The Shortage of Hospitals in Rural Tennessee, U.S.A

1. **Introduction**

1.1 Background

Hospitals are big business in Tennessee (TN). Three of the top ten largest hospital corporations in the U.S. are headquartered in the state. However, hospitals in rural areas of TN are closing forcing citizens to travel sometimes up to an hour or more to get health care that requires more than a doctor's visit.

1.2 Problem

Having to travel long distances for health care has several consequences that affect not only the patient, but also the patient's family.

In an emergency such as a stroke, heart attack, or traumatic accident, time is an important factor. Getting the proper medical care quickly can lessen symptoms and help to ensure a more successful patient recovery.

Having routine medical services conveniently located to you means you may be more likely to get testing and other preventive medical services thus increasing health and quality of life.

Family and family members wishing to visit a friend or loved one in the hospital are more likely to visit the patient when the hospital is closer to their home.

1.3 Interest

County government leaders need to ensure that their citizenry is well protected. County business leaders should be concerned since a lack of a hospital may cause business development to locate elsewhere. Hospital corporations may see these areas as healthcare deserts open to business opportunity.

2. **Data acquisition and cleaning**

2.1 The list of Tennessee hospitals came from downloading a csv file provided by the Tennessee Department of Safety and Homeland Security.

Tennessee county demographic information was scraped from en.wikipedia.org/wiki/List_of_Tennessee_locations_by_per_capita_income and then merged with the hospital list based upon with a left join to the hospital list. Demographic data that represented dollar amounts was imported a text, so the $ was removed, and the column type set to float.

The determination of whether a county is rural or not is based upon the US Department of Agriculture Rural-Urban Continuum Code. A category of 9 represents "Nonmetro - Completely rural or less than 2,500 urban population, not adjacent to a metro area" and an 8 represents "Nonmetro - Completely rural or less than 2,500 urban population, adjacent to a metro area". Population, Per Capita Income, and Median Household Income were used to formulate data as well.

Groups 1, 2, and 3 were combined into a single group to aid in map development.

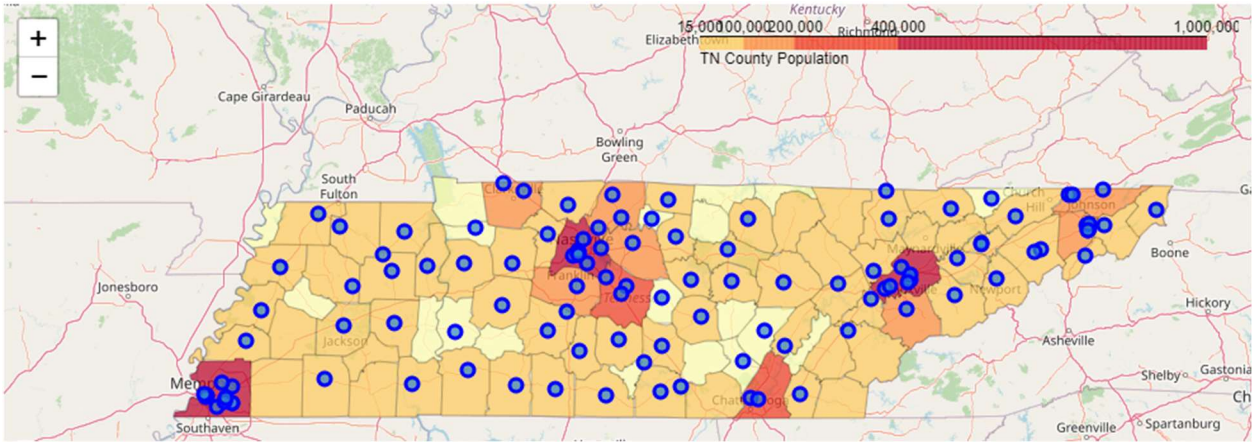
The county RUCC code dataframe was then merged with the county hospital dataframe to provide a dataframe of combined values and a column named Hospital that is either True when the county has a hospital or False when there is not a hospital present.

