Load and clean ASA24 data

Contents

Introduction

In this tutorial, we will use mock data from the VVKAJ dataset that was created with ASA24. VVKAJ stands for Vegetarian, Vegan, Keto, American, Japanese and was designed because these different eating patterns reflect differences that are often seen in real data. This mock dataset contains dietary for 15 mock participants who report dietary data while following the 5 different dietary patterns (VVKAJ) for three days. There are a total of XX dietary records in this dataset.

ASA24 data includes the following files: _Items.csv, _INS.csv, _Responses.csv, _TNS.csv, Totals.csv, and TS.csv. Refer to the ASA24 Reserchers' website for specific explanations for each file, but for the purpose of this tutorial, we focus on using the Items.csv, which has all the food items reported by the participants.

In this script, you will:

1. Use Metadata 1 to filter out individuals;
2. Remove users that has only a small number of totals (days of record) - if you know which one to remove; and
3. Look for outliers in your totals by nutrient consumed on each day.

Load functions and packages

Name the path to DietR directory where input files are pulled.

```
main_wd <- "~/GitHub/DietR"
```

Load the necessary functions.

```
source("lib/specify_data_dir.R")
source("lib/load_clean_ASA24.R")
source("lib/Food_tree_scripts/format.foods_2.r")
source("lib/QCOutliers.R")
source("lib/average.by.R")
```

You can come back to the main directory by:

```
setwd(main_wd)
```

Load ASA24 Items data

Specify the directory where the data is.

```
SpecifyDataDirectory(directory.name= "eg_data/VVKAJ/")
```

Load your unprocessed (raw) food items-level data (as downloaded from the ASA24 study website). The csv file will be loaded as a dataframe in R and be named as items_raw.

```
items_raw <- read.csv("Raw_data/VVKAJ_Items.csv", sep = ",", header=T)
```

`items_raw` has a column called "Food_Description", but this needs to be changed to "Main.food.description". Change the column name.

```
names(items_raw)[names(items_raw) == "Food_Description"] <- "Main.food.description"
```

Check if any column names match with "Main.food.description". If there is a match, it will be printed.

```
names(items_raw)[names(items_raw) == "Main.food.description"]
```

```
## [1] "Main.food.description"
```

> [**NOTE**] The numbers in the square brackets of the output indicate the sequential number of each element to help count the number of elements.

Save the items file as a .txt file. This command saves the object “items_raw” as a .txt file with the specified filename using the write.table function.

```
write.table(items_raw, "VVKAJ_Items.txt", sep="\t", row.names=F)
```

Special characters common in food names in dietary data such as ", ', ,, % may interfere correct data loading in R; thus, we replace them with an underscore _.

Format foods so that special characters will be replaced with “_“.”_f” stands for “formatted”.

```
FormatFoods(input_fn =  "VVKAJ_Items.txt",
            output_fn = "VVKAJ_Items_f.txt")
```

> [**NOTE**] It is best practice to avoid overwriting your raw data. Always save formatted/manipulated versions as a new file as described above.

Load the Items_f.txt file to take a look at it.
You need the `quote=""` and `colClasses="character"` arguments to ignore quotation marks (do not regard them as a cell separator) and to load all the columns as characters so that FoodID will keep the trailing “.0”.

```
items_f <- read.delim("VVKAJ_Items_f.txt", quote="", colClasses="character")
```

All special characters in the items data should have been replaced with an underscore in the Main.food.description column, the 3rd from the last column of the items_f. We can confirm that by using the head function, which shows the first six rows of the specified dataset by default. In this website version of tutorial, output is folded for visual clarity. You can click and expand the results.

```
head(items_f)
```

Click to expand output

```
##                            RecallRecId UserName
## 1 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 2 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 3 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 4 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 5 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 6 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
##                                 UserID RecallNo RecallAttempt RecallStatus
## 1 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 2 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 3 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 4 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 5 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 6 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
##   IntakeStartDateTime IntakeEndDateTime ReportingDate Lang Occ_No
## 1      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 2      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 3      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 4      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 5      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      2
## 6      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      3
##          Occ_Time Occ_Name EatWith WatchTVuseComputer Location FoodNum FoodType
## 1  11/5/2021 7:00        1    <NA>               <NA>        1       1        1
## 2  11/5/2021 7:00        1    <NA>               <NA>        1       2        1
## 3  11/5/2021 7:00        1    <NA>               <NA>        1       3        2
## 4  11/5/2021 7:00        1    <NA>               <NA>        1       4        1
## 5 11/5/2021 10:00        7    <NA>               <NA>        1       5        1
## 6 11/5/2021 12:00        3    <NA>               <NA>        1       6        1
##   FoodSrce CodeNum FoodCode ModCode HowMany SubCode PortionCode FoodAmt    KCAL
## 1     <NA>       1 92400000       0     9.6       0       30001   297.6 124.992
## 2     <NA>       2 57231250       0       1       0       10205     104  419.12
## 3     <NA>       3 91302010       0     0.5       0       21000    10.5   31.92
## 4     <NA>       4 63200200       0     0.5       0       10205    98.5  50.235
## 5     <NA>       5 94000100       0     8.4       0       30000     252       0
## 6     <NA>       6 32103000       0     0.5       0       10205     111  285.27
##      PROT    TFAT     CARB      MOIS ALC   CAFF THEO     SUGR   FIBE  CALC
## 1       0   0.744 30.83136 265.93536   0 26.784    0 29.58144      0 2.976
## 2    9.36   10.92    75.92     5.408   0      0    0   16.016 10.088 40.56
## 3  0.0315       0    8.652    1.7955   0      0    0   8.6226  0.021  0.63
## 4  0.4137  0.6304 11.98745  85.29115   0      0    0  8.32325 2.6595  7.88
## 5       0       0        0   251.748   0      0    0        0      0  7.56
## 6 11.3886 25.6854   1.1211   70.9956   0      0    0   1.1211      0 46.62
##      IRON   MAGN   PHOS   POTA   SODI    ZINC     COPP   SELE     VC     VB1
## 1 0.05952      0 26.784  14.88  8.928 0.26784 0.020832 0.2976      0       0
## 2  17.992 101.92 307.84 350.48 298.48   2.392    0.416 43.472  0.208   0.728
## 3  0.0441   0.21   0.42   5.46   0.42  0.0231  0.00378  0.084 0.0525       0
## 4  0.1773  4.925 10.835  53.19  0.985 0.06895 0.032505 0.0985 2.4625 0.03152
## 5       0   2.52      0      0  10.08  0.0252   0.0252      0      0       0
## 6    1.11   8.88 157.62 116.55 420.69  0.9657  0.01554 27.861      0 0.06105
##        VB2     NIAC      VB6   FOLA     FA    FF   FDFE   VB12   VARA    RET
## 1        0        0        0      0      0     0      0      0      0      0
## 2    0.832    9.984     1.04 199.68 179.92 19.76 326.56  3.016  447.2  447.2
## 3  0.00399 0.012705  0.00252   0.21      0  0.21   0.21      0      0      0
## 4 0.036445   0.5122 0.058115  6.895      0 6.895  6.895      0   1.97      0
## 5        0        0        0      0      0     0      0      0      0      0
## 6  0.45954  0.05661  0.10878  39.96      0 39.96  39.96 1.0101 135.42 134.31
##    BCAR ACAR CRYP LYCO     LZ   ATOC     VK  CHOLE     SFAT S040 S060    S080
## 1     0    0    0    0      0      0      0      0        0    0    0       0
## 2   5.2    0 1.04    0 142.48 0.7696  2.288      0     1.04    0    0       0
## 3     0    0    0    0      0      0      0      0        0    0    0       0
## 4 27.58    0    0    0  66.98 0.4728 16.154      0 0.052205    0    0       0
## 5     0    0    0    0      0      0      0      0        0    0    0       0
## 6  11.1    0 9.99    0 316.35 1.6317 35.631 340.77  5.44344    0    0 0.00222
##      S100    S120    S140     S160     S180     MFAT    M161     M181    M201
## 1       0       0       0        0        0        0       0        0       0
## 2       0 0.00208 0.00208  0.80392  0.20592    6.032 0.00312  5.98936 0.03744
## 3       0       0       0        0        0        0       0        0       0
## 4       0       0       0 0.030535 0.010835 0.089635 0.00197 0.085695       0
## 5       0       0       0        0        0        0       0        0       0
## 6 0.00222 0.00222 0.04329  3.80397  1.40859  7.27938 0.29415    6.882 0.07881
##      M221     PFAT     P182    P183 P184     P204    P205 P225    P226  VITD
## 1       0        0        0       0    0        0       0    0       0     0
## 2 0.00104    3.224  3.05136 0.17264    0       0       0    0       0 1.976
## 3       0        0        0       0    0       0       0    0       0     0
## 4       0 0.274815 0.164495 0.11032    0       0       0    0       0     0
## 5       0        0        0       0    0       0       0    0       0     0
## 6 0.00777 10.96014  9.55599 1.21545    0 0.14319 0.00444    0 0.03441 1.998
##     CHOLN VITE_ADD B12_ADD F_TOTAL F_CITMLB F_OTHER F_JUICE V_TOTAL V_DRKGR
## 1  0.8928        0       0       0        0       0       0       0       0
## 2  29.952        0   3.016       0        0       0       0       0       0
## 3   0.231        0       0       0        0       0       0       0       0
## 4  5.0235        0       0 0.65995  0.65995       0       0       0       0
## 5       0        0       0       0        0       0       0       0       0
## 6 268.398        0       0       0        0       0       0       0       0
##   V_REDOR_TOTAL V_REDOR_TOMATO V_REDOR_OTHER V_STARCHY_TOTAL V_STARCHY_POTATO
## 1             0              0             0               0                0
## 2             0              0             0               0                0
## 3             0              0             0               0                0
## 4             0              0             0               0                0
## 5             0              0             0               0                0
## 6             0              0             0               0                0
##   V_STARCHY_OTHER V_OTHER V_LEGUMES G_TOTAL G_WHOLE G_REFINED PF_TOTAL
## 1               0       0         0       0       0         0        0
## 2               0       0         0  2.7456   2.652    0.0936   0.7384
## 3               0       0         0       0       0         0        0
## 4               0       0         0       0       0         0        0
## 5               0       0         0       0       0         0        0
## 6               0       0         0       0       0         0   1.8093
##   PF_MPS_TOTAL PF_MEAT PF_CUREDMEAT PF_ORGAN PF_POULT PF_SEAFD_HI PF_SEAFD_LOW
## 1            0       0            0        0        0           0            0
## 2            0       0            0        0        0           0            0
## 3            0       0            0        0        0           0            0
## 4            0       0            0        0        0           0            0
## 5            0       0            0        0        0           0            0
## 6            0       0            0        0        0           0            0
##   PF_EGGS PF_SOY PF_NUTSDS PF_LEGUMES D_TOTAL D_MILK D_YOGURT D_CHEESE    OILS
## 1       0      0         0          0       0      0        0        0       0
## 2       0      0    0.7384          0       0      0        0        0  6.0736
## 3       0      0         0          0       0      0        0        0       0
## 4       0      0         0          0       0      0        0        0       0
## 5       0      0         0          0       0      0        0        0       0
## 6  1.8093      0         0          0       0      0        0        0 15.8508
##   SOLID_FATS ADD_SUGARS A_DRINKS FoodComp
## 1          0    7.05312        0        1
## 2          0      3.432        0        1
## 3          0    2.05275        0        1
## 4          0          0        0        1
## 5          0          0        0        1
## 6     4.8507          0        0        1
##                                       Main.food.description
## 1                                            Soft_drink_NFS
## 2 Cereal_Post_Great_Grains_Double_Pecan_Whole_Grain_Cereal_
## 3                                                     Honey
## 4                                        Berries_frozen_NFS
## 5                                                 Water_tap
## 6                            Egg_salad_made_with_mayonnaise
##                                    Old.Main.food.description     FoodID
## 1                                            Soft drink, NFS 92400000.0
## 2 Cereal (Post Great Grains Double Pecan Whole Grain Cereal) 57231250.0
## 3                                                      Honey 91302010.0
## 4                                       Berries, frozen, NFS 63200200.0
## 5                                                 Water, tap 94000100.0
## 6                            Egg salad, made with mayonnaise 32103000.0
```

