# Is Access to Electricity Linked to Hunger in Guatemala?

Chiara Brust

2023-03-23

## Introduction

The focus of this project is to see if there's a correlation between access (or lack thereof) to electricity and hunger in Guatemala. My first thoughts are that there will be an association since lack of access to a power grid could also indicate lack of access to a stable food supply. However, I'm unsure if this hypothesis is correct because it's just as likely that rural areas that are disconnected from the central power grid will grow their own food and be perfectly prepared agriculturally. Therefore, this project seeks to understand if there is an association in mean access to electricity and population hunger indicators by department.

## Data

This data comes from the SDG Data Alliance website (https://www.sdg.org), which acts as a repository of all available country data relating to the Sustainable Development Goals. I chose data from Guatemala and picked data related to achieving the SDGs Zero Hunger and Affordable and Clean Energy. These files include information on the percent of people in each region within Guatemala who have access to electricity along with the percent of people with hunger indicators such as anemia. In addition, the data contains shapefiles for the country, along with GeoIDs for each department within Guatemala.

The following GitHub repository contains all the files used, including the shapefiles and statistical data: https://github.com/kiki852/N741-Final-Project.git

```r
# Load libraries
library(tidyverse) # Manipulating data
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2

## Warning: package 'ggplot2' was built under R version 4.2.2

## Warning: package 'dplyr' was built under R version 4.2.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(readr) # Importing .csv files
library(sf) # Working with shapefiles
```

```
## Linking to GEOS 3.9.1, GDAL 3.4.3, PROJ 7.2.1; sf_use_s2() is TRUE
```

```r
library(tmap) # Plotting
```

```
## Warning: package 'tmap' was built under R version 4.2.2
```

```r
library(knitr) # For creating a table while knitting
```

```
## Warning: package 'knitr' was built under R version 4.2.2
```

```r
library(kableExtra) # For modifying kable tables
```

```
## Warning: package 'kableExtra' was built under R version 4.2.2
```

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 3 > 1' in coercion to 'logical(1)'
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

## Bringing in the hunger and electricity data

```r
# Import data
hunger_data <- read_csv("Inputs/Zero_Hunger.csv")
```

```
## Rows: 22 Columns: 27
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): GlobalID, GeoID, GeoLevel, SURVEY
## dbl  (9): OBJECTID, Reporting_Year, Shape__Area, Shape__Length, SH_STA_ANEM,...
## lgl (14): GeoName, SH_STA_ANEM_PREG_REPORTINGTYPE_R_P, SH_STA_ANEM_REPORTING...
##
## 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.
```

```r
energy_data <- read.csv("Inputs/Affordable_and_Clean_Energy.csv")


# View data
summary(hunger_data)
```

```
##    OBJECTID       GlobalID         Reporting_Year    GeoID
##  Min.   : 1.00   Length:22          Min.   :2015   Length:22
##  1st Qu.: 6.25   Class :character   1st Qu.:2015   Class :character
##  Median :11.50   Mode  :character   Median :2015   Mode  :character
##  Mean   :11.50                      Mean   :2015
##  3rd Qu.:16.75                      3rd Qu.:2015
##  Max.   :22.00                      Max.   :2015
##    GeoLevel          GeoName        Shape__Area        Shape__Length
##  Length:22          Mode:logical   Min.   :5.757e+08   Min.   : 137440
##  Class :character   NA's:22        1st Qu.:2.048e+09   1st Qu.: 260642
##  Mode  :character                  Median :2.747e+09   Median : 341954
##                                    Mean   :5.378e+09   Mean   : 399720
##                                    3rd Qu.:4.576e+09   3rd Qu.: 450489
##                                    Max.   :3.947e+10   Max.   :1096371
##     SURVEY           SH_STA_ANEM    SH_STA_ANEM_PREG
##  Length:22          Min.   : 8.50   Min.   : 8.50
##  Class :character   1st Qu.:11.95   1st Qu.:11.95
##  Mode  :character   Median :15.70   Median :15.70
##                     Mean   :15.10   Mean   :15.10
##                     3rd Qu.:17.20   3rd Qu.:17.20
##                     Max.   :22.10   Max.   :22.10
##  SH_STA_ANEM_PREG_REPORTINGTYPE_R_P SH_STA_ANEM_REPORTINGTYPE_R_P
##  Mode:logical                       Mode:logical
##  NA's:22                            NA's:22
##
##
##
##
##   SH_STA_STNT    SH_STA_STNT_LOCATION_R_P SH_STA_STNT_LOCATION_T_P
##  Min.   :18.70   Mode:logical             Mode:logical
##  1st Qu.:34.58   NA's:22                  NA's:22
##  Median :45.60
##  Mean   :45.65
##  3rd Qu.:55.40
##  Max.   :70.00
##  SH_STA_STNT_LOCATION_U_P  SH_STA_WAST     SH_STA_WAST_LOCATION_R_P
##  Mode:logical             Min.   :0.0000   Mode:logical
##  NA's:22                  1st Qu.:0.5000   NA's:22
##                           Median :0.7000
##                           Mean   :0.7318
##                           3rd Qu.:0.9750
##                           Max.   :1.6000
##  SH_STA_WAST_LOCATION_T_P SH_STA_WAST_LOCATION_U_P SH_STA_WAST_SEX_T_P
##  Mode:logical             Mode:logical             Mode:logical
##  NA's:22                  NA's:22                  NA's:22
##
##
##
##
##   SN_STA_OVWGT   SN_STA_OVWGT_LOCATION_R_P SN_STA_OVWGT_LOCATION_T_P
##  Min.   :2.900   Mode:logical              Mode:logical
##  1st Qu.:3.900   NA's:22                   NA's:22
##  Median :4.700
##  Mean   :4.918
```

```
##   3rd Qu.:5.675
##   Max.   :8.500
##   SN_STA_OVWGT_LOCATION_U_P SN_STA_OVWGT_SEX_T_P
##   Mode:logical              Mode:logical
##   NA's:22                   NA's:22
##
##
##
##
```

```
summary(energy_data)
```

```
##     OBJECTID       GlobalID          Reporting_Year    GeoID
##   Min.   : 1.00   Length:22          Min.   :2015   Length:22
##   1st Qu.: 6.25   Class :character   1st Qu.:2015   Class :character
##   Median :11.50   Mode  :character   Median :2015   Mode  :character
##   Mean   :11.50                      Mean   :2015
##   3rd Qu.:16.75                      3rd Qu.:2015
##   Max.   :22.00                      Max.   :2015
##    GeoLevel          GeoName         Shape__Area        Shape__Length
##   Length:22          Mode:logical   Min.   :5.757e+08   Min.   : 137440
##   Class :character   NA's:22        1st Qu.:2.048e+09   1st Qu.: 260642
##   Mode  :character                  Median :2.747e+09   Median : 341954
##                                     Mean   :5.378e+09   Mean   : 399720
##                                     3rd Qu.:4.576e+09   3rd Qu.: 450489
##                                     Max.   :3.947e+10   Max.   :1096371
##     SURVEY            EG_ACS_ELEC    EG_ACS_ELEC_LOCATION_R_P
##   Length:22          Min.   :50.00   Mode:logical
##   Class :character   1st Qu.:80.75   NA's:22
##   Mode  :character   Median :90.50
##                      Mean   :86.46
##                      3rd Qu.:94.92
##                      Max.   :99.30
##   EG_ACS_ELEC_LOCATION_T_P EG_ACS_ELEC_LOCATION_U_P
##   Mode:logical             Mode:logical
##   NA's:22                  NA's:22
##
##
##
##
```

I'm noticing that in the 'hunger_data' dataframe, the variables 'SH_STA_ANEM' and 'SH_STA_ANEM_PREG' have the exact same summary statistics. We also have several blank columns with no data, so I'll have to clean up this dataframe and select only the variables of interest. I'm only going to work with the `SH_STA_ANEM` variable and assume that this is the correct data for percent of women with anemia whether they are pregnant or not since this information is not specified in the source data webpage. This will have to be noted as potentially incorrect at the end of this analysis though.

```
# Selecting only columns of interest and renaming them into something more meaningful
hunger_data <- hunger_data %>%
  select(4, 10, 14, 18, 23) %>%
  rename(percent_anemic= SH_STA_ANEM,
         percent_child_stunt= SH_STA_STNT,
```

```
         percent_child_wasted= SH_STA_WAST,
         percent_child_ovrwgt= SN_STA_OVWGT)

hunger_data
```

```
## # A tibble: 22 x 5
##    GeoID           percent_anemic percent_child_stunt percent_child_wa~1 perce~2
##    <chr>                    <dbl>               <dbl>              <dbl>   <dbl>
##  1 GUDHS2015410011            9.5                18.7                0.9     3.5
##  2 GUDHS2015410013           20.1                29.1                1.6     4.5
##  3 GUDHS2015410014           12.8                42.4                0.9     8.5
##  4 GUDHS2015410015            8.5                56.5                0.4     5.9
##  5 GUDHS2015410016           15.4                26.9                1.1     2.9
##  6 GUDHS2015410017           18.3                33.6                0.6     4.8
##  7 GUDHS2015410018           16.7                65.6                0       4.6
##  8 GUDHS2015410019           12.4                70                  0.5     4.9
##  9 GUDHS2015410020           16.9                48.8                1       4.8
## 10 GUDHS2015410021           17.2                39.6                1.1     3.9
## # ... with 12 more rows, and abbreviated variable names
## #   1: percent_child_wasted, 2: percent_child_ovrwgt
```

```
energy_data <- energy_data %>%
  select(4, 10) %>%
  rename(percent_with_elec= EG_ACS_ELEC)

energy_data
```

```
##              GeoID percent_with_elec
## 1  GUDHS2015410011              99.3
## 2  GUDHS2015410013              90.5
## 3  GUDHS2015410014              98.2
## 4  GUDHS2015410015              95.3
## 5  GUDHS2015410016              96.4
## 6  GUDHS2015410017              89.4
## 7  GUDHS2015410018              95.2
## 8  GUDHS2015410019              95.7
## 9  GUDHS2015410020              94.1
## 10 GUDHS2015410021              93.9
## 11 GUDHS2015410022              92.8
## 12 GUDHS2015410023              91.6
## 13 GUDHS2015410024              80.7
## 14 GUDHS2015410025              80.9
## 15 GUDHS2015410026              72.2
## 16 GUDHS2015410027              50.0
## 17 GUDHS2015410028              70.3
## 18 GUDHS2015410029              77.8
## 19 GUDHS2015410030              85.7
## 20 GUDHS2015410031              78.7
## 21 GUDHS2015410032              82.9
## 22 GUDHS2015410033              90.5
```

