

Food Nutrition Analysis

Abstract:-

Food and nutrition are the way that we get fuel, providing energy for our bodies. We need to replace nutrients in our bodies with a new supply every day.

Water is an important component of nutrition.

The examples of nutrients are:- carbohydrates, proteins, fats, etc.

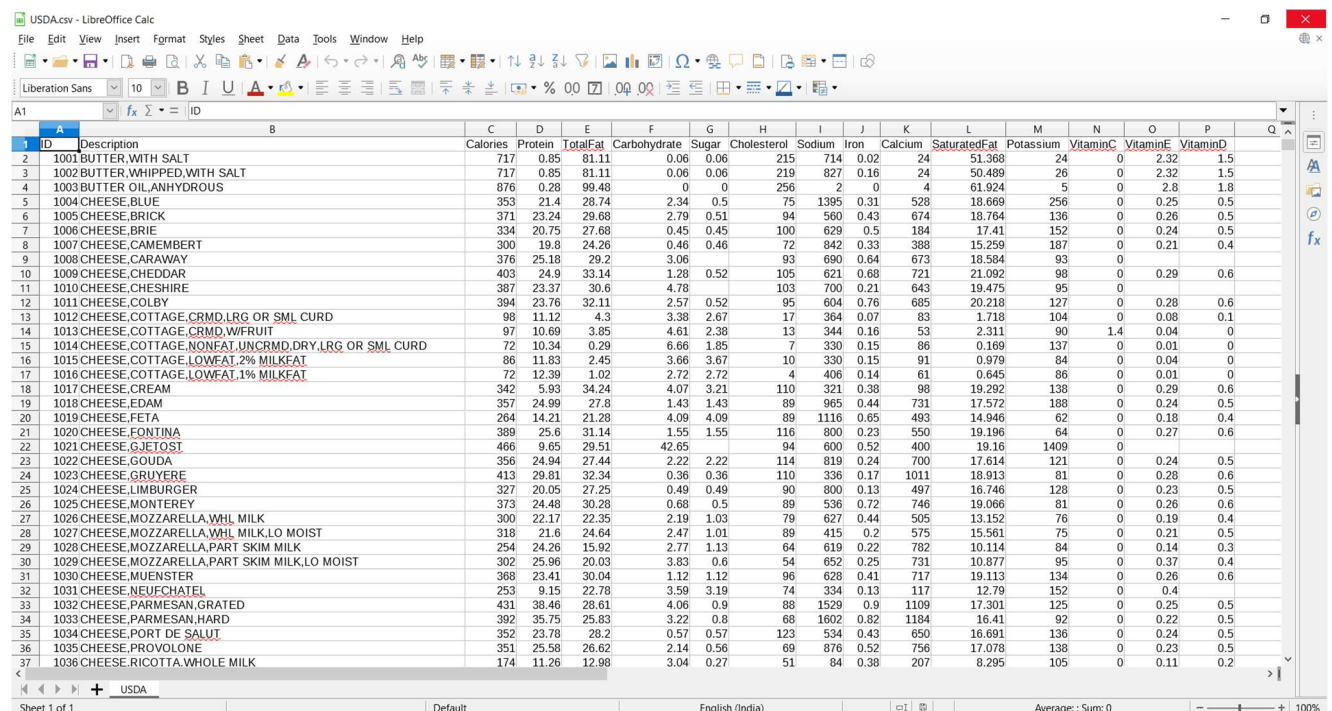
To maintain a good health vitamins and minerals are very important.

For pregnant women and adults over 50, vitamins such as vitamin D and minerals such as calcium and iron are important to consider when choosing foods to eat, as well as possible dietary supplements.

Here, we've used a standard dataset of food items with nutrients named USDA.csv.

This Dataset consist of foods with nutrients available in that food such as Fats, Carbohydrates, Vitamins, Proteins, Calories, Et al.

The figure below is a glimpse of this USDA.csv dataset.