3. Exploratory Data Analysis

3.1 Find which counties do not have hospitals.

Since data provided by the TN Department of Homeland Security provided us with a list of hospitals with the county name, we needed to join that to a list of TN counties. An outer left join was used for this purpose. Counties without hospitals i.e. NaN in the hospital related columns were identified with a value of False in a column called Hospital. This merged table contained several columns that were not going to be used for this exercise: 'X','Y','HospitalID','Expr1','EPCPhone','ESRI_OID', 'RadioCode1','EPC','ebola_ttx','GlobalID','NDMSFacility','DMATTeamSite','RadioCode1'. The majority of these codes are used for contacting or identifying the hospital which is not necessary for our purposes. Columns 'X', and 'Y' were duplicate coordinates.

In order to help visualize and areas that might have no hospitals, A list of TN Counties and their geographic coordinates were download to build the state and county map with blue dots representing hospitals.



From a visual perspective, population is a decent determinate of whether a county will have a hospital.

3.2 Demographics

This dataframe was further added to by joining it with the county RUCC dataframe based upon county. For analysis purposes, only a hospital value of true or false, the county name and number data was kept in this data frame:

	Hospital	County	Per_Cap_Income	Median_House_Income	Median_Family_Income	Population	RUCC
0	True	Williamson	27925.0	87832.0	100407.0	183182	3.0
1	True	Wilson	27814.0	60678.0	70092.0	113993	3.0
2	True	Davidson	27780.0	45668.0	56084.0	626681	3.0
12	True	Knox	27349.0	46759.0	62272.0	432226	3.0
19	True	Loudon	27046.0	49343.0	59044.0	48556	3.0

TN has a total of 95 counties with statistics as seen below:

	Per_Cap_Income	Median_House_Income	Median_Family_Income	Population	RUCC
count	95.000000	95.000000	95.000000	95.000000	95.000000
mean	19954.094737	38490.431579	47349.884211	66801.105263	5.094737
std	3463.837888	8769.767246	9526.613977	125680.219353	2.068372
min	11813.000000	23125.000000	31495.000000	5077.000000	3.000000
25%	17628.000000	32990.500000	41250.000000	18101.500000	3.000000
50%	19600.000000	36813.000000	45956.000000	31807.000000	6.000000
75%	21514.500000	41780.500000	51194.500000	56443.000000	6.000000
max	27925.000000	87832.000000	100407.000000	927644.000000	9.000000

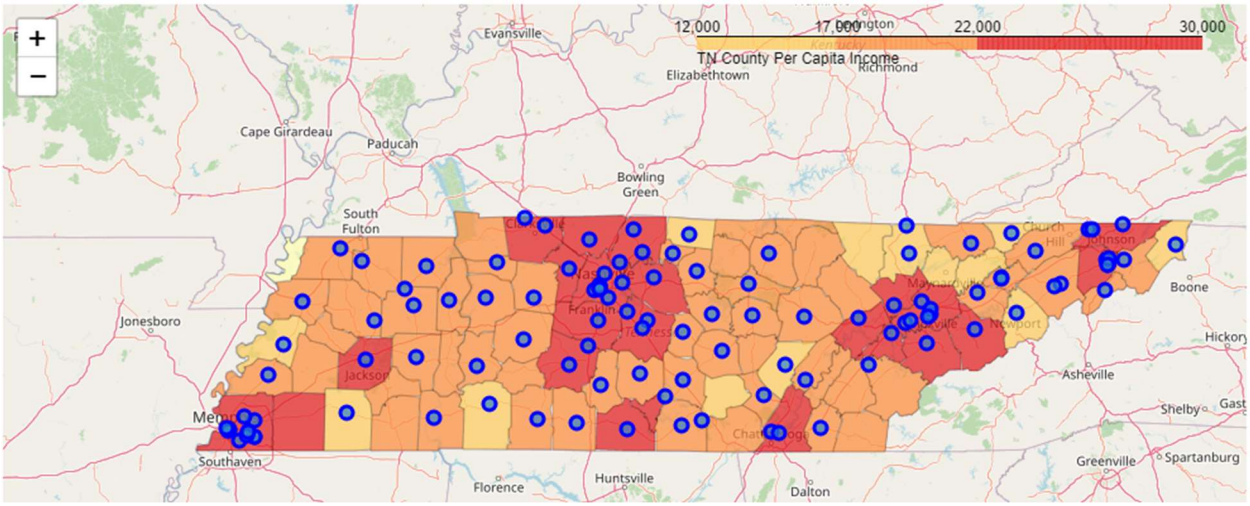
There are 25 TN counties without hospitals for a total of %26.32.

