Article 1: Digital Trace Data as Indicators of Social Data

Validating Google Trends for use in Scientific Research

Kelsey Gonzalez

21 March, 2022

Research Methodology

Table 1: New York Air Quality Measurements

Validated.Data.Source	Type	Dates	Google.Trends.Used
Behaviors and Attitudes			
General Social Survey	Cross-Sectional	2010 - 2020	
Vaccine Hesitancy for COVID-19	Cross-Sectional	March 3 – 15, 2021	Search Topics: 'Covid-19 vaccine', 'Coronavirus (Disease)', 'Coronavi
Mask-Wearing Survey Data	Cross-Sectional	July 2 - 14, 2020	Search Topics: 'Coronavirus (Disease)', 'Coronavirus (Virus)', 'Cloth F
Health			
Covid Rates	Longitudinal	Every Monday, 2020 - 2021	'Covid-19', 'Coronavirus', 'Taste Loss', 'Smell Loss'
County Suicide Rates	Longitudinal	Yearly 2010-2020	Search Topics: 'Suicide', 'Depression', Search Term:' Suicide Hotline'
Political			
American National Election Survey	Cross-Sectional	2020	
Presidential Election Results	Cross Sectional	2016 & 2020	Search Topics: 'Hilary Clinton', 'Donald Trump', 'Joe Biden'

Measures

vacc_hes 3148

from 4.99% to 32.33%.

mask (range = 0.10% to

55.80%)

health

covid 0 to a maximum of 1460.46 for each Monday from January 27, 2020 through December 27, 2021. There are 285986 cases across 3136 counties and 101 dates.

suicide There are 34683 total cases, resulting from 34617 observations of 3147 counties. Missing data were interpolated using.

Measures range from 0.034 to 53.254.

political

avaiable for 3147 counties in 2016

and 3118 counties in 2020.

with the lowest percent at 3.09% and the highest at 92.15%.

Table 2: Correlation Results (#t

measure	variable	trend1	trend2	trend3	trend4	tren
Vaccine Hesitancy						
		Covid-19 vaccine	Coronavirus (Disease)	Coronavirus (Virus)	Vaccine	
Pearson's R Correlation	Vaccine Hesitancy	-0.3866	-0.4657	-0.3561	-0.3994	
Mask Attitudes						
		Coronavirus (Disease)	Coronavirus (Virus)	Cloth Face Mask	Mask	Civil a
Pearson's R Correlation	Mask Rare	-0.1398	-0.1343	0.146	0.1339	-0.117

measure	variable	trend1	trend2	trend3	trend4	tren
Covid Rates						
		covid_19	smell_loss	taste_loss		
Pearson's R Correlation	Covid Rate	-0.1012	0.2835	0.2851		
repeated measures correlation coefficient	Covid Rate	-0.099***	0.322***	0.313***		
Suicide Rates						
		suicide	depression	suicide_hotline		
Pearson's R Correlation	Suicide Rate	0.0845	0.0273	0.0618		
repeated measures correlation coefficient	Suicide Rate	0.055***	0.142***	0.131***		
2016 Presidential Votes						
		Hilary Clinton	Donald Trump			
Pearson's R Correlation	2016 Votes for Clinton	-0.1663	0.1663			
Pearson's R Correlation	2016 Votes for Trump	0.1682	-0.1682			
2020 Presidential Votes						
		Joe Biden	Donald Trump			
Pearson's R Correlation	2020 Votes for Biden	-0.0286	0.0286			
Pearson's R Correlation	2020 Votes for Trump	0.0356	-0.0356			

