

Mimic-Fasting Diet Plan

Sandra Kilpatrick

July 21, 2015

Set directory and load libraries

```
setwd("~/DataScience/NutritionDatabase")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##   filter
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(httr)
library(ggplot2)
library(tidyr)
library(extrafont)
```

```
## Registering fonts with R
```

Read files into R

```
parse_file <- function(x) {
  base_url <- "http://www.ars.usda.gov/SP2UserFiles/Place/12354500/Data/SR27/asc/"
  resp <- GET(paste0(base_url, x))

  text <- content(resp, "text", encoding = "ISO-8859-1")
  read.delim(text = text, sep = "^", quote = "~", na.strings = c("^", "~"),
             header = FALSE, stringsAsFactors = FALSE) %>% tbl_df()
}
```

Create food

```

food <- parse_file("FOOD_DES.txt")
names(food) <- c("nbd_no", "fdgrp_cd", "long_desc", "shrt_desc", "comname", "manufacname",
"survey", "ref_desc", "refuse", "sciname", "n_factor", "pro_factor", "fat_factor",
"cho_factor")
food$survey <- food$survey == "Y"
food$nbd_no=as.character(food$nbd_no)
food$fdgrp_cd=as.character(food$fdgrp_cd)

```

Create food group

```

food_group <- parse_file("FD_GROUP.txt")
names(food_group) <- c("fdgrp_cd", "fdgrp_desc")
food_group$fdgrp_cd=as.character(food_group$fdgrp_cd)

```

Create nutrient

```

nutrient <- parse_file("NUT_DATA.txt")
names(nutrient) <- c("nbd_no", "nutr_no", "nutr_val", "num_data_pts", "std_error",
"src_cd", "deriv_cd", "ref_ndb_no", "add_nutr_mark", "num_studies", "min",
"max", "df", "low_eb", "up_eb", "stat_cmt", "addmod_date", "cc")
nutrient$nbd_no=as.character(nutrient$nbd_no)
nutrient$nutr_no=as.character(nutrient$nutr_no)
nutrient$fortified[nutrient$add_nutr_mark == ""] <- NA

```

Create nutrient definition file

```

nutrient_def <- parse_file("NUTR_DEF.txt")
names(nutrient_def) <- c("nutr_no", "units", "tagname", "nutrdesc", "num_dec", "sr_order")
nutrient_def$nutr_no=as.character(nutrient_def$nutr_no)

```

Create weight file

```

weight <- parse_file("WEIGHT.txt")
names(weight) <- c("nbd_no", "seq", "amount", "measure_desc", "gm_wgt",
"num_data_pts", "std_dev")
weight$nbd_no=as.character(weight$nbd_no)

```

Explore the dataset

```
glimpse(food)
```

```
## Observations: 8618
## Variables:
## $ nbd_no      (chr) "1001", "1002", "1003", "1004", "1005", "1006", "1...
## $ fdgrp_cd    (chr) "100", "100", "100", "100", "100", "100", "100", "...
## $ long_desc   (chr) "Butter, salted", "Butter, whipped, with salt", "B...
## $ shrt_desc   (chr) "BUTTER,WITH SALT", "BUTTER,WHIPPED,WITH SALT", "B...
## $ comname     (chr) "", "", "", "", "", "", "", "", "", "", "", "", "", "...
## $ manufacname (chr) "", "", "", "", "", "", "", "", "", "", "", "", "", "...
## $ survey      (lgl) TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, FALSE, T...
## $ ref_desc    (chr) "", "", "", "", "", "", "", "", "", "", "", "", "", "...
## $ refuse      (int) 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ sciname     (chr) "", "", "", "", "", "", "", "", "", "", "", "", "", "...
## $ n_factor    (dbl) 6.38, 6.38, 6.38, 6.38, 6.38, 6.38, 6.38, 6.38, NA...
## $ pro_factor  (dbl) 4.27, 4.27, 4.27, 4.27, 4.27, 4.27, 4.27, 4.27, NA...
## $ fat_factor  (dbl) 8.79, 8.79, 8.79, 8.79, 8.79, 8.79, 8.79, 8.79, NA...
## $ cho_factor  (dbl) 3.87, 3.87, 3.87, 3.87, 3.87, 3.87, 3.87, 3.87, NA...
```

```
glimpse(food_group)
```

```
## Observations: 25
## Variables:
## $ fdgrp_cd    (chr) "100", "200", "300", "400", "500", "600", "700", "8...
## $ fdgrp_desc  (chr) "Dairy and Egg Products", "Spices and Herbs", "Baby...
```

```
glimpse(nutrient)
```

```
## Observations: 654572
## Variables:
## $ nbd_no      (chr) "1001", "1001", "1001", "1001", "1001", "1001", ...
## $ nutr_no     (chr) "203", "204", "205", "207", "208", "221", "255",...
## $ nutr_val    (dbl) 0.85, 81.11, 0.06, 2.11, 717.00, 0.00, 15.87, 0....
## $ num_data_pts (int) 16, 580, 0, 35, 0, 0, 522, 0, 0, 0, 0, 0, 17, 18...
## $ std_error   (dbl) 0.074, 0.065, NA, 0.054, NA, NA, 0.061, NA, NA, ...
## $ src_cd      (int) 1, 1, 4, 1, 4, 7, 1, 7, 7, 4, 4, 4, 1, 1, 1, 1, ...
## $ deriv_cd    (chr) "", "", "NC", "", "NC", "", "", "Z", "Z", "", "N...
## $ ref_ndb_no  (int) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ add_nutr_mark (chr) "", "", "", "", "", "", "", "", "", "", "", ...
## $ num_studies (int) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ min         (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ max         (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ df          (int) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ low_eb      (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ up_eb       (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ stat_cmt     (chr) "", "", "", "", "", "", "", "", "", "", "", ...
## $ addmod_date  (chr) "11/1976", "11/1976", "11/1976", "11/1976", "08/...
## $ cc          (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ fortified   (lgl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
```

```
glimpse(nutrient_def)
```

```
## Observations: 150
## Variables:
## $ nutr_no   (chr) "203", "204", "205", "207", "208", "209", "210", "211..."
## $ units     (chr) "g", "g", "g", "g", "kcal", "g", "g", "g", "g", "g", ...
## $ tagname    (chr) "PROCNT", "FAT", "CHOCDF", "ASH", "ENERC_KCAL", "STAR..."
## $ nutrdesc   (chr) "Protein", "Total lipid (fat)", "Carbohydrate, by dif..."
## $ num_dec    (int) 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 1, 2, 2, 0, 0, 0, 2, ...
## $ sr_order   (int) 600, 800, 1100, 1000, 300, 2200, 1600, 1700, 1800, 19...
```

```
glimpse(weight)
```

```
## Observations: 15228
## Variables:
## $ nbd_no      (chr) "1001", "1001", "1001", "1001", "1002", "1002", "...
## $ seq         (int) 1, 2, 3, 4, 1, 2, 3, 4, 1, 2, 1, 2, 3, 1, 2, 3, 4...
## $ amount      (dbl) 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ measure_desc (chr) "pat (1\" sq, 1/3\" high)", "tbsp", "cup", "stick..."
## $ gm_wgt      (dbl) 5.00, 14.20, 227.00, 113.00, 3.80, 9.40, 151.00, ...
## $ num_data_pts (int) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ std_dev     (dbl) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
```

