**Public Attitudes Towards Electric Vehicles in the United States**

Sathvika Anugu, Chelsea Bradford, Tim Judge, Akash Loka, Mukul Mokkapati

Data Science, UMBC

DATA602: Introduction to Data Analysis and Machine Learning

Dr. Tony Diana

May 12, 2023

**Abstract**

Electric vehicles are considered by many to be the next frontier of transportation. In recent years, electric vehicles have become more popular, common, and ever present in American society. But how do people feel about electric vehicles? Are they as sought after as manufacturers say they are? Should the United States government invest in electric vehicles? Public sentiment around electric vehicles was evaluated and studied in this report. Sentiment Analysis was conducted through Python to evaluate newspaper articles about electric vehicles and record the public opinion regarding this next wave of technology.

The goal was to prove that public opinion on electric vehicles is driven by gas prices, charging stations, and cost and that these three factors are also crucial to government support for the vehicles as well. The sentiment analysis evaluated regions of the United States in their positive and negative feelings towards electric vehicles and the subjectivity of the articles written, before diving deeper to perform more complex analyses using transformers and models in Python.

**Introduction**

The transportation industry is responsible for a significant amount of greenhouse gas emissions, which are harmful to the environment and the general public's health. Electric vehicles (EVs), a cleaner and more environmentally friendly substitute for conventional gasoline-powered vehicles, have been thought to be a possible solution to address this problem. Although EV adoption is currently at a modest level, consumer behavior is significantly influenced by public opinion. Understanding public opinion on EVs is therefore essential for identifying barriers to and potential for increased adoption. Researchers, decision-makers, and industry stakeholders have paid close attention to the growing uptake of electric cars (EVs). Understanding public attitudes and perceptions of EVs is essential given the rising need for environmentally friendly transportation. Sentiment analysis, a computational method for identifying and extracting views and attitudes from textual data, might offer insightful information regarding the public's perception of EVs. This method can offer insightful information about the factors affecting the public's perception and behavior toward EVs. The purpose of this report is to investigate the sentiment analysis of EVs in the US and provide an answer to the research questions: What is the US public's opinion regarding electric vehicles? What are the influencing factors affecting public opinion on electric vehicles?

Understanding public attitudes towards EVs is crucial for manufacturers and policymakers aiming to promote their use and create more sustainable transportation systems. Numerous studies have examined American attitudes regarding EVs and found that these opinions are influenced by things like environmental concerns, technology perceptions, social norms, and demographic considerations. Several studies indicated that views toward EVs were positively correlated with familiarity with them and environmental concerns, but negatively correlated with fear of distance and a lack of electrical infrastructure for charging. In a similar vein, it was also discovered that while worries about cost and dependability had a negative effect, positive sentiments regarding EVs were related to concepts of energy independence and sustainability. Government programs and focused marketing campaigns can also support efforts to promote the adoption of EVs.

Overall, increasing EV adoption and developing more sustainable transportation networks depend on understanding consumer perceptions of EVs. This paper offers a summary of research looking at American attitudes regarding electric vehicles in variousregions of the United States, stressing the different factors influencing these attitudes.

**Literature Review**

To fully understand what is at stake in evaluating electric vehicle sentiment, research was done to provide a basic understanding of the topic. This question is being asked by many to predict the future of personal vehicles in the United States, and evaluate the validity of electric vehicles as a solution to climate change. The literature reviewed gave a lot of information to consider in developing sentiment analysis.

Over 1,000 Californians were polled as part of another study that was published in the Journal of Transport Geography (Wang et al., 2019) to find out how they felt about electric vehicles. According to the study, attitudes were shown to be positively associated with familiarity with electric cars and environmental concerns, and adversely associated with worries about range anxiety and a lack of infrastructure for charging. The study also found that attitudes varied by demographic factors such as age, income, and education.

Over 1,000 Austin, Texas residents were polled as part of a third study that was published in the International Journal of Sustainable Transportation (Bansal and Kockelman, 2017) to find out how they felt about electric vehicles. According to the study, attitudes were shown to be positively associated with perceptions of energy independence and environmental concerns, and adversely connected with worries about cost and reliability. According to the study, sentiments varied according to demographic parameters like age, income, and education.

These studies demonstrate that a variety of factors, such as environmental concerns, perceptions of technology, social norms, and demographic considerations, have an impact on American attitudes about electric cars. For manufacturers and legislators hoping to encourage the use of electric vehicles and create more sustainable transportation systems, it's critical to comprehend these elements.

Due to their potential to minimize greenhouse gas emissions and dependence on fossil fuels, electric vehicles (EVs) are growing in popularity. Despite these advantages, EV adoption has been slow, and a number of factors, such as attitudes toward electric cars, have been noted as adoption barriers. The authors of a study by Schoenau-Fog et al. (2021) explore how consumers' perceptions of product qualities affect their attitudes regarding EVs. They discovered that factors like range anxiety and the accessibility of charging infrastructure have a big impact on consumers' sentiments about EVs (Schoenau-Fog et al., 2021). According to the report, in order to expand the adoption of EVs, governments and EV producers should concentrate on resolving these issues.

The willingness of American consumers to pay for various forms of alternative fuel vehicles, including electric automobiles, was examined in a study that was published in the journal Transportation Research Part D: Transport and Environment (Hidrue et al., 2011). According to the study, people would be prepared to pay more for electric cars if they believed they would save money on fuel and that it would be better for the environment.

Similar to this research, Yalcinkaya and Alkaya's study from 2021 examines the variables affecting customers' opinions regarding EVs. The authors discovered that customer attitudes about EVs are positively influenced by environmental concerns, perceived benefits of EVs, and government incentives, whereas views are adversely affected by worries about battery life and restricted driving range (Yalcinkaya & Alkaya, 2021). According to the study, government initiatives and targeted marketing campaigns can boost the use of EVs.

Overall, these studies show that a range of factors, such as perceived product features, environmental concerns, governmental legislation, and demographic factors influence customer views about EVs. Increased adoption of EVs can be achieved by addressing issues with range anxiety, the availability of charging infrastructure, and battery life as well as by emphasizing the environmental advantages of EVs through targeted advertising and government incentives. For manufacturers and lawmakers seeking to promote the use of electric vehicles and create more sustainable transportation systems, it's critical that they understand these aspects.

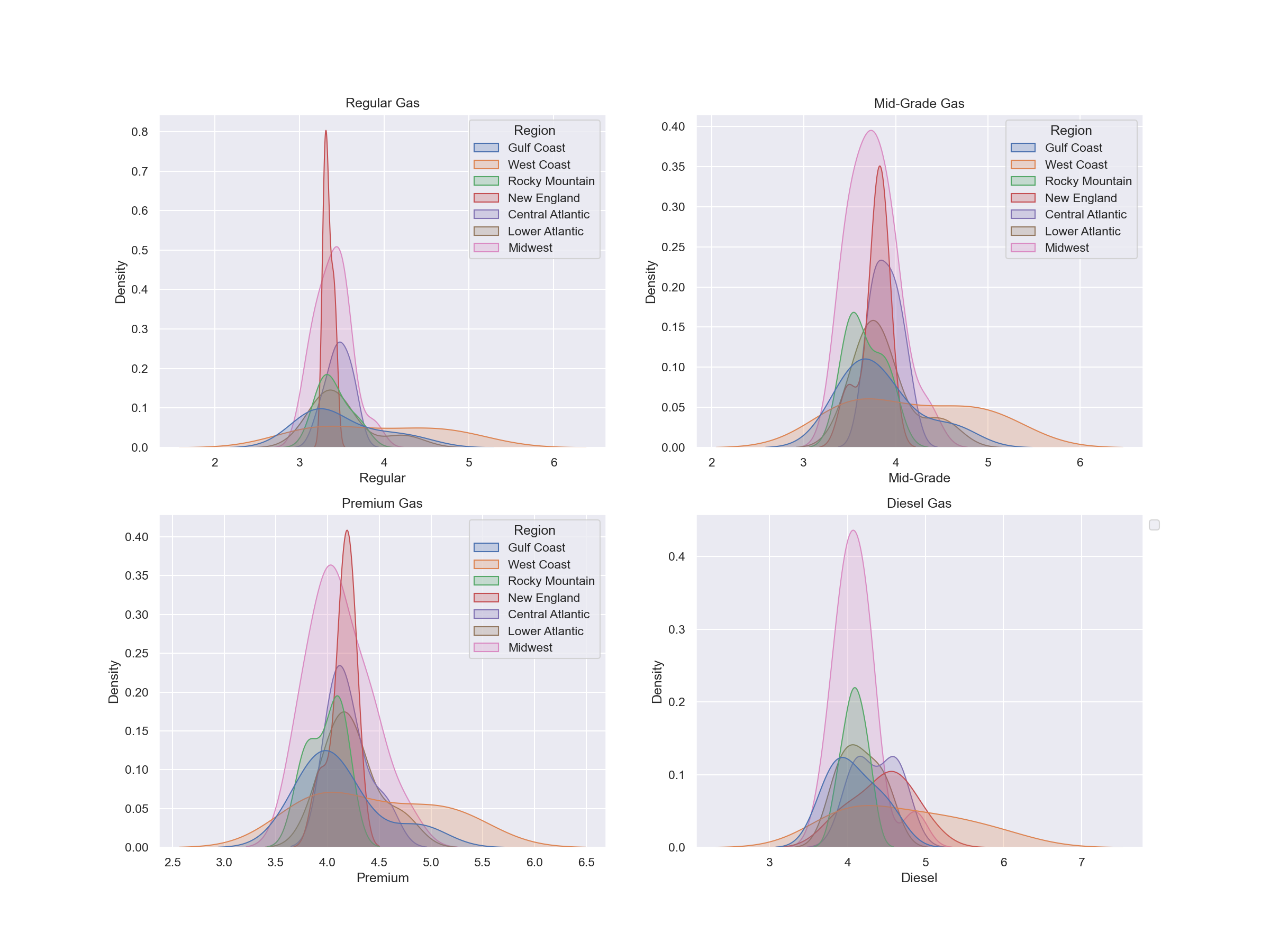
**Independent Variable Review**

The research and review of the question, lead to lots of information besides the article sentiments. There are a lot of other factors to consider when reviewing the public opinion on electric vehicles, and the independent variables found are some of those factors. Independent variables are the facts that wouldn’t change any analysis done, but could support or disprove the analysis. Things like the number of charging stations, the number of registered electric vehicles and more are some of the variables considered. All of the independent variable data found was evaluated by region and state, so that the analysis could be compared to the sentiment analysis work done. The regional breakdown can be found in Figure 6 located in the methodology section.

