title: “Obesity prevalence in Ghana” author: “Vasco Ayere Avoka” date: “2024-06-16” output: html_document: theme: bootswatch: flatly date: “2024-06-16” —

R Markdown

The dataset used for this analysis is the 2012 Ghana Senior High School public use dataset. This dataset contains on young people’s health behavior and protective factors related to the leading causes of morbidity and mortality among children and adults worldwide. The dataset is saved in a CSV (Comma-Separated Values) format, which is a type of delimited flat file where the delimiter is a comma. CSV files are commonly opened with spreadsheet programs like Microsoft Excel or text editors.

```
# Read the dataset into R

data <- read_csv("C:/Users/New User/Downloads/2012GHA_seniorhigh_public_use.csv")
```

```
## Rows: 1984 Columns: 24
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## dbl (24): Age, Sex, Grade, Height, Weight, Fruits last 30 days, Vegetables l...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Rename columns for better readability
colnames(data) <- c("Age", "Sex", "Grade", "Height", "Weight", "Fruits_Last_30_Days",  "Veget
ables_Last_30_Days", "Fast_Food_Last_7_Days",
                           "Days_Active_60_Min_Plus_7_Past_Days", "Walk_or_Bike_to_School_Past
_7_Days",
                           "Days_Went_to_PE_Each_Week",
                           "Ate_Fruits_2+_Times_Per_Week_Past_30_Days", "Ate_Vegetables_3+_Tim
es_Per_Day_Last_30_Days",
                            "Ate_Fast_Food_3+_Times_Per_Day_Last_30_Days",
                            "Active_60+_For_More_Than_5_Days_Past_7_Days",
                           "Walk_Bike_0_Days_to_School_Last_7_Days", "3+_Days_PE_Each_Week",
"Overweight", "Obese",
                           "Underweight", "Ate_5+_Fruits_Vegetables_30_Days", "Physically_Acti
ve_All_7_Days",
                           "Attended_PE_5+_Each_Week")


# Convert relevant columns to numeric (if not already)
numeric_columns <- c("Age", "Sex", "Grade", "Height", "Weight", "Fruits_Last_30_Days",
                     "Vegetables_Last_30_Days", "Fast_Food_Last_7_Days",
                     "Days_Active_60_Min_Plus_7_Past_Days", "Walk_or_Bike_to_School_Past_7_Da
ys",
                     "Days_Went_to_PE_Each_Week",
                     "Ate_Fruits_2+_Times_Per_Week_Past_30_Days", "Ate_Vegetables_3+_Times_Pe
r_Day_Last_30_Days",
                      "Ate_Fast_Food_3+_Times_Per_Day_Last_30_Days",
                      "Active_60+_For_More_Than_5_Days_Past_7_Days",
                     "Walk_Bike_0_Days_to_School_Last_7_Days", "3+_Days_PE_Each_Week", "Overw
eight", "Obese",
                     "Underweight", "Ate_5+_Fruits_Vegetables_30_Days", "Physically_Active_Al
l_7_Days",
                     "Attended_PE_5+_Each_Week")


data[numeric_columns] <- lapply(data[numeric_columns], as.numeric)


# Display the data
head(data)
```

```
## # A tibble: 6 × 24
##     Age   Sex Grade Height Weight Fruits_Last_30_Days Vegetables_Last_30_Days
##   <dbl> <dbl> <dbl>  <dbl>  <dbl>               <dbl>                     <dbl>
## 1     6     1     4   1.66     59                   3                         4
## 2     5     1     4   1.69     60                   2                         7
## 3     5     1     4   1.71     65                   3                         2
## 4     6     1     4   1.73     50                   4                         4
## 5     3     2     4   1.71     79                   2                         3
## 6     4     2     4   1.52     47                   3                         3
## # ℹ 17 more variables: Fast_Food_Last_7_Days <dbl>,
## #   Days_Active_60_Min_Plus_7_Past_Days <dbl>,
## #   Walk_or_Bike_to_School_Past_7_Days <dbl>, Days_Went_to_PE_Each_Week <dbl>,
## #   `Ate_Fruits_2+_Times_Per_Week_Past_30_Days` <dbl>,
## #   `Ate_Vegetables_3+_Times_Per_Day_Last_30_Days` <dbl>,
## #   `Ate_Fast_Food_3+_Times_Per_Day_Last_30_Days` <dbl>,
## #   `Active_60+_For_More_Than_5_Days_Past_7_Days` <dbl>, …
```

```
ncol(data)
```

```
## [1] 24
```

```
nrow(data)
```

```
## [1] 1984
```

```r
# Create a table with column names and descriptions
library(knitr)

# Create a data frame with column descriptions
column_descriptions <- data.frame(
  Column_Name = c("Age", "Sex", "Grade", "Height", "Weight", "Fruits_Last_30_Days", "Vegetabl
es_Last_30_Days", "Fast_Food_Last_7_Days",
