EV Power - Lab 4 Project Report

Lucas Zhou

Example Solution 1

Introduction

The importance of protecting the environment never fails to becoming the focus of human beings living in modern age. The mission of seeking ways to improve the environmental condition is a crucial role for researchers in whatever area of study. For us Statistician (or at least Statistic Student), what we can contribute is using our professional knowledge of data manipulation and proposing insightful conclusions based on data.

In particular, this project asks us to look into the energy data of every states in the US from 2021 to 2023. From EDA, we are able to demystify some common misconception and ferret out overlooked features. In this report, after cleaning the dirty and chaotic data sets and joining them together, I will answer mainly three research questions:

1. How has the share of renewable energy changed from 2021–2023 across states?
2. What is the share of electricity that comes from clean sources by state?
3. Are EV registrations concentrated in states with cleaner energy mixes?

Starting this project late until the last day, I spent the whole afternoon finishing the data cleaning, and the whole evening drawing out figures and formulating the report. Clearly, the heavy workload of the projects and problem sheets far exceed my expectations of a 3 unit course, on par with the workload of notoriously 4 unit CS61B, or cosplaying the formidable Stat215a. Even so, it was a great pleasure of doing this project, where I takes a lot. With the assisstance of Gemini AI, I was amazed by the power of R language in drawing out such an incredible map filled with my data. Thanks for Professor Andrew Bray and all the GSIs that gave us this valuable practice opportunity.

Part 0: libraries

Part 1: Defining Research Question

See Introduction.

