Jacob Feldgoise

jfeldgoi

36-207A

April 17, 2018

Religious Fervor
Analyzing the significance of religious practice

1. Introduction

Understanding the significance of constituent demographics is incredibly important to politicians and policymakers alike. These individuals constantly apply statistical tools to surveys in order to effectively understand the political perspectives of target populations. How fervently an individual practices their religion—what will be called "Degree of Religious Practice"—is one key aspect of political science. This study seeks to answer one primary research question and two secondary research questions. The primary research question is to determine whether or not there exists a correlation between the degree of religious practice and household income. The secondary research questions are as follows: (1) Is there a correlation between the degree of religious practice and geographic location within the United States? (2) Is there a correlation between the degree of religious practice and age?

My hypothesis for the primary question is that degree of religious practice will be strongly negatively correlated with household income; as household income increases, degree of religious practice will tend to decrease. My hypothesis for the first secondary question is that degree of religious practice will be moderately correlated with location within the US. My hypothesis for the second secondary question is that degree of religious practice will be moderately positively correlated with age; as the age of the respondent increases, degree of religious practice will tend to increase.

2. Methods

The dataset used by this study, called "religion_survey_results", contains the responses of approximately 1000 respondents to a FiveThirtyEight survey of religious preference and practice. The dataset was obtained from FiveThirtyEight's profile on Kaggle. The unit of analysis for this dataset is the individual respondents to the survey. This dataset contains 48 categorical variables and no quantitative variables. 44 of the variables (questions in the survey) ask the respondent about the frequency of their religious observance as well as personal comfort with regard to various aspects of religion. The survey also questions respondents about how others interact with the religious tendencies of the respondent. The remaining 4 variables are the respondents' age range, gender, household income, and region (location) in the U.S.

In order to answer my research question, an aggregate quantitative variable was created to assess the degree of religious practice for each respondent. The dataset provides the foundations for this variable with a number of variables (questions) asking about the frequency of practice for various religious activities. The survey provided 6 potential answers to these questions ranging from "never" to "more than daily." Each of the relevant categorical variables was recoded to a 0-6 quantitative scale. I set missing values such as "not applicable" to zero. All respondents who chose not to respond to every question were removed from the analyses. Then, the newly-formed quantitative variables were added together in order to generate a single measure of religious practice for each respondent.

The analyses also used variables that describe the age range, household income, and geographical region of the respondents. It did not make sense to convert the "US region" variable to a quantitative variable, so it was left it alone. The specified household incomes are intended to subdivide the respondents within the categories of low income, middle class, and high income. Furthermore, the age ranges were specifically chosen by the surveyor to identify young, middle-aged, and old respondents. Since the divisions are not equal for either variable, neither age nor household income were recoded into quantitative variables. This means that the dataset is not conducive to scatterplots or similar bivariate quantitative analyses.

First, univariate analyses were performed on the degree of religious practice, age, U.S. region, and household income variables. Specifically, this study analyzes the distributions of these variables to determine their significance, using histograms and bar graphs when appropriate. It is

important to recognize that the limited number of categories may hide the true shapes of the the qualitative variable distributions. Then, bivariate analyses were performed to individually compare degree of religious practice to age, U.S. region, and household income. Side-by-side box plots were analyzed and ANOVA were conducted for each bivariate distribution.

3. Results

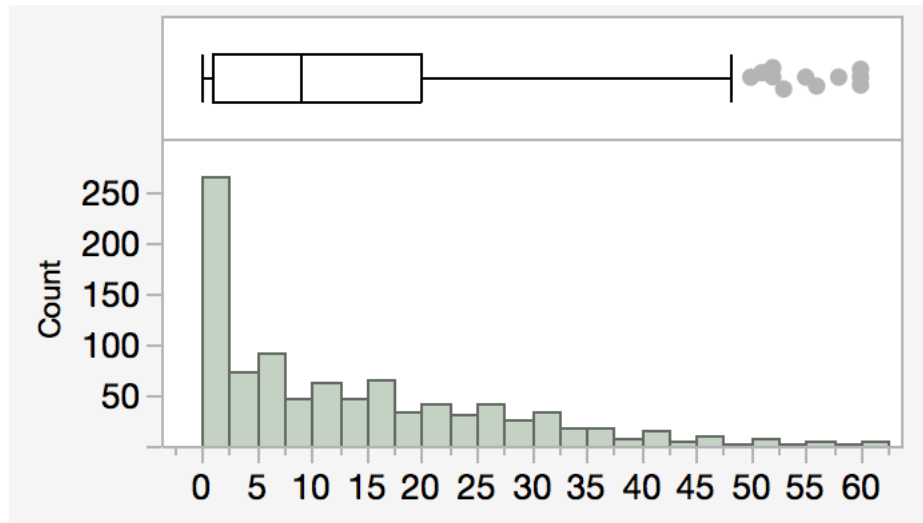*Figure 1: Degree of Religious Practice Histogram*

