Identification of Spark Cases

Maggie Walters

June 18, 2017

Document Synopsis

This document aims to identify patterns/trends in the spark cases within each county. The protocol will be as follows:

- Create a matrix (**index_week_mat**) including the first 100 cases within each county as well as the age of that specific case, and the week and county in which it occurred.
- Identify those which are in the same district.
- Look for any similarities in age patterns at the beginning of an outbreak.

Creating index_week_mat

index_week_mat will be created by pulling out the index of the first 100 cases in county i. The county name and week in which the case occured will also be recorded.

Preliminary Analysis:

Some cases do not have more than 100 cases, so they have relatively small case loads. Could change to doing counties with more than 20 cases, which would amount to approximately 80 counties. Again, I am coming across the question of **what is the minimum number of cases necessary for a disease outbreak to be considered an epidemic?**

Examine ages of first 100 cases and timing

AGE was added as a variable to the index_week_mat in order to examine age distribution patterns.

```
#fill in age
AGE <- rep(NA, 38700)
index_week_mat <- cbind(AGE, index_week_mat)


for(i in 1:38700){
  x <- index_week_mat[i,2]
  x <- as.numeric(as.character(x))
  index_week_mat[i,1] <- data_ordered$AGE[x]
}
```

## Ordering the cases in ascending order of week of onset

```
index_week_mat <- index_week_mat[order(index_week_mat$WEEK),]
```
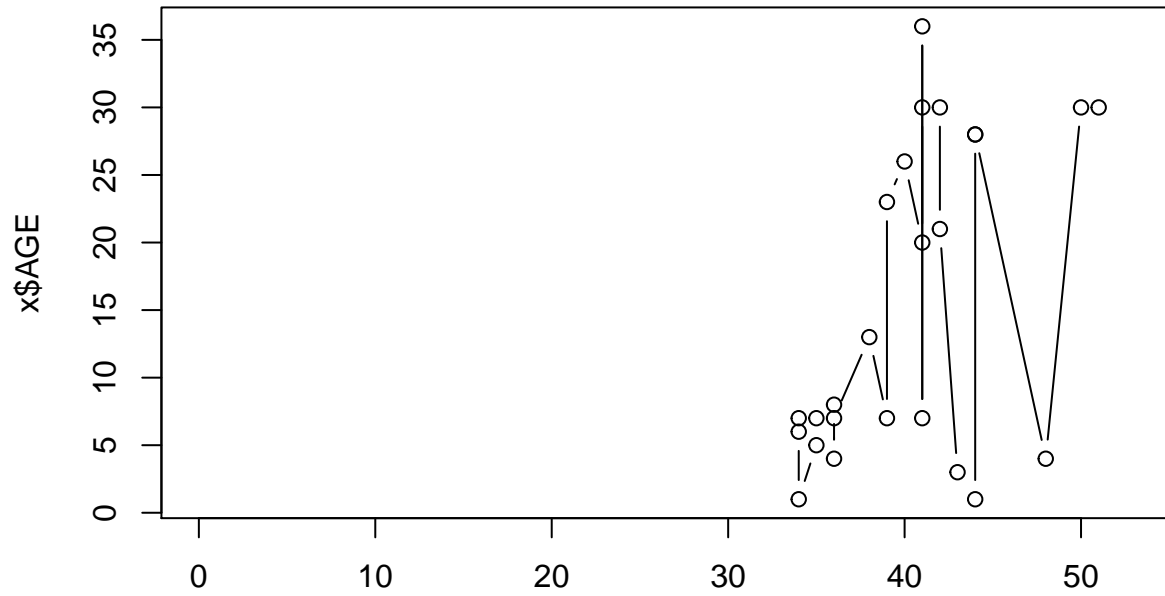
Plot age by week for each county

```r
#need to remove rows 6759 to 38700, just going to make a new matrix called the same thing
which.na <- which(is.na(index_week_mat$WEEK))

index_week_mat <- subset(index_week_mat, !is.na(index_week_mat$NUM) & !is.na(index_week_mat$AGE)
                         & !is.na(index_week_mat$COUNTY) & !is.na(index_week_mat$WEEK))
index_week_mat$WEEK <- as.numeric(index_week_mat$WEEK)
index_week_mat$AGE <- as.numeric(index_week_mat$AGE)
for(i in 1:length(county_vec)){
  x <- subset(index_week_mat, index_week_mat$COUNTY == county_vec[i])
  x$AGE <- as.numeric(as.character(x$AGE))
  plot(x$WEEK, x$AGE, type = "b", main = county_vec[i], xlim=c(0,53))
}
```

SAO PAULO
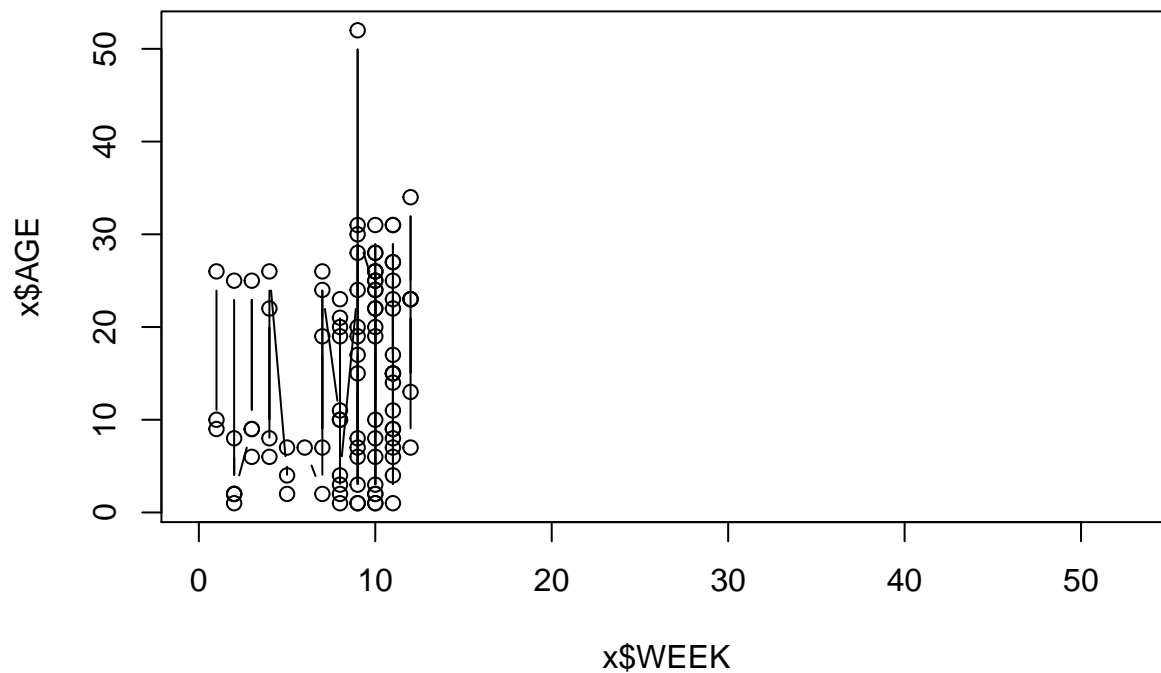
ARTUR NOGUEIRA

JAGUARIUNA
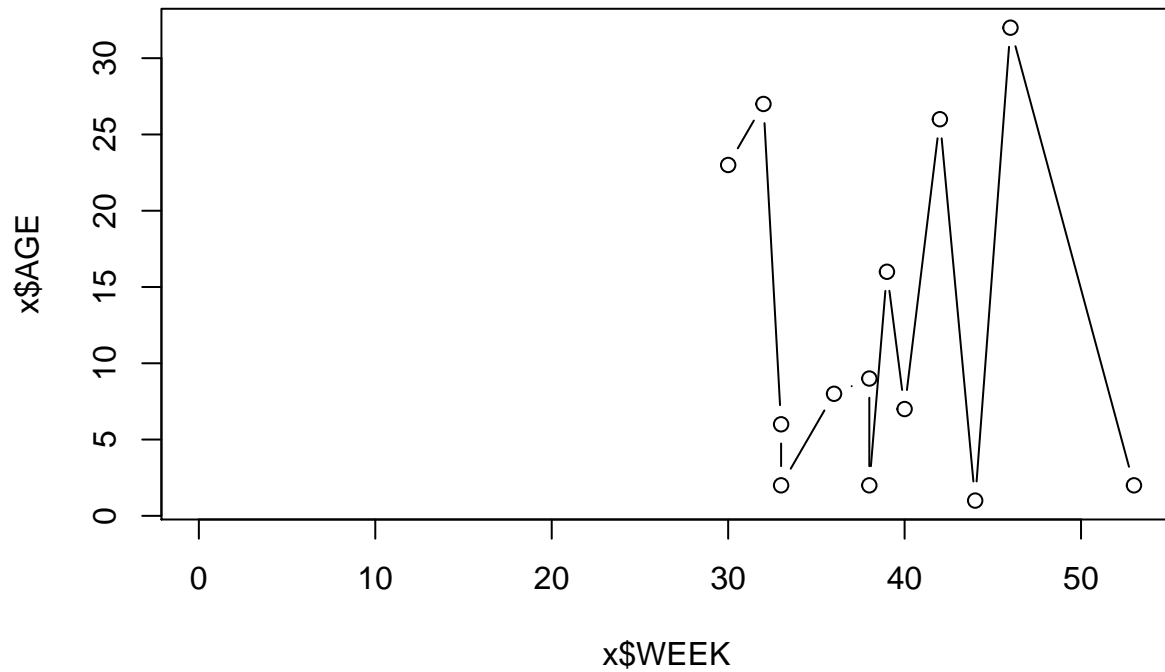
LINS

BOTUCATU

MOGI GUACU

x$WEEK
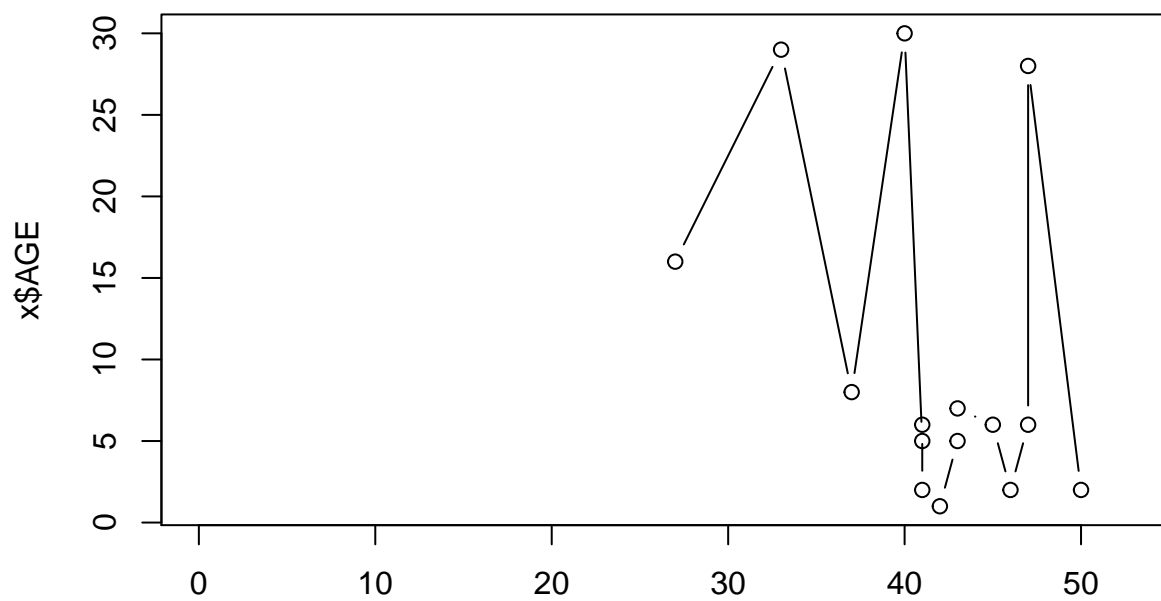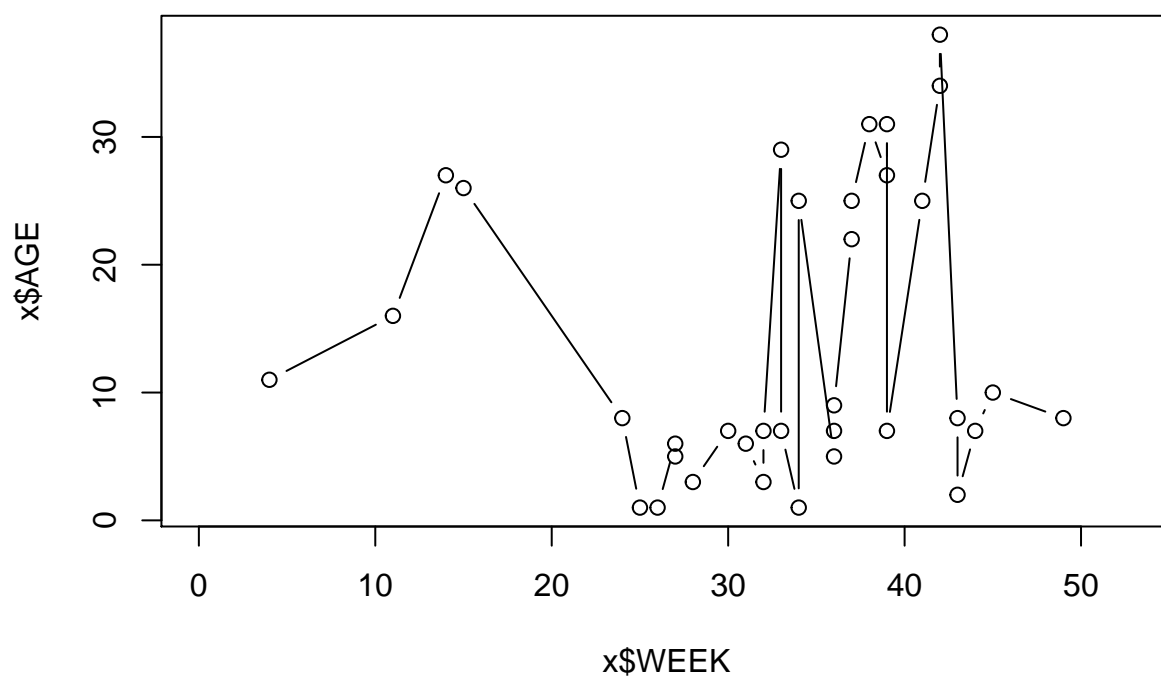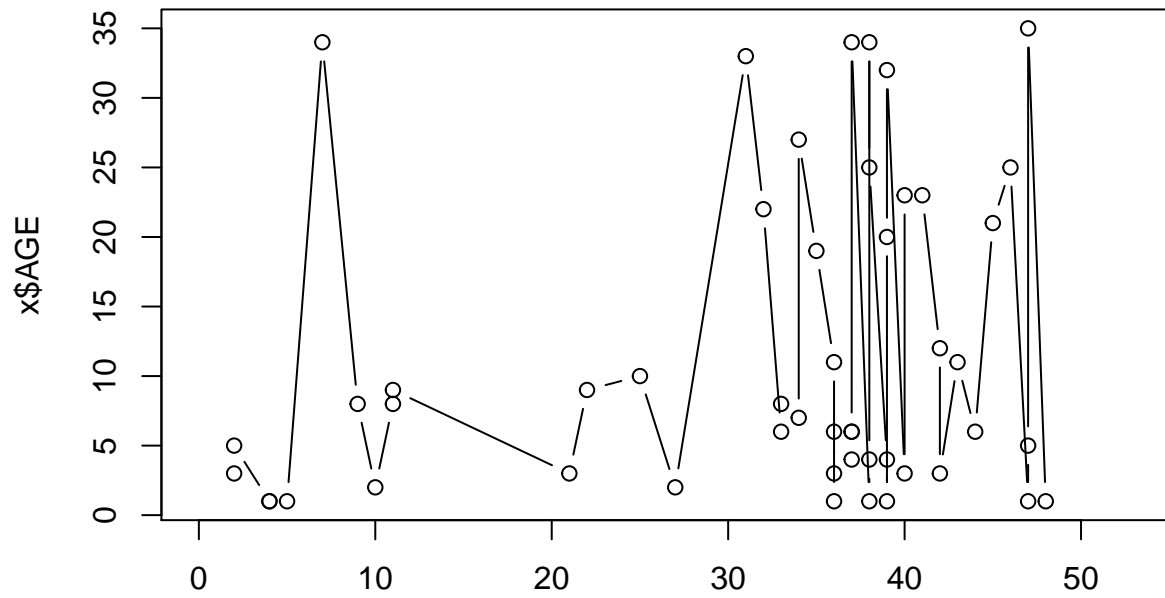
ENGENHEIRO COELHO

x$WEEK

BAURU

x$WEEK

SAO CAETANO DO SUL

x$WEEK

6
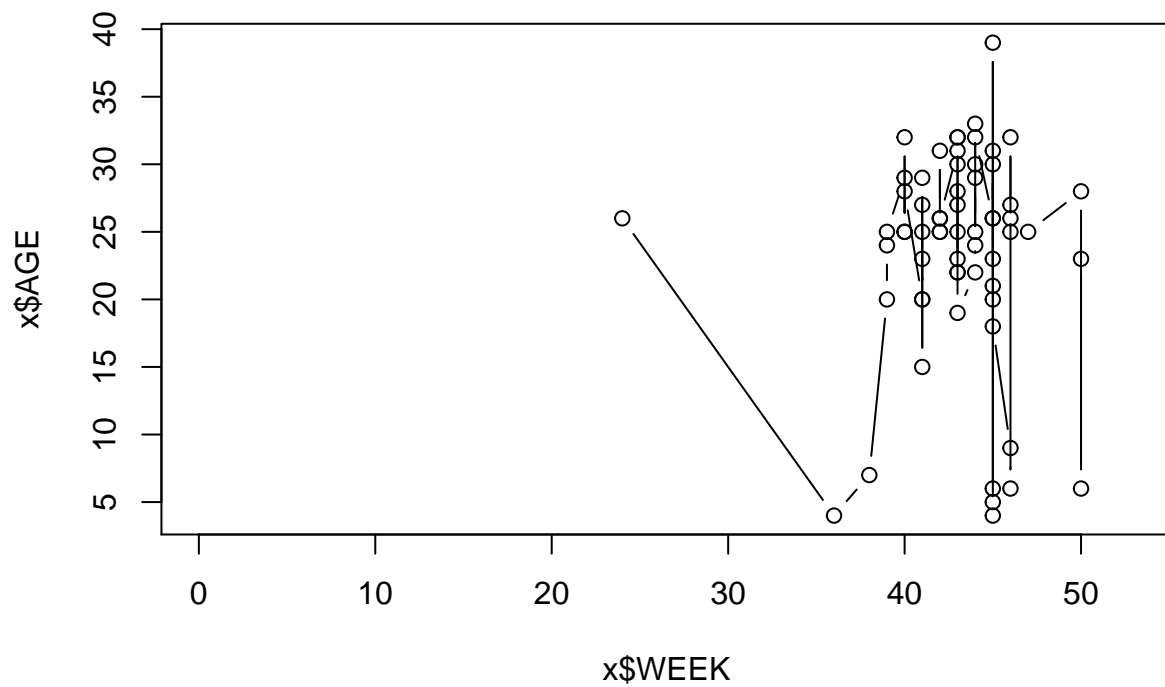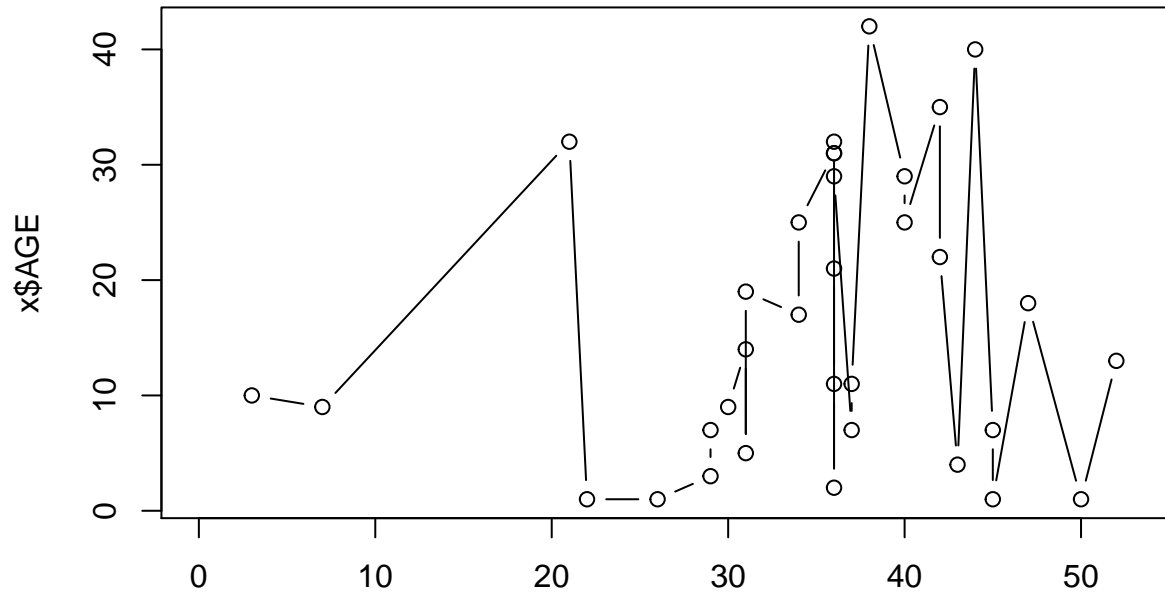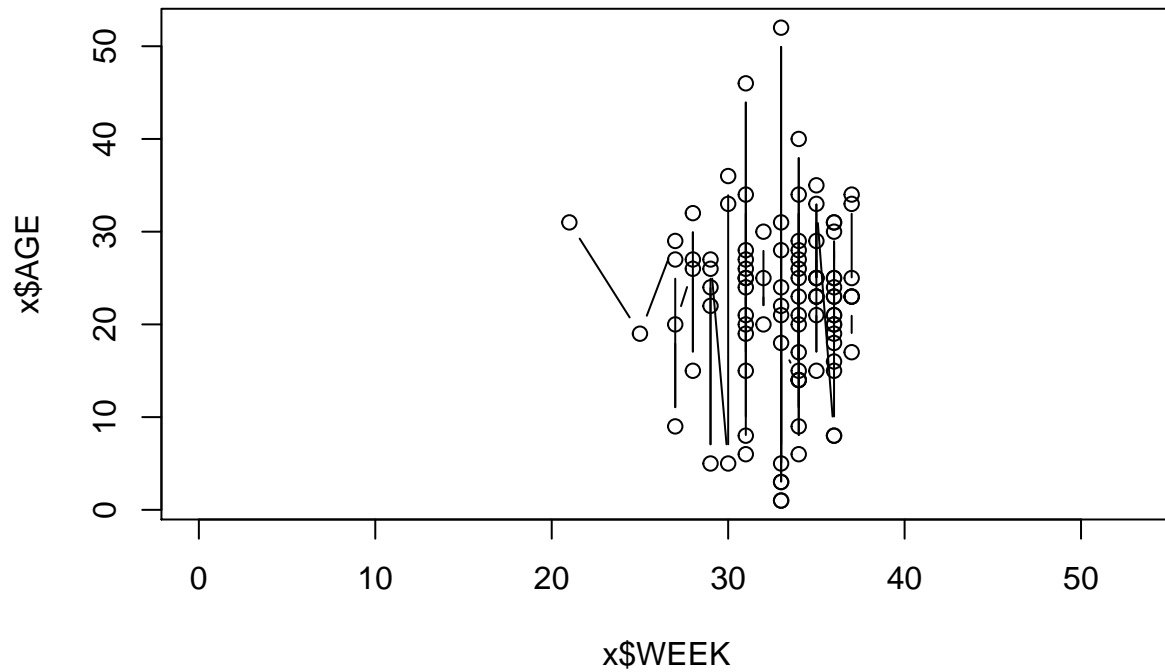
SAO CARLOS
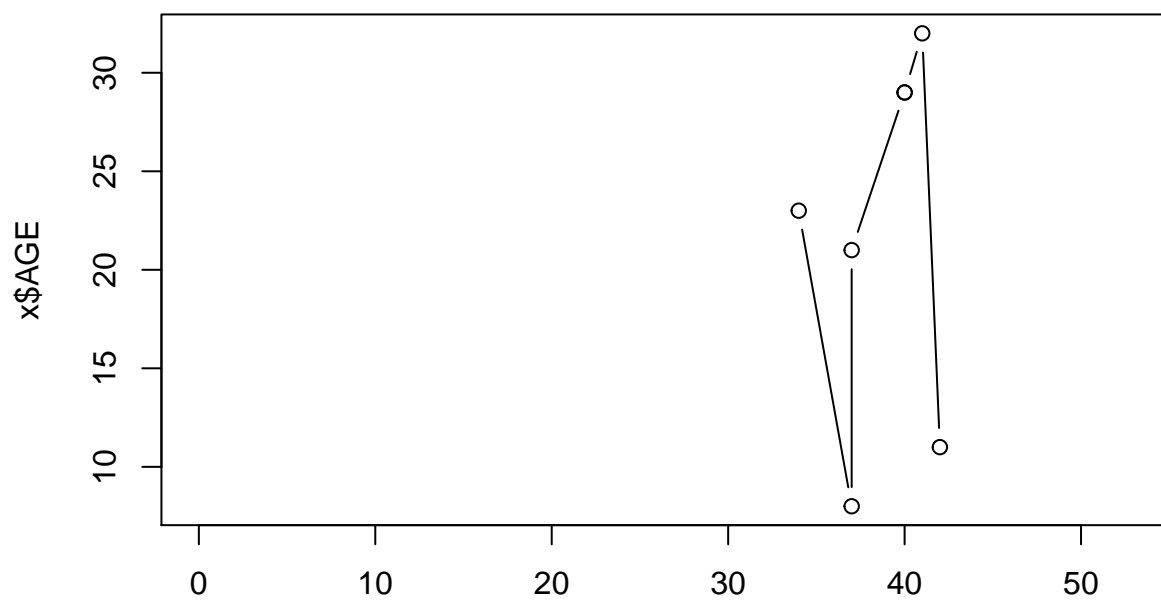
HOLAMBRA

CRUZEIRO

x$WEEK

MORRO AGUDO

x$WEEK

OSASCO

x$WEEK

DIADEMA

x$WEEK
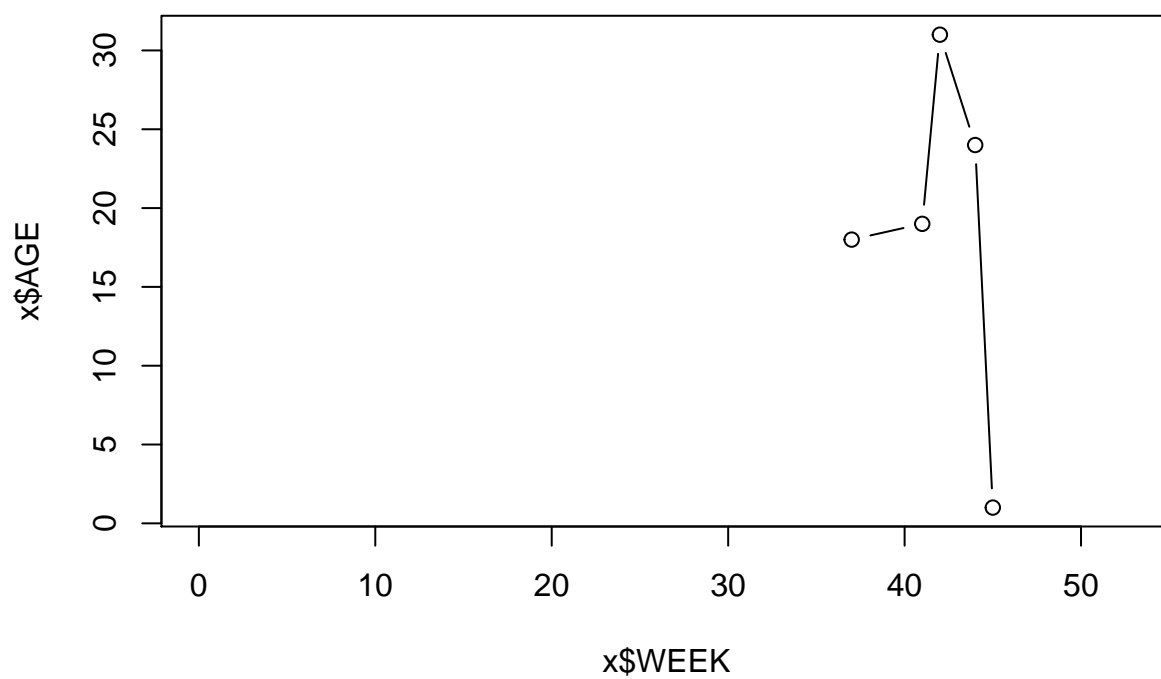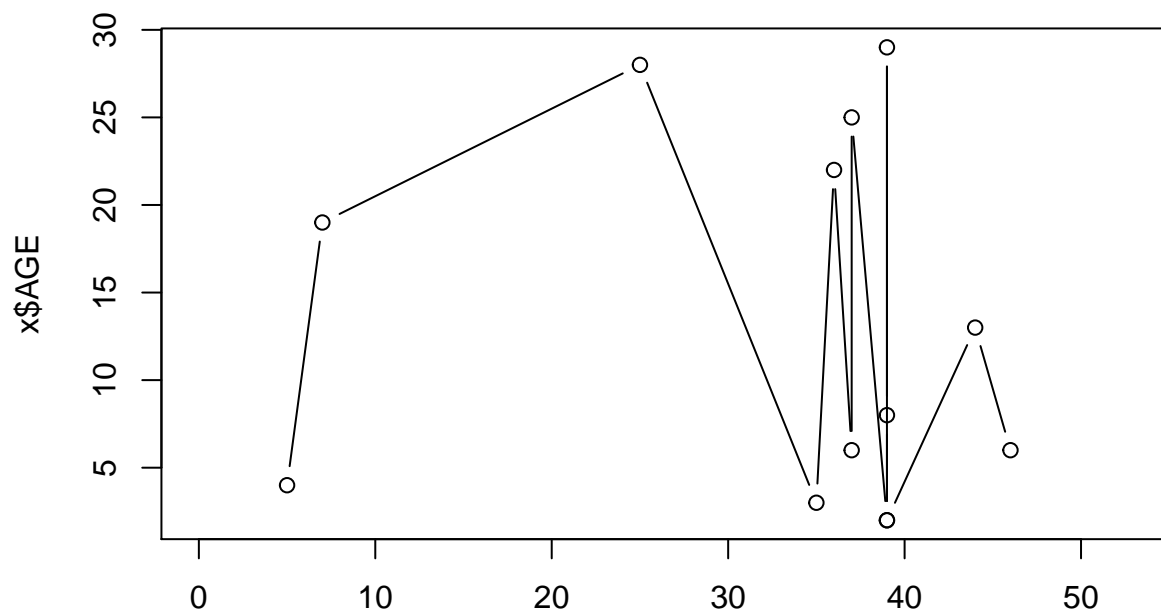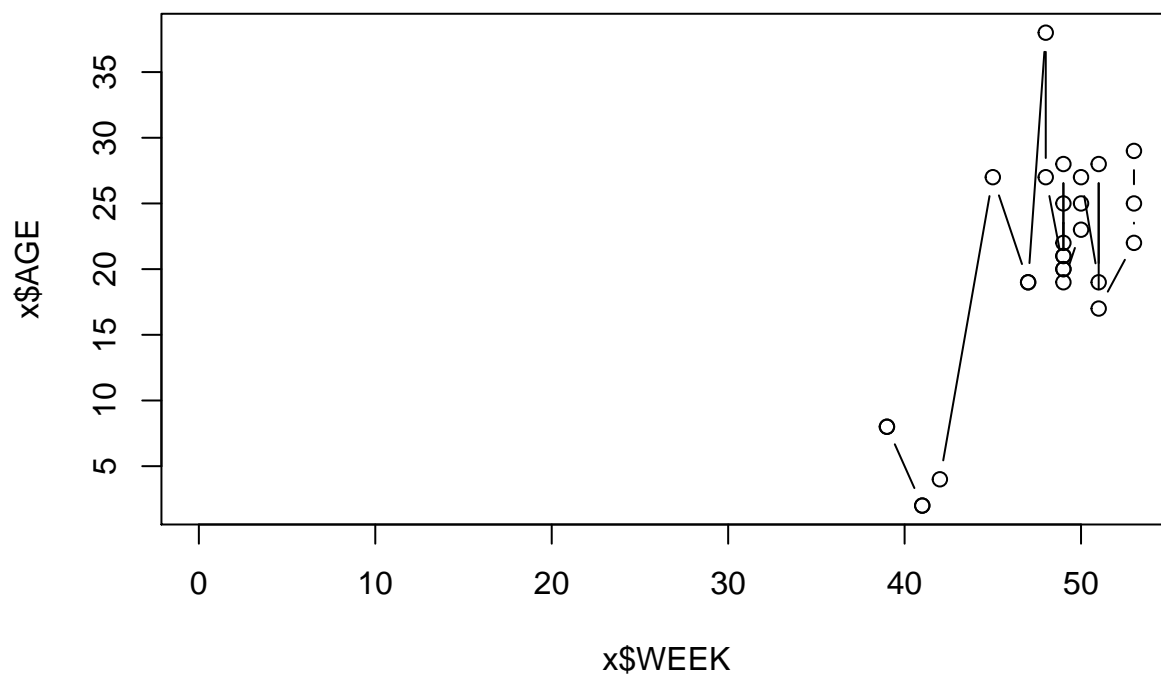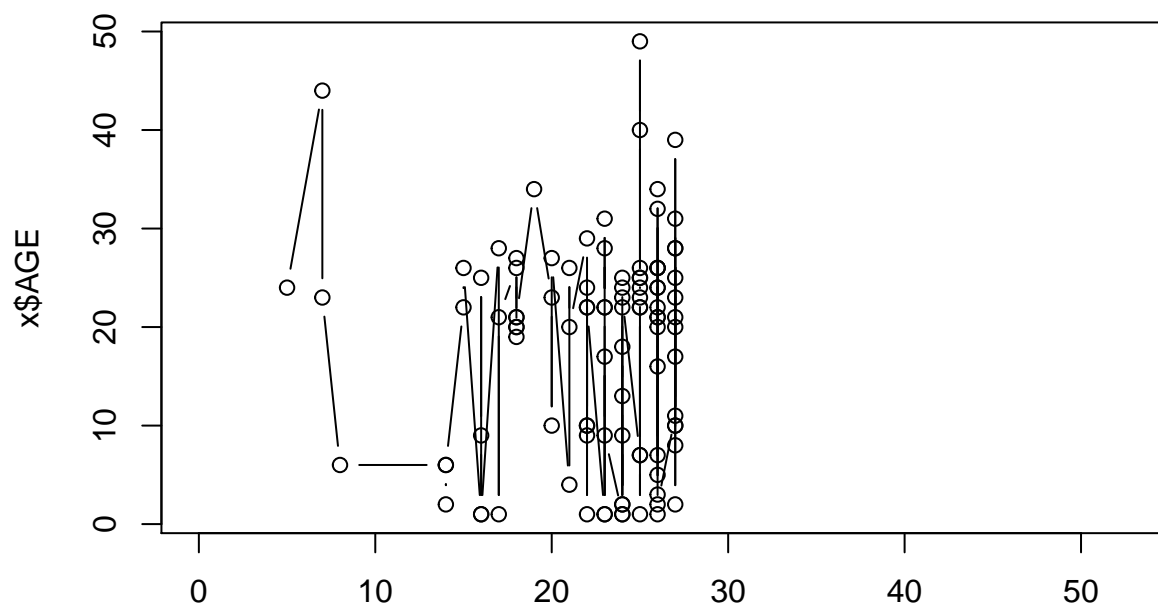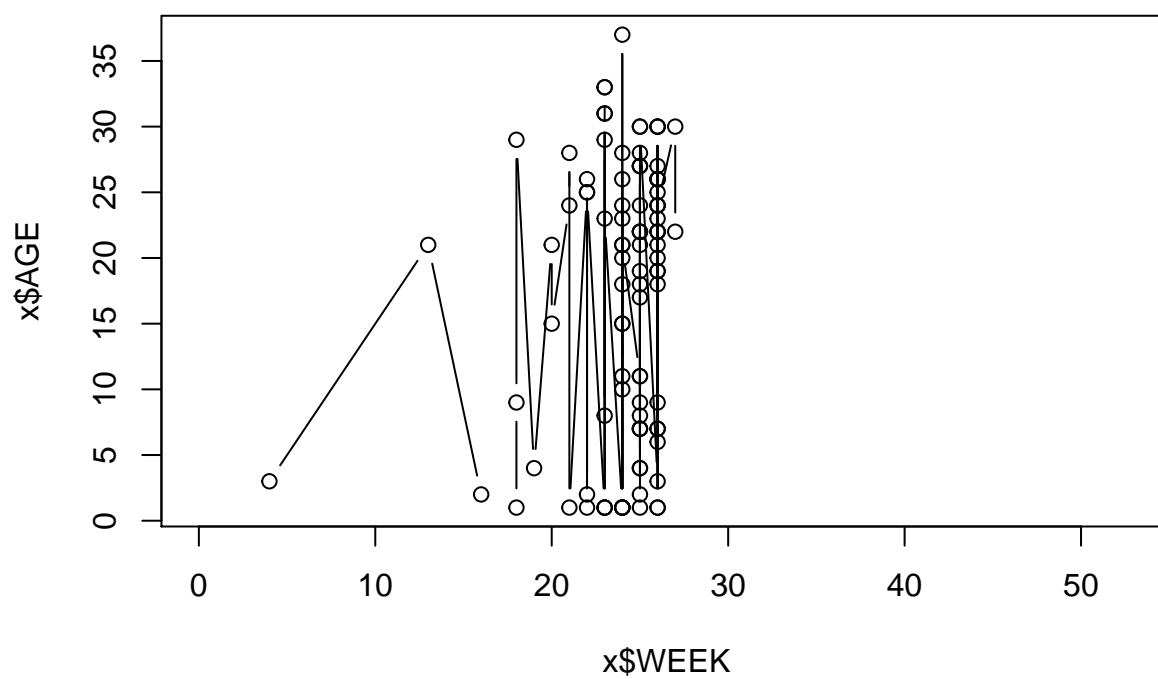
POA
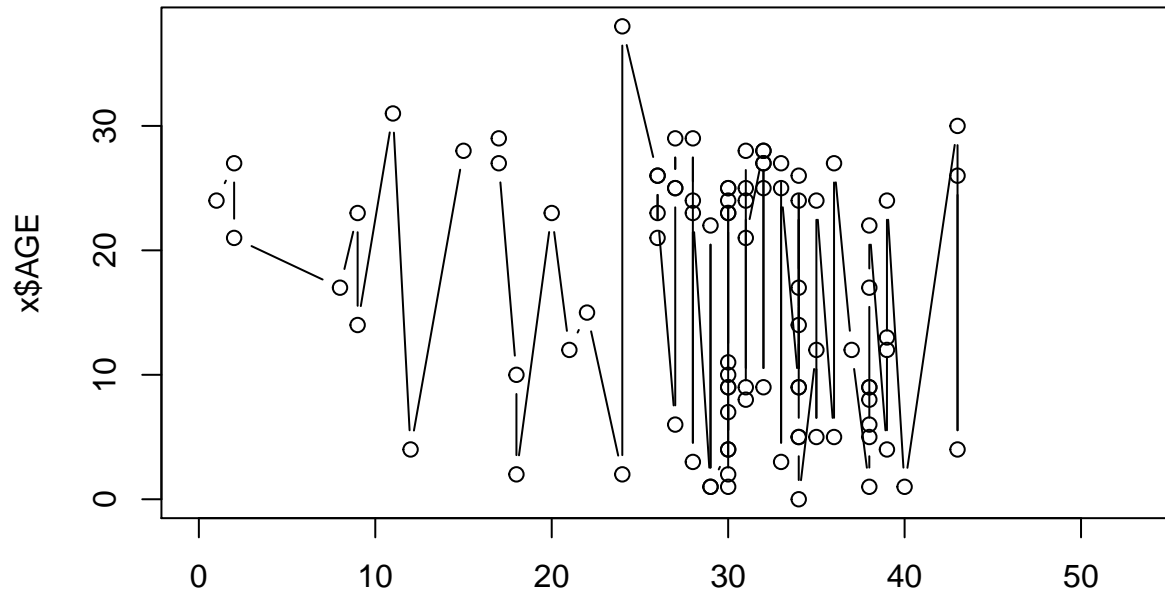
SAO BERNARDO DO CAMPO

GUARULHOS

x$WEEK

AGUAI

x$WEEK
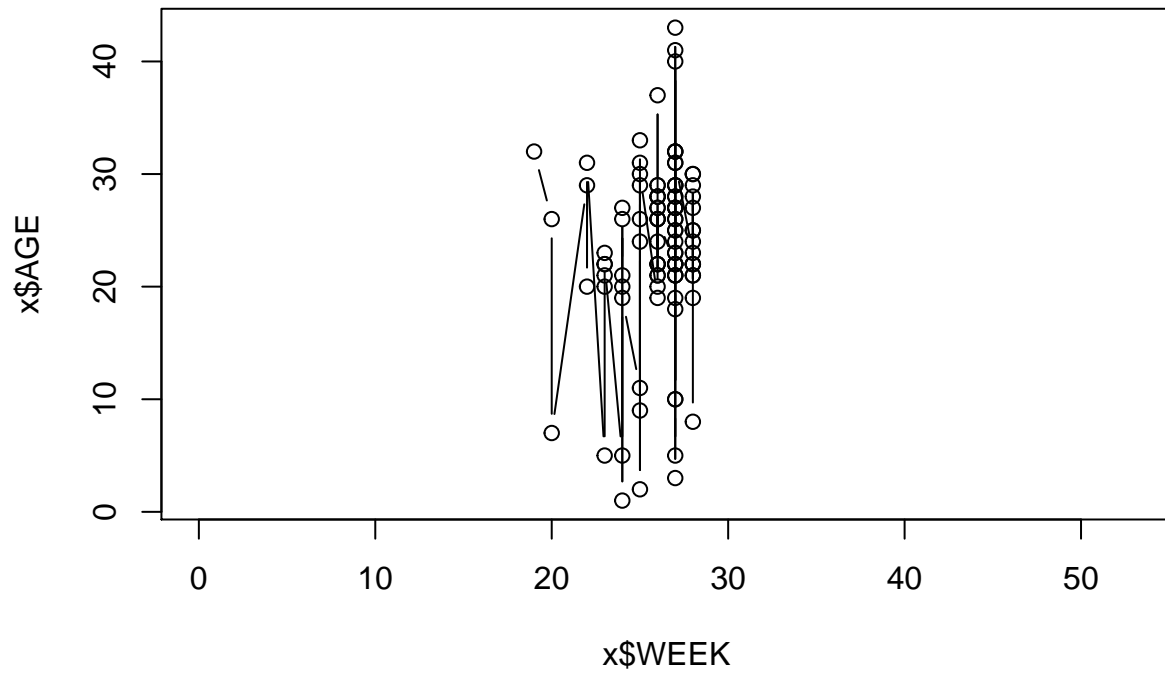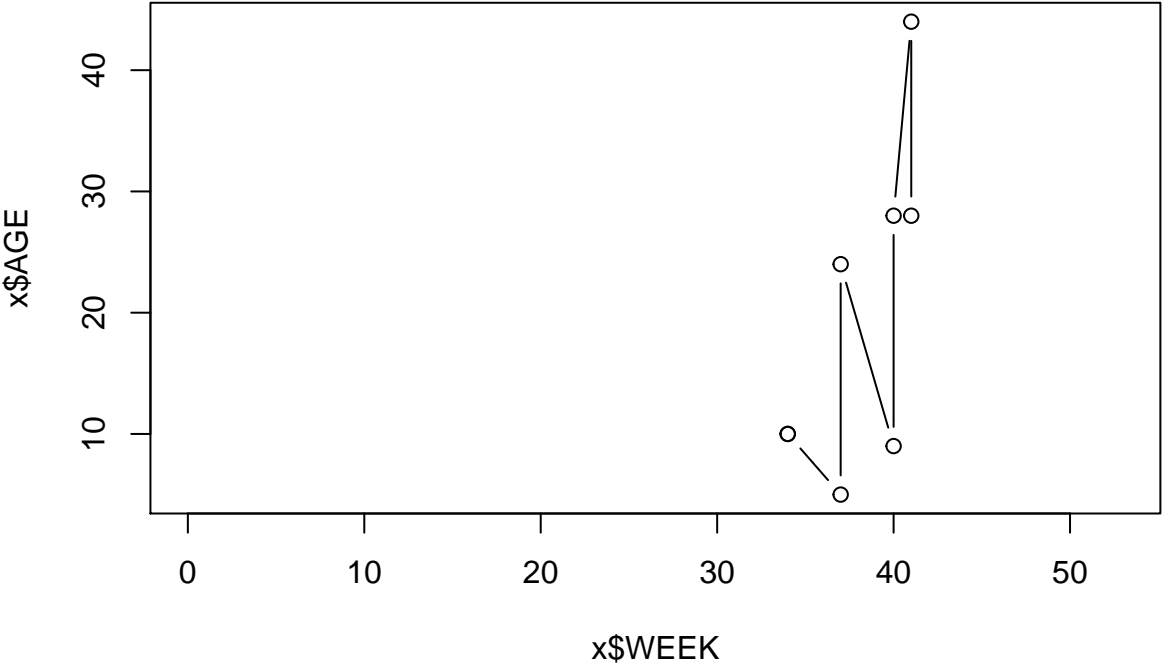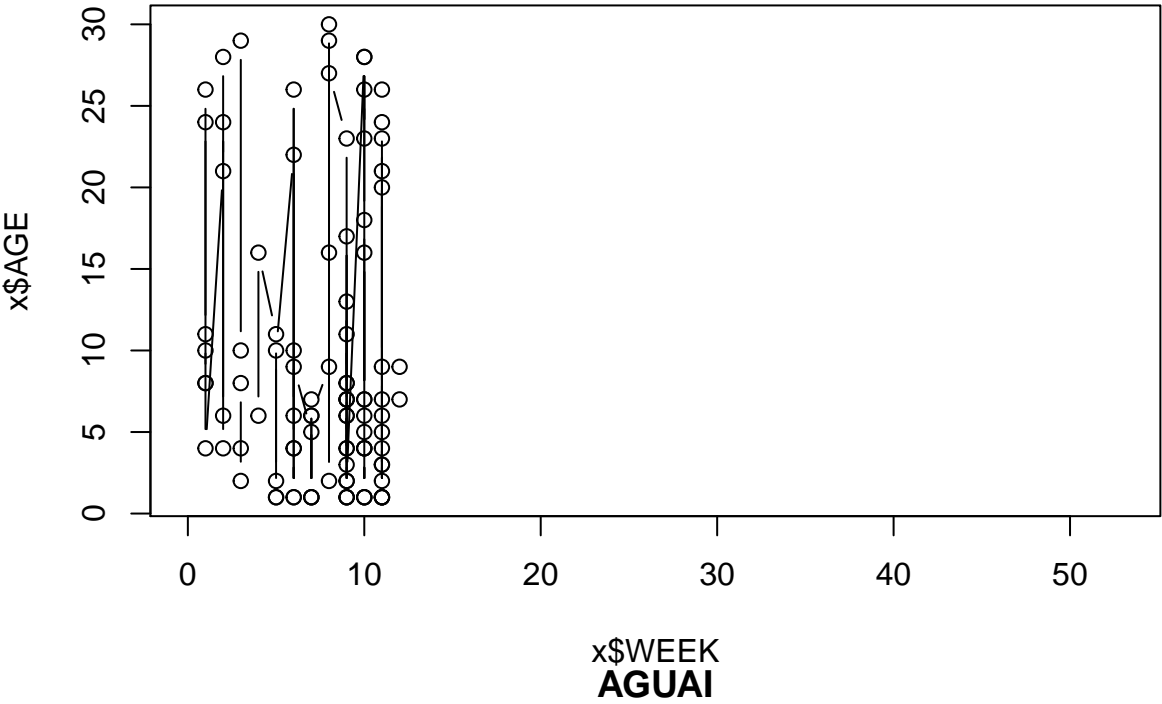
CAJAMAR
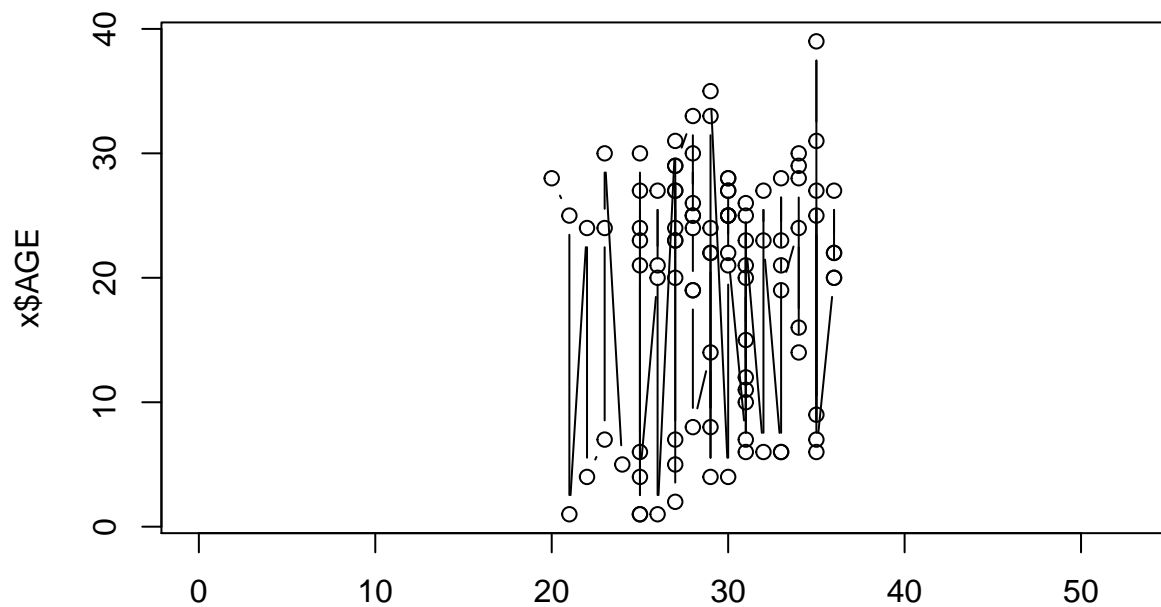
SANTO ANDRE
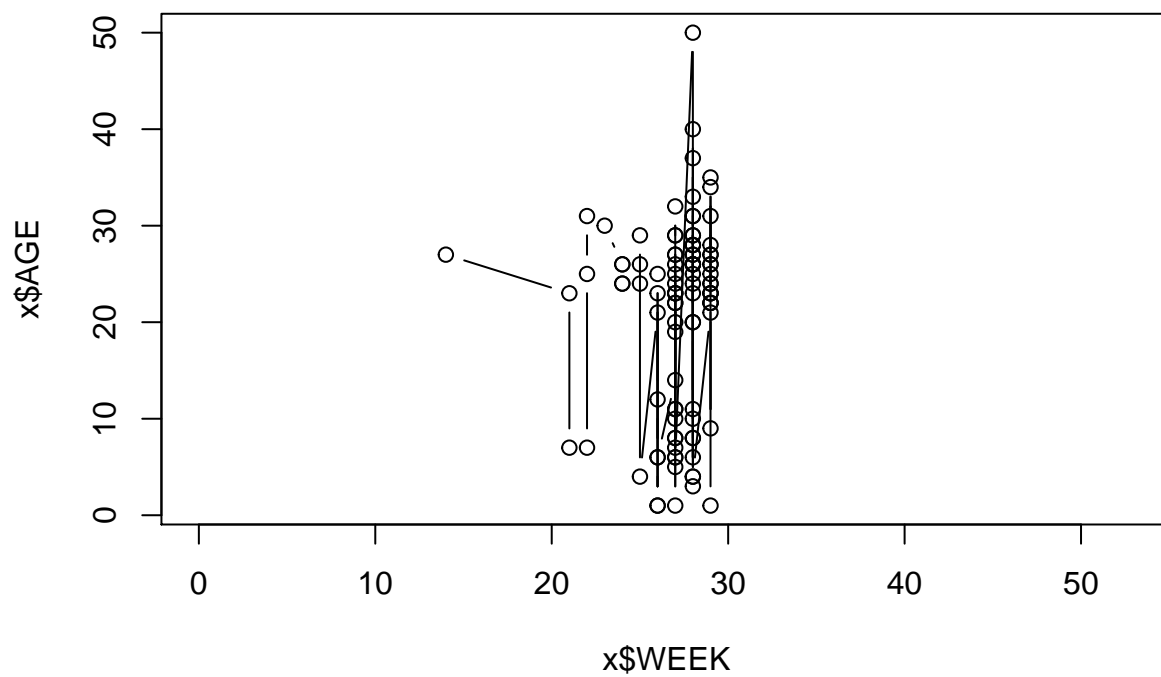
FRANCA
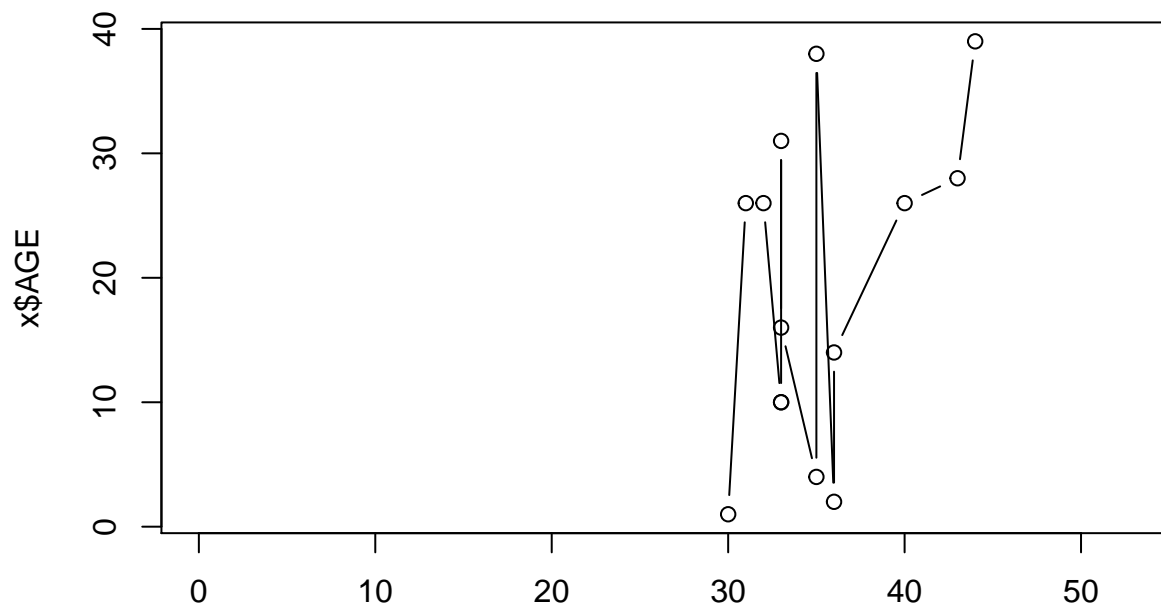
CASA BRANCA
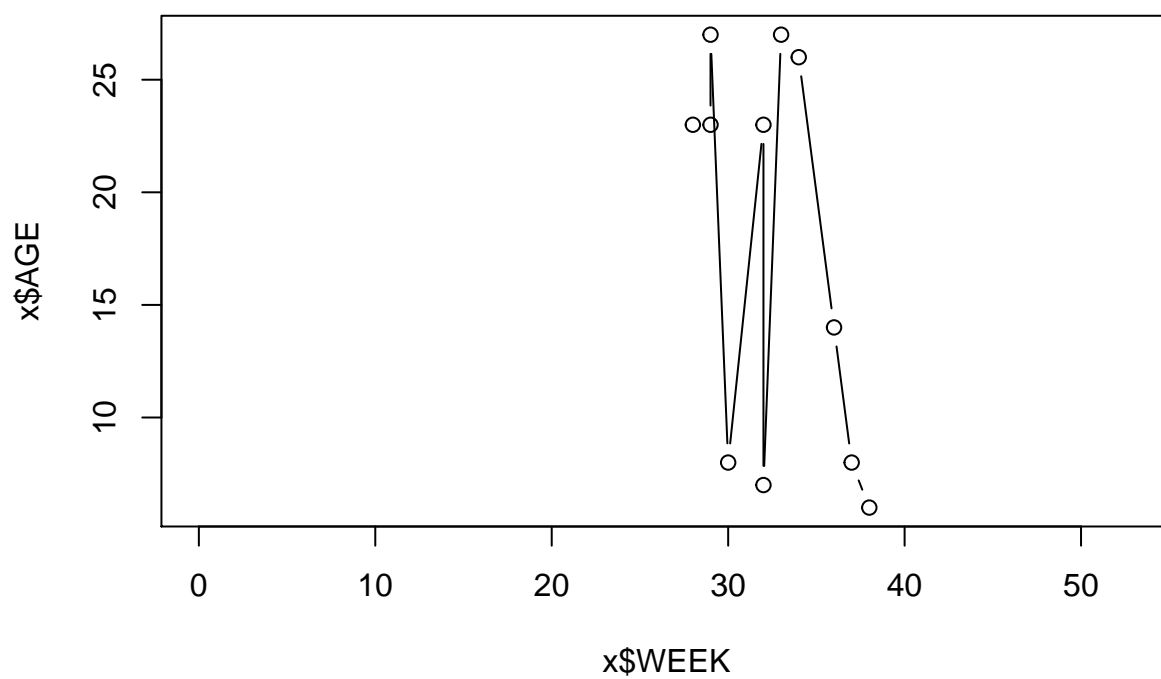
PRESIDENTE PRUDENTE

MARILIA
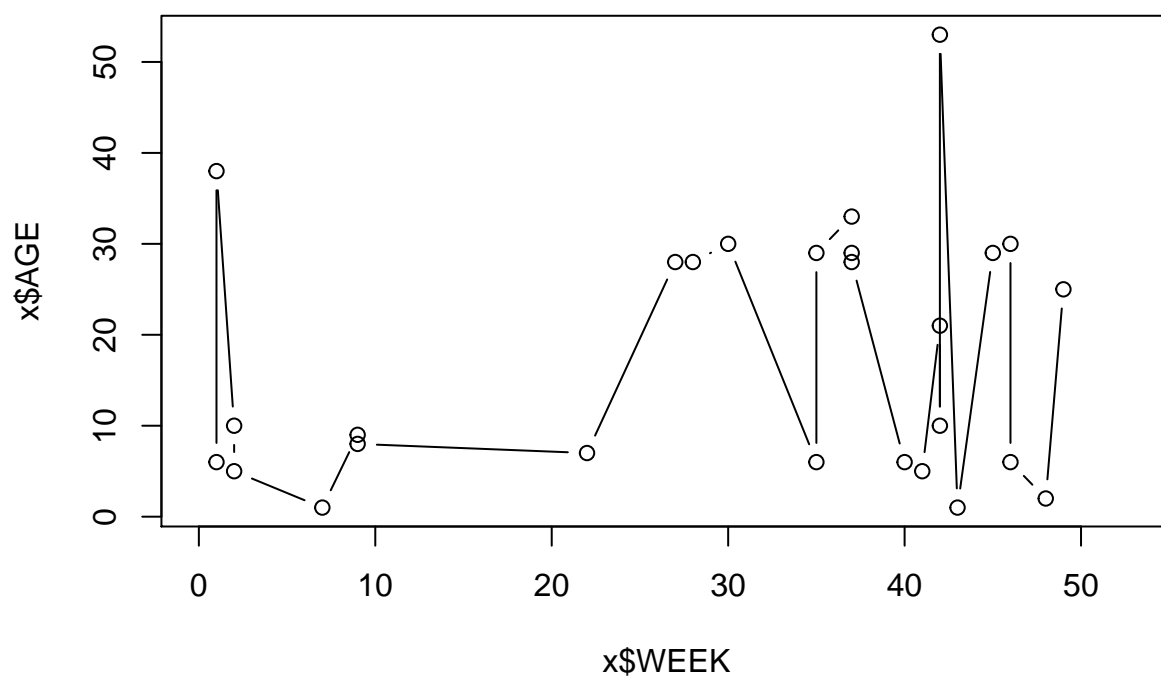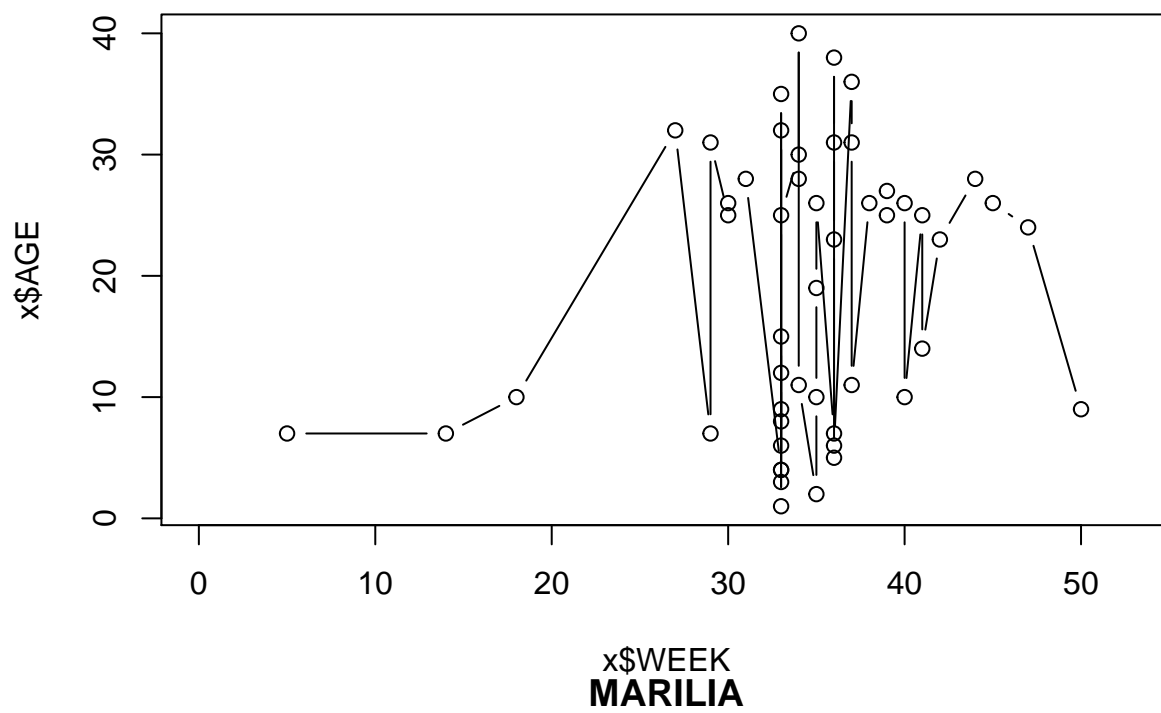
ALFREDO MARCONDES
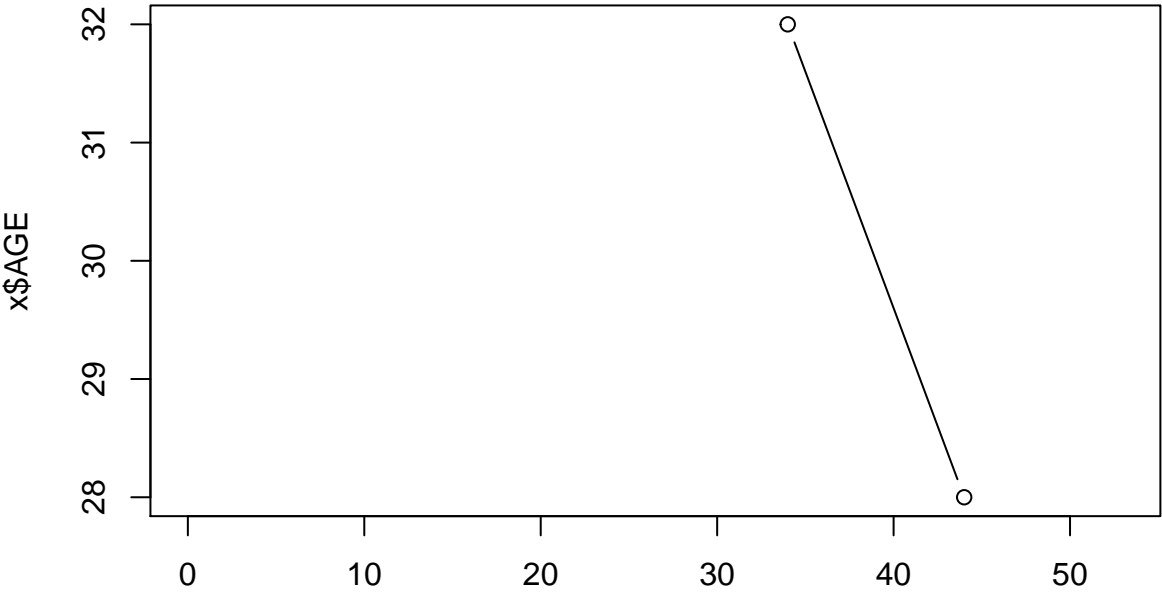
CRISTAIS PAULISTA
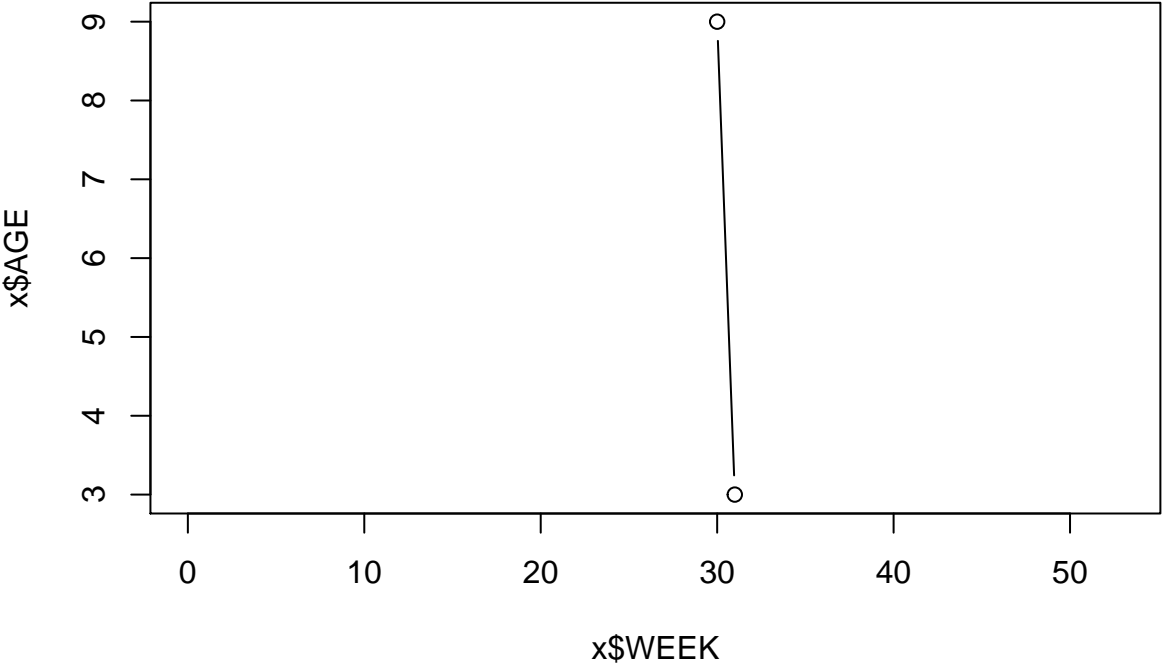
FERRAZ DE VASCONCELOS
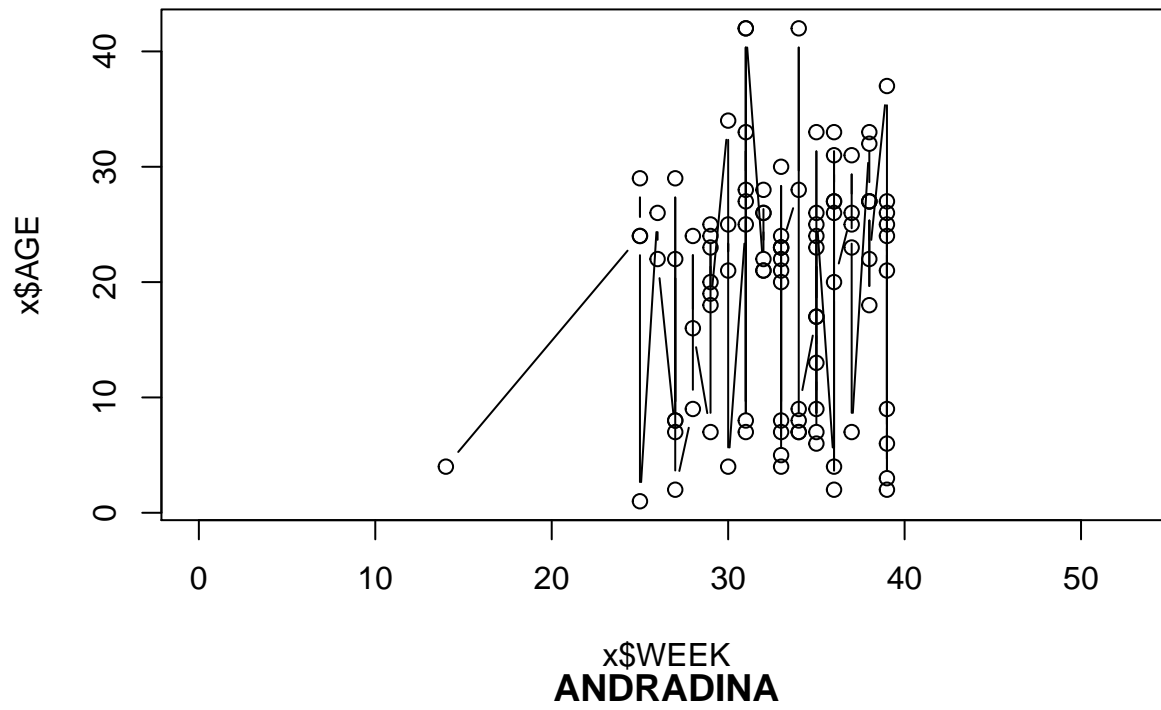
ANDRADINA
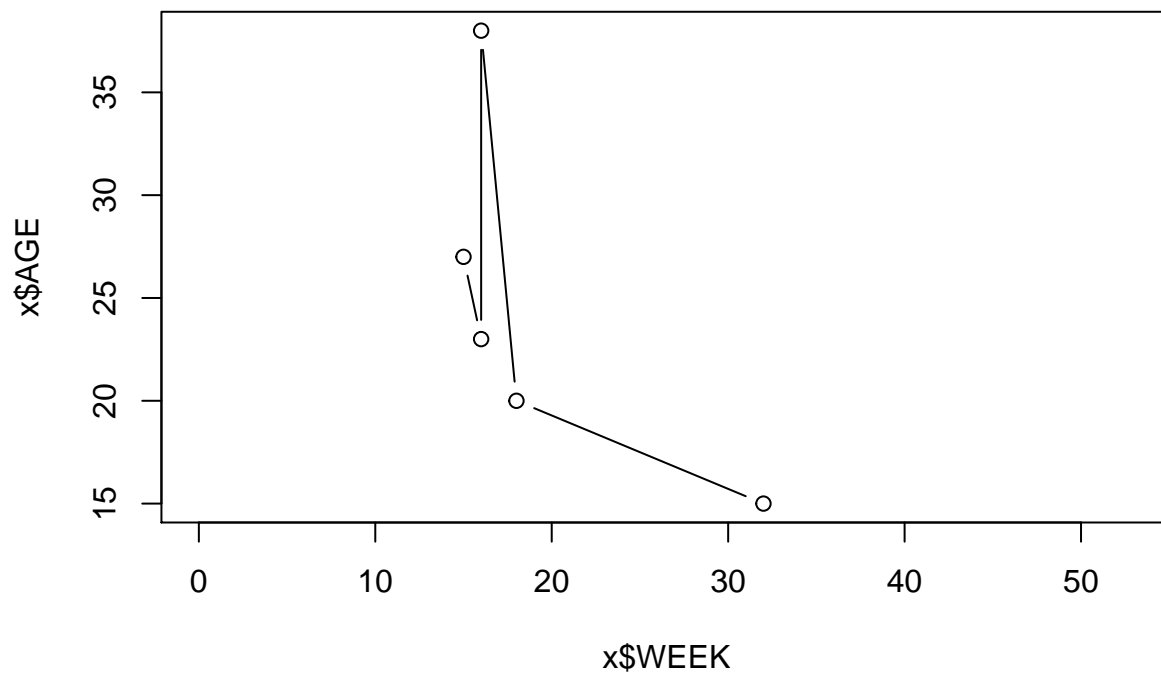
PRAIA GRANDE
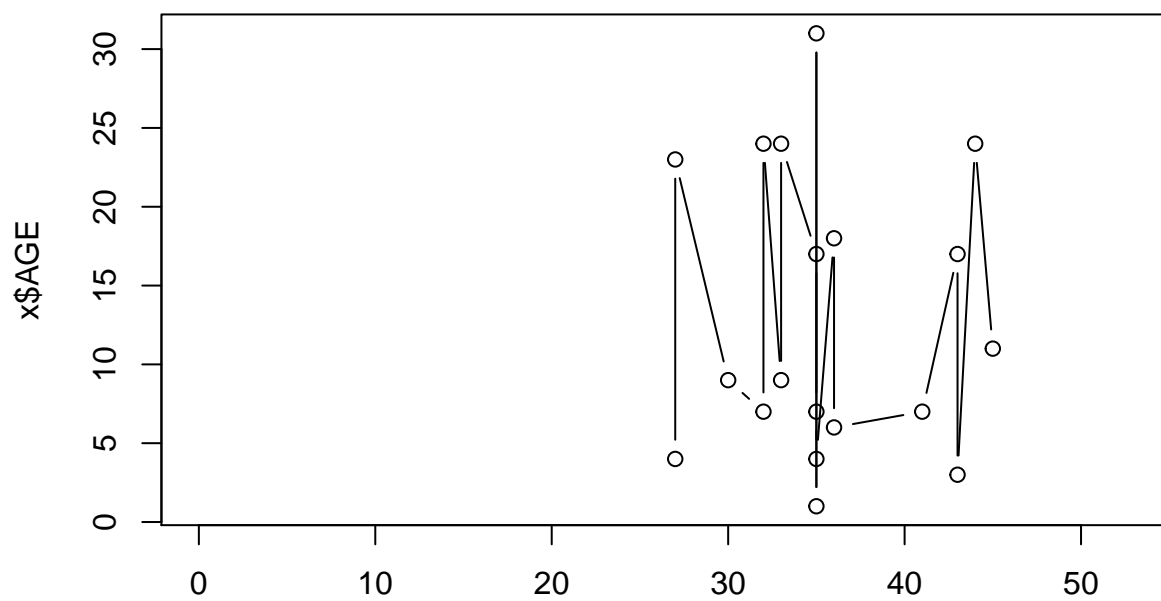
SABINO
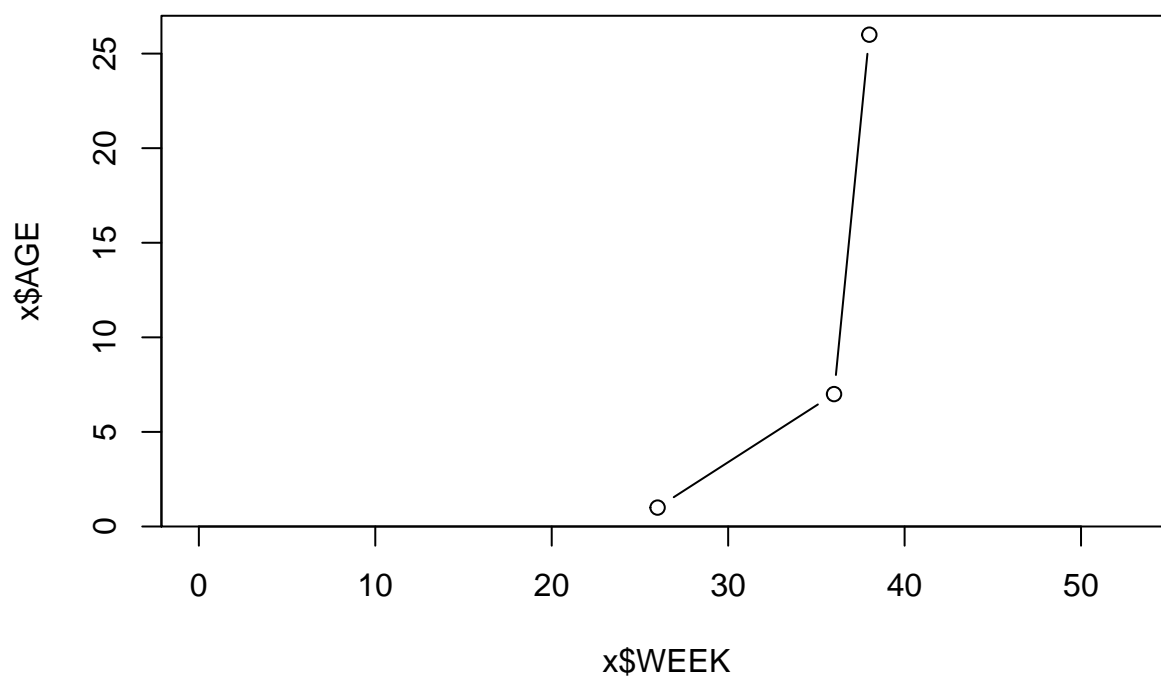
GUARUJA
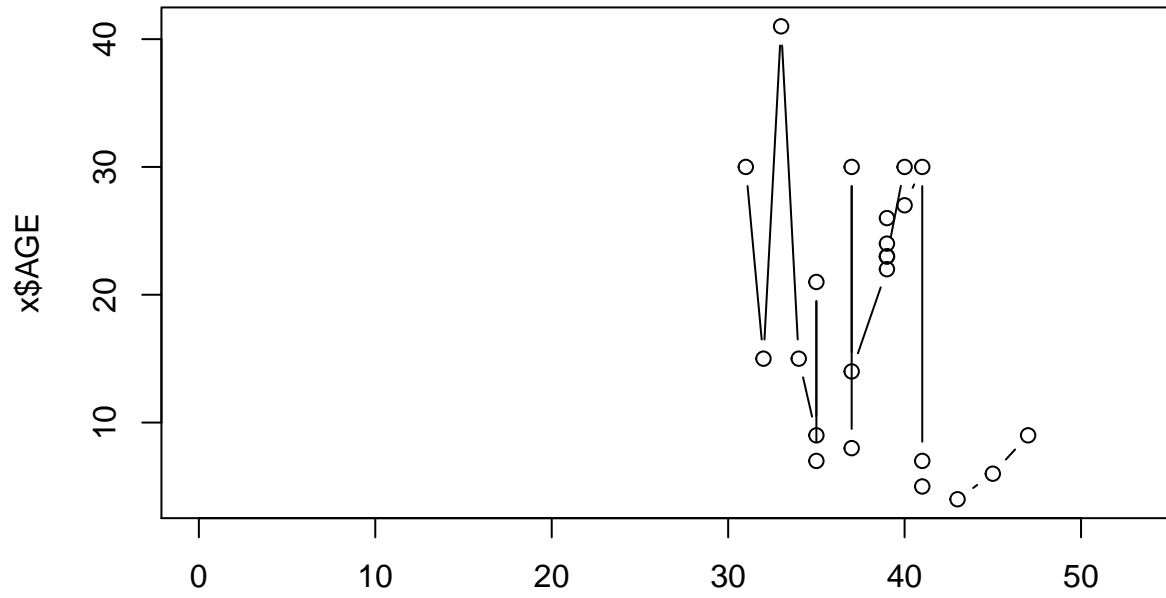
CUBATAO
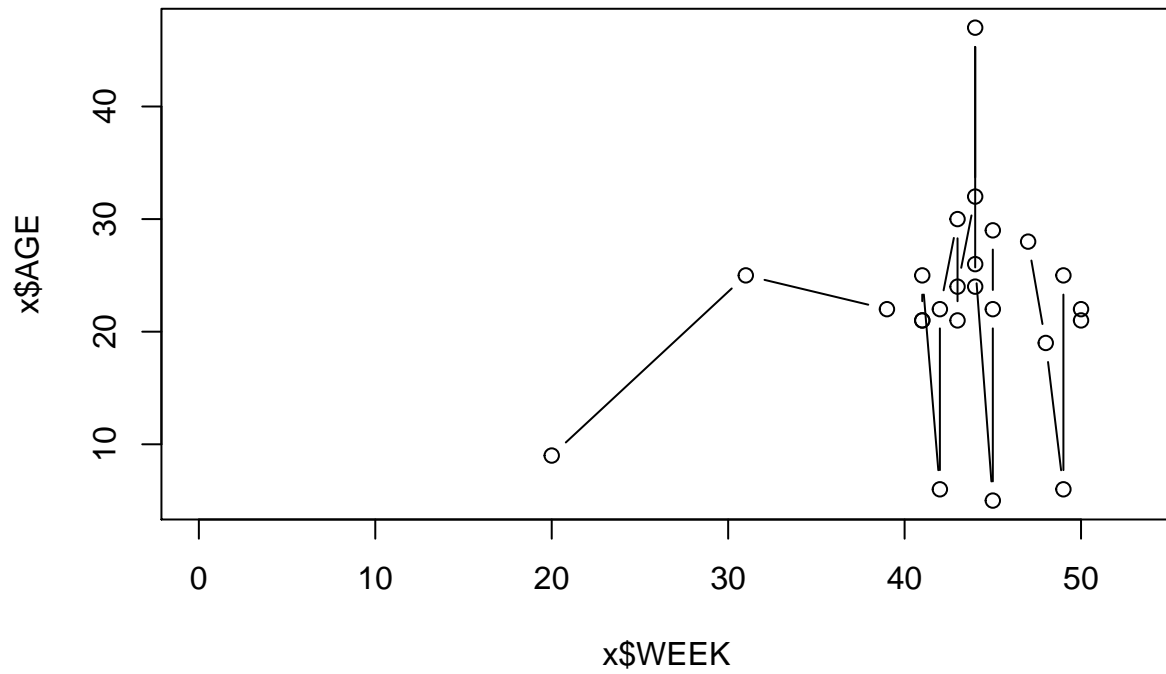
SAO JOAO DA BOA VISTA
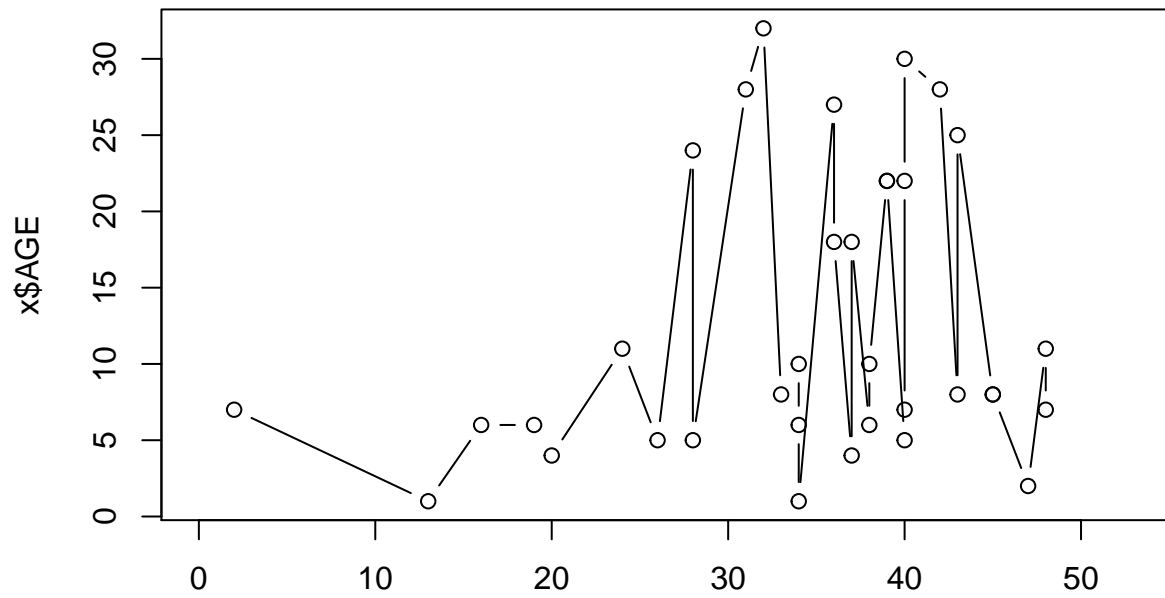
x$WEEK

ALVARES MACHADO

x$WEEK

19

ITAQUAQUECETUBA

x$WEEK

GUARATINGUETA

x$WEEK

SANTOS

x$WEEK

CAMPINAS

x$WEEK

PIRAJU

x$WEEK

ARARAQUARA

x$WEEK
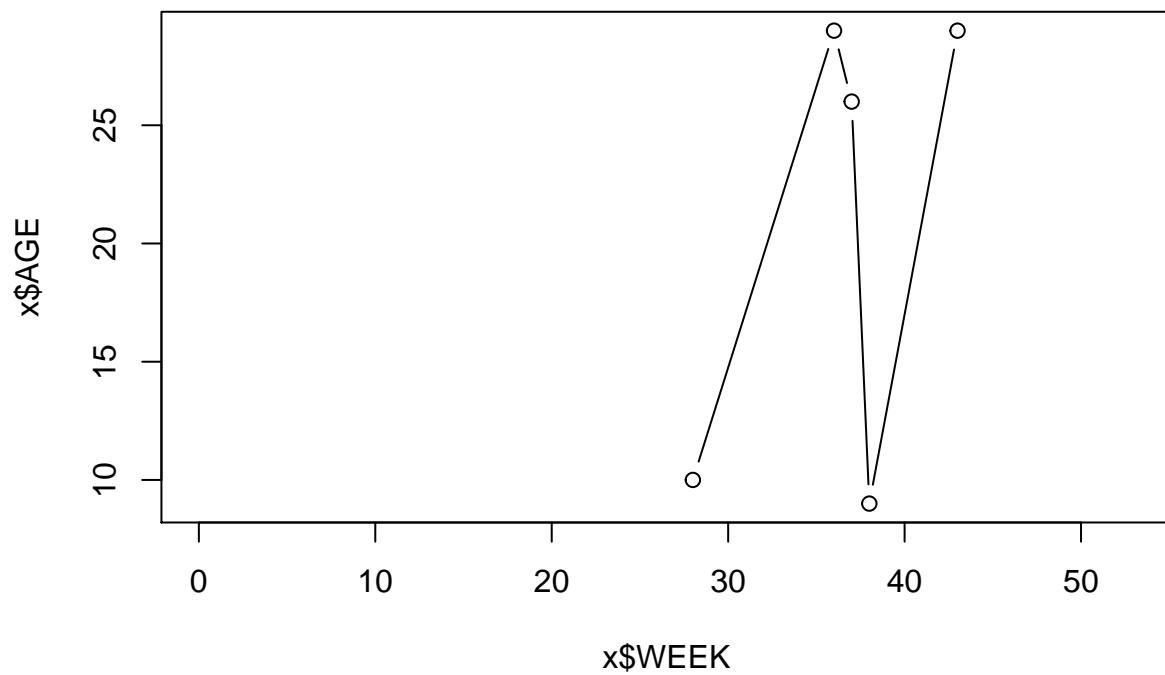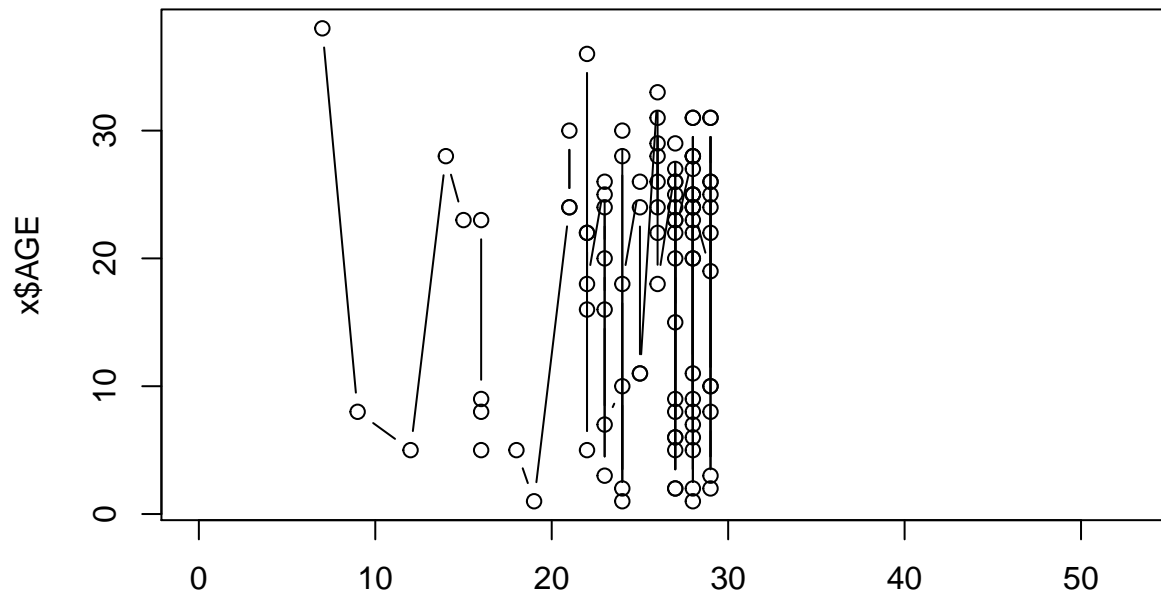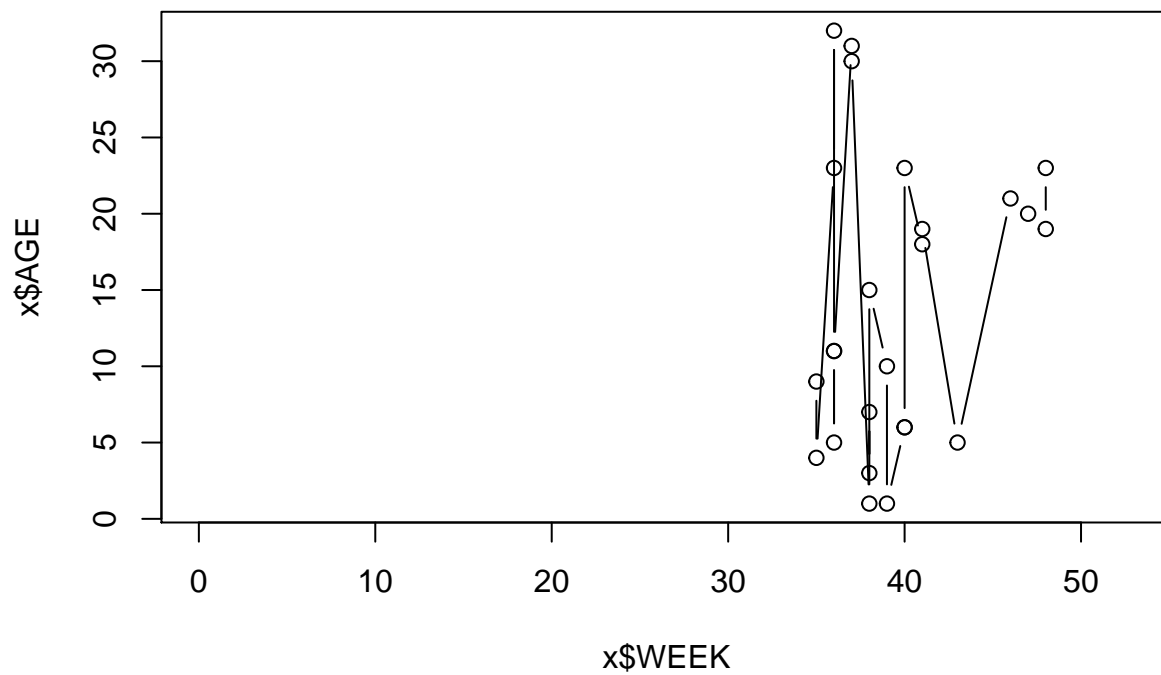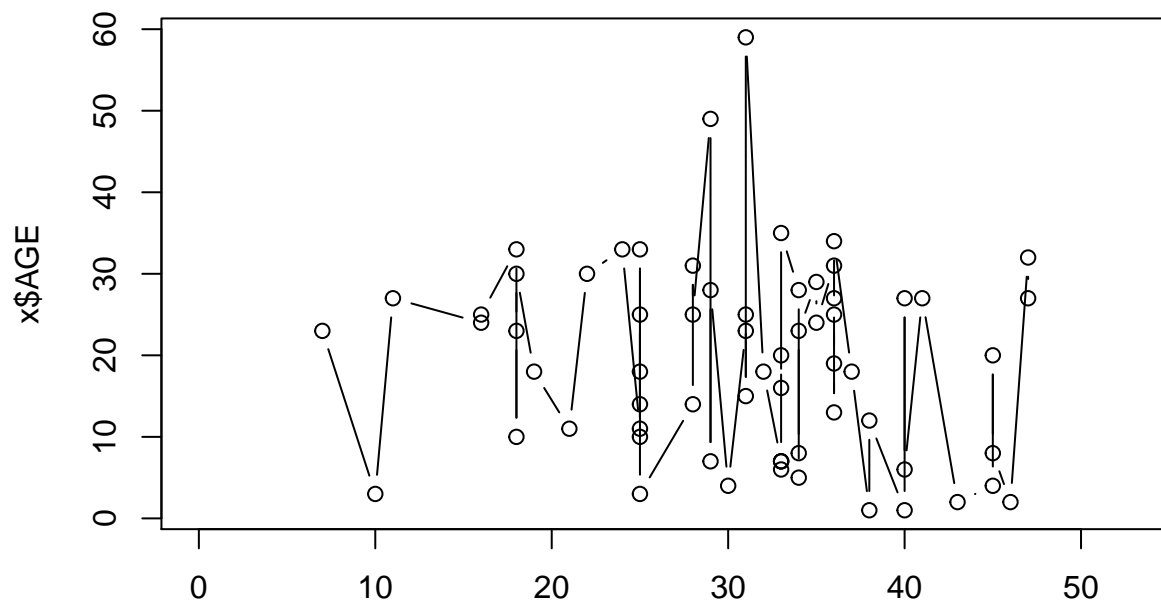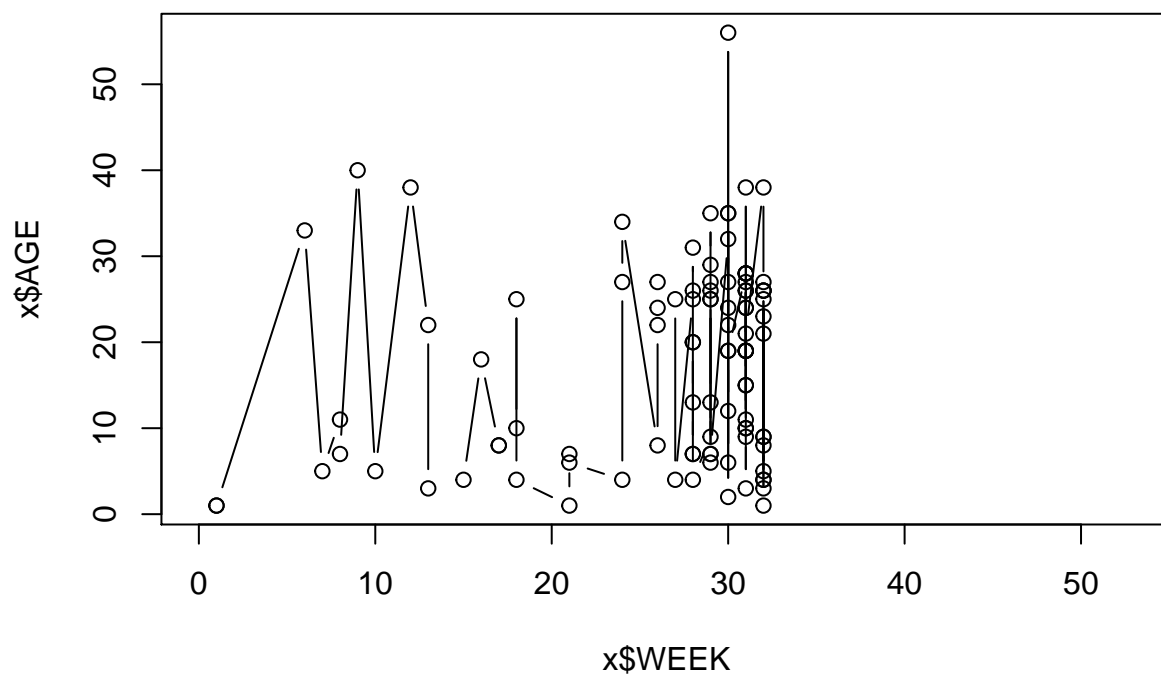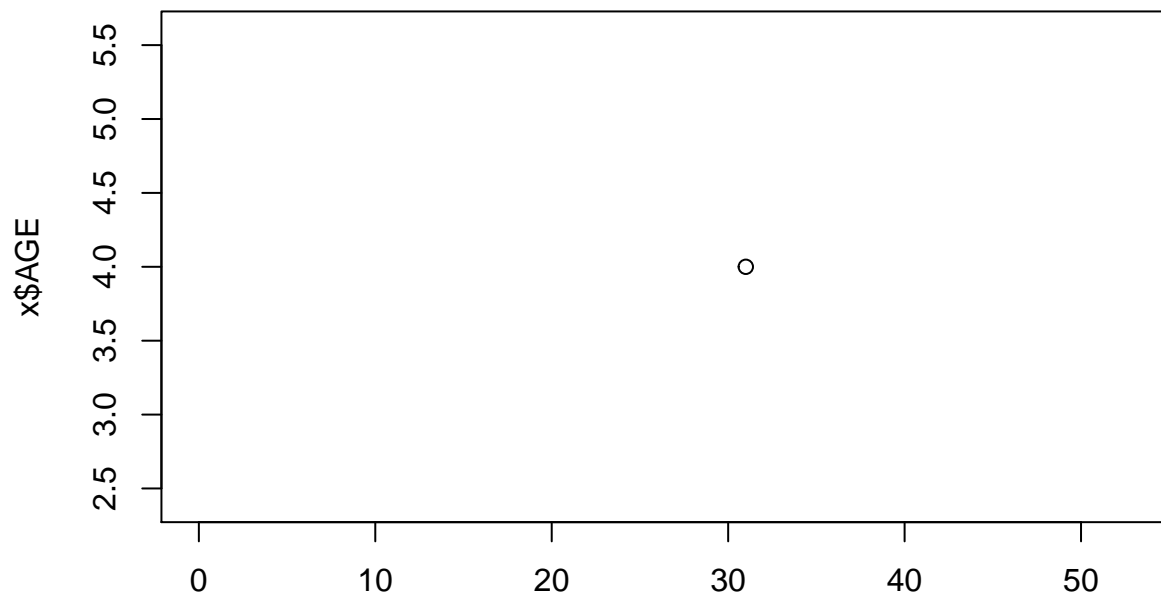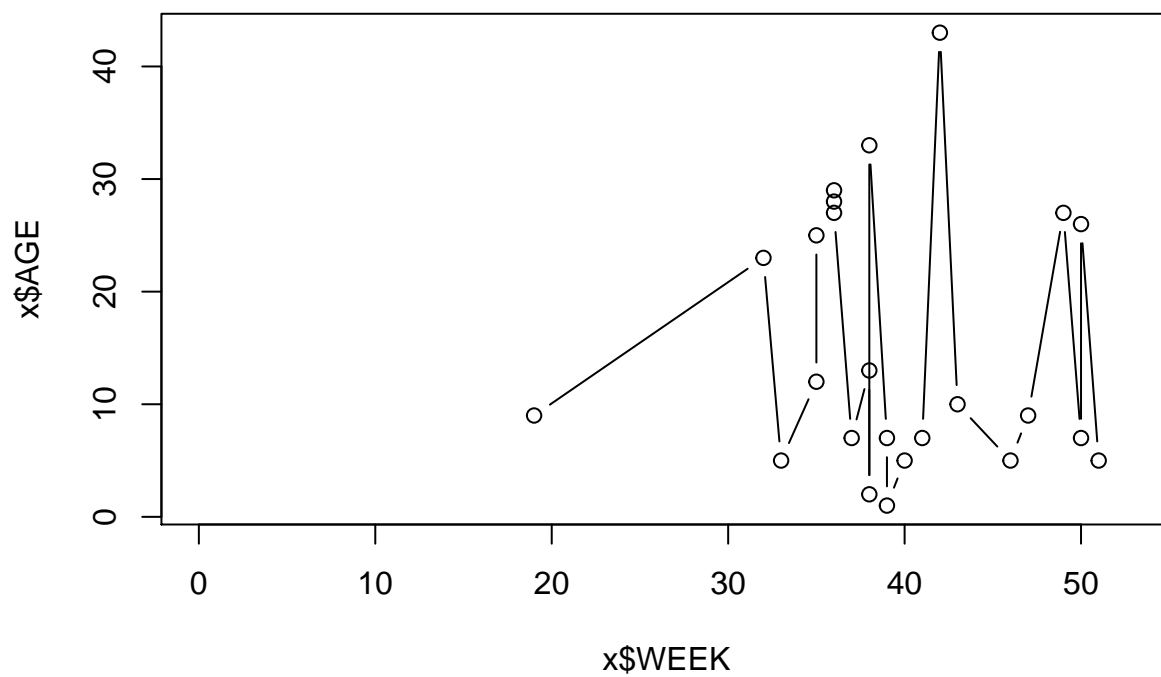
DOIS CORREGOS
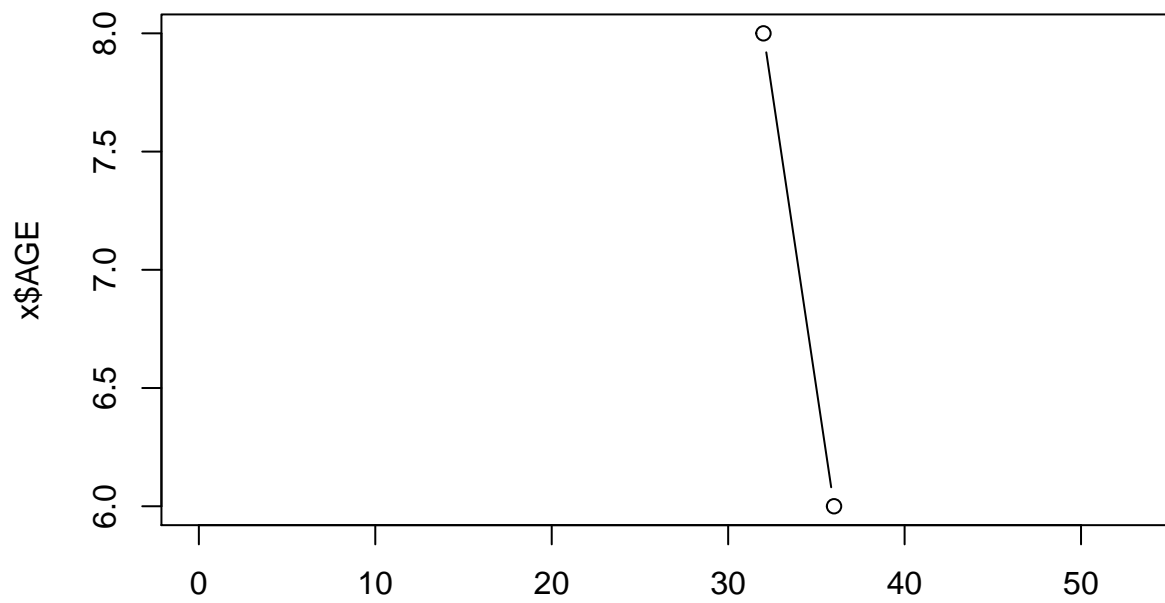
MIRANDOPOLIS

MOGI MIRIM

x$WEEK

ITUVERAVA

x$WEEK
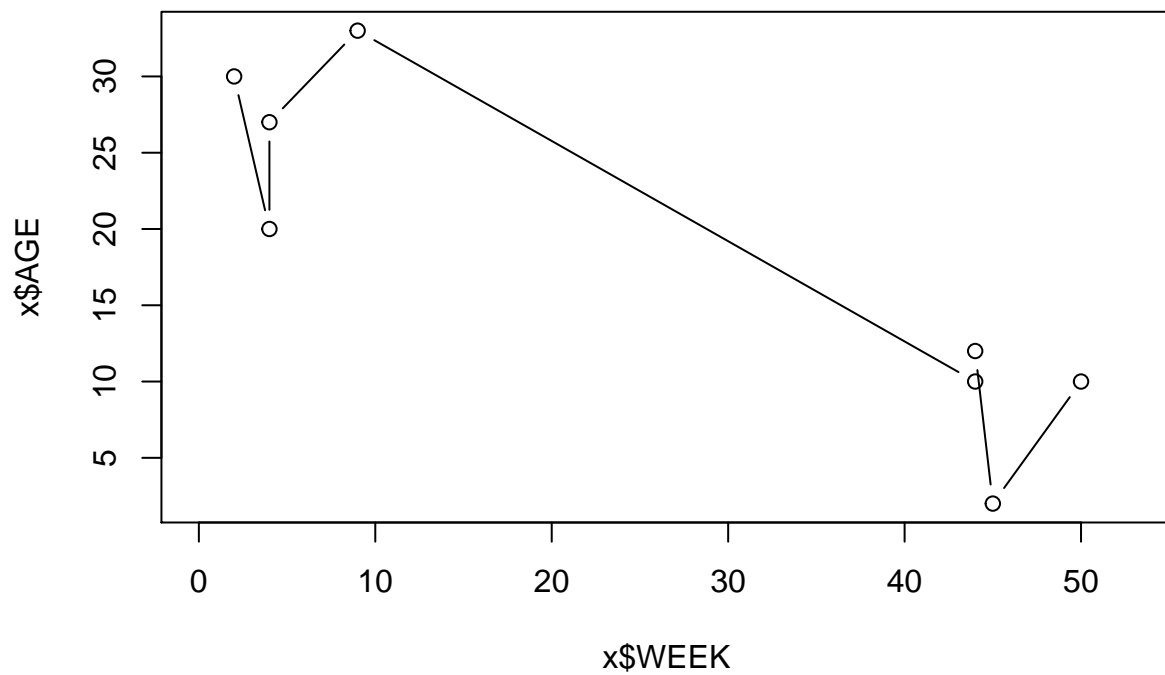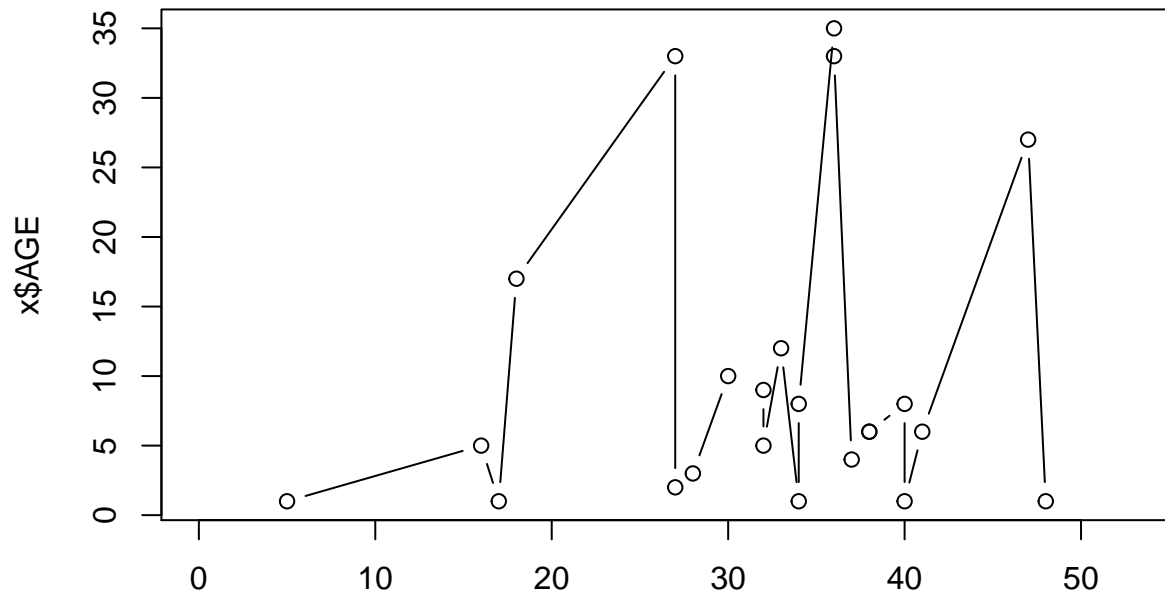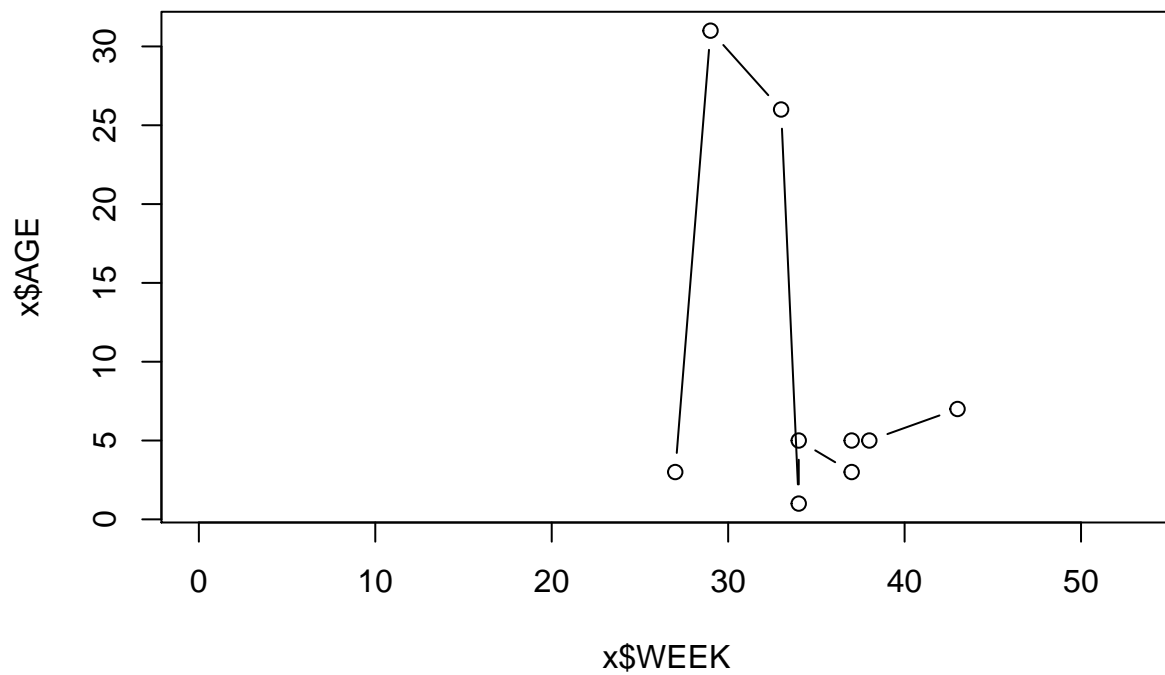
ITAPUI
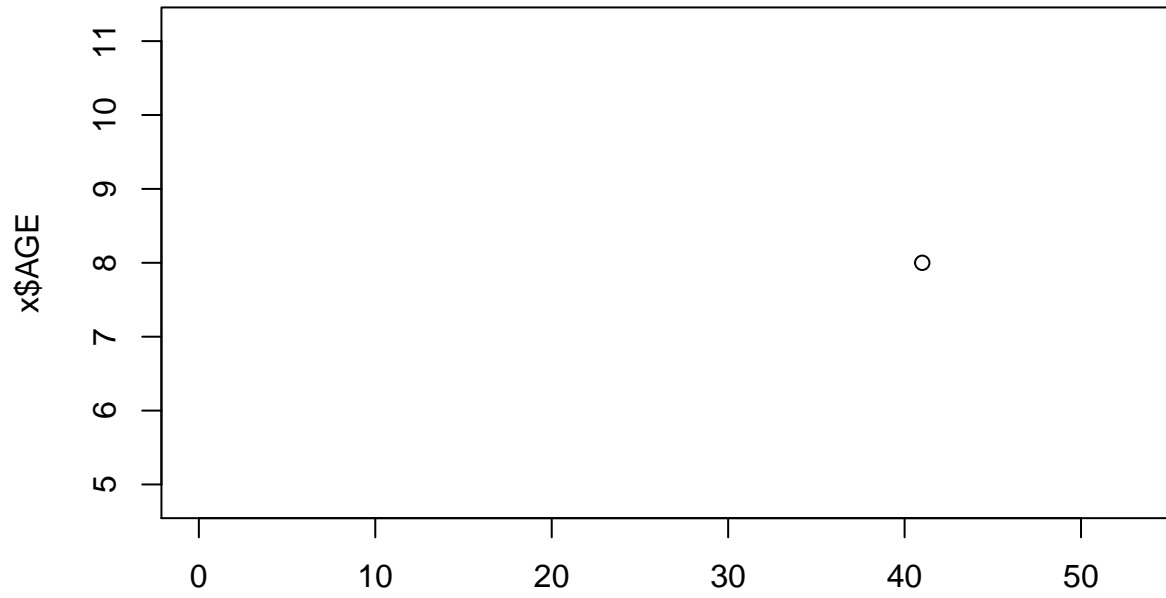
JANDIRA