Part 2: Data Preparation and Cleaning

Before cleaning, we have the raw data as below.

```
  State Energy_Source Renewable_Use_2021
1    AK       Biomass              ≈3153
```

```
2    AK    Geothermal          186 MMBtu
3    AK    Hydropower         5763 about
4    AK  Solar Energy                ~45
5    AK   Wind Energy           451 USD
6    AL       Biomass        198543 est.
```

```
  State Energy_Source Renewable_Use_2022
1    AK       Biomass              ≈3846
2    AK    Geothermal               $186
3    AK    Hydropower               $5846
4    AK  Solar Energy                 ~57
5    AK   Wind Energy               $475
6    AL       Biomass         193932 USD
```

```
  State Energy_Source Renewable_Use_2023
1    AK       Biomass           3404 kWh
2    AK    Geothermal              186.0
3    AK    Hydropower               6051
4    AK  Solar Energy                 67
5    AK   Wind Energy                380
6    AL       Biomass         189040 kWh
```

```
  Total.energy.average.price..dollars.per.million.Btu...
1                                                    ,,,
2                                  State,2021,2022,2023
3                AK,$20.03 per MMBtu,$27.33,$23.84 est.
4                   AL,about 17.85 USD,23.37 USD,≈21.11
5                     AR,$18.42,$23.84 per MMBtu,$21.76
6                       AZ,≈25.07,31.72 USD,about 30.28
```

```
           Energy_Source     AK     AL     AR     AZ      CA     CO     CT
1                   Coal  18694 309791 216123 160299   28244 252442   2880
2           Natural Gas† 395590 739891 360545 484962 2172757 509970 305184
3        Petroleum (BTU) 261094 583042 328271 606862 2959389 497788 284788
4                nuclear      0 480115 141372 329868  171842      0 179551
5 total_renewable_energy   9597 239817  89714  99266  810020 103955  49306
     DC     DE      FL     GA     HI     IA     ID      IL     IN     KS
KY
1     0   4542  200193 203870  12566 264419   3051  522809 753557 219031
548443
2 28336  82708 1591864 773889    133 383424 135176 1088485 869328 291797
365875
3 18439 113641 1748346 922503 223014 408385 188263 1136797 712427 339006
584011
```

```
4     0      0  307811 354085      0      0      0 1011555      0  89426
0
5  2487   7150  297291 289113  20134 389787  74428  224106 157324 135551
71744
       LA     MA     MD     ME     MI     MN     MO     MS     MT     NC
ND
1   95856      0  69186   1588 436203 179055 616413  64446 122765 222501
361811
2 1862349 404301 299282  57233 950364 523812 293633 576903  87105 637553
191168
3 1840835 503312 433791 163991 814081 561731 607276 384328 176686 884299
168682
4  179886      0 156369      0 358114 147286  44766 122771      0 449675
0
5  135905  75370  52732  95141 194075 216113  88879  66134  56334 196973
92653
      NE     NH     NJ     NM     NV      NY      OH     OK     OR      PA
1 216298   3259  12586 133228  35910    5370  575920 131695   1303  485193
2 191008  60116 697019 285809 305212 1359437 1294814 745911 305665 1868137
3 237214 142030 749892 262885 286548 1237451 1028000 517408 317322 1047658
4  71758 102789 293494      0      0  325141  182330      0      0  791587
5 158275  38479  70039  62210  63647  263977  146858 177087 225544  179589
      RI     SC     SD     TN      TX     UT     VA    VT     WA     WI     WV
1      0 162628  21589 225784  968401 276159  68603     0  36943 286760 633582
2 105473 349990  96787 413554 4773076 274420 699927 13801 384769 561076 277002
3  76464 508147 119505 713210 6783182 304823 795296 72241 711662 533390 205005
4      0 560782      0 368461  419363      0 297972     0  88764 103979      0
5  11798 143796 127382 135841  654199  36050 174615 21430 394052 145936  26427
      WY       US
1 376971 10548957
2 161580 31688203
3 146274 35250685
4      0  8130913
5  37734  7646167
```

```
     Energy_Source     AK     AL     AR     AZ      CA     CO     CT    DC
1 coal Consumption  18615 297654 211724 154007   30049 233256      0     0
2      Natural-Gas 437916 787300 398099 468038 2131372 524890 307212 30174
3  petroleum (btu) 263335 578431 327813 594859 3017944 538413 302881 18000
4  Nuclear Energy†      0 442093 149654 333738  183814      0 172018     0
5 total_renewables  10410 232035  90825 101215  880995 114917  49084  2622
      DE      FL     GA     HI     IA     ID      IL     IN     KS     KY
1   1846  171953 180888   7680 227866   1881  496983 719238 226712 523276
2  89674 1659544 809618    159 434374 141924 1134781 913401 318779 402534
