Study of the Texas County Health Rankings - 2020

Christine Orosco – Final Project

DATASET

- Texas County Health Rankings - 2020
- Produced by the University of Wisconsin Population Health Institute [1]
- Purpose is to help counties understand the influences affecting their population's health.
- Communities use the rankings to garner support for local health initiatives.

[1] University of Wisconsin Population Health Institute. (2020, January 1). 2020 Texas Report. Retrieved May 28, 2020, from https://www.countyhealthrankings.org/reports/state-reports/2020-texas-report

DATASET CONTENTS

- Contains 16 categories of health, economic, social, physical environment, demographic measures.
- Categories grouped into Ranked and Unranked measurements.
- Ranked Measures used to compute the county rankings
- Rankings compare against US measurements and State minimum and maximum measurements.

STUDY QUESTIONS

- 254 Counties in Texas with majority of population in urban counties.
- Compare factors between rural and urban counties
  - *Is there a significant statistical difference between rural and urban single parent households?*
  - *Are high school graduation rates along with reading and writing scores better in urban schools?*
  - *Do country people have less health concerns and live longer than city dwellers? – green acres is the place for me, farm living is where I wanna be…*

STUDY DATASET

- Measures from the unranked group, per county
  - *Single Parent Households percentage*
  - *Total county population counts*
  - *High School Graduation Rates with 3rd Average Math and Reading scores*
  - *Physician and Uninsured health counts*
  - *Quality of life factor counts:*
    - *Life Expectancy, Food Insecurity, Insufficient Sleep, Housing Cost Burden, Poor Health*

VARIABLE DEFINITIONS

- Single Parent Households percentage -- percentage per county.
- Total county population counts -- numeric value
- High School Graduation Rates. – percentage rate of high school graduations per county
- 3rd Average Math and Reading scores. -- numeric value representing the average 3rd grade reading and math levels. Using number 3 as the base. Level above 3 shows 3.1, 3.2. Level below shows 2.9, 2.7 etc.
- Physician -- number of physicians per county
- Uninsured health counts -- numeric count of people (adults and children) without insurance per county
- Quality of life survey counts. -- numeric counts per factor of positive responses per county.

ASSUMPTIONS AND VARIABLES PER QUESTION

- Rural = counties with population <= 50,0000
- Urban = counties with populations > 50,000
- Did not count rural populations within Urban counties. Could not determine the number of singles within these populations.
- Variables per question:
  - *Question1 : percent_single, population and county variables*
  - *Question2: grad_rate, read, maths*
  - *Question 3: Life_Expectancy, food_Insecurity, poor_sleep, Cost_burden, poor_health, Physicians, Uninsured*

QUESTION 1 – RESULTS

- CDF shows the percentage of counties and the associated percentage of single parent houesholds.
- No outliers detected
- Descriptive Statistics show no significant statistical difference
- Dsitribution of single parent households similar in both samples
- Cohen's-d Single Parent Households Rural vs Urban = 0.029857843619705685.
- Rural Mode = 32 Urban Mode = 37

QUESTION 1 – DESCRIPTIVE STATISTICS

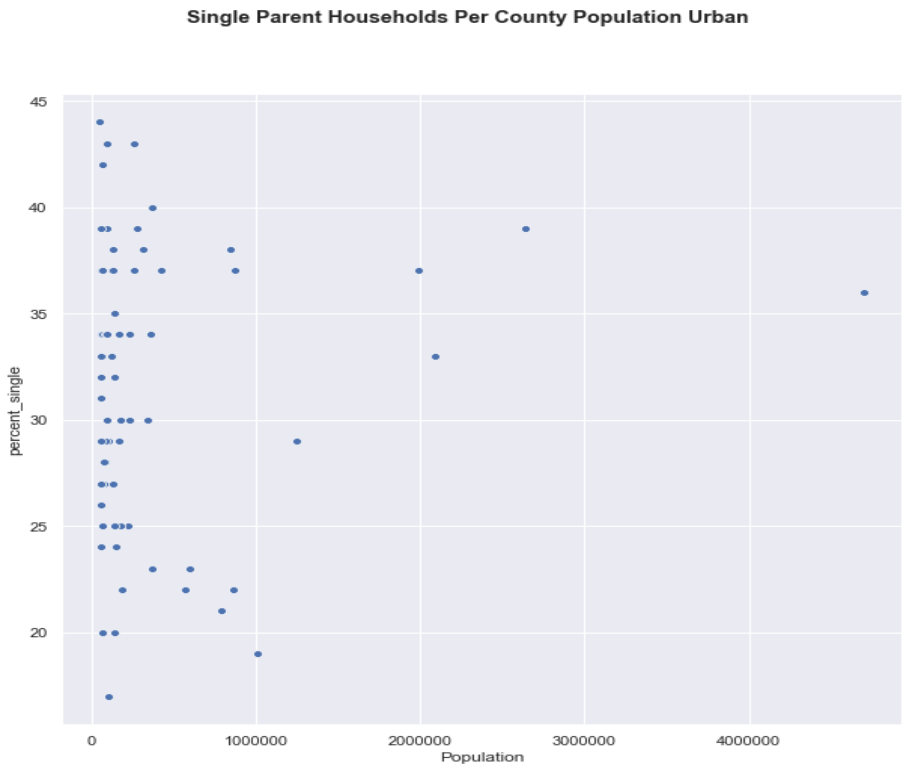
```
rural.describe()
```

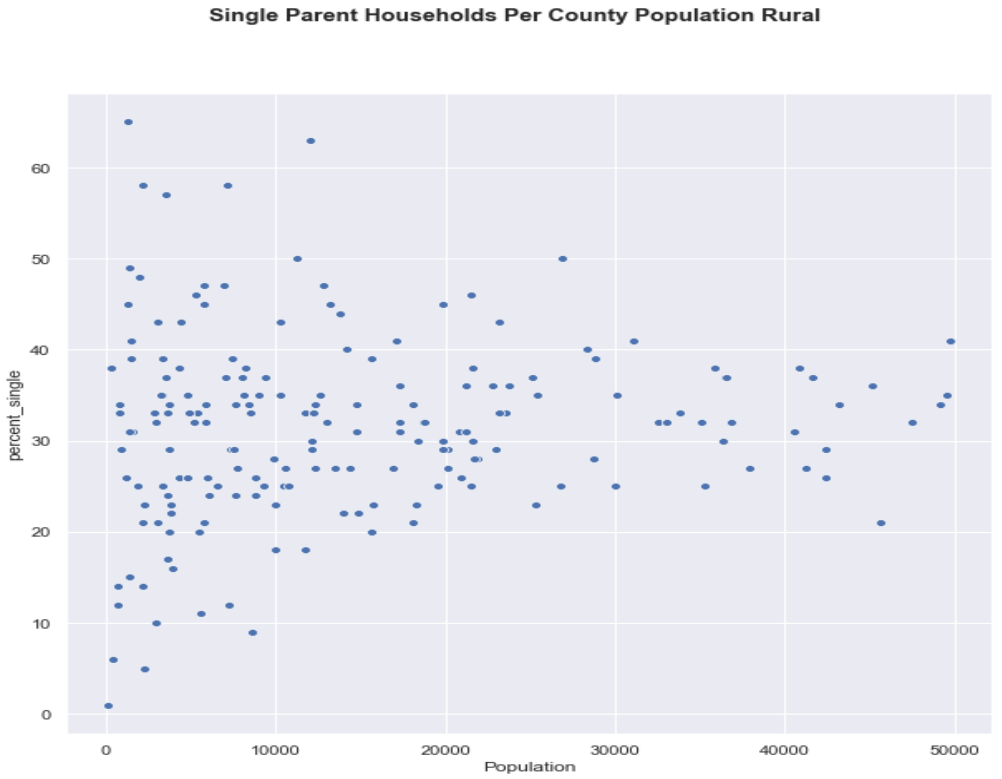
	nbr_sing_household	nbr_household	percent_single	Rural	Population
count	186.000000	186.000000	186.000000	186.000000	186.000000
mean	1122.634409	3468.086022	31.521505	7986.478495	14770.026882
std	1049.116656	3022.640709	10.032492	7027.797094	12605.944718
min	0.000000	61.000000	1.000000	82.000000	152.000000
25%	321.000000	1044.000000	25.250000	2954.000000	4470.000000
50%	768.000000	2715.000000	32.000000	4983.500000	10989.000000
75%	1685.000000	5078.250000	37.000000	11931.250000	21421.000000
max	4875.000000	14001.000000	65.000000	31172.000000	49728.000000

```
urban.describe()
```

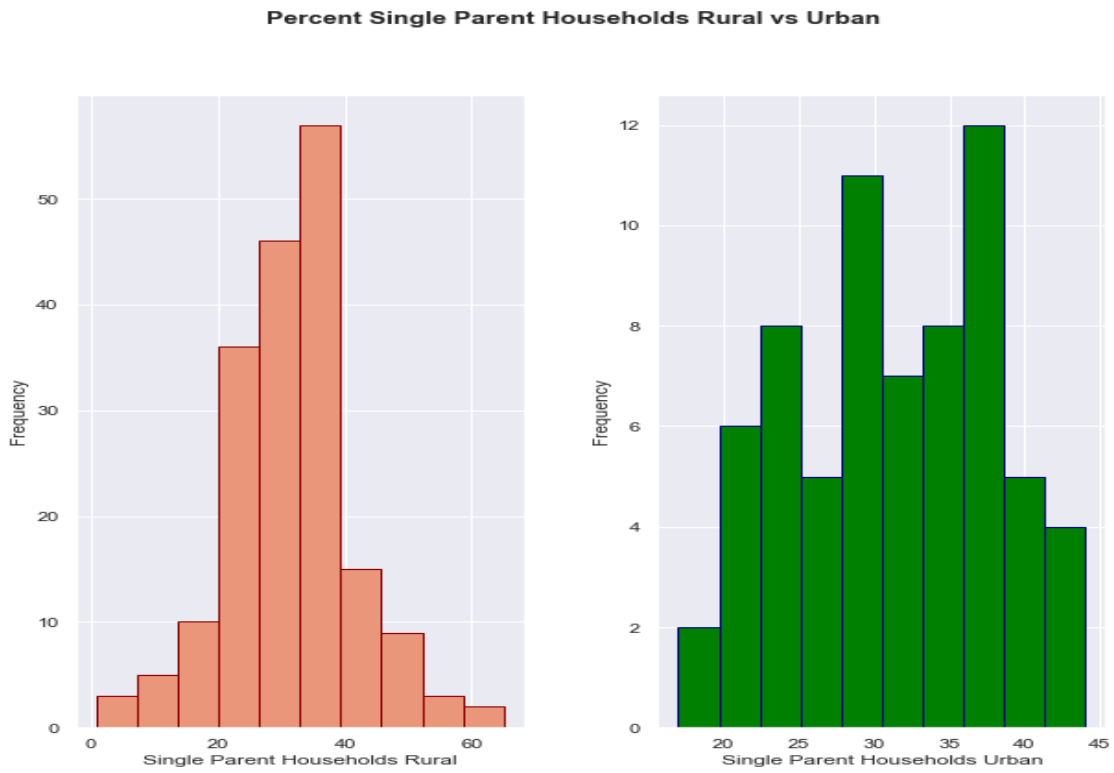
	nbr_sing_household	nbr_household	percent_single	Rural	Population
count	68.000000	6.800000e+01	68.000000	68.000000	6.800000e+01
mean	31809.117647	9.693757e+04	31.264706	34735.838235	3.816856e+05
std	68062.434466	1.882317e+05	6.571348	18480.051069	7.236031e+05
min	2392.000000	9.762000e+03	17.000000	5022.000000	5.003100e+04
25%	5136.750000	1.750350e+04	26.750000	19776.500000	6.795200e+04
50%	9656.500000	3.161000e+04	32.000000	32598.000000	1.330275e+05
75%	25979.750000	8.180650e+04	37.000000	47476.750000	3.150315e+05
max	445154.000000	1.231476e+06	44.000000	103571.000000	4.698619e+06

