automatization\_notebook\_02

2024-10-27

# Чтение данных

В этом варианте нужно использовать датасет food.

data <- as\_tibble(read.csv("./data/raw/food.csv"))

# Выведите общее описание данных

data |> glimpse()

## Rows: 7,083  
## Columns: 38  
## $ Category <chr> "Milk", "Milk", "Milk", "Milk", "Milk",…  
## $ Description <chr> "Milk, human", "Milk, NFS", "Milk, whol…  
## $ Nutrient.Data.Bank.Number <int> 11000000, 11100000, 11111000, 11111100,…  
## $ Data.Alpha.Carotene <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Data.Beta.Carotene <int> 7, 4, 7, 7, 7, 1, 0, 3, 1, 3, 1, 2, 1, …  
## $ Data.Beta.Cryptoxanthin <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Data.Carbohydrate <dbl> 6.89, 4.87, 4.67, 4.46, 4.67, 5.19, 4.8…  
## $ Data.Cholesterol <int> 14, 8, 12, 14, 12, 5, 2, 8, 5, 8, 5, 3,…  
## $ Data.Choline <dbl> 16.0, 17.9, 17.8, 16.0, 17.8, 17.4, 16.…  
## $ Data.Fiber <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Data.Lutein.and.Zeaxanthin <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Data.Lycopene <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ Data.Niacin <dbl> 0.177, 0.110, 0.105, 0.043, 0.105, 0.11…  
## $ Data.Protein <dbl> 1.03, 3.34, 3.28, 3.10, 3.28, 3.38, 3.4…  
## $ Data.Retinol <int> 60, 58, 31, 28, 31, 58, 137, 83, 58, 83…  
## $ Data.Riboflavin <dbl> 0.036, 0.137, 0.138, 0.105, 0.138, 0.14…  
## $ Data.Selenium <dbl> 1.8, 1.9, 1.9, 2.0, 1.9, 2.1, 2.1, 1.8,…  
## $ Data.Sugar.Total <dbl> 6.89, 4.89, 4.81, 4.46, 4.81, 4.96, 4.8…  
## $ Data.Thiamin <dbl> 0.014, 0.057, 0.056, 0.020, 0.056, 0.05…  
## $ Data.Water <dbl> 87.50, 89.04, 88.10, 88.20, 88.10, 89.7…  
## $ Data.Fat.Monosaturated.Fat <dbl> 1.658, 0.426, 0.688, 0.999, 0.688, 0.21…  
## $ Data.Fat.Polysaturated.Fat <dbl> 0.497, 0.065, 0.108, 0.128, 0.108, 0.03…  
## $ Data.Fat.Saturated.Fat <dbl> 2.009, 1.164, 1.860, 2.154, 1.860, 0.56…  
## $ Data.Fat.Total.Lipid <dbl> 4.38, 1.99, 3.20, 3.46, 3.20, 0.95, 0.1…  
## $ Data.Major.Minerals.Calcium <int> 32, 126, 123, 101, 123, 126, 204, 126, …  
## $ Data.Major.Minerals.Copper <dbl> 0.052, 0.001, 0.001, 0.010, 0.001, 0.00…  
## $ Data.Major.Minerals.Iron <dbl> 0.03, 0.00, 0.00, 0.05, 0.00, 0.00, 0.0…  
## $ Data.Major.Minerals.Magnesium <int> 3, 12, 12, 5, 12, 12, 11, 12, 12, 12, 1…  
## $ Data.Major.Minerals.Phosphorus <int> 14, 103, 101, 86, 101, 103, 101, 103, 1…  
## $ Data.Major.Minerals.Potassium <int> 51, 157, 150, 253, 150, 159, 166, 159, …  
## $ Data.Major.Minerals.Sodium <int> 17, 39, 38, 3, 38, 39, 52, 39, 39, 39, …  
## $ Data.Major.Minerals.Zinc <dbl> 0.17, 0.42, 0.41, 0.38, 0.41, 0.43, 0.4…  
## $ Data.Vitamins.Vitamin.A...RAE <int> 61, 59, 32, 29, 32, 58, 137, 83, 58, 83…  
## $ Data.Vitamins.Vitamin.B12 <dbl> 0.05, 0.56, 0.54, 0.36, 0.54, 0.61, 0.3…  
## $ Data.Vitamins.Vitamin.B6 <dbl> 0.011, 0.060, 0.061, 0.034, 0.061, 0.06…  
## $ Data.Vitamins.Vitamin.C <dbl> 5.0, 0.1, 0.0, 0.9, 0.0, 0.0, 1.0, 0.2,…  
## $ Data.Vitamins.Vitamin.E <dbl> 0.08, 0.03, 0.05, 0.08, 0.05, 0.02, 0.0…  
## $ Data.Vitamins.Vitamin.K <dbl> 0.3, 0.2, 0.3, 0.3, 0.3, 0.1, 0.0, 0.2,…

# Очистка данных

1. Уберите переменные, в которых пропущенных значений больше 20% или уберите субъектов со слишком большим количеством пропущенных значений. Или совместите оба варианта. Напишите обоснование, почему вы выбрали тот или иной вариант:

**Обоснование**: Проверим количество пропущенных значений в каждом субъекте и по каждой переменной разными способами:

#data |> sum(!is.na())  
data %>% summarise(across(everything(), ~sum(is.na(.x)), .names = "na\_{col}"))

## # A tibble: 1 × 38  
## na\_Category na\_Description na\_Nutrient.Data.Bank.Number na\_Data.Alpha.Carotene  
## <int> <int> <int> <int>  
## 1 0 0 0 0  
## # ℹ 34 more variables: na\_Data.Beta.Carotene <int>,  
## # na\_Data.Beta.Cryptoxanthin <int>, na\_Data.Carbohydrate <int>,  
## # na\_Data.Cholesterol <int>, na\_Data.Choline <int>, na\_Data.Fiber <int>,  
## # na\_Data.Lutein.and.Zeaxanthin <int>, na\_Data.Lycopene <int>,  
## # na\_Data.Niacin <int>, na\_Data.Protein <int>, na\_Data.Retinol <int>,  
## # na\_Data.Riboflavin <int>, na\_Data.Selenium <int>,  
## # na\_Data.Sugar.Total <int>, na\_Data.Thiamin <int>, na\_Data.Water <int>, …

По столбцам (переменным) нет пропущенных значений.

data %>%   
 mutate(row\_id = row\_number()) %>%  
 rowwise() %>%  
 mutate(na\_count = sum(is.na(across(everything())))) %>%  
 ungroup() %>%  
 select(row\_id, na\_count) %>%  
 filter(na\_count != 0)

## # A tibble: 0 × 2  
## # ℹ 2 variables: row\_id <int>, na\_count <int>

Соответственно, по строкам (субъектам) также нет пропущенных.

sum(is.na(data))

## [1] 0

В датасете нет пропущенных значений.

1. Переименуйте переменные в человекочитаемый вид (что делать с пробелами в названиях?);

data <- data %>%  
 rename\_with(function(x) x %>% str\_replace("Data.", "") %>% str\_replace\_all("\\.", "\_"))   
data

## # A tibble: 7,083 × 38  
## Category Description Nutrient\_Bank\_Number Alpha\_Carotene Beta\_Carotene  
## <chr> <chr> <int> <int> <int>  
## 1 Milk Milk, human 11000000 0 7  
## 2 Milk Milk, NFS 11100000 0 4  
## 3 Milk Milk, whole 11111000 0 7  
## 4 Milk Milk, low sodium,… 11111100 0 7  
## 5 Milk Milk, calcium for… 11111150 0 7  
## 6 Milk Milk, calcium for… 11111160 0 1  
## 7 Milk Milk, calcium for… 11111170 0 0  
## 8 Milk Milk, reduced fat… 11112110 0 3  
## 9 Milk Milk, acidophilus… 11112120 0 1  
## 10 Milk Milk, acidophilus… 11112130 0 3  
## # ℹ 7,073 more rows  
## # ℹ 33 more variables: Beta\_Cryptoxanthin <int>, Carbohydrate <dbl>,  
## # Cholesterol <int>, Choline <dbl>, Fiber <dbl>, Lutein\_and\_Zeaxanthin <int>,  
## # Lycopene <int>, Niacin <dbl>, Protein <dbl>, Retinol <int>,  
## # Riboflavin <dbl>, Selenium <dbl>, Sugar\_Total <dbl>, Thiamin <dbl>,  
## # Water <dbl>, Fat\_Monosaturated\_Fat <dbl>, Fat\_Polysaturated\_Fat <dbl>,  
## # Fat\_Saturated\_Fat <dbl>, Fat\_Total\_Lipid <dbl>, …

Для читаемого вида было убрано “Data.” с начала строки, а все точки заменены на “*”. Пробелов в названиях нет, но при наличии можно было бы также заменить на ”*”.

1. В соответствии с описанием данных приведите переменные к нужному типу (numeric или factor);

data <- data %>%  
 mutate(across(where(is.character), function(x) as.factor(x))) %>%  
 mutate(across(where(is.double), function(x) as.numeric(x))) %>%  
 mutate(across(where(is.integer), function(x) as.numeric(x)))   
   
data$Nutrient\_Bank\_Number <- as.character(data$Nutrient\_Bank\_Number)

Теперь у нас данные либо в типе factor, либо в numeric. Nutrient\_Bank\_Number по своей сути обозначает идентификатор, и на мой взгляд было бы логичнее отнести его к категориальным переменным.

1. Отсортируйте данные по возрасту по убыванию;

Так как переменной возраста в этом датасете нет, сортируем по углеводам по убыванию.

data\_sorted <- data %>%  
 arrange(desc(Carbohydrate))

1. Сохраните в файл outliers.csv субъектов, которые являются выбросами (например, по правилу трёх сигм) — это необязательное задание со звёздочкой;
2. Отфильтруйте датасет так, чтобы остались только Rice и Cookie (переменная Category и есть группирующая);

data\_sorted |> filter(Category == "Rice" | Category == "Cookie")

## # A tibble: 243 × 38  
## Category Description Nutrient\_Bank\_Number Alpha\_Carotene Beta\_Carotene  
## <fct> <fct> <chr> <dbl> <dbl>  
## 1 Cookie Cookie, fortune 53222010 0 0  
## 2 Cookie Cookie, tea, Japa… 53246000 0 0  
## 3 Cookie Cookie, rice, baby 53803300 0 0  
## 4 Cookie Cookie, marshmall… 53226500 0 0  
## 5 Cookie Cookie, marshmall… 53226600 0 0  
## 6 Cookie Cookie, peanut bu… 53234250 0 0  
## 7 Cookie Cookie, meringue 53228000 0 0  
## 8 Cookie Cookie, cone shel… 53222020 0 0  
## 9 Cookie Cookie, Pizzelle 53236000 0 0  
## 10 Cookie Cookie, vanilla s… 53243050 0 0  
## # ℹ 233 more rows  
## # ℹ 33 more variables: Beta\_Cryptoxanthin <dbl>, Carbohydrate <dbl>,  
## # Cholesterol <dbl>, Choline <dbl>, Fiber <dbl>, Lutein\_and\_Zeaxanthin <dbl>,  
## # Lycopene <dbl>, Niacin <dbl>, Protein <dbl>, Retinol <dbl>,  
## # Riboflavin <dbl>, Selenium <dbl>, Sugar\_Total <dbl>, Thiamin <dbl>,  
## # Water <dbl>, Fat\_Monosaturated\_Fat <dbl>, Fat\_Polysaturated\_Fat <dbl>,  
## # Fat\_Saturated\_Fat <dbl>, Fat\_Total\_Lipid <dbl>, …

1. Присвойте получившийся датасет переменной “cleaned\_data”.

cleaned\_data <- data\_sorted |> filter(Category == "Rice" | Category == "Cookie")

# Сколько осталось переменных?

ncol(data)

## [1] 38

ncol(cleaned\_data)

## [1] 38

Количество переменных не изменилось, так как по ним мы не фильтровали.

