R Project Milestone 3

Moyra Rasheed, Courtney Coon, Jarett Maycott

2022-10-28

```
## 'summarise()' has grouped output by 'county'. You can override using the
## '.groups' argument.
```

Visualization 1:

Table

```
# library(kableExtra)

# table<-merged_data%>%
#  rowwise() %>%
#  mutate(number_highs= sum(c_across(2:6) == "High priority", na.rm = TRUE),
#         number_mediums= sum(c_across(2:6) == "Medium priority", na.rm = TRUE),
#         temp_rank=(number_highs*2)+number_mediums
#         )%>%
#  ungroup()%>%
#  arrange(desc(temp_rank))%>%
#  select(-c(number_highs, number_mediums,temp_rank))%>%
#  slice(1:10)
# head(table)
#
# kable(table,
#       col.names = c("County","Chronic disease mortality burden",
#                     "Previous spending on projects",
#                     "Population density", "Median age of population",
#                     "% population that are renters"),
#       caption="Top 10 Counties ranked by need for oshpd projects.",
#       booktabs=TRUE,
#       align='lccccc')%>%
#  kable_styling(latex_options="scale_down")
#
#  table
```
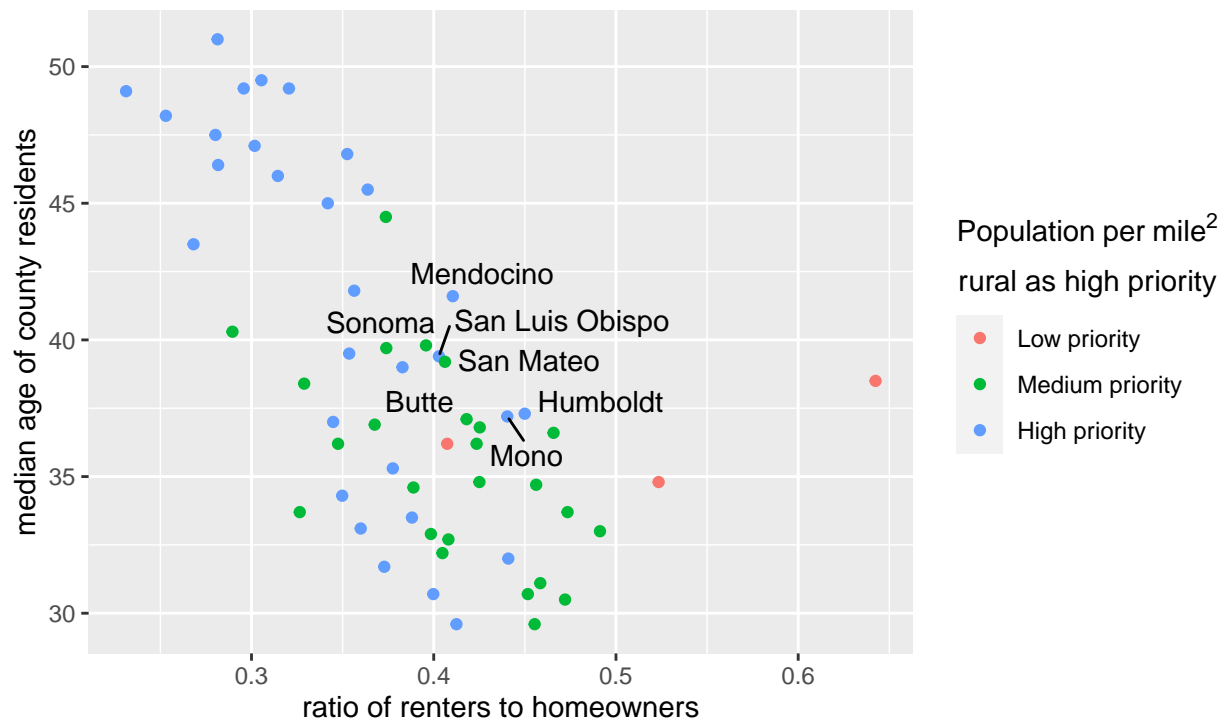
Visualization 2:

Using demographic data to rank counties

```
## renter ratio median = 39%
## median age median = 37.05
## population density 1st quantile (low cutoff) = 25.887
## population density 3rd quantile (high cutoff) = 333.485
ggplot(data = merged_data, aes(x = renter_ratio, y = med_age)) +
geom_point(data = merged_data, aes(x = renter_ratio, y = med_age,
                                   color = pop12_sqmi_CAT)) +
geom_text_repel(aes(label=ifelse((med_age > 37 & renter_ratio > 0.39
    & (pop12_sqmi_CAT=="High priority"| pop12_sqmi_CAT=="Medium priority")),
    county, "")))+
  labs(title = "Demographic data with priority counties identified:",
subtitle = "counties with high median age (>37yo), high ratio of renters (>39%),
and low or medium population density (<333 people/sqmi)",
      x = "ratio of renters to homeowners",
      y = "median age of county residents",
      color =
        bquote(atop(Population~per~mile^{"2"}, "rural as high priority")))+
  theme(plot.title=element_text(hjust=0.5),
        plot.subtitle=element_text(hjust=0.5))
```


