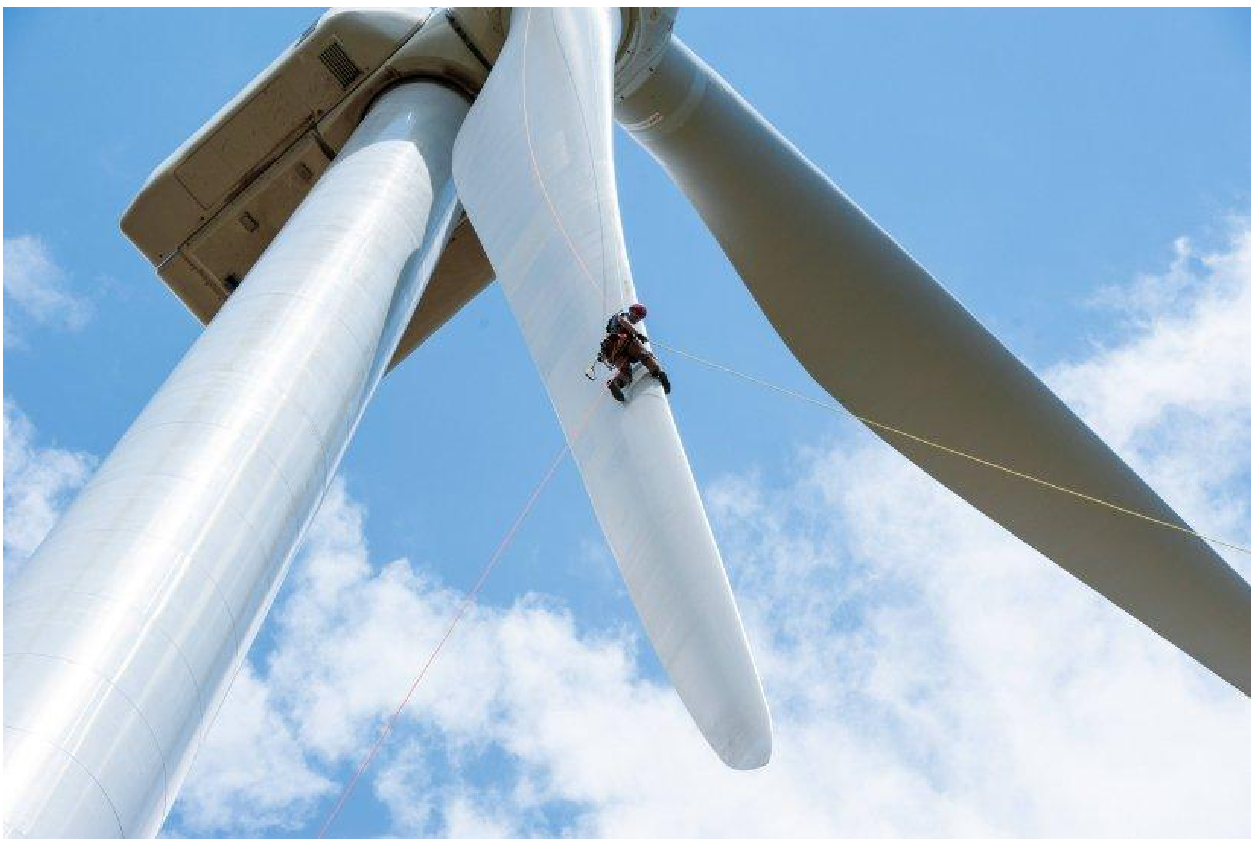
An Analysis of
Wind Farms in the Contiguous United States

Introduction

Renewable energies are an inexhaustible, pollutant free source of power. Wind energy, also referred to as eolic energy, is the most efficient source of renewable energy. Wind turbines are 50% efficient compared to the 23% efficiency of solar panels, offering an economically viable abatement toward climate change. Wind turbines capture wind with large blades attached to a hub. As wind currents rotate these blades, which are attached to a rotor, the rotor spins to generate electricity. Wind turbines installed in a group at a specific location are referred to as “wind farms”.



www.energy.gov: **Wind turbines are soaring to record sizes. The average rotor diameter of turbines installed in 2018 grew to 379 feet, up 141 percent since 1998–1999. | *NREL***

Research Questions

Given the potential of this energy source, and an estimated 67,000 turbines currently installed in the United States; it is of interest to investigate the location attributes and power production from wind farms and turbines currently installed in this country.

This project asks the questions:

1. What is the landscape of wind farms and turbines for the contiguous United States?
2. What is the landscape of megawatt production across the contiguous United states?
3. Are there seasonal trends by state or region?
4. Are there significant increases in megawatt production with retrofitting?

Data Description

To answer these questions, data was gathered with eia (Energy Information Administration) and USGS (United States Geological Survey) APIs (Application Programming Interfaces), and combined into a dataset. The data was filtered to contain only wind farms and turbines located in the contiguous United States (hereafter referred to as the U.S.). Variables utilized from USGS and eia in the analysis are described below.

USGS:

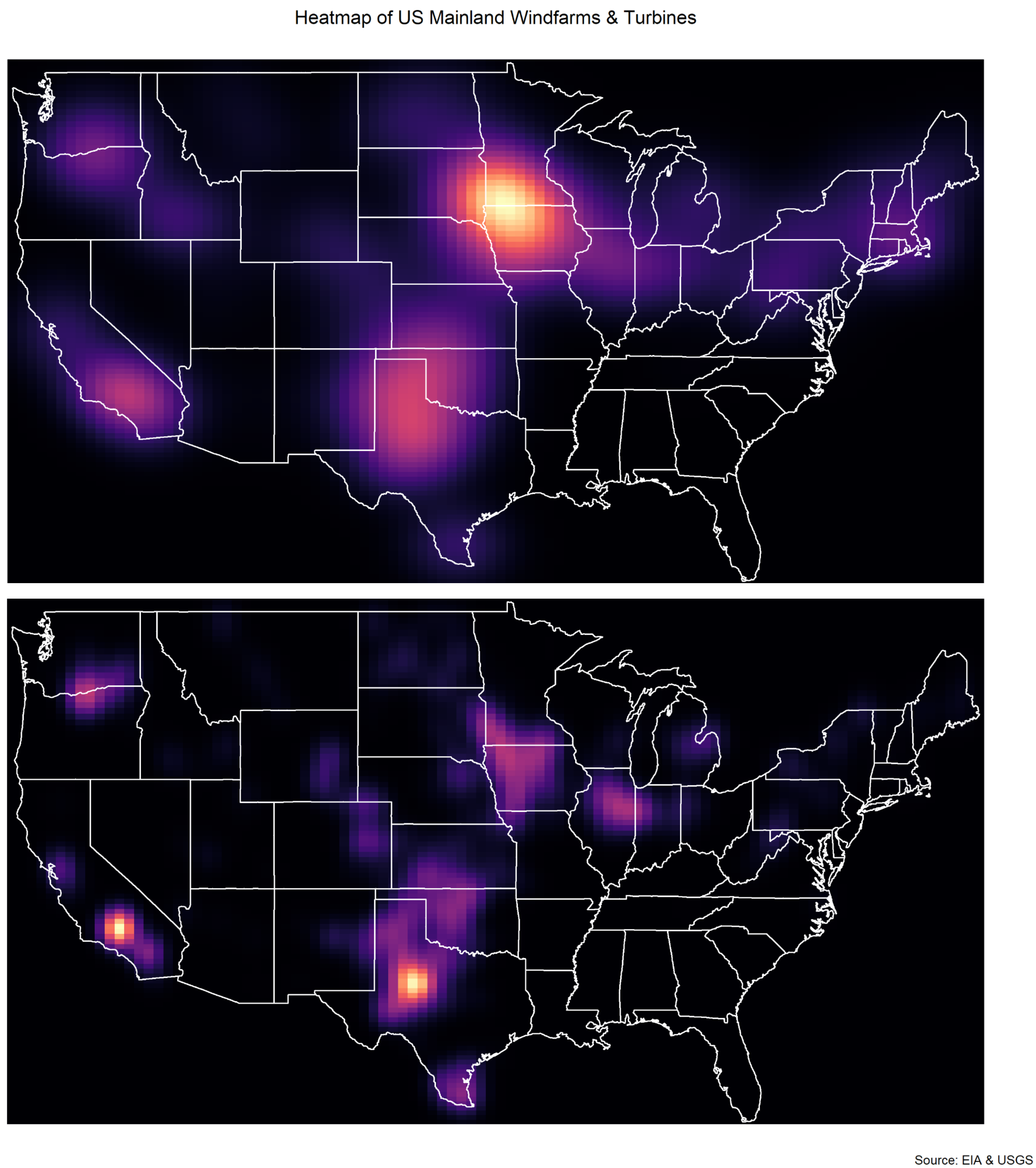
- Latitude - Latitude of the turbine location, in decimal degrees.
- Longitude - Longitude of the turbine location, in decimal degrees.
- Elevation - Turbine elevation in meters.
- Year Operational - Year in which the turbine became operational and providing power.
- Retrofit - Indicator of whether the turbine has been partially retrofitted after initial construction (e.g., rotor and/or nacelle replacement).
- Year Retrofitted - Year in which the turbine was partially retrofitted.
- t_cap - number (integer) Turbine rated capacity - stated output power at rated wind speed from manufacturer, ACP, and/or internet resources in kilowatts (kW).
- t_hh - number (float) Turbine hub height in meters (m).
- t_rd - number (float) Turbine rotor diameter in meters (m).
- t_rsa - number (float) Turbine rotor swept area in square meters (m2).
- t_ttlh - number (float) Turbine total height from ground to tip of a blade at its apex in meters (m).