Add a human-readable sample identifier (SampleID) with a desired prefix, and save it as a .txt file. SampleIDs are IDs unique to each combination of users and day and represent days of dietary intake in this dataset.

```
AddSampleIDtoItems(input.fn="VVKAJ_Items_f.txt", user.name="UserName", recall.no="RecallNo",
                   prefix="vvkaj.", out.fn="VVKAJ_Items_f_id.txt")
```

Load the formatted Items file with SampleID added.

```
items_f_id <- read.delim("VVKAJ_Items_f_id.txt", quote="", colClasses="character")
```

A combination of the specified prefix and sequential number (vvkaj.00001) should be added in the SampleID column, the first column of the items_f_id dataframe. You will probably need to scroll up the output a little bit in the console to view the first column.

```
head(items_f_id)
```

Click to expand output

```
##      SampleID                          RecallRecId UserName
## 1 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 2 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 3 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 4 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 5 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
## 6 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9 VVKAJ101
##                                 UserID RecallNo RecallAttempt RecallStatus
## 1 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 2 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 3 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 4 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 5 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 6 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
##   IntakeStartDateTime IntakeEndDateTime ReportingDate Lang Occ_No
## 1      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 2      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 3      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 4      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 5      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      2
## 6      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      3
##          Occ_Time Occ_Name EatWith WatchTVuseComputer Location FoodNum FoodType
## 1  11/5/2021 7:00        1    <NA>               <NA>        1       1        1
## 2  11/5/2021 7:00        1    <NA>               <NA>        1       2        1
## 3  11/5/2021 7:00        1    <NA>               <NA>        1       3        2
## 4  11/5/2021 7:00        1    <NA>               <NA>        1       4        1
## 5 11/5/2021 10:00        7    <NA>               <NA>        1       5        1
## 6 11/5/2021 12:00        3    <NA>               <NA>        1       6        1
##   FoodSrce CodeNum FoodCode ModCode HowMany SubCode PortionCode FoodAmt    KCAL
## 1     <NA>       1 92400000       0     9.6       0       30001   297.6 124.992
## 2     <NA>       2 57231250       0       1       0       10205     104  419.12
## 3     <NA>       3 91302010       0     0.5       0       21000    10.5   31.92
## 4     <NA>       4 63200200       0     0.5       0       10205    98.5  50.235
## 5     <NA>       5 94000100       0     8.4       0       30000     252       0
## 6     <NA>       6 32103000       0     0.5       0       10205     111  285.27
##       PROT    TFAT     CARB      MOIS ALC   CAFF THEO     SUGR   FIBE  CALC
## 1       0   0.744 30.83136 265.93536   0 26.784    0 29.58144      0 2.976
## 2    9.36   10.92    75.92     5.408   0      0    0   16.016 10.088 40.56
## 3  0.0315       0    8.652    1.7955   0      0    0   8.6226  0.021  0.63
## 4  0.4137  0.6304 11.98745  85.29115   0      0    0  8.32325 2.6595  7.88
## 5       0       0        0   251.748   0      0    0        0      0  7.56
## 6 11.3886 25.6854   1.1211   70.9956   0      0    0   1.1211      0 46.62
##      IRON   MAGN   PHOS   POTA   SODI    ZINC     COPP   SELE     VC     VB1
## 1 0.05952      0 26.784  14.88  8.928 0.26784 0.020832 0.2976      0       0
## 2  17.992 101.92 307.84 350.48 298.48   2.392    0.416 43.472  0.208   0.728
## 3  0.0441   0.21   0.42   5.46   0.42  0.0231  0.00378  0.084 0.0525       0
## 4  0.1773  4.925 10.835  53.19  0.985 0.06895 0.032505 0.0985 2.4625 0.03152
## 5       0   2.52      0      0  10.08  0.0252   0.0252      0      0       0
## 6    1.11   8.88 157.62 116.55 420.69  0.9657  0.01554 27.861      0 0.06105
##        VB2     NIAC      VB6   FOLA     FA    FF   FDFE   VB12   VARA    RET
## 1        0        0        0      0      0     0      0      0      0      0
## 2    0.832    9.984     1.04 199.68 179.92 19.76 326.56  3.016  447.2  447.2
## 3  0.00399 0.012705  0.00252   0.21      0  0.21   0.21      0      0      0
## 4 0.036445   0.5122 0.058115  6.895      0 6.895  6.895      0   1.97      0
## 5        0        0        0      0      0     0      0      0      0      0
## 6  0.45954  0.05661  0.10878  39.96      0 39.96  39.96 1.0101 135.42 134.31
##    BCAR ACAR CRYP LYCO     LZ   ATOC     VK  CHOLE     SFAT S040 S060    S080
## 1     0    0    0    0      0      0      0      0        0    0    0       0
## 2   5.2    0 1.04    0 142.48 0.7696  2.288      0     1.04    0    0       0
## 3     0    0    0    0      0      0      0      0        0    0    0       0
## 4 27.58    0    0    0  66.98 0.4728 16.154      0 0.052205    0    0       0
## 5     0    0    0    0      0      0      0      0        0    0    0       0
## 6  11.1    0 9.99    0 316.35 1.6317 35.631 340.77  5.44344    0    0 0.00222
##      S100    S120    S140     S160     S180     MFAT    M161     M181    M201
## 1       0       0       0        0        0        0       0        0       0
## 2       0 0.00208 0.00208  0.80392  0.20592    6.032 0.00312  5.98936 0.03744
## 3       0       0       0        0        0        0       0        0       0
## 4       0       0       0 0.030535 0.010835 0.089635 0.00197 0.085695       0
## 5       0       0       0        0        0        0       0        0       0
## 6 0.00222 0.00222 0.04329  3.80397  1.40859  7.27938 0.29415    6.882 0.07881
##      M221     PFAT     P182    P183 P184    P204    P205 P225    P226  VITD
## 1       0        0        0       0    0       0       0    0       0     0
## 2 0.00104    3.224  3.05136 0.17264    0       0       0    0       0 1.976
## 3       0        0        0       0    0       0       0    0       0     0
## 4       0 0.274815 0.164495 0.11032    0       0       0    0       0     0
## 5       0        0        0       0    0       0       0    0       0     0
## 6 0.00777 10.96014  9.55599 1.21545    0 0.14319 0.00444    0 0.03441 1.998
##    CHOLN VITE_ADD B12_ADD F_TOTAL F_CITMLB F_OTHER F_JUICE V_TOTAL V_DRKGR
## 1 0.8928        0       0       0        0       0       0       0       0
## 2 29.952        0   3.016       0        0       0       0       0       0
## 3   0.231        0       0       0        0       0       0       0       0
## 4  5.0235        0       0 0.65995  0.65995       0       0       0       0
## 5       0        0       0       0        0       0       0       0       