Data Mining Project - Visualization Code

Configuring Data

```r
#Consumption Data
energy_consump = read_xlsx("energy_consumption_by_state.xlsx", sheet = 2, skip = 2)
any(is.na(energy_consump)) #checks for missing values
```

```
## [1] FALSE
```

```r
ec = energy_consump %>% pivot_longer(c('1960':'2019'), names_to = "year",
                                     values_to = "consumption") #turns year headers into a column/varia

#Isolates US totals
us_ec = energy_consump %>%
  pivot_longer(c('1960':'2019'), names_to = "year", values_to = "consumption") %>%
  filter(State == "US")
#Removes us from states data
ec = ec[!(ec$State == "US"),]

#Population Data
state_population = read_csv("total_population.csv")
```

```
## Warning: Missing column names filled in: 'X62' [62], 'X63' [63], 'X64' [64],
## 'X65' [65], 'X66' [66], 'X67' [67], 'X68' [68], 'X69' [69], 'X70' [70],
## 'X71' [71], 'X72' [72], 'X73' [73], 'X74' [74], 'X75' [75], 'X76' [76],
## 'X77' [77], 'X78' [78], 'X79' [79], 'X80' [80], 'X81' [81], 'X82' [82],
## 'X83' [83], 'X84' [84], 'X85' [85], 'X86' [86], 'X87' [87], 'X88' [88],
## 'X89' [89], 'X90' [90], 'X91' [91], 'X92' [92], 'X93' [93], 'X94' [94],
## 'X95' [95], 'X96' [96], 'X97' [97], 'X98' [98], 'X99' [99], 'X100' [100],
## 'X101' [101]
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   State = col_character(),
##   X62 = col_logical(),
##   X63 = col_logical(),
##   X64 = col_logical(),
##   X65 = col_logical(),
##   X66 = col_logical(),
##   X67 = col_logical(),
##   X68 = col_logical(),
##   X69 = col_logical(),
```

```
##   X70 = col_logical(),
##   X71 = col_logical(),
##   X72 = col_logical(),
##   X73 = col_logical(),
##   X74 = col_logical(),
##   X75 = col_logical(),
##   X76 = col_logical(),
##   X77 = col_logical(),
##   X78 = col_logical(),
##   X79 = col_logical(),
##   X80 = col_logical()
##   # ... with 21 more columns
## )


## See spec(...) for full column specifications.
```

```r
state_population = state_population %>% select(1:61)
any(is.na(state_population)) #checks for missing values
```

```
## [1] FALSE
```

```r
spop = state_population %>% pivot_longer(c(`1960`:`2019`), names_to = "year",
                                         values_to = "population")

us_pop = state_population %>%
  pivot_longer(c(`1960`:`2019`), names_to = "year", values_to = "population") %>%
  filter(State == "US")
#Removes us from states data
spop = spop[!(spop$State == "US"),]

#Integrating Energy and population data
key_data = ec %>% inner_join(spop, by = c("State", "year"))
us_key_data = us_ec %>% inner_join(us_pop, by = c("State", "year"))
key_data = mutate(key_data, cons_per_pop = consumption / population)
us_key_data = mutate(us_key_data, cons_per_pop = consumption / population)

#Shows overall consumption trend in US
ggplot(us_key_data) +
  geom_line(mapping = aes(year, consumption, group = 1)) +
  geom_area(mapping = aes(year, consumption, group = 1),
            lwd = 1, col = "black", fill = "red", alpha = .6) +
  scale_x_discrete(breaks = c(1960, 1970, 1980, 1990, 2000, 2010, 2019)) +
  scale_y_continuous(label = scales::comma) +
  ggtitle("US Energy Consumption (billions BTU)") +
  xlab("Year") + ylab("Consumption") +
  theme_classic() +
  theme(plot.title = element_text(hjust = 0.5)) +
  theme(plot.title = element_text(size=9),
        axis.title.x = element_text(size=9), axis.title.y = element_text(size=9))  +
  ggsave("us_cons.png", width = 3.5, height = 2.5)
```

