ANALYZING DIALECTAL DIFFERENCES IN RELATION TO GEOGRAPHY IN THE AMERICAN SOUTH

by

LEAH M. DUDLEY

A Thesis Submitted to the University of Georgia's Center for Undergraduate Research Opportunities in Fulfillment of the CURO Graduation Distinction

Athens, Georgia

2019

ACKNOWLEDGEMENTS

This thesis is the result of two years of independent research within the University of Georgia's department of linguistics. I would like to thank many people who helped me along this journey and allowed me to realize my passion for research within linguistics. Of my professors, I would like to thank Dr. Sandra Kotzor, who I conducted my first research project on phonetics with. This led to my discovery of my love for the subfield of sociophonetics, my primary field of research. I would also like to thank my reader and mentor Dr. Peggy Renwick for introducing me to quantitative methods within the humanities, which led me to pursuing a Digital Humanities Certificate. Above all I would like to thank Dr. William Kretzschmar. Dr. Kretzschmar met me while I was working at the Linguistic Atlas my sophomore year and has allowed me to conduct independent research under his guidance ever since. He has been the guiding force in many of my undergraduate decisions, from becoming a 2018 CURO Summer Fellow to applying to graduate schools. I will forever be grateful for his kindness in allowing me to work under him and discover the amazing field of dialect.

On a non-academic note, I would also like to thank those around me. I extend gratitude to my parents especially, for funding my undergraduate degree and supporting my ambitions to obtain a PhD in a field they hadn't heard of until I started majoring in it. I would like to thank my sister for always having a positive outlook even when I didn't believe in myself. Finally, I would like to thank my friends and extended family for offering me endless support, advice, and reviews of my writing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The southeast region of the United States, colloquially referred to as 'the South', is arguably one of the most distinct regions of the country. In the era after the Civil War, the South was disparate from the rest of America in part due to "its case-like system of race relations, its agriculturally based economy and its relatively slow industrialization and urbanization, its fundamentalist religion, and its feeling of separateness from the rest of the nation" (Pederson 1974). This dissimilarity from the rest of America manifested itself in a regional dialect that was highly contrastive to the rest of the country, what most would aptly call 'a Southern accent'. However, generalizing this entire region as having a singular accent is an assertion any dialectologist would bristle against. They would instead claim that this area is made up of smaller regional accents. How these regional accents should be parsed is nonetheless a huge point of contention.

I chose to analyze Southern speech by analyzing existing linguistic collections available through my prior transcription and research on the Linguistic Atlas Project, or LAP (Pederson, 1968). The corpora I chose were the Linguistic Atlas of Gulf States (LAGS) and the Digital Archive of Southern Speech (DASS). LAGS is part of the larger LAP (Kretzschmar et. al, 2011). The LAP was created in the 1930's as a way to accurately collect regional speaker data in order to document language variation and is still an ongoing project. These corpora all had the same goal- collect salient lexical samples from multiple speakers within a region. It was made sure that speakers were native to their area and had not traveled elsewhere for an extended period of time in order to collect the most regionally specific data. It was also made sure that the interviews were exhaustive, going over hundred-page questionnaires made to elicit tokens that had the potential to be region-specific. By the end of the initial field recordings, LAGS had 911 primary speakers, but over 1000 speakers total, from the Southern United States (Kretzschmar et. al, 2011).

More recently, DASS was created to be a rigorous sample of 64 speakers from the larger LAGS corpus (Kretzschmar et. al, 2013). This goal was achieved by producing fully transcribed interviews of these participants, who were selected to best represent the demographics they came from. These transcriptions are especially beneficial, as they allow us to look at the context of a word in the speaker's conversation instead of just as a data point. This insight on the data of LAGS via the subset DASS also included measuring phonetic formant data from these interviews, instead of merely the phonetic transcriptions of each speaker's response. This again allows for a more nuanced understanding of dialectal variation, as we can view formant levels instead of just the phonemes that are elicited.

In my own research on LAMSAS in prior courses I found what one would typically expect, that certain lexical items were denser in the North than in the South and vice versa. However, one of the more surprising findings I had while browsing this corpus was that there were certain lexical items that seemed to be divided by the Appalachian Mountains. The best of this is the usage of 'mosquito hawk' as a lexical variant for 'dragon fly', as seen in Figure 1. Though this was not the South, it showed a potentially interesting trend.

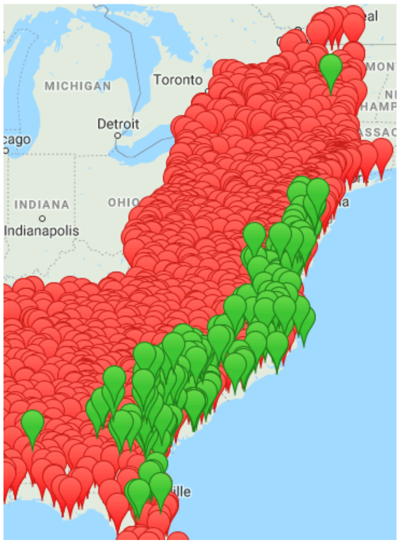

Figure 1

This led to me forming the question whether dialect was not simply influenced by the state, city, or county someone was from, but by the geography of their surroundings. It is feasible to believe that speakers from the Coast versus the Plains would utilize different lexicons. Lee Pederson, the director of the LAGS project, seemed to share the same opinion of this being a possibility:

> "The settlement of the Lower South, however, was distinctive as its terrain. The basic patterns were established by the expansion of the plantation culture from the Low Country in the southeast and the extension of the Highland culture from the north and east" (Pederson 16, 1974).

Here Pederson lays the groundwork to my thesis. The population of the South was initially stratified by different social groups, specifically those from the Highlands and those from the Low Country. The varieties subsequently found in these geographical areas are likely to reflect the social differences. Additionally, up until roughly the 1950's, different terrains meant different jobs available in the community (Beck et. al, 2007). As the South was originally an agrarian area slow to industrialization, many people held jobs that were unique to the terrain they inhabited. Farmers would gather in the Plains, fishermen the Coast, and cotton pickers the Delta. The creation of these communities that not only shared a history, but an industry, would very likely lead to a difference in dialect between them due to the jargon of their jobs and the tight-knit communities that were formed. Thus, I came to the conclusion that the human geography of the South likely influenced the dialect of these areas, both lexically and phonologically.

Though these geographical regions are well defined, and LAGS splits them into the regions of Piedmont, Piney Woods, Coast, Plains, Highlands, and Delta, it is much harder to reliably define the region an accent inhabits (Fig 2).

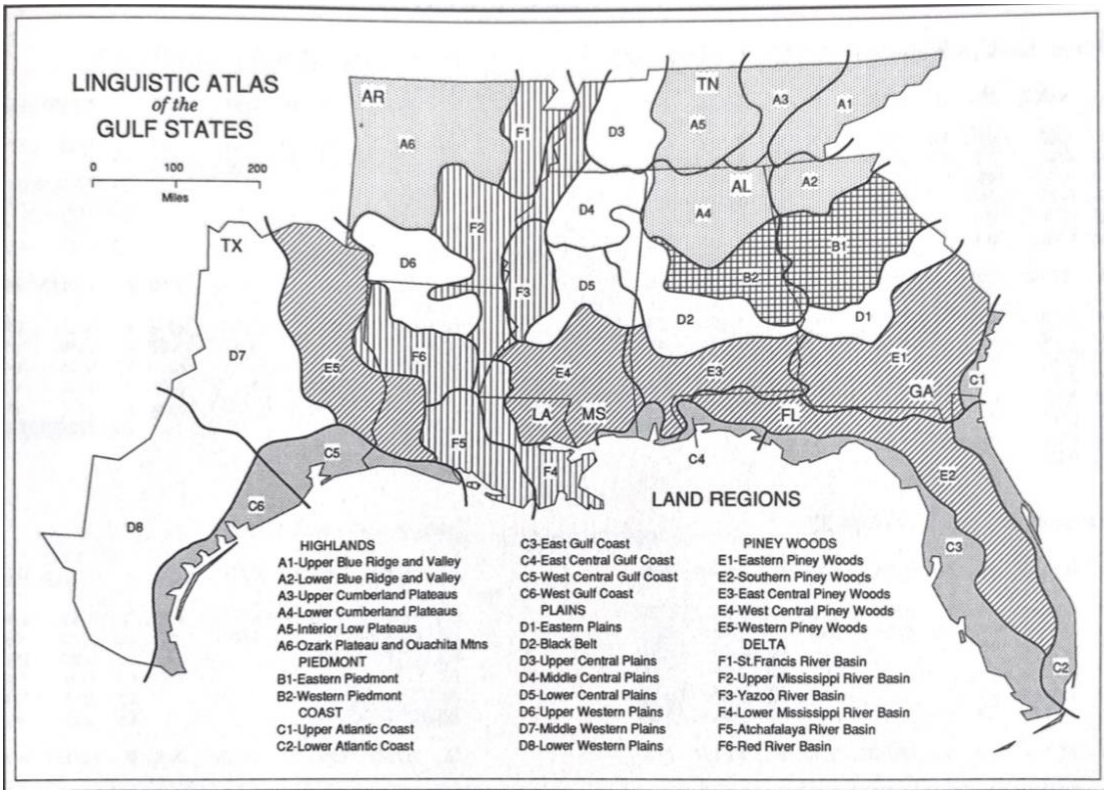

Figure 2

Defining it is not as simple as "finding the correct, or even defensible, boundaries of a region," as any assertion of hard dialect boundaries is bound to be flawed (Kretzschmar 130, 2003). For this reason, linguists researching dialect, such as Kretzschmar, tend to prefer to map dialect with gradients in order to show the variation possible within a region. This also implies that areas that are not shaded have the slight possibility of displaying certain dialectal variations. Figures 3 through 5 of the question "15 minutes til the next hour" are apt examples of this, showing that unshaded areas still have a 0 to .24 chance of displaying a certain lexical item. Thus, in this paper I am not attempting to assert that geographical region is the be-all-end-all of dialect variation, simply that it is one of the multiple variables that reliably influences dialect.