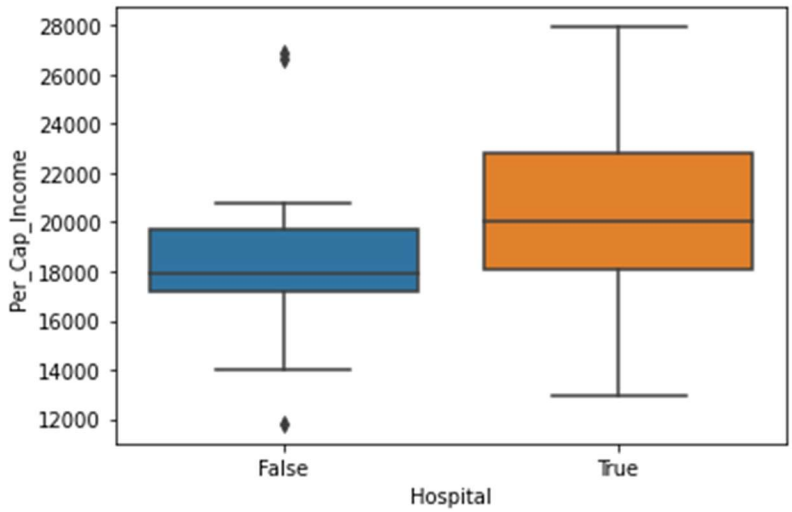
	Per_Cap_Income	Median_House_Income	Median_Family_Income	Population	RUCC
count	25.000000	25.000000	25.000000	25.000000	25.000000
mean	18409.840000	34402.200000	43347.360000	17679.960000	6.240000
std	3217.424911	6515.983809	6780.212576	10580.052286	2.454248
min	11813.000000	24700.000000	32394.000000	5077.000000	3.000000
25%	17160.000000	30193.000000	39767.000000	11753.000000	3.000000
50%	17883.000000	33506.000000	41948.000000	16825.000000	6.000000
75%	19742.000000	36772.000000	46670.000000	21987.000000	9.000000
max	26898.000000	56729.000000	63186.000000	52266.000000	9.000000

There are 70 TN Counties with hospitals for a total of %73.68.

	Per_Cap_Income	Median_House_Income	Median_Family_Income	Population	RUCC
count	70.000000	70.000000	70.000000	70.000000	70.000000
mean	20505.614286	39950.514286	48779.357143	84344.371429	4.685714
std	3401.777597	9043.666955	9988.577433	142453.976415	1.757327
min	12907.000000	23125.000000	31495.000000	6819.000000	3.000000
25%	18088.000000	34357.000000	42732.500000	26332.750000	3.000000
50%	20030.500000	38641.500000	47840.000000	40884.000000	4.000000
75%	22772.750000	43953.750000	52913.750000	71448.500000	6.000000
max	27925.000000	87832.000000	100407.000000	927644.000000	8.000000

