Latest_Analysis

Alyssa Monda

8/14/2017

Initial Set up

Load packages and set the seed for reproducibility

```
#Downloaded packages
library(phyloseq)
library(dada2)
library(ggplot2)
library(biom)
library(devtools)
library(knitr)
library(mixOmics)
library(tidyverse)
library(vegan)
library(mixOmics)
library(readxl)
library(tableone)
library(ReporteRs)
library(magrittr)
library(plyr)
library(RColorBrewer)
library(DESeq2)
library(edgeR)
library(dplyr); packageVersion("dplyr")
```

```
## [1] '0.5.0'
```

```
#If error in phyloseq load visit this page <https://joey711.github.io/phyloseq/install.html>

set.seed(08152017)
```

Import Meta Data

Import Meta Data as DataFrame

Inorder to input using this code, ensure the file path matches the path to the file on your computer. The file listed is found in the “Latest” folder of the AG data within the sub-folder containing the meta data. Download it to your computer and rewrite the code.

```
library(readr)
Meta <- read_delim("~/Desktop/ag-cleaned.txt", #file path name
    "\t", escape_double = FALSE, col_types = cols(AGE_YEARS = col_number(),
        BMI = col_number(), BMI_CORRECTED = col_number(), "#SampleID" = col_character(), `ENA-BASE-COUNT` =col_nu
mber(),HEIGHT_CM = col_number(), VIOSCREEN_FIBER = col_number(),
        VIOSCREEN_LOW_FAT_DAIRY_SERVING = col_number(),
        VIOSCREEN_LYSINE = col_number(),
        VIOSCREEN_MANNITOL = col_number(),
        VIOSCREEN_SFA100 = col_double(), #capric acid
        VIOSCREEN_SFA40 = col_double(),#SFA4:0 Butyric Acid
        VIOSCREEN_SFA170 = col_double(), #margaric acid
        VIOSCREEN_VEG5_DAY = col_number(),
        VIOSCREEN_VEGETABLE_SERVINGS = col_number(),
        chao1_1k = col_double()), trim_ws = TRUE)
```

Data Exploration for demographic information

Here I have summarized only a few of the variables in the dataset. In order to find summaries of other variables replace the variable name in the code.

Univariate Exploration

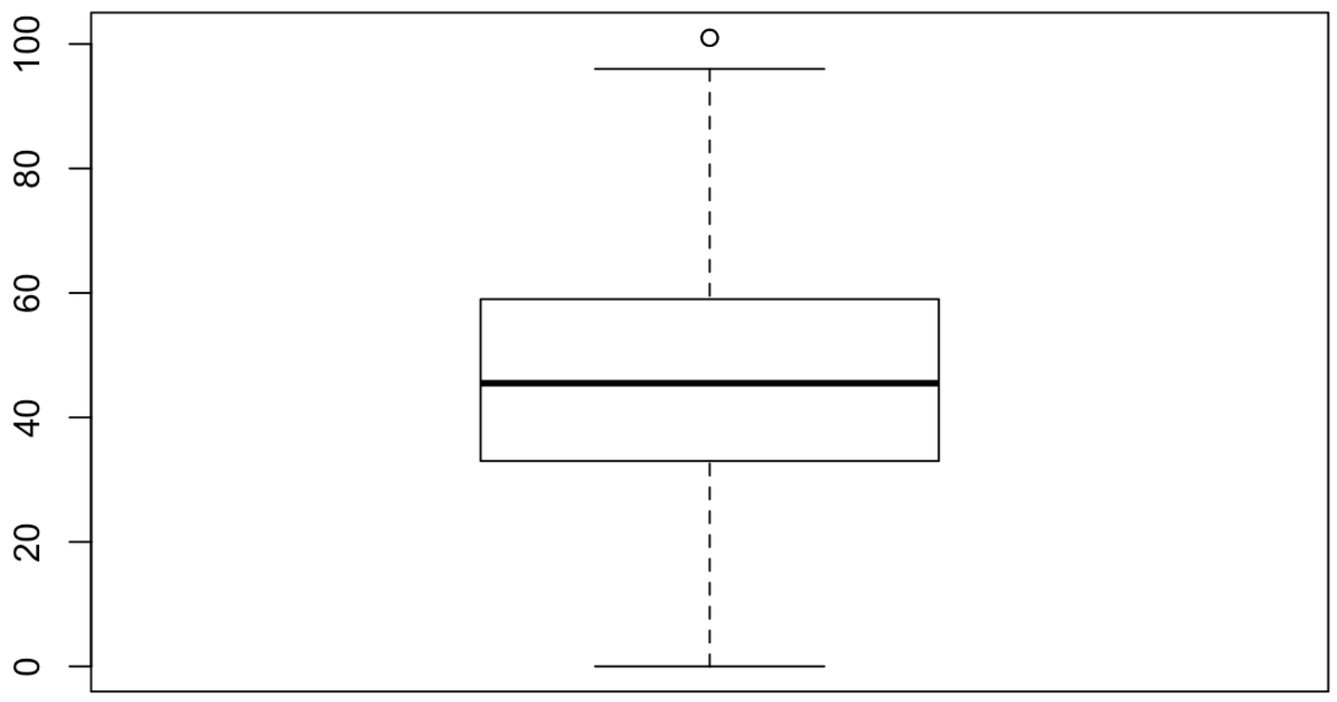
```
#Summary Statistics for continuous variables.
summary(Meta$AGE_YEARS)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   33.00   45.50   44.75   59.00  101.00     580
```

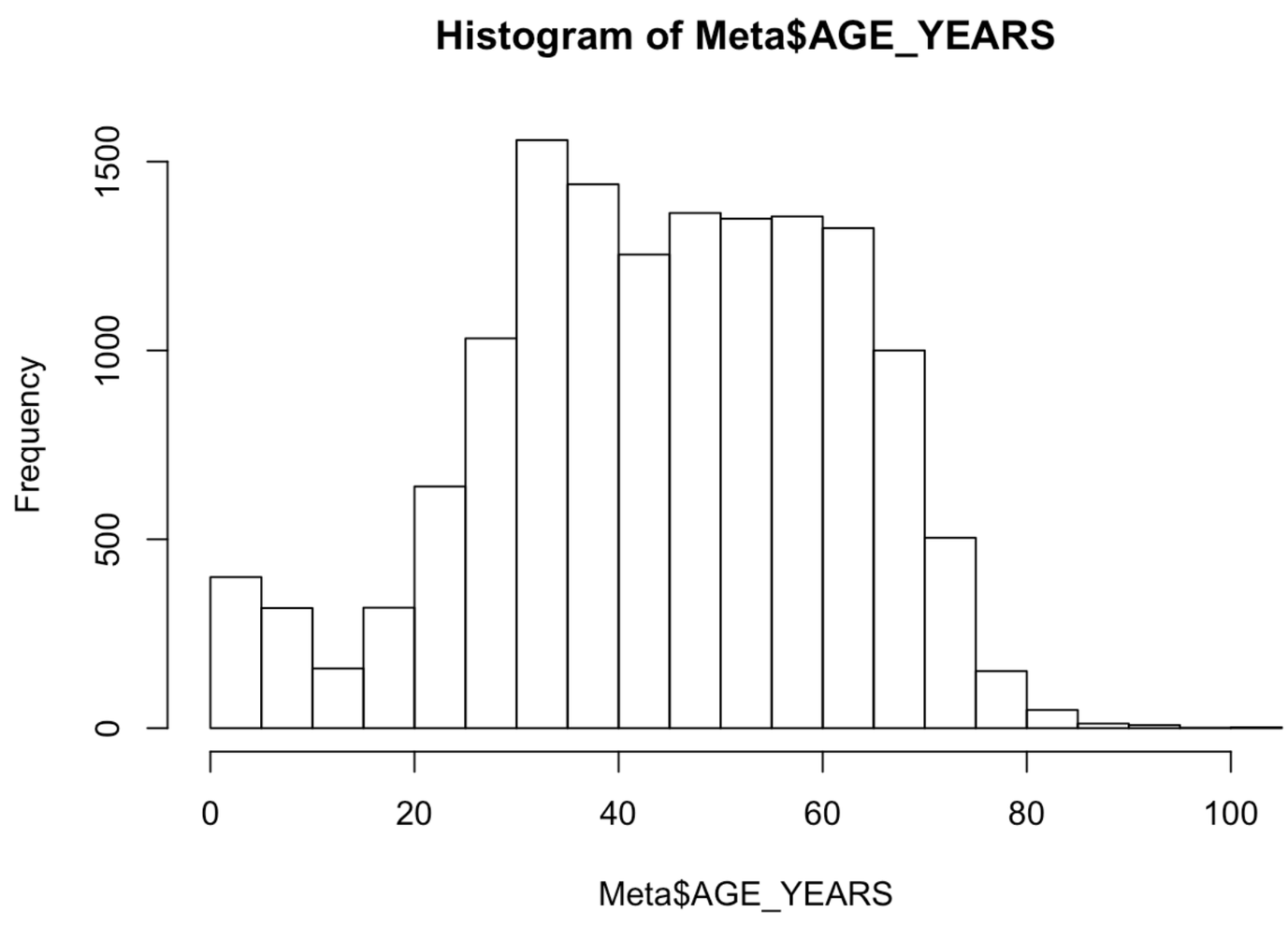
```
summary(Meta$BMI_CORRECTED) #BMI as listed in variable "BMI" listed has implausible values the corrected version has 8213 missing values and a max BMI of 76.89
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    9.63   20.81   23.24   23.90   26.32   76.89    8213
```

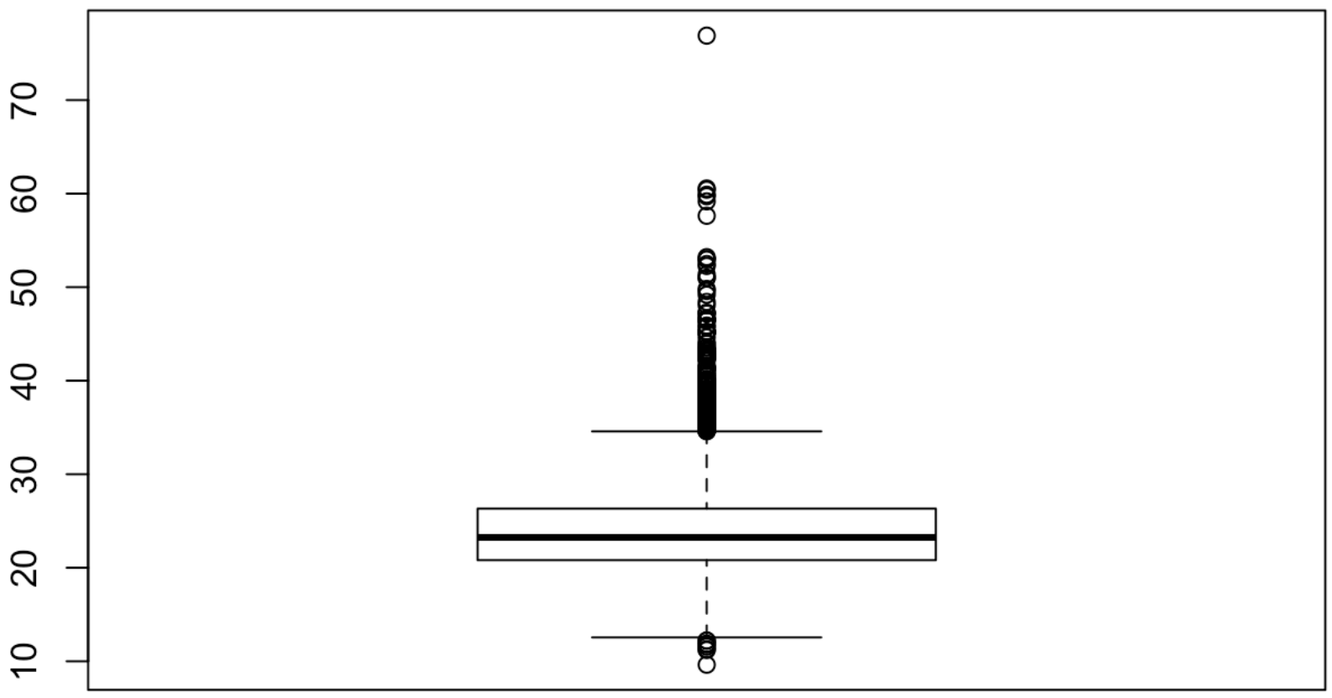
```
#Data Visualization
boxplot(Meta$AGE_YEARS)
```



