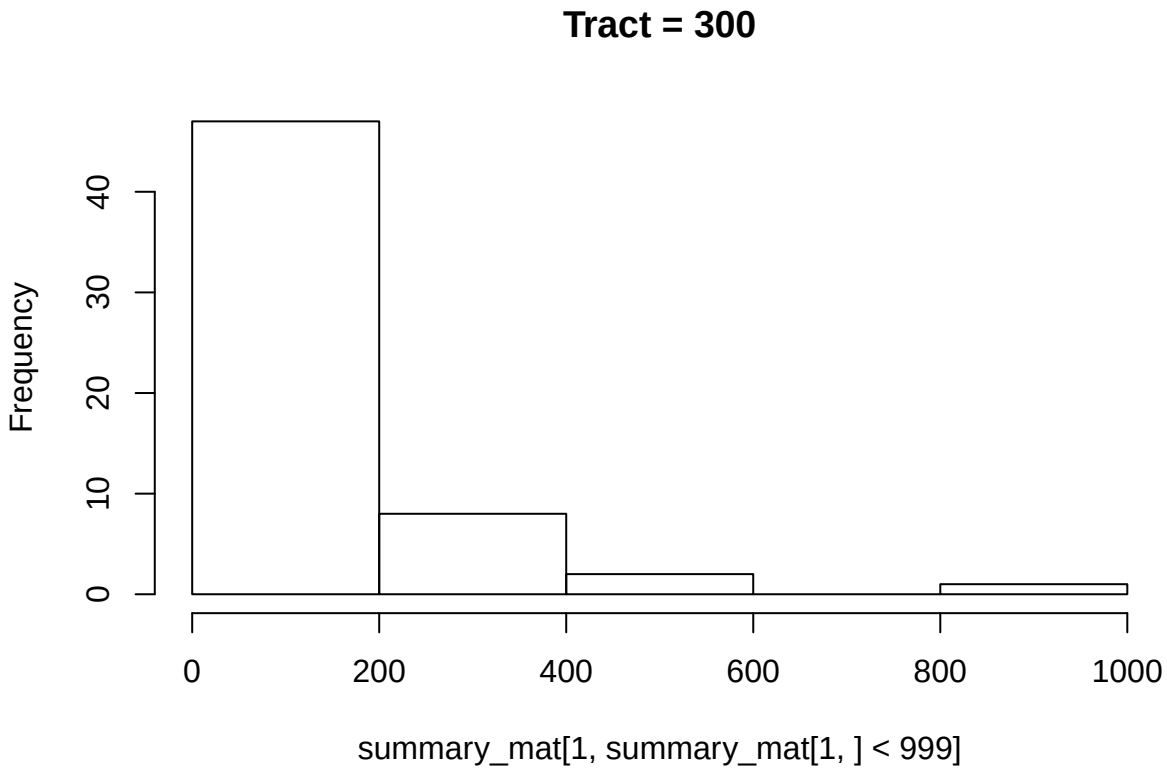

```
## [1] "Among total 100 simulated data sets, 19 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "67.0864197531772" "50.0167965615616"
## [1] "Tract = 100"
```



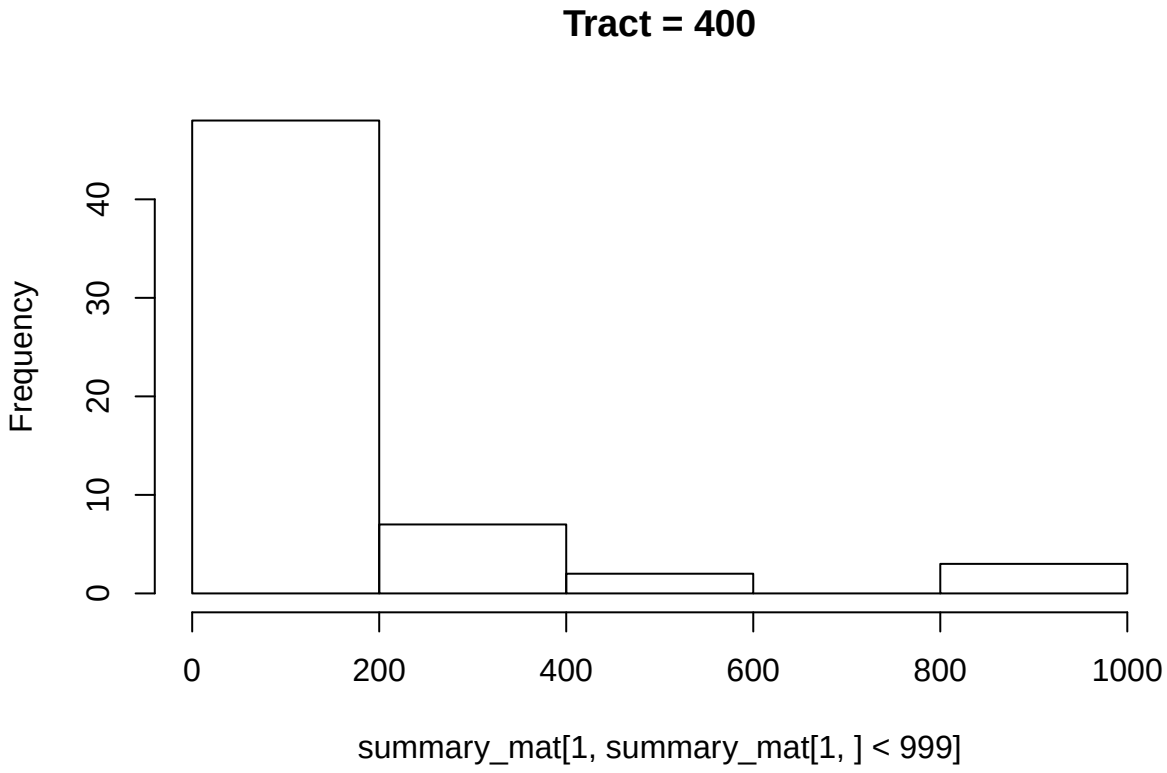
```
## [1] "Among total 100 simulated data sets, 23 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "121.597402597374" "111.189768998766"
## [1] "Tract = 200"
```



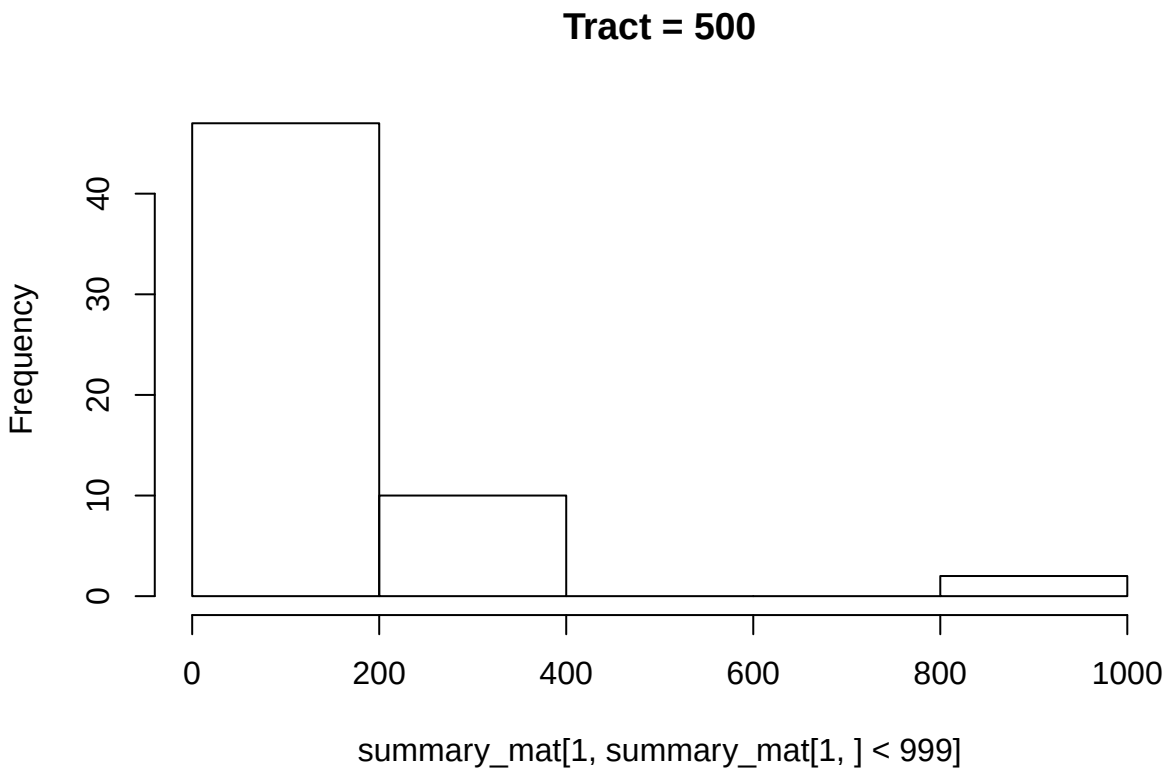

```
## [1] "Among total 100 simulated data sets, 38 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "111.590163934384" "109.950955285812"
## [1] "Tract = 300"
```




```
## [1] "Among total 100 simulated data sets, 42 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "124.293103448246" "157.010001761835"
## [1] "Tract = 400"
```

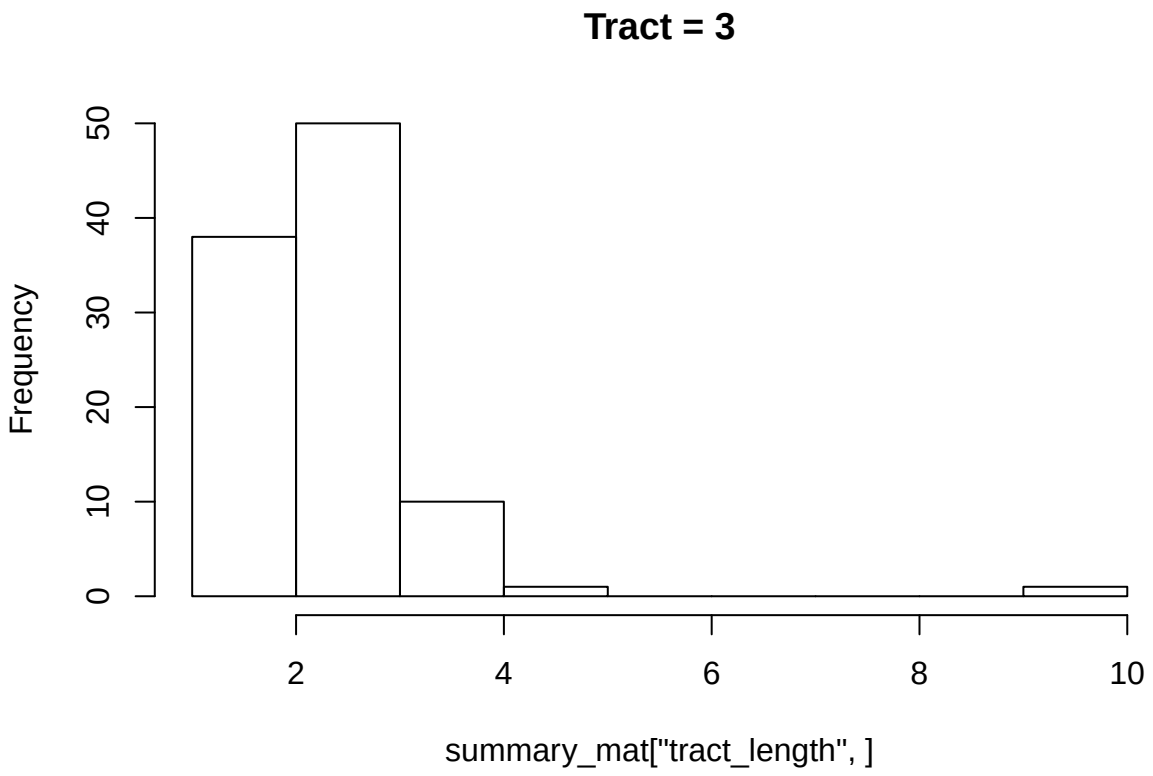


```
## [1] "Among total 100 simulated data sets, 40 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "153.333333333315" "219.73201449035"
## [1] "Tract = 500"
```

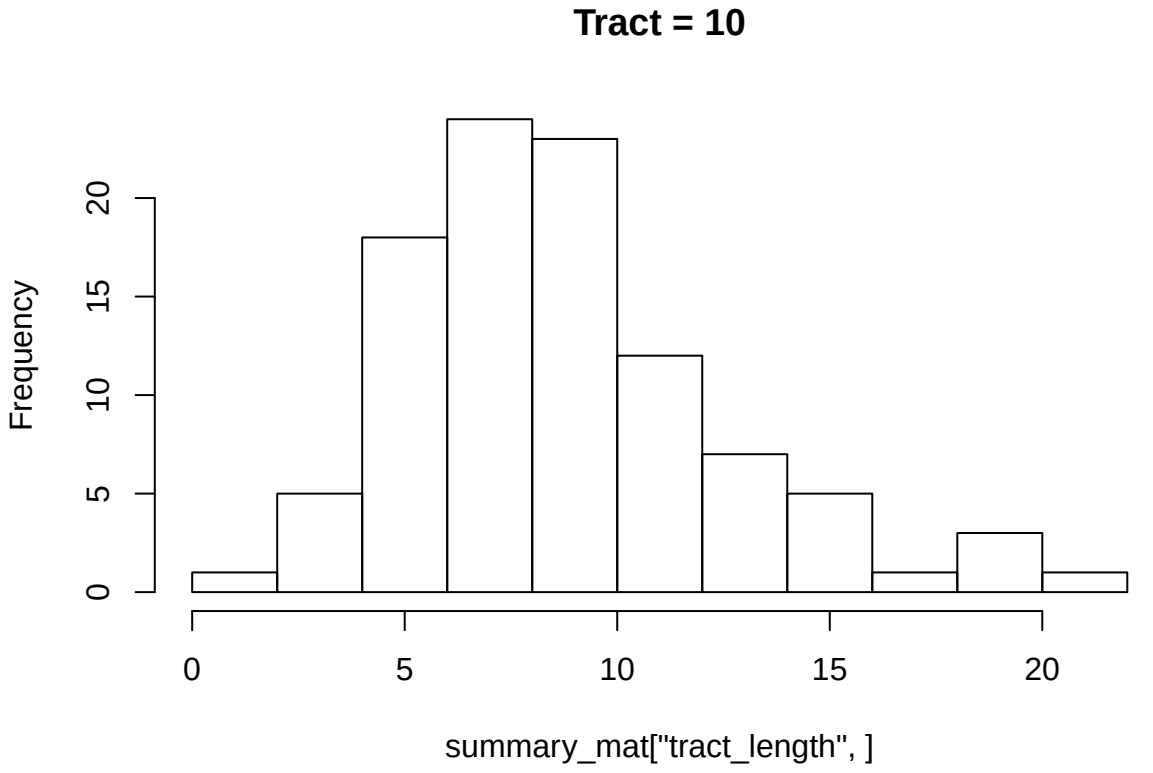



```
## [1] "Among total 100 simulated data sets, 41 datasets stuck at 1000"
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "136.30508474575" "187.058284687408"
```

```r
# PSJS results
for(tract in Tract.list){
  summary_mat <- get(paste("PSJS_Tract_", toString(tract), "_summary", sep = ""))
  # histogram of inferred tract length
  hist(summary_mat["tract_length", ], main = paste("Tract = ", toString(tract), sep = ""))
  print(matrix(c("mean", mean(summary_mat["tract_length", ]),
          "sd", sd(summary_mat["tract_length", ])), 2, 2))
}
```

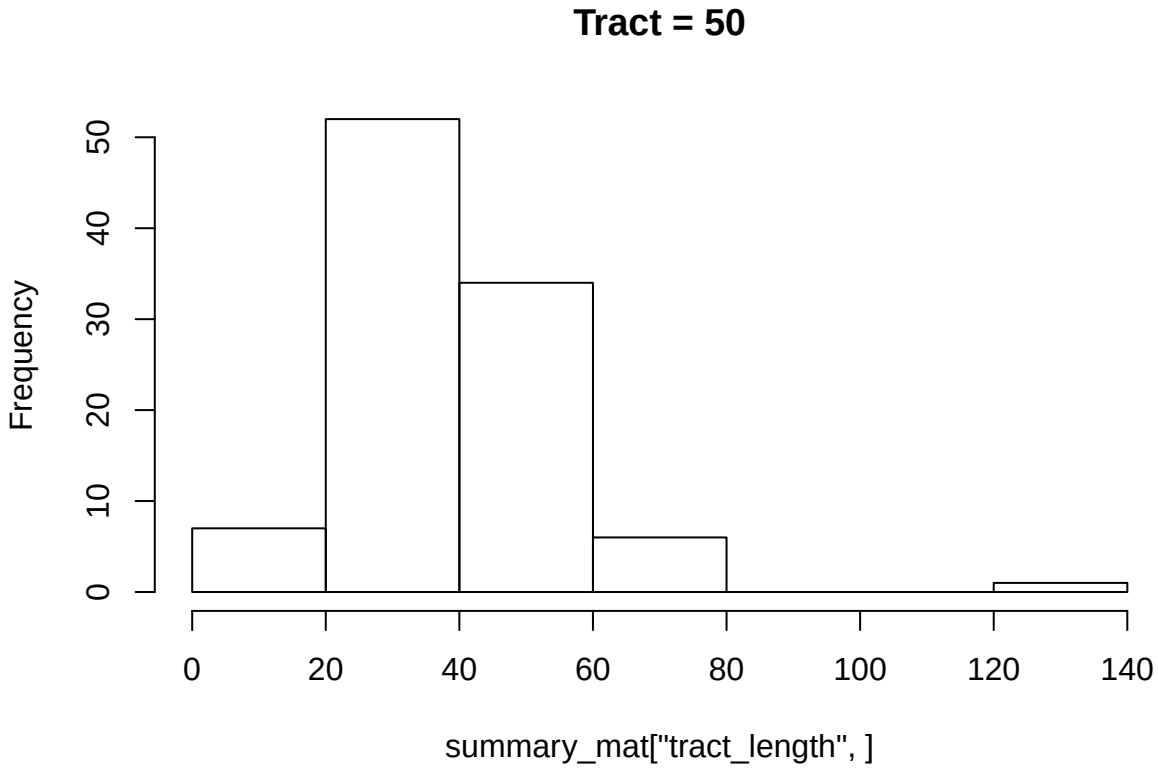



```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "2.29718013607812" "1.02874535788003"
```

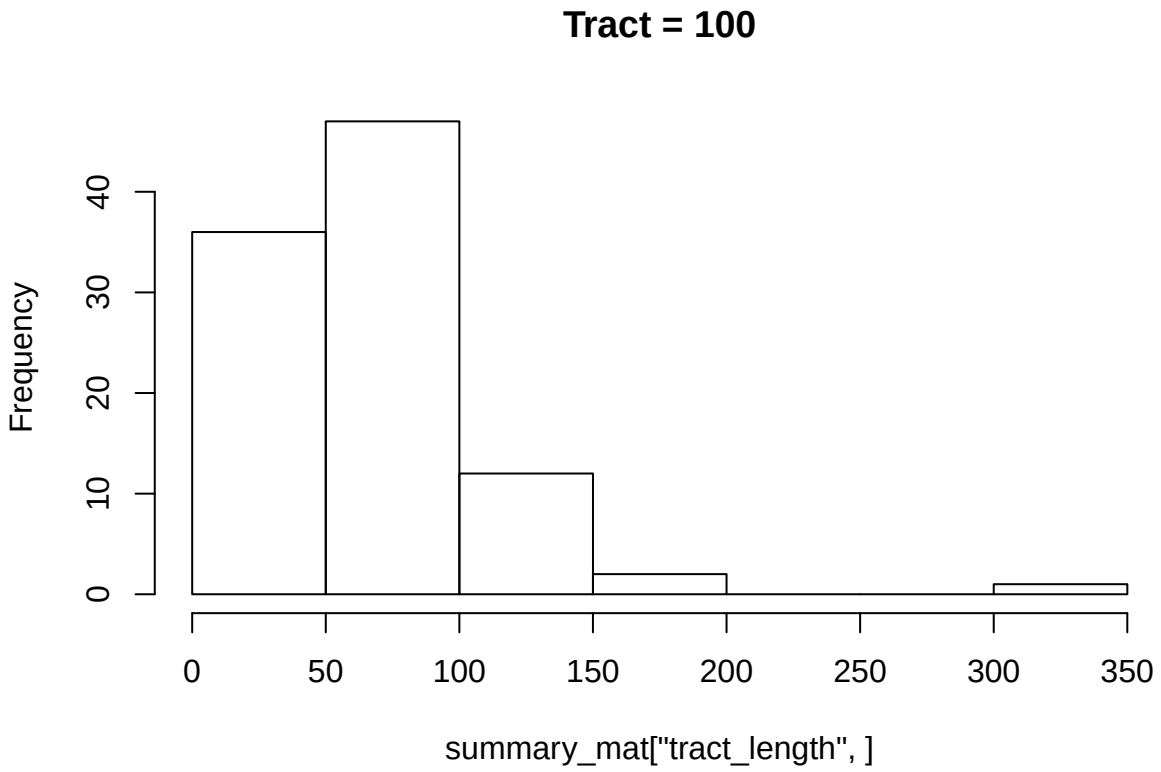



```
##      [,1]               [,2]
```

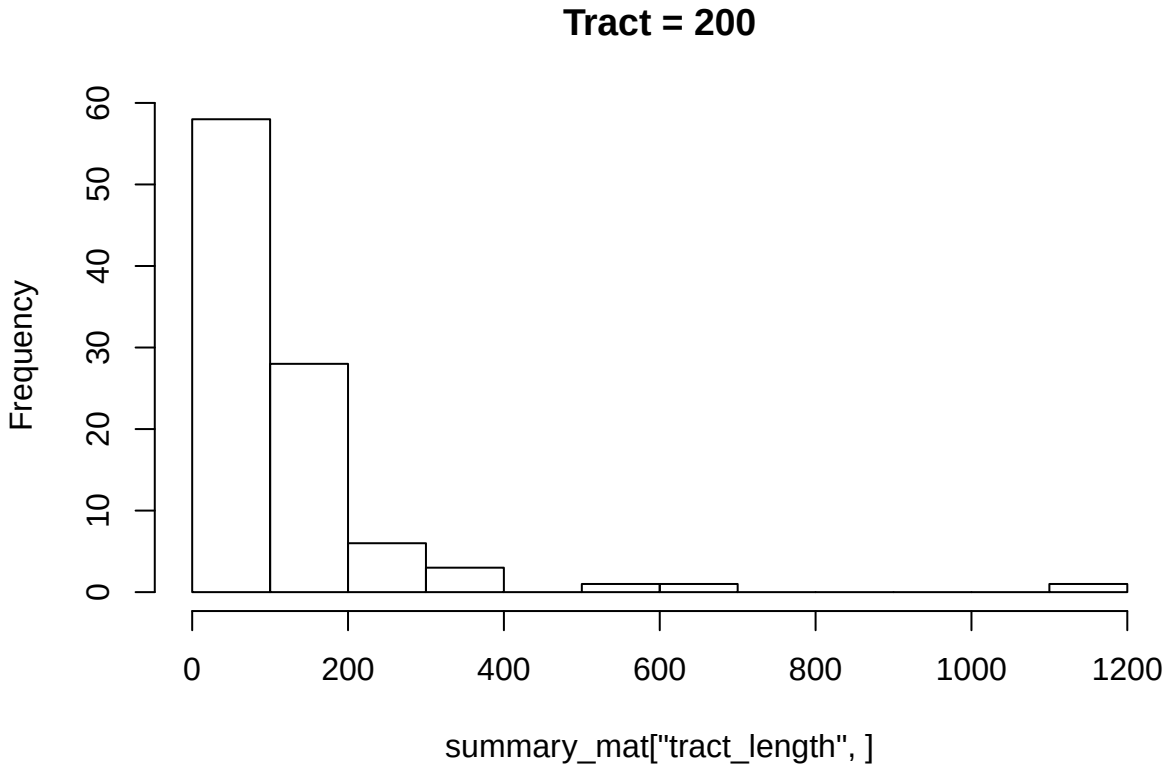
```
## [1,] "mean"             "sd"
## [2,] "8.72620746076138" "3.72714198660433"
```




```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "38.1001820929933" "16.207953647501"
```

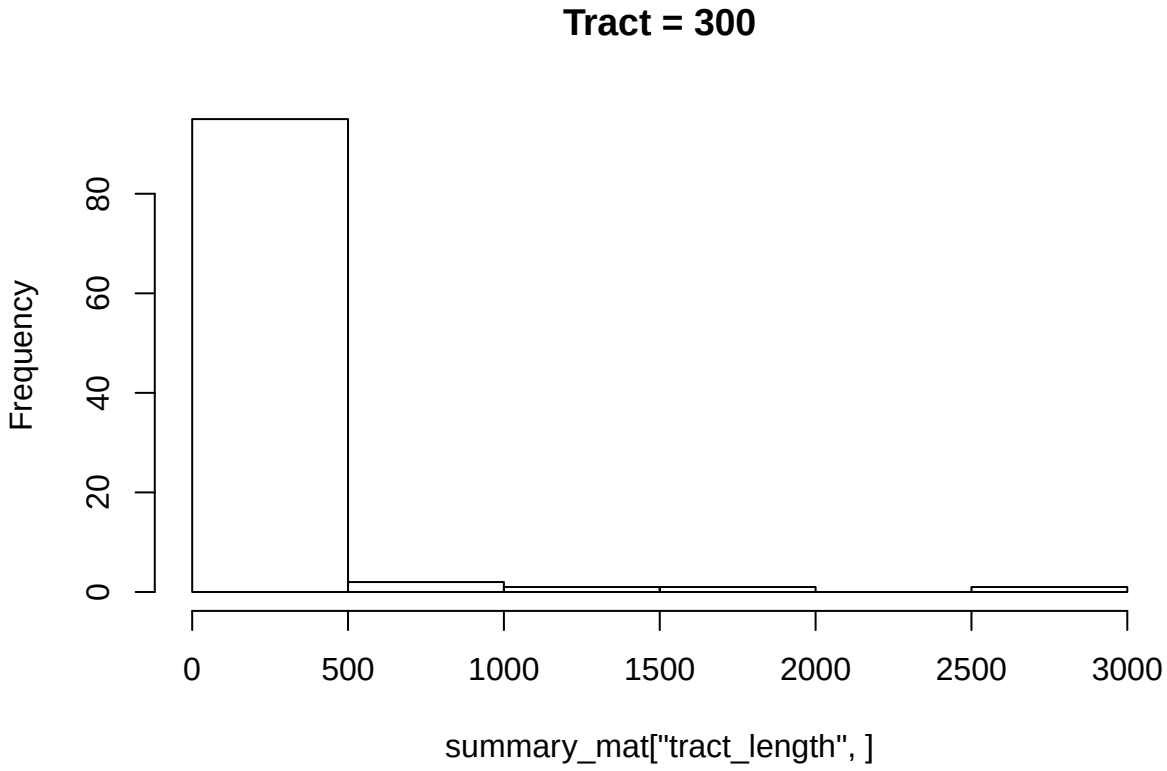



```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "67.3065398387664" "40.4358020083188"
```

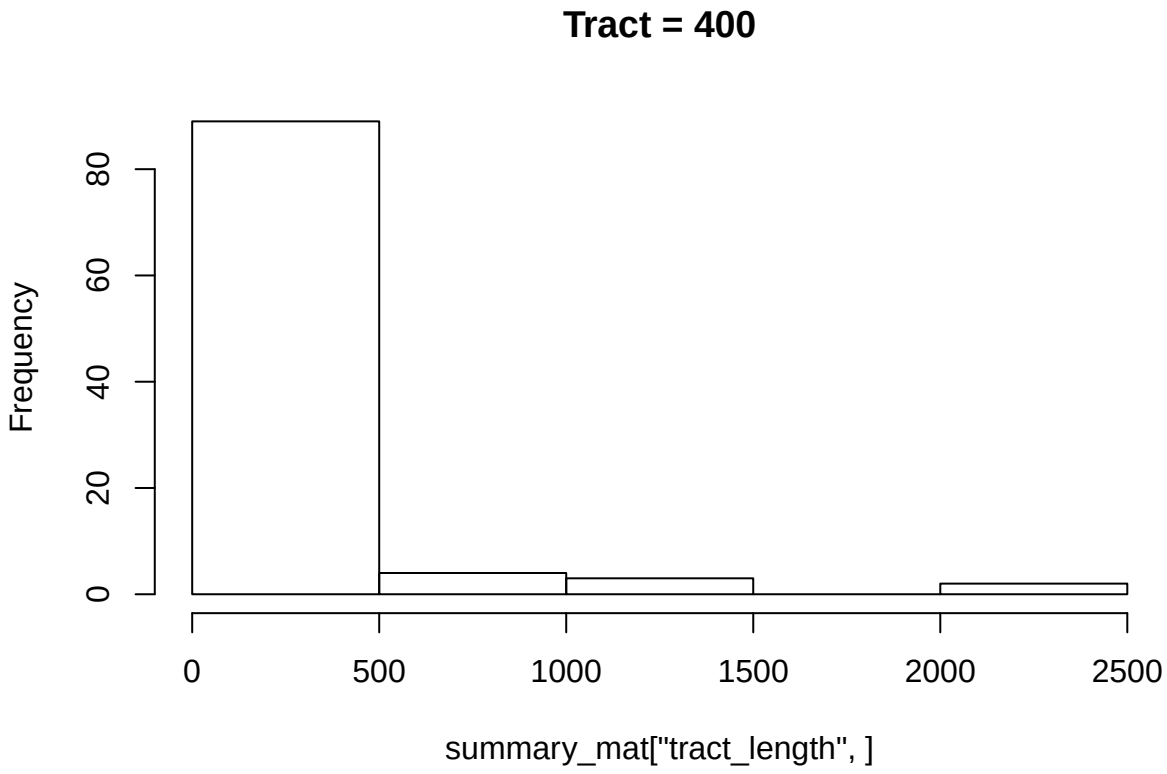



```
##      [,1]               [,2]
```

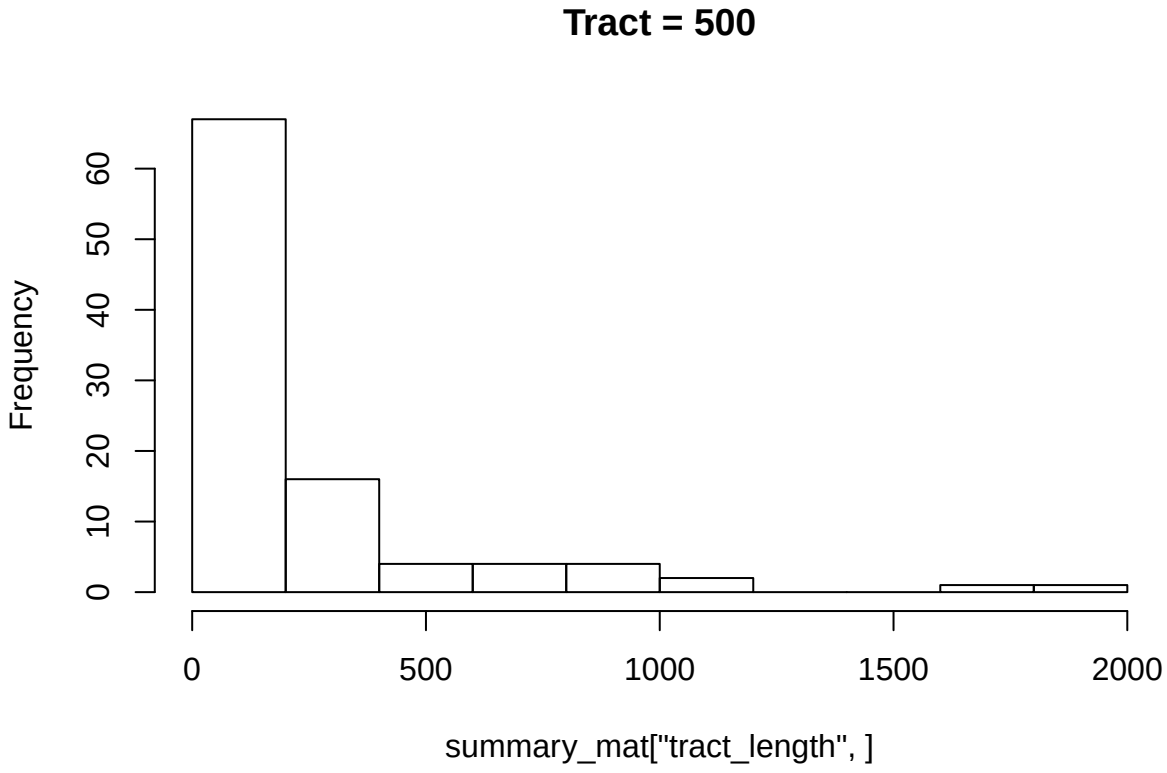
```
## [1,] "mean"             "sd"
## [2,] "123.447492290421" "145.046663199282"
```




```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "189.995520518941" "356.357994056025"
```



