EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.2     ✔ tibble    3.3.0
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.1.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to
become errors
```

```
library(leaflet)
library(sf)
```

```
Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

Part 1: Defining Research Question

Chosen Question: Are electric vehicles in high-EV states actually charging with clean energy, or are they running on coal?

This question directly addresses whether EVs are truly environmentally friendly by examining the energy mix used to power them. While EVs produce zero direct emissions, if they're charged using electricity generated primarily from coal - the dirtiest fossil fuel - their environmental benefit is significantly reduced. By analyzing the relationship between EV registrations and coal usage across states, we can identify which states have high EV adoption paired with low coal dependency (truly green EVs) versus states where high EV numbers coincide with high coal usage (potentially undermining the environmental benefits). This analysis directly answers the project's core question about whether the electricity used to charge EVs actually comes from clean sources.

Part 2: Data Preparation and Cleaning

```r
renew_2023 <- read_csv("data/renew-use-2023.csv")
```

```
Rows: 260 Columns: 3
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): State, Energy_Source, Renewable_Use_2023

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.
```

```r
total_2023 <- read_csv("data/total-use-2023.csv")
```

```
Rows: 5 Columns: 53
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr  (1): Energy_Source
dbl (52): AK, AL, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, ID, IL, IN, KS...

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.
```

```r
ev_reg <- read_csv("data/ev-registrations-by-state-2023.csv")
```

```
New names:
Rows: 54 Columns: 2
── Column specification
──────────────────────────────────────────────────────── Delimiter: "," chr
(2): electric vehicle registrations_by_state (2023), ...2
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.
• `` -> `...2`
```

```r
renew_clean <- renew_2023 |>
  rename_with(tolower) |>
  mutate(state = str_to_upper(state)) |>
  mutate(renewable_use_2023 = str_remove_all(renewable_use_2023, " kWh| MWh")) |>
  mutate(renewable_use_2023 = as.numeric(renewable_use_2023)) |>
  filter(state != "US")

total_clean <- total_2023 |>
  pivot_longer(cols = -Energy_Source,
               names_to = "state",
               values_to = "usage") |>
  mutate(state = str_to_upper(state)) |>
  mutate(energy_source = str_to_lower(Energy_Source)) |>
  select(state, energy_source, usage) |>
  filter(state != "US")

ev_clean <- ev_reg |>
```

```
  slice(-1:-2) |>
  rename(state = 1, ev_count = 2) |>
  mutate(ev_count = str_remove_all(ev_count, "#|~|EVs| ")) |>
  mutate(ev_count = as.numeric(ev_count)) |>
  filter(state != "TOTAL") |>
  mutate(state = state.abb[match(state, state.name)]) |>
  mutate(state = ifelse(is.na(state) & grepl("Columbia", state, ignore.case = TRUE),
                        "DC", state))


head(renew_clean)
```

```
# A tibble: 6 × 3
  state energy_source renewable_use_2023
  <chr> <chr>                      <dbl>
1 AK    Biomass                     3404
2 AK    Geothermal                   186
3 AK    Hydropower                  6051
4 AK    Solar Energy                  67
5 AK    Wind Energy                  380
6 AL    Biomass                   189040
```

```
head(total_clean)
```

```
# A tibble: 6 × 3
  state energy_source  usage
  <chr> <chr>          <dbl>
1 AK    coal_usage     18414
2 AL    coal_usage    224926
3 AR    coal_usage    180262
4 AZ    coal_usage    137885
5 CA    coal_usage     28746
6 CO    coal_usage    204826
```

```
head(ev_clean)
```

```
# A tibble: 6 × 2
  state ev_count
  <chr>    <dbl>
1 AL       13047
2 AK        2697
3 AZ       89798
4 AR        7108
5 CA     1256646
6 CO       90083
```

