EV Power - Lab 4 Project Report

Example Solution 1

Part 0: libraries

```
library(tidyr)
library(ggplot2)
library(dplyr)
```

```

Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':

    filter, lag
```

```
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
library(readr)
library(readxl)
library(stringr)
```

```
total_use_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.
```

```
ev_registrations_by_state_2023      <-      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`
```

Part 1: Defining Research Question

Chosen Question: Do states with higher renewable energy usage have more EV registrations in 2023??

```
head(total_use_2023)
```

```
# A tibble: 5 × 53
  Energy_Source        AK     AL     AR     AZ     CA     CO     CT    DC     DE
  <chr>             <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl> <dbl>  <dbl>
1 coal_usage        18414 224926 180262 137885 2.87e4 204826      0     0    338
2 NaturalGas       448087 775747 399566 537151 2.15e6 525446 304924 26236  84387
3 petroleum (BTU)  270391 565754 327465 599712 3.00e6 514174 292864 17292 110721
4 nuclear-energy †      0 476392 156492 329474 1.85e5      0 142873     0      0
5 total renewable…  10087 222189  87277 108445 1.07e6 115061  48981  2795   8041
# ℹ 43 more variables: FL <dbl>, GA <dbl>, HI <dbl>, IA <dbl>, ID <dbl>,
#   IL <dbl>, IN <dbl>, KS <dbl>, KY <dbl>, LA <dbl>, MA <dbl>, MD <dbl>,
#   ME <dbl>, MI <dbl>, MN <dbl>, MO <dbl>, MS <dbl>, MT <dbl>, NC <dbl>,
#   ND <dbl>, NE <dbl>, NH <dbl>, NJ <dbl>, NM <dbl>, NV <dbl>, NY <dbl>,
#   OH <dbl>, OK <dbl>, OR <dbl>, PA <dbl>, RI <dbl>, SC <dbl>, SD <dbl>,
#   TN <dbl>, TX <dbl>, UT <dbl>, VA <dbl>, VT <dbl>, WA <dbl>, WI <dbl>,
#   WV <dbl>, WY <dbl>, US <dbl>
```

Part 2: Data Preparation and Cleaning

```
use_2023 = total_use_2023 |> pivot_longer(cols = -1 , names_to = "States",
values_to = "Energy_usage")

use_2023 = use_2023 |>
  group_by(Energy_Source)|>
  filter(Energy_Source == "total renewable-energy")|>
  arrange(desc(Energy_usage))|>
  slice(-1)|>
```

```
  drop_na()|>
  rename(State = 'States')


print(use_2023)
```

```
# A tibble: 51 × 3
# Groups:   Energy_Source [1]
   Energy_Source          State Energy_usage
   <chr>                  <chr>        <dbl>
 1 total renewable-energy CA         1065179
 2 total renewable-energy TX          791211
 3 total renewable-energy IA          414801
 4 total renewable-energy WA          365956
 5 total renewable-energy GA          291462
 6 total renewable-energy FL          286306
 7 total renewable-energy NY          272967
 8 total renewable-energy IL          245703
 9 total renewable-energy OR          236062
10 total renewable-energy MN          223864
# ℹ 41 more rows
```

```

```

```
ev_new_2023= ev_registrations_by_state_2023 |>
  rename(States = "electric vehicle registrations_by_state (2023)",
         Count = "...2"  )|>
  drop_na()

ev_new_2023 = ev_new_2023|>
  slice(-1)|>
  mutate(
   Count = str_remove_all(Count, "[^0-9]"),
   Count = as.numeric(Count)
)
```

```
ev_new_2023_states = ev_new_2023|>
mutate(
    States = str_to_title(str_trim(States)),
    States = state.abb[match(States, state.name)])
```

```
ev_new_2023_states = ev_new_2023_states |> drop_na()|>
  rename(State = States)
```

```
ev_new_2023_states
```

```
# A tibble: 50 × 2
   State   Count
   <chr>   <dbl>
 1 AL      13047
 2 AK       2697
 3 AZ      89798
 4 AR       7108
 5 CA    1256646
 6 CO      90083
 7 CT      31557
 8 DE       8435
 9 FL     254878
10 GA      92368
# ℹ 40 more rows
```

