Analaysis of Public Housing in County Census Tracts and Proximity to CE(s)

Original Analysis
By: D. White
Date: 2019-11-13

KNITR File: Report_Logistic_Model.Rnw

```
## [1] "Generated on: 2020-04-23 16:49:58"
```

Methods: A multi-step data processing routine was performed for each county to integrate spatial and tabular data sets. Census tract data were spatially joined with CEs, Section-8 voucher housing units, and Low-Income Housing Tax Credit (LIHTC) units.

Processing Steps:
HUD Housing Data
1. Section-8 Voucher Data downloaded and aggregated from HUD website for each county (2000-2008)(see Assisted_Housing_Data.csv)
Section-8 Voucher data are spatially located by census tract. Primary data attributes included a census tract "CODE", "YEAR", "NUMBER_REPORTED", and others. Census tract data were a multi-year report for each census tract in the county (2000, 2004, and 2008). Each county was saved as a separate CSV file (COUNTY_ABREV_HUD.CSV).
2. LIHTC Data downloaded and aggregated from HUD website for each county (2000-2008) (The Low-Income Housing Tax Credit Affordable housing data.csv) LIHTC data are spatially located by address. Primary data attributes included a census tract "HUD ID Number", "Year", "Project Name", "Project Address", "Total Low-Income Units", "Total Number of Units", and others. Data files were modified for GIS processing (attribute labels cleaned up, shortned, characters removed, etc.) and added to a GIS.
LIHTC data were geocoded using the "ArcGIS World Geocoding Service". A total of 238 nationwide LIHTC addresses were geocoded. Except for two locations, 236 addresses were positively matched. A separate analysis was performed on the two other locations that were ties. The locations were verified and a final shapefile was created. Using the Select by Attribute tool in ArcGIS, each county was exported as a separate shapefile (County_Tax_Credit.shp) from the naitonwide data set.

County Census Tracts:
County census tract data were downloaded from the IPUMS/NHGIS (IPUMS.org) for each county from the 2000 Census (see folder "nghis0045_csv"). A data merge was performed in R using the spatial package (sp:::merge) between Section-8 CSV file and the Census Tract Shapefile for each county to generate a aggregated working file using a common census tract code as unique identifier ("GISJoin"). The resulting file in R that was filtered for the maximum available year that was 2008 for each county (Figure 1).

HUD-Aggregated Tables:
Section-8 and LIHTC were aggregated into a common spatial data set. The geocoded LIHTC county shapefile was imported into R and duplicate check was run to remove potential replicate LIHTC HUD IDs. Data were aggregated to the final observed year for each county, unlike Section-8, LIHTC data were summed across time to reflect the final year of available LIHTC units for each county. Section-8 and LIHTC were combined using a spatial join (sf:::st_join) in R. The resulting spatial file realized census tract data with a total number of Section-8 housing units and multiple observations of LIHTC units per census tract. Thus, a final data set needed to collapse reported LIHTC into a single observation for each census tract. Further

processing generated "clean" unique Section-8 and LIHTC summed by county census tract (Figure 2 and 3, repspectively).

The data processing by county can be found in the path below.
C:\Users\whitedl\Documents\R_Code\HUD_Project_Code\CNTYNAME_Public_Housing_CE.R

Results
Descriptive Statistics:
Total census tracts per county are presented in Figure 1. With the exception of SAC, all counties had fewer than 100 census tracts. Douglas county had the group minimum of 8. Sacramento had a group maximum of 279. Explanatory variables had high zero counts with respect to prescence per census tract (Figure 4 and 5). Both groups show a large number of zeros, but LIHTC units were generally found in fewer census tracts. Section 8 units were more abundant than LIHTC units (Figures 6 and 7). However, LIHTC appear to have a higher density in census tracts (Figure 8). Sacramento had the largest total number of Section 8 and LIHTC units. Douglas county had the fewest Section 8 units and zero LIHTC units.

Model:
A poisson logistic regression was performed on an all county data set to test the probability of a CE located in census tract dependent upon HUD housing counts (Tables 1-4). Table 5 presents the results of a zero-inflation model, however, an overdispersion test indicates that overdispersion was not present in the GLM-Poisson model (Table 4), and geneally not an issue for the county models. Tables 6-36 show results of the logistic-poisson model for each individual county. ALB and DGL were included in the county global model, but were not run as discrete county models due to small sample sizes.

Parameter coefficient estimates for the global model (all counties) (CE presence/abscence = LIHTC + Sec 8) were significant (Table 1). A positive coefficent will indicate that an increase in public housing is associated with an increase in the probability of a CE in a given Census Tract. A negative coefficent will imply that increased numbers of public housing units are associated with a decreased probability of a CE in a census tract. Section 8 units had a signficant negative effect (-0.007)($P<.01$) and LIHTC units had a signficant positive effect (0.002)($P<.05$). A significant negative constant was reported (-0.738)($P<.01$). Overdispersion was not indicted (Table 4).

Sacramento and York had significant coefficient estimates for Section 8 and LIHTC units. For Sacramento, Section 8 units had a signficant negative effect (-0.016)($P<.05$) and LIHTC units had a signficant postitive effect (0.003)($P<.01$). A significant negative constant was reported (-1.970)($P<.01$) (Table 30). Overdispersion was not indicted (Table 33). For York county, Section 8 units had a signficant negative effect (-0.016)($P<.05$) and LIHTC units had a signficant postitive effect (0.010)($P<.1$) (Table 42). A significant negative constant was reported (-0.401)($P<.1$). Overdispersion was not indicted (Table 44). For all other counties parameter coefficients were found to not be significant.