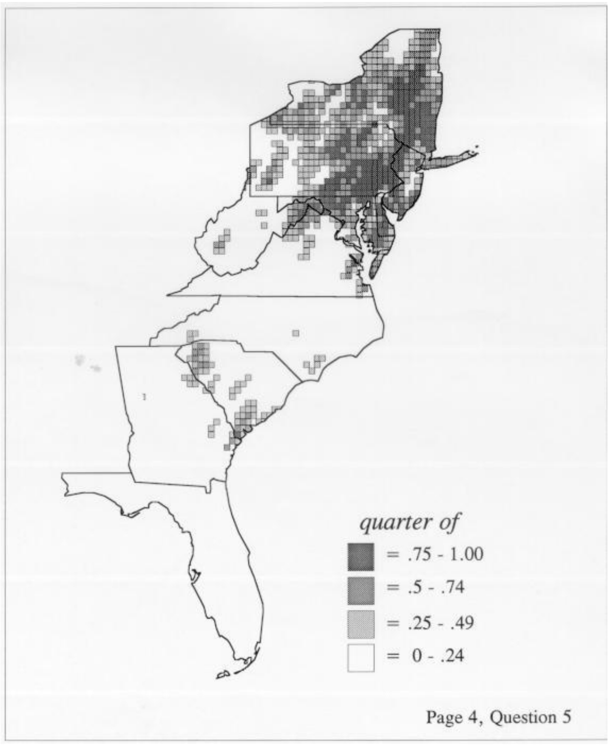

Figure 3

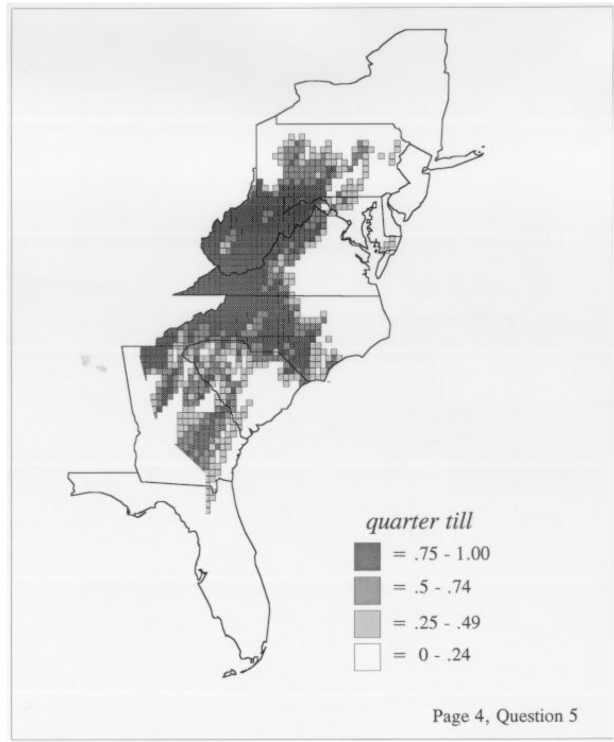

Figure 4

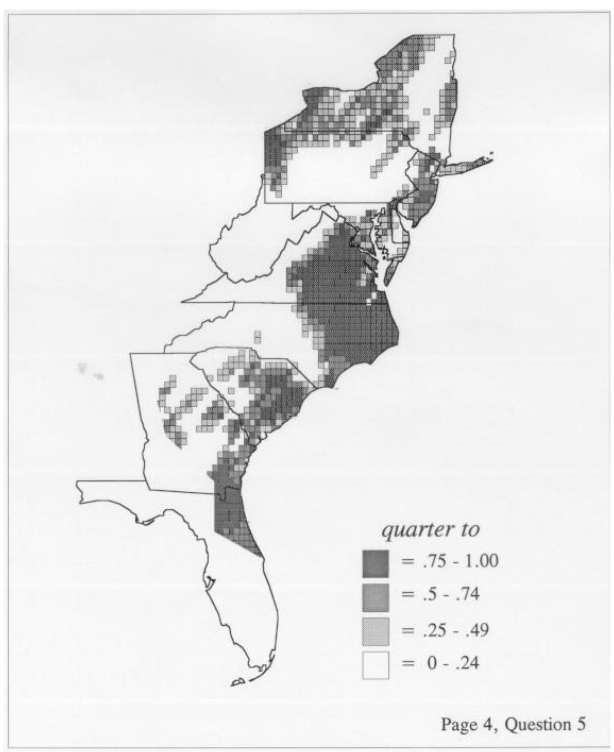

Figure 5

Finally, it should be of note that all of these corpora include data collected in the 20th century, and many of the speakers in it were born in the 19th century. While these corpora are important for the breadth of information they offer to researchers, the results derived from them should not be applied to the dialects found within the modern South. Instead, these findings should be treated as an analysis of a historical corpus.

METHODS

As stated before, I will explore the dialect of the South through both lexical and phonetic means. The lexical means currently consist of visualization of data only, as that is the most effective way to see the geographical spread of tokens. I hope to expand my understanding of the lexical data in the future via a creation of a corpus out of DASS (see Future Work). The phonetic data was processed using tools available in RStudio. This data processing for vowel formants, compared to the visualization of lexical items, allows for a more objective understanding of the data and the role different variables play. Since DASS is the only regional American corpus that has acoustic data available to researchers, that is the only corpus I analyzed via this method. Within DASS, the formants of 15 different vowel classes are charted. I specify vowel classes because some vowels charted are monophthongs and some are diphthongs. The vowel classes, their IPA equivalent, their position of voicing, and word examples are as follows in Table 1.

Table 1

Vowel	IPA	Position	Examples
AA	ɑː	open / back	larger, park
AE	æ	near open / front	dance, maps
AH	ə	mid / central	lunch, mother
AO	ɒ/ɔː	open & open mid / back	walked, corn
AW	aʊ	open / front → near close / near back	found, county
AY	aɪ	open / front → close / front	nice, bible
EH	e/ɛ	close mid & open mid / front	yellow, best
ER	ər	mid / central rhotic	person, doctor
EY	eɪ	close mid / front → near close / front	Cajun, paper
IH	ɪ	near close / near front	six, dinner
IY	iː	close / front	streets, anything
OW	əʊ	mid / central → near close / near back	don’t, told
OY	ɔi/oi	open mid / back → close / front	point, boiled
UH	ʊ	near close / near back	bushes, sugar
UW	uː	close / back	student, usually

However, only a handful of these vowel classes are notably changed in modern Southern speech. One of the most notable phonetic markers of Southern English is the shifting of the diphthong aɪ to a “glideless” long vowel, aː (Thomas, 2008) This is also found, to a lesser extent, in the diphthong ɔi/oi shifting to the monophthong ɔ/o before the alveolar lateral approximant /l/. Other notable markers are the shift of eɪ to ɛi/æi, the loss of rhoticity in ər, and the diphthongization of æ to æə, ɪ to ɪə, e to eiə, and iː to ɪi (304 – 305, 307). Due to this, I will focus my research largely on the results of the vowels AE, AY, EY, EH, IH, IY, and OY. I have purposefully excluded the vowel ER (the only rhotic vowel in DASS) from my analysis. This is because rhoticity in accents is shown by a markedly low F3, and in my own experience has been notoriously difficult to chart in relation to independent variables (Yan & Vaseghi, 2003, Dudley, 2019).

To chart the difference in vowel classes between geographical regions, I chose to analyze the F1 and F2 of each vowel class. As F1 relates to height of vowel and F2 relates to backness of vowel, running a statistical analysis on both of these formants will give us a fairly good understanding of whether their pronunciations change between speakers, and due to what factors. I also analyzed these formants separately, as the values are very different from one another and running a combined analysis would require me to include a predictor that accounts for these differences, which takes much trial and error. While both raw formant values and Lobanov-normalized values, which adjusts formant values on a speaker-by-speaker basis, were available to me, I chose simply to go with the raw data (Labov et. al, 2013). While the goal of Lobanov-normalized data is to remove physiological differences while keeping sociolinguistic differences, I ultimately came to the conclusion that working with the raw formant data would be most beneficial. For one, the merit of normalizing data is a topic of debate within the linguistic community. I also considered that keeping the data in its initial state would make its results more easily comparable to similar studies, as these other studies may work with raw acoustic data too or normalize them in a different way.

In order to find the model of best fit for these selected vowels I ran an lmer() model on the F1 and F2 of each of these vowels. I selected lmer() because it is one of the most comprehensive statistical models offered in RStudio, and DASS is an incredibly large and complex data set. This lmer() test contained all independent variables I hypothesized would have an influence on these formant levels.

1. Speaker (as a random variable)
2. Land_region
    a. Coast
    b. Delta
    c. Plains
    d. Piney Woods
    e. Piedmont
    f. Highlands
3. Duration
4. Sex
    a. Male
    b. Female
5. Ethnicity
    a. Black
    b. Non-black

6. Age_level

    a. 13 – 45

    b. 46 – 65

    c. 66 – 76

    d. 77 – 99

While I included multiple variables, the main variable I was focused on was land_region. This is the term used within the data to note the geographical region of the speaker, and is the variable which I have based the crux of my thesis on. I included the other variables here specifically because they have a strong possibility of influencing the data. For example, male speakers usually have lower formant values. There were other geographic variables that I deleted from the data, specifically state and locality (urban versus rural). I deleted these variables after testing them and finding them to be too similar to my land_region variable, resulting in collinearity. I also chose not to include more specific speaker-level data because I believed these would overfit the data. Examples of too-specific data were county, city, and birth year, as there were often only one or two speakers that shared the same value.

Though DASS is composed of a sample of speakers from the much larger LAGS corpus, DASS is still a very large data set. The sum of tokens within the data set equaled about 2 million, with 878,600 visible in the RStudio table. In order to show the distribution of tokens between vowel class and geographical region, Table 2 contains the number of tokens for each of these intersections as well as the sum of tokens for each variable.

Table 2

	AE	AY	EY	EH	IH	IY	OY	Total
Coast	6603	5048	6351	8241	8700	8766	261	43970
Delta	16244	11112	14518	17748	21031	16519	576	97748
Plains	20470	15297	18332	22451	23005	21387	701	121643

Piney Woods	14729	10653	12865	15183	16218	12596	536	82780
Piedmont	3537	2376	3027	4042	4560	3962	115	21619
Highlands	13898	9624	13031	16904	18247	15646	489	87839
Total	75481	54110	68124	84569	91761	78876	2678	455959

One of the most glaring differences in this data set is how few tokens OY has compared to the rest of the data. Though this is a large difference that may possibly have an impact on the way the data is processed, it is still a substantial number of tokens to work with, so I will keep it in my analysis of DASS. I also feel it is important to note that the regions Piedmont and Coast have a markedly lower number of tokens than the other regions. This is proportional with the LAGS data, as those are the smallest regional areas, but this could result in overfitted results in my data.

After identifying these independent variables I created an lmer() model that included these variables, then ran the model to see the significance codes for the *p*-values of each variable. The lowest *p*-value possible to have a variable be considered significant was $p > .05$. I also checked to make sure the Akaike Information Criterion, or AIC, was adequately low in the model of best fit compared to the other models. This is because the most complex model is not necessarily the best one, and AIC compares quality of models to one another and penalizes models that have too many variables. This allows us to analyze the tradeoff between goodness and complexity of a model (Levshina, 2015).

After creating the model of best fit, I ran the model through a bootstrapping test. This ensures that the model I created was not overfitted to the data I was using and could be adequately applied to other similar data (Levshina, 2015). In order for a model to be considered not overfitted, the optimism slope value must be > .05.

I intended to analyze these models via a Shapiro-Wilk test as well in order to check for normalcy of the data, but the sample size was too large to be able to run the test, as Shapiro-Wilk has a limit of 5,000 data tokens. Therefore I ran a Breusch-Pagan test (ncvTest() in R) to test for non-normalcy and heteroscedasticity, as it could handle larger data sets.

RESULTS

The below results show summary of the models run on these vowel classes rounded to the fifth decimal place, as well as the optimism result for overfitting via bootstrapping. Table 3 shows the key for the significance codes assigned to each *p*-value. Any *p*-value with a symbol next to it is considered statistically significant.

