

101C Final EDA

Andy Shen

11/30/2020

```
rm(list = ls())
library(tidyverse)
library(BASS)
library(TBASS)

training <- read.csv("/Users/andyshen/Desktop/Git/101C-Final/training.csv")
test <- read.csv("/Users/andyshen/Desktop/Git/101C-Final/test.csv")

dates <- training$PublishedDate
head(dates)

## [1] "4/17/2020 10:38" "8/31/2020 9:56"  "8/16/2020 12:15" "8/22/2020 9:00"
## [5] "8/22/2020 14:18"  "7/24/2020 21:16"

split_dates <- strsplit(dates, "[/: ]")
head(split_dates)

## [[1]]
## [1] "4"      "17"     "2020"   "10"     "38"
##
## [[2]]
## [1] "8"      "31"     "2020"   "9"      "56"
##
## [[3]]
## [1] "8"      "16"     "2020"   "12"     "15"
##
## [[4]]
## [1] "8"      "22"     "2020"   "9"      "00"
##
## [[5]]
## [1] "8"      "22"     "2020"   "14"     "18"
##
## [[6]]
## [1] "7"      "24"     "2020"   "21"     "16"

split_dates <- lapply(split_dates, as.numeric)
years <- function(date) {
  date[3]
}

yrs <- lapply(split_dates, years)
unique(yrs) # Only 2020

## [[1]]
## [1] 2020
```

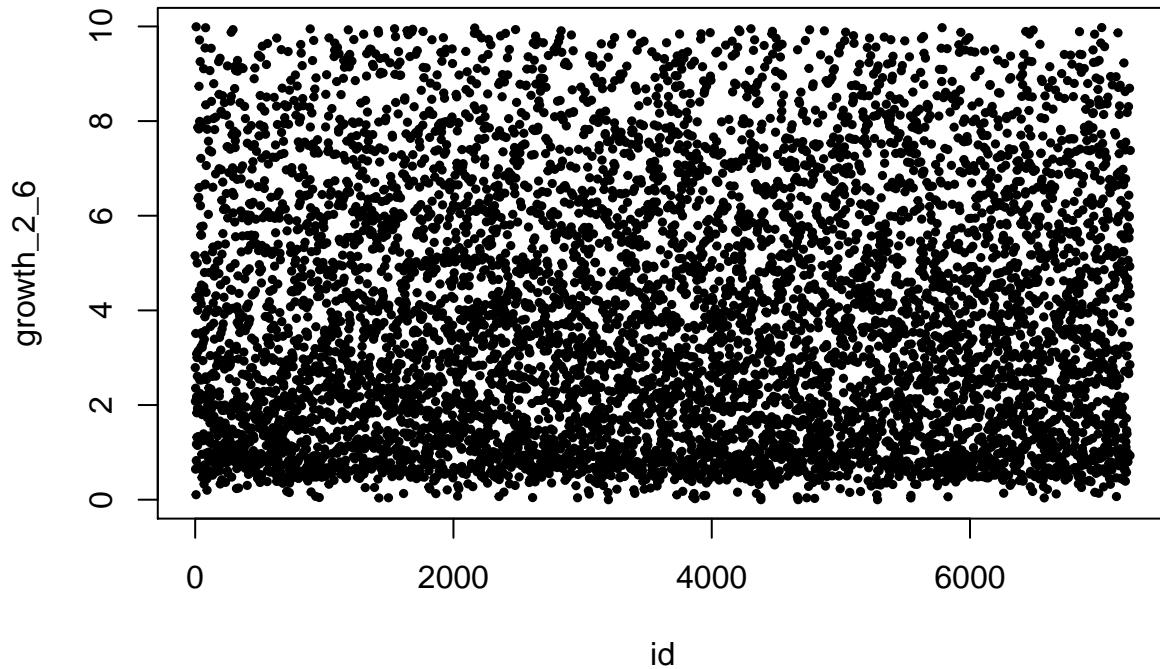
```

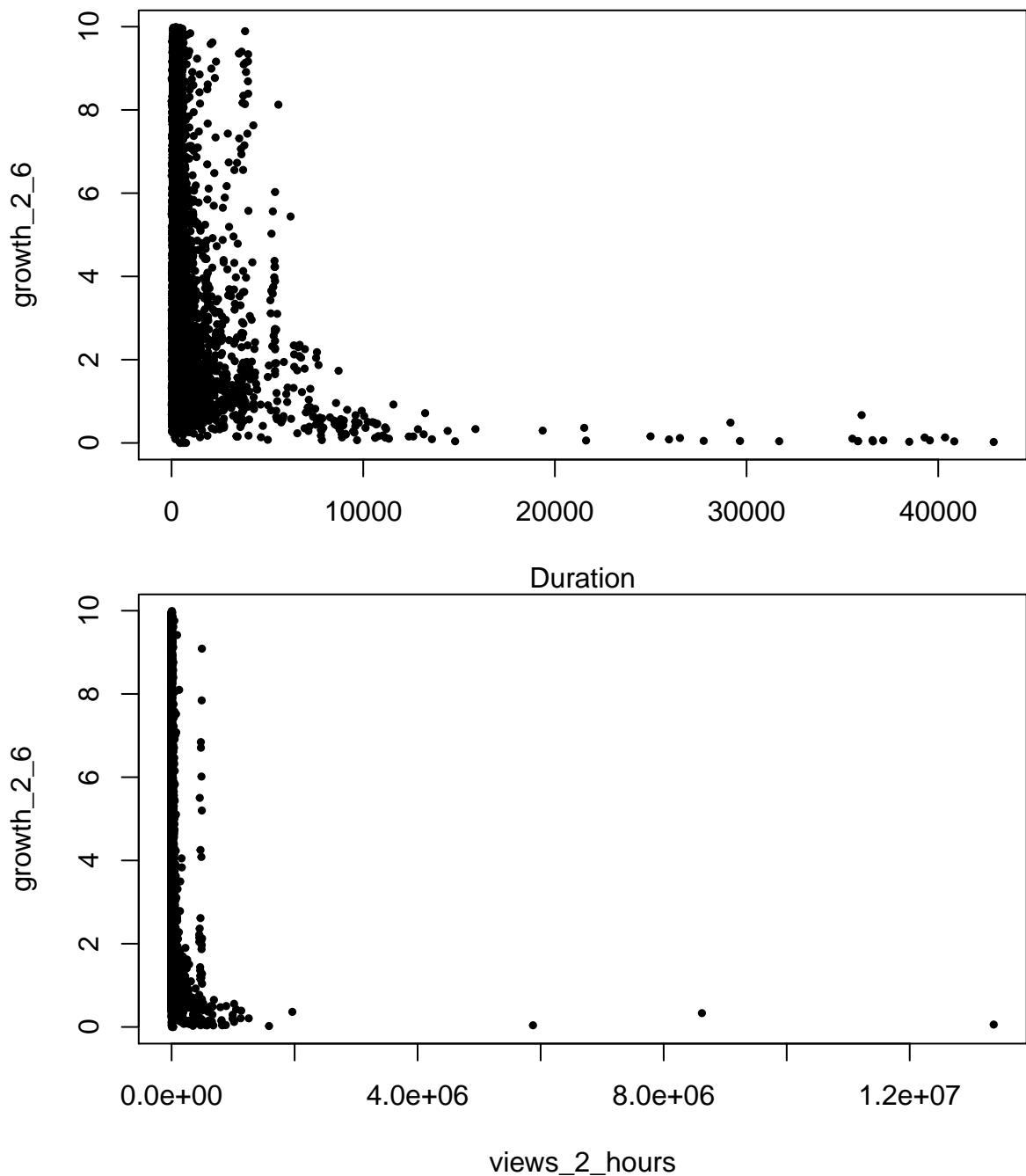
new_time <- function(old_date) {
  old_date <- as.numeric(old_date)
  month <- old_date[1]
  day <- old_date[2]
  hr <- old_date[4]
  minute <- old_date[5]
  month_days <- c(31, 29, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31)
  complete_months <- month - 1
  if (complete_months > 0) {
    complete_month_days <- sum(month_days[1:complete_months])
  } else {
    complete_month_days <- 0
  }
  total_days <- complete_month_days + day
  pre_hrs <- total_days * 24
  total_hrs <- pre_hrs + hr
  pre_min <- total_hrs * 60
  final_time <- pre_min + minute
  final_time
}
processed_dates <- vapply(split_dates, new_time, numeric(1))
training$PublishedDate <- processed_dates

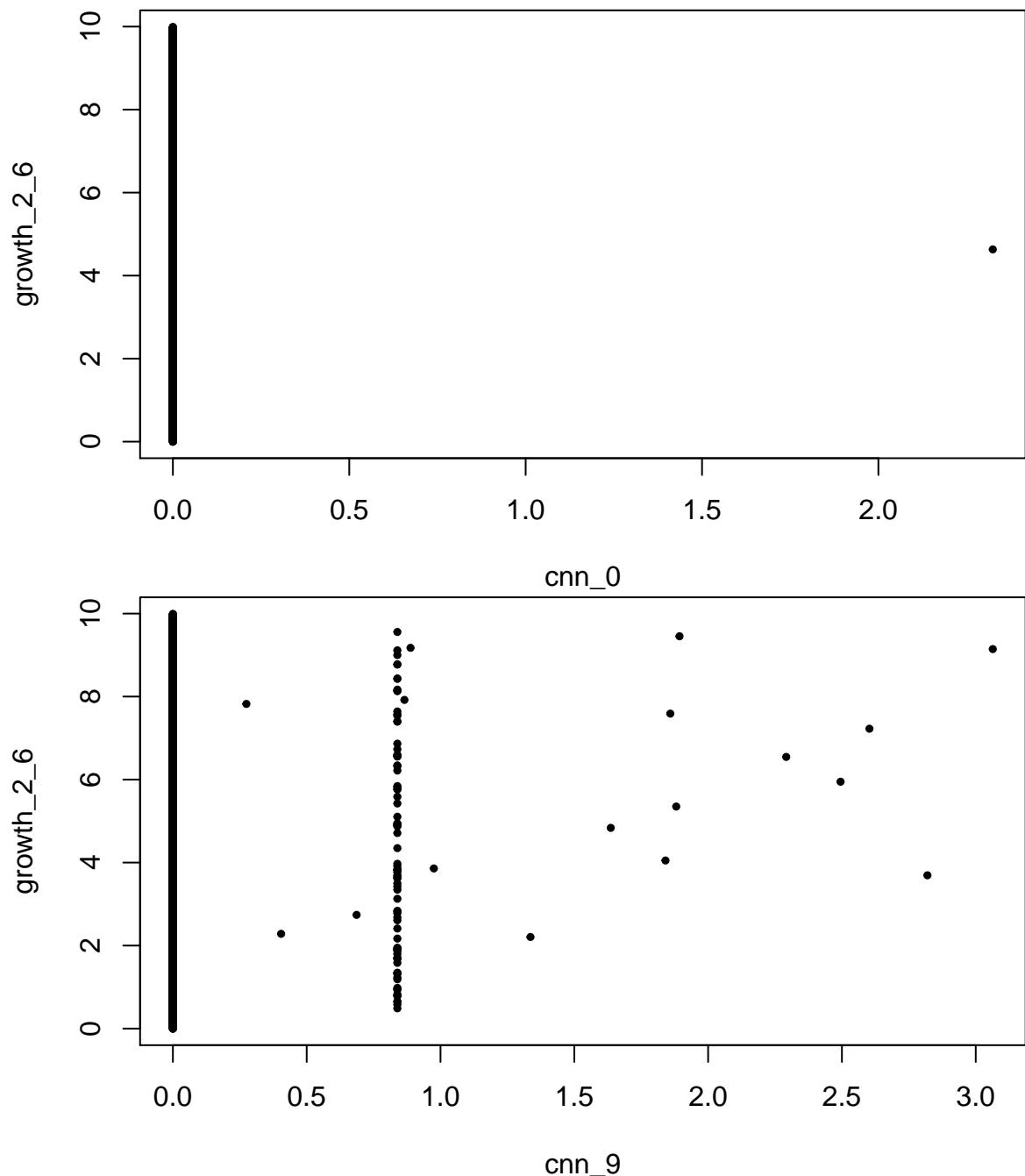
ind <- sample(1:nrow(training), size = floor(0.8 * nrow(training)), replace = F)
train <- training[ind,]
ttest <- training[-ind,]

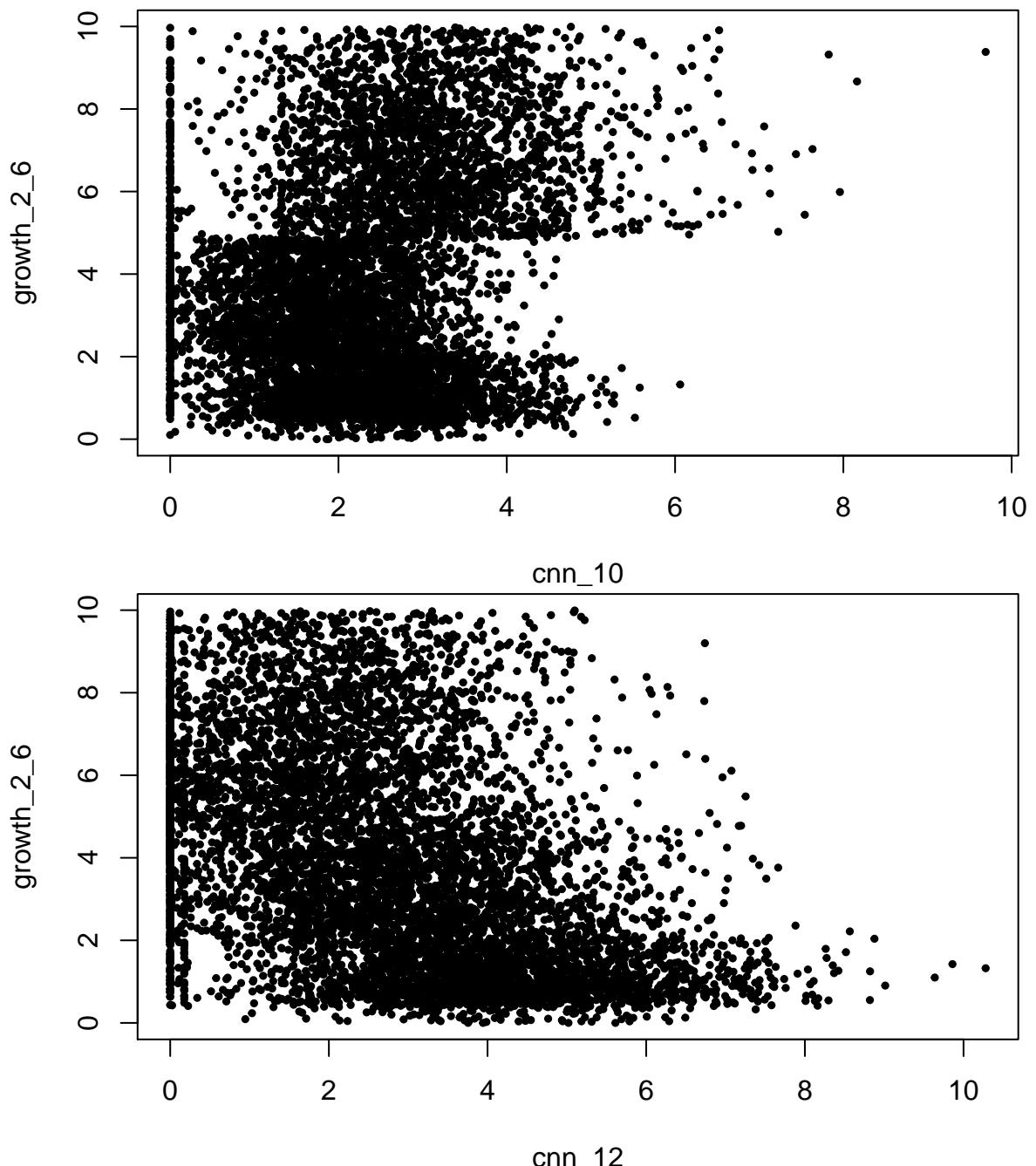
d <- training %>% select(-contains("hog")) %>% select(-PublishedDate)
plot(growth_2_6 ~ ., data = d, cex = 0.5, pch = 19)

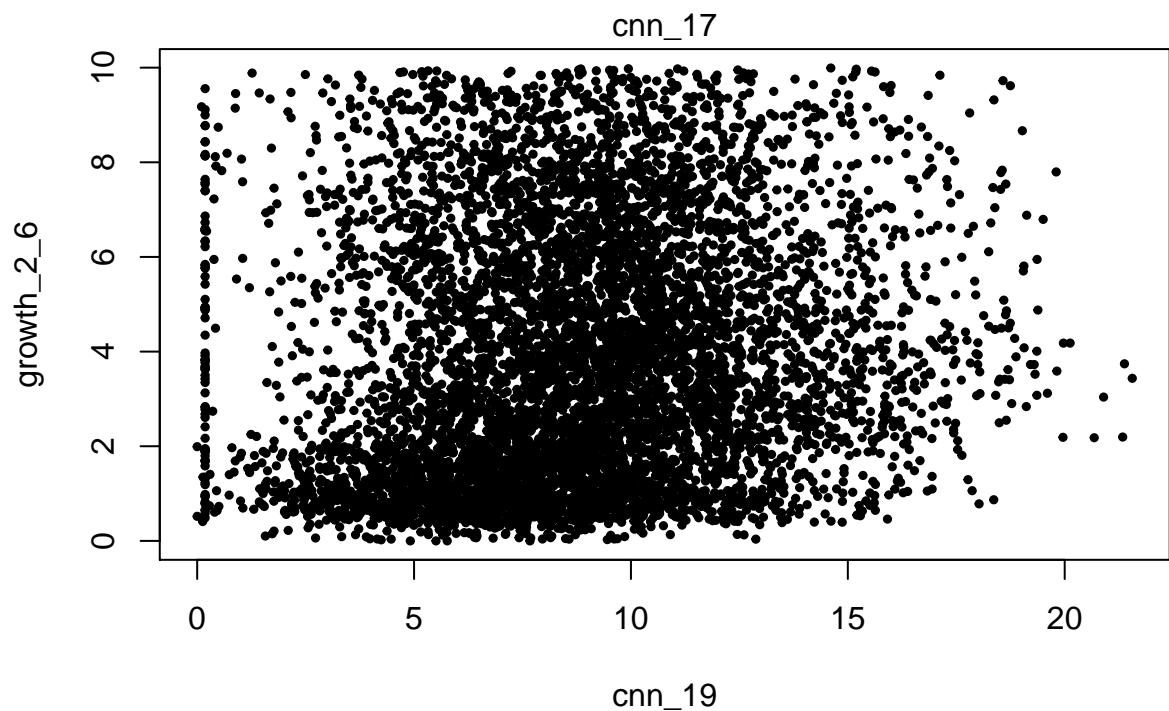
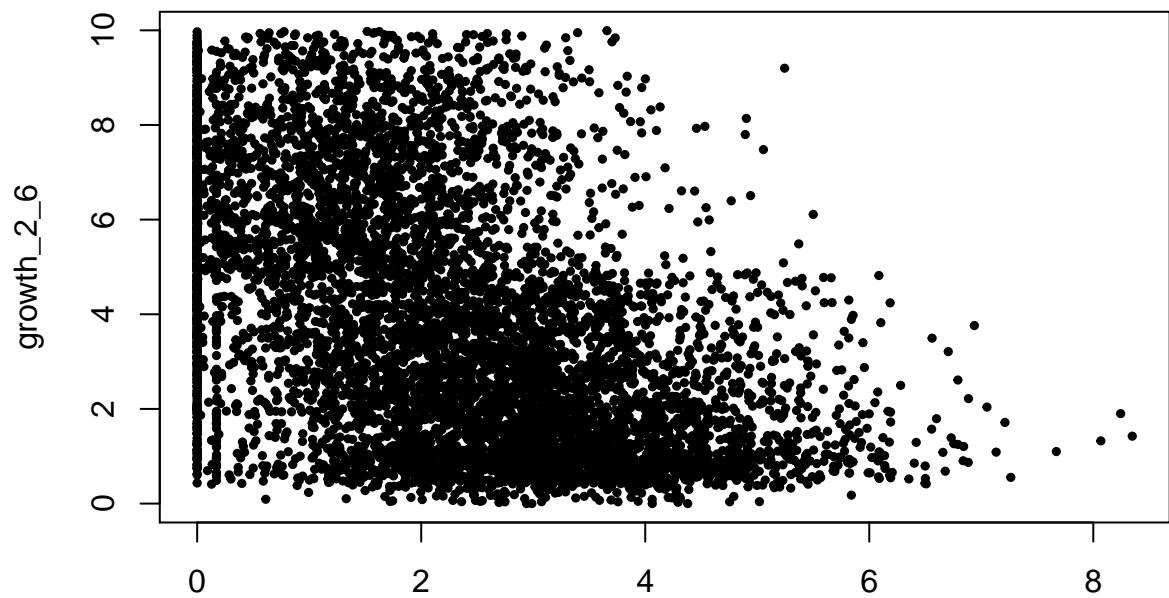
```

