

# Test Write Up

Megan

2023-07-13

## Contents

<b>Title: Preventable Hospital Stays</b>	<b>1</b>
<b>Executive Summary:</b>	<b>1</b>
<b>Introduction</b>	<b>2</b>
<b>Motivation</b>	<b>2</b>
<b>Social Economic Factors:</b>	<b>2</b>
Variables that matter . . . . .	6
<b>Health Behaviors</b>	<b>6</b>
Variables Definition: . . . . .	6
Pre-processing Methods: . . . . .	6
Analysis . . . . .	6
Hypothesis . . . . .	6
<b>Exploratory Data Analysis</b>	<b>6</b>
Bivariate Chloropleth . . . . .	7
<b>Modeling</b>	<b>9</b>
Ex: Linear Reg . . . . .	9
Description of Model . . . . .	9
Prediction Result for Linear Reg . . . . .	9
Ex: Random Forest . . . . .	9
Description of Model . . . . .	9
Prediction Result for Random Forest . . . . .	9
CART . . . . .	9
Description of Model . . . . .	9
Prediction Result for CART . . . . .	9
Diagnonstics . . . . .	10
<b>Conclusion</b>	<b>10</b>

## Title: Preventable Hospital Stays

### Executive Summary:

The objective of this research is to identify some of the factors that play a big role in preventable hospital stays in the United States by county, some of those sub factors include: Health Behaviors, and Social and Economic factors. A lot of these factors can differ from county to county.

## Introduction

The data was collected and analyzed from [countyhealthrankings.org](http://countyhealthrankings.org). We researched income inequality, unemployment and high school completion rates to see if it affects the number of preventable hospital stays of certain racial groups at the county level.

## Motivation

We are studying how income inequality, unemployment, and high school completion affect the quantity of preventable hospital stays. By researching this, we may uncover patterns in quality of life, access to care, and socio-economic factors. We aspire to model examples that result in improved disparities and predict quality of life.

## Social Economic Factors:

We examined the rate of preventable hospital stays with Unemployment, High School Completion, and Income inequality.

Results: (insert graphs of Unemployment)

(insert graphs of High school Completion)

(insert graphs of Income inequality)

```
library(readr)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
#SCATTERPLOT
```

```
data = read_csv("data/raw/analytic_data2023_0.csv", skip = 1)
```

```
## Rows: 3194 Columns: 720
```

```
## -- Column specification -----
## Delimiter: ","
## chr   (5): statecode, countycode, fipscode, state, county
## dbl (569): year, county_ranked, v001_rawvalue, v001_numerator, v001_denomina...
## lgl (146): v002_numerator, v002_denominator, v036_numerator, v036_denominato...
##
## 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.
```

```
#had to skip first line in order to get the headers (no longer need to use backticks to reference column)
```

```
#all 50 for 3 variables
```

```
data_all = data %>% select(4,5,153:162, 193:197, 183:187, 218:222)
```

```
#all 50 for asian
```