Table 3

p-value	code
0	***
0.001	**
0.01	*
0.05	.
0.1	

Data: DASSAE

Table 4

F1

<table>
<tr><td></td><td>"Estimate"</td><td>"Std. Error"</td><td>"df"</td><td>"t value"</td><td>"Pr(>|t|)"</td></tr>
<tr><td>"(Intercept)"</td><td>715.8721</td><td>27.61936</td><td>64.53993</td><td>25.91921</td><td>0 ***</td></tr>
<tr><td>"land_regionDelta"</td><td>-111.76575</td><td>27.74271</td><td>63.952</td><td>-4.02865</td><td>0.00015 ***</td></tr>
<tr><td>"land_regionHighlands"</td><td>-19.3678</td><td>28.35321</td><td>63.49383</td><td>-0.68309</td><td>0.49704</td></tr>
<tr><td>"land_regionPiedmont"</td><td>-39.38508</td><td>36.06574</td><td>63.48799</td><td>-1.09204</td><td>0.27894</td></tr>
<tr><td>"land_regionPiney Woods"</td><td>-71.51573</td><td>26.88486</td><td>64.61365</td><td>-2.66007</td><td>0.00984 **</td></tr>
<tr><td>"land_regionPlains"</td><td>-38.08411</td><td>28.10211</td><td>63.37535</td><td>-1.3552</td><td>0.18016</td></tr>
<tr><td>"age_level46-65 years old"</td><td>-19.24764</td><td>22.39556</td><td>63.39069</td><td>-0.85944</td><td>0.39334</td></tr>
<tr><td>"age_level66-76 years old"</td><td>-22.83244</td><td>18.54297</td><td>65.87072</td><td>-1.23133</td><td>0.22258</td></tr>
<tr><td>"age_level77-99 years old"</td><td>6.72171</td><td>18.25506</td><td>64.64109</td><td>0.36821</td><td>0.71392</td></tr>
<tr><td>"sexM"</td><td>-78.83917</td><td>14.91397</td><td>65.18055</td><td>-5.28626</td><td>0 ***</td></tr>
<tr><td>"ethnicityNon-Black"</td><td>20.94694</td><td>13.01597</td><td>67.94237</td><td>1.60933</td><td>0.11218</td></tr>
<tr><td>"dur"</td><td>187.82996</td><td>5.60805</td><td>75433.56956</td><td>33.49291</td><td>0 ***</td></tr>
</table>

Table 5

F2

<table>
<tr><td></td><td>"Estimate"</td><td>"Std. Error"</td><td>"df"</td><td>"t value"</td><td>"Pr(>|t|)"</td></tr>
</table>

"(Intercept)"	1841.98706	55.72297	64.14164	33.05616	0 ***
"land_regionDelta"	-1.93279	55.91936	63.3225	-0.03456	0.97254
"land_regionHighlands"	139.17283	57.10895	62.7255	2.43697	0.01766 *
"land_regionPiedmont"	173.41111	72.64314	62.72504	2.38716	0.02001 *
"land_regionPiney Woods"	105.24879	54.24586	64.19487	1.94022	0.05675 .
"land_regionPlains"	32.36664	56.59253	62.56928	0.57192	0.56942
"age_level46-65 years old"	24.3751	45.10171	62.59023	0.54045	0.59081
"age_level66-76 years old"	-53.78188	37.48505	65.83266	-1.43476	0.15609
"age_level77-99 years old"	59.28572	36.83539	64.25085	1.60948	0.11241
"sexM"	-235.00207	30.11825	64.95172	-7.80265	0 ***
"ethnicityNon-Black"	66.69578	26.39366	68.7499	2.52696	0.0138 *
"dur"	336.95055	13.43218	75439.34462	25.08533	0 ***

Table 6

DASSAY

F1

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	715.8721	27.61936	64.53993	25.91921	0 ***
"land_regionDelta"	-111.76575	27.74271	63.952	-4.02865	0.00015 ***
"land_regionHighlands"	-19.3678	28.35321	63.49383	-0.68309	0.49704
"land_regionPiedmont"	-39.38508	36.06574	63.48799	-1.09204	0.27894
"land_regionPiney Woods"	-71.51573	26.88486	64.61365	-2.66007	0.00984 **
"land_regionPlains"	-38.08411	28.10211	63.37535	-1.3552	0.18016
"age_level46-65 years old"	-19.24764	22.39556	63.39069	-0.85944	0.39334
"age_level66-76 years old"	-22.83244	18.54297	65.87072	-1.23133	0.22258
"age_level77-99 years old"	6.72171	18.25506	64.64109	0.36821	0.71392
"sexM"	-78.83917	14.91397	65.18055	-5.28626	0 ***
"ethnicityNon-Black"	20.94694	13.01597	67.94237	1.60933	0.11218
"dur"	187.82996	5.60805	75433.56956	33.49291	0 ***

Table 7

F2

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	1626.82429	49.38637	62.06903	32.94075	0 ***
"land_regionDelta"	-122.92324	49.52607	61.09733	-2.48199	0.01583 *
"land_regionHighlands"	-2.22077	50.56226	60.46082	-0.04392	0.96511
"land_regionPiedmont"	8.69204	64.31881	60.47386	0.13514	0.89295
"land_regionPiney Woods"	41.68568	48.05372	61.96999	0.86748	0.38903
"land_regionPlains"	-16.19661	50.09314	60.25771	-0.32333	0.74756
"age_level46-65 years old"	32.23018	39.92257	60.28015	0.80732	0.42266

"age_level66-76 years old"	-12.58548	33.2627	63.8631	-0.37837	0.70641
"age_level77-99 years old"	31.08715	32.67051	62.27924	0.95154	0.34501
"sexM"	-172.76315	26.68667	62.75208	-6.47376	0 ***
"ethnicityNon-Black"	36.11216	23.46258	67.20931	1.53914	0.12847
"dur"	11.41928	10.9116	54067.74362	1.04653	0.29532

DASSEY

Table 8

F1

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	646.65821	28.12114	63.84356	22.99545	0 ***
"land_regionDelta"	-90.26876	28.25123	63.29852	-3.19522	0.00218 **
"land_regionHighlands"	10.39701	28.87926	62.89154	0.36002	0.72004
"land_regionPiedmont"	2.80592	36.74297	62.93856	0.07637	0.93937
"land_regionPiney Woods"	-44.49406	27.36873	63.88063	-1.62573	0.10893
"land_regionPlains"	-6.89469	28.62509	62.78625	-0.24086	0.81045
"age_level46-65 years old"	-8.61769	22.8132	62.811	-0.37775	0.70689
"age_level66-76 years old"	-32.32545	18.89308	65.32211	-1.71097	0.09183 .
"age_level77-99 years old"	-6.52434	18.61108	64.23804	-0.35056	0.72706
"sexM"	-89.76333	15.17366	64.31126	-5.91573	0 ***
"ethnicityNon-Black"	20.77639	13.25223	67.20553	1.56777	0.12163
"dur"	211.55363	5.56733	68070.85917	37.99911	0 ***

Table 9

F2

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	1841.2994	50.14702	63.79328	36.71802	0 ***
"land_regionDelta"	-44.82005	50.25935	62.65327	-0.89178	0.37593
"land_regionHighlands"	30.23697	51.29136	61.93179	0.58951	0.55766
"land_regionPiedmont"	104.0654	65.27282	62.04851	1.59431	0.11595
"land_regionPiney Woods"	54.85203	48.80479	63.71406	1.12391	0.26527
"land_regionPlains"	-21.93919	50.81759	61.73935	-0.43172	0.66745
"age_level46-65 years old"	-1.72972	40.50426	61.78931	-0.0427	0.96607
"age_level66-76 years old"	-5.86643	33.87691	66.18396	-0.17317	0.86305
"age_level77-99 years old"	76.81834	33.23488	64.34921	2.31138	0.02403 *
"sexM"	-178.94692	27.10615	64.54429	-6.60171	0 ***
"ethnicityNon-Black"	40.38618	23.94484	70.24632	1.68663	0.09611 .
"dur"	303.3025	14.40344	68082.00822	21.05764	0 *

DASSEH

Table 10

F1

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	621.54684	23.21737	64.58913	26.77077	0 ***
"land_regionDelta"	-96.64682	23.32073	63.99211	-4.14425	1e-04 ***
"land_regionHighlands"	-6.86009	23.83859	63.57501	-0.28777	0.77446
"land_regionPiedmont"	-24.95928	30.33102	63.63275	-0.8229	0.41364
"land_regionPiney Woods"	-63.99138	22.59165	64.5731	-2.83252	0.00615 **
"land_regionPlains"	-23.11999	23.62935	63.47466	-0.97844	0.33157
"age_level46-65 years old"	-18.35829	18.83009	63.47694	-0.97494	0.33329
"age_level66-76 years old"	-9.42429	15.6033	66.14392	-0.60399	0.54791
"age_level77-99 years old"	9.72717	15.37069	65.05471	0.63284	0.52906
"sexM"	-64.96547	12.52329	64.97467	-5.18757	0 ***
"ethnicityNon-Black"	17.22342	10.94559	68.0825	1.57355	0.12023
"dur"	150.92032	5.42128	84522.98259	27.83851	0 ***

Table 11

F2

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	1804.64704	50.07594	64.86735	36.03821	0 ***
"land_regionDelta"	6.92556	50.21025	63.8152	0.13793	0.89073
"land_regionHighlands"	123.04796	51.26469	63.16149	2.40025	0.01934 *
"land_regionPiedmont"	196.52732	65.23853	63.27285	3.01244	0.00372 **
"land_regionPiney Woods"	97.693	48.71992	64.74001	2.0052	0.04913 *
"land_regionPlains"	47.86433	50.8001	63.00204	0.94221	0.34968
"age_level46-65 years old"	32.16631	40.48254	63.00502	0.79457	0.42985
"age_level66-76 years old"	-49.05743	33.79014	67.12905	-1.45183	0.15121
"age_level77-99 years old"	31.83883	33.19109	65.4839	0.95926	0.34096
"sexM"	-243.81363	27.0379	65.41512	-9.01748	0 ***
"ethnicityNon-Black"	54.18412	23.83043	70.6338	2.27374	0.02602 *
"dur"	644.64909	15.49168	84534.67626	41.61261	0 ***

DASSIH

Table 12

F1

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	586.2988	25.7009	64.91094	22.81239	0 ***

"land_regionDelta"	-101.03844	25.79589	64.12366	-3.91684	0.00022 ***
"land_regionHighlands"	-1.39754	26.34922	63.55376	-0.05304	0.95787
"land_regionPiedmont"	-17.77083	33.51441	63.53569	-0.53024	0.59779
"land_regionPiney Woods"	-69.26021	25.01327	64.91071	-2.76894	0.00732 **
"land_regionPlains"	-14.663	26.11271	63.41042	-0.56153	0.57642
"age_level46-65 years old"	-11.8274	20.80696	63.38713	-0.56843	0.57175
"age_level66-76 years old"	-4.33477	17.27839	66.49771	-0.25088	0.80268
"age_level77-99 years old"	5.83394	16.99112	65.05298	0.34335	0.73244
"sexM"	-56.13743	13.88683	65.66603	-4.0425	0.00014 ***
"ethnicityNon-Black"	11.97001	12.15872	69.28587	0.98448	0.32831
"dur"	-221.52991	7.79268	91707.35184	-28.42796	0 ***

Table 13

F2

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	1726.15801	56.82722	65.06991	30.37555	0 ***
"land_regionDelta"	58.44761	56.96042	63.94176	1.02611	0.30871
"land_regionHighlands"	139.3088	58.12702	63.19728	2.39663	0.01952 *
"land_regionPiedmont"	210.49972	73.93186	63.18429	2.84721	0.00594 **
"land_regionPiney Woods"	120.77921	55.30395	64.98525	2.18392	0.03258 *
"land_regionPlains"	70.57712	57.59128	63.00731	1.22548	0.22495
"age_level46-65 years old"	46.7472	45.88761	62.97334	1.01873	0.31223
"age_level66-76 years old"	-37.96732	38.29848	67.05974	-0.99135	0.32508
"age_level77-99 years old"	62.67299	37.57677	65.21654	1.66787	0.10014
"sexM"	-231.76836	30.74141	66.00851	-7.53929	0 ***
"ethnicityNon-Black"	66.30782	27.06908	71.09976	2.44958	0.01677 *
"dur"	1691.15895	20.68626	91711.59243	81.75275	0 ***

