age_distribution_of_cases

Maggie Walters

June 22, 2017

Document Synopsis

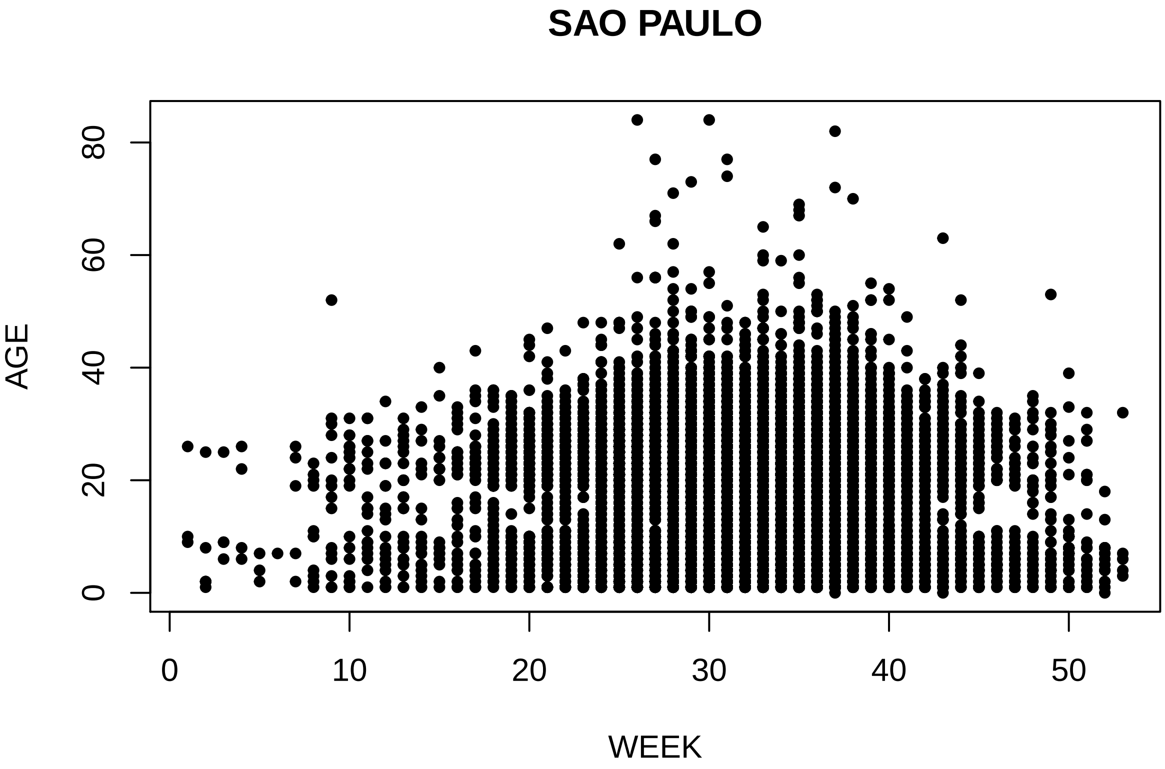
This document aims to visually and quantitatively describe the different age distribution of cases in the counties represented in spm.data.long.csv. Densities of cases for age groups will be calculated and then compared between counties.

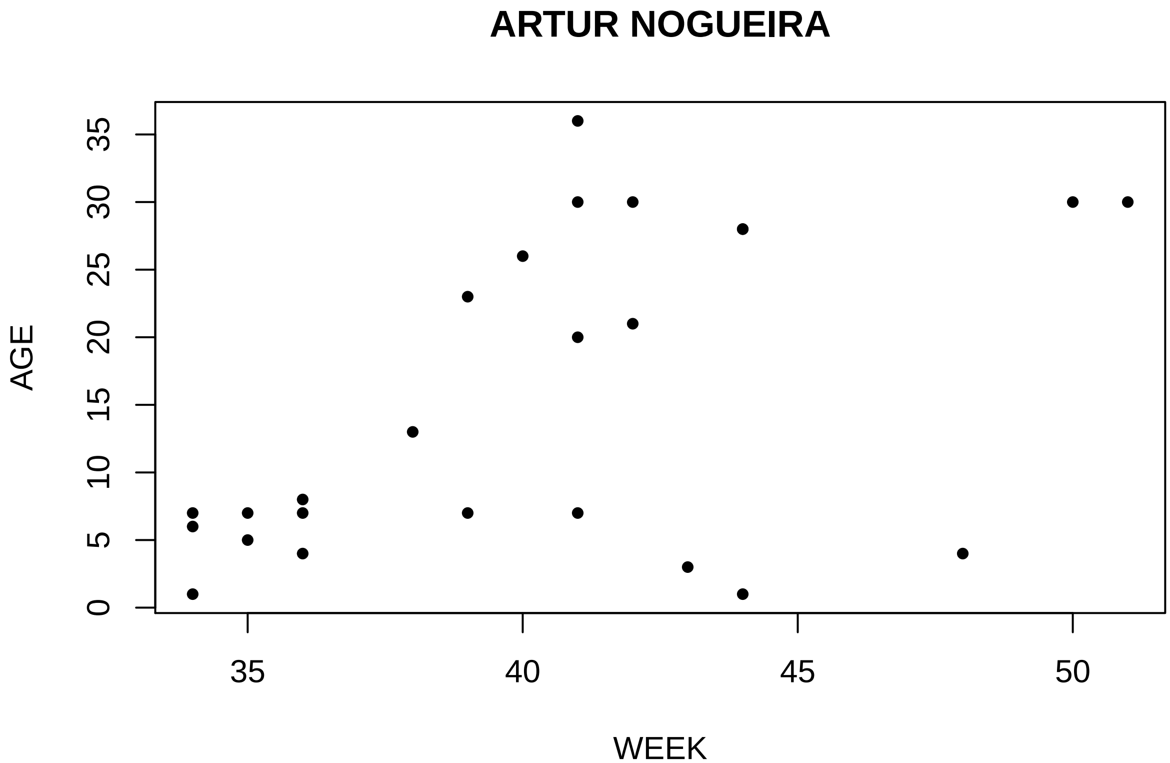
The difference between counties will be used in order to explore possible predictive relationships between population size or urbanness of a county and it's density of cases in different age classes.

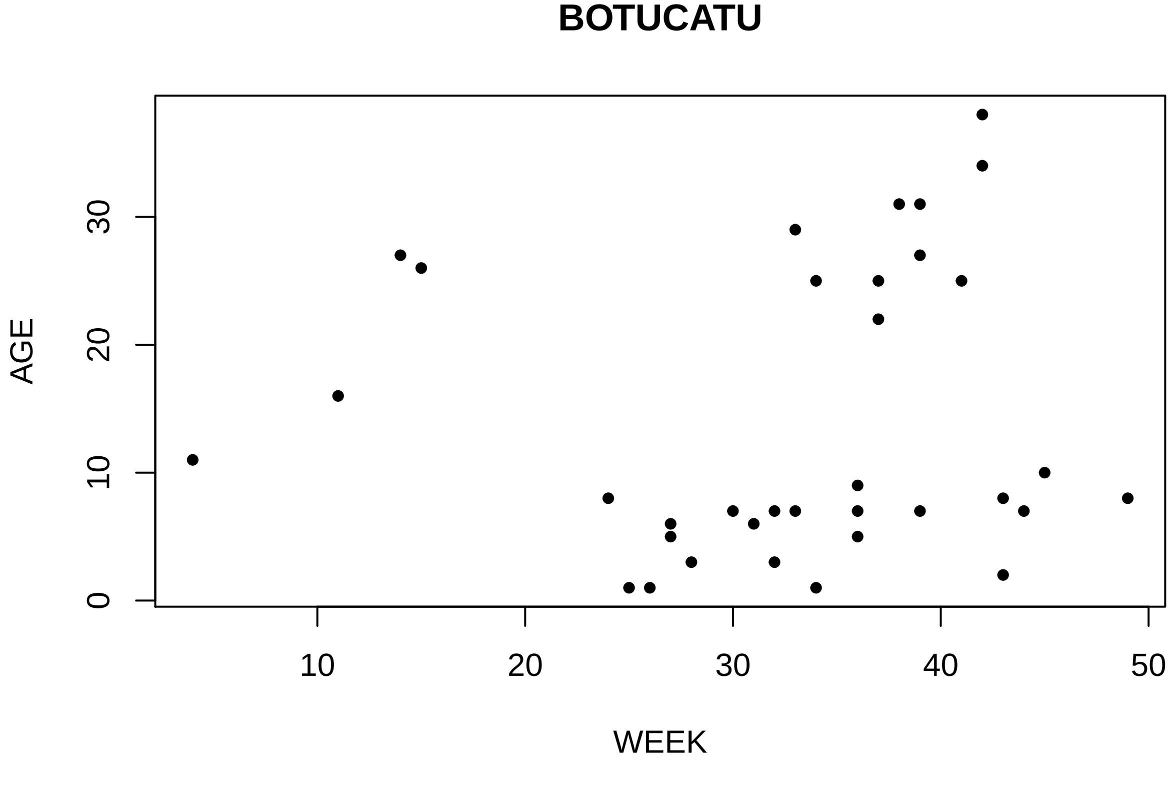
Age distribution scatter plots by county

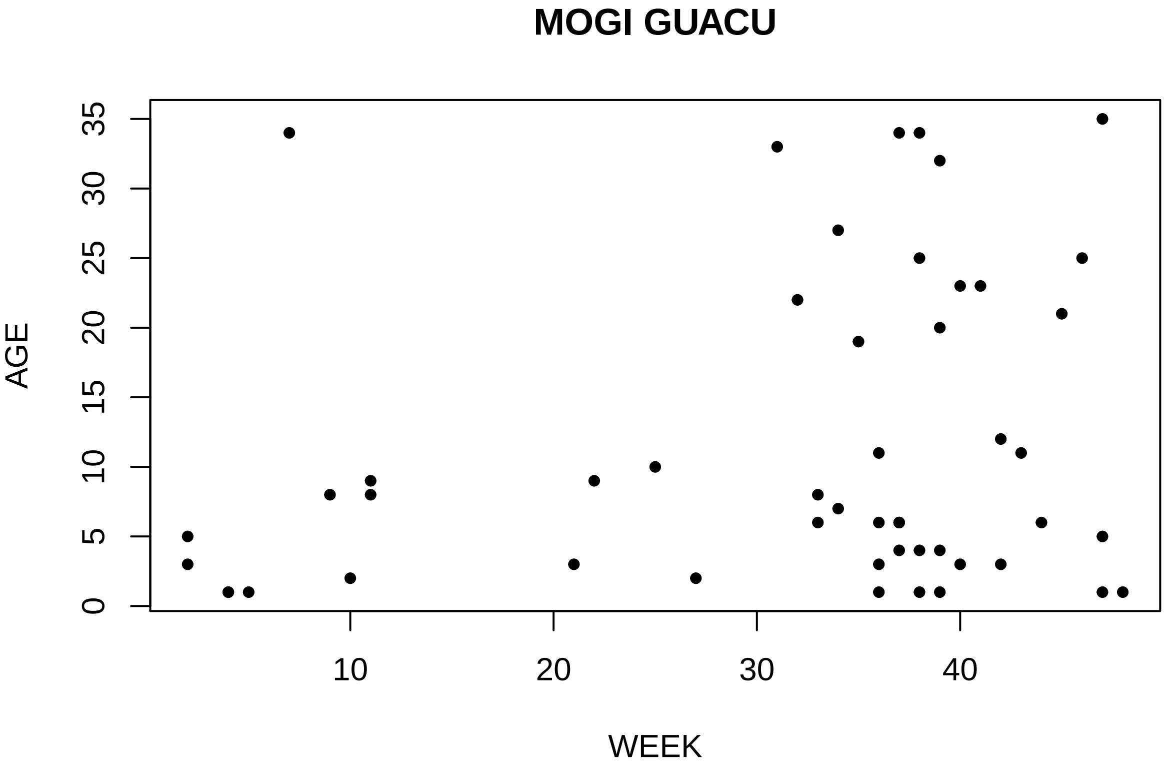
Scatter plots will be created in order to visually identify the age groups which frequently have a low density of cases. Depending on results, this may be limited to counties with number of cases above a certain threshold.

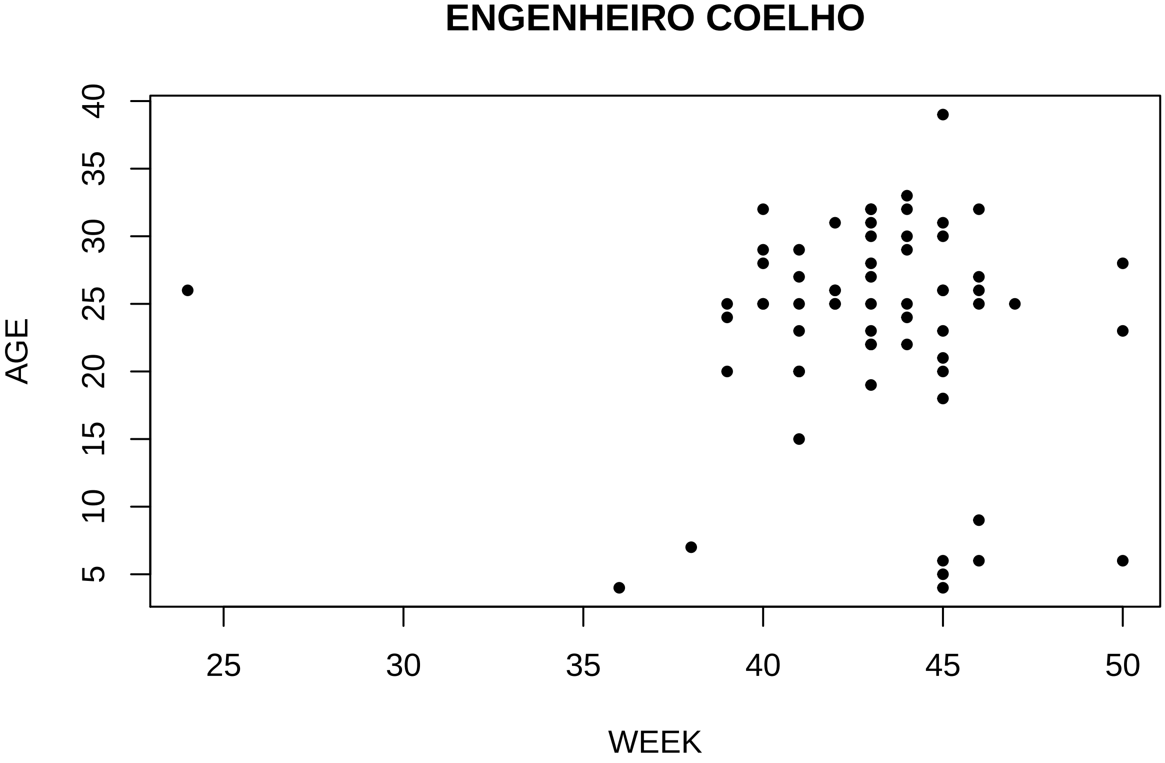
```
for(i in 1:length(county_vec)){
  subset_county <- subset(data, data$COUNTY == county_vec[i])
  if(length(subset_county$AGE) > 20){
    plot(subset_county$WEEK, subset_county$AGE, pch = 20,
         main = county_vec[i], xlab = "WEEK", ylab = "AGE")
  }
}
```