# Сколько осталось случаев?

nrow(data)

## [1] 7083

nrow(cleaned\_data)

## [1] 243

После фильтрации осталось 243 случая.

# Есть ли в данных идентичные строки?

duplicated\_rows\_dplyr <- cleaned\_data |>  
 group\_by\_all() |>  
 filter(n() > 1) |>  
 ungroup()

Идентичных строк в данных нет.

# Сколько всего переменных с пропущенными значениями в данных и сколько пропущенных точек в каждой такой переменной?

cleaned\_data %>%   
 mutate(row\_id = row\_number()) %>%  
 rowwise() %>%  
 mutate(na\_count = sum(is.na(across(everything())))) %>%  
 ungroup() %>%  
 select(row\_id, na\_count) %>%  
 filter(na\_count != 0)

## # A tibble: 0 × 2  
## # ℹ 2 variables: row\_id <int>, na\_count <int>

Пропущенных значений нет.

# Таблица 1 - Количественные переменные

# Описательные статистики

## Количественные переменные

1. Рассчитайте для всех количественных переменных для каждой группы (Category):

1.1) Количество значений;

1.2) Количество пропущенных значений;

1.3) Среднее;

1.4) Медиану;

1.5) Стандартное отклонение;

1.6) 25% квантиль и 75% квантиль;

1.7) Интерквартильный размах;

1.8) Минимум;

1.9) Максимум;

1.10) 95% ДИ для среднего - задание со звёздочкой.

cleaned\_data <- cleaned\_data %>%  
 mutate(across(where(is.character), function(x) as.factor(x))) %>%  
 mutate(across(where(is.double), function(x) as.numeric(x))) %>%  
 mutate(across(where(is.integer), function(x) as.numeric(x)))   
  
statistics <- list(  
 `Количество субъектов` = ~as.character(length(.x)),  
 `Количество (есть данные)` = ~as.character(sum(!is.na(.x))),  
 `Количество пропущенных` = ~as.character(sum(is.na(.x))),  
 `Ср. знач.` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", as.character(round(mean(.x, na.rm = TRUE), 2))),  
 `Станд. отклон.` = ~ifelse(sum(!is.na(.x)) < 3, "Н/П\*", as.character(round(sd(.x, na.rm = TRUE), 2))),  
 `95% ДИ для среднего` = ~ifelse(sum(!is.na(.x)) < 3, "Н/П\*",   
 paste0(round(mean(.x, na.rm = TRUE) - 1.96 \* sd(.x, na.rm = TRUE) / sqrt(sum(!is.na(.x))), 2),  
 " - ",  
 round(mean(.x, na.rm = TRUE) + 1.96 \* sd(.x, na.rm = TRUE) / sqrt(sum(!is.na(.x))), 2))),  
 `мин. - макс.` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*",   
 paste0(round(min(.x, na.rm = TRUE), 2), " - ", round(max(.x, na.rm = TRUE), 2))),  
 `Минимум` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", as.character(round(min(.x, na.rm = TRUE), 2))),  
 `Максимум` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", as.character(round(max(.x, na.rm = TRUE), 2))),  
 `25% квантиль` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", as.character(round(quantile(.x, 0.25, na.rm = TRUE), 2))),  
 `75% квантиль` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", as.character(round(quantile(.x, 0.75, na.rm = TRUE), 2))),  
 `Медиана` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", as.character(round(median(.x, na.rm = TRUE), 2)))  
)  
  
cleaned\_data %>%  
 select(Category, where(is.numeric)) %>%  
 group\_by(Category) %>%  
 summarise(across(where(is.numeric), statistics, .names = "{.col}\_\_{.fn}")) %>%  
 pivot\_longer(!Category) %>%  
 separate(name, into = c("Переменная", "Статистика"), sep="\_\_") %>%  
 rename(`Значение` = value) %>%  
 flextable() %>%  
 theme\_box() %>%  
 align(align = "center", part="all") %>%  
 merge\_v(c("Category", "Переменная"))

## Warning: Expected 2 pieces. Additional pieces discarded in 24 rows [349, 350, 351, 352,  
## 353, 354, 355, 356, 357, 358, 359, 360, 769, 770, 771, 772, 773, 774, 775, 776,  
## ...].