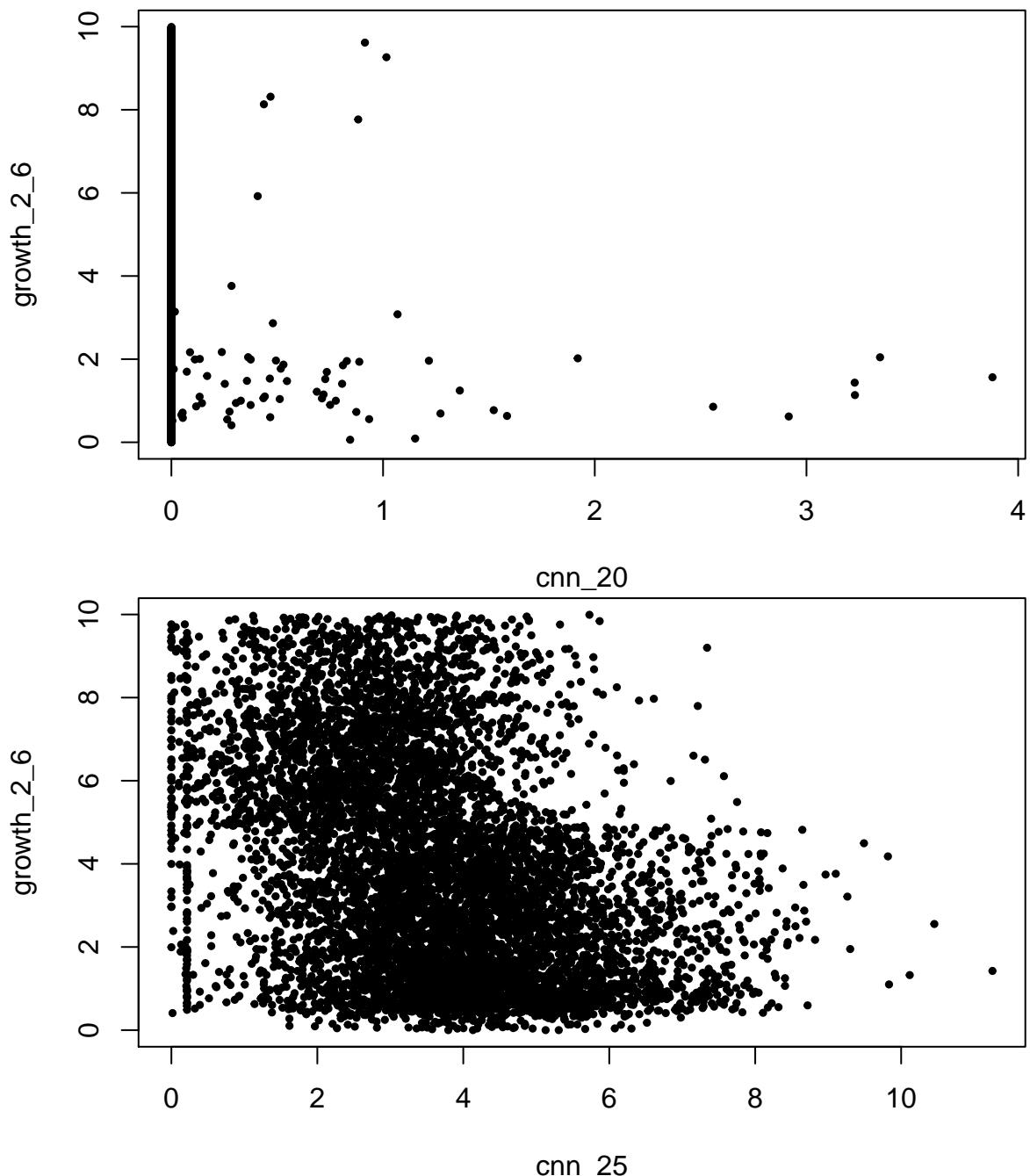


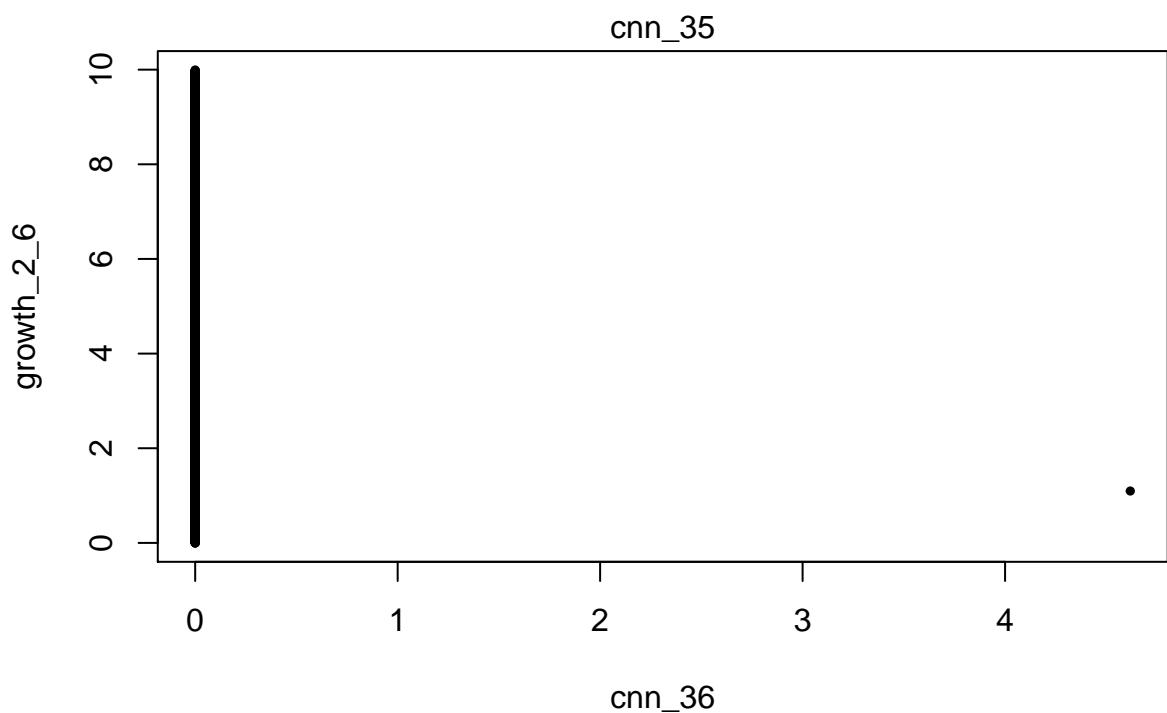
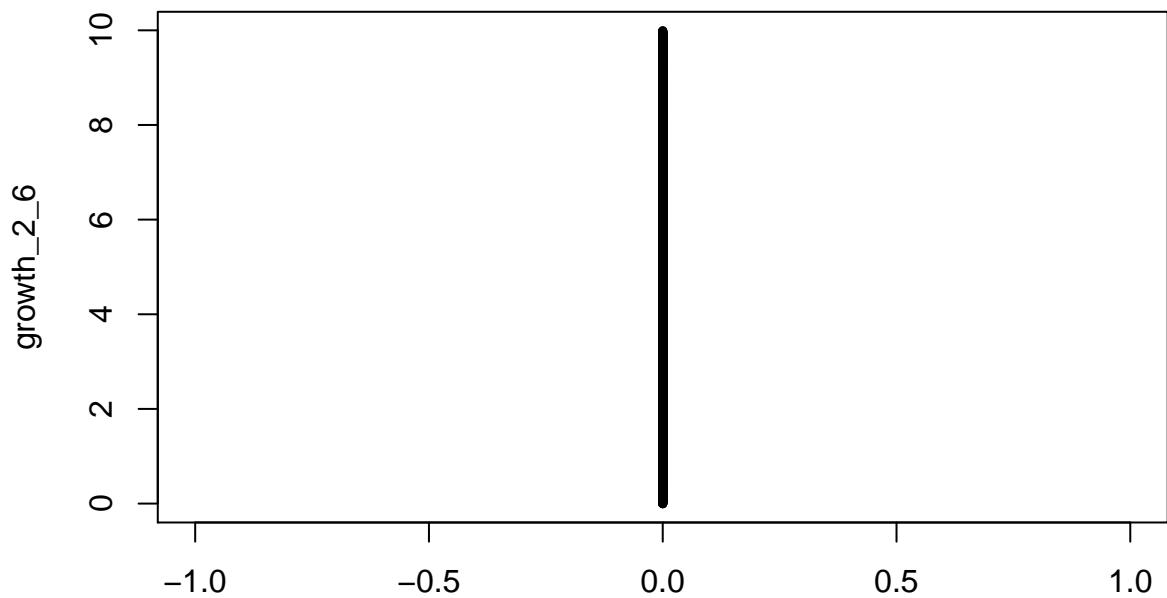


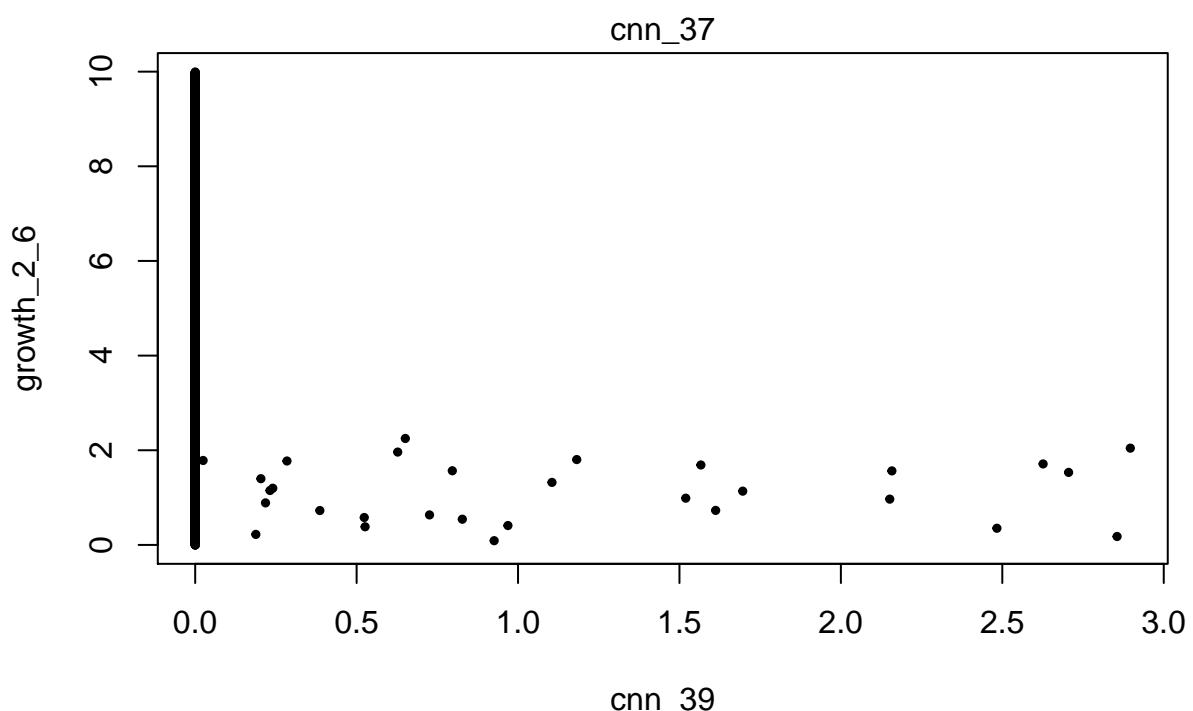
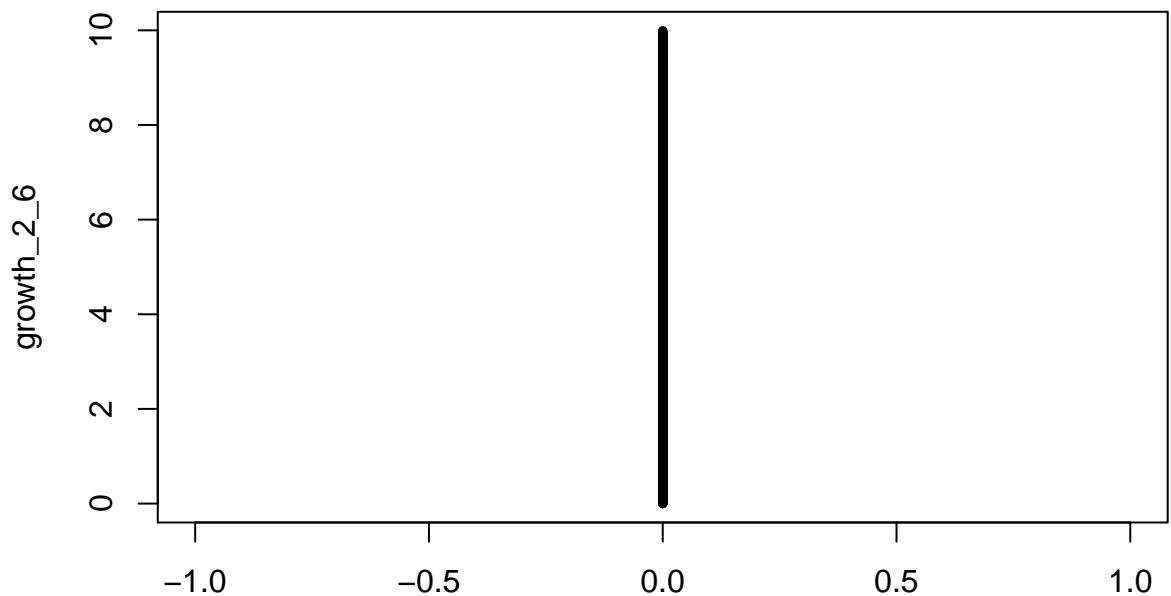