0
## 6 268.398        0       0       0        0       0       0       0       0
##   V_REDOR_TOTAL V_REDOR_TOMATO V_REDOR_OTHER V_STARCHY_TOTAL V_STARCHY_POTATO
## 1             0              0             0               0                0
## 2             0              0             0               0                0
## 3             0              0             0               0                0
## 4             0              0             0               0                0
## 5             0              0             0               0                0
## 6             0              0             0               0                0
##   V_STARCHY_OTHER V_OTHER V_LEGUMES G_TOTAL G_WHOLE G_REFINED PF_TOTAL
## 1               0       0         0       0       0         0        0
## 2               0       0         0  2.7456   2.652    0.0936   0.7384
## 3               0       0         0       0       0         0        0
## 4               0       0         0       0       0         0        0
## 5               0       0         0       0       0         0        0
## 6               0       0         0       0       0         0   1.8093
##   PF_MPS_TOTAL PF_MEAT PF_CUREDMEAT PF_ORGAN PF_POULT PF_SEAFD_HI PF_SEAFD_LOW
## 1            0       0            0        0        0           0            0
## 2            0       0            0        0        0           0            0
## 3            0       0            0        0        0           0            0
## 4            0       0            0        0        0           0            0
## 5            0       0            0        0        0           0            0
## 6            0       0            0        0        0           0            0
##   PF_EGGS PF_SOY PF_NUTSDS PF_LEGUMES D_TOTAL D_MILK D_YOGURT D_CHEESE    OILS
## 1       0      0         0          0       0      0        0        0       0
## 2       0      0    0.7384          0       0      0        0        0  6.0736
## 3       0      0         0          0       0      0        0        0       0
## 4       0      0         0          0       0      0        0        0       0
## 5       0      0         0          0       0      0        0        0       0
## 6  1.8093      0         0          0       0      0        0        0 15.8508
##   SOLID_FATS ADD_SUGARS A_DRINKS FoodComp
## 1          0    7.05312        0        1
## 2          0      3.432        0        1
## 3          0    2.05275        0        1
## 4          0          0        0        1
## 5          0          0        0        1
## 6     4.8507          0        0        1
##                                       Main.food.description
## 1                                            Soft_drink_NFS
## 2 Cereal_Post_Great_Grains_Double_Pecan_Whole_Grain_Cereal_
## 3                                                     Honey
## 4                                        Berries_frozen_NFS
## 5                                                 Water_tap
## 6                            Egg_salad_made_with_mayonnaise
##                                     Old.Main.food.description     FoodID
## 1                                             Soft drink, NFS 92400000.0
## 2 Cereal (Post Great Grains Double Pecan Whole Grain Cereal) 57231250.0
## 3                                                       Honey 91302010.0
## 4                                        Berries, frozen, NFS 63200200.0
## 5                                                  Water, tap 94000100.0
## 6                             Egg salad, made with mayonnaise 32103000.0
```

Ensure your items file has the expected dimensions (number of rows x number of columns, shown as number of obs. and number of variables) in the environment window of R Studio. Or by using dim(items_f_id) and dim(items_raw). Note that items_f_id has 3 more columns than items_raw because new columns of FoodID, Old.Main.food.description, and SampleID have been added.

```
  dim(items_f_id)
```

```
## [1] 779 133
```

```
  dim(items_raw)
```

```
## [1] 779 130
```

Use `individuals_to_remove.txt` to filter out users marked as Remove = yes

Load your metadata that has information about which UserName(s) to remove.

```
ind_to_rm <- read.delim("individuals_to_remove.txt")
```

Metadata for this purpose (ind_to_rm) has UserName and which one to be removed.

```
ind_to_rm
```

```
##    UserName Remove
## 1  VVKAJ101
## 2  VVKAJ102
## 3  VVKAJ103
## 4  VVKAJ104
## 5  VVKAJ105
## 6  VVKAJ106
## 7  VVKAJ107
## 8  VVKAJ108
## 9  VVKAJ109
## 10 VVKAJ110
## 11 VVKAJ111
## 12 VVKAJ112
## 13 VVKAJ113
## 14 VVKAJ114
## 15 VVKAJ115
## 16 VVKAJ116    yes
## 17 VVKAJ117
```

Show which has "yes" in the "Remove" column.

```
subset(ind_to_rm, Remove == "yes")
```

```
##    UserName Remove
## 16 VVKAJ116    yes
```

As shown in the console, the user named “VVKAJ116” is marked to be removed. VVKAJ116 has only 1 day of data, which may not be complete, thus it is marked as an individual to remove. However, be careful when deleting a datapoint from your study and never remove individuals from the raw dataset, to ensure you can always go back and include them if desired.

Remove the specified individuals.
The output will be saved as a text file with the specified name. This assumes the usernames are in UserName column, and will print which user(s) will be removed.

```
RemoveRows(data=items_f_id, metadata.file= ind_to_rm,
           output.name= "VVKAJ_Items_f_id_s.txt")
```

```
## 1 row(s) below are to be removed:
##    UserName Remove
## 16 VVKAJ116    yes
```

Load the output for further processing.

```
items_f_id_s <- read.delim("VVKAJ_Items_f_id_s.txt", quote="", colClasses="character")
```

Show unique usernames in items_f_id_s and confirm “VVKAJ116” has been removed.

```
unique(items_f_id_s$UserName)
```

```
##  [1] "VVKAJ101" "VVKAJ102" "VVKAJ103" "VVKAJ104" "VVKAJ105" "VVKAJ106"
##  [7] "VVKAJ107" "VVKAJ108" "VVKAJ109" "VVKAJ110" "VVKAJ111" "VVKAJ112"
## [13] "VVKAJ113" "VVKAJ114" "VVKAJ115" "VVKAJ117"
```

Merge individuals’ metadata to items

`ind_metadata` has the participants’ gender, age, height, weight, BMI, and Waist.Circumference, etc. If desired, this individual-specific information can be added to items data.

Load ind_metadata.txt.

```
ind_metadata <- read.table("ind_metadata.txt", sep="\t", header=T)
```

Look at what the metadata has.

```
head(ind_metadata)
```

```
##   UserName       Diet Gender Age Weight Height      BMI Waist.Circumference
## 1 VVKAJ101 Vegetarian      M  31     79    186 22.83501                  80
## 2 VVKAJ102      Vegan      F  60     73    163 27.47563                  90
## 3 VVKAJ103       Keto      M  43     81    175 26.44898                  72
## 4 VVKAJ104   American      F  25     85    169 29.76086                  89
## 5 VVKAJ105   Japanese      M  71     60    169 21.00767                  75
## 6 VVKAJ106 Vegetarian      F  53     68    159 26.89767                  85
```

This includes information on the removed individual, VVKAJ116, but it will not be used if VVKAJ116 is not in the items data.