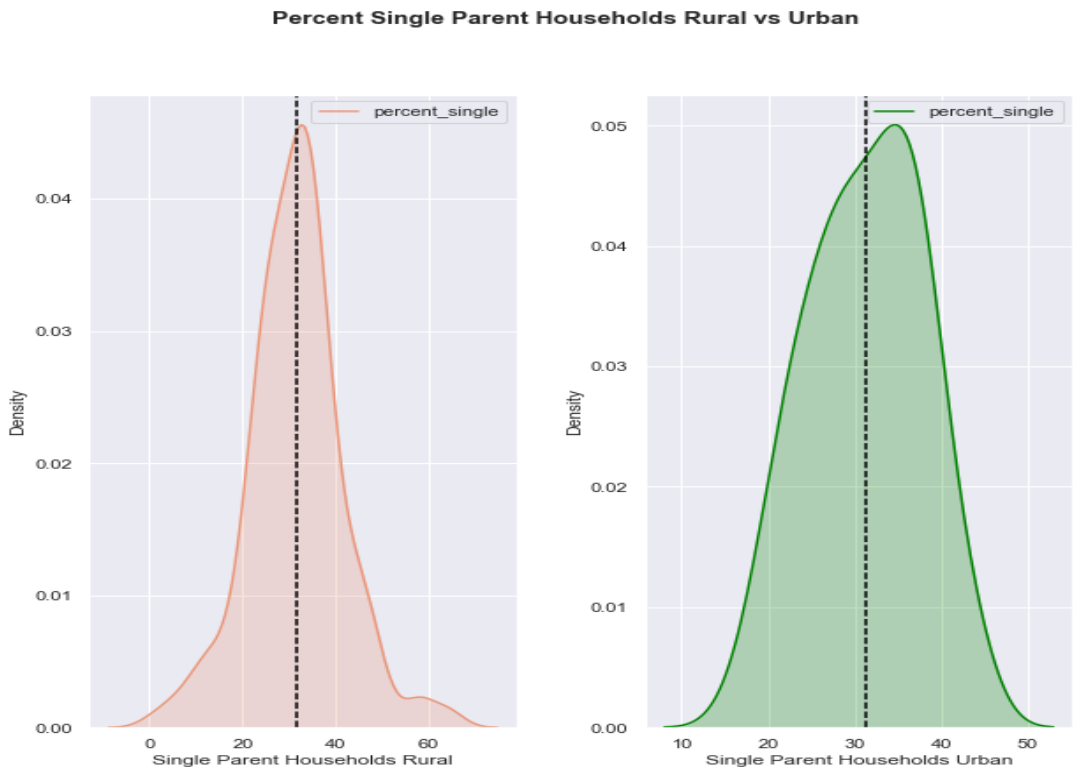
QUESTION 1: SINGLE PARENT HOUSEHOLDS
SCATTER PLOTS



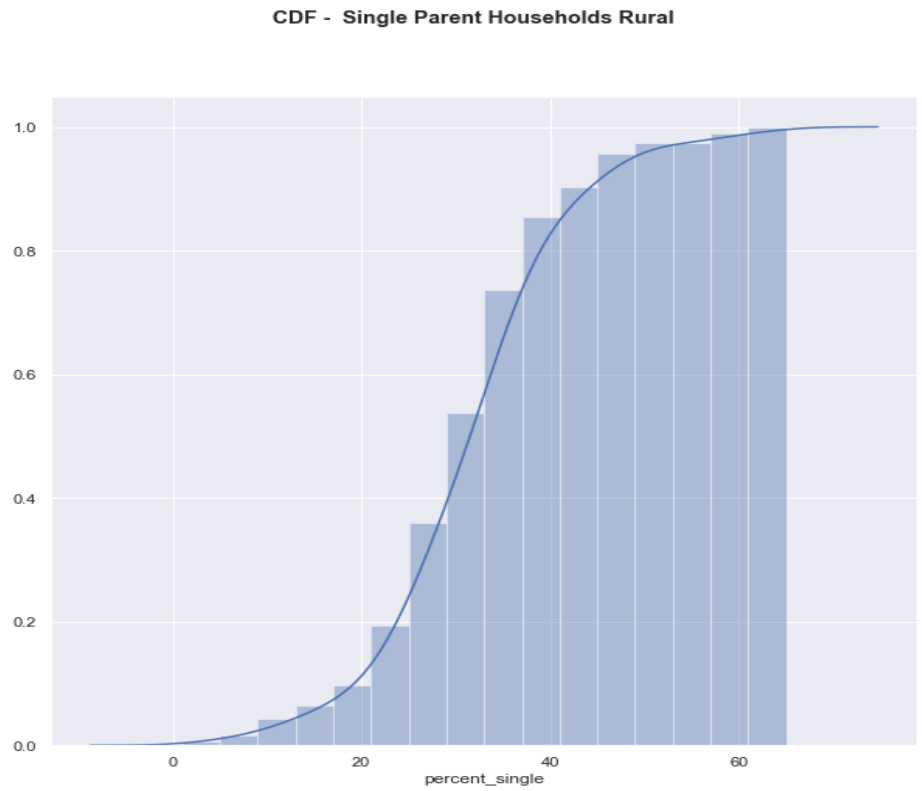


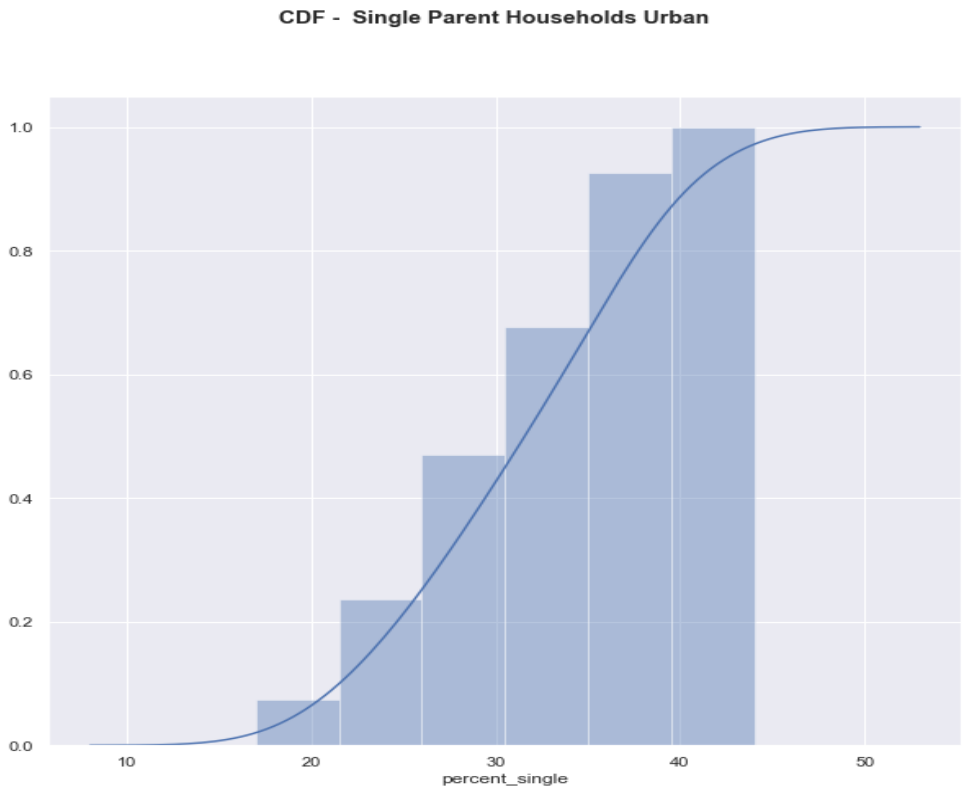
QUESTION 1: SINGLE PARENT HOUSEHOLDS
HISTOGRAM AND KDE





QUESTION 1: SINGLE PARENT HOUSEHOLDS
CDF PLOTS





QUESTION 2 – RESULTS

- No significant difference between Rural and Urban
- Urban has slightly higher 3rd reading level
- Each have a 98 graduation rate
- One reason for high graduation rate is standardized testing and push to pass the test.
- According to the results Rural schools are just as good as Urban schools when it comes to high school graduation, Math and Reading levels.
- Nice to have: How many graduates went on to college.
- Should have: Used the “Some college” variable
- No outliers detected

QUESTION 2 – DESCRIPTIVE STATISTICS

- Mode Rural vs Urban Reading Levels: 98/98
- Mode Rural vs Urban Reading Levels: 2.7/2.9
- Mode Rural vs Urban Math: 3.0/3.0
- Cohens d for High School Graduation Rates Rural vs Urban: 0.4826872508130733
- Cohens d for 3rd Grade Reading Levels Rural vs Urban: -0.1954009724967160 4
- Cohens d for 3rd Grade Math Levels Rural vs Urban: -0.2524115783211429

QUESTION 2 – DESCRIPTIVE STATISTICS

```
hs_df_rural.describe()
```

	Grad_Rate	Math	Read
count	186.000000	186.000000	186.000000
mean	94.392473	2.941398	2.785484
std	5.939492	0.279684	0.237678
min	62.000000	2.100000	2.000000
25%	93.000000	2.700000	2.700000
50%	96.000000	2.900000	2.800000
75%	98.000000	3.100000	2.900000
max	100.000000	4.000000	3.400000

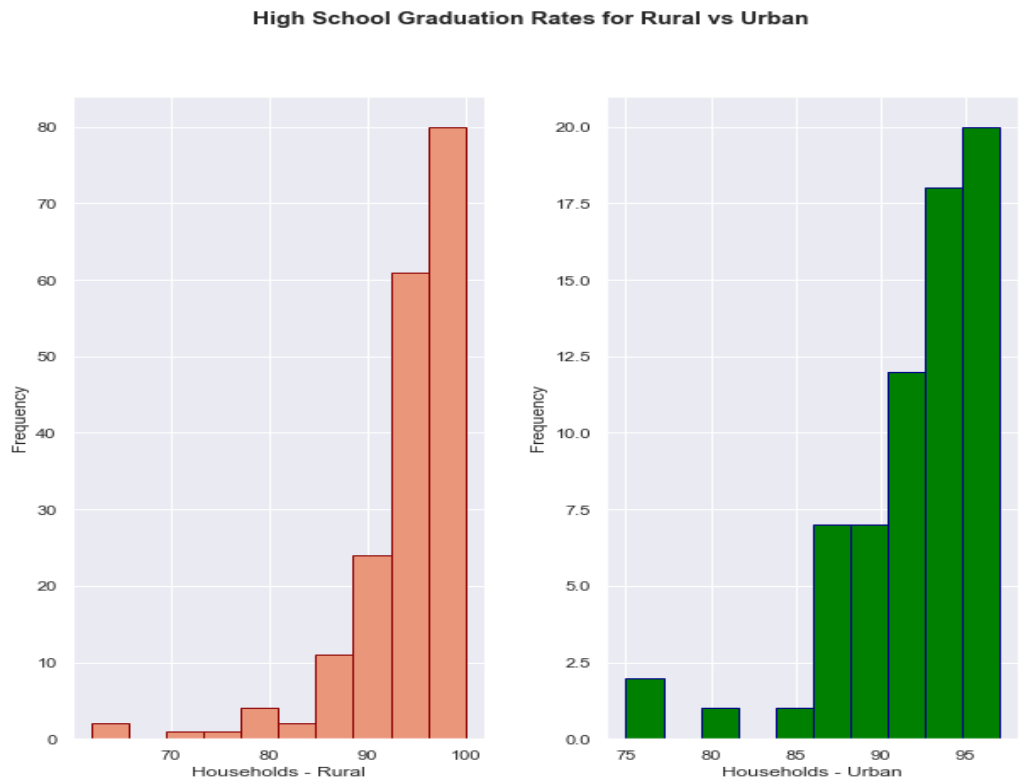
```
hs_df_urban.describe()
```

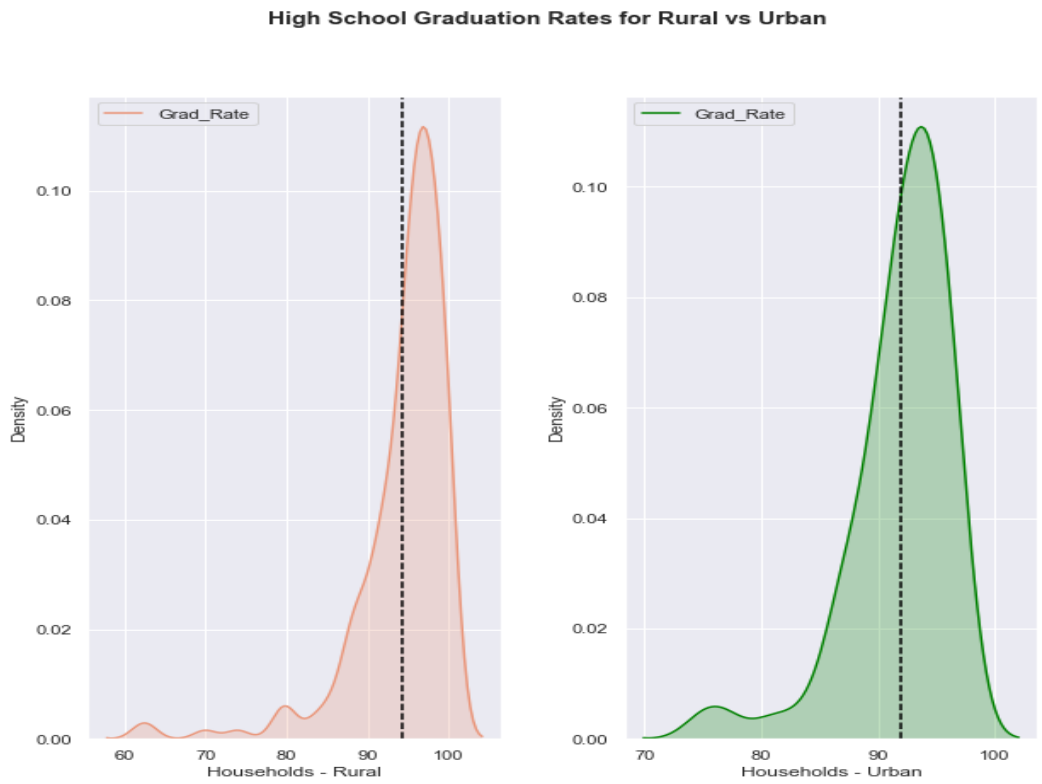
	Grad_Rate	Math	Read
count	68.000000	68.000000	68.000000
mean	91.926471	3.085294	2.894118
std	4.341113	0.197948	0.176941
min	75.000000	2.600000	2.500000
25%	90.000000	3.000000	2.800000
50%	93.000000	3.100000	2.900000
75%	95.000000	3.200000	3.000000
max	97.000000	3.600000	3.400000

