Impact of Social deprivation on Charity Care Burden

Ismat Ara Khan

2024-10-28

Abstract

Objective:

The aim of the project is to analyze the relationship between charity care burden and socioeconomic deprivation as measured by the Social Deprivation Index (SDI) and related factors across U.S. hospitals. The project seeks to understand how various factors, such as teaching status, urban-rural classification, ownership type, and bed size, influence the charity care burden that hospitals incur.

Methods:

The project involves merging and cleaning data from multiple sources, including hospital, Medicare, ZIP code, and poverty data, to analyze charity care burden across hospitals. Weighted averages of socioeconomic factors like the Social Deprivation Index (SDI) are calculated for each hospital, with total cases as weights. Quantile Generalized Additive Models (qgam) and linear quantile regression are used to model the relationship between charity care burden and weighted SDI score across various quantiles. Visualizations, including scatter plots and bar plots, help illustrate these relationships. Model diagnostics and hypothesis tests are used to assess the significance of predictors and the overall model fit.

Results:

The results show a significant positive relationship between the weighted Social Deprivation Index (SDI) score and charity care burden, particularly for hospitals with higher burdens. Hospitals serving more socioeconomically deprived populations tend to incur greater uncompensated care costs. Factors such as urban vs. rural location, teaching status, ownership type, and bed size also influence charity care burden, with nonprofit and urban hospitals displaying distinct patterns. The models explain a moderate amount of variability, with qgam results highlighting stronger effects in higher quantiles. Overall, the findings suggest that hospitals in more deprived areas face greater financial strain from uncompensated care.

Loading required packages and libraries

```r
# Install and load necessary packages if not already installed
if (!require(qgam)) install.packages("qgam", dependencies = TRUE)
if (!require(ggplot2)) install.packages("ggplot2", dependencies = TRUE)
if (!require(arsenal)) install.packages("arsenal", dependencies = TRUE)
if (!require(tibble)) install.packages("tibble", dependencies = TRUE)
if (!require(quantreg)) install.packages("quantreg", dependencies = TRUE)  # For linear quantile regres
if (!require(mgcv)) install.packages("mgcv", dependencies = TRUE)  # For generalized additive models
```

```r
# Load libraries
library(qgam)
library(mgcv)  # For s() and qgam()
library(ggplot2)
library(dplyr)
library(quantreg)
library(tidyr)
library(arsenal)
library(tibble)
options(repr.plot.width = 15, repr.plot.height = 8)
```

Introduction:

Healthcare systems, particularly hospitals, play a vital role in providing medical services to populations with varying socioeconomic backgrounds. In the United States, many hospitals incur significant charity care burden, which refers to the provision of uncompensated care to patients who are unable to pay for services. This burden is often disproportionately borne by hospitals serving economically disadvantaged communities, where social and economic deprivation may limit patients' ability to afford healthcare. The Social Deprivation Index (SDI), a composite measure of various socioeconomic factors, serves as a useful tool to quantify the level of deprivation in the populations served by these hospitals.

Understanding the relationship between socioeconomic deprivation and charity care burden is crucial for informing health policy, resource allocation, and financial support for hospitals. This study aims to investigate how weighted SDI scores and other hospital characteristics, such as teaching status, urban-rural classification, and ownership type, influence the charity care burden across U.S. hospitals. By employing advanced statistical techniques like quantile generalized additive models (qgam) and linear quantile regression, this research provides a comprehensive analysis of the factors driving charity care burden, offering insights into the challenges hospitals face in underserved areas.

Data Origin and Description

The data for this study was sourced from multiple publicly available datasets, ensuring a comprehensive analysis of the relationship between charity care burden and socioeconomic deprivation across U.S. hospitals. The primary sources of data include:

- Hospital Data: This dataset includes detailed information on charity care burden and various hospital characteristics, such as teaching status, urban-rural classification, and ownership type.
- Medicare Data: It provides data on hospital service areas, including medicare provider numbers and other relevant identifiers.
- ZIP Code-to-ZCTA Crosswalk: This dataset maps ZIP codes to ZCTAs (ZIP Code Tabulation Areas), facilitating the alignment of socioeconomic data with hospital service areas.

4.Poverty Data: This includes various socioeconomic indicators, such as poverty levels, education attainment, and employment status, which are used to calculate the Social Deprivation Index (SDI).

5.Bed Size Data:Contains information on the number of beds and hospital size categories, which is crucial for analyzing the impact of hospital size on charity care burden.

Methodology:

Data Integration and Cleaning

Multiple datasets that we used here,i.e. hospital data, Medicare data, ZIP code-to-ZCTA crosswalk, poverty data, and bed size data.Hospital data, Medicare data, ZIP code-to-ZCTA crosswalk, poverty data, and bed size data. These datasets were merged based on common identifiers such as ZIP code and Medicare provider number to ensure alignment of the Social Deprivation Index (SDI) and related factors with hospital data. The steps involved are as follows:

Merging Data:

- Merge ZIP code data with Medicare data using the ZIP_CD_OF_RESIDENCE and ZIP_CODE columns.
- Merge the resulting dataset with poverty data using the zcta and ZCTA5_FIPS columns.
- Convert the MEDICARE_PROV_NUM column to numeric to ensure compatibility for merging.
- Merge the combined dataset with hospital data using the MEDICARE_PROV_NUM and CCN columns.
- Finally, merge the dataset with bed size data using the MEDICARE_PROV_NUM and CCN columns.

Data Cleaning:

Handle non-numeric values and remove missing or invalid entries. Convert relevant columns (e.g., TOTAL_CASES, TOTAL_CHARGES, TOTAL_DAYS_OF_CARE) to numeric, ensuring no non-numeric characters are present. Remove rows with missing values in critical columns to ensure data integrity.

Data Transformation:

Weighted averages of socioeconomic variables (e.g., SDI score, poverty score) are calculated for each hospital, with the total number of cases used as weights. This ensures that hospitals with more cases have a larger influence on the averages.After this transformation we filtered out rows with negative or infinite values in the charity_care_burden column which ensure all relevant columns are complete and free of missing values.

Input all data

```
library(readr)
hospital_data = read_csv("C:/Users/Ismat/OneDrive/Desktop/fall24/Stat consulting/Project 1/2020-2022 Di
hospital1_data = read_csv("C:/Users/Ismat/OneDrive/Desktop/fall24/Stat consulting/Project 1/2022 5-Year
medicare_data = read_csv("C:/Users/Ismat/OneDrive/Desktop/fall24/Stat consulting/Project 1/Hospital_Serv
zip_data = read_csv("C:/Users/Ismat/OneDrive/Desktop/fall24/Stat consulting/Project 1/ZIP Code to ZCTA C
poverty_data = read_csv("C:/Users/Ismat/OneDrive/Desktop/fall24/Stat consulting/Project 1/rgcsdi-2015-20
bed_size_data = read_csv("C:/Users/Ismat/OneDrive/Desktop/fall24/Stat consulting/Project 1/Hospital Char
```

Merging Zip data with medicare data by using “ZIP_CD_OF_RESIDENCE” = “ZIP_CODE”

```
#  ZIP_CODE in zip_data and ZIP_CD_OF_RESIDENCE in medicare_data are the columns to merge on
zip_medicare_merged <- medicare_data %>%
  left_join(zip_data, by = c("ZIP_CD_OF_RESIDENCE" = "ZIP_CODE"))
```

Again merging proverty data with previous merged data by using “zcta” = “ZCTA5_FIPS”

```
#  Merge the poverty data based on ZCTA (from the merged dataset)
# 'zcta' in zip_medicare_merged corresponds to 'ZCTA5_FIPS' in poverty_data
zip_medicare_poverty_merged <- zip_medicare_merged %>%
  left_join(poverty_data, by = c("zcta" = "ZCTA5_FIPS"))
```