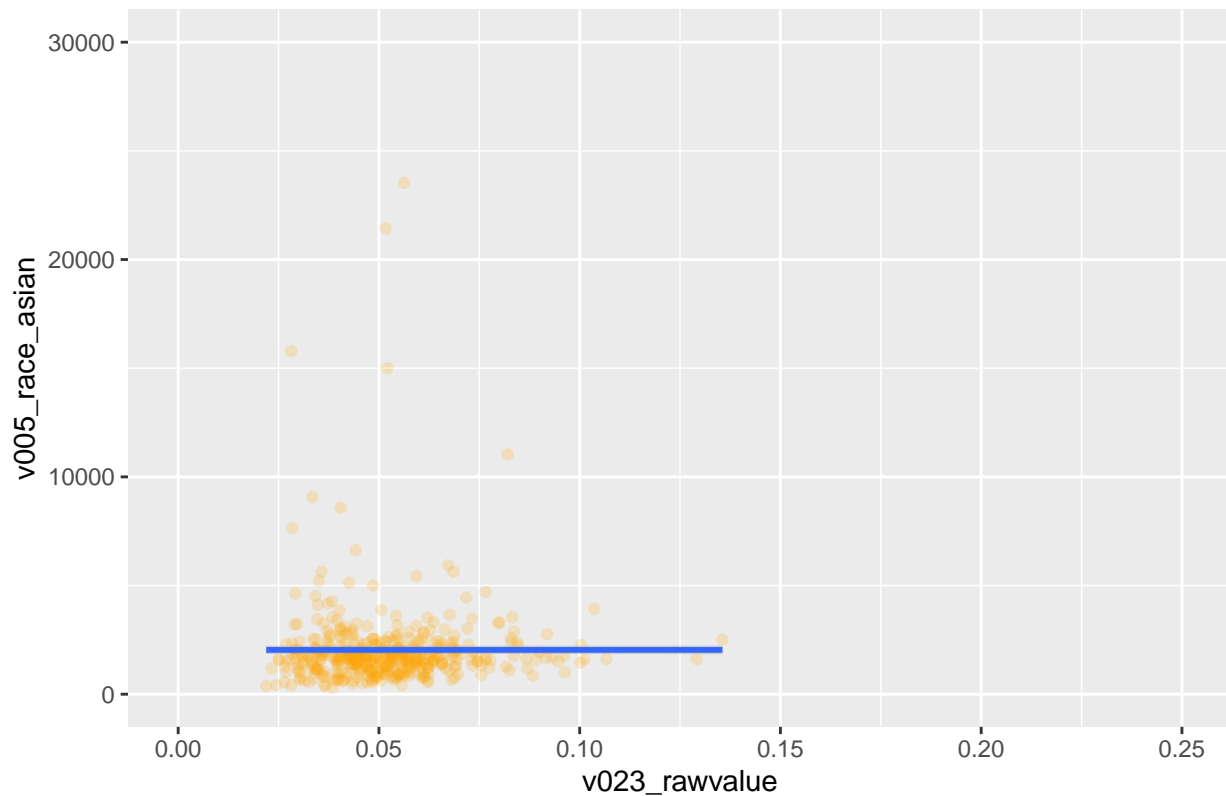
```
ggplot(data_all,aes(x = v023_rawvalue, y = v005_race_asian )) + geom_point(alpha = 0.2, color = "orange"
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 2760 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 2760 rows containing missing values (`geom_point()`).
```