QUESTION 2 – CORRELATION COEFFICIENTS

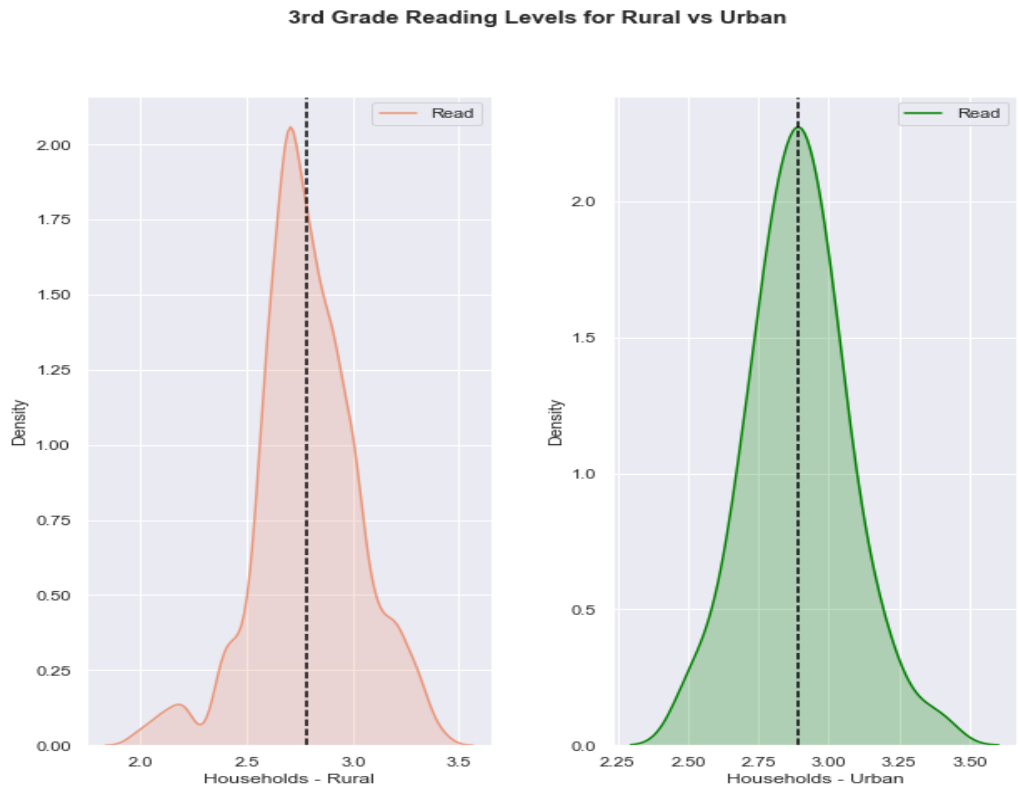
```
Pearson correlation coefficient Rural :
                Read        Math   Grad_Rate
Read        1.000000    0.739300    0.192063
Math        0.739300    1.000000    0.144729
Grad_Rate   0.192063    0.144729    1.000000

Pearson correlation coefficient Urban:
                Read        Math   Grad_Rate
Read        1.000000    0.875327    0.424968
Math        0.875327    1.000000    0.377365
Grad_Rate   0.424968    0.377365    1.000000

Spearman correlation coefficient Rural :
                Read        Math   Grad_Rate
Read        1.000000    0.725848    0.278714
Math        0.725848    1.000000    0.173201
Grad_Rate   0.278714    0.173201    1.000000

Spearman correlation coefficient Urban :
                Read        Math   Grad_Rate
Read        1.000000    0.838901    0.438993
Math        0.838901    1.000000    0.317306
Grad_Rate   0.438993    0.317306    1.000000
```