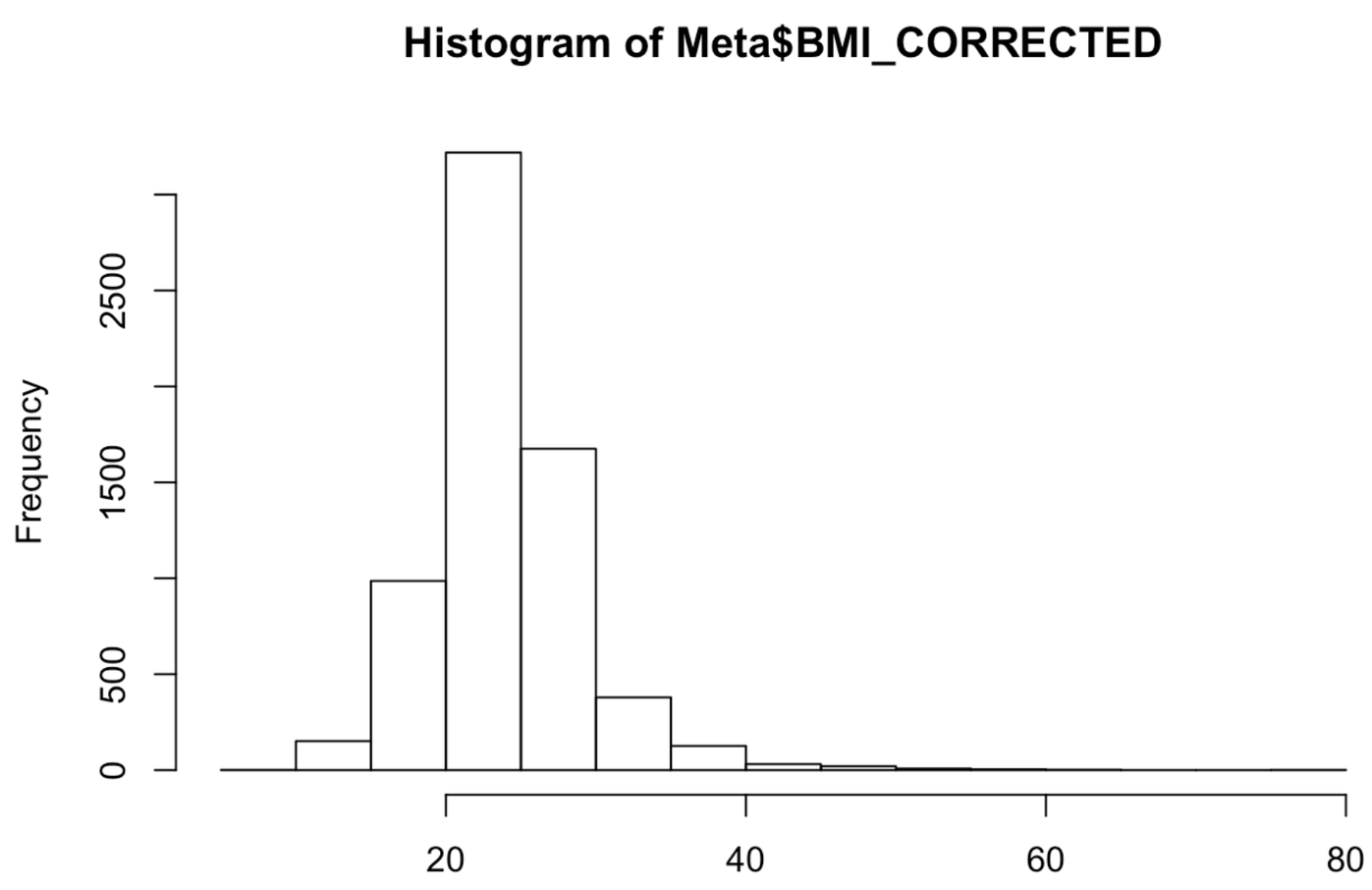

```
hist(Meta$AGE_YEARS)
```



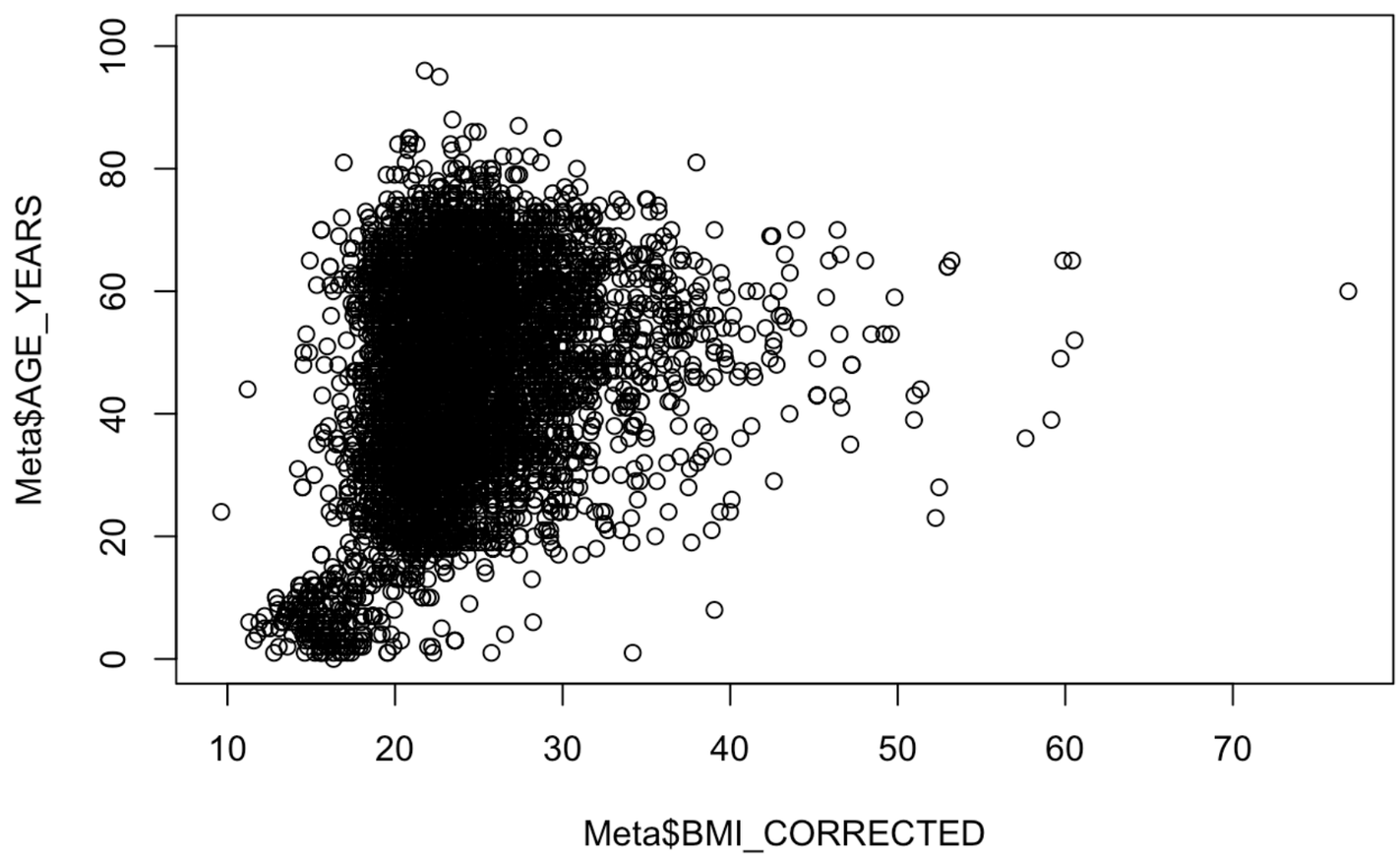

```
boxplot(Meta$BMI_CORRECTED)
```




```
hist(Meta$BMI_CORRECTED)
```




```
#Scatterplot BMI*AGE to observe for any trends
plot(Meta$BMI_CORRECTED, Meta$AGE_YEARS)
```




```
#Summary Of Categorical Variables

Categ_BodySite<- summary(as.factor(Meta$BODY_SITE))
Categ_BodySite
```

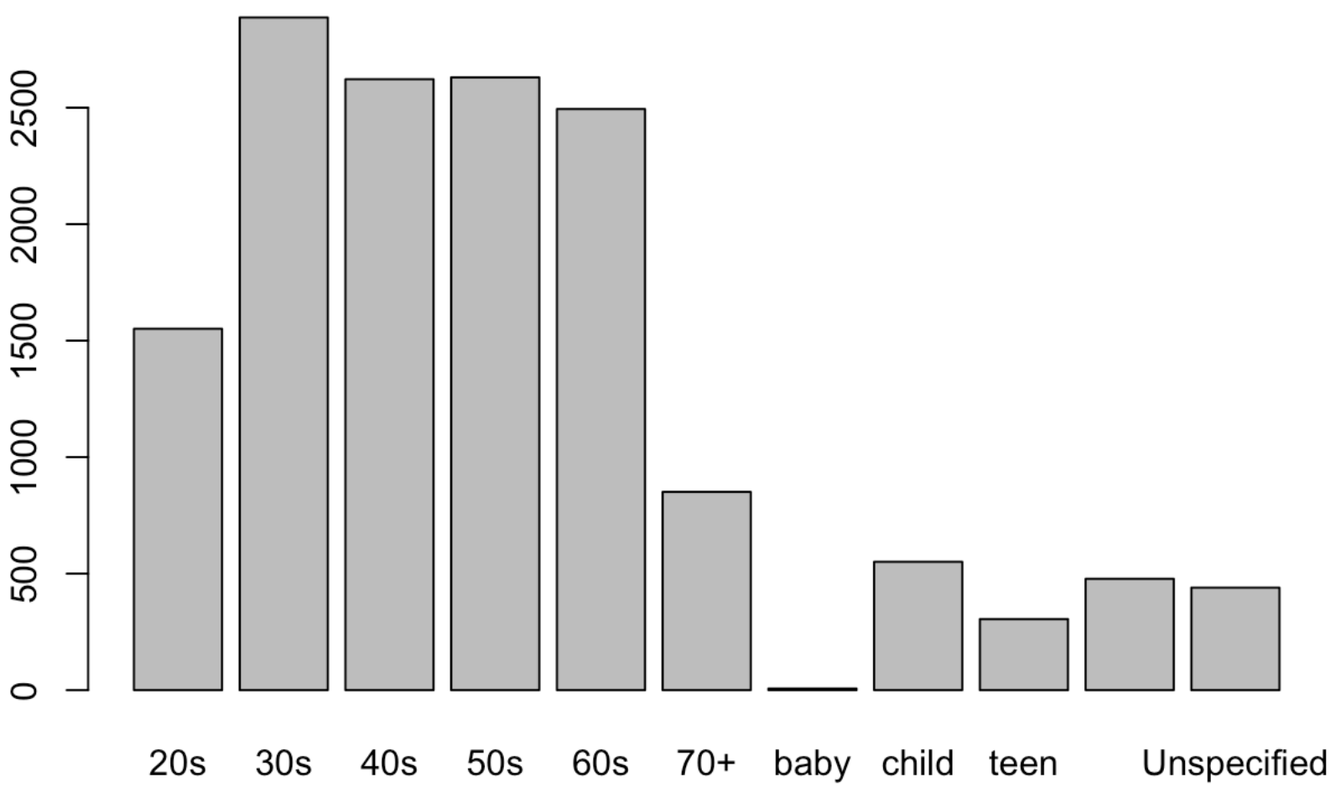
```
##        UBERON:feces      UBERON:nostril UBERON:skin of hand
##               12811                 192                 454
## UBERON:skin of head       UBERON:tongue
##                 323                1036
```

```
barplot(Categ_BodySite)
```


```
Categ_Age<- summary(as.factor(Meta$AGE_CAT))
Categ_Age
```

```
##         20s         30s         40s         50s         60s         70+
##        1551        2887        2622        2630        2494         851
##        baby       child        teen     Unknown Unspecified
##           7         551         305         478         440
```

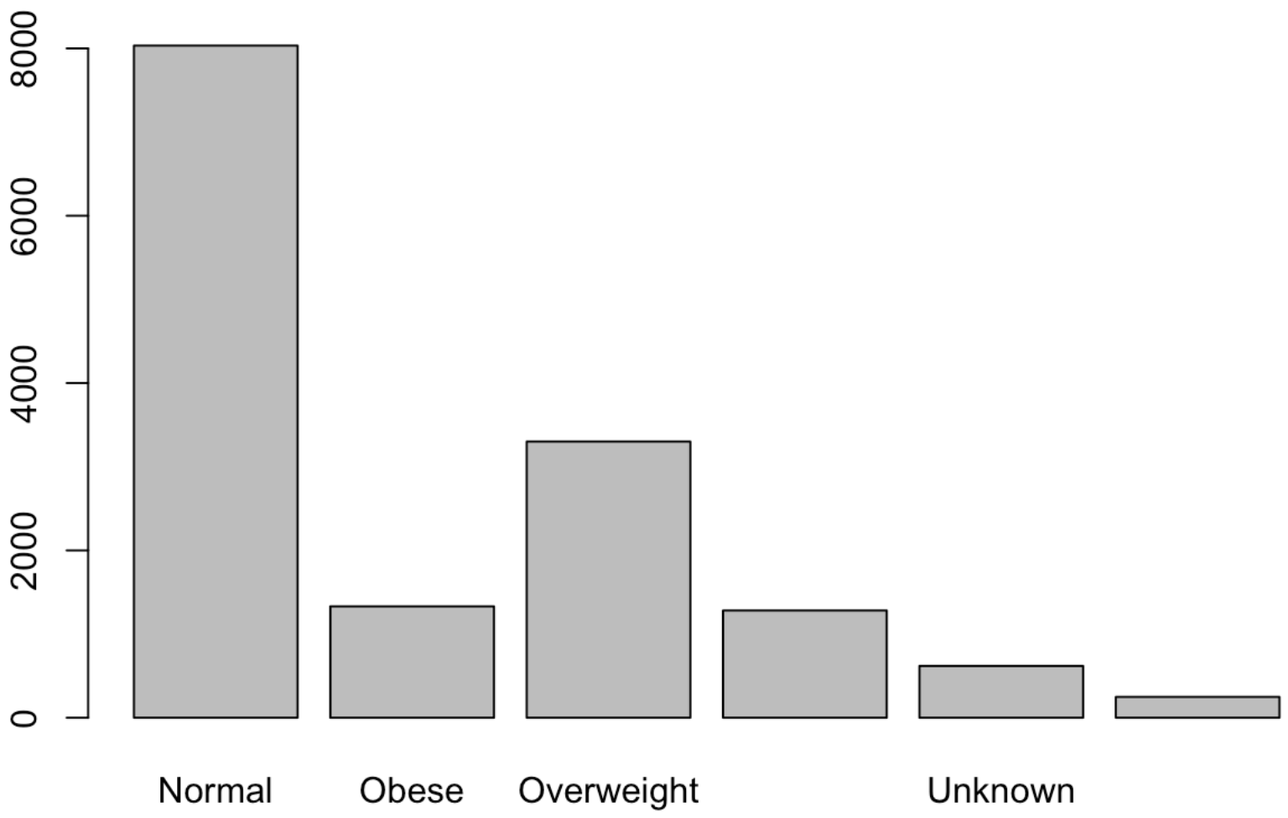
```
barplot(Categ_Age)
```




```
Categ_BMI<- summary(as.factor(Meta$BMI_CAT))
Categ_BMI
```

```
##      Normal       Obese  Overweight Underweight     Unknown Unspecified
##        8034        1331        3301        1282         619         249
```

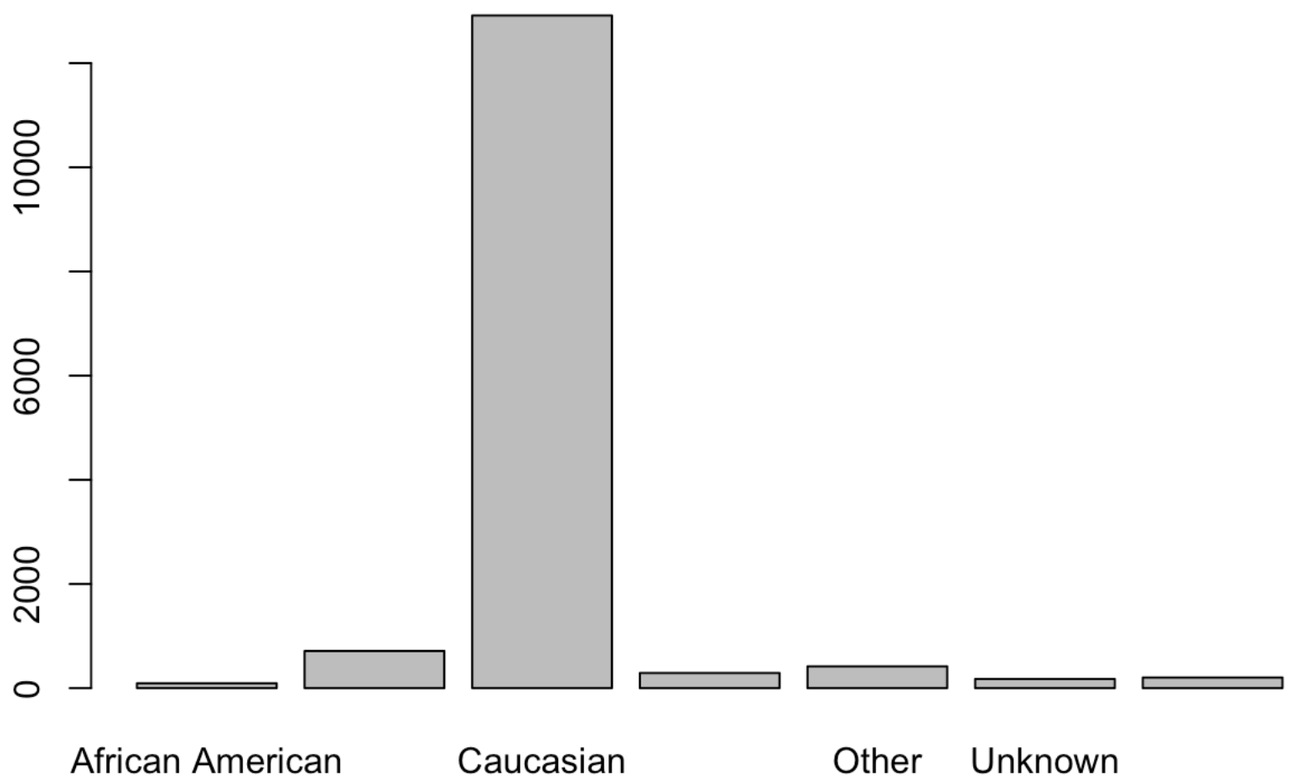
```
barplot(Categ_BMI)
```




```
Categ_Race<- summary(as.factor(Meta$RACE))
Categ_Race
```

```
##          African American Asian or Pacific Islander
##                        94                       714
##                 Caucasian                  Hispanic
##                     12916                       292
##                     Other                   Unknown
##                       419                       177
##               Unspecified
##                       204
```

```
barplot(Categ_Race)
```