## Bringing in the shapefile

```r
# Import data
guat_sf <- st_read("Inputs/Zero_Hunger.shp")
```

```
## Reading layer `Zero_Hunger' from data source
##   `C:\Users\Chiara\Documents\Graduate School\Spring 2023\Big Data\Project\N741-Final-Project\Inputs\Z
##   using driver `ESRI Shapefile'
## Simple feature collection with 22 features and 27 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -92.24123 ymin: 13.74108 xmax: -88.22414 ymax: 17.81569
## Geodetic CRS:  WGS 84
```

```r
# View the shapefile
guat_sf
```

```
## Simple feature collection with 22 features and 27 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -92.24123 ymin: 13.74108 xmax: -88.22414 ymax: 17.81569
## Geodetic CRS:  WGS 84
## First 10 features:
##    OBJECTID                             GlobalID Reporting_           GeoID
## 1         1 db381b29-73a4-4f95-8ba9-7865ea2a6263       2015 GUDHS2015410011
## 2         2 b2ec4d89-f69f-45b1-ab17-38cc1887ef45       2015 GUDHS2015410013
## 3         3 b0cbbbe9-fe45-479d-a3fa-5eadea18098c       2015 GUDHS2015410014
## 4         4 5d84af77-b576-4669-8ecf-2b68270d20cf       2015 GUDHS2015410015
## 5         5 91fe8183-f463-4a0f-900c-166b6f949899       2015 GUDHS2015410016
## 6         6 eab3f10d-a6e0-4d55-8147-61d74dd2d5c0       2015 GUDHS2015410017
## 7         7 ab805abd-235e-49c1-8e35-538c834feb3f       2015 GUDHS2015410018
## 8         8 3aa3aeb2-8452-4b06-b9cd-e190e2fb7575       2015 GUDHS2015410019
## 9         9 9f37976b-ce13-4e5f-8152-07c470718f24       2015 GUDHS2015410020
## 10       10 62a09829-656f-4517-a30f-7f749b709452       2015 GUDHS2015410021
##              GeoLevel GeoName Shape__Are Shape__Len    SURVEY SH_STA_ANE
## 1  GuatemalaAdmin2015    <NA> 2369652832   286624.7 GU2015DHS        9.5
## 2  GuatemalaAdmin2015    <NA> 1976825316   233570.5 GU2015DHS       20.1
## 3  GuatemalaAdmin2015    <NA>  575705241   137440.1 GU2015DHS       12.8
## 4  GuatemalaAdmin2015    <NA> 2004038172   241296.5 GU2015DHS        8.5
## 5  GuatemalaAdmin2015    <NA> 4821088739   410838.3 GU2015DHS       15.4
## 6  GuatemalaAdmin2015    <NA> 3374850825   359330.0 GU2015DHS       18.3
## 7  GuatemalaAdmin2015    <NA> 1254908638   200752.1 GU2015DHS       16.7
## 8  GuatemalaAdmin2015    <NA> 1160593214   194074.3 GU2015DHS       12.4
## 9  GuatemalaAdmin2015    <NA> 2295227111   347005.3 GU2015DHS       16.9
## 10 GuatemalaAdmin2015    <NA> 2570666925   367114.5 GU2015DHS       17.2
##    SH_STA_A_1 SH_STA_A_2 SH_STA_A_3 SH_STA_STN SH_STA_S_1 SH_STA_S_2 SH_STA_S_3
## 1         9.5       <NA>       <NA>       18.7       <NA>       <NA>       <NA>
## 2        20.1       <NA>       <NA>       29.1       <NA>       <NA>       <NA>
## 3        12.8       <NA>       <NA>       42.4       <NA>       <NA>       <NA>
## 4         8.5       <NA>       <NA>       56.5       <NA>       <NA>       <NA>
## 5        15.4       <NA>       <NA>       26.9       <NA>       <NA>       <NA>
## 6        18.3       <NA>       <NA>       33.6       <NA>       <NA>       <NA>
```

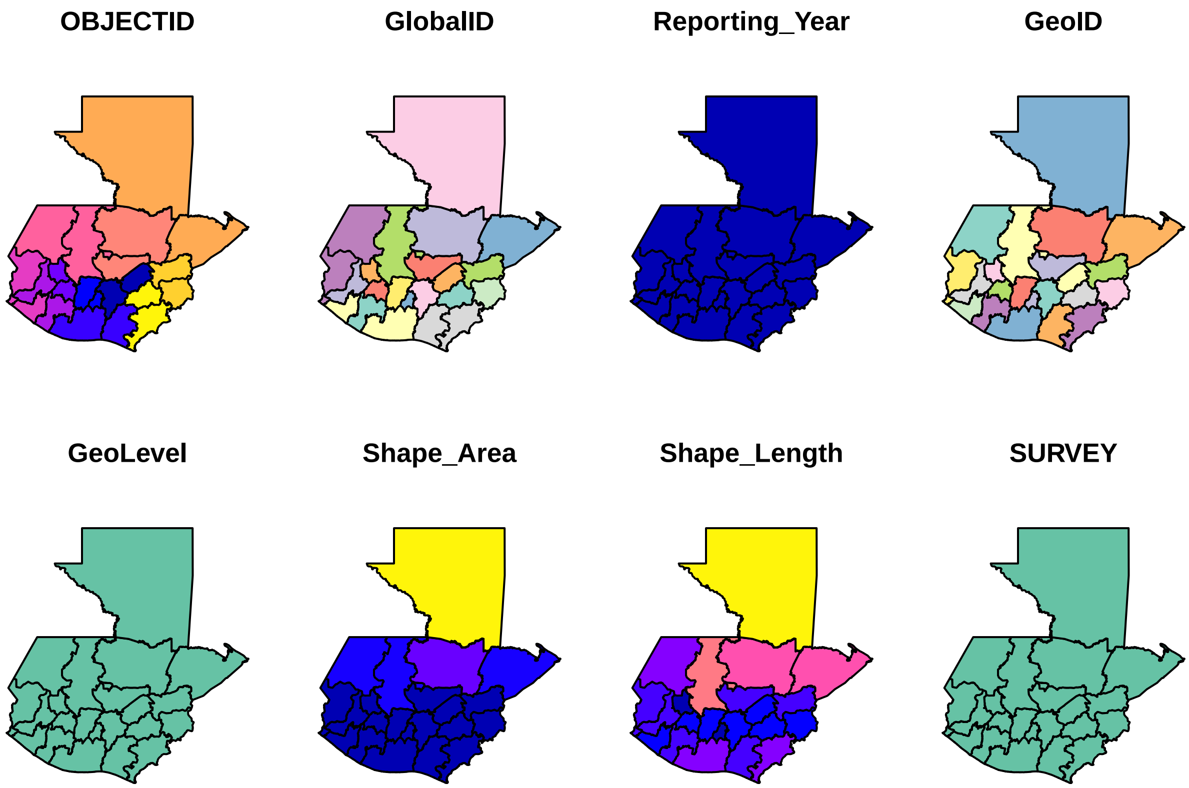
```
## 7        16.7       <NA>       <NA>       65.6       <NA>       <NA>       <NA>
## 8        12.4       <NA>       <NA>       70.0       <NA>       <NA>       <NA>
## 9        16.9       <NA>       <NA>       48.8       <NA>       <NA>       <NA>
## 10       17.2       <NA>       <NA>       39.6       <NA>       <NA>       <NA>
##    SH_STA_WAS SH_STA_W_1 SH_STA_W_2 SH_STA_W_3 SH_STA_W_4 SN_STA_OVW SN_STA_O_1
## 1         0.9       <NA>       <NA>       <NA>       <NA>        3.5       <NA>
## 2         1.6       <NA>       <NA>       <NA>       <NA>        4.5       <NA>
## 3         0.9       <NA>       <NA>       <NA>       <NA>        8.5       <NA>
## 4         0.4       <NA>       <NA>       <NA>       <NA>        5.9       <NA>
## 5         1.1       <NA>       <NA>       <NA>       <NA>        2.9       <NA>
## 6         0.6       <NA>       <NA>       <NA>       <NA>        4.8       <NA>
## 7         0.0       <NA>       <NA>       <NA>       <NA>        4.6       <NA>
## 8         0.5       <NA>       <NA>       <NA>       <NA>        4.9       <NA>
## 9         1.0       <NA>       <NA>       <NA>       <NA>        4.8       <NA>
## 10        1.1       <NA>       <NA>       <NA>       <NA>        3.9       <NA>
##    SN_STA_O_2 SN_STA_O_3 SN_STA_O_4                       geometry
## 1        <NA>       <NA>       <NA> MULTIPOLYGON (((-90.39979 1...
## 2        <NA>       <NA>       <NA> MULTIPOLYGON (((-89.96163 1...
## 3        <NA>       <NA>       <NA> MULTIPOLYGON (((-90.73572 1...
## 4        <NA>       <NA>       <NA> MULTIPOLYGON (((-90.99836 1...
## 5        <NA>       <NA>       <NA> MULTIPOLYGON (((-90.87102 1...
## 6        <NA>       <NA>       <NA> MULTIPOLYGON (((-90.29827 1...
## 7        <NA>       <NA>       <NA> MULTIPOLYGON (((-91.14753 1...
## 8        <NA>       <NA>       <NA> MULTIPOLYGON (((-91.31595 1...
## 9        <NA>       <NA>       <NA> MULTIPOLYGON (((-91.51455 1...
## 10       <NA>       <NA>       <NA> MULTIPOLYGON (((-91.48823 1...
```

Noting that the coordinate reference system has already been set to WGS 84.

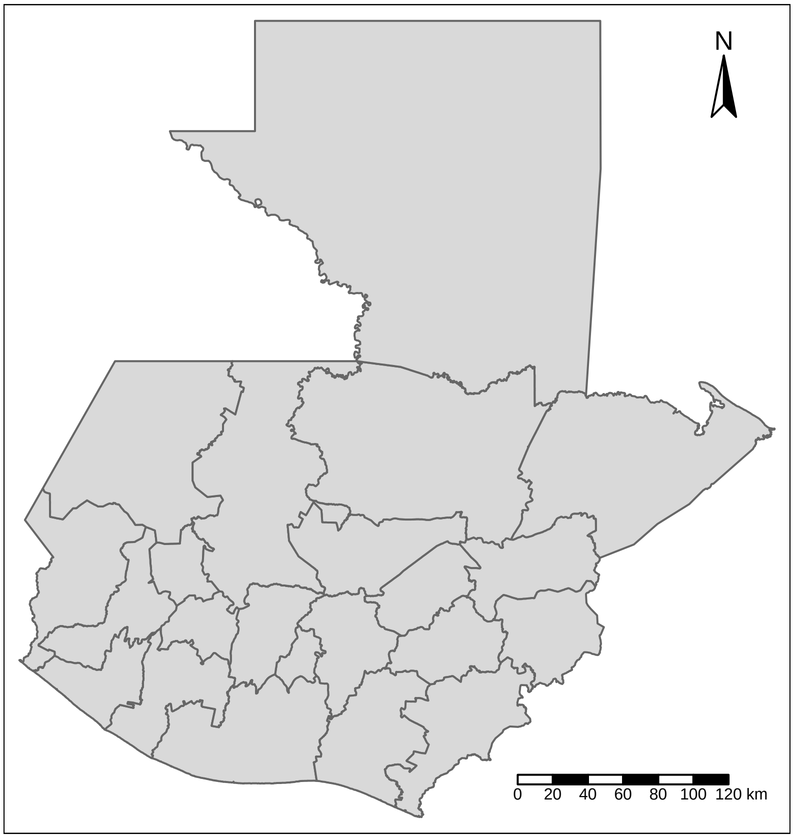
It appears that the hunger data has already been joined with this shapefile, but the variables are not exactly the same. Since I've already cleaned up our 'hunger_data' dataframe, I'm going to select only the shapefile-relevant variables and clean up this 'guat_sf' dataframe. I'm also going to change some of the variable names such as 'shape_are' to match the corresponding variables in 'energy_data' and 'hunger_data'.

```
# Select only variables of interest
guat_sf <- guat_sf %>%
  select(1:5, 7:9, 28) %>%
  rename(Reporting_Year= Reporting_,
         Shape_Area= Shape__Are,
         Shape_Length= Shape__Len)


# Plotting the shapefile variables
plot(guat_sf)
```




```
# Plotting just the country
tm_shape(guat_sf)+
  tm_polygons() +
  tm_compass(
    type= "arrow",
    position= c("right", "top")
  ) +
  tm_scale_bar()
```