Converting MEDICARE_PROV_NUM column to numeric before next merging with hospital data by using “MEDICARE_PROV_NUM” = “CCN”, and then finally merge with the bed size data

```
#  Check for non-numeric values in MEDICARE_PROV_NUM
non_numeric_rows <- zip_medicare_poverty_merged %>%
  filter(!grepl("^[0-9]+$", MEDICARE_PROV_NUM))

#  Clean MEDICARE_PROV_NUM and convert to numeric
zip_medicare_poverty_merged <- zip_medicare_poverty_merged %>%
  mutate(MEDICARE_PROV_NUM = as.numeric(gsub("[^0-9]", "", MEDICARE_PROV_NUM)))
```

```r
# Step 6: Merge hospital data based on the Medicare provider number (CCN and MEDICARE_PROV_NUM)
final_data_burden <- zip_medicare_poverty_merged %>%
  left_join(hospital_data, by = c("MEDICARE_PROV_NUM" = "CCN"))

# Convert MEDICARE_PROV_NUM in final_with_burden and CCN in CCN_bed to numeric, if necessary
 bed_size_data<- bed_size_data %>%
  mutate(CCN = as.numeric(CCN))
# Check the column names in CCN_bed
colnames(bed_size_data)
```

```
##  [1] "CCN"                    "critical_access"       "teaching_status"
##  [4] "urban_rural"            "referral_center"       "num_beds"
##  [7] "type_of_service"        "state_code"            "ownership_type_simple"
## [10] "bed_size_simple"
```

```r
# Assuming the correct column names are found to be 'BEDS' and 'TOTAL_BEDS'
final_data <- final_data_burden %>%
  left_join(bed_size_data %>% select(CCN, num_beds, bed_size_simple, type_of_service), by = c("MEDICARE_
```

Convert TOTAL_CASES, TOTAL_CHARGES, and TOTAL_DAYS_OF_CARE to numeric.

```r
#This will strip out any non-numeric characters before converting to numeric, minimizing the introducti
final_data <- final_data %>%
  mutate(TOTAL_CASES = as.numeric(gsub("[^0-9.]", "", TOTAL_CASES)),
         TOTAL_CHARGES = as.numeric(gsub("[^0-9.]", "", TOTAL_CHARGES)),
         TOTAL_DAYS_OF_CARE = as.numeric(gsub("[^0-9.]", "", TOTAL_DAYS_OF_CARE)))
```

Removing all rows that contain NA

```r
# Filter out rows with missing TOTAL_CASES
final_data2 <- final_data %>%
  dplyr::filter(!is.na(TOTAL_CASES))

# Filter out rows with missing MEDICARE_PROV_NUM
final_data2 <- final_data2 %>%
  dplyr::filter(!is.na(MEDICARE_PROV_NUM))
```

```r
# Remove rows with na
final_data3 <- na.omit(final_data2)
# View the cleaned final data
head(final_data3)
```

```
## # A tibble: 6 x 39
##   MEDICARE_PROV_NUM ZIP_CD_OF_RESIDENCE TOTAL_DAYS_OF_CARE TOTAL_CHARGES
##               <dbl> <chr>                            <dbl>         <dbl>
## 1             10001 32420                              148       1808644
## 2             10001 32421                               88       1094057
## 3             10001 32423                              218       2553675
## 4             10001 32425                              917      10887590
## 5             10001 32426                              118        947259
## 6             10001 32428                              788      10092714
## # i 35 more variables: TOTAL_CASES <dbl>, PO_NAME <chr>, STATE <chr>,
## #   ZIP_TYPE <chr>, zcta <chr>, zip_join_type <chr>, ZCTA5_population <dbl>,
## #   SDI_score <dbl>, PovertyLT100_FPL_score <dbl>,
## #   Single_Parent_Fam_score <dbl>, Education_LT12years_score <dbl>,
## #   HHNo_Vehicle_score <dbl>, HHRenter_Occupied_score <dbl>,
## #   HHCrowding_score <dbl>, Nonemployed_score <dbl>, sdi <dbl>,
## #   pct_Poverty_LT100 <dbl>, pct_Single_Parent_Fam <dbl>, ...
```

Let us check the number of unique Medicare IDs and the ZCTA counts

```
## Check unique medicare IDs & unique ZCTAs

# Step 2.1: Count unique MEDICARE_PROV_NUMs for each ZIP_CD_OF_RESIDENCE
mCCN_by_zip <- final_data2 %>%
  group_by(ZIP_CD_OF_RESIDENCE) %>%
  summarise(Unique_mCCN_by_zip = n_distinct(MEDICARE_PROV_NUM)) %>%
  ungroup()
print(mCCN_by_zip)
```

```
## # A tibble: 28,644 x 2
##    ZIP_CD_OF_RESIDENCE Unique_mCCN_by_zip
##    <chr>                            <int>
##  1 00601                                6
##  2 00602                                9
##  3 00603                                9
##  4 00604                                1
##  5 00605                                5
##  6 00606                                5
##  7 00610                                6
##  8 00611                                3
##  9 00612                               11
## 10 00613                                5
## # i 28,634 more rows
```

```
# Step 2.2: Get the total count of distinct MEDICARE_PROV_NUMs across all ZIP_CD_OF_RESIDENCE
total_mCCN_count <- n_distinct(final_data2$MEDICARE_PROV_NUM)
# Print the total distinct MEDICARE_PROV_NUMs
print(total_mCCN_count)
```

```
## [1] 6316
```

Calculating weighted score for selected columns and join them with the merged data

```
# Install and load necessary packages
if(!require(qgam)) install.packages("qgam", dependencies=TRUE)
library(qgam)
library(dplyr)

#After adding new variables, create new weighted data

new_weighted_avg_by_hospital <- final_data2 %>%
  group_by(MEDICARE_PROV_NUM) %>%
  mutate(total_cases = sum(TOTAL_CASES)) %>%
  mutate(weight = TOTAL_CASES / total_cases) %>%
  mutate(weighted_SDI_score = weight * SDI_score,
         weighted_PovertyLT100_FPL_score = weight * PovertyLT100_FPL_score,
         weighted_Single_Parent_Fam_score = weight * Single_Parent_Fam_score,
         weighted_Education_LT12years_score = weight * Education_LT12years_score,
         weighted_HHNo_Vehicle_score = weight * HHNo_Vehicle_score,
         weighted_HHRenter_Occupied_score = weight * HHRenter_Occupied_score,
         weighted_HHCrowding_score = weight * HHCrowding_score,
         weighted_Nonemployed_score = weight * Nonemployed_score) %>%
   summarise(
    #across(everything(), first),
    weighted_SDI_score = sum(weighted_SDI_score),
    weighted_PovertyLT100_FPL_score = sum(weighted_PovertyLT100_FPL_score),
    weighted_Single_Parent_Fam_score = sum(weighted_Single_Parent_Fam_score),
    weighted_Education_LT12years_score = sum(weighted_Education_LT12years_score),
    weighted_HHNo_Vehicle_score = sum(weighted_HHNo_Vehicle_score),
    weighted_HHRenter_Occupied_score = sum(weighted_HHRenter_Occupied_score),
    weighted_HHCrowding_score = sum(weighted_HHCrowding_score),
    weighted_Nonemployed_score = sum(weighted_Nonemployed_score),
    critical_access = first(critical_access),
    teaching_status = first(teaching_status),
    urban_rural = first(urban_rural),
    referral_center = first(referral_center),
    ownership_type_simple = first(ownership_type_simple),
    medicaid_caseload = first(medicaid_caseload),
    disproportionate_percentage = first(disproportionate_percentage),
    charity_care_burden = first(charity_care_burden),
    uncomp_burden = first(uncomp_burden),
    total_cases = first(total_cases),
    num_beds = first(num_beds),
    bed_size_simple = first(bed_size_simple),
    type_of_service = first(type_of_service)
  ) %>%
  ungroup()
#write.csv(new_weighted_avg_by_hospital,file = "new_weighted_data.csv",row.names = FALSE)
```