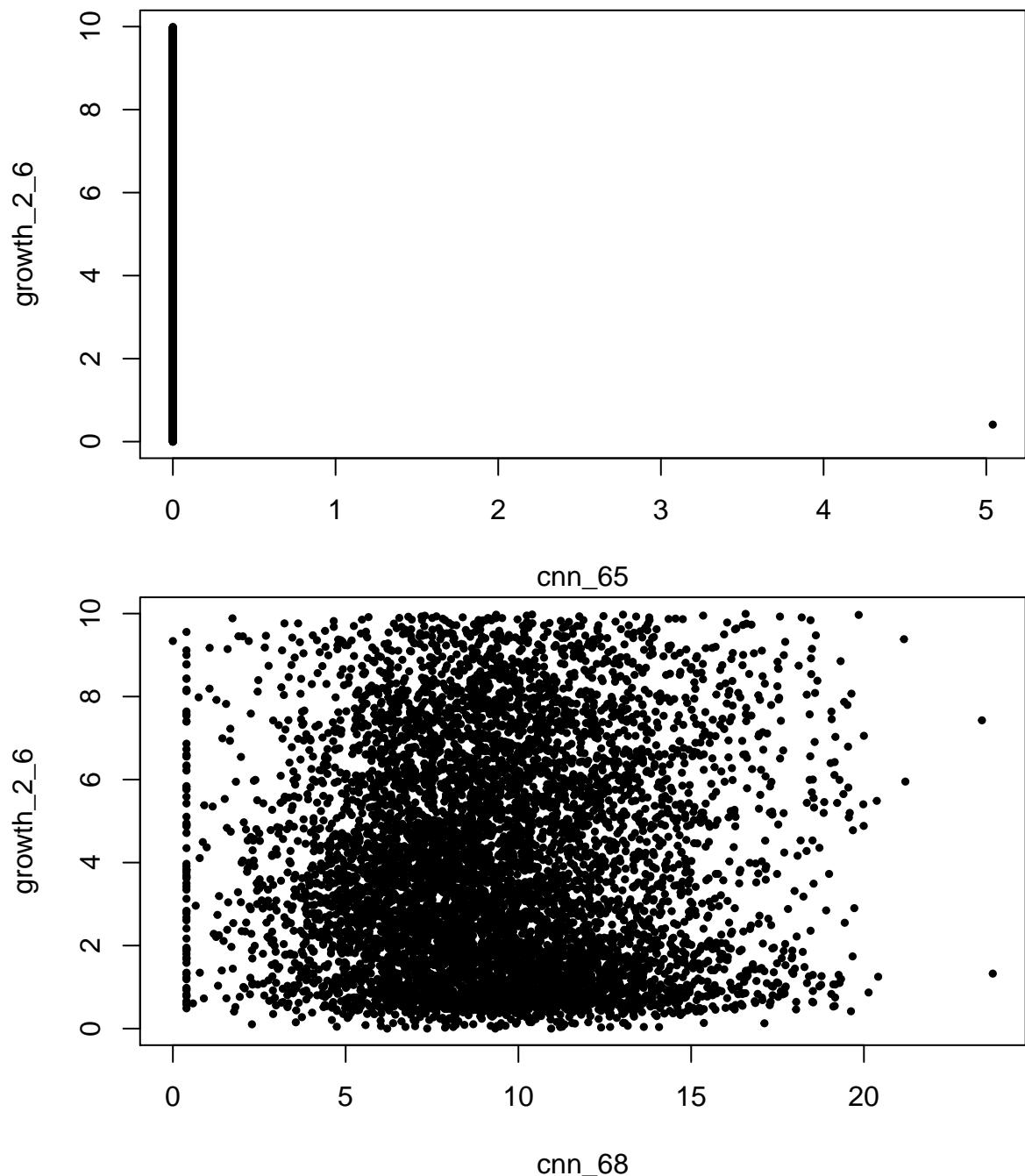


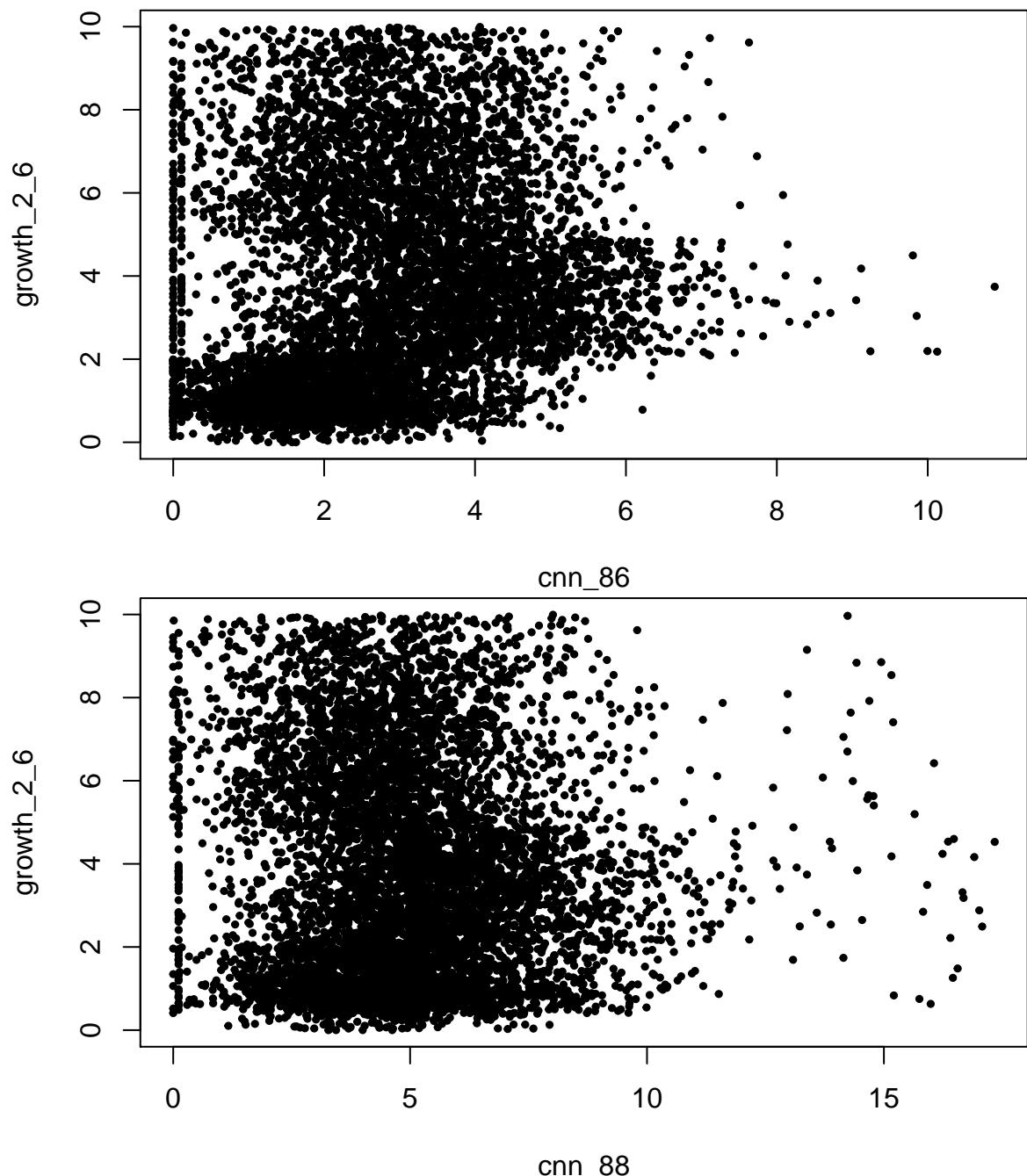


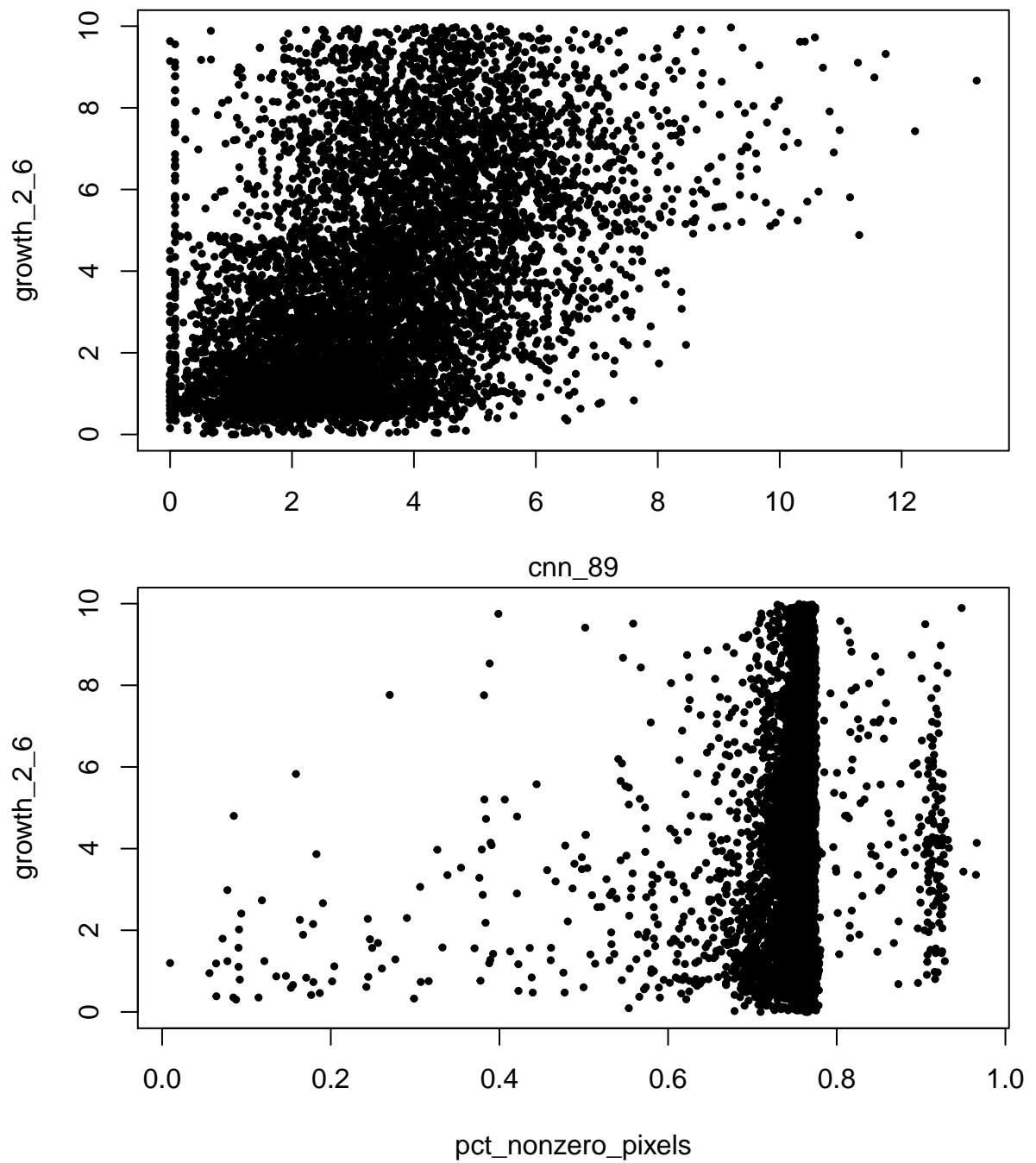


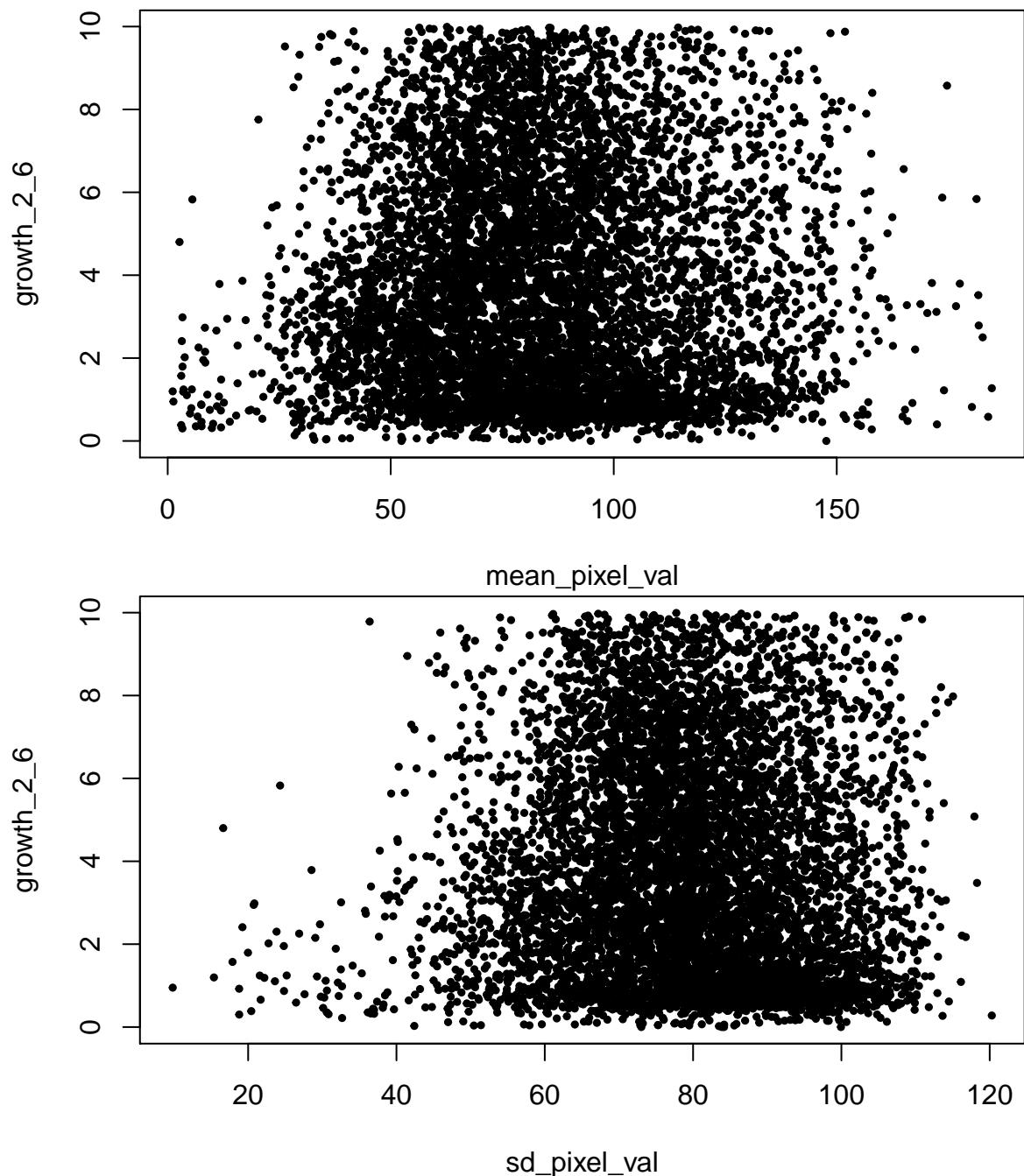


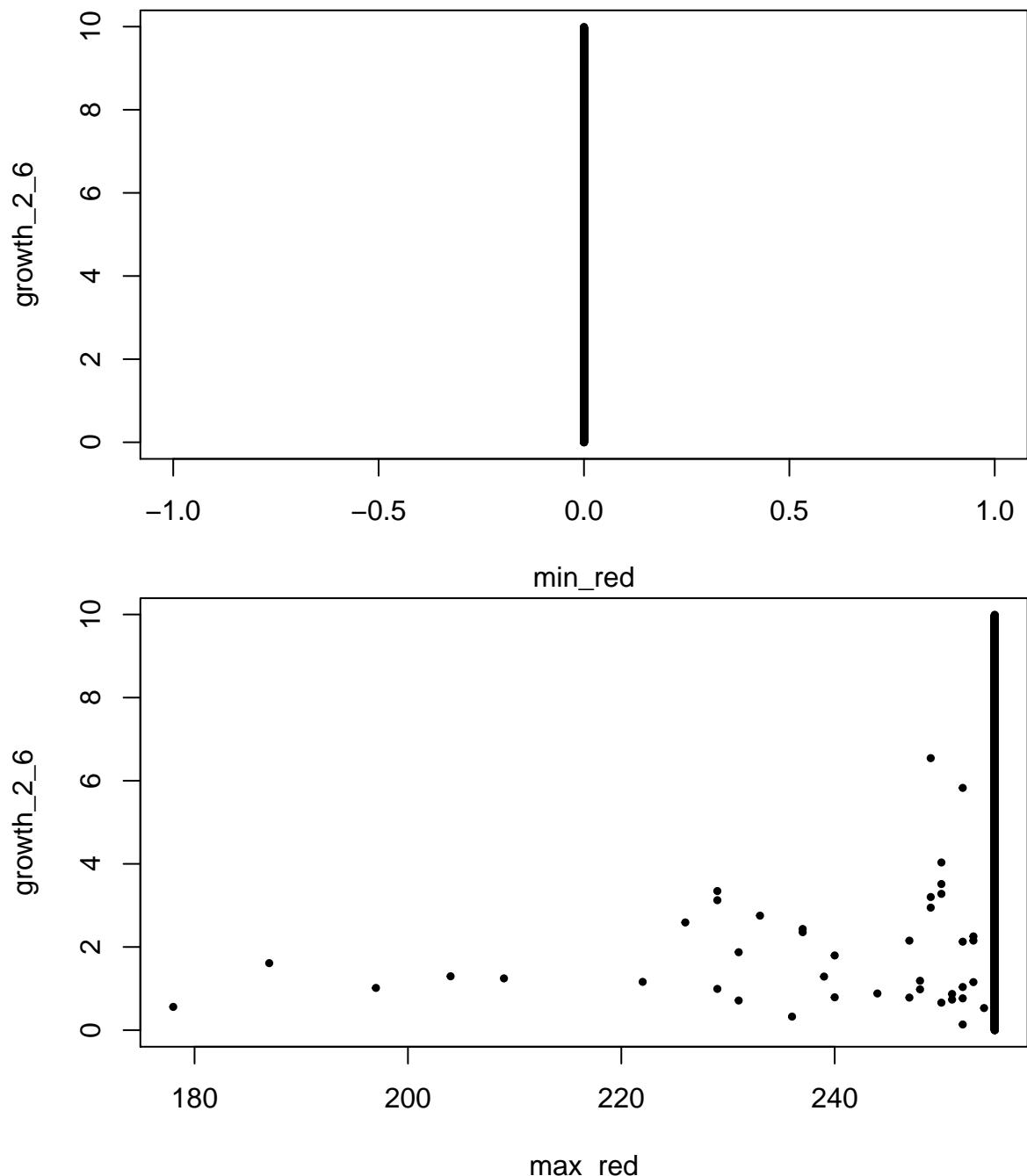


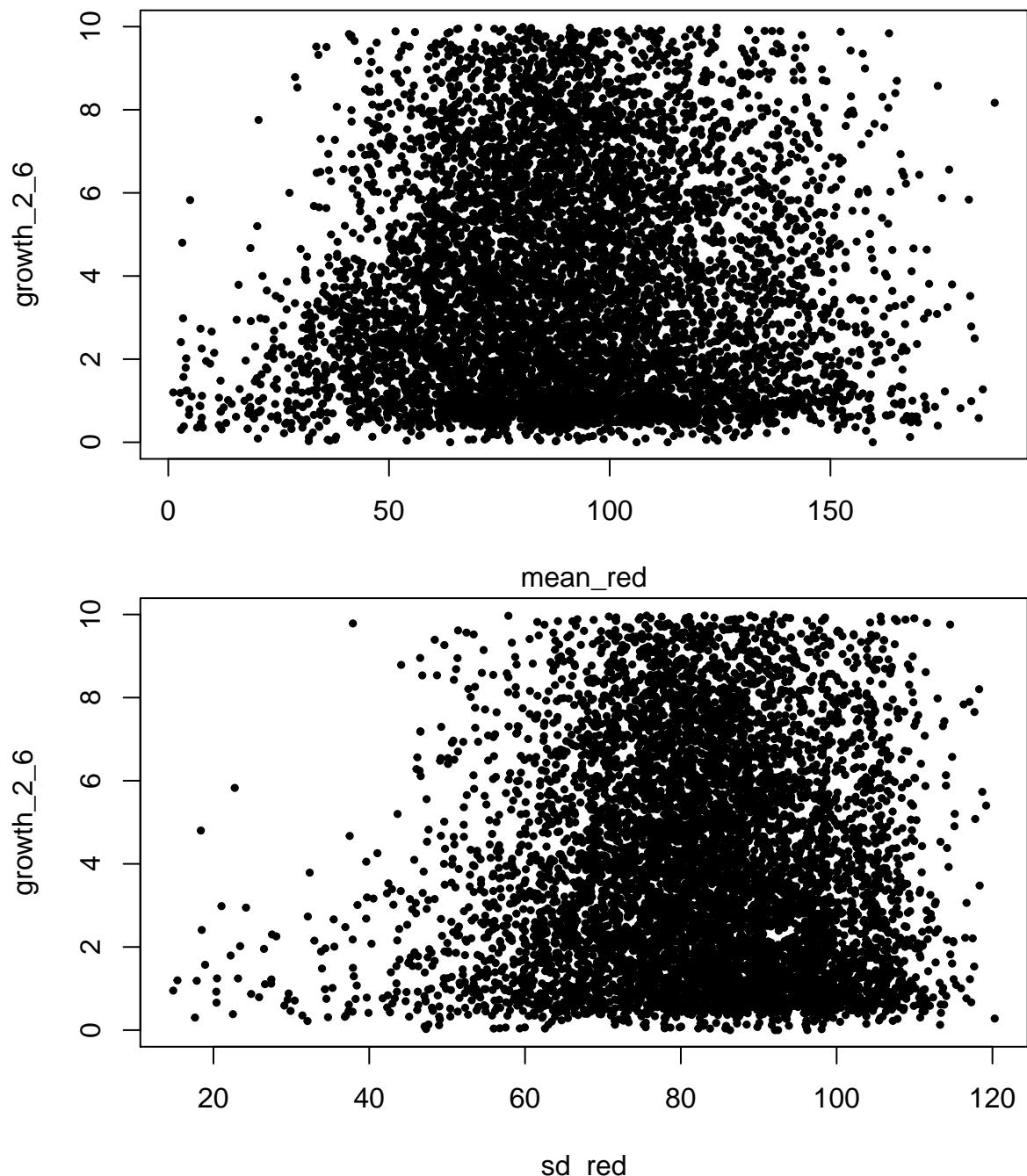


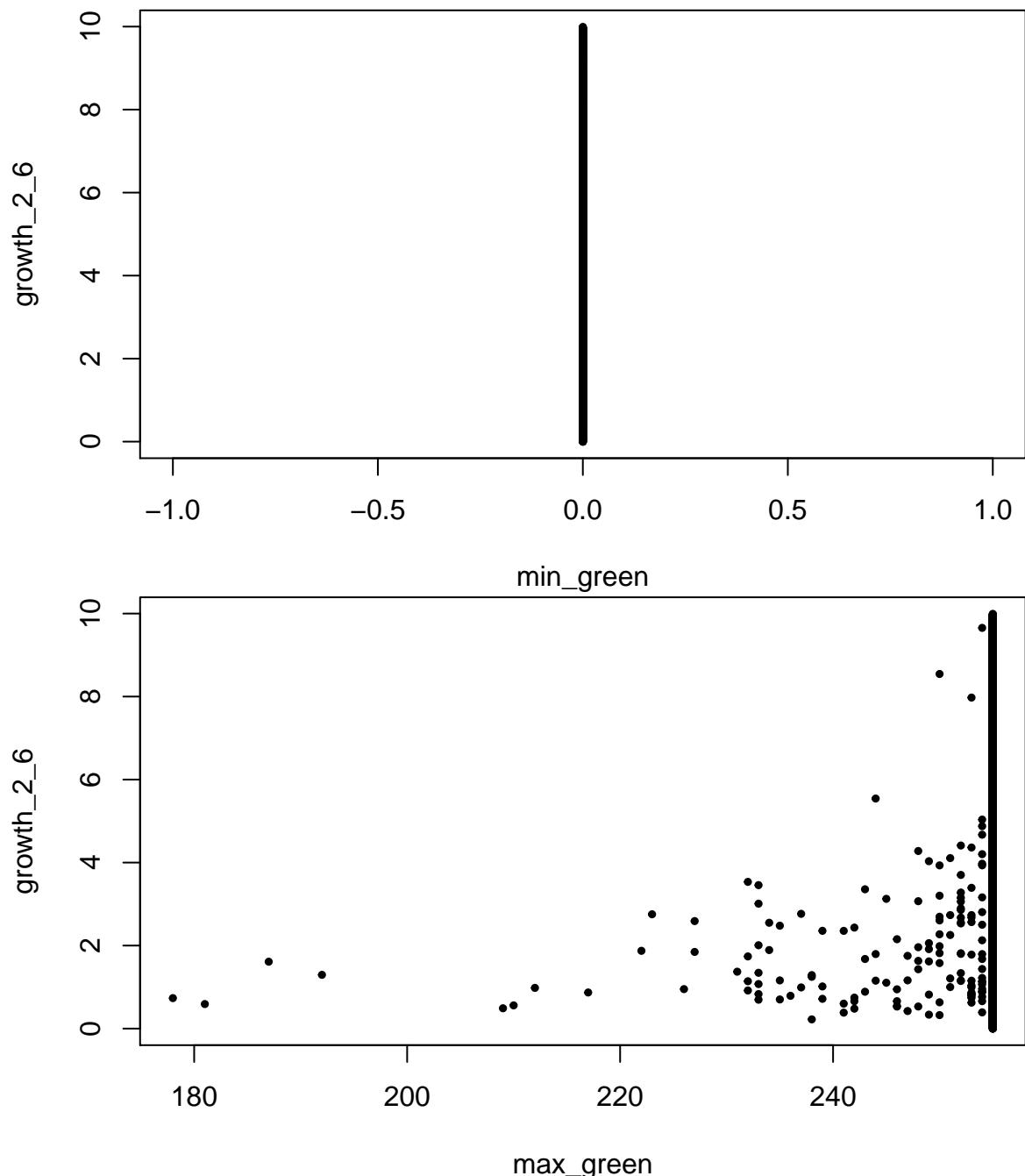


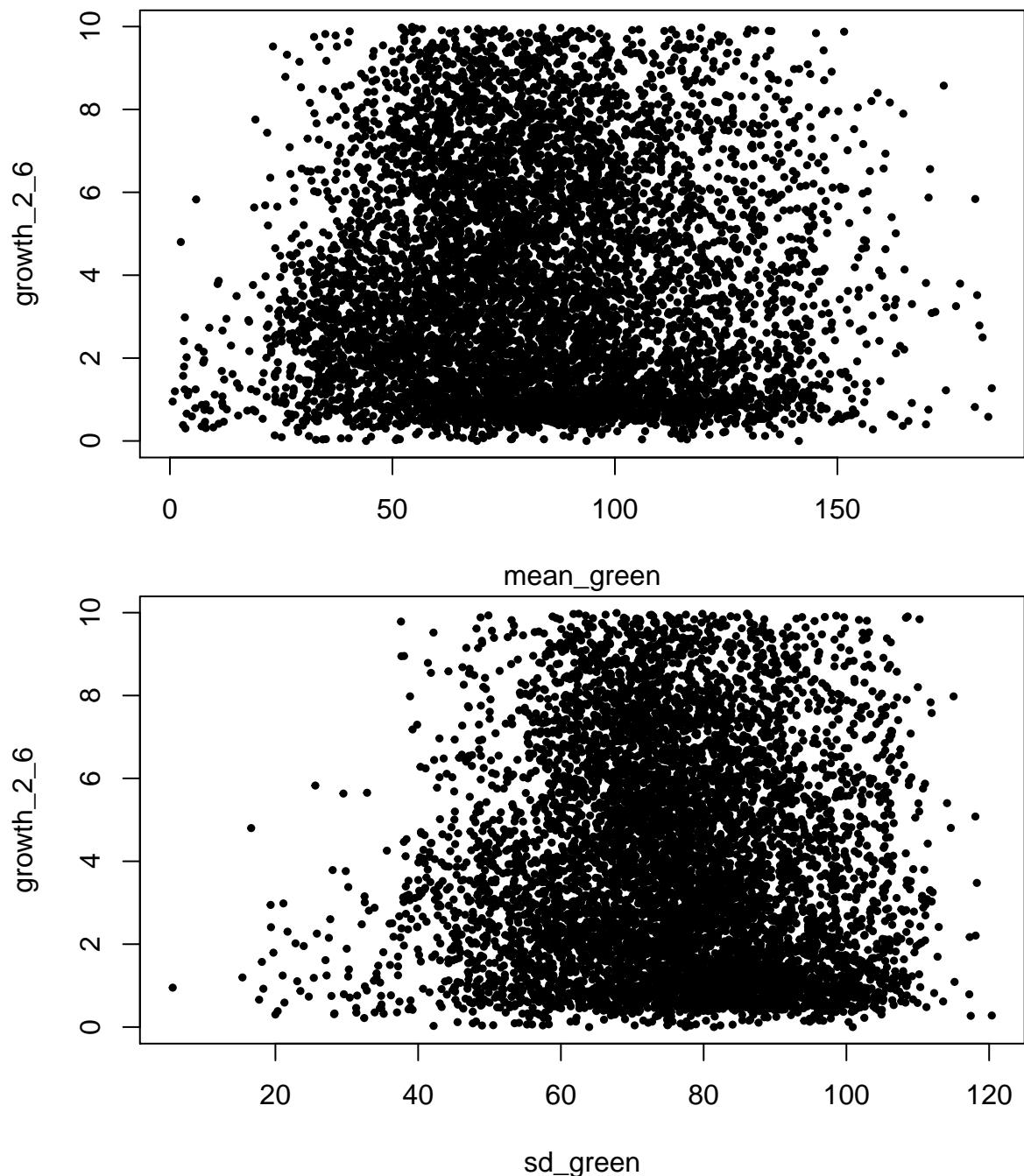


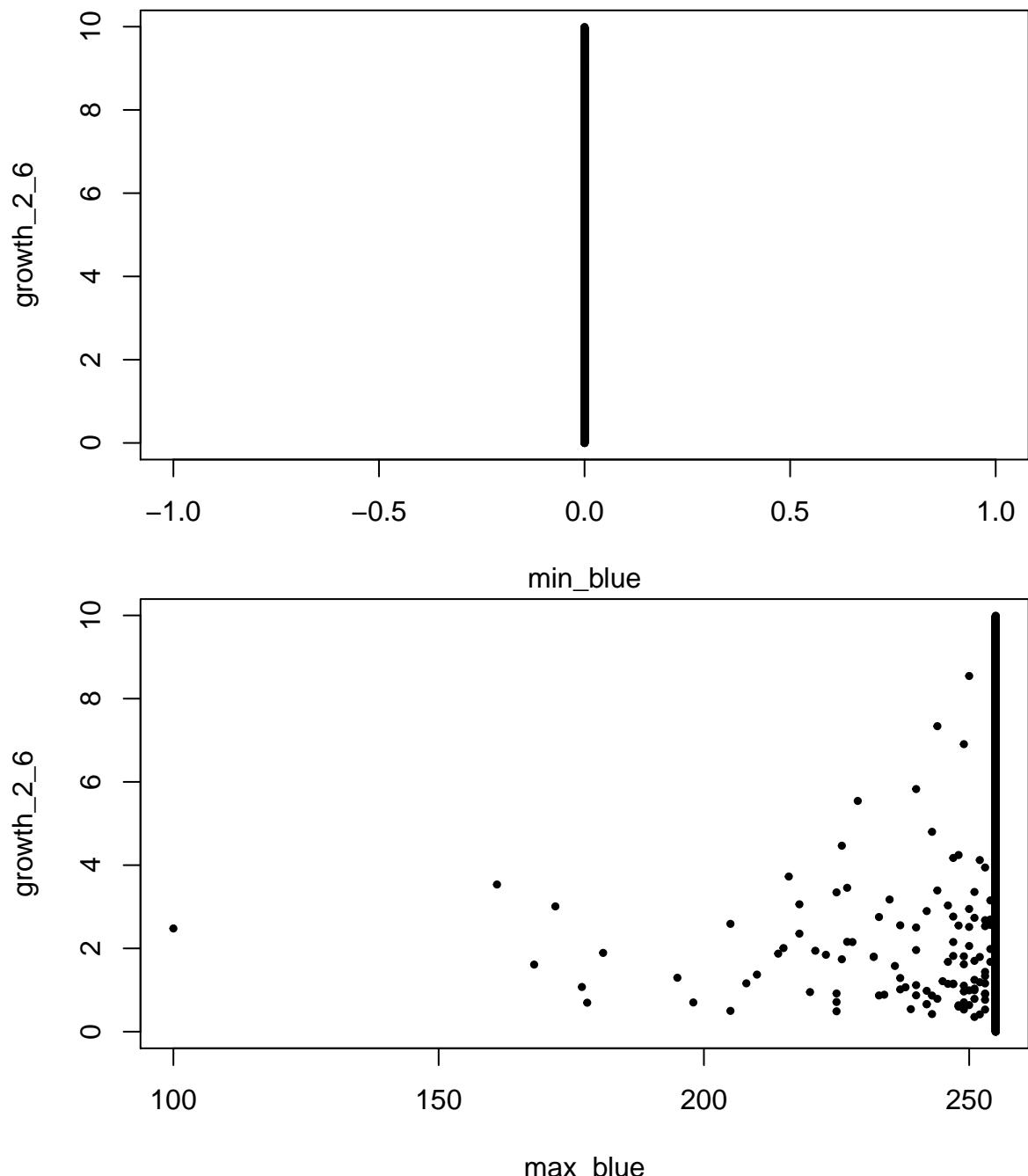


