Supplementary Online File 1: A comparison of COVID-19 outbreaks across US Combined Statistical Areas using new methods for estimating R_0 and social distancing behaviour

Ludovica Luisa Vissat [a], Nir Horvitz [a], Rachael Phillips [b], Zhongqi Miao [a], Whitney Mgbara [a], Yue You [b], Richard Salter [c], Alan Hubbard [b], Wayne M. Getz [*,a,b,d]

[a] Department of Environmental Science, Policy, and Management, UC Berkeley, CA 94720, USA

[b] Division Environmental Health Sciences, UC Berkeley, CA 94720, USA

[c] Computer Science Department, Oberlin College, Oberlin, Ohio, OH 44074, USA

[d] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban 4000, South Africa

[*] Corresponding author: Wayne M. Getz, wgetz@berkeley.edu

Contents

1 Introduction

In this supplementary online file we report the two criteria for the choice of CSAs of interest (Section 2) and for day 0 of the epidemic outbreak (Section 3). We then present the final choice of CSAs (Section 4) and some considerations on the calculation of R_0 (Section 5). We describe the choice of parameters for the NMB-DASA web app (Section 6) followed by the clustering procedure to extract weaker/stronger social distancing CSA responses and some additional clustering analysis (Section 7).

2 Criterion: choice of CSAs

As explained in the paper, we ranked the CSA by considering their total population, the number of cases and the number of deaths for each CSA, extracting data related to the COVID-19 epidemic from January 21, 2020 (which is the date of first case in the US) until September 5, 2020 (data downloaded from https://usafacts.org/). In particular, we ranked every CSA by population, by number of cases per 1000 people and by number of deaths per cases. We sum these 3 rankings and used the total rank to choose the CSAs of interest. By looking at this total ranking values, we chose the first 32 CSAs (by lowest ranking), given a gap in the ranking between these first CSAs and the following ones (140 CSAs), as shown in Figure 1 (total rank in the last column).

CSA_Code	CSA_Name	Population	Cases	Deaths	Cases per 1,000 people	Deaths per cases (%)	Rank Population	Rank Cases	Rank Death	Total rank
408	New York-Newark, NY-NJ-CT-PA	22589036	595839	47272	26.3773540402521	7.9%	1	26	2.0	29.0
406	New Orleans-Metairie-Hammond, LA-M	1507017	46757	1811	31.026192803399	3.9%	37	12	24.0	73.0
176	Chicago-Naperville, IL-IN-WI	9825325	207864	7527	21.1559414065184	3.6%	3	53	31.0	87.0
148	Boston-Worcester-Providence, MA-RI-NI	8287710	145451	9528	17.5502038560712	6.6%	6	76	6.0	88.0
365	McAllen-Edinburg, TX	933340	30690	1301	32.8819079863715	4.2%	61	9	19.0	89.0
429	Phoenix-Mesa, AZ	5002221	145975	3229	29.182037339014	2.2%	13	15	72.0	100.0
428	Philadelphia-Reading-Camden, PA-NJ-D	7209620	111602	7006	15.4795953184773	6.3%	9	86	7.0	102.0
370	Miami-Port St. Lucie-Fort Lauderdale, F	6889936	291586	5359	42.3205672737744	1.8%	10	3	96.0	109.0
220	Detroit-Warren-Ann Arbor, MI	5341994	74600	5610	13.9648228732567	7.5%	12	98	4.0	114.0
548	Washington-Baltimore-Arlington, DC-MI	9814928	175048	5481	17.834873572175	3.1%	4	73	40.0	117.0
154	Brownsville-Harlingen-Raymondville, T	444521	22233	817	50.0156348069045	3.7%	94	2	30.0	126.0
348	Los Angeles-Long Beach, CA	18711436	406782	8777	21.739753164856	2.2%	2	51	78.0	131.0
294	Indianapolis-Carmel-Muncie, IN	2457286	37931	1657	15.436135639075	4.4%	28	87	17.0	132.0
484	San Antonio-New Braunfels-Pearsall, TX	2571266	54565	1392	21.2210638650377	2.6%	25	52	56.0	133.0
508	Shreveport-Bossier City-Minden, LA	433046	12168	469	28.0986315541536	3.9%	95	19	25.0	139.0
122	Atlanta--Athens-Clarke County--Sandy S	6853392	155277	3094	22.6569558548526	2.0%	11	43	86.0	140.0
298	Jackson-Vicksburg-Brookhaven, MS	674340	18899	530	28.0259216419017	2.8%	74	20	47.0	141.0
476	St. Louis-St. Charles-Farmington, MO-IL	2907648	49406	1414	16.99174040324	2.9%	20	79	44.0	143.0
318	Lafayette-Opelousas-Morgan City, LA	620679	22727	578	36.616350802911	2.5%	81	5	57.0	143.0
288	Houston-The Woodlands, TX	7253193	162538	3135	22.4091651773226	1.9%	8	46	90.0	144.0
332	Las Vegas-Henderson, NV	2313238	60329	1171	26.0798932059736	1.9%	29	27	89.0	145.0
278	Hartford-East Hartford, CT	1470083	17823	1790	12.1238052545332	10.0%	41	113	1.0	155.0
536	Tucson-Nogales, AZ	1093777	24204	631	22.128825162716	2.6%	54	49	52.0	155.0
163	Cape Coral-Fort Myers-Naples, FL	1197501	32796	650	27.387033497258	2.0%	48	22	87.0	157.0
378	Minneapolis-St. Paul, MN-WI	4027861	60723	1588	15.0757436763582	2.6%	16	90	51.0	157.0
273	Greenville-Spartanburg-Anderson, SC	1475235	28989	751	19.6504285757862	2.6%	40	62	55.0	157.0
388	Montgomery-Selma-Alexander City, AL	461516	15307	371	33.1667807833315	2.4%	91	8	62.0	161.0
412	North Port-Sarasota, FL	1063906	22109	609	20.7809712512196	2.8%	56	56	49.0	161.0
216	Denver-Aurora, CO	3617927	41398	1556	11.4424641514326	3.8%	17	119	27.0	163.0
204	Corpus Christi-Kingsville-Alice, TX	535257	18817	422	35.1550750387197	2.2%	86	7	71.0	164.0
192	Columbia-Orangeburg-Newberry, SC	963048	24642	539	25.5875096568395	2.2%	60	31	73.0	164.0
422	Orlando-Lakeland-Deltona, FL	4160646	94297	1534	22.6640286147872	1.6%	15	42	107.0	164.0
184	Cleveland-Akron-Canton, OH	3586918	32877	1417	9.16580752612689	4.3%	18	138	18.0	174.0
279	Hattiesburg-Laurel, MS	253330	7428	210	29.3214384399795	2.8%	121	14	45.0	180.0
384	Monroe-Ruston, LA	247003	7967	225	32.2546689716319	2.8%	123	11	46.0	180.0
368	Memphis-Forrest City, TN-MS-AR	1371039	40006	561	29.1793304202142	1.4%	43	16	125.0	184.0
324	Lake Charles-Jennings, LA	241777	8815	229	36.4592165507885	2.6%	125	6	53.0	184.0