ID	Description	Calories	Protein	Total Fat	Carbohydrate	Sugar	Cholesterol	Sodium	Iron	Calcium	Saturated Fat	Potassium	Vitamin C	Vitamin E	Vitamin D
1001	BUTTER, WITH SALT	717	0.85	81.11	0.06	0.06	215	714	0.02	24	51.368	24	0	2.32	1.5
1002	BUTTER, WHIPPED, WITH SALT	717	0.85	81.11	0.06	0.06	219	827	0.16	24	50.489	26	0	2.32	1.5
1003	BUTTER OIL, ANHYDROUS	876	0.28	99.48	0	0	256	2	0	4	61.924	5	0	2.8	1.8
1004	CHEESE, BLUE	353	21.4	28.74	2.34	0.5	75	1395	0.31	528	18.669	256	0	0.25	0.5
1005	CHEESE, BRICK	371	23.24	29.68	2.79	0.51	94	560	0.43	674	18.764	136	0	0.26	0.5
1006	CHEESE, BRIE	334	20.75	27.68	0.45	0.45	100	629	0.5	184	17.41	152	0	0.24	0.5
1007	CHEESE, CAMEMBERT	300	19.8	24.26	0.46	0.46	72	842	0.33	388	15.259	187	0	0.21	0.4
1008	CHEESE, CARAWAY	376	25.18	29.2	3.06		93	690	0.64	673	18.584	93	0		
1009	CHEESE, CHEDDAR	403	24.9	33.14	1.28	0.52	105	621	0.68	721	21.092	98	0	0.29	0.6
1010	CHEESE, CHESHIRE	387	23.37	30.6	4.78		103	700	0.21	643	19.475	95	0		
1011	CHEESE, COLBY	394	23.76	32.11	2.57	0.52	95	604	0.76	685	20.218	127	0	0.28	0.6
1012	CHEESE, COTTAGE, CRMD, LRG OR SML CURD	98	11.12	4.3	3.38	2.67	17	364	0.07	83	1.718	104	0	0.08	0.1
1013	CHEESE, COTTAGE, CRMD, W/FRUIT	97	10.89	3.95	4.61	2.38	13	344	0.16	53	2.311	90	1.4	0.04	0
1014	CHEESE, COTTAGE, NONFAT, UNCRMD, DRY, LRG OR SML CURD	72	10.34	0.29	6.66	1.85	7	330	0.15	86	0.169	137	0	0.01	0
1015	CHEESE, COTTAGE, LOWFAT, 2% MILKFAT	86	11.83	2.45	3.66	3.67	10	330	0.15	91	0.979	84	0	0.04	0
1016	CHEESE, COTTAGE, LOWFAT, 1% MILKFAT	72	12.39	1.02	2.72	2.72	4	406	0.14	61	0.645	86	0	0.01	0
1017	CHEESE, CREAM	342	5.93	34.24	4.07	3.21	110	321	0.38	98	19.292	138	0	0.29	0.6
1018	CHEESE, EDAM	357	24.99	27.8	1.43	1.43	89	965	0.44	731	17.572	188	0	0.24	0.5
1019	CHEESE, FETA	264	14.21	21.28	4.09	4.09	89	1116	0.65	493	14.946	62	0	0.18	0.4
1020	CHEESE, FONTINA	389	25.6	31.14	1.55	1.55	116	800	0.23	550	19.196	64	0	0.27	0.6
1021	CHEESE, GJETOST	466	9.65	29.51	42.65		94	600	0.52	400	19.16	1409	0		
1022	CHEESE, GOUDA	356	24.94	27.44	2.22	2.22	114	819	0.24	700	17.614	121	0	0.24	0.5
1023	CHEESE, GRUYERE	413	29.81	32.34	0.36	0.36	110	336	0.17	1011	18.913	81	0	0.28	0.6
1024	CHEESE, LMBURGER	327	20.05	27.25	0.49	0.49	90	800	0.13	497	16.746	128	0	0.23	0.5
1025	CHEESE, MONTEREY	373	24.48	30.28	0.68	0.5	89	536	0.72	746	19.066	81	0	0.26	0.6
1026	CHEESE, MOZZARELLA, WHL MILK	300	22.17	22.35	2.19	1.03	79	627	0.44	505	13.152	76	0	0.19	0.4
1027	CHEESE, MOZZARELLA, WHL MILK, LO MOIST	318	21.6	24.64	2.47	1.01	89	415	0.2	575	15.561	75	0	0.21	0.5
1028	CHEESE, MOZZARELLA, PART SKIM MILK	254	24.26	15.92	2.77	1.13	64	619	0.22	782	10.114	84	0	0.14	0.3
1029	CHEESE, MOZZARELLA, PART SKIM MILK, LO MOIST	302	25.96	20.03	3.83	0.6	54	652	0.25	731	10.877	95	0	0.37	0.4
1030	CHEESE, MUESNTER	368	23.41	30.04	1.12	1.12	96	628	0.41	717	19.113	134	0	0.26	0.6
1031	CHEESE, NEUFCHATEL	253	9.15	22.78	3.59	3.19	74	334	0.13	117	12.79	152	0	0.4	
1032	CHEESE, PARMESAN, GRATED	431	38.46	28.61	4.06	0.9	88	1529	0.9	1109	17.301	125	0	0.25	0.5
1033	CHEESE, PARMESAN, HARD	392	35.75	25.83	3.22	0.8	68	1602	0.82	1184	16.41	92	0	0.22	0.5
1034	CHEESE, PORT DE SALUT	352	23.78	28.2	0.57	0.57	123	534	0.43	650	16.691	136	0	0.24	0.5
1035	CHEESE, PROVOLONE	351	25.58	26.62	2.14	0.56	69	876	0.52	756	17.078	138	0	0.23	0.5
1036	CHEESE, RICOTTA, WHOLE MILK	174	11.26	12.98	3.04	0.27	51	84	0.38	207	8.295	105	0	0.11	0.2