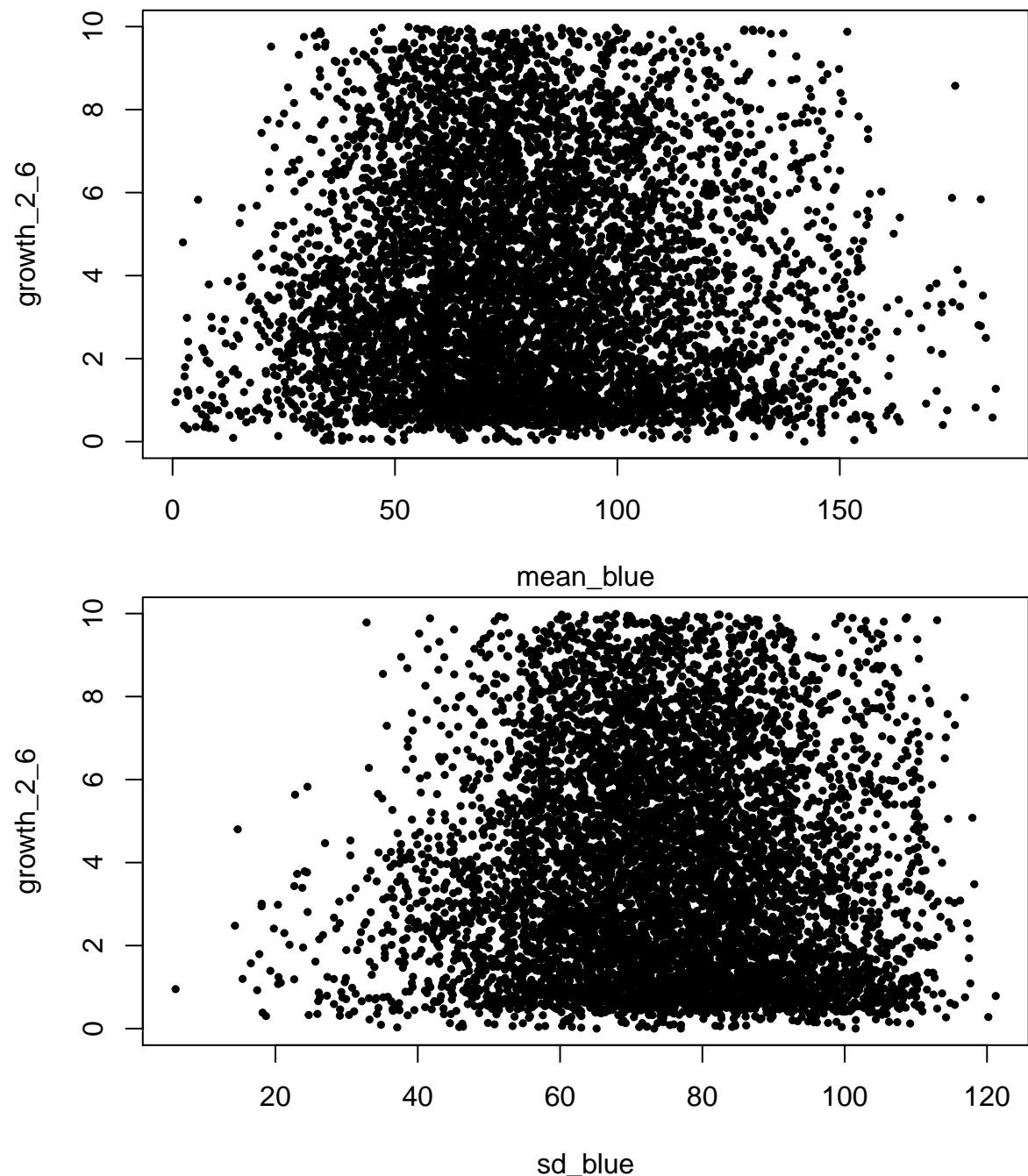


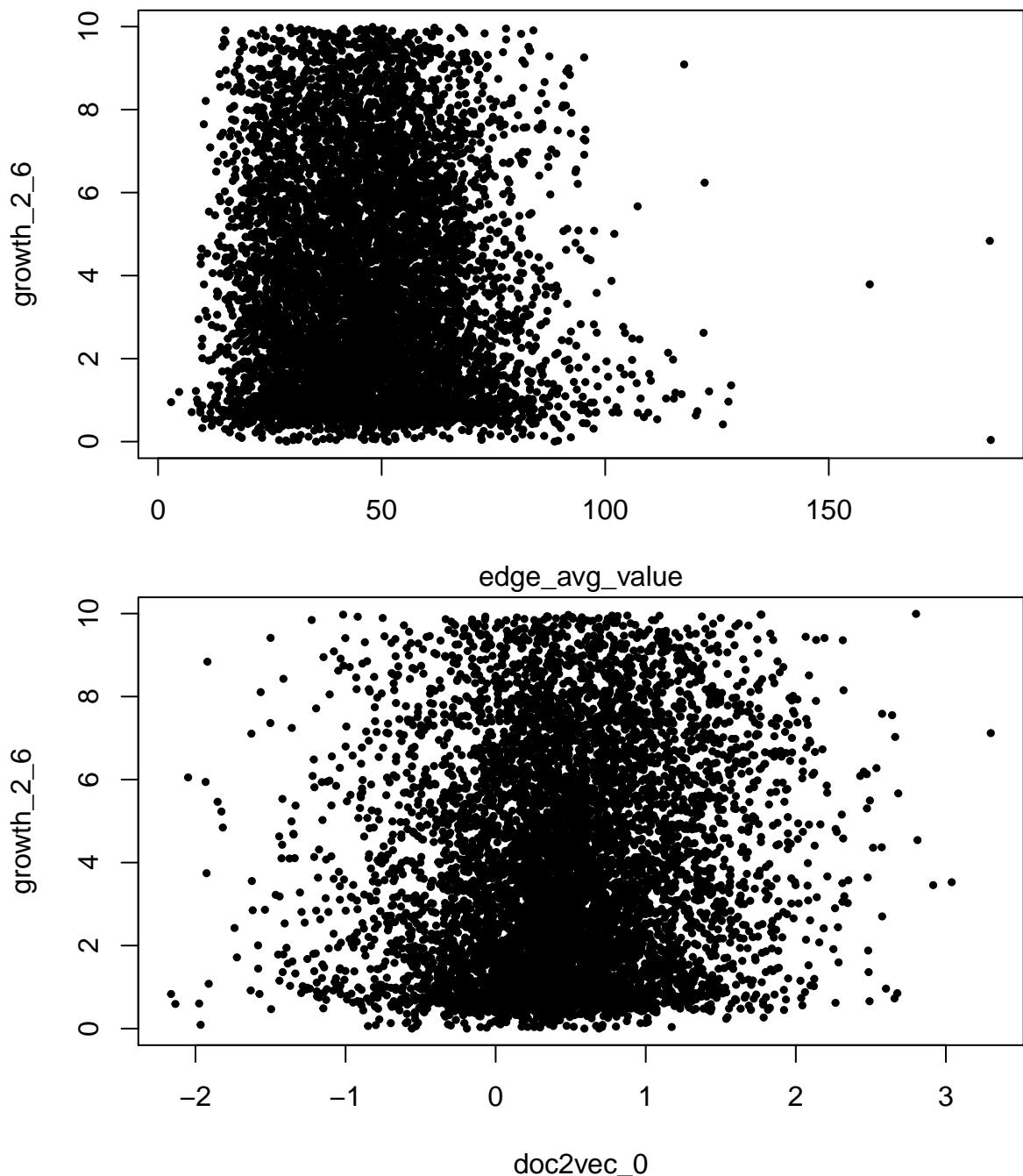


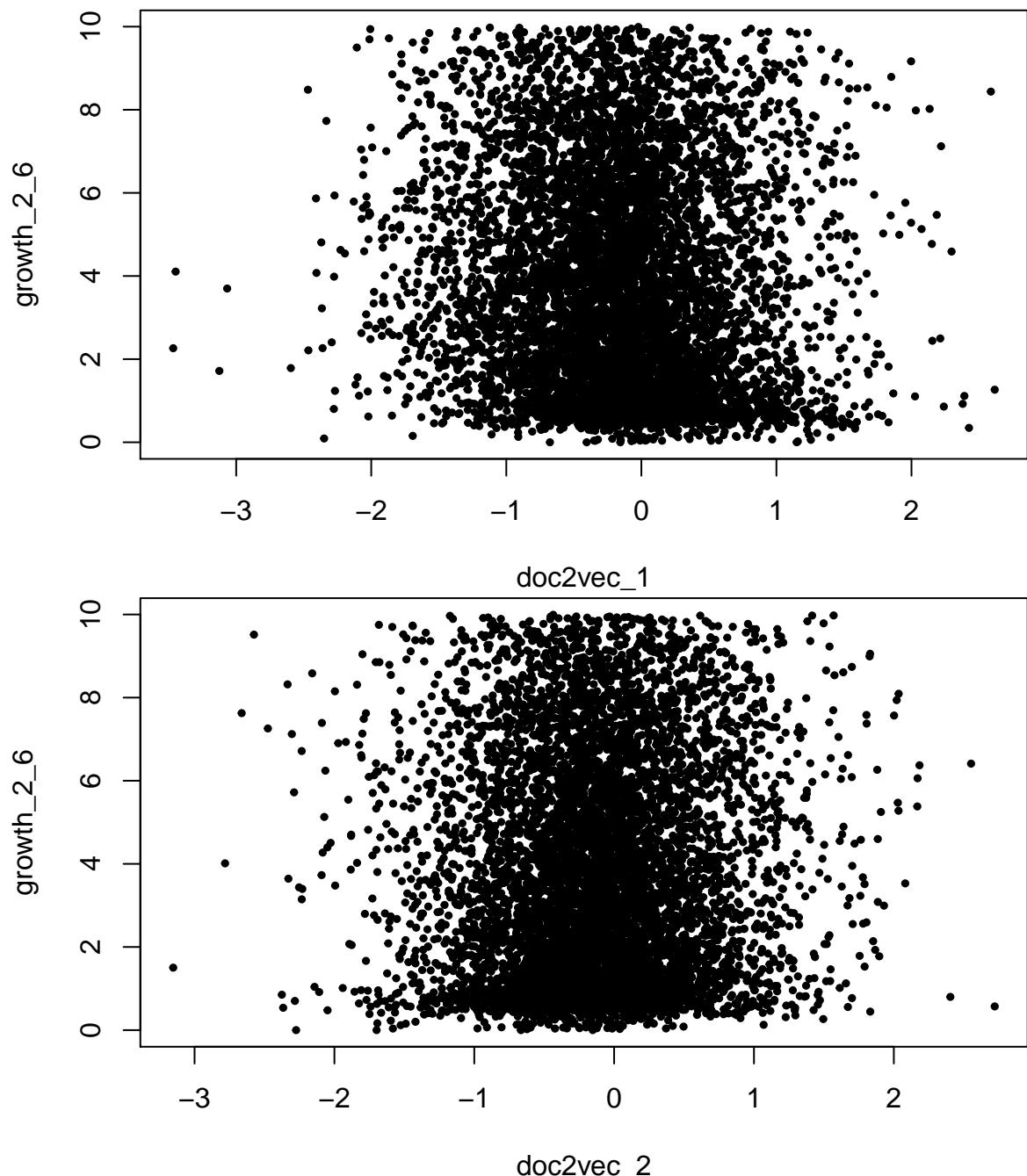


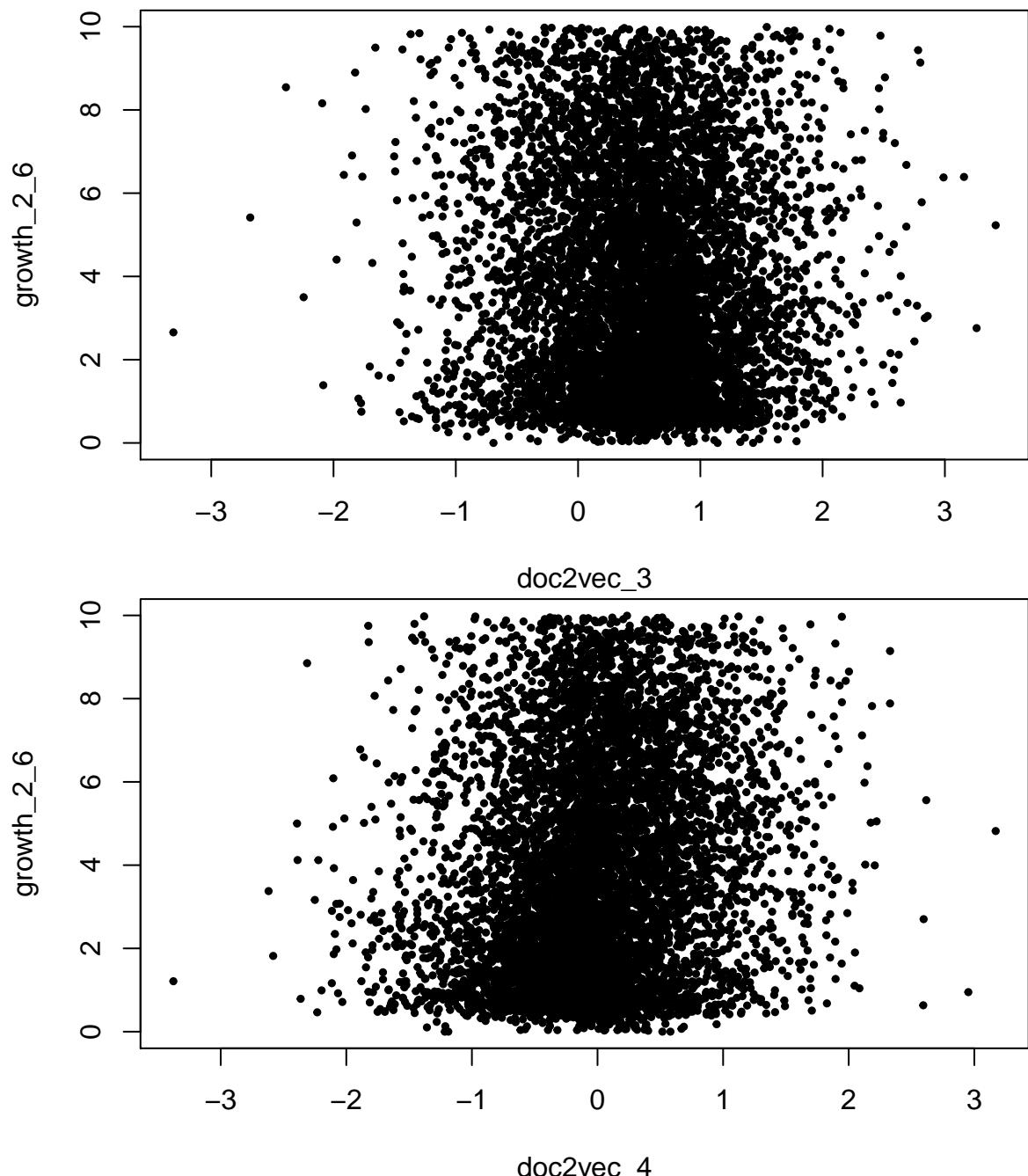


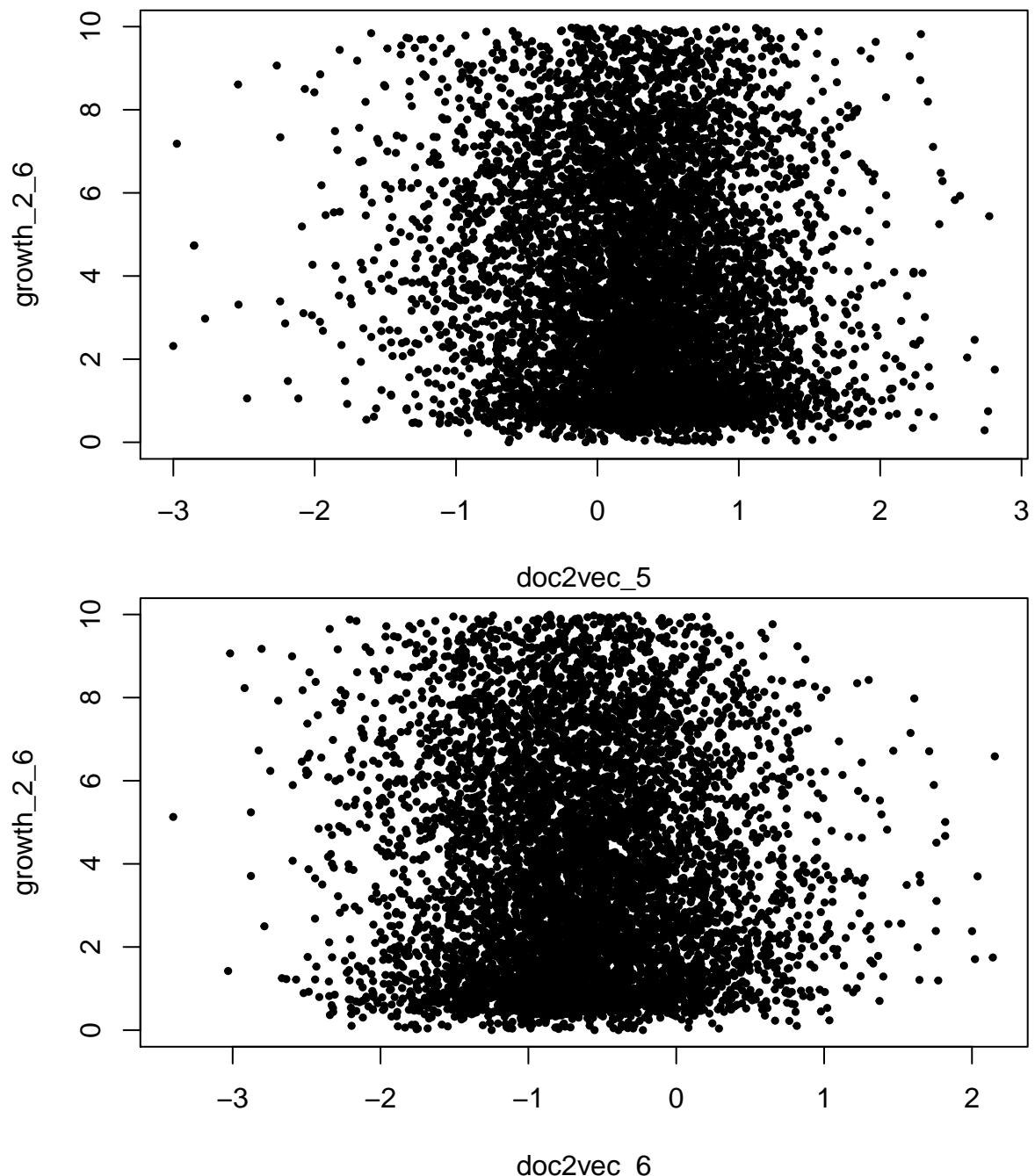


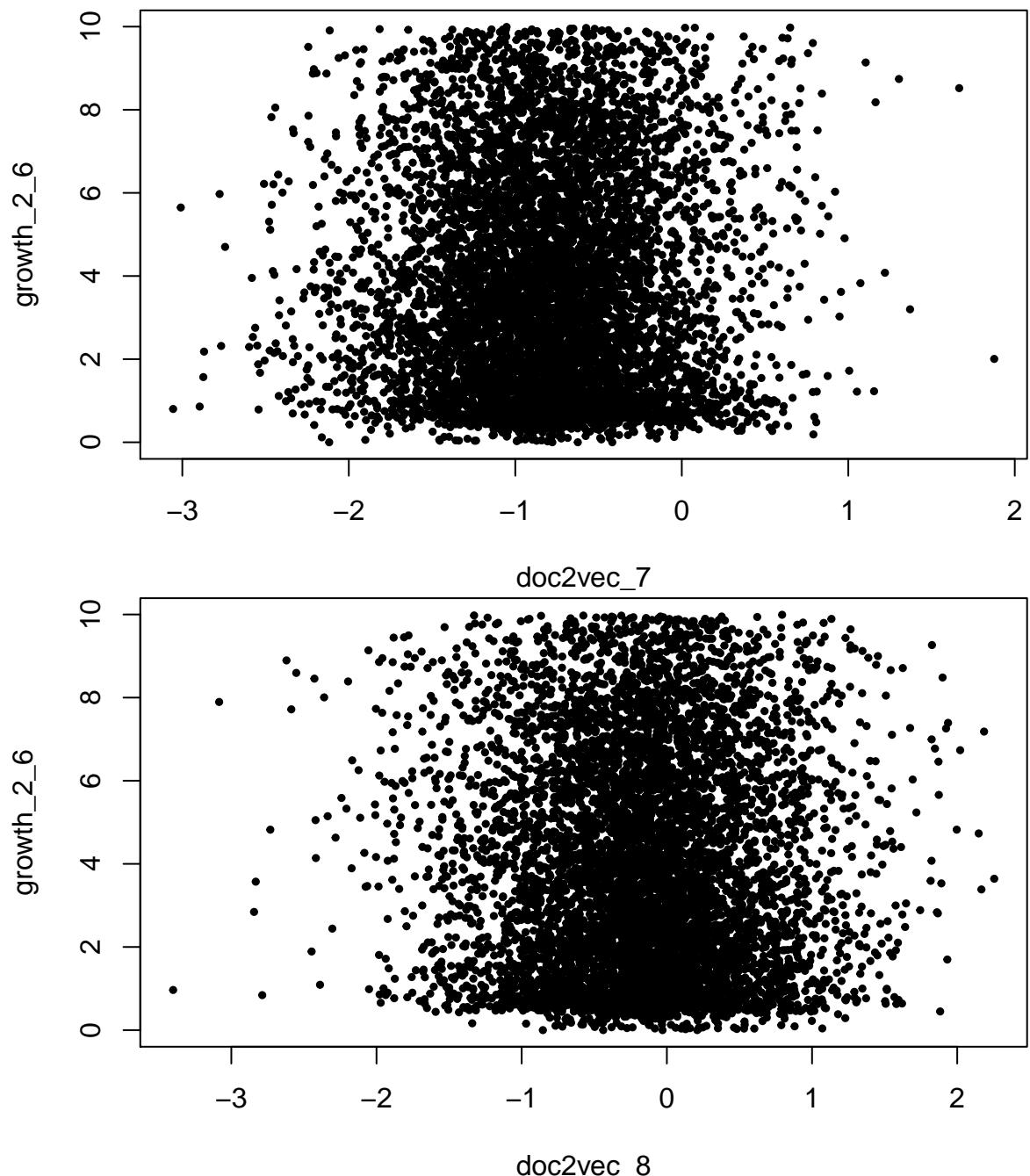


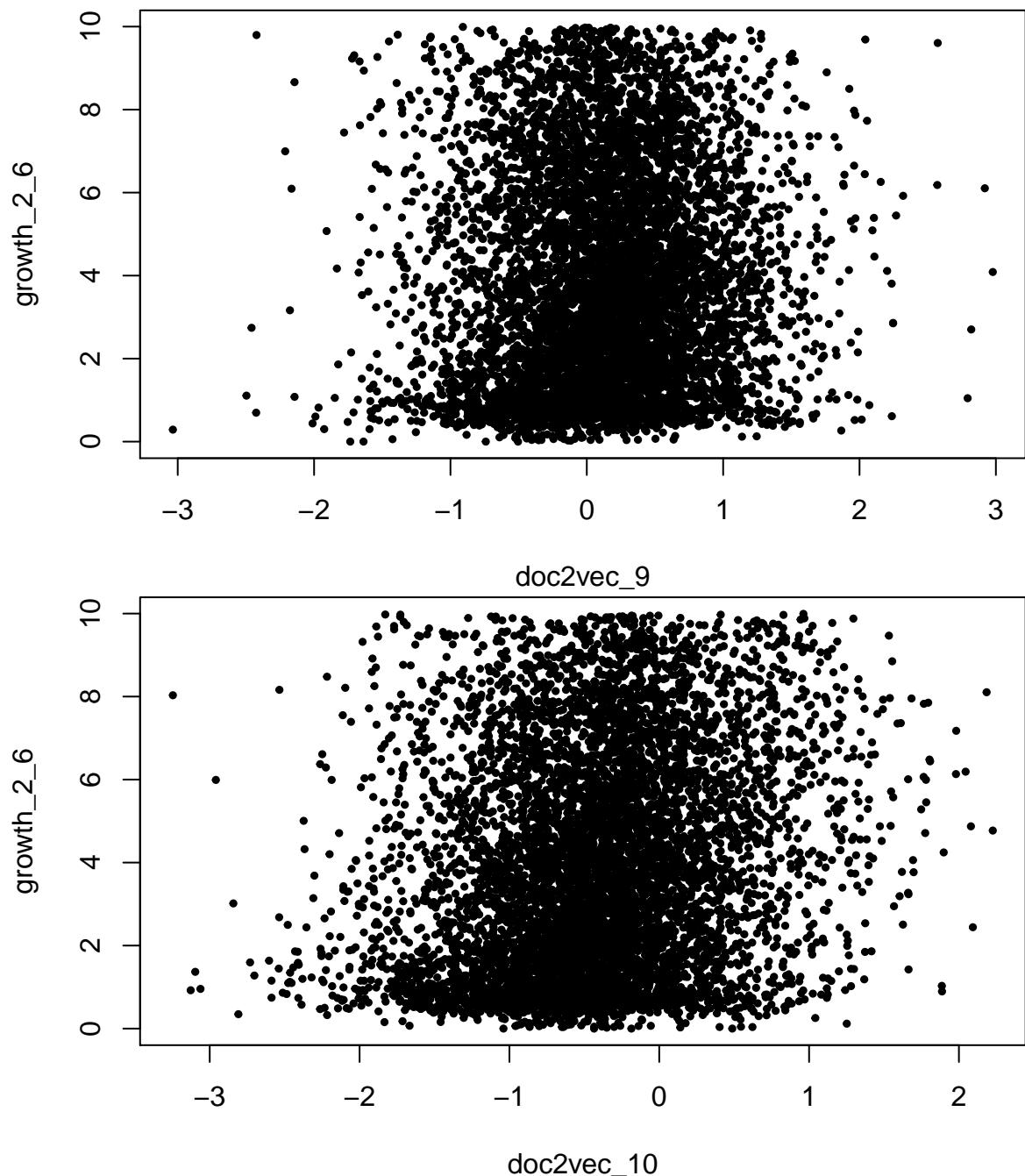


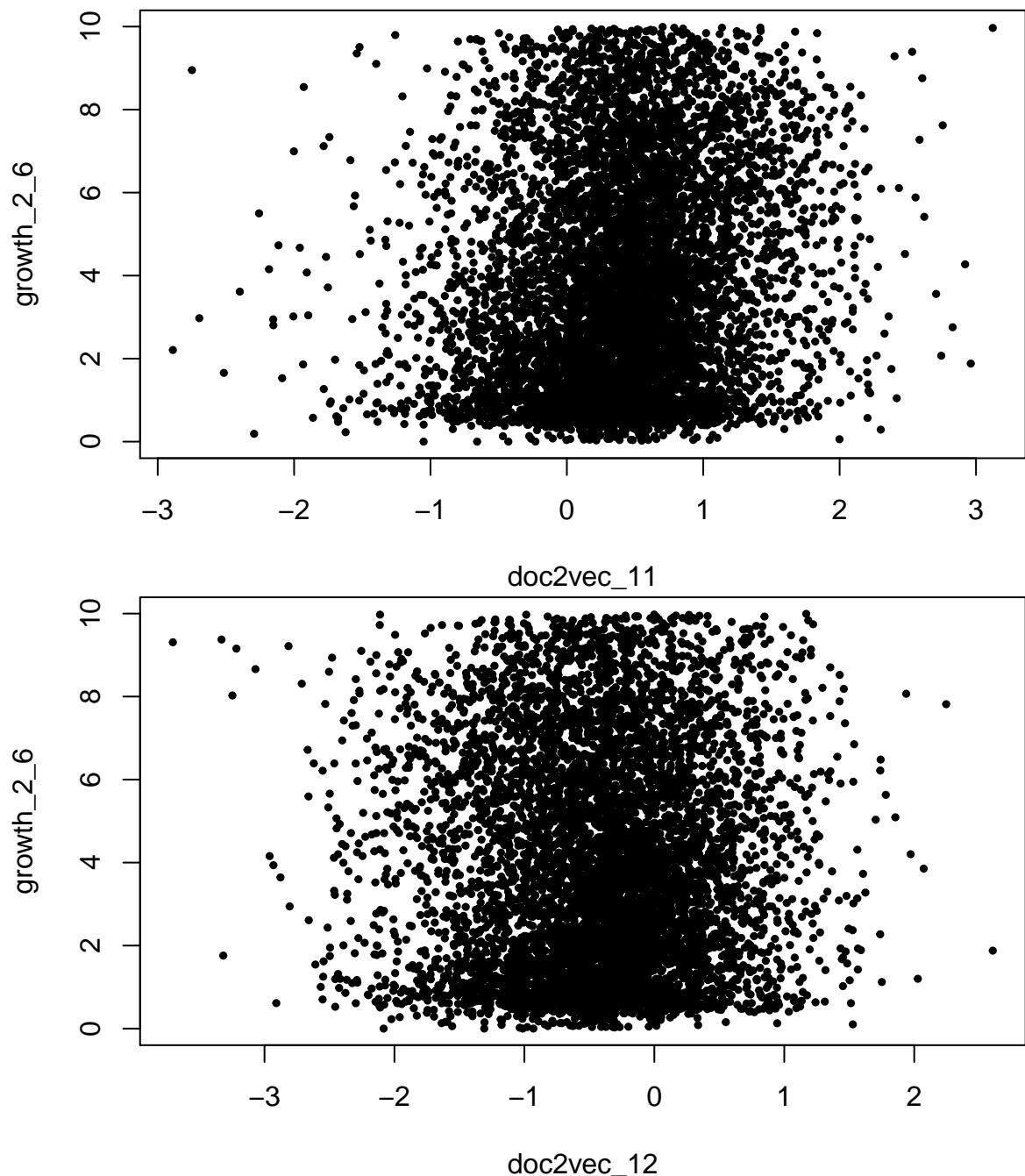