## Problems with Department Information

I've realized that there's not a column that lists the department names for each `GeoID` or `OBJECTID`. When searching through the website, there's also no information on which region the GeoIDs correspond to. I looked online and it appears that these GeoIDs are unique for this dataset and have no meaning to external sources, so I will have to plot the shapefile and include the `OBJECTID` label for each region before comparing this to a publicly available map of Guatemala and manually create a column that corresponds to each department and `OBJECTID`.

```
# Compute centroids and create a new data frame
# This is needed to add a label of each `OBJECTID` on top of the middle of each department region
centroids <- st_centroid(guat_sf)
```

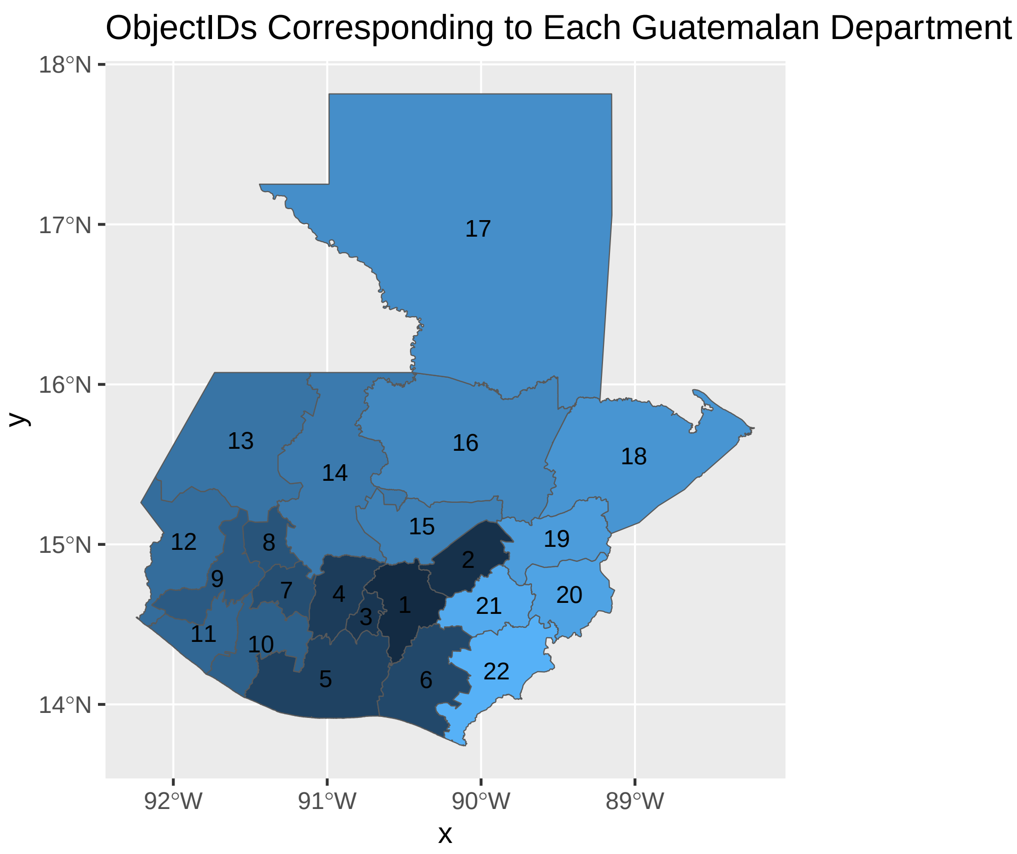
```
## Warning in st_centroid.sf(guat_sf): st_centroid assumes attributes are constant
## over geometries of x
```

```
centroids_df <- data.frame(OBJECTID = guat_sf$OBJECTID,
                           x = st_coordinates(centroids)[, 1],
                           y = st_coordinates(centroids)[, 2])

# Plot the map with `OBJECTID` labels
ggplot() +
  geom_sf(data= guat_sf, aes(fill = OBJECTID)) +
```

```
  geom_text(data = centroids_df, aes(label = OBJECTID, x = x, y = y), size = 3) +
  guides(fill=FALSE) +
  ggtitle("ObjectIDs Corresponding to Each Guatemalan Department")
```

```
## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as
## of ggplot2 3.3.4.
```




Embedding an image of Guatemalan departments from geology.com

## Joining Department values with our shapefile

```
# Create a lookup table of ObjectIDs and department names
lookup_table <- data.frame(
  OBJECTID = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
               12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22),
  Department = c("Guatemala", "El Progreso", "Antigua", "Chimaltenango", "Escuintla", "Cuilapa", "Solola
                 "Totonicapan", "Quetzaltenango", "Mazatenango", "Retalhuleu", "San Marcos", "Huehuetena
                 "Quiche", "Baja Verapaz", "Alta Verapaz", "Peten", "Izabal", "Zacapa", "Chiquimula", "
                 "Jutiapa"))

# Join the lookup table to the shapefile
guat_sf_dpt <- left_join(guat_sf, lookup_table, by = "OBJECTID")
```

## Joining the hunger and electricity data with the shapefile

```
# Joining data
data_sf <- full_join(guat_sf_dpt, energy_data, by= "GeoID") %>%
  full_join(hunger_data, by= "GeoID")

# Check out the data
str(data_sf)
```

```
## Classes 'sf' and 'data.frame':   22 obs. of  15 variables:
##  $ OBJECTID            : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ GlobalID            : chr  "db381b29-73a4-4f95-8ba9-7865ea2a6263" "b2ec4d89-f69f-45b1-ab17-38cc18
##  $ Reporting_Year      : int  2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
##  $ GeoID               : chr  "GUDHS2015410011" "GUDHS2015410013" "GUDHS2015410014" "GUDHS2015410015
##  $ GeoLevel            : chr  "GuatemalaAdmin2015" "GuatemalaAdmin2015" "GuatemalaAdmin2015" "Guatema
##  $ Shape_Area          : num  2.37e+09 1.98e+09 5.76e+08 2.00e+09 4.82e+09 ...
##  $ Shape_Length        : num  286625 233570 137440 241297 410838 ...
##  $ SURVEY              : chr  "GU2015DHS" "GU2015DHS" "GU2015DHS" "GU2015DHS" ...
##  $ Department          : chr  "Guatemala" "El Progreso" "Antigua" "Chimaltenango" ...
##  $ percent_with_elec   : num  99.3 90.5 98.2 95.3 96.4 89.4 95.2 95.7 94.1 93.9 ...
##  $ percent_anemic      : num  9.5 20.1 12.8 8.5 15.4 18.3 16.7 12.4 16.9 17.2 ...
##  $ percent_child_stunt : num  18.7 29.1 42.4 56.5 26.9 33.6 65.6 70 48.8 39.6 ...
##  $ percent_child_wasted: num  0.9 1.6 0.9 0.4 1.1 0.6 0 0.5 1 1.1 ...
##  $ percent_child_ovrwgt: num  3.5 4.5 8.5 5.9 2.9 4.8 4.6 4.9 4.8 3.9 ...
##  $ geometry            :sfc_MULTIPOLYGON of length 22; first list element: List of 1
##   ..$ :List of 1
##   .. ..$ : num [1:1617, 1:2] -90.4 -90.4 -90.4 -90.4 -90.4 ...
##   ..- attr(*, "class")= chr [1:3] "XY" "MULTIPOLYGON" "sfg"
##  - attr(*, "sf_column")= chr "geometry"
##  - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",..: NA NA NA NA NA NA NA NA NA NA ...
##   ..- attr(*, "names")= chr [1:14] "OBJECTID" "GlobalID" "Reporting_Year" "GeoID" ...
```

All looks good. Now it's time to explore the data

## Exploratory Data Analysis

### Summary Statistics

```
data_sf %>%
  st_drop_geometry() %>%
  select(10:14) %>%
  rename(
    `% with Electricity` = percent_with_elec,
    `% Anemic` = percent_anemic,
    `% Children Stunted` = percent_child_stunt,
    `% Children Wasted` = percent_child_wasted,
    `% Children Overweight` = percent_child_ovrwgt
  ) %>%
  summary() %>%
```

Table 1: Summary Statistics for Hunger and Energy in Guatemala

| | % with Electricity | % Anemic | % Children Stunted | % Children Wasted | % Children Overweight |
|---|---|---|---|---|---|
| | Min. :50.00 | Min. : 8.50 | Min. :18.70 | Min. :0.0000 | Min. :2.900 |
| | 1st Qu.:80.75 | 1st Qu.:11.95 | 1st Qu.:34.58 | 1st Qu.:0.5000 | 1st Qu.:3.900 |
| | Median :90.50 | Median :15.70 | Median :45.60 | Median :0.7000 | Median :4.700 |
| | Mean :86.46 | Mean :15.10 | Mean :45.65 | Mean :0.7318 | Mean :4.918 |
| | 3rd Qu.:94.92 | 3rd Qu.:17.20 | 3rd Qu.:55.40 | 3rd Qu.:0.9750 | 3rd Qu.:5.675 |
| | Max. :99.30 | Max. :22.10 | Max. :70.00 | Max. :1.6000 | Max. :8.500 |

