Find events relative to the same ID in two tables (by dates)

```
library(data.table)


set.seed(200)
```

first table

```
date_table_1 <- data.table(hetu = paste0("hetu_",sample(x = 1:4, size = 12,replace = T)),
                        date1 = Sys.Date() + sample(x = 1:40, size = 12,replace = T),
                        what1 = paste0("value_",sample(x = letters[1:14], size = 12,replace = T)))
```

second table

```
date_table_2 <- data.table(hetu = paste0("hetu_",sample(x = 1:4, size = 12,replace = T)),
                        date2 = Sys.Date() + sample(x = 1:40, size = 12,replace = T),
                        what2 = paste0("value_",sample(x = LETTERS[1:14], size = 12,replace = T)))
```

find all what1 before what2 (without allow.cartesian = T doesnot work)

```
merge(date_table_1,
  date_table_2,
  by = "hetu",
  allow.cartesian = T)[date1<date2]
```

if want first(min) what2 after what1 (see that for date2 can be more than one date1)

```
merge(date_table_1,
  date_table_2,
  by = "hetu",
  allow.cartesian = T)[date1<date2][,.SD[which.min(date2)],list(hetu,date1)]
```

as previous, but take the last(max) date1 (if there are multiple)

```
merge(date_table_1,
  date_table_2,
  by = "hetu",
  allow.cartesian =
      T)[date1<date2][,.SD[which.min(date2)],list(hetu,date1)][,.SD[which.max(date1)],list(hetu,date2)]
```

if want last(max) what1 beforewhat2 (see that for date1 can be more than one date2)

```
merge(date_table_1,
  date_table_2,
  by = "hetu",
  allow.cartesian = T)[date1<date2][,.SD[which.max(date1)],list(hetu,date2)]
```

as previous, but take the first(min) date2 (if there are multiple)

```
merge(date_table_1,
  date_table_2,
  by = "hetu",
  allow.cartesian =
      T)[date1<date2][,.SD[which.max(date1)],list(hetu,date2)][,.SD[which.min(date2)],list(hetu,date1)]
```

Mark words containing some regexp and replace this place as "////" word "\\\\" library(stringr) begin_mark <-"////"

```
end_mark     <- "\\\\\\\\"
word_to_look <- "diagn"
this_string  <- "DIag was very Hoito-tapahtumadiagnoosit and (Diagnoosi)."
str_replace_all(this_string,
            regex(paste0("([a-zääö\\-]*",word_to_look,"[a-zääö\\-]*)"),TRUE),
            paste0(begin_mark,"\\1",end_mark))
# result
[1] "DIag was very ////Hoito-tapahtumadiagnoosit\\\\ and (////Diagnoosi\\\\)."


library(RPostgreSQL)
library(data.table)
library(ggplot2)
```

Load data into R

```
drv    <- dbDriver("PostgreSQL")

con_pg <- dbConnect(drv, dbname="ktp",user = "ktp", host="ktppg.vsshp.net",port=5432 )

diagnoosit <- data.table(dbGetQuery(con_pg," SELECT * from stage_uraods.mv_palvelu_diagnoosi where
    yksikko_nimi ~ '^K' and dg_syy_koodi is not null limit 10000"))

tilastot <-diagnoosit[,.N,list(kliniikka = yksikko_nimi,
                               diagnoosi =   substr(dg_syy_koodi,1,3),
                               vuosi     = year(alkuhetki_pvm))][N>20][order(diagnoosi)]
ggplot() + geom_bar(data = tilastot,

                aes(x = kliniikka,
                    y = N,
                    fill = diagnoosi),
                stat = "identity") + facet_grid(vuosi~.)
```