Demographic data with priority counties identified:

counties with high median age (>37yo), high ratio of renters (>39%), and low or medium population density (<333 people/sqmi)

Visualization 3:

Using mortality and investment data to rank counties

```r
## make data set with continuous data and ranking factor for the demographic
## data in the first figure
second_fig_data_temp<-merged_data%>%
  select(c("county", "pop12_sqmi_CAT", "med_age_CAT", "renter_ratio_CAT"))%>%
 rowwise() %>%
 mutate(number_highs= sum(c_across(2:4) == "High priority", na.rm = TRUE),
        number_mediums= sum(c_across(2:4) == "Medium priority", na.rm = TRUE),
        temp_rank=(number_highs*2)+number_mediums
        )%>%
  ungroup()%>%
  select(c("county", "temp_rank"))

second_fig_data_final<-full_join(second_fig_data_temp, merged_data, by="county")

## make the figure
## summed chronic disease mortality median = 13413
## log(summed chronic disease mortality median) = log(13413) = 9.50398
## summed total cost median = 5961782208
## log(summed total cost median) = log(5961782208) = 22.50864
ggplot(data = second_fig_data_final,
       aes(x = log(summed_chronic_dis_mort), y = log(summed_total_cost))) +
geom_point(data = second_fig_data_final,
           aes(x = log(summed_chronic_dis_mort), y = log(summed_total_cost),
                                 color = as.factor(temp_rank))) +
guides(color = guide_legend(reverse=TRUE))+
 geom_text_repel(aes(label=ifelse(
   (summed_chronic_dis_mort >= 13413 & summed_total_cost<=5961782208
   & temp_rank >1), county, "")), max.overlaps = Inf)+
labs(title = "Additional data with priority counties identified:",
 subtitle = "counties with high chronic disease mortality,
low previous investment, and high priority based on demographics",
       x = "Log(total dollars invested in previous projects)",
       y = "Log(number of chronic disease mortality events \n between 2015-2020)",
       color = "priority ranking \n score based on \n demographics data") +
   theme(plot.title=element_text(hjust=0.5),
         plot.subtitle=element_text(hjust=0.5))
```

Additional data with priority counties identified:
counties with high chronic disease mortality,
low previous investment, and high priority based on demographics
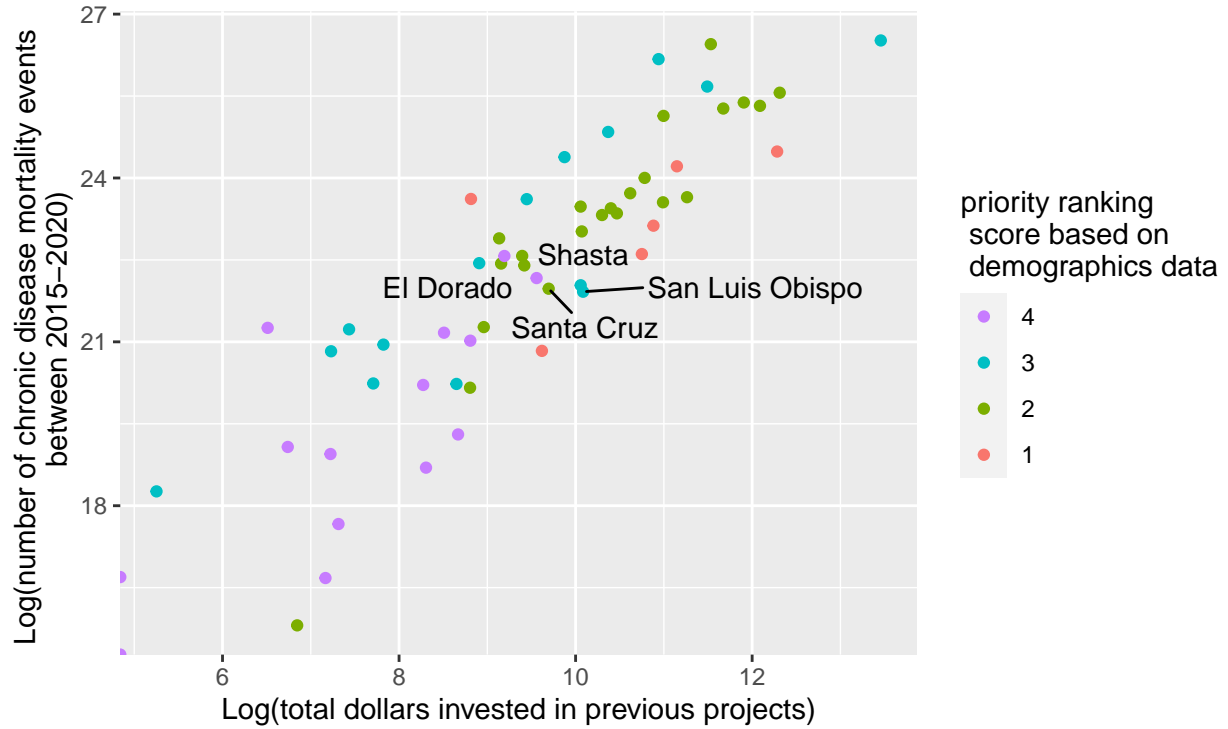
##PROBLEM - demographics ranking score is not matching with figure