

class sorting

Doe Emmanuel

09/21/2020

R Markdown

This is an R Markdown document for easy readability. Codes are also attached

```
#set working directory
setwd("C:/Users/EMMANUEL/Documents/my projects")

#load readxl function
library(readxl)

#Import excel dataset
excel_sheets("listalphaorder2021.xlsx")
```

```
## [1] "Sheet1"
```

```
read_excel("listalphaorder2021.xlsx")
```

```
## New names:
```

```
## * ' -> ...1
## * ' -> ...2
## * ' -> ...4
## * ' -> ...5
## * ' -> ...6
```

```
## # A tibble: 423 x 6
```

```
##   ...1 ...2 'HONEY HOME CHILDREN'S NUR. AND~ ...4 ...5 ...6
##   <chr> <chr> <chr> <chr> <chr> <chr>
## 1 <NA> <NA> 11 NURU ONIWO ST. AGUDA SURULER~ <NA> <NA> <NA>
## 2 <NA> <NA> <NA> <NA> <NA> <NA>
## 3 <NA> <NA> LIST OF PUPILS FOR 1st TERM 202~ <NA> <NA> <NA>
## 4 <NA> <NA> <NA> <NA> <NA> <NA>
## 5 <NA> <NA> <NA> <NA> <NA> <NA>
## 6 S/No NAME CLASS SCHOLA~ ON REB~ COMPUT~
## 7 1 ABAH CHIAMAKA NURSERY 2A -- CHIOMA NA <NA> KN0001~
## 8 2 ABAH OGOCHUKWU PRIMARY 3C -- 00 NA <NA> BS0002~
## 9 3 ADEBAYO AMIRA~ PRIMARY 5D -- 00 NA <NA> BS0001~
## 10 4 ADEBOWALE OLA~ PRIMARY 5B -- ENIONG NA <NA> BS0001~
## # ... with 413 more rows
```

```
alphaorderlist <- read_excel("listalphaorder2021.xlsx", skip = 8)
alphaorderlist
```

```
## # A tibble: 417 x 6
##   'S/No' NAME          CLASS          SCHOLARSHIP 'ON REBATE?' 'COMPUTER No'
##   <dbl> <chr>          <chr>          <chr>        <lgl>        <chr>
## 1      1 ABAH CHIAMAKA  NURSERY 2A --- NA          NA          KN0001820
## 2      2 ABAH OGOCHUKWU PRIMARY 3C --- NA          NA          BS0002110
## 3      3 ADEBAYO AMIRAT ~ PRIMARY 5D --- NA          NA          BS0001910
## 4      4 ADEBOWALE OLAOL~ PRIMARY 5B --- NA          NA          BS0001871
## 5      5 ADELEKE ADEDAYO~ PRIMARY 5C --- NA          NA          BS0001842
## 6      6 ADESANYA DAVID  PRIMARY 6A --- NA          NA          BS0001775
## 7      7 ADESOYE IBRAHEEM PRIMARY 1A --- <NA>      NA          BS0002221
## 8      8 ADETUNMBI DEMIL~ PRIMARY 5A --- NA          NA          BS0001900
## 9      9 ADETUNMBI FUNMI~ PRIMARY 3A --- NA          NA          BS0002038
## 10     10 ADEWOLE DAVID  PRIMARY 6C --- NA          NA          BS0001735
## # ... with 407 more rows
```

```
#Load libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
library(stringr)

#rename serial number column
alphaorderlist <- alphaorderlist %>% rename(s_no = 'S/No')

#separte the class teachers from the class
donn <- alphaorderlist %>%
  separate(CLASS, into = c("class", "teacher"), sep = "--")
donn
```

```