3 112026 1815529 940579 241994 423592 190635 1138141 699235 346852 580349
4      0  321468 356001      0      0      0 1032989      0  93844      0
5   7402  304605 293237  20471 421784  78406  248541 170986 151788  77517
```

```
       LA     MA     MD     ME      MI     MN     MO     MS     MT     NC
1   96914      0  61932   1269  423504 184517 566940  66214 131345 163029
2 2087166 432442 310133  62559 1087716 535010 322547 617855  93971 747187
3 1663129 529154 411842 166724  820709 568916 606374 383366 177009 906477
4  168889      0 154742      0  271788 153546  92724  89856      0 445547
5  138209  80700  51255  93867  206811 229769  95312  66614  60644 198165
      ND     NE     NH     NJ     NM     NV      NY      OH     OK     OR
1 369340 223571   3864   6199 138077  35835    6143  539587 106855   1066
2 198986 199260  60176 755048 301279 302315 1403401 1422175 772405 297591
3 170390 237556 149025 769751 255571 303234 1321362 1031807 521629 315400
4      0  58702 114108 295875      0      0  280133  175806      0      0
5  96024 168382  39863  73187  77286  72734  269884  155282 189654 237768
       PA    RI     SC     SD     TN      TX     UT     VA    VT     WA     WI
1  435540     0 150973  24769 204725  932569 237870  67739     0  42238 232501
2 1936985 93829 361249  98288 440017 5007366 287076 665869 14046 381886 622144
3 1108074 78260 495616 118593 707095 6582173 322387 798162 71534 725931 535483
4  795783     0 568055      0 372319  434709      0 294606     0 102929 105285
5  182051 13264 145328 129978 116472  751680  37369 185638 22009 418470 150890
      WV     WY       US
1 536642 390303  9885694
2 281657 172450 33361871
3 201769 145723 35330835
4      0      0  8061020
5  28391  42079  8107353
```

```
           Energy_Source     AK     AL     AR     AZ      CA     CO     CT
1             coal_usage  18414 224926 180262 137885   28746 204826      0
2             NaturalGas 448087 775747 399566 537151 2154533 525446 304924
3        petroleum (BTU) 270391 565754 327465 599712 2996168 514174 292864
4       nuclear-energy †      0 476392 156492 329474  185192      0 142873
5 total renewable-energy  10087 222189  87277 108445 1065179 115061  48981
     DC     DE      FL     GA     HI     IA     ID      IL     IN     KS
KY
1     0    338  129387 177521      0 201276   1144  342683 613533 184614
481815
2 26236  84387 1673836 787361    152 446677 154150 1101064 921814 309427
369986
3 17292 110721 1835394 980546 251676 404172 189553 1134461 695709 345807
584722
4     0      0  312935 390663      0      0      0 1019691      0 107675
0
5  2795   8041  286306 291462  21046 414801  77128  245703 172891 140268
72603
       LA     MA     MD     ME      MI     MN     MO     MS     MT     NC
1   58224      0  30349   1295  287490 148968 442901  49606 130059 153784
2 2055504 386946 304669  61045 1104234 536789 316512 630107  96777 662302
3 1620038 525647 429784 177091  810789 567072 612625 378072 173283 900241
```

```
4  127634      0 156610      0  292615 124626  95947 122807      0 442493
5  138982  81559  53711  89444  198459 223864  90412  67305  58470 186804
      ND     NE     NH     NJ     NM     NV      NY      OH     OK     OR
1 325716 195602   1838      0  75182  29284    4823  413577  63787    652
2 220768 206276  59589 721282 337083 301655 1346622 1448857 860217 327164
3 169307 233599 147387 787262 251686 296155 1341811 1009729 515440 313013
4      0  72391  99658 296162      0      0  287690  169392      0      0
5  92154 164502  38988  74408  80278  74878  272967  153083 185378 236062
       PA     RI     SC     SD     TN      TX     UT     VA    VT     WA
WI
1  307604      0 162323  22246 202367  805600 174315  46785     0  49523
219995
2 1937041 112499 346881  99752 396870 5284670 298976 655997 13001 403038
565025
3 1132958  76844 507146 114623 702827 6752349 324640 807547 70235 718277
525386
4  787083      0 581365      0 396522  425186      0 310037     0  88163
101204
5  178035  13579 142486 126540 115678  791211  39674 183979 22209 365956
150965
      WV     WY       US
1 472309 366098  8169673
2 309019 181395 33609104
3 206969 143944 35460356
4      0      0  8098974
5  28370  38474  8187317
```

```
  electric.vehicle.registrations_by_state..2023.          X
1
2                                          STATE Count-EVs
3                                        Alabama     #13047
4                                         Alaska      ~2697
5                                        Arizona      89798
6                                       Arkansas   7108 EVs
```