eia:

- Megawatt Production - A farm’s megawatt production for a given month.
- Region - A given wind farm's geographic region.
- Capacity - Wind farm capacity, in megawatts.

Geographic Locations of Wind Farms and Turbines

To investigate the landscape of wind farms in the U.S., heat maps of wind farm and wind turbine locations were generated.



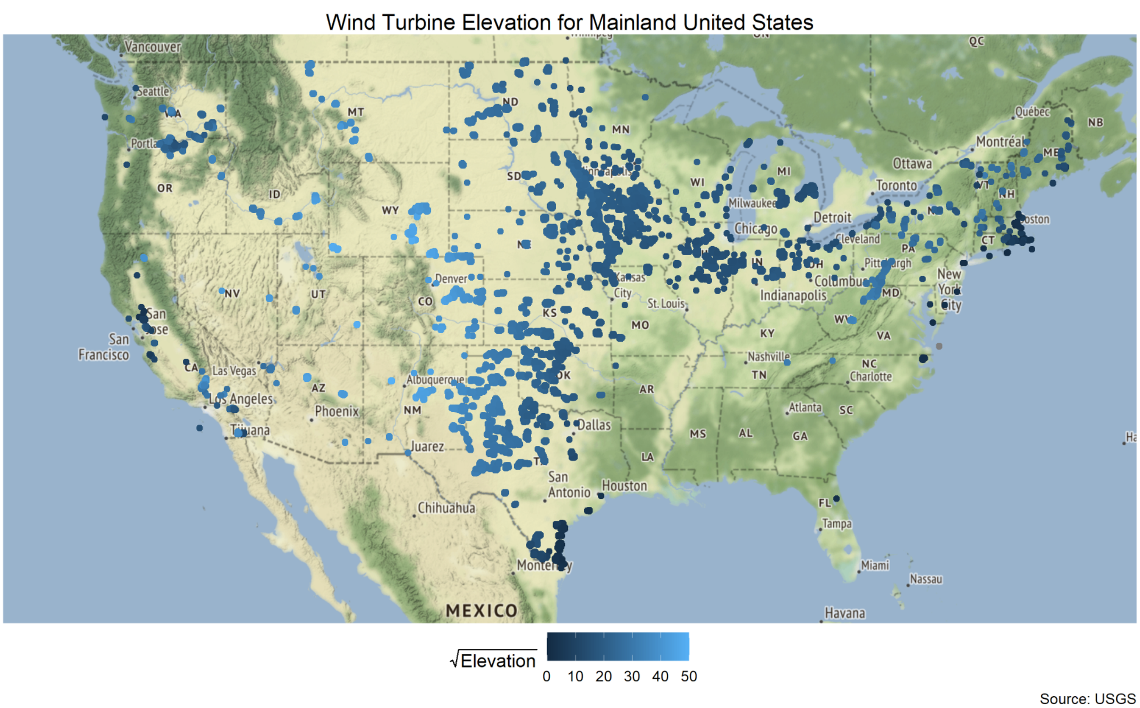
The first heat map displays the location of wind farms within the U.S. This heat map indicates the greatest concentration of wind farms in the midwest, along the borders of Minnesota, Iowa, South Dakota, and Nebraska. Notice the apparent absence of wind farms in the southeastern U.S. As California (CA) and Texas (TX) are famous for their wind power density, it is not surprising to see a concentration of wind farms in these locations. However, it is unexpected that the largest density would occur in the midwest.

To investigate this further, a second heat map, showing wind turbine locations was constructed. This second map shows a large concentration of wind turbines in California (CA) and Texas (TX), as expected, with a smaller, though still noticeable concentration at the borders, of Minnesota, Iowa, South Dakota, and Nebraska; a smaller concentration at the borders of Illinois and Indiana as well as the borders of Washington and Oregon; and an even smaller concentration in Wyoming and Colorado.