Cleaning data by filtering (charity__care__burden >= 0 & is.finite(charity__care__burden)) so that we can get rid of all values that are negative and infinity

```
# Step 1: Remove negative and infinite values from charity_care_burden
new_cleaned_data <- new_weighted_avg_by_hospital %>%
  dplyr::filter(charity_care_burden >= 0 & is.finite(charity_care_burden))
head(new_cleaned_data)
```

```
## # A tibble: 6 x 22
##   MEDICARE_PROV_NUM weighted_SDI_score weighted_PovertyLT100_FPL_score
##               <dbl>              <dbl>                           <dbl>
## 1             10001               69.1                            77.1
## 2             10005               63.8                            78.9
## 3             10006               52.8                            61.6
## 4             10007               61.1                            66.3
## 5             10008               53.5                            53.1
## 6             10011               51.0                            60.5
## # i 19 more variables: weighted_Single_Parent_Fam_score <dbl>,
## #   weighted_Education_LT12years_score <dbl>,
## #   weighted_HHNo_Vehicle_score <dbl>, weighted_HHRenter_Occupied_score <dbl>,
## #   weighted_HHCrowding_score <dbl>, weighted_Nonemployed_score <dbl>,
## #   critical_access <chr>, teaching_status <chr>, urban_rural <chr>,
## #   referral_center <chr>, ownership_type_simple <chr>,
## #   medicaid_caseload <dbl>, disproportionate_percentage <dbl>, ...
```

Exploratory Data Analysis (EDA):

EDA provides initial insights into data patterns, helps clean and check data quality, and clarifies variable distributions and relationships. It supports informed decision-making by highlighting trends, anomalies, and correlations, guiding appropriate model selection and feature engineering. In this part from EDA we can go through a quich check of the effect of other pedictor variables on charity care burden including weighted SDI score

Visualization:

Two scatter plot has been constructed : First one of them with all categorical variables along with weighted SDI score and another for all weighted score to show the effect of every variables on charity care burden.

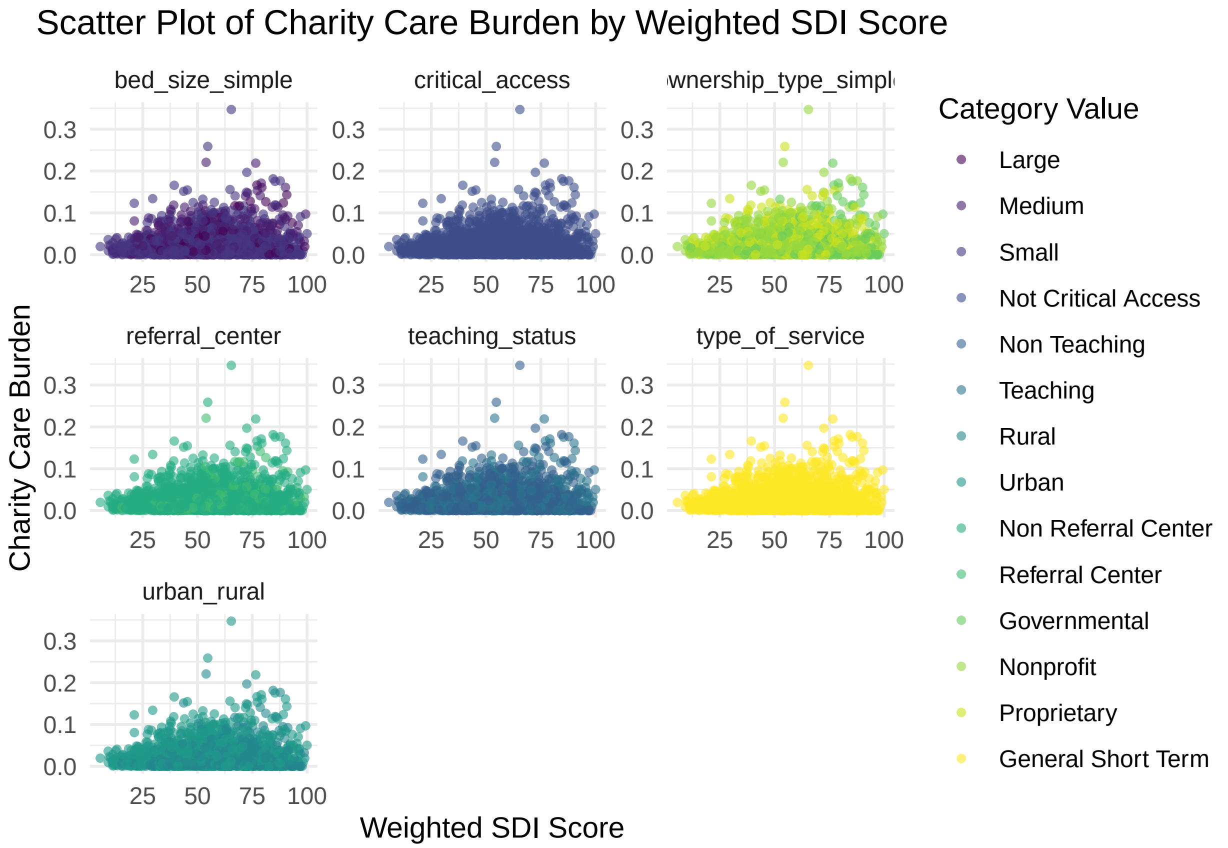
```r
# Load necessary libraries
library(ggplot2)
library(dplyr)
library(tidyr)
library(viridis)
```

```
## Warning: package 'viridis' was built under R version 4.3.3
```

```
## Loading required package: viridisLite
```

```
## Warning: package 'viridisLite' was built under R version 4.3.3
```

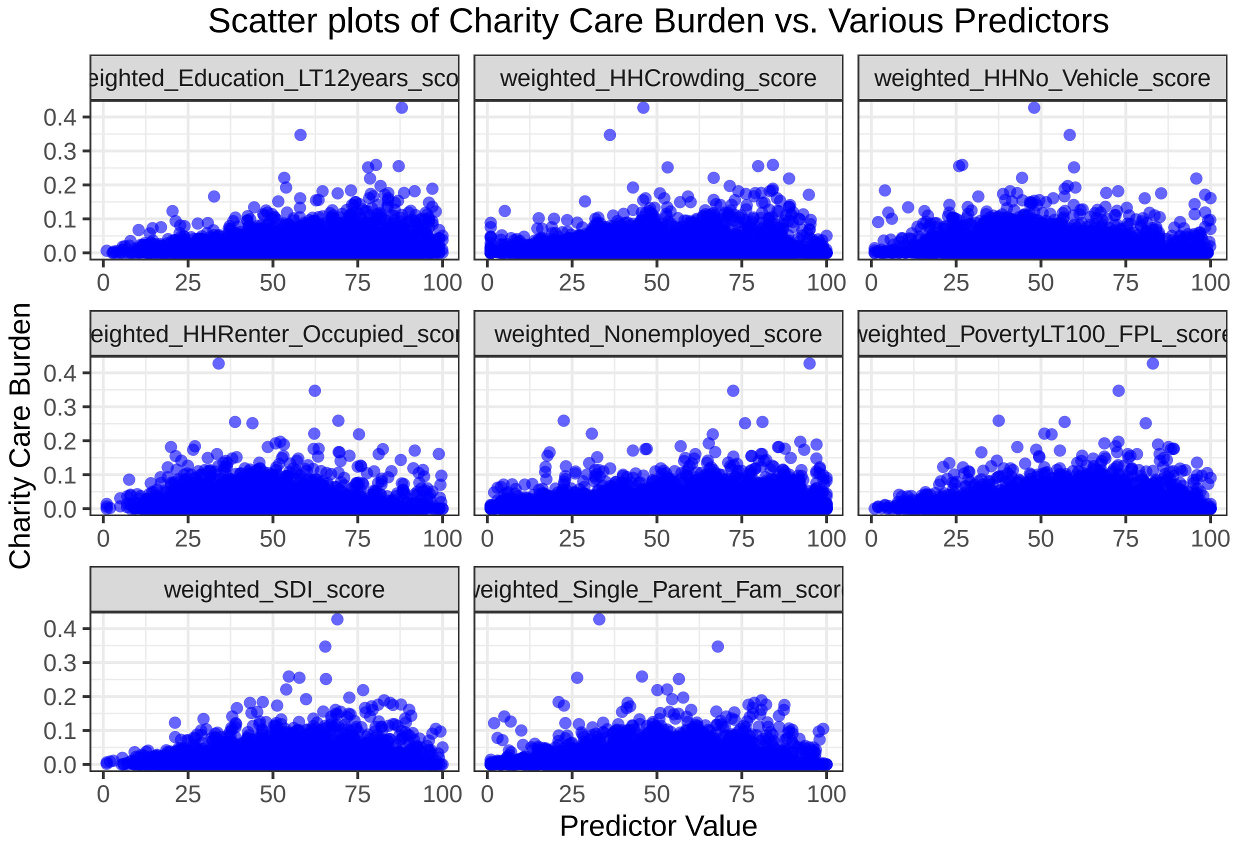
```r
# Remove rows with any NA values in the relevant columns
cleaned_data <- new_weighted_avg_by_hospital %>%
  filter(complete.cases(.))

# Define the list of categorical variables for plotting
categorical_vars <- c("bed_size_simple",
                      "critical_access",
                      "teaching_status",
                      "urban_rural",
                      "referral_center",
                      "ownership_type_simple",
                      "type_of_service")

# Convert categorical variables to factors
cleaned_data[categorical_vars] <- lapply(cleaned_data[categorical_vars], factor)

# Pivot the data to long format for easier faceting
long_data <- cleaned_data %>%
  pivot_longer(cols = all_of(categorical_vars),
               names_to = "Category",
               values_to = "Value")

# Create a scatter plot with facets for each categorical variable
ggplot(long_data, aes(x = weighted_SDI_score, y = charity_care_burden, color = Value)) +
  geom_point(alpha = 0.6, size = 1) +
  labs(
    title = "Scatter Plot of Charity Care Burden by Weighted SDI Score",
    x = "Weighted SDI Score",
    y = "Charity Care Burden",
    color = "Category Value"
  ) +
  scale_color_viridis_d() +  # Change palette here
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5)) +
  facet_wrap(~ Category, scales = "free")  # Facet by Category
```