Contents:-

- Introduction
- Methods
- Implementation, results and discussion
- Conclusion

Introduction

Here, by analysing the dataset,

- We'll determine which are the food items that provides us the maximum calories, proteins, total fat, carbohydrates, sugar and cholesterol
- We'll also determine what are the good sources of nutrients such as good sources of calories, proteins, carbohydrates and what are the high cholesterol food items that are to be avoided
- Determine what are the low fat and high fat food items
- Determine the statistics for each nutrient i.e summary(minimum, mean, maximum, etc)
- Also, the frequency distribution of food items with their calorie values
- Obtain relationship between fat and proteins
- Distribution of food items with their sugar values
- Determine the food items that are to be taken which provides the recommended daily intake of nutrients
- Determine the daily mineral intake by an individual based on the gender and age
- Determine the relative contribution of each food to the total intake of a given nutrient

Methods

Steps:-

1. Installing the packages NutrienTrackeR and ggplot2
2. Importing the packages using library function:
`library(NutrienTrackeR)`
`library(ggplot2)`
3. Importing the dataset/csv file and creating a data frame
`dataset<-read.csv("USDA.csv")`
4. Selecting data with 1000 entries and 10 columns as the actual data is too large to be used
`fd<-dataset[1:1000,1:10]`
5. Now 1st determining the food items which provides maximum calories by determining the row/entry which has the max calories
`cal<-which.max(fd$Calories)`
6. Followed by determining the name of the item with the given row
`fd$Description[cal]`
7. Determine the summary of calories
`summary(fd$Calories)`
8. Showing the frequency distribution of items with their calorie values by plotting a histogram
`hist(fd$Calories)`
9. Determine the sources which gives good amount of calories by selecting a subset of data that satisfies our constraint
`subset(fd$Description,fd$Calories>=200 & fd$Calories<500)`
10. Repeating steps 5,6,7 and 9 for Proteins by replacing Calories with Proteins and 9th step by changing lower bound to 12 and upper bound to 25
11. Repeating steps 5,6,7 for Fats by replacing Calories with TotalFat
12. Determine the relationship between fat and proteins by making use of scatter plot
`plot(x=fd$Protein,y=fd$TotalFat,xlab="Protein",ylab="Fat",main="Fat vs Protein")`

13. Repeating steps 5,6,7 and 9 for Carbohydrates by replacing Calories with Carbohydrates and 9th step by changing lower bound to 15 and upper bound to 65
14. Repeating steps 5,6,7 and 9 for Sugar by replacing Calories with Sugar
15. Show the distribution of food items based in their sugar value by plotting a boxplot
`boxplot(fd$Sugar,ylab="Sugar(mg)")`
16. Repeating steps 5,6,7 and 9 for Cholesterol by replacing Calories with Cholesterol and 9th step by changing lower bound to 175 and upper bound to 485
17. Determine the food items based on the recommended daily intake of nutrients
`subset(fd$Description,fd$Protein<=50 & fd$Sugar<=90 & fd$Carbohydrate<=130 & fd$Sodium<=2.3)`
18. Determine the daily mineral intake by an individual based on the gender and age by using the Diet balance function of NutrienTrackerR package
`daily_intake <- dietBalance(my_daily_food = sample_diet_USDA, food_database = "USDA", age = 27, gender = "female")`
19. Last but not the least, using the function `nutrientPiePlot()` of the package `ggplot2` that generates a pie-chart showing the relative contribution of each food to the total intake of a given nutrient.
`nutrientPiePlot(daily_intake, nutrient_name = "Iron, Fe (mg)")`

