STA- 6707 Multivariate Methods Spring 2025
Final SAS Project

Instructor: Dr. Amin Raid
Author: Murpys Duran Mendez

Introduction

Homelessness and housing instability remain pressing social issues in the United States. The dataset provided for this project includes eviction rates and homelessness counts across the country.

In this project, we apply descriptive statistics, correlation analysis, Principal Components Analysis (PCA), and Factor Analysis to explore patterns and underlying connections in the data. This allows us to demonstrate the use of multivariate statistical techniques while examining the relationships among homelessness, evictions, and related socioeconomic factors.

The data

The dataset used in this project contains information collected from multiple states, providing a basis for the analyses that follow. The study is focused on some of the variables presented below.

- county
- totalhomeless
- county_population
- eviction_filings
- renting_household_population
- z_med_rent
- z_med_inc
- pct_white
- pct_african_american
- pct_latinx
- pct_smokers
- pct_obese
- pct_unemployed
- pct_high_school_graduation
- violent_crime_rate
- pct_poverty
- pop_per_sq_mile

The complete dataset was imported into SAS from a CSV file, and a smaller version was sampled for the states of Oklahoma, Arkansas, Georgia, Ohio, Texas, and Montana. Figure 1 displays a sample portion of the dataset as it appears in SAS.

Total rows: 3108 Total columns: 28

	fips	state	county	year	totalhomeless	county_population	eviction_filings
1	48427	Texas	Starr County	2017	43	60968	16
2	48323	Texas	Maverick County	2017	37	54258	94
3	48479	Texas	Webb County	2017	65	250304	712
4	48505	Texas	Zapata County	2017	12	14018	7
5	46113	South Dakota	Shannon County	2017	7	13586	1
6	48507	Texas	Zavala County	2017	8	11677	15
7	48047	Texas	Brooks County	2017	7	7223	8
8	48215	Texas	Hidalgo County	2017	427	774769	2095
9	48247	Texas	Jim Hogg County	2017	4	5300	7
10	46121	South Dakota	Todd County	2017	7	9612	1
11	35031	New Mexico	McKinley County	2017	15	71492	92
12	48131	Texas	Duval County	2017	6	11782	9
13	48061	Texas	Cameron County	2017	277	406220	1005
14	36005	New York	Bronx County	2017	16136	1385108	86888
15	48127	Texas	Dimmit County	2017	7	9996	12
16	55078	Wisconsin	Menominee County	2017	7	4232	0
17	13063	Georgia	Clayton County	2017	406	259424	17695

Figure 1

Table 1 presents the frequency distribution of observations across the selected states in the dataset.

The FREQ Procedure

state	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Arkansas	75	10.58	75	10.58
Georgia	159	22.43	234	33.00
Montana	56	7.90	290	40.90
Ohio	88	12.41	378	53.31
Oklahoma	77	10.86	455	64.17
Texas	254	35.83	709	100.00

Table 1.

The table shows that Texas accounts for almost a third of the work data, followed by Georgia with 22.43 percent. While the rest of the states contribute modestly to the numbers.

Basic Statistics

To better understand the dataset, we computed basic descriptive statistics for all variables. These summaries provide an overview of central tendencies, variability, and overall data distribution and range.

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
fips	709	32353	16273	22938627	5001	48507
year	709	2017	0	1430053	2017	2017
totalhomeless	709	66.61495	203.71454	47230	0	2451
county_population	709	76201	237490	54026402	82.00000	4092459
eviction_filings	709	958.27362	4385	679416	0	59046
renting_household_population	709	11502	42708	8154952	15.00000	739211
z_med_rent	709	0.22000	0.08405	155.98043	0.01196	0.58696
z_med_inc	709	0.24480	0.10095	173.56551	0.04145	0.72479
pct_white	709	0.67699	0.21157	479.98643	0.00728	0.99825
pct_african_american	709	0.11134	0.14925	78.93765	0	0.72291
pct_latinx	709	0.15815	0.20173	112.12900	0	0.99200
primary_care_physician_rate	709	0.09751	0.05630	69.13735	0	0.28438
pct_single_parent_households	709	0.34076	0.10393	241.59755	0	1.00000
pct_smokers	709	0.18017	0.02948	127.74307	0.11764	0.32205
pct_obese	709	0.31083	0.03873	220.37700	0.15500	0.45500
pct_unemployed	709	0.05352	0.01665	37.94857	0.02081	0.13596
pct_high_school_graduation	709	0.88185	0.07524	625.23515	0.32556	1.00000
violent_crime_rate	709	0.15271	0.11403	108.27000	0	1.00000
z_air_rseihazard	709	0.01031	0.06043	7.30744	0	1.00000
z_land_rseihazard	709	0.0000471	0.0005506	0.03340	0	0.01275
z_water_rseihazard	709	0.00172	0.03767	1.21696	0	1.00000
republican_voting_pct	709	0.71134	0.15003	504.33741	0.13916	0.96753
pct_poverty	709	0.97843	0.09410	693.70796	0.65354	1.63415
pop_per_sq_mile	709	0.0003747	0.0009376	0.26564	0	0.00808
lng	709	-93.37170	8.51049	-66201	-115.40512	-80.74863
lat	709	34.96246	4.72419	24788	26.13427	48.78378

Descriptive statistics of the count of homeless population (totalhomelss):

Variable: totalhomeless

Moments			
N	709	Sum Weights	709
Mean	66.6149506	Sum Observations	47230
Std Deviation	203.714545	Variance	41499.6157
Skewness	7.0002738	Kurtosis	59.3776922
Uncorrected SS	32527952	Corrected SS	29381727.9
Coeff Variation	305.809045	Std Error Mean	7.65066031

Basic Statistical Measures			
Location		Variability	
Mean	66.61495	Std Deviation	203.71454
Median	17.00000	Variance	41500
Mode	5.00000	Range	2451
		Interquartile Range	31.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	8.707085	Pr > |t|	<.0001
Sign	M	352	Pr >= |M|	<.0001
Signed Rank	S	124080	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	2451
99%	1162
95%	270
90%	128
75% Q3	38
50% Median	17
25% Q1	7
10%	3
5%	2
1%	1
0% Min	0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	677	1411	512
0	606	1484	682
0	590	1727	308
0	586	2155	470
0	391	2451	556

Descriptive statistics of the count of eviction notices filed (eviction_filings):

Author: Murpys Duran Mendez

Variable: eviction_filings

Moments			
N	709	Sum Weights	709
Mean	958.273625	Sum Observations	679416
Std Deviation	4385.29404	Variance	19230803.9
Skewness	9.12233121	Kurtosis	98.6978887
Uncorrected SS	1.42665E10	Corrected SS	1.36154E10
Coeff Variation	457.624412	Std Error Mean	164.693175

Basic Statistical Measures			
Location		Variability	
Mean	958.2736	Std Deviation	4385
Median	75.0000	Variance	19230804
Mode	0.0000	Range	59046
		Interquartile Range	284.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	5.818539	Pr > |t|	<.0001
Sign	M	345	Pr >= |M|	<.0001
Signed Rank	S	119197.5	Pr >= |S|	<.0001