3.3 Per Capita Income as an indicator of No hospital vs Hospital





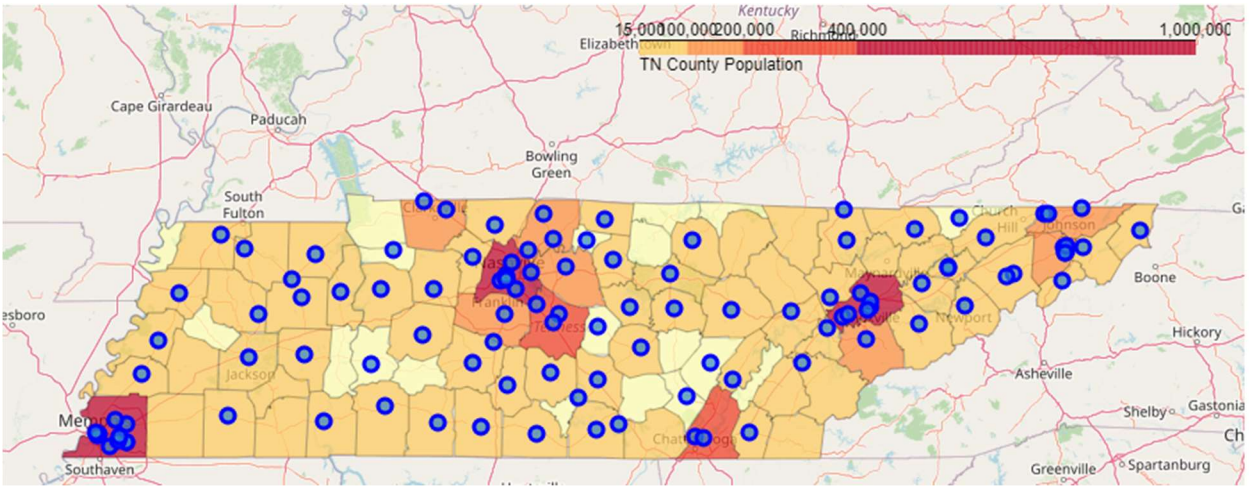
Based upon Per Capita Income (Per_Cap_Income) means of $18,410 for counties without a hospital and $20,506 for counties with a hospital. It would appear that per capita income does play a role in whether a county will have a hospital or not.

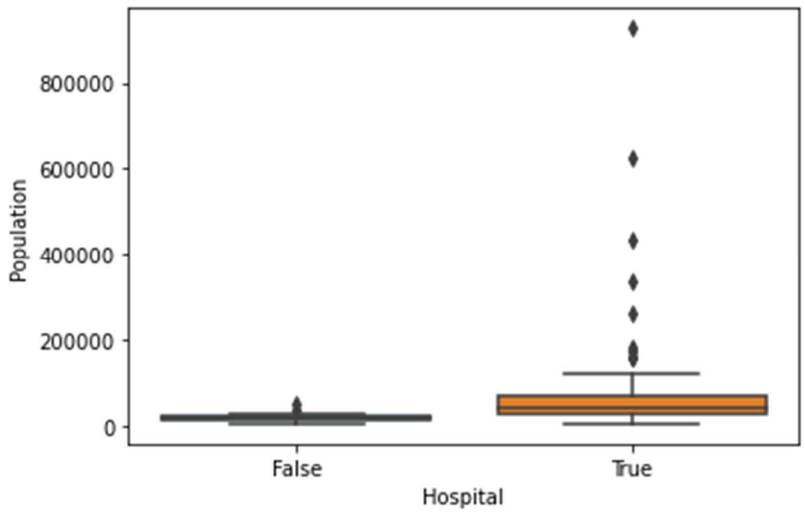
Using the Pearson Correlation Coefficient to see if the data is linear and whether the data provided supports our hypothesis that a lower per capita income does give an indication that a county may not have a hospital.

The Pearson Correlation Coefficient of Per Capita Income is 0.2678429650411144 with a P-value of P = 0.00868572882196481.

A Pearson Coefficient of .2678 would indicate that there isn't a strong linear association between per capita income and whether a county has a hospital or not. However, a P-value of .00857 indicates that per capita income can be a good indicator of whether a hospital will exist in a county or not.