One of the first independent variables that impacted the analysis is the average gas prices in the US. This information gave the average gas price per state, per type of gas. All four types of gas were analyzed in probability density plots, to better summarize the variety of prices in states across the regions. The average gas price information was found from AAA. The information is updated daily on their site. The information was pulled for analysis around April 6, 2023. One can find the plot below in Figure 1.

Figure 1

*Average Gas Prices Per Region for Four Types of Gasoline*

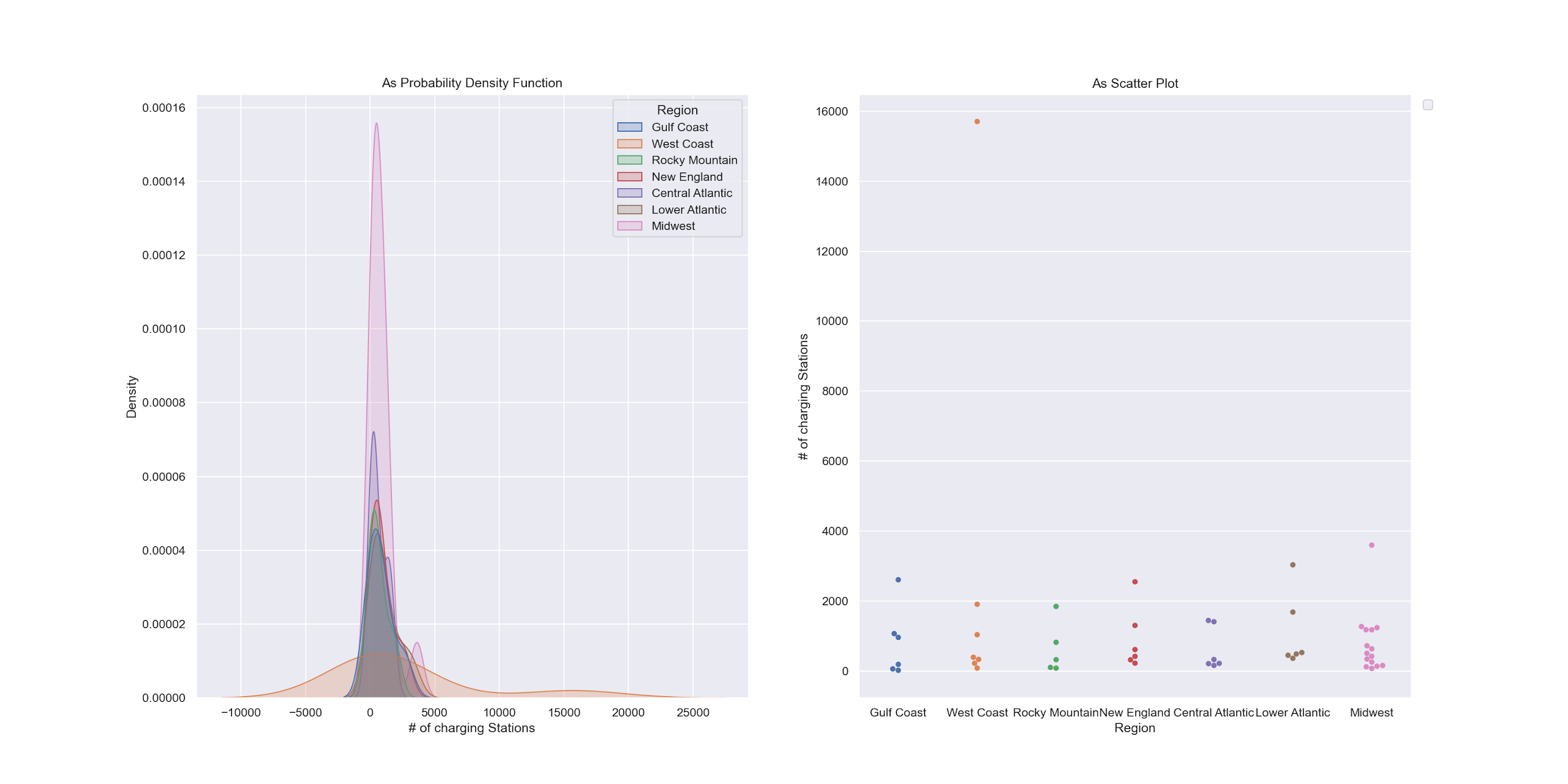


One can see the New England and the Midwest regions pretty consistently had the most consistent gas prices with higher densities than others.

The next independent variable considered is the number of electric vehicle charging stations in the various states. This variable listed the number of charging stations in each state. This data is from USAFACTS.org, which published the information in November 2022, having updated parts in March 2023. When looking at this data, one notices that this is a severe outlier in the data. One point is significantly higher than the rest of the data set and it is in the West Coast Region. This point belongs to the state of California, which has over 15,000 charging stations available. One can see the number of electric charging stations in Figure 2 below. This figure depicts the number of charging stations as a scatter plot and as a probability density plot.

Figure 2

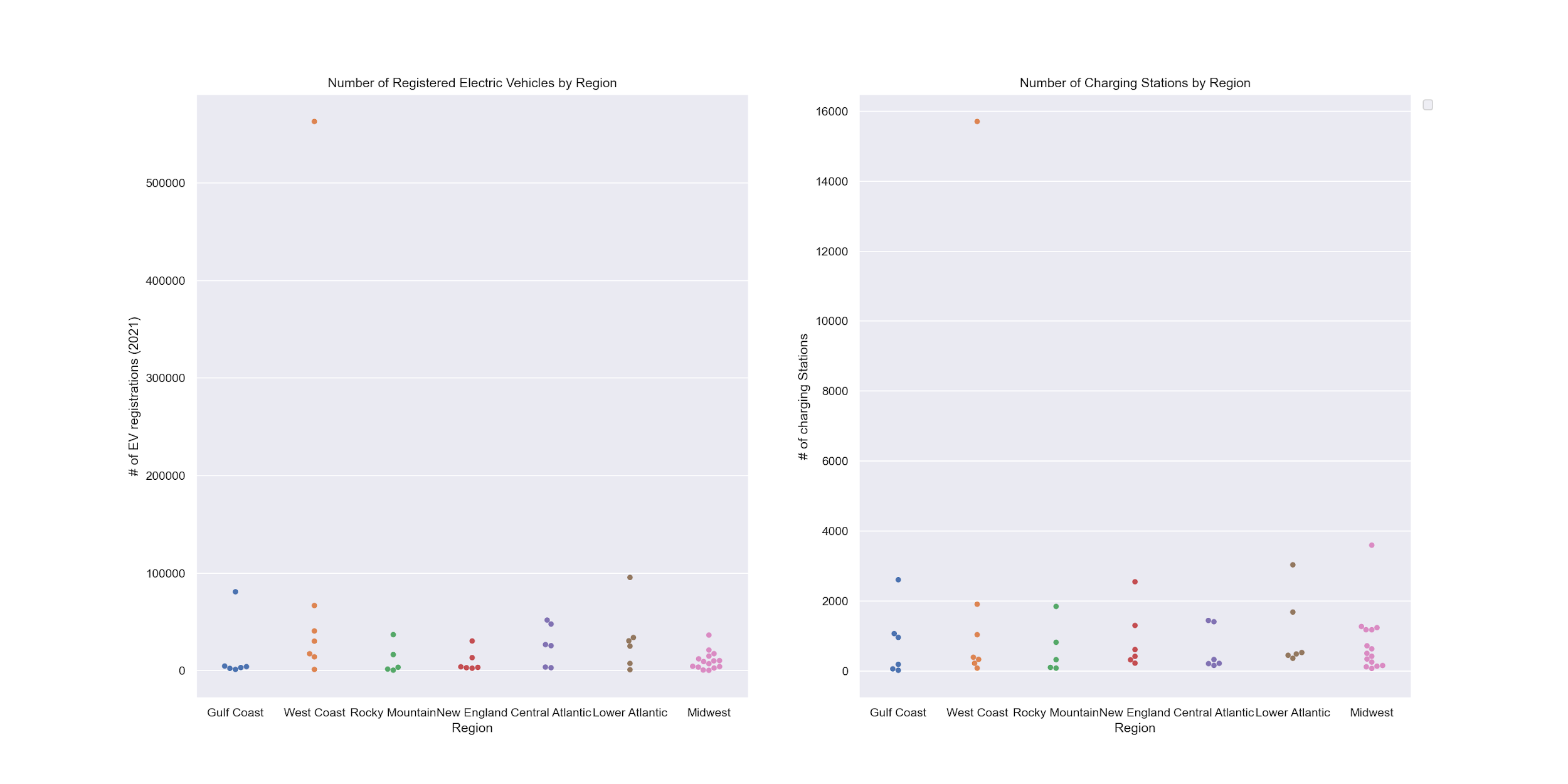
*Number of Electric Vehicle Charging Stations by Regions*



While this number is interesting, it is more interesting when compared to the number of registered electric vehicles in each state. Presumably the states with more charging stations would have more registered electric vehicles in them. And this information is mostly true as evident in Figure 3 below. This figure depicts the number of charging stations per region and next to it is the number of registered electric vehicles per region in the year 2021. There is the same outlier in both sets of data for the state of California. California has 563,070 registered electric vehicles in 2021 and 15,706 charging stations in it. This information was also found from USAFACTS.org.