US Energy Consumption (billions BTU)
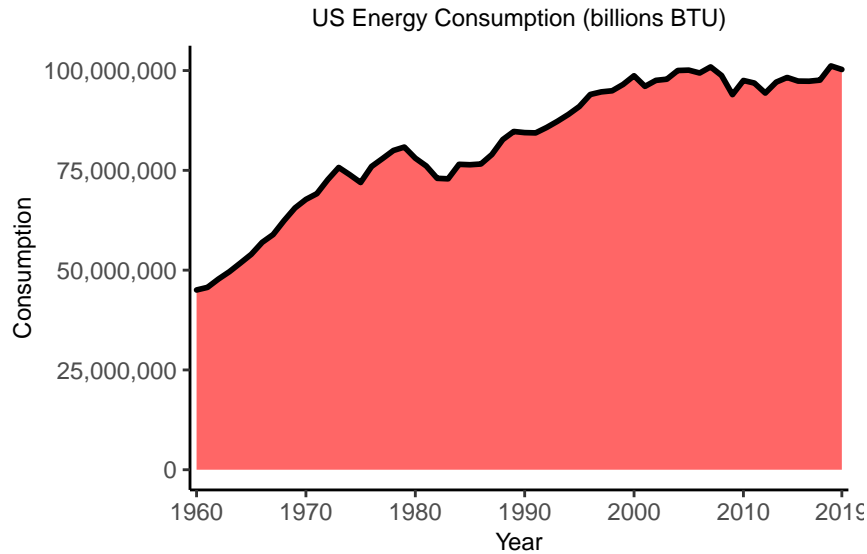
```r
#Regress on population
# pop_fit = lm(consumption~population, us_key_data)
# plot(us_key_data$population, us_key_data$consumption, main = "US Consumption vs Population (w/ Trend
# abline(pop_fit, col = "red")

co2_em = read_xlsx("co2_emissions2.xlsx")
```

```
## New names:
## * `` -> ...1
```

```r
co2_em = co2_em %>% select(-1)
any(is.na(co2_em)) #checks for missing values
```

```
## [1] FALSE
```

```r
co2_em = co2_em %>%
  pivot_longer(c(`2000`:`2018`), names_to = "year", values_to = "co2_em")

mapd = key_data %>% mutate(state = state.name[match(State, state.abb)])
mapd = mapd %>% rename("st_abbr" = "State")
mapd = mapd %>% rename("State" = "state")
mapd = co2_em %>% inner_join(mapd, by = c("State", "year"))
mapd = mapd %>% rename("state" = "State")

#Climate Data
tavg = read_csv("TAVG.csv") #avg temp
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##    .default = col_double(),
##    State = col_character()
## )
```

```
## See spec(...) for full column specifications.

tavg = tavg %>% select("State", "2007":"2018")
any(is.na(tavg)) #checks for missing values


## [1] FALSE

tavg = tavg %>%
  pivot_longer(c('2007':'2018'), names_to = "year", values_to = "tavg")

tmin = read_csv("TMIN.csv") #avg min temp


## Warning: Missing column names filled in: 'X1' [1]


## Parsed with column specification:
## cols(
##   .default = col_double(),
##   State = col_character()
## )
## See spec(...) for full column specifications.

tmin = tmin %>% select("State", "2007":"2018")
any(is.na(tmin)) #checks for missing values


## [1] FALSE

tmin = tmin %>%
  pivot_longer(c('2007':'2018'), names_to = "year", values_to = "tmin")

prcp = read_csv("PRCP.csv") #precipitaiton


## Warning: Missing column names filled in: 'X1' [1]


## Parsed with column specification:
## cols(
##   .default = col_double(),
##   State = col_character()
## )
## See spec(...) for full column specifications.

prcp = prcp %>% select("State", "2007":"2018")
any(is.na(prcp)) #checks for missing values


## [1] FALSE

prcp = prcp %>%
  pivot_longer(c('2007':'2018'), names_to = "year", values_to = "prcp")

snow = read_csv("SNOW.csv") #snowfall
```

```
## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   State = col_character()
## )
## See spec(...) for full column specifications.
```

```r
snow = snow %>% select("State", "2007":"2018")
any(is.na(snow)) #checks for missing values
```

```
## [1] FALSE
```

```r
snow = snow %>%
  pivot_longer(c('2007':'2018'), names_to = "year", values_to = "snow")

climate_data = tavg %>% inner_join(tmin, by = c("State", "year"))
climate_data = climate_data %>% inner_join(prcp, by = c("State", "year"))
climate_data = climate_data %>% inner_join(snow, by = c("State", "year"))
climate_data = climate_data %>% mutate(state = state.name[match(State, state.abb)])
```