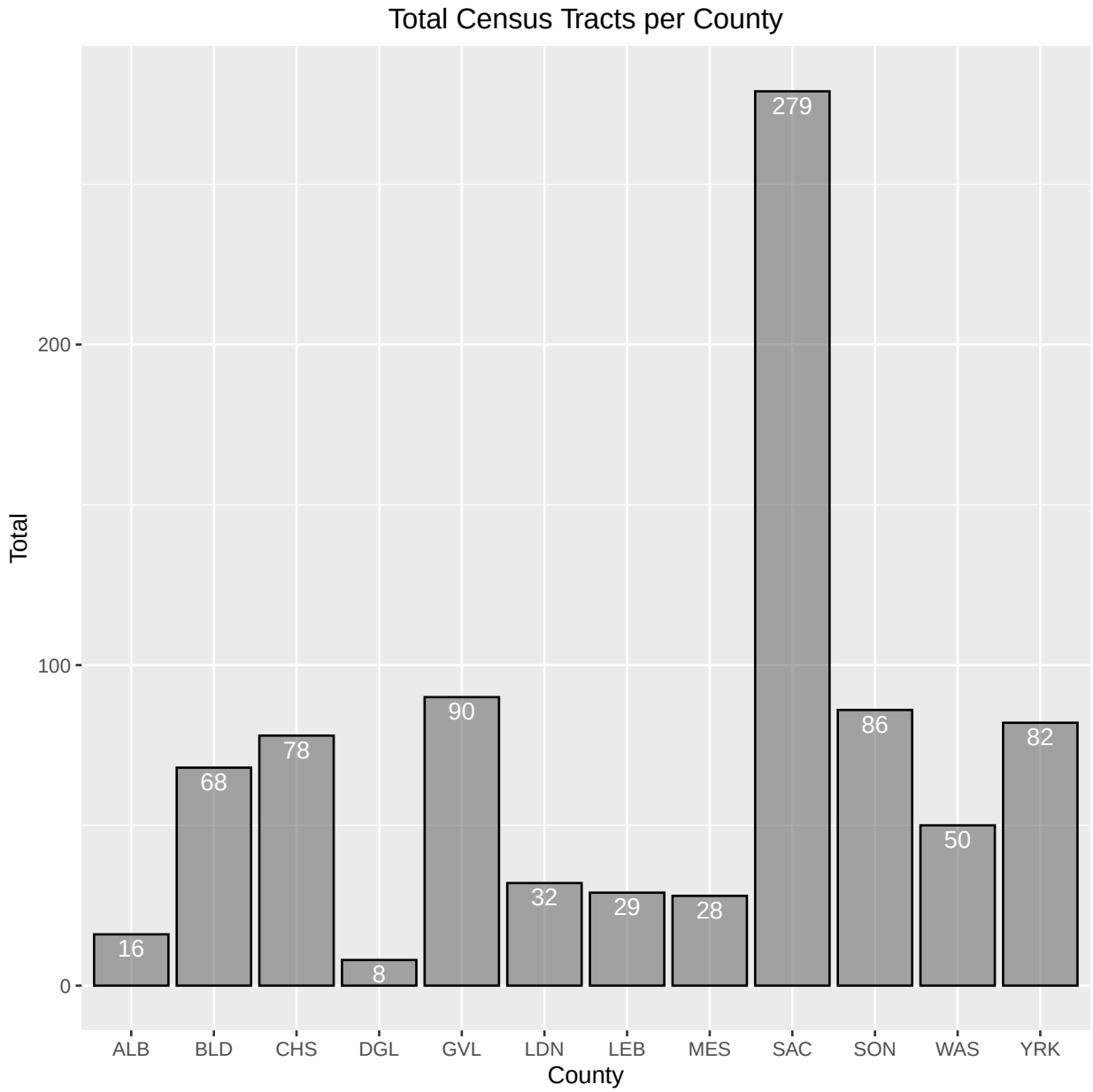

Figure 1: Total Census Tracts by County

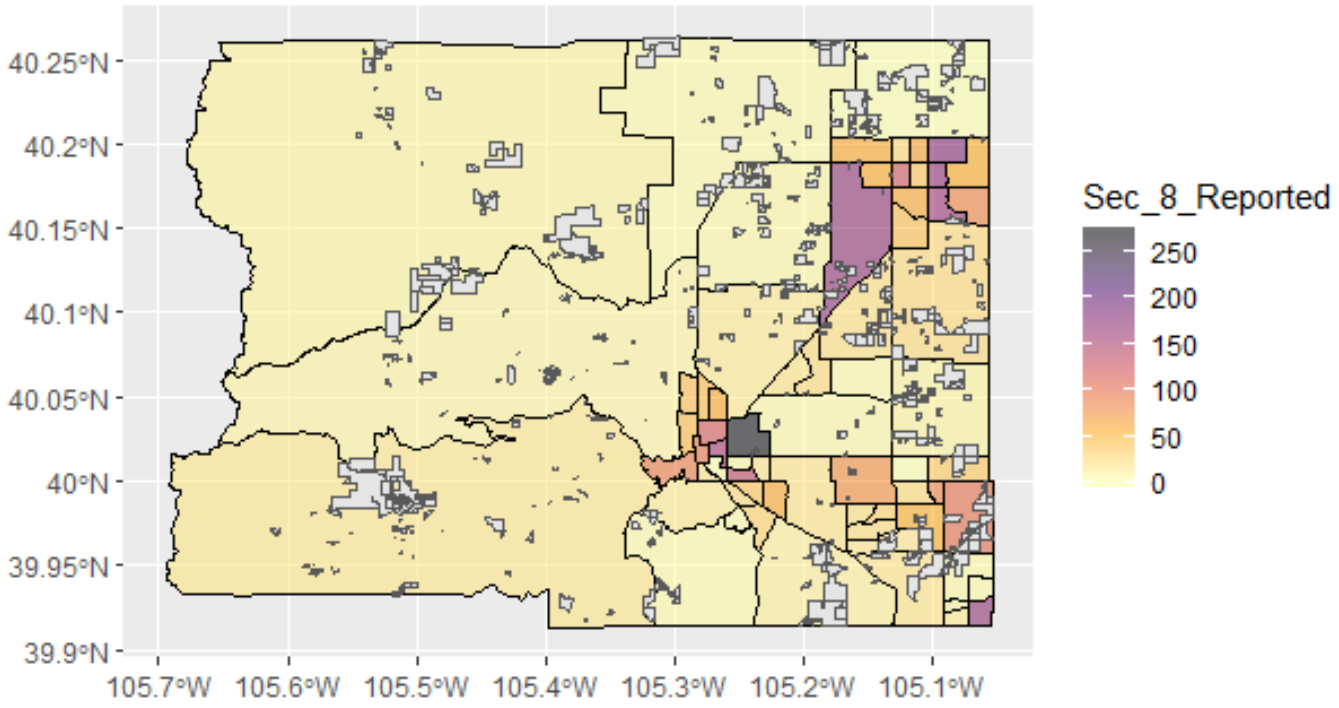

Figure 2: Section 8 Vouchers by Census Tract in Boulder, CO with CE Overlay

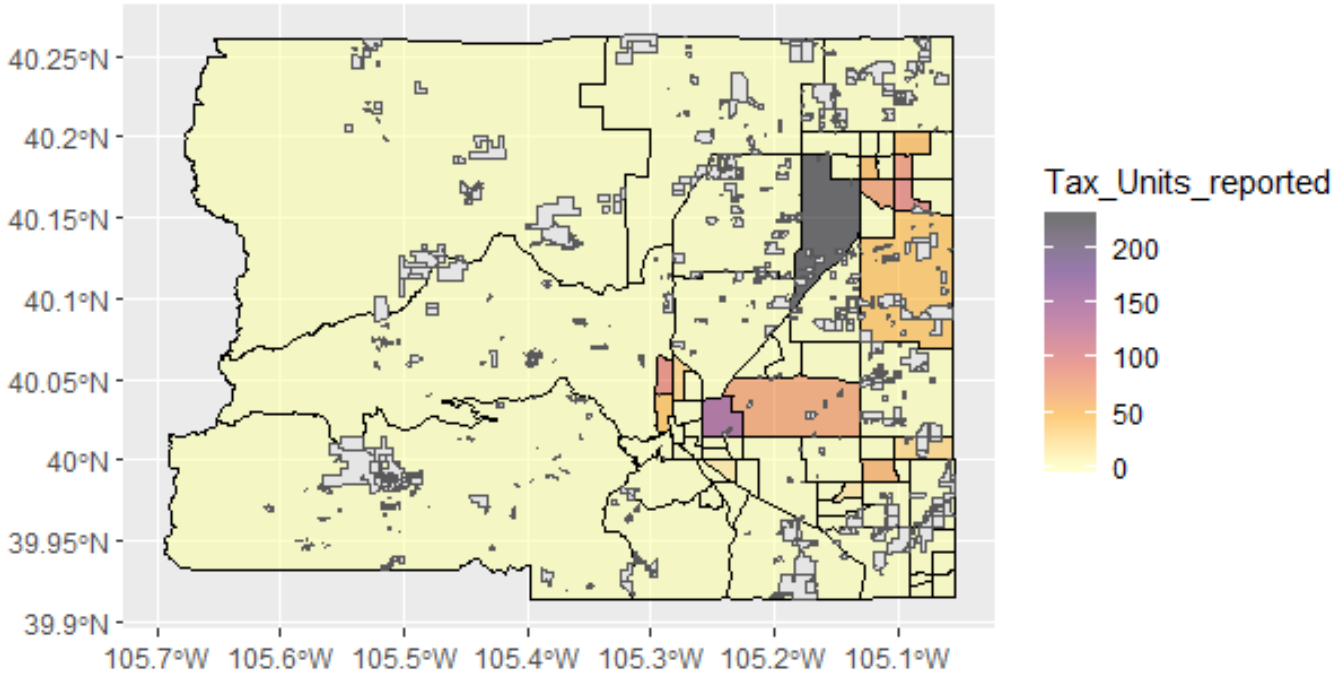

Figure 3: LIHTC by Census Tract in Boulder, CO with CE Overlay

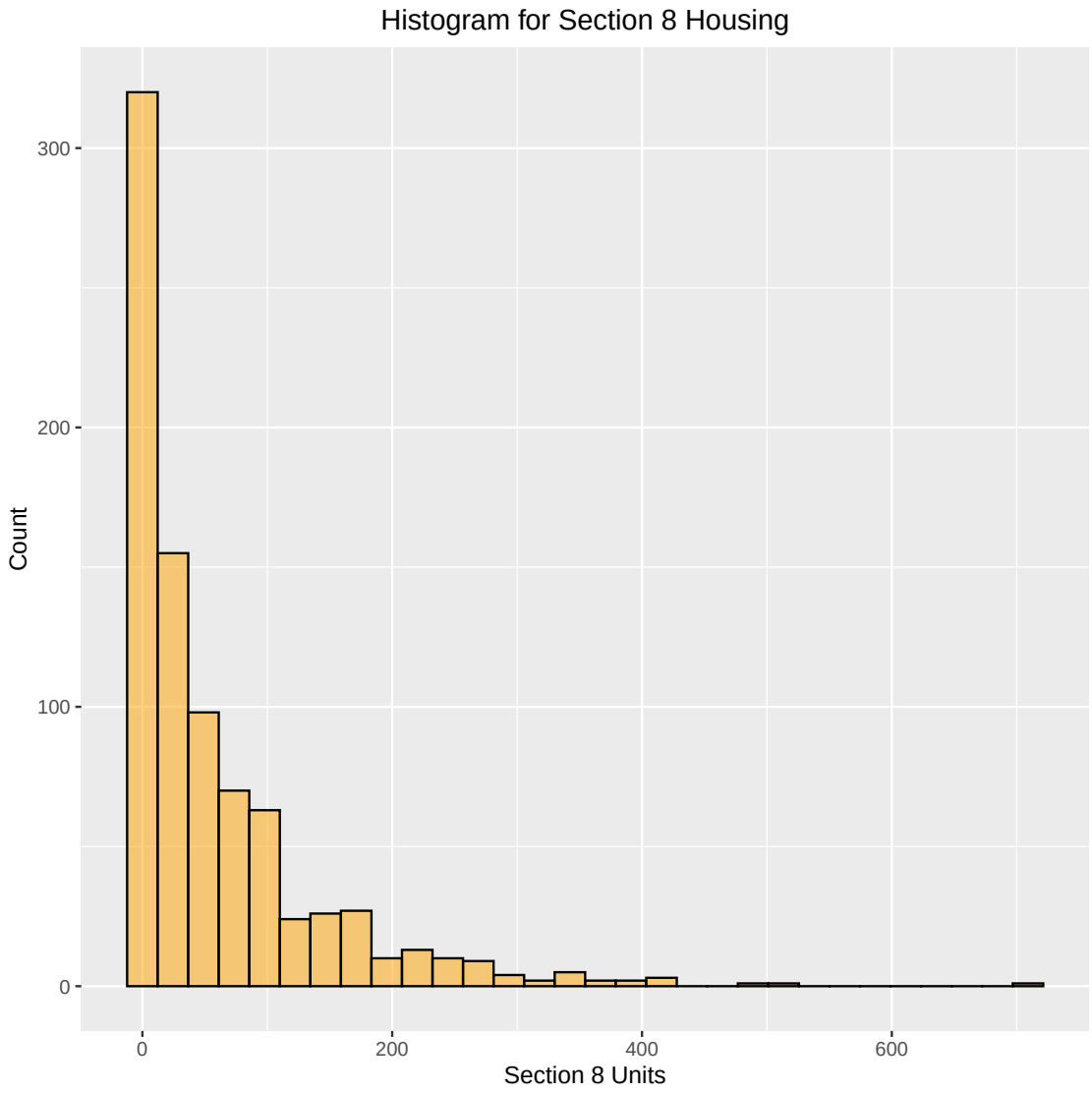

Figure 4: Histogram of Section 8 Units

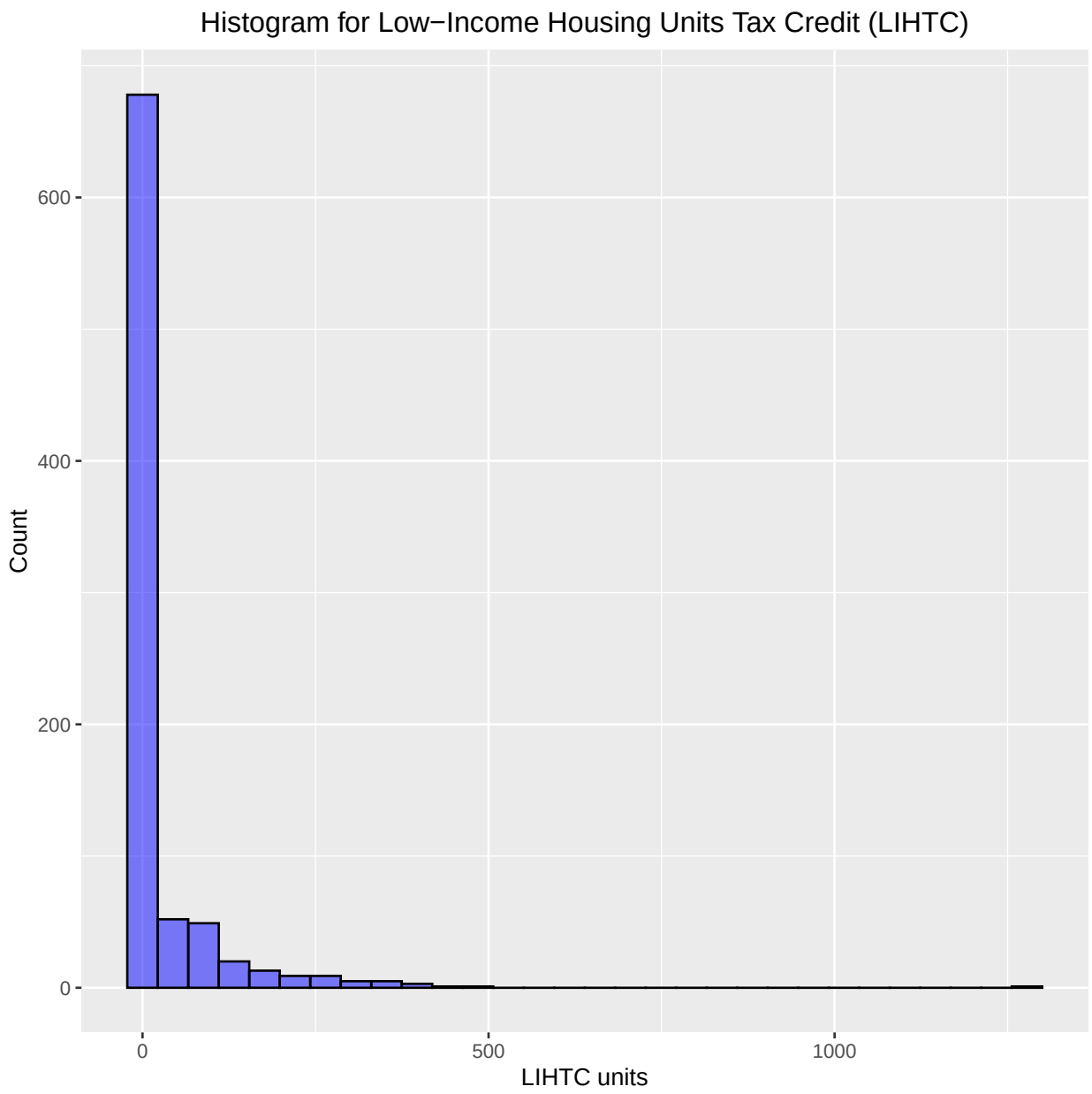

Figure 5: Histogram of LIHTC Units

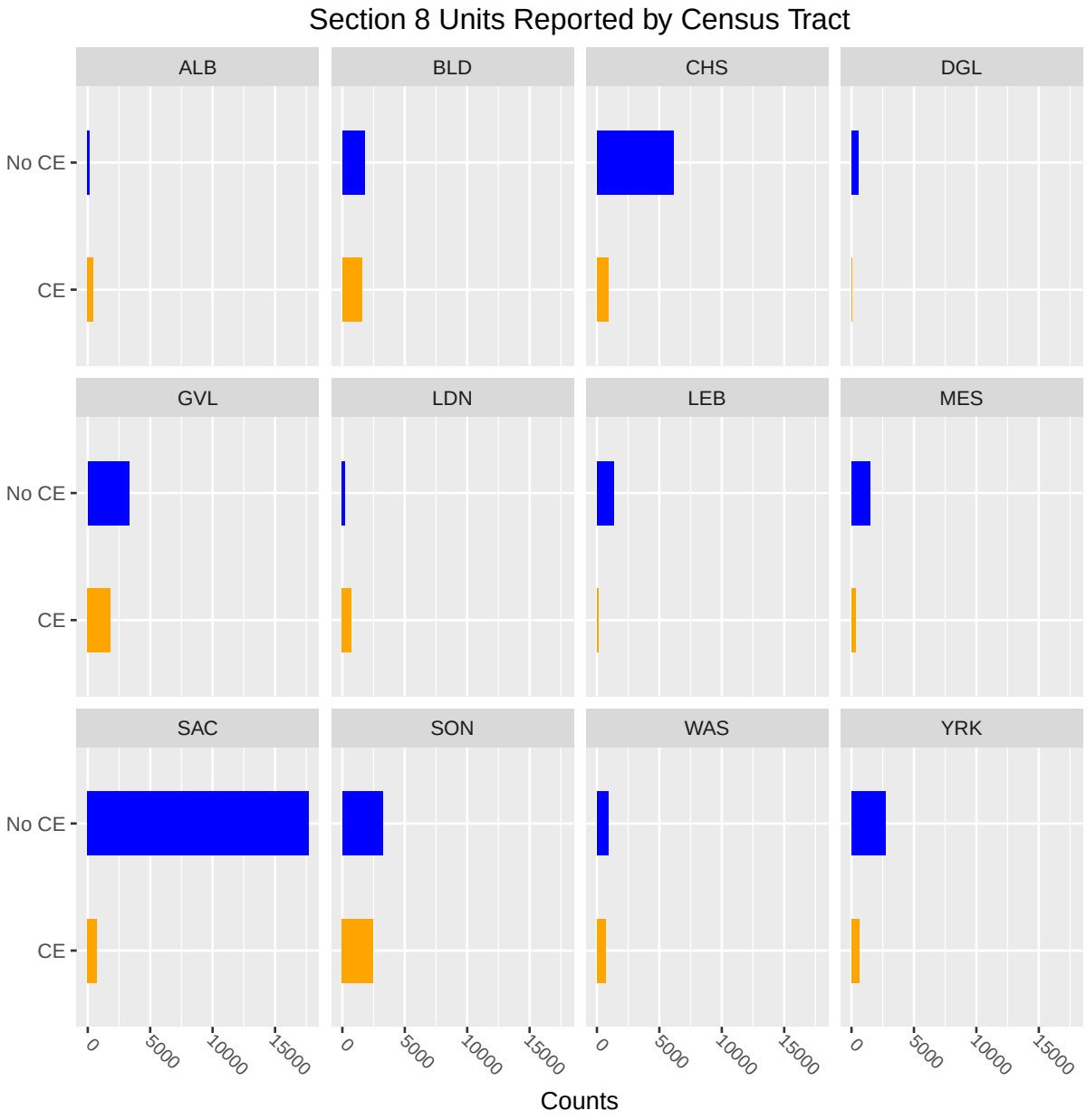

Figure 6: Section 8 Units Reported by Census Tract

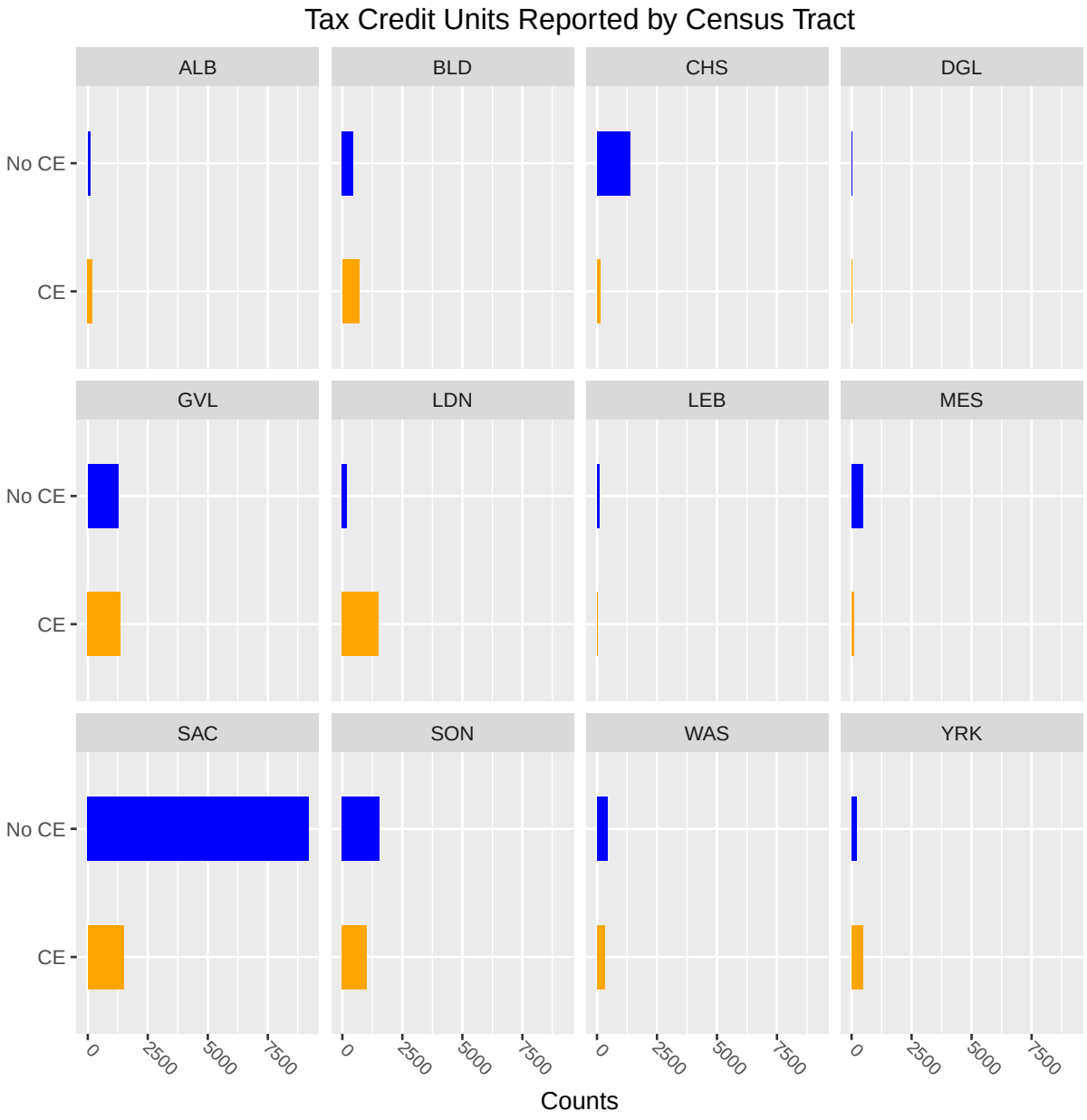

Figure 7: LIHTC Units Reported by Census Tract