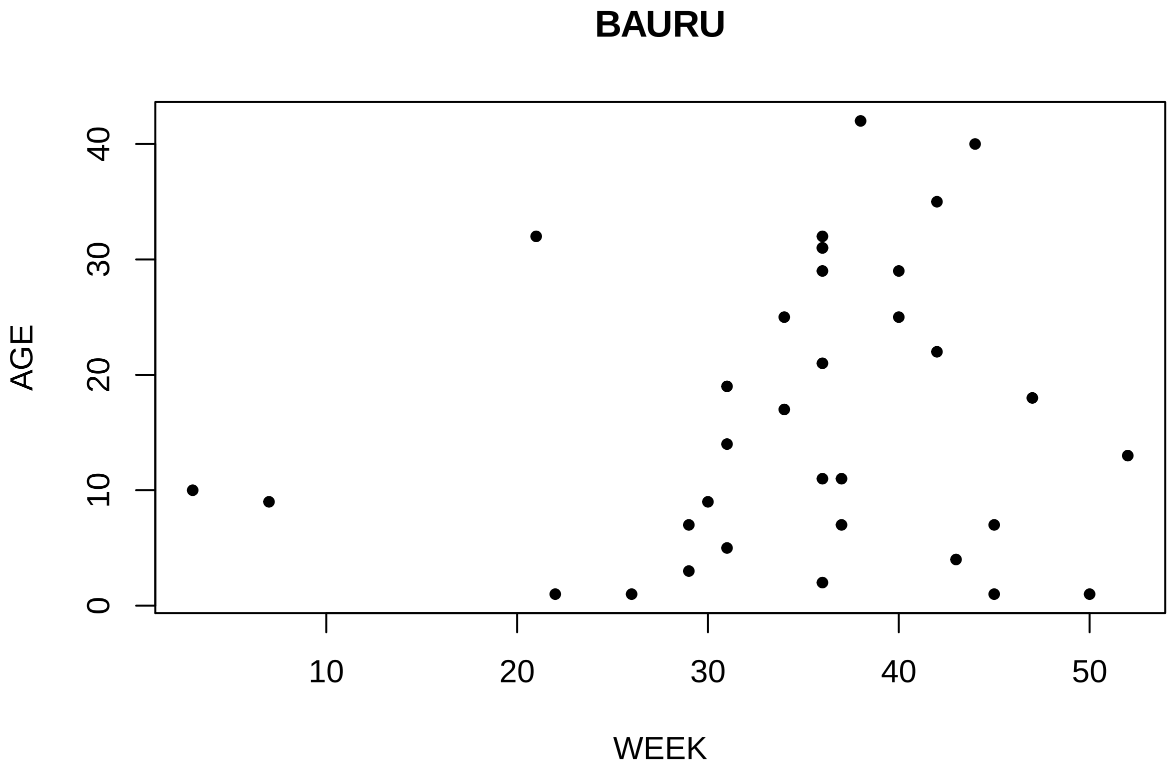


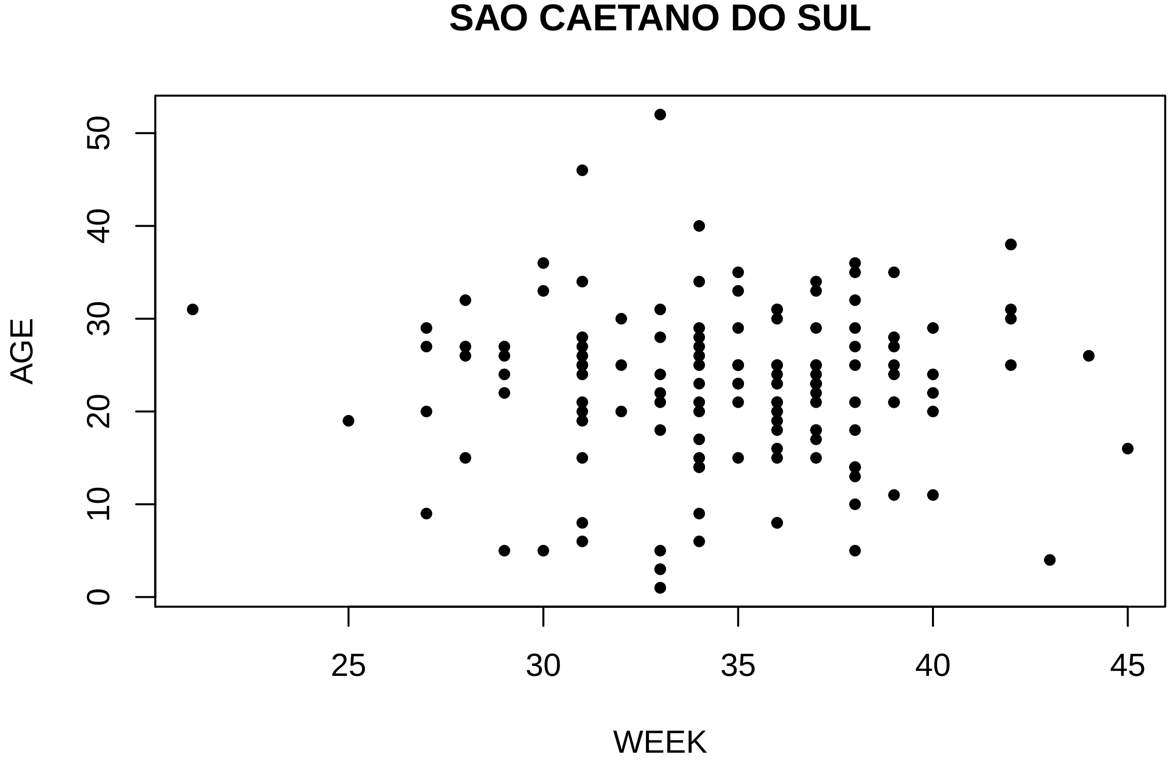

ARTUR NOGUEIRA
AGE
0
5
10
15
20
25
30
35
35
40
45
50
WEEK


BOTUCATU
AGE
0
10
20
30
10
20
30
40
50
WEEK

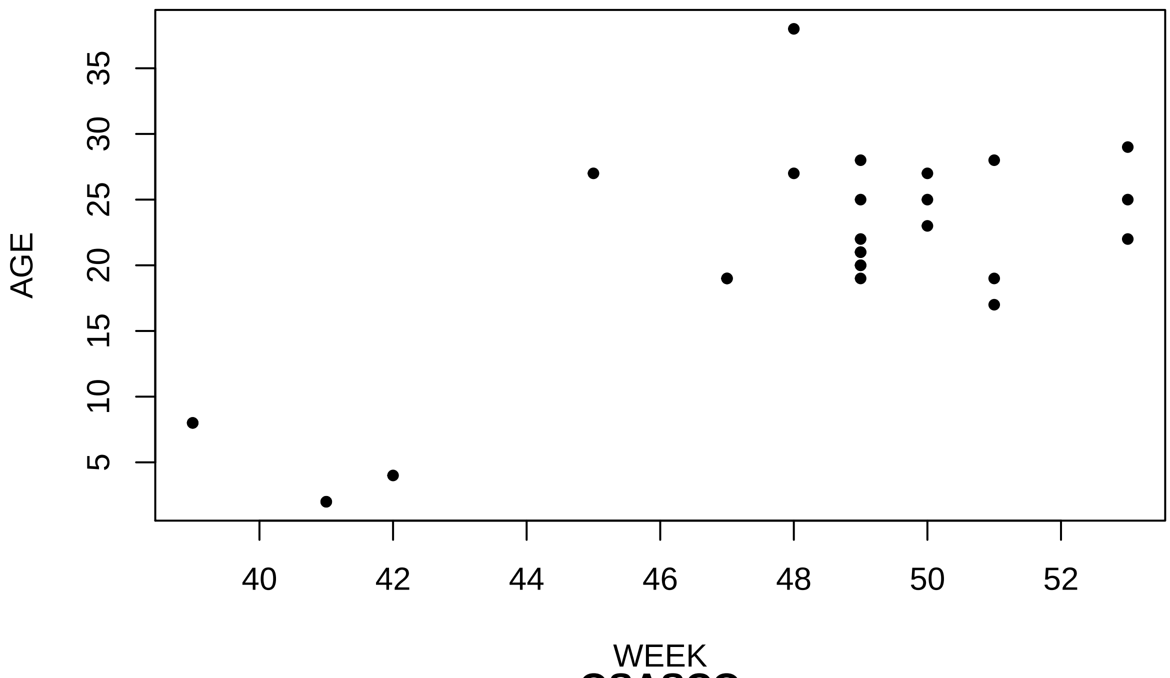

MOGI GUACU
AGE
0
5
10
15
20
25
30
35
10
20
30
40
WEEK


ENGENHEIRO COELHO
AGE
5
10
15
20
25
30
35
40
25
30
35
40
45
50
WEEK

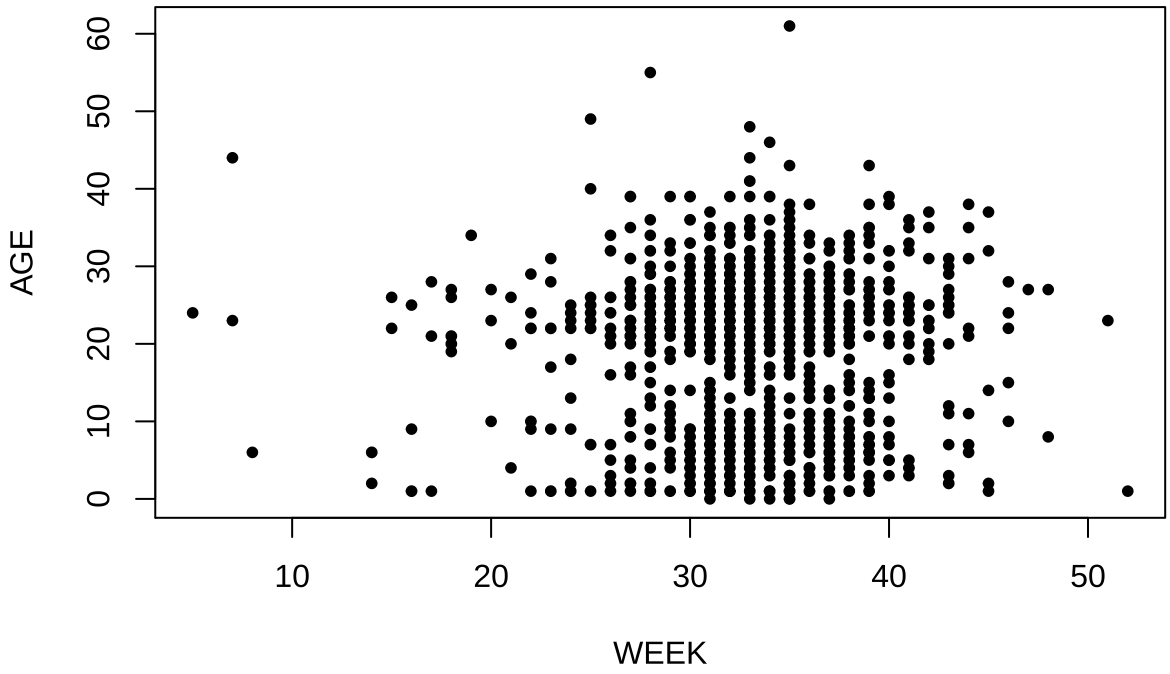

BAURU
AGE
WEEK
0
10
20
30
40
10
20
30
40
50

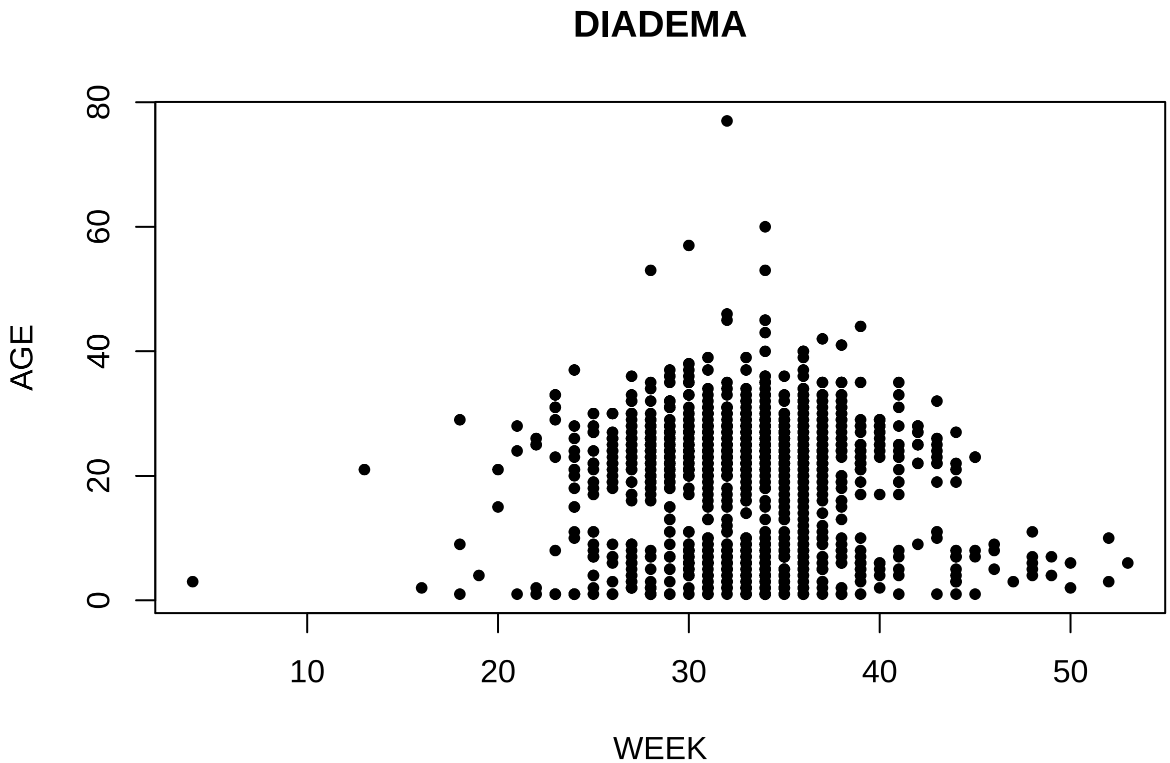

SAO CAETANO DO SUL
AGE
WEEK
0
10
20
30
40
50
25
30
35
40
45

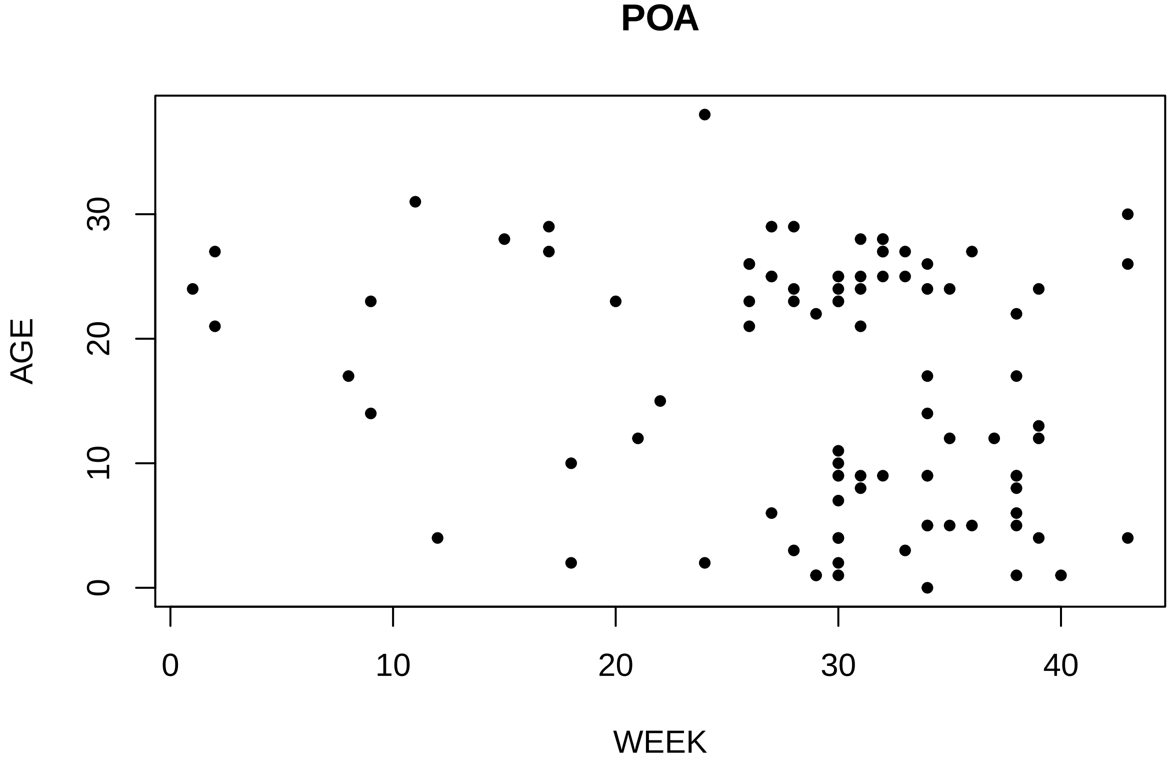
MORRO AGUDO

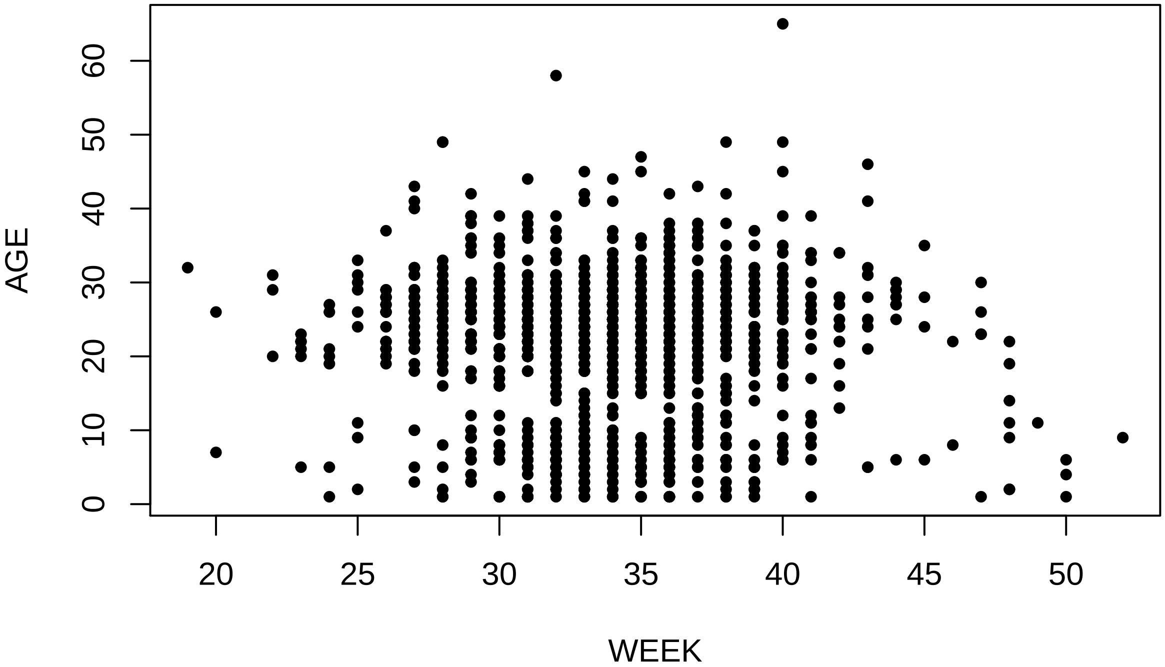


OSASCO

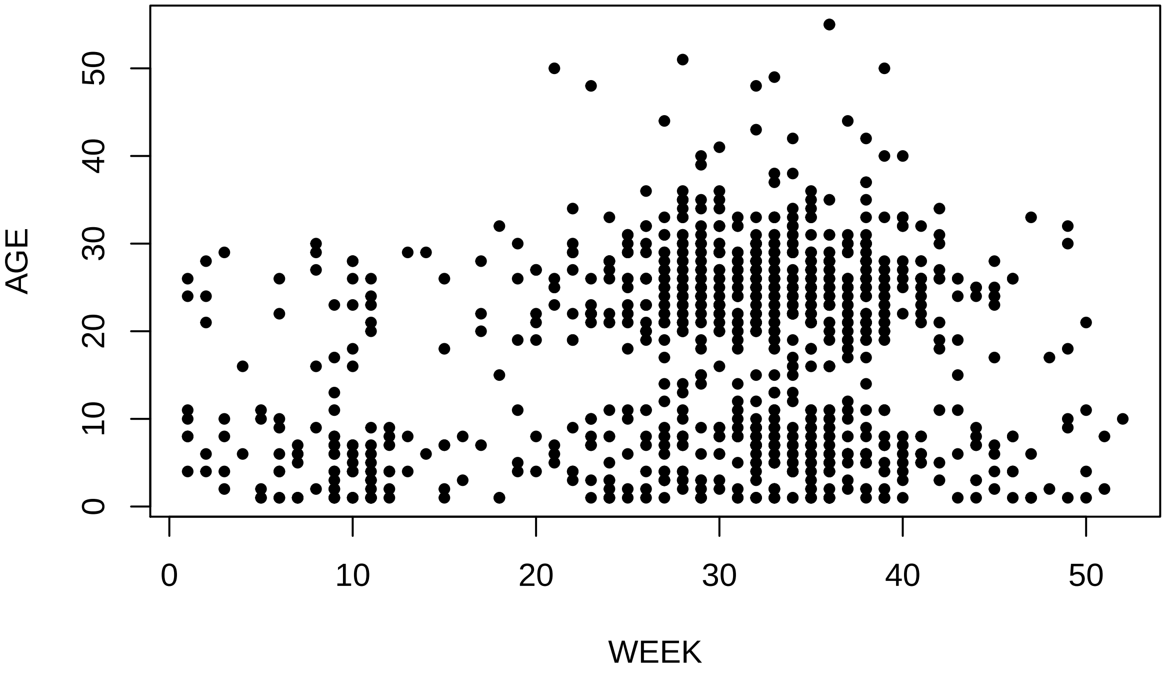



DIADEMA
AGE
WEEK
0
20
40
60
80
10
20
30
40
50

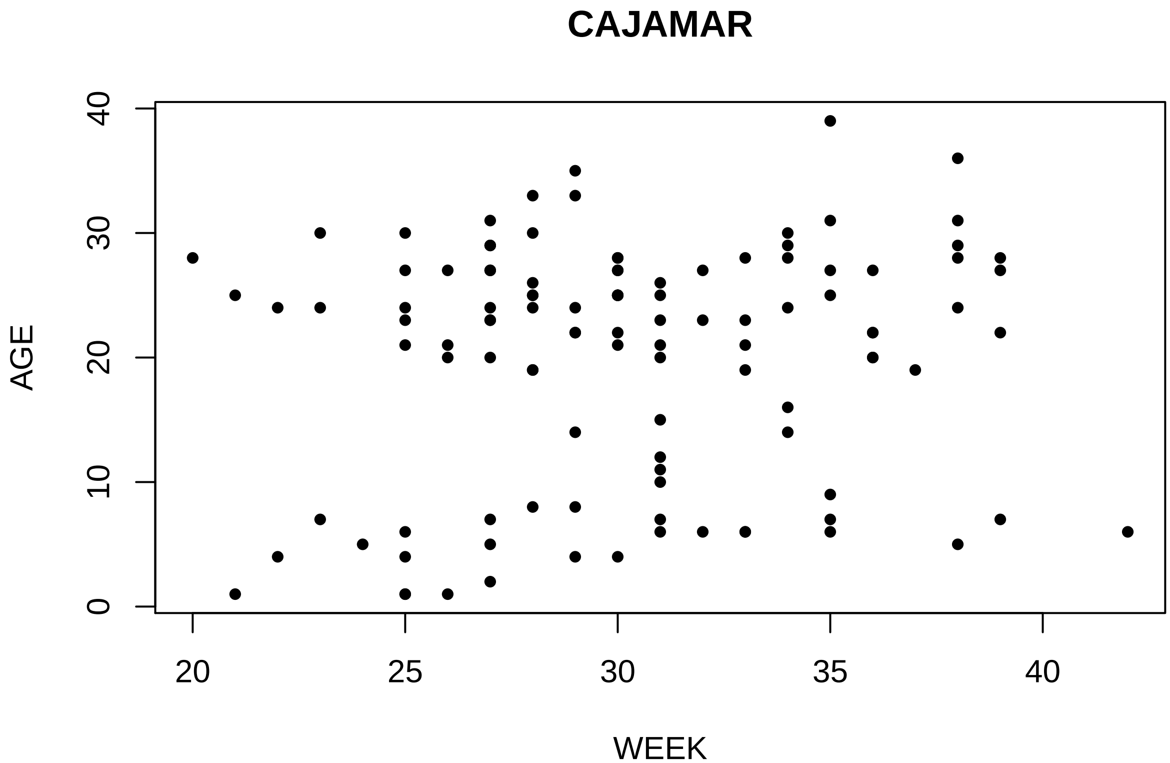

POA
AGE
WEEK
0
10
20
30
0
10
20
30
40



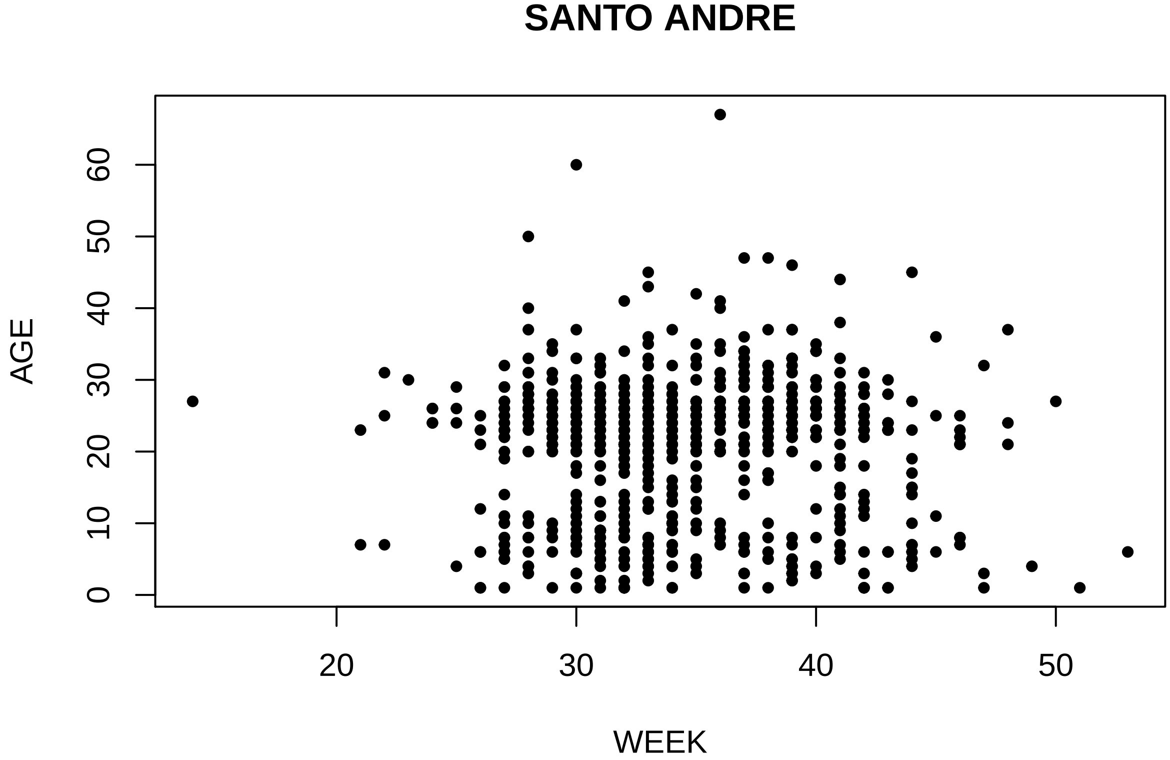
SAO BERNARDO DO CAMPO



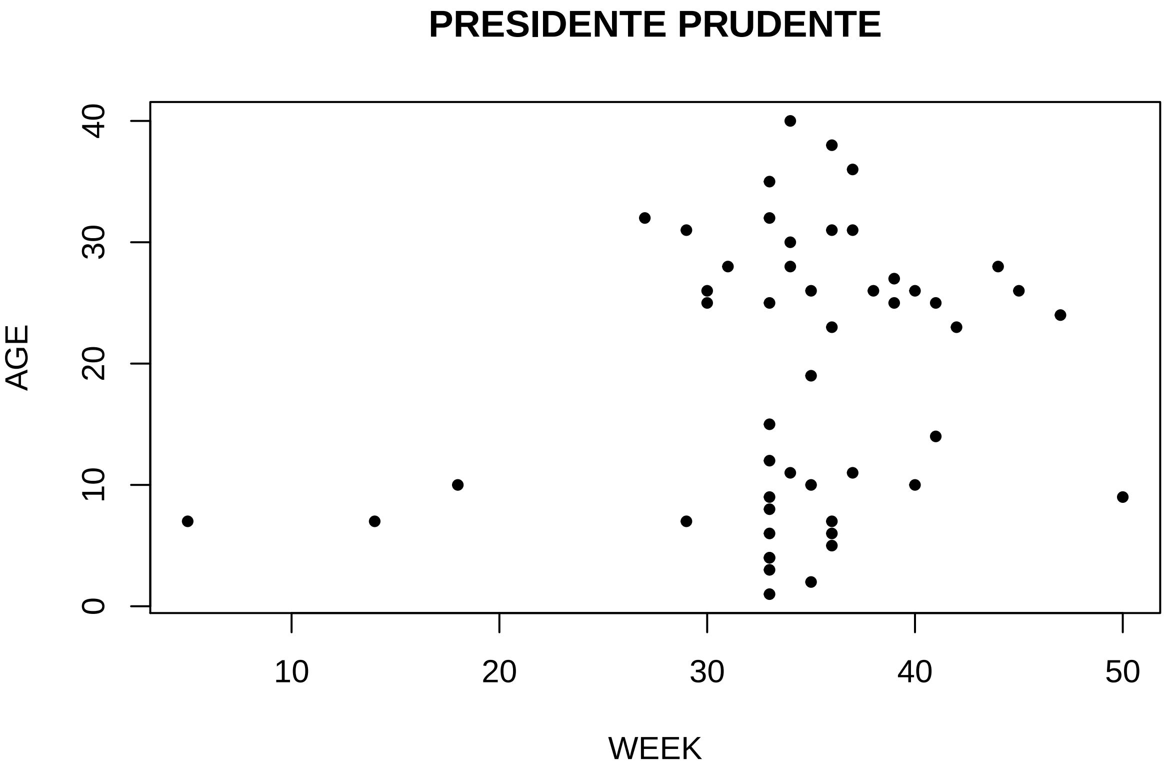
GUARULHOS

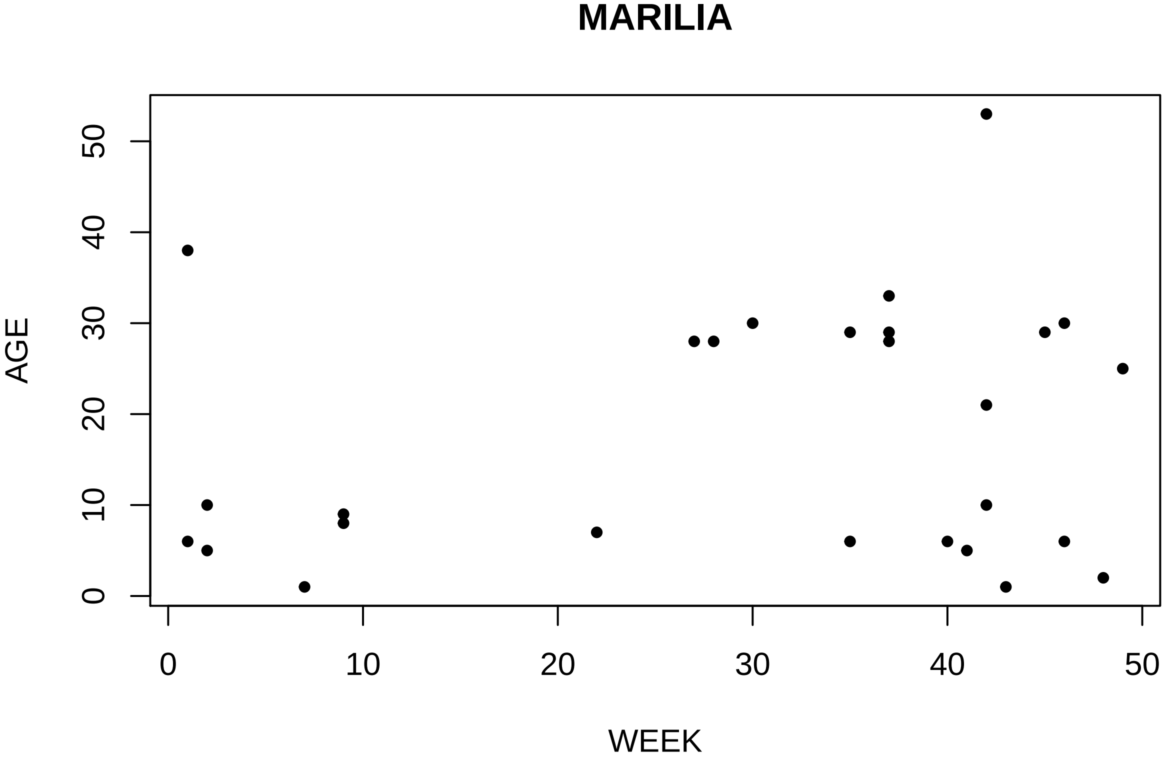


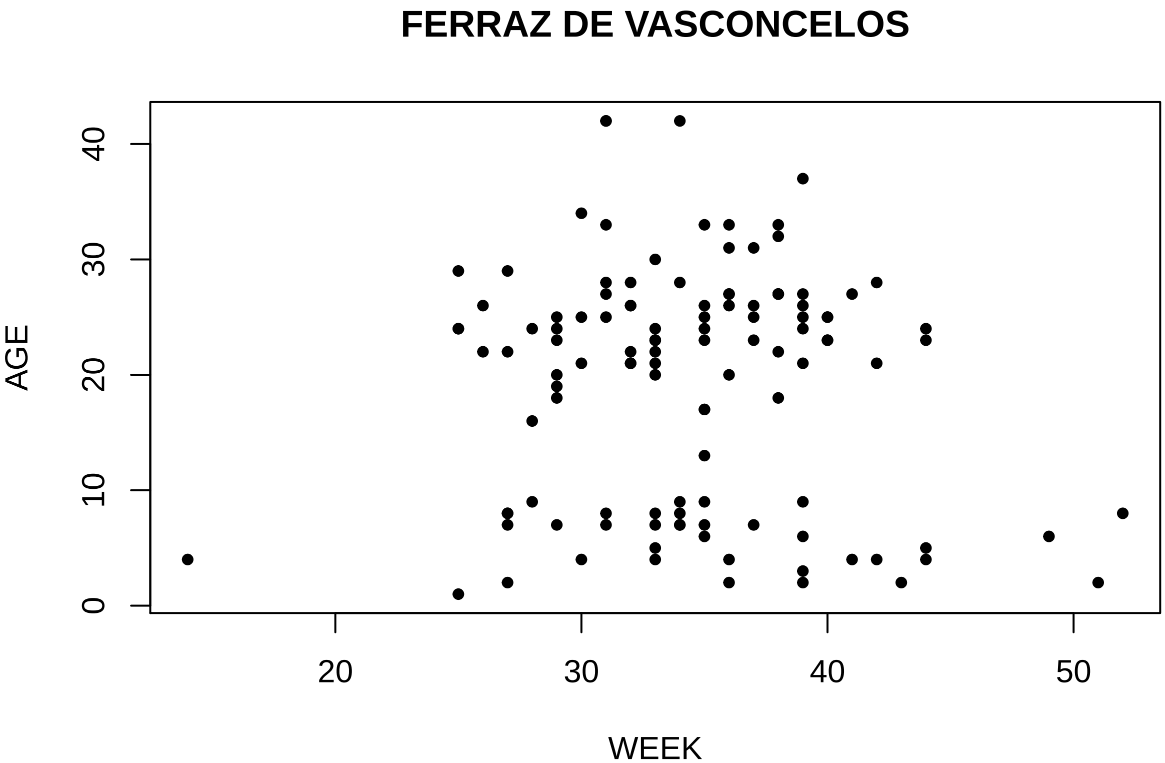
CAJAMAR

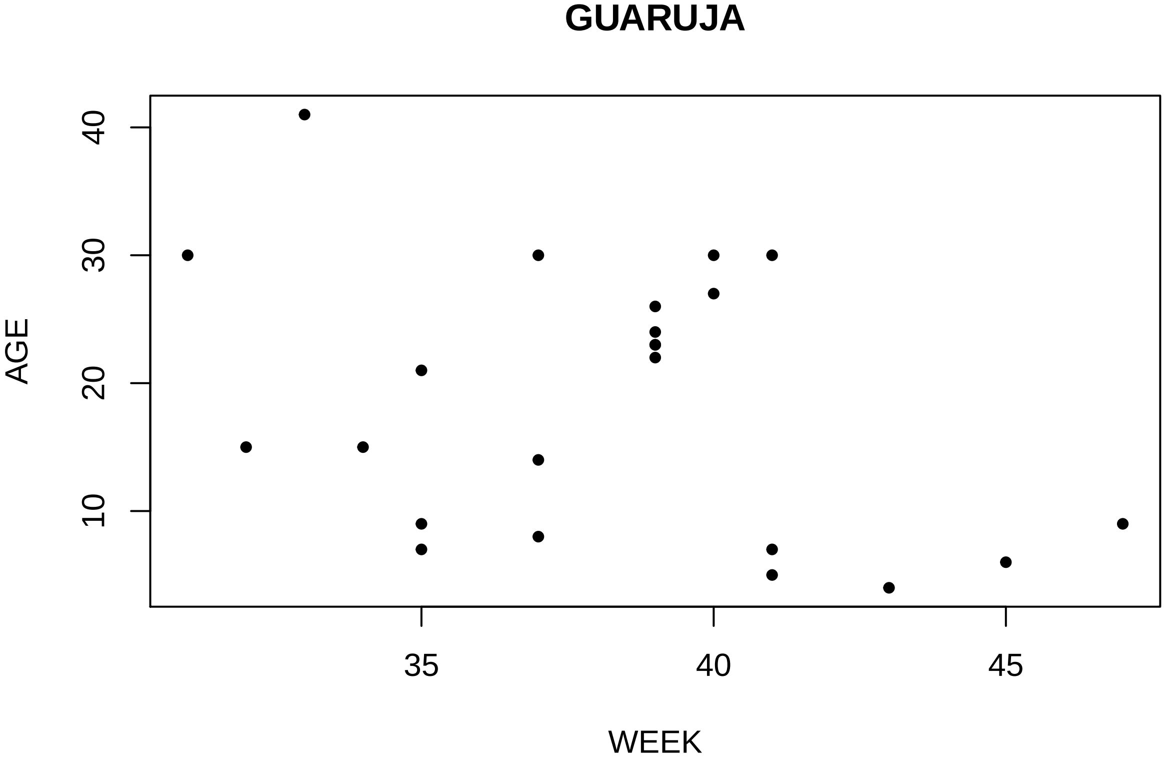


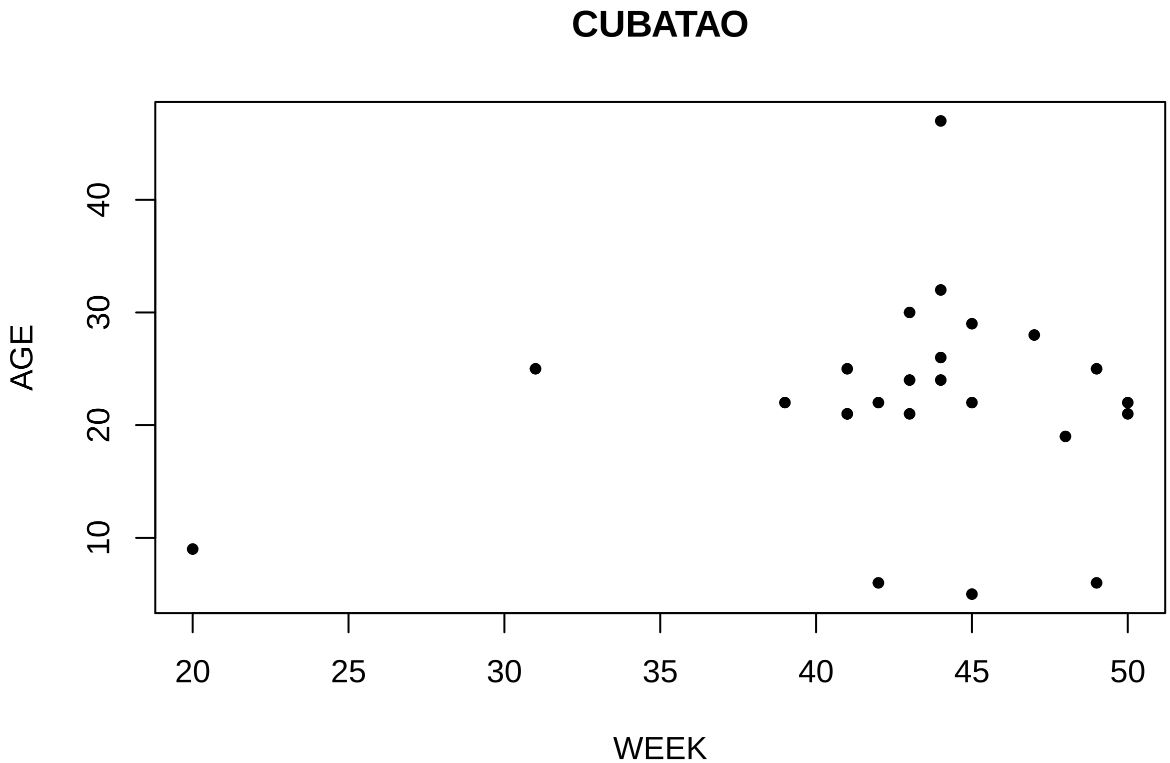
SANTO ANDRE


PRESIDENTE PRUDENTE
AGE
0
10
20
30
40
10
20
30
40
50
WEEK

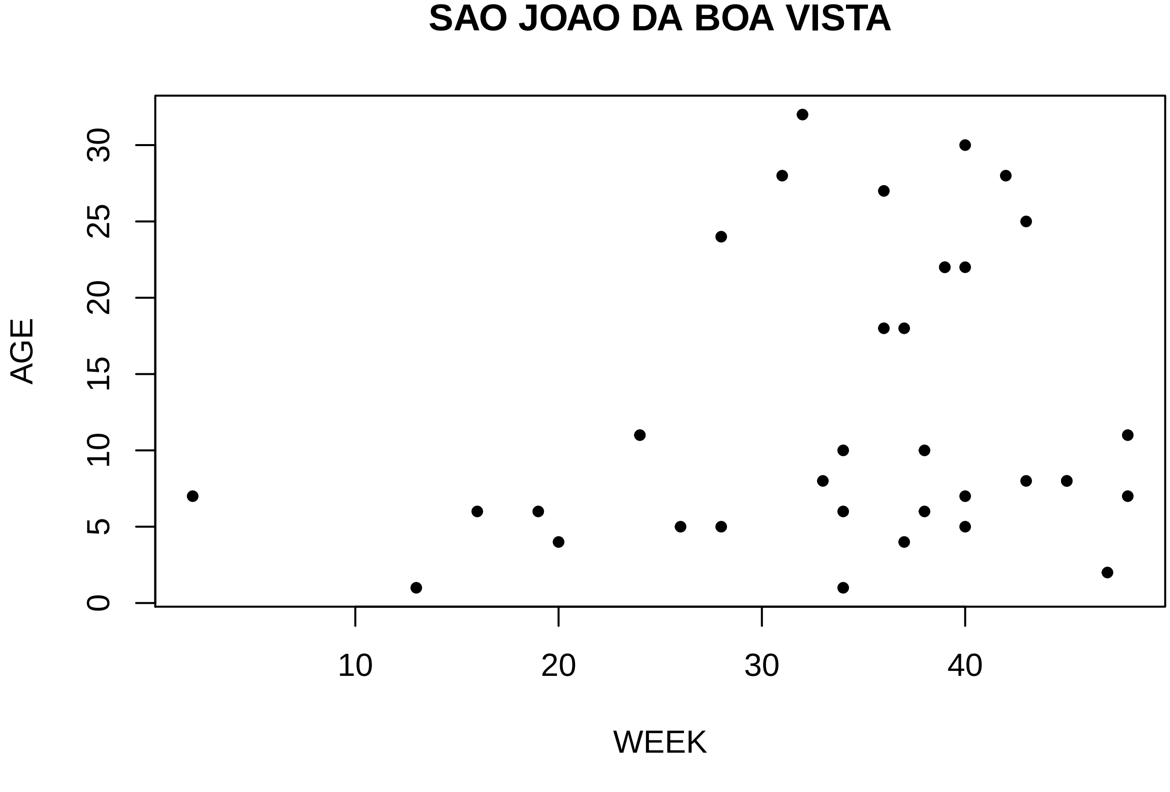

MARILIA
AGE
0
10
20
30
40
50
0
10
20
30
40
50
WEEK


FERRAZ DE VASCONCELOS
AGE
0
10
20
30
40
WEEK
20
30
40
50

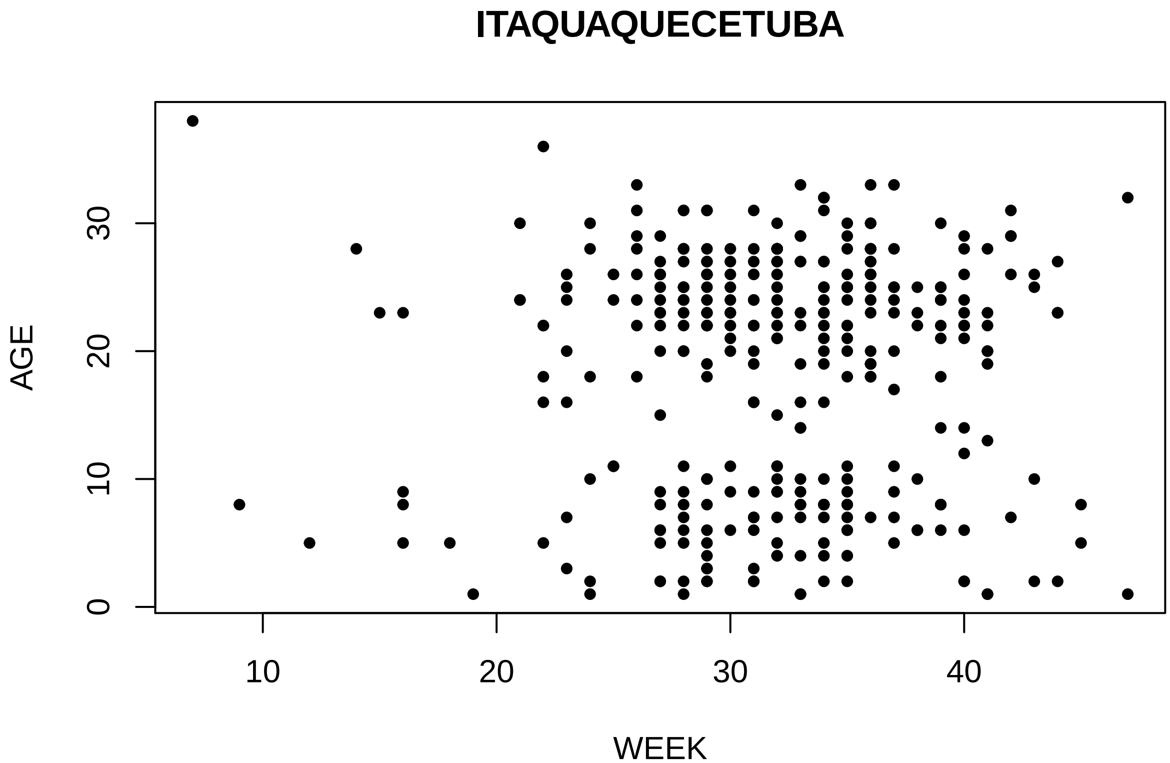

GUARUJA
AGE
10
20
30
40
WEEK
35
40
45

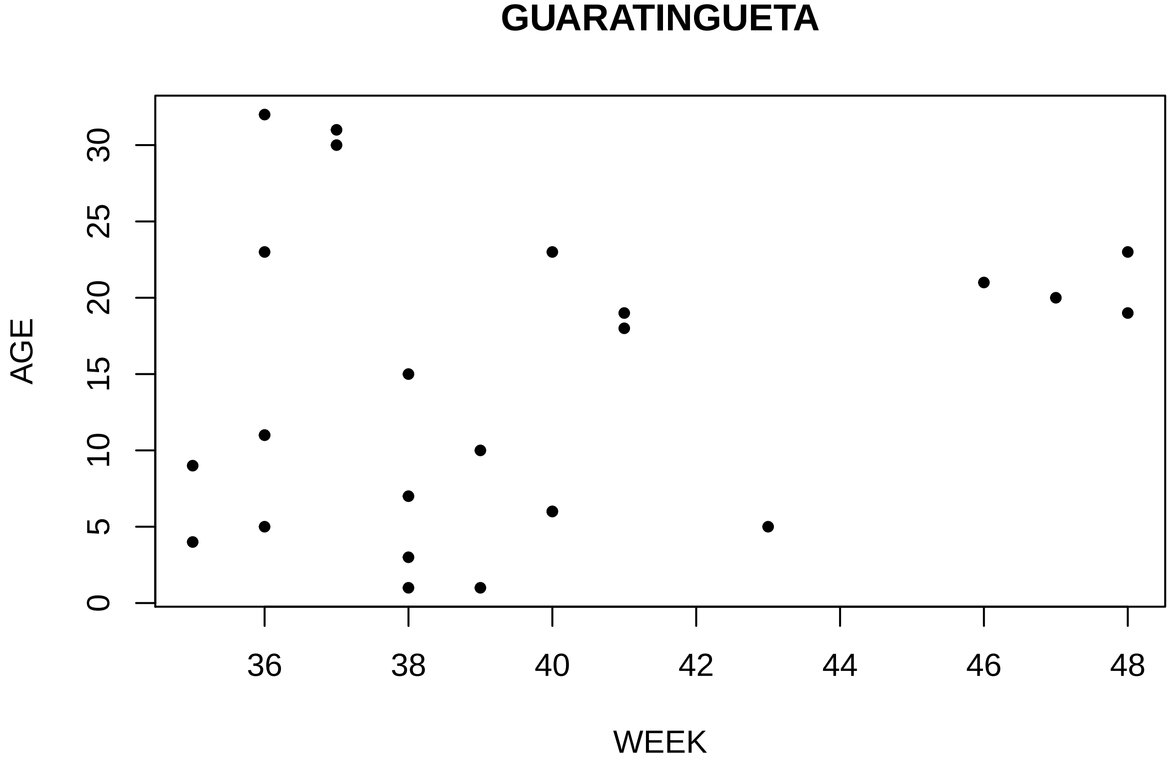


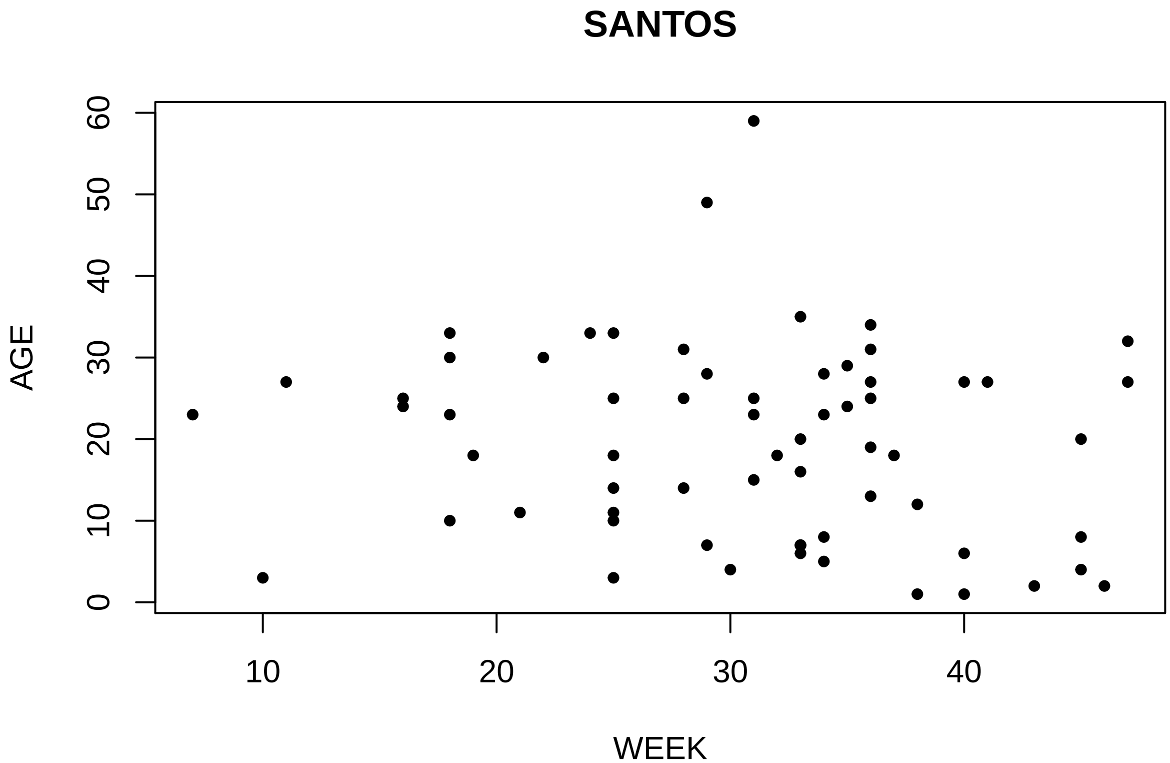
CUBATAO



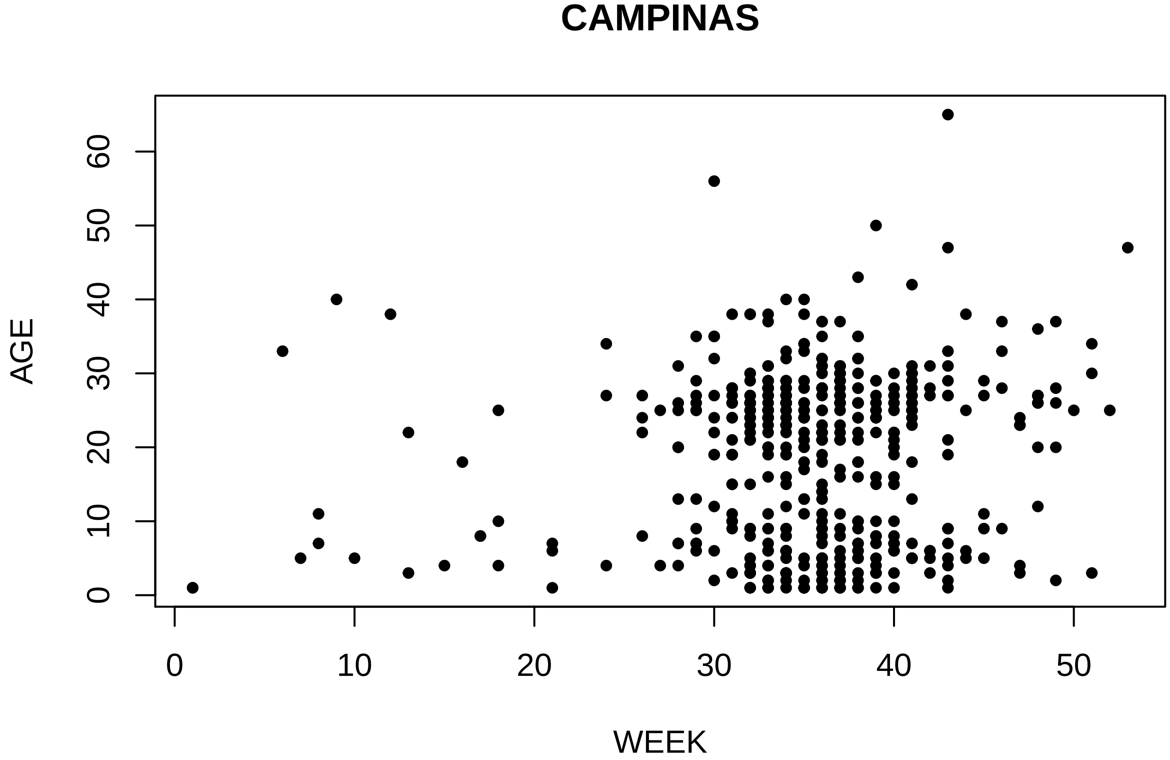
SAO JOAO DA BOA VISTA


ITAQUAQUECETUBA
AGE
0
10
20
30
10
20
30
40
WEEK

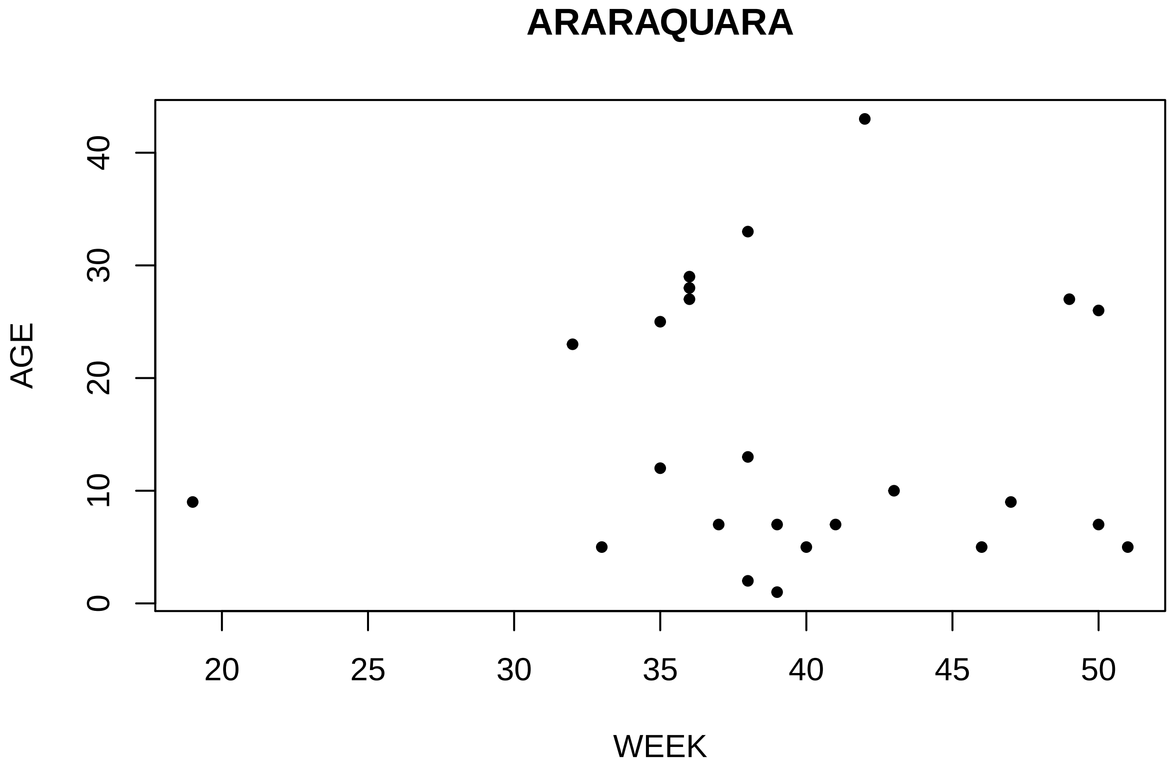

GUARATINGUETA
AGE
0
5
10
15
20
25
30
36
38
40
42
44
46
48
WEEK

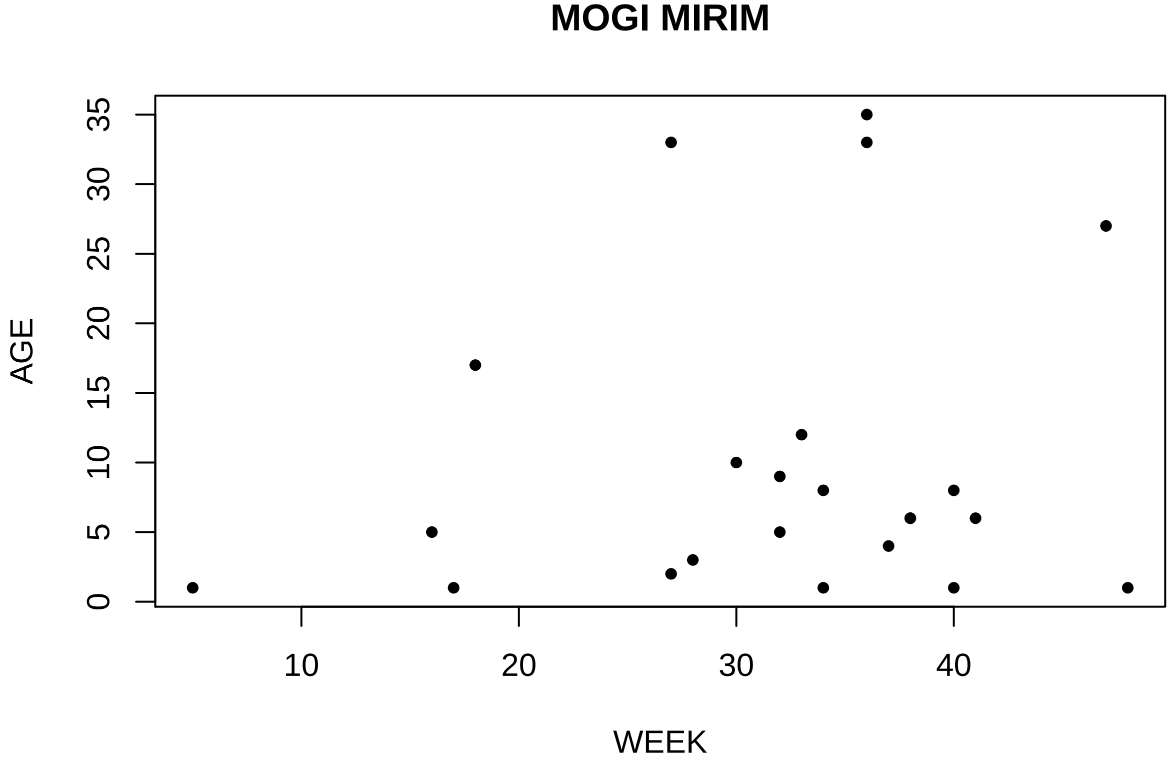


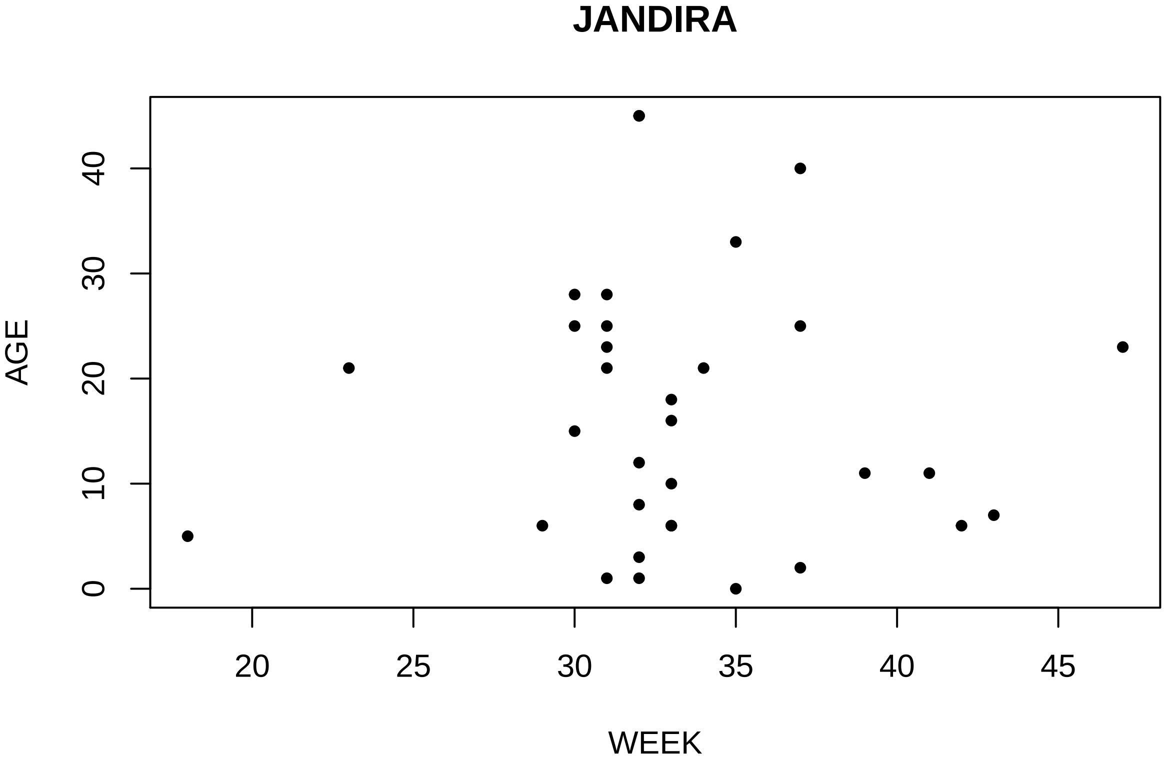
SANTOS

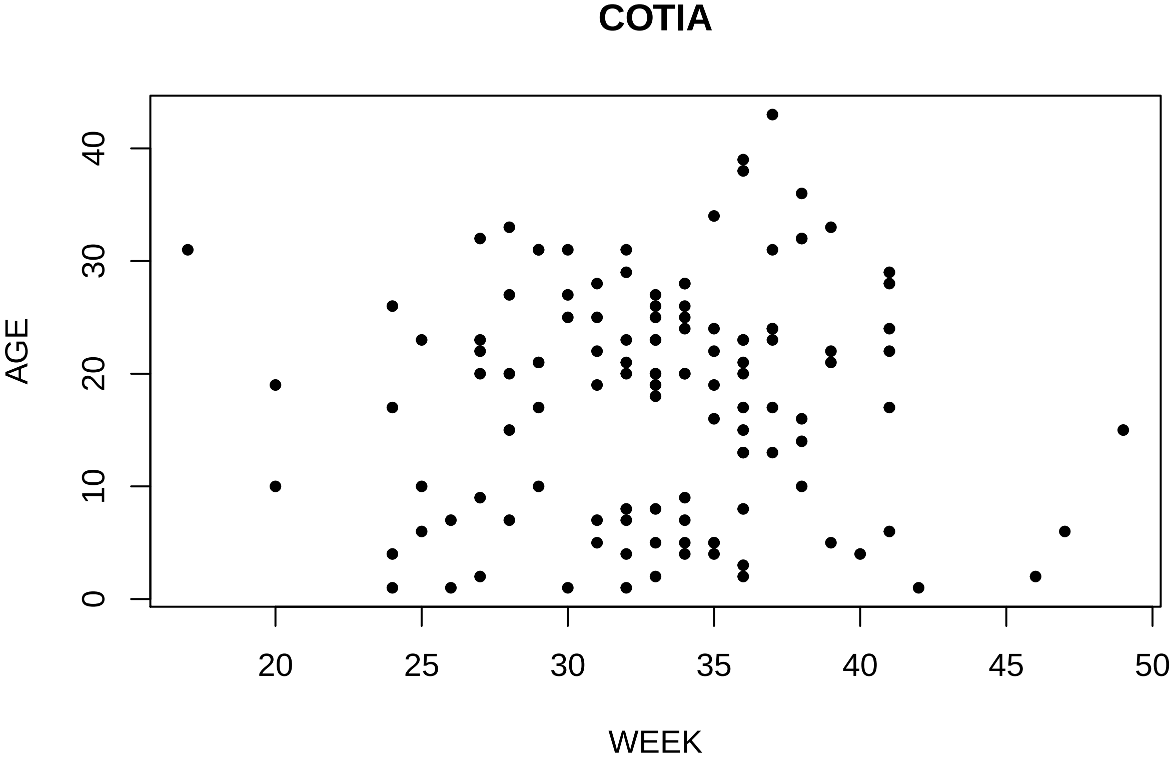


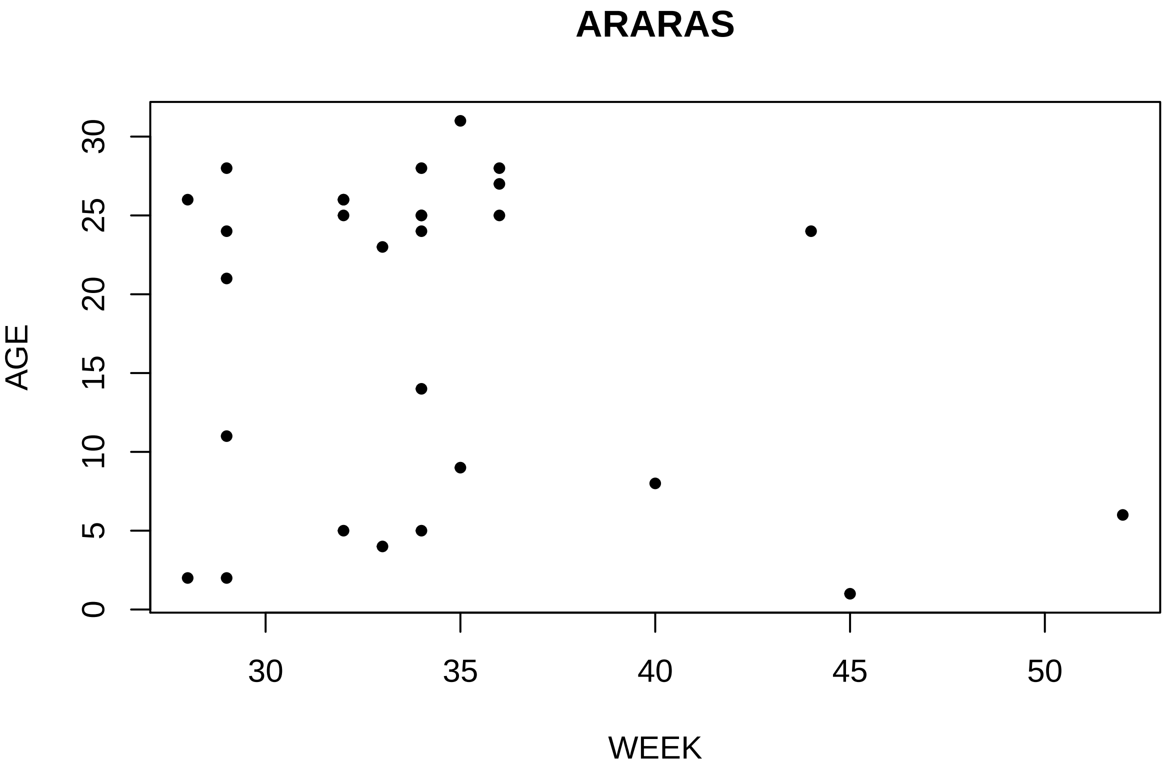
CAMPINAS


ARARAQUARA
AGE
0
10
20
30
40
20
25
30
35
40
45
50
WEEK

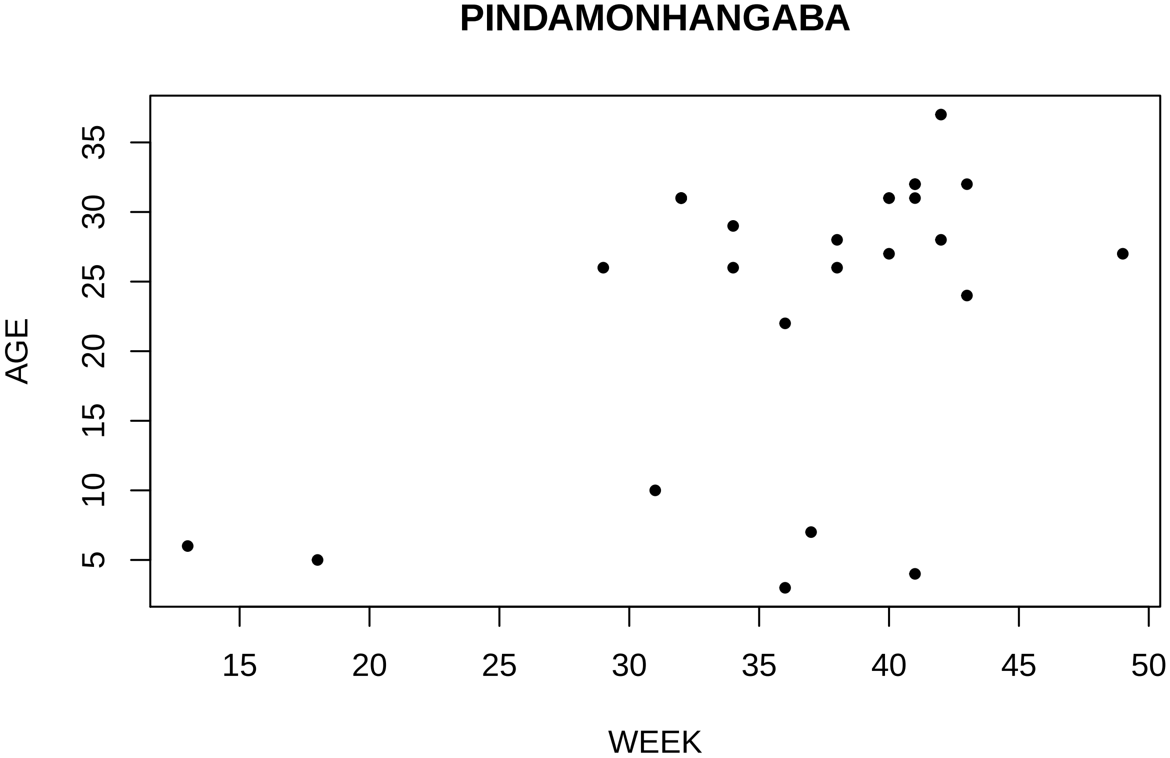

MOGI MIRIM
AGE
0
5
10
15
20
25
30
35
10
20
30
40
WEEK


JANDIRA
AGE
0
10
20
30
40
20
25
30
35
40
45
WEEK


COTIA
AGE
0
10
20
30
40
20
25
30
35
40
45
50
WEEK

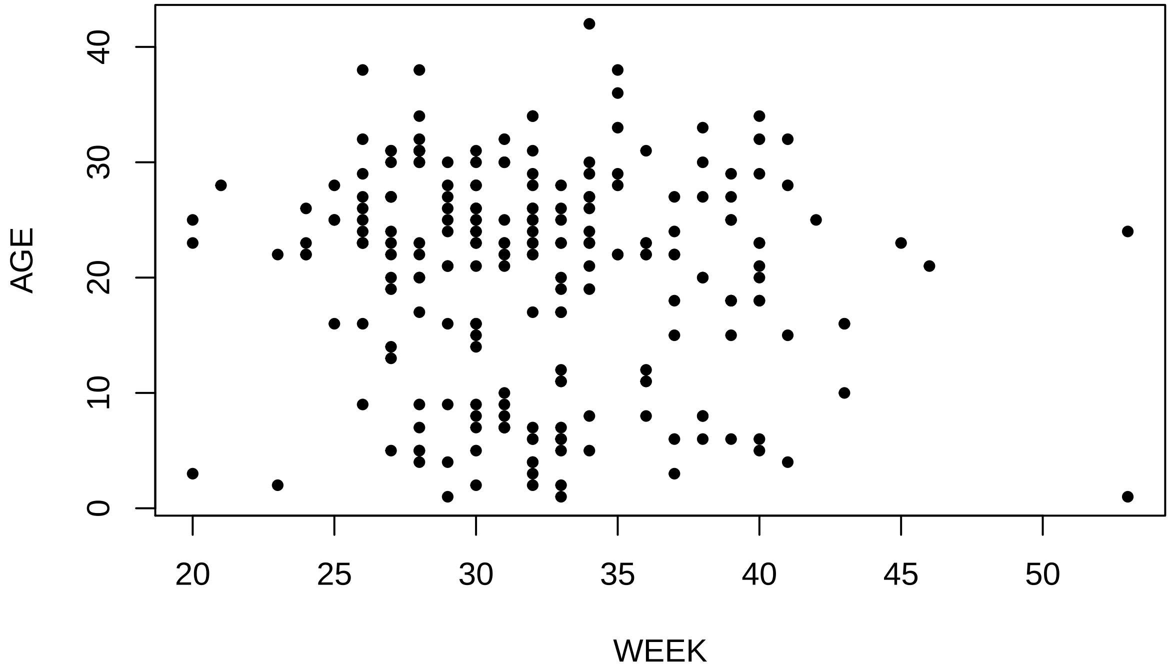


ARARAS
AGE
0
5
10
15
20
25
30
30
35
40
45
50
WEEK

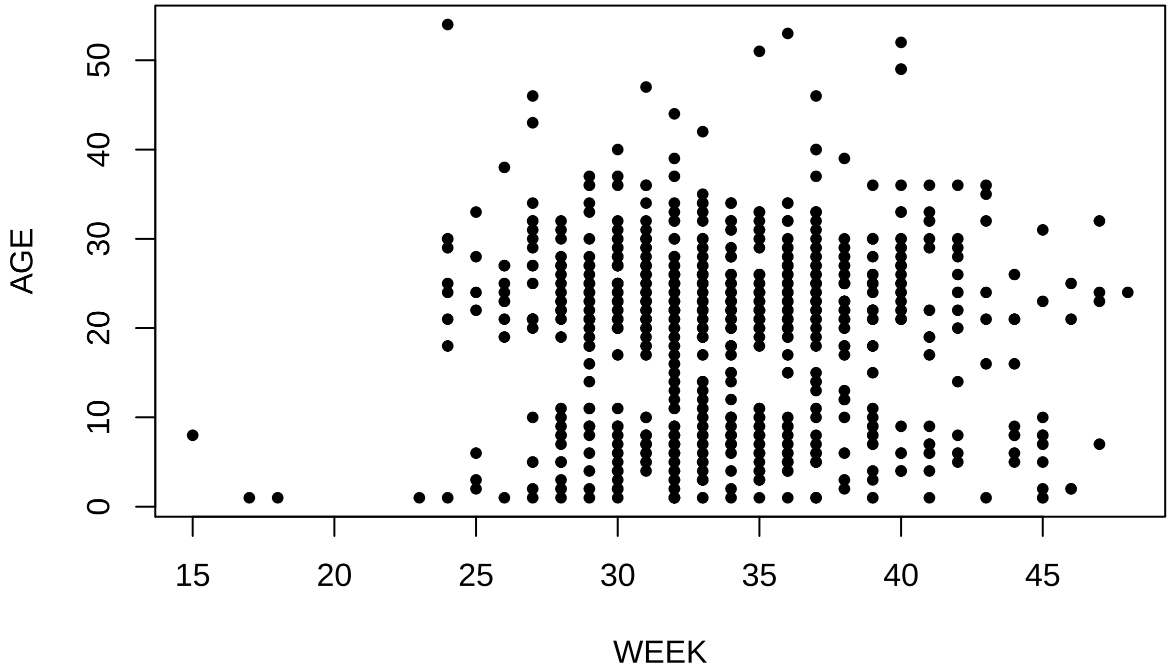


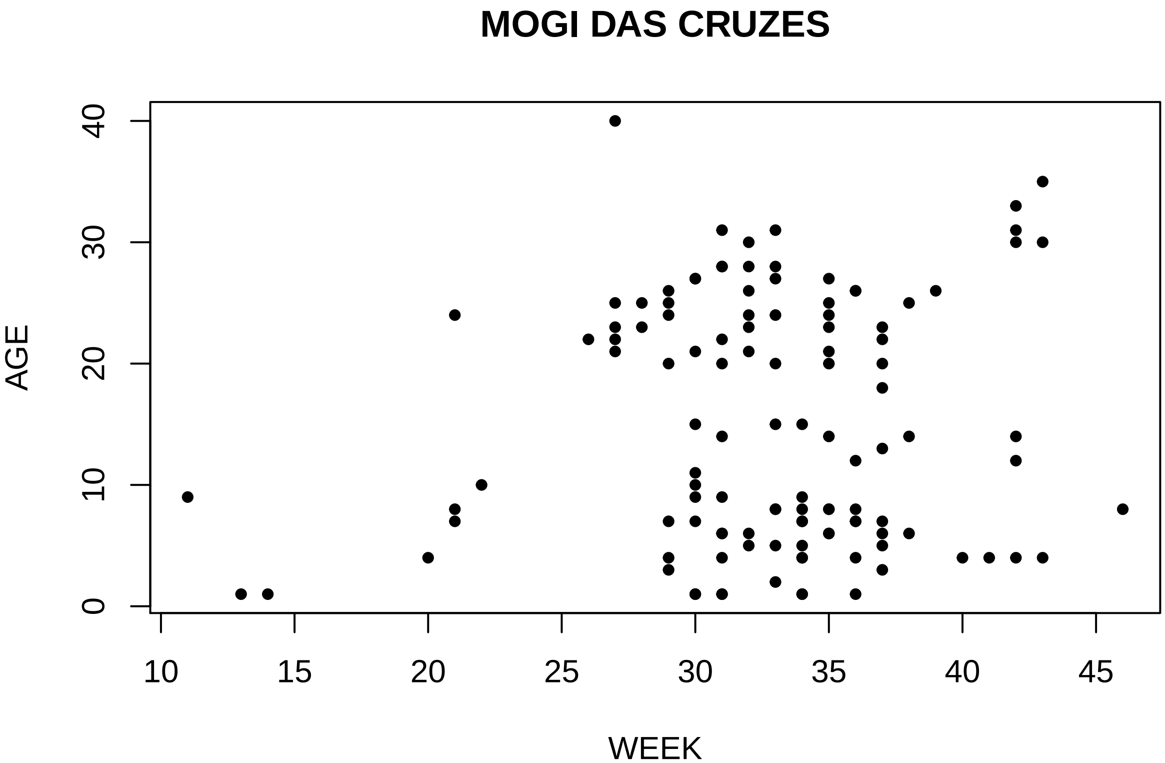
PINDAMONHANGABA
AGE
5
10
15
20
25
30
35
15
20
25
30
35
40
45
50
WEEK