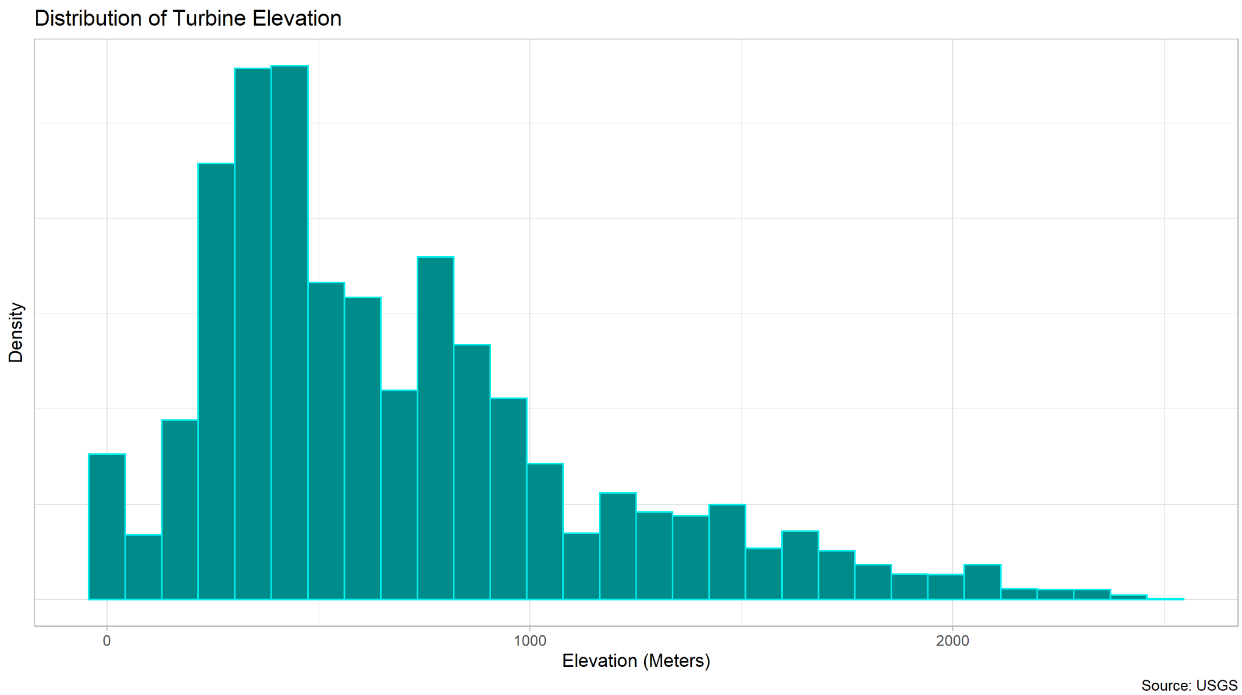
Potential reasons for the divergence of these graph's conclusions could be the existence of smaller wind farms in the midwest, and/or the potential that wind farms cross state borders, which may split the farm due to regulatory or other influences.

Turbine Elevation

As turbines rely heavily on wind currents, and one would expect these currents to vary with elevation, a map was created to determine the distribution of wind turbine elevation.



This map does not show any strong homogeneity. The majority of turbines appear to have a relatively low elevation: below 400 m. To investigate further, a histogram showing the distribution of turbine elevation on the density scale was constructed.



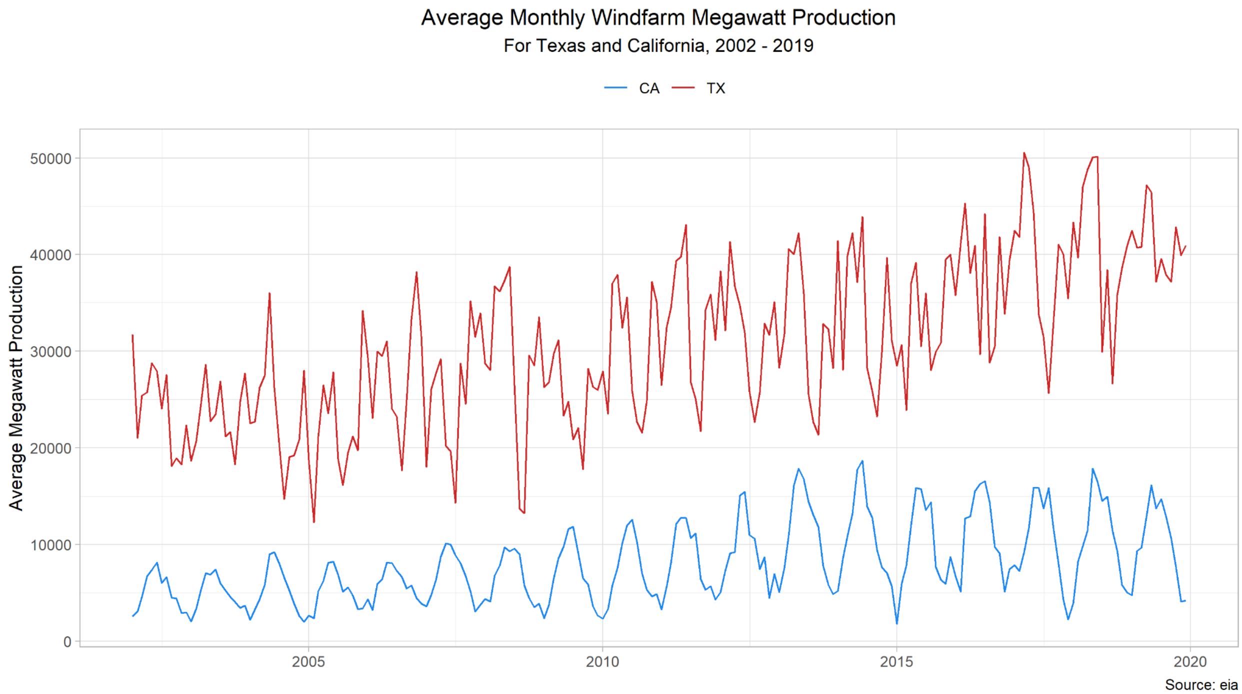
This histogram shows that few turbines are installed at sea level. The data is right skewed, indicating that most turbines are higher than sea level, but not particularly high, with a mean and median of 687 m and 561m, respectively. The peak of the density closely matches the elevation of Pittsburgh (372.8 m). Most turbines are located at elevations higher than Los Angeles, CA (93 m) and Dallas, TX (131 m). A KS test returns a p value of approximately zero, and we can, therefore, reject the null hypothesis and conclude that the histogram does not follow a normal distribution.

Wind Farm Production vs Time

We want to investigate the time series data between California and Texas, and motivate whether their trends apply to all turbine locations. In this section, we will be using California and Texas because they demonstrate the highest concentration of wind power density in the United States. By starting with the two variables with the most significant results, we can aim to generalize our findings for the entire country.