Figure 1 shows a univariate analysis (histogram and corresponding box plot) of the aggregate "Degree of Religious Practice" variable. It has a five number summery of 0, 1, 8, 20, 60. The distribution also has a standard deviation of 12.5, which suggests that it is highly variable. Since this is an aggregate variable, individual values to not have explicit meaning. Therefore, the center (median) is not be particularly relevant in a univariate distribution, however, the mean will later be crucial for our bivariate analyses. The distribution is strongly skewed to the right, with approximately 10 outliers in its tail. This distribution essentially tells us that most people do not practice religion to a significant degree, but a few people practice their religion with great zeal. A composite table including additional information about the distribution of the "degree of religious practice" variable is located in the appendix (Table A.1).

Figure 2 shows a univariate analysis (bar graph) of the respondents' household incomes. The distribution is likely centered in the $50,000-74,999 bar. This distribution has six categories

ranging from "$0 to $9,999" to "$200,000 and up" on either extreme. The distribution is generally skewed to the right, but occurrences also decrease at the lower extremes. It seems likely that the distribution might show a bell curve if number of bars were increased. An interesting feature of the distribution is the surprisingly high number of respondents who chose the "$200,000 and up" category. This category is likely overpopulated because there are no additional categories to encompass higher degrees of wealth. Again, if the number of categories were increased, this feature would likely not occur.
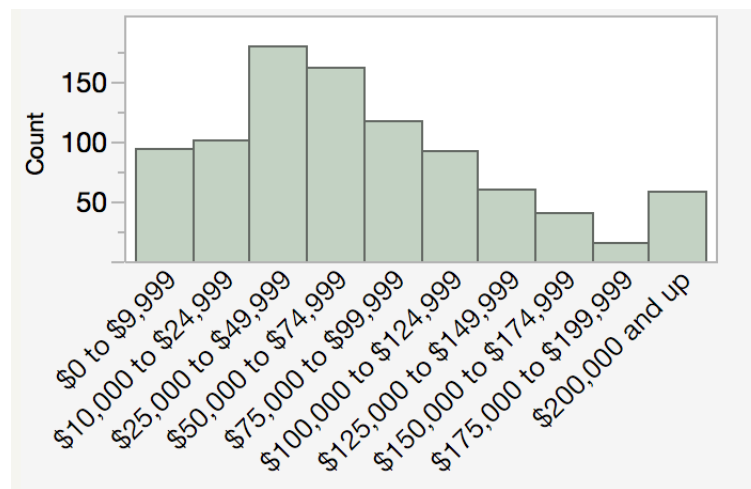