3.4 Population as an indicator of No hospital vs Hospital





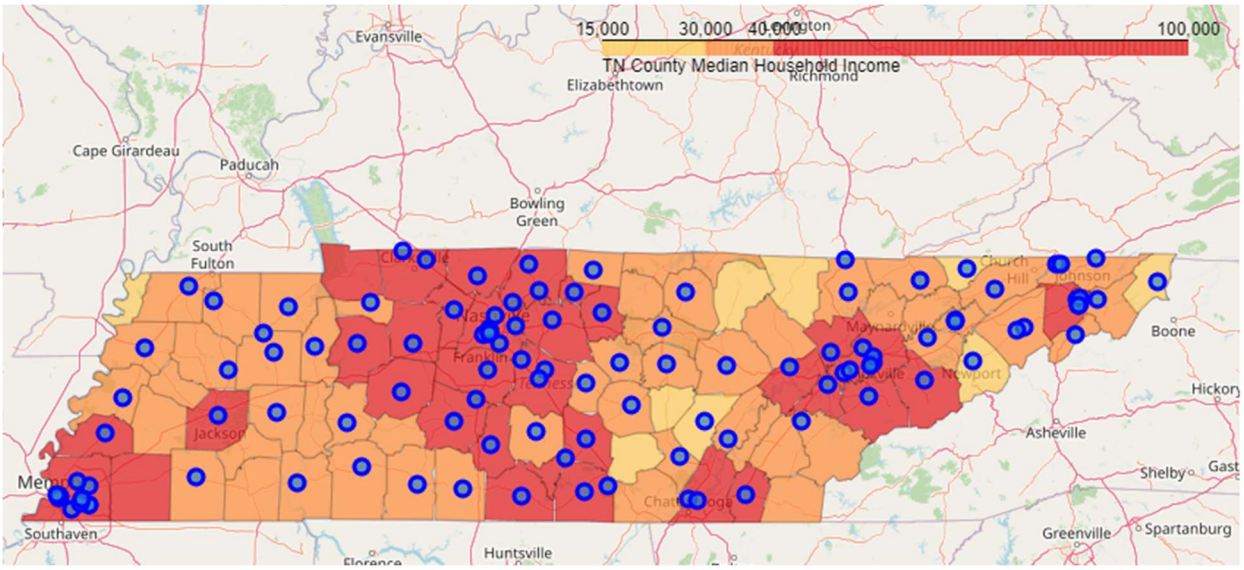
Based upon Population means of 17,680 for counties without a hospital and 84,344 for counties with a hospital. It would appear that population does play a role in whether a county will have a hospital or not.

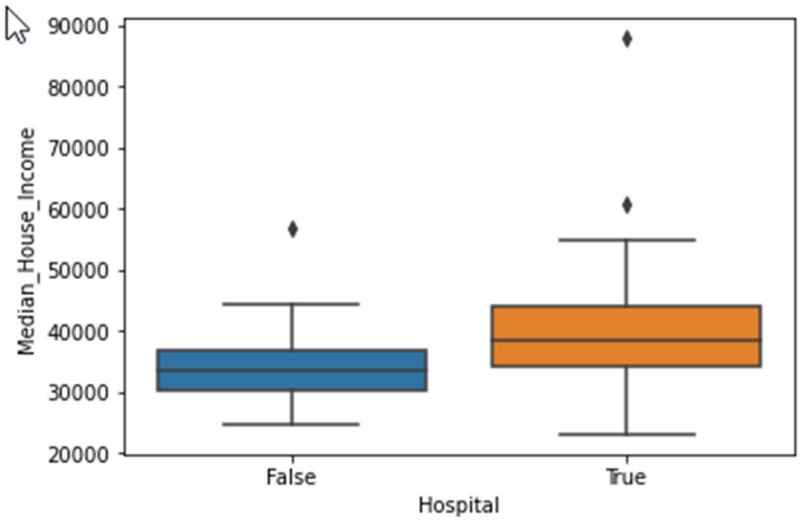
Using the Pearson Correlation Coefficient to see if the data is linear and whether the data provided supports our hypoth esis that a lower per capita income does give an indication that a county may not have a hospital.

The Pearson Correlation Coefficient of Population is 0.23481206832803075 with a P-value of P = 0.021992850478864623

A Pearson Coefficient of .2348 would indicate that there isn't a strong linear association between per capita income and whether a county has a hospital or not. However, a P-value of .022 indicates that population can be a good indicator of whether a hospital will exist in a county or not.