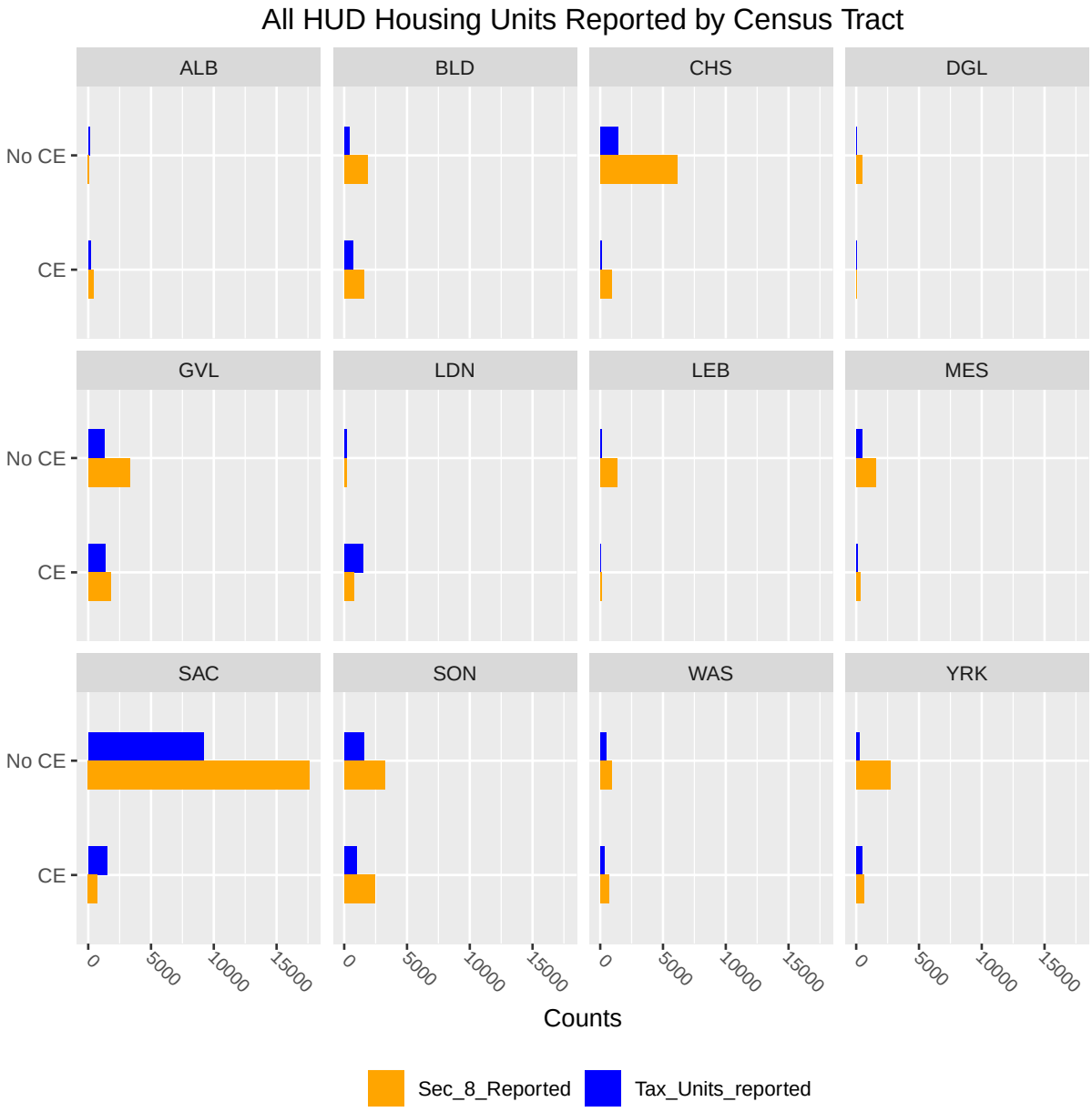

Figure 8: Housing Units Reported by Census Tract

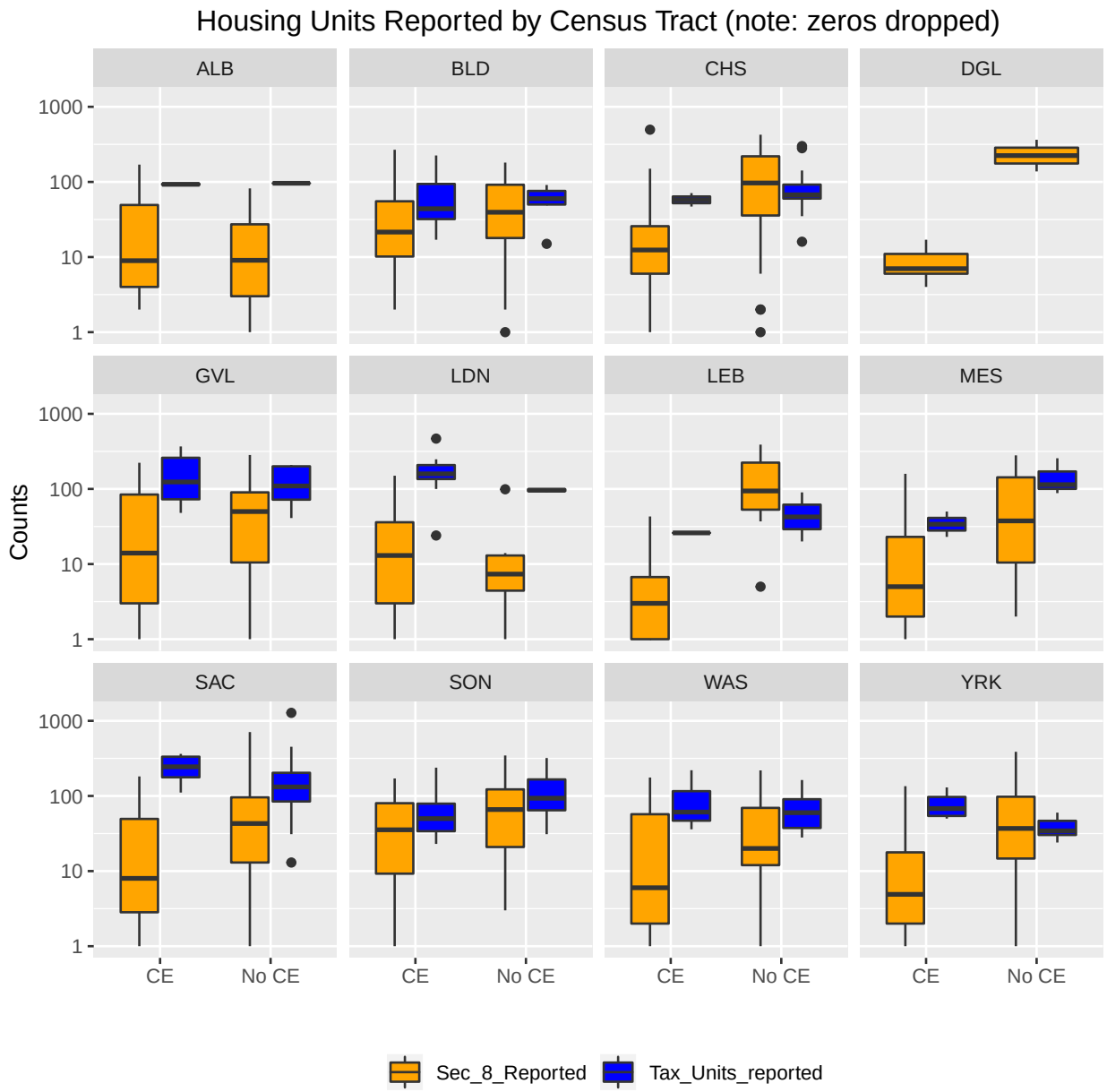

Figure 9: Housing Units Reported by Census Tract

Table 1: All Counties: Logistic Regression Results (GLM, Poisson) : HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.007***
	(0.001)
Tax_Units_reported	0.002**
	(0.001)
Constant	-0.738***
	(0.070)
Observations	846
Log Likelihood	-594.702
Akaike Inf. Crit.	1,195.404
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: All Counties: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	22.480	25.683	4.319	13.400	31.561	40.641
Resid. Df	3	844.000	1.000	843	843.5	844.5	845