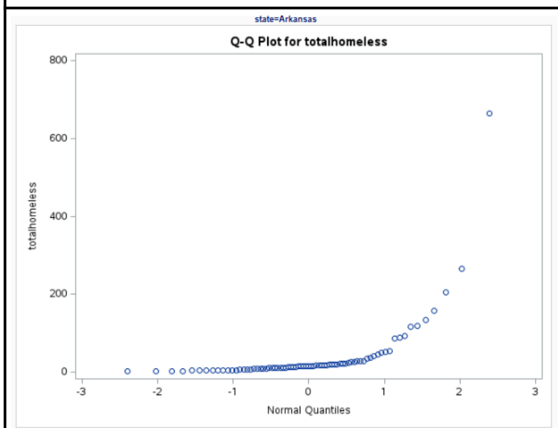
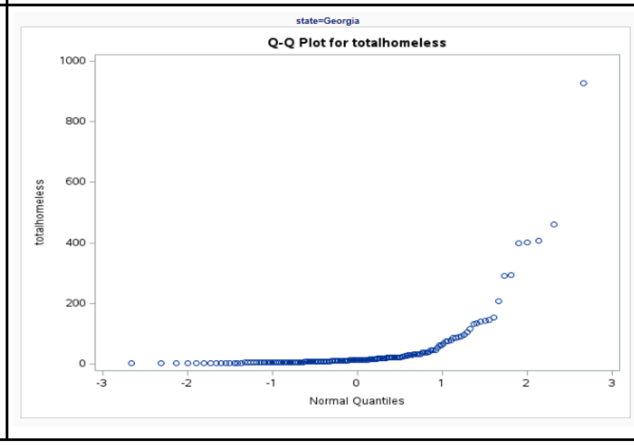
Quantiles (Definition 5)	
Level	Quantile
100% Max	59046
99%	22681
95%	3295
90%	1409
75% Q3	302
50% Median	75
25% Q1	18
10%	6
5%	3
1%	0
0% Min	0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0	677	26033	675
0	672	37430	118
0	671	37445	512
0	652	57103	135
0	635	59046	556

The data shows a high contrast in the distribution of both homelessness and evictions. For homelessness, the 0th percentile is 0, indicating no homelessness at the lower end, while the 100th percentile reaches 2,451, showing a significant upper range. This demonstrates a large disparity between the lowest and highest values in the homelessness data. Similarly, for evictions, the 0th percentile is 0, but the maximum value soars to 59,046, further highlighting the stark contrast between the low and high values in this variable. These large differences in the extremes suggest a high level of variability in both homelessness and evictions across the dataset.

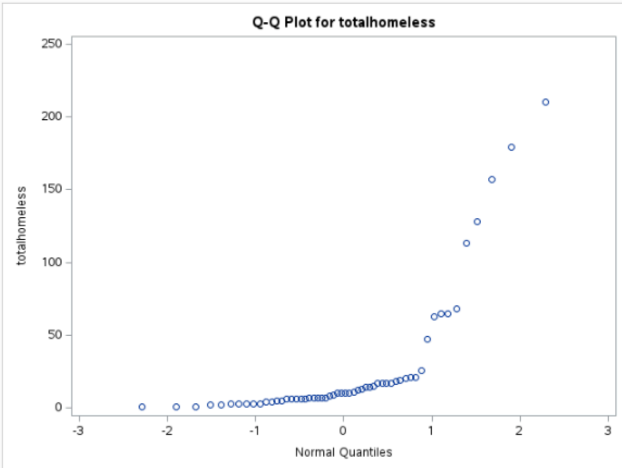
Testing Normality

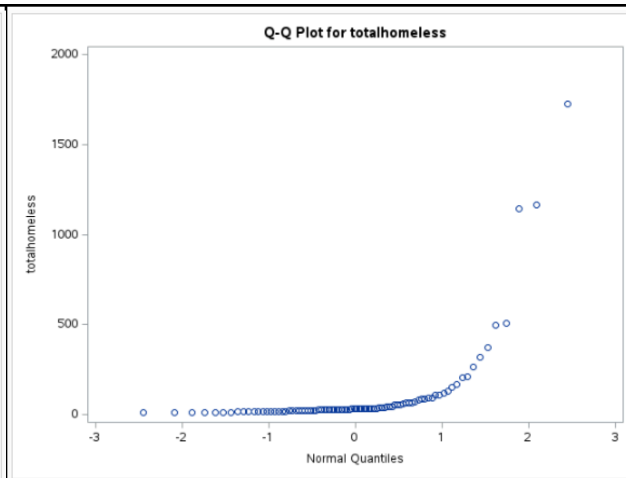
The normality test for the variable totalhomeless indicates that the data is not normally distributed across the states. The Shapiro-Wilk test produced a *p-value* less than 0.0001 in all cases, strongly rejecting the null hypothesis of normality. Additionally, the QQ plots further support this conclusion, as the points deviate significantly from the expected straight line, suggesting a non-normal distribution.

Arkansas	*Georgia*
	

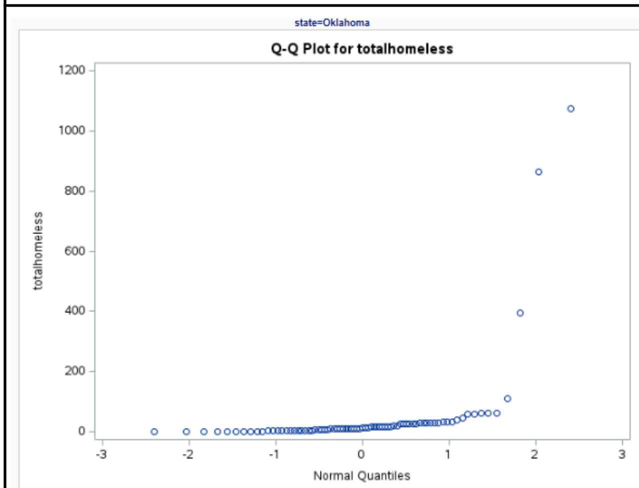
Montana	*Ohio*

Author: Murpys Duran Mendez

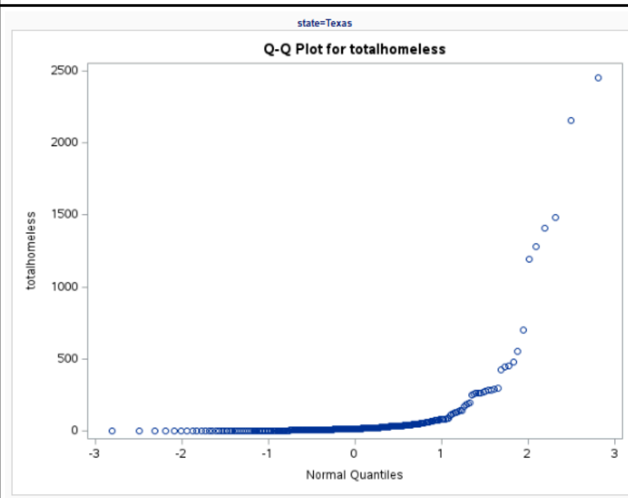




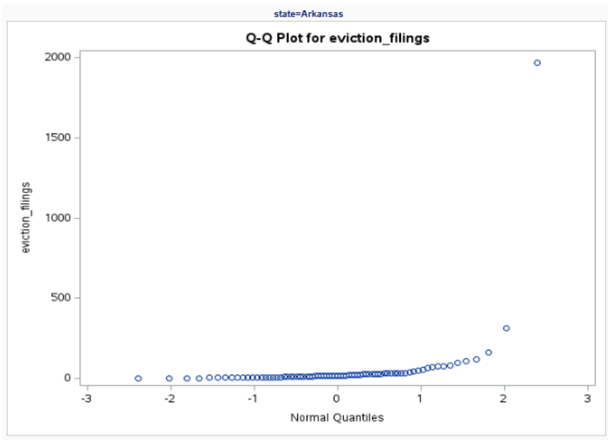
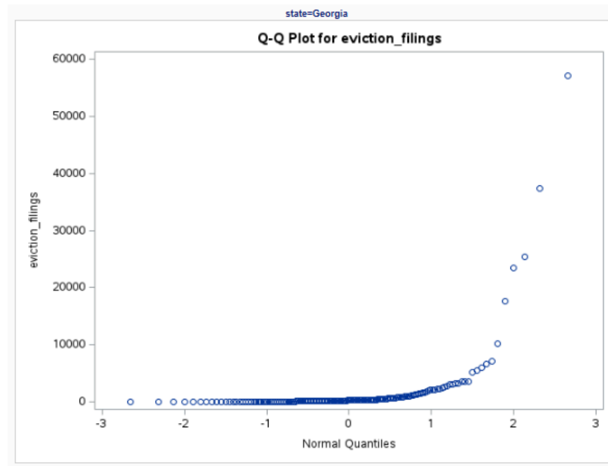
Oklahoma



