CSE 6242 - Final Report

Team 85: Min Kook Cho, Rithikaa Madhavan, Aaron Na, Jason Tourville, Carmen Yu

Introduction

Climate change is one of the major issues dominating headlines and policy discussions. It is estimated as a major factor contributing to higher energy demands (Sofian et al., 2004). As the world continues to experience extreme climates, research estimates that by 2050, baseline global energy demands will be two to three times larger than in 2010 (Ruijeven et al., 2019). Some estimates reveal this value to be nearly 700 QBtu (Raimi et al., 2023). Hot, humid, and tropical climates will see more impact, (Fonseca et al., 2020), experiencing up to 50 hotter days than normal (Ruijeven et al., 2019).

Notably, climate changes will be further exacerbated by the growing population, especially in the United States. Cohen et. al (2010) found a feedback loop between population and climate. In addition, the US population has a history of growing especially near coasts, which is vulnerable to natural disasters. (Cohen et. al, 2010). Hence it is important to recognize the relation between climate change, regional energy demand, and population growth.

Problem Definition

The main goal is to analyze climate change and its effects on energy demand and how renewable energy could address that demand. Historical temperature data across the US will be analyzed to determine future temperature patterns. This will be translated into population-weighted degree days to measure future energy demand. This demand will be forecasted for up to 2040 and used to propose state-wide renewable energy implementations.

Literature Survey

As energy is essential in daily activities, it is important for region-based allocation. Though renewables are a big source of interest for energy due to climate change, it remains a small fraction of overall energy resources. Its increasing feasibility given the rising fossil fuel costs, declining renewable energy costs, and policy changes are in favor of renewable energy (Timmons et al, 2014). It is predicted that it can meet two-thirds of US energy demand but would require large investments to accelerate the implementation. (Gielen et al., 2019)

Initial research on common renewable energies considers several sources. Case studies and data on solar energy detail solar panel efficacy and the effect of tilt angles on energy production (Hanif, 2012). In accordance, models such as the multi-input support regression (Nageem, 2017) will be referenced to accurately forecast solar power creation. For biofuel, to benefit from CO2 emission reduction to offset the transportation industry, areas to consider are energy source placement, land amount required for power plants, and more (Prasad et al, 2013).

Geothermal and ocean energy are two other promising renewable energy sources. Geothermal taps into Earth's natural heat reservoirs, harnessing steam and hot water beneath the surface to drive turbines to produce electricity. Direct usage and geothermal energy heat pumps have both increased over the past 25 years by ten folds (Lund, 2021). It is forecasted to be 50,000 TWh by 2050, up from 26,800 TWh in 2020 (RV, 2023). Ocean energy utilizes tidal turbines, wave energy converters, and ocean thermal energy conversion systems to capture energy from wave motions, tides, and temperature and salinity gradients (Melikoglu, 2018).

Previous researchers, such as (Sullivan et al., 2015) model energy demand through electric loads and temperatures. They found clear correlations between temperature deviations to energy demand. Additional studies and predictions reliably use hourly to annual meteorological data from high-density stations, long-term historical station observation data, population data, and more (Li et al, 2020). Other approaches such as Bayesian model averaging or Gaussian process regression ensure that prediction data are foundational and accurate (Mansfield et al 2020). Hence, climate data can confidently predict energy demand via climate change trends.

Methodology

Intuition and Innovations

We used temperature and population data from each US state to estimate energy demands by calculating degree days based on the expected temperature degree change anticipated by the year 2040. From there, we would connect the energy demand to visualize the scale of renewable energy implementations that would be required to match that demand. The project potentially pays off by informing key stakeholders and the government about the future of energy sources, which includes real estate companies, investment firms, etc.

Many models have estimated the future change in climate and energy demands. However, creating an interactive user interface using Tableau will create greater visual appeal, understanding, and promote climate/energy awareness. Our model predictions and interactive user interface will provide filtering abilities to obtain region-specific insights. Stakeholders would view appropriate renewable energy source options, its required number of sources, and plan for energy haul across the US. Since this is an interactive user interface, we would be able to measure user interactions as a potential measure of user reach with this project.

