Data Management for Diet & Nutrition dataset.

Pushpak Sunil Rane

Data Management

```
library(tidyverse)
library(tidyr)
library(dplyr)
```

```
raw_data <- read.csv(here::here("Data/country_environment.csv"))
raw_data1 <- read.csv(here::here("Data/country_adolescent.csv"))
raw_data2 <- read.csv(here::here("Data/country_diet.csv"))
merge_data <- merge(raw_data,raw_data1,raw_data2,
              by.x = c("country","region"),
              by.y = c("country","region"),
              by.z = c("country","region"),all = TRUE)
```

```
mydata  <-  merge_data %>%  filter(region != "") %>% drop_na(planetary_impacts_Cropland,plane
head(mydata)
```

```
      country region iso3.x iso3.1 disaggregation.x disagg.value.x
1 Afghanistan   Asia    AFG    AFG       food.group          sugar
2 Afghanistan   Asia    AFG    AFG       food.group          sugar
3 Afghanistan   Asia    AFG    AFG       food.group      fruit_veg
4 Afghanistan   Asia    AFG    AFG       food.group      fruit_veg
5 Afghanistan   Asia    AFG    AFG       food.group      beef_lamb
6 Afghanistan   Asia    AFG    AFG       food.group      beef_lamb
    subregion.x             section.x environmental_impacts_land_use
1 Southern Asia Environmental Impacts                         2.07634
2 Southern Asia Environmental Impacts                         2.07634
```

```
3 Southern Asia Environmental Impacts                       3.47439
4 Southern Asia Environmental Impacts                       3.47439
5 Southern Asia Environmental Impacts                     511.84080
6 Southern Asia Environmental Impacts                     511.84080
  environmental_impacts_freshwater_use
1                            0.8492800
2                            0.8492800
3                            4.9636600
4                            4.9636600
5                            0.7410889
6                            0.7410889
  environmental_impacts_ghg_emissions_lifecycle
1                                       1.56301
2                                       1.56301
3                                       3.39083
4                                       3.39083
5                                      22.01810
6                                      22.01810
  environmental_impacts_nitrogen_application
1                                   13.48210
2                                   13.48210
3                                   28.60970
4                                   28.60970
5                                    7.83279
6                                    7.83279
  environmental_impacts_phosphorus_application planetary_impacts_Cropland
1                                     2.266780                     2.14531
2                                     2.266780                     2.14531
3                                     5.017480                     6.06917
4                                     5.017480                     6.06917
5                                     1.243977                     4.74712
6                                     1.243977                     4.74712
  planetary_impacts_Freshwater planetary_impacts_Greenhouse.gases
1                    10.988400                             0.51024
2                    10.988400                             0.51024
3                     9.367690                             1.26095
4                     9.367690                             1.26095
5                     1.843905                           101.40700
6                     1.843905                           101.40700
  planetary_impacts_Nitrogen planetary_impacts_Phosphorus iso3.y
1                    5.66280                      4.40048    AFG
2                    5.66280                      4.40048    AFG
3                   10.47100                      8.15182    AFG
```

```
4                   10.47100                      8.15182    AFG
5                    4.11008                      3.57787    AFG
6                    4.11008                      3.57787    AFG
  disaggregation.y disagg.value.y   subregion.y
1              sex          Girls Southern Asia
2              sex           Boys Southern Asia
3              sex          Girls Southern Asia
4              sex           Boys Southern Asia
5              sex          Girls Southern Asia
6              sex           Boys Southern Asia
                                        section.y adolescent_obesity_2000
1 Child and adolescent (aged 5-19) nutrition status                    0.53
2 Child and adolescent (aged 5-19) nutrition status                    0.67
3 Child and adolescent (aged 5-19) nutrition status                    0.53
4 Child and adolescent (aged 5-19) nutrition status                    0.67
5 Child and adolescent (aged 5-19) nutrition status                    0.53
6 Child and adolescent (aged 5-19) nutrition status                    0.67
  adolescent_obesity_2001 adolescent_obesity_2002 adolescent_obesity_2003
1                    0.59                    0.67                    0.75
2                    0.74                    0.82                    0.90
3                    0.59                    0.67                    0.75
4                    0.74                    0.82                    0.90
5                    0.59                    0.67                    0.75
6                    0.74                    0.82                    0.90
  adolescent_obesity_2004 adolescent_obesity_2005 adolescent_obesity_2006
1                    0.85                    0.95                    1.07
2                    0.99                    1.09                    1.20
3                    0.85                    0.95                    1.07
4                    0.99                    1.09                    1.20
5                    0.85                    0.95                    1.07
6                    0.99                    1.09                    1.20
  adolescent_obesity_2007 adolescent_obesity_2008 adolescent_obesity_2009
1                    1.19                    1.33                    1.49
2                    1.31                    1.44                    1.58
3                    1.19                    1.33                    1.49
4                    1.31                    1.44                    1.58
5                    1.19                    1.33                    1.49
6                    1.31                    1.44                    1.58
  adolescent_obesity_2010 adolescent_obesity_2011 adolescent_obesity_2012
1                    1.66                    1.84                    2.05
2                    1.73                    1.89                    2.07
3                    1.66                    1.84                    2.05
4                    1.73                    1.89                    2.07
```

```
5                    1.66                    1.84                    2.05
6                    1.73                    1.89                    2.07
  adolescent_obesity_2013 adolescent_obesity_2014 adolescent_obesity_2015
1                    2.27                    2.52                    2.78