                  "Days_Active_60_Min_Plus_7_Past_Days", "Walk_or_Bike_to_School_Past_7_Day
s",
                  "Days_Went_to_PE_Each_Week", "Ate_Fruits_2+_Times_Per_Week_Past_30_Days",
                  "Ate_Vegetables_3+_Times_Per_Day_Last_30_Days", "Ate_Fast_Food_3+_Times_Per
_Day_Last_30_Days",
                  "Active_60+_For_More_Than_5_Days_Past_7_Days", "Walk_Bike_0_Days_to_School_
Last_7_Days",
                  "3+_Days_PE_Each_Week", "Overweight", "Obese", "Underweight", "Ate_5+_Fruit
s_Vegetables_30_Days",
                  "Physically_Active_All_7_Days", "Attended_PE_5+_Each_Week"),
  Description = c("Age of the student", "Sex of the student", "Grade of the student", "Height
of the student in cm",
                  "Weight of the student in kg", "Number of times the student ate fruits in t
he last 30 days",
                  "Number of times the student ate vegetables in the last 30 days",
                  "Number of times the student ate fast food in the last 7 days",
                  "Number of days the student was active for 60+ minutes in the past 7 days",
                  "Number of days the student walked or biked to school in the past 7 days",
                  "Number of days the student attended PE each week",
                  "Number of times the student ate fruits 2+ times per week in the past 30 da
ys",
                  "Number of times the student ate vegetables 3+ times per day in the last 30
days",
                  "Number of times the student ate fast food 3+ times per day in the last 30
days",
                  "Number of days the student was active for 60+ minutes for more than 5 days
in the past 7 days",
                  "Number of days the student walked or biked to school 0 days in the last 7
days",
                  "Number of days the student attended PE 3+ times each week", "Indicator if
the student is overweight",
                  "Indicator if the student is obese", "Indicator if the student is underweig
ht",
                  "Number of times the student ate 5+ fruits and vegetables in the last 30 da
ys",
                  "Indicator if the student was physically active all 7 days",
                  "Number of days the student attended PE 5+ times each week")
)

print(column_descriptions)
```

```
##                                    Column_Name
## 1                                          Age
## 2                                          Sex
## 3                                        Grade
## 4                                       Height
## 5                                       Weight
## 6                           Fruits_Last_30_Days
## 7                       Vegetables_Last_30_Days
## 8                         Fast_Food_Last_7_Days
## 9           Days_Active_60_Min_Plus_7_Past_Days
## 10           Walk_or_Bike_to_School_Past_7_Days
## 11                    Days_Went_to_PE_Each_Week
## 12    Ate_Fruits_2+_Times_Per_Week_Past_30_Days
## 13 Ate_Vegetables_3+_Times_Per_Day_Last_30_Days
## 14  Ate_Fast_Food_3+_Times_Per_Day_Last_30_Days
## 15  Active_60+_For_More_Than_5_Days_Past_7_Days
## 16       Walk_Bike_0_Days_to_School_Last_7_Days
## 17                          3+_Days_PE_Each_Week
## 18                                    Overweight
## 19                                         Obese
## 20                                   Underweight
## 21             Ate_5+_Fruits_Vegetables_30_Days
## 22                 Physically_Active_All_7_Days
## 23                     Attended_PE_5+_Each_Week
##                                                                                Descr
iption
## 1                                                                            Age of the s
tudent
## 2                                                                            Sex of the s
tudent
## 3                                                                          Grade of the s
tudent
## 4                                                                   Height of the student
in cm
## 5                                                                   Weight of the student
in kg
## 6                                     Number of times the student ate fruits in the last 3
0 days
## 7                                  Number of times the student ate vegetables in the last 3
0 days