## # A tibble: 417 x 7
##   s_no NAME          class      teacher SCHOLARSHIP 'ON REBATE?' 'COMPUTER No'
##   <dbl> <chr>          <chr>      <chr>      <chr>        <lgl>        <chr>
## 1      1 ABAH CHIAMAKA "NURSERY~ " CHIOM~ NA          NA          KN0001820
## 2      2 ABAH OGOCHUK~ "PRIMARY~ " 00"    NA          NA          BS0002110
## 3      3 ADEBAYO AMIR~ "PRIMARY~ " 00"    NA          NA          BS0001910
## 4      4 ADEBOWALE OL~ "PRIMARY~ " ENION~ NA          NA          BS0001871
## 5      5 ADELEKE ADED~ "PRIMARY~ " BAMID~ NA          NA          BS0001842
```

```
## 6      6 ADESANYA DAV~ "PRIMARY~ " AYO"    NA          NA          BS0001775
## 7      7 ADESOYE IBRA~ "PRIMARY~ " SHOGB~ <NA>    NA          BS0002221
## 8      8 ADETUNMBI DE~ "PRIMARY~ " SALAM~ NA          NA          BS0001900
## 9      9 ADETUNMBI FU~ "PRIMARY~ " OPARA" NA          NA          BS0002038
## 10     10 ADEWOLE DAVID "PRIMARY~ " OO"      NA          NA          BS0001735
## # ... with 407 more rows
```

```
#Remove class alphabets from classes
alphaorderclassonly <-donn %>%
  mutate(class = str_remove(class, "[A-Z]\\s$"))
alphaorderclassonly
```

```
## # A tibble: 417 x 7
##   s_no NAME          class teacher SCHOLARSHIP 'ON REBATE?' 'COMPUTER No'
##   <dbl> <chr>          <chr> <chr> <chr> <lgl> <chr>
## 1      1 ABAH CHIAMAKA NURSER~ " CHIOMA" NA          NA          KN0001820
## 2      2 ABAH OGOCHUKWU PRIMAR~ " OO"      NA          NA          BS0002110
## 3      3 ADEBAYO AMIRA~ PRIMAR~ " OO"      NA          NA          BS0001910
## 4      4 ADEBOWALE OLA~ PRIMAR~ " ENIONG" NA          NA          BS0001871
## 5      5 ADELEKE ADEDA~ PRIMAR~ " BAMIDE~ NA          NA          BS0001842
## 6      6 ADESANYA DAVID PRIMAR~ " AYO"      NA          NA          BS0001775
## 7      7 ADESOYE IBRAH~ PRIMAR~ " SHOGBE~ <NA>    NA          BS0002221
## 8      8 ADETUNMBI DEM~ PRIMAR~ " SALAMI" NA          NA          BS0001900
## 9      9 ADETUNMBI FUN~ PRIMAR~ " OPARA" NA          NA          BS0002038
## 10     10 ADEWOLE DAVID PRIMAR~ " OO"      NA          NA          BS0001735
## # ... with 407 more rows
```

```
alphaorderclassonly <- alphaorderclassonly %>%
  separate(NAME, into = c("Surname", "First_name", "Other_names"), sep = " ")
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 10 rows [71, 72, 73,
## 143, 182, 349, 381, 413, 414, 415].
```

```
## Warning: Expected 3 pieces. Missing pieces filled with 'NA' in 315 rows [1, 2,
## 4, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, ...].
```

```
alphaorderclassonly
```

```