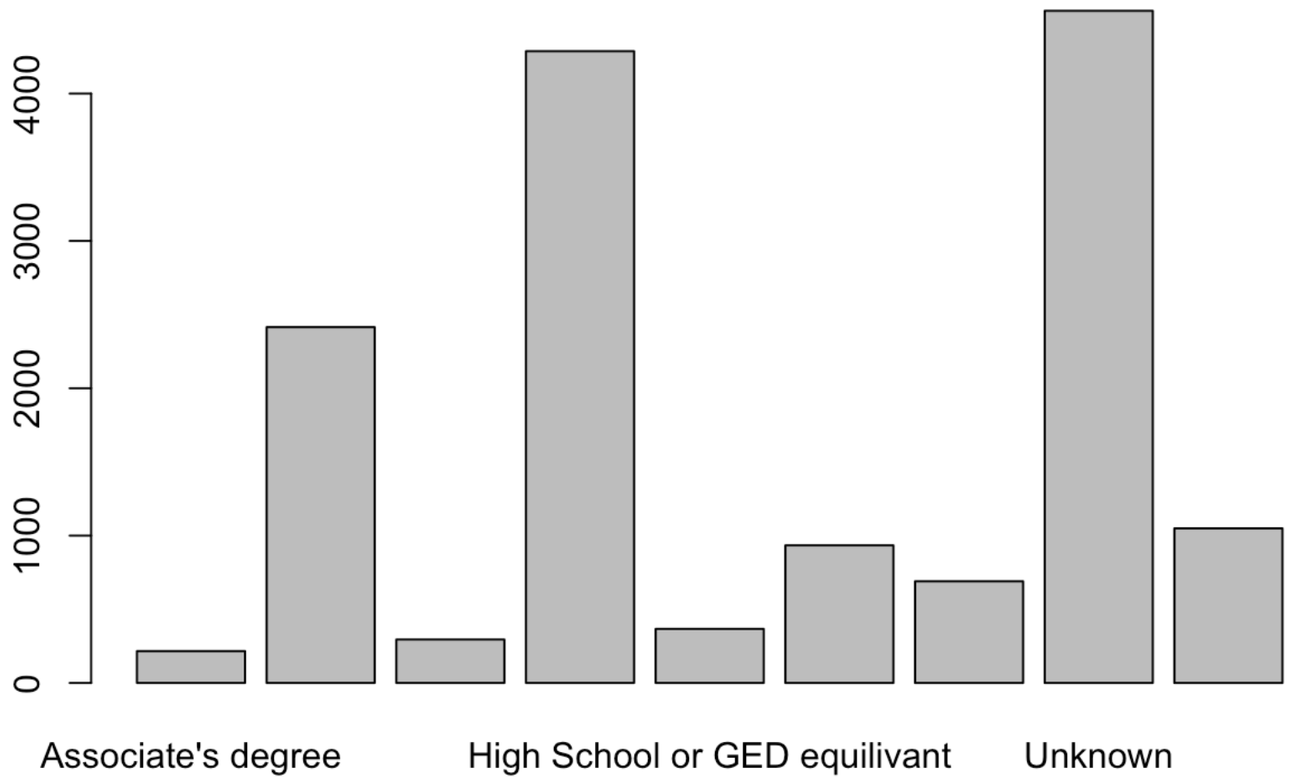

```
Categ_Sex<- summary(as.factor(Meta$SEX))
Categ_Sex
```

```
##      female        male       other     Unknown unspecified
##        7423        6639          14         517         223
```

```
Categ_Education<- summary(as.factor(Meta$LEVEL_OF_EDUCATION))
Categ_Education
```

```
##                   Associate's degree                    Bachelor's degree
##                                  216                                 2415
##         Did not complete high school     Graduate or Professional degree
##                                  295                                 4288
##        High School or GED equilivant    Some college or technical school
##                                  367                                  934
## Some graduate school or professional                              Unknown
##                                  690                                 4562
##                           Unspecified
##                                 1049
```

```
barplot(Categ_Education)
```



```
Categ_Region<- summary(as.factor(Meta$COUNTRY_RESIDENCE))
Categ_Region
```

```
##                     Australia                       Austria
##                            94                             9
##                       Belgium                        Belize
##                            15                             1
##                        Brazil                        Canada
##                             2                            61
##                         China                       Croatia
##                             2                             2
##                Czech Republic                       Denmark
##                             4                             5
##                       Finland                        France
##                             3                            28
##                       Georgia                       Germany
##                             1                            38
##                        Greece                     Hong Kong
##                             2                             7
##                       Iceland                         India
##                             1                            24
##                       Ireland                   Isle of Man
##                            52                             4
##                        Israel                         Italy
##                             2                            17
##                         Japan                        Jersey
##                             5                             2
##                        Malta                         Mexico
##                             1                             2
##                       Morocco                    Netherlands
##                            35                            15
##                   New Zealand                       no_data
##                             5                          7262
##                        Norway                      Portugal
##                             9                             4
##                   Puerto Rico                       Romania
##                             1                             1
##                  Saudi Arabia                        Serbia
##                             1                             5
##                     Singapore                      Slovakia
##                             3                             3
##                         Spain                        Sweden
##                             6                            30
##                   Switzerland Tanzania, United Republic of
##                            61                             2
##                      Thailand         United Arab Emirates
##                             2                             4
##                  United Kingdom                United States
##                          1712                          2982
##                       Unknown                   Unspecified
##                          1073                          1216
```

```
Categ_Birth_Country<- summary(as.factor(Meta$COUNTRY_OF_BIRTH))
Categ_Birth_Country
```

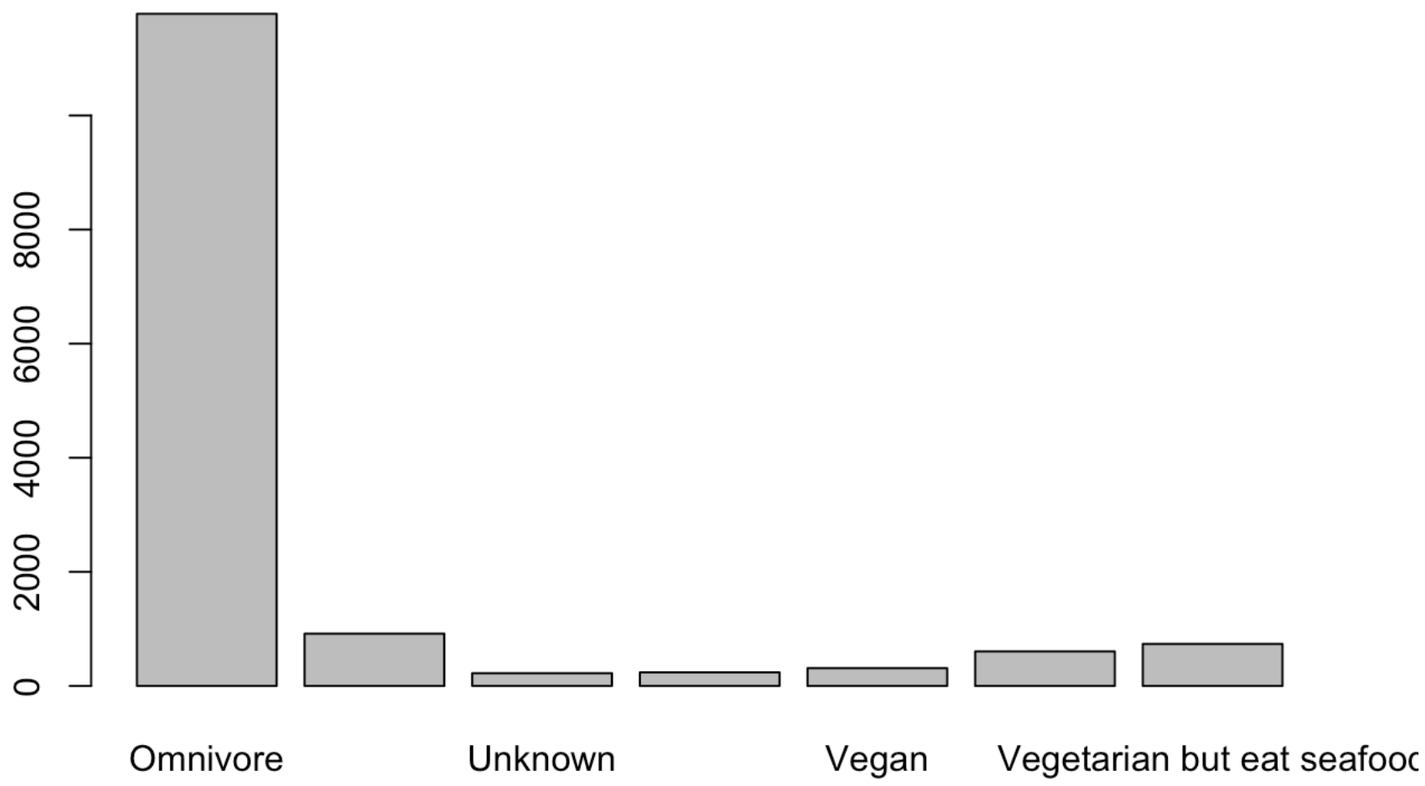
```
##                            United States                           United Kingdom
##                                     9312                                     2566
##                                Australia                                   Canada
##                                      341                                      323
##                                  Germany                              Unspecified
##                                      185                                      171
##                                    India                                  Unknown
##                                      150                                      138
##                                  Belgium                                   France
##                                      114                                      100
##                                  Ireland                                   Poland
##                                       98                                       57
##                                    Italy                              Switzerland
##                                       56                                       54
##                             South Africa                                   Sweden
##                                       52                                       51
##                 Iran, Islamic Republic of                              New Zealand
##                                       49                                       47
##                                    China                                   Brazil
##                                       46                                       45
##                              Netherlands                                    Spain
##                                       44                                       44
##                                Hong Kong                       Russian Federation
```

```
##                                39                                38
##                             Japan                            Mexico
##                                32                                31
##                       Afghanistan                  Korea, Republic of
##                                26                                26
##                           Denmark                           Hungary
##                                22                                22
##                            Norway                           Austria
##                                22                                21
##                       Philippines                            Israel
##                                21                                20
##                    Czech Republic                          Bulgaria
##                                19                                18
##         Taiwan, Province of China                          Malaysia
##                                17                                16
##                            Greece                       Puerto Rico
##                                15                                15
##                           Romania                          Thailand
##                                14                                14
##                           Finland                          Portugal
##                                12                                12
##                           Morocco                          Pakistan
##                                11                                11
##                         Singapore                           Ukraine
##                                11                                11
##                           Croatia                         Indonesia
##                                10                                10
##                          Colombia                            Cyprus
##                                10                                10
##                          Colombia                            Cyprus
##                                 9                                 9
##                          Zimbabwe                            Serbia
##                                 9                                 8
##                         Sri Lanka                          Viet Nam
##                                 8                                 8
##                         Argentina                           Lebanon
##                                 7                                 7
##      Tanzania, United Republic of                            Turkey
##                                 7                                 7
##                            Zambia                           Belarus
##                                 7                                 6
##                              Cuba                           Nigeria
##                                 6                                 6
##                          Slovakia                           Jamaica
##                                 6                                 5
##                         Lithuania                              Peru
##                                 5                                 5
##                         Venezuela                        Bangladesh
##                                 5                                 4
##                             Chile                Dominican Republic
##                                 4                                 4
##                             Egypt                       El Salvador
##                                 4                                 4
##                          Ethiopia                            Kuwait
##                                 4                                 4
##                            Latvia                             Malta
##                                 4                                 4
##                      Saudi Arabia United States Minor Outlying Islands
##                                 4                                 4
##                 Bosnia and Herzegovina                         Guatemala
##                                 3                                 3
##                           Iceland                             Kenya
##                                 3                                 3
##                          Paraguay                            Uganda
##                                 3                                 3
##                    United Arab Emirates                           Albania
##                                 3                                 2
##                           Algeria                          Azerbaijan
##                                 2                                 2
##                           Bermuda                        Costa Rica
##                                 2                                 2
##                           Ecuador                Libyan Arab Jamahiriya
##                                 2                                 2
##                        Luxembourg                            Malawi
##                                 2                                 2
##                         Mauritius                            Panama
##                                 2                                 2
##                   Syrian Arab Republic                          (Other)
##                                 2                                18
```

```
Categ_Diet<- summary(as.factor(Meta$DIET_TYPE))
Categ_Diet
```

```
##                          Omnivore Omnivore but do not eat red meat
##                             11783                             917
##                           Unknown                     Unspecified
##                               224                             238
##                             Vegan                      Vegetarian
##                               312                             605
##       Vegetarian but eat seafood
##                               737
```

```
barplot(Categ_Diet)
```