Visualizations

Non-Climate Visuals

```
mapd1 = mapd %>% filter(year == 2018)
plot_usmap(data = mapd1, values = "consumption", color = "red") +
  #scale_fill_continuous(name = "Consumption (2019)", label = scales::comma) +
  theme(legend.position = "right") +
  ggtitle("Energy Consumption in 2018") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  scale_fill_viridis_c(direction = -1, name = "Billions BTU", label = scales::comma) +
  ggsave("2018_ec.png", width = 3.5, height = 2.5)
```


Energy Consumption in 2018
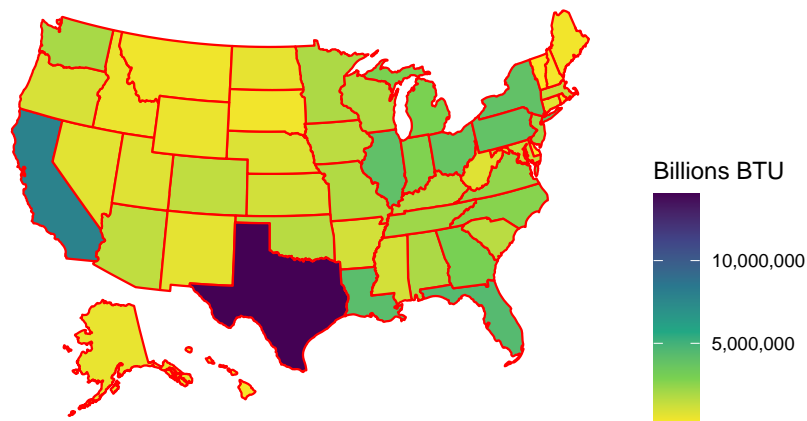
```
# mapd2 = mapd %>% filter(year == 2007)
# plot_usmap(data = mapd2, values = "consumption", color = "red") +
#   theme(legend.position = "right") +
#   ggtitle("Energy Consumption in 2007") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   scale_fill_viridis_c(direction = -1, name = "Billions BTU", label = scales::comma) +
#   ggsave("2007_ec.png", width = 3.5, height = 2.5)

plot_usmap(data = mapd1, values = "population", color = "red") +
  theme(legend.position = "right") +
  ggtitle("Population in 2018") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  scale_fill_viridis_c(direction = -1, name = "Tens of Thousands", label = scales::comma) +