After performing the monumental efforts on data cleaning, I am able to get the frames as below.

```
  State_abb  2021  2022  2023      State
1        AK 20.03 20.03 20.03     Alaska
2        AL 17.85 17.85 17.85    Alabama
3        AR 18.42 18.42 18.42   Arkansas
4        AZ 25.07 25.07 25.07    Arizona
5        CA 28.44 28.44 28.44 California
6        CO 20.64 20.64 20.64   Colorado
```

```
       State State_abb   count
1    Alabama        AL   13047
2     Alaska        AK    2697
3    Arizona        AZ   89798
4   Arkansas        AR    7108
5 California        CA 1256646
6   Colorado        CO   90083
```

```
  State_abb       Source Price_2021
1        AK      Biomass       3153
2        AK   Geothermal        186
3        AK   Hydropower       5763
4        AK Solar Energy         45
5        AK  Wind Energy        451
6        AL      Biomass     198543
```

```
  State_abb       Source Price_2022
1        AK      Biomass       3846
2        AK   Geothermal        186
3        AK   Hydropower       5846
4        AK Solar Energy         57
5        AK  Wind Energy        475
6        AL      Biomass     193932
```

```
  State_abb       Source Price_2023
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
```

```
# A tibble: 6 × 3
  Energy_Source State_abb Price_2021
  <fct>         <chr>          <int>
1 Coal          AK             18694
2 Coal          AL            309791
3 Coal          AR            216123
4 Coal          AZ            160299
5 Coal          CA             28244
6 Coal          CO            252442
```

```
# A tibble: 6 × 3
  Energy_Source State_abb Price_2022
```

```
  <fct>         <chr>          <int>
1 Coal          AK              18615
2 Coal          AL             297654
3 Coal          AR             211724
4 Coal          AZ             154007
5 Coal          CA              30049
6 Coal          CO             233256
```

```
# A tibble: 6 × 3
  Energy_Source State_abb Price_2023
  <fct>         <chr>          <int>
1 Coal          AK              18414
2 Coal          AL             224926
3 Coal          AR             180262
4 Coal          AZ             137885
5 Coal          CA              28746
6 Coal          CO             204826
```

Part 3: Joining / Pivoting Datasets for Analysis

This part will be done along with the next part on EDA.

Part 4: Mapping Visualization

Question 1: How has the share of renewable energy changed from 2021–2023 across states?

