Supplementary Information for

Evidence and Theory for Lower Rates of Depression in Larger U.S. Urban Areas

A. J. Stier, K. E. Schertz, Nak Won Rim, C. Cardenas-Iniguez, B. B. Lahey, L. M. A. Bettencourt, M. G. Berman

[1]To whom correspondence should be addressed. E-mail: andrewstier@uchicago.edu, bermanm@uchicago.edu

This PDF file includes:

Figs. S1 to S6 (not allowed for Brief Reports)
Tables S1 to S8 (not allowed for Brief Reports)

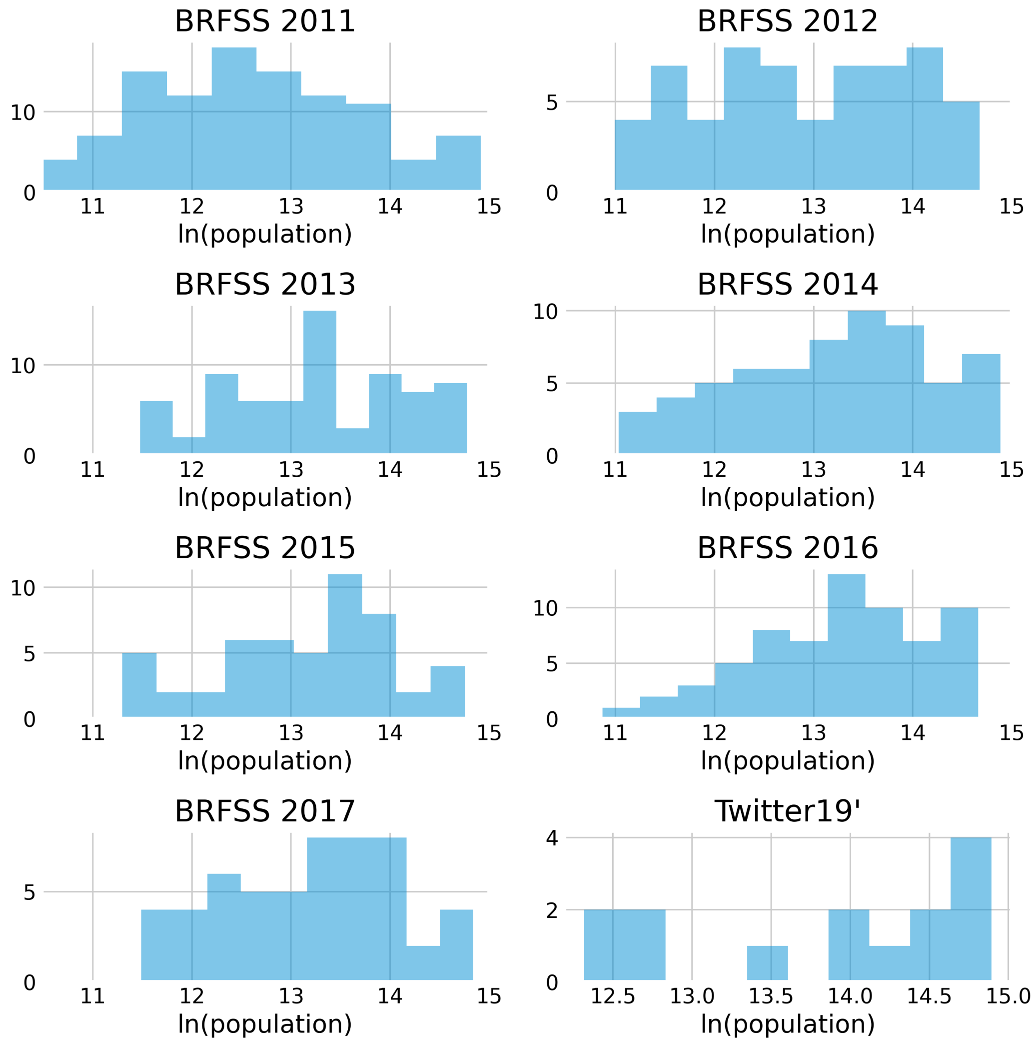

Fig. S1. Histograms of the detected change points for all window sizes in BRFSS data and Twitter19'. We used a covariate discriminant method (see Methods) to non-parametrically detect changes in the joint distribution of depression rates and population, under the assumption that BRFSS report methods might induce an artificial change. For the Twitter19' dataset, the detection of change points primarily at the edges of the population range is indicative of finite edge effects rather than a true change in the joint distribution of depression rates and city size. Compare to BRFSS data where change points are detected in the middle of the population range.