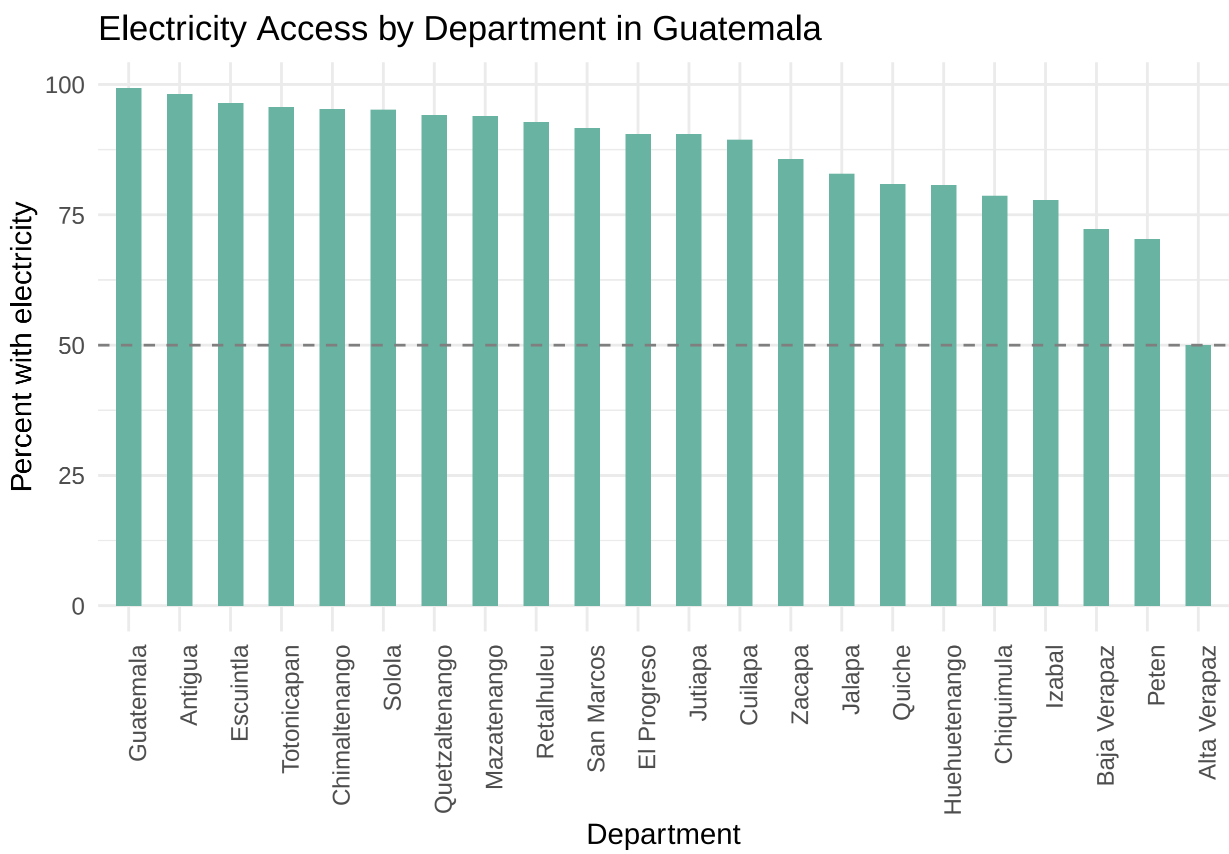
```
  kable(caption = "Summary Statistics for Hunger and Energy in Guatemala") %>%
  kable_styling(latex_options = c("striped", "scale_down"),
                full_width = FALSE)
```

## Comparing Departments

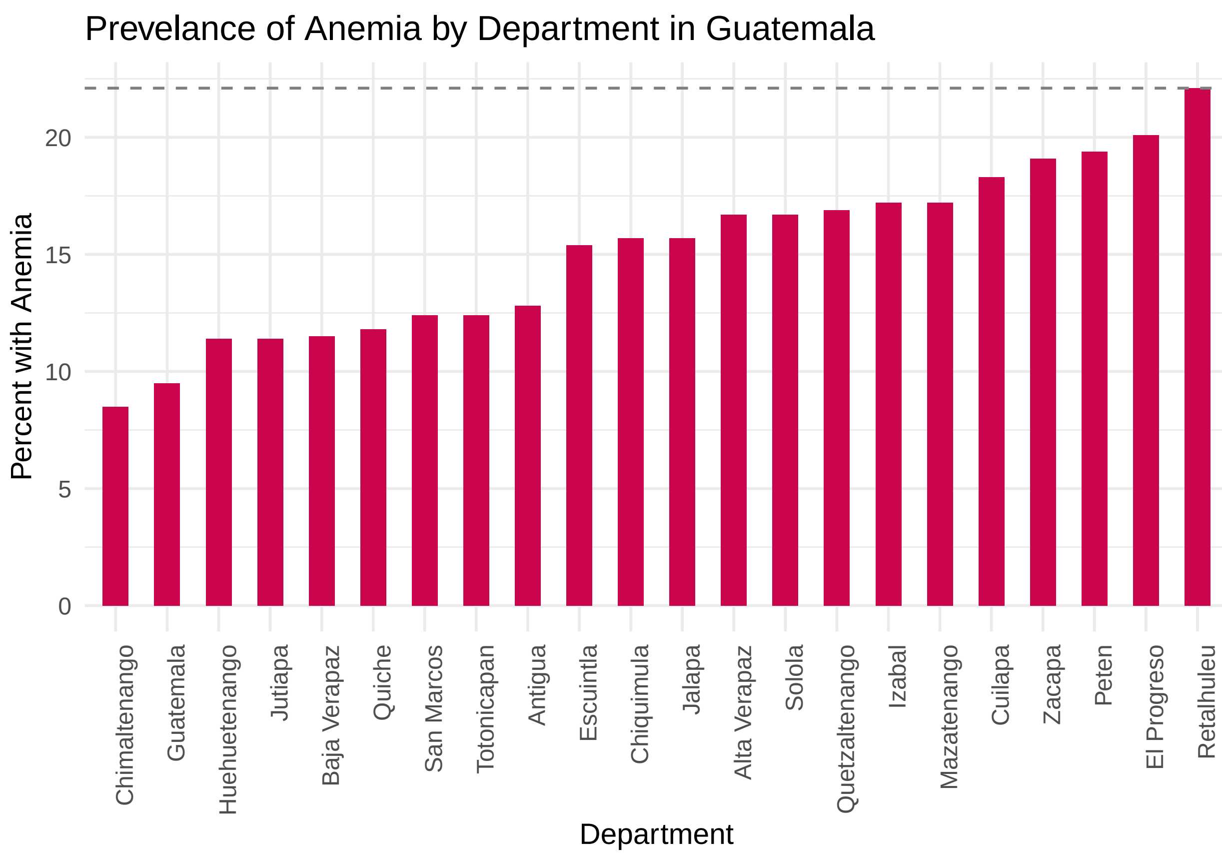
I'm going to create bar graphs of each department with each electricity/hunger variable

```
# Electricity
ggplot(data_sf) +
  geom_bar(aes(x= reorder(Department, -percent_with_elec), y= percent_with_elec),
           stat="identity", fill = "#69b3a2", width = 0.5) +
  labs(x = "Department", y = "Percent with electricity",
       title = "Electricity Access by Department in Guatemala") +
  theme_minimal() +
  geom_hline(yintercept = 50, size = 0.5, color = "grey50", linetype = "dashed") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

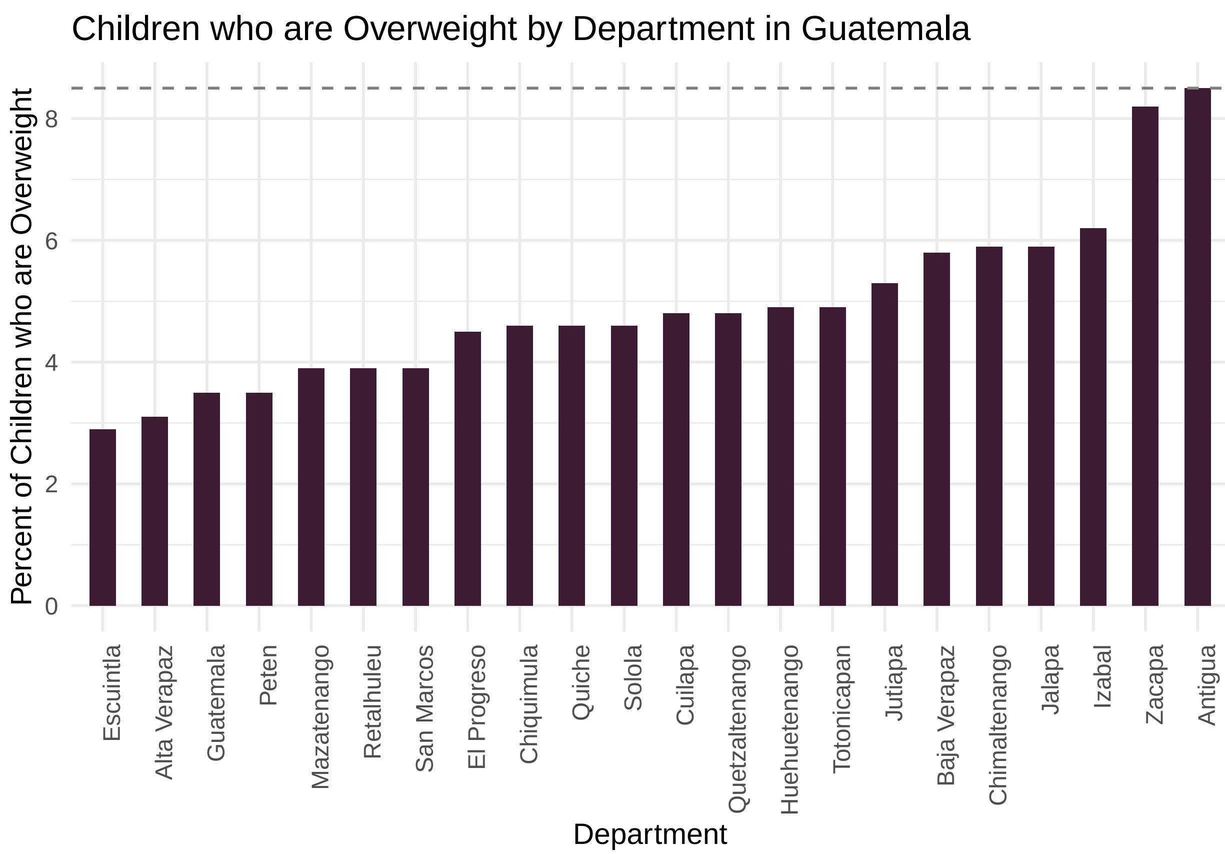
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
```