The above scatter plots analysis reveals a positive relationship between weighted SDI scores and charity care burden, with hospitals in more socioeconomically deprived areas showing higher burdens. Larger hospitals, particularly those in urban settings, teaching hospitals, and nonprofits, bear a greater charity care burden as SDI scores increase, suggesting that these factors amplify the effects of deprivation. The plot also highlights certain outliers where some hospitals, even with moderate SDI scores, incur high burdens, possibly due to local needs or unique hospital characteristics. This analysis underscores the combined impact of SDI, bed size, and hospital characteristics on charity care burden.

```
#DESCRIPTIVE ANALYSIS
# Install and load ggplot2 if not already installed
if (!require(ggplot2)) install.packages("ggplot2", dependencies=TRUE)
library(ggplot2)
library(tidyr)


# Define a list of variables to plot against charity_care_burden
predictor_vars <- c("weighted_SDI_score",
                    "weighted_PovertyLT100_FPL_score",
                    "weighted_Single_Parent_Fam_score",
                    "weighted_Education_LT12years_score",
                    "weighted_HHNo_Vehicle_score",
                    "weighted_HHRenter_Occupied_score",
                    "weighted_HHCrowding_score",
                    "weighted_Nonemployed_score")

# Reshape data to long format for ggplot2
long_data <- new_weighted_avg_by_hospital %>%
  pivot_longer(cols = all_of(predictor_vars),
               names_to = "predictor",
               values_to = "predictor_value")
# Plot with ggplot2 using facet_wrap
ggplot(long_data, aes(x = predictor_value, y = charity_care_burden)) +
  geom_point(color = "blue", alpha = 0.6) +
  labs(
    title = "Scatter plots of Charity Care Burden vs. Various Predictors",
    x = "Predictor Value",
    y = "Charity Care Burden"
  ) +
  theme_bw() +
  theme(plot.title = element_text(hjust = 0.5)) +
  facet_wrap(~ predictor, scales = "free_x")  # Free x-scale to handle different variable ranges
```

```
## Warning: Removed 12471 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



Theses scatter plots emphasize that socioeconomic indicators—such as deprivation, poverty, family structure, education, and housing stability—are all positively associated with charity care burden, underscoring the complex social factors contributing to hospitals' financial challenges in disadvantaged areas.

Desscriptive Statistics Table:

Following table contains the descriptive statistics of some important predictor variables

```
# Create the tableby summary
library(arsenal)
library(tibble)
summary_table <- tableby( ~ critical_access +
                            teaching_status +
                            urban_rural +
                            referral_center +
                            ownership_type_simple +
                            medicaid_caseload +
                            disproportionate_percentage +
                            charity_care_burden +
                            bed_size_simple+
                            uncomp_burden +
                            weighted_SDI_score +
                            weighted_PovertyLT100_FPL_score +
                            weighted_Single_Parent_Fam_score +
                            weighted_Education_LT12years_score +
                            weighted_HHNo_Vehicle_score +
                            weighted_HHRenter_Occupied_score +
                            weighted_HHCrowding_score +
                            weighted_Nonemployed_score,
                          data = new_cleaned_data)

