CEMA Internship Task

Humphrey Kinoti

2023-07-20

A copy of this project is in: link to my github repo

Instructions

You have been provided with a dataset which contains monthly data for children <5 years, disaggregated at a county level for the period January 2021 to June 2023.

Dataset description

The dataset contains the following variables:

- `period`: months from January 2021 to June 2023
- `county`: the 47 counties in Kenya
- `Total Dewormed`: Total number of children dewormed
- `Acute Malnutrition`: Number of children <5 years with acute malnutrition
- `stunted 0-6 months`, `stunted 6-23 months`, `stunted 24-59 months`: Number of children stunted
- `diarrhea cases`: Number of children <5 years with diarrhea
- `underweight 0-6 months`, `underweight 6-23 months`, `underweight 24-59 months`: Number of children who are underweight

Objectives

Your task is to: -

- Conduct exploratory data analysis
- State an appropriate research question you would want to answer from the data
- Carry out appropriate data analysis to address the question you have stated above

Explanatory Data Analysis

Import Data

As the data is in CSV format, we will import data using `read_csv()` function in the `readr` package.

```
library(tidyverse)
cema_data <- read_csv("data/cema_internship_task_2023.csv")
```

After the importation, Let's have a look of how the data is structured to have a better understanding of the data. In this case, I will use the `skim` function from `skimr` package.

```r
library(skimr)
skim(cema_data) %>%
  dplyr::select(skim_type, skim_variable, n_missing,
                character.min, character.max, character.empty,
                character.n_unique,
                numeric.mean, numeric.sd, numeric.p0, numeric.p25, numeric.p50, numeric.p75, numeric.p1
```

Table 1: Data summary

Name	cema_data
Number of rows	1410
Number of columns	11
Column type frequency:	
character	2
numeric	9
Group variables	None

Variable type: character

skim_variable	n_missing	min	max	empty	n_unique
period	0	6	6	0	30
county	0	11	22	0	47

Variable type: numeric

skim_variable	n_missing	mean	sd	p0	p25	p50	p75	p100
Total Dewormed	0	11457.92	25372.43	97	2454.50	4564.5	8222.50	392800
Acute Malnutrition	355	125.40	266.49	1	15.00	39.0	143.50	4123
stunted 6-23 months	11	280.16	380.55	1	69.50	159.0	328.50	4398
stunted 0-<6 months	19	139.79	280.24	1	36.50	84.0	157.00	7900
stunted 24-59 months	14	110.77	193.40	1	22.00	50.0	114.25	3169
diarrhoea cases	0	2813.38	2161.90	198	1464.25	2158.0	3335.25	15795
Underweight 0-<6 months	0	223.47	228.53	6	87.00	162.5	272.75	1937
Underweight 6-23 months	0	652.26	669.58	16	249.00	456.0	791.75	5348
Underweight 24-59 Months	0	305.74	538.46	1	51.25	120.5	311.00	4680