Part 3: Joining / Pivoting Datasets for Analysis

```
combined =ev_new_2023_states|>
  left_join(use_2023, by='State')|>
  rename(Count_of_ev = Count)


combined
```

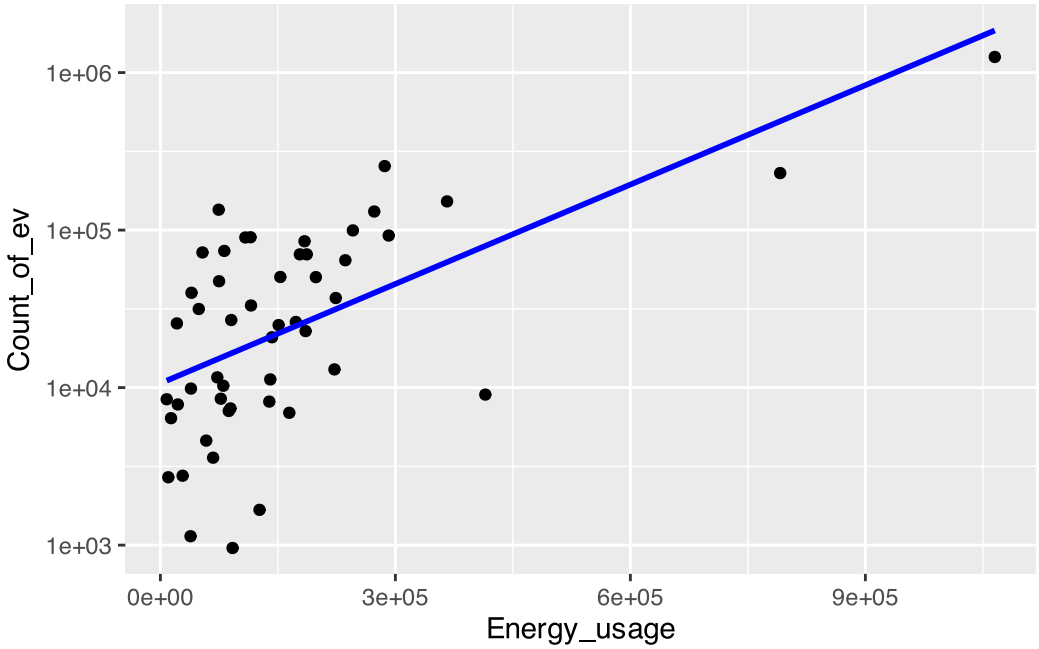
```
# A tibble: 50 × 4
   State Count_of_ev Energy_Source          Energy_usage
   <chr>       <dbl> <chr>                         <dbl>
 1 AL          13047 total renewable-energy       222189
 2 AK           2697 total renewable-energy        10087
 3 AZ          89798 total renewable-energy       108445
 4 AR           7108 total renewable-energy        87277
 5 CA        1256646 total renewable-energy      1065179
 6 CO          90083 total renewable-energy       115061
 7 CT          31557 total renewable-energy        48981
 8 DE           8435 total renewable-energy         8041
 9 FL         254878 total renewable-energy       286306
10 GA          92368 total renewable-energy       291462
# ℹ 40 more rows
```

or(combined$Count_of_ev, combined$Energy_usage, use = “complete.obs”)

Part 4: Mapping Visualization

```
ggplot(combined, aes(x = Energy_usage, y = Count_of_ev)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  scale_x_continuous() +
  scale_y_log10()
```

```
`geom_smooth()` using formula = 'y ~ x'
```




```
  labs(title = "EV Registrations vs Renewable Energy Usage",
       x = "Total Renewable Energy Usage (kWh)",
       y = "EV Registrations")
```

```
<ggplot2::labels> List of 3
 $ x    : chr "Total Renewable Energy Usage (kWh)"
 $ y    : chr "EV Registrations"
 $ title: chr "EV Registrations vs Renewable Energy Usage"
```

```
state_coords <- data.frame(
  State = state.abb,
  lat = state.center$y,
  lng = state.center$x
)
```