Figure 3

*Number of Charging Stations by Region and Number of Registered Electric Vehicles by Region*

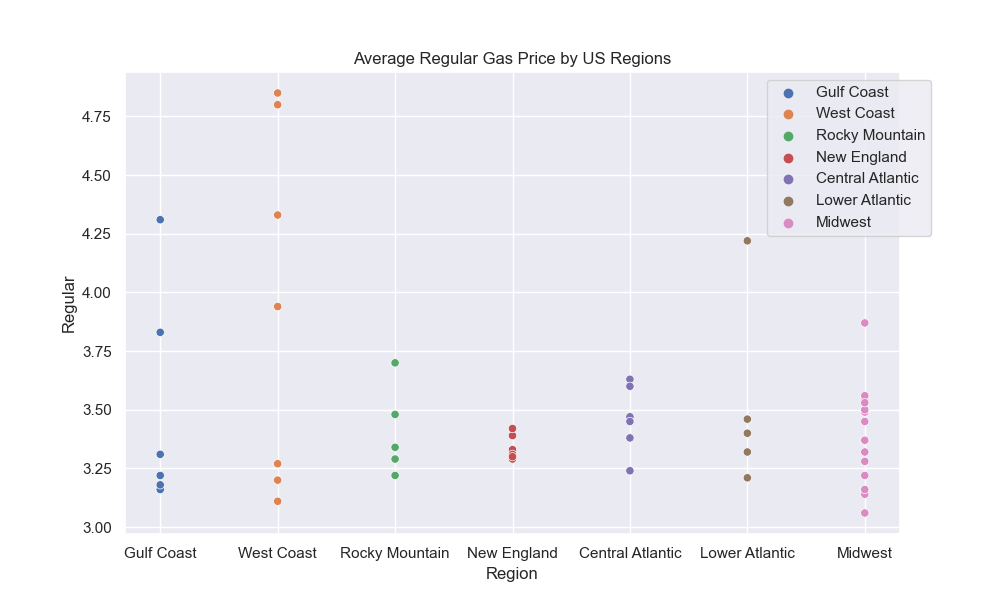


The data also supports the idea that the number of charging stations is not an anticipatory number. If the United States was fully invested in providing the infrastructure for electric vehicles nationwide, then there would be more charging stations than vehicles in these earlier stages. It was noted in the literature review that there is anxiety among drivers around getting an electric vehicle and not having anywhere to charge it. Resolving this problem would mean a significant investment in building more charging stations across the country.

The independent variables were also interesting when comparing the average gas prices with this information on charging stations. When considering Regular gasoline, the average price of gas across the nation can be seen in Figure 4, which is a scatterplot of the data. In this figure one can see that the West Coast region has some of the highest average gas prices, and they also have more charging stations and electric vehicles.

Figure 4

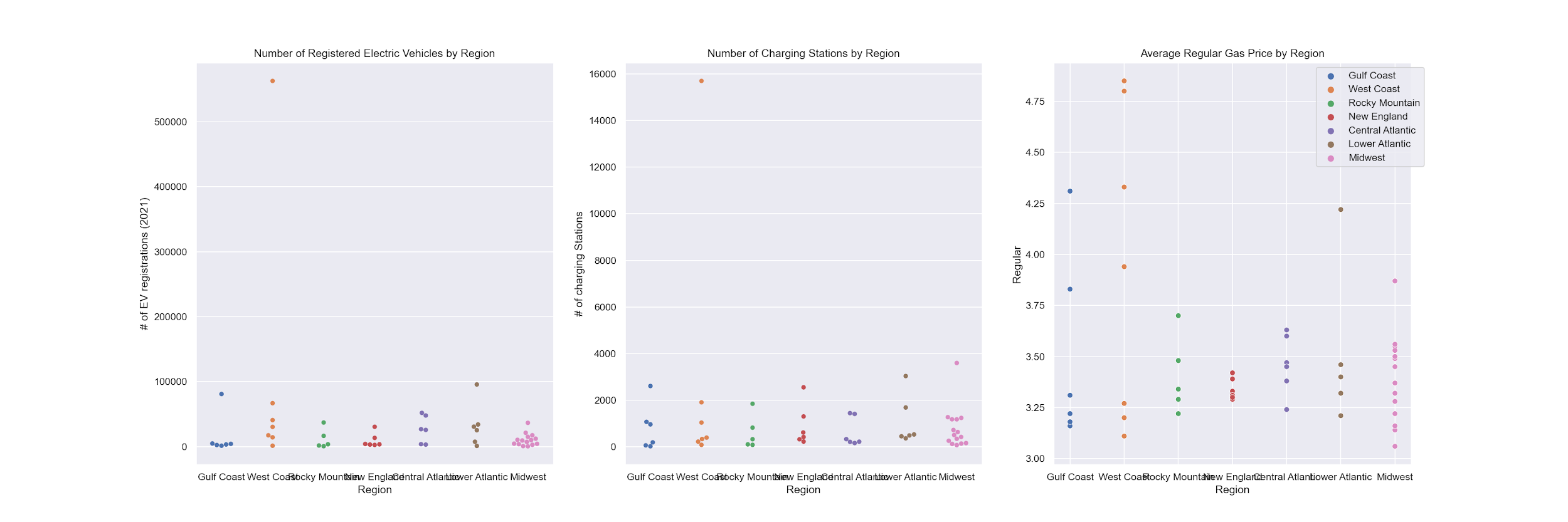
*Average Regular Gas Price by United States Regions*



Most notably from the independent variables, there was a lot of consistency in across variables. Reach of the regions lined up similarly, with a lot of the data across the nation sitting in similar areas and a few outliers in each category. The number of registered vehicles, the number of charging stations, and the average regular gas price can all be seen together in the figure below.

Figure 5

*Comparative Independent Variables*



Given the independent variables found, it really opened the conversation for what the results and topics could be most widely considered in the article review. The independent variables were helpful in evaluating the results of the sentiment analysis and explaining some of what was found.

**Methodology**

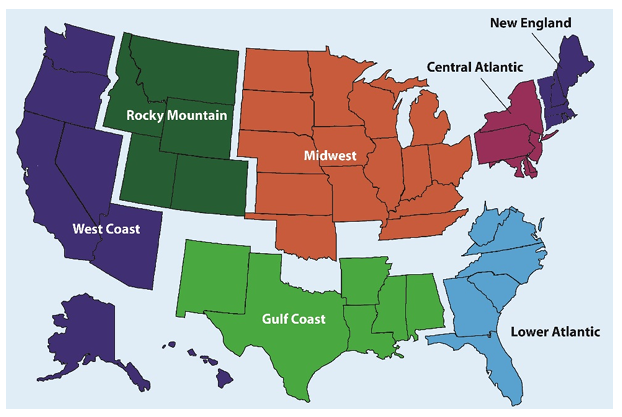
To explore this topic, the data analysis involved two significant steps, the collection of the data and the analysis of the data. These two areas were not independent of each other. The type of data collected would influence the analysis that needed to be done and certain analyses, that would be better for answering the proposed question, would be better analyzed with certain types of data. There were lots of options available to consider. Social media, blogs, articles, videos, and more all discuss electric vehicles. There was a lot of data to consider, evaluate, and learn from to look into the sentiment of electric vehicles in the United States. And the analysis chosen would need to fit with the data type chosen. To look at sentiment, it was determined that using articles would be the best source to evaluate and sentiment analysis using an LDA would be best.

There was originally the thought to look at Twitter and evaluate tweets on the topic of electric vehicles. Twitter is a great source for sentiment since it is full of people expressing their thoughts and opinions on things. However, to be able to do this type of analysis with Twitter, one needs a student developer license. One of these licenses was applied for but it was never granted. Had there been tweets to consider, it would have been interesting to compare the Twitter sentiment to the published article sentiment.

It was determined that the focus would be on looking at sentiment through published articles, specifically in newspapers, to evaluate the thoughts and opinions of the American public. But looking at the country as a whole would not be as thorough of an analysis as if individual states were considered. So it was determined that the United States would be broken into 7 regions to evaluate the sentiment within a region and between regions. One can find a map of the regions in Figure 6 below. This regional breakdown was created by the Department of Energy in their January 2023 Alternative Fuel Price Report. The regions would be able to help breakdown how sentiment varied throughout the US. Doing analysis state by state would not have been as fruitful, as not all local papers focused only on their state. Using a regional approach was faster and more accurate to evaluate sentiment. This map also matched the independent data gathered so comparisons could be made within regions as well.

Figure 6

*Map of United States Regions for Analysis*



With the regions broken down, the next step was to look at newspapers within each region to find articles related to electric vehicles. Every digital newspaper had a search bar and this was used to search the term “electric vehicles” in the search engine. The results were then sorted through, looking for the most relevant articles, but otherwise there was minimal filtering done. The goal was to find approximately 100 articles per region, but not all the articles first found were viable so the number of articles scraped per region varies. All the article weblinks were organized into a .csv or .txt file to be fed into a Jupyter notebook to be scraped. Some papers were easier to gather multiple articles from than others, due to things like paywalls, limited free articles, and coding that didn’t allow for the web scraper to work properly. So some areas had an abundance of newspapers used, and others used just two or three papers.

To scrap the articles, Beautiful Soup was the package used. Beautiful Soup is a package that “makes it easy to scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree” (Beautifulsoup4). Using this library, the output would be a csv that included the text of the article headline and the article itself, as well as the region and reporting paper. The Beautiful Soup library was used to find the URLs of articles through the ’href’ tag and it stored them in a txt file. The .txt file was then used to locate the text in the article by using the paragraph tag and finally it was exported into a CSV. This analysis was done regionally.