2                    2.26                    2.47                    2.69
3                    2.27                    2.52                    2.78
4                    2.26                    2.47                    2.69
5                    2.27                    2.52                    2.78
6                    2.26                    2.47                    2.69
  adolescent_obesity_2016 adolescent_overweight_2000 adolescent_overweight_2001
1                    3.07                       3.24                       3.48
2                    2.94                       2.77                       3.00
3                    3.07                       3.24                       3.48
4                    2.94                       2.77                       3.00
5                    3.07                       3.24                       3.48
6                    2.94                       2.77                       3.00
  adolescent_overweight_2002 adolescent_overweight_2003
1                       3.74                       4.02
2                       3.24                       3.51
3                       3.74                       4.02
4                       3.24                       3.51
5                       3.74                       4.02
6                       3.24                       3.51
  adolescent_overweight_2004 adolescent_overweight_2005
1                       4.32                       4.64
2                       3.79                       4.08
3                       4.32                       4.64
4                       3.79                       4.08
5                       4.32                       4.64
6                       3.79                       4.08
  adolescent_overweight_2006 adolescent_overweight_2007
1                       4.97                       5.34
2                       4.40                       4.74
3                       4.97                       5.34
4                       4.40                       4.74
5                       4.97                       5.34
6                       4.40                       4.74
  adolescent_overweight_2008 adolescent_overweight_2009
1                       5.72                       6.13
2                       5.10                       5.48
3                       5.72                       6.13
4                       5.10                       5.48
5                       5.72                       6.13
```

```
6                       5.10                       5.48
  adolescent_overweight_2010 adolescent_overweight_2011
1                       6.57                       7.03
2                       5.88                       6.31
3                       6.57                       7.03
4                       5.88                       6.31
5                       6.57                       7.03
6                       5.88                       6.31
  adolescent_overweight_2012 adolescent_overweight_2013
1                       7.52                       8.04
2                       6.77                       7.26
3                       7.52                       8.04
4                       6.77                       7.26
5                       7.52                       8.04
6                       6.77                       7.26
  adolescent_overweight_2014 adolescent_overweight_2015
1                       8.60                       9.18
2                       7.78                       8.32
3                       8.60                       9.18
4                       7.78                       8.32
5                       8.60                       9.18
6                       7.78                       8.32
  adolescent_overweight_2016 adolescent_thinness_2000 adolescent_thinness_2001
1                        9.8                    13.09                    12.97
2                        8.9                    27.30                    27.06
3                        9.8                    13.09                    12.97
4                        8.9                    27.30                    27.06
5                        9.8                    13.09                    12.97
6                        8.9                    27.30                    27.06
  adolescent_thinness_2002 adolescent_thinness_2003 adolescent_thinness_2004
1                    12.84                    12.70                    12.57
2                    26.82                    26.58                    26.33
3                    12.84                    12.70                    12.57
4                    26.82                    26.58                    26.33
5                    12.84                    12.70                    12.57
6                    26.82                    26.58                    26.33
  adolescent_thinness_2005 adolescent_thinness_2006 adolescent_thinness_2007
1                    12.43                    12.29                    12.15
2                    26.08                    25.83                    25.57
3                    12.43                    12.29                    12.15
4                    26.08                    25.83                    25.57
5                    12.43                    12.29                    12.15
6                    26.08                    25.83                    25.57
```

```
  adolescent_thinness_2008 adolescent_thinness_2009 adolescent_thinness_2010
1                    12.02                    11.88                    11.74
2                    25.31                    25.05                    24.78
3                    12.02                    11.88                    11.74
4                    25.31                    25.05                    24.78
5                    12.02                    11.88                    11.74
6                    25.31                    25.05                    24.78
  adolescent_thinness_2011 adolescent_thinness_2012 adolescent_thinness_2013
1                    11.61                    11.47                    11.33
2                    24.50                    24.22                    23.93
3                    11.61                    11.47                    11.33
4                    24.50                    24.22                    23.93
5                    11.61                    11.47                    11.33
6                    24.50                    24.22                    23.93
  adolescent_thinness_2014 adolescent_thinness_2015 adolescent_thinness_2016
1                    11.19                    11.05                    10.92
2                    23.64                    23.35                    23.05
3                    11.19                    11.05                    10.92
4                    23.64                    23.35                    23.05
5                    11.19                    11.05                    10.92
6                    23.64                    23.35                    23.05
  adolescent_thinness_2019
1                       NA
2                       NA
3                       NA
4                       NA
5                       NA
6                       NA
```