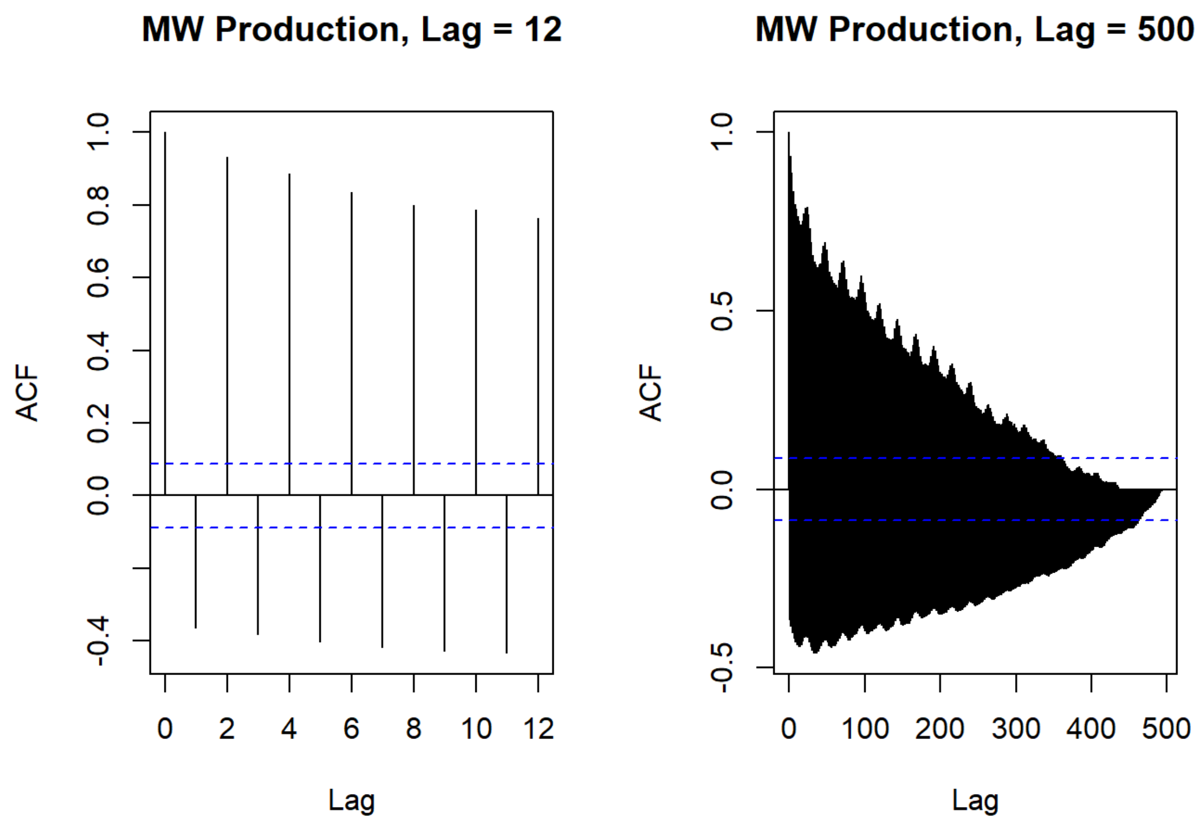
In particular, we are researching: Do different areas of the United States follow the same trends for megawatt production? If not, what seasonality trends can we find out about each state?

We will begin by comparing the simple time series data between California and Texas:



We notice that across all time periods, Texas consistently outperforms California in megawatt production. However, their trends appear to be fairly similar, where each state's peaks and troughs occur around the same time.

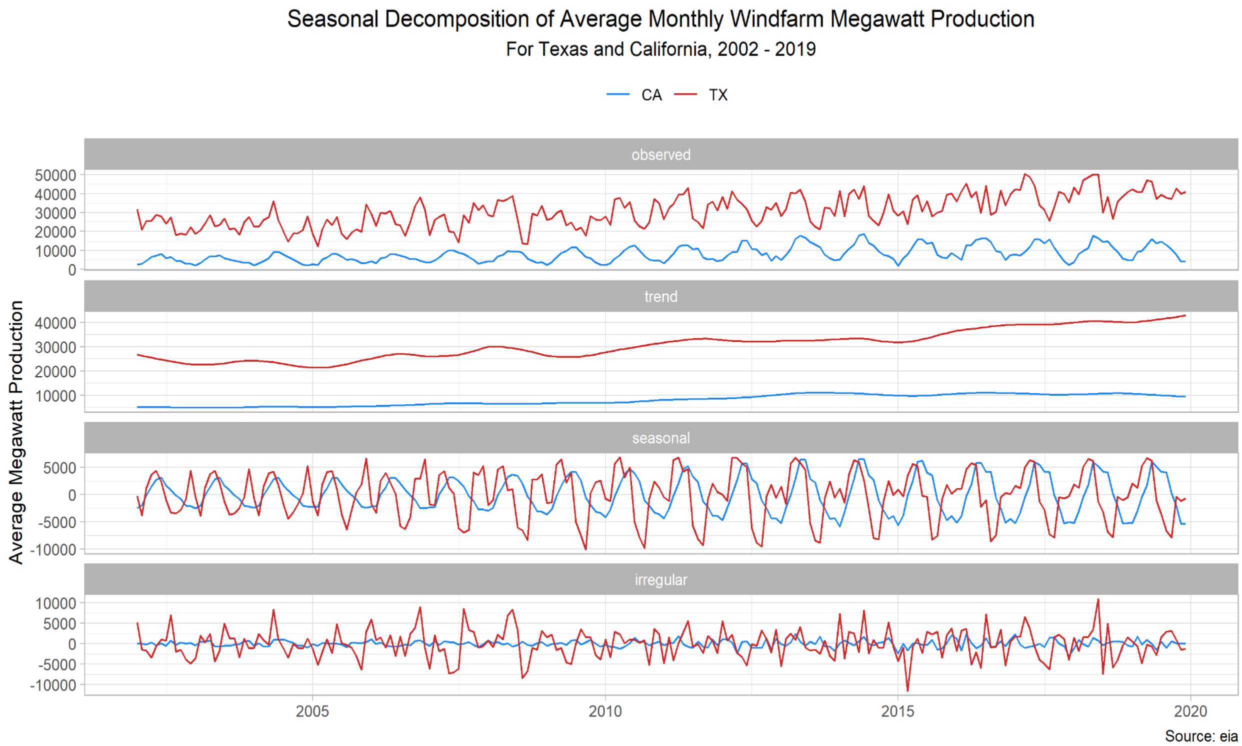
In order to determine if a trend exists across the entire country, we will analyze the autocorrelation of the combined data of California and Texas megawatt production. We do this because California and Texas make up a majority of wind farm production, and we will afterwards analyze the differences between the two states using a seasonal decomposition graph.



Our plots show that there are significant autocorrelations with lags of less than around 350 months. These significant autocorrelations would indicate that there is a strong relationship between megawatts produced during one month with respect to its previous months. However, we see that this relationship gets weaker over time.

An interesting note is that our autocorrelations are significant despite taking both positive and negative values. This observation can be explained by potential seasonal factors, where production in one season is correlated either positively or negatively with the production in another season.

Let's create a seasonal decomposition graph of this dataset to explore their seasonal trends. More particularly, we will separate the two lines by plant state locations.



We see that Texas and California follow similar trend patterns. However, it may be helpful to notice that the global trend of megawatt production in Texas in 2000-2005 saw a drastic increase in production compared to California.

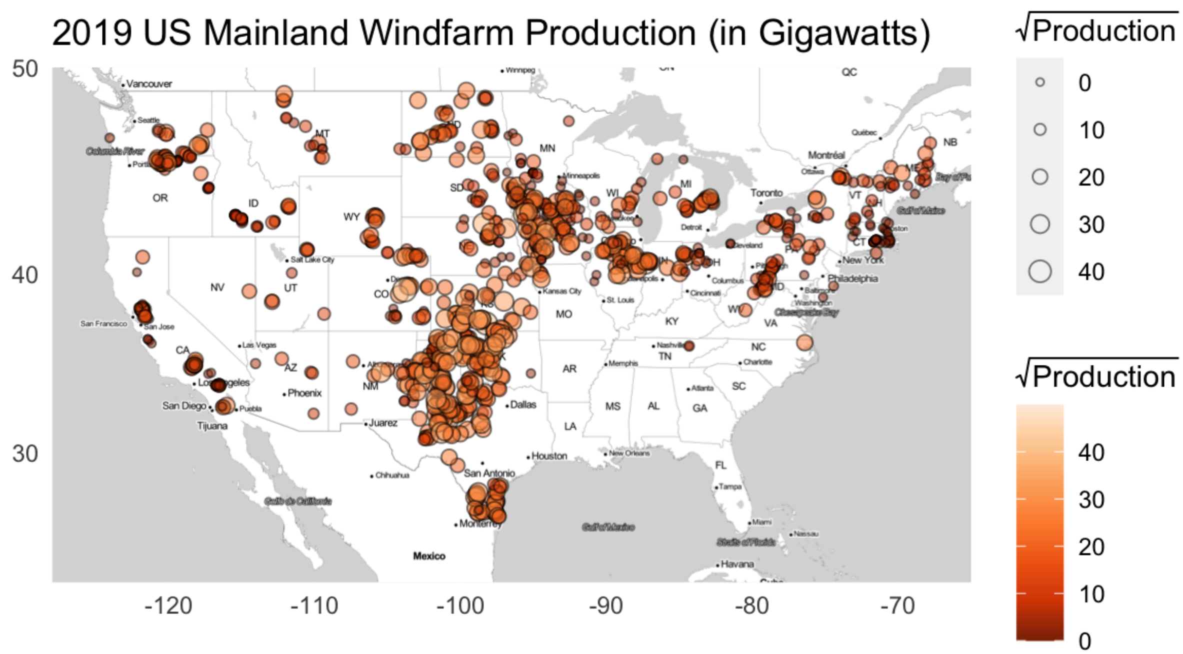
Looking at each state's seasonal trends, there are three notable differences between California and Texas:

1. The heights of both trend lines seem to increase over time. This may indicate that over time, our wind farms are more dependent on seasonality, while retaining this trend past 2008.
2. Looking closely at Texas vs California's seasonal production, we see that Texas is much more noisy than California's graph. In fact, for every seasonal cycle that California goes through, Texas goes through about two cycles. This observation, however, diminishes over time, where we notice that in recent years Texas slowly begins to "morph" their cycle's shapes to be more similar to California's. This stabilization is interesting to note because we can then generalize this finding to show that other states are also beginning to follow the same seasonal cycles as one another.
3. Lastly, the peaks of Texas are slightly prior in time compared to the peaks of California. We suspect that this is due to the geographic location of the two states. This will motivate us to further investigate what about a geographic location determines their megawatt production, which we will cover in our next section.

Wind Farm Production vs Location

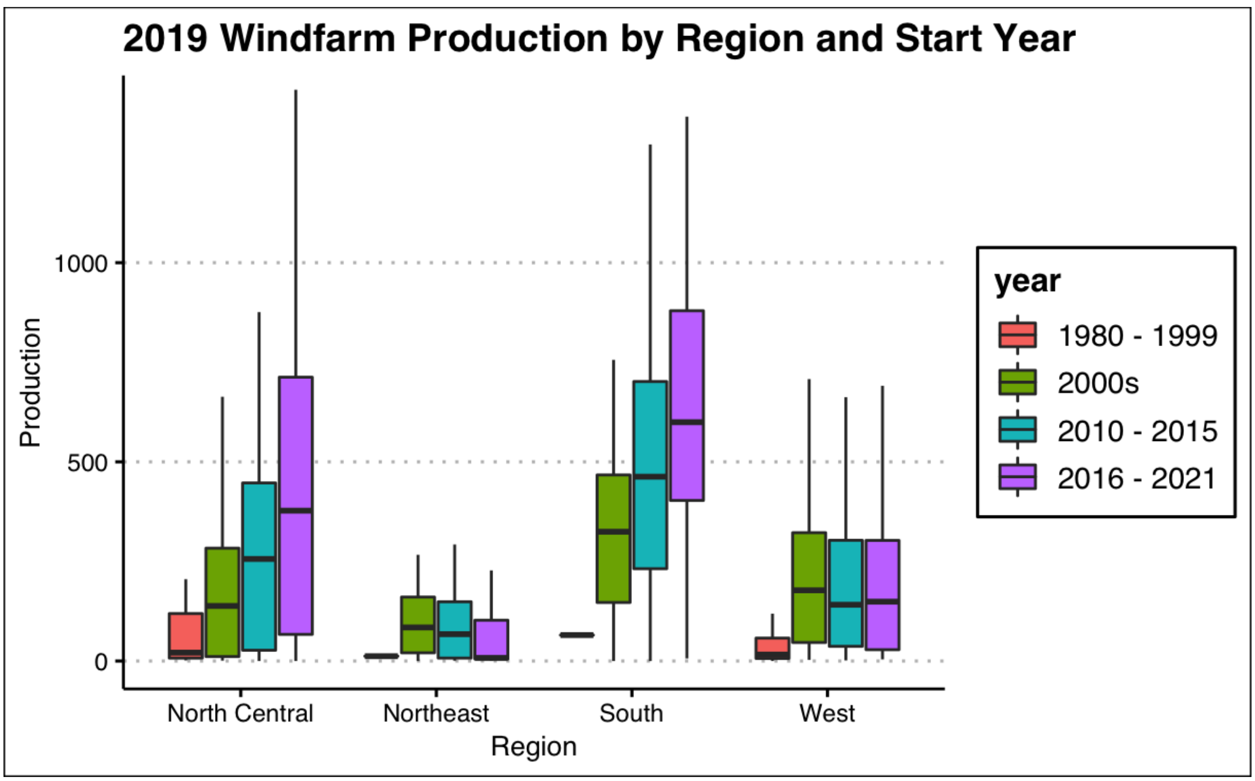
After comparing the trends of California and Texas wind farm productions, we want to investigate how wind farm production is related to geographical location of the wind farms and other factors.

To analyze the production from each wind farm, we aggregate the monthly Megawatt production for each wind farm into yearly Gigawatt production, and use the annual production data in 2019 as it includes the most wind farms. First, we plot a scatterplot of the wind farm production overlayed on top of the map of the wind farm locations.



In the production map, the size of the circle representing each wind farm is proportional to its annual production in 2019. As we can see, the wind farms with high production are mostly in the South, particularly the states of Texas, Oklahoma, and Kansas. Some wind farms in some northern states also have relatively high productions, namely Iowa and Illinois. In comparison, although there are many wind farms in the West (Washington, California) and the Northeast, in general they have lower productions. In comparison to the wind farm heat map from earlier, where there are a higher number of wind farms in Southern California and Northern Texas, we can see in this production map, the wind farms in Texas have much higher production than the ones in California, which is also supported by the analysis earlier on the monthly production trend comparison.

Aside from the state and location of the wind farms, there may be other factors that affect wind farm productions. Particularly, we look at the variable “Year Operational”, which is the year that the turbines in the wind farm became operational and began providing power. We produce side-by-side boxplots showing the distribution of 2019 wind farm production conditioned on the location of the wind farm and the year it became operational. Instead of using the longitude and latitude, we put each wind farm location into 4 regions (North Central, Northeast, South, West). We also grouped the “Year Operational” variable into 4 categories (1980-1999, 2000s, 2010-2015, after 2016) with roughly equal number of wind farms. Even though each category has a similar number of wind farms, the range of each category is decreasing (20 years, 10 years, 6 years, 4 years), indicating that the number of new wind farms that began providing power each year is increasing.



The side-by-side boxplots confirm that Southern and North Central states have higher productions in 2019 than Western and Northeastern states. For Southern and North Central states, wind farms that became operational in later years have higher productions than wind farms that began operations earlier. This does not seem to be the case for Western and Northeastern states, as wind farms with different starting years seem to have a similar distribution of production. This also provides an explanation for the difference in trend in California and Texas wind farm productions. Although each year there are new wind farms that start providing power, the newer wind farms in Texas are more efficient and provide more power, while the newer wind farms in California provide similar power than older wind farms, therefore the increase in average production is slightly faster for Texas wind farms compared to California wind farms.

