Predictors of Weight in Romanian Infants: Weight-for-age z-scores (WAZ)

Adrian Rus

2023-10-28

1. Research Hypotheses/Questions/Objectives

The objective of the present study was to explore the relationship between infants' weight and various demographics (i.e., infants' mother location, marital status, social economic status, and age; and infants being born on term, age, anemia status, rank into the family and dietary patterns).

Identifying the connection between such factors could help practitioners in developing public health policies and interventions in Romania.

2. Data Exploration

2.1. Setting the Working Directory

Check the working directory

```
getwd()
```

```
## [1] "C:/Users/adria/Documents/waz_infants/script"
```

2.2. For setting a new working directory use the `setwd()`

```
setwd("C:\\Users\\adria\\Documents\\waz_infants\\data")
```

List Files in Directory To see a list of all files in the current working directory, you can use

```
list.files()
```

```
## [1] "waz_infants_script.R"
## [2] "waz_infants_script_final_2023-7-21.R"
## [3] "waz_infants_script_markdown.html"
## [4] "waz_infants_script_markdown.Rmd"
## [5] "waz_iv_levels.R"
```

2.3. Import and check the dataset

```
library(readr)
waz_final <- read_csv("C:/Users/adria/Documents/waz_infants/data/waz_final.csv")
```

```
## Rows: 1532 Columns: 38
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (21): cMDD, cMMF, cMAD, Anemia2, Anemia4, Location, County, cAgeM3c, cGe...
## dbl (17): cID, Hemoglobin, cAgeM, cRank.numerical, mAge, cWeightB, cWeightN,...
##
## 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.
```

```
View(waz_final)
```