Implementation, Results and Discussion

Code:-

```
library(NutrienTrackerR)
```

```
library(ggplot2)
```

```
#importing the dataset and creating the dataframe
```

```
dataset<-read.csv("USDA.csv")
```

```
#selecting data with 1000 entries and 9 columns as the actual data is too large
```

```
fd<-dataset[1:1000,1:10]
```

```
#determining the row/entry which has the max calories
```

```
cal<-which.max(fd$Calories)
```

```
fd$Description[cal] #determining the name of the item with the given row
```

```
summary(fd$Calories) #summary of calories i.e what is the max, min value of Calories  
found, avg,mode,etc
```

```
hist(fd$Calories) #shows the frequency distribution of items with their calorie values
```

```
subset(fd$Description,fd$Calories>=200 & fd$Calories<500) #sources which gives good  
amount of calories
```

```
prtn<-which.max(fd$Protein)
```

```
fd$Description[prtn]
```

```
summary(fd$Protein)
```

```
subset(fd$Description,fd$Protein>=12 & fd$Protein<=25) #good source of proteins
```

```
tfmx<-which.max(fd$TotalFat)
```

```
fd$Description[tfmx] #food item with max fat
```

```
tfmi<-which.min(fd$TotalFat)
```

```
fd$Description[tfmi] #food item with min fat
```

```
summary(fd$TotalFat)
```

```
plot(x=fd$Protein,y=fd$TotalFat,xlab="Protein",ylab="Fat",main="Fat vs Protein")
```

```
#scatter plot to express the relationship between protein and fat graphically
```

```
carb<-which.max(fd$Carbohydrate)
```

```
fd$Description[carb]
```

```
summary(fd$Carbohydrate)
```

```
subset(fd$Description,fd$Carbohydrate>=15 & fd$Carbohydrate<=65) #good source of  
carbohydrates
```



```
sg<-which.max(fd$Sugar)
```

```
fd$Description[sg]
```

```
summary(fd$Sugar)
```

```
boxplot(fd$Sugar,ylab="Sugar(mg)") #distribution of food items based on their sugar value
```

```
chol<-which.max(fd$Cholesterol)
```

```
fd$Description[chol]
```

```
summary(fd$Cholesterol)
```

```
subset(fd$Description,fd$Cholesterol>=175 & fd$Cholesterol<=485) #high cholesterol food  
to avoid
```

```
#food items based on the recommended daily intake of nutrients
```

```
subset(fd$Description,fd$Protein<=50 & fd$Sugar<=90 & fd$Carbohydrate<=130 &  
fd$Sodium<=2.3)
```

#The function dietBalance() calculates the daily nutrient intake of an individual and compares it with the NIH recommendations (recommended dietary allowances (RDA))

```
daily_intake <- dietBalance(my_daily_food = sample_diet_USDA, food_database =  
"USDA", age = 27, gender = "female")
```

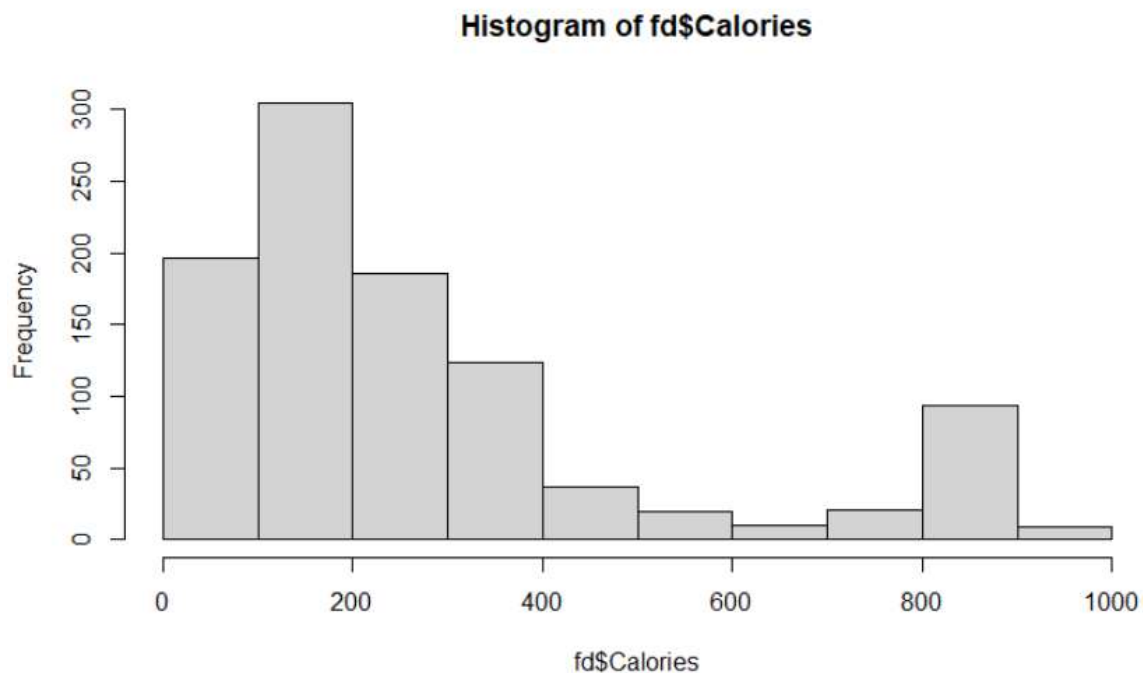
#The function nutrientPiePlot() generates a pie-chart showing the relative contribution of each food to the total intake of a given nutrient.

```
nutrientPiePlot(daily_intake, nutrient_name = "Iron, Fe (mg)")
```