| **Category** | **Переменная** | **Статистика** | **Значение** |
| --- | --- | --- | --- |
| Cookie | Alpha\_Carotene | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 8.25 |
| Станд. отклон. | 80.99 |
| 95% ДИ для среднего | -7.62 - 24.12 |
| мин. - макс. | 0 - 810 |
| Минимум | 0 |
| Максимум | 810 |
| 25% квантиль | 0 |
| 75% квантиль | 0 |
| Медиана | 0 |
| Beta\_Carotene | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 17.43 |
| Станд. отклон. | 102.62 |
| 95% ДИ для среднего | -2.68 - 37.54 |
| мин. - макс. | 0 - 1002 |
| Минимум | 0 |
| Максимум | 1002 |
| 25% квантиль | 0 |
| 75% квантиль | 1 |
| Медиана | 0 |
| Beta\_Cryptoxanthin | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.38 |
| Станд. отклон. | 1.98 |
| 95% ДИ для среднего | -0.01 - 0.77 |
| мин. - макс. | 0 - 18 |
| Минимум | 0 |
| Максимум | 18 |
| 25% квантиль | 0 |
| 75% квантиль | 0 |
| Медиана | 0 |
| Carbohydrate | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 68.09 |
| Станд. отклон. | 6.95 |
| 95% ДИ для среднего | 66.72 - 69.45 |
| мин. - макс. | 50.52 - 84 |
| Минимум | 50.52 |
| Максимум | 84 |
| 25% квантиль | 63.87 |
| 75% квантиль | 72.22 |
| Медиана | 67.78 |
| Cholesterol | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 8.68 |
| Станд. отклон. | 26.92 |
| 95% ДИ для среднего | 3.4 - 13.96 |
| мин. - макс. | 0 - 221 |
| Минимум | 0 |
| Максимум | 221 |
| 25% квантиль | 0 |
| 75% квантиль | 2 |
| Медиана | 0 |
| Choline | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 13.28 |
| Станд. отклон. | 12.36 |
| 95% ДИ для среднего | 10.85 - 15.7 |
| мин. - макс. | 0 - 92 |
| Минимум | 0 |
| Максимум | 92 |
| 25% квантиль | 6.7 |
| 75% квантиль | 14.6 |
| Медиана | 11.05 |
| Fiber | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 2.67 |
| Станд. отклон. | 2.5 |
| 95% ДИ для среднего | 2.18 - 3.16 |
| мин. - макс. | 0 - 16.1 |
| Минимум | 0 |
| Максимум | 16.1 |
| 25% квантиль | 1.48 |
| 75% квантиль | 3 |
| Медиана | 2.1 |
| Lutein\_and\_Zeaxanthin | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 23.15 |
| Станд. отклон. | 38.93 |
| 95% ДИ для среднего | 15.52 - 30.78 |
| мин. - макс. | 0 - 266 |
| Минимум | 0 |
| Максимум | 266 |
| 25% квантиль | 3 |
| 75% квантиль | 29 |
| Медиана | 8 |
| Lycopene | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.02 |
| Станд. отклон. | 0.2 |
| 95% ДИ для среднего | -0.02 - 0.06 |
| мин. - макс. | 0 - 2 |
| Минимум | 0 |
| Максимум | 2 |
| 25% квантиль | 0 |
| 75% квантиль | 0 |
| Медиана | 0 |
| Niacin | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 2.72 |
| Станд. отклон. | 2.07 |
| 95% ДИ для среднего | 2.31 - 3.12 |
| мин. - макс. | 0.05 - 15.97 |
| Минимум | 0.05 |
| Максимум | 15.97 |
| 25% квантиль | 1.75 |
| 75% квантиль | 3.14 |
| Медиана | 2.23 |
| Protein | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 5.34 |
| Станд. отклон. | 2.16 |
| 95% ДИ для среднего | 4.92 - 5.76 |
| мин. - макс. | 2.19 - 12.5 |
| Минимум | 2.19 |
| Максимум | 12.5 |
| 25% квантиль | 3.8 |
| 75% квантиль | 6.2 |
| Медиана | 5 |
| Retinol | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 20.13 |
| Станд. отклон. | 57.16 |
| 95% ДИ для среднего | 8.93 - 31.33 |
| мин. - макс. | 0 - 270 |
| Минимум | 0 |
| Максимум | 270 |
| 25% квантиль | 0 |
| 75% квантиль | 7 |
| Медиана | 0.5 |
| Riboflavin | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.28 |
| Станд. отклон. | 0.33 |
| 95% ДИ для среднего | 0.21 - 0.34 |
| мин. - макс. | 0.04 - 3.23 |
| Минимум | 0.04 |
| Максимум | 3.23 |
| 25% квантиль | 0.19 |
| 75% квантиль | 0.29 |
| Медиана | 0.23 |
| Selenium | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 8.13 |
| Станд. отклон. | 4.98 |
| 95% ДИ для среднего | 7.16 - 9.11 |
| мин. - макс. | 0 - 24.7 |
| Минимум | 0 |
| Максимум | 24.7 |
| 25% квантиль | 4.75 |
| 75% квантиль | 10 |
| Медиана | 6.3 |
| Sugar\_Total | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 32.71 |
| Станд. отклон. | 12.45 |
| 95% ДИ для среднего | 30.27 - 35.15 |
| мин. - макс. | 0 - 80.35 |
| Минимум | 0 |
| Максимум | 80.35 |
| 25% квантиль | 24.66 |
| 75% квантиль | 41.11 |
| Медиана | 32.9 |
| Thiamin | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.27 |
| Станд. отклон. | 0.18 |
| 95% ДИ для среднего | 0.24 - 0.3 |
| мин. - макс. | 0 - 1.46 |
| Минимум | 0 |
| Максимум | 1.46 |
| 25% квантиль | 0.18 |
| 75% квантиль | 0.3 |
| Медиана | 0.26 |
| Water | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 7.89 |
| Станд. отклон. | 5.79 |
| 95% ДИ для среднего | 6.75 - 9.02 |
| мин. - макс. | 0.7 - 24.83 |
| Минимум | 0.7 |
| Максимум | 24.83 |
| 25% квантиль | 3.6 |
| 75% квантиль | 11.5 |
| Медиана | 5.76 |
| Fat\_Monosaturated\_Fat | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 6.08 |
| Станд. отклон. | 3.23 |
| 95% ДИ для среднего | 5.44 - 6.71 |
| мин. - макс. | 0 - 16.08 |
| Минимум | 0 |
| Максимум | 16.08 |
| 25% квантиль | 3.29 |
| 75% квантиль | 8.5 |
| Медиана | 6.07 |
| Fat\_Polysaturated\_Fat | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 4.12 |
| Станд. отклон. | 2.62 |
| 95% ДИ для среднего | 3.61 - 4.63 |
| мин. - макс. | 0 - 11.49 |
| Минимум | 0 |
| Максимум | 11.49 |
| 25% квантиль | 2.26 |
| 75% квантиль | 5.75 |
| Медиана | 3.26 |
| Fat\_Saturated\_Fat | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 6.04 |
| Станд. отклон. | 4.58 |
| 95% ДИ для среднего | 5.14 - 6.94 |
| мин. - макс. | 0 - 20.1 |
| Минимум | 0 |
| Максимум | 20.1 |
| 25% квантиль | 3.05 |
| 75% квантиль | 8.09 |
| Медиана | 4.61 |
| Fat\_Total\_Lipid | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 17.33 |
| Станд. отклон. | 7.12 |
| 95% ДИ для среднего | 15.93 - 18.72 |
| мин. - макс. | 0.34 - 35.3 |
| Минимум | 0.34 |
| Максимум | 35.3 |
| 25% квантиль | 13.1 |
| 75% квантиль | 22.55 |
| Медиана | 17.95 |
| Major\_Minerals\_Calcium | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 33.18 |
| Станд. отклон. | 35.29 |
| 95% ДИ для среднего | 26.26 - 40.1 |
| мин. - макс. | 0 - 263 |
| Минимум | 0 |
| Максимум | 263 |
| 25% квантиль | 12 |
| 75% квантиль | 46 |
| Медиана | 28 |
| Major\_Minerals\_Copper | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.2 |
| Станд. отклон. | 0.11 |
| 95% ДИ для среднего | 0.18 - 0.23 |
| мин. - макс. | 0.02 - 0.58 |
| Минимум | 0.02 |
| Максимум | 0.58 |
| 25% квантиль | 0.11 |
| 75% квантиль | 0.27 |
| Медиана | 0.18 |
| Major\_Minerals\_Iron | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 3.03 |
| Станд. отклон. | 1.91 |
| 95% ДИ для среднего | 2.66 - 3.41 |
| мин. - макс. | 0 - 12.14 |
| Минимум | 0 |
| Максимум | 12.14 |
| 25% квантиль | 2.2 |
| 75% квантиль | 3.21 |
| Медиана | 2.58 |
| Major\_Minerals\_Magnesium | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 30.81 |
| Станд. отклон. | 16.97 |
| 95% ДИ для среднего | 27.48 - 34.14 |
| мин. - макс. | 6 - 101 |
| Минимум | 6 |
| Максимум | 101 |
| 25% квантиль | 15.5 |
| 75% квантиль | 40.25 |
| Медиана | 31 |
| Major\_Minerals\_Phosphorus | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 98.48 |
| Станд. отклон. | 47.59 |
| 95% ДИ для среднего | 89.15 - 107.81 |
| мин. - макс. | 5 - 277 |
| Минимум | 5 |
| Максимум | 277 |
| 25% квантиль | 66 |
| 75% квантиль | 122.25 |
| Медиана | 94 |
| Major\_Minerals\_Potassium | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 170.69 |
| Станд. отклон. | 90.42 |
| 95% ДИ для среднего | 152.97 - 188.41 |
| мин. - макс. | 39 - 504 |
| Минимум | 39 |
| Максимум | 504 |
| 25% квантиль | 110 |
| 75% квантиль | 207 |
| Медиана | 149 |
| Major\_Minerals\_Sodium | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 336.08 |
| Станд. отклон. | 123.71 |
| 95% ДИ для среднего | 311.83 - 360.33 |
| мин. - макс. | 9 - 580 |
| Минимум | 9 |
| Максимум | 580 |
| 25% квантиль | 256.75 |
| 75% квантиль | 403.75 |
| Медиана | 350 |
| Major\_Minerals\_Zinc | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.64 |
| Станд. отклон. | 0.29 |
| 95% ДИ для среднего | 0.58 - 0.69 |
| мин. - макс. | 0.03 - 1.6 |
| Минимум | 0.03 |
| Максимум | 1.6 |
| 25% квантиль | 0.45 |
| 75% квантиль | 0.75 |
| Медиана | 0.63 |
| Vitamins\_Vitamin\_A | \_RAE | 100 |
| \_RAE | 100 |
| \_RAE | 0 |
| \_RAE | 21.96 |
| \_RAE | 58.92 |
| \_RAE | 10.41 - 33.51 |
| \_RAE | 0 - 270 |
| \_RAE | 0 |
| \_RAE | 270 |
| \_RAE | 0 |
| \_RAE | 8 |
| \_RAE | 1 |
| Vitamins\_Vitamin\_B12 | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.06 |
| Станд. отклон. | 0.11 |
| 95% ДИ для среднего | 0.04 - 0.08 |
| мин. - макс. | 0 - 0.75 |
| Минимум | 0 |
| Максимум | 0.75 |
| 25% квантиль | 0 |
| 75% квантиль | 0.07 |
| Медиана | 0.01 |
| Vitamins\_Vitamin\_B6 | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.16 |
| Станд. отклон. | 0.61 |
| 95% ДИ для среднего | 0.04 - 0.28 |
| мин. - макс. | 0 - 5.9 |
| Минимум | 0 |
| Максимум | 5.9 |
| 25% квантиль | 0.03 |
| 75% квантиль | 0.09 |
| Медиана | 0.06 |
| Vitamins\_Vitamin\_C | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.34 |
| Станд. отклон. | 1.24 |
| 95% ДИ для среднего | 0.1 - 0.58 |
| мин. - макс. | 0 - 9.1 |
| Минимум | 0 |
| Максимум | 9.1 |
| 25% квантиль | 0 |
| 75% квантиль | 0.1 |
| Медиана | 0 |
| Vitamins\_Vitamin\_E | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 1.31 |
| Станд. отклон. | 1.02 |
| 95% ДИ для среднего | 1.11 - 1.51 |
| мин. - макс. | 0 - 4.32 |
| Минимум | 0 |
| Максимум | 4.32 |
| 25% квантиль | 0.26 |
| 75% квантиль | 2.12 |
| Медиана | 1.08 |
| Vitamins\_Vitamin\_K | Количество субъектов | 100 |
| Количество (есть данные) | 100 |
| Количество пропущенных | 0 |
| Ср. знач. | 8.41 |
| Станд. отклон. | 8.73 |
| 95% ДИ для среднего | 6.7 - 10.13 |
| мин. - макс. | 0 - 34.8 |
| Минимум | 0 |
| Максимум | 34.8 |
| 25% квантиль | 3.8 |
| 75% квантиль | 8 |
| Медиана | 5.85 |
| Rice | Alpha\_Carotene | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 125.06 |
| Станд. отклон. | 211.16 |
| 95% ДИ для среднего | 90.45 - 159.67 |
| мин. - макс. | 0 - 926 |
| Минимум | 0 |
| Максимум | 926 |
| 25% квантиль | 0 |
| 75% квантиль | 193 |
| Медиана | 4 |
| Beta\_Carotene | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 310.92 |
| Станд. отклон. | 469.49 |
| 95% ДИ для среднего | 233.97 - 387.87 |
| мин. - макс. | 0 - 2044 |
| Минимум | 0 |
| Максимум | 2044 |
| 25% квантиль | 6 |
| 75% квантиль | 415 |
| Медиана | 74 |
| Beta\_Cryptoxanthin | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 1.25 |
| Станд. отклон. | 5.79 |
| 95% ДИ для среднего | 0.3 - 2.2 |
| мин. - макс. | 0 - 31 |
| Минимум | 0 |
| Максимум | 31 |
| 25% квантиль | 0 |
| 75% квантиль | 0 |
| Медиана | 0 |
| Carbohydrate | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 21.95 |
| Станд. отклон. | 3.21 |
| 95% ДИ для среднего | 21.42 - 22.47 |
| мин. - макс. | 15.03 - 32.5 |
| Минимум | 15.03 |
| Максимум | 32.5 |
| 25% квантиль | 19.82 |
| 75% квантиль | 23.22 |
| Медиана | 21.36 |
| Cholesterol | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 2.51 |
| Станд. отклон. | 6.65 |
| 95% ДИ для среднего | 1.42 - 3.6 |
| мин. - макс. | 0 - 47 |
| Минимум | 0 |
| Максимум | 47 |
| 25% квантиль | 0 |
| 75% квантиль | 1 |
| Медиана | 0 |
| Choline | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 7.63 |
| Станд. отклон. | 4.22 |
| 95% ДИ для среднего | 6.93 - 8.32 |
| мин. - макс. | 1.1 - 25 |
| Минимум | 1.1 |
| Максимум | 25 |
| 25% квантиль | 3.8 |
| 75% квантиль | 9.9 |
| Медиана | 7.6 |
| Fiber | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 1.24 |
| Станд. отклон. | 0.7 |
| 95% ДИ для среднего | 1.13 - 1.35 |
| мин. - макс. | 0.1 - 4.3 |
| Минимум | 0.1 |
| Максимум | 4.3 |
| 25% квантиль | 0.9 |
| 75% квантиль | 1.7 |
| Медиана | 1.1 |
| Lutein\_and\_Zeaxanthin | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 115.45 |
| Станд. отклон. | 134.62 |
| 95% ДИ для среднего | 93.38 - 137.51 |
| мин. - макс. | 0 - 600 |
| Минимум | 0 |
| Максимум | 600 |
| 25% квантиль | 0 |
| 75% квантиль | 165 |
| Медиана | 88 |
| Lycopene | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 79.46 |
| Станд. отклон. | 193.81 |
| 95% ДИ для среднего | 47.7 - 111.23 |
| мин. - макс. | 0 - 849 |
| Минимум | 0 |
| Максимум | 849 |
| 25% квантиль | 0 |
| 75% квантиль | 0 |
| Медиана | 0 |
| Niacin | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 1.56 |
| Станд. отклон. | 0.49 |
| 95% ДИ для среднего | 1.48 - 1.64 |
| мин. - макс. | 0.29 - 2.55 |
| Минимум | 0.29 |
| Максимум | 2.55 |
| 25% квантиль | 1.21 |
| 75% квантиль | 2.04 |
| Медиана | 1.43 |
| Protein | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 2.89 |
| Станд. отклон. | 1.2 |
| 95% ДИ для среднего | 2.7 - 3.09 |
| мин. - макс. | 1.16 - 7.86 |
| Минимум | 1.16 |
| Максимум | 7.86 |
| 25% квантиль | 2.2 |
| 75% квантиль | 2.8 |
| Медиана | 2.47 |
| Retinol | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 4.69 |
| Станд. отклон. | 10.62 |
| 95% ДИ для среднего | 2.95 - 6.43 |
| мин. - макс. | 0 - 71 |
| Минимум | 0 |
| Максимум | 71 |
| 25% квантиль | 0 |
| 75% квантиль | 0 |
| Медиана | 0 |
| Riboflavin | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.05 |
| Станд. отклон. | 0.03 |
| 95% ДИ для среднего | 0.05 - 0.06 |
| мин. - макс. | 0.01 - 0.16 |
| Минимум | 0.01 |
| Максимум | 0.16 |
| 25% квантиль | 0.03 |
| 75% квантиль | 0.07 |
| Медиана | 0.05 |
| Selenium | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 5.28 |
| Станд. отклон. | 1.74 |
| 95% ДИ для среднего | 5 - 5.57 |
| мин. - макс. | 0.8 - 14.5 |
| Минимум | 0.8 |
| Максимум | 14.5 |
| 25% квантиль | 4.45 |
| 75% квантиль | 5.7 |
| Медиана | 5.1 |
| Sugar\_Total | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.87 |
| Станд. отклон. | 1 |
| 95% ДИ для среднего | 0.7 - 1.03 |
| мин. - макс. | 0.05 - 8.76 |
| Минимум | 0.05 |
| Максимум | 8.76 |
| 25% квантиль | 0.38 |
| 75% квантиль | 0.98 |
| Медиана | 0.74 |
| Thiamin | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.15 |
| Станд. отклон. | 0.05 |
| 95% ДИ для среднего | 0.14 - 0.16 |
| мин. - макс. | 0.01 - 0.3 |
| Минимум | 0.01 |
| Максимум | 0.3 |
| 25% квантиль | 0.12 |
| 75% квантиль | 0.17 |
| Медиана | 0.14 |
| Water | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 71.72 |
| Станд. отклон. | 4.62 |
| 95% ДИ для среднего | 70.97 - 72.48 |
| мин. - макс. | 51.52 - 83.14 |
| Минимум | 51.52 |
| Максимум | 83.14 |
| 25% квантиль | 69.8 |
| 75% квантиль | 74.48 |
| Медиана | 72.59 |
| Fat\_Monosaturated\_Fat | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.84 |
| Станд. отклон. | 0.56 |
| 95% ДИ для среднего | 0.75 - 0.93 |
| мин. - макс. | 0.03 - 2.61 |
| Минимум | 0.03 |
| Максимум | 2.61 |
| 25% квантиль | 0.32 |
| 75% квантиль | 1.06 |
| Медиана | 0.96 |
| Fat\_Polysaturated\_Fat | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.73 |
| Станд. отклон. | 0.46 |
| 95% ДИ для среднего | 0.66 - 0.81 |
| мин. - макс. | 0.03 - 2.56 |
| Минимум | 0.03 |
| Максимум | 2.56 |
| 25% квантиль | 0.3 |
| 75% квантиль | 1.05 |
| Медиана | 0.91 |
| Fat\_Saturated\_Fat | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.68 |
| Станд. отклон. | 1.38 |
| 95% ДИ для среднего | 0.45 - 0.91 |
| мин. - макс. | 0.03 - 15.17 |
| Минимум | 0.03 |
| Максимум | 15.17 |
| 25% квантиль | 0.23 |
| 75% квантиль | 0.5 |
| Медиана | 0.41 |
| Fat\_Total\_Lipid | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 2.4 |
| Станд. отклон. | 1.98 |
| 95% ДИ для среднего | 2.08 - 2.72 |
| мин. - макс. | 0.09 - 17.23 |
| Минимум | 0.09 |
| Максимум | 17.23 |
| 25% квантиль | 0.89 |
| 75% квантиль | 2.71 |
| Медиана | 2.42 |
| Major\_Minerals\_Calcium | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 19.69 |
| Станд. отклон. | 27.38 |
| 95% ДИ для среднего | 15.2 - 24.18 |
| мин. - макс. | 2 - 133 |
| Минимум | 2 |
| Максимум | 133 |
| 25% квантиль | 8 |
| 75% квантиль | 16.5 |
| Медиана | 11 |
| Major\_Minerals\_Copper | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.08 |
| Станд. отклон. | 0.03 |
| 95% ДИ для среднего | 0.07 - 0.08 |
| мин. - макс. | 0.03 - 0.25 |
| Минимум | 0.03 |
| Максимум | 0.25 |
| 25% квантиль | 0.06 |
| 75% квантиль | 0.09 |
| Медиана | 0.08 |
| Major\_Minerals\_Iron | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.92 |
| Станд. отклон. | 0.77 |
| 95% ДИ для среднего | 0.79 - 1.04 |
| мин. - макс. | 0.14 - 5.01 |
| Минимум | 0.14 |
| Максимум | 5.01 |
| 25% квантиль | 0.55 |
| 75% квантиль | 1 |
| Медиана | 0.81 |
| Major\_Minerals\_Magnesium | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 21.45 |
| Станд. отклон. | 10.55 |
| 95% ДИ для среднего | 19.73 - 23.18 |
| мин. - макс. | 5 - 39 |
| Минимум | 5 |
| Максимум | 39 |
| 25% квантиль | 12 |
| 75% квантиль | 32 |
| Медиана | 16 |
| Major\_Minerals\_Phosphorus | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 68.73 |
| Станд. отклон. | 29.22 |
| 95% ДИ для среднего | 63.94 - 73.52 |
| мин. - макс. | 8 - 139 |
| Минимум | 8 |
| Максимум | 139 |
| 25% квантиль | 41 |
| 75% квантиль | 91.5 |
| Медиана | 74 |
| Major\_Minerals\_Potassium | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 87.96 |
| Станд. отклон. | 37.09 |
| 95% ДИ для среднего | 81.88 - 94.04 |
| мин. - макс. | 10 - 219 |
| Минимум | 10 |
| Максимум | 219 |
| 25% квантиль | 69 |
| 75% квантиль | 105.5 |
| Медиана | 87 |
| Major\_Minerals\_Sodium | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 238.12 |
| Станд. отклон. | 91.79 |
| 95% ДИ для среднего | 223.07 - 253.16 |
| мин. - макс. | 69 - 510 |
| Минимум | 69 |
| Максимум | 510 |
| 25% квантиль | 178 |
| 75% квантиль | 265 |
| Медиана | 201 |
| Major\_Minerals\_Zinc | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.57 |
| Станд. отклон. | 0.2 |
| 95% ДИ для среднего | 0.54 - 0.6 |
| мин. - макс. | 0.22 - 1.37 |
| Минимум | 0.22 |
| Максимум | 1.37 |
| 25% квантиль | 0.43 |
| 75% квантиль | 0.69 |
| Медиана | 0.52 |
| Vitamins\_Vitamin\_A | \_RAE | 143 |
| \_RAE | 143 |
| \_RAE | 0 |
| \_RAE | 35.92 |
| \_RAE | 46.89 |
| \_RAE | 28.23 - 43.6 |
| \_RAE | 0 - 209 |
| \_RAE | 0 |
| \_RAE | 209 |
| \_RAE | 2.5 |
| \_RAE | 57.5 |
| \_RAE | 20 |
| Vitamins\_Vitamin\_B12 | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.02 |
| Станд. отклон. | 0.07 |
| 95% ДИ для среднего | 0.01 - 0.03 |
| мин. - макс. | 0 - 0.52 |
| Минимум | 0 |
| Максимум | 0.52 |
| 25% квантиль | 0 |
| 75% квантиль | 0 |
| Медиана | 0 |
| Vitamins\_Vitamin\_B6 | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.1 |
| Станд. отклон. | 0.03 |
| 95% ДИ для среднего | 0.09 - 0.1 |
| мин. - макс. | 0.03 - 0.18 |
| Минимум | 0.03 |
| Максимум | 0.18 |
| 25% квантиль | 0.09 |
| 75% квантиль | 0.12 |
| Медиана | 0.1 |
| Vitamins\_Vitamin\_C | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 2.9 |
| Станд. отклон. | 4.37 |
| 95% ДИ для среднего | 2.19 - 3.62 |
| мин. - макс. | 0 - 19.5 |
| Минимум | 0 |
| Максимум | 19.5 |
| 25% квантиль | 0 |
| 75% квантиль | 3.6 |
| Медиана | 0.8 |
| Vitamins\_Vitamin\_E | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 0.31 |
| Станд. отклон. | 0.16 |
| 95% ДИ для среднего | 0.29 - 0.34 |
| мин. - макс. | 0.02 - 0.75 |
| Минимум | 0.02 |
| Максимум | 0.75 |
| 25% квантиль | 0.18 |
| 75% квантиль | 0.44 |
| Медиана | 0.3 |
| Vitamins\_Vitamin\_K | Количество субъектов | 143 |
| Количество (есть данные) | 143 |
| Количество пропущенных | 0 |
| Ср. знач. | 5.45 |
| Станд. отклон. | 5.79 |
| 95% ДИ для среднего | 4.5 - 6.4 |
| мин. - макс. | 0 - 28.1 |
| Минимум | 0 |
| Максимум | 28.1 |
| 25% квантиль | 2.3 |
| 75% квантиль | 6.7 |
| Медиана | 3.7 |

