Univariate Graphing Assignment

Pushpak Sunil Rane

Describing Distribution

Load necessary libraries needed.

```
# Load libraries
library(sjPlot); library(ggplot2); library(ggpubr); library(here);library(gtsummary)

# Import data
load(here("data/environmental_impact_clean.Rdata"))
```

Environmental_impacts_freshwater_use

"environmental_impacts_freshwater_use" variable represents environmental impact of using freshwater (use in km3) on diet and nutrition in certain regions.

```
summary(region$environmental_impacts_freshwater_use)
```

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
   0.0003    0.4669    2.3623   30.6096   11.1830 1699.1279
```

```
IQR(region$environmental_impacts_freshwater_use)
```

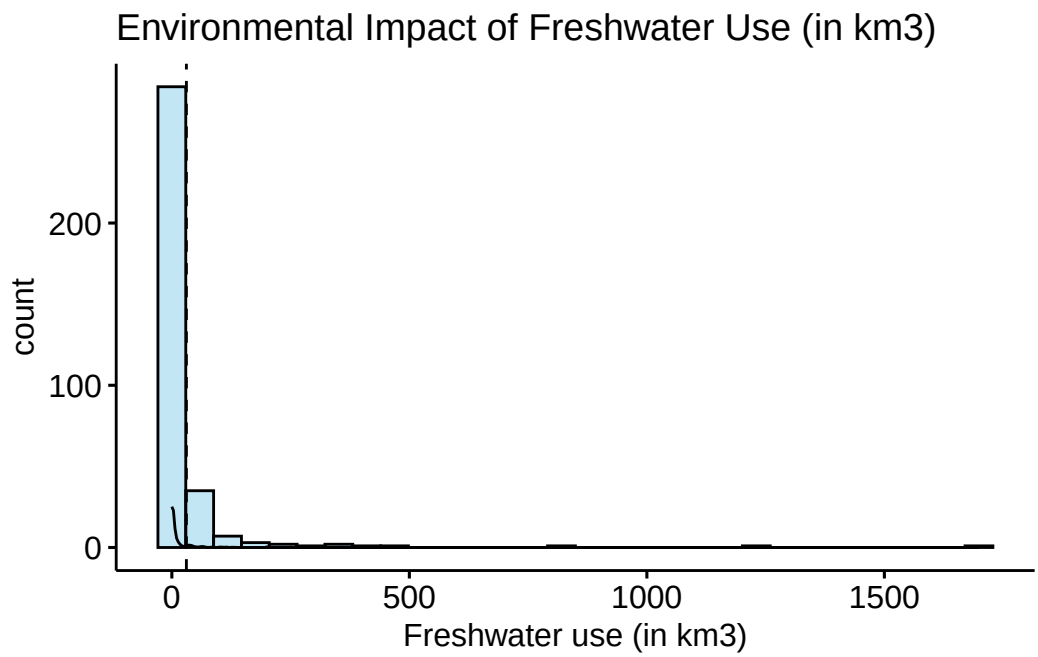
```
[1] 10.71601
```

```
sd(region$environmental_impacts_freshwater_use)
```

```
[1] 133.1164
```

Histogram for variable “environmental_impacts_freshwater_use” in km3 units.

```
# this function is from the ggpubr package
gghistogram(data=region, x = "environmental_impacts_freshwater_use", add = "mean",
            fill = "skyblue", add_density = TRUE,
            xlab = "Freshwater use (in km3)",
            title = "Environmental Impact of Freshwater Use (in km3)")
```



The histogram titled “Environmental Impact of Freshwater Use (in km³)” depicts the distribution of impact of freshwater use on diet and nutrition in various regions with 351 observations and . The data seems to be right-skewed, indicating more frequent lower values with a long tail of higher values. The mean of environmental impact of freshwater use is around 30.6096 km³, with the median of 2.3623 km³. The interquartile range (IQR) is 10.71601 which spans from 0.4669 km3 to 11.1830 km³, showing where the middle 50% of the data lies that is middle of data from variable. The range of the data is huge with a standard deviation of 133.1164 km³. There seem some outliers that are values outside this range or due to data entry errors.

Environmental_impacts_land_use

“environmental_impacts_land_use” variable represents environmental impact of land (use in Mkm2) on diet and nutrition in certain regions.