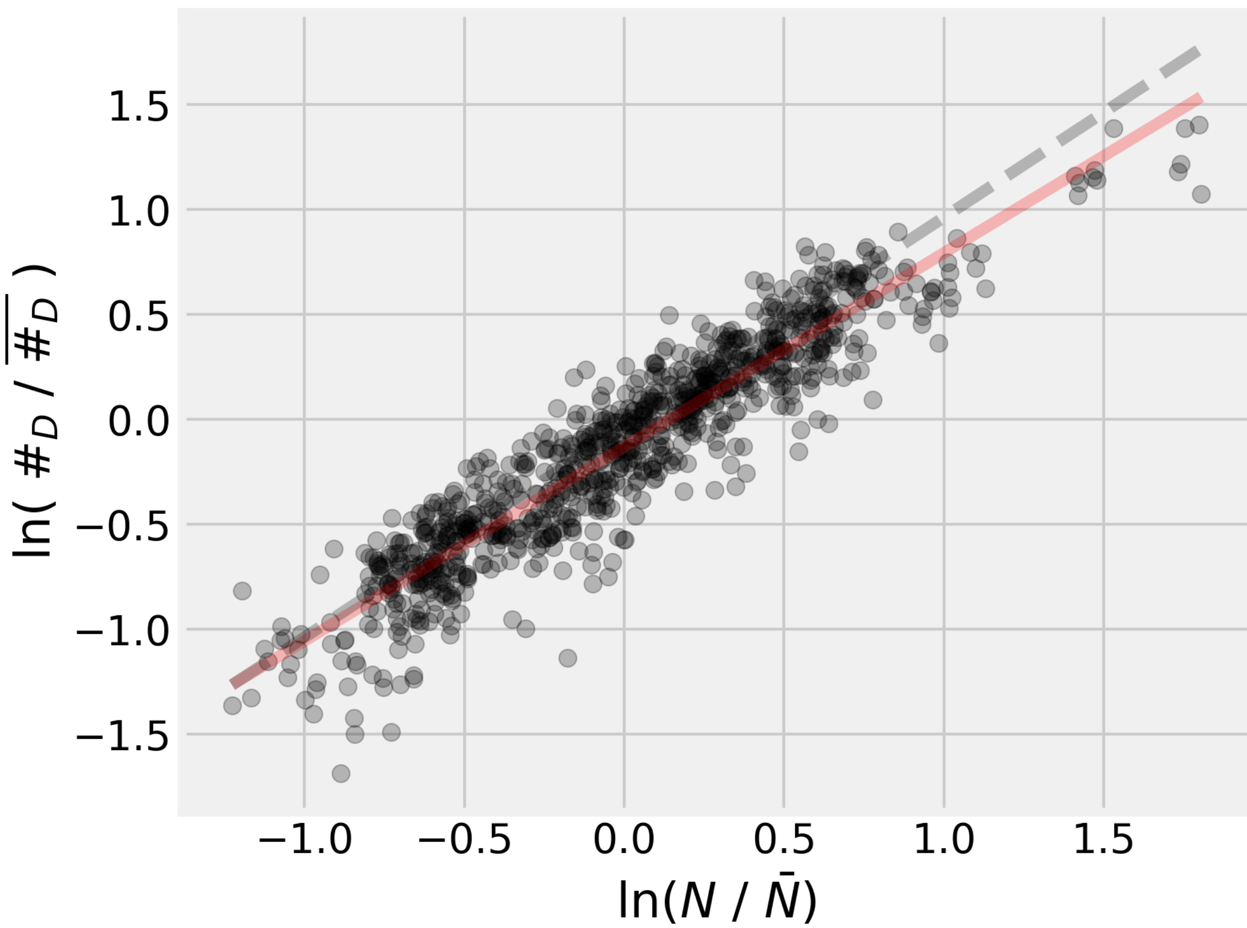

Fig. S2. Pooling BRFSS data across years for all cities results in a scaling exponent of $\beta = 0.926$ (95% CI = [0.903, 0.950]), consistent with lower depression rates in larger cities.