Approach

The main metric used for energy demand in this analysis is degree days, which is a measurement of how warm/cool a location is based on the average temperature throughout the day in comparison to base temperature. According to US Energy Information and Administration (*Degree-days - U.S. Energy Information Administration (EIA)*, n.d.), the base temperature typically used for the US is 65F. Heating degree days (HDD) are how much colder it is than 65F while cooling degree days (CDD) are how much hotter it is than 65F. For example, an average daily temperature of 40F is 25 HDD, but an average daily temperature of 80F is 15 CDD. In addition, Kennard et al. (2022) suggests considering population in forecasting values, and that calculating population weighted degree days provides a broader scope on energy distribution.For this analysis, the temperature data gathered will forecast the temperature in 2040 through time series models for each state. This will provide the temperatures to be used in calculating the yearly total HDD and CDD. Individual states will be weighted by the US Census Regions' populations. This captures a larger area of potentially similar states.

To represent possible renewable energy implementation, solar and wind energy was chosen to represent two possible energy sources, as other sources such as hydroelectricity and geothermal require further geological investigation on how probable those sources of energy could be established in each state. While this is also applicable to solar and wind, these two energy sources have more flexibility in being implemented. For solar, the forecasted temperatures will be used to estimate the efficiency of its energy output, as Dubey et al. (2013) noted that increasing temperatures have an effect on how effective solar panels will work in generating power. As for wind, this will be estimated using the average capacity of a wind turbine. This along with the estimated energy demands will allow us to visualize the scale of possible renewable energy required to be implemented to meet the future energy demand.

Data Collection/Extraction

Temperature data was derived from the National Centers for Environmental Information website. To align with the population dataset timeframe, we downloaded 2010-2024 US data. A JavaScript library Puppeteer was utilized to control a Chromium browser. It iterated over the fifty states to download the relevant csv files while selecting appropriate data columns. A written python script casted files into a Pandas dataframe and then into a cleaned csv file per state. The resulting files were then inserted into DataGrip for storage, manipulations, and export.

Data Storage & Warehousing

In DataGrip, a stored procedure was created to speed up the time process of cleaning the data and create individual state tables. These state tables consist of monthly max temperature values for years 2010-2024. After all fifty state tables were created, these tables were joined to start the process of initial modeling.

Modeling

For the progress report, initial modeling intended to produce temporary results for visualization and to identify potential data difficulties and forecasting values. The data was split into a 70% training and validation set, and a 30% test set as values from the later years. Only the training set has been used for initial modeling thus far, as the test set will be used as part of the evaluations on the performance of the final models.

For each state, the ARIMA model used monthly temperatures for time series analysis and forecasting values into 2040. Hyperparameter tuning was carried out with the training and validation set through randomized search to increase the range of hyperparameters without causing a large runtime. Three values from the range [0,5] were selected for p and q, while d used a range of [0,2]. Mean-squared error (MSE) was calculated along with Akaike Information Criterion (AIC) to evaluate training performance and identify potential problem states. The optimal hyperparameter found from randomized search was then used to fit the model and forecast temperatures up to the year 2040. This was then transformed into a dataset of unweighted degree days that contains the State, Year, CDD and HDD. MSE, AIC, and tuned hyperparameter values are in the Appendix.

In addition, it was important to incorporate population as heavily populated states will have a heavier impact as more people means more energy used. Using the population data, population ratio was calculated for each state relative to the total US population spanning from 2010-2040. Population ratio was then used to weight the degree day values of each state accordingly.