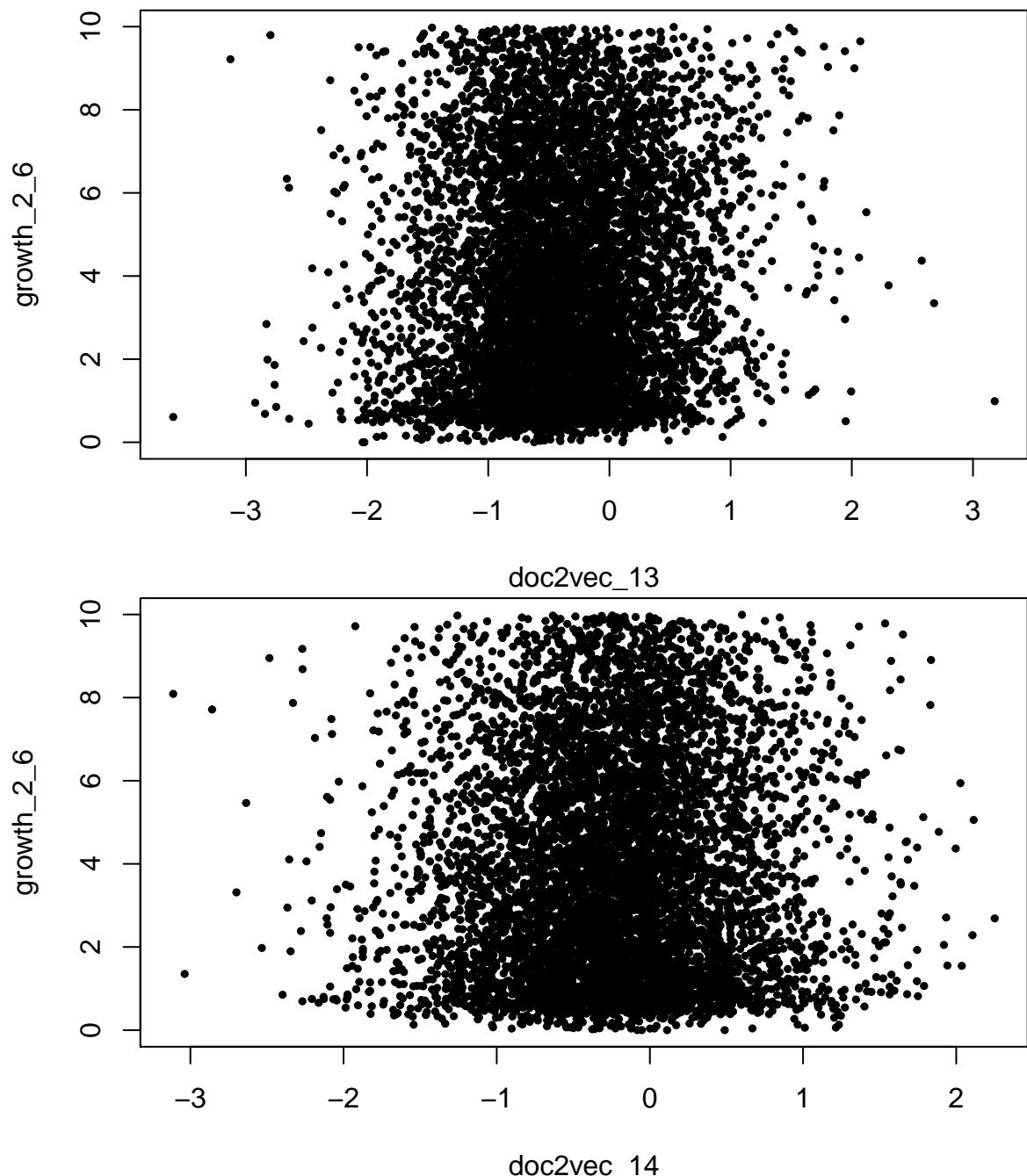


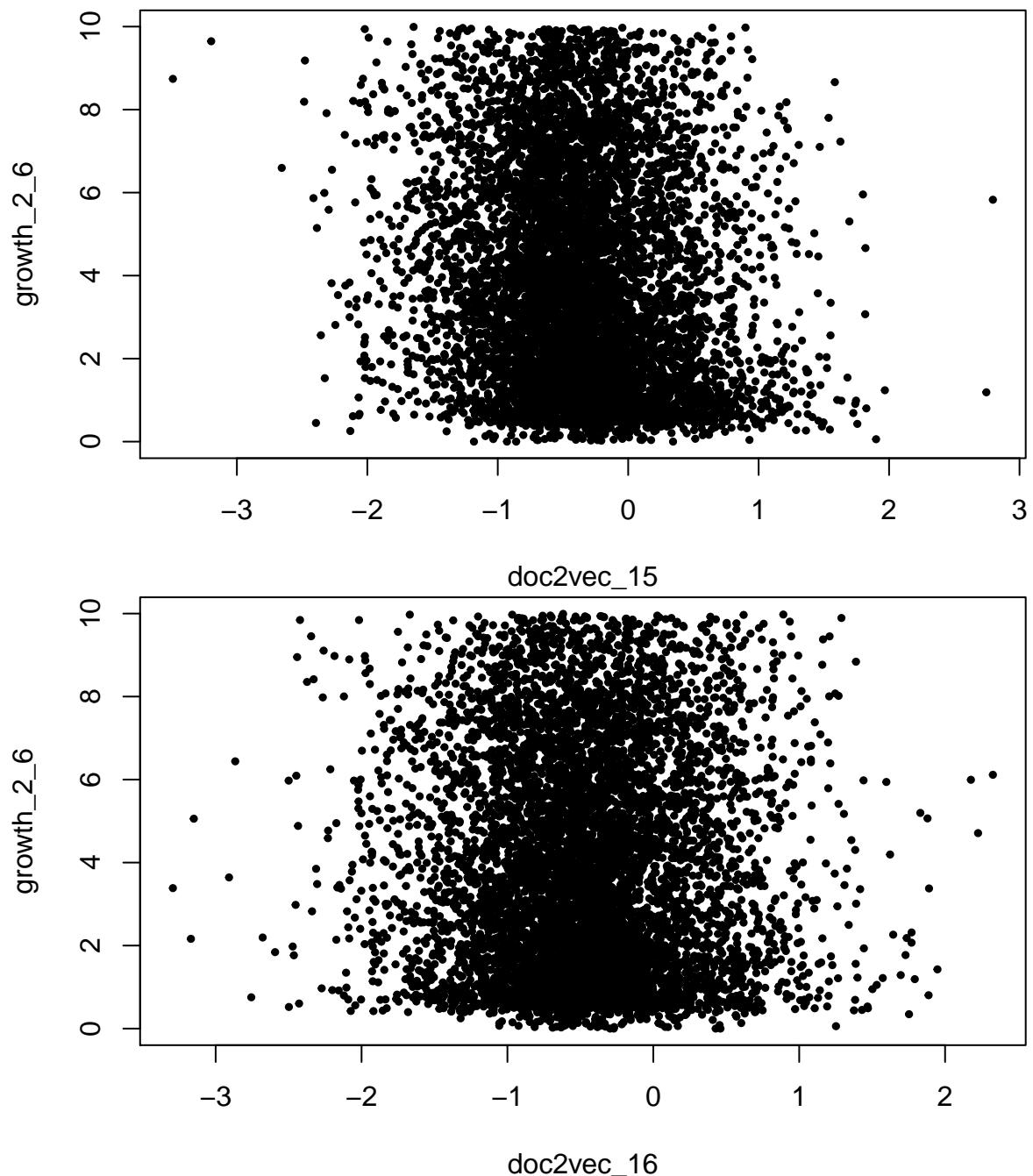


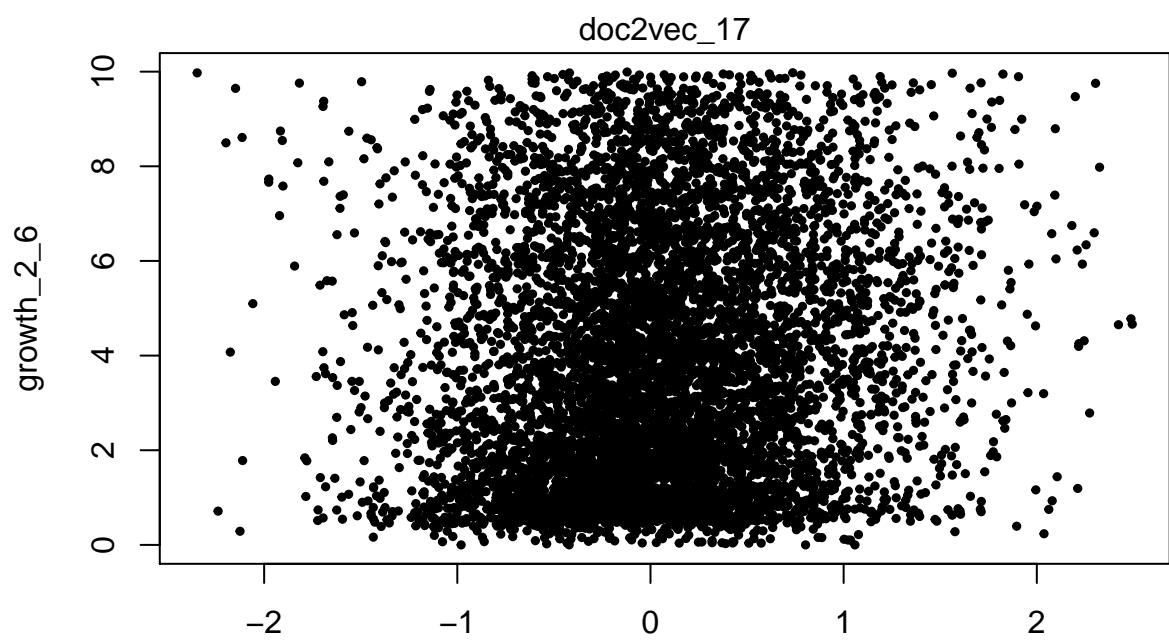
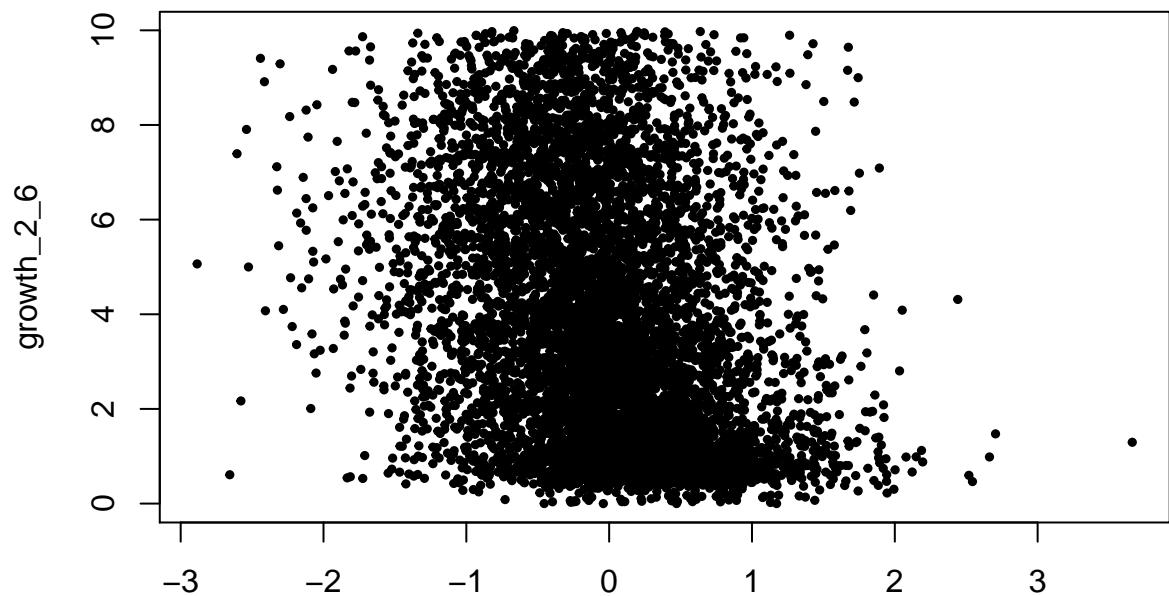




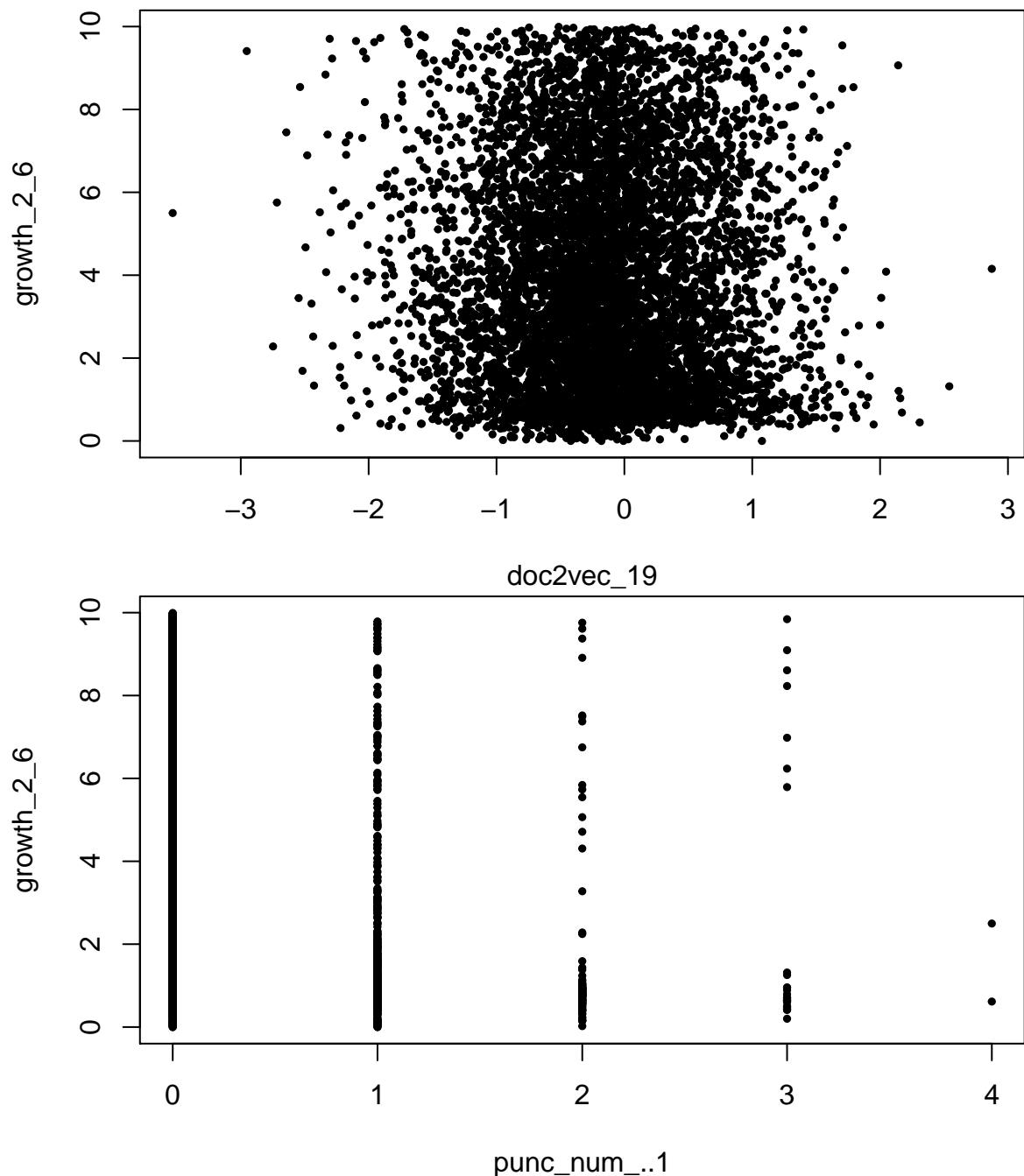


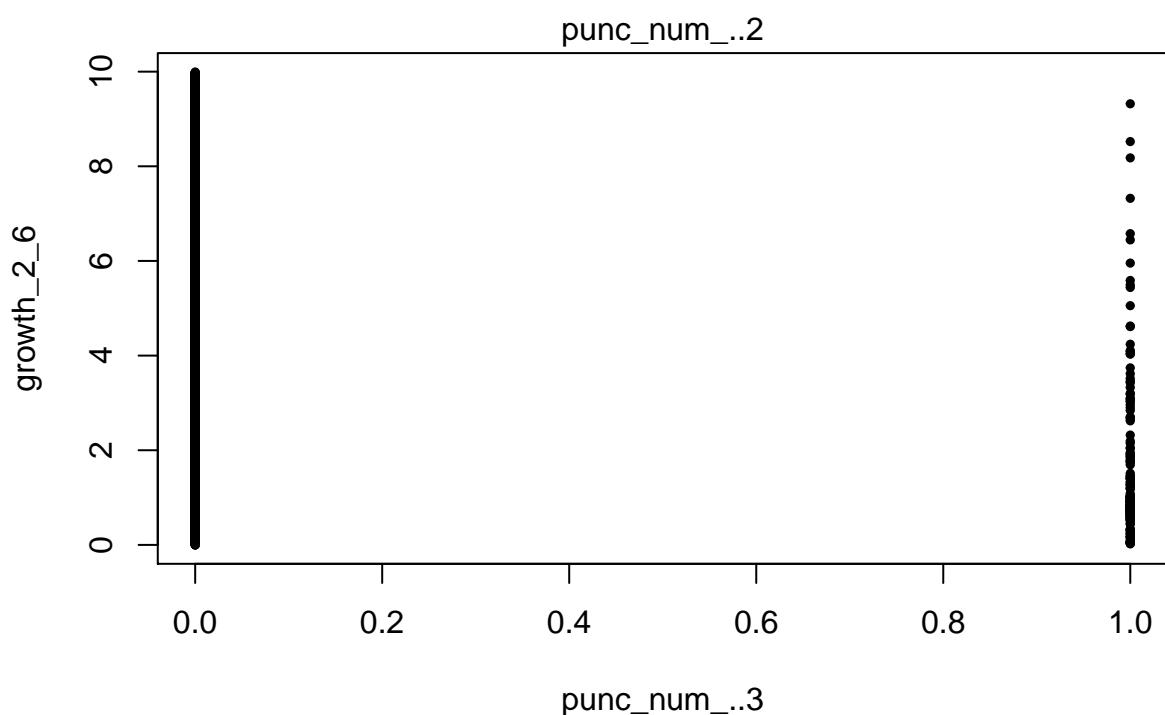
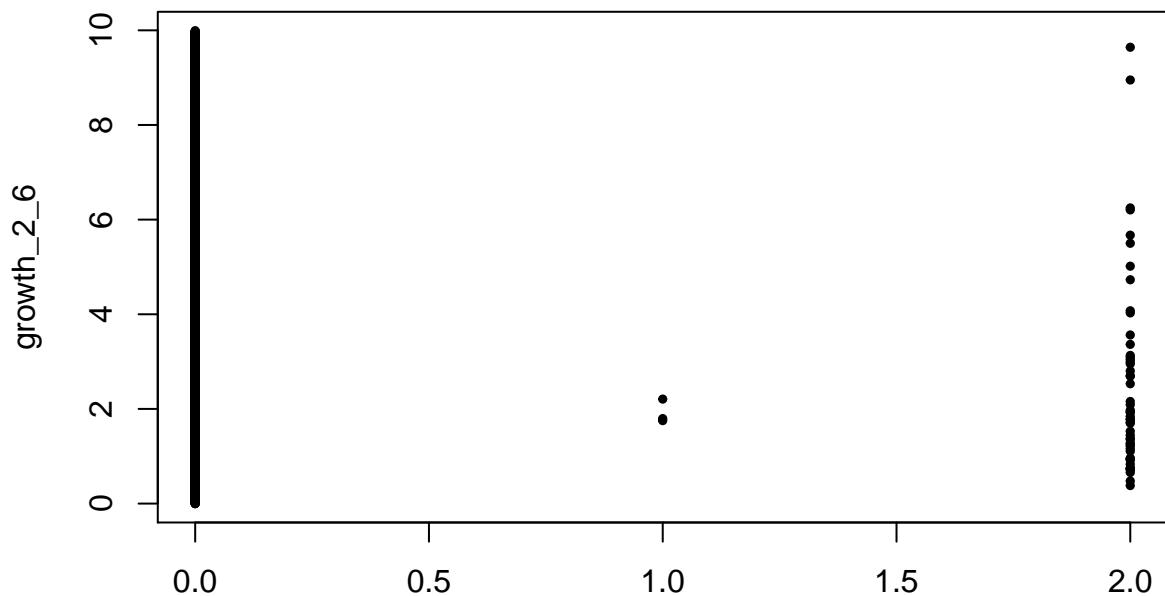


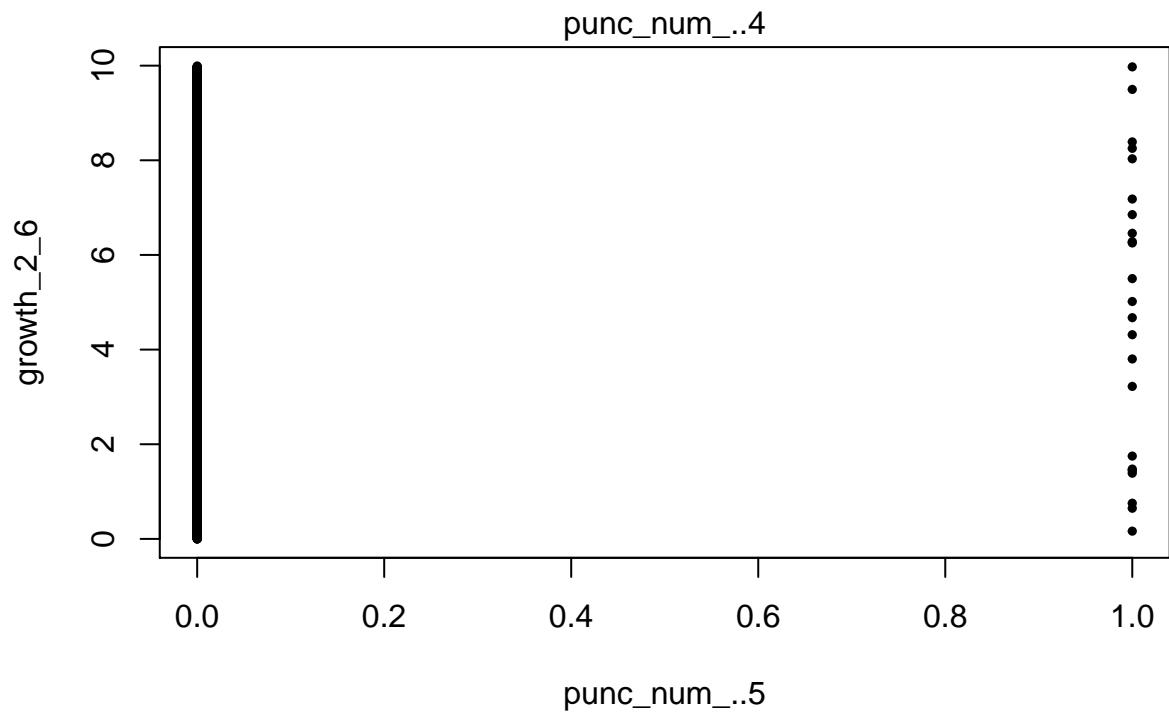
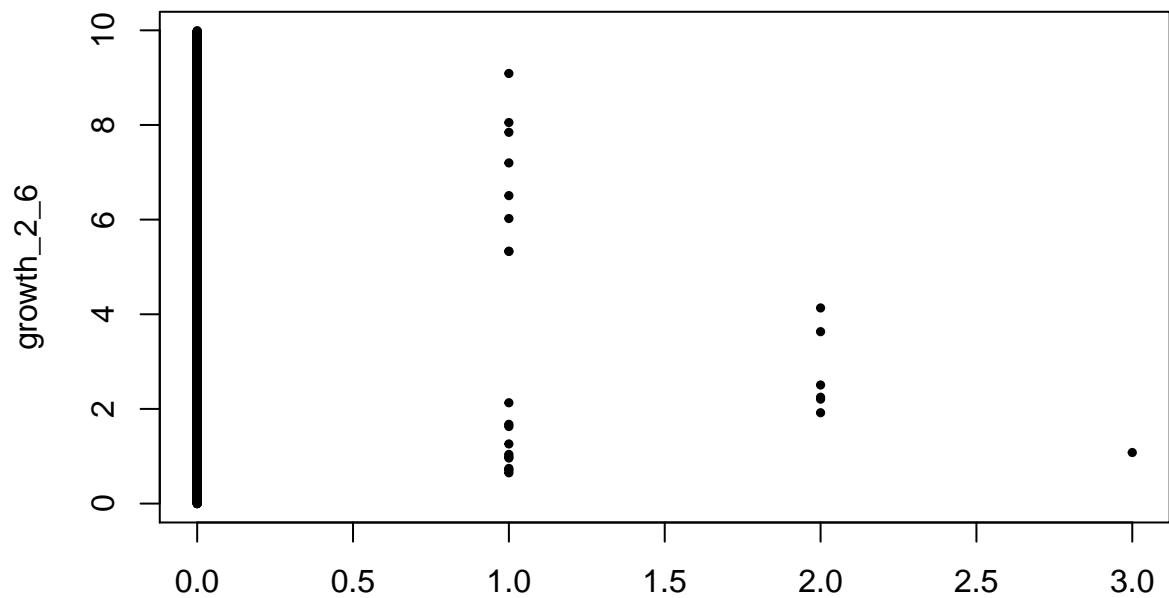


