Hospital Funding and Health Equity

Meredith Klashman & Ruben Vargas

12/13/2021

Background

Problem Statement

The Department of Health Equity has recently funded a public-private partnership to improve rural California facilities that have been underfunded by OSHPD in the past 5 years (2014-2019). In this study, we aim to identify and evaluate the five counties best suited to receive this funding based on prioritized demographic metrics: high percentage of non-homeowners, large number of aging individuals, and small population size.

Methods

A mortality dataset was used to analyze the classification and numbers of deaths that occurred in each county between 2014 and 2019. To clean the data, all NA values were replaced with 0. An additional variable was created to classify the deaths as from chronic or acute conditions. Chronic conditions were defined as the following diseases: Alzheimer's disease, Chronic lower respiratory diseases Diabetes mellitus, Essential hypertension and hypertensive renal disease, Chronic liver disease and cirrhosis, and Parkinson's disease. The dataset was filtered for chronic conditions, metrics that represented the total population, and geographic occurrences. The sum of mortality from chronic conditions was calculated by grouping by year and county.

A county demographic dataset was used to analyze demographics of particular interest for project funding (population size, average age of population, home ownership status). The dataset was collected in 2012. A metric calculating the ratio of renters to owners in each county was created by dividing the sum of renters by the sum of owners for each line item.

To select 5 counties of focus, the population per square mile was filtered to less than 30 miles (the lowest quartile of population per square mile in the data). The median age was filtered to greater than 30 to target the counties with the highest quartile of median age (represents counties with an aging population). The ratio of renters to owners was filtered for greater than 0.6 to target counties within the highest quartile of rentership.

The hospital cost closure dataset was used to determine OSHPD's history of closures and financial investment in different counties. The data spans the years 2013 to 2021. Data was filtered for projects that were in closure.