  ggsave("2018_pop.png", width = 3.5, height = 2.5)
```
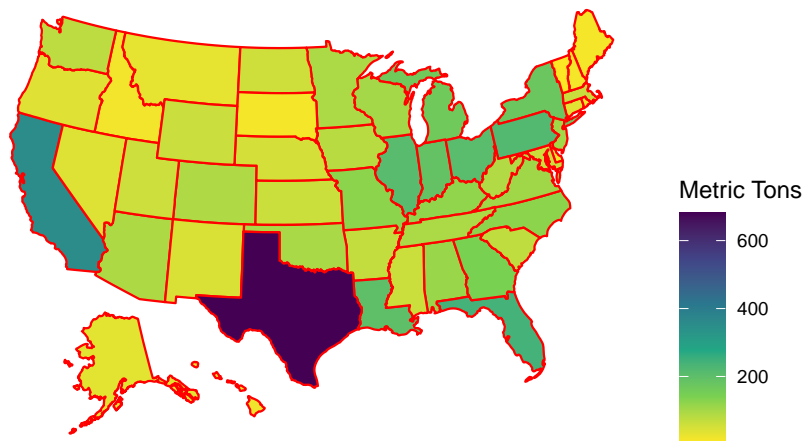
Population in 2018
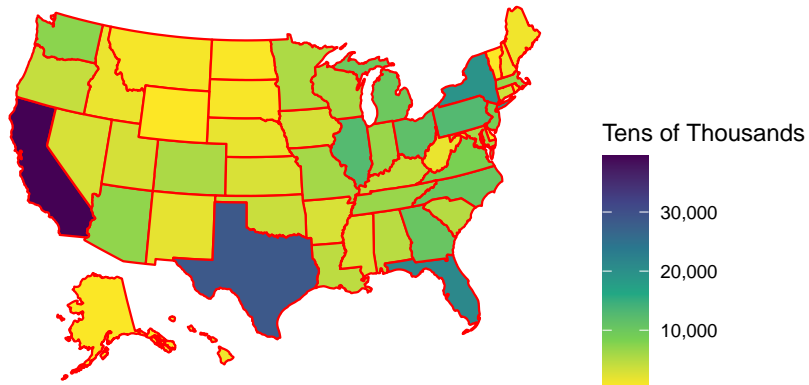

```
# plot_usmap(data = mapd2, values = "population", color = "red") +
#   theme(legend.position = "right") +
#   ggtitle("Population in 2007") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   scale_fill_viridis_c(direction = -1, name = "Tens of Thousands", label = scales::comma) +
#   ggsave("2007_pop.png", width = 3.5, height = 2.5)

plot_usmap(data = mapd1, values = "co2_em", color = "red") +
  theme(legend.position = "right") +
  ggtitle("CO2 Emissions in 2018") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  scale_fill_viridis_c(direction = -1, name = "Metric Tons", label = scales::comma) +
  ggsave("2018_co2.png", width = 3.5, height = 2.5)
```
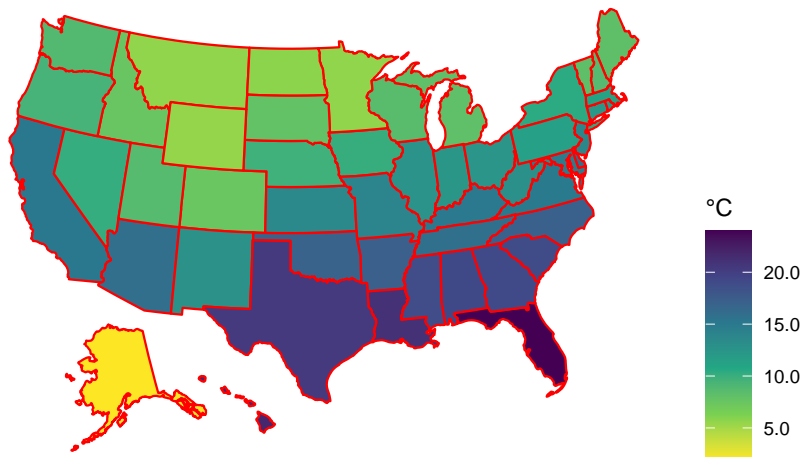
CO2 Emissions in 2018

```
# plot_usmap(data = mapd2, values = "co2_em", color = "red") +
#   theme(legend.position = "right") +
#   ggtitle("CO2 Emissions in 2007") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   scale_fill_viridis_c(direction = -1, name = "Metric Tons", label = scales::comma) +
#   ggsave("2007_co2.png", width = 3.5, height = 2.5)
```