The next step was to analyze the articles and pre-processes them to calculate the sentiment of each. To do that, a text-blob was run on each article to extract the tokens (i.e., break it up into individual words) and help in identifying the parts of speech. Then it was time to remove the whitespaces and parentheses. This is the data clean up that needs to be done to complete sentiment analysis.

Now is when numerical analysis began to occur. The next step was to calculate the subjectivity and polarity. Polarity and subjectivity are two of the most significant characteristics of sentiment analysis. They assist to establish the overall sentiment communicated in a piece of text. Polarity relates to the text's emotional tone or feeling, which might be positive, negative, or neutral. Polarity is often assessed on a scale of -1 to +1 in sentiment analysis, with negative values indicating negative sentiment, positive values indicating positive emotion, and zero suggesting neutral opinion. On the other hand, subjectivity refers to the degree to which a statement is subjective or objective. Subjective statements represent personal ideas, emotions, or sentiments, whereas objective statements are facts that are free of personal prejudice or emotion. Subjectivity is often quantified in sentiment analysis on a scale of 0 to 1, with a score of 0 indicating totally objective statements and a score of 1 indicating completely subjective remarks. Subjectivity and polarity are both significant aspects of sentiment analysis since they assist to assess the overall sentiment conveyed in a piece of text. The polarity and subjectivity were calculated using Textblob. Textblob uses a Lexicon based approach where each word in the text is assigned a polarity, or sentiment, score between -1 and 1. The polarity and subjectivity were calculated per article per region and added to the dataframe They were also graphed in a probability density function to compare how these numbers fluctuate between either the source paper or the state. Many of the polarity and subjectivity graphs can be seen in further figures.

Once polarity and subjectivity were acquired, it was time to perform sentiment analysis on the headlines and articles using transformers and pretrained models. Transformers are neural networks that use attention in order to be able to review longer sequences of information without running out of memory. To score the article headline, the model “bert-base-multilingual-uncased” was used, which is a BERT model. BERT stands for Bidirectional Encoder Representations from Transformers. The wrapper classed TFAutoModelForSequenceClassification allows for easy loading of pre-trained models for sequence classification tasks, while AutoTokenizer is used to load the appropriate tokenizer for the model. BERT can only handle documents of up to 512 tokens without needing to break the documents into smaller sections or windows; therefore a deferment model designed for longer documents, the AllenAI Longformer base-4096 model, was used to determine the sentiment of the articles. The longformer model uses a combination of window attention and global attention in order to analyze documents with up to 4,096 tokens. The pipeline() method from the Transformers library was then used to set up the pipeline for sentiment analysis. This pipeline takes in a model and tokenizer and automatically applies the necessary pre-processing and post-processing steps to perform sentiment analysis on text input. This provided scores of each headline and the articles in the dataframe.

At this point, the polarity, subjectivity, headline score, and article score were all computed. These numbers are compared by each region to better shape the further analysis. The polarity, subjectivity, headline score, and article score were all graphed as Kernel Density Estimate, which is a type of probability density function. In evaluating the graphs, it was determined that the peaks are typically the median values, meaning the most likely value and the higher the peak the more that value was seen. The more spread out the curve the more range the data had and the less similarity of the score amongst the evaluating parameters.

It was then time to move onto topic modeling. Before analyzing the topics in the articles, the texts had to be prepared through several steps of preprocessing. First, all of the words had to be tokenized and nltk was used to remove stop words. Additional stop words were added that were specific to the area of research such as gazette, copyright, and subscription. Next, all non-alpha numeric characters and single character words were removed from the text and each of the words were lemmatized. Finally, bigrams were created and each bigram appeared a minimum of twenty times to the tokenized text.

In order to model the topics of the articles in the study, Latent Dirichlet Allocation, or LDA, model was used. The first step in creating an LDA model is determining the correct number of topics. There are a number of different approaches for finding the optimal number of topics for an LDA model, including analyzing perplexity and coherence. Different approaches have different pros and cons. The advantage to assessing models based on topic coherence, is that topic coherence is meant to reflect how well a topic can be understood by a human. In order to measure coherence this study used the UMass coherence score. UMass defines coherence as

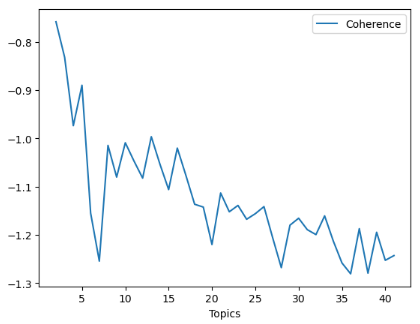
(Zvornicanin, 2023)

In this equation D(wi, wj) stands for how often word i and word j simultaneously appear within a topic, while D(wi) represents how often word I appears alone.

We measured coherence using the Coherence model in the Gensim library. As can be seen in Fig. 2, the coherence scores trended downwards as the number of topics increased with slight spikes along the way. The highest coherence score occurs with only two topics.

Figure 8

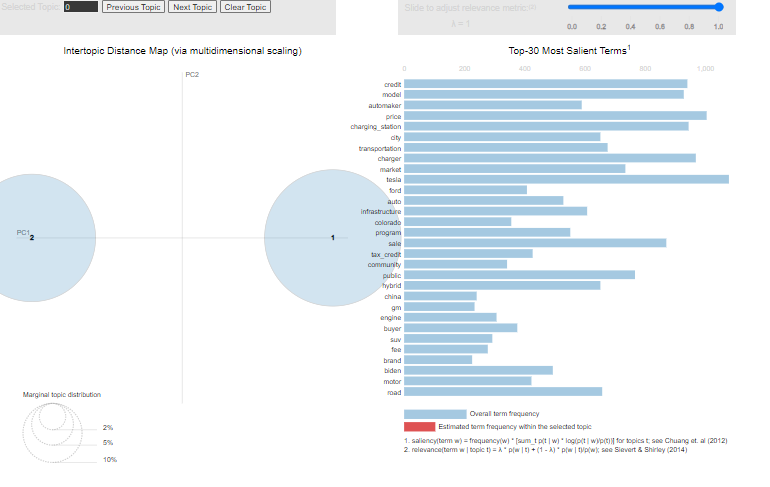
*Coherence by Number of Topics*



Now it was time to look into the topics and what significance they play to the research question. Using the LDA model, a sample size of 2 topics was looked at and the model was trained for a few iterations to look into the most prominent topics. Then one must use the pyLDAvis package to display the interactive results of the topics as seen in Figure 9 below

Figure 9

*Screenshot of the Interactive Topic Model*



It was then time to calculate the perplexity of the LDA model. Perplexity is the measure used to evaluate the performance of the probabilistic models like language and topic models. A lower perplexity score indicates that the model is better at predicting the words in the test set, and thus better at generalization of the performance.After looking at perplexity, it was time to move on to calculate the coherence of the model. Coherence is a metric used to evaluate the quality of the model. It measures the degree of semantic similarity between the top words in the topic, based on the co-occurrence of the corpus. A higher coherence score indicates that the model has learned more coherent and interpretable topics with stronger semantic relationships between the top words. Finally, the final topic(s) are distributed across all the data points in the dataframe, so every article was assigned a final topic. From there the data frame was exported into a csv, to process it further and analyze the outcomes.

**Outcomes**

Using the methodology described a lot of information was available for analysis. Polarity, subjectivity, headline score, and article score were great for comparing between regions to evaluate sentiment. And the topics revealed by the LDA model really gave some great insight to the most important topics around electric vehicles. The information presented really gave a solid answer to the research question,

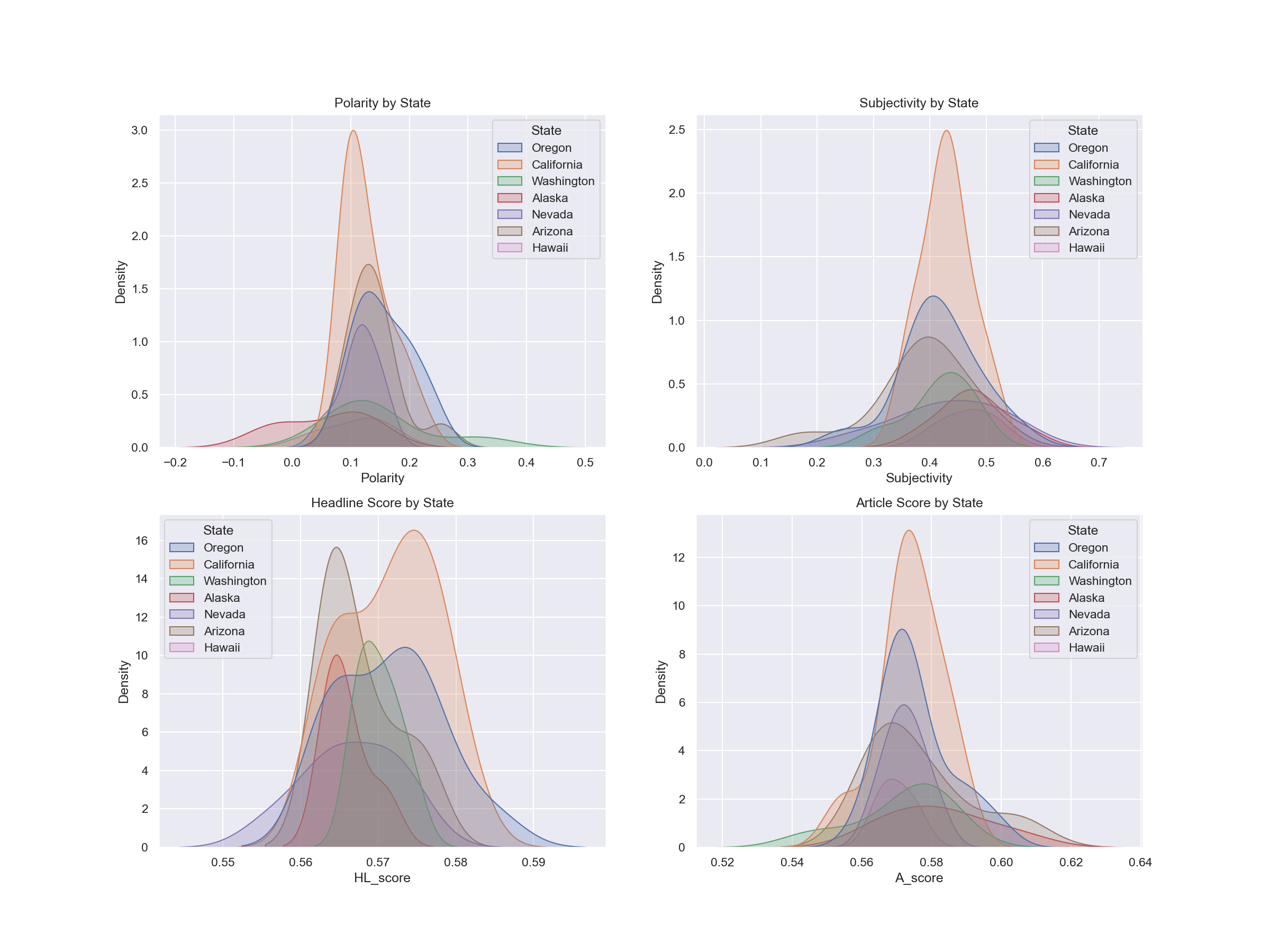
The polarity, subjectivity, headline score, and article score for each region were first considered. In each region, these topics were either sorted based on Source or the State from which the paper originated, to see how these things compared throughout the region. One wanted to see if there were any extreme outliers in a region, if there was consistency in a region, and so forth. These values were compared on a Kernel Density Estimate, a type of probability density function, which shows the most likely scores at the peak of a curve. To read these graphs, one needs to know that where the peak is highest is where the score, value, etc. is most likely to occur. The higher the peak, the more likely that number is to occur.