Resid. Dev	3	593.830	24.805	577.404	579.563	602.043	622.364
Pr(>Chi)	2	0.019	0.027	0.000	0.009	0.028	0.038

Table 3: All Counties: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-594.702	-617.182	44.960	0.036	0.052	0.067

Table 4: Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-11.262	1	-0.362	0	greater	Overdispersion test	model1_Poisson

Table 5: All Counties: Logistic Regression Results, Zero Inflation Model (Poisson Distribution): HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.008***
	(0.001)
Tax_Units_reported	0.003***
	(0.001)
Constant	-0.745***
	(0.070)
Observations	846
Log Likelihood	-592.625
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: BLD Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.005
	(0.004)
Tax_Units_reported	0.005
	(0.005)
Constant	-0.478**
	(0.215)
Observations	68
Log Likelihood	-58.659
Akaike Inf. Crit.	123.318
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: BLD: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	0.859	0.250	0.682	0.770	0.947	1.036
Resid. Df	3	66.000	1.000	65	65.5	66.5	67
Resid. Dev	3	44.236	0.865	43.318	43.836	44.695	45.036
Pr(>Chi)	2	0.359	0.071	0.309	0.334	0.384	0.409

Table 8: BLD: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-58.659	-59.518	1.718	0.014	0.025	0.030