To answer this question, I firstly join the three *cleaned_renew* data frames, named and shown as q1. I then adjust this joint data frame into a format better for future plotting, named and shown as analysis_q1.

```
  State_abb       Source Price_2023 Price_2021 Price_2022
1        AK      Biomass       3404       3153       3846
2        AK   Geothermal        186        186        186
3        AK   Hydropower       6051       5763       5846
4        AK Solar Energy         67         45         57
5        AK  Wind Energy        380        451        475
6        AL      Biomass     189040     198543     193932
```

```
# A tibble: 6 × 6
  State_abb Source      Year Energy_Amount Total_Energy_Per_Year  Share
  <chr>     <fct>      <dbl>         <dbl>                 <dbl>  <dbl>
1 AK        Biomass     2023          3404                 10088 0.337
2 AK        Biomass     2021          3153                  9598 0.329
3 AK        Biomass     2022          3846                 10410 0.369
4 AK        Geothermal  2023           186                 10088 0.0184
```

```
5 AK        Geothermal  2021          186                 9598 0.0194
6 AK        Geothermal  2022          186                10410 0.0179
```

I then use the ggplot to generate a multi-facets bar plot, where each facet represents a state, in which shows the change in proportion of renewable energy across the three years.

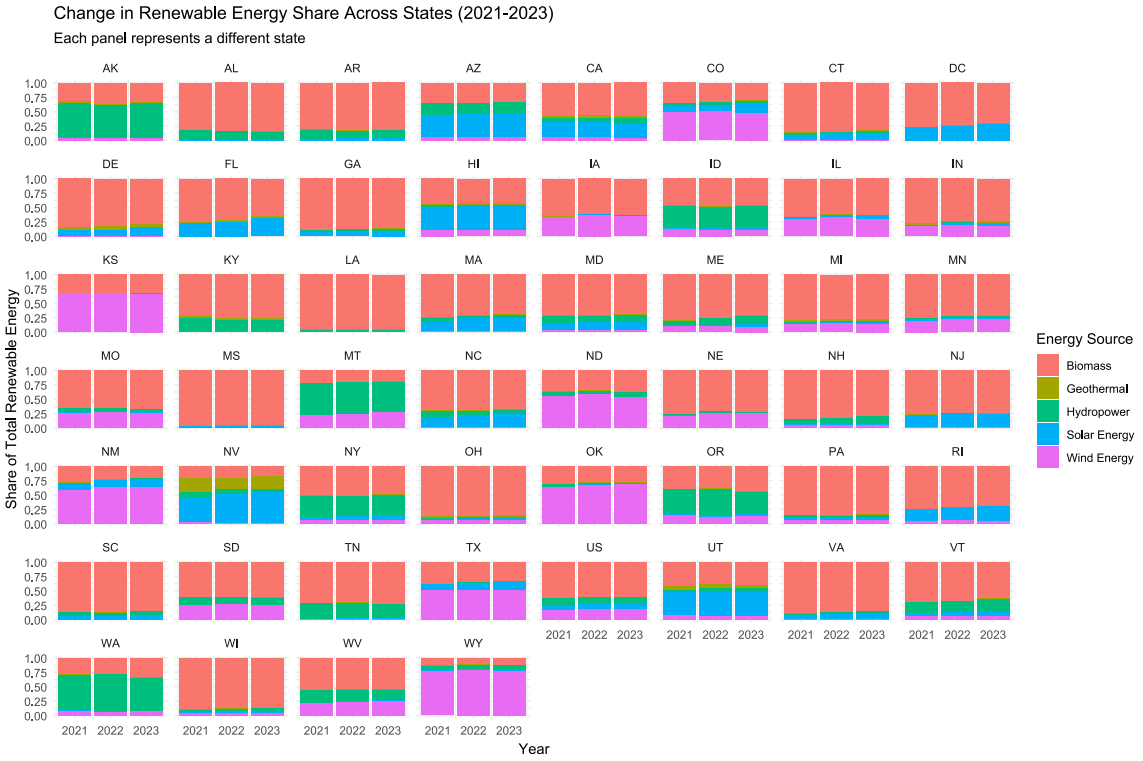
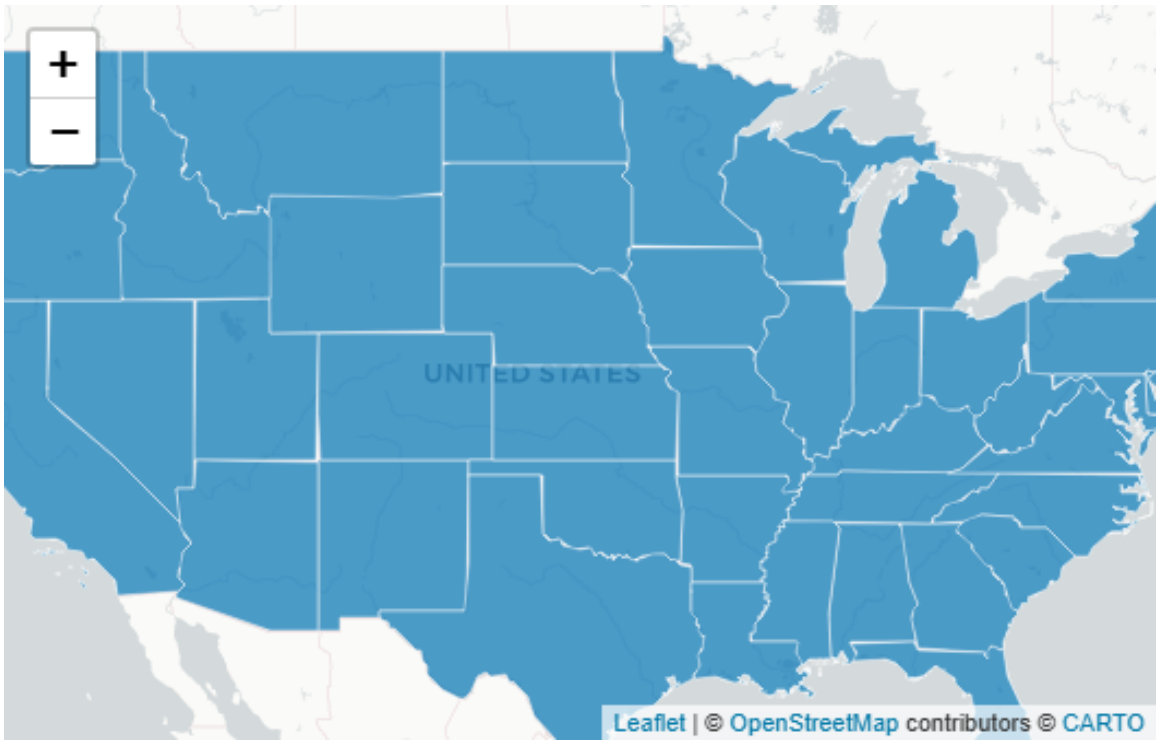