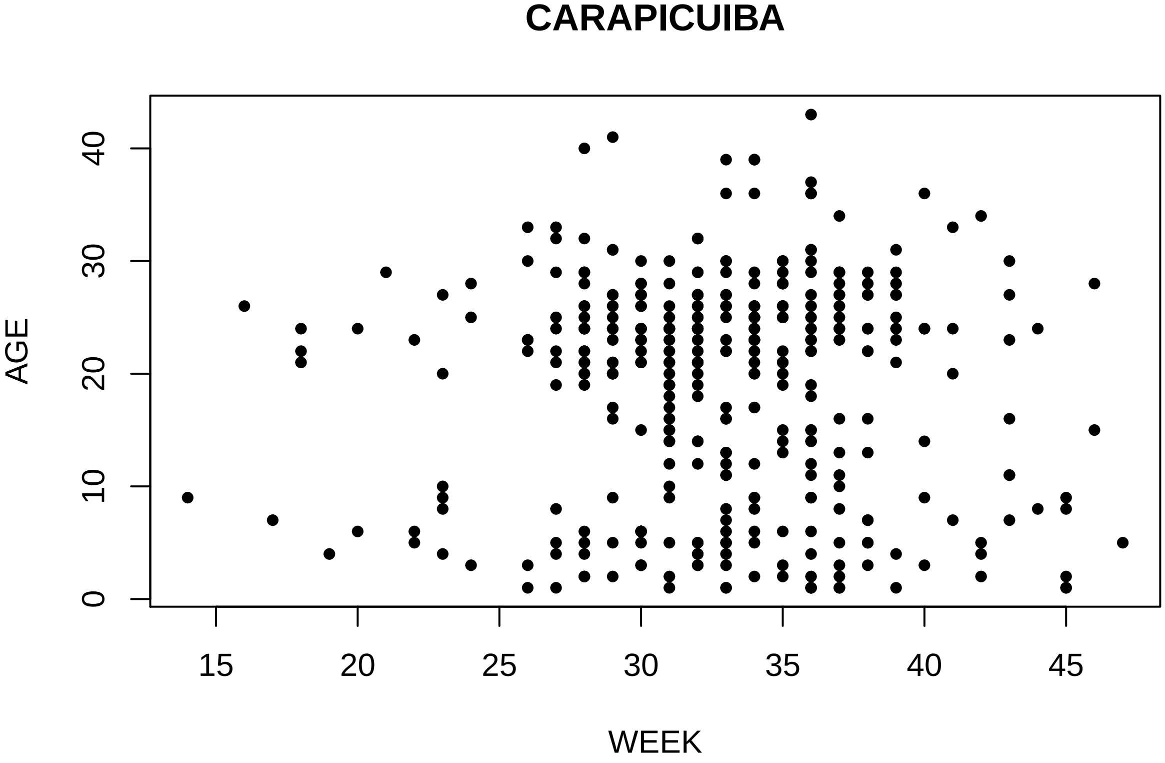
FRANCO DA ROCHA

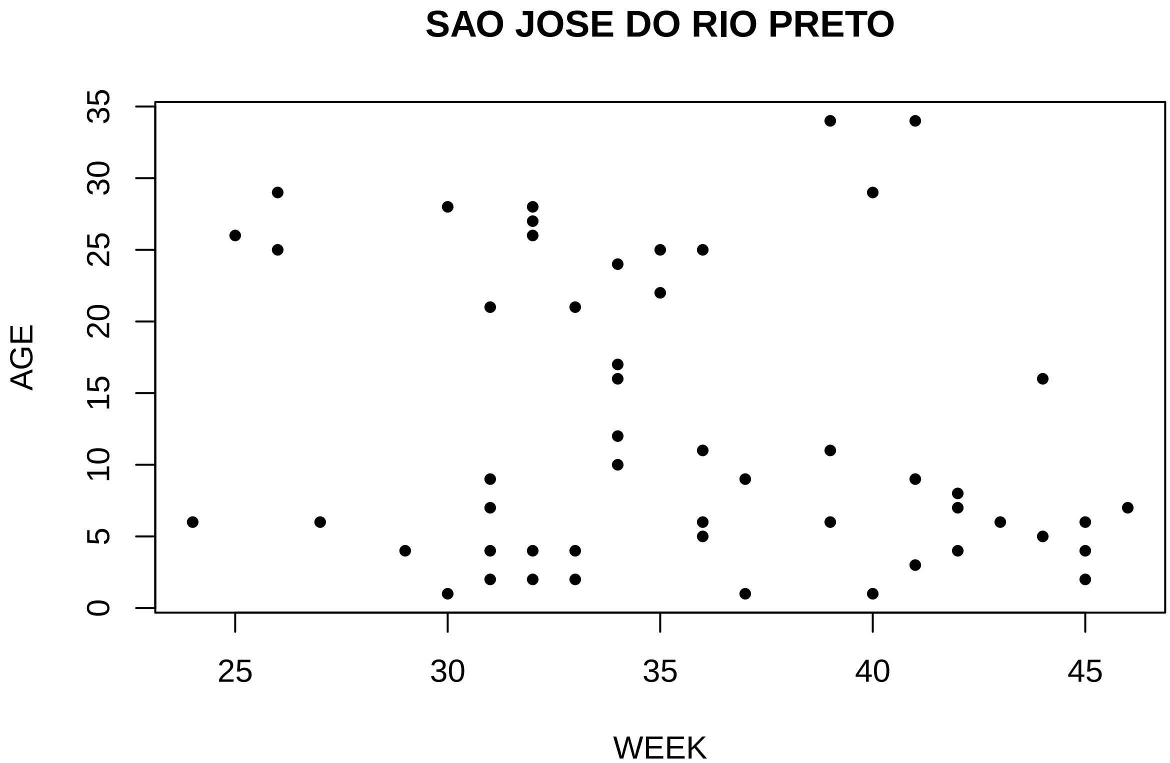


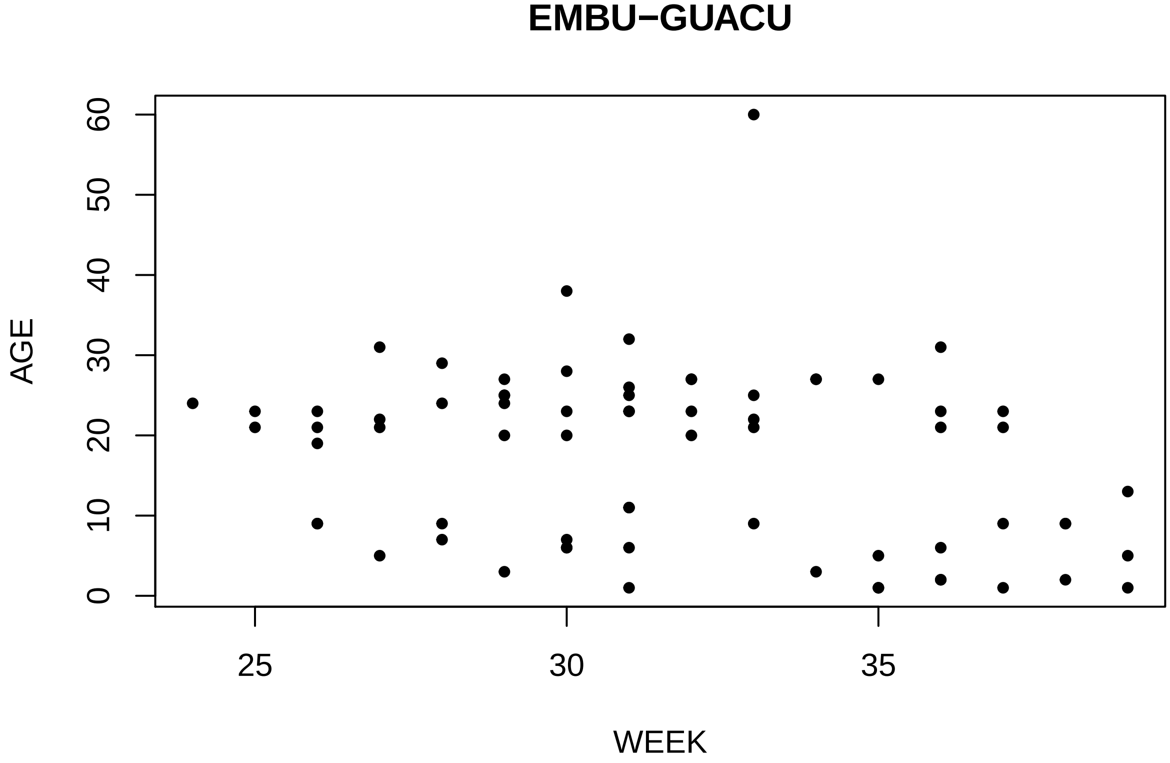
MAUA

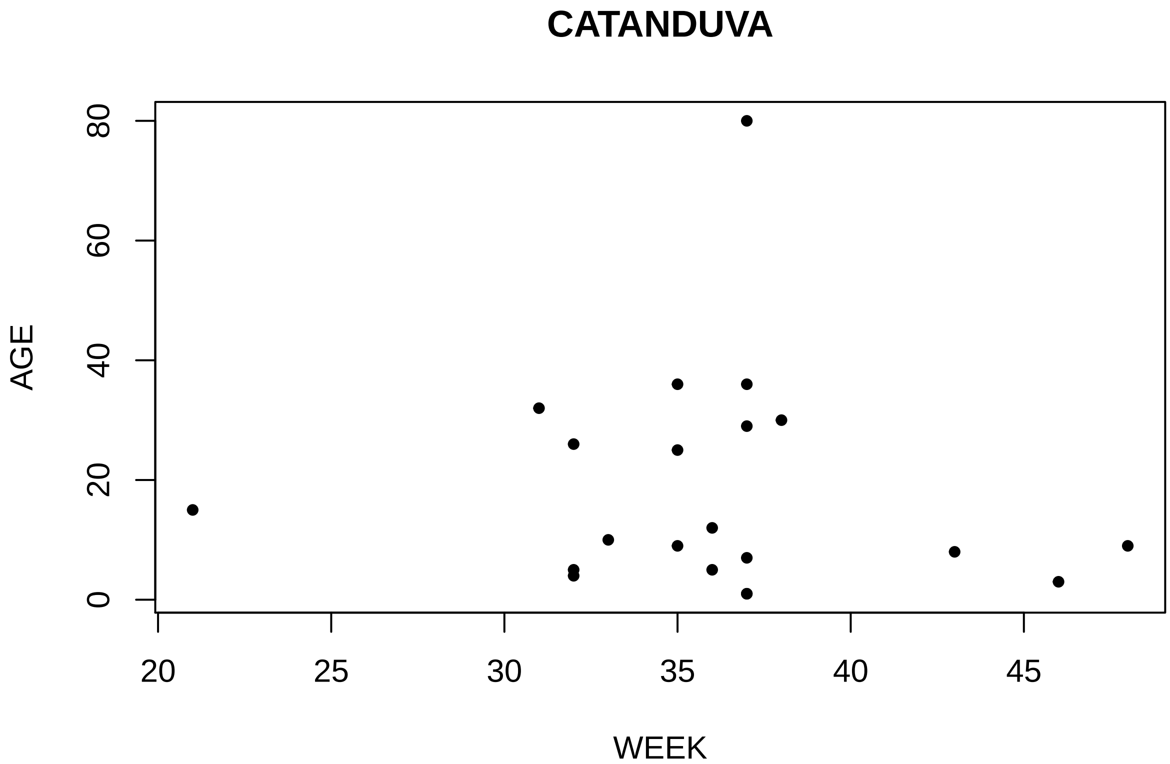


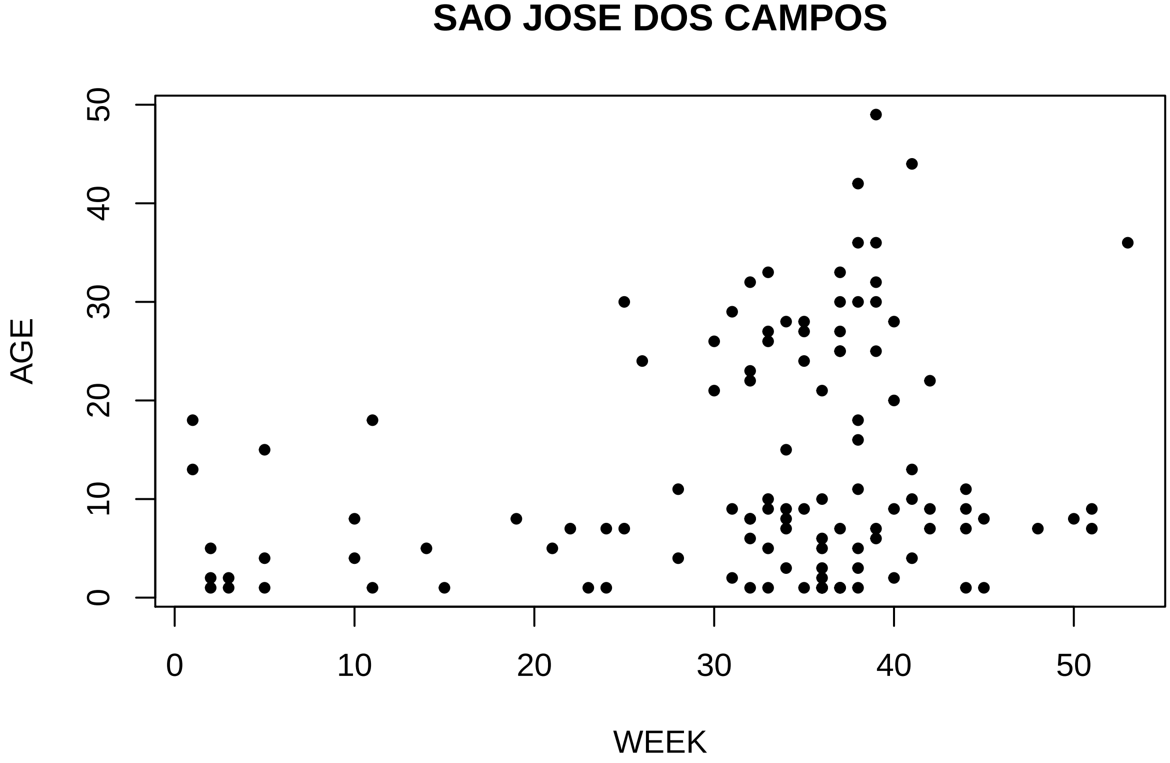

MOGI DAS CRUZES
AGE
0
10
20
30
40
10
15
20
25
30
35
40
45
WEEK

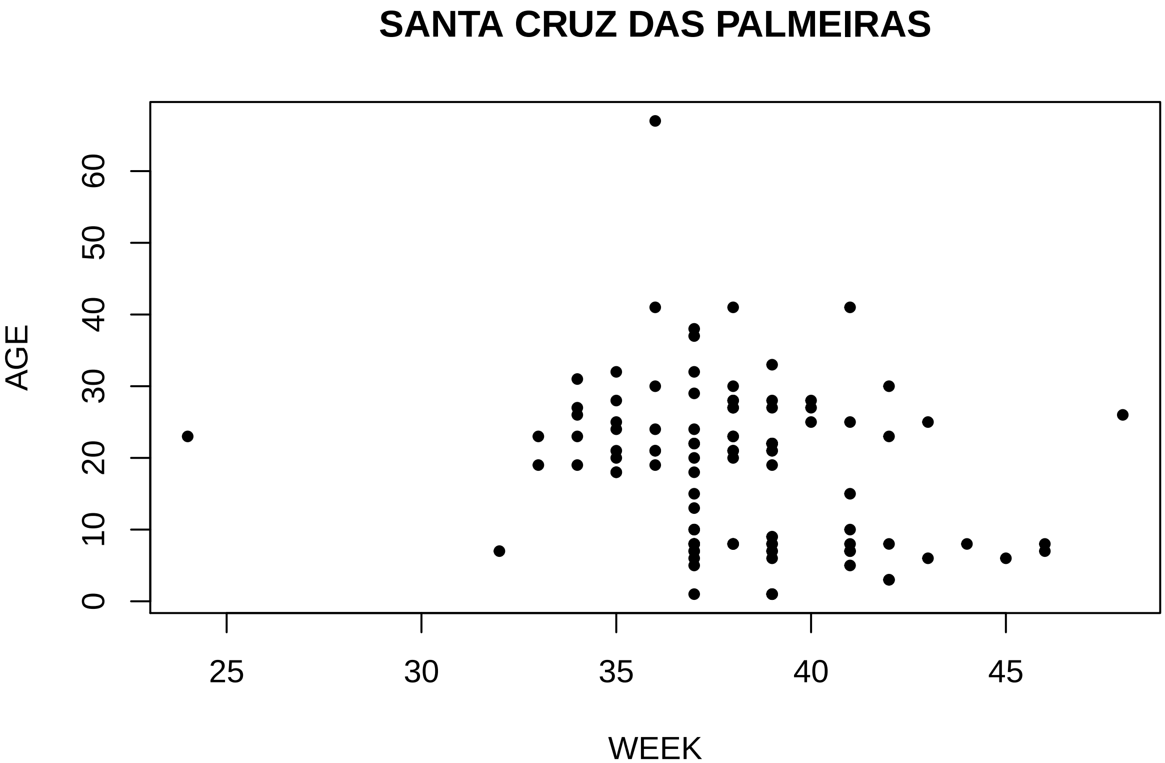

CARAPICUIBA
AGE
0
10
20
30
40
15
20
25
30
35
40
45
WEEK

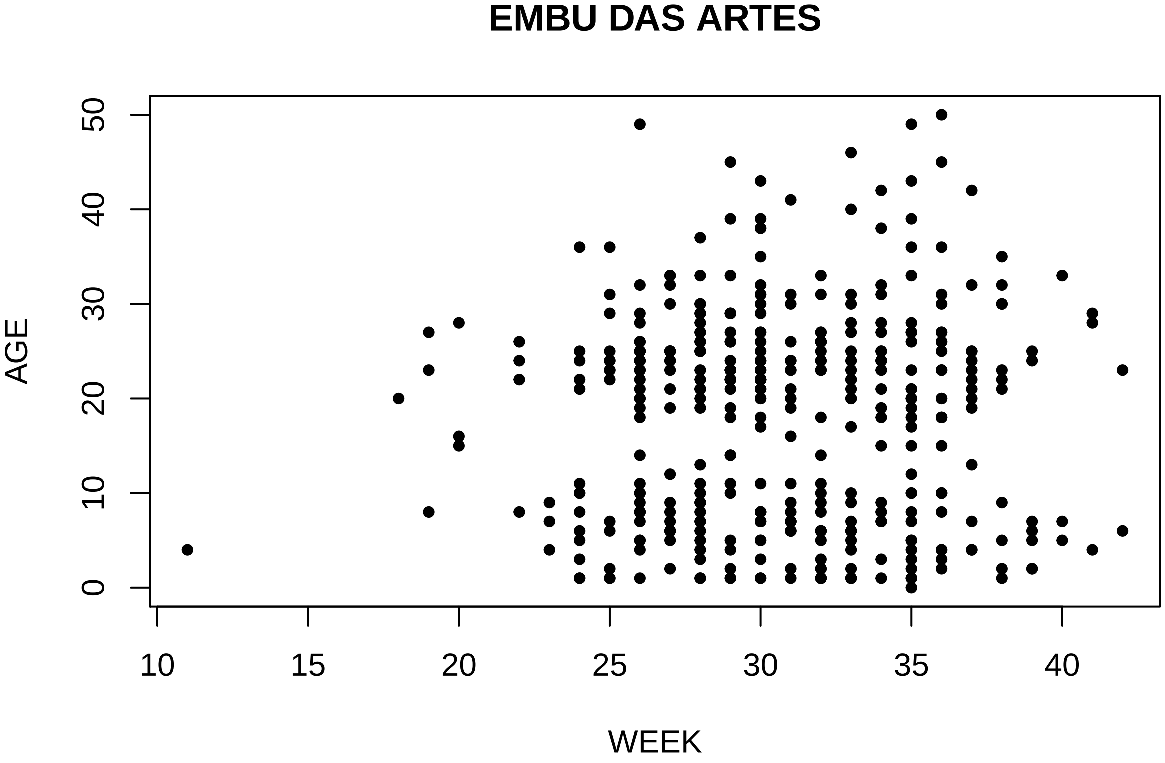

SAO JOSE DO RIO PRETO
AGE
0
5
10
15
20
25
30
35
25
30
35
40
45
WEEK

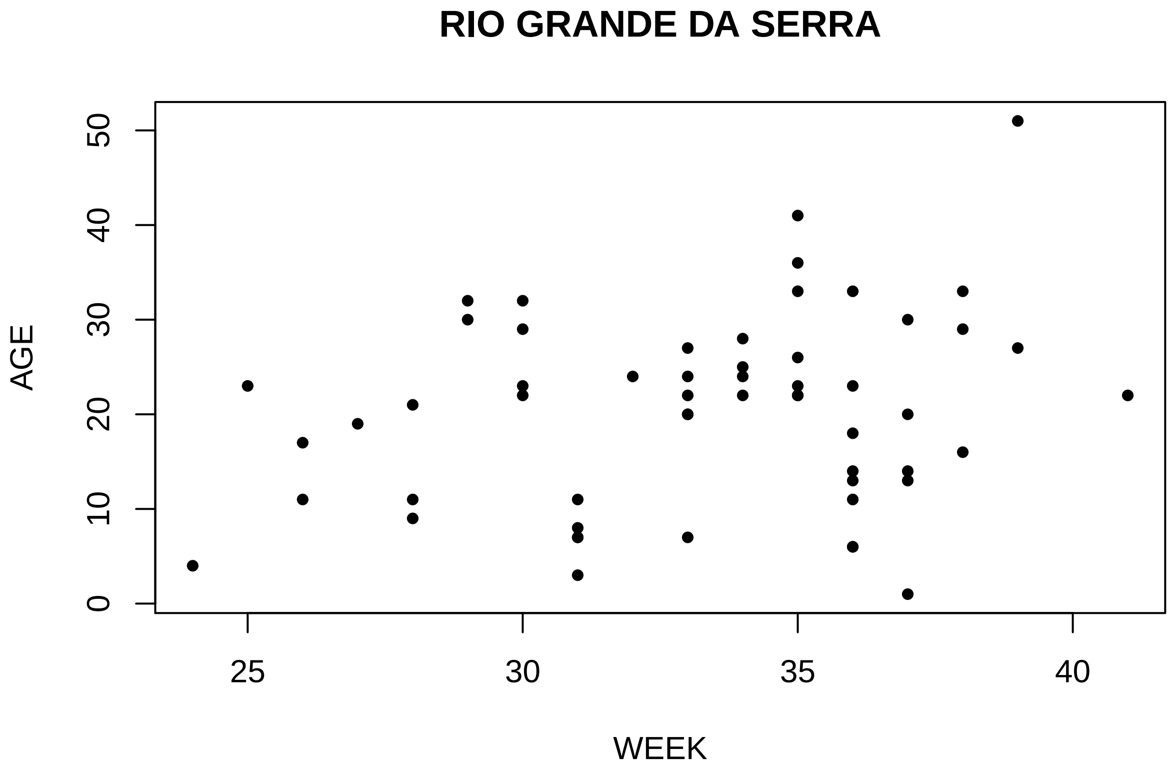

EMBU−GUACU
AGE
0
10
20
30
40
50
60
25
30
35
WEEK

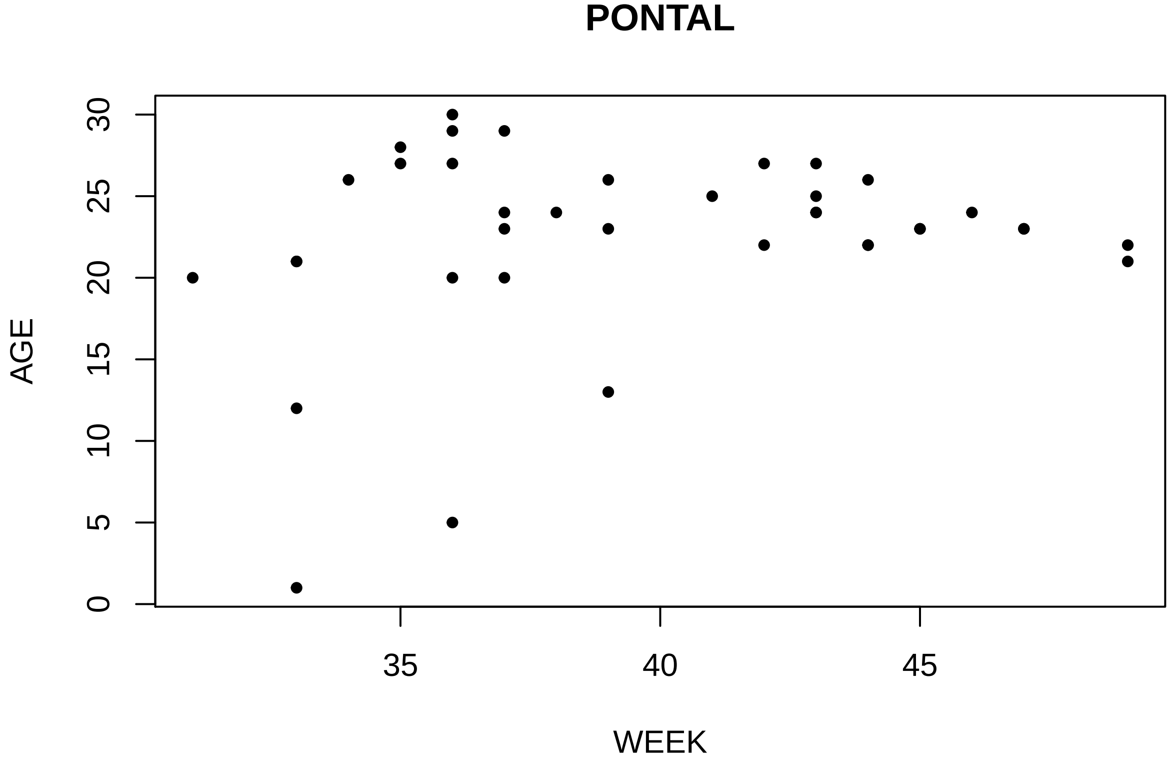

CATANDUVA
AGE
0
20
40
60
80
20
25
30
35
40
45
WEEK

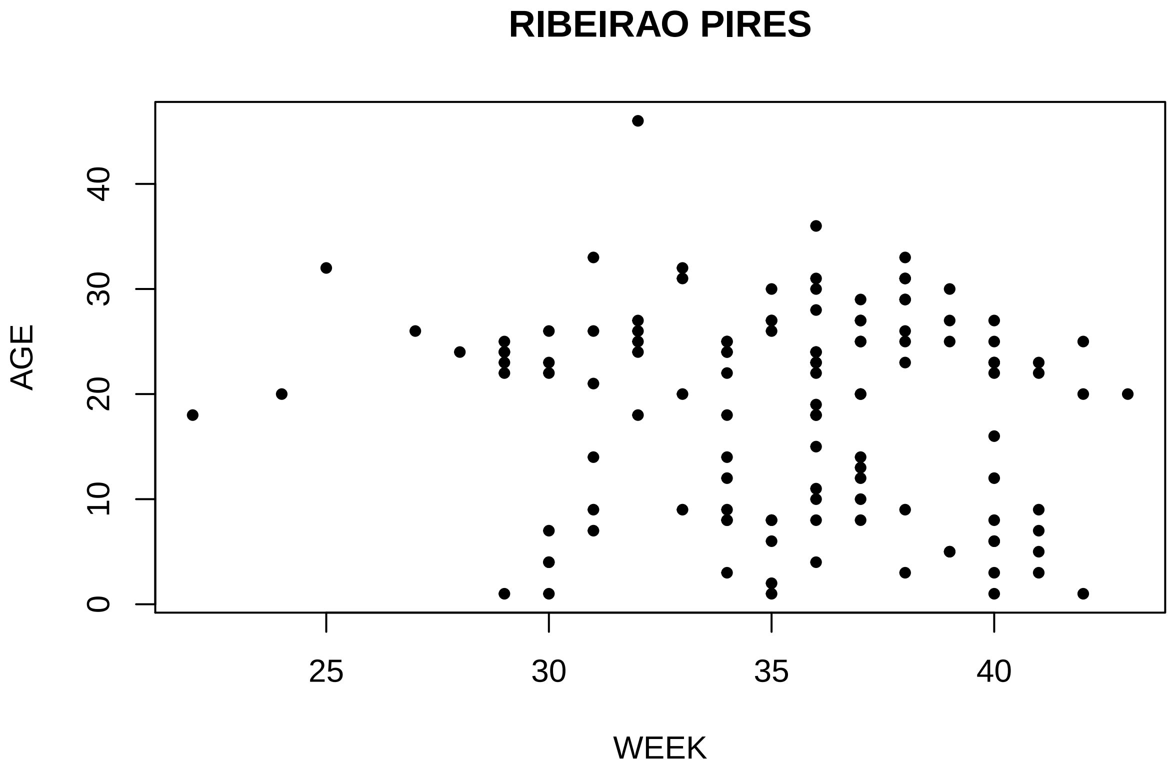

SAO JOSE DOS CAMPOS
AGE
0
10
20
30
40
50
0
10
20
30
40
50
WEEK

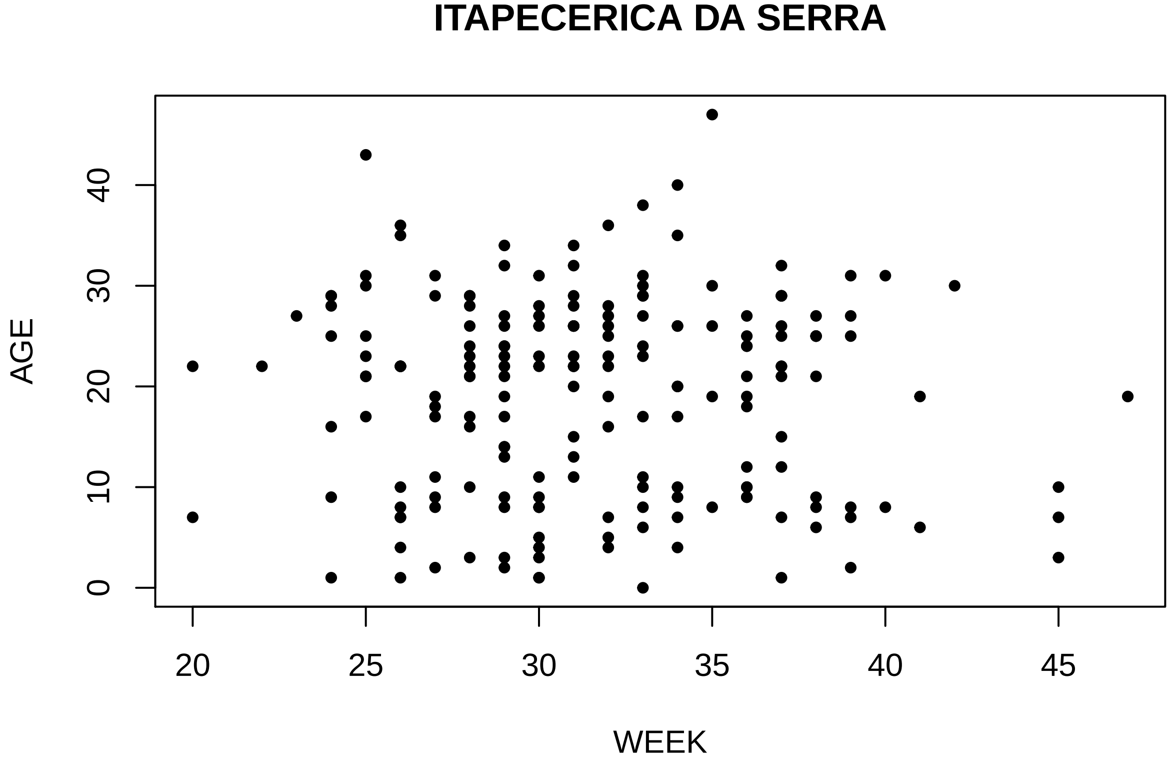

SANTA CRUZ DAS PALMEIRAS
AGE
0
10
20
30
40
50
60
25
30
35
40
45
WEEK


EMBU DAS ARTES
AGE
0
10
20
30
40
50
10
15
20
25
30
35
40
WEEK

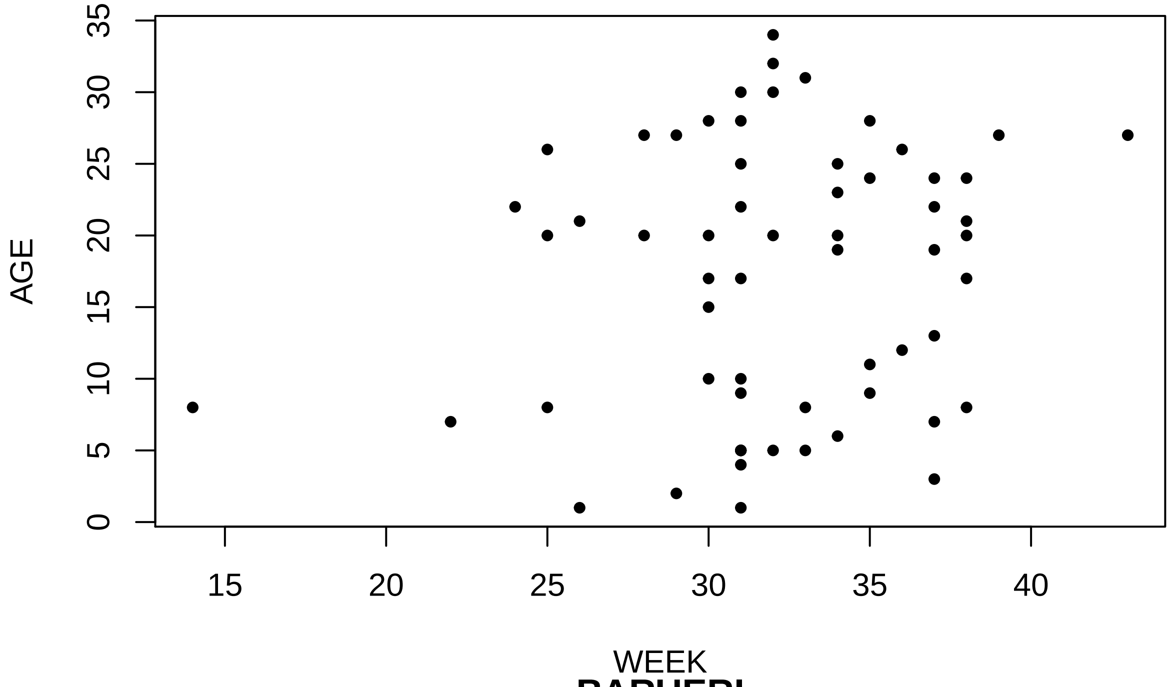

RIO GRANDE DA SERRA
AGE
0
10
20
30
40
50
25
30
35
40
WEEK


PONTAL
AGE
0
5
10
15
20
25
30
35
40
45
WEEK

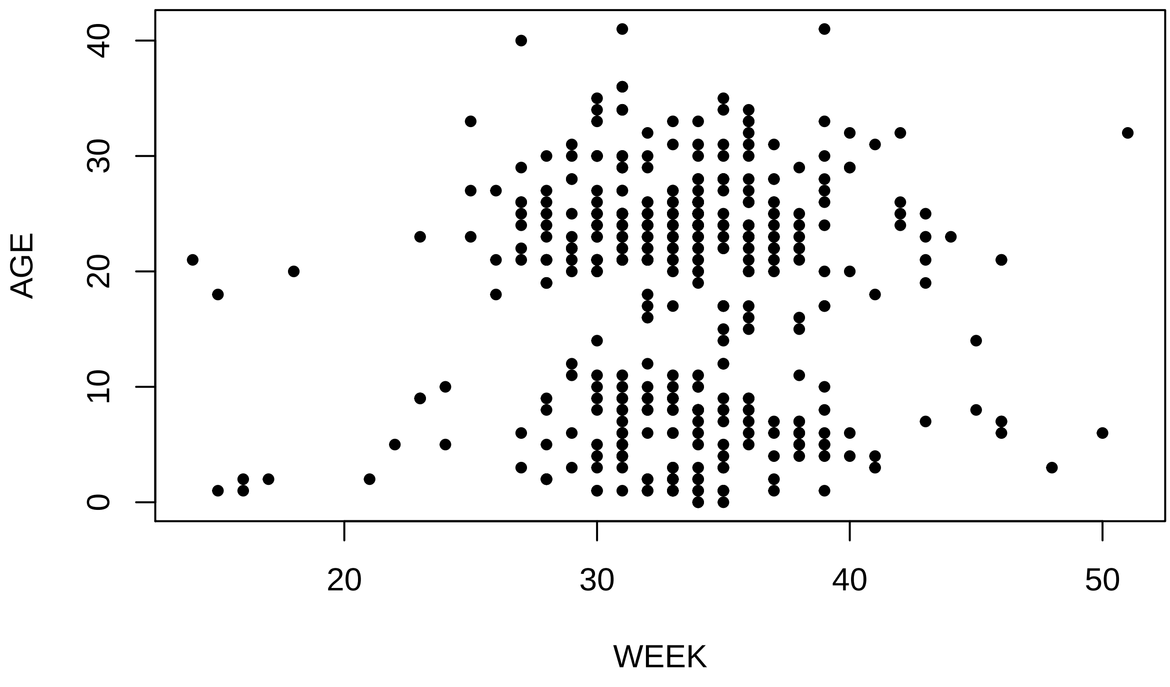

RIBEIRAO PIRES
AGE
0
10
20
30
40
WEEK
25
30
35
40

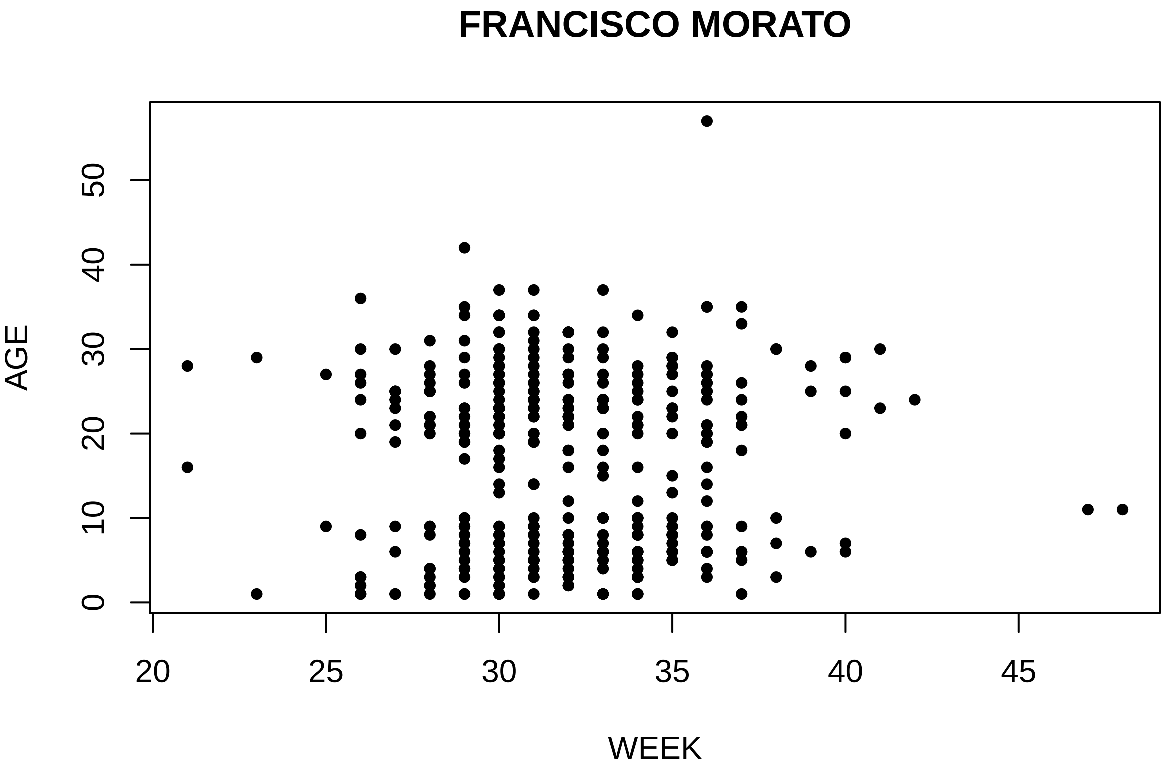

ITAPECERICA DA SERRA
AGE
0
10
20
30
40
WEEK
20
25
30
35
40
45

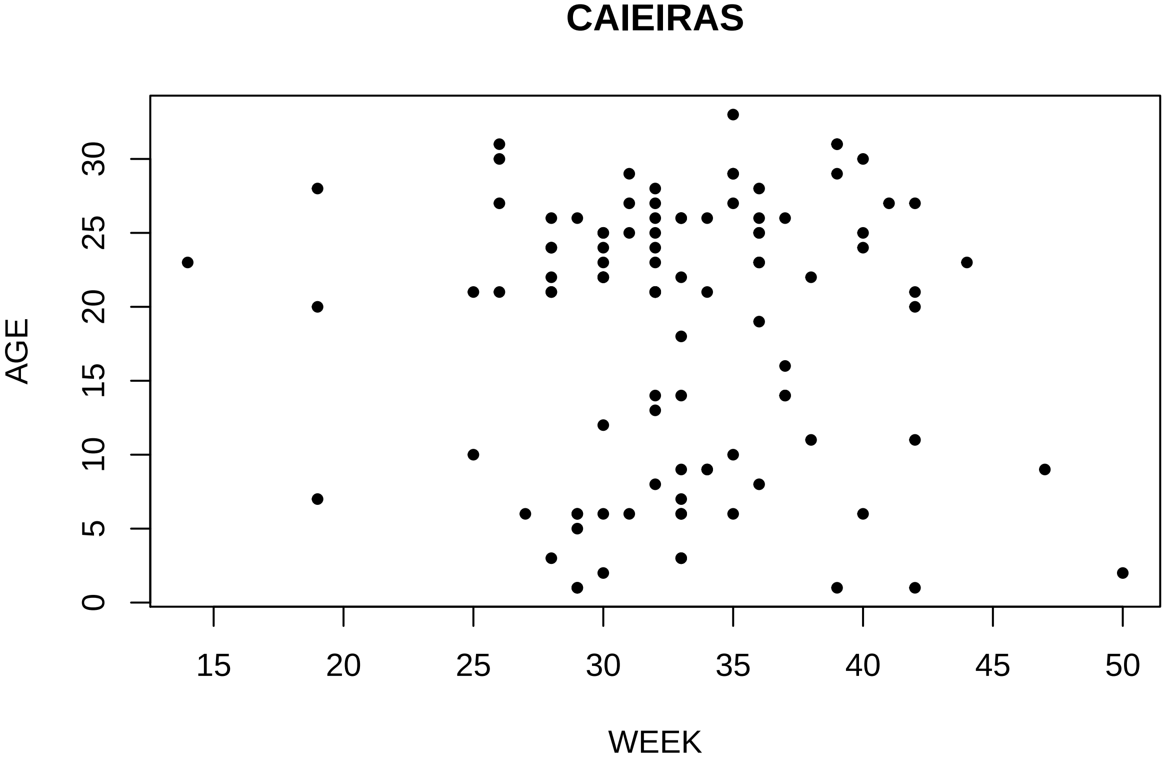
SANTANA DE PARNAIBA

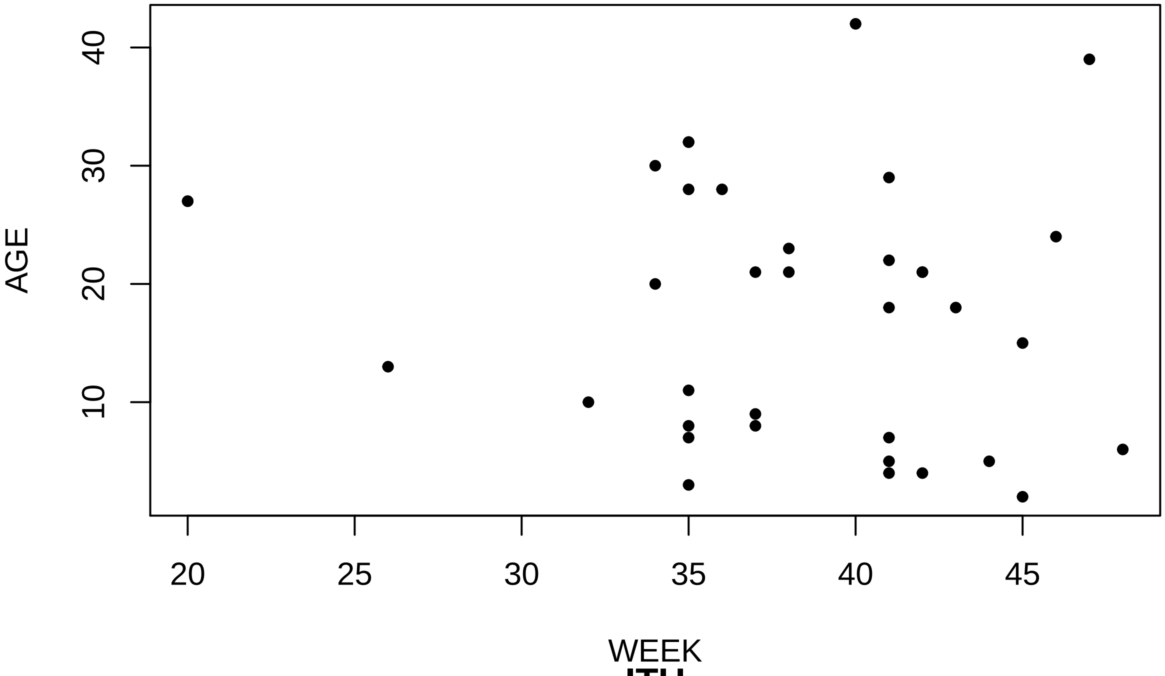


BARUERI

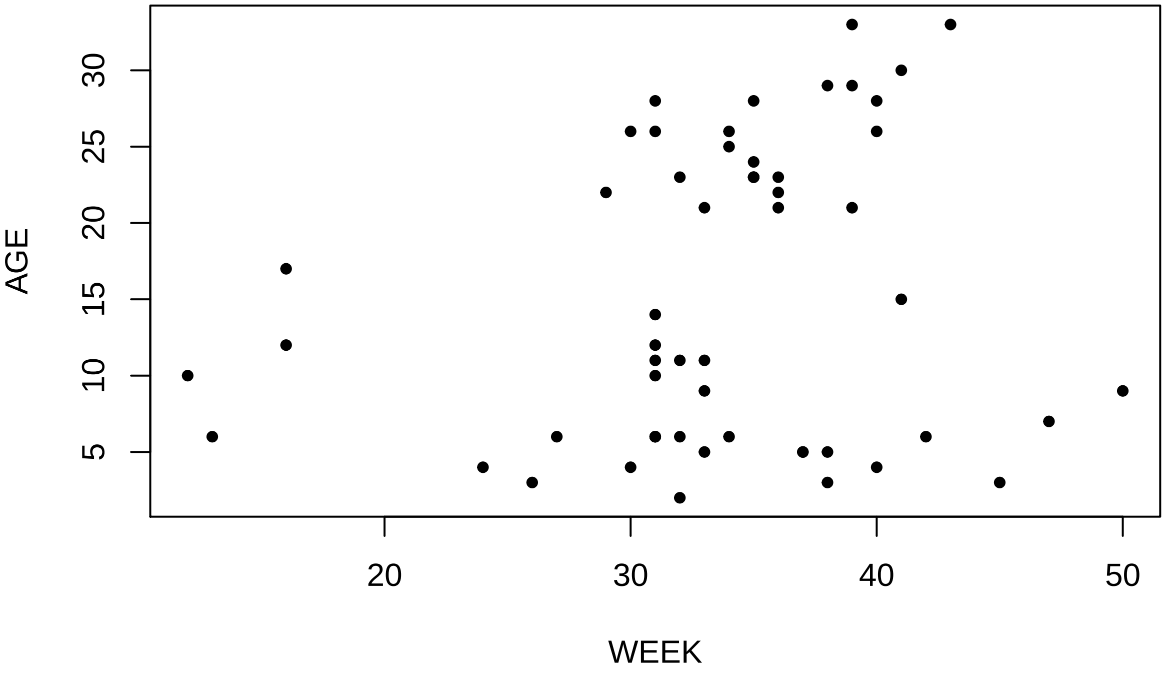



FRANCISCO MORATO
AGE
0
10
20
30
40
50
20
25
30
35
40
45
WEEK

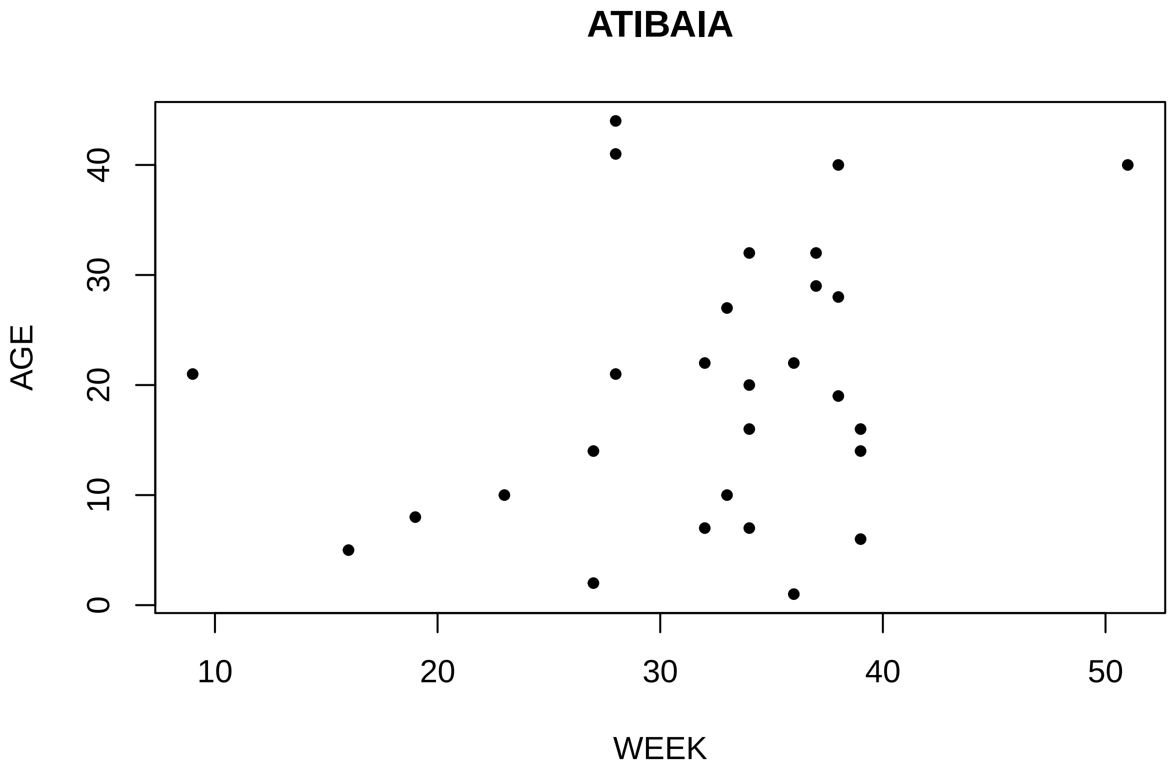

CAIEIRAS
AGE
0
5
10
15
20
25
30
15
20
25
30
35
40
45
50
WEEK

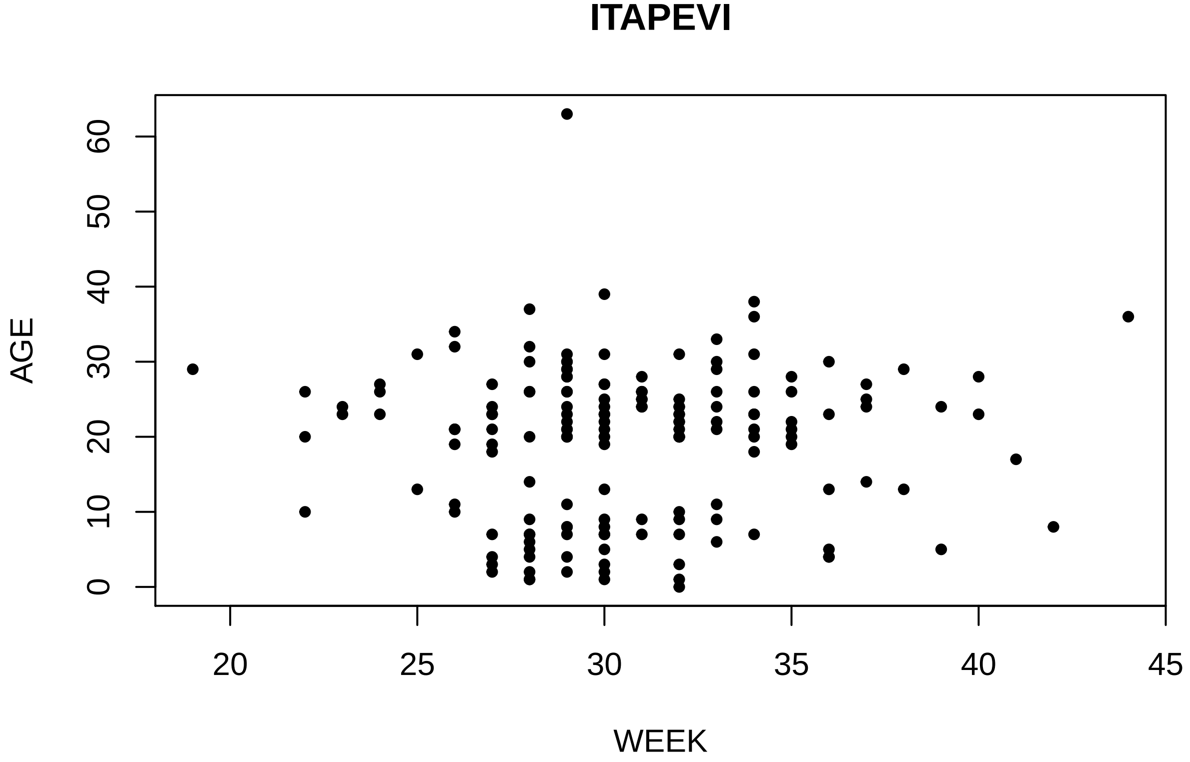

JACAREI

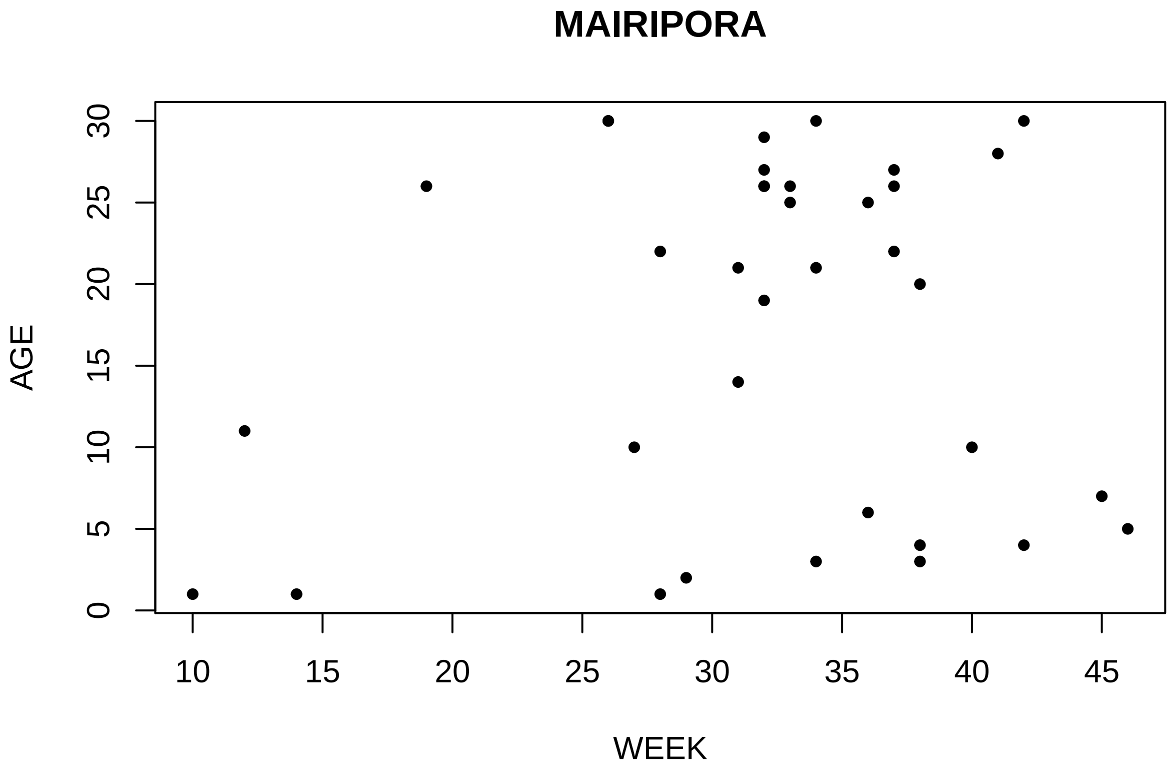
AGE
10
20
30
40
20
25
30
35
40
45
WEEK

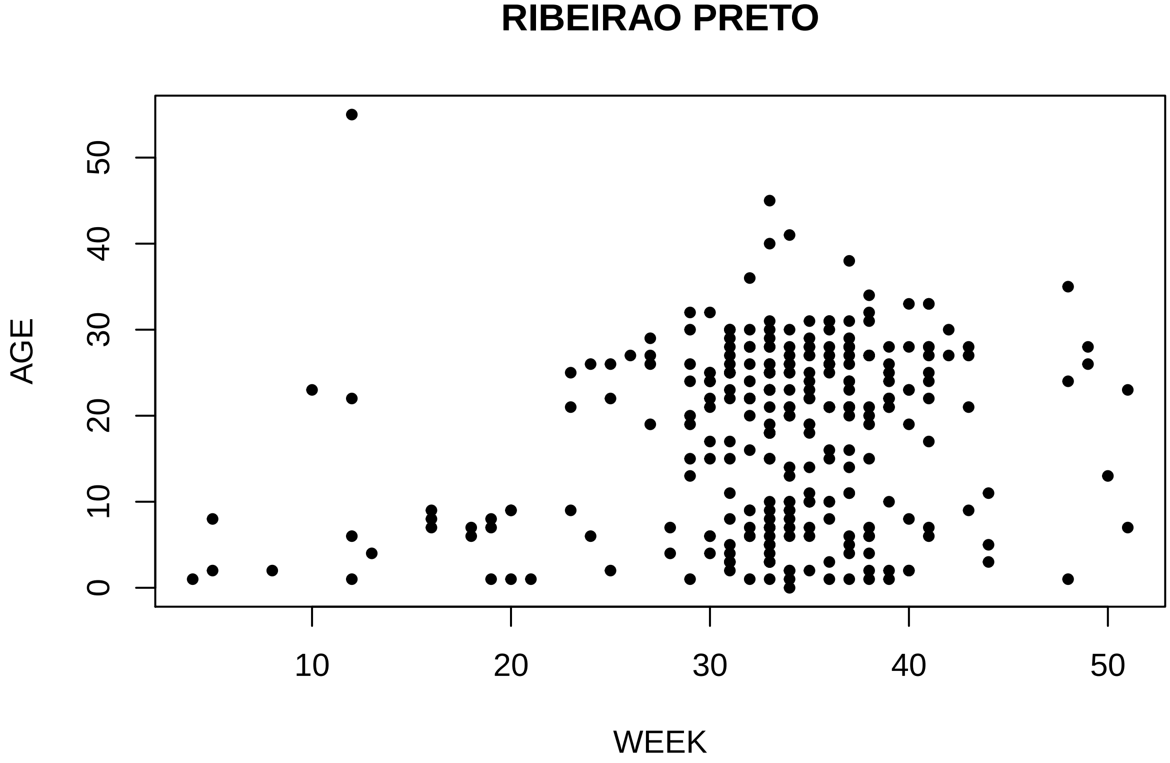

ITU

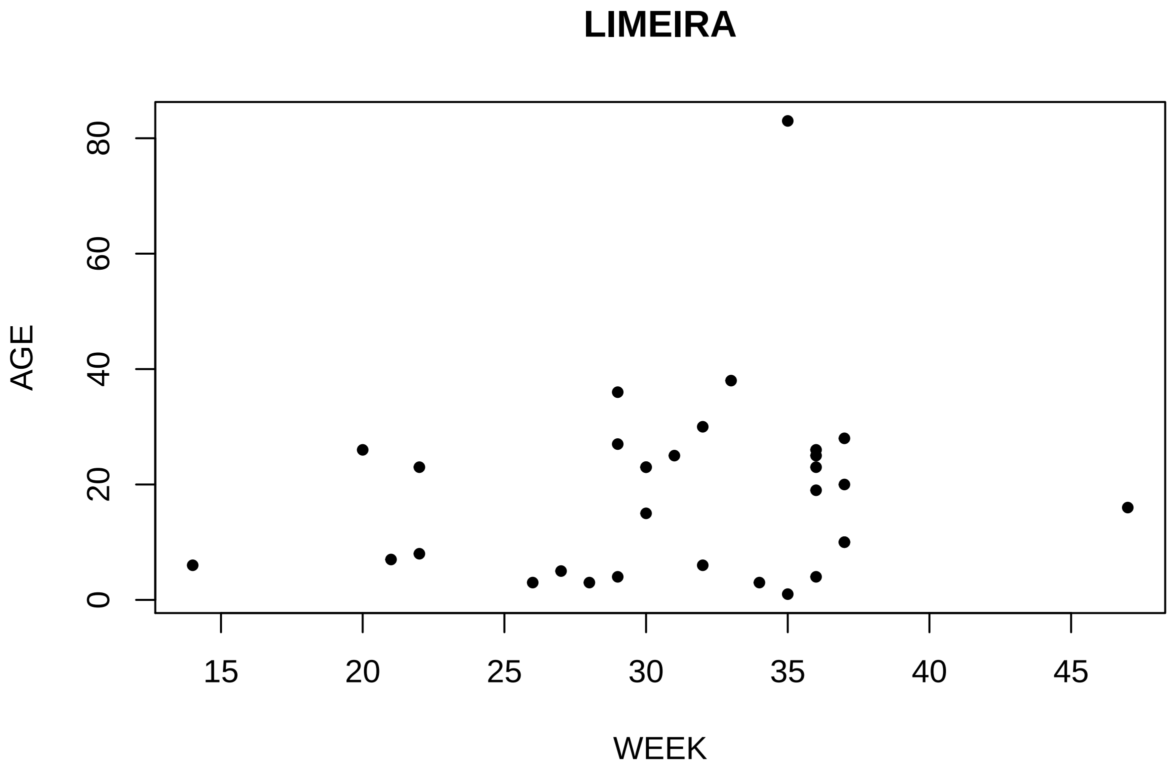
AGE
5
10
15
20
25
30
20
30
40
50
WEEK

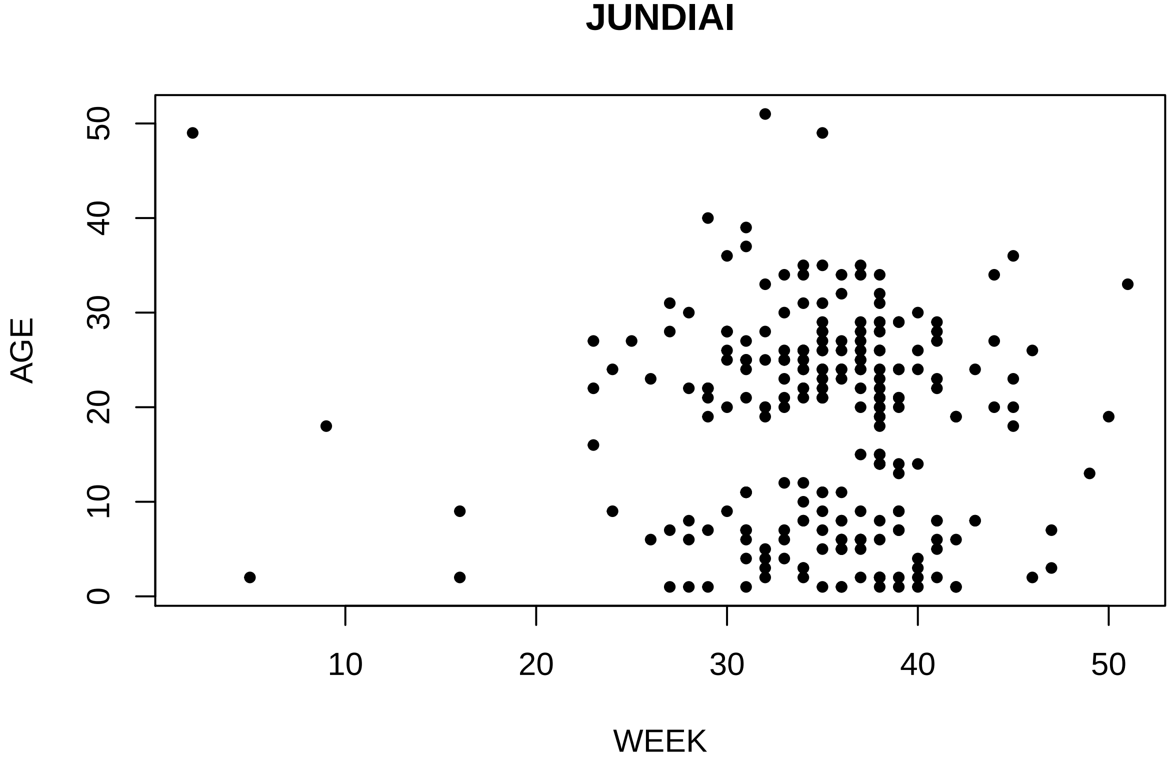

ATIBAIA
AGE
0
10
20
30
40
10
20
30
40
50
WEEK

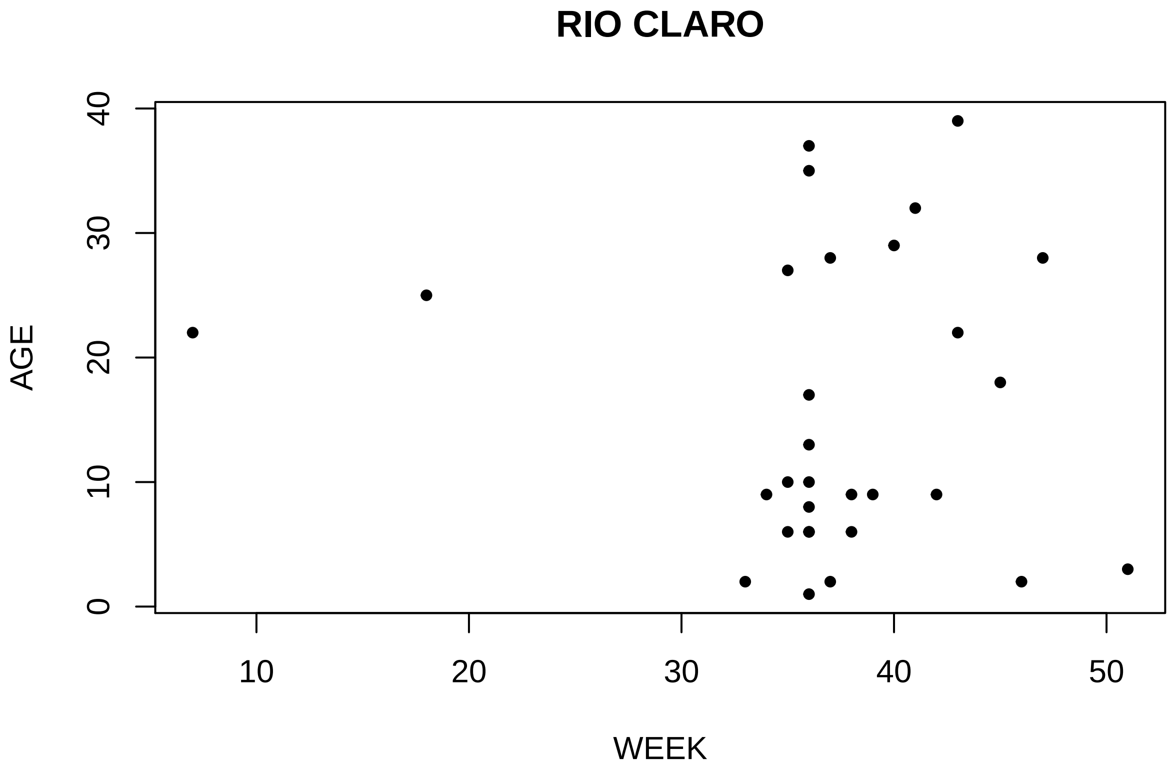

ITAPEVI
AGE
0
10
20
30
40
50
60
20
25
30
35
40
45
WEEK

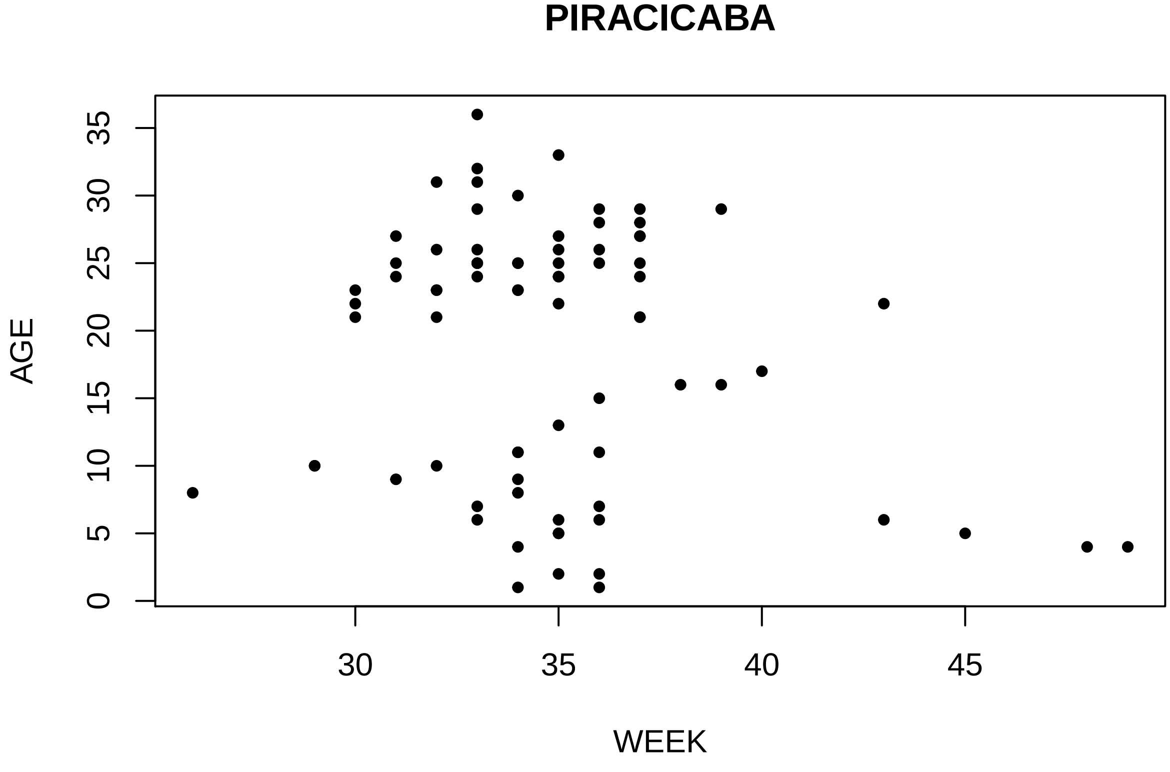

MAIRIPORA
AGE
0
5
10
15
20
25
30
10
15
20
25
30
35
40
45
WEEK

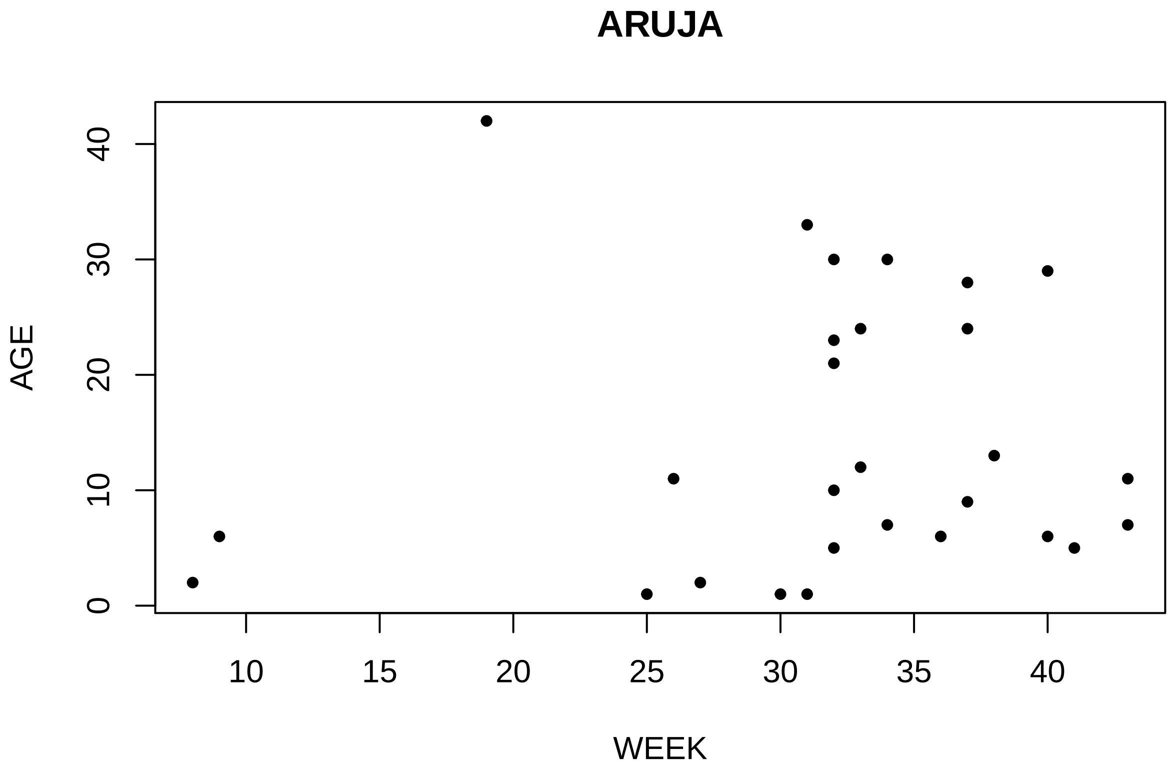

RIBEIRAO PRETO
AGE
0
10
20
30
40
50
10
20
30
40
50
WEEK

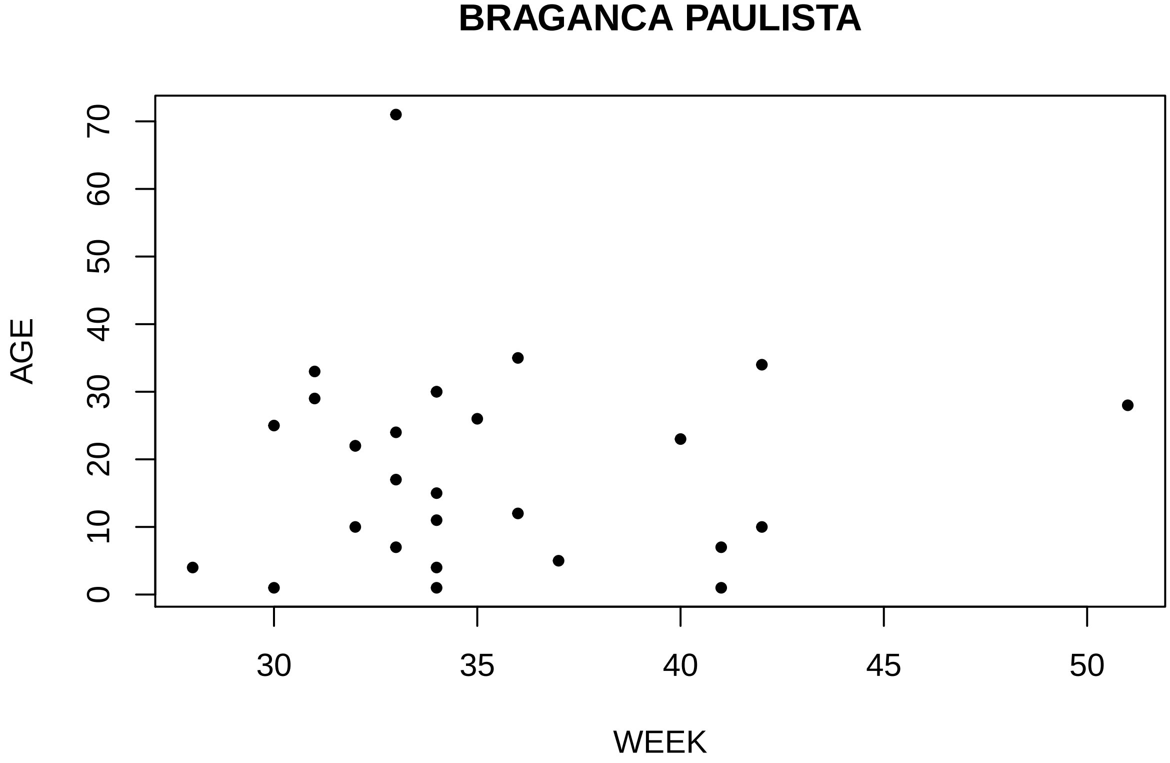

LIMEIRA
AGE
0
20
40
60
80
WEEK
15
20
25
30
35
40
45