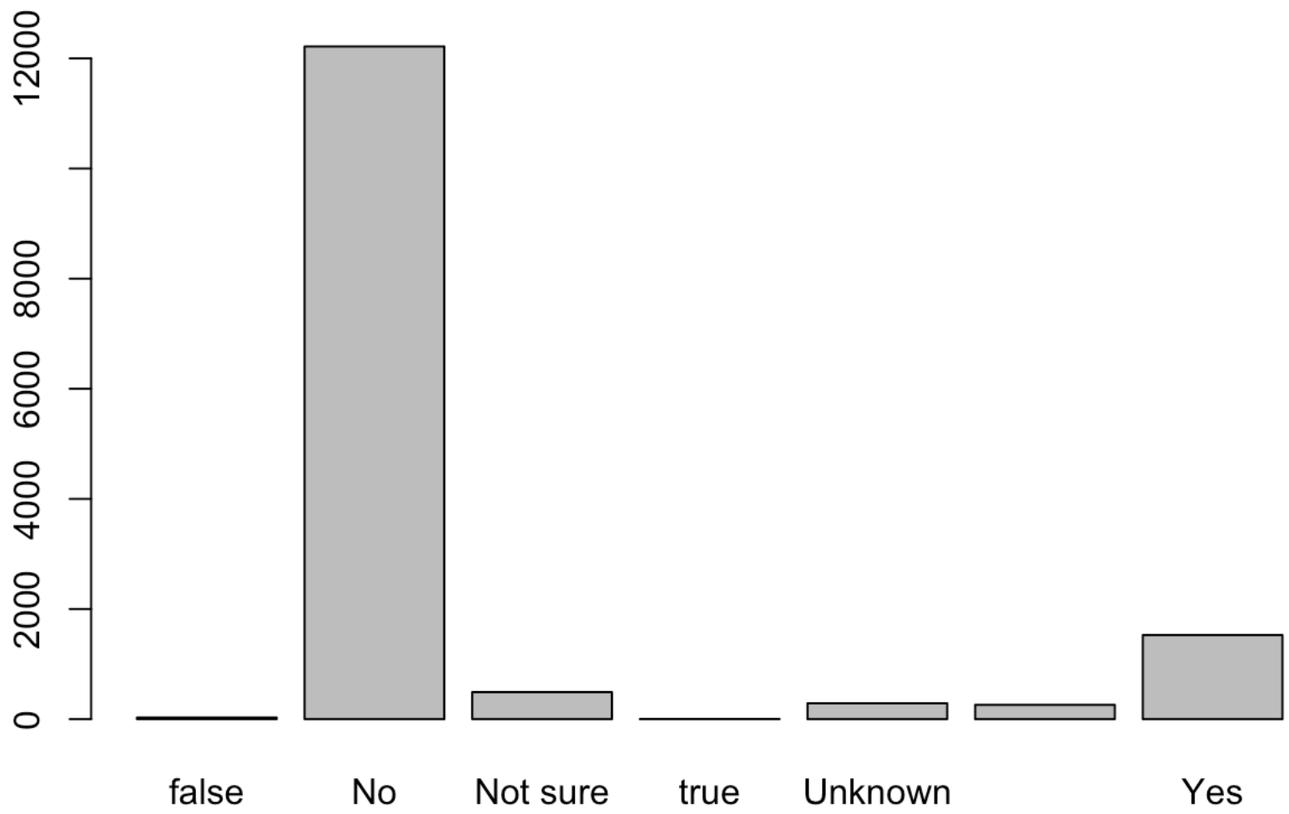

```
Categ_Pregnant<- summary(as.factor(Meta$PREGNANT))
Categ_Pregnant
```

```
##       false          No    Not sure        true     Unknown Unspecified
##         600       10066          41           1        3367         672
##         Yes
##          69
```

```
Categ_Delivery<- summary(as.factor(Meta$CSECTION))
Categ_Delivery
```

```
##       false          No    Not sure        true     Unknown Unspecified
##          30       12217         491           2         289         260
##         Yes
##        1527
```

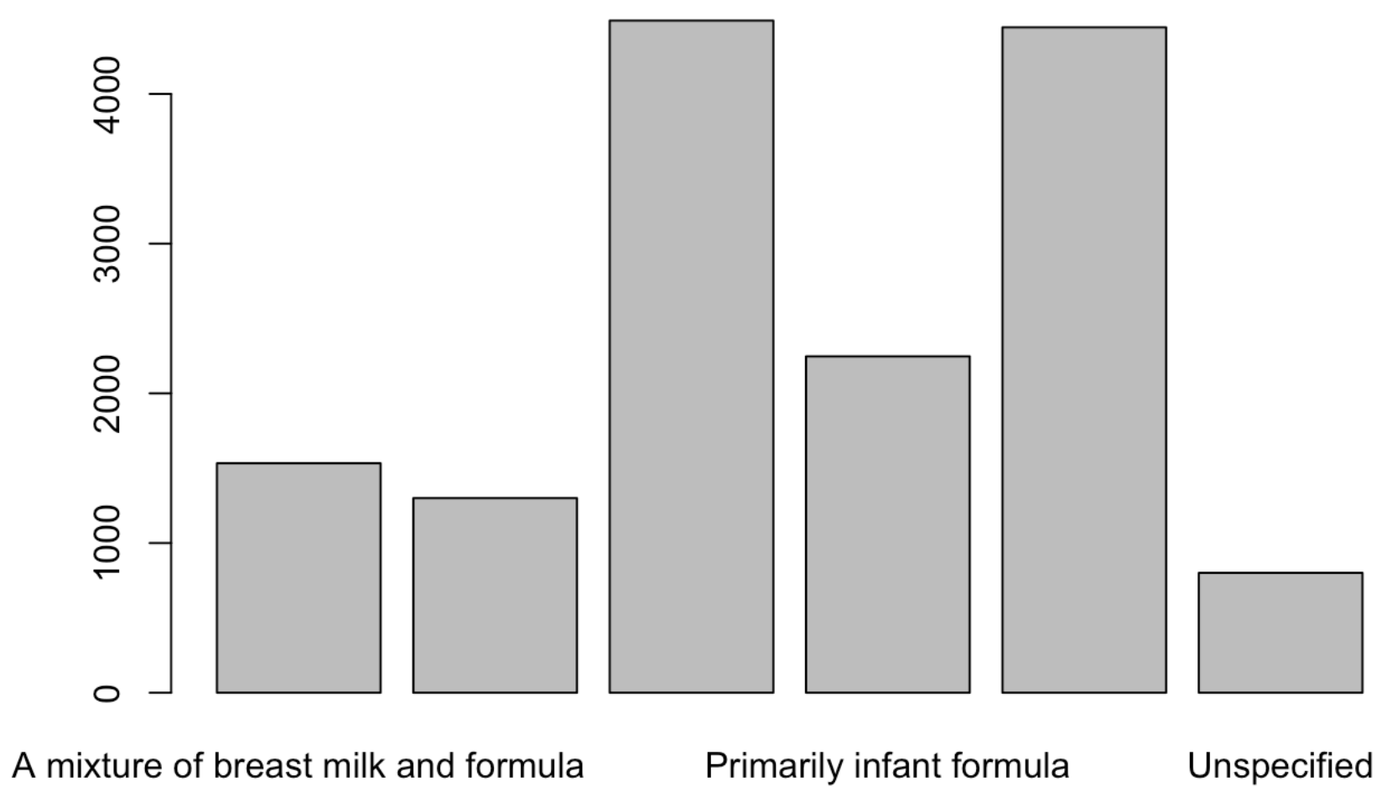
```
barplot(Categ_Delivery) #CSECTION=TRUE
```




```
Categ_Fed<- summary(as.factor(Meta$FED_AS_INFANT))
Categ_Fed
```

```
## A mixture of breast milk and formula                          Not sure
##                                 1533                              1300
##                Primarily breast milk             Primarily infant formula
##                                 4490                              2247
##                              Unknown                       Unspecified
##                                 4445                               801
```

```
barplot(Categ_Fed)
```



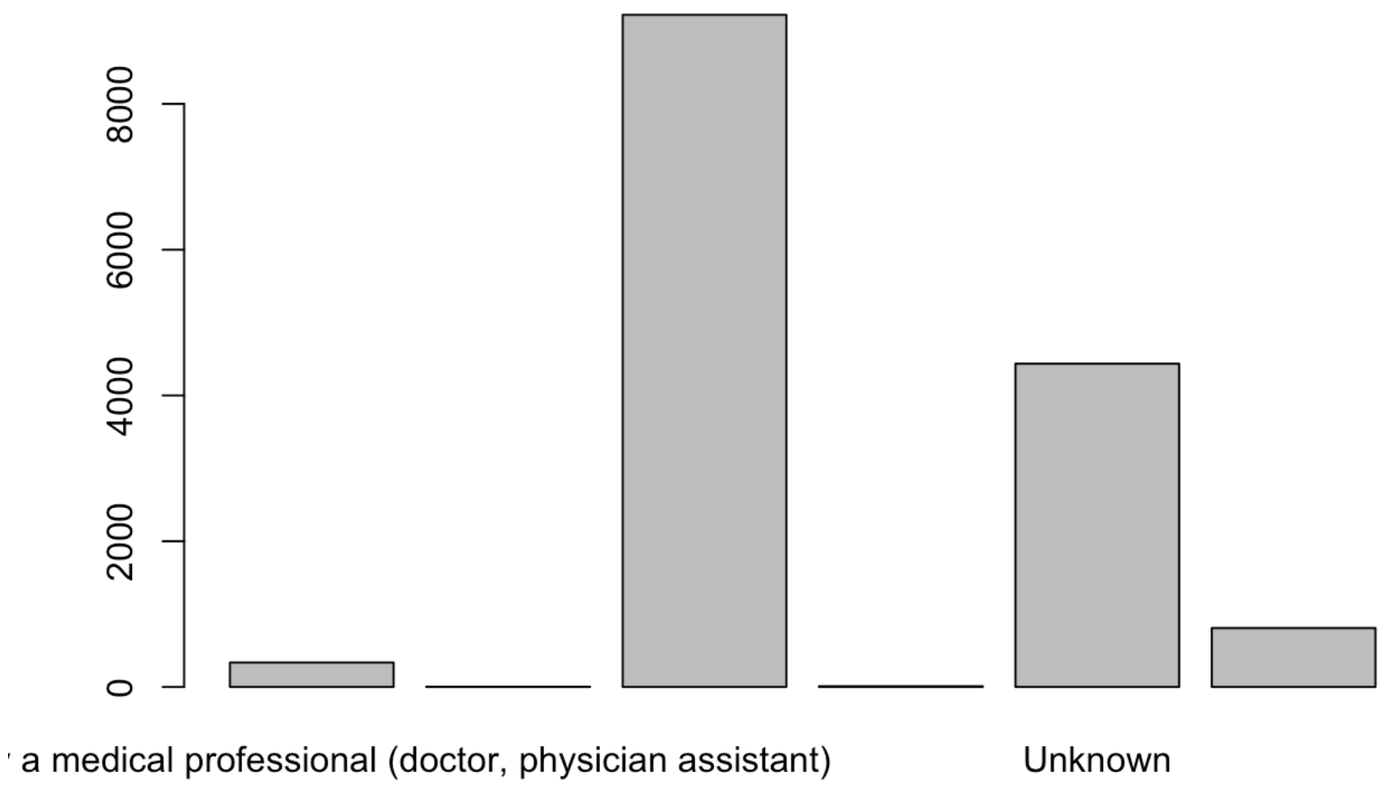

```
Categ_Healthy<- summary(as.factor(Meta$SUBSET_HEALTHY))
Categ_Healthy
```

```
## false False  true  True
##  5566  2995  3842  2413
```

```
Categ_CVD<- summary(as.factor(Meta$CARDIOVASCULAR_DISEASE))
Categ_CVD
```

```
## Diagnosed by a medical professional (doctor, physician assistant)
##                                                                335
##                  Diagnosed by an alternative medicine practitioner
##                                                                  3
##                                       I do not have this condition
##                                                               9223
##                                                       Self-diagnosed
##                                                                 10
##                                                            Unknown
##                                                               4436
##                                                        Unspecified
##                                                                809
```

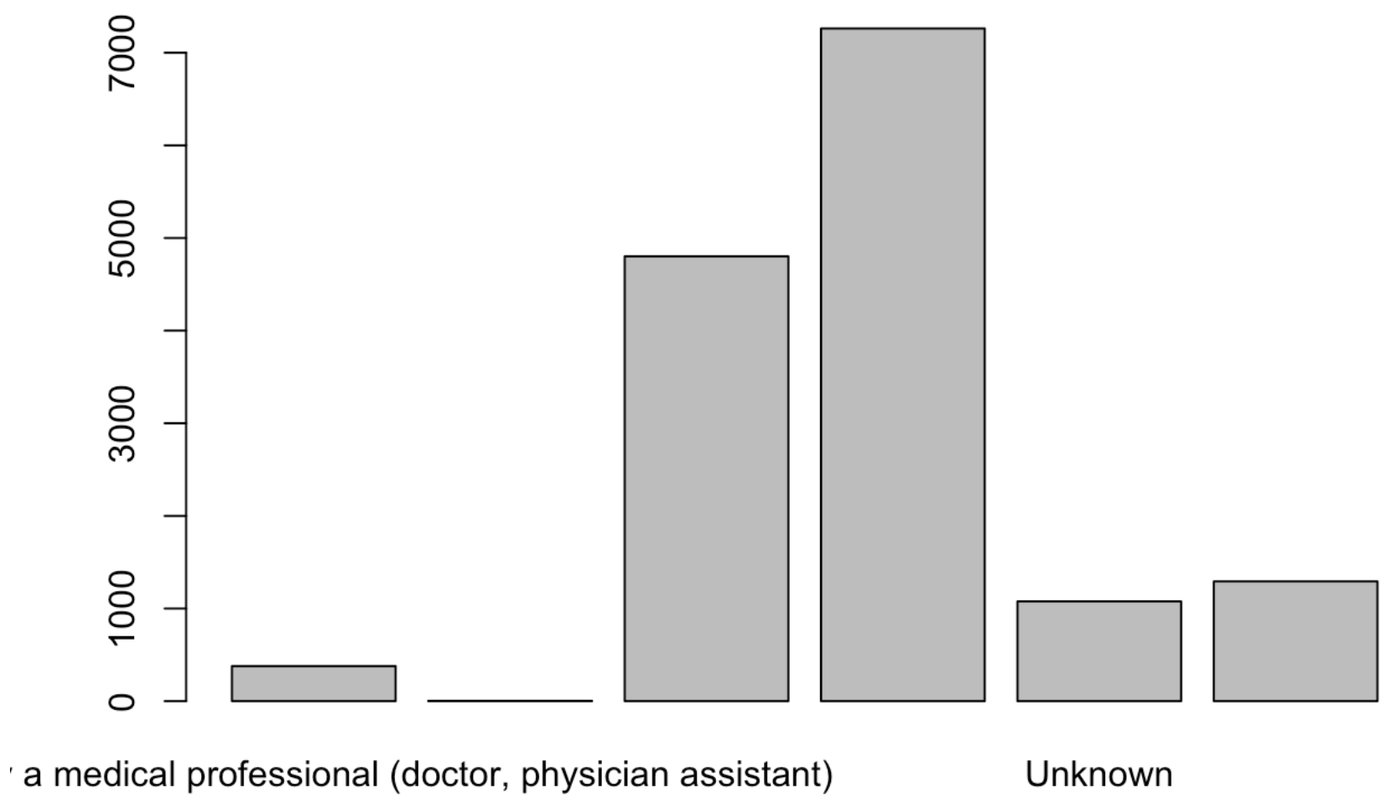
```
barplot(Categ_CVD)
```




```
Categ_Cancer<- summary(as.factor(Meta$CANCER))
Categ_Cancer
```

```
## Diagnosed by a medical professional (doctor, physician assistant)
##                                                                379
##                  Diagnosed by an alternative medicine practitioner
##                                                                  3
##                                       I do not have this condition
##                                                               4802
##                                                            no_data
##                                                               7262
##                                                            Unknown
##                                                               1077
##                                                        Unspecified
##                                                               1293
```

```
barplot(Categ_Cancer)
```




```
Categ_Diabetes<- summary(as.factor(Meta$DIABETES))
Categ_Diabetes
```