DASSIY

Table 14

F1

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	530.06799	26.9141	64.55375	19.69481	0 ***
"land_regionDelta"	-88.79106	27.03257	63.94802	-3.2846	0.00166 **
"land_regionHighlands"	6.41954	27.63017	63.50855	0.23234	0.81702
"land_regionPiedmont"	1.00641	35.14791	63.51454	0.02863	0.97725
"land_regionPiney Woods"	-54.5816	26.18919	64.55002	-2.08413	0.04111 *
"land_regionPlains"	-7.24912	27.3873	63.40467	-0.26469	0.79211
"age_level46-65 years old"	-12.01844	21.82234	63.37927	-0.55074	0.58375
"age_level66-76 years old"	-7.78039	18.05236	65.66728	-0.43099	0.66789

"age_level77-99 years old"	6.95992	17.77936	64.52244	0.39146	0.69675
"sexM"	-58.71897	14.52261	65.02548	-4.04328	0.00014 ***
"ethnicityNon-Black"	14.91588	12.66493	67.57708	1.17773	0.24304
"dur"	-225.05998	5.53659	78825.15784	-40.64959	0 ***

Table 15

F2

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	1857.72582	67.36942	64.93399	27.57521	0 ***
"land_regionDelta"	88.65991	67.62563	64.17217	1.31104	0.19452
"land_regionHighlands"	152.638	69.09089	63.64043	2.20923	0.03076 *
"land_regionPiedmont"	229.30156	87.89008	63.65133	2.60896	0.01131 *
"land_regionPiney Woods"	175.64377	65.55409	64.9076	2.67937	0.00934 **
"land_regionPlains"	52.05826	68.47653	63.51376	0.76024	0.44993
"age_level46-65 years old"	42.0586	54.56105	63.48207	0.77085	0.44365
"age_level66-76 years old"	-44.65679	45.23446	66.25545	-0.98723	0.32712
"age_level77-99 years old"	74.91187	44.50237	64.87977	1.68332	0.09712 .
"sexM"	-216.82785	36.36807	65.48967	-5.96204	0 ***
"ethnicityNon-Black"	84.29091	31.7913	68.66499	2.65138	0.00995 **
"dur"	1255.9059	15.50898	78828.48635	80.97927	0 ***

DASSOY

Table 16

F1

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	653.55933	26.65867	73.88936	24.51583	0 ***
"land_regionDelta"	-88.66958	25.73364	64.81105	-3.44567	0.001 **
"land_regionHighlands"	2.87633	25.96824	64.0777	0.11076	0.91215
"land_regionPiedmont"	-1.20262	33.18882	65.52566	-0.03624	0.9712
"land_regionPiney Woods"	-57.41323	25.53204	68.37133	-2.24867	0.02776 *
"land_regionPlains"	-11.36173	25.72673	63.95942	-0.44163	0.66025
"age_level46-65 years old"	-11.92593	20.42949	63.57972	-0.58376	0.56145
"age_level66-76 years old"	-10.12855	18.04111	68.0525	-0.56141	0.57636
"age_level77-99 years old"	-3.39725	17.47346	67.82336	-0.19442	0.84643
"sexM"	-85.24132	14.43592	70.68683	-5.90481	0 ***
"ethnicityNon-Black"	22.92013	14.07	88.20572	1.62901	0.10688
"dur"	16.58671	21.87266	2649.37138	0.75833	0.44832

Table 17

F2

	"Estimate"	"Std. Error"	"df"	"t value"	"Pr(>|t|)"
"(Intercept)"	1370.23015	47.44231	76.13958	28.88203	0 ***
"land_regionDelta"	-135.2005	44.53941	63.01224	-3.03553	0.00349 **
"land_regionHighlands"	-84.33692	44.92881	63.57967	-1.87712	0.06509 .
"land_regionPiedmont"	-115.01882	57.769	66.70749	-1.99101	0.05058 .
"land_regionPiney Woods"	-102.91438	44.66512	67.31478	-2.30413	0.02431 *
"land_regionPlains"	-71.44845	44.46673	62.87245	-1.60678	0.11311
"age_level46-65 years old"	6.4238	35.31513	63.66723	0.1819	0.85624
"age_level66-76 years old"	-23.93058	31.41046	64.20771	-0.76187	0.44893
"age_level77-99 years old"	-1.29301	30.41916	64.12022	-0.04251	0.96623
"sexM"	-135.9247	25.37503	68.81878	-5.35663	0 ***
"ethnicityNon-Black"	-22.14502	25.54764	81.0035	-0.86681	0.3886
"dur"	-657.4053	54.22583	2669.06334	-12.12347	0 ***

In these tables we can see which of the variables chosen had the most consistent and profound impact on the data via the *p*-value. The sex of the speaker was an influential factor in every single model produced, which was unsurprising. As stated in my methods section, men tend to have deeper voices and thus lower formant values, resulting in a strong correlation between the speaker's sex and their formant heights. Duration was also shown to influence the F1 and F2 of a vowel 12 out of 14 times. This signals what is likely a strong correlation between the duration of a vowel class and the height or backness of it.

However, my primary focus in these models was to see if the variable of land_region had an impact on the formant values. Table 18 shows the *p*-value significance for all geographical regions except for Coast, which was the first factor in the .

Table 18

	Significant	Not Significant
Delta	9	5
Highlands	5	9
Piedmont	5	9
Piney Woods	11	3
Plains	0	14

This shows us an interesting split in the geographical region of a speaker and how affected their formant values were by the region they lived in. Among the five, the area with the most significantly different vowels is the Piney Woods area. Looking back to Figure 2, we see that this region covers the southern portion of Georgia, Alabama, and Mississippi, as well as the inland areas of Florida and East Texas. The geographical region with the second most significantly different vowels was the Delta, which follows the Mississippi River through most of Louisiana, as well as Arkansas, Mississippi, and Tennessee. Highlands and Piedmont were only found to be significant in 5 out of 14 models, and Plains was found to be significant in none. This is disappointing, but from the results of Piney Woods and Delta we can assume that at least some geographical regions have a pronounced impact on dialect.

While the reason for these geographical areas having more of an influence on dialect are no doubt complex, I have a hypothesis as to why the Delta region is so markedly different. A majority of the Delta is based in Louisiana, which has a rich history that is distinct from a majority of the south. Southern Louisiana specifically is known for its Cajun and Creole heritage, which has a language and culture distinct from the rest of the South due to its bilingual French history (Dubois & Horvath 36, 2003). This distinct dialect and accent could have possibly impacted the acoustic formants of the speakers interviewed, resulting in the statistical significance found. This is all speculative, as the Delta is much larger than South Louisiana, but it is a possibility that could be looked at for why the acoustic data is so different. Unfortunately, I do not have a strong hypothesis as to why the Piney Woods area has such strong significance within the data.

Since the data yielded such interesting results, I decided to plot the phonetic data on a map of the South to see what the formants looked like visually. Figures 6 through 19 show us the mapping of F1 and F2 averages from DASS's data. I have mapped the F1 and F2 for each vowel separately because F1 and F2 are independent from one another. They also have entirely different values, which would make charting them on the same map complicated and ineffective. I also chose to show formant averages for each place on the map. Data are not linear so this is not a perfect solution, since it averages the distributions of each speaker. However, this is the simplest way to show the vowel distribution in this region of America, as these dots would overlap one another if shown geographically. Charting each of the tens of thousands of points on this map would be as ineffective as charting F1 and F2 on the same map.