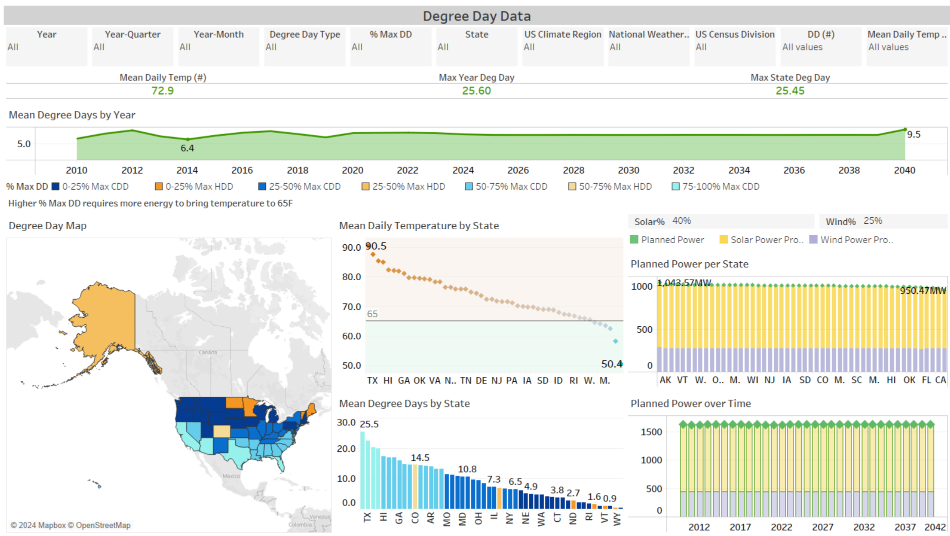
Renewables Estimation

Using the degree day predictions along with energy consumption estimates from the [US Energy Information Administration](), energy consumed in Btu for each state for each month was estimated. This was translated to the power needed for each state, assuming a duration of 365 days a year, for the full 24 hours a day. While this an idealistic assumption, this is useful as it gives the minimal basis of power needed if everything is running smoothly, serving as a useful baseline to expand upon for non-ideal conditions.

Solar is then estimated using a standard 5MW farm, which is noted by the US Energy Information Administration (2019) as the most common size solar farm capacity in the US. Since temperature affects the efficiency of the panels (Dubey et al., 2013), the efficiency of the power capacity of the solar farm was scaled to the temperatures that were forecasted by the models. This was done by using a base temperature of 77F and utilizing a temperature coefficient (Boston Solar, n.d.), denoting the %/F that the panel loses efficiency for every F above 77F. This way, the solar power output is tailored to the temperatures estimated. As for wind, a 3MW turbine was chosen as it was noted to be around the average capacity installed in the US in 2021-2022.

Tableau Visualization

The created visualization lets users filter on the aforementioned fields and some US Region categories. The charts show a choropleth map based on the DD, showing which states require more energy to bring the temperature back to 65F. Other charts show the Mean Daily Temperature (MDT), and Mean DD by Year and State. For the MDD by Year and MDT, the axis does not include 0 for better pattern display. Finally, column charts show Solar and Wind power usage by Year and State based on user parameter sliders. Each component can also be used as a filter object. The MDT measure uses a diverging palette with 65F at the center. For map categorical colors, CDD uses blue and HDD uses orange. Higher powered renewable energy plants would be placed in lighter colored regions. The public dashboard is linked in the Appendix.

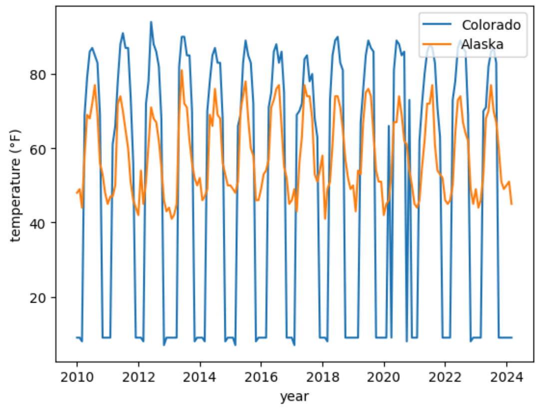


Experiments/Evaluation

Model Performance Evaluation

This was done via using a 70% train and validation to 30% test split to verify confidence in model deployment and hyperparameter tuning. Metrics such as mean squared error (MSE) and Akaike Information Criterion (AIC) assessed the performance of the ARIMA models of each state. These metrics were chosen as MSE represents the average error of the forecasts, and AIC for quality of fit and model complexity. The goal was to have lower MSE and AIC in the hyperparameter tuning and similar magnitudes of training MSE and AIC values across all states. Model robustness for future deployment would be further confirmed by running a test set on the tuned model and using MSE to check that it returns in similar magnitudes. A table with all values returned for every state is in the Appendix.