Climate Visuals

Full

```
cd1 = climate_data %>% filter(year == 2018)
plot_usmap(data = cd1, values = "tavg", color = "red") +
  theme(legend.position = "right") +
  ggtitle("Avg Temperature in 2018") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  scale_fill_viridis_c(direction = -1, name = "°C", label = scales::comma) +
  ggsave("2018_temp.png", width = 3.5, height = 2.5)
```


Avg Temperature in 2018

```
# cd2 = climate_data %>% filter(year == 2007)
# plot_usmap(data = cd2, values = "tavg", color = "red") +
#   theme(legend.position = "right") +
#   ggtitle("Avg Temperature in 2007") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   scale_fill_viridis_c(direction = -1, name = "°C", label = scales::comma) +
#   ggsave("2007_temp.png", width = 3.5, height = 2.5)

plot_usmap(data = cd1, values = "prcp", color = "red") +
  theme(legend.position = "right") +
  ggtitle("Precipitation in 2018") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
```

```
    scale_fill_viridis_c(direction = -1, name = "Inches", label = scales::comma) +
    ggsave("2018_prcp.png", width = 3.5, height = 2.5)
```

Precipitation in 2018
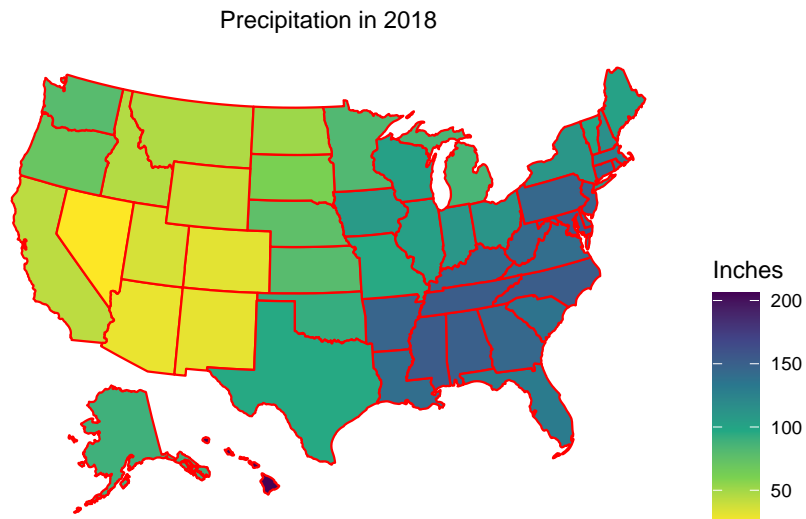

```
# plot_usmap(data = cd2, values = "prcp", color = "red") +
#   theme(legend.position = "right") +
#   ggtitle("Precipitation in 2007") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   scale_fill_viridis_c(direction = -1, name = "Inches", label = scales::comma) +
#   ggsave("2007_prcp.png", width = 3.5, height = 2.5)
```

Regional

```
# Midwest
plot_usmap(data = mapd1, values = "consumption", color = "red", include = .midwest_region) +
  scale_fill_continuous(name = "Billions BTU", label = scales::comma, trans = "reverse") +
  theme(legend.position = "right") +
  guides(fill = guide_colorbar(reverse = TRUE)) +
  ggtitle("Energy Consumption in Midwest (2018)") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  ggsave("2018_mw_ec.png", width = 3.5, height = 2.5)
```
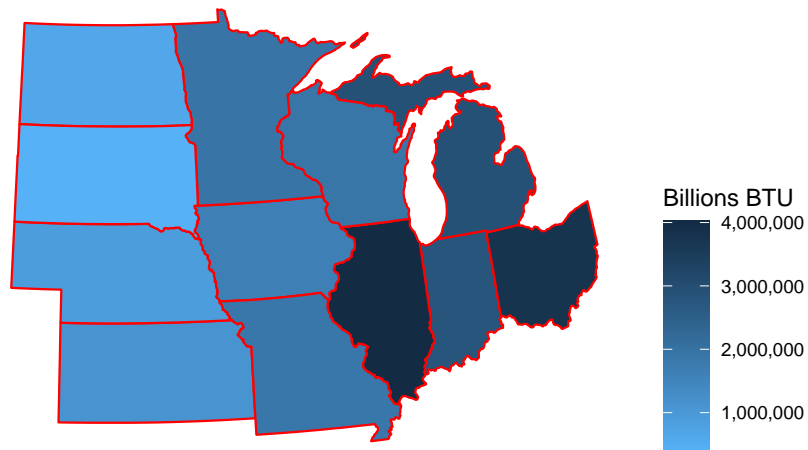
Energy Consumption in Midwest (2018)