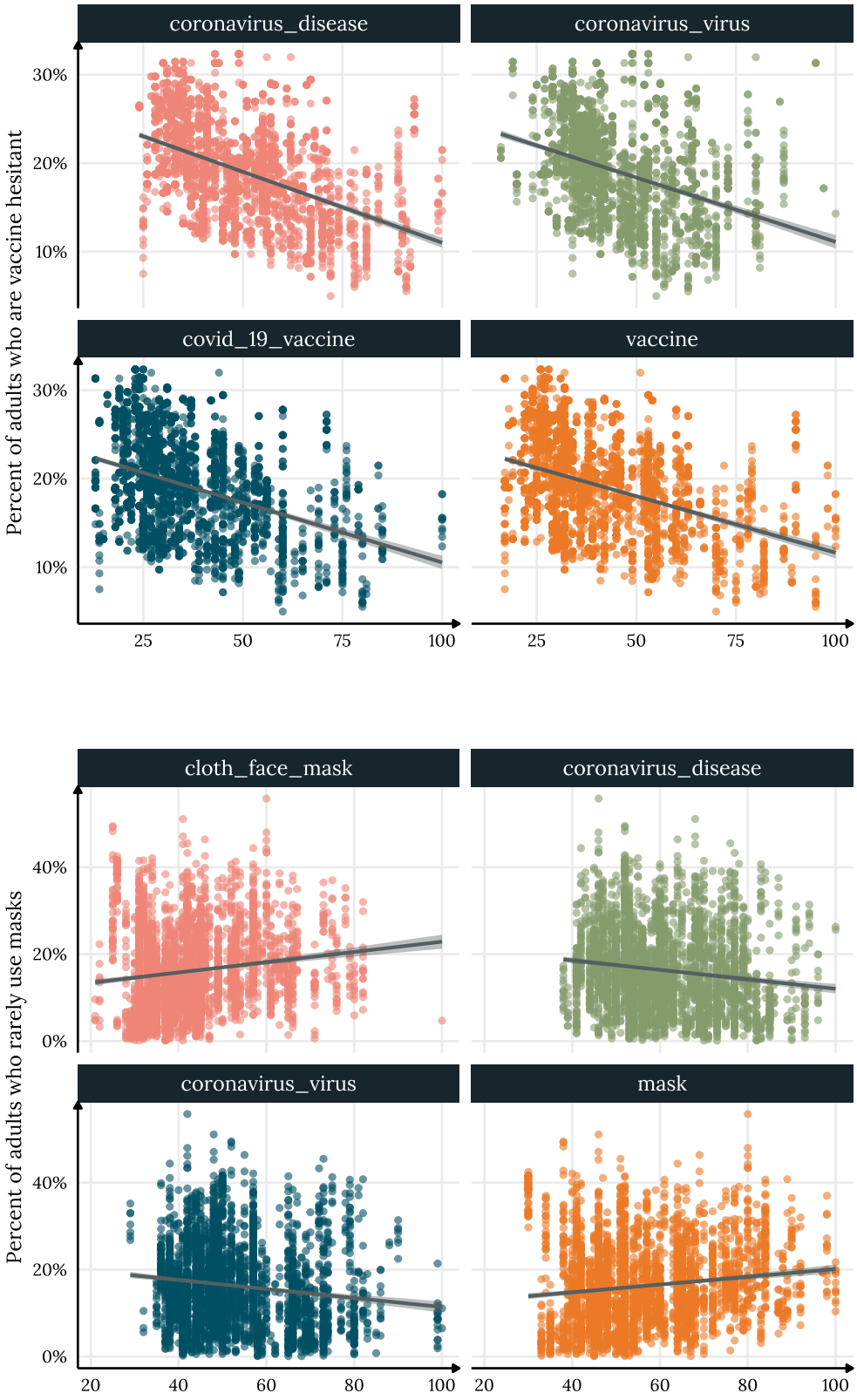

coronavirus_disease
coronavirus_virus
covid_19_vaccine
vaccine
Percent of adults who are vaccine hesitant
30%
20%
10%
25
50
75
100
cloth_face_mask
coronavirus_disease
coronavirus_virus
mask
Percent of adults who rarely use masks
40%
20%
0%
20
40
60
80
100