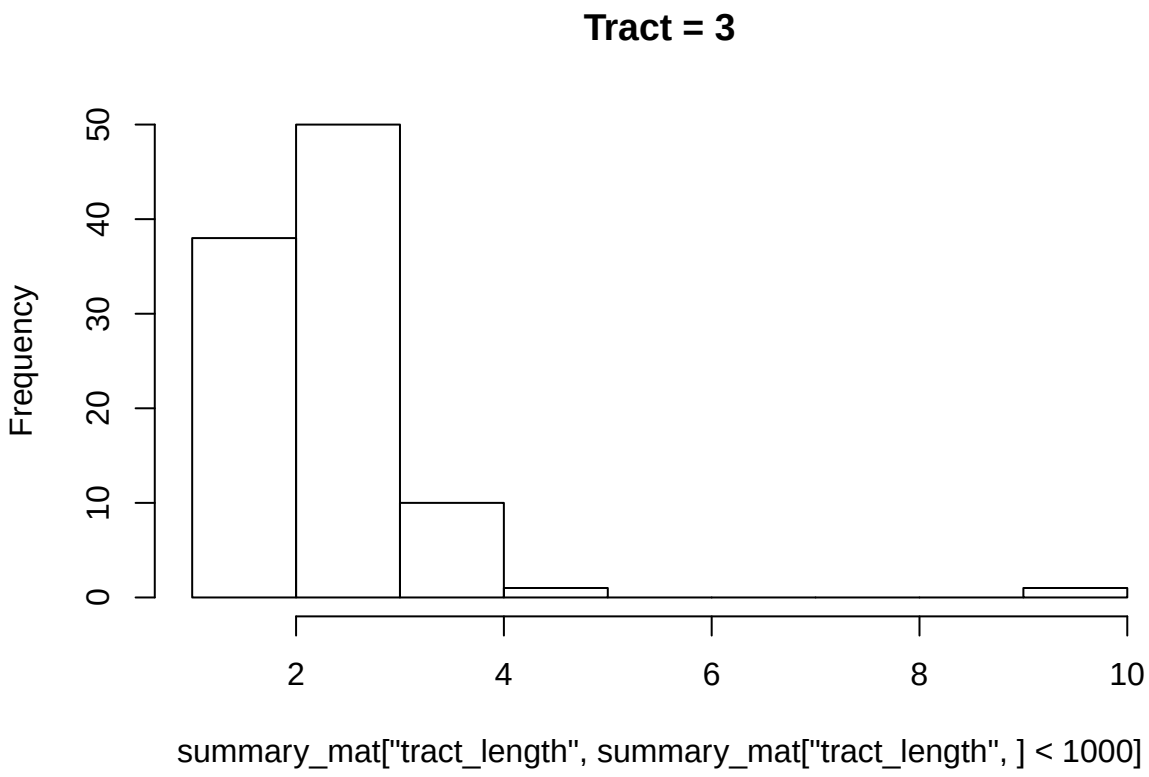

```
##      [,1]              [,2]
## [1,] "mean"            "sd"
## [2,] "217.14487788415" "377.699962735285"
```



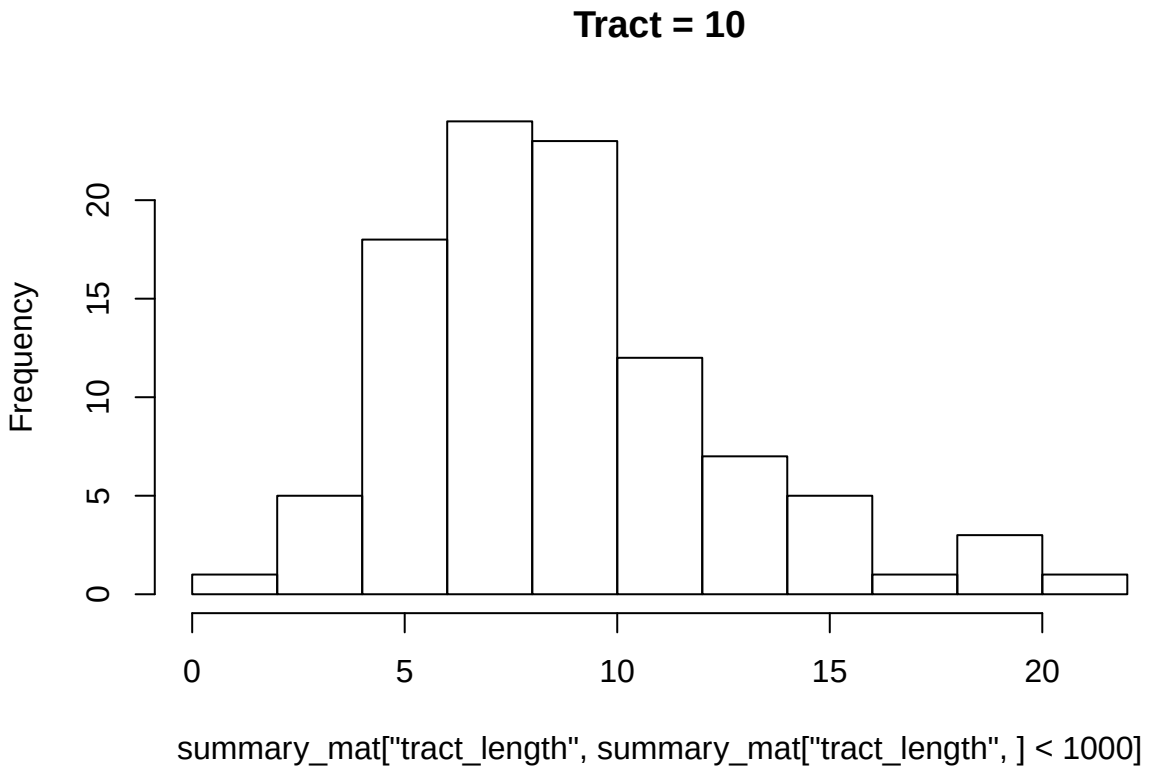

```
##      [,1]              [,2]
```

```
## [1,] "mean"             "sd"
## [2,] "239.190848695252" "346.723408519857"
```

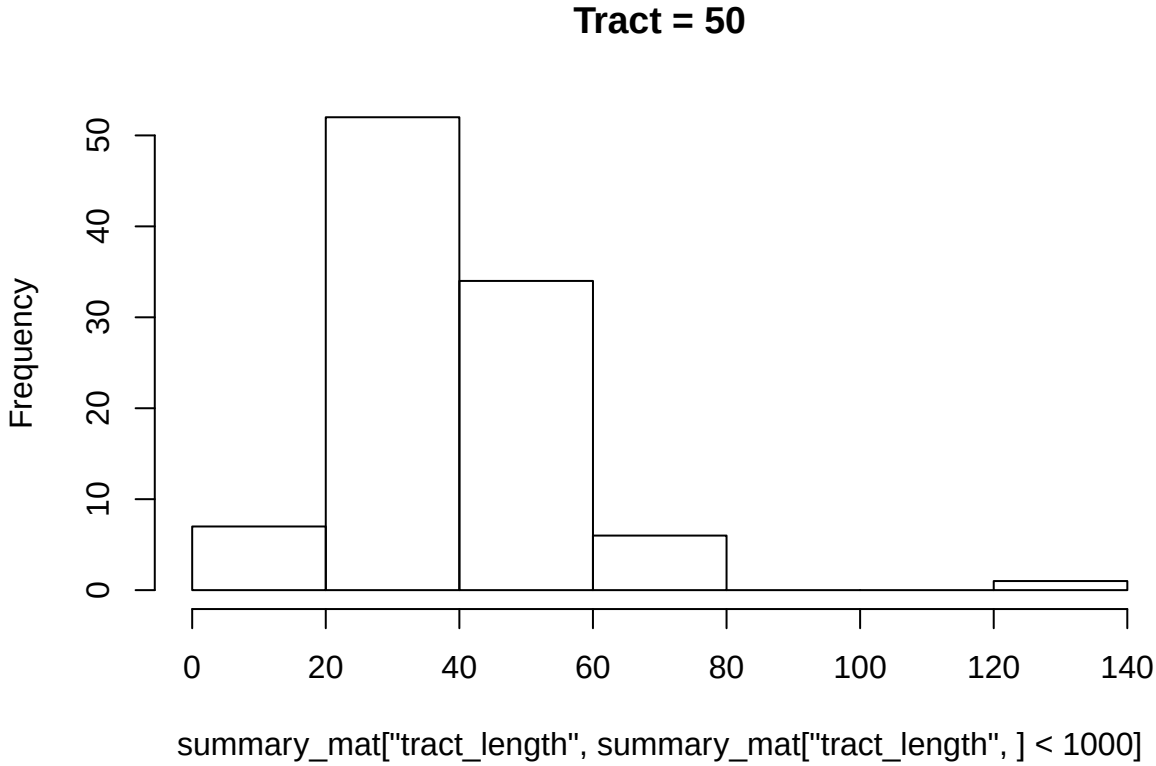
```
# exclude suspiciously long inferred tract length, plot again
for(tract in Tract.list){
  summary_mat <- get(paste("PSJS_Tract_", toString(tract), "_summary", sep = ""))
  # histogram of inferred tract length
  hist(summary_mat["tract_length", summary_mat["tract_length", ] < 1000.0], main = paste("Tract = ", to
  print(matrix(c("mean", mean(summary_mat["tract_length", summary_mat["tract_length", ] < 1000.0]),
         "sd", sd(summary_mat["tract_length", summary_mat["tract_length", ] < 1000.0])), 2, 2))
}
```




```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "2.29718013607812" "1.02874535788003"
```

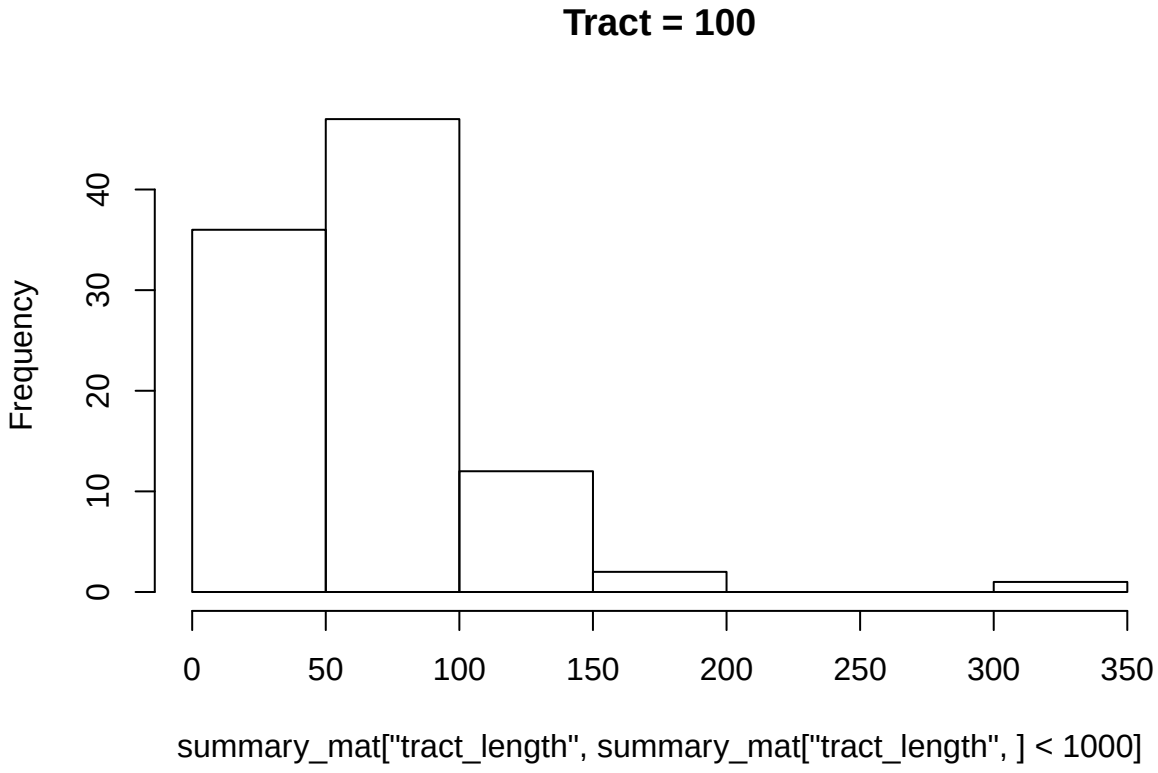



```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "8.72620746076138" "3.72714198660433"
```

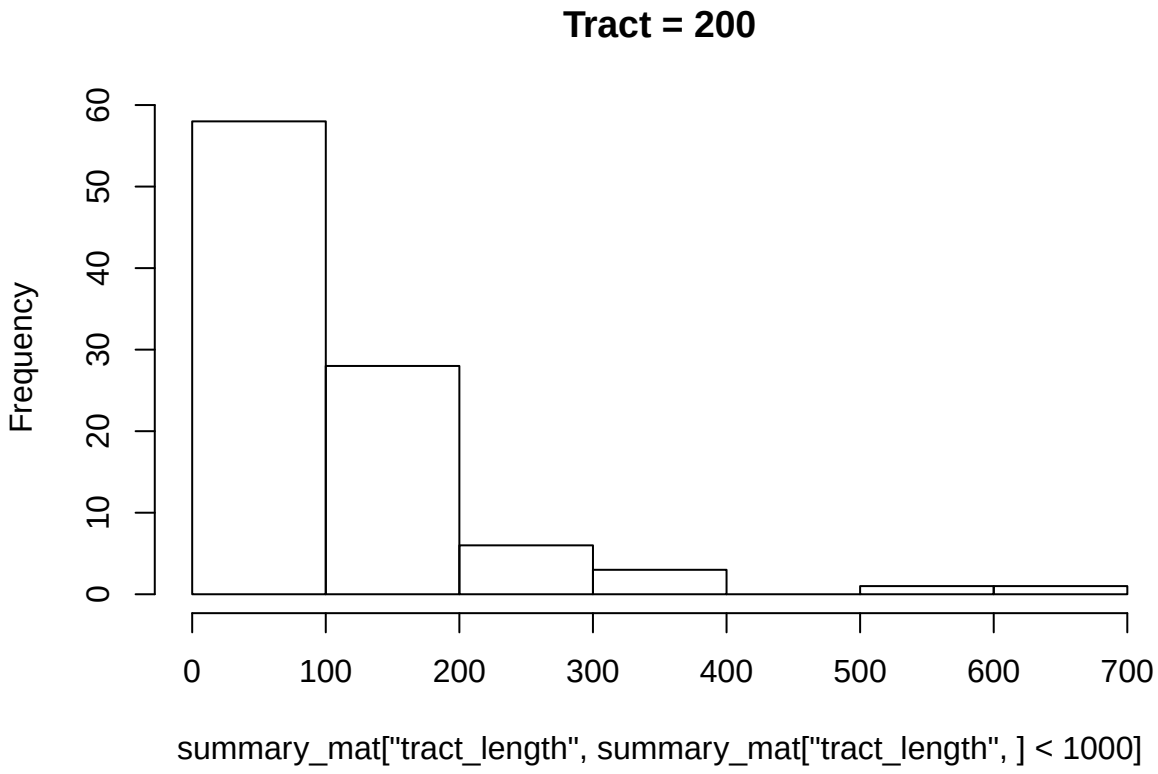



```
##      [,1]               [,2]
```

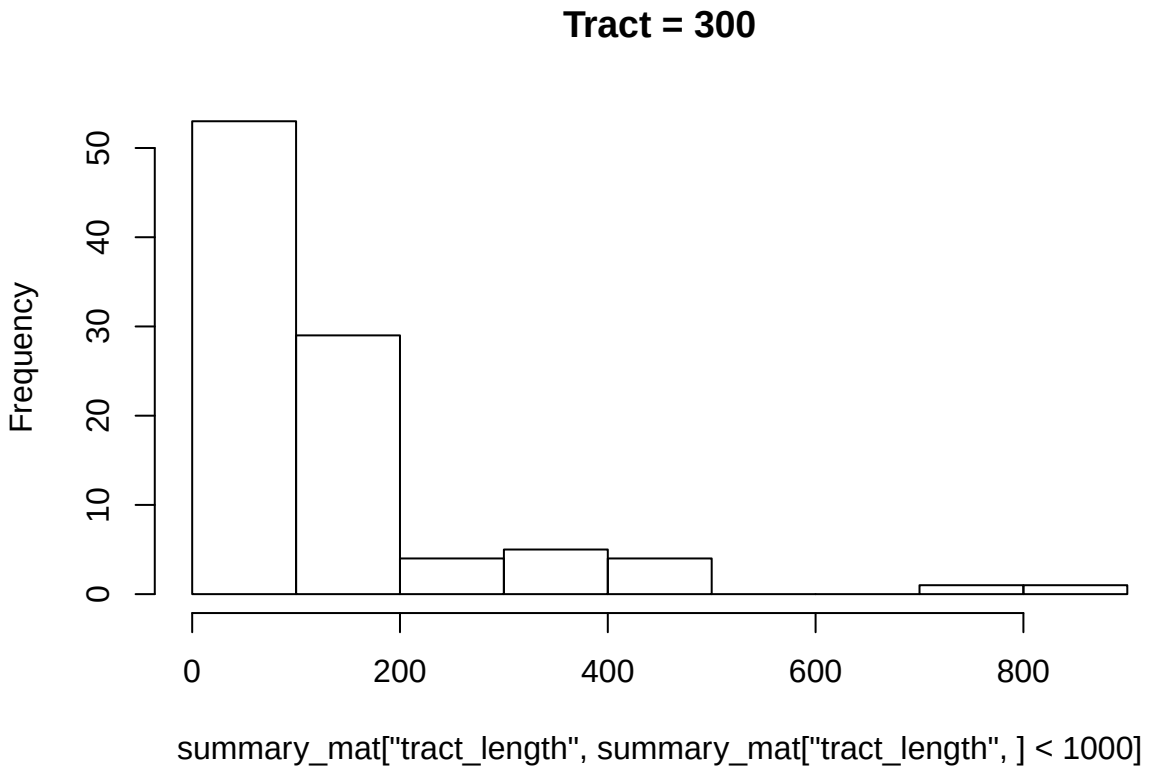
```
## [1,] "mean"             "sd"
## [2,] "38.1001820929933" "16.207953647501"
```




```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "67.3065398387664" "40.4358020083188"
```

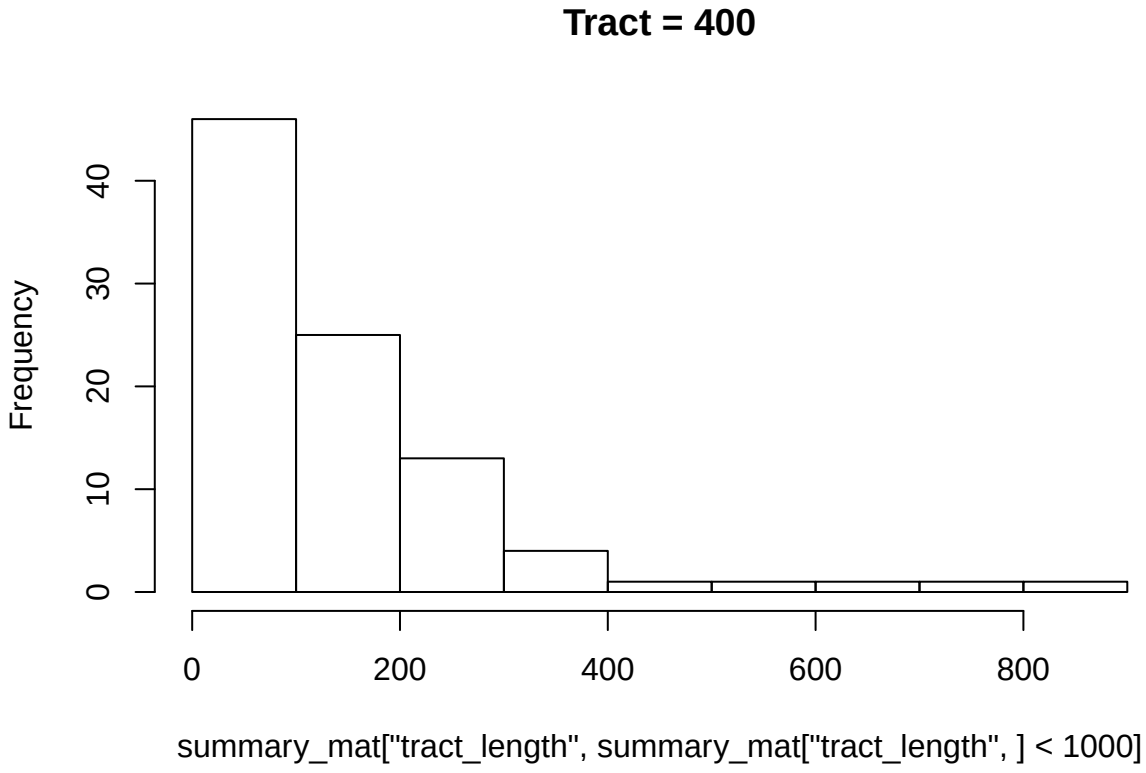



```
##      [,1]                [,2]
## [1,] "mean"              "sd"
## [2,] "113.216002236222" "104.364024061081"
```

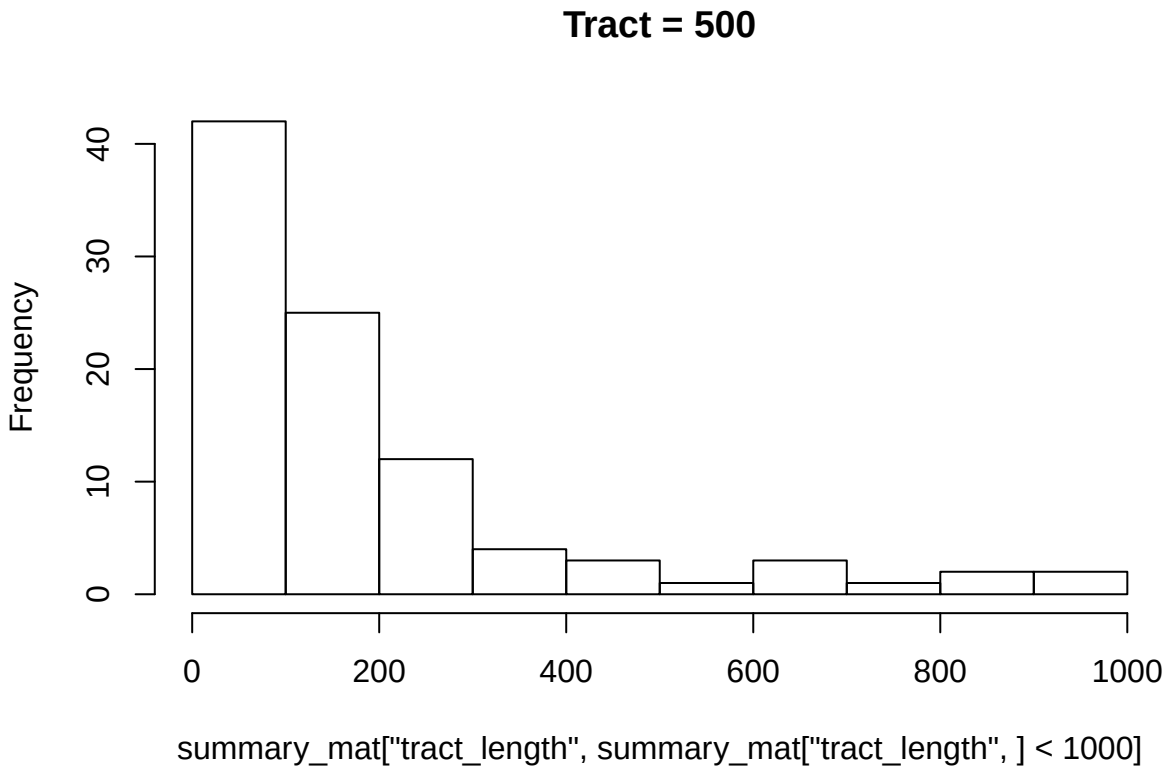



```
##      [,1]                [,2]
```

```
## [1,] "mean"             "sd"
## [2,] "135.843591947529" "146.132608739573"
```




```
##      [,1]               [,2]
## [1,] "mean"             "sd"
## [2,] "142.57022420328" "154.570518796298"
```

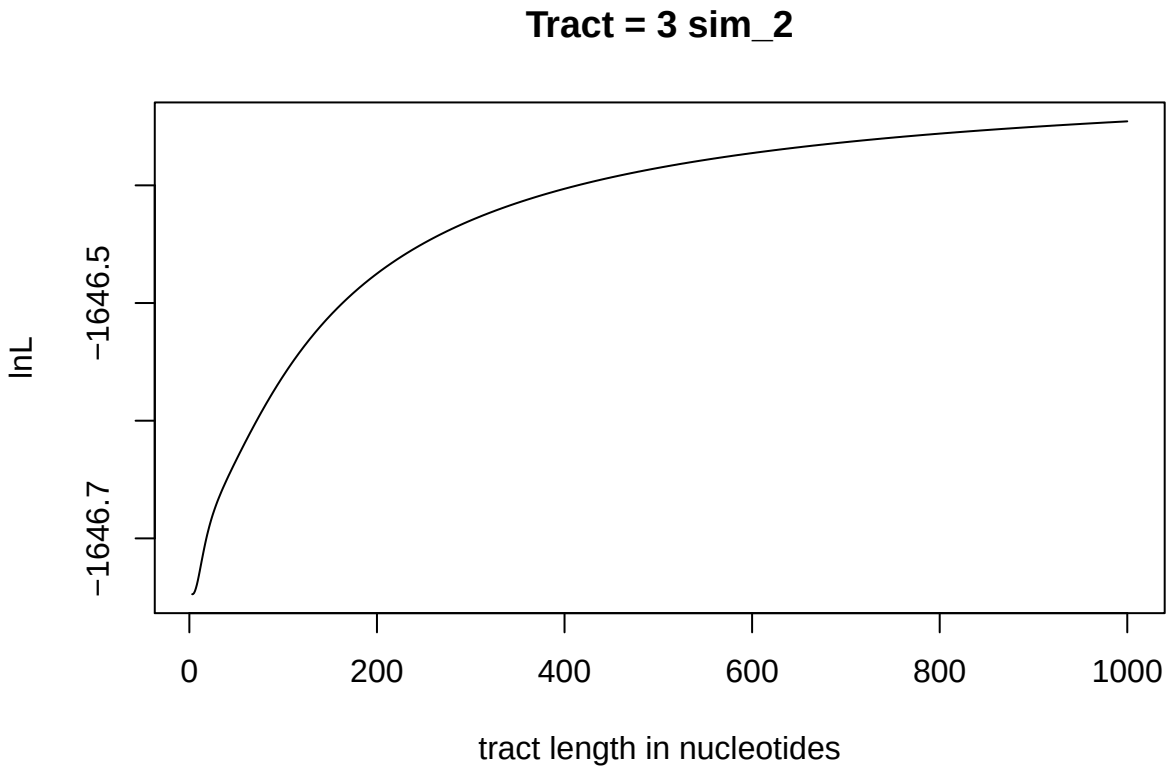


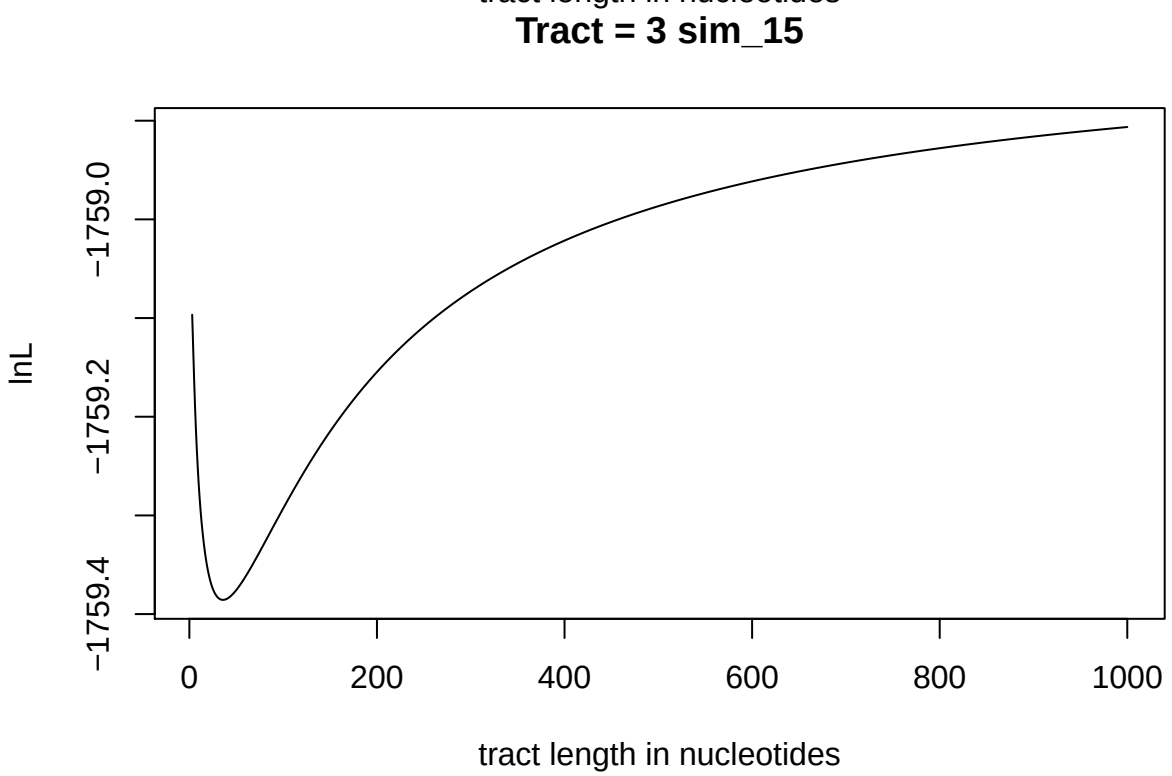

```
##      [,1]                [,2]
## [1,] "mean"              "sd"
## [2,] "186.868349944308" "222.75685619742"
```

A plot of HMM surface that infer tract length at boundary from each tract length condition

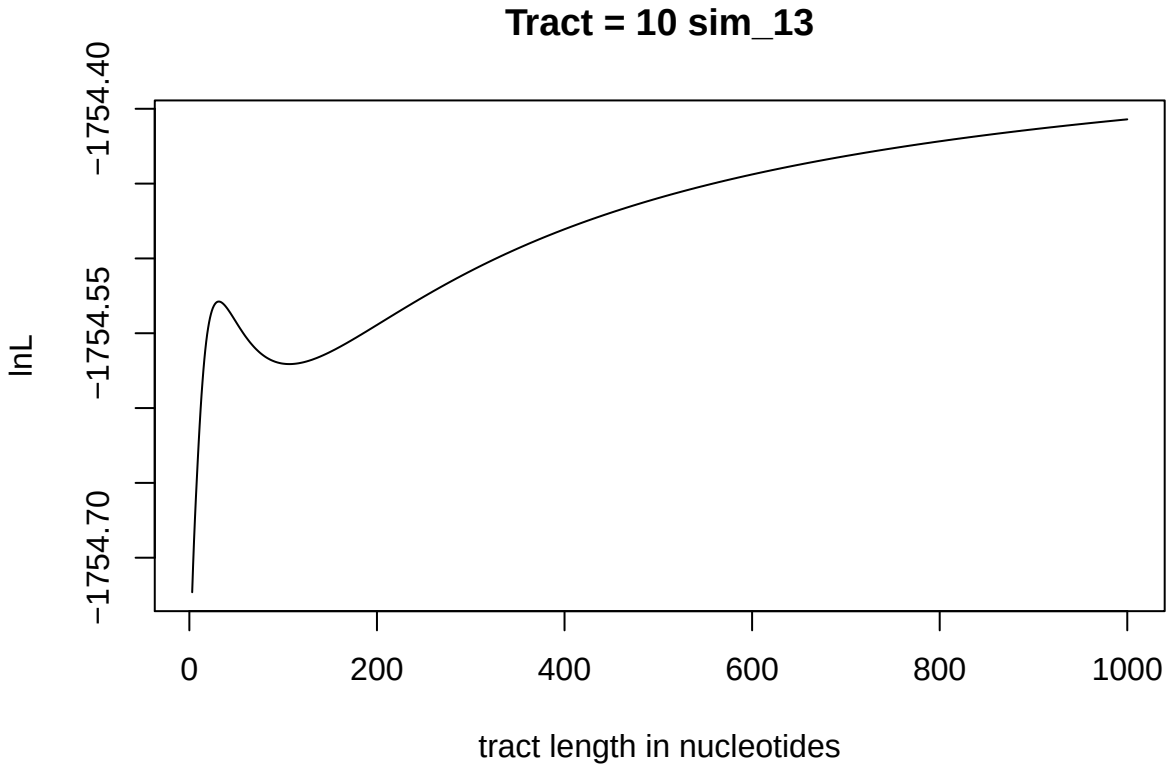
```
# plot the first two datasets that HMM inferred tract length stuck at boundary of 1000.0
for (tract in Tract.list){
  print(paste("Tract = ", toString(tract), sep = ""))
  summary_mat <- get(paste("HMM_Tract_", toString(tract), "_plot", sep = ""))
  count = 0
  for(iter in 1:99){
    if(summary_mat[1, iter] > 999. & count < 2){
      to.plot <- read.table(paste("./plot/Tract_", toString(tract), ".0/", colnames(summary_mat)[iter],
                                  "/HMM_YDR418W_YEL054C_lnL_", colnames(summary_mat)[iter],
                                  "_1D_surface.txt", sep = ""))
      plot(3.0*exp(-to.plot[, 1]), to.plot[, 2], main = paste("Tract = ", toString(tract), " ",colnames
           ylab = "lnL")
      count = count + 1
    }
  }
}
```