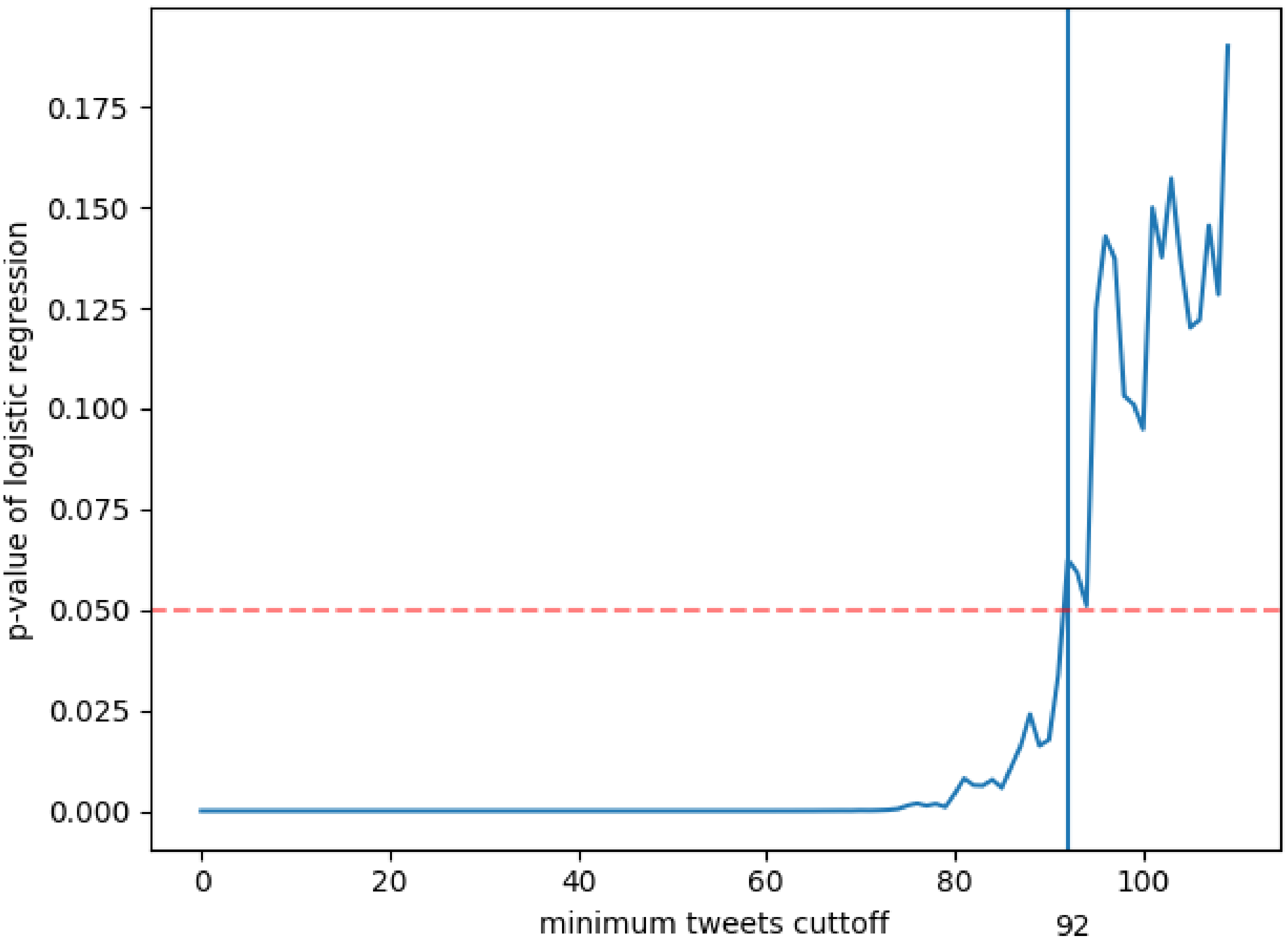

Fig. S3. Users with lower numbers of tweets are more likely to have depressive sentiment in their tweets. When using an exclusion criteria of less that 92 tweets a logistic regression model significantly distinguishes individuals with depressive sentiment from individuals without depressive sentiment.