Figure 10 displays this data for the West Coast Region. In this region, the comparison was done by state, since there were fewer articles per paper found and the region includes Alaska and Hawaii which would have extreme opinions compared to the contiguous United States. In this figure one can see that California had more narrow and higher peaks than the other states. What is notable in this area can be summarized in the following points:

* Polarity was mostly positive, with Alaska having negative or neutral regards
* Subjectivity was most significant between 0.4 and 0.5 for the region, meaning sentiment was in the middle of neutrality. The papers weren’t objective or subjective and had equal components of both.
* Headline scores varied the most widely. Which is to be expected. Headlines are supposed to be short and attention grabbing. Nuances can be left out when something is quick.
* Article scores varied the least of any range of data presented. With a less than 0.1 range, it showed that the articles really hovered in the middle.

Figure 10

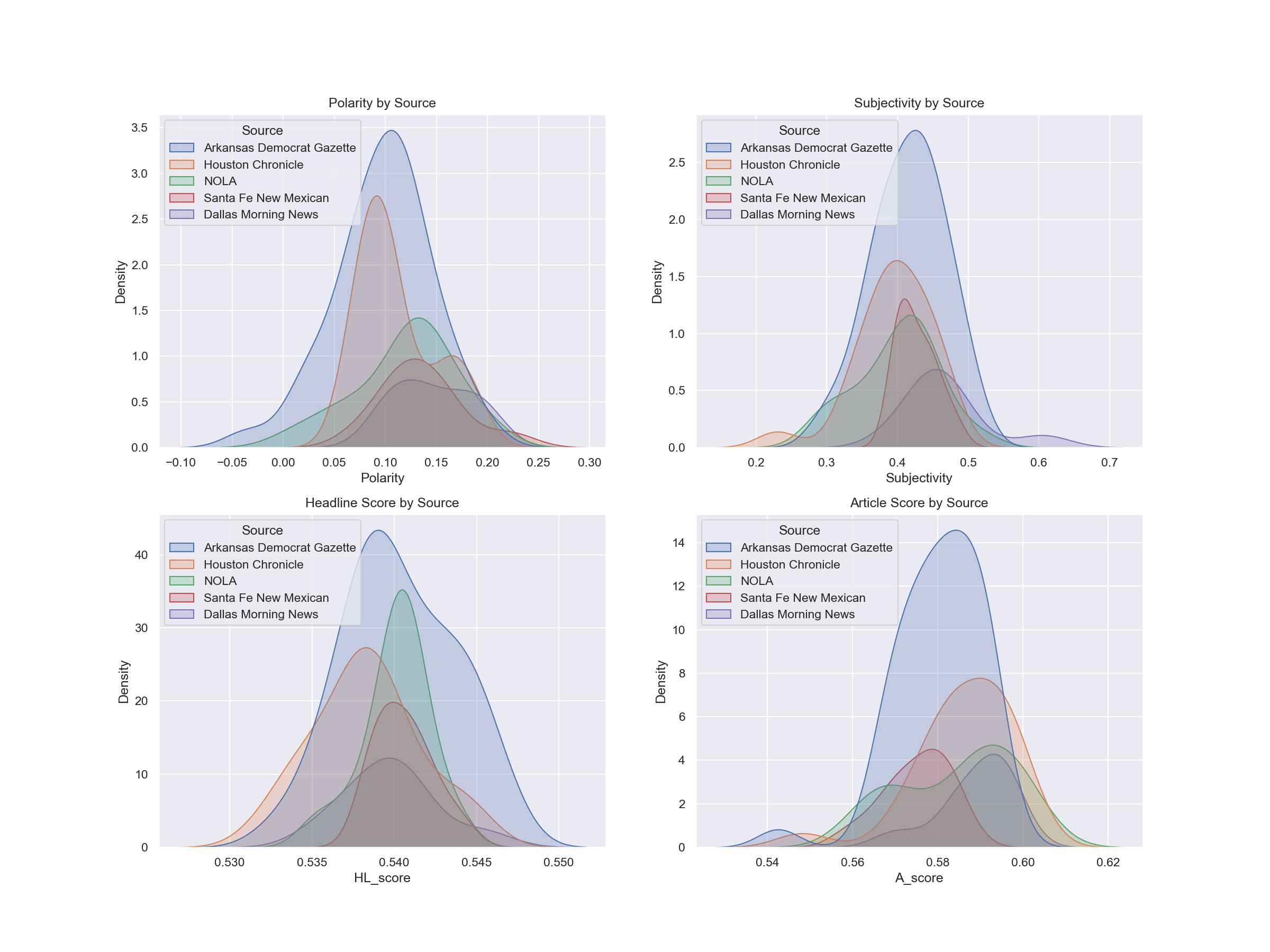
*Polarity, Subjectivity, Headline Score, and Article Score for the West Coast Compared by State*



The Gulf Coast Region had a variety of papers used so the dataset was quite interesting to evaluate. One of the papers used in evaluating the region, the Arkansas Democrat Gazette, had results fitting a democratic newspaper. In the politicization of electric vehicles, democrats have been more positive and supportive of electric vehicles. So one would think that the paper would have more positive polarity and possibly less subjectivity. However that is not necessarily what is shown in Figure 11. One actually sees that the polarity and subjectivity are in line with the other papers of the region. But three are much higher density scores for this paper than others. This means that the paper was very confident and consistent in how they wrote about electric vehicles.

Figure 11

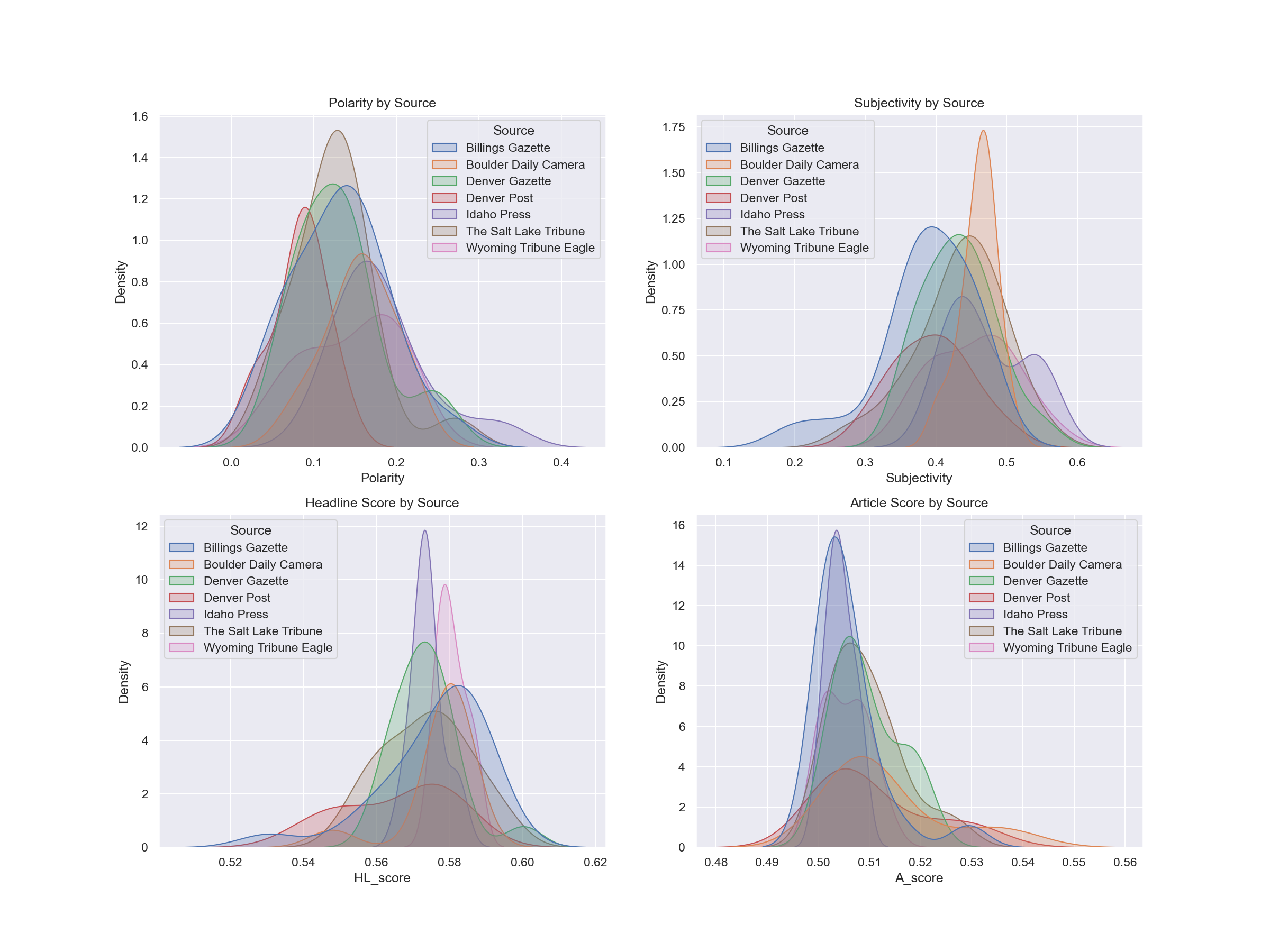
*Polarity, Subjectivity, Headline Score, and Article Score for the Gulf Coast Compared by Source*