Table 9: BLD Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-4.550	1.000	-0.544	0	greater	Overdispersion test	BLD_model1

Table 10: CHS Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.005
	(0.004)
Tax_Units_reported	-0.005
	(0.011)
Constant	-0.931***
	(0.260)
Observations	78
Log Likelihood	-45.890
Akaike Inf. Crit.	97.780
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 11: CHS: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	2.666	3.512	0.183	1.424	3.907	5.149
Resid. Df	3	76.000	1.000	75	75.5	76.5	77
Resid. Dev	3	51.618	3.027	49.780	49.872	52.537	55.112
Pr(>Chi)	2	0.346	0.457	0.023	0.185	0.508	0.669

Table 12: CHS: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-45.890	-48.556	5.332	0.055	0.066	0.093

Table 13: CHS Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-2.778	0.997	-0.269	0	greater	Overdispersion test	CHS_model1

Table 14: GVL Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.003
	(0.003)
Tax_Units_reported	0.002
	(0.002)
Constant	-0.771***
	(0.203)
Observations	90
Log Likelihood	-69.907
Akaike Inf. Crit.	145.814
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 15: GVL: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	0.858	0.237	0.690	0.774	0.941	1.025
Resid. Df	3	88.000	1.000	87	87.5	88.5	89
Resid. Dev	3	64.727	0.863	63.814	64.326	65.184	65.529
Pr(>Chi)	2	0.359	0.067	0.311	0.335	0.382	0.406