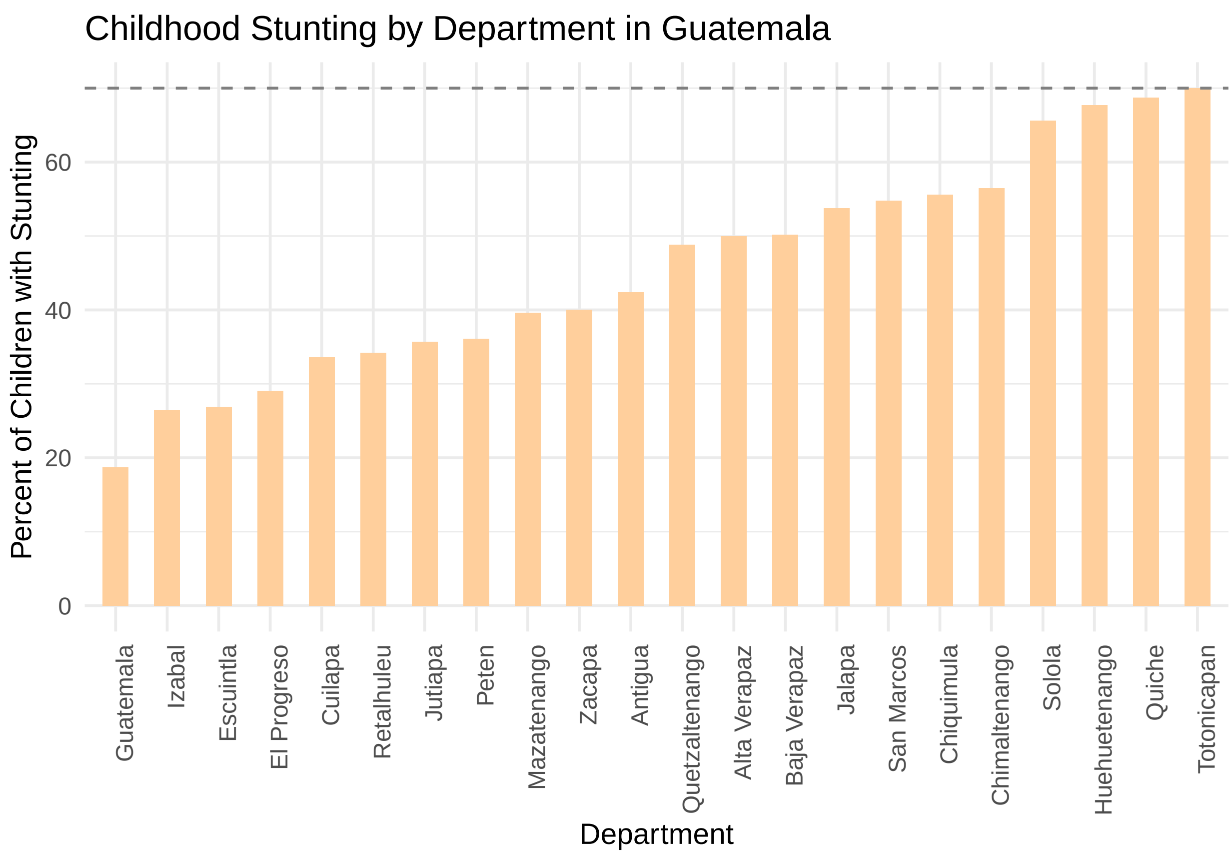

```
# Anemia
ggplot(data_sf) +
  geom_bar(aes(x= reorder(Department, percent_anemic), y= percent_anemic),
           stat="identity", fill = "#CA054D", width = 0.5) +
  labs(x = "Department", y = "Percent with Anemia",
       title = "Prevelance of Anemia by Department in Guatemala") +
  theme_minimal() +
  geom_hline(yintercept = 22.1, size = 0.5, color = "grey50", linetype = "dashed") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



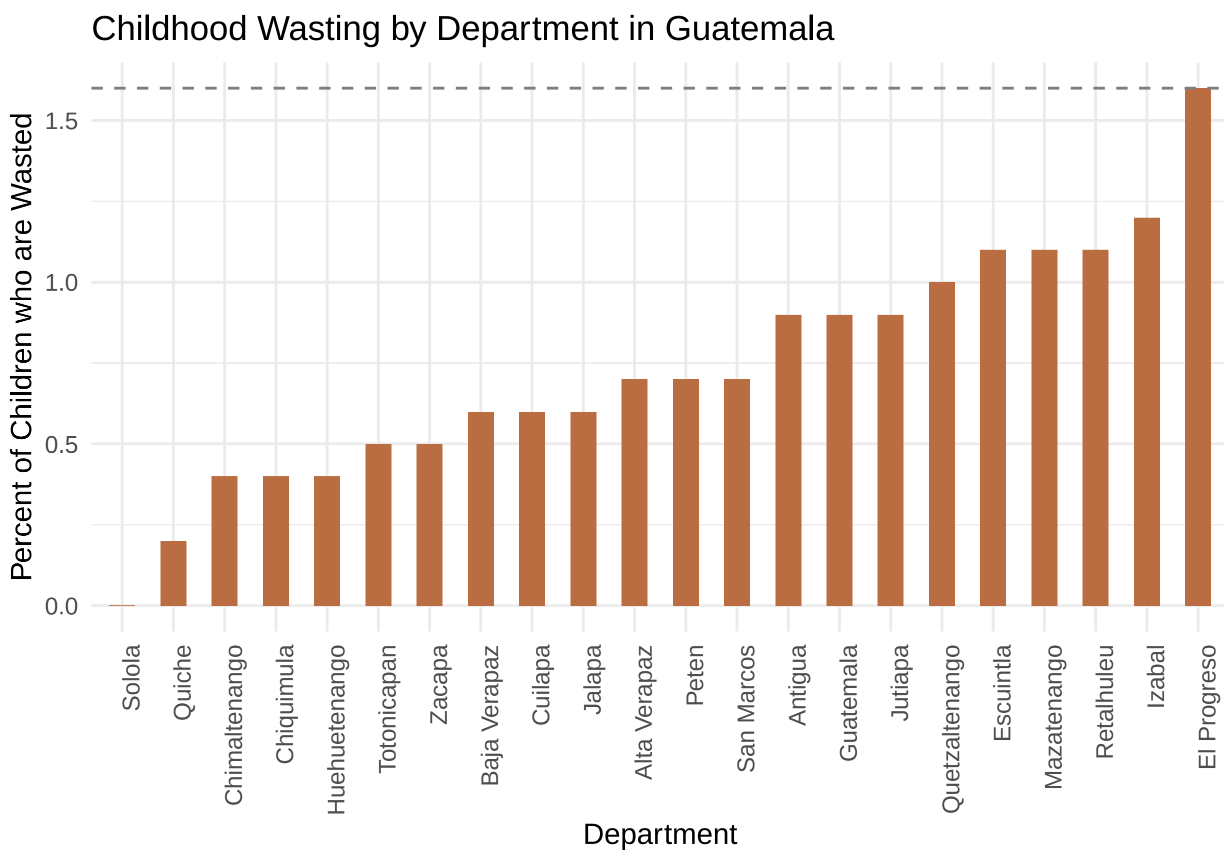

```
# Children Overweight
ggplot(data_sf) +
  geom_bar(aes(x= reorder(Department, percent_child_ovrwgt), y= percent_child_ovrwgt),
           stat="identity", fill = "#3B1C32", width = 0.5) +
  labs(x = "Department", y = "Percent of Children who are Overweight",
       title = "Children who are Overweight by Department in Guatemala") +
  theme_minimal() +
  geom_hline(yintercept = 8.5, size = 0.5, color = "grey50", linetype = "dashed") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```




```r
# Childhood Stunting
ggplot(data_sf) +
  geom_bar(aes(x= reorder(Department, percent_child_stunt), y= percent_child_stunt),
           stat="identity", fill = "#FFCF9C", width = 0.5) +
  labs(x = "Department", y = "Percent of Children with Stunting",
       title = "Childhood Stunting by Department in Guatemala") +
  theme_minimal() +
  geom_hline(yintercept = 70, size = 0.5, color = "grey50", linetype = "dashed") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```




```r
# Childhood Wasting
ggplot(data_sf) +
  geom_bar(aes(x= reorder(Department, percent_child_wasted), y= percent_child_wasted),
           stat="identity", fill = "#B96D40", width = 0.5) +
  labs(x = "Department", y = "Percent of Children who are Wasted",
       title = "Childhood Wasting by Department in Guatemala") +
  theme_minimal() +
  geom_hline(yintercept = 1.6, size = 0.5, color = "grey50", linetype = "dashed") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