The next area considered was the Rocky Mountain area. This area had much more positive polarity in it than some of the others. The polarity graph, seen in Figure 12, has the x-axis only going to -0.1, while for the other regions it showed much more negative numbers. But polarity still was peaking around 0.1 and 0.2 for all the papers in this region, which is consistent with the other regions seen.

Figure 12

*Polarity, Subjectivity, Headline Score, and Article Score for the Rocky Mountain Compared by Source*



Continuing to move across the United States, the next region considered is the Midwest. The Midwest region saw an outlier in the Chicago Tribune. The density for the polarity and subjectivity of the Chicago Tribune was significantly higher than the other articles. This is due in part to the number of articles taken from the Chicago Tribune and for the consistency in the results from the Chicago Tribune articles.

Figure 13

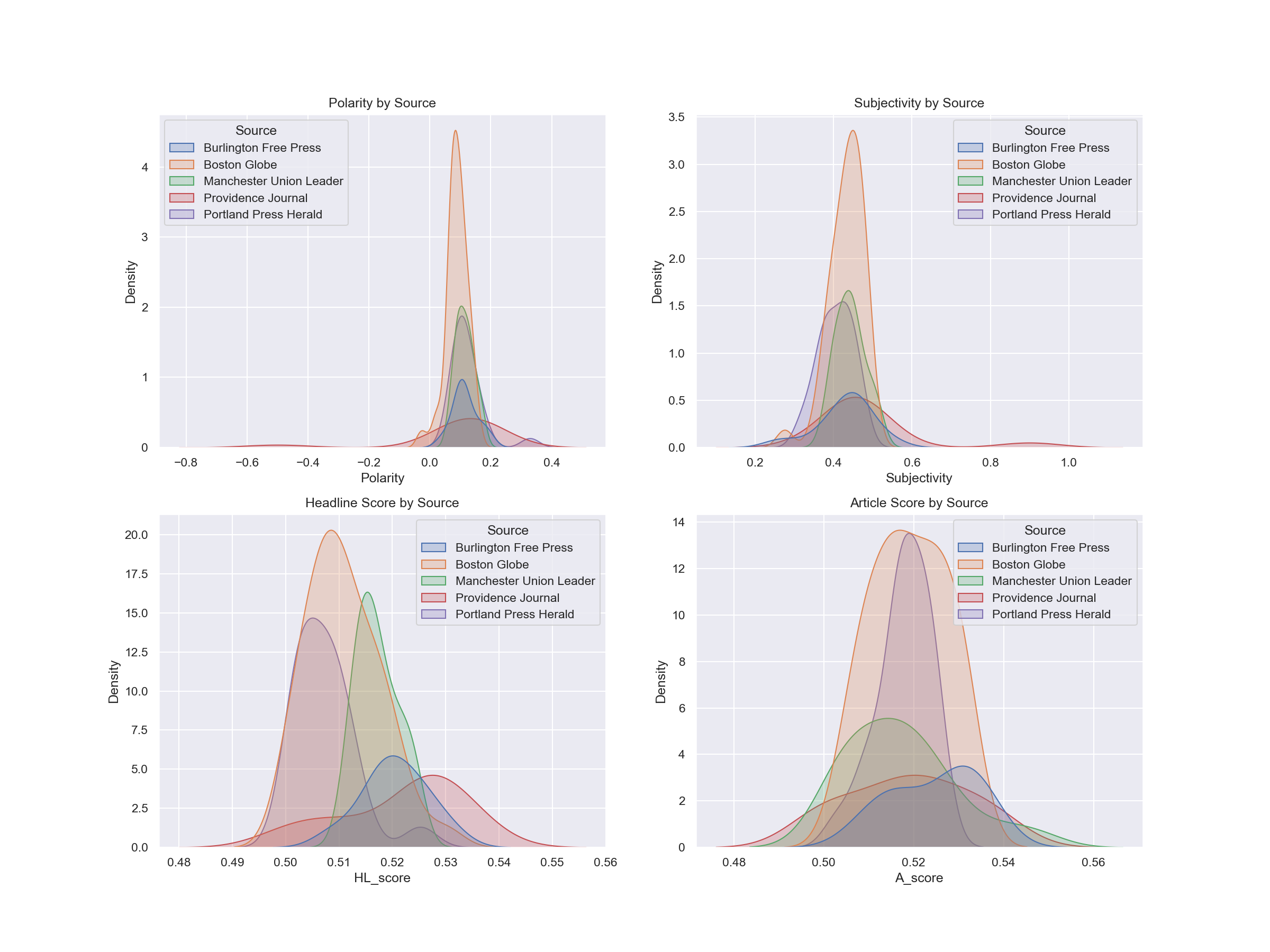
*Polarity, Subjectivity, Headline Score, and Article Score for the Midwest Compared by Source*



There are two more regions evaluated by source before the data was looked at nationally. In Figure 14 one can see the New England region analysis. The Boston Globe had very narrow curves, meaning the results were very centered around the median, meaning consistency in the sentiment and polarity of the articles from the Boston Globe.

Figure 14

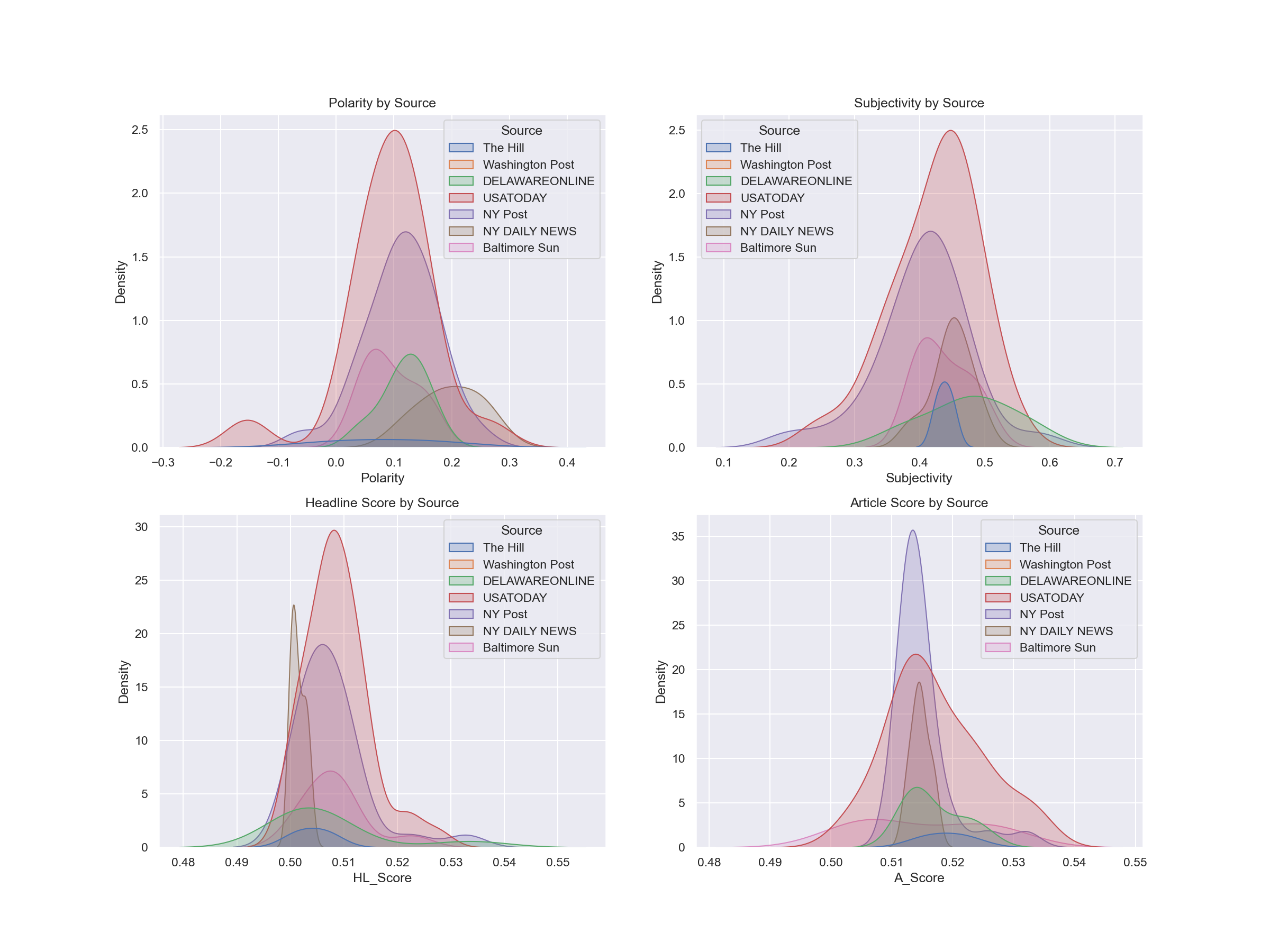
*Polarity, Subjectivity, Headline Score, and Article Score for the New England Compared by Source*



When looking at the Mid-Atlantic region, the articles were compared between the two newspapers that were used to gather articles. Since the articles and the information collected were more comprehensive and integrated between the states of the region, comparing the scores based on state didn’t make sense. The Mid-Atlantic region was not exciting or eye-opening in its regional analysis as one can see in Figure 15 below.