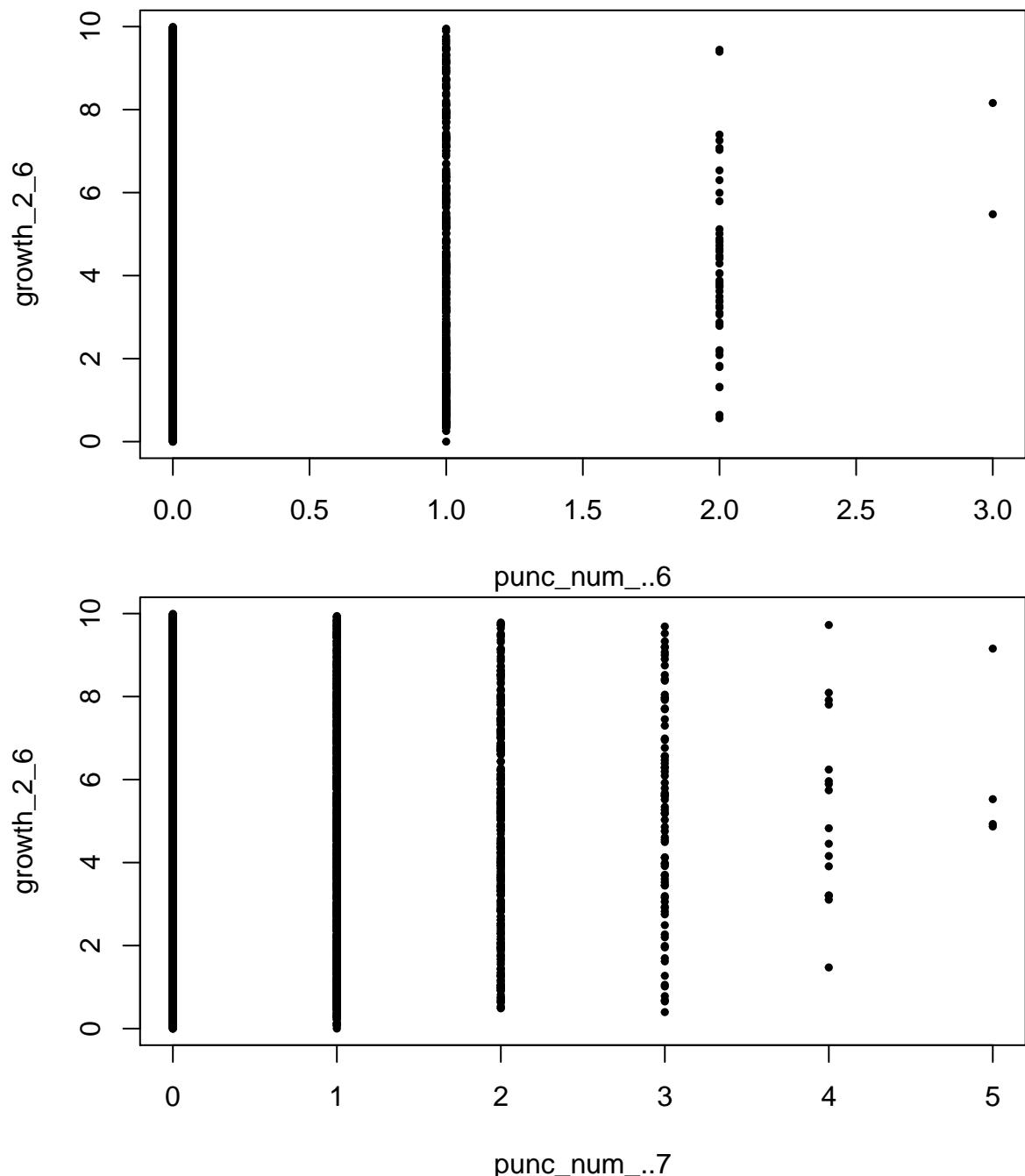


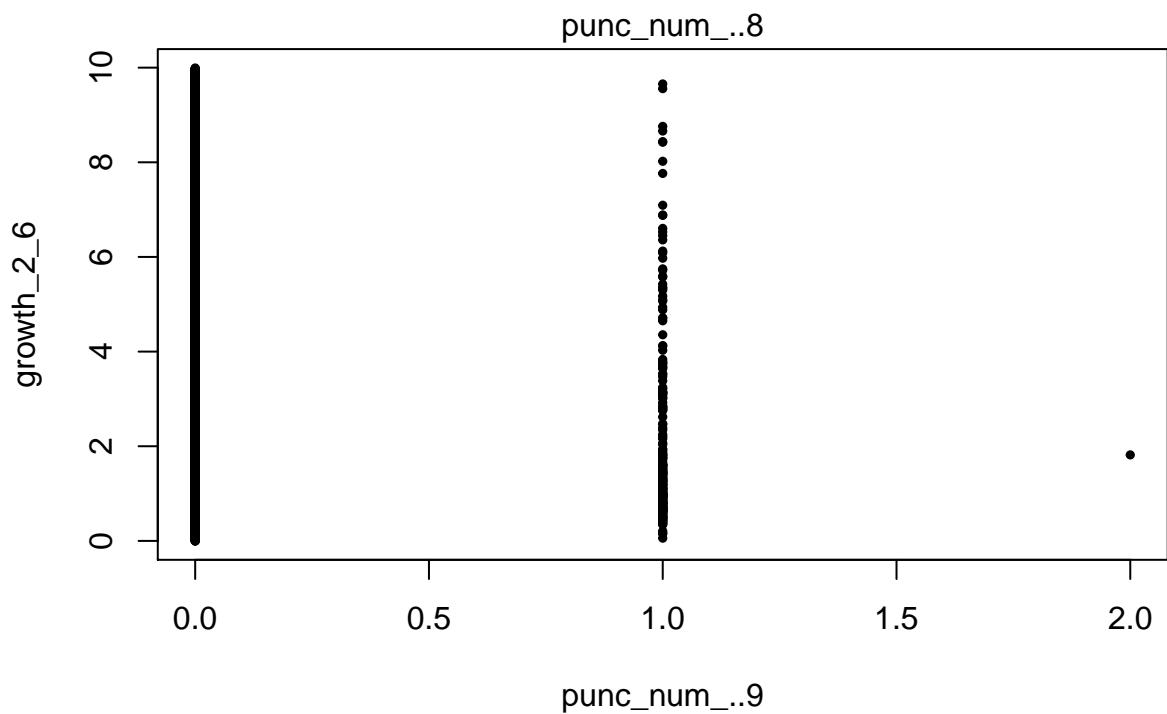
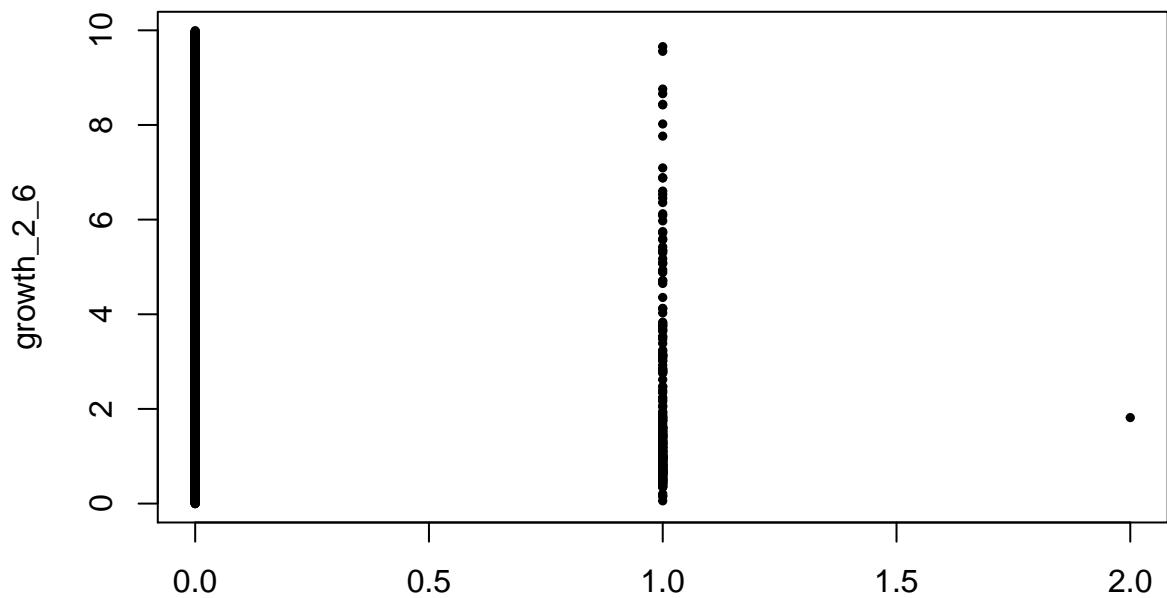
doc2vec_18

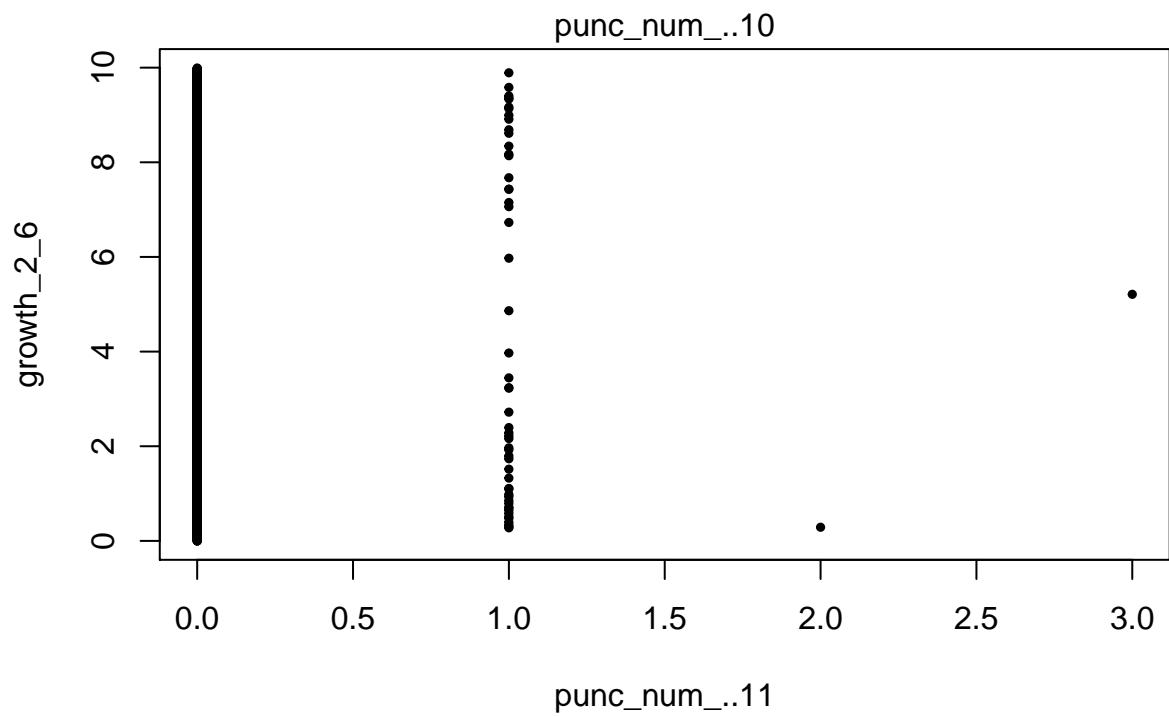
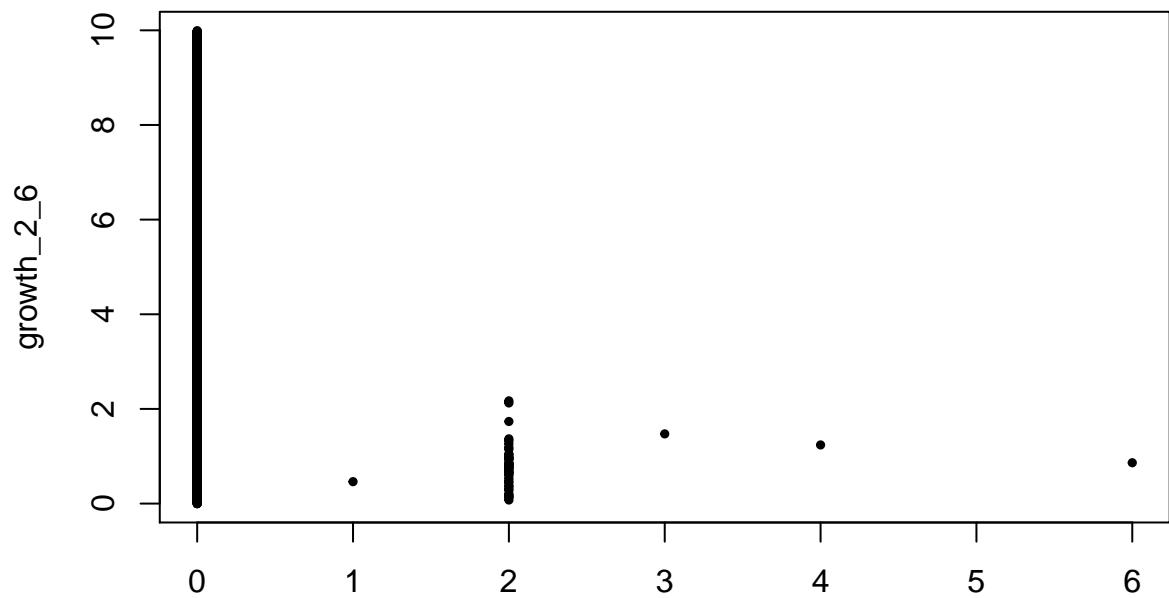


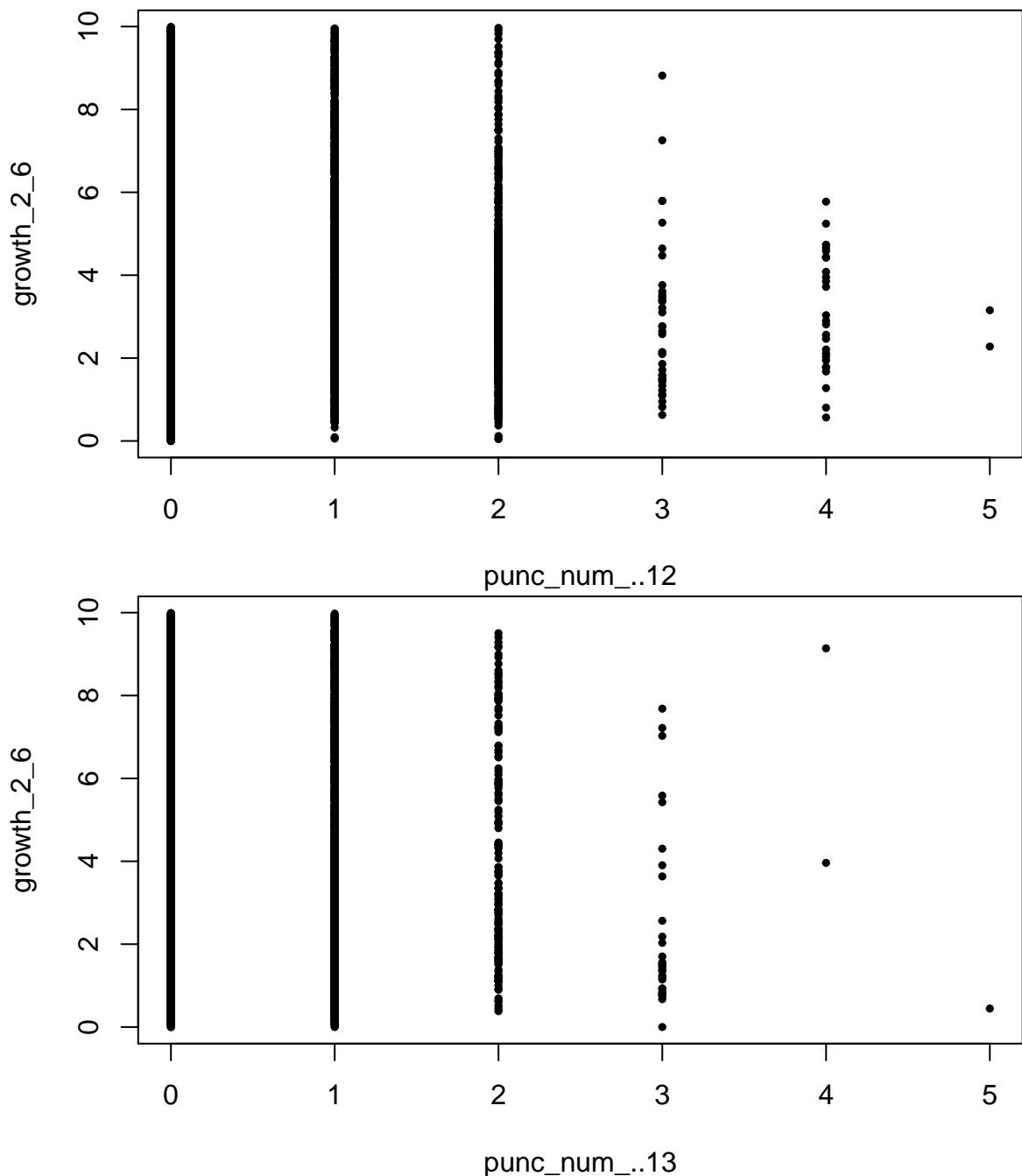


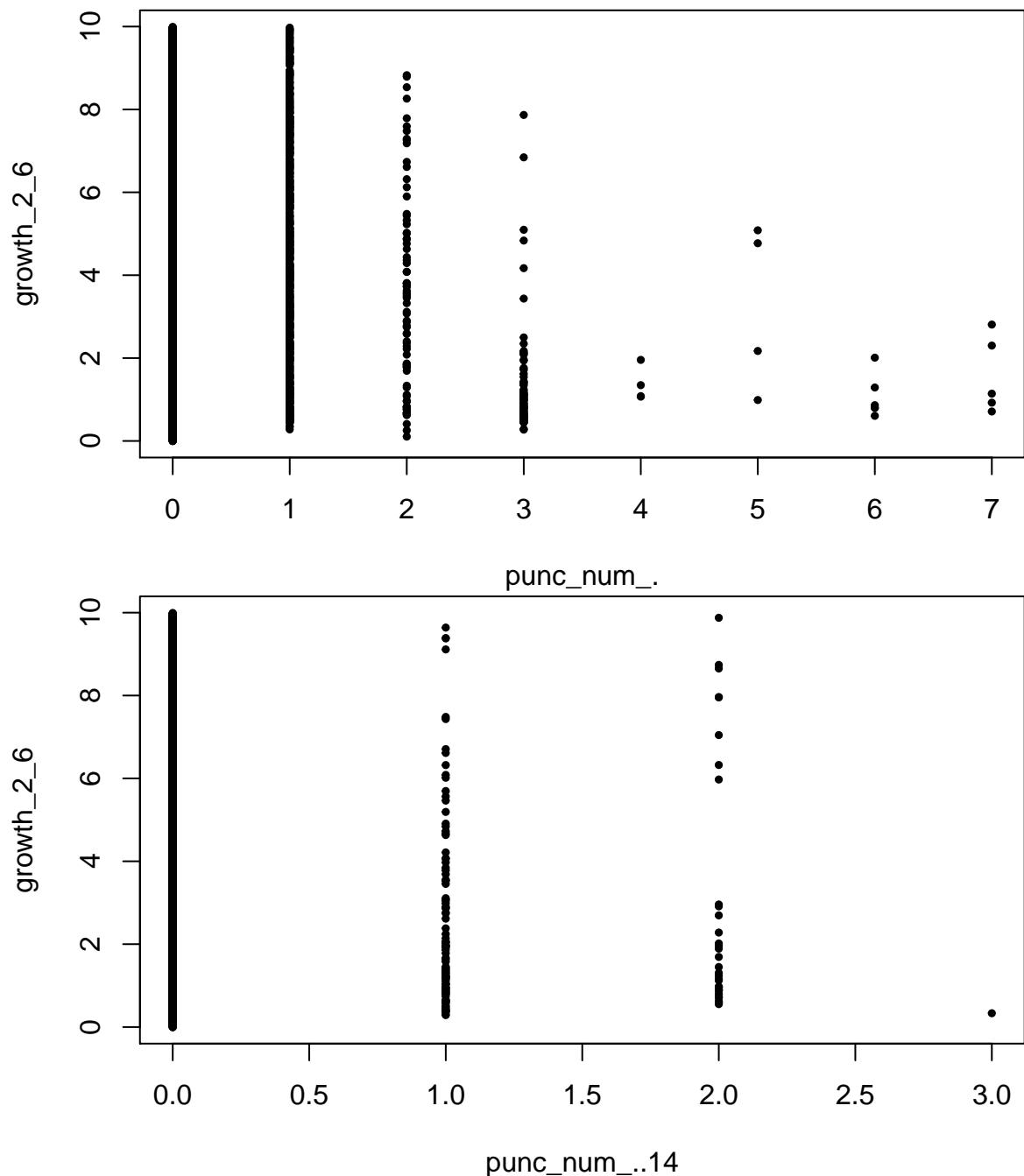


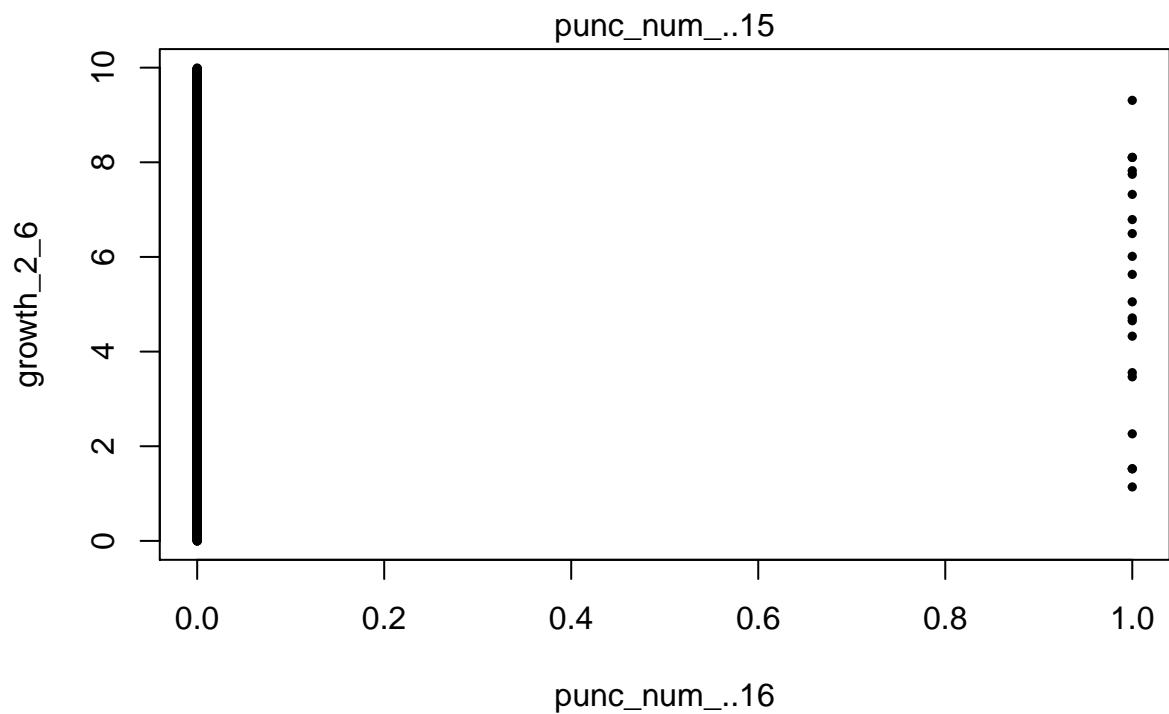
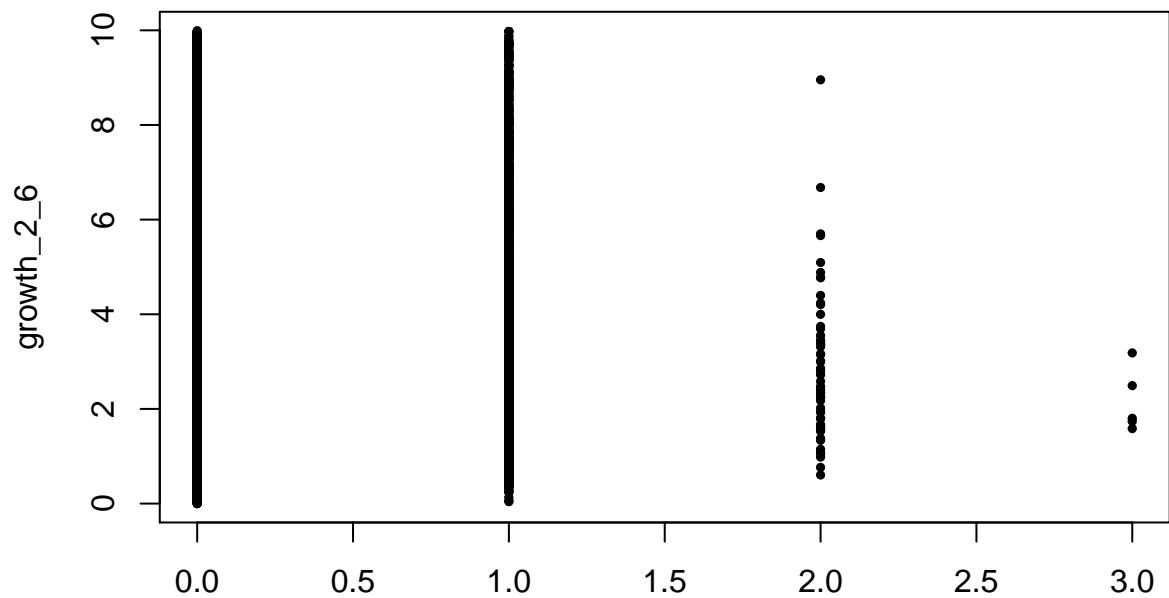


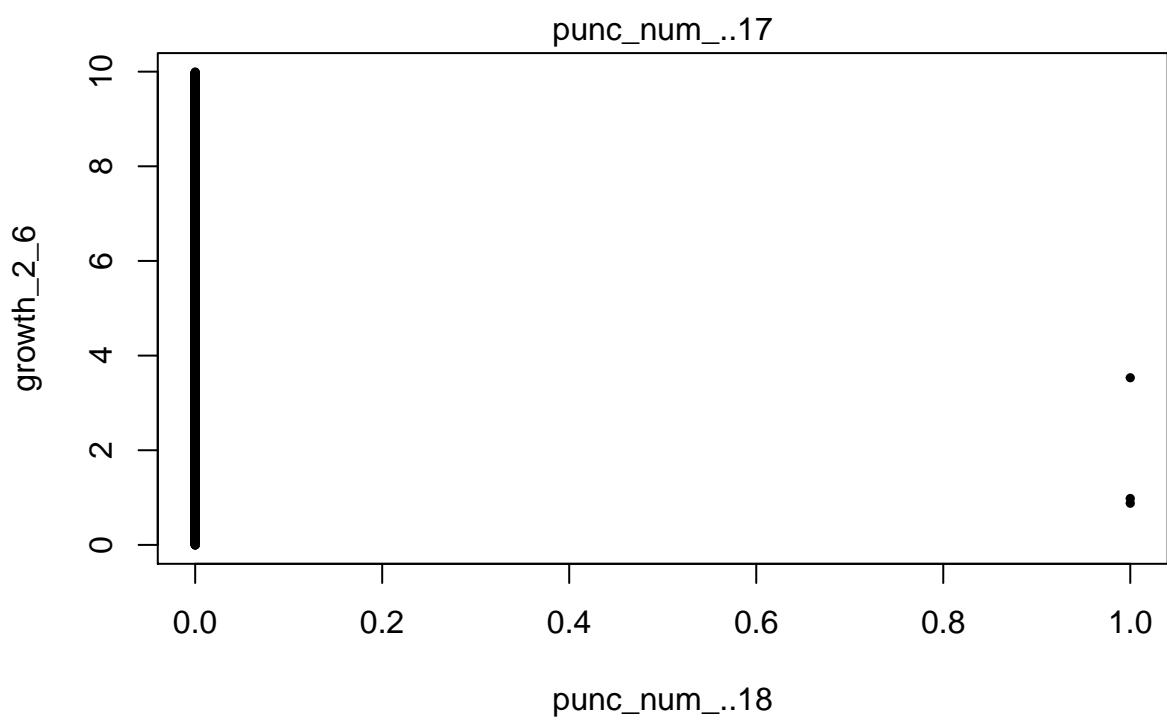
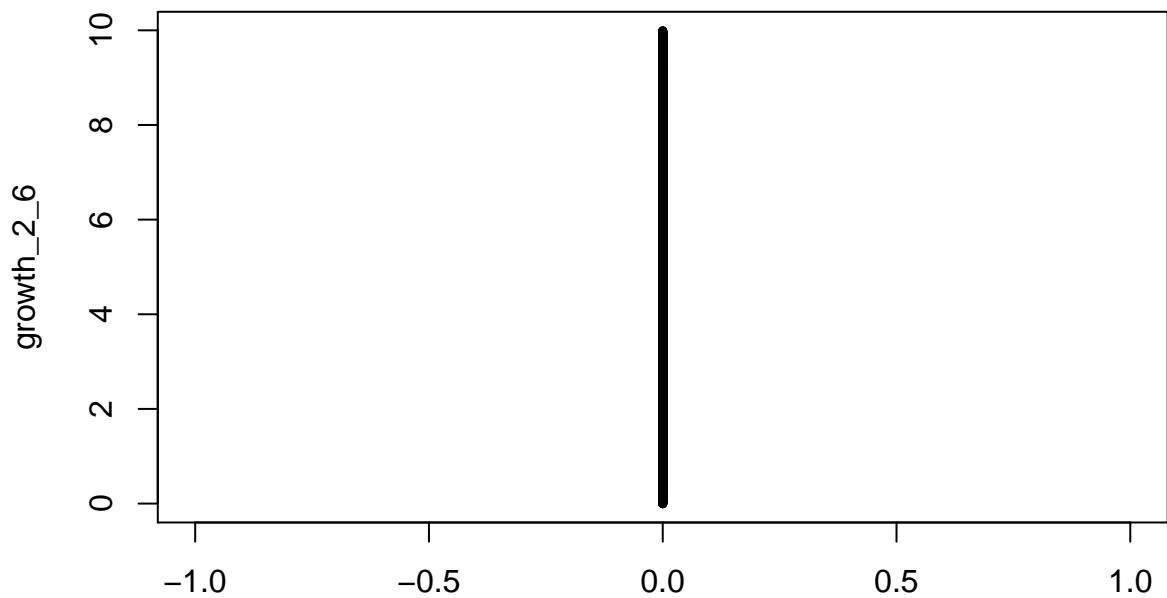