QUESTION 2: HIGH SCHOOL GRADUATION RATES
HISTOGRAM AND KDE

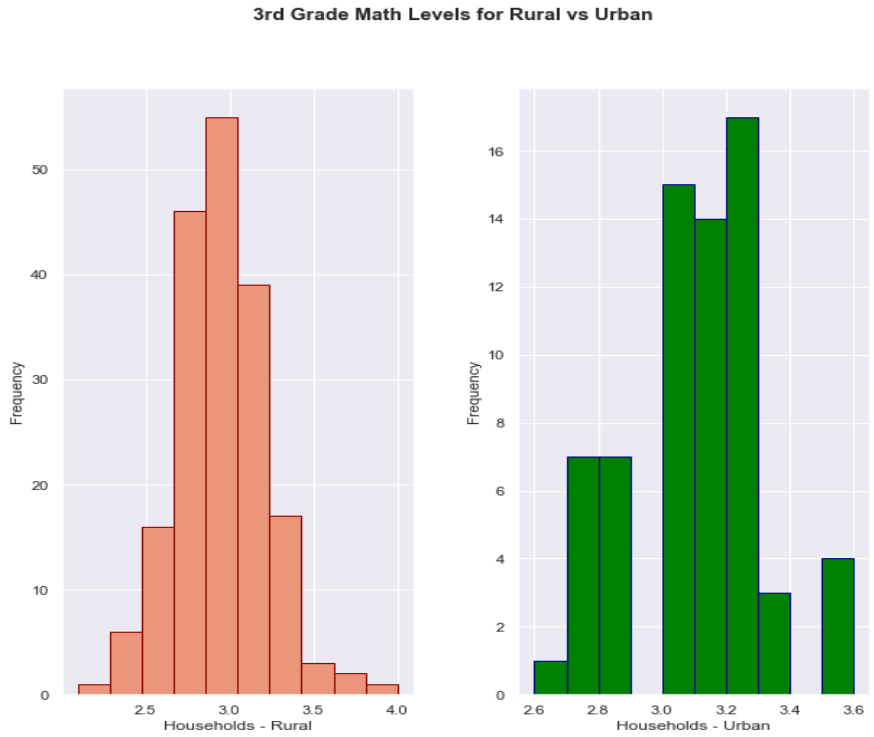


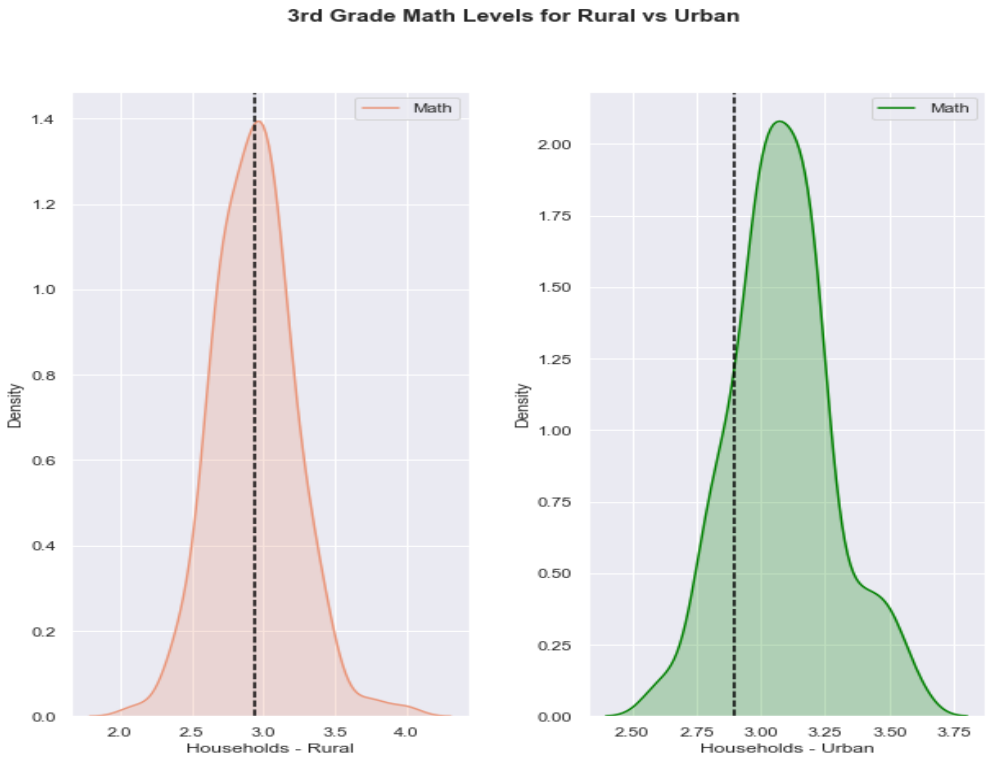


QUESTION 2: 3RD GRADE READING LEVELS
HISTOGRAM AND KDE

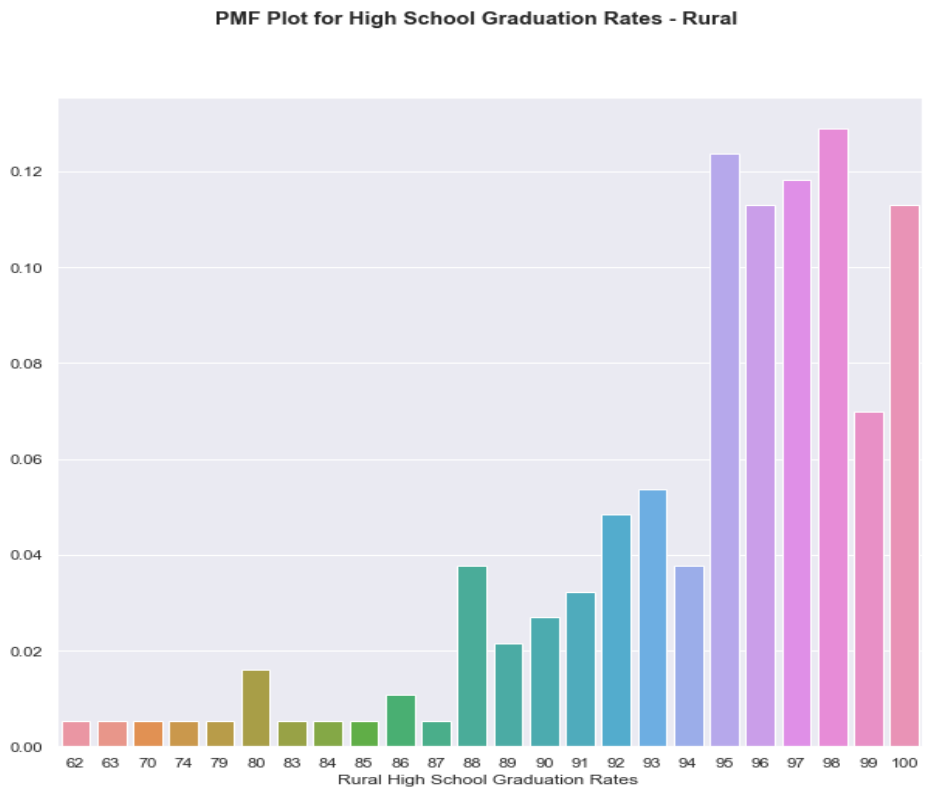


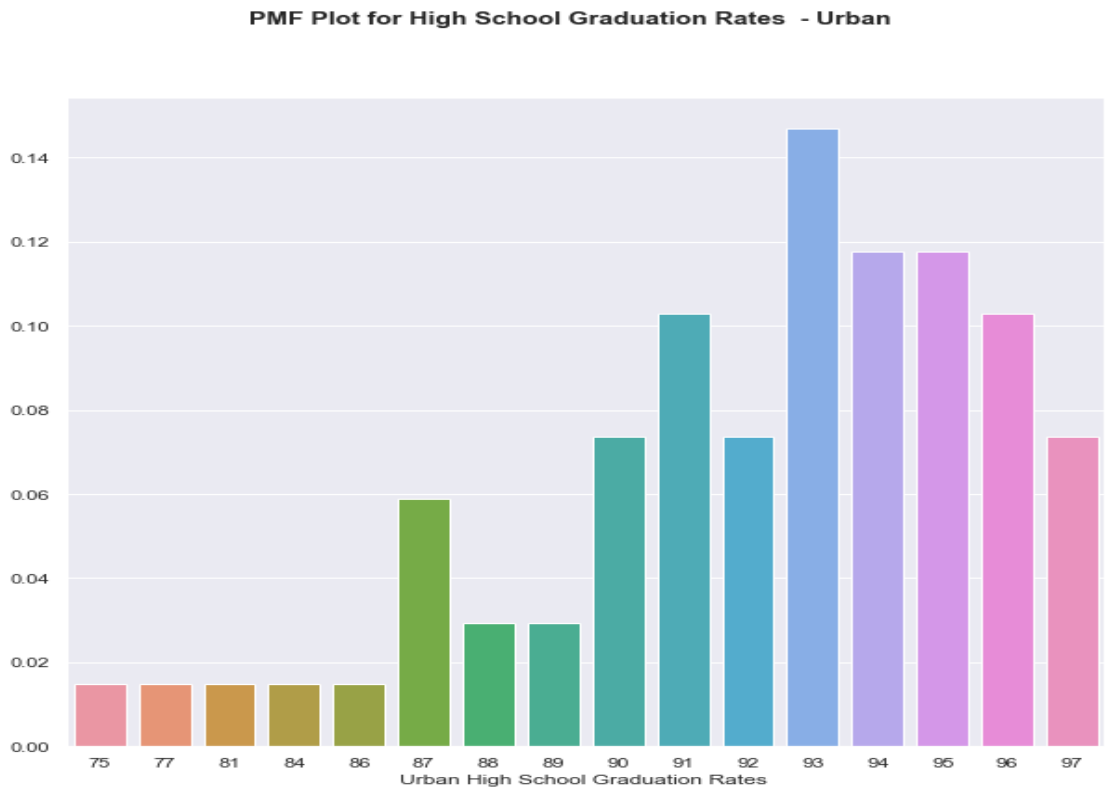
QUESTION 2: 3RD GRADE MATH LEVELS
HISTOGRAM AND KDE



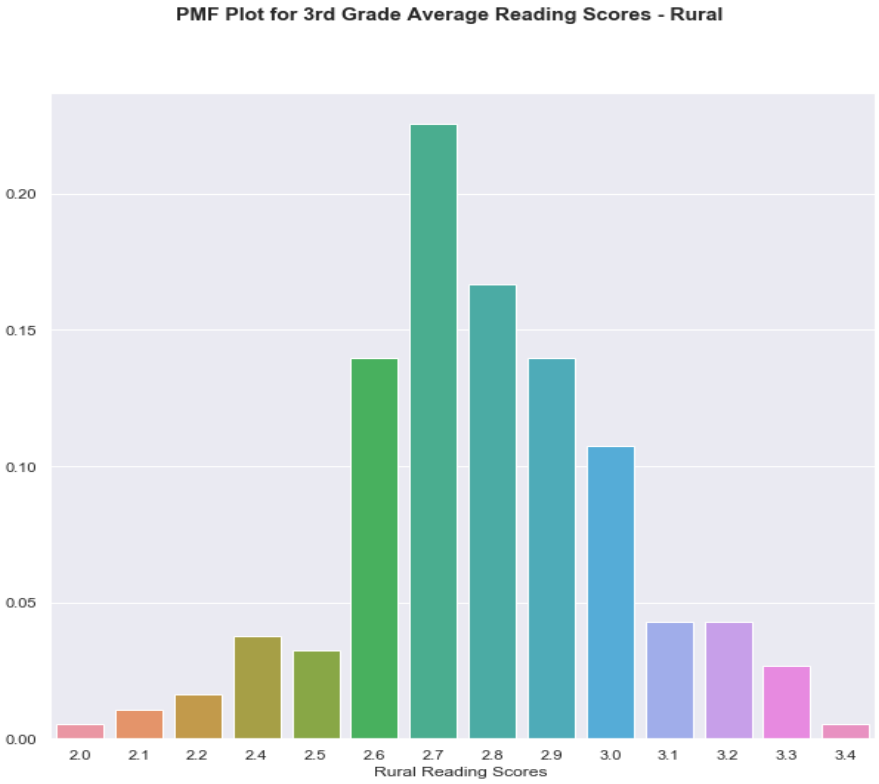


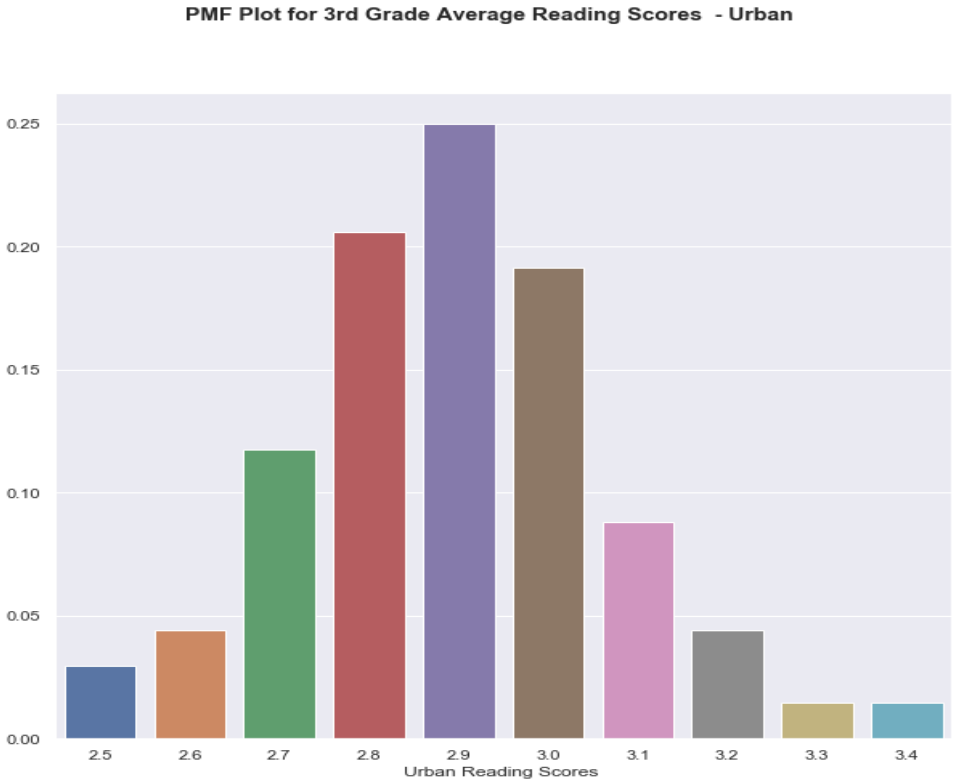
QUESTION 2: HIGH SCHOOL GRADUATION RATES
PMF



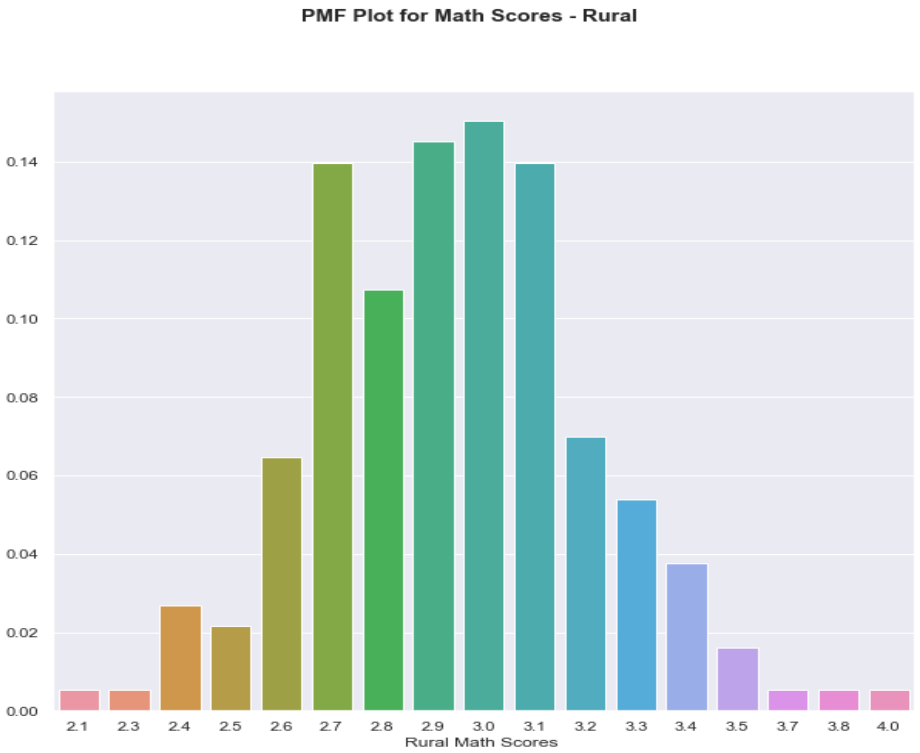


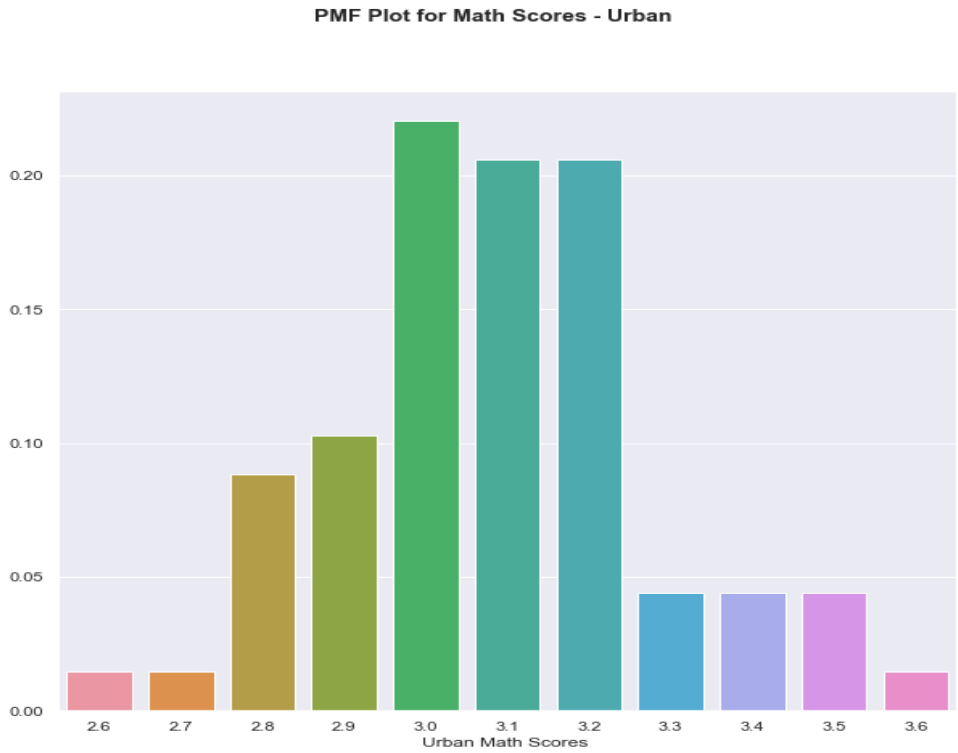
QUESTION 2: 3RD GRADE READING LEVELS
PMF