3.5 Median Household Income as an indicator of No hospital vs Hospital





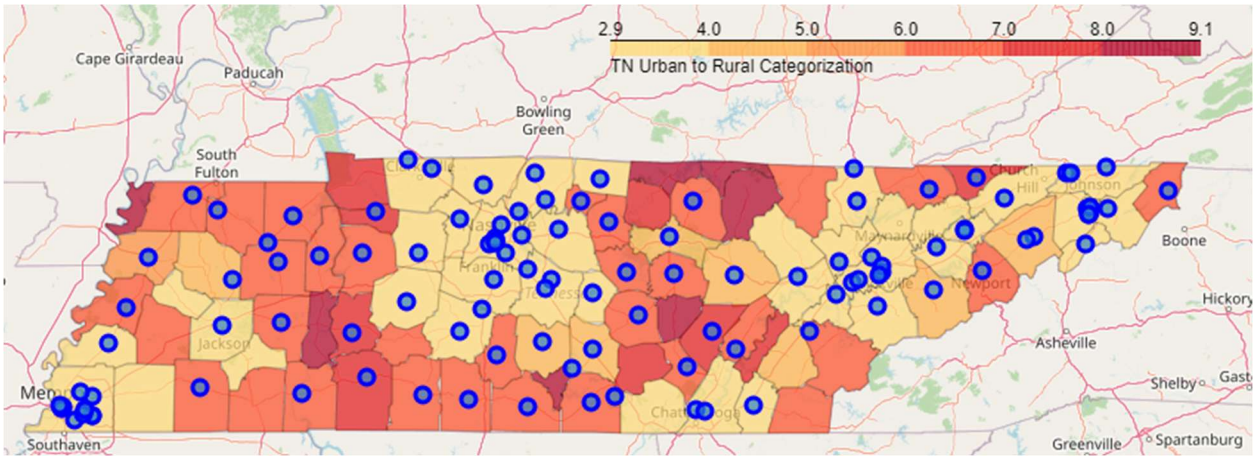
Based upon Median Household Income means of $34,402 for counties without a hospital and $39,951 for counties with a hospital. It would appear that Median Household Income does play a role in whether a county will have a hospital or not.

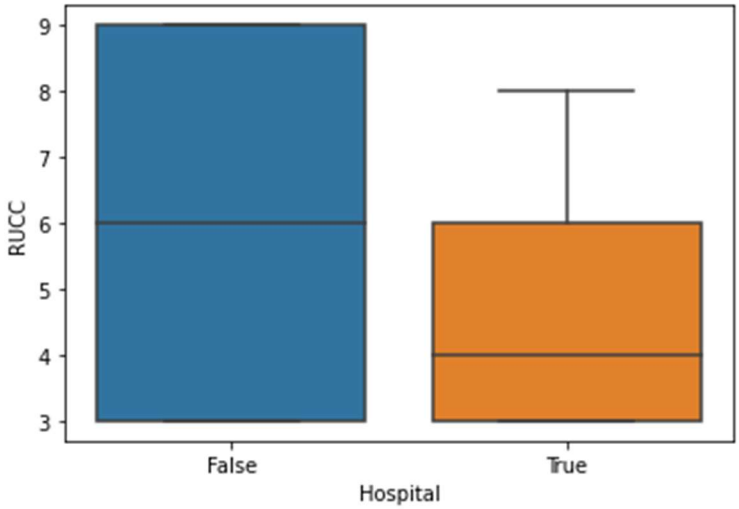
Using the Pearson Correlation Coefficient to see if the data is linear and whether the data provided supports our hypoth esis that a lower per capita income does give an indication that a county may not have a hospital.

The Pearson Correlation Coefficient of Median Household Income is 0.28006980022153116 with a P-value of P = 0.0059 79795046228489

A Pearson Coefficient of .2800 would indicate that there isn't a strong linear association between household income and whether a county has a hospital or not. However, a P-value of .00598 indicates that median household income can be a good indicator of whether a hospital will exist in a county or not.