Join the datasets and create algorithms

```
fastdiet <- food %>%
  full_join(food_group)%>%
  full_join(nutrient)%>%
  full_join(nutrient_def)%>%
  inner_join(weight) %>%
  #Select the variables for exploration
  select(food_id=nbd_no,fdgrp_id=fdgrp_cd,desc=shrt_desc,food_group=fdgrp_desc,nutrient_id=nutr_no,nutr...)
  #Filter food groups we don't care about
  filter(food_group!="Baby Foods",food_group!="American Indian/Alaska Native Foods",food_group!="Fast F...)
  #filter for only the nutrients we are exploring
  filter(nutrient_id==203|nutrient_id==204|nutrient_id==205|nutrient_id==208) %>%
  #create new variables
  mutate(nutrients_per_gram=nutrient_value/100,nutrients_per_serving=nutrients_per_gram*grams)

## Joining by: "fdgrp_cd"
## Joining by: "nbd_no"
## Joining by: "nutr_no"
## Joining by: c("nbd_no", "num_data_pts")
```

Data exploration

```

#Create theme for charts
#Basic
theme.diet_chart <-
  theme(legend.position = "none") +
  theme(plot.title = element_text(size=26, family="Trebuchet MS", face="bold", hjust=0, color="#666666") +
  theme(axis.title = element_text(size=18, family="Trebuchet MS", face="bold", color="#666666")) +
  theme(axis.title.y = element_text(angle=0))
#Scatterplot
theme.diet_chart_SCATTER <- theme.diet_chart +
  theme(axis.title.x = element_text(hjust=0, vjust=-.5))
#Histogram
theme.diet_chart_HIST <- theme.diet_chart +
  theme(axis.title.x = element_text(hjust=0, vjust=-.5))
#Small multiple
theme.diet_chart_SMALLM <- theme.diet_chart +
  theme(panel.grid.minor = element_blank()) +
  theme(strip.text.x = element_text(size=16, family="Trebuchet MS", face="bold"))

```

Basic exploration

Scatterplot

```
{r} #ggplot(data=fastdiet, aes(x=nutrient, y=nutrient_value))  
+ # geom_point(alpha=.4, size=4, color="#880011") + # ggtitle("FastDiet")  
+ # labs(x="Nutrient", y="Nutrient Value") + # theme.diet_chart_SCATTER  
#
```

Histogram

```
{r} #ggplot(data=fastdiet, aes(x=food_id)) + # geom_histogram(fill="#880011")  
+ # ggtitle("Foods") + # labs(x="Foods", y="Count of Foods") +  
# theme.diet_chart_HIST #
```

Barplot

```
{r} #fastdiet %>% # filter(nutrient_value >725 & nutrient_value  
<1090) %>% # ggplot(aes(x= as.factor(nutrient_value))) + #  
geom_bar(fill="#880011") + # labs(x="Nutrient Value") + # theme.diet_chart_BAR  
#
```

small multiples

```
{r} #ggplot(data=fastdiet, aes(x=nutrient_value)) + # geom_histogram(fill="#880011")  
+ # ggtitle("Histogram of Nutrients") + # labs(x="Nutrient  
Value", y="Count Records") + # facet_wrap(~nutrient) + # theme.diet_chart_MULTIPLE  
#
```