LESSON 3.. COLUMN SELECTION

NGAIRA

2024-01-20

## R Markdown

# Data Loading

## Rows: 1410 Columns: 11  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): period, county  
## dbl (9): Total Dewormed, Acute Malnutrition, stunted 6-23 months, stunted 0-...  
##   
## 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.

## **COLUMN SELECTION**

## ONLY WORKING WITH THE FIRST 30 COUNTIES

new <-cema[1:30,]  
new <- subset(new, select = -c(period) )

## REPLACE MISSING VALUES WITH ZEROS

new[is.na(new)] = 0

# Count the NAs

# Count the number of missing values in each column  
colSums(is.na(new))

## county Total Dewormed Acute Malnutrition   
## 0 0 0   
## stunted 6-23 months stunted 0-<6 months stunted 24-59 months   
## 0 0 0   
## diarrhoea cases Underweight 0-<6 months Underweight 6-23 months   
## 0 0 0   
## Underweight 24-59 Months   
## 0

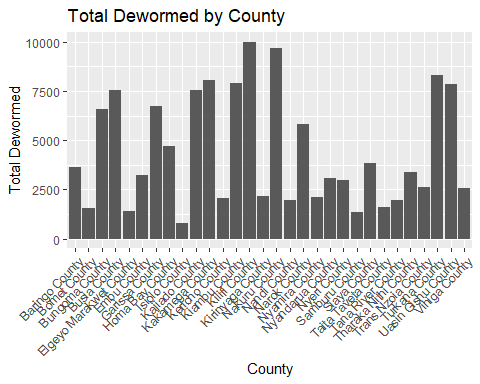
#SUMMARY OF THE NEWLY DEFINED DATAFRAME

new<-as.data.frame(new)  
summary(new)

## county Total Dewormed Acute Malnutrition stunted 6-23 months  
## Length:30 Min. : 790 Min. : 0.00 Min. : 1.00   
## Class :character 1st Qu.:2075 1st Qu.: 0.25 1st Qu.: 97.25   
## Mode :character Median :3311 Median : 25.00 Median : 161.50   
## Mean :4437 Mean : 59.43 Mean : 272.60   
## 3rd Qu.:7337 3rd Qu.: 78.75 3rd Qu.: 302.25   
## Max. :9991 Max. :250.00 Max. :1690.00   
## stunted 0-<6 months stunted 24-59 months diarrhoea cases  
## Min. : 3.00 Min. : 0.00 Min. : 906   
## 1st Qu.: 37.25 1st Qu.: 17.50 1st Qu.:1665   
## Median : 91.00 Median : 41.50 Median :2349   
## Mean :118.63 Mean : 112.07 Mean :2970   
## 3rd Qu.:151.25 3rd Qu.: 96.25 3rd Qu.:4195   
## Max. :505.00 Max. :1194.00 Max. :7420   
## Underweight 0-<6 months Underweight 6-23 months Underweight 24-59 Months  
## Min. : 13.0 Min. : 86.0 Min. : 9.0   
## 1st Qu.: 88.5 1st Qu.: 351.5 1st Qu.: 53.5   
## Median :162.0 Median : 541.0 Median : 115.0   
## Mean :205.9 Mean : 730.2 Mean : 380.8   
## 3rd Qu.:248.8 3rd Qu.: 703.2 3rd Qu.: 263.0   
## Max. :723.0 Max. :3465.0 Max. :3940.0

## NEW DATA VISUALISATION

# \*Plot Total Dewormed by County\*  
ggplot(new, aes(x = county, y = `Total Dewormed`)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Total Dewormed by County", x = "County", y = "Total Dewormed") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



## Scatter plot of Acute Malnutrition vs. Total Dewormed

ggplot(new, aes(x = `Total Dewormed`, y = `Acute Malnutrition`)) +  
 geom\_point() +  
 labs(title = "Scatter plot of Acute Malnutrition vs. Total Dewormed",  
 x = "Total Dewormed", y = "Acute Malnutrition")

