Is Solar Energy A Political Issue

Tuckman, Jonah

Introduction:

My research question is whether or not the solar output of each state can be dictated or influenced by political affiliation. My original thought was a related but different question, but based on a poor representation of overall data, I pivoted my scope of searching within the solar energy field and landed on this question. The question that I initially hoped to explore was whether headquarter location of top 10 (in total company valuation) American oil companies affected statewide solar output. The hiccup I ran into when exploring this was that over 70% of these companies had their headquarters in Texas. This lead to a poor dataset as the question was now comparing the solar output of 51 states compared to headquarters located in 3.
After I realized this I changed the scope of my question to explore whether solar output statewide could see a correlation between the positive output and political voting patterns.
The idea of solar energy is one that is globally impactful for all states and people, independent of political affiliation. Thus I wish to explore this issue and identify a trend between political leaning, to see if politics are getting in the way of the nationwide greater good.
Positive aspects of solar energy include a limited reliance on oil (and thus foreign policy implications as the US would not need to send soldiers to wars fought over dictating oil-rich regions), climate change reduction (the fossil fuel emissions would be limited if cars, trains, planes and more were running off of solar power rather than currently harmful, emission heavy, tendencies), fiscal advantages (government stipends for those who opt into a switch into renewable energy) and many more.
I found interest in this question after deciding to work for a company over the summer who works towards using Artificial Intelligence to advance the output and profitability of solar fields. Doing so will allow the business side of solar field generation to become more profitable and beneficial, and with big money backing it allow for an appropriate adversary to big oil companies. Once understanding this aspect of solar energy I began to wonder whether one aspect still holding us back could be political ideals.

Data:

Original Data Set: https://catalog.data.gov/dataset/energy-generation-by-state-and-technology-2009/resource/bb0da868-f498-4c47-b61f-da217897198f

My initial data was from a 2009 study of Energy Generation by Source and State alongside the results from the 2008 general election broken into statewide results. I initially sliced this Energy Generation csv in order to solely include State and Solar output. The original dataset includes all types of energy output (coal, petroleum, solar, wind, etc) but I decided to limit this study to solely solar output. A thought I had for a later extension of this project would be to broaden the exploration and create a boolean variable for all renewable energy output (if any of the renewable energy categories are positive then RENEWABLE = 1, else RENEWABLE = 0). This would encompass states who have a positive renewable output that does not include solar, thus making our exploration more accurate.
Once I had a set of States and Solar output, I pulled data of voting records and merged this into the current dataset by state name. Now my data includes State, Voting History, and Solar output. From here I did a variety of things to be used later including splitting into separate Democratic and Republican sets, adding a boolean variable of Solar Positivity to the original set and adding a boolean variable of Republican to the original set.

Exploratory data analysis:

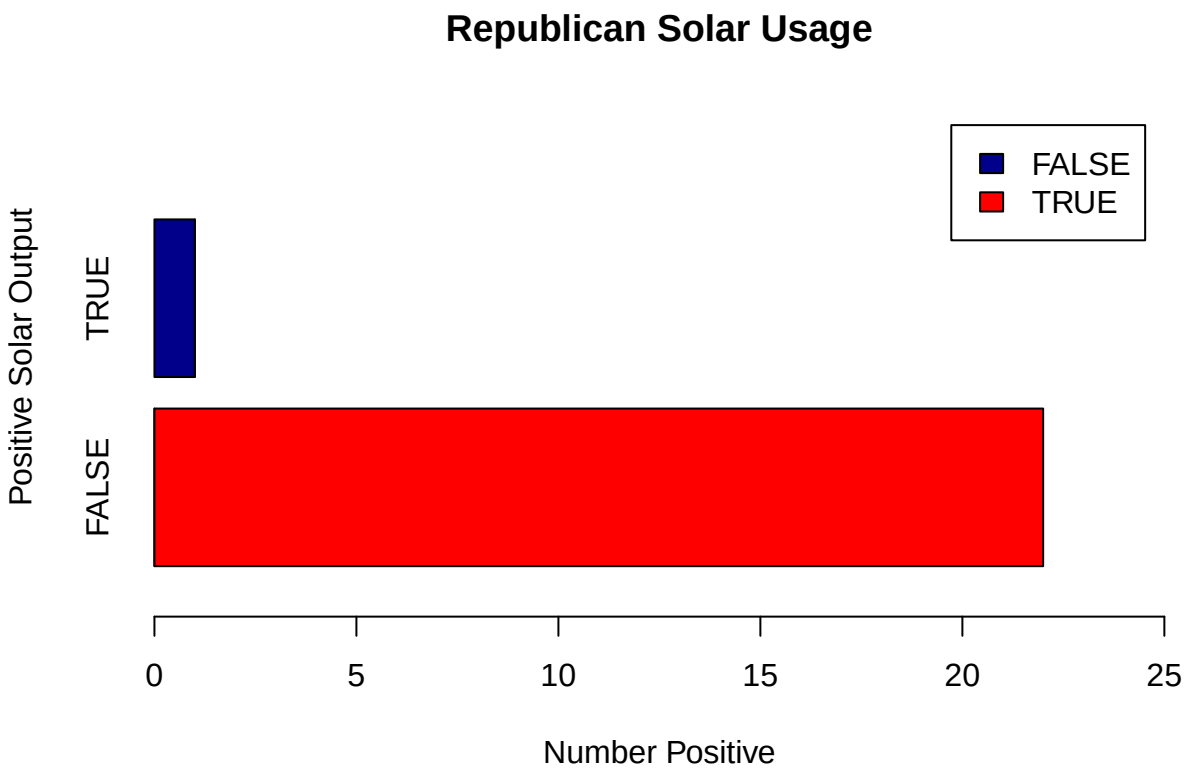
Party Distribution Percentages

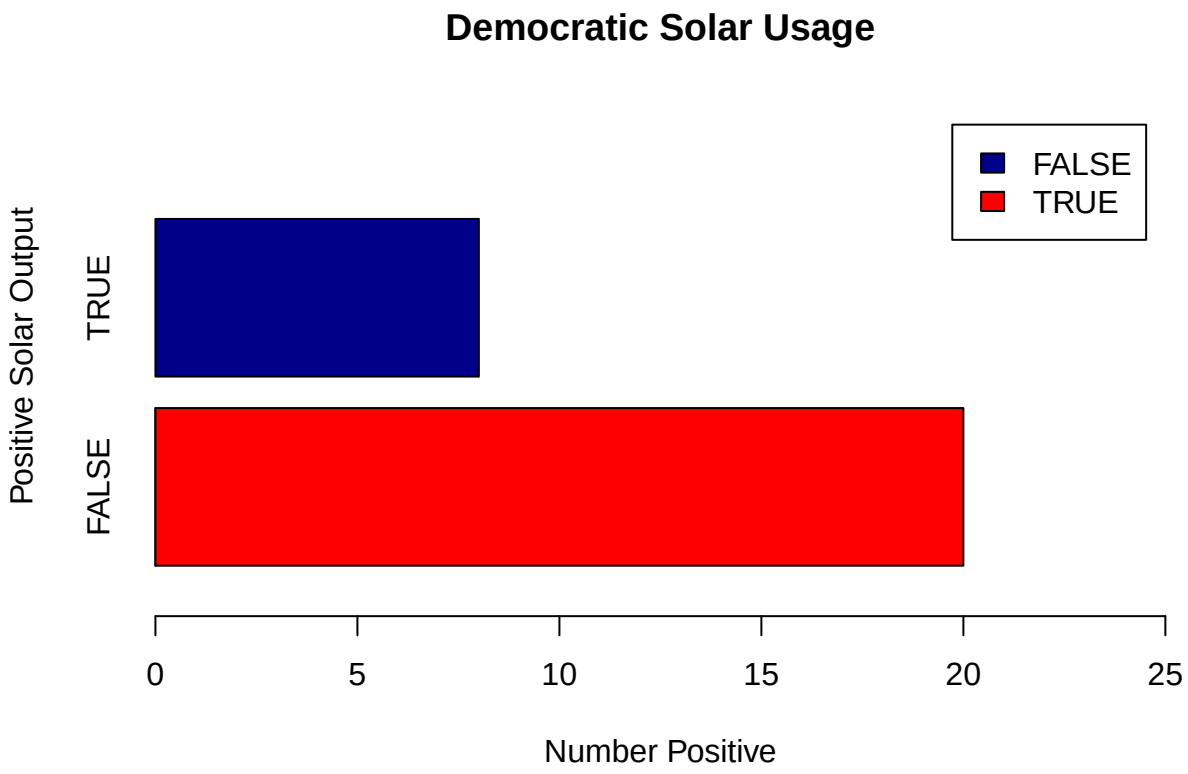
```
## [1] "Democrat Positive Solar Output Percentage:  28.571429"
```

```
## [1] "Republican Positive Solar Output Percentage:  4.347826"
```