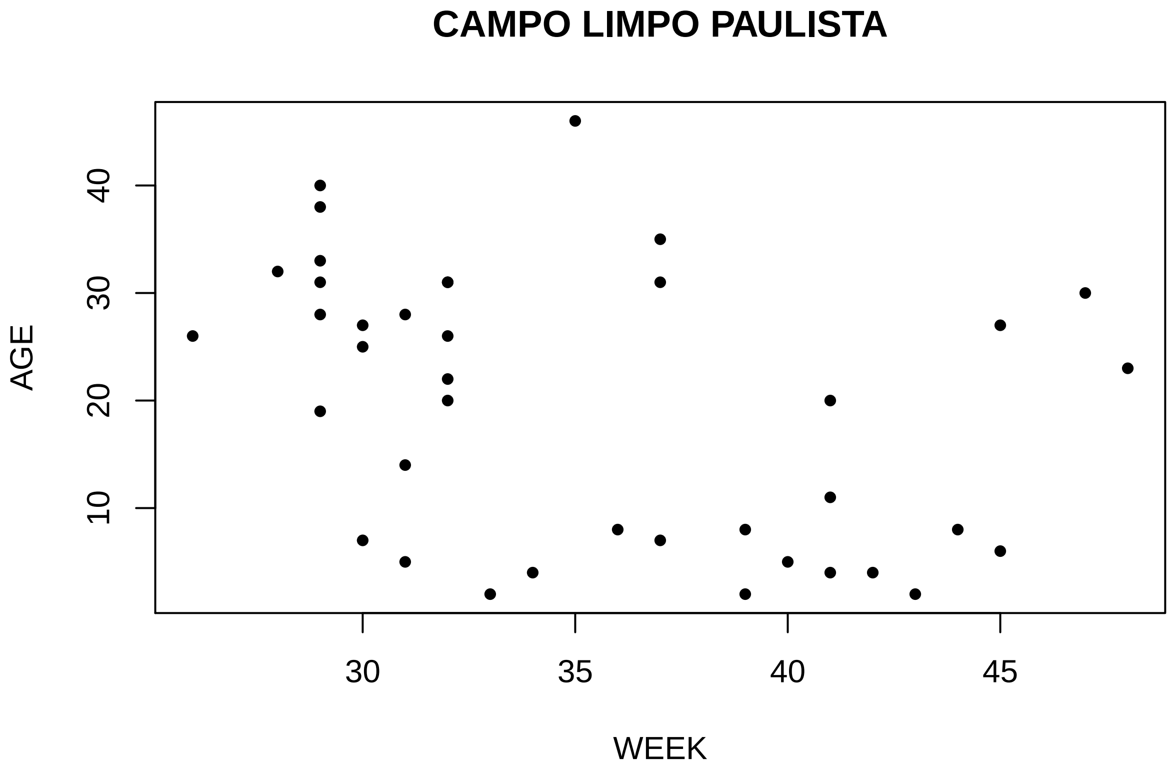

JUNDIAI
AGE
0
10
20
30
40
50
WEEK
10
20
30
40
50


RIO CLARO

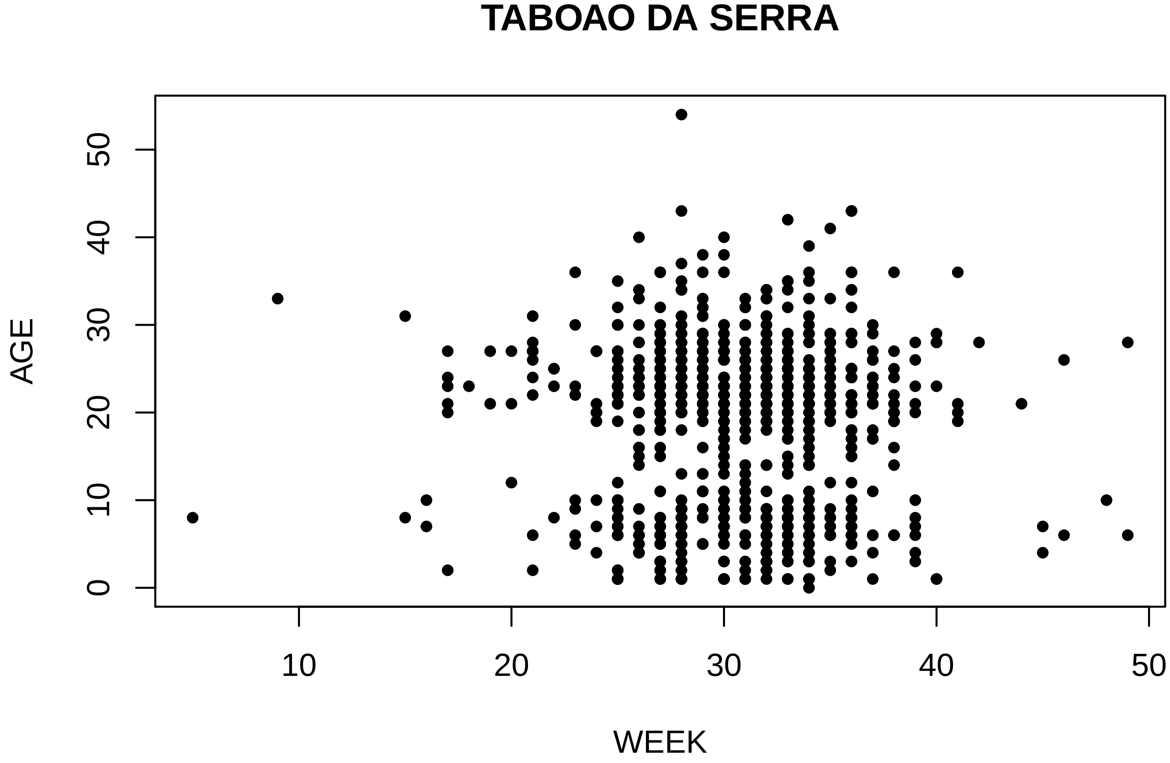

PIRACICABA


ARUJA
AGE
0
10
20
30
40
10
15
20
25
30
35
40
WEEK

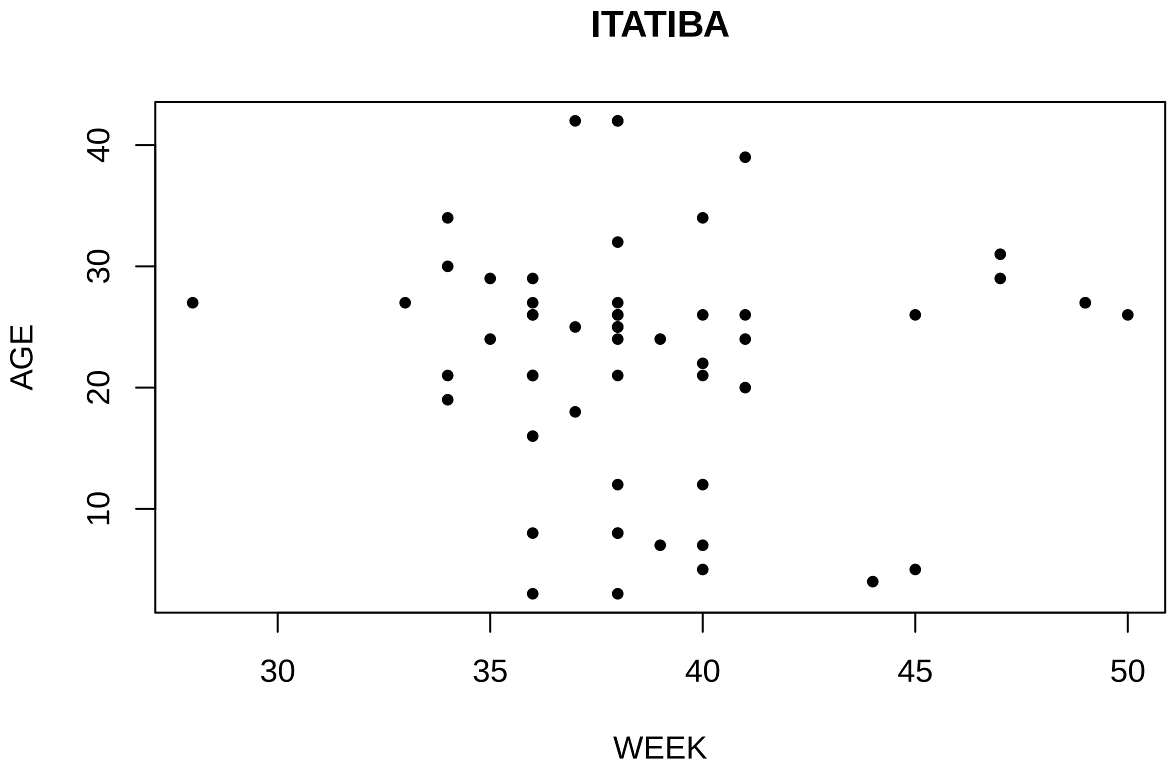

BRAGANCA PAULISTA
AGE
0
10
20
30
40
50
60
70
30
35
40
45
50
WEEK

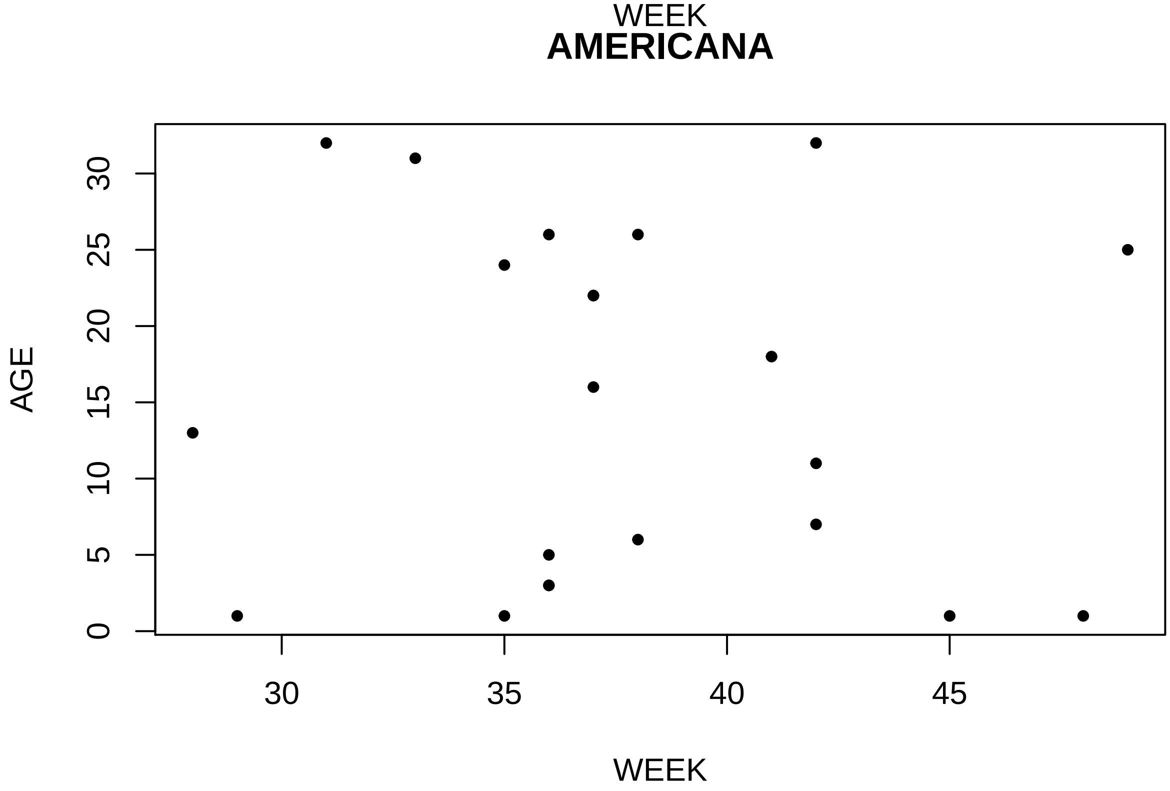


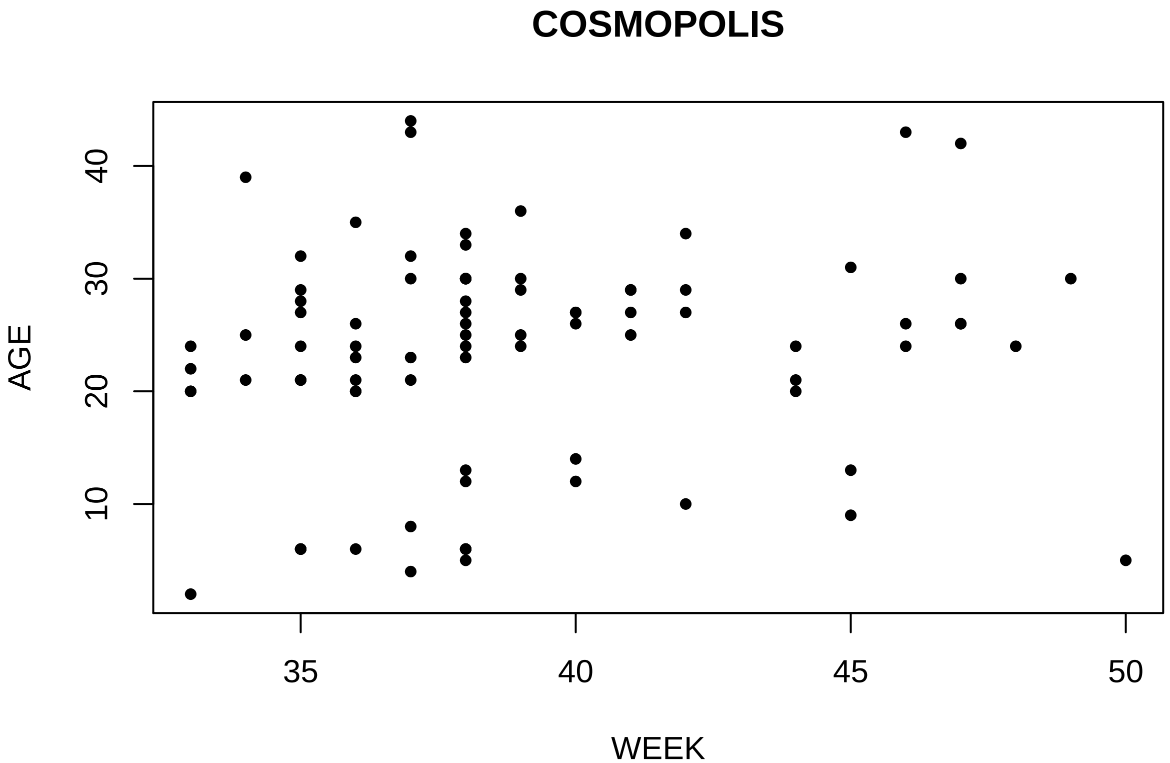
CAMPO LIMPO PAULISTA

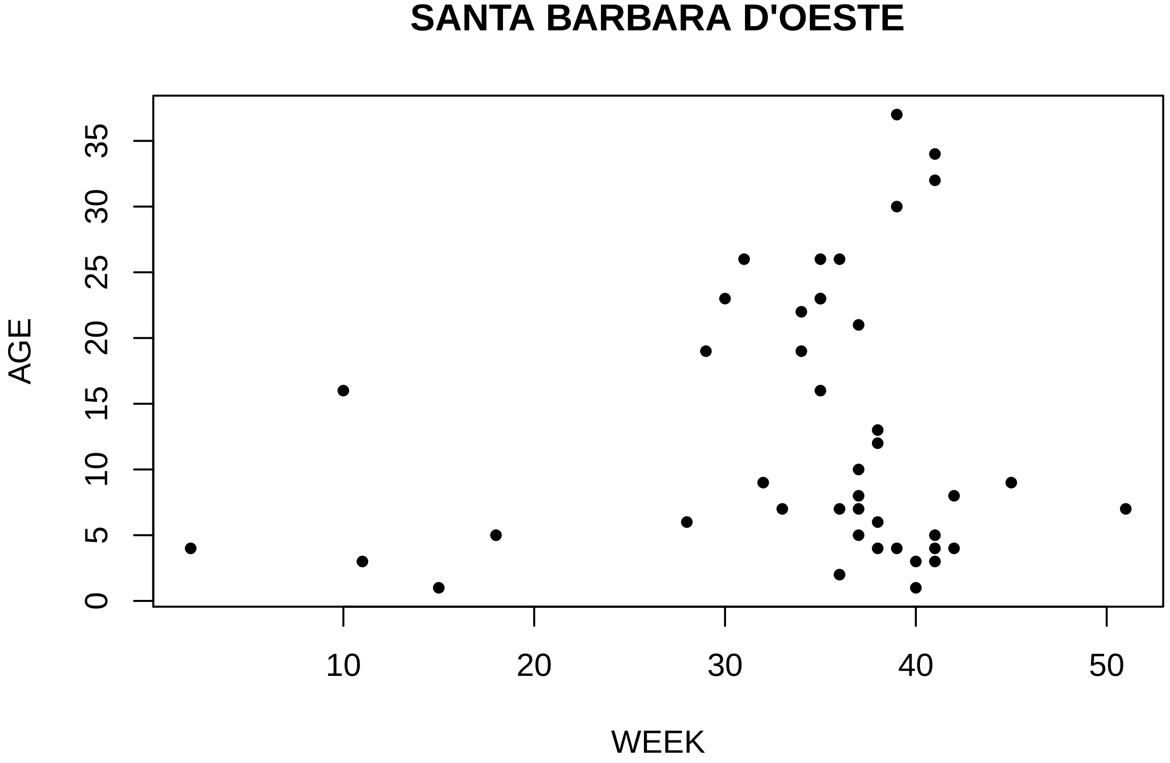


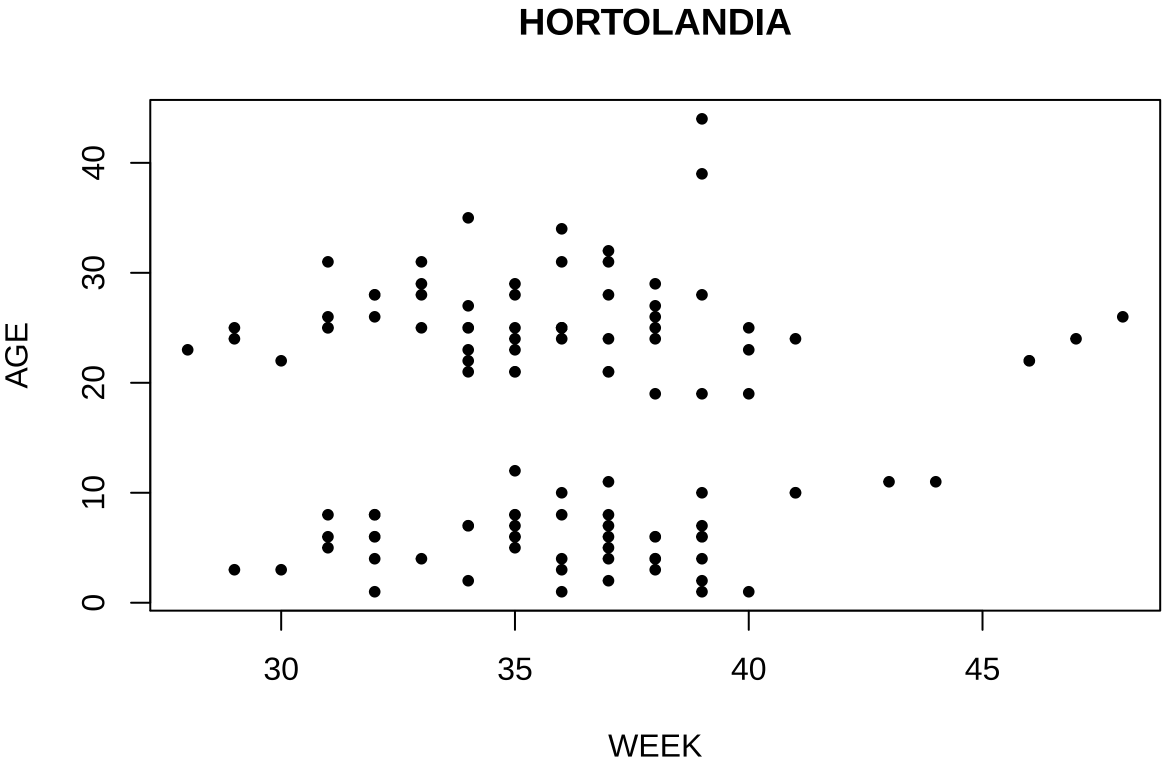
TABOAO DA SERRA

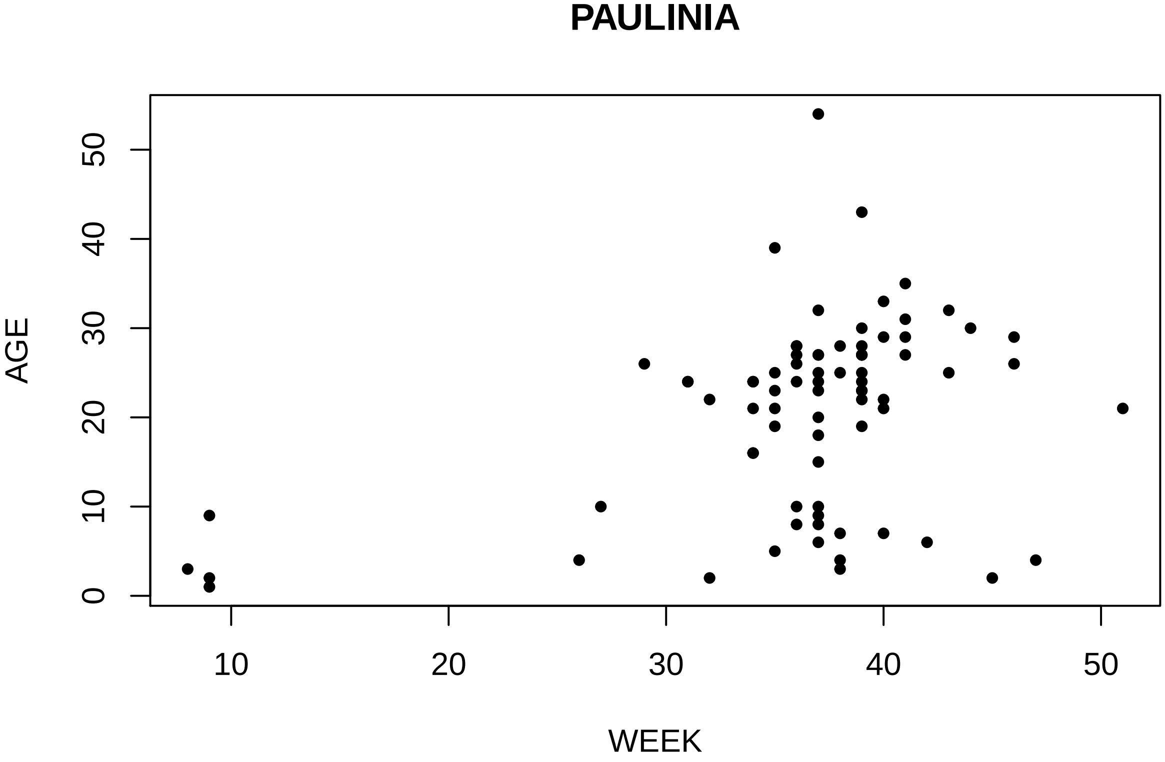

ITATIBA

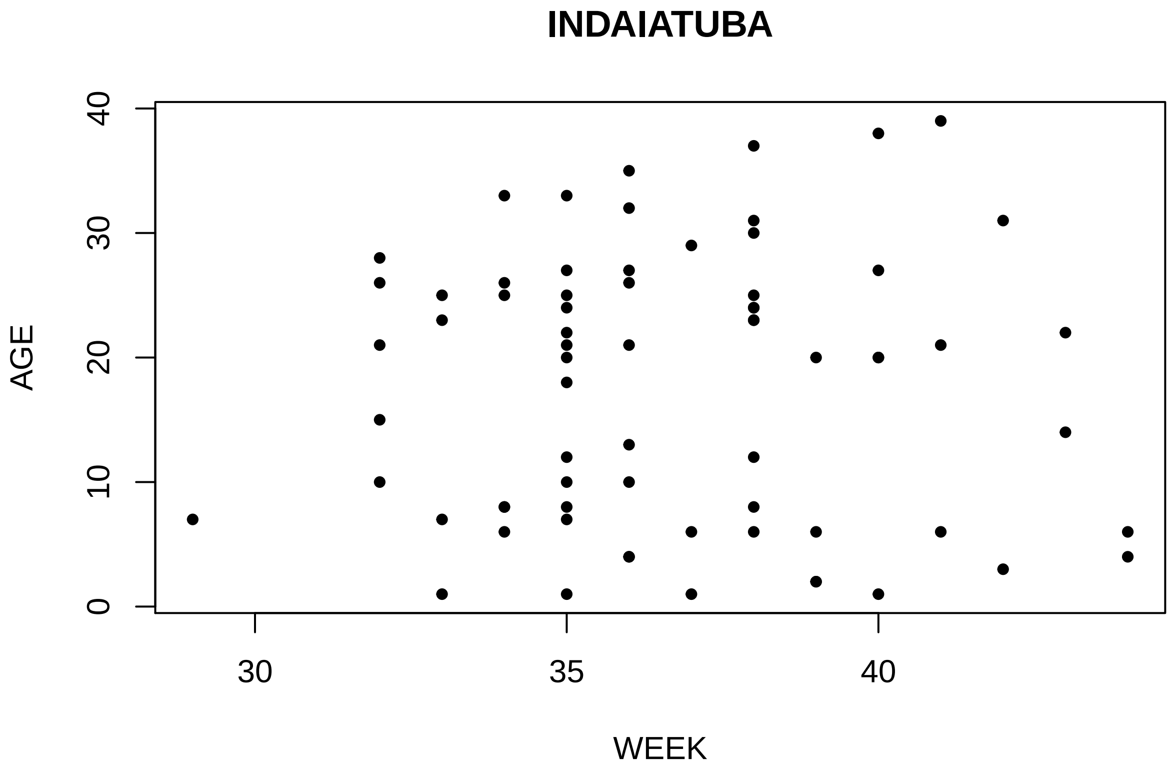

AMERICANA

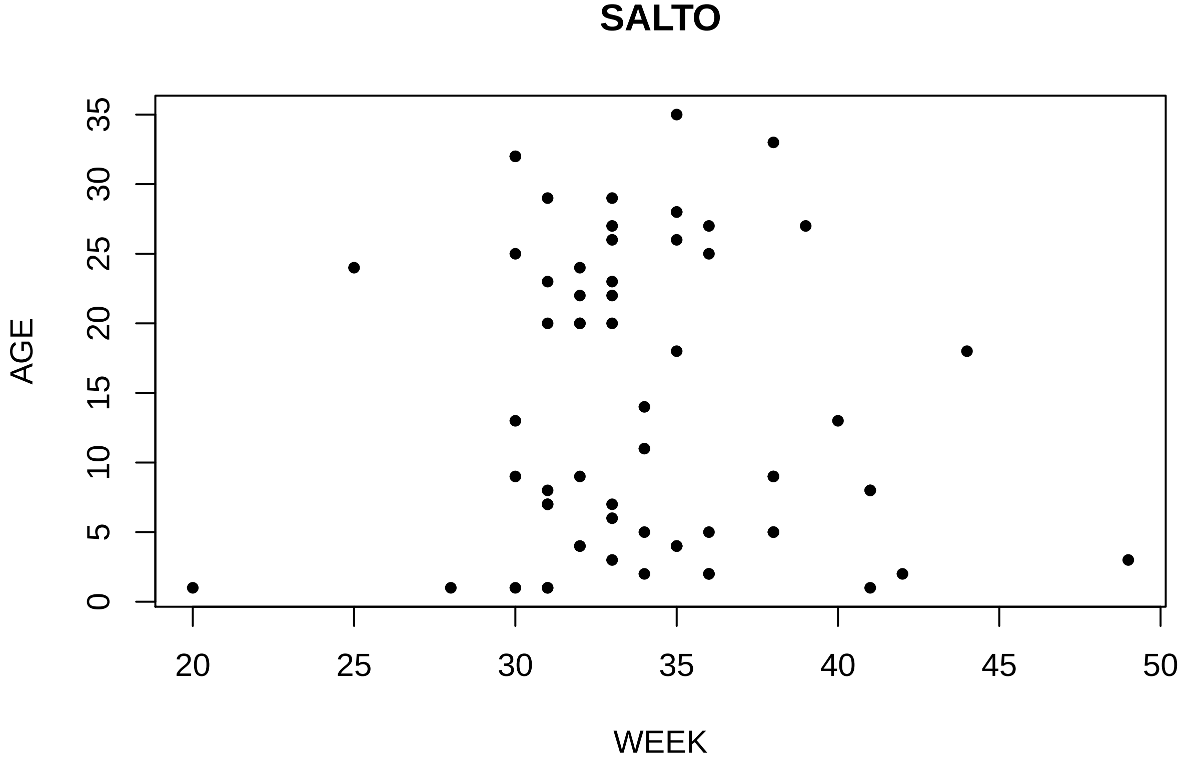

COSMOPOLIS
AGE
10
20
30
40
35
40
45
50
WEEK

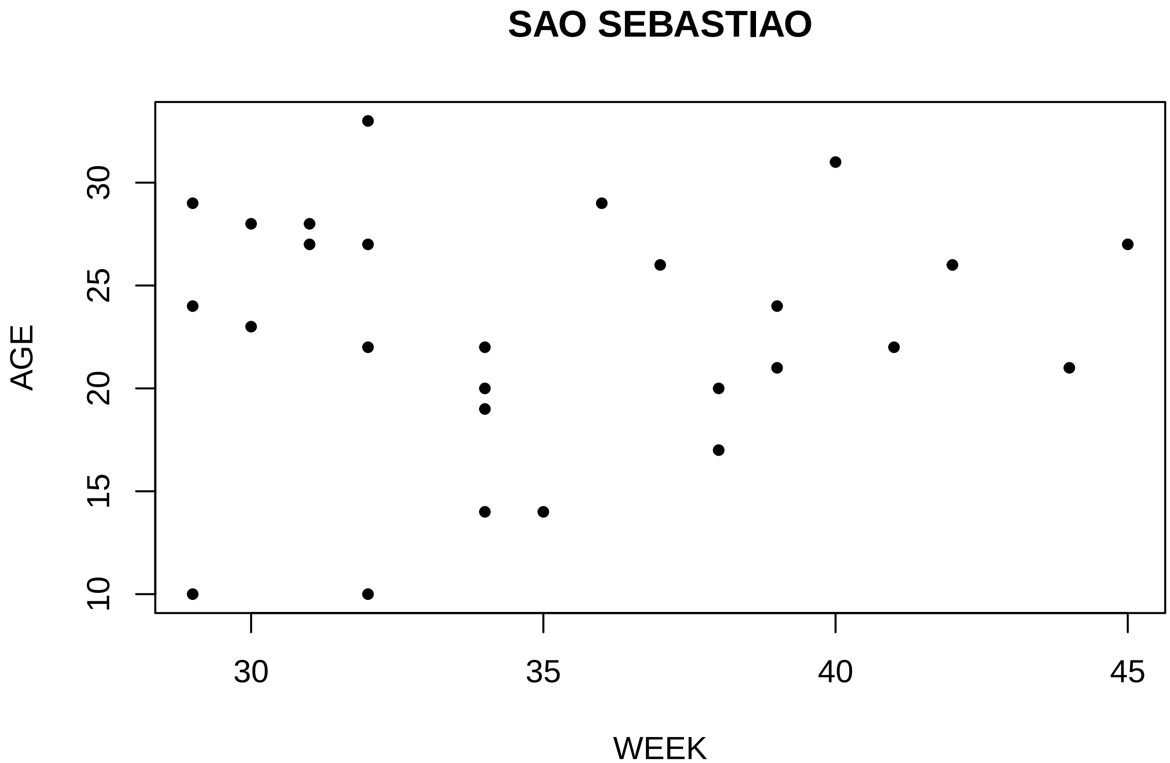

SANTA BARBARA D'OESTE
AGE
0
5
10
15
20
25
30
35
10
20
30
40
50
WEEK

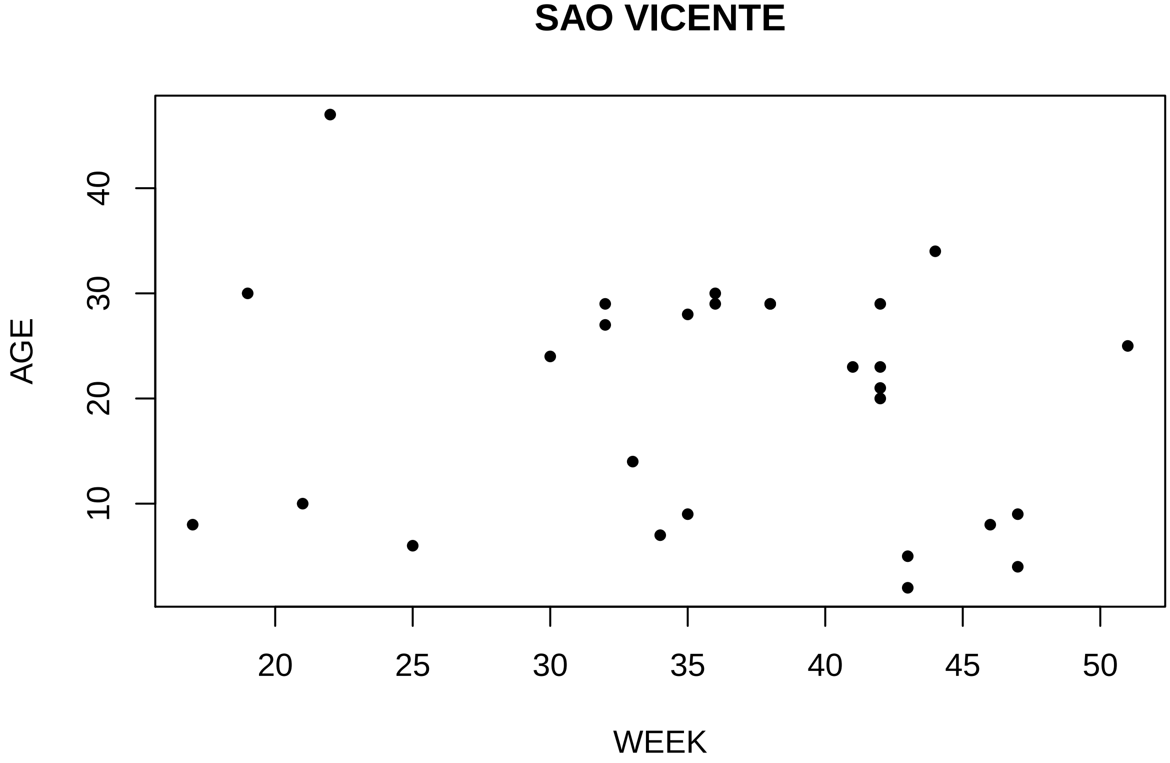

HORTOLANDIA
AGE
0
10
20
30
40
30
35
40
45
WEEK

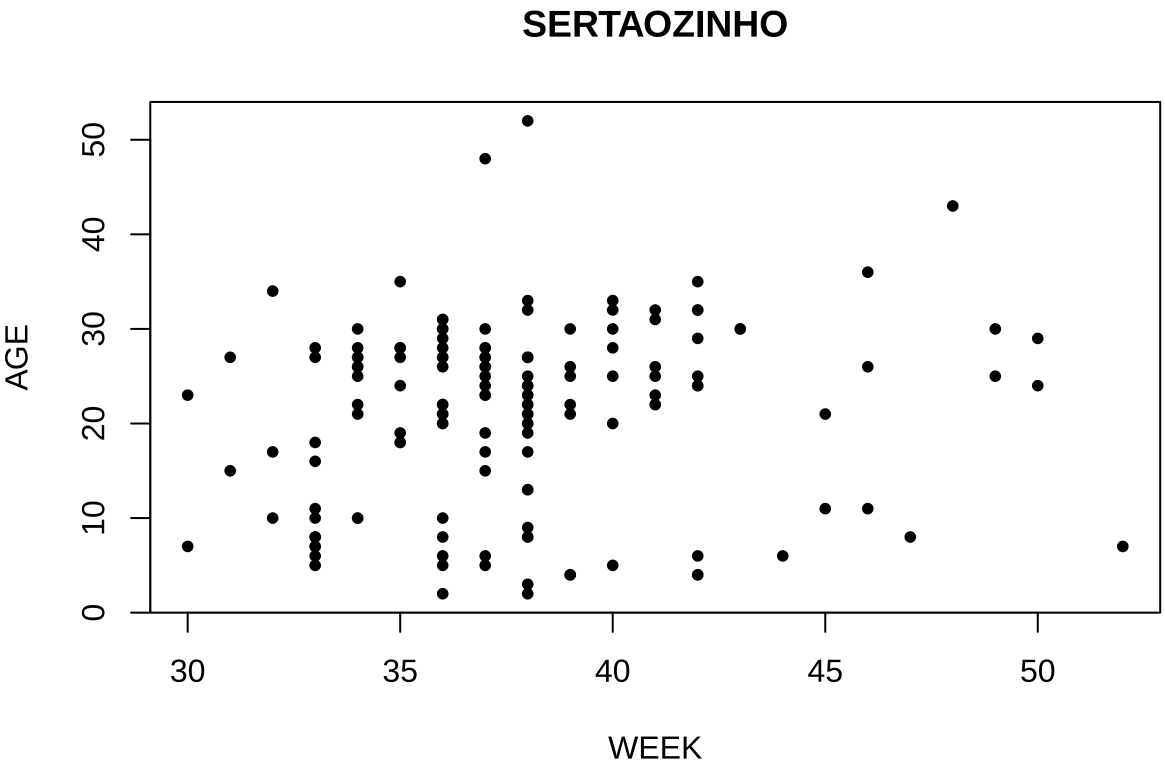

PAULINIA
AGE
0
10
20
30
40
50
10
20
30
40
50
WEEK

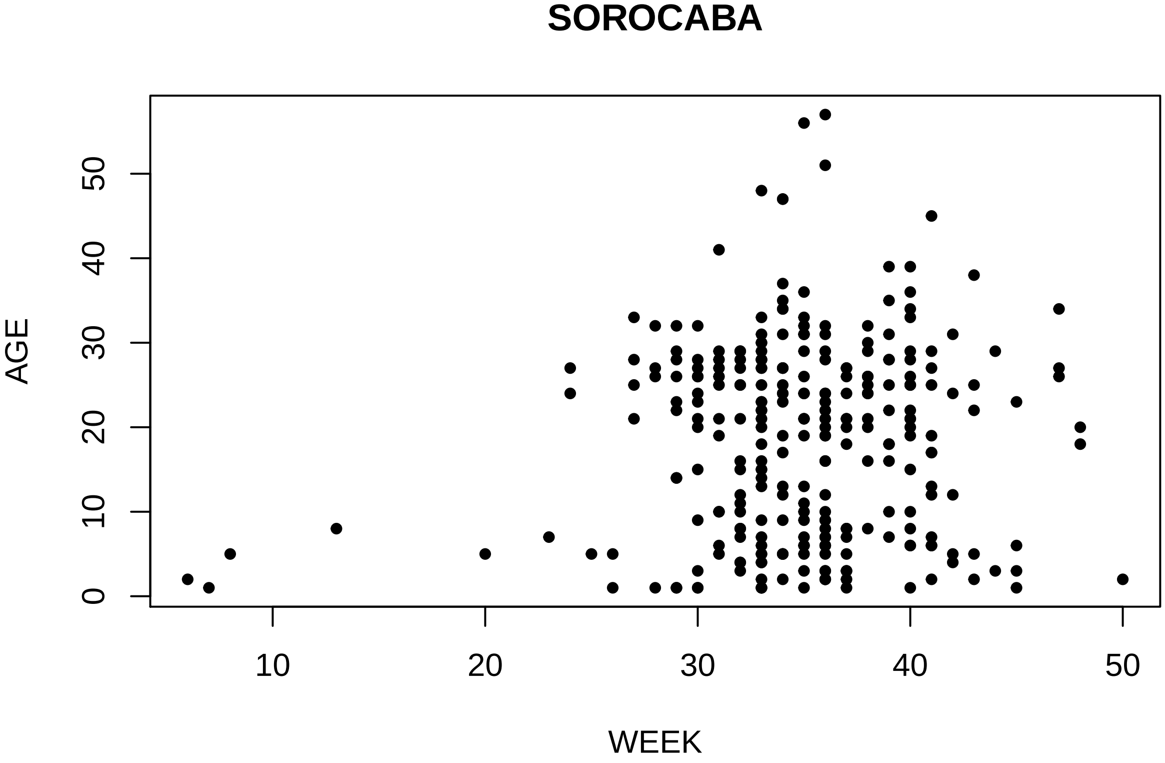

INDAIATUBA
AGE
0
10
20
30
40
30
35
40
WEEK

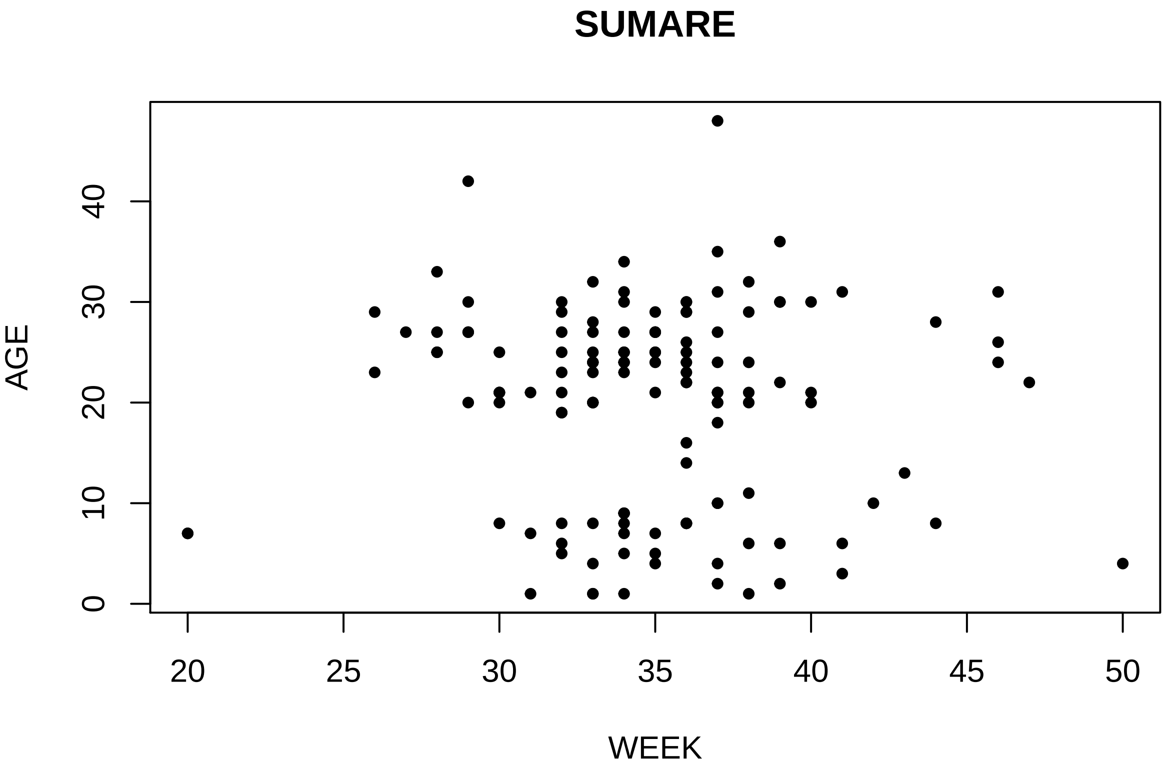

SALTO
AGE
0
5
10
15
20
25
30
35
20
25
30
35
40
45
50
WEEK

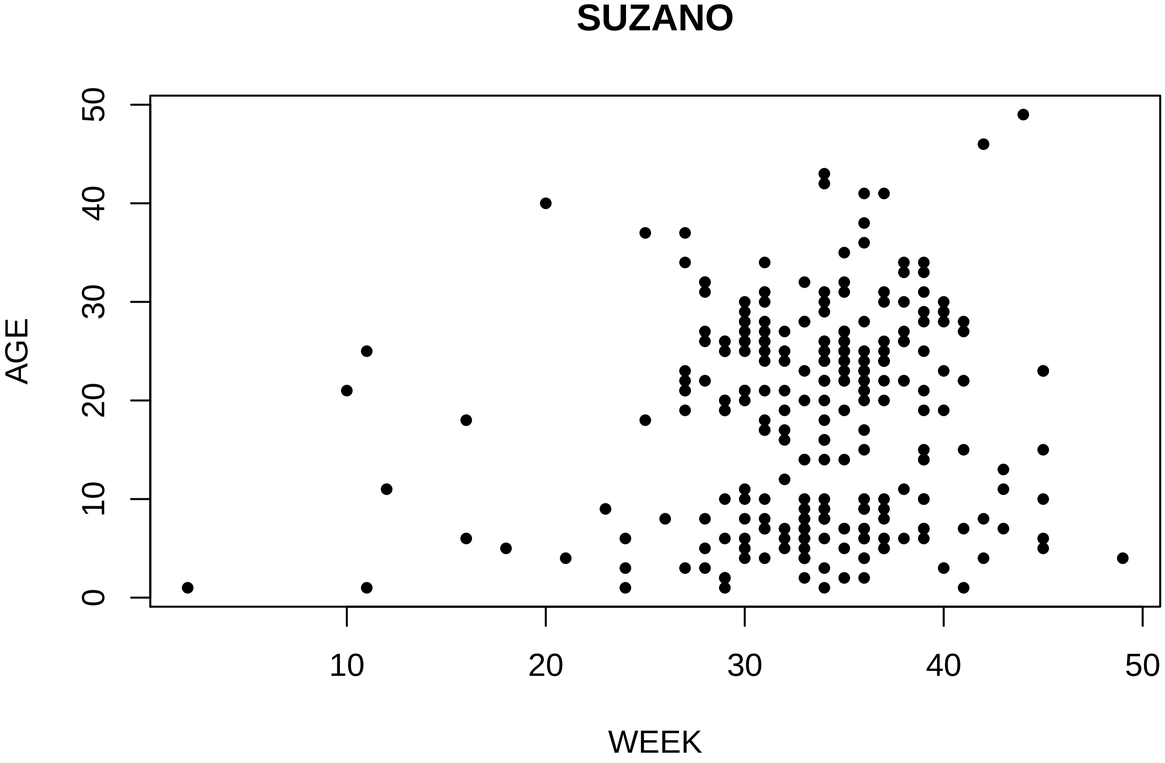

SAO SEBASTIAO
AGE
10
15
20
25
30
30
35
40
45
WEEK

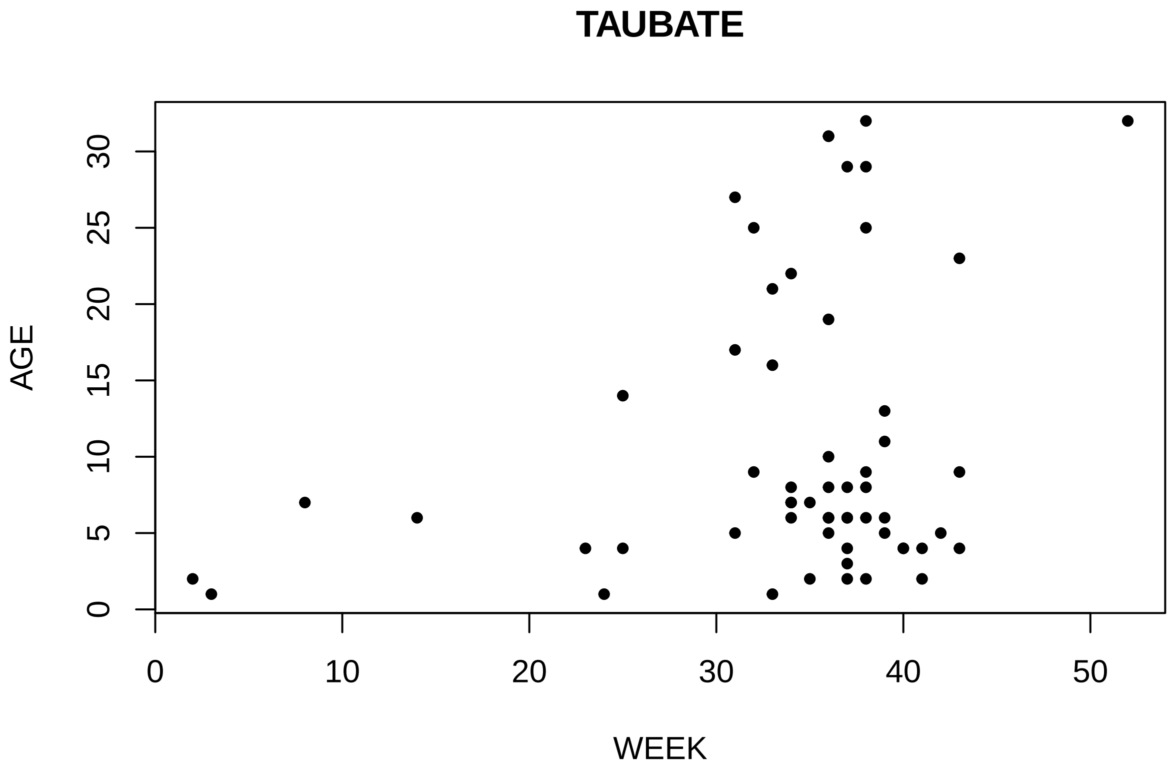

SAO VICENTE
AGE
10
20
30
40
20
25
30
35
40
45
50
WEEK

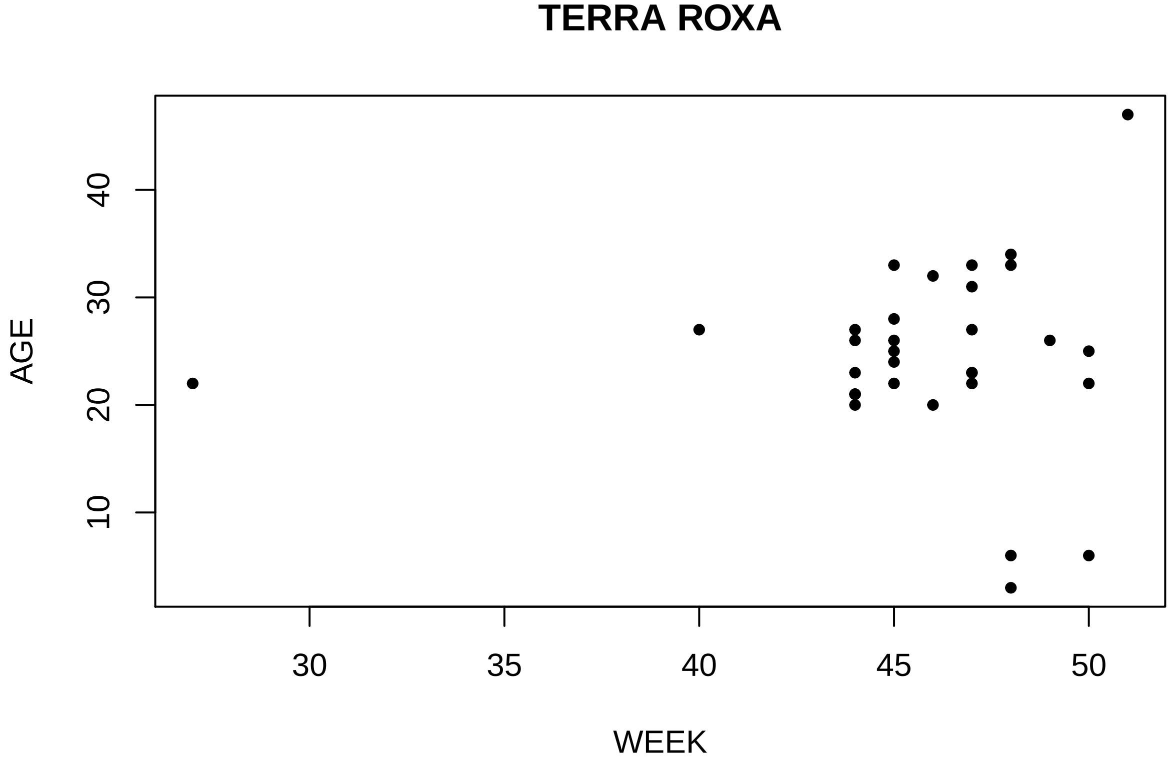

SERTAOZINHO
AGE
0
10
20
30
40
50
30
35
40
45
50
WEEK

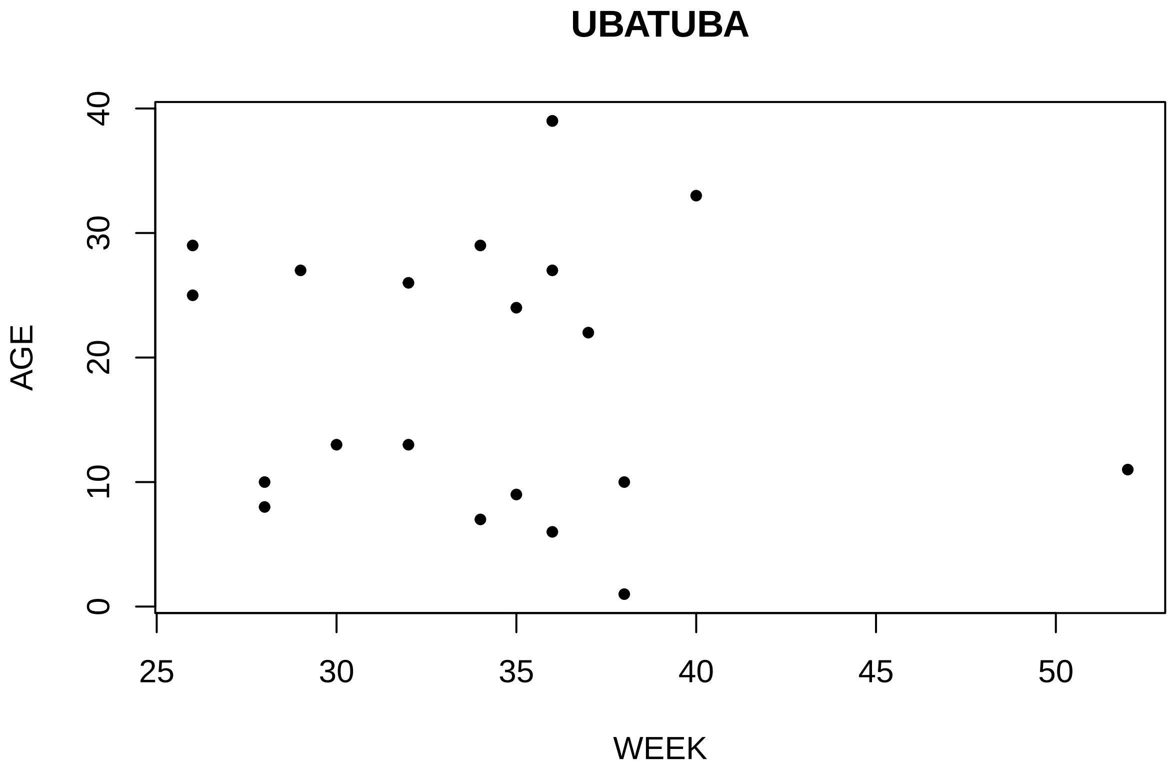

SOROCABA
AGE
0
10
20
30
40
50
10
20
30
40
50
WEEK

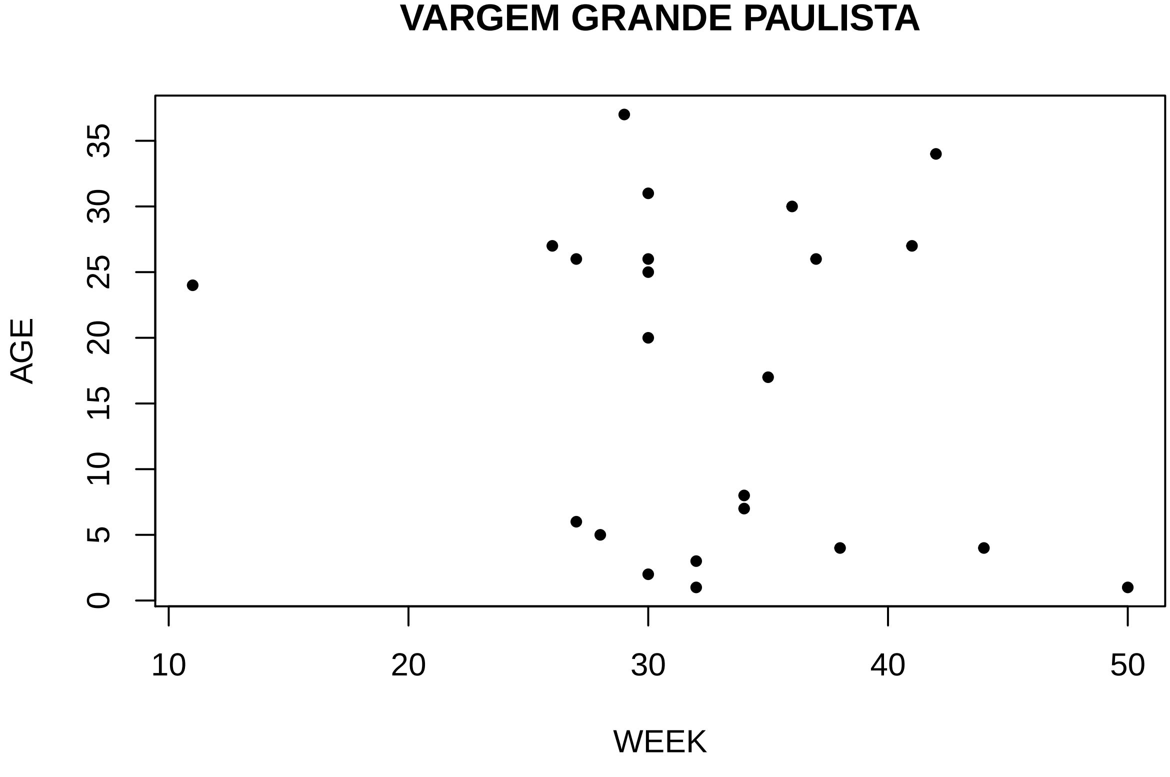

SUMARE
AGE
0
10
20
30
40
WEEK
20
25
30
35
40
45
50

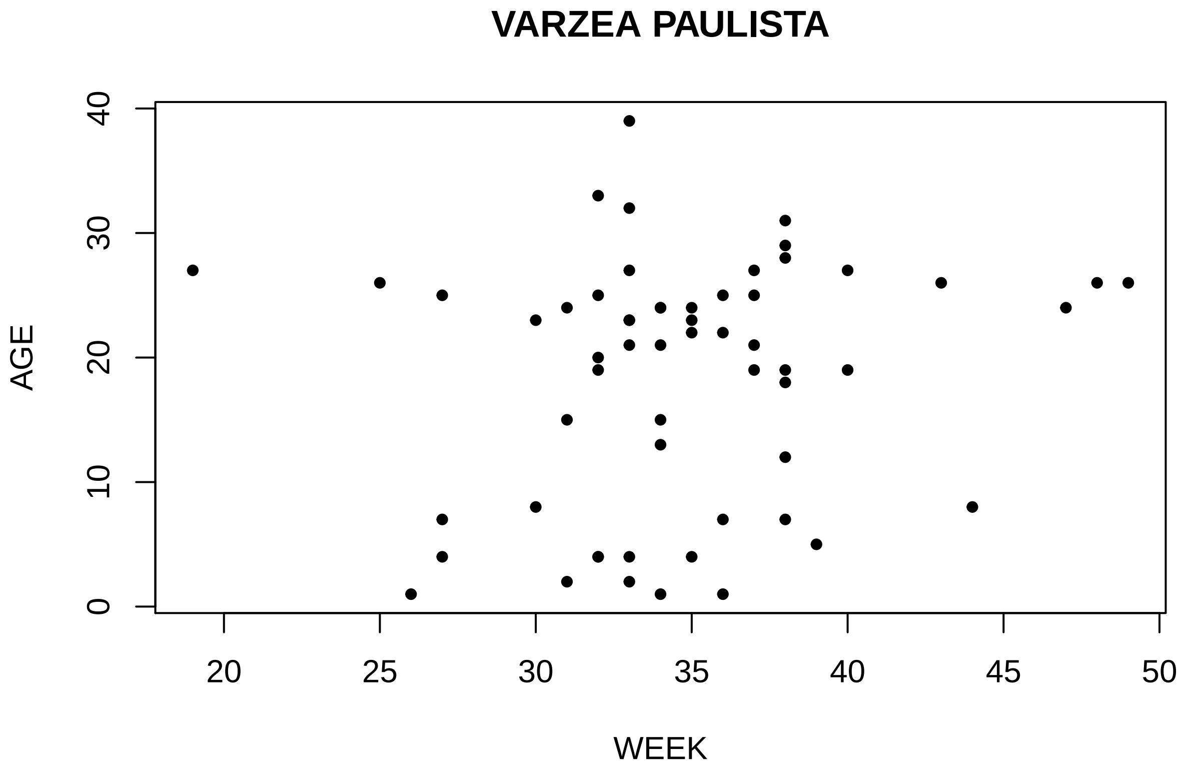

SUZANO
AGE
0
10
20
30
40
50
WEEK
10
20
30
40
50

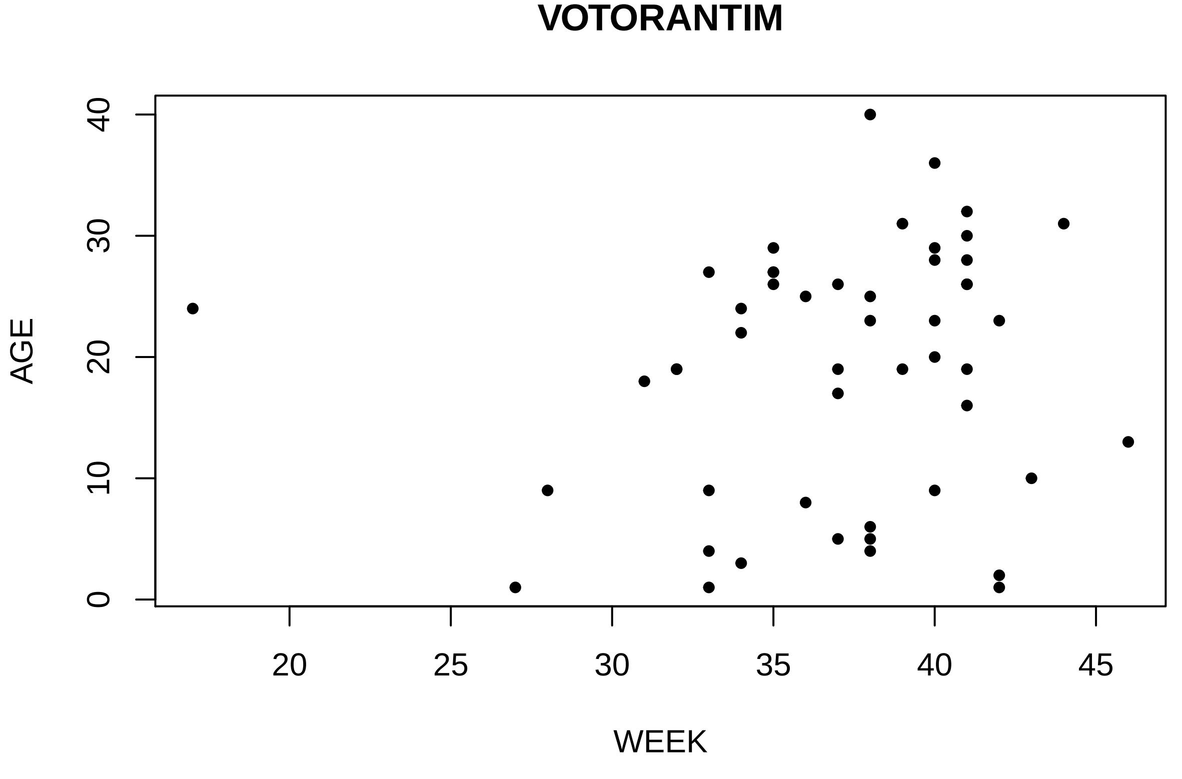

TAUBATE
AGE
0
5
10
15
20
25
30
0
10
20
30
40
50
WEEK


TERRA ROXA
AGE
10
20
30
40
30
35
40
45
50
WEEK


UBATUBA


VARGEM GRANDE PAULISTA


VARZEA PAULISTA
AGE
0
10
20
30
40
WEEK
20
25
30
35
40
45
50


VOTORANTIM
AGE
0
10
20
30
40
WEEK
20
25
30
35
40
45

Preliminary Analysis

Confidence intervals for age group densities

Created two_matrix, three_matrix, four_matrix, and five_matrix which contain the density of cases in age groups of the corresponding span for each county. Density was just calculated as the number of cases in that specific class divided by the number of cases in that county total.