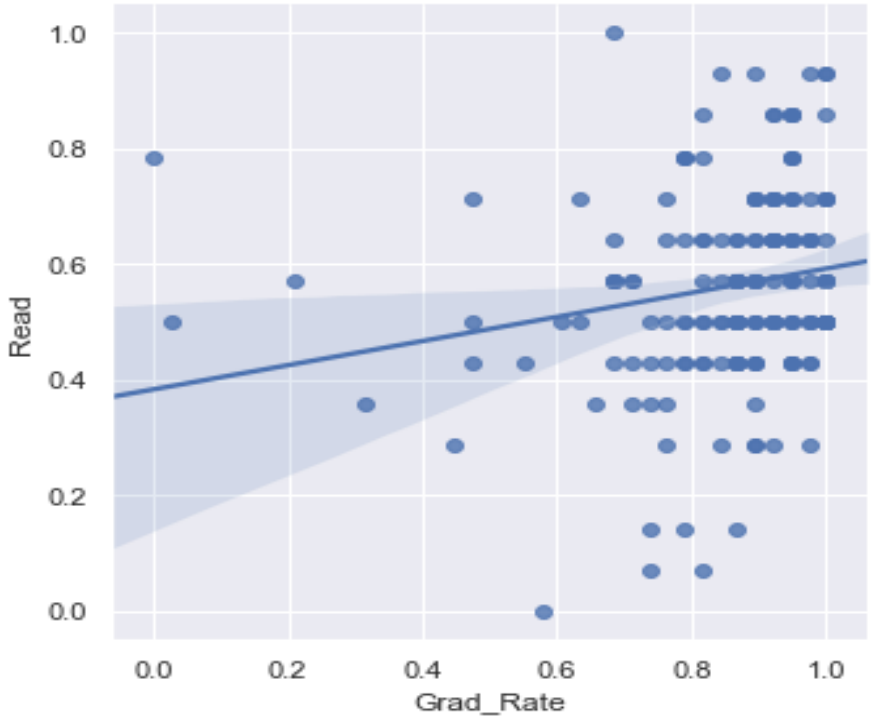
QUESTION 2: 3RD GRADE MATH LEVELS
PMF



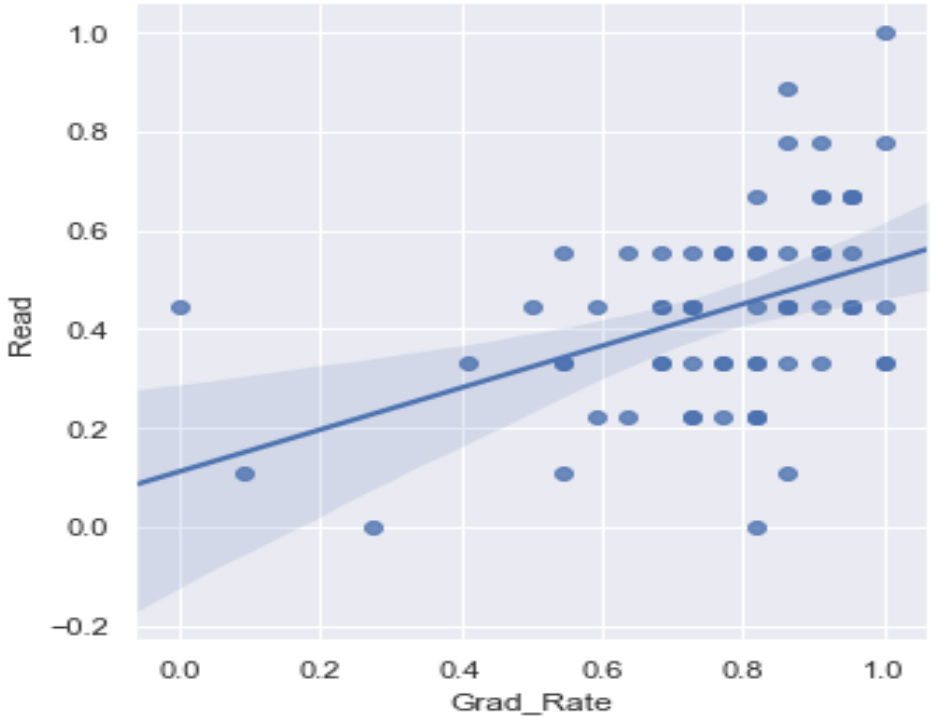


QUESTION 2: GRADUATION RATES AND READING LEVELS
LINEAR REGRESSION

Rural

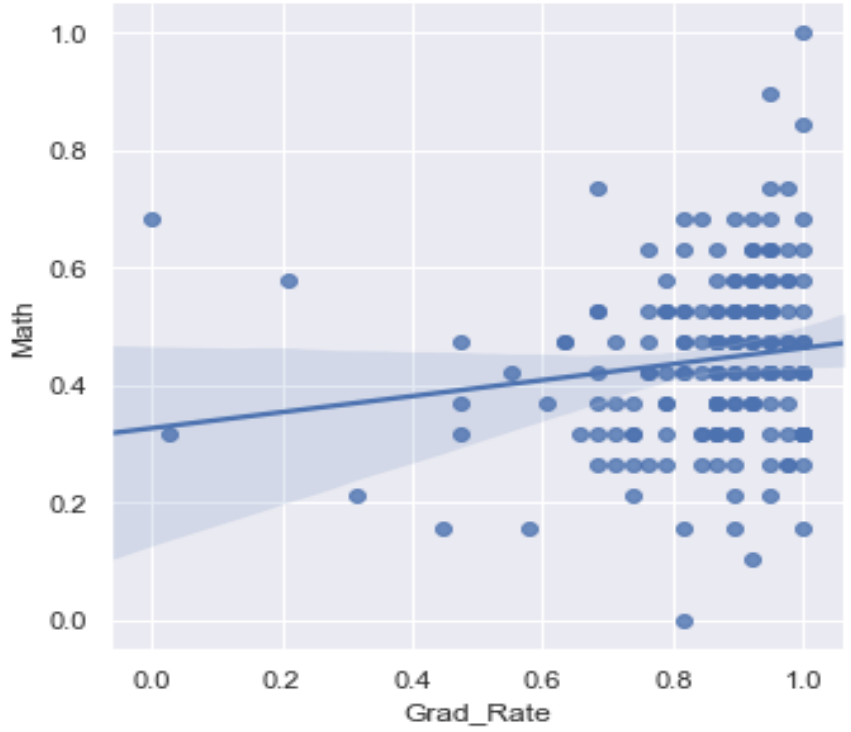


Urban

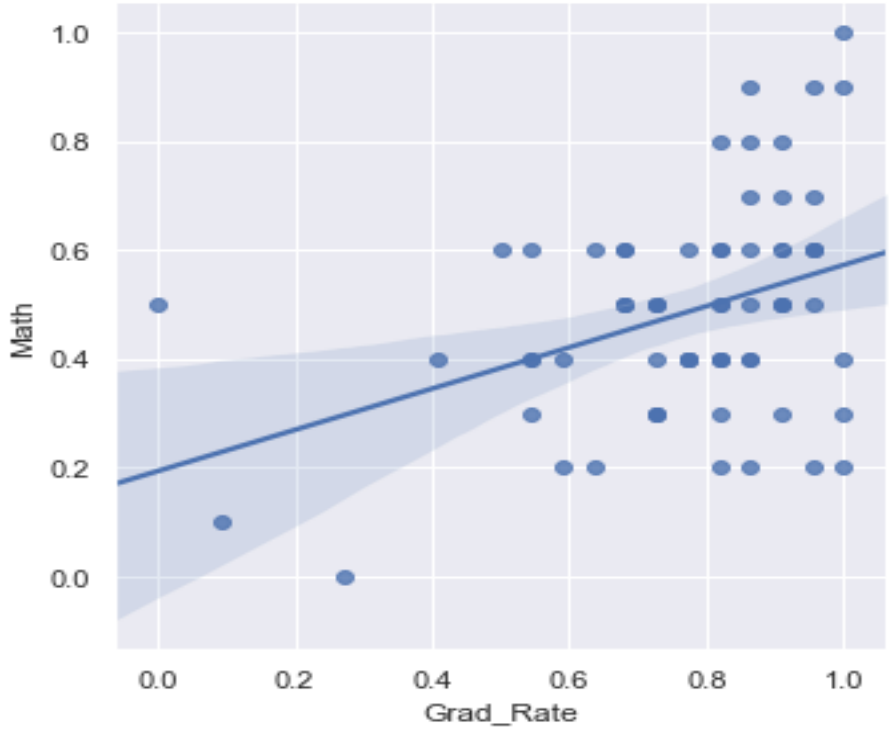


QUESTION 2: GRADUATION RATES AND MATH LEVELS
LINEAR REGRESSION

Rural



Urban



SCENARIO 3 – RESULTS

- Life Expectancy:
  - *Urban sample had close positive correlation with poor health.*
  - *Rural sample has a small negative correlation with all factors.*
- Poor health:
  - *Close positive correlation with all factors especially lack of sleep.*
  - *Similar results for both samples*

QUESTION 3 – DESCRIPTIVE STATISTICS
RURAL

RURAL

	Population	poor_health	poor_sleep	Uninsured_health	\
count	186.000000	186.000000	186.000000	186.000000	
mean	14770.026882	3064.365591	4692.655914	2444.564516	
std	12605.944718	2673.077929	4040.692191	2083.528517	
min	152.000000	23.000000	50.000000	27.000000	
25%	4470.000000	857.250000	1418.500000	812.750000	
50%	10989.000000	2335.500000	3414.500000	1827.500000	
75%	21421.000000	4372.750000	6814.750000	3549.000000	
max	49728.000000	14489.000000	16987.000000	9222.000000	

	Life_Expectancy	Physicians	food_Insecure	Cost_Burden
count	186.000000	186.000000	186.000000	186.000000
mean	77.393548	5.317204	2178.870968	496.537634
std	2.376930	6.139057	2018.025879	505.374179
min	71.900000	0.000000	10.000000	0.000000
25%	75.800000	1.000000	572.500000	120.000000
50%	77.300000	3.000000	1490.000000	336.000000
75%	78.975000	8.000000	3105.000000	656.750000
max	89.700000	38.000000	9970.000000	2451.000000

QUESTION 3 – DESCRIPTIVE STATISTICS
URBAN

URBAN

	Population	poor_health	poor_sleep	Uninsured_health	\
count	6.800000e+01	68.000000	6.800000e+01	68.000000	
mean	3.816856e+05	74733.455882	1.260119e+05	62499.058824	
std	7.236031e+05	141018.568160	2.434585e+05	135041.924584	
min	5.003100e+04	8414.000000	1.508600e+04	6687.000000	
25%	6.795200e+04	14305.500000	2.276400e+04	11253.500000	
50%	1.330275e+05	23471.000000	4.166000e+04	18019.500000	
75%	3.150315e+05	68266.000000	1.077918e+05	41986.500000	
max	4.698619e+06	885812.000000	1.593969e+06	908742.000000	

	Life_Expectancy	Physicians	food_Insecure	Cost_Burden
count	68.000000	68.000000	68.000000	68.000000
mean	78.020588	239.191176	54094.852941	16881.529412
std	2.294286	462.947283	110360.809809	36246.466308
min	73.300000	6.000000	4210.000000	1182.000000
25%	76.200000	23.750000	10707.500000	2361.000000
50%	78.300000	74.000000	17410.000000	4829.500000
75%	79.425000	232.500000	47455.000000	14472.000000
max	83.000000	2742.000000	739120.000000	240521.000000

QUESTION 3 – CORRELATION COEFFICIENTS
RURAL

Pearson correlation coefficient Rural :

	Population	poor_health	poor_sleep	Uninsured_health	\
Population	1.000000	0.949782	0.997272	0.968661	
poor_health	0.949782	1.000000	0.961490	0.949579	
poor_sleep	0.997272	0.961490	1.000000	0.968746	
Uninsured_health	0.968661	0.949579	0.968746	1.000000	
Life_Expectancy	-0.169082	-0.157482	-0.179557	-0.147392	
Physicians	0.802969	0.716882	0.785917	0.765350	
food_Insecure	0.949488	0.866604	0.948382	0.901662	
Cost_Burden	0.941897	0.886859	0.938154	0.904360	

	Life_Expectancy	Physicians	food_Insecure	Cost_Burden
Population	-0.169082	0.802969	0.949488	0.941897
poor_health	-0.157482	0.716882	0.866604	0.886859
poor_sleep	-0.179557	0.785917	0.948382	0.938154
Uninsured_health	-0.147392	0.765350	0.901662	0.904360
Life_Expectancy	1.000000	-0.047904	-0.253418	-0.171704
Physicians	-0.047904	1.000000	0.738383	0.807168
food_Insecure	-0.253418	0.738383	1.000000	0.927326
Cost_Burden	-0.171704	0.807168	0.927326	1.000000

QUESTION 3 – CORRELATION COEFFICIENTS
URBAN

Pearson correlation coefficient Urban :

	Population	poor_health	poor_sleep	Uninsured_health	\
Population	1.000000	0.986811	0.999027	0.983531	
poor_health	0.986811	1.000000	0.989802	0.985677	
poor_sleep	0.999027	0.989802	1.000000	0.985493	
Uninsured_health	0.983531	0.985677	0.985493	1.000000	
Life_Expectancy	0.343420	0.330408	0.330367	0.293629	
Physicians	0.985984	0.961588	0.980984	0.950631	
food_Insecure	0.989885	0.968455	0.989285	0.984261	
Cost_Burden	0.992718	0.978460	0.991086	0.986810	

	Life_Expectancy	Physicians	food_Insecure	Cost_Burden
Population	0.343420	0.985984	0.989885	0.992718
poor_health	0.330408	0.961588	0.968455	0.978460
poor_sleep	0.330367	0.980984	0.989285	0.991086
Uninsured_health	0.293629	0.950631	0.984261	0.986810
Life_Expectancy	1.000000	0.368765	0.283458	0.299450
Physicians	0.368765	1.000000	0.971714	0.978157
food_Insecure	0.283458	0.971714	1.000000	0.993071
Cost_Burden	0.299450	0.978157	0.993071	1.000000

QUESTION 3 – OLS MODEL
LIFE EXPECTANCY

RURAL

OLS Regression Results

Dep. Variable:	life	R-squared:	0.212
Model:	OLS	Adj. R-squared:	0.123
Method:	Least Squares	F-statistic:	2.376
Date:	Thu, 28 May 2020	Prob (F-statistic):	0.0417
Time:	18:54:36	Log-Likelihood:	49.587
No. Observations:	60	AIC:	-85.17
Df Residuals:	53	BIC:	-70.51
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.2873	0.024	12.035	0.000	0.239	0.335
health	-0.0988	0.396	-0.249	0.804	-0.894	0.696
sleep	0.0571	0.487	0.117	0.907	-0.920	1.034
uninsured	0.1549	0.314	0.494	0.623	-0.474	0.784
docs	0.3275	0.200	1.634	0.108	-0.074	0.729
food	-0.4867	0.362	-1.344	0.185	-1.213	0.240
burden	0.0758	0.341	0.223	0.825	-0.608	0.759

Omnibus:	2.409	Durbin-Watson:	2.265
Prob(Omnibus):	0.300	Jarque-Bera (JB):	1.619
Skew:	0.220	Prob(JB):	0.445
Kurtosis:	3.674	Cond. No.	54.7

–

URBAN

OLS Regression Results

Dep. Variable:	life1	R-squared:	0.492
Model:	OLS	Adj. R-squared:	0.434
Method:	Least Squares	F-statistic:	8.541
Date:	Thu, 28 May 2020	Prob (F-statistic):	1.67e-06
Time:	18:54:37	Log-Likelihood:	24.084
No. Observations:	60	AIC:	-34.17
Df Residuals:	53	BIC:	-19.51
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.4333	0.029	14.940	0.000	0.375	0.491
health1	0.0309	1.951	0.016	0.987	-3.883	3.944
sleep1	4.2851	2.546	1.683	0.098	-0.822	9.393
uninsured1	0.4955	1.930	0.257	0.798	-3.375	4.366
docs1	3.2563	0.914	3.562	0.001	1.422	5.090
food1	-8.2411	2.112	-3.902	0.000	-12.477	-4.005
burden1	0.4006	1.819	0.220	0.827	-3.249	4.050

Omnibus:	0.808	Durbin-Watson:	2.328
Prob(Omnibus):	0.668	Jarque-Bera (JB):	0.555
Skew:	-0.236	Prob(JB):	0.758
Kurtosis:	2.998	Cond. No.	161.