Texas

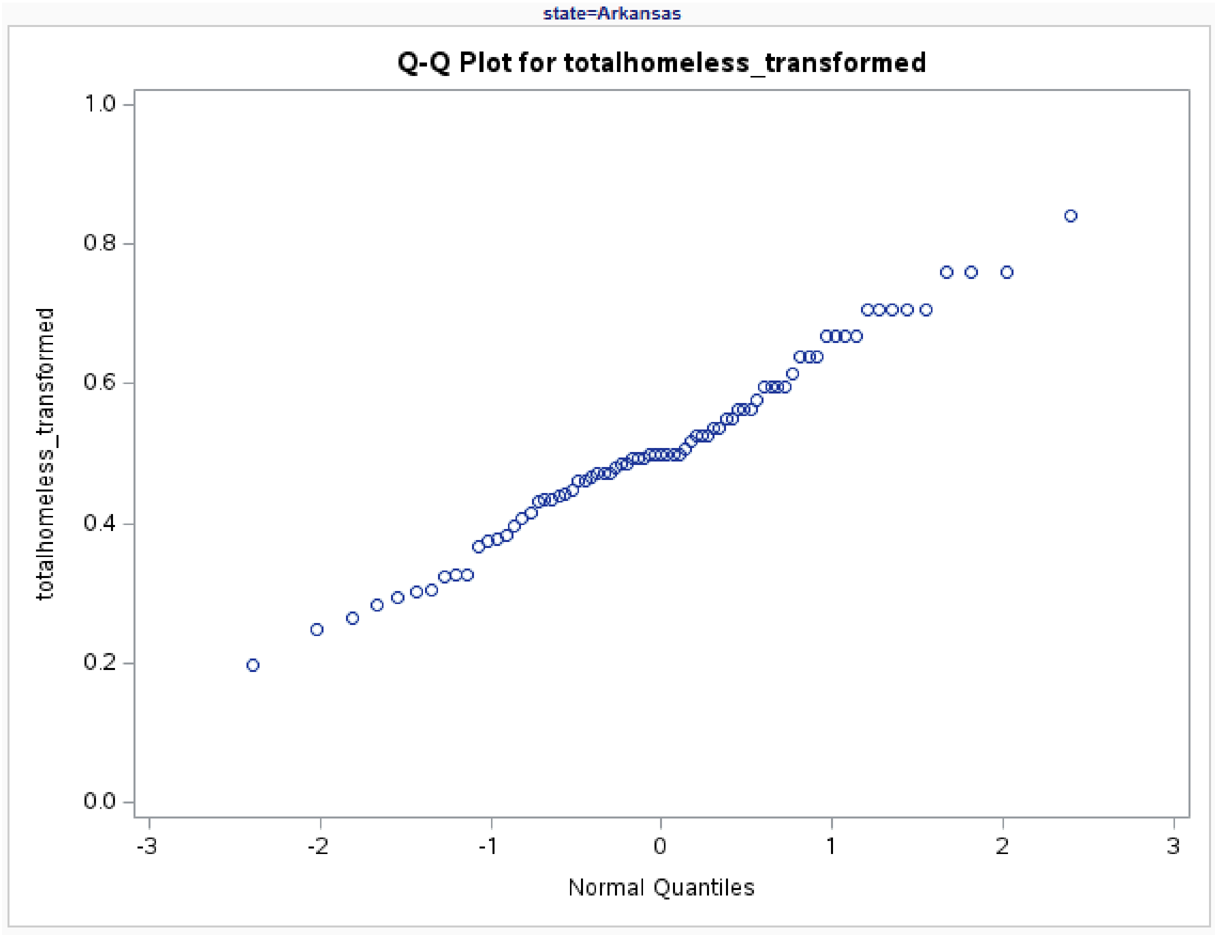


The normality test for the variable eviction_filings also indicates that the data is not normally distributed. The Shapiro-Wilk tests resulted in a *p-value* less than 0.0001 for all states, rejecting the null hypothesis of normality. The QQ plots show a clear deviation from the expected straight line, suggesting that the data does not follow a normal distribution. Additionally, the histograms are visibly skewed, further supporting the conclusion that the distribution of evictions is not normal. We present two states below as examples.

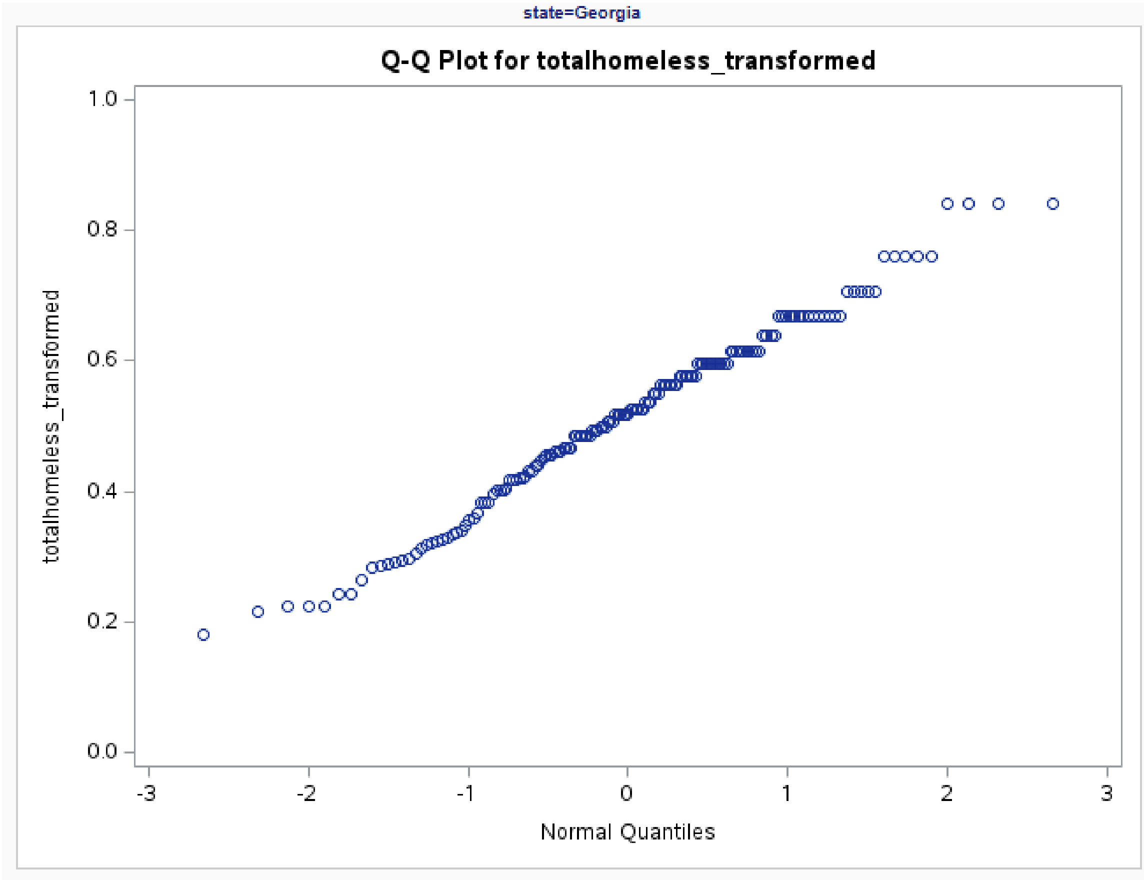
Arkansas	*Georgia*
	

After applying Box-Cox and other transformations, we obtained an approximation to the normal distribution, though the Shapiro-Wilks test still indicates non-normality for some states. This is likely caused by the presence of high values that skew the data.