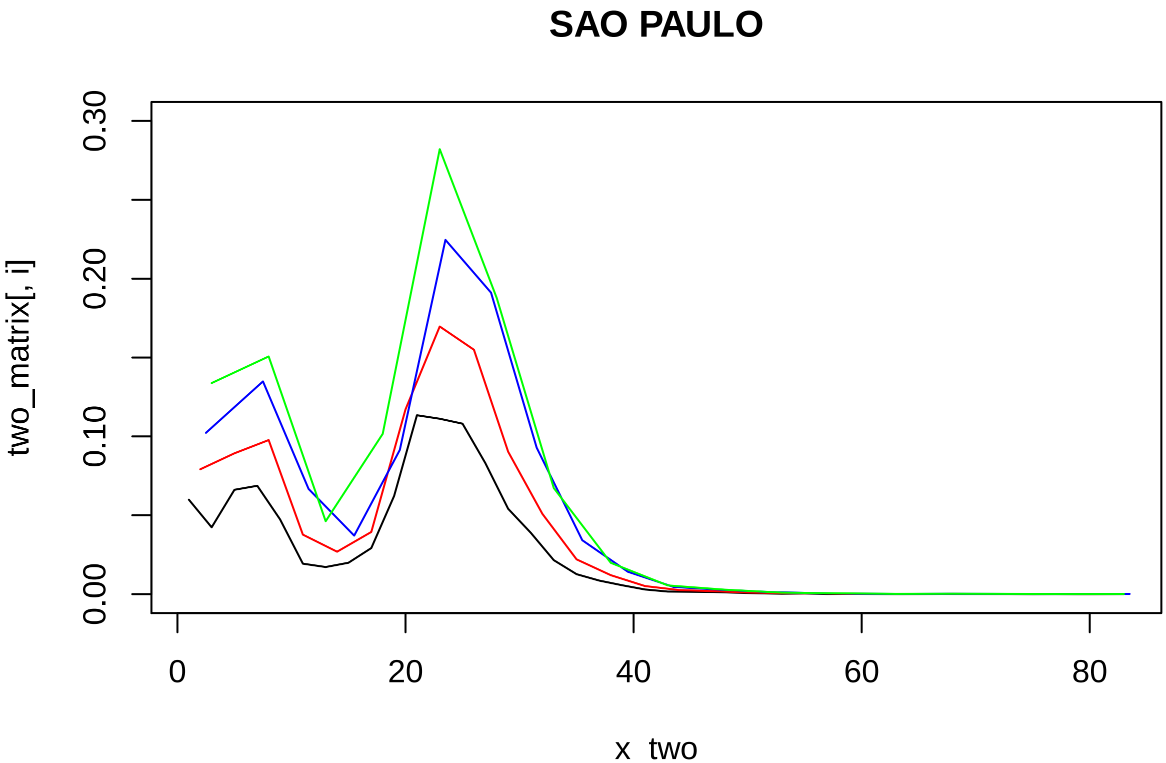
```r
#find appropriate age windows, look at age groups with ranges ranging from 2 to 5 years
density_age <- function(max.age, size){
  amount <- round(max.age / size)
  amount.vec <- rep(NA, amount)
  amount.vec <- as.data.frame(amount.vec)
  amount.vec[1,] <- length(which(subset_county$AGE <= size))
  for(i in 2:amount){
    before <- as.integer(sum(amount.vec[1: i - 1,]))
    amount.vec[i,] <- length(which(subset_county$AGE <= (size * i))) - before
  }
  density.vec <- amount.vec / length(subset_county$AGE)
  return(density.vec)
}

#matrix for age group of 2's
two_matrix <- matrix(rep(NA, round(84/2) * length(county_vec)), ncol = length(county_vec))
two_matrix <- as.data.frame(two_matrix)
colnames(two_matrix) <- county_vec
row.names(two_matrix) <- c("<2", "2-4", "4-6", "6-8", "8-10",
                           "10-12", "12-14", "14-16", "16-18", "18-20",
                           "20-22", "22-24", "24-26", "26-28", "28-30",
                           "30-32", "32-34", "34-36", "36-38", "38-40",
                           "40-42", "42-44", "44-46", "46-48", "48-50",
                           "50-52", "52-54", "54-56", "56-58", "58-60",
                           "60-62", "62-64", "64-66", "66-68", "68-70",
                           "70-72", "72-74", "74-76", "76-78", "78-80",
                           "80-82", "82-84")

for(j in 1:length(county_vec)){
  subset_county <- subset(data, data$COUNTY == county_vec[j])
  max.age <- max(subset_county$AGE)
  density <- unlist(density_age(max.age, 2))
  length_d <- length(density)
  density <-  c(density, rep(NA, (42 - length_d)))
  two_matrix[,j] <- density
}
#matrix for age groups of 3's
three_matrix <- matrix(rep(NA, round(84/3) * length(county_vec)), ncol = length(county_vec))
three_matrix <- as.data.frame(three_matrix)
colnames(three_matrix) <- county_vec
rownames(three_matrix) <- c("<3", "3-6", "6-9",
                            "9-12", "12-15", "15-18",
                            "18-21", "21-24", "24-27",
                            "27-30", "30-33", "33-36",
                            "36-39", "39-42", "42-45",
                            "45-48", "48-51", "51-54",
                             "54-57", "57-60", "60-63",
                             "63-66", "66-69", "69-72",
                             "72-75", "75-78", "78-81",
                             "81-84")

for(j in 1:length(county_vec)){
  subset_county <- subset(data, data$COUNTY == county_vec[j])
  max.age <- max(subset_county$AGE)
  density <- unlist(density_age(max.age, 3))
  length_d <- length(density)
  density <-  c(density, rep(NA, (28 - length_d)))
  three_matrix[,j] <- density
}
#matrix for age groups of 4's
four_matrix <- matrix(rep(NA, 21 * length(county_vec)), ncol = length(county_vec))
four_matrix <- as.data.frame(four_matrix)
colnames(four_matrix) <- county_vec
row.names(four_matrix) <- c("<4", "4-8", "8-12", "12-16",
                            "16-20", "20-24", "24-28", "28-32",
                            "32-36", "36-40", "40-44", "44-48",
                            "48-52", "52-56", "56-60", "60-64",
                            "64-68", "68-72", "72-76", "76-80",
                            "80-84")

for(j in 1:length(county_vec)){
  subset_county <- subset(data, data$COUNTY == county_vec[j])
  max.age <- max(subset_county$AGE)
  density <- unlist(density_age(max.age, 4))
  length_d <- length(density)
  density <-  c(density, rep(NA, (21 - length_d)))
  four_matrix[,j] <- density
}
#matrix for age groups of 5's
five_matrix <- matrix(rep(NA, round(84/5) * length(county_vec)), ncol = length(county_vec))
five_matrix <- as.data.frame(five_matrix)
colnames(five_matrix) <- county_vec
row.names(five_matrix) <- c("<5", "5-10", "10-15", "15-20",
                           "20-25", "25-30", "30-35", "35-40",
                           "40-45", "45-50", "50-55", "55-60",
                           "60-65", "65-70", "70-75", "75-80",
                           "80-85")

for(j in 1:length(county_vec)){
  subset_county <- subset(data, data$COUNTY == county_vec[j])
  max.age <- max(subset_county$AGE)
  density <- unlist(density_age(max.age, 5))
  length_d <- length(density)
  density <-  c(density, rep(NA, (17 - length_d)))
  five_matrix[,j] <- density
}
```

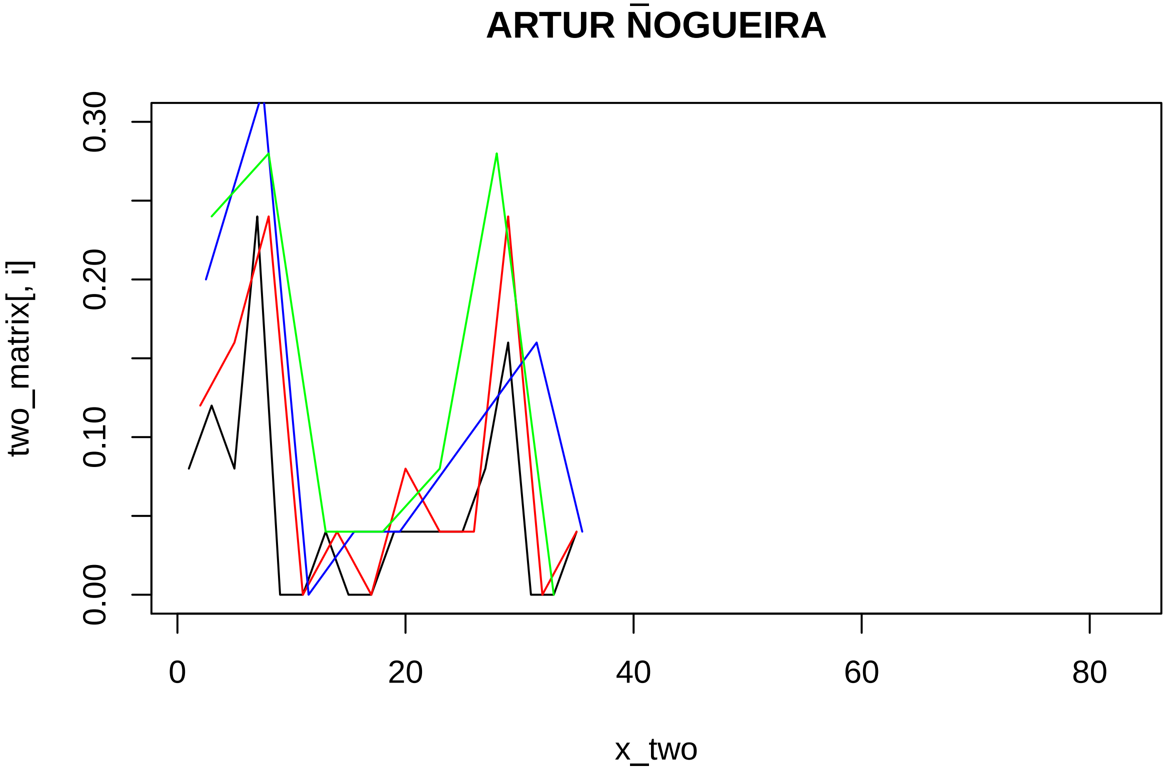
Preliminary Analysis

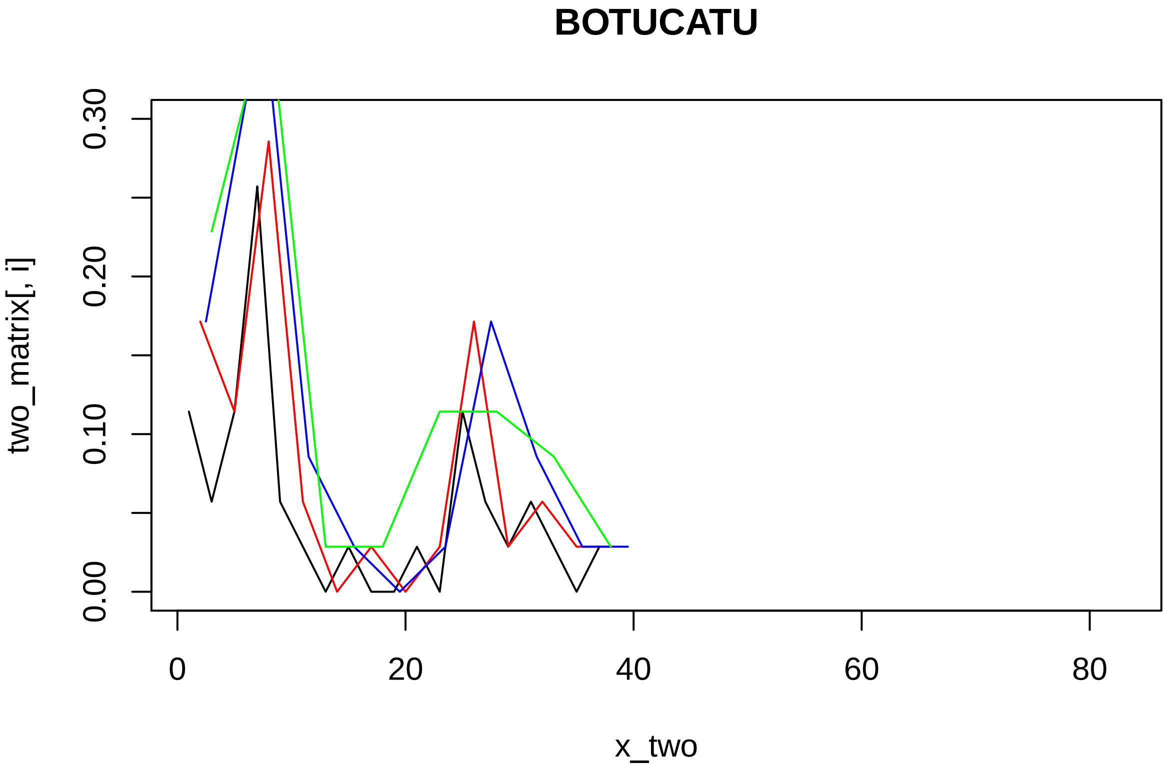
It appears that the size of the window does have a siginficant (colloquially) effect on the density in that close age classes have pretty different densities. Going to compare densities on a line graph.

Comparison of windows

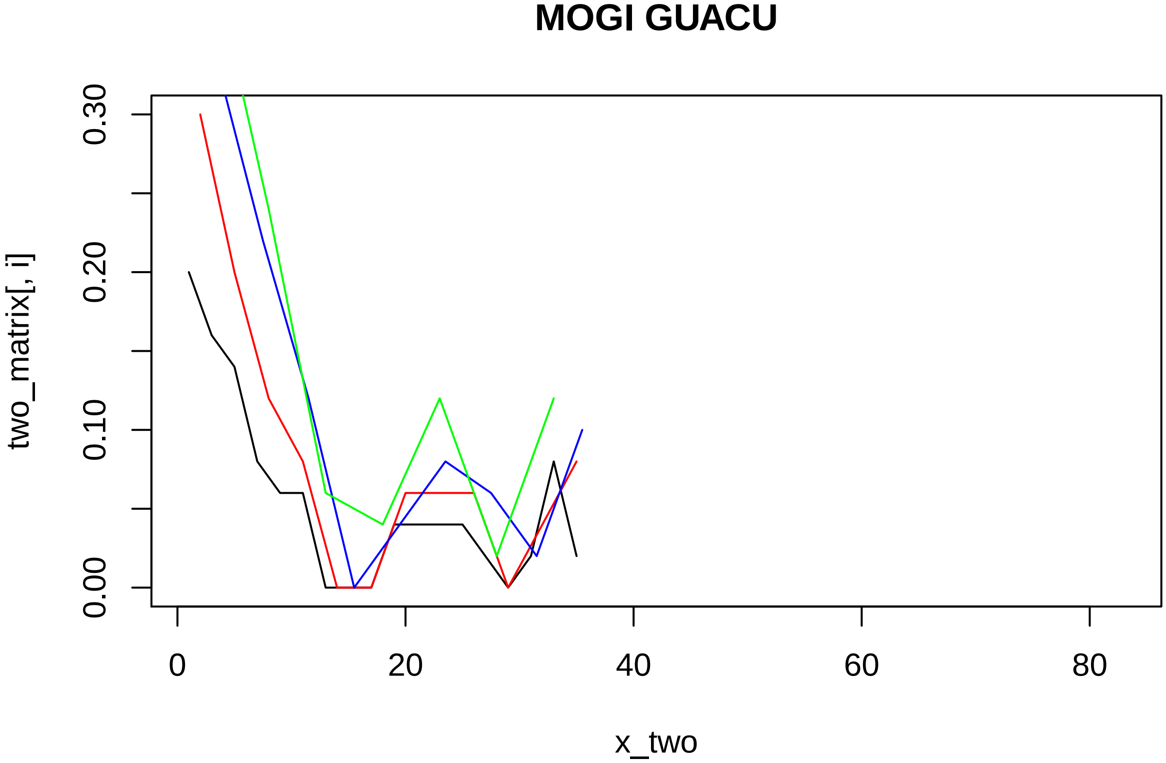
```r
#x axis for twos
x_two <- rep(NA, 42)
x_two[1] <- 1
for(i in 2:42){
  x_add <- seq(1:41)
  x_two[i] <- i + x_add[i - 1]
}
#x axis for threes
x_three <- rep(NA, 28)
x_three[1] <- 2
for(i in 2:28){
  x_add <- seq(1:28)
  x_three[i] <- 2 * i + x_add[i] -1
}
#x axis for fours
x_four <- rep(NA, 21)
x_four[1] <- 2.5
for(i in 2:21){
  x_add <- seq(1:21)
  x_four[i] <- 3 * i + x_add[i] + 0.5 -1
}
#x axis for fives
x_five <- rep(NA, 17)
x_five[1] <- 3
for(i in 2:17){
  x_add <- seq(1:17)
  x_five[i] <- 4 * i + x_add[i] -2
}

for(i in 1:length(county_vec)){
  subset_county <- subset(data, data$COUNTY == county_vec[i])
  if(length(subset_county$AGE) > 20){
    {plot(x_two, two_matrix[,i], type = "l", ylim = c(0, 0.3), main = county_vec[i])
     lines(x_three, three_matrix[,i], col = "red")
     lines(x_four, four_matrix[,i], col = "blue")
     lines(x_five, five_matrix[,i], col = "green")}
  }
}
```