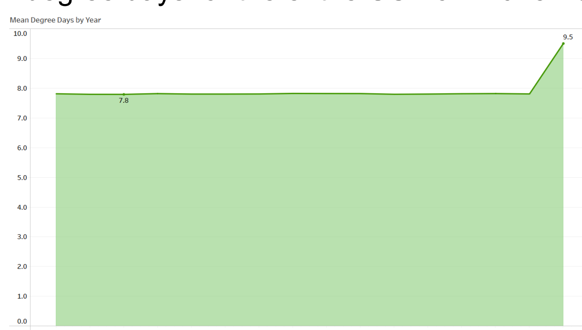
The AIC and MSE values for all states are similar magnitudes. However, there is one exception: Colorado. The MSE is about 2 magnitudes higher than the rest of the MSE values for every state and has the highest AIC value across the models. Upon further investigation, it was found that the data for Colorado had consistently recorded incredibly low temperatures for its winter months and displayed some inconsistent patterns. The following plot below plots the temperatures recorded for Colorado in comparison to Alaska, the northernmost state.



As seen in the plot, Colorado temperatures have a larger temperature range. In 2020-2021 there are also some inconsistent temperature patterns. Steps were taken to re-evaluate whether this was a problem with data collection and processing. Despite efforts to re-scrape the data in hopes that it was a scraping error, the results were the same. This is hypothesized that there is some issue with the data itself, such as an instrumentation error that was never caught. Imputation was considered but this was behavior consistent almost every year. For the future, more work and investigation should be done on alternative datasets of Colorado's temperature for comparison, and potentially a different model for it.

Visualization Evaluation

Utilizing Tableau, the forecasted values were plotted and visualized to see the impact over the years and if any additional patterns and trends would appear. The following plot below shows the projected mean degree days for the entire US from 2025-2040.



As seen in the chart, it appears flat across the year with an odd spike at the end of 2040. While it indicates consistent predictions and energy demands, it does not capture the expectation of rising energy demand as predicted in the literature survey. More research in time-series models indicated difficulties with long-term forecasting and is currently a large topic of research. For example, Wang et al. (2023) researches on potentially improving an ARIMA model for an ultra long time series, which could be investigated in the future to improve this model. Another problem may be that the training data from 2010-2024 is a slightly smaller interval than 2025-2040, and is only one time period of representation.

To visualize the impact of renewable energy implementation, a slider option was added to see how each source could contribute to the energy demand of each state which gave a stronger impact of scale. The sliders represent the percentages of energy demand that the renewable energy source is meant to generate. While it is common sentiment that "more renewables are needed", this gives an idea of how much expansion is really needed for renewables to become a more significant percentage of energy generation.