2.3.1. Basic inspection of dataset

```
# View the first 6 rows
head(waz_final)
```

```
## # A tibble: 6 × 38
##   cMDD  cMMF       cMAD     cID Hemoglobin Anemia2 Anemia4 Location County cAgeM
##   <chr> <chr>      <chr>  <dbl>      <dbl> <chr>   <chr>   <chr>    <chr>  <dbl>
## 1 No    Aqequate   Inade…     2       11   No (11… No ane… Rural    BZ     11.7
## 2 No    Aqequate   Inade…     8        8.1 Yes (<… Modera… Rural    BZ     11.2
## 3 No    Inadequate Inade…     9       11.5 No (11… No ane… Rural    BZ     11.9
## 4 No    Aqequate   Inade…    10       12.2 No (11… No ane… Rural    BZ     10.5
## 5 No    Aqequate   Inade…    11       11.6 No (11… No ane… Rural    BZ     10.4
## 6 No    Aqequate   Inade…    16       13   No (11… No ane… Urban    BZ      6.13
## # ℹ 28 more variables: cAgeM3c <chr>, cGender <chr>, cRank.numerical <dbl>,
## #   cRank3c <chr>, cRank <chr>, cBirthT <chr>, mAge <dbl>, mAge2c <chr>,
## #   mSES3c <chr>, mMarital <chr>, mMarital2 <chr>, mEdu3c <chr>, mEthnic <chr>,
## #   mEthnic2 <chr>, mSmoke <chr>, cWeightB <dbl>, cWeightN <dbl>,
## #   cHeightN <dbl>, cHageZ <dbl>, cHageZ.stunting <chr>, cWageZ <dbl>,
## #   cWHZ <dbl>, cBMIage <dbl>, cBMIageZ <dbl>, cWageZ_no_outliers <dbl>,
## #   cWageZ_no_outliers_robust <dbl>, mAge_no_outliers <dbl>, …
```

```
# View the last 6 rows
tail(waz_final)
```

```
## # A tibble: 6 × 38
##   cMDD  cMMF       cMAD     cID Hemoglobin Anemia2 Anemia4 Location County cAgeM
##   <chr> <chr>      <chr>  <dbl>      <dbl> <chr>   <chr>   <chr>    <chr>  <dbl>
## 1 Yes   Aqequate   Inade…  1418       11.1 No (11… No ane… Rural    DJ        12
## 2 Yes   Inadequate Inade…  1564       10.9 Yes (<… Mild a… Urban    HD        12
## 3 No    Inadequate Inade…  1693       11.1 No (11… No ane… Urban    GJ        12
## 4 Yes   Aqequate   Aqequ…  1913       10.1 Yes (<… Mild a… Rural    CT        12
## 5 No    Aqequate   Inade…  1934       11   No (11… No ane… Urban    CT        12
## 6 Yes   Aqequate   Aqequ…  2124       12.1 No (11… No ane… Urban    B         12
## # ℹ 28 more variables: cAgeM3c <chr>, cGender <chr>, cRank.numerical <dbl>,
## #   cRank3c <chr>, cRank <chr>, cBirthT <chr>, mAge <dbl>, mAge2c <chr>,
## #   mSES3c <chr>, mMarital <chr>, mMarital2 <chr>, mEdu3c <chr>, mEthnic <chr>,
## #   mEthnic2 <chr>, mSmoke <chr>, cWeightB <dbl>, cWeightN <dbl>,
## #   cHeightN <dbl>, cHageZ <dbl>, cHageZ.stunting <chr>, cWageZ <dbl>,
## #   cWHZ <dbl>, cBMIage <dbl>, cBMIageZ <dbl>, cWageZ_no_outliers <dbl>,
## #   cWageZ_no_outliers_robust <dbl>, mAge_no_outliers <dbl>, …
```

```
# View the structure of dataset
str(waz_final)
```

```
## spc_tbl_ [1,532 × 38] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ cMDD                     : chr [1:1532] "No" "No" "No" "No" ...
##  $ cMMF                     : chr [1:1532] "Aqequate" "Aqequate" "Inadequate" "Aqequate" ...
##  $ cMAD                     : chr [1:1532] "Inadequate" "Inadequate" "Inadequate" "Inadequat
e" ...
##  $ cID                      : num [1:1532] 2 8 9 10 11 16 17 19 20 21 ...
##  $ Hemoglobin               : num [1:1532] 11 8.1 11.5 12.2 11.6 13 11.6 9.2 12.8 12.1 ...
##  $ Anemia2                  : chr [1:1532] "No (11> g/dl)" "Yes (<11 g/dl)" "No (11> g/dl)"
"No (11> g/dl)" ...
##  $ Anemia4                  : chr [1:1532] "No anemia (11> g/dl)" "Moderate anemia (7-8.9 g/d
l)" "No anemia (11> g/dl)" "No anemia (11> g/dl)" ...
##  $ Location                 : chr [1:1532] "Rural" "Rural" "Rural" "Rural" ...
##  $ County                   : chr [1:1532] "BZ" "BZ" "BZ" "BZ" ...
##  $ cAgeM                    : num [1:1532] 11.7 11.2 11.9 10.5 10.4 ...
##  $ cAgeM3c                  : chr [1:1532] "6-11 months" "6-11 months" "6-11 months" "6-11 mo
nths" ...
##  $ cGender                  : chr [1:1532] "Female" "Female" "Female" "Male" ...
##  $ cRank.numerical          : num [1:1532] 3 1 2 1 1 1 1 2 2 1 ...
##  $ cRank3c                  : chr [1:1532] "Third and up" "First" "Second" "First" ...
##  $ cRank                    : chr [1:1532] "Third" "First" "Second" "First" ...
##  $ cBirthT                  : chr [1:1532] "At term (37-40 weeks)" "At term (37-40 weeks)" "B
efore term (27-36 weeks)" "At term (37-40 weeks)" ...
##  $ mAge                     : num [1:1532] 34 22 26 19 42 18 26 33 25 28 ...
##  $ mAge2c                   : chr [1:1532] "30-53 years" "15-29 years" "15-29 years" "15-29 y
ears" ...
##  $ mSES3c                   : chr [1:1532] "Medium" "Low" "Medium" "Medium" ...
##  $ mMarital                 : chr [1:1532] "Married" "Married" "Cohabitation" "Married" ...
##  $ mMarital2                : chr [1:1532] "Married" "Married" "Other" "Married" ...
##  $ mEdu3c                   : chr [1:1532] "No school/Gymnasium (finished or not)" "No schoo
l/Gymnasium (finished or not)" "No school/Gymnasium (finished or not)" "High/Technical School"
...
##  $ mEthnic                  : chr [1:1532] "Romanian" "Romanian" "Romanian" "Romanian" ...
##  $ mEthnic2                 : chr [1:1532] "Romanian" "Romanian" "Romanian" "Romanian" ...
##  $ mSmoke                   : chr [1:1532] "Not" "Not" "Not" "Not" ...
##  $ cWeightB                 : num [1:1532] 2200 3000 2500 3700 2150 3500 4050 3200 1700 2870
...
##  $ cWeightN                 : num [1:1532] 9 11 7 8.7 7.38 6.5 9.4 11 9.1 11 ...
##  $ cHeightN                 : num [1:1532] 80 89 70 73 71 68.8 74 77.3 72.5 77 ...
##  $ cHageZ                   : num [1:1532] 2.7591 6.5851 -1.2244 -0.0832 -0.9331 ...
##  $ cHageZ.stunting          : chr [1:1532] "No stunted" "No stunted" "No stunted" "No stunte
d" ...
##  $ cWageZ                   : num [1:1532] 0.11 1.813 -2.022 -0.596 -2.081 ...
##  $ cWHZ                     : num [1:1532] -1.474 -1.375 -1.951 -0.728 -2.206 ...
##  $ cBMIage                  : num [1:1532] 13.8 13.7 14 16 14.4 ...
##  $ cBMIageZ                 : num [1:1532] -2 -2.17 -1.83 -0.74 -2.16 ...
##  $ cWageZ_no_outliers       : num [1:1532] 0.11 1.813 -2.022 -0.596 -2.081 ...
##  $ cWageZ_no_outliers_robust: num [1:1532] 0.11 1.813 -2.022 -0.596 -2.081 ...
##  $ mAge_no_outliers         : num [1:1532] 34 22 26 19 42 18 26 33 25 28 ...
##  $ cHageZ_no_outliers       : num [1:1532] 2.7591 NA -1.2244 -0.0832 -0.9331 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   cMDD = col_character(),
```

```
##   ..   cMMF = col_character(),
##   ..   cMAD = col_character(),
##   ..   cID = col_double(),
##   ..   Hemoglobin = col_double(),
##   ..   Anemia2 = col_character(),
##   ..   Anemia4 = col_character(),
##   ..   Location = col_character(),
##   ..   County = col_character(),
##   ..   cAgeM = col_double(),
##   ..   cAgeM3c = col_character(),
##   ..   cGender = col_character(),
##   ..   cRank.numerical = col_double(),
##   ..   cRank3c = col_character(),
##   ..   cRank = col_character(),
##   ..   cBirthT = col_character(),
##   ..   mAge = col_double(),
##   ..   mAge2c = col_character(),
##   ..   mSES3c = col_character(),
##   ..   mMarital = col_character(),
##   ..   mMarital2 = col_character(),
##   ..   mEdu3c = col_character(),
##   ..   mEthnic = col_character(),
##   ..   mEthnic2 = col_character(),
##   ..   mSmoke = col_character(),
##   ..   cWeightB = col_double(),
##   ..   cWeightN = col_double(),
##   ..   cHeightN = col_double(),
##   ..   cHageZ = col_double(),
##   ..   cHageZ.stunting = col_character(),
##   ..   cWageZ = col_double(),
##   ..   cWHZ = col_double(),
##   ..   cBMIage = col_double(),
##   ..   cBMIageZ = col_double(),
##   ..   cWageZ_no_outliers = col_double(),
##   ..   cWageZ_no_outliers_robust = col_double(),
##   ..   mAge_no_outliers = col_double(),
##   ..   cHageZ_no_outliers = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
# Summarize the dataset
summary(waz_final)
```

```
##      cMDD               cMMF               cMAD                cID
##  Length:1532        Length:1532        Length:1532        Min.   :   2.0
##  Class :character   Class :character   Class :character   1st Qu.: 562.8
##  Mode  :character   Mode  :character   Mode  :character   Median :1081.5
##                                                           Mean   :1076.2
##                                                           3rd Qu.:1589.2
##                                                           Max.   :2127.0
##
##    Hemoglobin      Anemia2            Anemia4            Location
##  Min.   : 4.90   Length:1532        Length:1532        Length:1532
##  1st Qu.:10.20   Class :character   Class :character   Class :character
##  Median :11.10   Mode  :character   Mode  :character   Mode  :character
##  Mean   :10.97
##  3rd Qu.:11.80
##  Max.   :14.40
##
##     County              cAgeM         cAgeM3c            cGender
##  Length:1532        Min.   : 6.00   Length:1532        Length:1532
##  Class :character   1st Qu.:10.10   Class :character   Class :character
##  Mode  :character   Median :13.20   Mode  :character   Mode  :character
##                     Mean   :14.26
##                     3rd Qu.:18.60
##                     Max.   :23.97
##
##  cRank.numerical    cRank3c             cRank             cBirthT
##  Min.   : 1.000   Length:1532        Length:1532        Length:1532
##  1st Qu.: 1.000   Class :character   Class :character   Class :character
##  Median : 1.000   Mode  :character   Mode  :character   Mode  :character
##  Mean   : 1.683
##  3rd Qu.: 2.000
##  Max.   :11.000
##
##       mAge          mAge2c             mSES3c            mMarital
##  Min.   :15.00   Length:1532        Length:1532        Length:1532
##  1st Qu.:24.00   Class :character   Class :character   Class :character
##  Median :28.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   :28.02
##  3rd Qu.:32.00
##  Max.   :53.00
##
##   mMarital2            mEdu3c            mEthnic            mEthnic2
##  Length:1532        Length:1532        Length:1532        Length:1532
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     mSmoke             cWeightB       cWeightN         cHeightN
##  Length:1532        Min.   : 850   Min.   : 4.180   Min.   : 53.00
##  Class :character   1st Qu.:2900   1st Qu.: 8.535   1st Qu.: 72.00
##  Mode  :character   Median :3250   Median : 9.850   Median : 76.50
```

```
##                    Mean   :3222   Mean   : 9.883   Mean   : 76.79
##                    3rd Qu.:3600   3rd Qu.:11.000   3rd Qu.: 82.00
##                    Max.   :5260   Max.   :18.000   Max.   :100.00
##
##      cHageZ        cHageZ.stunting        cWageZ              cWHZ
##  Min.   :-7.6635   Length:1532        Min.   :-3.20207   Min.   :-5.15111
##  1st Qu.:-0.7407   Class :character   1st Qu.:-0.68796   1st Qu.:-0.81679
##  Median : 0.1535   Mode  :character   Median : 0.03308   Median :-0.07995
##  Mean   : 0.1609                      Mean   : 0.02086   Mean   :-0.03869
##  3rd Qu.: 1.1242                      3rd Qu.: 0.76291   3rd Qu.: 0.76978
##  Max.   :10.8441                      Max.   : 3.23849   Max.   : 5.44326
##                                       NA's   :17
##     cBMIage         cBMIageZ        cWageZ_no_outliers
##  Min.   :10.01   Min.   :-6.24525   Min.   :-2.84295
##  1st Qu.:15.17   1st Qu.:-0.94129   1st Qu.:-0.68314
##  Median :16.23   Median :-0.10064   Median : 0.03308
##  Mean   :16.41   Mean   :-0.07798   Mean   : 0.02324
##  3rd Qu.:17.49   3rd Qu.: 0.79188   3rd Qu.: 0.75837
##  Max.   :25.51   Max.   : 5.53067   Max.   : 2.90917
##                                     NA's   :30
##  cWageZ_no_outliers_robust mAge_no_outliers cHageZ_no_outliers
##  Min.   :-3.20207          Min.   :15.00    Min.   :-3.3959
##  1st Qu.:-0.68796          1st Qu.:24.00    1st Qu.:-0.7126
##  Median : 0.03308          Median :28.00    Median : 0.1635
##  Mean   : 0.02086          Mean   :27.99    Mean   : 0.1745
##  3rd Qu.: 0.76291          3rd Qu.:32.00    3rd Qu.: 1.0939
##  Max.   : 3.23849          Max.   :43.00    Max.   : 3.7662
##  NA's   :17                NA's   :2        NA's   :52
```