SAO PAULO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

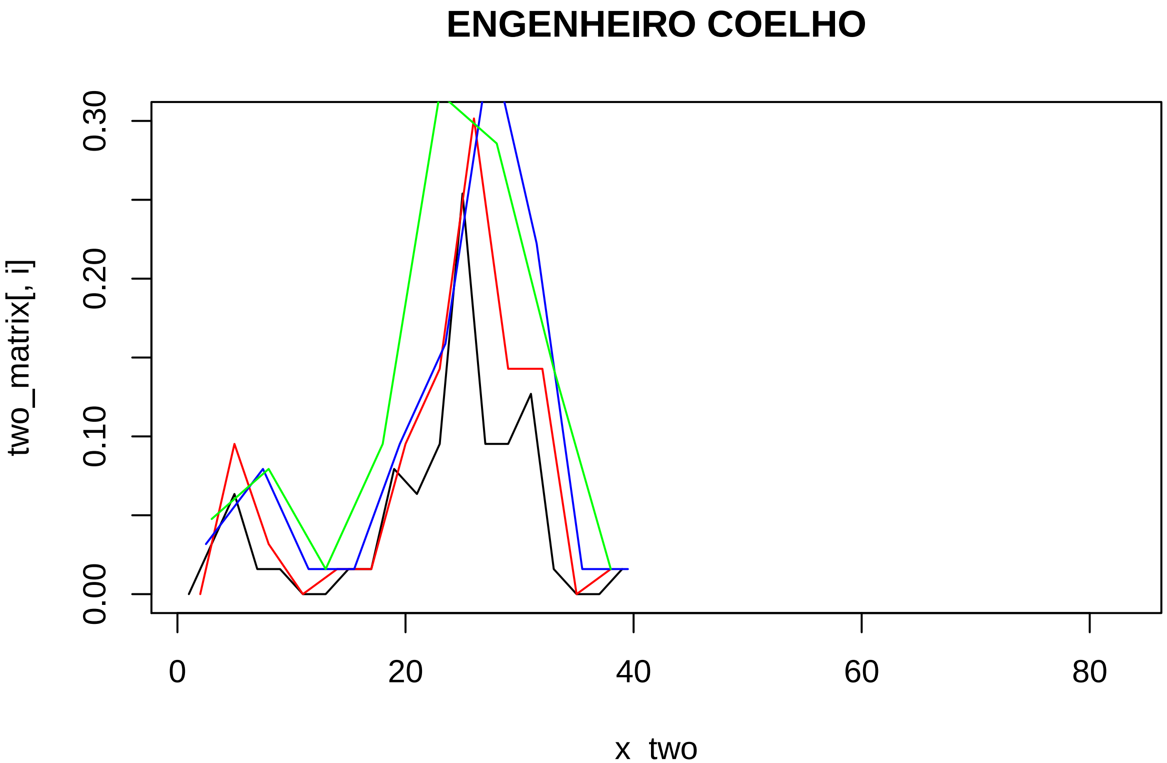

ARTUR NOGUEIRA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

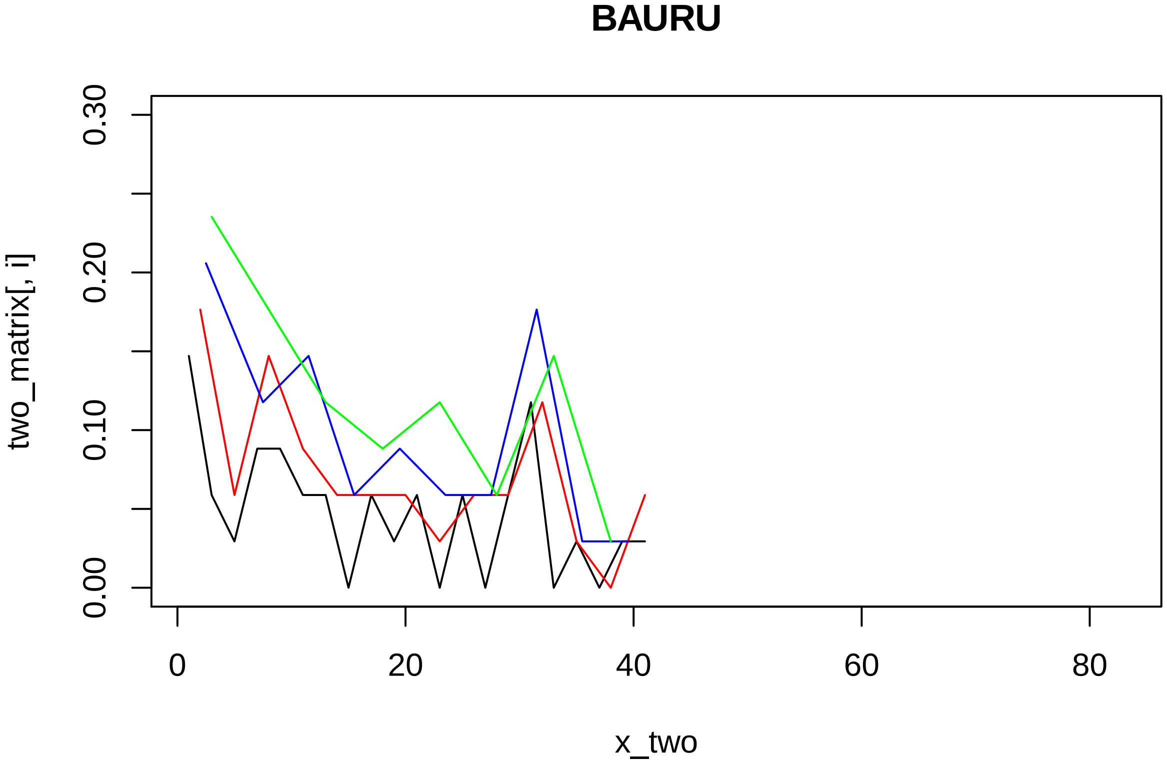


BOTUCATU

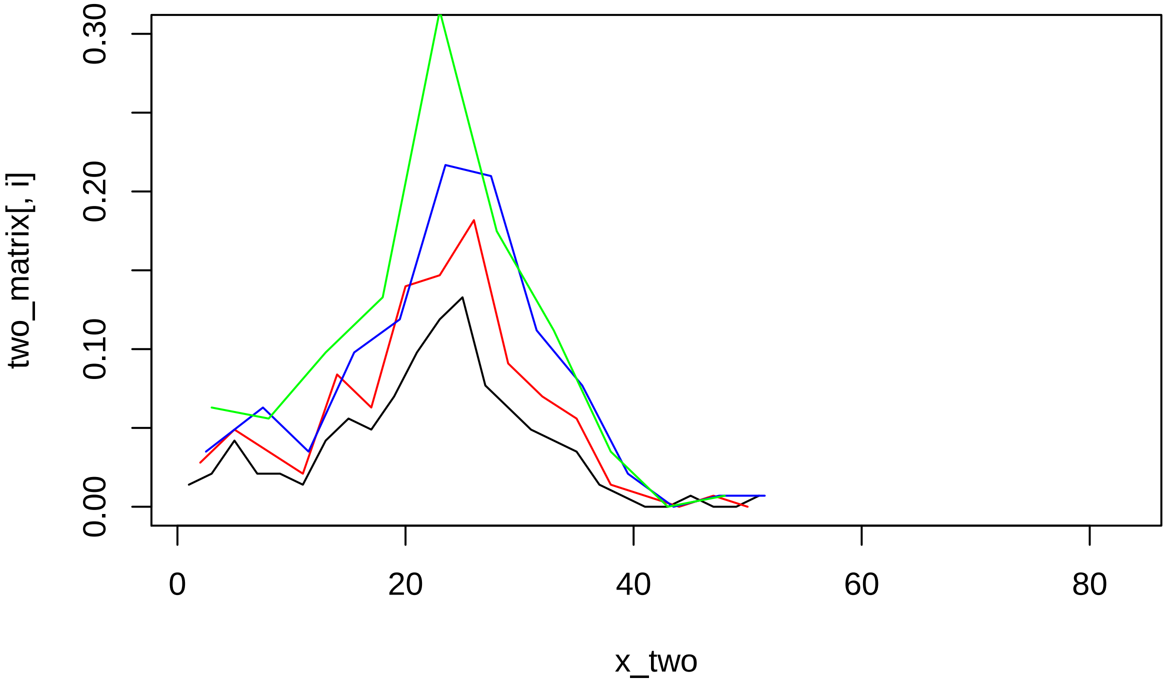


MOGI GUACU

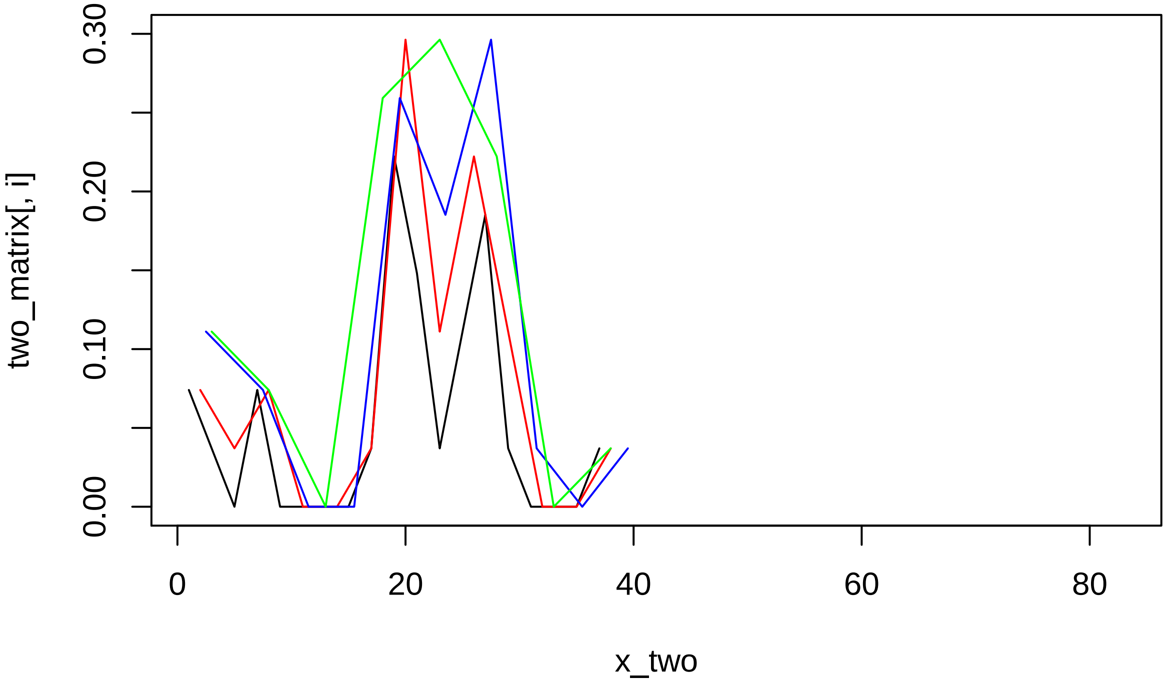

ENGENHEIRO COELHO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

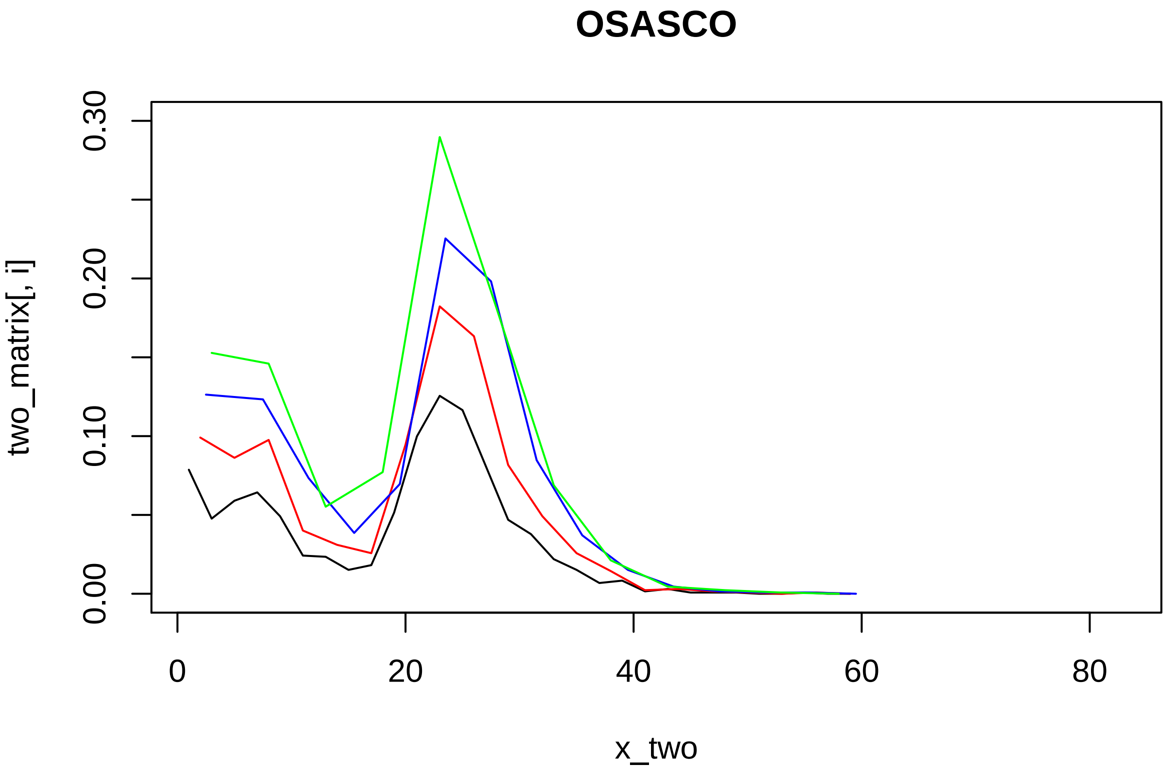

BAURU
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

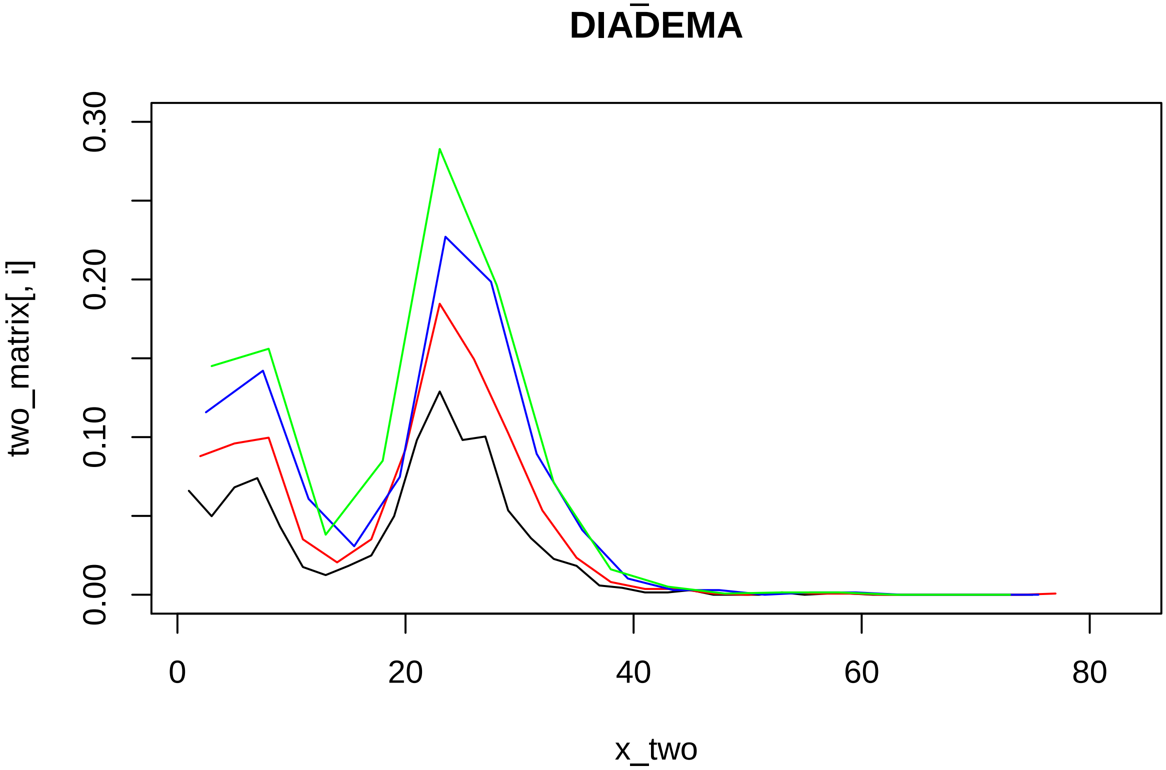
SAO CAETANO DO SUL

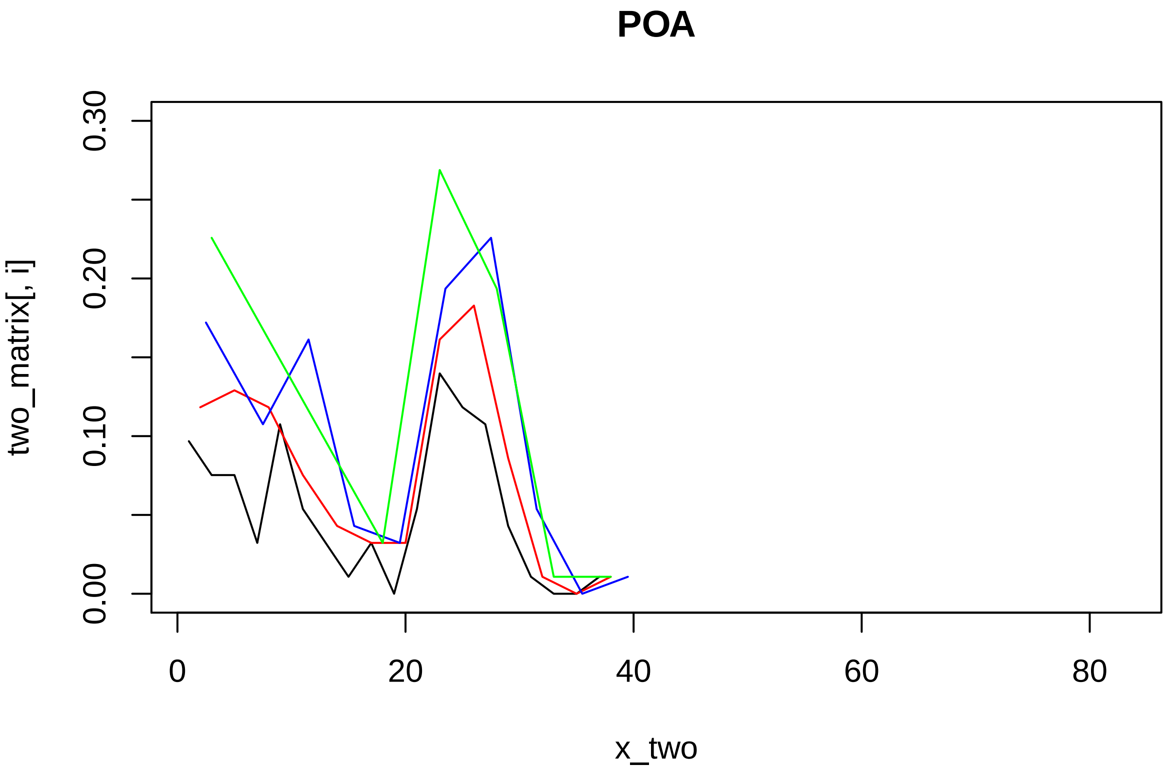


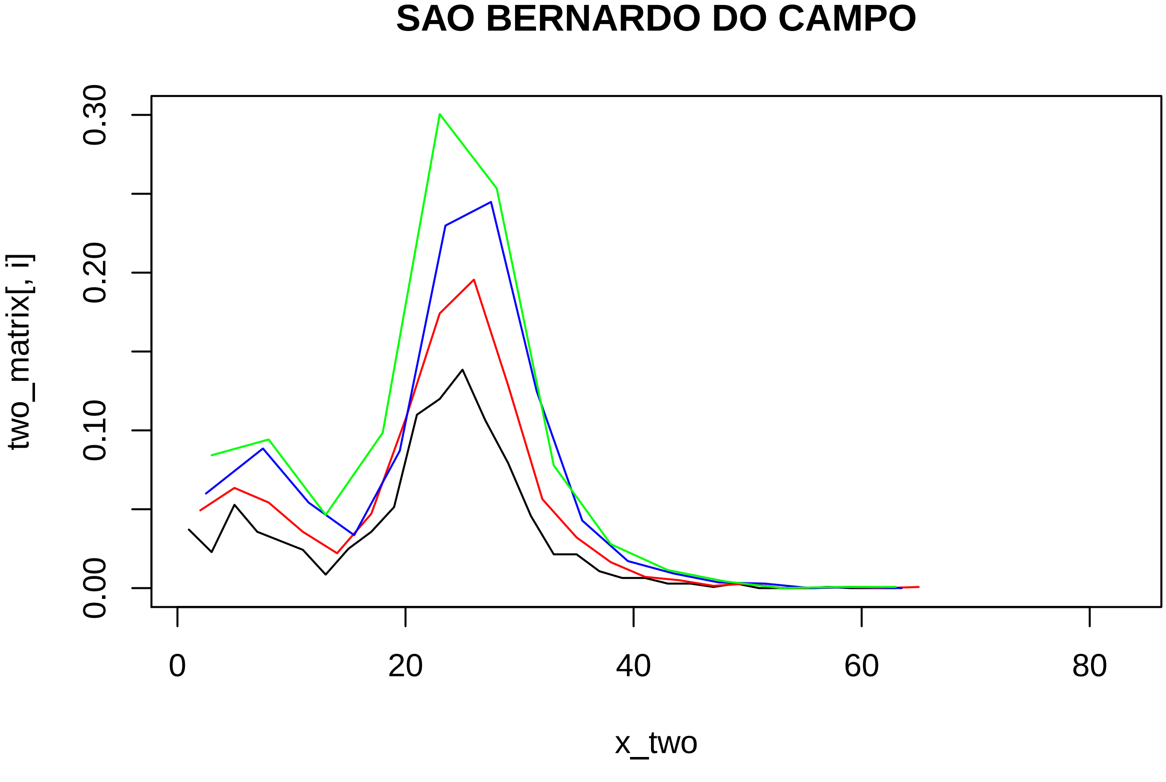
MORRO AGUDO

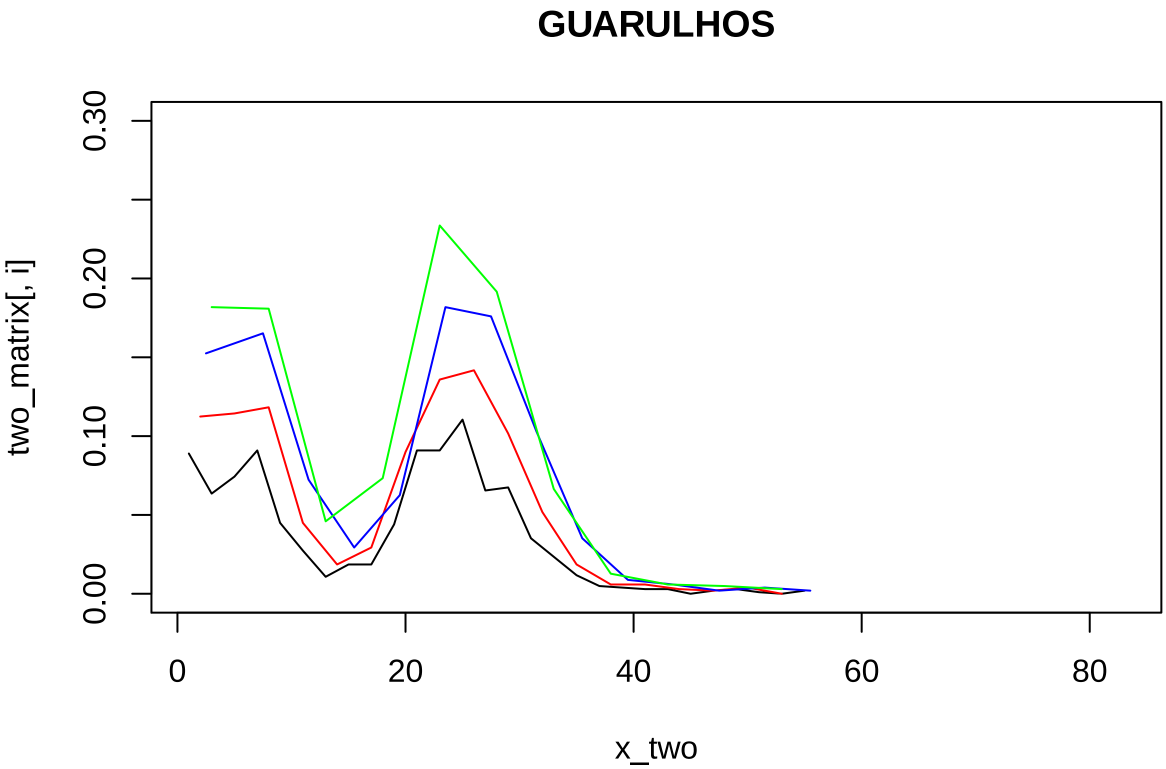


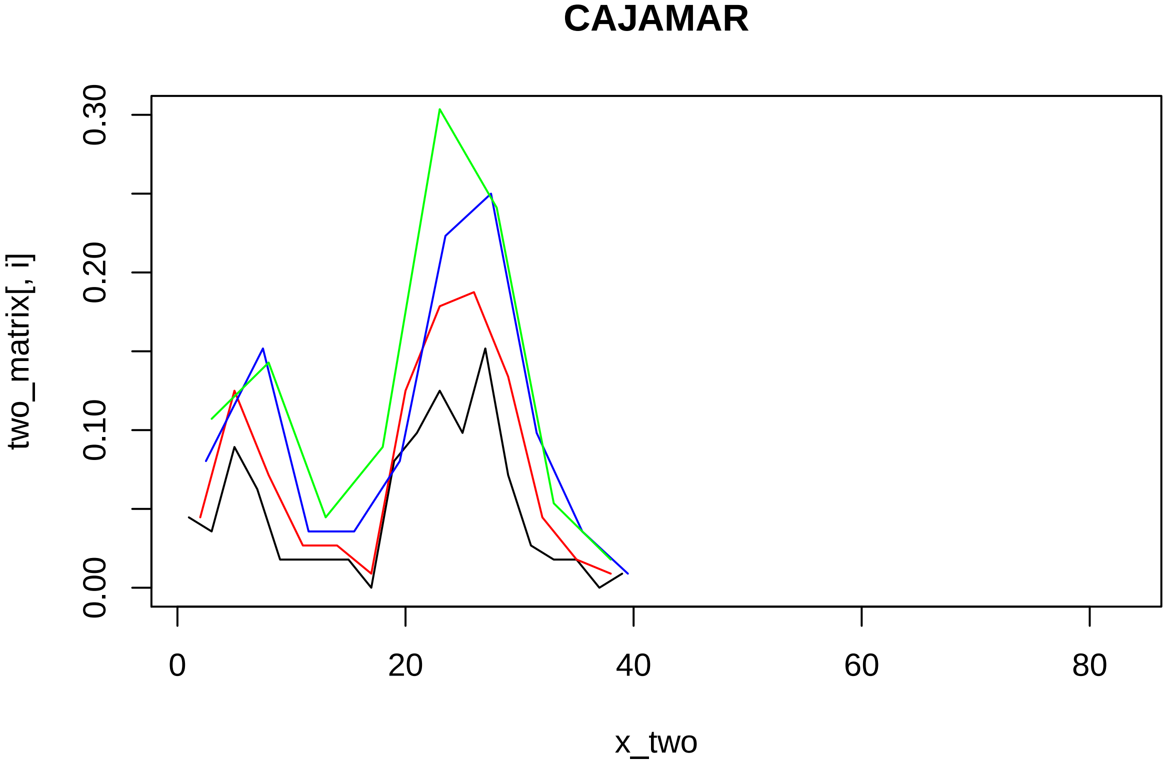

OSASCO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

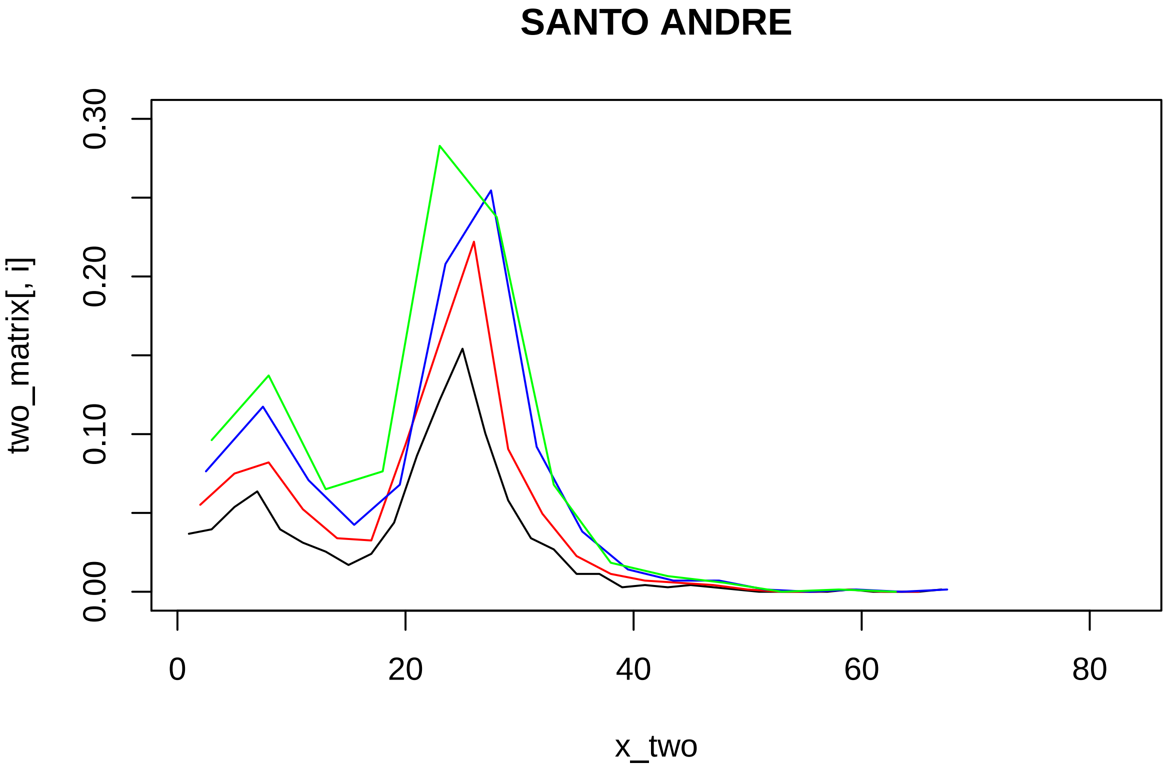

DIADEMA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

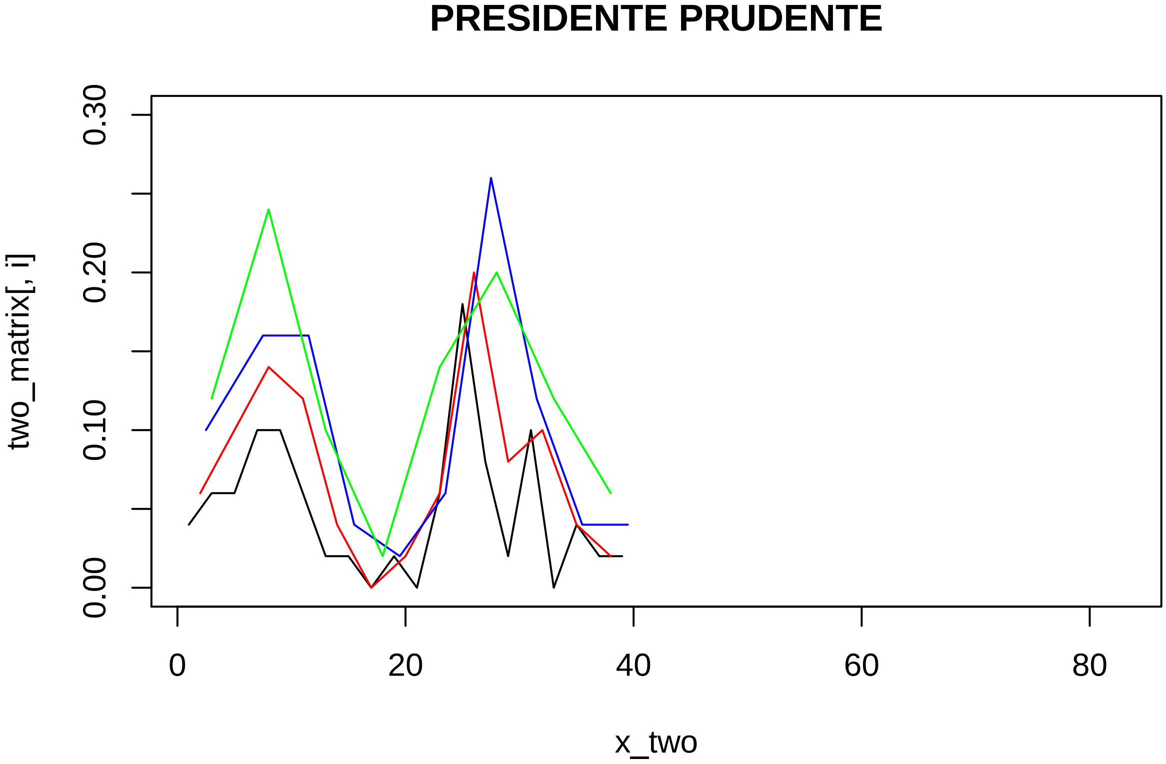

POA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

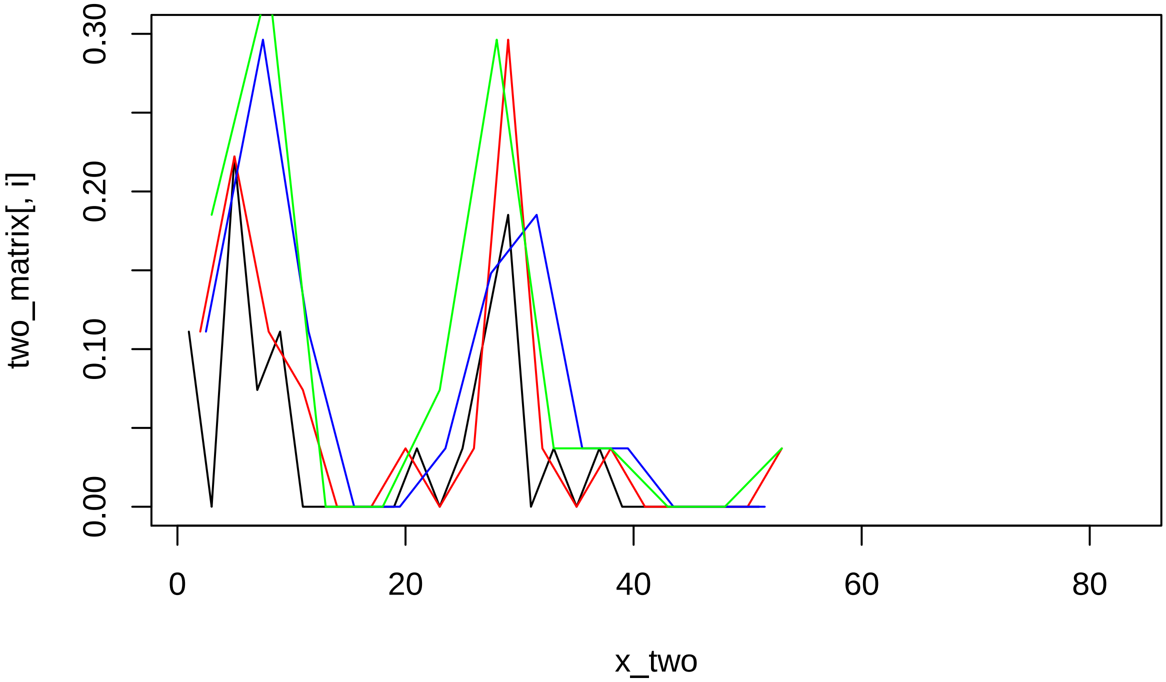

SAO BERNARDO DO CAMPO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


GUARULHOS
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

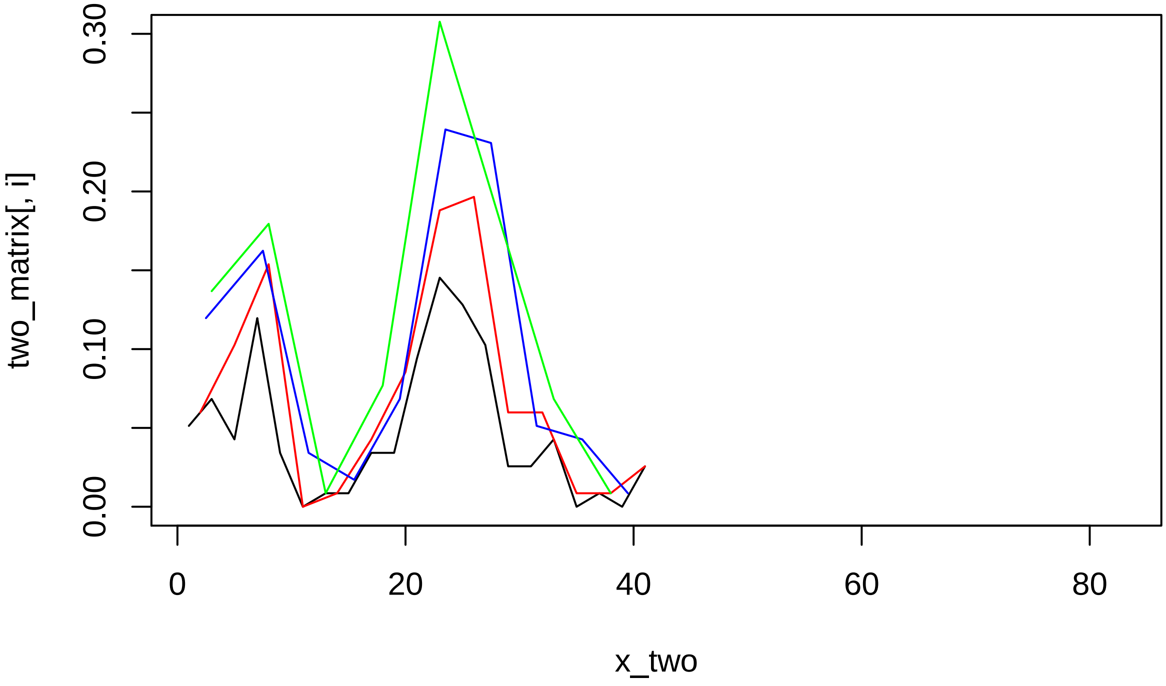

CAJAMAR
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


SANTO ANDRE
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

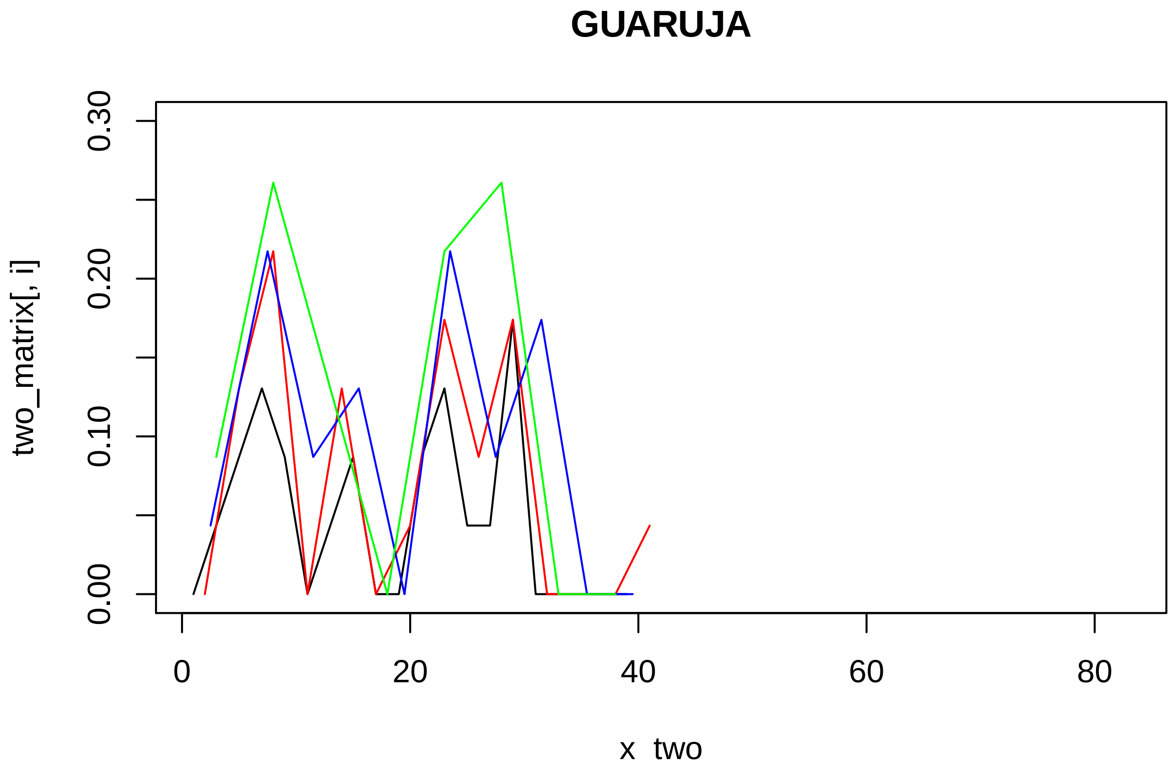

PRESIDENTE PRUDENTE
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

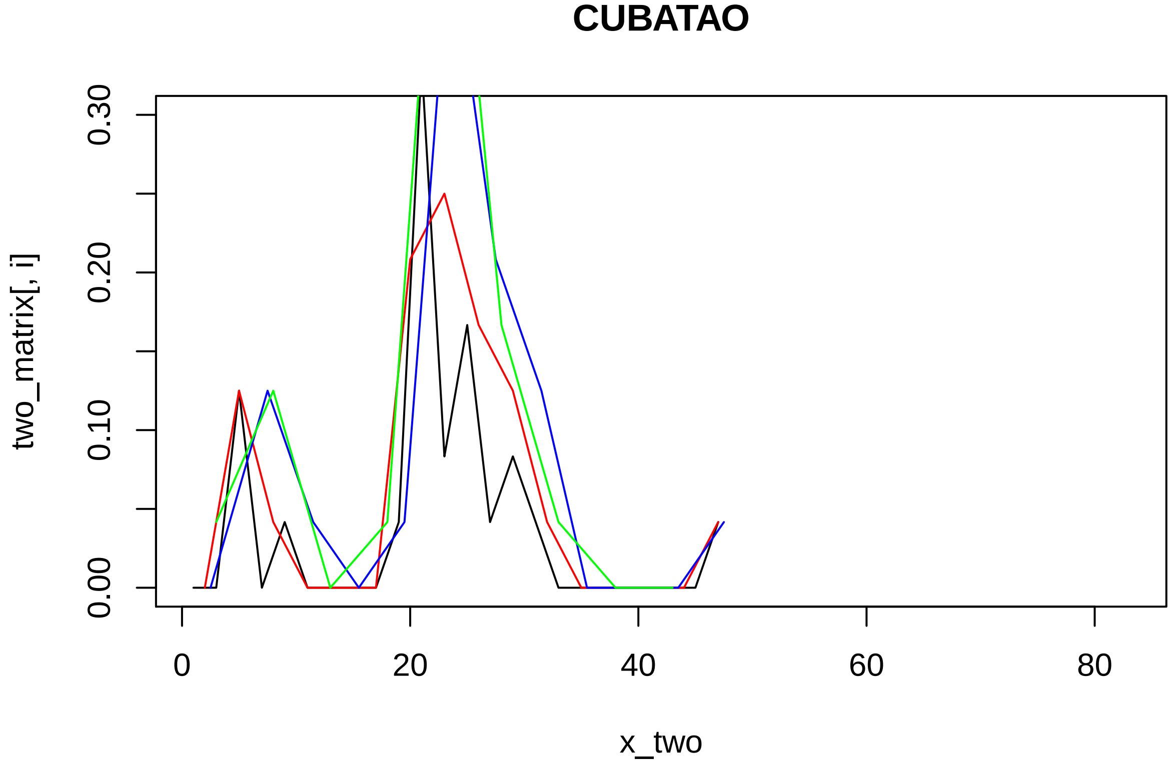


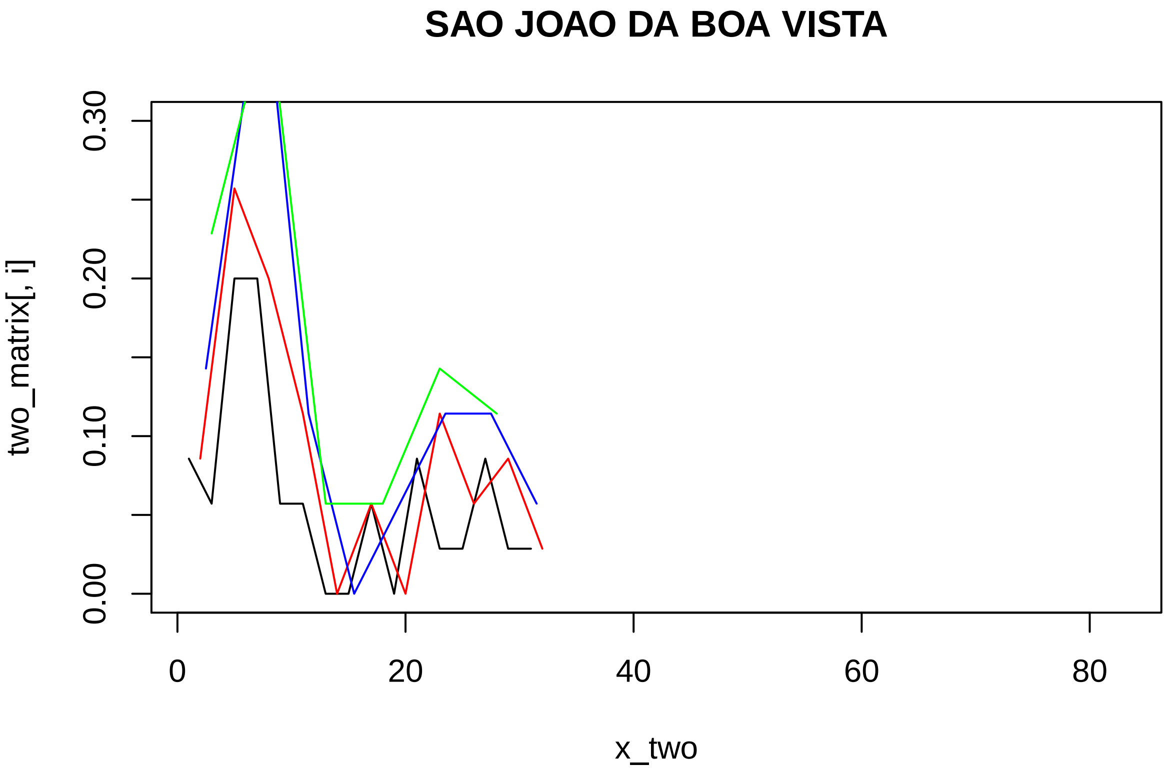
MARILIA

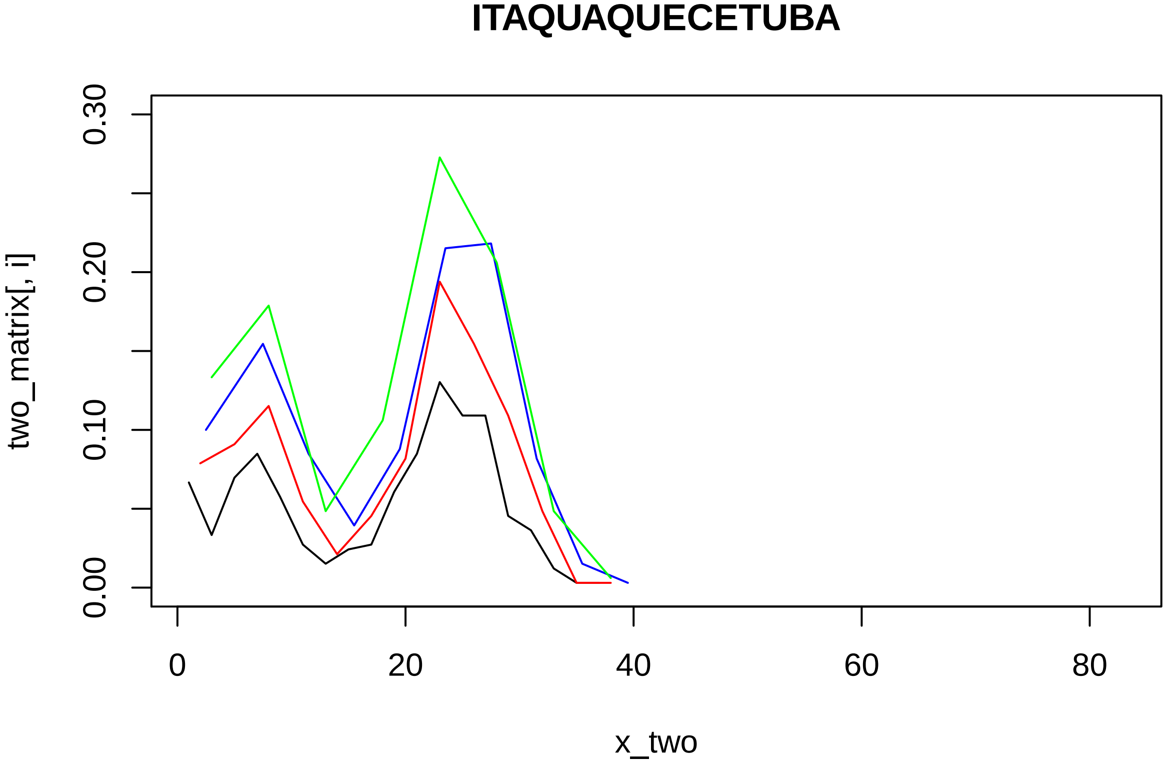


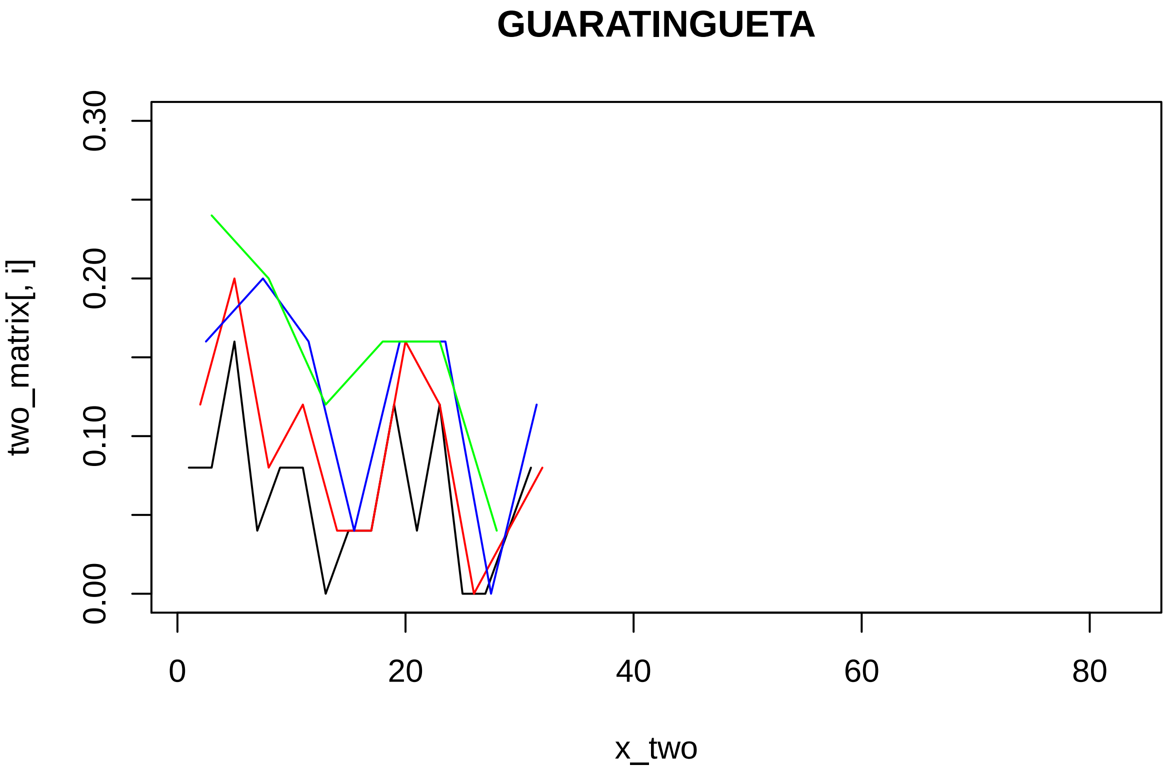
FERRAZ DE VASCONCELOS

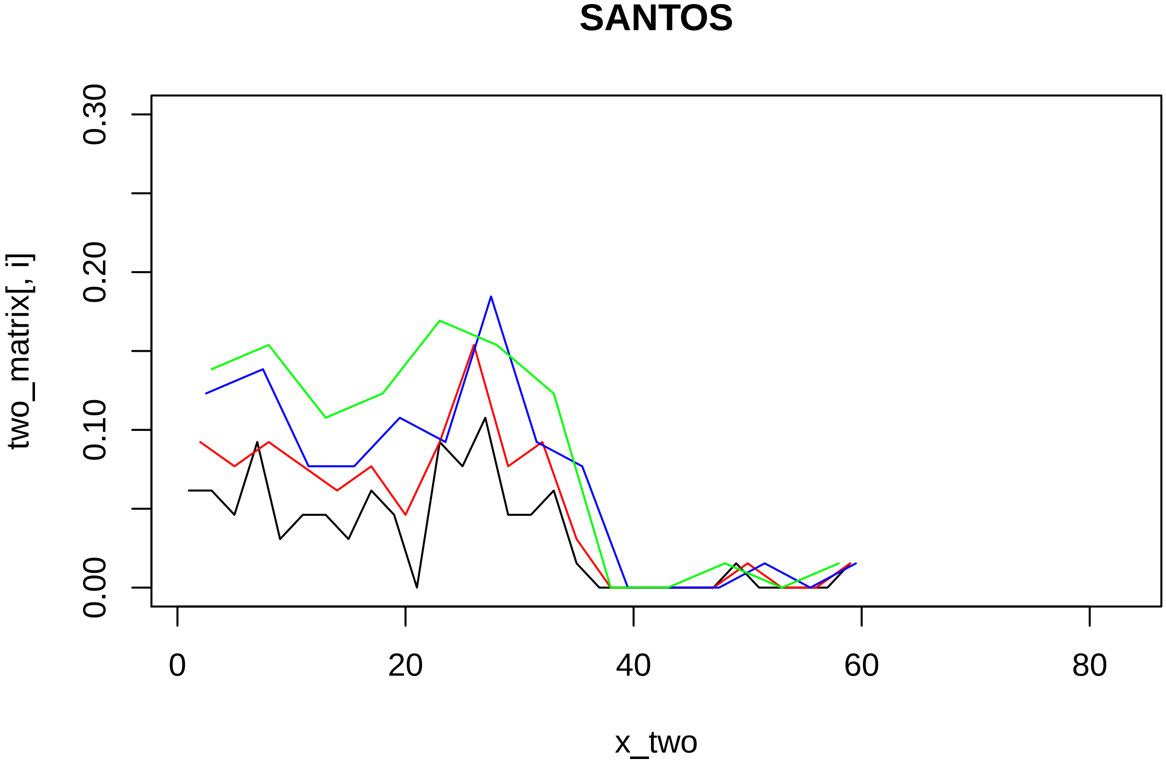

GUARUJA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


CUBATAO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

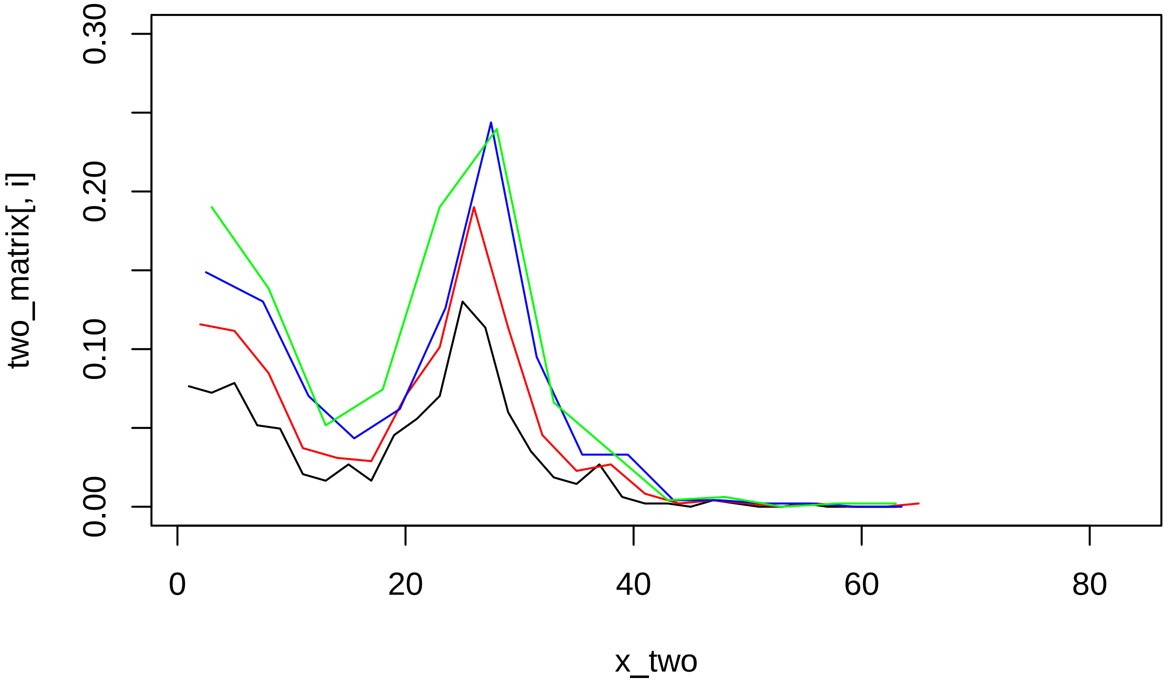

SAO JOAO DA BOA VISTA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


ITAQUAQUECETUBA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

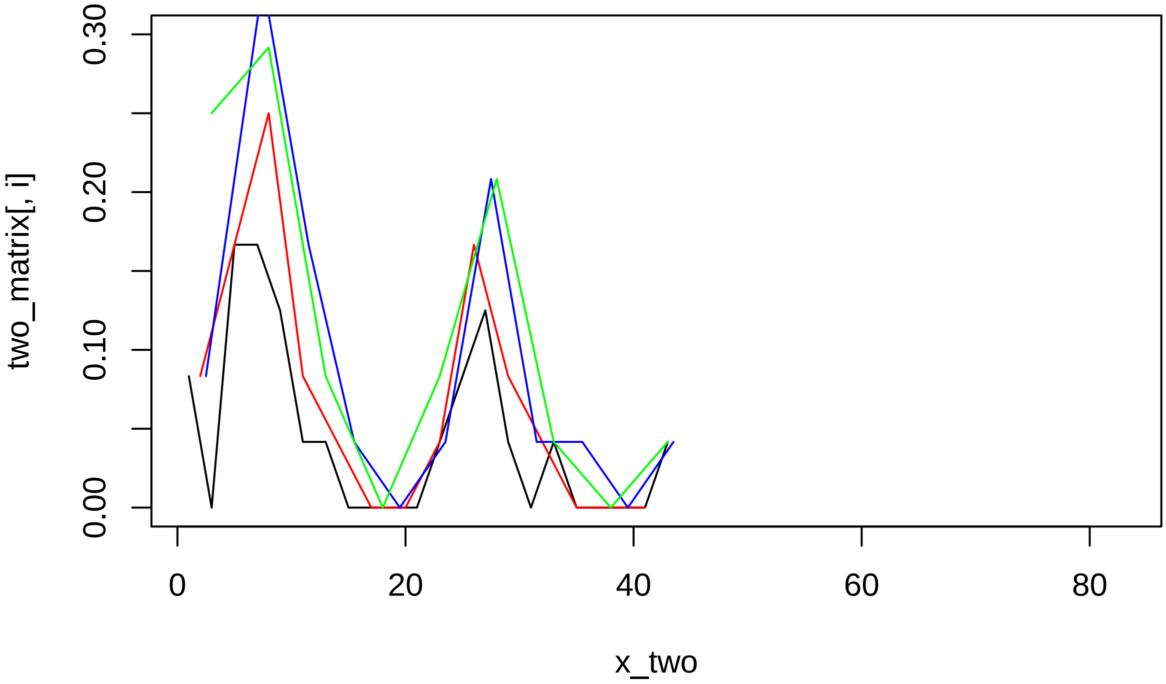

GUARATINGUETA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

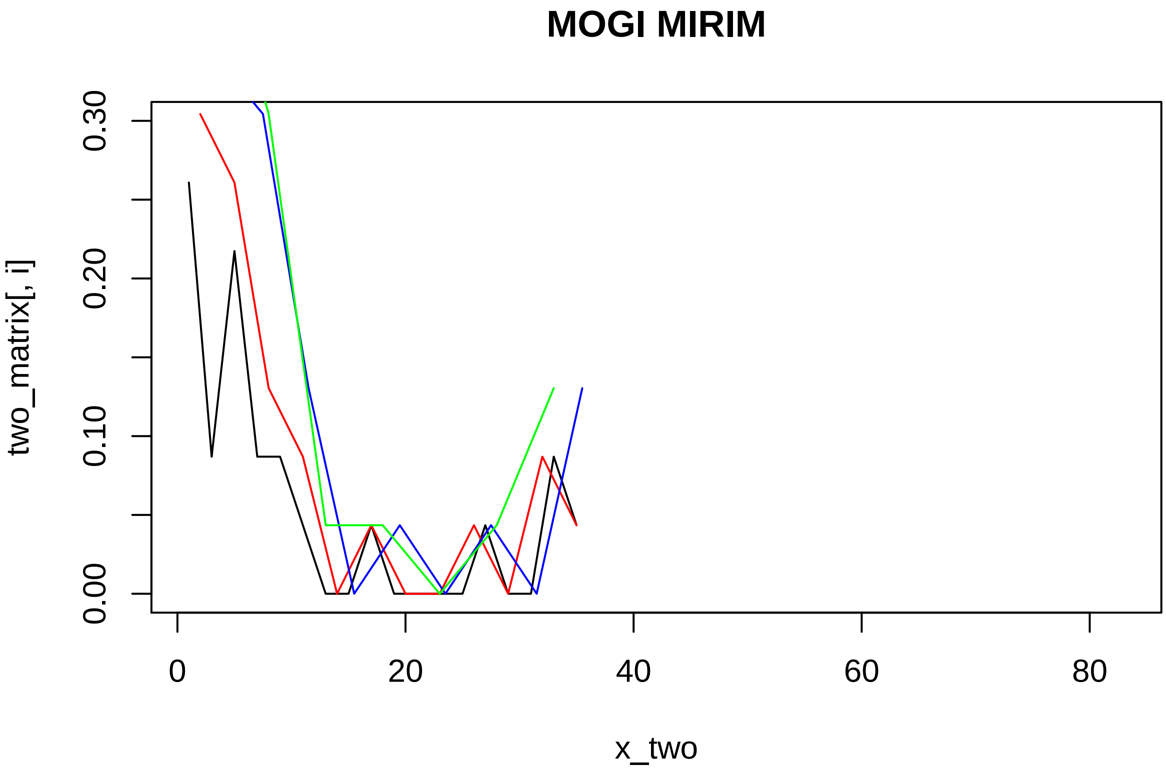

SANTOS
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

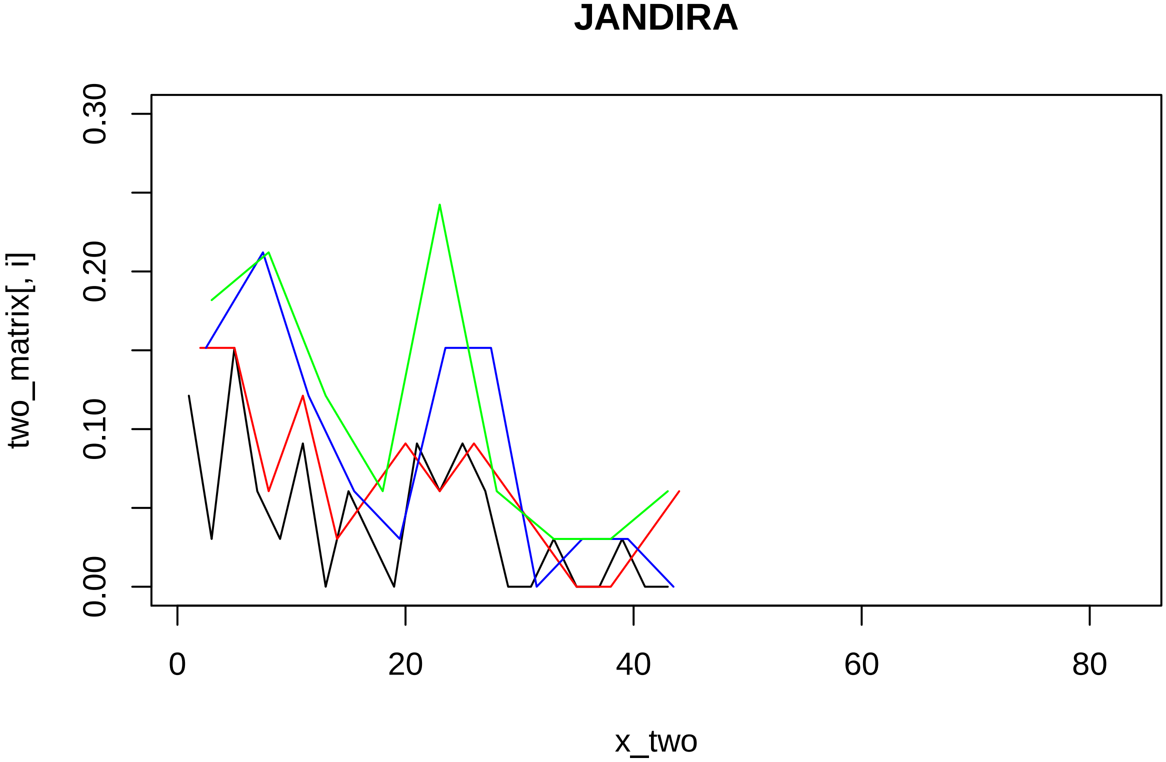
CAMPINAS

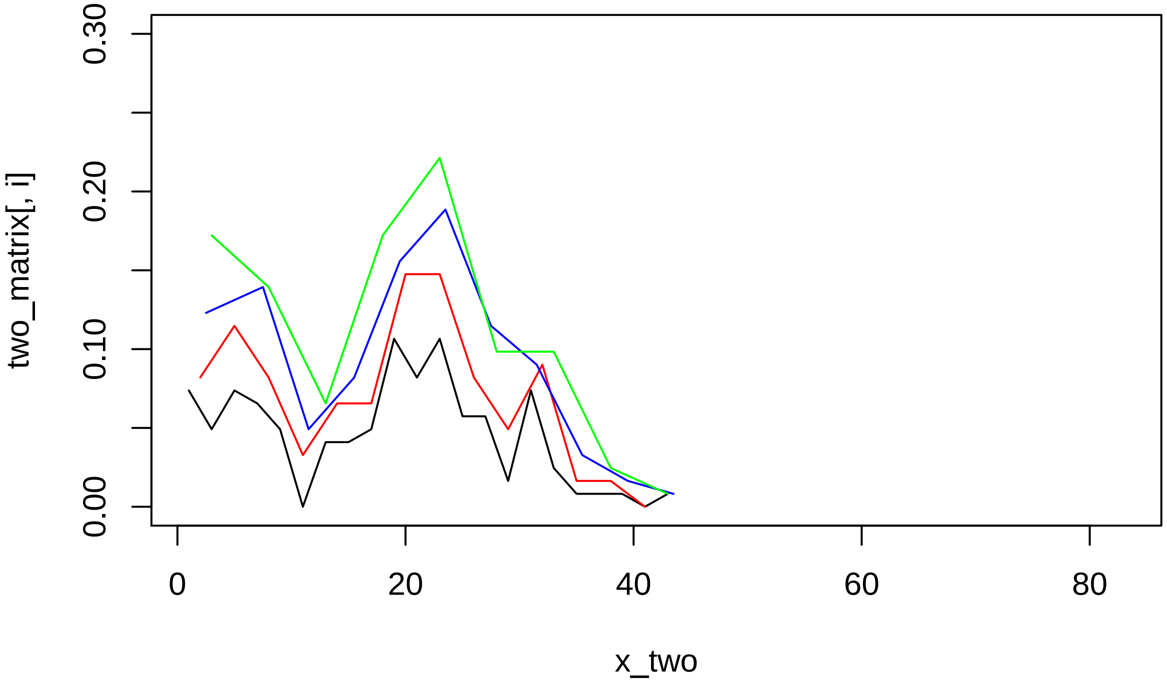


