Milestone 4

Meredith Klashman and Ruben Vargas

11/29/2021

Data Connection and Cleanup from Milestone 3

```
getwd()
```

```
## [1] "/Users/rubenvargas/Desktop/Public Health W251/PHW251_MK_RV"
```

```
# countydemo <- read.csv("/home/rstudio/PHW251_MK_RV/ca_county_demographics.csv",
                       # header=TRUE %>%
  #select(-c("X", "families", "hse_units", "ave_fam_sz", "vacant", "county_fips")) %>%
 # rename(County = name))

#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)

countydemo <- read.csv("/Users/rubenvargas/Desktop/Public Health W251/PHW251_MK_RV/ca_county_demographi
                       header = T) %>%
  select(-c("X", "families", "hse_units", "ave_fam_sz", "vacant", "county_fips")) %>%
  rename(County = name)

mortality <- read.csv("/Users/rubenvargas/Desktop/Public Health W251/PHW251_MK_RV/mort_by_county.csv",
                       header = T)

hospital_cost <- read.csv("/Users/rubenvargas/Desktop/Public Health W251/PHW251_MK_RV/oshpd_hospital_co
                      header = T)
```

```
#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

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

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

```
#Calculate Rent vs Homeowners Ratio

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

```
#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

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)

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(County, 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 4104 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 )
```

Milestone 4 Visualizations

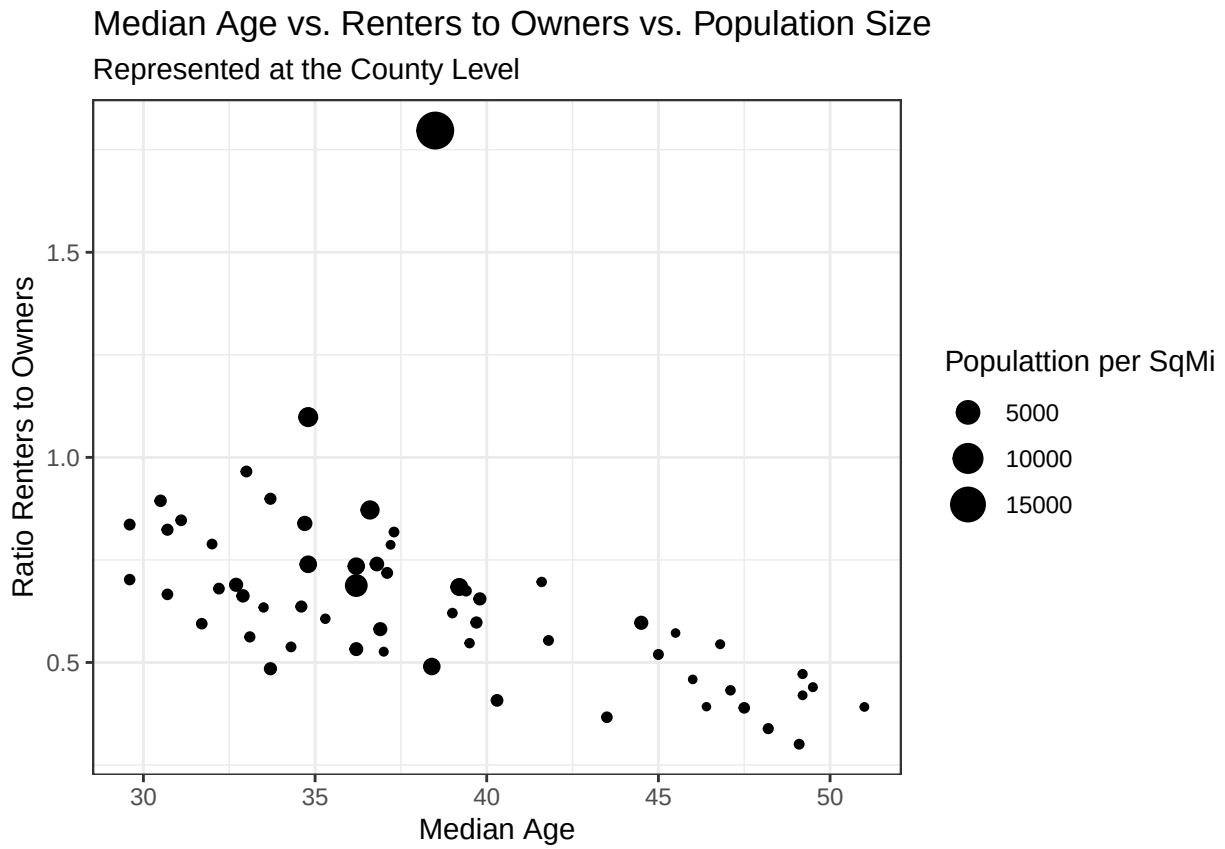
Join Datasets for Analysis

```
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)
```

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 = "Populattion per SqMi")
```



Interpretation: County Level Data indicates that Median Age has an inverse relationship with proportion of Renters: older populations have a higher proportion of home ownership. The largest populations tend to sit within the mode of the dataset in terms of both Median Age and Ratio of Renters to owners.

Figure 2: Table

```
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", "
  group_by(County) %>%
  mutate(total_mortality = sum(total_mortality))

kable(x2, digits = 2, booktabs = F, longtable = F, col.names = c("County", "Population", "Population pe
```

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

County	Population	Population per Square Mile	Median Age	Owners	Renters	Renter to Homeowner Ratio	Total Mortality
Mendocino	88094	25.08	41.6	20601	14344	0.70	51112
Mono	14418	4.60	37.2	3228	2540	0.79	0
Colusa	21780	18.83	33.5	4318	2738	0.63	3944
Del Norte	28685	28.30	39.0	6114	3793	0.62	14467
Glenn	28516	21.49	35.3	6100	3700	0.61	9395

Interpretation: Given the 5 counties that we determined to have 3 similar attributes of low population per square mile, high median age, and a high proportion of renters vs. homeowners, we see that the chronic mortality is quite significant in 4/5 counties. Mono County may have no results due to lack of responses (NA inputs) or due to how small the population in the county is.

Figure 3: Table of Choice

```
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_m

kable(x7, digits = 2, booktabs = F, longtable = F, col.names = c("County", "Population per Square Mile"
```

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	2464121218	931	51112
Mono	4.60	0.79	15525102	52	0
Colusa	18.83	0.63	3496981	128	3944
Del Norte	28.30	0.62	77987586	268	14467
Glenn	21.49	0.61	1887885	24	9395

Interpretation: We can use this chart to make recommendations that these 5 counties would benefit from more funding in OSHPD projects based on the minimal amount they have gotten in the passed and argue it would lessen chronic mortality.