```
summary(region$environmental_impacts_land_use)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
    0.00    10.12    57.97  1035.65   410.46 58062.30
```

```
IQR(region$environmental_impacts_land_use, na.rm = TRUE)
```

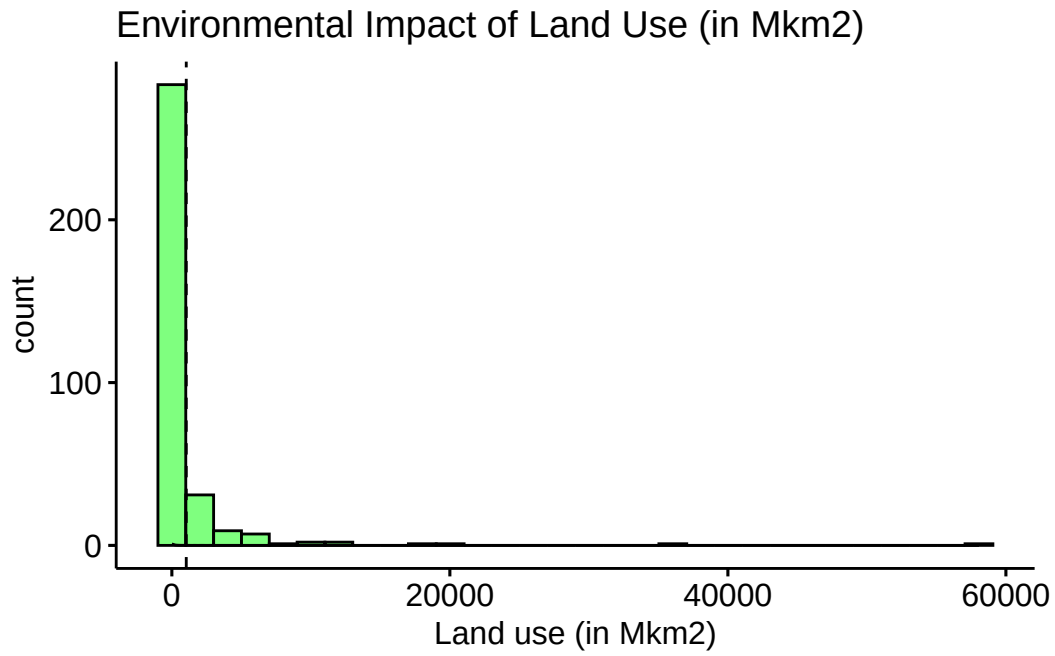
```
[1] 400.3397
```

```
sd(region$environmental_impacts_land_use, na.rm = TRUE)
```

```
[1] 4220.845
```

Histogram for variable "environmental_impacts_land_use" in Mkm2

```
# this function is from the ggpubr package
gghistogram(data=region, x = "environmental_impacts_land_use",
            add = "mean",
            fill = "green", add_density = TRUE,
            xlab = "Land use (in Mkm2)",
            title = "Environmental Impact of Land Use (in Mkm2)")
```



The histogram titled "Environmental Impact of Land Use (in Mkm2)" illustrates how impact of land use on diet and nutrition is distributed in million square kilometers (Mkm^2). The data appears to be right-skewed, suggesting that lower values are more common, with a long tail extending towards higher values. The average land use is approximately 1035.65 Mkm^2, while the median is slightly lower at 57.97 Mkm^2. The interquartile range (IQR) is 400.3397 which spans from 10.12 Mkm^2 to 410.46 Mkm^2, indicating where the middle 50% of the data falls. The overall range of the data is quite large with a standard deviation of 4220.845 Mkm^2. There are some outliers present, which could be values that fall outside this range or may result from data entry mistakes.

Region

"region" variable represents regions or places for analyzing impact of it by environmental factors on diet and nutrition.