Add this metadata of each participant to totals or items. 'NA' will be inserted to UserNames which are not in ind_metadata.

```
items_f_id_s_m <- merge(x=items_f_id_s, y=ind_metadata, by="UserName", all.x=T)
```

Check that the items data and metadata are merged.

```
head(items_f_id_s_m)
```

Click to expand output

```
##   UserName    SampleID                          RecallRecId
## 1 VVKAJ101 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9
## 2 VVKAJ101 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9
## 3 VVKAJ101 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9
## 4 VVKAJ101 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9
## 5 VVKAJ101 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9
## 6 VVKAJ101 vvkaj.00001 41fee4cf-783f-469b-aadf-62c7e2cd33a9
##                                 UserID RecallNo RecallAttempt RecallStatus
## 1 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 2 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 3 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 4 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 5 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
## 6 7bd05142-312e-4648-b1ac-fc258540af52        1             0            2
##   IntakeStartDateTime IntakeEndDateTime ReportingDate Lang Occ_No
## 1      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 2      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 3      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 4      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      1
## 5      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      2
## 6      11/5/2021 0:00   11/5/2021 23:59     11/6/2021    1      3
##          Occ_Time Occ_Name EatWith WatchTVuseComputer Location FoodNum FoodType
## 1  11/5/2021 7:00        1    <NA>               <NA>        1       1        1
## 2  11/5/2021 7:00        1    <NA>               <NA>        1       2        1
## 3  11/5/2021 7:00        1    <NA>               <NA>        1       3        2
## 4  11/5/2021 7:00        1    <NA>               <NA>        1       4        1
## 5 11/5/2021 10:00        7    <NA>               <NA>        1       5        1
## 6 11/5/2021 12:00        3    <NA>               <NA>        1       6        1
##   FoodSrce CodeNum FoodCode ModCode HowMany SubCode PortionCode FoodAmt    KCAL
## 1     <NA>       1 92400000       0     9.6       0       30001   297.6 124.992
## 2     <NA>       2 57231250       0       1       0       10205     104  419.12
## 3     <NA>       3 91302010       0     0.5       0       21000    10.5   31.92
## 4     <NA>       4 63200200       0     0.5       0       10205    98.5  50.235
## 5     <NA>       5 94000100       0     8.4       0       30000     252       0
## 6     <NA>       6 32103000       0     0.5       0       10205     111  285.27
##      PROT    TFAT     CARB      MOIS ALC   CAFF THEO     SUGR   FIBE  CALC
## 1       0   0.744 30.83136 265.93536   0 26.784    0 29.58144      0 2.976
## 2    9.36   10.92    75.92     5.408   0      0    0   16.016 10.088 40.56
## 3  0.0315       0    8.652    1.7955   0      0    0   8.6226  0.021  0.63
## 4  0.4137  0.6304 11.98745  85.29115   0      0    0  8.32325 2.6595  7.88
## 5       0       0        0   251.748   0      0    0        0      0  7.56
## 6 11.3886 25.6854   1.1211   70.9956   0      0    0   1.1211      0 46.62
##      IRON   MAGN   PHOS   POTA   SODI    ZINC     COPP   SELE     VC     VB1
## 1 0.05952      0 26.784  14.88  8.928 0.26784 0.020832 0.2976      0       0
## 2  17.992 101.92 307.84 350.48 298.48   2.392    0.416 43.472  0.208   0.728
## 3  0.0441   0.21   0.42   5.46   0.42  0.0231  0.00378  0.084 0.0525       0
## 4  0.1773  4.925 10.835  53.19  0.985 0.06895 0.032505 0.0985 2.4625 0.03152
## 5       0   2.52      0      0  10.08  0.0252   0.0252      0      0       0
## 6    1.11   8.88 157.62 116.55 420.69  0.9657  0.01554 27.861      0 0.06105
##        VB2     NIAC      VB6   FOLA     FA    FF   FDFE   VB12   VARA    RET
## 1        0        0        0      0      0     0      0      0      0      0
## 2    0.832    9.984     1.04 199.68 179.92 19.76 326.56  3.016  447.2  447.2
## 3  0.00399 0.012705  0.00252   0.21      0  0.21   0.21      0      0      0
## 4 0.036445   0.5122 0.058115  6.895      0 6.895  6.895      0   1.97      0
## 5        0        0        0      0      0     0      0      0      0      0
## 6  0.45954  0.05661  0.10878  39.96      0 39.96  39.96 1.0101 135.42 134.31
##    BCAR ACAR CRYP LYCO     LZ   ATOC     VK  CHOLE     SFAT S040 S060    S080
## 1     0    0    0    0      0      0      0      0        0    0    0       0
## 2   5.2    0 1.04    0 142.48 0.7696  2.288      0     1.04    0    0       0
## 3     0    0    0    0      0      0      0      0        0    0    0       0
## 4 27.58    0    0    0  66.98 0.4728 16.154      0 0.052205    0    0       0
## 5     0    0    0    0      0      0      0      0        0    0    0       0
## 6  11.1    0 9.99    0 316.35 1.6317 35.631 340.77  5.44344    0    0 0.00222
##      S100    S120    S140     S160     S180     MFAT    M161     M181    M201
## 1       0       0       0        0        0        0       0        0       0
## 2       0 0.00208 0.00208  0.80392  0.20592    6.032 0.00312  5.98936 0.03744
## 3       0       0       0        0        0        0       0        0       0
## 4       0       0       0 0.030535 0.010835 0.089635 0.00197 0.085695       0
## 5       0       0       0        0        0        0       0        0       0
## 6 0.00222 0.00222 0.04329  3.80397  1.40859  7.27938 0.29415    6.882 0.07881
##      M221     PFAT     P182    P183 P184    P204    P205 P225    P226  VITD
## 1       0        0        0       0    0       0       0    0       0     0
## 2 0.00104    3.224  3.05136 0.17264    0       0       0    0       0 1.976
## 3       0        0        0       0    0       0       0    0       0     0
## 4       0 0.274815 0.164495 0.11032    0       0       0    0       0     0
## 5       0        0        0       0    0       0       0    0       0     0
## 6 0.00777 10.96014  9.55599 1.21545    0 0.14319 0.00444    0 0.03441 1.998
##     CHOLN VITE_ADD B12_ADD F_TOTAL F_CITMLB F_OTHER F_JUICE V_TOTAL V_DRKGR
## 1  0.8928        0       0       0        0       0       0       0       0
## 2  29.952        0   3.016       0        0       0       0       0       0
## 3   0.231        0       0       0        0       0       0       0       0
## 4  5.0235        0       0 0.65995  0.65995       0       0       0       0
## 5       0        0       0       0        0       0       0       0       0
## 6 268.398        0       0       0        0       0       0       0       0
##   V_REDOR_TOTAL V_REDOR_TOMATO V_REDOR_OTHER V_STARCHY_TOTAL V_STARCHY_POTATO
## 1             0              0             0               0                0
## 2             0              0             0               0                0
## 3             0              0             0               0                0
## 4             0              0             0               0                0
## 5             0              0             0               0                0
## 6             0              0             0               0                0
##   V_STARCHY_OTHER V_OTHER V_LEGUMES G_TOTAL G_WHOLE G_REFINED PF_TOTAL
## 1               0       0         0       0       0         0        0
## 2               0       0         0  2.7456   2.652    0.0936   0.7384
## 3               0       0         0       0       0         0        0
## 4               0       0         0       0       0         0        0
## 5               0       0         0       0       0         0        0
## 6               0       0         0       0       0         0   1.8093
##   PF_MPS_TOTAL PF_MEAT PF_CUREDMEAT PF_ORGAN PF_POULT PF_SEAFD_HI PF_SEAFD_LOW
## 1            0       0            0        0        0           0            0
## 2            0       0            0        0        0           0            0
## 3            0       0            0        0        0           0            0
## 4            0       0            0        0        0           0            0
## 5            0       0            0        0        0           0            0
## 6            0       0            0        0        0           0            0
##   PF_EGGS PF_SOY PF_NUTSDS PF_LEGUMES D_TOTAL D_MILK D_YOGURT D_CHEESE    OILS
## 1       0      0         0          0       0      0        0        0       0
## 2       0      0    0.7384          0       0      0        0        0  6.0736
## 3       0      0         0          0       0      0        0        0       0
## 4       0      0         0          0       0      0        0        0       0
## 5       0      0         0          0       0      0        0        0       0
## 6  1.8093      0         0          0       0      0        0        0 15.8508
##   SOLID_FATS ADD_SUGARS A_DRINKS FoodComp
## 1          0    7.05312        0        1
## 2          0      3.432        0        1
## 3          0    2.05275        0        1
## 4          0          0        0        1
## 5          0          0        0        1
## 6     4.8507          0        0        1
##                                       Main.food.description
## 1                                            Soft_drink_NFS
## 2 Cereal_Post_Great_Grains_Double_Pecan_Whole_Grain_Cereal_
## 3                                                     Honey
## 4                                        Berries_frozen_NFS
## 5                                                 Water_tap
## 6                            Egg_salad_made_with_mayonnaise
##                                    Old.Main.food.description     FoodID
## 1                                            Soft drink, NFS 92400000.0
## 2 Cereal (Post Great Grains Double Pecan Whole Grain Cereal) 57231250.0
## 3                                                      Honey 91302010.0
## 4                                       Berries, frozen, NFS 63200200.0
## 5                                                 Water, tap 94000100.0
## 6                            Egg salad, made with mayonnaise 32103000.0
##         Diet Gender Age Weight Height      BMI Waist.Circumference
## 1 Vegetarian      M  31     79    186 22.83501                  80
## 2 Vegetarian      M  31     79    186 22.83501                  80
## 3 Vegetarian      M  31     79    186 22.83501                  80
## 4 Vegetarian      M  31     79    186 22.83501                  80
## 5 Vegetarian      M  31     79    186 22.83501                  80
## 6 Vegetarian      M  31     79    186 22.83501                  80
```