Figure 15

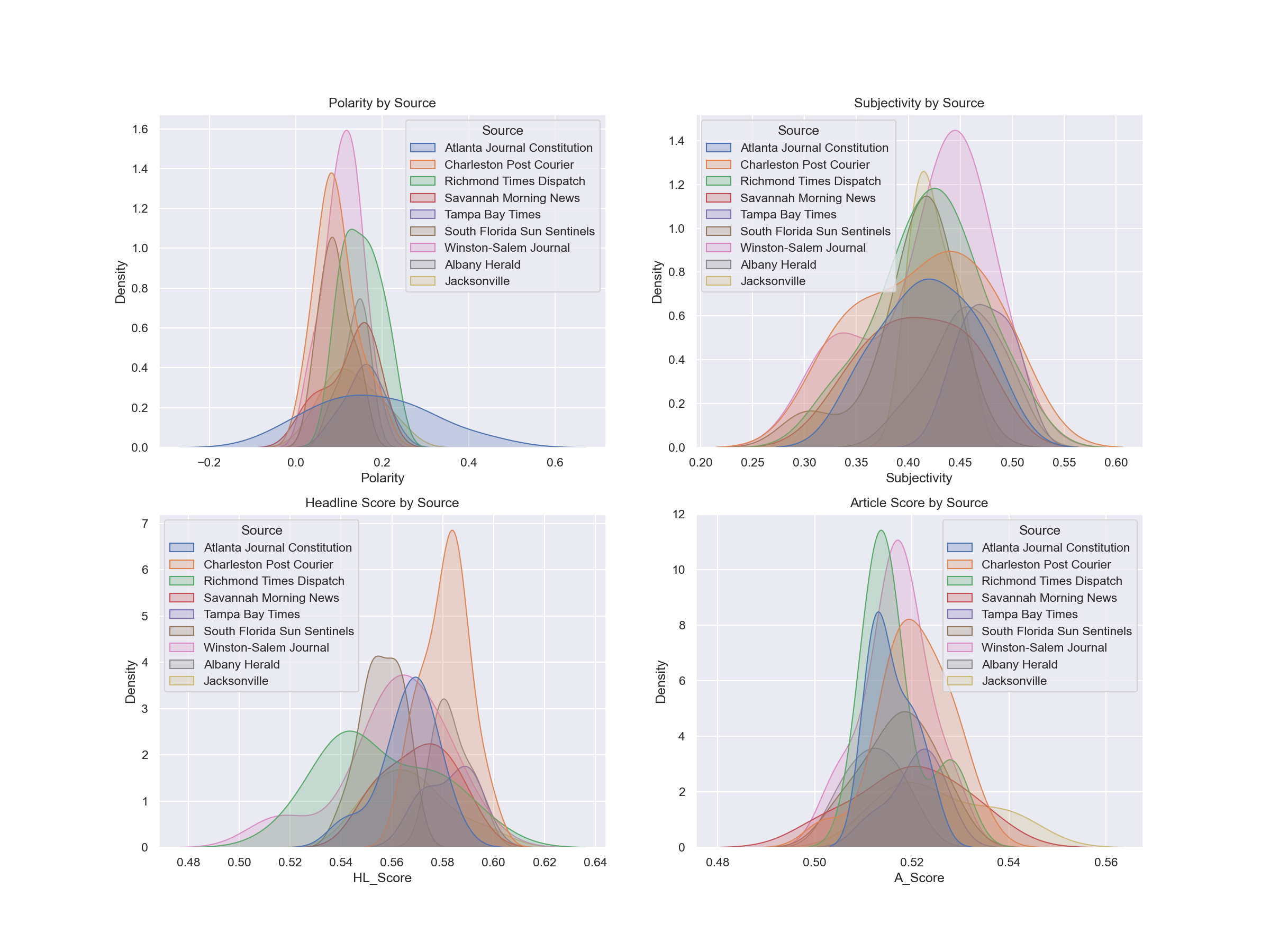
*Polarity, Subjectivity, Headline Score, and Article Score for the Mid-Atlantic Compared by Source*



The Lower Atlantic region was similarly simple in its analysis. There were some interesting bumps in the Kernel Density Plots, with the Jacksonville paper having some high negative polarity and some high positive subjectivity. The results for the Lower Atlantic region can be seen in Figure 16.

Figure 16

*Polarity, Subjectivity, Headline Score, and Article Score for the Lower Atlantic Compared by Source*



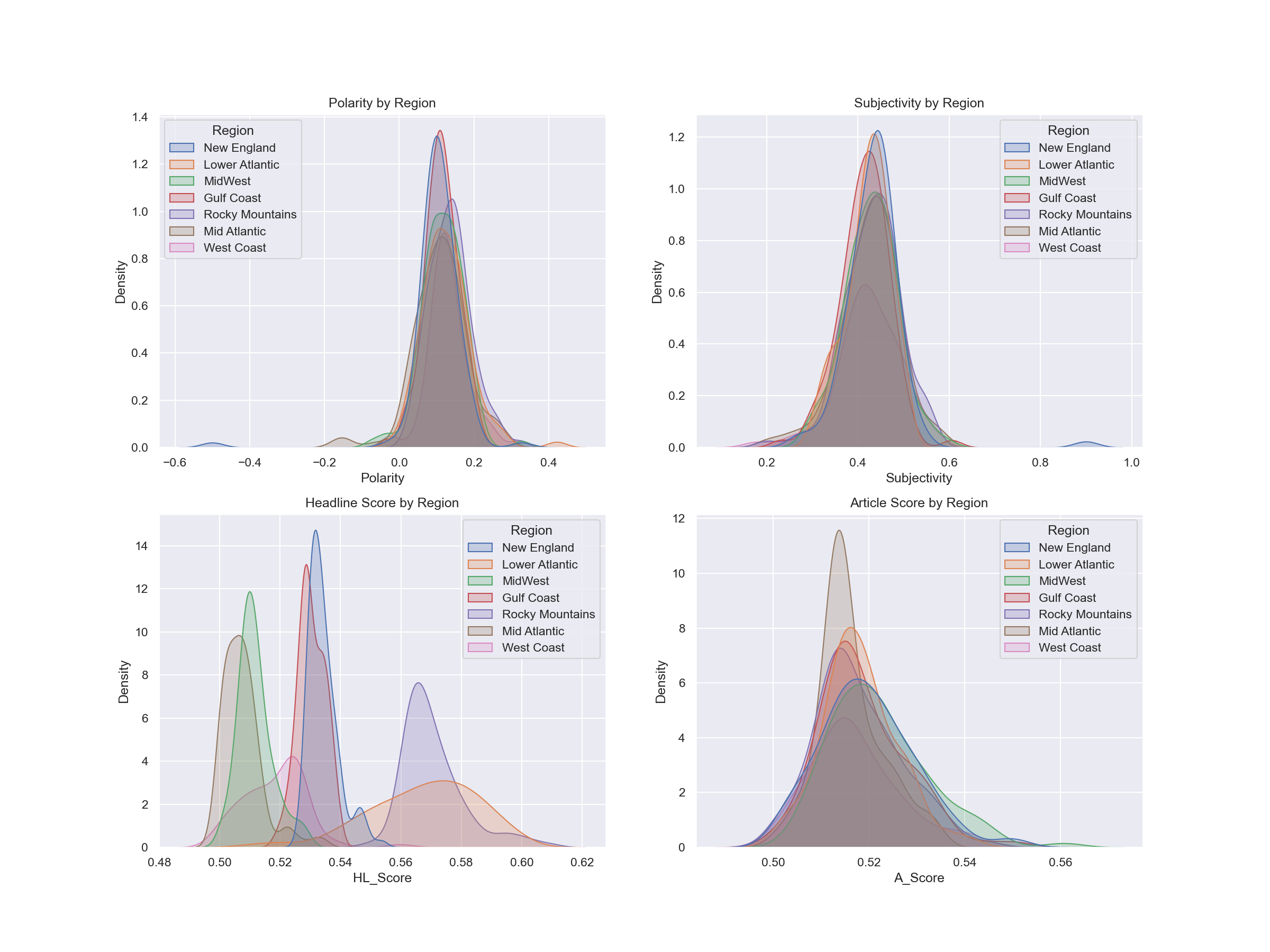
The regional analysis was done step by step to begin the understanding in the pursuit of sentiment analysis. Understanding how the regions looked impacted the features that should be looked out for when doing a national analysis. Would the smaller curves be more obvious on a national scale? Would the outliers be outliers in the national scale? These were all things to be considered when doing the national analysis, which was done next.

The national analysis broken by region can be found in Figure 17. There were a number of things noted in comparing the national dataset that are summarized as follows:

* Polarity is overall positive. Sitting at about 0.1, overall the polarity of articles around electric vehicles was seen as positive. There was some extreme negative polarity for a few regions.
* Subjectivity was just under 0.5 for all the regions. That means the articles weren’t too subjective and not too objective, and just neutral. There were a few outliers of high subjectivity, meaning the articles were not objective, in some regions.
* Head line scores were all over the place for each and every region of the United States. The overall range was between 0.48 and 0.62, so not the largest possible spread, but still a spread.