## Категориальные переменные

1. Рассчитайте для всех категориальных переменных для каждой группы (Category):

1.1) Абсолютное количество;

1.2) Относительное количество внутри группы;

1.3) 95% ДИ для доли внутри группы - задание со звёздочкой.

cleaned\_data %>%   
 select(Category, where(is.factor)) -> factor\_data  
group\_variable <- "Category"  
variable\_name <- "Nutrient\_Bank\_Number"  
  
factor\_data %>%  
 group\_by(.data[[group\_variable]], .data[[variable\_name]]) %>%  
 summarise(n = n(), .groups = 'drop') %>%  
 mutate(`Relative frequency` = (n / sum(n)) %>% round(4) %>% `\*`(100) %>% str\_c("%")) %>%  
  
 group\_by(.data[[group\_variable]]) %>%  
 mutate(`Relative frequency by group` = (n / sum(n)) %>% round(4) %>% `\*`(100) %>% str\_c("%")) %>%  
 ungroup() %>%  
 rename(`Absolute frequency` = n)

## # A tibble: 243 × 5  
## Category Nutrient\_Bank\_Number `Absolute frequency` `Relative frequency`  
## <fct> <fct> <int> <chr>   
## 1 Cookie 53200100 1 0.41%   
## 2 Cookie 53201000 1 0.41%   
## 3 Cookie 53202000 1 0.41%   
## 4 Cookie 53203000 1 0.41%   
## 5 Cookie 53203500 1 0.41%   
## 6 Cookie 53204000 1 0.41%   
## 7 Cookie 53204010 1 0.41%   
## 8 Cookie 53204100 1 0.41%   
## 9 Cookie 53204840 1 0.41%   
## 10 Cookie 53204860 1 0.41%   
## # ℹ 233 more rows  
## # ℹ 1 more variable: `Relative frequency by group` <chr>