## 8                                    Number of times the student ate fast food in the last
7 days
## 9                       Number of days the student was active for 60+ minutes in the past
7 days
## 10                       Number of days the student walked or biked to school in the past
7 days
## 11                                                Number of days the student attended PE eac
h week
## 12                   Number of times the student ate fruits 2+ times per week in the past 3
0 days
## 13               Number of times the student ate vegetables 3+ times per day in the last 3
0 days
## 14                Number of times the student ate fast food 3+ times per day in the last 3
0 days
## 15 Number of days the student was active for 60+ minutes for more than 5 days in the past
```

```
7 days
## 16                Number of days the student walked or biked to school 0 days in the last
7 days
## 17                                    Number of days the student attended PE 3+ times eac
h week
## 18                                                          Indicator if the student is over
weight
## 19                                                               Indicator if the student is
obese
## 20                                                         Indicator if the student is under
weight
## 21                  Number of times the student ate 5+ fruits and vegetables in the last 3
0 days
## 22                                    Indicator if the student was physically active all
7 days
## 23                                    Number of days the student attended PE 5+ times eac
h week
```

This data frame has 1984 rows and 24 columns. The names of the columns and a brief description of each are in the table below:

```
# Display the table
kable(column_descriptions)
```

Column_Name	Description
Age	Age of the student
Sex	Sex of the student
Grade	Grade of the student
Height	Height of the student in cm
Weight	Weight of the student in kg
Fruits_Last_30_Days	Number of times the student ate fruits in the last 30 days
Vegetables_Last_30_Days	Number of times the student ate vegetables in the last 30 days
Fast_Food_Last_7_Days	Number of times the student ate fast food in the last 7 days
Days_Active_60_Min_Plus_7_Past_Days	Number of days the student was active for 60+ minutes in the past 7 days
Walk_or_Bike_to_School_Past_7_Days	Number of days the student walked or biked to school in the past 7 days
Days_Went_to_PE_Each_Week	Number of days the student attended PE each week
Ate_Fruits_2+_Times_Per_Week_Past_30_Days	Number of times the student ate fruits 2+ times per week in the past 30 days
Ate_Vegetables_3+_Times_Per_Day_Last_30_Days	Number of times the student ate vegetables 3+ times per day in the last 30 days

Column_Name	Description
Ate_Fast_Food_3+_Times_Per_Day_Last_30_Days	Number of times the student ate fast food 3+ times per day in the last 30 days
Active_60+_For_More_Than_5_Days_Past_7_Days	Number of days the student was active for 60+ minutes for more than 5 days in the past 7 days
Walk_Bike_0_Days_to_School_Last_7_Days	Number of days the student walked or biked to school 0 days in the last 7 days
3+_Days_PE_Each_Week	Number of days the student attended PE 3+ times each week
Overweight	Indicator if the student is overweight
Obese	Indicator if the student is obese
Underweight	Indicator if the student is underweight
Ate_5+_Fruits_Vegetables_30_Days	Number of times the student ate 5+ fruits and vegetables in the last 30 days
Physically_Active_All_7_Days	Indicator if the student was physically active all 7 days
Attended_PE_5+_Each_Week	Number of days the student attended PE 5+ times each week

```r
library(dplyr)

library(dplyr)

# Clean column names
names(data) <- make.names(names(data))

# Check the structure of your dataframe
str(data)
```

```
## spc_tbl_ [1,984 × 24] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Age                                          : num [1:1984] 6 5 5 6 3 4 5 5 5 5 ...
##  $ Sex                                          : num [1:1984] 1 1 1 1 2 2 2 2 2 2 ...
##  $ Grade                                        : num [1:1984] 4 4 4 4 4 4 4 4 4 4 ...
##  $ Height                                       : num [1:1984] 1.66 1.69 1.71 1.73 1.71 1.5
2 1.62 1.6 1.52 1.64 ...
##  $ Weight                                       : num [1:1984] 59 60 65 50 79 47 51 50 44 5
9 ...
