week5 exercise

Xiang Li

2023-10-05

Exercise 2

```
x = c(3, 4, 6, 2, 4, 3, 8, 1, 3)
```

1

```
ma = function(x, t, k){
  s_t = max(c(t-k+1, 1))
  result = sum(x[s_t:t])/k
  return(result)
}
ma(x, 5, 3)
```

```
## [1] 4
```

```
ma(x, 2, 3)
```

```
## [1] 2.333333
```

2

See the solution of question 1.

3

```
for (t in 1:length(x)) {
  print(paste('The result of t=', t, ' is: ', ma(x, t, 3), sep = ''))
}
```

```
## [1] "The result of t=1 is: 1"
## [1] "The result of t=2 is: 2.33333333333333"
## [1] "The result of t=3 is: 4.33333333333333"
## [1] "The result of t=4 is: 4"
## [1] "The result of t=5 is: 4"
## [1] "The result of t=6 is: 3"
## [1] "The result of t=7 is: 5"
## [1] "The result of t=8 is: 4"
## [1] "The result of t=9 is: 4"
```

Exercise 3

```
x = c(3, 4, 6, 2, 4, 3, 8, 1, 3)
```

1

```
ma1 = function(x, t, k){
  if (t >= k) {
    result = mean(x[(t-k+1):t])
  }
  else {
    result = NA
  }
  return(result)
}
ma1(x, 5, 3)
```

```
## [1] 4
```

```
ma1(x, 2, 3)
```

```
## [1] NA
```

2

See the solution of question 1. ## 3

```
for (t in 1:length(x)) {
  print(paste('The result of t=', t, ' is: ', ma1(x, t, 3), sep = ''))
}
```

```
## [1] "The result of t=1 is: NA"
## [1] "The result of t=2 is: NA"
## [1] "The result of t=3 is: 4.33333333333333"
## [1] "The result of t=4 is: 4"
## [1] "The result of t=5 is: 4"
## [1] "The result of t=6 is: 3"
## [1] "The result of t=7 is: 5"
## [1] "The result of t=8 is: 4"
## [1] "The result of t=9 is: 4"
```

Exercise 4

```r
set.seed(4)
r = 100
A = matrix(NA, nrow = r, ncol = 5)
A[ , 1:2] = rpois(2*r, 6)
A[ , 3] = rgamma(r, shape = 2, scale = 1)
A[ , 4:5] = rbinom(r, 2, 0.6)
```

1

```r
for (c in 1:ncol(A)) {
  print(paste('The mean of column ', c, ' is: ', mean(A[, c]), sep = ''))
}
```

```
## [1] "The mean of column 1 is: 6.38"
## [1] "The mean of column 2 is: 5.85"
## [1] "The mean of column 3 is: 1.99823933401354"
## [1] "The mean of column 4 is: 1.25"
## [1] "The mean of column 5 is: 1.25"
```

2

```r
mr = c(rep(NA, nrow(A)))
for (r in 1:nrow(A)) {
  mr[r] = mean(A[r, ])
}
mr
```

```
##   [1] 3.291189 2.272466 2.640902 3.111095 3.626159 2.785598 3.287181 3.515471
##   [9] 5.013893 3.110099 3.349365 3.179451 2.672720 3.795876 3.119570 3.667153
##  [17] 3.750258 3.420071 5.334524 2.823049 2.168831 4.313738 4.170264 3.922343
##  [25] 3.057442 5.011309 4.680214 3.940540 3.171850 3.249605 3.128821 2.071929
##  [33] 3.800002 2.472575 2.986846 4.915670 2.733382 3.562307 3.904345 1.419388
##  [41] 4.458485 3.205480 3.487084 3.835591 3.389630 2.832274 3.915139 4.057130
##  [49] 3.440784 2.912621 2.545943 2.636555 5.399106 4.273108 3.336074 3.208792
##  [57] 2.317662 2.400785 3.434188 3.591113 4.200431 3.417701 3.966343 3.057654
##  [65] 3.173224 3.525466 1.651295 2.847296 3.137991 4.321664 2.176795 4.481994
##  [73] 2.871325 3.040652 2.476331 2.141397 3.193488 3.632953 2.846659 3.463482
##  [81] 3.393715 3.698968 3.409908 4.833123 3.480056 1.833386 2.970404 4.638822
##  [89] 3.500193 5.074874 2.475508 3.390183 4.076886 1.689733 2.497737 3.116681
##  [97] 3.756823 3.038489 2.319318 2.620802
```