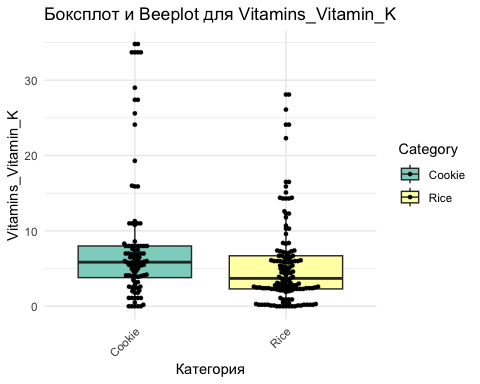
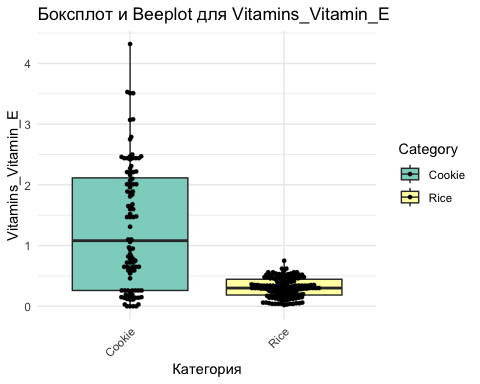
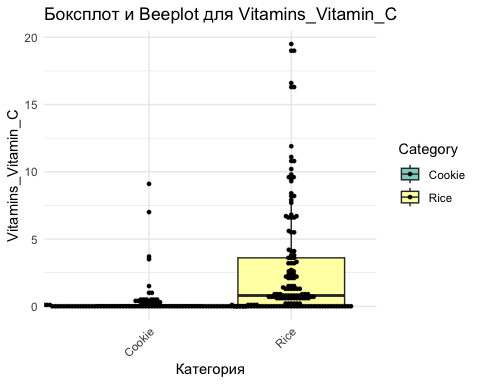
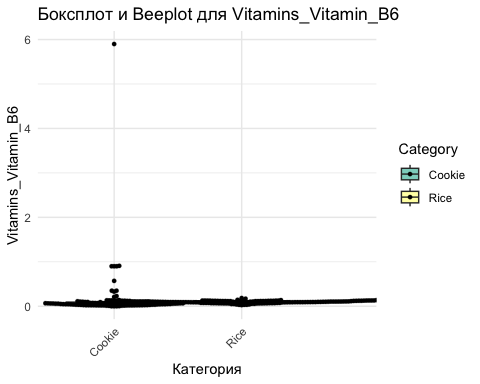
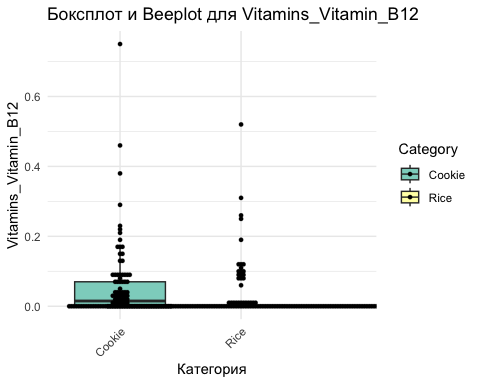
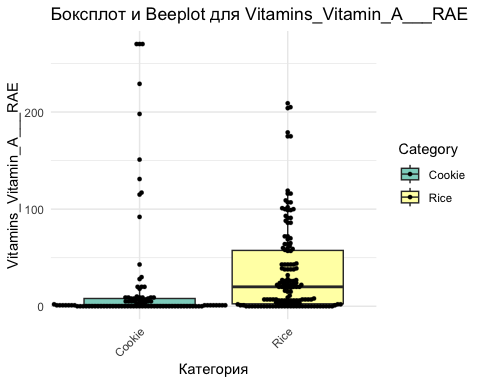
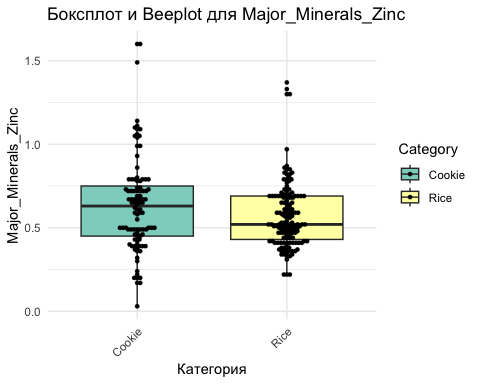
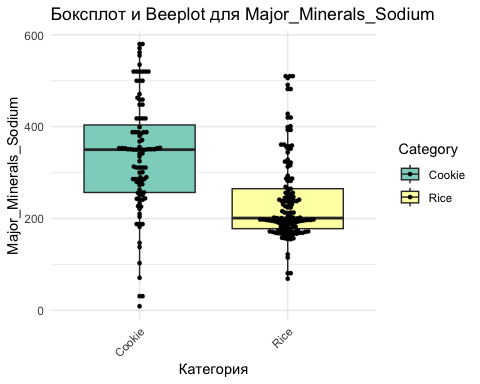
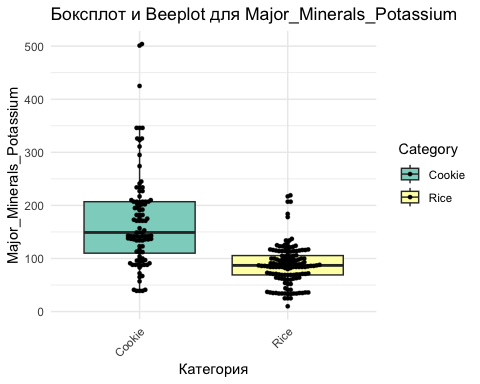
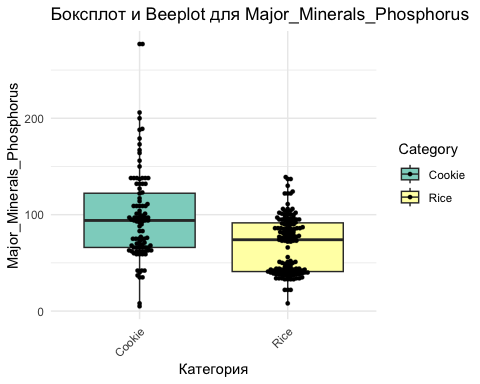
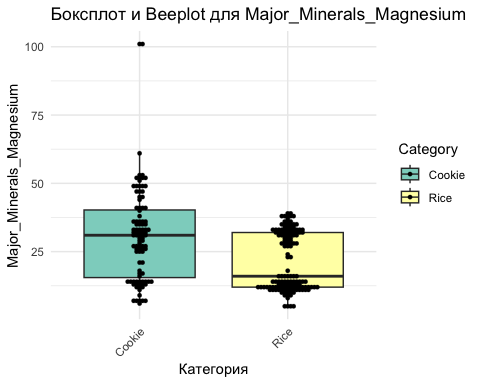
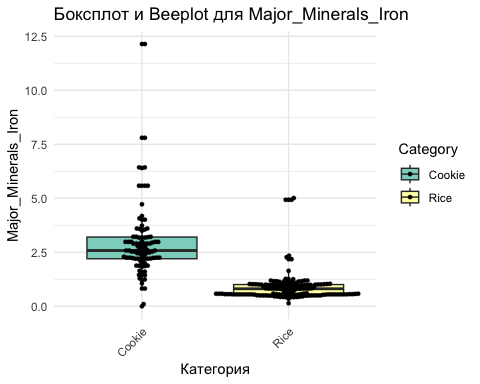
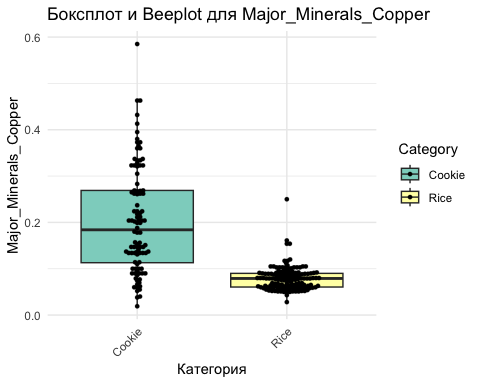
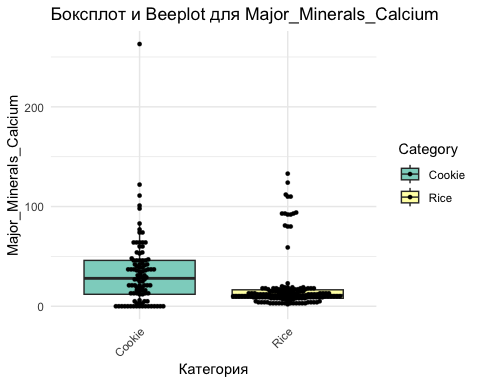
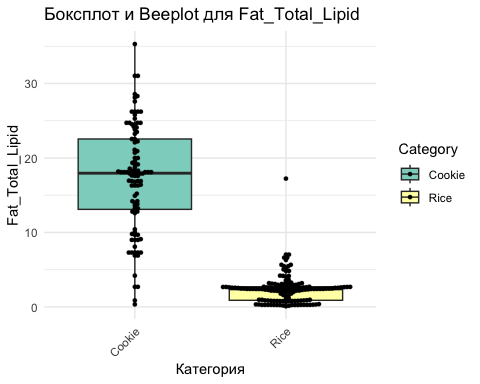
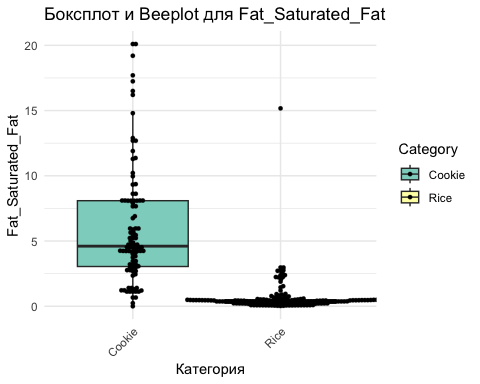
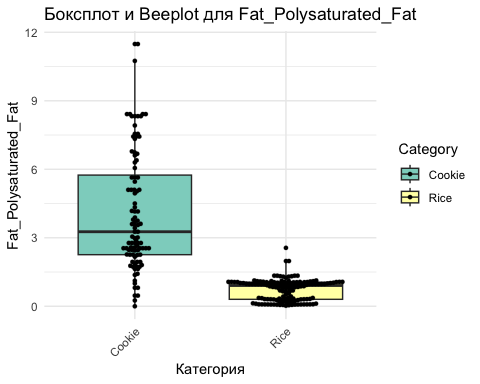
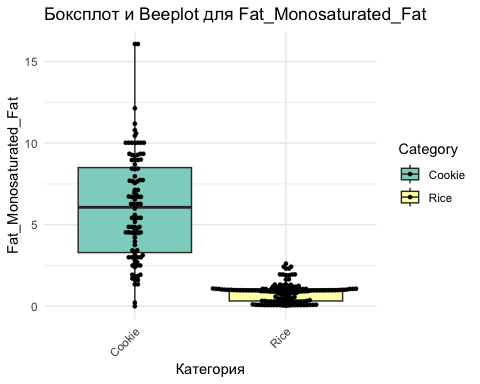
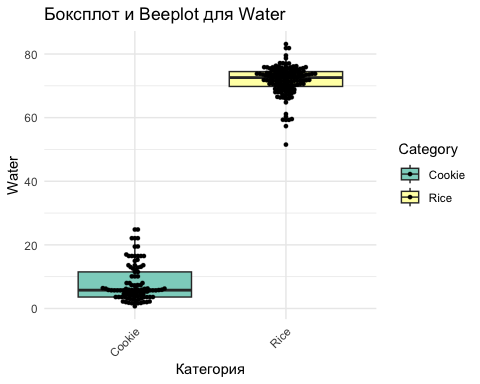
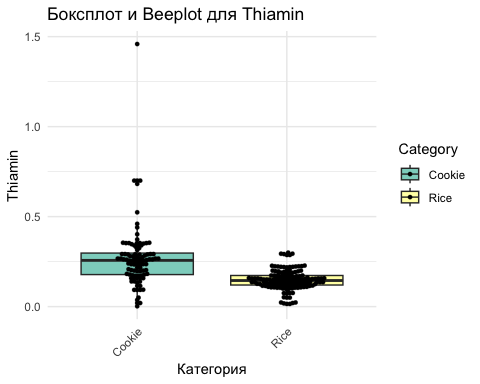
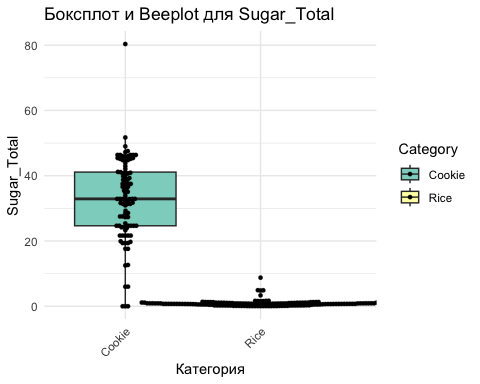
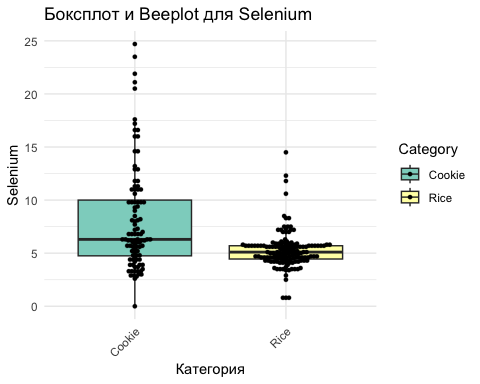
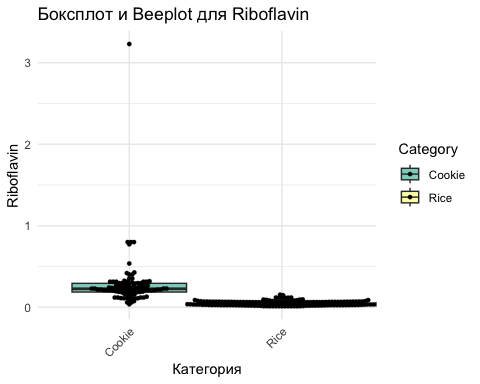
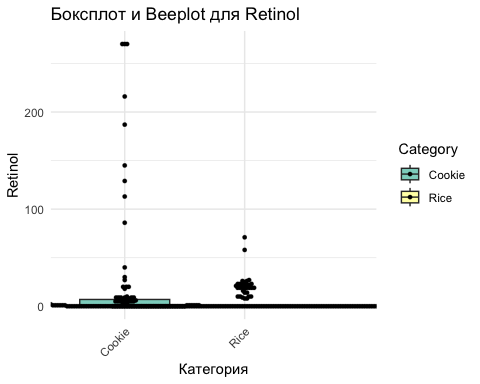
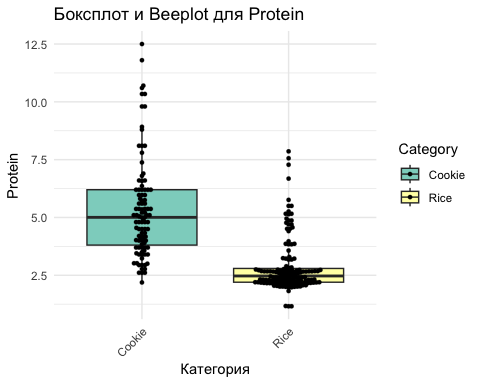
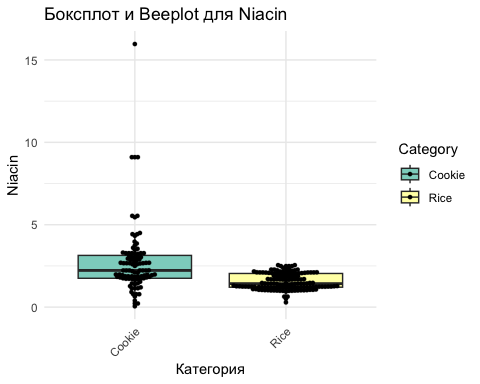
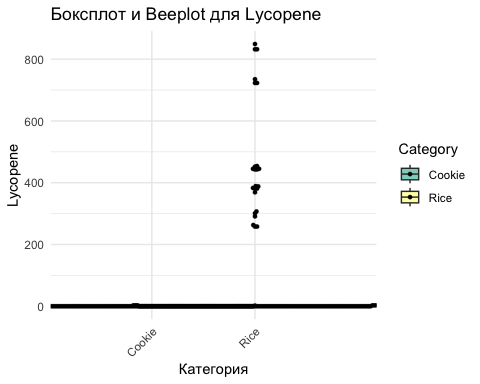
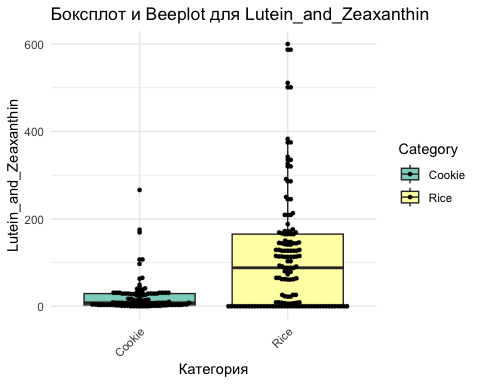
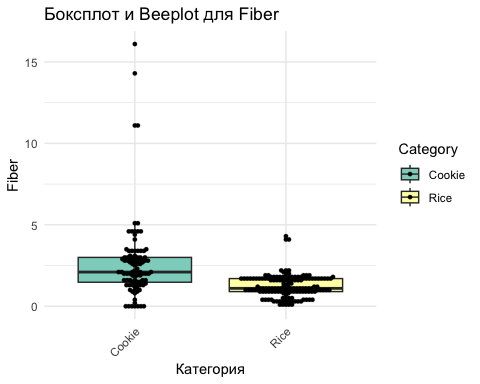
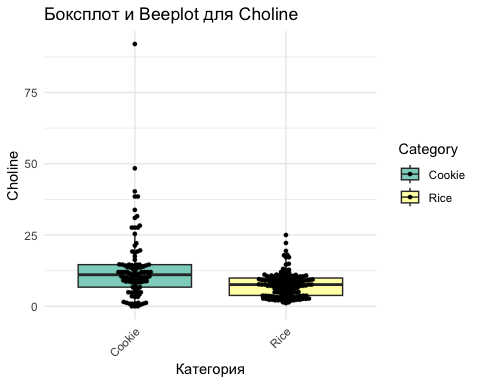
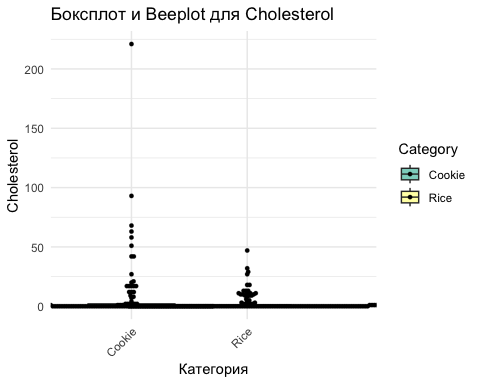
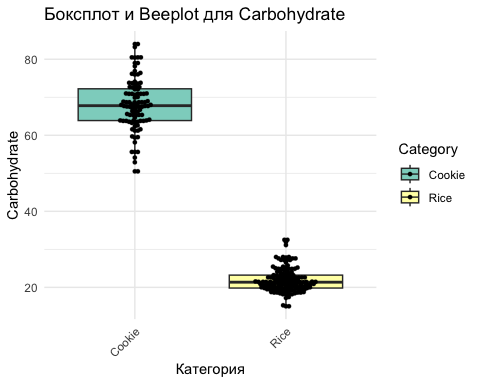
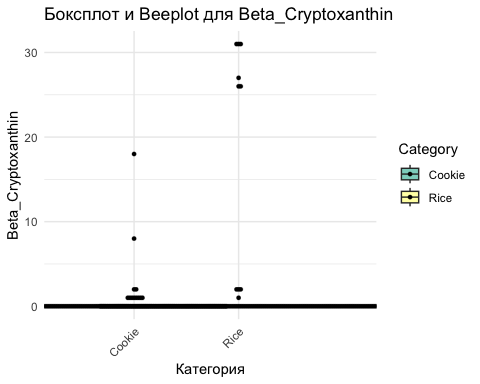
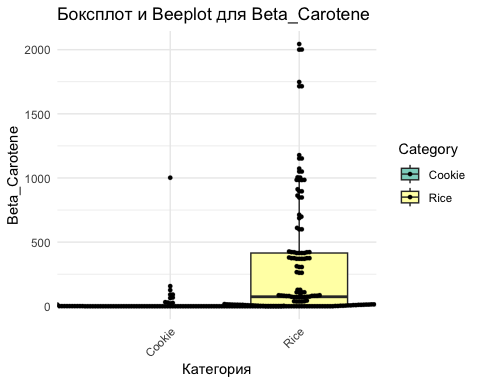
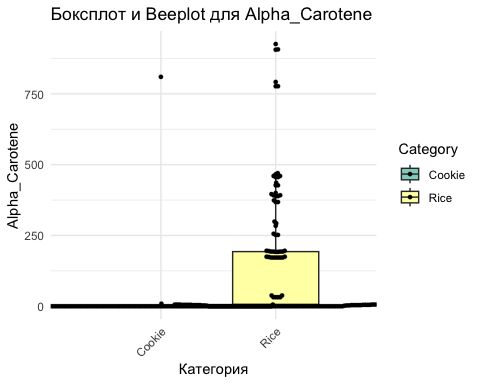
В данном датасете каждый случай включается только один раз, поэтому все частоты одинаковы.