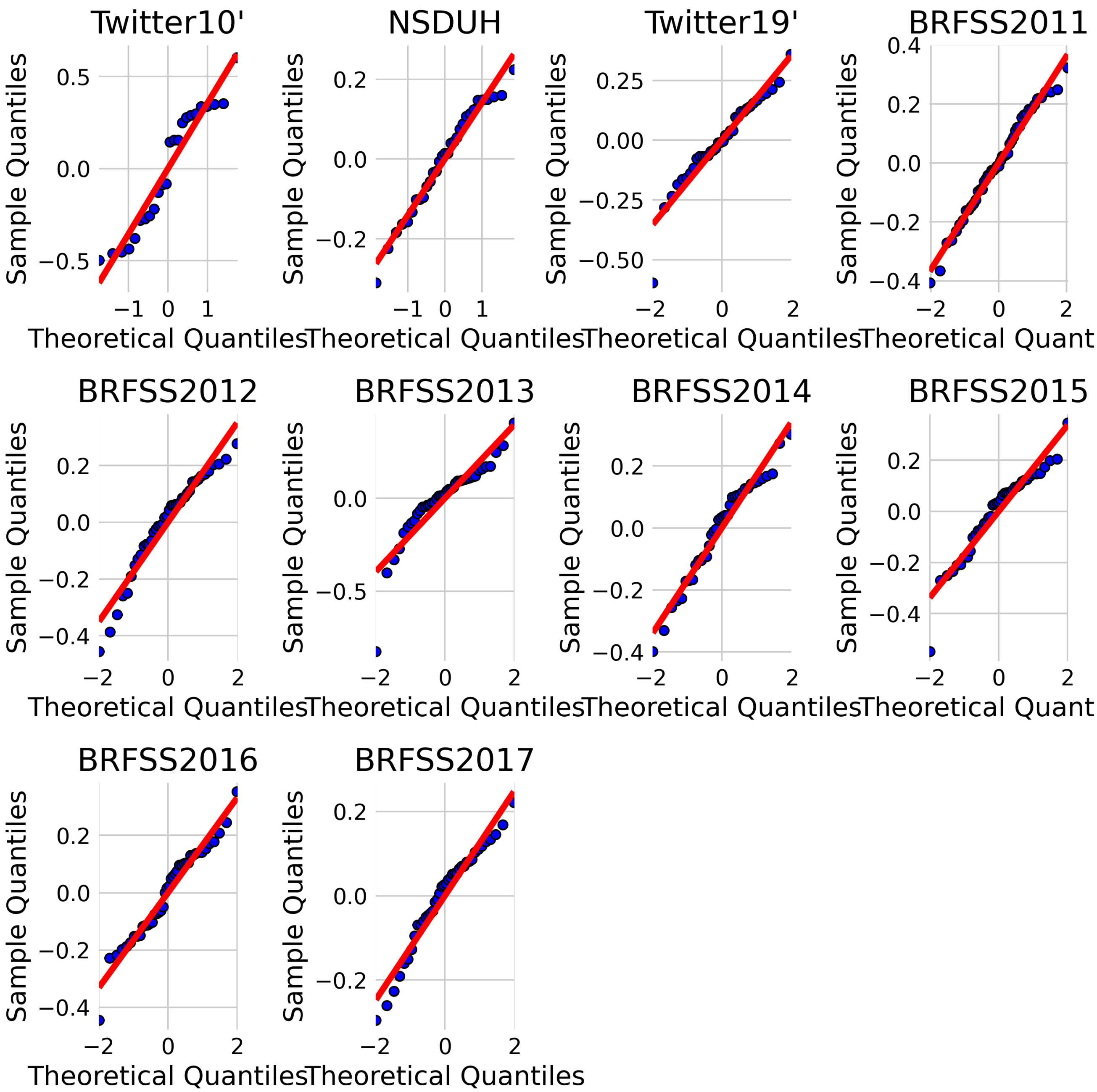

Fig. S4. QQ plots of the residuals of the OLS model. No significant deviations are observed indicating that the residuals are approximately normally distributed and the linear model is appropriate.