### unemployment vs preventable stays for asians



```
#all 50 for black
```

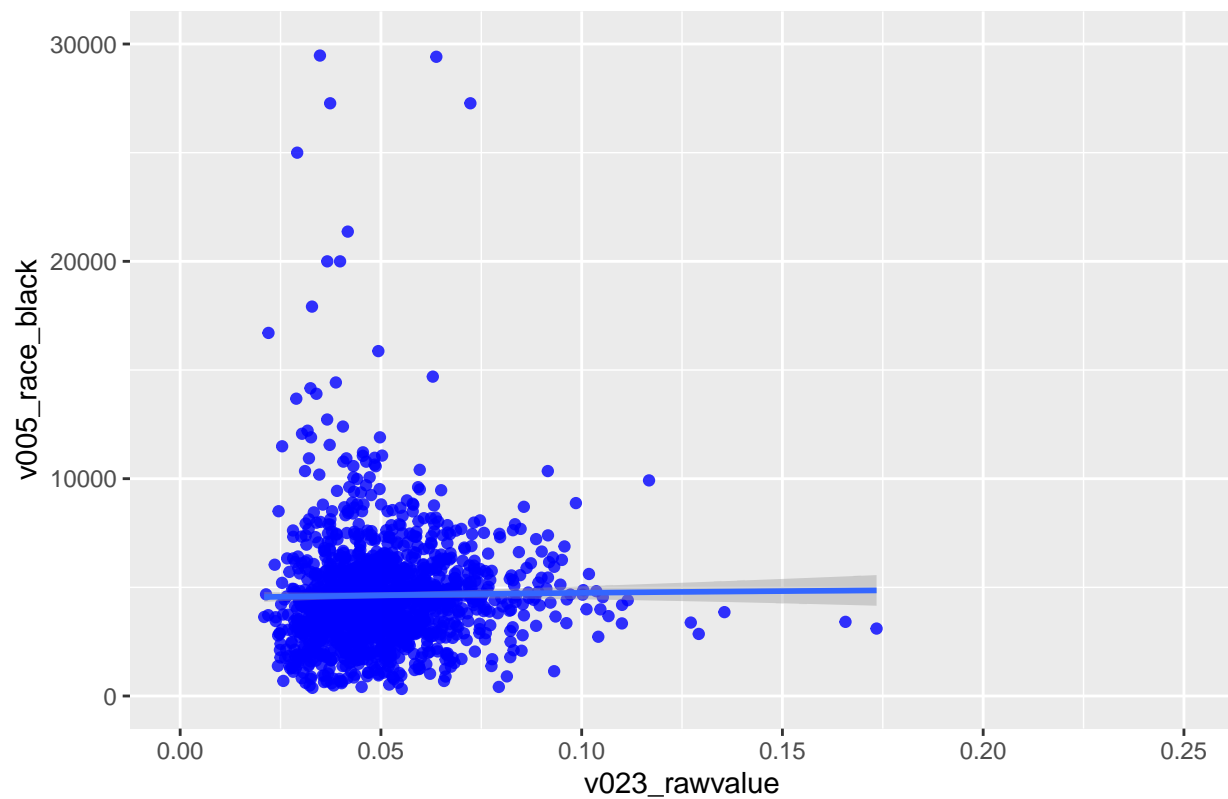
```
ggplot(data_all,aes(x = v023_rawvalue, y = v005_race_black )) + geom_point(alpha = 0.8, color = "blue"
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 1766 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 1766 rows containing missing values (`geom_point()`).
```

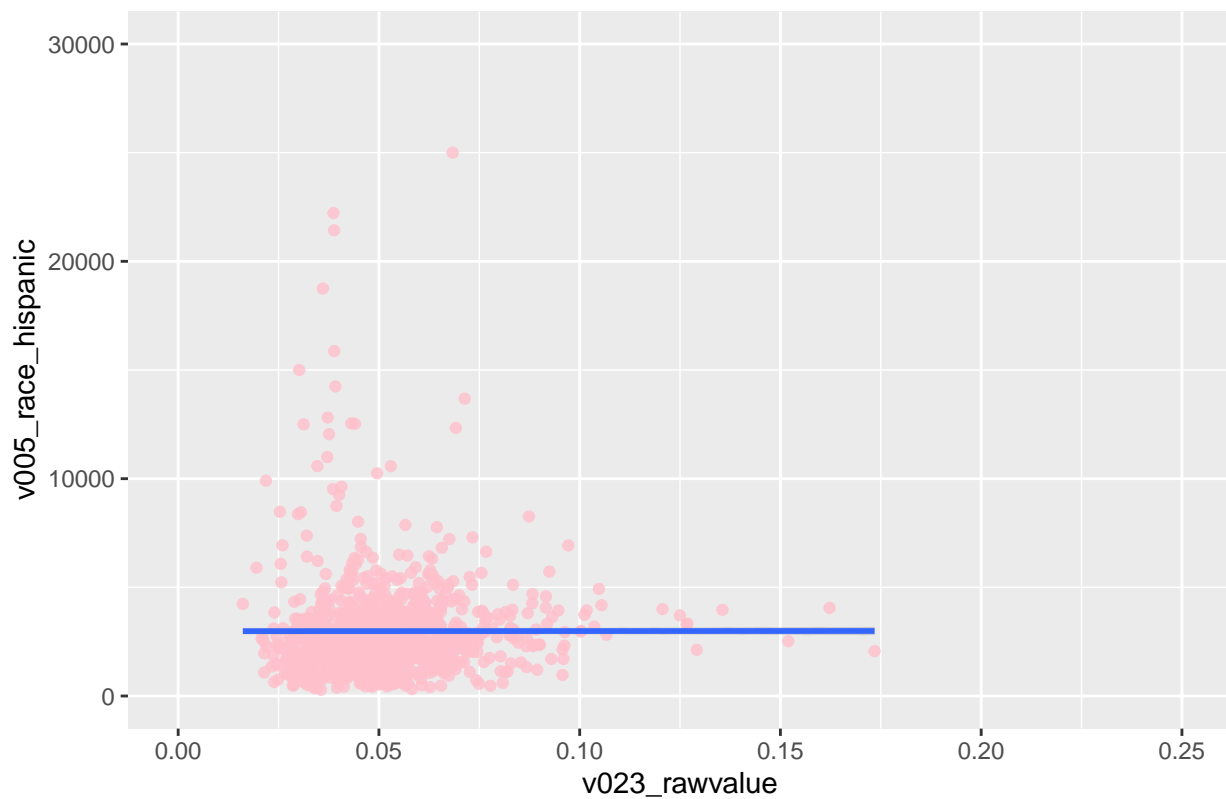
unemployment vs preventable stays for blacks



```
#all 50 for hispanic
ggplot(data_all,aes(x = v023_rawvalue, y = v005_race_hispanic )) + geom_point(alpha = 0.8, color = "pink")

## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
## Warning: Removed 2222 rows containing non-finite values (`stat_smooth()`).
## Warning: Removed 2222 rows containing missing values (`geom_point()`).
```