2.3.2. Check the tidiness of the dataset

```
# Checking the missing values.
# Tidy data should have a consistent structure, but it can contain missing values.
sum(is.na(waz_final))
```

```
## [1] 119
```

```
# Tidy data should ideally not have duplicated rows unless the repetition is meaningful.
sum(duplicated(waz_final))
```

```
## [1] 0
```

```
# Column names should be clear and descriptive but not too lengthy.
names(waz_final)
```

```
##  [1] "cMDD"                    "cMMF"
##  [3] "cMAD"                    "cID"
##  [5] "Hemoglobin"              "Anemia2"
##  [7] "Anemia4"                 "Location"
##  [9] "County"                  "cAgeM"
## [11] "cAgeM3c"                 "cGender"
## [13] "cRank.numerical"         "cRank3c"
## [15] "cRank"                   "cBirthT"
## [17] "mAge"                    "mAge2c"
## [19] "mSES3c"                  "mMarital"
## [21] "mMarital2"               "mEdu3c"
## [23] "mEthnic"                 "mEthnic2"
## [25] "mSmoke"                  "cWeightB"
## [27] "cWeightN"                "cHeightN"
## [29] "cHageZ"                  "cHageZ.stunting"
## [31] "cWageZ"                  "cWHZ"
## [33] "cBMIage"                 "cBMIageZ"
## [35] "cWageZ_no_outliers"      "cWageZ_no_outliers_robust"
## [37] "mAge_no_outliers"        "cHageZ_no_outliers"
```

2.3.3. Check the data types and dimensionality

```
# Tells you whether the object is a data frame, matrix, vector, list, etc.
class(waz_final)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```

```
# Returns the class (type) of each column in the data frame.
sapply(waz_final, class)
```

```
##                    cMDD                   cMMF                   cMAD
##             "character"            "character"            "character"
##                     cID             Hemoglobin                Anemia2
##               "numeric"              "numeric"            "character"
##                Anemia4               Location                 County
##             "character"            "character"            "character"
##                   cAgeM                cAgeM3c                cGender
##               "numeric"            "character"            "character"
##          cRank.numerical                cRank3c                  cRank
##               "numeric"            "character"            "character"
##                cBirthT                   mAge                  mAge2c
##             "character"              "numeric"            "character"
##                 mSES3c                mMarital                mMarital2
##             "character"            "character"            "character"
##                 mEdu3c                mEthnic                mEthnic2
##             "character"            "character"            "character"
##                 mSmoke                cWeightB                cWeightN
##             "character"              "numeric"              "numeric"
##                cHeightN                   cHageZ          cHageZ.stunting
##               "numeric"              "numeric"            "character"
##                  cWageZ                   cWHZ                 cBMIage
##               "numeric"              "numeric"              "numeric"
##                cBMIageZ       cWageZ_no_outliers cWageZ_no_outliers_robust
##               "numeric"              "numeric"              "numeric"
##         mAge_no_outliers       cHageZ_no_outliers
##               "numeric"              "numeric"
```