3.6 Rural-Urban Continuum Code (RUCC) as an indicator of No hospital vs Hospital





Based upon RUCC mean of 6.24 for counties without a hospital and 4.68 for counties with a hospital. It would appear that RUCC score does play a role in whether a county will have a hospital or not.

Using the Pearson Correlation Coefficient to see if the data is linear and whether the data provided supports our hypoth esis that a lower per capita income does give an indication that a county may not have a hospital.

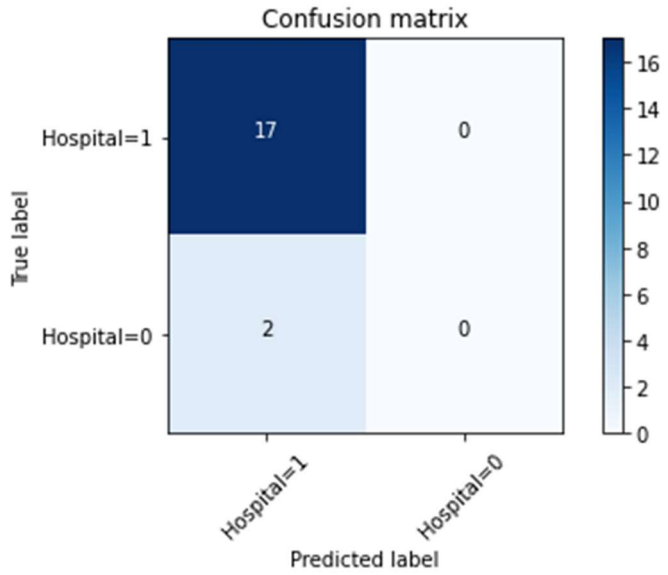
The Pearson Correlation Coefficient of Rural Code is -0.33265607692173693 with a P-value of P = 0.0009886315428535 2

A Pearson Coefficient of .3327 would indicate that there isn't a strong linear association between RUCC and whether a county has a hospital or not. However, a P-value of .00099 indicates that RUCC can be a good indicator of whether a hospital will exist in a county or not and is in fact the best determinate.

3.7 Logistic Regression to verify RUCC as a reliable indicator

3.7.1 Using the RUCC score

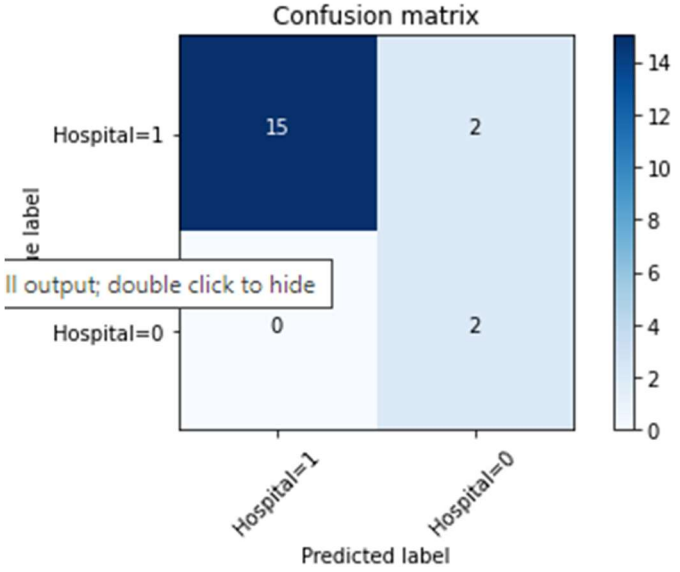
Using RUCC alone, we were able to correctly predict that a county had a hospital 17 our 17 times with an F1 score calculated to be 0.94. Counties without a hospital were incorrectly calculated as having a hospital 2 out of 2 and a calculated F1 score of 0.0:



	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.89	1.00	0.94	17
accuracy			0.89	19
macro avg	0.45	0.50	0.47	19
weighted avg	0.80	0.89	0.85	19