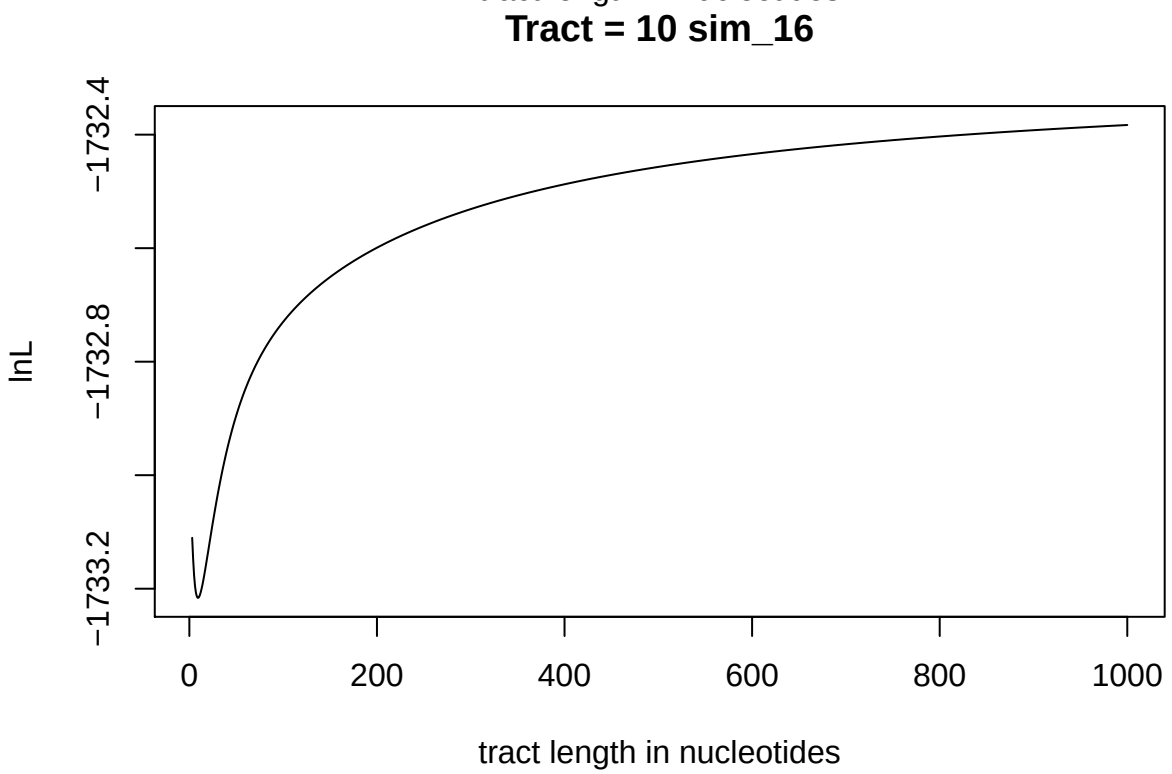
```
## [1] "Tract = 3"
```


Tract = 3 sim_2
lnL
−1646.5
−1646.7
0
200
400
600
800
1000
tract length in nucleotides

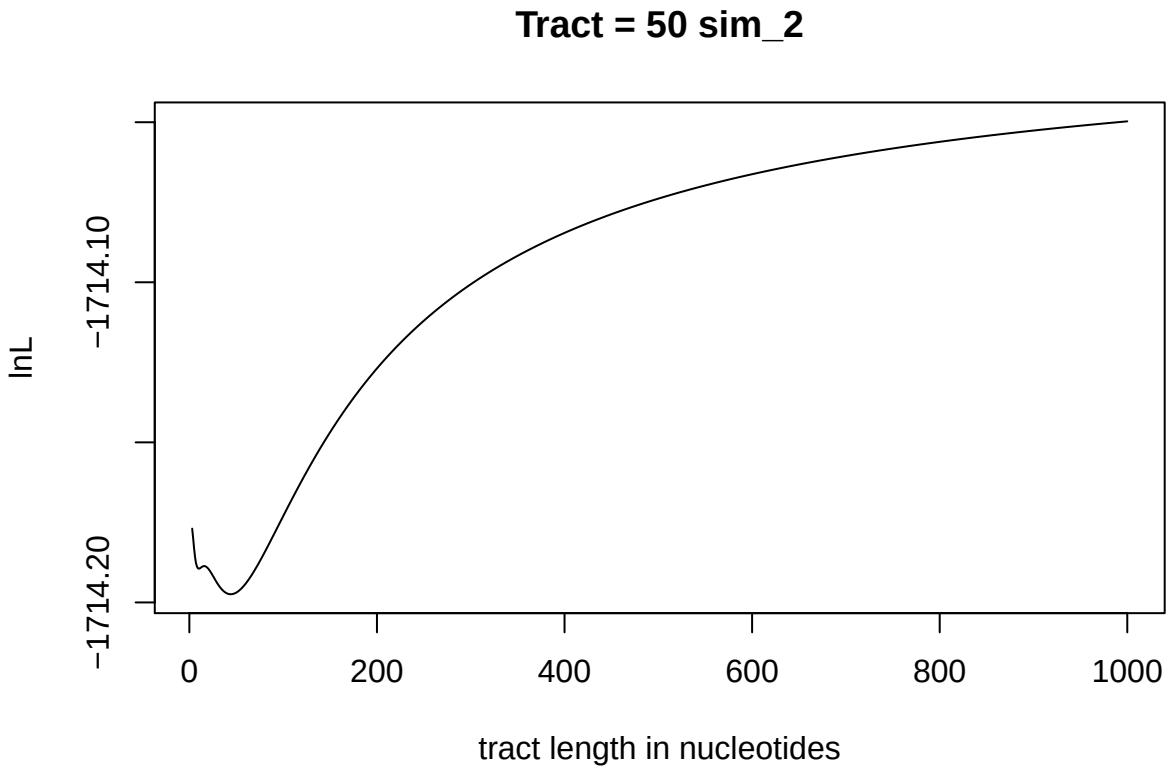

Tract = 3 sim_15
lnL
−1759.0
−1759.2
−1759.4
0
200
400
600
800
1000
tract length in nucleotides

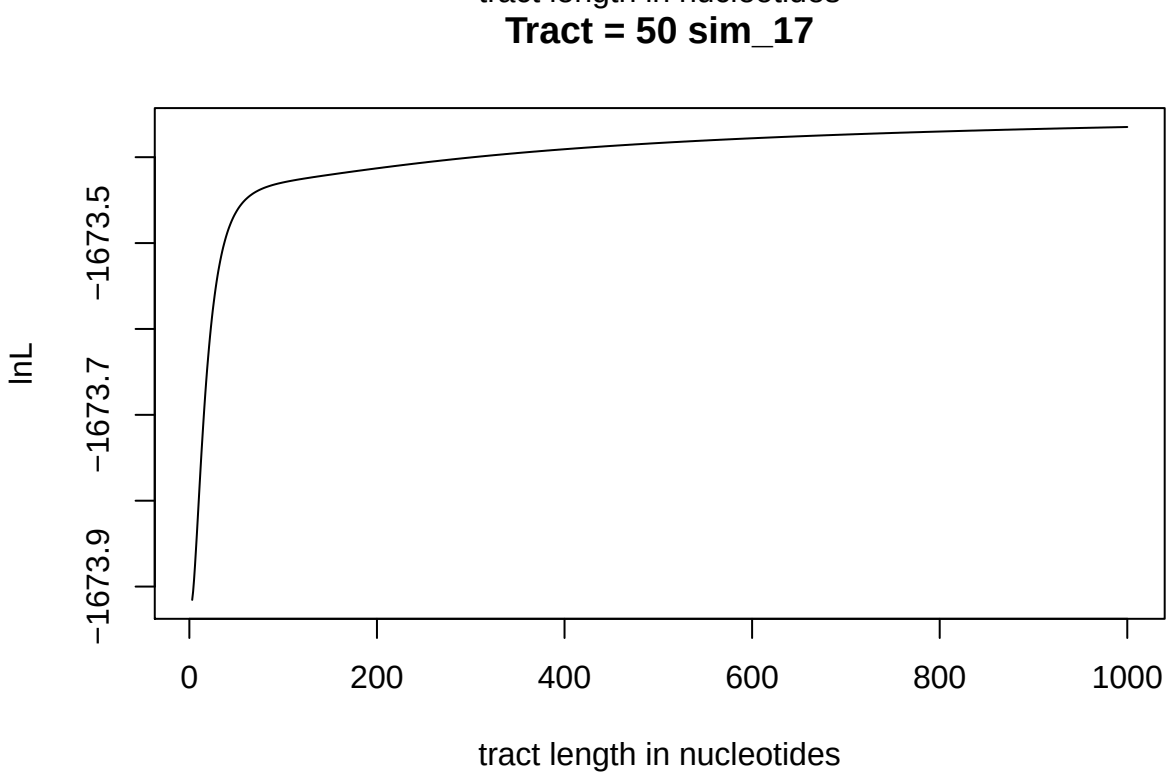
```
## [1] "Tract = 10"
```


Tract = 10 sim_13
lnL
−1754.40
−1754.55
−1754.70
0
200
400
600
800
1000
tract length in nucleotides

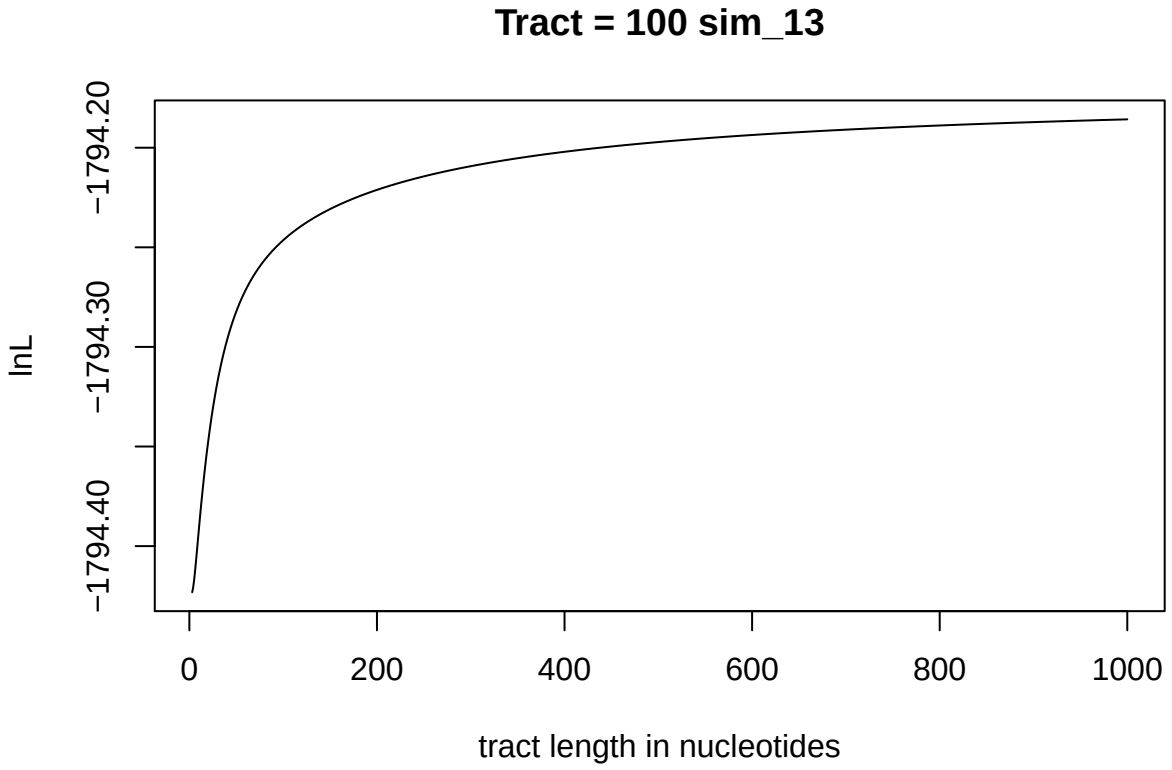

Tract = 10 sim_16
lnL
−1732.4
−1732.8
−1733.2
0
200
400
600
800
1000
tract length in nucleotides

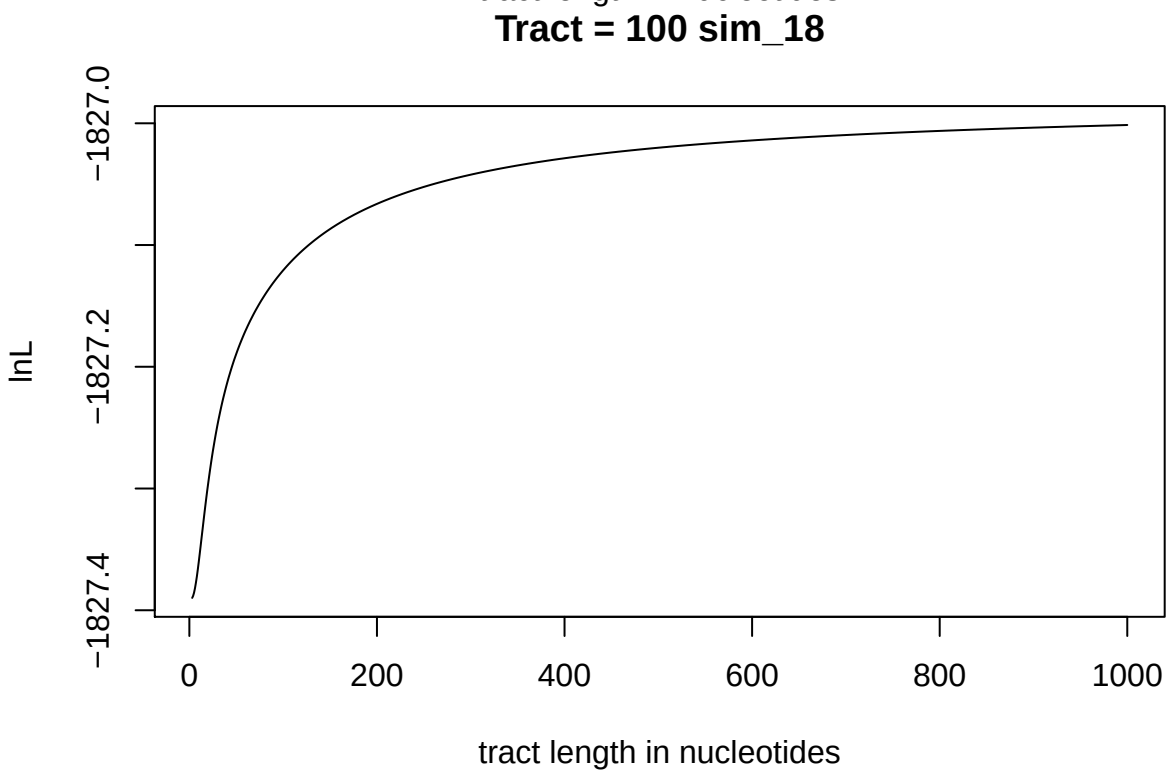
```
## [1] "Tract = 50"
```


Tract = 50 sim_2
lnL
−1714.10
−1714.20
0
200
400
600
800
1000
tract length in nucleotides

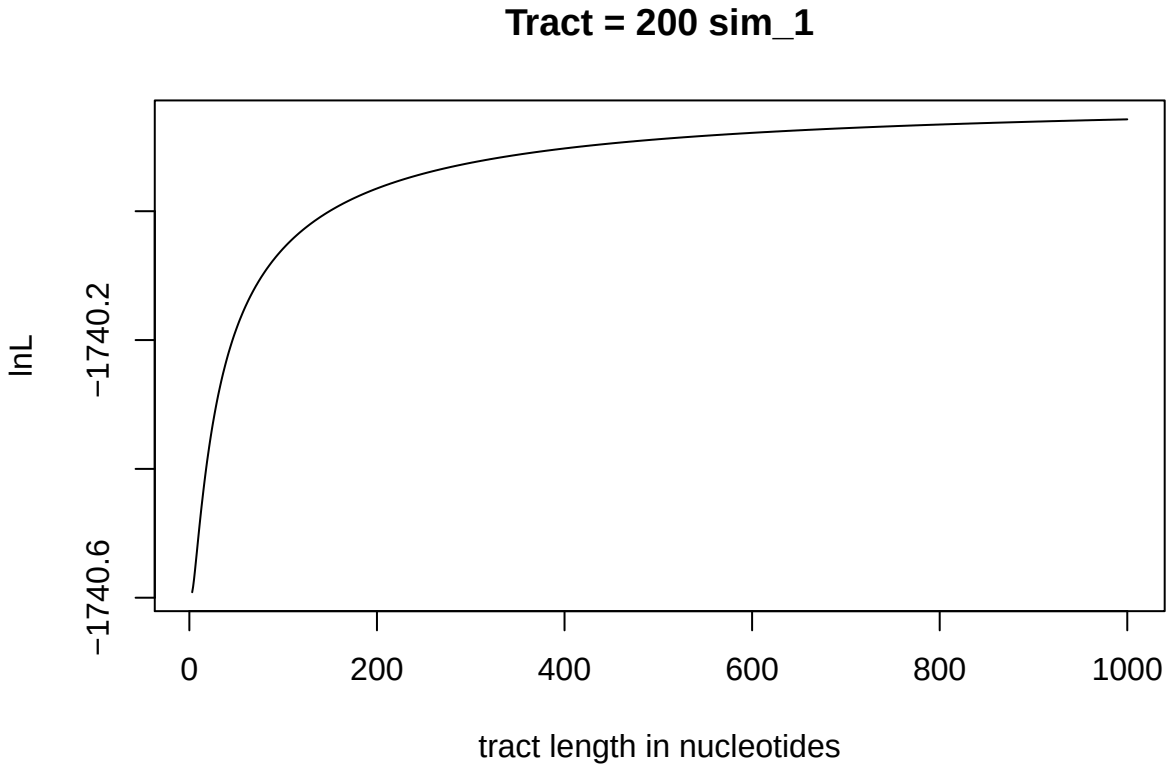

Tract = 50 sim_17
lnL
−1673.5
−1673.7
−1673.9
0
200
400
600
800
1000
tract length in nucleotides

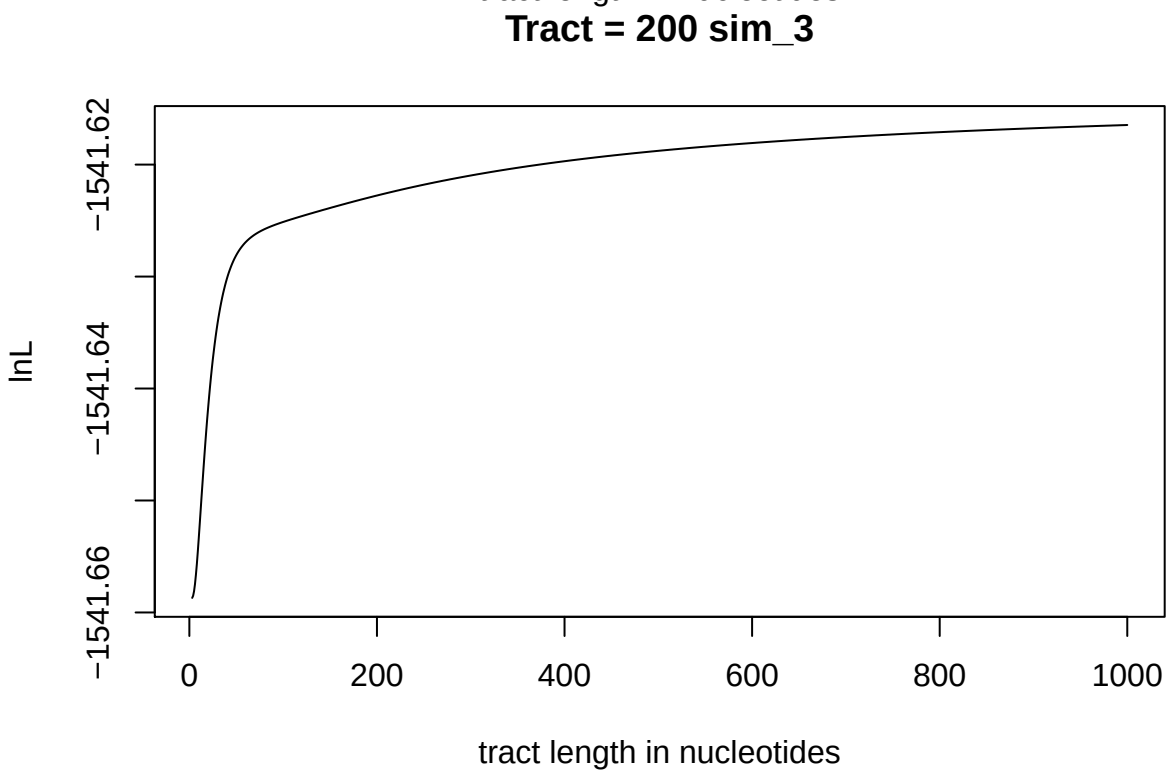
```
## [1] "Tract = 100"
```


Tract = 100 sim_13
lnL
−1794.20
−1794.30
−1794.40
0
200
400
600
800
1000
tract length in nucleotides

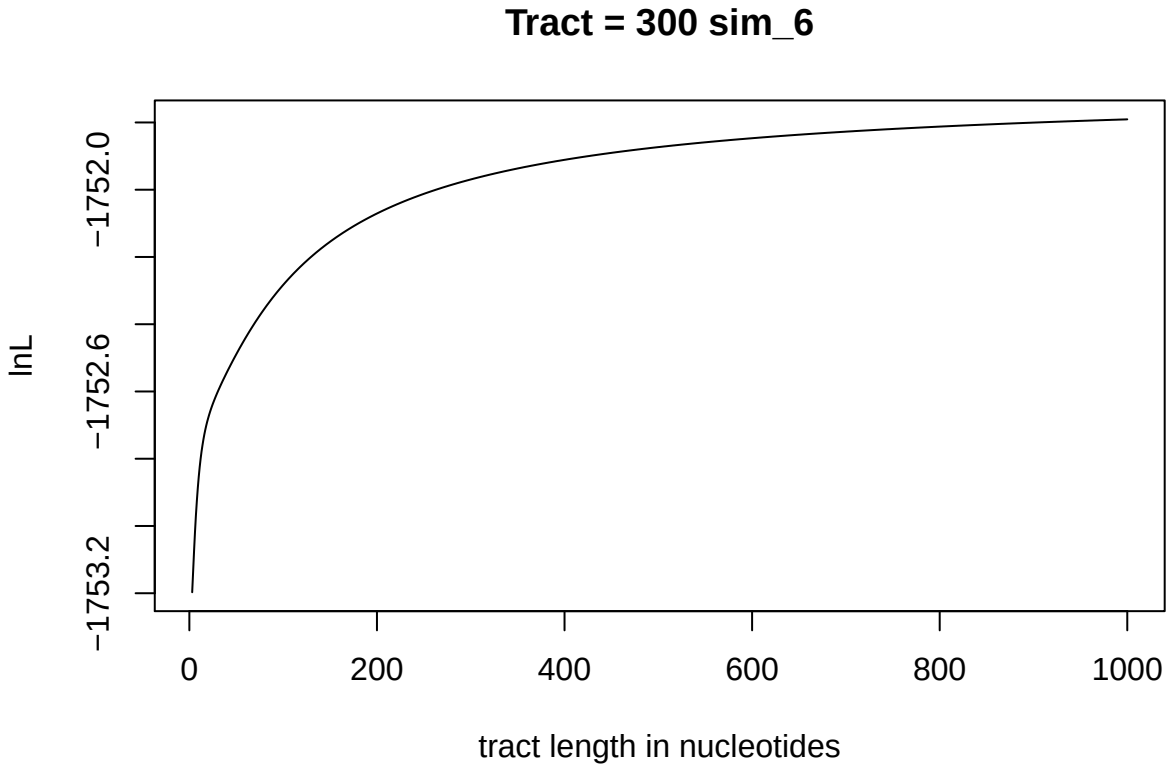

Tract = 100 sim_18
lnL
−1827.0
−1827.2
−1827.4
0
200
400
600
800
1000
tract length in nucleotides

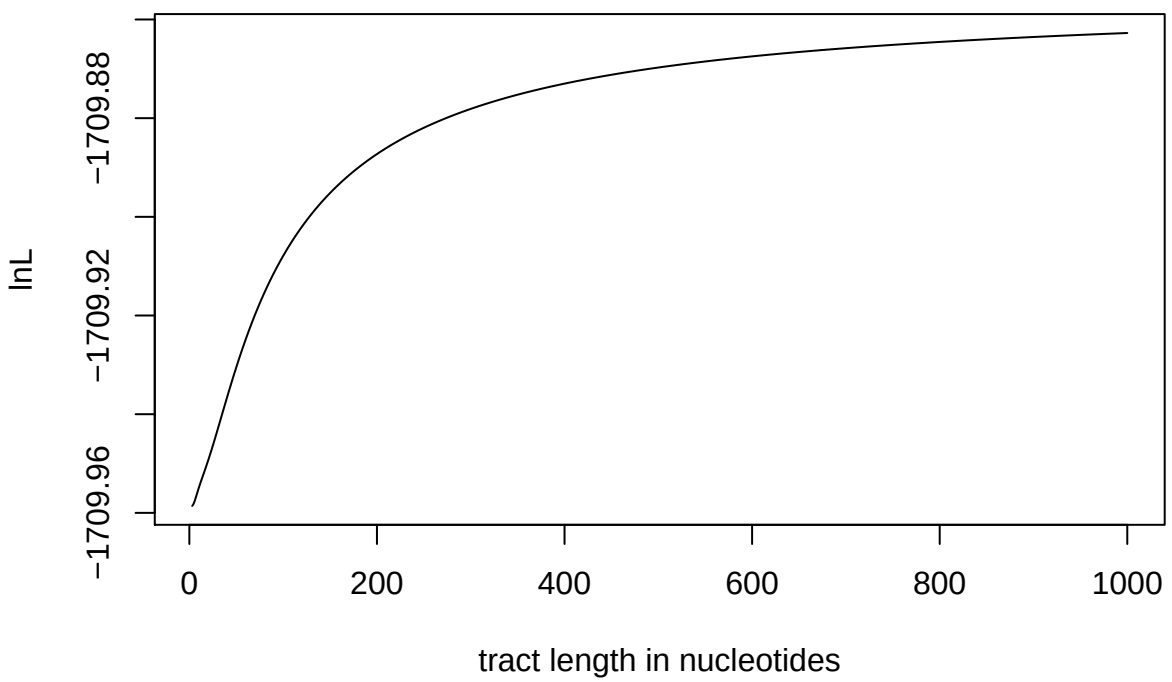
```
## [1] "Tract = 200"
```


Tract = 200 sim_1
lnL
−1740.2
−1740.6
0
200
400
600
800
1000
tract length in nucleotides

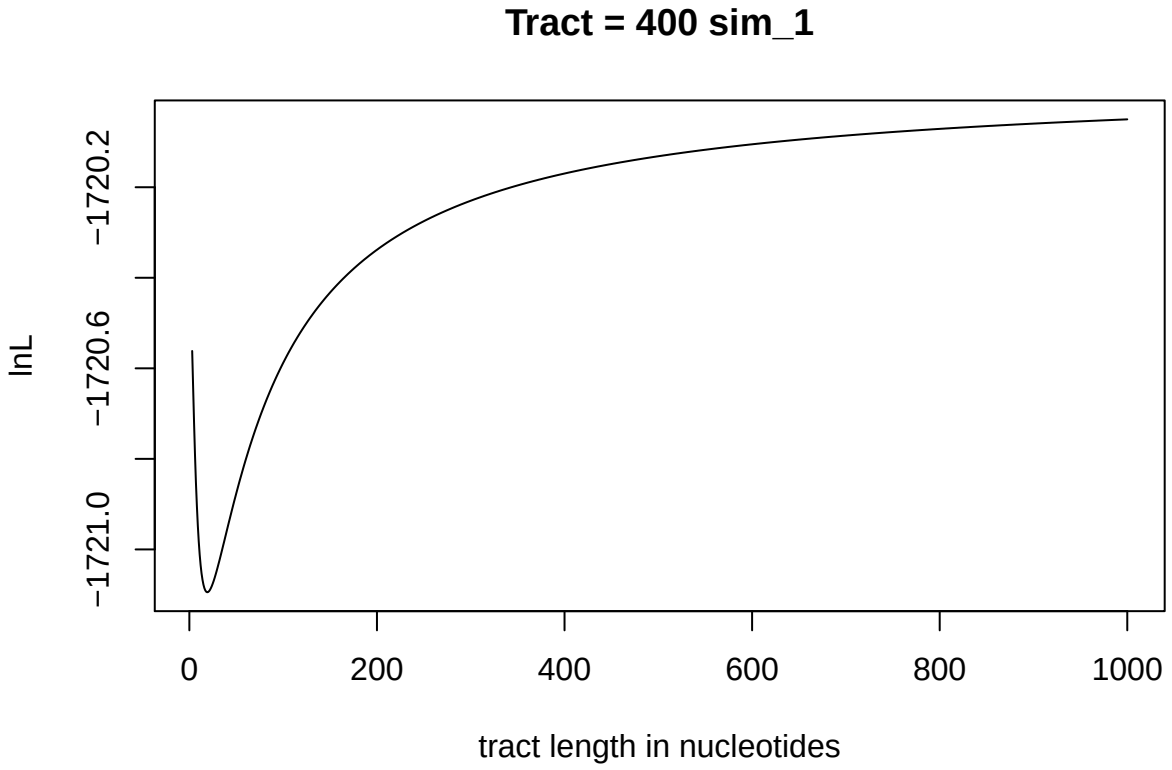

Tract = 200 sim_3
lnL
−1541.62
−1541.64
−1541.66
0
200
400
600
800
1000
tract length in nucleotides

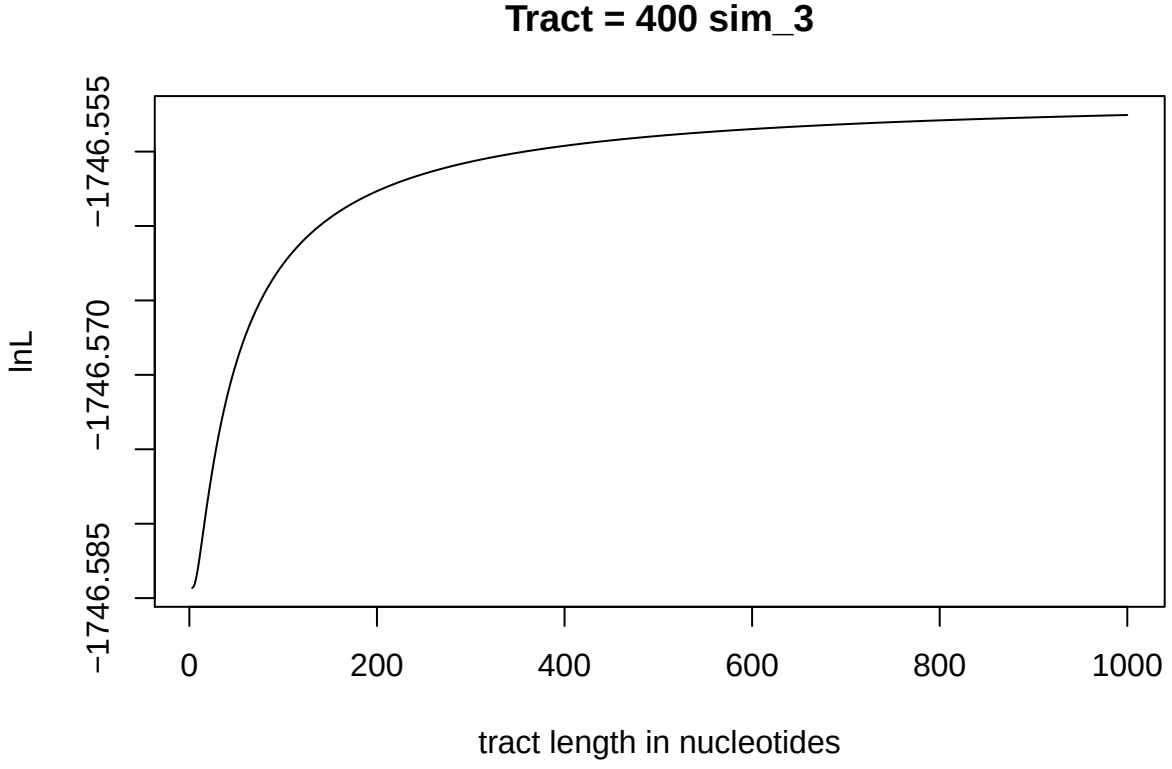
```
## [1] "Tract = 300"
```


Tract = 300 sim_6
lnL
−1752.0
−1752.6
−1753.2
0
200
400
600
800
1000
tract length in nucleotides

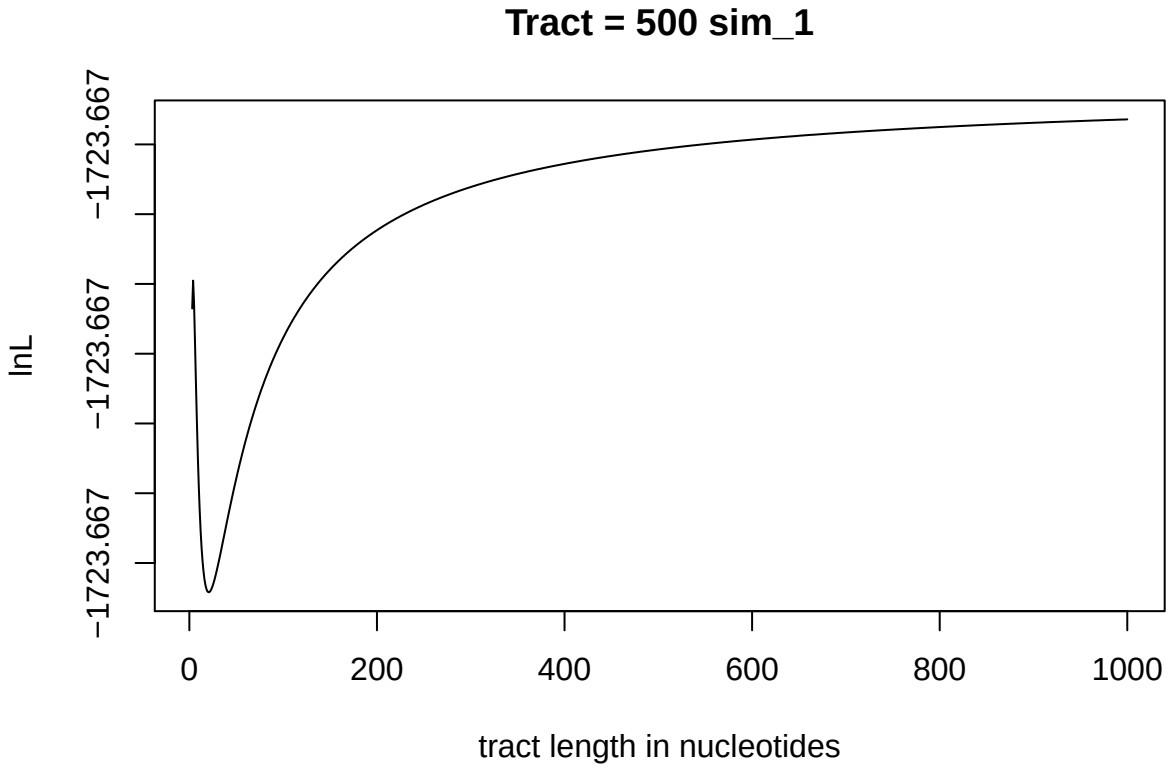

Tract = 300 sim_13
lnL
−1709.88
−1709.92
−1709.96
0
200
400
600
800
1000
tract length in nucleotides