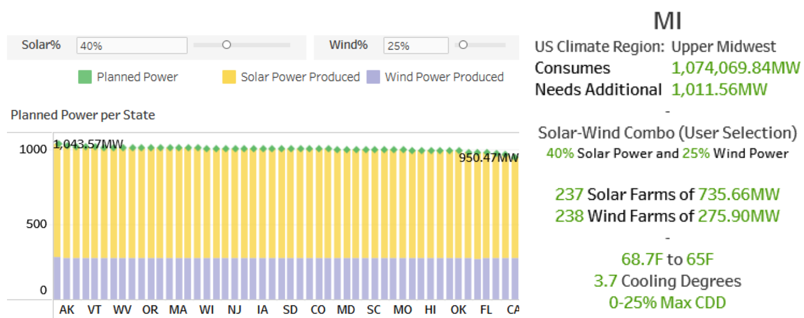


Tableau and User Reach

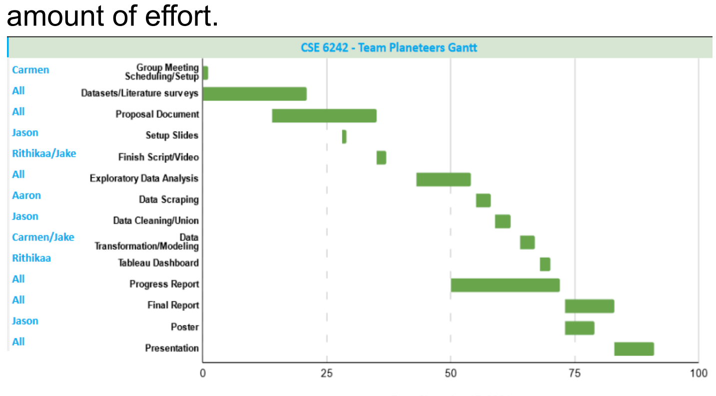
As Tableau Public exists as a public interface, user interactivity can be tracked through Tableau Viz views to measure the outreach of the project's potential impact. As more stakeholders use the sharing option, the view count will increase, allowing for public visibility and awareness. Additionally, technical users and interested users will have the ability to download the workbook as it is updated with the latest data to customize it to their needs, further employing reusability and reproducibility. Finally, this allows for users to utilize the current project for any future work and extensions to further expand the scope of the analysis.

Conclusion and Results

With temperature and climate changes, global warming remains a prevalent issue, resulting in higher global energy demand. Hence, this project undertook projecting energy demand for the US up to the year 2040. For this, data for climate and population were scraped using JavaScript's Puppeteer and transformed in DataGrip through SQL. Using ARIMA's time-series model, the resulting data predicted temperatures which were then used to create monthly population-weighted degree days per state. Data for renewable wind and solar were integrated to obtain power consumption and renewable energy production. This allowed Tableau to display granular insights for stakeholders such as the Government, climate activists, businesses, and the general public. Through this end-to-end process, users can simulate and visualize the scale of renewable plants required and its associated power needed to run for each state. Future work could delve into renewables investment and environmental infrastructure to better plan for specific use-cases such as addressing how the US population has a history of growing especially near coasts. To enhance and expand on the project scope, more renewable energy plants such as hydropower and geothermal data can be added. Nuclear energy could also be considered.

Plan of Activities and Gantt

Activities were carried out as planned in the progress report and all team members contributed a similar amount of effort.



References

Boston Solar. (n.d.). *How temperature & shade affect solar panel efficiency | Boston Solar*. https://www.bostonsolar.us/solar-blog-resource-center/blog/how-do-temperature-and-shade-affect-solar-panel-efficiency/

Cohen, J. E. (2010). Population and Climate Change. *Proceedings of the American Philosophical Society*, *154*(2), 158–182. https://www.jstor.org/stable/pdf/41000096.pdf?refreqid=fastly-default%3A040a1dddabb1432e335233e56b721a68&ab_segments=&origin=&initiator=&acceptTC=1

Degree-days - U.S. Energy Information Administration (EIA). (n.d.). https://www.eia.gov/energyexplained/units-and-calculators/degree-days.php

Dubey, S., Sarvaiya, J. N., & Seshadri, B. (2013). Temperature Dependent Photovoltaic (PV) Efficiency and Its Effect on PV Production in the World – A Review. *Energy Procedia*, *33*, 311–321. https://doi.org/10.1016/j.egypro.2013.05.072

Fonseca, J. A., Nevat, I., & Peters, G. W. (2020). Quantifying the uncertain effects of climate change on building energy consumption across the United States. *Applied Energy*, *277*, 115556. https://doi.org/10.1016/j.apenergy.2020.115556

Gielen, D., & Boshell, F. (2019). The Role of Renewable Energy in the Global Energy Transformation. *Energy Strategy Reviews*, *24*(24), 38–50. ScienceDirect. https://doi.org/10.1016/j.esr.2019.01.006