```
# Returns the number of columns.
ncol(waz_final)
```

```
## [1] 38
```

```
# Returns the number of rows and columns in a data frame.
dim(waz_final)
```

```
## [1] 1532   38
```

```
# Select only numerical columns

numerical_columns <- waz_final[, sapply(waz_final, is.numeric)]

head(numerical_columns) # Show the first few rows of the numerical columns
```

```
## # A tibble: 6 × 17
##     cID Hemoglobin cAgeM cRank.numerical  mAge cWeightB cWeightN cHeightN
##   <dbl>      <dbl> <dbl>           <dbl> <dbl>    <dbl>    <dbl>    <dbl>
## 1     2       11   11.7                3    34     2200     9        80
## 2     8        8.1 11.2                1    22     3000    11        89
## 3     9       11.5 11.9                2    26     2500     7        70
## 4    10       12.2 10.5                1    19     3700     8.7      73
## 5    11       11.6 10.4                1    42     2150     7.38     71
## 6    16       13    6.13               1    18     3500     6.5      68.8
## # ℹ 9 more variables: cHageZ <dbl>, cWageZ <dbl>, cWHZ <dbl>, cBMIage <dbl>,
## #   cBMIageZ <dbl>, cWageZ_no_outliers <dbl>, cWageZ_no_outliers_robust <dbl>,
## #   mAge_no_outliers <dbl>, cHageZ_no_outliers <dbl>
```

```
View(numerical_columns) # It shows all the variables
```

```
# Select only categorical columns (factor or character)
categorical_columns <- waz_final[, sapply(waz_final, function(col) is.factor(col) || is.characte
r(col))]
# Show the first few rows of the categorical columns
head(categorical_columns)
```

```
## # A tibble: 6 × 21
##   cMDD  cMMF       cMAD  Anemia2 Anemia4 Location County cAgeM3c cGender cRank3c
##   <chr> <chr>      <chr> <chr>   <chr>   <chr>    <chr>  <chr>   <chr>   <chr>
## 1 No    Aqequate   Inad… No (11… No ane… Rural    BZ     6-11 m… Female  Third …
## 2 No    Aqequate   Inad… Yes (<… Modera… Rural    BZ     6-11 m… Female  First
## 3 No    Inadequate Inad… No (11… No ane… Rural    BZ     6-11 m… Female  Second
## 4 No    Aqequate   Inad… No (11… No ane… Rural    BZ     6-11 m… Male    First
## 5 No    Aqequate   Inad… No (11… No ane… Rural    BZ     6-11 m… Male    First
## 6 No    Aqequate   Inad… No (11… No ane… Urban    BZ     6-11 m… Female  First
## # ℹ 11 more variables: cRank <chr>, cBirthT <chr>, mAge2c <chr>, mSES3c <chr>,
## #   mMarital <chr>, mMarital2 <chr>, mEdu3c <chr>, mEthnic <chr>,
## #   mEthnic2 <chr>, mSmoke <chr>, cHageZ.stunting <chr>
```

2.3.4. Specific variable content

```
# Shows unique values in a specific column.

unique(waz_final$mAge2c)
```

```
## [1] "30-53 years" "15-29 years"
```

```
# Allows to identify the unique data in all columns

unique_values_all_columns <- lapply(waz_final, unique)