Figure 17

*National Polarity, Subjectivity, Headline Score, and Article Score per Region*

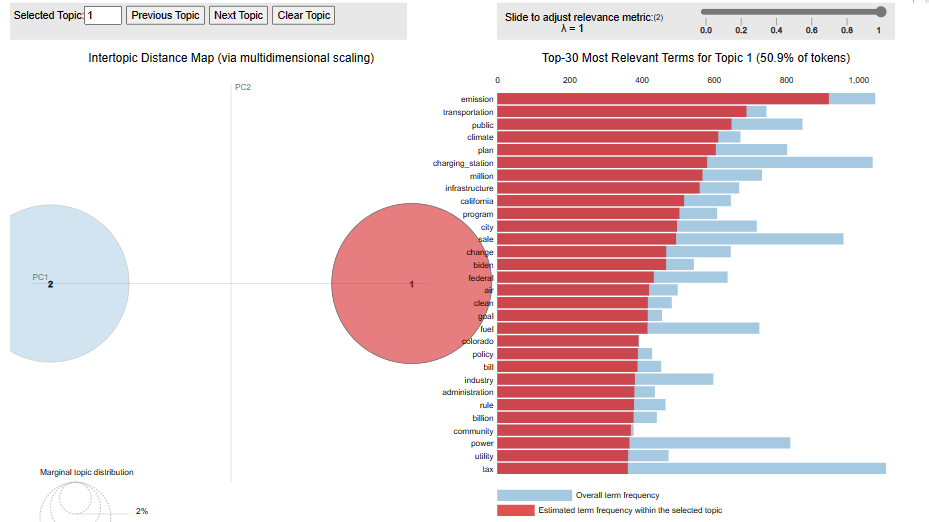


This information at the national level was very eye opening. The fact that polarity and subjectivity were so close for the whole nation was not expected. The variety in head line scores were also so interesting. Both of these things could probably be explained by the very nature of journalism. These are published works and while they seek to inform individuals, they are trying to get people to subscribe and spend money. Headlines are meant to be catchy and attention grabbing. They may be intense, shocking, dramatic, or exciting to get the reader's attention, encourage viewership and in turn gather more subscribers. The similarity in subjectivity is again a product of journalism. Respected journalists attempt to remain objective, minimize their subjectivity, and present fact. The more sharply objective pieces were most likely one of the op-eds or opinion pieces gathered.

Upon running the Gensim analysis, the perplexity and coherence indicated that 2 topics were the most relevant to the data collected. The Gensim LDA model has a visualizer that allows a visual look at the topics and their relevance. In Figure 18, one can see the first topic of relevance. This topic is referred to as climate. In this topic, the most relevant terms included emission, transportation, climate, and public. This topic was not surprising based on the independent variables, or supplemental articles that were considered and evaluated. These topic words relate to the environmental impacts of electric vehicles. In the reports by the Department of Energy charging stations and emissions are common topics discussed.

Figure 18

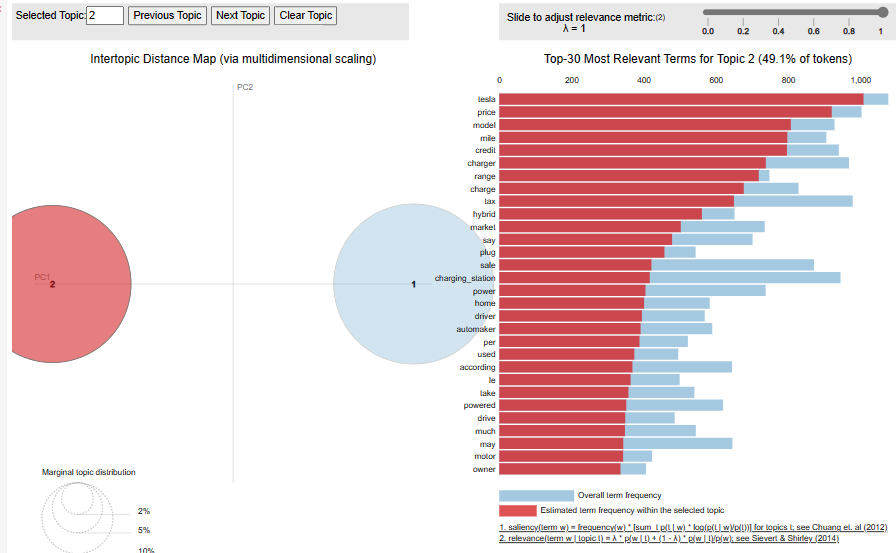
*Gensim Visualizer of First Model Topic: climate*



The second topic, Tesla, seen in Figure 19, has some of its key terms as tesla, price, model, and charger. The charger term is relevant because Tesla chargers are not universal to all electric vehicles and can only be used for a Tesla. Tesla also has various models at various price points. Finding this topic to be the second most prevalent was not indicated in any of the independent variables or literature reviewed prior to analysis. While Tesla is known as a big player in the electric vehicle market, the conversation has shifted in recent years to include other EV manufacturers and look at the holistic electric vehicle area.

Figure 19

*Gensim Visualizer of Second Model Topic: climate*



The analysis presented does offer some answers to the proposed research questions: What is the US public’s opinion regarding electric vehicles? What are the influencing factors affecting public opinion on electric vehicles? The national sentiment can be seen in Figure 17, where the subjectivity, polarity, and article scores are all present. Nationally, there is positive polarity and subjectivity regarding electric vehicles. The articles' scores peak between 0.51 and 0.52, meaning the confidence of the article labels (which were almost exclusively at 1) was pretty confident given the amount of articles reviewed. This all says that the US public is generally more positive about electric vehicles. They are interested, they want to know more, they possibly want one, and there isn’t any extreme resistance. The influencing factors are found through the topics. The tesla topic can also represent price and market availability; consumers are concerned about what is available and the price of an electric vehicle. The climate topic indicates public interest in the clean energy implications of electric vehicles.

It feels safe to say that at this point, the US public has a positive opinion of electric vehicles and their influencing factors are the market availability and climate change. This may seem reductive, but this is after only analyzing one type of source, published articles. Further analysis with other source types would be very interesting to evaluate and see how it compares to the data collected here. One would think that with Twitter usually being the direct output of an individual’s feelings, there would be much more polarizing data and less objective facts. Comparing individual feelings to the compiled fact-driven articles would create a very interesting analysis and reflection on what is truly the public opinion.

**Conclusion**

Sentiment Analysis of US opinions toward electric vehicles demonstrated that using only one type of source can lead to monotonous and simple results. The medium or source is very influential in how sentiment analysis is evaluated. Using published articles to evaluate sentiment was informational, but did not necessarily reveal any new conclusions or information not previously known or suspected. The United States is generally positive towards electric vehicles and is focused on the climate impacts and market and this is not a bad place to be at this moment.

**References**

AAA. (n.d.). *State gas price averages*. AAA Gas Prices. https://gasprices.aaa.com/state-gas-price-averages/

Bansal, P., & Kockelman, K. M. (2017). Understanding the influence of household characteristics and attitudes on electric vehicle usage in Texas. International Journal of Sustainable Transportation, 11(6), 434-447.

*Beautifulsoup4*. PyPI. (n.d.). https://pypi.org/project/beautifulsoup4/

Clean Cities, & Bourbon, E., Alternative Fuels Data Center (2023). Retrieved April 2023, from https://afdc.energy.gov/files/u/publication/alternative\_fuel\_price\_report\_january\_2023.pdf

Hidrue, M. K., Parsons, G. R., Kempton, W., & Gardner, M. P. (2011). Willingness to pay for electric vehicles and their attributes. Transportation Research Part D: Transport and Environment, 16(8), 1-7.

*How many electric vehicle charging stations are there in the US?*. USA Facts. (2023, March 28). https://usafacts.org/articles/how-many-electric-vehicle-charging-stations-are-there-in-the-us/?utm\_source=google&utm\_medium=cpc&utm\_campaign=ND-Environment&gclid=Cj0KCQjwuLShBhC\_ARIsAFod4fJU72aUh51VV-nFVGxgGO8dF5j9\_z8EziN9ZVxYvB8Wq4ed2kI34rMaAkK8EALw\_wcB

Kapadia, S. (2022, December 24). Evaluate Topic Models: Latent Dirichlet Allocation (LDA). *Medium*. https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0

U.S. Department of Energy. (n.d.). *Alternative fueling station counts by State*. Alternative Fuels Data Center: Alternative Fueling Station Counts by State. https://afdc.energy.gov/stations/states#:~:text=Station%20Counts%20by%20State%20and%20Fuel%20Type%20,%20%20797%20%2015%20more%20rows%20

Schoenau-Fog, H., Fitch-Roy, O., Chauhan, A., & Foxon, T. (2021). Product qualities, EV attitudes: Exploring the influence of consumers’ perceptions of product qualities on attitudes towards electric vehicles. Transportation Research Part D: Transport and Environment, 94, 102845. https://doi.org/10.1016/j.trd.2021.102845.

Shah, P. (2021, December 15). Sentiment Analysis using TextBlob - Towards Data Science. Medium. https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524

Wang, J., Wang, Y., Li, X., Li, J., & Zhang, X. (2019). Attitude towards electric vehicles: An investigation of public perception and demand characteristics in California. Journal of Transport Geography, 74, 184-195.

Yalcinkaya, G., & Alkaya, E. (2021). Understanding customer attitudes towards electric vehicles: A review of factors affecting customer perceptions of electric vehicles. Energy Reports, 7, 3495-3504. <https://doi.org/10.1016/j.egyr.2021.08.002>.

Zvornicanin, E., & Zvornicanin, E. (2023). When Coherence Score Is Good or Bad in Topic Modeling? | Baeldung on Computer Science. *Baeldung on Computer Science*. https://www.baeldung.com/cs/topic-modeling-coherence-score