A look at electricity usage in California by county.

Data from California Energy Commission & Department of Finance

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(patchwork)
```

```
## Warning: package 'patchwork' was built under R version 4.3.3
```

```r
library(maps)
```

```
## Warning: package 'maps' was built under R version 4.3.3
```

```
##
## Attaching package: 'maps'
##
## The following object is masked from 'package:purrr':
##
##     map
```

```r
library(mapproj)
```

```
## Warning: package 'mapproj' was built under R version 4.3.3
```

```r
CA_electricity_county <- read_csv("ElectricityByCounty.csv")
```

```
## Rows: 174 Columns: 36
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## chr  (2): County, Sector
## dbl (34): 2022, 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, ...
##
## 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.
```

```
CA_2020_pop <- read_csv("CA_county_pop_2020.csv")
```

```
## Rows: 58 Columns: 2
## -- Column specification ------------------------------------------------
## Delimiter: ","
## chr (1): County
## dbl (1): Population
##
## 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.
```

```
CA_electricity_county_pivot <- CA_electricity_county |>
  mutate("Total Usage" = NULL) |>
  pivot_longer(
    cols = c("1990":"2022"),
    names_to = "year",
    values_to = "GWh"
  ) |>
  transform(year = as.numeric(year))

number_of_counties = length(unique(CA_electricity_county_pivot$County))
```

Importing data

   The CEC (California Energy Commission) makes available county level electricity usage in the state in .csv form. Each county is broken up by three yearly observations. Commercial and non-commercial usage of electricity are recorded separately, the last observation for every year is a summation of commercial and non-commercial use. The CEC also provides a "total usage" by county variable which will be dropped for this analyses as it can be easily derived and can create data tidying issues down the line.

   The Information provided by the CEC also needs to be pivoted as it records yearly sums in a horizontal fashion which does not work well with R. By pivoting the data a separate row can be created for each county by year and usage. This makes the data easier for R to understand. Lastly the year variable was not recognized as a double so specifying to R that it is a numeric value instead of a character is beneficial.

   After applying these transformations we get a list of every county by year and for each year there are now three observations which correspond to commercial, non-commercial and, total use.

Preliminary Findings

   The data set includes observations for 58 counties which is the number of counties in the state. It also only records consumption data for the past 22 years.

```
if (require("maps")) {
  ca_county <- map_data("county", "california")
  ca_map <- ggplot(ca_county, aes(x = long, y = lat, group = group)) +
    geom_polygon(fill = "beige", colour = "lightgreen")

  ca_map + coord_map() +
    labs(title = "California Map",
         subtitle = "with county lines.",
         x = "Longitude",
         y = "Latitude") +
```
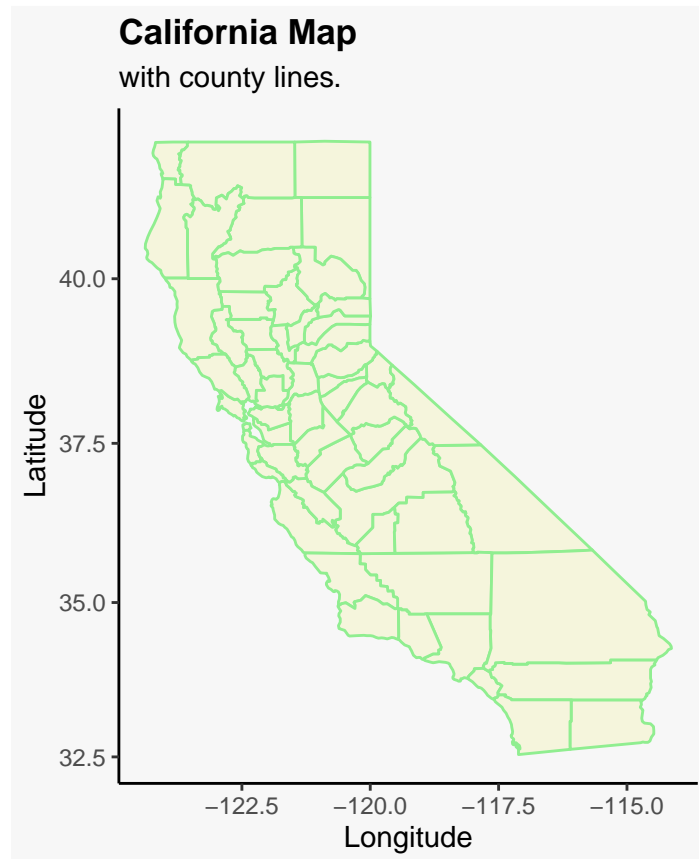
```
    theme_classic() +
    theme(plot.title = element_text(face = "bold")) +
    theme(panel.background = element_rect(fill = "grey97")) +
    theme(plot.background = element_rect("grey97")) +
    theme(legend.background = element_rect("grey99"))
}
```