Each dataset was merged by county name, and analyzed by comparative analysis using tables or graphical plots. The project's desired population was prioritized for visualization.

```
getwd()
```

```
## [1] "/home/rstudio/PHW251_MK_RV"
```

```
countydemo <- read.csv("/home/rstudio/PHW251_MK_RV/ca_county_demographics.csv",
                       header=TRUE)
mortality <- read.csv("/home/rstudio/PHW251_MK_RV/mort_by_county.csv", header =
                        TRUE)
hospital_cost <- read.csv("/home/rstudio/PHW251_MK_RV/oshpd_hospital_cost.csv",
                          header = TRUE)

#Edit Mortality to replace NA with 0

mortality[is.na(mortality)] = 0

#Create a Variable to Indicate Chronic Condition

#Chronic Health Conditions defined as the following:
    #Alzheimer's disease
    #Chronic lower respiratory diseases
    #Diabetes mellitus
    #Essential hypertension and hypertensive renal disease
    #Chronic liver disease and cirrhosis
    #Parkinson's disease

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
```

```
mortality <- mortality %>%
  mutate(Chronic = case_when(Cause_Desc %in% c("Alzheimer's disease",
                    "Chronic lower respiratory diseases", "Diabetes mellitus",
                    "Essential hypertension and hypertensive renal disease",
                    "Chronic liver disease and cirrhosis",
                    "Parkinson's disease") ~ "Yes",
                    TRUE ~ "No"))

#subtotal for mortality from chronic in each county per year

library(tidyr)
mortality_chronic <- mortality %>%
              filter (Geography_Type == "Occurrence") %>%
              filter(Chronic == "Yes")%>%
              filter(Strata == "Total Population") %>%
              group_by(County,Year) %>%
              mutate(County_Year_Chronic_Mortality = sum(Count))
```

```r
#Calculate Rent vs Homeowners Ratio

library(dplyr)

countydemo <- countydemo %>%
              mutate(RatioRenttoOwn = renter_occ/owner_occ) %>%
              mutate (County = name)


#locate 5 counties that share three common attributes: low population per
#square mile 'pop12_sqmi1', high median age
#'med_age', a high proportion of renters vs. homeowners

summary(countydemo$pop12_sqmi)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##     1.544    25.887   103.424   665.061   333.485 17398.354
```

```r
summary(countydemo$med_age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   29.60   33.70   37.05   38.49   43.08   51.00
```

```r
summary(countydemo$RatioRenttoOwn)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3007  0.5212  0.6272  0.6472  0.7382  1.7968
```

```r
countydemo5 <- countydemo %>%  filter (pop12_sqmi < 30) %>%
               filter (med_age > 30) %>%
               filter (RatioRenttoOwn > 0.6)


#locate the most recent account of OSHPD funding for projects that are in
#closure

hospital_cost_closure <- hospital_cost %>% filter(OSHPD.Project.Status ==
                                                    "In Closure")

#Optimize data for joining (edit so all county naming is same format)

library(tidyr)

hospital_cost_closure_clean <- hospital_cost_closure %>%
                      separate(County, c('Number','County_Name_1',
                      'County_Name_2', 'County_Name_3' ))
```

```
## Warning: Expected 4 pieces. Missing pieces filled with 'NA' in 11856 rows [1, 2,
## 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
hospital_cost_closure_clean [is.na(hospital_cost_closure_clean)] <- ""
hospital_cost_closure_clean <- hospital_cost_closure_clean %>% unite(County,
                               'County_Name_1','County_Name_2','County_Name_3',
                               sep = " ", remove = TRUE )

county_demo_clean <- countydemo %>% separate(name, c('County_Name_1',
                      'County_Name_2', 'County_Name_3' ))
```

```
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 57 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
county_demo_clean [is.na(county_demo_clean )] <- ""
county_demo_clean <- county_demo_clean %>%
                      unite(County,'County_Name_1','County_Name_2',
                      'County_Name_3', sep = " ", remove = TRUE )

mortality_chronic_clean <- mortality_chronic %>% separate(County, c(
                      'County_Name_1','County_Name_2', 'County_Name_3' ))
```

```
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 2052 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
mortality_chronic_clean [is.na(mortality_chronic_clean)] <- ""
mortality_chronic_clean <- mortality_chronic_clean %>%
                      unite(County,'County_Name_1','County_Name_2',
                      'County_Name_3', sep = " ", remove = TRUE )

Merged_Dataset <- left_join(hospital_cost_closure_clean, county_demo_clean,
                  by = "County", copy = TRUE)
Merged_Dataset <- left_join(Merged_Dataset, mortality_chronic_clean,
                  by = "County", copy = TRUE)

library(formattable)
hospital_cost_closure_clean$Total.Costs.of.OSHPD.Projects <-
  currency(hospital_cost_closure_clean$Total.Costs.of.OSHPD.Projects,
                    digits = 0L, big.mark = ",")
```

Results

Figure 1: Plot of Requested Population Parameters

```
library(ggplot2)

ggplot(data = county_demo_clean) + geom_point(mapping = aes(x = med_age, y = RatioRenttoOwn,size= pop12
)+theme_bw()+labs(x = "Median Age", y = "Ratio Renters to Owners", title =
"Median Age vs. Renters to Owners vs. Population Size", subtitle =
"Represented at the County Level", size = "Population per SqMi")
```