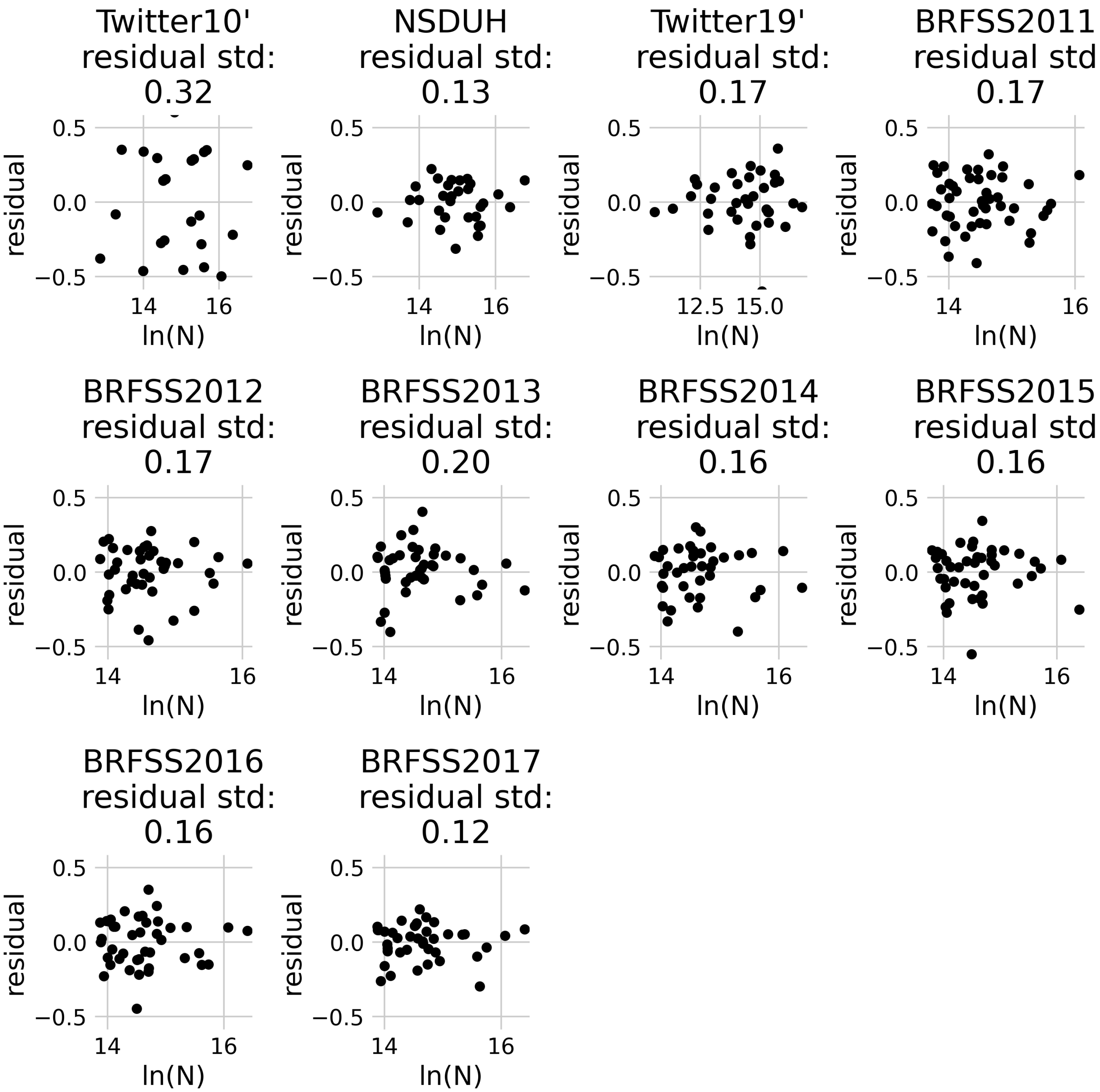

Fig. S5. Residuals from OLS models are not correlated with city size. In all datasets, residuals are not correlated with city size (Spearman-r minimum p-value = 0.44). Thus no corrections to estimates of β are required.

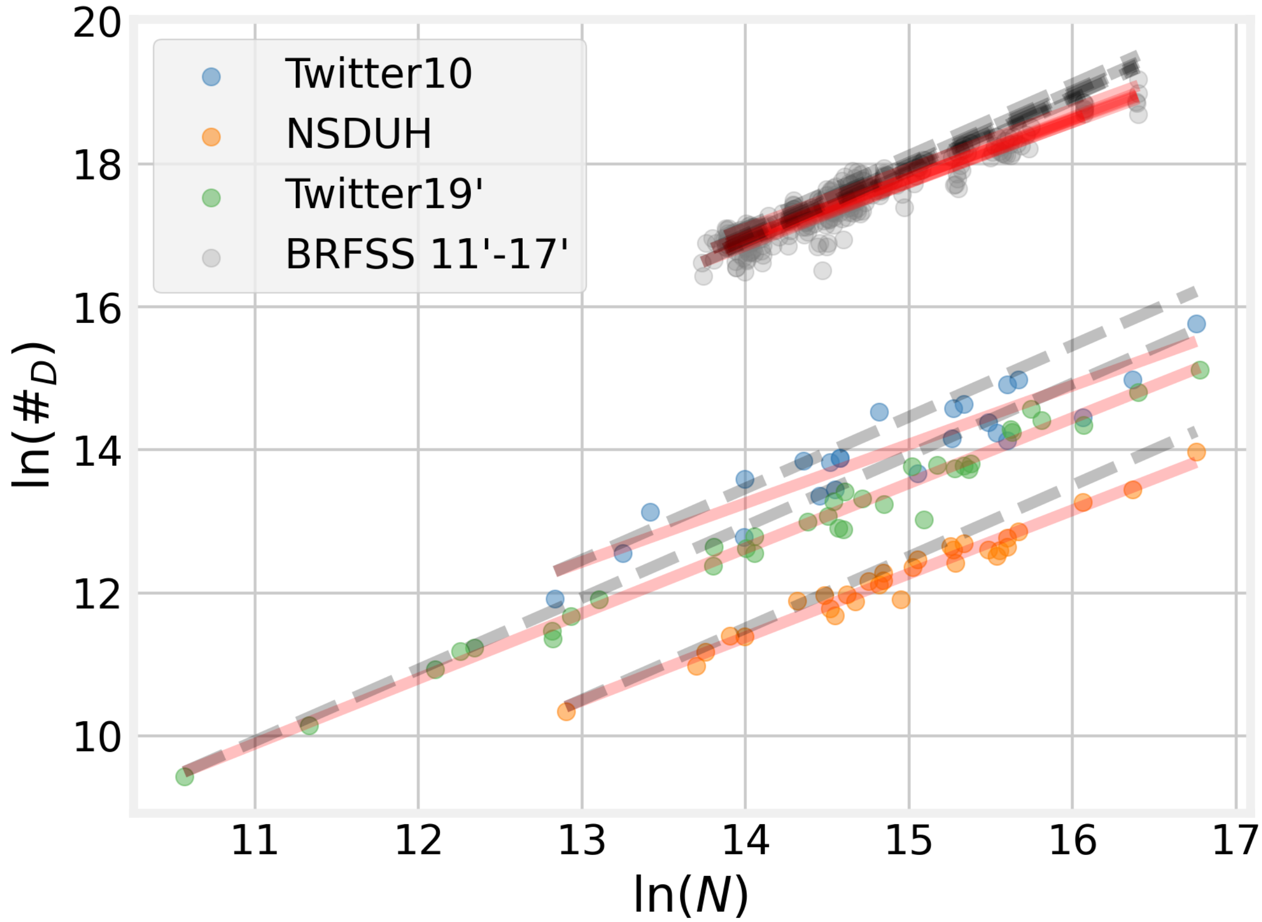

Fig. S6. OLS fit to each dataset. Sublinear scaling is observed across all datasets.