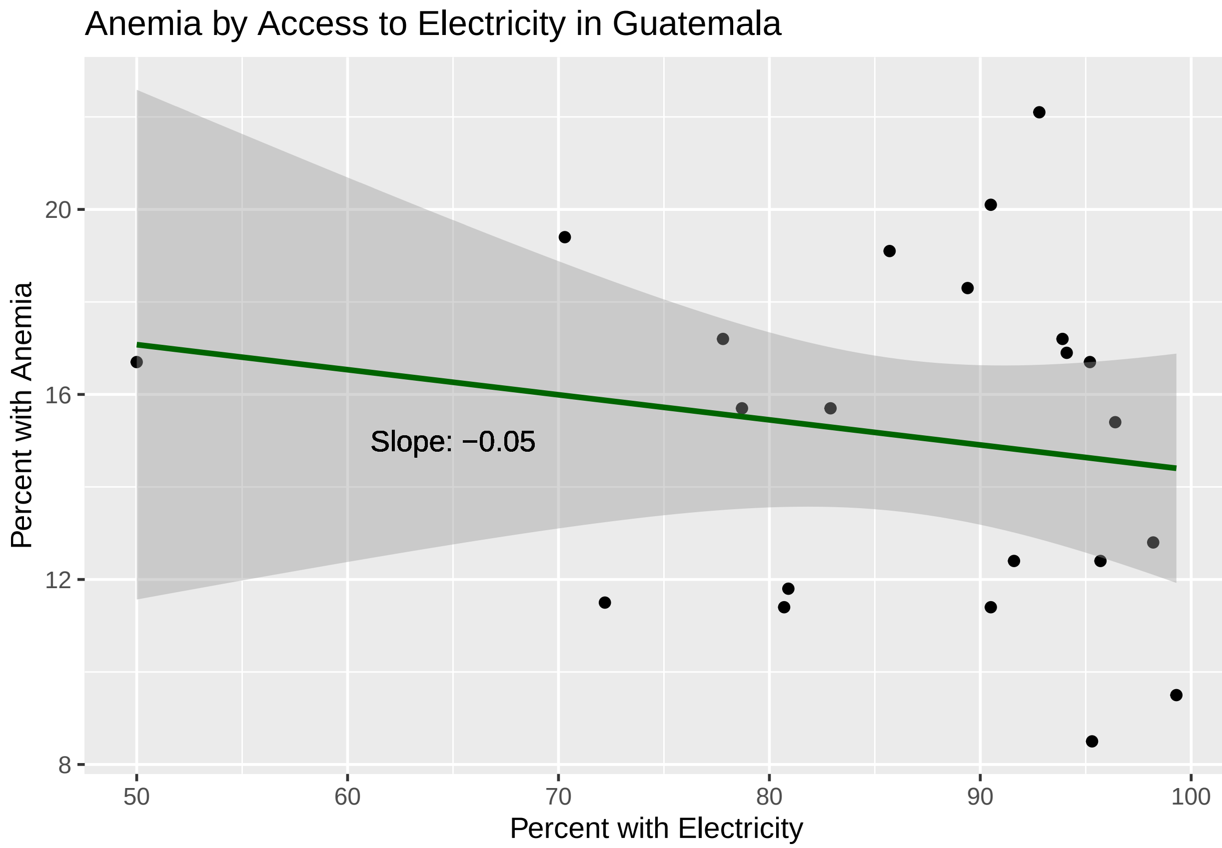

## Exploratory Comparison of Electricity Access with Each Variable of Interest

```
# Electricity and Anemia

# Get line of best fit
fit_anemia <- lm(percent_anemic ~ percent_with_elec, data = data_sf)


# Create the plot
ggplot(data_sf, aes(x = percent_with_elec, y = percent_anemic)) +
  geom_point() +
  geom_smooth(method = "lm", color= "darkgreen") +
  geom_text(x = 65, y = 15, label = paste0("Slope: ", round(coef(fit_anemia)[2], 2))) +
  ggtitle("Anemia by Access to Electricity in Guatemala") +
  labs (x= "Percent with Electricity", y= "Percent with Anemia")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



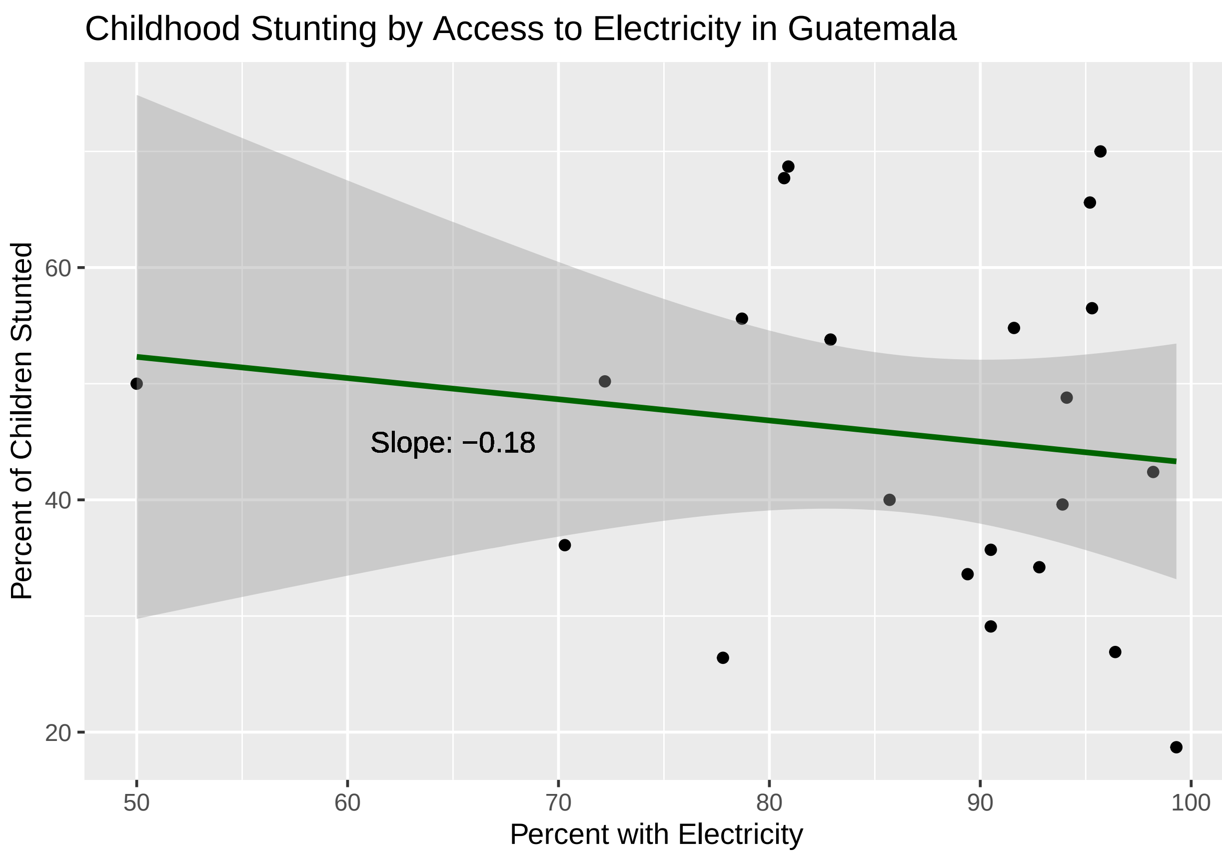

It looks like there is not a solid relationship between access to electricity and prevalence of anemia in this country. If anything, there is a downward trend that shows as more people get access to electricity, the lower the prevalence of anemia.

```
# Electricity and Childhood Stunting

# Get line of best fit
fit_stunt <- lm(percent_child_stunt ~ percent_with_elec, data = data_sf)


# Create the plot
ggplot(data_sf, aes(x = percent_with_elec, y = percent_child_stunt)) +
  geom_point() +
  geom_smooth(method = "lm", color= "darkgreen") +
  geom_text(x = 65, y = 45, label = paste0("Slope: ", round(coef(fit_stunt)[2], 2))) +
  ggtitle("Childhood Stunting by Access to Electricity in Guatemala") +
  labs(x = "Percent with Electricity", y = "Percent of Children Stunted")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



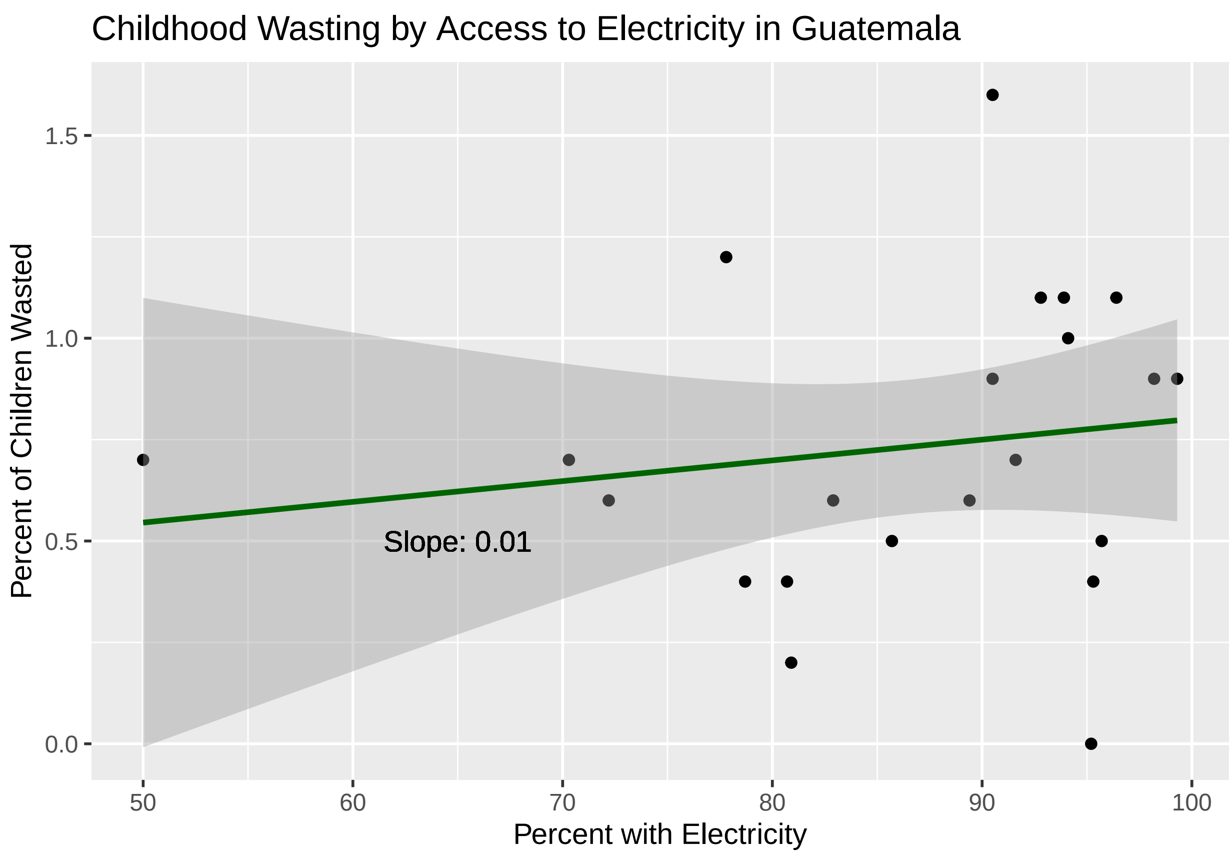

The trends in this plot look very similar to the plot of Anemia and Access to Electricity. There is not a strong relationship between the two variables, but the overall trend is a small downward slope.

```
# Electricity and Childhood Wasting

# Get line of best fit
fit_waste <- lm(percent_child_wasted ~ percent_with_elec, data = data_sf)


# Create the plot
ggplot(data_sf, aes(x = percent_with_elec, y = percent_child_wasted)) +
  geom_point() +
  geom_smooth(method = "lm", color= "darkgreen") +
  geom_text(x = 65, y = 0.5, label = paste0("Slope: ", round(coef(fit_waste)[2], 2))) +
  ggtitle("Childhood Wasting by Access to Electricity in Guatemala") +
  labs(x = "Percent with Electricity", y = "Percent of Children Wasted")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



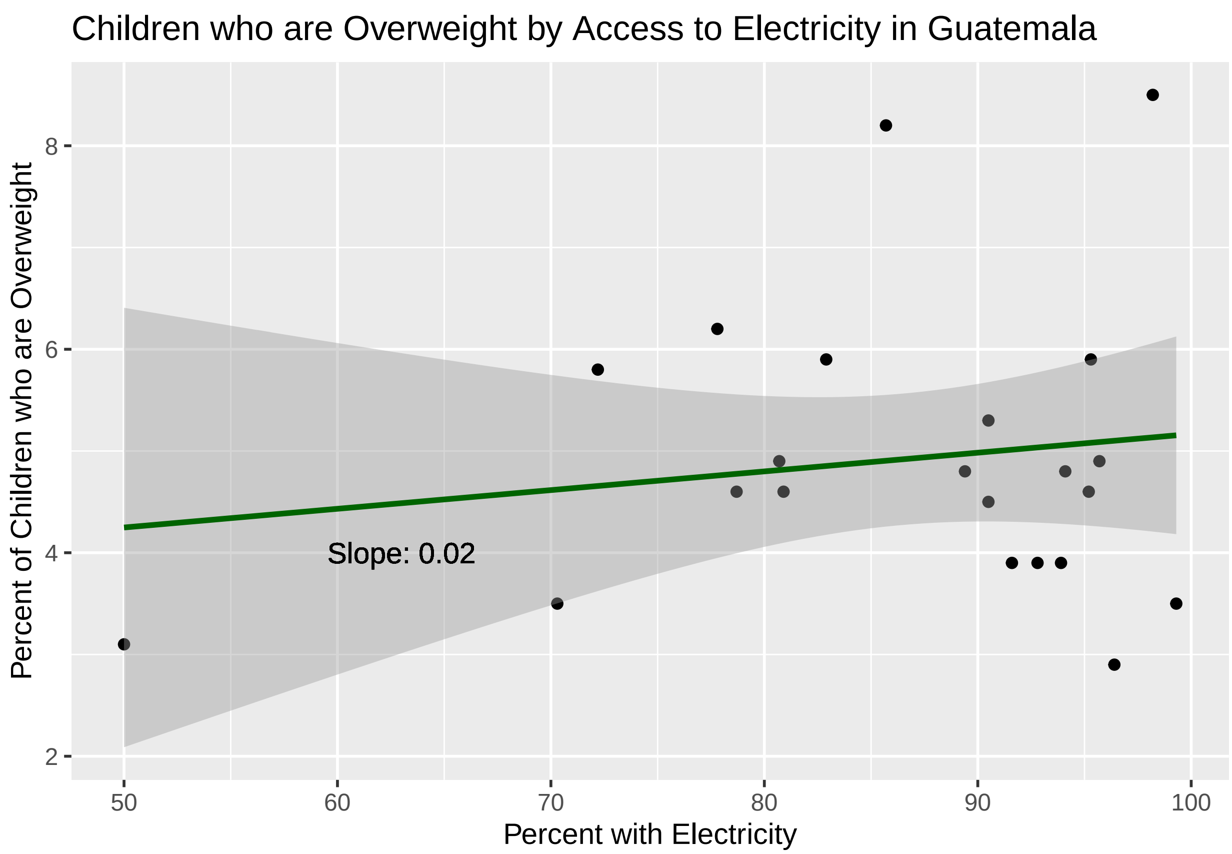

This is interesting- based on this plot, it appears that as access to electricity increases, the percent of children who are wasted also increases (albeit slightly). This is the opposite trend than what I was expecting. Although wasting refers to a short term period of malnutrition that results in a low BMI, stunting is very similar but refers to long term malnutrition instead of short term. Since both of these variables are similar, I would expect them to have similar trends.