Figure 2: Household Income Bar Graph

Figure 3 shows a univariate analysis (bar graph) of the respondents ages. Since this is a qualitative variable with only a few categories, the distribution analysis is not particularly informative. However, the distribution does suggest that "Age" is roughly uniform across the four age ranges. This distribution indicates that the survey succeeded in reaching roughly equal numbers of young, middle-aged, and (older) retired individuals.
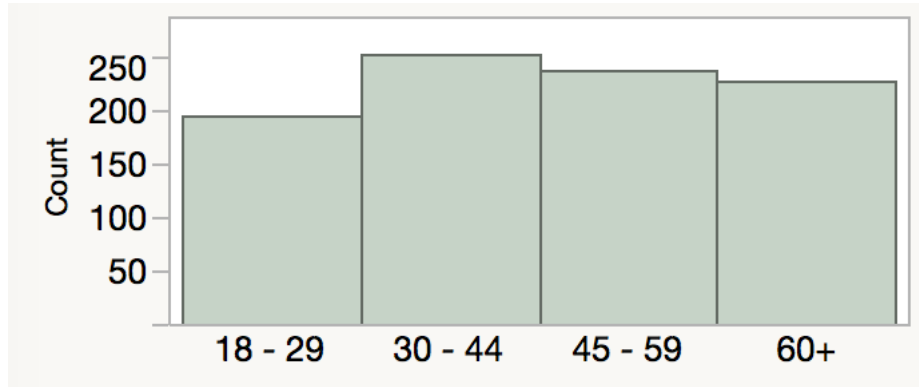
Figure 4 shows a univariate analysis (bar graph) of the respondents geographical location within the United States. The graph shows that most of the respondents live in the "East North Central" and "South Atlantic" regions. "East North Central" likely includes states such as Wisconsin, Michigan, Indiana, Illinois, Ohio, and maybe Minnesota. "South Atlantic" likely includes North Carolina, South Carolina, Georgia, Florida, and maybe Alabama. However, the graph also indicates large numbers of respondents from the mid-Atlantic and Pacific regions. Surprisingly, there were only about 50 respondents from New England, compared to roughly 175 respondents from the South Atlantic; a SRS of the U.S. population should have generated more responses from densely populated New England. All told, the survey seems to have captured a fairly representative sample of the U.S. population by region, even if data was obtained for fewer New Englanders than expected.
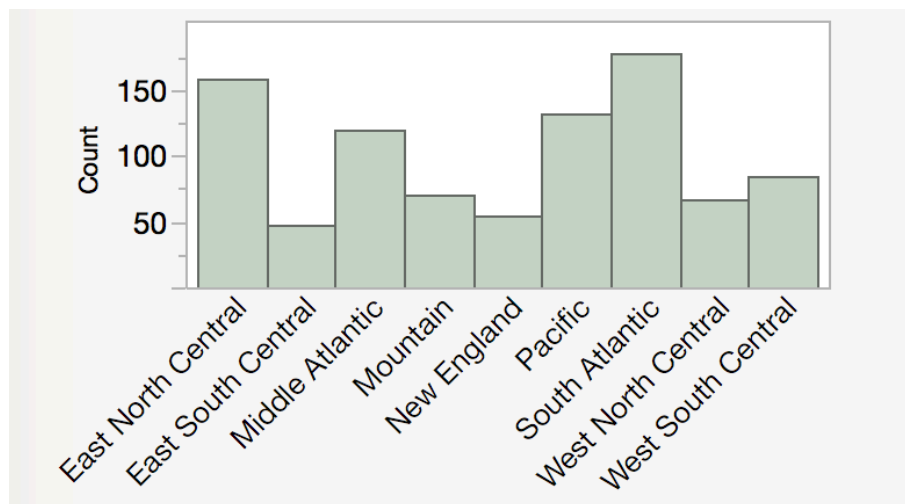
Figure 4: U.S. Region Bar Graph

Figure 5 shows a bivariate analysis of degree of religious freedom over household income. Box plots are used to show the changes to the median and spread of the distribution across each category of household income. The minimum values of each box plot are exactly the same: zero. However, the upper whiskers of the box plots decrease as household income increases. Furthermore, as household income increases, the number of outliers generally increases as well. This indicates that the centers of the distributions (mean and median) tend to shift downward as household income increases. Both IQRs and standard deviations (See Table A.2) for the distributions decrease as household income increases, but the trend is not particularly pronounced.

The null hypothesis is that the average value of degree of religious practice is the same for all categories of household income. The ANOVA test yielded a p-value less than 0.0001, so the null hypothesis is rejected at the 0.05 confidence level. Visually observing the medians and means of Figure 5 clearly shows that degree of religious practice varies depending on household income. Specifically, the medians and means seem to generally decrease as the household income of respondents increases. This visual analysis supports the initial hypothesis (stated in the introduction) that degree of religious practice and household income are negatively related.