3

```r
my_colMeans = function(mat) {
  result = c(rep(NA, ncol(mat)))
  for (c in 1:ncol(mat)) {
    result[c] = mean(mat[ , c])
  }
  return(result)
}
my_colMeans(A)
```

```
## [1] 6.380000 5.850000 1.998239 1.250000 1.250000
```

```r
colMeans(A)
```

```
## [1] 6.380000 5.850000 1.998239 1.250000 1.250000
```

4

```r
my_colmeanNstd = function(mat) {
  nr = ncol(mat)
  result = data.frame(column=1:nr, mean=c(rep(NA, nr)), standard_deviation=c(rep(NA, nr)))
  for (n in 1:nr) {
    result[n, 'mean'] = mean(mat[,n])
    result[n, 'standard_deviation'] = sd(mat[,n])
  }
  return(result)
}
my_colmeanNstd(A)
```

```
##   column     mean standard_deviation
## 1      1 6.380000          2.5654267
## 2      2 5.850000          2.3926445
## 3      3 1.998239          1.7092407
## 4      4 1.250000          0.7159792
## 5      5 1.250000          0.7159792
```

Exercise 5

1

```r
setwd('/Users/macbookpro/Desktop/learning materials/Statistical Computing with R')
data_df = read.csv('data/irish_polls.csv')
```

2

```r
data_df[data_df == 'Not Available'] = NA
to_decimal = function(x) {
  return(sub('%', '', x))
}
data_df[, 10:21] = lapply(data_df[, 10:21], to_decimal)
head(data_df, 5)
```

```
##              Polling.Firm        Commissioners Fieldwork.Start Fieldwork.End
## 1          Ireland Thinks Irish Mail on Sunday      2021-09-18    2021-09-18
## 2                   Red C        Business Post      2021-09-03    2021-09-09
## 3 Behaviour and Attitudes     The Sunday Times      2021-08-26    2021-09-08
## 4          Ireland Thinks Irish Mail on Sunday      2021-08-21    2021-08-21
## 5          Ireland Thinks Irish Mail on Sunday      2021-07-17    2021-07-17
##      Scope Sample.Size Sample.Size.Qualification Participation Precision
## 1 National        1000                  Provided          <NA>        1%
## 2 National        1031                  Provided          <NA>        1%
## 3 National         922                  Provided          <NA>        1%
## 4 National        1203                  Provided          <NA>        1%
## 5 National        1001                  Provided          <NA>        1%
##   Fine.Gael Fianna.Fáil Sinn.Féin Labour.Party Solidarity.People.Before.Profit
## 1        23          19        29            4                               4
## 2        28          13        29            5                               3
## 3        23          21        33            5                               2
## 4        24          15        30            6                               3
## 5        25          14        30            7                               4
##   Social.Democrats Green.Party Aontú Renua.Ireland Independent.Alliance
## 1                6           4     4          <NA>                 <NA>
## 2                5           4     2          <NA>                 <NA>
## 3                2           5     0          <NA>                 <NA>
## 4                6           4     3          <NA>                 <NA>
## 5                5           4     4          <NA>                 <NA>
##   Independents Other
## 1         <NA>  <NA>
## 2           10  <NA>
## 3            8  <NA>
## 4            9  <NA>
## 5            7  <NA>
```

3

```r
data_df[, 10:21] = as.numeric(unlist(data_df[, 10:21]))/100
head(data_df, 5)
```

```
##              Polling.Firm        Commissioners Fieldwork.Start Fieldwork.End
## 1          Ireland Thinks Irish Mail on Sunday      2021-09-18    2021-09-18
## 2                   Red C        Business Post      2021-09-03    2021-09-09
## 3 Behaviour and Attitudes     The Sunday Times      2021-08-26    2021-09-08
## 4          Ireland Thinks Irish Mail on Sunday      2021-08-21    2021-08-21
## 5          Ireland Thinks Irish Mail on Sunday      2021-07-17    2021-07-17
##      Scope Sample.Size Sample.Size.Qualification Participation Precision
## 1 National        1000                  Provided          <NA>        1%
## 2 National        1031                  Provided          <NA>        1%
## 3 National         922                  Provided          <NA>        1%
## 4 National        1203                  Provided          <NA>        1%
## 5 National        1001                  Provided          <NA>        1%
##   Fine.Gael Fianna.Fáil Sinn.Féin Labour.Party Solidarity.People.Before.Profit
## 1      0.23        0.19      0.29         0.04                            0.04
## 2      0.28        0.13      0.29         0.05                            0.03
## 3      0.23        0.21      0.33         0.05                            0.02
## 4      0.24        0.15      0.30         0.06                            0.03
## 5      0.25        0.14      0.30         0.07                            0.04
##   Social.Democrats Green.Party Aontú Renua.Ireland Independent.Alliance
## 1             0.06        0.04  0.04            NA                   NA
## 2             0.05        0.04  0.02            NA                   NA
## 3             0.02        0.05  0.00            NA                   NA
## 4             0.06        0.04  0.03            NA                   NA
## 5             0.05        0.04  0.04            NA                   NA
##   Independents Other
## 1           NA    NA
## 2         0.10    NA
## 3         0.08    NA
## 4         0.09    NA
## 5         0.07    NA
```

