**Multilinear Regression**

**Significant variables:** region, subregion, population, % of no diploma, % of vacancy, % of no job, % of poverty, % of med inc, % of change job, % of change business, % of whites, % of Hispanics, % of Asians, % of mixed or other races, area of land in square miles, and area of water in square miles

**Variation explained by significant variables:** 84.1%

**Nonsignificant variables:** % ofBlacks,population density for land in square miles, and population density for water in square miles

Call:

lm(formula = DCI\_Percentile ~ Region + SubRegion + Population +

No\_Diploma + Vacancy + No\_Job + Poverty + Med\_Inc + Chg\_Job +

Chg\_Biz + White + Hispanic + Asian + Mixed\_Other + Area\_Land +

Area\_Water, data = dataset)

Residuals:

Min 1Q Median 3Q Max

-131.573 -7.965 -0.277 8.139 315.169

Coefficients: (3 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.748e+01 7.626e-01 36.032 < 2e-16 \*\*\*

RegionNorthEast -1.874e+00 3.446e-01 -5.437 5.46e-08 \*\*\*

RegionSouth 1.992e+00 2.968e-01 6.711 1.97e-11 \*\*\*

RegionWest -1.511e+00 3.232e-01 -4.676 2.95e-06 \*\*\*

SubRegionEast South Central 8.793e-01 3.636e-01 2.418 0.015602 \*

SubRegionMid-Atlantic 3.173e+00 3.555e-01 8.927 < 2e-16 \*\*\*

SubRegionMountain 1.228e-01 3.782e-01 0.325 0.745359

SubRegionNew England NA NA NA NA

SubRegionPacific NA NA NA NA

SubRegionSouth Atlantic 5.868e-02 2.992e-01 0.196 0.844542

SubRegionWest North Central -1.723e+00 2.805e-01 -6.143 8.21e-10 \*\*\*

SubRegionWest South Central NA NA NA NA

Population -3.047e-05 5.606e-06 -5.435 5.54e-08 \*\*\*

No\_Diploma 6.206e-01 1.295e-02 47.907 < 2e-16 \*\*\*

Vacancy 1.106e+00 1.538e-02 71.952 < 2e-16 \*\*\*

No\_Job 5.536e-01 8.789e-03 62.987 < 2e-16 \*\*\*

Poverty 6.091e-01 1.196e-02 50.937 < 2e-16 \*\*\*

Med\_Inc -2.402e-01 2.905e-03 -82.713 < 2e-16 \*\*\*

Chg\_Job 3.587e-04 1.079e-04 3.324 0.000888 \*\*\*

Chg\_Biz -2.632e-01 3.125e-03 -84.248 < 2e-16 \*\*\*

White 3.882e-02 5.304e-03 7.318 2.58e-13 \*\*\*

Hispanic -2.964e-02 8.533e-03 -3.473 0.000515 \*\*\*

Asian 4.326e-02 1.528e-02 2.831 0.004650 \*\*

Mixed\_Other 5.613e-02 5.876e-03 9.553 < 2e-16 \*\*\*

Area\_Land 3.037e-03 3.746e-04 8.108 5.38e-16 \*\*\*

Area\_Water -1.301e-02 5.847e-03 -2.224 0.026130 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.51 on 26102 degrees of freedom

Multiple R-squared: 0.8411, Adjusted R-squared: 0.841

F-statistic: 6282 on 22 and 26102 DF, p-value: < 2.2e-16

**AIC using backwards stepwise selection:**

Df Sum of Sq RSS AIC

<none> 3458414 127684

- Area\_Water 1 656 3459070 127687 **🡨Least Important Variable**

- Asian 1 1062 3459476 127690

- Chg\_Job 1 1464 3459878 127693

- Hispanic 1 1598 3460013 127694

- Population 1 3913 3462328 127712

- White 1 7096 3465511 127736

- Area\_Land 1 8710 3467124 127748

- Mixed\_Other 1 12091 3470505 127773

- SubRegion 8 53416 3511831 128069

- No\_Diploma 1 304087 3762501 129884

- Poverty 1 343776 3802191 130158

- No\_Job 1 525666 3984081 131379

- Vacancy 1 685951 4144365 132409

- Med\_Inc 1 906460 4364875 133764

- Chg\_Biz 1 940420 4398835 133966 **🡨Most Important Variable**

**Criteria for classifying population density by land area in square miles**

=IF(AJ2<25,"0-25",IF(AJ2<50,"25-50",IF(AJ2<75,"50-75",IF(AJ2<100,"75-100",IF(AJ2<250,"100-250",IF(AJ2<500,"250-500",IF(AJ2<750,"500-750",IF(AJ2<1000,"750-1000",IF(AJ2<2500,"1000-2500",IF(AJ2<5000,"2500-5000",IF(AJ2<7500,"5000-7500",IF(AJ2<10000,"7500-10000","10000+"))))))))))))