str(unique_values_all_columns)
```

```
## List of 38
##  $ cMDD                    : chr [1:2] "No" "Yes"
##  $ cMMF                    : chr [1:2] "Aqequate" "Inadequate"
##  $ cMAD                    : chr [1:2] "Inadequate" "Aqequate"
##  $ cID                     : num [1:1532] 2 8 9 10 11 16 17 19 20 21 ...
##  $ Hemoglobin              : num [1:78] 11 8.1 11.5 12.2 11.6 13 9.2 12.8 12.1 10.4 ...
##  $ Anemia2                 : chr [1:2] "No (11> g/dl)" "Yes (<11 g/dl)"
##  $ Anemia4                 : chr [1:4] "No anemia (11> g/dl)" "Moderate anemia (7-8.9 g/dl)"
"Mild anemia (9-11 g/dl)" "Severe anemia (<6.9 g/dl)"
##  $ Location                : chr [1:2] "Rural" "Urban"
##  $ County                  : chr [1:16] "BZ" "BT" "SM" "IF" ...
##  $ cAgeM                   : num [1:391] 11.7 11.2 11.9 10.5 10.4 ...
##  $ cAgeM3c                 : chr [1:3] "6-11 months" "18-23 months" "12-17 months"
##  $ cGender                 : chr [1:2] "Female" "Male"
##  $ cRank.numerical         : num [1:10] 3 1 2 4 8 6 5 9 7 11
##  $ cRank3c                 : chr [1:3] "Third and up" "First" "Second"
##  $ cRank                   : chr [1:10] "Third" "First" "Second" "Forth" ...
##  $ cBirthT                 : chr [1:2] "At term (37-40 weeks)" "Before term (27-36 weeks)"
##  $ mAge                    : num [1:32] 34 22 26 19 42 18 33 25 28 35 ...
##  $ mAge2c                  : chr [1:2] "30-53 years" "15-29 years"
##  $ mSES3c                  : chr [1:3] "Medium" "Low" "High"
##  $ mMarital                : chr [1:4] "Married" "Cohabitation" "Unmarried" "Divorced/Separa
ted/Widow"
##  $ mMarital2               : chr [1:2] "Married" "Other"
##  $ mEdu3c                  : chr [1:3] "No school/Gymnasium (finished or not)" "High/Technic
al School" "College"
##  $ mEthnic                 : chr [1:4] "Romanian" "Rroma" "Hungarian" "German"
##  $ mEthnic2                : chr [1:2] "Romanian" "Other"
##  $ mSmoke                  : chr [1:3] "Not" "Yes" NA
##  $ cWeightB                : num [1:172] 2200 3000 2500 3700 2150 3500 4050 3200 1700 2870
...
##  $ cWeightN                : num [1:218] 9 11 7 8.7 7.38 6.5 9.4 9.1 10.5 10 ...
##  $ cHeightN                : num [1:134] 80 89 70 73 71 68.8 74 77.3 72.5 77 ...
##  $ cHageZ                  : num [1:1398] 2.7591 6.5851 -1.2244 -0.0832 -0.9331 ...
##  $ cHageZ.stunting         : chr [1:2] "No stunted" "Stunted"
##  $ cWageZ                  : num [1:1443] 0.11 1.813 -2.022 -0.596 -2.081 ...
##  $ cWHZ                    : num [1:1114] -1.474 -1.375 -1.951 -0.728 -2.206 ...
##  $ cBMIage                 : num [1:961] 13.8 13.7 14 16 14.4 ...
##  $ cBMIageZ                : num [1:1510] -2 -2.17 -1.83 -0.74 -2.16 ...
##  $ cWageZ_no_outliers      : num [1:1430] 0.11 1.813 -2.022 -0.596 -2.081 ...
##  $ cWageZ_no_outliers_robust: num [1:1443] 0.11 1.813 -2.022 -0.596 -2.081 ...
##  $ mAge_no_outliers        : num [1:31] 34 22 26 19 42 18 33 25 28 35 ...
##  $ cHageZ_no_outliers      : num [1:1349] 2.7591 NA -1.2244 -0.0832 -0.9331 ...
```

3. Dependent Variable (DV) and Data Processing

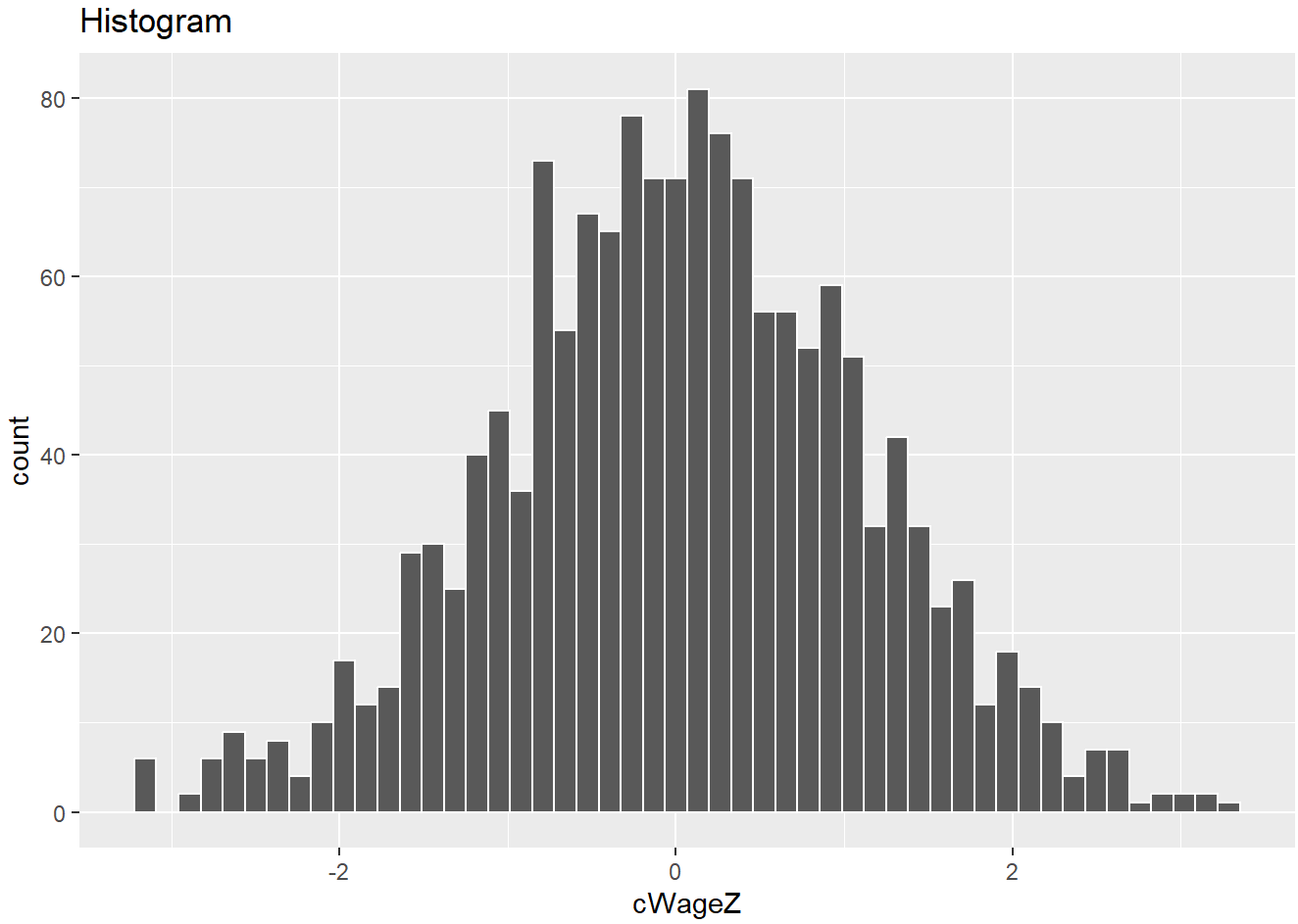
Dependent variable (DV): cWageZ (weight-for-age z-scores; WAZ)

Independent variable (IV):

3.1. Visual check of the DV

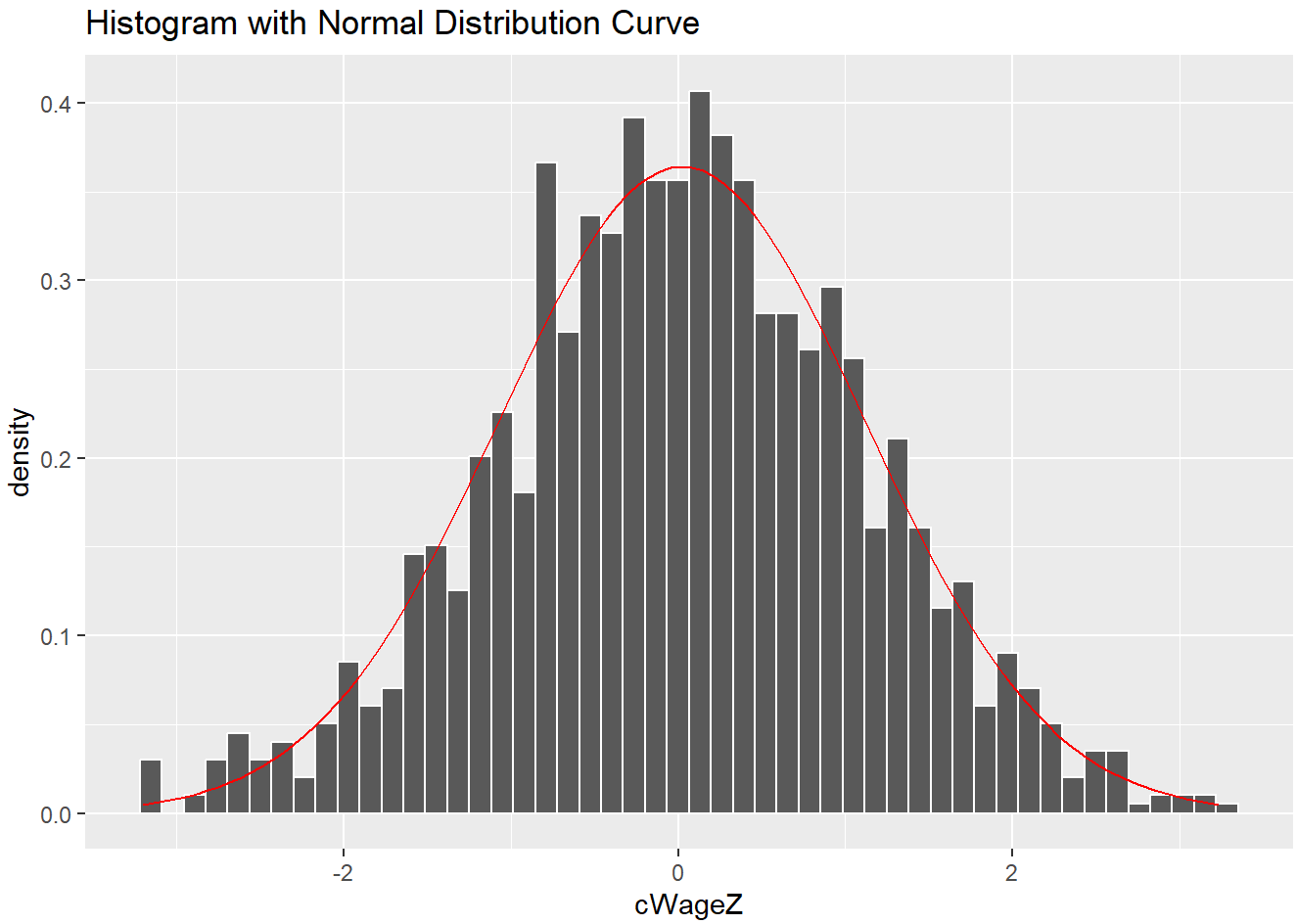
```
# Histogram for DV
library(ggplot2)
ggplot(waz_final, aes(x = cWageZ)) +
  geom_histogram(bins = 50, col= "white") +
  ggtitle("Histogram")
```