3.7.2 Using combined demographics provides a better indicator

By adding Per Capita Income, Median Household income, and Population to our logistic regression model, we were better able to determine whether a county had a hospital or not.



	precision	recall	f1-score	support
0	0.50	1.00	0.67	2
1	1.00	0.88	0.94	17
accuracy			0.89	19
macro avg	0.75	0.94	0.80	19
weighted avg	0.95	0.89	0.91	19

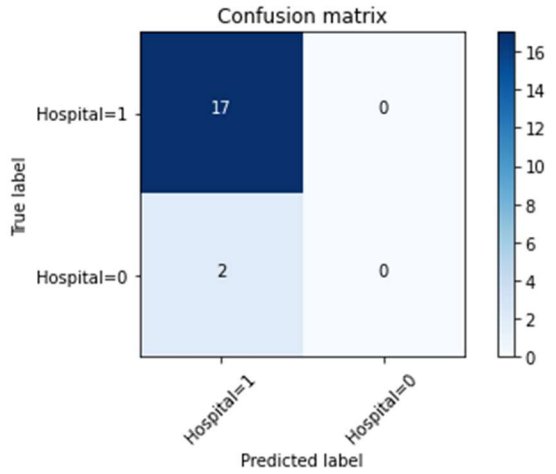
We were able to correctly predict that a county had a hospital 15 our 17 times with an F1 score calculated to be 0.94. Counties without a hospital were incorrectly calculated as having a hospital 2 out of 2 and a calculated F1 score of 0.0.

3.8 Support Vector Machine

3.8.1 Using the RUCC score

	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.89	1.00	0.94	17
accuracy			0.89	19
macro avg	0.45	0.50	0.47	19
weighted avg	0.80	0.89	0.85	19

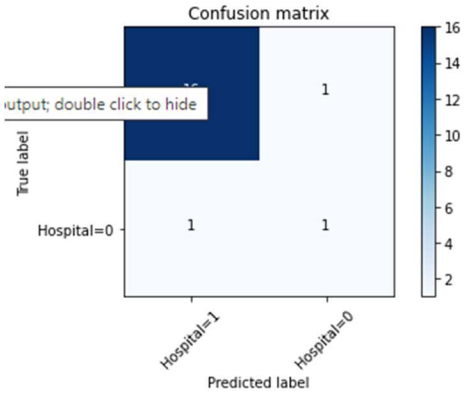
```
Confusion matrix, without normalization
[[17  0]
 [ 2  0]]
```



3.8.2 Using combined demographics provides a better indicator

	precision	recall	f1-score	support
0	0.50	0.50	0.50	2
1	0.94	0.94	0.94	17
accuracy			0.89	19
macro avg	0.72	0.72	0.72	19
weighted avg	0.89	0.89	0.89	19

```
Confusion matrix, without normalization
[[16  1]
 [ 1  1]]
```



3.8.3 Logistic Regression gives a stronger prediction indicator

Weighted Avg	Precision	Recall	F1-score	Support
RUCC Logistic Reg.	0.80	0.89	0.85	19
RUCC SVM	0.80	0.89	0.85	19
All Logistic Reg.	0.95	0.89	0.91	19
All SVM	0.89	0.89	0.89	19

4. Results

If you live in a Tennessee county that is classified as rural and not adjacent to a metropolitan area (RUCC=9), you do not have a hospital. Although a RUCC score of 9 is currently an indicator of no hospital, lower scores alone become problematic in predicting the presence of a hospital. Using the RUCC score along with population and income data allows us to better predict the absence of a hospital.

5. Discussion

Sparse populations and the lack of available capital together seem to be the best indicator of not having a hospital. Given a small hospital costs $52,200,000 to build, the economic reality is that it takes a certain amount of people with sufficient income to make a hospital economically viable.

6. Conclusion

There are 3006 counties in the U.S. so this study only took into account 3% of the counties available. A larger sample size might allow us to have a better understanding of what causes hospital deserts, and allow community leaders to act decisively on behalf of their citizens.