# Print the summary table
summary(summary_table)
```

```
##
##
## |                                       | Overall (N=5188) |
## |:--------------------------------------|:----------------:|
## |**critical_access**                    |                  |
## |&nbsp;&nbsp;&nbsp;Critical Access      |   1288 (24.8%)   |
## |&nbsp;&nbsp;&nbsp;Not Critical Access  |   3900 (75.2%)   |
## |**teaching_status**                    |                  |
## |&nbsp;&nbsp;&nbsp;Non Teaching         |   3786 (73.0%)   |
## |&nbsp;&nbsp;&nbsp;Teaching             |   1402 (27.0%)   |
## |**urban_rural**                        |                  |
## |&nbsp;&nbsp;&nbsp;Rural                |   2513 (48.4%)   |
## |&nbsp;&nbsp;&nbsp;Urban                |   2675 (51.6%)   |
## |**referral_center**                    |                  |
## |&nbsp;&nbsp;&nbsp;Non Referral Center  |   4474 (86.2%)   |
## |&nbsp;&nbsp;&nbsp;Referral Center      |   714 (13.8%)    |
## |**ownership_type_simple**              |                  |
## |&nbsp;&nbsp;&nbsp;Governmental         |   981 (18.9%)    |
## |&nbsp;&nbsp;&nbsp;Nonprofit            |   2799 (54.0%)   |
## |&nbsp;&nbsp;&nbsp;Proprietary          |   1408 (27.1%)   |
## |**medicaid_caseload**                  |                  |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            |  0.075 (0.103)   |
## |&nbsp;&nbsp;&nbsp;Range                |  0.000 - 0.761   |
## |**disproportionate_percentage**        |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       2279       |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            |  0.324 (0.160)   |
## |&nbsp;&nbsp;&nbsp;Range                |  0.001 - 1.733   |
## |**charity_care_burden**                |                  |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            |  0.018 (0.027)   |
## |&nbsp;&nbsp;&nbsp;Range                |  0.000 - 0.427   |
## |**bed_size_simple**                    |                  |
## |&nbsp;&nbsp;&nbsp;Large                |    391 (7.5%)    |
## |&nbsp;&nbsp;&nbsp;Medium               |   1156 (22.3%)   |
## |&nbsp;&nbsp;&nbsp;Small                |   3641 (70.2%)   |
## |**uncomp_burden**                      |                  |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            |  0.056 (0.062)   |
## |&nbsp;&nbsp;&nbsp;Range                |  -0.001 - 1.359  |
## |**weighted_SDI_score**                 |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       465        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 52.650 (18.944)  |
## |&nbsp;&nbsp;&nbsp;Range                | 1.000 - 100.000  |
## |**weighted_PovertyLT100_FPL_score**    |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       426        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 56.255 (19.969)  |
## |&nbsp;&nbsp;&nbsp;Range                | 1.000 - 100.000  |
## |**weighted_Single_Parent_Fam_score**   |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       426        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 51.935 (19.224)  |
## |&nbsp;&nbsp;&nbsp;Range                | 1.000 - 100.000  |
## |**weighted_Education_LT12years_score** |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       426        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 54.477 (19.960)  |
## |&nbsp;&nbsp;&nbsp;Range                | 1.000 - 100.000  |
## |**weighted_HHNo_Vehicle_score**        |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       426        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 52.755 (18.533)  |
## |&nbsp;&nbsp;&nbsp;Range                |  1.000 - 99.989  |
## |**weighted_HHRenter_Occupied_score**   |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       426        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 46.186 (17.317)  |
## |&nbsp;&nbsp;&nbsp;Range                | 1.000 - 100.000  |
## |**weighted_HHCrowding_score**          |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       426        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 45.413 (21.727)  |
## |&nbsp;&nbsp;&nbsp;Range                | 1.000 - 100.000  |
## |**weighted_Nonemployed_score**         |                  |
## |&nbsp;&nbsp;&nbsp;N-Miss               |       426        |
## |&nbsp;&nbsp;&nbsp;Mean (SD)            | 55.597 (25.226)  |
## |&nbsp;&nbsp;&nbsp;Range                | 1.000 - 100.000  |
```

The table provides counts and percentages for different types of hospitals based on critical access status, teaching status, urban-rural classification, referral center status, and ownership type. It includes mean, standard deviation, and range for various socioeconomic indicators like Medicaid caseload, disproportionate percentage, and weighted scores for SDI, poverty, single-parent families, education, vehicle ownership, renter occupancy, crowding, and nonemployment.The table summarizes the mean, standard deviation, and range for charity care burden and uncompensated burden across hospitals.

The table also indicates the number of missing values for certain variables, highlighting areas where data may be incomplete.

Modeling Approach:

Quantile Generalized Additive Models (QGAM)

Quantile Generalized Additive Models (QGAM) extend the framework of Generalized Additive Models (GAM) to allow for the modeling of different quantiles of the response variable. This is particularly useful when the relationship between predictors and the response variable varies across different points of the distribution, such as the median or other quantiles.

Generalized Additive Models (GAM):

GAMs are a flexible extension of generalized linear models (GLMs) that allow for non-linear relationships between the predictors and the response variable. The model is expressed as:

$$g(E(Y)) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p)$$

where (g) is the link function, $(E(Y))$ is the expected value of the response variable (Y), (β_0) is the intercept, and $(f_j(X_j))$ are smooth functions of the predictors (X_j).

Quantile Regression:

Quantile regression extends traditional regression models by estimating the conditional quantiles of the response variable, rather than the mean. This is useful for understanding the impact of predictors on different points of the distribution of the response variable. The quantile regression model for the (τ)-th quantile is given by:

$$Q_Y(\tau|X) = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2 + \cdots + \beta_p(\tau)X_p$$

where $(Q_Y(\tau|X))$ is the (τ)-th quantile of (Y) given (X).

Quantile Generalized Additive Models (QGAM):

QGAM combines the flexibility of GAMs with the robustness of quantile regression. It models the conditional quantiles of the response variable using smooth functions of the predictors. The QGAM for the (τ)-th quantile is expressed as:

$$Q_Y(\tau|X) = \beta_0(\tau) + f_1(X_1, \tau) + f_2(X_2, \tau) + \ldots + f_p(X_p, \tau)$$

where $(f_j(X_j, \tau))$ are smooth functions that can vary with the quantile (τ).

Implementation:

The charity care burden is modeled as the response variable, with predictors such as SDI score, hospital characteristics (teaching status, urban/rural status, ownership type), and total cases.

The qgam function is used to fit models for multiple quantiles (e.g., 0.25, 0.5, 0.9), allowing the analysis to capture how the relationship between SDI and charity care burden changes across different quantiles of the response variable.

Fit a quantile regression model using qgam

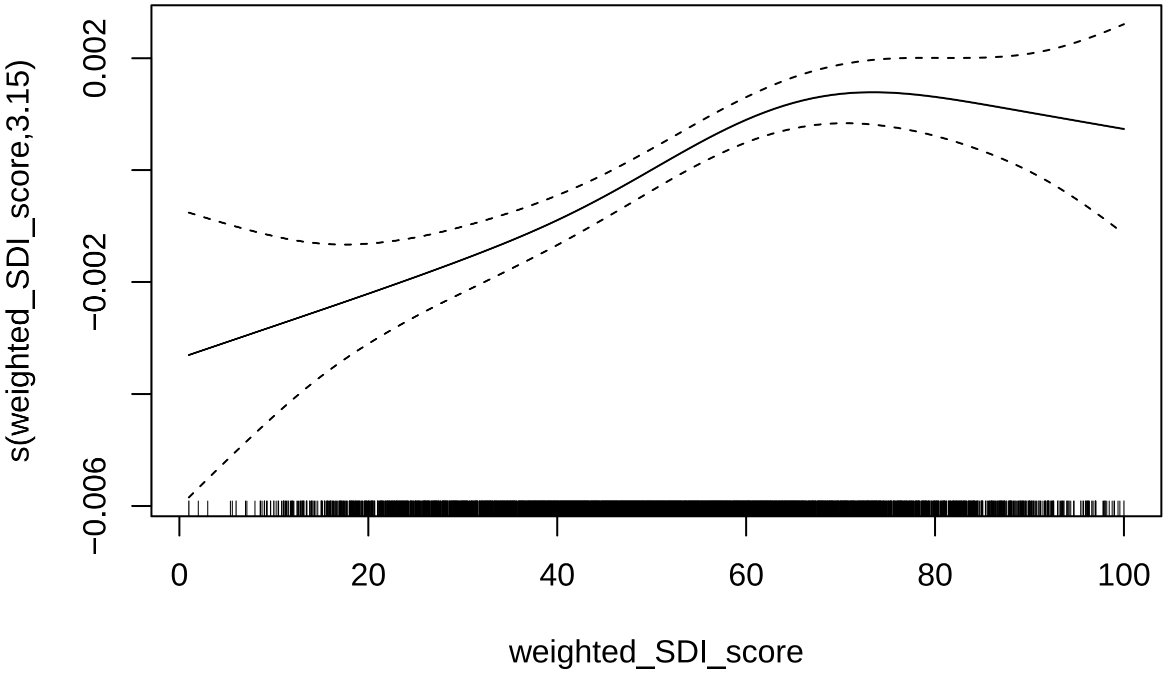
```r
# model for the median (0.5 quantile), can adjust this to other quantiles if needed.
library(qgam)
qgam_model <- qgam(charity_care_burden ~ s(weighted_SDI_score) +
                     critical_access +
                     teaching_status +
                     urban_rural +
                     referral_center +
                     ownership_type_simple +
                     type_of_service +
                     bed_size_simple+
                     total_cases,
                   data = new_cleaned_data, qu = 0.5)
```

```
## Estimating learning rate. Each dot corresponds to a loss evaluation.
## qu = 0.5.........done
```

```
# View the summary of the model
summary(qgam_model)
```

```
##
## Family: elf
## Link function: identity
##
## Formula:
## charity_care_burden ~ s(weighted_SDI_score) + critical_access +
##     teaching_status + urban_rural + referral_center + ownership_type_simple +
##     type_of_service + bed_size_simple + total_cases
##
## Parametric coefficients:
##                                                   Estimate Std. Error z value
## (Intercept)                                     -7.704e-03  2.910e-03  -2.648
## critical_accessNot Critical Access               3.266e-03  6.864e-04   4.758
## teaching_statusTeaching                         -2.544e-04  6.211e-04  -0.410
## urban_ruralUrban                                 2.825e-03  6.077e-04   4.649
## referral_centerReferral Center                  -1.860e-03  7.733e-04  -2.405
## ownership_type_simpleNonprofit                   1.602e-03  5.330e-04   3.006
## ownership_type_simpleProprietary                 1.997e-03  6.881e-04   2.902
## type_of_serviceChildren                          2.338e-04  3.822e-03   0.061
## type_of_serviceExtended Neoplastic Disease Care -2.256e-03  6.522e-03  -0.346
## type_of_serviceGeneral Long Term                -2.364e-04  2.532e-03  -0.093
## type_of_serviceGeneral Short Term                1.434e-02  2.489e-03   5.762
## type_of_serviceOther                             1.352e-02  6.121e-03   2.209
## type_of_servicePsychiatric                      -1.181e-04  2.517e-03  -0.047
## type_of_serviceRehabilitation                    3.835e-04  2.513e-03   0.153
## bed_size_simpleMedium                            5.875e-04  1.379e-03   0.426
## bed_size_simpleSmall                            -6.652e-04  1.618e-03  -0.411
## total_cases                                      3.775e-07  1.588e-07   2.377
##                                                 Pr(>|z|)
## (Intercept)                                      0.00810 **
## critical_accessNot Critical Access              1.96e-06 ***
## teaching_statusTeaching                          0.68209
## urban_ruralUrban                                3.34e-06 ***
## referral_centerReferral Center                   0.01616 *
## ownership_type_simpleNonprofit                   0.00265 **
## ownership_type_simpleProprietary                 0.00371 **
## type_of_serviceChildren                          0.95122
## type_of_serviceExtended Neoplastic Disease Care  0.72936
## type_of_serviceGeneral Long Term                 0.92562
## type_of_serviceGeneral Short Term               8.30e-09 ***
## type_of_serviceOther                                 0.02718 *
## type_of_servicePsychiatric                           0.96256
## type_of_serviceRehabilitation                        0.87872
## bed_size_simpleMedium                                0.67007
## bed_size_simpleSmall                                 0.68104
## total_cases                                          0.01744 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                          edf Ref.df Chi.sq p-value
## s(weighted_SDI_score) 3.149  3.959  41.66  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0897   Deviance explained = 24.3%
## -REML = -12176  Scale est. = 1         n = 4723
```