Billions BTU

4,000,000

3,000,000

2,000,000

1,000,000

```
# plot_usmap(data = mapd2, values = "consumption", color = "red", include = .midwest_region) +
#   scale_fill_continuous(name = "Billions BTU", label = scales::comma, trans = "reverse") +
#   theme(legend.position = "right") +
#   guides(fill = guide_colorbar(reverse = TRUE)) +
#   ggtitle("Energy Consumption in Midwest (2007)") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   ggsave("2007_mw_ec.png", width = 3.5, height = 2.5)

plot_usmap(data = cd1, values = "snow", color = "red", include = .midwest_region) +
  scale_fill_continuous(name = "Inches", label = scales::comma, trans = "reverse") +
  theme(legend.position = "right") +
  guides(fill = guide_colorbar(reverse = TRUE)) +
  ggtitle("Snowfall in Midwest (2018)") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  ggsave("2018_mw_snow.png", width = 3.5, height = 2.5)
```
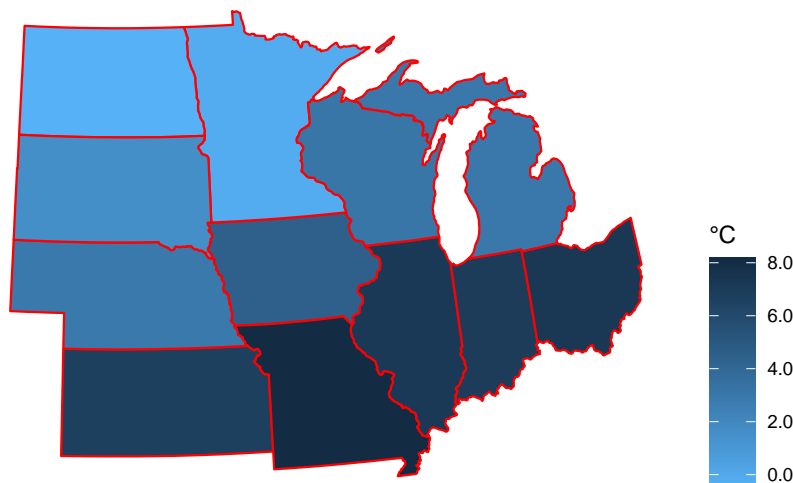
Snowfall in Midwest (2018)

Inches

200

150

100

50
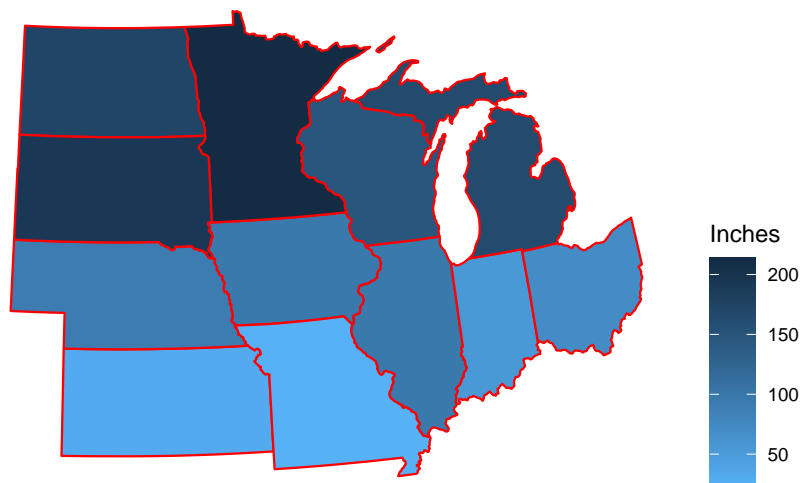
```
# plot_usmap(data = cd2, values = "snow", color = "red", include = .midwest_region) +
#   scale_fill_continuous(name = "Inches", label = scales::comma, trans = "reverse") +
#   theme(legend.position = "right") +
#   guides(fill = guide_colorbar(reverse = TRUE)) +
#   ggtitle("Snowfall in Midwest (2007)") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   ggsave("2007_mw_snow.png", width = 3.5, height = 2.5)

plot_usmap(data = cd1, values = "tmin", color = "red", include = .midwest_region) +
  scale_fill_continuous(name = "°C", label = scales::comma, trans = "reverse") +
  theme(legend.position = "right") +
  guides(fill = guide_colorbar(reverse = TRUE)) +
  ggtitle("Avg Daily Min Temperature in Midwest (2018)") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  ggsave("2018_mw_tmin.png", width = 3.5, height = 2.5)
```

Avg Daily Min Temperature in Midwest (2018)


```
# plot_usmap(data = cd2, values = "tmin", color = "red", include = .midwest_region) +
#   scale_fill_continuous(name = "°C", label = scales::comma, trans = "reverse") +
#   theme(legend.position = "right") +
#   guides(fill = guide_colorbar(reverse = TRUE)) +
#   ggtitle("Avg Daily Min Temperature in Midwest (2007)") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   ggsave("2007_mw_tmin.png", width = 3.5, height = 2.5)

# Southwest
plot_usmap(data = mapd1, values = "consumption", color = "red", include = c("Arizona", "New Mexico", "Te
  scale_fill_continuous(name = "Billions BTU", label = scales::comma, trans = "reverse") +
  theme(legend.position = "right") +
  guides(fill = guide_colorbar(reverse = TRUE)) +
  ggtitle("Energy Consumption in Southwest (2018)") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  ggsave("2018_sw_ec.png", width = 3.5, height = 2.5)
```