## # A tibble: 417 x 9
##   s_no Surname First_name Other_names class teacher SCHOLARSHIP 'ON REBATE?'
##   <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <lgl>
## 1      1 ABAH CHIAMAKA <NA> NURS~ " CHIO~ NA          NA
## 2      2 ABAH OGOCHUKWU <NA> PRIM~ " OO" NA          NA
## 3      3 ADEBAYO AMIRAT OLATOKUNBO PRIM~ " OO" NA          NA
## 4      4 ADEBOW~ OLAOLUWA <NA> PRIM~ " ENIO~ NA          NA
## 5      5 ADELEKE ADEDAYO ODINAKA PRIM~ " BAMI~ NA          NA
## 6      6 ADESAN~ DAVID <NA> PRIM~ " AYO" NA          NA
## 7      7 ADESOYE IBRAHEEM <NA> PRIM~ " SHOG~ <NA> NA
## 8      8 ADETUN~ DEMILADE <NA> PRIM~ " SALA~ NA          NA
## 9      9 ADETUN~ FUNMILADE <NA> PRIM~ " OPAR~ NA          NA
## 10     10 ADEWOLE DAVID <NA> PRIM~ " OO" NA          NA
## # ... with 407 more rows, and 1 more variable: 'COMPUTER No' <chr>
```

```
#check for distinct classes
alphaorderclassonly %>%
  count(class)
```

```
## # A tibble: 10 x 2
##   class      n
##   <chr>    <int>
## 1 KG 1      7
## 2 KG 2     23
## 3 NURSERY 1  23
## 4 NURSERY 2  32
## 5 PRIMARY 1  51
## 6 PRIMARY 2  46
## 7 PRIMARY 3  59
## 8 PRIMARY 4  41
## 9 PRIMARY 5  75
## 10 PRIMARY 6  60
```

```
#gets the count of people whose surname appear more than once(they are most likely siblings)
same_surname <- alphaorderclassonly %>%
  count(Surname) %>%
  filter(n > 1) %>%
  arrange(desc(n))
same_surname
```

```
## # A tibble: 111 x 2
##   Surname      n
##   <chr>    <int>
## 1 OKAFOR     12
## 2 OKEKE      7
## 3 OKONKWO    6
## 4 ETONIRU    5
## 5 ENIKANLOGBON 4
## 6 NWACHUKWU  4
## 7 NWAOKOLO   4
## 8 NWODOH     4
## 9 OKOYE      4
## 10 OLUFEMI   4
## # ... with 101 more rows
```

```
#summing helps to know the total number of students in this category
sum(same_surname$n)
```

```
## [1] 293
```

```
#this shows(filters) the students with siblings
siblings <- alphaorderclassonly %>%
  filter(Surname %in% same_surname$Surname)
siblings
```

```
## # A tibble: 293 x 9
```

```
##      s_no Surname First_name Other_names class teacher SCHOLARSHIP 'ON REBATE?'
##      <dbl> <chr>   <chr>       <chr>      <chr> <chr>   <chr>      <lg1>
##  1      1 ABAH     CHIAMAKA  <NA>      NURS~ " CHIO~ NA      NA
##  2      2 ABAH     OGOCHUKWU <NA>      PRIM~ " OO"  NA      NA
##  3      8 ADETUN~ DEMILADE  <NA>      PRIM~ " SALA~ NA      NA
##  4      9 ADETUN~ FUNMILADE <NA>      PRIM~ " OPAR~ NA      NA
##  5     10 ADEWOLE DAVID      <NA>      PRIM~ " OO"  NA      NA
##  6     11 ADEWOLE DORATHY  <NA>      NURS~ " BETT~ NA      NA
##  7     12 ADIEZE DAVID      <NA>      PRIM~ " OMOT~ <NA>    NA
##  8     13 ADIEZE FLOURISH  AMANDA    NURS~ " BETT~ NA      NA
##  9     15 AGIM    CHIMAMANDA <NA>      PRIM~ " EMMA~ NA      NA
## 10     16 AGIM    PHILIP    <NA>      PRIM~ " OO"  NA      NA
## # ... with 283 more rows, and 1 more variable: 'COMPUTER No' <chr>
```