Furthermore, as a quick way to look at the metadata of only the selected individuals, you can subset the metadata to just the usernames present in the analysis dataset (items_f_id_s) using the `%in%` operator.

```
ind_metadata_s <- ind_metadata[ind_metadata$UserName %in% items_f_id_s$UserName, ]
```

Use the tail function to show the last six rows of ind_metadata_s. You can see that the last individual in this metadata is now VVKAJ117, and that VVKAJ116, which was not in items_f_id_s, has been omitted.

```
tail(ind_metadata_s)
```

Click to expand output

```
##    UserName       Diet Gender Age  Weight Height      BMI Waist.Circumference
## 11 VVKAJ111 Vegetarian      M  60 74.9000  164.1 27.81408                78.0
## 12 VVKAJ112      Vegan      F  48 69.8992  158.0 28.00000                95.5
## 13 VVKAJ113       Keto      M  43 80.0000  183.0 23.88844                98.0
## 14 VVKAJ114   American      F  62 90.7235  161.0 35.00000               110.0
## 15 VVKAJ115   Japanese      M  22 69.6348  174.0 23.00000                70.0
## 17 VVKAJ117      error      M  23 76.0000  175.0 24.81633                80.0
```

Save the merged dataframe as a .txt file.

```
write.table(items_f_id_s_m, "VVKAJ_Items_f_id_s_m.txt", sep="\t", row.names=F, quote=F)
```

Generate new totals file from the items file

Use one of the input files saved above as an input for calculating totals for. Specify which columns have usernames and Recall.No., which has the recorded days.

```
GenerateTotals(inputfn =   "VVKAJ_Items_f_id_s_m.txt",
               User.Name = "UserName",
               Recall.No = "RecallNo",
               outfn =     "VVKAJ_Tot.txt")
```

Load the total file generated above.

```
new_totals <- read.table("VVKAJ_Tot.txt", header=T, sep="\t")
```

The number of rows should be {No. of users x No. days}. For the example data, 16 users x 3 days = 48 rows (observations).

```
nrow(new_totals)
```

```
## [1] 48
```

View the new_totals.

```
head(new_totals)
```

Click to expand output

```
##     User_Day UserName RecallNo  FoodAmt      KCAL     PROT     TFAT     CARB
## 1 VVKAJ101_1 VVKAJ101        1 2402.500 2314.9040 80.15720 82.97721 337.7973
## 2 VVKAJ101_2 VVKAJ101        2 1671.650  913.5955 35.13461 37.42791 118.5702
## 3 VVKAJ101_3 VVKAJ101        3 1920.588 1604.8050 48.29881 70.08057 215.7318
## 4 VVKAJ102_1 VVKAJ102        1 1957.475 1440.1995 38.54926 80.38863 152.8296
## 5 VVKAJ102_2 VVKAJ102        2 1858.067 1508.9957 46.25199 56.89753 212.5225
## 6 VVKAJ102_3 VVKAJ102        3 2049.400 1834.2306 58.89500 87.53381 211.8413
##       MOIS ALC   CAFF THEO      SUGR     FIBE      CALC      IRON     MAGN
## 1 1879.619   0 26.784 0.00 144.20847 52.62550  889.8455 31.282000 589.3865
## 2 1467.035   0 76.800 0.00  47.67748 21.04413  619.2521  7.366001 273.2332
## 3 1572.133   0 30.240 3.36 111.81870 23.27856 1427.6826 14.628896 320.9569
## 4 1669.286   0 50.400 5.04  25.75794 33.90378  912.4538 11.797247 284.4450
## 5 1521.262   0 22.680 2.52  55.29288 31.25067 1144.7673 14.180575 366.8420
## 6 1666.272   0  0.000 0.00  31.41001 29.35169  700.0389 14.009471 373.1694
##        PHOS     POTA     SODI      ZINC     COPP      SELE       VC      VB1
## 1 1787.4540 4615.400 2298.662 11.776515 1.935984 113.91525 209.0321 1.882480
## 2  794.2740 2093.175 1058.073  5.470231 1.325496  58.42284 147.7427 0.983826
## 3 1249.7137 2571.690 2069.723  8.953749 1.565804  83.81562 137.9377 1.316055
## 4  808.2387 2974.354 2550.088  5.641533 1.278044  42.76070 179.4785 1.137953
## 5 1043.4290 4686.336 2840.423  6.408068 1.427889  38.26618 157.1404 2.909776
## 6 1164.2706 4653.953 4285.981  6.745011 2.066325  72.05487 117.7118 4.570952
##         VB2     NIAC      VB6     FOLA      FA       FF      FDFE     VB12
## 1 2.9713065 18.72793 3.829823 719.3660 179.920 539.4460  846.2460  5.17202
## 2 1.2894892 12.24912 1.386882 372.7493  85.070 287.6793  432.4393  2.23350
## 3 2.0125987 14.49594 4.120610 918.5919 601.800 316.7919 1340.1919 13.23145
## 4 0.9494982 10.92889 1.637870 459.3215  52.025 410.3265  495.5415  1.80990
## 5 1.3127483 13.76033 2.560617 534.1780  20.200 515.9980  548.3180  3.97461
## 6 1.2672856 18.25922 2.540221 506.4956  26.700 481.8156  525.2356  3.70270
##        VARA      RET      BCAR      ACAR      CRYP     LYCO        LZ      ATOC
## 1 1261.6665 673.1850  7004.547  106.7165  56.11800   13.350 14273.582  9.851310
## 2 1412.8761  82.3500 13464.676 4921.4081  81.65188 1174.626  2444.209  4.628156
## 3 1111.7093 504.9250  6476.761 1614.4750  19.37550    4.704  5437.084 16.765260
## 4  310.6182 165.1115  1551.208  344.1073  68.67350 7582.125  1340.894 18.987955
## 5 1265.3238 222.6810 10784.821 3226.1183 198.74500 3380.685  4032.658 16.267538
## 6  344.2824 144.1200  2307.391  143.9169  12.66863 3374.064  3388.637  9.870323
##         VK   CHOLE      SFAT     S040       S060      S080      S100       S120
## 1 674.6513 375.967 18.696057 0.310902 0.11065100 0.1176030 0.2303450 0.18606100
## 2 176.1320  18.300  8.035444 0.137250 0.13725000 0.1401794 0.1440906 0.16560063
## 3 356.2514  64.265 21.762310 0.461520 0.35581000 0.6559880 0.7907570 3.08172600
## 4 193.4643  63.277 12.393455 0.063991 0.04301325 0.0834395 0.1062550 0.06083225
## 5 271.3043  53.980 10.073571 0.086720 0.06900500 0.2074300 0.2042550 0.43357300
## 6 270.4181  65.830 17.110502 0.318160 0.16680000 0.2767274 0.3239814 0.26321963
##        S140      S160     S180     MFAT      M161     M181      M201       M221
## 1 1.2467585 11.448329 4.130060 28.14091 0.7393160 26.99568 0.2349650 0.00881000
## 2 0.5821610  4.955292 1.532696 15.32206 0.3639200 14.79923 0.1110759 0.00346125
## 3 2.6981170  9.237105 3.415553 24.67066 0.4838810 23.65191 0.1867220 0.00864600
## 4 0.1955330  8.961791 2.439567 34.62131 0.9667318 33.33719 0.2755755 0.00450000
## 5 0.5196547  6.309740 1.897680 22.20331 0.4462590 21.47338 0.2112680 0.02609700
## 6 0.8637470 10.355075 3.764050 33.49530 0.3996655 32.60066 0.3411139 0.00658725
##       PFAT      P182     P183 P184     P204    P205    P225    P226      VITD
## 1 28.95861 25.201469 3.525870    0 0.157937 0.00444 0.00000 0.03441  4.881200
## 2 10.24234  9.135821 1.092069    0 0.001315 0.00000 0.00000 0.00000  2.379000
## 3 18.75467 16.930229 1.803171    0 0.006573 0.00000 0.00000 0.00000 10.193000
## 4 28.41416 24.796558 3.534361    0 0.038890 0.00000 0.00000 0.00909  3.375425
## 5 20.44748 17.914195 2.412754    0 0.026967 0.00000 0.00000 0.01881  4.985000
## 6 31.17042 27.616109 3.344330    0 0.038623 0.00032 0.00208 0.00670  2.102000
##      CHOLN VITE_ADD  B12_ADD  F_TOTAL F_CITMLB  F_OTHER F_JUICE  V_TOTAL
## 1 572.5637  0.00000  3.01600 4.037445  0.65995 3.035495  0.3420 3.656365
## 2 143.2537  0.00000  1.41000 0.640500  0.30625 0.323750  0.0105 2.727494
## 3 195.7137  8.34700 10.79370 1.530100  1.01160 0.505000  0.0135 1.906206
## 4 182.4953  6.85640  1.53720 0.000000  0.00000 0.000000  0.0000 4.280535
## 5 212.2748  7.19922  3.21321 0.000000  0.00000 0.000000  0.0000 7.783067
## 6 219.7563  0.00000  2.83410 0.000000  0.00000 0.000000  0.0000 7.084731
##     V_DRKGR V_REDOR_TOTAL V_REDOR_TOMATO V_REDOR_OTHER V_STARCHY_TOTAL
## 1 1.6658400      0.000000        0.00000      0.000000       0.0000000
## 2 0.4840000      1.177100        0.04860      1.128500       0.0000000
## 3 1.0143750      0.300500        0.00000      0.300500       0.0000000
## 4 0.0099900      1.485350        1.48535      0.000000       0.1288625
## 5 0.6186667      1.857575        0.84840      1.009175       3.3200000
## 6 0.7412500      0.848400        0.84840      0.000000       4.6146000
##   V_STARCHY_POTATO V_STARCHY_OTHER   V_OTHER V_LEGUMES G_TOTAL G_WHOLE
## 1           0.0000       0.0000000 1.9905250  1.256750  5.4384  2.6520
## 2           0.0000       0.0000000 1.0663938  0.105000  2.6500  1.7669
## 3           0.0000       0.0000000 0.5913312  0.135000  4.4072  4.2542
## 4           0.0000       0.1288625 2.6563325  1.085300  5.0421  1.5522
## 5           3.3200       0.0000000 1.9868250  0.285000  2.2405  1.0285
## 6           4.6146       0.0000000 0.8804812  0.192375  1.6050  0.0000
##   G_REFINED PF_TOTAL PF_MPS_TOTAL PF_MEAT PF_CUREDMEAT PF_ORGAN PF_POULT
## 1    2.7864 3.111700       0.0000       0       0.0000        0        0
## 2    0.8831 3.316731       0.0000       0       0.0000        0        0
## 3    0.1530 2.501000       0.0000       0       0.0000        0        0
## 4    3.4899 0.644600       0.0000       0       0.0000        0        0
## 5    1.2120 2.226680       0.0000       0       0.0000        0        0
## 6    1.6050 4.697931       0.2016       0       0.2016        0        0
##   PF_SEAFD_HI PF_SEAFD_LOW PF_EGGS PF_SOY PF_NUTSDS PF_LEGUMES D_TOTAL  D_MILK
## 1           0            0  1.8093 0.0000 1.3024000   4.935150 1.34064 0.21168
## 2           0            0  0.0000 1.9200 1.3967312   0.420000 0.75030 0.75030
## 3           0            0  0.0000 0.0000 2.5010000   0.540000 2.39525 0.88850
## 4           0            0  0.3030 0.0000 0.3416000   4.351050 0.15150 0.15150
## 5           0            0  0.2020 1.6660 0.3586800   1.155000 0.57980 0.57980
## 6           0            0  0.2020 3.3558 0.9385313   0.772875 0.70760 0.58900
##   D_YOGURT D_CHEESE     OILS SOLID_FATS ADD_SUGARS A_DRINKS
## 1  0.00000   0.5544 46.76440  14.582140   12.53787        0
## 2  0.00000   0.0000 19.10275   4.932700    3.95306        0
## 3  1.50675   0.0000 36.48778  18.798000   10.32484        0
## 4  0.00000   0.0000 69.07104   2.439577    1.72710        0
## 5  0.00000   0.0000 44.07012   3.130800    5.21406        0
## 6  0.00000   0.0140 65.58897   9.271000    0.80064        0
```