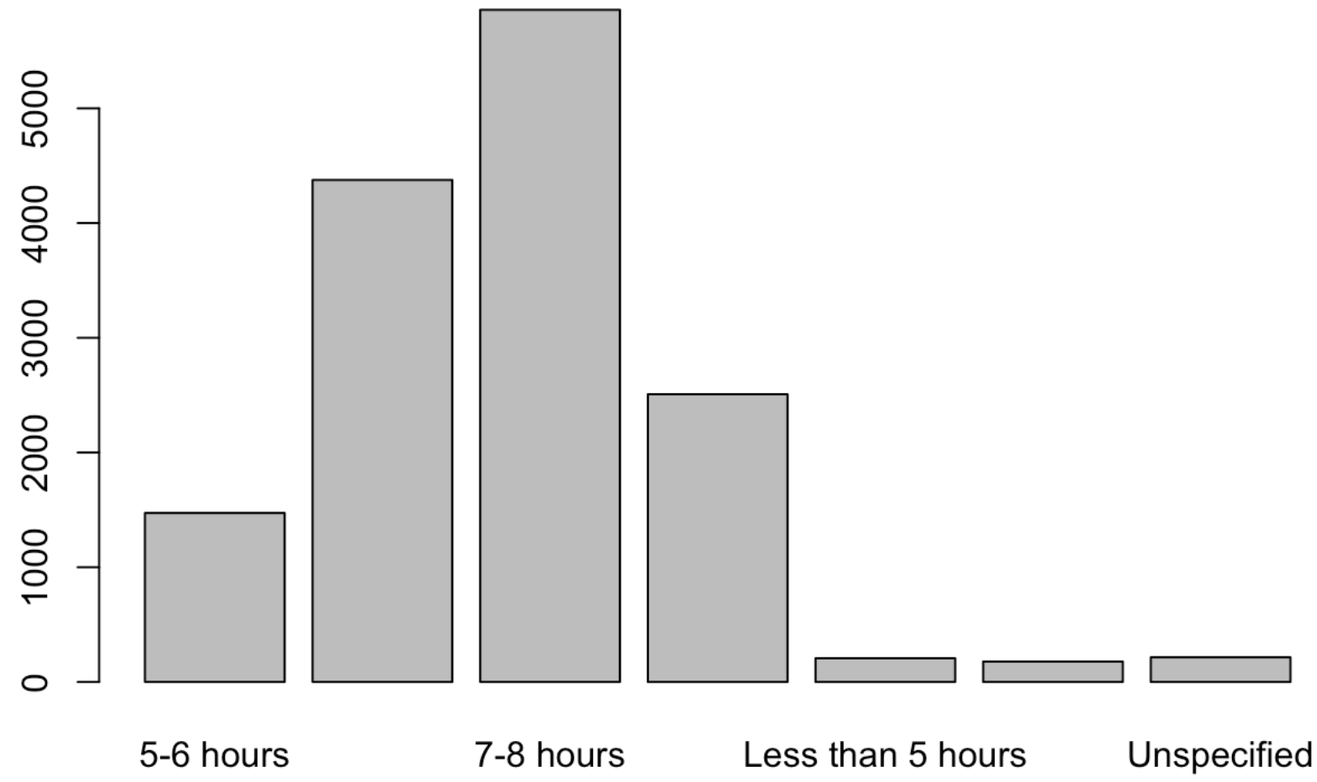
```
## Diagnosed by a medical professional (doctor, physician assistant)
##                                                               203
##                 Diagnosed by an alternative medicine practitioner
##                                                                 3
##                                      I do not have this condition
##                                                             13934
##                                                      Self-diagnosed
##                                                                12
##                                                           Unknown
##                                                               396
##                                                       Unspecified
##                                                               268
```

```
barplot(Categ_Diabetes)
```


```
Categ_Sleep<- summary(as.factor(Meta$SLEEP_DURATION))
Categ_Sleep
```

```
##          5-6 hours          6-7 hours          7-8 hours   8 or more hours
##               1473               4376               5859              2508
## Less than 5 hours            Unknown        Unspecified
##                207                178                215
```

```
barplot(Categ_Sleep)
```



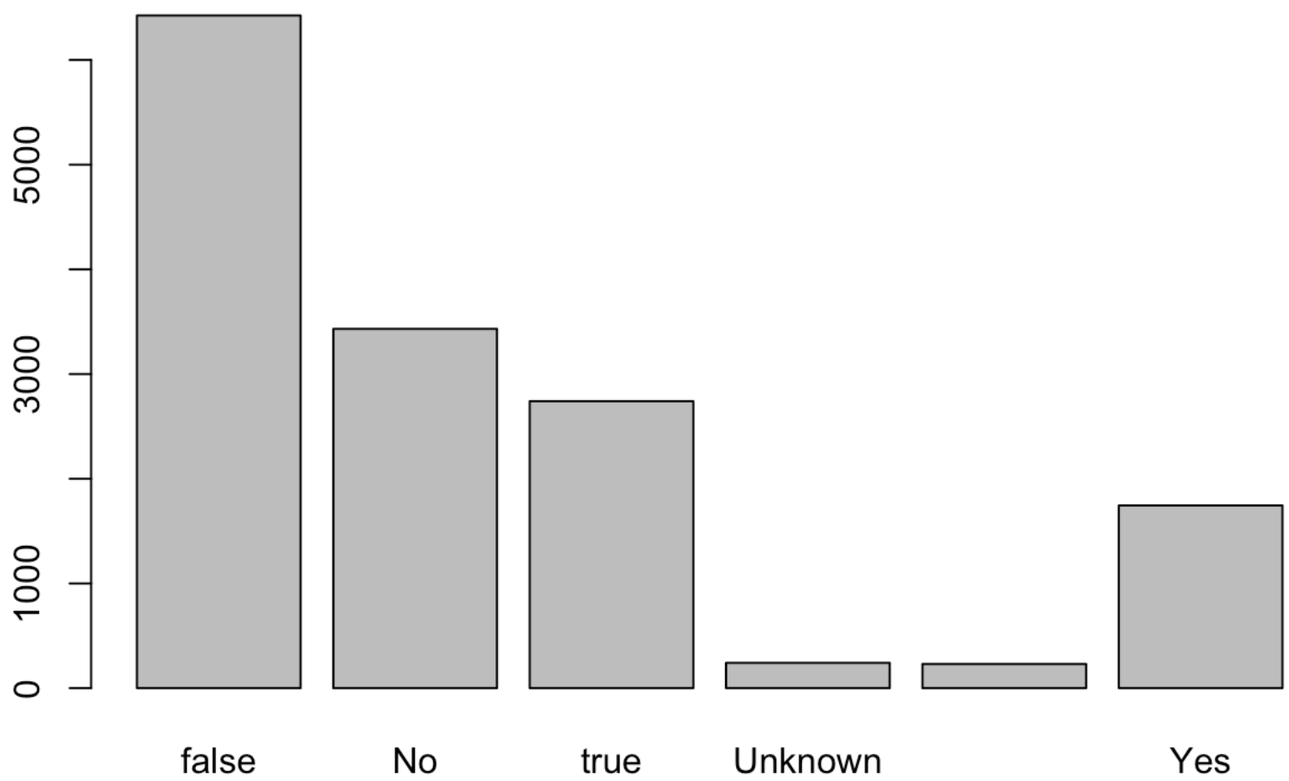

```
Categ_Pet<- summary(as.factor(Meta$PETS_OTHER))
Categ_Pet
```

```
##       false          No     no_data        true     Unknown Unspecified
##        1049        3602        7262         142        1077        1254
##         Yes
##         430
```

```
Categ_Dog<- summary(as.factor(Meta$DOG))
Categ_Dog
```

```
##       false          No        true     Unknown Unspecified         Yes
##        6425        3432        2741         242         231        1745
```

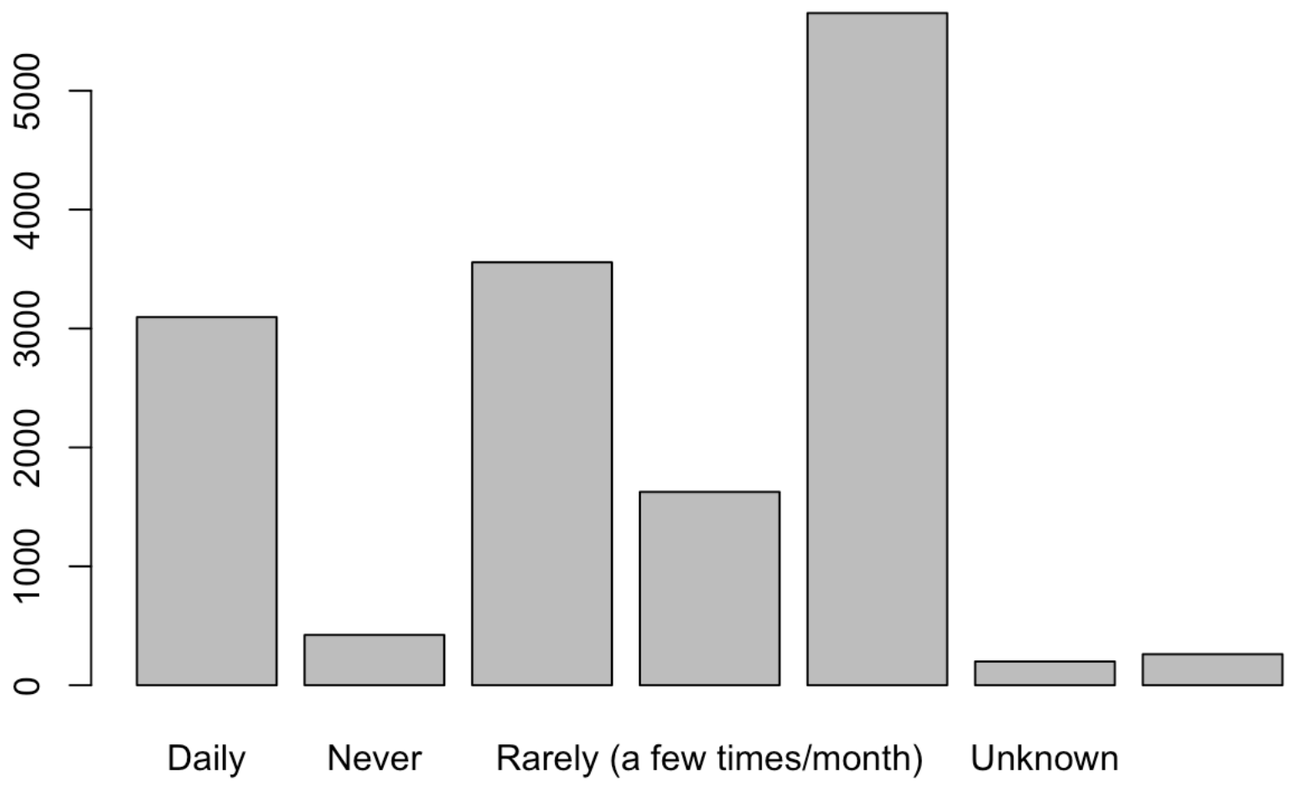
```
barplot(Categ_Dog)
```




```
Categ_Exercise<- summary(as.factor(Meta$EXERCISE_FREQUENCY))
Categ_Exercise
```

```
##                          Daily                          Never
##                           3096                            423
## Occasionally (1-2 times/week)    Rarely (a few times/month)
##                           3557                           1626
##    Regularly (3-5 times/week)                        Unknown
##                           5653                            200
##                     Unspecified
##                            261
```

```
barplot(Categ_Exercise)
```



Breakdown by Age Split

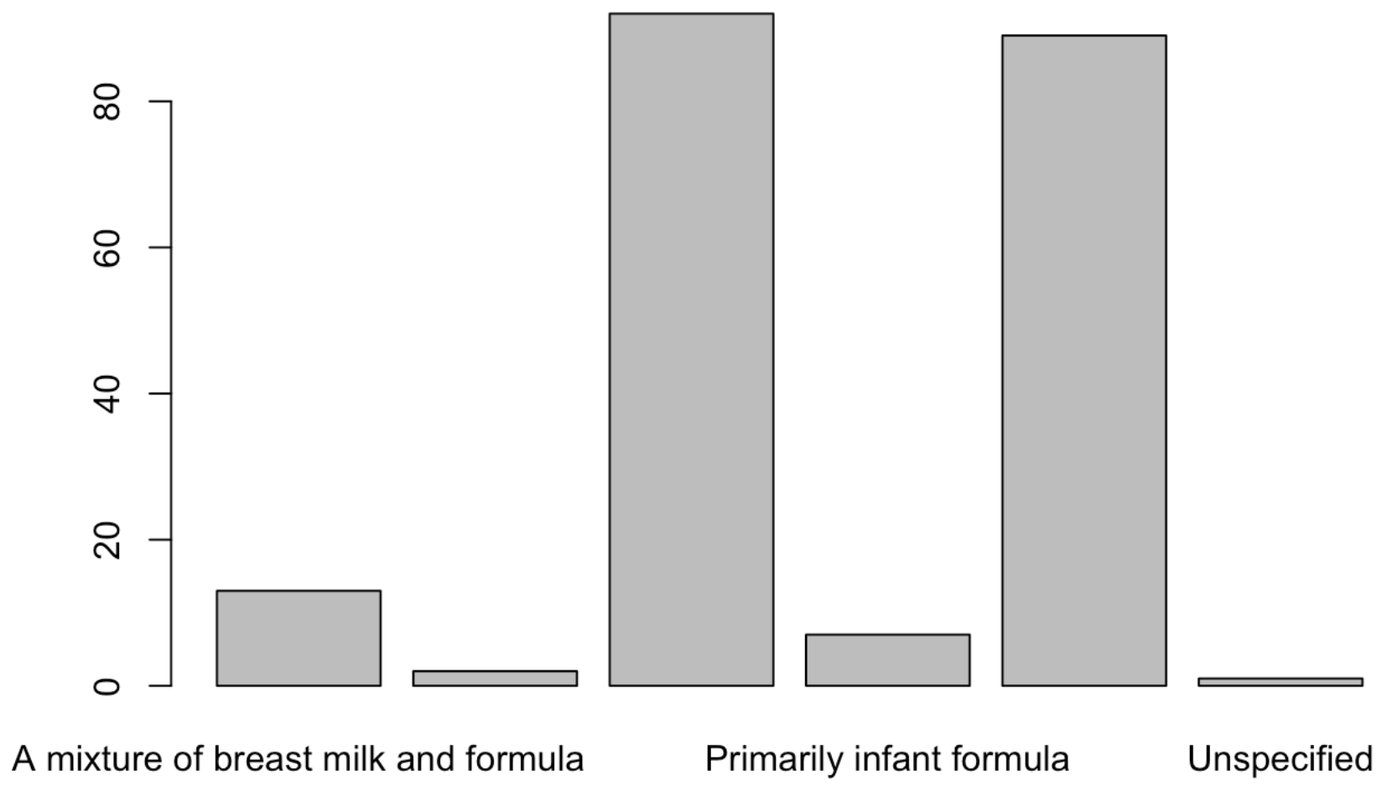
```
#Split Meta for Under age 3
Meta3 <-subset(Meta, AGE_YEARS<3)
summary(Meta3$AGE_YEARS)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   2.000   1.662   2.000   2.000
```

```
#Under 3 and what they are fed
Categ_Fed3<- summary(as.factor(Meta3$FED_AS_INFANT))
Categ_Fed3
```

```
## A mixture of breast milk and formula                          Not sure
##                                   13                                 2
##                Primarily breast milk            Primarily infant formula
##                                   92                                 7
##                              Unknown                       Unspecified
##                                   89                                 1
```

```
barplot(Categ_Fed3)
```



