1

In this problem you will help social planner to figure out how many children a given women has at a particular date.

- Social planner does not know how many children a given women has at a particular date but she knows whether she paid childcare benefits to a given woman at a particular date.
- She knows that if in a given month there is 1 in the data, then in this month woman received childcare benefits;
- On the other hand, if in a given month there is 0 in the data then in that month woman did not receive childcare benefits;
- She also remembers that sometimes in the data there are mistakes - if the difference between two subsequent sequences of 1s is less than 5 (that is there are less than 5 0s in between two subsequent sequences of 1s), then a woman received childcare benefits for the same child and it is not the case that a new child was born

Your task is to help social planner to understand how many children a women has at any given point in time To do that you need to write a function, which implements what a social planner knows:

1. As inputs your function should take a vector of 0s and 1s;
2. Every time you see a sequence of 1s in the data you need to increase the number of children by 1;
3. Be careful with the two subsequent sequences of 1s, where the difference between them is less than 5 (i.e. when there are less than 5 0s in between them, then it is the same child and not a new child);
4. To help you social planner provides some examples of what your function should return:
    - Input: c(0,0,0,1,1,1,0,0)
    - Output: 0 0 0 1 1 1 1 1
    - Input: c(1,1,1,1,0,0,0,0)
    - Output: 1 1 1 1 1 1 1 1
    - Input: c(0,0,0,0,1,1,1,1,0,0,0,0,0,1,1,1)
    - Output: 0 0 0 0 1 1 1 1 1 1 1 1 1 2 2 2
    - Input: c(0,0,0,0,1,1,1,1,0,0,1,1,0,0,0,1,1,0,0,0,0,1,1,0,0,0,0,0,1)
    - Output: 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
    - Input: c(0,1,0,0,0,0,1,1,0,0)
    - Output: 0 1 1 1 1 1 1 1 1 1
5. Functions, which might be helpful:
    - rle();
    - diff();
    - cumsum();
    - which().

2

- In the folder hw1p5 you will find data on public procurement for one of Russian regions. More specifically you will find data on notifications;

Questions:

1. Write a function, which converts a given archive in the notifications folder into a list of data.tables. Each data.table should contain following fields:

- notificationNumber;
- versionNumber;

- createDate;
- placingWay.code;
- placingWay.name;
- order.placer.regNum;
- lots.lot.products.product.code;
- lots.lot.customerRequirements.customerRequirement.maxPrice;
- Name of the child.

  If there are several fields with the same name please join them together using “&&&&”.

2. Apply this function over all archives in the notifications folder. Note that notifications folder also contains a daily subfolder. You should consider zip files located in the daily subfolder too.
3. Convert this list of lists into (by doing necessary flattening before) into one big data.table. Keep only those observations, where the name of the child equals notificationZK or notificationEA.
4. Keep unique observations by notificationNumber, versionNumber and createDate.
5. Calculate the following:
    - Sum of lots.lot.customerRequirements.customerRequirement.maxPrice in 2011, 2012 and 2013;
    - Number of auction procedures in 2011, 2012 and 2013;
    - notificationNumber where the buyer bought the most goods (as proxied by the number of lots.lot.products.product.code);
    - Distribution of placingWay.code by year.