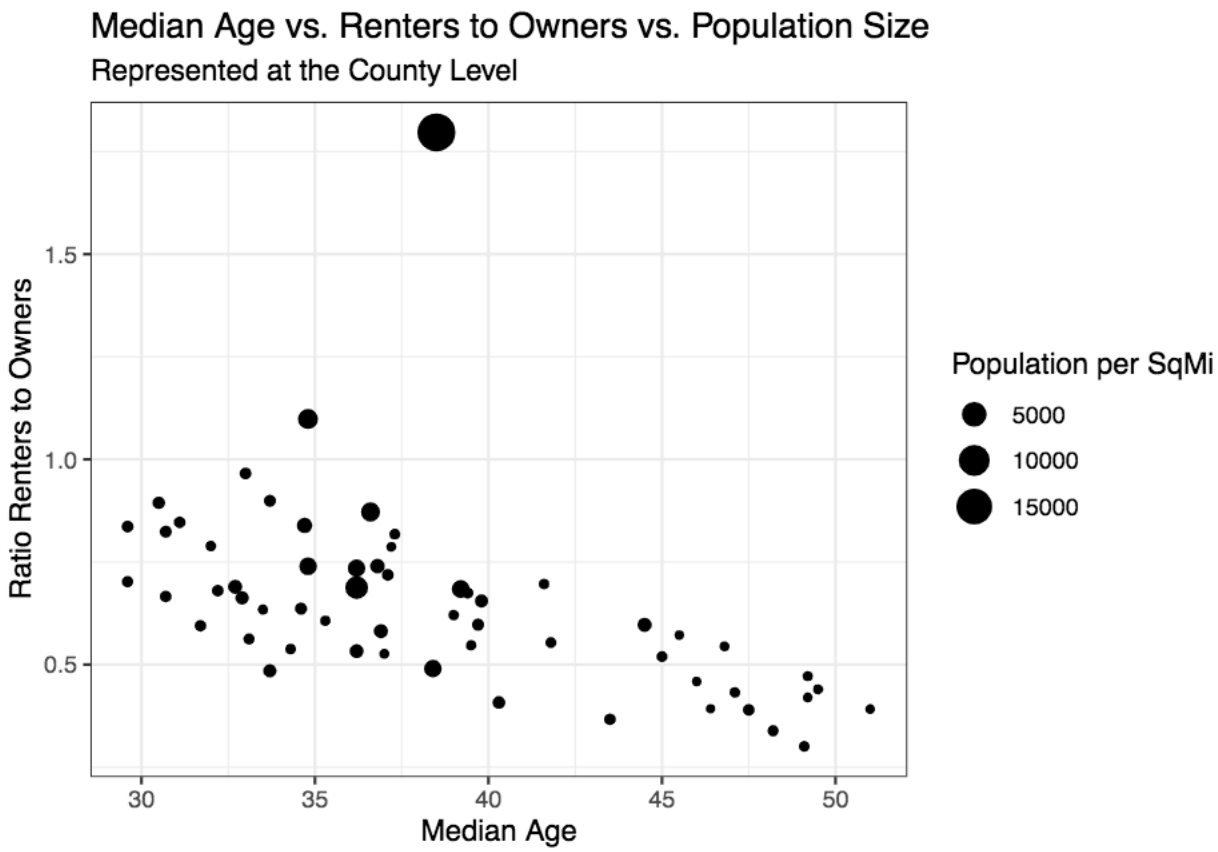

Figure 2: Table of Chronic Mortality for Selected Counties

```
library(knitr)

Merged_Dataset$County <- trimws(Merged_Dataset$County, which = c("right"))
x <- Merged_Dataset %>% filter(County == c("Mendocino", "Mono", "Colusa",
                                           "Del Norte", "Glenn"))
```

```
## Warning in County == c("Mendocino", "Mono", "Colusa", "Del Norte", "Glenn"):
## longer object length is not a multiple of shorter object length
```

```
x1 <- x %>% group_by(County) %>%
  summarize(total_mortality = sum(Count))
x2 <- left_join(countydemo5, x1, by = "County") %>%
  select(c("County", "pop2012", "pop12_sqmi", "med_age", "owner_occ",
           "renter_occ", "RatioRenttoOwn", "total_mortality")) %>%
  group_by(County) %>%
  mutate(total_mortality = sum(total_mortality))
kable(x2, digits = 2, booktabs = F, longtable = F, col.names = c("County",
      "Population", "Population per Square Mile", "Median Age", "Home Owners",
      "Renters", "Renter to Homeowner Ratio", "Total Mortality"),
      caption = "2014-2019 Chronic Mortality and
      Demographic Data of 5 at Risk Counties")
```

Table 1: 2014-2019 Chronic Mortality and Demographic Data of 5 at Risk Counties

County	Population	Population per Square Mile	Median Age	Home Owners	Renters	Renter to Homeowner Ratio	Total Mortality
Mendocino	88094	25.08	41.6	20601	14344	0.70	23096
Mono	14418	4.60	37.2	3228	2540	0.79	0
Colusa	21780	18.83	33.5	4318	2738	0.63	1574
Del Norte	28685	28.30	39.0	6114	3793	0.62	6692
Glenn	28516	21.49	35.3	6100	3700	0.61	1134