Output:-

```
> library(NutrientTracker)  
> library(ggplot2)  
> #importing the dataset and creating the dataframe  
> dataset<-read.csv("USDA.csv")  
> #selecting data with 1000 entries and 9 columns as the actual data is too large  
> fd<-dataset[1:1000,1:10]  
> #determining the row/entry which has the max calories  
> cal<-which.max(fd$Calories)  
> fd$Description[cal] #determining the name of the item with the given row  
[1] "FAT,BEEF TALLOW"  
> summary(fd$Calories) #summary of calories i.e what is the max, min value of Calories found, avg,mode,etc  
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.        
    0     119     200     285     345     902        
> hist(fd$Calories) #shows the frequency distribution of items with their calorie values  
> subset(fd$Description,fd$Calories>=200 & fd$Calories<500) #sources which gives good amount of calories  
[1] "CHEESE,BLUE"  
[2] "CHEESE,BRICK"  
[3] "CHEESE,BRIE"  
[4] "CHEESE,CAMEMBERT"  
[5] "CHEESE,CARAWAY"
```

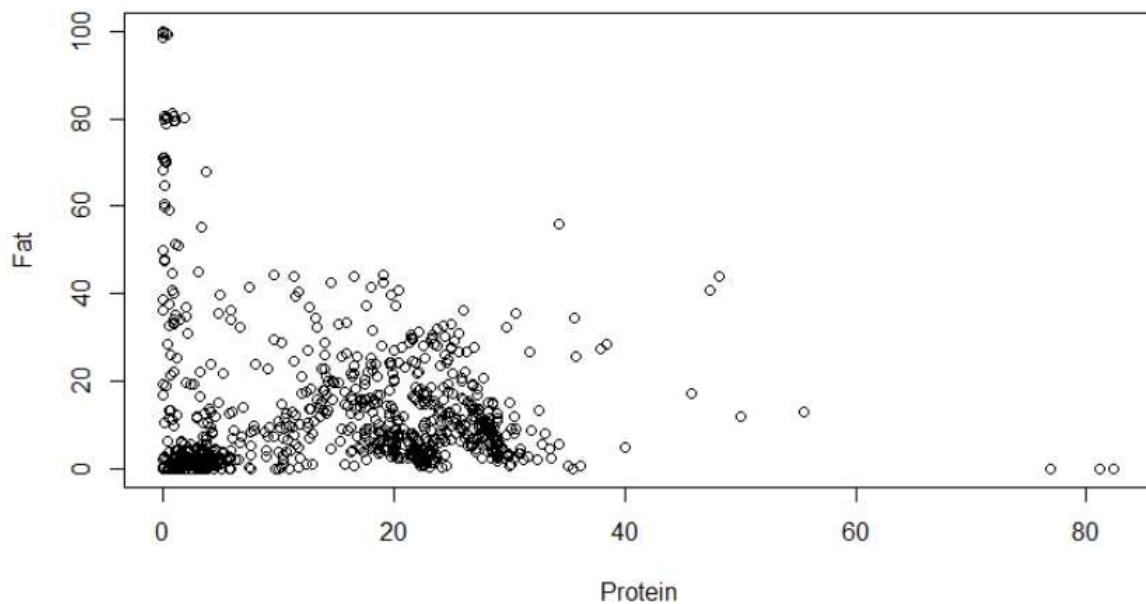
```
[32] "CHEESE,PAST PROCESS,AMERICAN,FORT W/ VITAMIN D"  
[33] "CHEESE,PAST PROCESS,PIMENTO"  
[34] "CHEESE,PAST PROCESS,SWISS"  
[35] "CHEESE FD,COLD PK,AMERICAN"  
[36] "CHEESE FD,PAST PROCESS,AMERICAN,VITAMIN D FORT"  
[37] "CHEESE FD,PAST PROCESS,SWISS"  
[38] "CHEESE SPRD,PAST PROCESS,AMERICAN"  
[39] "CREAM,FLUID,LT WHIPPING"  
[40] "CREAM,FLUID,HVY WHIPPING"  
[41] "CREAM,WHIPPED,CRM TOPPING,PRESSURIZED"  
[42] "DESSERT TOPPING,PRESSURIZED"  
[43] "DESSERT TOPPING,SEMI SOLID,FRZ"  
[44] "SOUP CRM,IMITN,CULTURED"  
[45] "MILK,DRY,WHL,W/ ADDED VITAMIN D"  
[46] "MILK,DRY,NONFAT,REG,WO/ ADDED VIT A & VITAMIN D"  
[47] "MILK,DRY,NONFAT,INST,W/ ADDED VIT A & VITAMIN D"  
[48] "MILK,DRY,NONFAT,CA RED"  
[49] "MILK,BUTTERMILK,DRIED"  
[50] "MILK,CND,COND,SWTND"  
[51] "WHEY ACID DRIED"
```



```
> prtn<-which.max(fd$Protein)
> fd$Description[prtn]
[1] "EGG,WHITE,DRIED,PDR,GLUCOSE RED"
> summary(fd$Protein)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00   1.96   12.19   13.14   22.73   82.40
> subset(fd$Description,fd$Protein>=12 & fd$Protein<=25) #good source of proteins
[1] "CHEESE,BLUE"
[2] "CHEESE,BRICK"
[3] "CHEESE,BRIE"
[4] "CHEESE,CAMEMBERT"
[5] "CHEESE,CHEDDAR"
[6] "CHEESE,CHESHIRE"
[7] "CHEESE,COLBY"
[8] "CHEESE,COTTAGE,LOWFAT,1% MILKFAT"
[9] "CHEESE,EDAM"
[10] "CHEESE,FETA"
[11] "CHEESE,GOUDA"
[12] "CHEESE,LIMBURGER"
[13] "CHEESE,MONTEREY"
```