LARANJAL PAULISTA

REGISTRO

MIRACATU

x$WEEK

SANTA ERNESTINA

x$WEEK

GUARA

AGUDOS
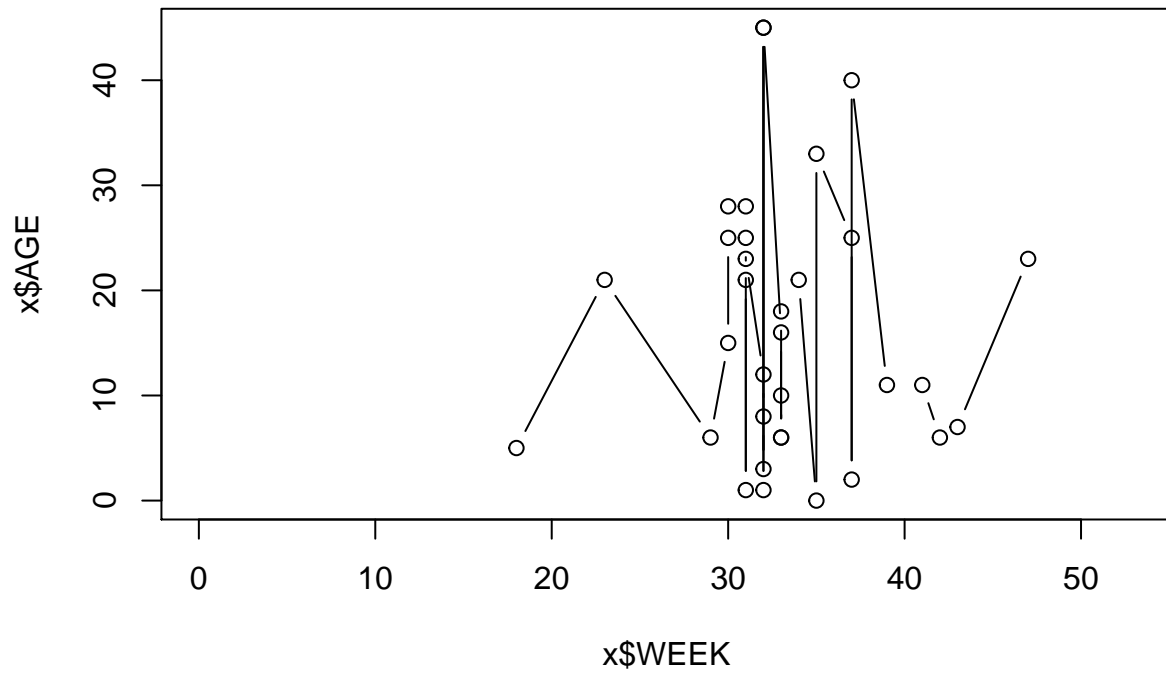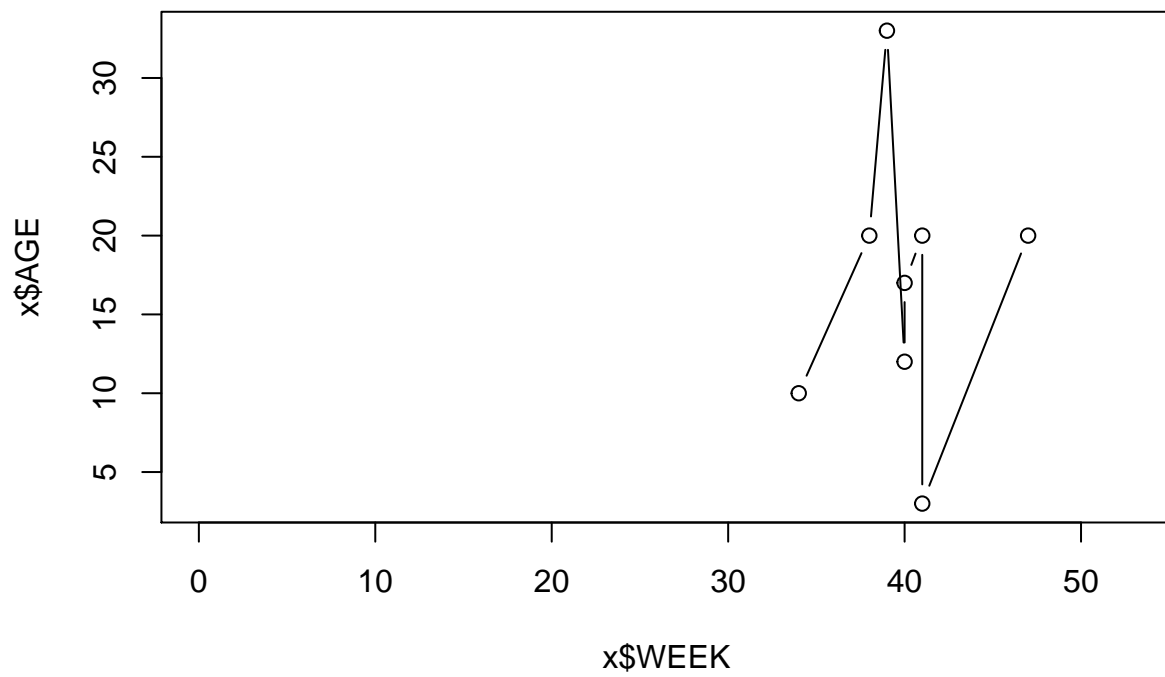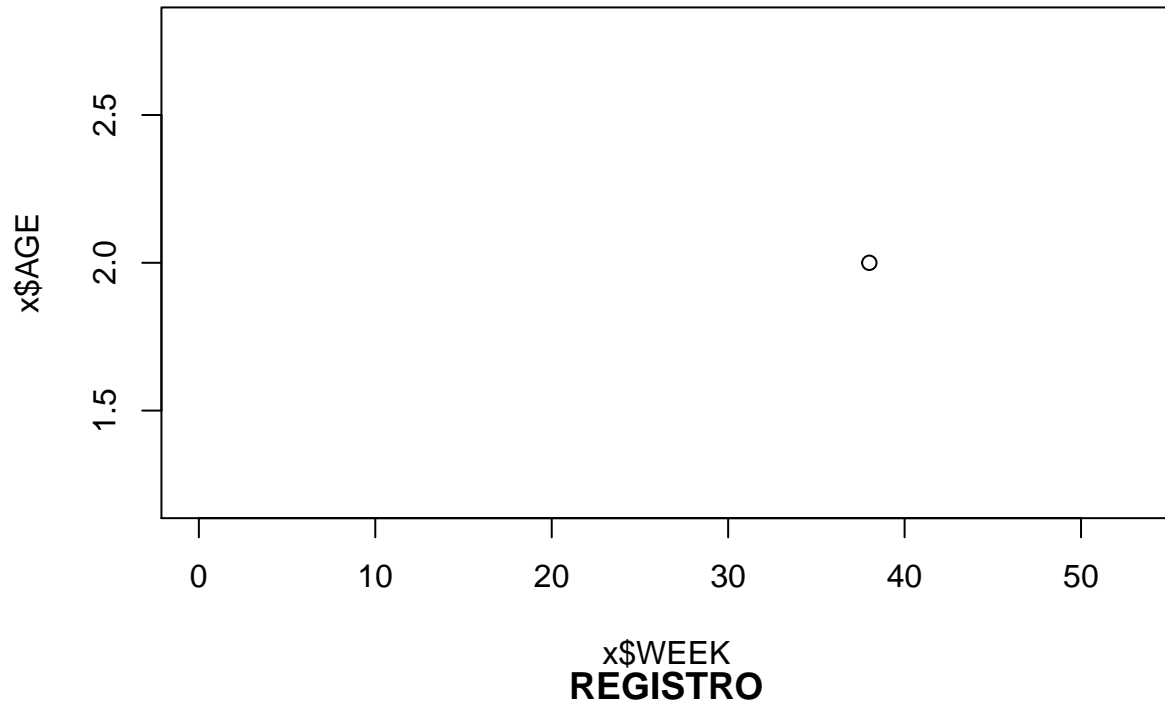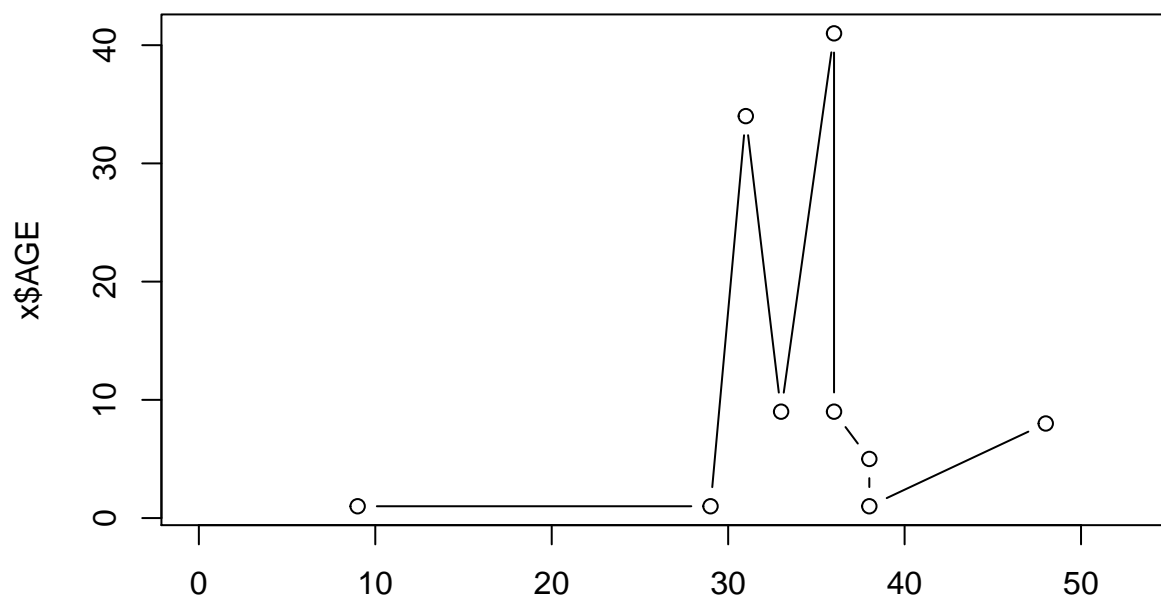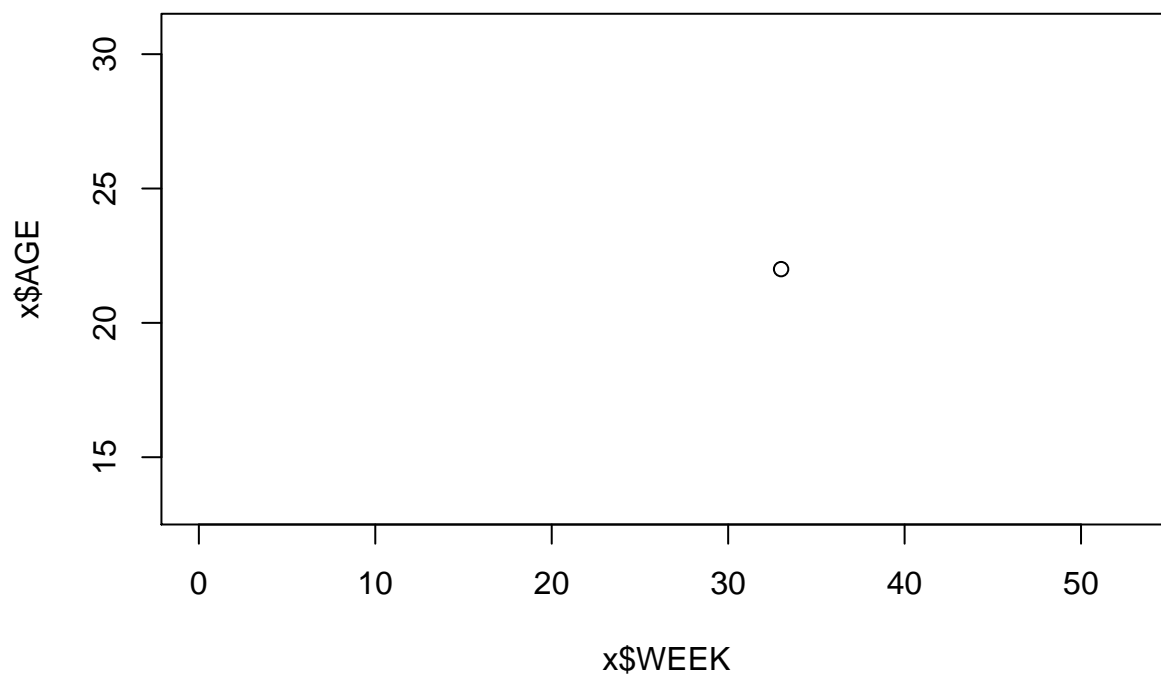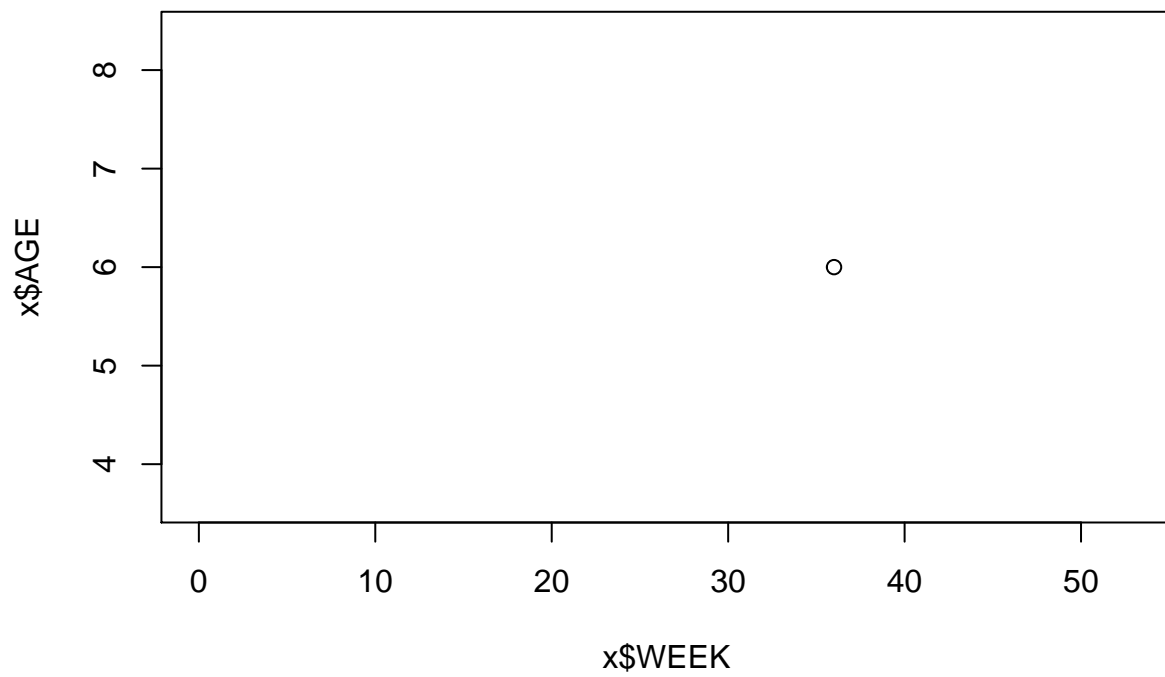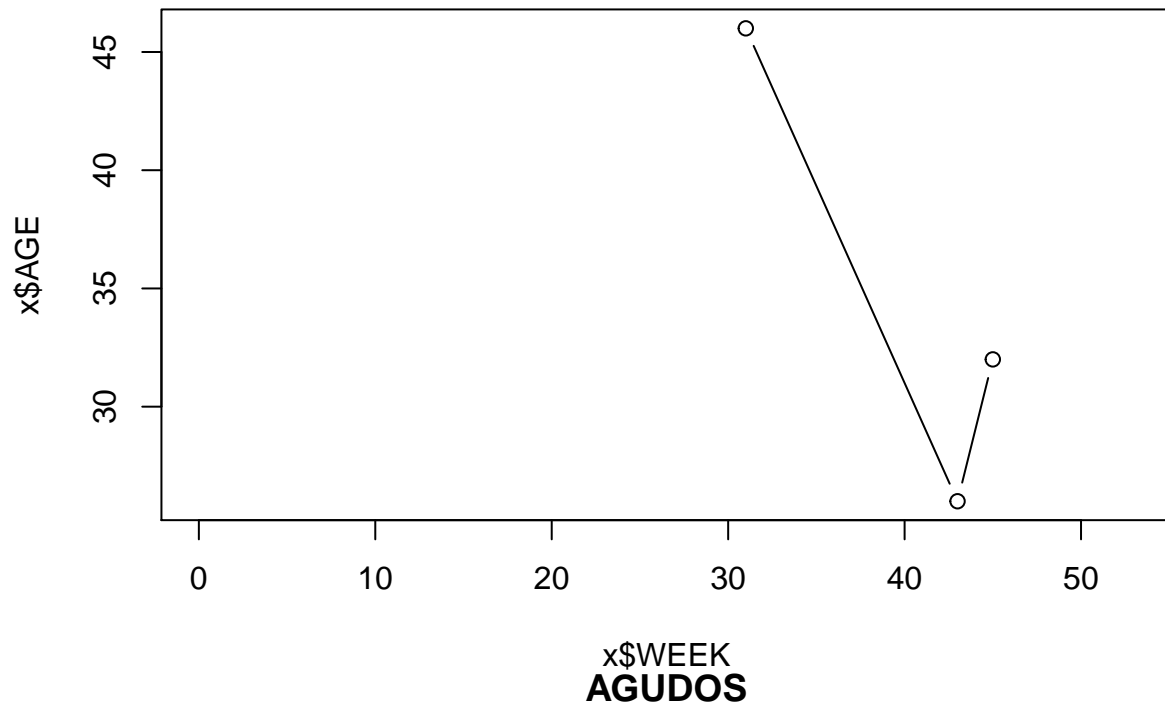
COTIA
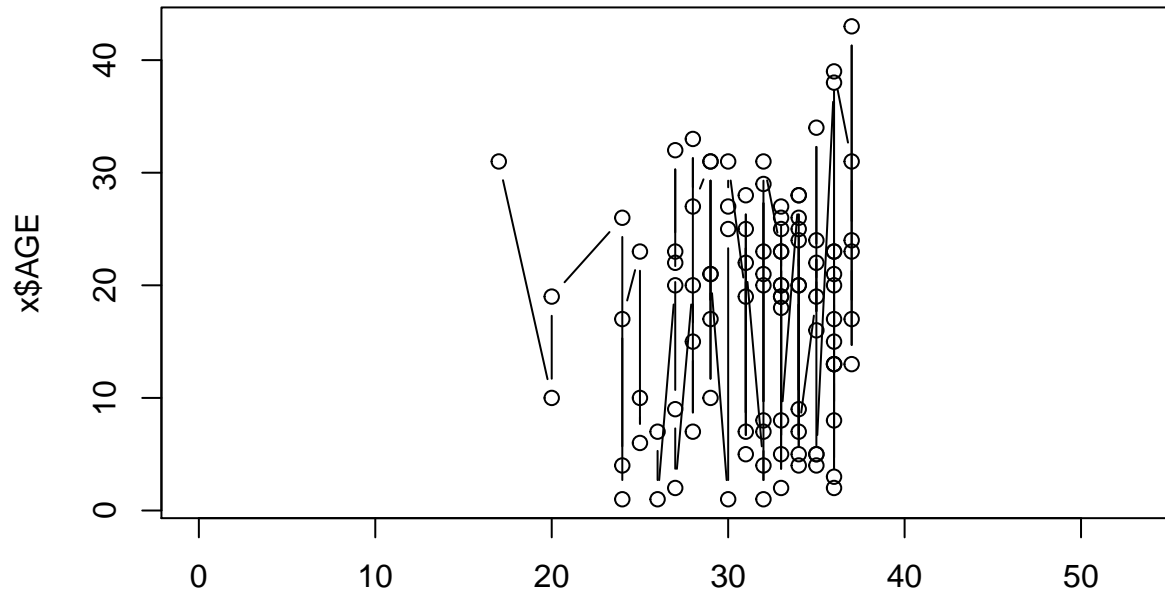
ARARAS
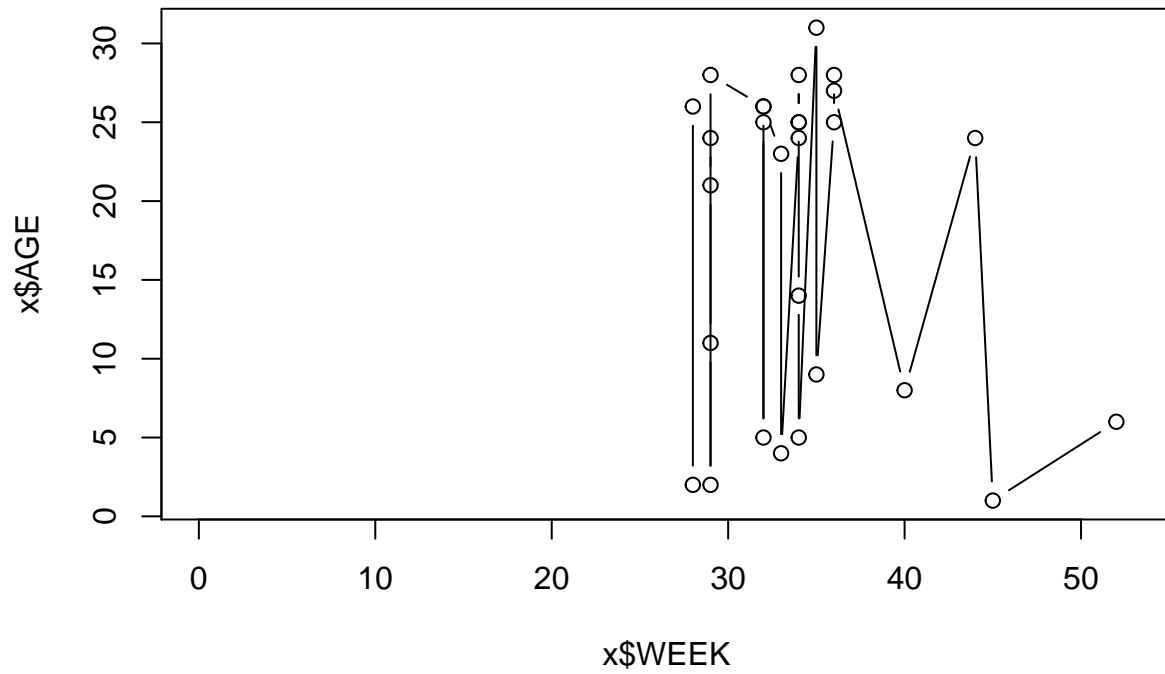
ITAPORANGA
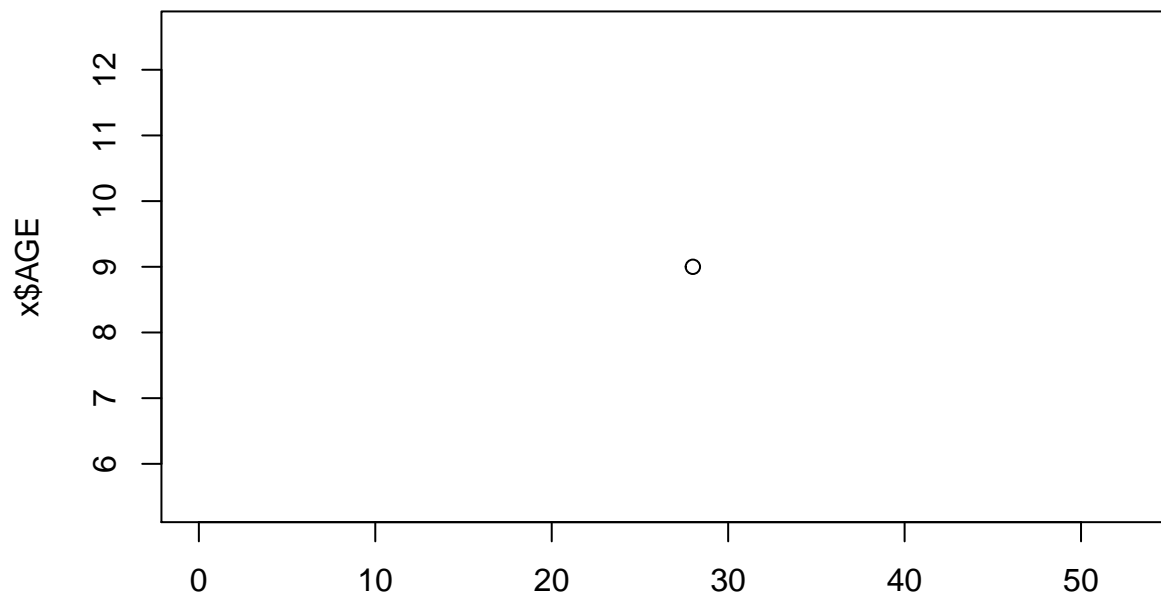
LENCOIS PAULISTA
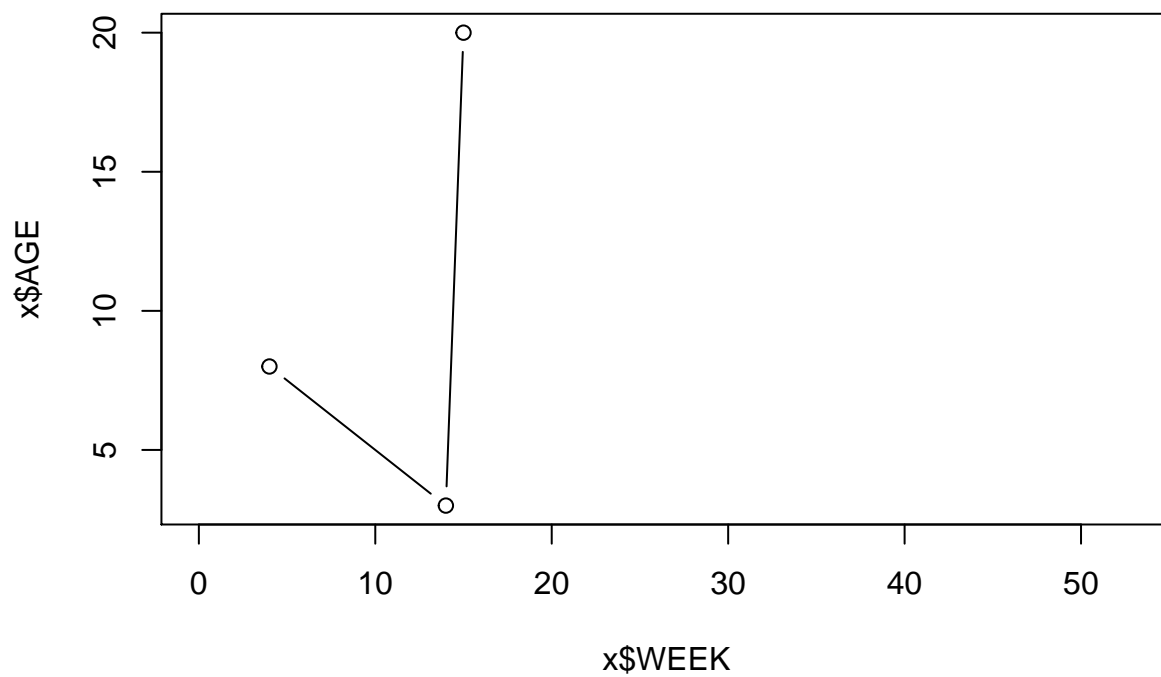
BERTIOGA
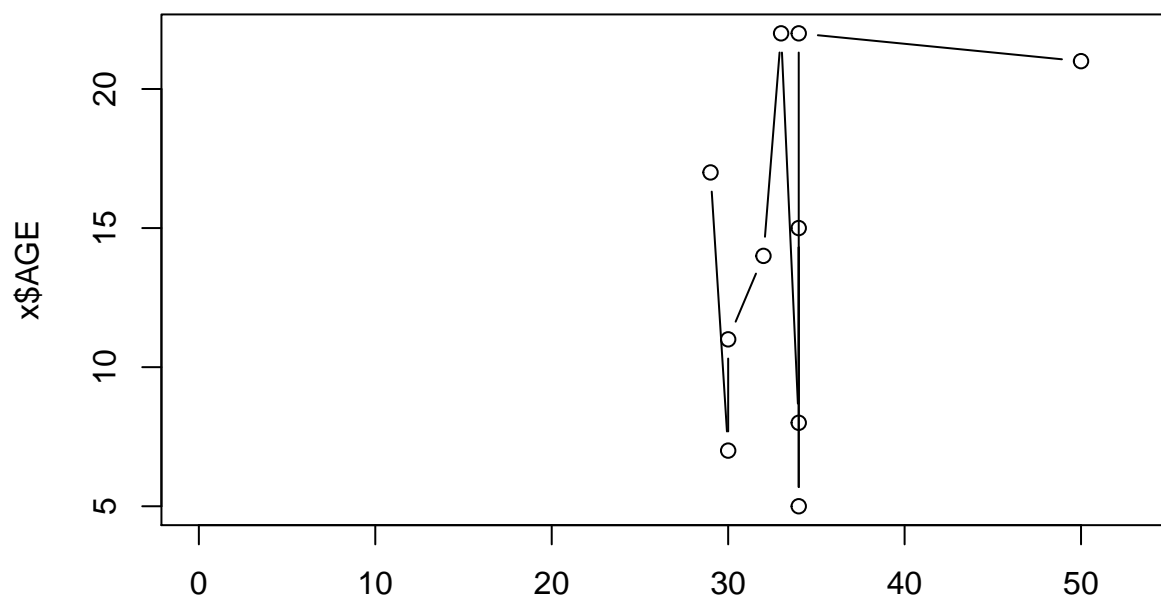
PARDINHO
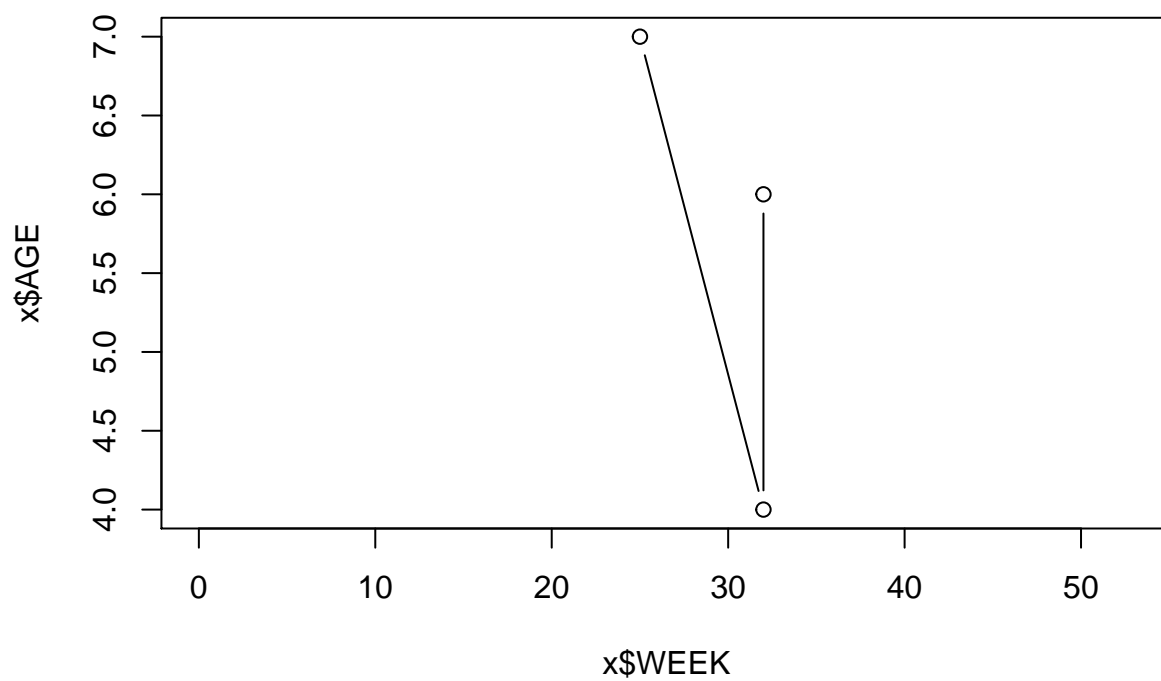
BARRETOS
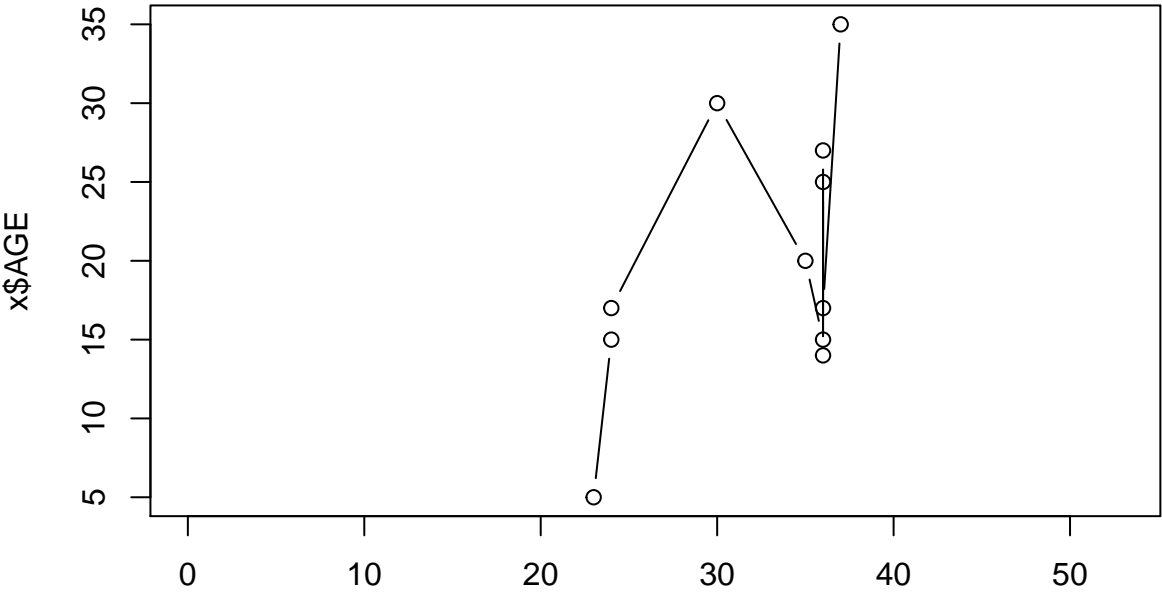
AREIOPOLIS
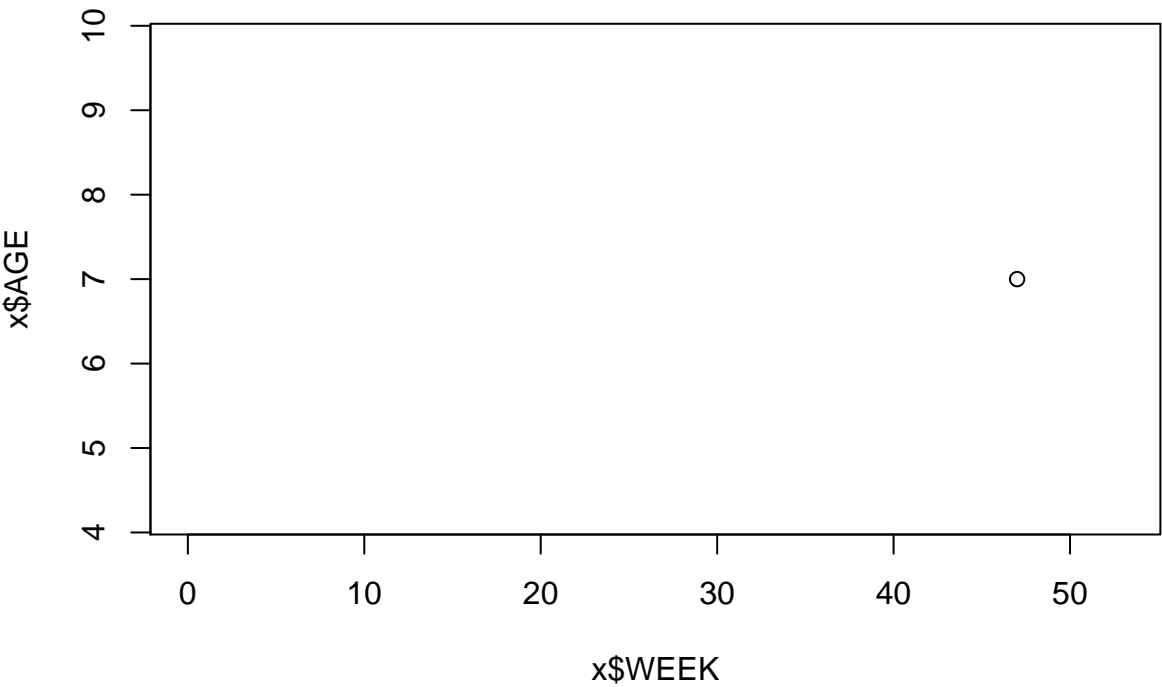
PINDAMONHANGABA

CONCHAS

JAU

x$WEEK

BRODOWSKI

x$WEEK

34
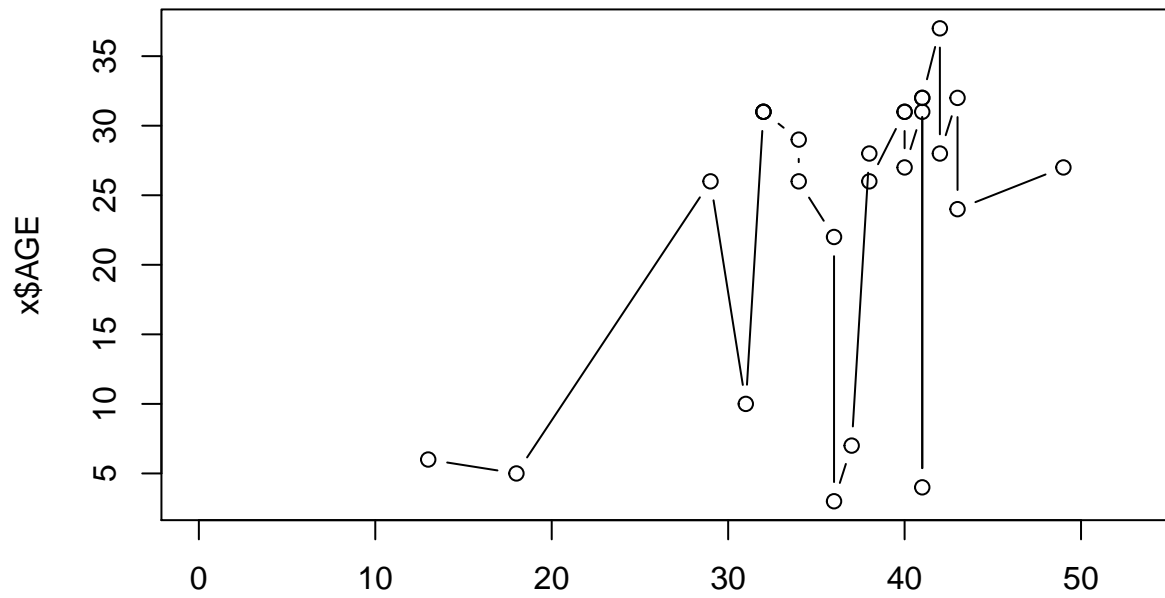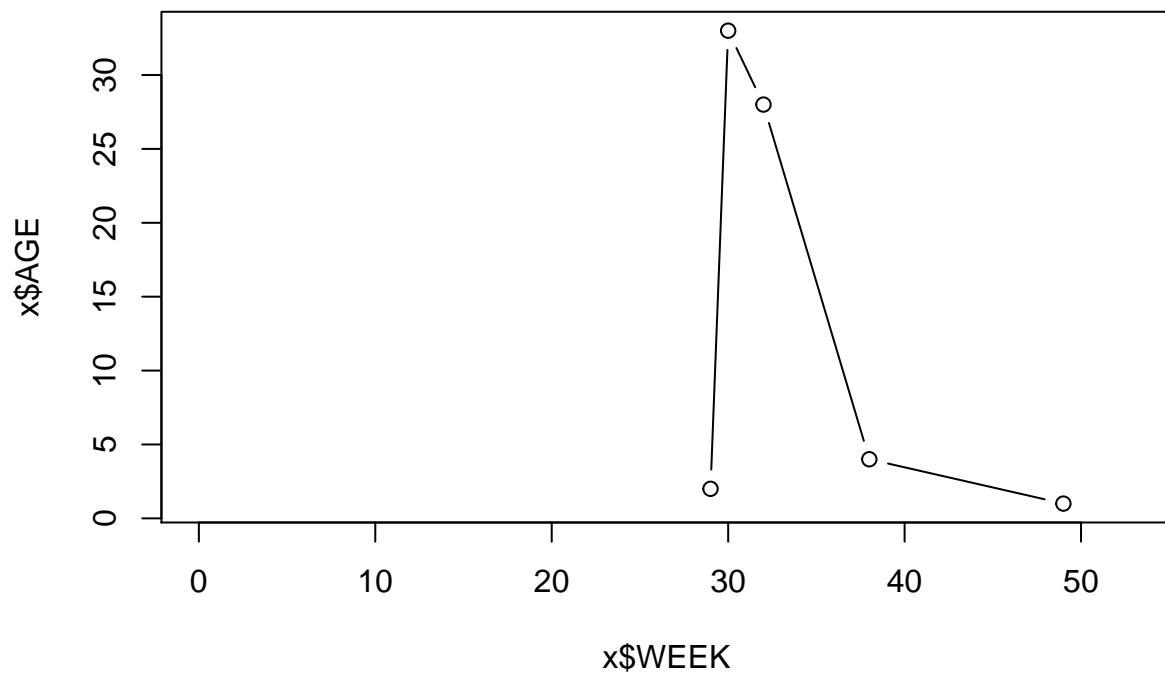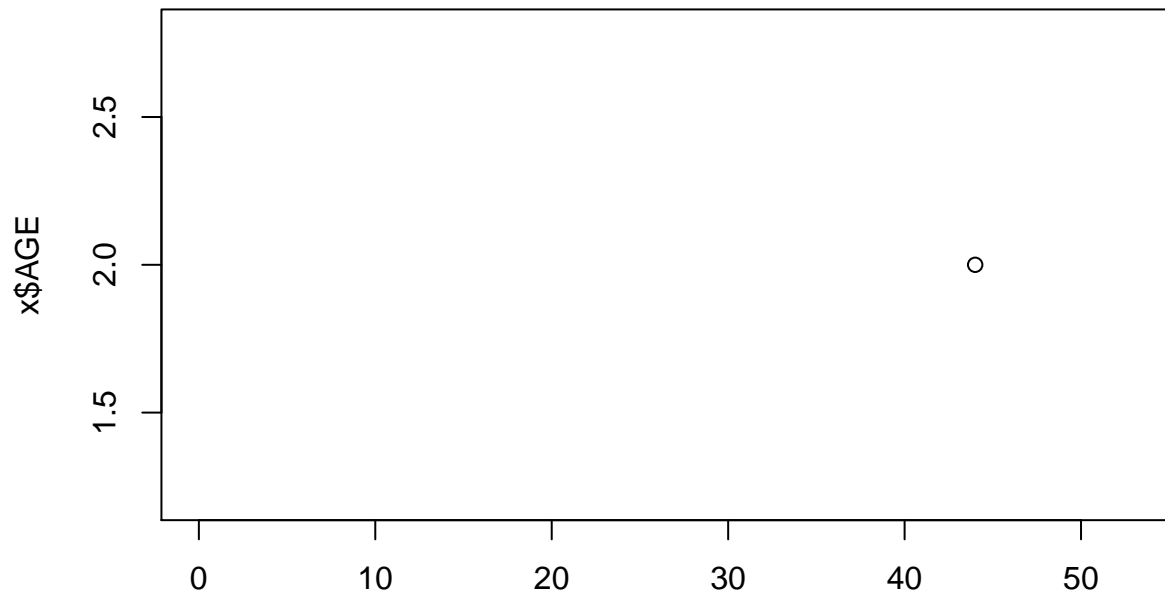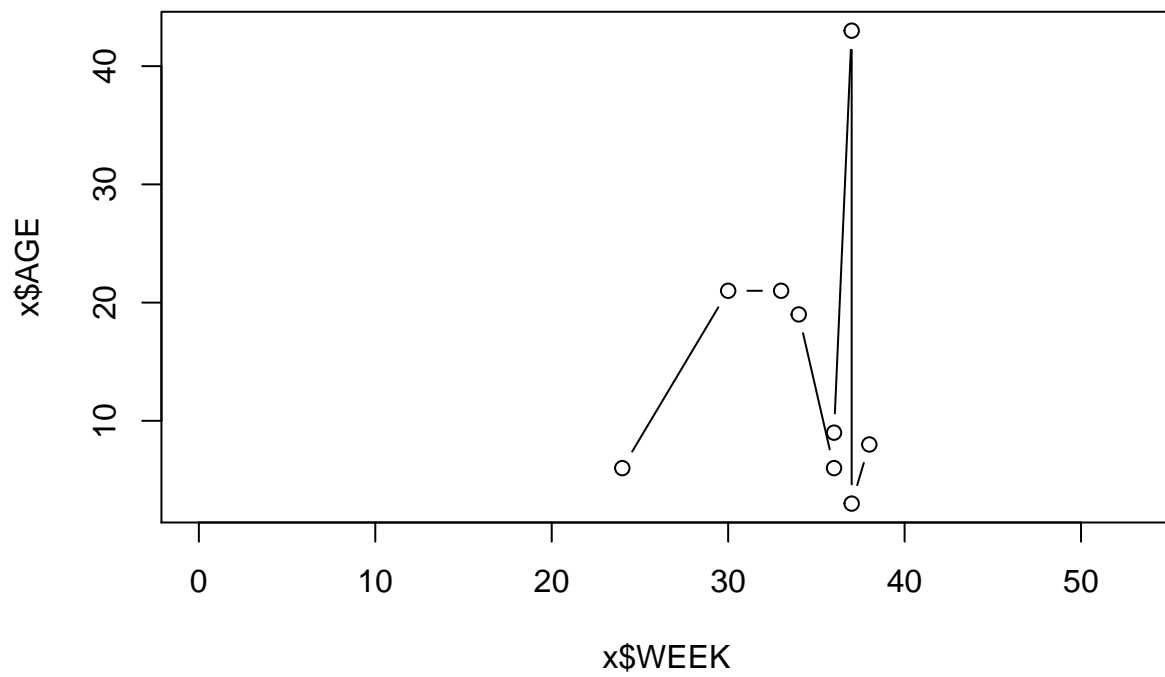
CAJURU
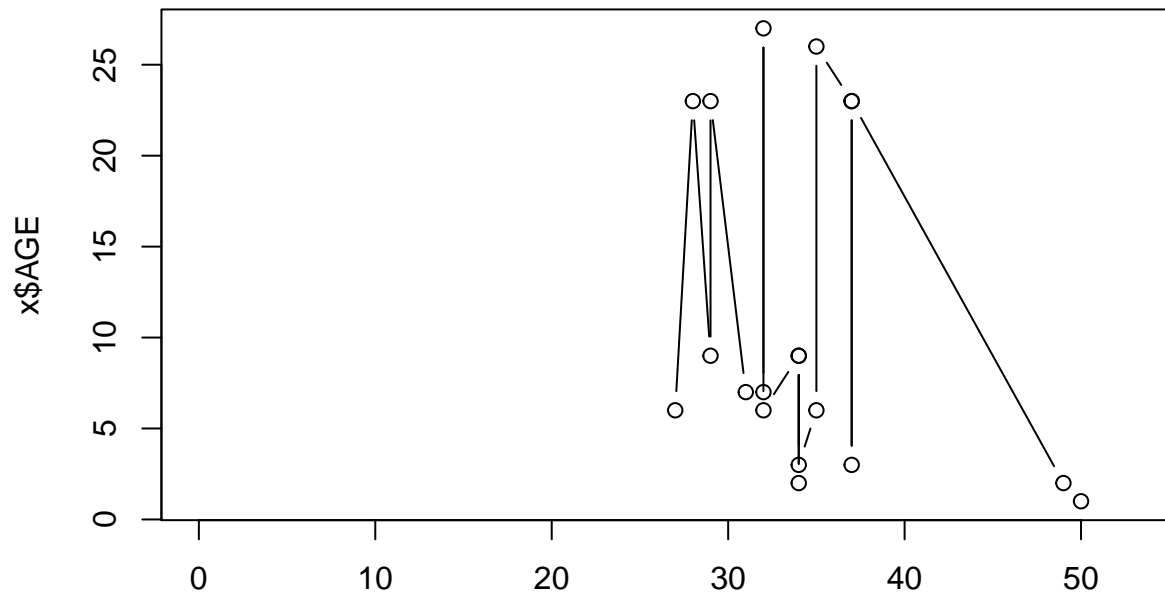
DUMONT
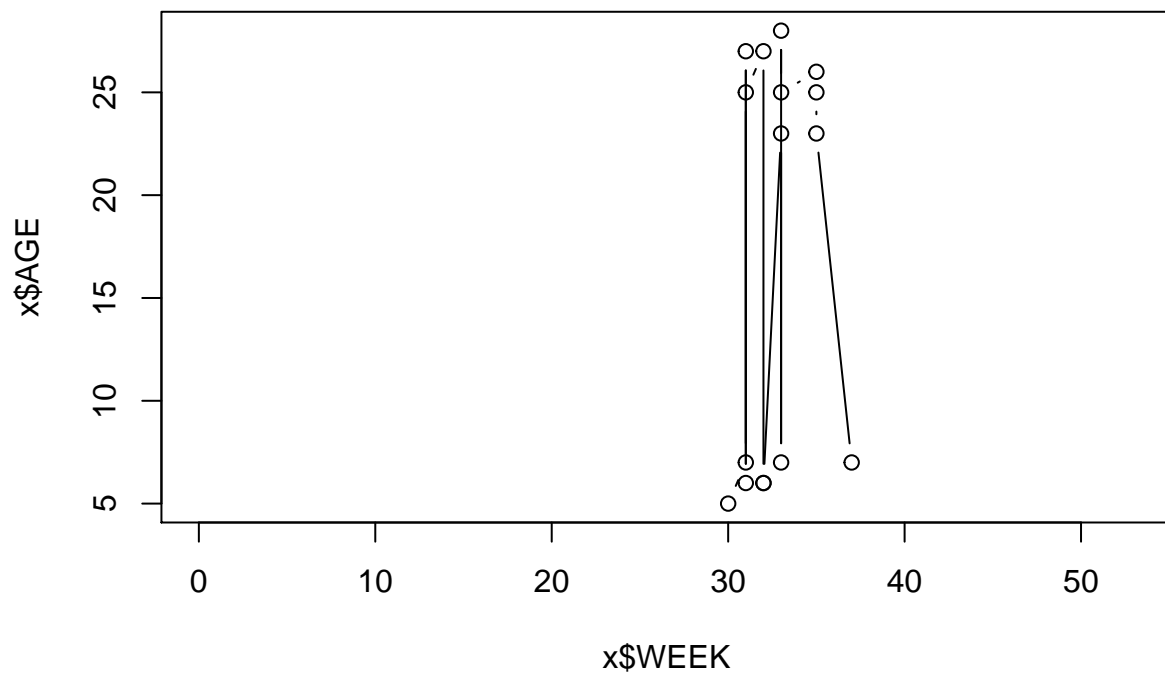
FRANCO DA ROCHA
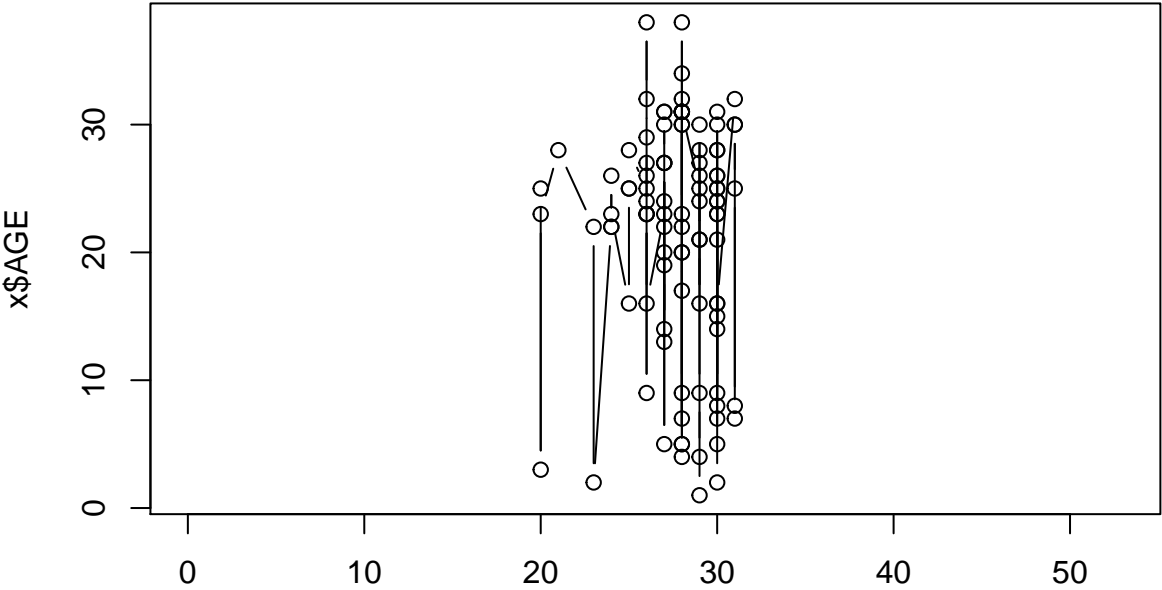
ALVINLANDIA
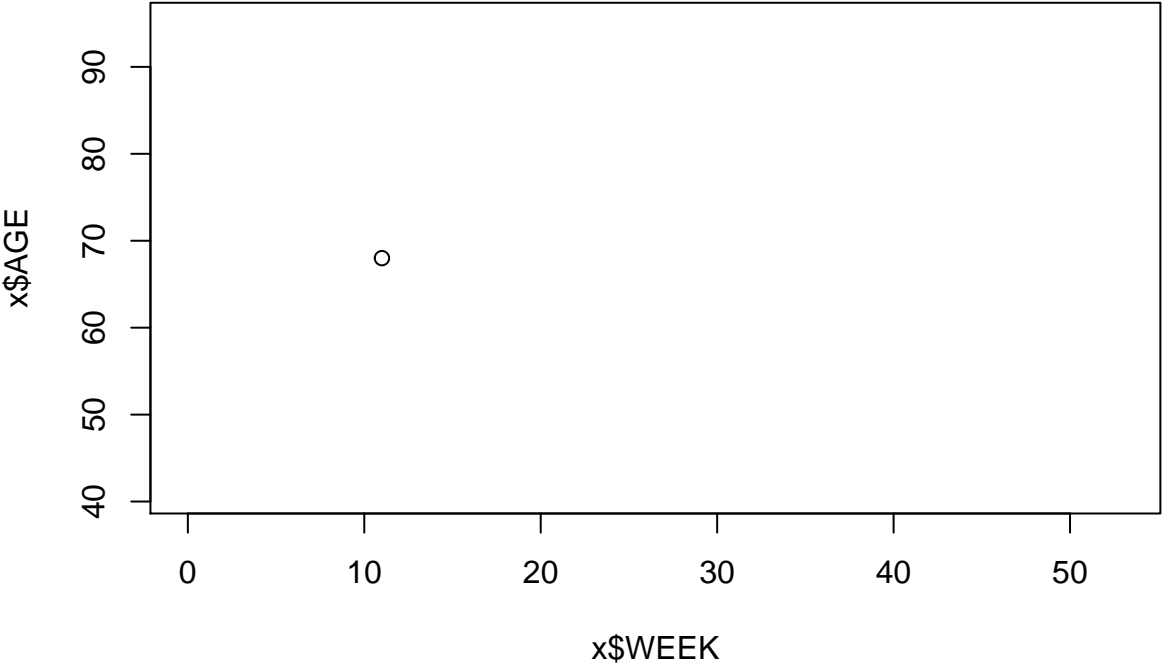
DRACENA
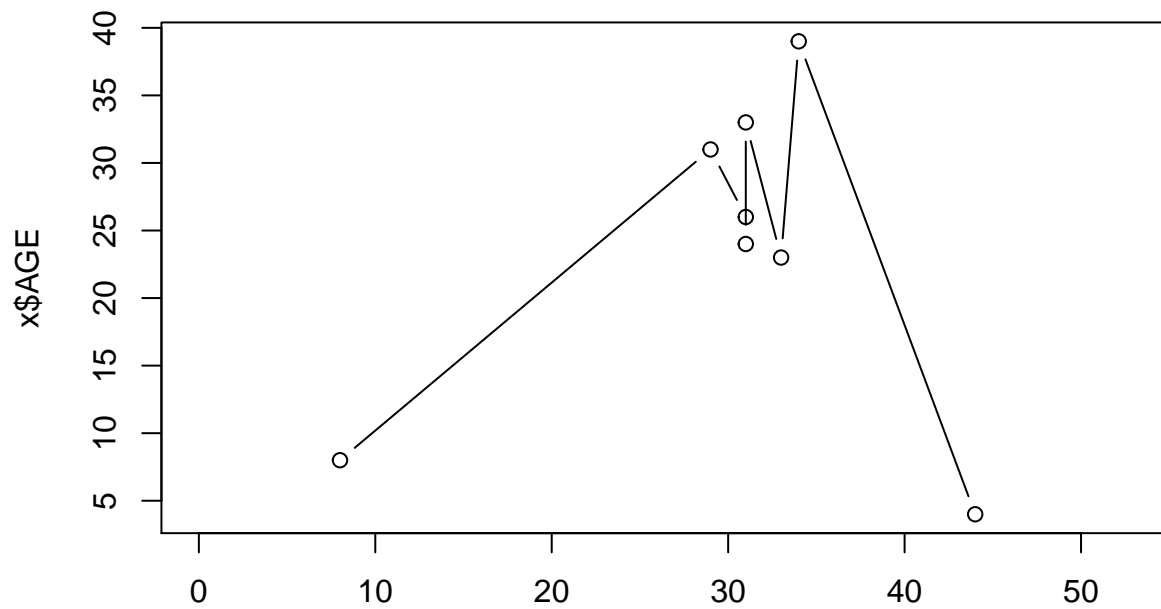
x$WEEK

MAUA

x$WEEK
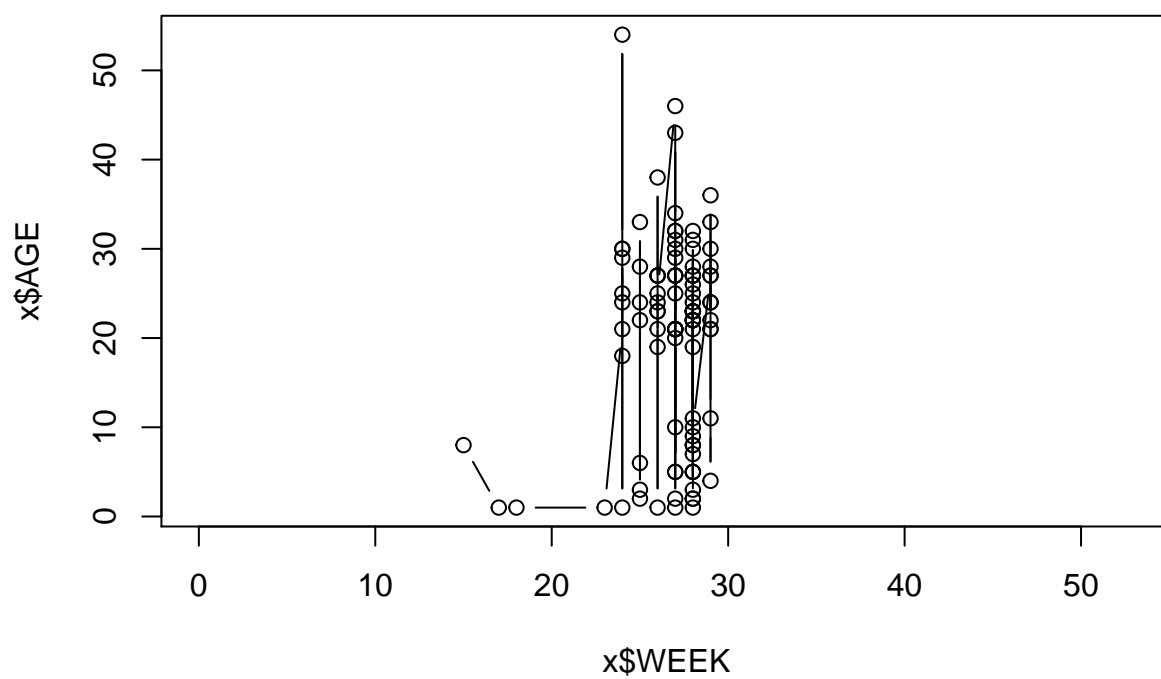
PARANAPANEMA
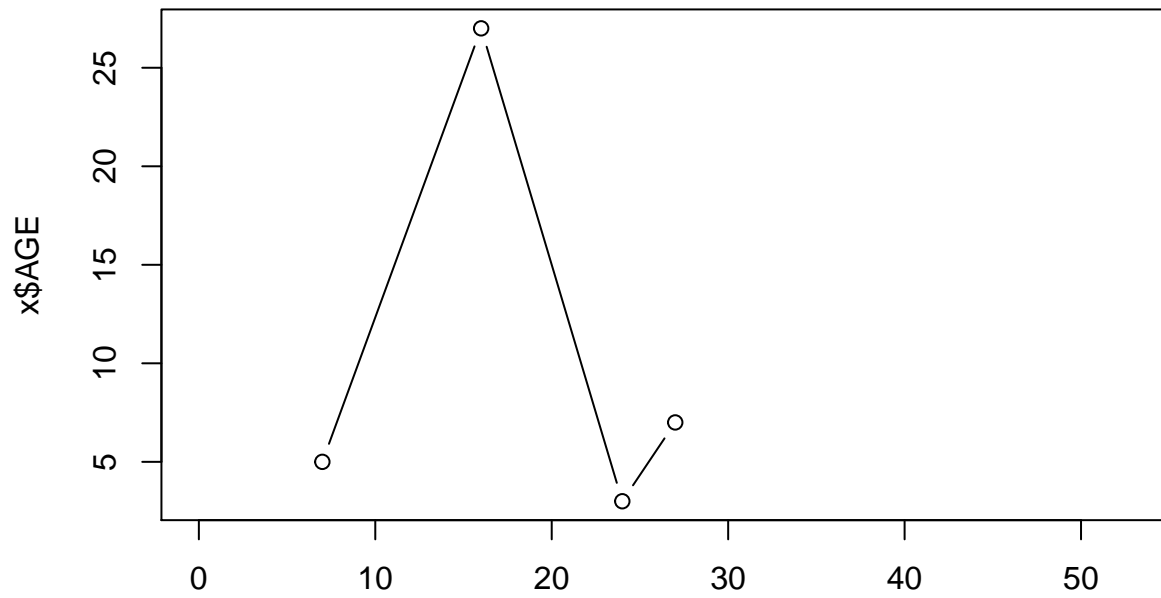
MOGI DAS CRUZES
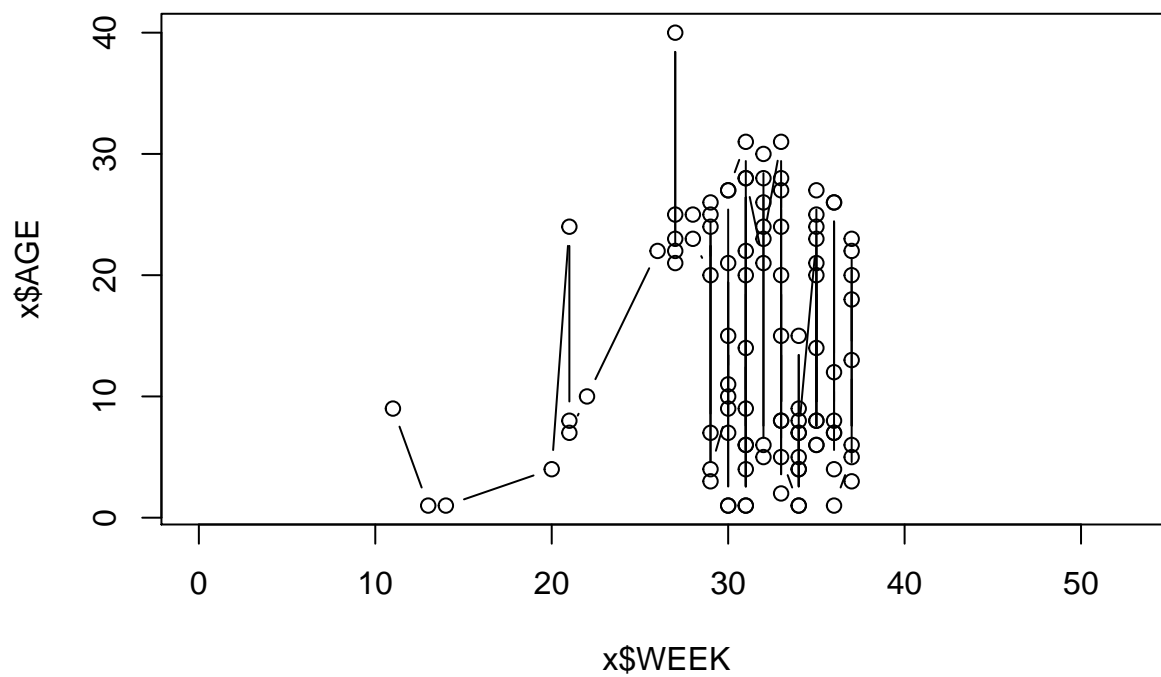
CARAPICUIBA
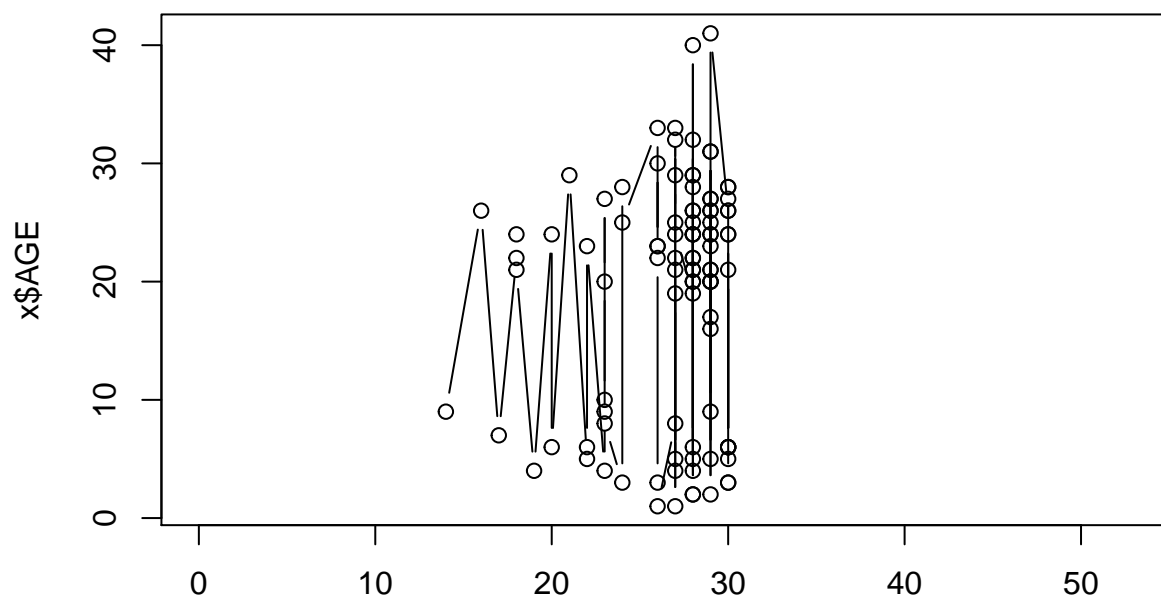
SAO JOSE DO RIO PRETO

EMBU–GUACU

x$WEEK

MIRASSOL

x$WEEK

40
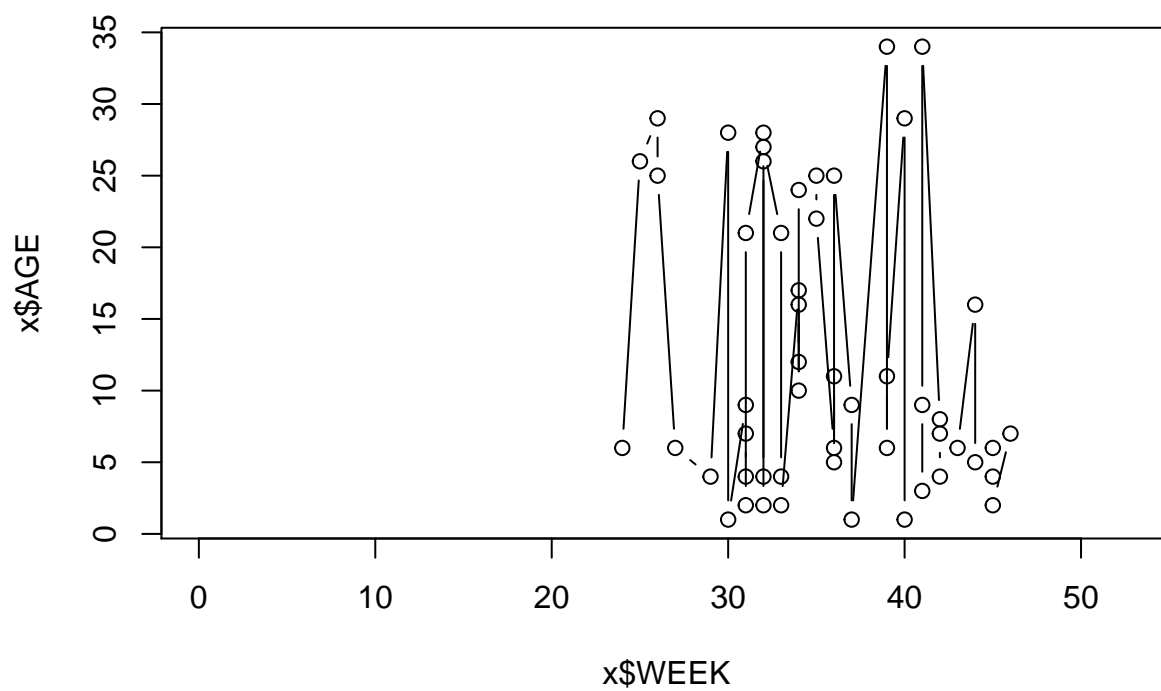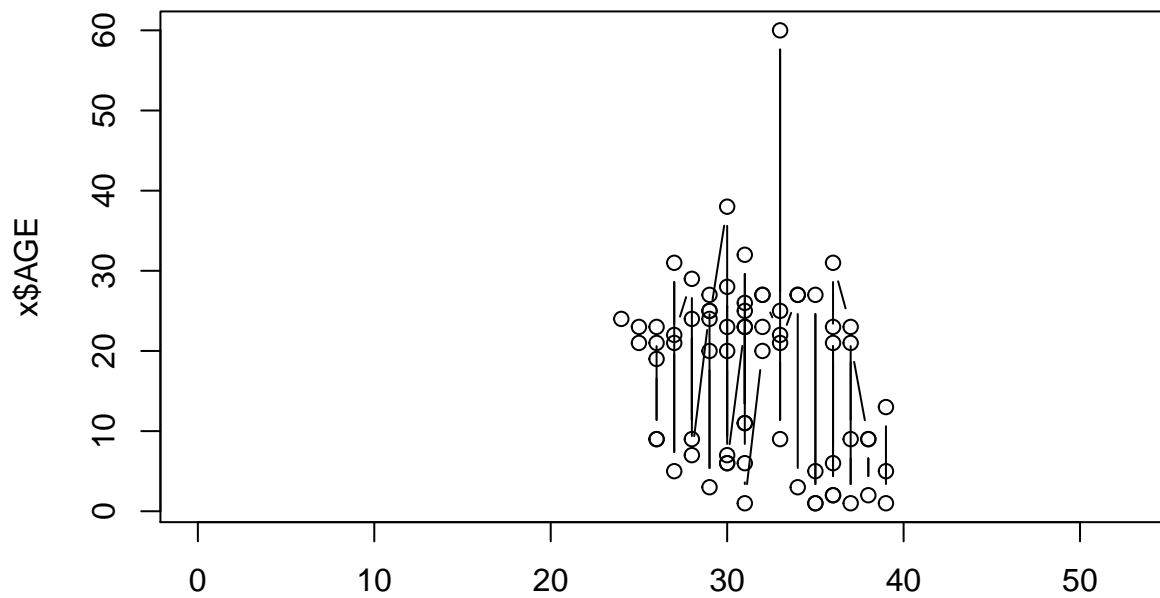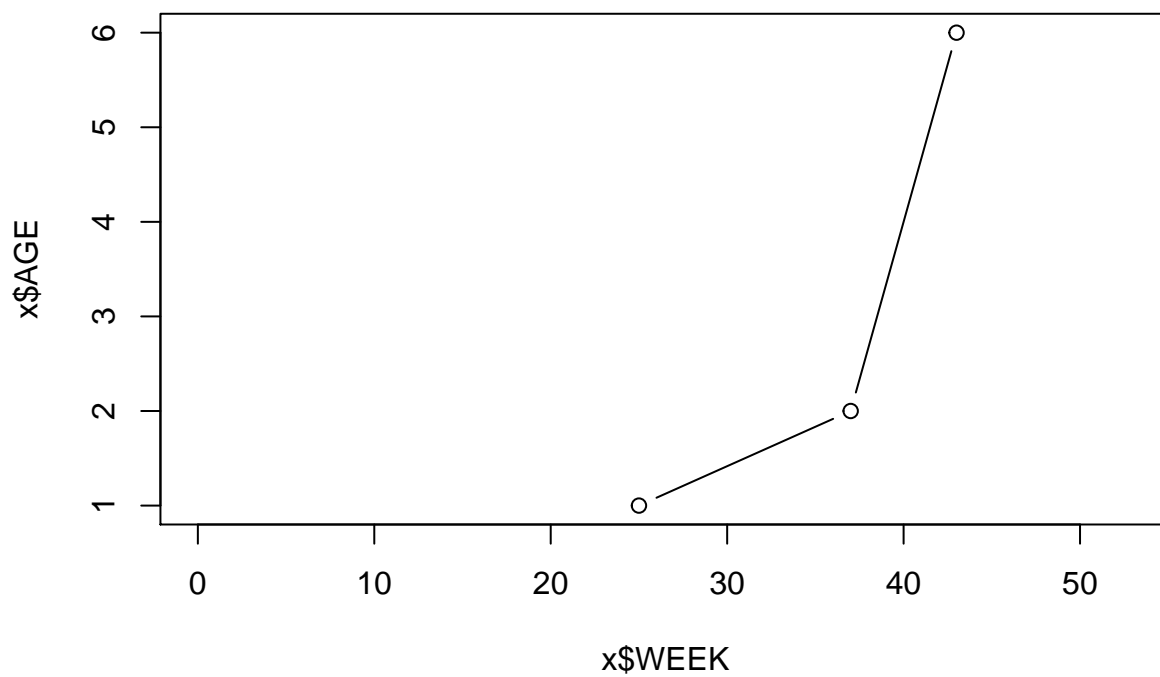
CATANDUVA
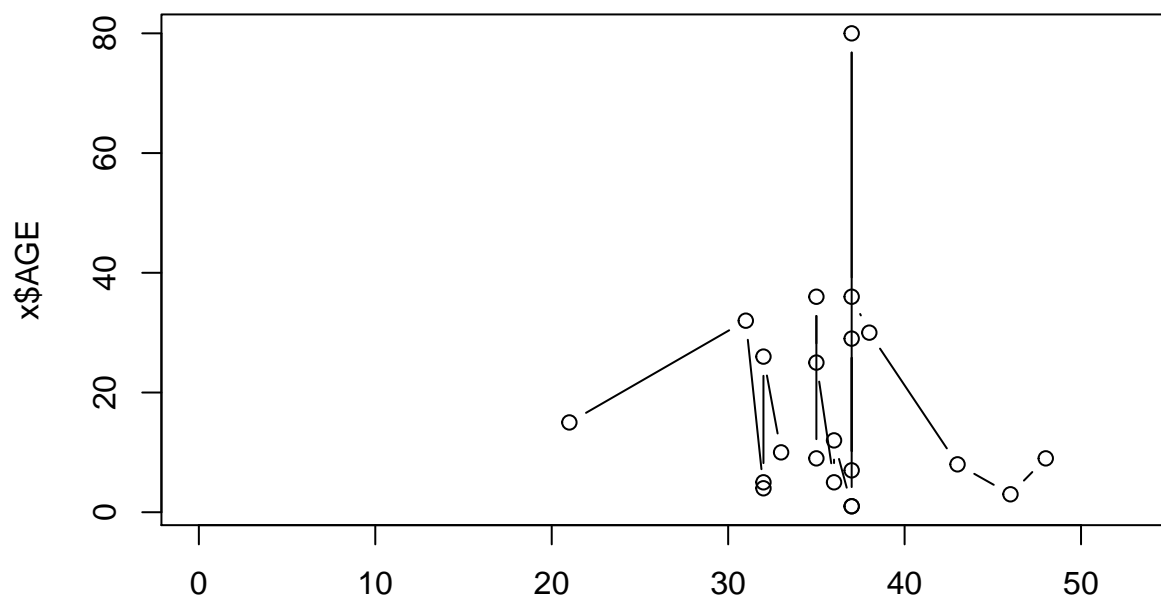
JARDINOPOLIS

IPERO

x$WEEK
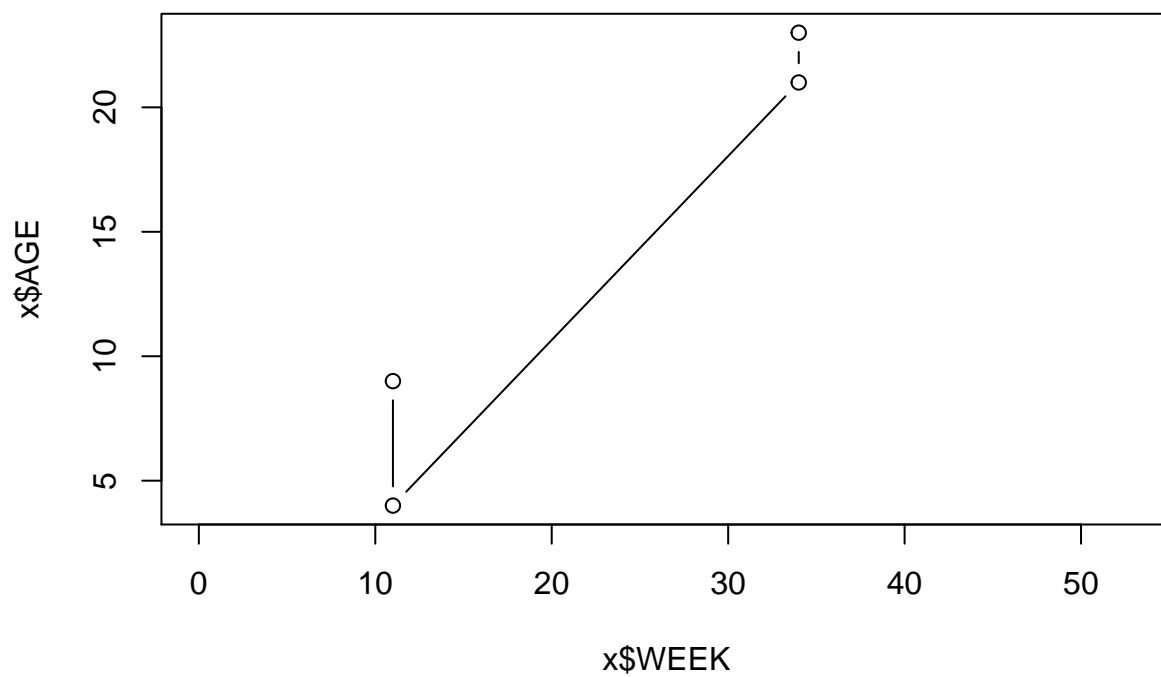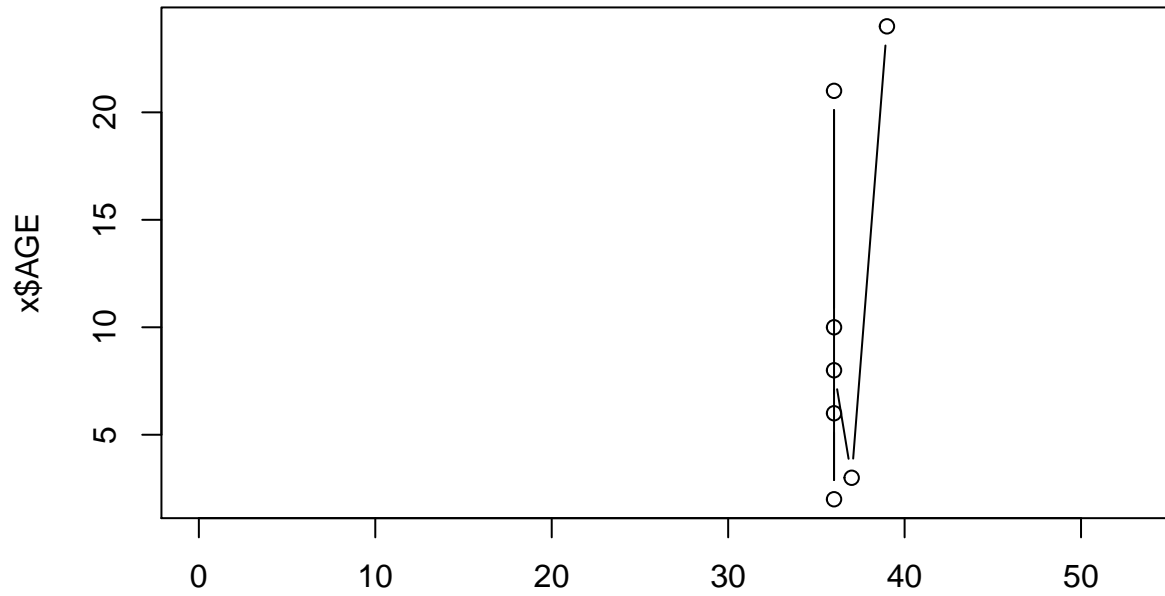
SAO JOSE DOS CAMPOS

x$WEEK
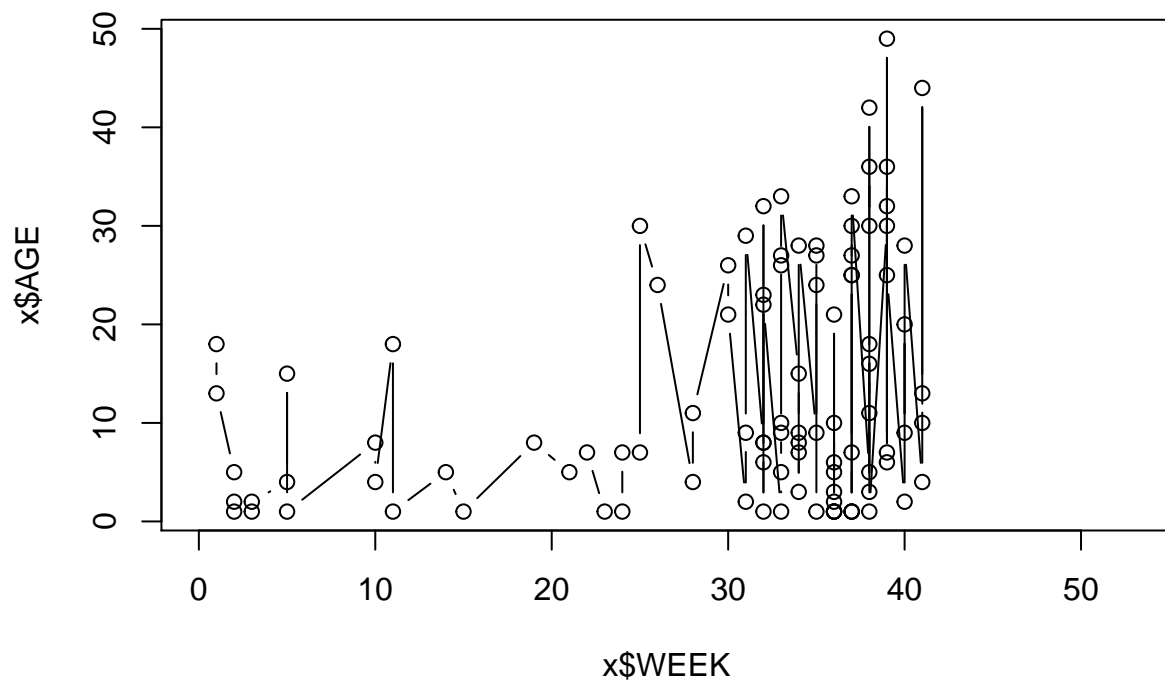
SANTA CRUZ DAS PALMEIRAS

x$WEEK

EMBU DAS ARTES

x$WEEK
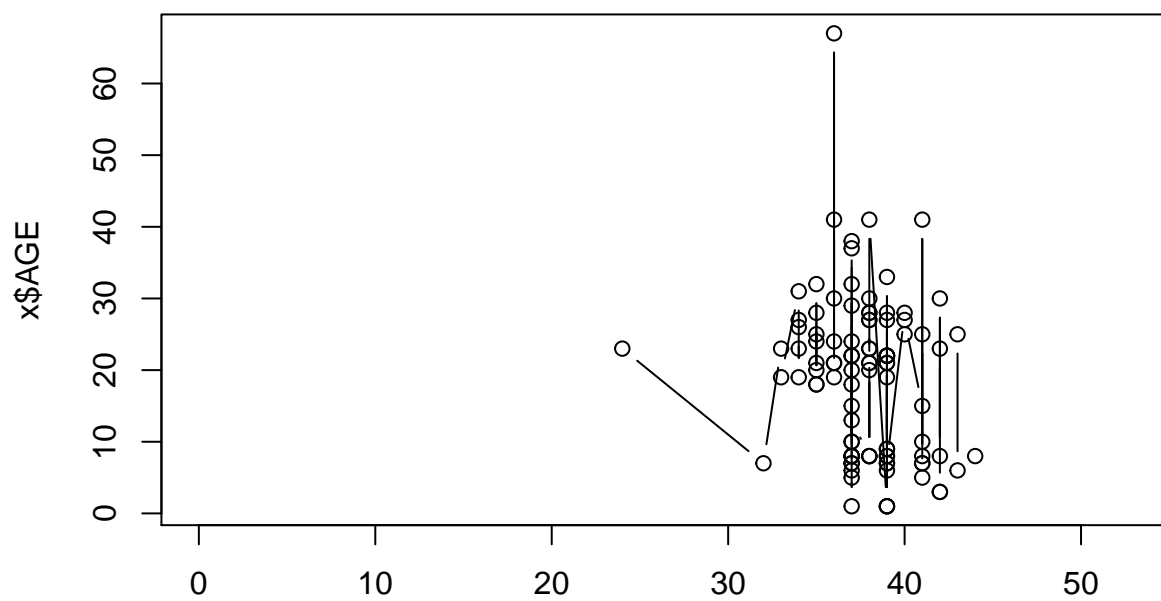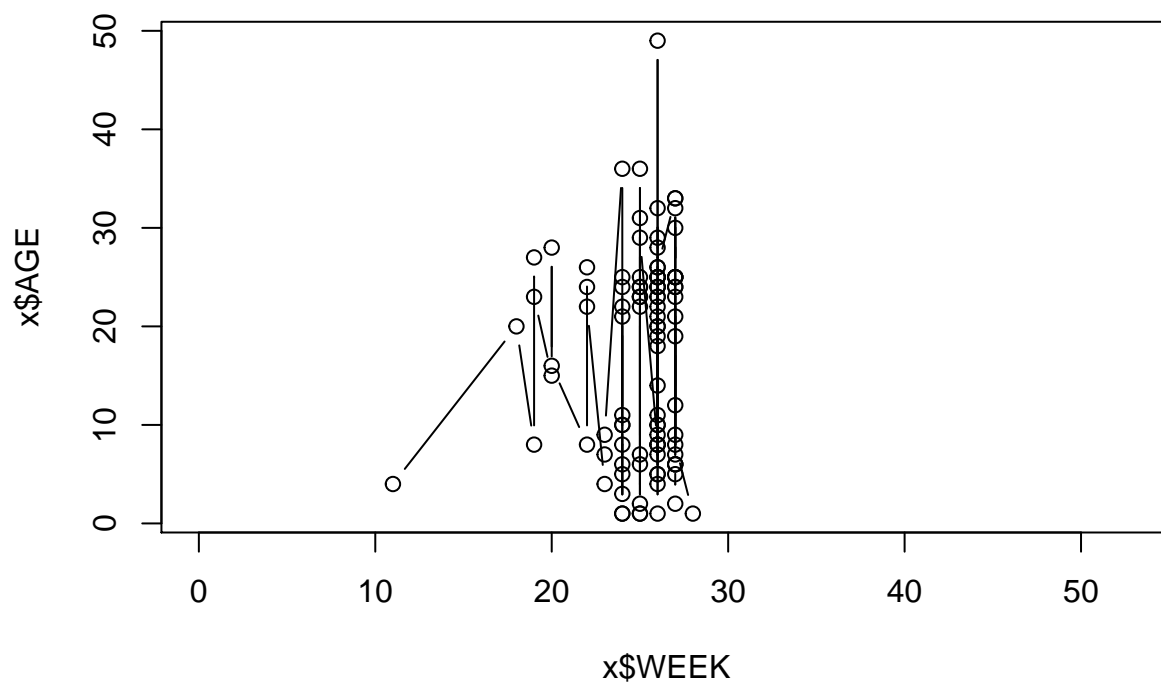
RIO GRANDE DA SERRA

AVARE
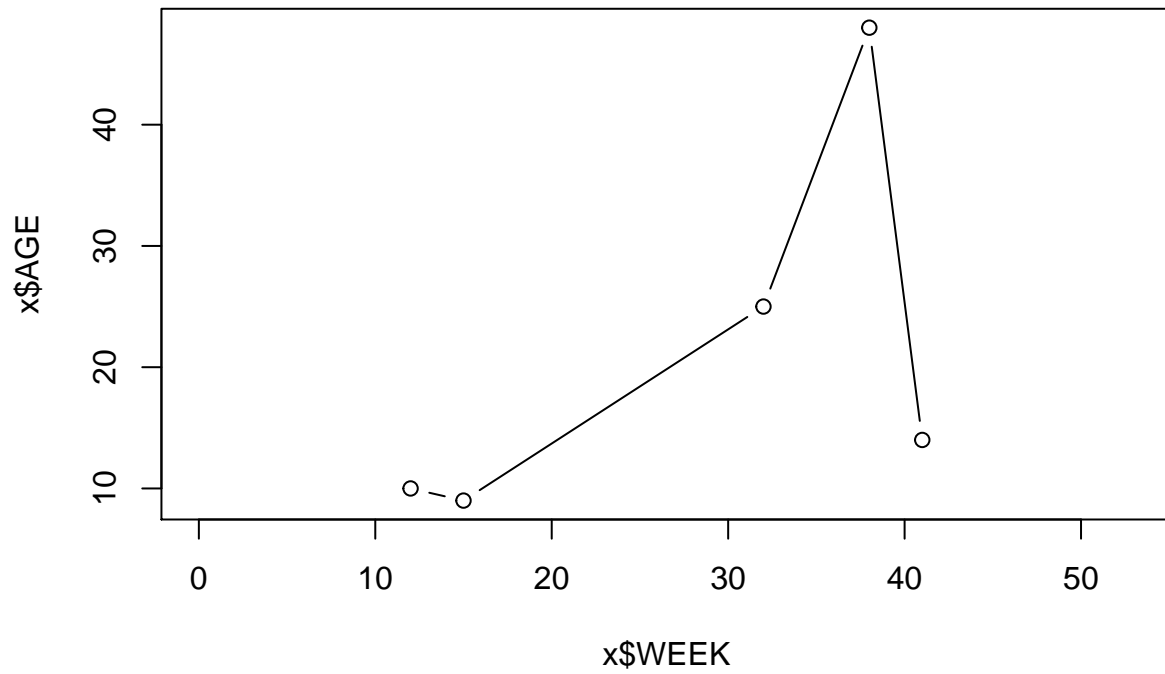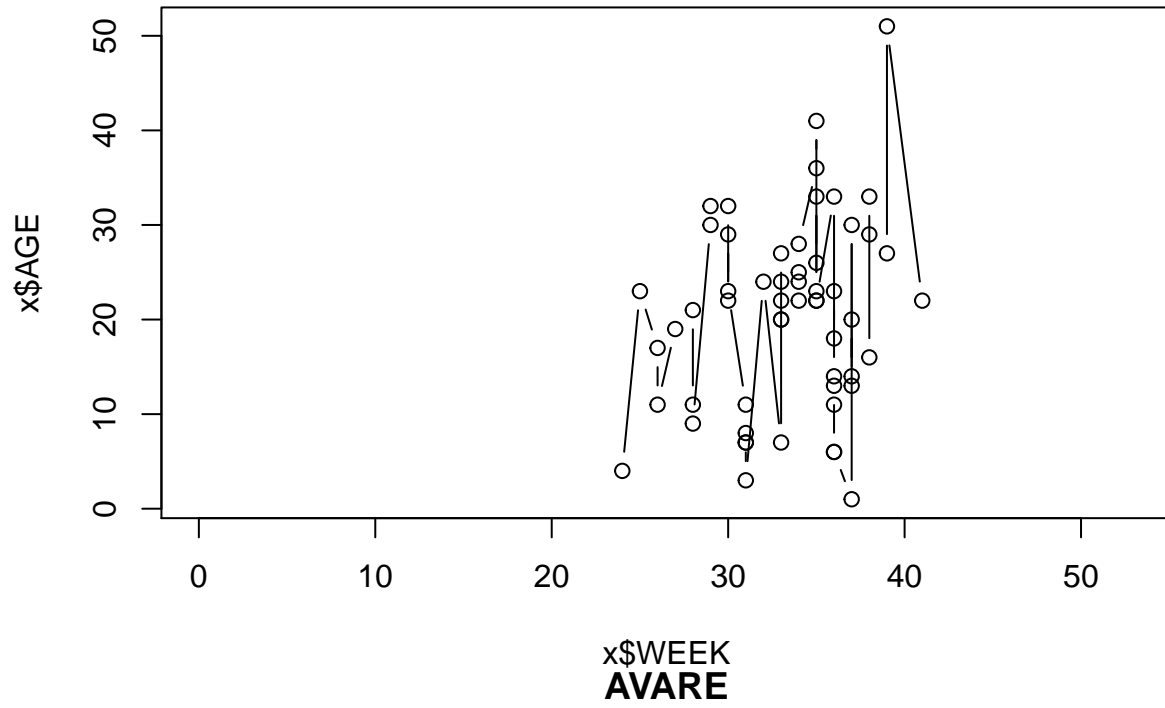
PEDERNEIRAS
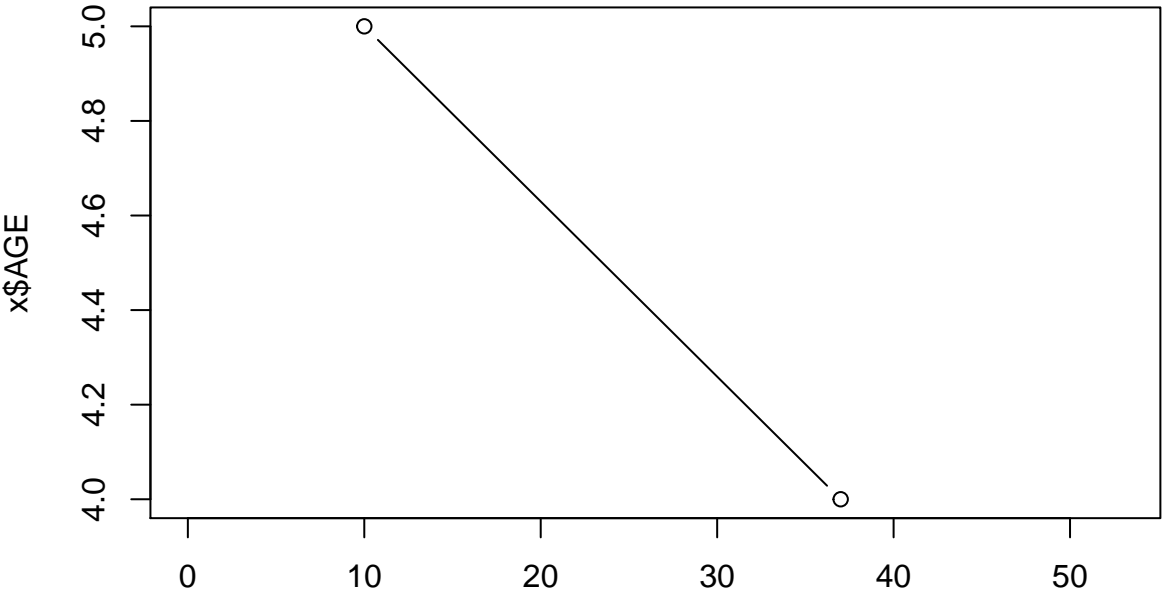
SAO MANUEL
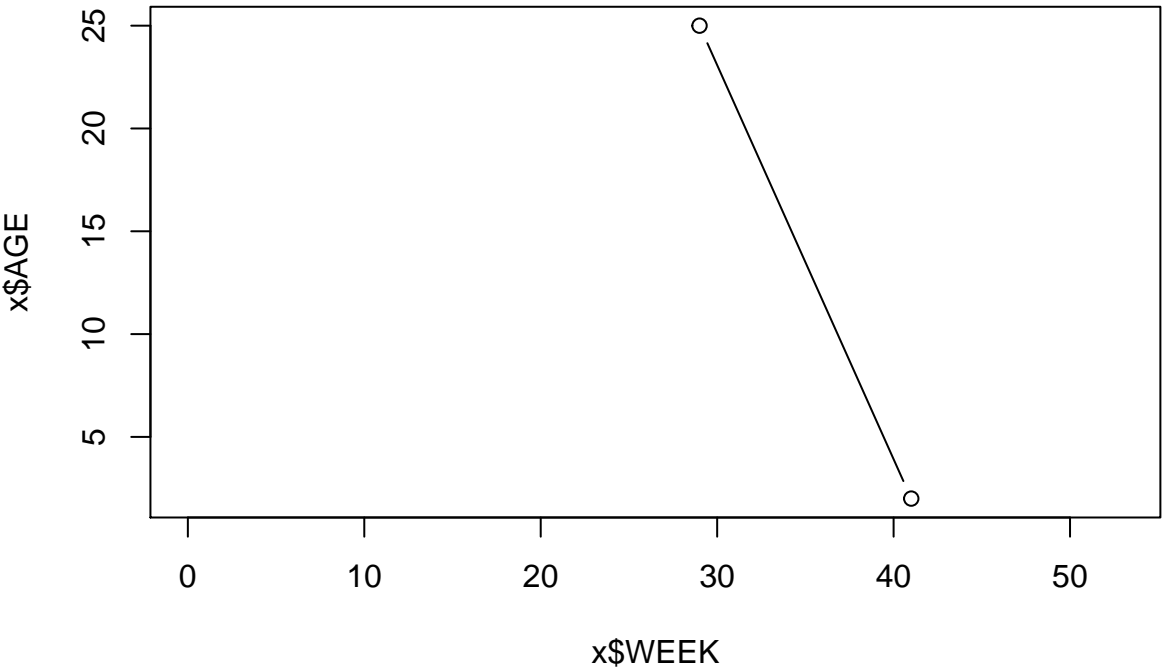
PONTAL
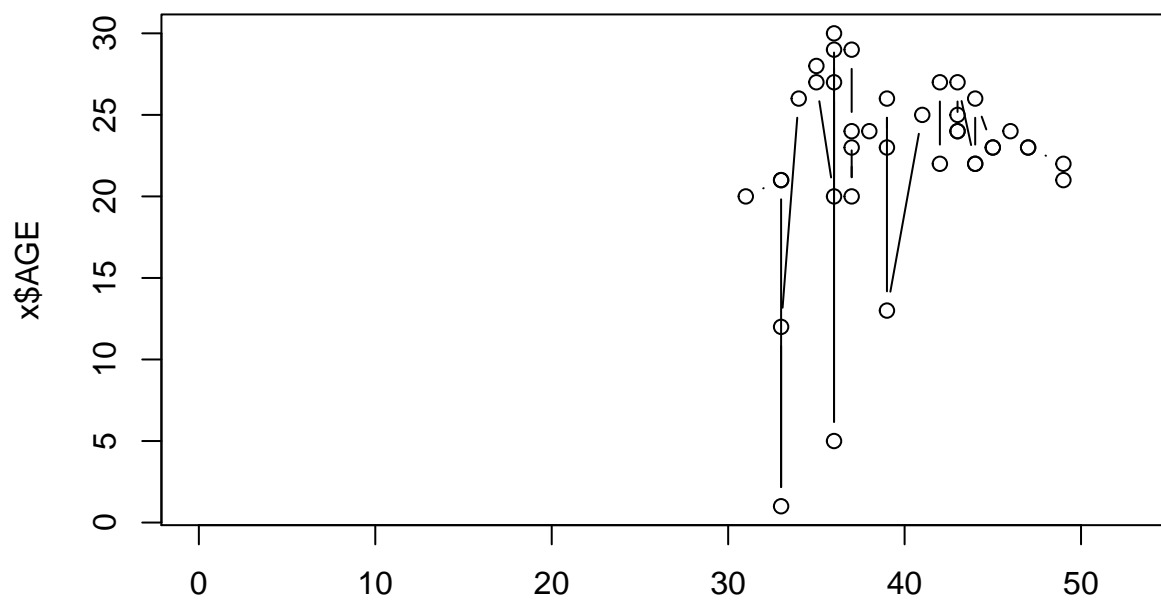
AGUAS DA PRATA
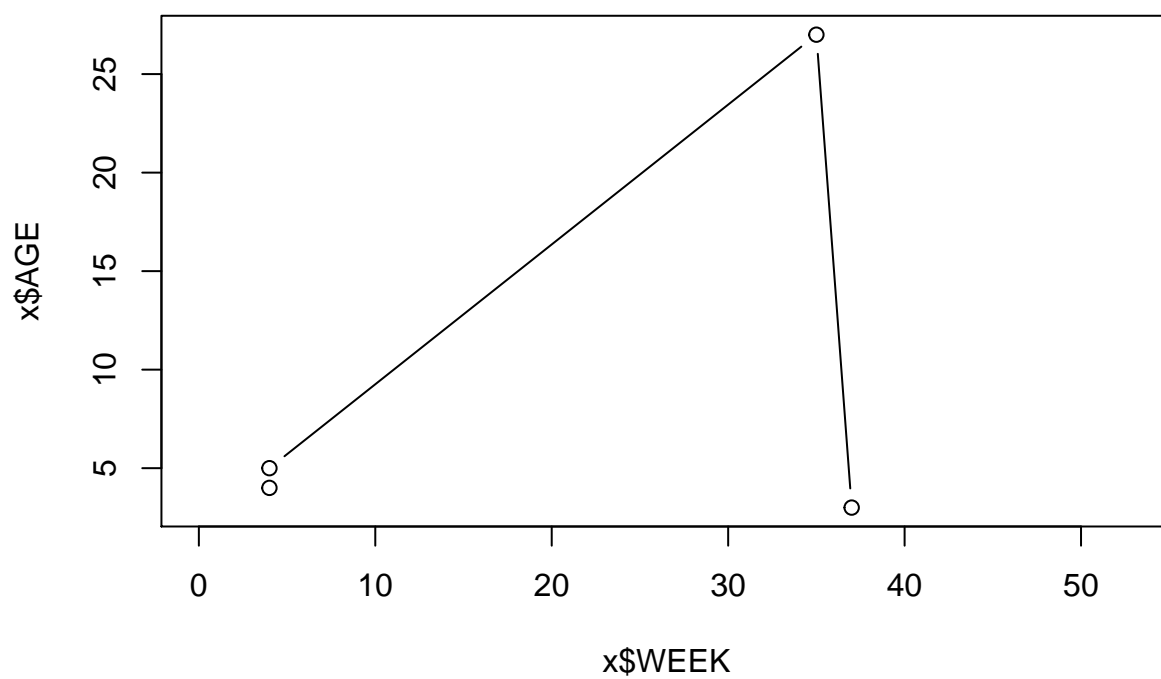
RIBEIRAO PIRES
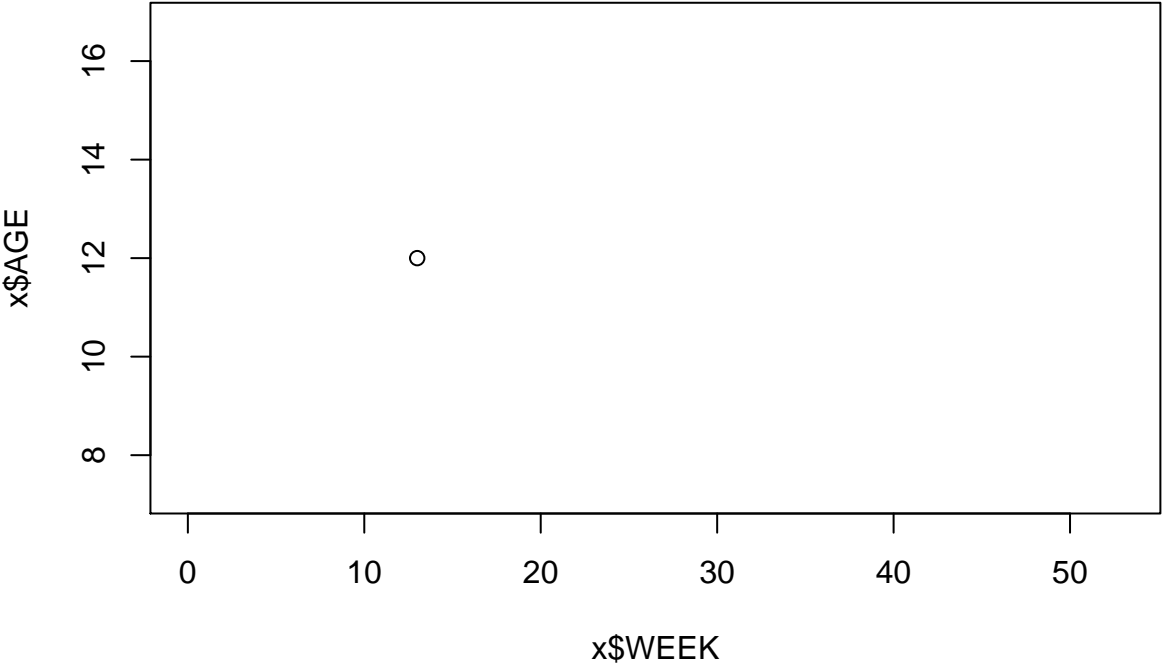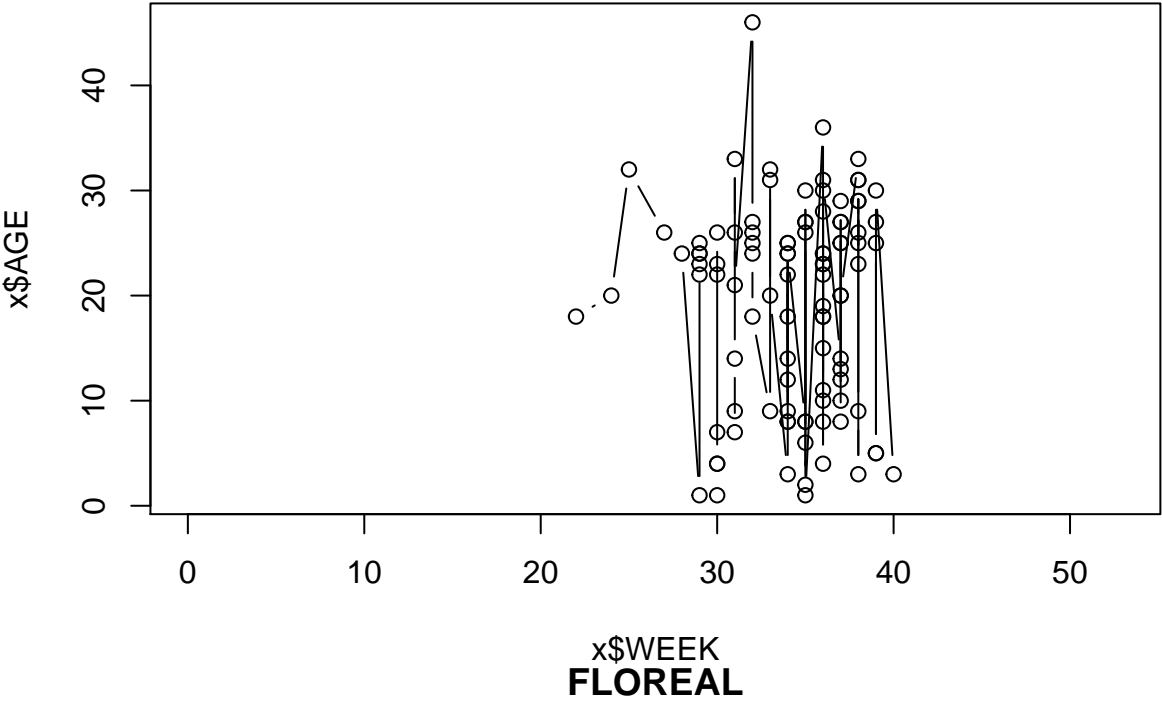
FLOREAL

ITAPECERICA DA SERRA

x$WEEK

SANTANA DE PARNAIBA

x$WEEK

BARUERI

x$WEEK

FRANCISCO MORATO

x$WEEK

49

CAIEIRAS

x$WEEK

COLINA

x$WEEK
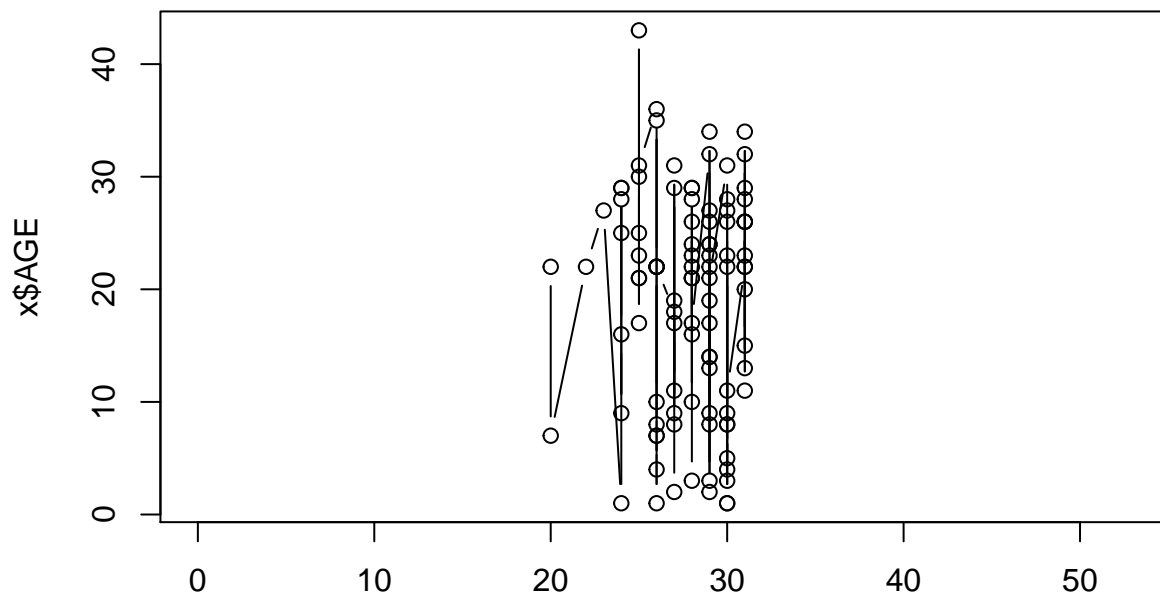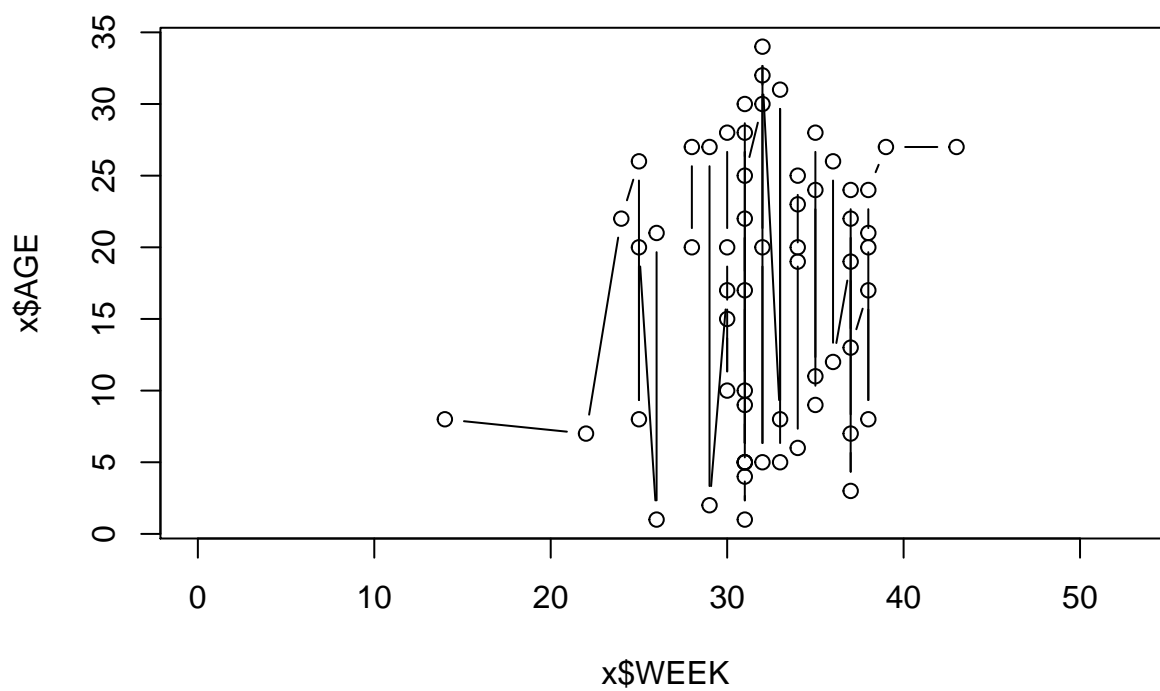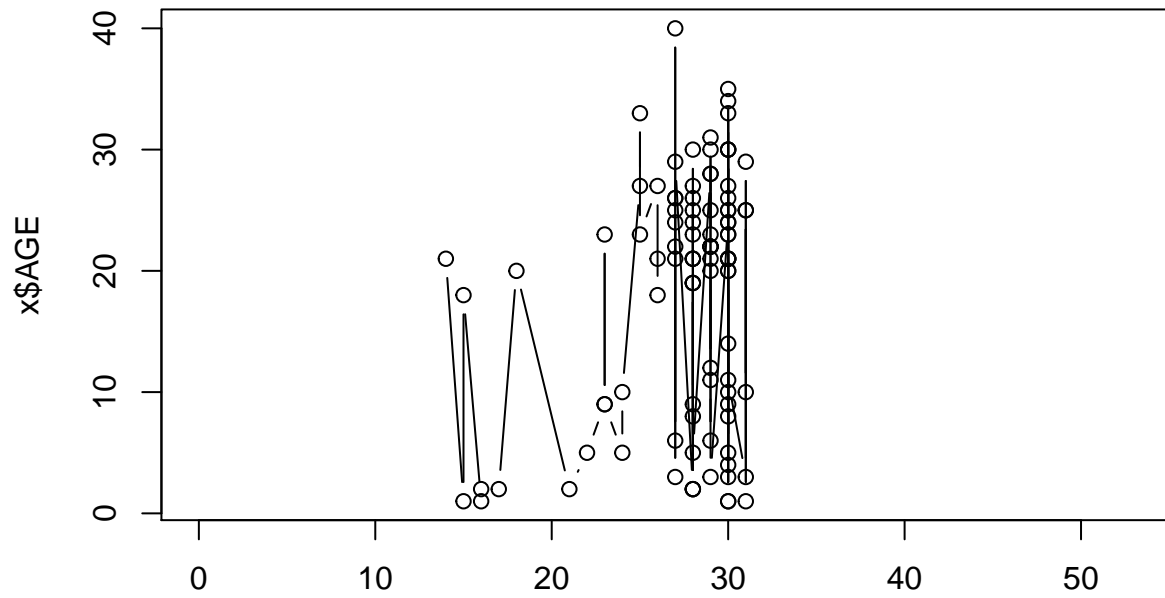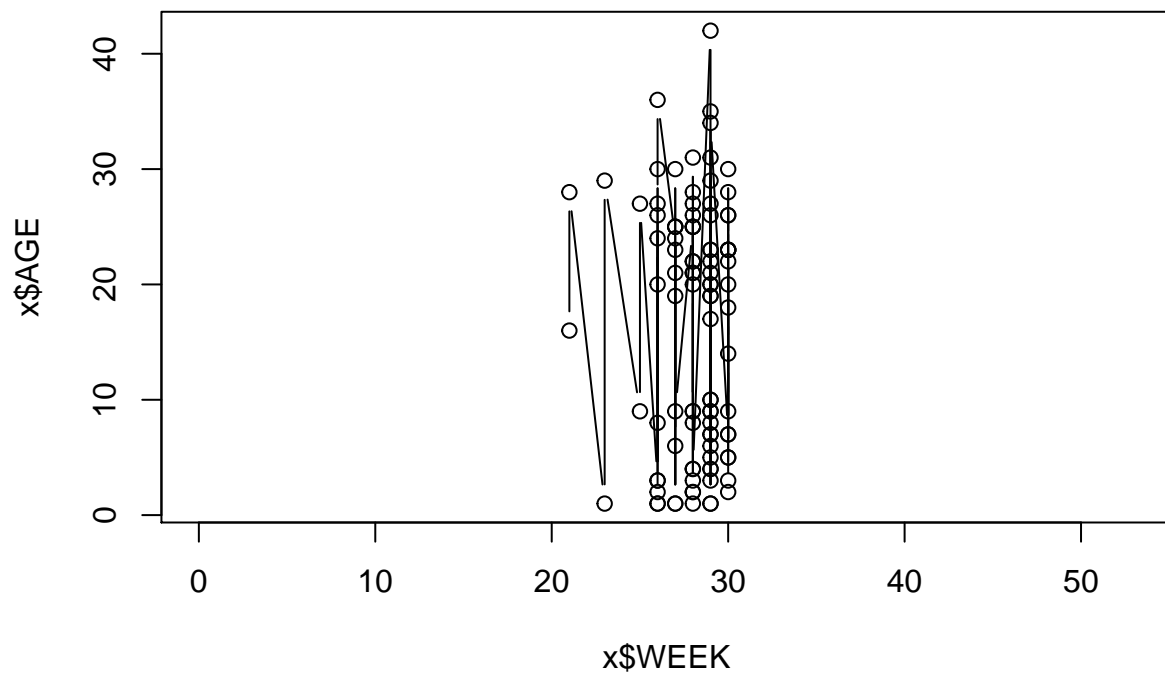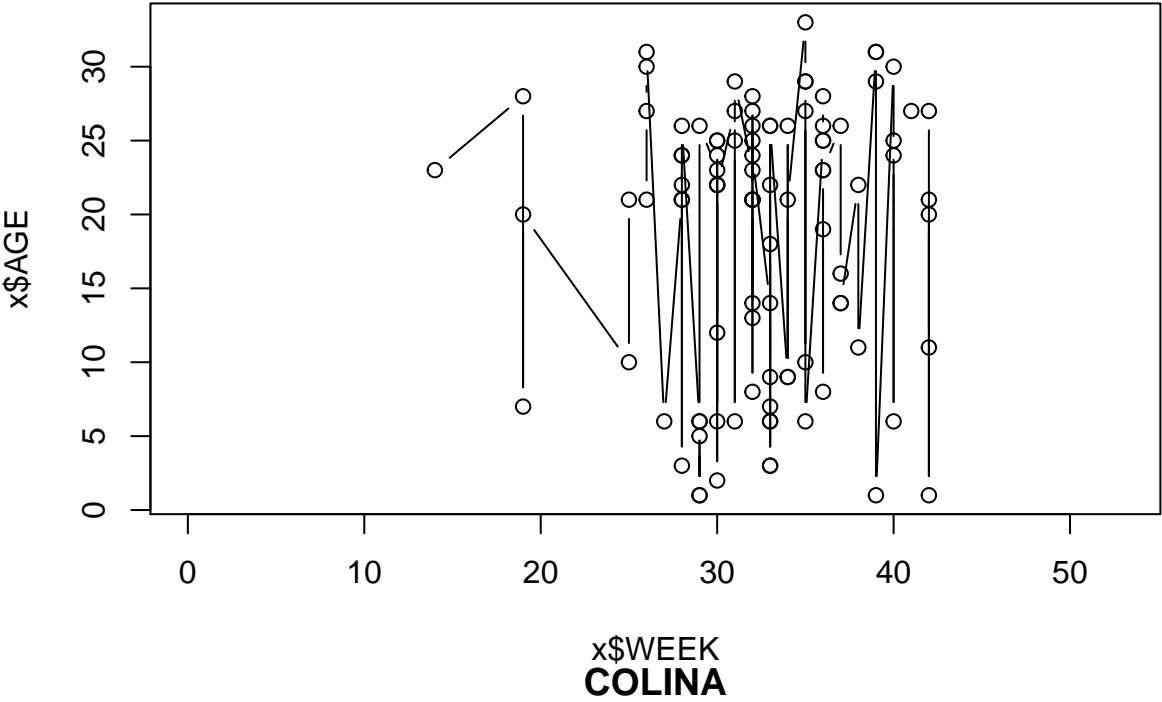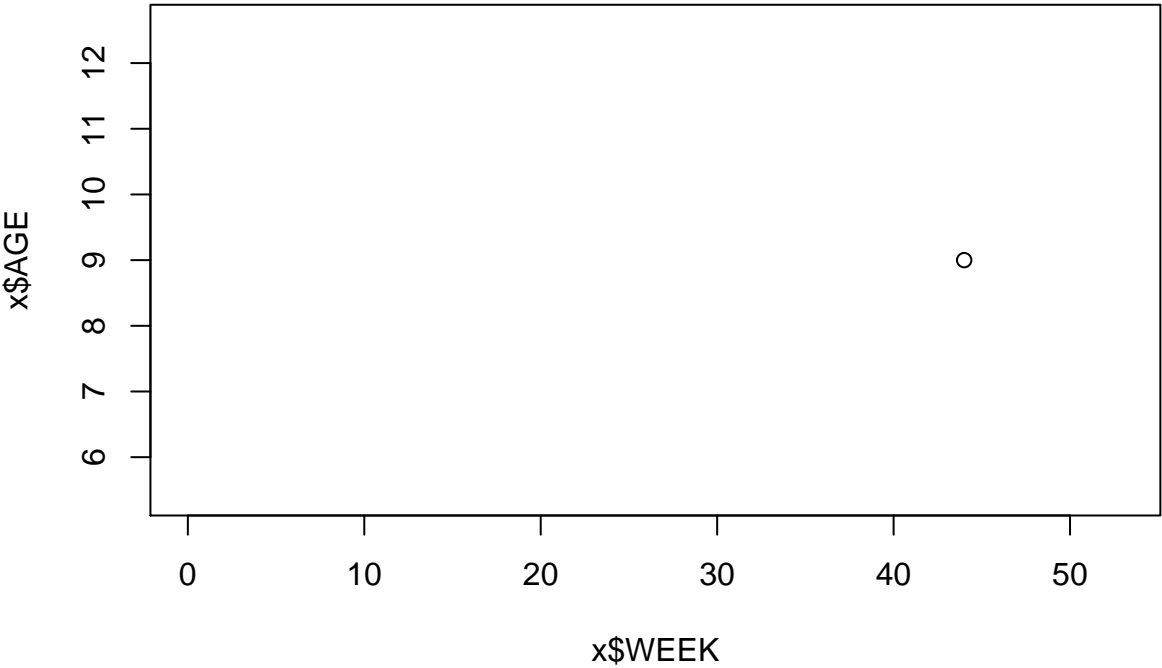
CAMPOS NOVOS PAULISTA
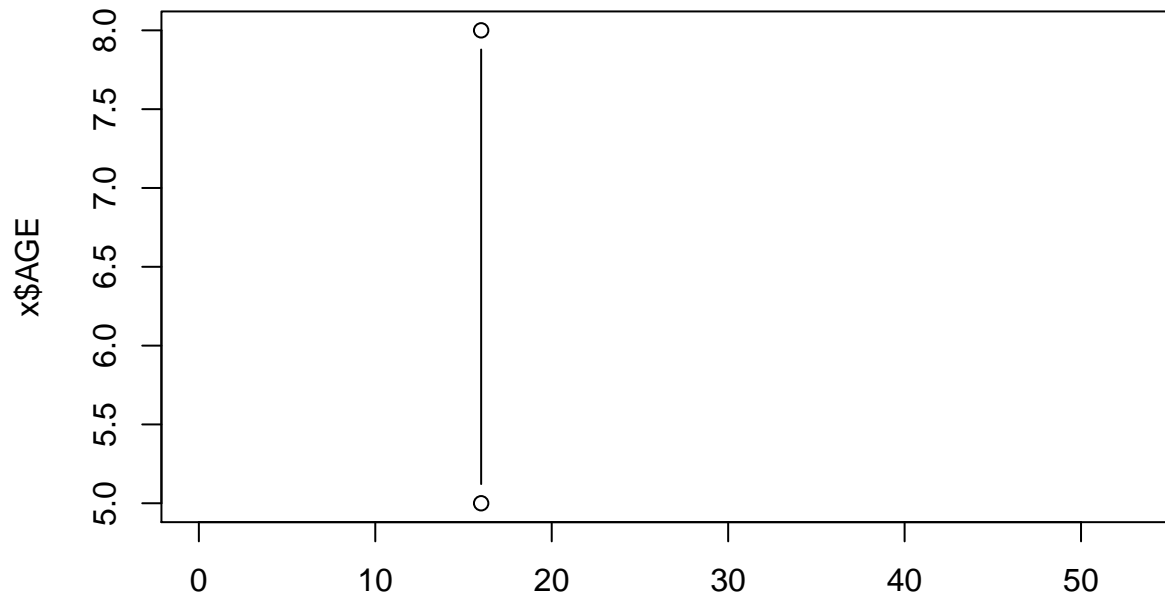
BARRA DO TURVO
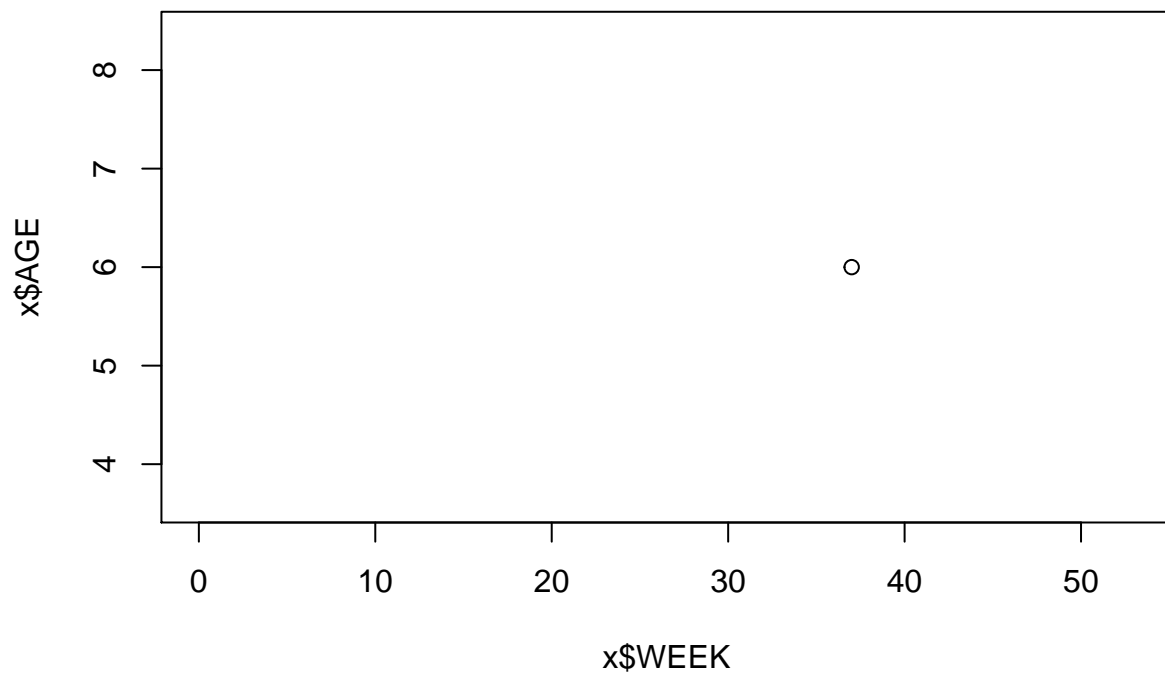
PRESIDENTE ALVES
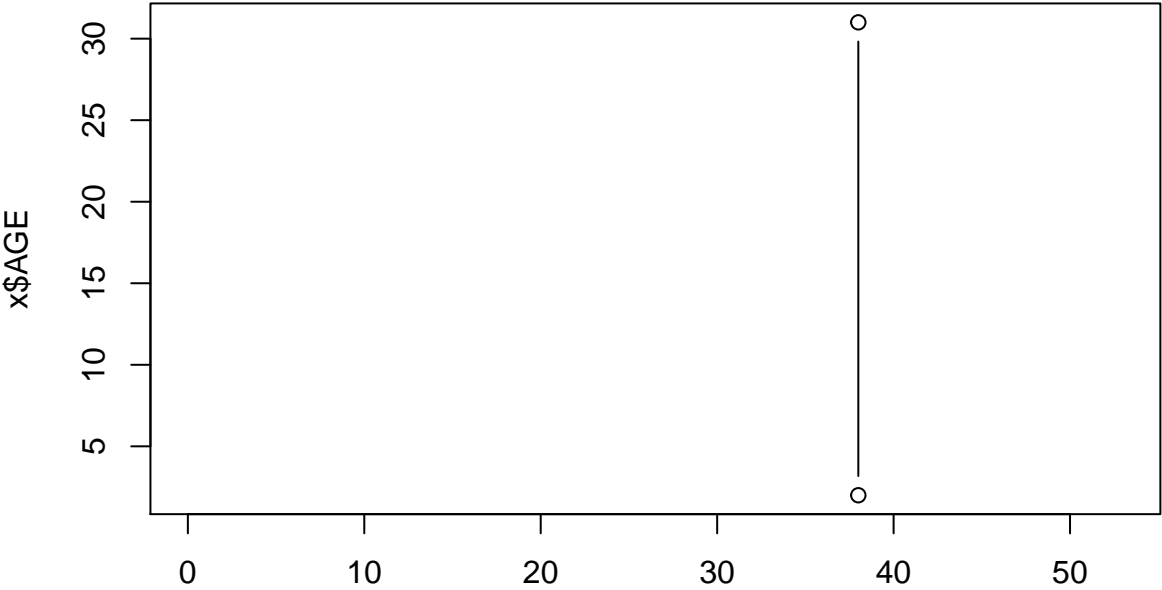
GUARARAPES
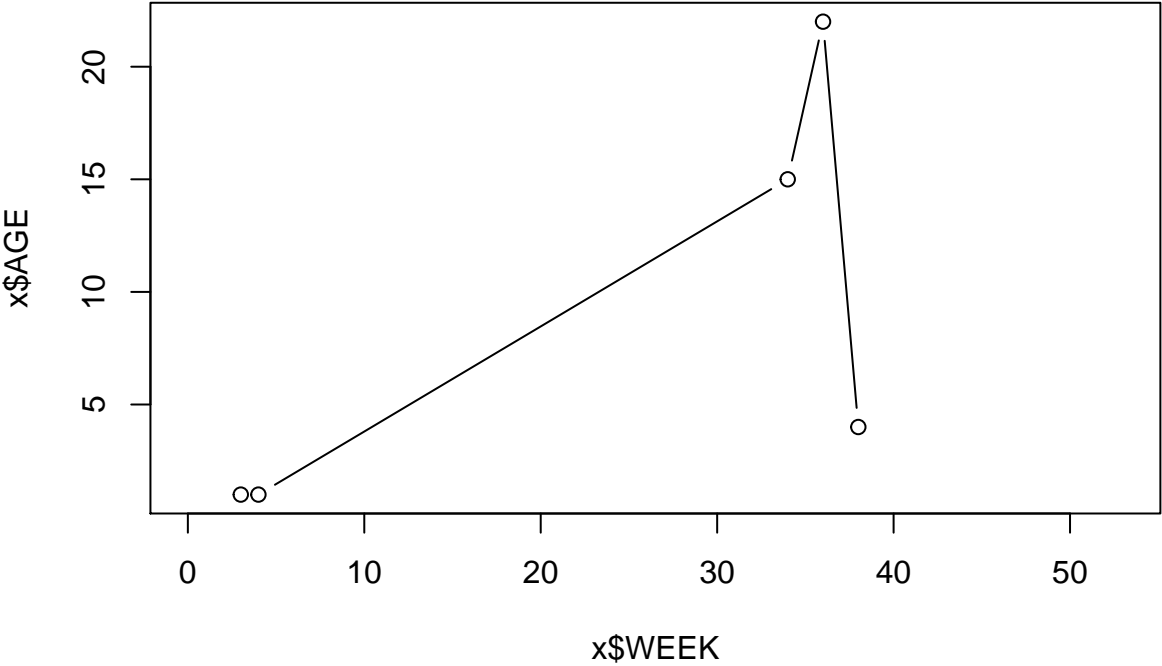
ITAPIRA

BIRIGUI

JACAREI

x$WEEK

OSCAR BRESSANE

x$WEEK
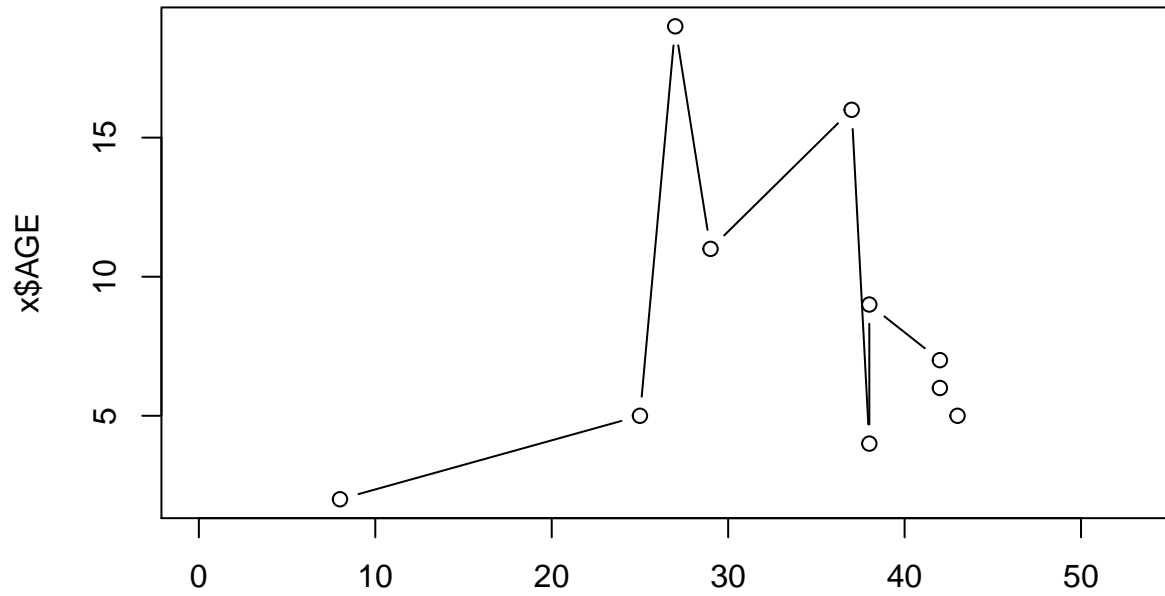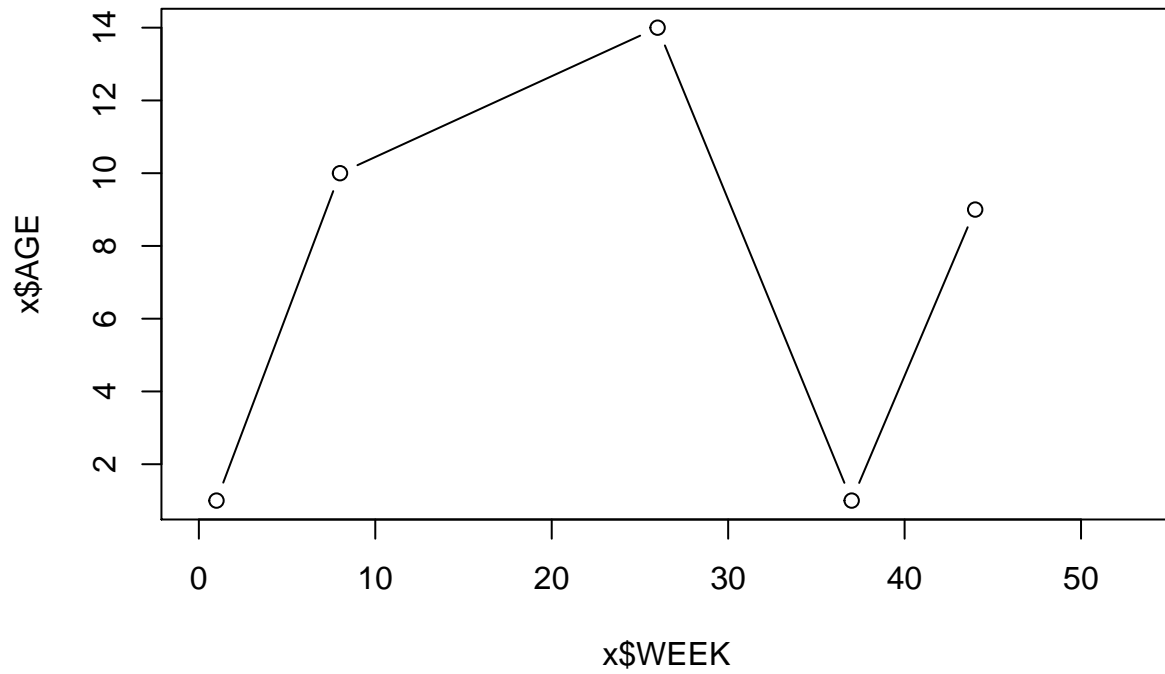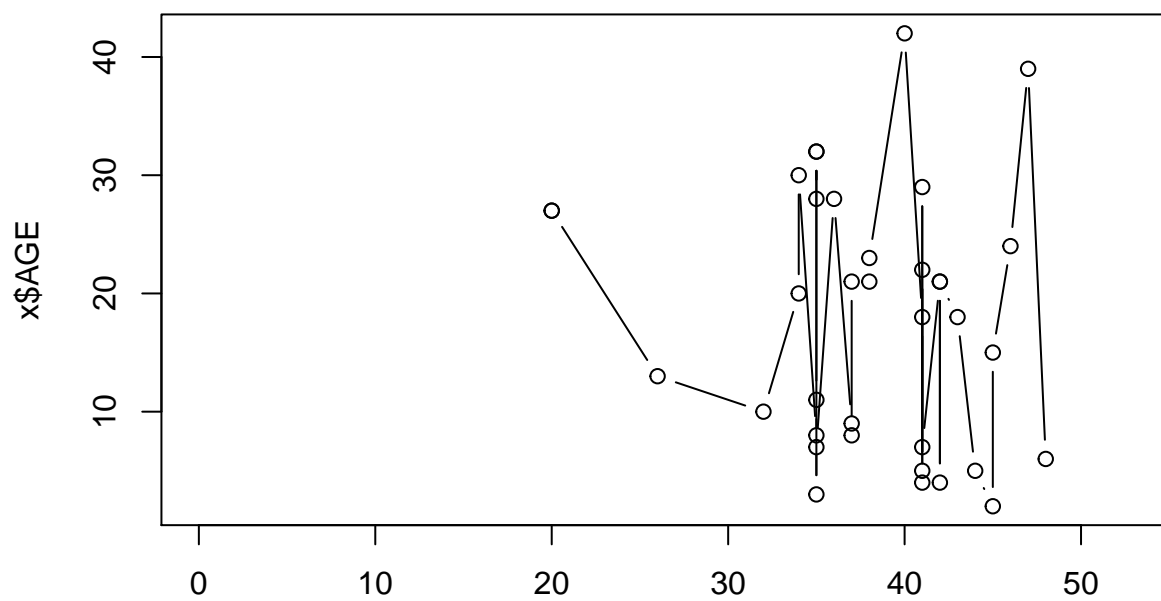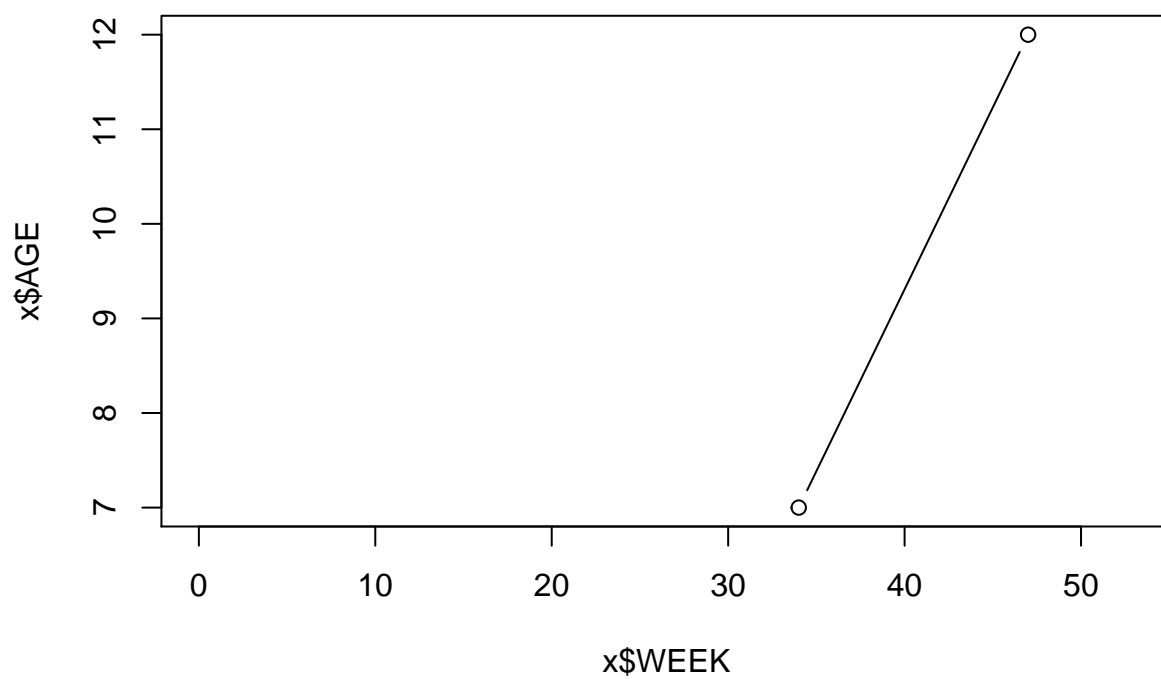
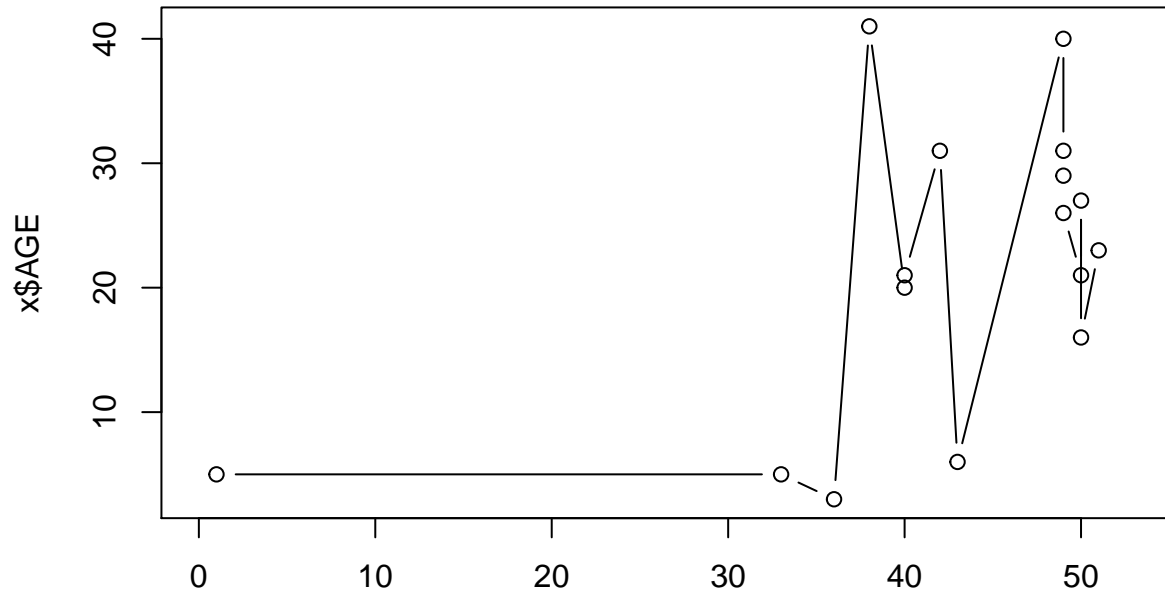**AURIFLAMA**
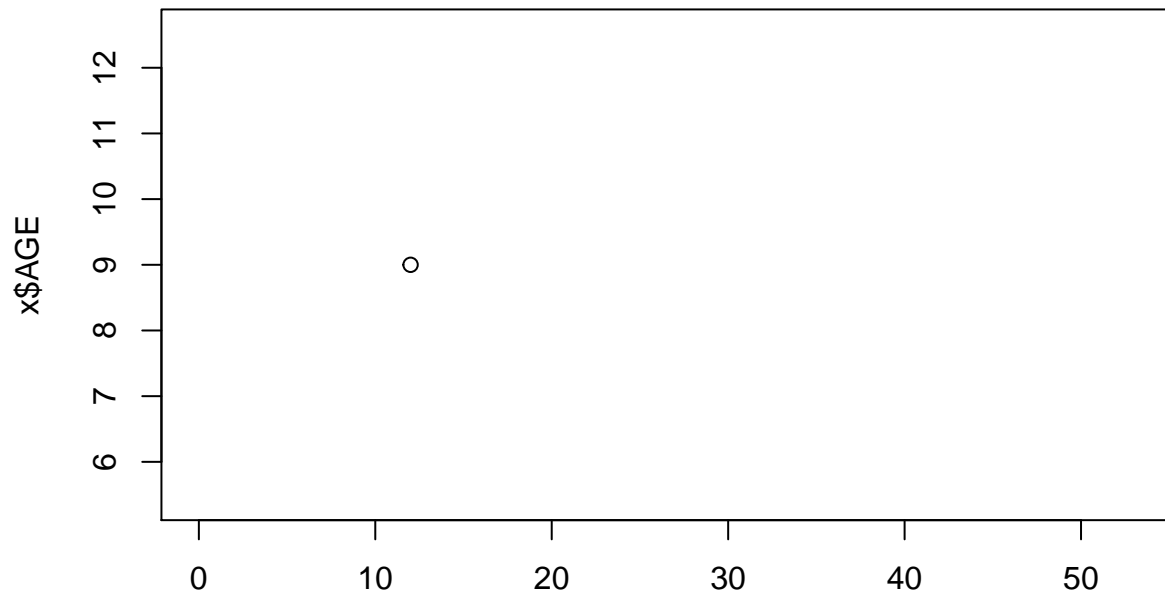
GABRIEL MONTEIRO
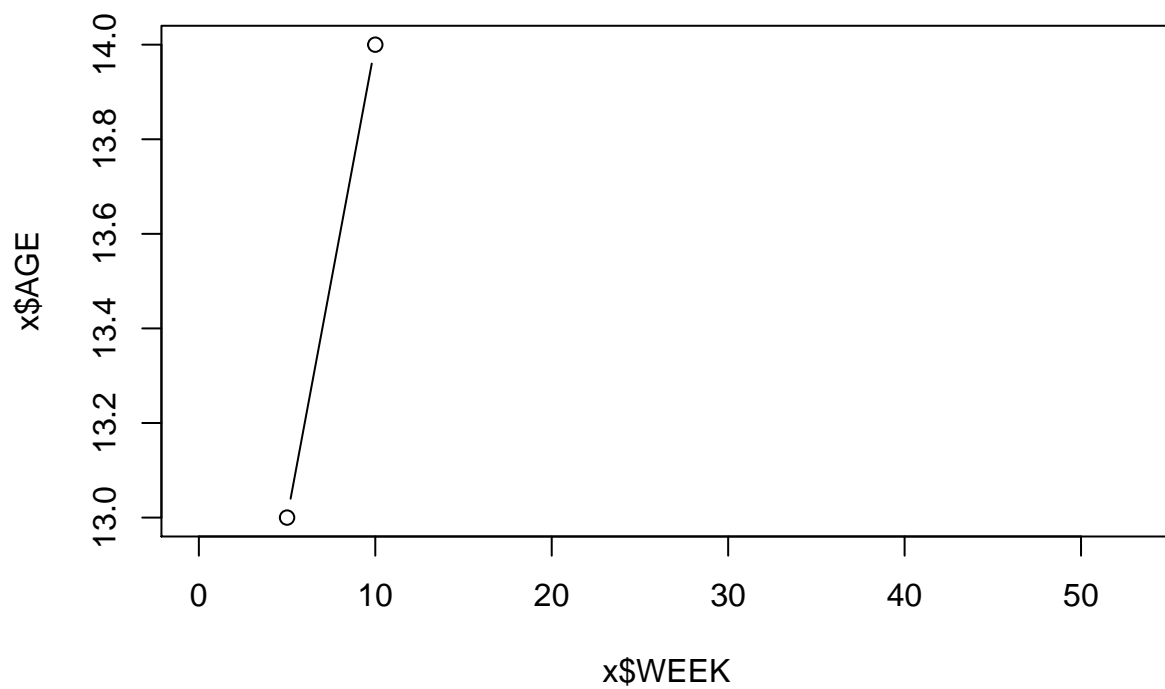
DESCALVADO
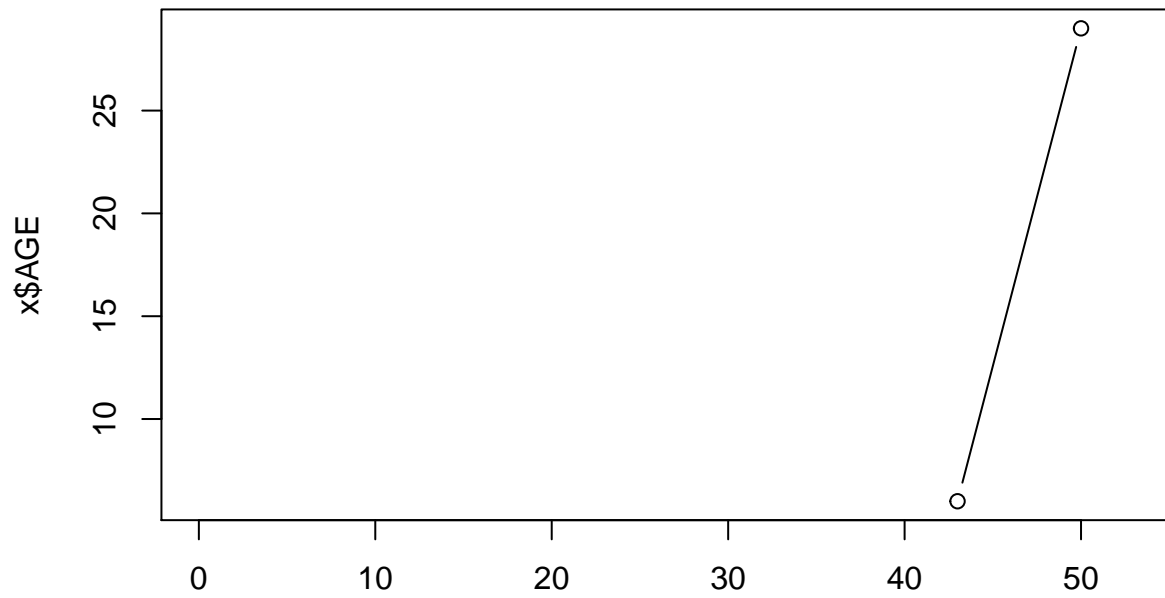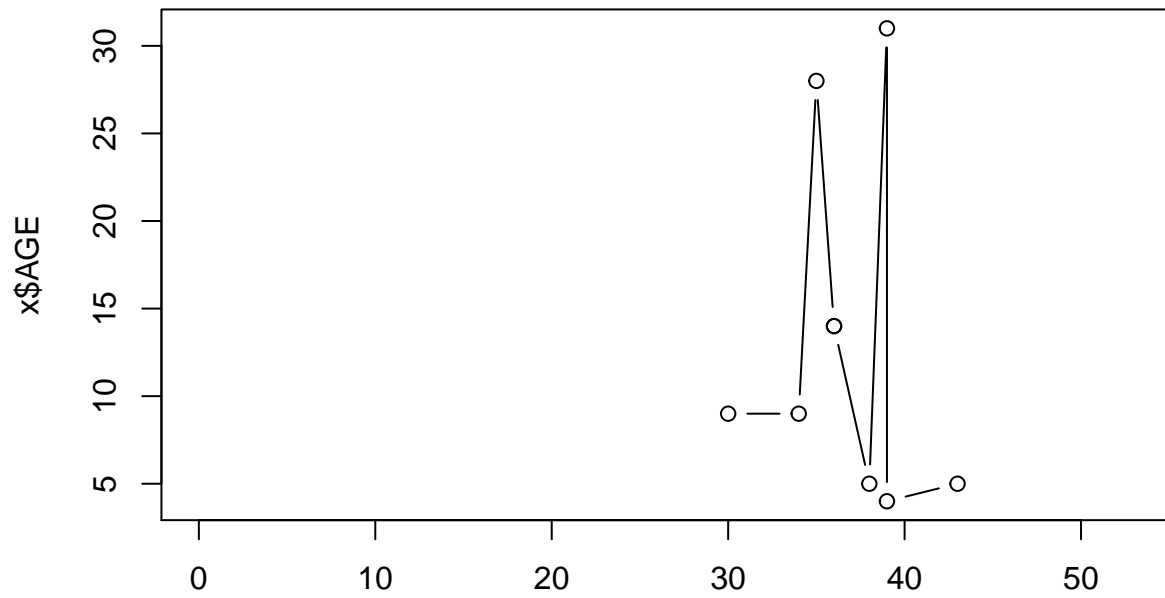