This shows the percentage of each political party affiliated state that has a positive solar output. 28.57% of the democratic states have a positive solar output and only 4.35% of the Republican states have a positive solar output. This shows a large difference in output based on broken groups.

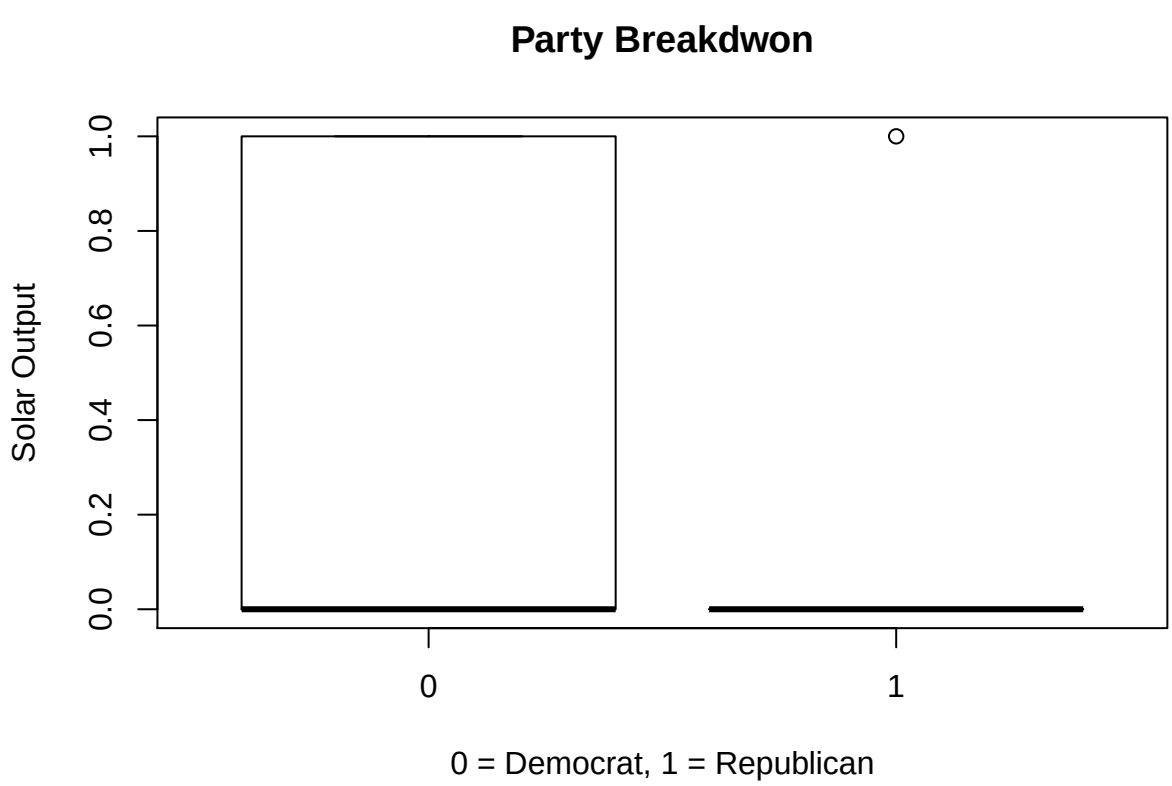
Party Distribution Barplot





In these plots, we see the republican portion alongside the democratic portion on the same horizontal range. This allows us to see the difference in internal to party breakdowns of solar output. We can see that within each party, a higher portion of democratic states have a positive solar output when compared to that of Republican states.