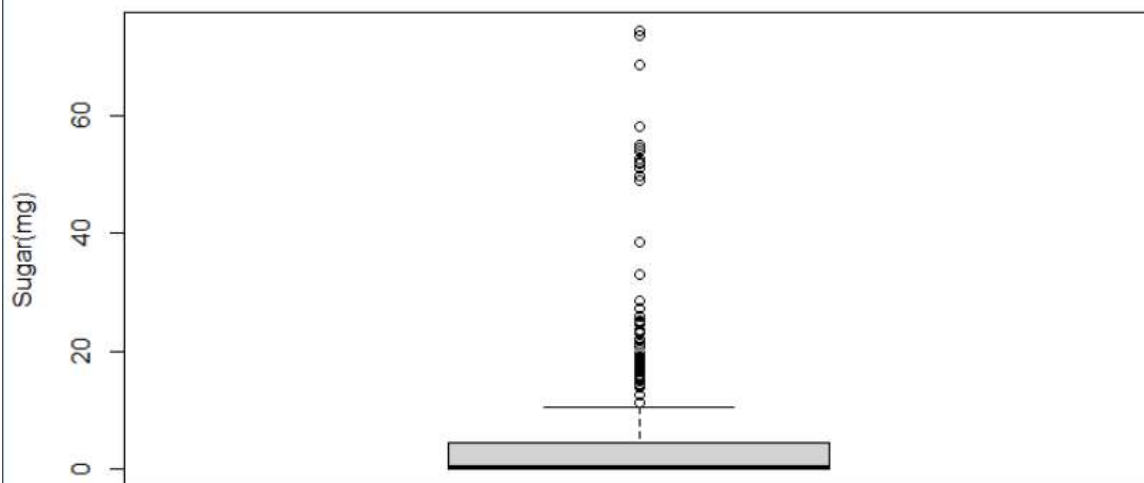
```
[319] "SOUP,CHICK BROTH OR BOUILLON,DRY,DRY"
[320] "SOUP,BEEF BROTH,CUBED,DRY"
[321] "SOUP,BF NOODLE,DRY,MIX"
[322] "SOUP,CHICK BROTH OR BOUILLON,DRY"
[323] "SOUP,CHICK BROTH CUBES,DRY"
[324] "SAUCE,CHS,DRY,PDR"
[325] "GRAVY,UNSPEC TYPE,DRY"
[326] "SOUP,CHICK NOODLE,DRY,MIX"
[327] "SAUCE,SOFRITO,PREP FROM RECIPE"
> tfmx<-which.max(fd$TotalFat)
> fd$Description[tfmx] #food item with max fat
[1] "FAT,BEEF TALLOW"
> tfmi<-which.min(fd$TotalFat)
> fd$Description[tfmi] #food item with min fat
[1] "EGG,WHITE,RAW,FROZEN"
> summary(fd$TotalFat)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.118   9.075  22.241  24.378 100.000
> plot(x=fd$Protein,y=fd$TotalFat,xlab="Protein",ylab="Fat",main="Fat vs Protein")
>
```