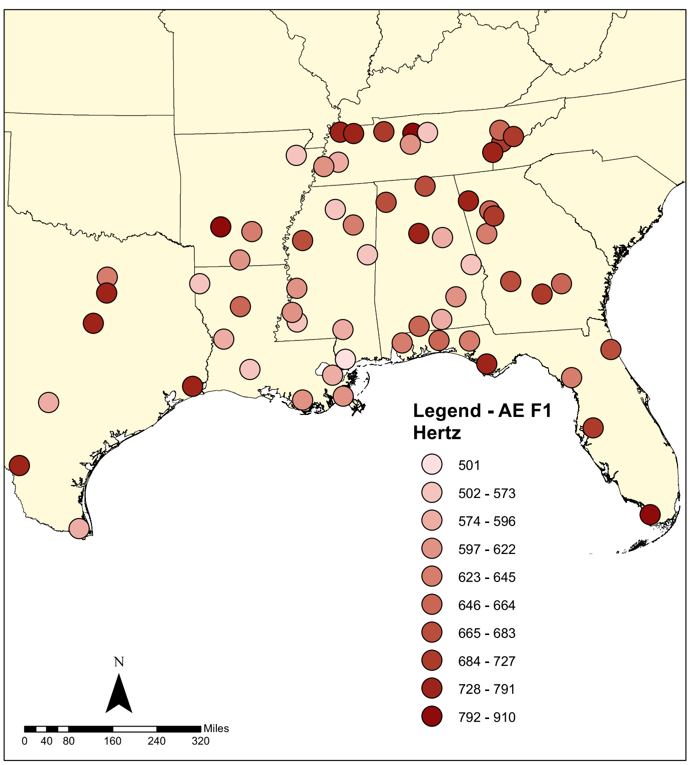

Figure 6

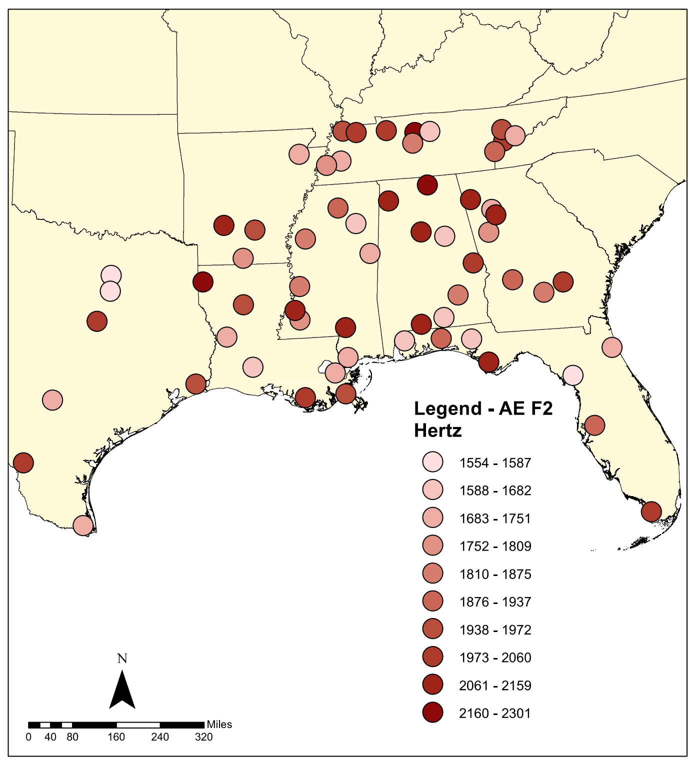

Figure 7

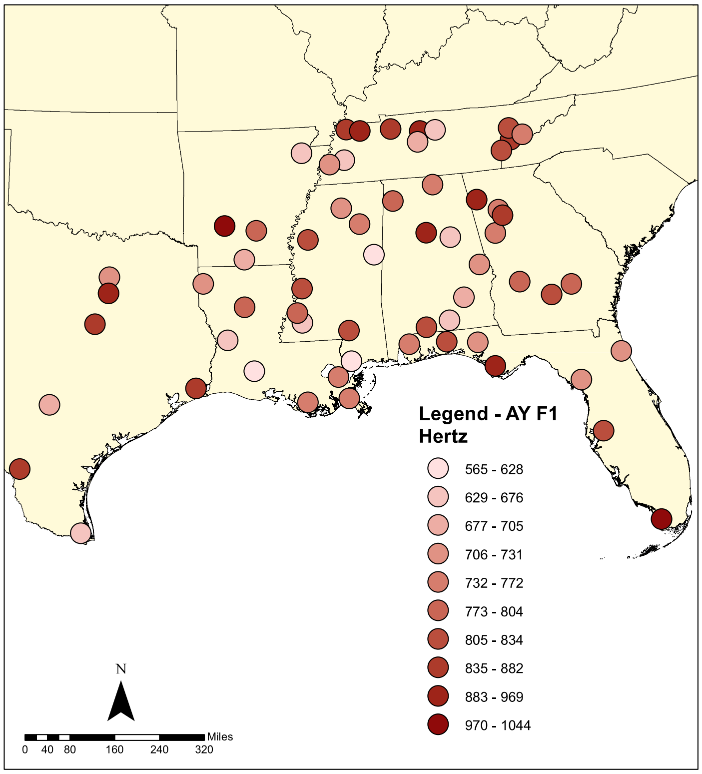

Figure 8

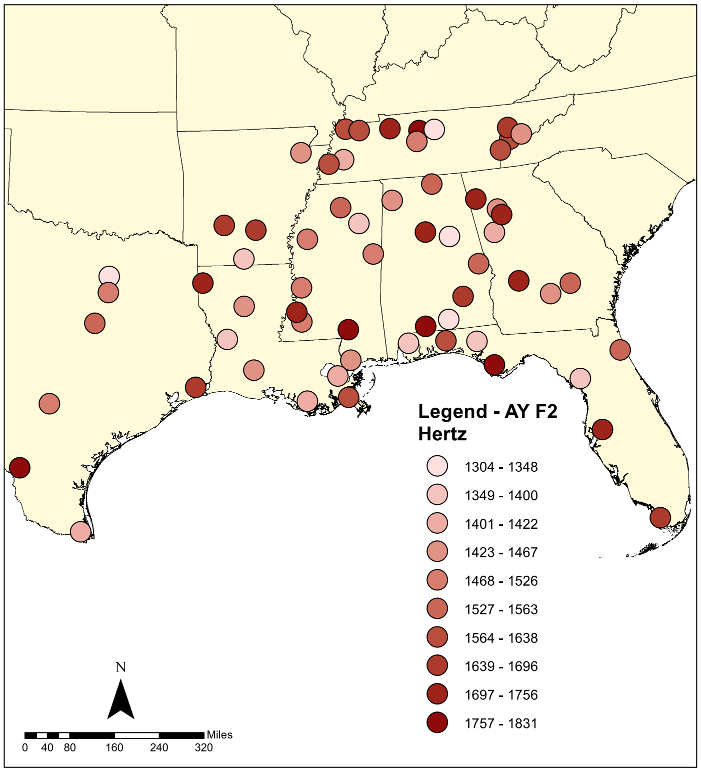

Figure 9

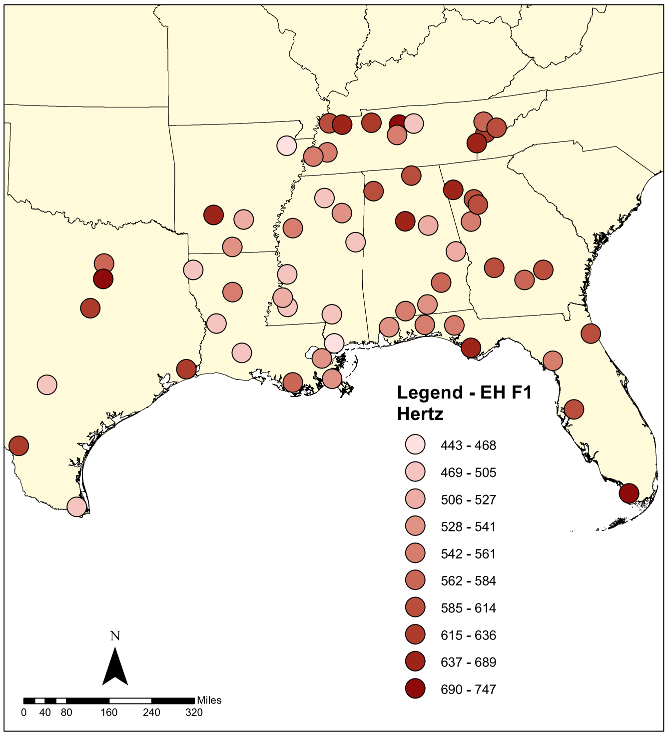

Figure 10

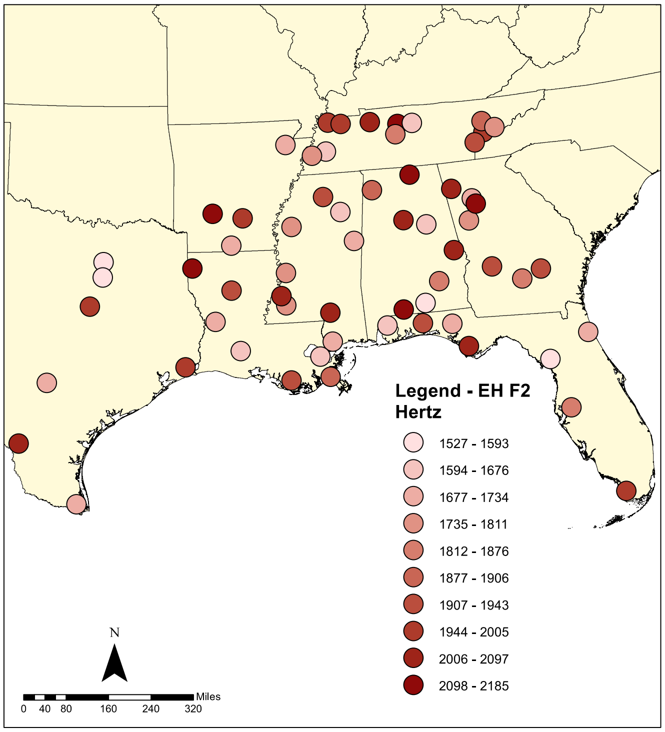

Figure 11

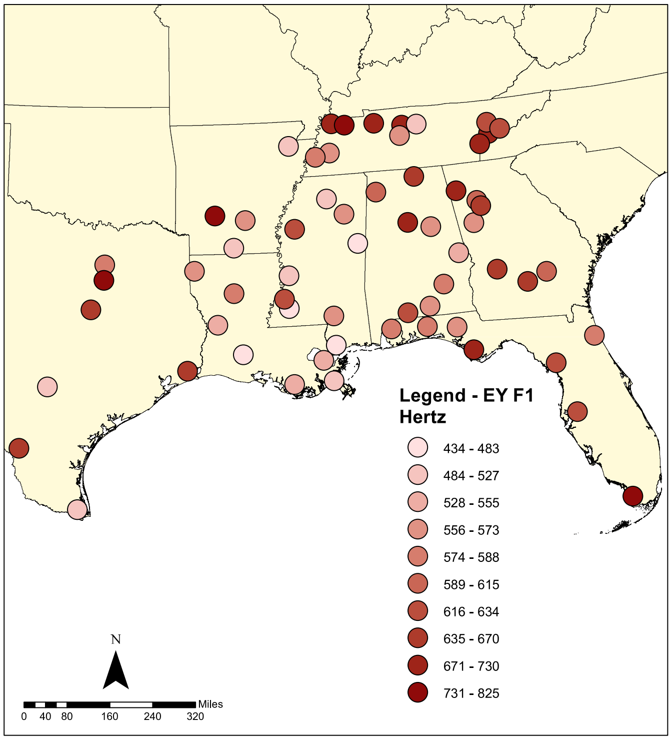

Figure 12

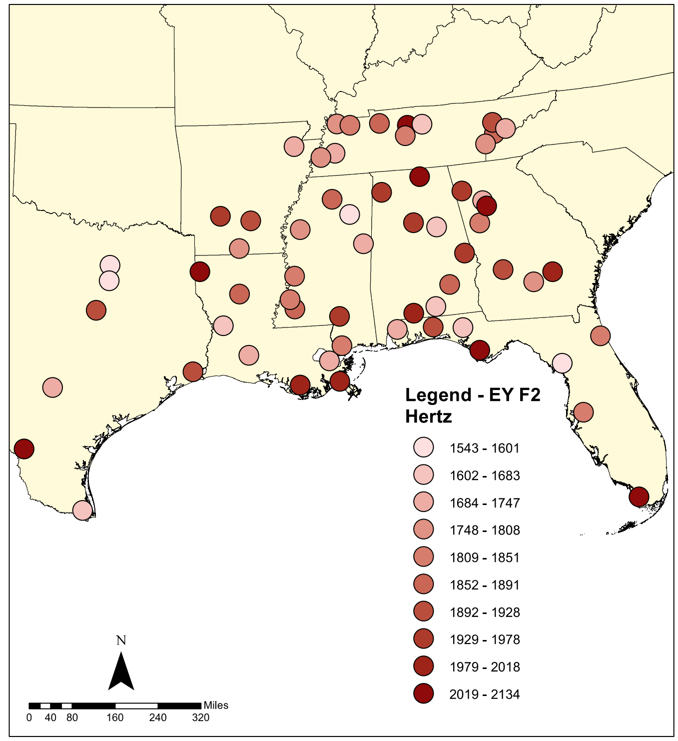

Figure 13

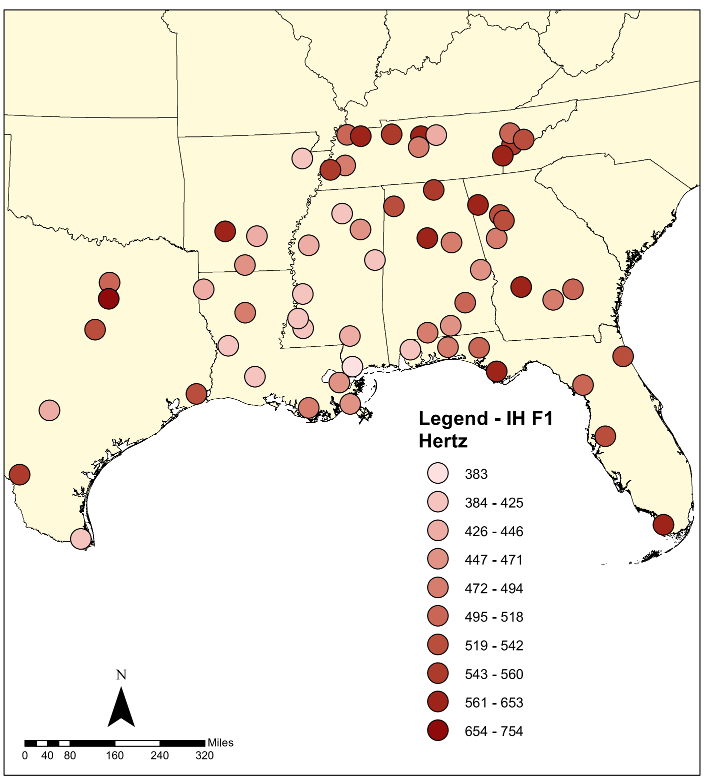

Figure 14

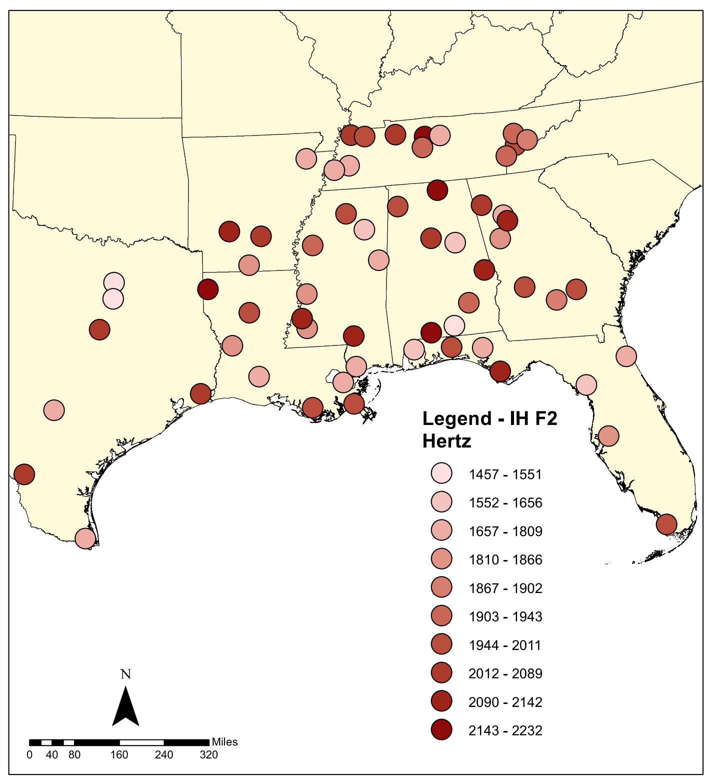

Figure 15

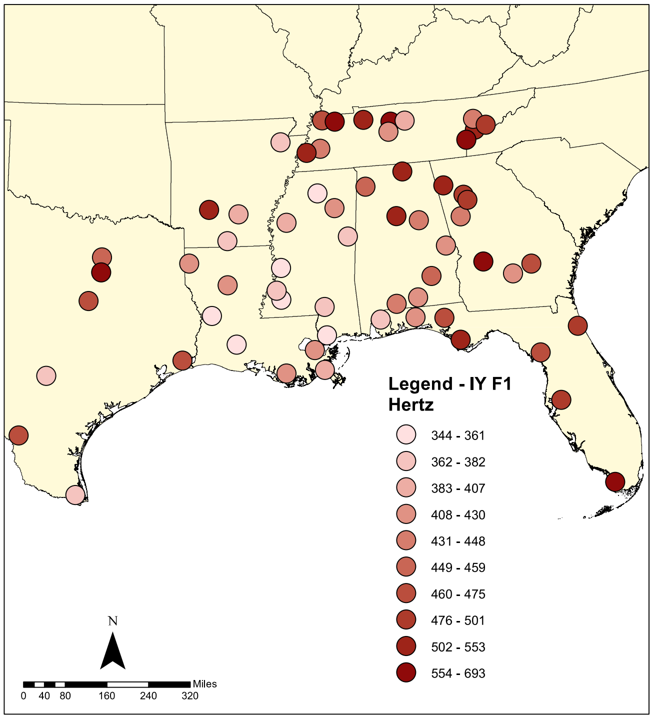

Figure 16

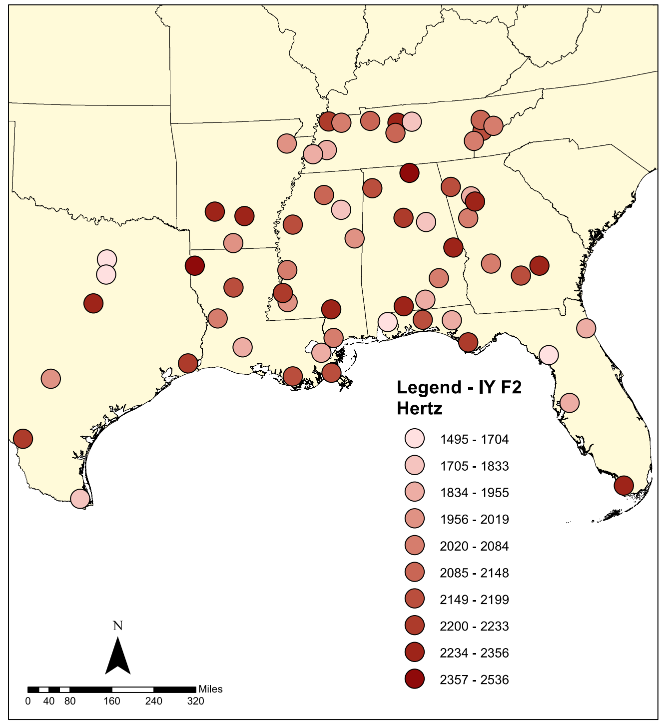

Figure 17

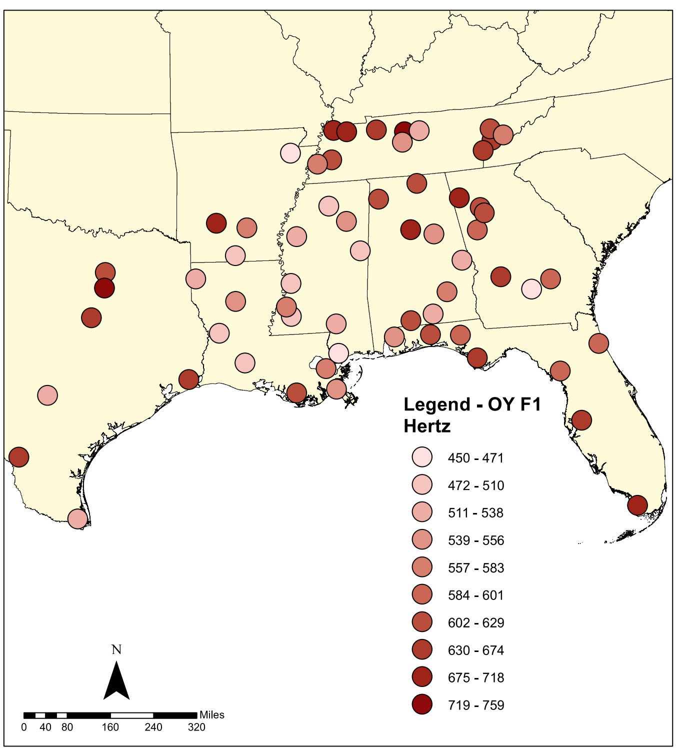

Figure 18

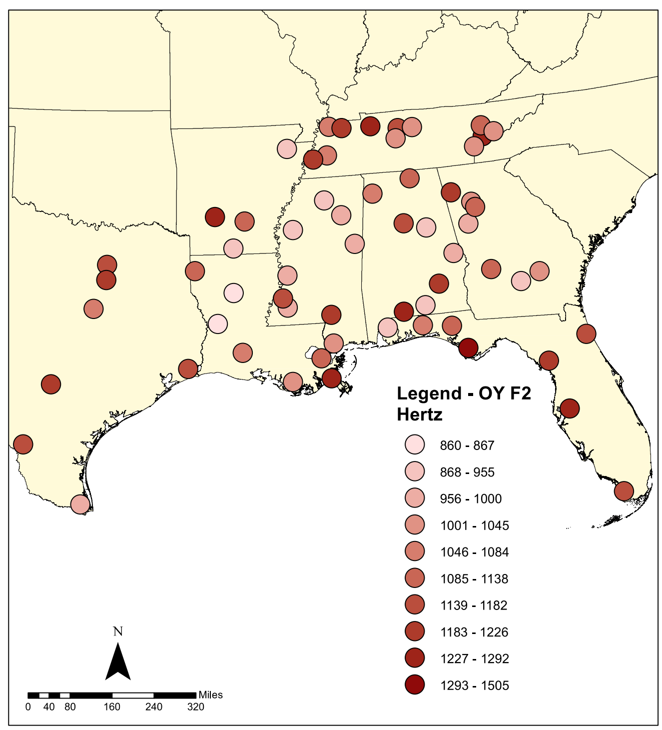

Figure 19

These maps allow us to visualize the data distribution of F1 and F2 for these different vowel classes. For example, take Figure 16. This figure shows lower F1 utterances taking place around the lower Mississippi River area, especially compared to the Highlands and Plains. This leads us to an inference that the IY vowel is pronounced higher than in these other areas. If we corroborate this with Table 14 we do see that the Delta area has a significant effect on the pronunciation of the data. Similar comparisons can be made between the maps given and their respective statistical tables to see vowel heightening and lowering as it relates to the geographical region of a speaker.

While it is great that I found models with statistically significant variables, it is important to see if these models are overfitted to the data. In order to do this, I ran my aforementioned bootstrapping test to see what the optimism scores came out to be.

Table 19

DASSAE

F1

[1] land_region age_level sex ethnicity dur