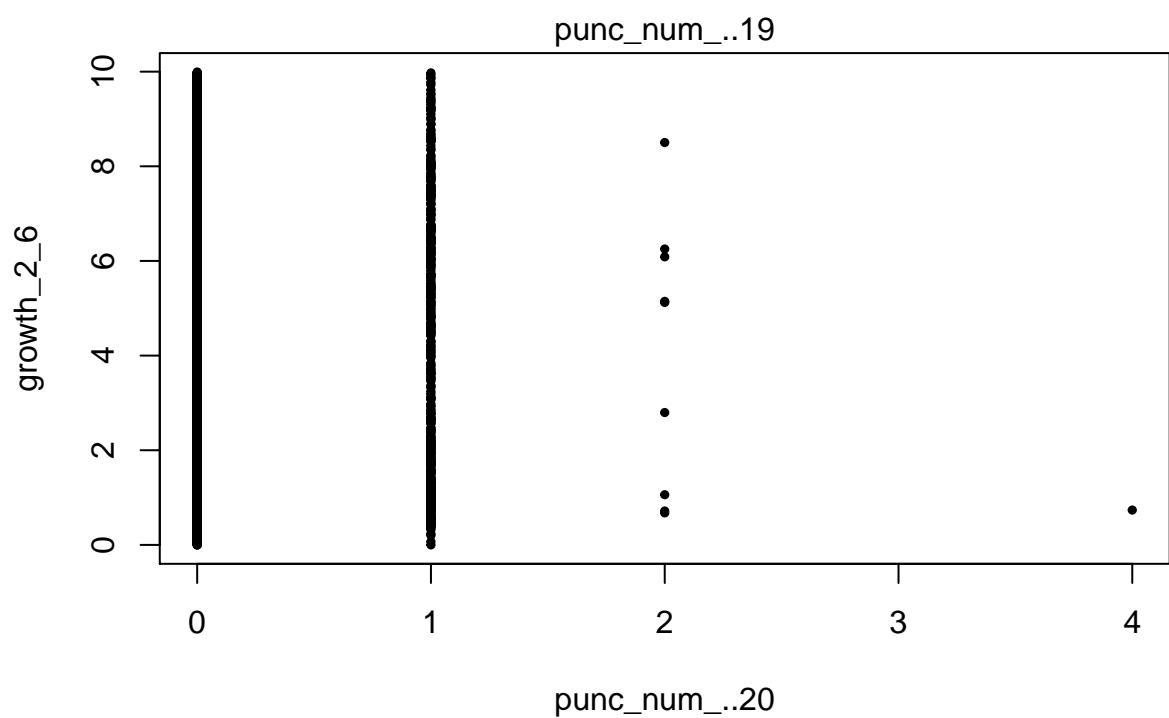
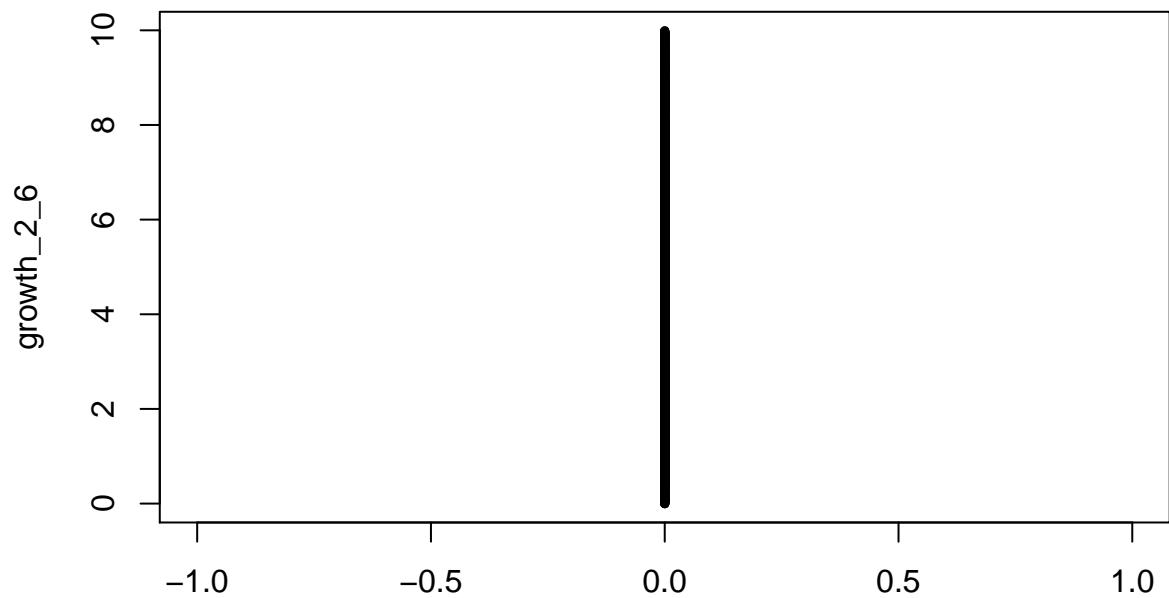


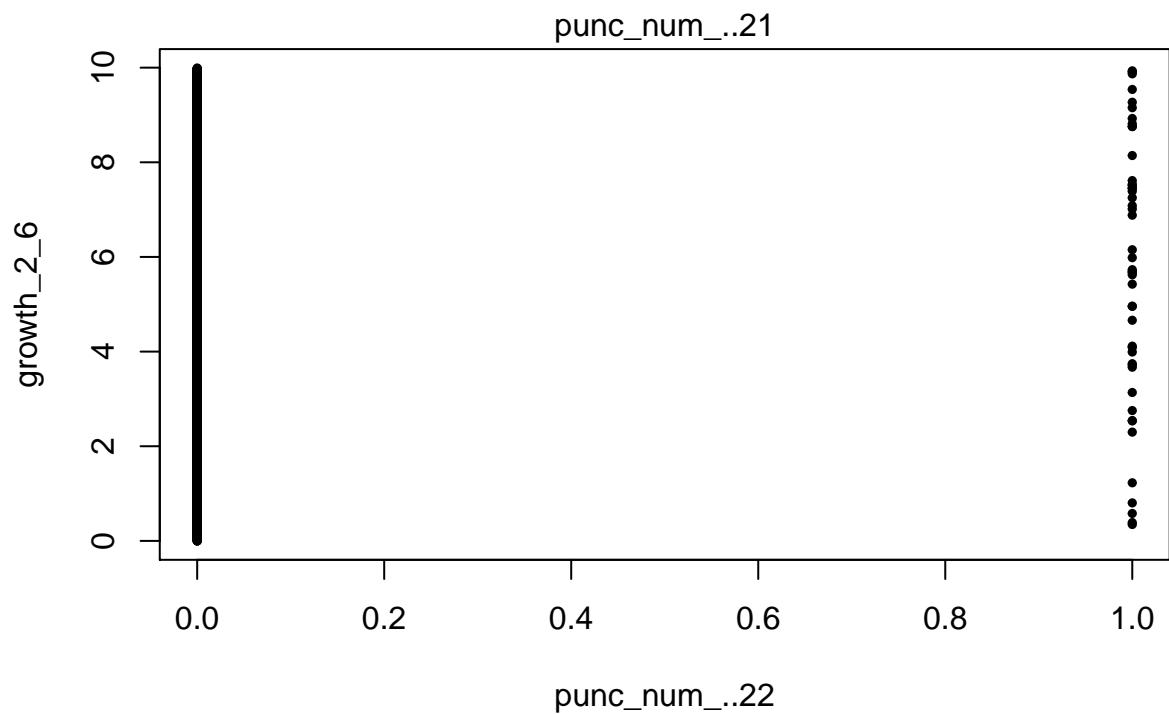
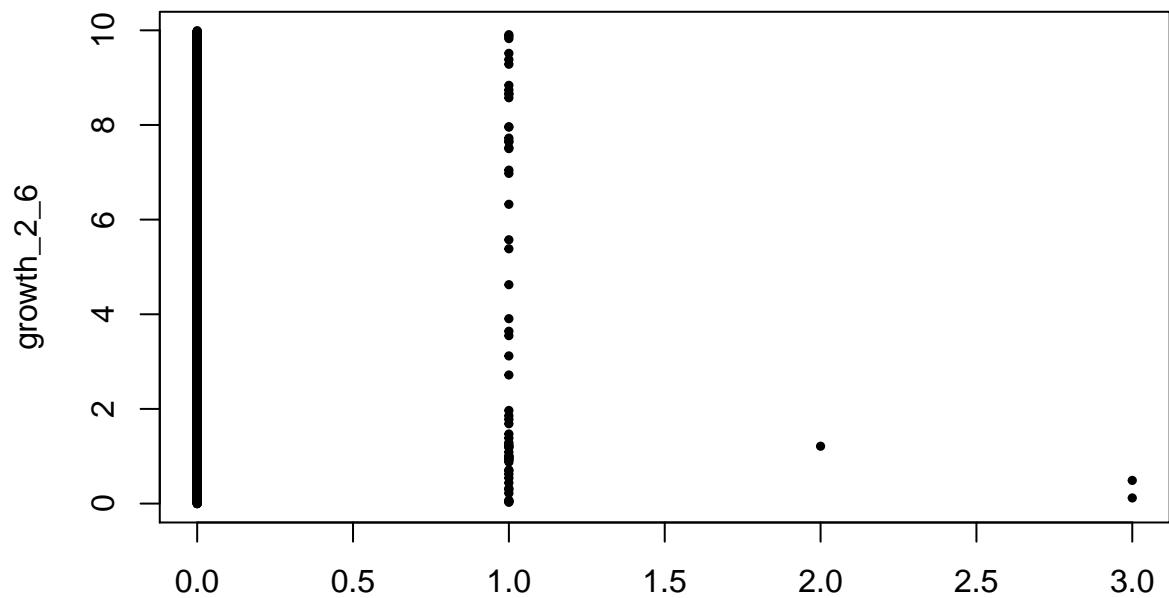


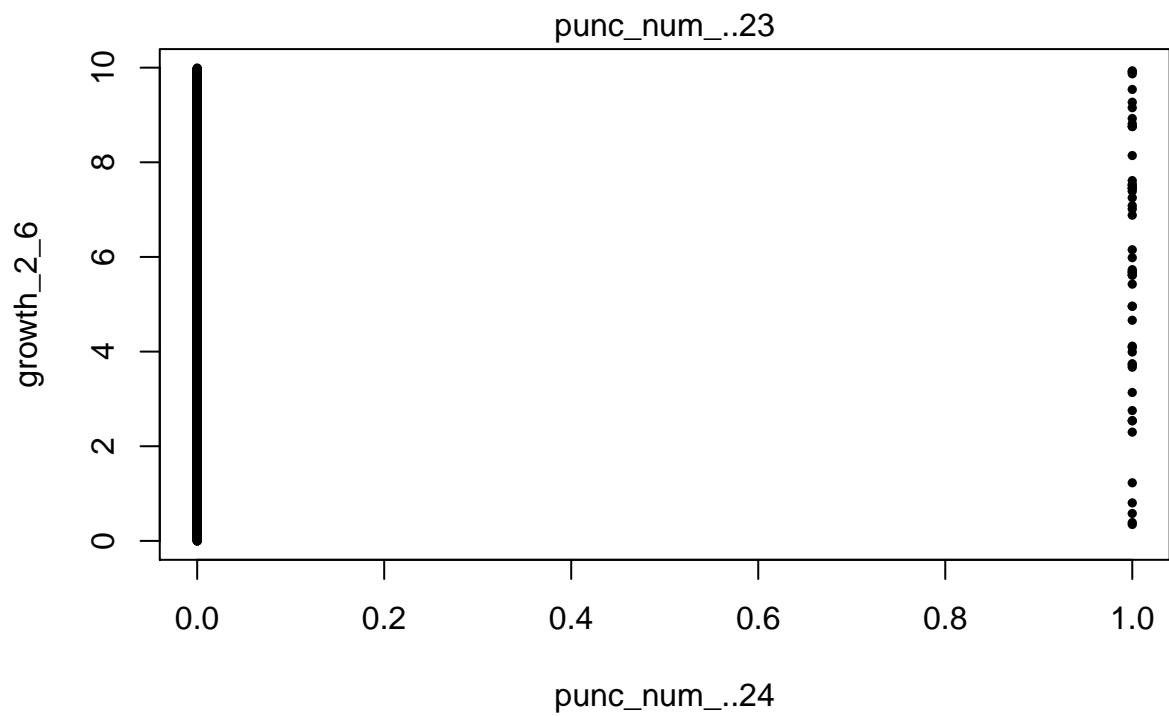
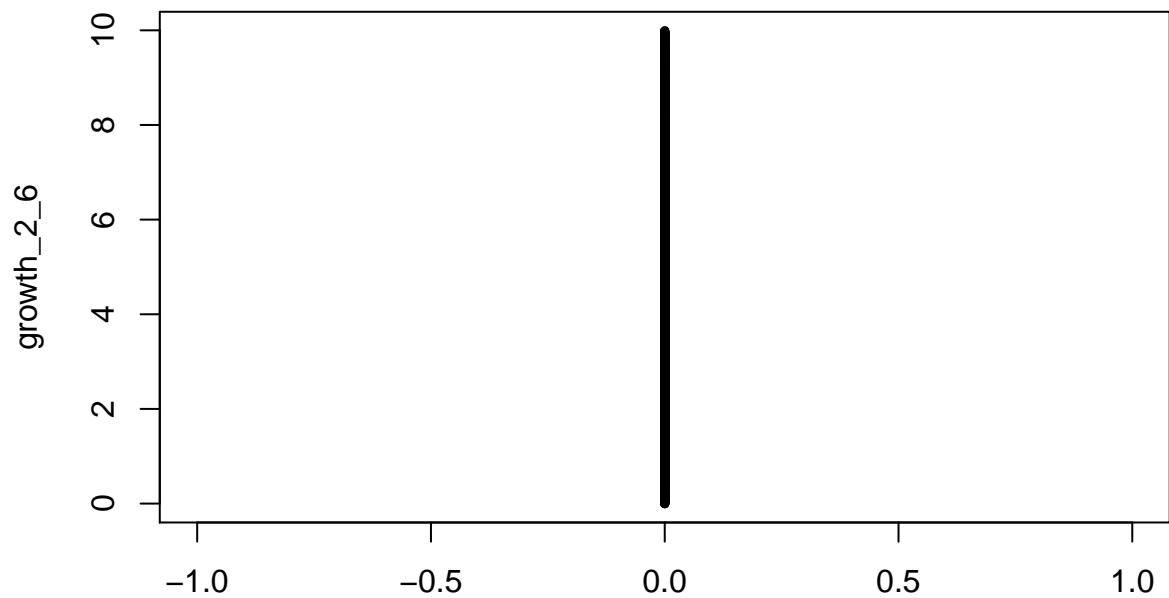


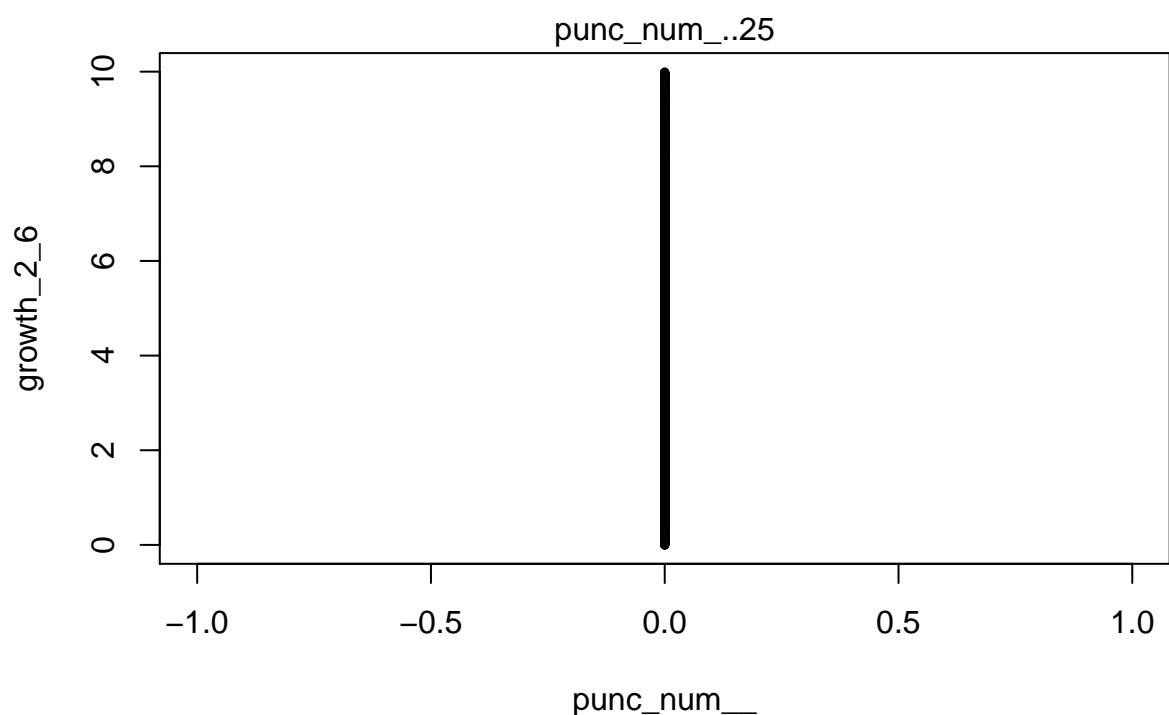
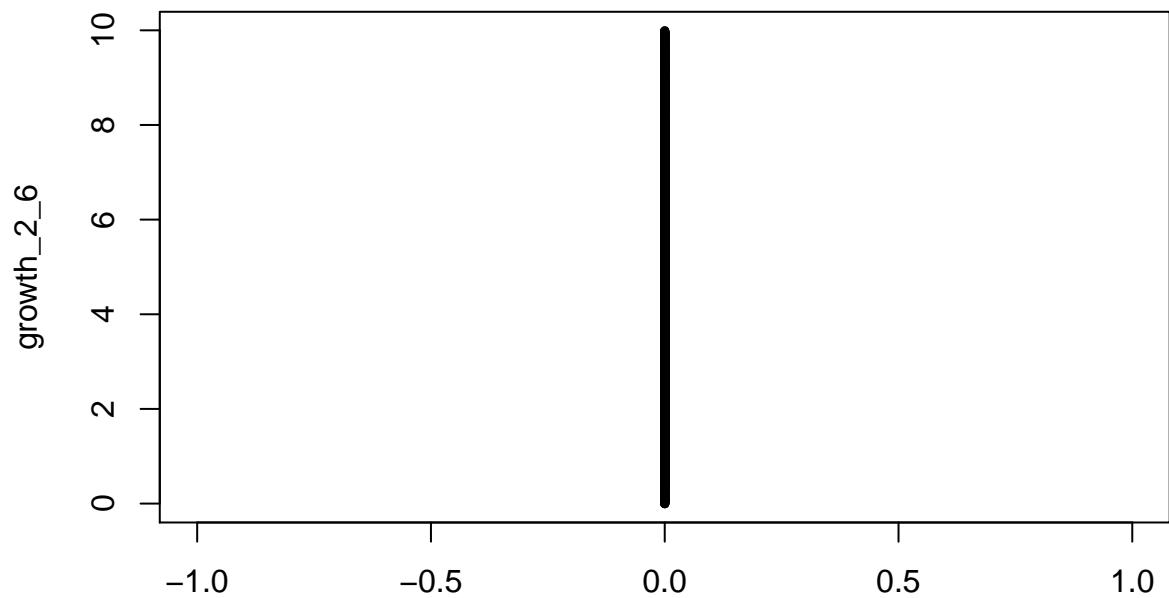


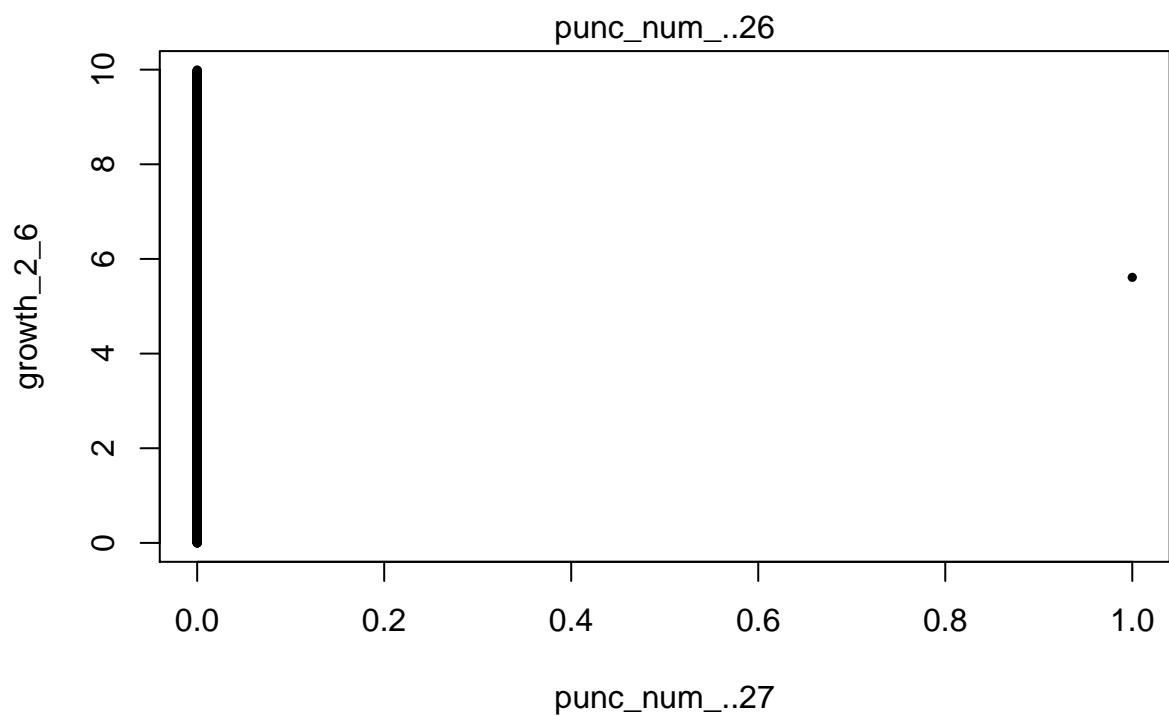
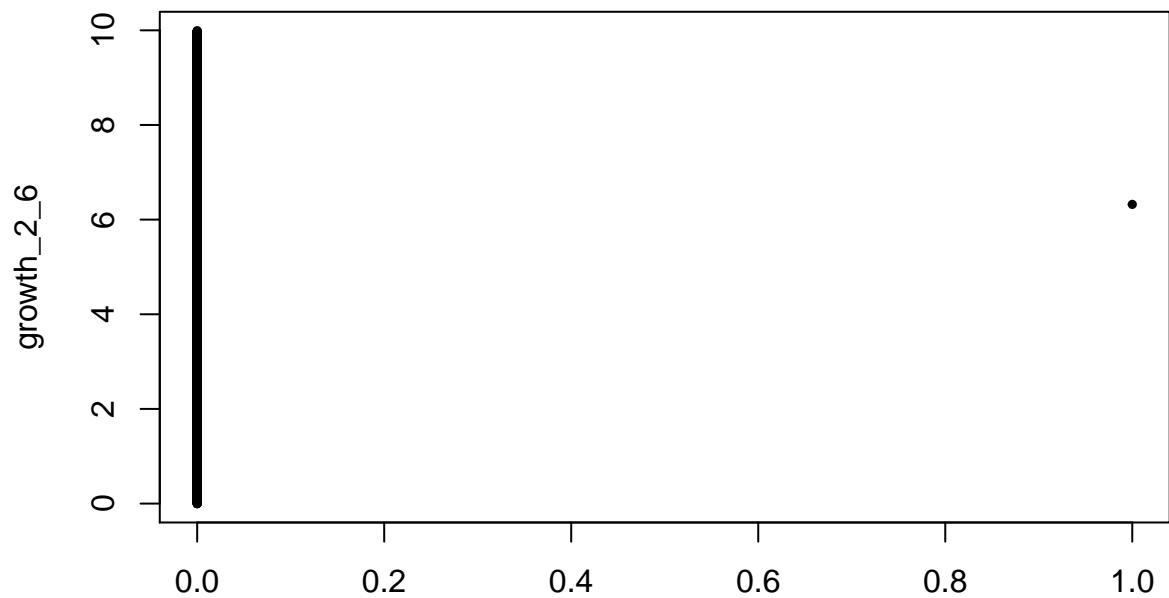