Add the participants' metadata back to totals

Load ind_metadata.txt if you have not done so.

```
ind_metadata <- read.table("ind_metadata.txt", sep="\t", header=T)
```

Add this metadata of each participant to totals.'NA' will be inserted to UserNames which are not in ind_metadata.

```
new_totals_m <- merge(x=new_totals, y=ind_metadata, by="UserName", all.x=T)
```

Check that the items data and metadata are merged.

```
head(new_totals_m)
```

Click to expand output

```
##   UserName   User_Day RecallNo  FoodAmt      KCAL     PROT     TFAT     CARB
## 1 VVKAJ101 VVKAJ101_1        1 2402.500 2314.9040 80.15720 82.97721 337.7973
## 2 VVKAJ101 VVKAJ101_2        2 1671.650  913.5955 35.13461 37.42791 118.5702
## 3 VVKAJ101 VVKAJ101_3        3 1920.588 1604.8050 48.29881 70.08057 215.7318
## 4 VVKAJ102 VVKAJ102_1        1 1957.475 1440.1995 38.54926 80.38863 152.8296
## 5 VVKAJ102 VVKAJ102_2        2 1858.067 1508.9957 46.25199 56.89753 212.5225
## 6 VVKAJ102 VVKAJ102_3        3 2049.400 1834.2306 58.89500 87.53381 211.8413
##       MOIS ALC   CAFF THEO      SUGR     FIBE      CALC      IRON     MAGN
## 1 1879.619   0 26.784 0.00 144.20847 52.62550  889.8455 31.282000 589.3865
## 2 1467.035   0 76.800 0.00  47.67748 21.04413  619.2521  7.366001 273.2332
## 3 1572.133   0 30.240 3.36 111.81870 23.27856 1427.6826 14.628896 320.9569
## 4 1669.286   0 50.400 5.04  25.75794 33.90378  912.4538 11.797247 284.4450
## 5 1521.262   0 22.680 2.52  55.29288 31.25067 1144.7673 14.180575 366.8420
## 6 1666.272   0  0.000 0.00  31.41001 29.35169  700.0389 14.009471 373.1694
##        PHOS     POTA     SODI      ZINC     COPP      SELE       VC      VB1
## 1 1787.4540 4615.400 2298.662 11.776515 1.935984 113.91525 209.0321 1.882480
## 2  794.2740 2093.175 1058.073  5.470231 1.325496  58.42284 147.7427 0.983826
## 3 1249.7137 2571.690 2069.723  8.953749 1.565804  83.81562 137.9377 1.316055
## 4  808.2387 2974.354 2550.088  5.641533 1.278044  42.76070 179.4785 1.137953
## 5 1043.4290 4686.336 2840.423  6.408068 1.427889  38.26618 157.1404 2.909776
## 6 1164.2706 4653.953 4285.981  6.745011 2.066325  72.05487 117.7118 4.570952
##         VB2     NIAC      VB6     FOLA      FA       FF      FDFE     VB12
## 1 2.9713065 18.72793 3.829823 719.3660 179.920 539.4460  846.2460  5.17202
## 2 1.2894892 12.24912 1.386882 372.7493  85.070 287.6793  432.4393  2.23350
## 3 2.0125987 14.49594 4.120610 918.5919 601.800 316.7919 1340.1919 13.23145
## 4 0.9494982 10.92889 1.637870 459.3215  52.025 410.3265  495.5415  1.80990
## 5 1.3127483 13.76033 2.560617 534.1780  20.200 515.9980  548.3180  3.97461
## 6 1.2672856 18.25922 2.540221 506.4956  26.700 481.8156  525.2356  3.70270
##        VARA      RET      BCAR      ACAR      CRYP     LYCO        LZ      ATOC
## 1 1261.6665 673.1850  7004.547  106.7165  56.11800   13.350 14273.582  9.851310
## 2 1412.8761  82.3500 13464.676 4921.4081  81.65188 1174.626  2444.209  4.628156
## 3 1111.7093 504.9250  6476.761 1614.4750  19.37550    4.704  5437.084 16.765260
## 4  310.6182 165.1115  1551.208  344.1073  68.67350 7582.125  1340.894 18.987955
## 5 1265.3238 222.6810 10784.821 3226.1183 198.74500 3380.685  4032.658 16.267538
## 6  344.2824 144.1200  2307.391  143.9169  12.66863 3374.064  3388.637  9.870323
##         VK   CHOLE      SFAT     S040       S060      S080      S100       S120
## 1 674.6513 375.967 18.696057 0.310902 0.11065100 0.1176030 0.2303450 0.18606100
## 2 176.1320  18.300  8.035444 0.137250 0.13725000 0.1401794 0.1440906 0.16560063
## 3 356.2514  64.265 21.762310 0.461520 0.35581000 0.6559880 0.7907570 3.08172600
## 4 193.4643  63.277 12.393455 0.063991 0.04301325 0.0834395 0.1062550 0.06083225
## 5 271.3043  53.980 10.073571 0.086720 0.06900500 0.2074300 0.2042550 0.43357300
## 6 270.4181  65.830 17.110502 0.318160 0.16680000 0.2767274 0.3239814 0.26321963
##        S140      S160     S180     MFAT      M161     M181      M201       M221
## 1 1.2467585 11.448329 4.130060 28.14091 0.7393160 26.99568 0.2349650 0.00881000
## 2 0.5821610  4.955292 1.532696 15.32206 0.3639200 14.79923 0.1110759 0.00346125
## 3 2.6981170  9.237105 3.415553 24.67066 0.4838810 23.65191 0.1867220 0.00864600
## 4 0.1955330  8.961791 2.439567 34.62131 0.9667318 33.33719 0.2755755 0.00450000
## 5 0.5196547  6.309740 1.897680 22.20331 0.4462590 21.47338 0.2112680 0.02609700
## 6 0.8637470 10.355075 3.764050 33.49530 0.3996655 32.60066 0.3411139 0.00658725
##       PFAT      P182     P183 P184     P204    P205    P225    P226      VITD
## 1 28.95861 25.201469 3.525870    0 0.157937 0.00444 0.00000 0.03441  4.881200
## 2 10.24234  9.135821 1.092069    0 0.001315 0.00000 0.00000 0.00000  2.379000
## 3 18.75467 16.930229 1.803171    0 0.006573 0.00000 0.00000 0.00000 10.193000
## 4 28.41416 24.796558 3.534361    0 0.038890 0.00000 0.00000 0.00909  3.375425
## 5 20.44748 17.914195 2.412754    0 0.026967 0.00000 0.00000 