Figure 5: Degree of Religious Practice by Household Income: Side-by-side Box Plots
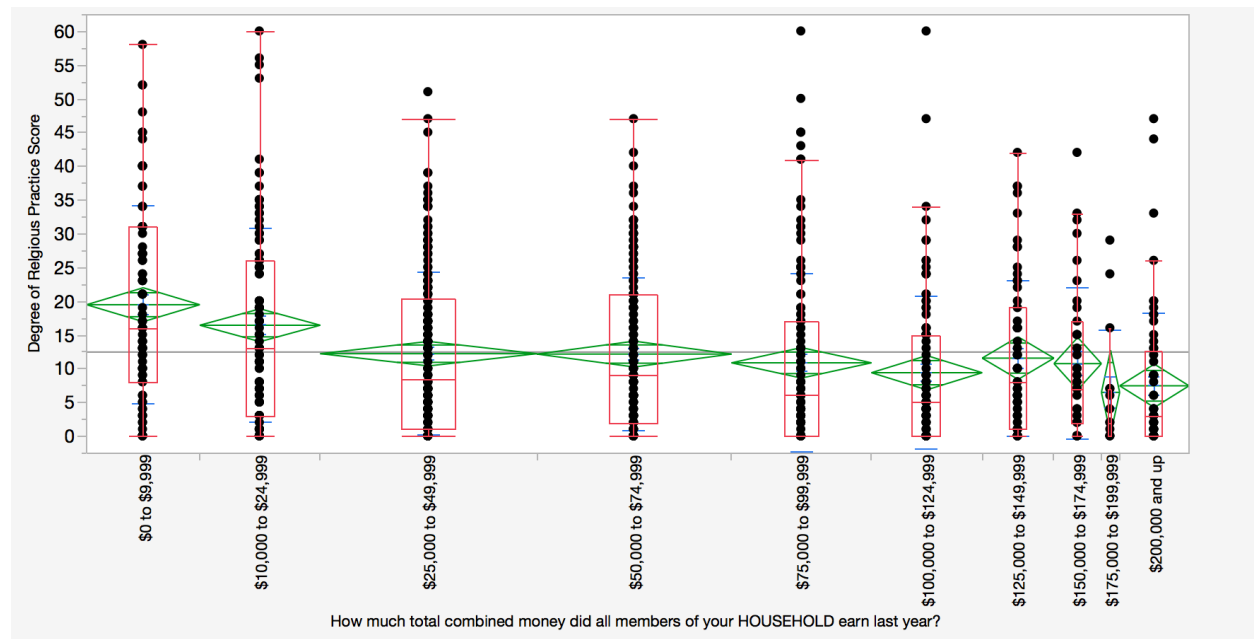
Figure 6 shows a bivariate analysis of degree of religious freedom over age. Box plots are used to show the changes to the median and spread of the distribution across each category of age. The minimum values of each box plot are exactly the same: zero. However, the upper whiskers of the box plots decreases as age increases. The maximum values still remain roughly the same for each age group because of outliers to the upper extreme of degree of religious practice. This indicates that the centers of the distributions (mean and median) tend to shift downward as age increases. The standard deviations of the box plots seem to follow no trend (See Table A.3), however, the IQRs do slightly but consistently decrease as age increases (See Table A. 4).

The null hypothesis is that the average value of degree of religious practice is the same for all categories of age. The ANOVA test yielded a p-value of 0.0182, so the null hypothesis is rejected at the 0.05 confidence level. Visually observing the medians and means of Figure 6 clearly shows that degree of religious practice varies depending on age, despite that there exists no observable linear trend. Specifically, Americans of ages 45-59 will, on average, practice their religion to a greater degree than Americans of any other age group. This visual analysis rejects the initial hypothesis that degree of religious practice will be moderately positively correlated with age.

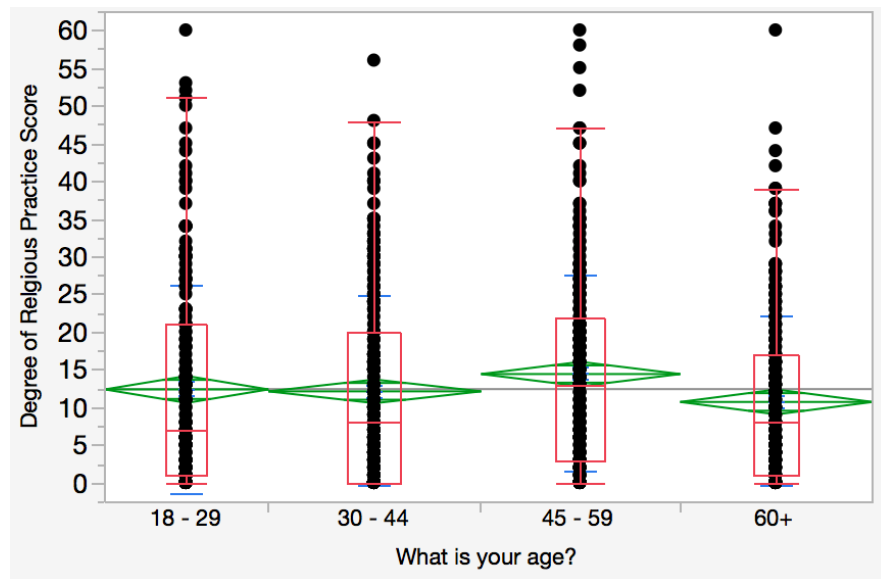*Figure 6: Degree of Religious Practice by Age: Side-by-side Box Plots*