4

```
rec_data_df = data_df[1:10, ]
```

5

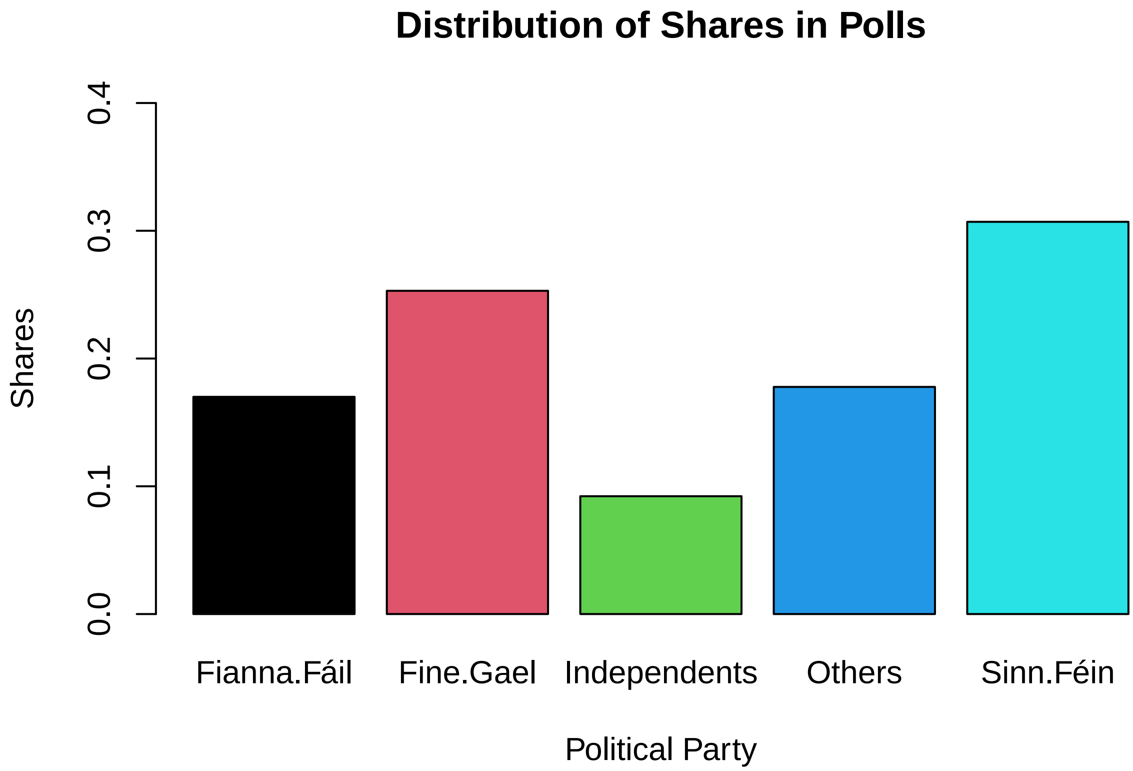
```
mean_v = colMeans(rec_data_df[, 10:20], na.rm = TRUE)
high_party = mean_v[(mean_v > 0.06) & (! is.na(mean_v))]
names(high_party)
```

```
## [1] "Fine.Gael"    "Fianna.Fáil"  "Sinn.Féin"    "Independents"
```

```
mean_df = data.frame(party=c(names(high_party), 'Others'), mean_shares=c(unname(high_party), 1-sum(high_
```