Table 3: Linear Regression Results for Vaccine Hesitancy

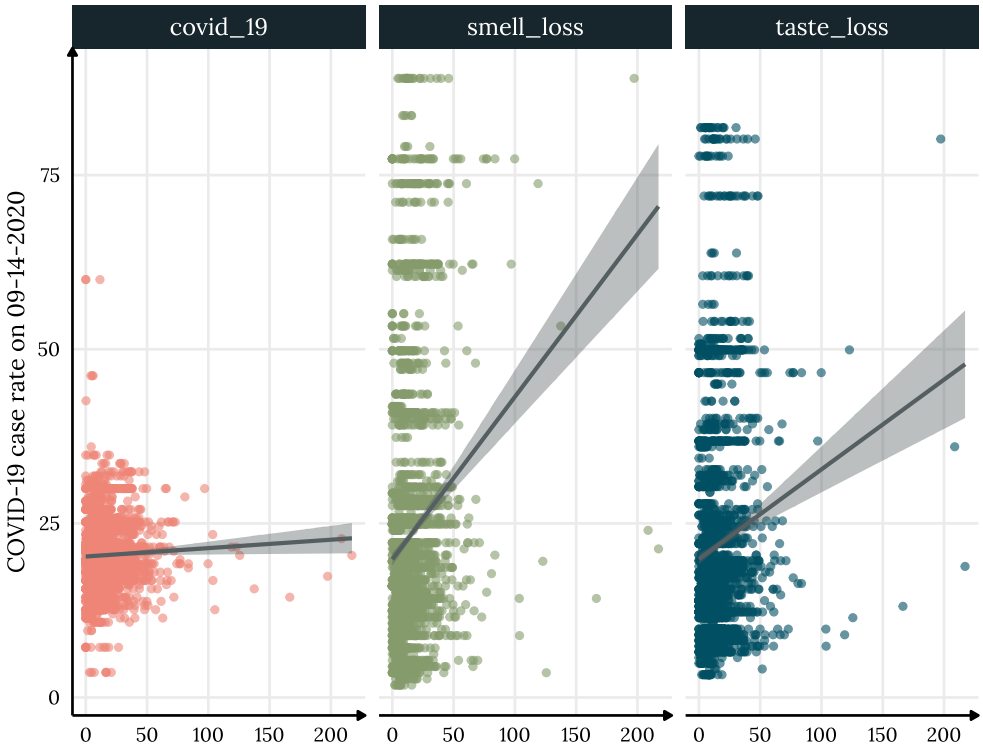
	Model 1	Model 2	Model 3
(Intercept)	0.191***	0.191***	0.191***
	(0.001)	(0.001)	(0.001)
covid_19_vaccine	−0.006+		−0.011**
	(0.004)		(0.003)
vaccine	0.021***		0.023***
	(0.005)		(0.005)
coronavirus_disease	−0.037***		−0.027***
	(0.003)		(0.003)
coronavirus_virus	−0.003*		0.000
	(0.001)		(0.001)
total_pop		−0.005***	−0.004***
		(0.001)	(0.001)
pop_density		−0.004***	−0.003***
		(0.001)	(0.001)
unemployment_rate		−0.002*	0.000
		(0.001)	(0.001)
over_65		−0.013***	−0.010***
		(0.001)	(0.001)
poverty_rate		0.005**	0.004**
		(0.002)	(0.001)
median_income		−0.021***	−0.016***
		(0.001)	(0.001)
broadband		−0.008***	−0.005***
		(0.001)	(0.001)
Num.Obs.	3148	3133	3133
R2	0.233	0.350	0.411
R2 Adj.	0.232	0.349	0.409
Log.Lik.	5163.343	5405.127	5557.567
F	238.587	240.876	197.733

* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed test).

Table 4: Linear Regression Results for Rare Mask Usage

	Model 1	Model 2	Model 3
(Intercept)	0.162***	0.163***	0.162***
	(0.002)	(0.002)	(0.002)
coronavirus_disease	−0.011***		−0.012***
	(0.003)		(0.003)
coronavirus_virus	−0.007*		0.003
	(0.003)		(0.003)
cloth_face_mask	−0.003		−0.008*
	(0.003)		(0.003)
mask	0.019***		0.016***
	(0.003)		(0.003)
total_pop		−0.012***	−0.011***
		(0.002)	(0.002)
pop_density		−0.003+	−0.003*
		(0.002)	(0.002)
unemployment_rate		−0.028***	−0.027***
		(0.002)	(0.002)
over_65		−0.006**	−0.005*
		(0.002)	(0.002)
poverty_rate		−0.010**	−0.010**
		(0.003)	(0.003)
median_income		−0.036***	−0.033***
		(0.003)	(0.003)
broadband		−0.006*	−0.006*
		(0.003)	(0.003)
Num.Obs.	3136	3133	3121
R2	0.049	0.176	0.189
R2 Adj.	0.047	0.174	0.186
Log.Lik.	2970.506	3184.261	3202.189
F	40.072	95.547	65.970

* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed test).