0.01881  4.985000
## 6 31.17042 27.616109 3.344330    0 0.038623 0.00032 0.00208 0.00670  2.102000
##      CHOLN VITE_ADD  B12_ADD  F_TOTAL F_CITMLB  F_OTHER F_JUICE  V_TOTAL
## 1 572.5637  0.00000  3.01600 4.037445  0.65995 3.035495  0.3420 3.656365
## 2 143.2537  0.00000  1.41000 0.640500  0.30625 0.323750  0.0105 2.727494
## 3 195.7137  8.34700 10.79370 1.530100  1.01160 0.505000  0.0135 1.906206
## 4 182.4953  6.85640  1.53720 0.000000  0.00000 0.000000  0.0000 4.280535
## 5 212.2748  7.19922  3.21321 0.000000  0.00000 0.000000  0.0000 7.783067
## 6 219.7563  0.00000  2.83410 0.000000  0.00000 0.000000  0.0000 7.084731
##     V_DRKGR V_REDOR_TOTAL V_REDOR_TOMATO V_REDOR_OTHER V_STARCHY_TOTAL
## 1 1.6658400      0.000000        0.00000      0.000000       0.0000000
## 2 0.4840000      1.177100        0.04860      1.128500       0.0000000
## 3 1.0143750      0.300500        0.00000      0.300500       0.0000000
## 4 0.0099900      1.485350        1.48535      0.000000       0.1288625
## 5 0.6186667      1.857575        0.84840      1.009175       3.3200000
## 6 0.7412500      0.848400        0.84840      0.000000       4.6146000
##   V_STARCHY_POTATO V_STARCHY_OTHER   V_OTHER V_LEGUMES G_TOTAL G_WHOLE
## 1           0.0000       0.0000000 1.9905250  1.256750  5.4384  2.6520
## 2           0.0000       0.0000000 1.0663938  0.105000  2.6500  1.7669
## 3           0.0000       0.0000000 0.5913312  0.135000  4.4072  4.2542
## 4           0.0000       0.1288625 2.6563325  1.085300  5.0421  1.5522
## 5           3.3200       0.0000000 1.9868250  0.285000  2.2405  1.0285
## 6           4.6146       0.0000000 0.8804812  0.192375  1.6050  0.0000
##   G_REFINED PF_TOTAL PF_MPS_TOTAL PF_MEAT PF_CUREDMEAT PF_ORGAN PF_POULT
## 1    2.7864 3.111700       0.0000       0       0.0000        0        0
## 2    0.8831 3.316731       0.0000       0       0.0000        0        0
## 3    0.1530 2.501000       0.0000       0       0.0000        0        0
## 4    3.4899 0.644600       0.0000       0       0.0000        0        0
## 5    1.2120 2.226680       0.0000       0       0.0000        0        0
## 6    1.6050 4.697931       0.2016       0       0.2016        0        0
##   PF_SEAFD_HI PF_SEAFD_LOW PF_EGGS PF_SOY PF_NUTSDS PF_LEGUMES D_TOTAL  D_MILK
## 1           0            0  1.8093 0.0000 1.3024000   4.935150 1.34064 0.21168
## 2           0            0  0.0000 1.9200 1.3967312   0.420000 0.75030 0.75030
## 3           0            0  0.0000 0.0000 2.5010000   0.540000 2.39525 0.88850
## 4           0            0  0.3030 0.0000 0.3416000   4.351050 0.15150 0.15150
## 5           0            0  0.2020 1.6660 0.3586800   1.155000 0.57980 0.57980
## 6           0            0  0.2020 3.3558 0.9385313   0.772875 0.70760 0.58900
##   D_YOGURT D_CHEESE     OILS SOLID_FATS ADD_SUGARS A_DRINKS       Diet Gender
## 1  0.00000   0.5544 46.76440  14.582140   12.53787        0 Vegetarian      M
## 2  0.00000   0.0000 19.10275   4.932700    3.95306        0 Vegetarian      M
## 3  1.50675   0.0000 36.48778  18.798000   10.32484        0 Vegetarian      M
## 4  0.00000   0.0000 69.07104   2.439577    1.72710        0      Vegan      F
## 5  0.00000   0.0000 44.07012   3.130800    5.21406        0      Vegan      F
## 6  0.00000   0.0140 65.58897   9.271000    0.80064        0      Vegan      F
##   Age Weight Height      BMI Waist.Circumference
## 1  31     79    186 22.83501                  80
## 2  31     79    186 22.83501                  80
## 3  31     79    186 22.83501                  80
## 4   60     73    163 27.47563                    90
## 5   60     73    163 27.47563                    90
## 6   60     73    163 27.47563                    90
```

Save the merged dataframe as a .txt file.

```
write.table(new_totals_m, "VVKAJ_Tot_m.txt", sep="\t", row.names=F, quote=F)
```

Calculate the mean of totals/participant

Calculate the mean of the totals data across all the days for each participant.

```
AverageBy(data= new_totals, by= "UserName", start.col= "FoodAmt", end.col= "A_DRINKS",
          outfn="VVKAJ_Tot_mean.txt")
```

Load the output for further processing.

```
new_totals_mean <- read.table("VVKAJ_Tot_mean.txt", header=T, sep="\t")
```

The number of rows should be equal to the number of users. This example data has 16 users, so there should be 16 rows of mean totals.

```
nrow(new_totals_mean)
```

```
## [1] 16
```

Add the participants' metadata to the mean totals

Load ind_metadata.txt if you have not done so.

```
ind_metadata <- read.table("ind_metadata.txt", sep="\t", header=T)
```

Add this metadata of each participant in the mean totals. 'NA' will be inserted to UserNames which are not in ind_metadata.

```
new_totals_mean_m <- merge(x=new_totals_mean, y=ind_metadata, by="UserName", all.x=T)
```

Check that the mean totals and the users' metadata are merged.

```
head(new_totals_mean_m, 1)
```