## unemployment vs preventable stays for hispanics



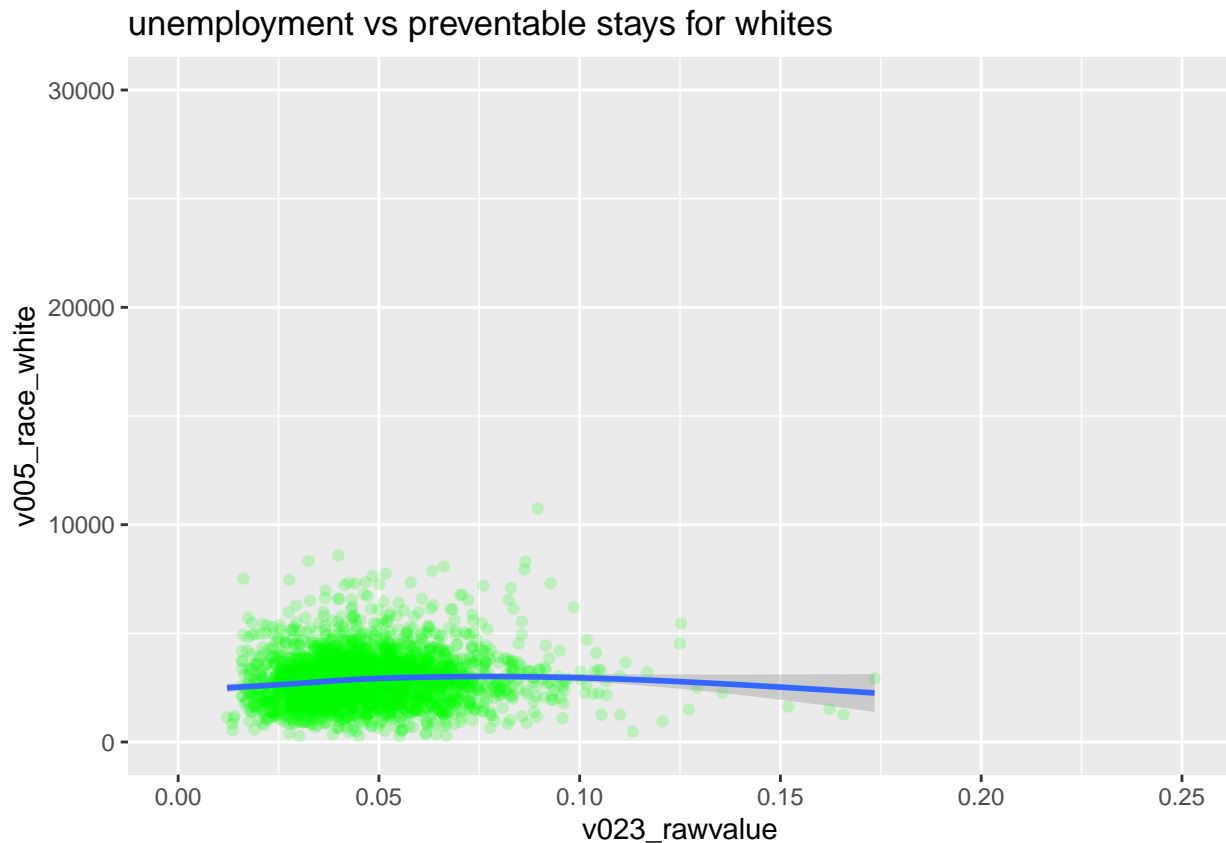
*#all 50 for white*

```
ggplot(data_all, aes(x = v023_rawvalue, y = v005_race_white )) + geom_point(alpha = 0.2, color = "green"
```

```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 105 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 105 rows containing missing values (`geom_point()`).
```