Table 16: GVL: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-69.907	-70.764	1.715	0.012	0.019	0.024

Table 17: GVL Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-4.091	1.000	-0.422	0	greater	Overdispersion test	GVL_model1

Table 18: LDN Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	0.001
	(0.006)
Tax_Units_reported	0.001
	(0.002)
Constant	-0.469*
	(0.266)
Observations	32
Log Likelihood	-30.002
Akaike Inf. Crit.	66.004
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 19: LDN: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	0.241	0.083	0.182	0.212	0.271	0.300
Resid. Df	3	30.000	1.000	29	29.5	30.5	31
Resid. Dev	3	16.225	0.244	16.004	16.095	16.336	16.487
Pr(>Chi)	2	0.626	0.061	0.584	0.605	0.648	0.669

Table 20: LDN: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-30.002	-30.243	0.483	0.008	0.015	0.018

Table 21: LDN Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-4.213	1.000	-0.687	0	greater	Overdispersion test	LDN_model1

Table 22: LEB Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.032
	(0.020)
Tax_Units_reported	0.017
	(0.040)
Constant	0.004
	(0.241)
Observations	29
Log Likelihood	-22.141
Akaike Inf. Crit.	50.283
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 23: LEB: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	5.290	7.260	0.156	2.723	7.857	10.424
Resid. Df	3	27.000	1.000	26	26.5	27.5	28
Resid. Dev	3	7.861	6.064	4.283	4.361	9.651	14.863
Pr(>Chi)	2	0.347	0.489	0.001	0.174	0.520	0.693

Table 24: LEB: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-22.141	-27.431	10.580	0.193	0.306	0.360

Table 25: LEB Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-6.024	1	-0.690	0	greater	Overdispersion test	LEB_model1

Table 26: MES Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.009
	(0.009)
Tax_Units_reported	0.003
	(0.014)
Constant	-0.657*
	(0.373)
Observations	28
Log Likelihood	-19.084
Akaike Inf. Crit.	44.168
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 27: MES: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	1.212	1.647	0.047	0.630	1.795	2.377
Resid. Df	3	26.000	1.000	25	25.5	26.5	27
Resid. Dev	3	18.992	1.386	18.168	18.192	19.404	20.592
Pr(>Chi)	2	0.475	0.498	0.123	0.299	0.652	0.828

Table 28: MES: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-19.084	-20.296	2.424	0.060	0.083	0.108

Table 29: MES Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-2.030	0.979	-0.357	0	greater	Overdispersion test	MES_model1

Table 30: SAC Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.016**
	(0.007)
Tax_Units_reported	0.003***
	(0.001)
Constant	-1.970***
	(0.269)
Observations	279
Log Likelihood	-74.861
Akaike Inf. Crit.	155.722
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 31: SAC: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	5.541	0.895	4.908	5.224	5.857	6.173
Resid. Df	3	277.000	1.000	276	276.5	277.5	278
Resid. Dev	3	109.051	5.553	103.722	106.176	111.716	114.803
Pr(>Chi)	2	0.020	0.010	0.013	0.016	0.023	0.027

Table 32: SAC: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-74.861	-80.402	11.081	0.069	0.039	0.089

Table 33: SAC Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-2.778	0.997	-0.269	0	greater	Overdispersion test	CHS_model1

Table 34: SON Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.004
	(0.003)
Tax_Units_reported	-0.001
	(0.003)
Constant	-0.375*
	(0.201)
Observations	86
Log Likelihood	-74.130
Akaike Inf. Crit.	154.260
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 35: SON: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	1.267	1.748	0.031	0.649	1.885	2.504
Resid. Df	3	84.000	1.000	83	83.5	84.5	85
Resid. Dev	3	55.115	1.454	54.260	54.276	55.543	56.795
Pr(>Chi)	2	0.487	0.528	0.114	0.300	0.674	0.860

Table 36: SON: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-74.130	-75.397	2.535	0.017	0.029	0.035

Table 37: SON Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-5.168	1.000	-0.547	0	greater	Overdispersion test	SON_model1

Table 38: WAS Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.003
	(0.005)
Tax_Units_reported	-0.00005
	(0.007)
Constant	-0.524**
	(0.223)
Observations	50
Log Likelihood	-43.350
Akaike Inf. Crit.	92.700
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 39: WAS: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	0.287	0.406	0.0001	0.143	0.430	0.574
Resid. Df	3	48.000	1.000	47	47.5	48.5	49
Resid. Dev	3	32.892	0.331	32.700	32.700	32.987	33.274
Pr(>Chi)	2	0.722	0.386	0.449	0.585	0.858	0.994

Table 40: WAS: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-43.350	-43.637	0.574	0.007	0.011	0.014

Table 41: WAS Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-3.829	1.000	-0.540	0	greater	Overdispersion test	WAS_model1

Table 42: YRK Regression Results: HUD Housing

	Dependent variable:
	CE_Present
Sec_8_Reported	-0.016**
	(0.007)
Tax_Units_reported	0.010*
	(0.006)
Constant	-0.401**
	(0.180)
Observations	82
Log Likelihood	-64.841
Akaike Inf. Crit.	135.683
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 43: YRK: Analysis of Deviance

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Df	2	1.000	0.000	1.000	1.000	1.000	1.000
Deviance	2	5.916	4.893	2.456	4.186	7.646	9.376
Resid. Df	3	80.000	1.000	79	79.5	80.5	81
Resid. Dev	3	48.446	6.244	43.683	44.911	50.827	55.515
Pr(>Chi)	2	0.060	0.081	0.002	0.031	0.088	0.117

Table 44: YRK: McFadden Statistic:similar to R2

llh	llhNull	G2	McFadden	r2ML	r2CU
-64.841	-70.757	11.832	0.084	0.134	0.163

Table 45: YRK Overdisperson Test

	statistic	p.value	estimate	null.value	alternative	method	data.name
z	-2.778	0.997	-0.269	0	greater	Overdispersion test	CHS_model1