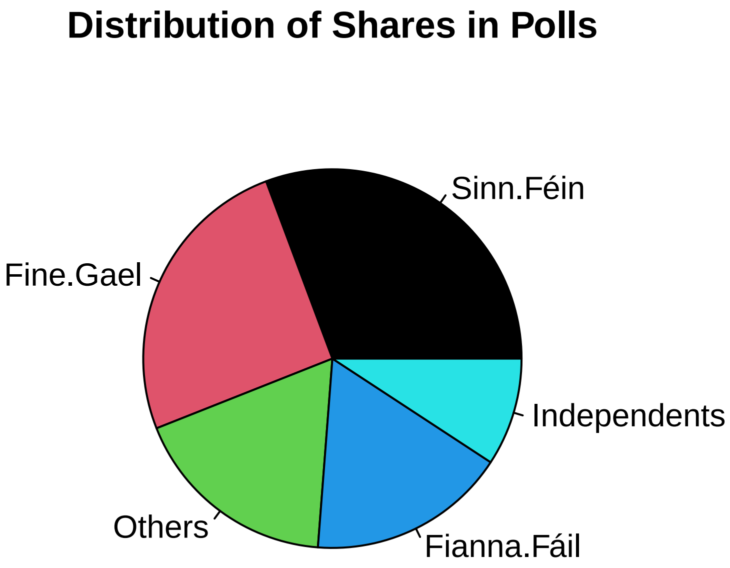
6

```
sorted_mean_df = mean_df[order(mean_df$mean_shares, decreasing = TRUE), ]
title = 'Distribution of Shares in Polls'
barplot(mean_shares~party, data = sorted_mean_df, col=1:nrow(sorted_mean_df), main = title, ylim = c(0,
```



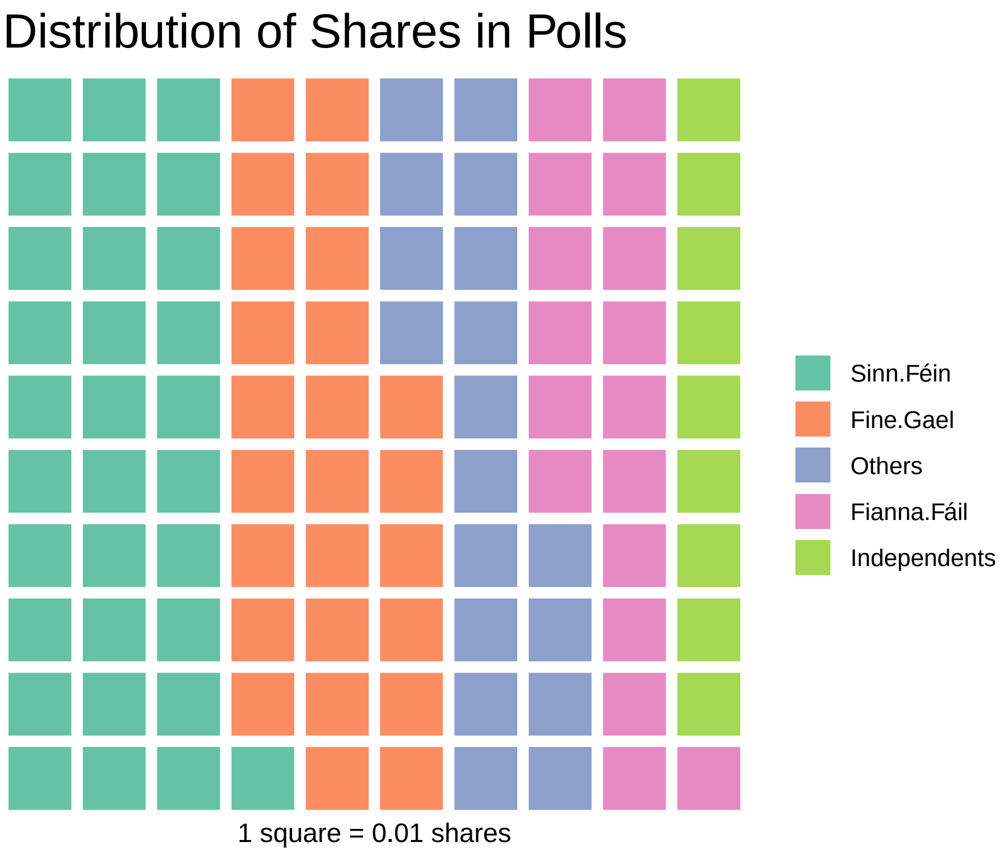
7

```
pie(sorted_mean_df$mean_shares, labels = sorted_mean_df$party, col = 1:5, main = title)
```