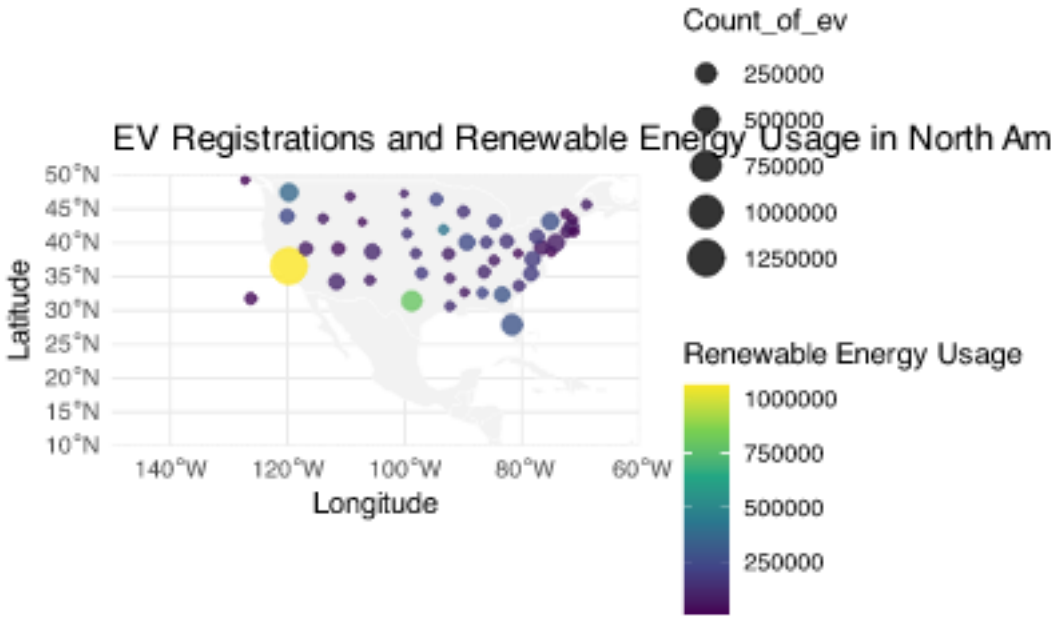
```r
combined_with_coords <- combined |>
  left_join(state_coords, by = c("State" = "State"))

print(combined_with_coords)
```

```
# A tibble: 50 × 6
   State Count_of_ev Energy_Source          Energy_usage   lat    lng
   <chr>       <dbl> <chr>                         <dbl> <dbl>  <dbl>
 1 AL          13047 total renewable-energy       222189  32.6  -86.8
 2 AK           2697 total renewable-energy        10087  49.2 -127.
 3 AZ          89798 total renewable-energy       108445  34.2 -112.
 4 AR           7108 total renewable-energy        87277  34.7  -92.3
 5 CA        1256646 total renewable-energy      1065179  36.5 -120.
 6 CO          90083 total renewable-energy       115061  38.7 -106.
 7 CT          31557 total renewable-energy        48981  41.6  -72.4
 8 DE           8435 total renewable-energy         8041  38.7  -75.0
 9 FL         254878 total renewable-energy       286306  27.9  -81.7
10 GA          92368 total renewable-energy       291462  32.3  -83.4
# ℹ 40 more rows
```

```r
library(rnaturalearth)
map_north_america <- ne_countries(continent = "north america", returnclass = "sf")
```

```r
ggplot() +
  geom_sf(data = map_north_america, fill = "gray95", color = "white") +
  geom_point(
    data = combined_with_coords,
    aes(x = lng, y = lat, size = Count_of_ev, color = Energy_usage),
    alpha = 0.8
  ) +
  scale_color_viridis_c(name = "Renewable Energy Usage") +
  coord_sf(xlim = c(-150, -60), ylim = c(10, 50), expand = FALSE) +
  labs(
   title = "EV Registrations and Renewable Energy Usage in North America (2023)",
    x = "Longitude",
    y = "Latitude"
  ) +
  theme_minimal() +
  theme(panel.background = element_blank())
```


EV Registrations and Renewable Energy Usage in North Am
Latitude
50°N
45°N
40°N
35°N
30°N
25°N
20°N
15°N
10°N
140°W
120°W
100°W
80°W
60°W
Longitude
Count_of_ev
250000
500000
750000
1000000
1250000
Renewable Energy Usage
1000000
750000
500000
250000