# Визуализация

## Количественные переменные

1. Для каждой количественной переменной сделайте боксплоты по группам. Расположите их либо на отдельных рисунках, либо на одном, но читаемо;
2. Наложите на боксплоты beeplots - задание со звёздочкой.
3. Раскрасьте боксплоты с помощью библиотеки RColorBrewer.

numeric\_vars <- names(cleaned\_data)[sapply(cleaned\_data, is.numeric)]  
  
palette1 <- brewer.pal(8, "Set3")   
palette2 <- brewer.pal(8, "Set1")   
palette3 <- brewer.pal(8, "Set2")  
  
palette <- c(palette1, palette2, palette3)  
  
if (length(unique(cleaned\_data$Category)) > length(palette)) {  
 palette <- rep(palette, length.out = length(unique(cleaned\_data$Category)))  
}  
  
plots <- lapply(numeric\_vars, function(var) {  
 ggplot(cleaned\_data, aes(x = Category, y = .data[[var]], fill = Category)) +  
 geom\_boxplot(outlier.shape = NA) +   
 geom\_beeswarm(aes(color = Category), size = 1.5, dodge.width = 0.75, stroke = 0.1, color = "black") +   
 scale\_fill\_manual(values = palette) +   
 scale\_color\_manual(values = palette) +   
 labs(title = paste("Боксплот и Beeplot для", var),  
 x = "Категория",  
 y = var) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))   
})  
  
for (plot in plots) {  
 print(plot)  
}



## Категориальные переменные

1. Сделайте подходящие визуализации категориальных переменных. Обоснуйте, почему выбрали именно этот тип.

Визуализация категориальных переменных в данном датасете не имеет смысла, так как мы имеем только переменные Category, Description, Nutrient\_Bank\_Number. Последние две переменные - уникальны и встречаются только один раз.

# Статистические оценки

## Проверка на нормальность

1. Оцените каждую переменную на соответствие нормальному распределению с помощью теста Шапиро-Уилка. Какие из переменных являются нормальными и как как вы это поняли?