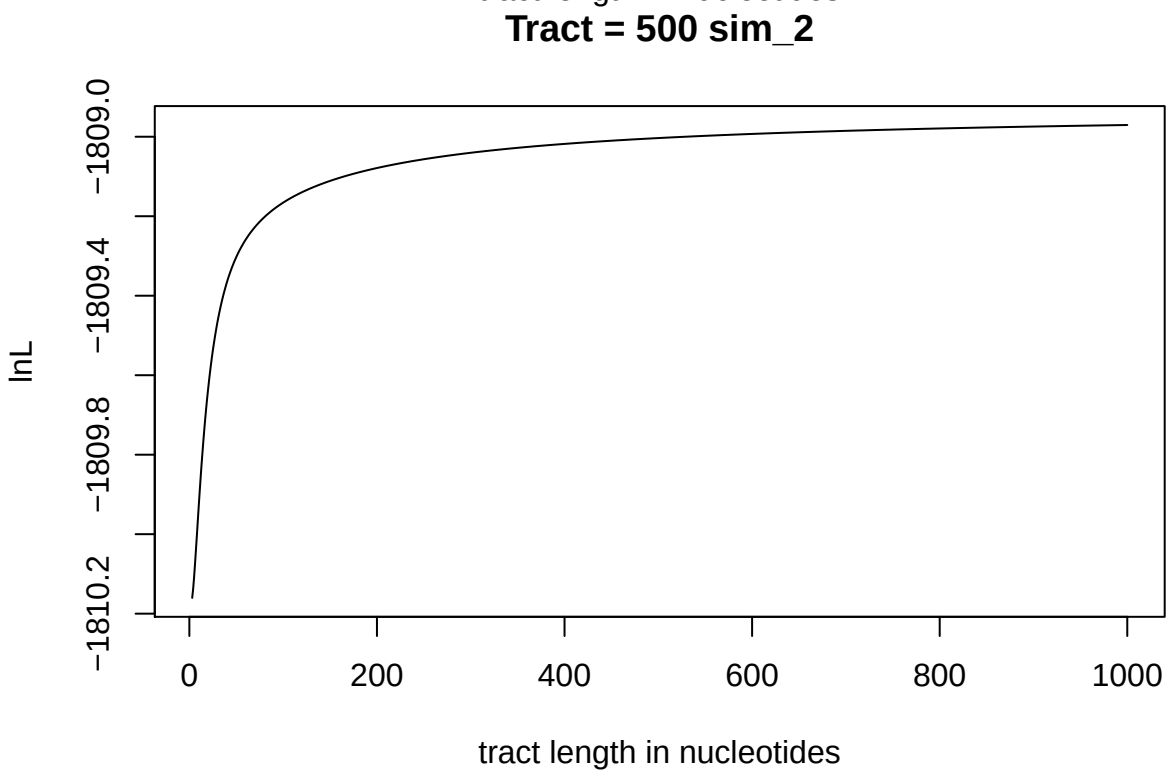
```
## [1] "Tract = 400"
```


Tract = 400 sim_1
lnL
−1720.2
−1720.6
−1721.0
0
200
400
600
800
1000
tract length in nucleotides


Tract = 400 sim_3
lnL
−1746.555
−1746.570
−1746.585
0
200
400
600
800
1000
tract length in nucleotides

```
## [1] "Tract = 500"
```

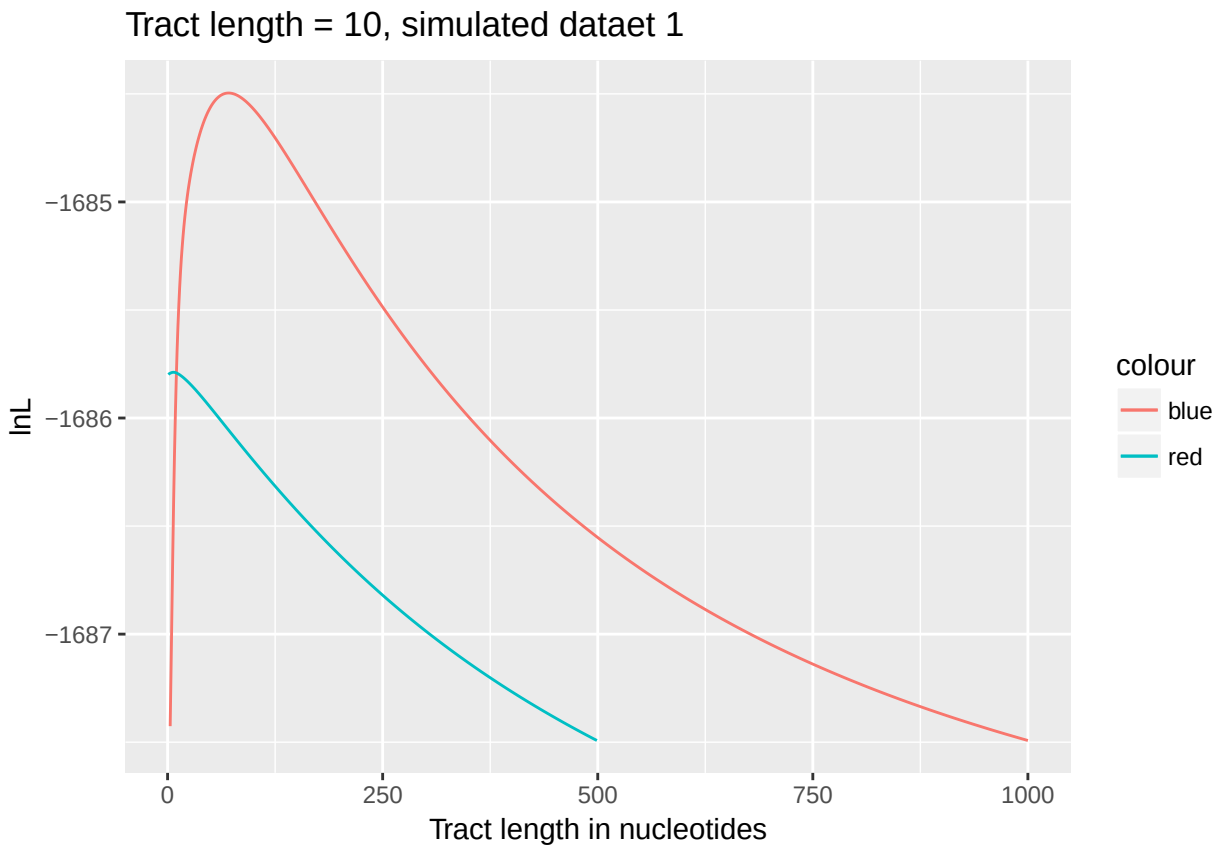




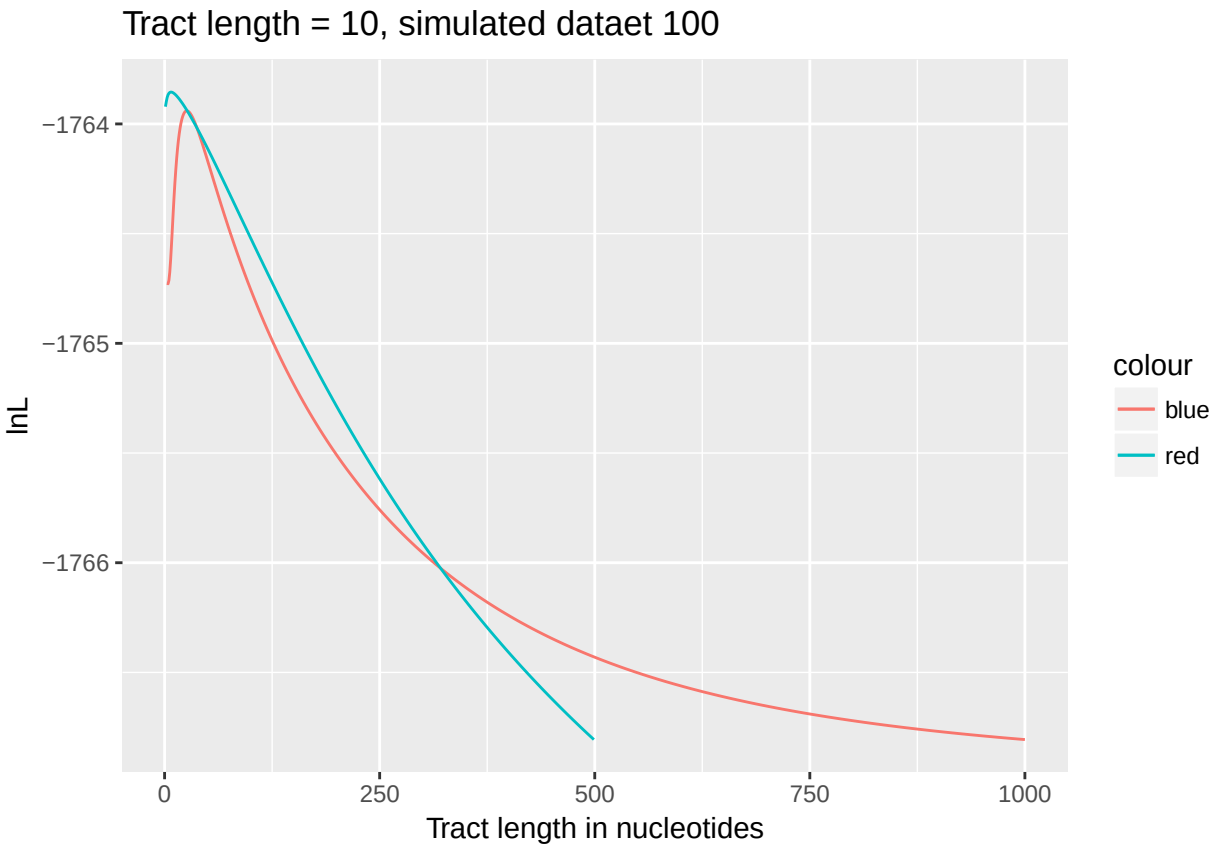
Now plot the two plots of lnL

```
library(ggplot2)
# Tract length = 10
hmm.plot <- read.table("./plot/Tract_10.0/sim_1/HMM_YDR418W_YEL054C_lnL_sim_1_1D_surface.txt")
PSJS.plot <- read.table("./plot/Tract_10.0/sim_1/PSJS_HKY_rv_sim_1_Tract_10.0_lnL_1D_surface.txt")
```

```
ggplot(mapping = aes(x = 3.0*exp(-hmm.plot[,1]), y = hmm.plot[, 2], colour = "blue")) + geom_line() +
  geom_line(aes(x = exp(-PSJS.plot[,1]), y = PSJS.plot[, 2]/488 + min(hmm.plot[, 2]) - min(PSJS.plot[, 2
                colour = "red")) +
  xlab("Tract length in nucleotides") +
  ylab("lnL") +
  ggtitle("Tract length = 10, simulated dataet 1")
```




```
hmm.plot <- read.table("./plot/Tract_10.0/sim_100/HMM_YDR418W_YEL054C_lnL_sim_100_1D_surface.txt")
PSJS.plot <- read.table("./plot/Tract_10.0/sim_100/PSJS_HKY_rv_sim_100_Tract_10.0_lnL_1D_surface.txt")
ggplot(mapping = aes(x = 3.0*exp(-hmm.plot[,1]), y = hmm.plot[, 2], colour = "blue")) + geom_line() +
  geom_line(aes(x = exp(-PSJS.plot[,1]), y = PSJS.plot[, 2]/488 + min(hmm.plot[, 2]) - min(PSJS.plot[, 2
                colour = "red")) +
  xlab("Tract length in nucleotides") +
  ylab("lnL") +
  ggtitle("Tract length = 10, simulated dataet 100")
```



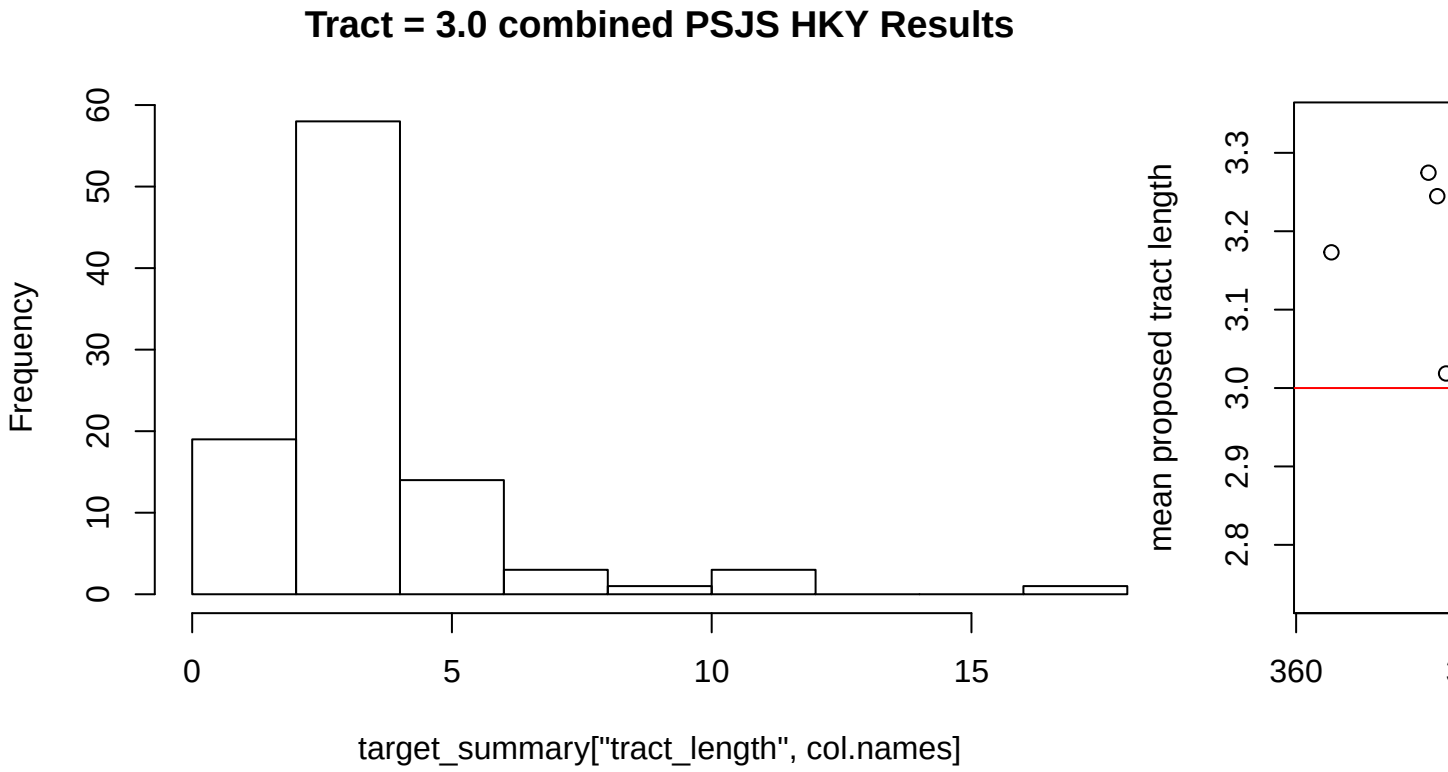
Now see how estimates from the two approaches differ from the actual mean tract length in each simulated data set.

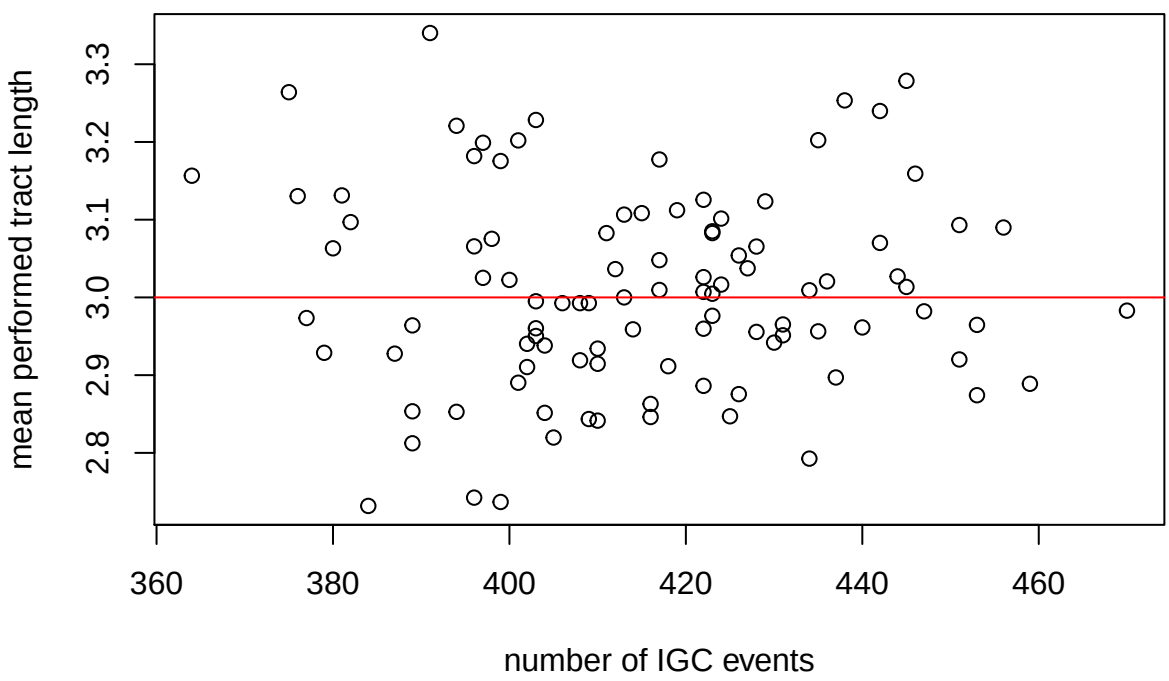
```
# for(tract in Tract.list){
#   sim.info <- get(paste("sim.tract.", toString(tract), sep = ""))
#   # Show mean and sd
#   print(matrix(c("proposed mean", mean(sim.info["mean proposed tract length", ]),
#           "performed mean", mean(sim.info["mean performed tract length", ]),
#           "geometric mean", tract,
#           "proposed sd", mean(sim.info["sd proposed tract length",], na.rm = TRUE),
#           "performed sd", mean(sim.info["sd performed tract length",], na.rm = TRUE),
#           "geometric sd", sqrt(tract^2-tract*3.0)), 2, 6))
#   hmm.info <- get(paste("HMM_Tract_", toString(tract), "_plot", sep = ""))
#   PSJS.info <- get(paste("PSJS_Tract_", toString(tract), "_summary", sep = ""))
#   shared.col <- intersect(colnames(hmm.info), colnames(PSJS.info))
#
#   # Now show the ratio of HMM estimated tract / actual mean tracts in simulation
#   hmm.ratio <- hmm.info[1, shared.col]/sim.info[1, shared.col]
#   hist(hmm.ratio, main = paste("HMM ratio Tract = ", toString(tract), sep = ""))
#   print(matrix(c("HMM mean", mean(hmm.ratio), "HMM sd", sd(hmm.ratio)), 2, 2))
#
#   # Now show the ratio of PSJS estimated tract / actual mean tracts in simulation
#   PSJS.ratio <- PSJS.info["tract_length", shared.col]/sim.info[1, shared.col]
#   hist(PSJS.ratio, main = paste("PSJS ratio Tract = ", toString(tract), sep = ""))
#   print(matrix(c("PSJS mean", mean(PSJS.ratio), "PSJS sd", sd(PSJS.ratio)), 2, 2))
#
# }
```

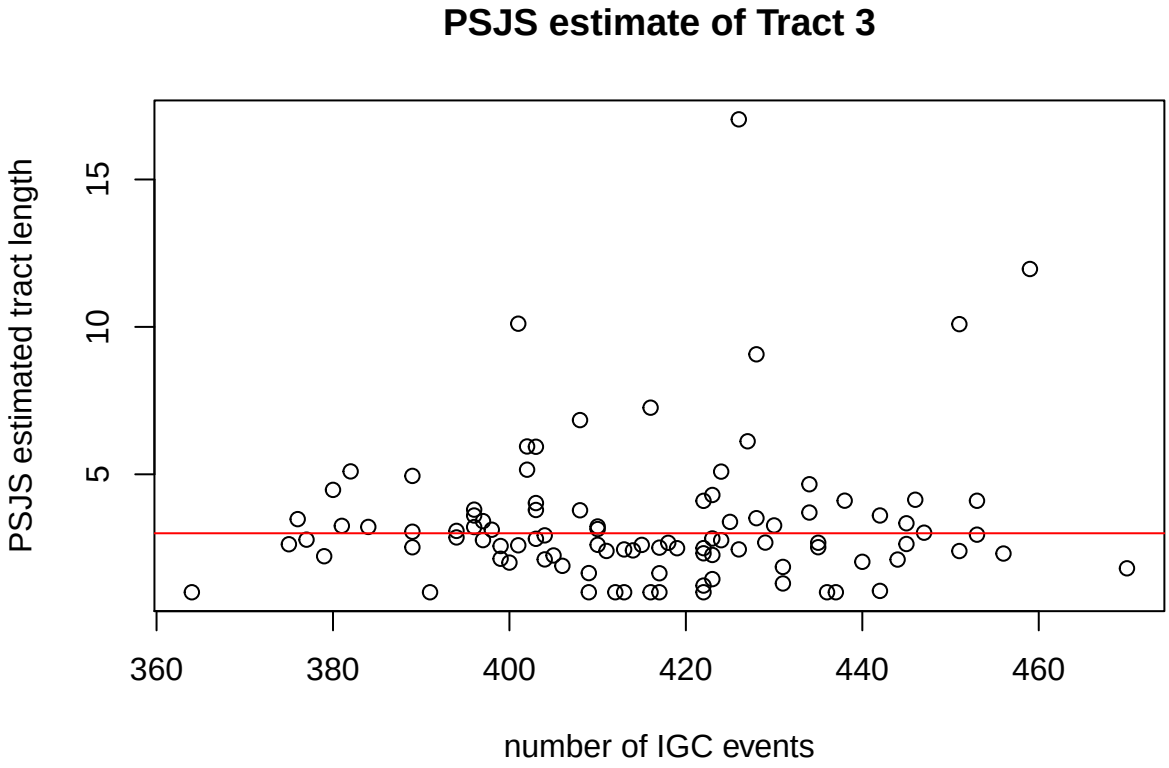
09152017

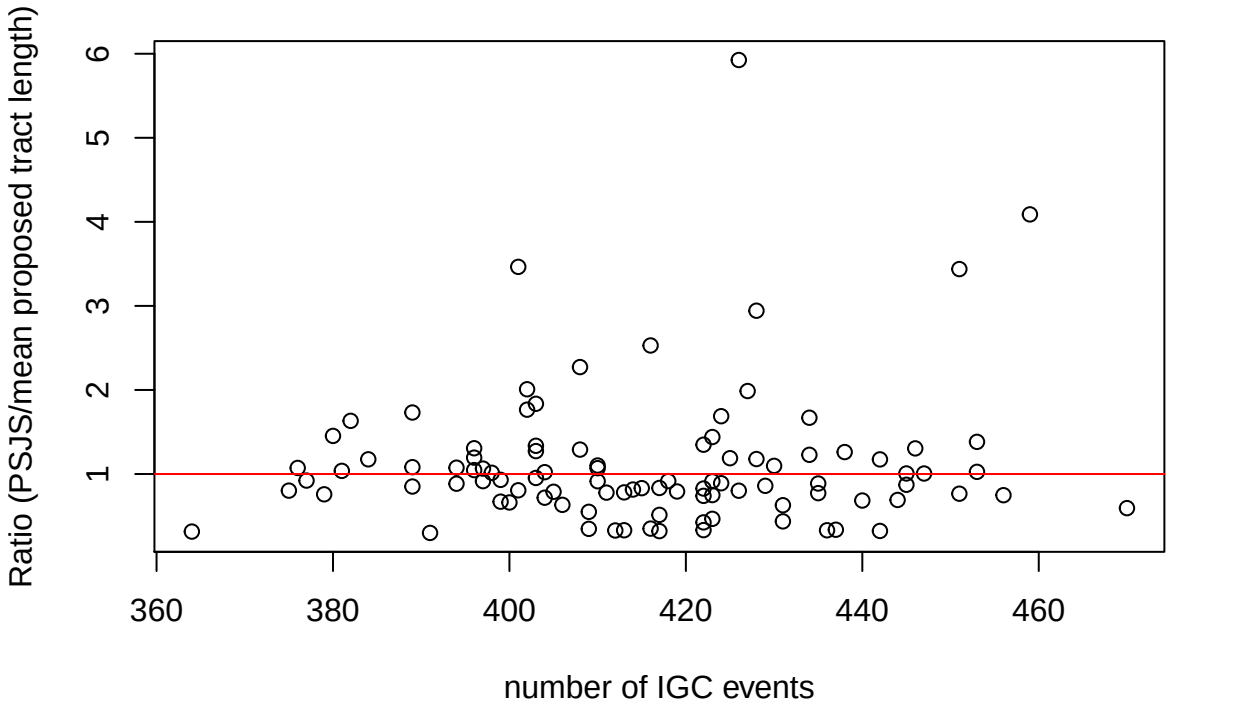
Show the PSJS estimated results for Simulated datasets with estimated tau

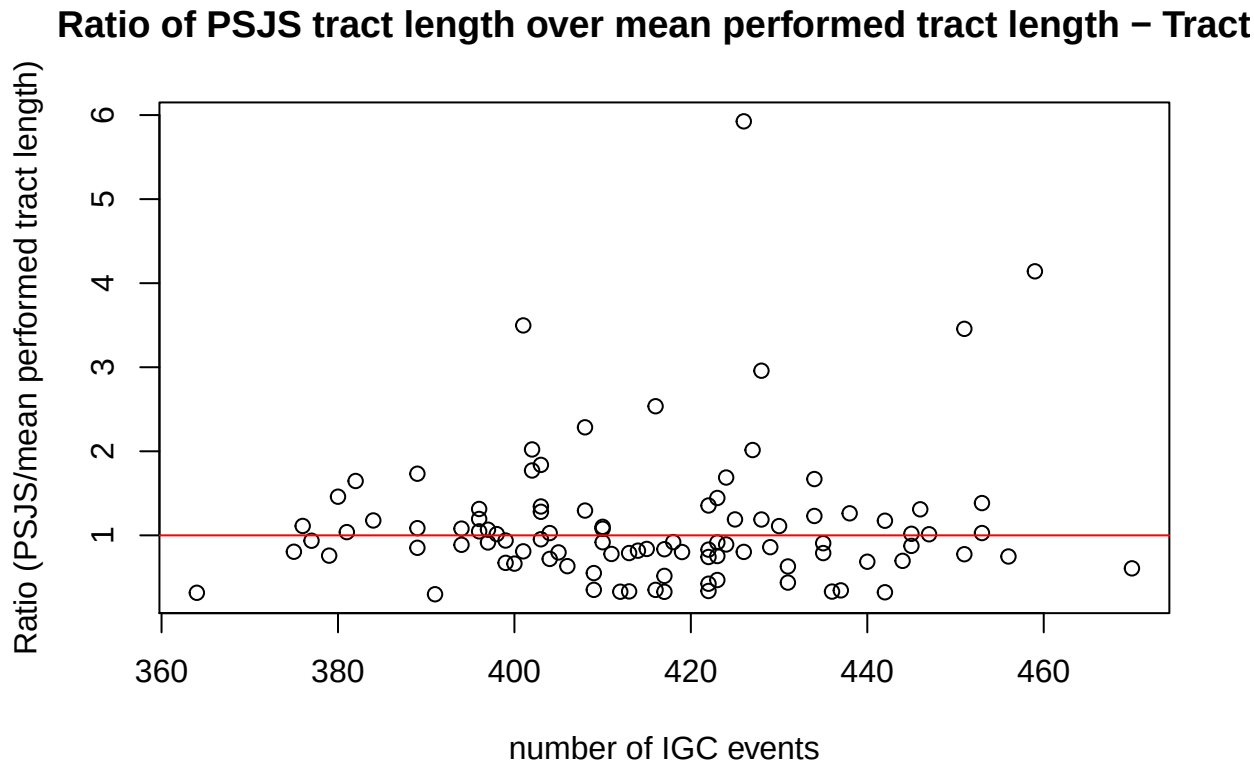
```
for(tract in Tract.list){
  target_summary <- get(paste("PSJS_HKY_Tract_", toString(tract), "_combined_summary", sep = ""))
  col.names <- target_summary["tract_length", ] < 10*tract
  hist(target_summary["tract_length", col.names],
       main = paste("Tract = ", toString(tract), ".0 combined PSJS HKY Results", sep = ""))
  sim_info <- get(paste("sim.tract.", toString(tract), sep = ""))
  plot(sim_info["num IGC", ], sim_info["mean proposed tract length", ],
       main = paste("Simulation info of Tract ", toString(tract), sep = ""),
       xlab = "number of IGC events", ylab = "mean proposed tract length")
  abline(h = tract, col = "red")

  plot(sim_info["num IGC", ], sim_info["mean performed tract length", ],
       main = paste("Simulation Info of Tract ", toString(tract), sep = ""),
       xlab = "number of IGC events", ylab = "mean performed tract length")
  abline(h = tract, col = "red")

  plot(sim_info["num IGC", colnames(target_summary)[col.names]], target_summary["tract_length", col.nam
       main = paste("PSJS estimate of Tract ", toString(tract), sep = ""),
       xlab = "number of IGC events", ylab = "PSJS estimated tract length")
  abline(h = tract, col = "red")

  plot(sim_info["num IGC", colnames(target_summary)[col.names]],
       target_summary["tract_length", col.names] / sim_info["mean proposed tract length", colnames(targ
       main = paste(" Ratio of PSJS tract length over mean proposed tract length - Tract ", toString(tra
       xlab = "number of IGC events", ylab = "Ratio (PSJS/mean proposed tract length)")
  abline(h = 1.0, col = "red")

  plot(sim_info["num IGC", colnames(target_summary)[col.names]],
       target_summary["tract_length", col.names] / sim_info["mean performed tract length", colnames(tar
       main = paste(" Ratio of PSJS tract length over mean performed tract length - Tract ", toString(t
       xlab = "number of IGC events", ylab = "Ratio (PSJS/mean performed tract length)")
  abline(h = 1.0, col = "red")

  print(paste("Tract = ", toString(tract), ".0 combined PSJS HKY Results", sep = ""))
  print(matrix(c("Total samples", sum(col.names),
                 "mean", mean(target_summary["tract_length", col.names]),
                 "sd", sd(target_summary["tract_length", col.names])), 2, 3))
}
```


Tract = 3.0 combined PSJS HKY Results
Frequency
0
10
20
30
40
50
60
0
5
10
15
target_summary["tract_length", col.names]
mean proposed tract length
2.8
2.9
3.0
3.1
3.2
3.3
360

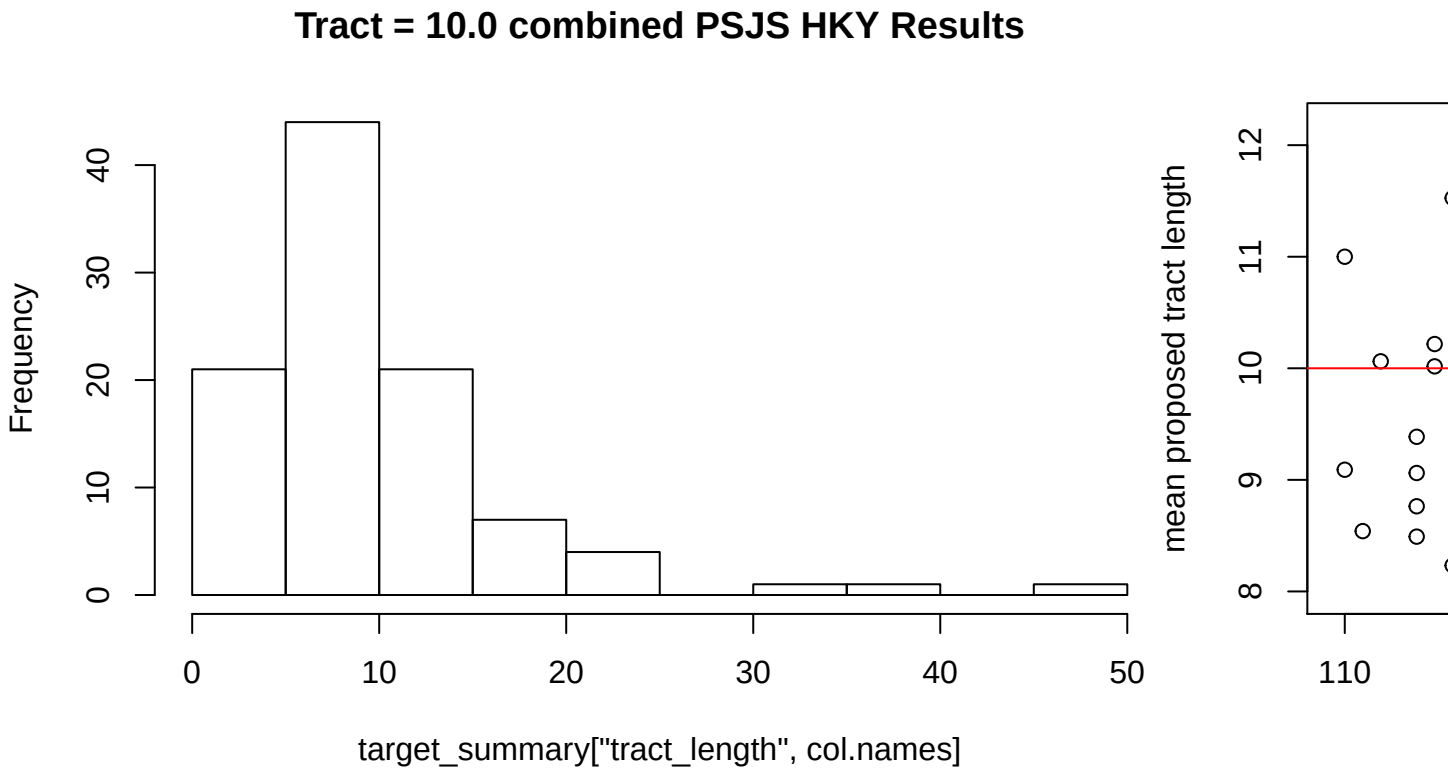

Simulation Info of Tract 3
mean performed tract length
2.8
2.9
3.0
3.1
3.2
3.3
360
380
400
420
440
460
number of IGC events