Party Split Boxplot



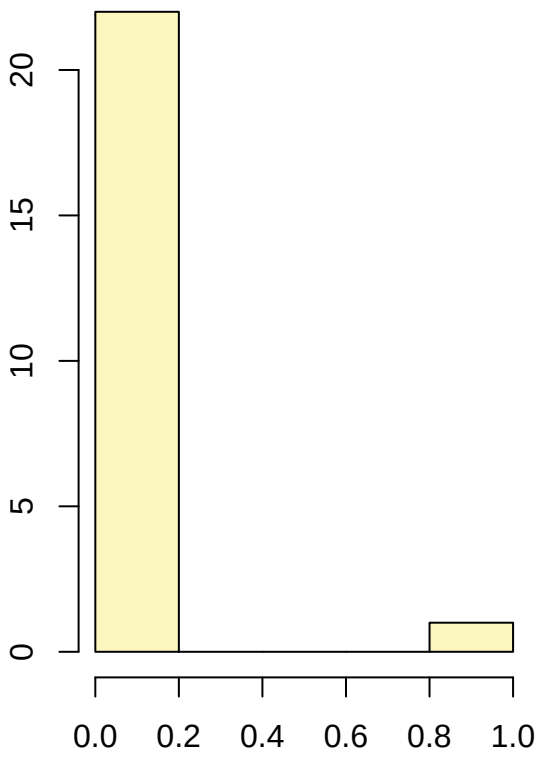
Although the above boxplot is not the prettiest representation, it does share something important. As we know, in a box plot the lower quartile to the upper quartile (the box) shows us a range of values that encompasses the 25th to the 75th quartiles. Thus this includes the median as well as the middle 50% of values.
In the boxplot shown above we can see that within the Republican portion of this, this middle 50% and the bottom 25% (the bottom 75%) is all at the value 0. The maximum is shown at 1 which indicates that there are indeed republican states that have a positive solar output, but this also indicates that at least 75% of these Republican states have no solar output.
In the democrat boxplot, we see a median at 0, which means that less than 50% of the democrat states have positive solar output, but we see the upper quartile at 1. This indicates that more than 25% of democrat aligned states have a positive solar output.

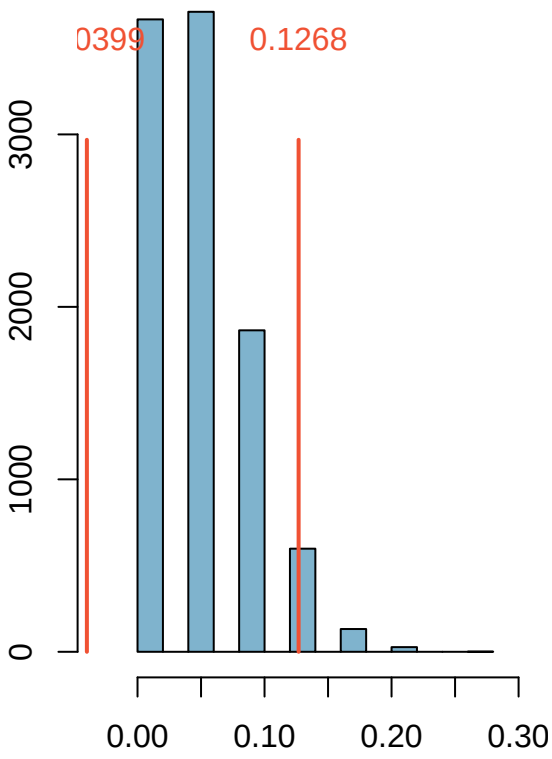
Inferences For Parties

```
## Warning: Ignoring success since y are numerical.
```

```
## Single mean
## Summary statistics:
```

```
## mean = 0.0435 ;  sd = 0.2085 ;  n = 23
```