Figure 1: Change in Renewable Energy Share Across States (2021-2023)

However, since there are quite a lot of states and the change in proportion across years is small, it is quite hard to extract any understandable information from the figure. In this case, a better figure could be drawn by putting those data into the US map. With the help of Gemini, I finally able to create an interactive map which contains all the data in the previous multi-facets bar chart, which is so amazing.



Question 2: What is the share of electricity that comes from clean sources by state?

To answer this question, I join the three *cleaned_total_use* data frames named q2. Likewise, I reformat this new data frame, and calculate Share of the clean energy.

```
# A tibble: 6 × 6
  Energy_Source    State_abb  Year Energy_Amount Total_Energy_Per_Year  Share
  <fct>            <chr>     <dbl>         <dbl>                 <dbl>  <dbl>
1 Total_Renewables AK         2021          9597                684975 0.0140
2 Total_Renewables AK         2022         10410                730276 0.0143
3 Total_Renewables AK         2023         10087                746979 0.0135
4 Total_Renewables AL         2021        239817               2352656 0.102
5 Total_Renewables AL         2022        232035               2337513 0.0993
6 Total_Renewables AL         2023        222189               2265008 0.0981
```

This time, we create another bar chart, which shows the proportion of electricity that comes from each state.

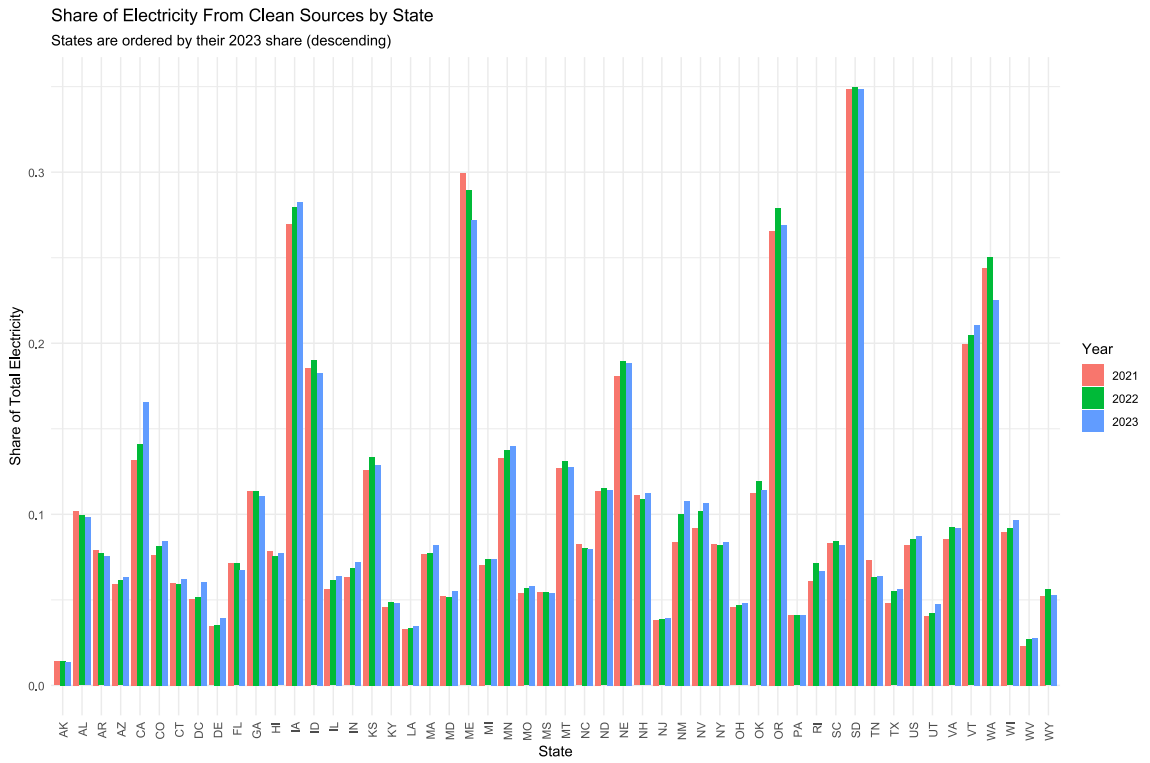
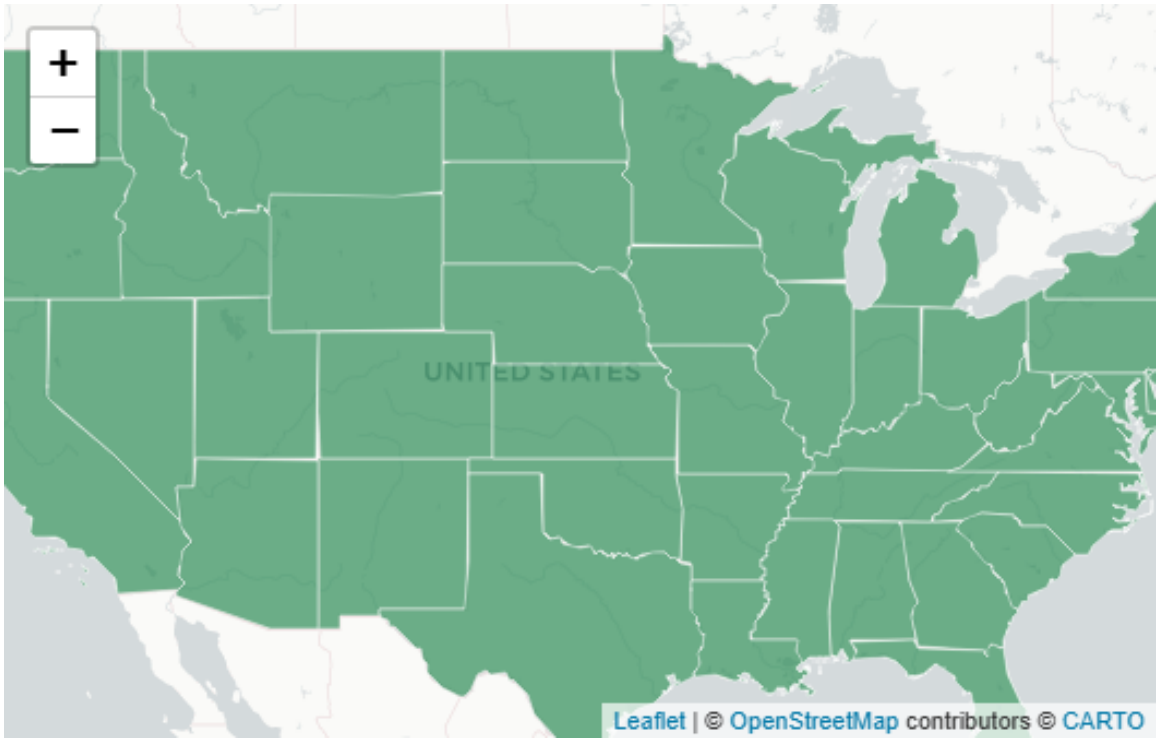