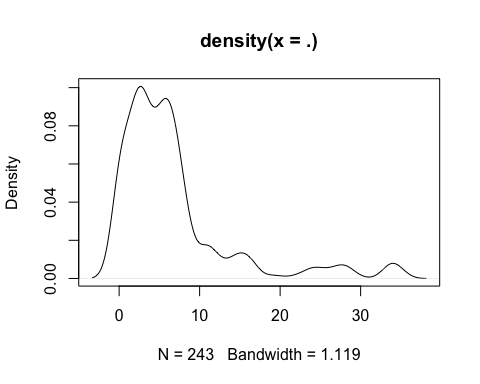
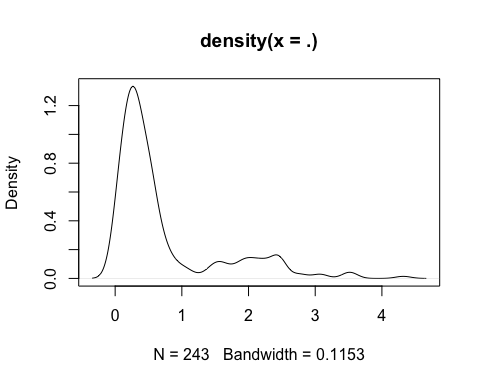
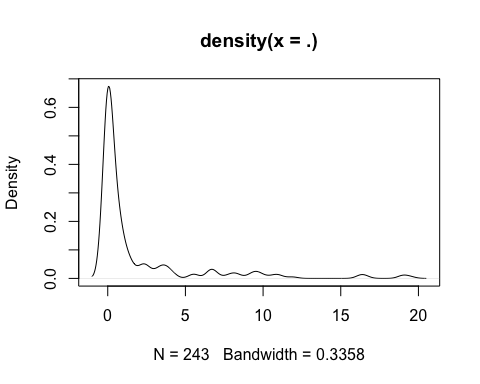
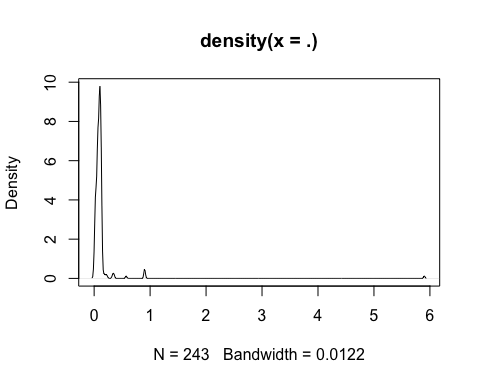
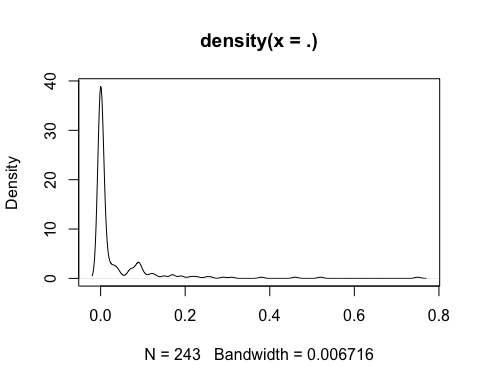
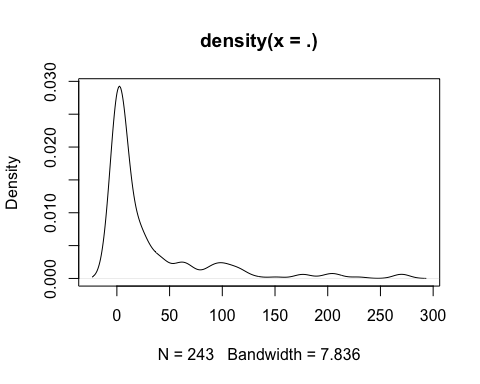
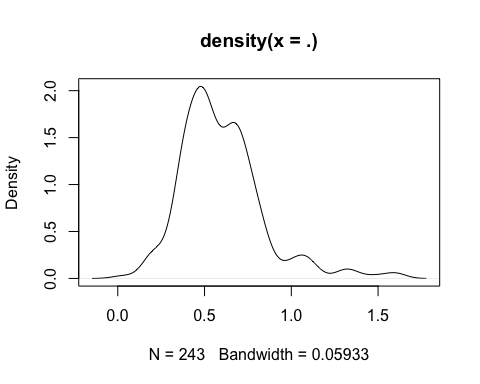
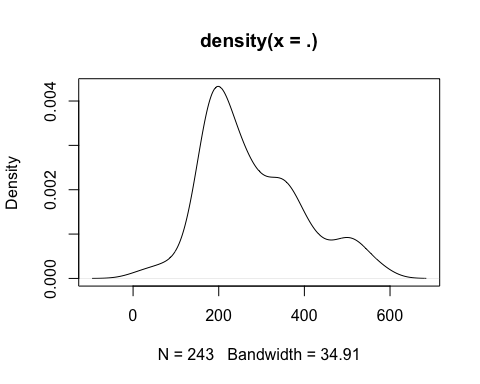
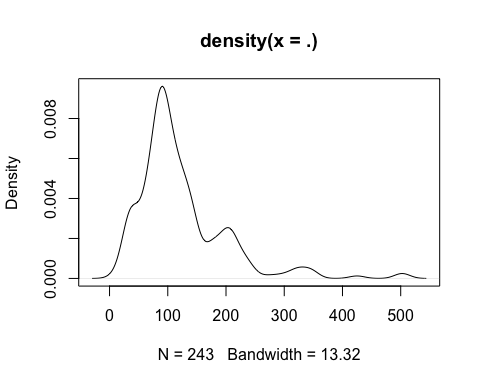
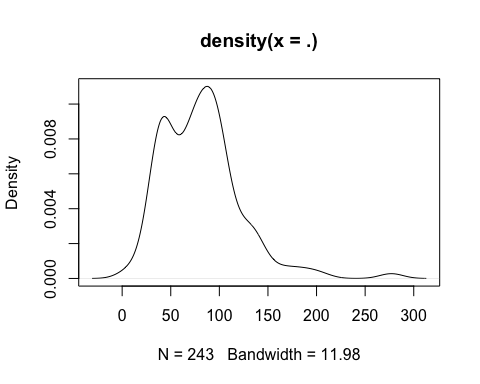
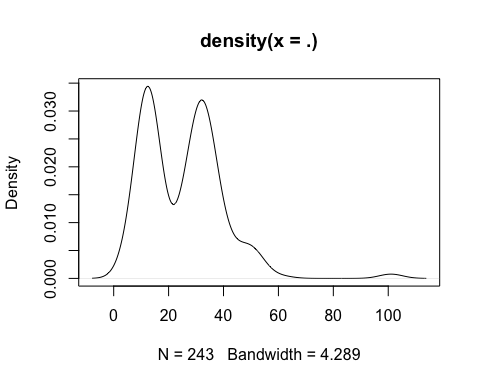
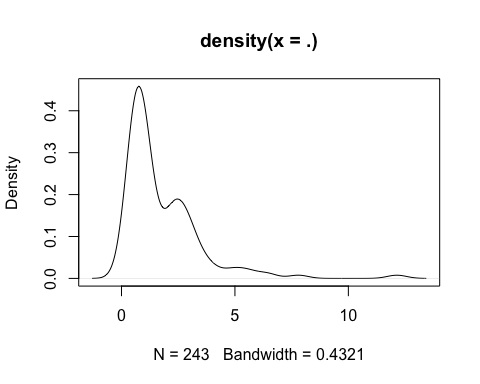
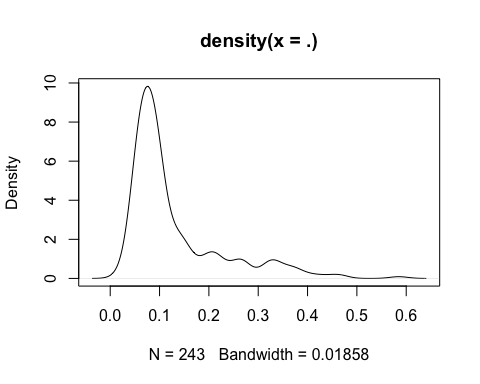
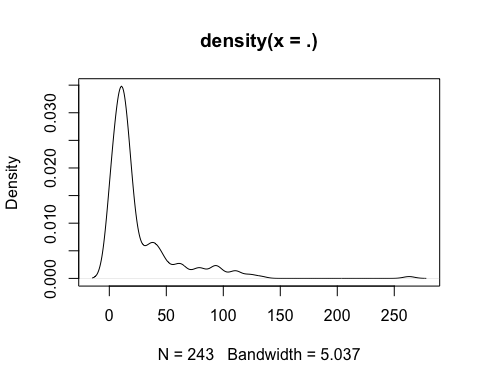
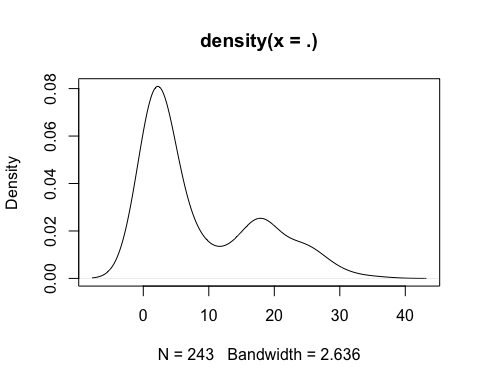
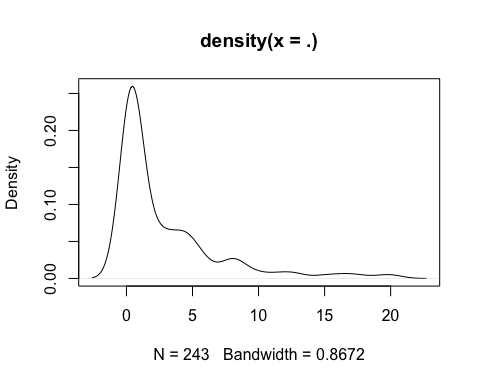
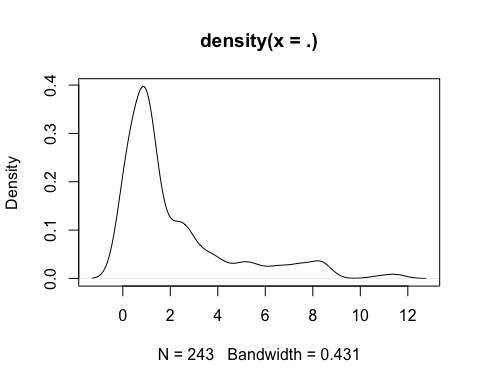
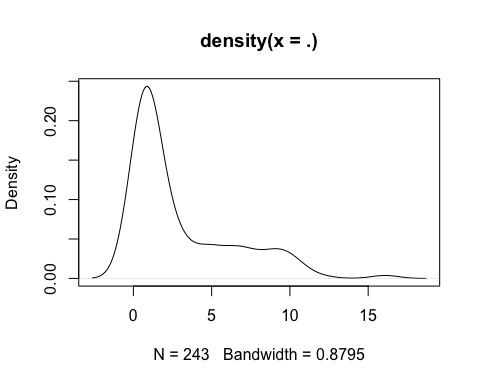
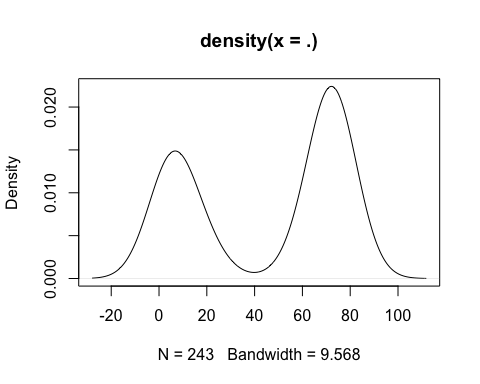
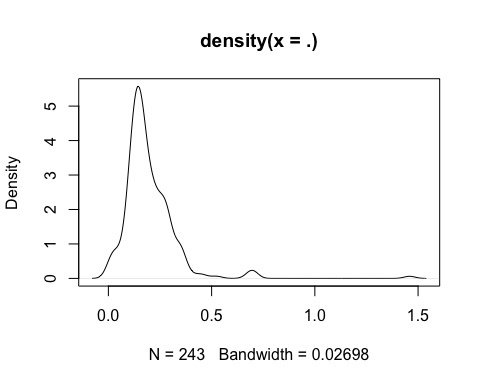
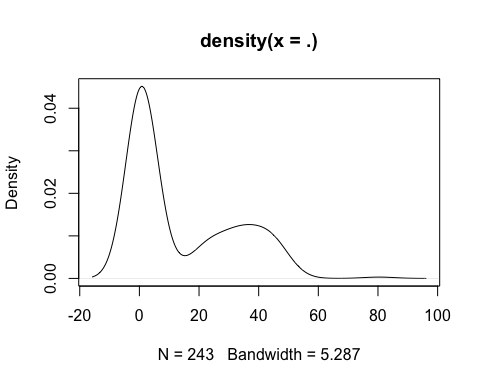
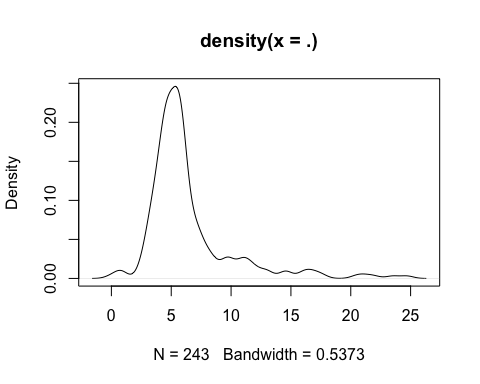
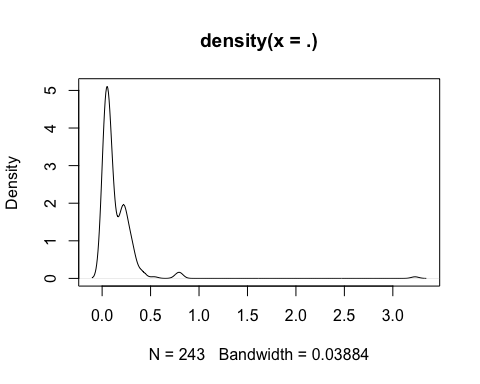
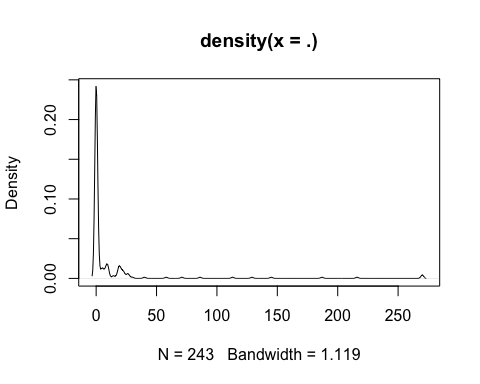
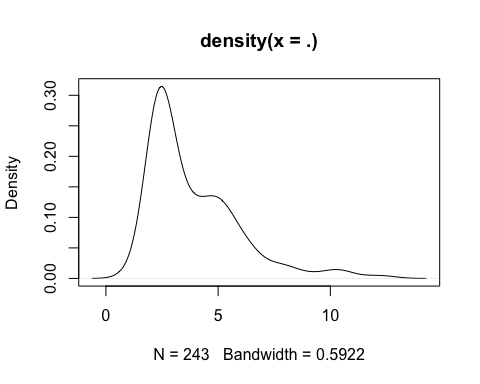
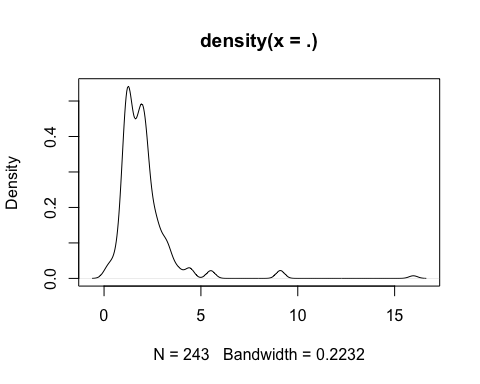
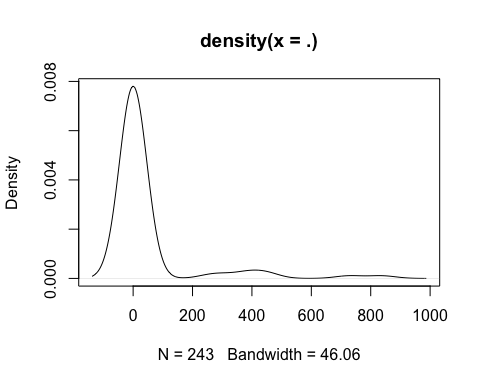
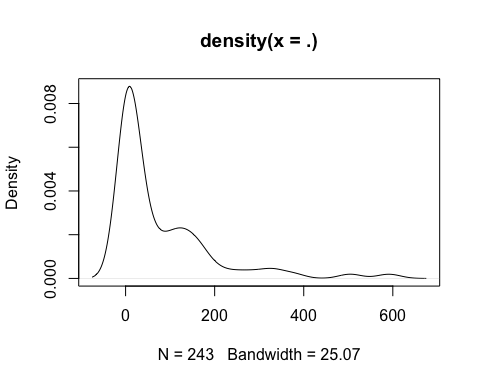
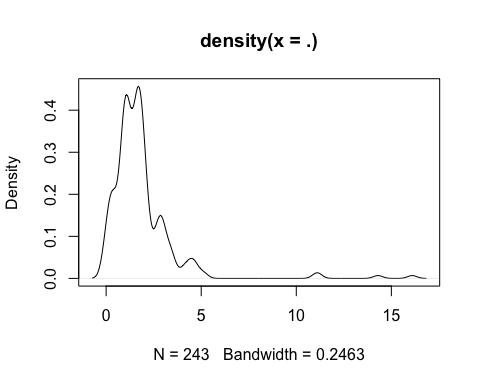
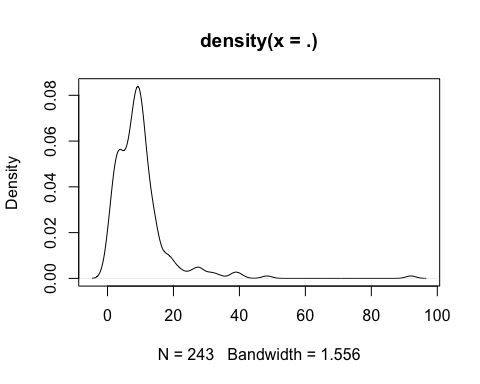
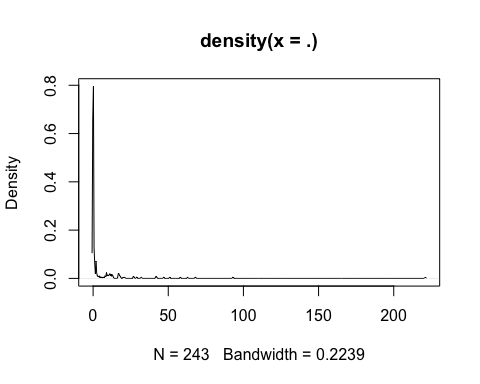
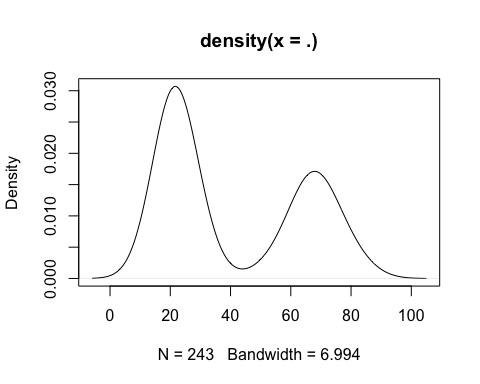
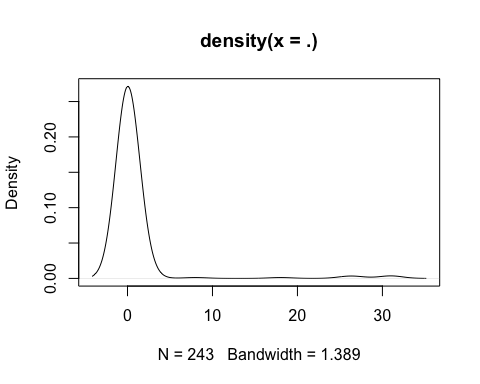
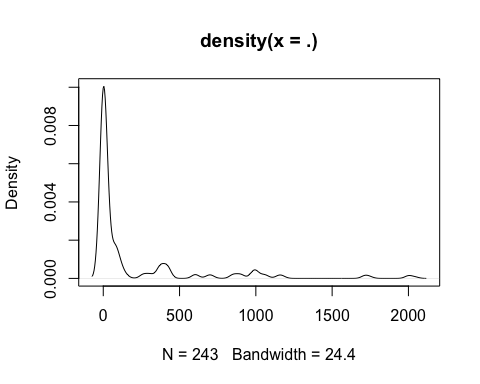
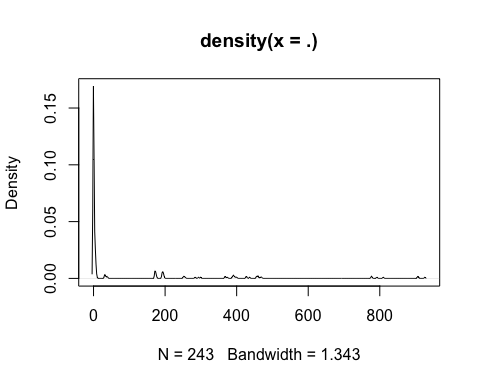
cleaned\_data %>%   
 select(where(is.numeric)) %>%  
 map(function(x) x %>% shapiro.test %>% .$p.value %>% `<`(0.05)) %>%  
 enframe() %>%  
 mutate(across(value, function(x) ifelse(value == TRUE, "Распределение отлично от нормального", "Распределение нормальное")))

