

Updating Literacy Estimates in the Oklahoma City Metro

Olivia Campos, Alkin Huggins, Nelly Sime

Advisor: Tracy L. Morris

SCHOLAR, Department of Mathematics and Statistics, University of Central Oklahoma



Introduction

- Project SCHOLAR¹ (Statistical Consulting Help for Organizational Leaders and Academic Researchers) is an undergraduate student statistical consulting service at UCO. SCHOLAR students work under the supervision of faculty mentors from the Department of Mathematics and Statistics on projects submitted by researchers on campus as well as off campus.
- This year's group was asked by the Oklahoma City (OKC) Metro Literacy Coalition² to update estimates of literacy levels in the OKC metro area.
- Demographic variables were collected from the 2000 US Census³ for all counties in KY, MD, MA, MO, NY, and OK, and were used to develop a model for predicting the 2003 State Assessment of Adult Literacy (SAAL)⁴ estimates.
- This model was applied to demographic data collected from the 2015 American Community Survey³ for every zip code in the OKC metro area, to obtain more recent estimates of below basic proficiency in literacy.

Methods

The statistical analysis was completed in three stages.

1. Univariate analysis:

- Demographic variables from the 2000 US Census³ for each county in KY, MD, MA, MO, NY, and OK, were checked one at a time to determine which variables are significantly related to the 2003 SAAL⁴ estimates.
- This was done by looking at the p-value, residual plot, and QQ-plot for each demographic variable.
- Several variables, including the response variable, were log-transformed to achieve linearity.
- Demographic variables with a p-value less than 0.25 were considered for inclusion in the multiple regression model.
- The univariate analysis was done using proc reg in SAS v. 9.4.

2. Multiple linear regression:

- Before performing the multiple regression, the remaining demographic variables were checked for collinearity.
- Multiple linear regression with stepwise selection was performed on the group of variables remaining after the check for collinearity. Summary statistics for the variables remaining after stepwise selection are presented in Table 1.
- Interaction and quadratic terms were assessed for inclusion in the model. Diagnostic plots and parameter estimates for the final model are presented in Figure 1 and Table 3, respectively.

- The final model was developed using proc reg in SAS v. 9.4.

3. Prediction:

- The model was applied to demographic variables from the 2015 American Community Survey³ for each zip code in the OKC metro area. Summary statistics for these variables are shown in Table 2.
- Heat maps for the resulting estimates were constructed for OK county and the OKC metro area. These are displayed in Figures 2 and 3, respectively
- The choroplethr package⁵ in R v. 3.4.1 was used to draw the heat maps.

Results

Table 1. Summary statistics (%) for the 2000 US Census data for states included in the 2003 SAAL (n=412)

Variable	Mean	Std Dev	Minimum	Maximum
Black	4.8	7.5	0.0	64.3
Hispanic	2.7	4.3	0.3	48.4
HS Graduate	75.4	9.0	49.2	93.1
Unemployment	3.4	1.1	0.8	8.2
Urban	38.3	30.1	0.0	100.0
Literacy	12.1	4.4	4.0	46.0

Table 2. Summary statistics (%) for the 2015 American Community Survey data for all zip codes in the OKC metro area (n=96)

Variable	Mean	Std Dev	Minimum	Maximum
Black	10.1	16.5	0.0	80.4
Hispanic	9.5	11.7	0.0	59.0
HS Graduate	87.0	8.6	52.0	100.0
Unemployment	3.5	1.9	0.0	9.4
Urban	57.5	43.8	0.0	100.0
Literacy	13.2	9.5	5.0	58.7