```r
# Plot the results if needed
plot(qgam_model)
```



```r
library(ggplot2)
library(dplyr)
library(qgam)


# Fit the QGAM model
```

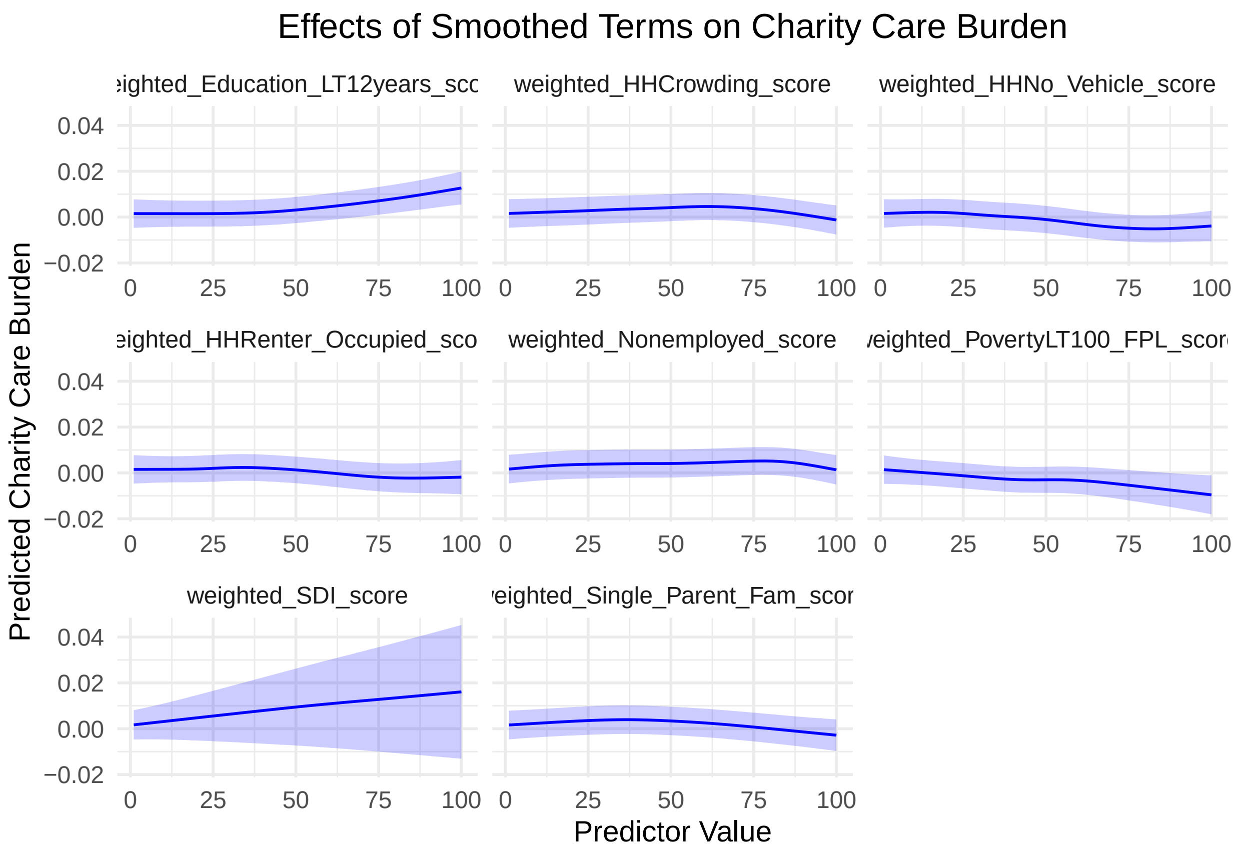
```r
qgam_model <- qgam(
  charity_care_burden ~
    s(weighted_SDI_score) +
    s(weighted_PovertyLT100_FPL_score) +
    s(weighted_Single_Parent_Fam_score) +
    s(weighted_Education_LT12years_score) +
    s(weighted_HHNo_Vehicle_score) +
    s(weighted_HHRenter_Occupied_score) +
    s(weighted_HHCrowding_score) +
    s(weighted_Nonemployed_score),
  data = new_cleaned_data,
  qu = 0.5
)
```

```
## Estimating learning rate. Each dot corresponds to a loss evaluation.
## qu = 0.5................done
```

```r
# View the summary of the model
summary(qgam_model)
```

```
##
## Family: elf
## Link function: identity
##
## Formula:
## charity_care_burden ~ s(weighted_SDI_score) + s(weighted_PovertyLT100_FPL_score) +
##     s(weighted_Single_Parent_Fam_score) + s(weighted_Education_LT12years_score) +
##     s(weighted_HHNo_Vehicle_score) + s(weighted_HHRenter_Occupied_score) +
##     s(weighted_HHCrowding_score) + s(weighted_Nonemployed_score)
##
## Parametric coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.0090619  0.0002237    40.5   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                         edf Ref.df Chi.sq  p-value
## s(weighted_SDI_score)                 1.470  1.816  2.085 0.406214
## s(weighted_PovertyLT100_FPL_score)    3.840  4.834 11.340 0.054929 .
## s(weighted_Single_Parent_Fam_score)   3.235  4.090 20.796 0.000396 ***
## s(weighted_Education_LT12years_score) 3.087  3.928 23.323 0.000113 ***
## s(weighted_HHNo_Vehicle_score)        4.264  5.309 32.412 6.75e-06 ***
## s(weighted_HHRenter_Occupied_score)   3.852  4.839 13.714 0.015867 *
## s(weighted_HHCrowding_score)          3.955  4.905 25.407 0.000118 ***
## s(weighted_Nonemployed_score)         4.685  5.748 14.475 0.019044 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0559   Deviance explained = 17.1%
## -REML = -11731  Scale est. = 1         n = 4723
```

```r
# Extract predictions for each smooth term
terms <- c("weighted_SDI_score", "weighted_PovertyLT100_FPL_score",
           "weighted_Single_Parent_Fam_score", "weighted_Education_LT12years_score",
           "weighted_HHNo_Vehicle_score", "weighted_HHRenter_Occupied_score",
           "weighted_HHCrowding_score", "weighted_Nonemployed_score")

# Extract predictions for each smooth term
predictions <- lapply(terms, function(term) {
  # Create a new data frame with only the current term varying, others set to 0
  new_data <- data.frame(matrix(0, nrow = 100, ncol = length(terms)))
  names(new_data) <- terms
  new_data[[term]] <- seq(min(new_cleaned_data[[term]], na.rm = TRUE),
                          max(new_cleaned_data[[term]], na.rm = TRUE), length.out = 100)

  # Get predictions for the current term
  preds <- predict(qgam_model, new_data, se.fit = TRUE)
  data.frame(
    term = term,
    x = new_data[[term]],
    fit = preds$fit,
    lower = preds$fit - 1.96 * preds$se.fit,  # 95% CI lower bound
    upper = preds$fit + 1.96 * preds$se.fit   # 95% CI upper bound
  )
})