##  $ Fruits_Last_30_Days                          : num [1:1984] 3 2 3 4 2 3 3 2 2 2 ...
##  $ Vegetables_Last_30_Days                      : num [1:1984] 4 7 2 4 3 3 4 3 4 4 ...
##  $ Fast_Food_Last_7_Days                        : num [1:1984] 7 1 1 2 1 2 3 1 3 6 ...
##  $ Days_Active_60_Min_Plus_7_Past_Days          : num [1:1984] 4 1 6 7 1 8 4 4 8 5 ...
##  $ Walk_or_Bike_to_School_Past_7_Days           : num [1:1984] 8 1 2 1 1 8 8 6 1 6 ...
##  $ Days_Went_to_PE_Each_Week                    : num [1:1984] 6 1 1 1 1 2 3 1 1 NA ...
##  $ Ate_Fruits_2._Times_Per_Week_Past_30_Days    : num [1:1984] 2 2 2 2 2 2 2 2 2 2 ...
##  $ Ate_Vegetables_3._Times_Per_Day_Last_30_Days: num [1:1984] 2 2 2 1 2 2 2 2 2 2 ...
##  $ Ate_Fast_Food_3._Times_Per_Day_Last_30_Days : num [1:1984] 2 1 2 2 2 2 2 2 2 2 ...
##  $ Active_60._For_More_Than_5_Days_Past_7_Days : num [1:1984] 1 2 2 2 2 2 2 2 2 1 ...
##  $ Walk_Bike_0_Days_to_School_Last_7_Days       : num [1:1984] 2 2 1 1 2 1 2 2 1 2 ...
##  $ X3._Days_PE_Each_Week                        : num [1:1984] 2 1 2 1 1 2 2 2 1 2 ...
##  $ Overweight                                   : num [1:1984] 1 2 2 2 2 2 2 2 2 NA ...
##  $ Obese                                        : num [1:1984] 2 2 2 2 1 2 2 2 2 2 ...
##  $ Underweight                                  : num [1:1984] 2 2 2 2 1 2 2 2 2 2 ...
##  $ Ate_5._Fruits_Vegetables_30_Days             : num [1:1984] 2 2 2 2 2 2 2 2 2 2 ...
##  $ Physically_Active_All_7_Days                 : num [1:1984] 2 1 2 2 2 2 2 2 2 2 ...
##  $ Attended_PE_5._Each_Week                     : num [1:1984] 2 2 2 2 2 1 2 2 1 2 ...
##  $ NA.                                          : num [1:1984] 1 2 2 2 2 2 2 2 2 NA ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Age = col_double(),
##   ..   Sex = col_double(),
##   ..   Grade = col_double(),
##   ..   Height = col_double(),
##   ..   Weight = col_double(),
##   ..   `Fruits last 30 days` = col_double(),
##   ..   `Vegetables last 30 days` = col_double(),
##   ..   `fast food last 7 days` = col_double(),
##   ..   `days active 6o mins plus 7 past days` = col_double(),
##   ..   `walk or bike to school past 7 days` = col_double(),
##   ..   `days went to PE each week` = col_double(),
##   ..   `always hungry last 30 days` = col_double(),
##   ..   `ate fruits 2+ times per week past 30 days` = col_double(),
##   ..   `ate vegetabes 3+ times per day last 30 days` = col_double(),
##   ..   `ate fast food 3+ times per day last 30 days` = col_double(),
##   ..   `active 60+ for more than 5 days for past 7 days` = col_double(),
##   ..   `walk/bike 0 days to school last 7 days` = col_double(),
##   ..   `3+ days PE each week` = col_double(),
##   ..   overweight = col_double(),
##   ..   obese = col_double(),
##   ..   underweight = col_double(),
##   ..   `ate 5+ fruits/vegetables 30 days` = col_double(),
##   ..   `physically active all 7 days` = col_double(),
##   ..   `attended PE 5+ each week` = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```r
# Convert columns to character first to ensure na_if works correctly
data <- data %>%
  mutate(
    Age = as.character(Age),
    Height = as.character(Height),
    Weight = as.character(Weight)
  )

# Replace empty strings with NA
data <- data %>%
  mutate(
    Age = na_if(Age, ""),
    Height = na_if(Height, ""),
    Weight = na_if(Weight, "")
  )

# Convert columns to numeric (if they are not already)
data <- data %>%
  mutate(
    Age = as.numeric(Age),
    Height = as.numeric(Height),
    Weight = as.numeric(Weight)
  )

# Calculate summaries
summary_stats <- data %>%
  summarise(
    min_Age = min(Age, na.rm = TRUE),
    max_Age = max(Age, na.rm = TRUE),
    mean_age = mean(Age, na.rm = TRUE),
    min_height = min(Height, na.rm = TRUE),
    max_height = max(Height, na.rm = TRUE),
    mean_height = mean(Height, na.rm = TRUE),
    min_weight = min(Weight, na.rm = TRUE),
    max_weight = max(Weight, na.rm = TRUE),
    mean_weight = mean(Weight, na.rm = TRUE)
  )

print(summary_stats)
```

```
## # A tibble: 1 × 9
##   min_Age max_Age mean_age min_height max_height mean_height min_weight
##     <dbl>   <dbl>    <dbl>      <dbl>      <dbl>       <dbl>      <dbl>
## 1       2       8     7.15       1.36       1.94        1.65         32
## # ℹ 2 more variables: max_weight <dbl>, mean_weight <dbl>
```