Figure 1. Diagnostic plots for the final model

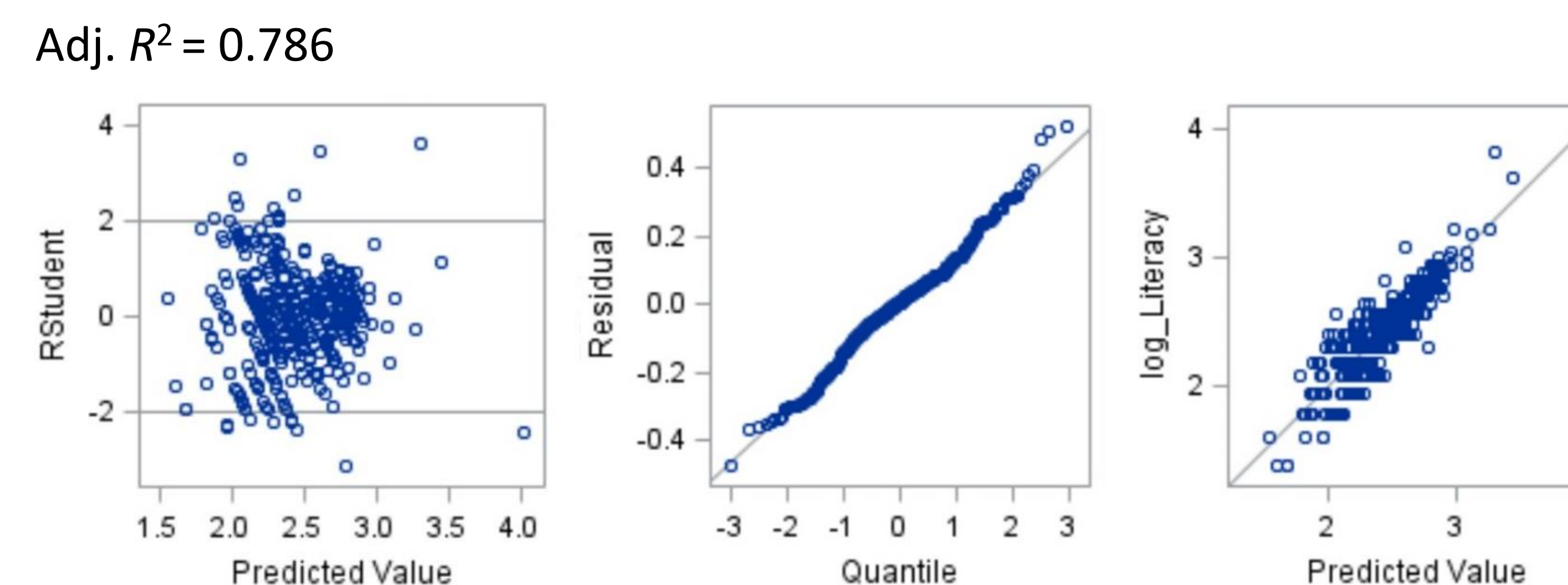


Table 3. Parameter estimates for the final model

Variable	Estimate	Std Err of Estimate	Test Statistic	p-value
Intercept	4.313	0.290	14.86	<0.001
ln(Hispanic)	-0.032	0.058	-0.56	0.578
ln(Black)	0.026	0.031	0.86	0.393
HS Graduate	-0.029	0.004	-7.67	<0.001
Unemployment	-0.303	0.067	-4.54	<0.001
Urban	0.011	0.003	4.13	<0.001
South	0.379	0.040	9.50	<0.001
ln(Hispanic)*Urban	0.0017	0.0004	3.70	<0.001
ln(Hispanic)*South	-0.083	0.032	-2.59	0.010
ln(Black)*South	-0.057	0.020	-2.84	0.005
HS Grad*Unemp	0.0046	0.0009	5.23	<0.001
HS Grad*Urban	-0.0002	0.000003	-5.80	<0.001
(ln(Hispanic)) ²	0.057	0.018	3.10	0.002
(ln(Black)) ²	0.018	0.009	2.03	0.043