Figure 7 shows a bivariate analysis of degree of religious freedom over age. Box plots are used to show the changes to the median and spread of the distribution across each category of age. Again, the minimum values of each box plot are exactly the same: zero. The upper whiskers of the box plots do vary significantly by region. Specifically, the "West South Central" region has the highest upper whisker. Five of the nine regions have outliers that reside above the rest of their respective distributions. The standard deviations and IQRs of the box plots seem to follow no trends (See Tables A.5 and A.6).

The null hypothesis is that the average value of degree of religious practice is the same for all U.S. geographical regions. The ANOVA test yielded a p-value less than 0.0001, so the null hypothesis is rejected at the 0.05 confidence level. Visually observing the medians and means of Figure 7 clearly shows that degree of religious practice varies depending on geographical region, despite that there exists no observable linear trend. Specifically, Americans who live in the "East South Central" or "West South Central" region will, on average, practice their religion to a greater degree than Americans in other parts of the country. Notably, individuals who take religious practice to the extreme (represented by outliers) are not located in "East South Central" or "West South Central." This signifies that the means of those two regions would be even more significant if we chose to exclude outliers. This visual analysis supports the initial hypothesis that a correlation exists between the degree of religious practice and geographic location within the United States.

Figure 7: Degree of Religious Practice by U.S. Region: Side-by-side Box Plots
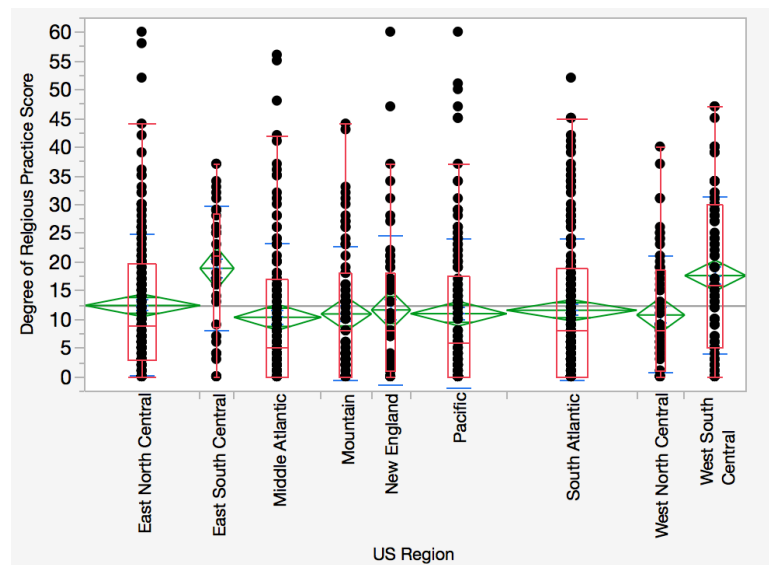
4. Discussion

The bivariate analyses performed above showed that degree of religious practice does vary depending on household income, age, and U.S. geographical region. First, to answer the primary research question, it seems highly probable that degree of religious practice and household income are negatively related. This finding shows that political scientists and politicians can use the average wealth of a district to roughly approximate degree of religious practice relative to districts of differing wealth; low income communities will tend to practice religion to a greater degree than high income communities.

The answer to the first secondary research question is that a relationship does seem to exist between the degree of religious practice and geographic location within the United States. Specifically, Americans who live in the "East South Central" or "West South Central" region will, on average, practice their religion to a greater degree than Americans in other parts of the country. Surprisingly, this analysis did not include "South Atlantic" in the group of regions where individuals practice their religion (over average) to a greater degree than the rest of the country. The answer to the second secondary research question is that degree of religious practice does vary depending on age. Specifically, Americans of ages 45-59 will, on average, practice their religion to a greater degree than Americans of any other age group.