Part 3: Joining / Pivoting Datasets for Analysis

```r
renewable_total <- renew_clean |>
  group_by(state) |>
  summarise(total_renewable = sum(renewable_use_2023, na.rm = TRUE))

coal_usage <- total_clean |>
  filter(energy_source == "coal_usage") |>
  select(state, coal_usage = usage)

total_energy <- total_clean |>
  group_by(state) |>
  summarise(total_energy_use = sum(usage, na.rm = TRUE))

analysis_data <- renewable_total |>
  left_join(coal_usage, by = "state") |>
  left_join(total_energy, by = "state") |>
  left_join(ev_clean, by = "state") |>
  mutate(
    renewable_pct = (total_renewable / total_energy_use) * 100,
    coal_pct = (coal_usage / total_energy_use) * 100
  ) |>
  mutate(state_full = tolower(state.name[match(state, state.abb)])) |>
  mutate(state_full = ifelse(state == "DC", "district of columbia", state_full))

head(analysis_data)
```

```
# A tibble: 6 × 8
  state total_renewable coal_usage total_energy_use ev_count renewable_pct
  <chr>           <dbl>      <dbl>            <dbl>    <dbl>         <dbl>
1 AK              10088      18414           746979     2697          1.35
2 AL             222189     224926          2265008    13047          9.81
3 AR              87277     180262          1151062     7108          7.58
4 AZ             108445     137885          1712667    89798          6.33
5 CA            1065179      28746          6429818  1256646         16.6 
6 CO             115062     204826          1359507    90083          8.46
# ℹ 2 more variables: coal_pct <dbl>, state_full <chr>
```

```r
summary(analysis_data[c("coal_pct", "renewable_pct", "ev_count")])
```

```
    coal_pct      renewable_pct       ev_count      
 Min.   : 0.000   Min.   : 1.351   Min.   :    959  
 1st Qu.: 0.488   1st Qu.: 5.715   1st Qu.:   8221  
 Median : 6.556   Median : 7.964   Median :  25833  
 Mean   :10.238   Mean   :10.329   Mean   :  70948  
 3rd Qu.:13.900   3rd Qu.:11.408   3rd Qu.:  71645  
 Max.   :50.157   Max.   :34.844   Max.   :1256646  
                                   NA's   :1        
```

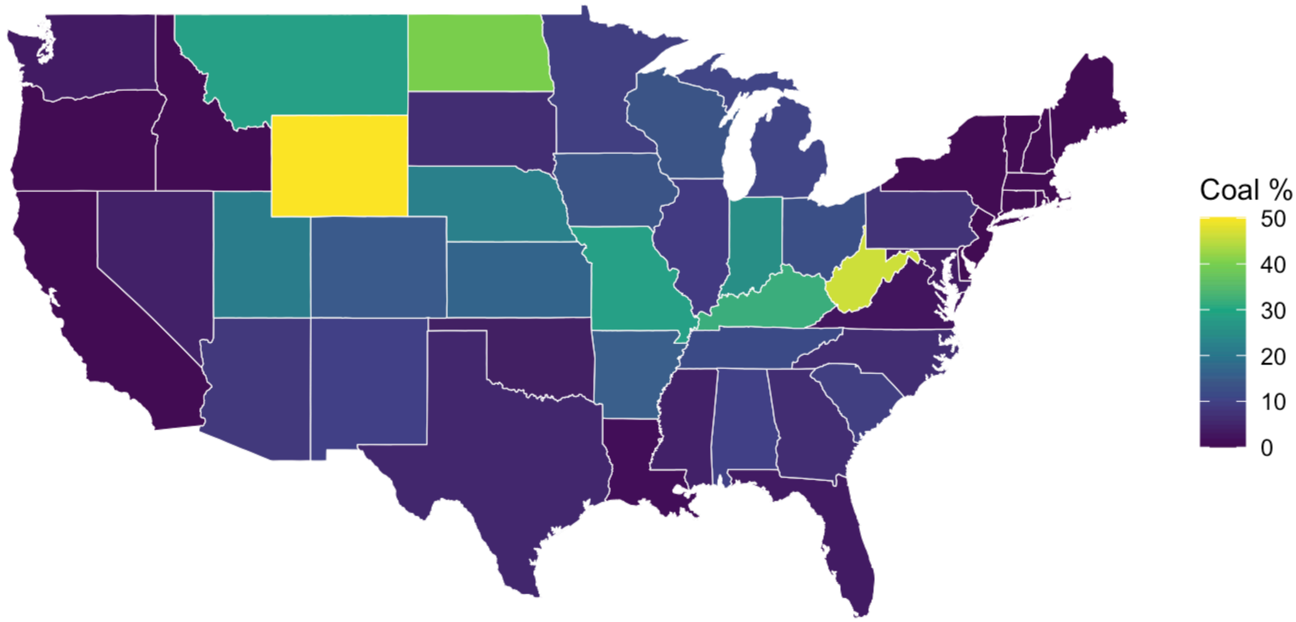
Part 4: Mapping Visualization

```r
us_states <- map_data("state")

map_data <- us_states %>%
  left_join(analysis_data, by = c("region" = "state_full"))

head(map_data)
```