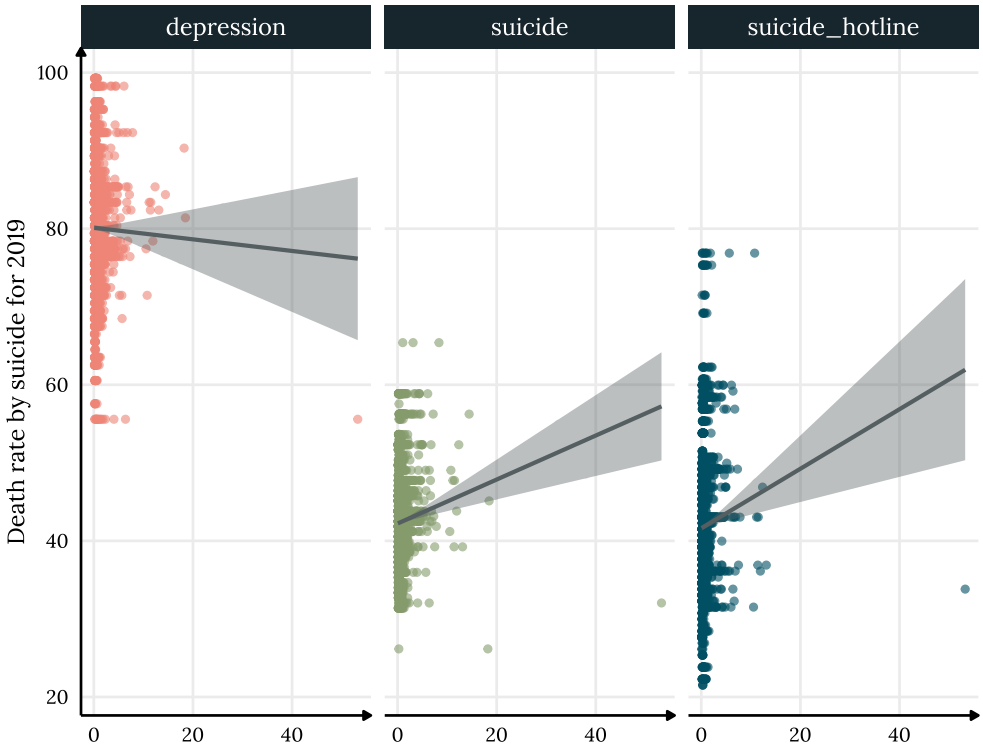
Random effect variances not available. Returned R2 does not account for random effects. Random effect variances not available. Returned R2 does not account for random effects.

Table 5: Hierarchical Model for Covid Case Rates

	Model 1	Model 2	Model 3
(Intercept)	−532.306***	−712.023***	−534.608***
	(8.271)	(6.009)	(8.268)
covid_19	1.031***		1.024***
	(0.082)		(0.082)
smell_loss	7.373***		7.374***
	(0.082)		(0.082)
taste_loss	6.344***		6.347***
	(0.078)		(0.078)
covid_rate_fips_mean	0.813***	0.982***	0.879***
	(0.019)	(0.010)	(0.020)
date	0.029***	0.038***	0.029***
	(0.000)	(0.000)	(0.000)
SD (Intercept)	5.196	0.002	4.984
SD (Observations)	29.478	31.736	29.441
total_pop		0.064	0.409**
		(0.066)	(0.140)
pop_density		0.010	0.463***
		(0.067)	(0.123)
unemployment_rate		0.049	0.725***
		(0.080)	(0.155)
over_65		−0.116	−0.055
		(0.074)	(0.139)
poverty_rate		0.054	0.351
		(0.118)	(0.221)
median_income		0.031	0.907***
		(0.114)	(0.214)
broadband		0.073	0.437*
		(0.097)	(0.182)
Num.Obs.	248 346	284 666	247 655
R2 Marg.	0.201	0.081	0.200
R2 Cond.		0.081	

Random intercept per county

* p < .05. ** p < .01. *** p < .001 (two-tailed test).



Random effect variances not available. Returned R2 does not account for random effects. Random effect variances not available. Returned R2 does not account for random effects. Random effect variances not available. Returned R2 does not account for random effects.

Table 6: Hierarchical Model for Suicide Rates

	Model 1	Model 2	Model 3
(Intercept)	−7.364***	−10.046***	−9.683***
	(0.359)	(0.342)	(0.454)
suicide	−0.001		−0.001
	(0.001)		(0.001)
depression	−0.001+		−0.001
	(0.001)		(0.001)
suicide_hotline	0.002***		0.002***
	(0.001)		(0.001)
year	0.004***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)
SD (Intercept)	1.460	1.587	1.459
SD (Observations)	0.050	0.051	0.050
total_pop		−0.032***	−0.027**
		(0.010)	(0.009)
pop_density		−0.037**	−0.033**
		(0.011)	(0.011)
unemployment_rate		−0.001	−0.001
		(0.001)	(0.001)
over_65		−0.010***	−0.014***
		(0.002)	(0.002)
poverty_rate		−0.005***	−0.003**
		(0.001)	(0.001)
median_income		−0.014***	−0.015***
		(0.002)	(0.002)
Num.Obs.	30 008	33 582	30 001
R2 Marg.	0.058	0.582	0.564
R2 Cond.			

Random intercept per county

* p < .05. ** p < .01. *** p < .001 (two-tailed test).