8

```
library(ggplot2)
library(waffle)
x = round(sorted_mean_df$mean_shares/0.01)
names(x) = sorted_mean_df$party
waffle(x, rows = 10, xlab = '1 square = 0.01 shares', title = title)
```



9

```
pdf(file="plot/week5_Exercise5_6.pdf")
barplot(mean_shares~party, data = sorted_mean_df, col=1:nrow(sorted_mean_df), main = title, ylim = c(0,
dev.off()
```

```
## pdf
##   2
```

```
pdf(file="plot/week5_Exercise5_7.pdf")
pie(sorted_mean_df$mean_shares, labels = sorted_mean_df$party, col = 1:5, main = title)
dev.off()
```

```
## pdf
##   2
```

```
pdf(file="plot/week5_Exercise5_8.pdf")
waffle(x, rows = 10, xlab = '1 square = 0.01 shares', title = title)
dev.off()
```

```
## pdf
##   2
```

10

```
jpeg(file="plot/week5_Exercise5_6.jpg", quality = 90)
barplot(mean_shares~party, data = sorted_mean_df, col=1:nrow(sorted_mean_df), main = title, ylim = c(0,
dev.off()
```

```
## pdf
##   2
```

```
jpeg(file="plot/week5_Exercise5_7.jpg", quality = 90)
pie(sorted_mean_df$mean_shares, labels = sorted_mean_df$party, col = 1:5, main = title)
dev.off()
```

```
## pdf
##   2
```

```
jpeg(file="plot/week5_Exercise5_8.jpg", quality = 90)
waffle(x, rows = 10, xlab = '1 square = 0.01 shares', title = title)
dev.off()
```

```
## pdf
##   2
```

Exercise 6

1

```
library(psych)
```

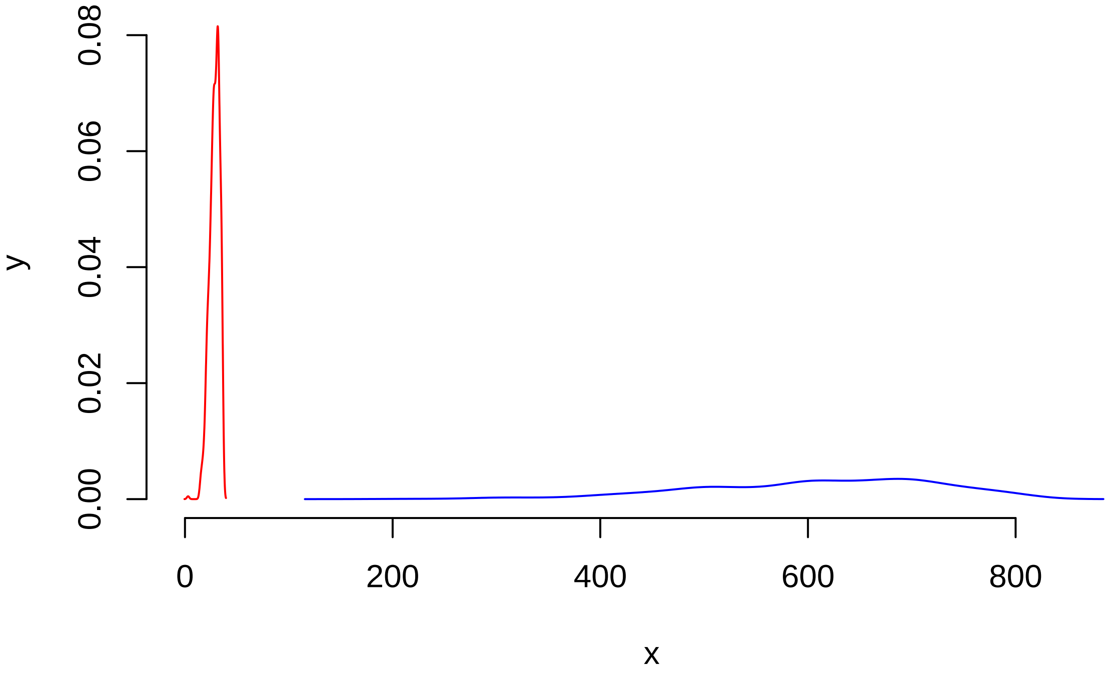
```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
data(sat.act)
act_df = as.data.frame(sat.act)
```