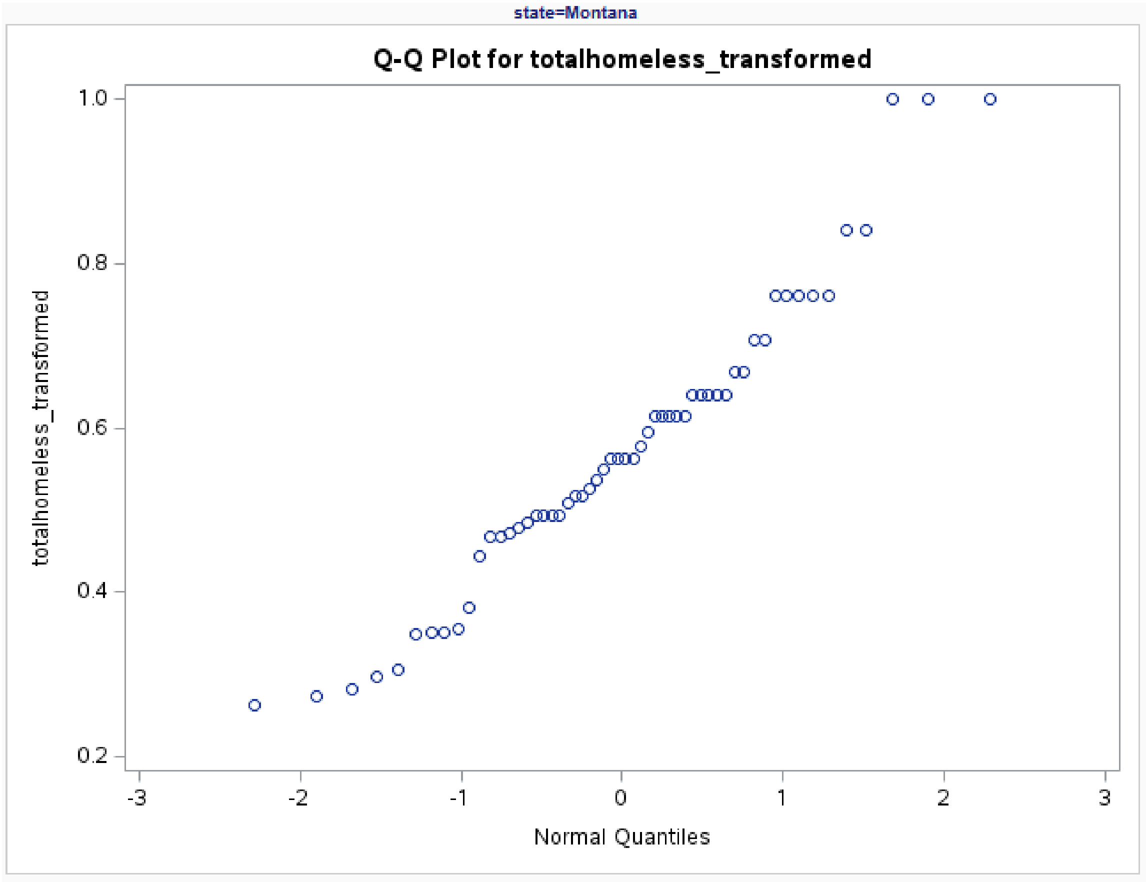
Author: Murpys Duran Mendez



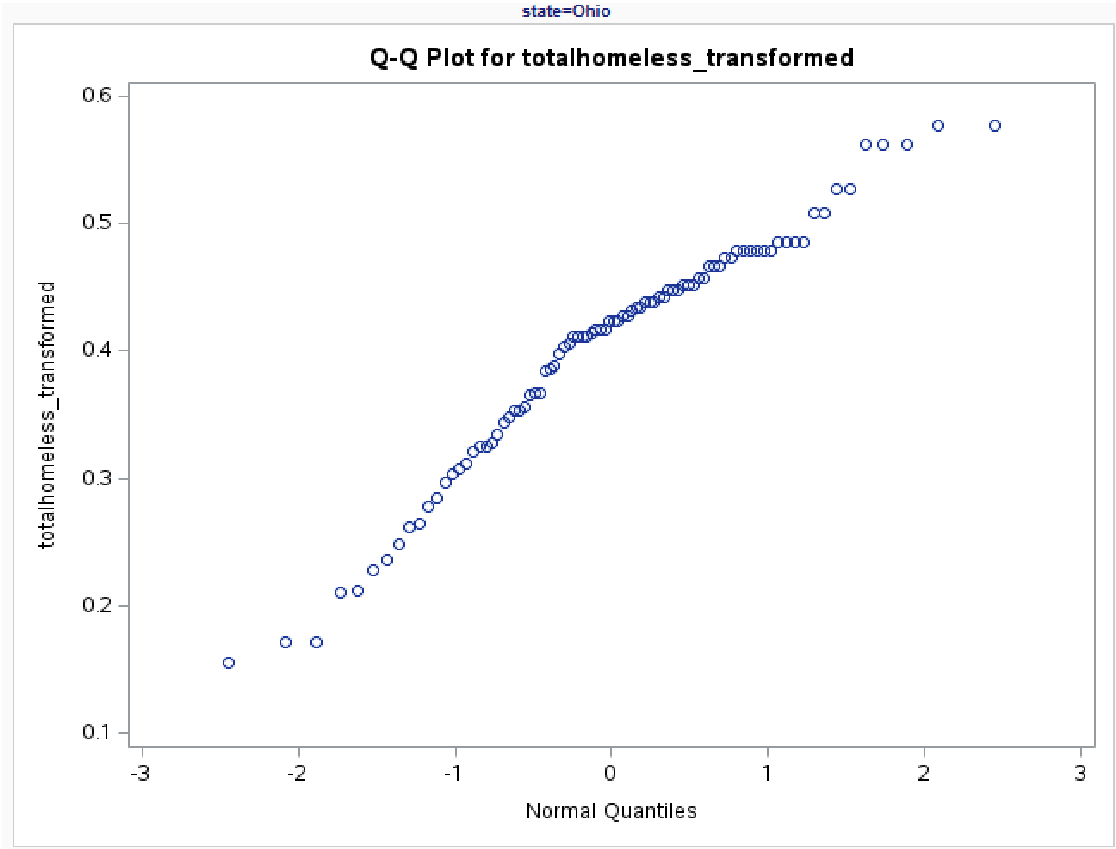
Author: Murpys Duran Mendez



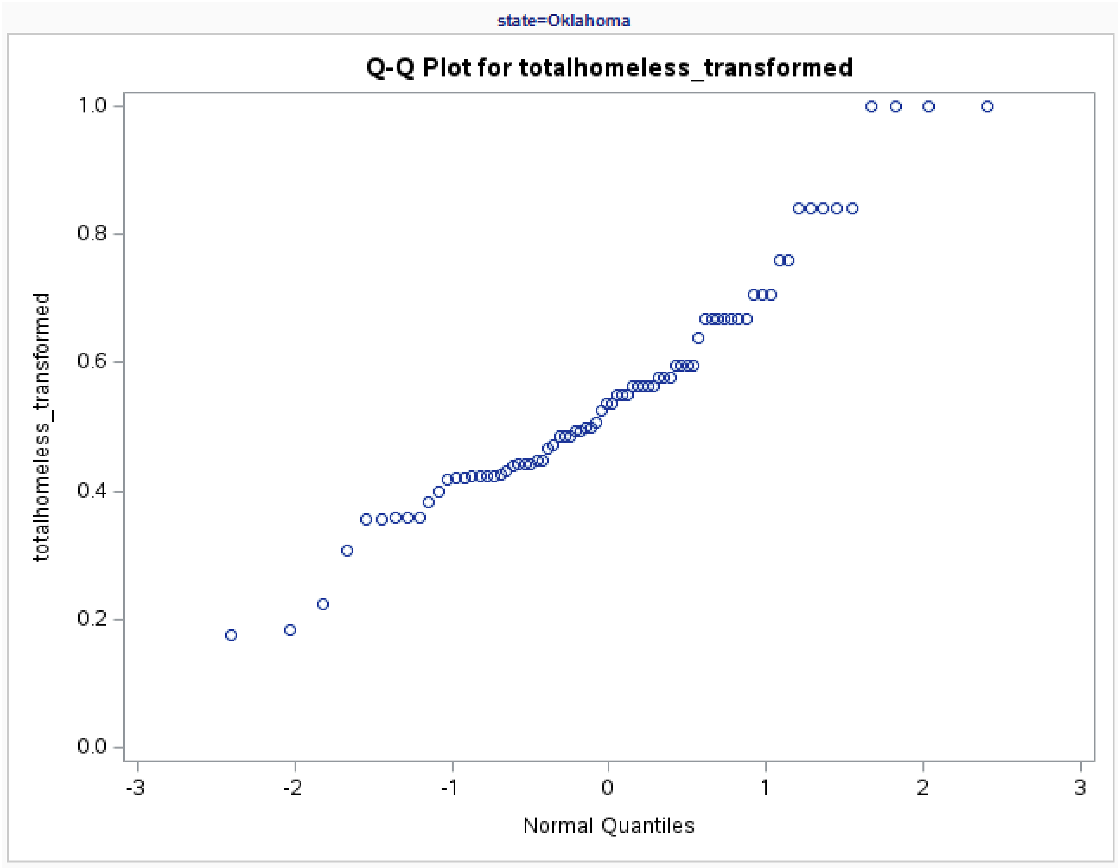
Author: Murpys Duran Mendez



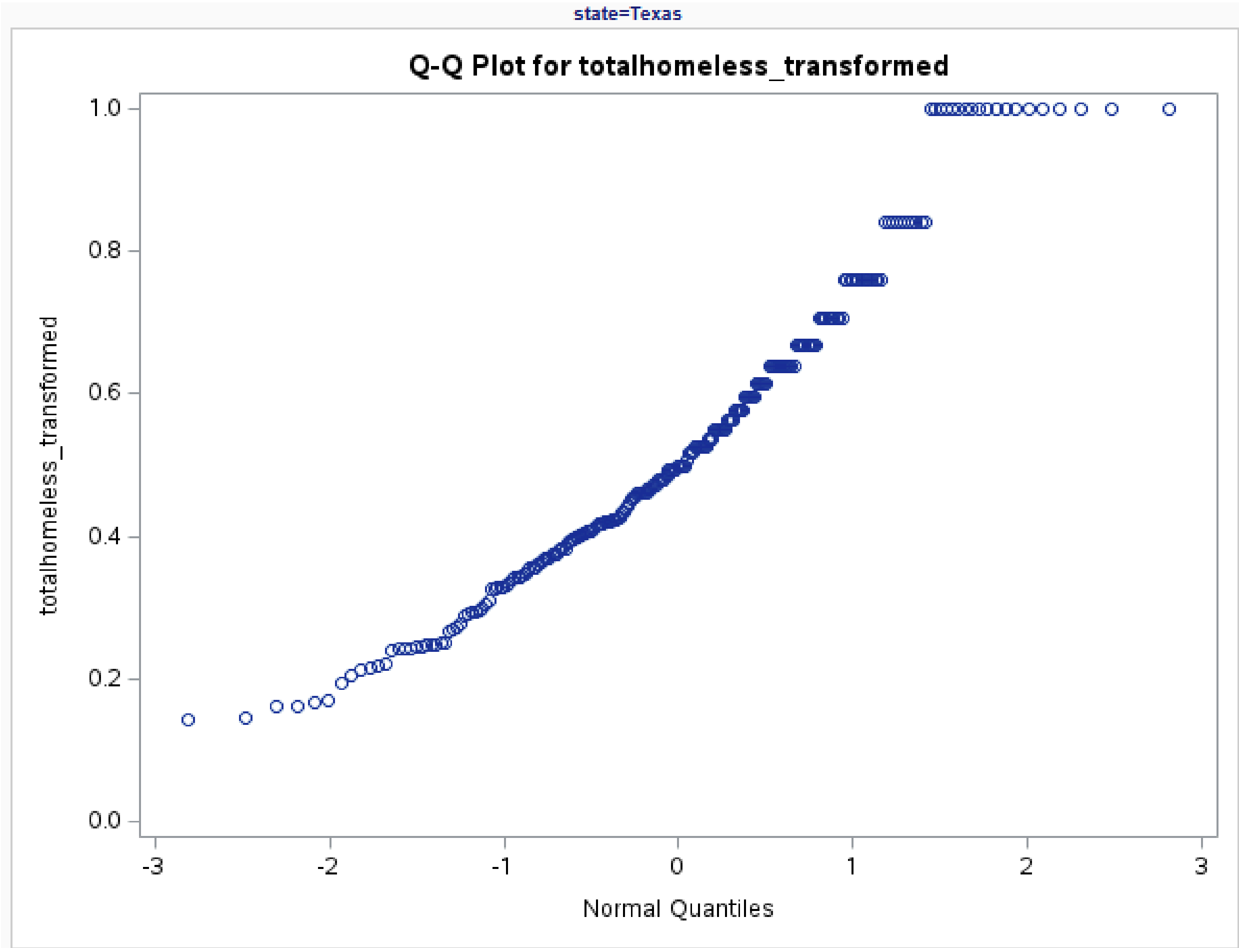
Author: *Murpys Duran Mendez*



Author: Murpys Duran Mendez



Author: Murpys Duran Mendez



Author: *Murpys Duran Mendez*

We conducted a **one-way ANOVA** to examine whether average homelessness levels differ significantly across U.S. states. The analysis revealed a statistically significant difference in mean totalhomeless values between states, $F(49, 698) = 10.09$, $p < .0001$, indicating that homelessness varies meaningfully by state. A post hoc Tukey test (α = 0.05) identified **Ohio** as significantly