Figure 3: Table of OSHPD Funding for Selected Counties

```
x3 <- x %>% group_by(County) %>%
  summarize(total_mortality = sum(Count))
hospital_cost_closure_clean$County <- trimws(hospital_cost_closure_clean$County,
                                             which = c("right"))
x4 <- left_join(x3, hospital_cost_closure_clean, by = "County") %>%
  select(c("County", "Total.Costs.of.OSHPD.Projects",
           "Number.of.OSHPD.Projects", "total_mortality")) %>%
  group_by(County) %>%
  summarize(total_cost = sum(Total.Costs.of.OSHPD.Projects),
            total_projects = sum(Number.of.OSHPD.Projects))
x5 <- left_join(x4, x3, by = "County")
x6 <- left_join(x2, x5, by = "County")
x7 <- x6 %>% select(c("County", "pop12_sqmi", "RatioRenttoOwn", "total_cost", "total_projects", "total_
kable(x7, digits = 2, booktabs = F, longtable = F, col.names = c("County",
    "Population per Square Mile", "Renter to Homeowner Ratio",
    "Total Cost of OSHPD Projects", "Number of OSHPD Projects",
    "Total Chronic Mortality"),
    caption = "OSHPD Funding for Projects that are in Closure in 5 Targeted Counties between 2014 and 2
```

Table 2: OSHPD Funding for Projects that are in Closure in 5 Targeted Counties between 2014 and 2019

County	Population per Square Mile	Renter to Homeowner Ratio	Total Cost of OSHPD Projects	Number of OSHPD Projects	Total Chronic Mortality
Mendocino	25.08	0.70	$2,464,121,218	931	23096
Mono	4.60	0.79	$15,525,102	52	0
Colusa	18.83	0.63	$3,496,981	128	1574
Del Norte	28.30	0.62	$77,987,586	268	6692
Glenn	21.49	0.61	$1,887,885	24	1134

Interpretation & Discussion

Figure 1 shows an inverse relationship between median age of the county population and the proportion of renters to homeowners within those counties. The smaller proportion of renters to homeowners as the median age increases is possibly due to the fact that older populations more than likely are homeowners. It should also be noted that the larger populations, in terms of population per square mile, tend to be in the middle of the data in both overall median age as well as renter to owner ratio.

Upon further filtering of the data, we found 5 counties that shared the three similar attributes that we are using to evaluate which counties should receive additional funding. The attributes are as follows: low population per square mile, high overall median age, and a high proportion of renters to homeowners. This information is outlined in Figure 2. We also included the total mortality caused by chronic diseases with this information so we were able to see how large of an impact chronic disease had within these smaller populated counties. The mortality from chronic diseases between the years 2014 to 2019 in 4 out of the 5 counties of interest is quite large. We do not see any records of mortality from chronic diseases in Mono county, but this does not seem realistic (especially over the span of 5 years). We concluded that low mortality rates can be attributed to a lack of responses (NA inputs) which were replaced with 0 in the data cleanup process.

Additionally, the county is the smallest in terms of population per square mile, which could have been a factor leading to incomplete data/results.

Our final figure incorporates OSHPD funding within the counties of interest for projects that had been labeled as "In Closure" between the years 2014 to 2019. Although it may look as though there has been significant funding to these counties, one must keep in mind that this is over a 5 year time period. We also chose to include the number of projects that were in each county to contextualize on average how much funding was given per projects. Given the population per square mile and proportion of renters to homeowners, in conjunction with total chronic mortality between the years 2014 and 2019, we strongly recommend that Colusa, Del Norte and Glenn counties be strongly considered for the funding. We also recommend that Mono county be considered despite the discrepancy in chronic mortality, as they are a county that has a small population per square mile, a large ratio of renters to homeowners and minimal funding in OSHPD projects over the past 5 years.