STEVE_ODETTE_GEORGE

2023-07-21

libraries

```
library('tidyverse')

library('ggplot2')
install.packages("viridis")

library(sf)
library(ggplot2)
library(viridis)
```

Load data

```
data <- read.csv("path/to/file.csv")

epidemiology_df <- read.csv('cema_internship_task_2023.csv')
```

EXPLORATORY DATA ANAYSIS

RESEARCH QUESTION

Let us get a summary of the data to help provide a leading research question

```
summary(epidemiology_df)
```

From the summary:

There are missing values that will need to be addressed before further analysis. These include:

- 11 missing values in the acute malnutrition column
- 19 missing values in the Stunted Children (6-23 months) column
- 355 missing values in the Diarrhoea Cases
- Missing values in the underweight columns
- In removing the missing values:

```
epidemiology_df1 <- na.omit(epidemiology_df)

summary(epidemiology_df1)
```

Based on the above summary:

- There seem to be a wide variation in the number of children dewormed; from a in of 616, to a max of 392800, anda mean of 13014.
- This points to a large variation accross counties and time periods
- There seem to be a possible wide variation in children with acute mulnutrition, from 1 to over 4,000, with an average of around 125.

- Clearly, some counties have higher levels/rates of accute mulnutrition than others. Some time periods as well
- Similar patterns as seen above have been replicated in number of stunted, underweight and diarrhea cases.
- Based on these eagles eyeview;

Primary question: What is the trend in the number of children <5 years with acute malnutrition across Kenyan counties from January 2021 to June 2023?

Rationale: If acute malnutrition is a significant problem in Kenya and policy and interventions are governmnent priorities:

We can identify hotspots for acute mulnutrition and put more efforts there.

Secondary question:What is the trend in the number of children <5 years dewormed across Kenyan counties from January 2021 to June 2023?

And if the primary question is coupled with the secondary question, we can judge to what extent do we need to go with the access to deworming?

-----EDA----

Let us get a sneak peak of how the variables play per county first

```
dewormed_bar <- ggplot(epidemiology_df1, aes(x = Acute.Malnutrition, y = county)) +

  geom_bar(stat = "identity", fill = "steelblue") +

  xlab("Total Dewormed") +

  ylab("County") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +

  ggtitle("Total Dewormed per County")

dewormed_bar

# NOTES: Bomet, kisii, Bungoma, Baringo, Nyamira, Nyandarua, Siaya, Taita Taveta, and Uasin Gishu count

mulnutrition_bar <- ggplot(epidemiology_df1, aes(x = Acute.Malnutrition, y = county)) +

  geom_bar(stat = "identity", fill = "steelblue") +

  xlab("Acute malnutrition") +

  ylab("County") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +

  ggtitle("Cases of acute mulnutrition per County")

mulnutrition_bar
```

```r
# NOTES: We have more cases of acute malnutrition in Wajir county with significantly few cases in count

diarhe_bar <- ggplot(epidemiology_df1, aes(x = diarrhoea.cases, y = county)) +

  geom_bar(stat = "identity", fill = "steelblue") +

  xlab("Diarrhoea cases") +

  ylab("County") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +

  ggtitle("Cases of diarrhoea per County")

diarhe_bar

# NOTES: We have more cases of diarrhea in Nairobi, county. Other leading counties are Turkana, Nakuru,

# Counties that had lower rates of deworming have relatively lower rates of diarrhoea.


# HOW IS STUNTED GROWTH DISTRIBUTED ACCROSS THE COUNTIES?

stunted_data <- epidemiology_df1 %>%

  select(county, stunted.0..6.months, stunted.6.23.months, stunted.24.59.months)

stunted_data_long <- stunted_data %>%

  gather(age_group, value, -county)

ggplot(stunted_data_long, aes(x = value, y = county, fill = age_group)) +

  geom_bar(stat = "identity") +

  xlab("County") +

  ylab("Number of Stunted Children") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +

  ggtitle("Distribution of Stunted Growth Across Age Groups per County") +

  scale_fill_manual(values = c("#f9a8a8", "#f3cd6a", "#9fcc8e"),

                    labels = c("0-6 months", "6-23 months", "24-59 months"))
```

NOTES: There variations in how ages groups have a stunted growth accross the counties. In Nairobi (had the highest rates of diarrhea), for instance, the rates of stunted growth is high among those aged 24-59 months than 0-6 months, or 6-23 months. In Wajir county (had the highest rates of deworming, malnutrition), the highest rates of stunted growth are among those aged 6-23 months.