Figure 2: Share of Electricity From Clean Sources by State

Likewise, feed these data into the combination with the US map with the help of Gemini.

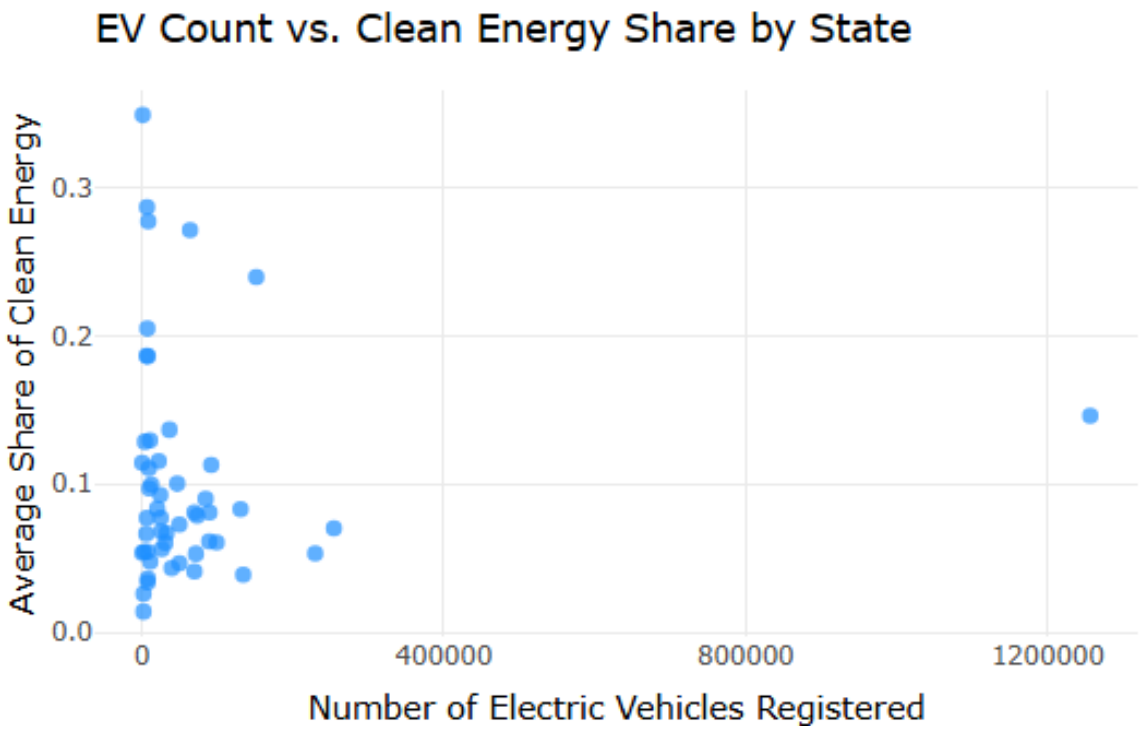


Question 3: Are EV registrations concentrated in states with cleaner energy mixes?

After getting some general senses of our data and the energy situations for every state, it would be helpful to get into some specific interrogations. Using our data from previous part, I join the data frame *q2* and *cleaned_ev_registrations*. Removing redundant columns, I get the following result.