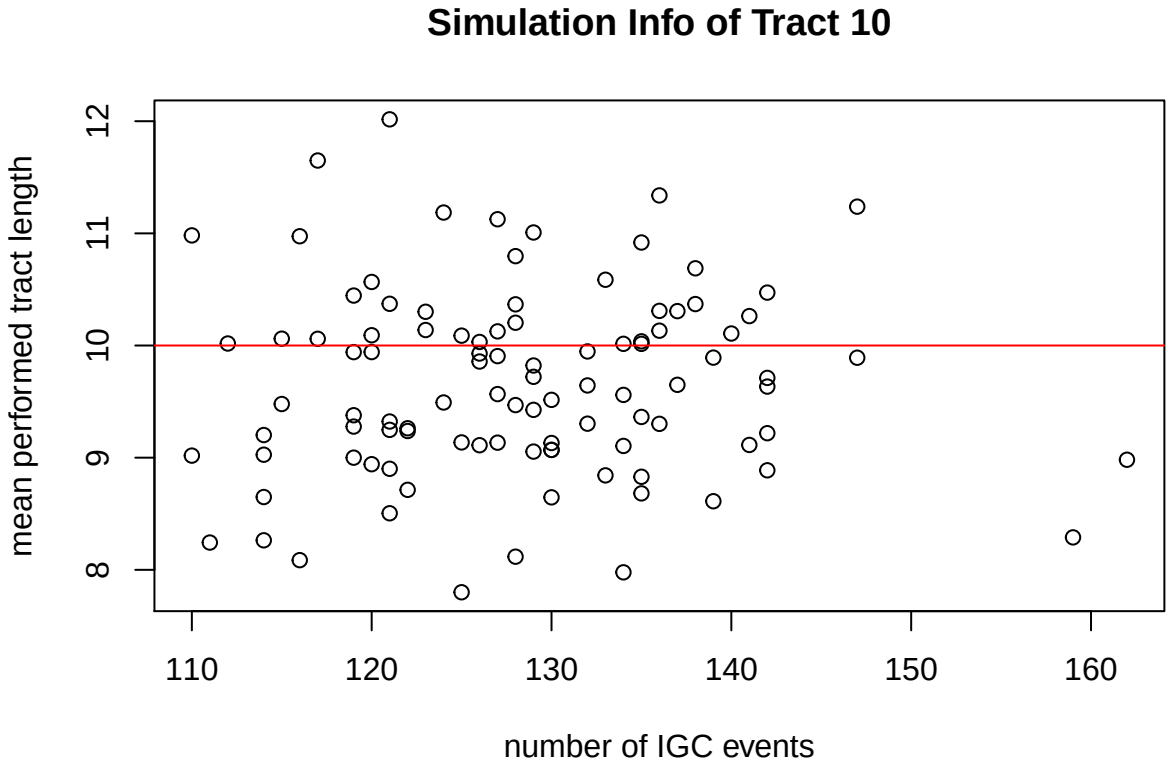

PSJS estimate of Tract 3
PSJS estimated tract length
5
10
15
360
380
400
420
440
460
number of IGC events

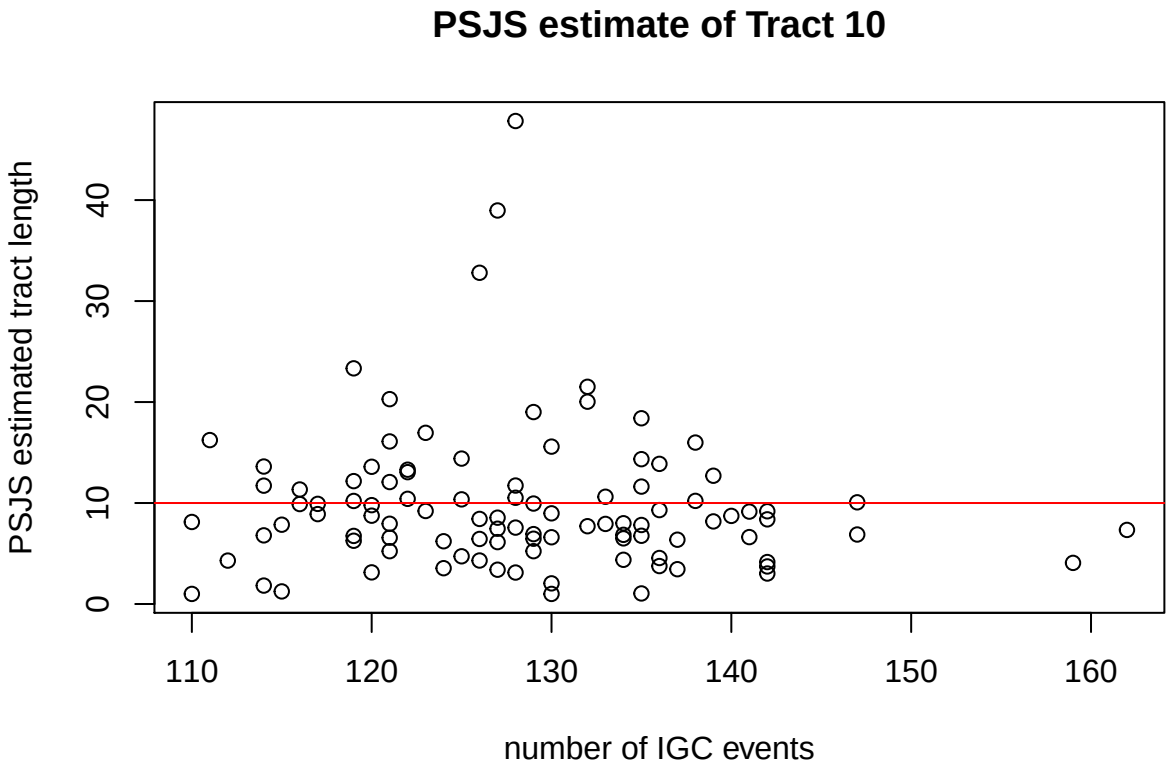

Ratio of PSJS tract length over mean proposed tract length − Tract
Ratio (PSJS/mean proposed tract length)
1
2
3
4
5
6
360
380
400
420
440
460
number of IGC events

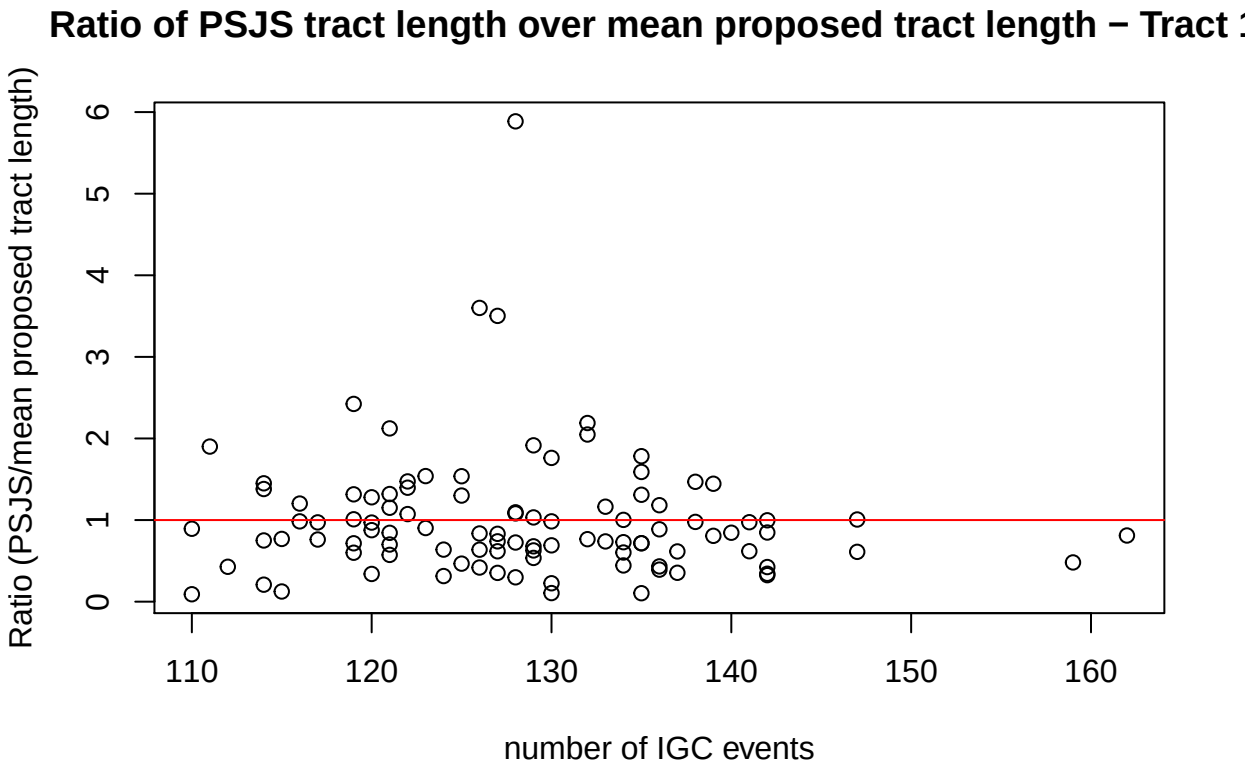


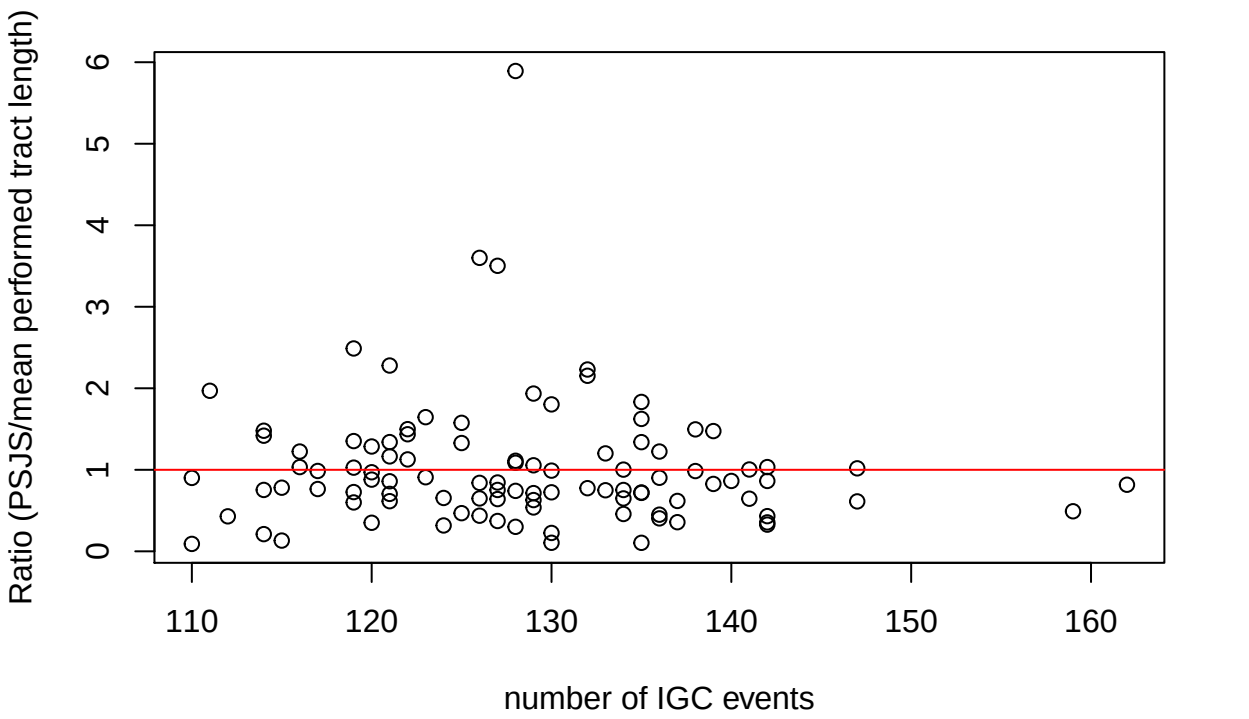

```
## [1] "Tract = 3.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "99"            "3.35636357555957" "2.40699716566058"
```



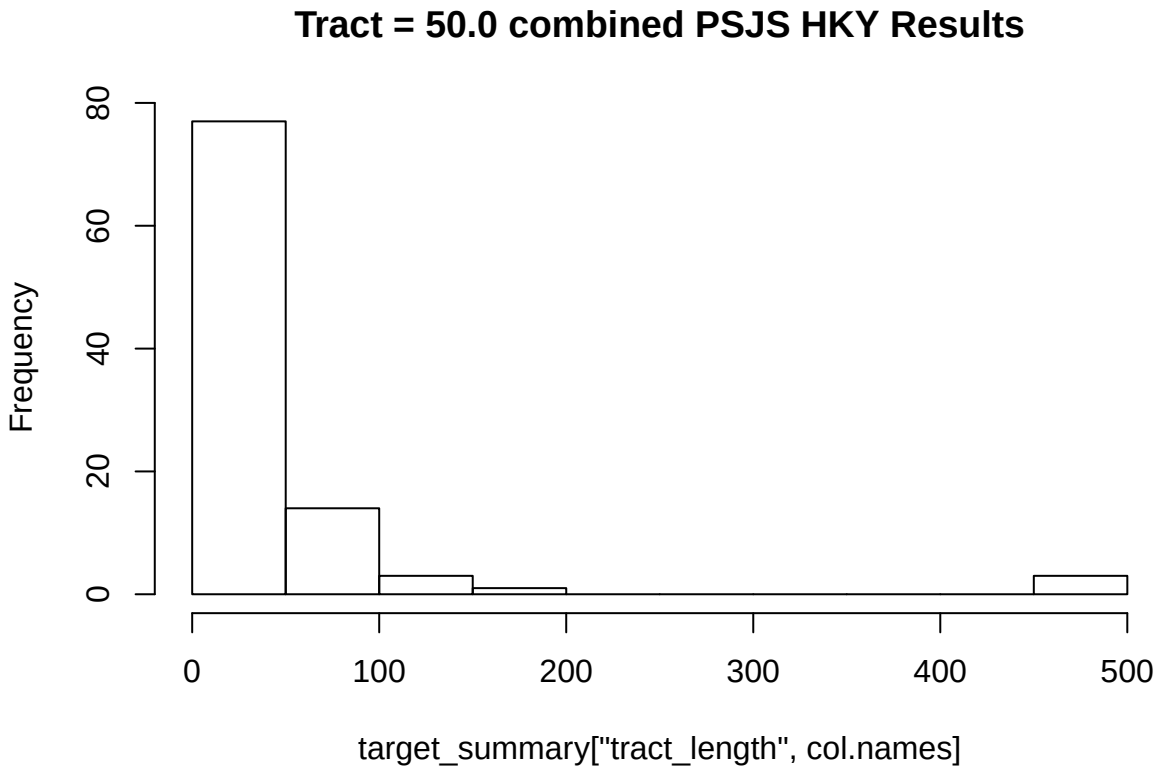

Simulation Info of Tract 10
mean performed tract length
8
9
10
11
12
110
120
130
140
150
160
number of IGC events

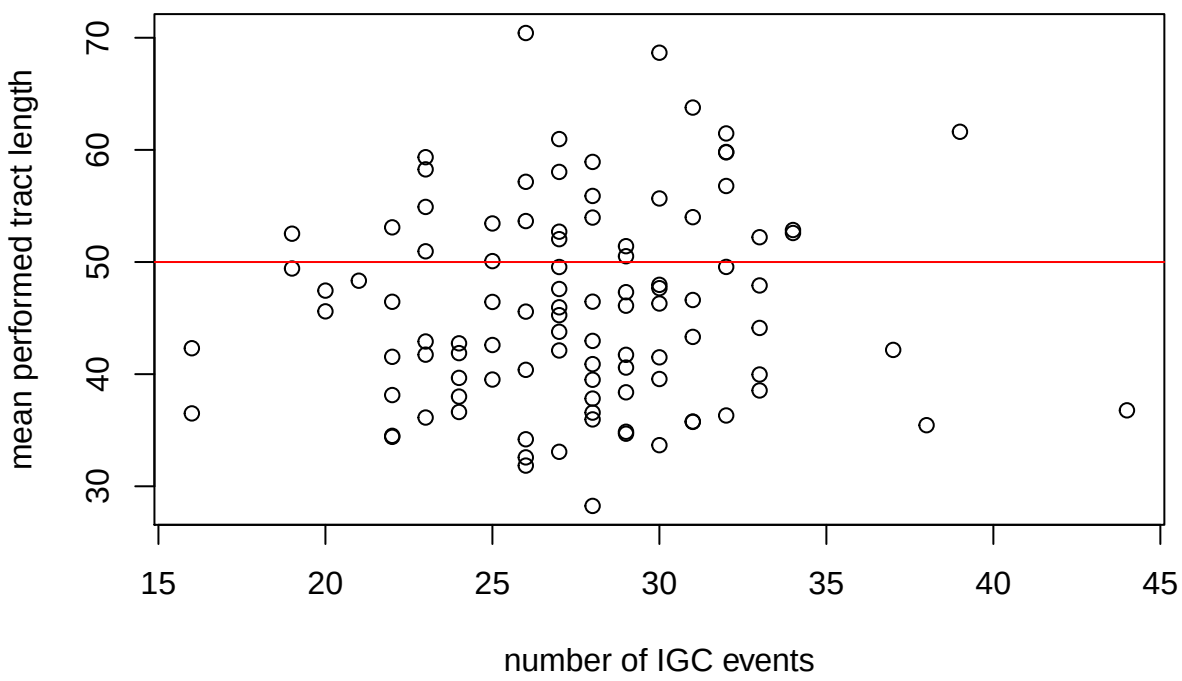

PSJS estimate of Tract 10
PSJS estimated tract length
0
10
20
30
40
110
120
130
140
150
160
number of IGC events

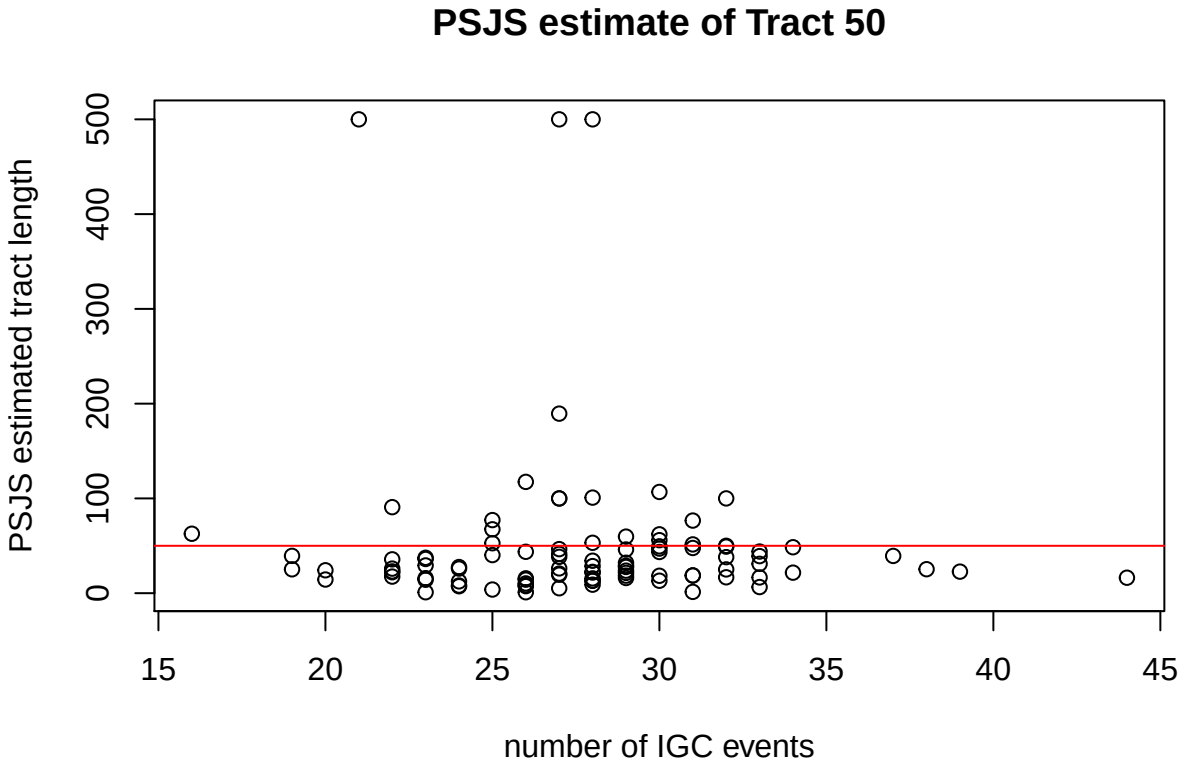


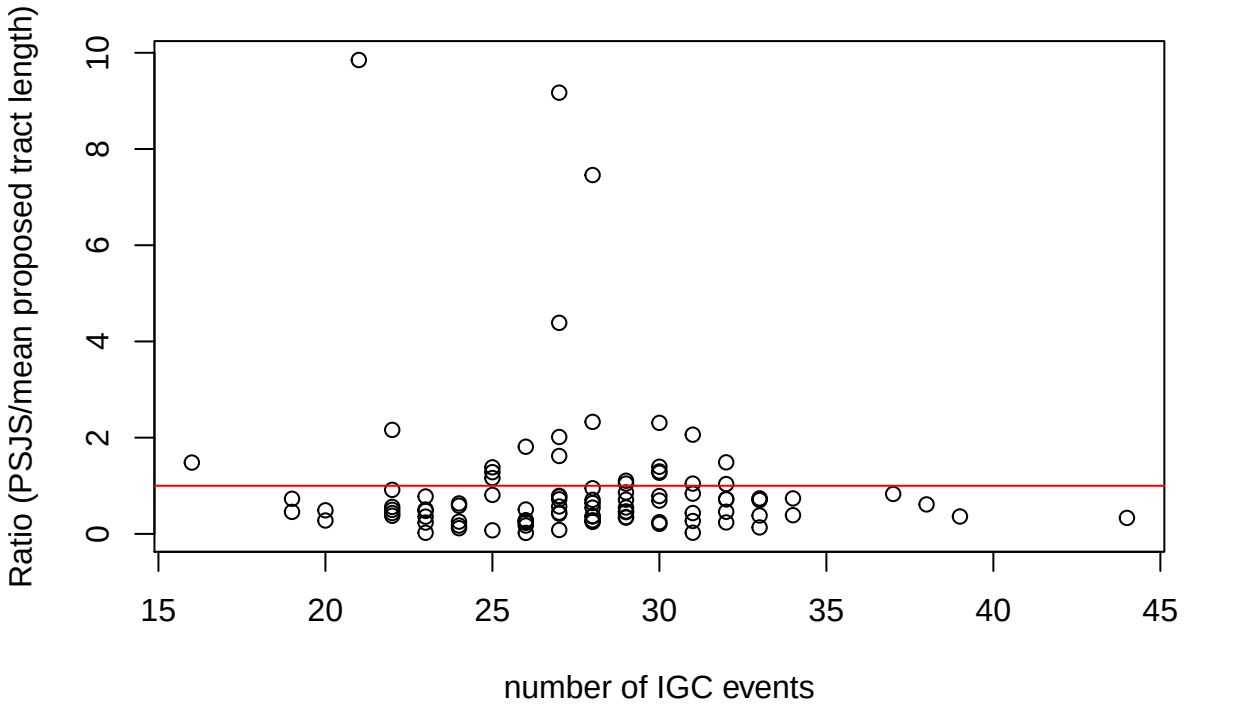


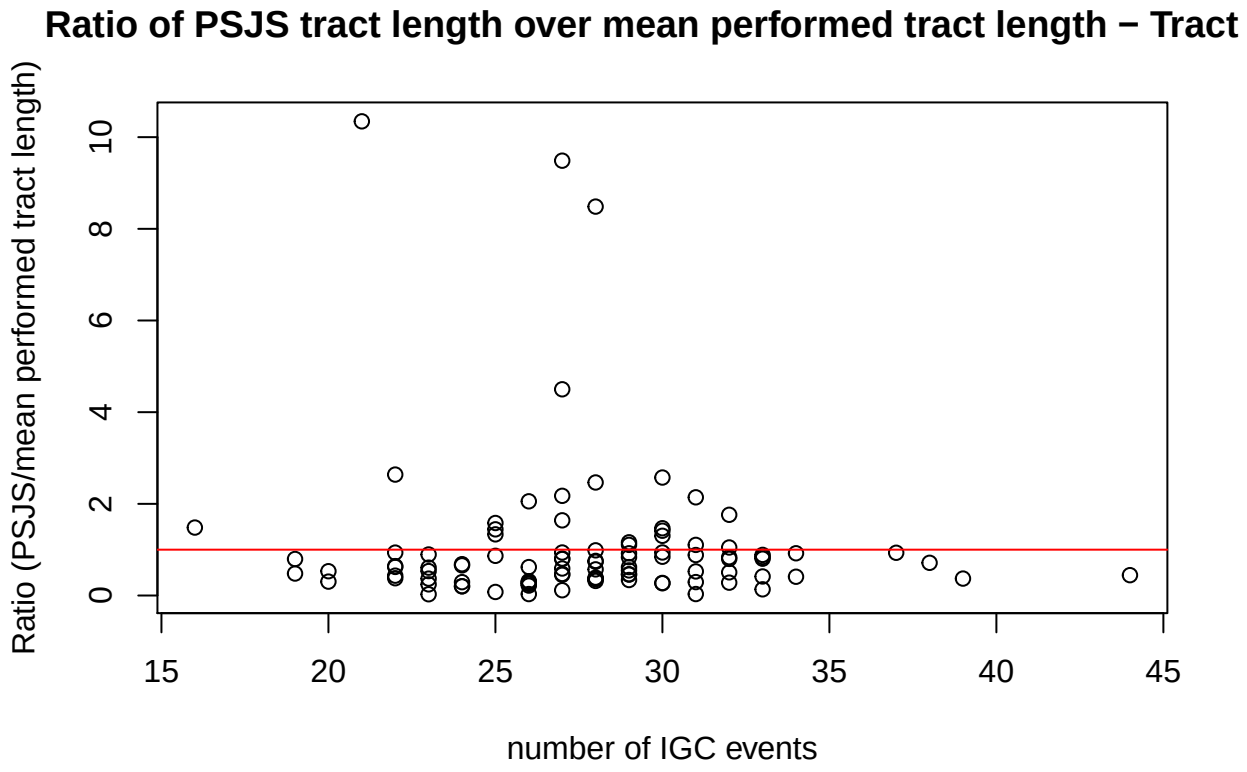

```
## [1] "Tract = 10.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "100"           "9.79420203503121" "7.17615180954532"
```


Tract = 50.0 combined PSJS HKY Results
Frequency
0
20
40
60
80
0
100
200
300
400
500
target_summary["tract_length", col.names]
mean proposed tract length
40
50
60
70
80
15

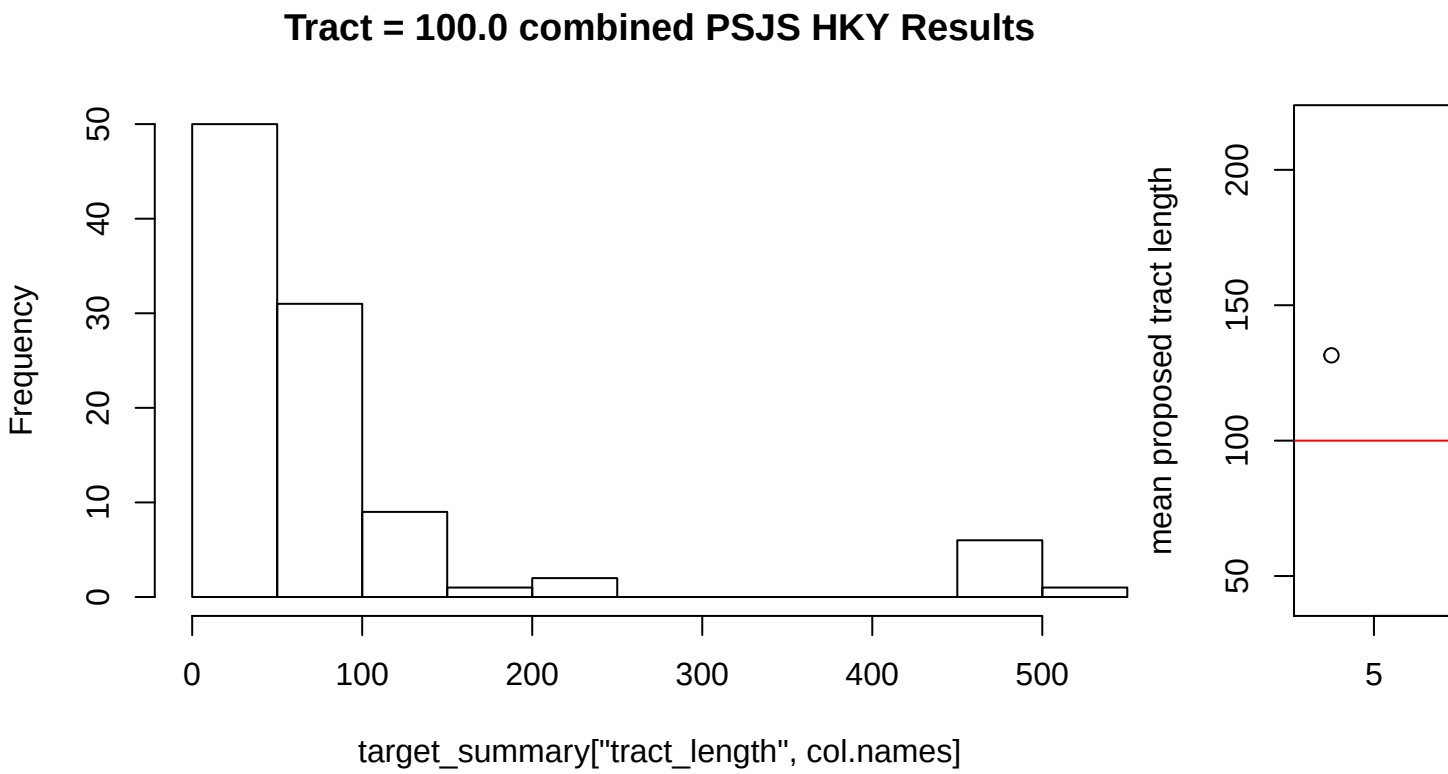

Simulation Info of Tract 50
mean performed tract length
30
40
50
60
70
15
20
25
30
35
40
45
number of IGC events

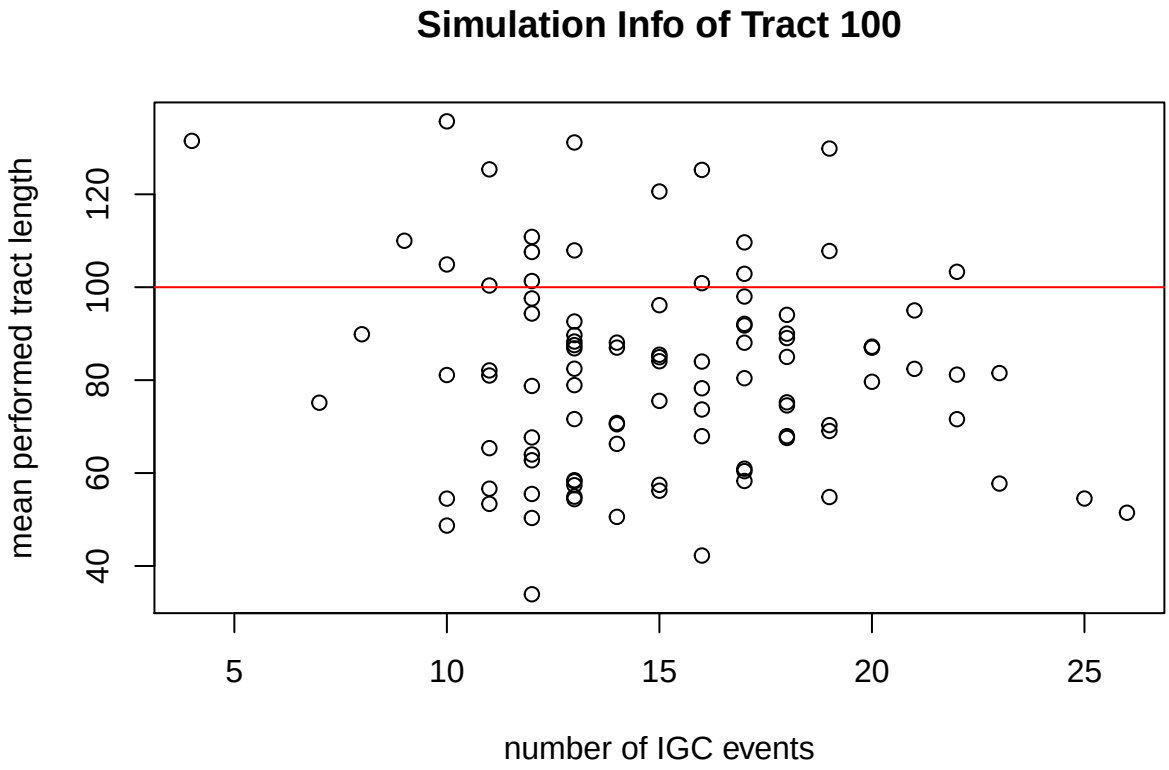

PSJS estimate of Tract 50
PSJS estimated tract length
0
100
200
300
400
500
15
20
25
30
35
40
45
number of IGC events