```r
# Electricity and Overweight Children

# Get line of best fit
fit_ovrwgt <- lm(percent_child_ovrwgt ~ percent_with_elec, data = data_sf)


# Create the plot
ggplot(data_sf, aes(x = percent_with_elec, y = percent_child_ovrwgt)) +
  geom_point() +
  geom_smooth(method = "lm", color= "darkgreen") +
  geom_text(x = 63, y = 4, label = paste0("Slope: ", round(coef(fit_ovrwgt)[2], 2))) +
  ggtitle("Children who are Overweight by Access to Electricity in Guatemala") +
  labs(x = "Percent with Electricity", y = "Percent of Children who are Overweight")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

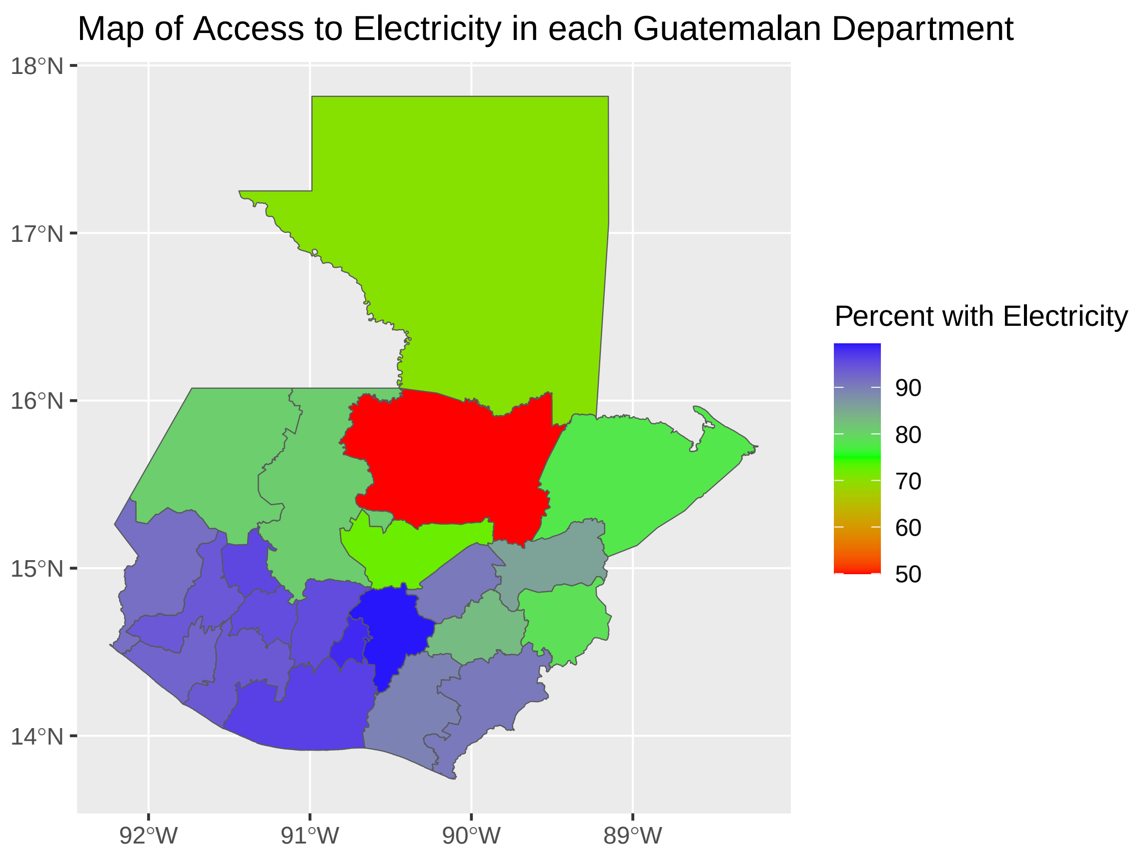



This plot illustrates that there is a slight upward trend between the two variables- as access to electricity increases, the percent of children who are overweight slightly increases as well.

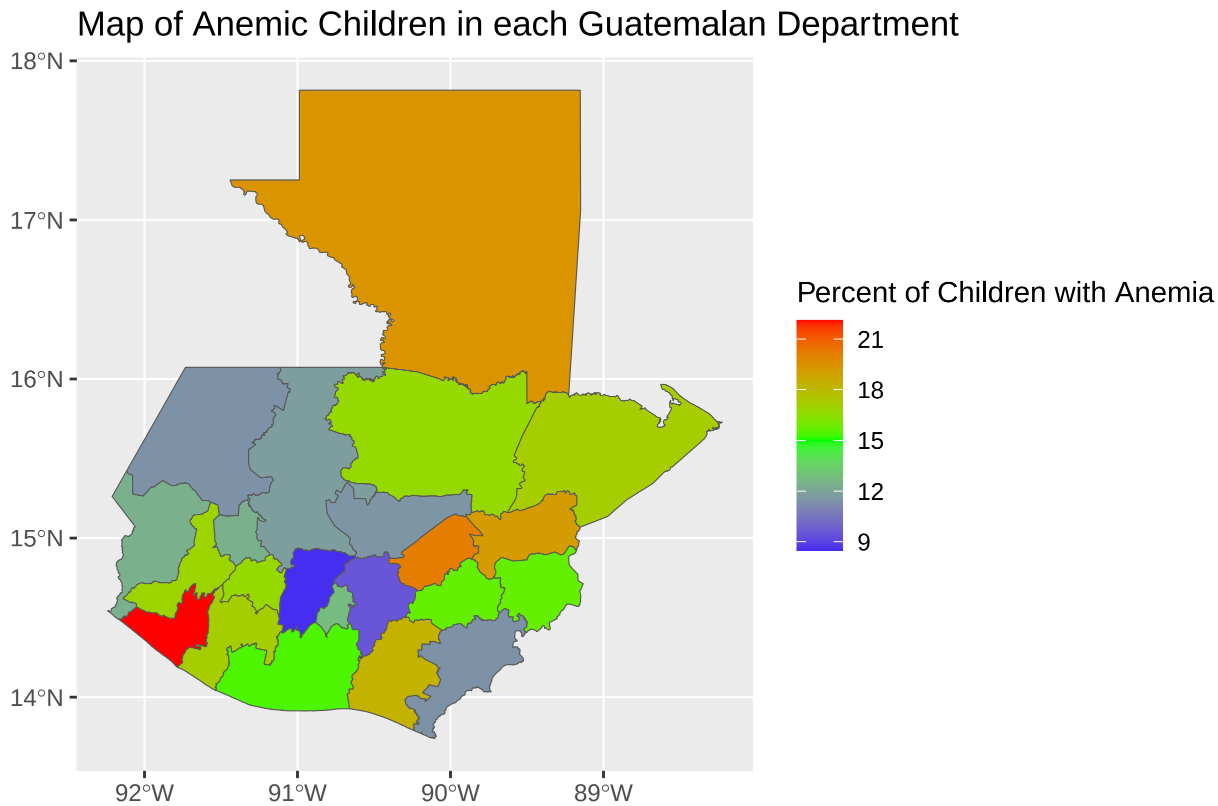
## Spatial Plots

The following are plots of each department's corresponding hunger and electricity indicators.

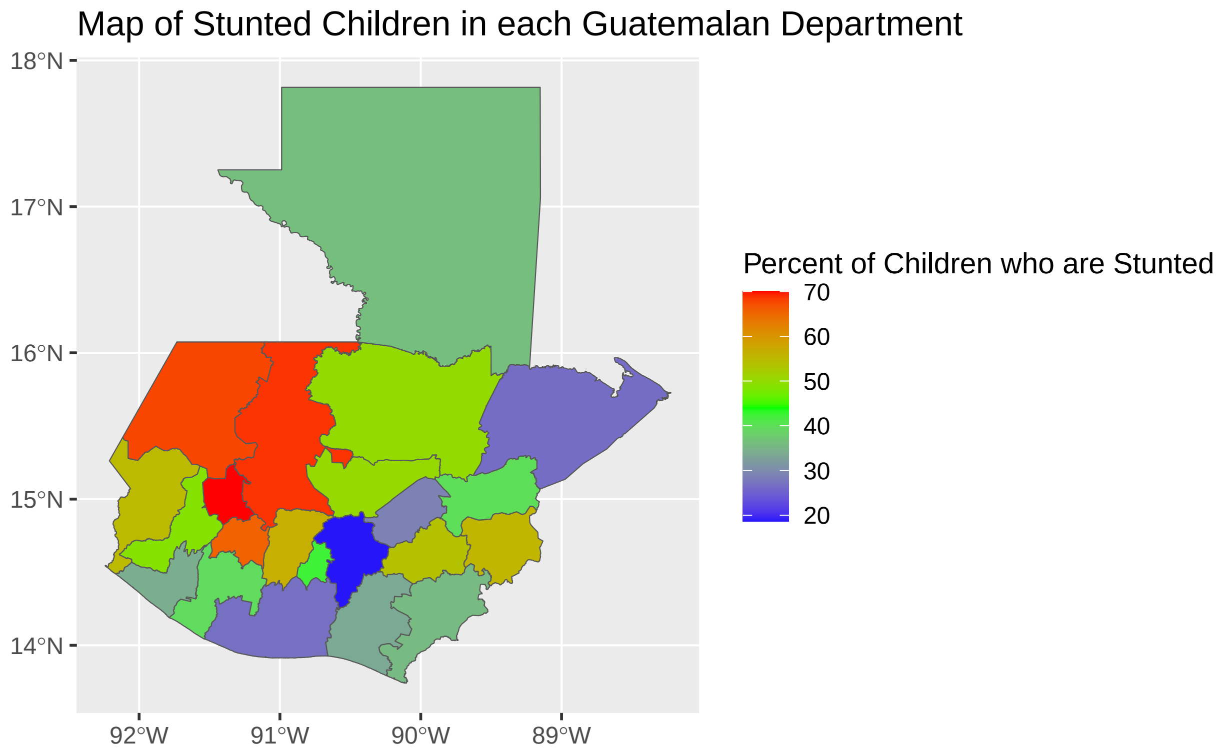
```
# Electricity
ggplot(data_sf) +
  geom_sf(aes(fill = percent_with_elec)) +
  scale_fill_gradient2(low = "red",
                       mid = "green",
                       high = "blue",
                       midpoint=75) +
  ggtitle("Map of Access to Electricity in each Guatemalan Department") +
  labs (fill= "Percent with Electricity")
```