APARECIDA
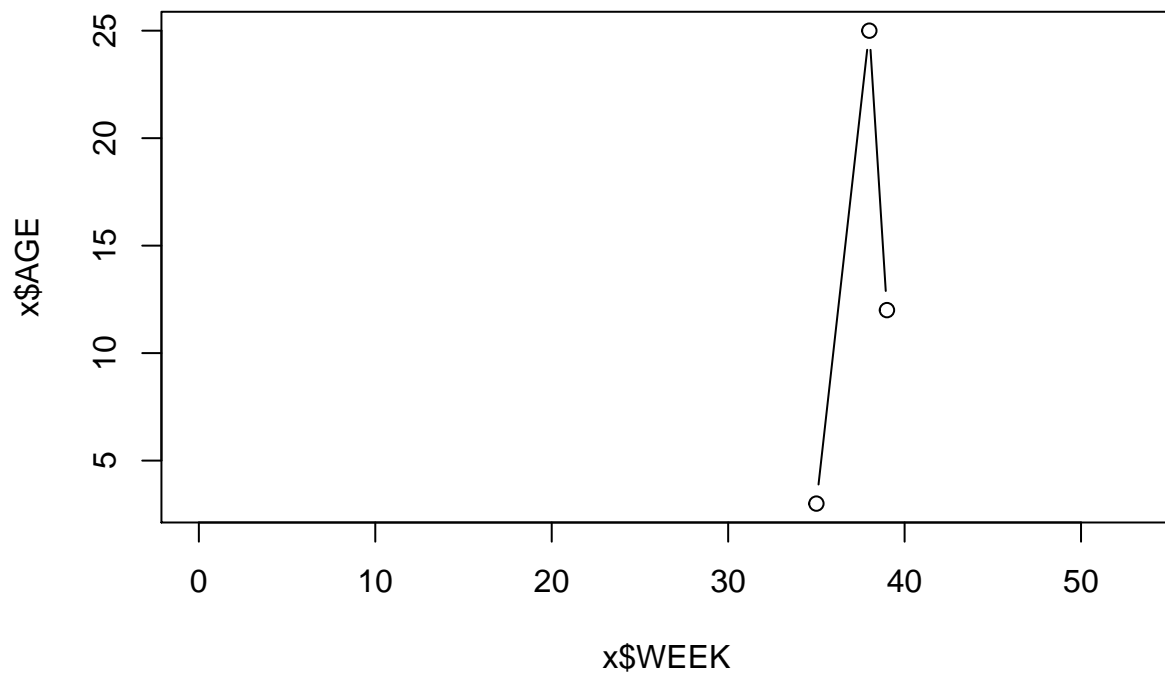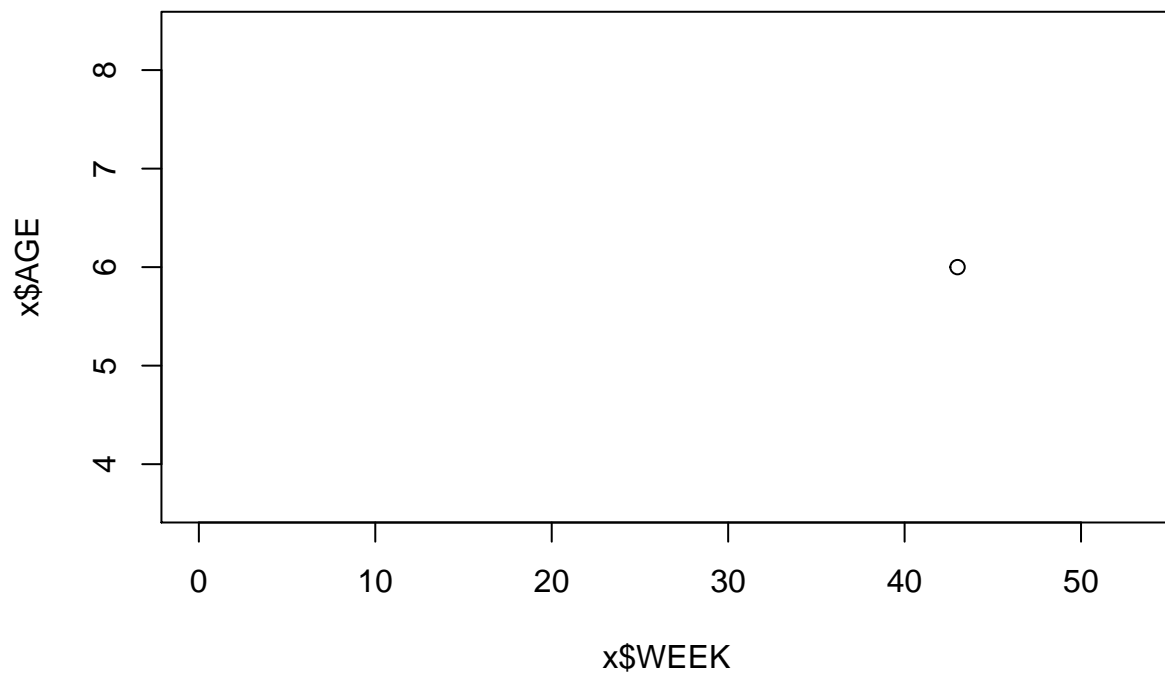
GUZOLANDIA
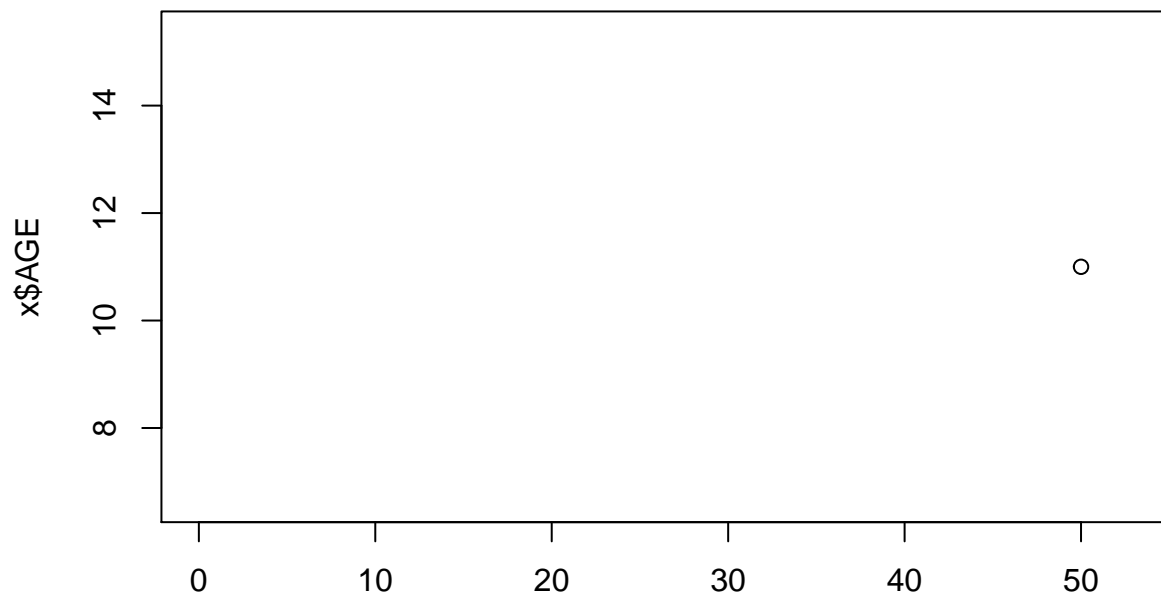
BEBEDOURO
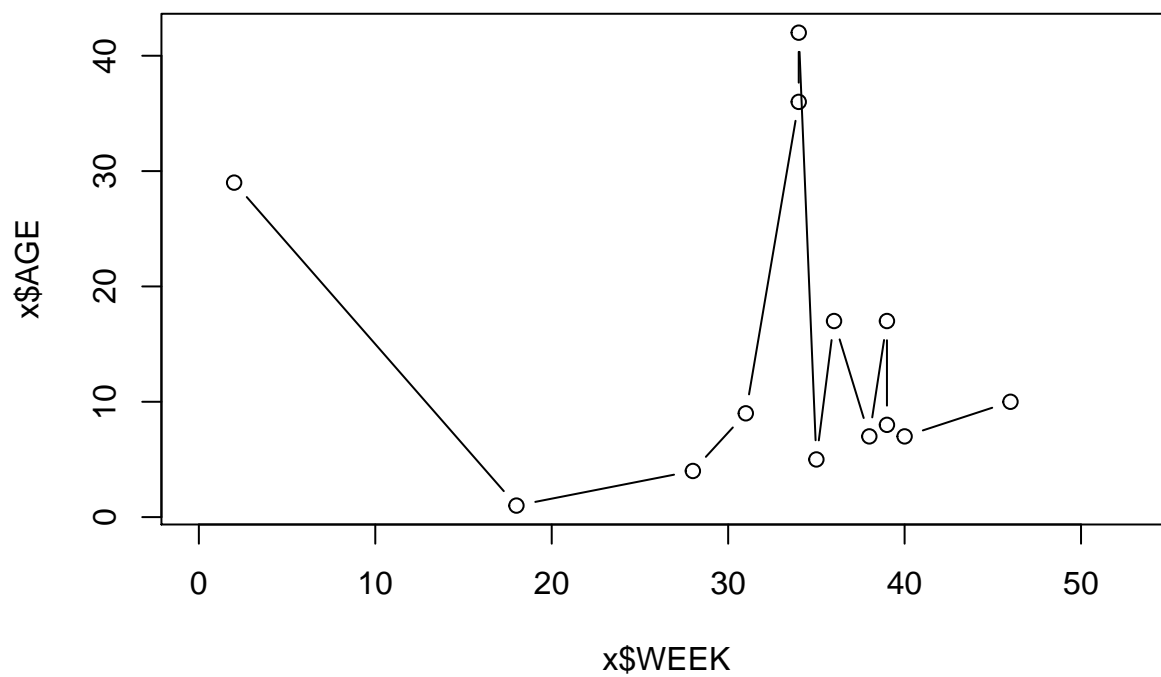
SANTA CRUZ DO RIO PARDO

ELDORADO
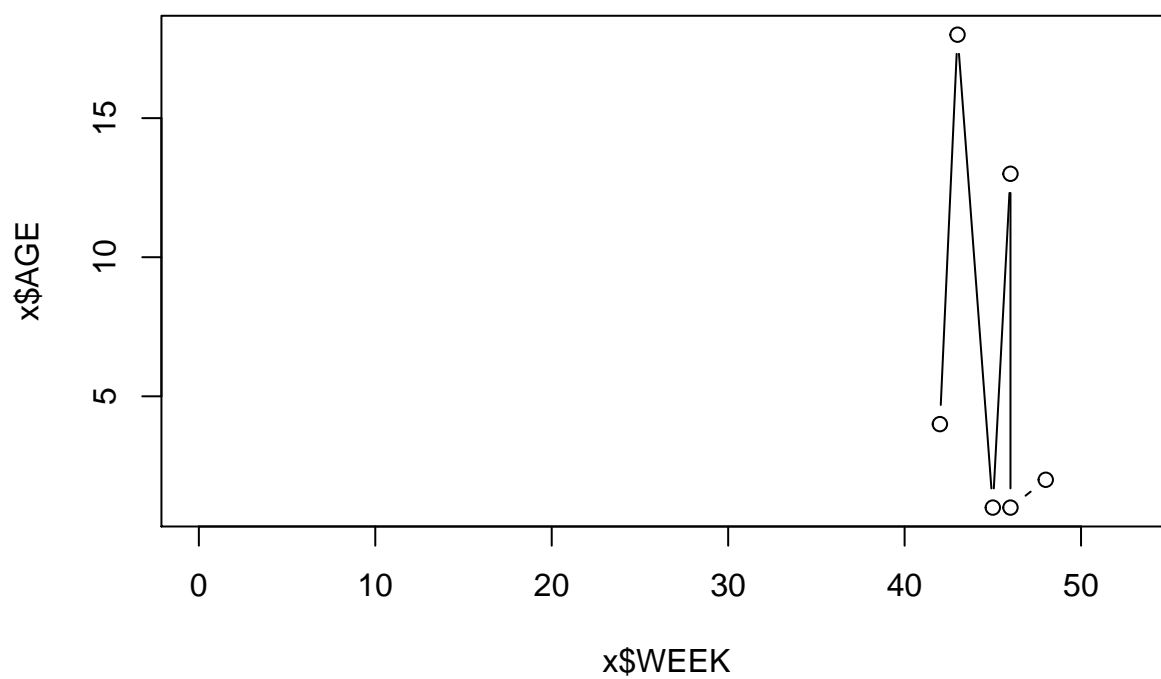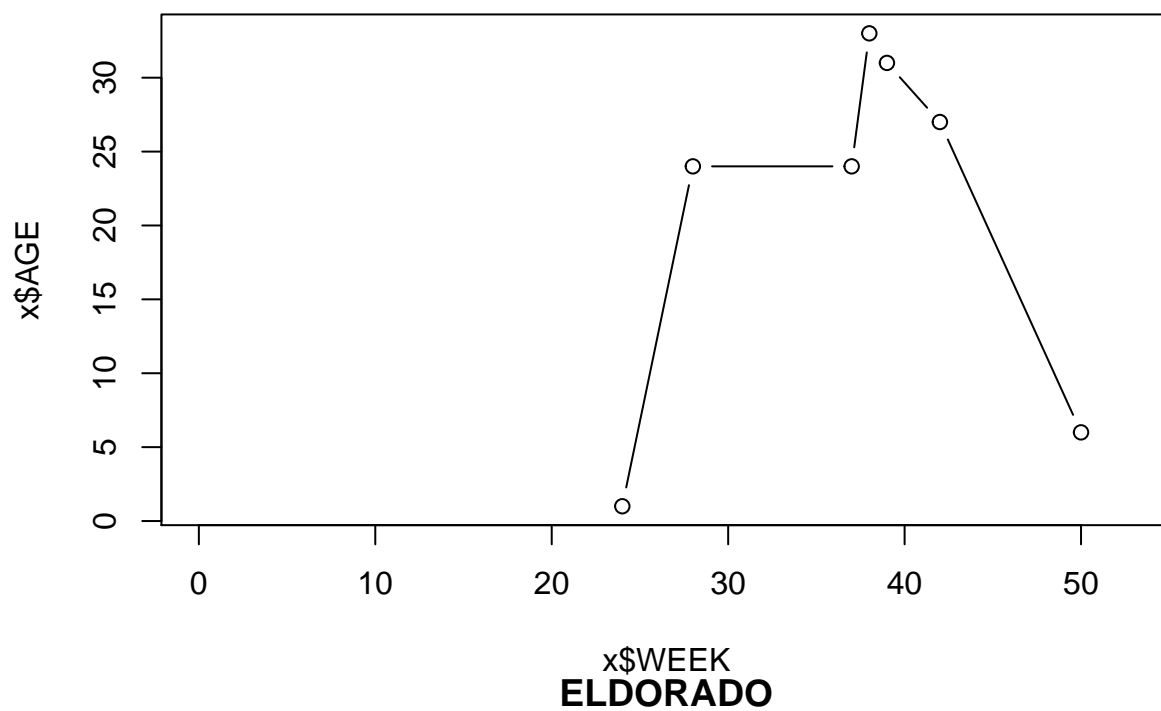
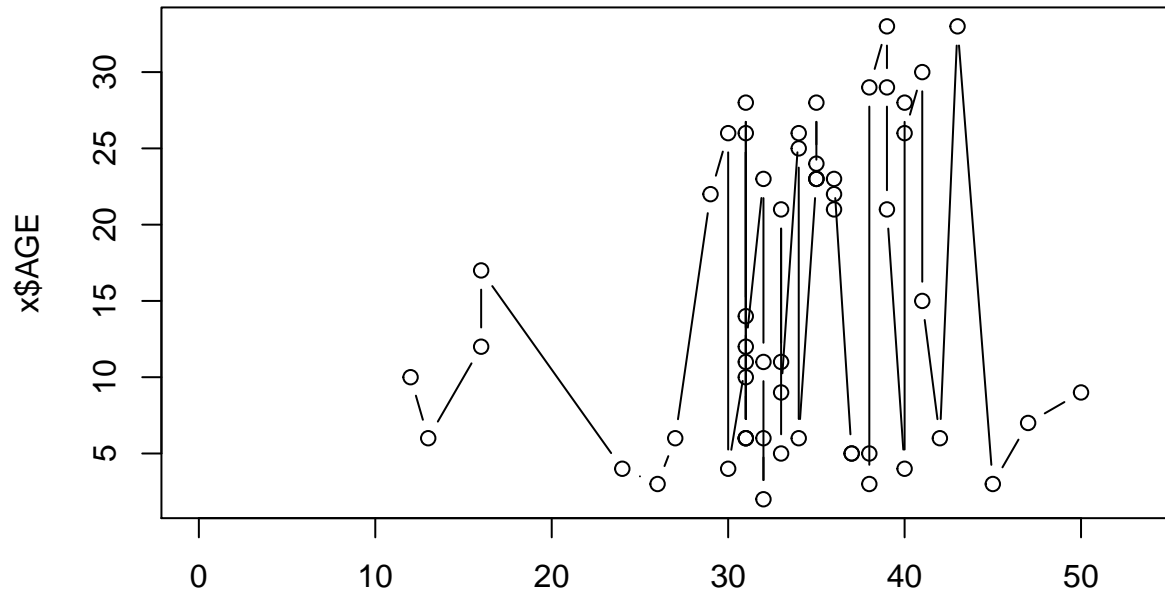**LEME**

COROADOS

CAJATI

x$WEEK
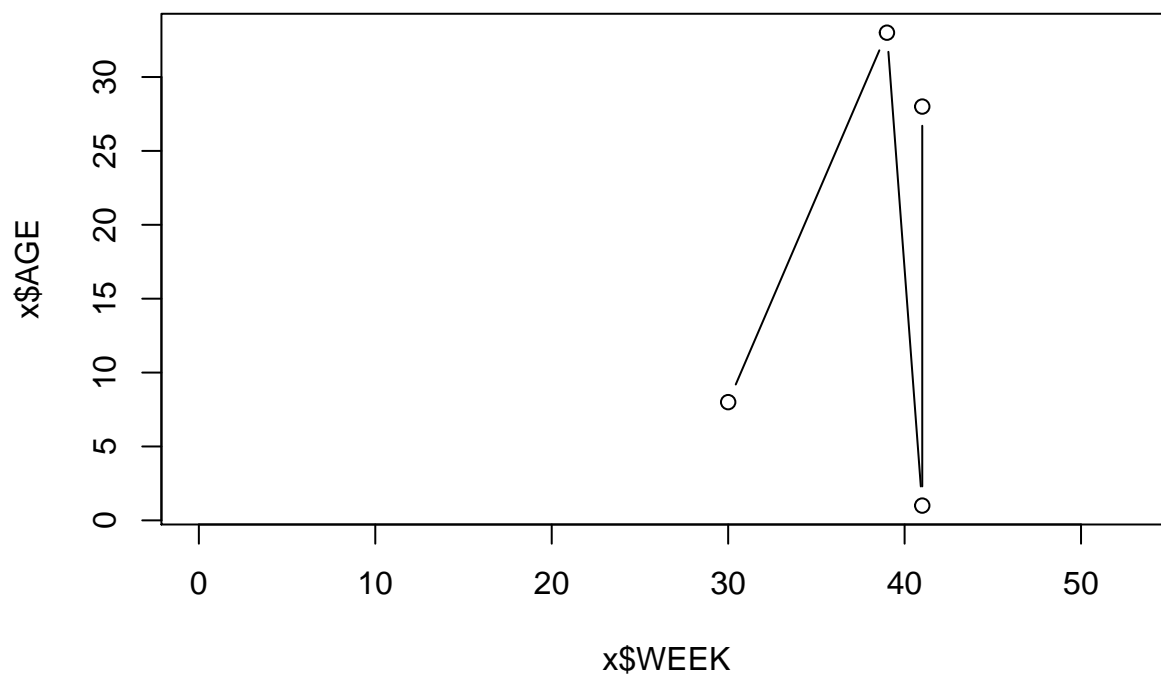
ESTRELA D'OESTE
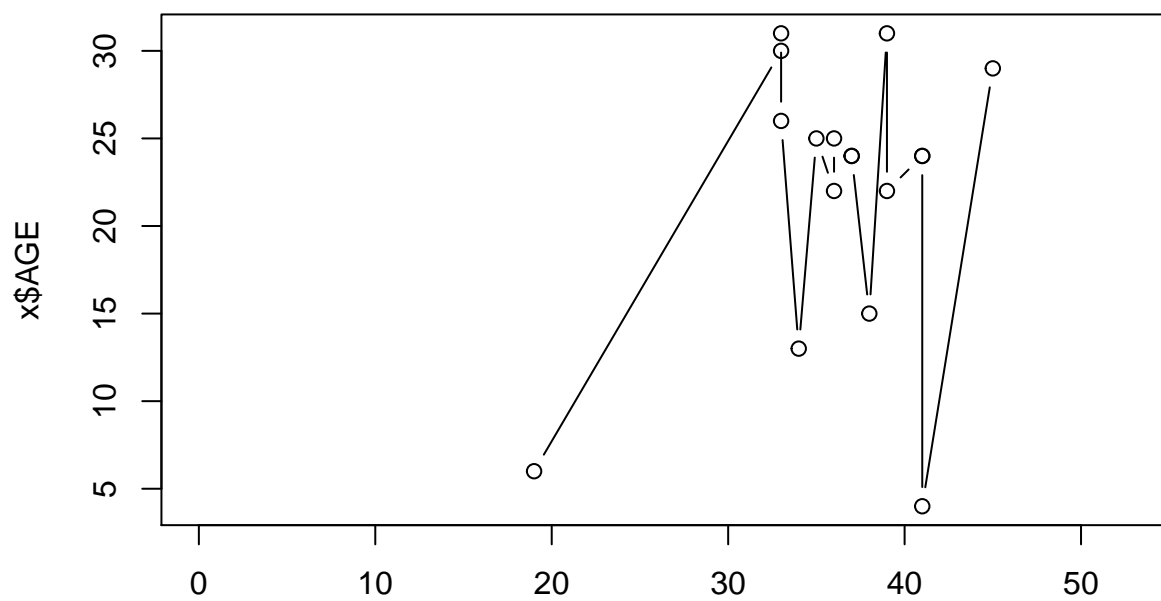

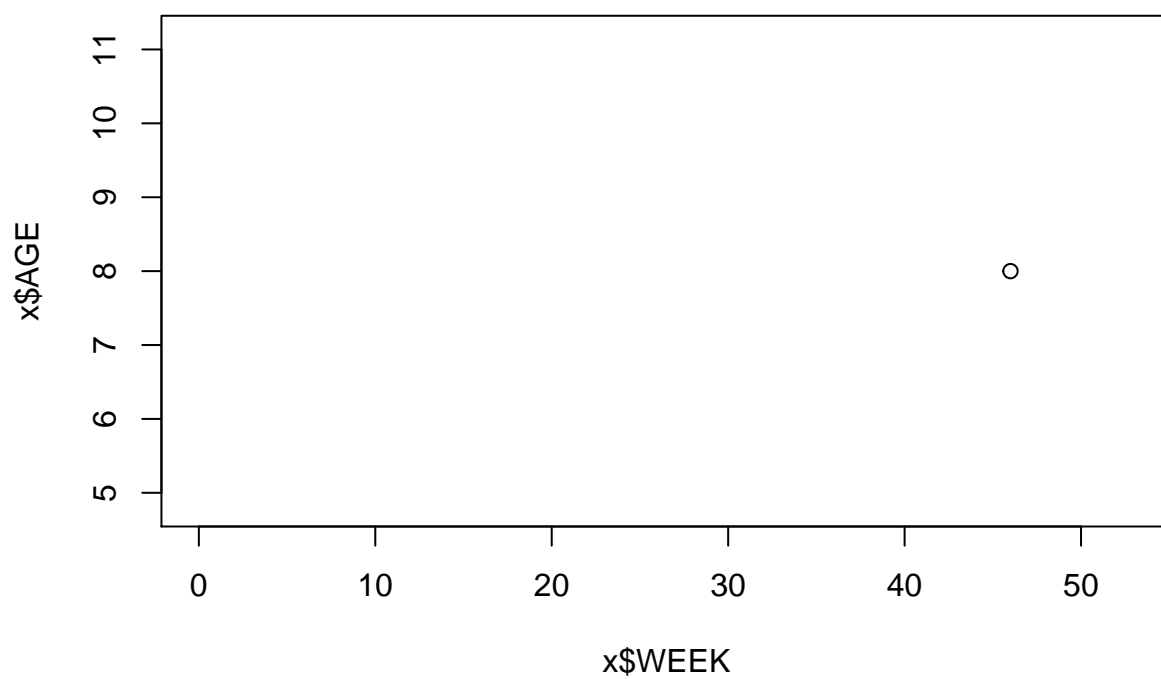x$WEEK

ITU

x$WEEK

CARAGUATATUBA

x$WEEK
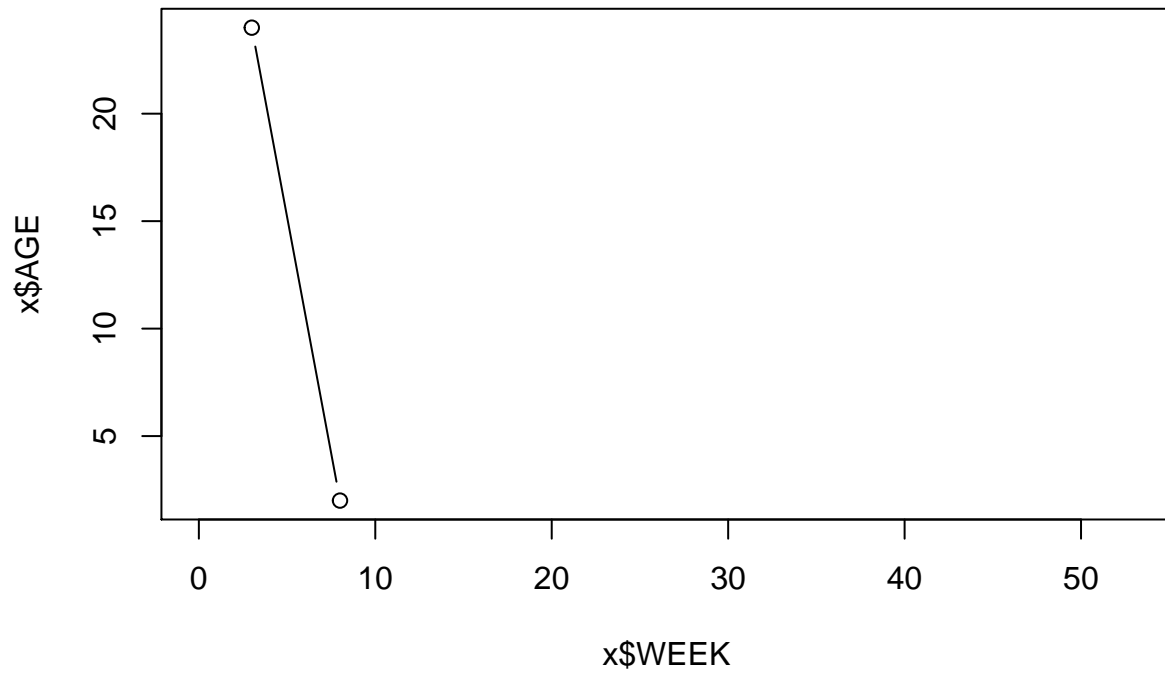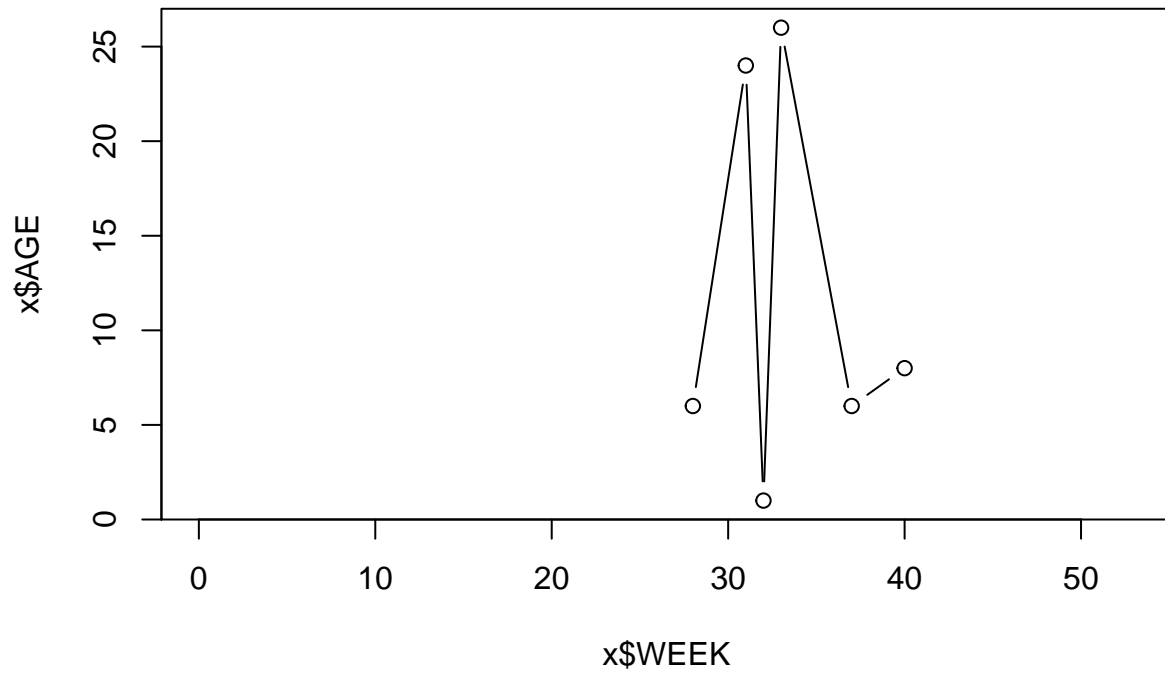
ITATINGA
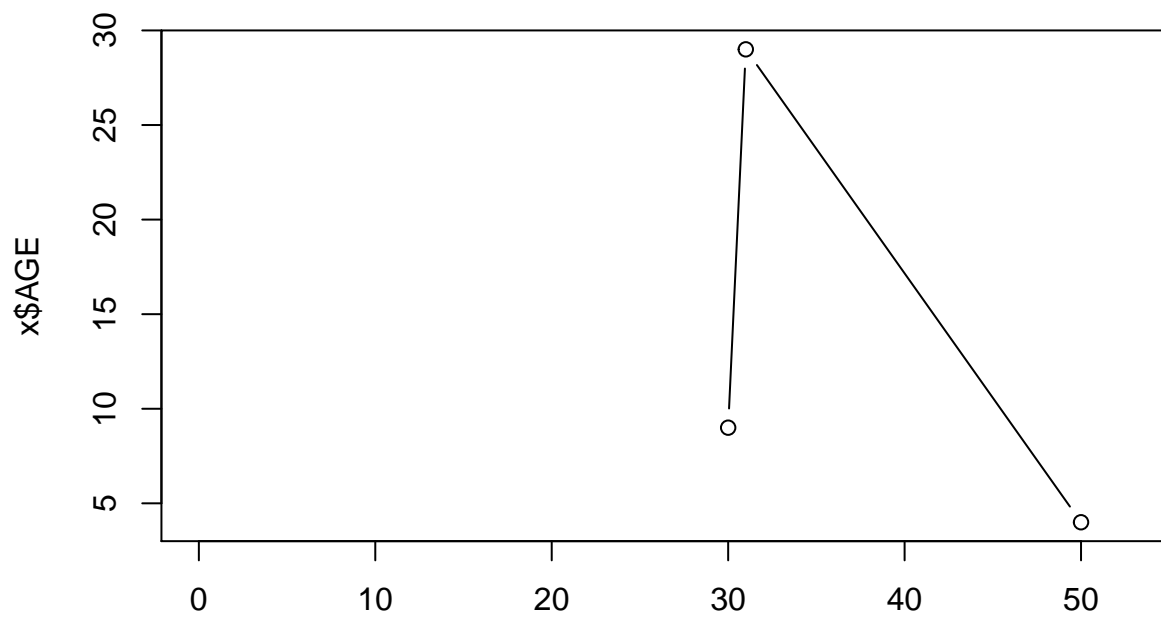
IGARACU DO TIETE
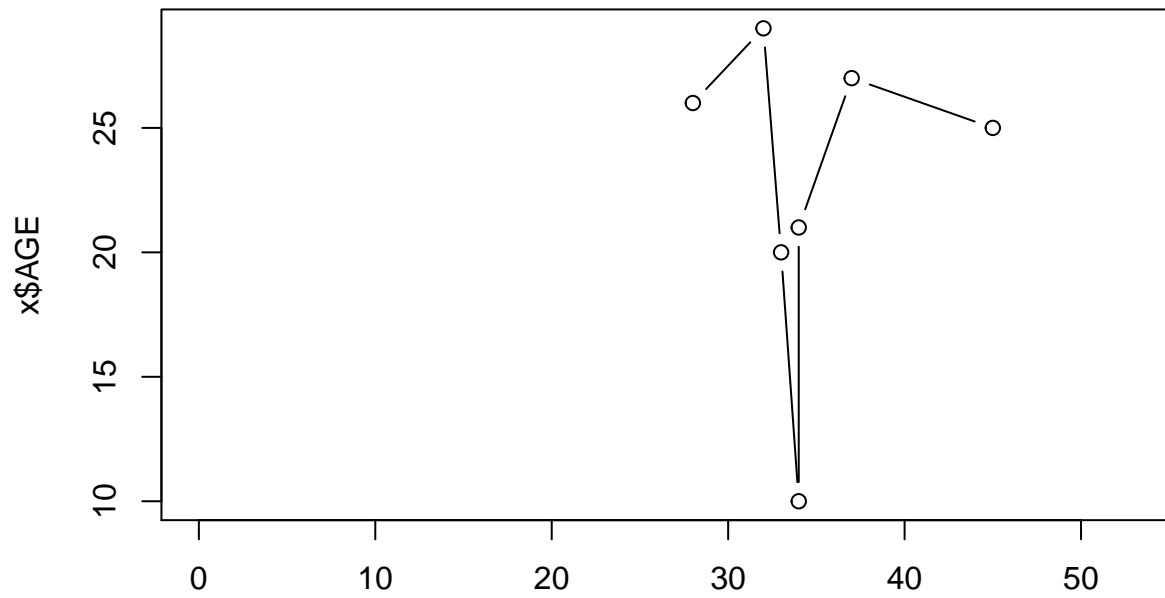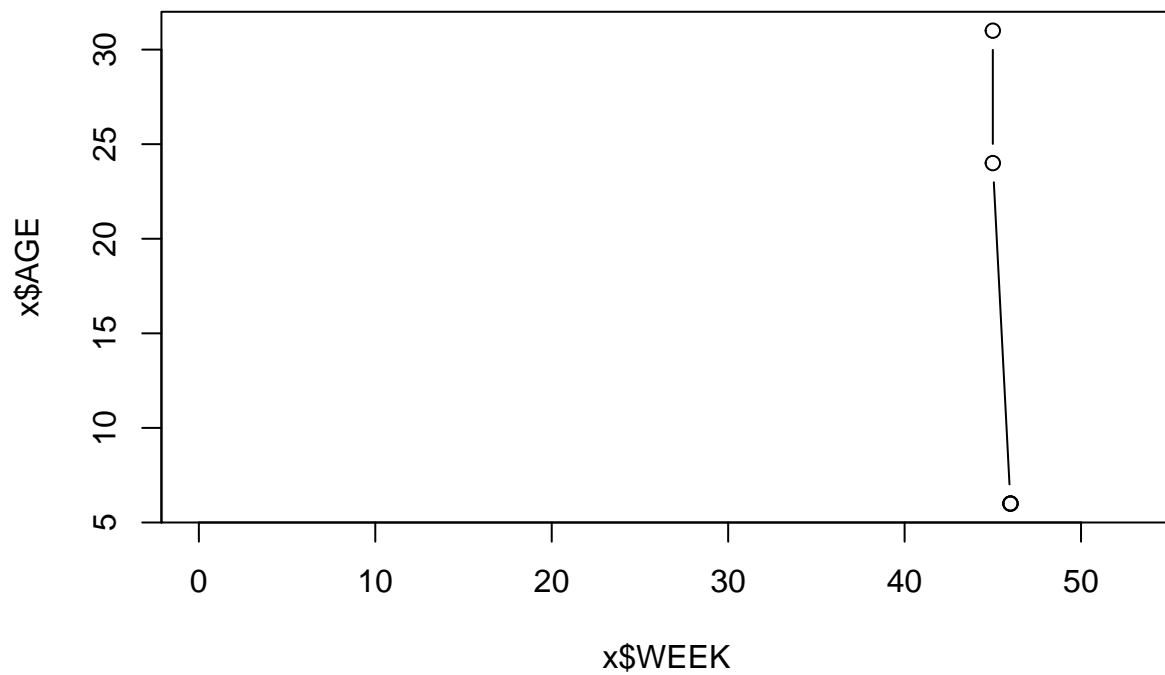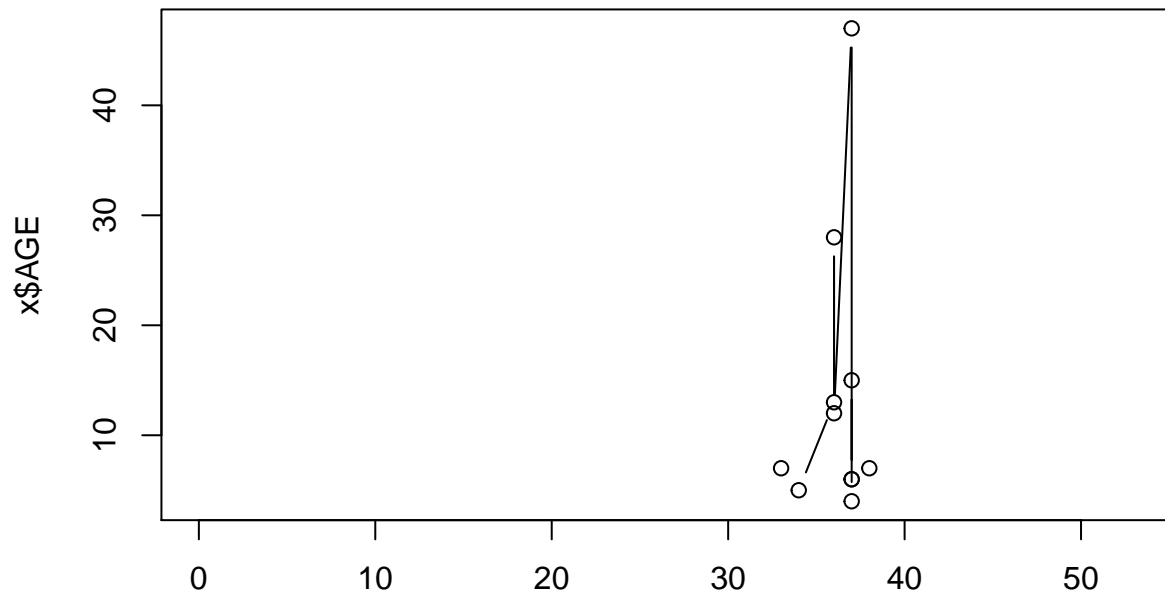
CAMPOS DO JORDAO

BENTO DE ABREU

ARACATUBA

PARAGUACU PAULISTA
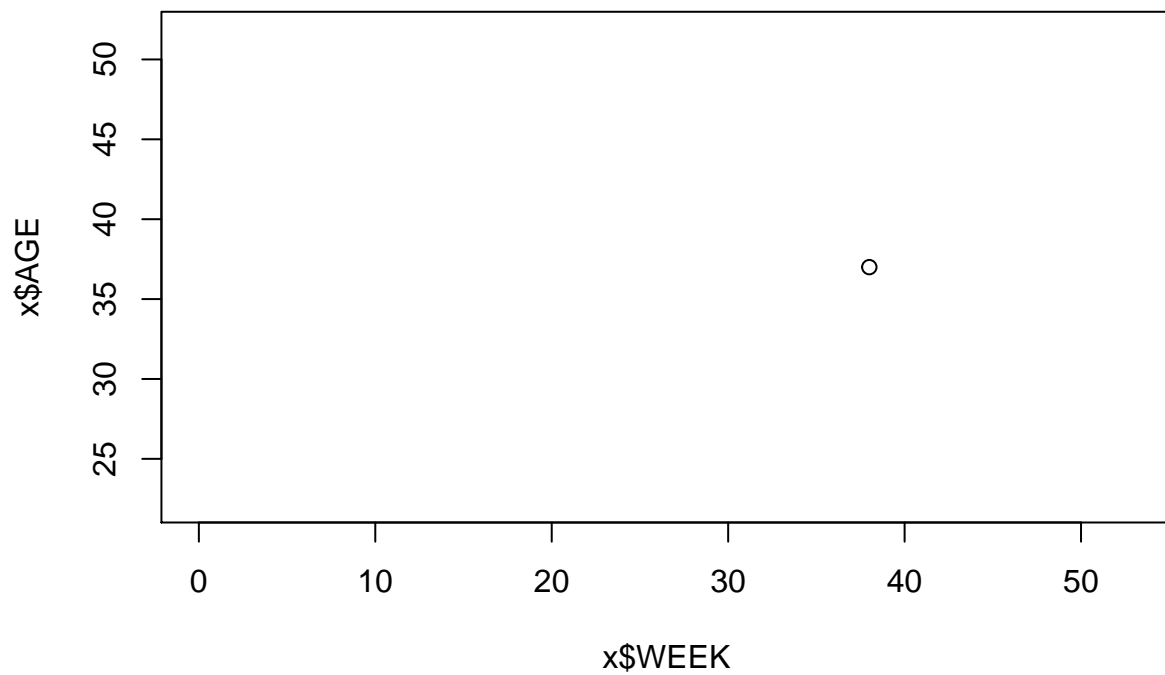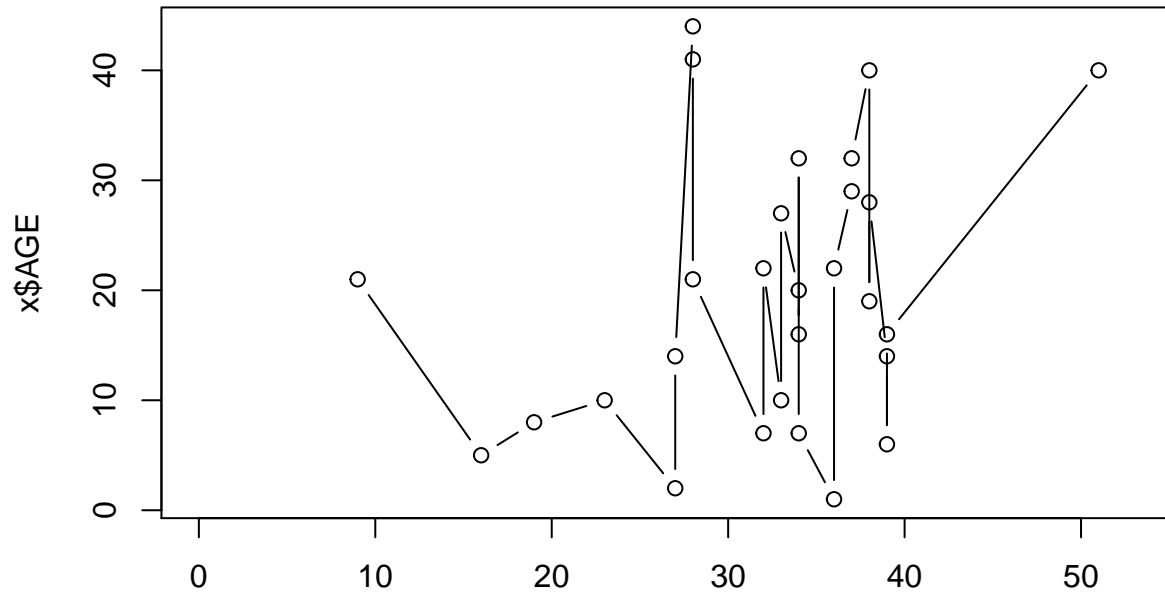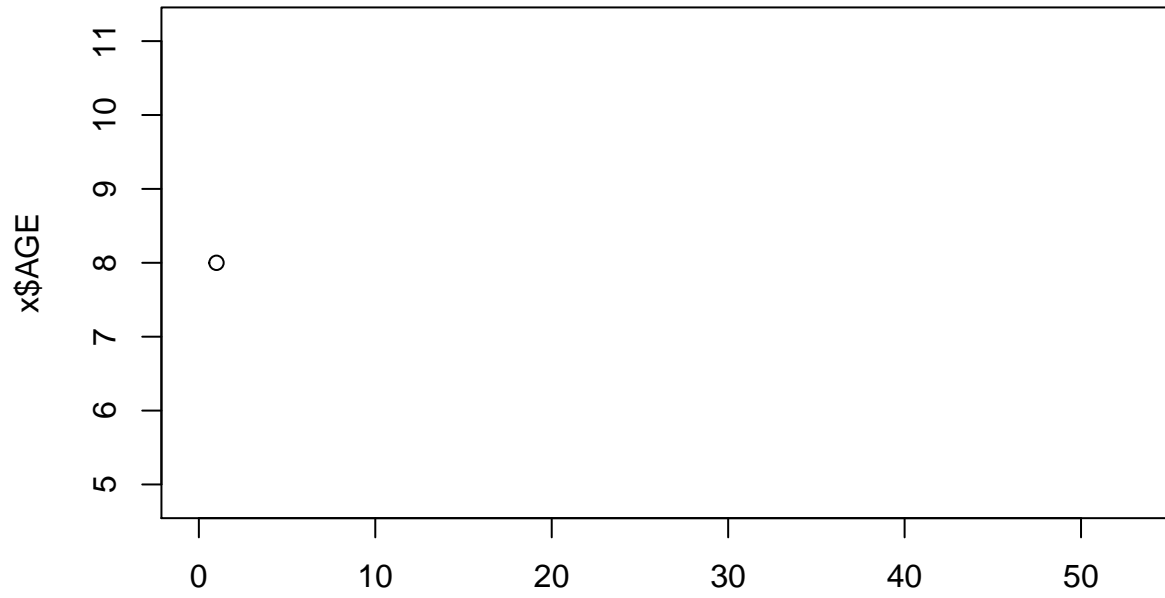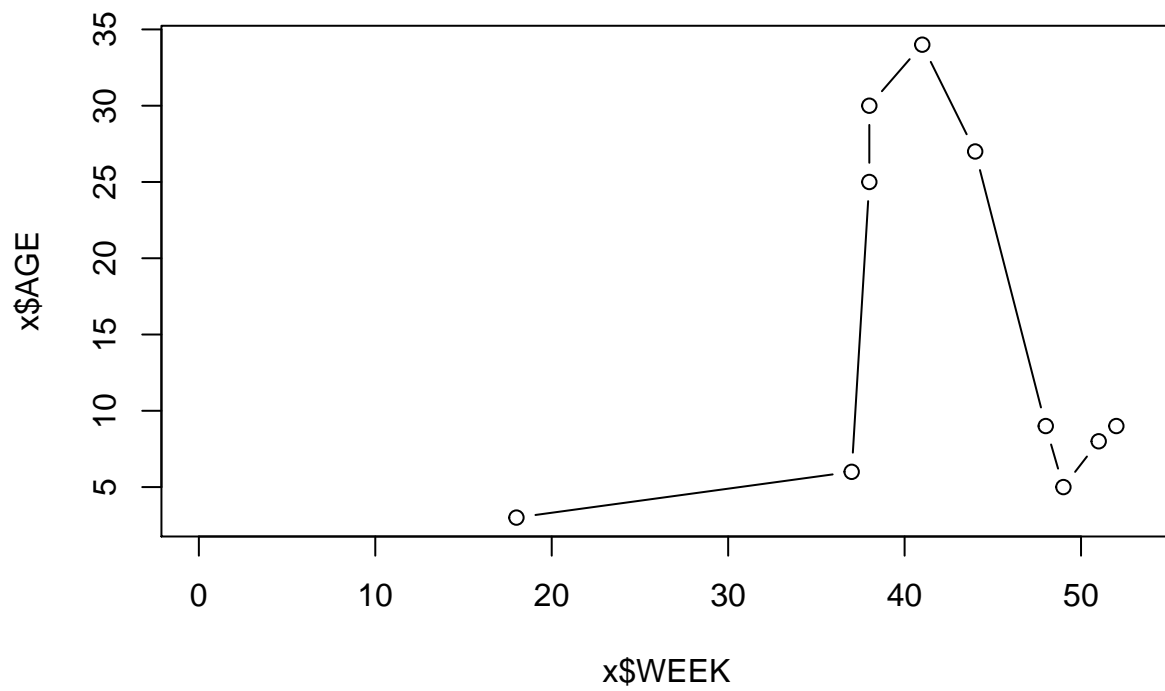
CACAPAVA
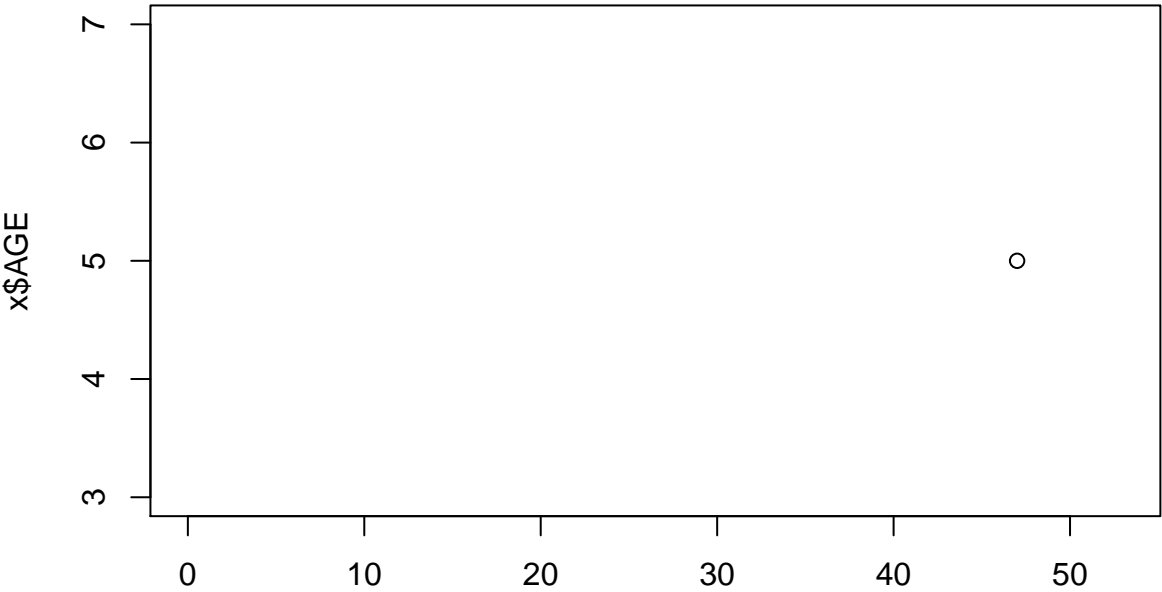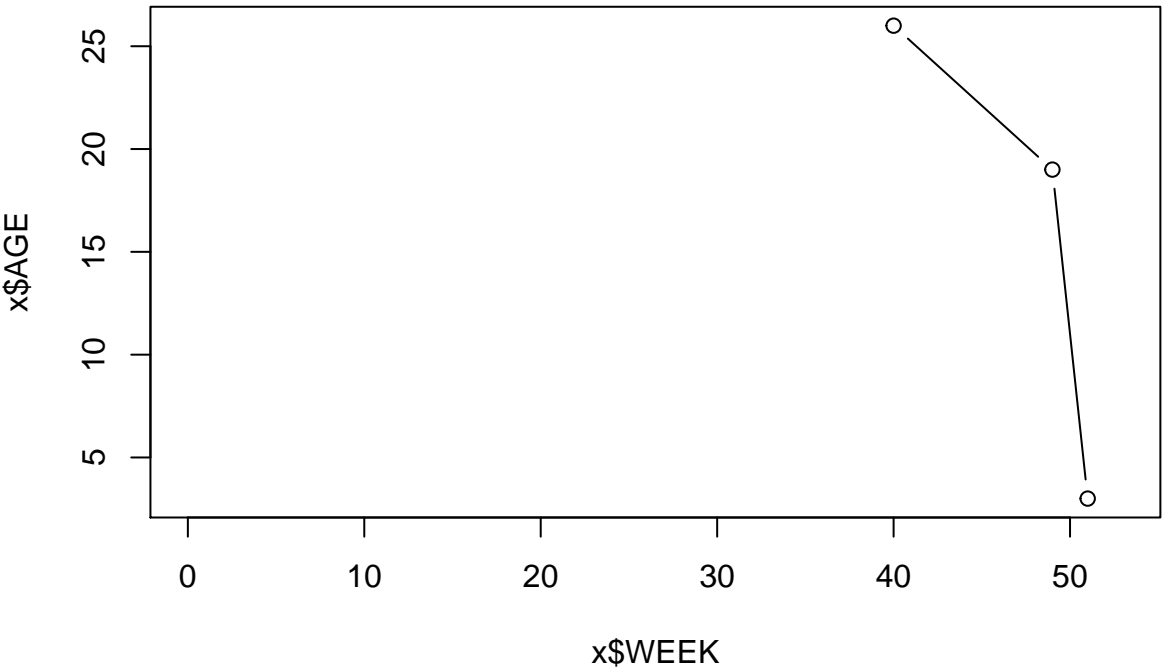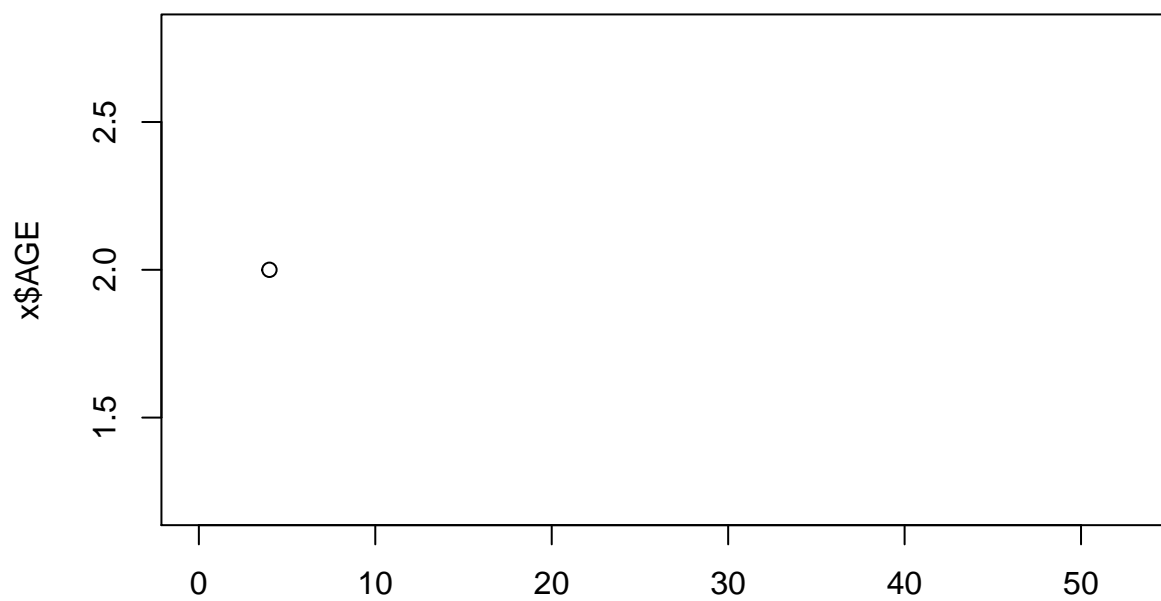

PORTO FERREIRA

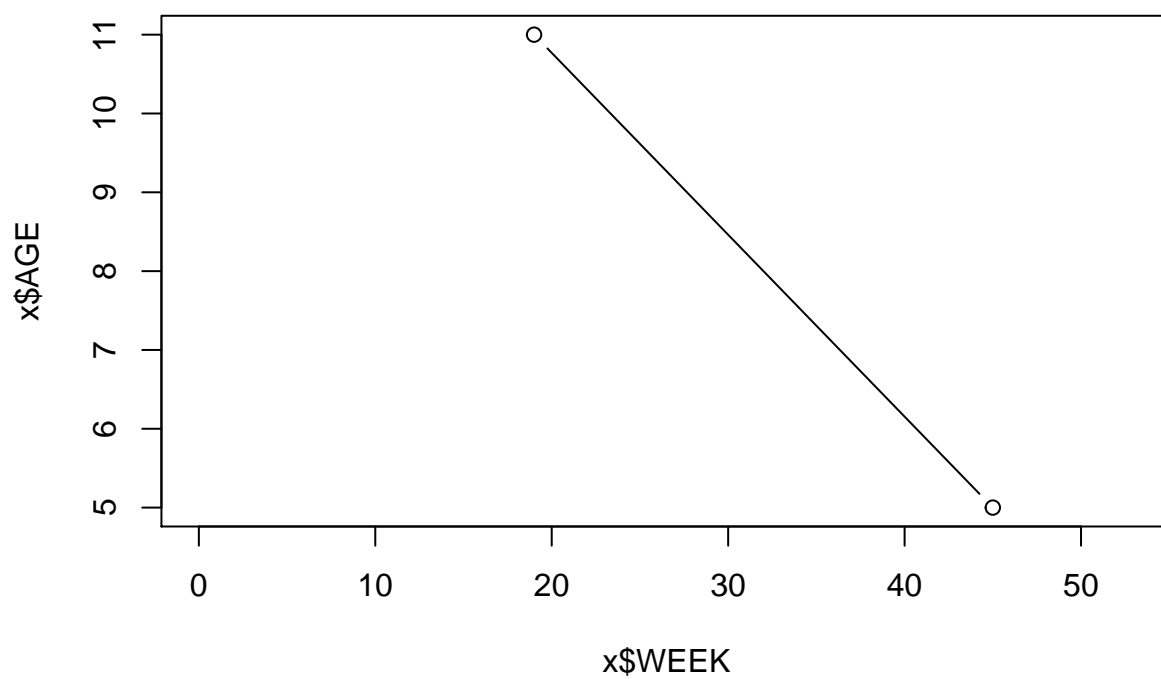**ATIBAIA**

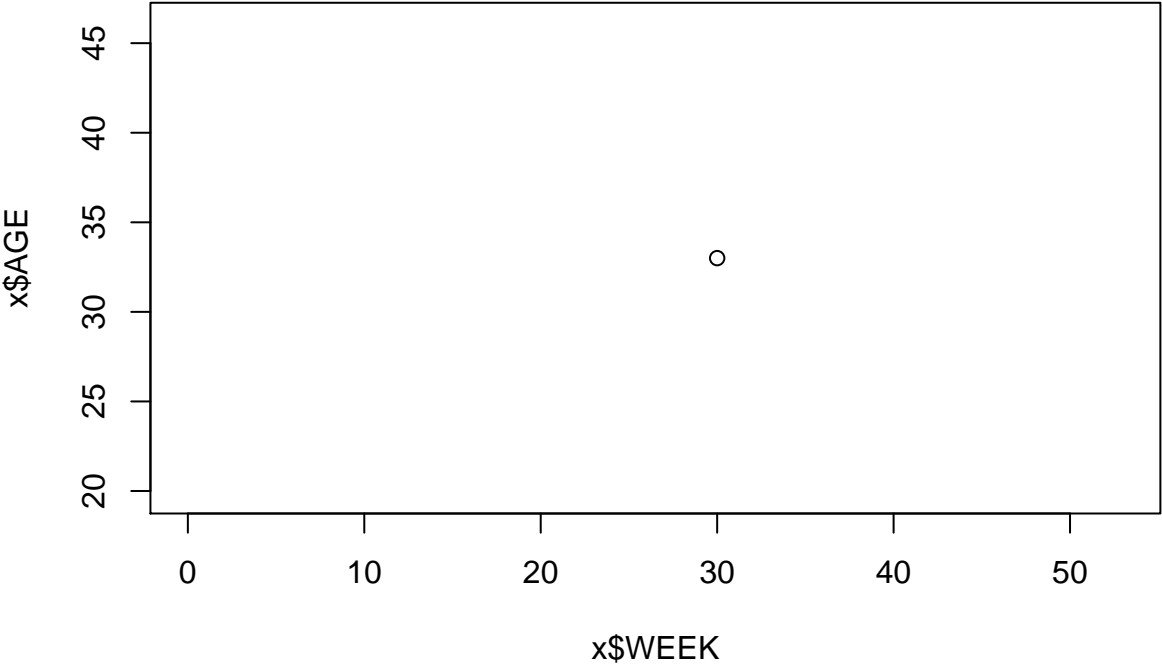**PENAPOLIS**
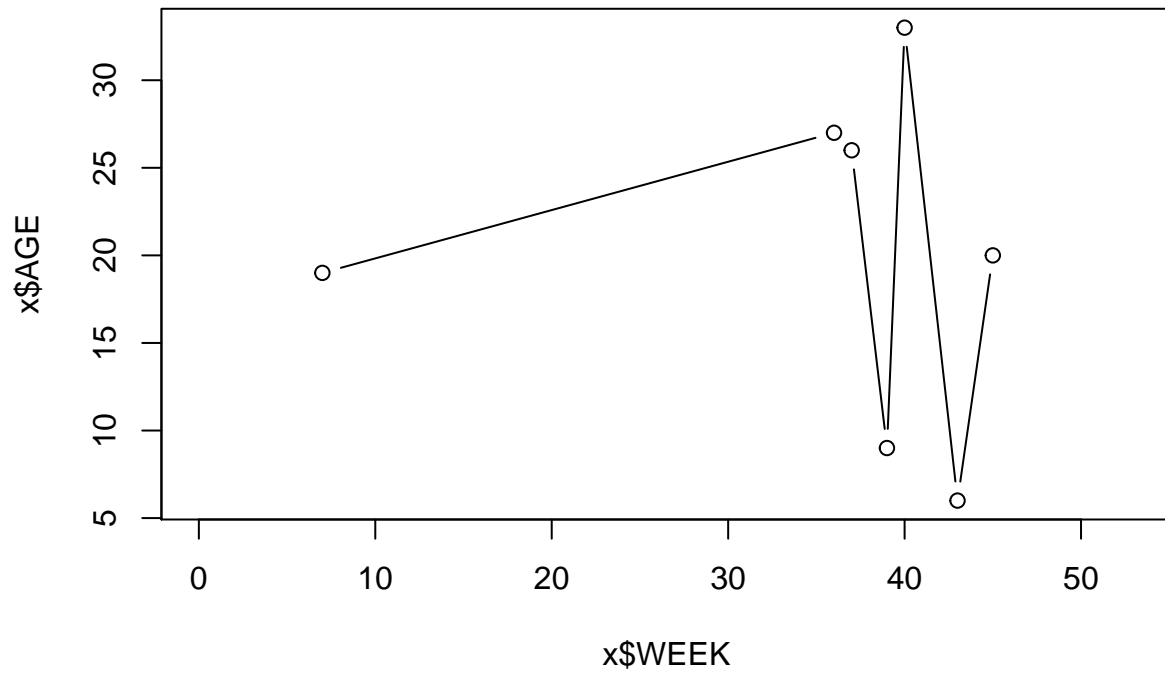
JACUPIRANGA

BARRA BONITA

JUNQUEIROPOLIS

NARANDIBA

BOFETE

GUAPIARA

SAO JOSE DO RIO PARDO

PEREIRAS
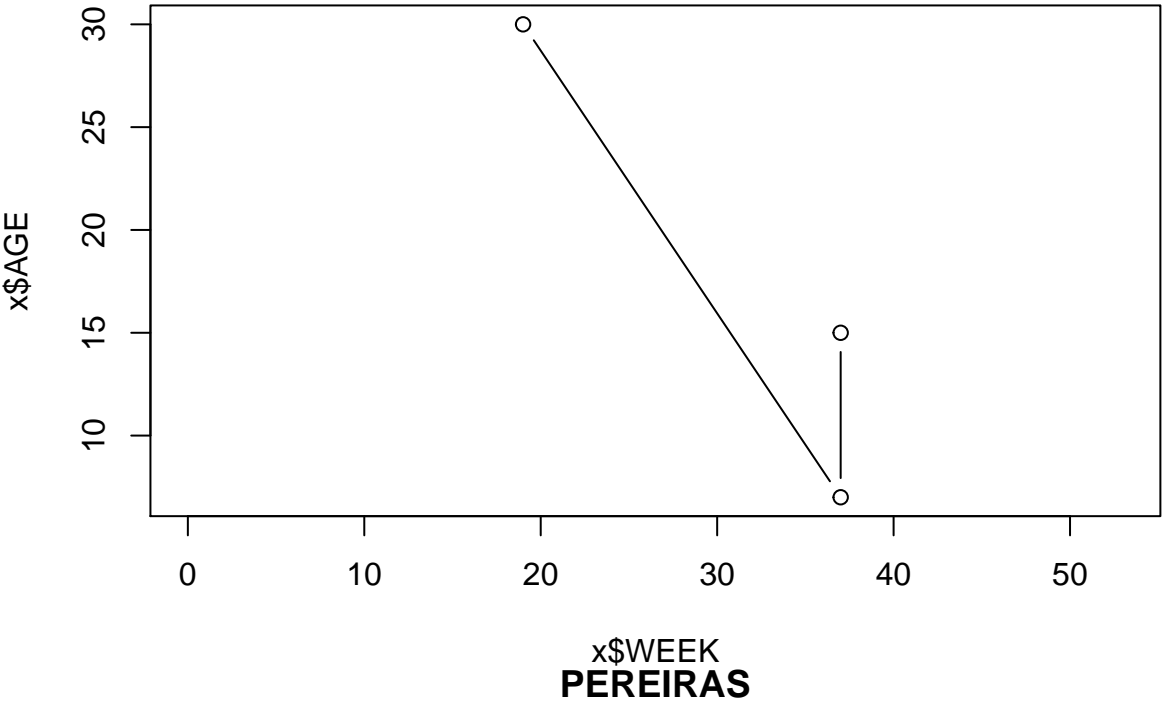
QUINTANA
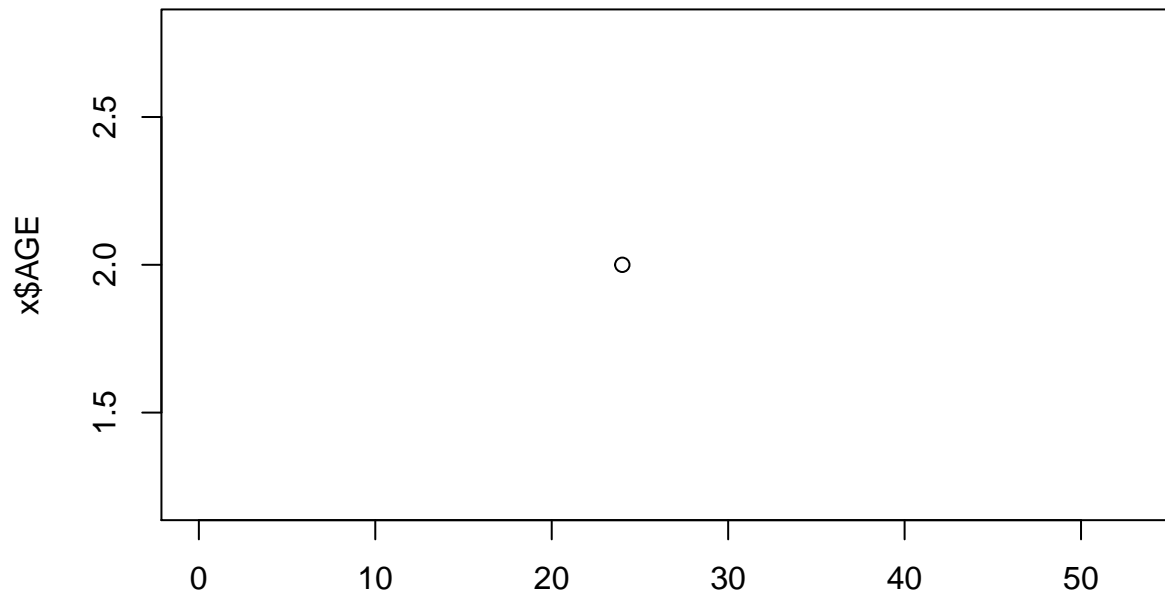
MONTE CASTELO

PANORAMA

x$WEEK

PAULICEIA

x$WEEK
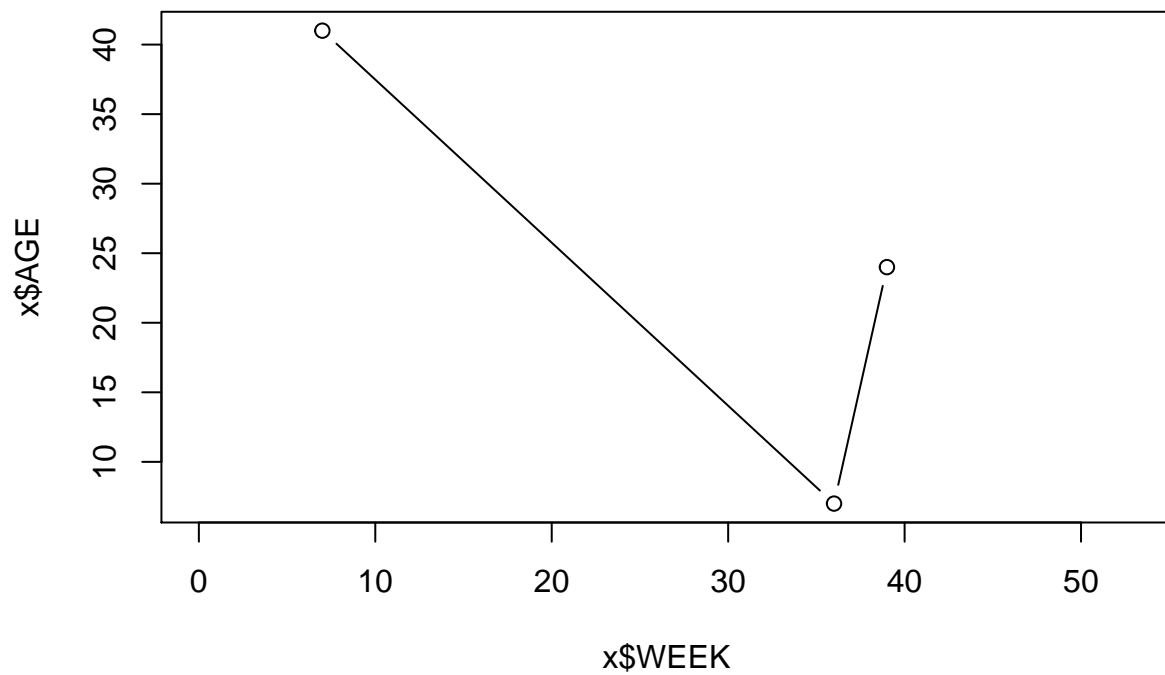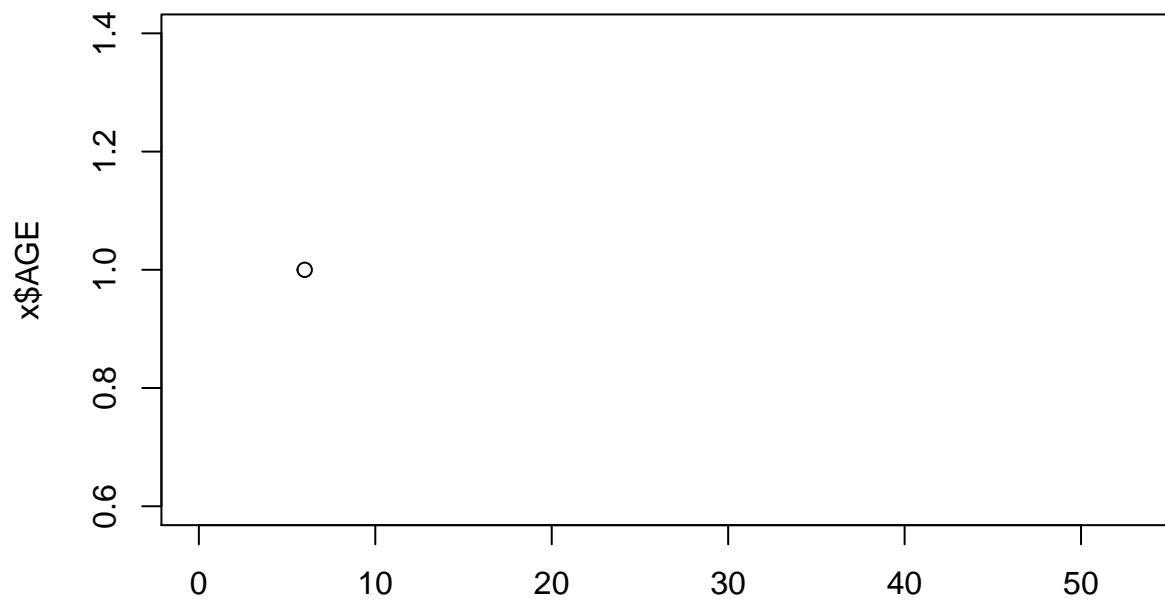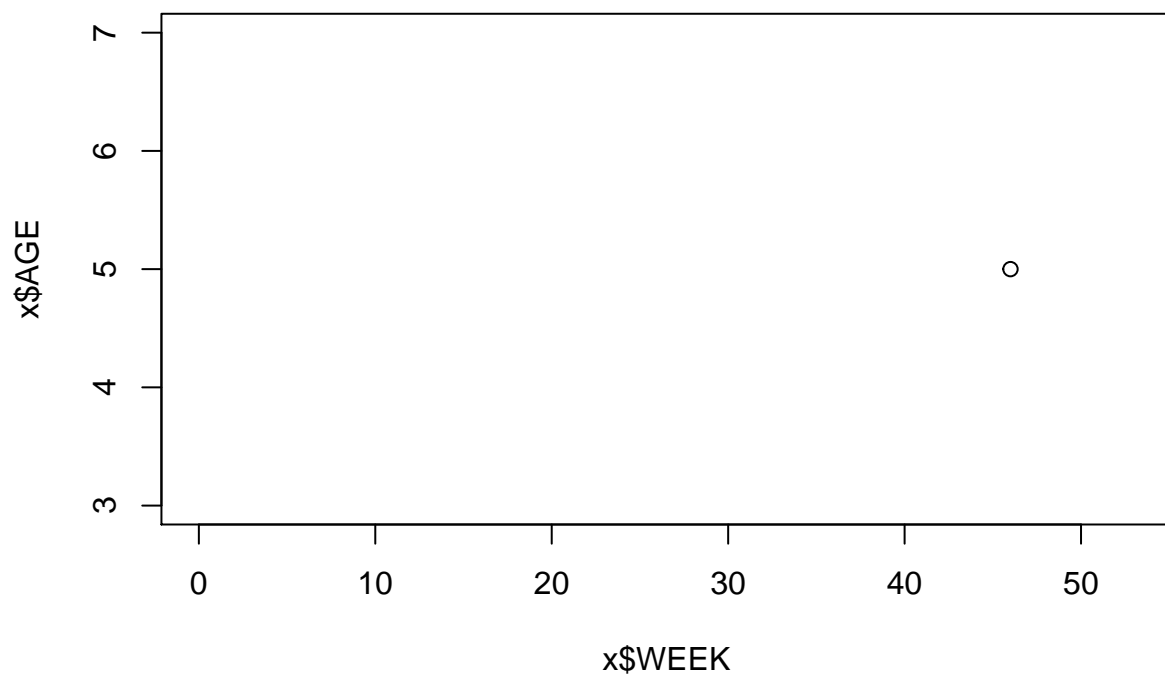
ITANHAEM
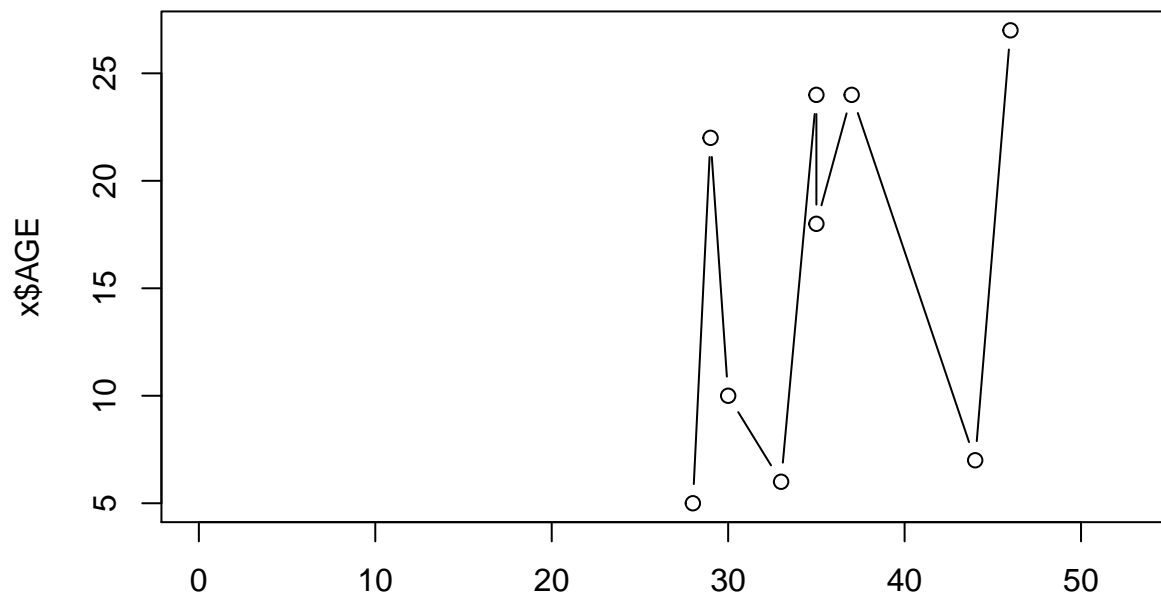
PRESIDENTE VENCESLAU
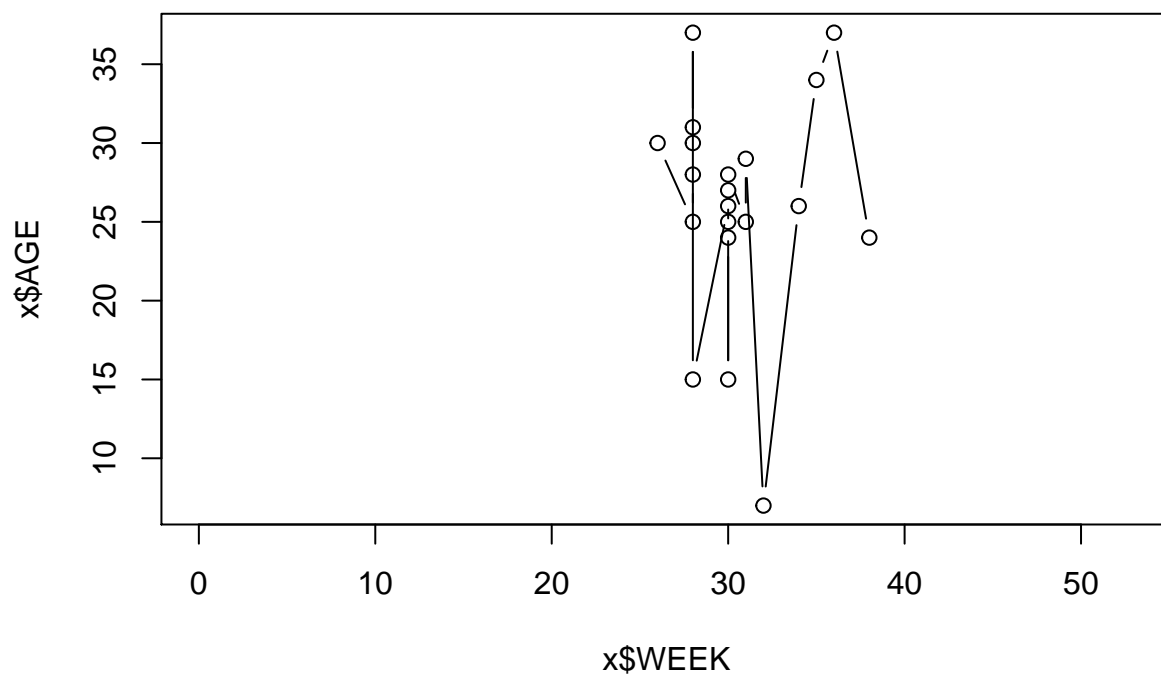
PIRAPOZINHO
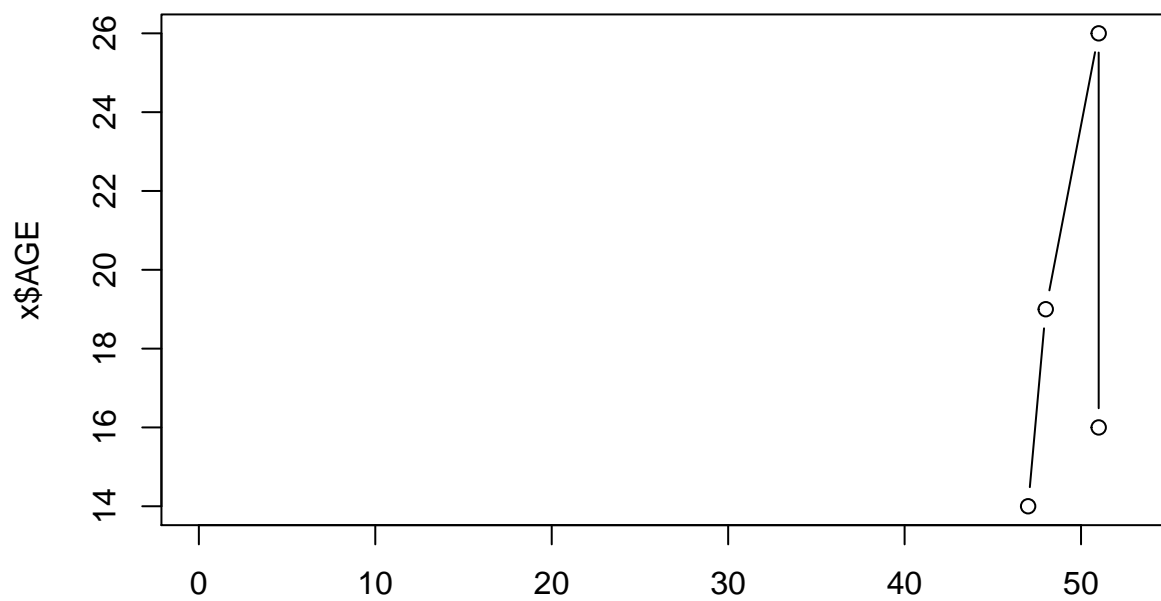
JABORANDI
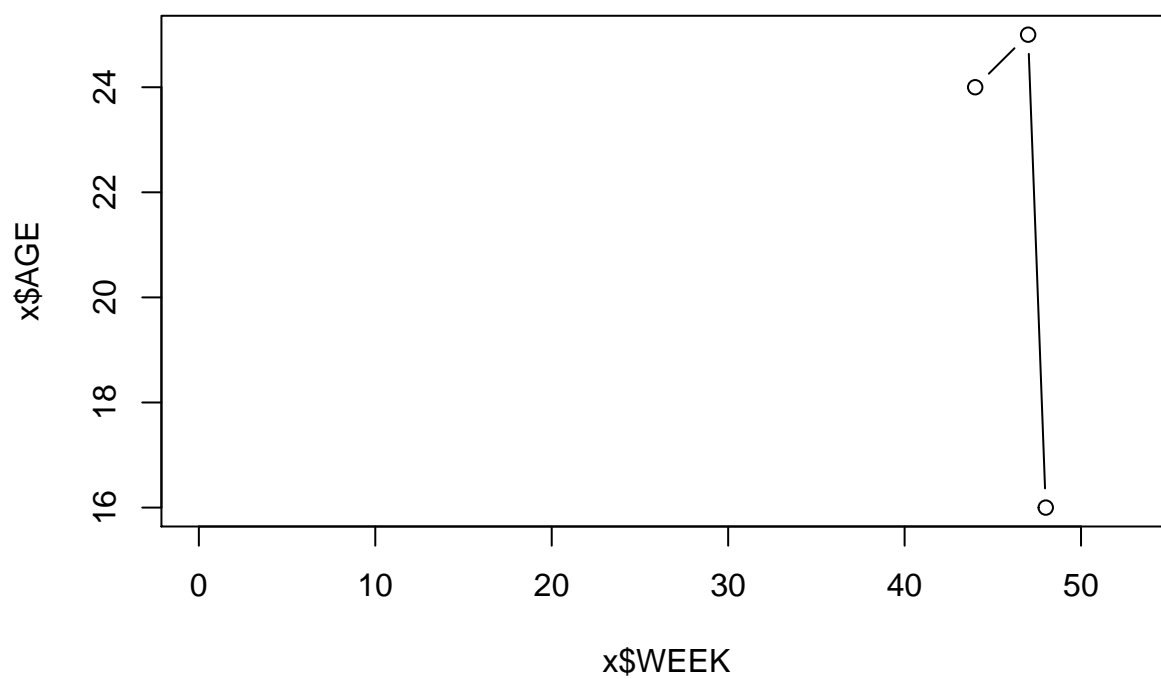
ESTIVA GERBI
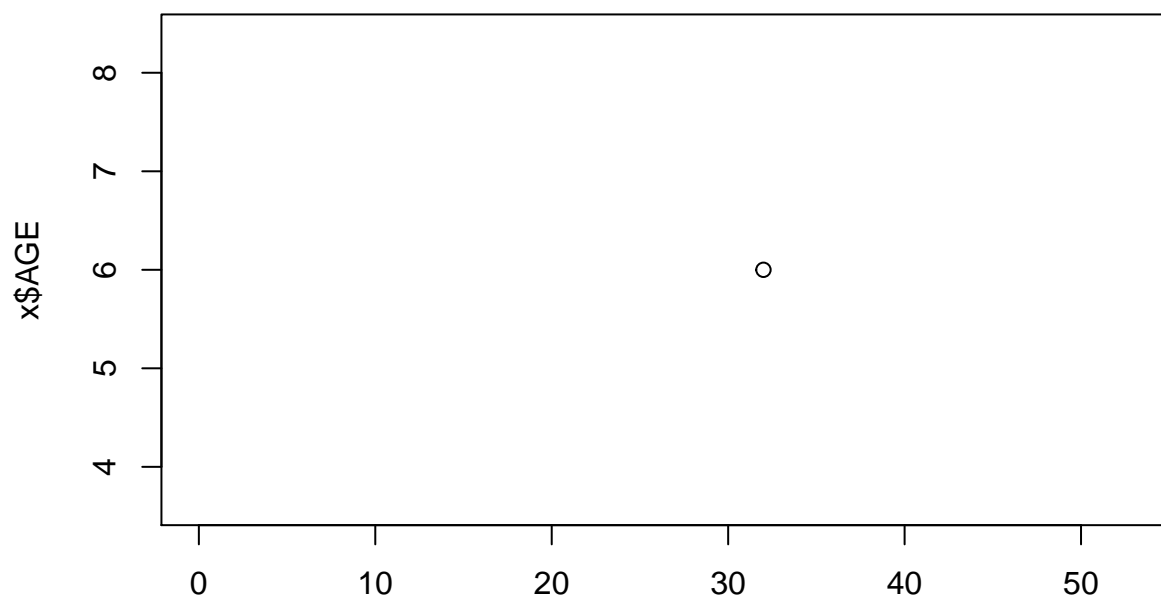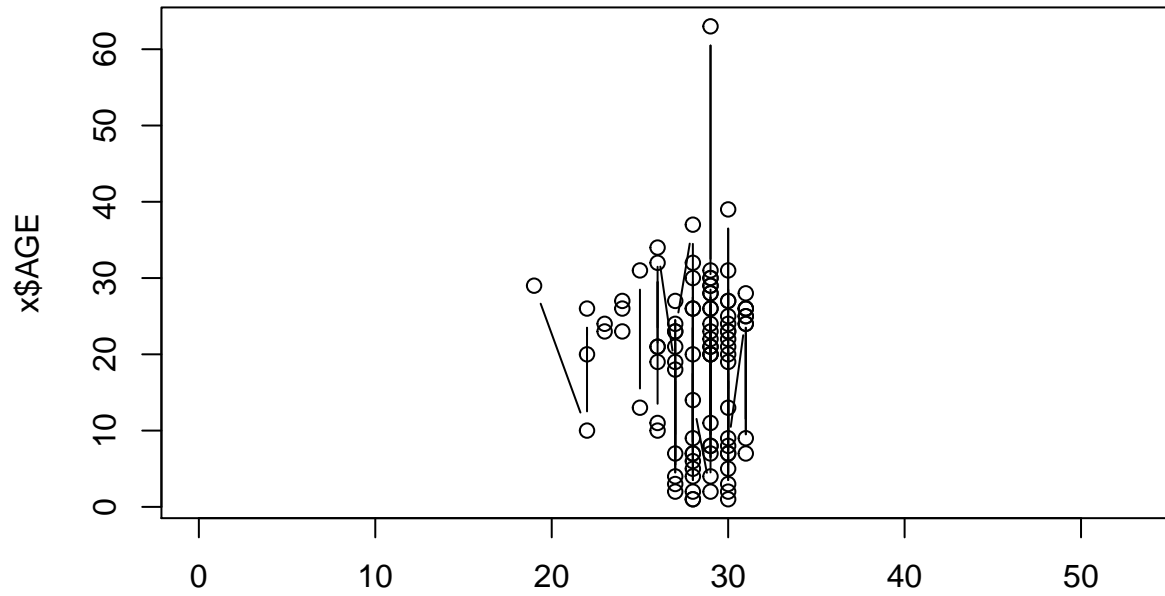

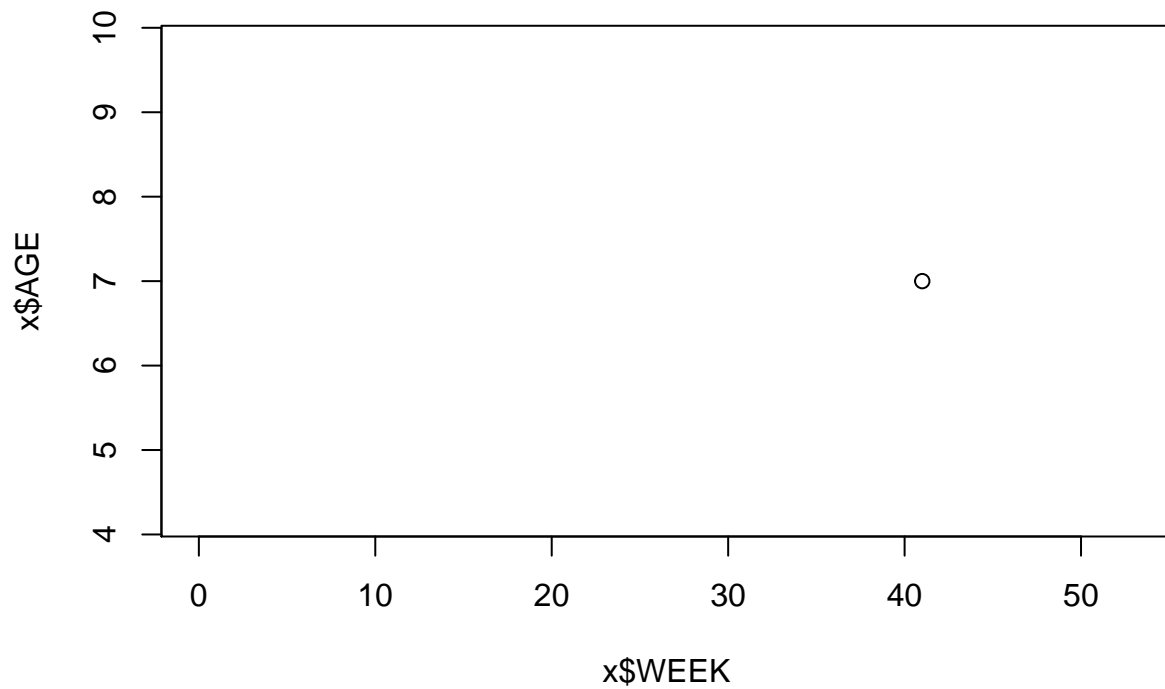**MONTEIRO LOBATO**
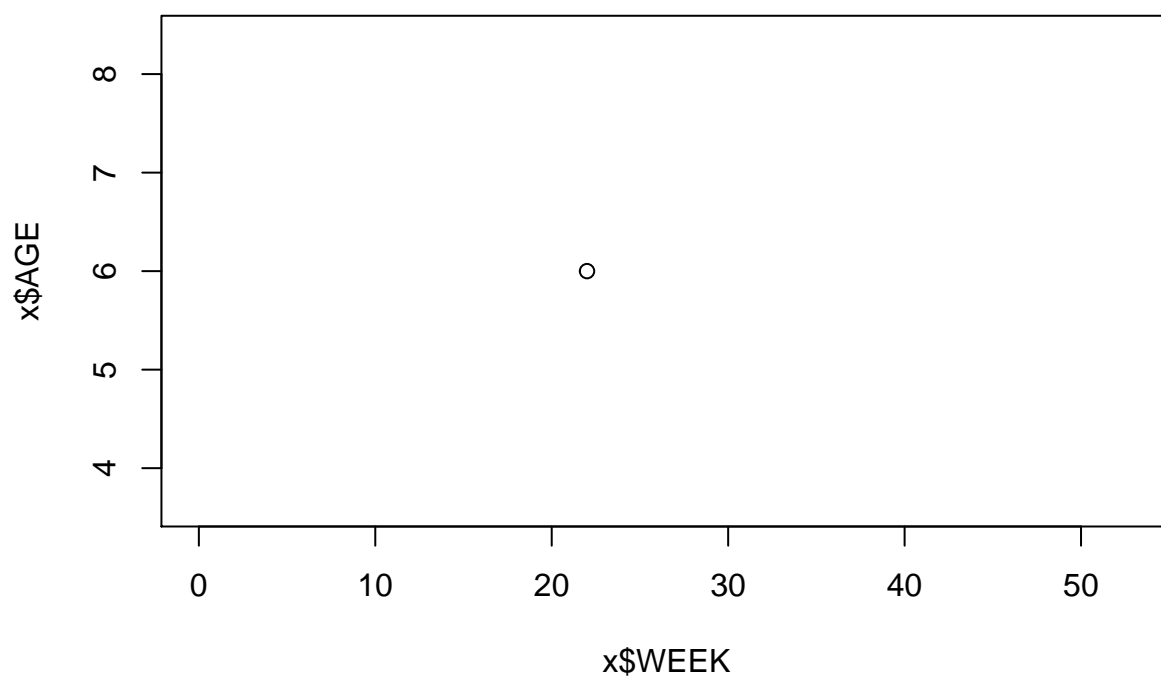
RIBEIRAO DOS INDIOS
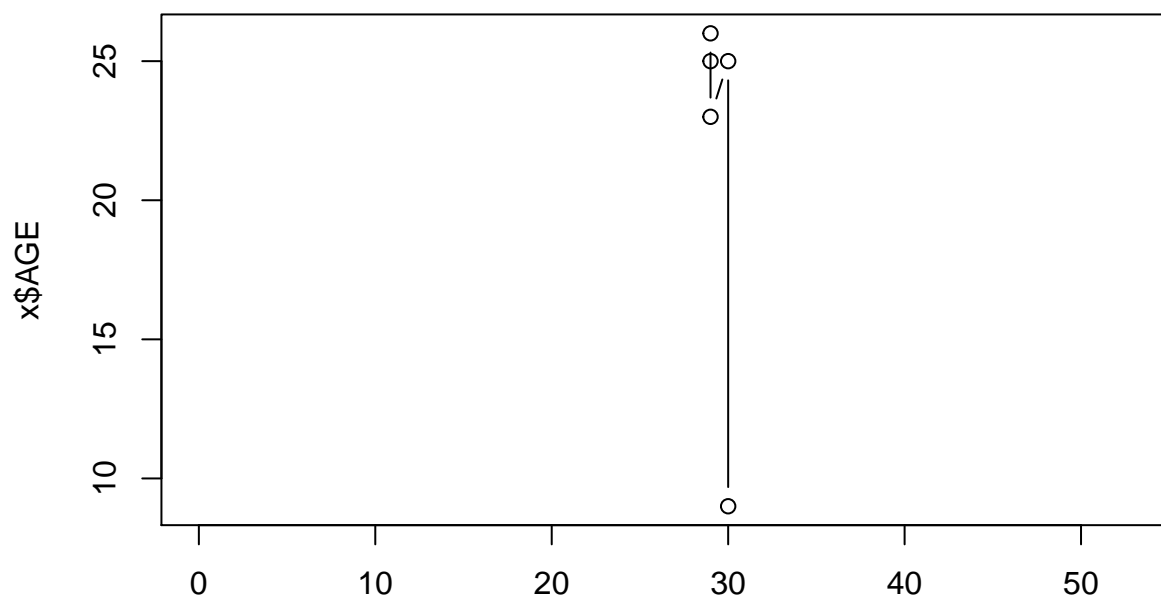
NOVA EUROPA
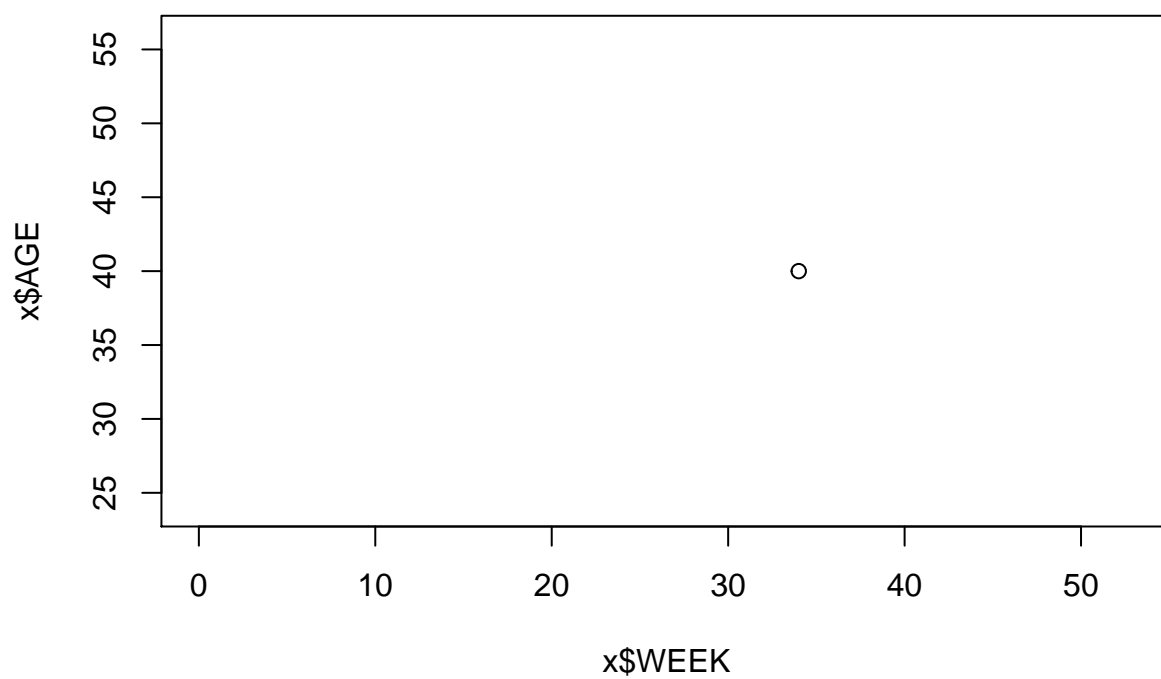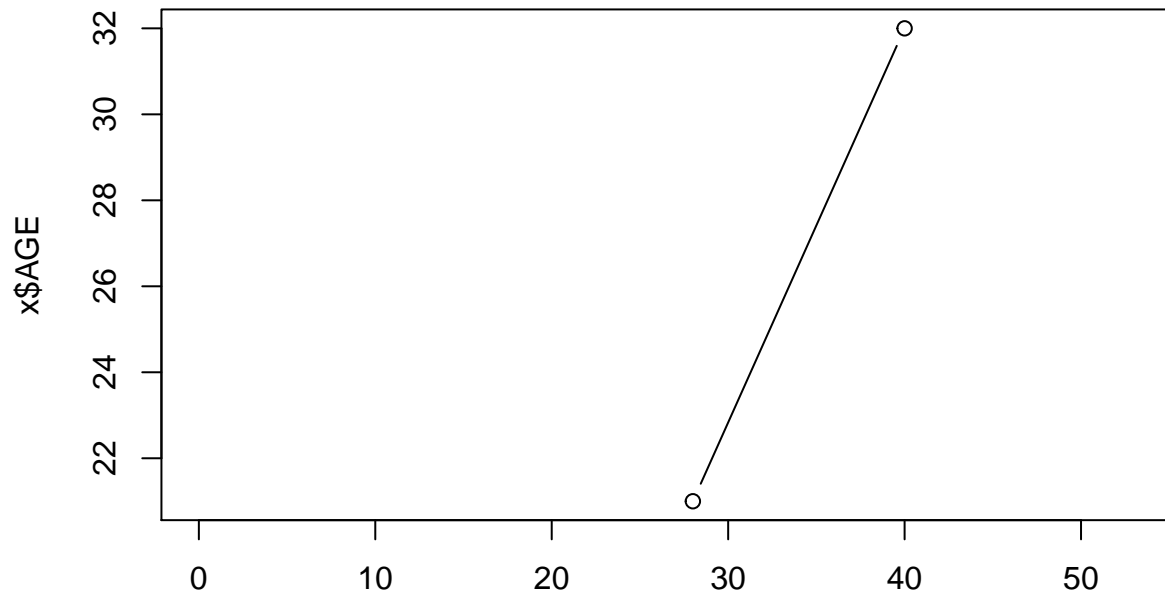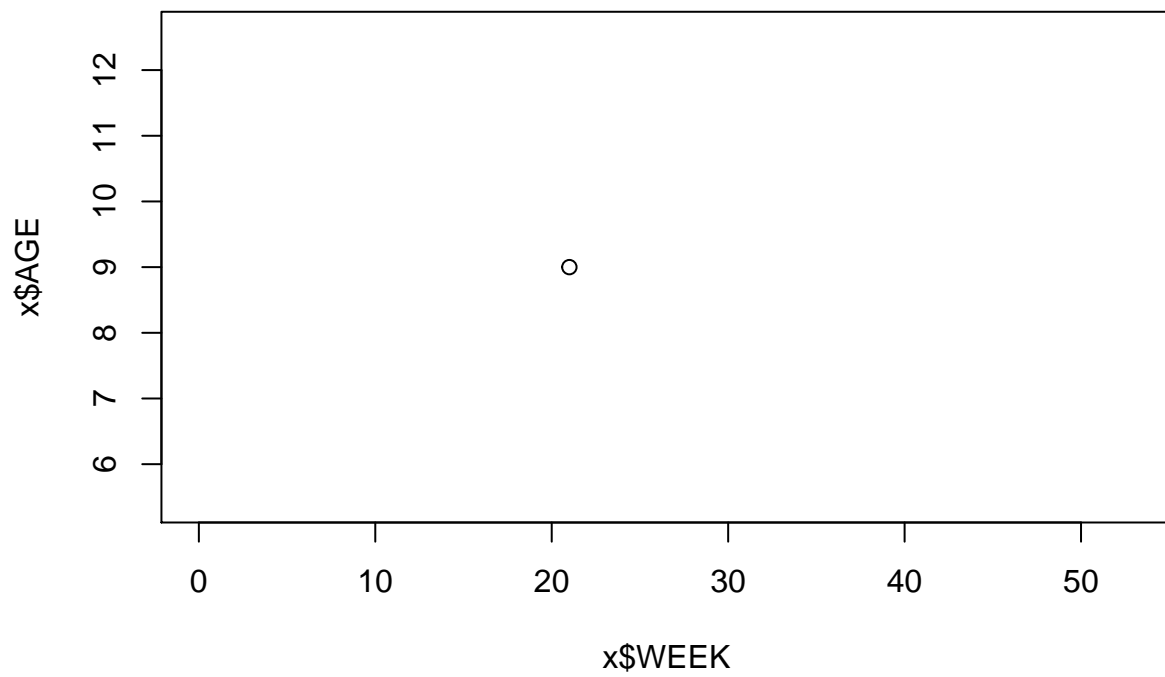
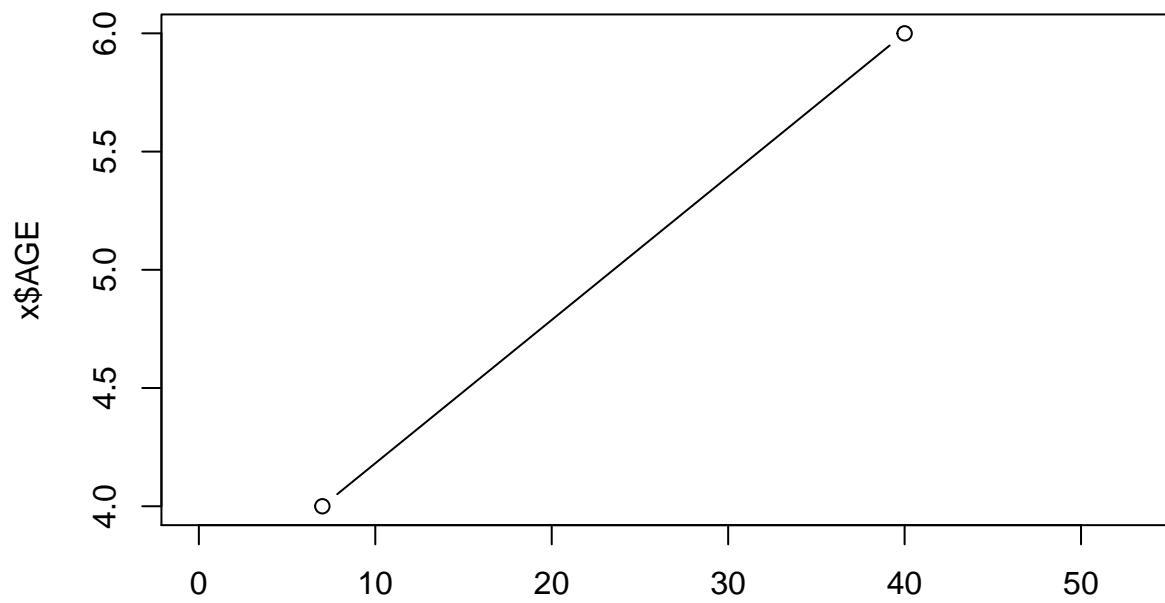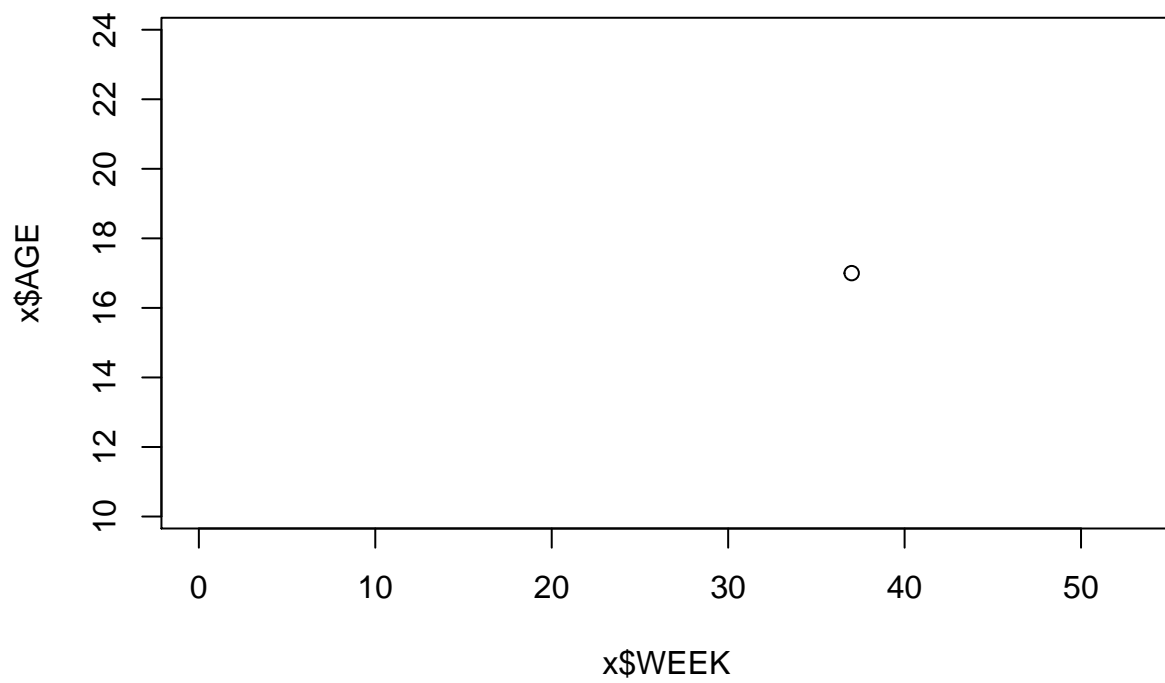**ALAMBARI**
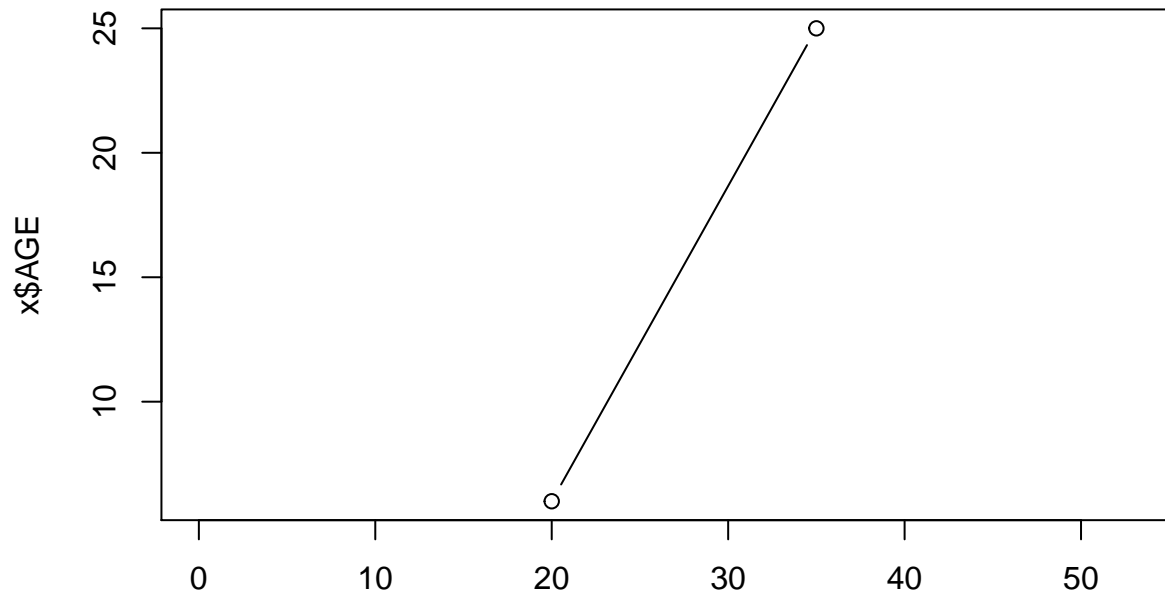
PERUIBE
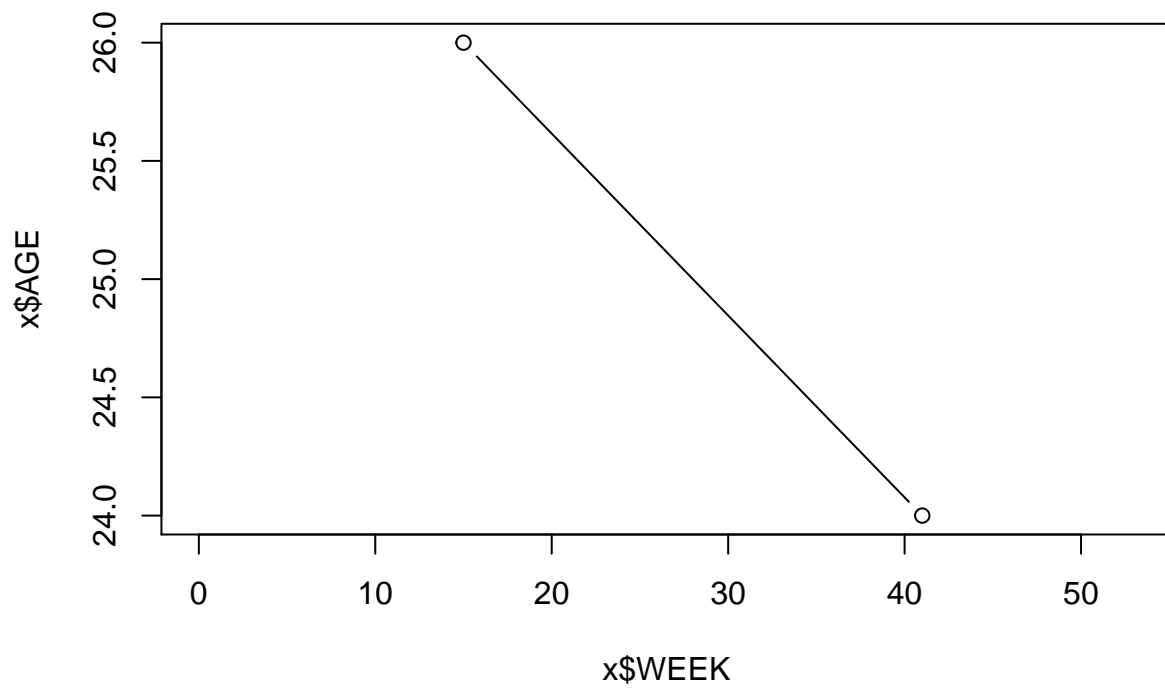
ITAPEVI

AGUAS DE SANTA BARBARA
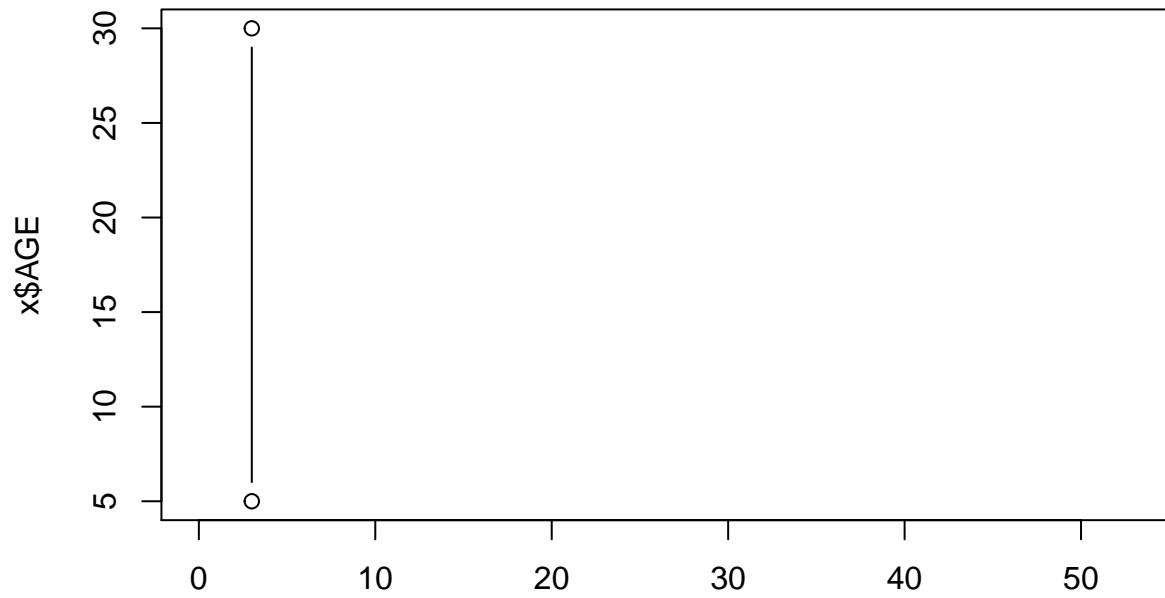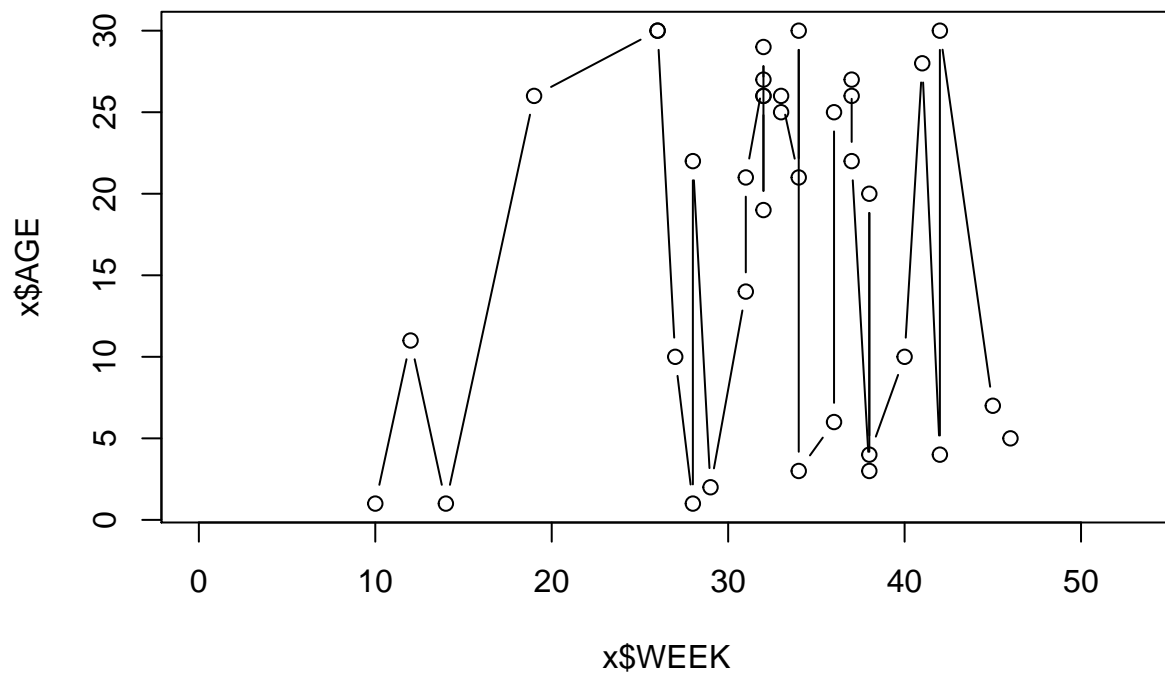
MOCOCA