Based on the factors of Unemployment, High School Completion, and Income Inequality we were not able to see a lot of contribution for Preventable Hospital stays, therefore we decided to move into: Community safety as it reflects not only violent acts in neighborhoods and homes, but also injuries caused unintentionally through accidents.

### Variables that matter

- PCP Ration
- Test

## Health Behaviors

During this research we also took into consideration Health Behaviors such as Drug Overdose Deaths with Preventable Hospital Stays, which contribute a big part of the findings. Number of drug poisoning deaths per 100,000 population. The 2023 County Health Rankings used data from 2018-2020 for this measure. Drug overdose deaths are a leading contributor to premature death and are largely preventable.

### Variables Definition:

### Pre-processing Methods:

### Analysis

### Hypothesis

## Exploratory Data Analysis

explain process with EDA ##Scatterplots put code chunks here with stuff

```
##Chloropleths put code chunks here with stuff
```

## Bivariate Chloropleth

```
access_care = data[, c("state", "fipscode", "county", "v005_rawvalue", "v005_numerator", "v005_denominator")]
```

```
##MAKE THIS TIDIER; JUST AN EXAMPLE CHUNK FOR NOW  
#BIVARIATE CHLOROPLETH
```

```
library(usmap)  
library(ggplot2)  
library(biscale)  
library(maps)  
library(usdata)
```

```
state_borders <- map_data("state") %>%  
mutate(state = state2abbr(region)) %>%  
select(-subregion)  
head(state_borders)
```

```
##      long      lat group order  region state  
## 1 -87.46201 30.38968     1     1  alabama  AL  
## 2 -87.48493 30.37249     1     2  alabama  AL  
## 3 -87.52503 30.37249     1     3  alabama  AL  
## 4 -87.53076 30.33239     1     4  alabama  AL  
## 5 -87.57087 30.32665     1     5  alabama  AL  
## 6 -87.58806 30.32665     1     6  alabama  AL
```

```
bi_chloro = access_care %>%  
# mutate(v004_other_data_1 = 1/ v004_other_data_1) %>%  
#filter(!is.na(v004_other_data_1) & !is.na(v062_other_data_1) & state != "US") %>%  
group_by(state)
```

```
#join  
#bi_chloro_plot <- state_borders %>%  
#left_join(bi_chloro,  
#by = "state")
```

```
library(cowplot)  
library(sf)
```

```
## Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE
```

```
library(biscale)  
library(tidycensus)  
library(sf)  
library(tidyr)
```

```
#county_geospatial_df =  