# Combine predictions into a single data frame for faceting
predictions_df <- bind_rows(predictions)

# Plot with ggplot2 and facet_wrap
ggplot(predictions_df, aes(x = x, y = fit)) +
  geom_line(color = "blue") +
  geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2, fill = "blue") +
  labs(title = "Effects of Smoothed Terms on Charity Care Burden",
       x = "Predictor Value",
       y = "Predicted Charity Care Burden") +
  facet_wrap(~ term, scales = "free_x") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```



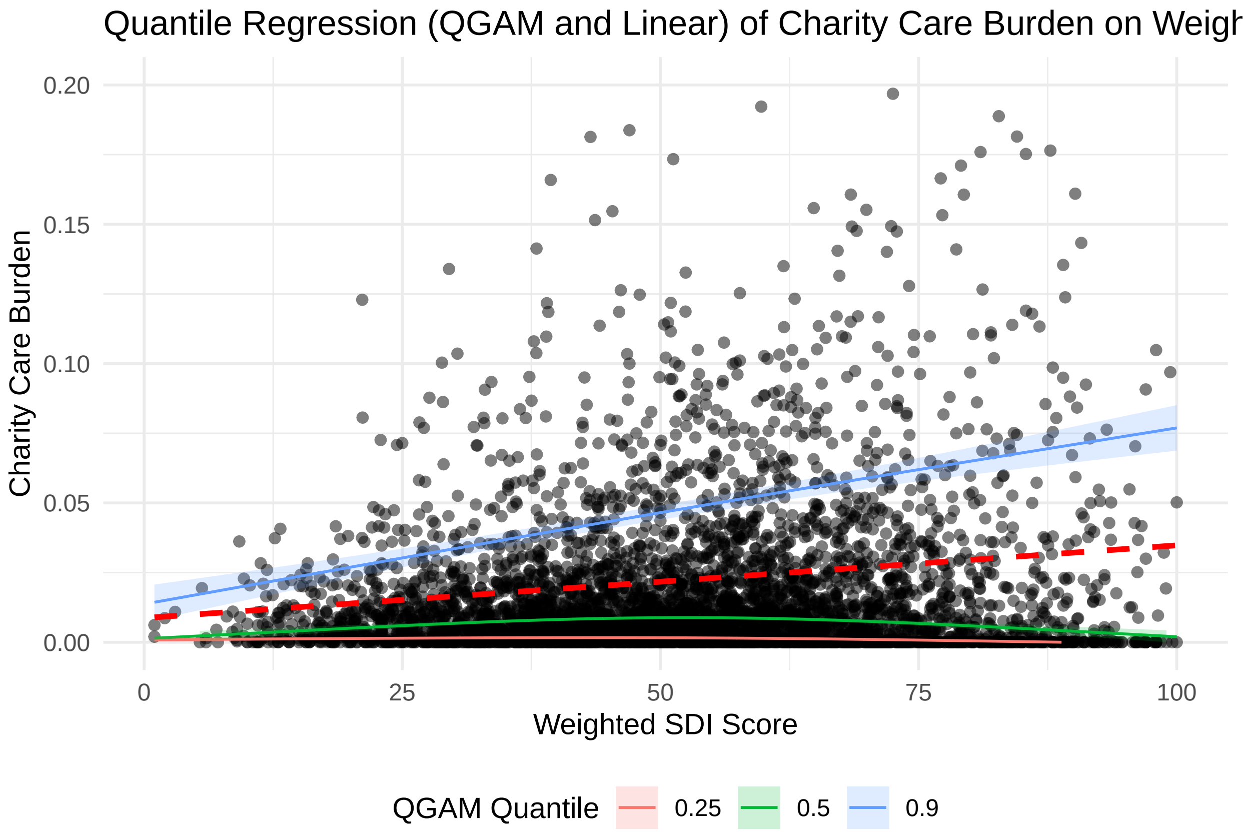
```r
# Fit quantile models for multiple quantiles (e.g., 0.25, 0.5, 0.90)
quantiles <- c(0.25, 0.5, 0.9)  # Specify the quantiles you want to model
models <- lapply(quantiles, function(q) {
  qgam(charity_care_burden ~ s(weighted_SDI_score, k=3),
       data = new_cleaned_data, qu = q)
})
```

```
## Estimating learning rate. Each dot corresponds to a loss evaluation.
## qu = 0.25................done
## Estimating learning rate. Each dot corresponds to a loss evaluation.
## qu = 0.5........done
## Estimating learning rate. Each dot corresponds to a loss evaluation.
## qu = 0.9.....................done
```

```r
# Predict values and confidence intervals for each model
prediction_data <- new_cleaned_data %>%
  dplyr::select(weighted_SDI_score) %>%
  distinct() %>%
  arrange(weighted_SDI_score)

predictions <- do.call(rbind, lapply(1:length(quantiles), function(i) {
  model <- models[[i]]
  pred <- predict(model, newdata = prediction_data, se.fit = TRUE)

  data.frame(
    weighted_SDI_score = prediction_data$weighted_SDI_score,
    fit = pred$fit,
    lower = pred$fit - 1.96 * pred$se.fit,  # 95% CI lower bound
    upper = pred$fit + 1.96 * pred$se.fit,  # 95% CI upper bound
    quantile = quantiles[i]
  )
}))


# Fit a linear quantile regression model for q=0.75
q <- 0.75
linear_qr_model <- rq(charity_care_burden ~ weighted_SDI_score, tau = q, data = new_cleaned_data)


# View the summary of the linear quantile regression model
summary(linear_qr_model)
```

```
##
## Call: rq(formula = charity_care_burden ~ weighted_SDI_score, tau = q,
##     data = new_cleaned_data)
##
## tau: [1] 0.75
##
## Coefficients:
##                    Value   Std. Error t value Pr(>|t|)
## (Intercept)        0.00868 0.00125    6.94933 0.00000
## weighted_SDI_score 0.00026 0.00003    9.01021 0.00000
```

```r
# Generate predictions from the linear quantile regression model
linear_qr_predictions <- data.frame(
  weighted_SDI_score = prediction_data$weighted_SDI_score,
  fit = predict(linear_qr_model, newdata = prediction_data)
)

# Plot with ggplot2, adding actual data points and both qgam and linear quantile regression lines
ggplot() +
  # Plot the actual data points
  geom_point(data = new_cleaned_data,
             aes(x = weighted_SDI_score, y = charity_care_burden),
             color = "black", alpha = 0.5) +

  # Plot the quantile regression lines from qgam
  geom_line(data = predictions, aes(x = weighted_SDI_score, y = fit, color = factor(quantile))) +

  # Plot the confidence interval ribbons from qgam
  geom_ribbon(data = predictions,
              aes(x = weighted_SDI_score, ymin = lower, ymax = upper, fill = factor(quantile)),
              alpha = 0.2) +

  # Add the linear quantile regression line for q=0.75
  geom_line(data = linear_qr_predictions, aes(x = weighted_SDI_score, y = fit),
            color = "red", linetype = "dashed", size = 1,
            inherit.aes = FALSE) +

  # Add labels and title
  labs(
    title = "Quantile Regression (QGAM and Linear) of Charity Care Burden on Weighted SDI Score",
    x = "Weighted SDI Score",
    y = "Charity Care Burden",
    color = "QGAM Quantile",
    fill = "QGAM Quantile"
  ) +
  ylim(c(0, .2)) +

  # Adjust the theme
  theme_minimal() +
  theme(legend.position = "bottom")
```



Results:

The results show a significant positive relationship between the weighted Social Deprivation Index (SDI) score and charity care burden, particularly for hospitals with higher burdens. Hospitals serving more socioeconomically deprived populations tend to incur greater uncompensated care costs. Factors such as urban vs. rural location, teaching status, ownership type, and bed size also influence charity care burden, with nonprofit and urban hospitals displaying distinct patterns.

The analysis indicates that larger hospitals tend to have a higher charity care burden compared to smaller hospitals. This is likely due to the higher volume of patients and the greater capacity to serve socioeconomically deprived populations. Larger hospitals, particularly those in urban areas and those with teaching status, face more significant financial strain from uncompensated care. The models explain a moderate amount of variability, with qgam results highlighting stronger effects in higher quantiles. Overall, the findings suggest that hospitals in more deprived areas and those with larger bed sizes face greater financial strain from uncompensated care.