Click to expand output

```
##   UserName  FoodAmt     KCAL    PROT     TFAT     CARB     MOIS ALC   CAFF THEO
## 1 VVKAJ101 1998.246 1611.101 54.5302 63.49523 224.0331 1639.596   0 44.608 1.12
##       SUGR     FIBE     CALC     IRON     MAGN     PHOS     POTA     SODI
## 1 101.2349 32.31606 978.9267 17.75897 394.5255 1277.147 3093.422 1808.819
##       ZINC     COPP     SELE       VC     VB1      VB2     NIAC      VB6
## 1 8.733498 1.609095 85.38457 164.9042 1.39412 2.091131 15.15766 3.112438
##       FOLA     FA       FF    FDFE    VB12     VARA      RET     BCAR   ACAR
## 1 670.2357 288.93 381.3057 872.959 6.87899 1262.084 420.1533 8981.995 2214.2
##       CRYP     LYCO       LZ     ATOC       VK   CHOLE    SFAT     S040
## 1 52.38179 397.5601 7384.958 10.41491 402.3449 152.844 16.1646 0.303224
##       S060      S080      S100     S120     S140     S160     S180     MFAT
## 1 0.201237 0.3045901 0.3883975 1.144463 1.509012 8.546908 3.026103 22.71121
##       M161     M181      M201        M221     PFAT     P182    P183 P184
## 1 0.529039 21.81561 0.1775876 0.006972417 19.31854 17.08917 2.14037    0
##       P204    P205 P225    P226     VITD    CHOLN VITE_ADD  B12_ADD  F_TOTAL
## 1 0.055275 0.00148    0 0.01147 5.817733 303.8437 2.782333 5.073233 2.069348
##    F_CITMLB  F_OTHER F_JUICE  V_TOTAL  V_DRKGR V_REDOR_TOTAL V_REDOR_TOMATO
## 1 0.6592667 1.288082   0.122 2.763355 1.054738     0.4925333         0.0162
##   V_REDOR_OTHER V_STARCHY_TOTAL V_STARCHY_POTATO V_STARCHY_OTHER  V_OTHER
## 1     0.4763333               0                0               0 1.216083
##   V_LEGUMES G_TOTAL  G_WHOLE G_REFINED PF_TOTAL PF_MPS_TOTAL PF_MEAT
## 1 0.4989167  4.1652 2.891033  1.274167 2.976477            0       0
##   PF_CUREDMEAT PF_ORGAN PF_POULT PF_SEAFD_HI PF_SEAFD_LOW PF_EGGS PF_SOY
## 1            0        0        0           0            0  0.6031   0.64
##   PF_NUTSDS PF_LEGUMES  D_TOTAL    D_MILK D_YOGURT D_CHEESE     OILS SOLID_FATS
## 1  1.733377    1.96505 1.495397 0.6168267  0.50225   0.1848 34.11831   12.77095
##   ADD_SUGARS A_DRINKS       Diet Gender Age Weight Height      BMI
## 1    8.93859        0 Vegetarian      M  31     79    186 22.83501
##   Waist.Circumference
## 1                  80
```

Save the merged dataframe as a .txt file.

```
write.table(new_totals_mean_m, "VVKAJ_Tot_mean_m.txt", sep="\t", row.names=F, quote=F)
```

Quality Control (QC) for the mean totals data

Totals data may contain outliers due to errors in dietary reporting. These errors may be due to omission or inaccurate over- or under-estimation of portion size, leading to improbable nutrient totals. ASA24 provides General Guidelines for Reviewing & Cleaning Data for identifying and removing suspicious records.

Here, we will identify records that contain values that fall outside typically observed ranges of kilocalories (KCAL), protein (PROT), total fat (TFAT), and vitamin C (VC). The ASA24 guide provides ranges of beta carotene (BCAR), too, however, outlier checking for BCAR is omitted in this tutorial but can be considered if you identify it as a nutrient that has a high variance in your study dataset.

Please note that your input dataframe (QCtotals) will be overwritten after each outlier removal.

Load your totals if necessary - to be used as input for QC.

```
new_totals_mean_m <- read.table("VVKAJ_Tot_mean_m.txt", sep="\t", header=T)
```

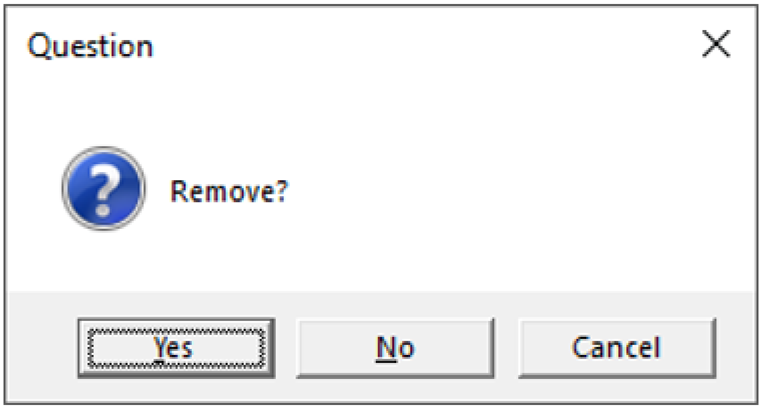
Define your totals dataset to be used as input.

```
QCtotals <- new_totals_mean_m
```

Flag if KCAL is <600 or >5700 → ask remove or not → if yes, remove those rows.

```
QCOutliers(input.data = QCtotals, target.colname = "KCAL", min = 600, max = 5700)
```

This function will print out rows that fall outside the specified min-max range, and a dialogue box will appear outside the R Studio (shown below), asking whether to remove them. You should make sure to review these records carefully to double-check if the removal is warranted. It is possible to have a valid record that could meet the threshold for removal. Only you will know if you can trust the record when working with real data.



If you find potential outlier(s) here, click "No", and view those total(s) with their other nutrient intake information by running the following;

```
KCAL_outliers <- subset(QCtotals, KCAL < 600 | KCAL > 5700)
```

```
# Sort the rows by KCAL and show only the specified variables.
KCAL_outliers[order(KCAL_outliers$KCAL, decreasing = T),
              c('UserName', 'KCAL', 'FoodAmt', 'PROT', 'TFAT', 'CARB')]
```

```
##    UserName     KCAL  FoodAmt     PROT     TFAT     CARB
## 16 VVKAJ117 5789.476 5639.525 194.1051 263.8755 674.0867
```

If you think it is a true outlier, then run the QCOutliers command for KCAL again, and click "Yes" to remove the outlier. Here for this tutorial, we will remove this individual.

```
QCOutliers(input.data = QCtotals, target.colname = "KCAL", min = 600, max = 5700)
```

```
## There are 1 observations with < 600 or > 5700 .
## Remove? (Yes/no/cancel)
## Outlier rows were removed; the cleaned data is saved as an object called "QCtotals".
##  15 rows remained.
```

Continue the QC process with other variables.

Flag if PROT is <10 or >240 → ask remove or not → if yes, remove those rows

```
QCOutliers(input.data = QCtotals, target.colname = "PROT", min = 10, max = 240)
```

```
## There are 0 observations with < 10 or > 240 .
## There are no outlier rows, but the input data was renamed as QCtotals.
##  15 rows remained.
```

Flag if TFAT is <15 or >230 → ask remove or not → if yes, remove those rows

```
  QCOutliers(input.data = QCtotals, target.colname = "TFAT", min = 15, max = 230)
```

```
## There are 0 observations with < 15 or > 230 .
## There are no outlier rows, but the input data was renamed as QCtotals.
##  15 rows remained.
```

Flag if VC (Vitamin C) is <5 or >400 → ask remove or not → if yes, remove those rows.

```
QCOutliers(input.data = QCtotals, target.colname = "VC", min = 5, max = 400)
```

```
## There are 0 observations with < 5 or > 400 .
## There are no outlier rows, but the input data was renamed as QCtotals.
##  15 rows remained.
```

Save as a .txt file.

```
write.table(QCtotals, "VVKAJ_Tot_mean_m_QCed.txt", sep="\t", quote=F, row.names=F)
```

Adjust totals and items after QC

Remove the QC-ed individual(s) from the totals to be consistent

In the previous section, we have removed individual(s) that did not pass the QC from mean total data. We will remove those individual(s) from the totals (before taking means of days), so that we will have the same individuals in the mean_total and total.

Among the individuals in new_totals_m, retain only those in QCtotals.

```
new_totals_m_QCed <- new_totals_m[ new_totals_m$UserName %in% QCtotals$UserName, ]
```

Save as a .txt file. This will be the total for each of the "QC-ed" individuals for each day, to be used for clustering analyses.

```
write.table(new_totals_m_QCed, "VVKAJ_Tot_m_QCed.txt", sep="\t", quote=F, row.names=F)
```

Similarly, remove the QC-ed individual(s) from the items to be consistent with the QC-ed averaged totals

Among the individuals in new_totals_m, pick up only those in QCtotals.

```
items_f_id_s_m_QCed <- items_f_id_s_m[ items_f_id_s_m$UserName %in% QCtotals$UserName, ]
```

Save as a .txt file. This will be the items for each of the “QC-ed” individuals for each day, to be used for ordination etc.

```
write.table(items_f_id_s_m_QCed, "VVKAJ_Items_f_id_s_m_QCed.txt", sep="\t", quote=F, row.names=F)
```

Come back to the main directory before you start running another script.

```
  setwd(main_wd)
```