```
## Warning: Removed 17 rows containing non-finite values (`stat_bin()`).
```



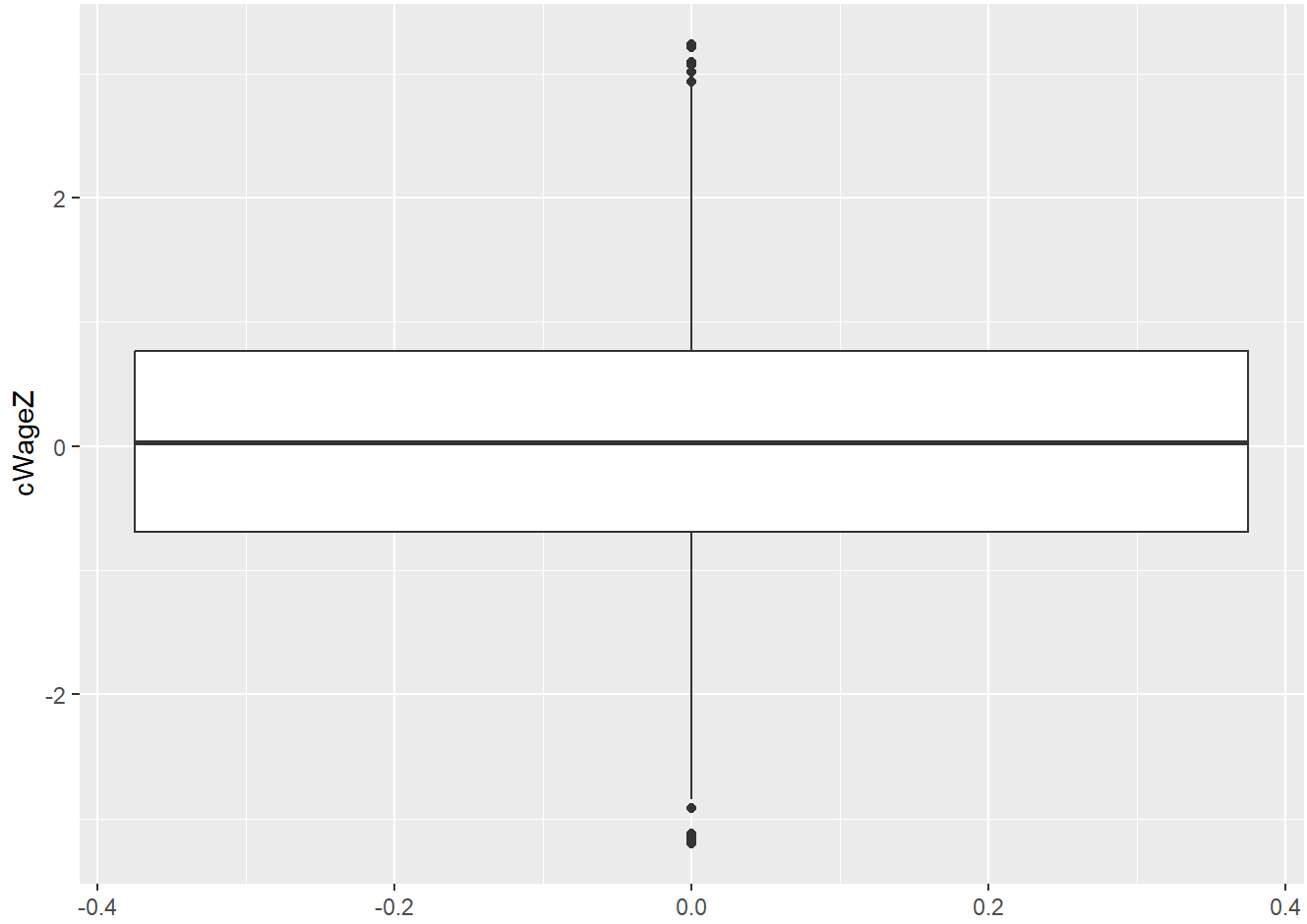

```
# Histogram adding the normal distribution curve
# Calculate mean and standard deviation of the data
data_mean <- mean(waz_final$cWageZ, na.rm = TRUE)
data_sd <- sd(waz_final$cWageZ, na.rm = TRUE)
# Create the ggplot
ggplot(waz_final, aes(x = cWageZ)) +
  geom_histogram(aes(y = after_stat(density)), bins = 50, col= "white") +
  stat_function(fun = dnorm, args = list(mean = data_mean, sd = data_sd), color = "red") +
  ggtitle("Histogram with Normal Distribution Curve")
```