Table S1. MSAs included in the analysis in the main text (Fig. 2.). Included MSAs are marked with an X.

MSA	Twitter10'	BFRSS	NSDUH	Twitter19'	2017 Population
Albuquerque, NM	-	-	X	-	912897
Ann Arbor, MI	-	-	-	X	369208
Atlanta-Sandy Springs-Alpharetta, GA	X	X	X	-	5874249
Augusta-Richmond County, GA-SC	X	-	-	-	600006
Austin-Round Rock-Georgetown, TX	-	X	-	-	2115230
Baltimore-Columbia-Towson, MD	X	X	X	-	2798587
Birmingham-Hoover, AL	-	X	-	-	1085750
Boston-Cambridge-Newton, MA-NH	X	-	X	-	4844597
Bridgeport-Stamford-Norwalk, CT	-	X	-	-	943457
Buffalo-Cheektowaga, NY	-	X	-	-	1129660
Charlotte-Concord-Gastonia, NC-SC	-	X	-	-	2549741
Chicago-Naperville-Elgin, IL-IN-WI	X	X	X	X	9520784
Cincinnati, OH-KY-IN	X	X	-	X	2202597
Cleveland-Elyria, OH	X	X	X	X	2058549
Columbus, OH	-	X	-	-	2082475
Dallas-Fort Worth-Arlington, TX	X	-	X	X	7340943
Denver-Aurora-Lakewood, CO	-	X	X	-	2892979
Detroit-Warren-Dearborn, MI	X	-	X	X	4321704
Fresno, CA	-	-	-	X	986542
Grand Rapids-Kentwood, MI	-	X	-	-	1063926
Gulfport-Biloxi, MS	-	-	-	X	412946
Hartford-East Hartford-Middletown, CT	-	X	-	-	1206719
Houma-Thibodaux, LA	-	-	-	X	209893
Houston-The Woodlands-Sugar Land, TX	X	X	X	X	6905695
Indianapolis-Carmel-Anderson, IN	X	X	-	-	2026723
Jacksonville, FL	-	X	-	-	1504841
Kansas City, MO-KS	X	X	X	X	2127259
Lafayette, LA	-	-	-	X	490107
Las Vegas-Henderson-Paradise, NV	-	X	X	X	2183310
Lock Haven, PA	-	-	-	X	38837
Los Angeles-Long Beach-Anaheim, CA	X	X	X	X	13298709
Louisville/Jefferson County, KY-IN	-	X	-	-	1260391
Macon-Bibb County, GA	-	-	-	X	229081
Manchester-Nashua, NH	-	-	X	-	413157
Memphis, TN-MS-AR	-	X	-	-	1339290
Miami-Fort Lauderdale-Pompano Beach, FL	X	X	X	X	6149687
Milwaukee-Waukesha, WI	-	X	-	-	1575151
Minneapolis-St. Paul-Bloomington, MN-WI	-	X	X	X	3577765
Montgomery, AL	X	-	-	-	374042
Nashville-Davidson–Murfreesboro–Franklin, TN	-	X	X	-	1875736
New Orleans-Metairie, LA	X	X	X	X	1270465
New York-Newark-Jersey City, NY-NJ-PA	X	-	X	X	19325698
Oklahoma City, OK	-	X	-	-	1383249
Opelousas, LA	-	-	-	X	83447
Orlando-Kissimmee-Sanford, FL	X	X	-	-	2512917
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	X	-	X	X	6078451
Phoenix-Mesa-Chandler, AZ	-	X	X	X	4761694
Pittsburgh, PA	-	X	X	-	2330283
Portland-Vancouver-Hillsboro, OR-WA	-	X	X	X	2456462
Poughkeepsie-Newburgh-Middletown, NY	X	-	-	-	673253
Providence-Warwick, RI-MA	-	X	-	-	1617057
Raleigh-Cary, NC	-	X	-	-	1334342
Richmond, VA	X	X	-	X	1269478
Riverside-San Bernardino-Ontario, CA	X	X	-	X	4570427
Rochester, NY	-	X	-	-	1071589
Sacramento-Roseville-Folsom, CA	-	X	-	-	2320381
Salt Lake City, UT	-	X	X	X	1205238
San Antonio-New Braunfels, TX	-	X	-	-	2474274
San Diego-Chula Vista-Carlsbad, CA	-	X	X	X	3325468