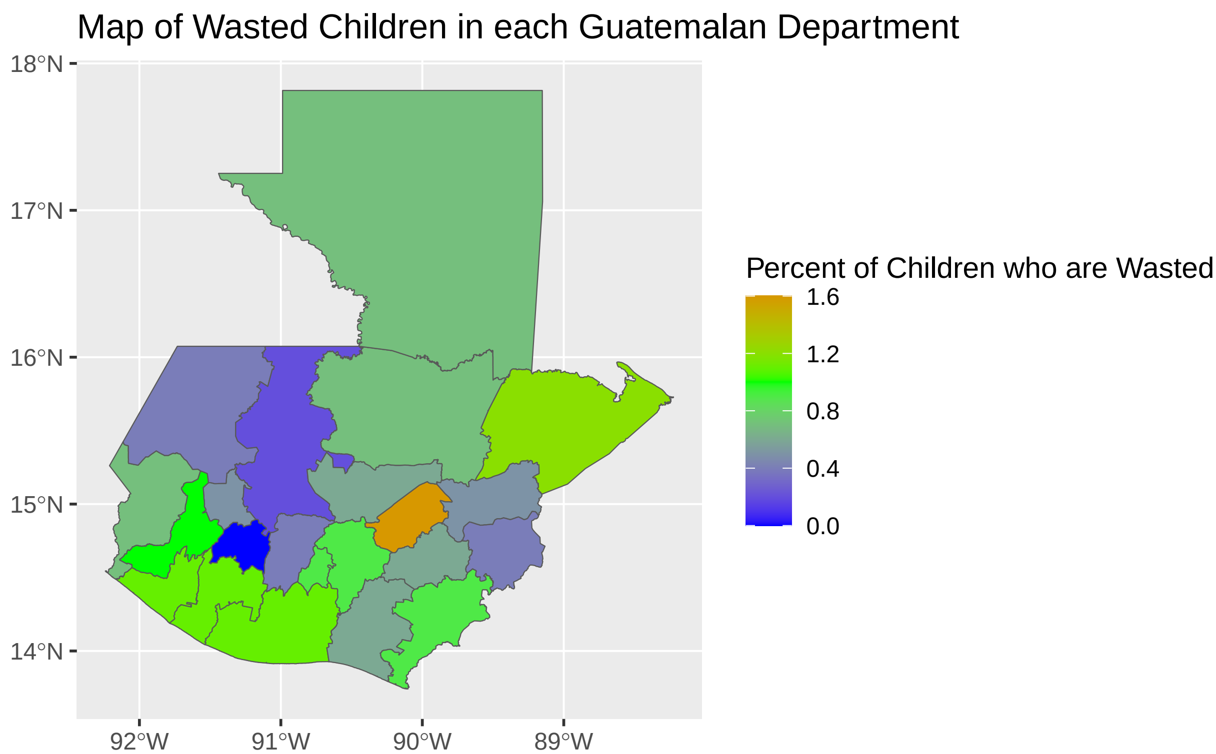

```r
# Anemia
ggplot(data_sf) +
  geom_sf(aes(fill = percent_anemic)) +
  scale_fill_gradient2(low = "blue",
                       mid = "green",
                       high = "red",
                       midpoint=15) +
  ggtitle("Map of Anemic Children in each Guatemalan Department") +
  labs (fill= "Percent of Children with Anemia")
```



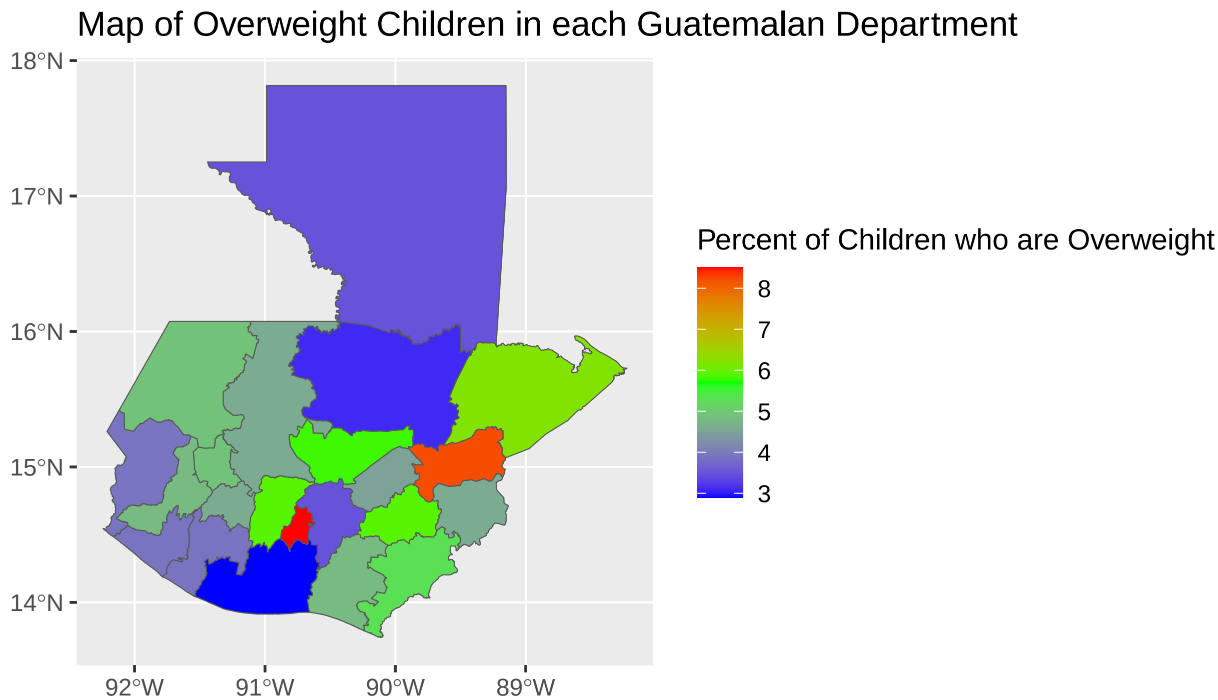

```r
# Childhood Stunting
ggplot(data_sf) +
  geom_sf(aes(fill = percent_child_stunt)) +
  scale_fill_gradient2(low = "blue",
                       mid = "green",
                       high = "red",
                       midpoint=44) +
  ggtitle("Map of Stunted Children in each Guatemalan Department") +
  labs (fill= "Percent of Children who are Stunted")
```




```r
# Childhood Wasting
ggplot(data_sf) +
  geom_sf(aes(fill = percent_child_wasted)) +
  scale_fill_gradient2(low = "blue",
                       mid = "green",
                       high = "red",
                       midpoint=1) +
  ggtitle("Map of Wasted Children in each Guatemalan Department") +
  labs (fill= "Percent of Children who are Wasted")
```




```
# Children Overweight
ggplot(data_sf) +
  geom_sf(aes(fill = percent_child_ovrwgt)) +
  scale_fill_gradient2(low = "blue",
                       mid = "green",
                       high = "red",
                       midpoint= 5.7) +
  ggtitle("Map of Overweight Children in each Guatemalan Department") +
  labs (fill= "Percent of Children who are Overweight")
```




Based on these maps, it appears that more people in the Southern regions of Guatemala tend to have more access to electricity than the Northern regions. This makes sense because the capital city of Antigua lies in the South near these regions, so more urban populations likely reside here compared to rural ones. For anemia, the pattern is less obvious. It seems that there are more anemic populations living on both the Southern and Northern departments. The spatial trend is also less clear for the percent of children who are stunted. For this variable, it appears that there are lower levels of stunting in the Southern-most and Western departments, but this is not a strong pattern. However, there is a cluster of departments that have relatively high levels of childhood stunting in the East. This is unexpected because there are lower levels of anemia in this same region. On the other hand, there is a higher prevalence of overweight children on the Eastern side of the country.

## Correlation Analysis

Before I run the correlation tests, I need to check my assumptions. The plots above show me that we have linear relationships that are homoscedastic. I noticed that there is an outlier at the 50% electricity mark that may affect our correlation analysis, so I'm going to filter that value out. I'm also going to test for normality to make sure that that assumption is satisfied.

```
# Filter out the outlier
data_sf_cropped <- data_sf %>%
  filter(percent_with_elec> 50)

# Test for normality using Shapiro-Wilk test
shapiro.test(data_sf_cropped$percent_with_elec)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_sf_cropped$percent_with_elec
## W = 0.91456, p-value = 0.06761
```

```
shapiro.test(data_sf_cropped$percent_anemic)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_sf_cropped$percent_anemic
## W = 0.96191, p-value = 0.5555
```

```
shapiro.test(data_sf_cropped$percent_child_stunt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_sf_cropped$percent_child_stunt
## W = 0.95365, p-value = 0.3985
```

```
shapiro.test(data_sf_cropped$percent_child_wasted)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_sf_cropped$percent_child_wasted
## W = 0.97747, p-value = 0.885
```

```
shapiro.test(data_sf_cropped$percent_child_ovrwgt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_sf_cropped$percent_child_ovrwgt
## W = 0.89768, p-value = 0.03156
```

When running a Shapiro-Wilk normality test, a p-value greater than 0.05 indicates that the variable of interest is normally distributed. After cropping the data to remove the outlier, it appears that every variable except `percent_child_ovrwgt` is normally distributed. So, I'll run a Pearson's correlation coefficient test on the normally distributed variables and a Spearman's rank correlation coefficient test on the variable that is not normally distributed.

```
# Pearson's correlation coefficient test

# Anemia
cor.test(data_sf_cropped$percent_with_elec, data_sf_cropped$percent_anemic, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data_sf_cropped$percent_with_elec and data_sf_cropped$percent_anemic
## t = -0.64702, df = 19, p-value = 0.5254
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5440315  0.3041367
## sample estimates:
##        cor
## -0.1468273
```

```
# Childhood stunting
cor.test(data_sf_cropped$percent_with_elec, data_sf_cropped$percent_child_stunt, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data_sf_cropped$percent_with_elec and data_sf_cropped$percent_child_stunt
## t = -0.59937, df = 19, p-value = 0.556
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5363691  0.3139233
## sample estimates:
##        cor
## -0.1362241
```

```
# Childhood wasting
cor.test(data_sf_cropped$percent_with_elec, data_sf_cropped$percent_child_wasted, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data_sf_cropped$percent_with_elec and data_sf_cropped$percent_child_wasted
## t = 0.92454, df = 19, p-value = 0.3668
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2462556  0.5866306
## sample estimates:
##       cor
## 0.2074882
```

```
# Spearman's rank correlation coefficient test

# Children Overweight
cor.test(data_sf_cropped$percent_with_elec, data_sf_cropped$percent_child_ovrwgt, method = "spearman")
```

```
## Warning in cor.test.default(data_sf_cropped$percent_with_elec,
## data_sf_cropped$percent_child_ovrwgt, : Cannot compute exact p-value with ties
```

```
##
##  Spearman's rank correlation rho
```

```
##
## data:  data_sf_cropped$percent_with_elec and data_sf_cropped$percent_child_ovrwgt
## S = 1804.1, p-value = 0.4573
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.1715042
```

For electricity and anemia, the p-value is greater than 0.05 (p-value= 0.5254). Therefore, there is not statistically significant relationship between these two variables.

For electricity and childhood stunting, the p-value is also greater than 0.05 (p-value = 0.556). Therefore, there is not statistically significant relationship between these two variables either.

We see the same lack of statistical significance between access to electricity and percent of children wasted since the p-value is less than 0.05 (p-value = 0.3668).

When running the correlation test for electricity and children who are overweight, the p-value was also greater than 0.05 (p-value = 0.4573), indicating that this correlation is also not statistically significant.

# Conclusion

According to our correlation analysis, it appears that there are no statistically significant relationships between hunger and access to electricity in Guatemala. However, our analysis still identified departments within the country that lack access to basic needs such as electricity and have relatively high levels of hunger indicators. Therefore, these results can now be used to prioritize departments that are most in need of assistance.