```
## Warning: Removed 17 rows containing non-finite values (`stat_bin()`).
```



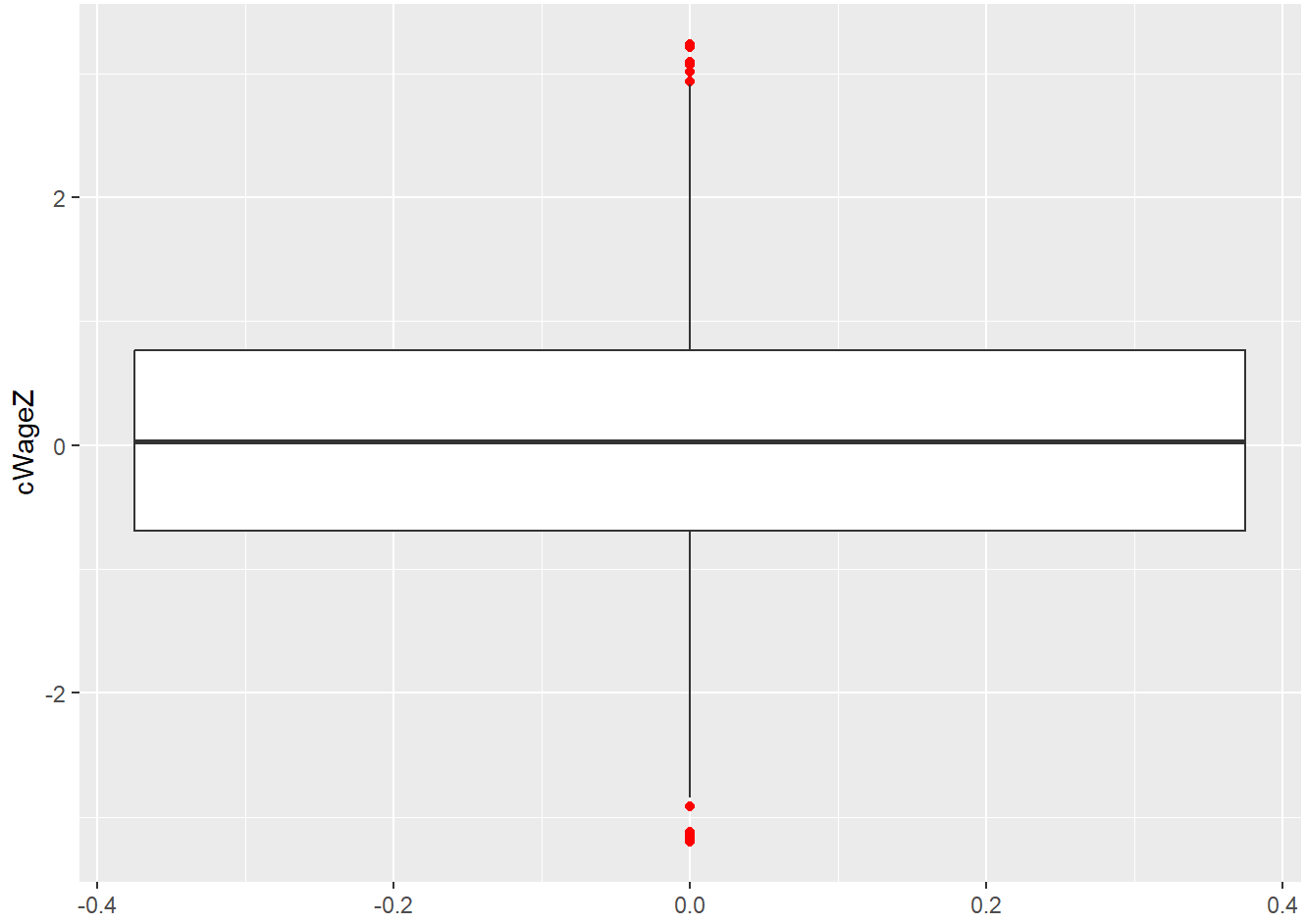

```
# Boxplot to identify outliers and understand the data's spread.
ggplot(waz_final, aes(y = cWageZ)) + geom_boxplot()
```

```
## Warning: Removed 17 rows containing non-finite values (`stat_boxplot()`).
```



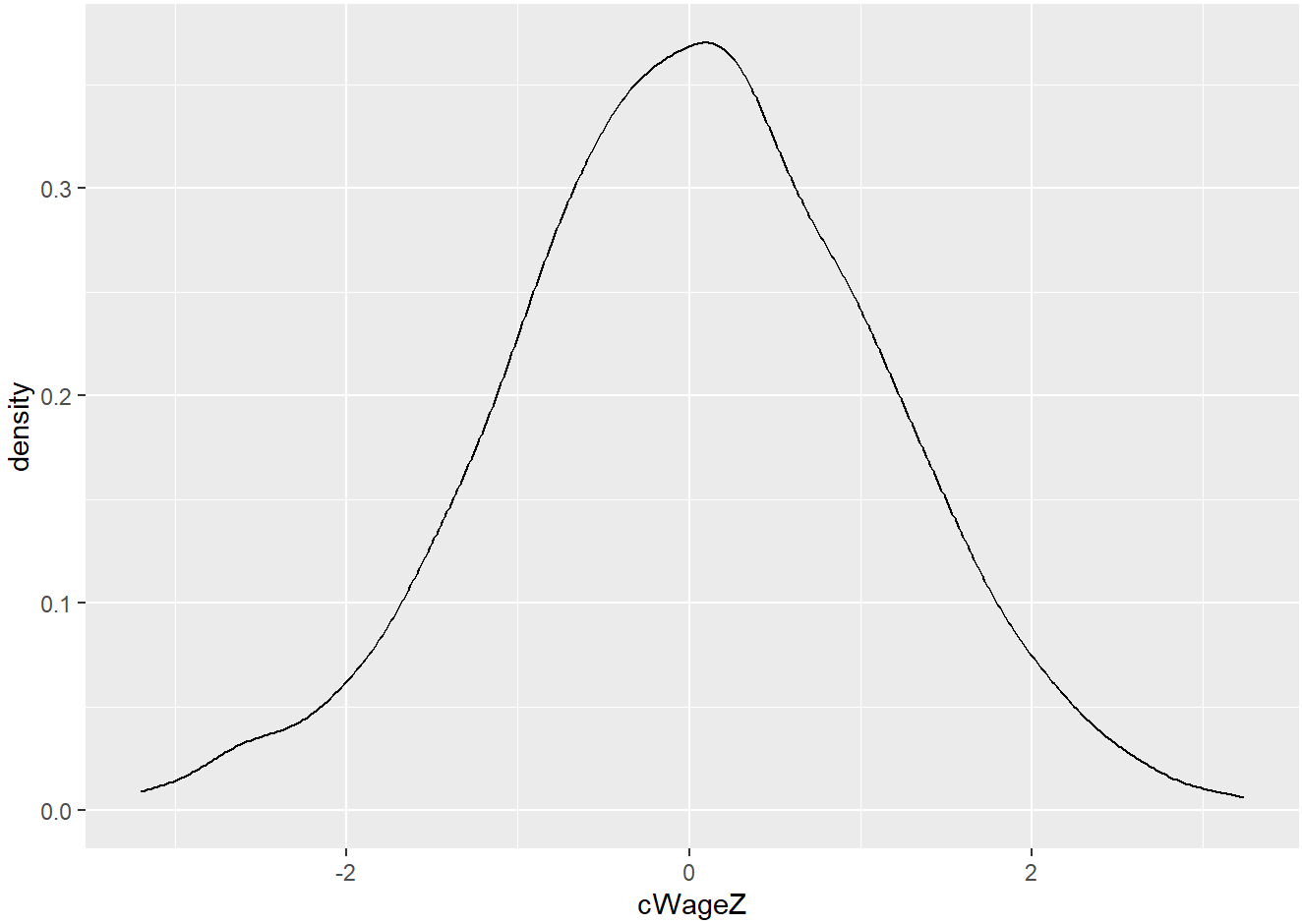