different from multiple states. Specifically, Ohio had significantly different mean homelessness levels compared to Montana, Oklahoma, Texas, Georgia, and Arkansas, as all these comparisons had 95% confidence intervals that excluded zero. This suggests Ohio's homelessness profile is notably distinct from those of several other states.

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
state	6	Arkansas Georgia Montana Ohio Oklahoma Texas

Number of Observations Read	709
Number of Observations Used	704

The ANOVA Procedure

Dependent Variable: totalhomeless_transformed

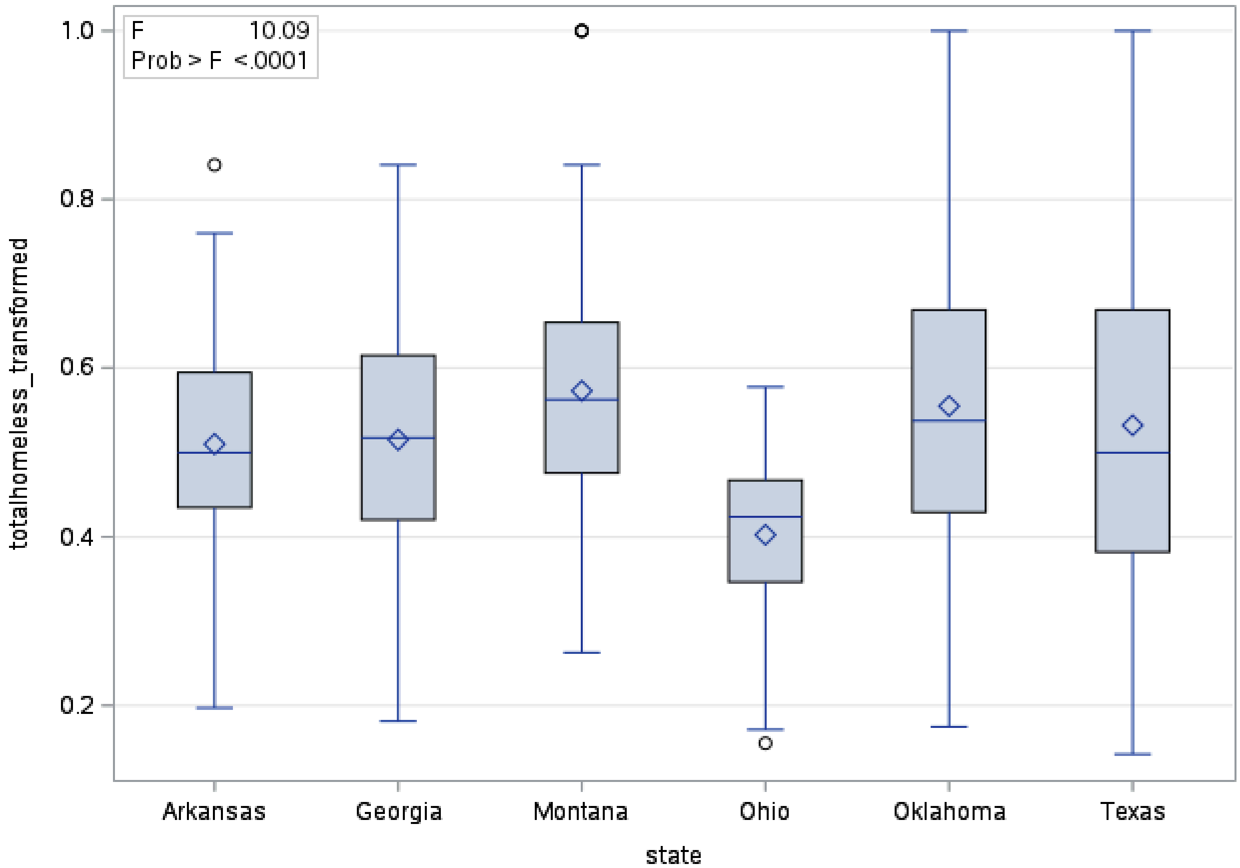
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1.49416721	0.29883344	10.09	<.0001
Error	698	20.66722530	0.02960921		
Corrected Total	703	22.16139252			