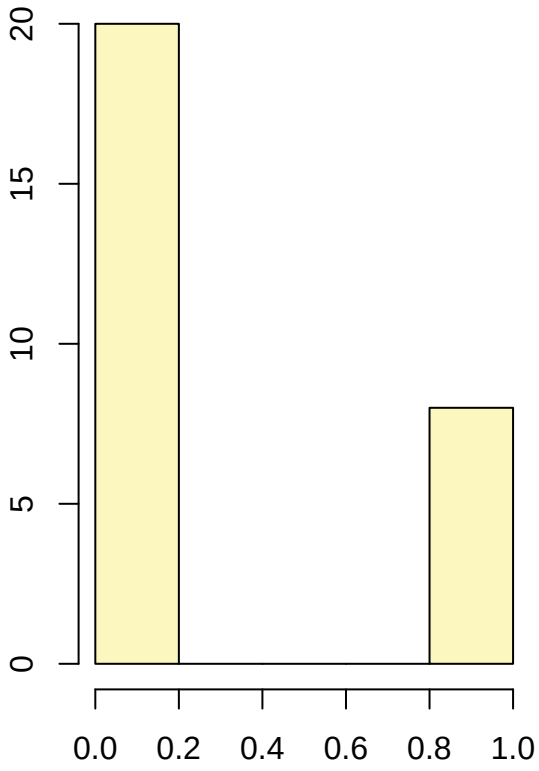


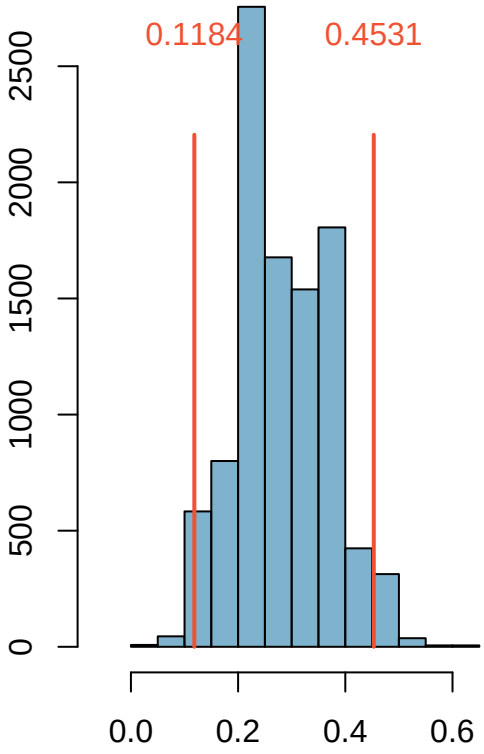

```
## Bootstrap method: Standard error; Boot. SE = 0.0425
## 95 % Bootstrap interval = ( -0.0399 , 0.1268 )

## Warning: Ignoring success since y are numerical.

## Single mean
## Summary statistics:

## mean = 0.2857 ;  sd = 0.46 ;  n = 28
```






```
## Bootstrap method: Standard error; Boot. SE = 0.0854
## 95 % Bootstrap interval = ( 0.1184 , 0.4531 )
```

```
## [1] "Condifence intervals: "
```

```
## [1] "Republican: 95 % Bootstrap interval = ( -0.0422 , 0.1331 )"
```

```
## [1] "Democrat: 95 % Bootstrap interval = ( 0.1195 , 0.4519 )"
```

In the republican 95% confidence bootstrap interval the confidence range has a negative minimum and includes 0 in the range. Clearly, our data could not be negative considering they are boolean responses, but the fact that 0 is included is very important as it shows that we are 95% confident that 0 is in the range of possible values for the mean of Republican Leaning States.
In the democrat data, this is not the case, 0 is not in the range thus we are 95% confident that the mean number of democrat leaning states with positive solar output is not 0. Although the range also does not include 1 meaning that we are also confident that not every democrat leaning state has positive solar output. Our takeaway from this is that, in a 95% confidence interval, we can use this confidence to state it is possible for no Republican-leaning states to have a positive solar output, and not possible that no democrat leaning states have a positive solar output.

SolarPositive ~ Republican T Test

```
##
```

```
##  Welch Two Sample t-test
##
## data:  statsData$SolarPositive by statsData$Republican
## t = 2.492, df = 39.186, p-value = 0.01704
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.04564886 0.43882319
## sample estimates:
## mean in group 0 mean in group 1
##      0.28571429      0.04347826
```

Conditions that need to be met to perform this T-Test. 1. The data is a random sample from the population. -> In this case the data is taken from the general population. The Solar output and voting records are not randomly collected but are both valid in their sources of collection. 2. The sampling distribution is nearly normal -> It is assumed that the sampling here is nearly normal. This can be seen in the prior tests of bootstrapping plots.
3. Individual observations are independent. -> In this case I am under the assumption that states do not coordinate voting patterns (a strong assumption) and that they do not coordinate solar output (medium strength assumption).