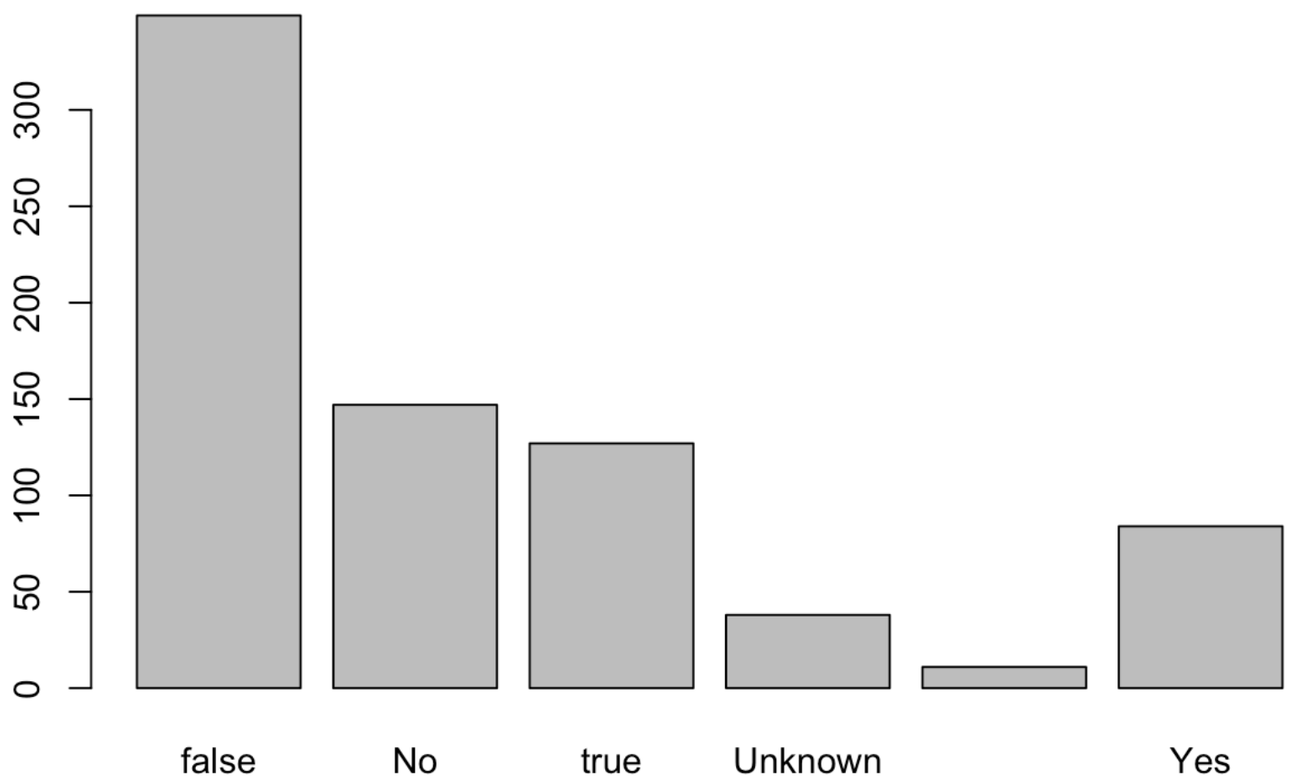

```
#Split Meta for Under age 12
Meta12 <-subset(Meta, AGE_YEARS<12)
summary(Meta12$AGE_YEARS)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   2.000   5.000   5.386   8.000  11.000
```

```
#Under 12 with Allergies
Categ_Allergies12<- summary(as.factor(Meta12$SEASONAL_ALLERGIES))
Categ_Allergies12
```

```
##       false          No        true     Unknown Unspecified         Yes
##         349         147         127          38          11          84
```

```
barplot(Categ_Allergies12)
```




```
#or
count(Meta12,'SEASONAL_ALLERGIES')
```

```
##   SEASONAL_ALLERGIES freq
## 1              false  349
## 2                 No  147
## 3               true  127
## 4            Unknown   38
## 5        Unspecified   11
## 6                Yes   84
```

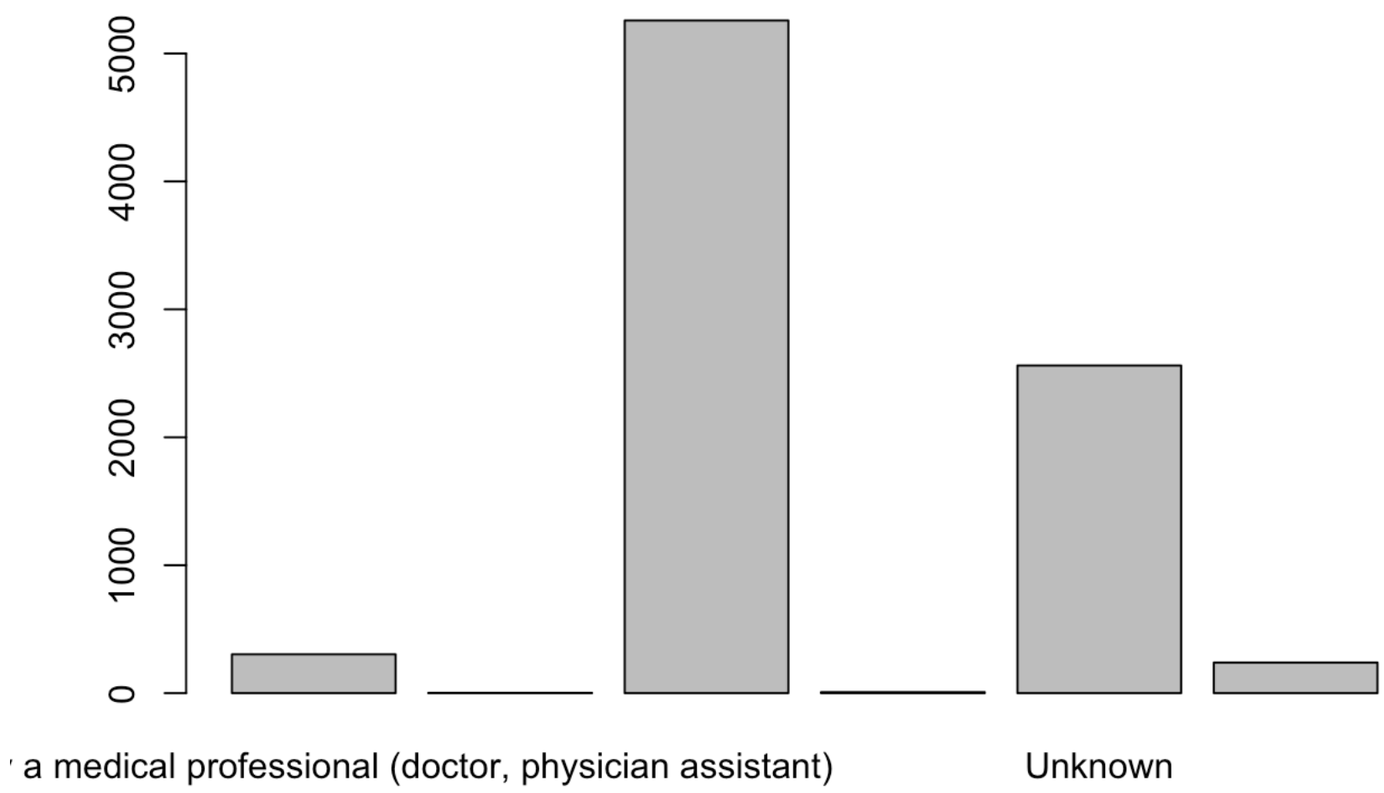
```
#Split Meta Over 40
Meta40 <-subset(Meta, AGE_YEARS>40)
summary(Meta40$AGE_YEARS)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   41.00   49.00   56.00   56.76   64.00  101.00
```

```
#Over 40 and CVD
Categ_CVD40<- summary(as.factor(Meta40$CARDIOVASCULAR_DISEASE))
Categ_CVD40
```

```
## Diagnosed by a medical professional (doctor, physician assistant)
##                                                               304
##                  Diagnosed by an alternative medicine practitioner
##                                                                 2
##                                        I do not have this condition
##                                                              5258
##                                                     Self-diagnosed
##                                                                 8
##                                                           Unknown
##                                                              2561
##                                                       Unspecified
##                                                               239
```

```
barplot(Categ_CVD40)
```



BiVariate Exploration

Allergies and _____ in participants under age 12

```
#Seasonal Allergies and Dogs
knitr:: kable (count(Meta12, c('SEASONAL_ALLERGIES', 'DOG')))
```

SEASONAL_ALLERGIES	DOG	freq
false	false	255
false	true	91
false	Unknown	3
No	No	119
No	Yes	28
true	false	75
true	true	51
true	Unknown	1
Unknown	false	27
Unknown	true	8
Unknown	Unknown	3
Unspecified	No	8
Unspecified	Yes	3
Yes	No	68
Yes	Yes	16

```
#or
#Meta12 %>% filter(SEASONAL_ALLERGIES=c("true", "Yes")) %>% summary(as.factor('DOG'))

#Seasonal Allergies and Other Pets
knitr:: kable (count(Meta12, c('SEASONAL_ALLERGIES', 'PETS_OTHER')))
```

SEASONAL_ALLERGIES	PETS_OTHER	freq
false	false	28
false	no_data	276
false	true	10
false	Unknown	35
No	No	113

No	Unknown	2
No	Unspecified	5
No	Yes	27
true	false	12
true	no_data	106
true	true	1
true	Unknown	8
Unknown	false	2
Unknown	no_data	32
Unknown	Unknown	4
Unspecified	No	7
Unspecified	Unspecified	1
Unspecified	Yes	3
Yes	No	74
Yes	Unknown	2
Yes	Unspecified	5
Yes	Yes	3

```
#Seasonal Allergies and Delivery Method
knitr:: kable (count(Meta12, c('SEASONAL_ALLERGIES', 'CSECTION')))
```

SEASONAL_ALLERGIES	CSECTION	freq
false	false	4
false	No	216
false	Not sure	1
false	Unknown	3
false	Yes	125
No	No	108
No	Yes	39
true	No	94
true	Not sure	2
true	Yes	31
Unknown	No	20
Unknown	Unknown	6
Unknown	Yes	12
Unspecified	No	7
Unspecified	Unspecified	2
Unspecified	Yes	2
Yes	No	58
Yes	Not sure	1
Yes	Yes	25

```
#Seasonal Allergies and Food as Infant
knitr:: kable (count(Meta12, c('SEASONAL_ALLERGIES', 'FED_AS_INFANT')))
```

SEASONAL_ALLERGIES	FED_AS_INFANT	freq
false	A mixture of breast milk and formula	39
false	Primarily breast milk	121
false	Primarily infant formula	14
false	Unknown	175

No	A mixture of breast milk and formula	19
No	Primarily breast milk	111
No	Primarily infant formula	16
No	Unspecified	1
true	A mixture of breast milk and formula	18
true	Not sure	2
true	Primarily breast milk	28
true	Primarily infant formula	21
true	Unknown	58
Unknown	A mixture of breast milk and formula	8
Unknown	Primarily breast milk	11
Unknown	Primarily infant formula	1
Unknown	Unknown	18
Unspecified	A mixture of breast milk and formula	2
Unspecified	Primarily breast milk	5
Unspecified	Primarily infant formula	2
Unspecified	Unspecified	2
Yes	A mixture of breast milk and formula	44
Yes	Not sure	1
Yes	Primarily breast milk	34
Yes	Primarily infant formula	5