Li, Y., Wang, W., Wang, Y., Xin, Y., He, T., & Zhao, G. (2021). A Review of Studies Involving the Effects of Climate Change on the Energy Consumption for Building Heating and Cooling. *International Journal of Environmental Research and Public Health*, *18*(1), 40. https://doi.org/10.3390/ijerph18010040

Lund, J. W., & Toth, A. N. (2020). Direct utilization of geothermal energy 2020 worldwide review. *Geothermics*, *90*, 101915. https://doi.org/10.1016/j.geothermics.2020.101915

Kennard, H., Oreszczyn, T., Mistry, M., & Hamilton, I. (2022). Population-weighted degree-days: The global shift between heating and cooling. *Energy and Buildings*, *271*, 112315. https://doi.org/10.1016/j.enbuild.2022.112315

Mansfield, L. A., Nowack, P. J., Kasoar, M., Everitt, R. G., Collins, W. J., & Voulgarakis, A. (2020). Predicting global patterns of long-term climate change from short-term simulations using machine learning. *Npj Climate and Atmospheric Science*, *3*(1). https://doi.org/10.1038/s41612-020-00148-5

Melikoglu, M. (2018). Current status and future of ocean energy sources: A global review. *Ocean Engineering*, *148*, 563–573. https://doi.org/10.1016/j.oceaneng.2017.11.045

Nageem, R., & R, J. (2017). Predicting the Power Output of a Grid-Connected Solar Panel Using Multi-Input Support Vector Regression. *Procedia Computer Science*, *115*, 723–730. https://doi.org/10.1016/j.procs.2017.09.143

Office of Energy Efficiency & Renewable Energy. (2023, August 24). *Wind Turbines: the Bigger, the Better*. Energy.gov. https://www.energy.gov/eere/articles/wind-turbines-bigger-better

Prasad, S., Kumar, A., & Muralikrishna, K. S. (2014). Biofuels production: A sustainable solution to combat climate change. *Indian Journal of Agricultural Sciences*, *84*(12). https://doi.org/10.56093/ijas.v84i12.45210

Raimi, D., Zhu, Y., Newell, R., Prest, B., & Bergman, A. (2023). *Global Energy Outlook 2023: Sowing the Seeds of an Energy Transition A Global Energy Outlook 2023: Sowing the Seeds of an Energy Transition*. https://media.rff.org/documents/Report_23-02_NPKKYws.pdf

Rohit R.V, Vipin Raj R, Dennis Cheruiyot Kiplangat, R. Veena, Jose, R., A.P. Pradeepkumar, & K. Satheesh Kumar. (2023). Tracing the evolution and charting the future of geothermal energy research and development. *Renewable & Sustainable Energy Reviews*, *184*, 113531–113531. https://doi.org/10.1016/j.rser.2023.113531

Sofian, A., Hooi Ren Lim, Heli Siti Halimatul Munawaroh, Ma, Z., Kit Wayne Chew, & Pau Loke Show. (2024). Machine learning and the renewable energy revolution: Exploring solar and wind energy solutions for a sustainable future including innovations in energy storage. *Sustainable Development*. https://doi.org/10.1002/sd.2885

Sullivan, P., Colman, J., & Kalendra, E. (2015). *Predicting the Response of Electricity Load to Climate Change*. https://www.nrel.gov/docs/fy15osti/64297.pdf

Timmons, D., Harris, J., & Roach, B. (2014). *The Economics of Renewable Energy A GDAE Teaching Module on Social and Environmental Issues in Economics*. https://www.bu.edu/eci/files/2019/06/RenewableEnergyEcon.pdf

U.S. Energy Information Administration. (2019, February 7). *Most U.S. utility-scale solar photovoltaic power plants are 5 megawatts or smaller - U.S. Energy Information Administration (EIA)*. https://www.eia.gov/todayinenergy/detail.php?id=38272#

van Ruijven, B. J., De Cian, E., & Sue Wing, I. (2019). Amplification of future energy demand growth due to climate change. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10399-3