QUESTION 3 – OLS MODEL
POOR HEALTH

RURAL

OLS Regression Results

Dep. Variable:	health	R-squared:	0.962
Model:	OLS	Adj. R-squared:	0.958
Method:	Least Squares	F-statistic:	226.4
Date:	Thu, 28 May 2020	Prob (F-statistic):	6.13e-36
Time:	18:55:34	Log-Likelihood:	113.20
No. Observations:	60	AIC:	-212.4
Df Residuals:	53	BIC:	-197.7
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0134	0.016	0.843	0.403	-0.018	0.045
life	-0.0119	0.048	-0.249	0.804	-0.107	0.084
sleep	0.7822	0.130	6.010	0.000	0.521	1.043
uninsured	0.3364	0.099	3.412	0.001	0.139	0.534
docs	-0.2776	0.060	-4.621	0.000	-0.398	-0.157
food	-0.5702	0.101	-5.662	0.000	-0.772	-0.368
burden	0.2940	0.111	2.648	0.011	0.071	0.517

Omnibus:	7.787	Durbin-Watson:	2.153
Prob(Omnibus):	0.020	Jarque-Bera (JB):	8.483
Skew:	0.546	Prob(JB):	0.0144
Kurtosis:	4.483	Cond. No.	41.8

URBAN

OLS Regression Results

Dep. Variable:	health1	R-squared:	0.995
Model:	OLS	Adj. R-squared:	0.994
Method:	Least Squares	F-statistic:	1720.
Date:	Thu, 28 May 2020	Prob (F-statistic):	7.26e-59
Time:	18:55:34	Log-Likelihood:	183.30
No. Observations:	60	AIC:	-352.6
Df Residuals:	53	BIC:	-337.9
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.0067	0.005	1.473	0.147	-0.002	0.016
life1	0.0002	0.010	0.016	0.987	-0.019	0.020
sleep1	0.9546	0.129	7.396	0.000	0.696	1.214
uninsured1	0.7020	0.096	7.328	0.000	0.510	0.894
docs1	-0.0116	0.072	-0.162	0.872	-0.155	0.132
food1	-0.5735	0.149	-3.845	0.000	-0.873	-0.274
burden1	-0.0463	0.128	-0.362	0.719	-0.303	0.210

Omnibus:	12.072	Durbin-Watson:	1.818
Prob(Omnibus):	0.002	Jarque-Bera (JB):	23.230
Skew:	0.543	Prob(JB):	9.03e-06
Kurtosis:	5.848	Cond. No.	140.

SCENARIO 3: HYPOTHESIS TESTING

Paired Student's t-test Rural and Urban

t= 6.79875785 5.51777929 6.56017156 6.5975982 -7.00952344 2.69525087 5.53527142 5.12666291

p = 5.92225149e-09 8.03380888e-07 1.49548235e-08 1.29351795e-08 2.60678077e-09 9.14884204e-03 7.52365180e-07 3.42963082e-06

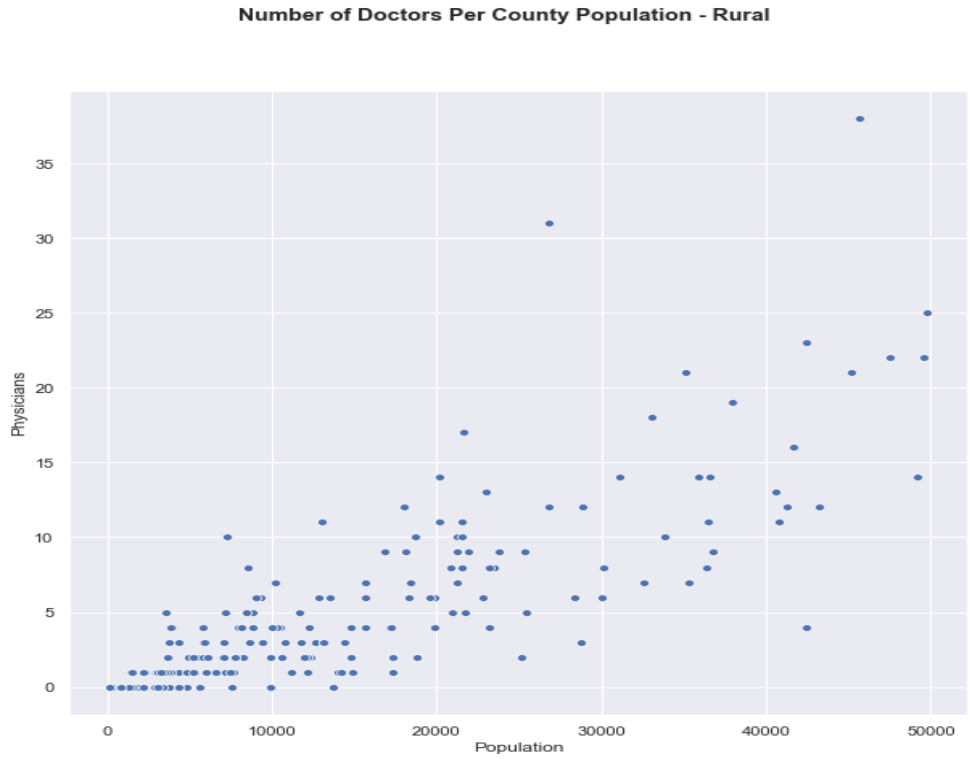
Student's t-test Rural and Urban

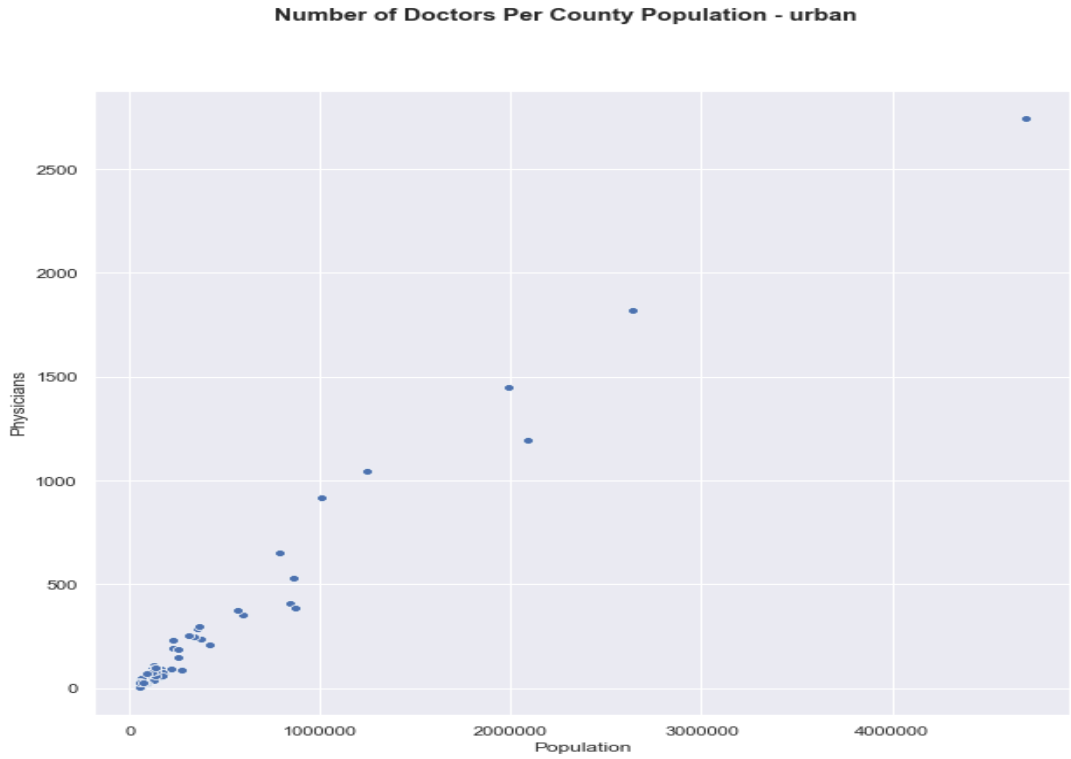
t= 6.79323141 5.31725255 6.55877629 6.79038486 -7.51255979 2.35885894 5.56395097 5.00468522

p = 7.85440886e-11 2.32278625e-07 3.05527250e-10 7.98651687e-11 1.00808010e-12 1.90950891e-02 6.72848316e-08 1.05122247e-06

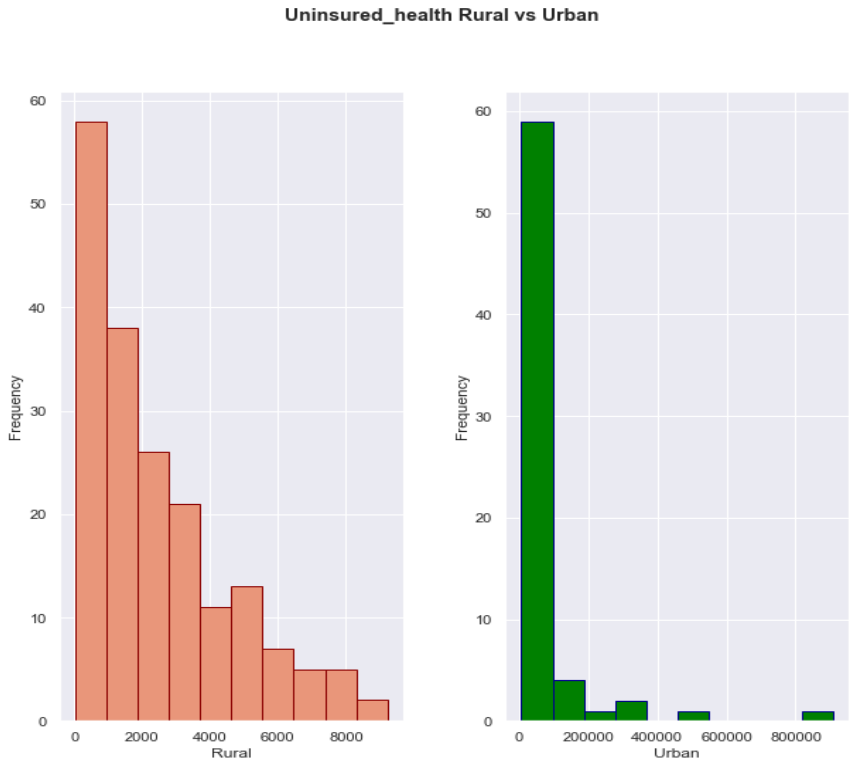
Variables: *Population, poor health, poor sleep, Uninsured, Life Expectancy, Physicians, food insecurity, Housing Cost Burden*

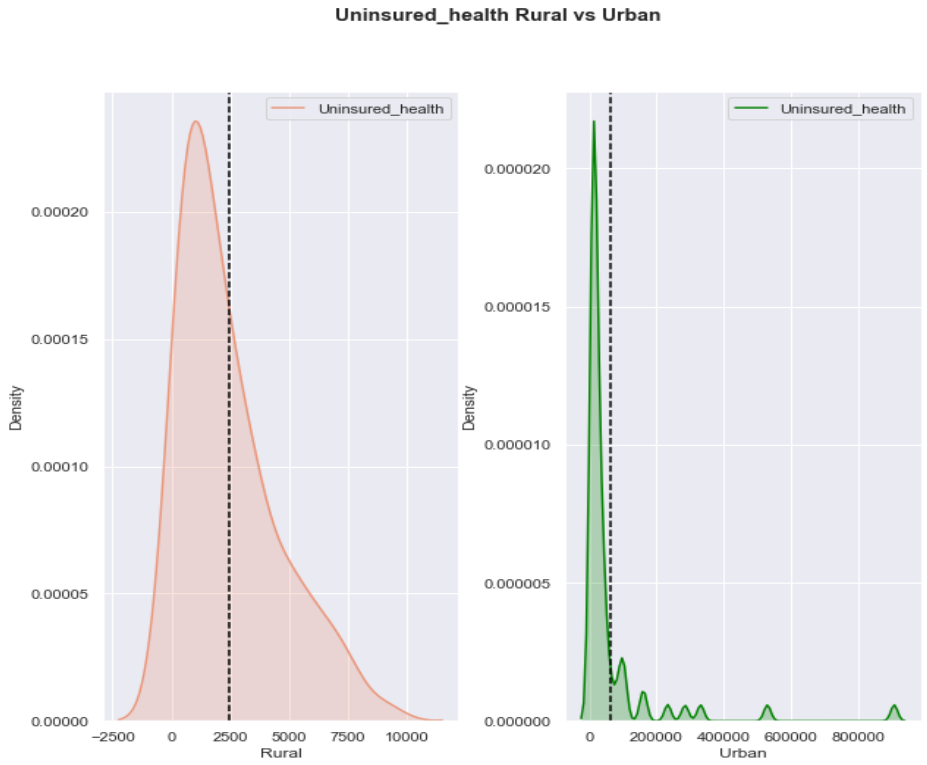
SCENARIO 3 – SCATTERPLOTS
PHYSICIANS PER COUNTY