San Francisco-Oakland-Berkeley, CA	-	X	X	X	4710693
San Jose-Sunnyvale-Santa Clara, CA	-	X	-	X	1993582
Seattle-Tacoma-Bellevue, WA	X	-	X	X	3884469
St. Louis, MO-IL	-	X	X	X	2805850
Tampa-St. Petersburg-Clearwater, FL	-	X	X	-	3091225
Trenton-Princeton, NJ	-	-	-	X	368602
Tucson, AZ	-	X	-	-	1027502
Tulsa, OK	-	X	X	X	991610
Virginia Beach-Norfolk-Newport News, VA-NC	X	X	-	X	1761305
Warner Robins, GA	-	-	-	X	180019
Washington-Arlington-Alexandria, DC-VA-MD-WV	-	-	X	-	6213246
Worcester, MA-CT	-	X	-	-	942303

Table S2. Estimates of the scaling exponent made with BRFSS data from smaller cities that were below the estimated change point for each year.

Dataset	β	95% CI	R^2	n
BRFSS2011	1.000	[0.960, 1.039]	0.952	128
BRFSS2012	1.001	[0.961, 1.040]	0.954	122
BRFSS2013	1.020	[0.969, 1.070]	0.953	81
BRFSS2014	1.034	[0.991, 1.077]	0.969	74
BRFSS2015	1.044	[0.996, 1.093]	0.966	67
BRFSS2016	0.967	[0.906, 1.028]	0.931	76
BRFSS2017	1.010	[0.962, 1.058]	0.959	76

Table S3. Robustness of scaling exponent estimates made with BRFSS data to variation in the city size below which data was excluded.

Dataset	β	95% CI
BRFSS2011	0.88	[0.87, 0.89]
BRFSS2012	0.85	[0.85, 0.87]
BRFSS2013	0.86	[0.85, 0.87]
BRFSS2014	0.83	[0.83, 0.84]
BRFSS2015	0.83	[0.82, 0.84]
BRFSS2016	0.83	[0.82, 0.84]
BRFSS2017	0.83	[0.83, 0.85]

Table S4. Scaling exponent estimates for all BFRSS data. No cities below the change point are excluded.

Dataset	β	95% CI	R^2	n
BRFSS2011	0.966	[0.942, 0.991]	0.974	172
BRFSS2012	0.956	[0.931, 0.982]	0.972	161
BRFSS2013	0.951	[0.920, 0.982]	0.968	122
BRFSS2014	0.959	[0.932, 0.987]	0.978	111
BRFSS2015	0.961	[0.932, 0.990]	0.976	109
BRFSS2016	0.941	[0.910, 0.972]	0.968	119
BRFSS2017	0.965	[0.939, 0.991]	0.980	116

Table S5. Robustness of scaling exponent estimates to variation in the minimum number of tweets required for inclusion in the Twitter analyses.

Minimum Tweets	β	95% CI	# MSAs
82	0.85	[0.75, 0.96]	31
83	0.85	[0.75, 0.95]	29
84	0.86	[0.75, 0.96]	29
85	0.87	[0.75, 0.98]	28
86	0.86	[0.75, 0.98]	28
87	0.83	[0.69, 0.97]	26
88	0.83	[0.68, 0.98]	25
89	0.80	[0.65, 0.95]	25
90	0.79	[0.63, 0.94]	24
91	0.80	[0.65, 0.96]	24
92	0.82	[0.67, 0.97]	24
93	0.85	[0.70, 0.99]	22
94	0.84	[0.69, 0.98]	22
95	0.83	[0.70, 0.95]	22
96	0.83	[0.70, 0.97]	22
97	0.84	[0.71, 0.98]	22
98	0.86	[0.71, 1.00]	22
99	0.84	[0.70, 0.98]	22
100	0.81	[0.65, 0.97]	22
101	0.86	[0.68, 1.04]	21

Table S6. Shapiro-Wilk test of normality on the OLS residuals for each dataset. The residuals from the BRFSS 2013 data fail this normality test due to one outlier city with an negative residual.

Dataset	statistic	p-value	n
Twitter10'	0.917	5.03e-02	24
NSDUH	0.970	5.26e-01	31
Twitter19	0.948	9.26e-02	36
BRFSS2011	0.977	4.79e-01	48
BRFSS2012	0.951	8.96e-02	39
BRFSS2013	0.873	3.47e-04	40
BRFSS2014	0.964	2.54e-01	38
BRFSS2015	0.951	7.38e-02	41
BRFSS2016	0.969	2.95e-01	43
BRFSS2017	0.959	1.55e-01	40

Table S7. Result of logistic regression models for each year of BRFSS data. In addition to city natural-log-population and the rate of population change from the previous year, we conditioned on income, race, and education. The income variable had 6 levels with a baseline of not reported or missing, followed by: less than $15k, $15-$25k, $25-$35k, $35-$55k, and greater than $50k. The education variable had 5 levels with a baseline of not reported followed by: no high-school, graduated high-school, attended college, and graduated college. The race variable had 4 levels with a baseline of White followed by: Black, Asian, and other/multi-racial.