```
table(region$region,useNA = "always")
```

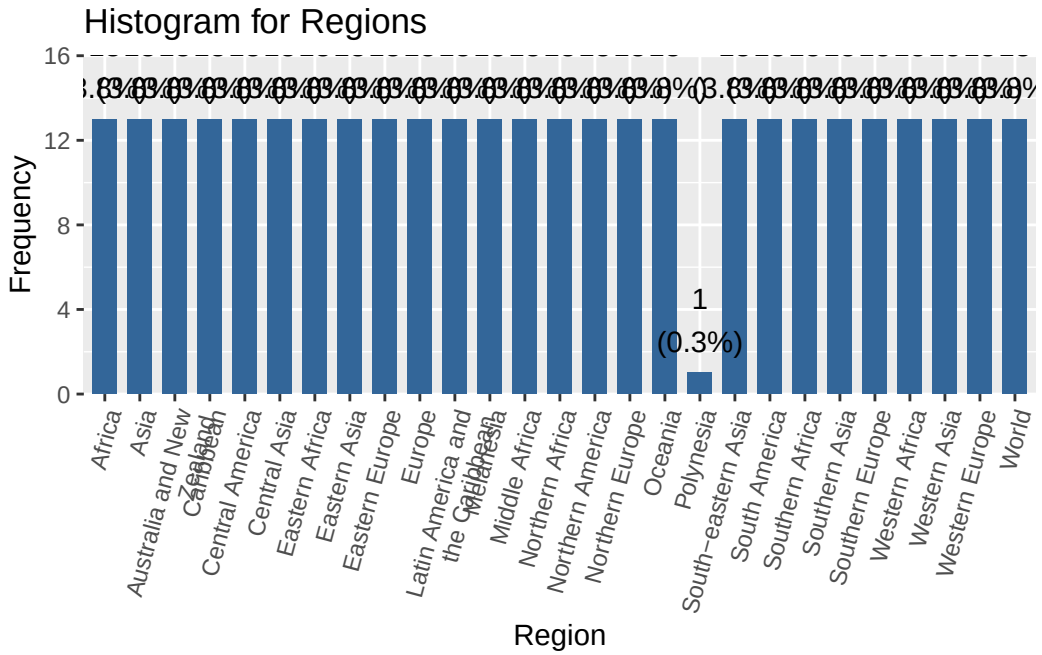
```
                          Africa                             Asia
                              13                               13
      Australia and New Zealand                        Caribbean
                              13                               13
                 Central America                     Central Asia
                              13                               13
                  Eastern Africa                     Eastern Asia
                              13                               13
                  Eastern Europe                           Europe
                              13                               13
Latin America and the Caribbean                        Melanesia
                              13                               13
                    Middle Africa                  Northern Africa
                              13                               13
                 Northern America                  Northern Europe
                              13                               13
                         Oceania                        Polynesia
                              13                                1
              South-eastern Asia                     South America
                              13                               13
                  Southern Africa                     Southern Asia
                              13                               13
                  Southern Europe                     Western Africa
```

```
                        13                         13
                Western Asia                Western Europe
                        13                         13
                     World                       <NA>
                        13                          0
```

```r
prop.table(table(region$region))
```

```
                    Africa                       Asia
                0.038348083                 0.038348083
      Australia and New Zealand                  Caribbean
                0.038348083                 0.038348083
             Central America                Central Asia
                0.038348083                 0.038348083
              Eastern Africa                Eastern Asia
                0.038348083                 0.038348083
              Eastern Europe                      Europe
                0.038348083                 0.038348083
Latin America and the Caribbean                  Melanesia
                0.038348083                 0.038348083
               Middle Africa              Northern Africa
                0.038348083                 0.038348083
            Northern America              Northern Europe
                0.038348083                 0.038348083
                    Oceania                    Polynesia
                0.038348083                 0.002949853
          South-eastern Asia               South America
                0.038348083                 0.038348083
              Southern Africa               Southern Asia
                0.038348083                 0.038348083
              Southern Europe               Western Africa
                0.038348083                 0.038348083
                Western Asia               Western Europe
                0.038348083                 0.038348083
                       World
                0.038348083
```