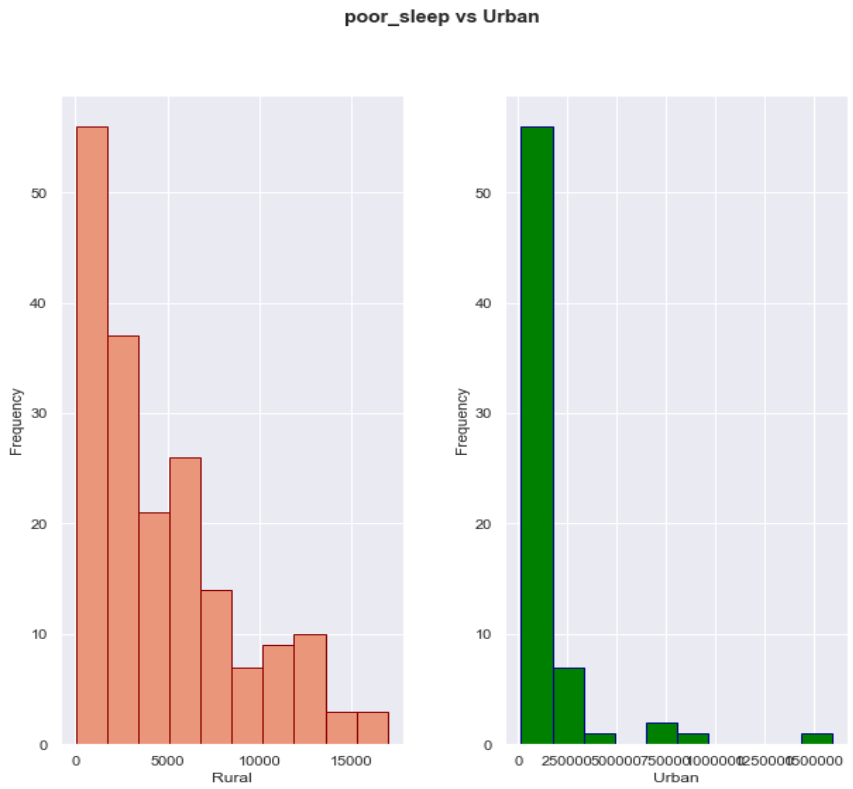


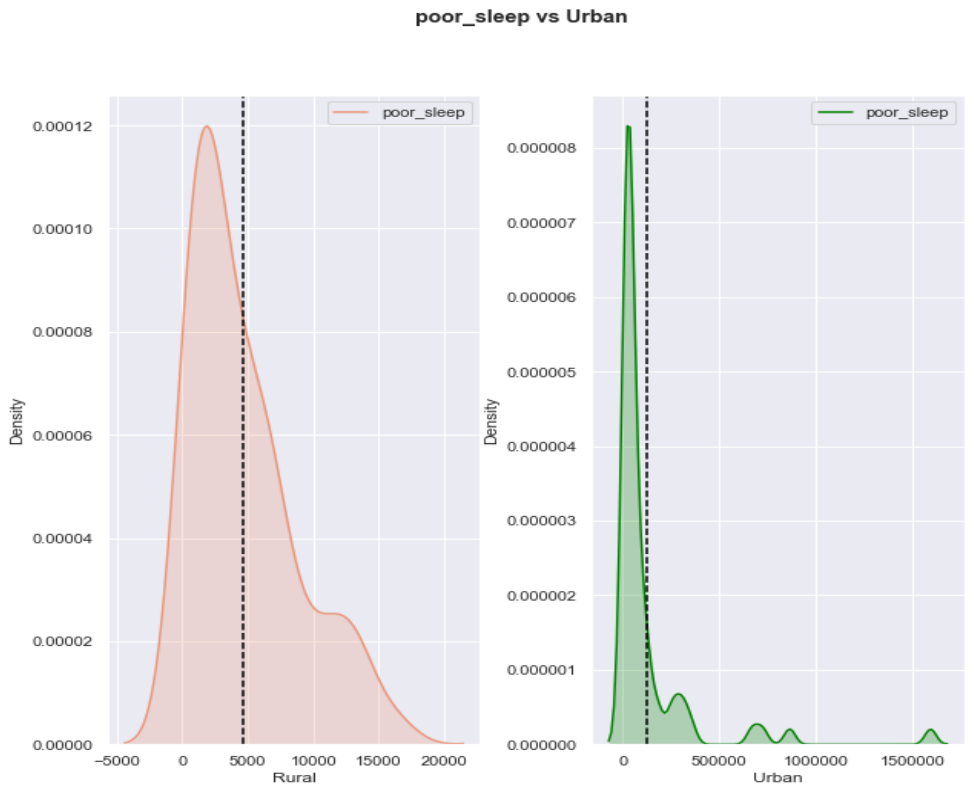
SCENARIO 3 – HISTOGRAMS AND KDE
UNINSURED



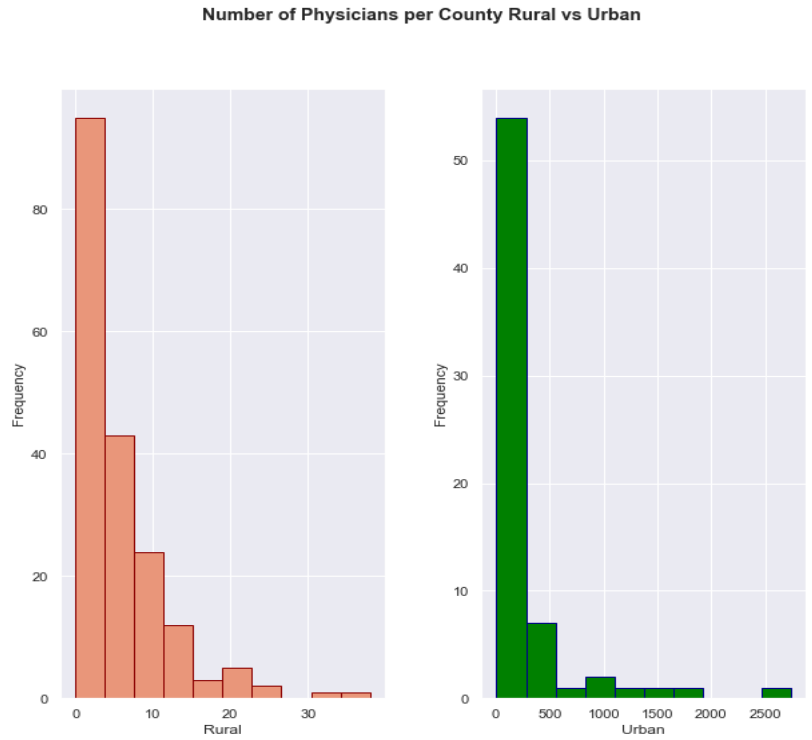


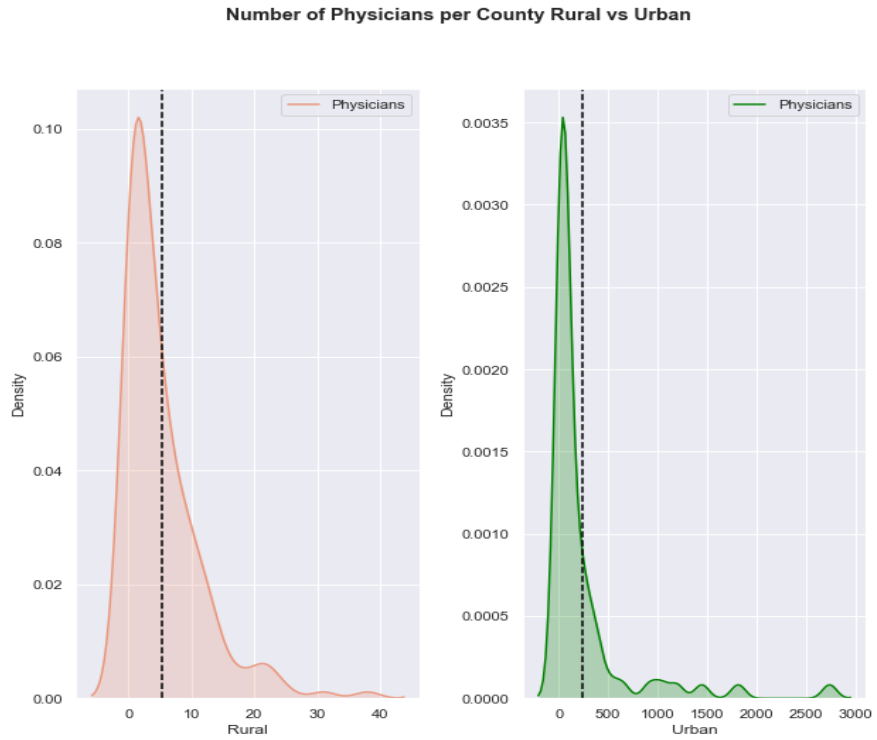
SCENARIO 3 – HISTOGRAMS AND KDE
POOR SLEEP