Turbine Retrofitting

A turbine is considered partially retrofitted if it has undergone rotor and/or nacelle replacement or installation of other new equipment. The common goal of retrofitting is to extend the lifespan of the turbine and improve efficiency and energy production if the turbine is fitted with a larger rotor diameter. In this section of the report, we are interested in turbines that have indicated to be retrofitted to answer the following questions:

1. Are there patterns in the locations and energy production trends of turbines that have been retrofitted?

2. Are turbines being retrofitted with equipment different from turbines that have not been retrofitted that would allow for increased energy production beyond the capabilities of their original construction? Can we observe this difference by clustering turbines by attributes such as rotor diameter, sweep area, hub height, turbine height?

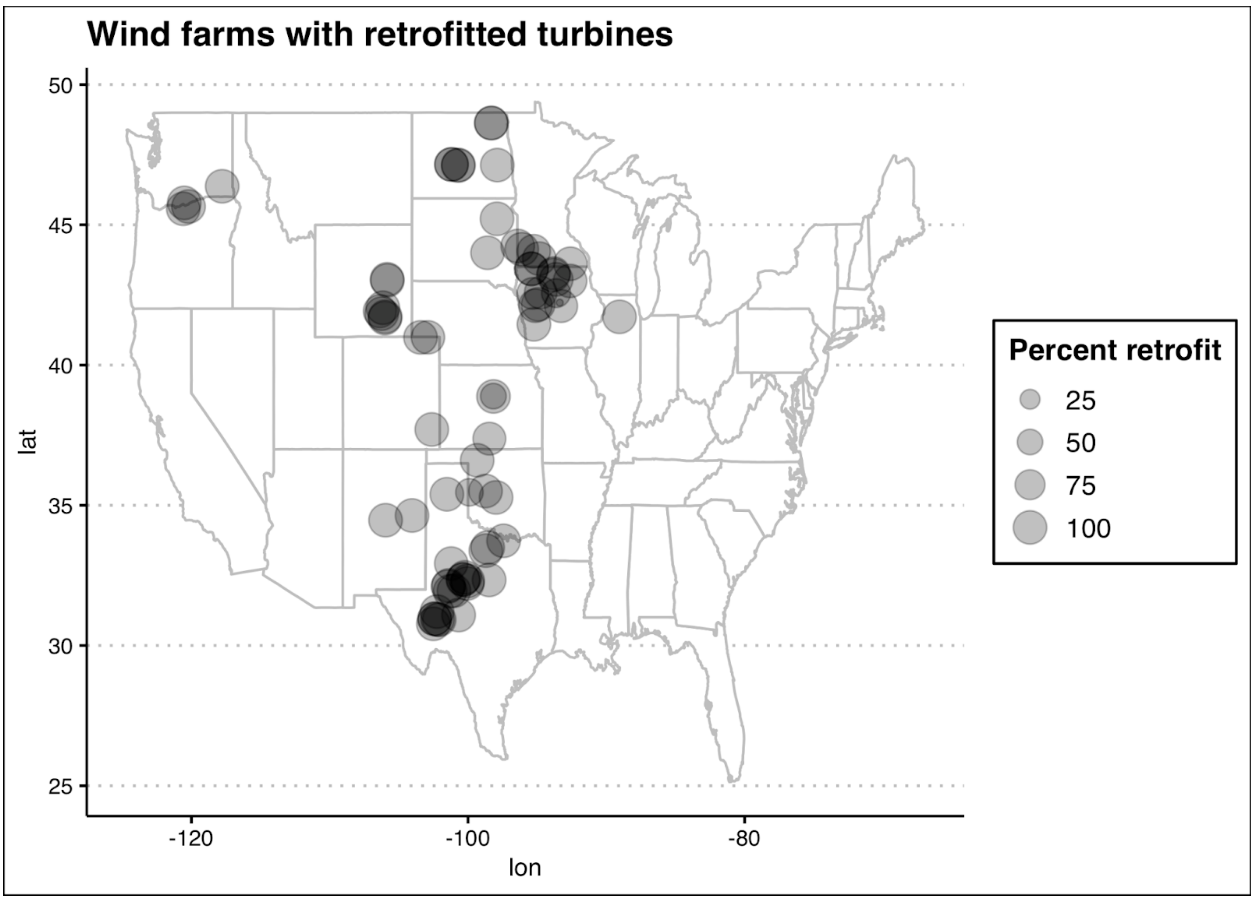
	Number of Turbines
Partially Retrofitted	5983
Never Retrofitted	63119
Total Turbines	69102

Around 8% of the turbines in the U.S. Wind Turbine Database are known to have been partially retrofitted after the turbine was constructed by the manufacturer.

We investigate the effects on energy production of retrofitting turbines. Specifically, we are interested in whether after retrofitting, energy production increases.

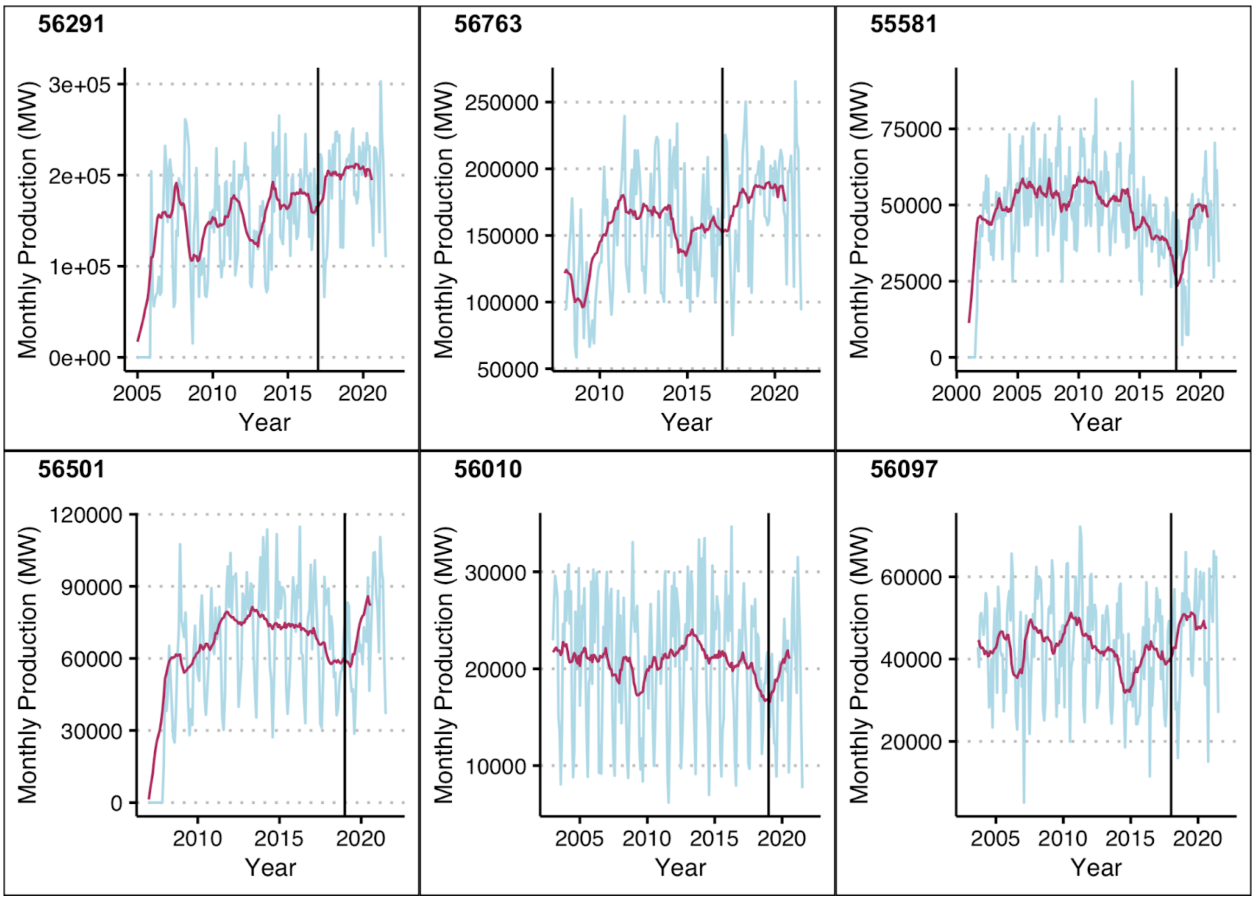
	Number of Wind Farms
Retrofitted Turbines	75
No Retrofitted Turbines	1092
Total Wind Farms	1167