```r
plot_frq(region$region) + labs(title = "Histogram for Regions", x = "Region",
                               y = "Frequency") +
  theme(axis.text.x = element_text(angle = 75, vjust = 1, hjust = 1))
```



The histogram titled "Histogram for Regions" illustrates how often different regions appear in the dataset. It reveals that almost all regions have same frequency count than Polynesia that has 1 count only. The data for region "Asia" is standing out as the leading region. Every region make 7% of data (i.e. number of observations (N=13)) besides Polynesia which makes 0.3% of data (i.e. number of observations (N=1))

Disagg_value

"*disagg_value*" variable represents diet items(fruits,vegetables,beef,etc) in particular region.

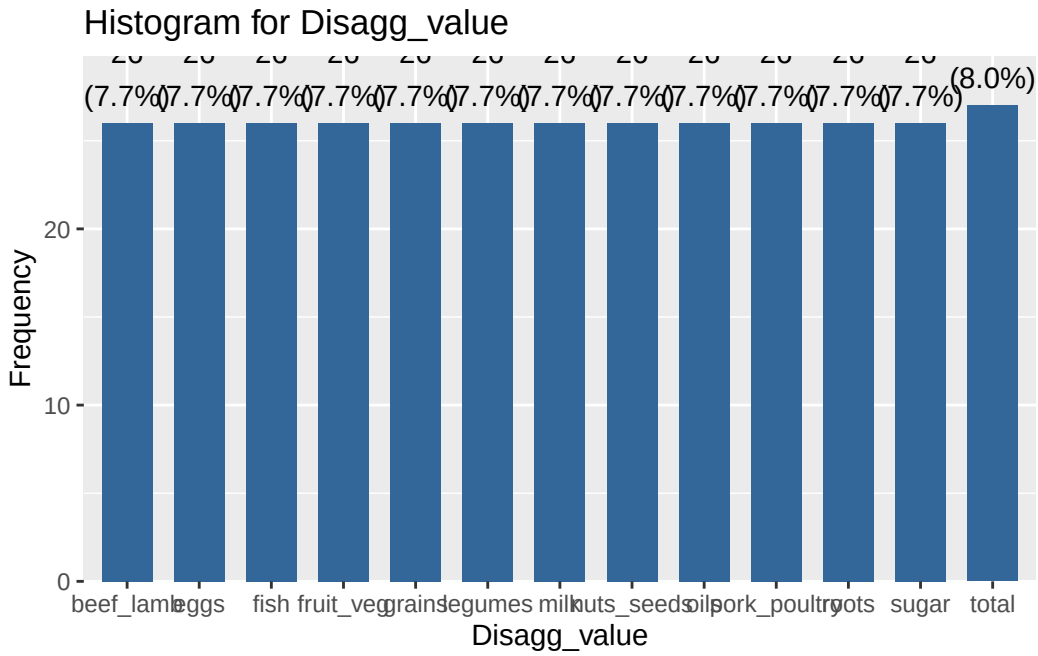
```
table(region$disagg_value,useNA = "always")
```

```
beef_lamb      eggs      fish fruit_veg    grains   legumes
       26        26        26        26        26        26
    milk nuts_seeds      oils pork_poultry   roots     sugar
       26        26        26        26        26        26
   total      <NA>
       27         0
```

```
prop.table(table(region$disagg_value))
```

```
  beef_lamb        eggs        fish   fruit_veg      grains     legumes
0.07669617  0.07669617  0.07669617  0.07669617  0.07669617  0.07669617
       milk   nuts_seeds        oils pork_poultry       roots       sugar
0.07669617  0.07669617  0.07669617  0.07669617  0.07669617  0.07669617
      total
0.07964602
```

```
plot_frq(region$disagg_value) + labs(title = "Histogram for Disagg_value", x = "Disagg_value"
```



The plot illustrates how often different disagg_value (i.e. diet) appear in the dataset. It reveals that almost all disagg_value have same frequency count than total which contains all other disagg_values that are there in that region having 27 which is higher than other. The data for region "total" is standing out as the leading value. Every value makes 7.7% of data (i.e. number of observations (N=26)) besides total which makes 8.0% of data (i.e. number of observations (N=27)).