This T-Test is built by looking at the solar positive and the Republican boolean variables. The two means we are testing is the mean which indicates the probability that a solar positive value of 1 is equal to a Republican value of one, in simpler terms if a state has a positive solar output, is this state Republican-aligned?
I ran this T-Test to determine if we can conclude that there is significant evidence of a factor resulting in this data, or if this output discrepancy is purely due to chance. The data resulted in a T-Value of 2.492 and a P-Value of 0.0174. This p-value is less than 0.05 thus we are able to reject the null. This means that we are able to understand that we would not be able to obtain this data if there was not some influence based on an external factor, in this case, political affiliation.

Inference:

Goal: To simulate the Republican portion of the data (23 states) and use null and alternative probabilities within their respective tests and count the number of solar positive occurrences.

Null Hypothesis: H0 <- The probability that a state in our Republican simulation will have a positive solar output is 9/51 (the overall probability of all states excluding political factors). This would mean that we could reasonably get this data with no information concerning the states political history.

Alternative Hypothesis: HA <- The probability that a state has a positive solar output is 4.25% (1/23) which is the overall probability that a state has a positive solar output given it is Republican.

When recreating vectors the size of the total Republican states, we are simulating our data vs the null hypothesis.

Simulation and T-Test

```
##
##  Welch Two Sample t-test
##
## data:  newFrame$nullValsVec and newFrame$testValsVec
## t = 45.058, df = 1498.8, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.888529 3.151471
```

```
## sample estimates:
## mean of x mean of y
##     4.015     0.995
```

After simulating two vectors 1000 times with the null probability (the probability a Republican state would have positive solar output independent of political alignment) and the alternative probability (the probability that a Republican state would have positive solar output given its political view), I ran a T-Test on these two vectors to find whether or not the null can be accepted.
After running this test I found a p-value of 2.2e-16 which is incredibly small. This would lead me to reject the null hypothesis and state that we are able to understand that this data could not be found solely based on the probability of positive solar output statewide. It is not enough to know the number of states who have solar output and use this probability in recreating the data, we also have to understand more factors, which in this case, is the political alignment.
When recreating the republican vector in my simulation, using the probability of a Republican state to have a solar output (4.35%), the simulation found a mean of 1.054 states with positive solar output, which is close to the true value in the data of 1. When using the probability that any state has positive solar output, the simulation found a mean of 4.090 which is much higher than that found using solely Republican probability information. This is based on the overall nationwide probability of 17% chance of finding a state with positive solar output.

Thus a further look at this data indicates that there is more to the eye than just nationwide probability of a state having positive solar output and thus we reject the null.

Conclusion:

When looking at this data, I was able to conclude that political factors, or some other unseen factor, do indeed affect the amount of solar output that a state sees. I can even go further and say that, based on probabilities of states containing a solar positive energy output within the groups, the data shows that democrat leaning states are more likely to have this positive output than Republican-leaning states.

When doing these tests I learned that, although solar energy has wide positives for all, there are political issues that are holding it back from widespread use nationwide. Within this understanding I was also able to learn the beauty the data exploration presents us with. I began this project with a question thought of while reading an unrelated article about solar energy. I read about the ways the solar field was changing and the good it could do for the world, and my mind brought me to the question 'Why is this not happening faster?'. Through data exploration, I was able to build a model that not only wondered the same thing but proved possible reasoning and shared factual evidence that there was indeed a valid question to be asked.

I believe that the solar industry, alongside a grouping of other renewables, will revolutionize how the world uses energy. This shift will encompass oil-dependent nations and drive the economies of those who are able to drive the change. The political issue at hand will need to be overlooked if we would like to take ourselves out of a dying and unhealthy industry into a fresh and new one whose outlook is to clean the world up.

As I mentioned above, I would like to recreate this test and broaden the scope of the comparison to not just positive solar output but positive renewable output. This will allow us to look deeper into the question of renewable energy and the factors that hold it back.

After understanding the impacts of the solar industry and through data seeing the limitations of it as well, I would hope to see these barriers broken in order to create an energy structure that is built and harvest, rather than mined and controlled, while also cleaning up our atmosphere slowly but importantly.