Categorical Variable 01

"country" variable represents regions or places for analyzing diet and impact of it by environmental factors.

```
str(mydata$region)
```

```
 chr [1:3956] "Asia" "Asia" "Asia" "Asia" "Asia" "Asia" "Asia" "Asia" ...
```

Does this match with the intended data type?

Yes, it matches the intended data type.

```
table(mydata$region,useNA = "always")
```

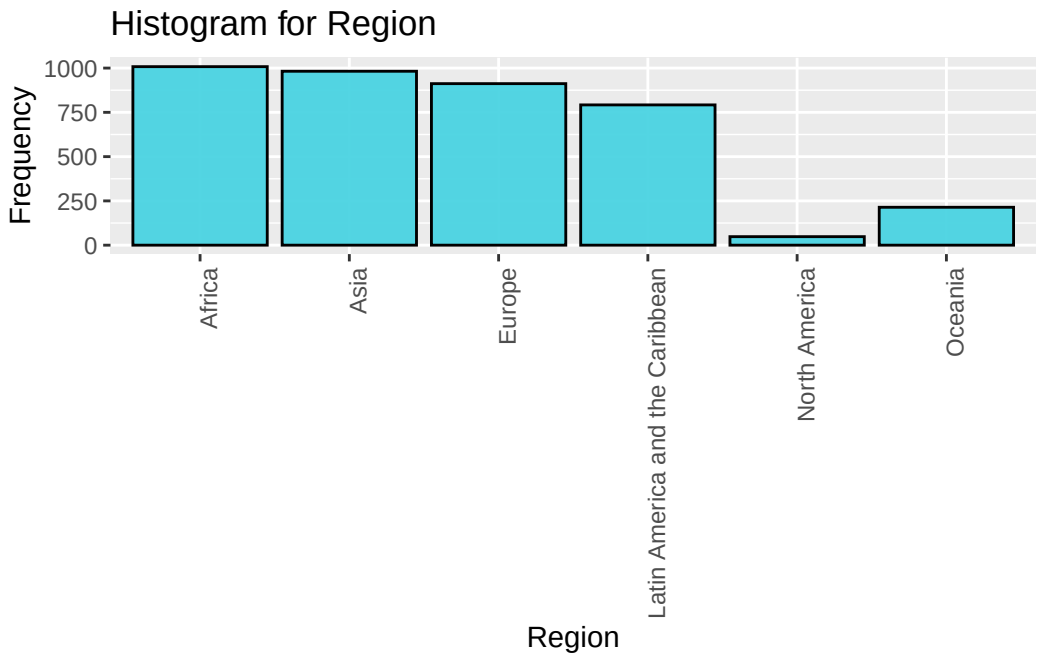
```
                         Africa                                Asia
                           1008                                 982
                         Europe Latin America and the Caribbean
                            912                                 792
                  North America                             Oceania
                             48                                 214
                           <NA>
                              0
```

There are 0 entries that do not match the expected categories of country.

```
summary(mydata$region)
```

```
   Length     Class      Mode
     3956 character character
```

```
library(ggplot2)
ggplot(mydata, aes(x = region)) +
  geom_histogram(stat = "count",fill = "#4cd4e2f5", color = "black") +
  labs(title = "Histogram for Region", x = "Region", y = "Frequency") +  theme(axis.text.x =
```



Categorical Variable 02

"*disagg.value*" variable represents diet items(fruits,vegetables,beef,etc) in particular country.

```
class(mydata$disagg.value.x)
```

```
[1] "character"
```

Does this match with the intended data type?