```
county_filter_norcal <-
  c("SAN FRANCISCO",
    "SANTA CLARA",
    "ALAMEDA",
    "SAN MATEO")
p1 <- CA_electricity_county_pivot |>
  filter(Sector == "Total") |>
  filter(County %in% county_filter_norcal) |>
  ggplot(aes(year, GWh, color = County)) +
  geom_line(linewidth = 1.4) +
  labs(
    title = "Norcal Select County Electricity Consumption",
    subtitle = "in GWh combined residential & commercial",
    x = "",
    y = ""
  ) +
  scale_color_viridis_d() +
  theme_classic() +
```
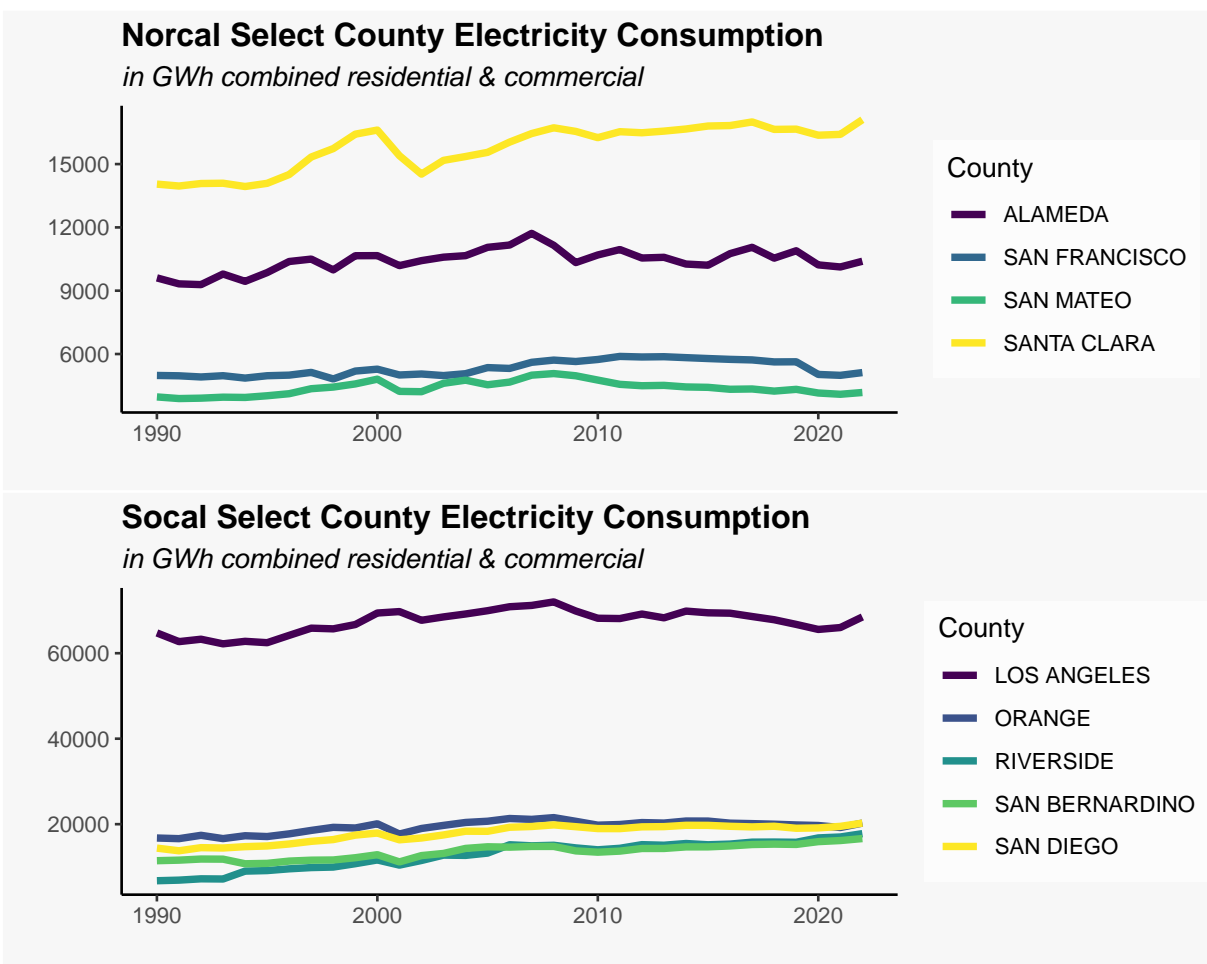
```r
    theme(plot.title = element_text(face = "bold")) +
    theme(plot.subtitle = element_text(face = "italic")) +
    theme(panel.background = element_rect(fill = "grey97")) +
    theme(plot.background = element_rect("grey97")) +
    theme(legend.background = element_rect("grey99"))