Nontheless, in most counties, there are higher rates of studented growth among children aged 24-59 months; this should be carefully noted for government interventions or additional studies

HOW ABOUT UNDERWEIGHT

```
underweight_data <- epidemiology_df1 %>%

  select(county, Underweight.0..6.months, Underweight.6.23.months, Underweight.24.59.Months)

stunted_data_long <- stunted_data %>%

  gather(age_group, value, -county)

ggplot(stunted_data_long, aes(x = value, y = county, fill = age_group)) +

  geom_bar(stat = "identity") +

  xlab("County") +

  ylab("Number of Stunted Children") +

  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +

  ggtitle("Distribution of Underweight Across Age Groups per County") +

  scale_fill_manual(values = c("#ff7f00", "#1f78b4", "#33a02c"),

                    labels = c("0-6 months", "6-23 months", "24-59 months"))

underweight_data
```

NOTES: A similar distribition as seen in stunted growth is depicted in underweight. More children aged 25-59 months are underweight compared to others. The case in Wajir is the same as seen in stunted growth.

```
shapefile_path <- "shapefiles"
gdf <- st_read(shapefile_path)

merged_data <- merge(gdf, df, by.x = "County", by.y = "County")
ggplot() +
  geom_sf(data = merged_data, aes(fill = Acute_Malnutrition)) +
  scale_fill_viridis() +
  theme_minimal() +
  labs(title = "Geographical Distribution of Acute Malnutrition per County")
```

```
shapefile_path <- "/home/odette/MDR_TB/STEVE_ODETTE_GEORGE/shapefiles"
gdf <- st_read(shapefile_path)

gdf$Name <- county_mapping$epidemiology_df_name[match(gdf$Name, county_mapping$gdf_name)]
county_mapping <- data.frame(
  gdf_name = c(
    "Mombasa", "Kwale", "Kilifi", "Tana River", "Lamu", "Taita Taveta", "Garissa",
    "Wajir", "Mandera", "Marsabit", "Isiolo", "Meru", "Tharaka Nithi", "Embu",
    "Kitui", "Machakos", "Makueni", "Nyandarua", "Nyeri", "Kirinyaga", "Muranga",
    "Kiambu", "Turkana", "West Pokot", "Samburu", "Trans Nzoia", "Uasin Gishu",
    "Elgeyo Marakwet", "Nandi", "Baringo", "Laikipia", "Nakuru", "Narok", "Kajiado",
    "Kericho", "Bomet", "Kakamega", "Vihiga", "Bungoma", "Busia", "Siaya", "Kisumu",
    "Homa Bay", "Migori", "Kisii", "Nyamira", "Nairobi"
```

```
  ),
  epidemiology_df_name = c(
    "Mombasa County", "Kwale County", "Kilifi County", "Tana River County",
    "Lamu County", "Taita Taveta County", "Garissa County", "Wajir County",
    "Mandera County", "Marsabit County", "Isiolo County", "Meru County",
    "Tharaka Nithi County", "Embu County", "Kitui County", "Machakos County",
    "Makueni County", "Nyandarua County", "Nyeri County", "Kirinyaga County",
    "Muranga County", "Kiambu County", "Turkana County", "West Pokot County",
    "Samburu County", "Trans Nzoia County", "Uasin Gishu County",
    "Elgeyo Marakwet County", "Nandi County", "Baringo County", "Laikipia County",
    "Nakuru County", "Narok County", "Kajiado County", "Kericho County",
    "Bomet County", "Kakamega County", "Vihiga County", "Bungoma County",
    "Busia County", "Siaya County", "Kisumu County", "Homa Bay County",
    "Migori County", "Kisii County", "Nyamira County", "Nairobi County"
  )
)

mean_malnutrition <- epidemiology_df1 %>%
  group_by(county) %>%
  summarise(mean_malnutrition = sum(Acute.Malnutrition, na.rm = TRUE))
mean_malnutrition
gdf_sf <- st_as_sf(gdf)

gdf_with_malnutrition <- left_join(gdf_sf, mean_malnutrition, by = c("Name" = "county"))

ggplot() +
  geom_sf(data = gdf_with_malnutrition, aes(fill = mean_malnutrition)) +
  scale_fill_viridis() +
  theme_minimal() +
  labs(title = "Acute Malnutrition by County")
gdf_with_malnutrition
```

The above map shows how malnutrition is distributed accross various counties.