#gets the count of people who dont share surname

```
dif_surname <- alphaorderclassonly %>%
  count(Surname)%>%
  filter(n==1)
dif_surname
```

```
## # A tibble: 124 x 2
##   Surname      n
##   <chr>    <int>
## 1 ADEBAYO      1
## 2 ADEBOWALE    1
## 3 ADELEKE      1
## 4 ADESANYA     1
## 5 ADESOYE      1
## 6 AGBASI       1
## 7 AGU          1
## 8 AJADI        1
## 9 AJAYI        1
## 10 AKAKASIAKA  1
## # ... with 114 more rows
```

#shows(filters) the students without siblings

```
no_siblings <- alphaorderclassonly %>%
  filter(Surname %in% dif_surname$Surname)
no_siblings
```

```
## # A tibble: 124 x 9
##      s_no Surname First_name Other_names class teacher SCHOLARSHIP 'ON REBATE?'
##      <dbl> <chr>   <chr>       <chr>      <chr> <chr>   <chr>      <lg1>
##  1      3 ADEBAYO AMIRAT     OLATOKUNBO PRIM~ " OO"  NA      NA
##  2      4 ADEBOW~ OLAOLUWA  <NA>      PRIM~ " ENIO~ NA      NA
##  3      5 ADELEKE ADEDAYO   ODINAKA    PRIM~ " BAMI~ NA      NA
##  4      6 ADESAN~ DAVID     <NA>      PRIM~ " AYO"  NA      NA
##  5      7 ADESOYE IBRAHEEM  <NA>      PRIM~ " SHOG~ <NA>    NA
##  6     14 AGBASI ZIKORAH   <NA>      PRIM~ " OPAR~ NA      NA
##  7     19 AGU     CHINEDU   <NA>      PRIM~ " OO"  NA      NA
##  8     20 AJADI   DEBORAH   <NA>      PRIM~ " OO"  NA      NA
##  9     21 AJAYI   ANJOLAOLU~ <NA>      NURS~ " OGUN~ NA      NA
## 10     26 AKAKAS~ DOMINION  <NA>      PRIM~ " OMOT~ <NA>    NA
## # ... with 114 more rows, and 1 more variable: 'COMPUTER No' <chr>
```

```

#Initial Grouping Process
alphaorderclassonly$groups <- NA
surname_counter <- 1

for(name in alphaorderclassonly$Surname){
  name_holder <- alphaorderclassonly$Surname
  if (name == name_holder){
    alphaorderclassonly$groups[[surname_counter]] <- "group a"
    next
    surname_counter <- surname_counter + 1
  }
  else{
    surname_counter <- surname_counter + 1
    name_holder <- alphaorderclassonly$Surname[surname_counter]
    if (name == name_holder){
      alphaorderclassonly$groups[[surname_counter]] <- "group b"
    }
    else{
      break
    }
  }
}
}

```

```

## Warning in if (name == name_holder) {: the condition has length > 1 and only the
## first element will be used

```

```

## Warning in if (name == name_holder) {: the condition has length > 1 and only the
## first element will be used

```

```

## Warning in if (name == name_holder) {: the condition has length > 1 and only the
## first element will be used

```

```
alphaorderclassonly
```

```

## # A tibble: 417 x 10
##   s_no Surname First_name Other_names class teacher SCHOLARSHIP 'ON REBATE?'
##   <dbl> <chr>   <chr>       <chr>      <chr> <chr>   <chr>      <lgl>
## 1     1  ABAH     CHIAMAKA    <NA>      NURS~ " CHIO~ NA      NA
## 2     2  ABAH     OGOCHUKWU  <NA>      PRIM~ " 00"  NA      NA
## 3     3  ADEBAYO AMIRAT     OLATOKUNBO PRIM~ " 00"  NA      NA
## 4     4  ADEBOW~ OLAOLUWA   <NA>      PRIM~ " ENIO~ NA      NA
## 5     5  ADELEKE ADEDAYO    ODINAKA    PRIM~ " BAMI~ NA      NA
## 6     6  ADESAN~ DAVID      <NA>      PRIM~ " AYO"  NA      NA
## 7     7  ADESOYE IBRAHEEM <NA>      PRIM~ " SHOG~ <NA> NA
## 8     8  ADETUN~ DEMILADE  <NA>      PRIM~ " SALA~ NA      NA
## 9     9  ADETUN~ FUNMILADE <NA>      PRIM~ " OPAR~ NA      NA
## 10    10 ADEWOLE DAVID    <NA>      PRIM~ " 00"  NA      NA
## # ... with 407 more rows, and 2 more variables: 'COMPUTER No' <chr>,
## #   groups <chr>

```