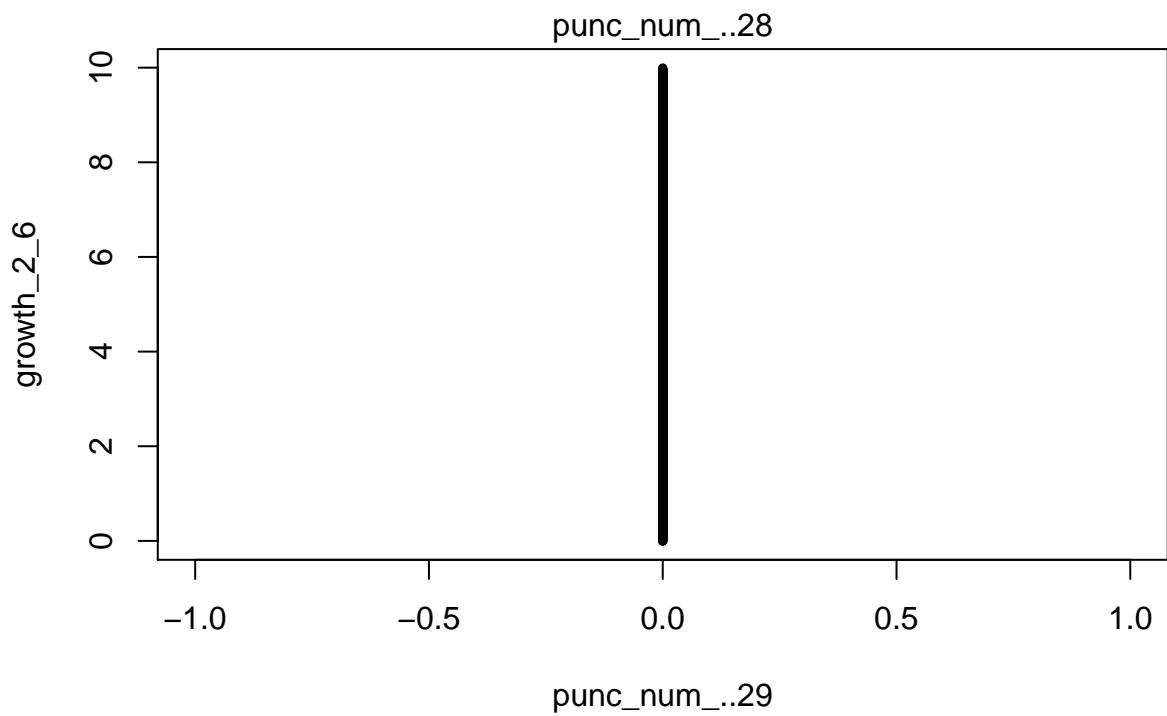
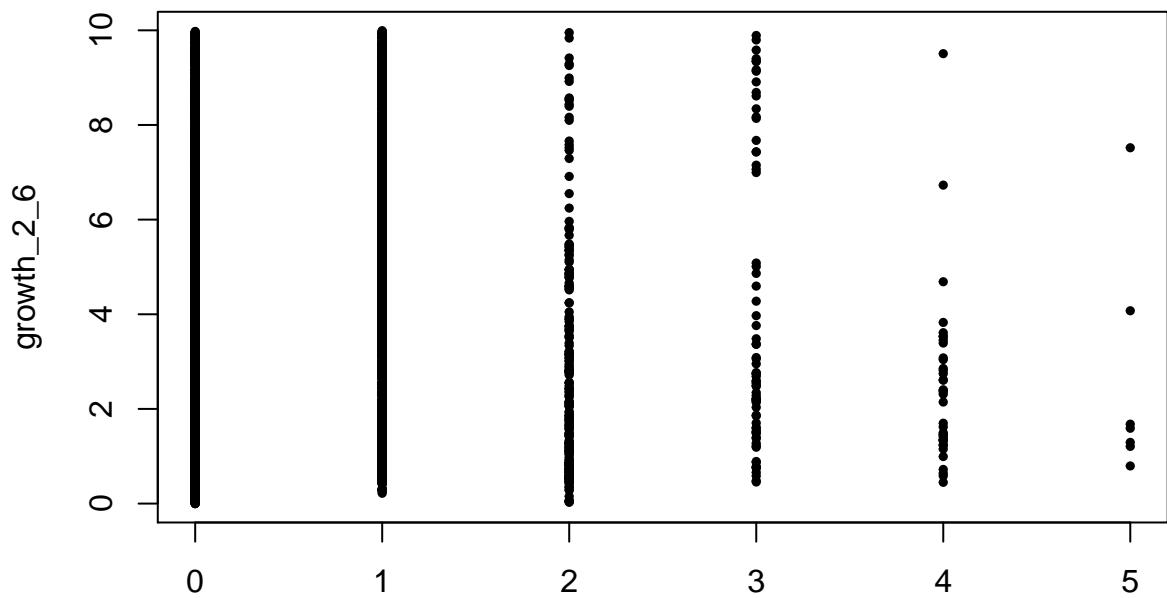


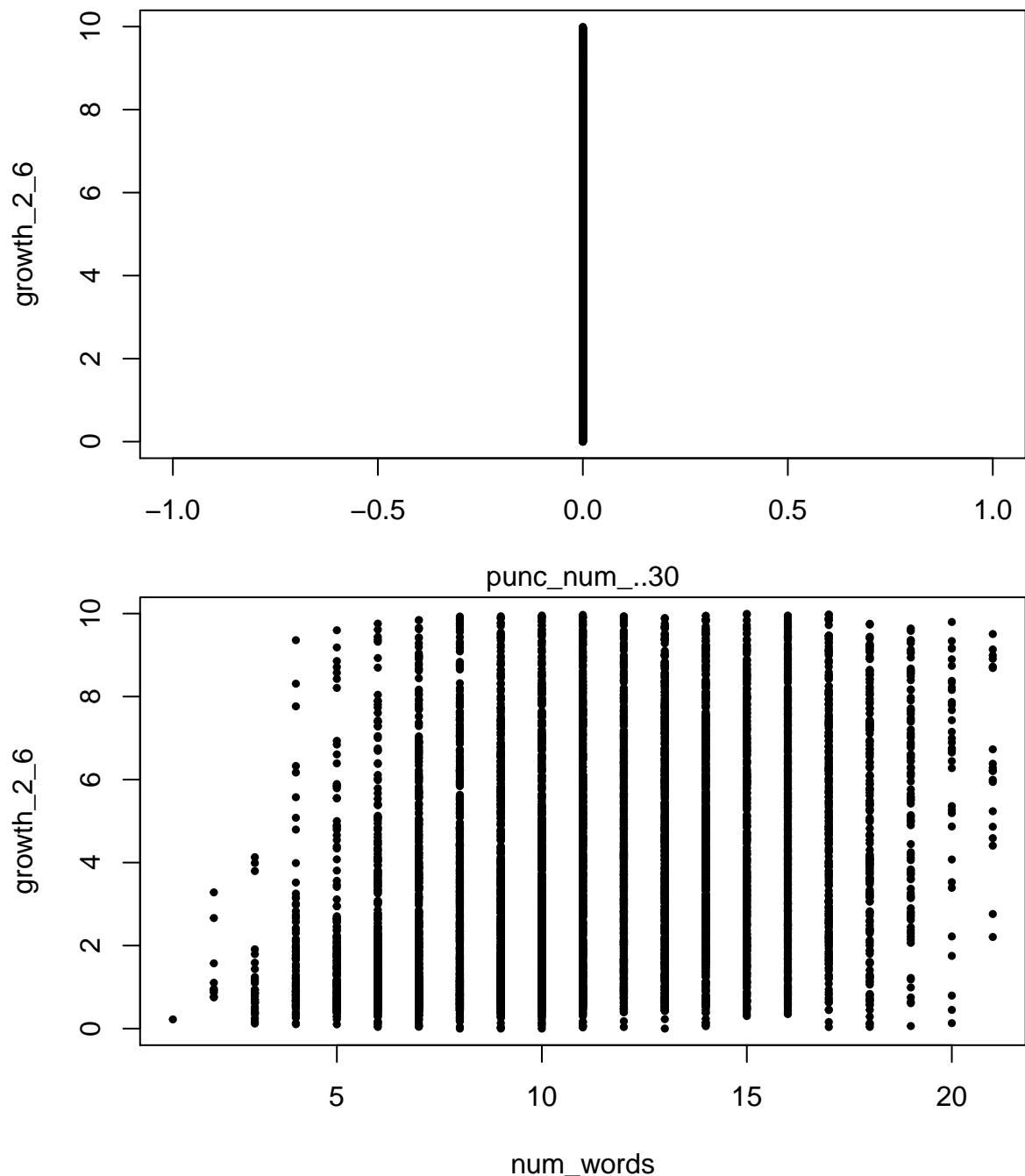


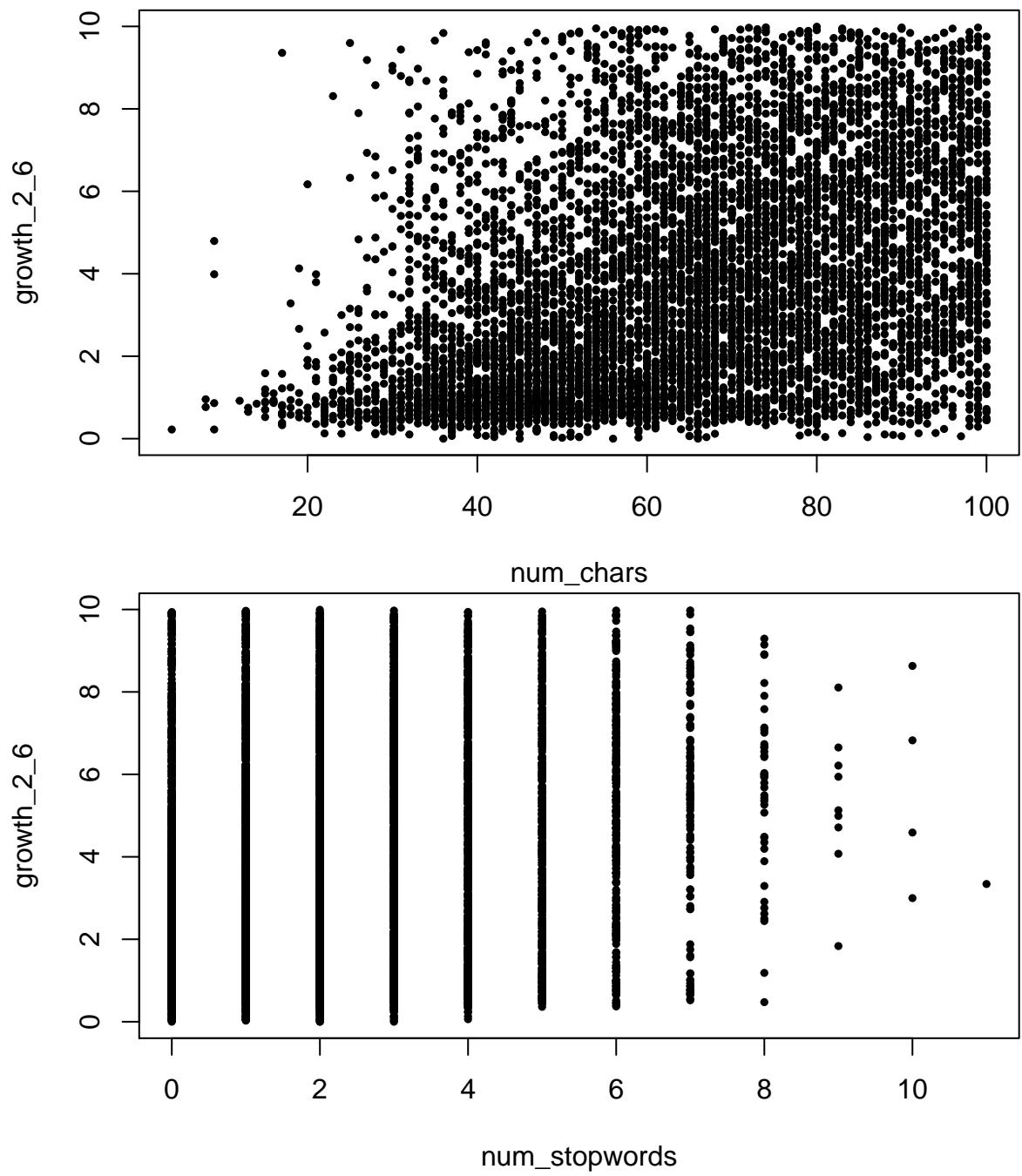


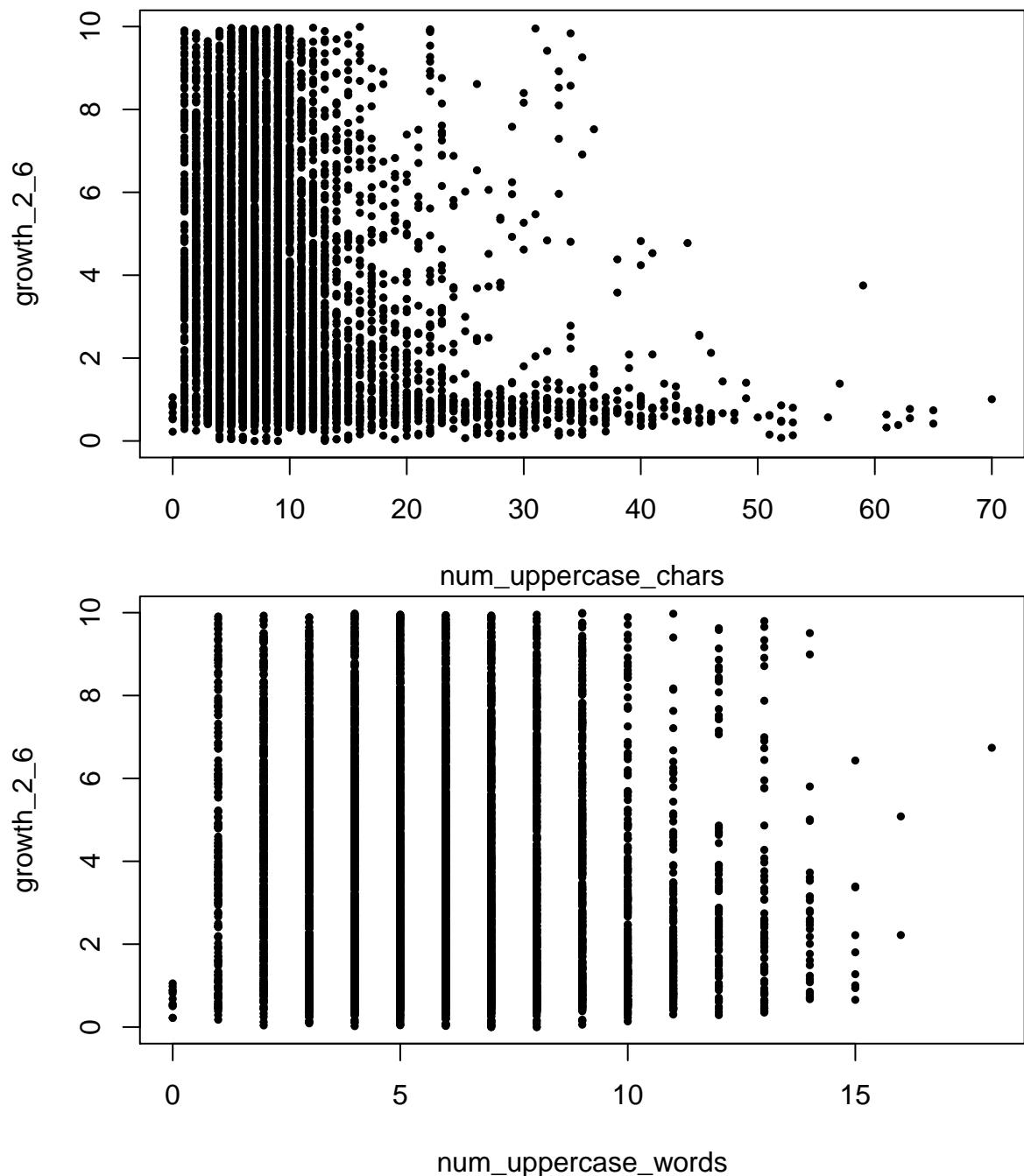


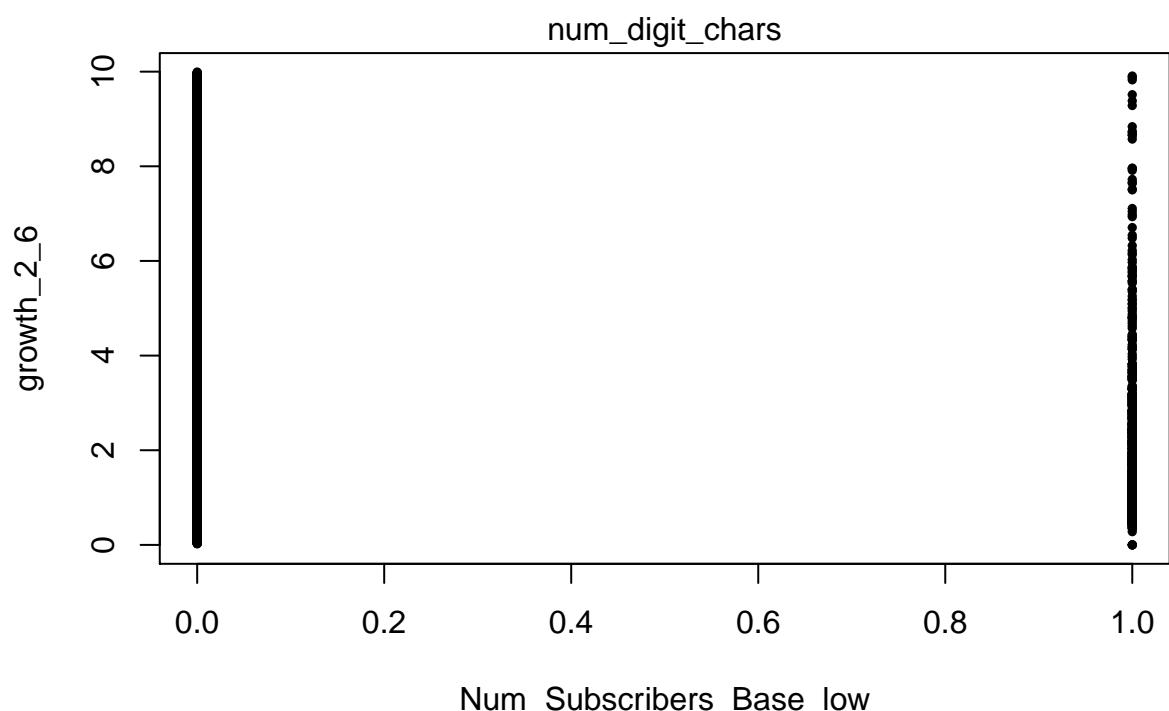
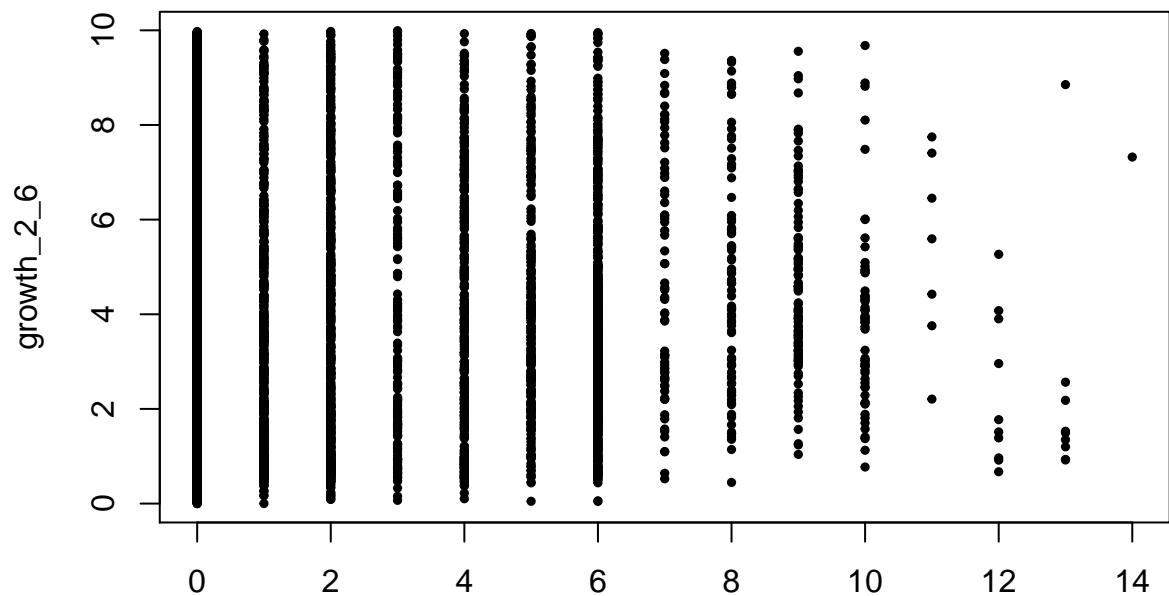


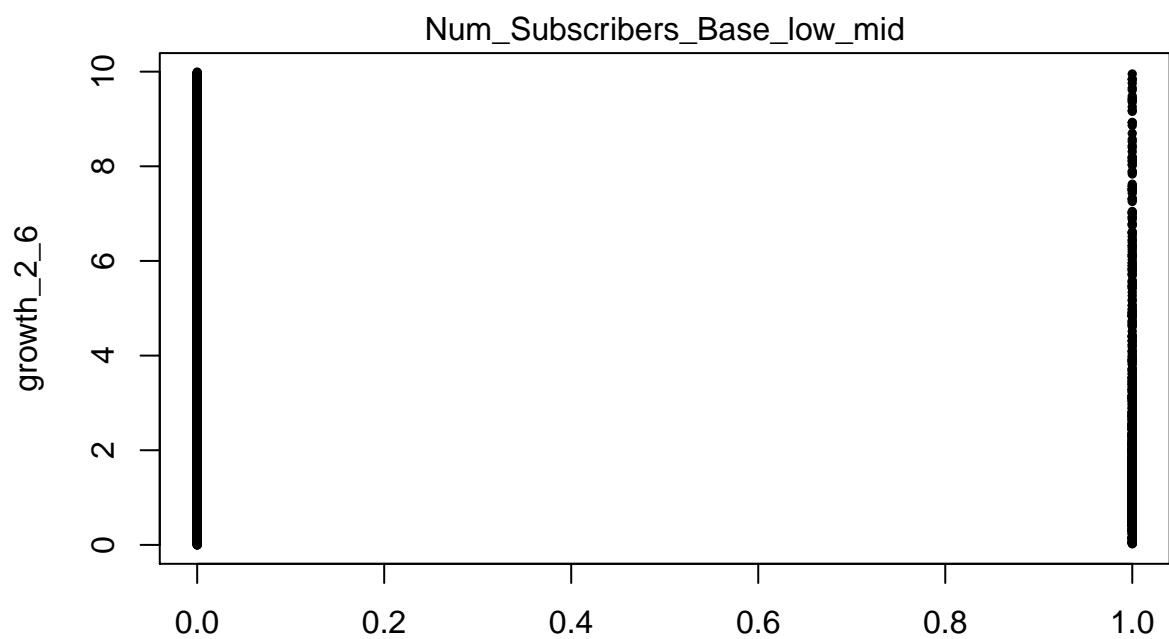
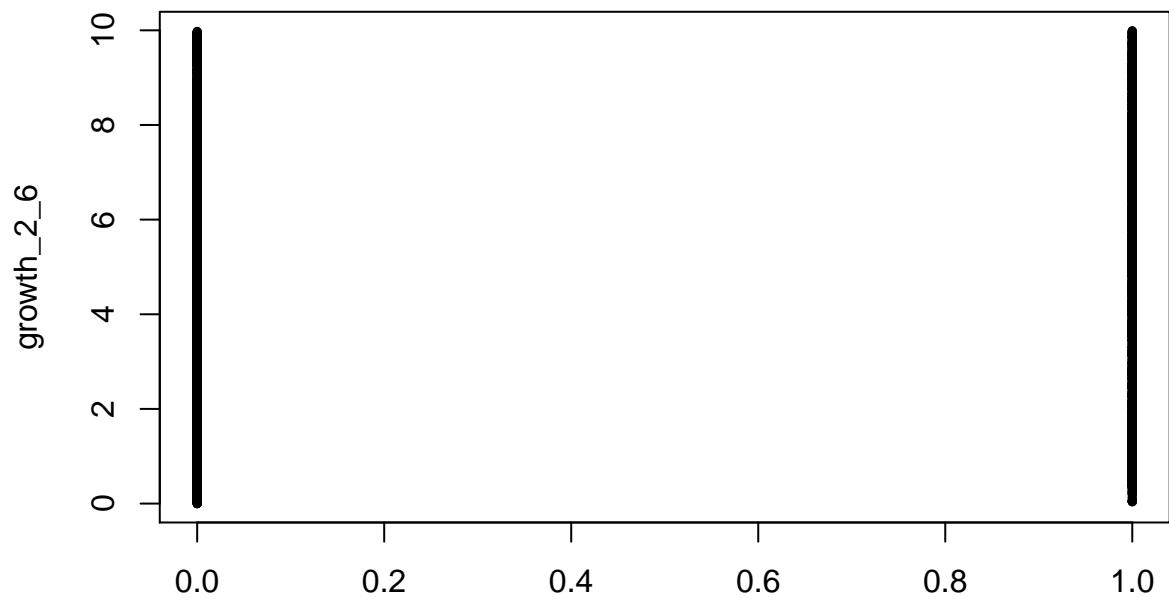












Num_Subscribers_Base_low_mid

Num_Subscribers_Base_low_mid

