Schools

Lucy M Li

14/11/2020

Read in data

```
bay_area_counties <- c("Alameda", "San Francisco", "San Mateo", "Santa Clara")
# SOC school types filter
soc_types <- c("08", "60", "61", "62", "63", "64", "65", "66", "67")
```

```
shapefile <- readOGR("California_School_District_Areas_2018-19-shp", "DistrictAreas1819") %>% spTransfo
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/Lucy/Google Drive lucymqli/Personal/relationship/Buying a house/Schools/California_S
## with 944 features
## It has 69 fields
```

```
shapefile <- shapefile[shapefile@data$CountyName %in% bay_area_counties, ]
shapefile_counties_df <- fortify(shapefile, region="CountyName")
shapefile_df <- fortify(shapefile)
```

```
ca_sch <- read_tsv("pubschls.txt") %>%
  filter(County %in% bay_area_counties, StatusType=="Active", SOC %in% soc_types) %>%
  mutate(Longitude=as.numeric(Longitude), Latitude=as.numeric(Latitude)) %>%
  filter(Longitude>=min(shapefile_df$long),
         Longitude<=max(shapefile_df$long),
         Latitude>=min(shapefile_df$lat),
         Latitude<=max(shapefile_df$lat))
ca_districts <- read_tsv("pubdistricts.txt") %>% filter(County %in% bay_area_counties)
```

```
# https://www.cde.ca.gov/ta/ac/cm/datafiles2019.asp
annual_data <- list()
for (x in list.files("annual_data", full.names=TRUE)) {
  var_name <- strsplit(basename(x), ".", fixed=TRUE)[[1]][1]
  eval(parse(text=paste0("annual_data$", var_name, " <- read_delim('", x, "', delim='",
                         ifelse(grepl("priority", x), "|", "\t"), "')")))
  col_name <- grep("cds", names(annual_data[[var_name]]), ignore.case=TRUE, value=TRUE)[1]
  annual_data[[var_name]] %<>% `[`(.[[col_name]] %in% ca_sch$CDSCode, )
}
```

School stats

```
# https://www.cde.ca.gov/ta/ac/cm/gradindicator19.asp
annual_data$graduation2019 %>%
  filter(studentgroup=="AS", !is.na(currdenom)) %>%
```

```
  select(cds, grad_rate=currnumer/currdenom)
```

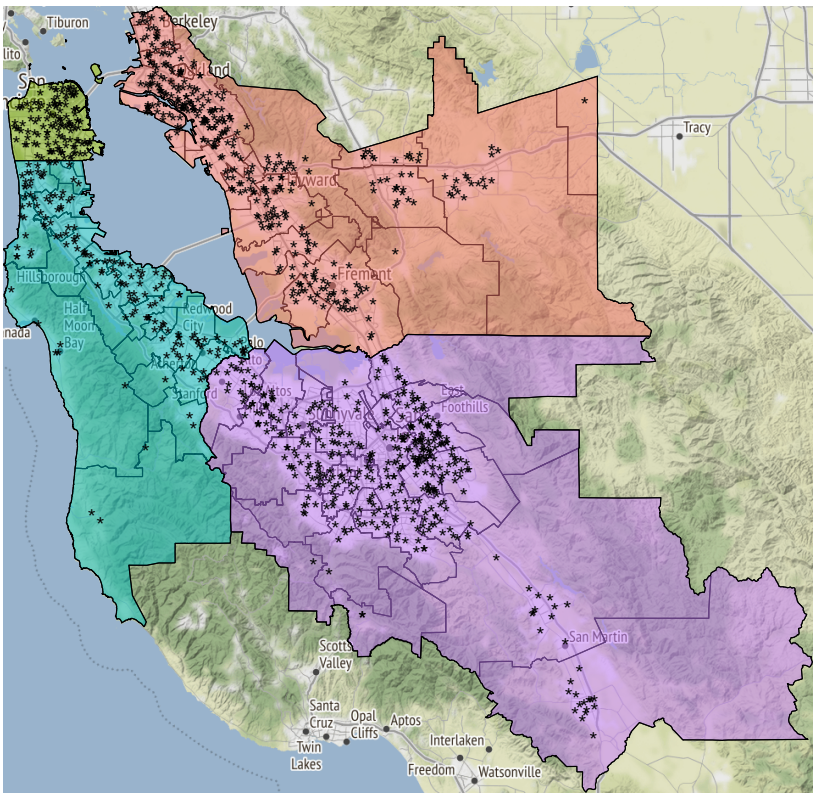
```
## # A tibble: 160 x 2
##    cds             grad_rate
##    <chr>               <dbl>
##  1 01100170112607         NA
##  2 01100170136101         NA
##  3 01611190106401         19
##  4 01611190119222         NA
##  5 01611190130229        173
##  6 01611190130609         NA
##  7 01611190132142         48
##  8 01611270130450         83
##  9 01611430131177         65
## 10 01611500132225        187
## # ... with 150 more rows
```

Maps

```
map <- get_map(location = c(min(shapefile_df$long),
                            min(shapefile_df$lat),
                            max(shapefile_df$long),
                            max(shapefile_df$lat)))
```

```
districts_plot <- ggmap(map, extent = "normal", maprange = FALSE) +
  theme_nothing() +
  geom_polygon(data = shapefile_df,
               aes(x=long, y=lat, group=group),
               color='black', size = .2, alpha=0) +
  geom_polygon(data = shapefile_counties_df,
                aes(x=long, y=lat, group=group, fill=id),
                color='black', size = .2, alpha=.5)
schools_plot <- districts_plot +
  geom_point(data = ca_sch,
             aes(x=Longitude, y=Latitude),
             size=2, shape='*')

schools_plot
```


Tiburon
Tracy
Fremont
Hillsborough
Redwood City
Half Moon Bay
Foothills
San Martin
Scotts Valley
Santa Cruz
Opal Cliffs
Aptos
Twin Lakes
Interlaken
Freedom
Watsonville