Module 1 - R

Rudy Fasano

3/8/2021

Chapter 3 - # 21-26

```r
library(ggplot2)
library(knitr)
library(dplyr)

setwd("D:/School/502/Week1/datasets/Website Data Sets/")
df=read.csv("nutrition_subset.csv")
```

Q21 Sorting saturated fat and commenting on validity

```r
### Q21 sort saturated_fat in descending order

sort_satfat<-df[order(-df$saturated_fat),]
```

Q22 Create new variable, saturated_fat_per_gram then merge with dataset

```r
df1 <- df
df1$saturated_fat_per_gram<-df1$saturated_fat/df1$weight_in_grams
```

Q22 - Part a - sort dataset in descending order then visualize the five highest values sorted in descending order of variable, saturated_fat_per_gram.

```r
df1<-sort_satfat_g<-df1[order(-df1$saturated_fat_per_gram),]
head(df1,n=5)
```

	food.item <fct>	▶
909	BUTTER; SALTED 1 TBSP	
910	BUTTER; UNSALTED 1 TBSP	
710	BUTTER; SALTED 1/2 CUP	
711	BUTTER; UNSALTED 1/2 CUP	
913	BUTTER; SALTED 1 PAT	
5 rows	1-2 of 6 columns		

Q22 - Part b - the foods with the most saturated fat per gram are both unsalted and salted butter (0.507).

Q23 Derive a new variable, cholesterol_per_gram

```
df1$cholesterol_fat_per_gram<-df1$cholesterol/df1$weight_in_grams
```

Q23 - part a - sort in descending order to find the 5 foods with highest values in cholesteral per gram.

```
df1<-chol_satfat_g<-df1[order(-df1$cholesterol_fat_per_gram),]
head(df1,n=5)
```

	food.item <fct>	▶
120	EGGS; RAW; YOLK 1 YOLK	
59	CHICKEN LIVER; COOKED 1 LIVER	
46	BEEF LIVER; FRIED 3 OZ	
168	EGGS; COOKED; FRIED 1 EGG	
185	EGGS; RAW; WHOLE 1 EGG	

5 rows | 1-2 of 7 columns

Q23 - part b - 1 raw egg yolk has the highest cholesterol per gram (12.53).

Q24 Standardize and produce outliers above/below the thresholds for the variable saturated_fat_per_gram

```
df1$saturated_fat_per_gramz<-scale(x=df1$saturated_fat_per_gram)
```

```
satfat_pergram_outliers<-df1[ which(df1$saturated_fat_per_gramz < -3 | df1$saturated_fat_per_gra
mz > 3),]
satfat_pergram_outliers
```

	food.item <fct>	▶
909	BUTTER; SALTED 1 TBSP	
910	BUTTER; UNSALTED 1 TBSP	
913	BUTTER; SALTED 1 PAT	
914	BUTTER; UNSALTED 1 PAT	
710	BUTTER; SALTED 1/2 CUP	
711	BUTTER; UNSALTED 1/2 CUP	

	food.item <fct>	
891	LARD 1 CUP	
900	LARD 1 TBSP	
921	IMITATION CREAMERS; POWDERED 1 TSP	
211	CHOCOLATE; BITTER OT BAKING 1 OZ	

1-10 of 15 rows | 1-2 of 8 columns   Previous   **1**   2   Next

Q24 cont. There are zero outliers below the lower limit threshold, however, there are 15 outliers present above the upper limit threshhold.

Q25 Standardize and produce outliers above the threshold for the variable cholesterol_per_gram

```
df1$cholesterol_fat_per_gramz<-scale(x=df1$cholesterol_fat_per_gram)
```

```
chol_pergram_outliers<-df1[ which(df1$cholesterol_fat_per_gramz < -3 | df1$cholesterol_fat_per_g
ramz > 3),]
chol_pergram_outliers
```

	food.item <fct>	
120	EGGS; RAW; YOLK 1 YOLK	
59	CHICKEN LIVER; COOKED 1 LIVER	
46	BEEF LIVER; FRIED 3 OZ	
168	EGGS; COOKED; FRIED 1 EGG	
185	EGGS; RAW; WHOLE 1 EGG	
187	EGGS; COOKED; HARD-COOKED 1 EGG	
186	EGGS; COOKED; POACHED 1 EGG	
190	EGGS; COOKED; SCRAMBLED/OMELET1 EGG	

8 rows | 1-2 of 9 columns

Q25 cont. There are a total of 8 outliers above the upper limit threshold for cholesterol_per_gram.

Q26 Record Indexing

```
n <- dim(df1)[1]
df1$Index <- c(1:n)
head(df1)
```

	food.item <fct>	▶
120	EGGS; RAW; YOLK 1 YOLK	
59	CHICKEN LIVER; COOKED 1 LIVER	
46	BEEF LIVER; FRIED 3 OZ	
168	EGGS; COOKED; FRIED 1 EGG	
185	EGGS; RAW; WHOLE 1 EGG	
187	EGGS; COOKED; HARD-COOKED 1 EGG	

6 rows | 1-2 of 10 columns