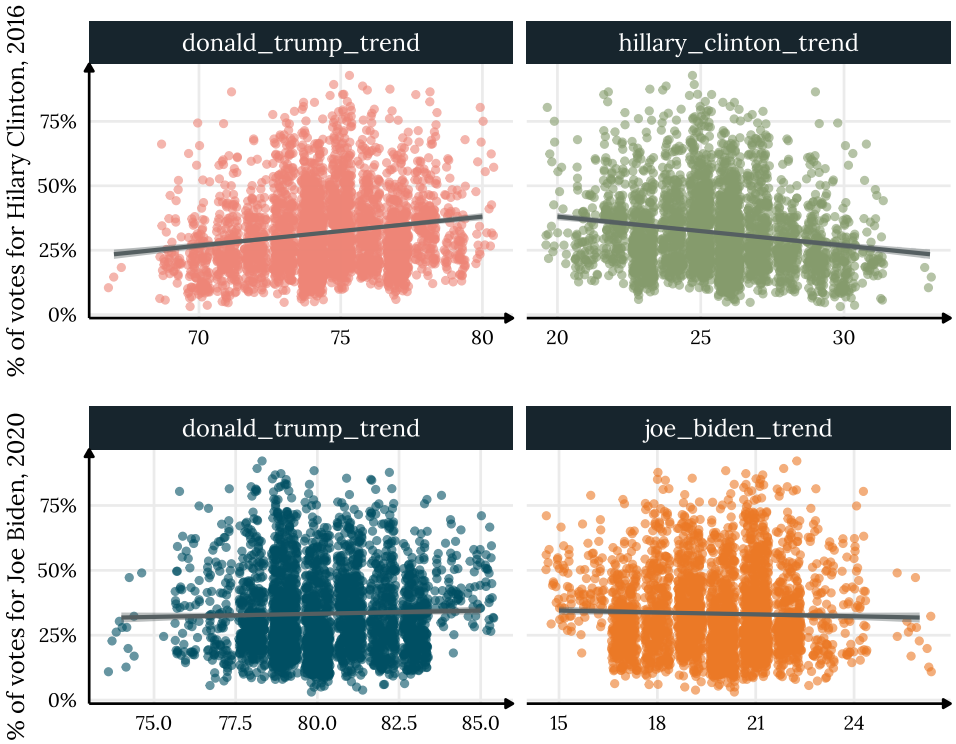

donald_trump_trend
hillary_clinton_trend
% of votes for Hilary Clinton, 2016
75%
50%
25%
0%
70
75
80
20
25
30
donald_trump_trend
joe_biden_trend
% of votes for Joe Biden, 2020
75%
50%
25%
0%
75.0
77.5
80.0
82.5
85.0
15
18
21
24

Table 7: Linear Regression Results for 2016 Presidential Election Results (Hilary Clinton Shown)

	Model 1	Model 2	Model 3
(Intercept)	0.318***	0.318***	0.318***
	(0.003)	(0.002)	(0.002)
hillary_clinton_trend	−0.025***		−0.019***
	(0.003)		(0.002)
total_pop		0.028***	0.028***
		(0.002)	(0.002)
pop_density		0.022***	0.022***
		(0.002)	(0.002)
unemployment_rate		0.047***	0.045***
		(0.003)	(0.003)
over_65		−0.012***	−0.014***
		(0.003)	(0.003)
poverty_rate		0.034***	0.034***
		(0.004)	(0.004)
median_income		0.054***	0.052***
		(0.004)	(0.004)
Num.Obs.	3147	3145	3145
R2	0.028	0.314	0.329
R2 Adj.	0.027	0.313	0.327
Log.Lik.	1483.493	2037.065	2071.025

Results predicting Donald J. Trump percentage largely equivalent and available upon request.

* $p < .05$. ** $p < .01$. *** $p < .001$ (two-tailed test).

Table 8: Linear Regression Results for 2020 Presidential Election Results (Joe Biden Shown)

	Model 1	Model 2	Model 3
(Intercept)	0.333***	0.333***	0.333***
	(0.003)	(0.002)	(0.002)
joe_biden_trend	−0.005		−0.014***
	(0.003)		(0.002)
total_pop		0.030***	0.029***
		(0.003)	(0.003)
pop_density		0.021***	0.022***
		(0.003)	(0.003)
unemployment_rate		0.044***	0.044***
		(0.003)	(0.003)
over_65		−0.007*	−0.008**
		(0.003)	(0.003)
poverty_rate		0.032***	0.033***
		(0.004)	(0.004)
median_income		0.069***	0.070***
		(0.004)	(0.004)
Num.Obs.	3118	3118	3118
R2	0.001	0.295	0.303
R2 Adj.	0.000	0.294	0.301
Log.Lik.	1291.161	1835.433	1852.002

Results predicting Donald J. Trump percentage largely equivalent and available upon request.
* p < .05. ** p < .01. *** p < .001 (two-tailed test).