county_filter_socal <-
  c("LOS ANGELES",
    "SAN BERNARDINO",
    "ORANGE",
    "SAN DIEGO",
    "RIVERSIDE")
p2 <- CA_electricity_county_pivot |>
  filter(Sector == "Total") |>
  filter(County %in% county_filter_socal) |>
  ggplot(aes(year, GWh, color = County)) +
  geom_line(linewidth = 1.4) +
  labs(
    title = "Socal Select County Electricity Consumption",
    subtitle = "in GWh combined residential & commercial",
    x = "",
    y = ""
  ) +
  scale_color_viridis_d() +
  theme_classic() +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "italic")) +
  theme(panel.background = element_rect(fill = "grey97")) +
  theme(plot.background = element_rect("grey97")) +
  theme(legend.background = element_rect("grey99"))

wrap_plots(p1 / p2) +
  plot_annotation(title = "Electricity Usage between the North and South",
                  caption = "Source: ElectricityByCounty.csv, CEC") +
  theme(plot.title = element_text(face = "bold")) +
  plot_layout()
```

Electricity Usage between the North and South

Norcal Select County Electricity Consumption

in GWh combined residential & commercial

County
— ALAMEDA
— SAN FRANCISCO
— SAN MATEO
— SANTA CLARA

Socal Select County Electricity Consumption

in GWh combined residential & commercial

County
— LOS ANGELES
— ORANGE
— RIVERSIDE
— SAN BERNARDINO
— SAN DIEGO

Source: ElectricityByCounty.csv, CEC

Looking at the Bay Area & Southern California

Looking at electricity consumption in Northern California, Santa Clara is the largest user of electricity followed by Alameda county. These are the two most populous county in Northern California so electricity consumption seems to be linked to the population. Correspondingly San Francisco county with a small geographical size and population use a smaller amount of energy.

Alameda and Santa Clara counties are comparable in in population to Riverside in southern California. Interestingly even with climate changes and increased electricity usage to pump water to the south, Riverside county still uses a comparable amount of electricity to a similar sized county in the north.

Of all the counties in the state, Los Angeles county is by bar and way the largest consumer of electricity. It is also the most populous county in the state with nearly 9 million people. This is three times larger than the next largest county San Diego. As a result Los Angeles uses over three times the amount of energy. While this may seem alarming the per capita usage of electricity by county will be calculated for 2020 which will level out outliers like Los Angeles County.