R-Square	Coeff Var	Root MSE	totalhomeless_transformed Mean
0.067422	33.39499	0.172073	0.515267

Source	DF	Anova SS	Mean Square	F Value	Pr > F
state	5	1.49416721	0.29883344	10.09	<.0001

Author: Murpys Duran Mendez

Distribution of Homelessness



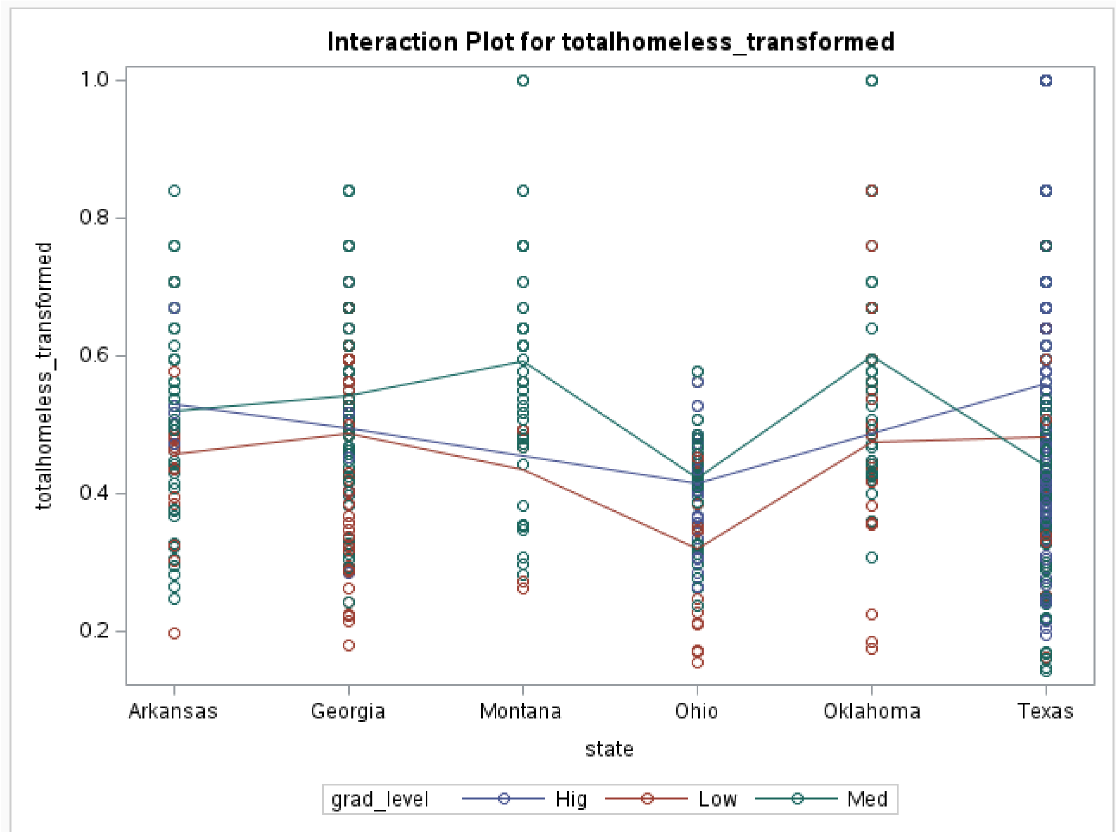
We used a MANOVA test to see if homelessness and evictions differ by **state**, **education level**, and their **interaction**. The results showed that both outcomes vary significantly between states (Wilks' Lambda = 0.693, $F = 27.08$, $p < .0001$). Education level also had a significant effect (Wilks' Lambda = 0.969, $F = 5.37$, $p = 0.0003$), meaning it's related to homelessness and evictions. Finally, the interaction between state and education level was also significant (Wilks' Lambda = 0.961, $F = 1.70$, $p = 0.0412$), showing that the effect of education level changes depending on the state.

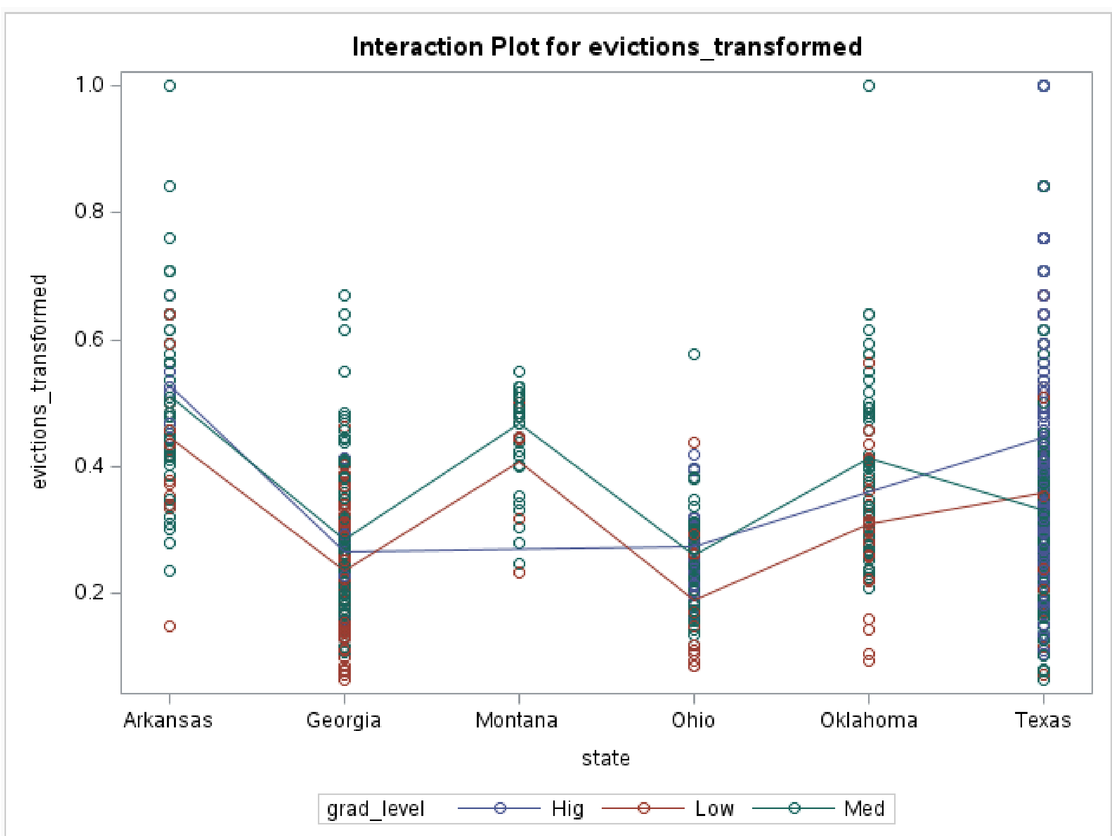
Author: *Murpys Duran Mendez*

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall grad_level Effect H = Type III SSCP Matrix for grad_level E = Error SSCP Matrix S=2 M=-0.5 N=335					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.96881369	5.37	4	1344	0.0003
Pillai's Trace	0.03130690	5.35	4	1346	0.0003
Hotelling-Lawley Trace	0.03206574	5.38	4	805.36	0.0003
Roy's Greatest Root	0.02754748	9.27	2	673	0.0001
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
NOTE: F Statistic for Wilks' Lambda is exact.					