## # A tibble: 35 × 2  
## name value   
## <chr> <chr>   
## 1 Alpha\_Carotene Распределение отлично от нормального  
## 2 Beta\_Carotene Распределение отлично от нормального  
## 3 Beta\_Cryptoxanthin Распределение отлично от нормального  
## 4 Carbohydrate Распределение отлично от нормального  
## 5 Cholesterol Распределение отлично от нормального  
## 6 Choline Распределение отлично от нормального  
## 7 Fiber Распределение отлично от нормального  
## 8 Lutein\_and\_Zeaxanthin Распределение отлично от нормального  
## 9 Lycopene Распределение отлично от нормального  
## 10 Niacin Распределение отлично от нормального  
## # ℹ 25 more rows

Если p-value меньше 0.05 - скорее всего, распределение не нормальное. В данном случае нет переменных с нормальным распределением.

1. Постройте для каждой количественной переменной QQ-плот. Отличаются ли выводы от теста Шапиро-Уилка? Какой метод вы бы предпочли и почему?

cleaned\_data %>%   
 select(where(is.numeric)) %>%  
 map(function(x) x %>% density() %>% plot())



## $Alpha\_Carotene  
## NULL  
##   
## $Beta\_Carotene  
## NULL  
##   
## $Beta\_Cryptoxanthin  
## NULL  
##   
## $Carbohydrate  
## NULL  
##   
## $Cholesterol  
## NULL  
##   
## $Choline  
## NULL  
##   
## $Fiber  
## NULL  
##   
## $Lutein\_and\_Zeaxanthin  
## NULL  
##   
## $Lycopene  
## NULL  
##   
## $Niacin  
## NULL  
##   
## $Protein  
## NULL  
##   
## $Retinol  
## NULL  
##   
## $Riboflavin  
## NULL  
##   
## $Selenium  
## NULL  
##   
## $Sugar\_Total  
## NULL  
##   
## $Thiamin  
## NULL  
##   
## $Water  
## NULL  
##   
## $Fat\_Monosaturated\_Fat  
## NULL  
##   
## $Fat\_Polysaturated\_Fat  
## NULL  
##   
## $Fat\_Saturated\_Fat  
## NULL  
##   
## $Fat\_Total\_Lipid  
## NULL  
##   
## $Major\_Minerals\_Calcium  
## NULL  
##   
## $Major\_Minerals\_Copper  
## NULL  
##   
## $Major\_Minerals\_Iron  
## NULL  
##   
## $Major\_Minerals\_Magnesium  
## NULL  
##   
## $Major\_Minerals\_Phosphorus  
## NULL  
##   
## $Major\_Minerals\_Potassium  
## NULL  
##   
## $Major\_Minerals\_Sodium  
## NULL  
##   
## $Major\_Minerals\_Zinc  
## NULL  
##   
## $Vitamins\_Vitamin\_A\_\_\_RAE  
## NULL  
##   
## $Vitamins\_Vitamin\_B12  
## NULL  
##   
## $Vitamins\_Vitamin\_B6  
## NULL  
##   
## $Vitamins\_Vitamin\_C  
## NULL  
##   
## $Vitamins\_Vitamin\_E  
## NULL  
##   
## $Vitamins\_Vitamin\_K  
## NULL

По графикам также видно, что переменные не имеют нормального распределения.

1. Ниже напишите, какие ещё методы проверки на нормальность вы знаете и какие у них есть ограничения.

Есть также тест Колмогорова-Смирнова (однако он чувствителен к выбросам и требует стандартизации), тест Лиллиефорса (не очень подходит для больших выборок), построение гистограммы и сравнение ее с кривой нормального распределения (не точный).

## Сравнение групп

1. Сравните группы (переменная **Category**) по каждой переменной (как количественной, так и категориальной). Для каждой переменной выберите нужный критерий и кратко обоснуйте его выбор в комментариях.

cleaned\_data %>%  
 select(where(is.numeric)) %>%  
 names() %>%  
 set\_names () %>%  
 map(function(x) wilcox.test(cleaned\_data[[x]] ~ cleaned\_data$Category)$p.value < 0.05) %>%  
 enframe() %>%  
 mutate(across(value, function(x) ifelse(value == TRUE, "Различие между группами есть", "Различие между группами не доказано"))) %>%  
 filter(value == "Различие между группами есть")

## # A tibble: 32 × 2  
## name value   
## <chr> <chr>   
## 1 Alpha\_Carotene Различие между группами есть  
## 2 Beta\_Carotene Различие между группами есть  
## 3 Carbohydrate Различие между группами есть  
## 4 Choline Различие между группами есть  
## 5 Fiber Различие между группами есть  
## 6 Lutein\_and\_Zeaxanthin Различие между группами есть  
## 7 Lycopene Различие между группами есть  
## 8 Niacin Различие между группами есть  
## 9 Protein Различие между группами есть  
## 10 Retinol Различие между группами есть  
## # ℹ 22 more rows

Так как у нас нет нормального распределения ни в одной из переменных, для количественных используем Wilcoxon test.

Для сравнения категориальных переменных используем тест Фишера. Однако с нашими данными, где каждое наблюдение по категориальным переменным встречается только один раз, это неинформативно. В данном датасете нет подходящих категориальных переменных для анализа.

В случае других переменных можно использовать нижеприведенный код, однако в данном случае он генерирует очень большой список всех случаев, поэтому я решила его оставить как комментарий.

#table(factor\_data[[group\_variable]], factor\_data[[variable\_name]]) #%>% fisher.test()

# Далее идут **необязательные** дополнительные задания, которые могут принести вам дополнительные баллы в том числе в случае ошибок в предыдущих

## Корреляционный анализ

1. Создайте корреляционную матрицу с визуализацией и поправкой на множественные сравнения. Объясните, когда лучше использовать корреляционные матрицы и в чём минусы и плюсы корреляционных исследований.

## Моделирование

1. Постройте регрессионную модель для переменной **Category**. Опишите процесс построения