Yes, it matches the intended data type.

```
table(mydata$disagg.value.x,useNA="always")
```

```
   beef_lamb         eggs         fish    fruit_veg       grains      legumes
```

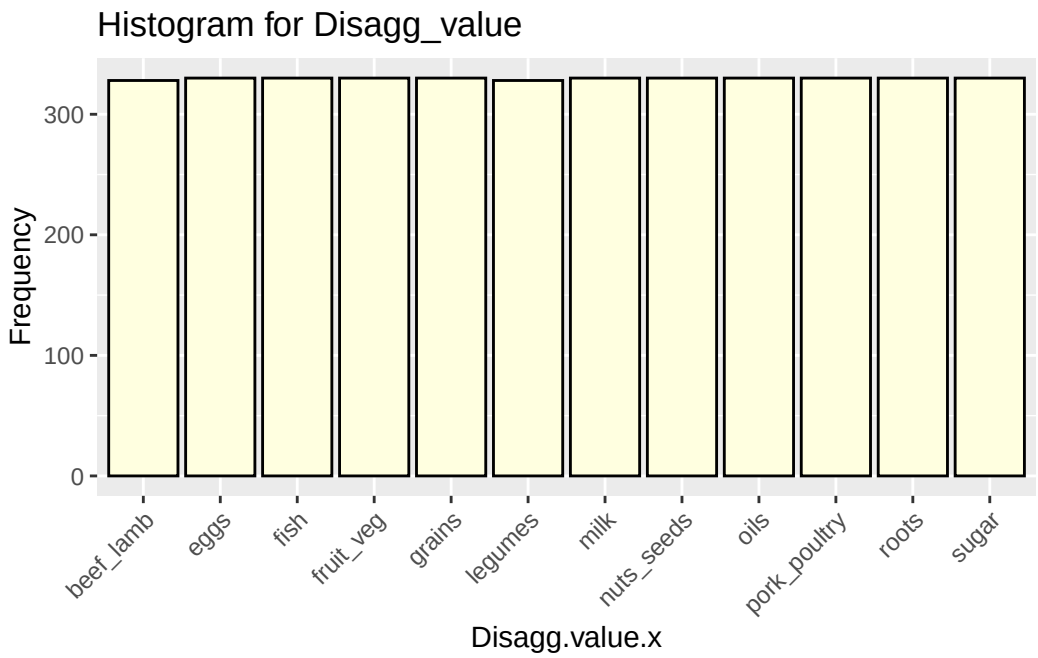
```
      328          330          330          330          330          328
     milk   nuts_seeds         oils pork_poultry        roots        sugar
      330          330          330          330          330          330
     <NA>
        0
```

There are 0 entries that do not match the expected categories of country.

```
summary(mydata$disagg.value.x)
```

```
   Length     Class      Mode
     3956 character character
```

```
# Create histogram for variable disagg_value using ggplot library
ggplot(mydata, aes(x = disagg.value.x)) +
  geom_histogram(stat = "count", fill = "lightyellow", color = "black") +
  labs(title = "Histogram for Disagg_value", x = "Disagg.value.x", y = "Frequency") + theme(a
```



Quantitative Variable 01

"environmental_impacts_freshwater_use" variable represents environmental impact of using freshwater on diet and nutrition in certain country.

```
class(mydata$planetary_impacts_Cropland)
```

```
[1] "numeric"
```

Does this match with the intended data type?

Yes, it matches the intended data type.

```
mydata$impact_cropland <- ifelse(mydata$planetary_impacts_Cropland > 21,
"less_impact_Cropland", "more_impact_Cropland")
table(mydata$impact_cropland, useNA="always")
```