```
CA_2020_pop_cleaned <- CA_2020_pop |>
  mutate(County = str_to_upper(County))
```

```r
CA_electricity_county_pivot_pop <- CA_electricity_county_pivot |>
  filter(year == 2020) |>
  filter(Sector == "Total") |>
  full_join(CA_2020_pop_cleaned, by = "County") |>
  mutate(per_capita_use_MWh = (GWh / Population) * 1000) |>
  arrange(desc(per_capita_use_MWh))


ggplot(CA_electricity_county_pivot_pop,
       aes(
         per_capita_use_MWh,
         reorder(County, per_capita_use_MWh),
         color = County
       )) +
  geom_vline(aes(xintercept = mean(per_capita_use_MWh)),
             color = "grey80",
             linewidth = 0.8
             ) +
  geom_point() +
  labs(
    title = "Per Capita County Electricity Usage in 2020",
    subtitle = "in MWh combined commercial & non-commercial use.",
    x = "MWh /  person",
    y = "",
    caption = "Source: ElectricityByCounty.csv, CEC
               E-2_Report_July-2010-2021_w.xlsx, CA DOF",
  ) +
  scale_color_viridis_d() +
  theme_classic() +
  theme(legend.position = "none") +
  theme(plot.title = element_text(face = "bold")) +
  theme(plot.subtitle = element_text(face = "italic")) +
  theme(panel.background = element_rect(fill = "grey97")) +
  theme(plot.background = element_rect("grey97")) +
  theme(legend.background = element_rect("grey99"))
```

Per Capita County Electricity Usage in 2020

in MWh combined commercial & non-commercial use.

Source: ElectricityByCounty.csv, CEC
E−2_Report_July−2010−2021_w.xlsx, CA DOF
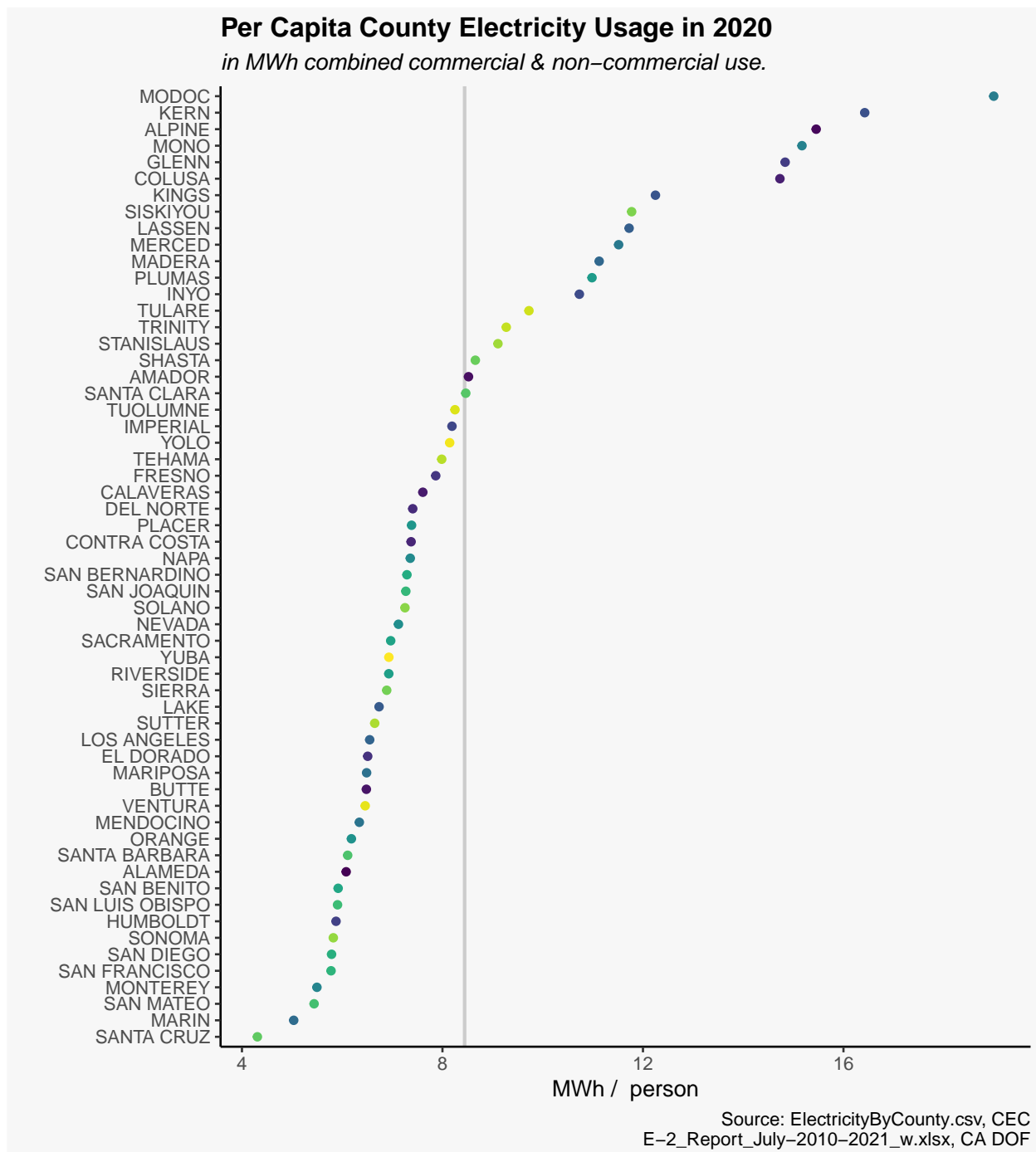
Per capita electricity use in 2020

To reduce the effects of outlier counties like Los Angeles instead of looking at energy consumption alone, it can be divided by the population of the county as seen earlier, population is a large determinant in a counties energy consumption. Accounting for the effects of population would show more information about the states energy consumption.