POMPEIA

GASTAO VIDIGAL

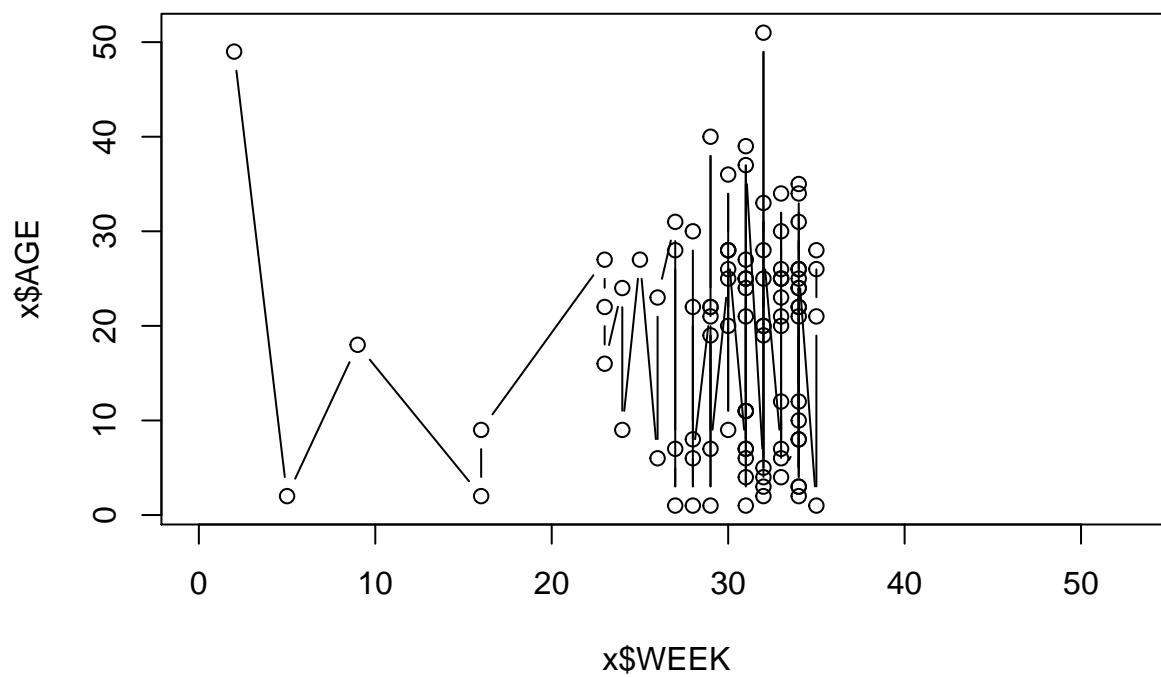# ROSANA

ANHEMBI

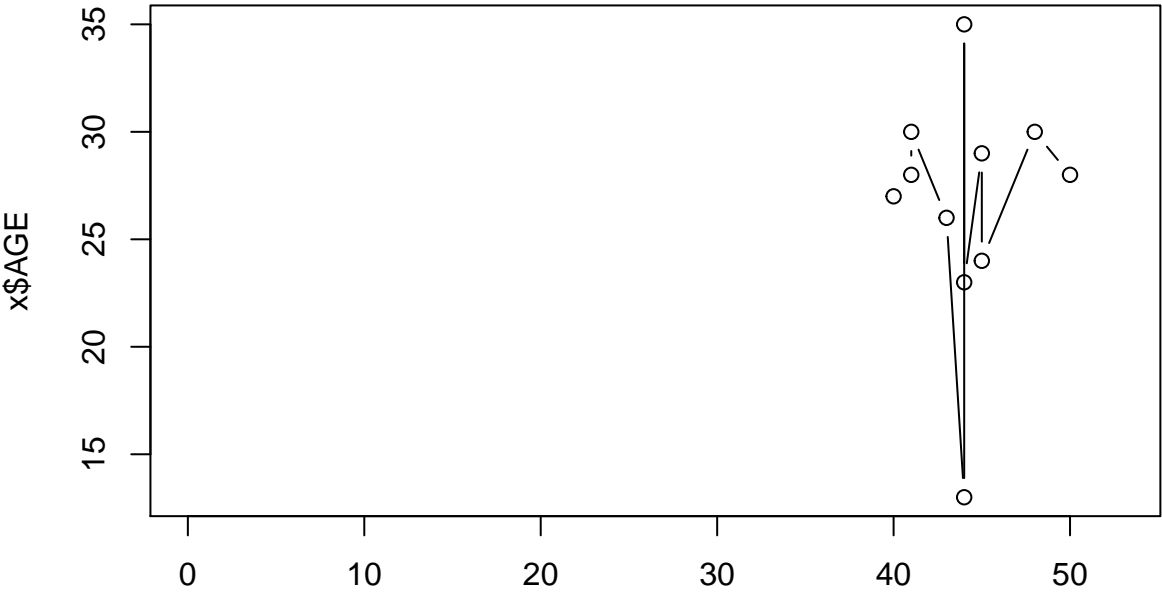**MAIRIPORA**

PORANGABA
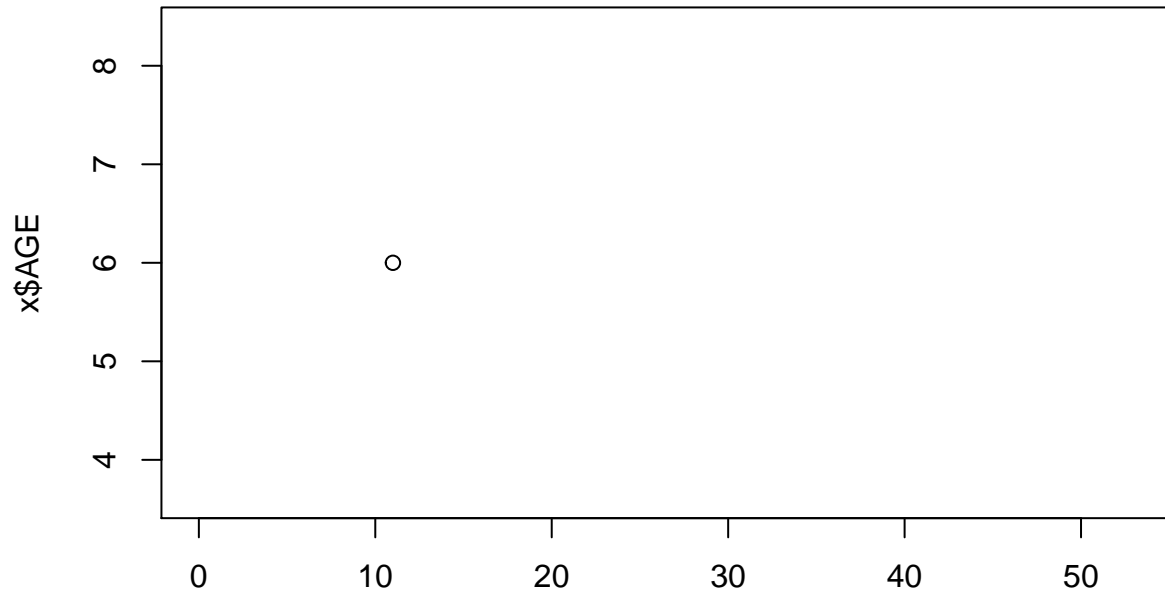
RIBEIRAO PRETO
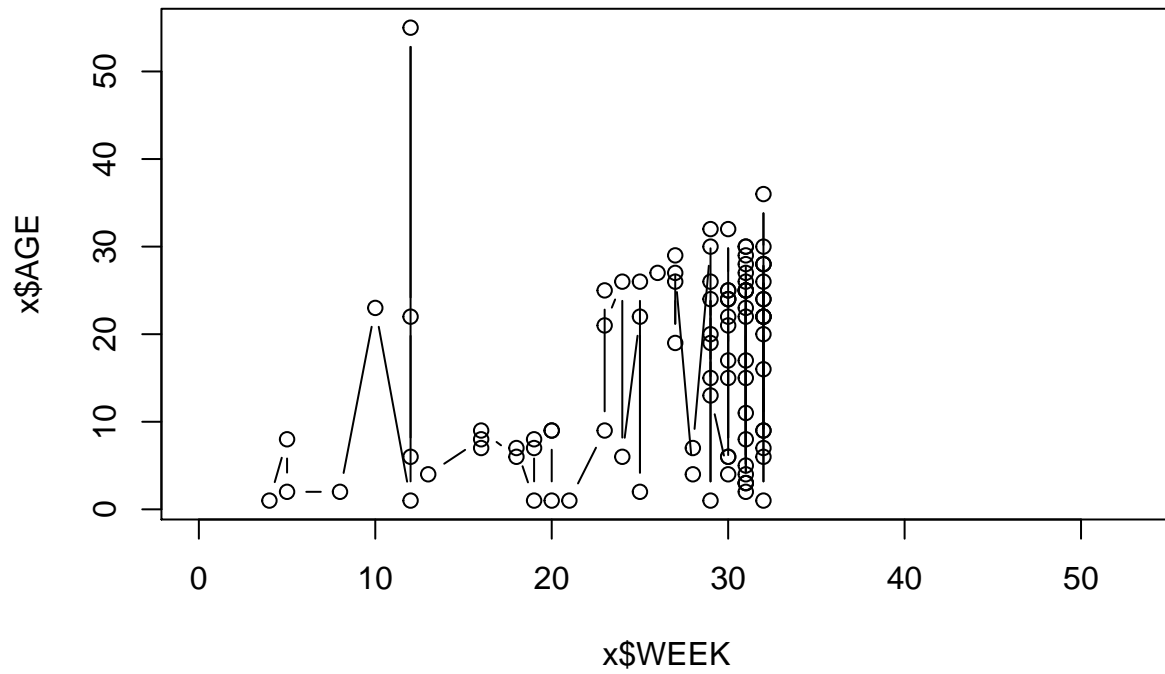
LORENA
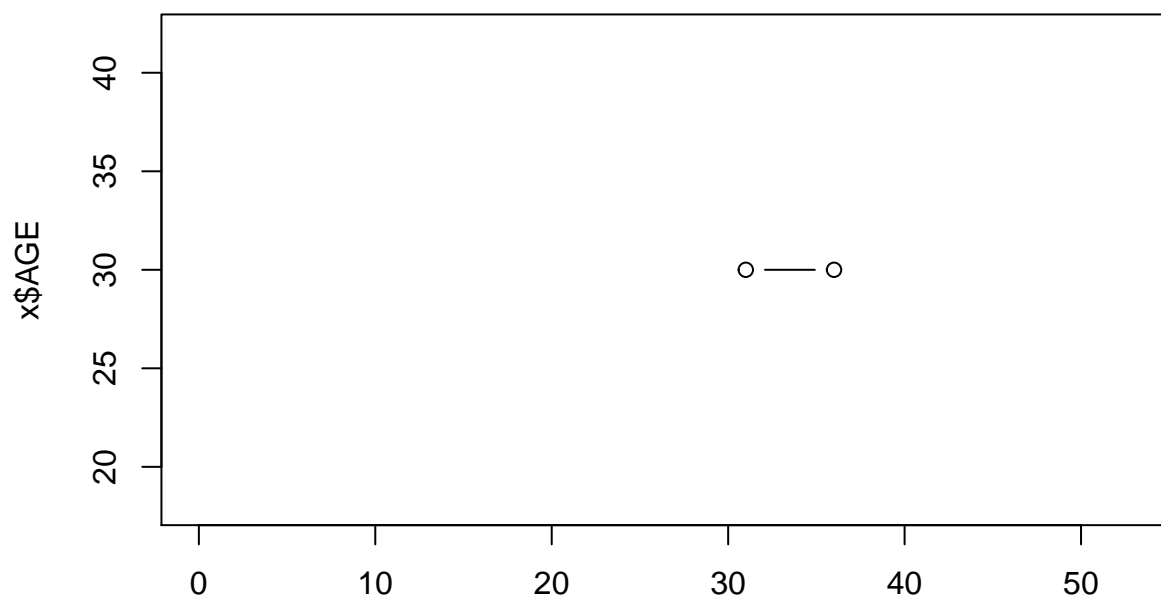
QUATA
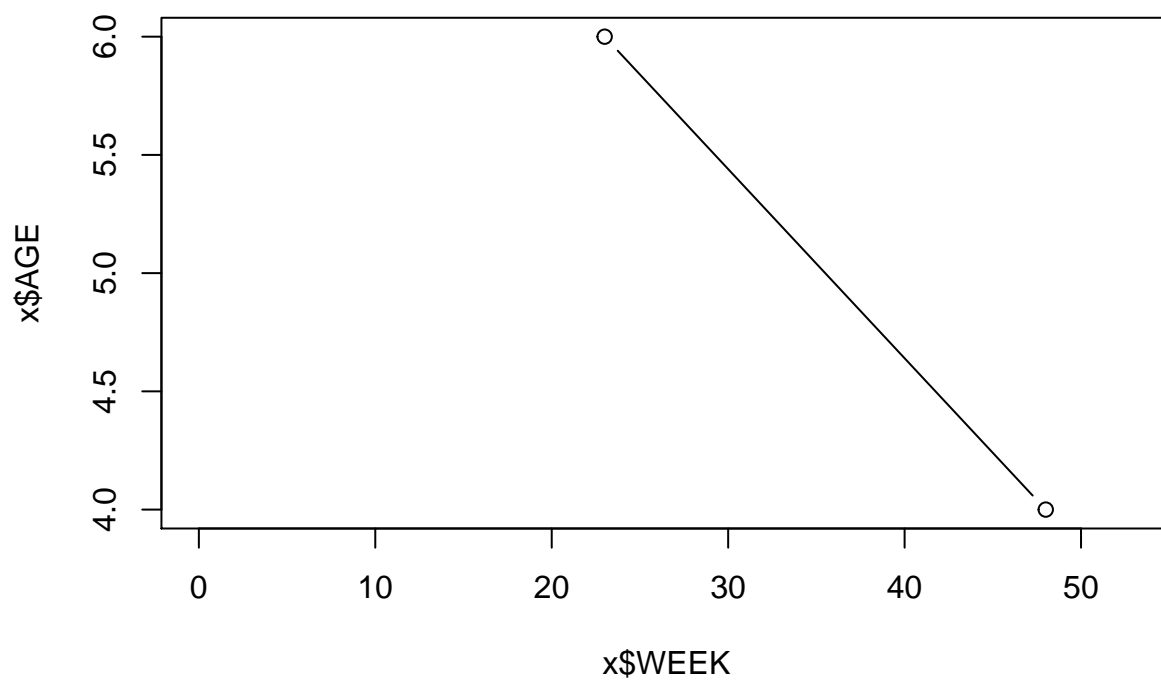
LIMEIRA
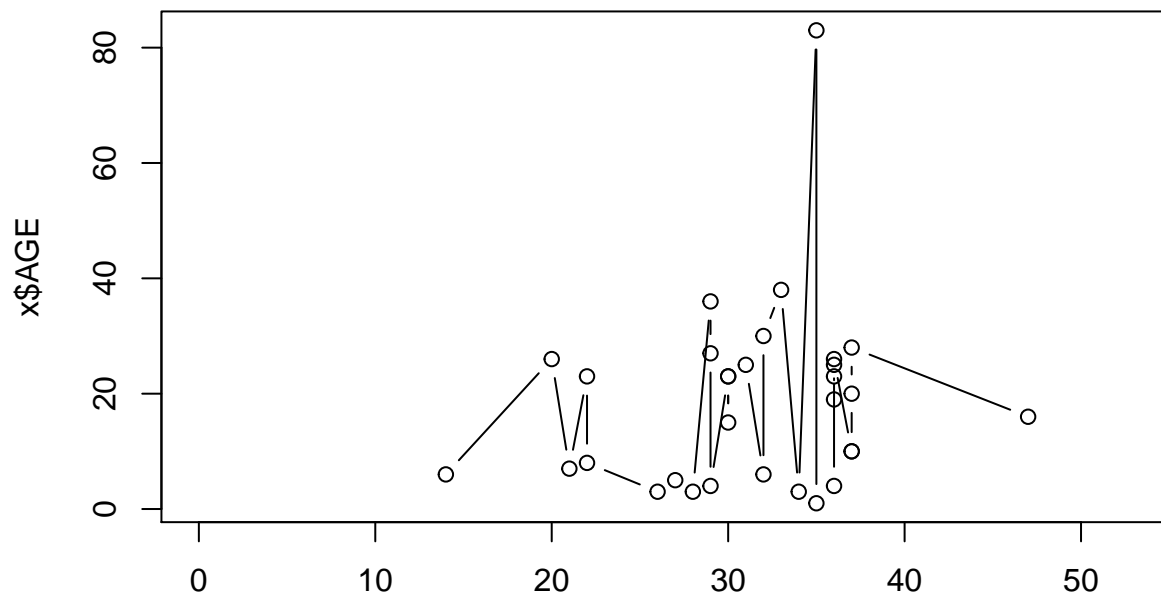
JUNDIAI

SANTA GERTRUDES

x$WEEK

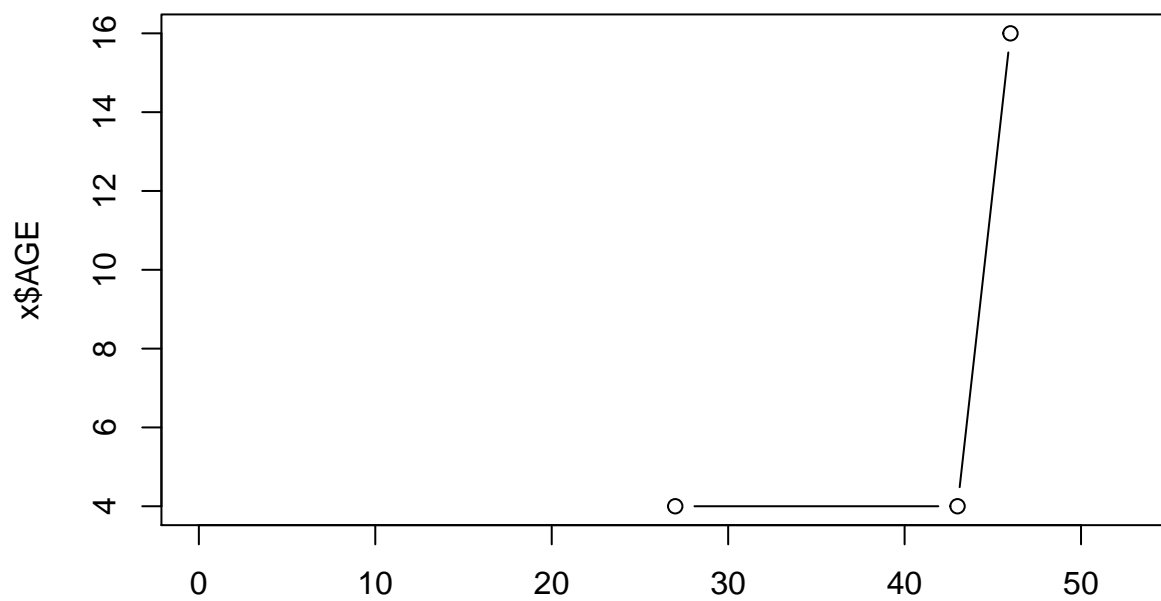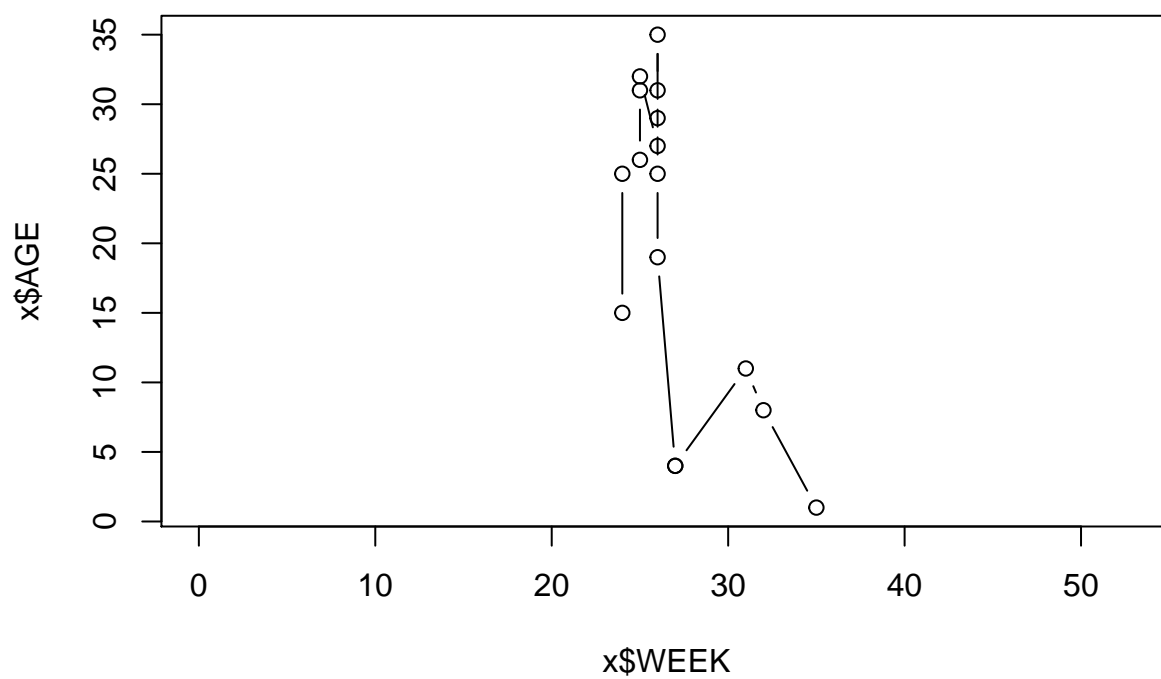**RIO CLARO**

x$WEEK

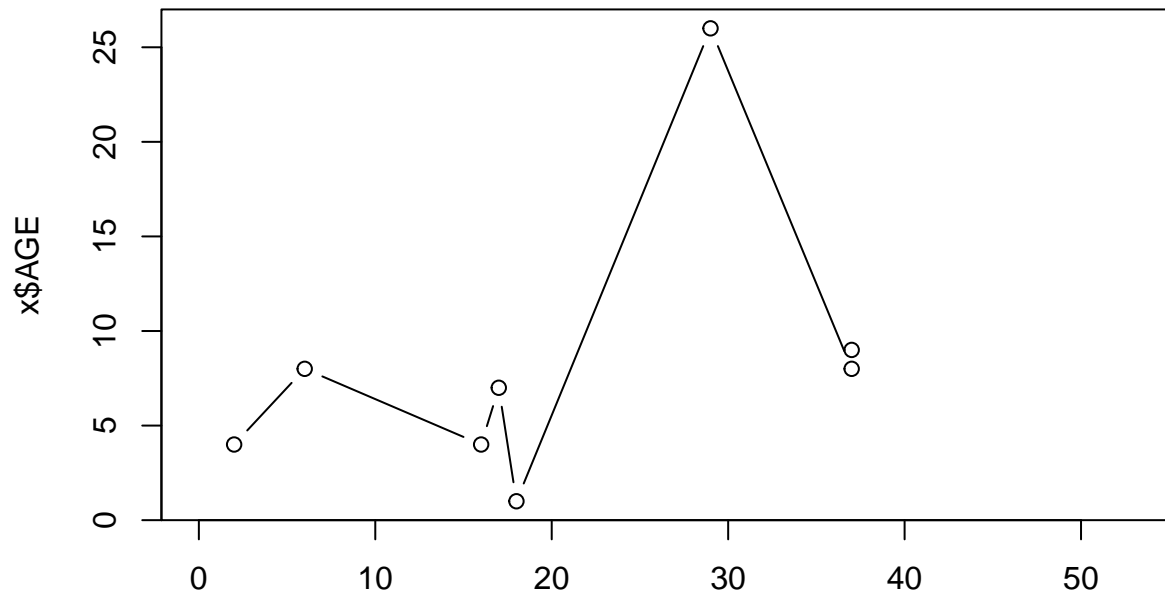# PIRACICABA

x$WEEK

PIRASSUNUNGA

x$WEEK
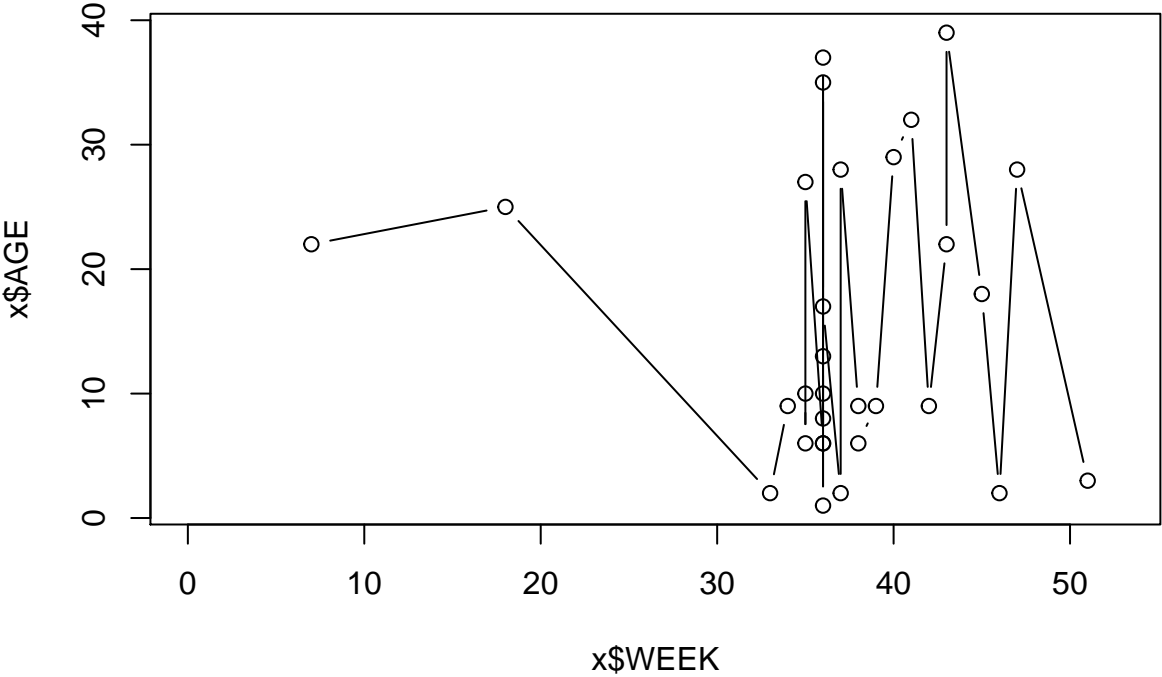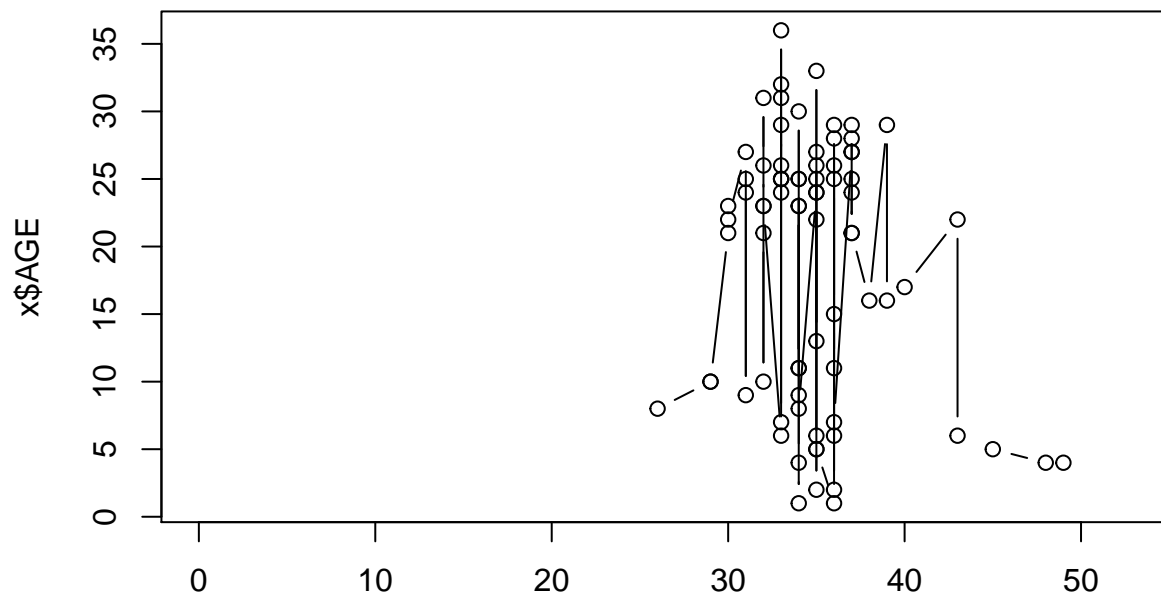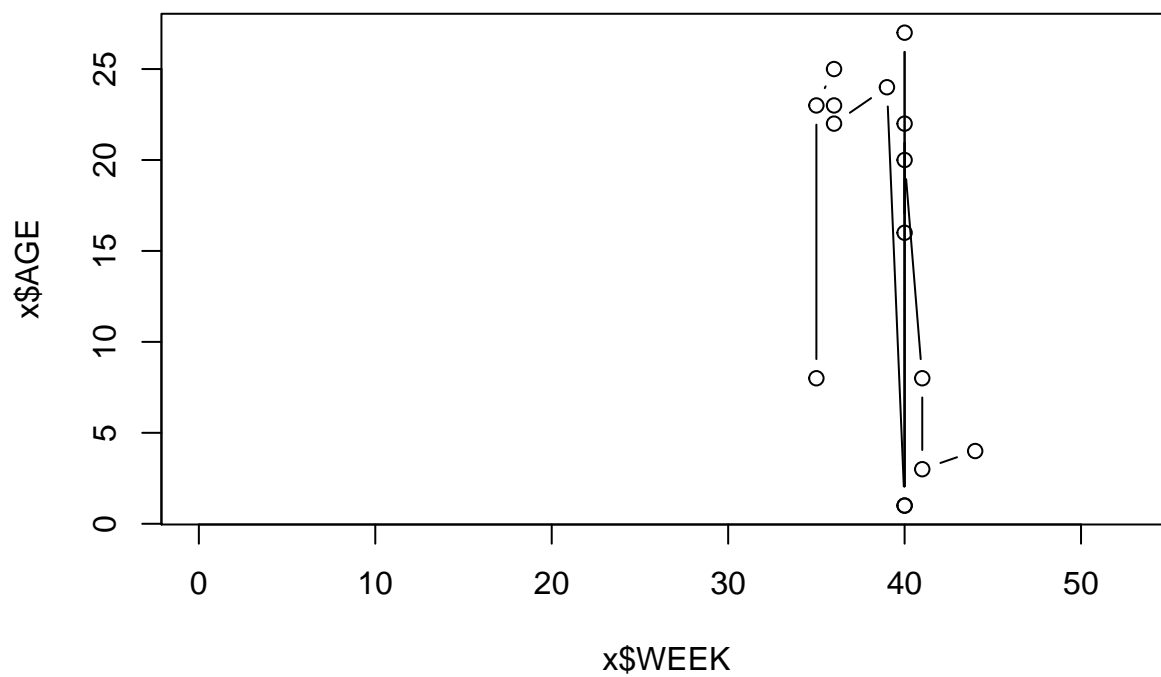
CORUMBATAI
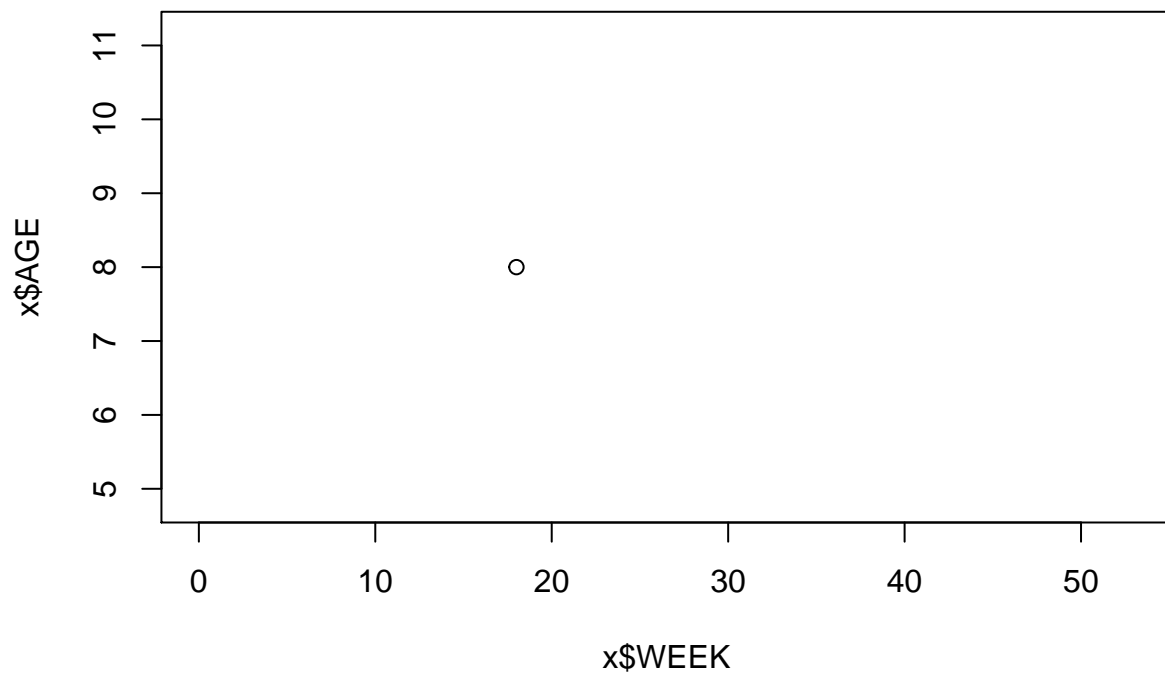

CESARIO LANGE
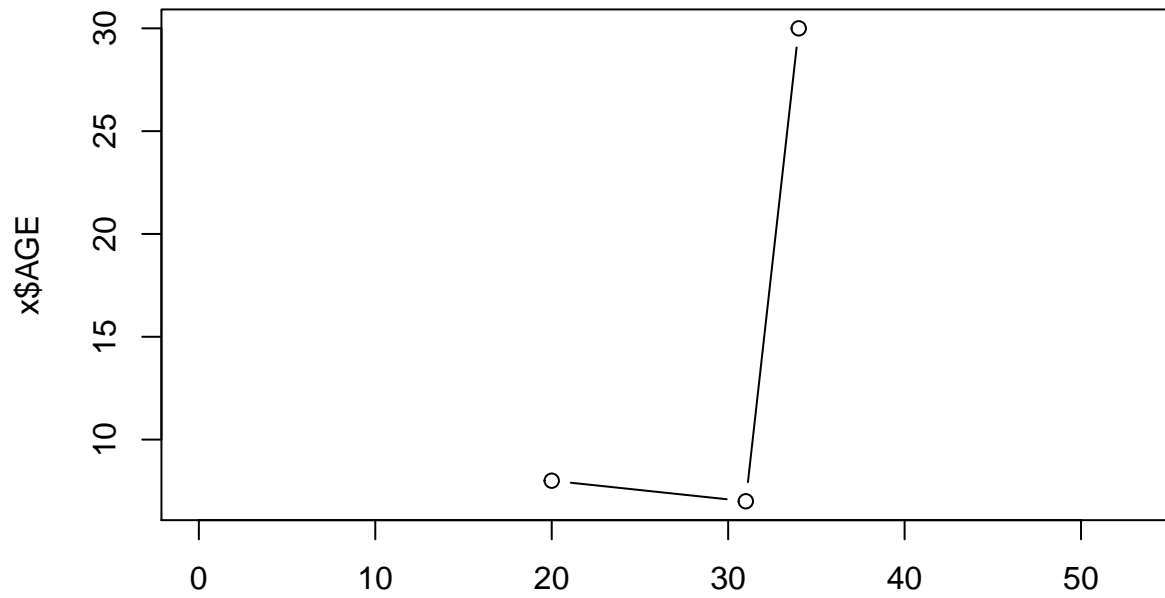
IRACEMAPOLIS

CHARQUEADA

ITIRAPINA

REGENTE FEIJO
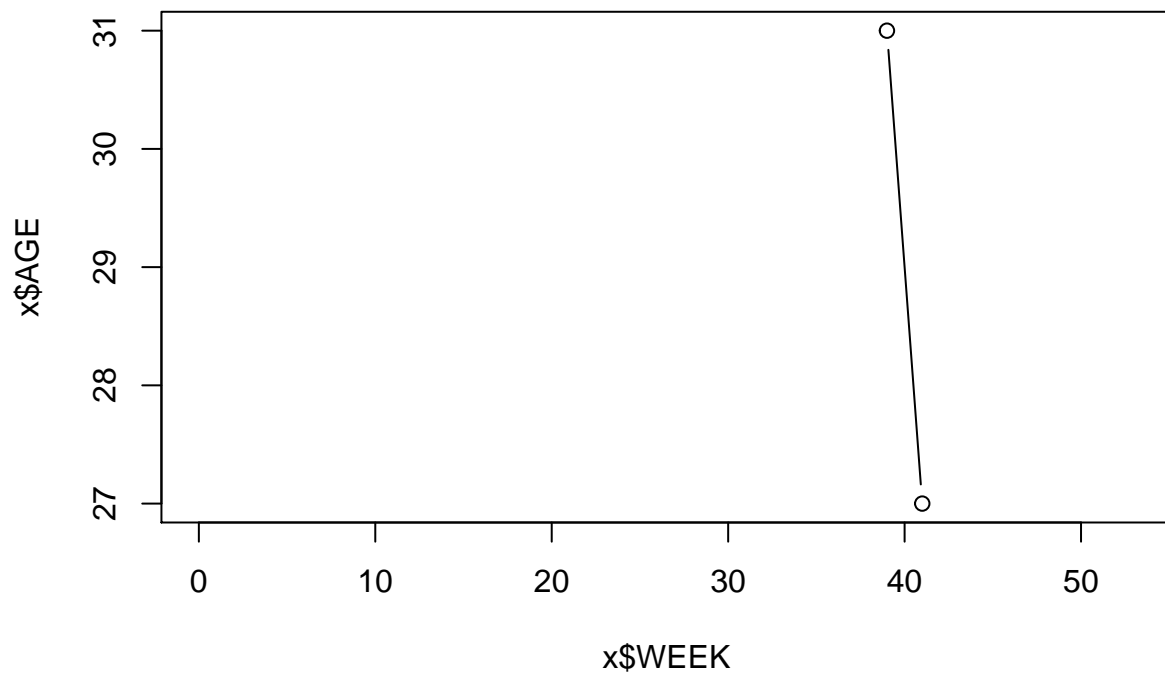
INDIANA
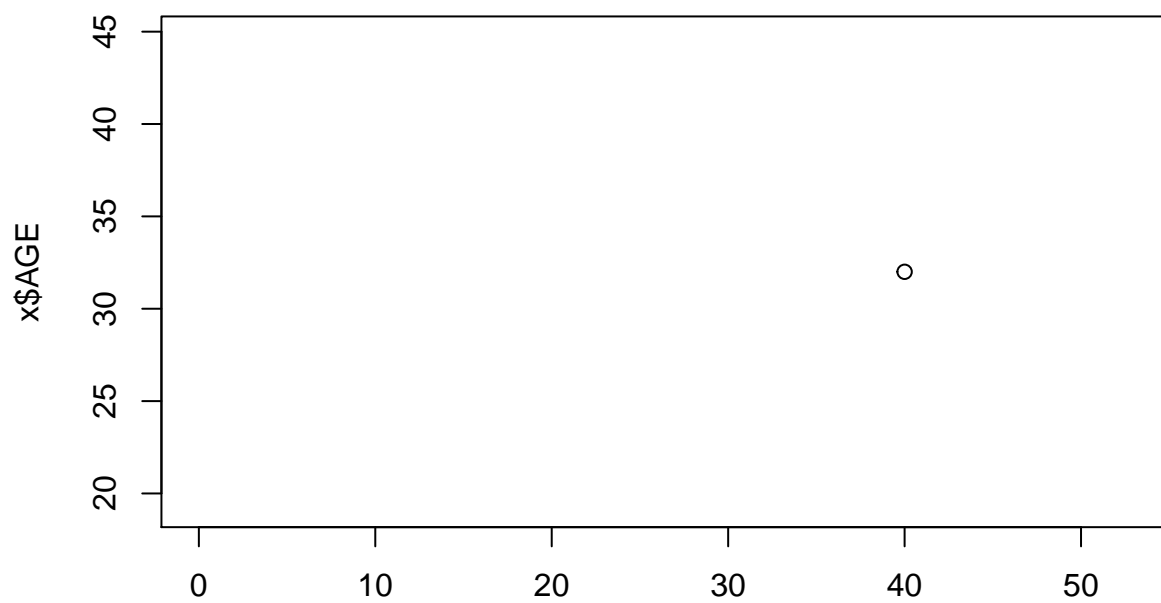
PRADOPOLIS
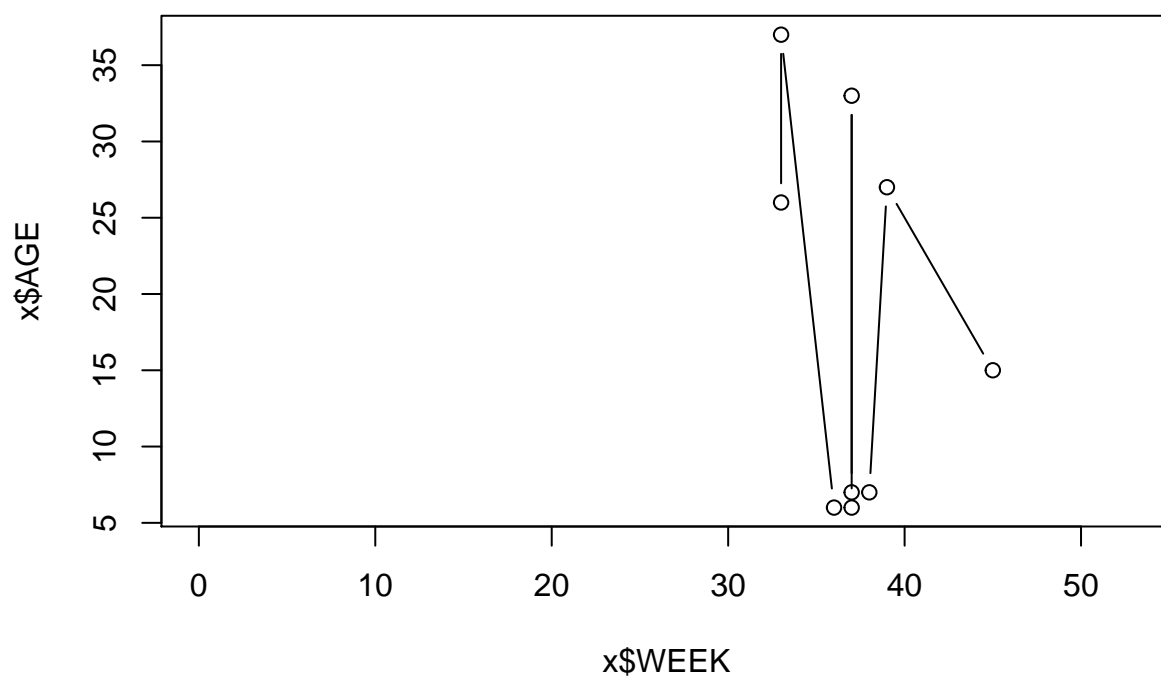
CRAVINHOS
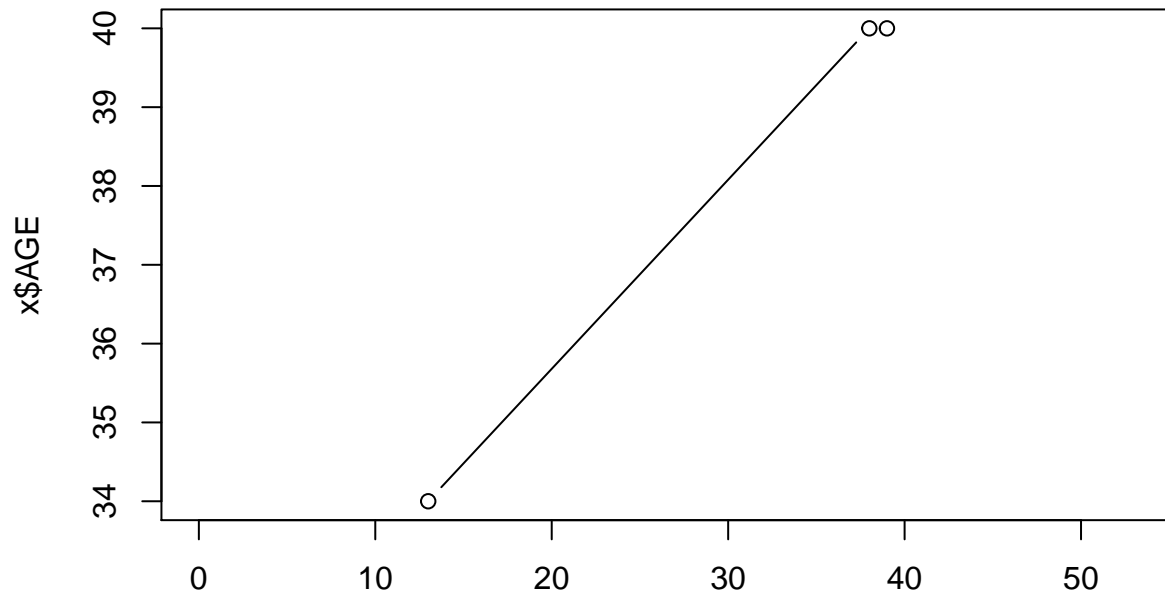
NOVA GRANADA
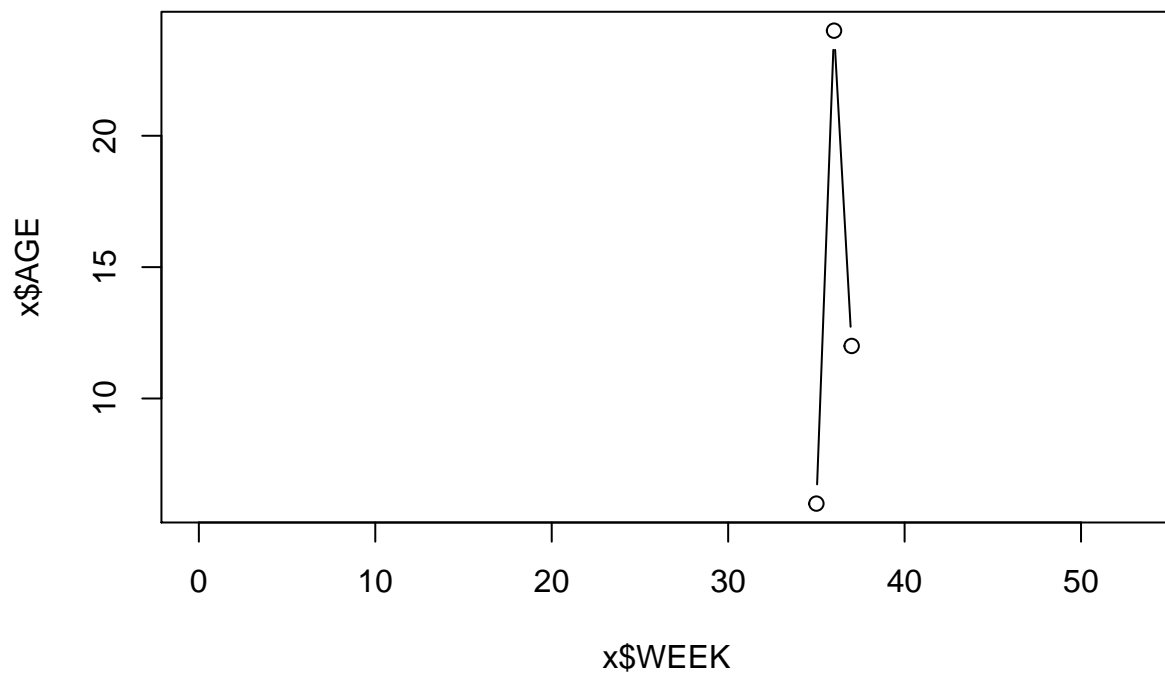
MONTE APRAZIVEL
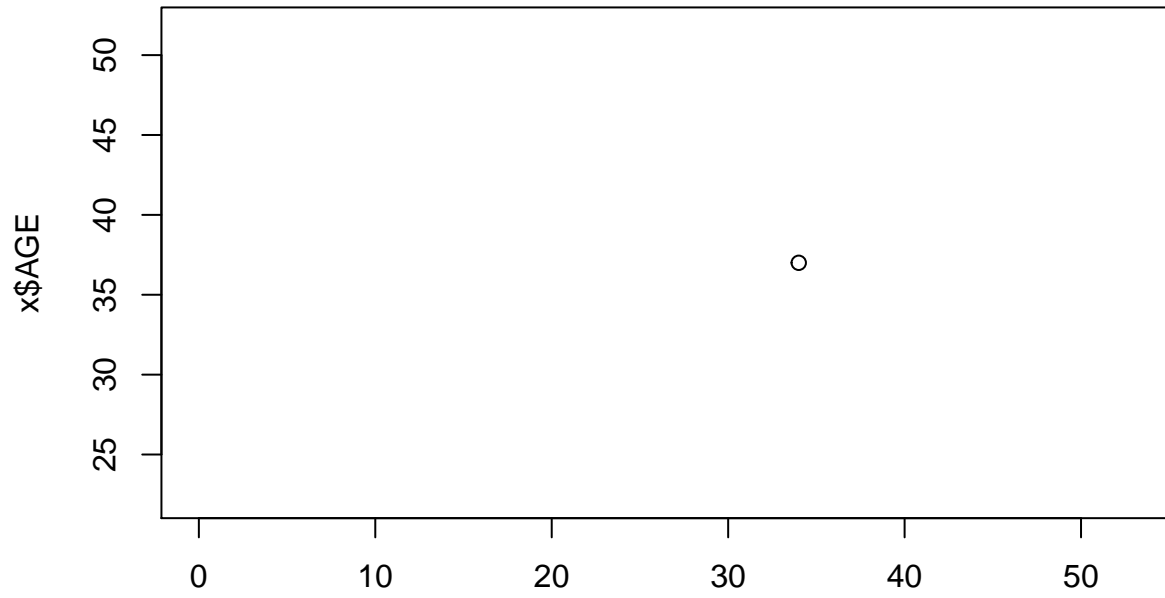
SANTA ADELIA
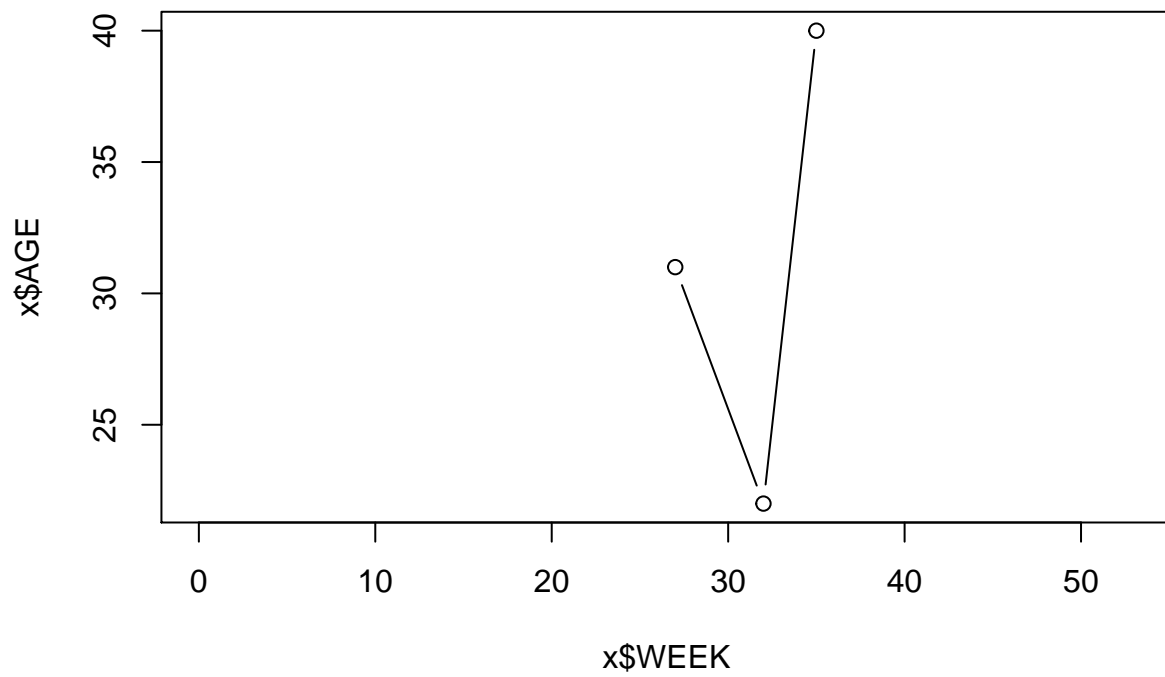
FERNANDOPOLIS
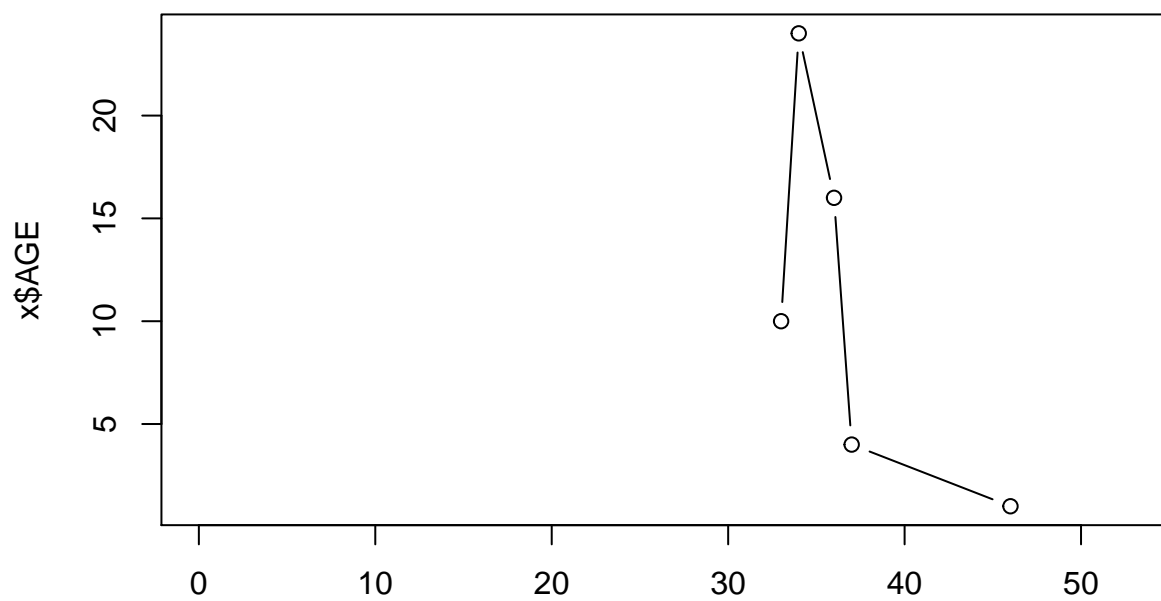
PINDORAMA
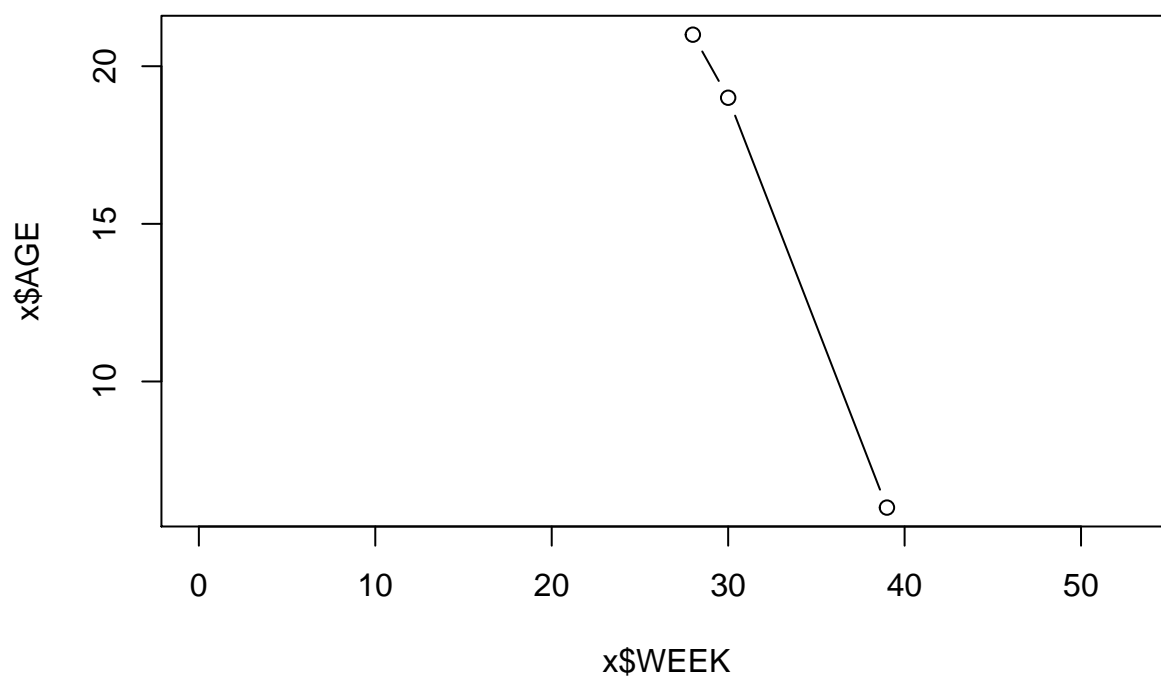
SANTA CLARA D'OESTE
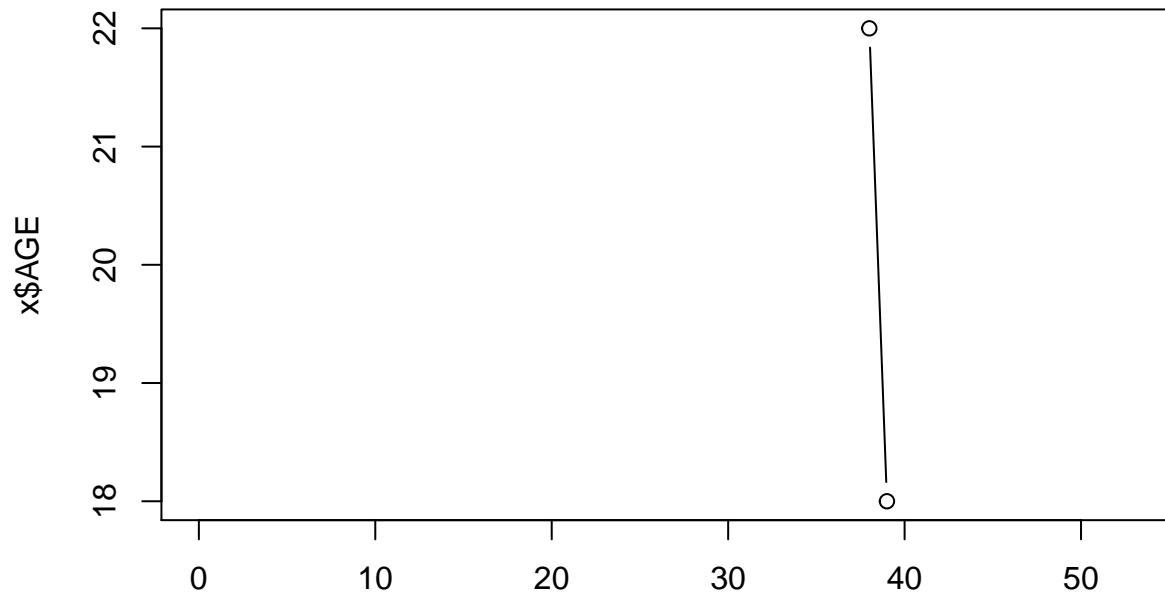
PALMEIRA D'OESTE
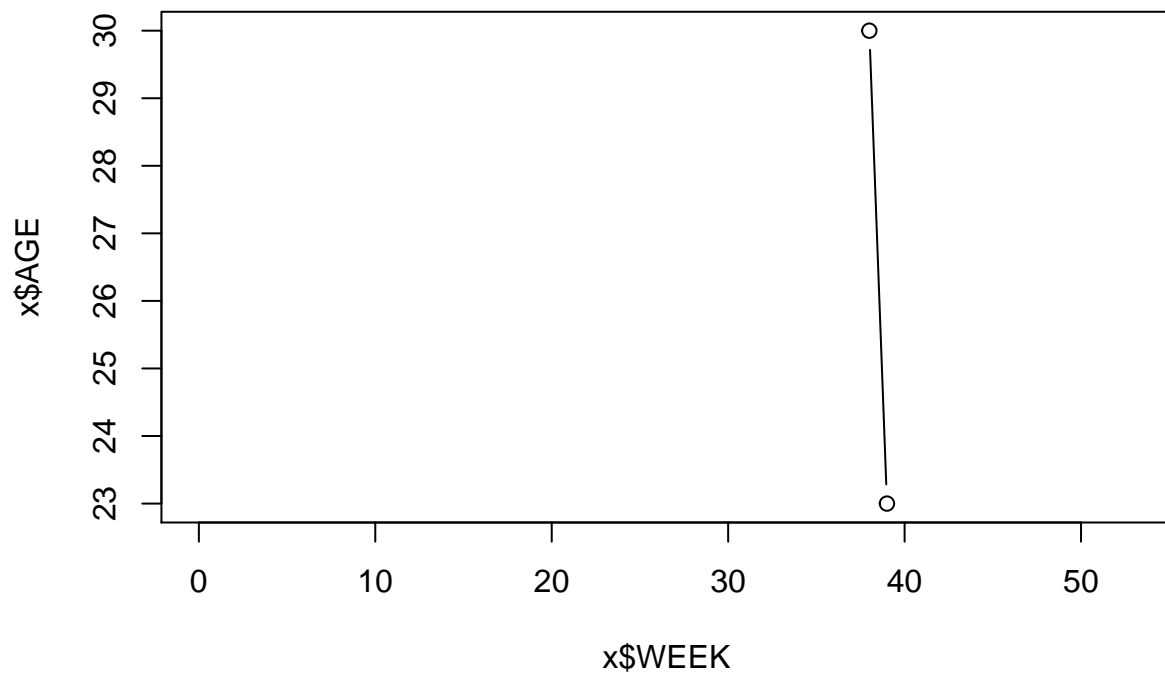
MIRA ESTRELA
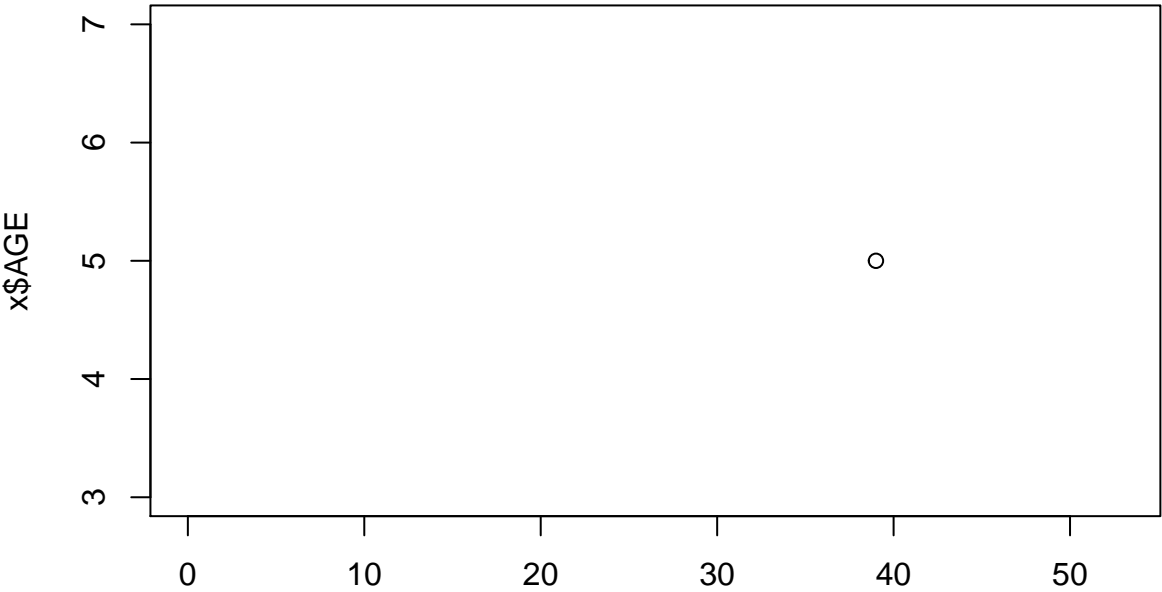
MACEDONIA
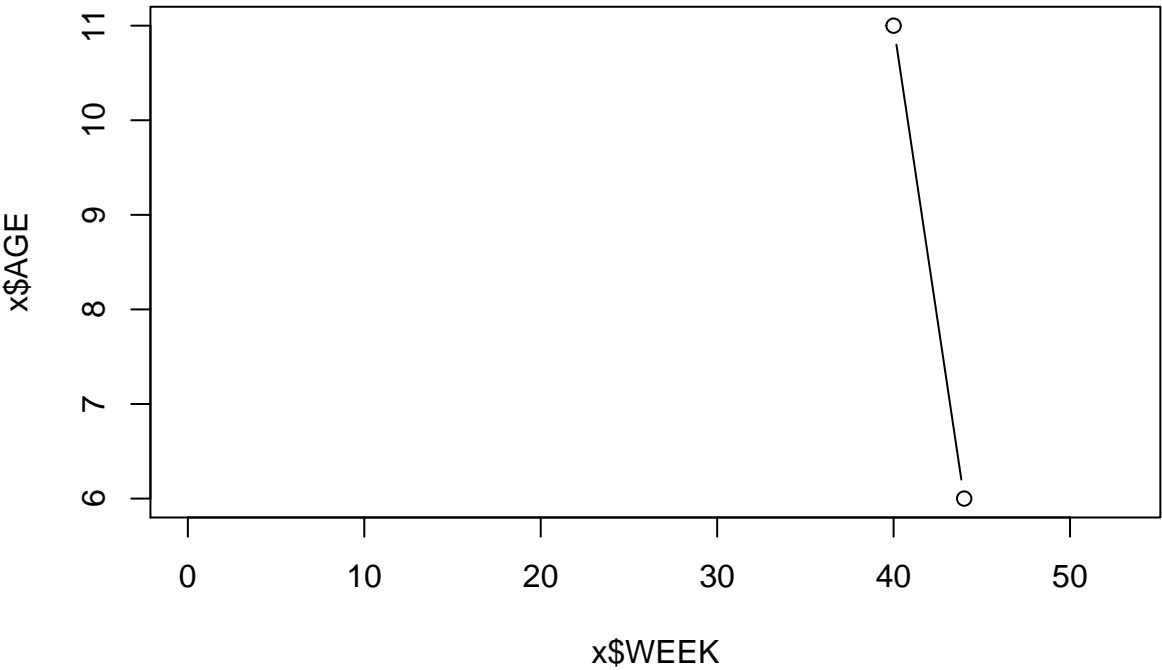
PARAISO
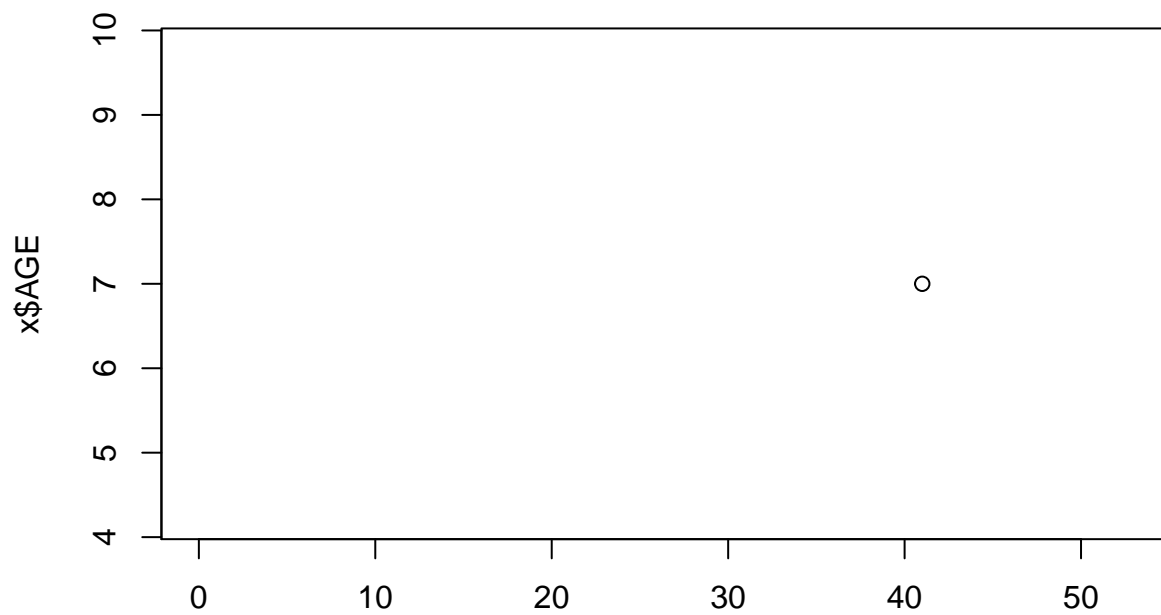
POLONI
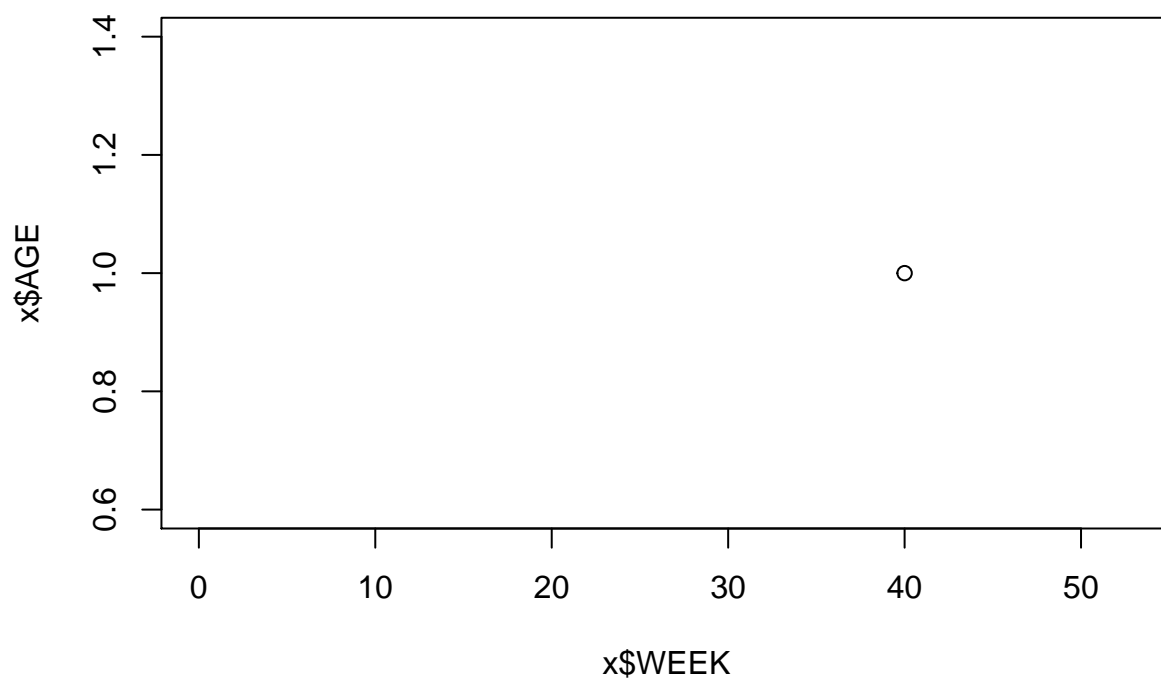
PEDRANOPOLIS
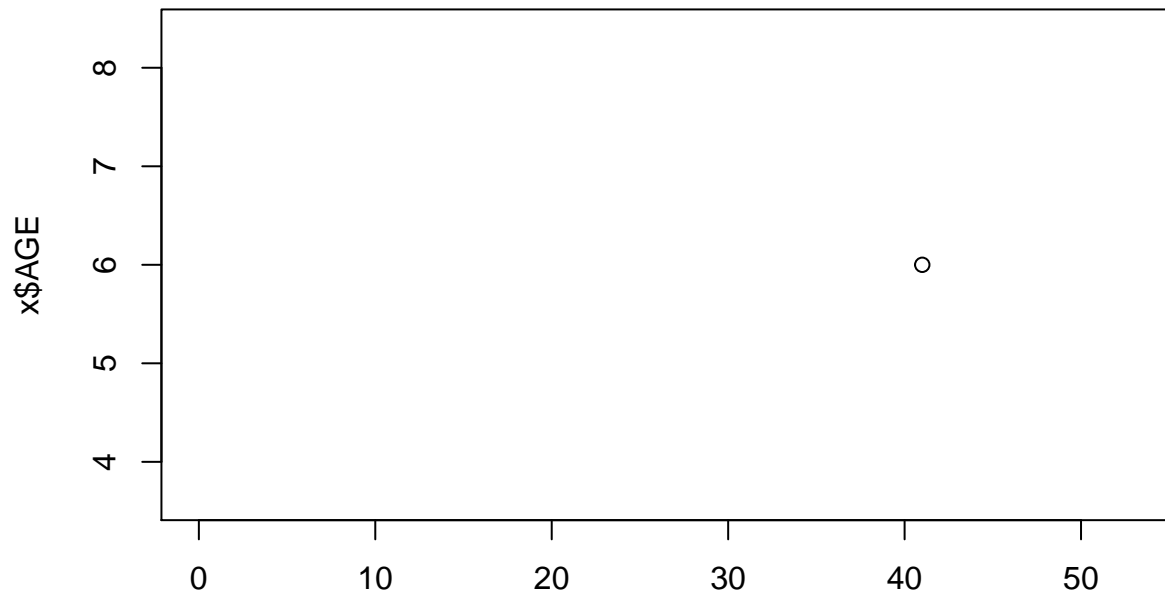
POTIRENDABA
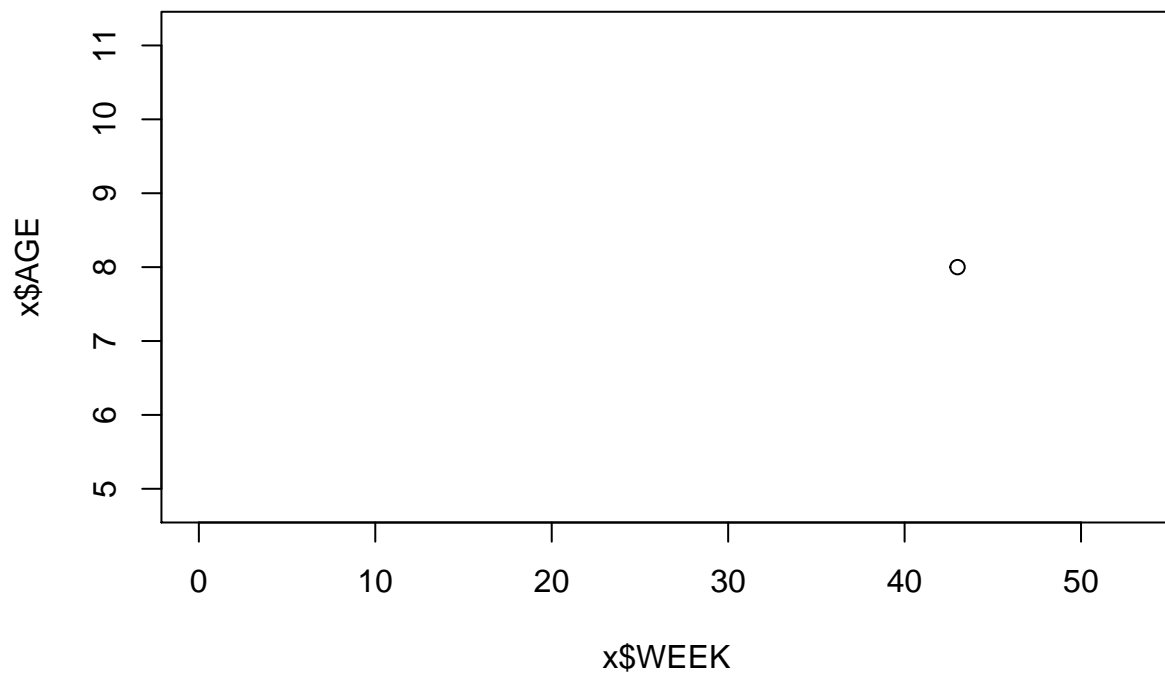
ARACOIABA DA SERRA
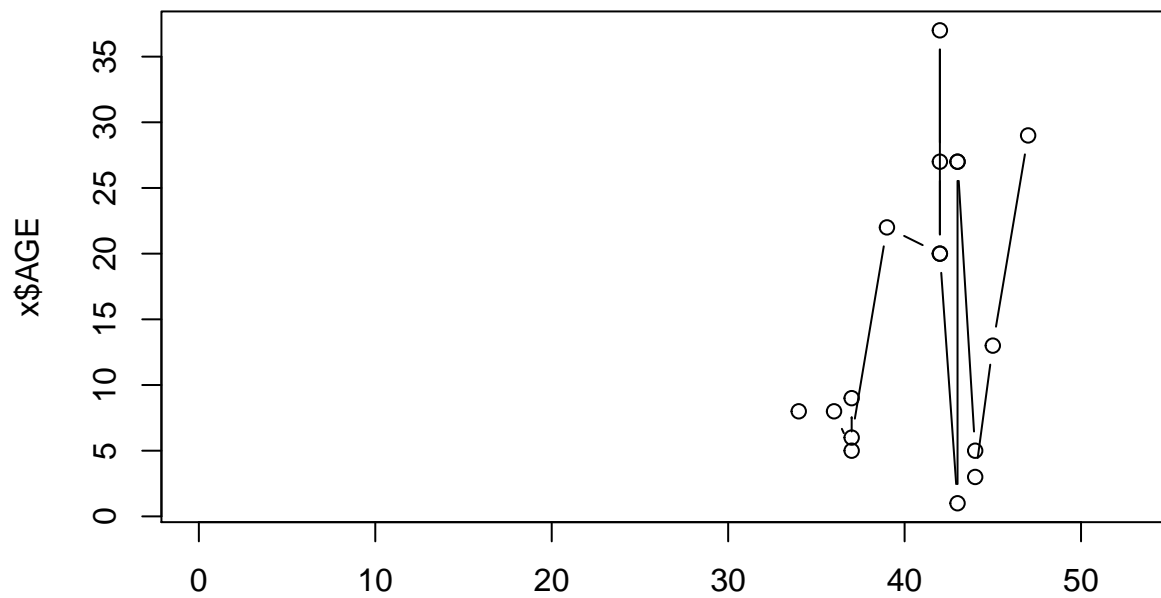
ARACARIGUAMA
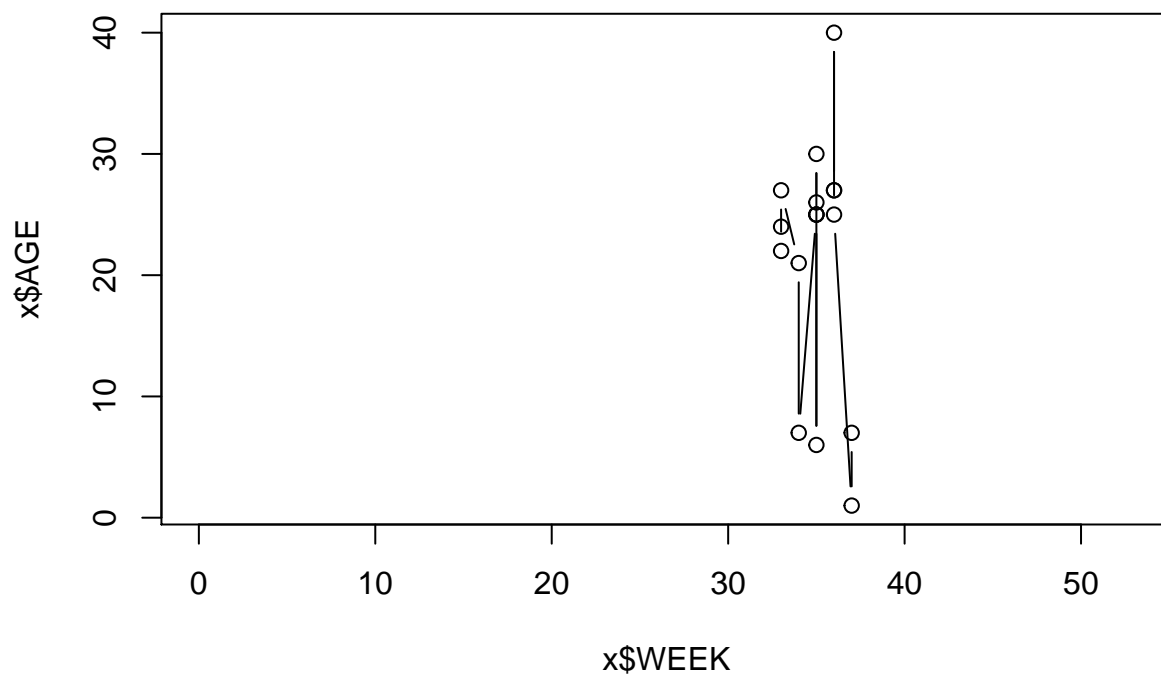
PIEDADE
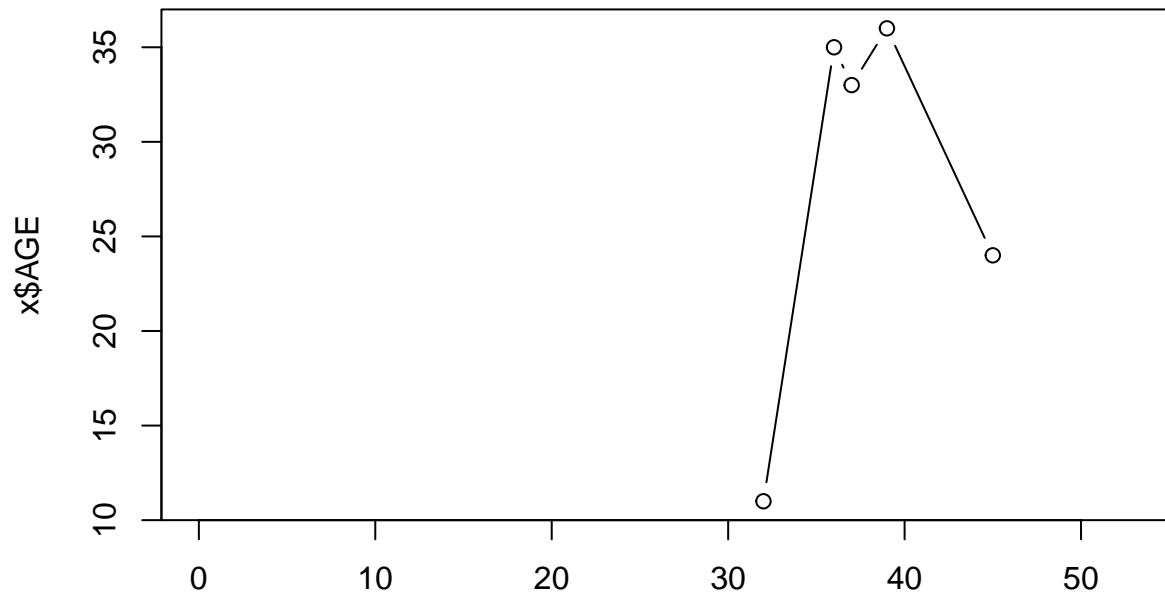
SALTO DE PIRAPORA
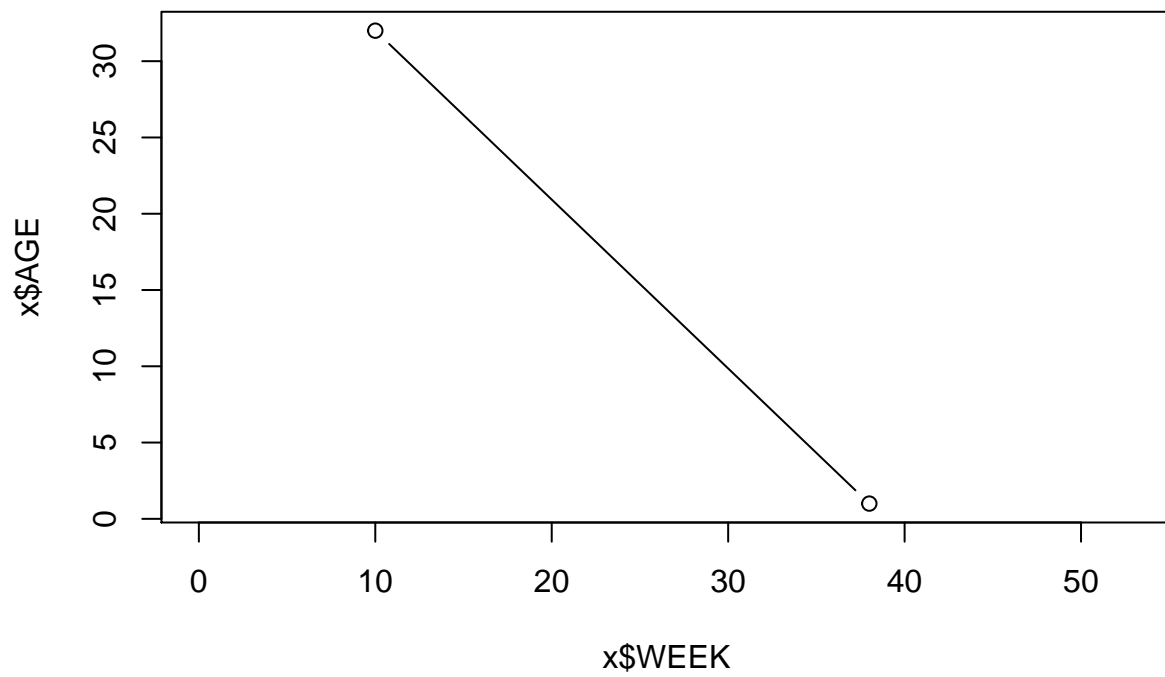
BURI
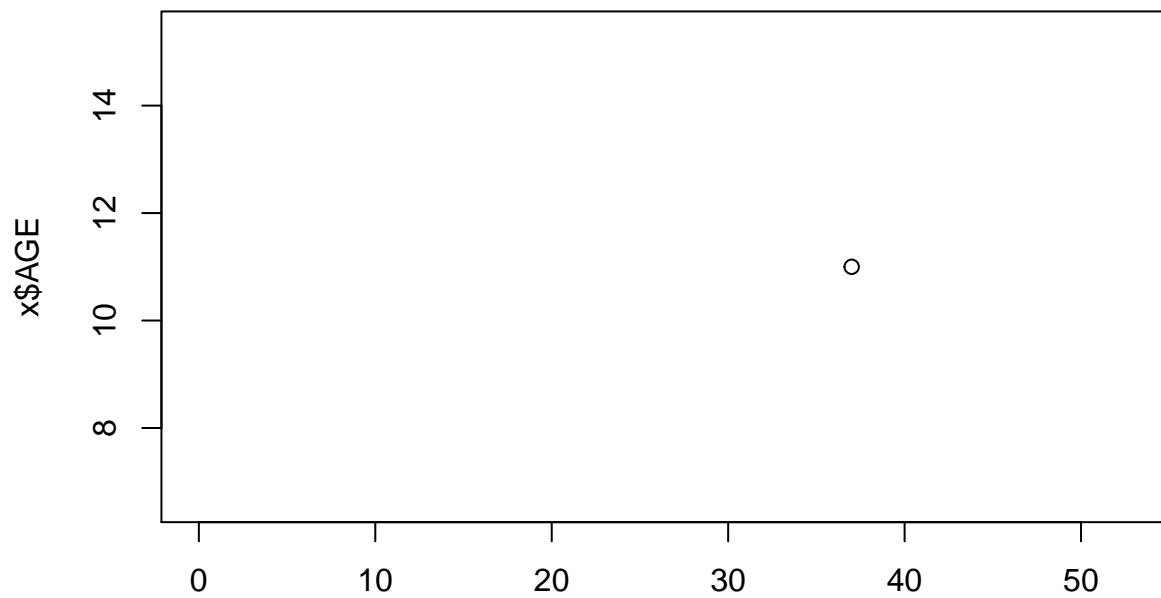
PIRAPORA DO BOM JESUS

MONTE MOR

x$AGE

x$WEEK

JARINU

x$AGE

x$WEEK
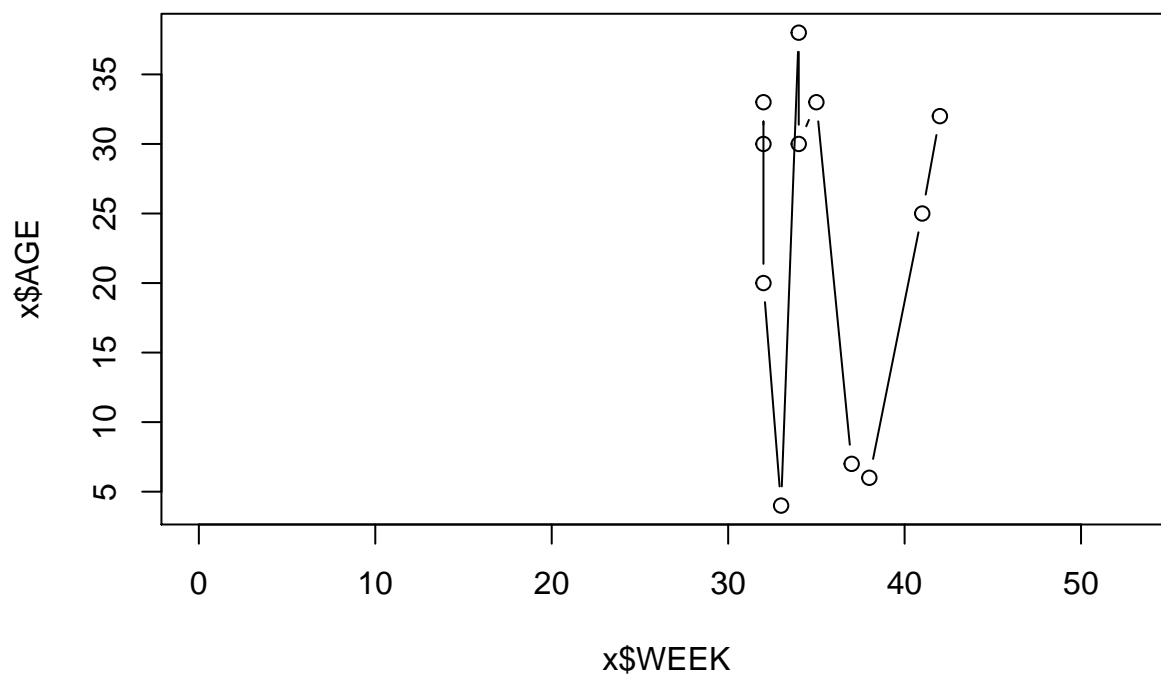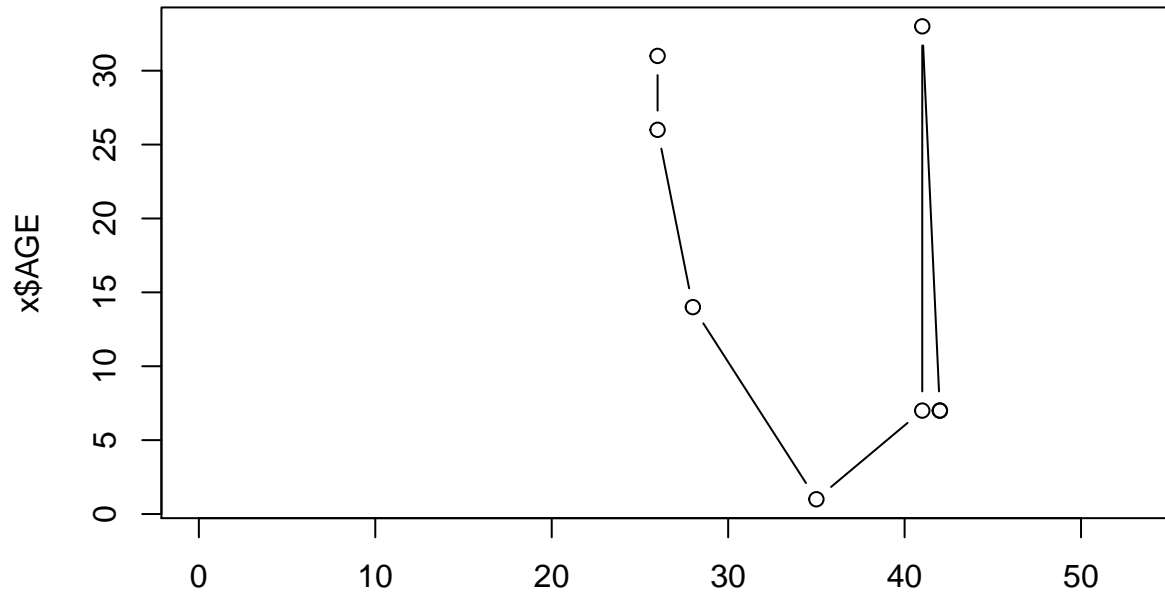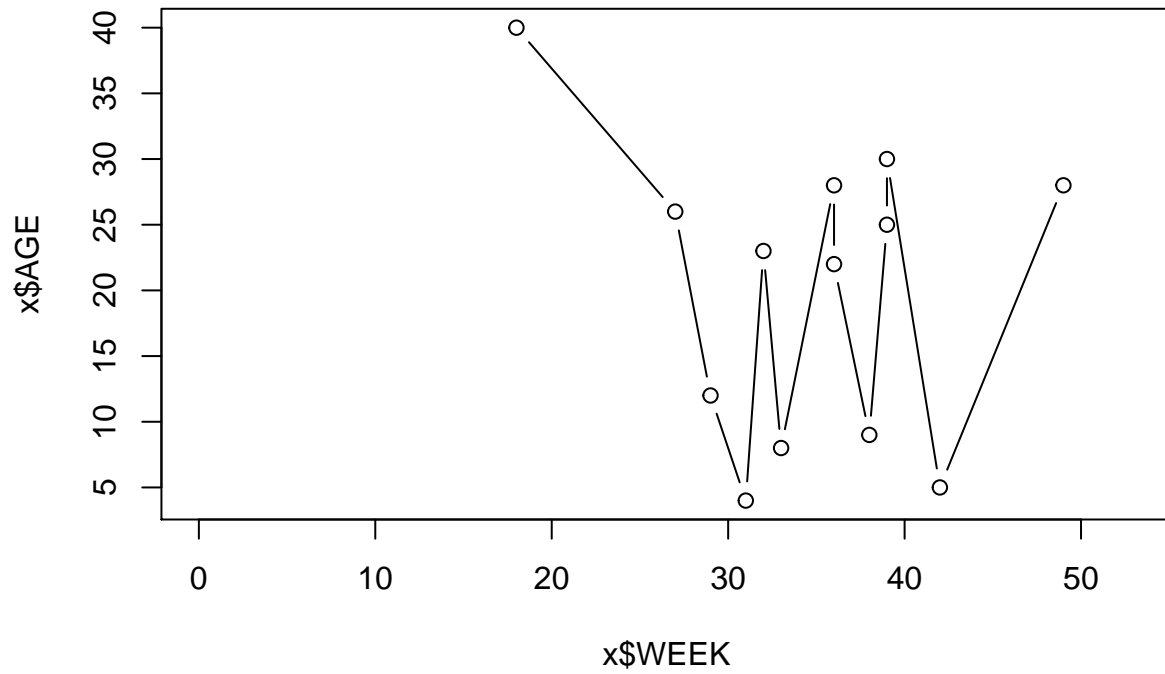
ITAPETININGA
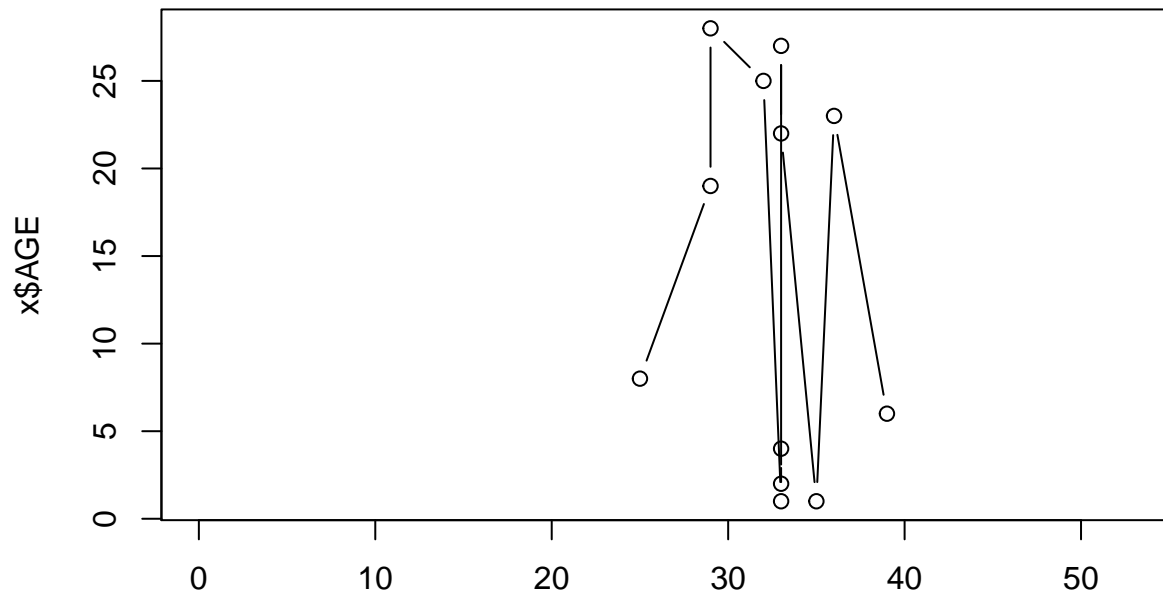
PEDREGULHO
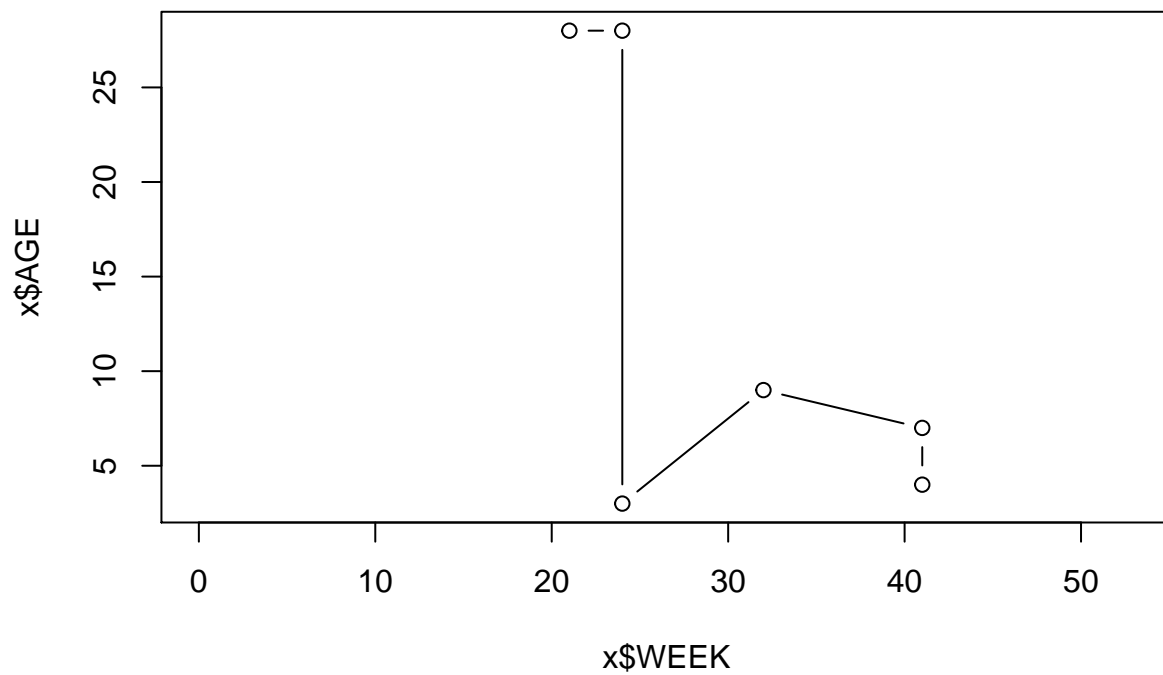
SALESOPOLIS
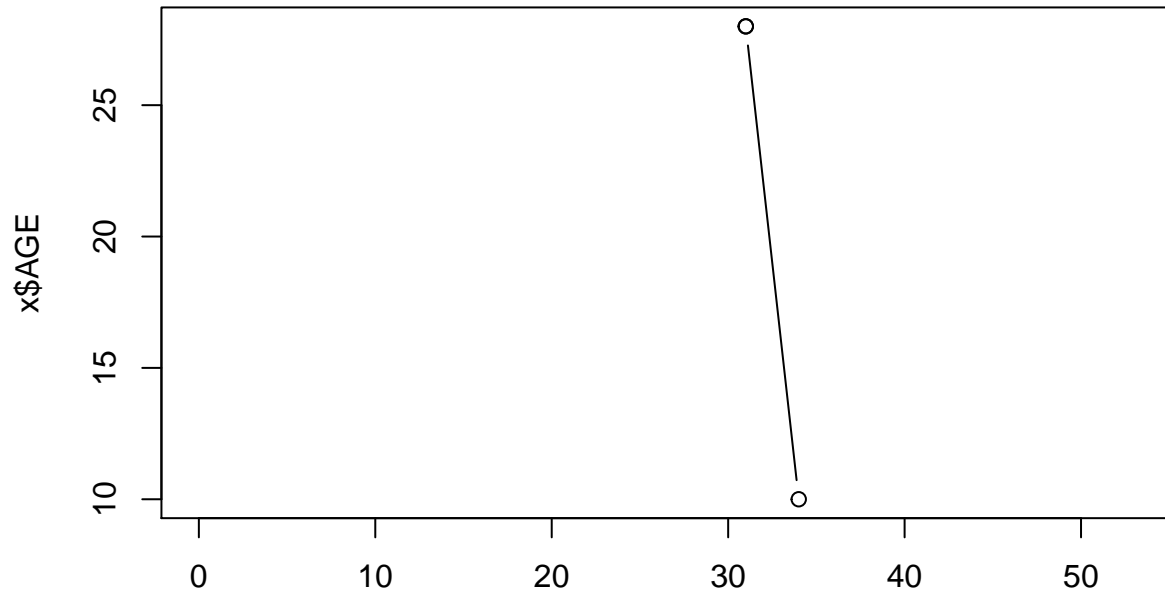
AMPARO
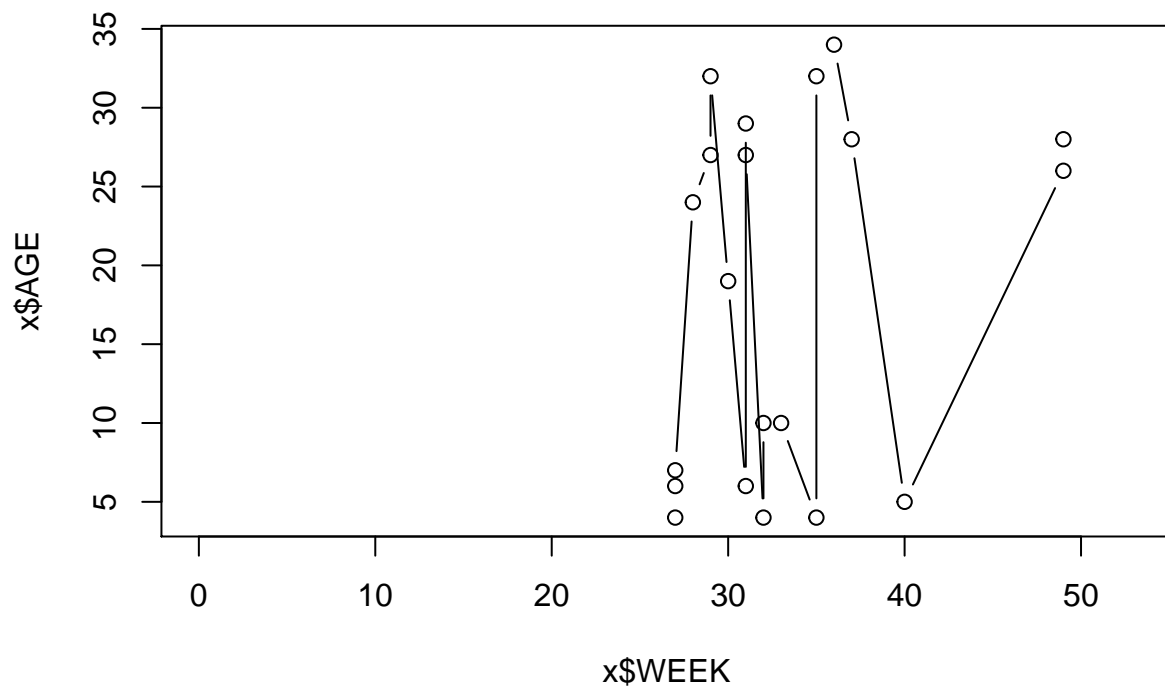
ARUJA
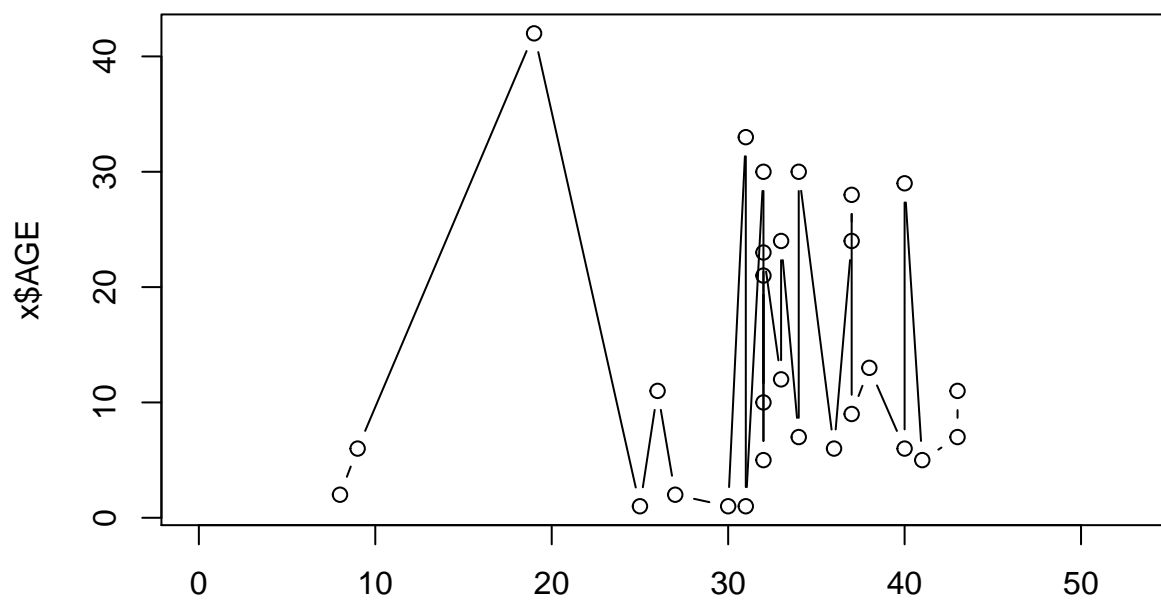
OURINHOS
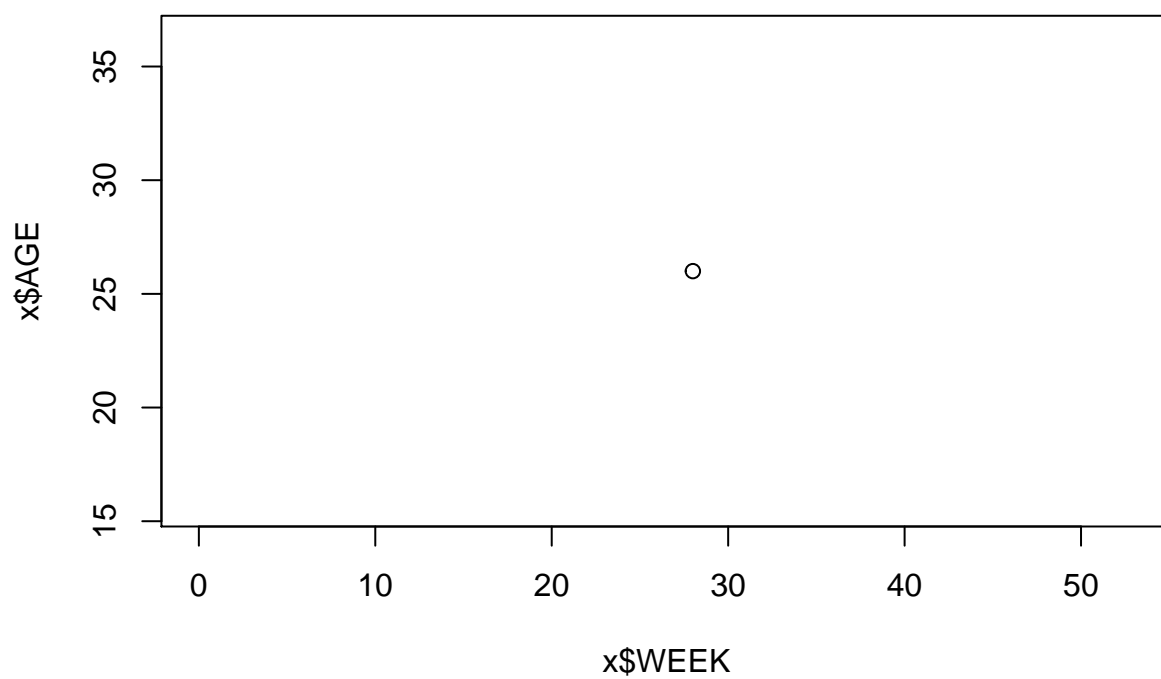
PIRAJUI
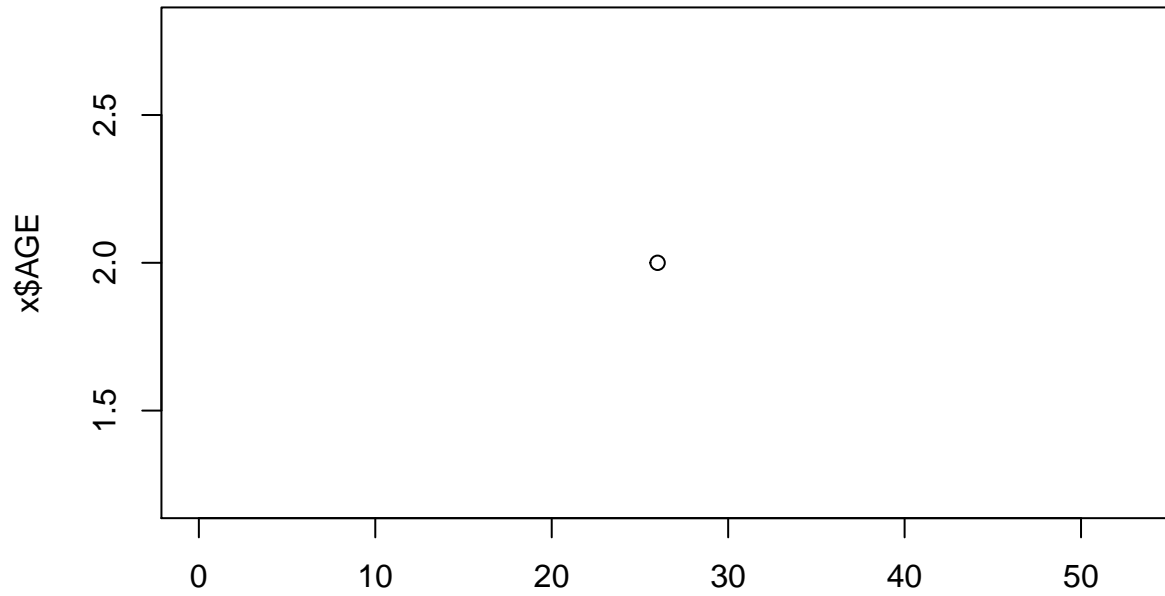
BRAGANCA PAULISTA
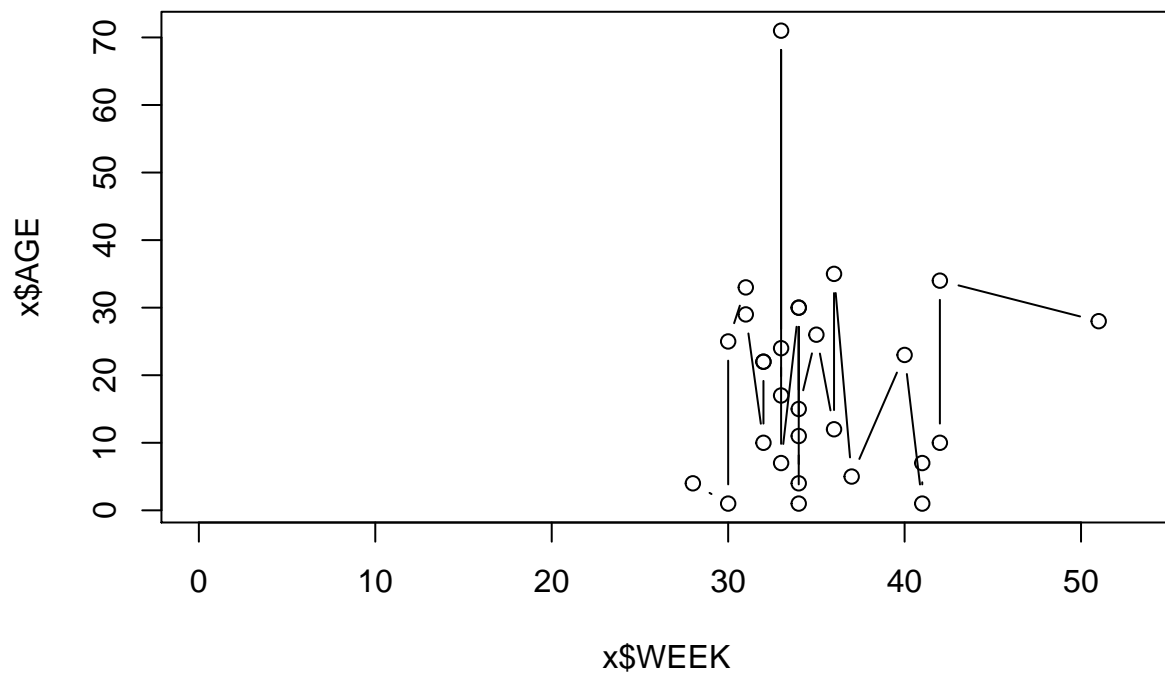
GUARAREMA
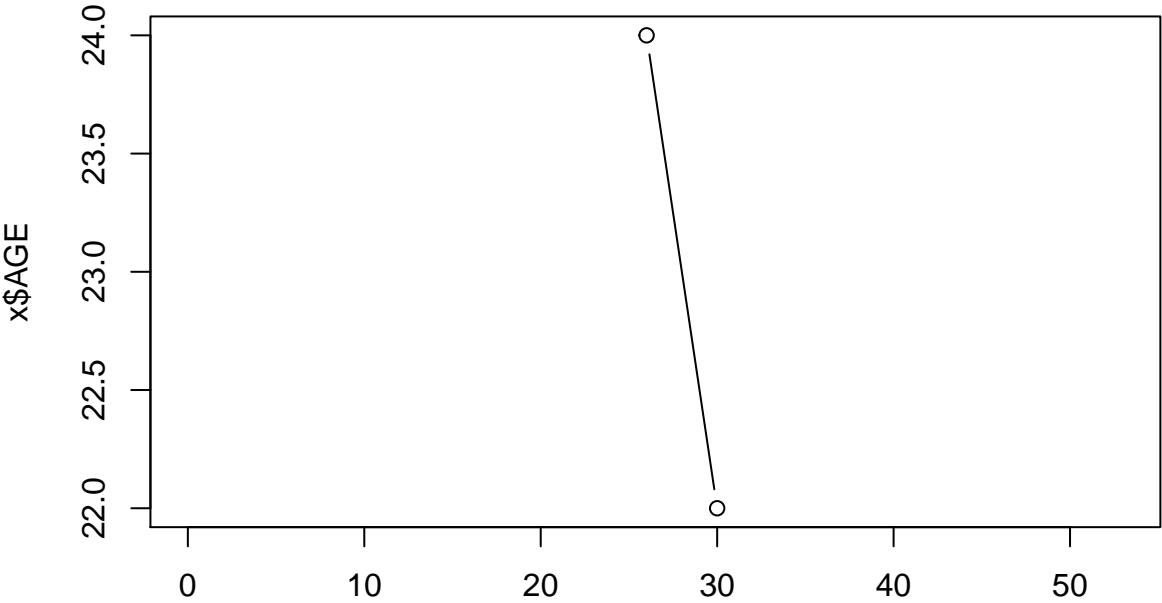
CAMPO LIMPO PAULISTA

TABOAO DA SERRA

IEPE

ITAPEVA

x$WEEK

NOVA ODESSA

x$WEEK

108
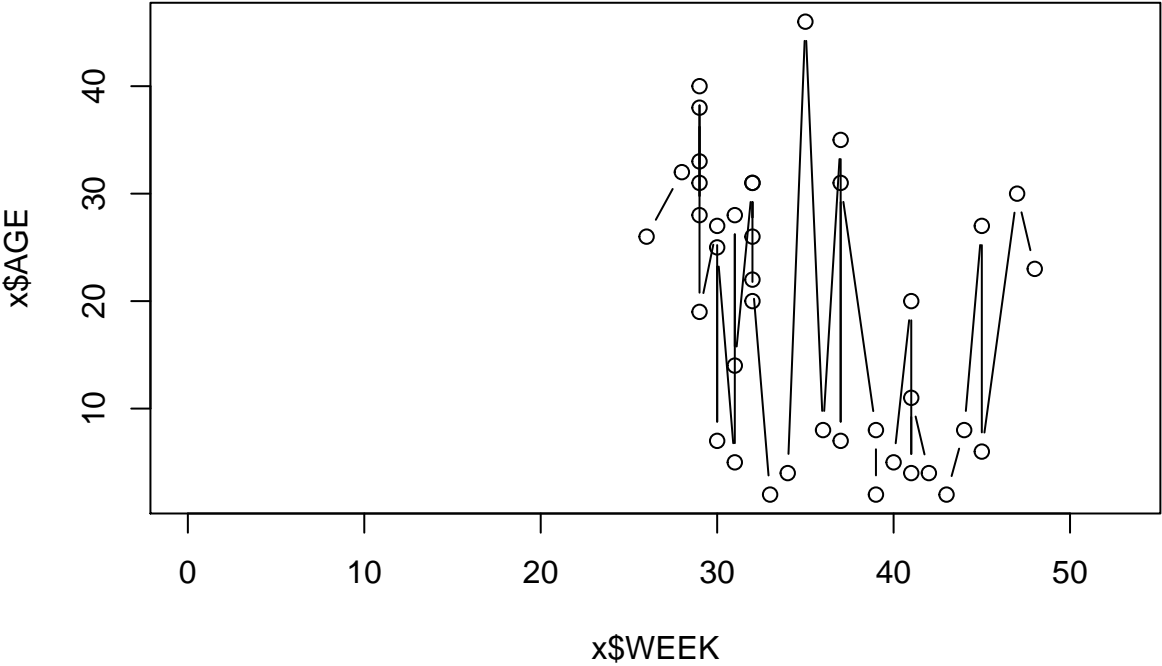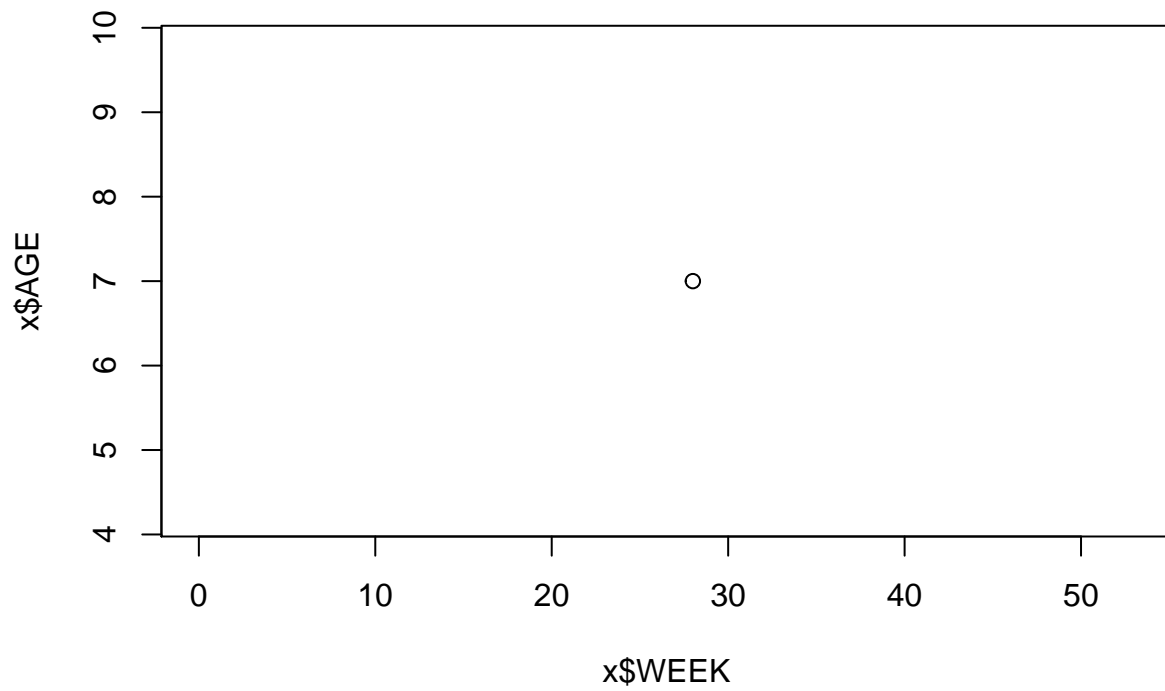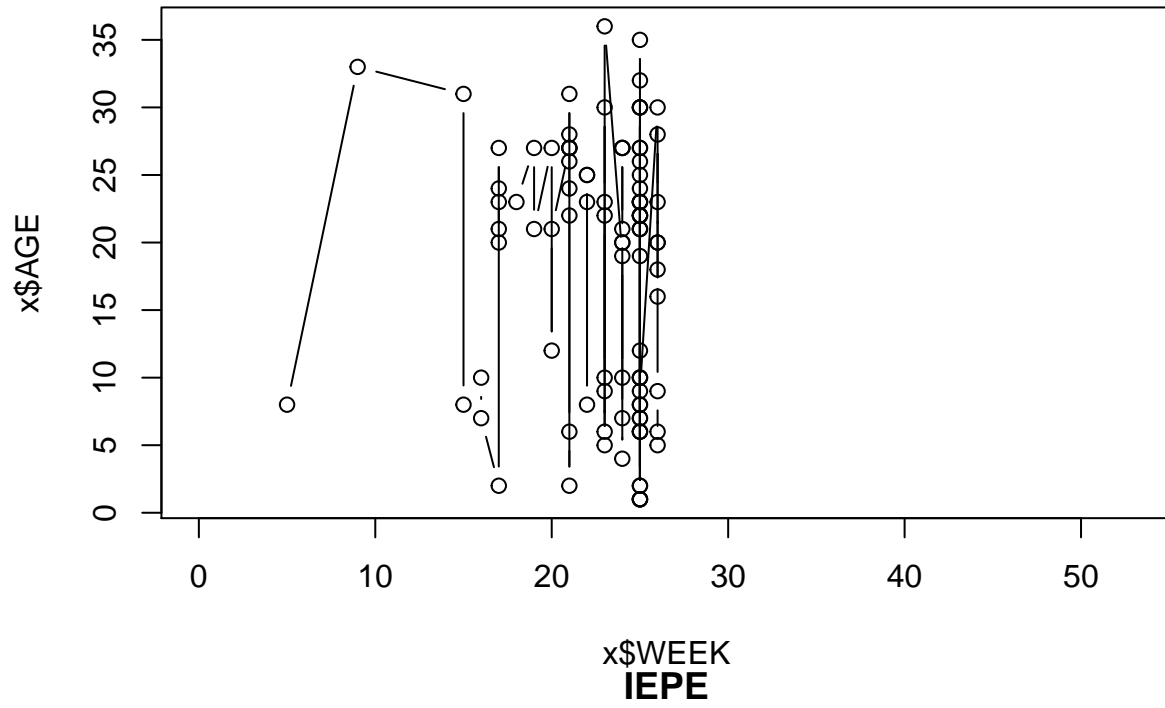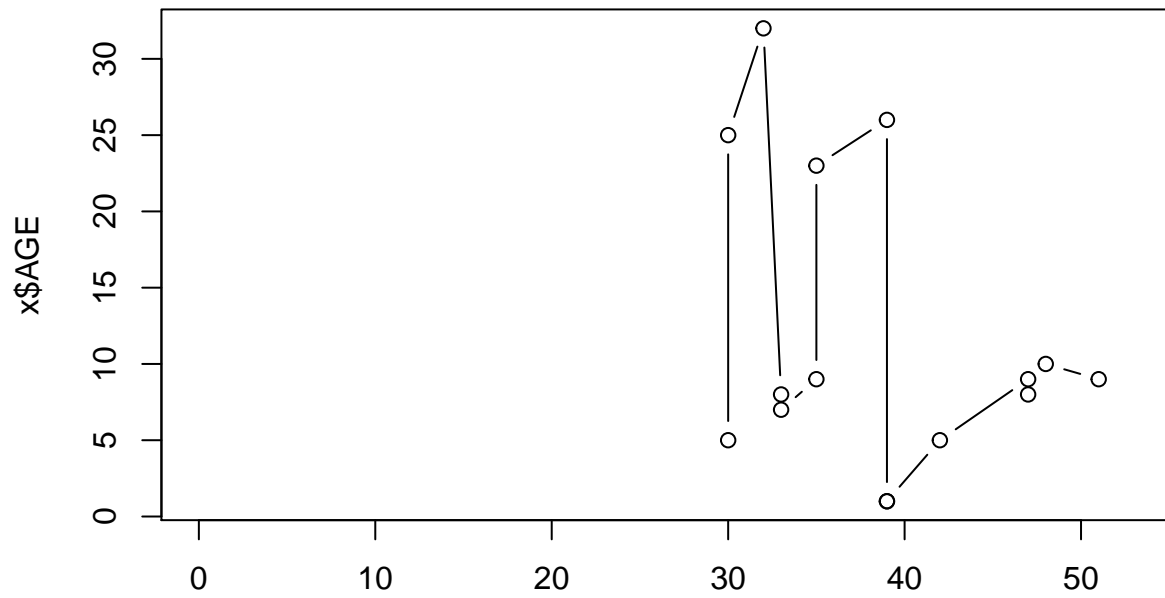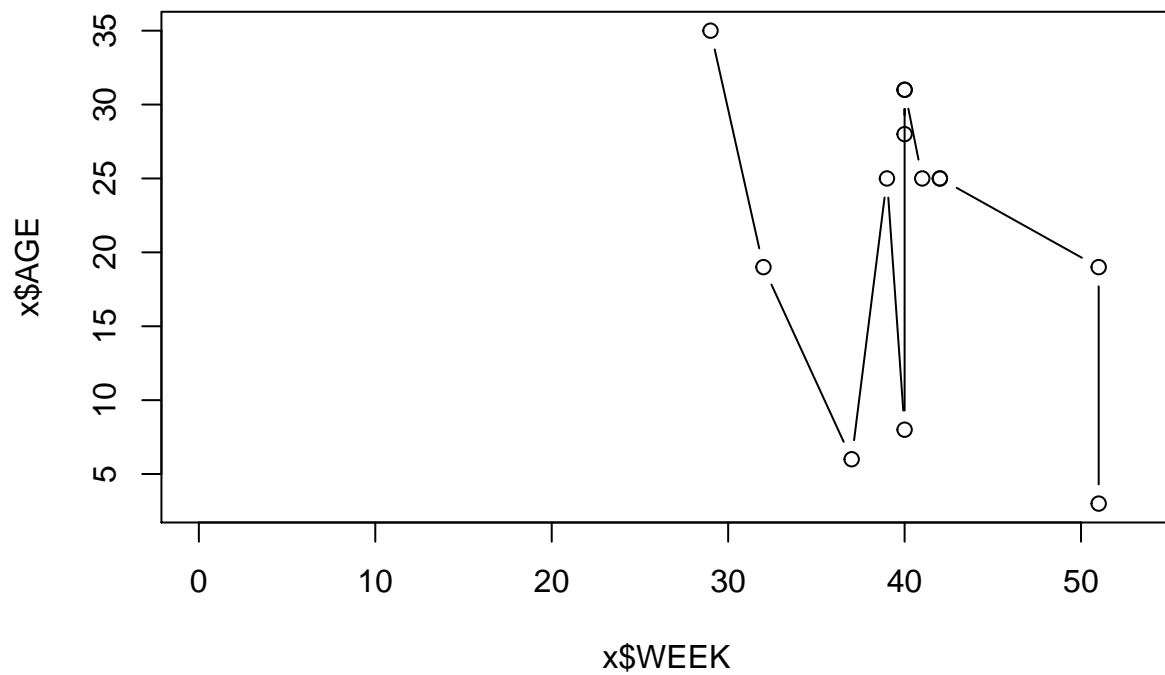
ITATIBA
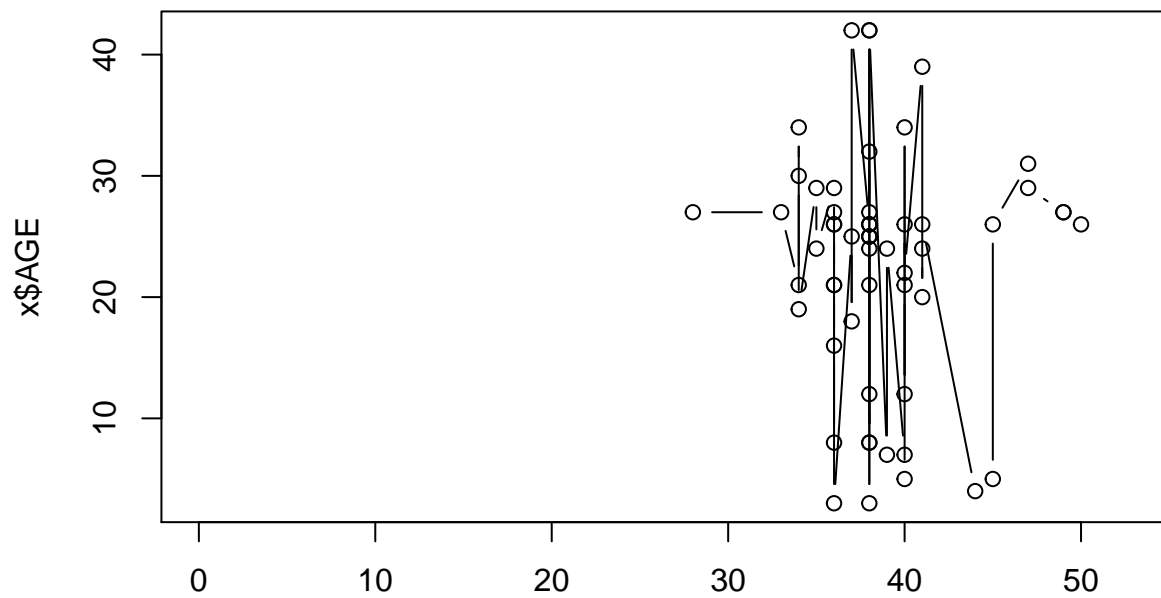
RANCHARIA
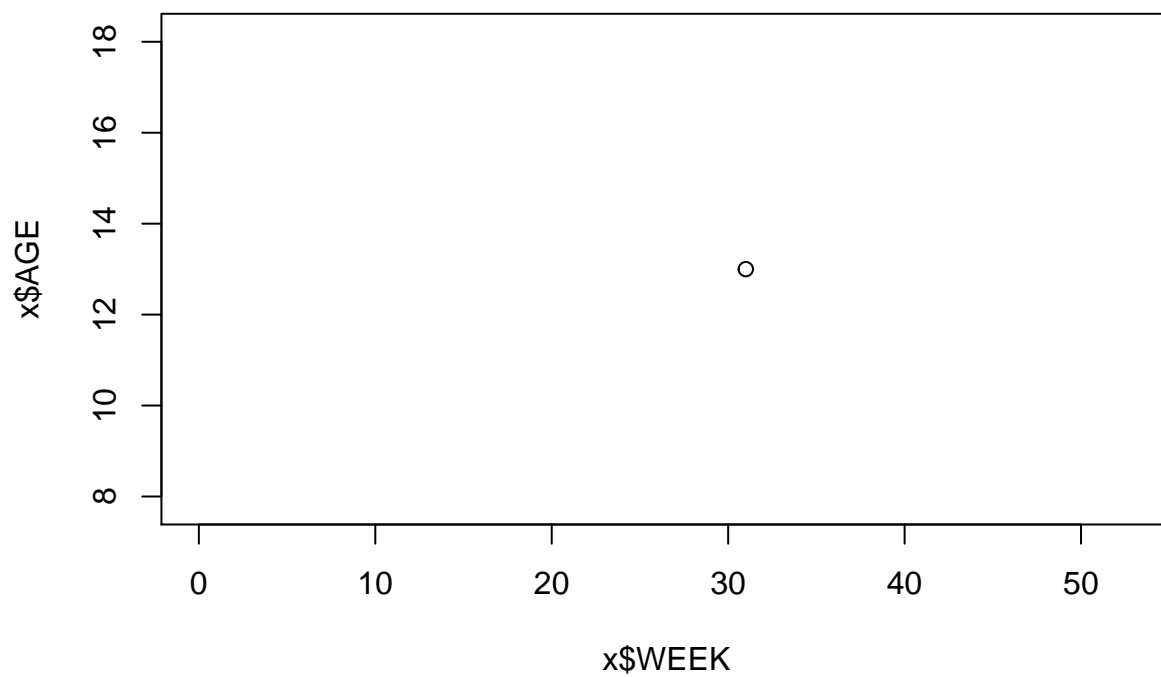
AMERICANA
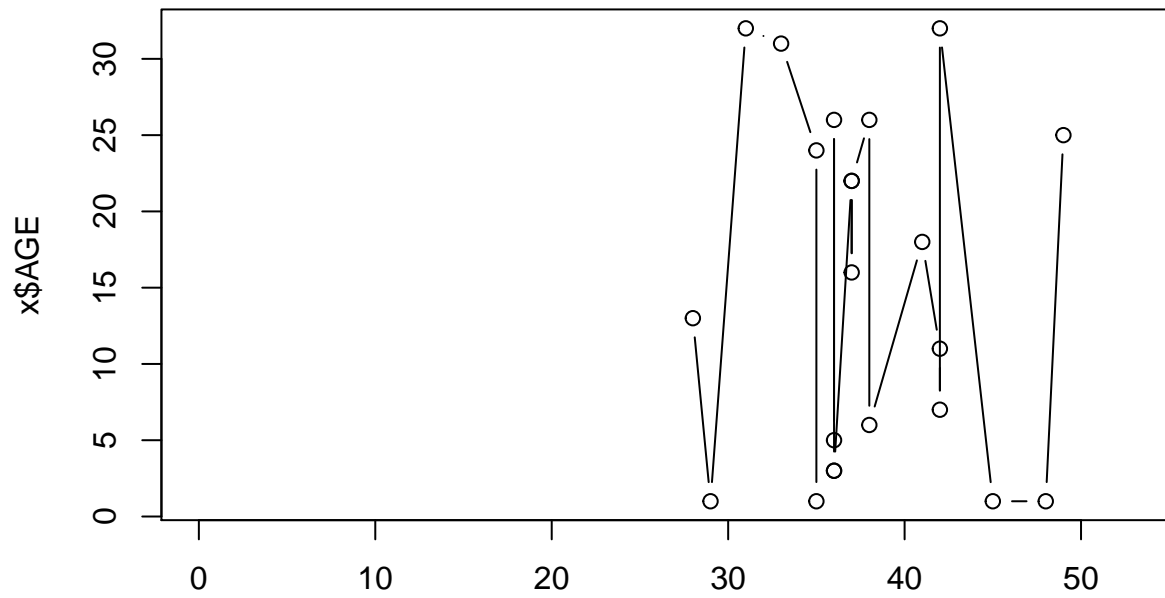
JABOTICABAL
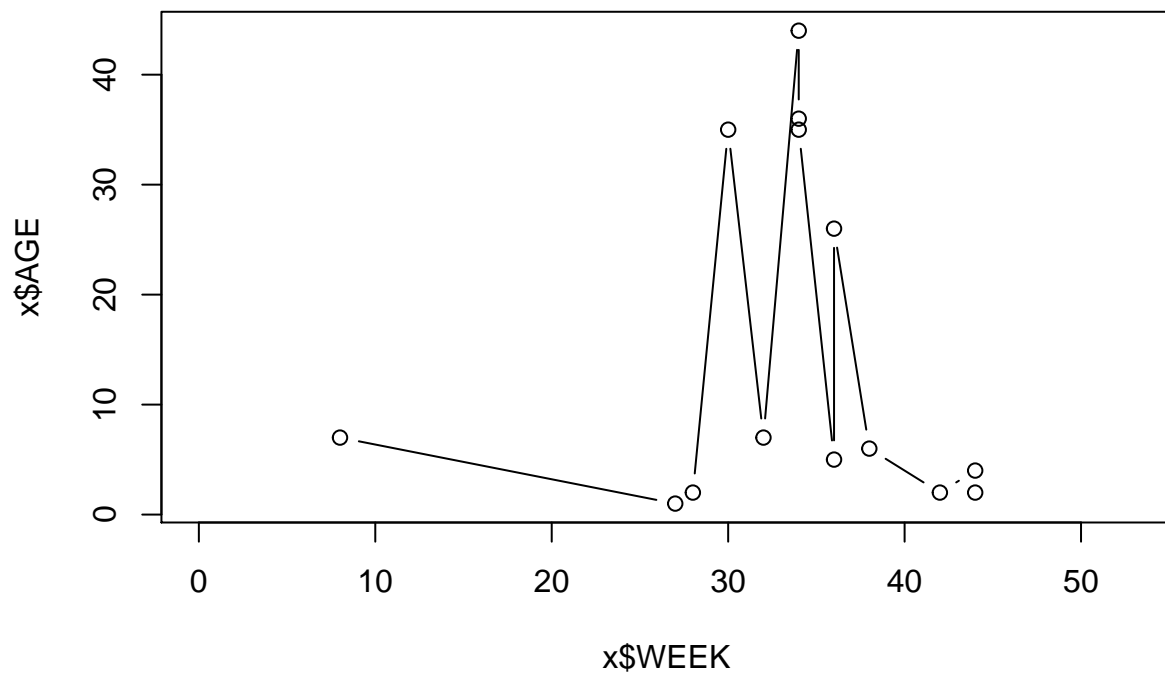
ANGATUBA
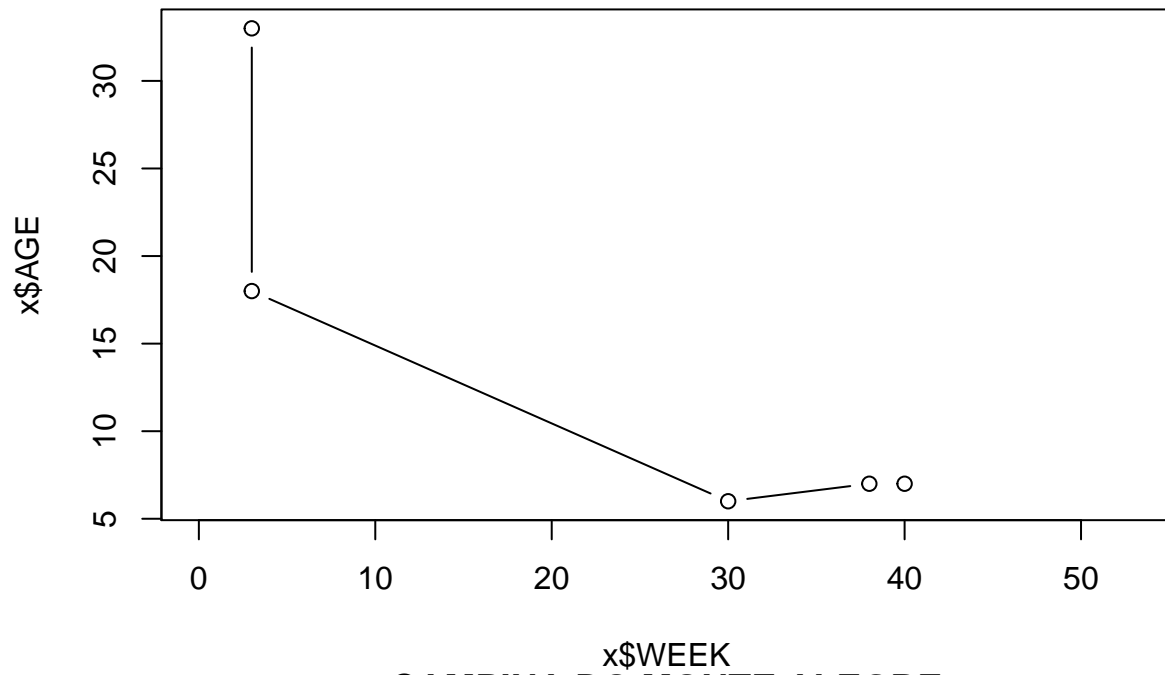
CAMPINA DO MONTE ALEGRE
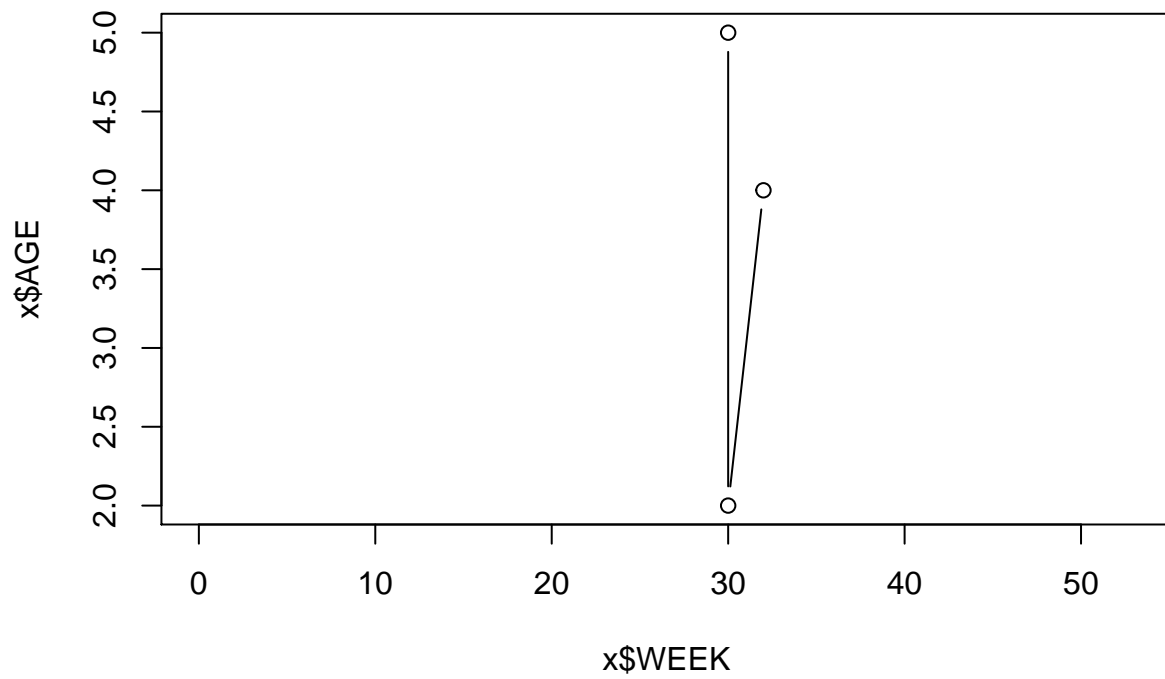
SALES
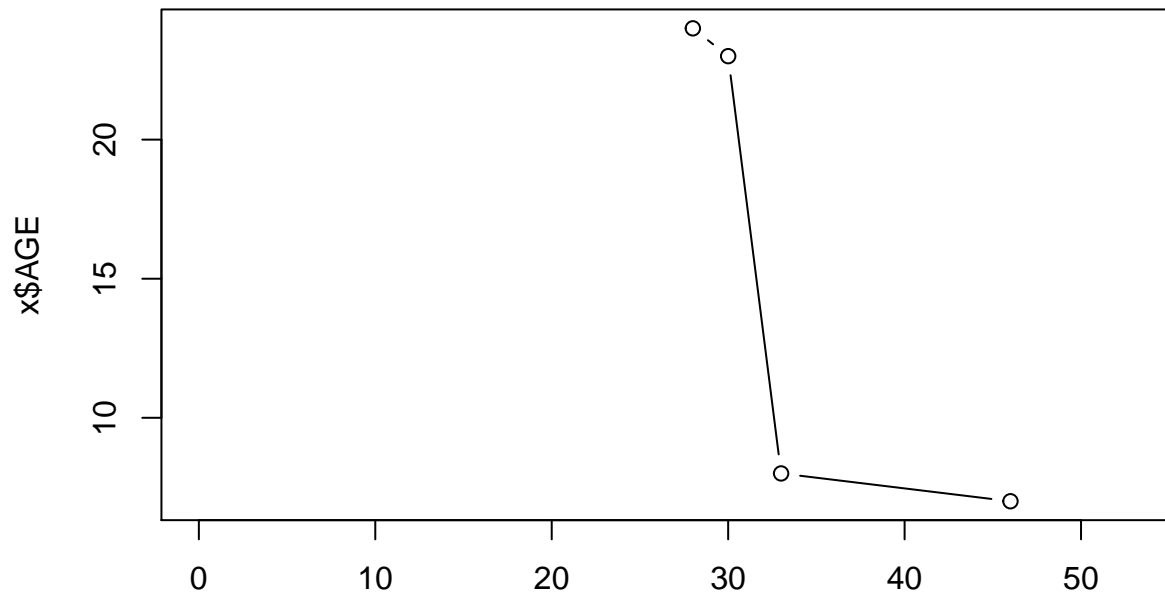
BADY BASSITT
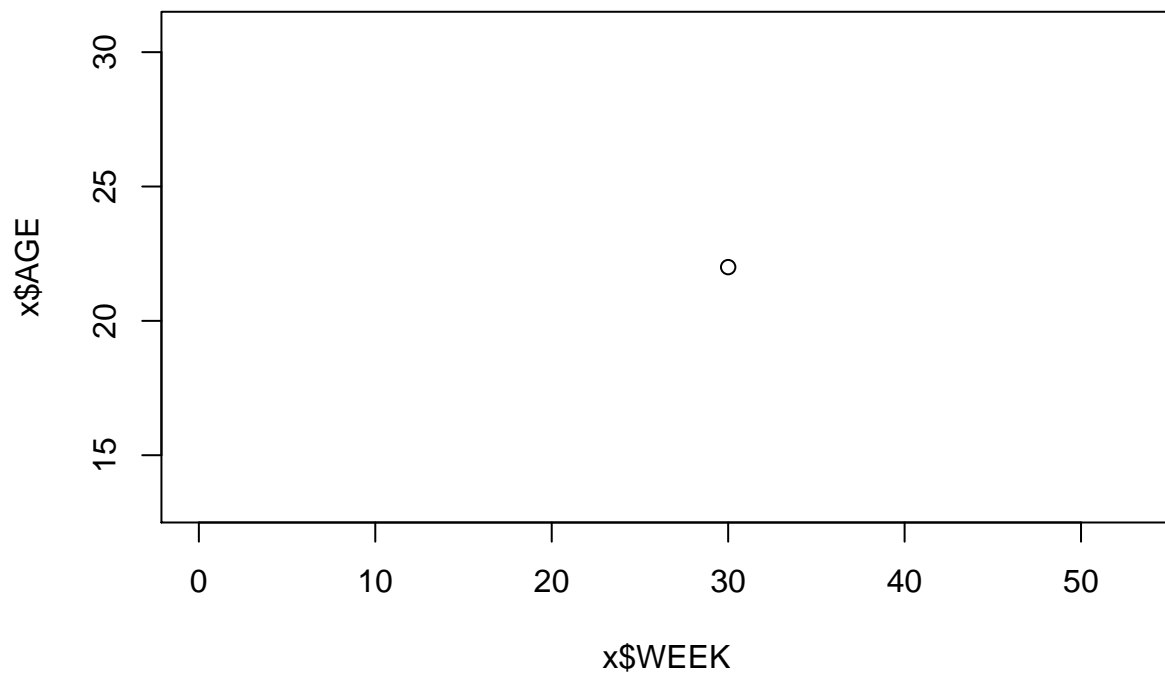
SANTA ISABEL
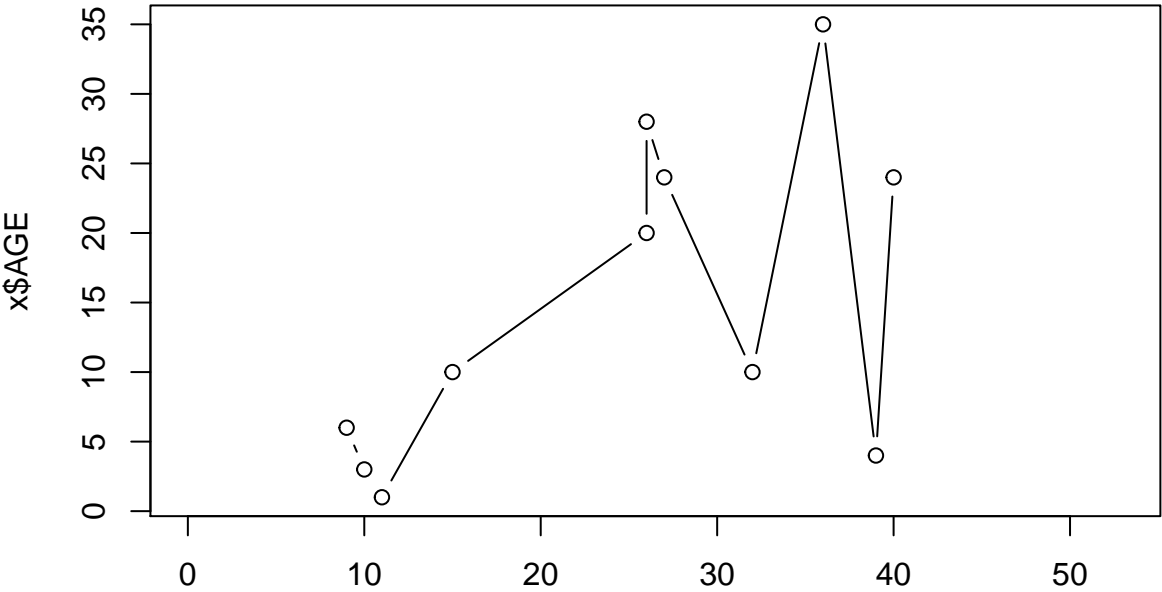
CERQUEIRA CESAR