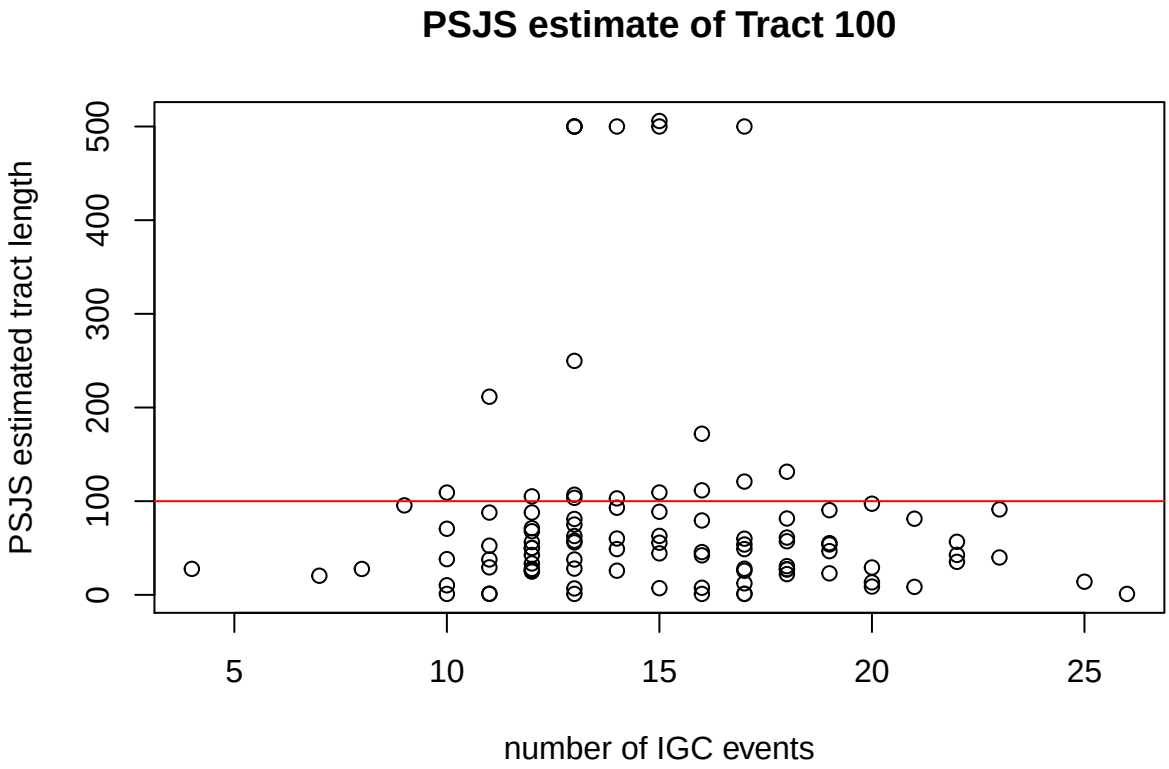

Ratio of PSJS tract length over mean proposed tract length − Tract
Ratio (PSJS/mean proposed tract length)
0
2
4
6
8
10
15
20
25
30
35
40
45
number of IGC events

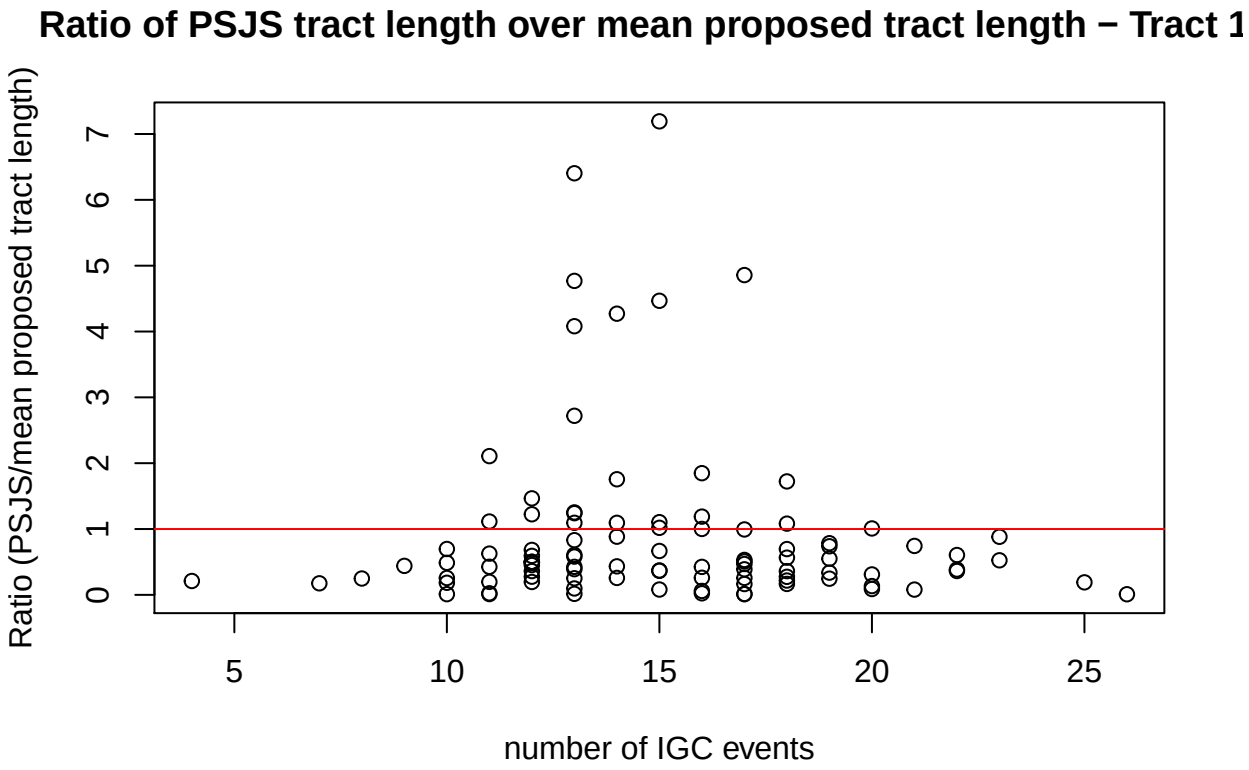


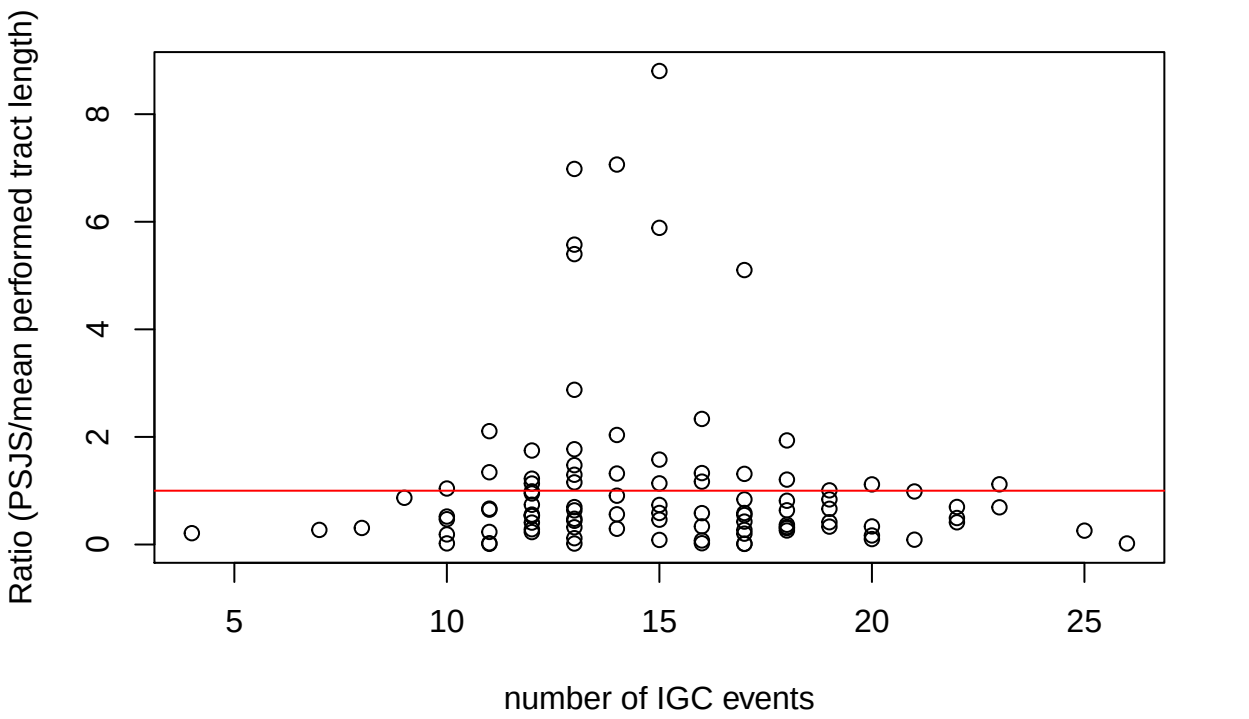

```
## [1] "Tract = 50.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "98"            "50.3095661868595" "85.6445738100097"
```



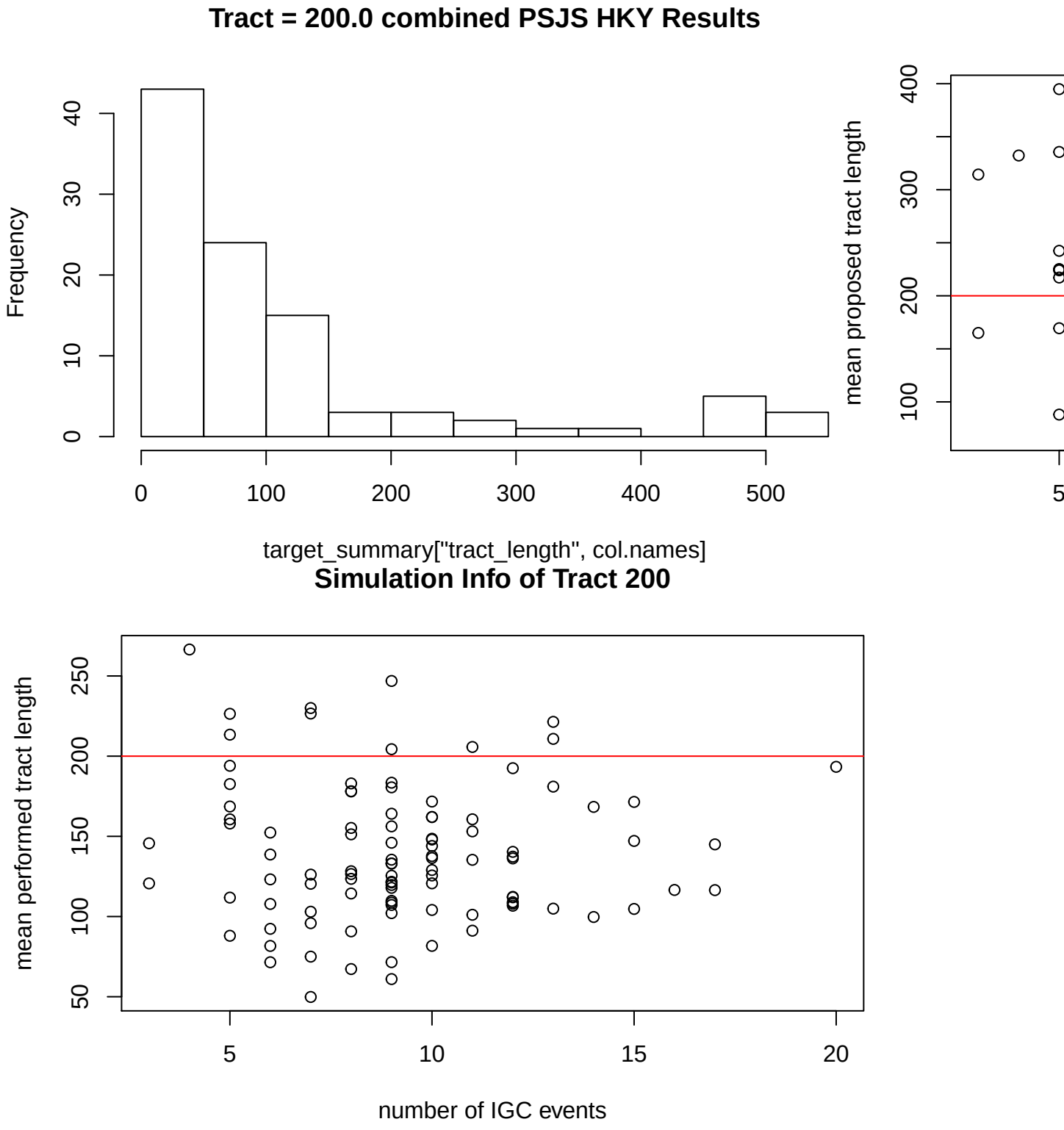

Simulation Info of Tract 100
mean performed tract length
40
60
80
100
120
5
10
15
20
25
number of IGC events

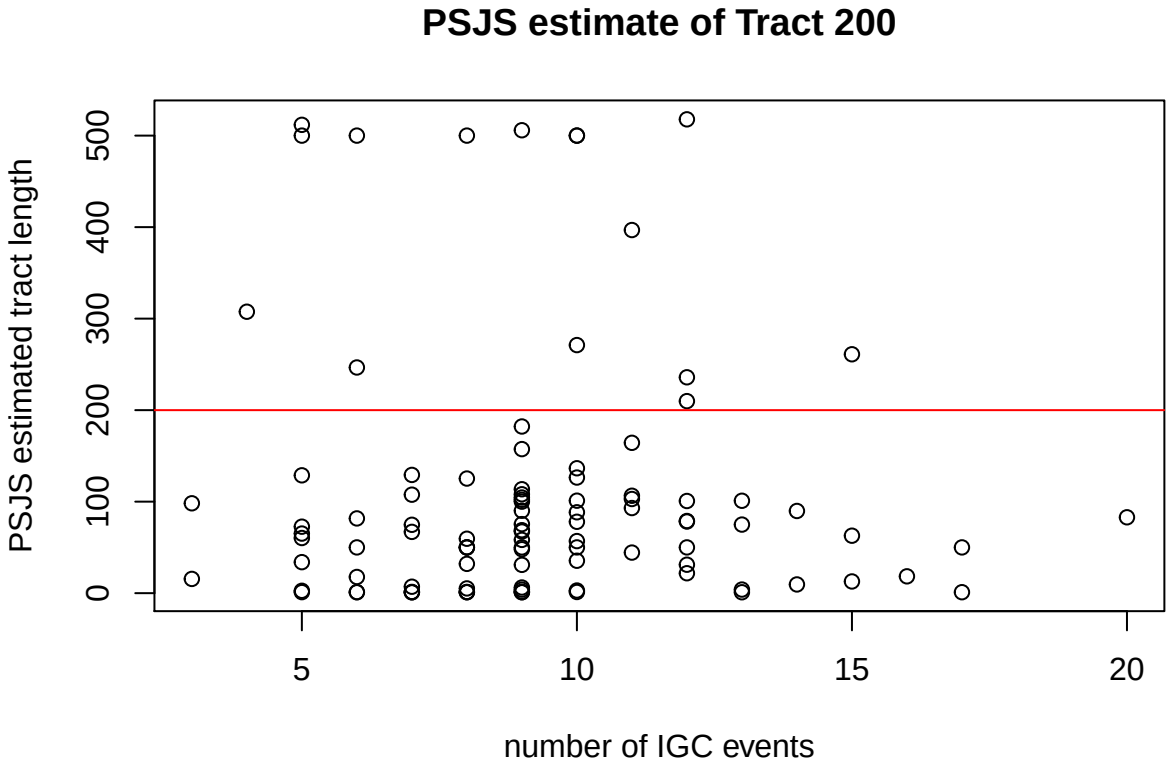

PSJS estimate of Tract 100
PSJS estimated tract length
0
100
200
300
400
500
5
10
15
20
25
number of IGC events

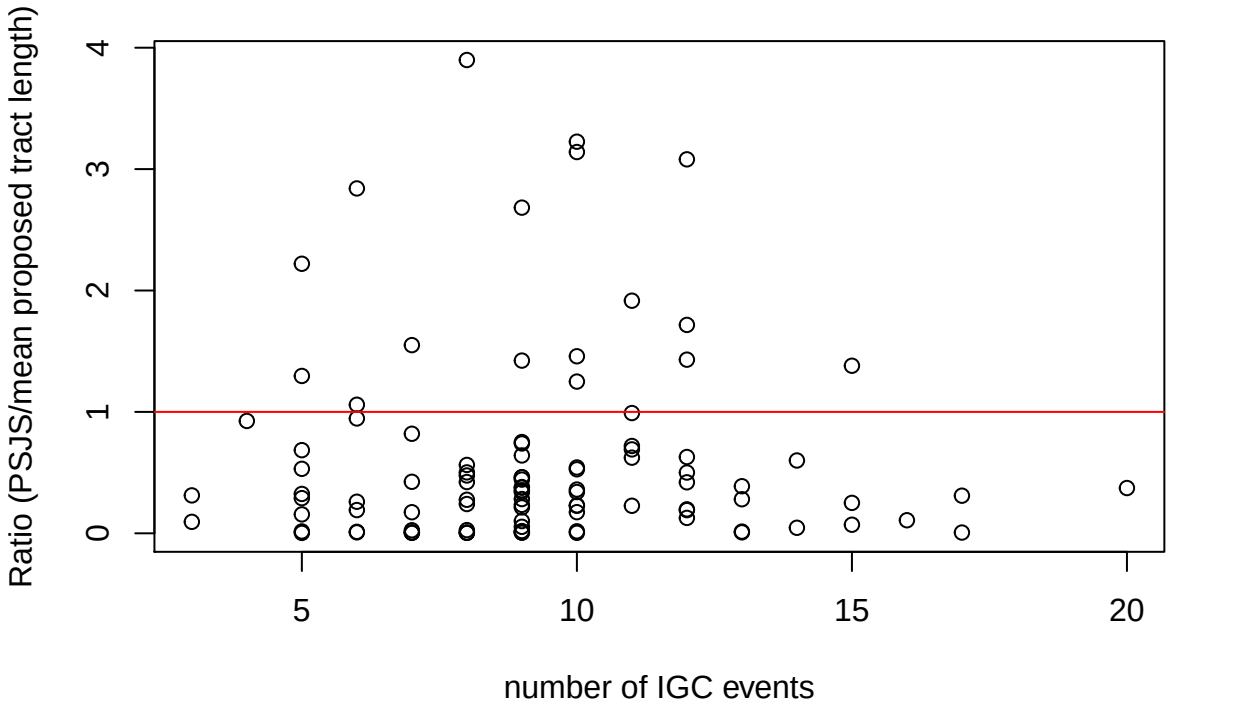


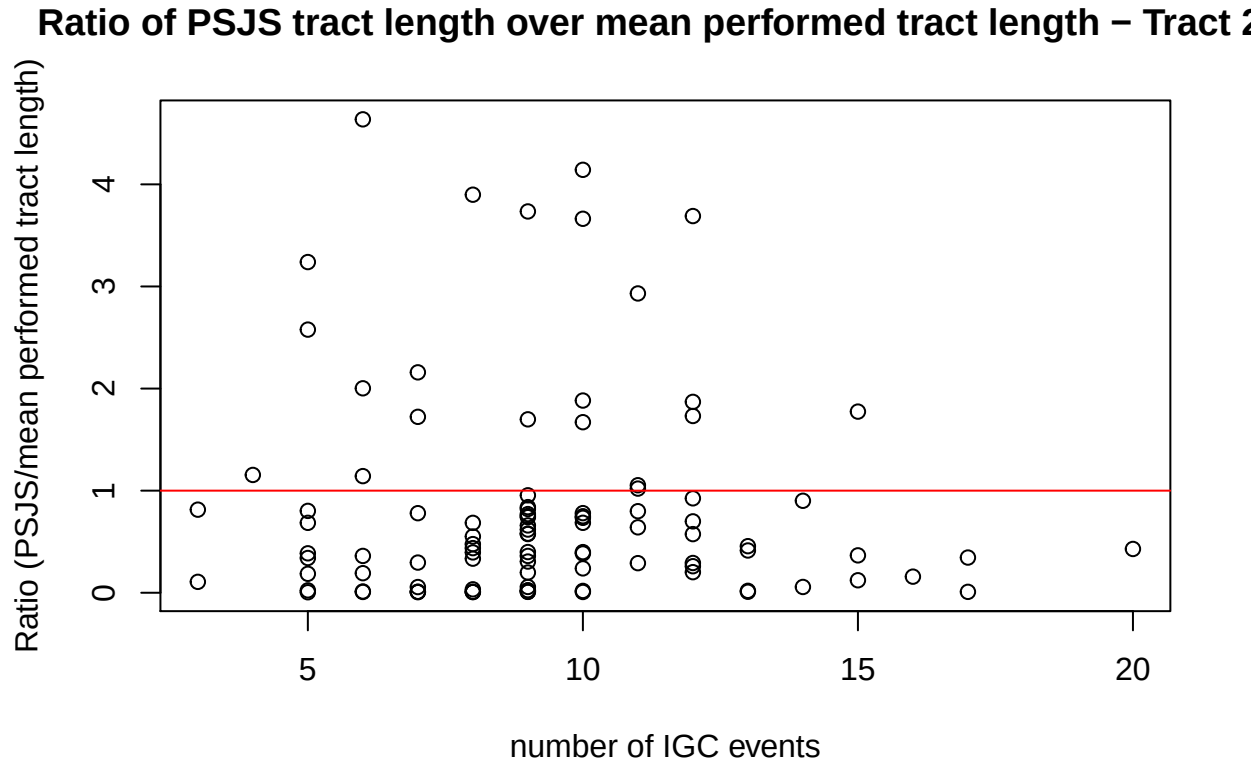



```
## [1] "Tract = 100.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "100"           "85.7367890914407" "122.076518037225"
```

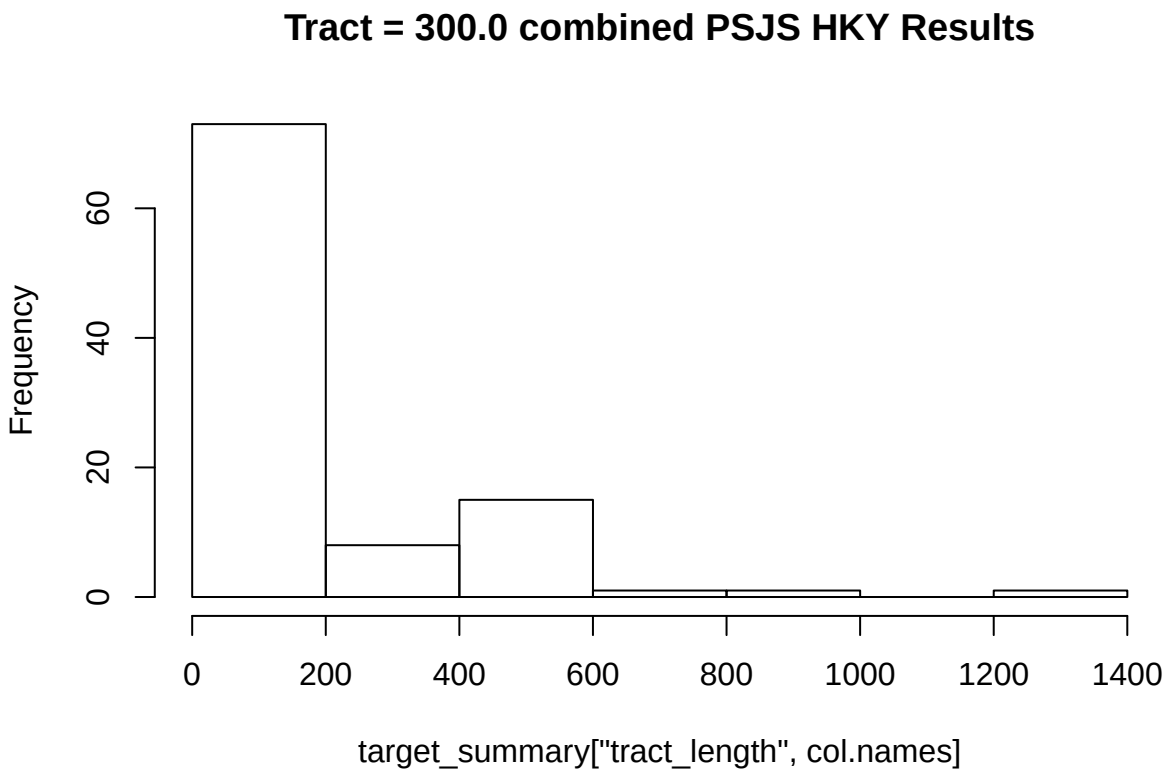

Tract = 200.0 combined PSJS HKY Results
Frequency
0
10
20
30
40
0
100
200
300
400
500
target_summary["tract_length", col.names]
mean proposed tract length
100
200
300
400
5
Simulation Info of Tract 200
mean performed tract length
50
100
150
200
250
5
10
15
20
number of IGC events

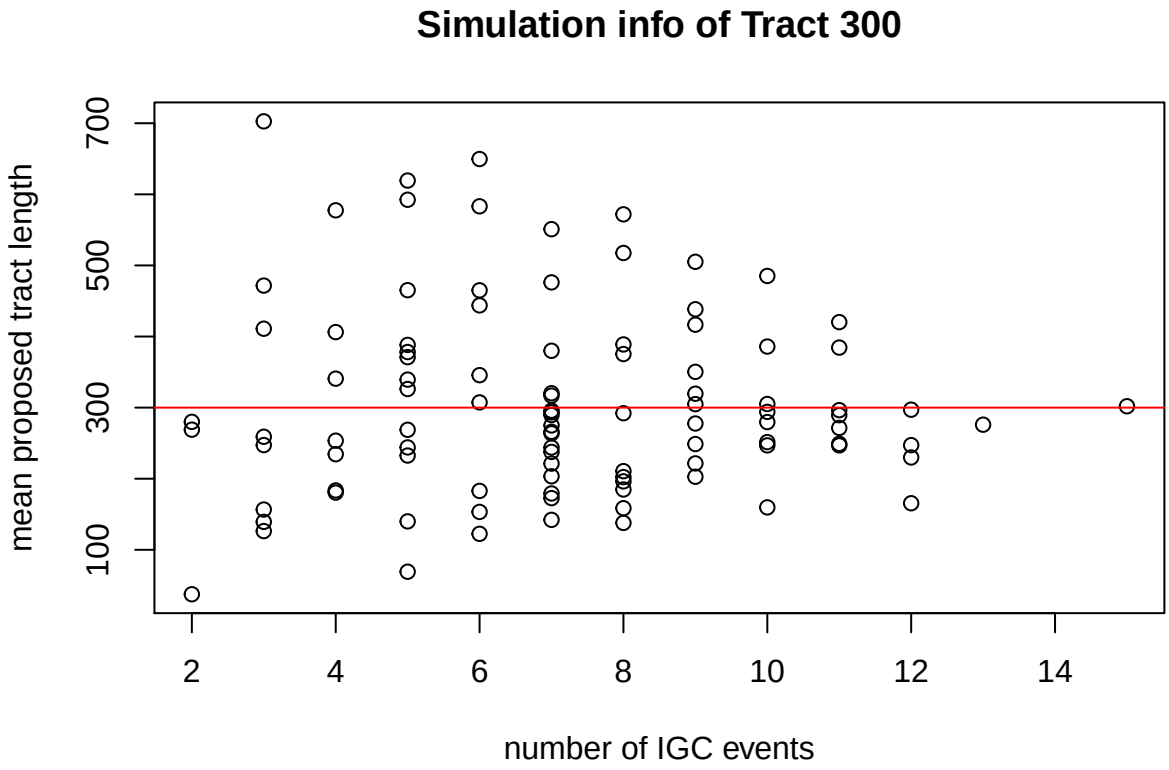

PSJS estimate of Tract 200
PSJS estimated tract length
0
100
200
300
400
500
5
10
15
20
number of IGC events

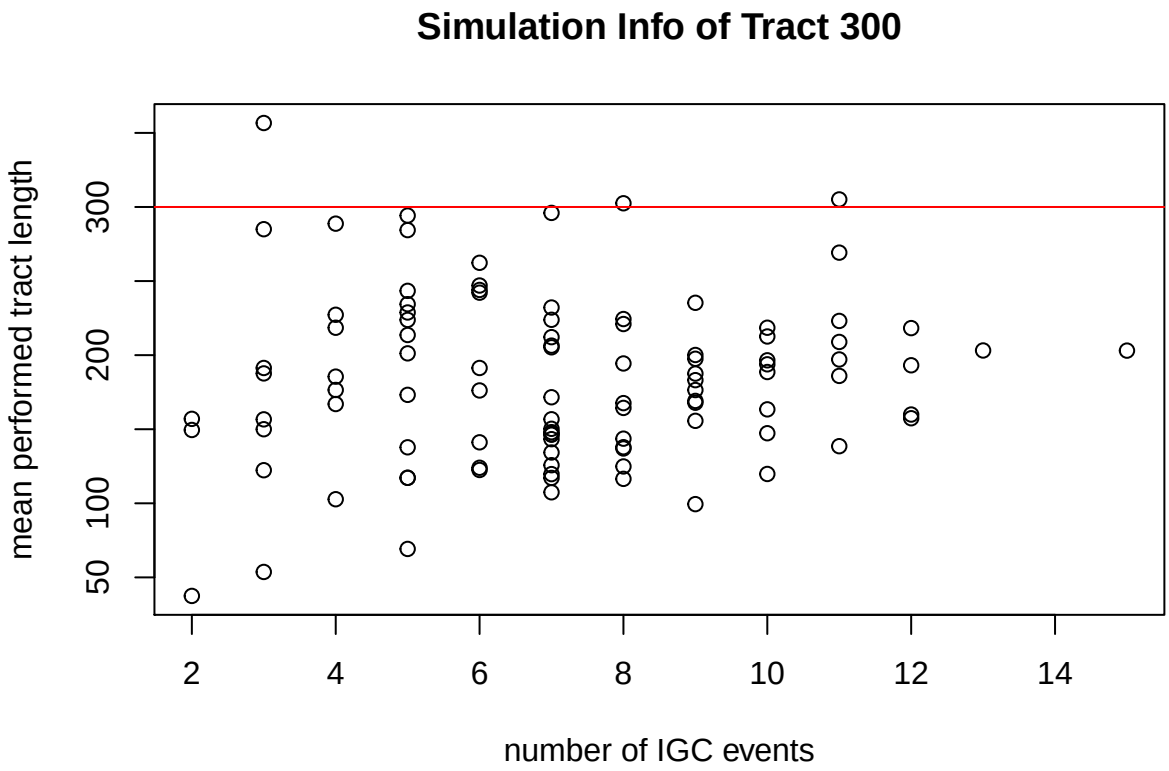

Ratio of PSJS tract length over mean proposed tract length − Tract 2
Ratio (PSJS/mean proposed tract length)
0
1
2
3
4
5
10
15
20
number of IGC events