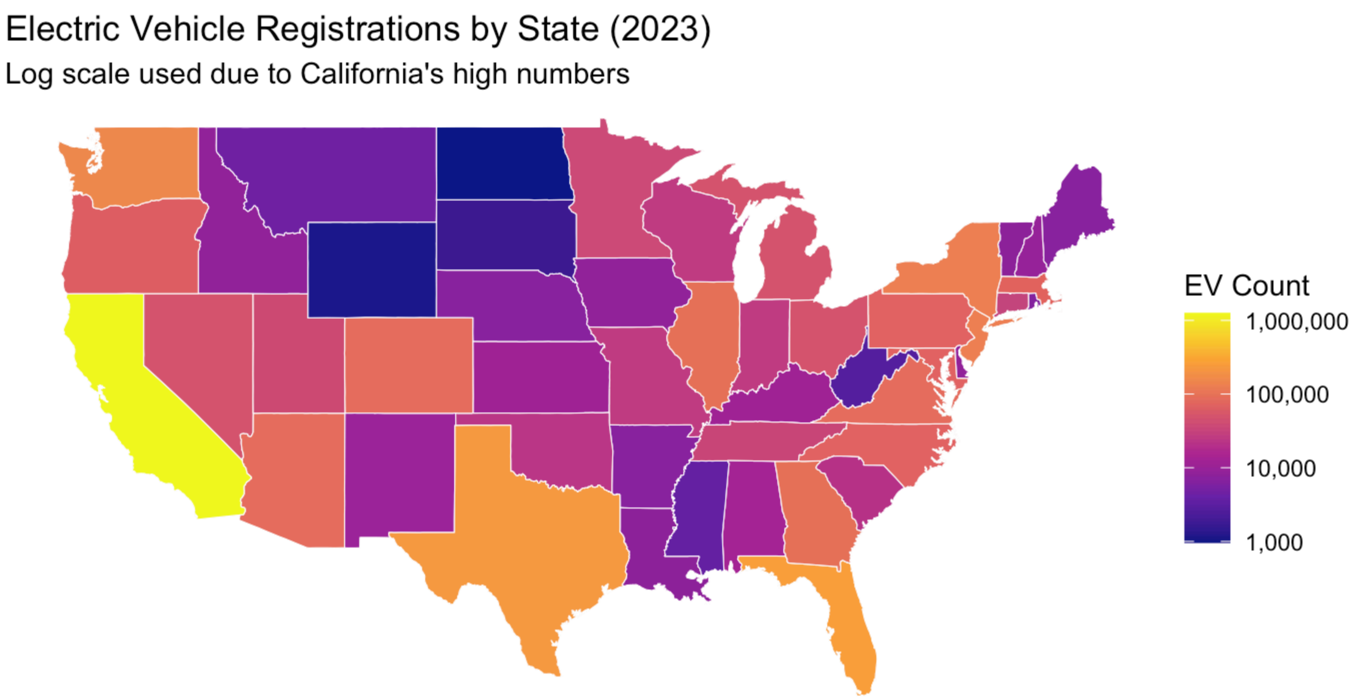
```
       long      lat group order  region subregion state total_renewable
1 -87.46201 30.38968     1     1 alabama      <NA>    AL          222189
2 -87.48493 30.37249     1     2 alabama      <NA>    AL          222189
3 -87.52503 30.37249     1     3 alabama      <NA>    AL          222189
4 -87.53076 30.33239     1     4 alabama      <NA>    AL          222189
5 -87.57087 30.32665     1     5 alabama      <NA>    AL          222189
6 -87.58806 30.32665     1     6 alabama      <NA>    AL          222189
  coal_usage total_energy_use ev_count renewable_pct coal_pct
1     224926          2265008    13047      9.809634 9.930473
2     224926          2265008    13047      9.809634 9.930473
3     224926          2265008    13047      9.809634 9.930473
4     224926          2265008    13047      9.809634 9.930473
5     224926          2265008    13047      9.809634 9.930473
6     224926          2265008    13047      9.809634 9.930473
```

```r
ggplot(map_data, aes(x = long, y = lat, group = group)) +
  geom_polygon(aes(fill = coal_pct), color = "white", linewidth = 0.2) +
  scale_fill_viridis_c(name = "Coal %") +
  coord_fixed(1.3) +
  theme_void() +
  labs(title = "Coal Usage as Percentage of Total Energy by State (2023)")
```