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall state Effect H = Type III SSCP Matrix for state E = Error SSCP Matrix S=2 M=1 N=335					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.69276666	27.08	10	1344	<.0001
Pillai's Trace	0.32169940	25.80	10	1346	<.0001
Hotelling-Lawley Trace	0.42260590	28.37	10	1005.3	<.0001
Roy's Greatest Root	0.36546962	49.19	5	673	<.0001
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
NOTE: F Statistic for Wilks' Lambda is exact.					

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall state*grad_level Effect H = Type III SSCP Matrix for state*grad_level E = Error SSCP Matrix S=2 M=2.5 N=335					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.96077961	1.70	16	1344	0.0412
Pillai's Trace	0.03944281	1.69	16	1346	0.0421
Hotelling-Lawley Trace	0.04058993	1.70	16	1096.1	0.0406
Roy's Greatest Root	0.03372584	2.84	8	673	0.0042
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
NOTE: F Statistic for Wilks' Lambda is exact.					





Author: Murpys Duran Mendez

Interpretation of the Interaction Plots

In most states, medium education levels are linked to the highest homelessness rates, while high education is associated with medium homelessness, and low education tends to show the lowest homelessness rates. For evictions, a similar pattern is observed, with medium education showing higher evictions compared to high and low education. However, Montana stands out, as it follows a different trend, where high education is linked to lower evictions rates.

Correlation Matrix

We calculated the correlation matrix to examine how the variables in the dataset relate to one another. The matrix shows the strength and direction of linear associations between each pair of variables. Identifying strong or weak correlations helps determine which variables may provide similar information and which contribute distinct variation—useful when preparing for Principal Component Analysis (PCA).

Fragment of the Correlation Matrix

	fips	year	totalhomeless	county_population	eviction_filings	renting_household_population	z_med_rent	z_med_inc	pct_white	pct_african_american
fips	1.00000	. .	0.09442 0.0119	0.08689 0.0207	-0.01671 0.6569	0.07437 0.0478	0.17069 <.0001	0.29731 <.0001	-0.13718 0.0002	-0.52190 <.0001
year	. .	. .	. .	. .	. .	. .	. .	. .	. .	. .
totalhomeless	0.09442 0.0119	. .	1.00000	0.90623 <.0001	0.72184 <.0001	0.89855 <.0001	0.39302 <.0001	0.23931 <.0001	-0.17450 <.0001	0.08347 0.0262
county_population	0.08689 0.0207	. .	0.90623 <.0001	1.00000	0.84775 <.0001	0.98934 <.0001	0.37774 <.0001	0.23003 <.0001	-0.17406 <.0001	0.08333 0.0265
eviction_filings	-0.01671 0.6569	. .	0.72184 <.0001	0.84775 <.0001	1.00000	0.86568 <.0001	0.35650 <.0001	0.18260 <.0001	-0.19416 <.0001	0.19926 <.0001
renting_household_population	0.07437 0.0478	. .	0.89855 <.0001	0.98934 <.0001	0.86568 <.0001	1.00000	0.34154 <.0001	0.18866 <.0001	-0.17297 <.0001	0.09870 0.0085
z_med_rent	0.17069 <.0001	. .	0.39302 <.0001	0.37774 <.0001	0.35650 <.0001	0.34154 <.0001	1.00000	0.74217 <.0001	-0.07513 0.0455	-0.02867 0.4459
z_med_inc	0.29731 <.0001	. .	0.23931 <.0001	0.23003 <.0001	0.18260 <.0001	0.18866 <.0001	0.74217 <.0001	1.00000	0.18732 <.0001	-0.30501 <.0001
pct_white	-0.13718 0.0002	. .	-0.17450 <.0001	-0.17406 <.0001	-0.19416 <.0001	-0.17297 <.0001	-0.07513 0.0455	0.18732 <.0001	1.00000	-0.41131 <.0001
pct_african_american	-0.52190 <.0001	. .	0.08347 0.0262	0.08333 0.0265	0.19926 <.0001	0.09870 0.0085	-0.02867 0.4459	-0.30501 <.0001	-0.41131 <.0001	1.00000
pct_latinx	0.50097 <.0001	. .	0.09362 0.0126	0.09441 0.0119	0.03193 0.3960	0.08348 0.0262	0.10096 0.0071	0.02496 0.5069	-0.71607 <.0001	-0.24872 <.0001
primary_care_physician_rate	-0.08555 0.0227	. .	0.28976 <.0001	0.24169 <.0001	0.22840 <.0001	0.24060 <.0001	0.26590 <.0001	0.15333 <.0001	0.03034 0.4198	0.08300 0.0271
pct_single_parent_households	-0.28585 <.0001	. .	0.03082 0.4126	0.02436 0.5172	0.07496 0.0460	0.04452 0.2364	-0.23226 <.0001	-0.48897 <.0001	-0.38335 <.0001	0.59554 <.0001
pct_smokers	-0.39195 <.0001	. .	-0.14416 0.0001	-0.13309 0.0004	-0.10361 0.0058	-0.11533 0.0021	-0.48119 <.0001	-0.56160 <.0001	-0.02154 0.5669	0.29595 <.0001
pct_obese	-0.37797 <.0001	. .	-0.13016 0.0005	-0.11387 0.0024	-0.11700 0.0018	-0.11936 0.0015	-0.28316 <.0001	-0.29738 <.0001	0.02731 0.4678	0.27677 <.0001
pct_unemployed	-0.41152 <.0001	. .	-0.11122 0.0030	-0.08415 0.0250	-0.03002 0.4248	-0.08108 0.0309	-0.32135 <.0001	-0.55525 <.0001	-0.25713 <.0001	0.49522 <.0001
pct_high_school_graduation	0.37277 <.0001	. .	-0.13816 0.0002	-0.12292 0.0010	-0.19669 <.0001	-0.13473 0.0003	0.01428 0.7042	0.18073 <.0001	0.09856 0.0086	-0.32534 <.0001

Principal Component Analysis

We ran a Principal Component Analysis (PCA) using six variables: *total homeless population*, *county population*, *population per square mile*, *percent in poverty*, *Zillow median rent*, and *percent unemployed*. These variables were chosen because they might be related and reflect conditions linked to homelessness. The aim was to reduce the number of variables while keeping most of the information, and to see if a few components could summarize the main patterns in the data.

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix					
	1	2	3	4	5
1	0.82711	0.32215	-0.16116	0.38939	0.18578
2	0.47751	-0.51472	0.59520	-0.38188	0.08337
3	-0.19979	0.49175	0.78366	0.31918	0.04761
4	0.14226	0.62074	-0.03684	-0.75223	-0.16503
5	0.16650	-0.06442	0.06543	0.18659	-0.96386

Rotated Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
totalhomeless	0.96194	0.10684	-0.05203	0.13826	-0.00555
county_population	0.96041	0.09285	-0.02732	0.12218	0.05691
pop_per_sq_mile	0.79508	0.10999	0.01299	0.24750	0.54186
pct_poverty	0.13573	0.94368	-0.14241	0.26368	0.03505
z_med_rent	0.23799	0.30371	-0.16257	0.90491	0.07598
pct_unemployed	-0.03047	-0.12790	0.98306	-0.12765	0.00124

Variance Explained by Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
2.5558473	1.0312713	1.0167426	0.9999848	0.3038881

The first factor has high loadings on homelessness, county_population, and population per square mile. This factor could indicate population density. The second factor reflects, significantly, only the “poverty” variable, without an important contribution from the rest of the variables. This variable would possibly be better situated in another variable set.