	index.orig	training	test	optimism	index.corrected	n
R-square	0.2326	0.2324	0.2324	-0.0001	0.2326	200
MSE	12011.5546	12002.6391	12013.5047	-10.8657	12022.4203	200
g	68.3798	68.3191	68.3659	-0.0468	68.4266	200
Intercept	0.0000	0.0000	-0.4024	0.4024	-0.4024	200
Slope	1.0000	1.0000	1.0007	-0.0007	1.0007	200

Table 20

F2

[1] land_region age_level sex ethnicity dur

	index.orig	training	test	optimism	index.corrected	n
R-square	0.2437	0.2440	0.2435	0.0005	0.2432	200
MSE	61273.7188	61232.5138	61283.6805	-51.1667	61324.8855	200
g	159.5319	159.6318	159.4933	0.1385	159.3934	200
Intercept	0.0000	0.0000	1.2175	-1.2175	1.2175	200
Slope	1.0000	1.0000	0.9993	0.0007	0.9993	200

Table 21

DASSAY

F1

[1] land_region age_level sex ethnicity dur

	index.orig	training	test	optimism	index.corrected	n
R-square	0.2908	0.2911	0.2907	0.0005	0.2904	200
MSE	16009.5569	15997.3245	16013.0937	-15.7692	16025.3261	200
g	91.8614	91.8879	91.8420	0.0459	91.8155	200
Intercept	0.0000	0.0000	0.4420	-0.4420	0.4420	200
Slope	1.0000	1.0000	0.9995	0.0005	0.9995	200

Table 22

F2

[1] land_region age_level sex ethnicity dur

	index.orig	training	test	optimism	index.corrected	n
R-square	0.2054	0.2056	0.2052	0.0004	0.2051	200
MSE	39513.2553	39511.2802	39523.0827	-11.8025	39525.0578	200
g	110.2917	110.3344	110.2395	0.0949	110.1969	200
Intercept	0.0000	0.0000	1.6703	-1.6703	1.6703	200
Slope	1.0000	1.0000	0.9989	0.0011	0.9989	200

Table 23

DASSEH

F1

[1] land_region age_level sex ethnicity dur

	index.orig	training	test	optimism	index.corrected	n
R-square	0.2188	0.2192	0.2187	0.0005	0.2183	200
MSE	10368.3402	10358.2275	10369.7593	-11.5318	10379.8720	200
g	61.6295	61.6623	61.6143	0.0480	61.5815	200
Intercept	0.0000	0.0000	0.4565	-0.4565	0.4565	200
Slope	1.0000	1.0000	0.9993	0.0007	0.9993	200