BOITUVA

x$WEEK
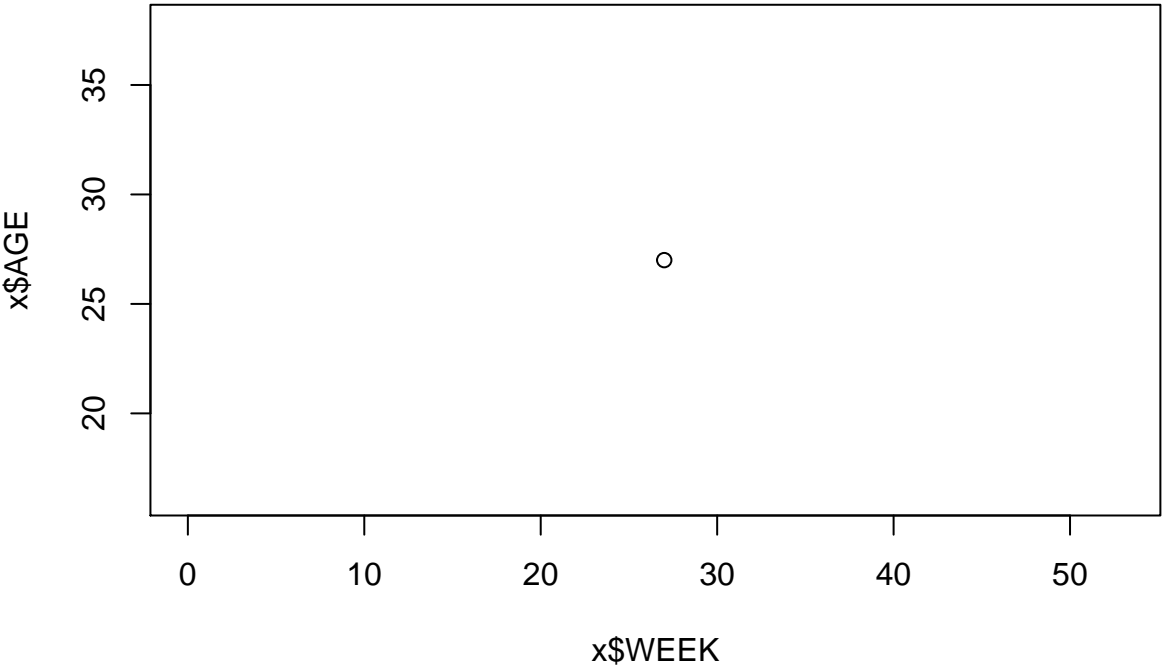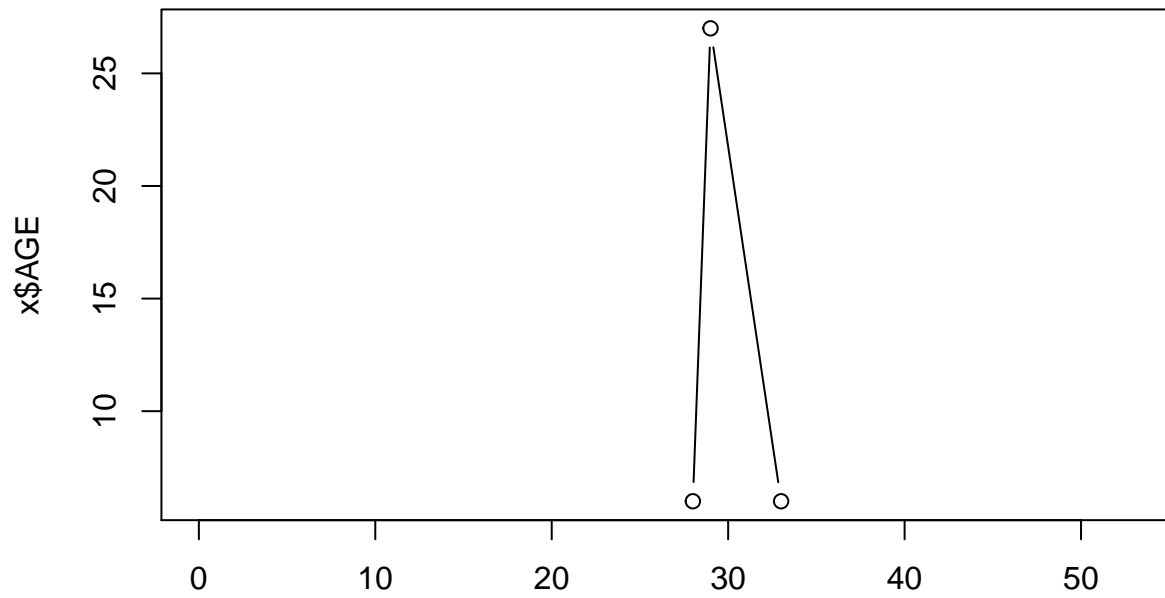
CABRALIA PAULISTA

x$WEEK
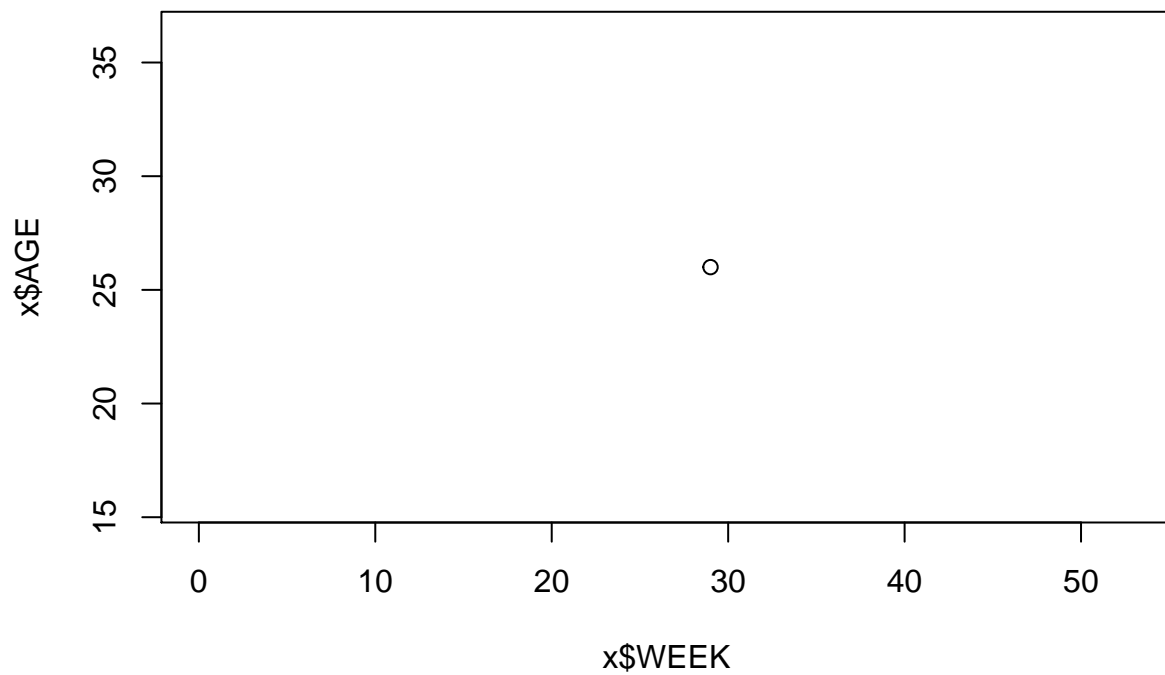
COSMOPOLIS
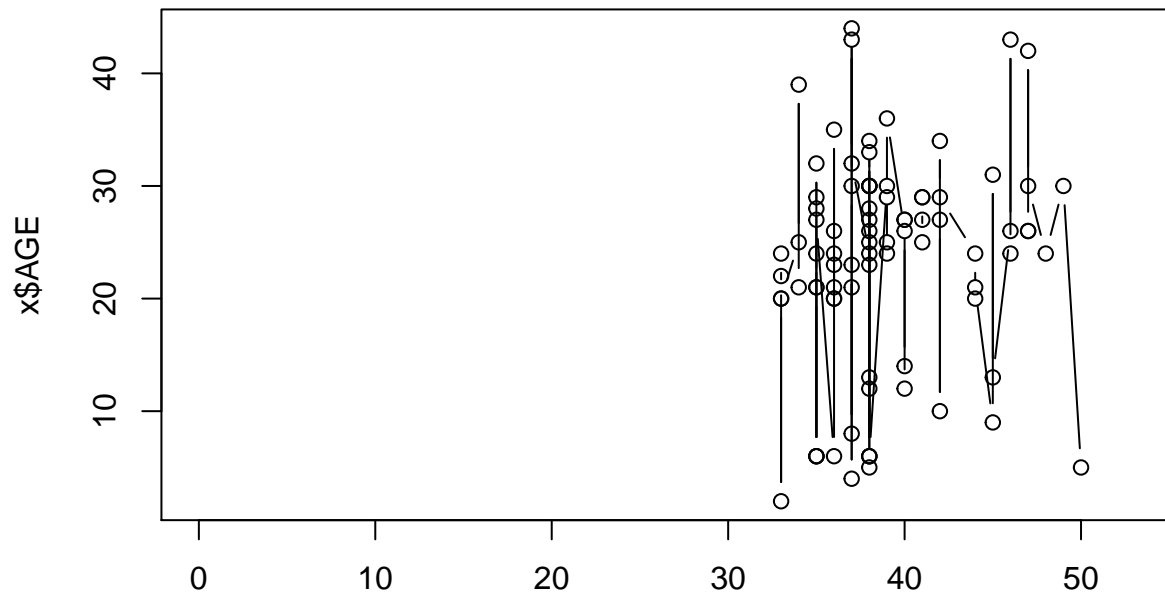
SANTA BARBARA D'OESTE
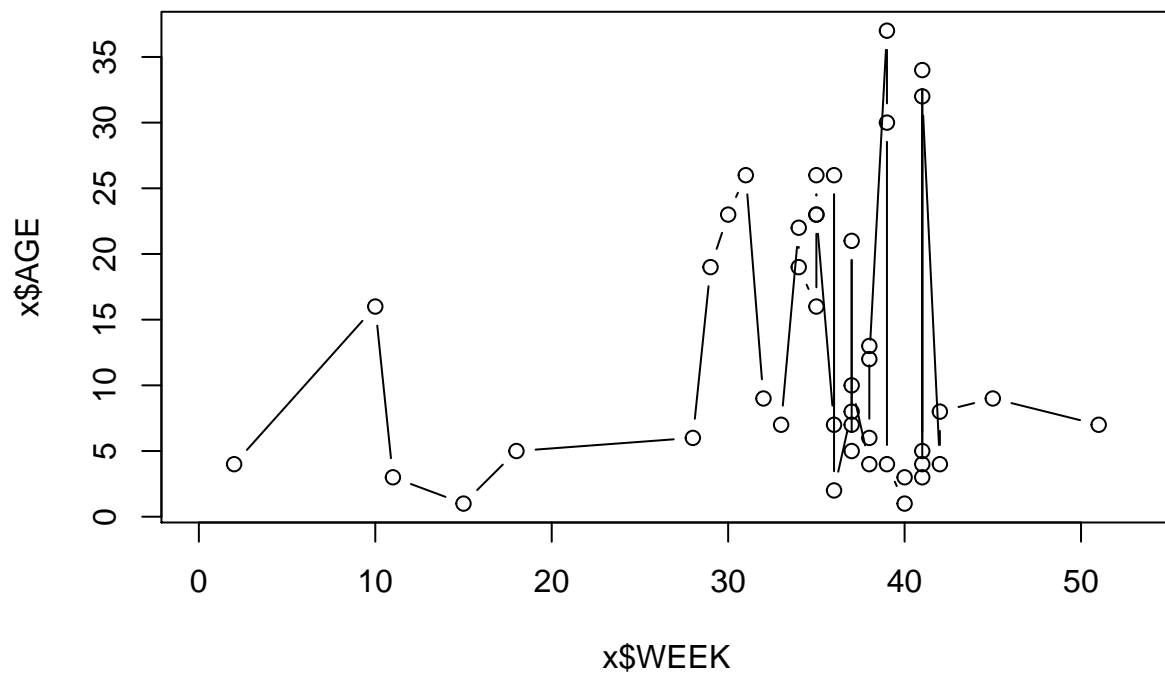
HORTOLANDIA
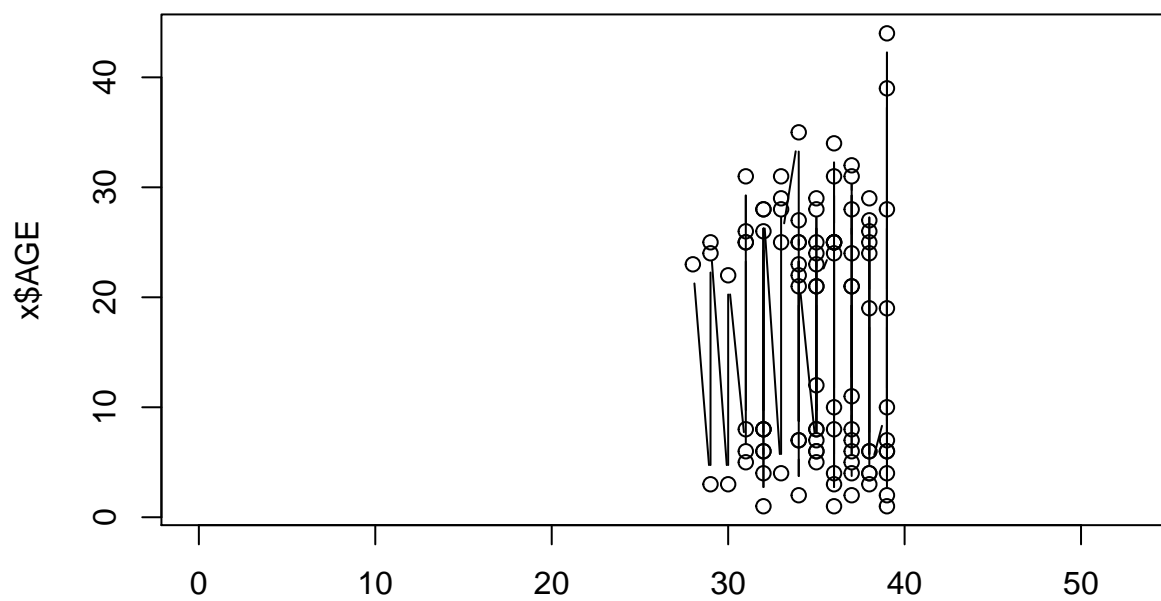
PARIQUERA-ACU

PAULINIA

x$WEEK

PROMISSAO

x$WEEK
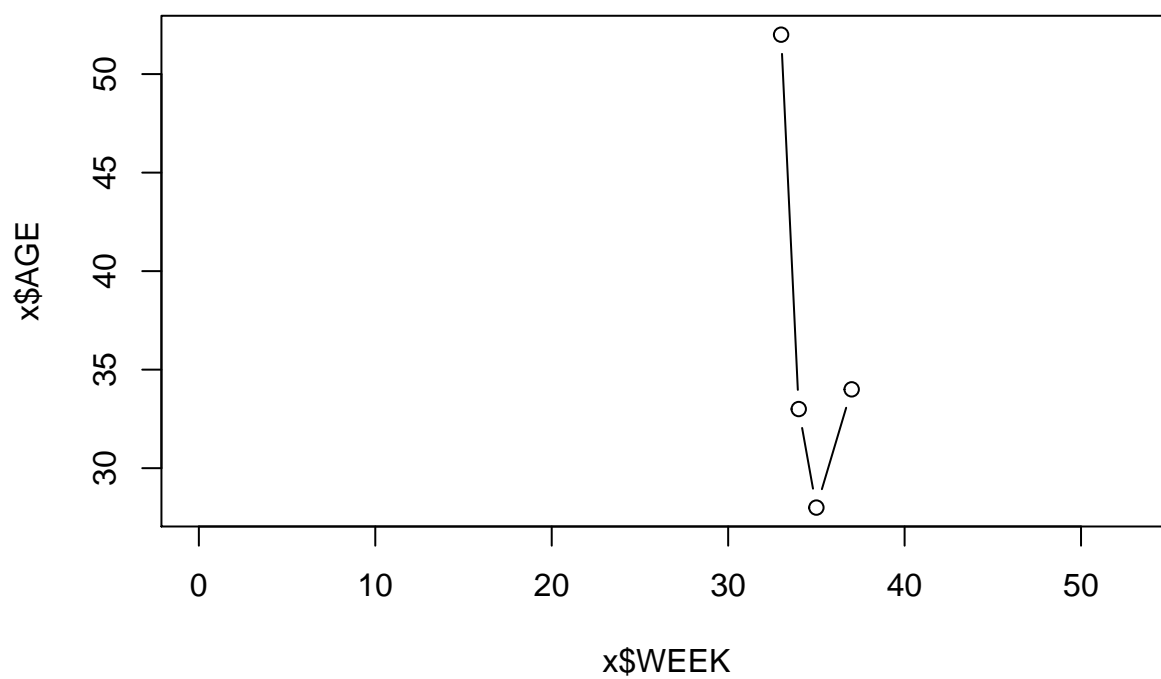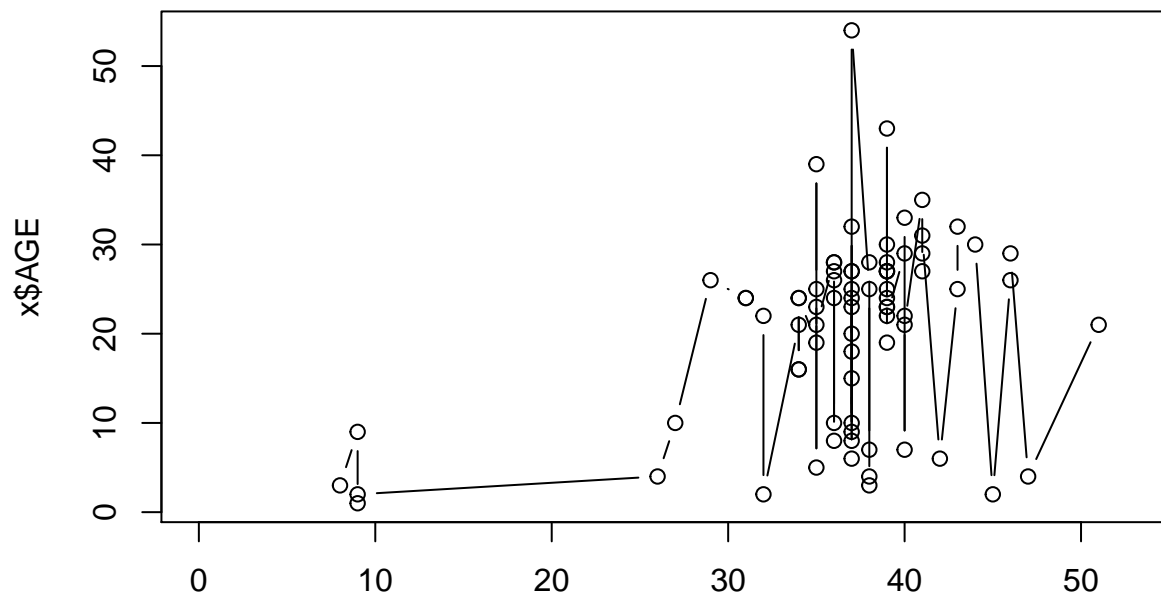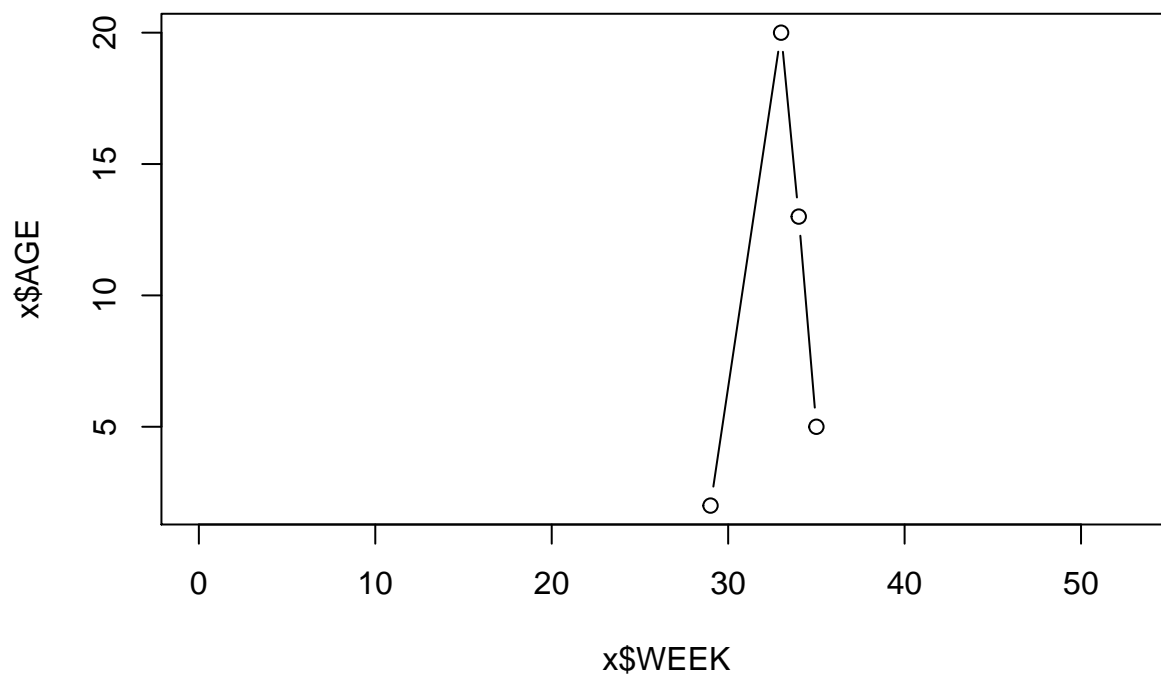
RIO DE JANEIRO
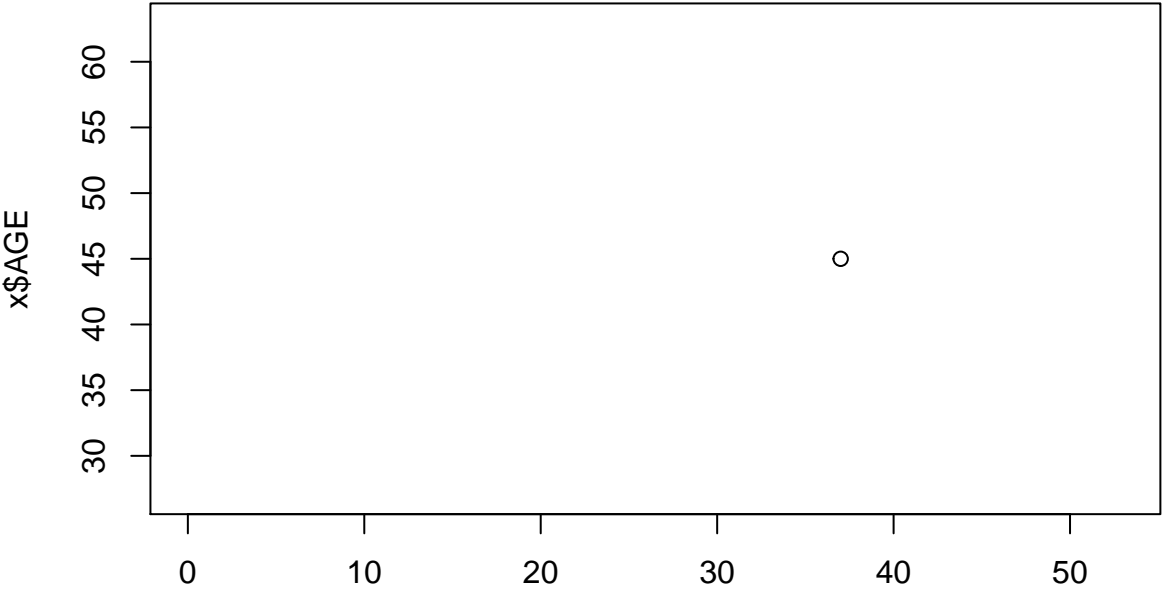
ALVARO DE CARVALHO
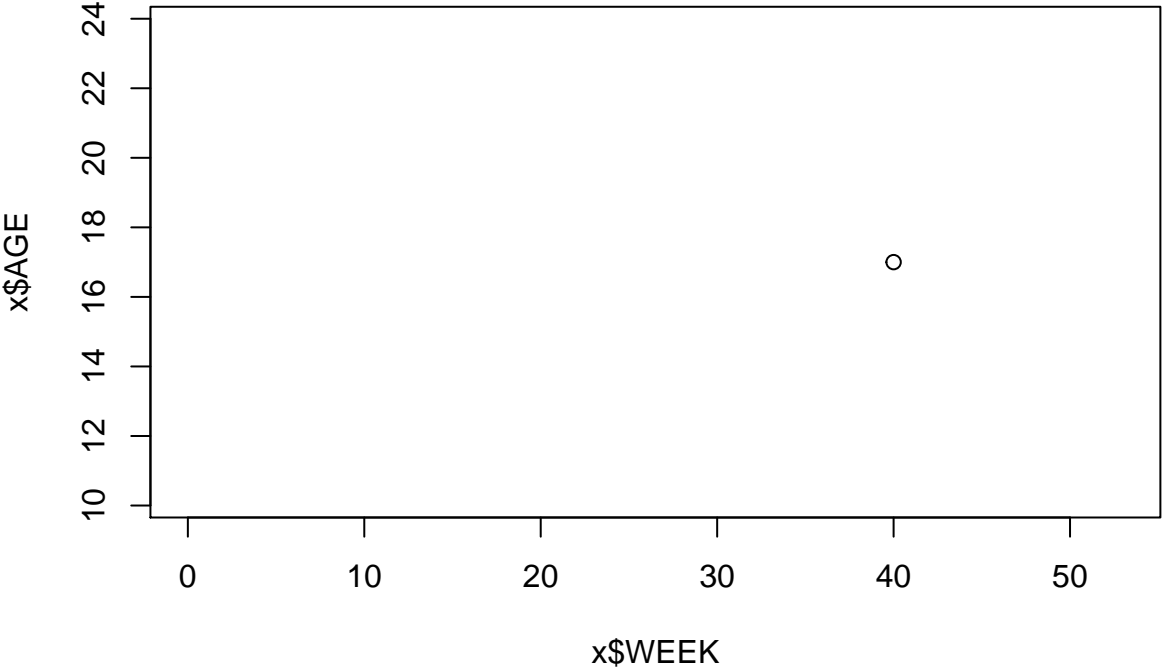
PINHALZINHO
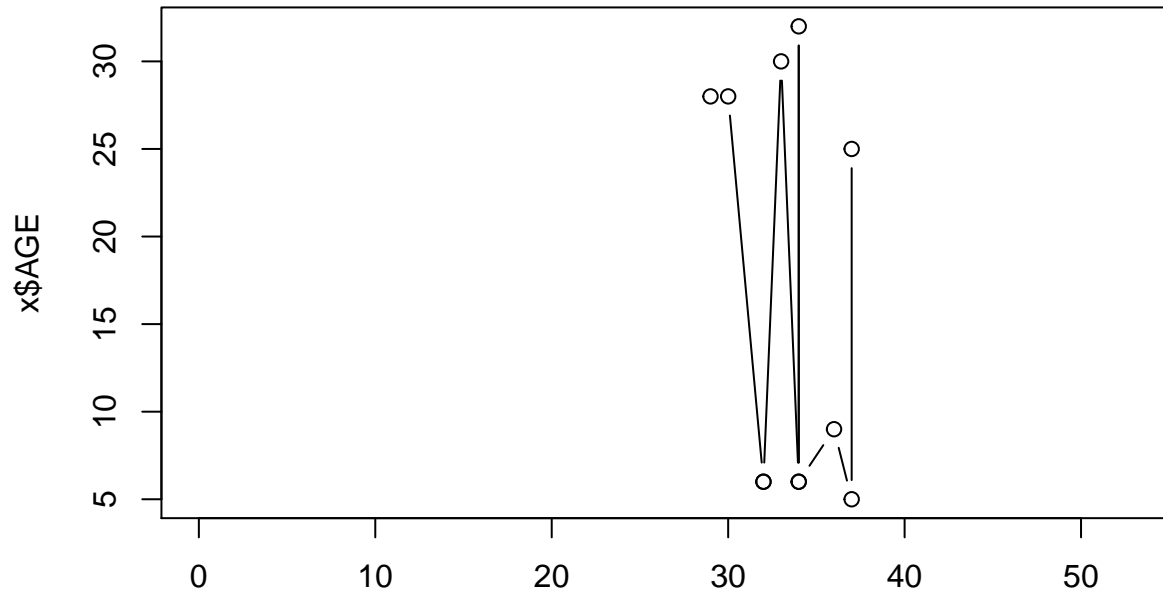
SAO LOURENCO DA SERRA

BATATAIS

x$WEEK

PILAR DO SUL

x$WEEK
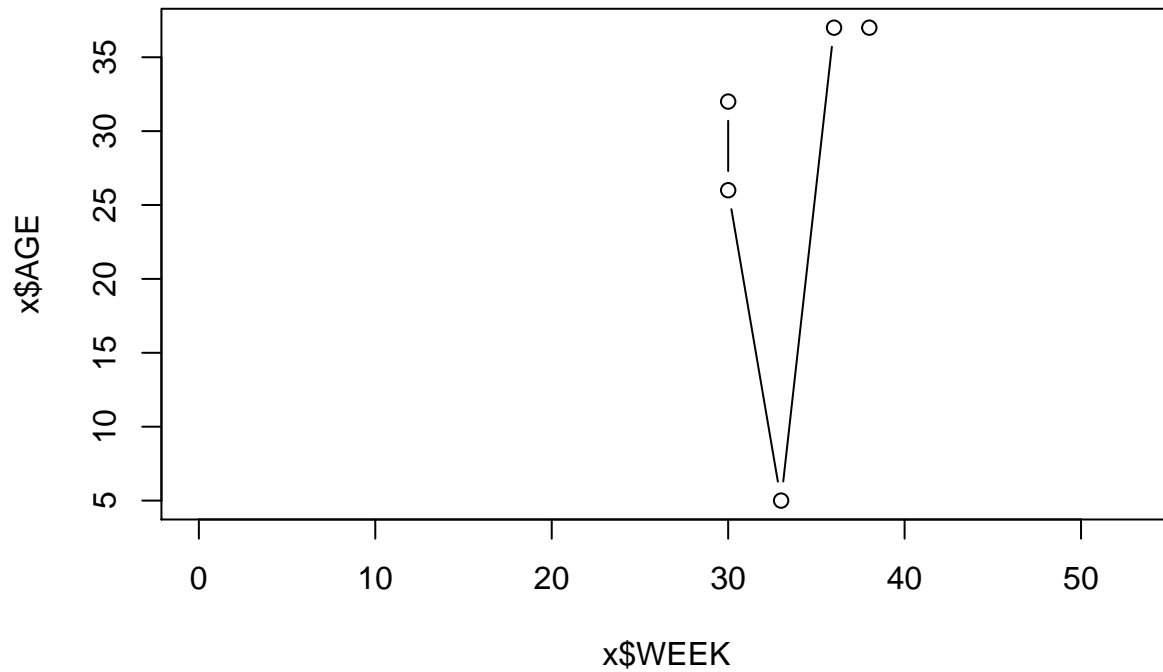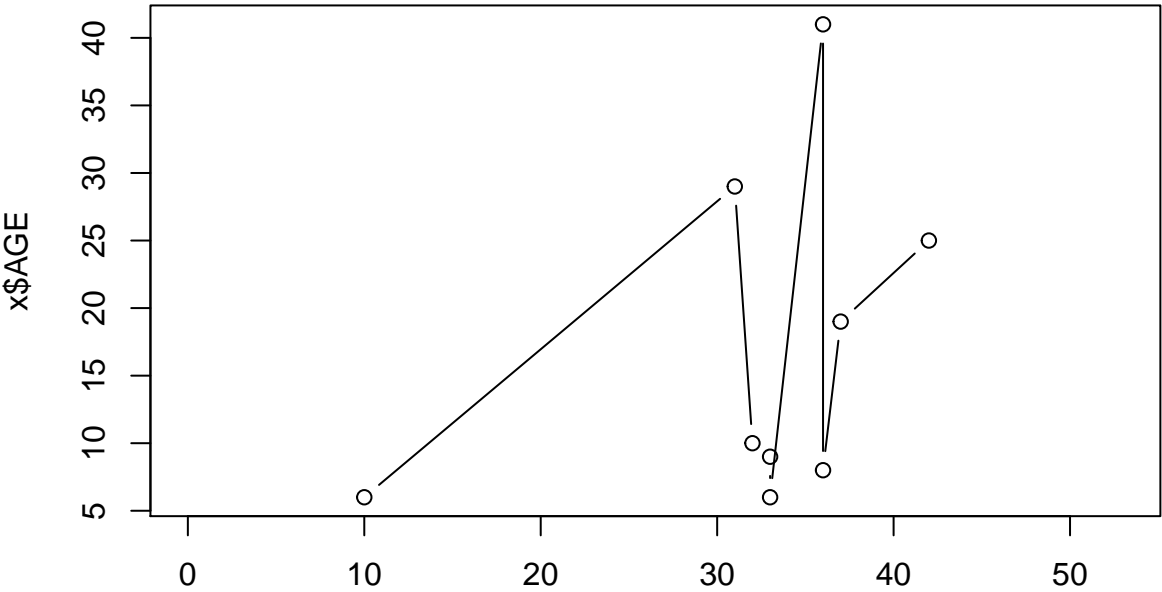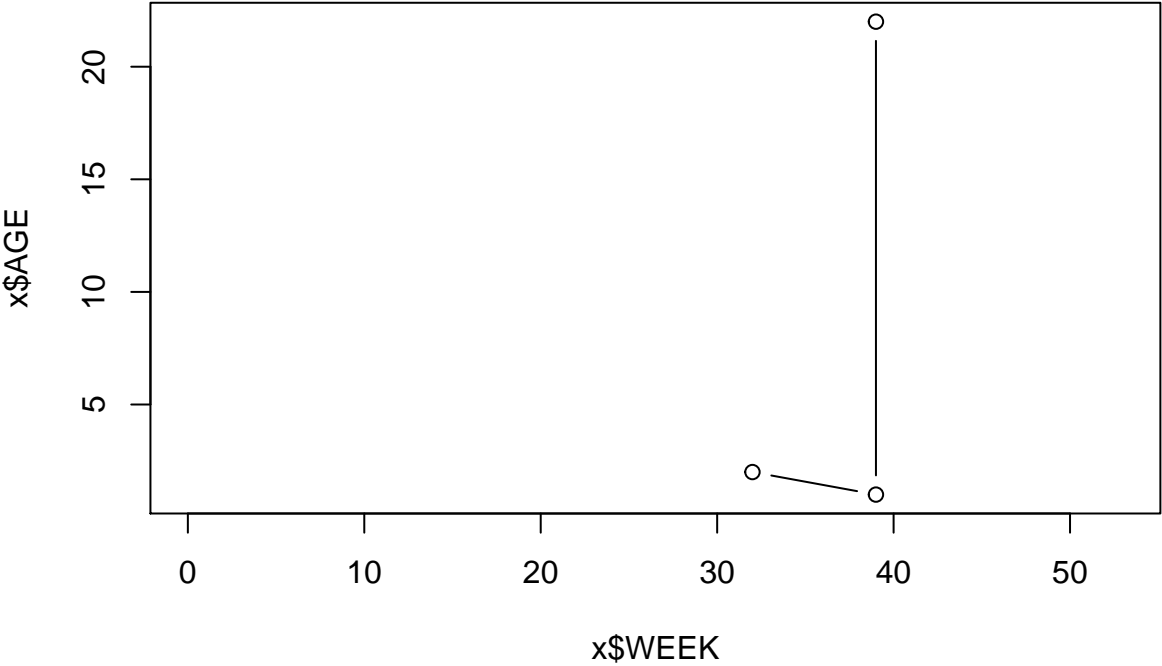
PEREIRA BARRETO
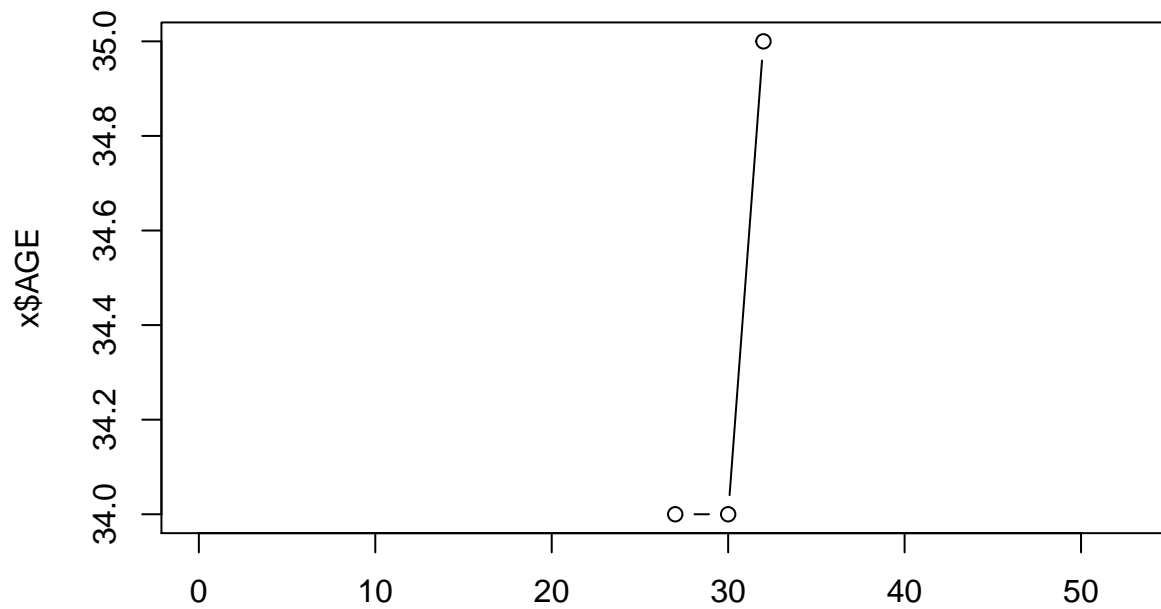
PIRACAIA
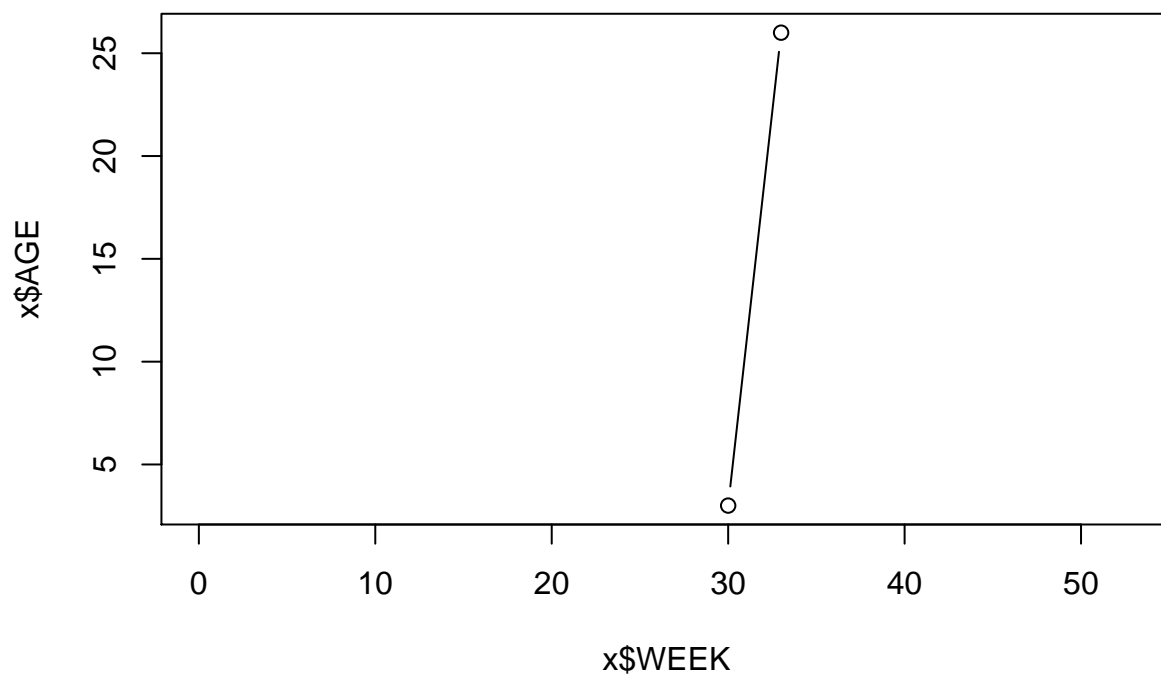
BOM JESUS DOS PERDOES

PITANGUEIRAS

122
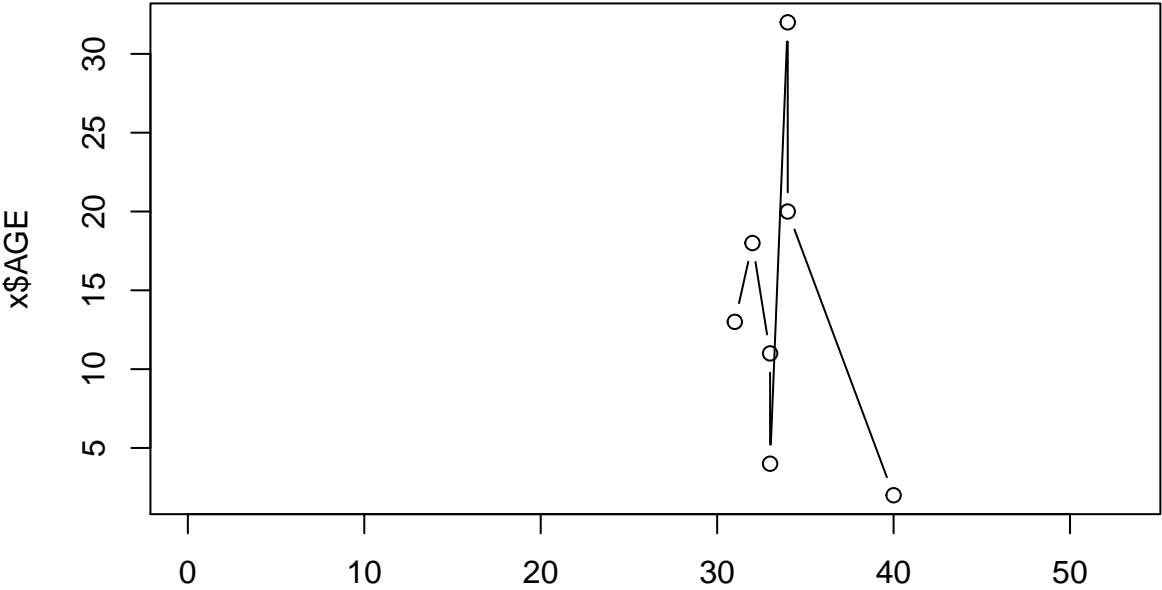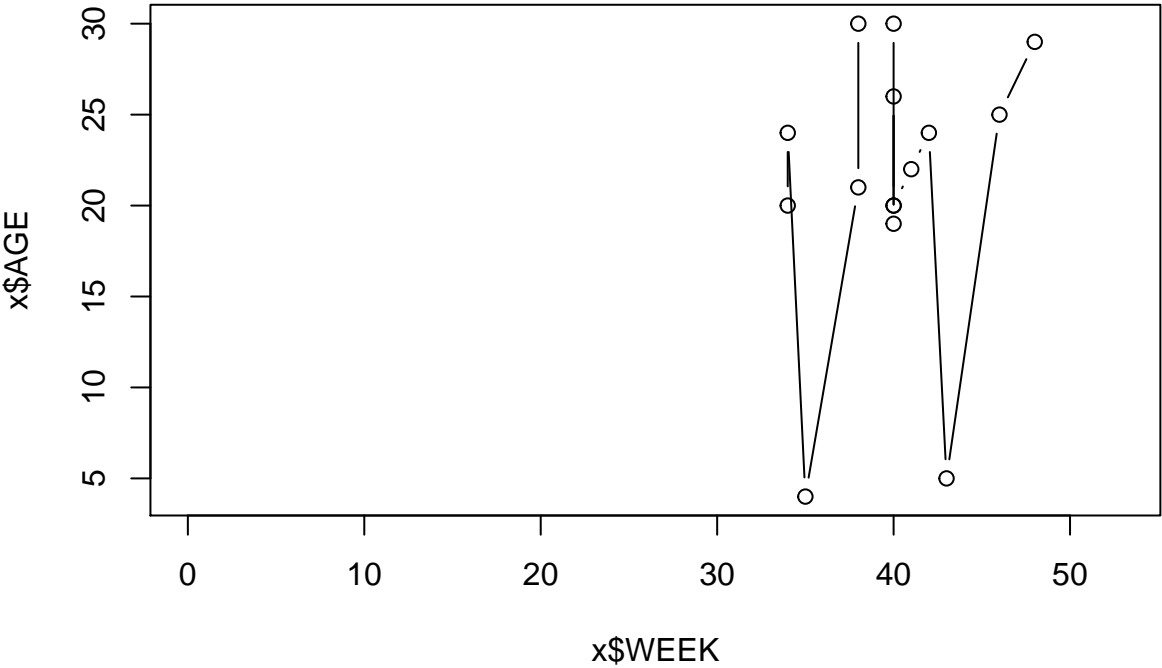
SANTO ANASTACIO
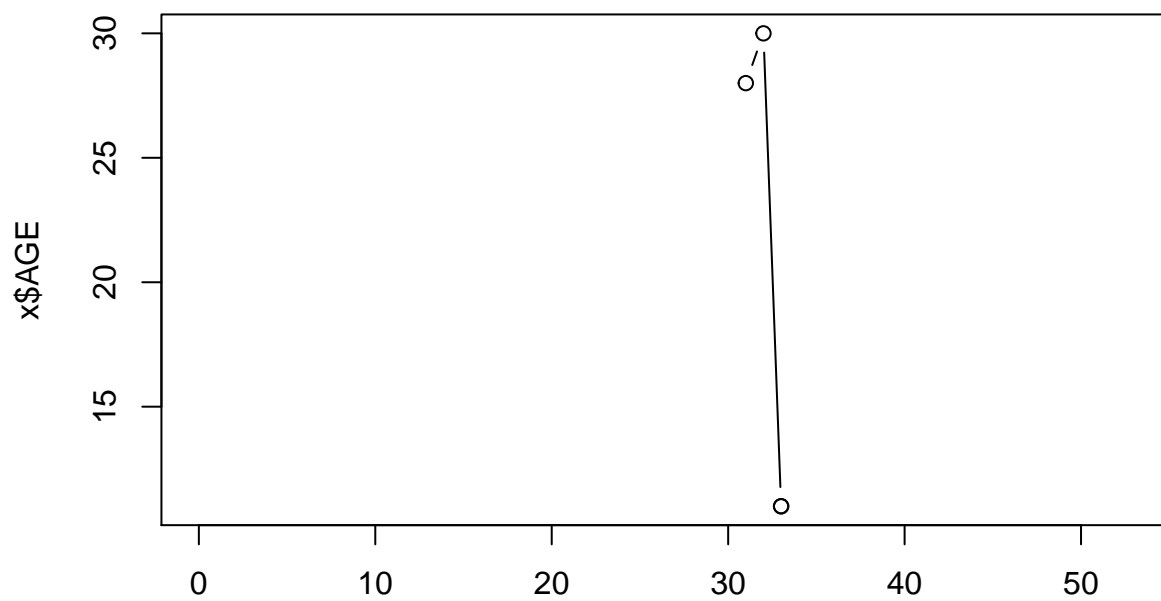
INDAIATUBA
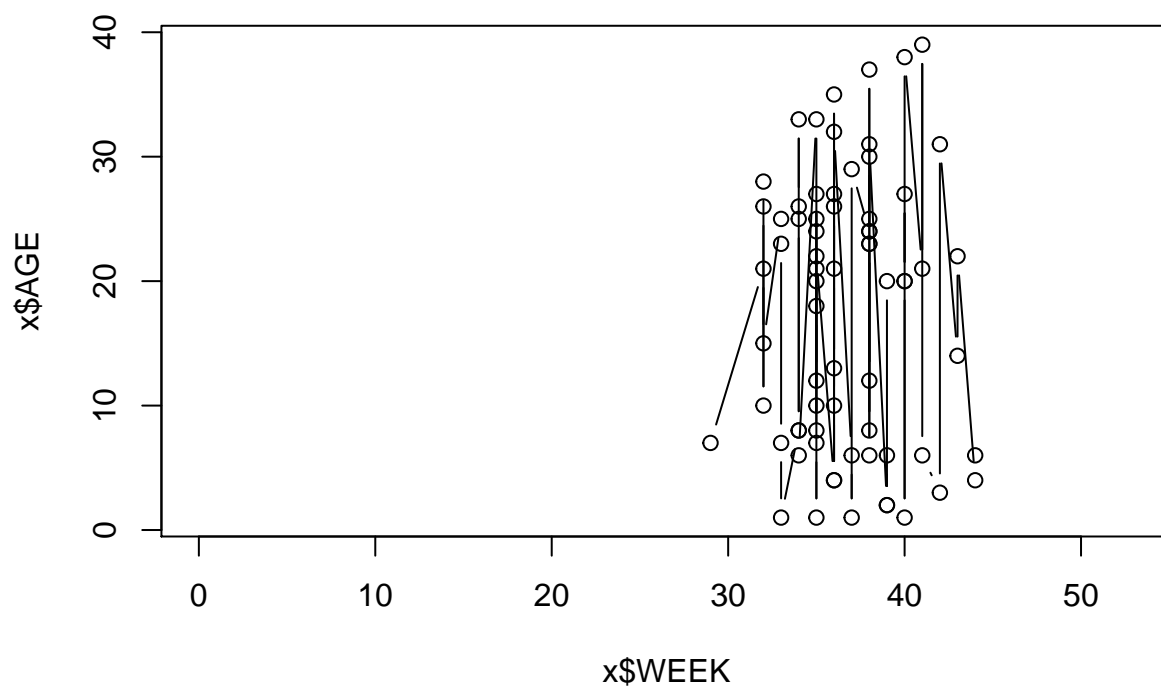
ITUPEVA
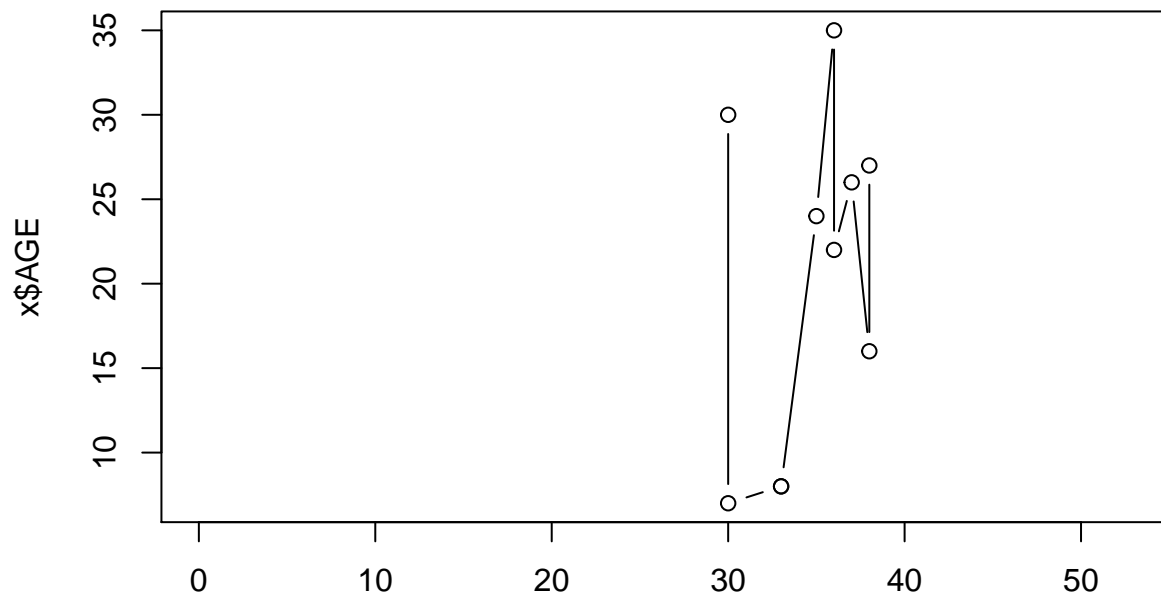
CORDEIROPOLIS
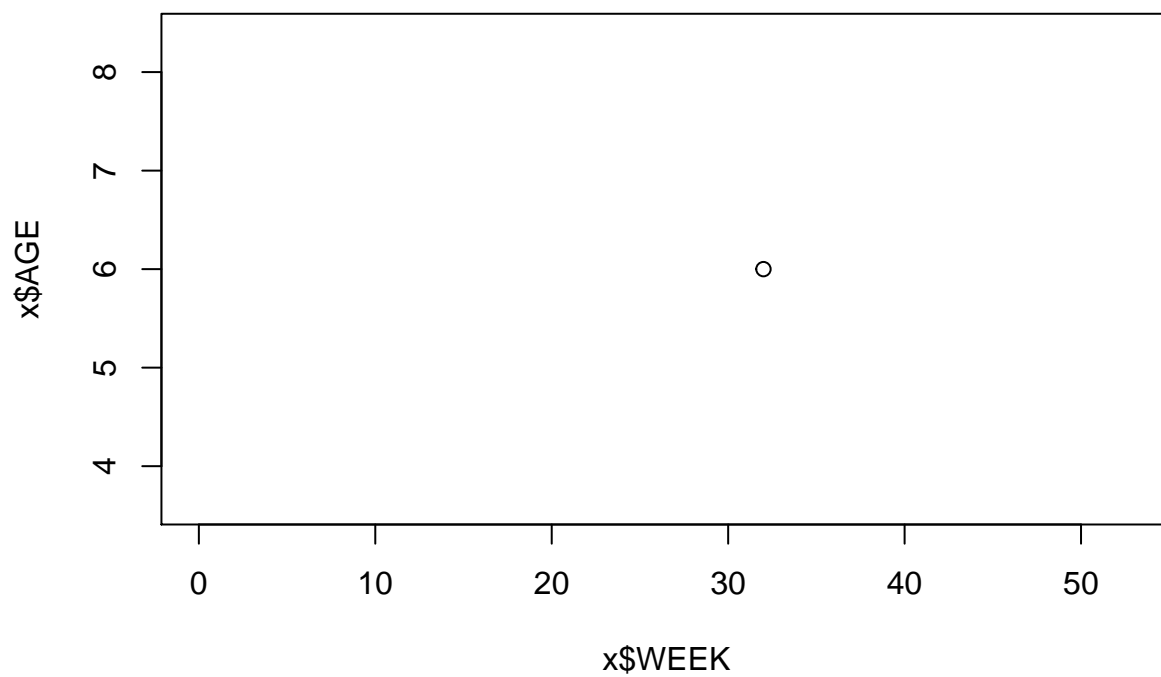
LUCELIA
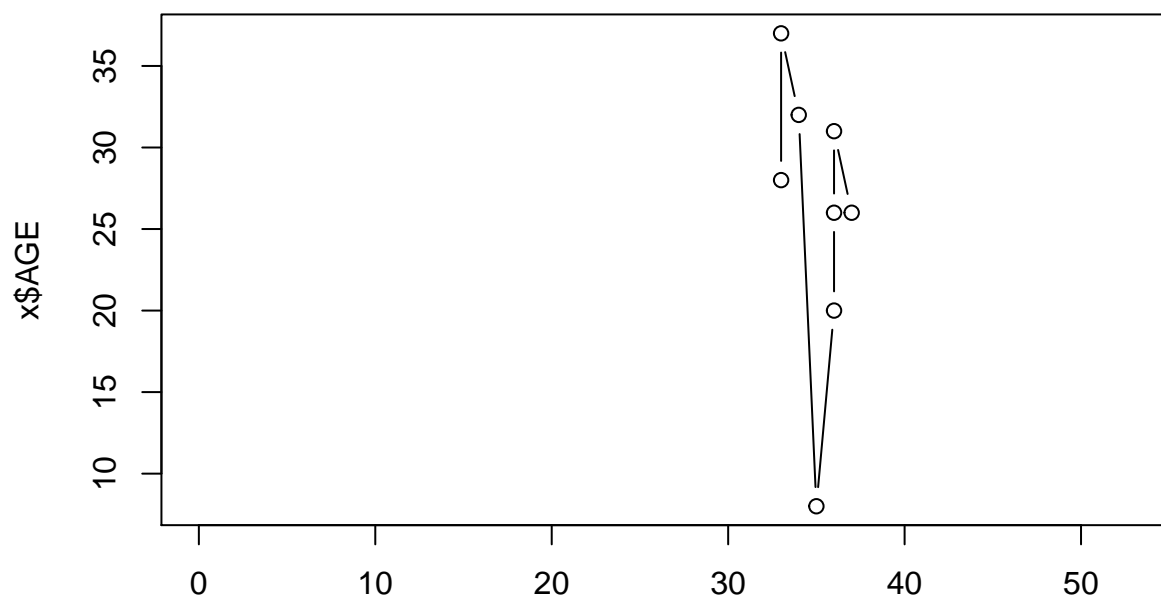
x$WEEK

ELIAS FAUSTO
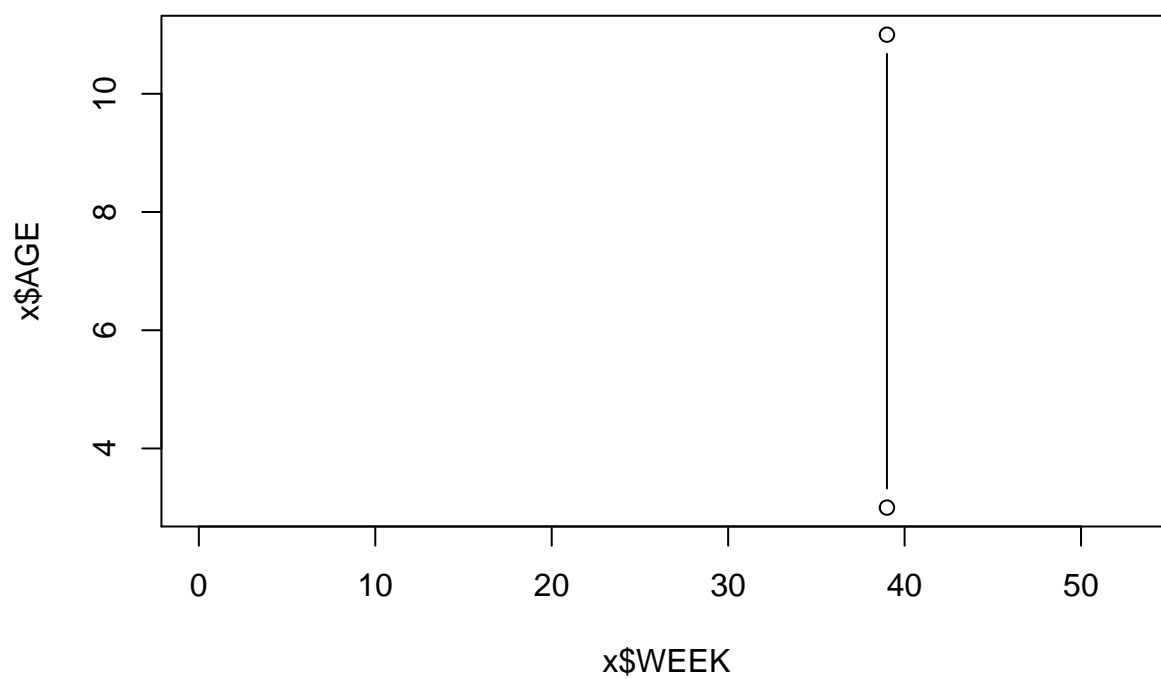
x$WEEK

CAPIVARI
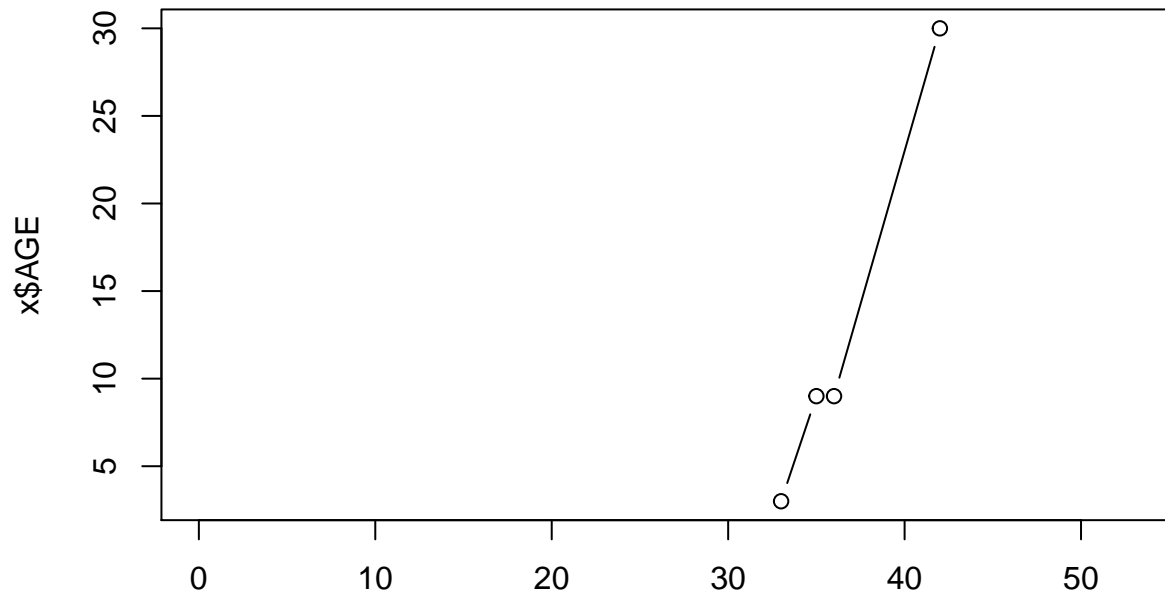
SALTINHO
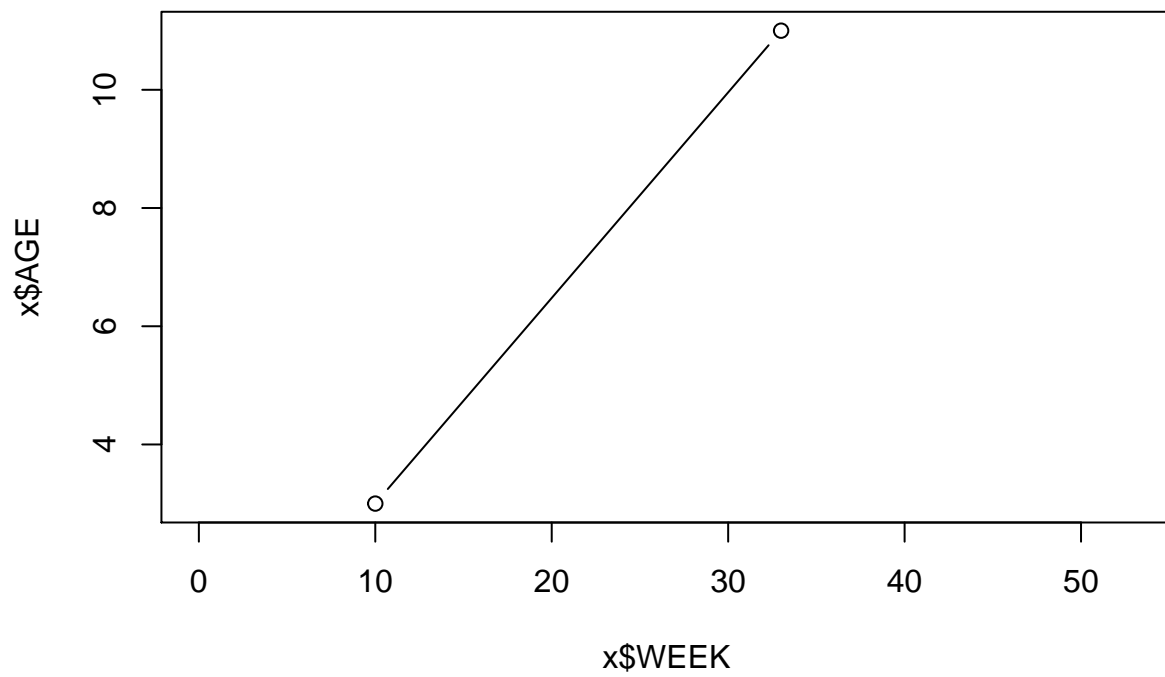
MOMBUCA
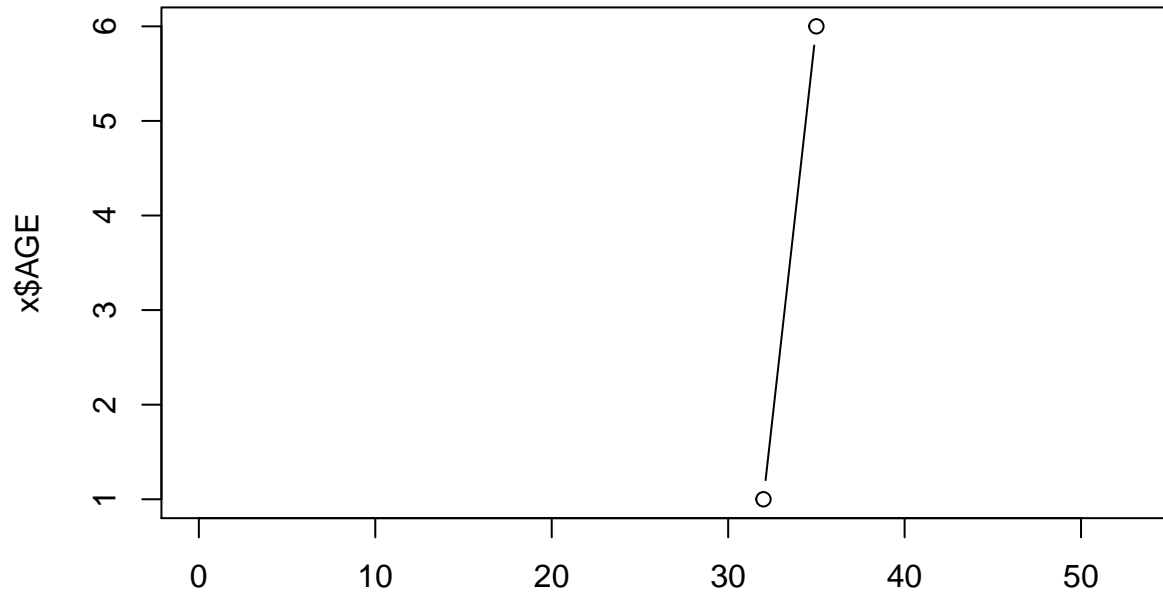
ALUMINIO
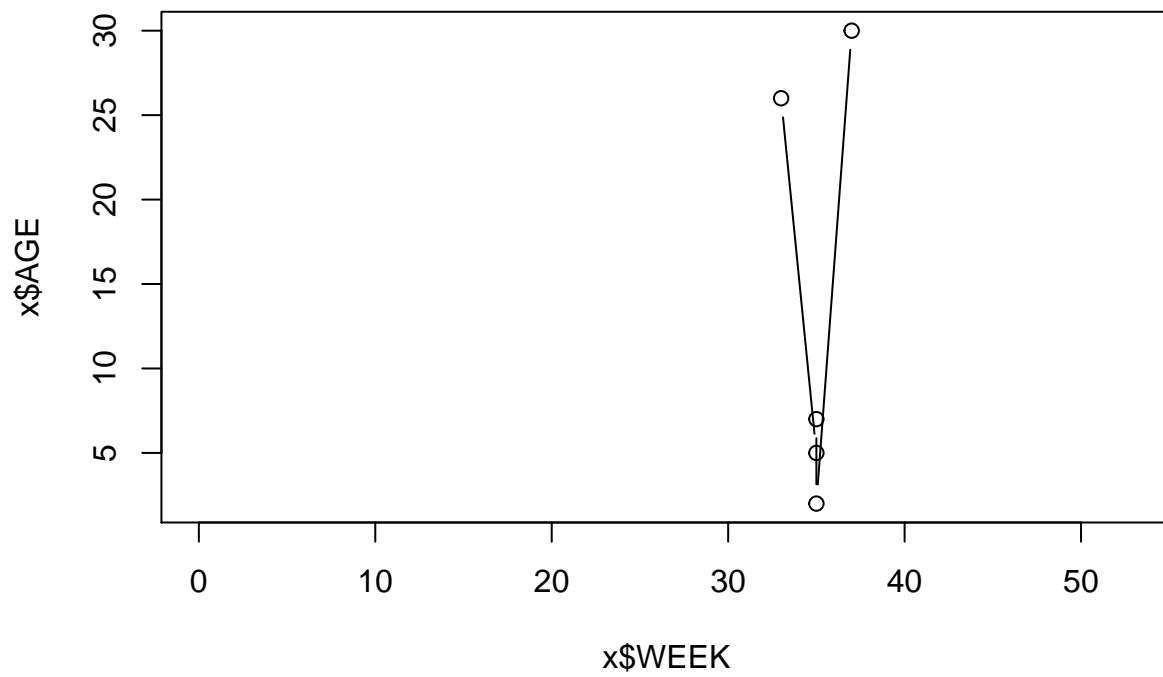
OURO VERDE
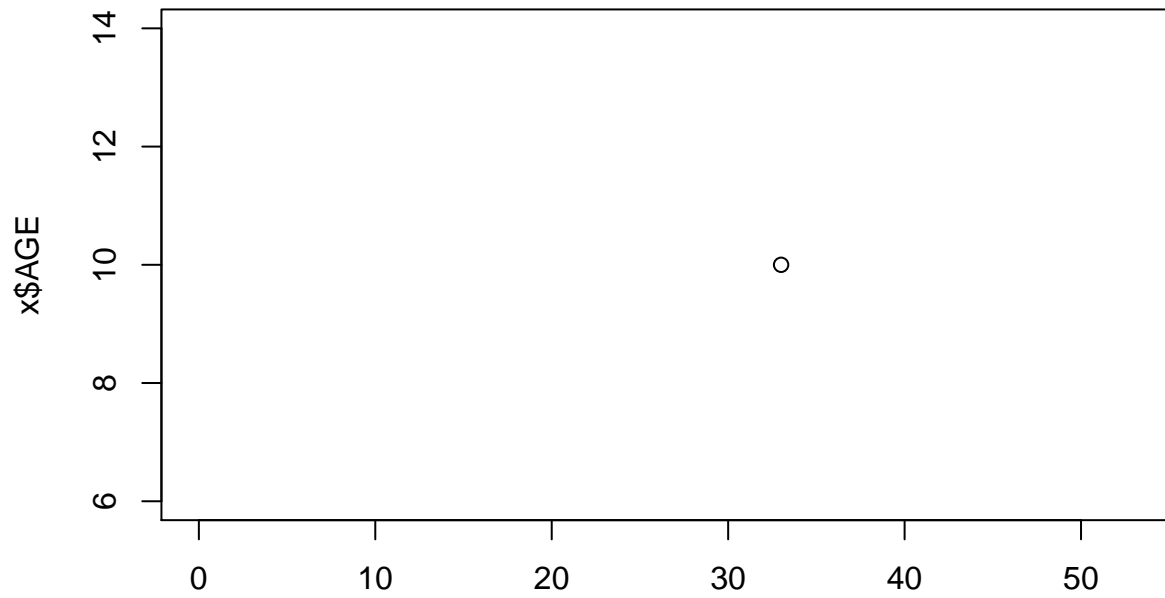
PEDREIRA
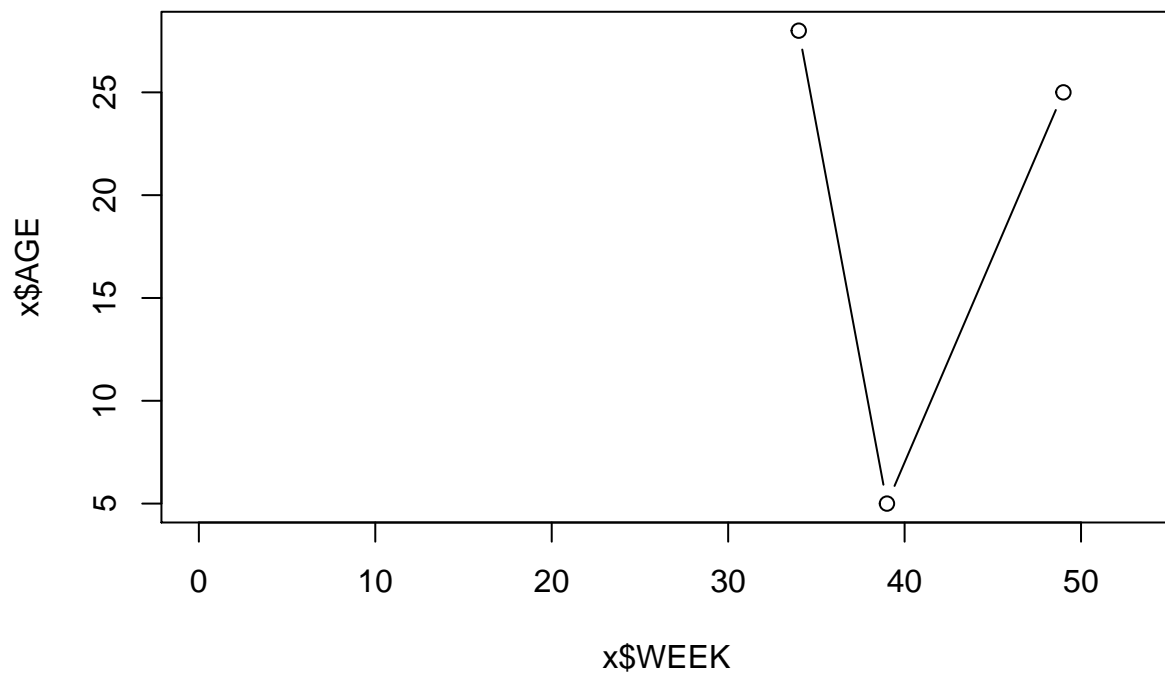
CONCHAL
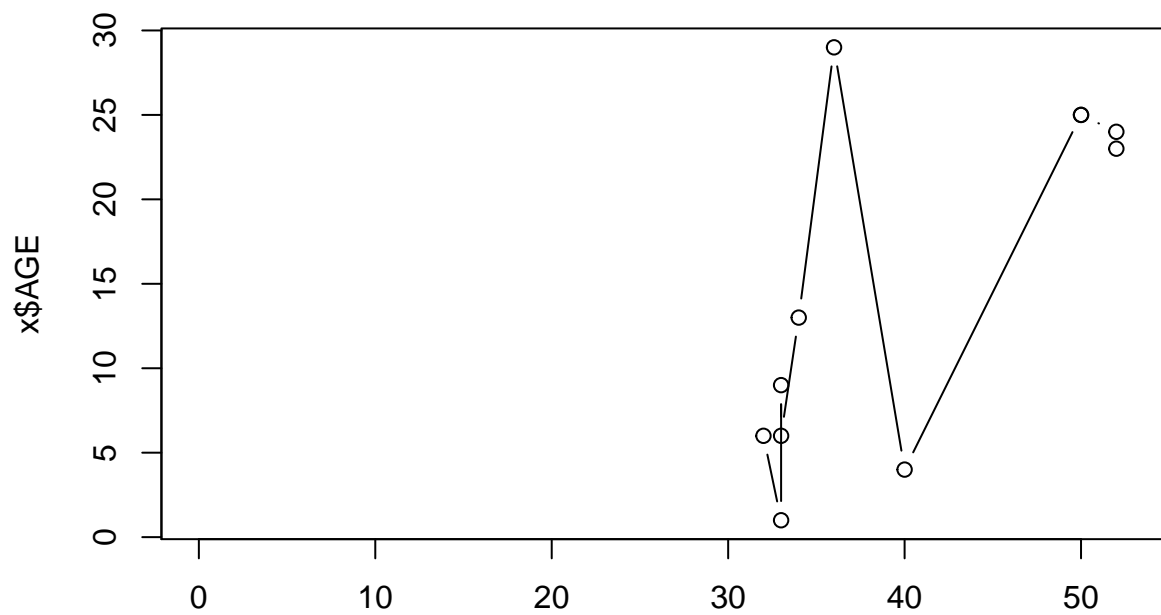
RIBEIRAO GRANDE
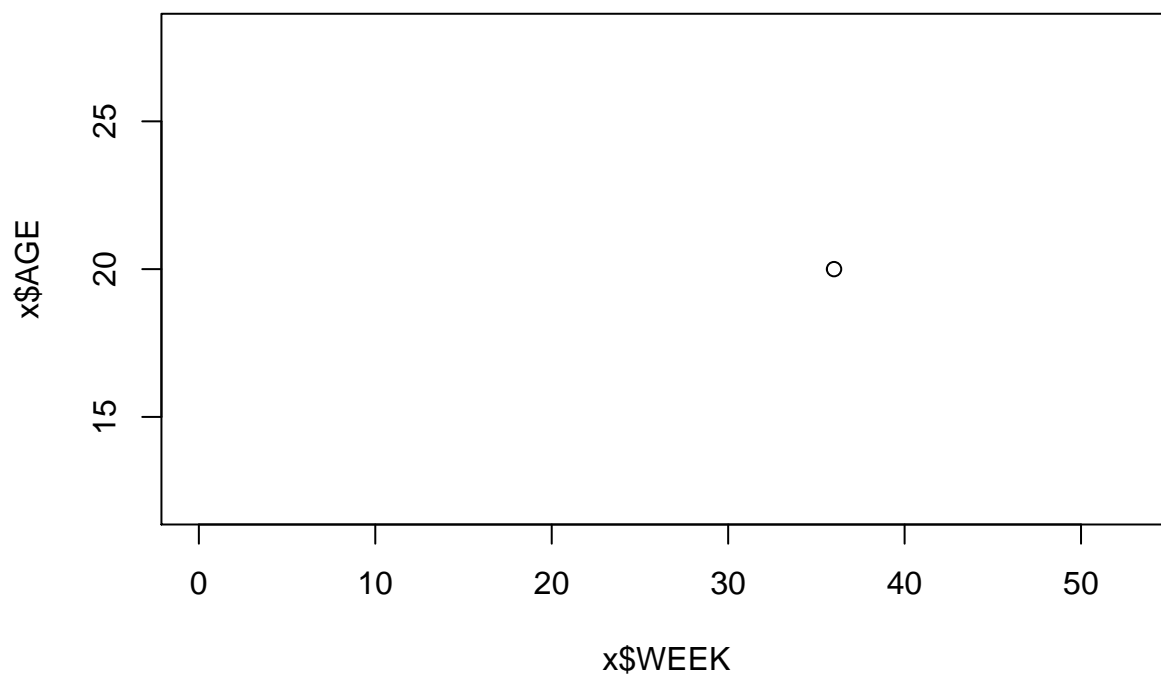
MAIRINQUE
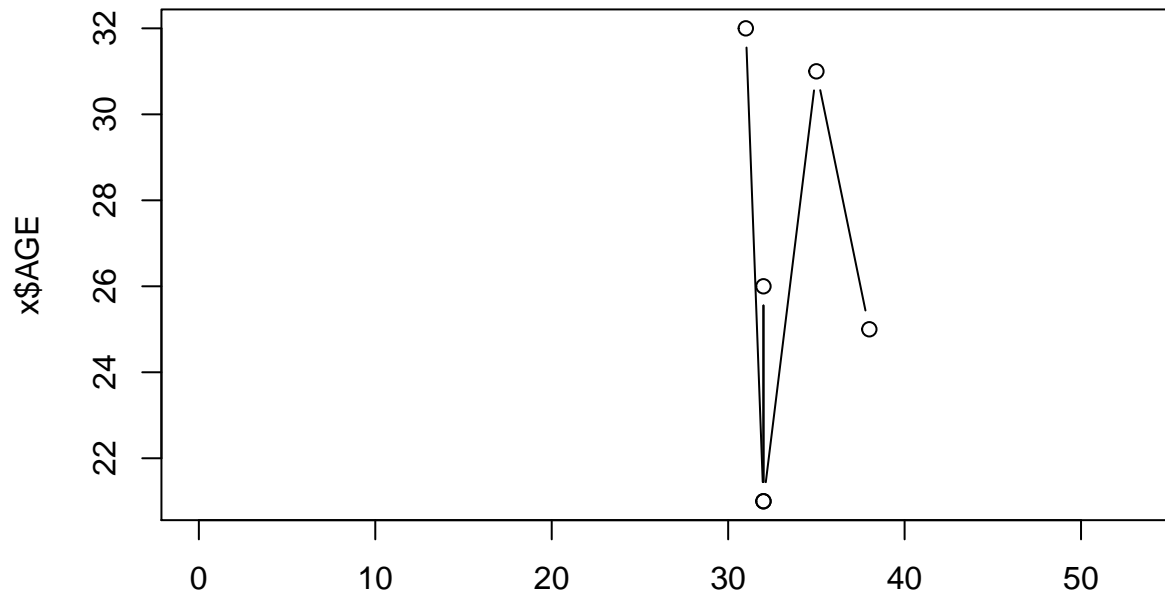
MONTE ALEGRE DO SUL
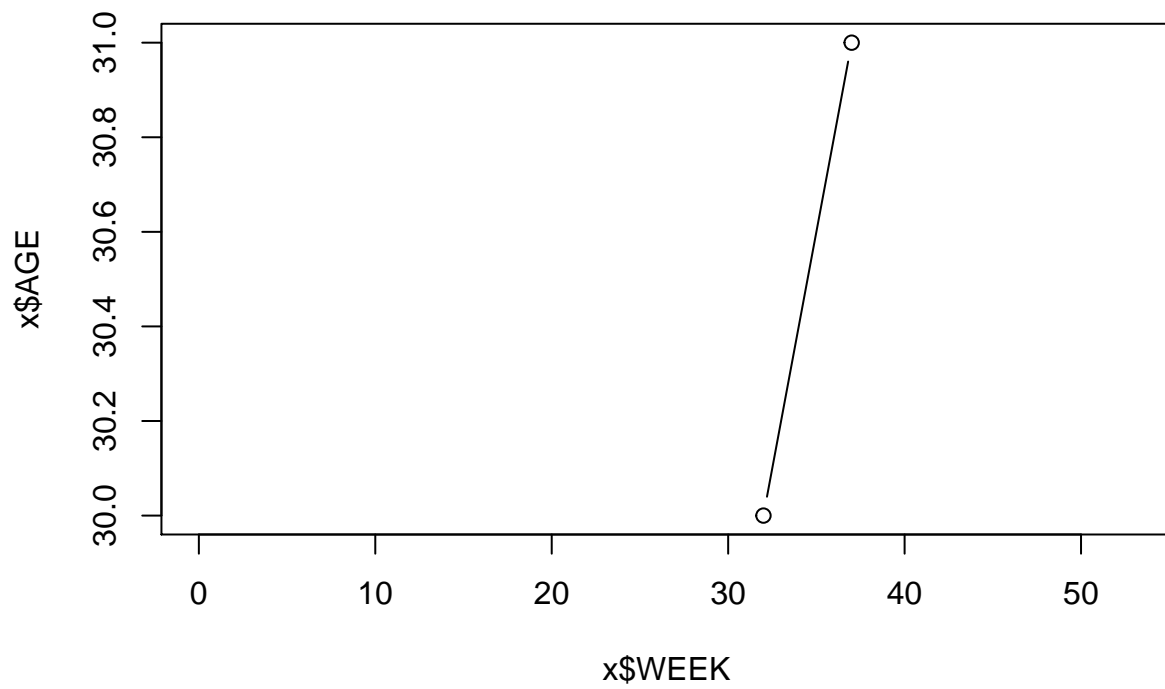
CABREUVA
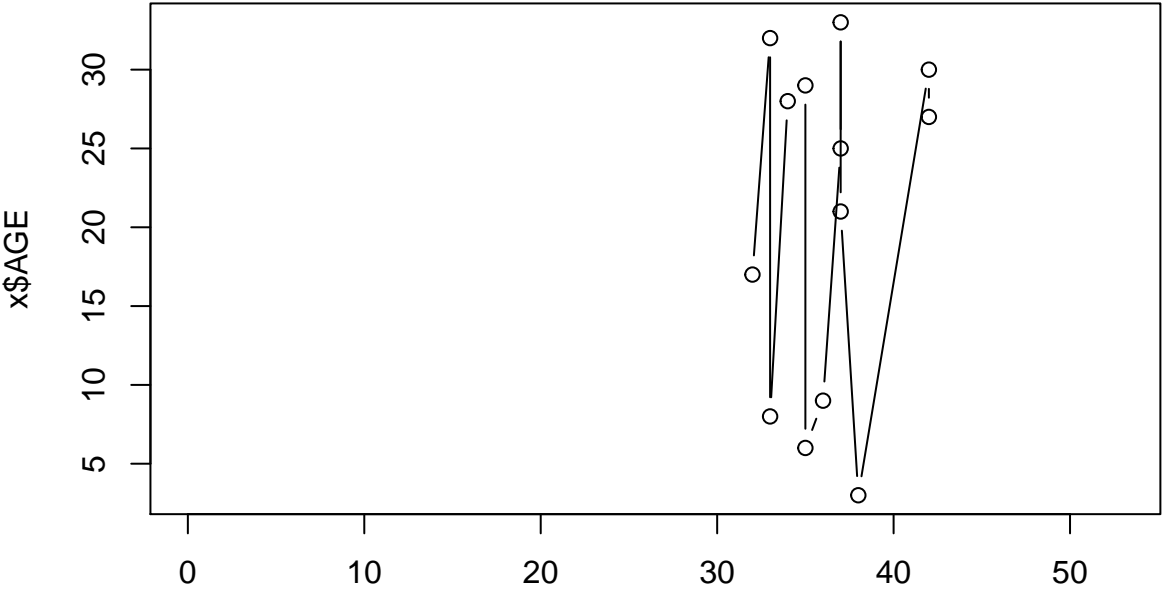
SALTO
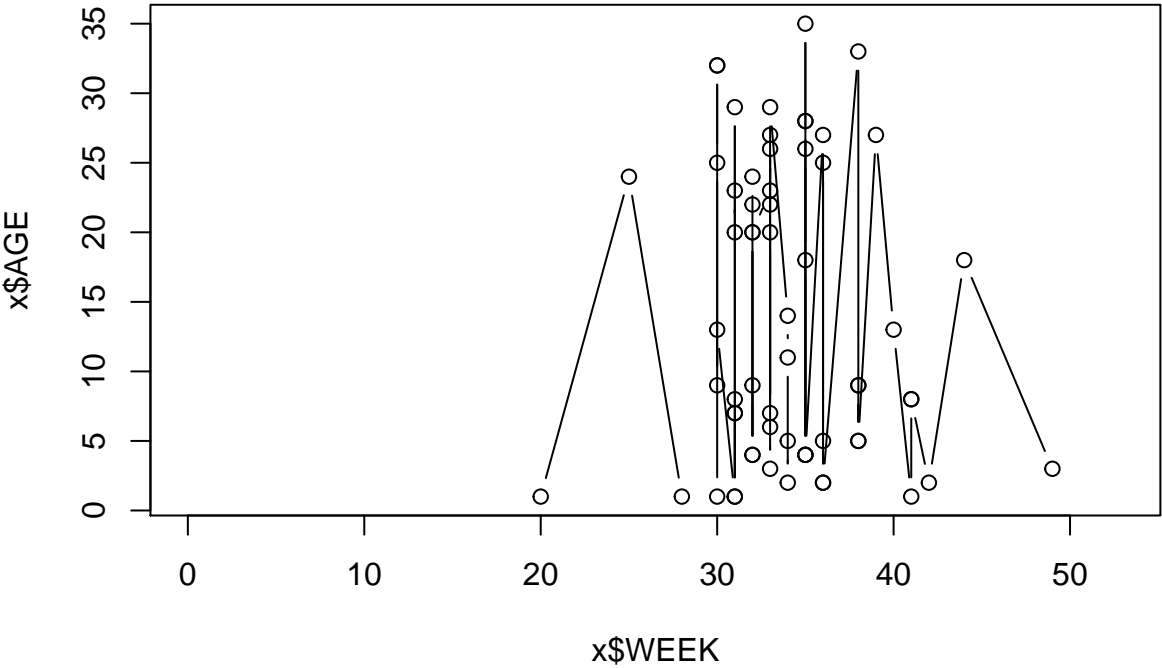
OSVALDO CRUZ
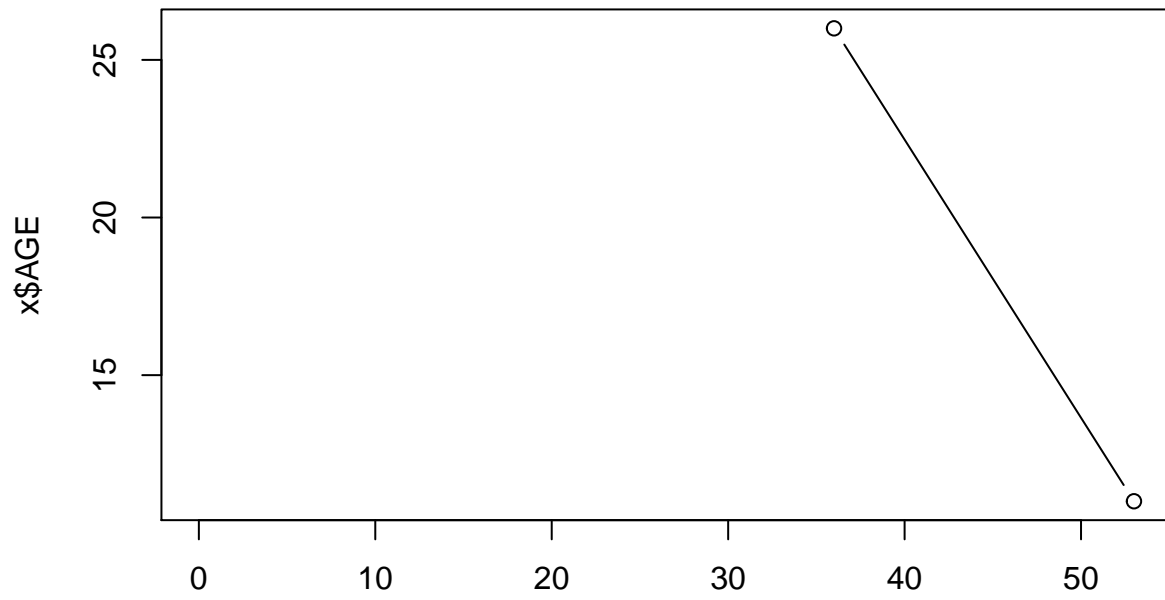
PARAIBUNA
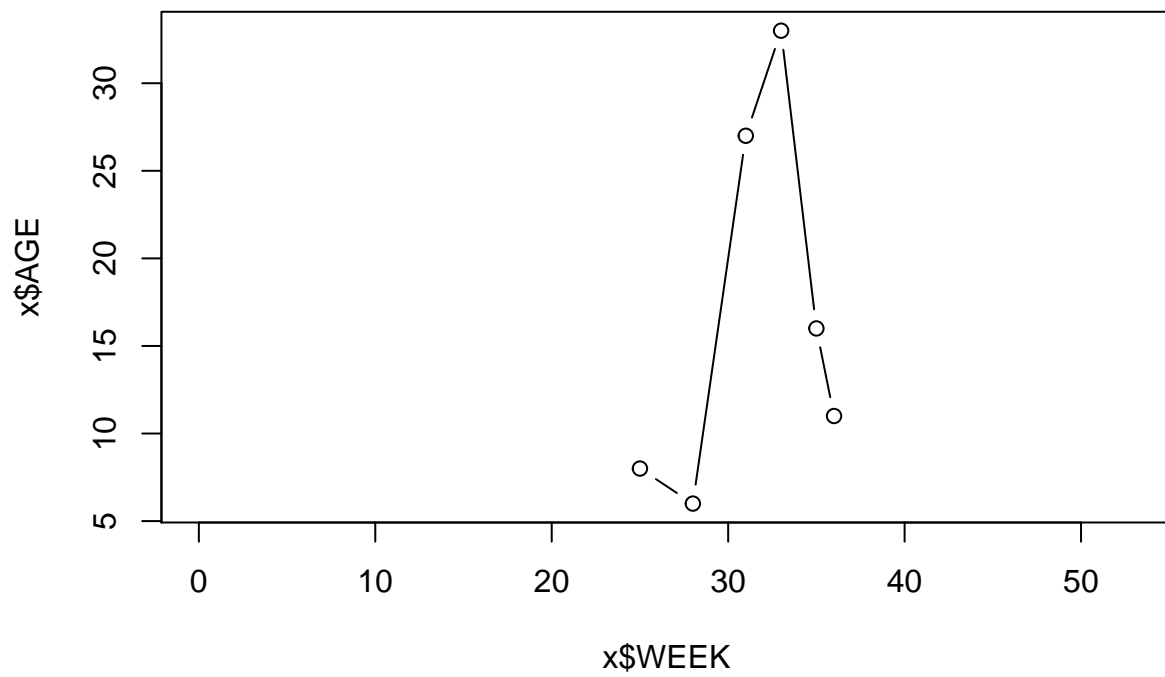
SANTA ALBERTINA
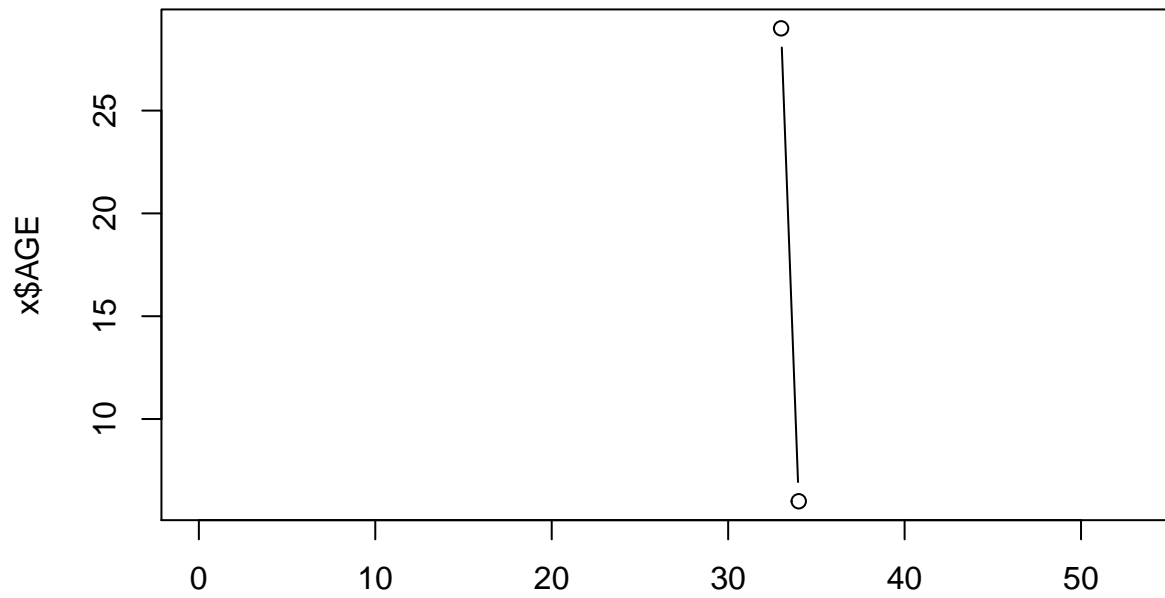
SAO JOAQUIM DA BARRA
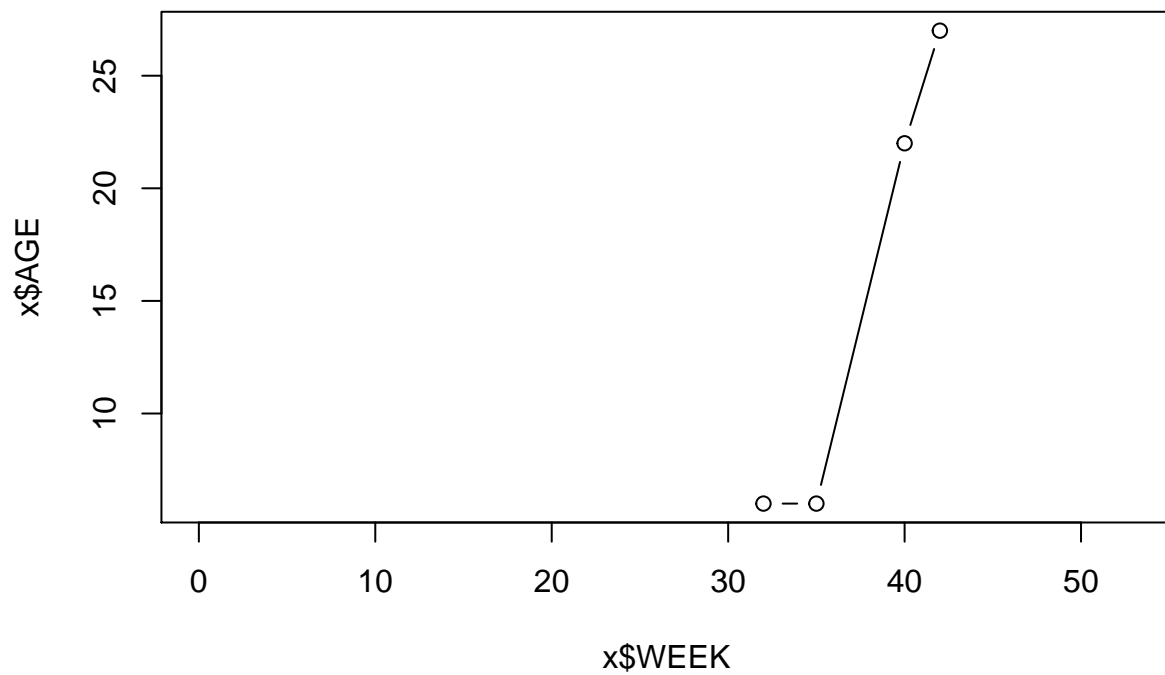
PRESIDENTE EPITACIO
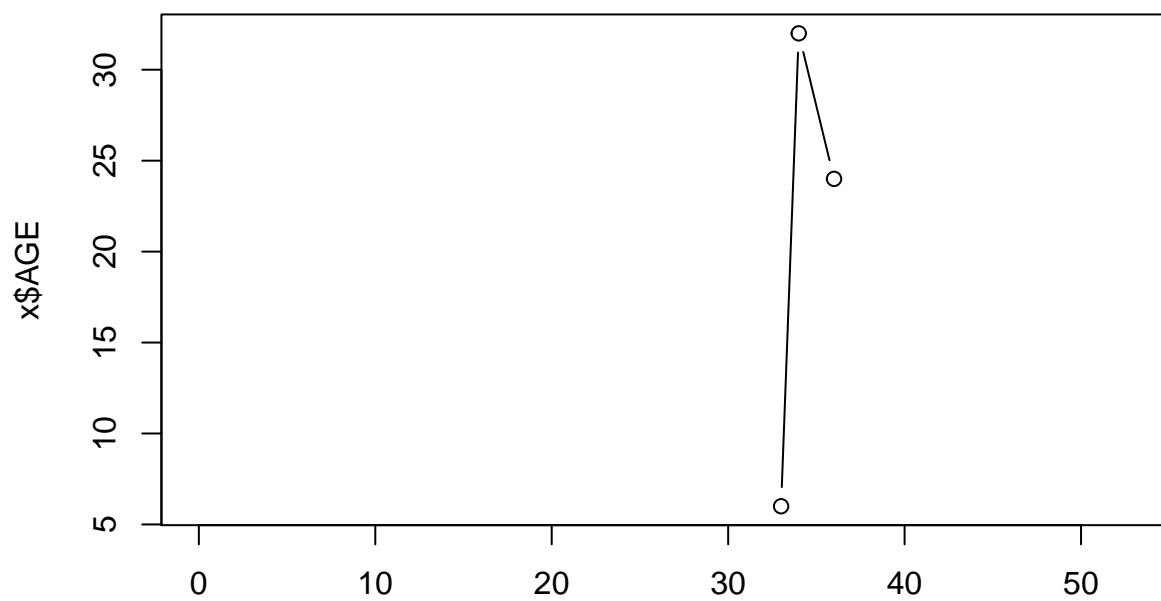
MATAO
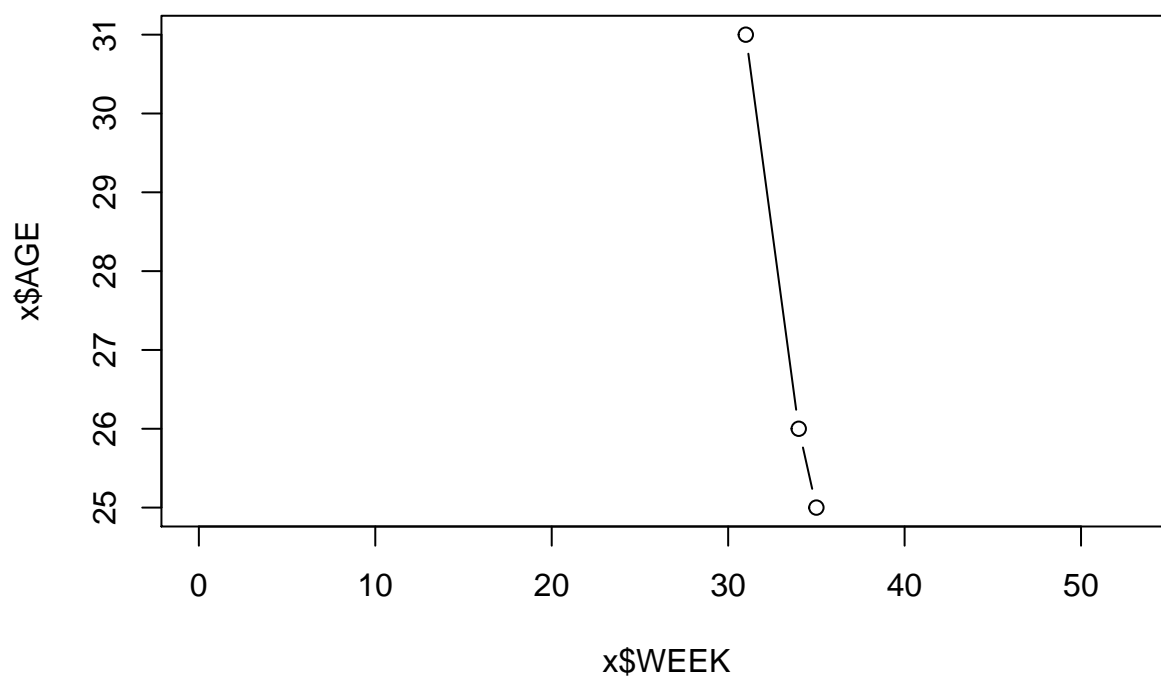
BORBOREMA
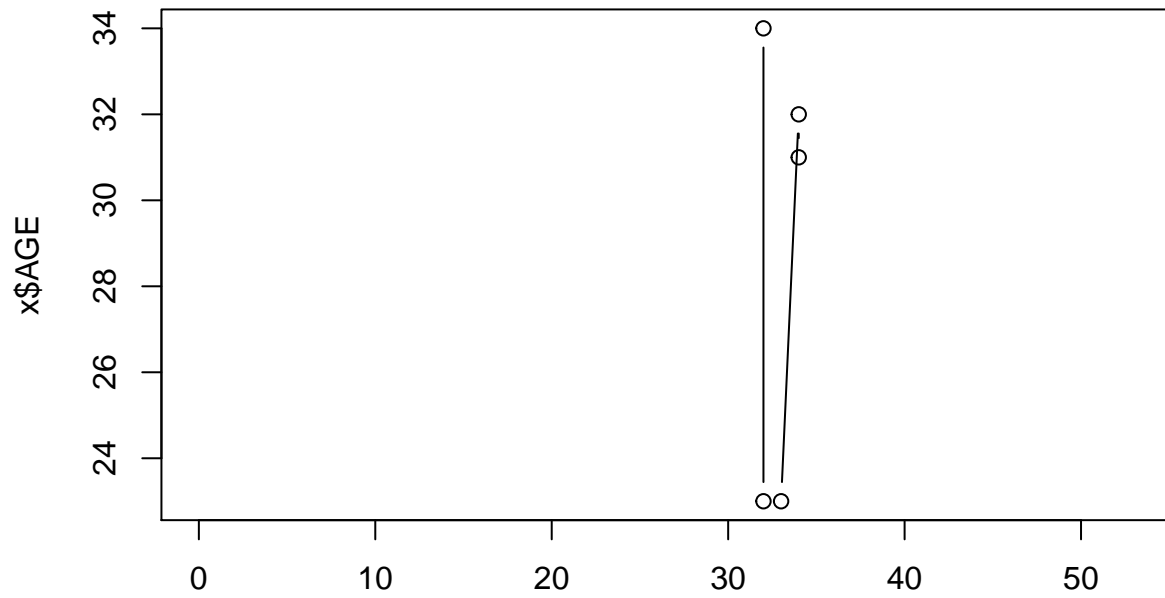
LOUVEIRA
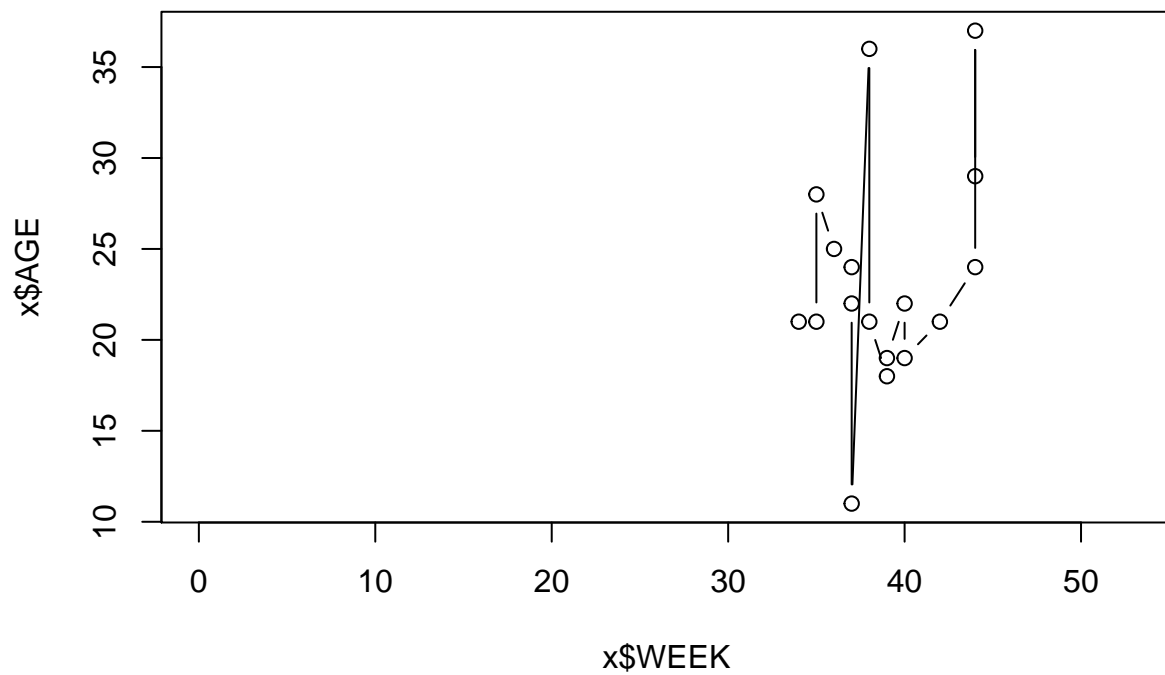
PRESIDENTE BERNARDES
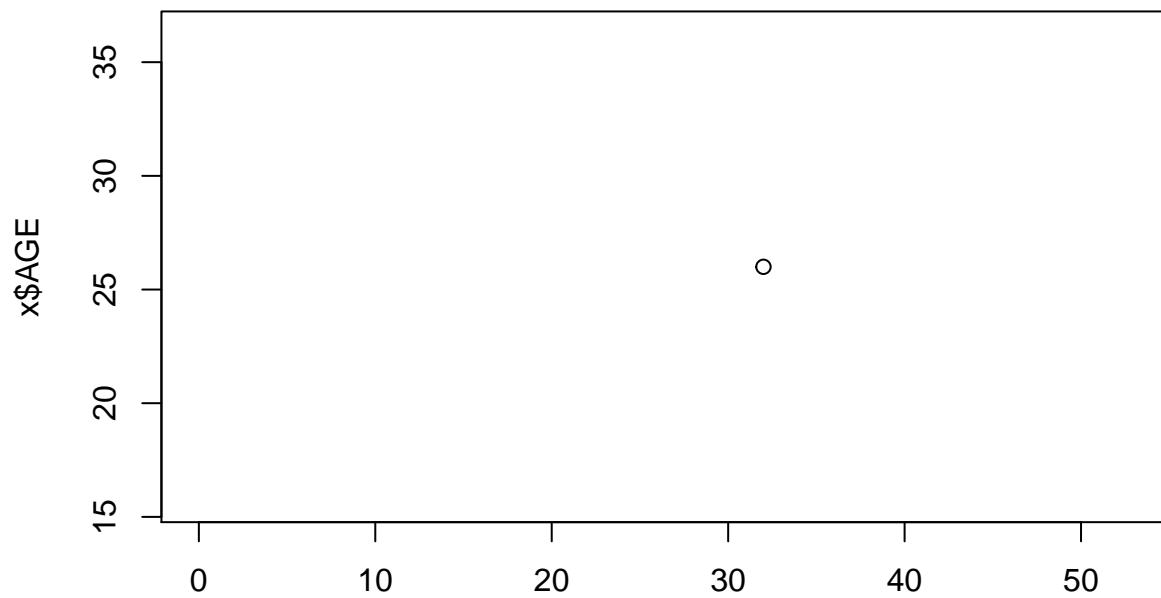
ITARARE
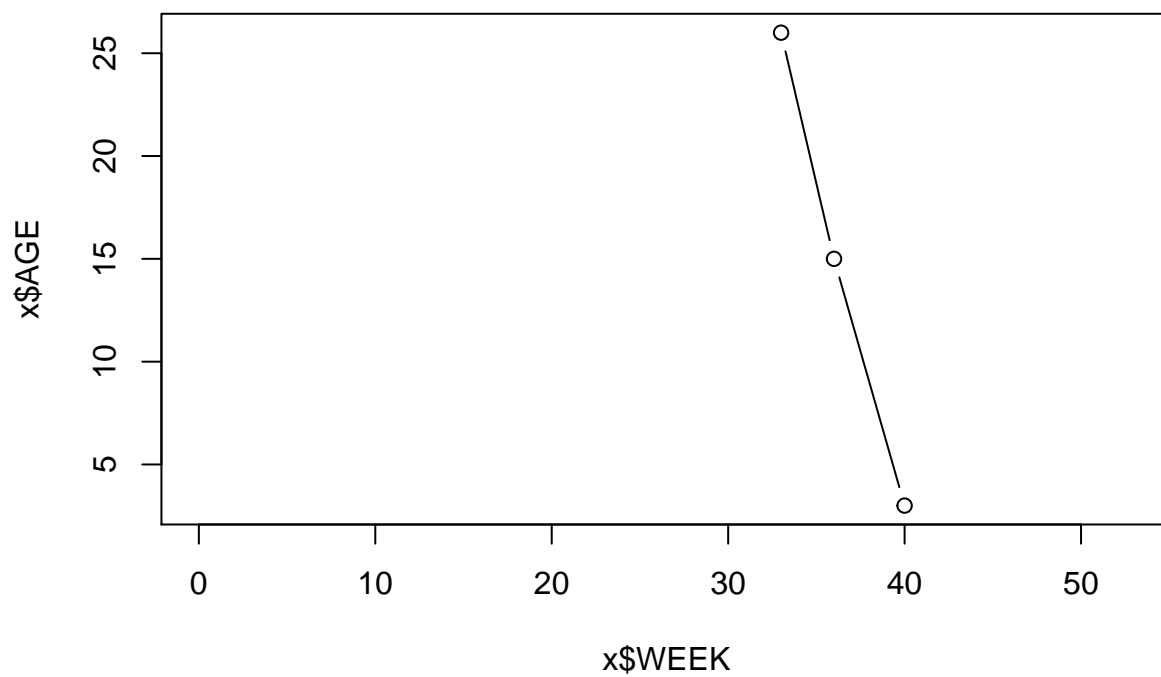
SANTO ANTONIO DA ALEGRIA

x$WEEK

BARRINHA

x$WEEK
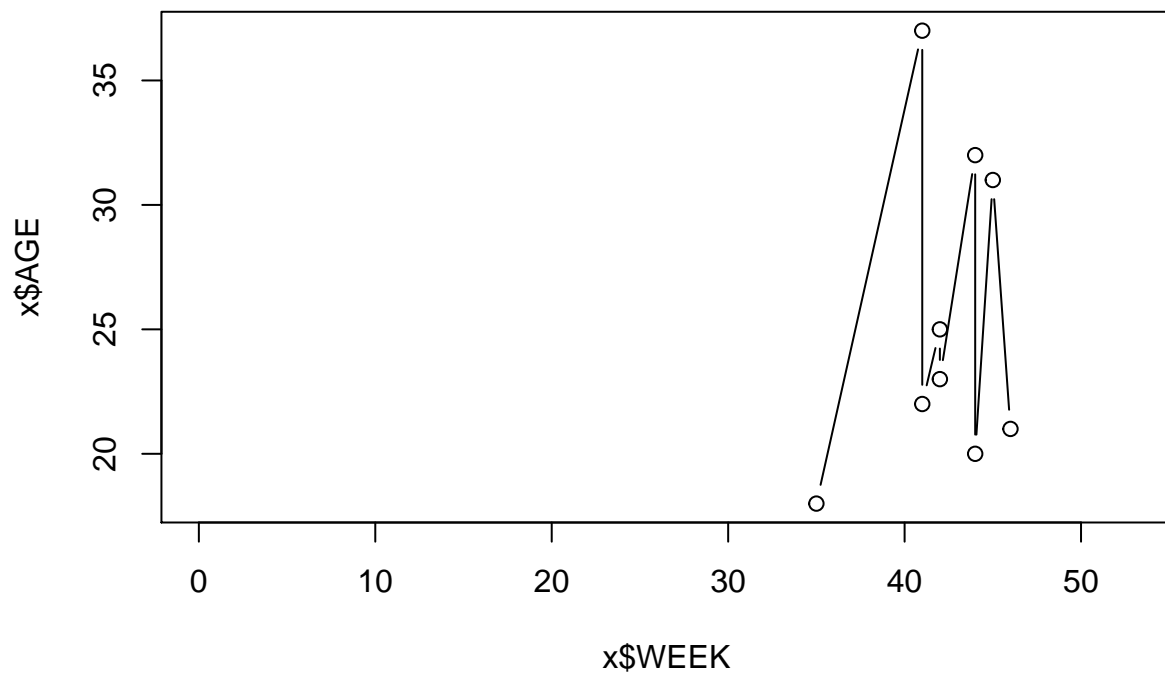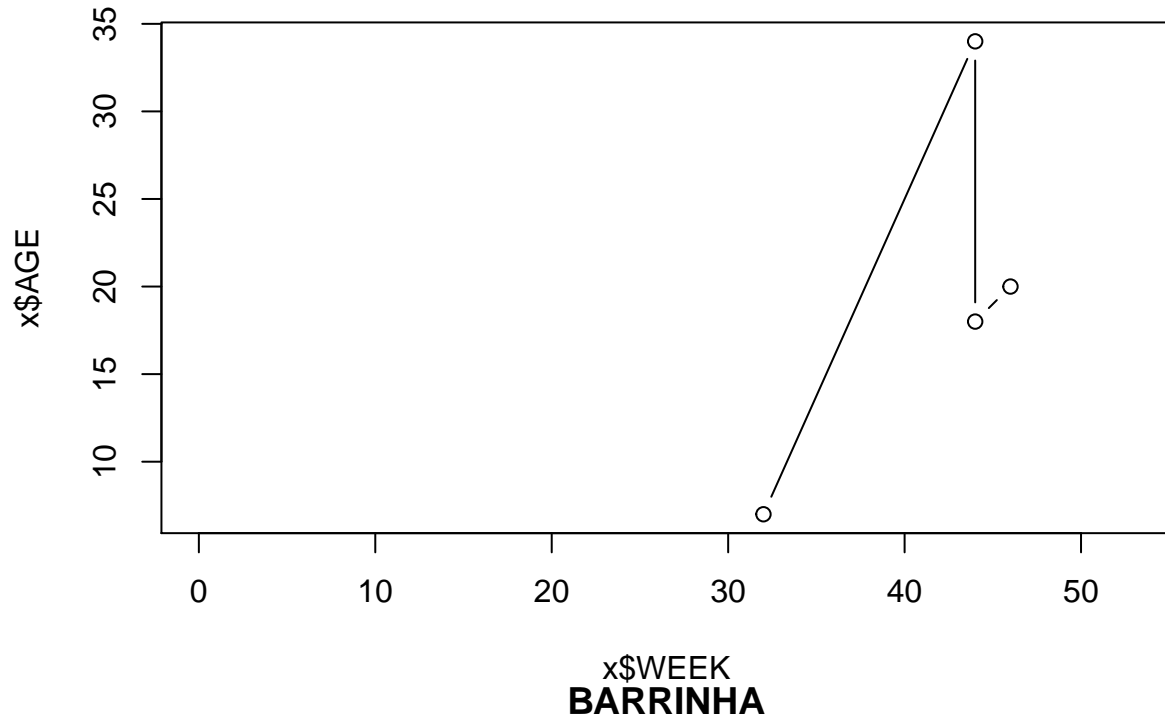
SANTA ROSA DE VITERBO