Based on the findings of this study, political campaigns that seek to convey religious messages should specifically target low income communities with high populations of 45-59 year-olds in the central southern region of the U.S. to achieve maximum impact. However, in order to fully confirm the findings of this study, further surveys should seek to obtain quantitative age and household income data. Although this may be difficult because of respondents' reluctance to state such private information, the benefits of t-tests would greatly supplement this work.

Appendix A

Table A.1: Quantiles and Summary Statistics for the "Religious Practice" Variable

Quantiles			Summary Statistics	
100.0%	maximum	60	Mean	12.50883
99.5%		56.93	Std Dev	12.708914
97.5%		44	Std Err Mean	0.4222254
90.0%		31	Upper 95% Mean	13.337485
75.0%	quartile	20	Lower 95% Mean	11.680175
50.0%	median	9	N	906
25.0%	quartile	1		
10.0%		0		
2.5%		0		
0.5%		0		
0.0%	minimum	0		

Table A.2: Summary Statistics for Degree of Religious Practice by Household Income Box Plots

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
$0 to $9,999	93	19.5269	14.6477	1.5189	16.510	22.544
$10,000 to $24,999	99	16.4848	14.3743	1.4447	13.618	19.352
$25,000 to $49,999	178	12.2697	11.9805	0.8980	10.498	14.042
$50,000 to $74,999	160	12.2125	11.2604	0.8902	10.454	13.971
$75,000 to $99,999	115	10.9130	13.1173	1.2232	8.490	13.336
$100,000 to $124,999	91	9.4505	11.3159	1.1862	7.094	11.807
$125,000 to $149,999	59	11.6102	11.5550	1.5043	8.599	14.621
$150,000 to $174,999	39	10.7949	11.1643	1.7877	7.176	14.414
$175,000 to $199,999	15	6.5333	9.2263	2.3822	1.424	11.643
$200,000 and up	57	7.5088	10.6656	1.4127	4.679	10.339

Table A.3: Summary Statistics for Degree of Religious Practice by Age

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
18 - 29	193	12.4663	13.8433	0.99646	10.501	14.432
30 - 44	251	12.2112	12.5745	0.79370	10.648	13.774
45 - 59	235	14.5064	13.0159	0.84906	12.834	16.179
60+	227	10.8062	11.2469	0.74648	9.335	12.277

Table A.4: Quantiles for Degree of Religious Practice by Age

Level	Minimum	10%	25%	Median	75%	90%	Maximum
18 - 29	0	0	1	7	21	33.2	60
30 - 44	0	0	0	8	20	31	56
45 - 59	0	0	3	13	22	32.4	60
60+	0	0	1	8	17	26.2	60

Table A.5: Summary Statistics for Degree of Religious Practice by U.S. Region

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
East North Central	156	12.4679	12.3708	0.9905	10.511	14.424
East South Central	46	18.9348	10.9044	1.6078	15.697	22.173
Middle Atlantic	118	10.3983	12.9173	1.1891	8.043	12.753
Mountain	68	10.9853	11.5713	1.4032	8.184	13.786
New England	53	11.6792	12.9686	1.7814	8.105	15.254
Pacific	130	11.0308	12.9021	1.1316	8.792	13.270
South Atlantic	176	11.6250	12.3495	0.9309	9.788	13.462
West North Central	65	10.8000	10.1538	1.2594	8.284	13.316
West South Central	83	17.6627	13.6474	1.4980	14.683	20.643

Table A.6: Quantiles for Degree of Religious Practice by U.S. Region

Level	Minimum	10%	25%	Median	75%	90%	Maximum
East North Central	0	0	3	9	19.75	30	60
East South Central	0	2.1	8.5	21	28.25	32.3	37
Middle Atlantic	0	0	0	5	17	30.1	56
Mountain	0	0	0	8	18	28.2	44
New England	0	0	1	8	18	29.8	60
Pacific	0	0	0	6	17.5	28.9	60
South Atlantic	0	0	0	8	19	32.3	52
West North Central	0	0	1	8	18.5	25	40
West South Central	0	0	5	16	30	37	47