Principal Component 2.

We ran a second Principal Component Analysis using a new set of variables: *percent smokers*, *percent single-parent households*, *primary care physician rate*, *percent obese*, *percent unemployed*, *violent crime rate*, *percent in poverty*, and *percent voting Republican*. The first component showed strong loadings for *percent voting Republican* (−0.80) and *percent unemployed* (0.80), indicating these variables contribute most to this dimension. Additionally, *percent single-parent households* had a moderate loading (0.59), suggesting it also plays a meaningful, though smaller, role in shaping the component. This PC could be cataloged as "**Economic Hardship**".

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix					
	1	2	3	4	5
1	0.63971	0.60821	0.36542	-0.29467	0.02234
2	0.29977	-0.39569	0.44589	0.43295	0.60606
3	-0.59790	0.61234	0.14168	0.17682	0.46498
4	-0.14447	0.07593	0.56253	0.49186	-0.64420
5	0.35008	0.30461	-0.57545	0.67270	-0.03147

Rotated Factor Pattern					
	Factor1	Factor2	Factor3	Factor4	Factor5
pct_smokers	0.30813	0.80588	0.11421	-0.23112	0.00500
pct_single_parent_households	0.58668	0.28702	0.46919	-0.19189	0.01223
primary_care_physician_rate	0.05223	-0.02525	0.06956	0.05701	0.97892
pct_obese	0.05143	0.93053	0.08243	-0.04925	-0.03838
pct_unemployed	0.79973	0.29523	-0.02958	-0.19409	-0.17174
violent_crime_rate	0.14516	0.09628	0.94899	0.02877	0.07838
pct_poverty	-0.08943	-0.17488	-0.00189	0.96559	0.06278
republican_voting_pct	-0.80482	0.01058	-0.21650	-0.09045	-0.25639

Variance Explained by Each Factor				
Factor1	Factor2	Factor3	Factor4	Factor5
1.7608922	1.7254618	1.1931518	1.0749479	1.0652583

Factor Analysis

We conducted a factor analysis using the following variables: *fips*, *total homeless population*, *county population*, *eviction filings*, *renting household population*, *percent* *white*, *percent African American*, and *percent Latinx*. Based on the analysis, four factors were extracted. The first factor explains the largest share of the variance (3.63), followed by the second (1.72), and the third (1.31). The fourth factor accounts for a very small portion of the variance (0.08), suggesting it may not provide meaningful insight. Most of the structure in the data appears to be captured by the first three factors.

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2	3	4
1	0.97784	-0.20713	0.03036	-0.00232
2	-0.15632	-0.81879	-0.55227	-0.01209
3	0.13925	0.53541	-0.83295	-0.01217
4	0.00207	-0.00387	-0.01674	0.99985

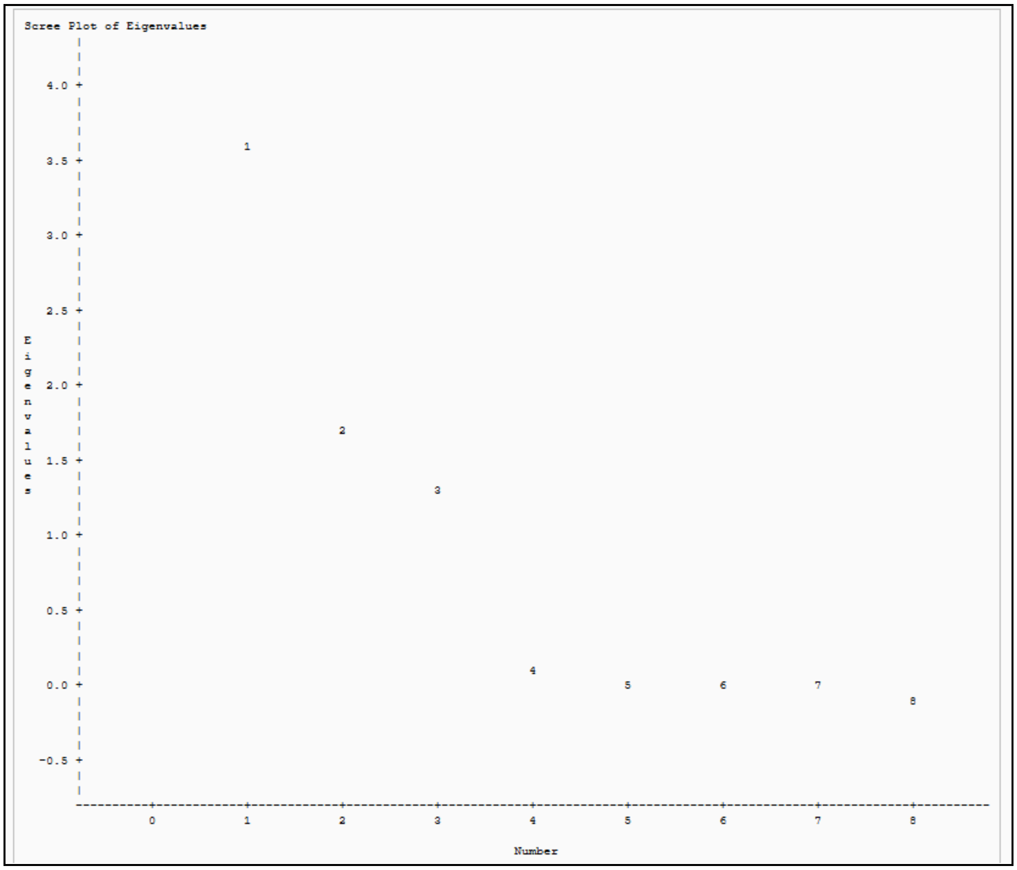
Rotated Factor Pattern

	Factor1	Factor2	Factor3	Factor4
fips	0.05455	-0.25916	-0.62752	-0.02726
totalhomeless	0.89543	-0.07144	-0.02692	-0.18623
county_population	0.99230	-0.06152	-0.02762	-0.02898
eviction_filings	0.85389	-0.07048	0.11983	0.19766
renting_household_population	0.99514	-0.05737	-0.00944	0.02392
pct_white	-0.12356	0.93148	-0.19918	0.01116
pct_african_american	0.09646	-0.20449	0.88603	-0.01693
pct_latinx	0.03024	-0.83118	-0.45297	0.01301

Variance Explained by Each Factor

Factor1	Factor2	Factor3	Factor4
3.5343329	1.6846399	1.4396112	0.0764864

The Scree Plot of the Eigenvalues shows the “elbow” forming after the first three factors.



Factor 1: Population and Housing Density

Strong loadings:

- county_population (0.99)
- renting_household_population (0.99)
- totalhomeless (0.90)
- eviction_filings (0.85)

Interpretation: This factor explains variance related to **population size and housing issues** — it reflects counties with large populations, high numbers of renters, homelessness, and evictions.

Author: Murpys Duran Mendez

Factor 2: Racial Composition (White vs. Latinx)

Strong loadings:

- `pct_white` (0.93)
- `pct_latinx` (−0.83)

Interpretation: This factor distinguishes areas by **racial/ethnic composition**, especially contrasting percent white and percent Latinx populations. A higher factor score may indicate a higher proportion of white residents and lower Latinx presence.

Factor 3: African American Population

Strong loading:

- `pct_african_american` (0.89)

Moderate negative:

- `pct_latinx` (−0.45)

Interpretation: This factor seems to capture variation in the **African American population**, distinct from the Latinx proportion. Since no other variables load highly here, this may be a more isolated demographic dimension.

Conlusion

Throughout this project, we explored different statistical methods to better understand the patterns and differences across U.S. counties in terms of social, economic, and demographic characteristics. Each technique added a layer of insight: we examined group differences, relationships among variables, and underlying structures that aren't immediately visible in raw data. Taken together, the analyses helped highlight how certain counties and states stand out, how variables cluster or relate, and which dimensions matter most in shaping local conditions.