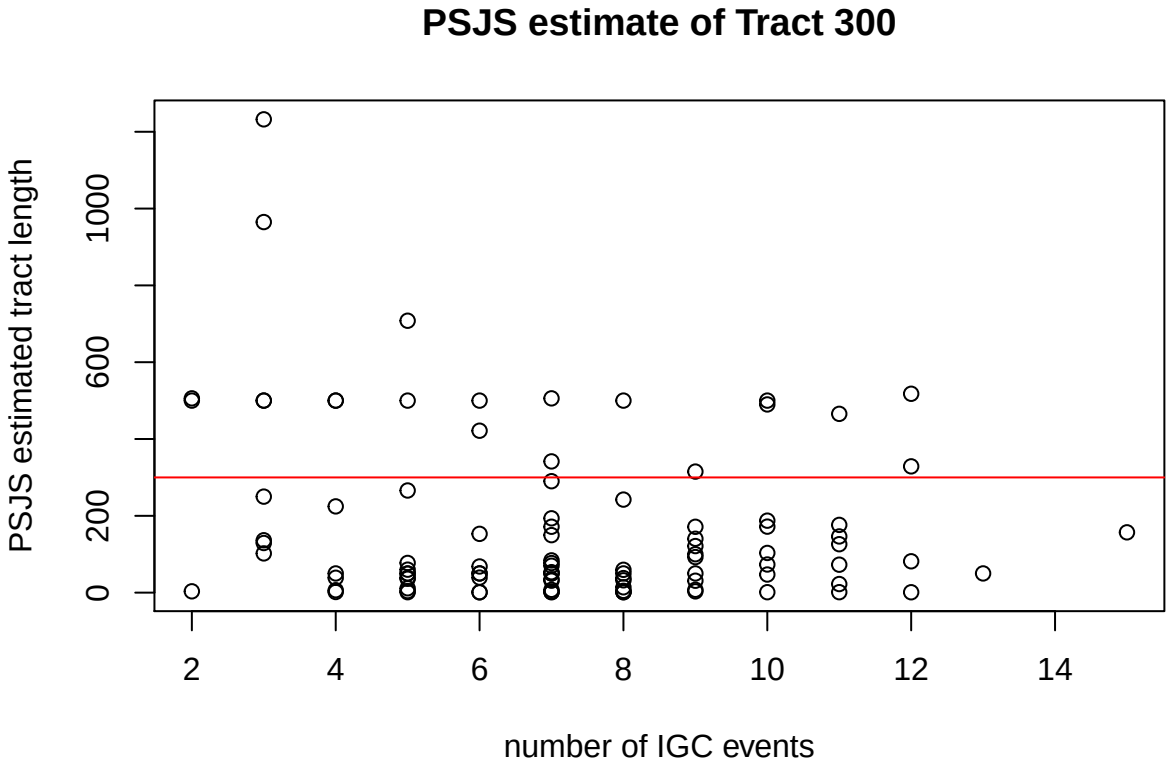


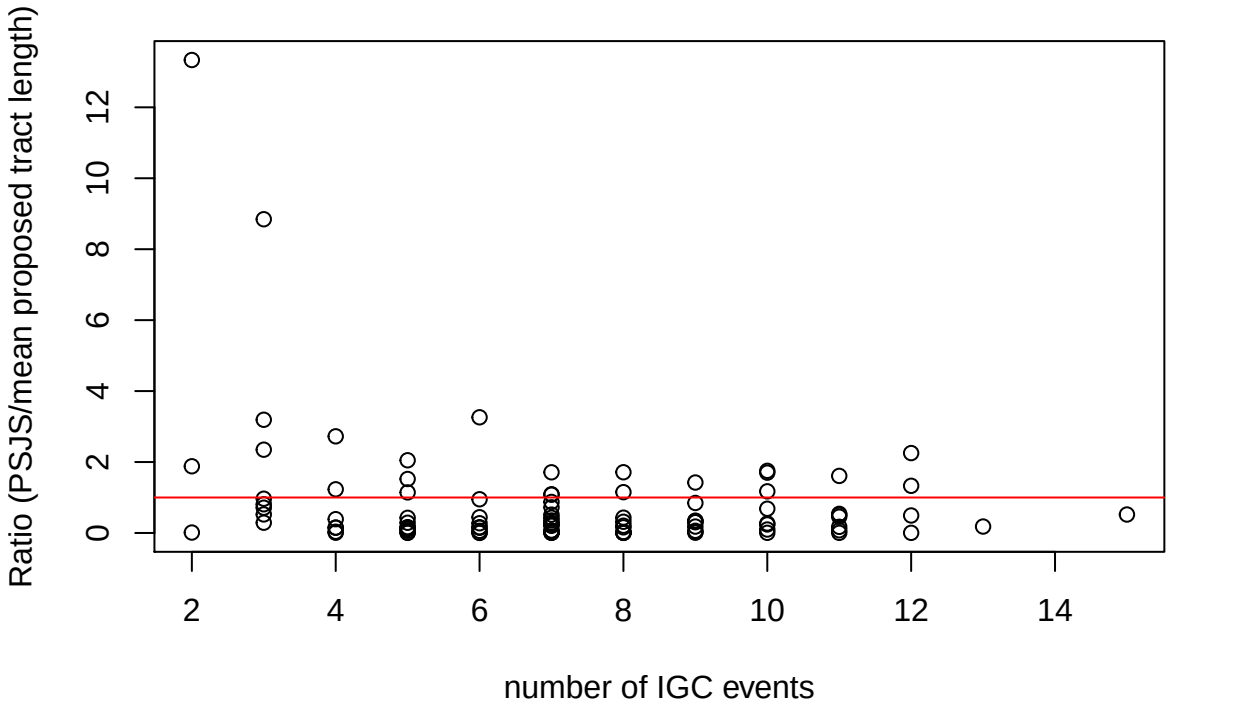

```
## [1] "Tract = 200.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "100"           "107.854006985987" "137.763222339365"
```

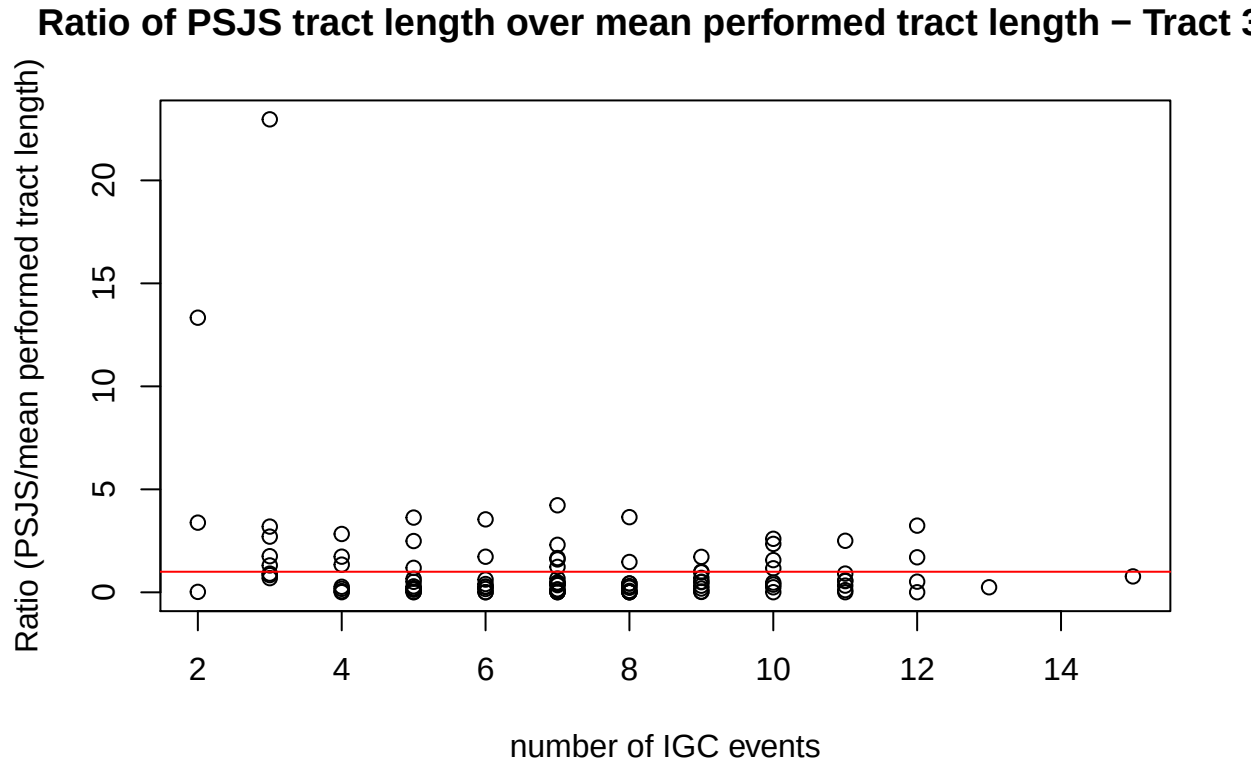



Simulation info of Tract 300
mean proposed tract length
100
300
500
700
2
4
6
8
10
12
14
number of IGC events

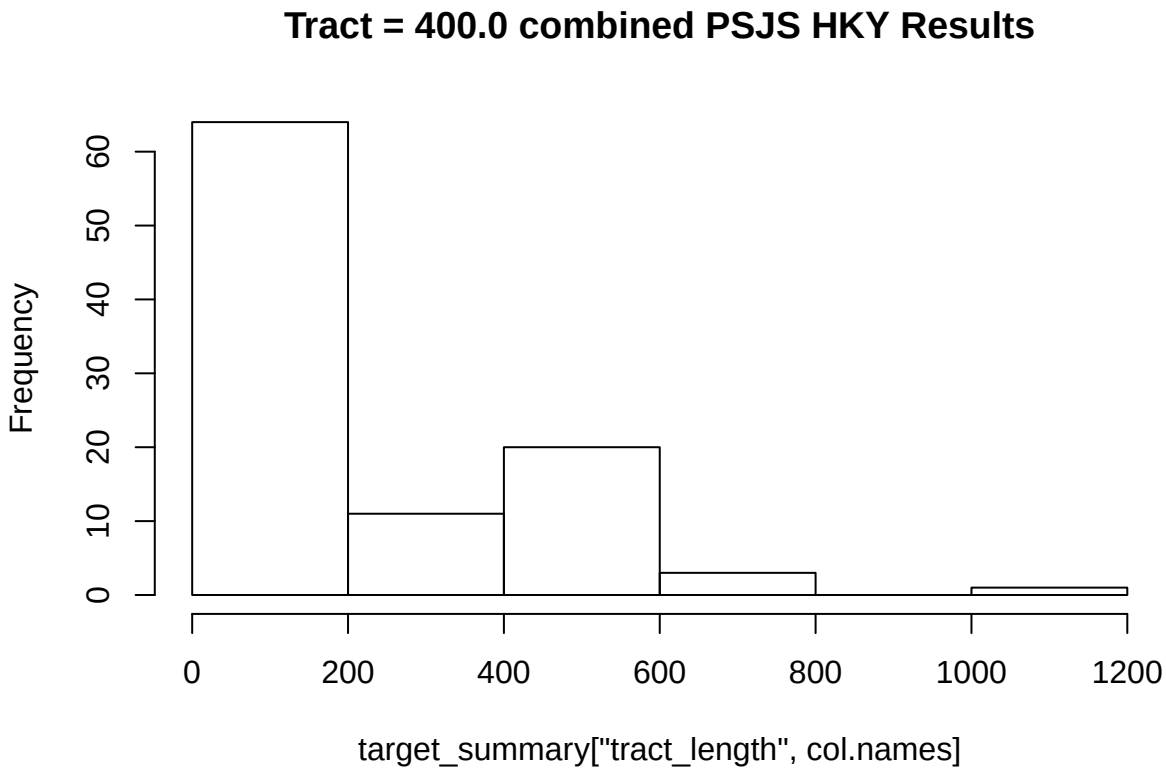

Simulation Info of Tract 300
mean performed tract length
50
100
200
300
2
4
6
8
10
12
14
number of IGC events

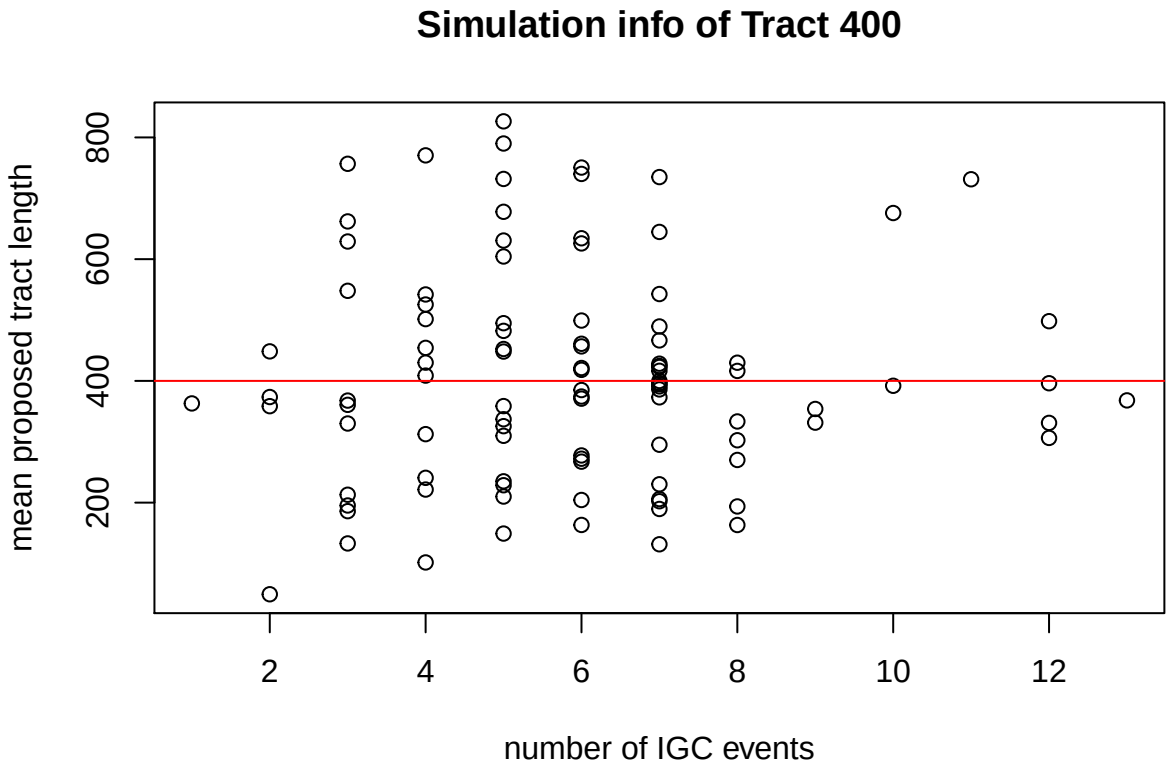

PSJS estimate of Tract 300
PSJS estimated tract length
0
200
600
1000
2
4
6
8
10
12
14
number of IGC events

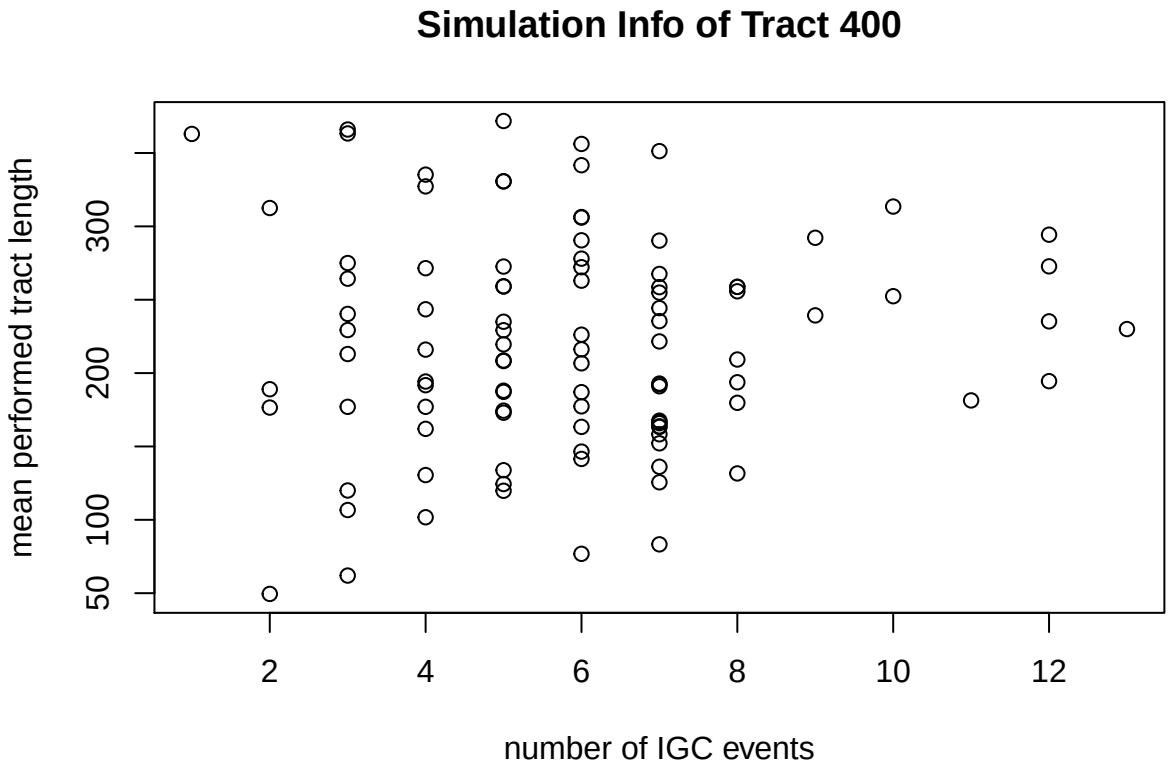

Ratio of PSJS tract length over mean proposed tract length − Tract 3
Ratio (PSJS/mean proposed tract length)
0
2
4
6
8
10
12
2
4
6
8
10
12
14
number of IGC events

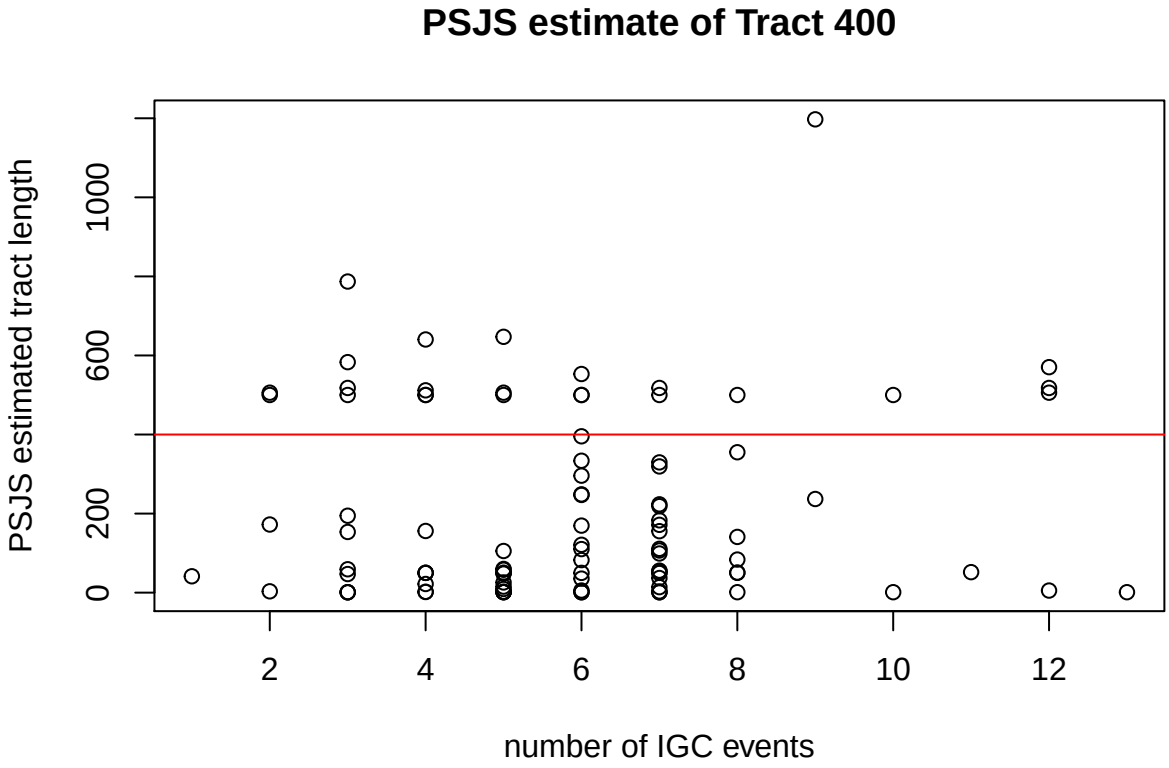


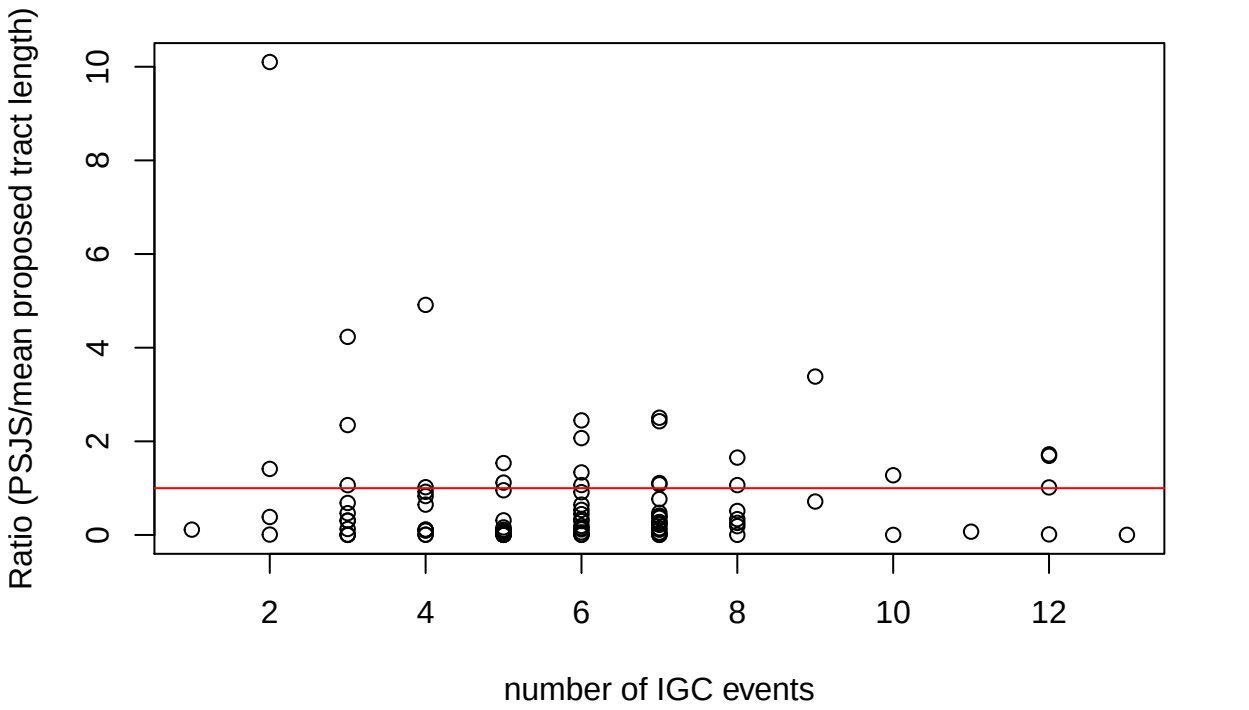

```
## [1] "Tract = 300.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "99"            "172.173660004663" "222.141070349374"
```

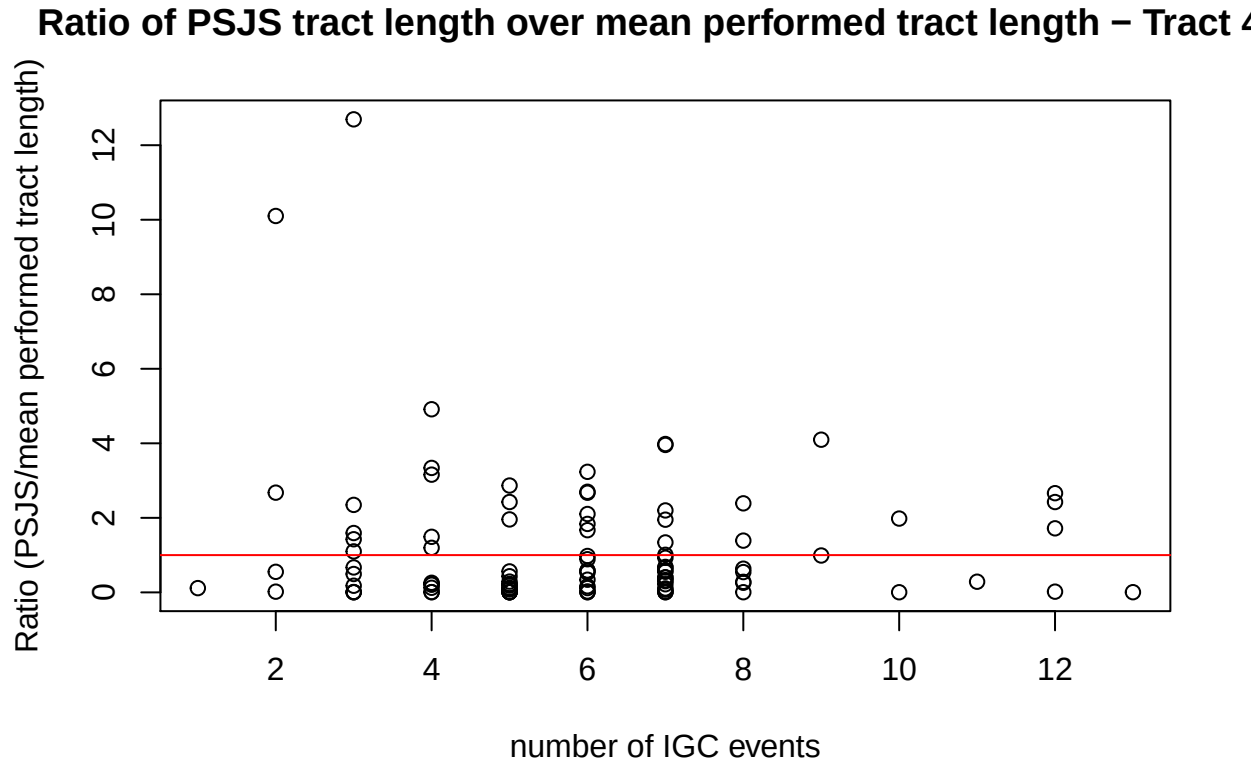



Simulation info of Tract 400
mean proposed tract length
200
400
600
800
2
4
6
8
10
12
number of IGC events

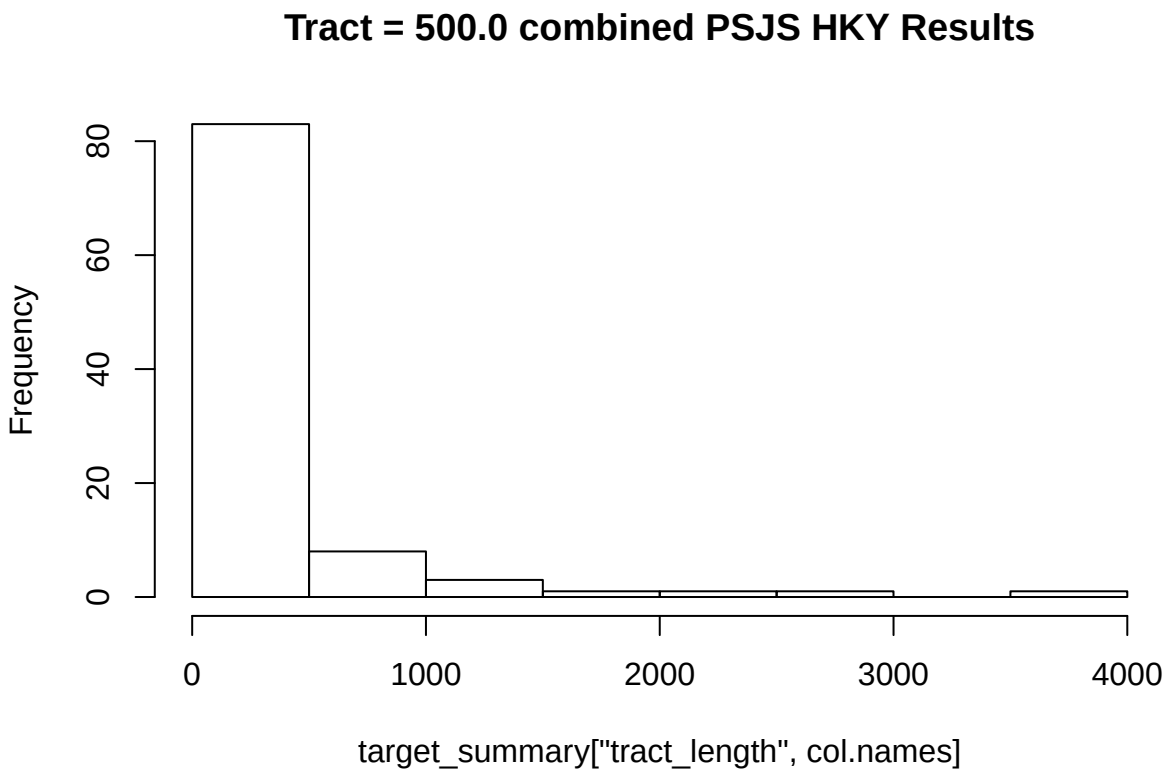

Simulation Info of Tract 400
mean performed tract length
50
100
200
300
2
4
6
8
10
12
number of IGC events

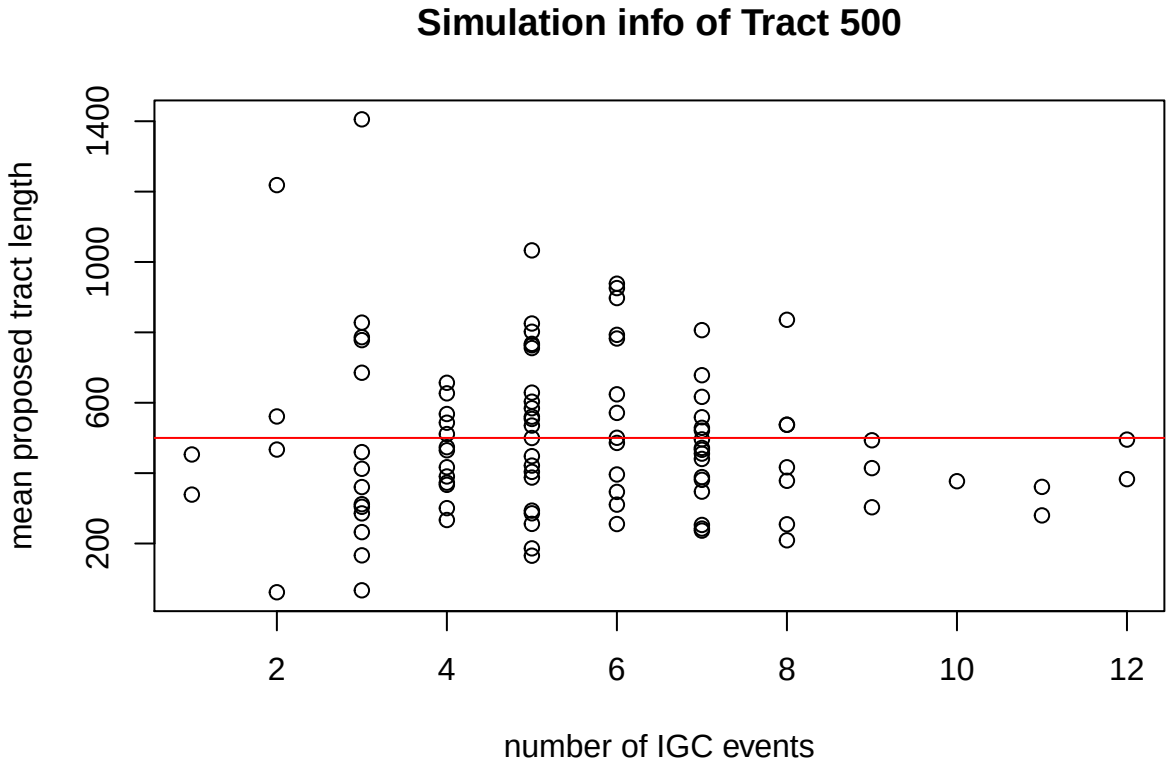

PSJS estimate of Tract 400
PSJS estimated tract length
0
200
600
1000
2
4
6
8
10
12
number of IGC events

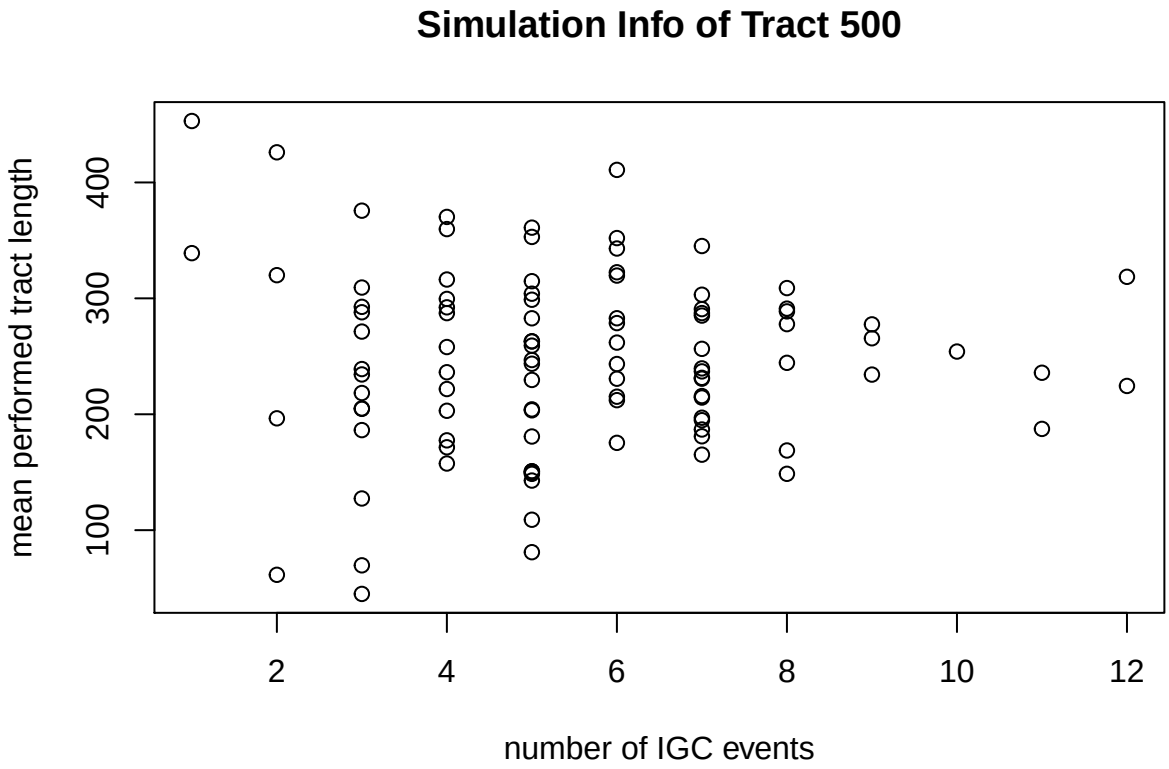

Ratio of PSJS tract length over mean proposed tract length − Tract 4
Ratio (PSJS/mean proposed tract length)
0
2
4
6
8
10
2
4
6
8
10
12
number of IGC events