IGUAPE
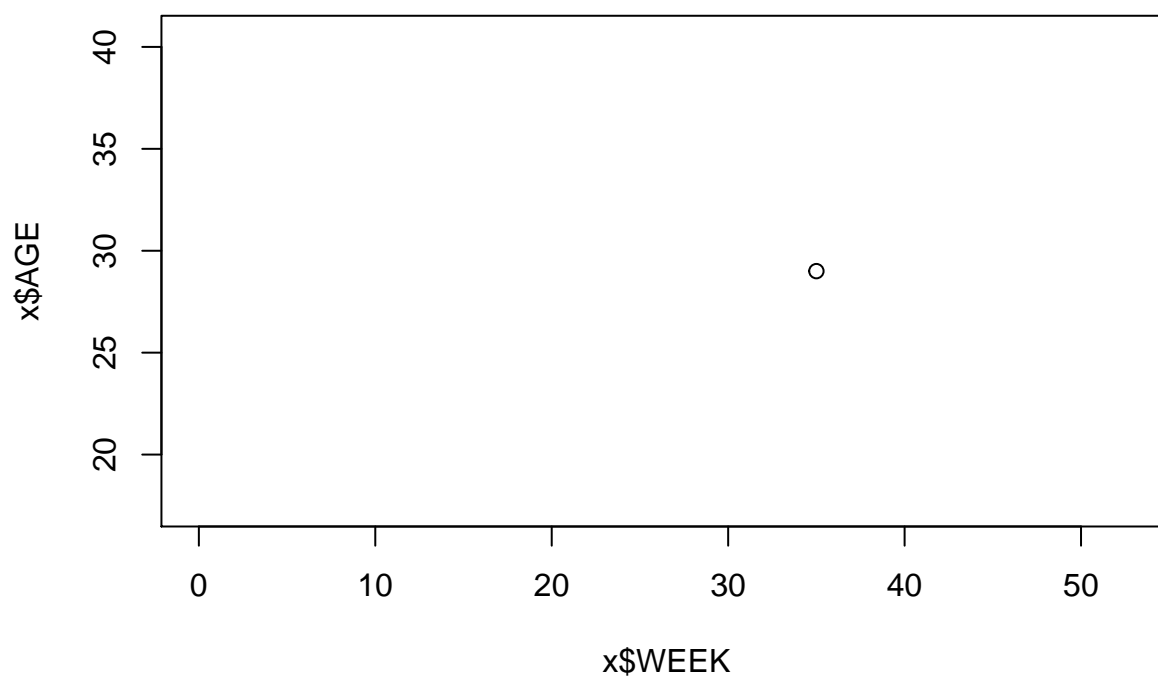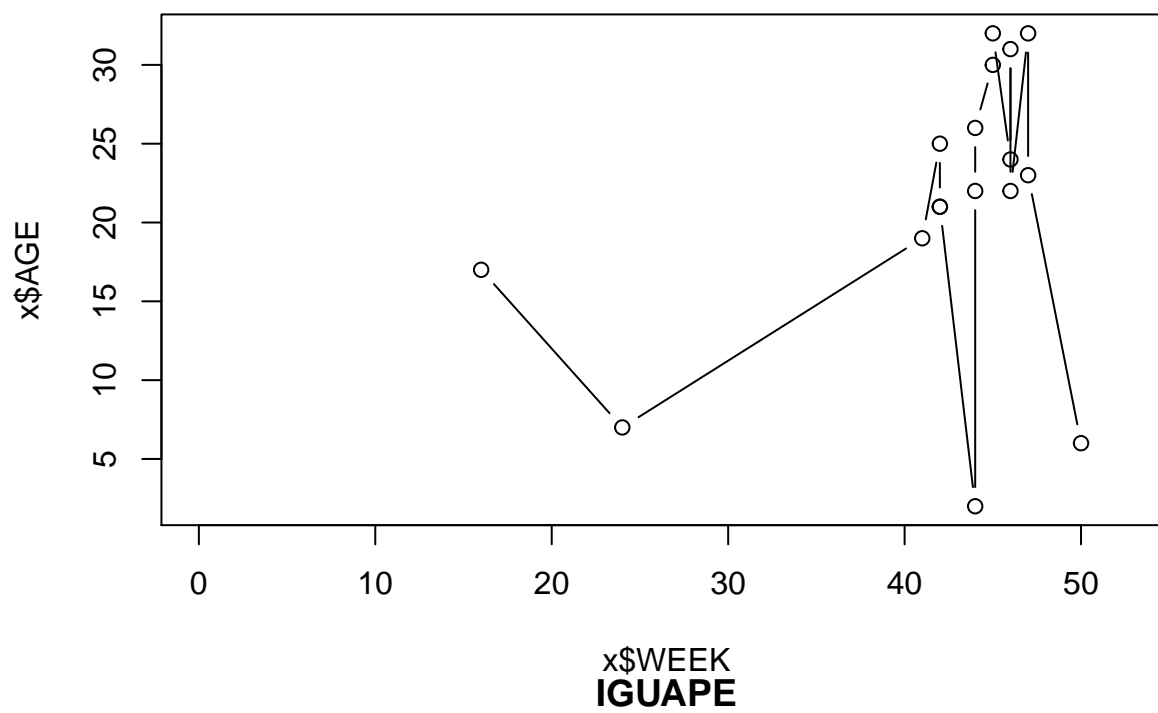
PORTO FELIZ

BRASILIA

139
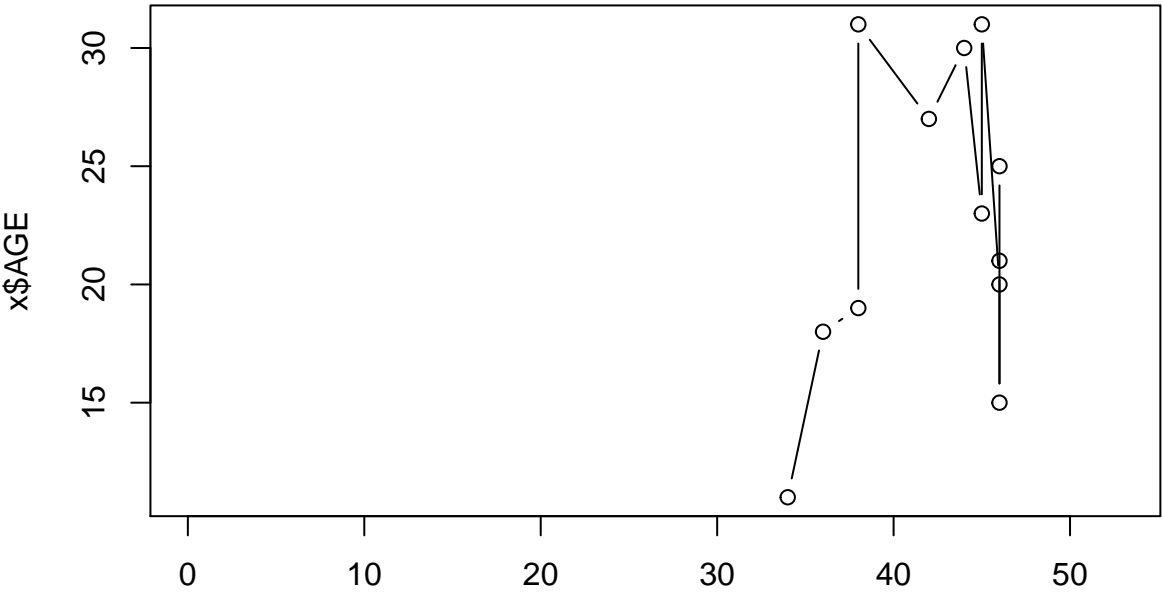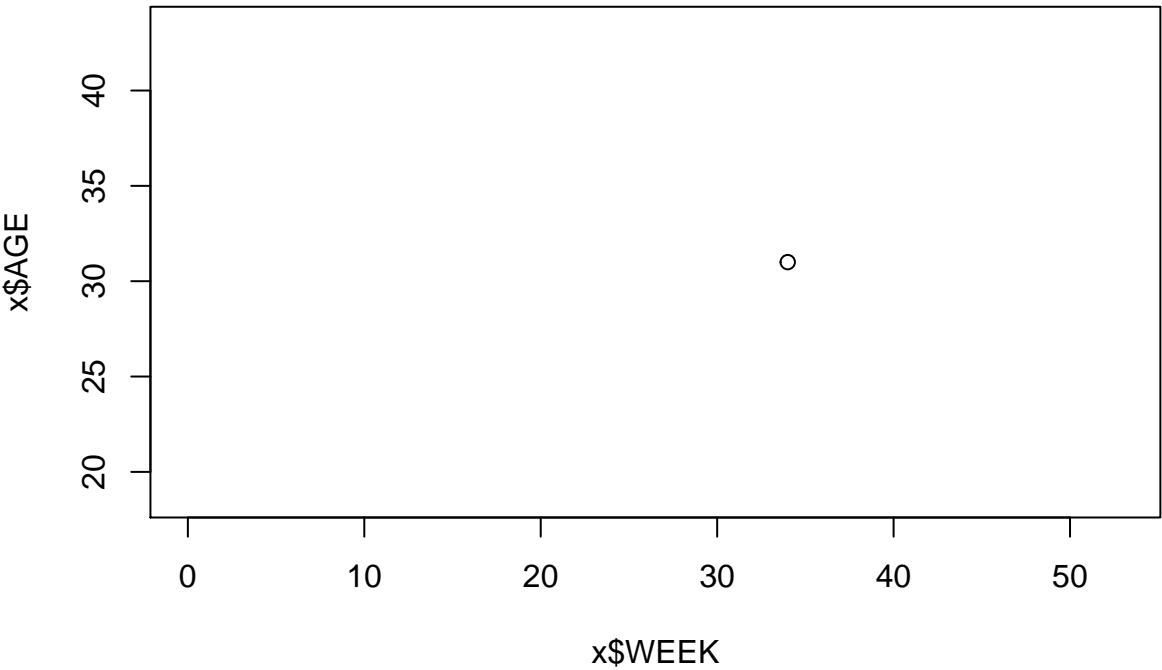
GUAREI
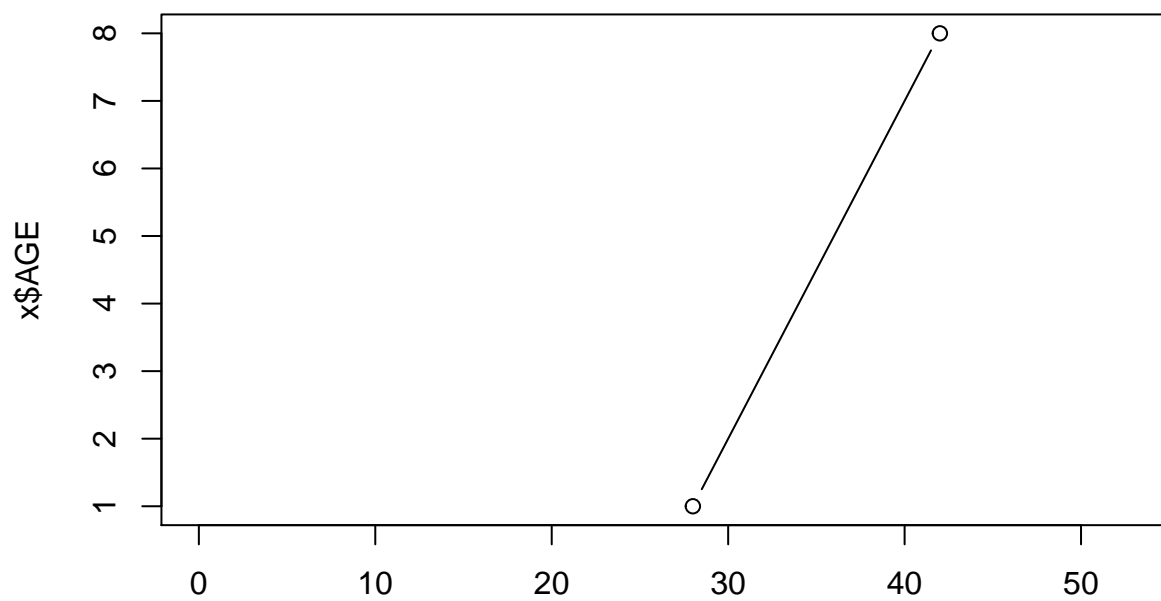
CERQUILHO

ITAPOLIS

x$WEEK

SAO PEDRO

x$WEEK
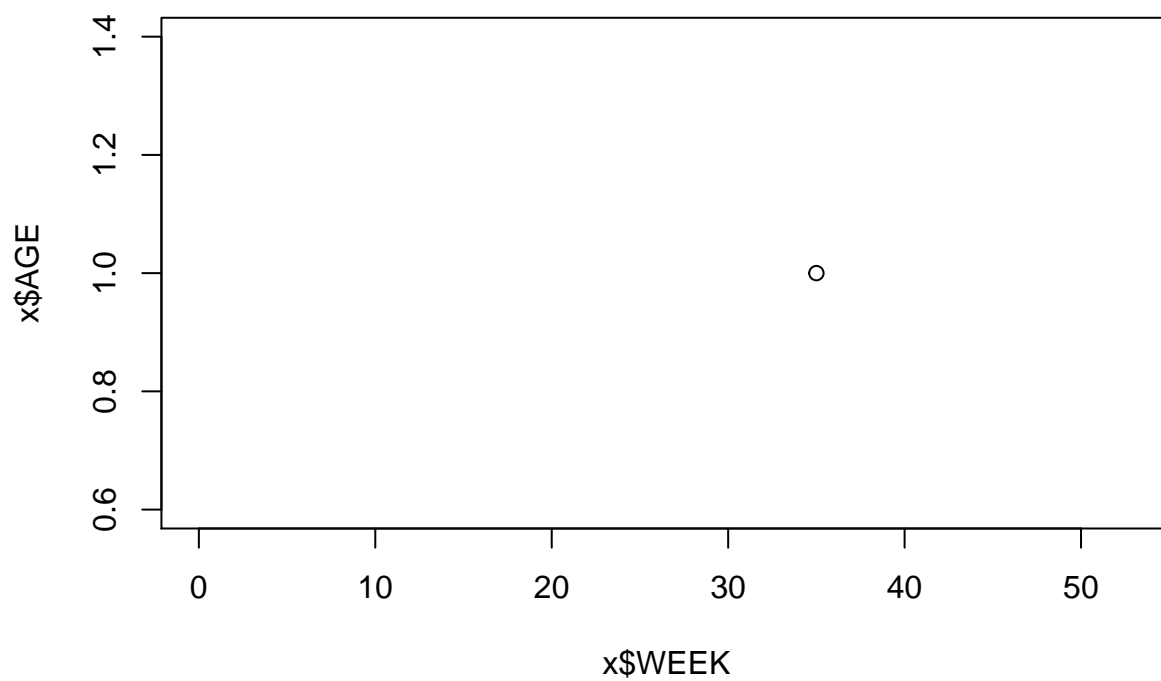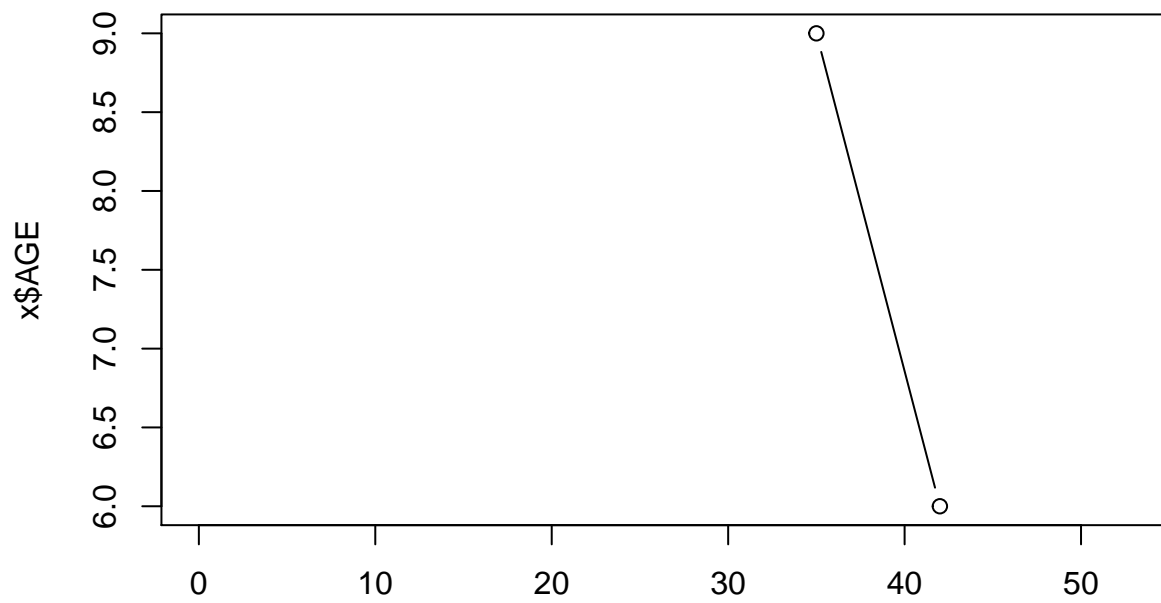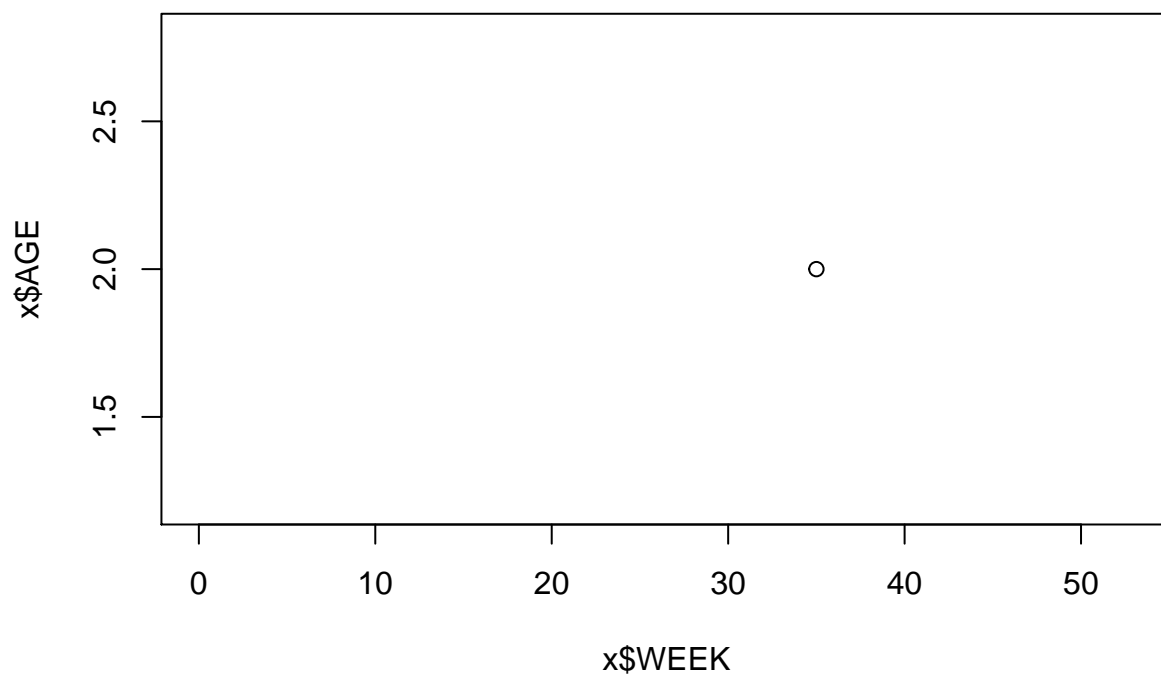
SAO ROQUE
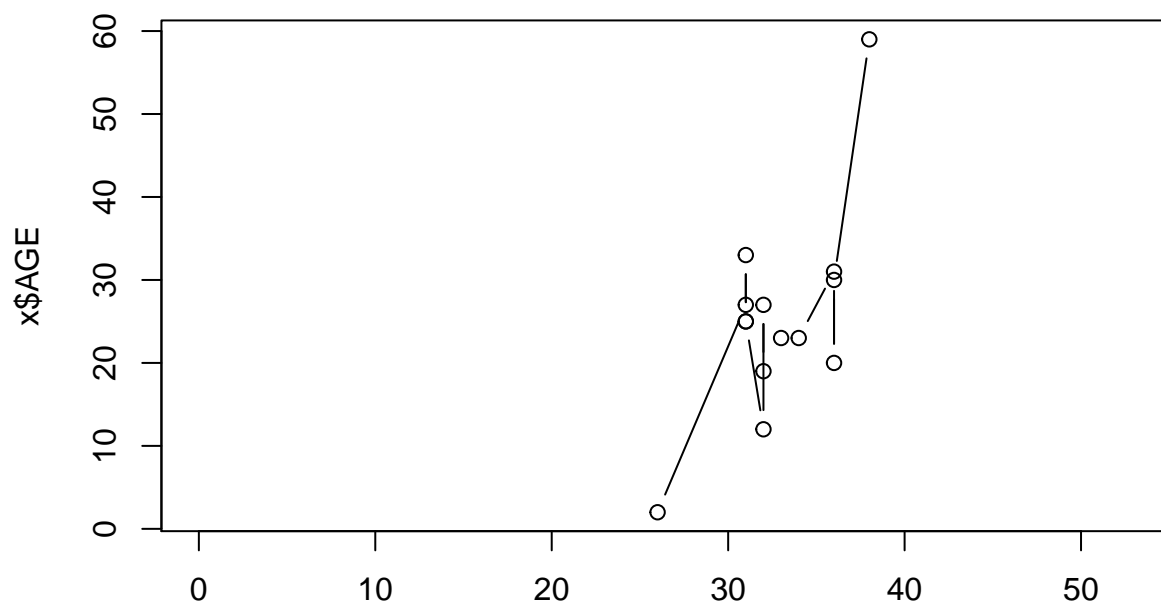
SAO SEBASTIAO
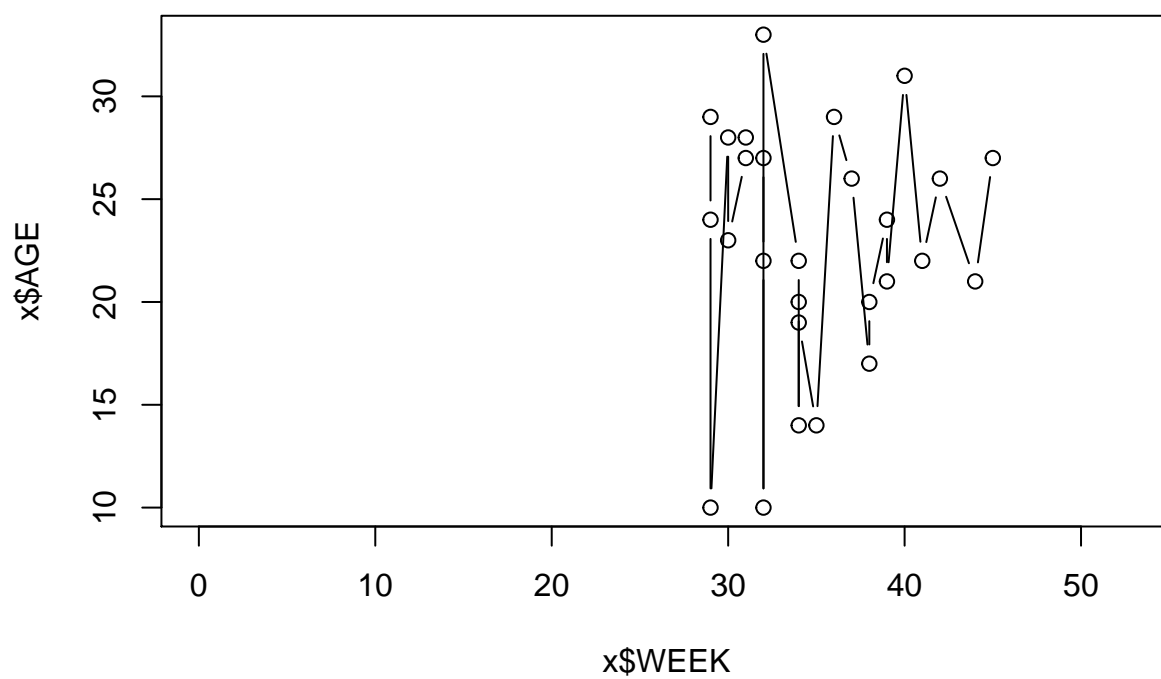
OUROESTE
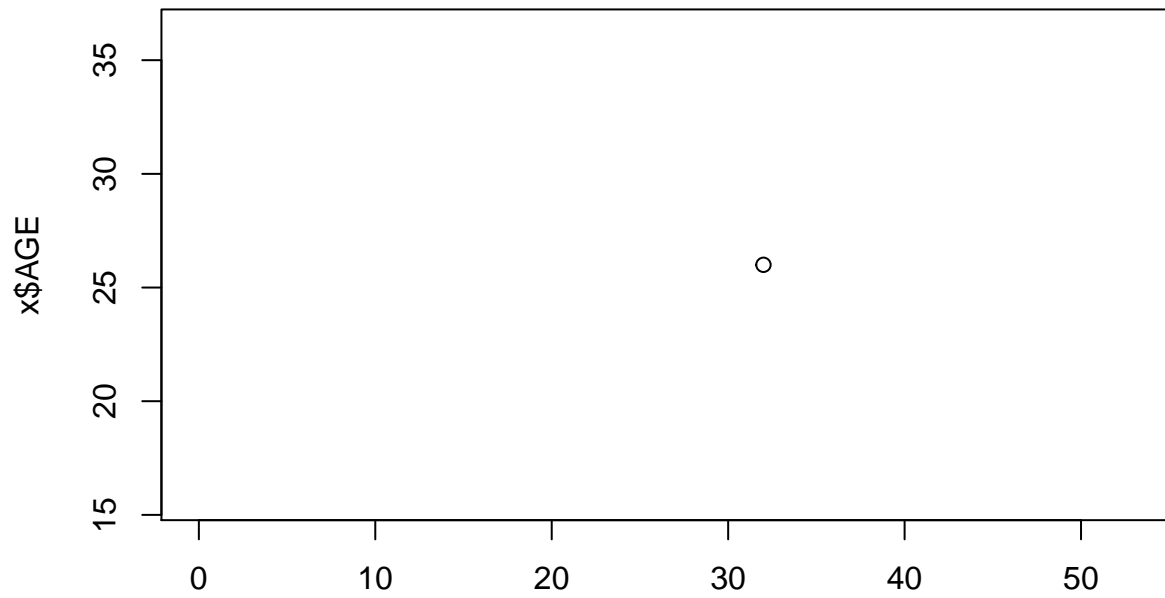
SAO SEBASTIAO DA GRAMA
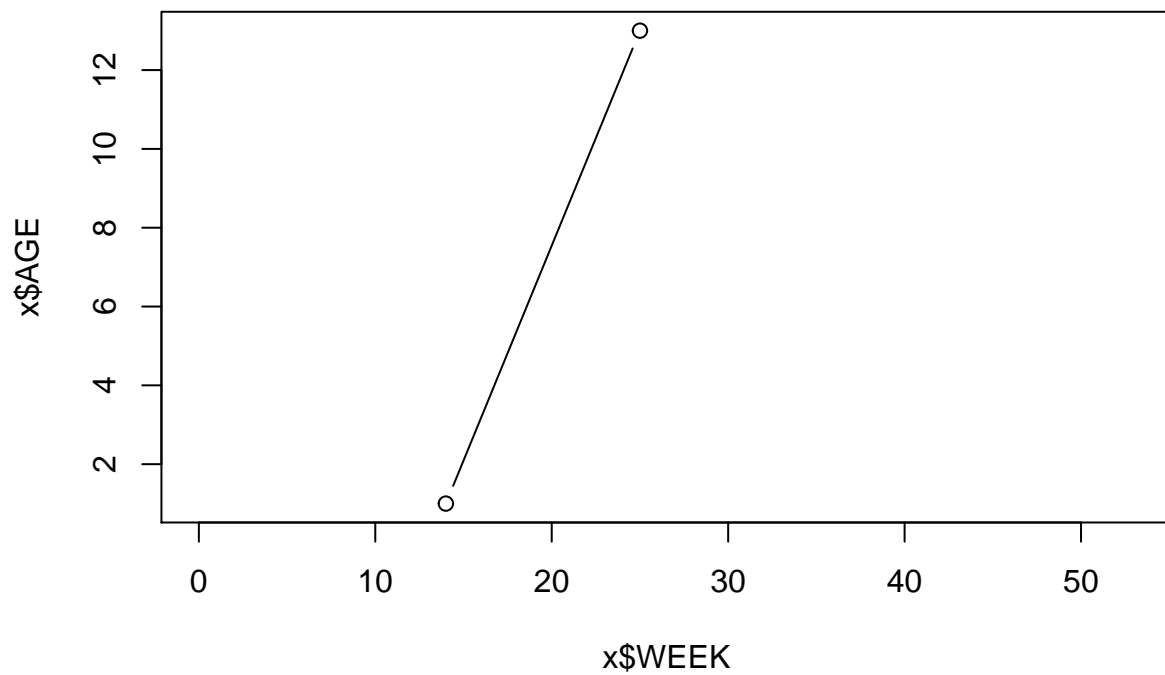
JUQUIA
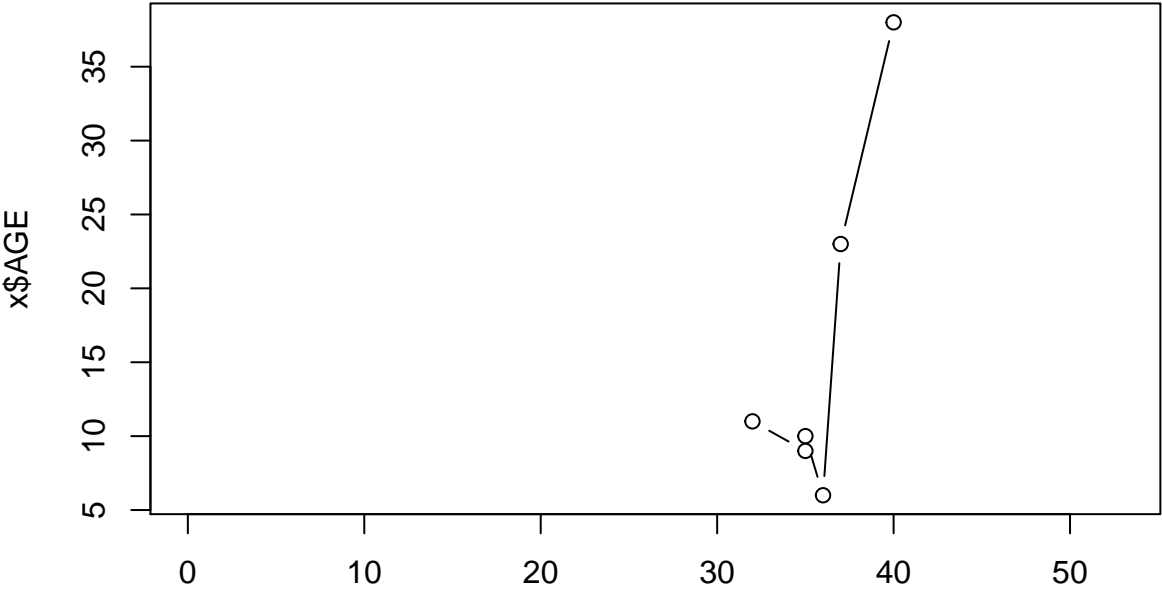
SAO SIMAO
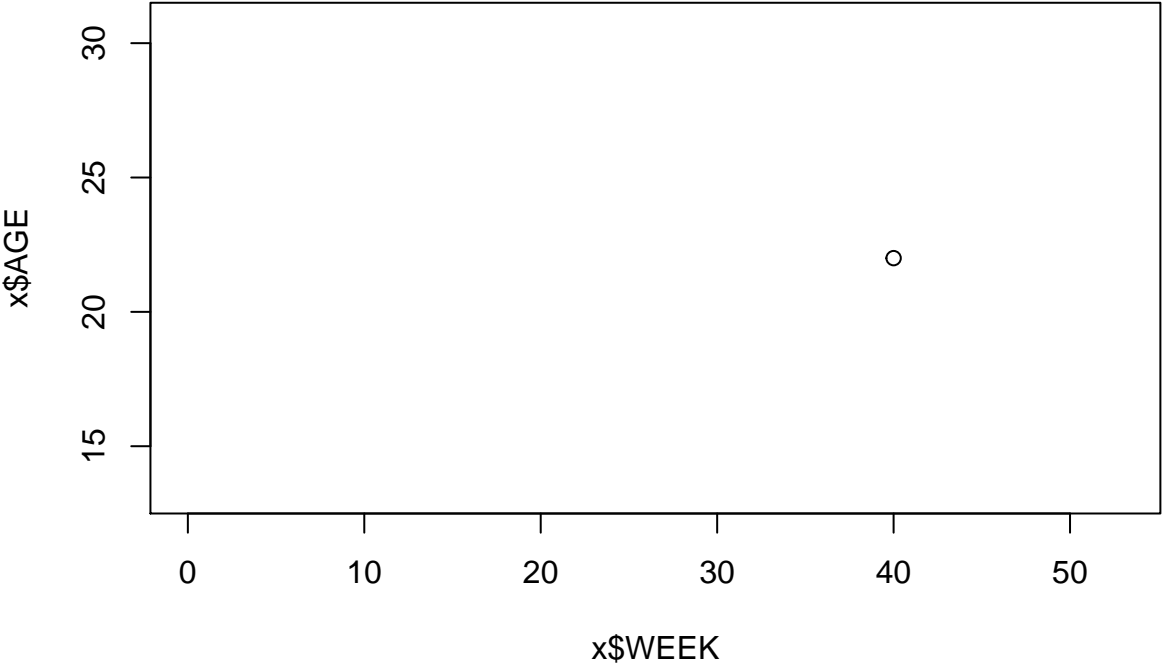
SAO VICENTE
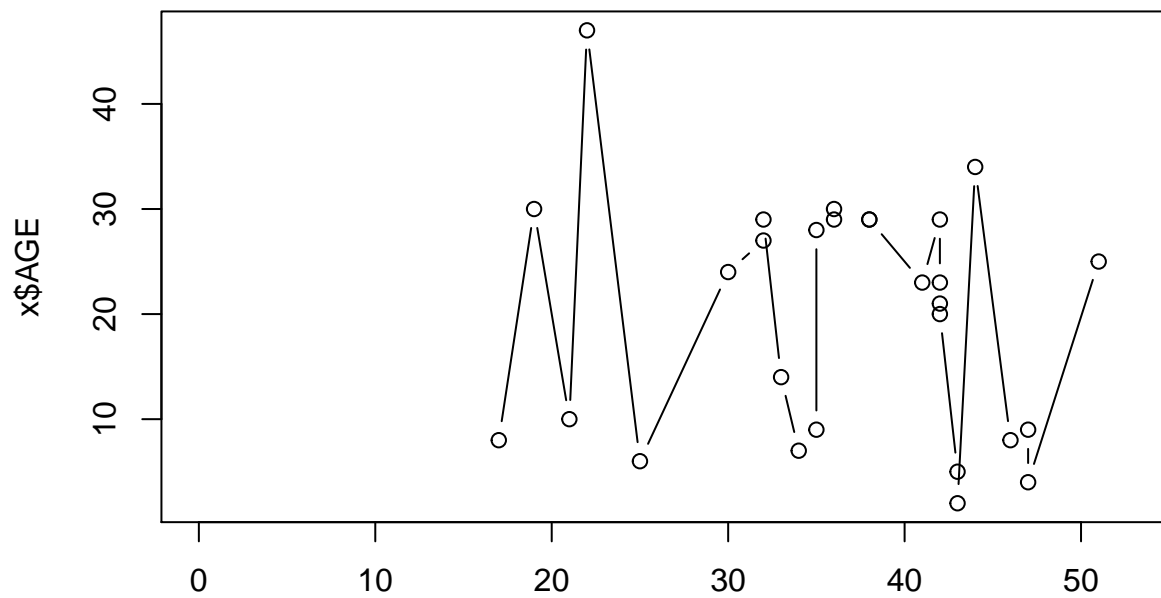
ILHA COMPRIDA
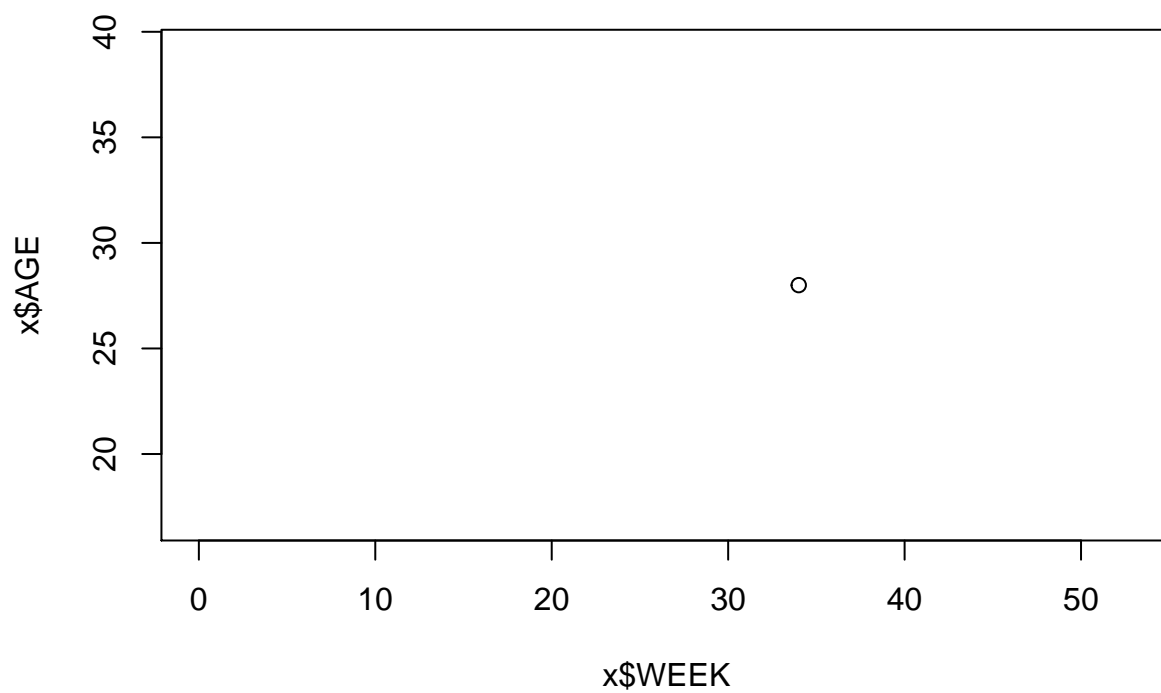
SARAPUI
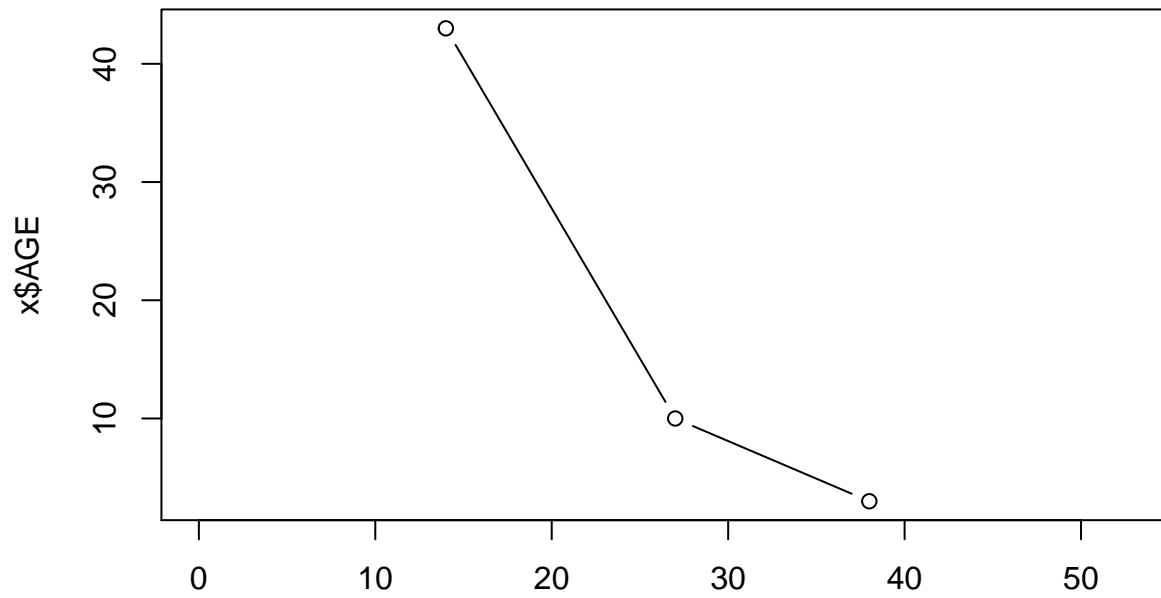
SARUTAIA

SERRA NEGRA

x$WEEK

SERRANA

x$WEEK
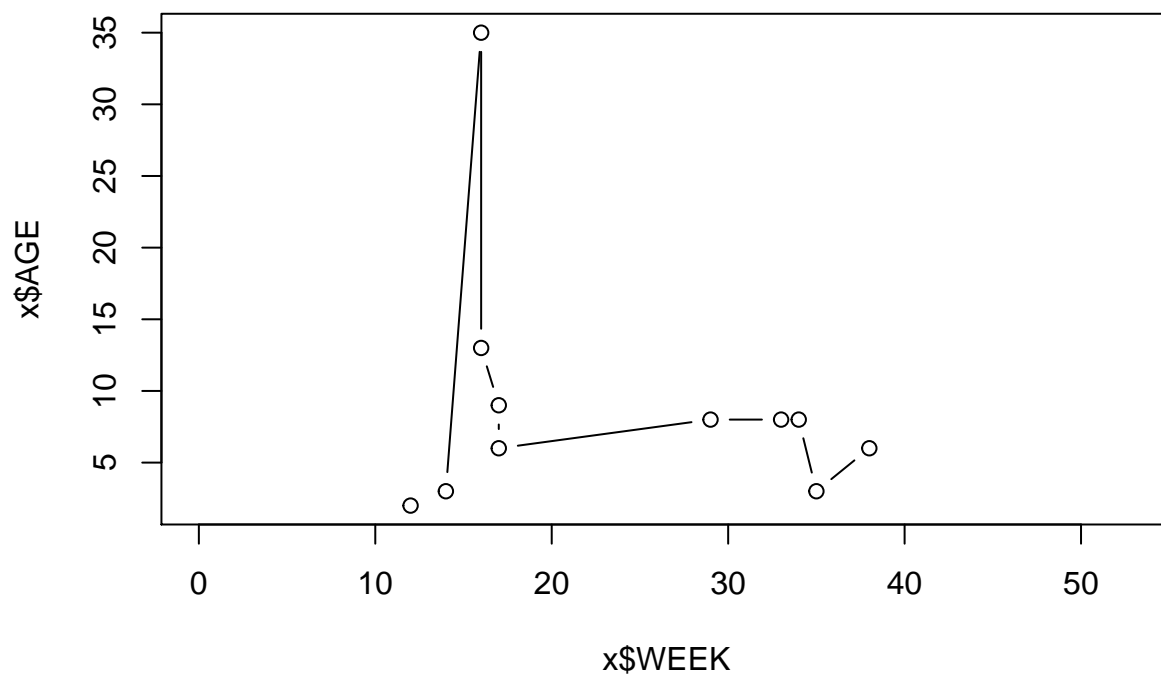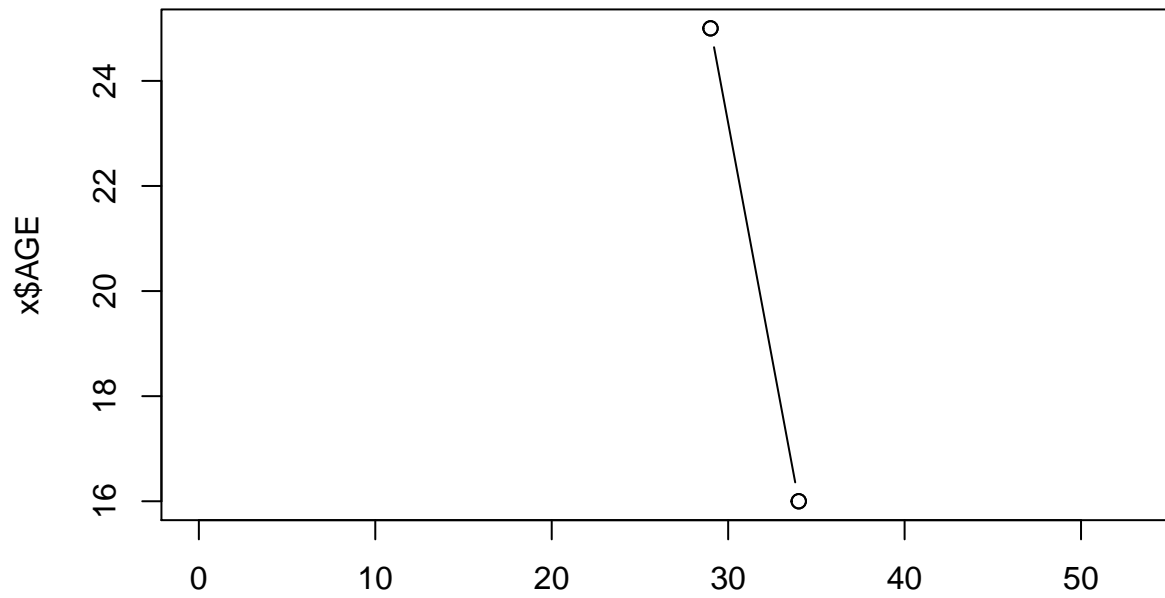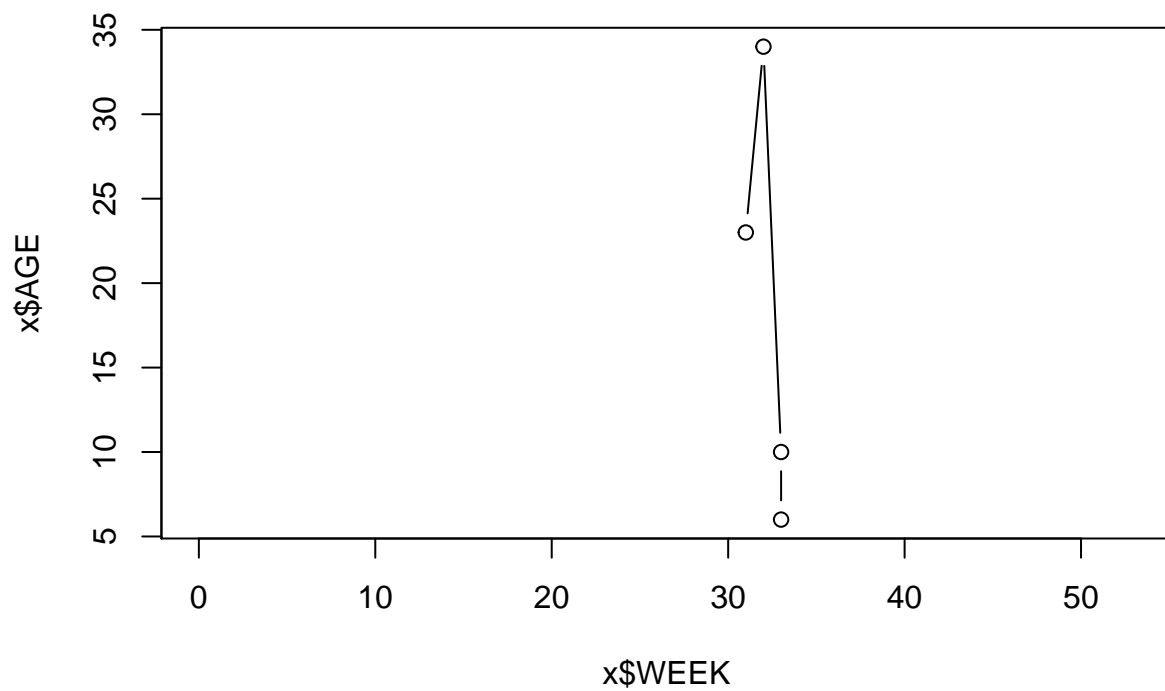
SERTAOZINHO
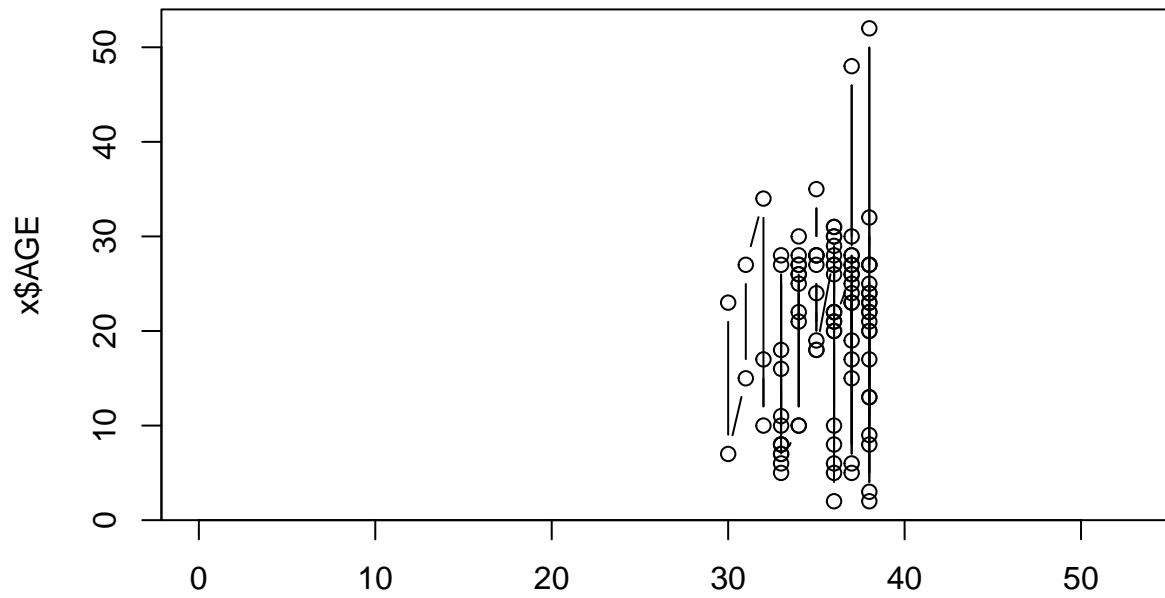
IBITINGA
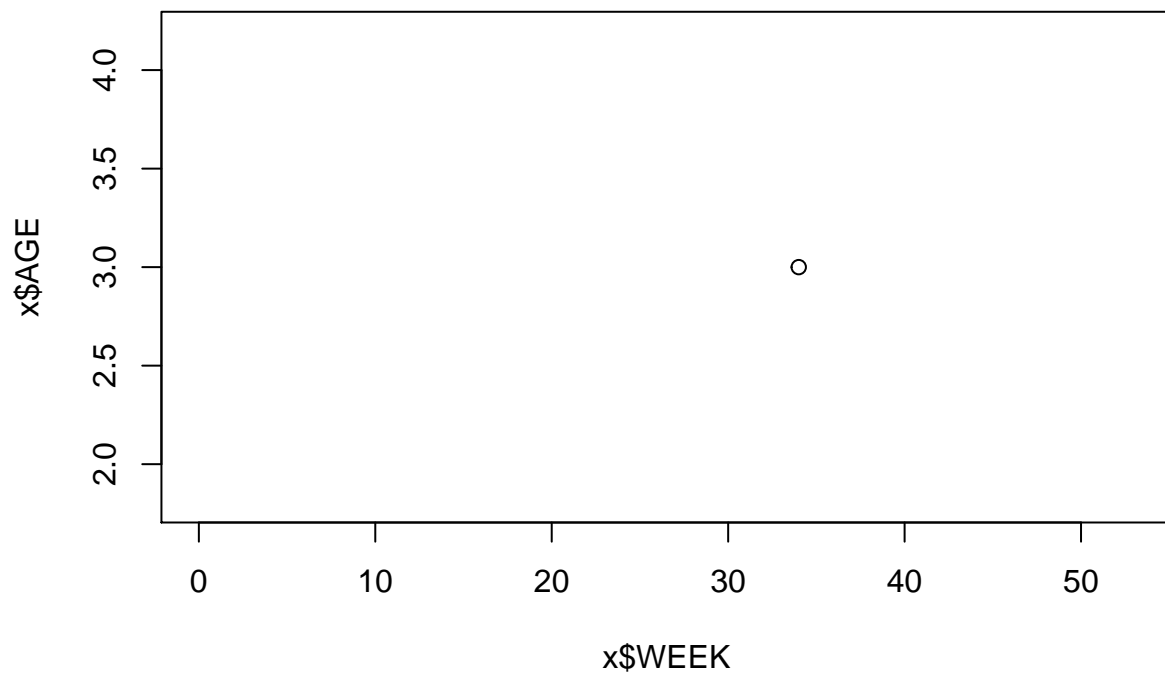
ALTINOPOLIS
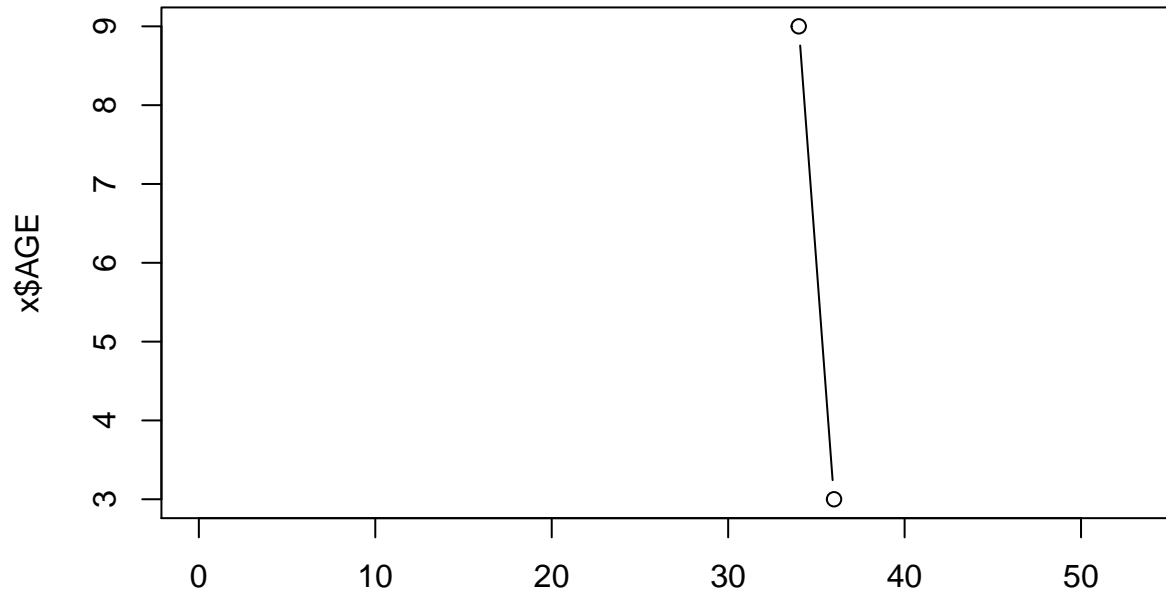
MONTE ALTO
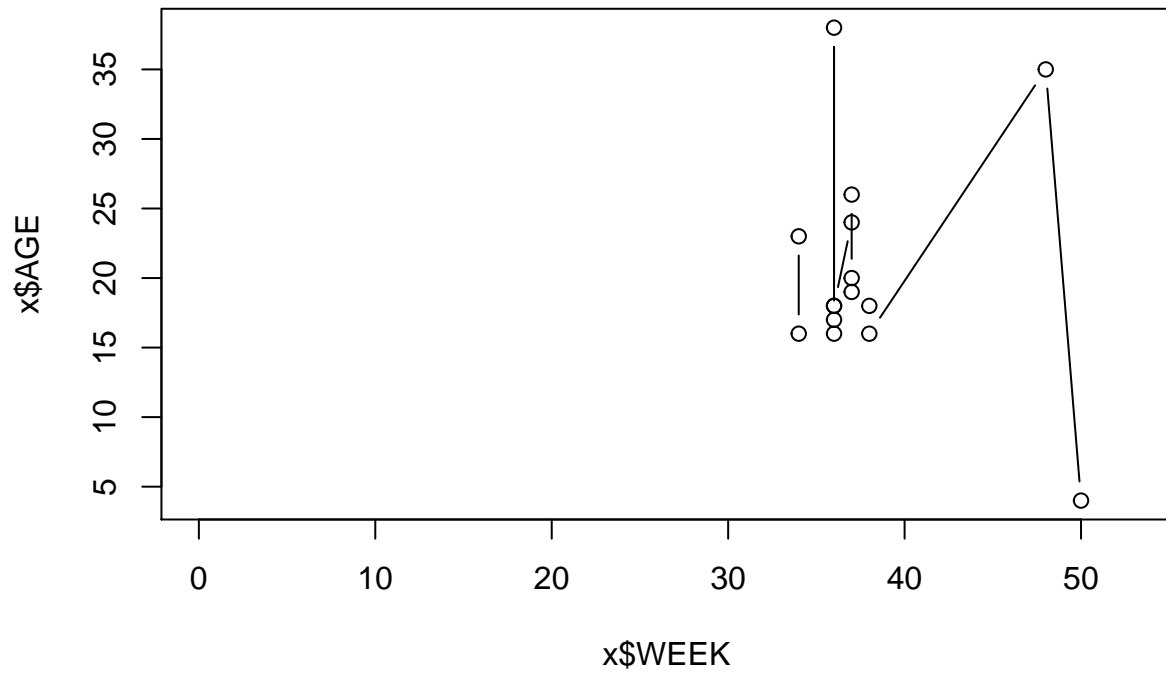
PEDRO DE TOLEDO
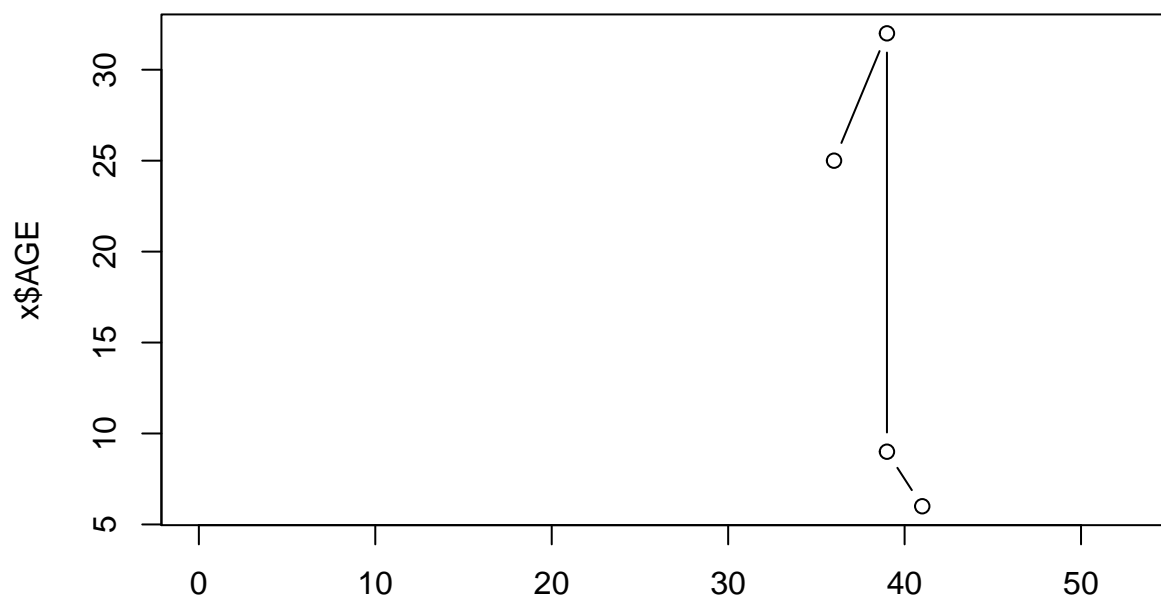
SETE BARRAS
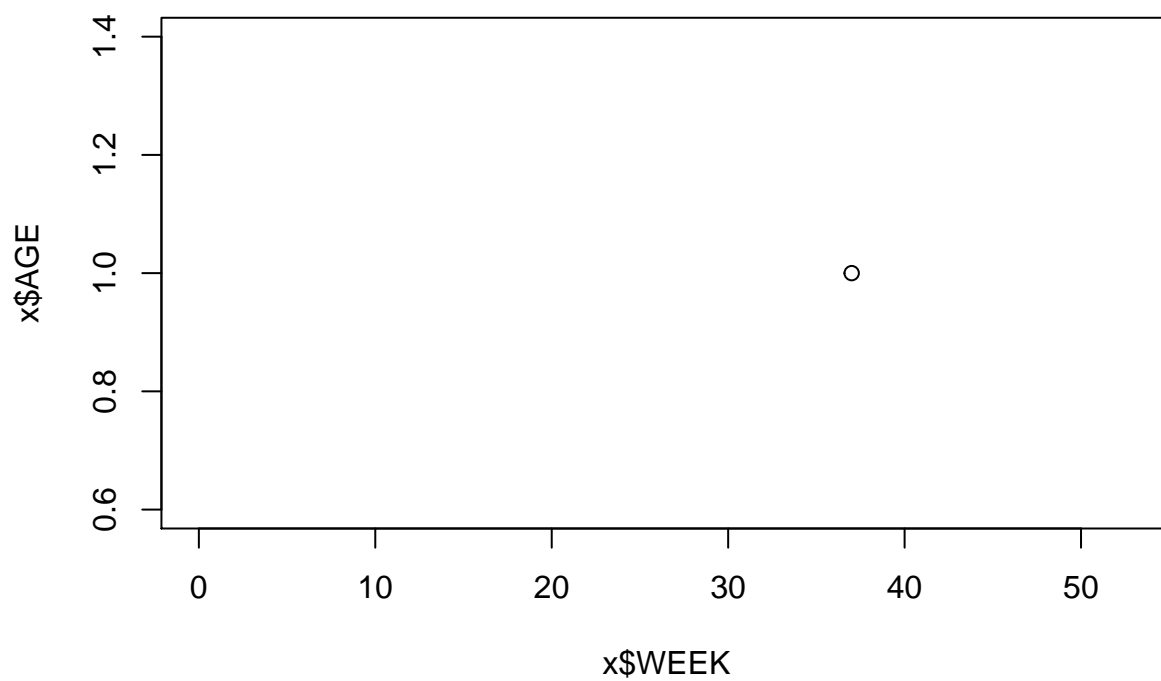
SILVEIRAS

SOCORRO

SOROCABA

x$WEEK

BARBOSA

x$WEEK

NOVA ALIANCA

x$WEEK

PEDRA BELA

x$WEEK

153

SUD MENNUCCI

x$WEEK

SUMARE

x$WEEK
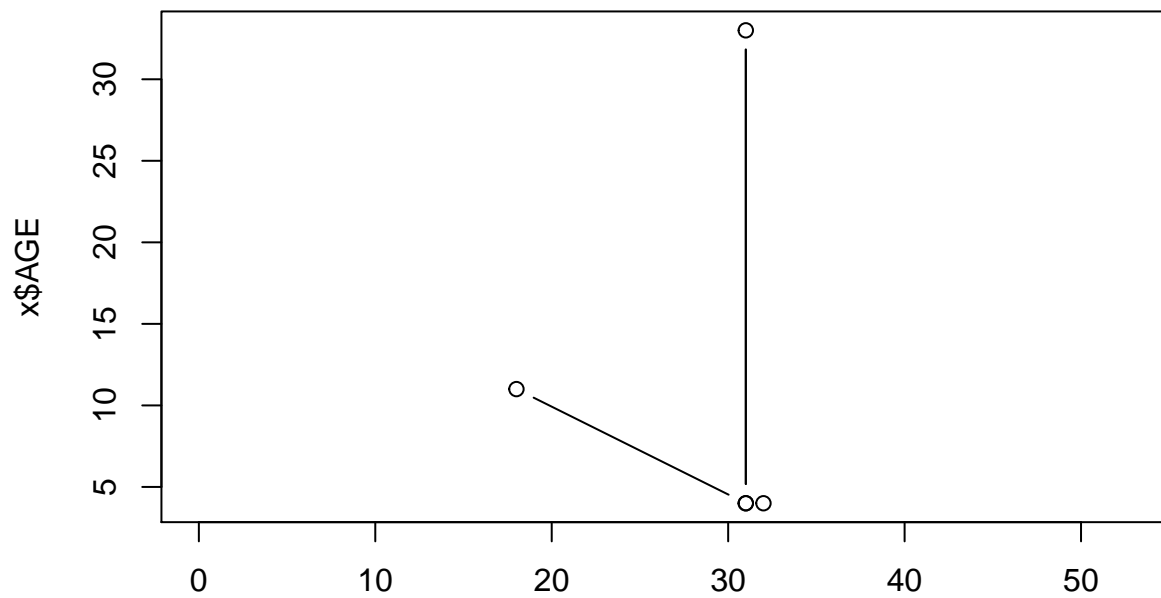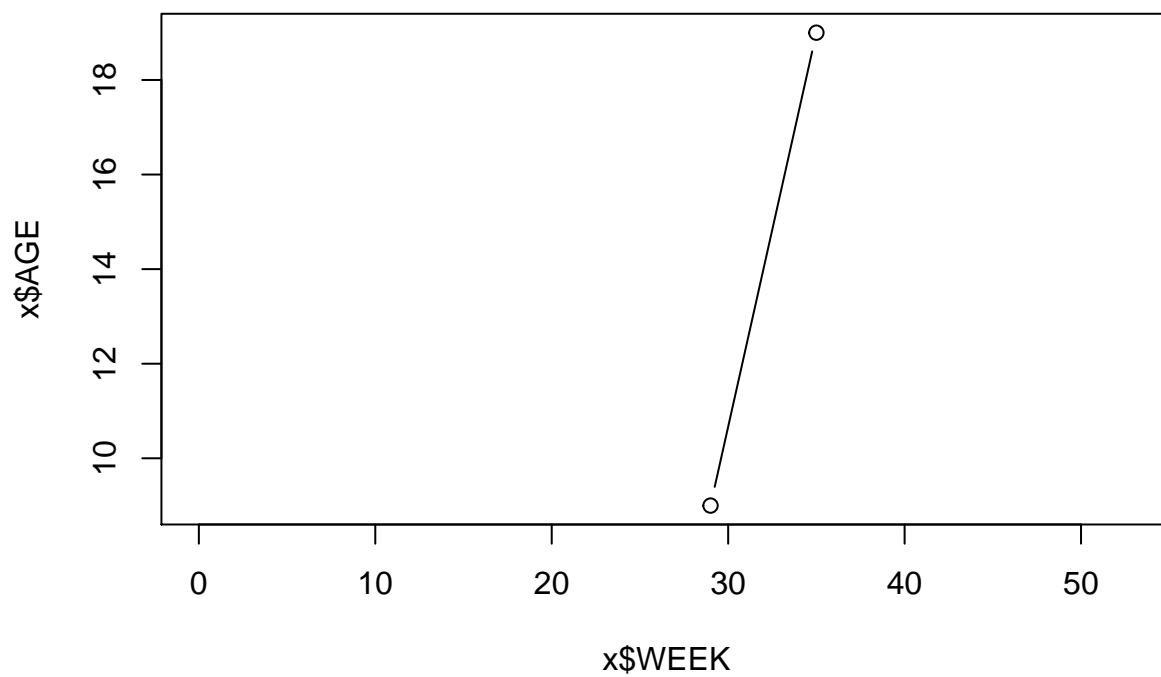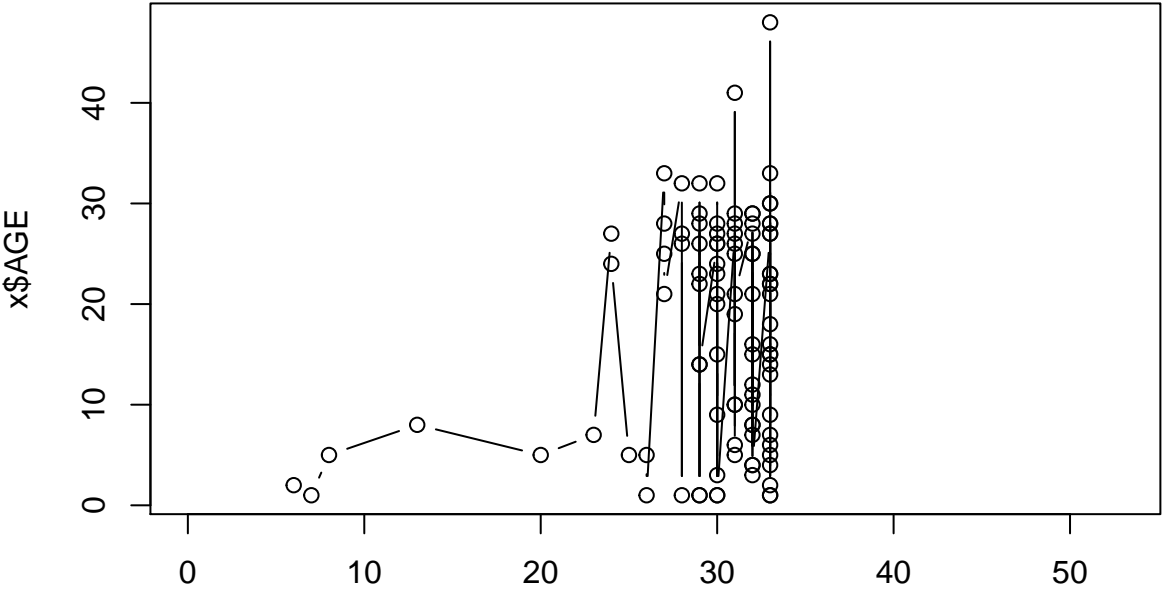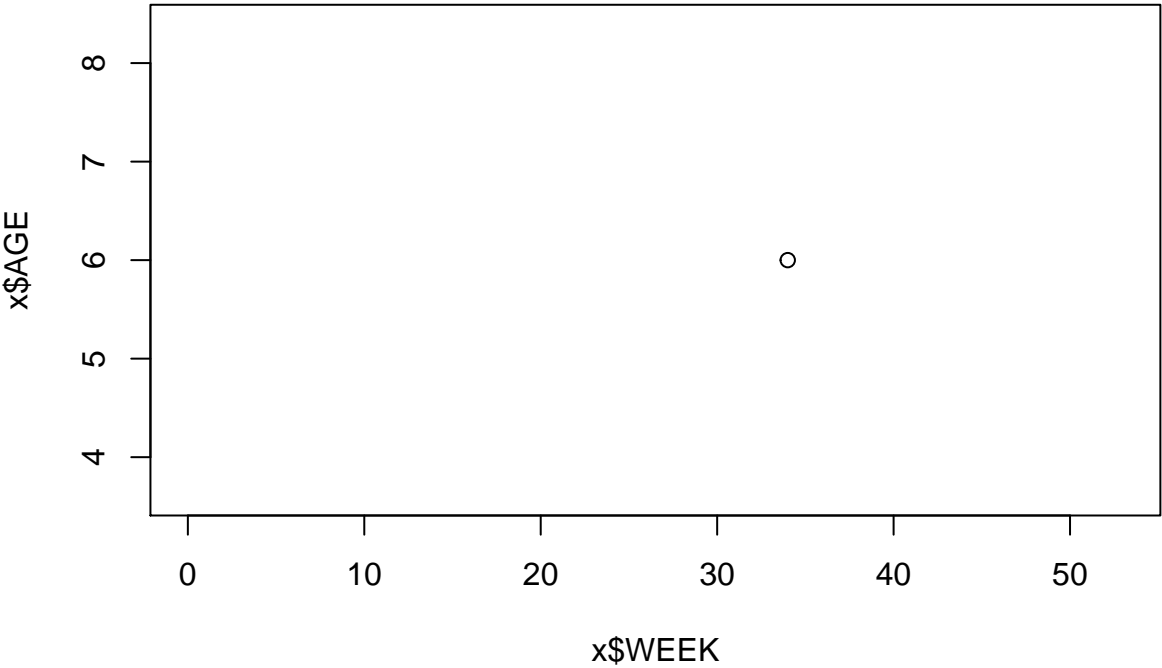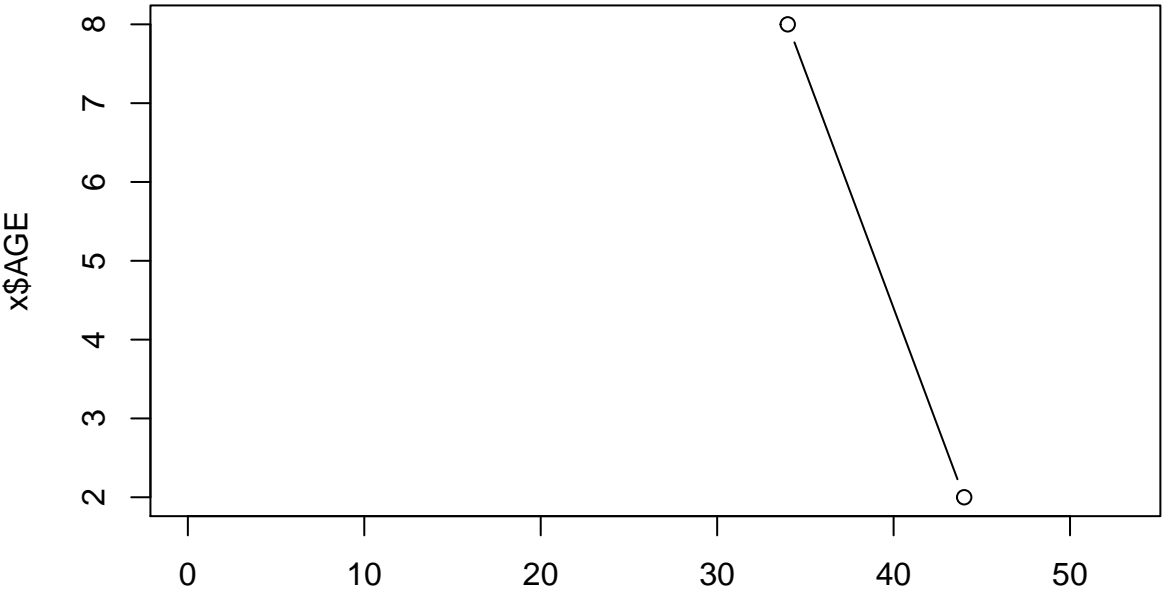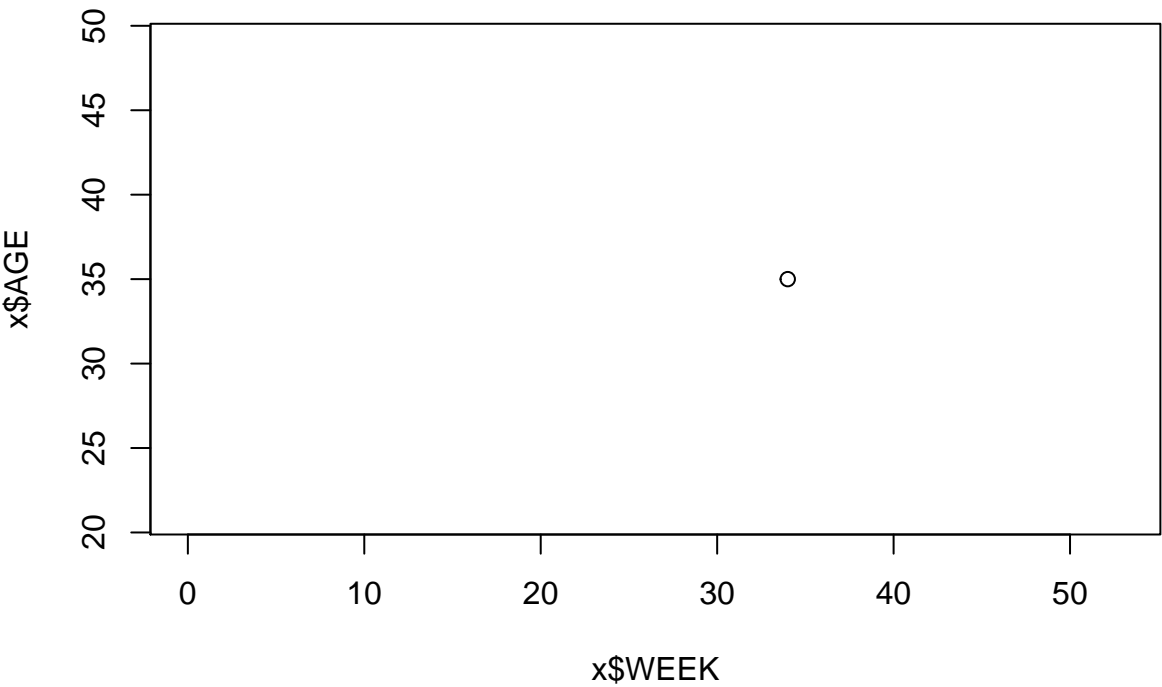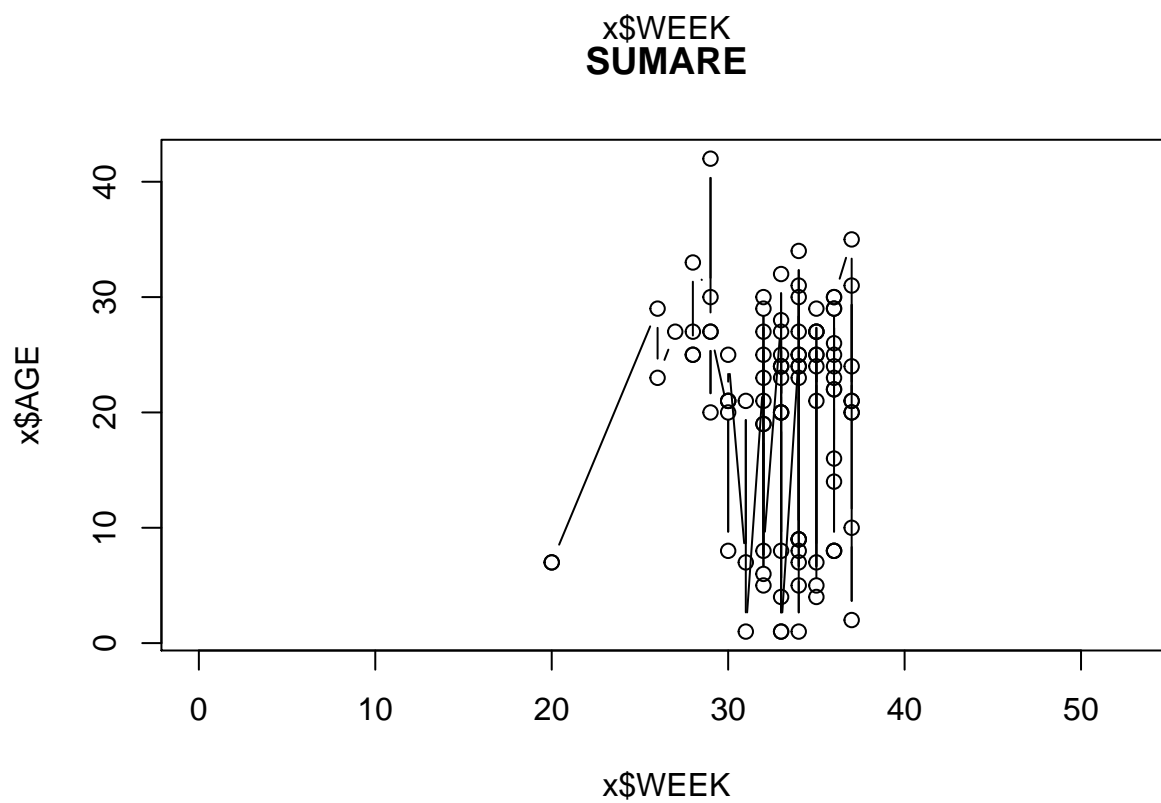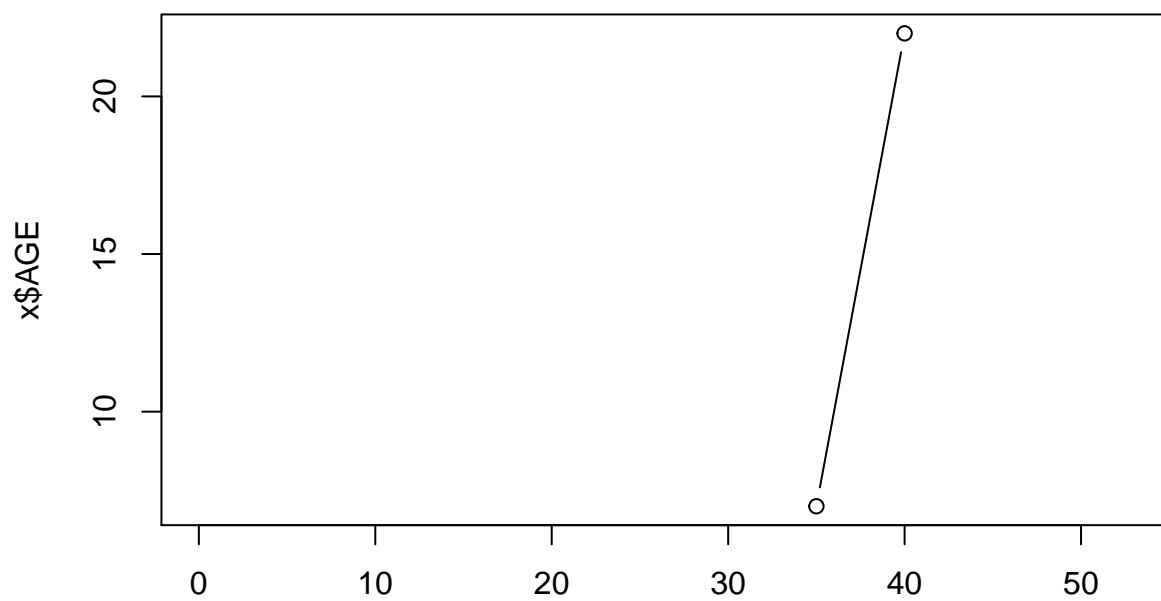
LUPERCIO
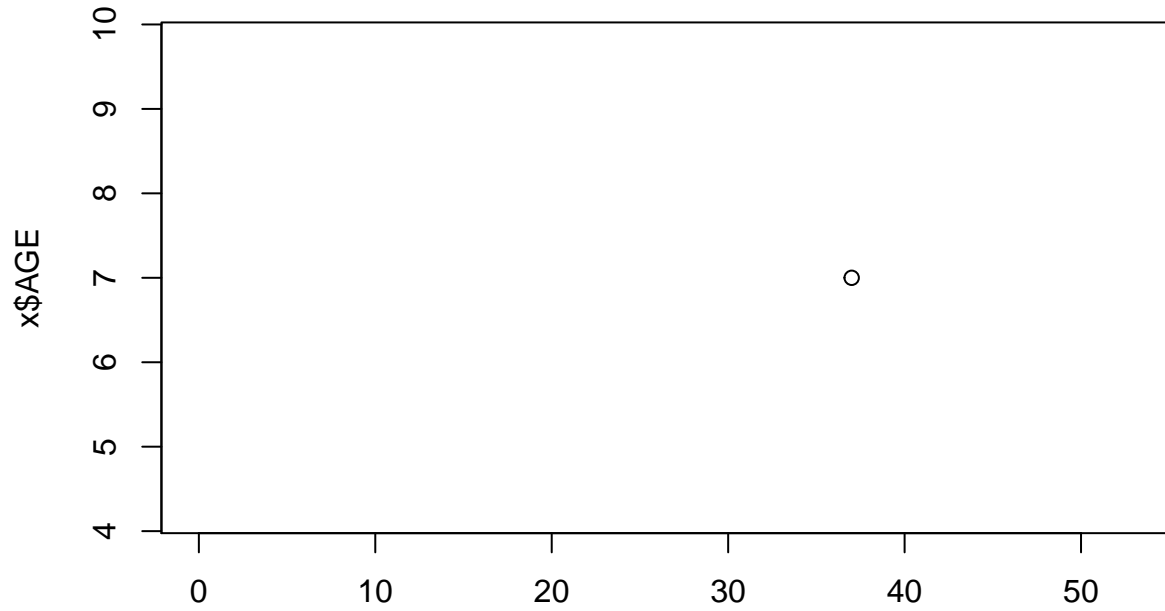
MONGAGUA
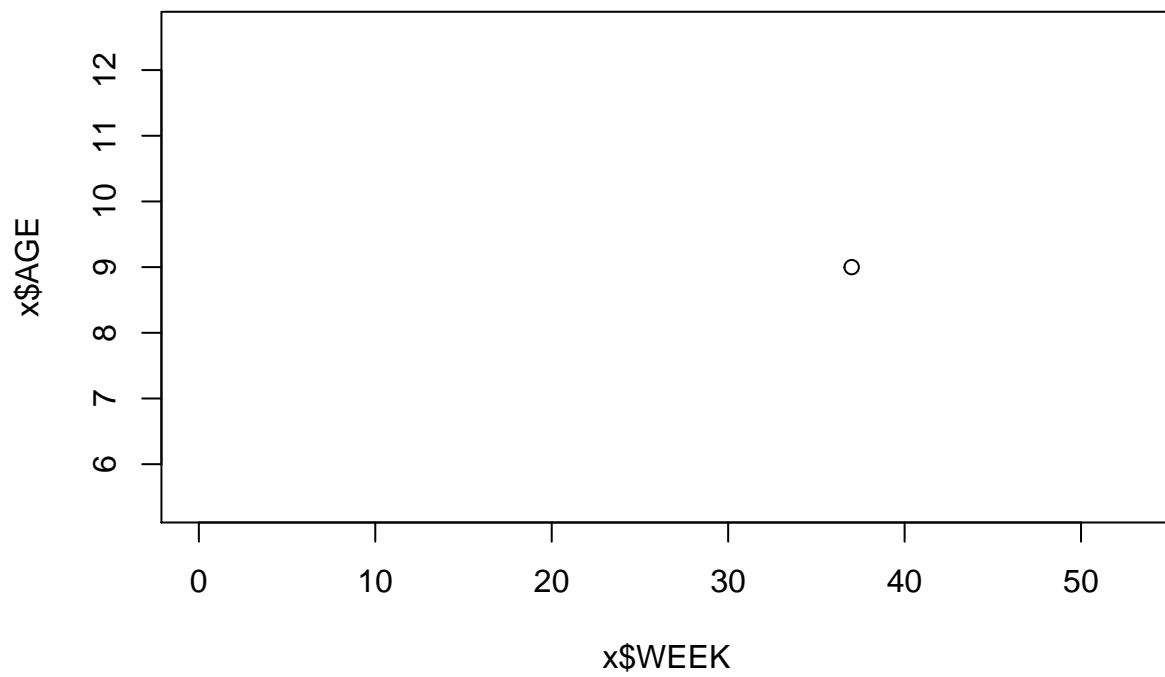
MORUNGABA
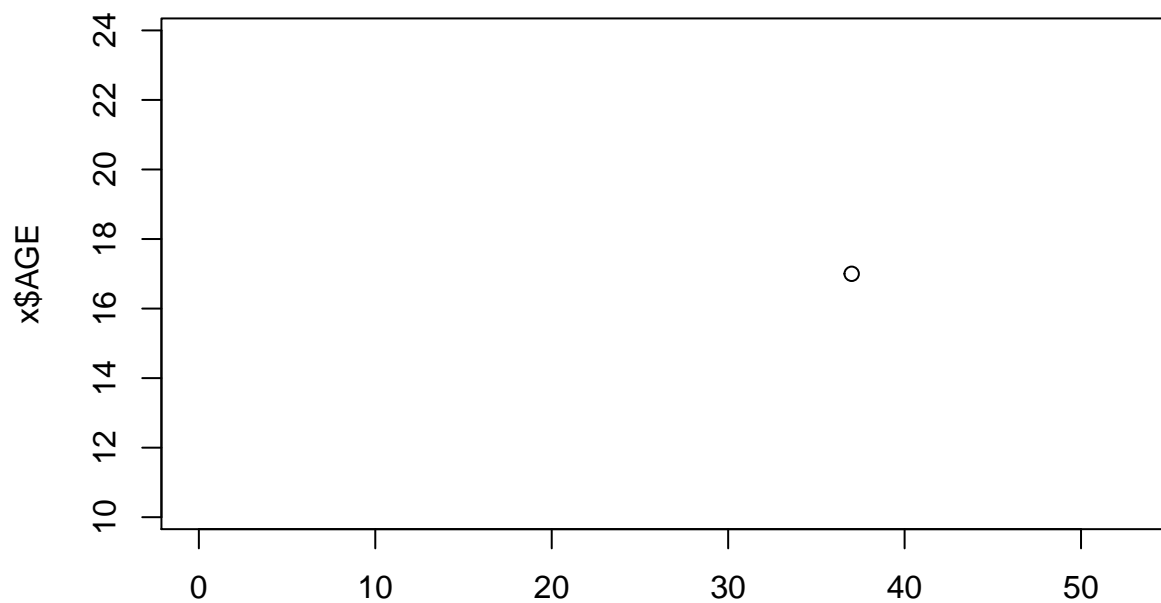
SUZANO
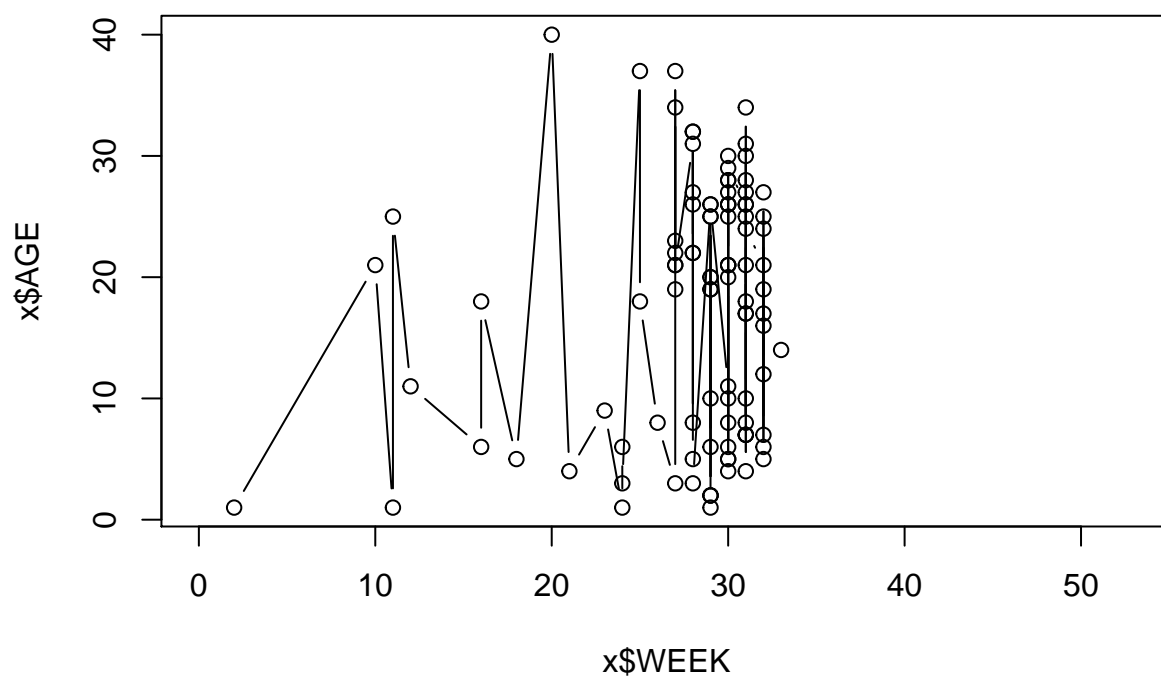
GUARANI D'OESTE
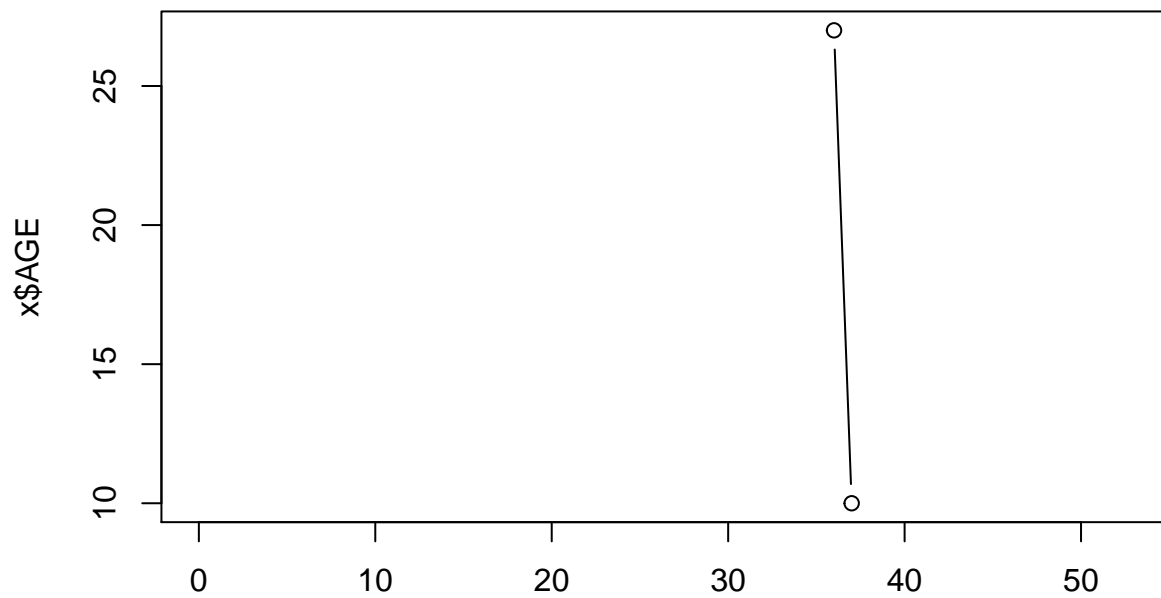
PRACINHA
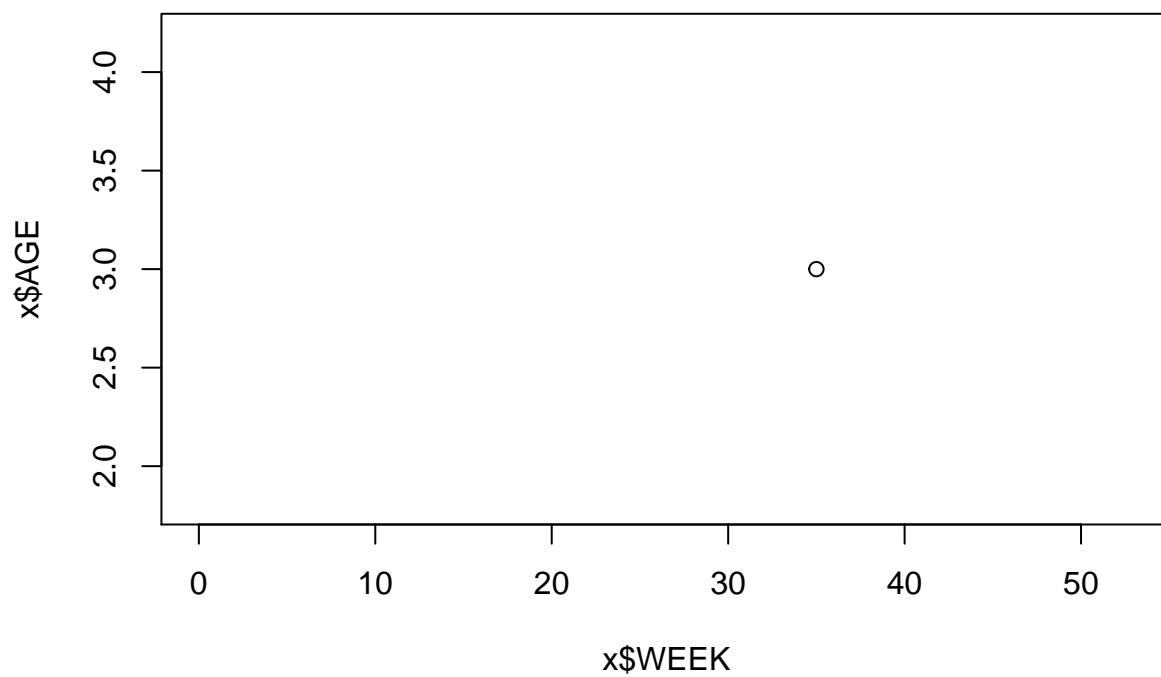
GALIA
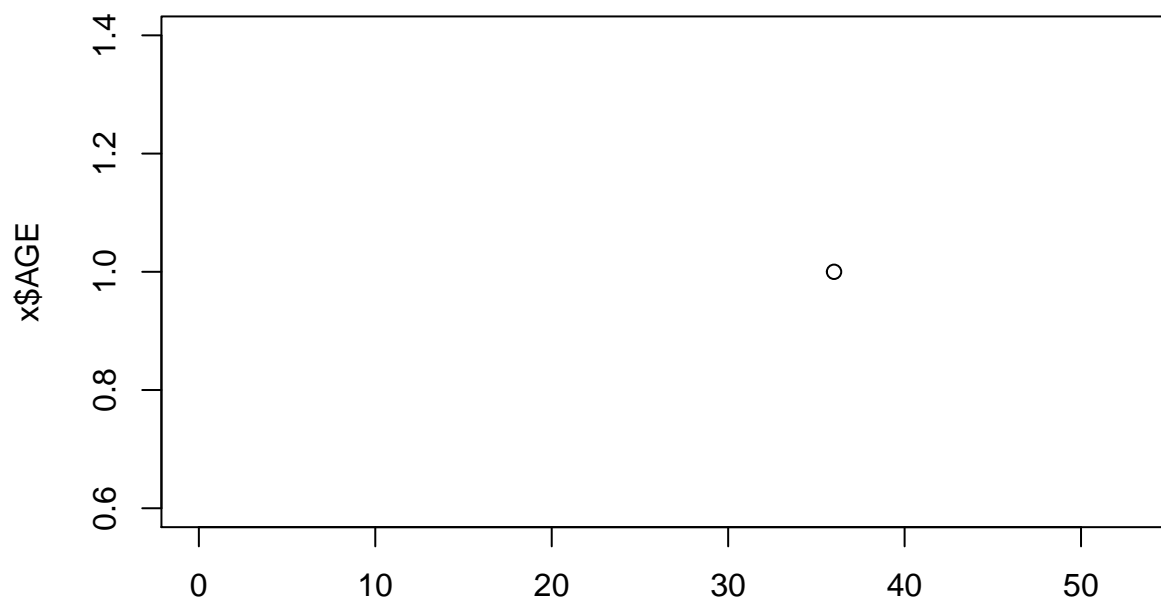
IPORANGA
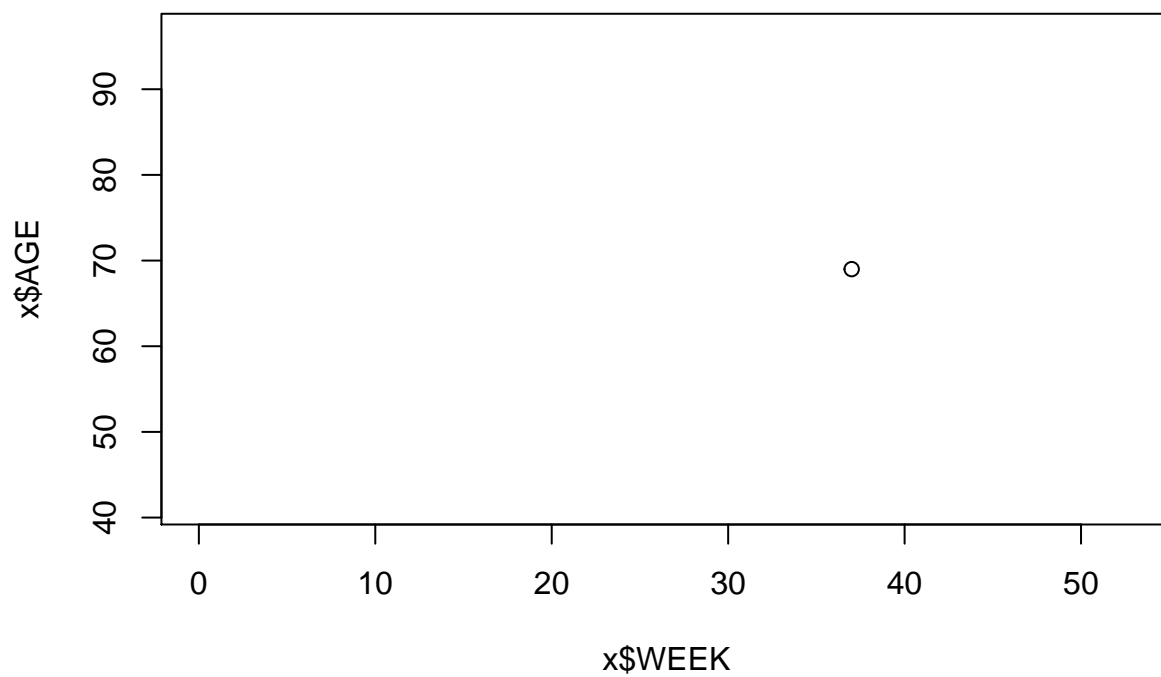
CAPELA DO ALTO
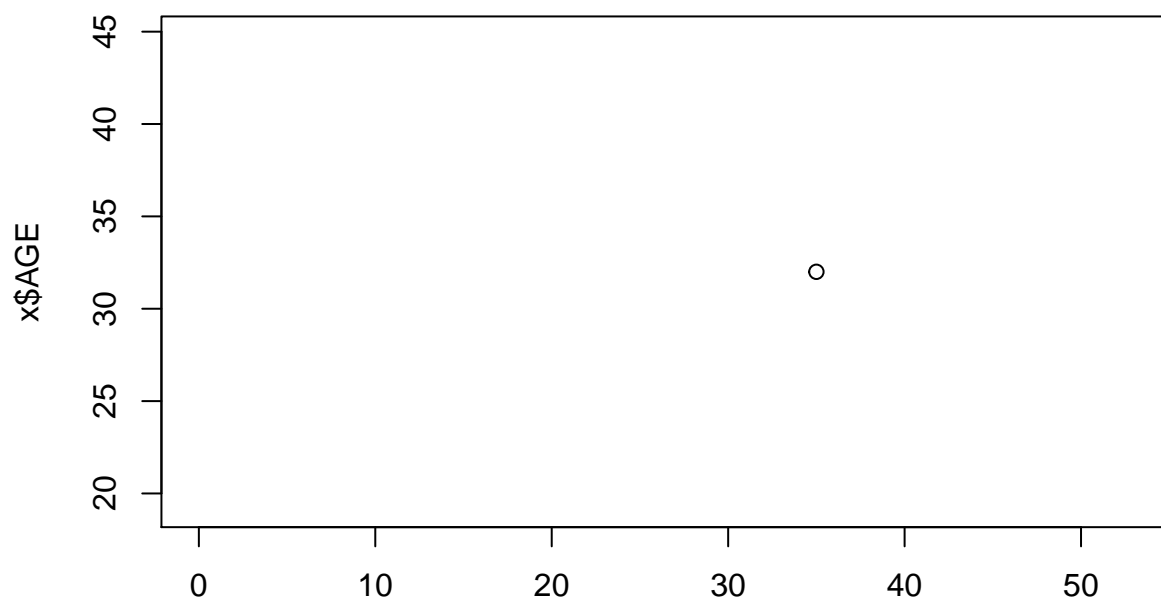
GARCA
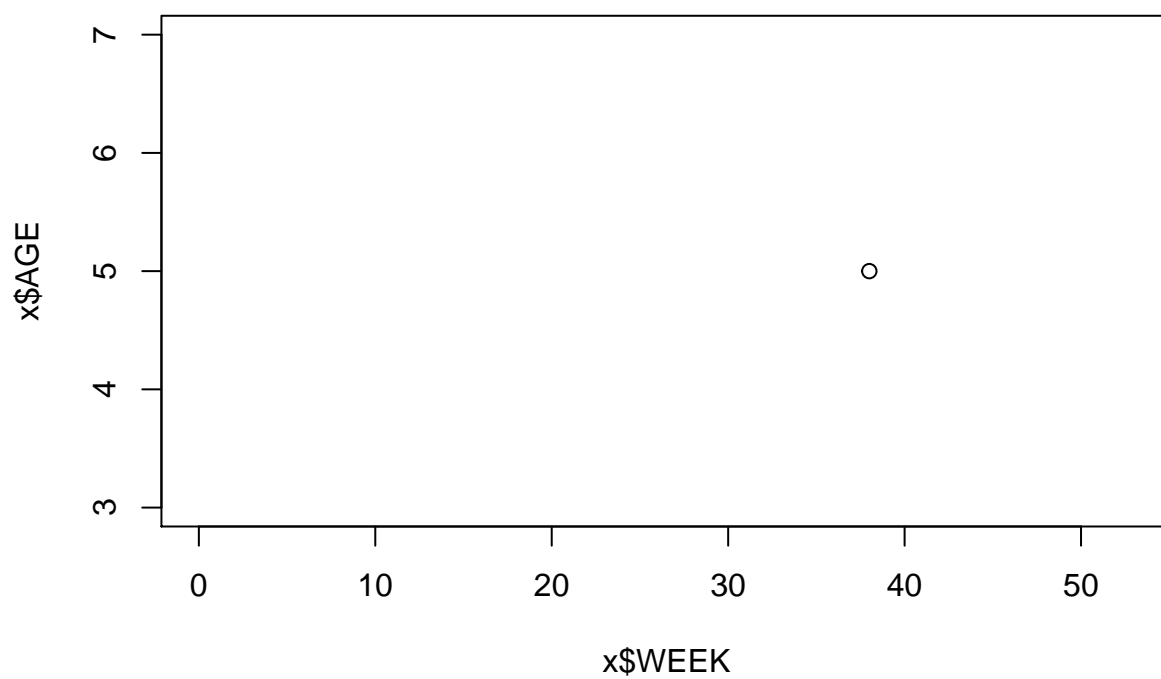
MERIDIANO
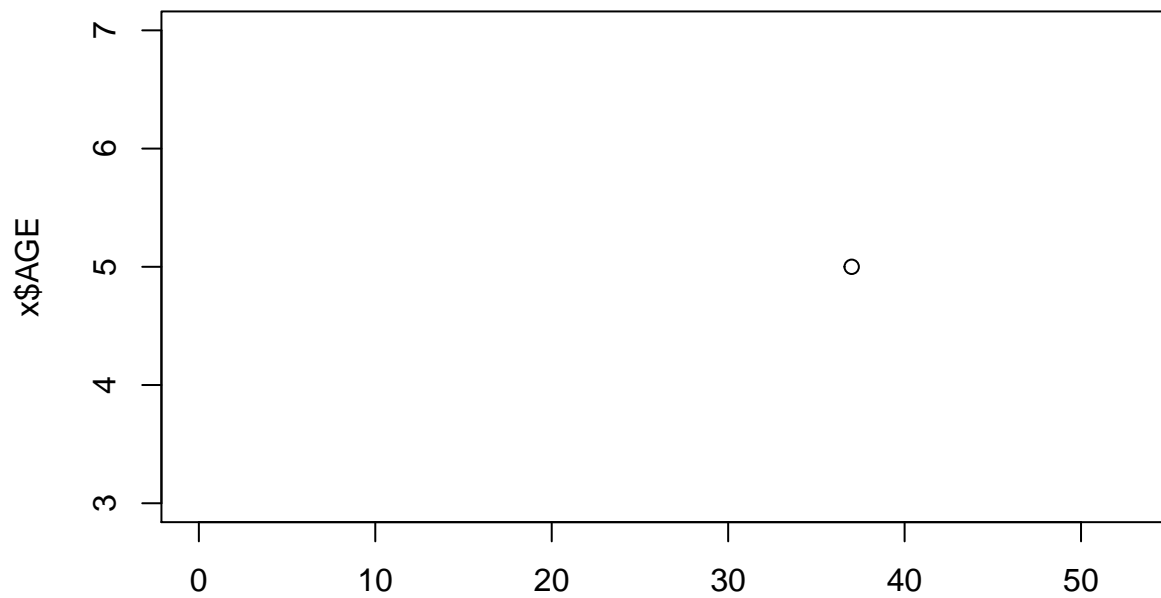
ILHA SOLTEIRA
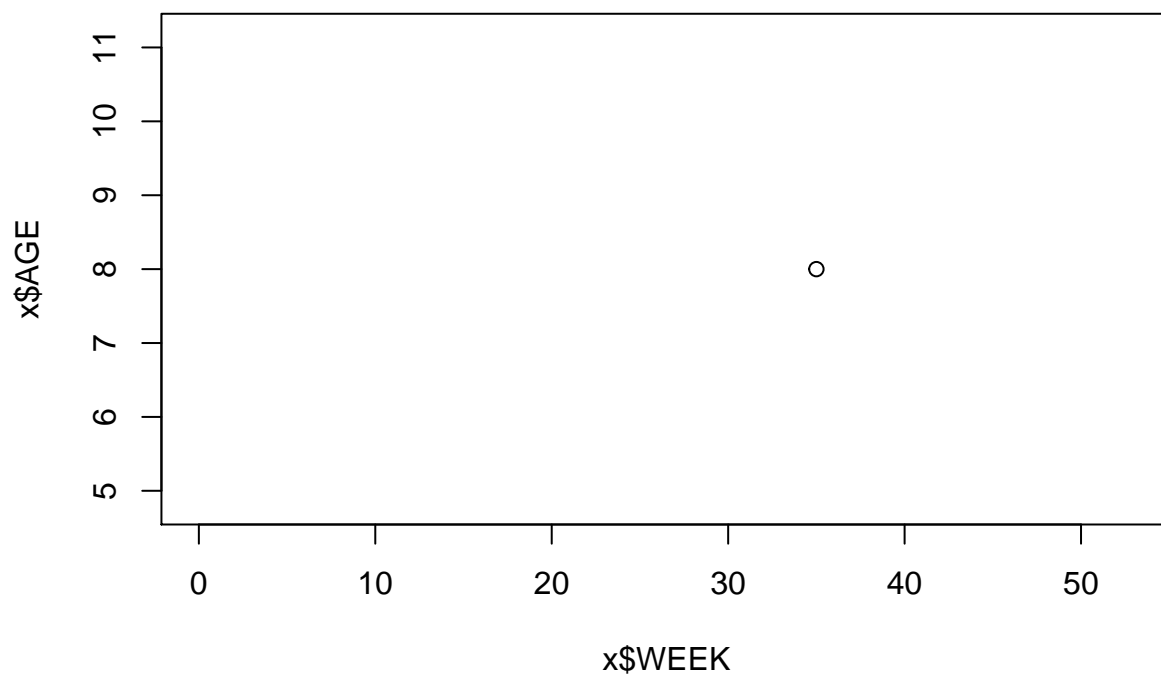
TAMBAU
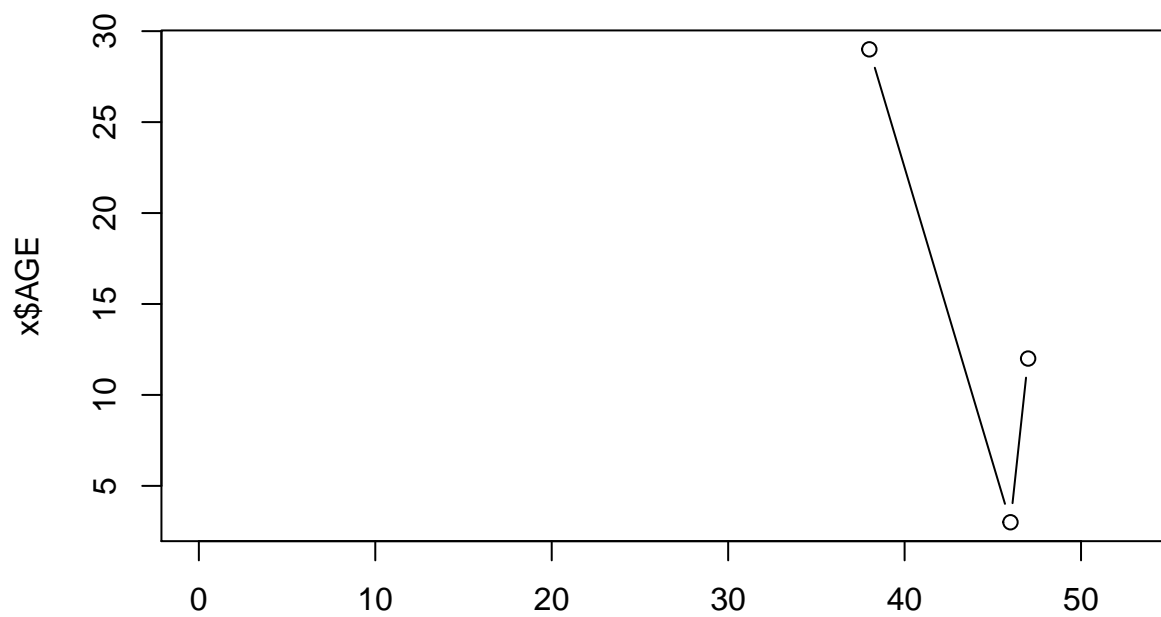
TANABI
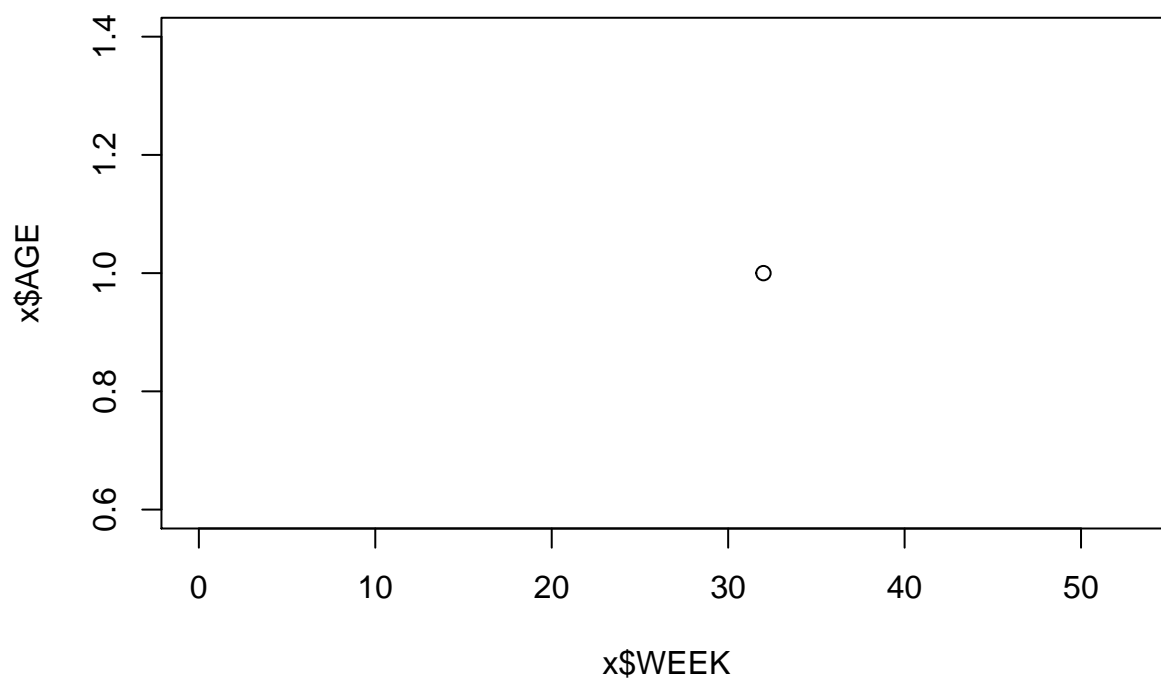
TAPIRATIBA
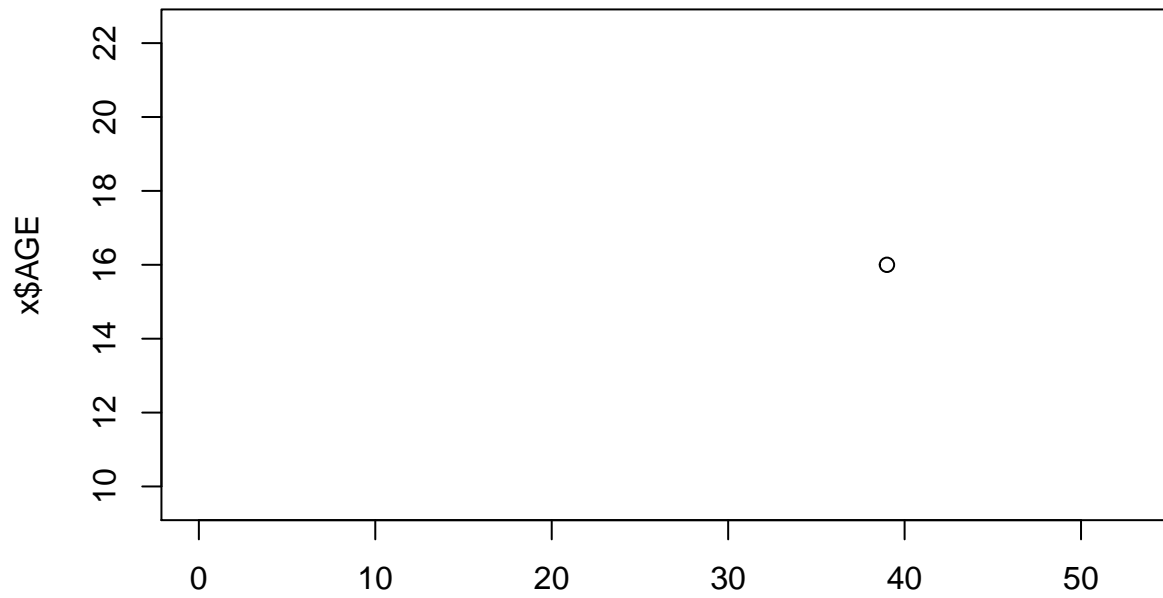
TAQUARAL
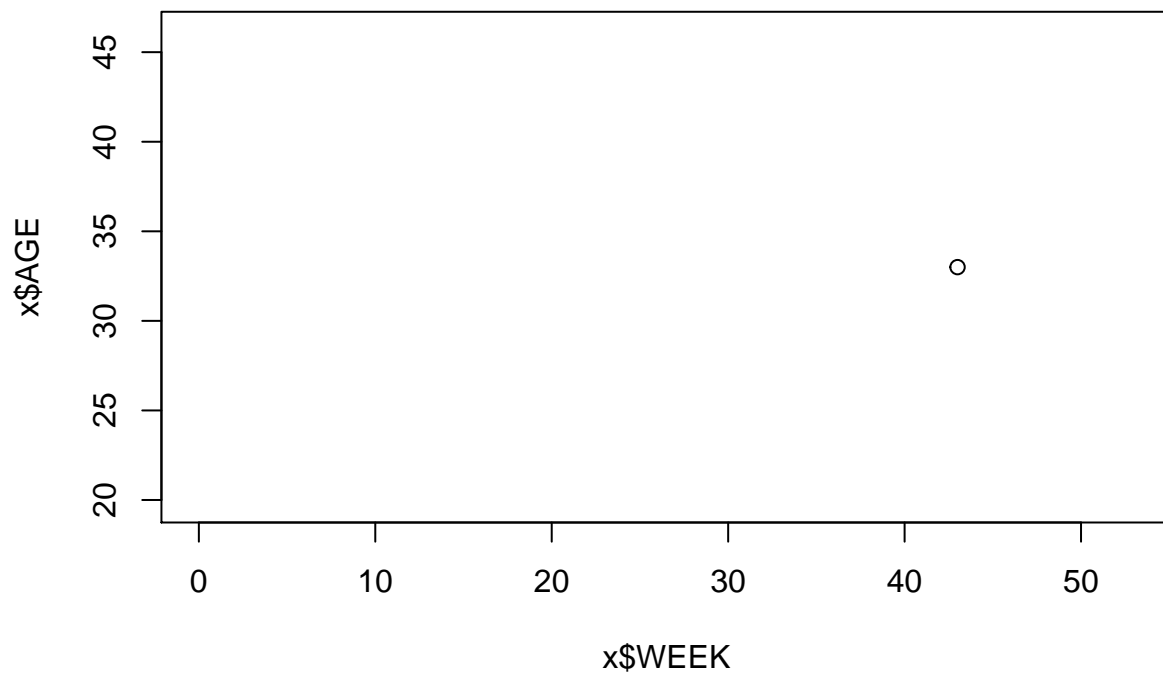
TAQUARITINGA
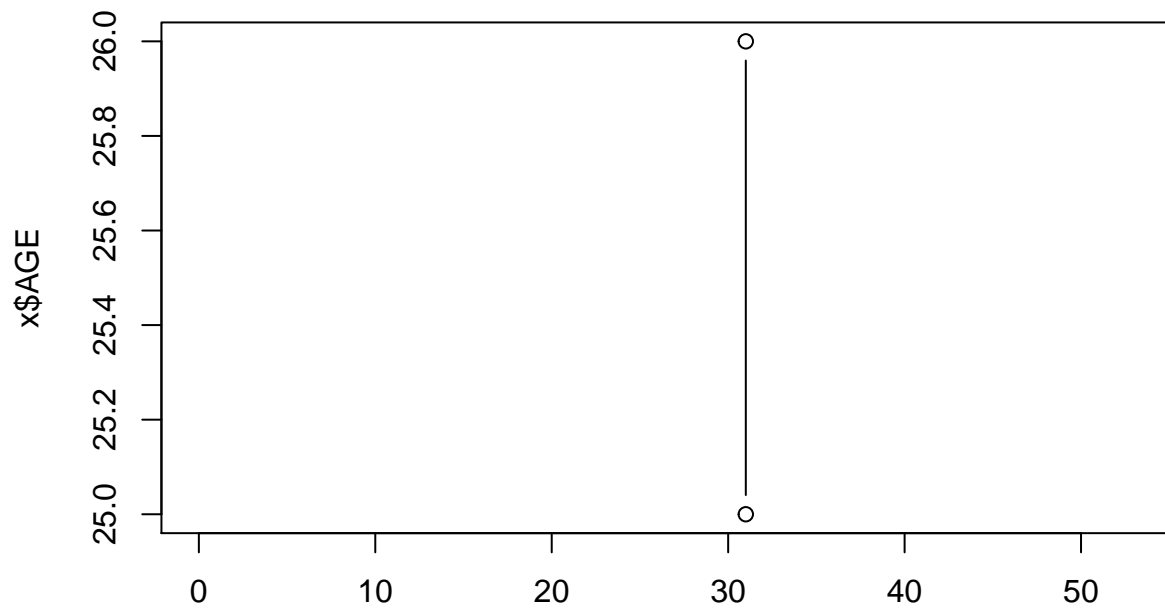
TAQUARITUBA
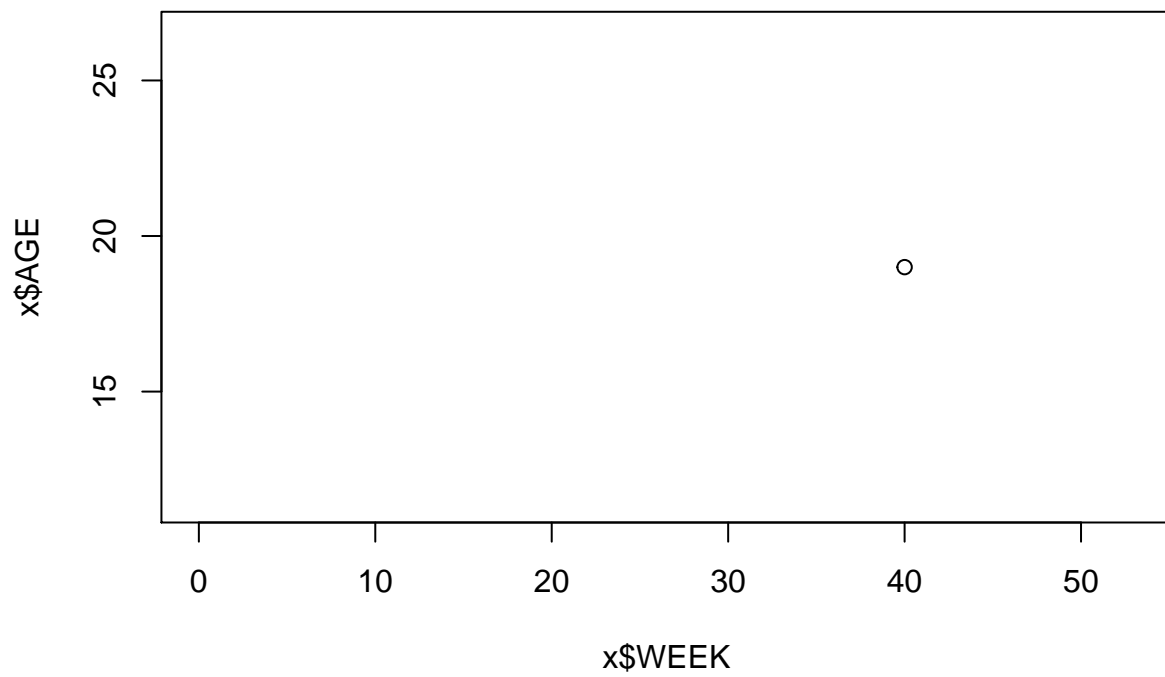
TARABAI
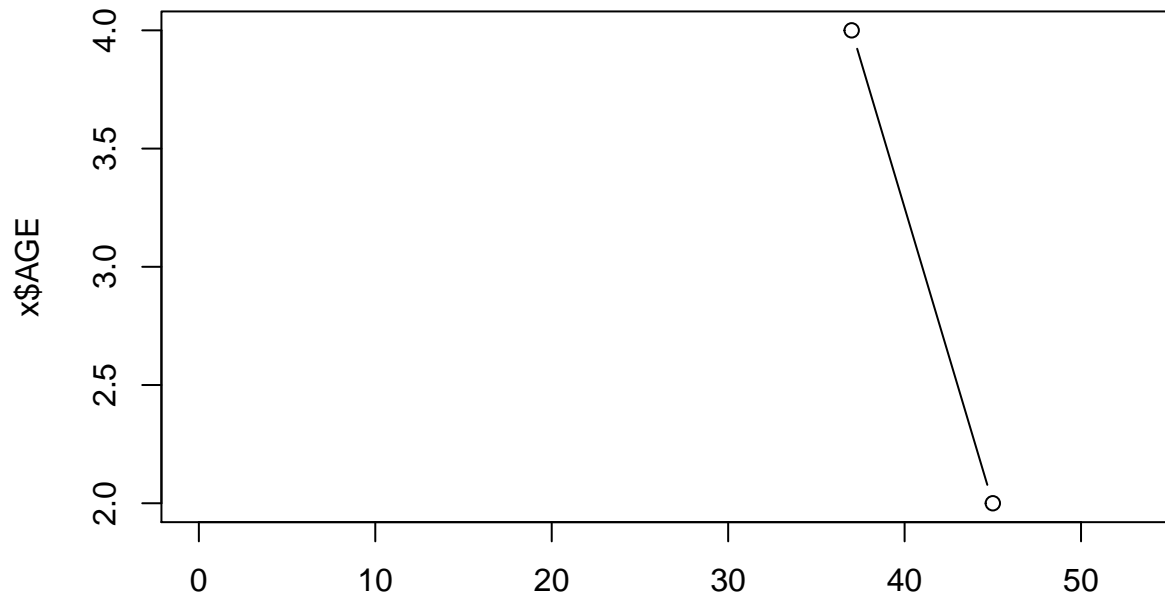
TATUI
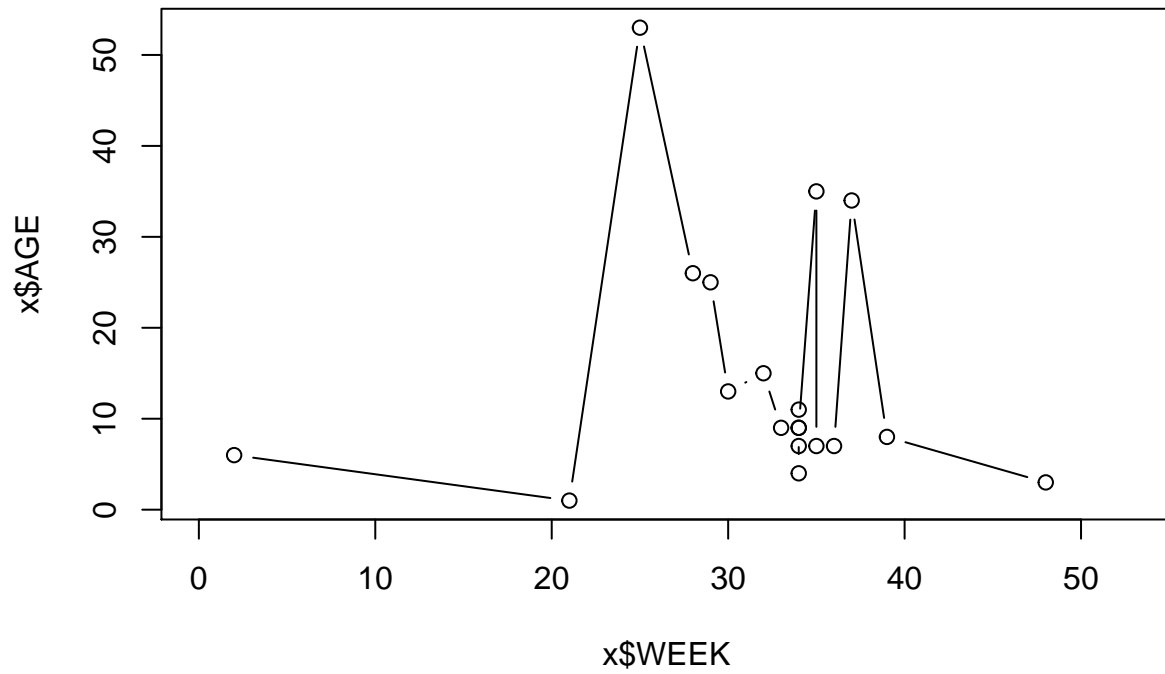
OLIMPIA
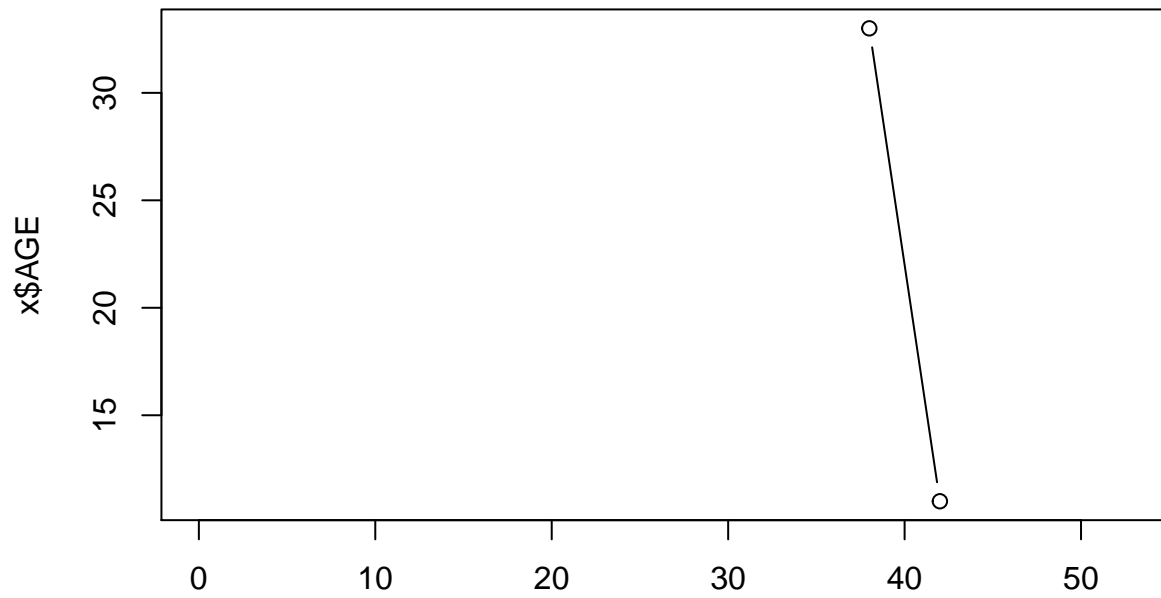
TAUBATE