Table 24

F2

```
[1] land_region age_level    sex          ethnicity    dur
           index.orig   training       test optimism index.corrected   n
R-square       0.2155     0.2157     0.2153   0.0004          0.2151 200
MSE        74495.9012 74422.0873 74507.0484 -84.9611      74580.8623 200
g            162.1468   162.1829   162.0992   0.0836        162.0632 200
Intercept      0.0000     0.0000     0.9358  -0.9358          0.9358 200
Slope          1.0000     1.0000     0.9995   0.0005          0.9995 200
```

Table 25

DASSEY

F1

```
[1] land_region age_level    sex          ethnicity    dur
           index.orig   training       test optimism index.corrected   n
R-square       0.2802     0.2805     0.2801   0.0004          0.2798 200
MSE        11936.6211 11932.5708 11938.7532  -6.1824      11942.8035 200
g             78.0602    78.0981    78.0426   0.0555         78.0047 200
Intercept      0.0000     0.0000     0.3901  -0.3901          0.3901 200
Slope          1.0000     1.0000     0.9993   0.0007          0.9993 200
```

Table 26

F2

```
[1] land_region age_level    sex          ethnicity    dur
           index.orig   training       test optimism index.corrected   n
R-square       0.1559     0.1562     0.1557   0.0005          0.1554 200
MSE        65169.6392 65133.4302 65182.4362 -49.0060      65218.6452 200
g            126.1844   126.2932   126.1256   0.1676        126.0168 200
Intercept      0.0000     0.0000     2.4303  -2.4303          2.4303 200
Slope          1.0000     1.0000     0.9987   0.0013          0.9987 200
```

Table 27

DASSIH

F1

```
[1] land_region age_level    sex          ethnicity    dur
           index.orig   training       test optimism index.corrected   n
R-square       0.1622     0.1622     0.1620   0.0001          0.1620 200
MSE        13006.0405 12995.3907 13007.6452 -12.2546      13018.2951 200
g             57.1122    57.0963    57.0971  -0.0009         57.1130 200
Intercept      0.0000     0.0000     0.0305  -0.0305          0.0305 200
Slope          1.0000     1.0000     1.0000   0.0000          1.0000 200
```

Table 28

F2

```
[1] land_region age_level    sex          ethnicity   dur
           index.orig   training        test optimism index.corrected   n
R-square       0.2296     0.2298      0.2295   0.0003          0.2293 200
MSE        81069.1133 81093.1968 81080.3481  12.8488      81056.2646 200
g            176.8406   176.9484    176.8002   0.1482        176.6924 200
Intercept      0.0000     0.0000      1.6063  -1.6063          1.6063 200
Slope          1.0000     1.0000      0.9992   0.0008          0.9992 200
```

Table 29

DASSIY

F1

```
[1] land_region age_level    sex          ethnicity   dur
           index.orig   training        test optimism index.corrected   n
R-square       0.1811     0.1813      0.1810   0.0004          0.1807 200
MSE        11212.4704 11213.6022 11214.1563  -0.5541      11213.0244 200
g             56.9245    56.9719     56.9092   0.0628         56.8617 200
Intercept      0.0000     0.0000      0.3614  -0.3614          0.3614 200
Slope          1.0000     1.0000      0.9992   0.0008          0.9992 200
```

Table 30

F2

```
[1] land_region age_level    sex          ethnicity   dur
           index.orig   training        test optimism index.corrected   n
R-square       0.2589     0.2589      0.2587   0.0002          0.2587 200
MSE        81013.0926 81016.6255 81026.3960  -9.7706      81022.8631 200
g            192.3911   192.4230    192.3550   0.0681        192.3230 200
Intercept      0.0000     0.0000      0.3548  -0.3548          0.3548 200
Slope          1.0000     1.0000      0.9998   0.0002          0.9998 200
```

Table 31

DASSOY

F1

```
[1] land_region age_level   sex
```

	index.orig	training	test	optimism	index.corrected	n
R-square	0.2941	0.2960	0.2911	0.0048	0.2892	200
MSE	8717.4746	8708.8505	8753.6923	-44.8417	8762.3164	200
g	68.0749	68.2753	67.7741	0.5012	67.5737	200
Intercept	0.0000	0.0000	3.5642	-3.5642	3.5642	200
Slope	1.0000	1.0000	0.9933	0.0067	0.9933	200

Table 32

F2

```
[1] land_region sex         ethnicity   dur
```

	index.orig	training	test	optimism	index.corrected	n
R-square	0.1596	0.1634	0.1572	0.0063	0.1534	200
MSE	44398.2082	44368.6702	44528.7639	-160.0937	44558.3019	200
g	104.6829	106.0062	104.0661	1.9401	102.7428	200
Intercept	0.0000	0.0000	17.7253	-17.7253	17.7253	200
Slope	1.0000	1.0000	0.9830	0.0170	0.9830	200

In these tables we are looking at the optimism score for R-square and Slope. From the bootstrapping of this data I found that none of the models were overfitted to the data. This is relieving but not surprising, as the data sample was so large. In order to check for heteroscedasticity, or that my data has equal variance from the line estimated in my models, I ran the Breusch-Pagan tests for each of my models, as shown below.

DASSAE

Table 33

F1 - Chisquare = 826.278, Df = 1, p = < 2.22e-16

F2 - Chisquare = 759.4459, Df = 1, p = < 2.22e-16

DASSAY

Table 34

F1 - Chisquare = 329.8354, Df = 1, p = < 2.22e-16

F2 - Chisquare = 331.6828, Df = 1, p = < 2.22e-16

DASSEH

Table 35

F1 - Chisquare = 175.8279, Df = 1, p = < 2.22e-16

F2 - Chisquare = 378.1065, Df = 1, p = < 2.22e-16

DASSEY

Table 36

F1 - Chisquare = 66.30199, Df = 1, p = 3.8687e-16

F2 - Chisquare = 366.494, Df = 1, p = < 2.22e-16

DASSIH

Table 37

F1 - Chisquare = 921.5451, Df = 1, p = < 2.22e-16

F2 - Chisquare = 785.2797, Df = 1, p = < 2.22e-16

DASSIY

Table 38

F1 - Chisquare = 497.3479, Df = 1, p = < 2.22e-16

F2 - Chisquare = 717.4345, Df = 1, p = < 2.22e-16

DASSOY

Table 39

F1 - Chisquare = 6.985301, Df = 1, p = 0.0082182

F2 - Chisquare = 5.96775, Df = 1, p = 0.01457

All of these results show similar findings- the data has a low p value, indicating heteroscedasticity. These two tests show us that the models I have created are not overfitted, but that the data within them is abnormally distributed. This is to be expected for most humanities data, so this finding is disappointing but not entirely surprising.

When comparing the phonetic data between regions, we could also see via vowel spaces plotted with GGPlot that there was a slight shift in pronunciations. As OY is my smallest dataset and less visually overwhelming, Figure 20 and 21 show the difference in the vowel plots of the Delta and Plains. I chose these two geographical locations to compare because the Delta region is one whose dialect is most influenced by geography, and Plains the least.

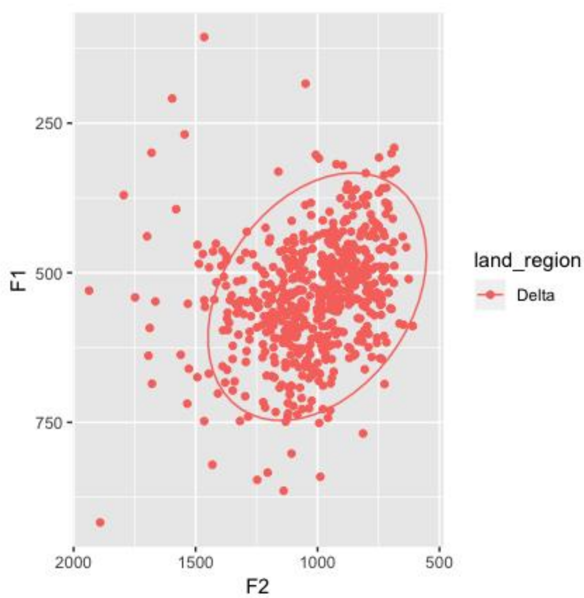

Figure 20

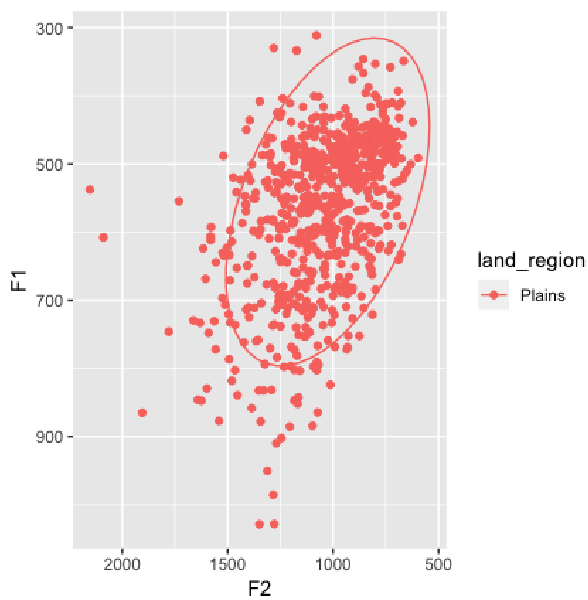

Figure 21

We see a distinct shift in the nucleus of this vowel class between these two regions, with a trend of the vowel's F1 being heightened in the Delta. Looking at larger datasets, it can be more difficult to determine if there is a distinct shift in the vowel's nucleus over geographical regions. To remedy this I created a Point Pattern Analysis plot from the Gazetteer of Southern Vowels (or GSV), an online source for DASS's acoustical data (Stanley 2019). This site does not allow comparison by geographical region, so instead I created a PPA plot of two demographically similar diverse speakers, one from the Plains and one from the Delta. Both speakers are non-black men aged 77 to 99, living in an urban area with 11+ years of schooling. Below are their plots for the AE vowel. The Delta speaker is Figure 22 and Plains is Figure 23.

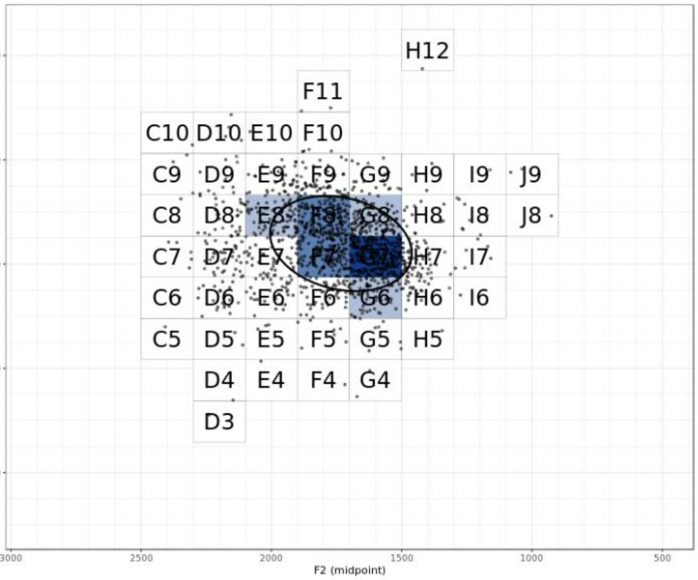

Figure 22

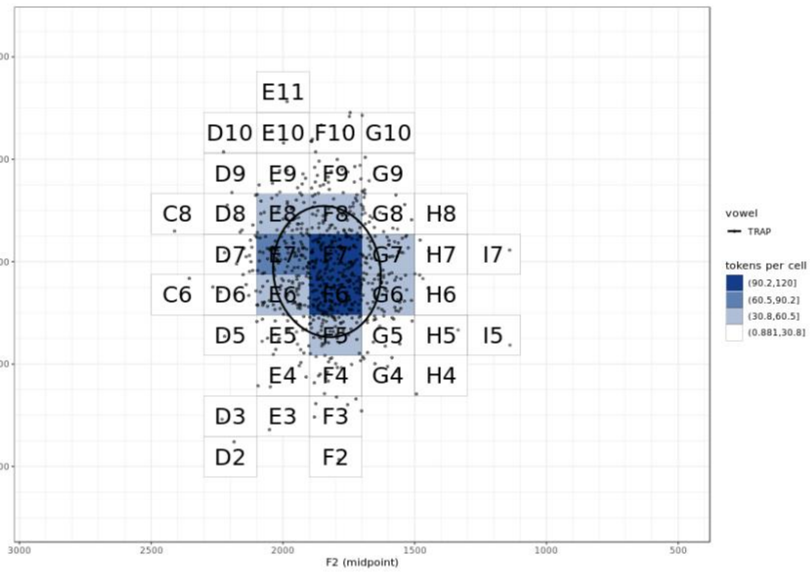

Figure 23

We can see from the difference in density of these PPA plots that the Plains speaker has a higher F2 than the Delta speaker. As I aimed to get a well-rounded view of this vowel, I also chose two female speakers from the Plains and Delta. Both speakers are non-black women aged 13 to 45, with 13+ years of schooling. The Delta speaker is Figure 24 and Plains is Figure 25.