Out of 1167 wind farms from the dataset, 75 wind farms have retrofitted turbines.



From the map above, we see the geographical locations of these 75 wind farms. Nearly all wind farms with retrofitted turbines retrofitted 100% of the turbines on their farm. The farms are generally located along the middle of the continental United States with majority clustered in West Texas and Iowa. This is consistent with prior research on wind energy production stating that the Midwest/Great Plains region along with Texas captures the majority of investment in wind energy throughout the past few years. If retrofitting turbines indeed increases energy production, perhaps this fact can explain a portion of why megawatt production from wind farms in Texas are consistently greater than those in California.

To investigate changes in energy production pre and post retrofit, we chose to visualize the energy production of the top six wind farms with the most retrofitted turbines. A vertical line indicates the first day of the year which turbines on the farm were retrofitted.

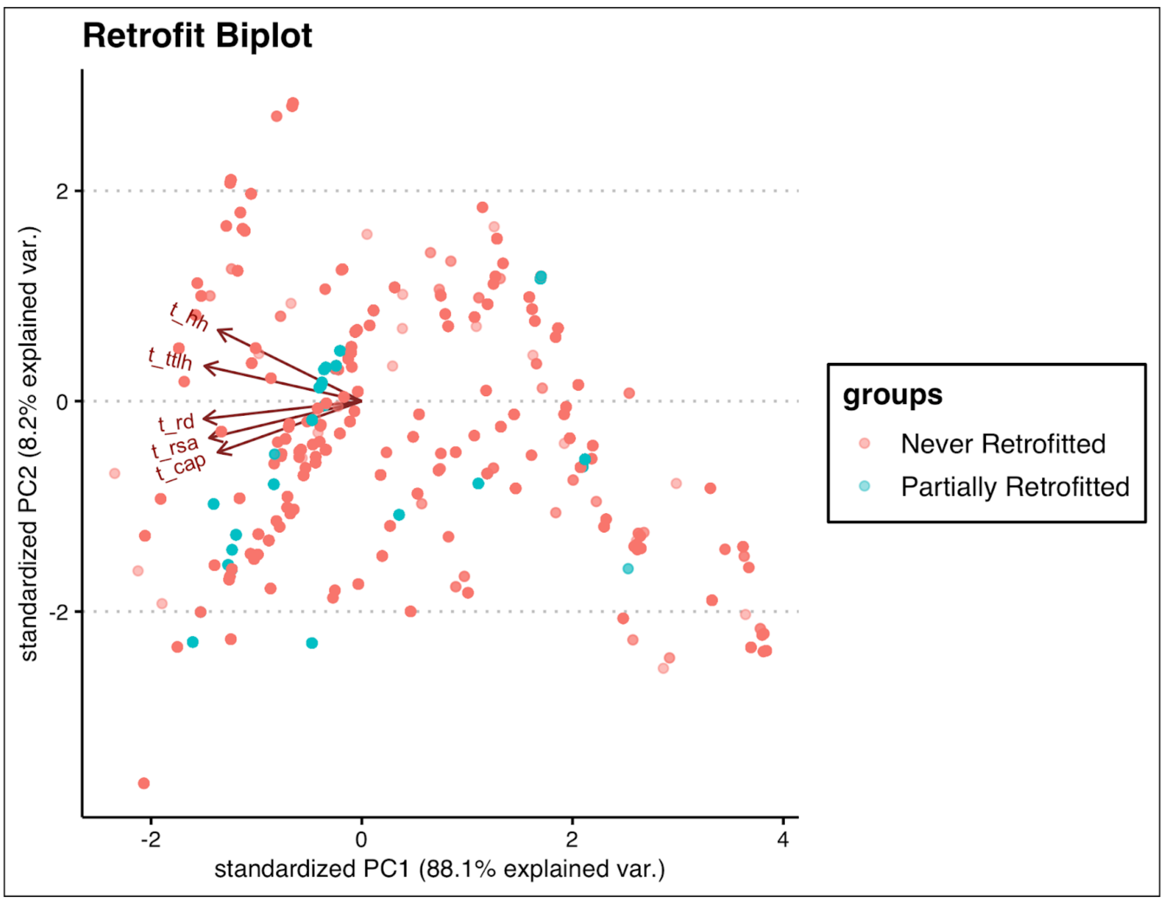


As shown by the dark line displaying a rolling average of monthly energy production, retrofit occurs after a period of decline in production and production post retrofit increases previous levels. Notice Farm 55581 as it shows a very complete story of turbine energy production lifespan. For around 15 years, energy production is relatively stable, then begins to decrease steeply until the year of retrofit when production climbs back to previous stable levels.

We conclude that retrofit happens after a period of decreasing energy production and the goal is to increase production back to a stable state thereby prolonging the lifespan of a turbine. Now we investigate whether retrofitted turbines are being fitted with larger equipment than original constructed parts by performing a principal component analysis on the quantitative variables of turbine physical attributes and energy production capacity to attempt to cluster retrofitted and never retrofitted turbines.

Clustering turbines by physical attributes

Now we explore the physical characteristics of the retrofitted and never retrofitted turbines, specifically the quantitative variables documenting energy production capacity (t_cap), hub height (t_hh), rotor diameter (t_rd), rotor sweep area (t_rsa), and turbine total height (t_ttlh). We are interested to see if retrofitted turbines are being fitted with replacements different from never retrofitted turbines that would allow for increased energy production such as larger rotors.



From the PCA plot of the first two principal components that together explain over 95% of the variation in the physical characteristics of turbines, we notice that retrofitted turbines and never retrofitted turbines are not well clustered with these principal components. Specifically, the majority of partially retrofitted turbines appear to fall along an increasing line against the vectors of coefficients of the variables. From the coefficient vectors, we see the variables are highly correlated with each other, especially turbine capacity (t_cap) with rotor sweep area (t_rsa) and rotor diameter. We also notice that the majority of retrofitted turbines have higher values of height, rotor measurements and capacity. This indicates that turbines with higher energy output capacity have larger rotors and are also taller as capacity is also highly correlated with turbine total height (t_ttlh) and turbine hub height (t_hh).