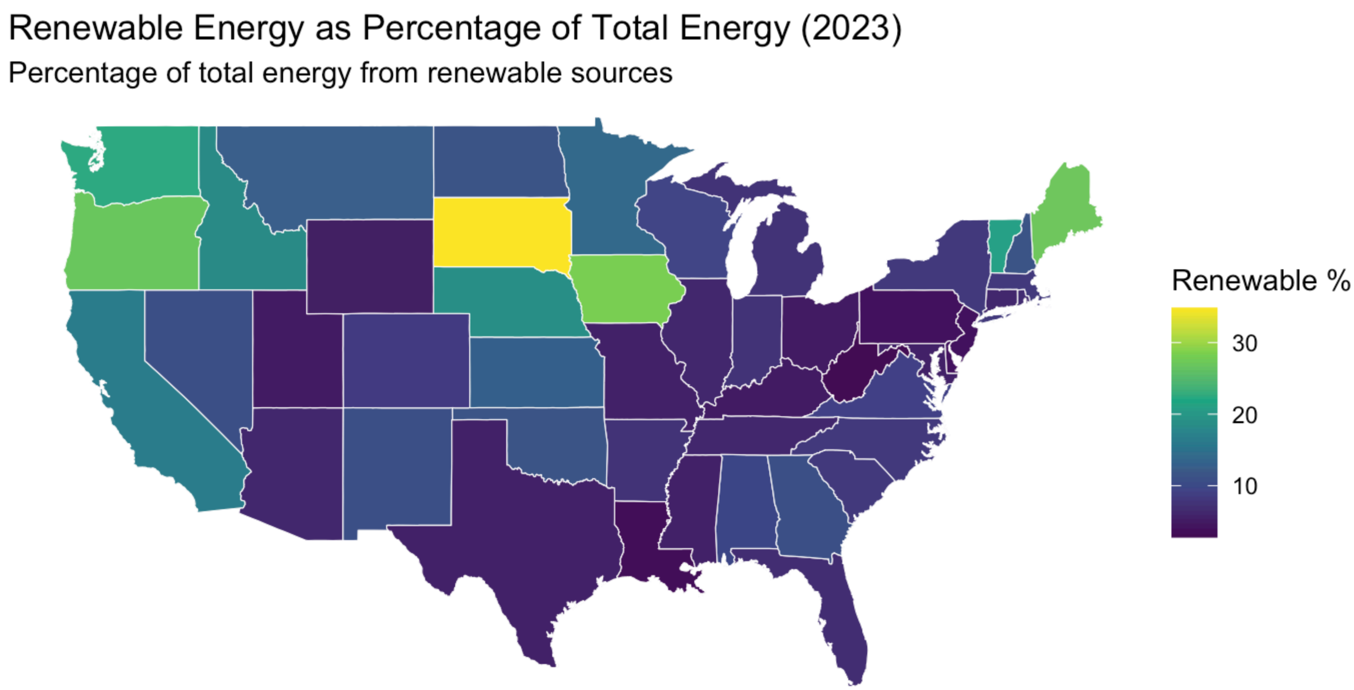
Coal Usage as Percentage of Total Energy by State (2023)




```r
ggplot(map_data, aes(x = long, y = lat, group = group)) +
  geom_polygon(aes(fill = ev_count), color = "white", linewidth = 0.2) +
  scale_fill_viridis_c(
    name = "EV Count",
    option = "plasma",
    trans = "log10",
    labels = scales::comma
  ) +
  coord_fixed(1.3) +
  theme_void() +
  labs(title = "Electric Vehicle Registrations by State (2023)",
       subtitle = "Log scale used due to California's high numbers")
```