Wang, X., Kang, Y., Hyndman, R. J., & Li, F. (2023). Distributed ARIMA models for ultra-long time series. *International Journal of Forecasting*, *39*(3), 1163–1184. https://doi.org/10.1016/j.ijforecast.2022.05.001

Appendix

Tableau Dashboard:
https://public.tableau.com/app/profile/rithikaa.madhavan/viz/DegreeDayDashboardFinal2010-2040/DegreeDayData?publish=yes

MSEs, AIC and Hyperparameters for each state ARIMA model:

State	Training_MSE	AIC	Test_MSE	Hyperparameters
AL	7.616766	478.416277	95.869174	(3, 1, 4)
AK	23.112637	582.267581	12.643878	(3, 0, 4)
AZ	11.417682	545.840411	44.567661	(3, 0, 1)
AR	13.201584	518.507035	12.080258	(2, 0, 3)
CA	13.865810	536.144536	20.620720	(2, 1, 3)
CO	403.559586	794.592858	336.984996	(2, 0, 4)
CT	21.965142	571.186851	23.876347	(2, 1, 4)
DE	12.637259	574.987496	21.100761	(3, 1, 4)
FL	2.867738	401.492697	3.287491	(2, 1, 3)
GA	7.797121	477.621095	8.360045	(2, 1, 3)
HI	2.446150	506.256298	1.237264	(2, 1, 3)
ID	16.024201	561.164808	16.569781	(3, 1, 3)
IL	27.708320	568.424985	27.403672	(2, 1, 4)

IN	40.337249	580.948527	17.148027	(2, 0, 3)
IA	38.534635	623.237752	22.758052	(3, 0, 3)
KS	19.848561	575.065797	15.429184	(3, 0, 3)
KY	27.001961	550.663963	10.910564	(2, 1, 4)
LA	5.504818	471.949125	8.455891	(3, 0, 4)
ME	9.948616	590.181228	22.866294	(3, 0, 3)
MD	21.435129	570.933798	18.662854	(2, 0, 4)
MA	22.345412	578.236469	15.580960	(3, 0, 4)
MI	28.523996	604.589558	28.238053	(2, 0, 3)
MN	36.463461	635.088397	41.764265	(2, 0, 3)
MS	12.654731	533.806510	9.503843	(3, 0, 3)
MO	18.662277	551.426386	18.624400	(2, 0, 3)
MT	23.061638	583.727662	19.489038	(3, 0, 3)
NE	21.512660	579.248711	23.015361	(3, 1, 4)
NV	16.520581	525.493323	20.097434	(2, 0, 4)
NH	16.107465	593.488986	21.433453	(3, 1, 3)
NJ	17.830598	575.314085	16.832046	(3, 0, 3)

NM	8.616048	496.566833	10.381158	(2, 1, 3)
NY	18.314313	590.527971	18.709507	(2, 0, 3)
NC	10.190026	523.227723	9.038698	(3, 0, 3)
ND	25.197742	613.672498	27.732883	(2, 0, 3)
OH	33.286729	571.495246	15.121477	(2, 1, 4)
OK	20.645668	552.914892	18.864423	(2, 1, 3)
OR	10.316363	519.368930	27.558733	(2, 0, 3)
PA	25.748477	580.507797	16.631412	(2, 1, 4)
RI	10.363573	580.422977	16.701425	(3, 0, 3)
SC	10.293139	508.102168	8.419033	(3, 1, 4)
SD	22.276263	622.636145	32.816772	(3, 0, 3)
TN	22.292906	528.016180	13.733627	(2, 0, 3)
TX	5.876155	479.850920	10.975897	(3, 1, 3)
UT	20.306926	557.566817	44.795839	(3, 1, 3)
VT	20.613784	614.430193	29.950487	(3, 0, 3)
VA	15.571360	577.828216	13.494902	(2, 0, 4)
WA	12.932174	580.541955	35.089710	(2, 1, 3)

WV	18.735493	540.821682	18.389882	(3, 0, 4)
WI	27.495172	627.717783	34.108792	(3, 0, 3)
WY	11.041220	577.032148	16.311608	(2, 0, 3)