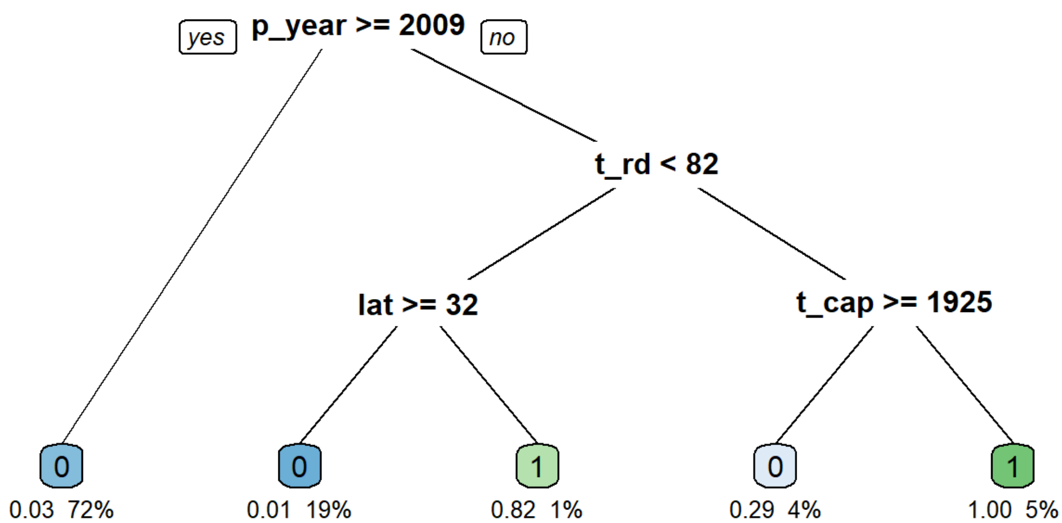
Since retrofit is not well clustered from our principal component analysis of turbine physical attributes, we suspect that retrofitted turbines' physical attributes are not very different from those turbines which have not been retrofitted. We interpret this to potentially indicate that turbines are not being retrofitted with equipment which are different from their original parts and that there may be other factors that better cluster retrofit such as the age of the turbine.

Using Decision Tree Classification to Predict Retrofit

Since our retrofit data presents difficulty with clustering, we want to use another method of determining if there is a difference between retrofitted and non-retrofitted data. To predict whether a turbine is retrofitted or not, we can construct a binary decision tree classifier, which will determine retrofit based on quantitative data such as turbine properties and turbine locations.

Our constructed decision tree using randomly sampled training data is as follows:

Predicting Retrofit using Turbine Data and Location



Note: Our constructed decision tree uses a maximum depth of 3. This hyperparameter is fine-tuned to avoid overfitting our data.

This graph shows us that the year a turbine became operational determines retrofit for a majority of the data, where turbines becoming operational past 2009 (p_year >= 2009) are classified to be not retrofitted, which accounts for 72% of our data. The other three features that our tree classifies on are: turbine rotor diameter (t_rd), latitude of the turbine (lat), and turbine capacity (t_cap).

Now we will check how well this decision tree classifies our testing dataset using a confusion matrix:

```
##
## retro_tree_predict     0     1
##                  0 18862   674
##                  1    49  1165
```

We see that our decision tree is very accurate in predicting retrofit, with only about a 3% error rate. In other words, our classifier will predict turbine retrofitting correctly for about 97% of our training data.

However, we need to note that there are a few limitations using this approach. Firstly, our classifier predicts a majority of retrofitting using the year a turbine became operational. As we collect more data over time, we expect our year "2009" to increase over time, since turbines constructed recently will begin to wear out over time. We can counter this by taking the age of a turbine rather than the year it was constructed. Secondly, our table shows us that we often mispredict retrofitted turbines as non-retrofitted, with a high error rate of about 1/3. This suggests that our classifier does a good job of predicting non-retrofitted turbines, but struggles with predicting retrofitted turbines.

Conclusion

This project examined the location, elevation, and megawatt production of wind turbines and wind farms in the contiguous United States. It was concluded that the majority of wind turbines are located in the states California and Texas, at elevations of peak density closely matching the elevation of Pittsburgh (372.8 m).

An analysis of megawatt production showed Texas has a consistently higher megawatt production than California.

The power production of a wind farm is correlated to its geographical location. In particular, we find that wind farms in the South, namely Texas, Oklahoma, and Kansas have the highest production, followed by wind farms in North Central states. In comparison, wind farms in the West and the Northeast have lower production. In addition to the location, the production of the wind farm is also related to the year it became operational. Notably, in Southern and North Central states, the newer wind farms in general have higher production.

Retrofitting wind turbines is known to be a way to extend the lifespan and increase energy production of the turbine. From our analyses, we conclude that retrofitted turbines are located in the north central and south central United States with the majority being in western Texas and Iowa. Additionally, turbines are commonly retrofitted during a period of energy production decline, and post retrofit energy production does climb back to previously stable levels. However, considering it is well known that physical characteristics such as rotor diameter/sweep area and height are important factors determining turbine energy production, principal component analysis on energy production capacity (t_cap), hub height (t_hh), rotor diameter (t_rd), rotor sweep area (t_rsa), and turbine total height (t_ttlh) did not cluster retrofit very well. We interpret this as retrofitted turbines are not being fitted with new equipment such as larger rotors or hubs that distinguish them from turbines that have not been retrofitted. Instead there may be other factors that better cluster retrofit such as turbine age. Perhaps this indicates that retrofitting is primarily to revive turbines to their newly constructed state rather than attempt to improve upon their original construction, or other variables in the turbine's construction could prevent retrofitting with larger equipment that would allow for increased capacity for energy production from their original construction.

Further Research

Further investigation into whether topography, which could influence accessibility of turbines for inspection and maintenance, could play a role in turbine farm location would be of additional interest, but would require additional data outside the scope of this project.

In regards to the relationship between wind farm production and the location of the wind farm, an interesting direction for further research is to investigate whether the relationship is causal, as in if all the other attributes of the turbines and the wind farm are the same, the location of the turbines would affect the production of the wind farms.

There are many questions on retrofitted wind turbines that investigating this one variable can expand to its own research project. Our data capture the turbines in a single moment of time which makes analyzing pre and post retrofit beyond the monthly energy production difficult. If annual inspection records exist with data on turbines physical characteristics, we would like to compare the same attributes used in our PCA analysis to explicitly investigate the changes after retrofit. With further data on the retrofitted components, we may investigate the monetary cost and energy production return of retrofit, whether it is an investment more wind farms beyond the Midwest/Great Plains and Texas regions should consider to improve their turbine efficiency and lifespan.

Additionally, by combining turbine location with results that retrofitting increases energy production after a period of production decrease but do not seem to be fitted with larger equipment than original parts, further research in how geographical characteristics effect turbine maintenance, inspection, and retrofit can help us learn more about maximizing turbine lifespan and energy production perhaps could be used to justify more investments in wind turbines.

Citations

Energy Information Administration (EIA) {https://www.eia.gov/}
United States Geological Survey (USGS) {https://www.usgs.gov/}
Jossi, Frank. “Industry Report: Midwest and Great Plains Lead Wind Energy Expansion.” Energy News Network, 19 Apr. 2017, https://energynews.us/2017/04/19/industry-report-midwest-and-great-plains-lead-wind-energy-expansion/.