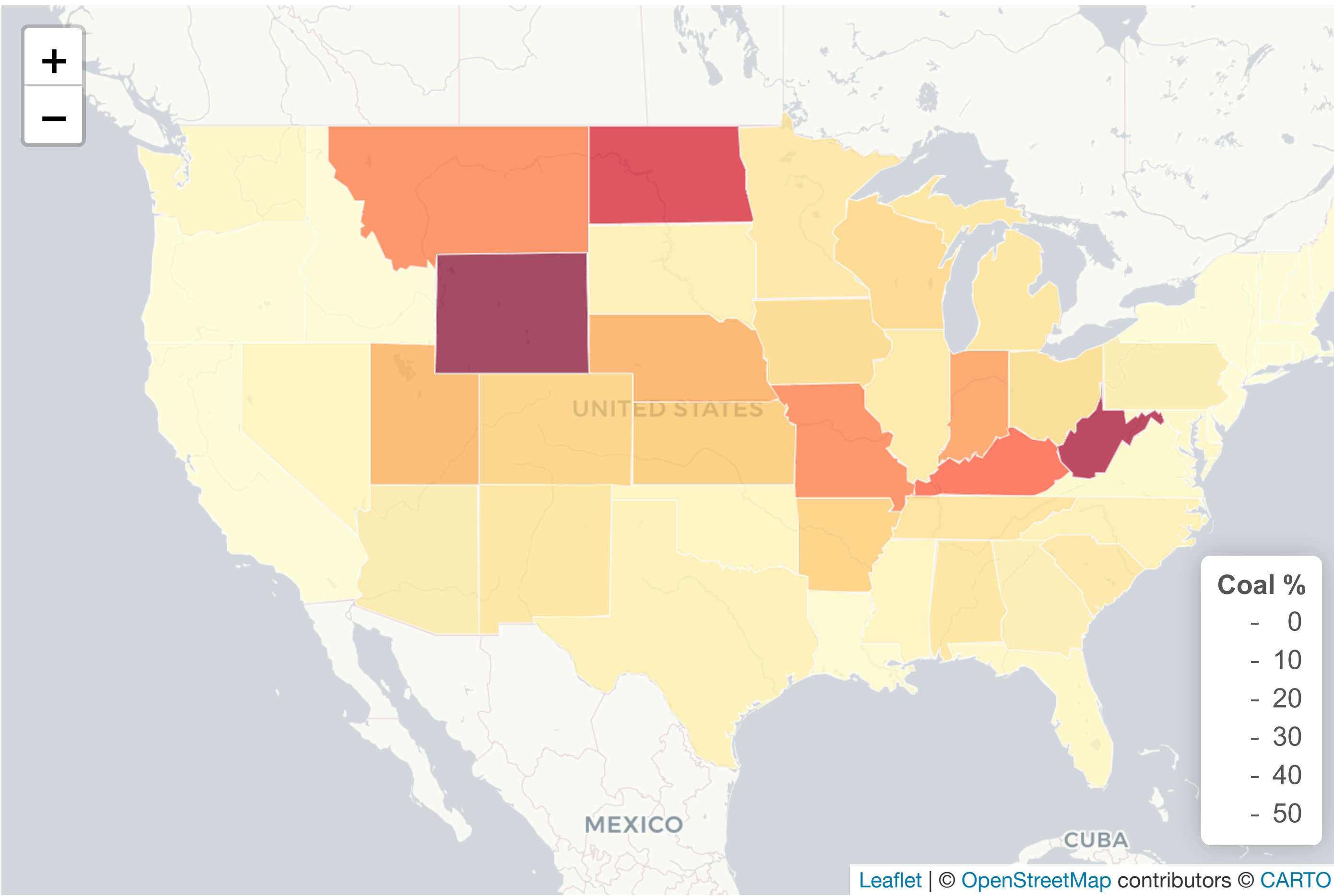

```r
ggplot(map_data, aes(x = long, y = lat, group = group)) +
  geom_polygon(aes(fill = renewable_pct), color = "white", linewidth = 0.2) +
  scale_fill_viridis_c(
    name = "Renewable %",
    option = "viridis"
  ) +
  coord_fixed(1.3) +
  theme_void() +
  labs(title = "Renewable Energy as Percentage of Total Energy (2023)",
       subtitle = "Percentage of total energy from renewable sources")
```




```r
us_states_sf <- st_as_sf(maps::map("state", plot = FALSE, fill = TRUE))

us_states_sf <- us_states_sf |>
  mutate(state_full = str_replace_all(ID, ":main", "")) |>
  left_join(analysis_data, by = "state_full")

leaflet(us_states_sf) |>
  addProviderTiles(providers$CartoDB.Positron) |>
  addPolygons(
    fillColor = ~colorNumeric("YlOrRd", coal_pct)(coal_pct),
    fillOpacity = 0.7,
    color = "white",
    weight = 1,
    label = ~paste0(
      state, "<br/>",
      "Coal: ", round(coal_pct, 1), "%<br/>",
      "EVs: ", scales::comma(ev_count)
    ) |> lapply(htmltools::HTML),
    highlightOptions = highlightOptions(
      weight = 3,
      color = "#666",
      fillOpacity = 0.9
    )
```

```
  ) |>
  addLegend(
    position = "bottomright",
    pal = colorNumeric("YlOrRd", us_states_sf$coal_pct),
    values = ~coal_pct,
    title = "Coal %",
    opacity = 0.7
  )
```

```
Warning: sf layer has inconsistent datum (+proj=longlat +ellps=clrk66 +no_defs).
Need '+proj=longlat +datum=WGS84'
```



Part 5: Analysis and Interpretation

Our analysis reveals important patterns regarding the environmental impact of electric vehicles across different states. By examining the relationship between EV adoption and energy sources, we found significant variation in how "clean" EVs actually are depending on location.

States with High Coal Dependency: West Virginia and Wyoming show coal usage exceeding 50% of total energy, meaning EVs in these states are primarily powered by one of the dirtiest energy sources. This raises questions about whether EV adoption in these regions truly reduces emissions.

Clean EV States: California, despite having the highest EV count (1.2+ million vehicles), maintains relatively low coal usage (~0.4%), suggesting that EVs in this state are indeed contributing to emission reductions. Washington and Oregon also show favorable combinations of high renewable energy and low coal dependency.

The Greenwashing Problem: Several states with moderate EV adoption show surprisingly high coal dependency, indicating that the environmental benefits of EVs are not uniform across the country. This directly answers our main research question: electric vehicles do not always run on clean energy - it depends heavily on the state's energy mix.

Conclusion: Whether an EV is truly "green" depends not just on the vehicle itself, but on where it's being charged. States investing in renewable energy infrastructure are seeing genuine environmental benefits from EV adoption, while others may be overstating their climate impact.