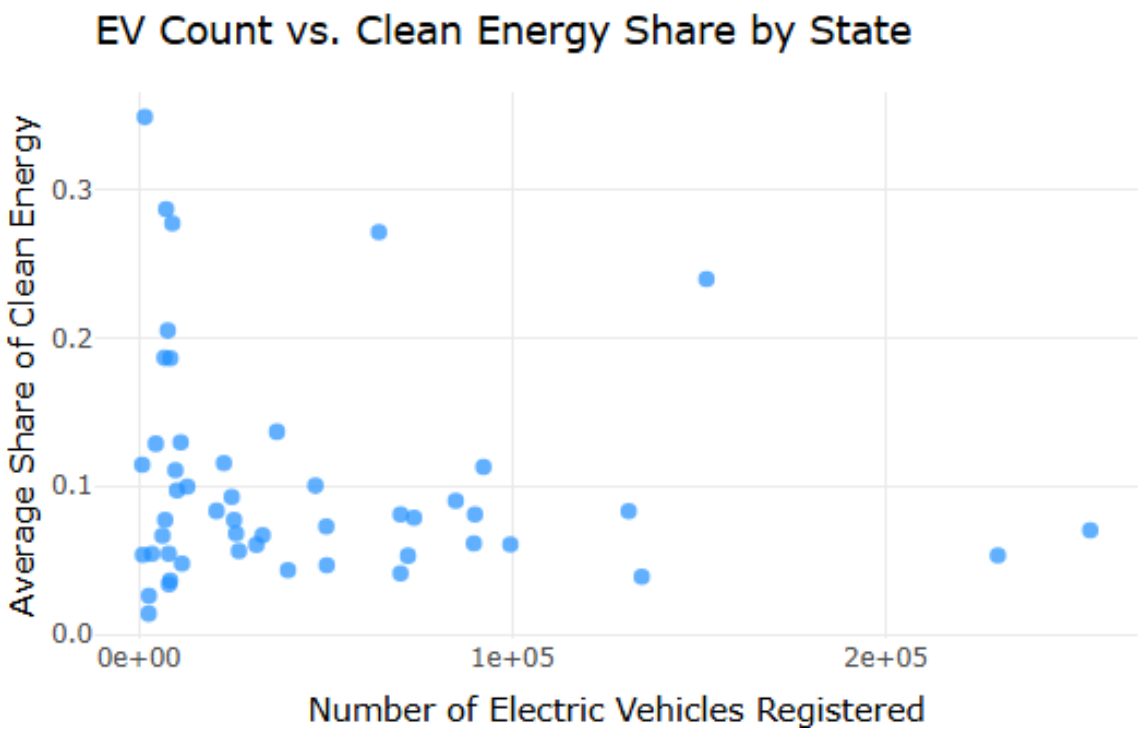
```
# A tibble: 6 × 3
  State_abb EV_count avg_share
  <chr>        <dbl>     <dbl>
1 AL           13047    0.0998
2 AK            2697    0.0139
3 AZ           89798    0.0612
4 AR            7108    0.0773
5 CA         1256646    0.146
6 CO           90083    0.0808
```

To visualize whether there is a relationship between clean energy proportion and EV registrations, I made a scatter plot withnumbers of EVs against the average clean energy shares among 2021, 2022, 2023. To make it better for visualization, I make it interactively to show the name of the state for each point, with the help of Gemini.



Nevertheless, I'm facing the issue that the *EV_count* for California are far more than other states, which distorts the figure. Therefore, treating it as an outlier by removing it, I can replot the figures as above.

```
# A tibble: 6 × 3
  State_abb EV_count avg_share
  <chr>        <dbl>     <dbl>
1 AL           13047    0.0998
2 AK            2697    0.0139
3 AZ           89798    0.0612
4 AR            7108    0.0773
5 CO           90083    0.0808
6 CT           31557    0.0604
```



The figures give us a general sense on the relationship between these two variables, but to make more justifiable claim, we need the help of testing statistics. Below, I use a linear model to fit the two variables, and calculating some characteristic values as shown.

```
Call:
lm(formula = avg_share ~ EV_count, data = q3)

Residuals:
     Min       1Q   Median       3Q      Max
-0.08716 -0.04693 -0.02359  0.01357  0.24791

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 1.010e-01  1.119e-02   9.028 5.35e-12 ***
EV_count    1.708e-08  5.893e-08   0.290    0.773
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07434 on 49 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.001712,  Adjusted R-squared:  -0.01866
F-statistic: 0.08402 on 1 and 49 DF,  p-value: 0.7731
```

```
    Pearson's product-moment correlation

data:  q3$avg_share and q3$EV_count
t = 0.28986, df = 49, p-value = 0.7731
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2369117  0.3133837
sample estimates:
       cor
0.04137317
```

From the summary, we get a p-value larger than 0.05. Also, the correlation coefficient is very insignificant. Therefore, we cannot reject our null hypothesis, that there is no clear linear relationship between clean energy shares and number of EVs registered.