2

3/4

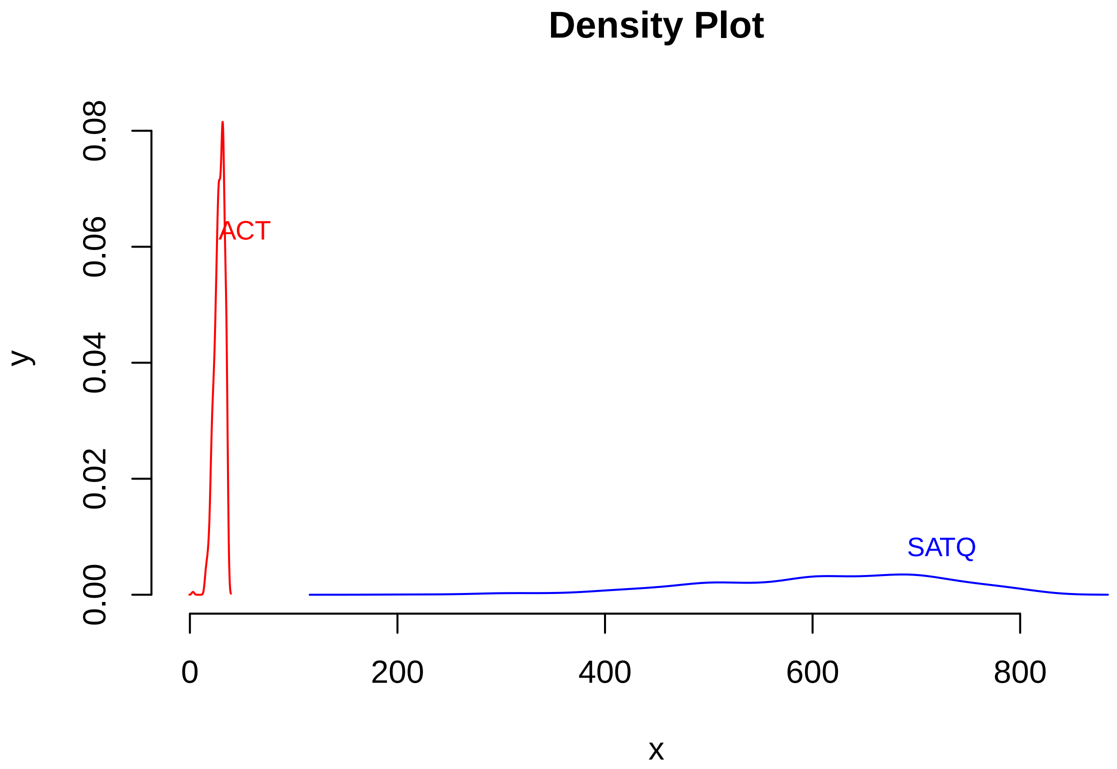
```
dens_ACT = density(act_df[(! is.na(act_df$ACT)), 'ACT'])
x_ACT = unname(unlist(dens_ACT[1]))
y_ACT = unname(unlist(dens_ACT[2]))
dens_SATQ = density(act_df[(! is.na(act_df$SATQ)), 'SATQ'])
x_SATQ = unname(unlist(dens_SATQ[1]))
y_SATQ = unname(unlist(dens_SATQ[2]))
plot(x_ACT, y_ACT, frame = FALSE, col = "red", type = 'l', xlab = "x", ylab = "y", xlim = c(-1, 900))
lines(x_SATQ, y_SATQ, col = "blue", type = 'l')
```



```
# legend("topright", legend=c("ACT", "SATQ"), col=c("red", "blue"), lty = c(1, 1), cex = 0.3, pt.cex =
```

5/6

```
plot(x_ACT, y_ACT, frame = FALSE, col = "red", type = 'l', xlab = "x", ylab = "y", xlim = c(-1, 900), ma
lines(x_SATQ, y_SATQ, col = "blue", type = 'l')
text(52.88124, 0.062899485, labels = 'ACT', col = "red", cex = 0.8)
text(724.40507, 0.008220942, labels = 'SATQ', col = "blue", cex = 0.8)
```


Density Plot
y
0.00
0.02
0.04
0.06
0.08
x
0
200
400
600
800
ACT
SATQ