Appendix0_10

```
# NDNS analysis, data management -----------------------------------------

# Change the data path accordingly ---------------------------------------
setwd("/home/wangcc-me/Downloads/UKDA-6533-stata11_se/stata11_se/") # on Ubuntu
library(epiDisplay)
library(plyr)
library(tidyverse)

# Read the data into memory ----------------------------------------------
library(haven)
data <- read_dta("ndns_rp_yr1-4a_foodleveldietarydata_uk.dta")
data56 <- read_dta("ndns_rp_yr5-6a_foodleveldietarydata.dta")
data78 <- read_dta("ndns_rp_yr7-8a_foodleveldietarydata.dta")
names(data)
names(data56)
names(data78)
names(data)[names(data) == "seriali"] <- "id"
names(data56)[names(data56) == "seriali"] <- "id"
names(data78)[names(data78) == "seriali"] <- "id"

# Extract the data we needed ---------------------------------------------
df14d <- data[, c(113, 1, 2, 3, 5, 6, 7, 8, 9, 21, 24, 55, 57, 58,
    59, 60, 61, 62, 63, 64)]
var <- names(df14d)
df56d <- data56 %>% select(var)
df78d <- data78 %>% select(var)
dfs1 <- rbind(df14d, df56d, df78d)
dfs2 <- dfs1[dfs1$Age >= 19, ]
rm(data, data56, data78)
dfs2

# Calculate the time (minute and hour) when they eat ---------------------

dfs2$MealTime_chr <- as.character(dfs2$MealTime)
dfs2$MealTime_hm <- unlist(strsplit(dfs2$MealTime_chr, " "))[c(FALSE,
    TRUE)]
dfs2$MealHourN <- as.numeric(unlist(strsplit(dfs2$MealTime_hm, ":"))[c(TRUE,
    FALSE, FALSE)])
dfs2$MealMinN <- as.numeric(unlist(strsplit(dfs2$MealTime_hm, ":"))[c(FALSE,
    TRUE, FALSE)])
dfs2$MealMinN0 <- (60 * dfs2$MealHourN) + dfs2$MealMinN
dfs3 <- dfs2[order(dfs2$id, dfs2$DayNo, dfs2$MealMinN0), ]
length(unique(dfs3$id))  ## number of participants = 6155

# Create a subset data with only the first observation of each
# participant --------
NDNS <- dfs3[!duplicated(dfs3$id), ]
with(NDNS, tab1(SurveyYear, graph = FALSE, decimal = 2))

# #SurveyYear :
```

```
            # Frequency Percent Cum. percent
# NDNS Year 1       801   13.01        13.01
# NDNS Year 2       812   13.19        26.21
# NDNS Year 3       782   12.71        38.91
# NDNS Year 4      1055   17.14        56.05
# NDNS Year 5       625   10.15        66.21
# NDNS Year 6       663   10.77        76.98
# NDNS Year 7       703   11.42        88.40
# NDNS Year 8       714   11.60       100.00
  # Total          6155  100.00       100.00


# create a variable combine id and day No ---------------------------------
dfs3 <- dfs3 %>%
mutate(id_dy = paste(id, DayNo, sep = "D"))

# For each subject, the total energy/carbohydrate intake for each eating
# time can be calculated --------
old <- Sys.time()
Energy <- ddply(dfs3, .(id_dy, id, SurveyYear, DayNo, Age, Sex,
                        DiaryDaysCompleted, MealHourN, DayofWeek),
                summarise,
                Tot_Energ = sum(EnergykJ),
                Tot_Carb = sum(Carbohydrateg),
                Tot_Sugar = sum(Totalsugarsg),
                Tot_Starch = sum(Starchg))
new <- Sys.time() - old
print(new)
# Time difference of 3.876385 mins

rm(df14d, df56d, df78d, dfs2)

# Calculate the energy from total carbohydrates ---------------------------
Energy <- Energy %>%