ARARAQUARA

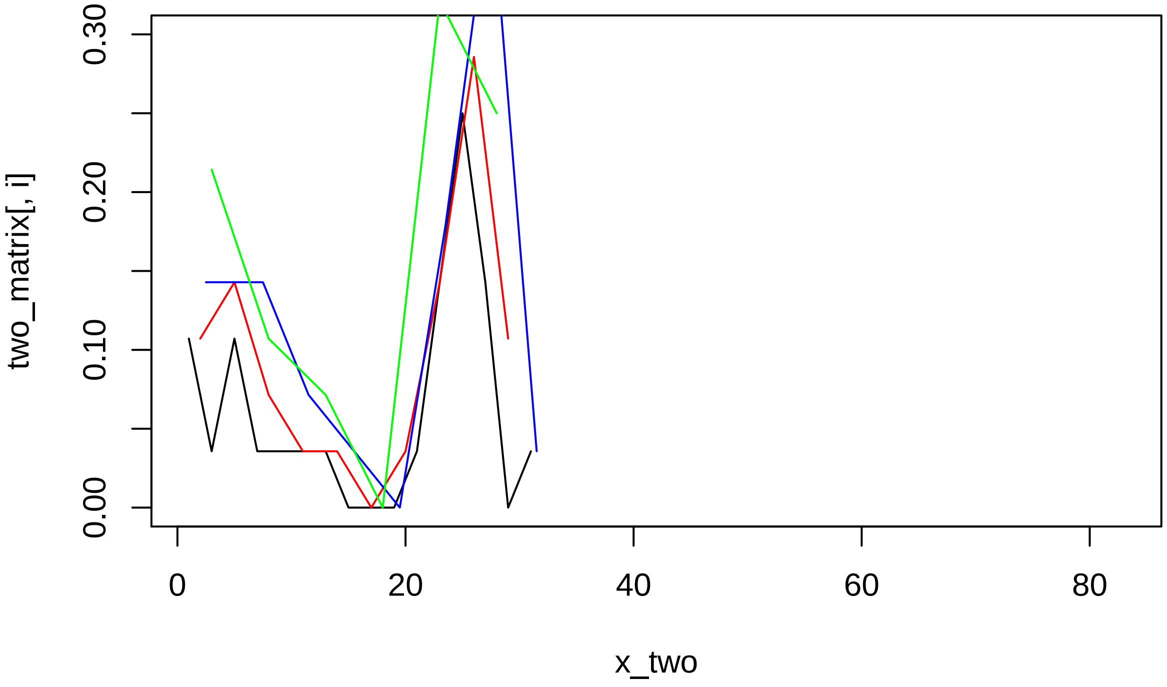



MOGI MIRIM
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

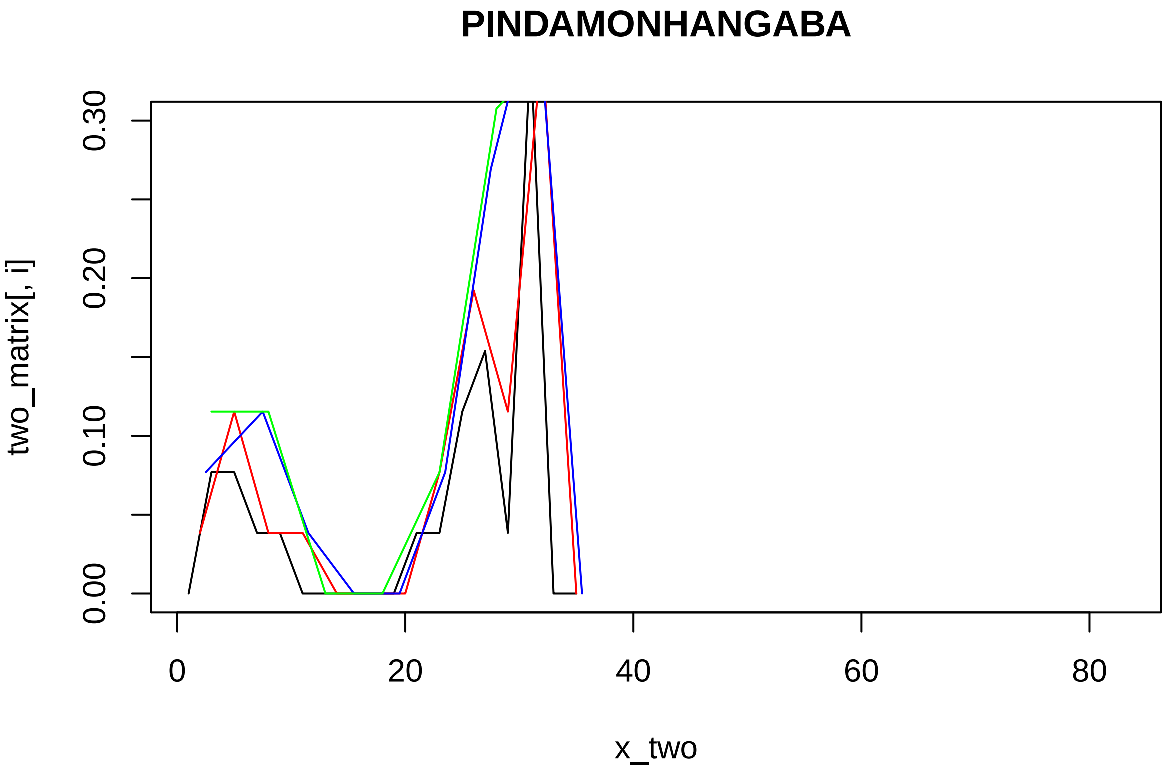

JANDIRA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

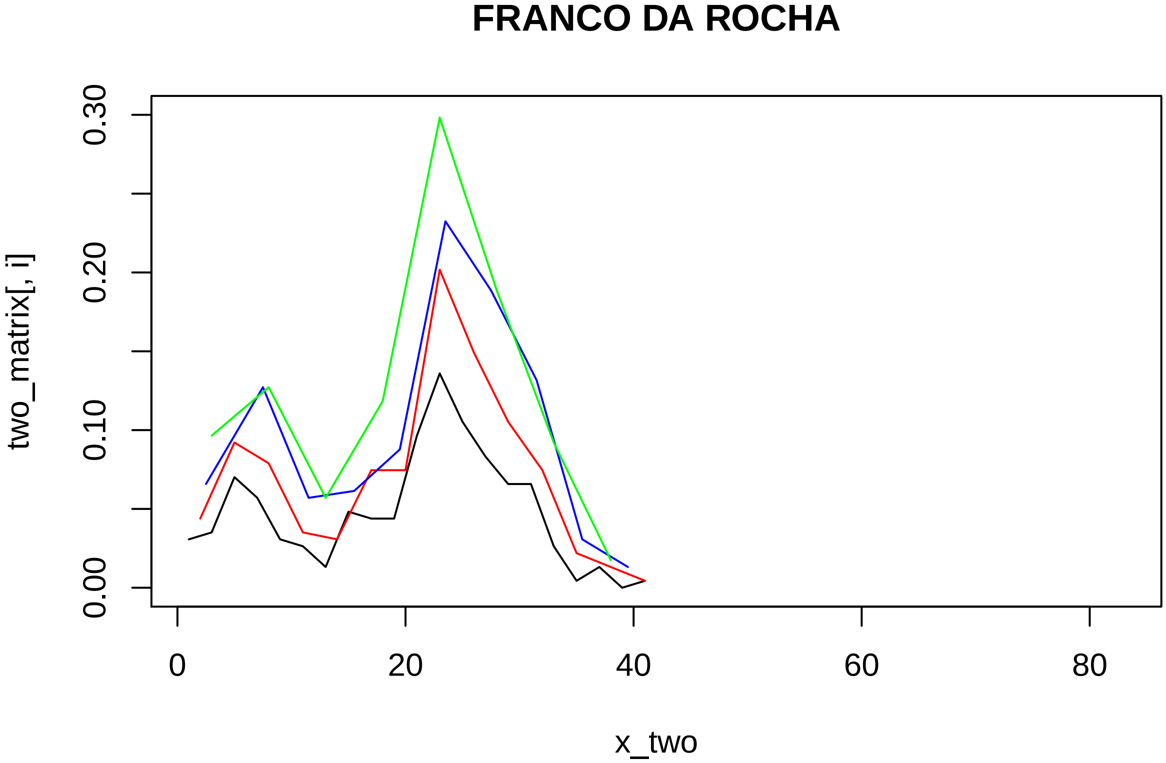


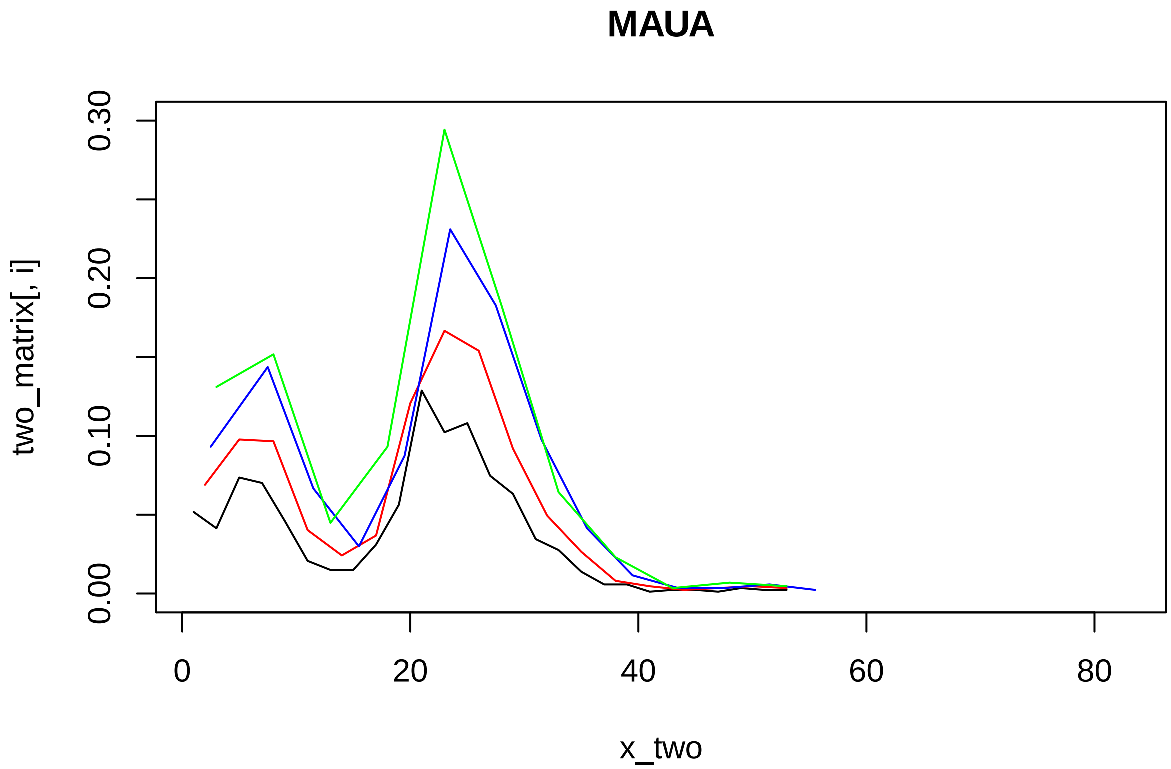
COTIA

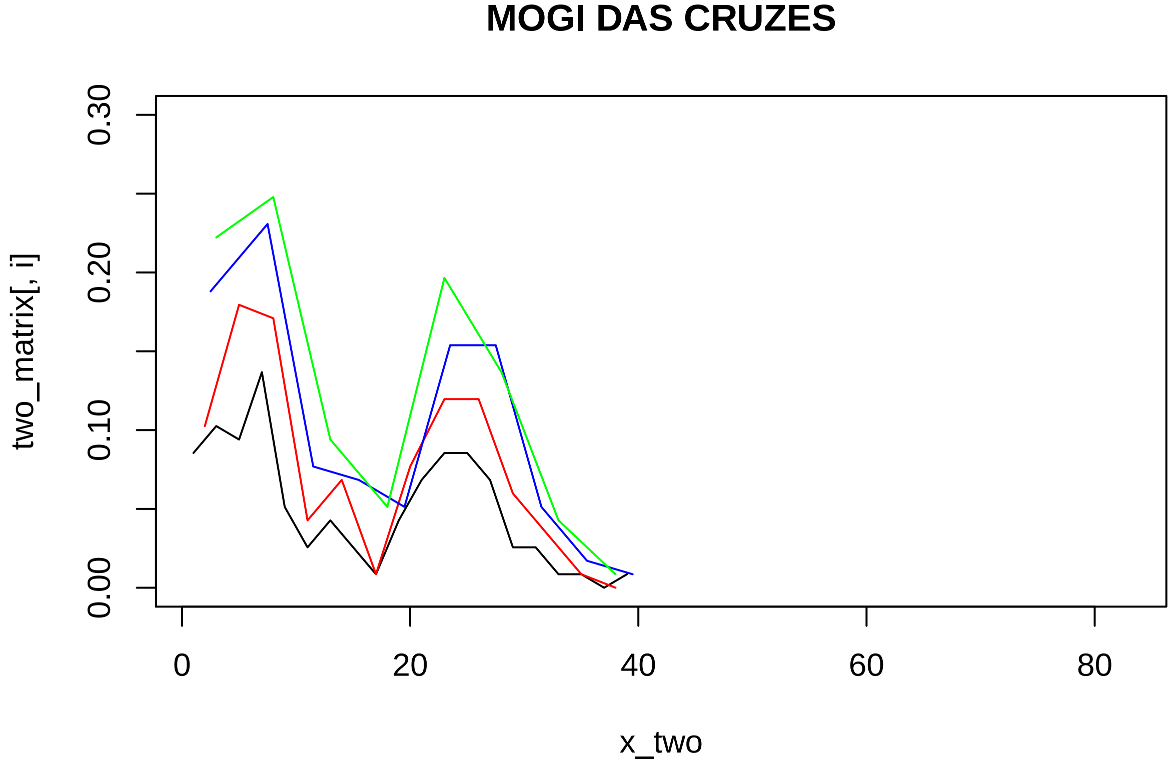


ARARAS

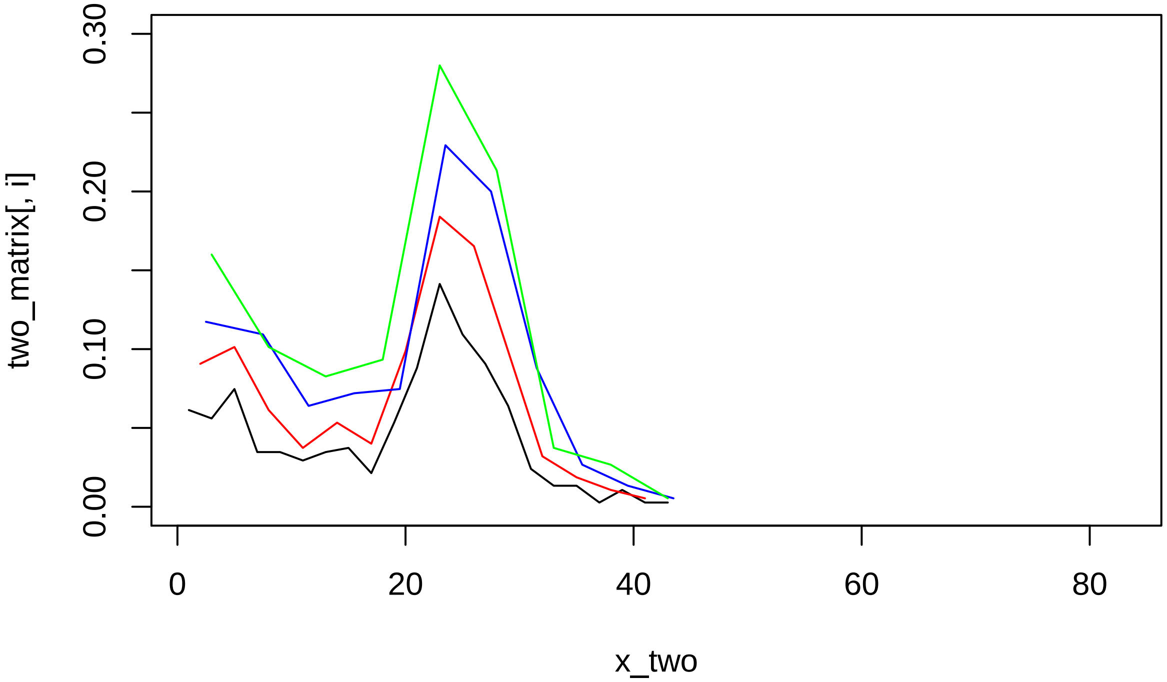

PINDAMONHANGABA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


FRANCO DA ROCHA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

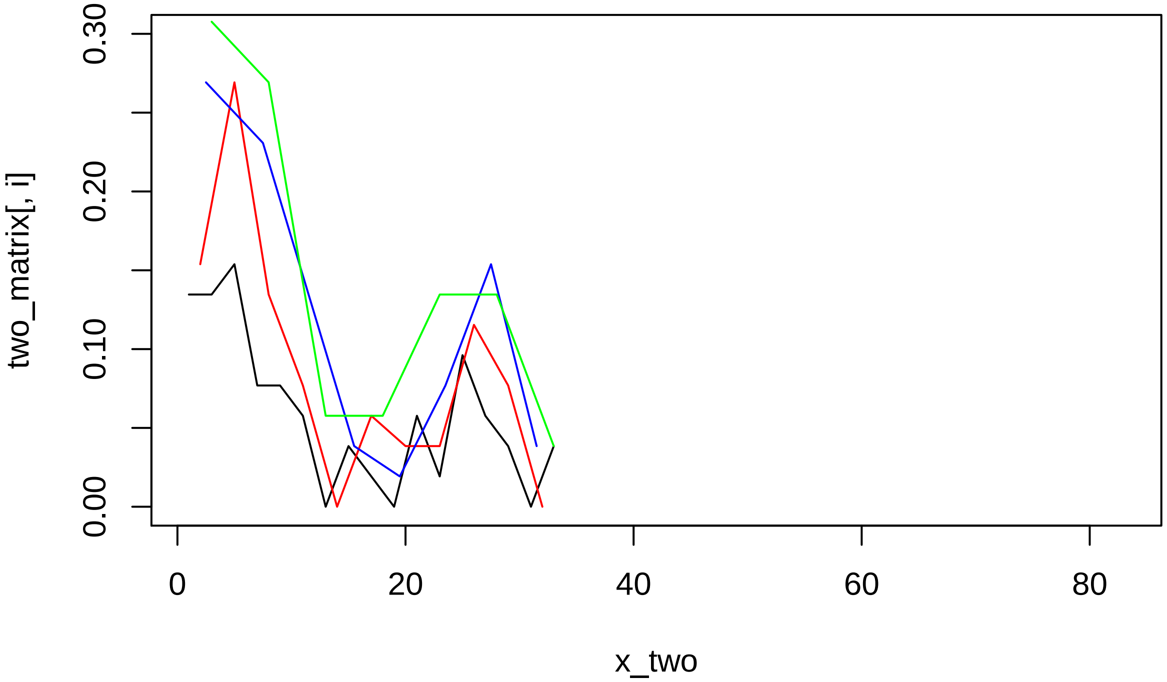

MAUA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

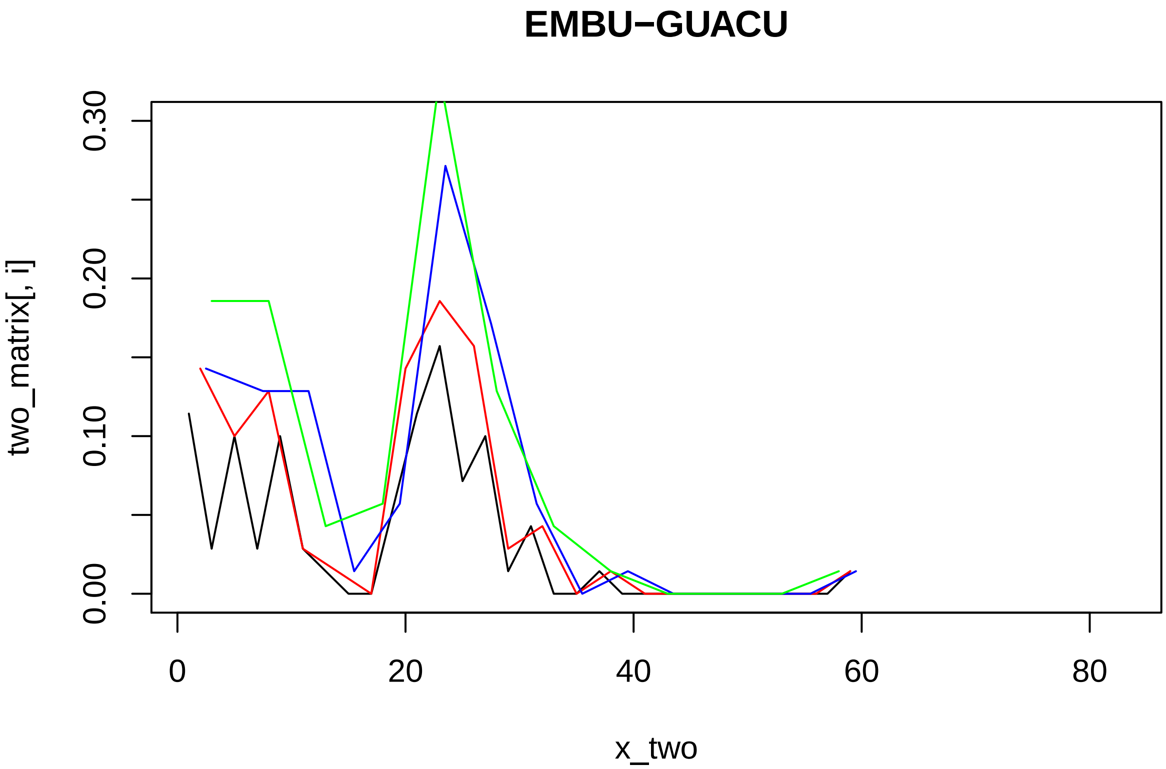

MOGI DAS CRUZES
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

CARAPICUIBA

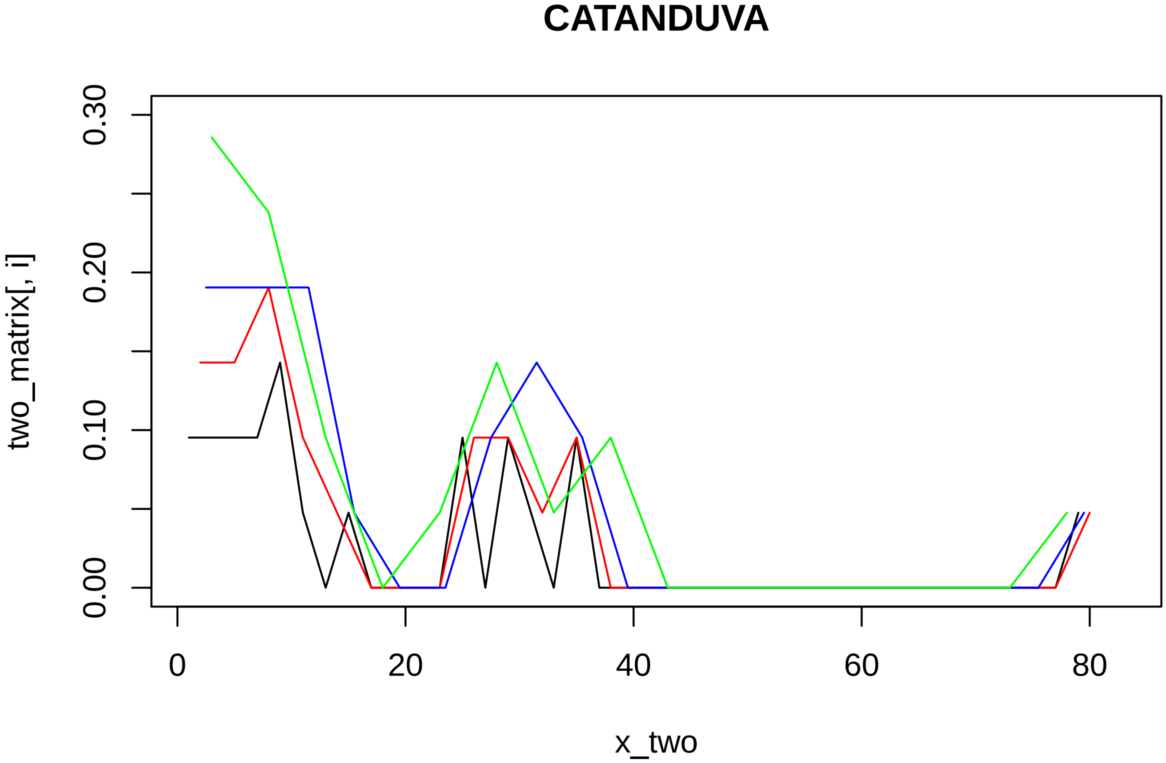


SAO JOSE DO RIO PRETO



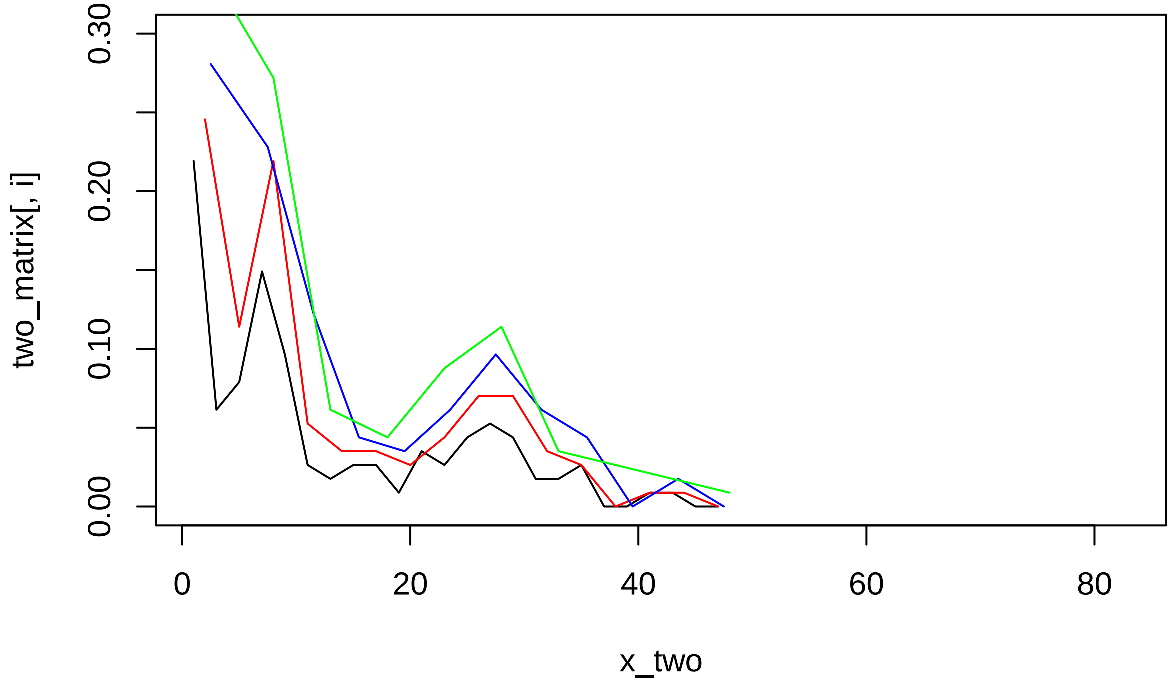


EMBU−GUACU

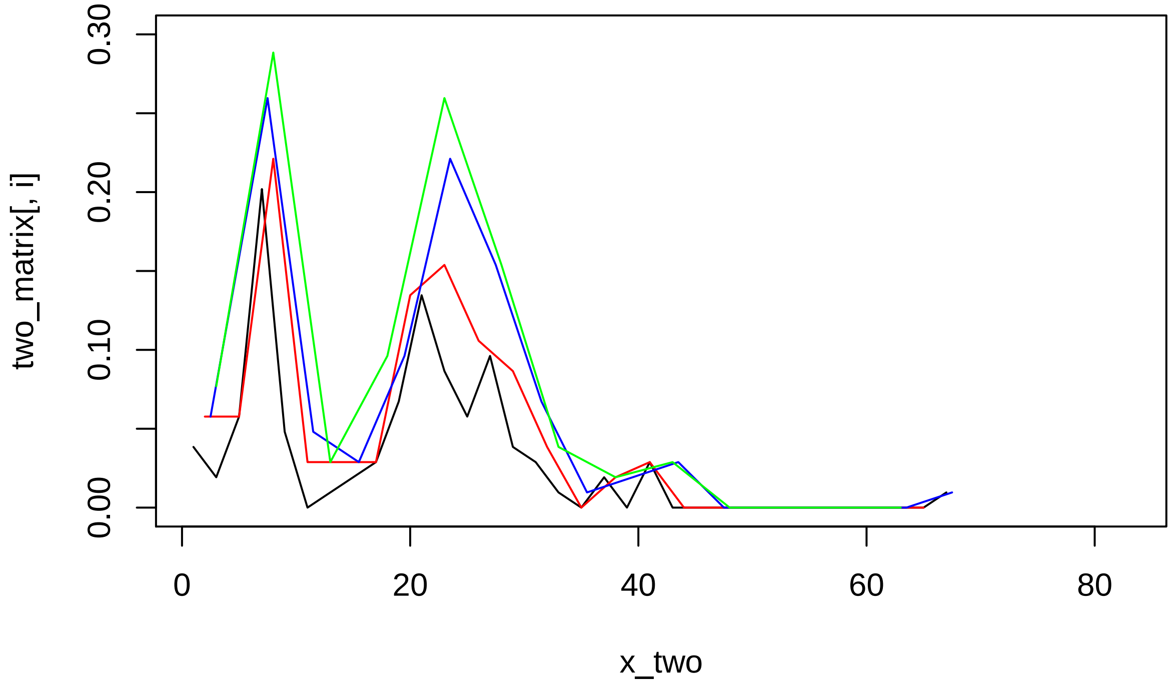


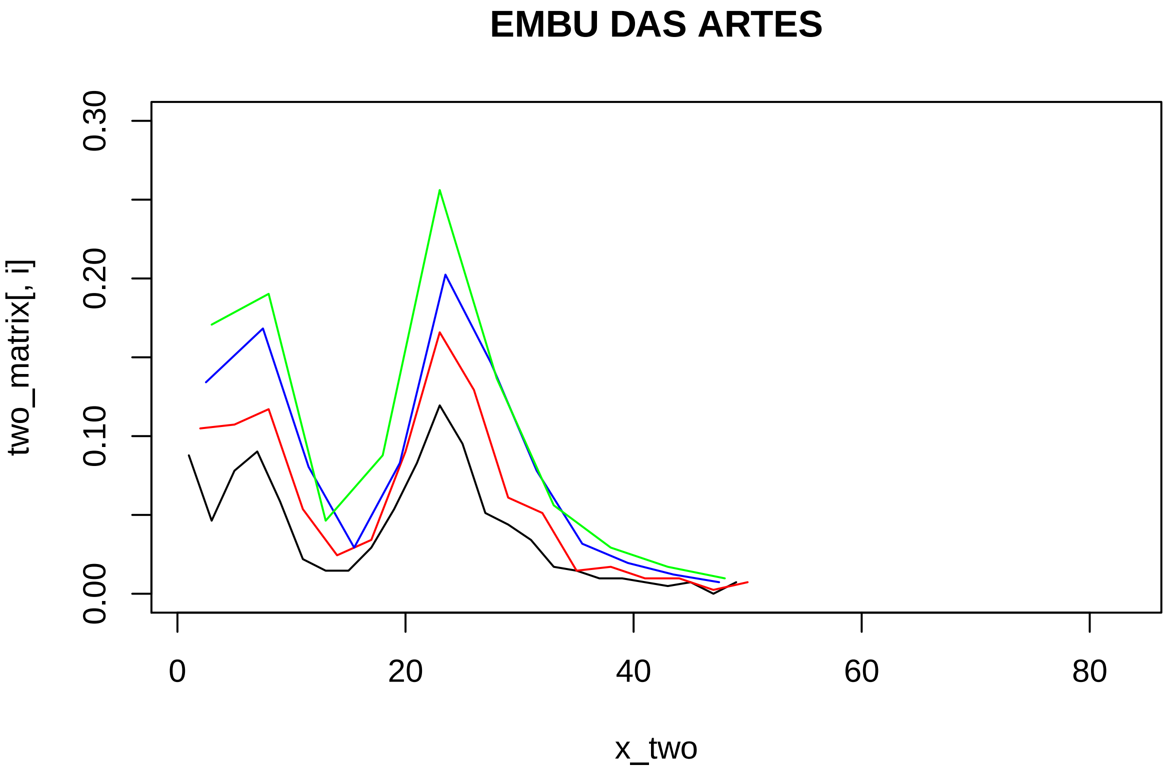
CATANDUVA

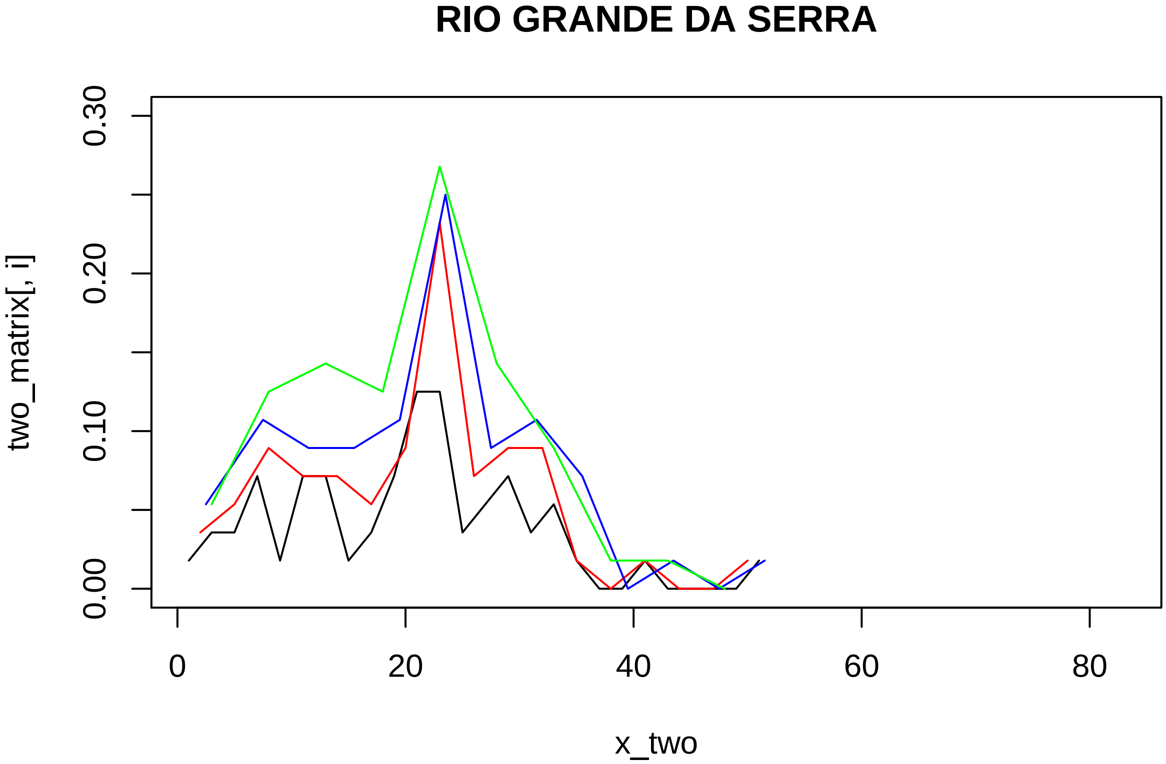
SAO JOSE DOS CAMPOS



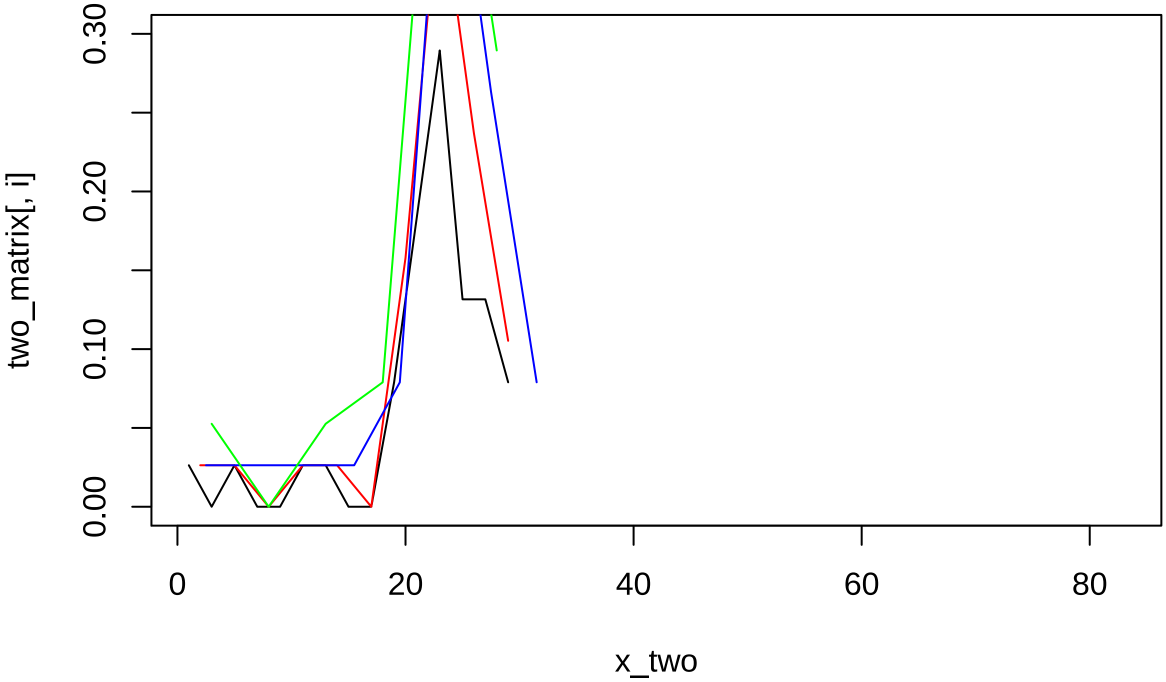
SANTA CRUZ DAS PALMEIRAS



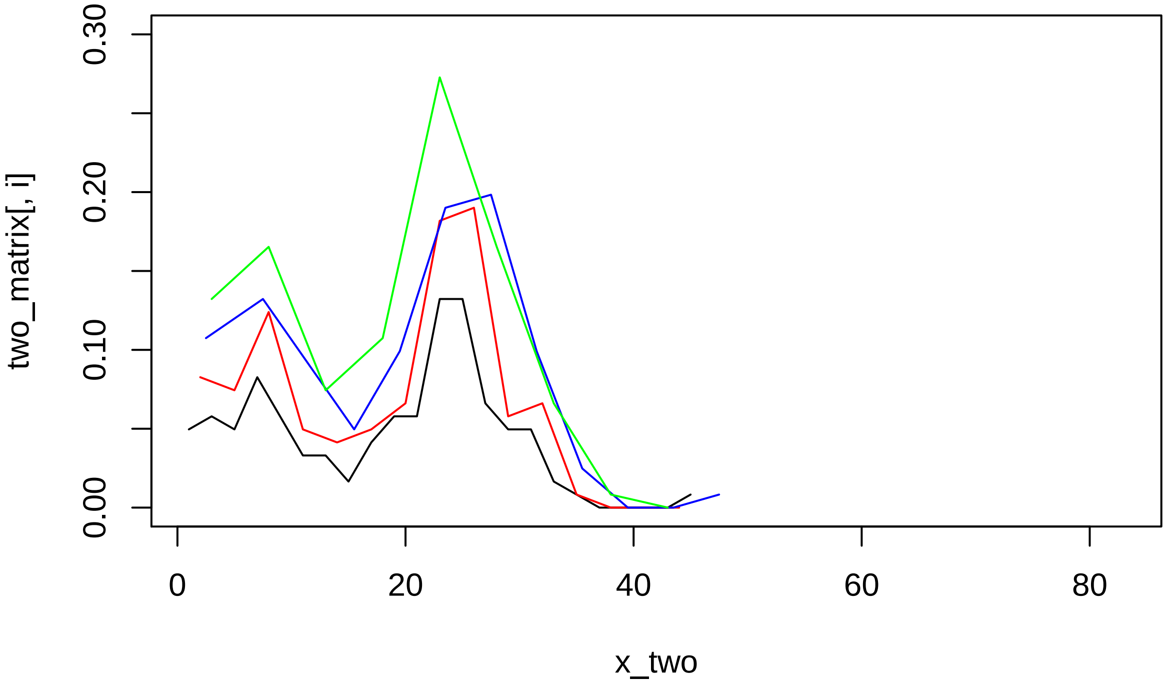

EMBU DAS ARTES

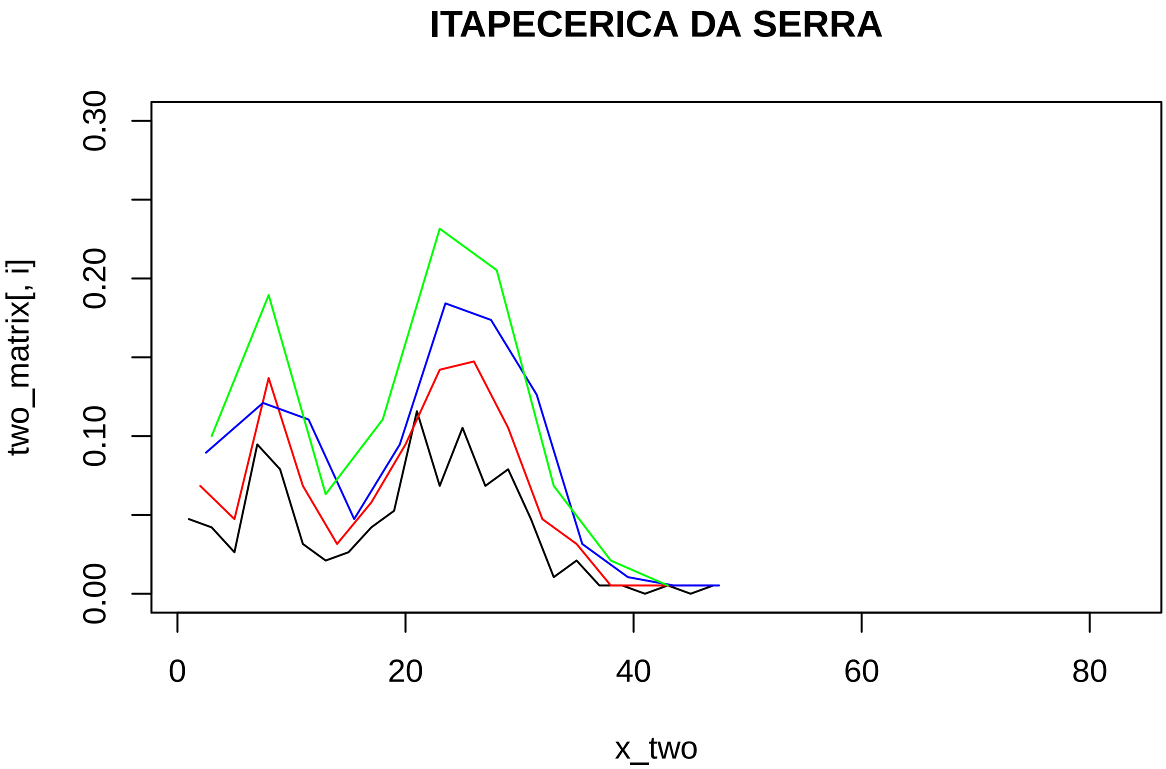

RIO GRANDE DA SERRA

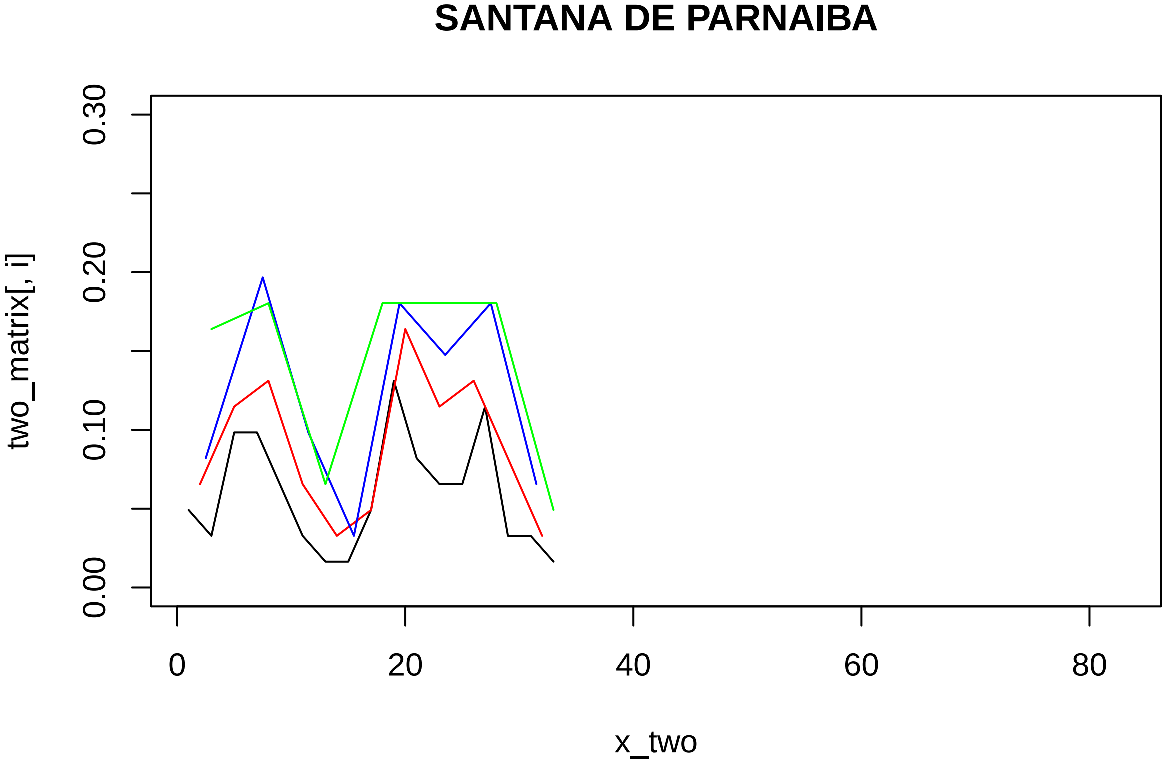
PONTAL

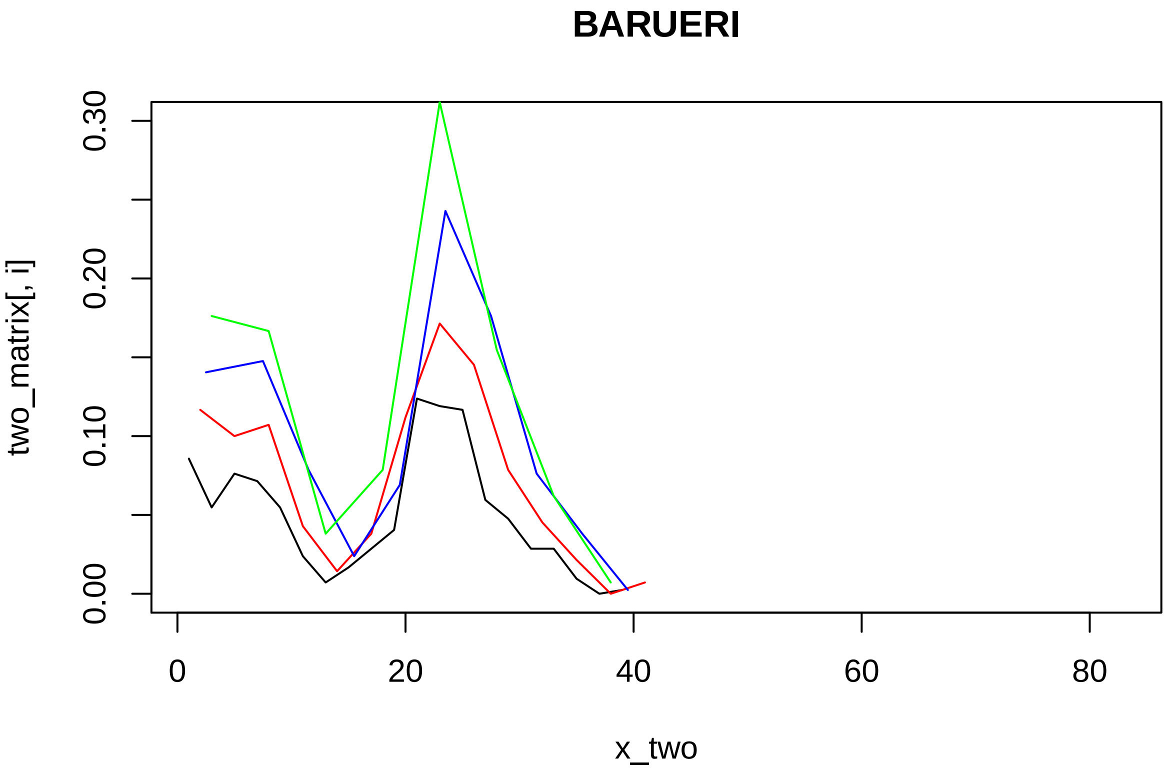


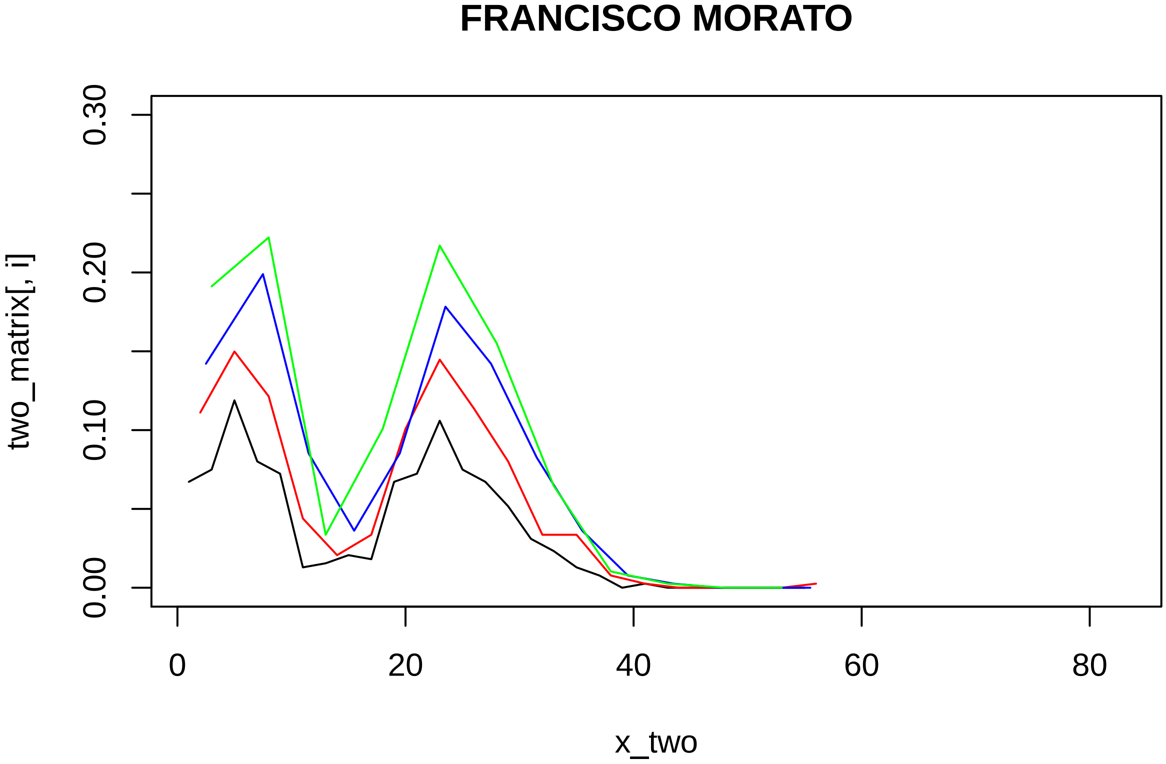
RIBEIRAO PIRES

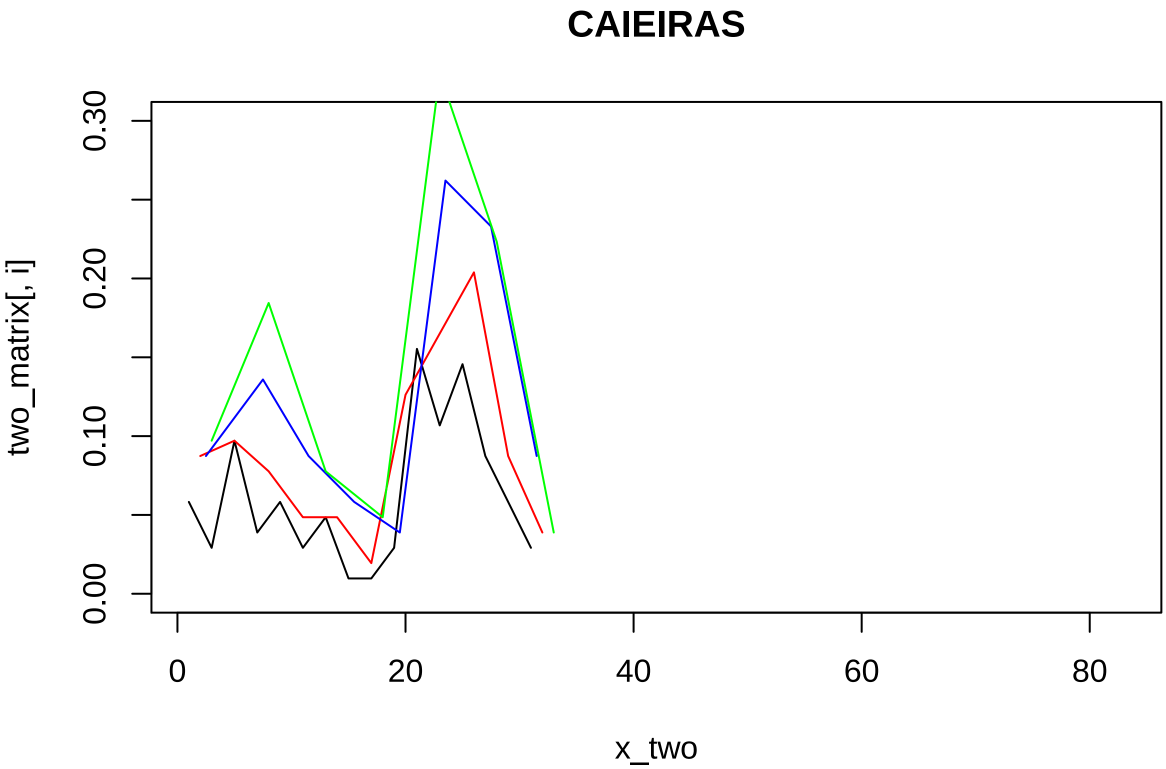


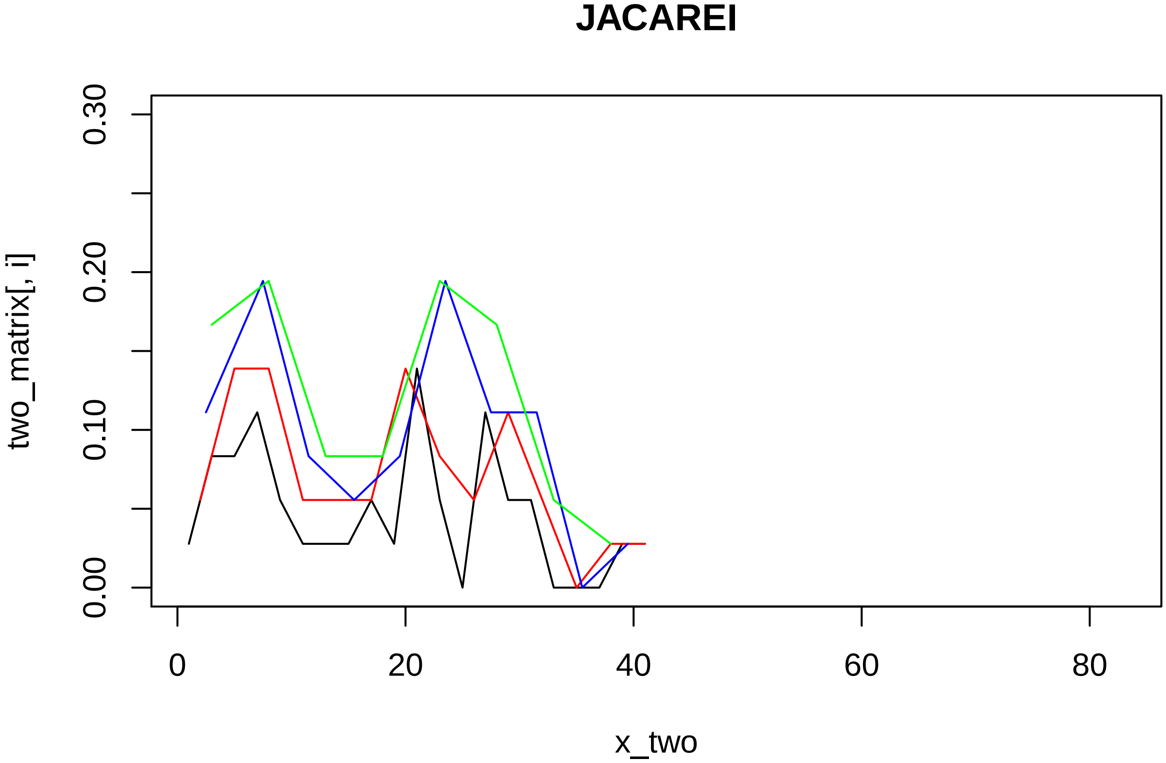

ITAPECERICA DA SERRA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


SANTANA DE PARNAIBA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

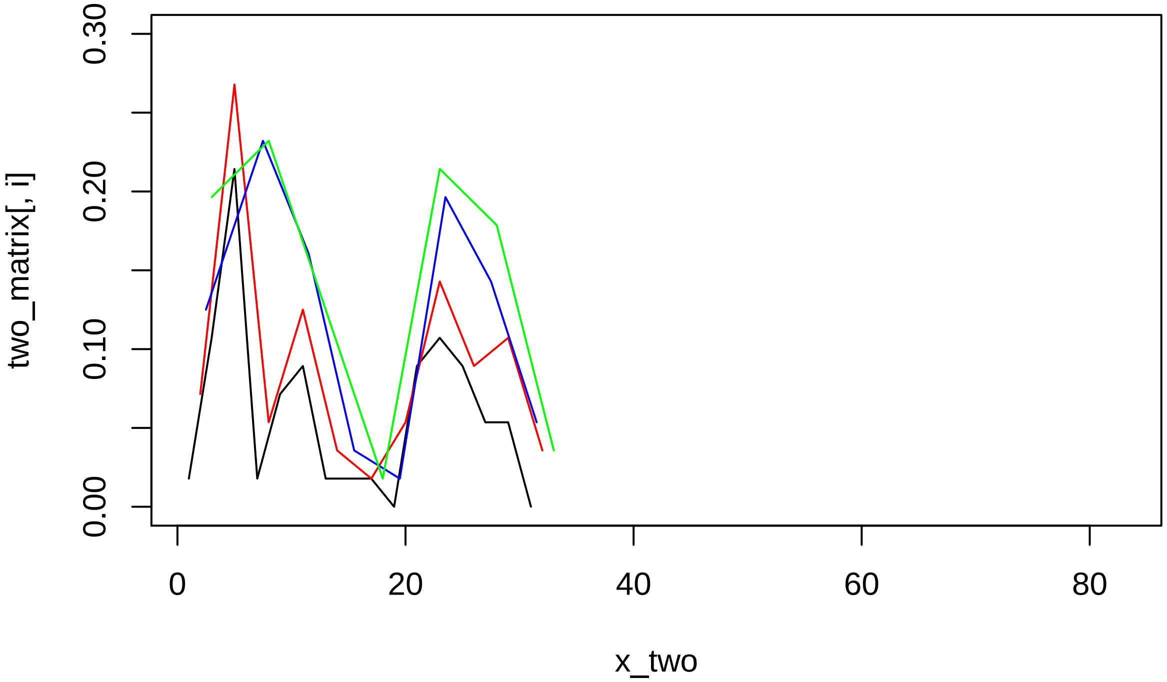

BARUERI
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


FRANCISCO MORATO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


CAIEIRAS
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

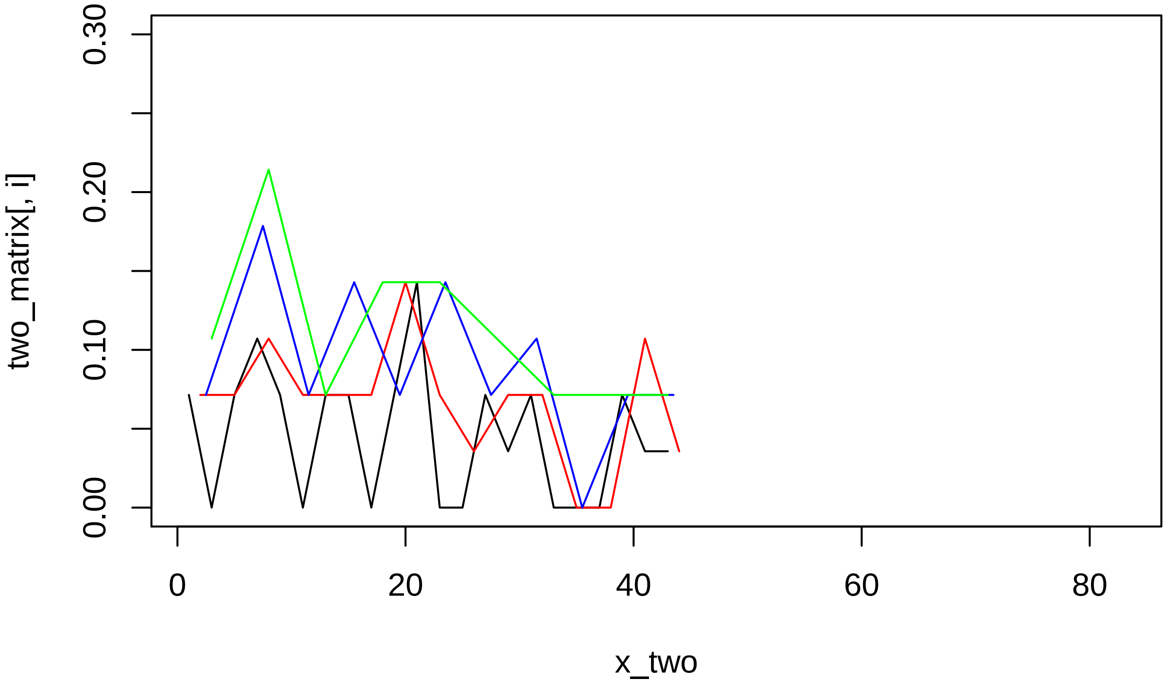

JACAREI
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

ITU

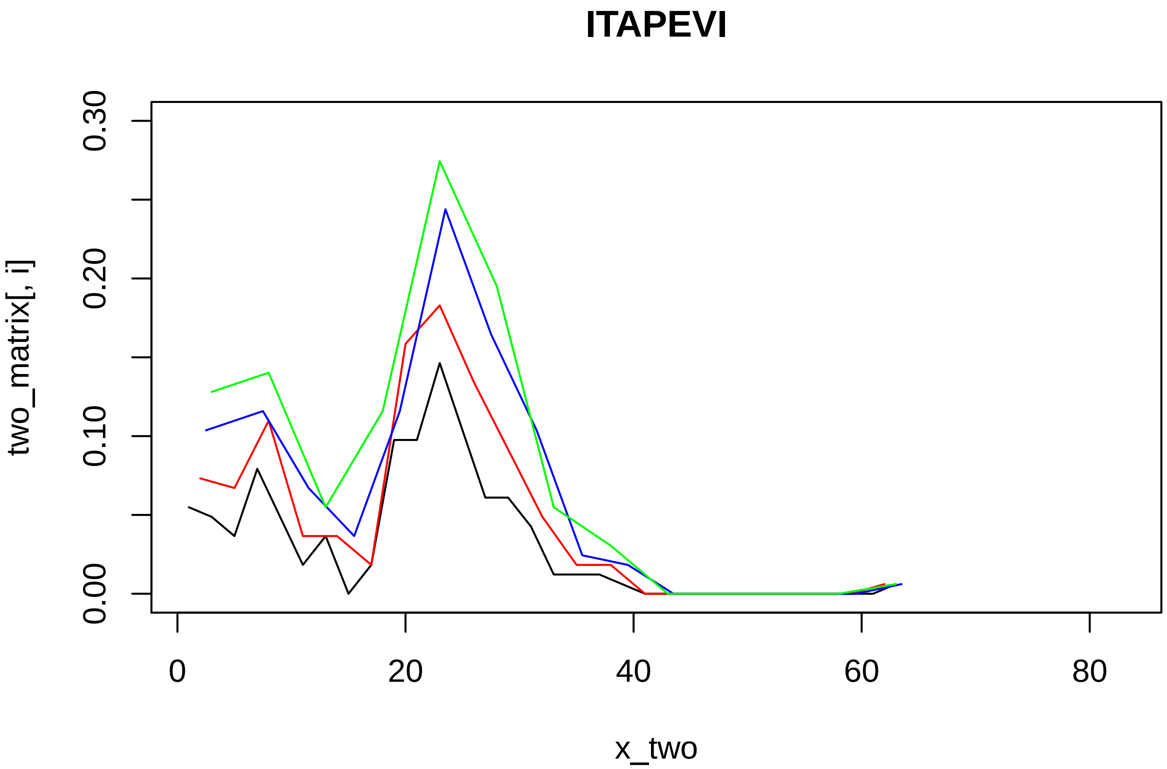


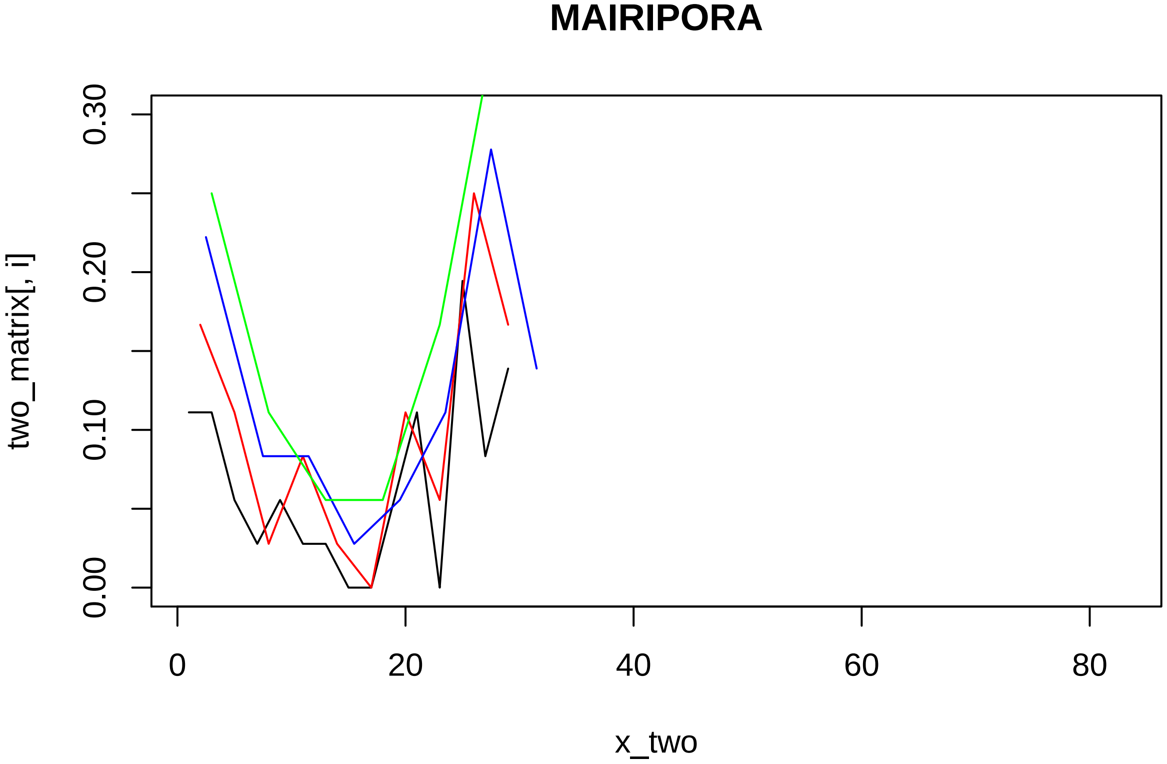
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

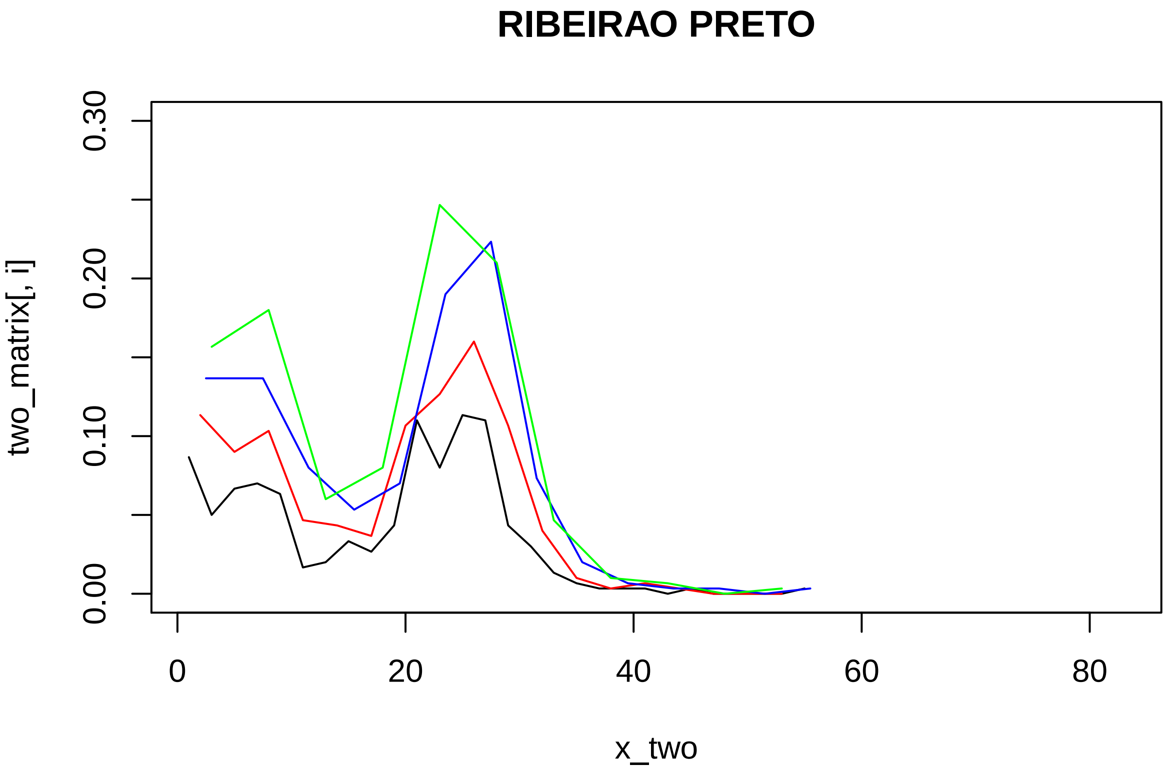
ATIBAIA

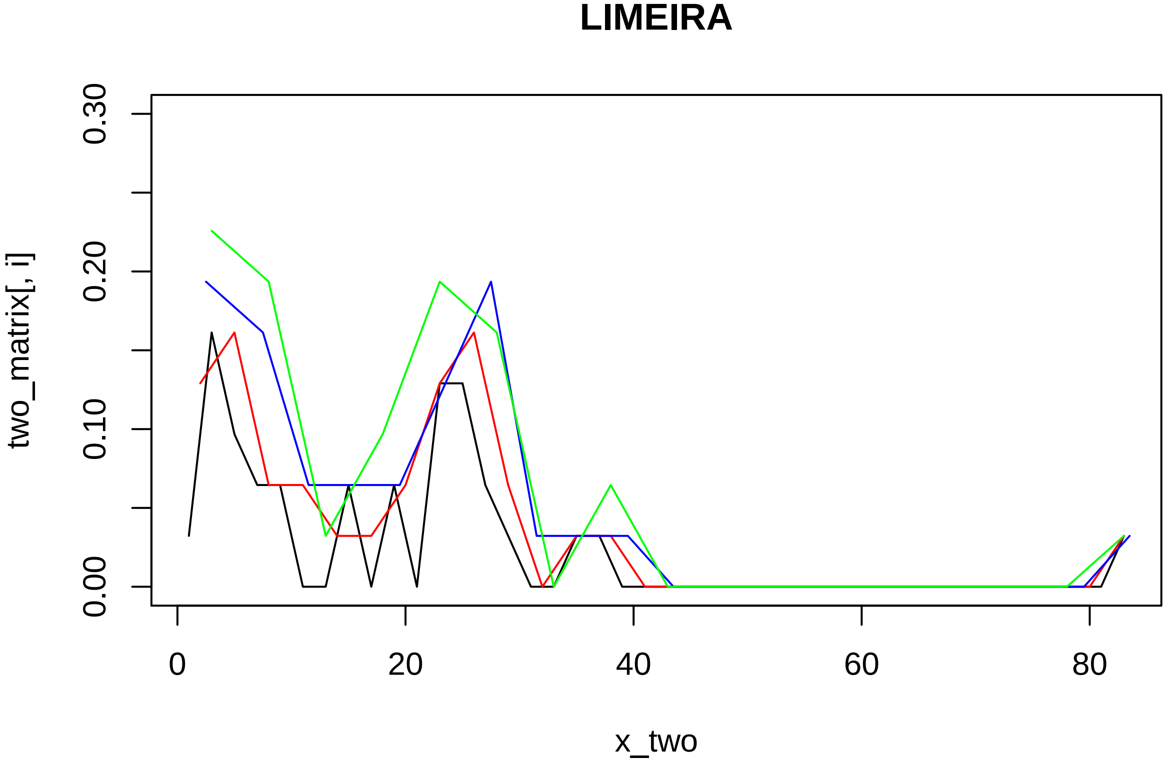


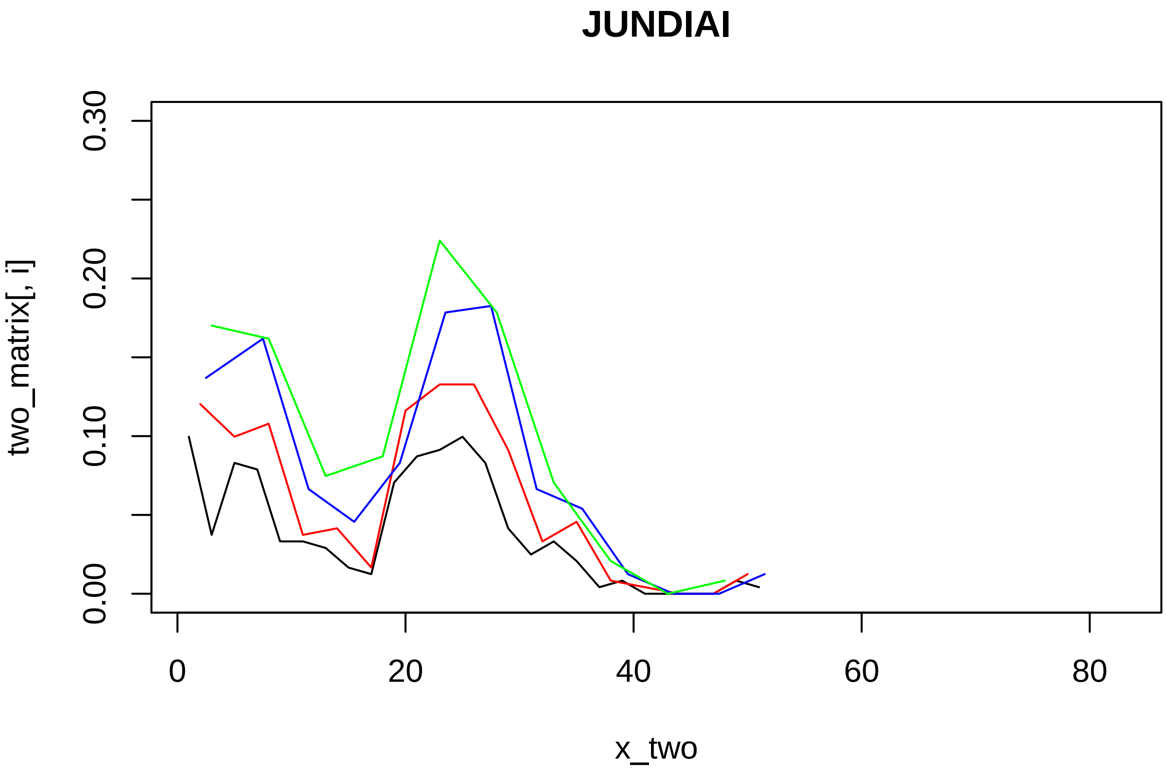
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