List of all variables available for further analysis

```
names(Meta)
```

```
##   [1] "#SampleID"
##   [2] "BarcodeSequence"
##   [3] "LinkerPrimerSequence"
##   [4] "VIOSCREEN_MANNITOL"
##   [5] "VIOSCREEN_SFA100"
##   [6] "NON_FOOD_ALLERGIES_BEESTINGS"
##   [7] "VIOSCREEN_FINISHED"
##   [8] "TONSILS_REMOVED"
##   [9] "WEIGHT_CHANGE"
##  [10] "VIOSCREEN_SFA170"
##  [11] "VIOSCREEN_VEGETABLE_SERVINGS"
##  [12] "VIOSCREEN_SORBITOL"
##  [13] "COUNTRY_OF_BIRTH"
##  [14] "PETS_OTHER"
##  [15] "MASTERMIX_LOT"
##  [16] "VIOSCREEN_SFA80"
##  [17] "VIOSCREEN_BETACAR"
##  [18] "VIOSCREEN_LOW_FAT_DAIRY_SERVING"
##  [19] "VIOSCREEN_NIACINEQ"
##  [20] "ORIG_NAME"
##  [21] "ALLERGIC_TO_OTHER"
##  [22] "VIOSCREEN_GAMMTOCO"
##  [23] "SPECIALIZED_DIET_I_DO_NOT_EAT_A_SPECIALIZED_DIET"
##  [24] "VIOSCREEN_LYSINE"
##  [25] "VIOSCREEN_ISOMALT"
##  [26] "LIVER_DISEASE"
##  [27] "BODY_PRODUCT"
##  [28] "VIOSCREEN_DATABASE"
##  [29] "VIOSCREEN_FIBINSO"
##  [30] "WHOLE_GRAIN_FREQUENCY"
##  [31] "ARTIFICIAL_SWEETENERS"
##  [32] "ENV_MATERIAL"
##  [33] "VIOSCREEN_SUCROSE"
##  [34] "ENA-BASE-COUNT"
##  [35] "VIOSCREEN_CLAC9T11"
##  [36] "SUBSET_HEALTHY"
##  [37] "VIOSCREEN_FIBER"
##  [38] "VIOSCREEN_LACTITOL"
```

```
##  [39] "VIOSCREEN_HEI_NON_JUICE_FRT"
##  [40] "VIOSCREEN_HEI_SCORE"
##  [41] "ALLERGIC_TO_UNSPECIFIED"
##  [42] "VIOSCREEN_OXALICM"
##  [43] "VIOSCREEN_ALPHACAR"
##  [44] "SALTED_SNACKS_FREQUENCY"
##  [45] "ALCOHOL_TYPES_UNSPECIFIED"
##  [46] "AGE_CAT"
##  [47] "VIOSCREEN_QUESTIONNAIRE"
##  [48] "VIOSCREEN_LYCOPENE"
##  [49] "VIOSCREEN_VEG5_DAY"
##  [50] "SUBSET_BMI"
##  [51] "MIGRAINE"
##  [52] "VIOSCREEN_HEI_MILK"
##  [53] "SPECIALIZED_DIET_EXCLUDE_REFINED_SUGARS"
##  [54] "VIOSCREEN_CLAT10C12"
##  [55] "VIOSCREEN_CARBO"
##  [56] "VIOSCREEN_HEI_MEAT_BEANS"
##  [57] "VIOSCREEN_HEI_SOL_FAT_ALC_ADD_SUG"
##  [58] "AGE_YEARS"
##  [59] "VIOSCREEN_GLUCOSE"
##  [60] "VIOSCREEN_HEI2010_REFINED_GRAINS"
##  [61] "LIBRARY_CONSTRUCTION_PROTOCOL"
##  [62] "VIOSCREEN_LUTZEAX"
##  [63] "EXERCISE_FREQUENCY"
##  [64] "HOST_COMMON_NAME"
##  [65] "EXTRACTION_ROBOT"
##  [66] "SPECIALIZED_DIET_KOSHER"
##  [67] "ALCOHOL_TYPES_SPIRITSHARD_ALCOHOL"
##  [68] "VIOSCREEN_SALAD_VEGETABLE_SERVINGS"
##  [69] "BOWEL_MOVEMENT_FREQUENCY"
##  [70] "VIOSCREEN_SFA200"
##  [71] "SIBO"
##  [72] "FERMENTED_PLANT_FREQUENCY"
##  [73] "ENV_BIOME"
##  [74] "SUGARY_SWEETS_FREQUENCY"
##  [75] "VIOSCREEN_HEI2010_DAIRY"
##  [76] "VIOSCREEN_FIBH2O"
##  [77] "DOG"
##  [78] "VIOSCREEN_GLAC"
##  [79] "VIOSCREEN_PHENYLAL"
##  [80] "VIOSCREEN_SFA140"
##  [81] "VIOSCREEN_TAGATOSE"
##  [82] "VIOSCREEN_HEI_WHL_GRAINS"
##  [83] "COLLECTION_DATE"
##  [84] "VIOSCREEN_VISIT"
##  [85] "VIOSCREEN_HEI2010_WHOLE_GRAINS"
##  [86] "VIOSCREEN_MULTI_CALCIUM_AVG"
##  [87] "ANONYMIZED_NAME"
##  [88] "VIOSCREEN_MULTIVITAMIN"
##  [89] "PREPARED_MEALS_FREQUENCY"
##  [90] "HOMECOOKED_MEALS_FREQUENCY"
##  [91] "VIOSCREEN_F_NJ_OTHER"
##  [92] "VIOSCREEN_D_MILK"
##  [93] "HEIGHT_UNITS"
##  [94] "VIOSCREEN_M_ORGAN"
##  [95] "VIOSCREEN_M_POULT"
##  [96] "VIOSCREEN_M_MPF"
##  [97] "VIOSCREEN_WHOLE_GRAIN_SERVINGS"
##  [98] "SPECIALIZED_DIET_EXCLUDE_NIGHTSHADES"
##  [99] "VIOSCREEN_HEI2010_PROTIEN_FOODS"
## [100] "SPECIALIZED_DIET_PALEODIET_OR_PRIMAL_DIET"
## [101] "ENV_FEATURE"
## [102] "VIOSCREEN_NIACIN"
## [103] "VIOSCREEN_NCCGLGR"
## [104] "VIOSCREEN_OXALIC"
## [105] "VIOSCREEN_HEI_FRUIT"
## [106] "ECONOMIC_REGION"
## [107] "VITAMIN_D_SUPPLEMENT_FREQUENCY"
## [108] "ADD_ADHD"
## [109] "READY_TO_EAT_MEALS_FREQUENCY"
## [110] "VIOSCREEN_MFA161"
## [111] "VIOSCREEN_SATOCO"
## [112] "VIOSCREEN_RIBOFLA"
## [113] "VIOSCREEN_STARCH"
## [114] "MENTAL_ILLNESS_TYPE_UNSPECIFIED"
## [115] "IBD"
## [116] "CAT"
## [117] "WELL_DESCRIPTION"
## [118] "PROCESSING_ROBOT"
## [119] "HOST_TAXID"
```

```
## [120] "VIOSCREEN_V_TOTAL"
## [121] "VIOSCREEN_V_ORANGE"
## [122] "IBS"
## [123] "VIOSCREEN_M_FRANK"
## [124] "COLLECTION_TIMESTAMP"
## [125] "VIOSCREEN_DELTTOCO"
## [126] "RUN_PREFIX"
## [127] "MULTIVITAMIN"
## [128] "EXPERIMENT_TITLE"
## [129] "CANCER"
## [130] "ACNE_MEDICATION_OTC"
## [131] "VIOSCREEN_FRUIT_SERVINGS"
## [132] "EXPERIMENT_DESIGN_DESCRIPTION"
## [133] "ALLERGIC_TO_I_HAVE_NO_FOOD_ALLERGIES_THAT_I_KNOW_OF"
## [134] "DIABETES"
## [135] "CDIFF"
## [136] "VIOSCREEN_SODIUM"
## [137] "RUN_CENTER"
## [138] "VIOSCREEN_BETATOCO"
## [139] "HEIGHT_CM"
## [140] "VIOSCREEN_FRIED_FOOD_SERVINGS"
## [141] "VIOSCREEN_FISH_SERVINGS"
## [142] "CHICKENPOX"
## [143] "VIOSCREEN_FRTSUMM"
## [144] "VIOSCREEN_GRAMS"
## [145] "ASSIGNED_FROM_GEO"
## [146] "VIOSCREEN_VALINE"
## [147] "EPILEPSY_OR_SEIZURE_DISORDER"
## [148] "BODY_SITE"
## [149] "VIOSCREEN_HISTIDIN"
## [150] "VIOSCREEN_BETAINE"
## [151] "VIOSCREEN_ALCOHOL_SERVINGS"
## [152] "FUNGAL_OVERGROWTH"
## [153] "FLOSSING_FREQUENCY"
## [154] "EXTRACTIONKIT_LOT"
## [155] "ONE_LITER_OF_WATER_A_DAY_FREQUENCY"
## [156] "SPECIALIZED_DIET_UNSPECIFIED"
## [157] "GEO_LOC_NAME"
## [158] "COSMETICS_FREQUENCY"
## [159] "HOST_SUBJECT_ID"
## [160] "VIOSCREEN_SFATOT"
## [161] "VIOSCREEN_POTASS"
## [162] "TM1000_8_TOOL"
## [163] "VIOSCREEN_PHYTIC"
## [164] "NON_FOOD_ALLERGIES_SUN"
## [165] "VIOSCREEN_PECTINS"
## [166] "CARDIOVASCULAR_DISEASE"
## [167] "DRINKS_PER_SESSION"
## [168] "VIOSCREEN_SFA180"
## [169] "VIOSCREEN_FOL_SYN"
## [170] "VIOSCREEN_F_NJ_TOTAL"
## [171] "VIOSCREEN_A_CAL"
## [172] "FED_AS_INFANT"
## [173] "MENTAL_ILLNESS"
## [174] "SPECIALIZED_DIET_WESTENPRICE_OR_OTHER_LOWGRAIN_LOW_PROCESSED_FO"
## [175] "NON_FOOD_ALLERGIES"
## [176] "WEIGHT_UNITS"
## [177] "VIOSCREEN_ACTIVITY_LEVEL"
## [178] "VIOSCREEN_CALORIES"
## [179] "COUNTRY"
## [180] "ACID_REFLUX"
## [181] "VIOSCREEN_MULTIVITAMIN_FREQ"
## [182] "VIOSCREEN_GENDER"
## [183] "COMMON_NAME"
## [184] "VIOSCREEN_A_BEV"
## [185] "VIOSCREEN_HEI2010_EMPTY_CALORIES"
## [186] "NON_FOOD_ALLERGIES_UNSPECIFIED"
## [187] "FLU_VACCINE_DATE"
## [188] "PREGNANT"
## [189] "BMI"
## [190] "CENSUS_REGION"
## [191] "SPECIALIZED_DIET_RAW_FOOD_DIET"
## [192] "TEETHBRUSHING_FREQUENCY"
## [193] "DEPRESSION_BIPOLAR_SCHIZOPHRENIA"
## [194] "FROZEN_DESSERT_FREQUENCY"
## [195] "LIVINGWITH"
## [196] "CLINICAL_CONDITION"
## [197] "VIOSCREEN_SCFV"
## [198] "TARGET_GENE"
## [199] "VIOSCREEN_LEGUMES"
## [200] "VIOSCREEN_HEI2010_SODIUM"
```

```
## [201] "VIOSCREEN_PROTOCOL"
## [202] "ALTITUDE"
## [203] "LAST_TRAVEL"
## [204] "DNA_EXTRACTED"
## [205] "VIOSCREEN_TRYPTOPH"
## [206] "VIOSCREEN_TOTALTFA"
## [207] "ROOMMATES"
## [208] "VIOSCREEN_ALCOHOL"
## [209] "VIOSCREEN_SACCHAR"
## [210] "VIOSCREEN_ERYTHR"
## [211] "VIOSCREEN_D_TOTAL"
## [212] "VIOSCREEN_F_TOTAL"
## [213] "SEX"
## [214] "ALLERGIC_TO_TREE_NUTS"
## [215] "VIOSCREEN_NON_FRIED_FISH_SERVINGS"
## [216] "ASD"
## [217] "VIVID_DREAMS"
## [218] "VIOSCREEN_HEI_OILS"
## [219] "LUNG_DISEASE"
## [220] "SAMP_SIZE"
## [221] "VIOSCREEN_ADD_SUG"
## [222] "VIOSCREEN_PFA225"
## [223] "VIOSCREEN_HEI2010_WHOLE_FRUIT"
## [224] "ALCOHOL_FREQUENCY"
## [225] "POULTRY_FREQUENCY"
## [226] "DOMINANT_HAND"
## [227] "VIOSCREEN_FAT"
## [228] "VIOSCREEN_PROLINE"
## [229] "VIOSCREEN_M_MEAT"
## [230] "ROOMMATES_IN_STUDY"
## [231] "VIOSCREEN_CALCIUM_AVG"
## [232] "VIOSCREEN_WEIGHT"
## [233] "SURVEY_ID"
## [234] "VIOSCREEN_G_WHL"
## [235] "SUBSET_ANTIBIOTIC_HISTORY"
## [236] "SCIENTIFIC_NAME"
## [237] "VIOSCREEN_PROTANIM"
## [238] "SMOKING_FREQUENCY"
## [239] "TM50_8_TOOL"
## [240] "VIOSCREEN_CALCIUM_SERVINGS"
## [241] "DEPTH"
## [242] "VIOSCREEN_BMI"
## [243] "VIOSCREEN_M_EGG"
## [244] "BREASTMILK_FORMULA_ENSURE"
## [245] "PRIMER_PLATE"
## [246] "PLATING"
## [247] "VIOSCREEN_FOL_NAT"
## [248] "LEVEL_OF_EDUCATION"
## [249] "MENTAL_ILLNESS_TYPE_DEPRESSION"
## [250] "VIOSCREEN_MAGNES"
## [251] "KIDNEY_DISEASE"
## [252] "VIOSCREEN_VITA_RE"
## [253] "COLLECTION_TIME"
## [254] "VIOSCREEN_VEGSUMM"
## [255] "SPECIALIZED_DIET"
## [256] "VIOSCREEN_SFA60"
## [257] "COLLECTION_MONTH"
## [258] "VITAMIN_B_SUPPLEMENT_FREQUENCY"
## [259] "TM300_8_TOOL"
## [260] "VIOSCREEN_PFA226"
## [261] "SUBSET_IBD"
## [262] "VIOSCREEN_HEI_SAT_FAT"
## [263] "VIOSCREEN_TOTFOLAT"
## [264] "VIOSCREEN_ARGININE"
## [265] "VIOSCREEN_HEI2010_FATTY_ACIDS"
## [266] "ALCOHOL_TYPES_SOUR_BEERS"
## [267] "SPECIALIZED_DIET_EXCLUDE_DAIRY"
## [268] "RUN_DATE"
## [269] "VIOSCREEN_ALPHTOCE"
## [270] "PHYSICAL_SPECIMEN_LOCATION"
## [271] "VIOSCREEN_PROTEIN"
## [272] "VIOSCREEN_ALPHTOCO"
## [273] "VIOSCREEN_GLTC"
## [274] "VIOSCREEN_MULTI_CALCIUM_DOSE"
## [275] "ALLERGIC_TO"
## [276] "VIOSCREEN_VITA_IU"
## [277] "MENTAL_ILLNESS_TYPE_PTSD_POSTTRAUMATIC_STRESS_DISORDER"
## [278] "NAIL_BITER"
## [279] "VIOSCREEN_CALCIUM_FREQ"
## [280] "ALCOHOL_TYPES_RED_WINE"
## [281] "ENA-CHECKLIST"
```

```
## [282] "GLUTEN"
## [283] "VIOSCREEN_SELENIUM"
## [284] "VIOSCREEN_TOTCLA"
## [285] "VIOSCREEN_MFA201"
## [286] "WEIGHT_KG"
## [287] "SUBSET_AGE"
## [288] "CONTRACEPTIVE"
## [289] "VIOSCREEN_PFA184"
## [290] "VIOSCREEN_PFA183"
## [291] "VIOSCREEN_PFA182"
## [292] "VIOSCREEN_D_TOT_SOYM"
## [293] "VIOSCREEN_PROTVEG"
## [294] "VIOSCREEN_DISCFAT_OIL"
## [295] "PCR_PRIMERS"
## [296] "MILK_SUBSTITUTE_FREQUENCY"
## [297] "VIOSCREEN_JUICE_SERVINGS"
## [298] "VIOSCREEN_FRT5_DAY"
## [299] "VIOSCREEN_HEIGHT"
## [300] "VIOSCREEN_PINITOL"
## [301] "VIOSCREEN_CALCIUM"
## [302] "LOWGRAIN_DIET_TYPE"
## [303] "VIOSCREEN_GENISTN"
## [304] "SEAFOOD_FREQUENCY"
## [305] "VIOSCREEN_SFA120"
## [306] "VIOSCREEN_HEI_GRAINS"
## [307] "ALLERGIC_TO_PEANUTS"
## [308] "QIITA_PREP_ID"
## [309] "PROBIOTIC_FREQUENCY"
## [310] "MENTAL_ILLNESS_TYPE_BULIMIA_NERVOSA"
## [311] "WELL_ID"
## [312] "VIOSCREEN_METHION"
## [313] "VIOSCREEN_HEI_VEG"
## [314] "VIOSCREEN_COPPER"
## [315] "OLIVE_OIL"
## [316] "VIOSCREEN_EER"
## [317] "VIOSCREEN_CHOLINE"
## [318] "VIOSCREEN_ISOLEUC"
## [319] "VIOSCREEN_STARTED"
## [320] "VIOSCREEN_G_TOTAL"
## [321] "VIOSCREEN_TFA161T"
## [322] "SKIN_CONDITION"
## [323] "CANCER_TREATMENT"
## [324] "ENV_PACKAGE"
## [325] "VIOSCREEN_XYLITOL"
## [326] "VIOSCREEN_BIOCHANA"
## [327] "VIOSCREEN_F_NJ_CITMLB"
## [328] "VIOSCREEN_HEI2010_VEG"
## [329] "VIOSCREEN_M_NUTSD"
## [330] "VIOSCREEN_PFATOT"
## [331] "TARGET_SUBFRAGMENT"
## [332] "TITLE"
## [333] "VIOSCREEN_V_OTHER"
## [334] "IBD_DIAGNOSIS"
## [335] "VIOSCREEN_WGRAIN"
## [336] "CENTER_PROJECT_NAME"
## [337] "VIOSCREEN_TFA182T"
## [338] "VIOSCREEN_COUMEST"
## [339] "ENA-SPOT-COUNT"
## [340] "EXERCISE_LOCATION"
## [341] "VIOSCREEN_M_FISH_LO"
## [342] "NON_FOOD_ALLERGIES_POISON_IVYOAK"
## [343] "VIOSCREEN_VITD2"
## [344] "OTHER_SUPPLEMENT_FREQUENCY"
## [345] "MENTAL_ILLNESS_TYPE"
## [346] "VIOSCREEN_RETINOL"
## [347] "SUBSET_DIABETES"
## [348] "ALLERGIC_TO_SHELLFISH"
## [349] "VIOSCREEN_MFATOT"
## [350] "VIOSCREEN_CYSTINE"
## [351] "FRUIT_FREQUENCY"
## [352] "VIOSCREEN_V_POTATO"
## [353] "VIOSCREEN_PANTOTHE"
## [354] "VIOSCREEN_RECNO"
## [355] "VIOSCREEN_SUBJECT_ID"
## [356] "STATE"
## [357] "HAS_PHYSICAL_SPECIMEN"
## [358] "VIOSCREEN_PHOSPHOR"
## [359] "VIOSCREEN_CALCIUM_FROM_DAIRY_SERVINGS"
## [360] "VIOSCREEN_MALTITOL"
## [361] "WATER_LOT"
## [362] "SPECIALIZED_DIET_FODMAP"
```

```
## [363] "VIOSCREEN_ASPARTIC"
## [364] "VIOSCREEN_V_STARCY"
## [365] "NON_FOOD_ALLERGIES_PET_DANDER"
## [366] "VIOSCREEN_SFA220"
## [367] "COUNTRY_RESIDENCE"
## [368] "SLEEP_DURATION"
## [369] "PUBLIC"
## [370] "QIITA_STUDY_ID"
## [371] "VIOSCREEN_F_CITMLB"
## [372] "VIOSCREEN_D_YOGURT"
## [373] "LACTOSE"
## [374] "COLLECTION_SEASON"
## [375] "VIOSCREEN_THREONIN"
## [376] "WHOLE_EGGS"
## [377] "TAXON_ID"
## [378] "BOWEL_MOVEMENT_QUALITY"
## [379] "VIOSCREEN_BCODEID"
## [380] "MENTAL_ILLNESS_TYPE_SUBSTANCE_ABUSE"
## [381] "VIOSCREEN_V_TOMATO"
## [382] "TYPES_OF_PLANTS"
## [383] "PROJECT_NAME"
## [384] "VIOSCREEN_HEI2010_FRUIT"
## [385] "PLATFORM"
## [386] "DEODORANT_USE"
## [387] "VIOSCREEN_DISCFAT_SOL"
## [388] "VIOSCREEN_TYROSINE"
## [389] "VIOSCREEN_WATER"
## [390] "VIOSCREEN_MFA141"
## [391] "VIOSCREEN_HEI2010_SCORE"
## [392] "VIOSCREEN_VITD"
## [393] "VIOSCREEN_TOTSUGAR"
## [394] "VIOSCREEN_VITC"
## [395] "MENTAL_ILLNESS_TYPE_SCHIZOPHRENIA"
## [396] "VIOSCREEN_NUTRIENT_RECOMMENDATION"
## [397] "VIOSCREEN_VITK"
## [398] "VIOSCREEN_SFA160"
## [399] "VIOSCREEN_FRUCTOSE"
## [400] "SPECIALIZED_DIET_HALAAL"
## [401] "SEASONAL_ALLERGIES"
## [402] "THYROID"
## [403] "VIOSCREEN_ADDSUGAR"
## [404] "VIOSCREEN_EMAIL"
## [405] "INSTRUMENT_MODEL"
## [406] "VIOSCREEN_RGRAIN"
## [407] "VIOSCREEN_V_DRKGR"
## [408] "VIOSCREEN_CALCIUM_DOSE"
## [409] "VIOSCREEN_F_OTHER"
## [410] "ELEVATION"
## [411] "VIOSCREEN_ADSUGTOT"
## [412] "VIOSCREEN_IRON"
## [413] "CSECTION"
## [414] "RACE"
## [415] "VIOSCREEN_ZINC"
## [416] "VIOSCREEN_AVCARB"
## [417] "VIOSCREEN_MFA221"
## [418] "VIOSCREEN_OMEGA3"
## [419] "REQUIRED_SAMPLE_INFO_STATUS"
## [420] "LONGITUDE"
## [421] "PRIMER_DATE"
## [422] "SAMPLE_PLATE"
## [423] "ALCOHOL_TYPES_WHITE_WINE"
## [424] "ALZHEIMERS"
## [425] "VIOSCREEN_SRVID"
## [426] "VIOSCREEN_LINE_GI"
## [427] "CENTER_NAME"
## [428] "VIOSCREEN_TGRAIN"
## [429] "SAMPLE_TYPE"
## [430] "NON_FOOD_ALLERGIES_DRUG_EG_PENICILLIN"
## [431] "MENTAL_ILLNESS_TYPE_ANOREXIA_NERVOSA"
## [432] "VIOSCREEN_GALACTOS"
## [433] "SUGAR_SWEETENED_DRINK_FREQUENCY"
## [434] "VIOSCREEN_SUCRLOSE"
## [435] "LAST_MOVE"
## [436] "VIOSCREEN_M_SOY"
## [437] "VIOSCREEN_SUCPOLY"
## [438] "PETS_OTHER_FREETEXT"
## [439] "CONSUME_ANIMAL_PRODUCTS_ABX"
## [440] "VIOSCREEN_ALANINE"
## [441] "LINKER"
## [442] "VIOSCREEN_MFA181"
## [443] "VIOSCREEN_M_FISH_HI"
```

```
## [444] "VIOSCREEN_TFA181T"
## [445] "VIOSCREEN_MALTOSE"
## [446] "ENV_MATTER"
## [447] "VIOSCREEN_VITD3"
## [448] "VIOSCREEN_ACESUPOT"
## [449] "VIOSCREEN_PROCDATE"
## [450] "VIOSCREEN_GLUTAMIC"
## [451] "VIOSCREEN_HEI2010_SEA_FOODS_PLANT_PROTIENS"
## [452] "VIOSCREEN_LEUCINE"
## [453] "MENTAL_ILLNESS_TYPE_BIPOLAR_DISORDER"
## [454] "ACNE_MEDICATION"
## [455] "VIOSCREEN_THIAMIN"
## [456] "IBD_DIAGNOSIS_REFINED"
## [457] "MILK_CHEESE_FREQUENCY"
## [458] "VIOSCREEN_ASPARTAM"
## [459] "ALCOHOL_TYPES_BEERCIDER"
## [460] "VIOSCREEN_VITD_IU"
## [461] "ANTIBIOTIC_HISTORY"
## [462] "HIGH_FAT_RED_MEAT_FREQUENCY"
## [463] "PHYSICAL_SPECIMEN_REMAINING"
## [464] "VIOSCREEN_USER_ID"
## [465] "VIOSCREEN_DOB"
## [466] "VIOSCREEN_HEI_DRK_G_ORG_VEG_LEG"
## [467] "VIOSCREEN_INOSITOL"
## [468] "SOFTENER"
## [469] "VIOSCREEN_FRIED_FISH_SERVINGS"
## [470] "VIOSCREEN_VITB12"
## [471] "VIOSCREEN_SFA40"
## [472] "BODY_HABITAT"
## [473] "VIOSCREEN_CHOLEST"
## [474] "VIOSCREEN_TIME"
## [475] "VIOSCREEN_VITE_IU"
## [476] "APPENDIX_REMOVED"
## [477] "DRINKING_WATER_SOURCE"
## [478] "VIOSCREEN_VITA_RAE"
## [479] "BMI_CORRECTED"
## [480] "DIABETES_TYPE"
## [481] "BMI_CAT"
## [482] "VIOSCREEN_MANGAN"
## [483] "VIOSCREEN_BETACRYP"
## [484] "DIET_TYPE"
## [485] "RED_MEAT_FREQUENCY"
## [486] "AUTOIMMUNE"
## [487] "BIRTH_YEAR"
## [488] "VIOSCREEN_AGE"
## [489] "VIOSCREEN_GLYCITN"
## [490] "VIOSCREEN_SCF"
## [491] "VIOSCREEN_GLYCINE"
## [492] "ALCOHOL_CONSUMPTION"
## [493] "VIOSCREEN_LACTOSE"
## [494] "ALCOHOL_TYPES"
## [495] "VIOSCREEN_VITB6"
## [496] "MEAT_EGGS_FREQUENCY"
## [497] "VIOSCREEN_D_CHEESE"
## [498] "SPECIALIZED_DIET_MODIFIED_PALEO_DIET"
## [499] "VIOSCREEN_NCCGLBR"
## [500] "VIOSCREEN_HEI2010_GREENS_BEANS"
## [501] "VIOSCREEN_CAFFEINE"
## [502] "VIOSCREEN_DAIDZEIN"
## [503] "VIOSCREEN_NITROGEN"
## [504] "VIOSCREEN_SERINE"
## [505] "VIOSCREEN_PFA204"
## [506] "VIOSCREEN_PFA205"
## [507] "VIOSCREEN_G_NWHL"
## [508] "VIOSCREEN_HEI_SODIUM"
## [509] "LATITUDE"
## [510] "VEGETABLE_FREQUENCY"
## [511] "SEQUENCING_METH"
## [512] "VIOSCREEN_JOULES"
## [513] "VIOSCREEN_FORMONTN"
## [514] "VIOSCREEN_SWEET_SERVINGS"
## [515] "EXPERIMENT_CENTER"
## [516] "POOL_FREQUENCY"
## [517] "VIOSCREEN_FOL_DEQV"
## [518] "PKU"
## [519] "VIOSCREEN_ASH"
## [520] "VIOSCREEN_METHHIS3"
## [521] "AGE_CORRECTED"
## [522] "SPECIALIZED_DIET_OTHER_RESTRICTIONS_NOT_DESCRIBED_HERE"
## [523] "VIOSCREEN_NATOCO"
## [524] "Description"
```

```
## [525] "SIMPLE_BODY_SITE"
## [526] "TITLE_ACRONYM"
## [527] "TITLE_BODY_SITE"
## [528] "HMP_SITE"
```