The state of California Department of Finance provides population estimates for the state on a yearly basis, the data presented in E-2 gives an estimation for every counties population from 2010 to 2020. For this

analyses, energy consumption in the year 2020 will be combined with population data from the same year to create a per capita MWH (Megawatt hour) consumption figure for each county. The mean consumption across all counties is highlighted as a gray vertical bar..

Graphing the per capita MWh consumption, outliers like Los Angeles disappear. When accounting for the population size of Los Angeles, a person living in Los Angeles county uses slightly less than the state average per capita consumption. This is a surprising finding as Los Angeles extensive network of aqueducts and reservoirs require vast amount of energy as much of the water in the region is imported. However the energy consumption of the states large scale water infrastructure might be more equitably shared amongst all counties than previously anticipated.

Another finding of per capita consumption is the out sized consumption that many smaller counties in the north of the state have. Modoc County only has a population of ~8700 people and total electricity usage is the fifth lowest in the state however the per capita use shows that their is a higher energy intensity per person in Modoc. The second highest per capita user of electricity in the state is Kern County. Kern is in the south and has a sizable population nearing 1 million in 2020. More research will have to be done to figure out why per capita usage is so high relative to other counties around Kern.

Data Sources:

County Electricity Data 1990 - 2022:

http://www.ecdms.energy.ca.gov/elecbycounty.aspx

E-2 California County Population Estimates and Components of Change by Year:

https://dof.ca.gov/forecasting/demographics/e-2-california-county-population-estimates-and-components-of-change-by-year/