Fat vs Protein



```
> carb<-which.max(fd$Carbohydrate)
> fd$Description[carb]
[1] "CINNAMON,GROUND"
> summary(fd$Carbohydrate)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  0.000   1.245   8.785   8.780  80.590
> subset(fd$Description,fd$Carbohydrate>=15 & fd$Carbohydrate<=65) #good source of carbohydrates
[1] "CHEESE,GJETOST"
[2] "CREAM SUBSTITUTE,POWDERED"
[3] "DESSERT TOPPING,POWDERED"
[4] "DESSERT TOPPING,PDR,1.5 OZ PREP W/1/2 CUP MILK"
[5] "DESSERT TOPPING,PRESSURIZED"
[6] "DESSERT TOPPING,SEMI SOLID,FRZ"
[7] "MILK,DRY,WHL,W/ ADDED VITAMIN D"
[8] "MILK,DRY,NONFAT,REG,WO/ ADDED VIT A & VITAMIN D"
[9] "MILK,DRY,NONFAT,INST,W/ ADDED VIT A & VITAMIN D"
[10] "MILK,DRY,NONFAT,CA RED"
[11] "MILK,BUTTERMILK,DRIED"
[12] "MILK,CND,COND,SWTND"
[13] "MILK SHAKES,THICK CHOC"

[139] "GRAVY,PORK,DRY,PDR"
[140] "GRAVY,UNSPEC TYPE,DRY"
[141] "SOUP,CHICK NOODLE,DRY,MIX"
[142] "SAUCE,MOLE POBLANO,DRY MIX,SINGLE BRAND"
[143] "SAUCE,BARBECUE"
[144] "SAUCE,PLUM,READY-TO-SERVE"
[145] "SOUP,TOMATO BISQUE,CND,COND"
[146] "SOUP,TOMATO,CND,COND"
[147] "SAUCE,HOISIN,RTS"
[148] "SAUCE,TERIYAKI,RTS,RED NA"
[149] "CAMPBELL SOUP COMPANY,RED&WHITE,25% LESS NA TOMATO SOUP,COND"
> sg<-which.max(fd$Sugar)
> fd$Description[sg]
[1] "WHEY,SWEET,DRIED"
> summary(fd$Sugar)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 0.000  0.000   0.400   4.341   4.455  74.460    297
> boxplot(fd$Sugar,ylab="Sugar(mg)") #distribution of food items based on their sugar value
> |
```



```
> chol<-which.max(fd$Cholesterol)
> fd$Description[chol]
[1] "EGG,YOLK,DRIED"
> summary(fd$Cholesterol)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
   0.00   1.00   46.50  72.15  88.00 2052.00    12
> subset(fd$Description,fd$Cholesterol>=175 & fd$Cholesterol<=485) #high cholesterol food to avoid
[1] "BUTTER,WITH SALT"
[2] "BUTTER,WHIPPED,WITH SALT"
[3] "BUTTER OIL,ANHYDROUS"
[4] "EGG,WHL,RAW,FRSH"
[5] "EGG,WHL,CKD,FRIED"
[6] "EGG,WHL,CKD,HARD-BOILED"
[7] "EGG,WHOLE,COOKED,OMELET"
[8] "EGG,WHL,CKD,POACHED"
[9] "EGG,WHL,CKD,SCMRBLD"
[10] "BUTTER,WITHOUT SALT"
[11] "EGG,WHOLE,RAW,FROZEN"
[12] "EGG,WHL,RAW,FRZ,SALTED"
[13] "ETSU OTL SALMON"
```

```
> subset(fd$Description,fd$Protein<=50 & fd$Sugar<=90 & fd$Carbohydrate<=130 & fd$Sodium<=2.3)
[1] "BUTTER OIL,ANHYDROUS"
[2] "VINEGAR,DISTILLED"
[3] "FAT,BEEF TALLOW"
[4] "LARD"
[5] "SHORTENING,HOUSEHOLD,PARTIALLY HYDROG SOYBN -COTTONSEED"
[6] "OIL,SOYBN,SALAD OR COOKING,(PARTIALLY HYDROGENATED)"
[7] "OIL,RICE BRAN"
[8] "OIL,WHEAT GERM"
[9] "OIL,PNUT,SALAD OR COOKING"
[10] "OIL,SOYBN,SALAD OR COOKING"
[11] "OIL,COCNT"
[12] "OIL,OLIVE,SALAD OR COOKING"
[13] "OIL,PALM"
[14] "OIL,SESAME,SALAD OR COOKING"
[15] "OIL,SUNFLOWER,LINOLEIC (LESS THAN 60%)"
[16] "SALAD DRSNG,HOME RECIPE,VINEGAR&OIL"
[17] "SALAD DRSNG,FRENCH DRSNG,COMM,REG,WO/ SALT"
[18] "OIL,COCOA BUTTER"
[19] "OIL,COTNSD,SALAD OR COOKING"
```

```

> daily_intake <- dietBalance(my_daily_food = sample_diet_USDA, food_database = "USDA", age = 27, gender = "female")
The results correspond to an average of 7 days
Total energy intake (kcal): 1795
The intake level of the following nutrients is below the RDA:

```

	nutrient	proportion(%RDA)	group
1	Water (g)	30.97156	macronutrient
2	Calcium, Ca (mg)	58.25710	mineral
3	Sodium, Na (mg)	59.84527	mineral
4	Potassium, K (mg)	67.10106	mineral
5	Iron, Fe (mg)	99.56111	mineral
6	Vitamin D (D2 + D3) (ug)	11.28667	vitamin
7	Vitamin E (alpha-tocopherol) (mg)	31.04000	vitamin
8	Vitamin A, RAE (ug)	45.67343	vitamin
9	Choline, total (mg)	85.32682	vitamin
10	Folate, DFE (ug)	96.62500	vitamin

```

The intake level of the following nutrients is above the TUIL:

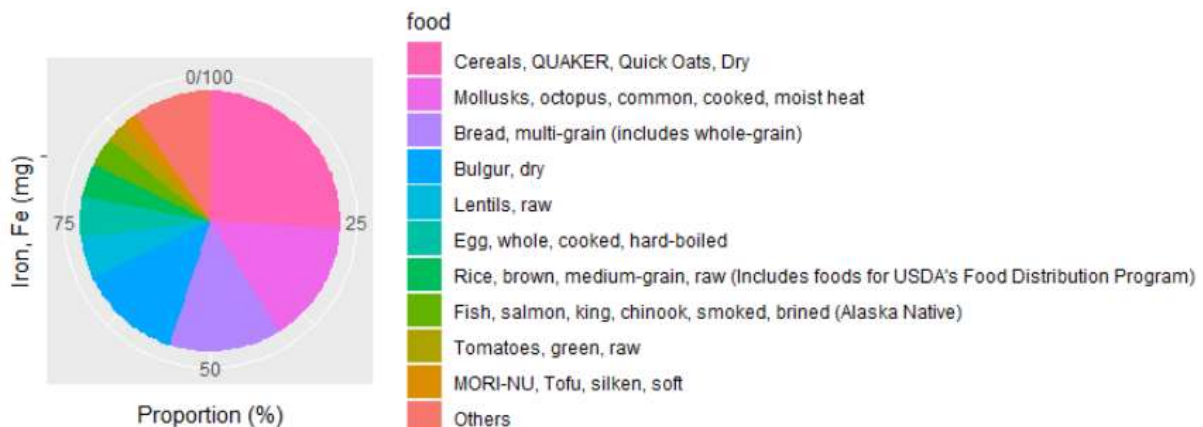
```

	nutrient	proportion(%RDA)	group
1	Magnesium, Mg (mg)	223.7326	mineral
2	Manganese, Mn (mg)	643.8333	mineral

```

> #The function nutrientPiePlot() generates a pie-chart showing the relative contribution of each food to the total intake of a given nutrient.
> nutrientPiePlot(daily_intake, nutrient_name = "Iron, Fe (mg)")
> |

```



Conclusion

Thus, after knowing the nutritious food items from the analysis and having them in our diet and avoiding the food that are harmful to our health, we can now maintain a good health and keep ourselves fit and energetic.

References:-

https://rpubs.com/ondiekhann/understanding_food_using_data

<https://cran.r-project.org/web/packages/NutrienTrackeR/vignettes/NutrienTrackeR.html>

<https://rpubs.com/JanpuHou/391136>