# tidycensus::get_acs(geography = "county",  
#                      variables = "B19013_001",  
#                      geometry = TRUE, year = 2021) %>%  
# select(-variable, -moe, -estimate) %>%  
# separate(col = 'NAME', into = c('county', 'state'), sep = ',') %>%  
# mutate(state = usdata::state2abbr(state)) %>%
```

```

# tigris::shift_geometry()
# Save acs data
#saveRDS(county_geospatial_df, "./data/external/county_acs.rds")
# Load acs data
county_geospatial_df = readRDS("./data/external/county_acs.rds")

n_color_dim = 2
data_bi1 = bi_class(bi_chloro, x = v005_rawvalue, y = v004_rawvalue, style = "quantile", dim = n_color_dim)

## Warning in classInt::classIntervals(.data[[var]], n = dim, style = style): var
## has missing values, omitted in finding classes

## Warning in classInt::classIntervals(.data[[var]], n = dim, style = style): var
## has missing values, omitted in finding classes

#CHECK THS CODE
#data_bi1 = data_bi1
#%>%
# rename(biVar = bi_class)
# mutate(biVar = bi_class) %>%
# select(-(1:4)) %>%
# select(-bi_class)

#data_bi1$bi_class = rep("3-3", nrow(data_bi1))
#join two dataframes
county_bivar = merge(data_bi1, county_geospatial_df, by = c("state", "county"))

description = readRDS("./data/processed/colDesc_analysis2023.rds")
#description[["v004_other_data_1"]] = "PCP Ratio"#"Population / PCPs."
map <- ggplot() +
  geom_sf(data = county_bivar, mapping = aes(fill = bi_class, geometry = geometry), color = "lightgray") +
  bi_scale_fill(pal = "GrPink", dim = n_color_dim, na.value="transparent") +
  labs(
    title = "PCP vs PHS Across US",) +
  bi_theme()

legend <- bi_legend(pal = "GrPink",
  dim = n_color_dim,
  xlab = sub(" raw value", "", description[["v005_rawvalue"]]),
  ylab = sub(" raw value", "", description[["v004_rawvalue"]]),
  size = 5)

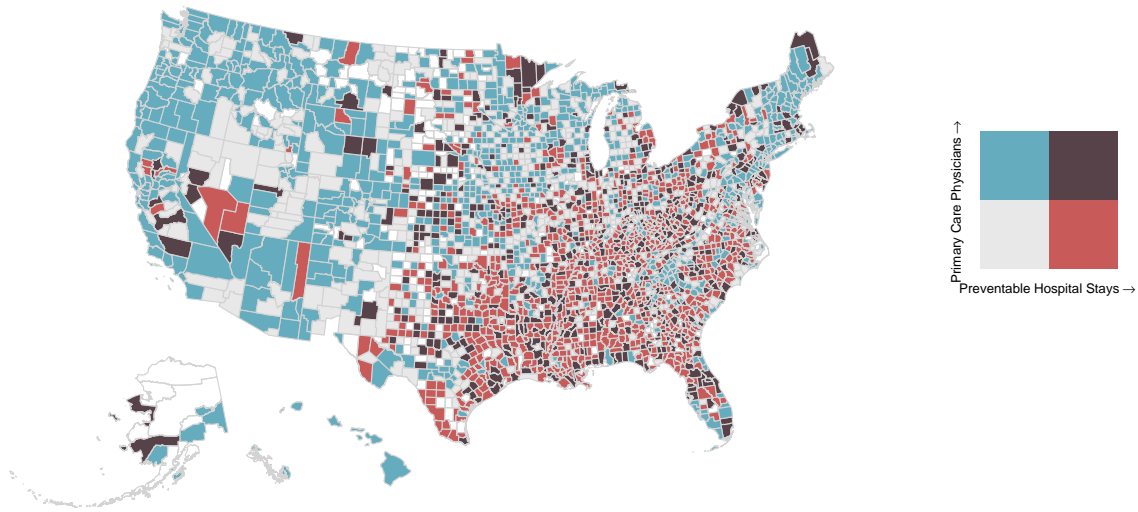
finalPlotCow = cowplot::plot_grid(map, legend, ncol = 2, rel_widths = c(0.8, 0.2))

finalPlotCow

```



# PCP vs PHS Across US



Write information that was gathered from EDA

## Modeling

### Ex: Linear Reg

inset code chunk with model

#### Description of Model

- include specific considerations
- tuning

#### Prediction Result for Linear Reg

### Ex: Random Forest

#### Description of Model

- include specific considerations
- tuning

#### Prediction Result for Random Forest

## CART

#### Description of Model

- include specific considerations
- tuning

#### Prediction Result for CART

##Comparing RMSE

**Dia<sup>gn</sup>onstics**

**Conclusion**