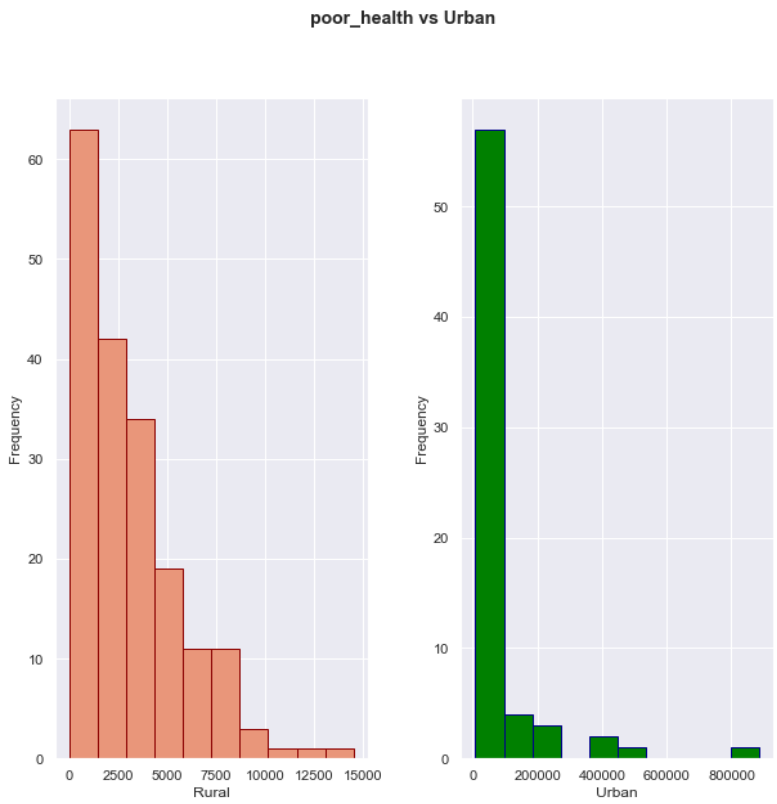


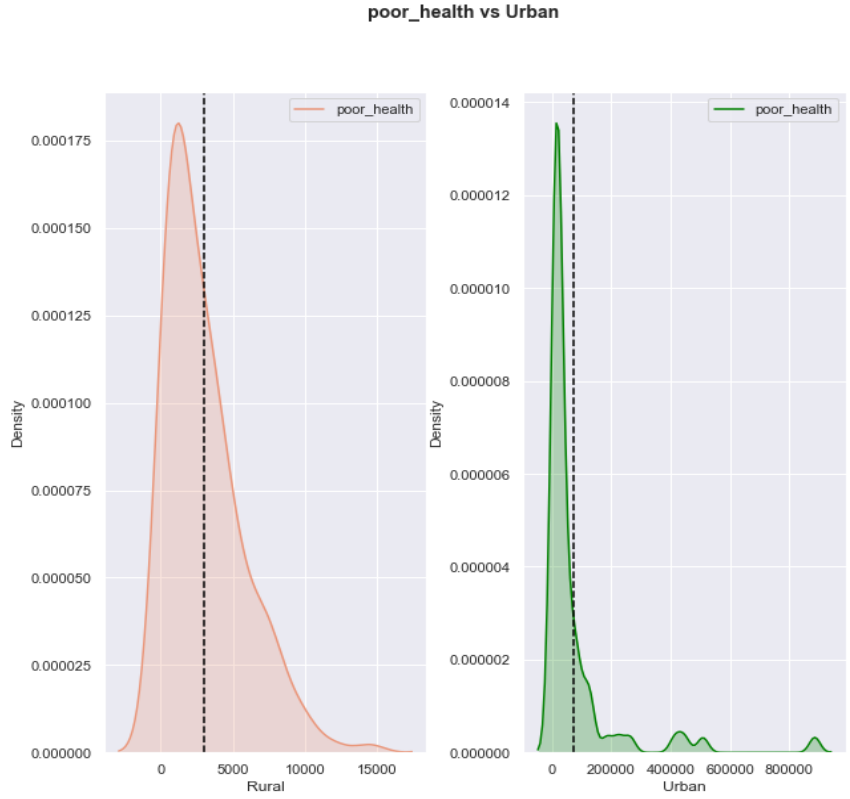
SCENARIO 3 – HISTOGRAMS AND KDE
NUMBER OF PHYSICIANS



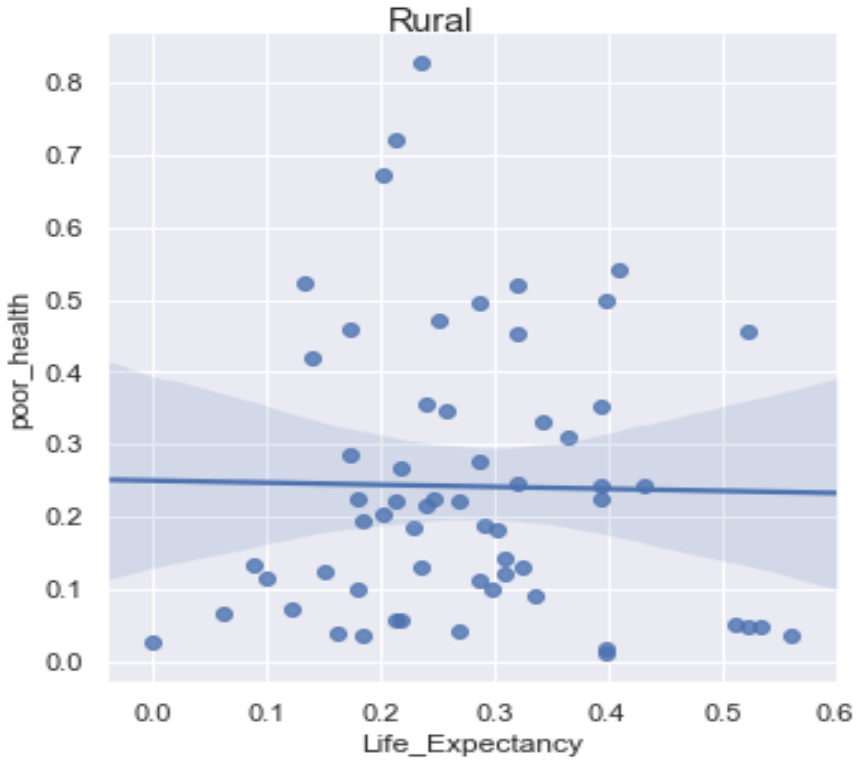


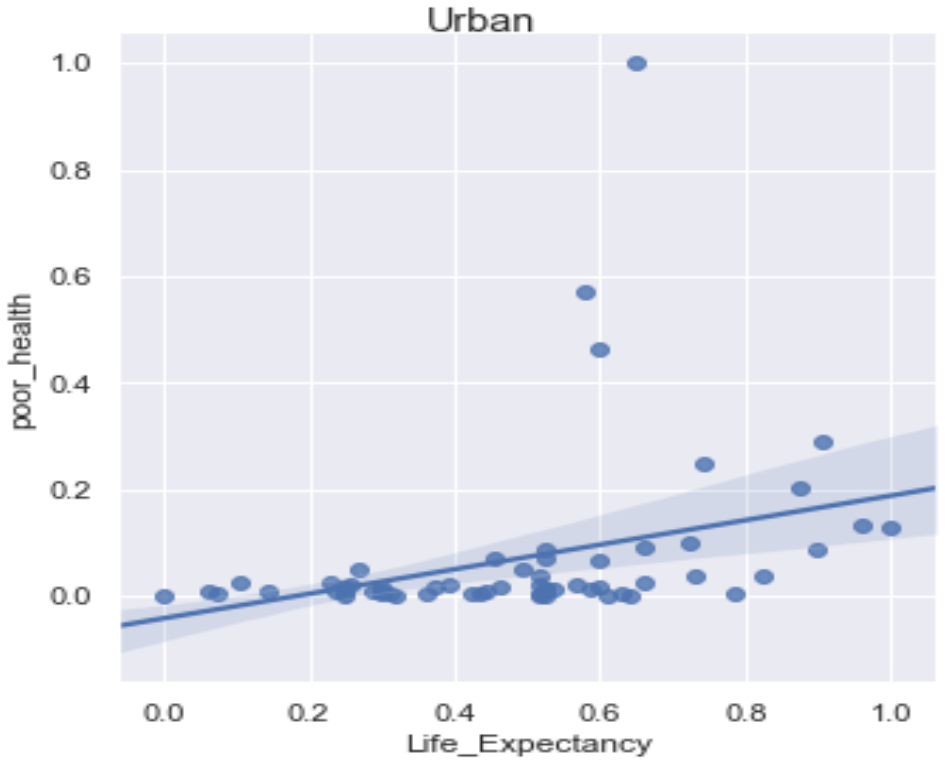
SCENARIO 3 – HISTOGRAMS AND KDE
POOR HEALTH



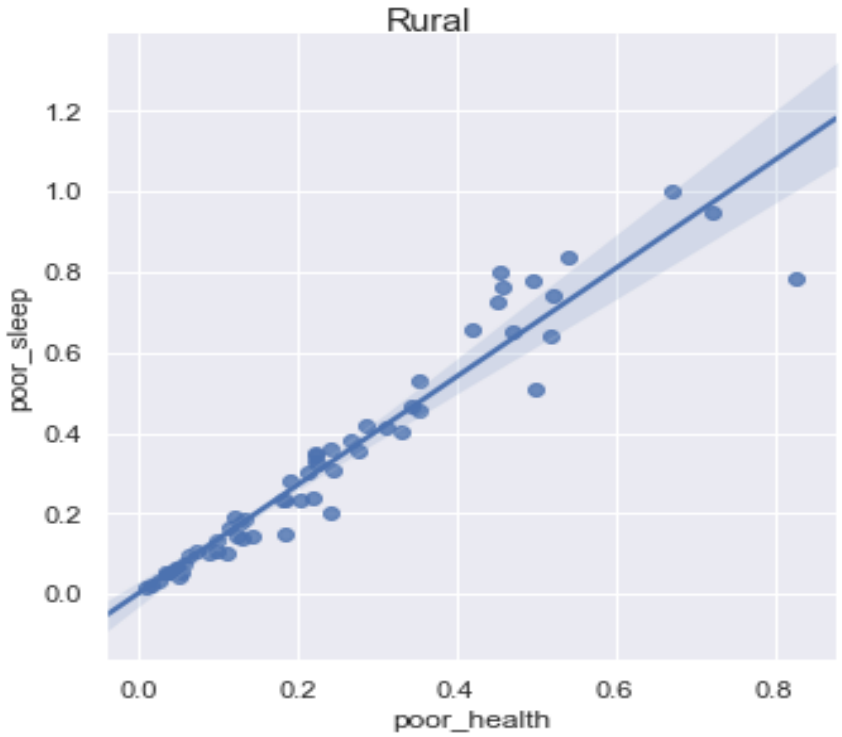


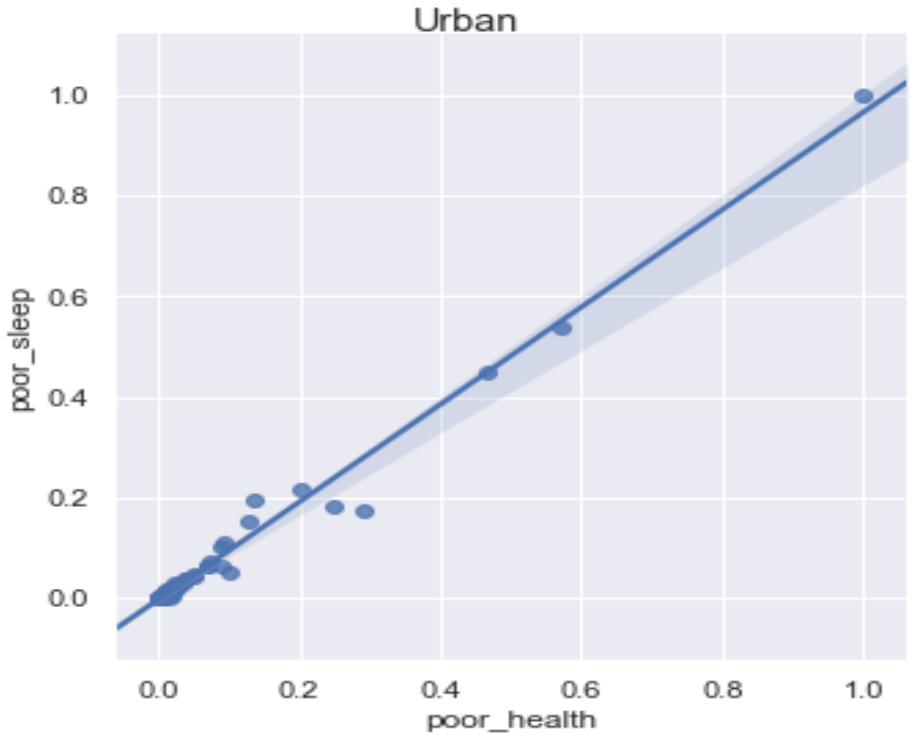
SCENARIO 3 – LINEAR REGRESSION PLOTS
LIFE EXPECTANCY AND POOR HEALTH



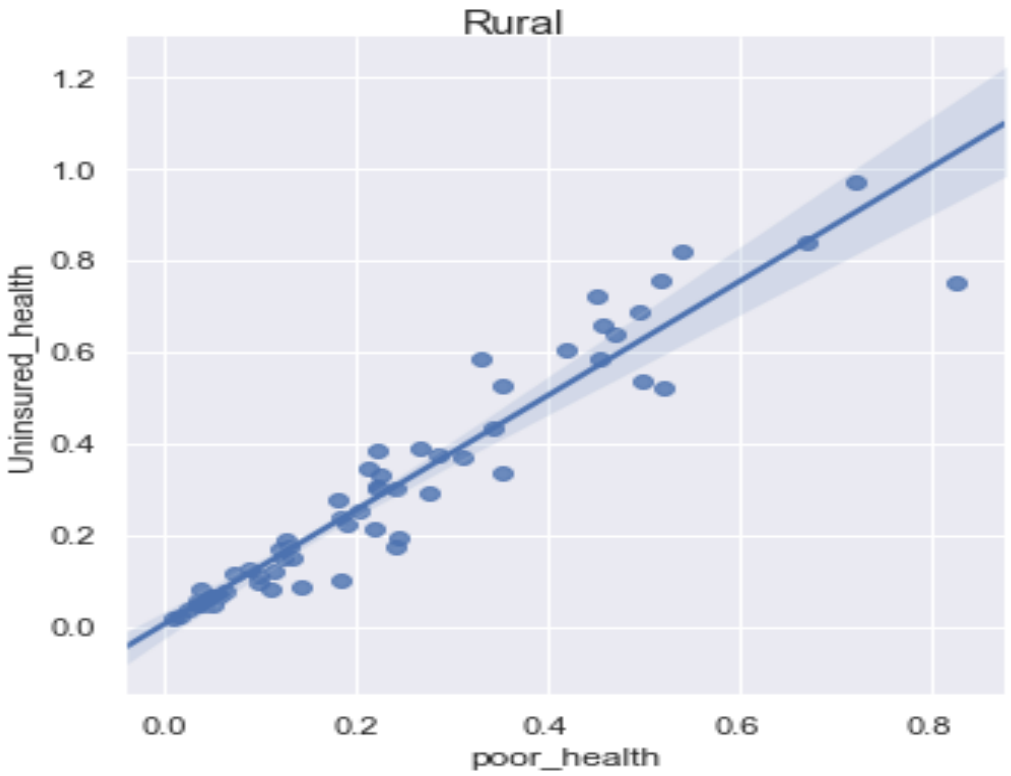


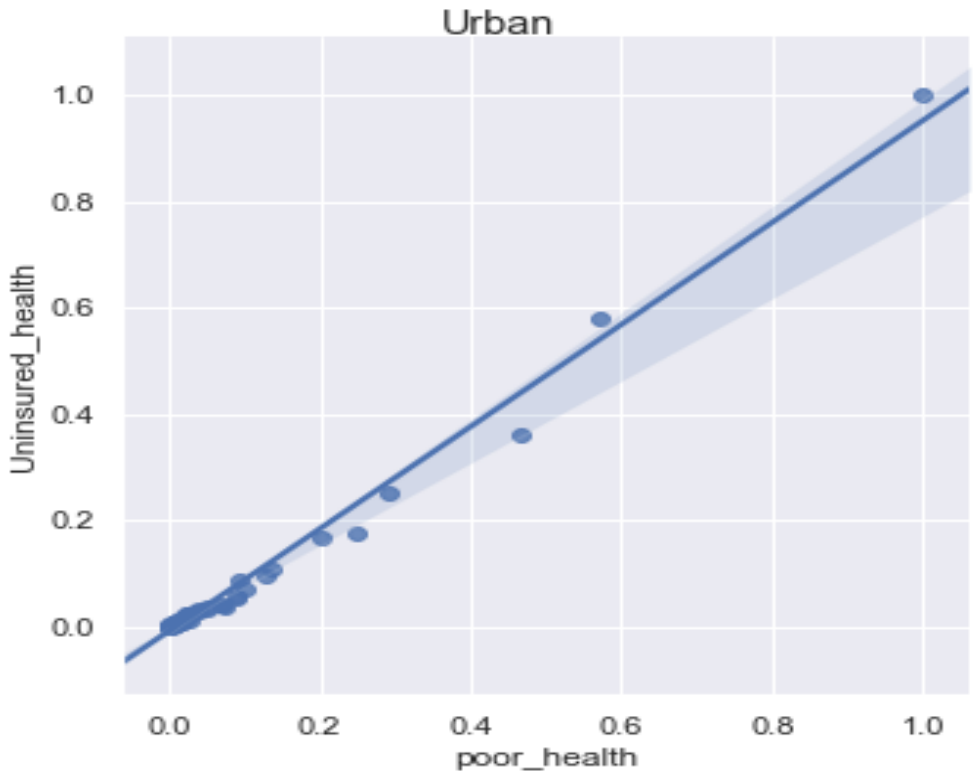
SCENARIO 3 – LINEAR REGRESSION PLOTS POOR HEALTH AND POOR SLEEP





SCENARIO 3 – LINEAR REGRESSION PLOTS
POOR HEALTH AND UNINSURED





SCENARIO 3 – LINEAR REGRESSION PLOTS POOR HEALTH AND FOOD INSECURITIES