TEODORO SAMPAIO

x$WEEK

TERRA ROXA

x$WEEK
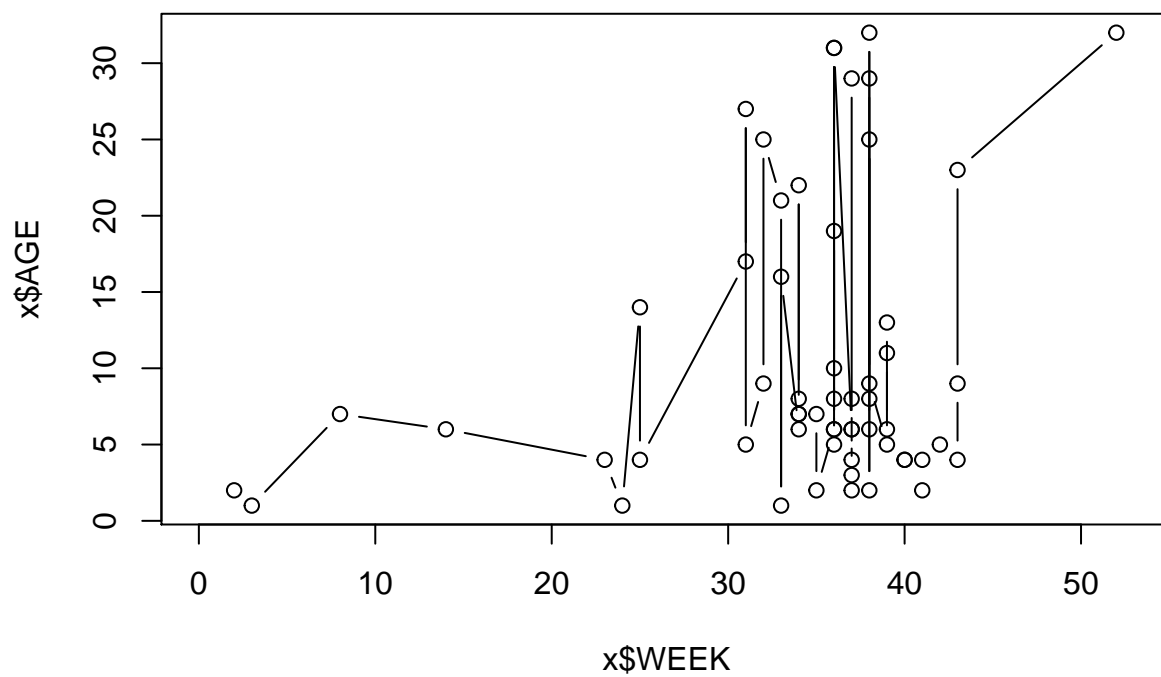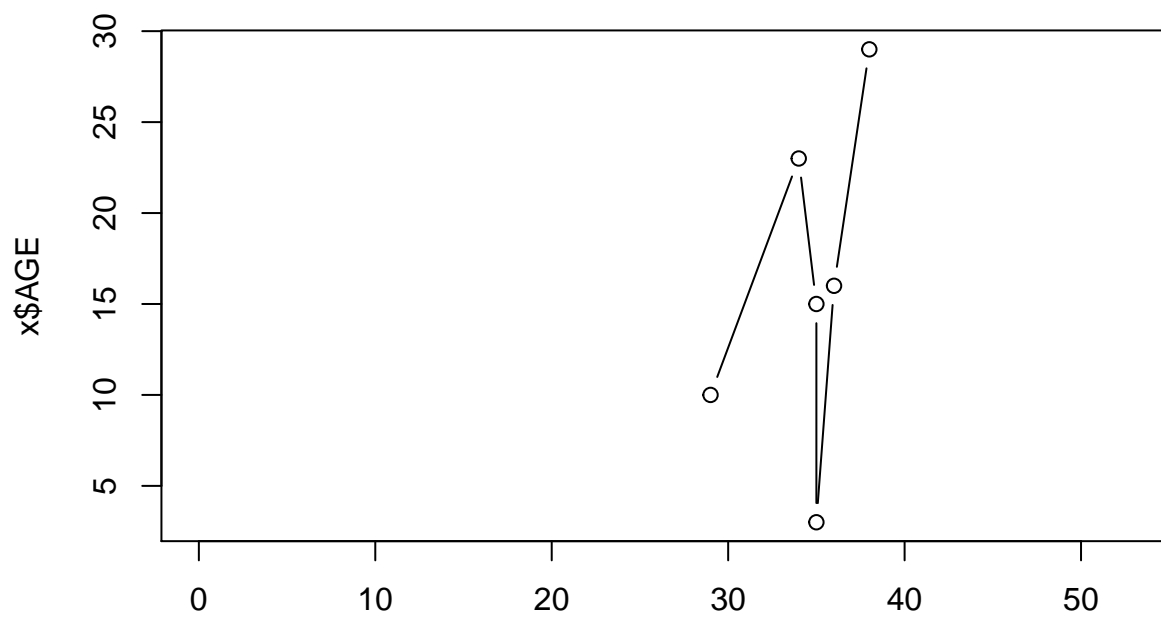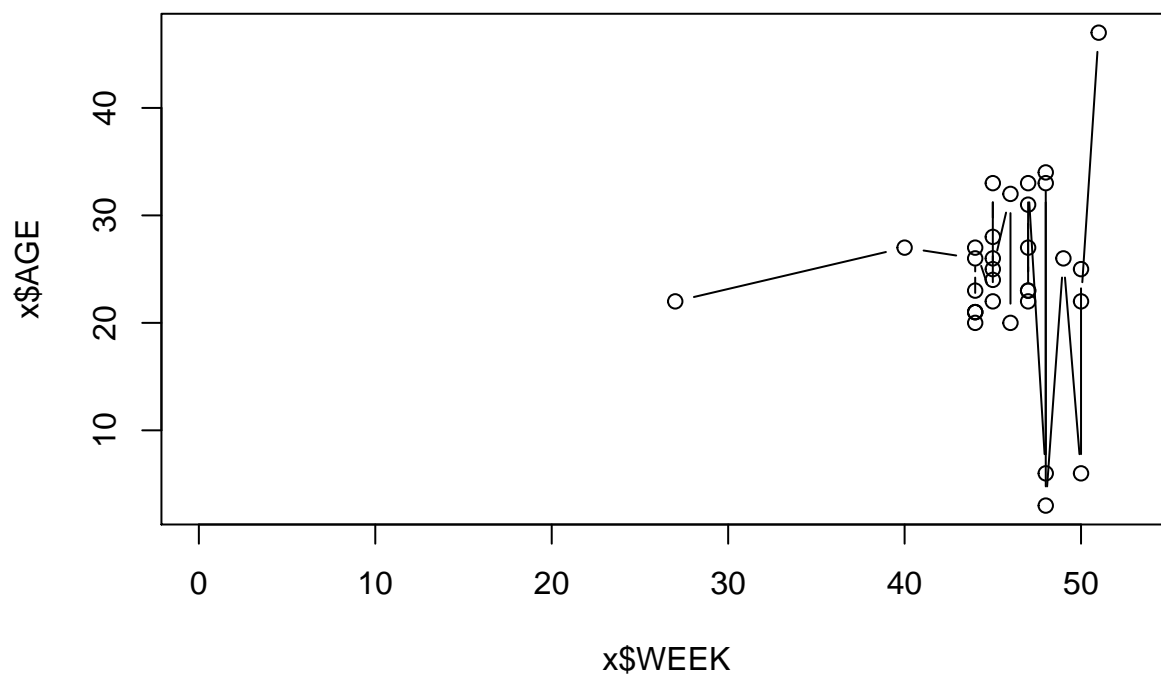
TIETE
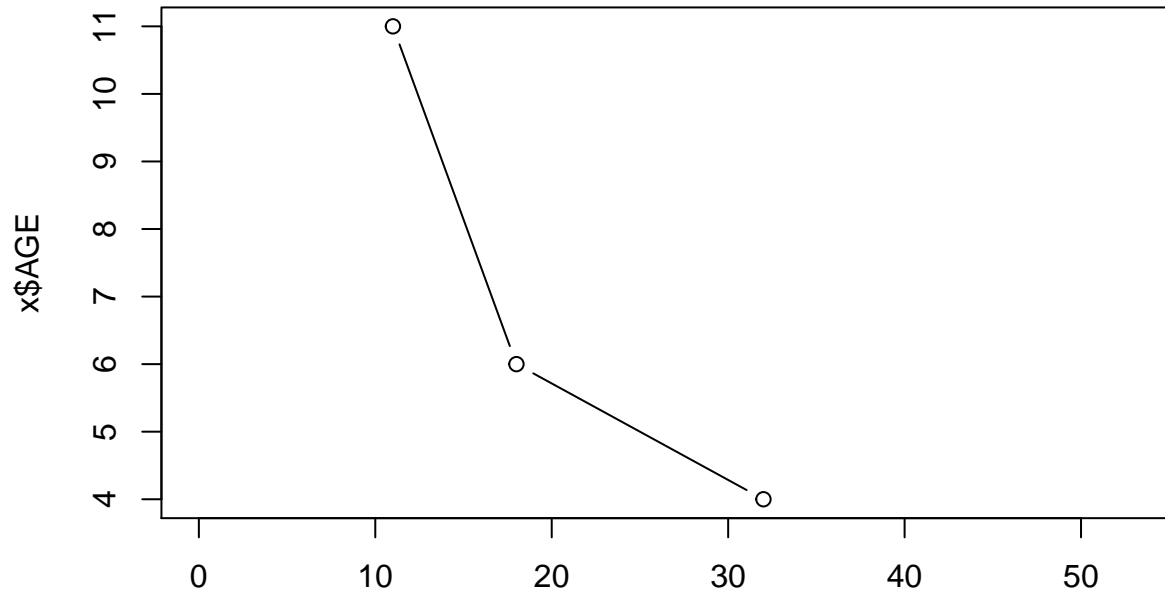
TIMBURI
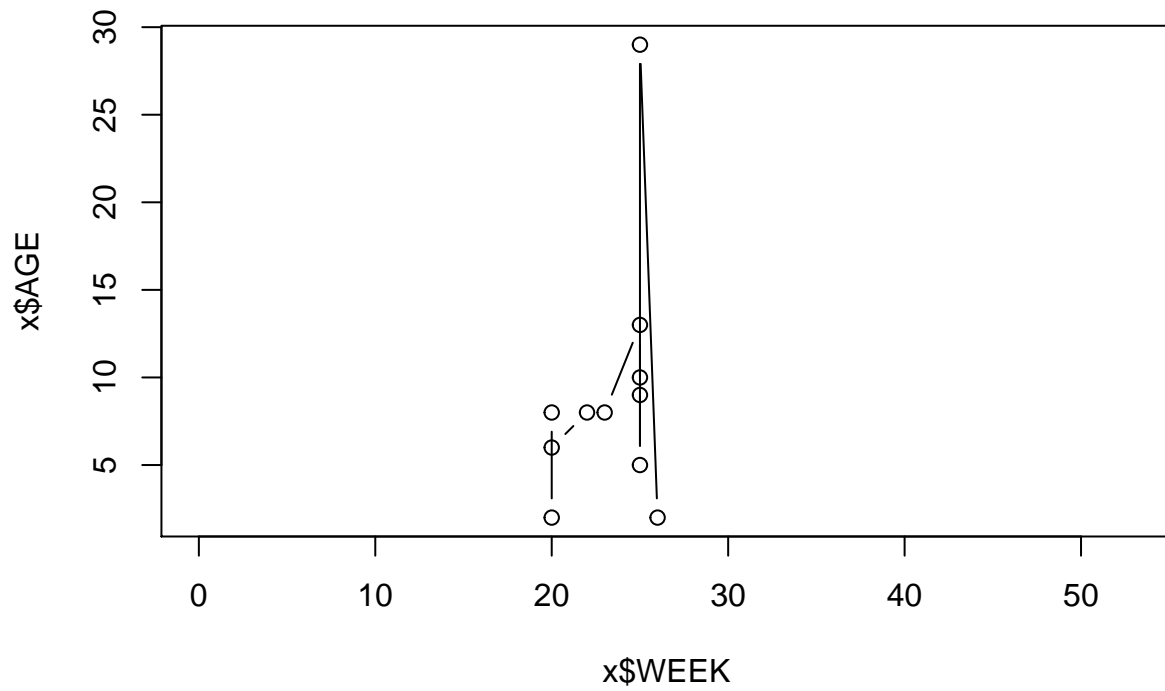
TORRINHA
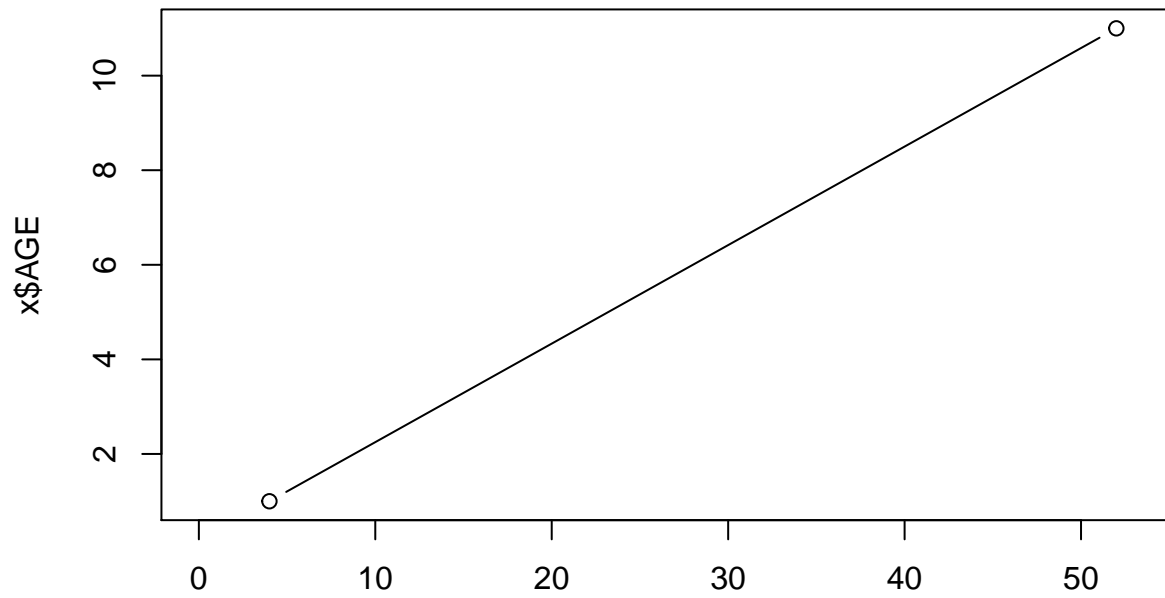
TREMEMBE
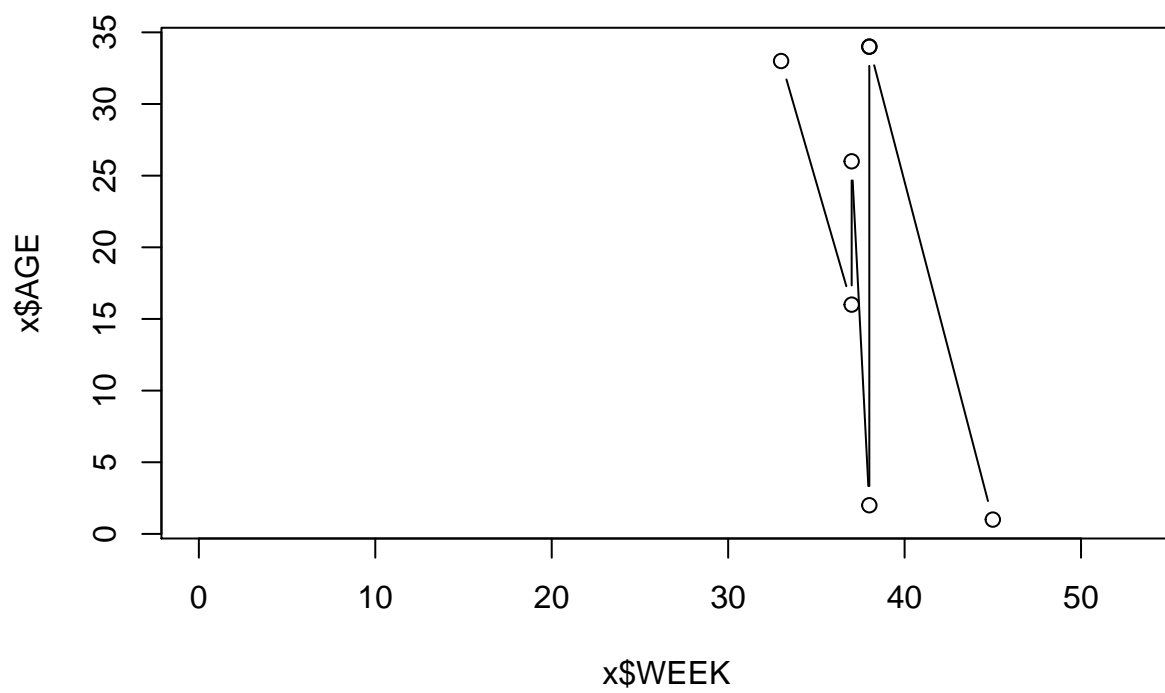
TUIUTI
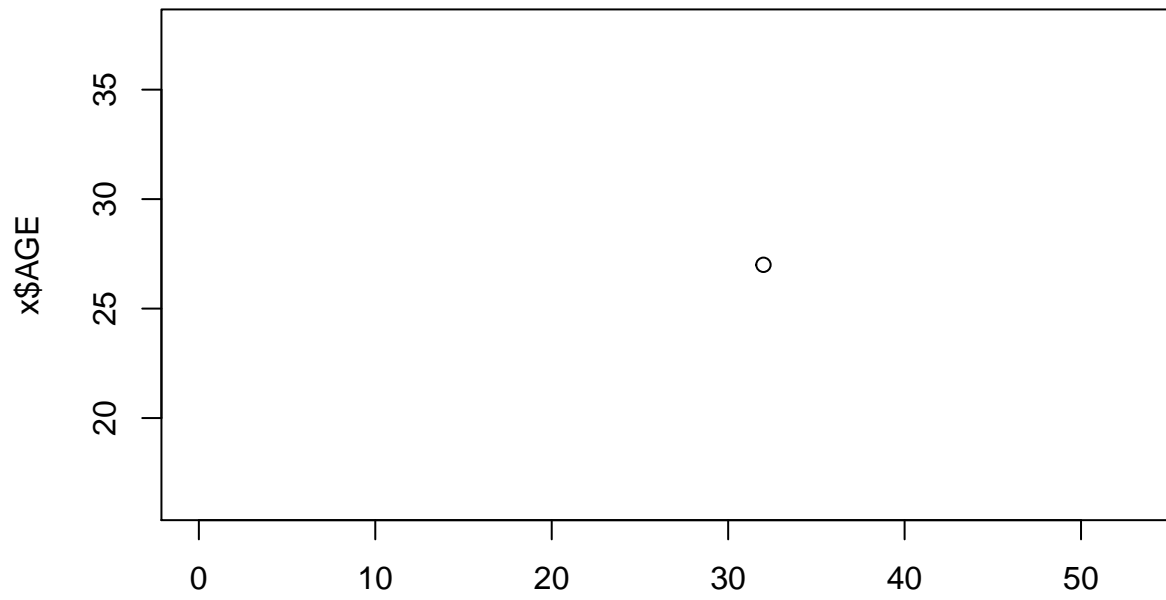
TUPA
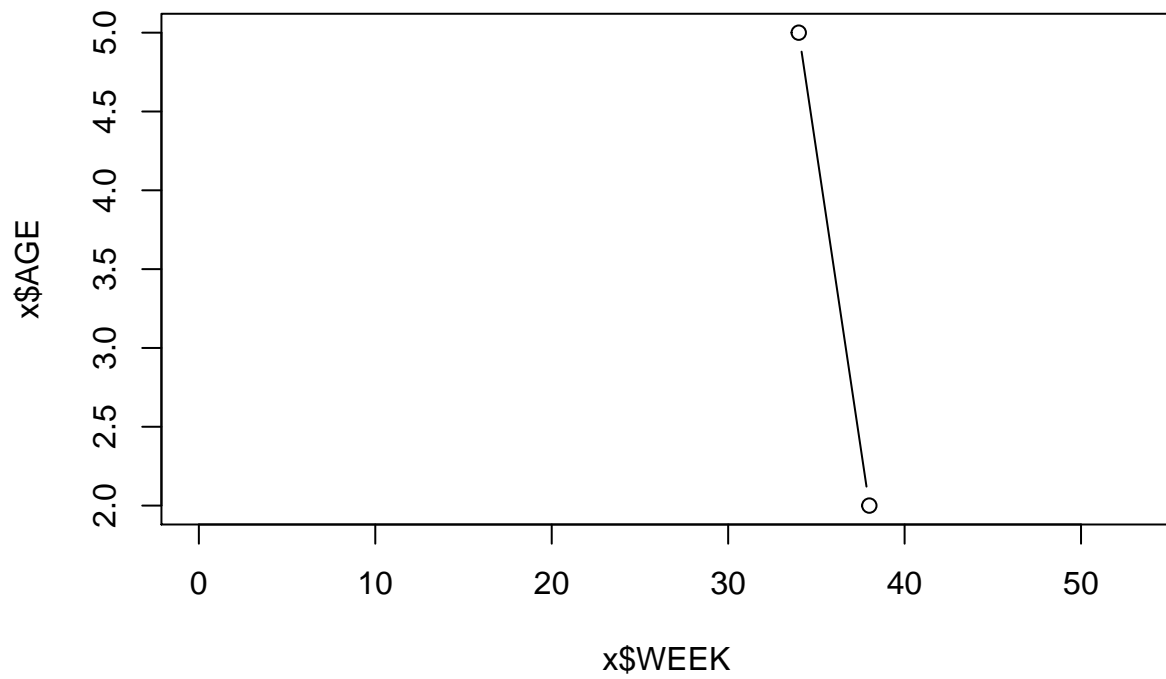
TUPI PAULISTA
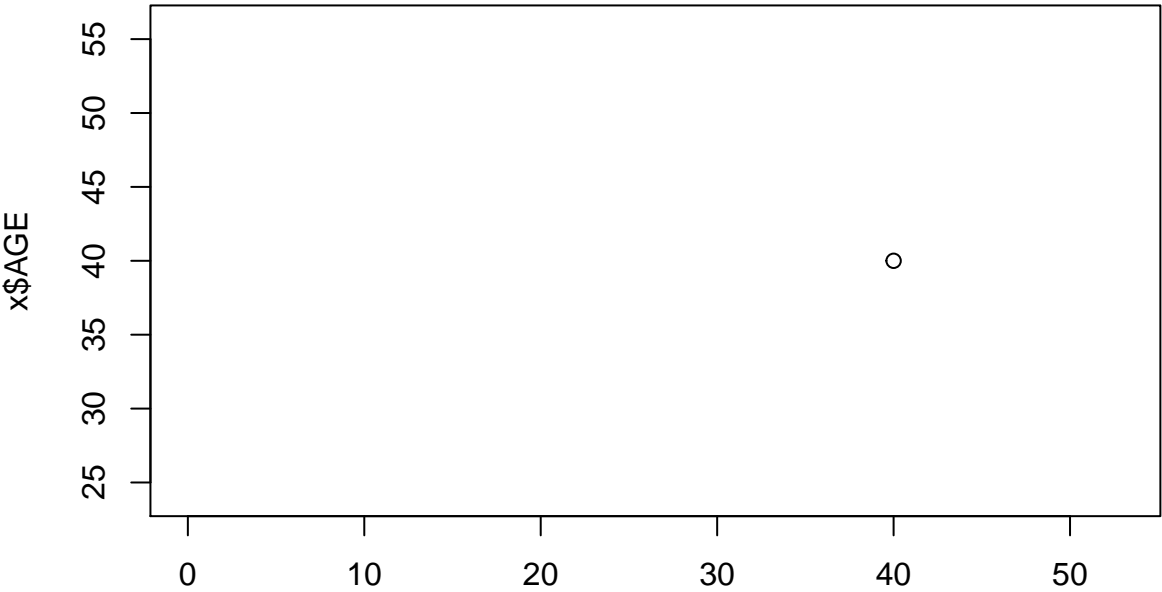
UBATUBA
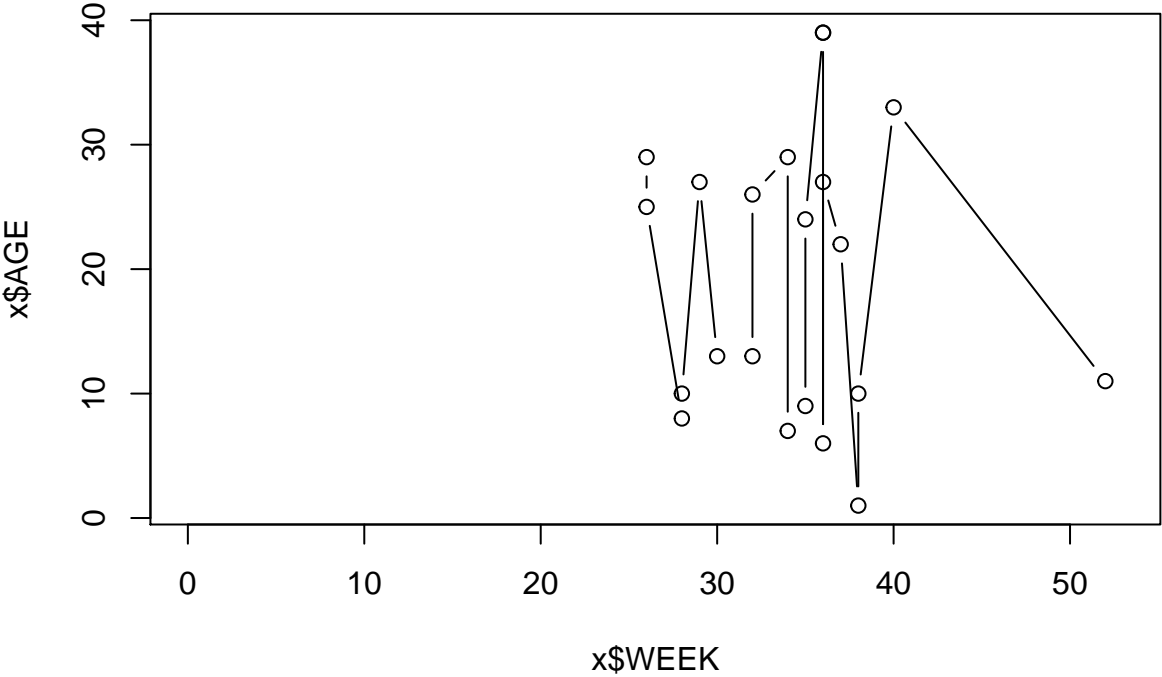
UBIRAJARA
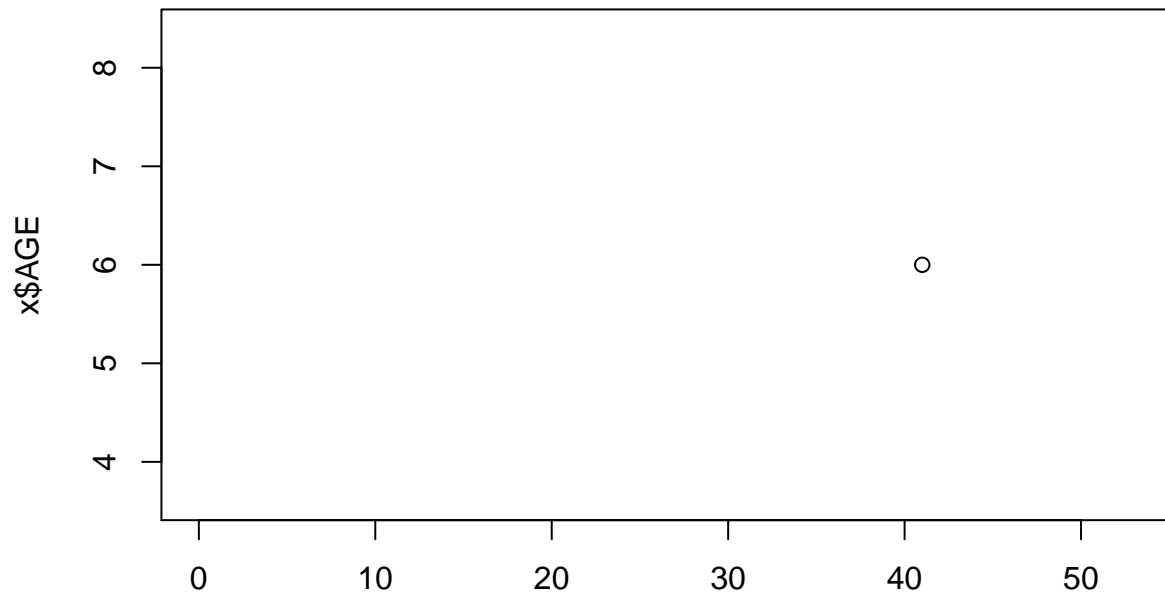
VALINHOS
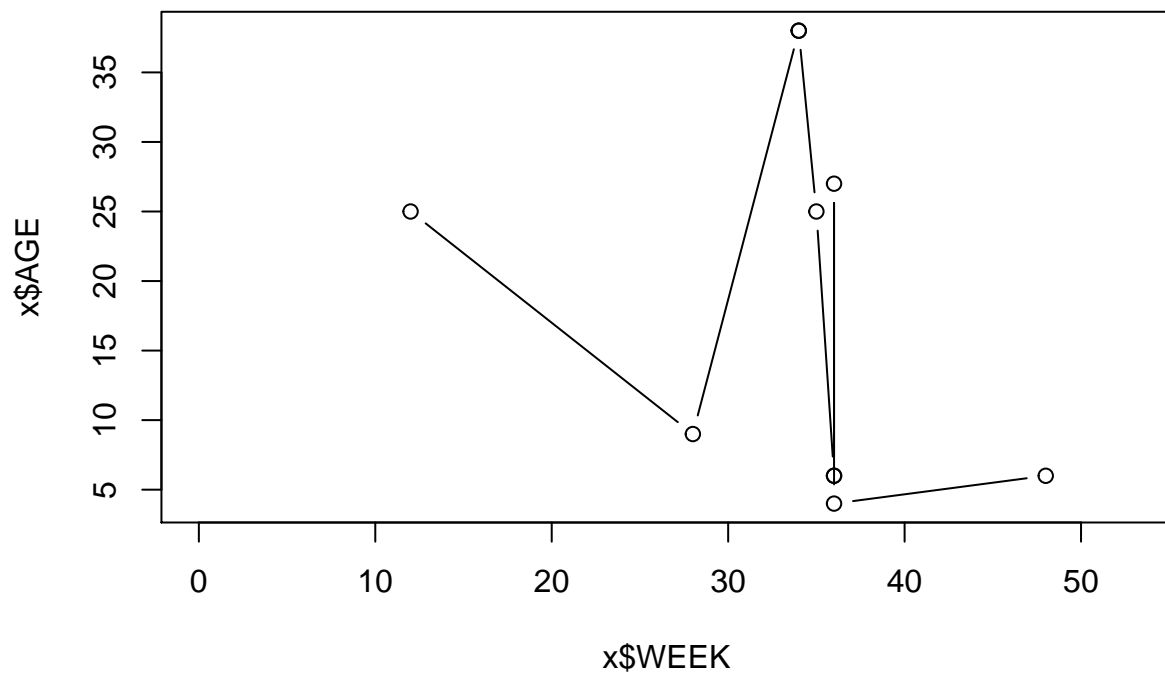
VALPARAISO
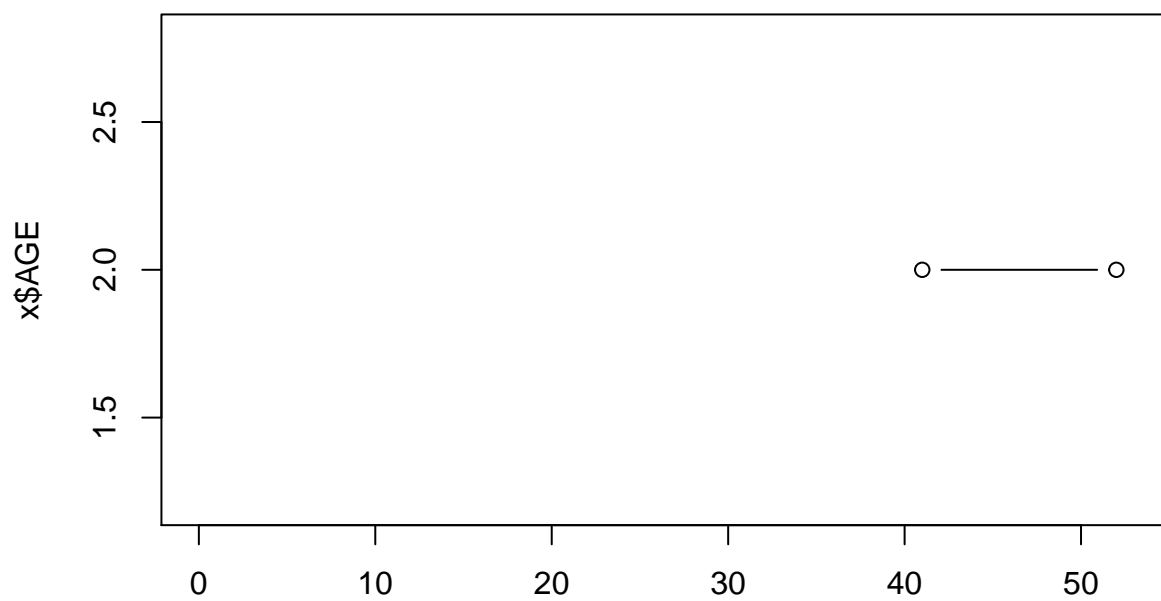
VARGEM GRANDE PAULISTA
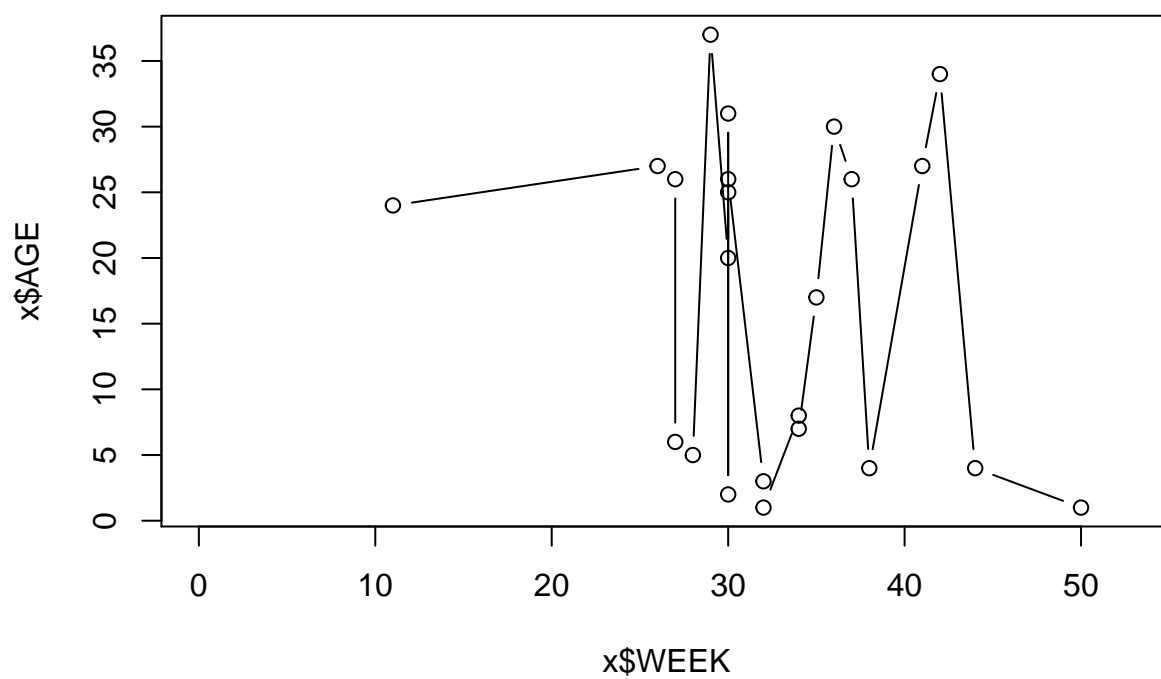
VARZEA PAULISTA
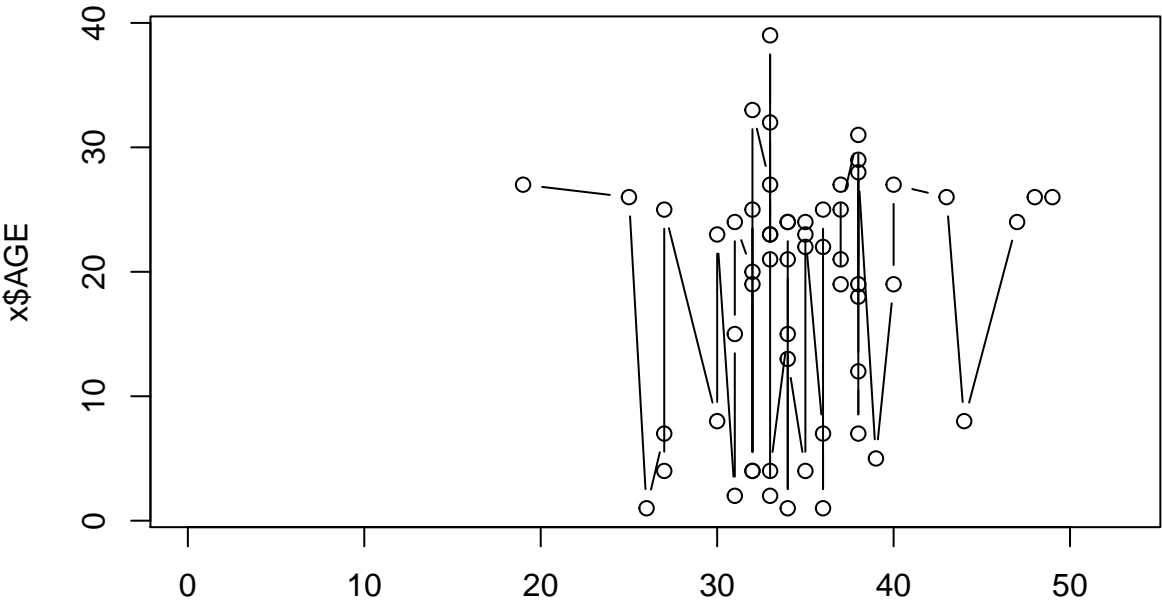
VERA CRUZ
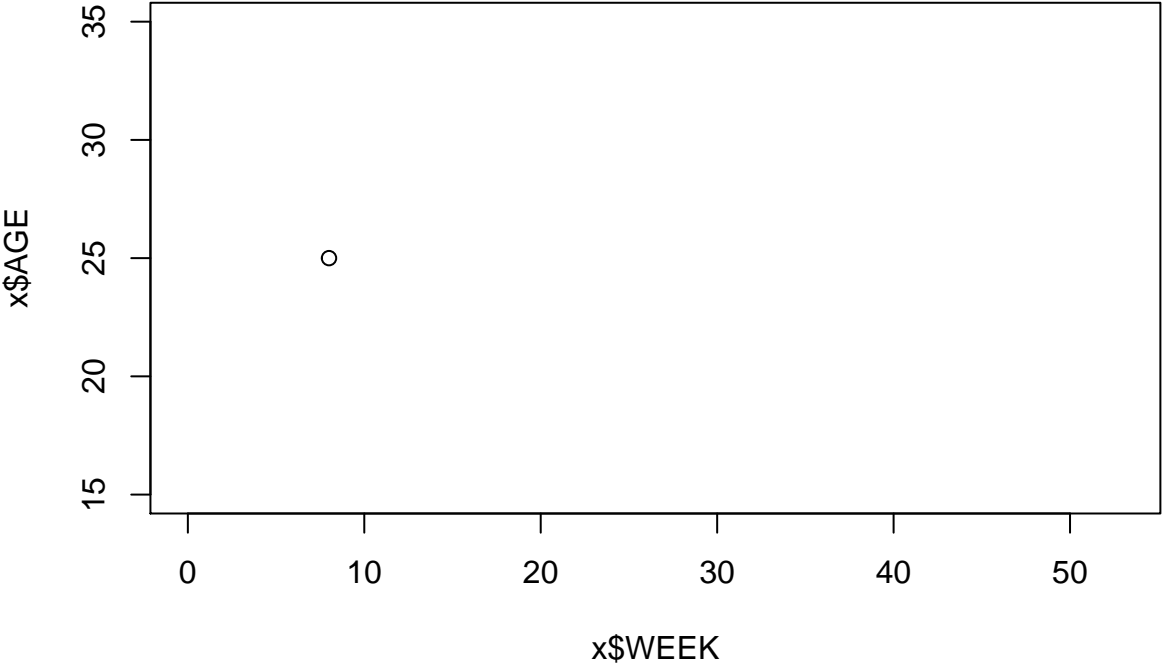
VINHEDO
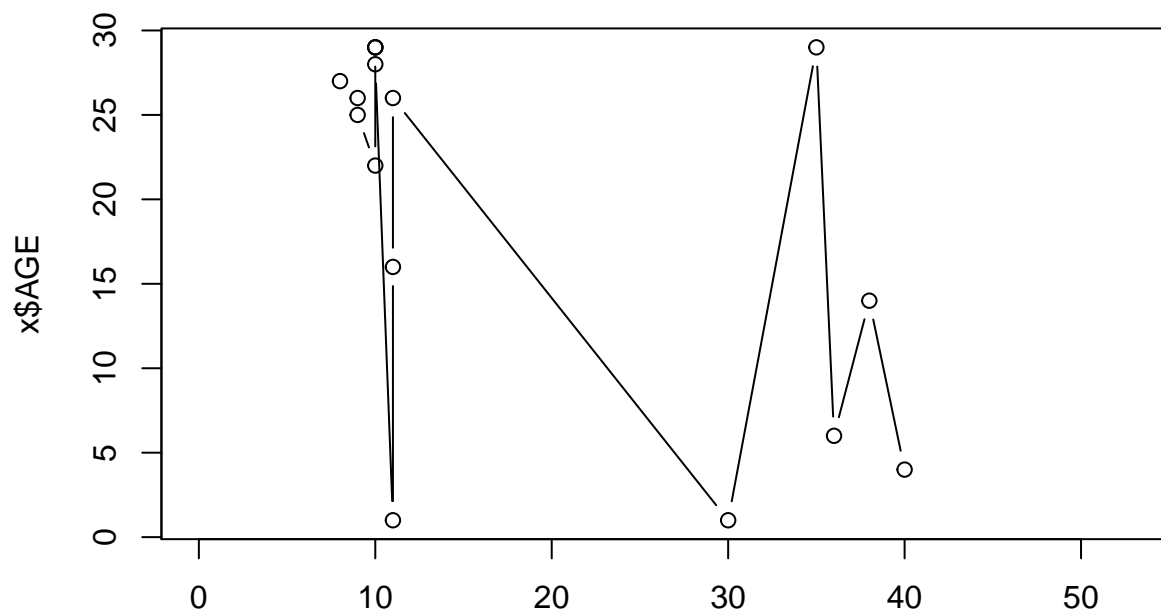
VIRADOURO
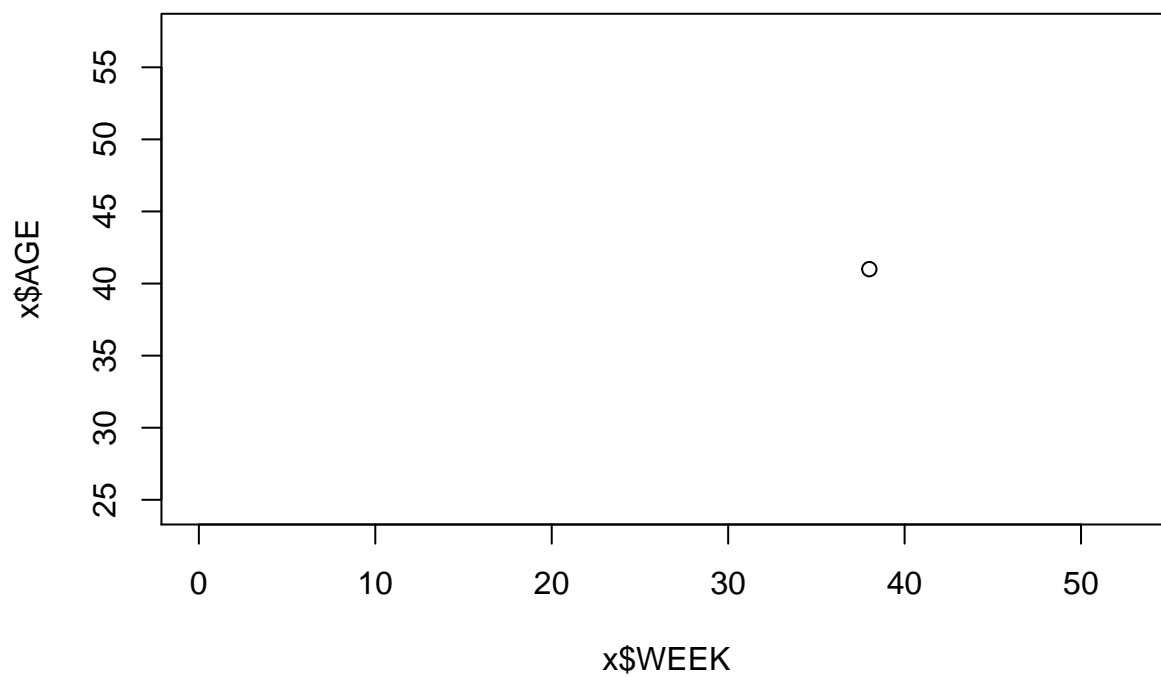
VITORIA BRASIL
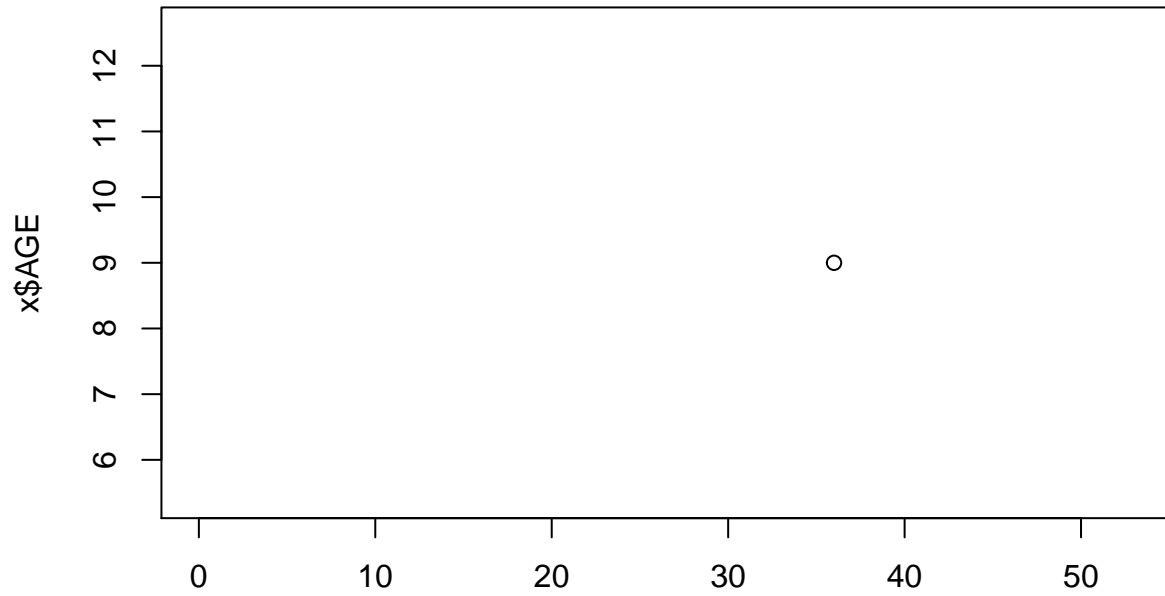
VOTORANTIM
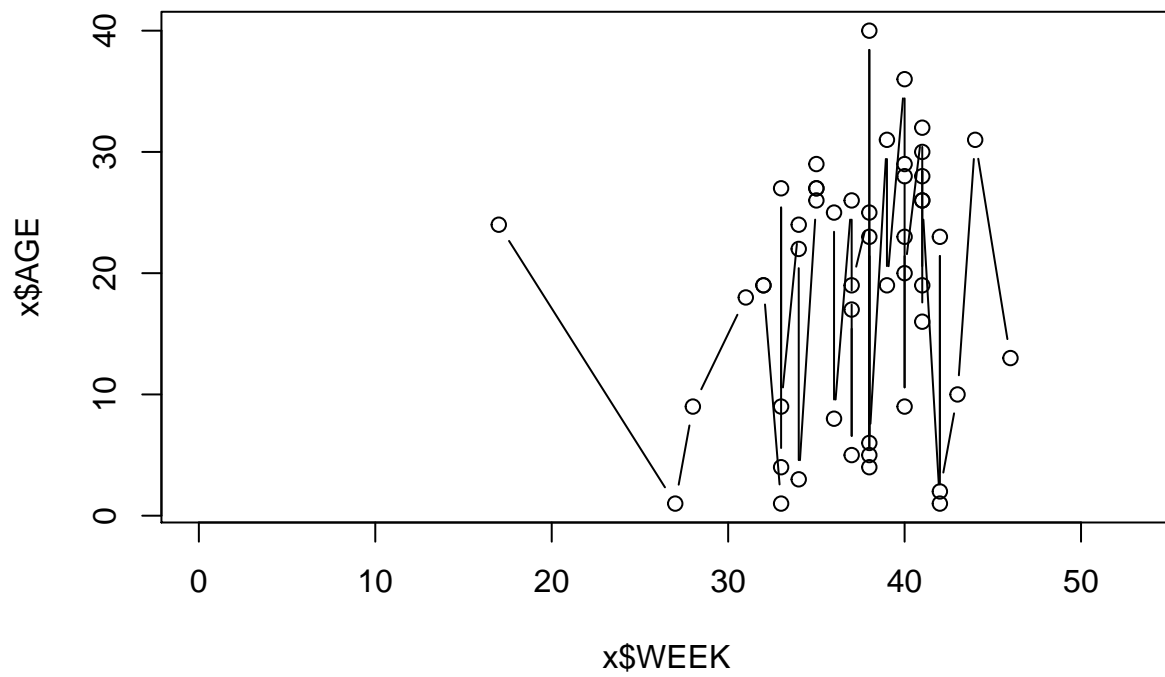
SANTA FE DO SUL

POPULINA

POTIM

x$WEEK

VOTUPORANGA

x$WEEK
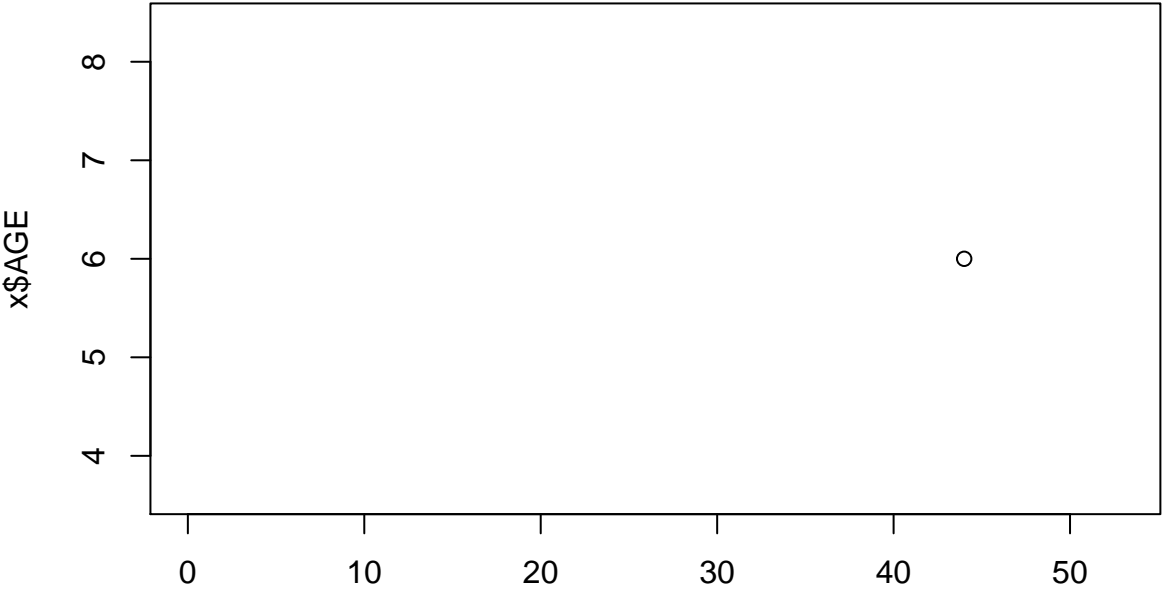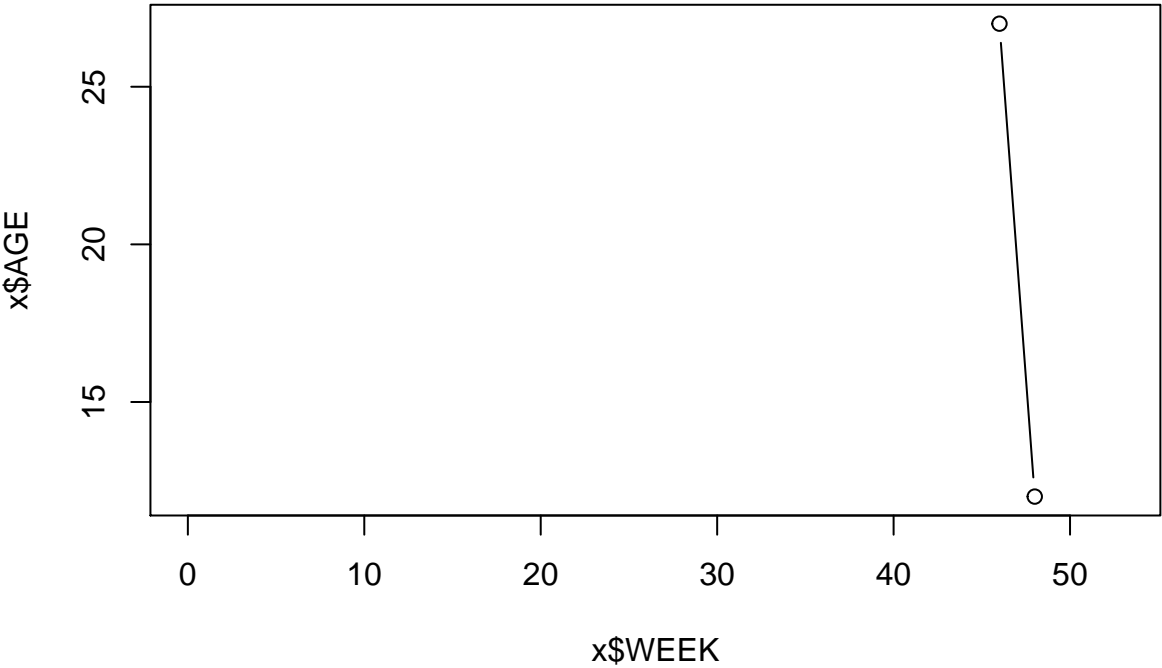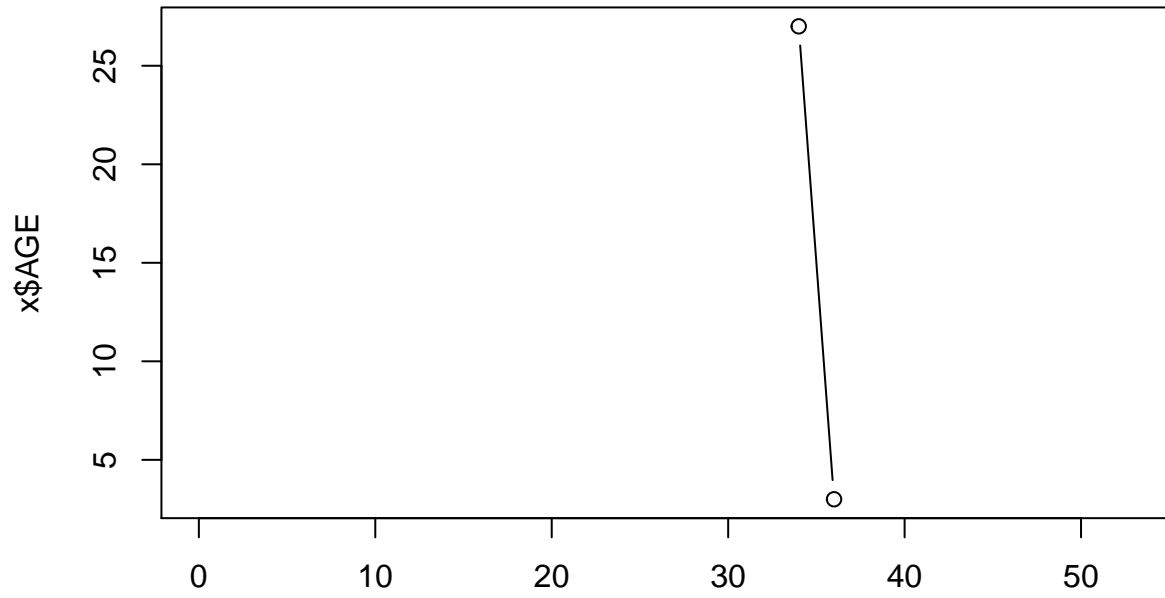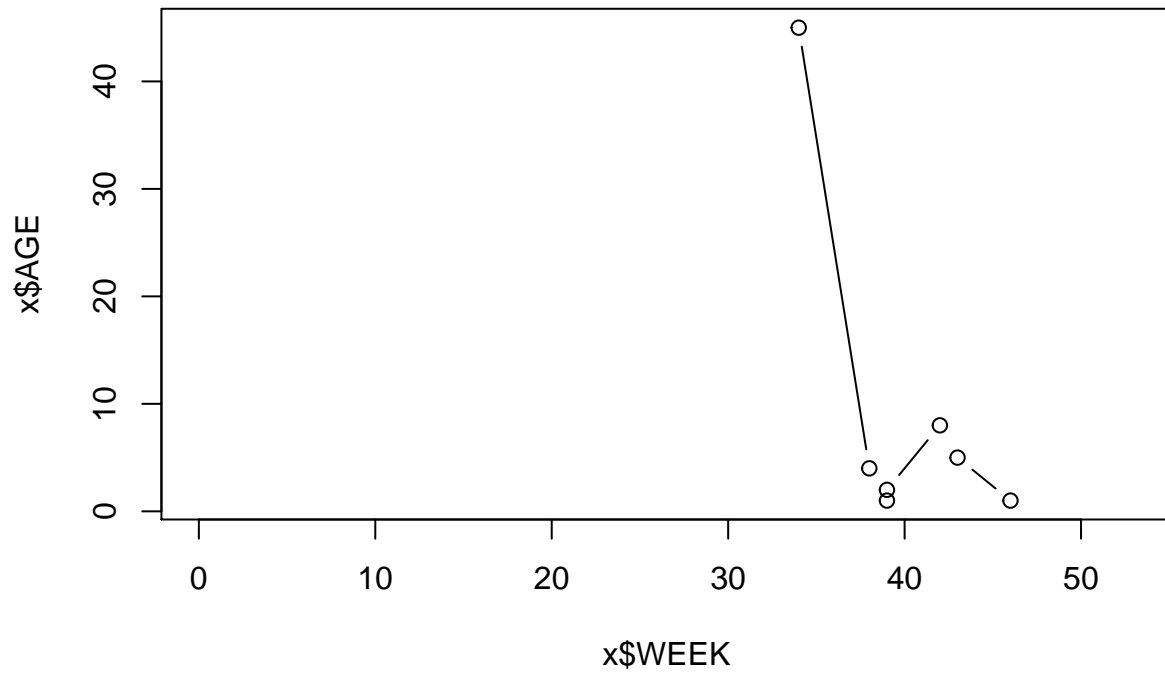
SANTO ANTONIO DO PINHAL

x$WEEK

ZACARIAS

x$WEEK
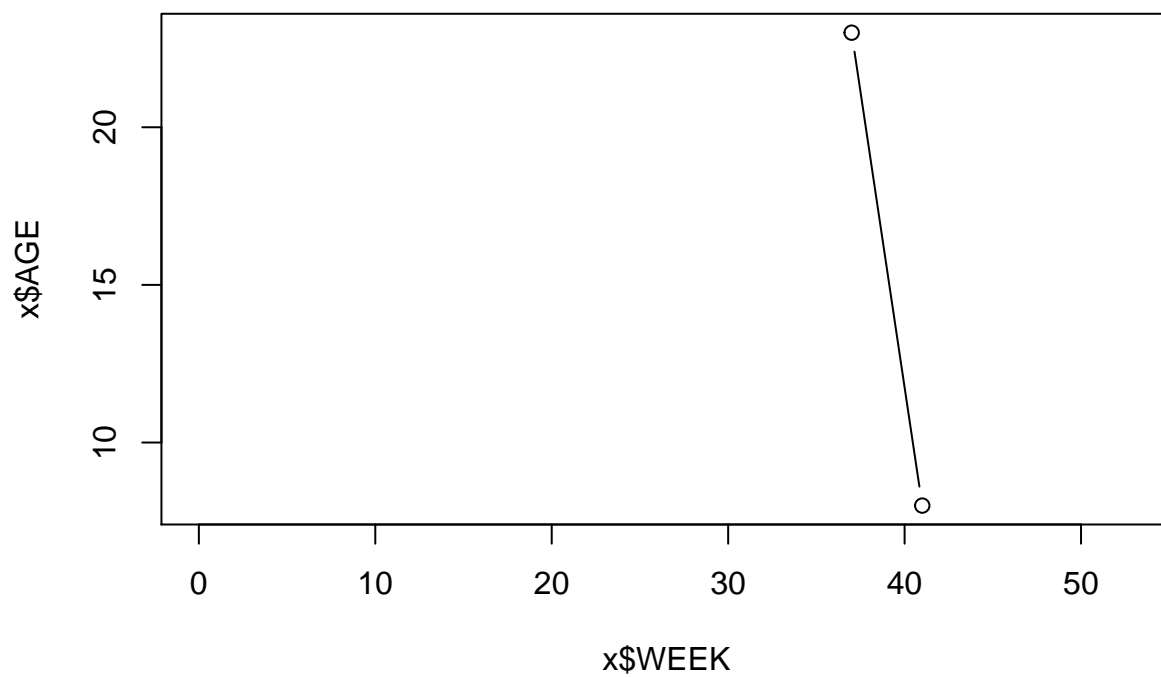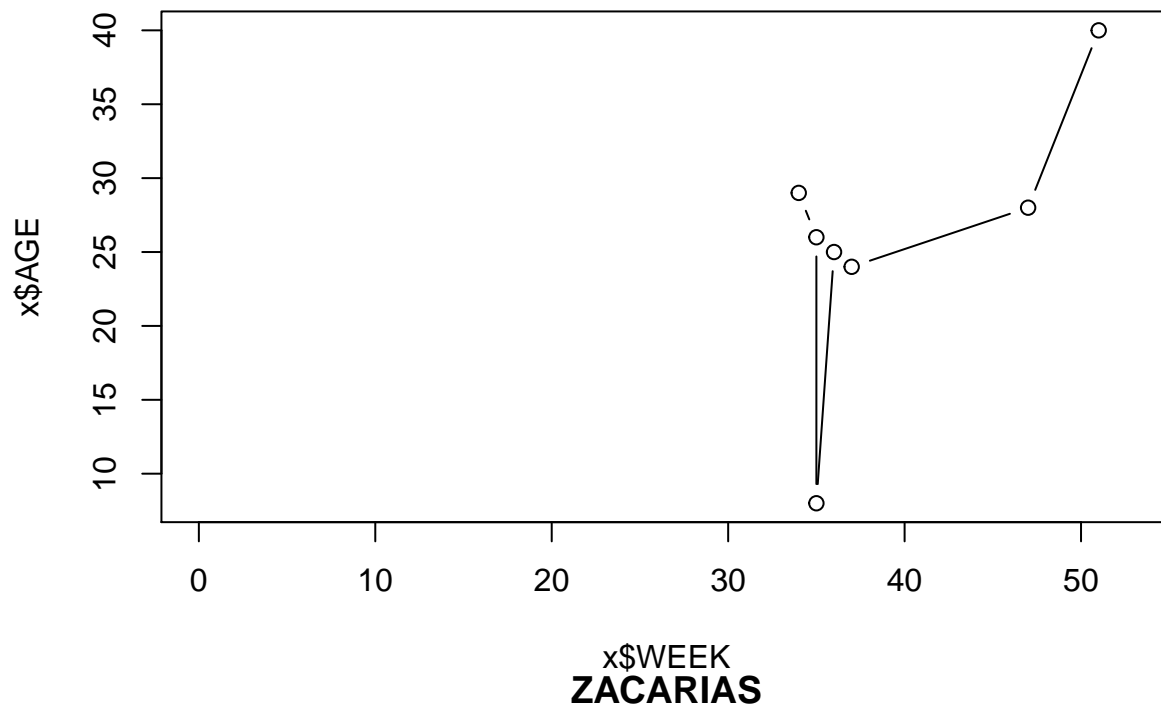
NOVO HORIZONTE
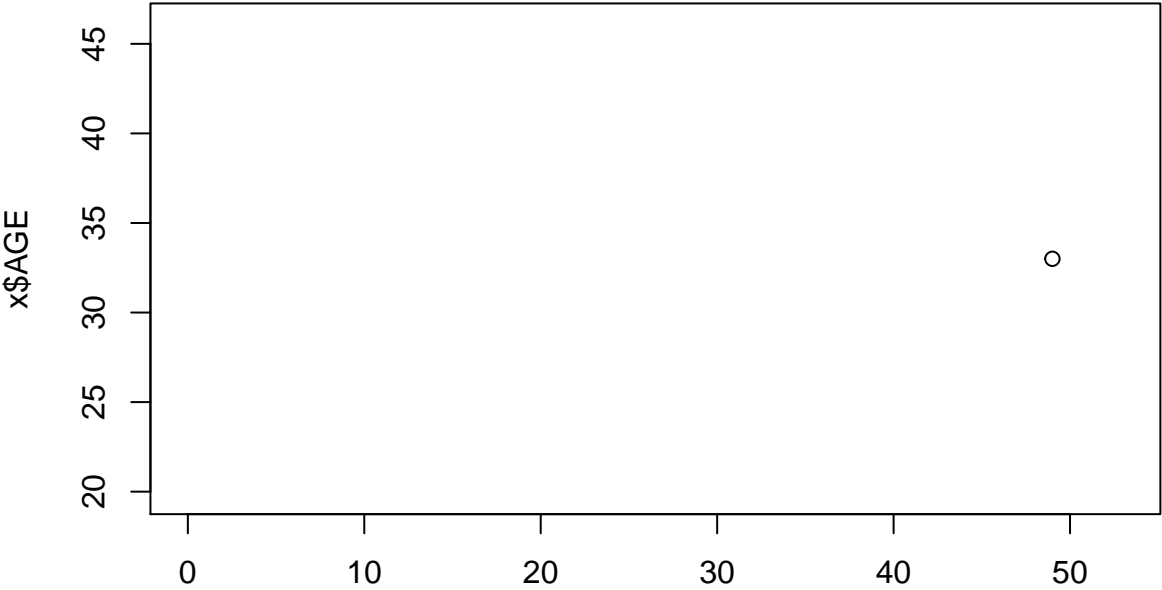
MENDONCA
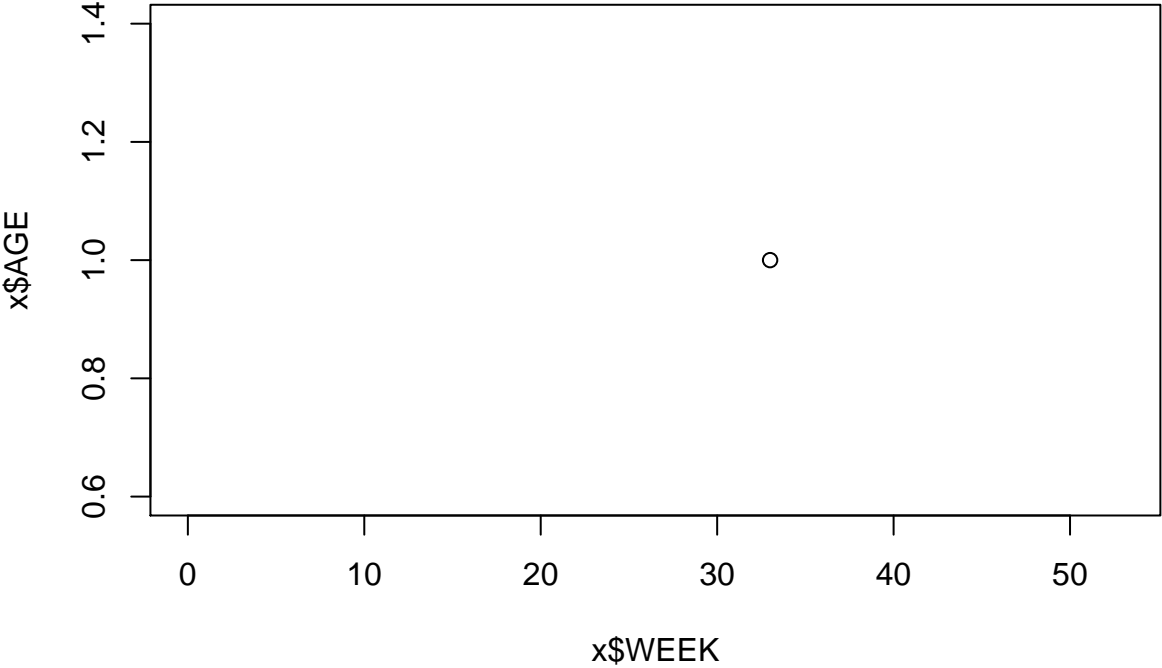
SAO BENTO DO SAPUCAI
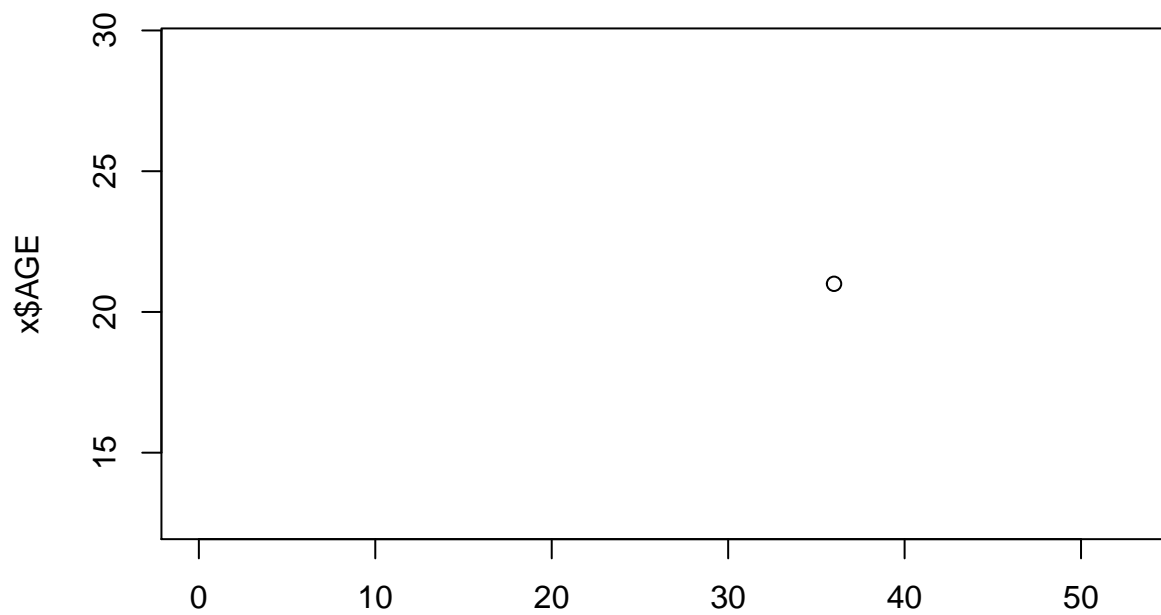
NAZARE PAULISTA

ILHABELA

x$WEEK

SAO FRANCISCO

x$WEEK
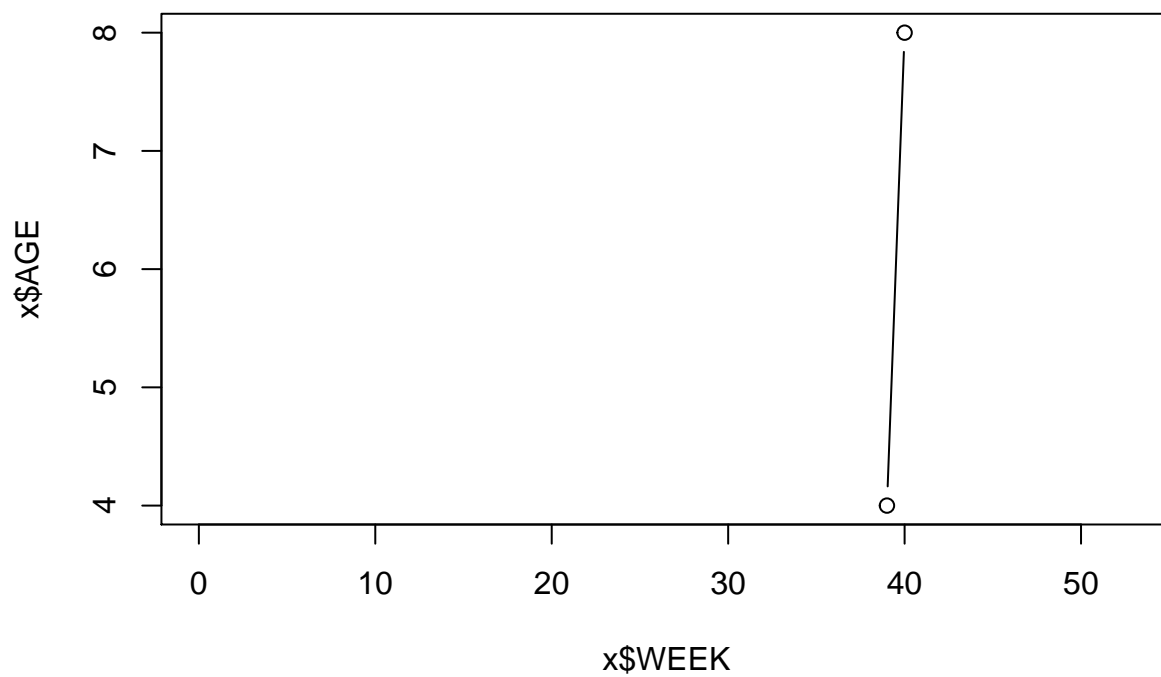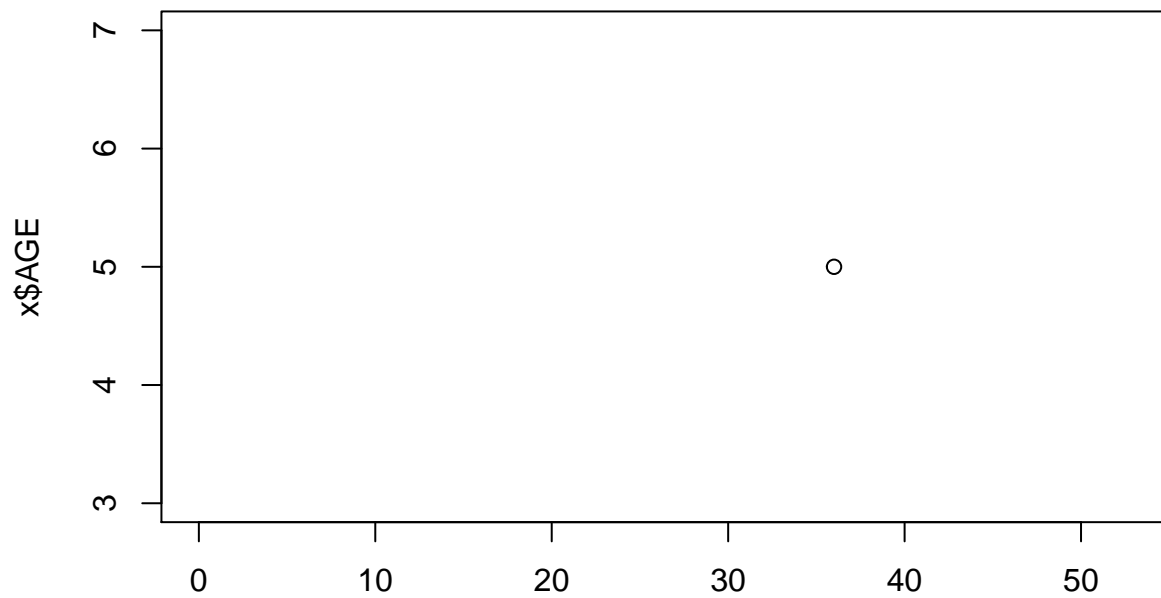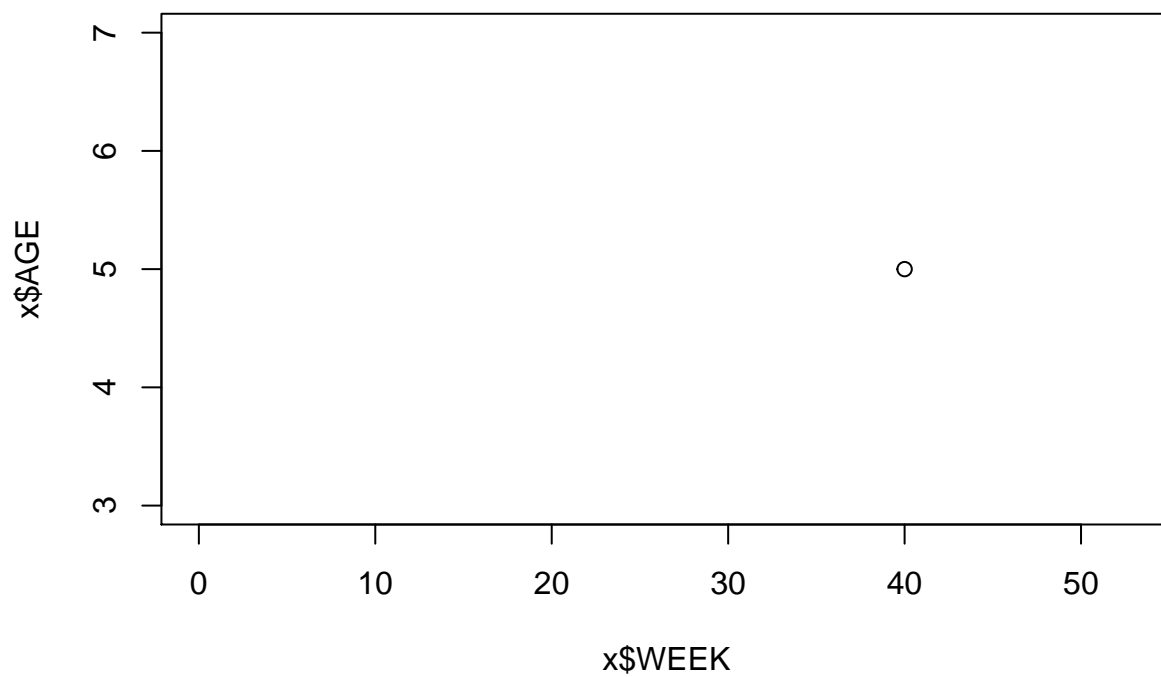
SANTA CRUZ DA CONCEICAO
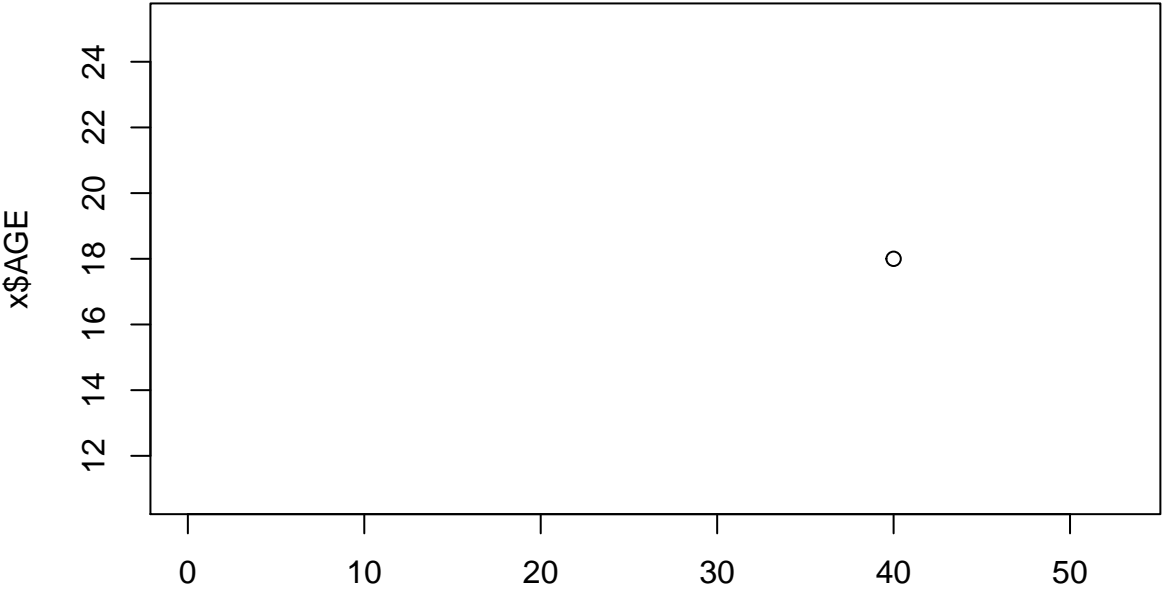
JERIQUARA

SALES OLIVEIRA

x$WEEK

IBIUNA

x$WEEK
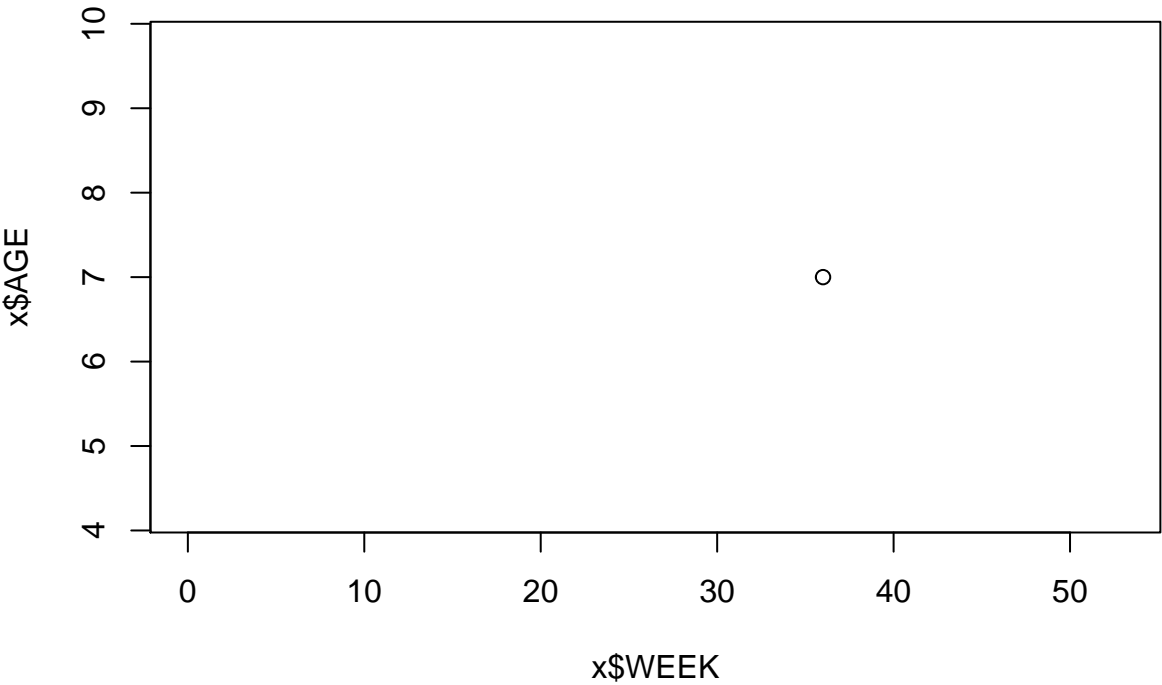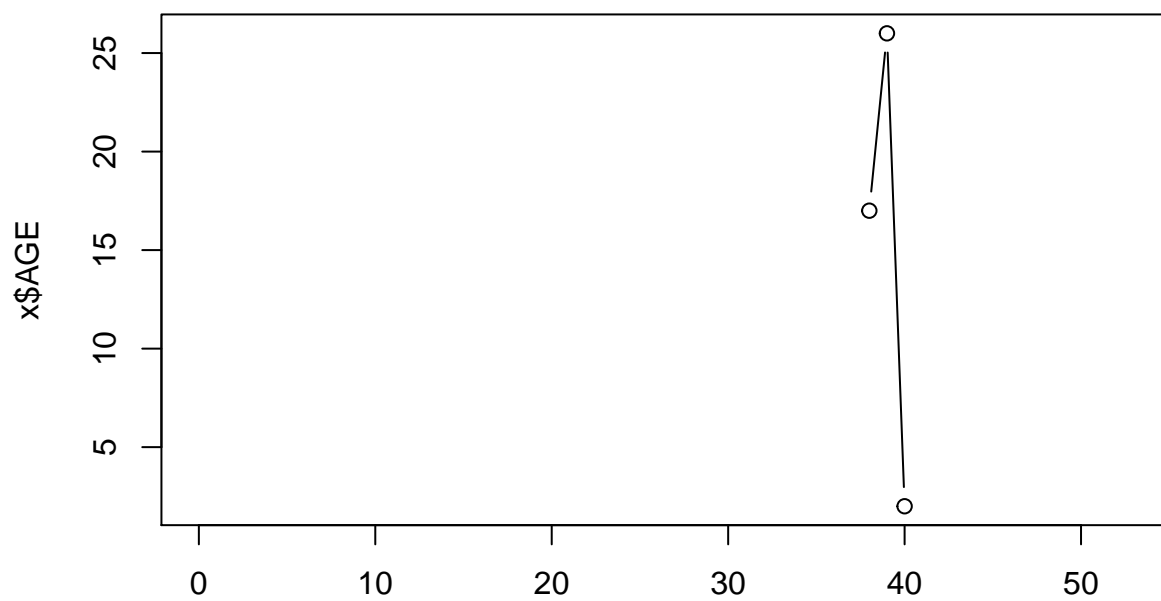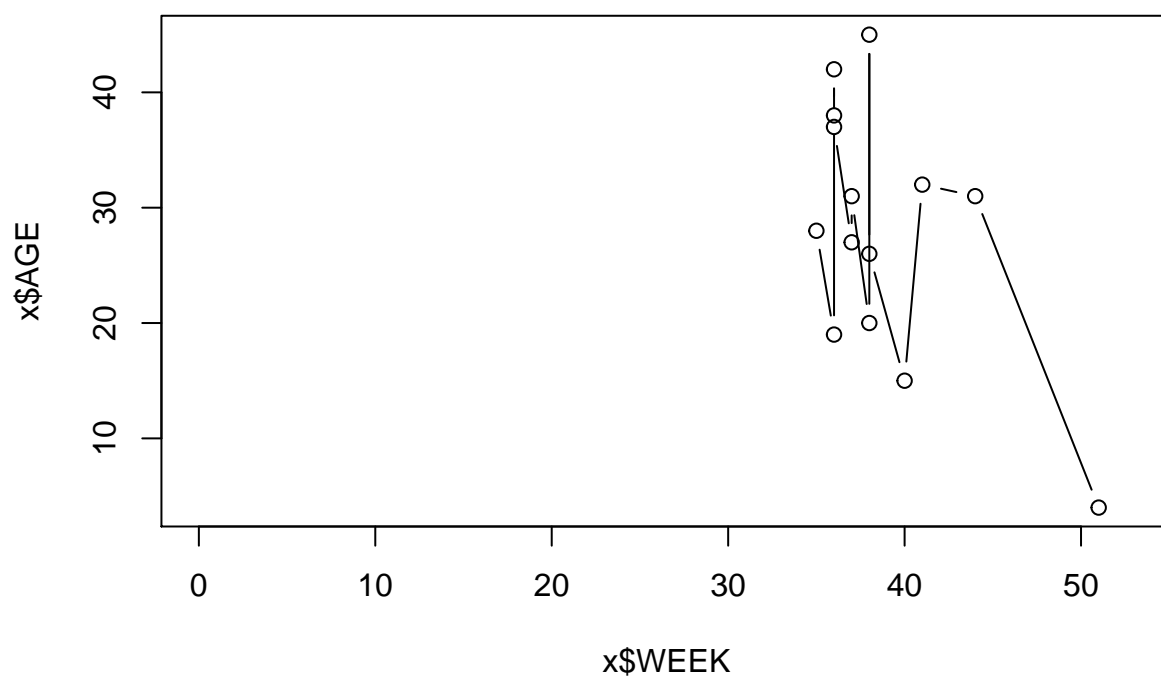
INDIAPORA
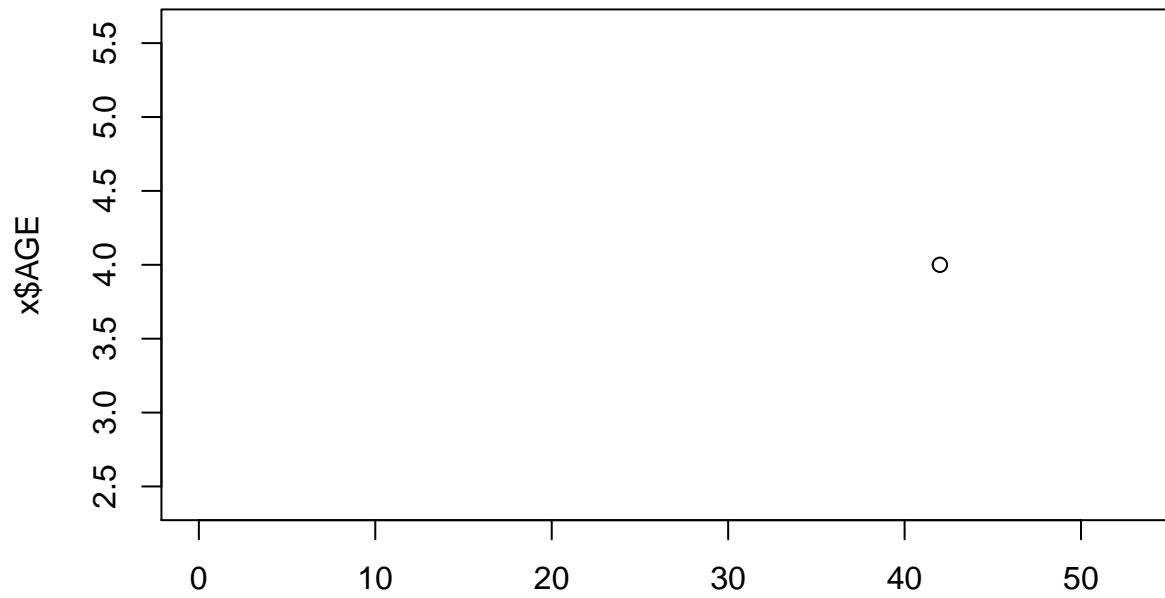
JUQUITIBA
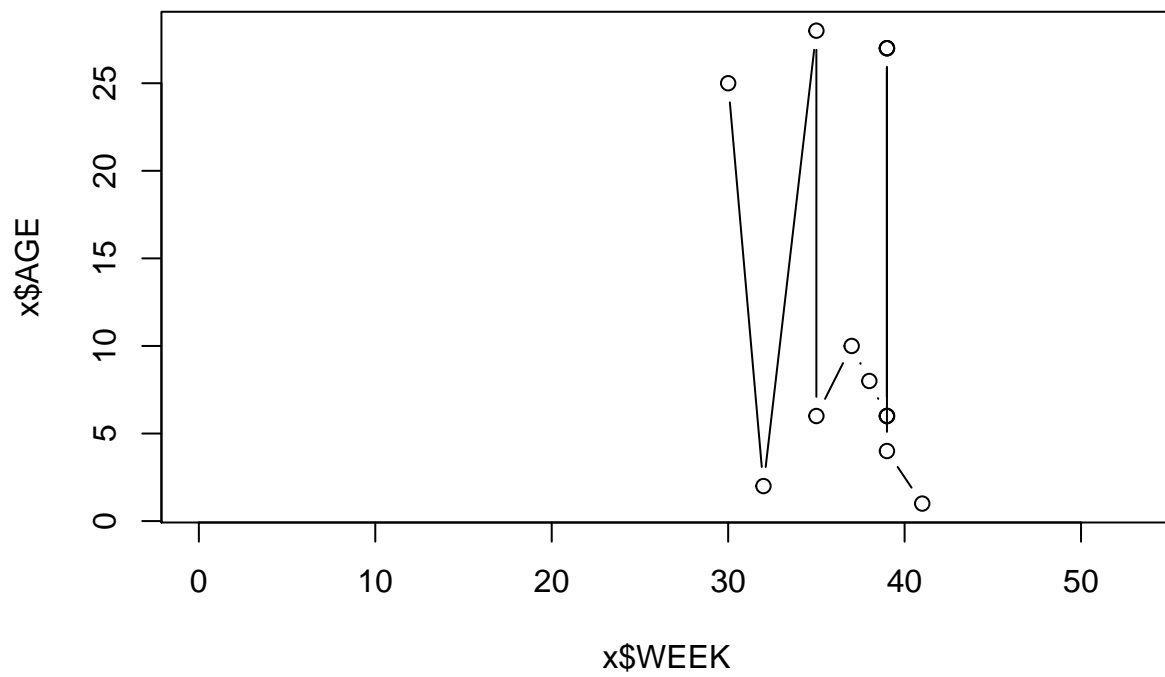
CAJOBI
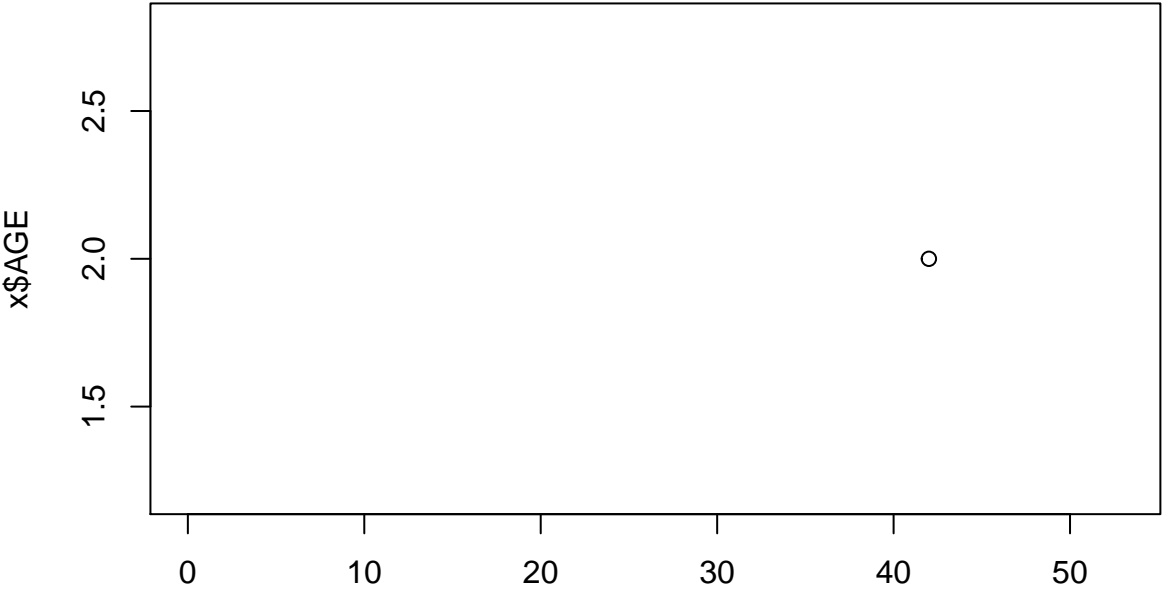
AMERICO DE CAMPOS
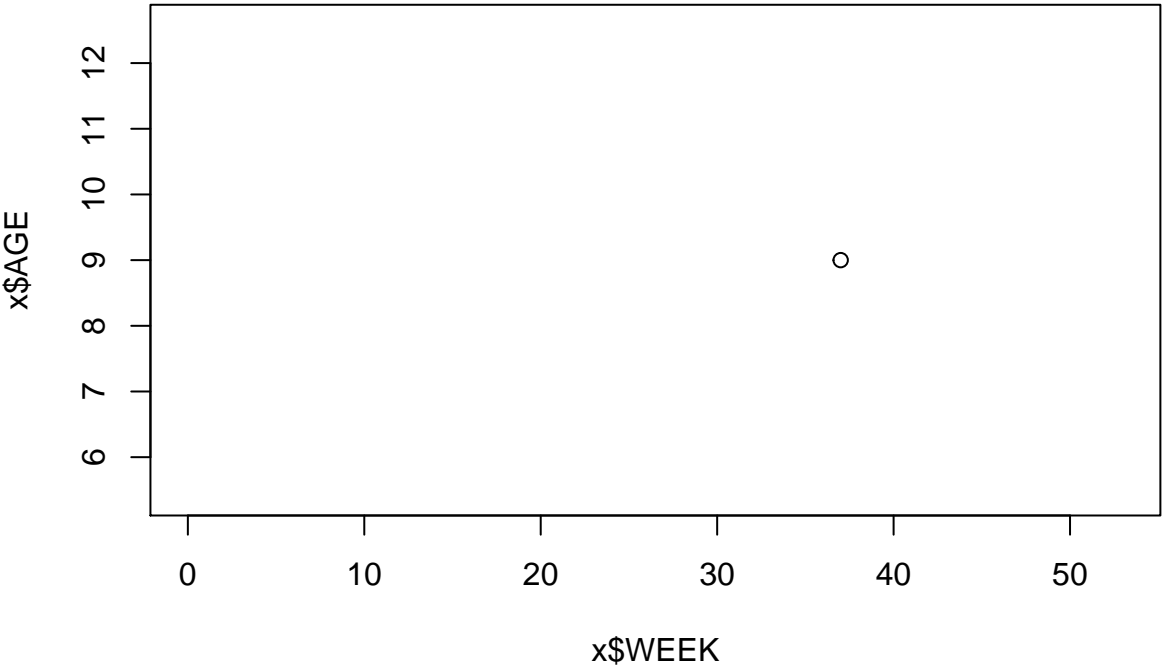
ARCO–IRIS
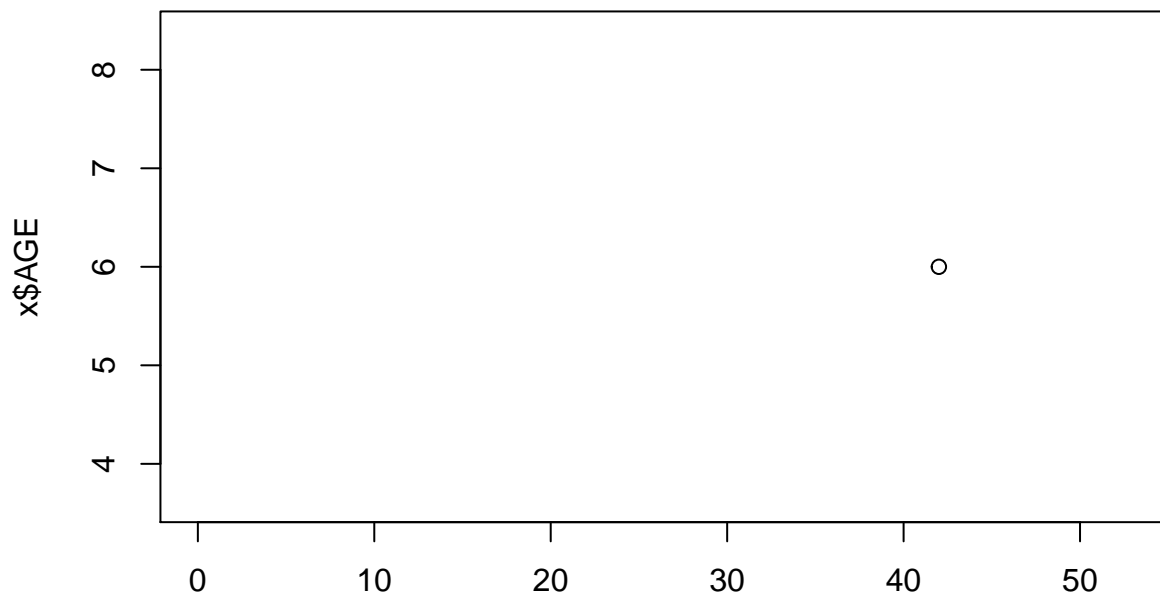
GUAPIACU
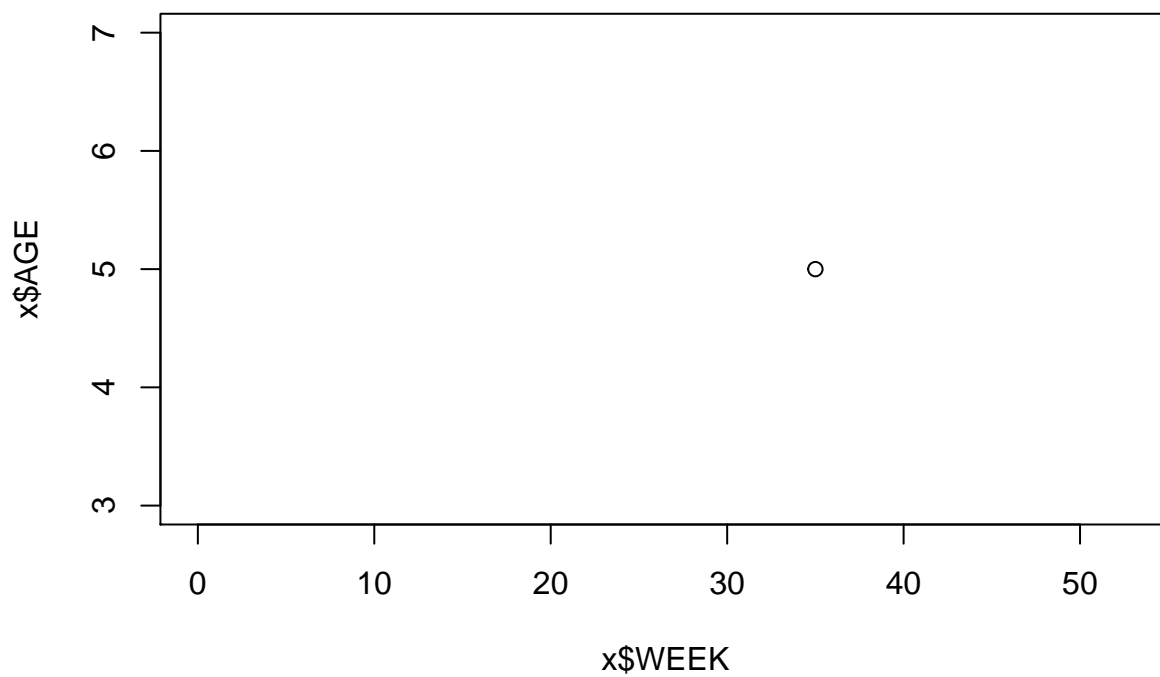
OLEO
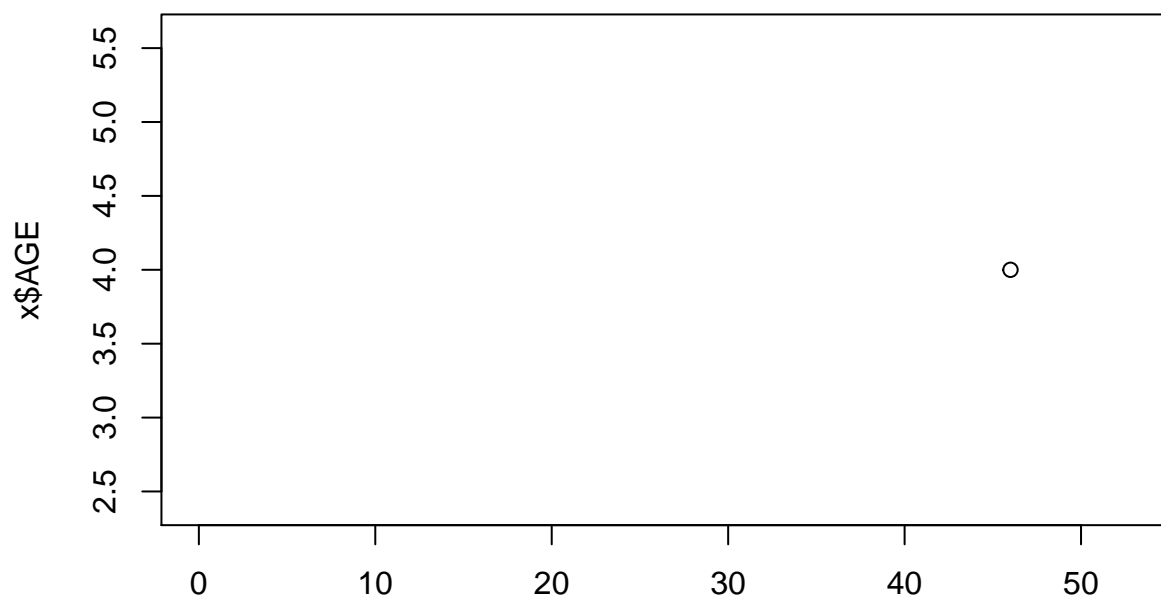
CHAVANTES
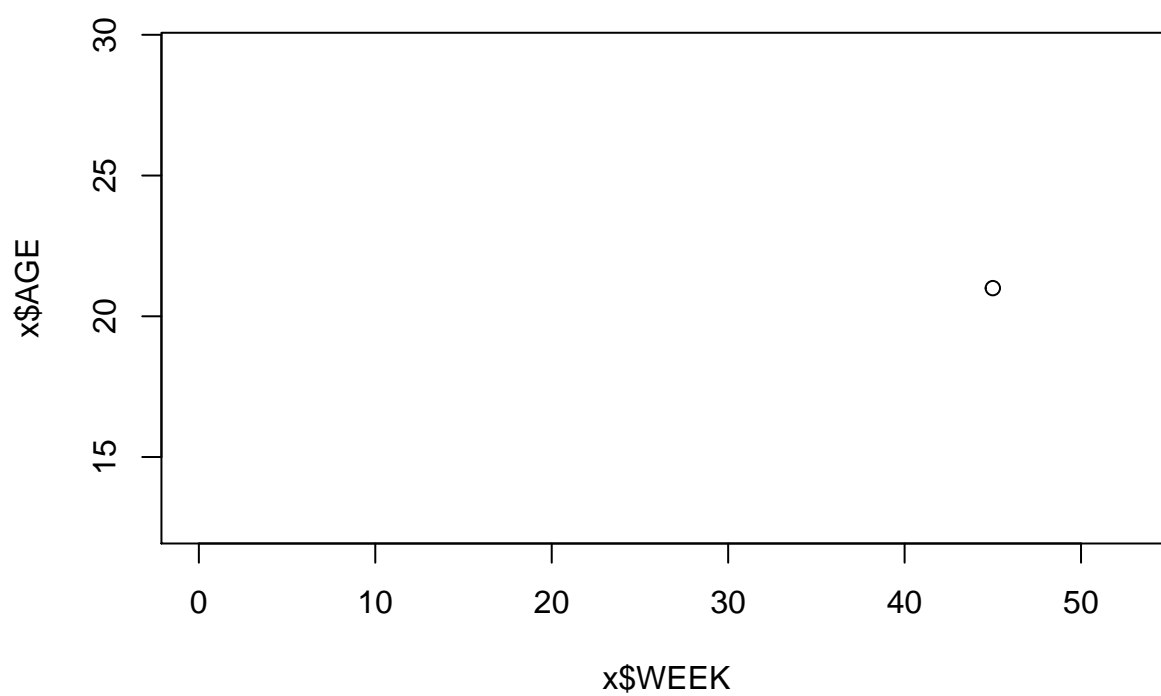
ESTRELA DO NORTE
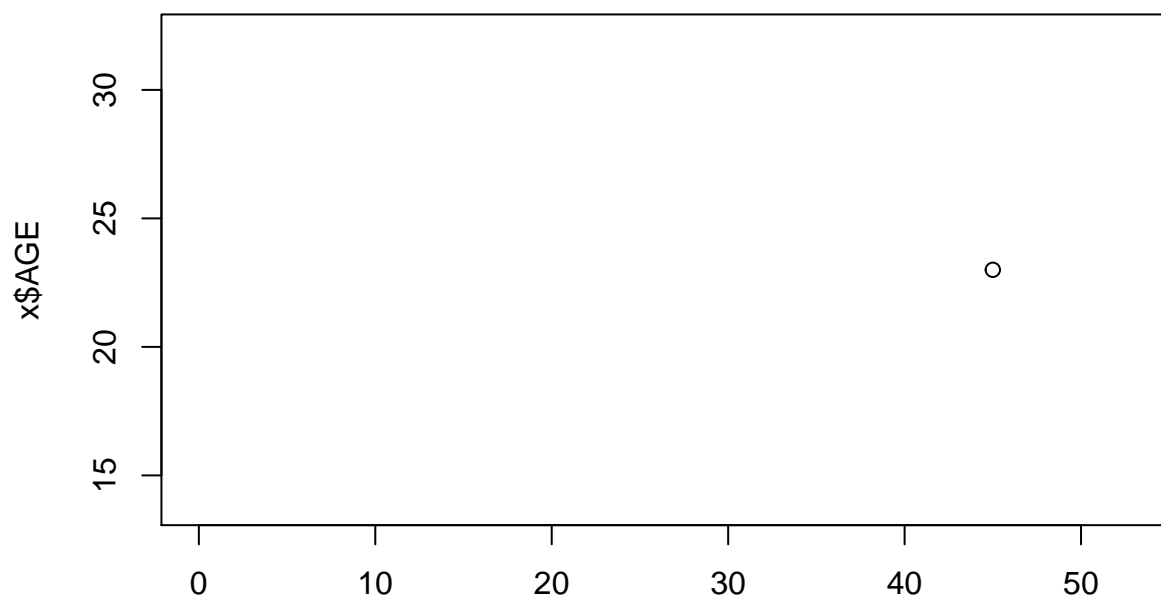
ASSIS
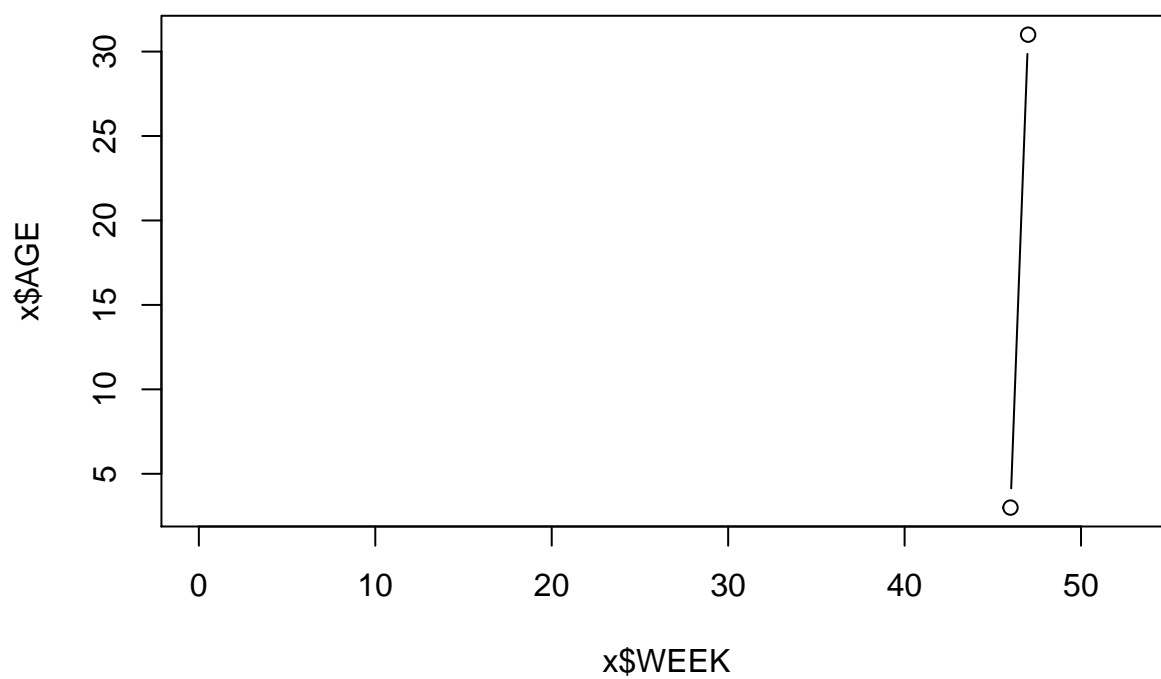
SANTO ANTONIO DE POSSE
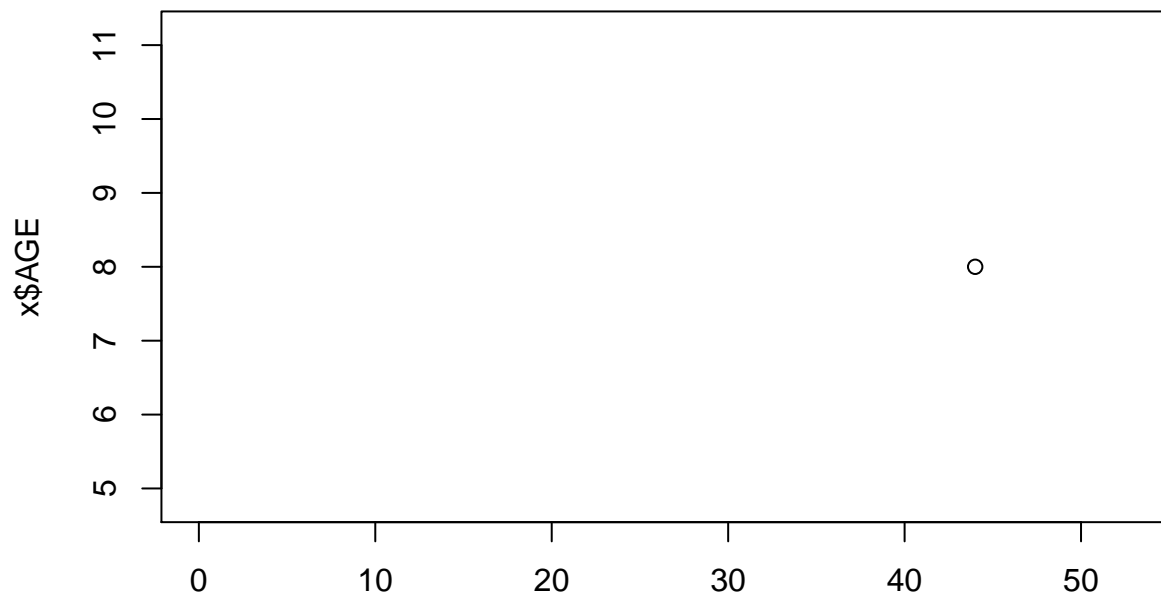
BIRITIBA−MIRIM
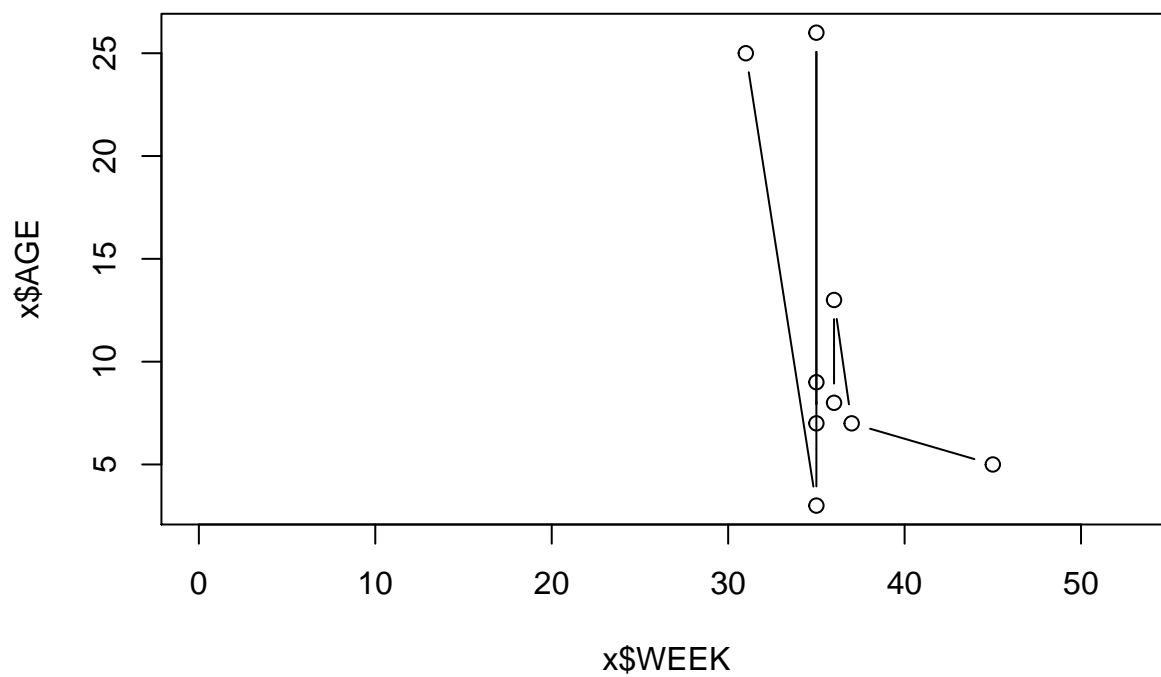
BOCAINA
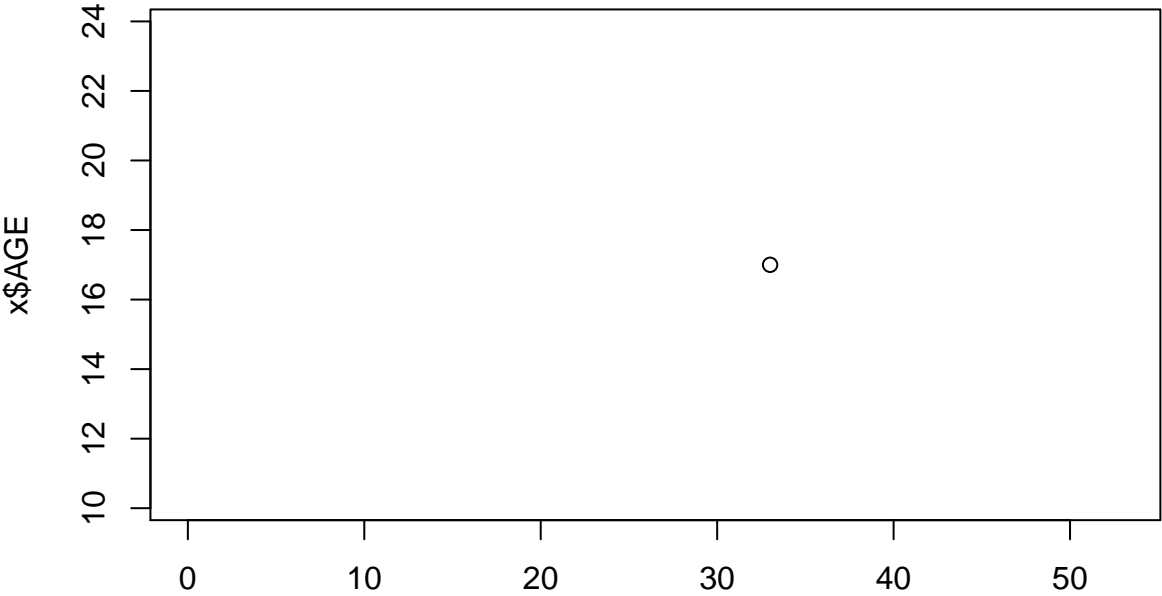
EMILIANOPOLIS
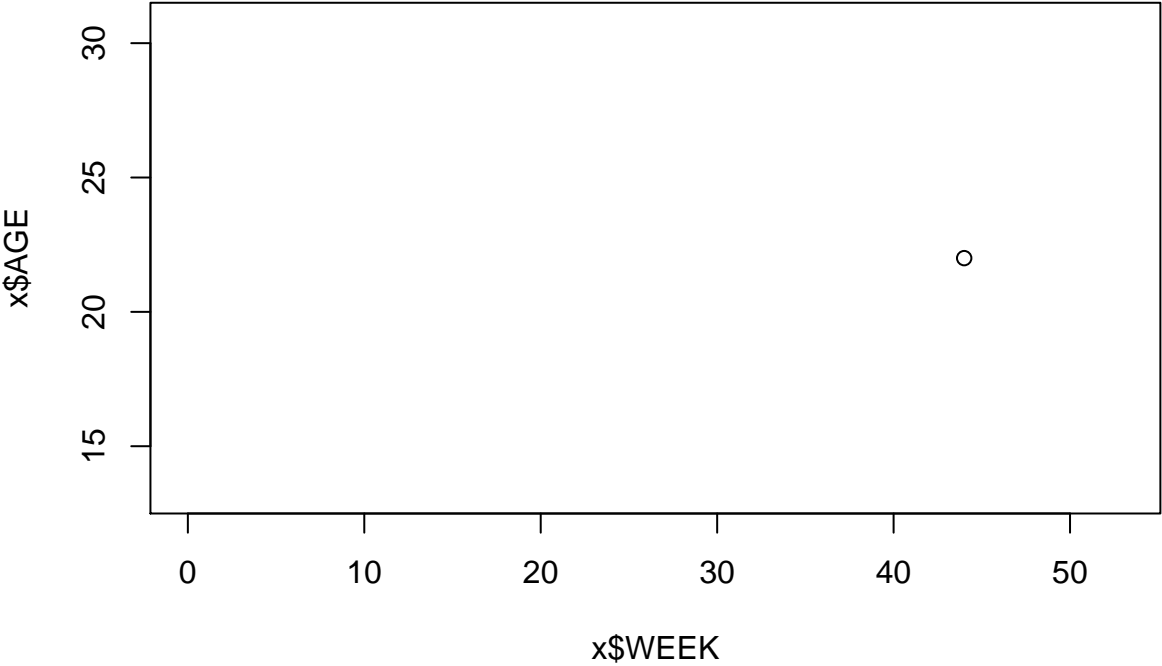
ADAMANTINA
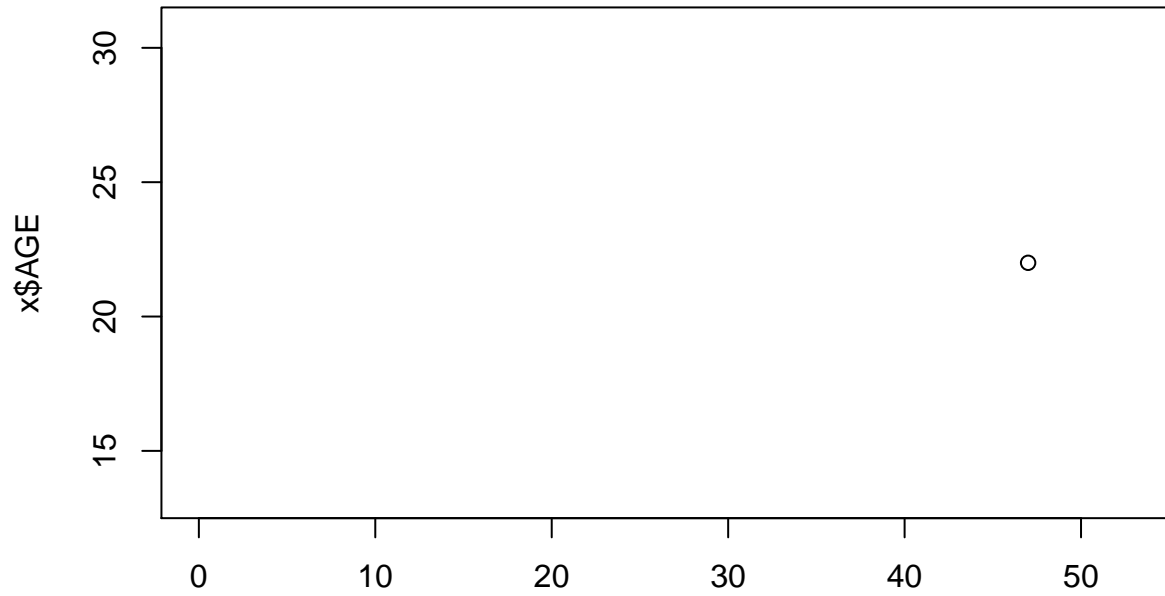
IPUA
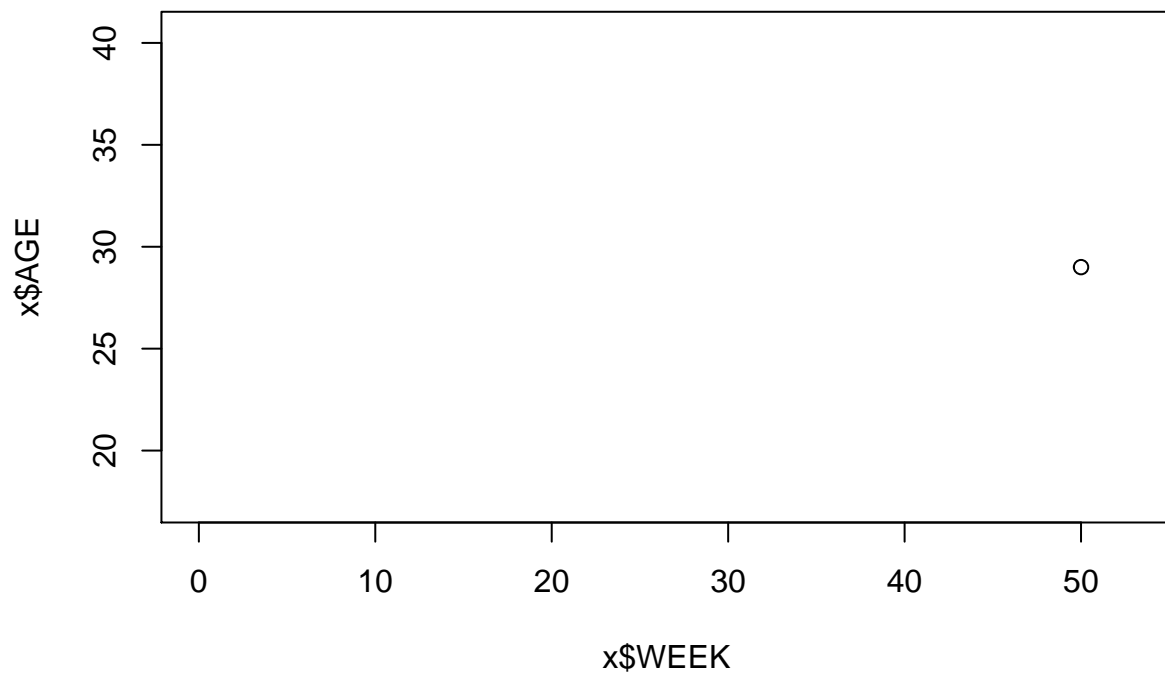
AGUAS DE LINDOIA
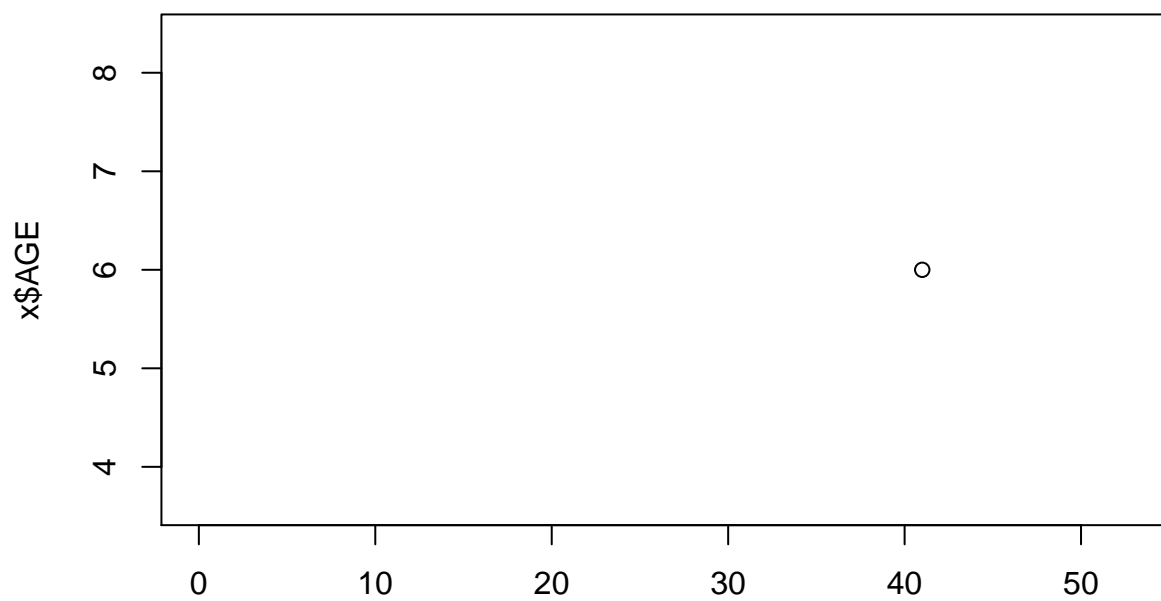
CAPAO BONITO
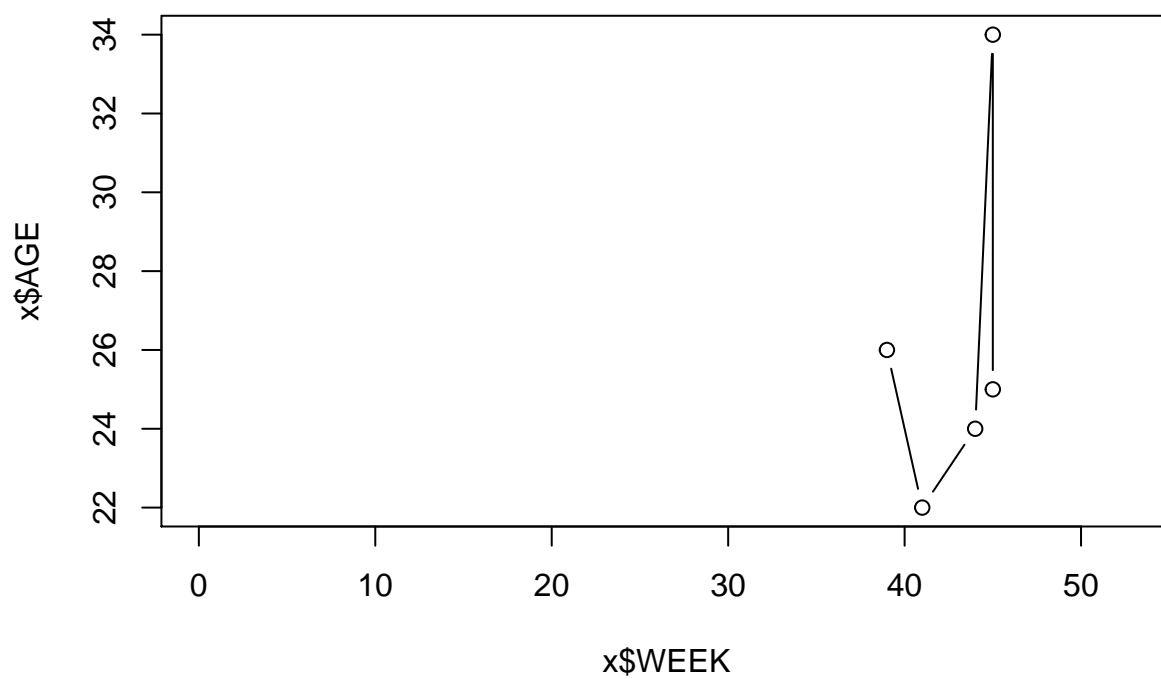
BARRA DO CHAPEU
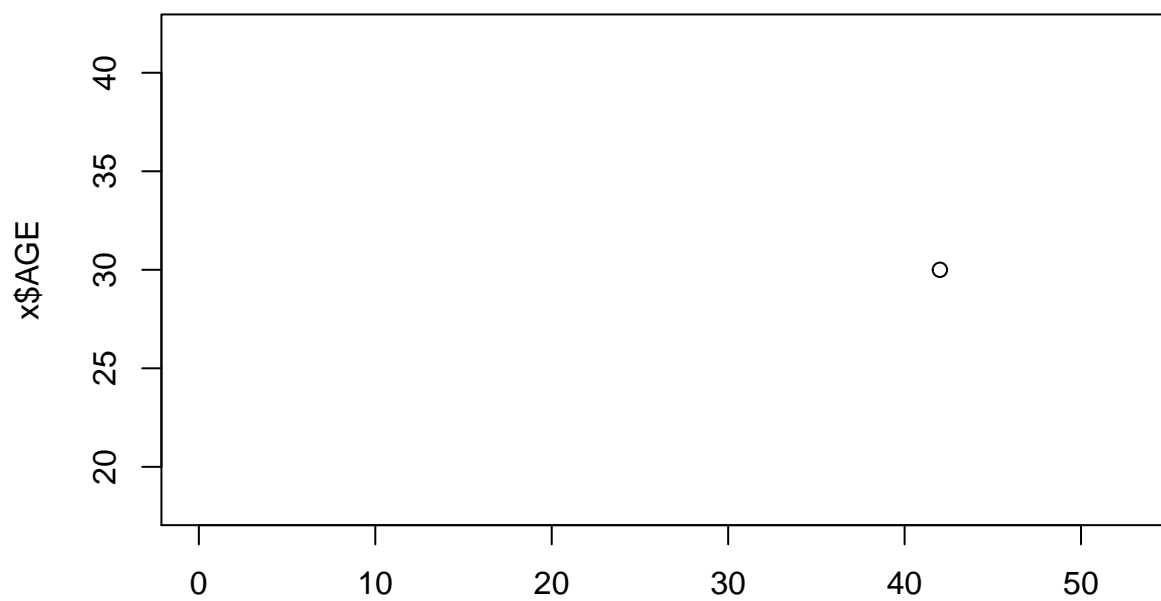
BOM SUCESSO DE ITARARE