```
ggplot(waz_final, aes(y = cWageZ)) +
  geom_boxplot(outlier.colour = "red")
```

```
## Warning: Removed 17 rows containing non-finite values (`stat_boxplot()`).
```



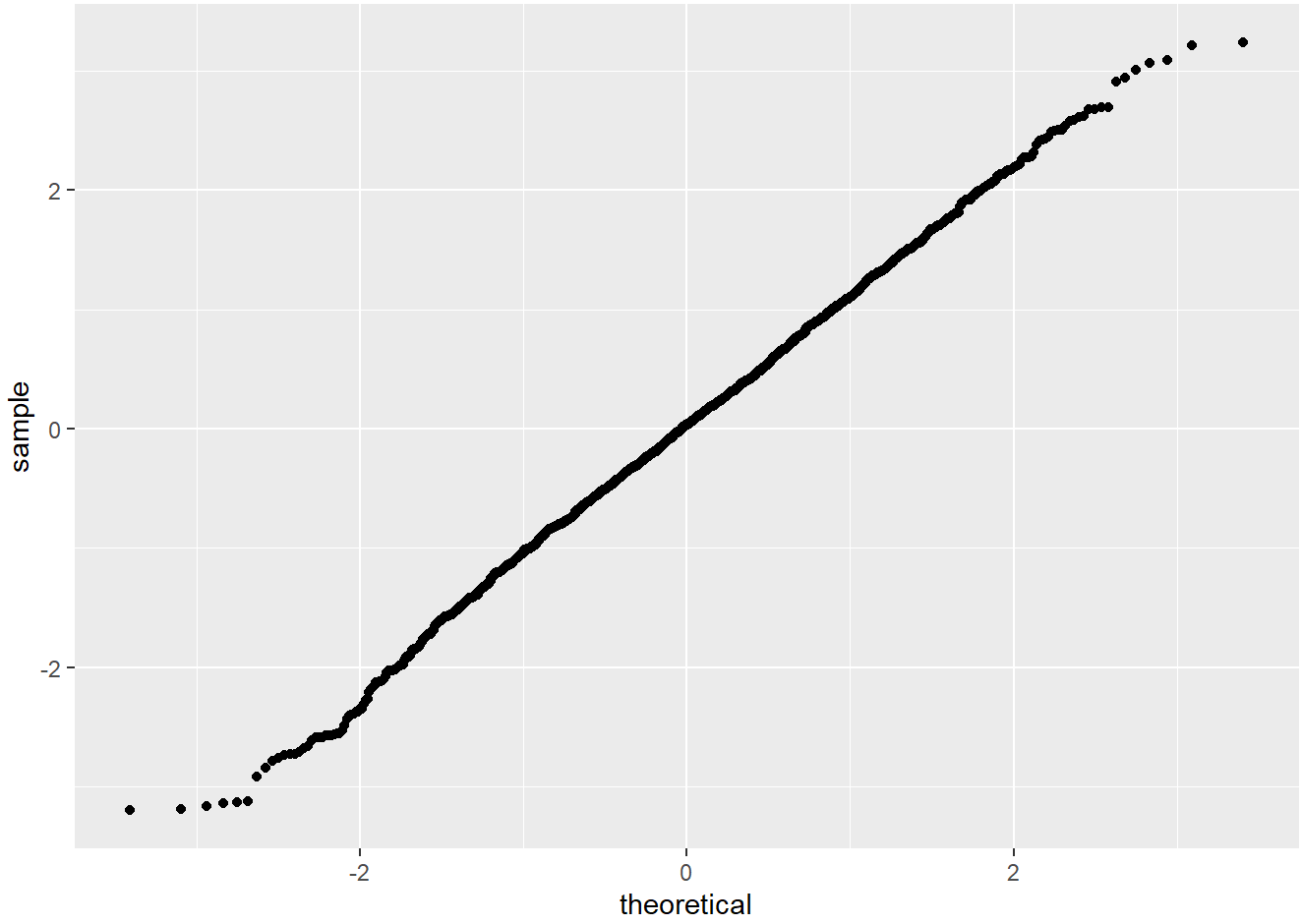

```
# Density Plot: To visualize the distribution.
ggplot(waz_final, aes(x = cWageZ)) + geom_density()
```

```
## Warning: Removed 17 rows containing non-finite values (`stat_density()`).
```




```
# Q-Q Plot: To assess normality.
ggplot(waz_final, aes(sample = cWageZ)) + stat_qq()
```

```
## Warning: Removed 17 rows containing non-finite values (`stat_qq()`).
```




```
#Pair Plot: To visualize relationships between multiple numerical variables, if applicable.

# Filter only numeric columns
# waz_final_numeric <- waz_final[, sapply(waz_final, is.numeric)]


# Create the pairs plot
# pairs(waz_final_numeric)
```

3.2. Compute central tendency, dispersion, skewness, kurtosis, and normality (Shapiro-Wilk P-value)

To be continued.