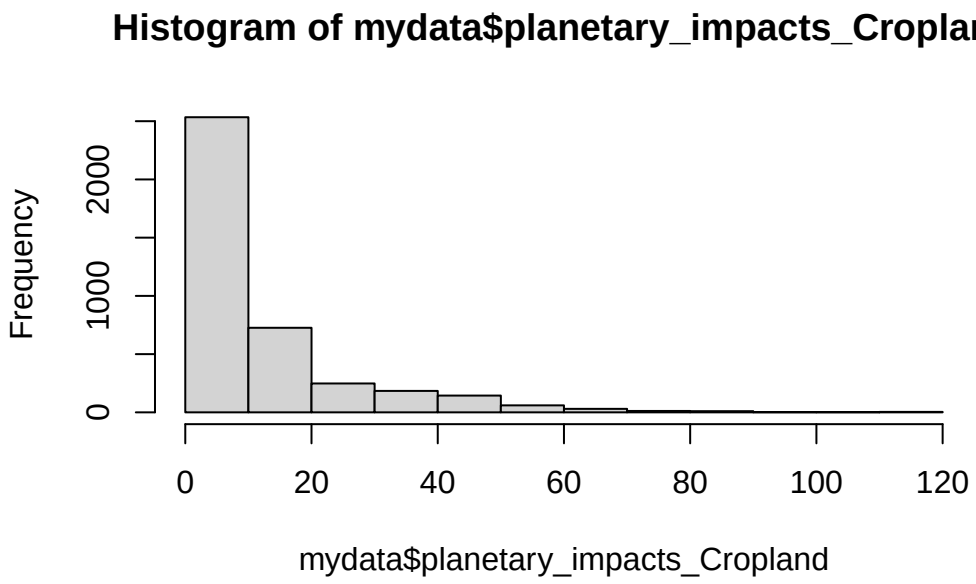
```
less_impact_Cropland more_impact_Cropland                 <NA>
                 662                  3294                    0
```

There are 4 entries that do not match the expected quantity. Mostly, either they are out of range or error values in the dataset. Otherwise, entries beside that 4 all are in range.

```
summary(mydata$planetary_impacts_Cropland)
```

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  0.00169   2.33686   6.55000  11.78541  14.45801 115.46900
```

```
hist(mydata$planetary_impacts_Cropland)
```



Quantitative Variable 02

"environmental_impacts_land_use" variable represents environmental impact of land on diet and nutrition in certain regions.

```
class(mydata$planetary_impacts_Freshwater)
```

```
[1] "numeric"
```

Does this match with the intended data type?

Yes, it matches the intended data type.

```
mydata$impact_freshwater <- ifelse(mydata$planetary_impacts_Freshwater > 30,
"less_impact_freshwater", "more_impact_freshwater")
table(mydata$impact_freshwater, useNA="always")
```

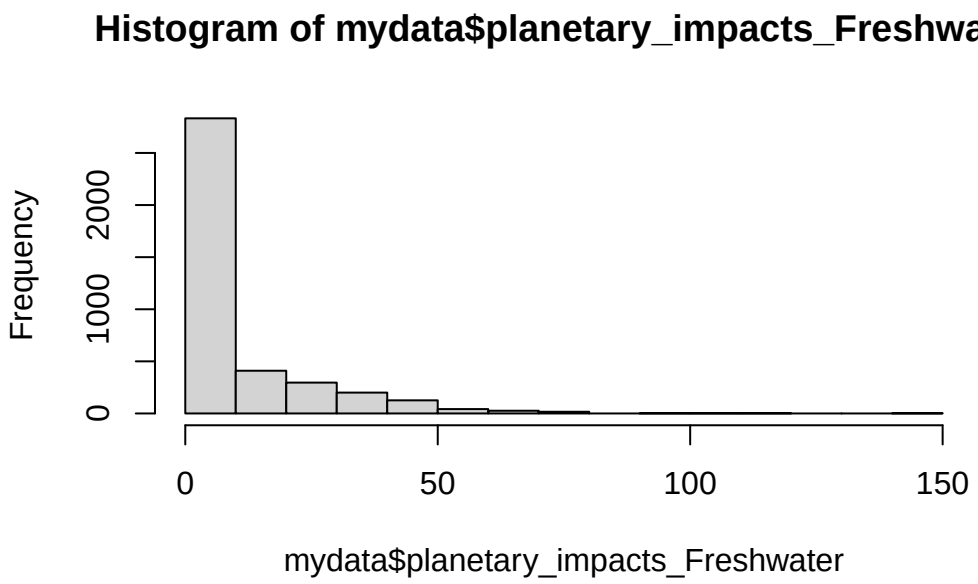
```
less_impact_freshwater more_impact_freshwater                    <NA>
                   418                    3538                       0
```

```
summary(mydata$planetary_impacts_Freshwater)
```

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  0.00207   1.20060   3.46803  10.00658  12.63189 140.65600
```

There are 16 entries that do not match the expected quantity. Mostly, either they are out of range or error values in the dataset. Otherwise, entries beside that 16 all are in range.

```
hist(mydata$planetary_impacts_Freshwater)
```




```
save(mydata, file=here::here("data/Data_clean.Rdata"))
```