ITABERA

x$WEEK

PORTO ALEGRE

x$WEEK
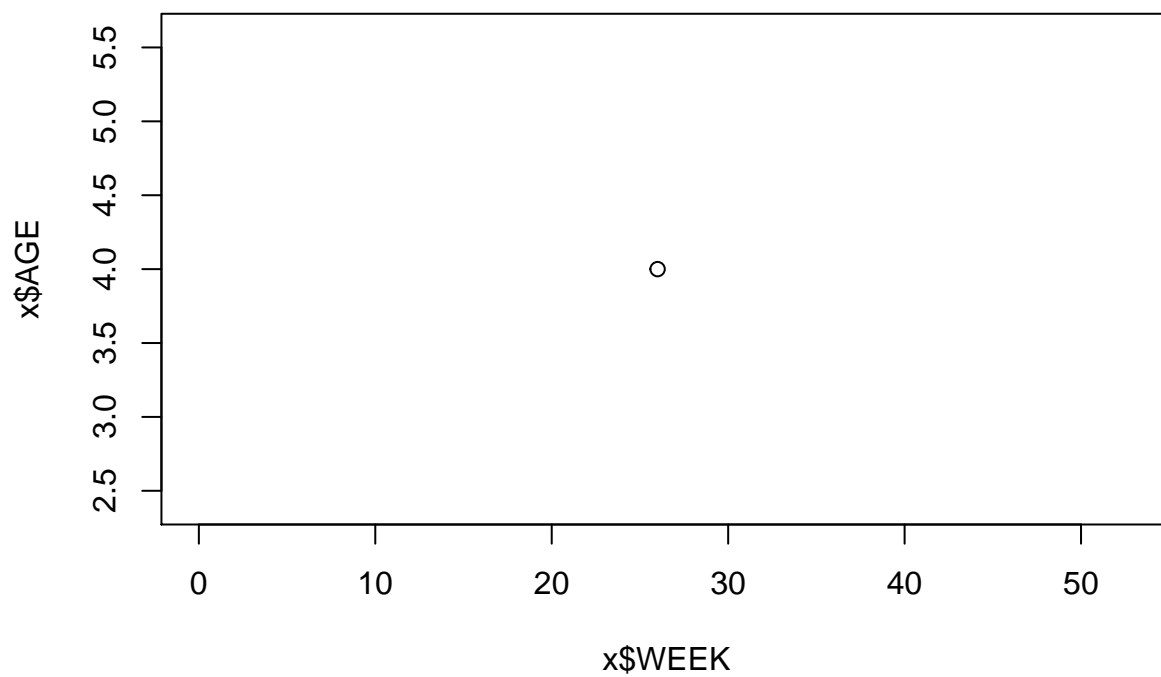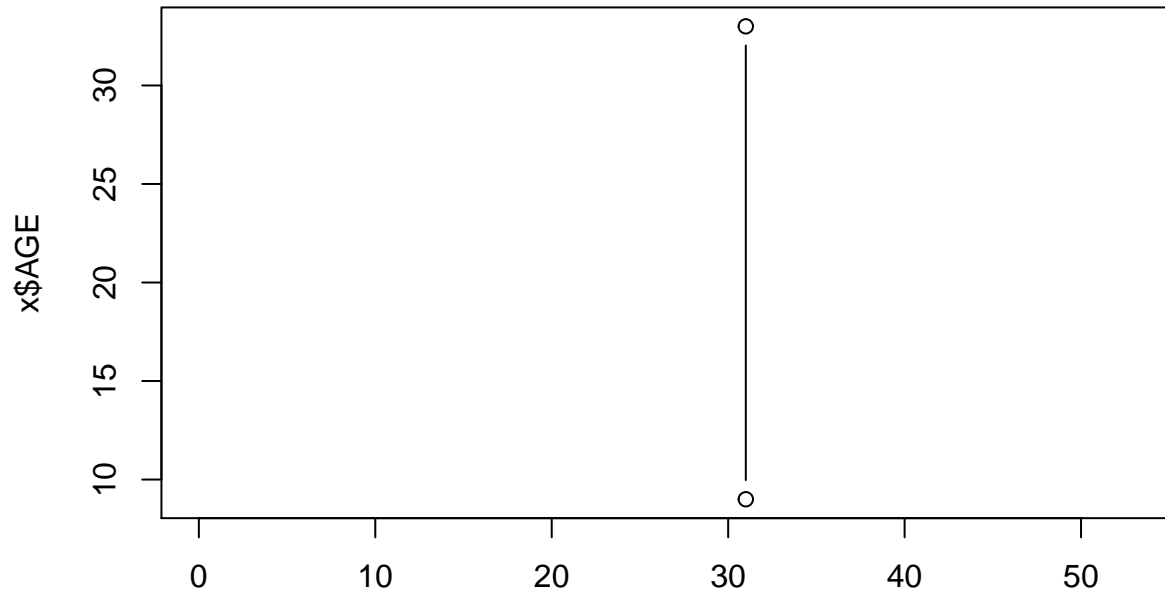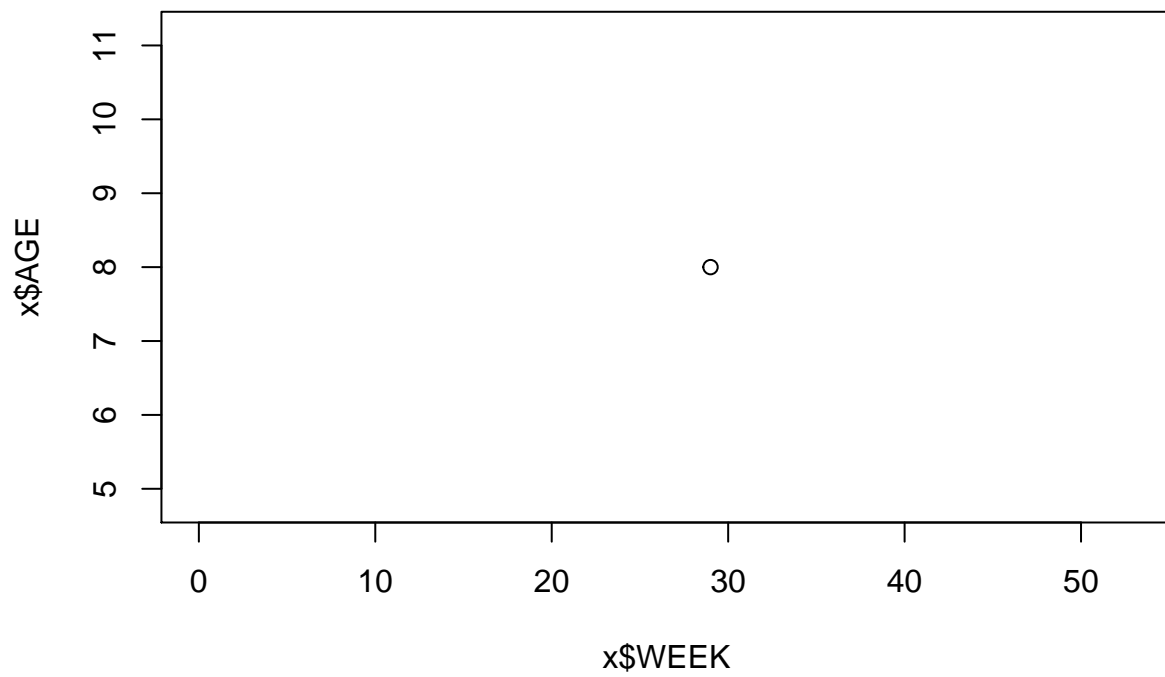
NOVA CAMPINA
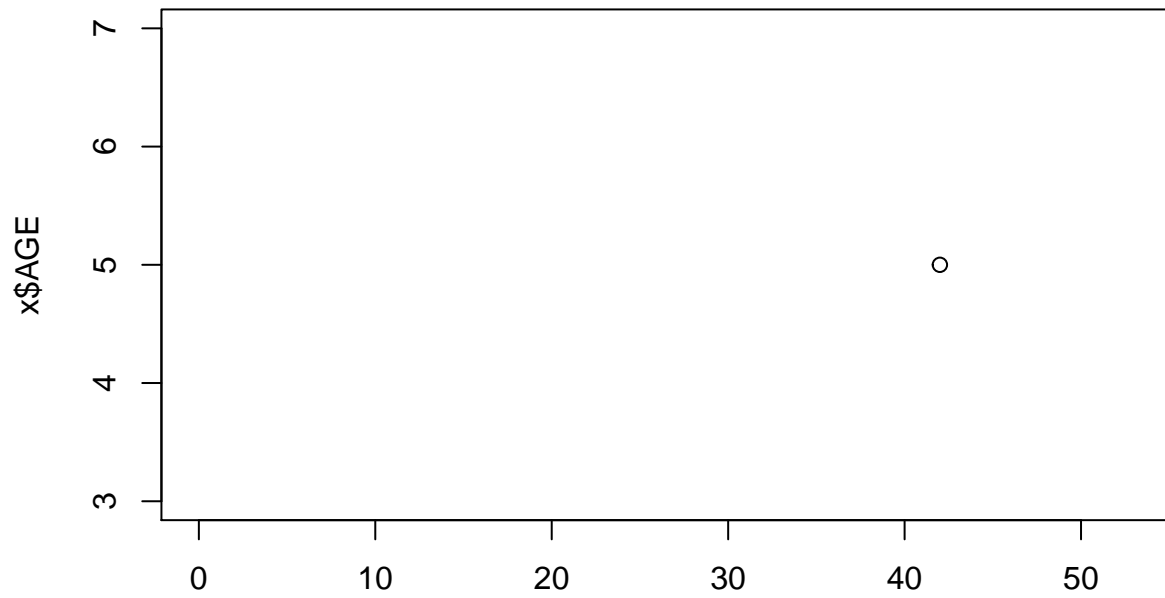
BURITAMA
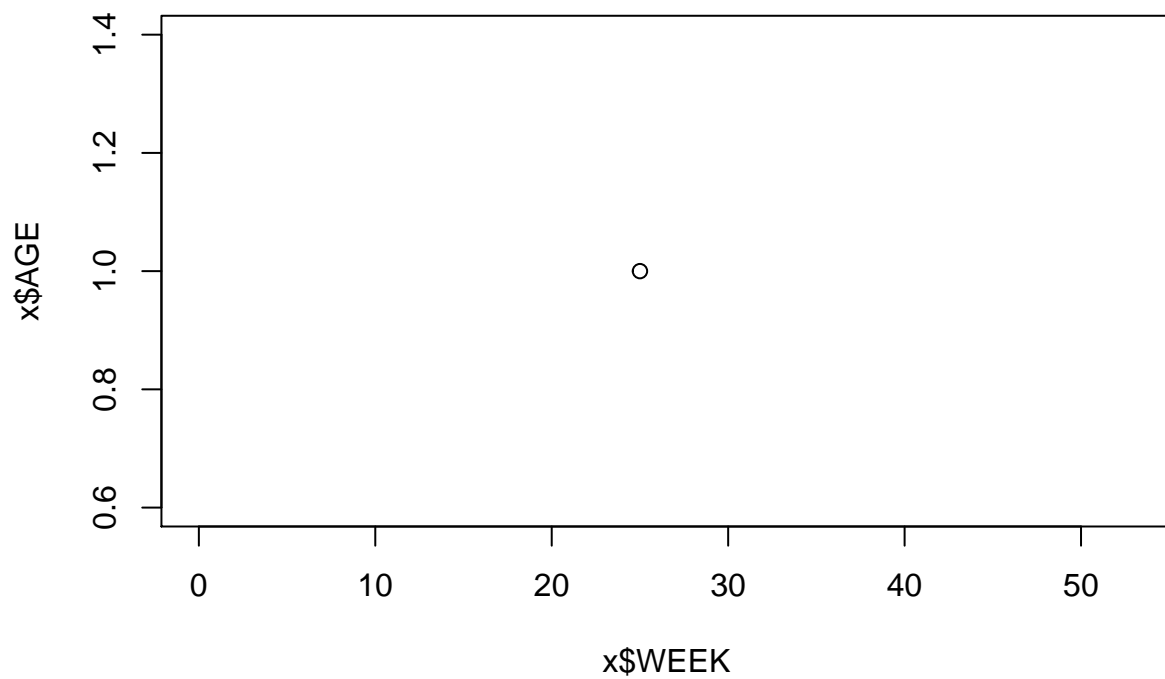
Preliminary analysis:

Graphs of counties with less than 10 cases are really unhelpful visually. Additionally, the the line graph seems to indicate causation between cases, which is not accurate.

Next step:

- If available, make all of the cases which occurred in the same district the same color.

Examination by district

Just realized that I only have district data for those in Sao Paulo county. I will make graphs for all of the information that I have available.

```
x <- subset(data, data$COUNTY == "SAO PAULO")
x <- subset(x, !is.na(x$DISTRICT))
test_district <- as.character(unique(x$DISTRICT))


sp.index_week_mat <- subset(index_week_mat, index_week_mat$COUNTY == county_vec[1])
sp.data <- data[as.numeric(as.character(sp.index_week_mat$NUM)),]
sp.data <- subset(sp.data, !is.na(sp.data$DISTRICT))
unique.district <- unique(sp.data$DISTRICT)

for(i in 1:length(unique.district)){
  dis.which <- which(sp.data$DISTRICT == unique.district[i])
  plot(sp.index_week_mat$WEEK[dis.which], sp.index_week_mat$AGE[dis.which],
     pch = 19,
     main = unique.district[i])
}
```

SAO MIGUEL

PEDREIRA

sp.index_week_mat$WEEK[dis.which]

BRASILANDIA
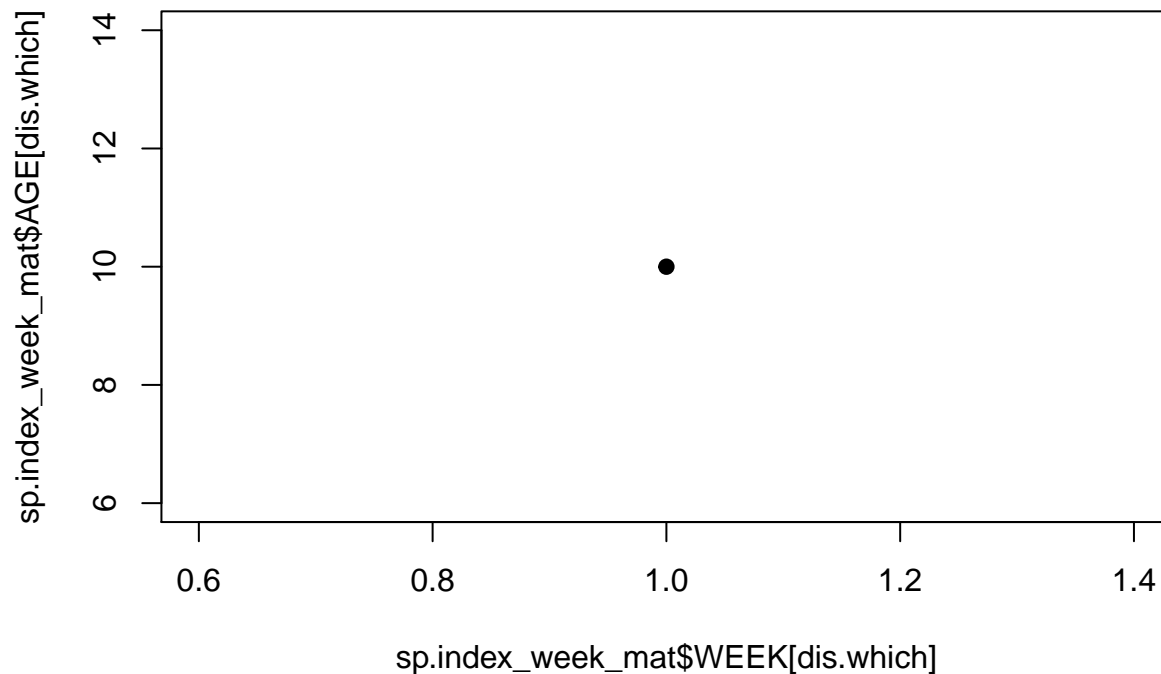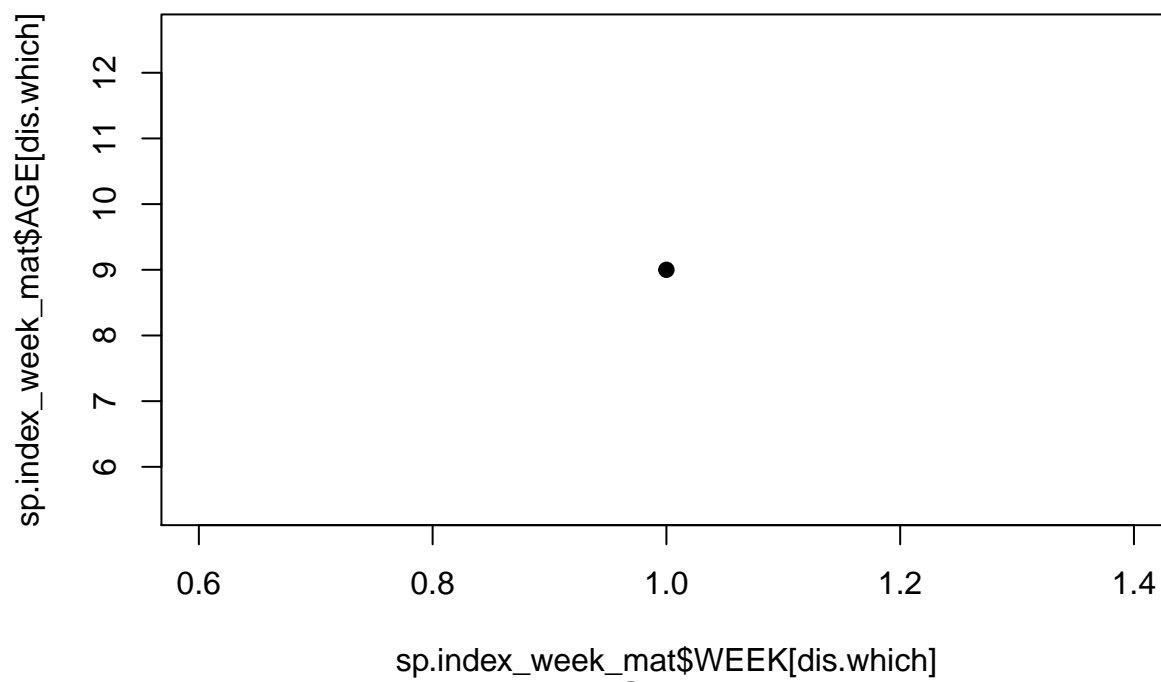
sp.index_week_mat$WEEK[dis.which]
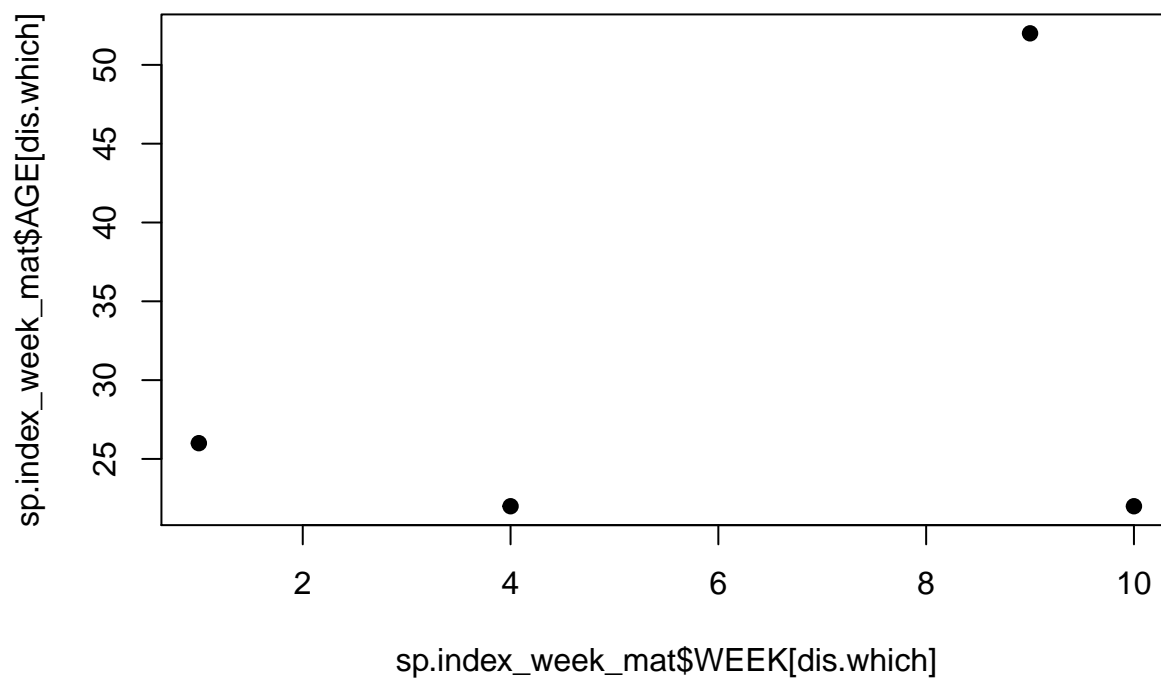
VILA MARIANA
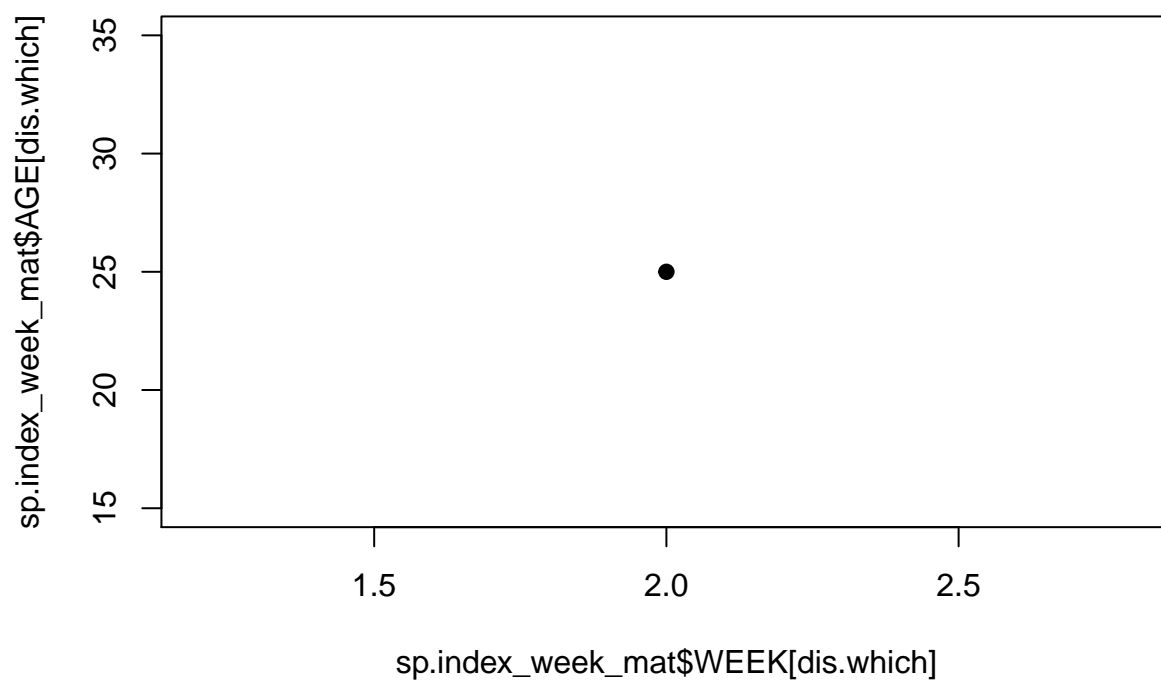
VILA SONIA

JARDIM SAO LUIS

SAUDE
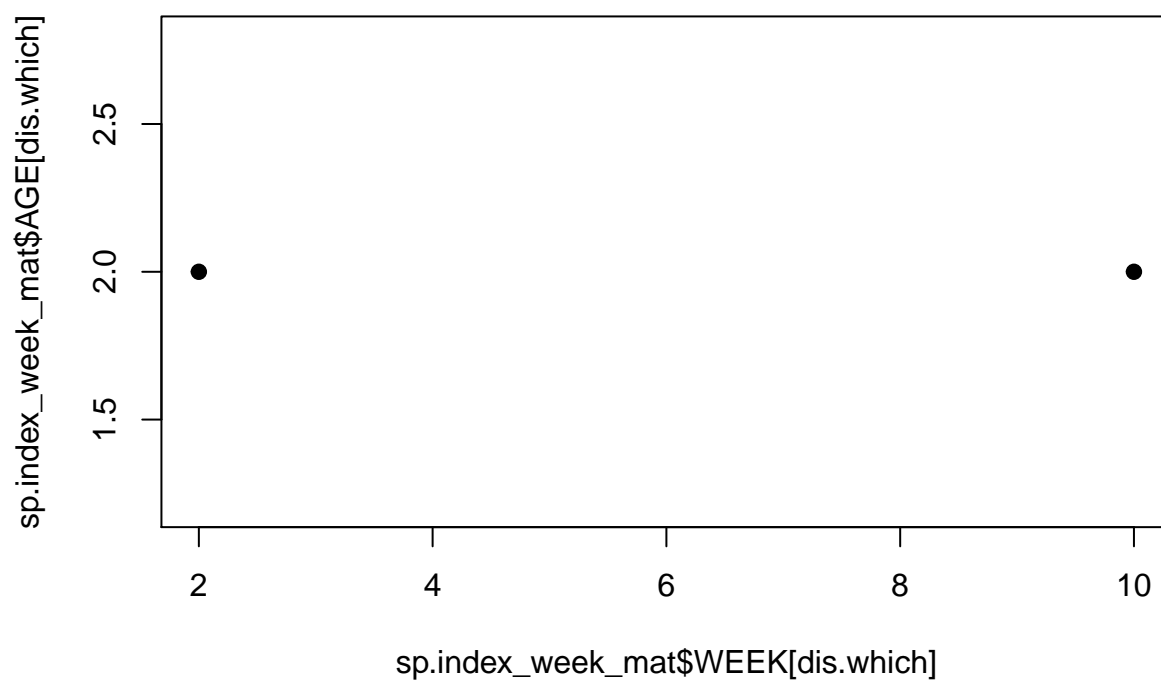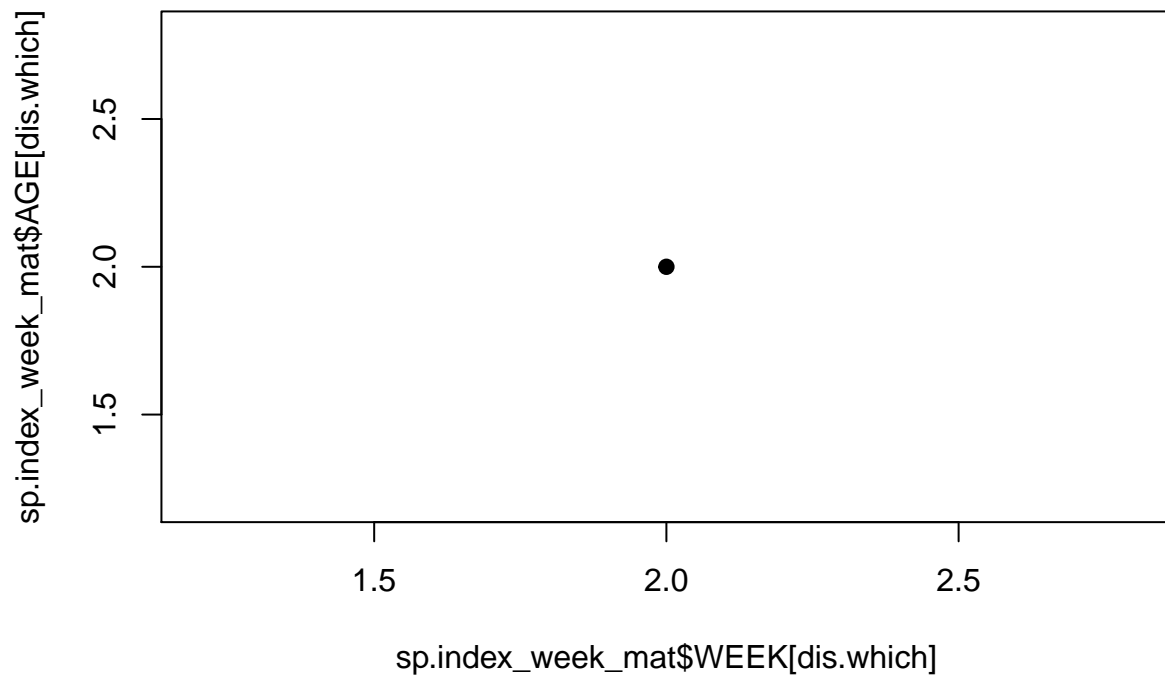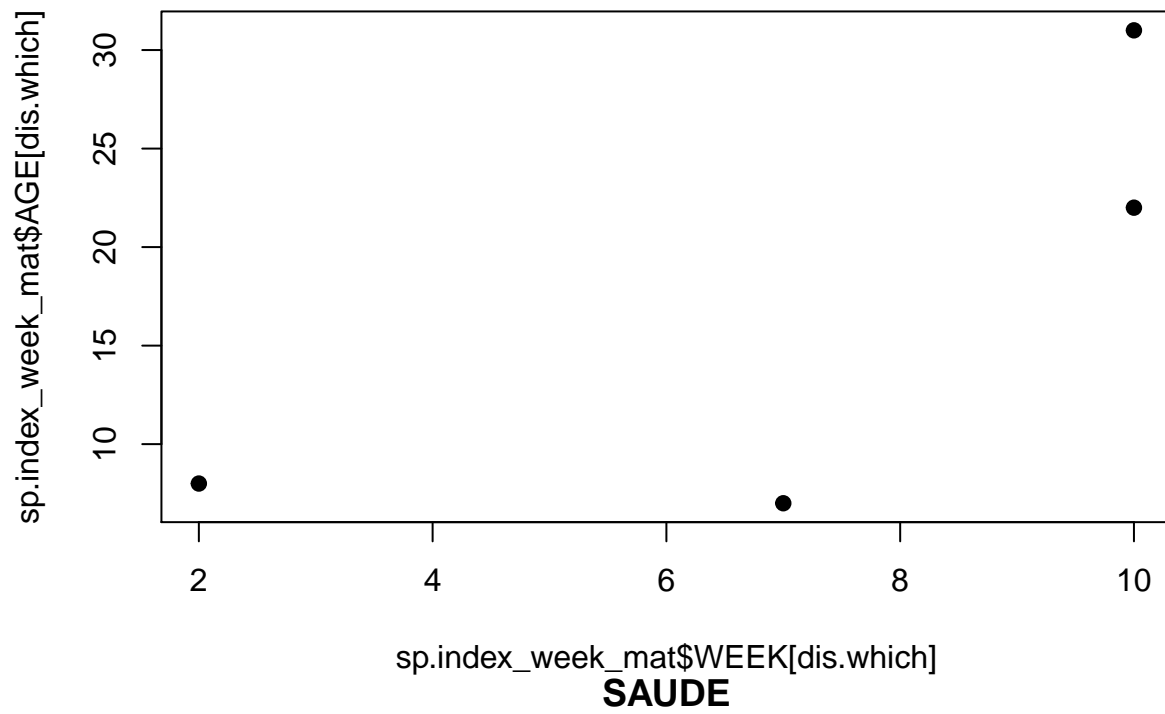
SACOMA

TREMEMBE

CIDADE ADEMAR

sp.index_week_mat$WEEK[dis.which]

RIO PEQUENO

sp.index_week_mat$WEEK[dis.which]

SANTO AMARO
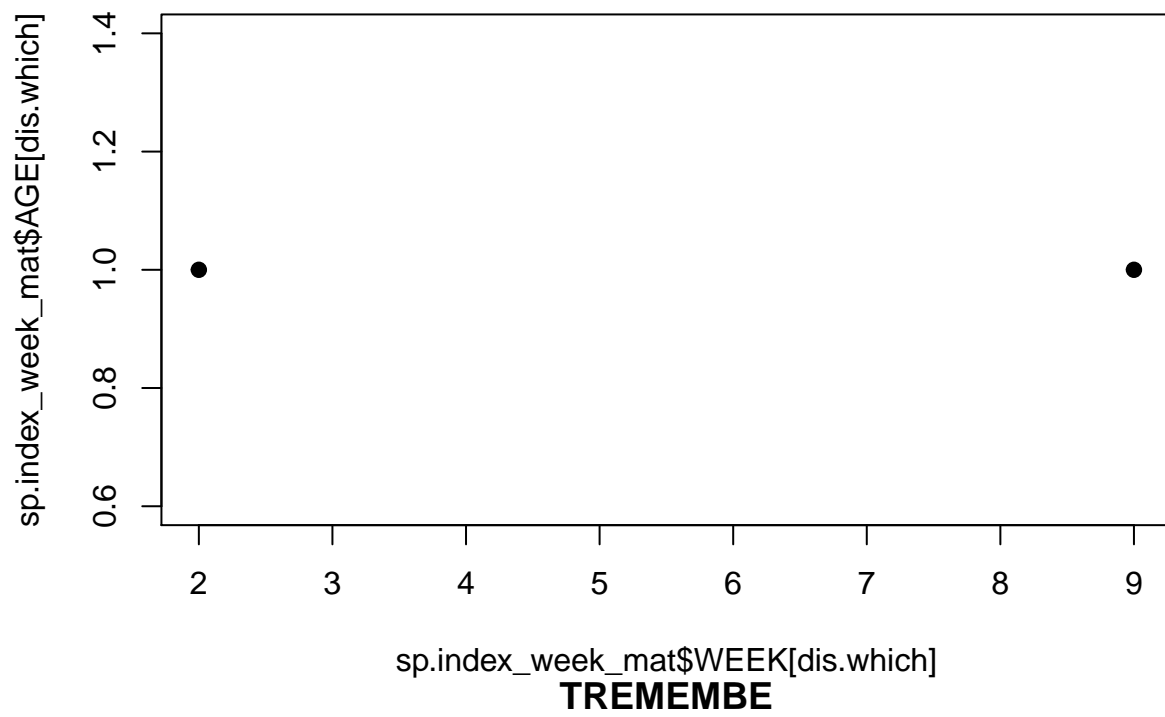
sp.index_week_mat$WEEK[dis.which]
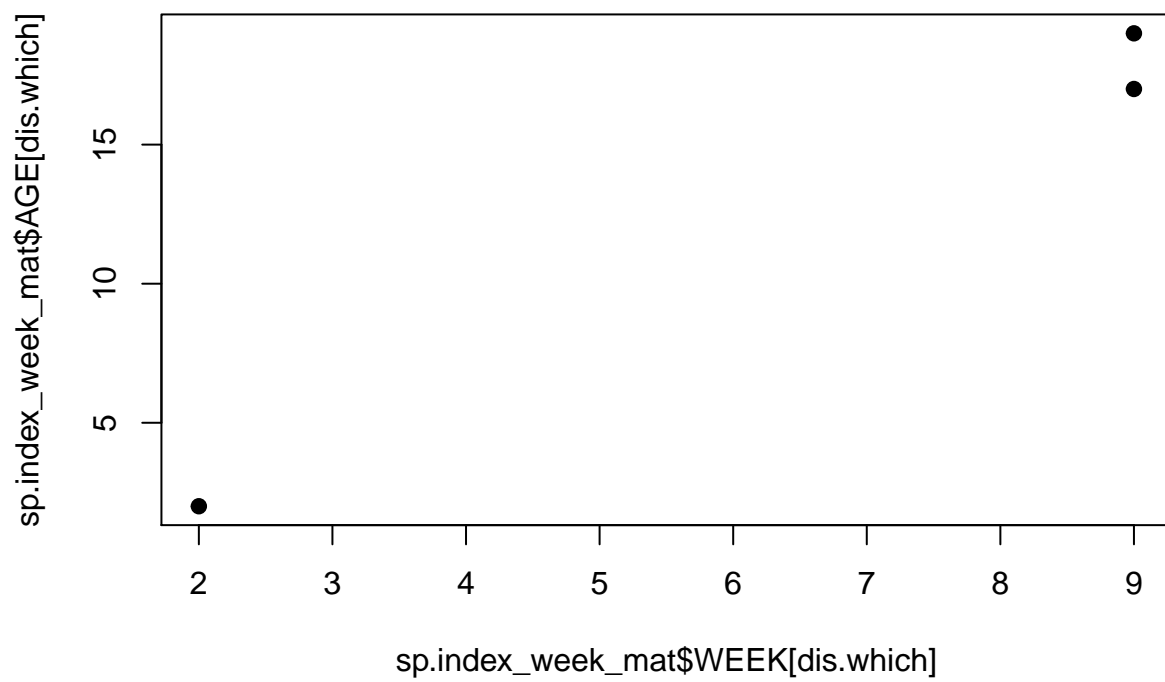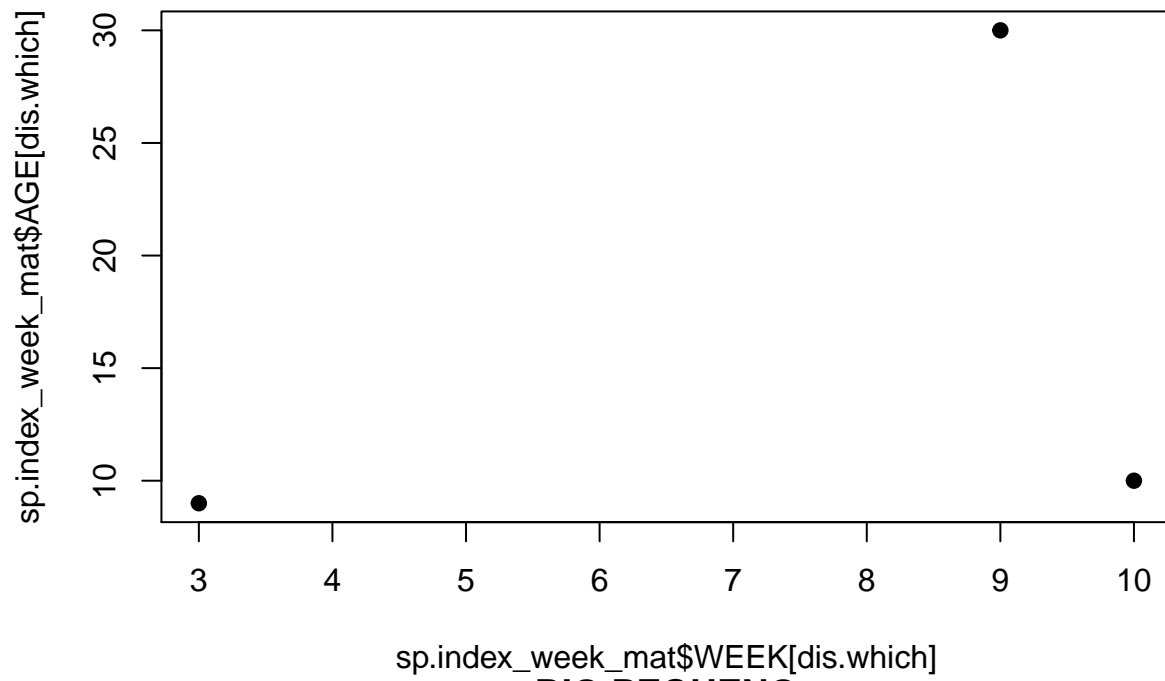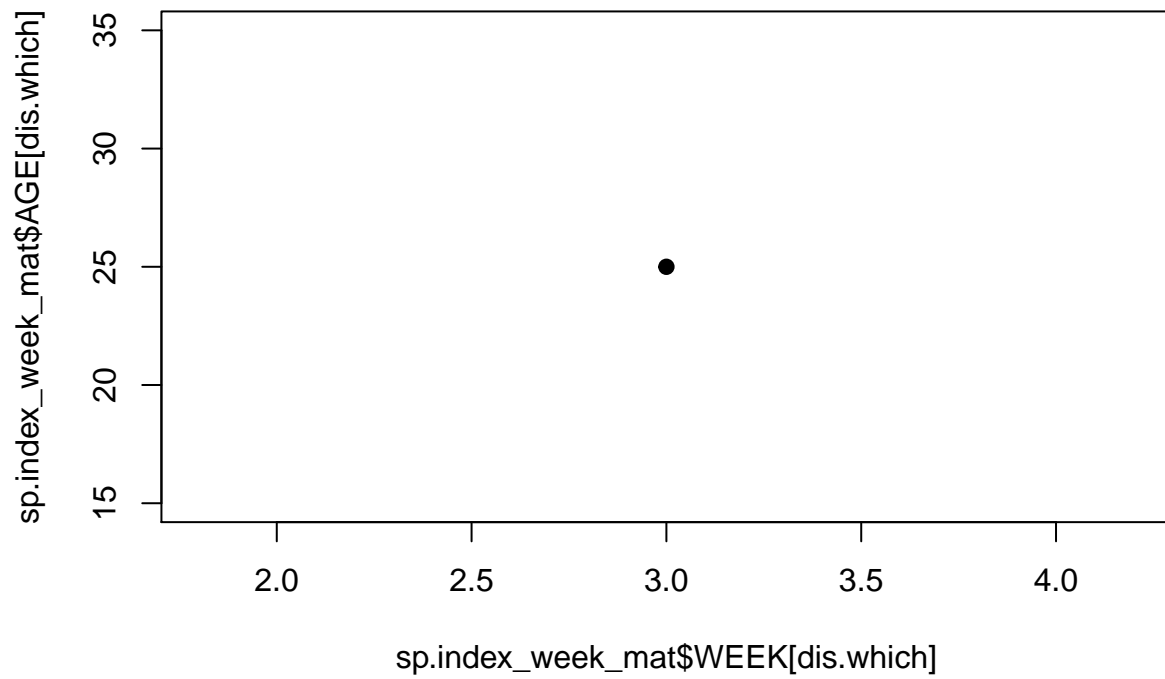
ITAQUERA

sp.index_week_mat$WEEK[dis.which]
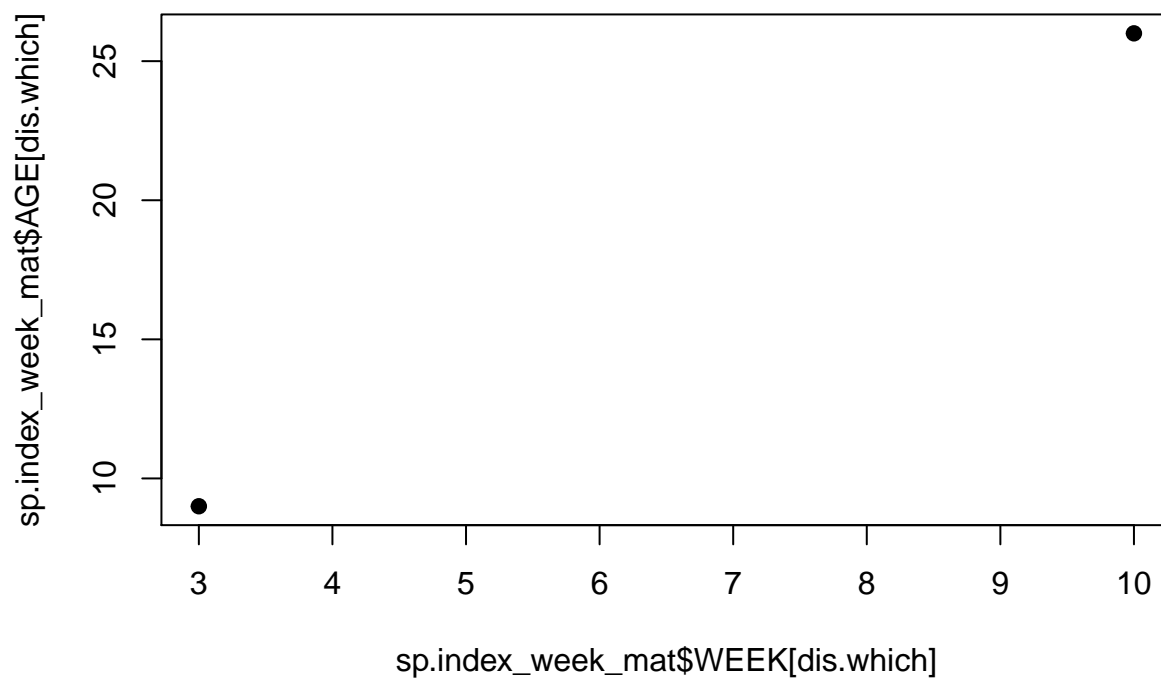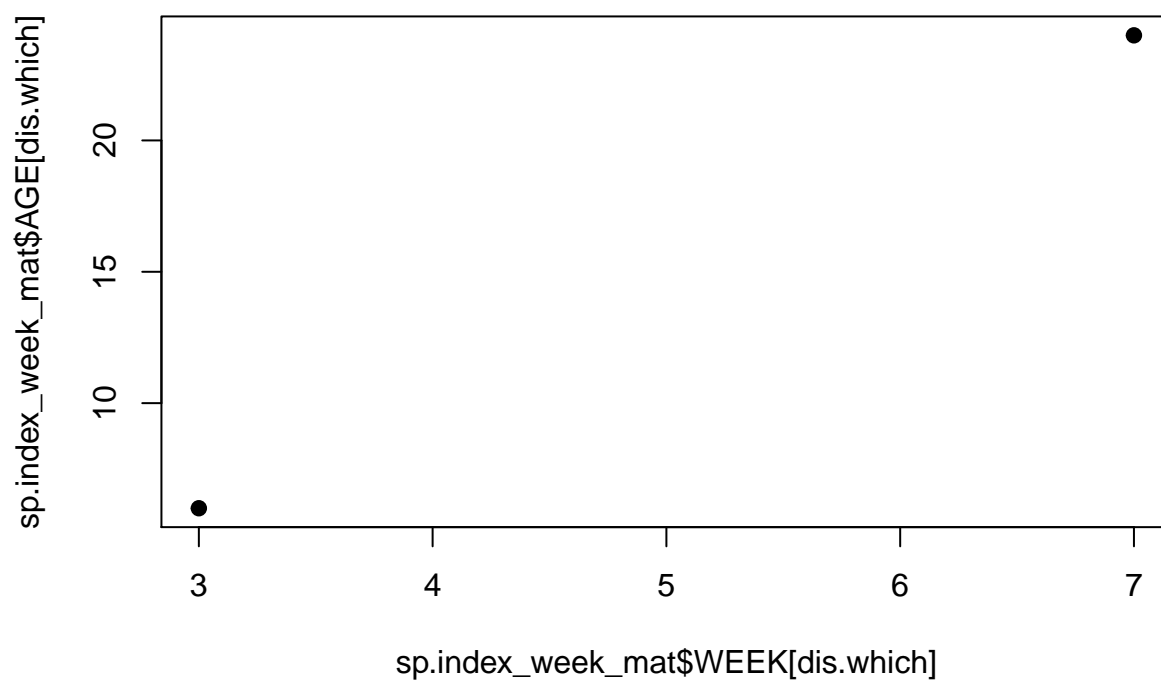
PIRITUBA
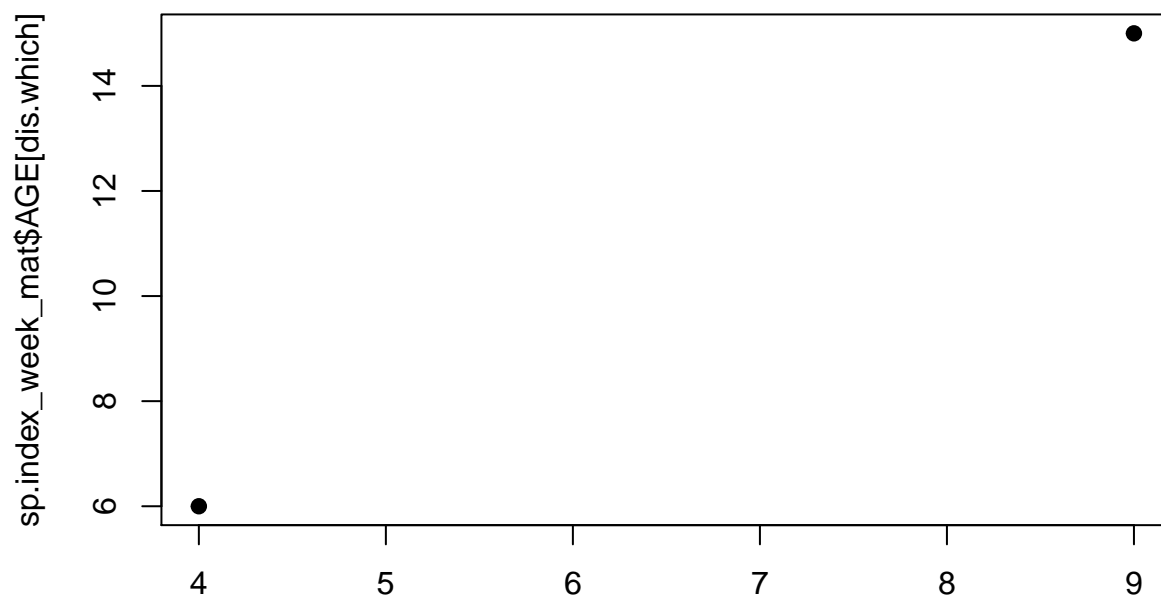
sp.index_week_mat$WEEK[dis.which]

CAMPO LIMPO

sp.index_week_mat$WEEK[dis.which]

CARRAO

sp.index_week_mat$WEEK[dis.which]
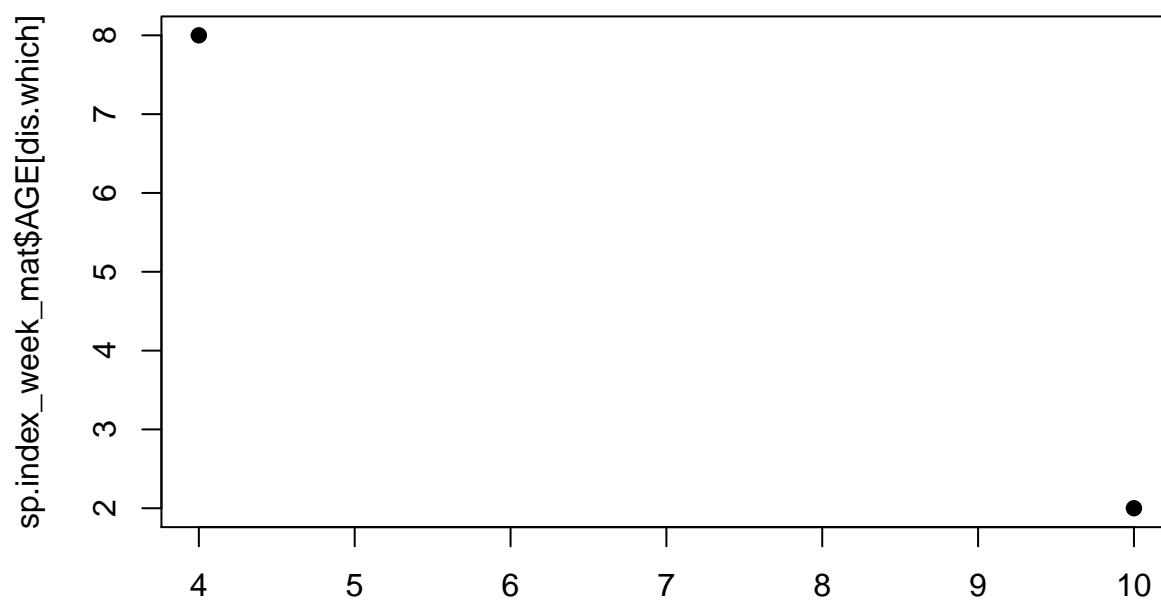
SAPOPEMBA

sp.index_week_mat$WEEK[dis.which]
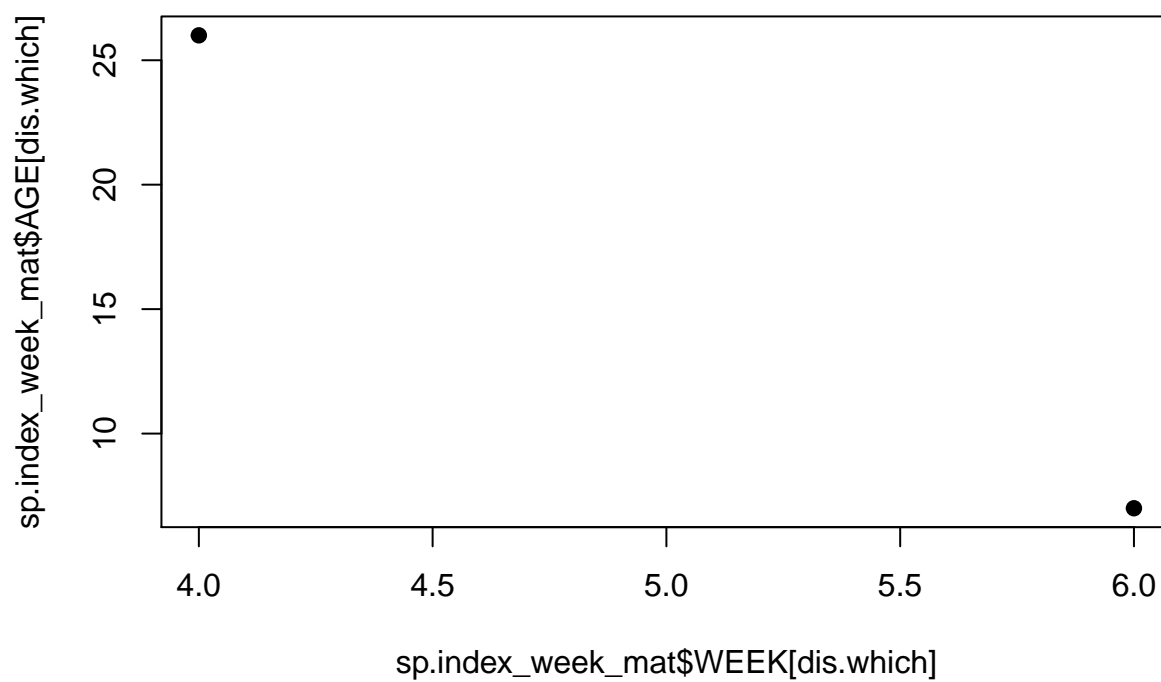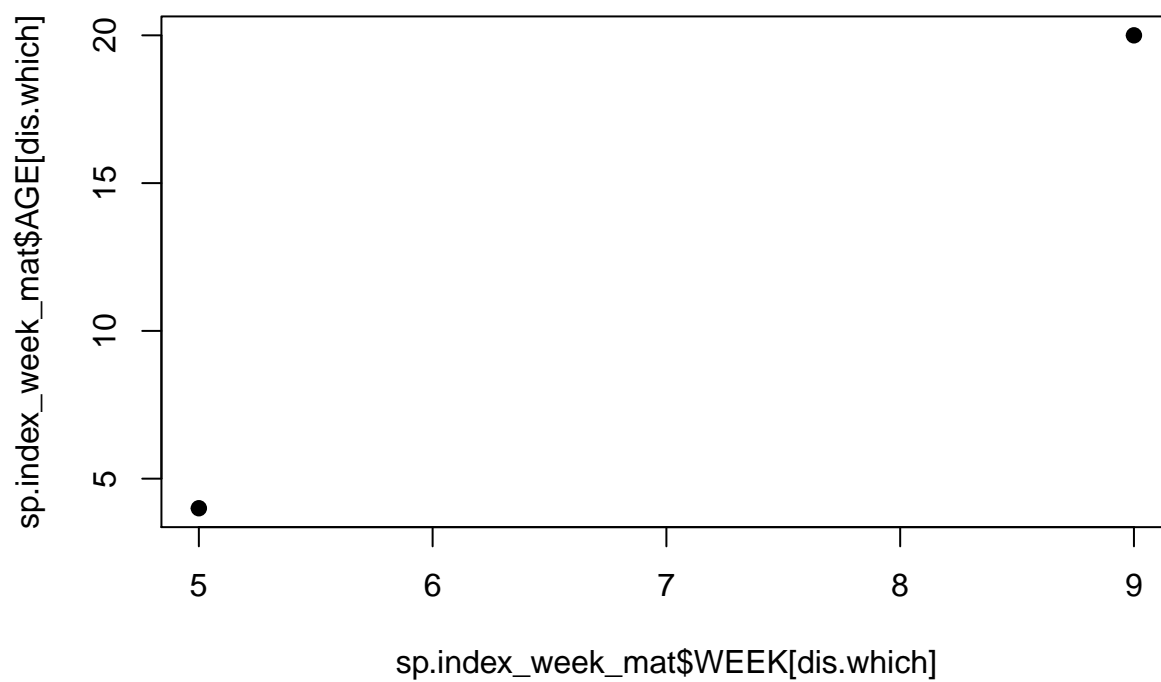
JABAQUARA
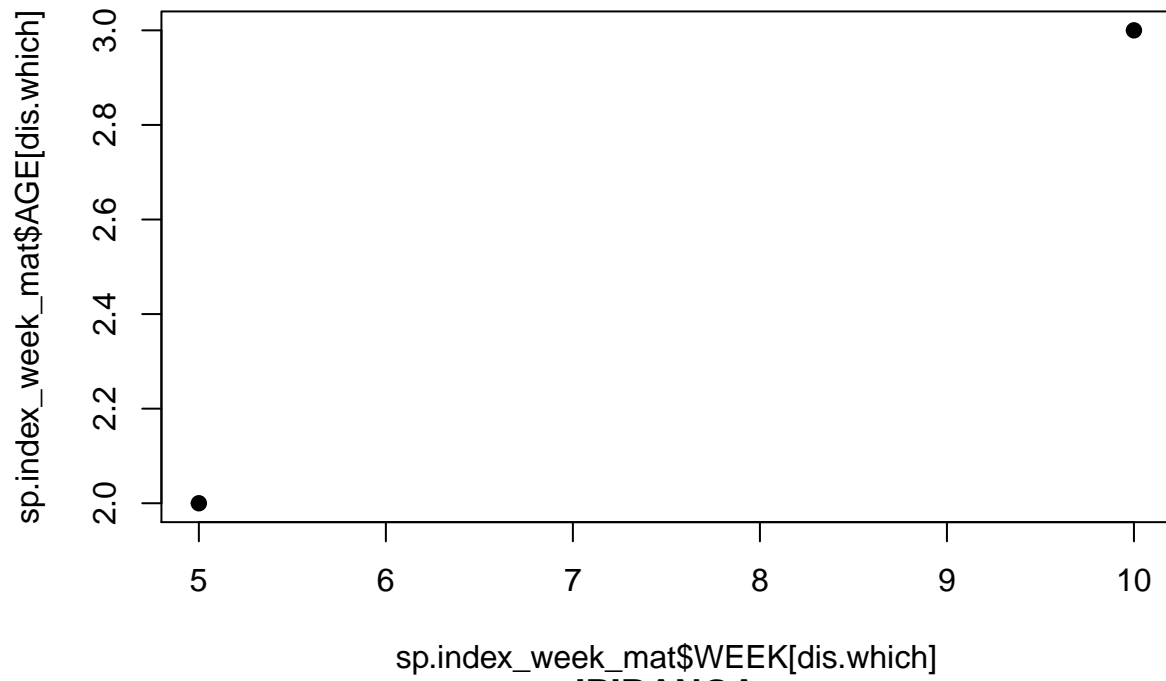
IPIRANGA
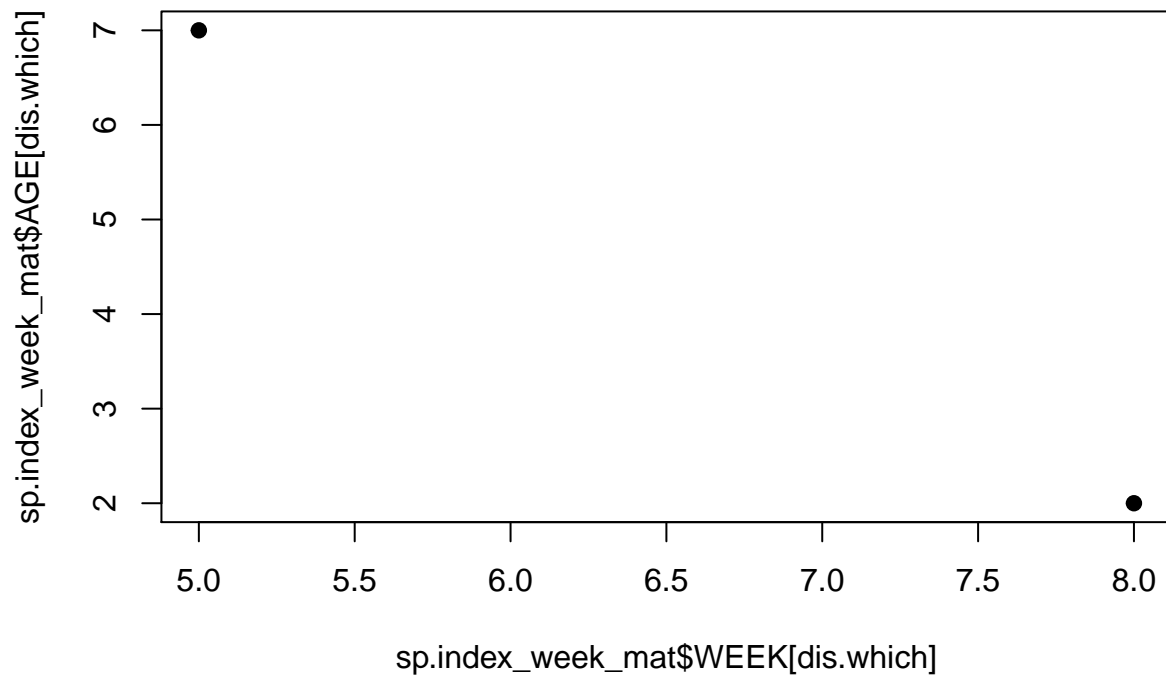
JACANA
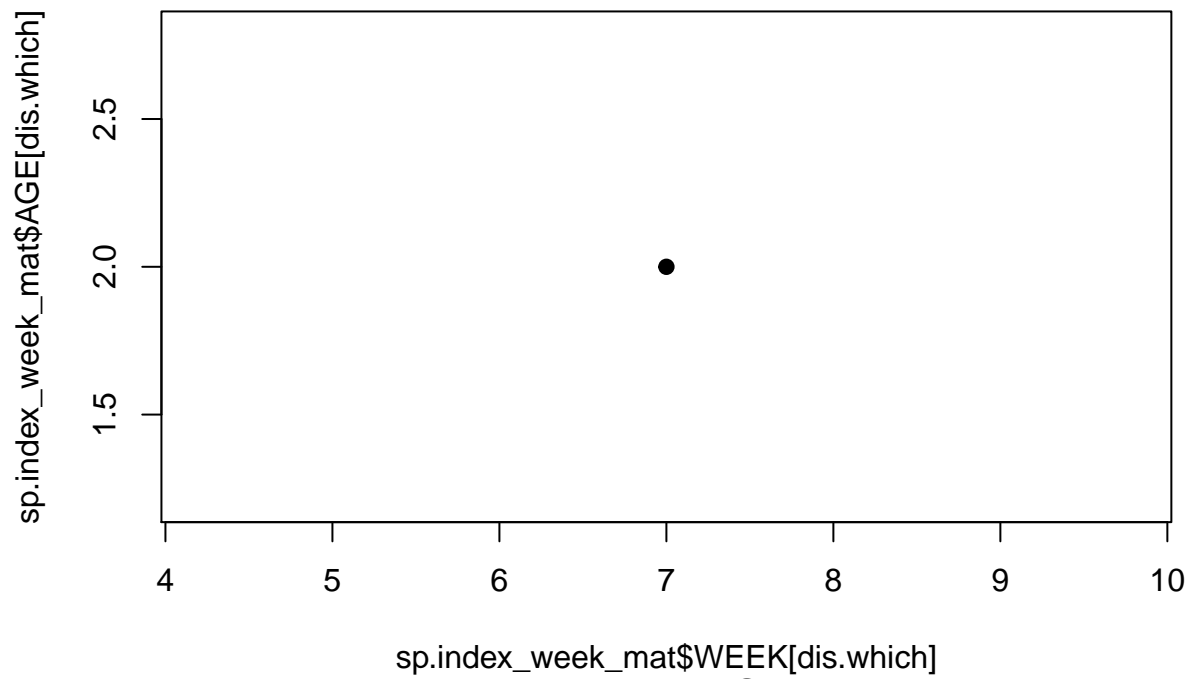
PERDIZES
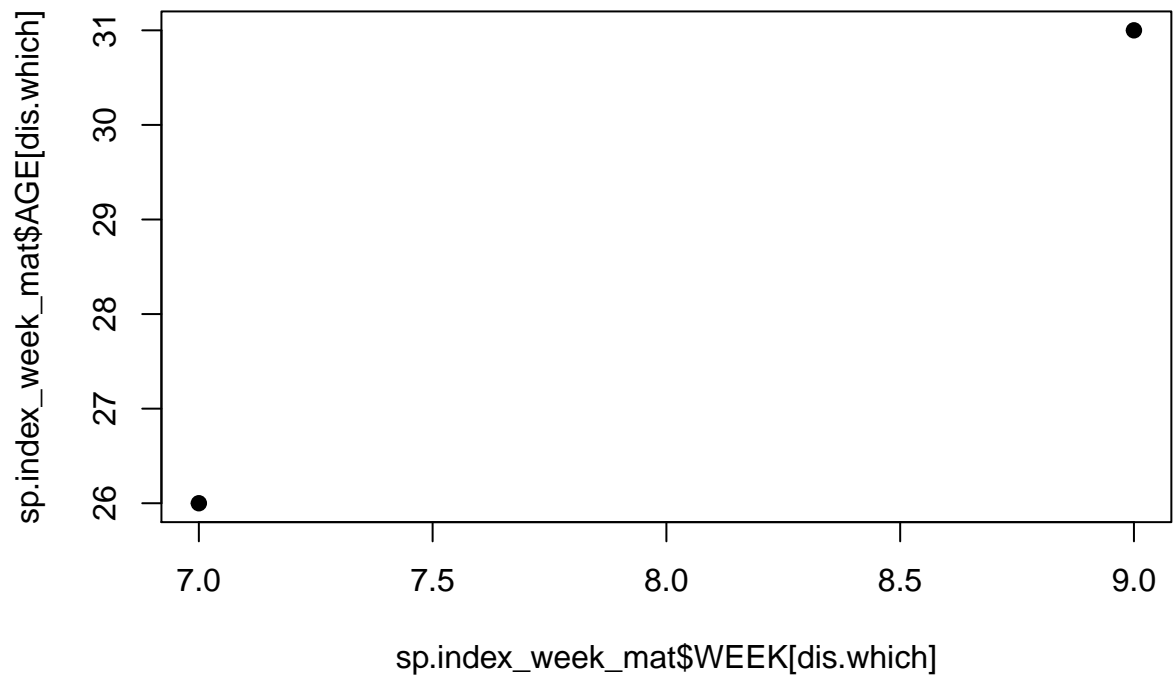
CIDADE DUTRA
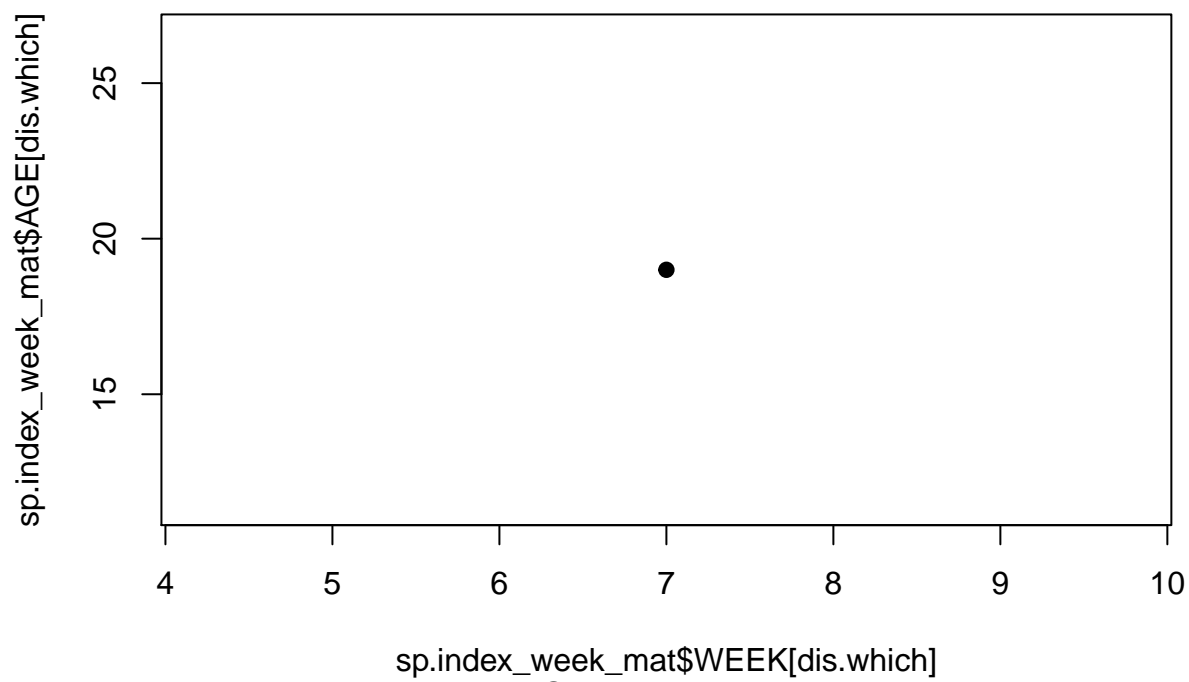
CAMPO BELO
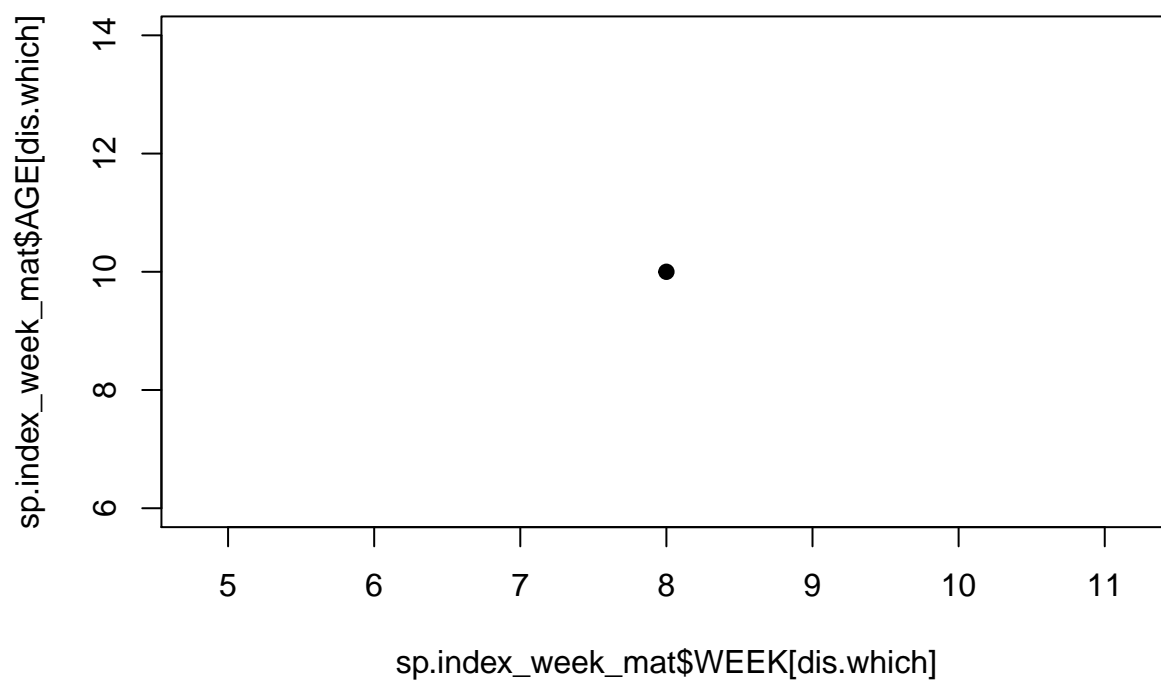
CIDADE LIDER
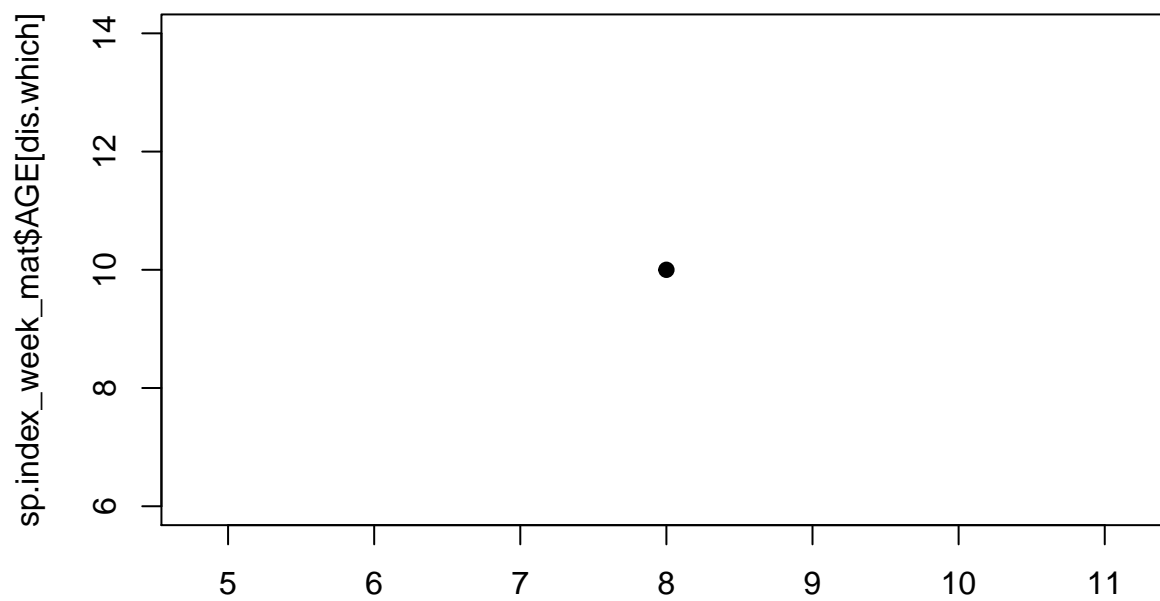
JARDIM HELENA
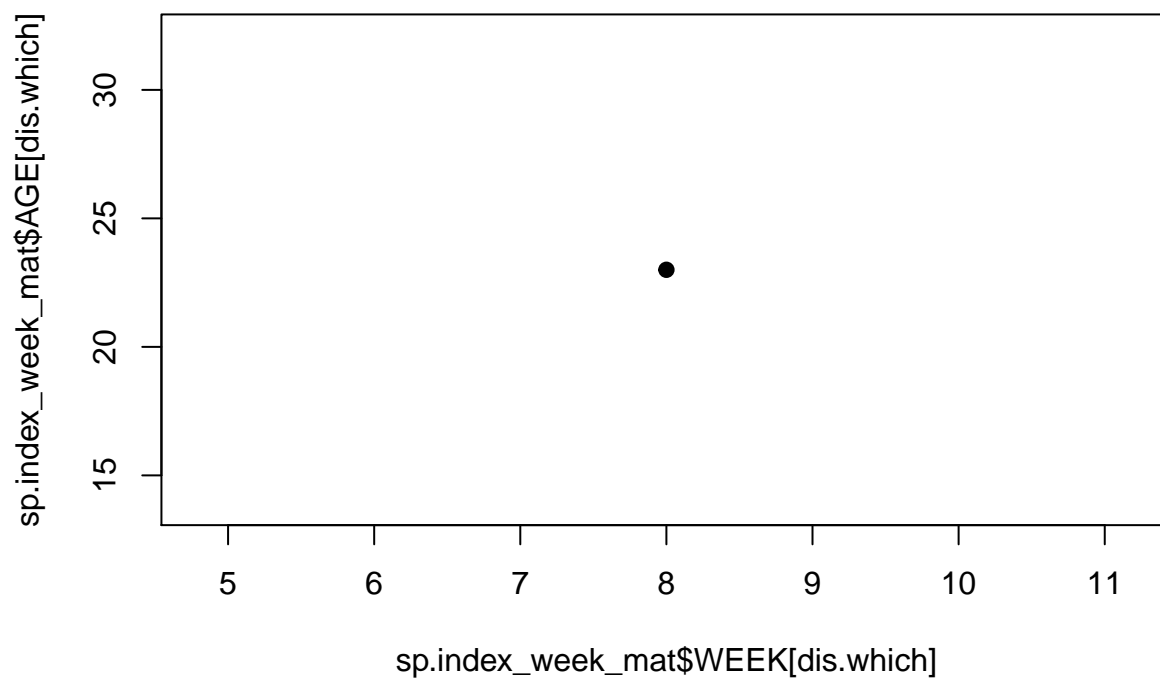
VILA MARIA

ITAIM PAULISTA

MOOCA

sp.index_week_mat$AGE[dis.which]

sp.index_week_mat$WEEK[dis.which]

SAO MATEUS

sp.index_week_mat$AGE[dis.which]

sp.index_week_mat$WEEK[dis.which]
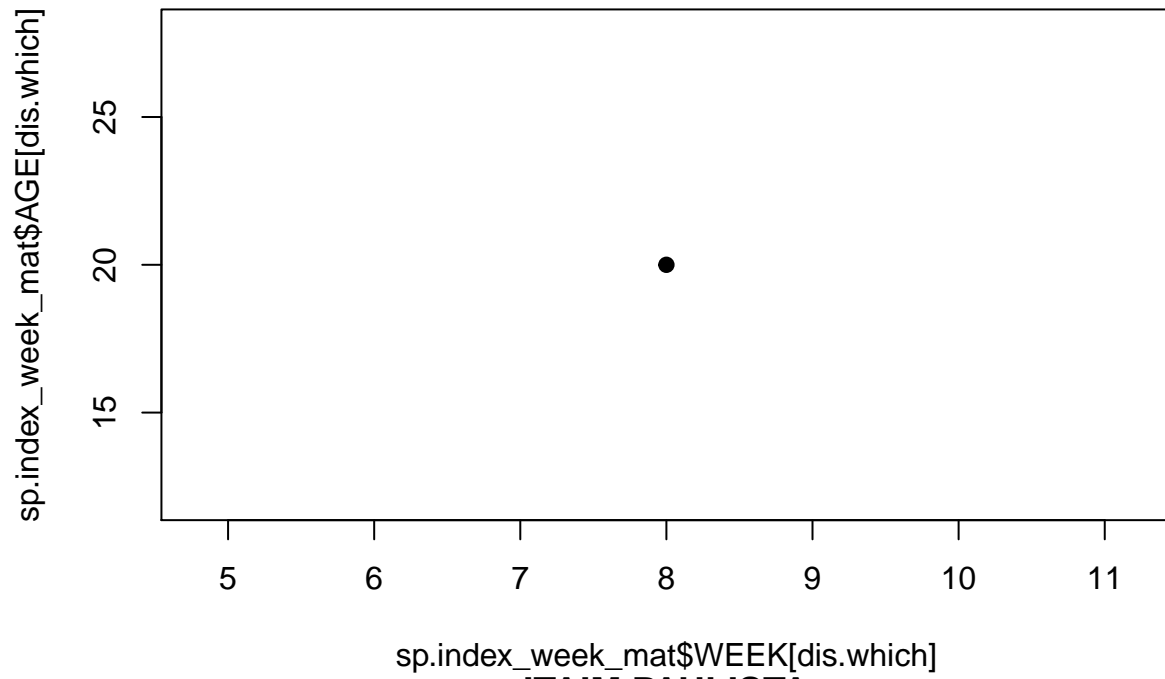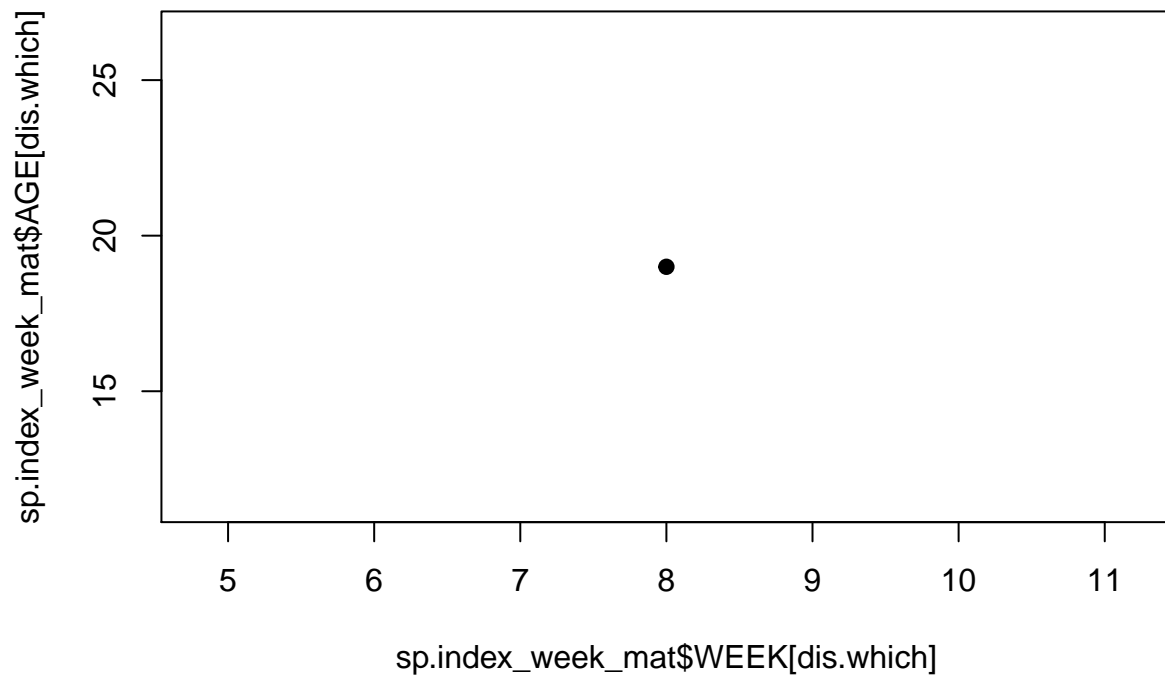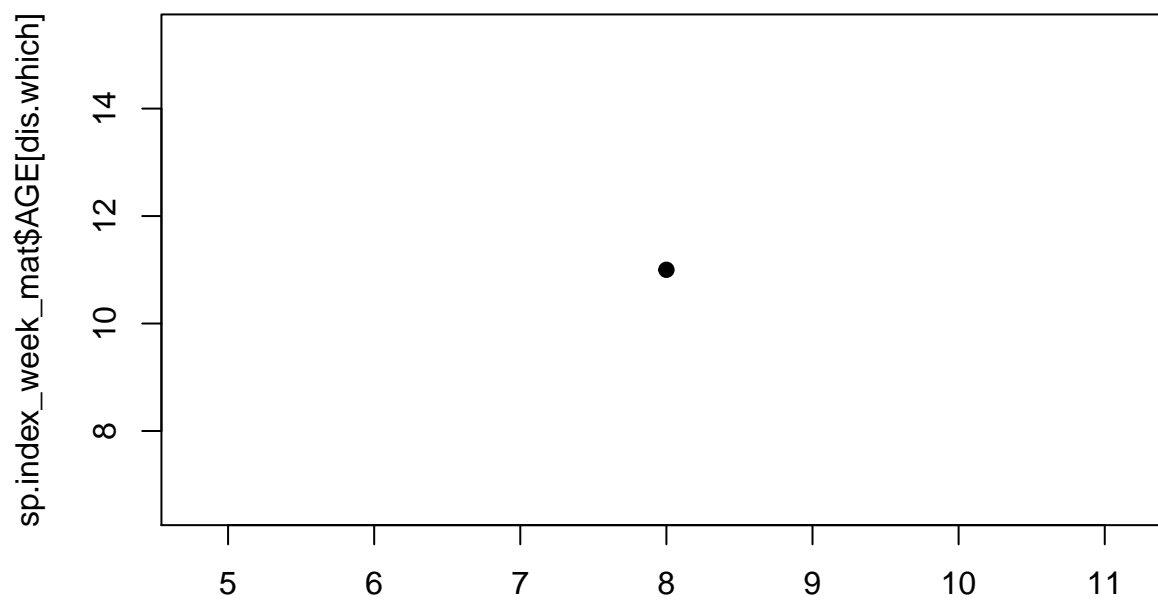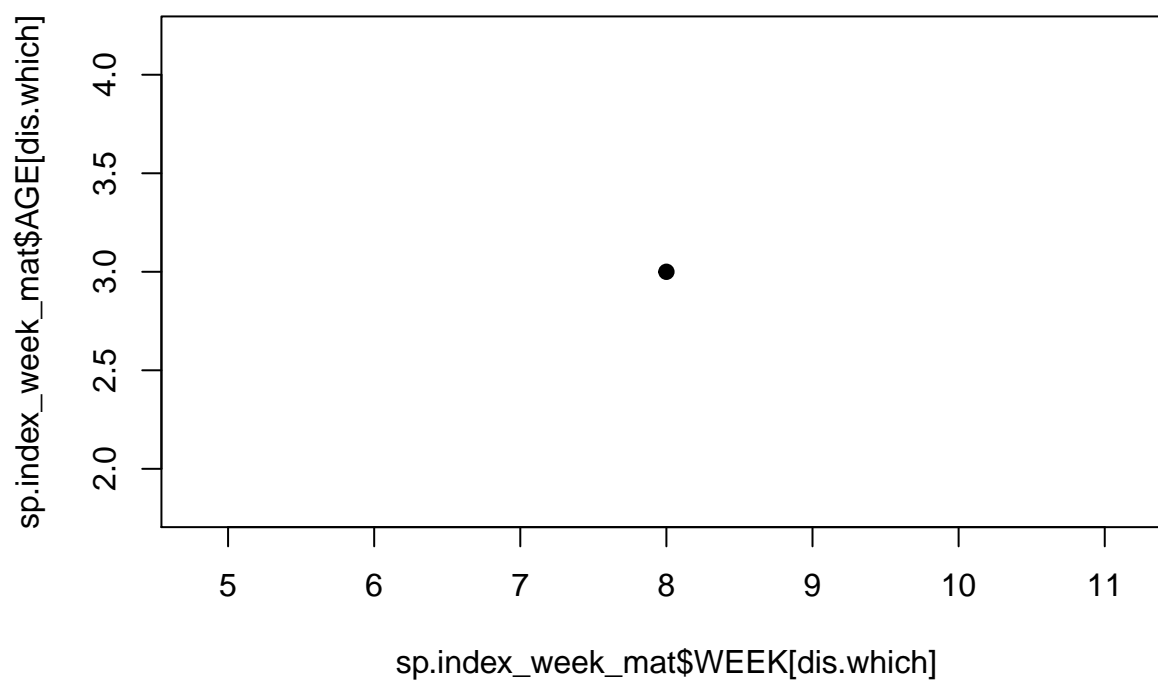
SANTANA
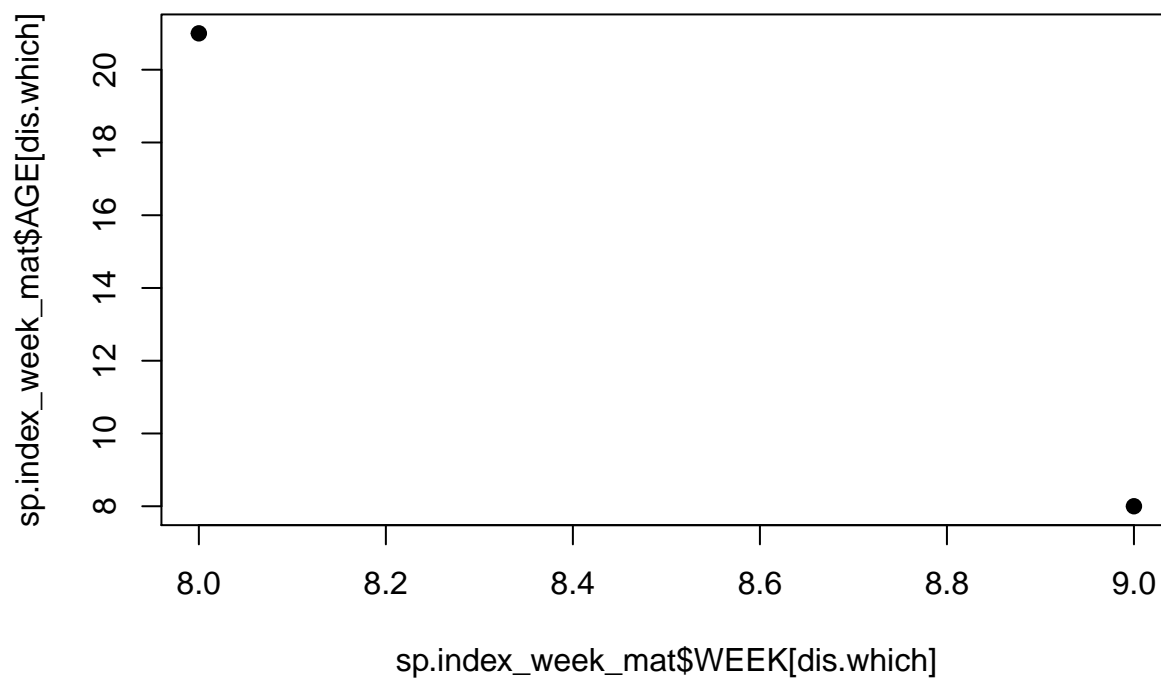
VILA PRUDENTE
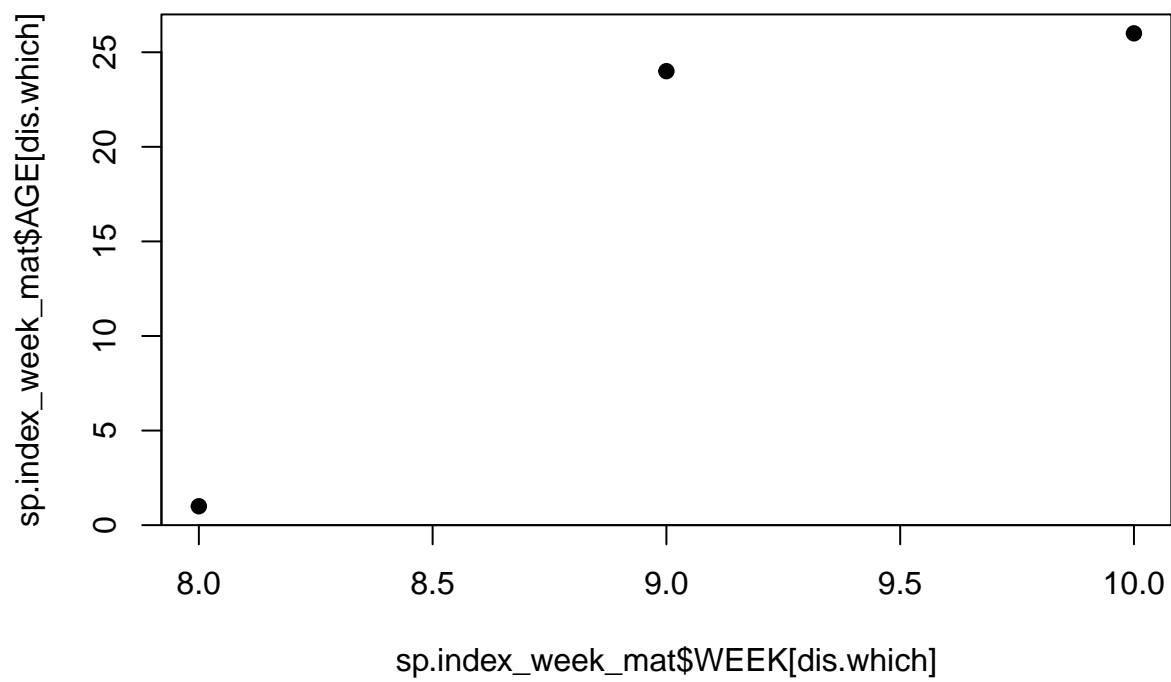
JARDIM ANGELA
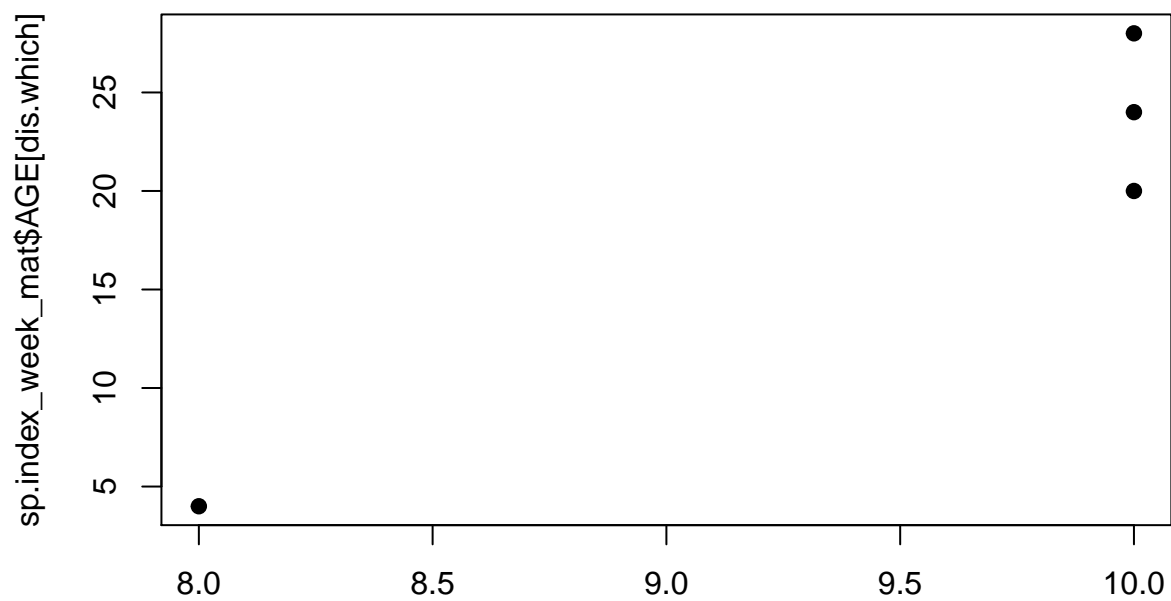
JOSE BONIFACIO
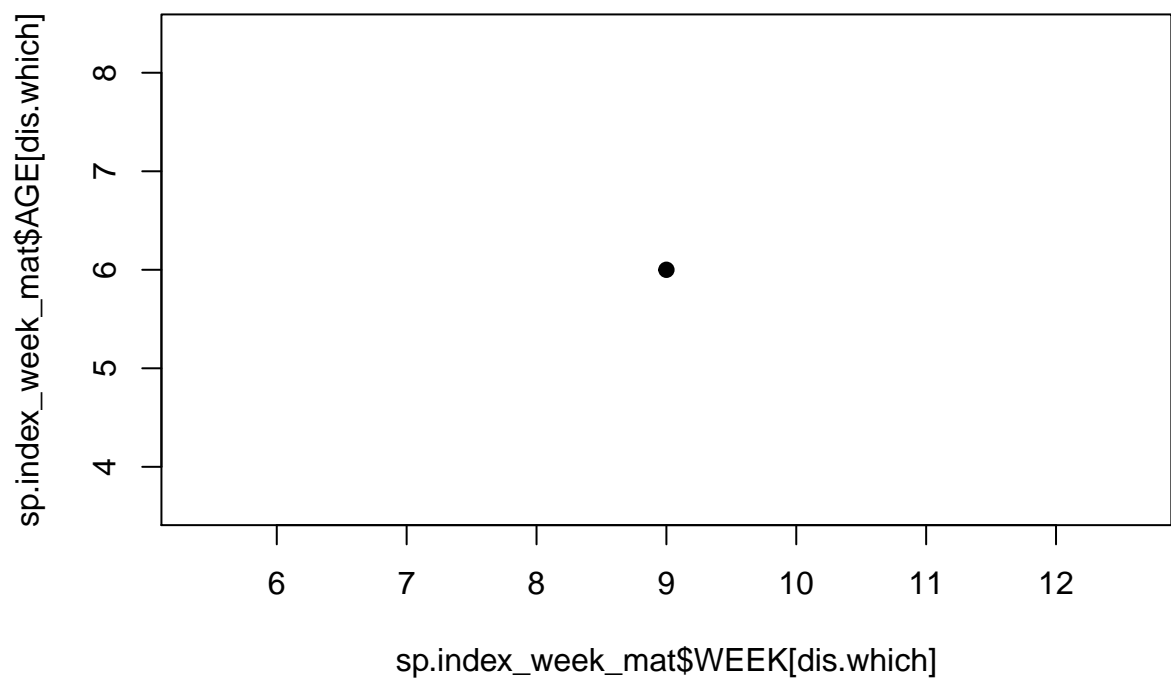
GRAJAU
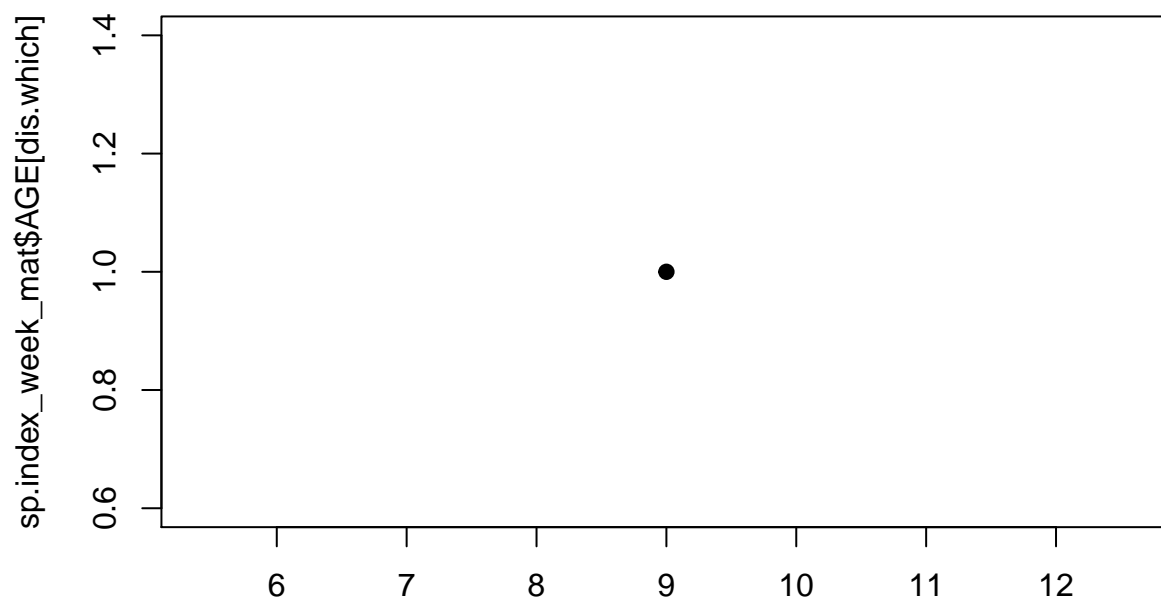
BELA VISTA
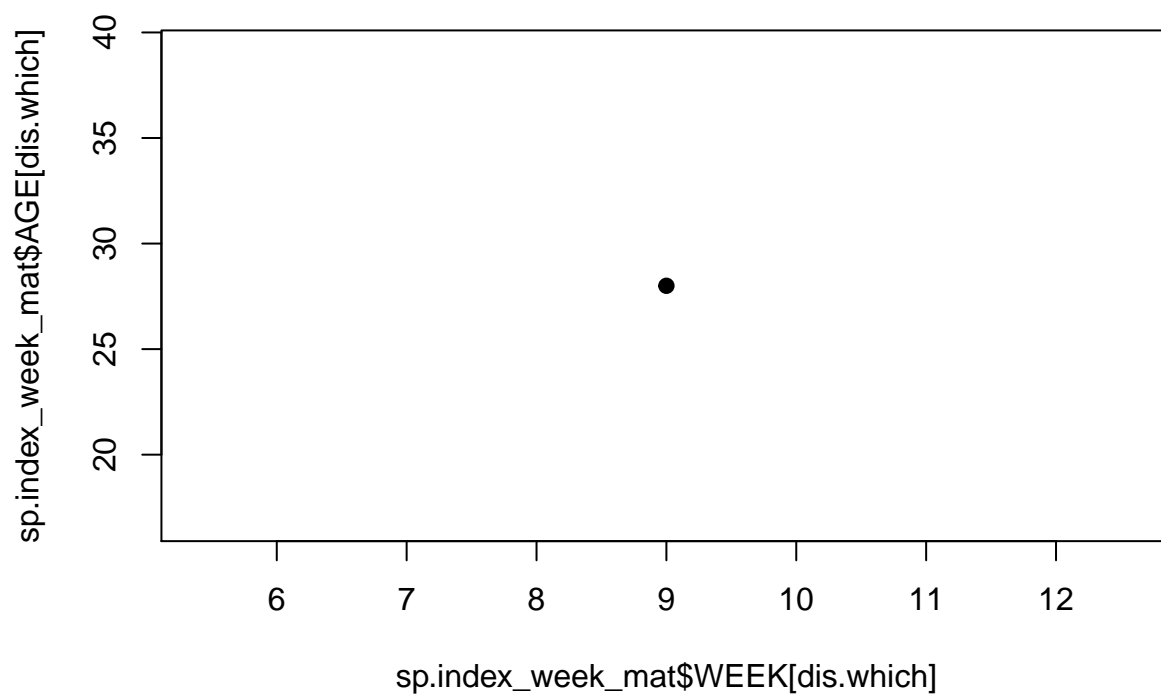
JARAGUA
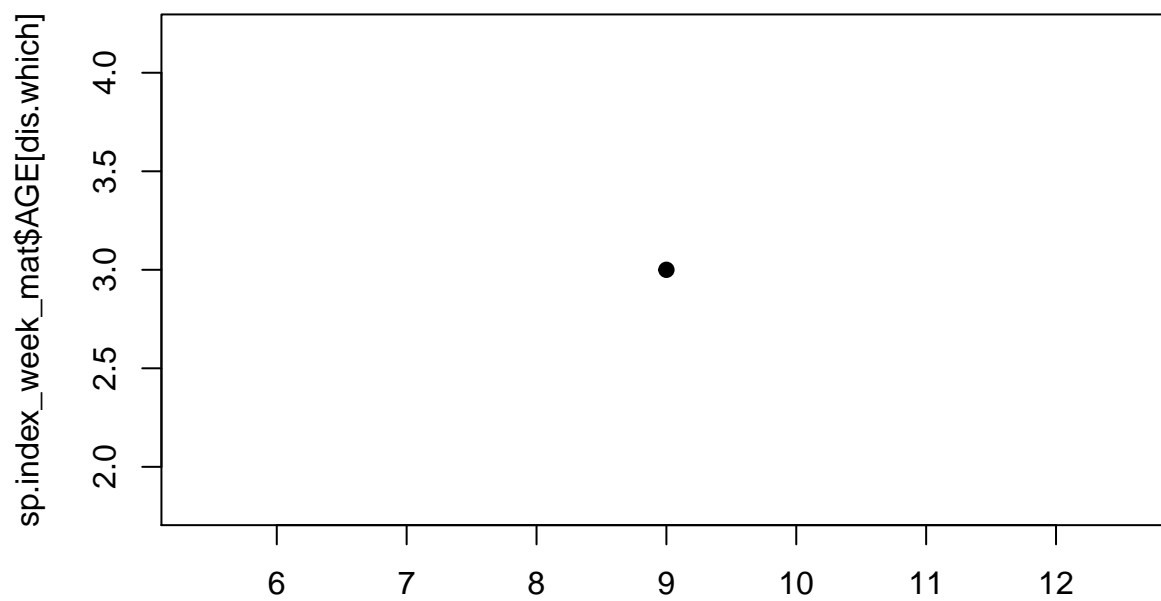
CAPAO REDONDO
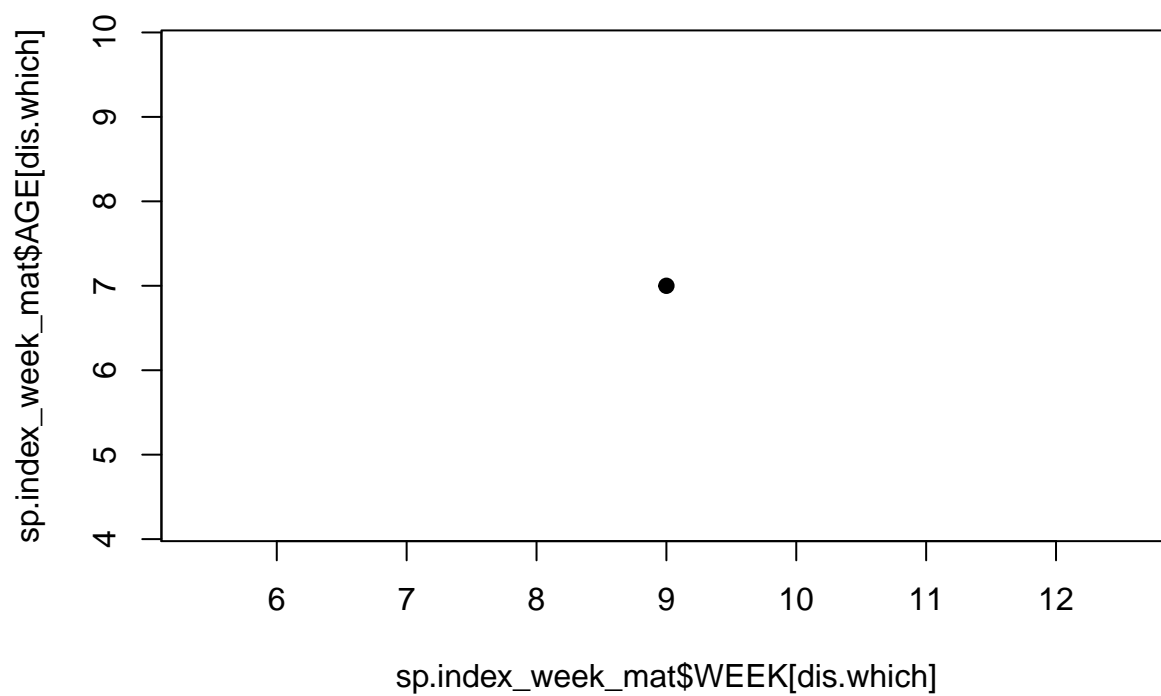
GUAIANASES
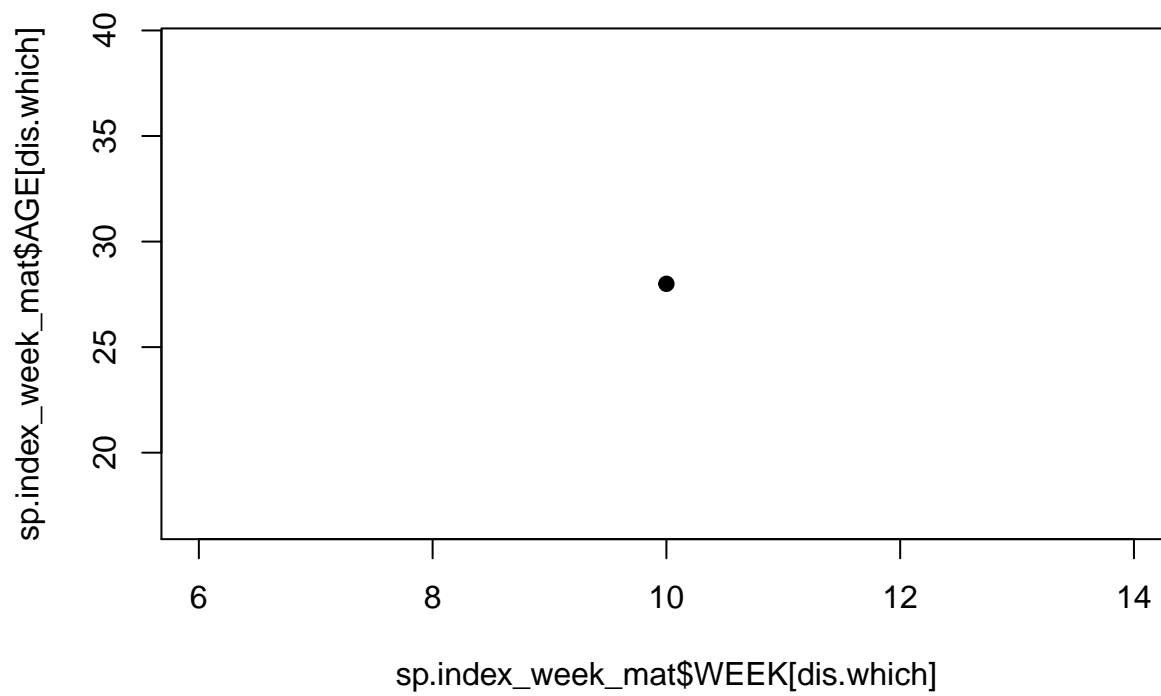
LAJEADO
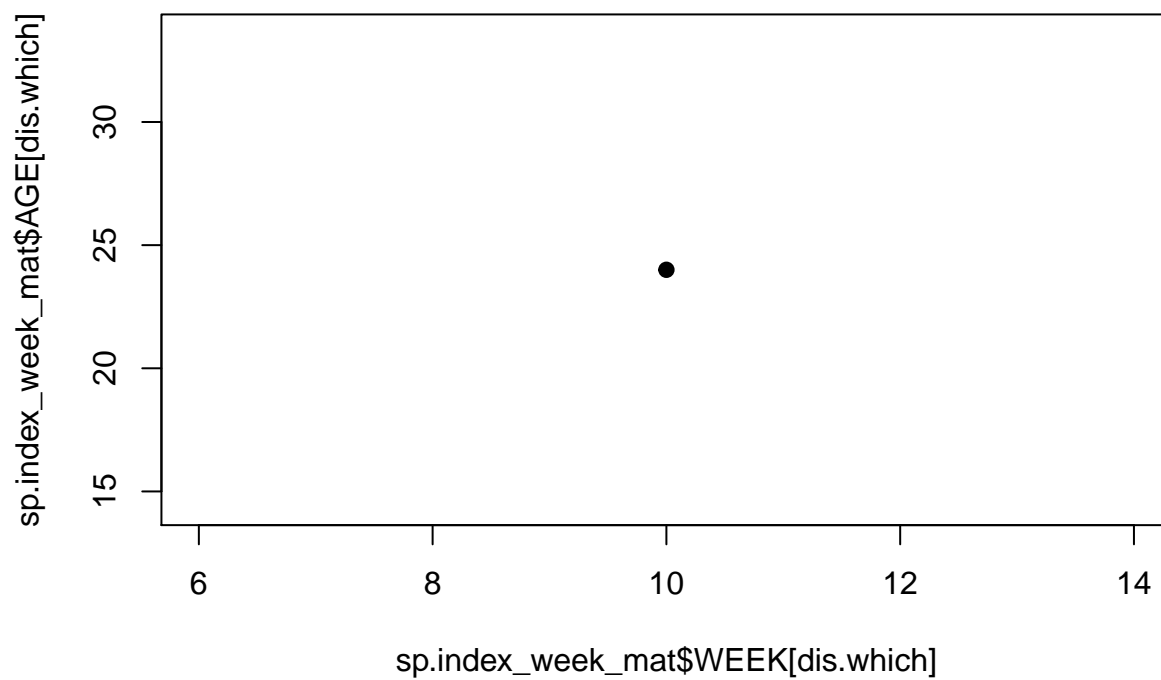
PINHEIROS
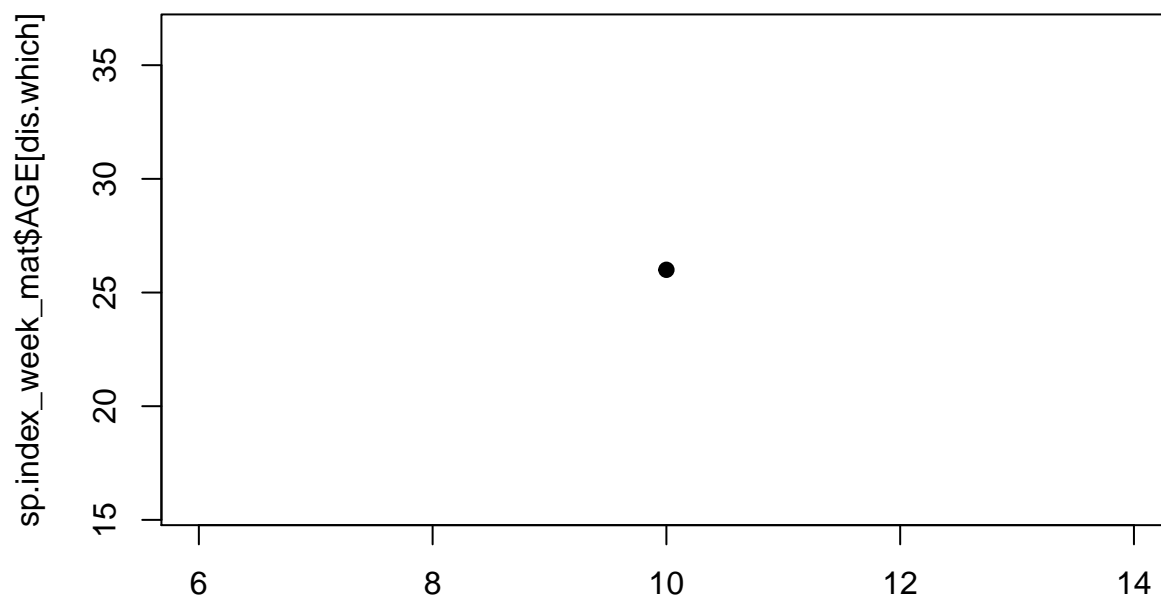
CASA VERDE
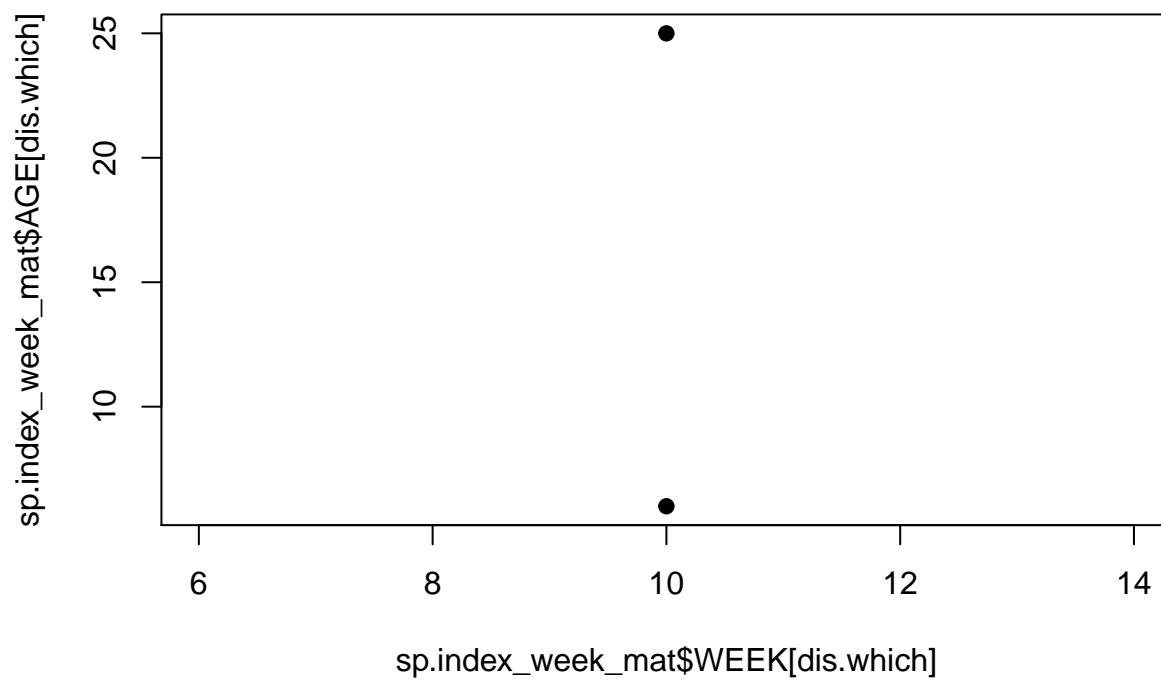
MOEMA
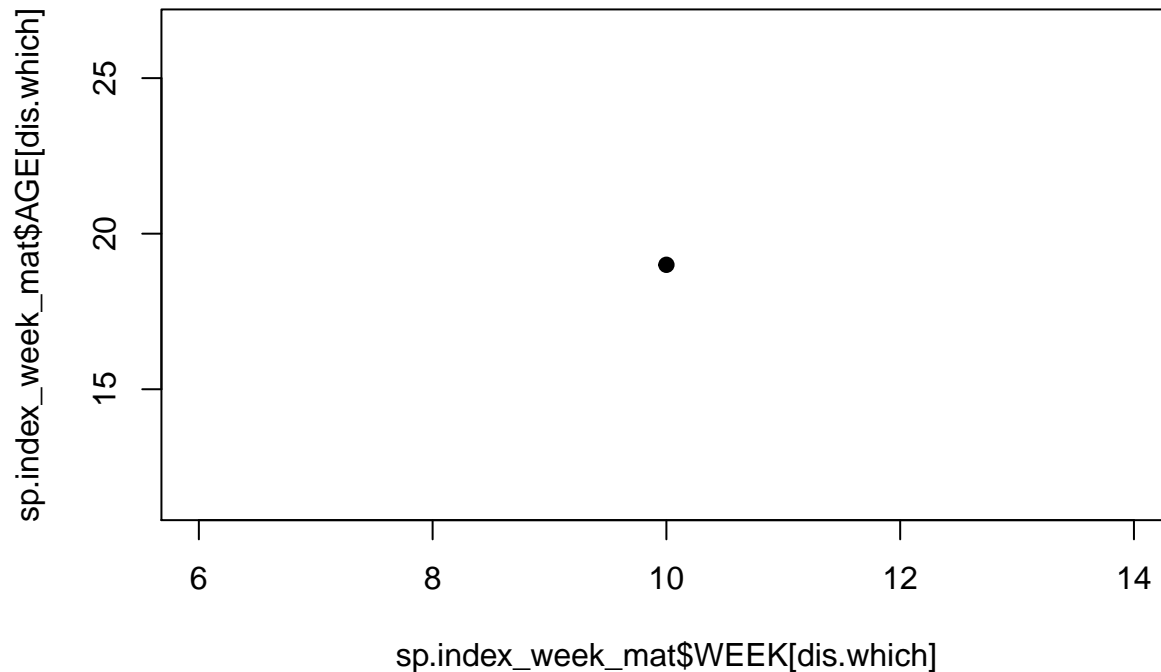
Preliminary analysis:

It does not seem that analysis via examination by district is going to be very informative. District is only available for Sao Paulo county, and on examination of the first one hundred cases by district for Sao Paulo county, there is not obvious trend.

Stage of epidemic in each county during each week of the epidemic

```
status_mat <- matrix(rep(NA, length(county_vec) * 53), ncol = length(county_vec))
colnames(status_mat) <- county_vec
rownames(status_mat) <- seq(1:53)

for(i in 1:length(county_vec)){
  for(j in 1:53){
    x <- subset(data, data$COUNTY == county_vec[i] & data$WEEK == j)
    status_mat[j,i] <- nrow(x)
  }
}

##contracting from someone in the same district
#i = county_vec[i]
#j = week
active_cases <- function(i, j){
  day_one <- 1 * status_mat[j-1,i] + 1 * status_mat[j-2,i] + (6/7) * status_mat[j-3,i]
  day_two <- 1 * status_mat[j-1,i] + 1 * status_mat[j-2,i] + (5/7) * status_mat[j-3,i] + (1/7) * status_
  day_three <- 1 * status_mat[j-1,i] + 1 * status_mat[j-2,i] + (4/7) * status_mat[j-3,i] + (2/7) * stat
  day_four <- 1 * status_mat[j-1,i] + 1 * status_mat[j-2,i] + (3/7) * status_mat[j-3,i] + (3/7) * status
```

```
  day_five <- 1 * status_mat[j-1,i] + 1 * status_mat[j-2,i] + (2/7) * status_mat[j-3,i] + (4/7) * status
  day_six <- 1 * status_mat[j-1,i] + 1 * status_mat[j-2,i] + (1/7) * status_mat[j-3,i] + (5/7) * status
  day_seven <- 1 * status_mat[j-1,i] + 1 * status_mat[j-2,i] + (6/7) * status_mat[j,i]
  days <- c(day_one, day_two, day_three, day_four, day_five, day_six, day_seven)
return(days)
}
```

```
install.packages("ggmap")
```

```
## Installing package into '/Users/maggiewalters/Library/R/3.3/library'
## (as 'lib' is unspecified)
```

```
## Warning: unable to access index for repository YOUR FAVORITE MIRROR/src/contrib:
##   scheme not supported in URL 'YOUR FAVORITE MIRROR/src/contrib/PACKAGES'
```

```
## Warning: package 'ggmap' is not available (for R version 3.3.3)
```

```
## Warning: unable to access index for repository YOUR FAVORITE MIRROR/bin/macosx/mavericks/contrib/3.3
##   scheme not supported in URL 'YOUR FAVORITE MIRROR/bin/macosx/mavericks/contrib/3.3/PACKAGES'
```

```
library(ggmap)
```

```
## Loading required package: ggplot2
```

```
distance <- function(to, from){
  distance <- mapdist(from = from, to = to)
  return(distance$km)
}
```

- Created the function **active_cases(i, j)**, which describes the amount of infectious cases in one county (i) at one time (j in weeks). Goal is to scale this value by population size in order to represent the chance that you'd come into contact with an infected individual.
- These values would also eventually be done to describe the chance of contacting someone from another county who is infected. The chance of being infected by someone in a different county would be scaled by distance from your county to the county in question, and possible immigration rates if possible.
  - Could also use whether the area is urban or rural as a scaling factor.
  - Urban centers would be scaled lower for those in rural area because rural to urban is more likely than urban to rural.
- Also created the function **distance(to, from)** which outputs the distance from one county to another.