Results

Figure 2. 2015 estimates of below basic proficiency in literacy for OK county

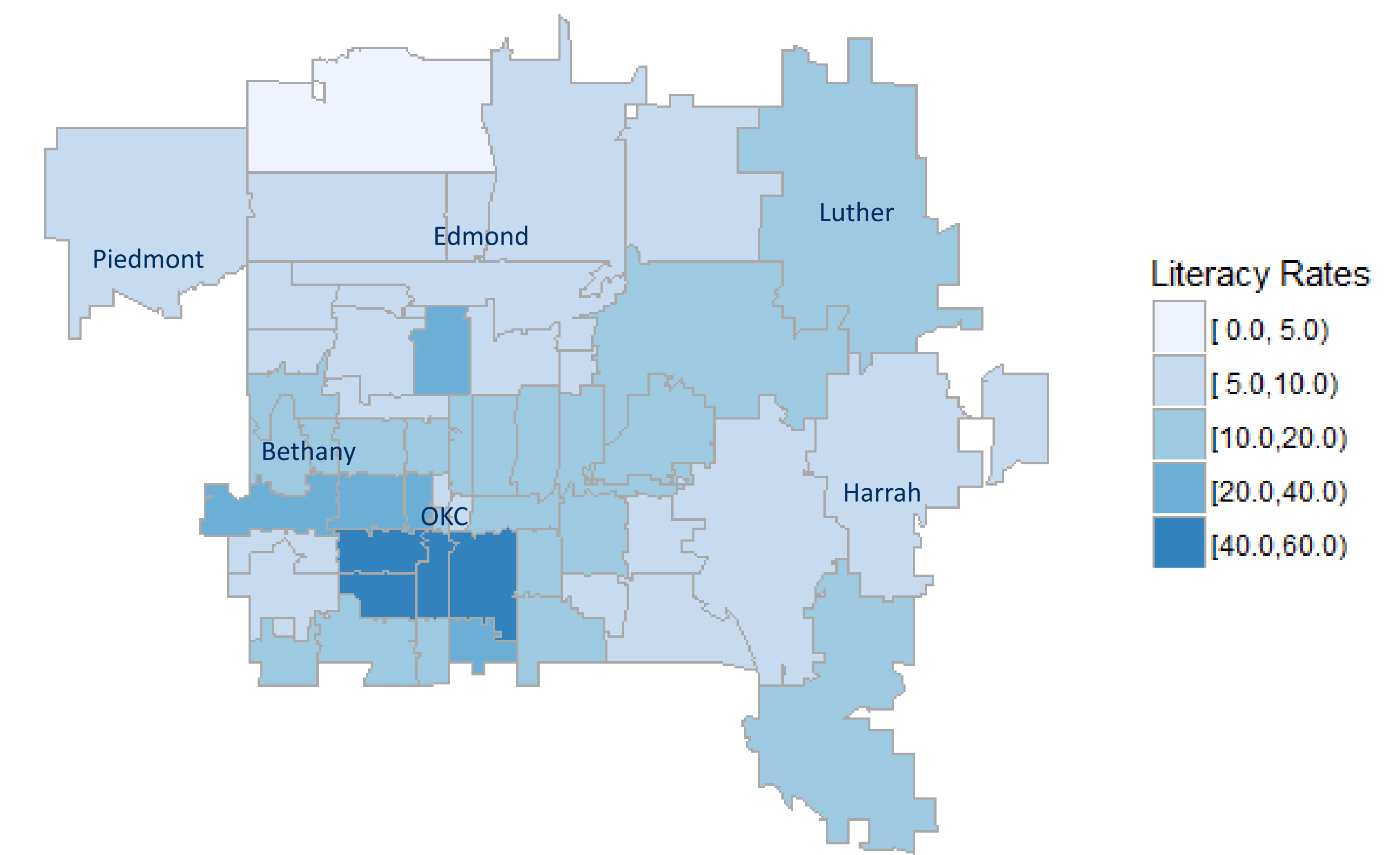
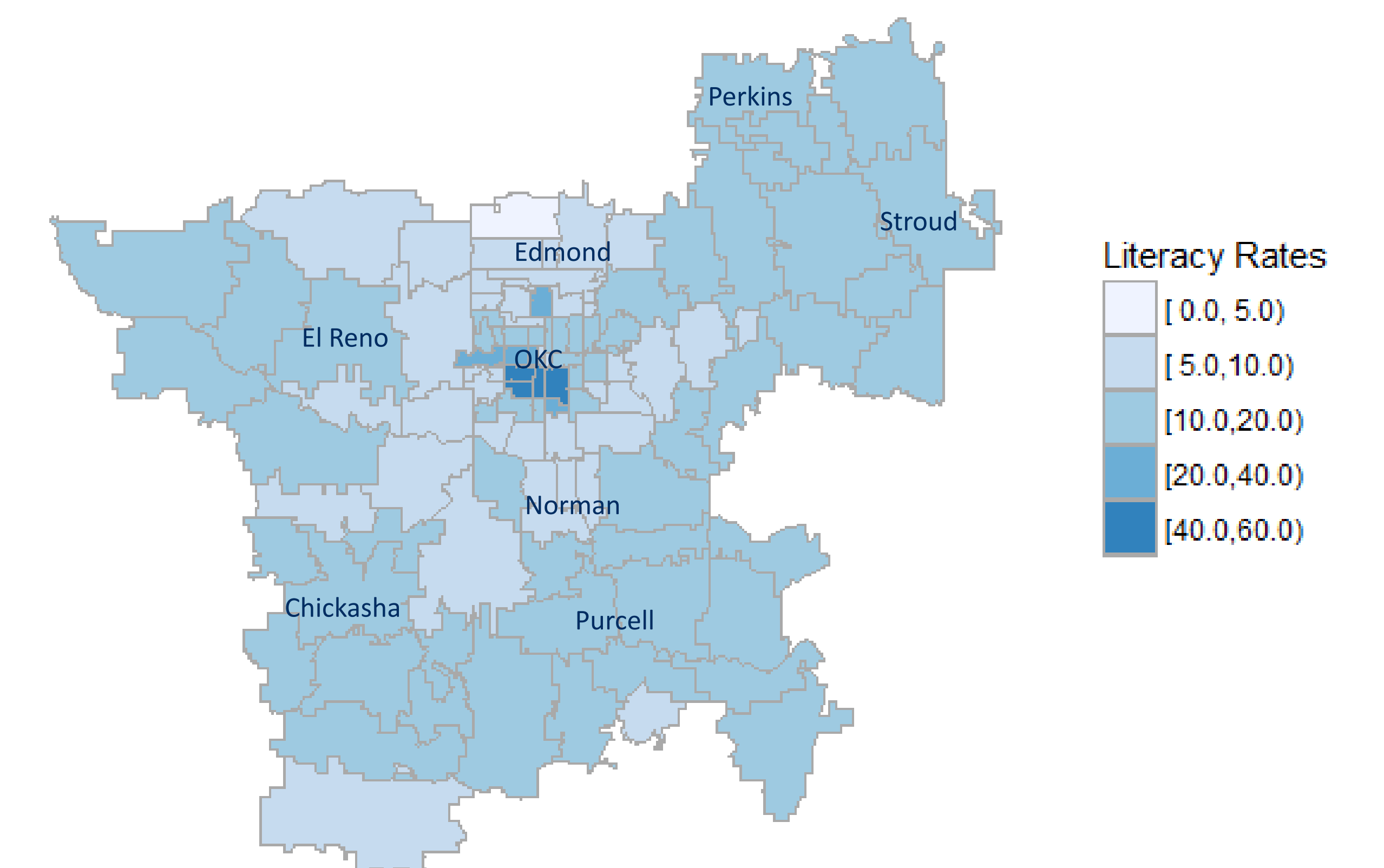


Figure 3. 2015 estimates of below basic proficiency in literacy for the OKC metropolitan area



Resources

- <http://www.math.uco.edu/resources/scholar/index.html>
- <https://okcliteracycoalition.org/>
- <https://factfinder.census.gov/>
- <https://nces.ed.gov/naal/saal.asp>
- <https://cran.r-project.org/web/packages/choroplethr/index.html>

Acknowledgements

- Partial support for this work was provided by the NSF's Improving Undergraduate STEM Education (IUSE) program under Award No. 1611732. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.
- We would like to thank Julie Serven and the OKC Metro Literacy Coalition for bringing this project to our attention.