ITAPEVI
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

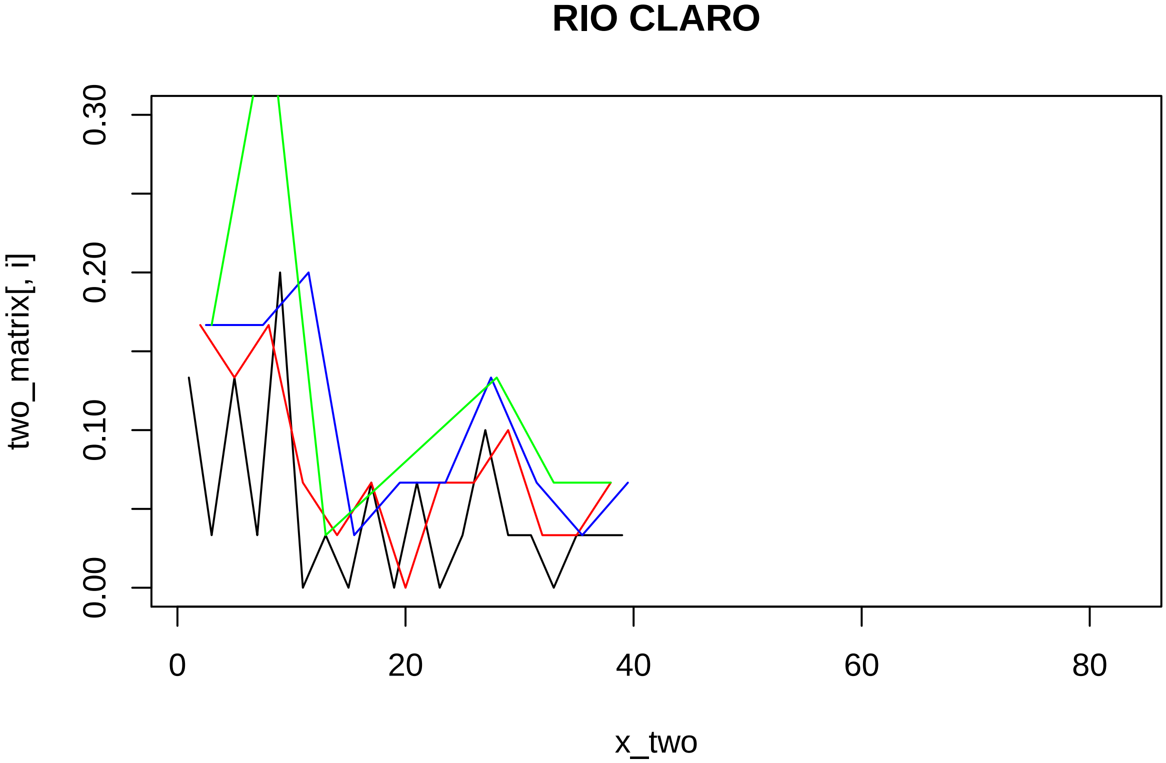

MAIRIPORA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


RIBEIRAO PRETO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

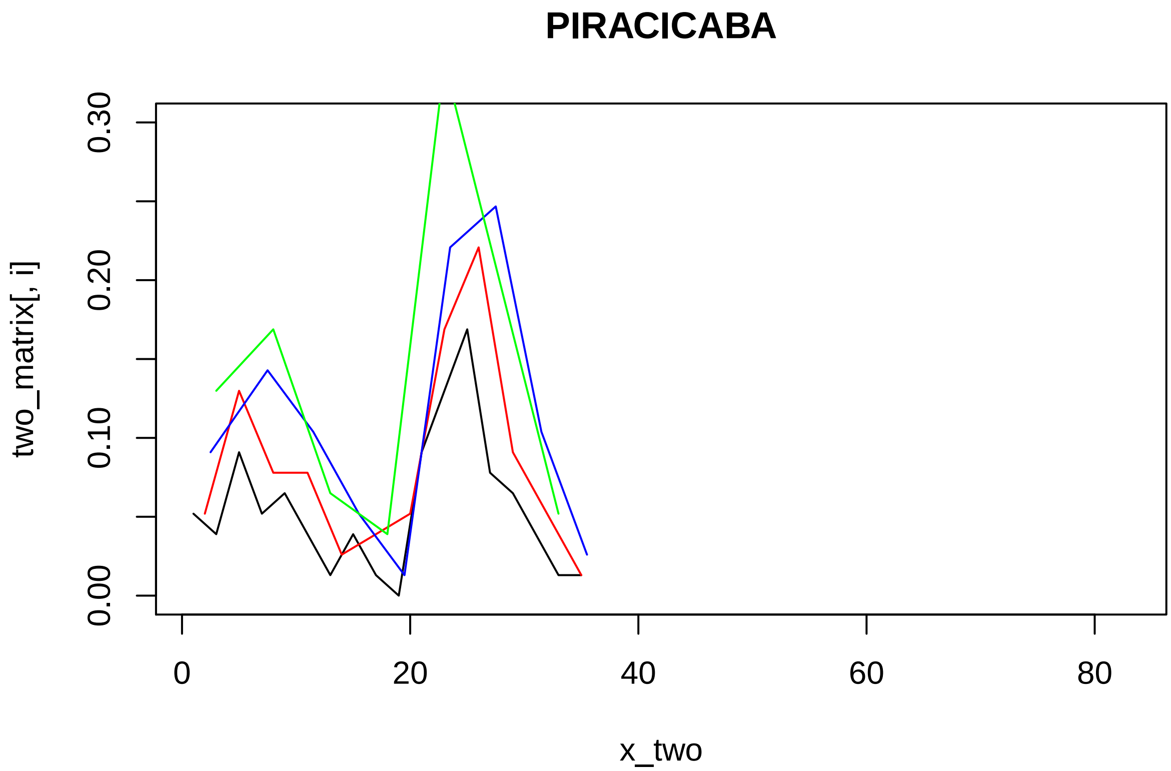

LIMEIRA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

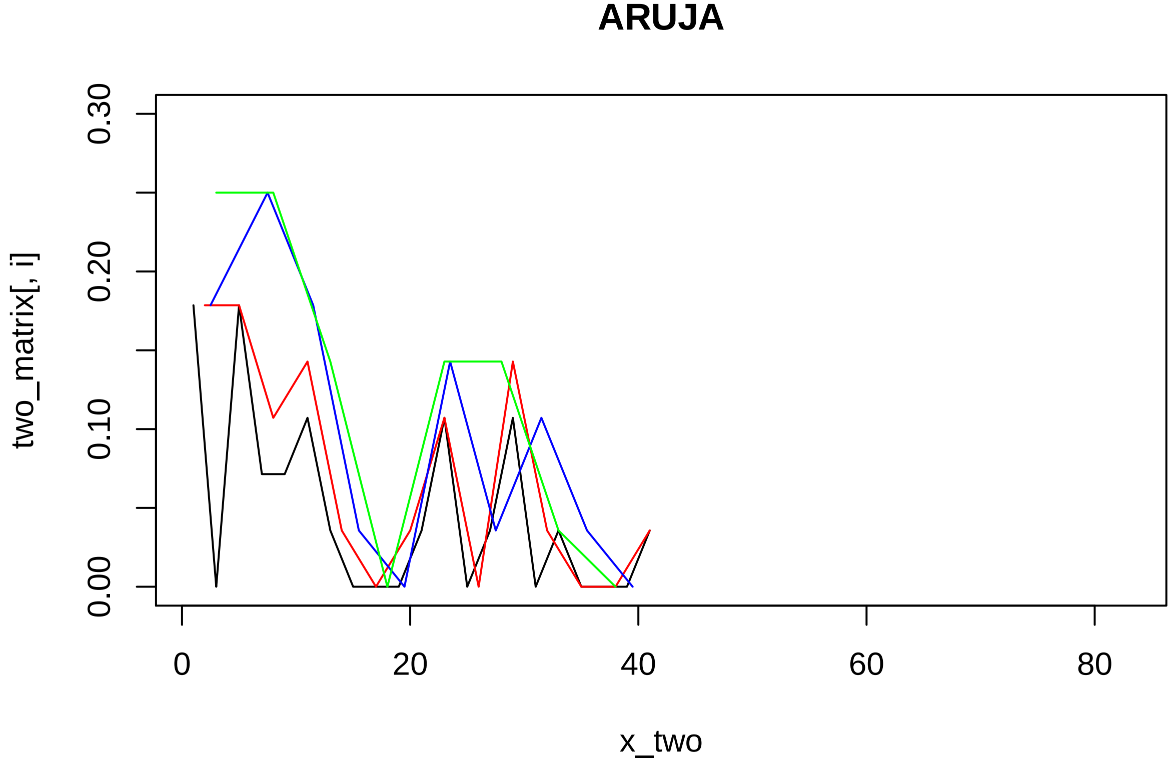


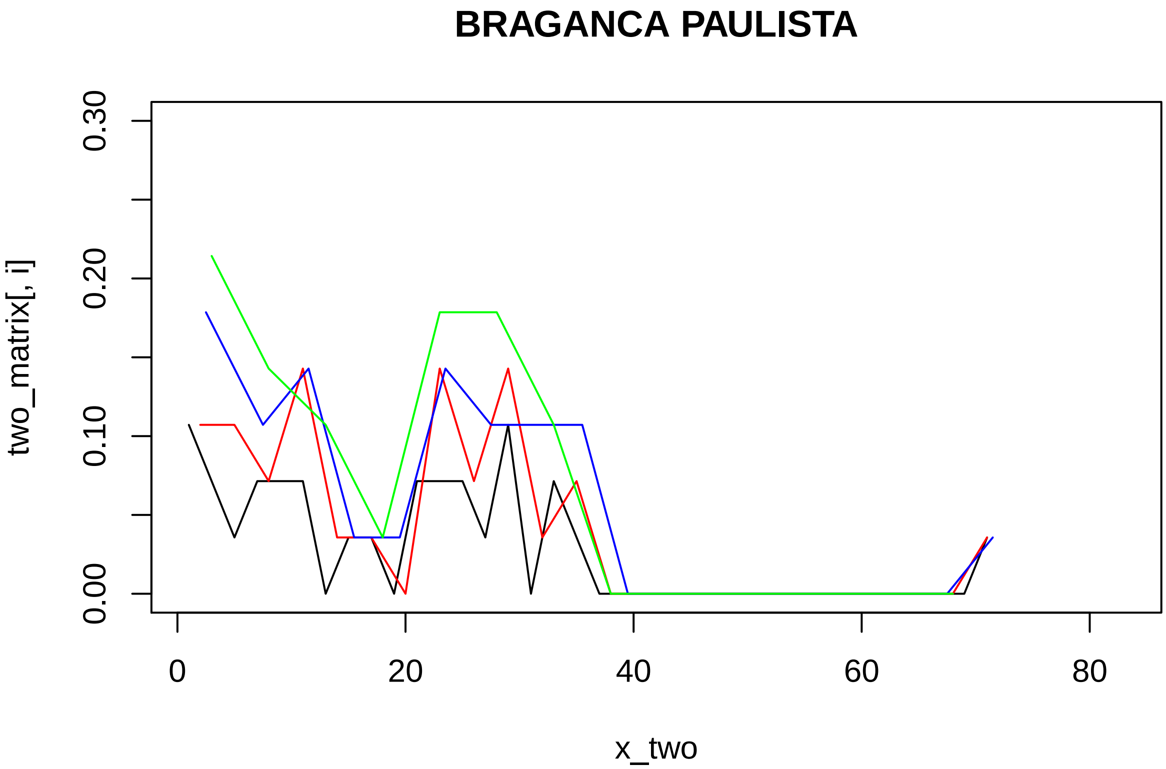
JUNDIAI

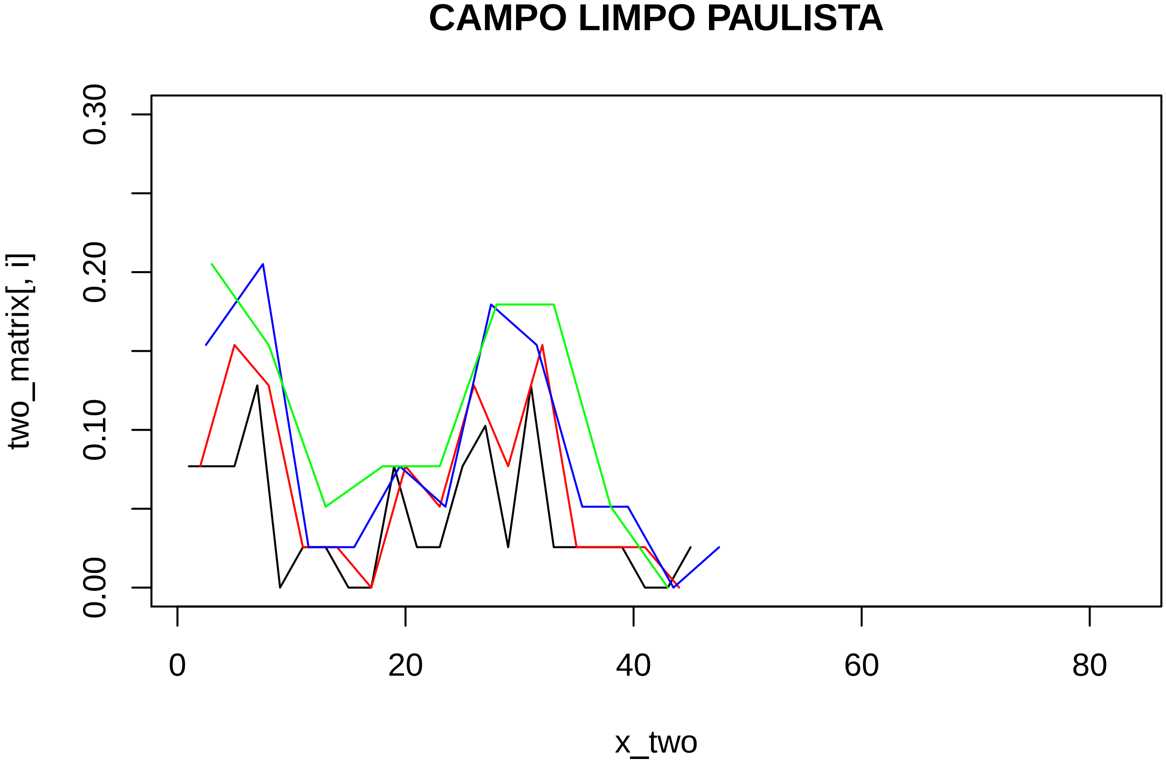


RIO CLARO

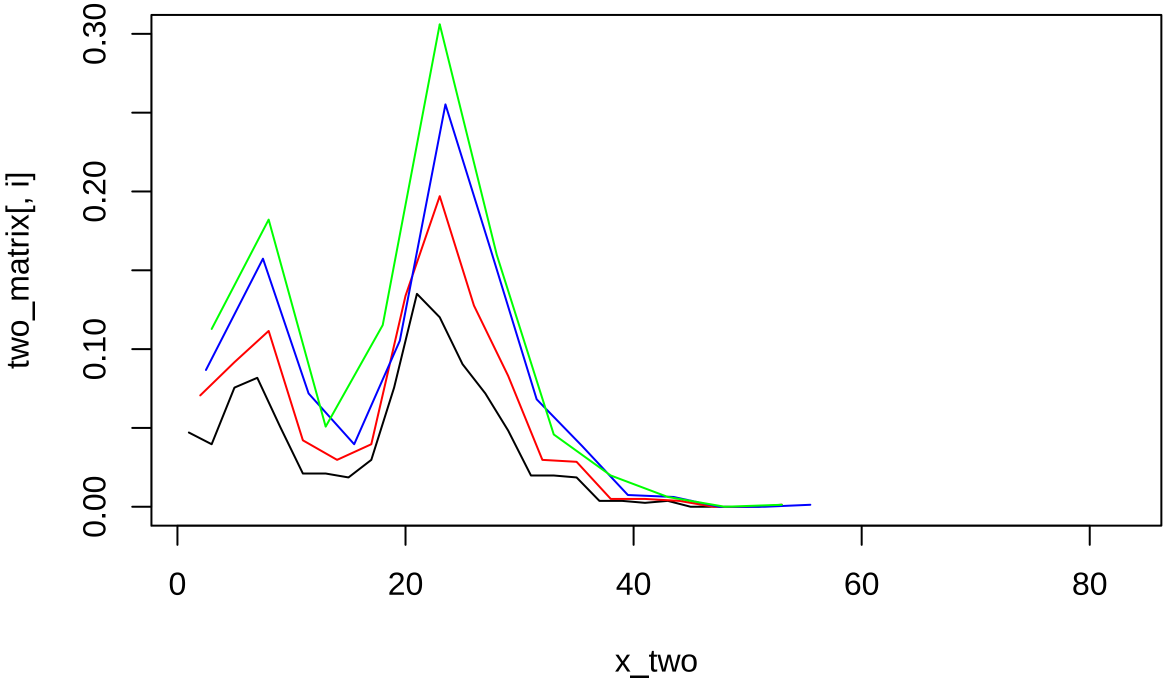

PIRACICABA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


ARUJA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

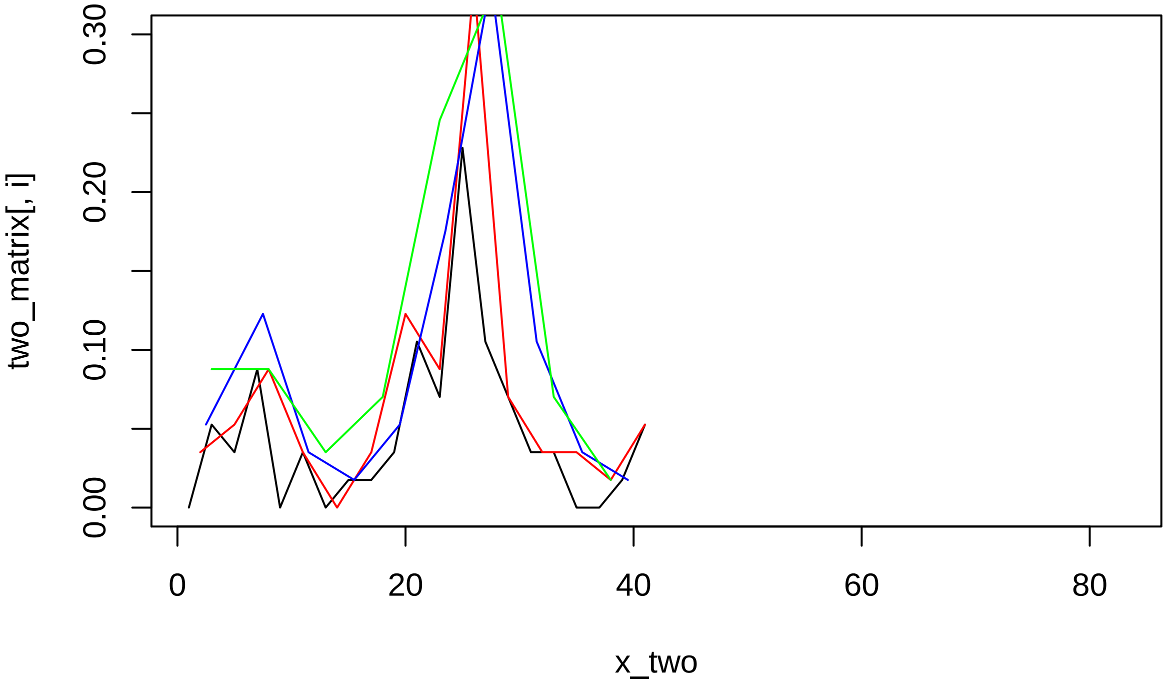

BRAGANCA PAULISTA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

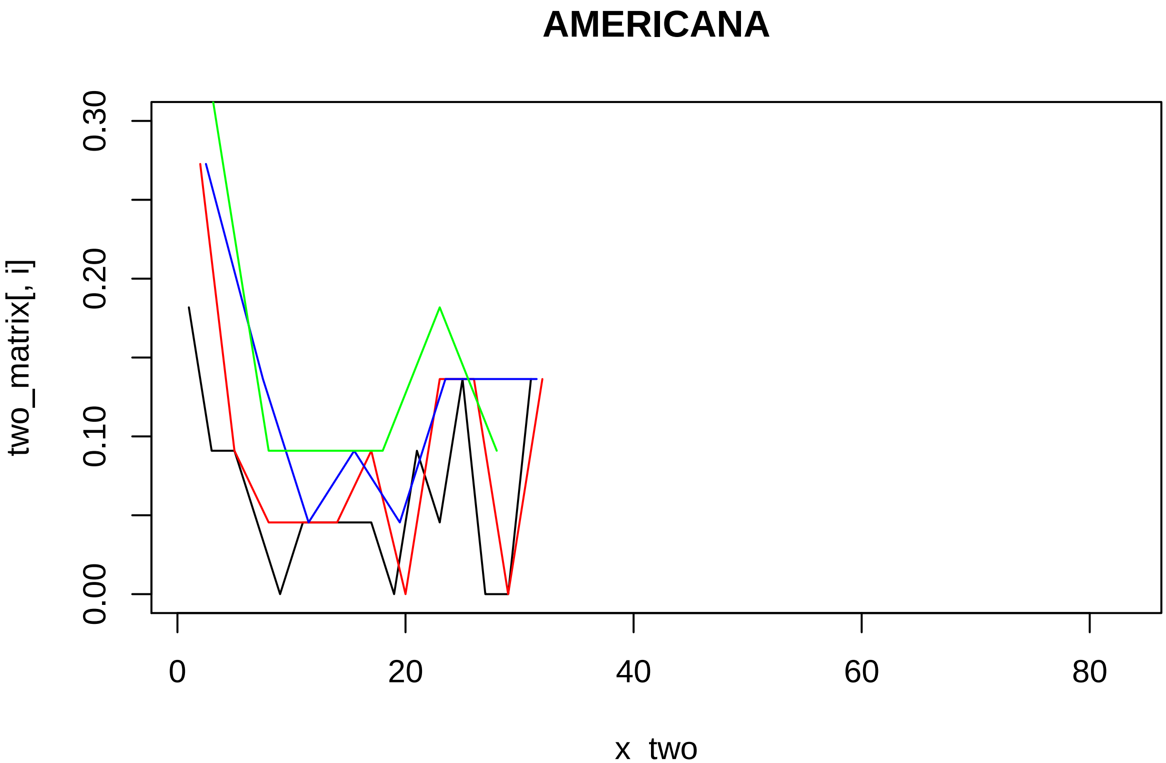

CAMPO LIMPO PAULISTA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

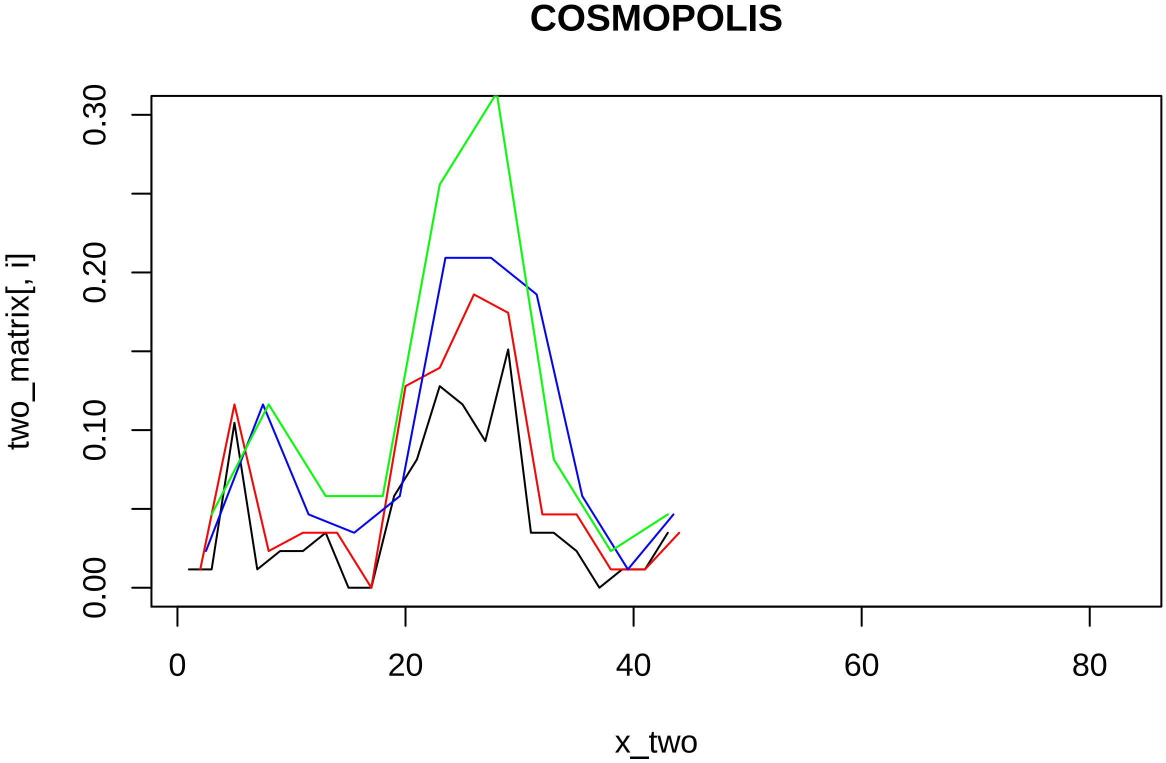
TABOAO DA SERRA

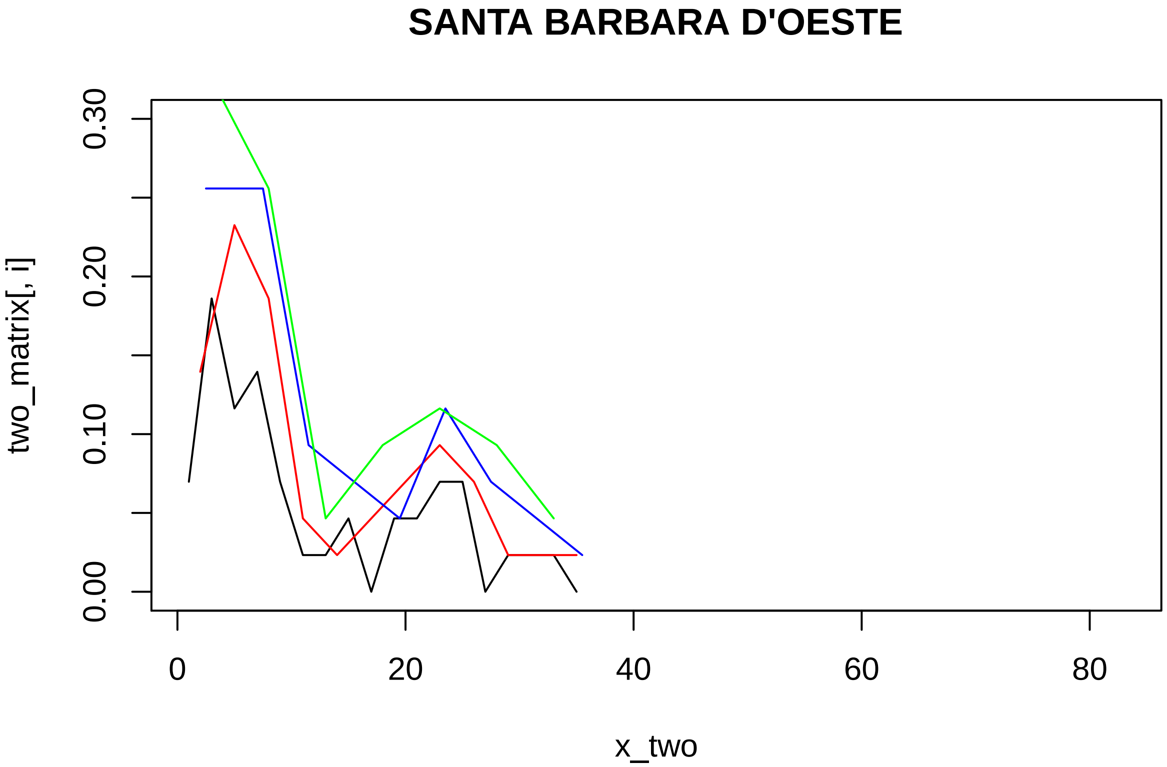


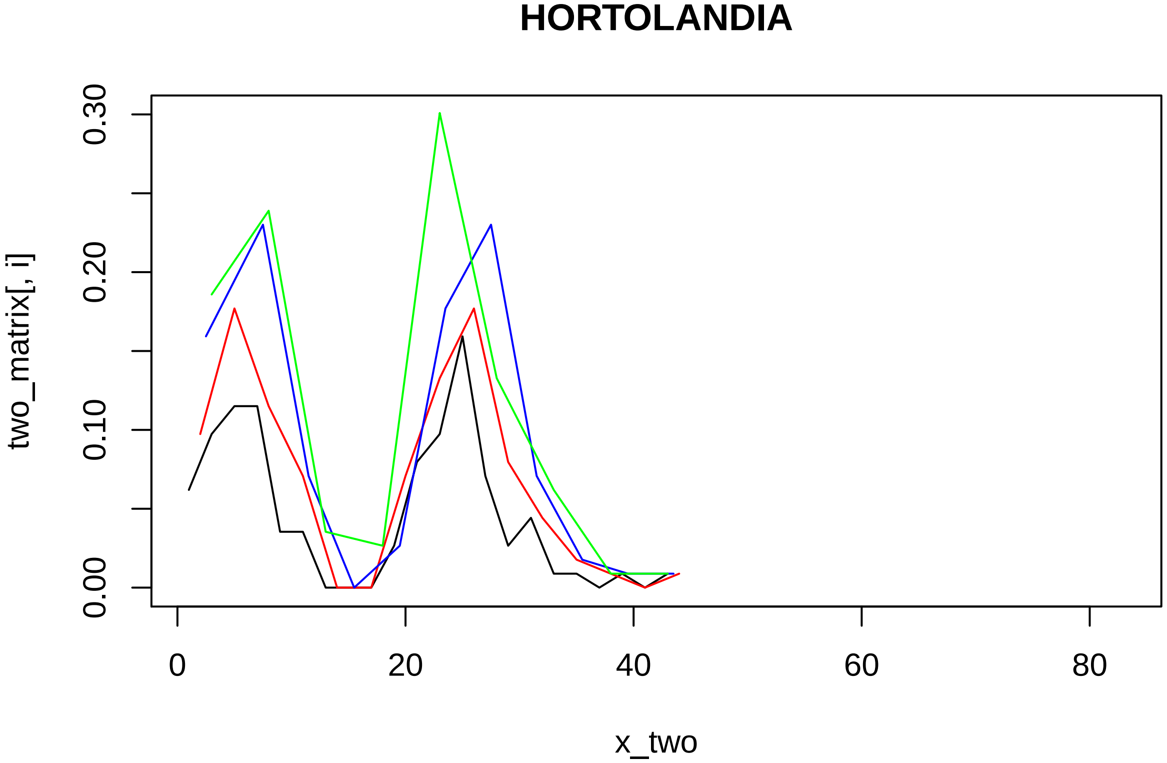
ITATIBA

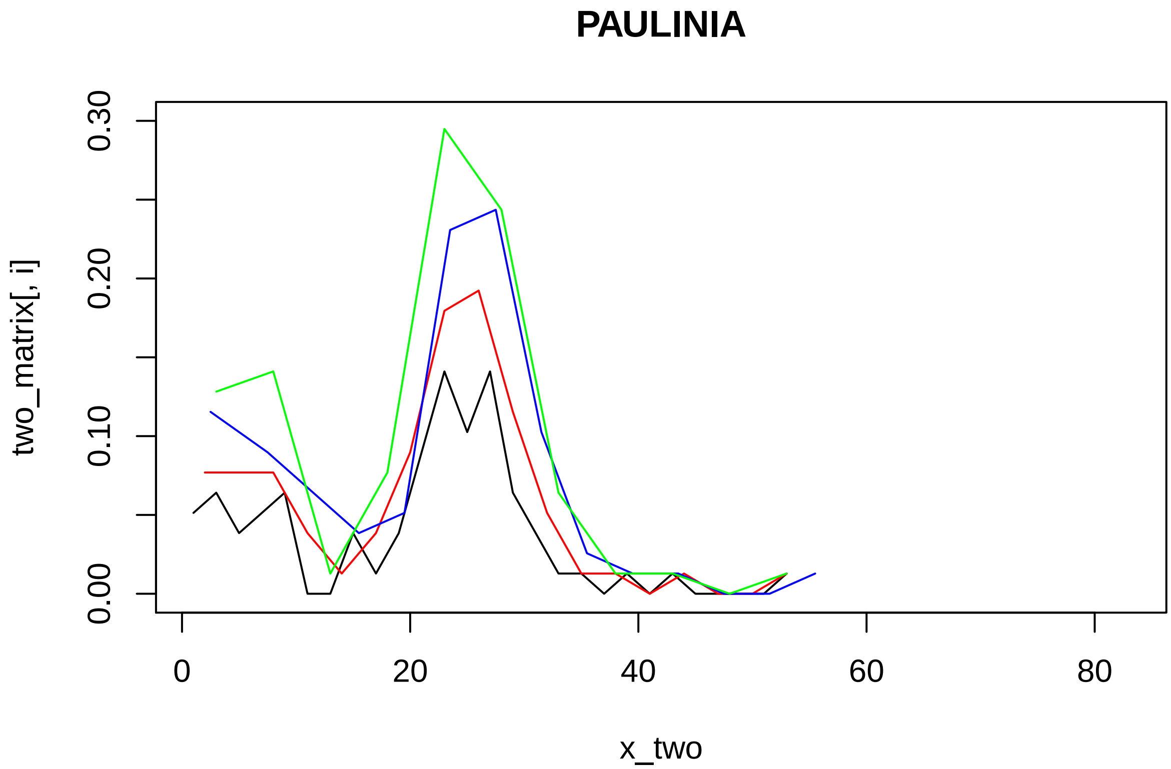


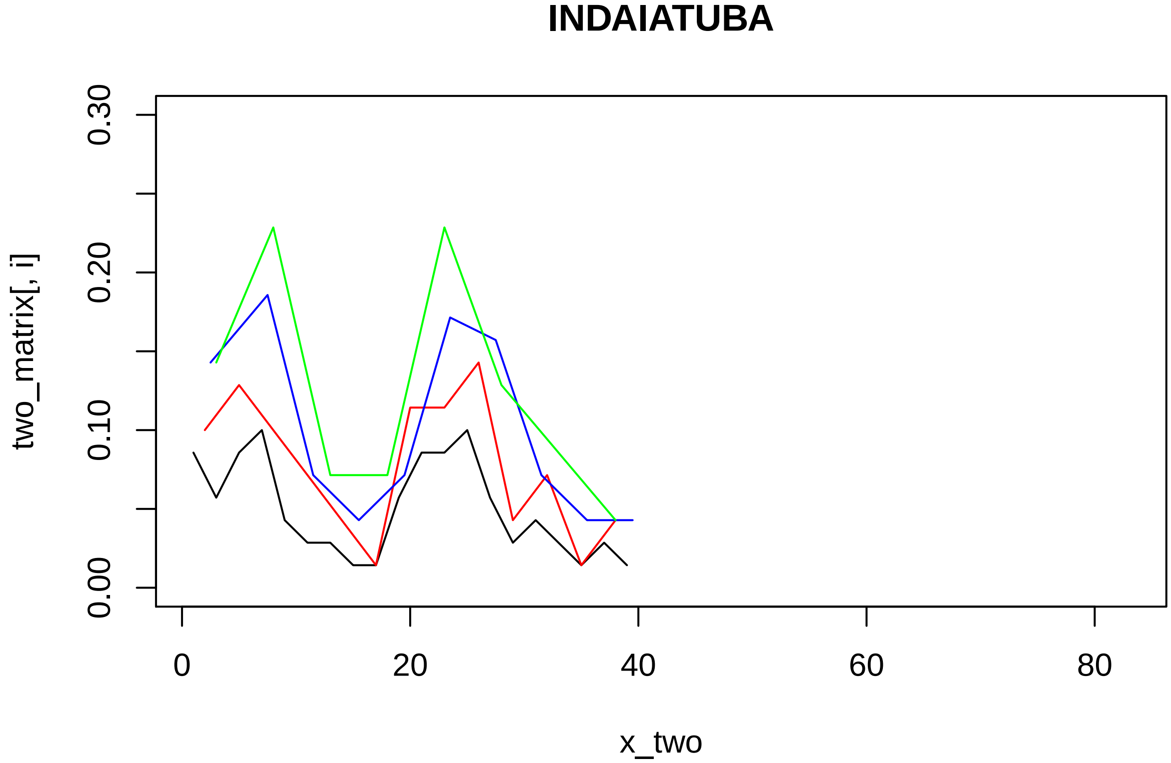

AMERICANA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

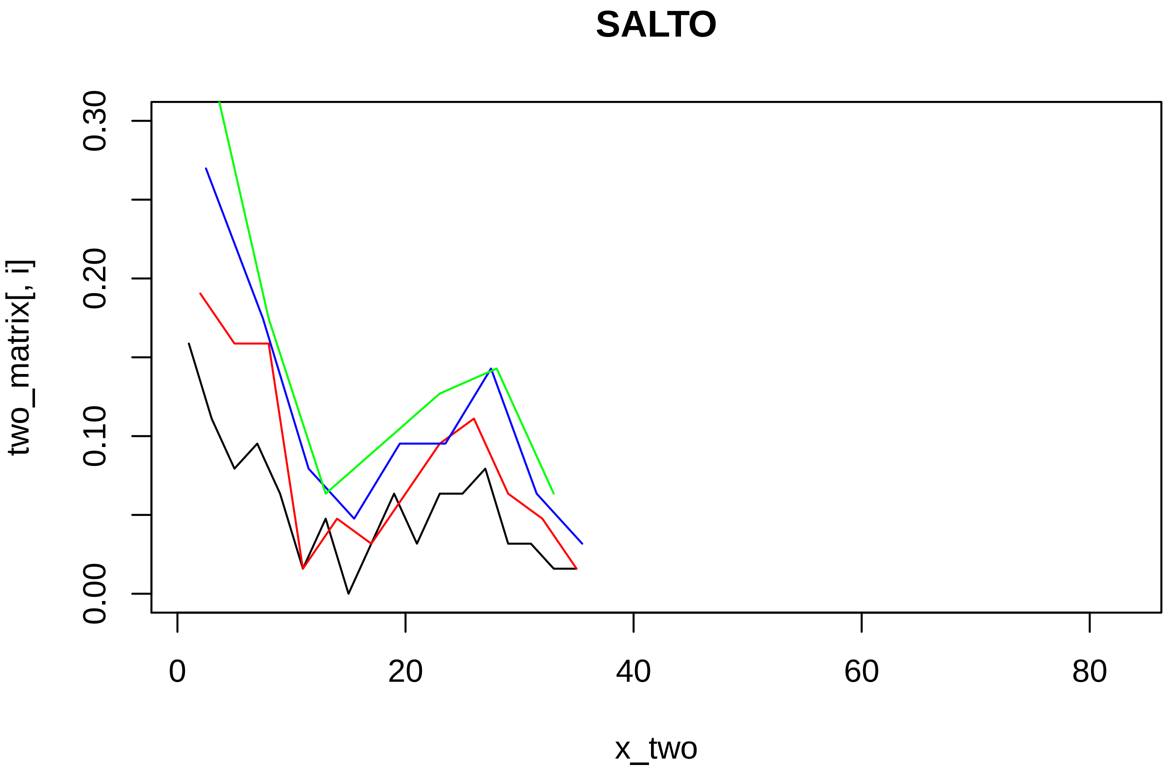

COSMOPOLIS
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

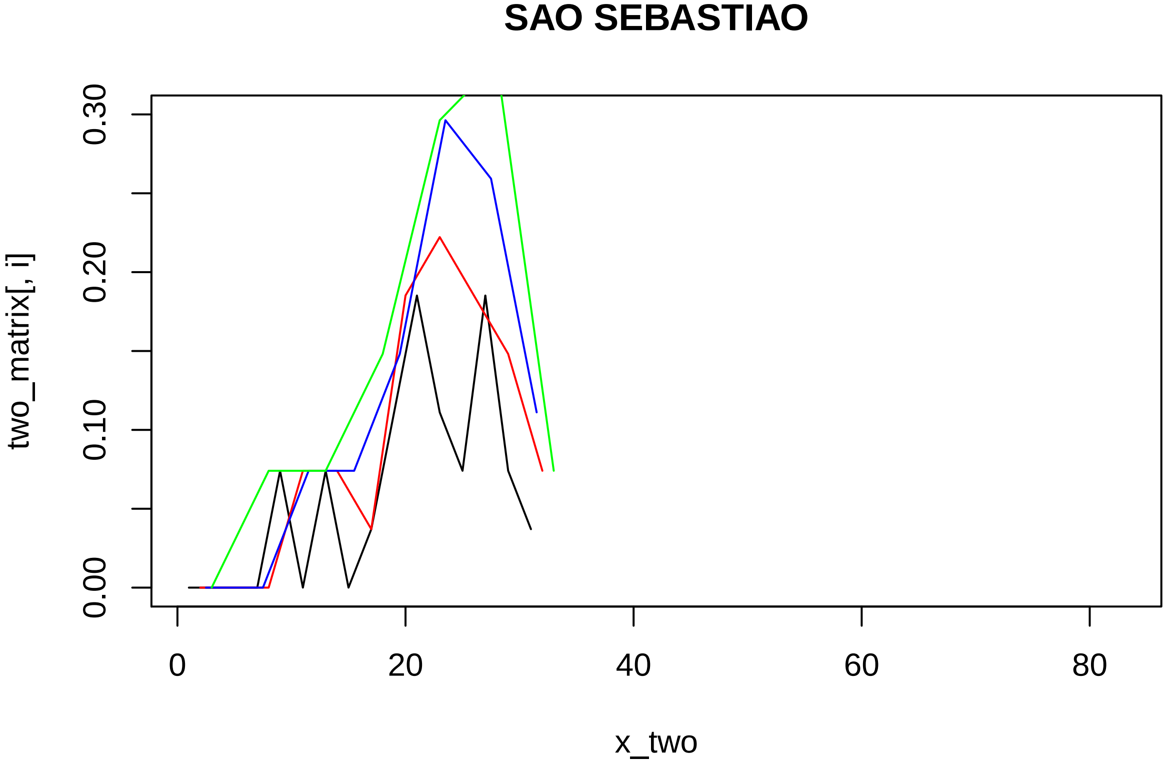

SANTA BARBARA D'OESTE
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

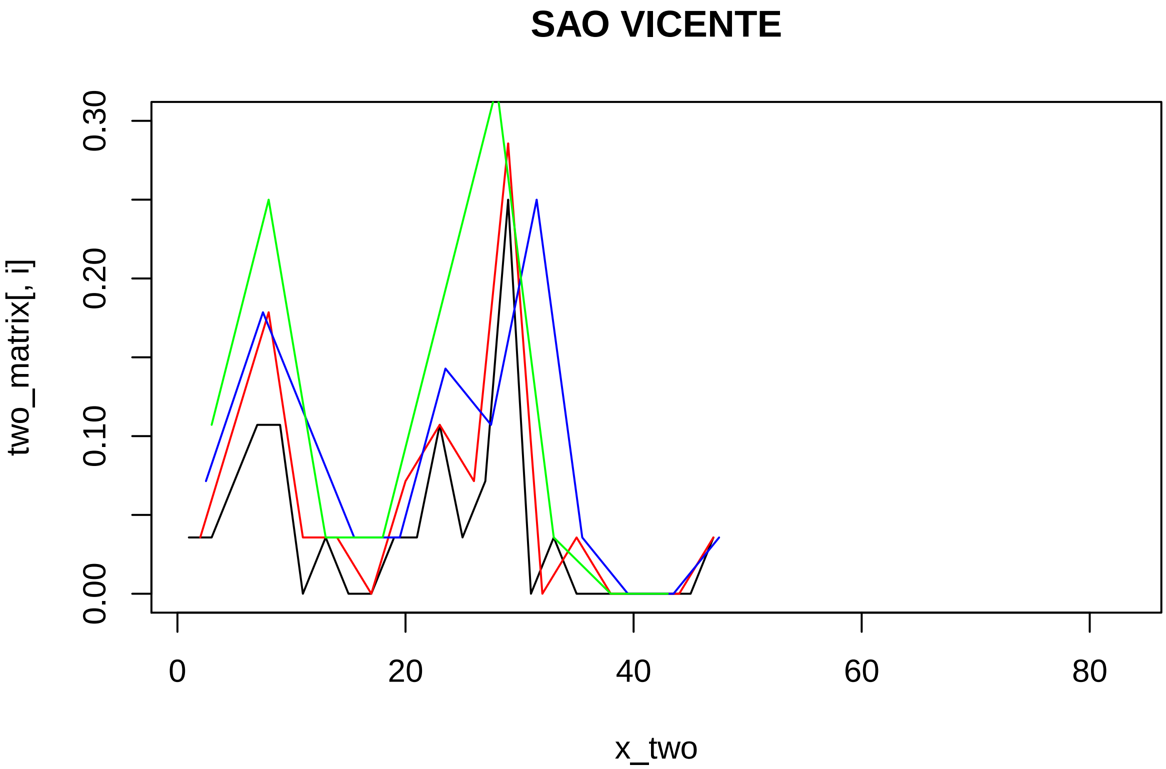

HORTOLANDIA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

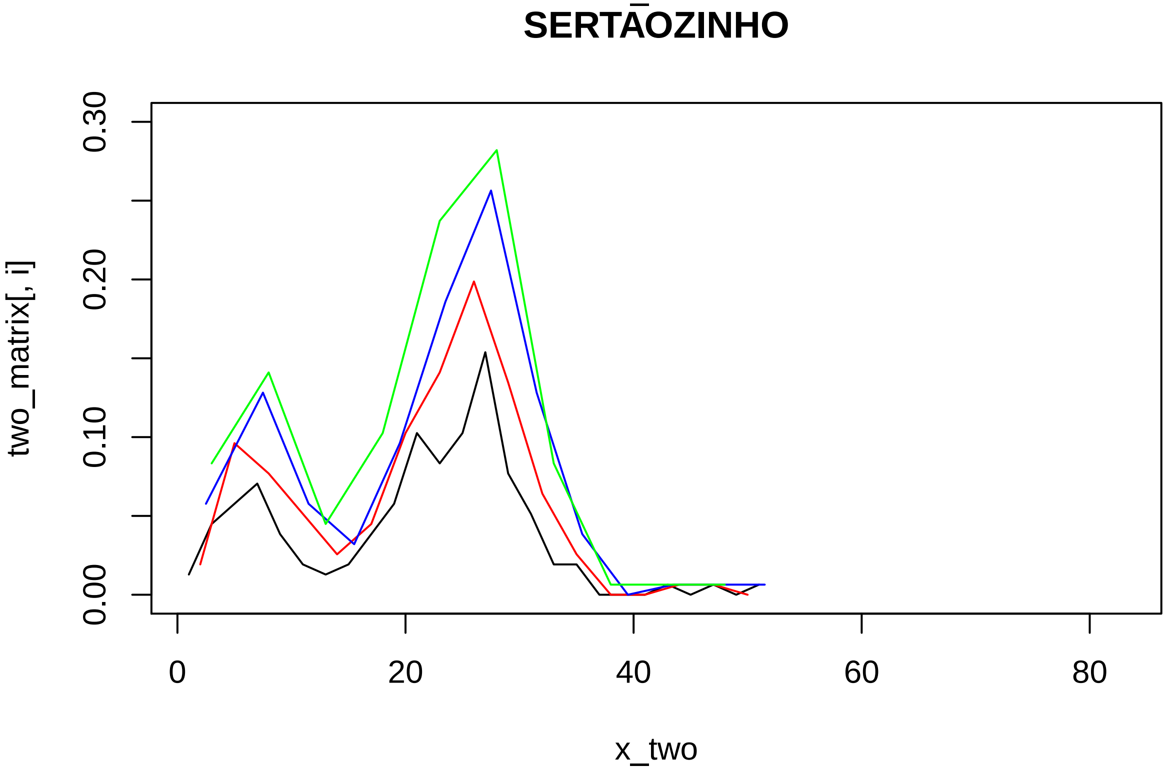

PAULINIA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

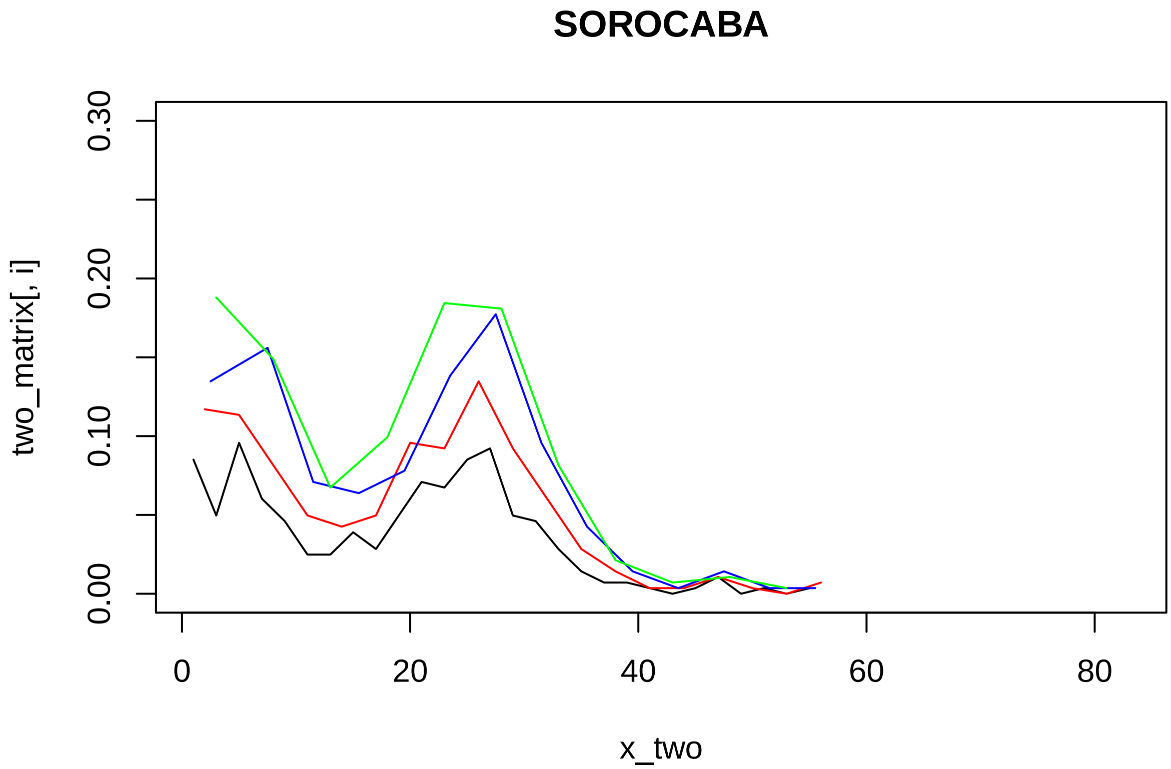

INDAIATUBA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

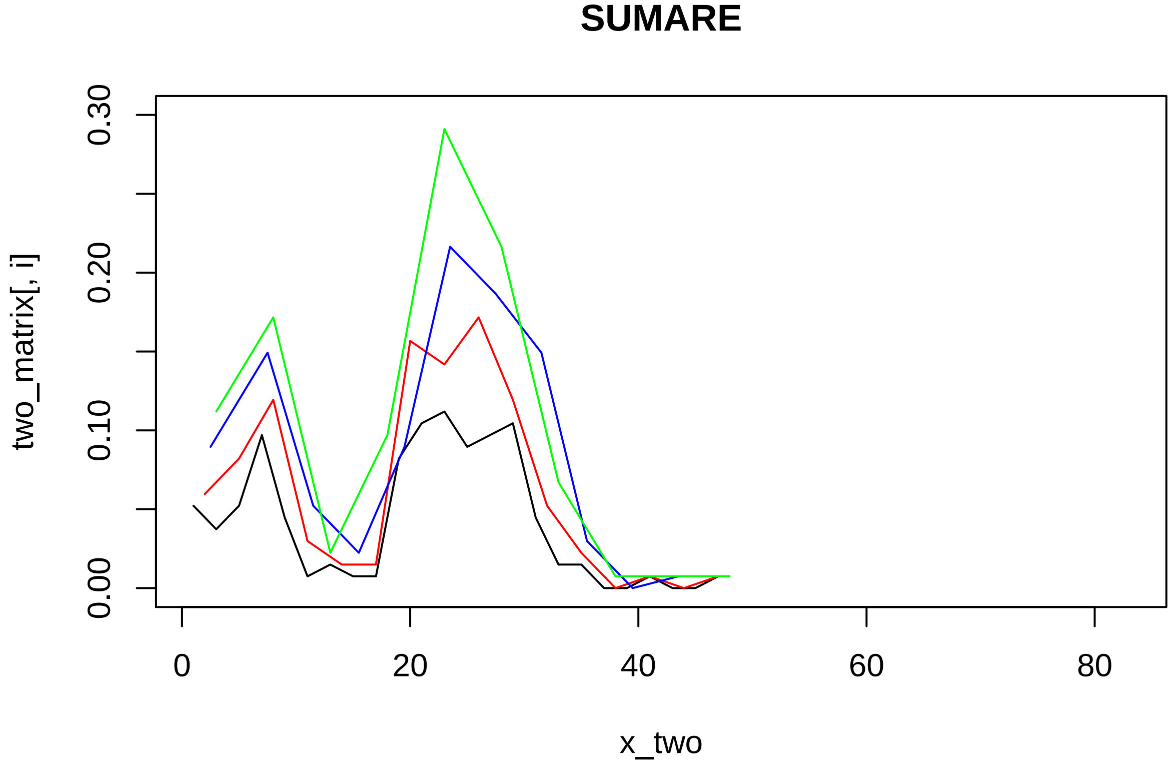

SALTO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

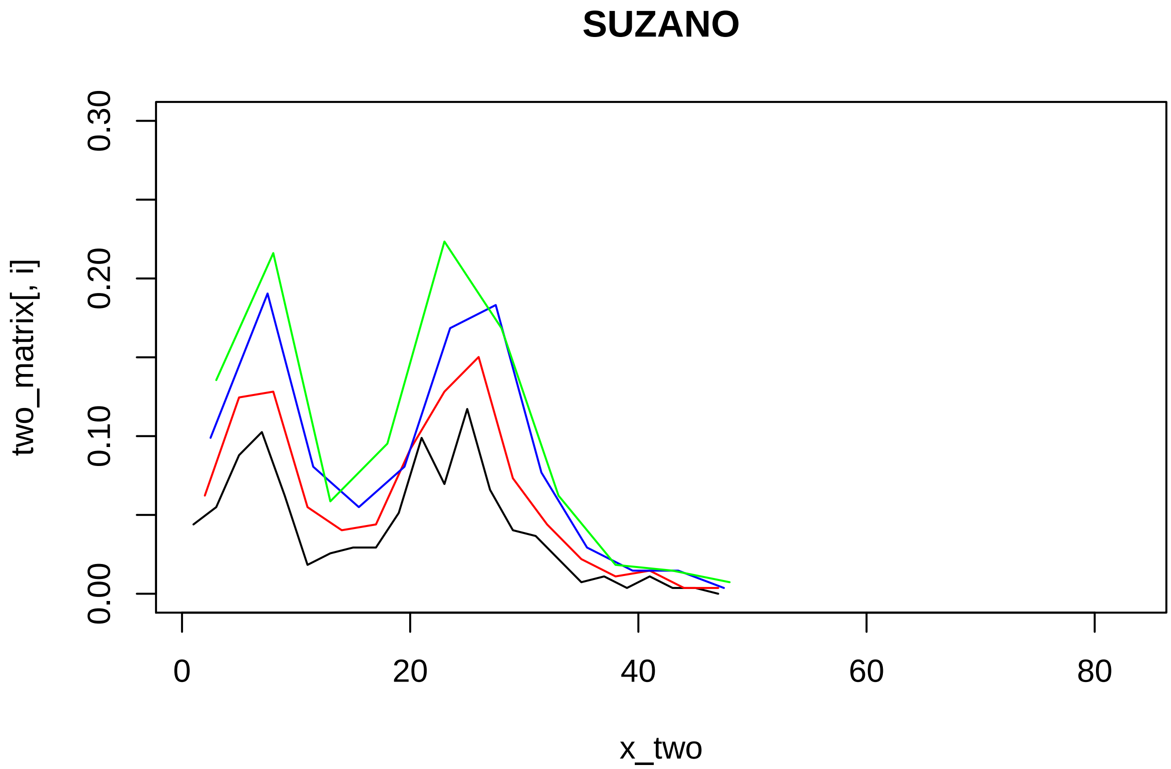

SAO SEBASTIAO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

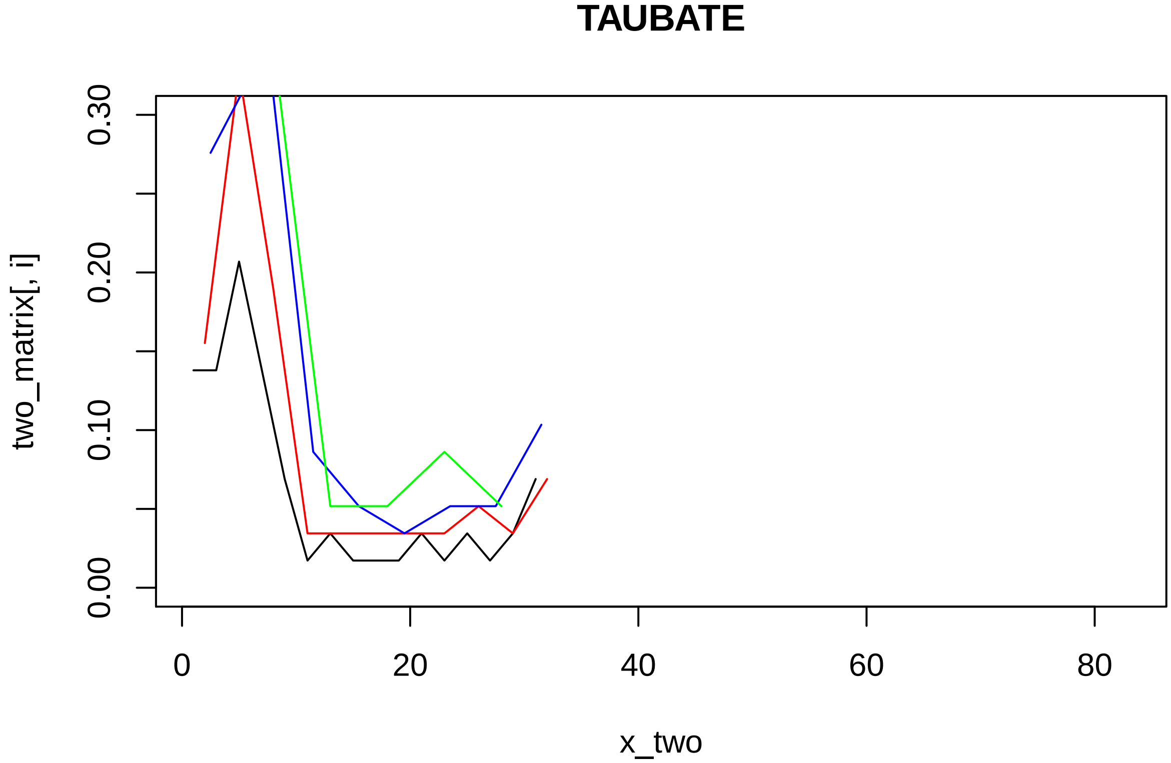

SAO VICENTE
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

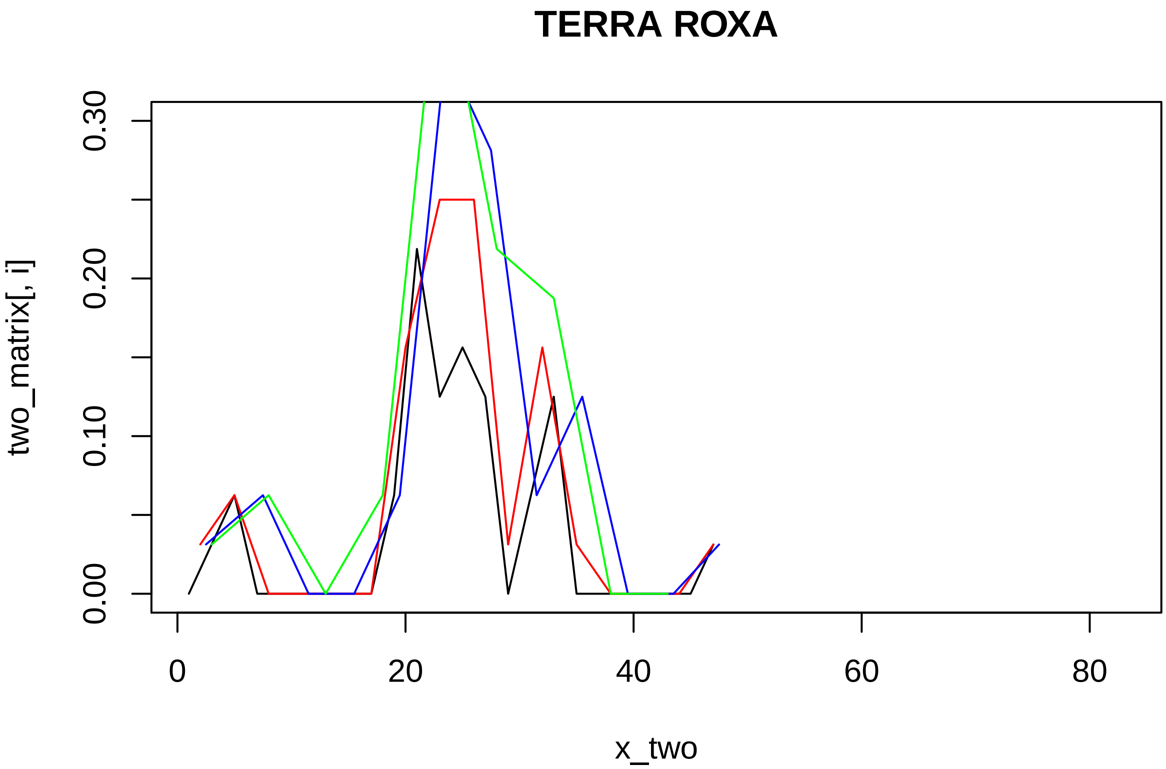

SERTAOZINHO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

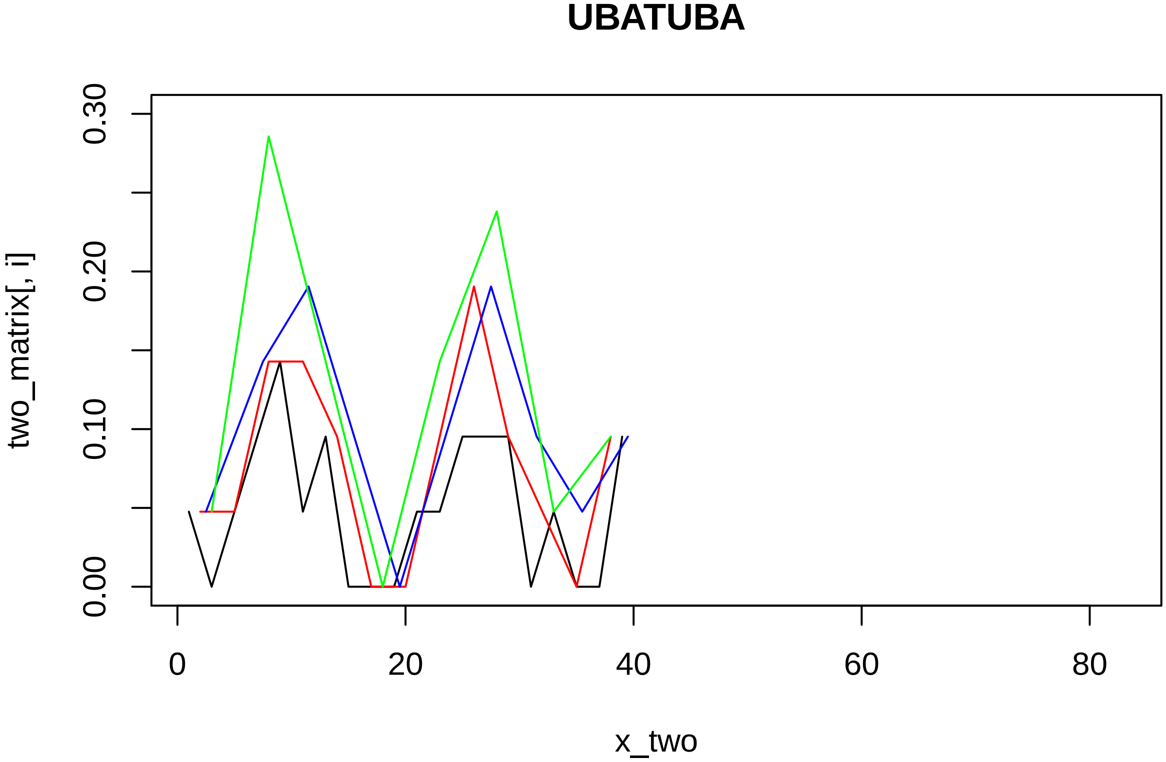

SOROCABA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

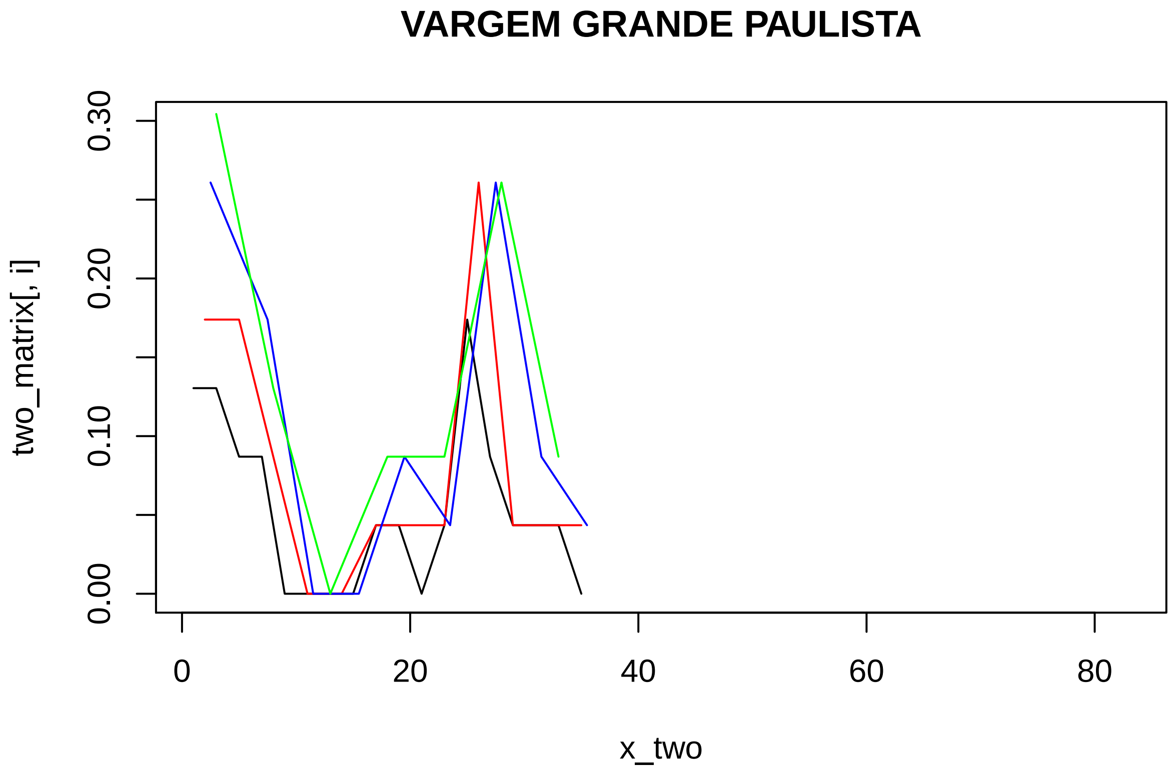

SUMARE
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

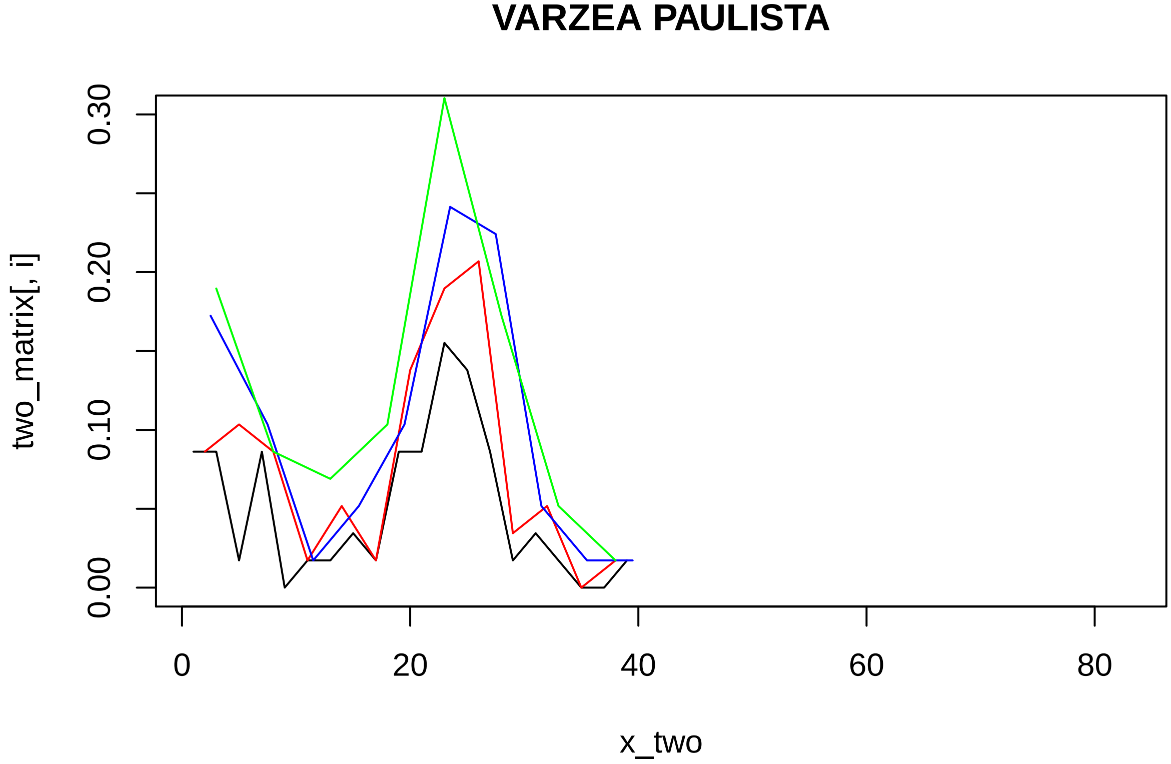

SUZANO
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two

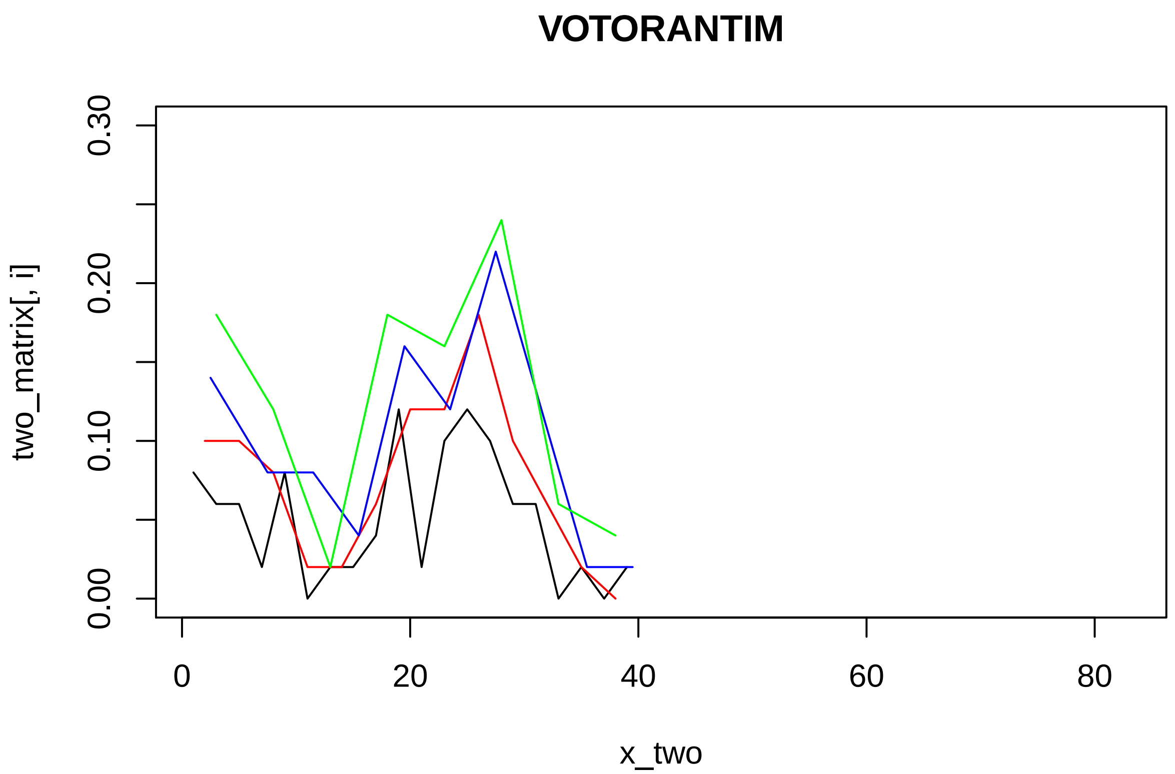

TAUBATE
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


TERRA ROXA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


UBATUBA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


VARGEM GRANDE PAULISTA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two


VARZEA PAULISTA
two_matrix[, i]
0.00
0.10
0.20
0.30
0
20
40
60
80
x_two



```
#two
mean_two <- rep(NA, 42)
sd_two <- rep(NA, 42)
for(i in 1:42){
  mean_two[i] <- rowMeans(two_matrix[i,], na.rm = TRUE)
  sd_two[i] <- apply(two_matrix, 1, sd, na.rm = TRUE)[i]
}

#three
mean_three <- rep(NA, 28)
sd_three <- rep(NA, 28)
for(i in 1:28){
  mean_three[i] <- rowMeans(three_matrix[i,], na.rm = TRUE)
  sd_three[i] <- apply(three_matrix, 1, sd, na.rm = TRUE)[i]
}

#four
mean_four <- rep(NA, 21)
sd_four <- rep(NA, 21)
for(i in 1:21){
  mean_four[i] <- rowMeans(four_matrix[i,], na.rm = TRUE)
  sd_four[i] <- apply(four_matrix, 1, sd, na.rm = TRUE)[i]
}

#five
mean_five <- rep(NA, 17)
sd_five <- rep(NA, 17)
for(i in 1:17){
  mean_five[i] <- rowMeans(five_matrix[i,], na.rm = TRUE)
  sd_five[i] <- apply(five_matrix, 1, sd, na.rm = TRUE)[i]
}
```

```
#RMSP = 1
urban <- subset(data, data$RMSP == "1")
urban_counties <- as.character(unique(urban$COUNTY))

which.urban <- rep(NA, length(urban_counties))
for(i in 1:length(urban_counties)){
 which.urban[i] <- which(county_vec[] == urban_counties[i])
}
two_urban_mat <- two_matrix[which.urban]
three_urban_mat <- three_matrix[which.urban]
four_urban_mat <- three_matrix[which.urban]
five_urban_mat <- five_matrix[which.urban]


#RMSP = 0
rural <- subset(data, data$RMSP == "0")
rural_counties <- as.character(unique(rural$COUNTY))

which.rural <- rep(NA, length(rural_counties))
for(i in 1:length(rural_counties)){
  which.rural[i] <- which(county_vec[] == rural_counties[i])
}
two_rural_mat <- two_matrix[which.rural]
two_rural <- matrix(rep(NA, 3 * 42 * length(rural_counties)), ncol = 3)
colnames(two_rural) <- c("COUNTY", "CLASS", "DENSITY")
two_rural <- as.data.frame(two_rural)
#fill counties
two_rural[,1] <- rep(rural_counties, 42)

#fill age class
two_rural[seq(1:length(rural_counties)),2] <- x_two[1]
for(i in 1:length(x_two)){
  x <- length(rural_counties) * i + seq(1:length(rural_counties))
  two_rural[x,2] <- x_two[i+1]
}

#fill density
for(i in 1:length(rural_counties)){
  two_rural[i,3] <- two_rural_mat[1,i]
}
for(i in 1:length(rural_counties)){
  for(j in 2:length(x_two)){
    two_rural[(i + 348 * (j - 1) ),3]  <- two_rural_mat[j,i]
  }
  }
two_rural$CLASS <- as.factor(two_rural$CLASS)
two_rural$COUNTY <- as.factor(two_rural$COUNTY)


three_rural_mat <- three_matrix[which.rural]
four_rural_mat <- three_matrix[which.rural]
```

```
five_rural_mat <- five_matrix[which.rural]
```

Preliminary analysis

Created a method (via for loops) to make a matrix which is compatible with ANOVAs.

Up next:

- Look for correlation or association between denstiy and class on both a rural and urban level.
  - This may be best to do with the five_rural_mat and five_urban_mat because they have fewer factor levels. Starting with urban may also be more promising because it will have fewer factors for county.
- Look into whether the moving windows/data smoothing can be used for the windows of age groups. Try to understand how this works.
- Plot row means with row sd error bars for the different windows.