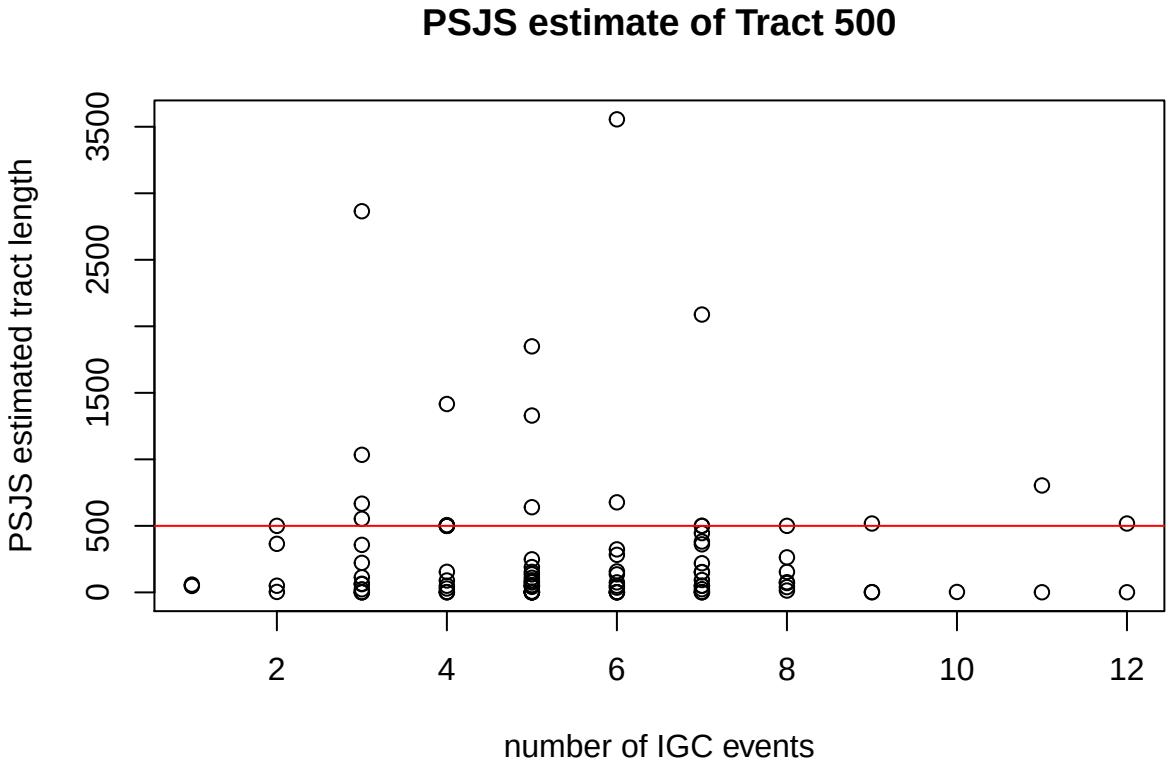


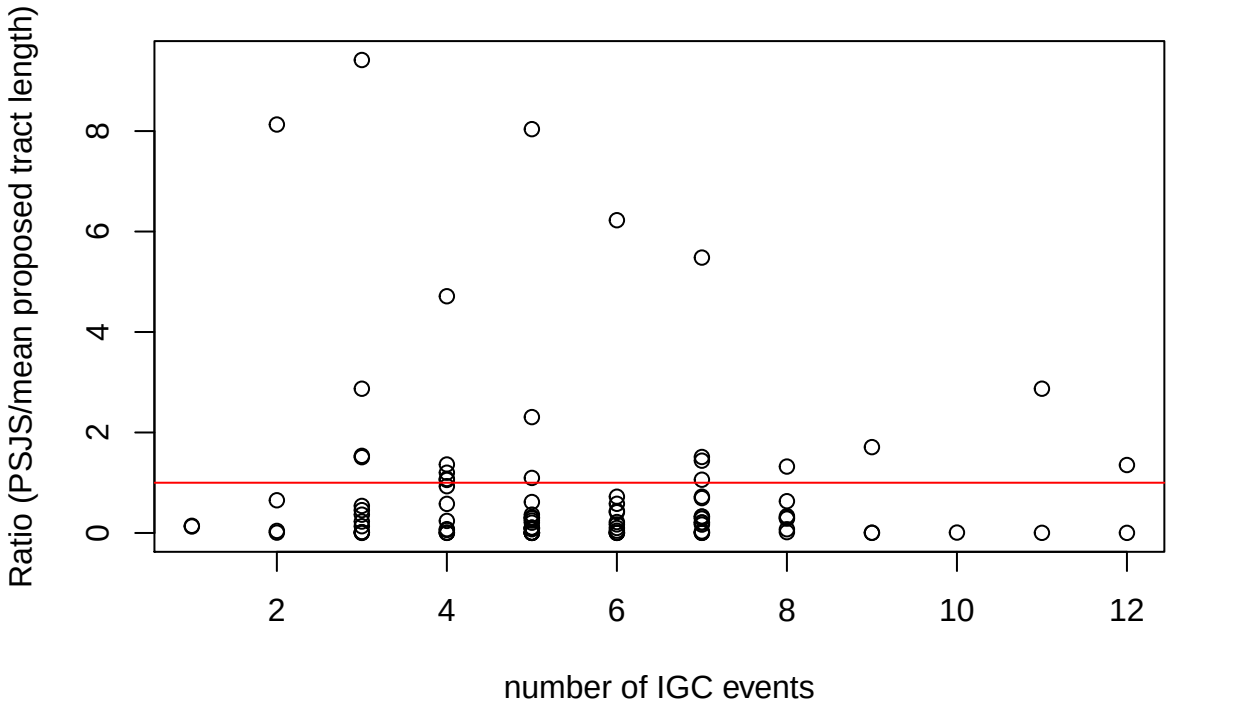

```
## [1] "Tract = 400.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "99"            "204.596413353198" "234.801497128067"
```

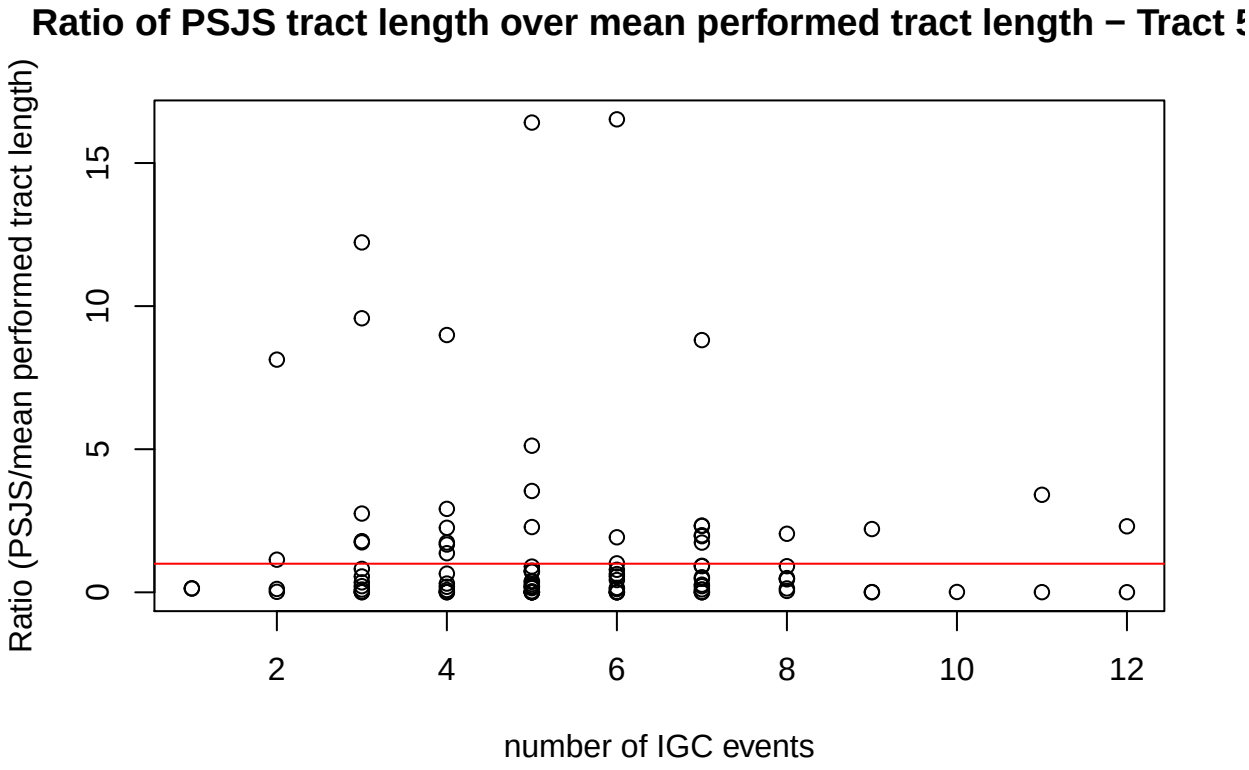



Simulation info of Tract 500
mean proposed tract length
200
600
1000
1400
2
4
6
8
10
12
number of IGC events


Simulation Info of Tract 500
mean performed tract length
100
200
300
400
2
4
6
8
10
12
number of IGC events

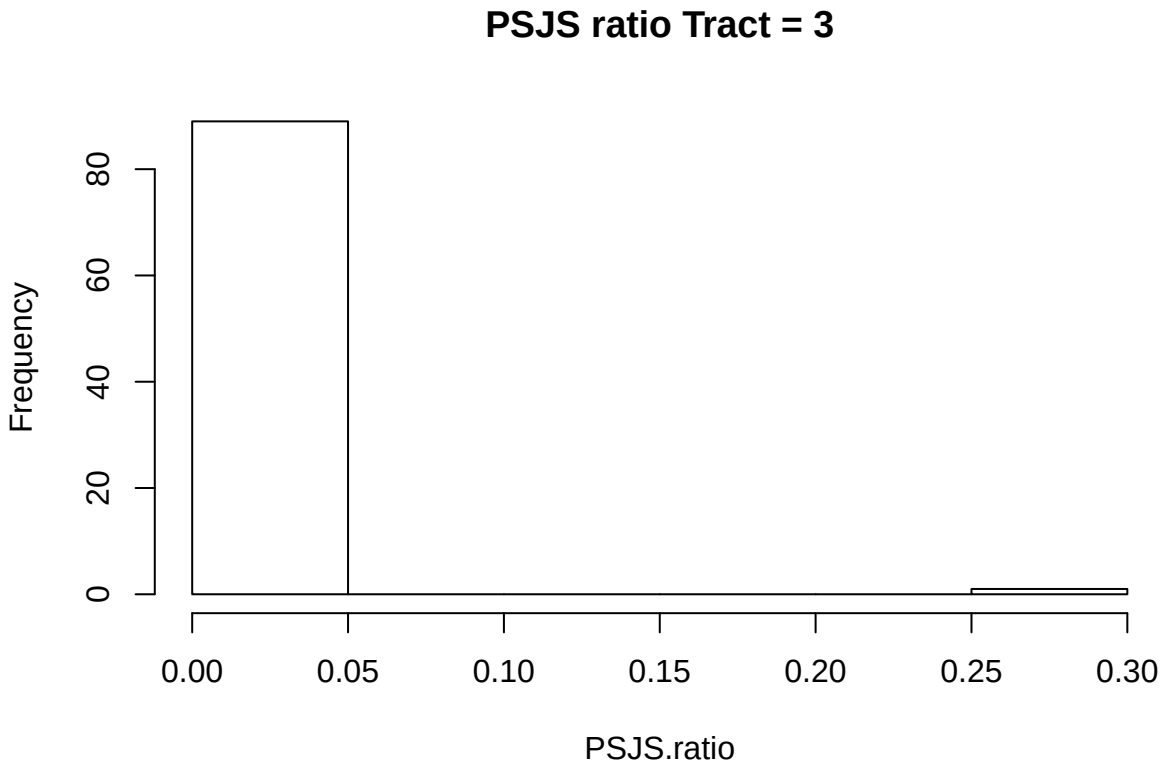

PSJS estimate of Tract 500
PSJS estimated tract length
0
500
1500
2500
3500
2
4
6
8
10
12
number of IGC events


Ratio of PSJS tract length over mean proposed tract length − Tract 5
Ratio (PSJS/mean proposed tract length)
0
2
4
6
8
2
4
6
8
10
12
number of IGC events

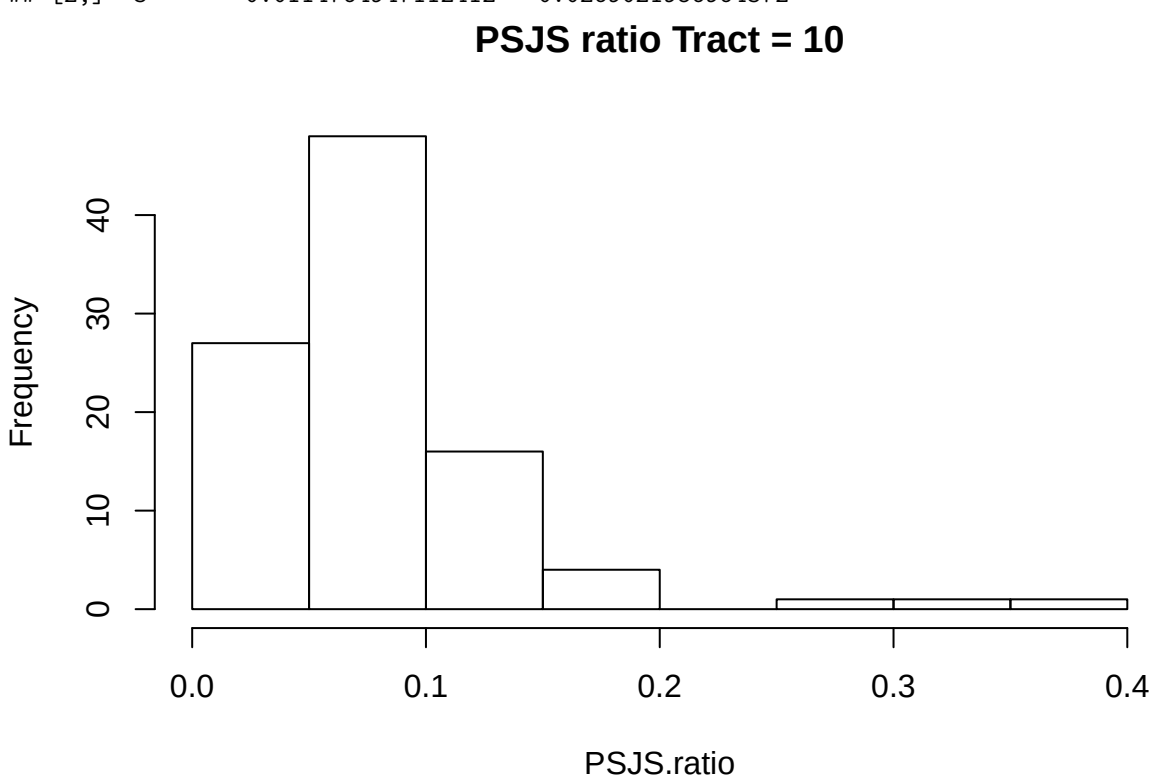



```
## [1] "Tract = 500.0 combined PSJS HKY Results"
##      [,1]            [,2]               [,3]
## [1,] "Total samples" "mean"             "sd"
## [2,] "98"            "302.363945646174" "565.81950081518"
```

```
for(tract in Tract.list){
  sim.info <- get(paste("sim.tract.", toString(tract), sep = ""))
  PSJS.info <- get(paste("PSJS_HKY_Tract_", toString(tract), "_combined_summary", sep = ""))
  shared.col <- colnames(PSJS.info)[PSJS.info["tract_length", ] < 10000 & PSJS.info["tract_length", ] >

  # Now show the ratio of PSJS estimated tract / actual mean tracts in simulation
  PSJS.ratio <- PSJS.info["tract_length", shared.col]/sim.info[1, shared.col]
  hist(PSJS.ratio, main = paste("PSJS ratio Tract = ", toString(tract), sep = ""))
  print(matrix(c("Tract", tract, "PSJS mean", mean(PSJS.ratio), "PSJS sd", sd(PSJS.ratio)), 2, 3))
}
```

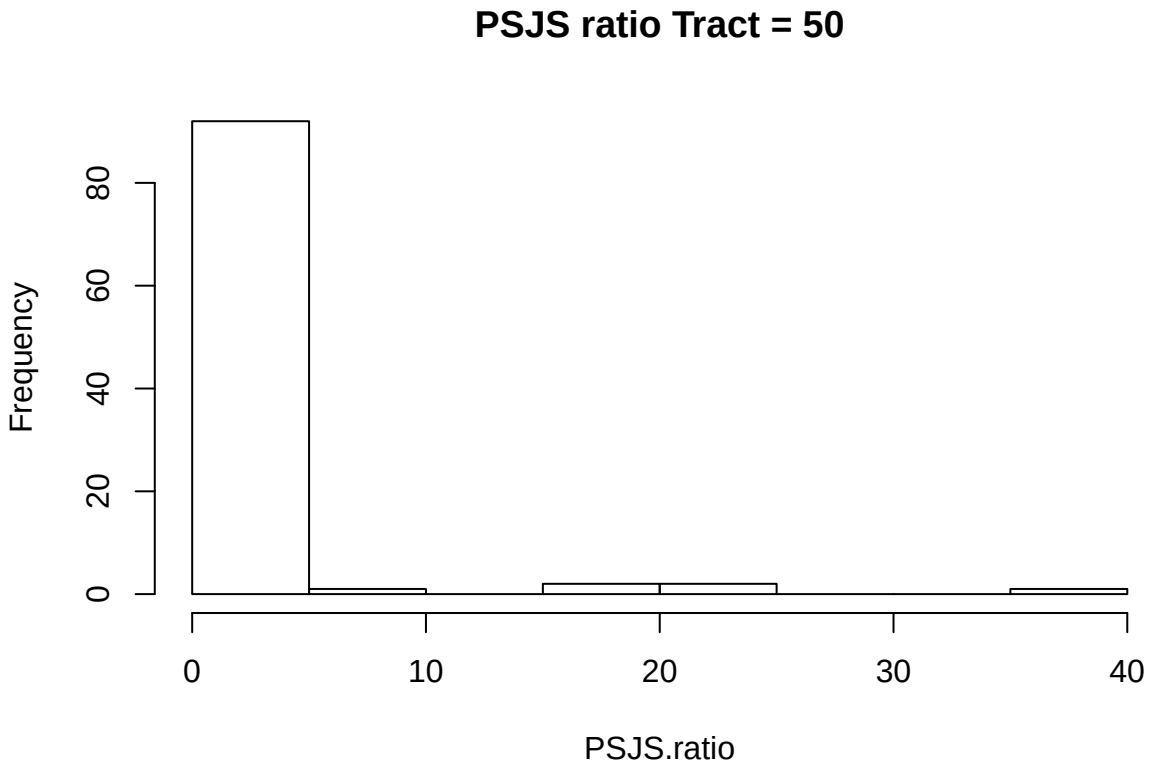



```
##      [,1]    [,2]                 [,3]
## [1,] "Tract" "PSJS mean"          "PSJS sd"
## [2,] "3"     "0.0114764947112412" "0.0269021956964372"
```

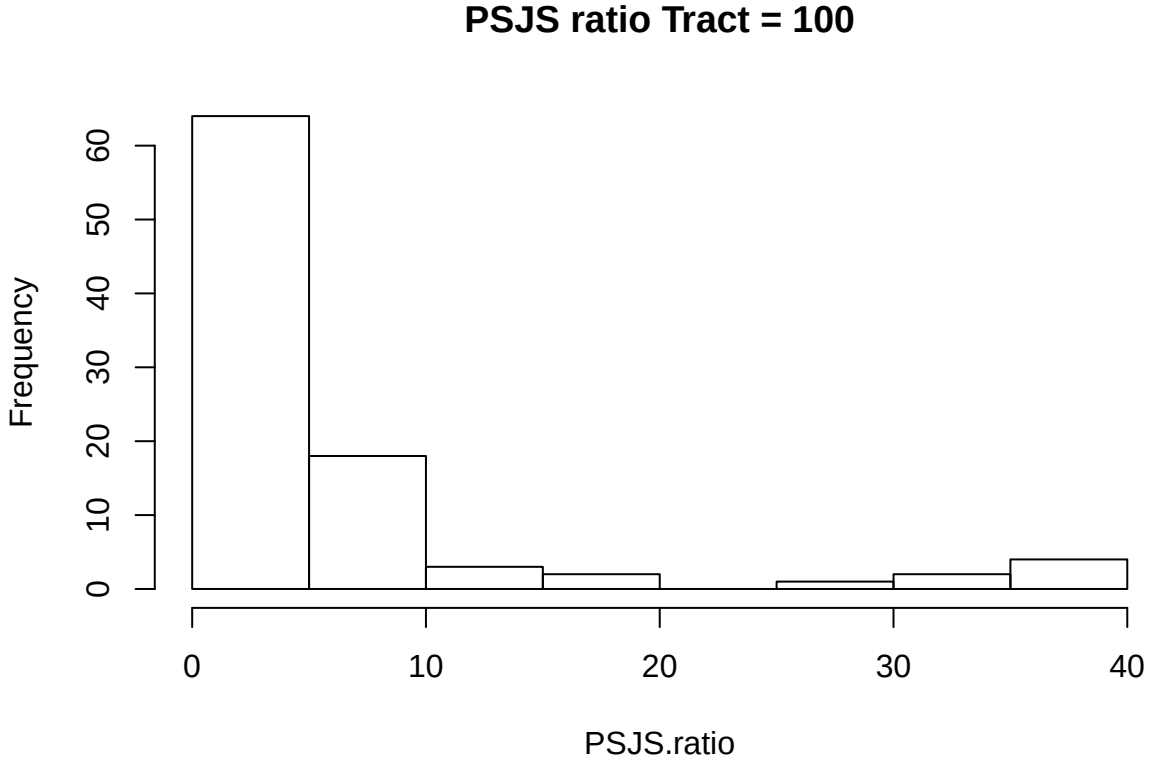



```
##      [,1]    [,2]                 [,3]
## [1,] "Tract" "PSJS mean"          "PSJS sd"
```

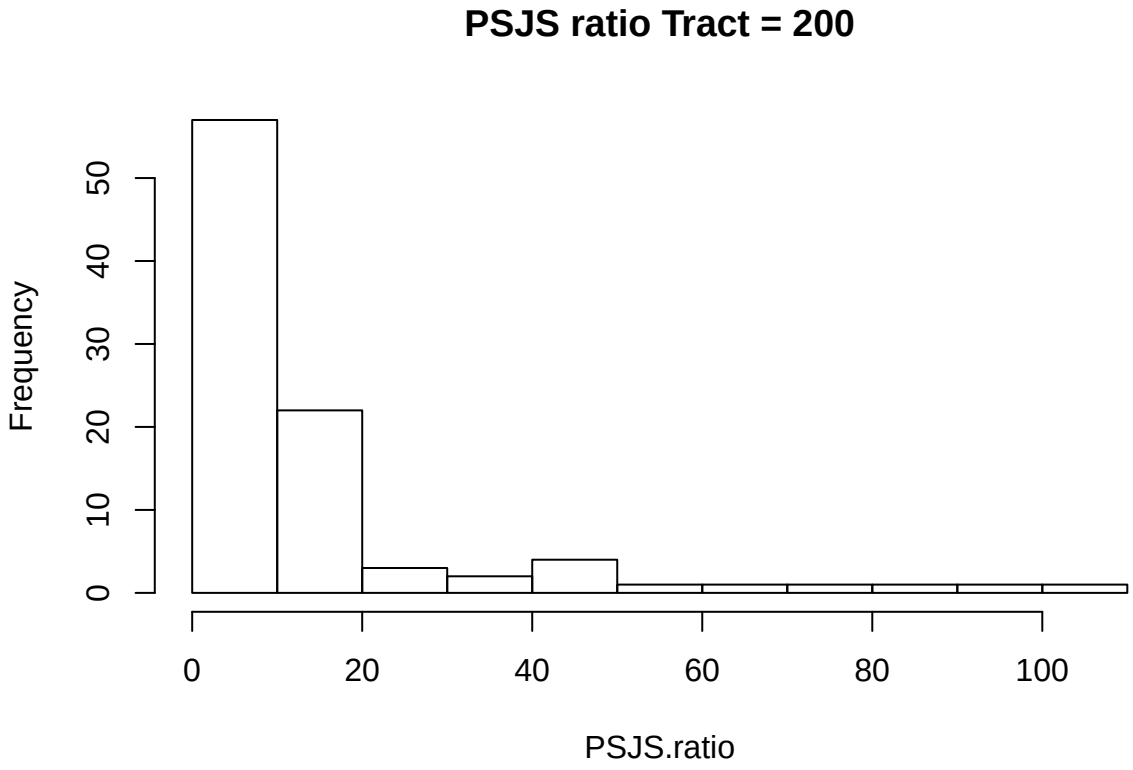
```
## [2,] "10"    "0.0784491503718006" "0.0568692826532482"
```



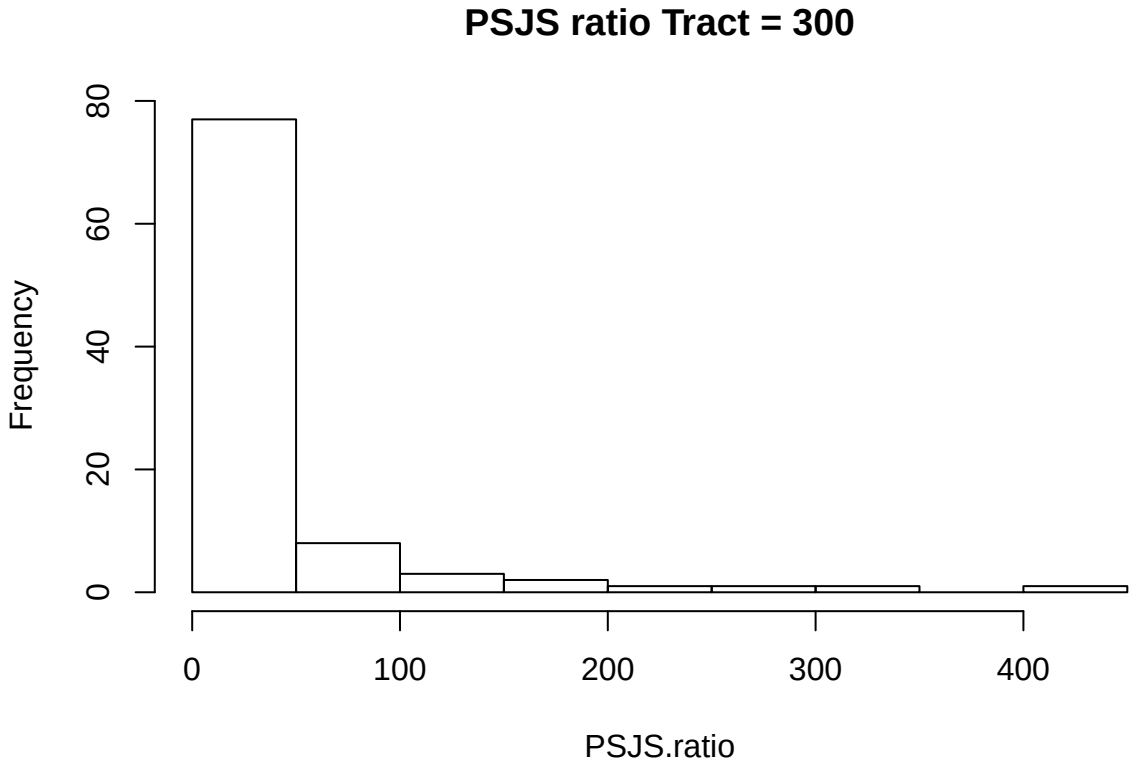

```
##      [,1]    [,2]               [,3]
## [1,] "Tract" "PSJS mean"        "PSJS sd"
## [2,] "50"    "2.48986487158233" "5.23413675636928"
```



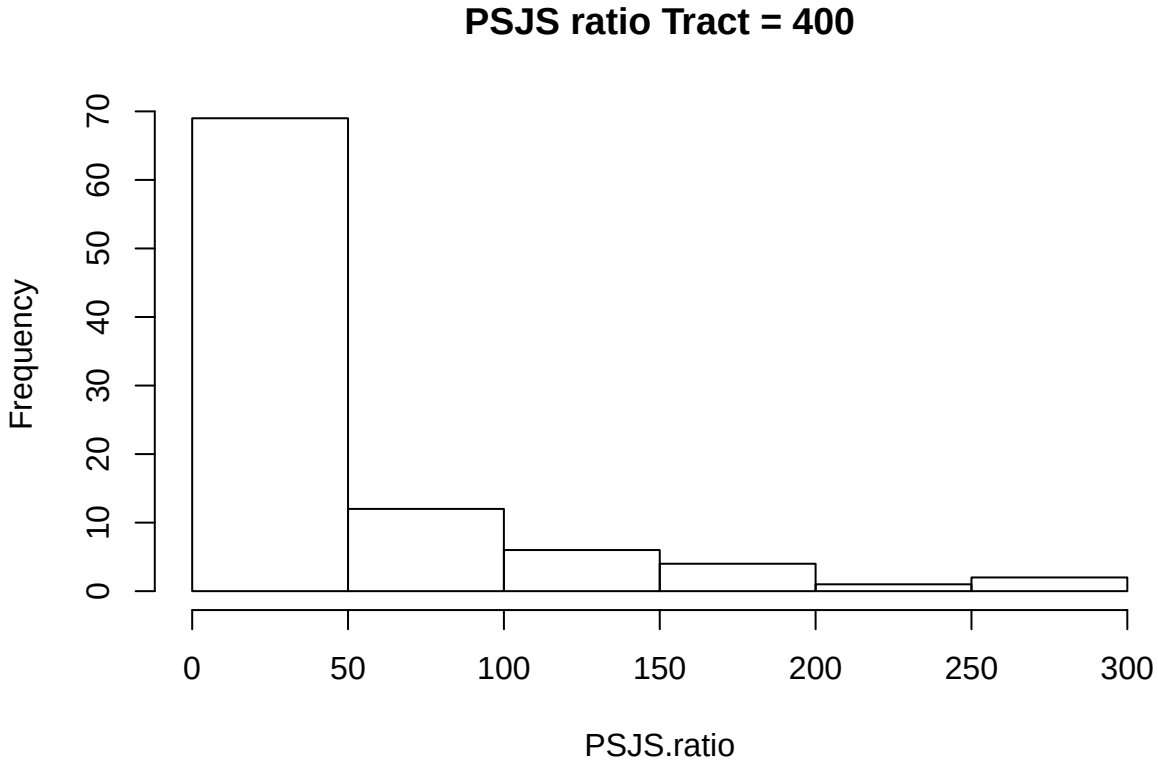
```
##      [,1]    [,2]               [,3]
## [1,] "Tract" "PSJS mean"        "PSJS sd"
## [2,] "100"   "6.51817135170871" "8.91221188549571"
```




```
##      [,1]    [,2]               [,3]
## [1,] "Tract" "PSJS mean"        "PSJS sd"
## [2,] "200"   "14.3815208218789" "20.6537471861418"
```

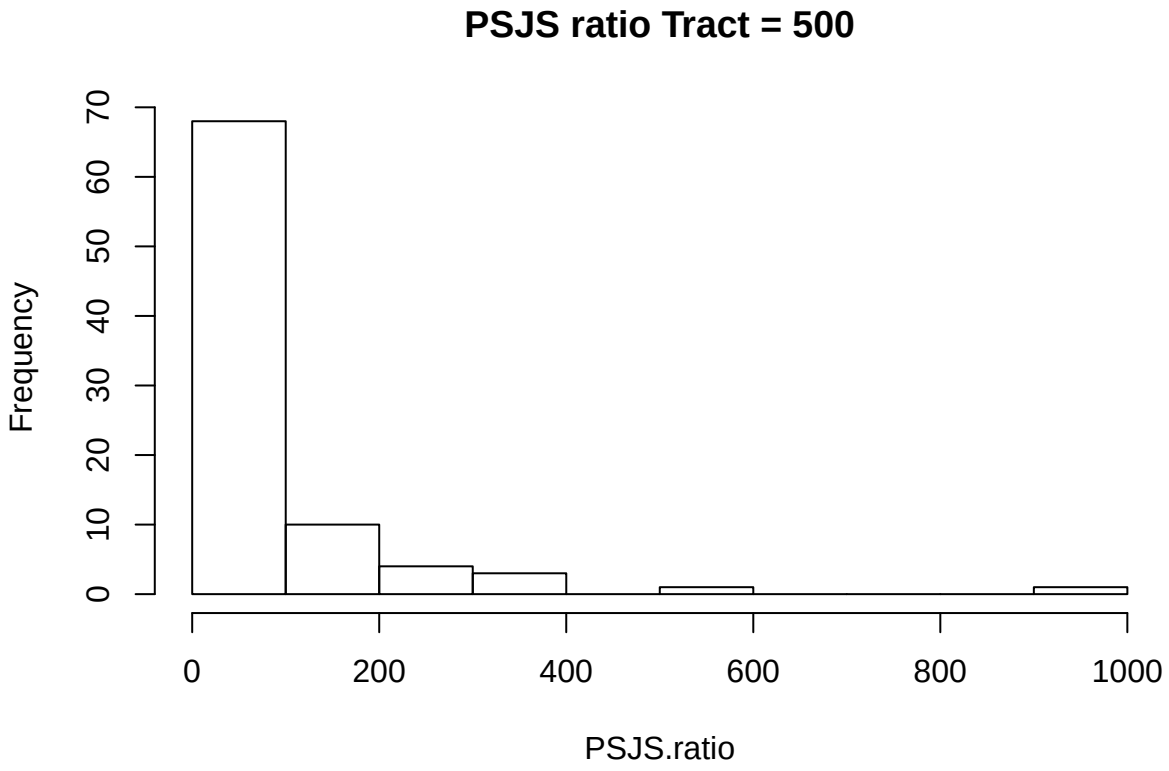



```
##      [,1]    [,2]               [,3]
## [1,] "Tract" "PSJS mean"        "PSJS sd"
## [2,] "300"   "37.8336862816968" "68.9900189948574"
```




```
##      [,1]    [,2]               [,3]
## [1,] "Tract" "PSJS mean"        "PSJS sd"
```

```
## [2,] "400"   "44.4451321387978" "59.1291162757734"
```




```
##      [,1]    [,2]               [,3]
## [1,] "Tract" "PSJS mean"        "PSJS sd"
## [2,] "500"   "75.2961192115813" "139.772577959362"
```

save workspace now

```
save.image("./SimulationStudy.RData")
```