Interactive Data Widget

Here is an intereactive and searchable version of the dataset. The same could also be done by exporting it to an excel file.

```
library(DT)
datatable(Meta, options = list(pageLength = 25))
```

Show 10 entries

Search:

	#SampleID	BarcodeSequence	LinkerPrimerSequence	VIOSCREEN_MANNITOL	VIOSCREEN_SFA100	NON_FOOD_
1	10317.000027826	no_data	no_data	0.496931478	0.88257555	false
2	10317.000027827	no_data	no_data			false
3	10317.000027824	ATATTGGCAGCC	GTGTGCCAGCMGCCGCGGTAA			No
4	10317.000027825	no_data	no_data			false
5	10317.000027822	GGCTTCGGAGCG	GTGTGCCAGCMGCCGCGGTAA	0.289205452	0.432219205	No
6	10317.000014400	no_data	no_data			false
7	10317.000027820	no_data	no_data			false
8	10317.000065774	GAGCGCCGAACA	GTGTGCCAGCMGCCGCGGTAA			No

9	10317.000065776	GTAAACGACTTG	GTGTGCCAGCMGCCGCGGTAA			No
10	10317.000065777	ACATAGCGGTTC	GTGTGCCAGCMGCCGCGGTAA			No
11	10317.000065770	GTCACCAATCCG	GTGTGCCAGCMGCCGCGGTAA			No
12	10317.000065771	GAGTCCGTTGCT	GTGTGCCAGCMGCCGCGGTAA			No
13	10317.000027828	no_data	no_data	0.203397254	0.040273975	false
14	10317.000027829	no_data	no_data			false
15	10317.000001386	no_data	no_data			false
16	10317.000001384	no_data	no_data			false
17	10317.000001380	no_data	no_data			false
18	10317.000001389	no_data	no_data			false
19	10317.000030085	GGACAAGTGCGA	GTGTGCCAGCMGCCGCGGTAA			No
20	10317.000030086	no_data	no_data	0.248109582	0.777780778	false

21	10317.000030087	no_data	no_data			false
22	10317.000030080	no_data	no_data			false
23	10317.000030081	no_data	no_data			false
24	10317.000030082	no_data	no_data	0.269287674	0.656739703	false
25	10317.000030083	no_data	no_data	0.115671236	0.296684934	false

Input OTU/Taxonomy

Create a phyloseq object to work with inside of the phyloseq package. To do this we will upload the ".biom" file then merge the metadata, OTU, taxonomy, and tree file. This will allow us to explore the diversity of the sample. Again check your file names and paths in order to use this code. (Note: ".biom" file is found in the "OTU" folder of the "latest" data. The "meta" folder appears to be missing the ".biom" file associated with the previously used metadata within the "Meta" folder)

```
##Import OTU file
#Data originally downloaded from <ftp://ftp.microbio.me/AmericanGut/latest>
#Make sure to pull the correct biom file to match the meta data down for analysis.

# To read in original .biom file use command below
file_path<- ("~/Desktop/otu_table.biom")
dat <- import_biom(file_path)

#To fix error saying "input string 1 is invalid in this locale" run command "Sys.setlocale(locale="C")" in consol
e and run command chunk again

#Make Phyloseq values for separate pieces of the biom file
#Taxonomy
taxonomy <- tax_table(dat)
#OTU
OTU_table<- otu_table(dat)

#MetaData
#Create Phyloseq object for sample data so it can be merged
SampleData<- sample_data(Meta)
SD<-sample_data(Meta)

#Import the tree corresponding to this data ### Will Not work
tree<-read_tree_greengenes("~/Desktop/97_otus.tree")

#Merge OTU, Taxonomy, and metadata into one phyloseq object called "ps"that can be used for analysis
Sample_data <- (SD)
sam_cov <- as.data.frame(Sample_data)
rownames(sam_cov) <- sam_cov$"#SampleID"
sd <- sample_data(sam_cov)
sample_names(sd) #Previous steps fix error in naming of SampleIds
otus <- otu_table(OTU_table, taxa_are_rows = TRUE)
sample_names(otus)
tt <- tax_table(taxonomy)
ps <- phyloseq(otus, sd, tt,tree) #Step to merge OTU,Taxonomy, and metadata
ps #View object and make sure it has all 3 parts
```

Summary of OTU

```
#Summary of OTU
ntaxa(ps)
rank_names(ps)
nsamples(ps)
sample_names(ps)[1:5]
otu_table(ps)[1:5, 1:5]
tax_table(ps)[1:5, 1:4]
taxa_names(ps)[1:10]
```