Quantile Generalized Additive Model (qgam) Results:

The results from the qgam models showing different quantiles (e.g., 25th, 50th, 90th percentiles) of the charity care burden are influenced by factors like:

- Weighted Social Deprivation Index (SDI) Score: There is a significant positive relationship between the weighted SDI score and charity care burden. Hospitals serving more socioeconomically deprived populations tend to incur greater uncompensated care costs. This effect is particularly pronounced in higher quantiles, indicating that hospitals with higher burdens are more affected by socioeconomic deprivation.
- Teaching Status: Teaching hospitals generally bear a higher charity care burden compared to non-teaching hospitals. This is likely due to their larger size, broader range of services, and mission to serve diverse populations, including those from socioeconomically deprived backgrounds.
- Urban vs. Rural Location: Urban hospitals face a higher charity care burden compared to rural hospitals. Urban hospitals often serve larger, more diverse populations, including a higher proportion of socioeconomically deprived individuals, leading to greater uncompensated care costs.
- Ownership Type: Nonprofit hospitals tend to have a higher charity care burden compared to proprietary (for-profit) and governmental hospitals. Nonprofit hospitals often have a mission to serve the community and provide care regardless of patients' ability to pay, which increases their uncompensated care costs.
- Bed Size: Larger hospitals tend to have a higher charity care burden compared to smaller hospitals. This is due to their higher patient volume and greater capacity to serve socioeconomically deprived populations. Larger hospitals, particularly those in urban areas and with teaching status, face more significant financial strain from uncompensated care.
- Critical Access Status: Hospitals that are not designated as critical access hospitals tend to have a higher charity care burden. Critical access hospitals, typically smaller and rural, receive specific funding and support to ensure their viability, which may reduce their uncompensated care burden.
- Referral Center Status: Referral centers tend to have a lower charity care burden compared to non-referral centers. Referral centers often receive patients referred for specialized care, which may involve higher reimbursement rates and lower uncompensated care costs.
- Total Cases: The volume of cases influences the charity care burden. Hospitals with a higher number of total cases tend to have a higher charity care burden, reflecting the increased demand for services and the likelihood of serving more socioeconomically deprived patients.

Significance of Predictors:

From the model summary, it is easy to identify which predictors are statistically significant. For example, here urban_rural_Urban and ownership_type_simple_Proprietary are significant predictor other than SDI score.

- (Intercept): The intercept value is -0.007704, which is the baseline level of charity care burden when all predictors are at their reference levels.
- Critical Access (Not Critical Access): The coefficient is 0.003266, indicating that hospitals not designated as critical access have a higher charity care burden. This effect is highly significant ($p < 0.001$).
- Teaching Status (Teaching): The coefficient is -0.0002544, suggesting a slight decrease in charity care burden for teaching hospitals, but this effect is not statistically significant ($p = 0.682$).
- Urban vs. Rural (Urban): The coefficient is 0.002825, indicating that urban hospitals have a higher charity care burden. This effect is highly significant ($p < 0.001$).
- Referral Center (Referral Center): The coefficient is -0.001860, suggesting that referral centers have a lower charity care burden. This effect is statistically significant ($p = 0.016$).
- Ownership Type (Nonprofit and Proprietary): Nonprofit hospitals have a coefficient of 0.001602 ($p = 0.003$), and proprietary hospitals have a coefficient of 0.001997 ($p = 0.004$), both indicating higher charity care burdens compared to governmental hospitals.

- Type of Service: Various types of services show different effects, but most are not statistically significant except for "General Short Term" (coefficient = 0.01434, $p < 0.001$) and "Other" (coefficient = 0.01352, $p = 0.027$).
- Bed Size (Medium and Small): The coefficients for medium and small bed sizes are not statistically significant.
- Total Cases: The coefficient is 3.775e-07, indicating a positive relationship with charity care burden, and this effect is statistically significant ($p = 0.017$).

Smooth Terms:

Weighted SDI Score: The smooth term for the weighted SDI score is highly significant ($p < 0.001$), indicating a strong non-linear relationship with charity care burden. Other smooth terms (e.g., weighted PovertyLT100__FPL__score, weighted Single__Parent__Fam__score, etc.) show varying levels of significance, with some being highly significant (e.g., weighted HHNo__Vehicle__score, $p < 0.001$).

Smooth Terms Plot:

The plot for the smooth term of the weighted SDI score shows a non-linear relationship with charity care burden. As the weighted SDI score increases, the charity care burden also increases, particularly at higher levels of SDI. Other smooth terms (e.g., weighted PovertyLT100__FPL__score, weighted Single__Parent__Fam__score) show varying patterns, indicating their respective impacts on charity care burden.

visualize qgam model:

The scatter plots and quantile regression lines effectively illustrates the complex relationship between socioeconomic deprivation and charity care burden across different quantiles. Key takeaways include:

- All quantile regression lines (0.25, 0.5, and 0.9) show a positive relationship between the weighted SDI score and charity care burden. This indicates that as the level of socioeconomic deprivation increases, the charity care burden also increases across all parts of the distribution.

- The slope of the quantile regression lines becomes steeper at higher quantiles (e.g., 0.9), indicating that the effect of socioeconomic deprivation on charity care burden is more pronounced for hospitals with higher burdens. This suggests that hospitals already facing high charity care burdens are more sensitive to increases in socioeconomic deprivation.

- The QGAM lines show a non-linear relationship, particularly at higher quantiles. This non-linearity suggests that the impact of socioeconomic deprivation on charity care burden is not constant and may vary depending on the level of deprivation and the existing burden on the hospital.

- The confidence intervals around the quantile regression lines provide a measure of the uncertainty in the estimates. Narrower intervals indicate more precise estimates, while wider intervals suggest greater uncertainty. The intervals are generally wider at higher quantiles, reflecting greater variability in the data for hospitals with higher charity care burdens.

- The red dashed line (linear quantile regression for the 0.75 quantile) provides a linear approximation of the relationship. While it captures the general trend, it may not fully account for the non-linearities observed in the QGAM lines, particularly at higher levels of deprivation.

Model Fit and Deviance Explained:

The R-squared values (or the deviance explained) from the models indicate how well the model explains the variability in the charity care burden. For instance, the qgam model might explain around 24% of the deviance, suggesting that while the model captures some important trends, there may be additional factors influencing the charity care burden.

The model explains a moderate portion of the variability in charity care burden, and visualizations of the plot of “Quantile Regression (QGAM and Linear) of Charity Care Burden on Weighted SDI Score” illustrates these relationships effectively.

The results ultimately shed light on how deprivation and hospital characteristics affect the amount of uncompensated care hospitals provide, which can inform policy or financial support strategies.

Conclusion:

The analysis of the relationship between charity care burden and socioeconomic factors, particularly the Social Deprivation Index (SDI), provides valuable insights into how hospitals across the United States are impacted by the communities they serve. By employing quantile generalized additive models (qgam) and linear quantile regression, this study captures the variability in charity care burden across different levels of deprivation and hospital characteristics.

The results show that hospitals serving more socioeconomically deprived populations, as indicated by higher weighted SDI scores, tend to face a greater charity care burden. This relationship is especially pronounced in hospitals with higher burdens (e.g., in the 75th and 90th percentiles), indicating that deprivation significantly affects hospitals’ ability to recover costs, leading to more uncompensated care.

References:

1. Robert, S. A., & House, J. S. (2000). Socioeconomic inequalities in health: Integrating individual-, community-, and societal-level theory and research. International Journal of Health Services, 30(3), 441–464.

2. Nikpay, S., Buchmueller, T., & Levy, H. (2016). Affordable Care Act Medicaid expansion reduced uninsured hospital stays in 2014. Health Affairs, 35(1), 106–110.

3. Braveman, P., & Gottlieb, L. (2014). The social determinants of health: It’s time to consider the causes of the causes. Public Health Reports, 129(1_suppl2), 19–31.

4. Koenker, R., & Hallock, K. F. (2001). Quantile regression: An introduction. Journal of Economic Perspectives, 15(4), 143–156.

5. Probst, J. C., Bellinger, J. D., & Walsemann, K. M. (2014). Higher risk of hospital admission among rural residents. The Journal of Rural Health, 30(2), 194–201.