  mutate(KJcarbo = Tot_Carb * 16) %>%
  mutate(CarKJpercentage = KJcarbo/Tot_Energ) %>%
  mutate(Carbo = cut(CarKJpercentage, breaks = c(0, 0.26, 0.75, 2),
        right = FALSE)) %>% mutate(Carbo2 = cut(CarKJpercentage, breaks = c(0,
    0.26, 2), right = FALSE))
Energy0 <- Energy[!(Energy$Tot_Energ == 0), ]
          # some food consumption does not contain any carbohydrates
Energy0$Carbo <- factor(Energy0$Carbo, labels = c("Low_carb", "Med_carb",
    "High_carb"))
Energy0$Carbo2 <- factor(Energy0$Carbo2, labels = c("Low_carb", "Med_or_high_carb"))

# Generate data sets for each day -----------------------------------------------------
dta_day1 <- Energy0 %>%
  filter(DayNo == 1) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))
```

```r
dta_day2 <- Energy0 %>%
  filter(DayNo == 2) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))

dta_day3 <- Energy0 %>%
  filter(DayNo == 3) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))

dta_day4 <- Energy0 %>%
  filter(DayNo == 4) %>%
  select(c("id", "Age",
    "Sex", "DayofWeek", "MealHourN", "Carbo", "Carbo2")) %>%
  mutate(DayofWeek = factor(DayofWeek,
    levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
        "Saturday", "Sunday")))
vecid1 <- unique(dta_day1$id) # n = 6153
vecid2 <- unique(dta_day2$id) # n = 6153
vecid3 <- unique(dta_day3$id) # n = 6151
vecid4 <- unique(dta_day4$id) # n = 6026

Noday1 <- setdiff(vecid, vecid1)  # two subjects did not have day 1 data
Noday2 <- setdiff(vecid, vecid2)  # two subjects did not have day 2 data
Noday3 <- setdiff(vecid, vecid3)  # four subjects did not have day 3 data
Noday4 <- setdiff(vecid, vecid4)  # 129 subjects did not have day 4 data

# Transform the data shape from long to wide ---------------------------------
dta_d1_wide <- dta_day1[, -7] %>%
  spread(key = MealHourN, value = Carbo)
names(dta_d1_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")

dta_d2_wide <- dta_day2[, -7] %>%
  spread(key = MealHourN, value = Carbo)
names(dta_d2_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")

dta_d3_wide <- dta_day3[, -7] %>%
  spread(key = MealHourN, value = Carbo)
names(dta_d3_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")

dta_d4_wide <- dta_day4[, -7] %>%
  spread(key = MealHourN, value = Carbo)
names(dta_d4_wide)[5:28] <- paste(rep("H", 24), 0:23, sep = "")

# recode NA to not eating ----------------------------------------------------
for (i in 5:ncol(dta_d1_wide))
  if (is.factor(dta_d1_wide[, i])) levels(dta_d1_wide[,
```

```
    i]) <- c(levels(dta_d1_wide[, i]), "Not_eating")

dta_d1_wide[is.na(dta_d1_wide)] <- "Not_eating"

for (i in 5:ncol(dta_d2_wide))
  if (is.factor(dta_d2_wide[, i])) levels(dta_d2_wide[,
    i]) <- c(levels(dta_d2_wide[, i]), "Not_eating")

dta_d2_wide[is.na(dta_d2_wide)] <- "Not_eating"

for (i in 5:ncol(dta_d3_wide))
  if (is.factor(dta_d3_wide[, i])) levels(dta_d3_wide[,
    i]) <- c(levels(dta_d3_wide[, i]), "Not_eating")

dta_d3_wide[is.na(dta_d3_wide)] <- "Not_eating"

for (i in 5:ncol(dta_d4_wide))
  if (is.factor(dta_d4_wide[, i])) levels(dta_d4_wide[,
    i]) <- c(levels(dta_d4_wide[, i]), "Not_eating")

dta_d4_wide[is.na(dta_d4_wide)] <- "Not_eating"
```