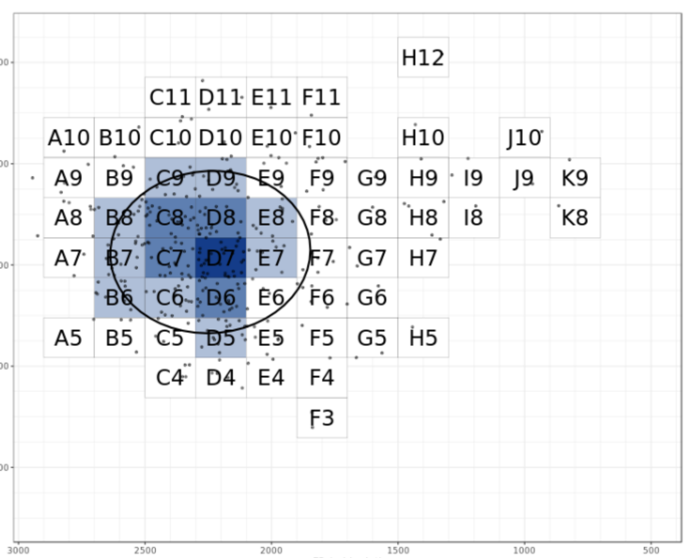

Figure 24

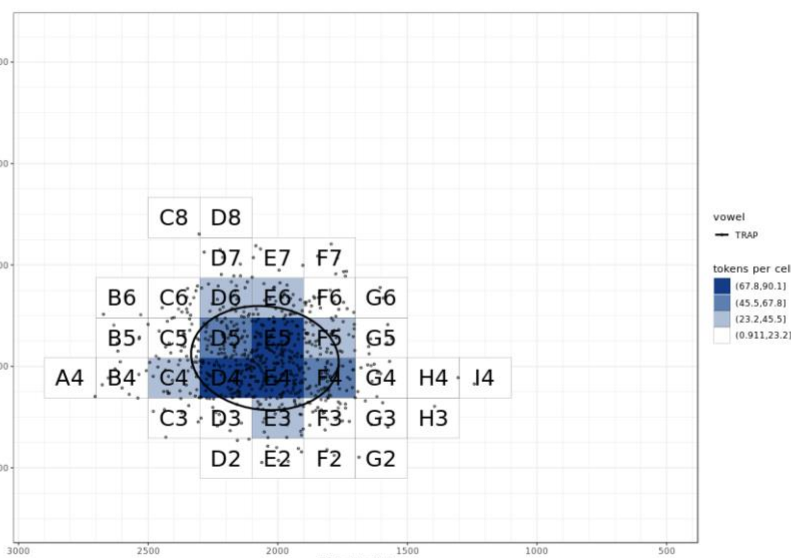

Figure 25

This PPA plot shows an even more drastic heightening of the F2. These visual comparisons, though not exhaustive, give us a visual representation of the shift in the vowel nucleus of demographically similar speakers from different geographical areas. This, along with the other data given in this paper, can lead us to believe that geography in certain areas has a small but distinct impact on the acoustic formants found in the people from that area.

In terms of lexical data, the visual trends I found were not quite as striking as my phonetic results, but aided my hypothesis nonetheless. The most notable example of this I could find was the term for hornet in LAGS. Figure 26 shows the usage "sweat bee", and Figure 27 shows the usage "honey bee".

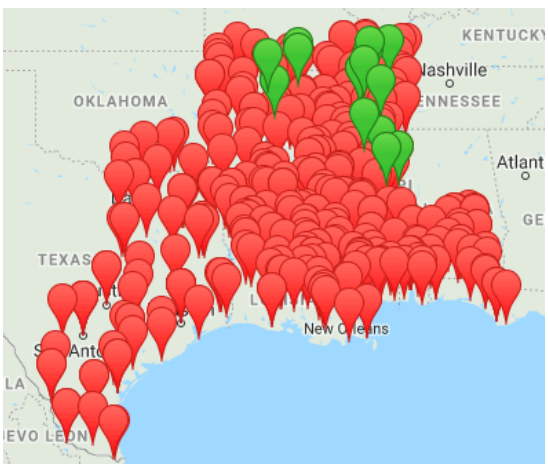

Figure 26

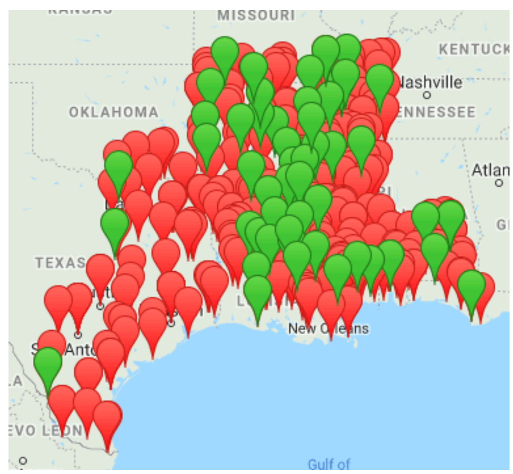

Figure 27

"Sweat bee" seems to be the usage in certain parts of northern Arkansas (green), with the Highland and Plains area using them. It is interesting that the Delta region avoids using this lexical item, though their neighbors in different geographical regions are not averse. "Honey bee" seems to have a more scattered usage, but upon closer inspection and comparison with the LAGS region map, you can see a line of usages going through the Delta region. While it is not a hard line of usage, that is to be expected in dialect studies. Though these examples are quite small, within the larger landscape of my phonological findings they are important to note.

CONCLUSIONS

From my statistical findings it is safe to say that certain geographical areas have a strong correlation on the phonology of the speakers inhabiting it. While some regions, particularly the Plains, are not proven to have any impact on dialect, the Delta and Piney Woods areas yield results that are hard to dispute. Why some geographical regions have a much stronger correlation than others is a topic that should absolutely be further explored in dialect studies.

Even if not all of the data I found turned up the results I was looking for, I was pleased to find my statistical analyses yielded results that were homoscedastic and not overfitted. Nevertheless, I believe this shows an important finding in the dialect landscape of the American South in this period. These findings could assert that physiography affects the tongue height in vowel pronunciations. The visual examples given, such as the maps of acoustic data and the vowel plots, further solidify the strong possibility of geography directly affecting dialect.

FURTHER WORK

While I attempted to chart the vowel trajectories of these speakers, I found that there was no simple way to accomplish this in RStudio other than averaging all of the speaker's data before putting it into the formula meant to chart vector lengths. I will continue to work towards finding a simpler way to do this, or at least a way to visualize the vowel trajectories of the speakers on a plot to show if one region has a tendency to diphthongize or monophthongize over another.

Another goal in my assessment was to look at the lexical differences in these regions as well as the phonological differences. At the completion of this thesis, the DASS lexical corpus was still being loaded onto UGA's server. Therefore, I am unable to run my more in-depth lexical analyses that I had hoped to do alongside my phonological analyses.

WORKS CITED

Beck, J., Randall, A. J., & Frandsen, W. J. (2007). Southern culture: an introduction. Carolina Academic Press.

Dubois, S., & Horvath, B. M. (2003). Verbal Morphology in Cajun Vernacular English: A Comparison with Other Varieties of Southern English. Journal of English Linguistics, 31(1), 34–59.

Dudley, Leah M. "Accent accommodation in American students studying in the UK." 6th Annual Linguistic Conference at the University of Georgia. University of Georgia, Athens. 5 Oct 2019. Poster Presentation.

Kretzschmar, William A., Jr. (2003). Mapping Southern English, 78(2), 130–149.

Kretzschmar, William A., Jr. (1993). Handbook of the linguistic atlas of the Middle and South Atlantic states. University of Chicago Press.

Kretzschmar, William A. Jr. (2011). Linguistic Atlas Project. Linguistic Atlas Project. http://www.lap.uga.edu/.

Kretzschmar, William A., Jr., & Schneider, E. W. (1996). Introduction to quantitative analysis of linguistic survey data : an atlas by the numbers. Sage Publications.

Kretzschmar Jr., W. A., Bounds, P., Hettel, J., Pederson, L., Juuso, I., Opas-Hänninen, L. L., & Seppänen, T. (2013). The Digital Archive of Southern Speech (DASS). Southern Journal of Linguistics, 37(2), 17–38.

Kretzschmar, William A. Jr., Margaret E. L. Renwick, Lisa M. Lipani, Michael L. Olsen, Rachel M. Olsen, Yuanming Shi & Joseph A. Stanley. (2019). Transcriptions of the Digital Archive of Southern Speech. http://www.lap.uga.edu/Projects/DASS2019/.

William Labov, Ingrid Rosenfelder, & Josef Fruehwald. (2013). One Hundred Years of Sound Change in Philadelphia: Linear Incrementation, Reversal, and Reanalysis. Language, 89(1), 30.

Levshina, N. (2015). How to do linguistics with R : data exploration and statistical analysis. John Benjamins Publishing Company.

Pederson, Lee, et al. Linguistic Atlas of the Gulf States Volume 1. University of Georgia Press, 1986.

Pederson, Lee, et al. Linguistic Atlas of the Gulf States Volume 5. University of Georgia Press, 1986.

Pederson, Lee & Kretzschmar, William A. Jr. Linguistic Atlas of the Gulf States (LAGS). http://www.lap.uga.edu/Site/LAGS.html.

Pederson, L. (1974). A manual for dialect research in the Southern States. University of Alabama Press.

Stanley, J. (2019). The Gazetteer of Southern Vowels (GSV). http://lap3.libs.uga.edu/u/jstanley/vowelcharts/.

Thomas, E. R. (2008). Rural Southern White Accents. In E. W. Schneider & B. Kortmann (Eds.), Varieties of English, 2: The Americas and the Caribbean (pp. 87–114). Mouton de Gruyter.

Qin Yan, & Vaseghi, S. (2003). Analysis, modelling and synthesis of formants of British, American and Australian accents. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP ’03)., Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP ’03). 2003 IEEE International Conference on, Acoustics, Speech, and Signal Processing, 1.