Energy Consumption in Southwest (2018)

Billions BTU

10,000,000

5,000,000
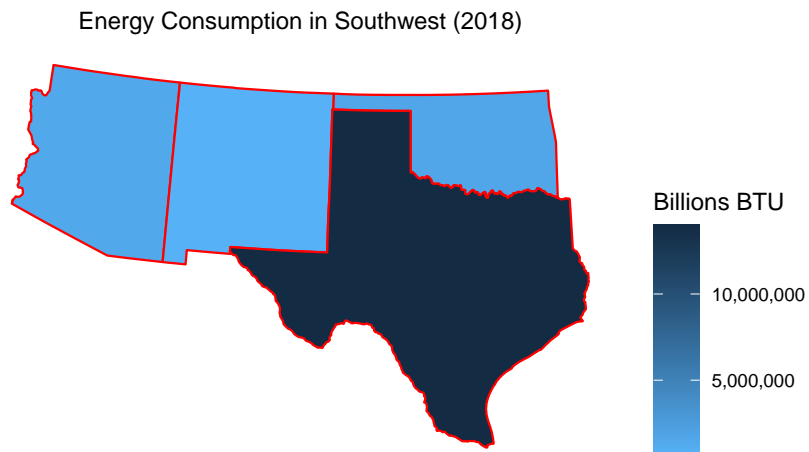
```
# plot_usmap(data = mapd2, values = "consumption", color = "red", include = c("Arizona", "New Mexico",
#   scale_fill_continuous(name = "Billions BTU", label = scales::comma, trans = "reverse") +
#   theme(legend.position = "right") +
#   guides(fill = guide_colorbar(reverse = TRUE)) +
#   ggtitle("Energy Consumption in Southwest (2007)") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   ggsave("2007_sw_ec.png", width = 3.5, height = 2.5)


plot_usmap(data = cd1, values = "tavg", color = "red", include = c("Arizona", "New Mexico", "Texas", "Ok
  scale_fill_continuous(name = "°C", label = scales::comma, trans = "reverse")+
  theme(legend.position = "right") +
  guides(fill = guide_colorbar(reverse = TRUE)) +
  ggtitle("Avg Temperature in Southwest (2018)") +
  theme(plot.title = element_text(hjust = 0.5, size = 9)) +
  ggsave("2018_sw_tavg.png", width = 3.5, height = 2.5)
```

Avg Temperature in Southwest (2018)
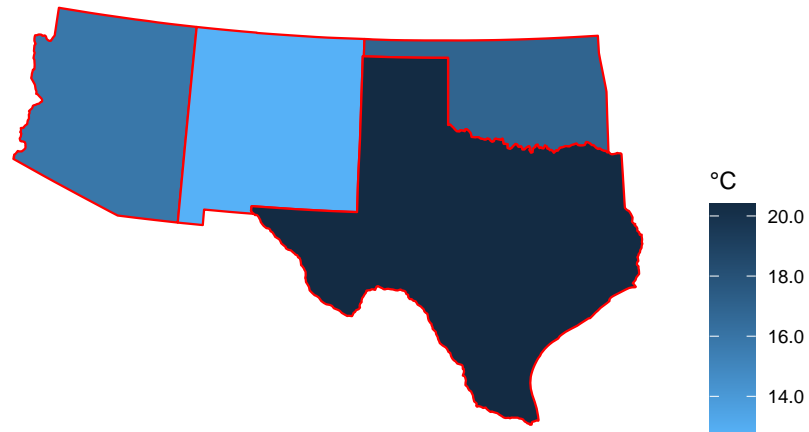

```
# plot_usmap(data = cd2, values = "tavg", color = "red", include = c("Arizona", "New Mexico", "Texas",
#   scale_fill_continuous(name = "°C", label = scales::comma, trans = "reverse") +
#   theme(legend.position = "right") +
#   guides(fill = guide_colorbar(reverse = TRUE)) +
#   ggtitle("Avg Temperature in Southwest (2007)") +
#   theme(plot.title = element_text(hjust = 0.5, size = 9)) +
#   ggsave("2007_sw_tavg.png", width = 3.5, height = 2.5)
```