	Dependent variable:						
	dep						
	(2017)	(2016)	(2015)	(2014)	(2013)	(2012)	(2011)
logpop	−0.104***	−0.115***	−0.115***	−0.108***	−0.124***	−0.092***	−0.118***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)
inc1	1.062***	1.098***	1.103***	1.125***	1.102***	1.101***	1.111***
	(0.033)	(0.032)	(0.033)	(0.032)	(0.031)	(0.034)	(0.031)
inc2	0.583***	0.613***	0.679***	0.598***	0.588***	0.601***	0.568***
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)	(0.032)	(0.030)
inc3	0.321***	0.288***	0.344***	0.270***	0.282***	0.292***	0.318***
	(0.033)	(0.034)	(0.034)	(0.034)	(0.034)	(0.036)	(0.034)
inc4	0.206***	0.174***	0.188***	0.157***	0.189***	0.144***	0.164***
	(0.031)	(0.031)	(0.032)	(0.032)	(0.032)	(0.035)	(0.033)
inc5	−0.093***	−0.130***	−0.055**	−0.114***	−0.086***	−0.138***	−0.112***
	(0.024)	(0.025)	(0.025)	(0.026)	(0.027)	(0.029)	(0.027)
edu1	0.626***	0.416***	0.786***	0.434***	0.460***	0.691***	0.167
	(0.159)	(0.151)	(0.172)	(0.105)	(0.129)	(0.168)	(0.136)
edu2	0.541***	0.238	0.686***	0.302***	0.302**	0.587***	0.024
	(0.157)	(0.149)	(0.170)	(0.102)	(0.128)	(0.167)	(0.135)
edu3	0.755***	0.450***	0.923***	0.511***	0.551***	0.779***	0.262*
	(0.157)	(0.149)	(0.170)	(0.102)	(0.127)	(0.167)	(0.134)
edu4	0.534***	0.261*	0.718***	0.320***	0.345***	0.618***	0.123
	(0.157)	(0.149)	(0.170)	(0.102)	(0.127)	(0.167)	(0.135)
rac2	−0.474***	−0.509***	−0.487***	−0.460***	−0.497***	−0.513***	−0.497***
	(0.027)	(0.028)	(0.029)	(0.027)	(0.027)	(0.029)	(0.028)
rac3	0.116*	0.108	−0.017	0.029	−0.014	−1.186***	−1.084***
	(0.063)	(0.068)	(0.069)	(0.068)	(0.069)	(0.097)	(0.086)
rac4	−0.327***	−0.371***	−0.392***	−0.336***	−0.335***	−0.069**	−0.074**
	(0.029)	(0.030)	(0.032)	(0.031)	(0.031)	(0.031)	(0.030)
pop_change_rate	−1.194	−1.803*	−1.427	−1.874*	−1.972	−2.987**	0.718
	(1.094)	(0.944)	(1.017)	(1.087)	(1.274)	(1.302)	(1.281)
Constant	−2.083***	−1.949***	−2.340***	−1.937***	−1.946***	−2.260***	−1.814***
	(0.157)	(0.148)	(0.170)	(0.101)	(0.127)	(0.166)	(0.134)
Observations	104,556	110,826	102,349	108,795	106,845	96,590	109,683
Log Likelihood	−50,461.720	−49,524.370	−48,143.410	−50,977.470	−51,100.640	−44,614.290	−49,923.300
Akaike Inf. Crit.	100,953.400	99,078.740	96,316.820	101,985.000	102,231.300	89,258.580	99,876.610
Note:							*p<0.1; **p<0.05; ***p<0.01

	coef	std err	t	P$>|t|$	[0.025	0.975]
Twitter 2010 Ln Population	-0.2284	0.107	-2.127	0.045	-0.452	-0.005
Twitter 2010 Population Change %	-184.8047	287.438	-0.643	0.527	-782.564	412.955
Twitter 2019 Ln Population	-0.0846	0.022	-3.842	0.001	-0.129	-0.040
Twitter 2019 Population Change %	-3.0574	4.446	-0.688	0.496	-12.102	5.987
NSDUH Ln Population	-0.0952	0.045	-2.108	0.044	-0.188	-0.003
NSDUH Population Change %	67.5237	130.243	0.518	0.608	-199.266	334.314

Table S8. Depression rates are not associated with year over year population change. Results from ordinary least squares fits with the rate of population change included.