Figure 1: CSA ranking and choice

3 Criterion: choice of day 0

We present now the algorithm (Algorithm 1) which implements the criterion to choose "day 0" (starting date of the exponential phase of community transmission). The output will be the ideal "day 0" to start the fit.

- Line 1: we require the CSA incidence data D as input.
- Line 2-3: As first steps, we calculate the 7-day lagged moving average of the incidence data (line 2), using the function M_7, and we round its value (line 3). The function M_7 calculates the 7-day lagged moving average:

$$m(i) = \frac{1}{7} \sum_{i}^{i+6} D(i)$$

  for $i = 0, ...,$ length(D)-6. Note that we round the values to have a curve that resembled a step function at the beginning, when cases were relatively small, to have a bigger error in the fit of the exponential curve. There were cases where the best fit was given by the very beginning of the time series, with a very few cases at the end, which was not ideal. With

Algorithm 1: "Day 0" criterion implementation

Input D
$m = M_7(D)$
$r = \text{round}(m)$
$v_{\text{ex}} = v_{\text{e}} = v_{\text{d}} = c()$
$j = \min\{i : (r[i] > 0)\}$
$k = j + 15$
for i in $j : k$ **do**
    $b = i$
    $e = i + 14$
    $l = lm(\log(r[b : e]) \sim c(1 : 15))$
    $v_{\text{ex}} = c(v_{\text{ex}}, \alpha)$ # α: coeff(l)[2]
    $\epsilon = \sum_{h=1}^{15} |\text{rl[h]}|$
    $v_{\text{e}} = c(v_{\text{e}}, \epsilon)$
    $v_{\text{d}} = c(v_{\text{d}}, i)$
if $\max(v_{\text{ex}}) > 0.25$ **then**
    $\alpha_{\text{m}} = 0.25$
else if $\max(v_{\text{ex}}) > 0.2$ **then**
    $\alpha_{\text{m}} = 0.2$
else
    return NA
$\text{ind} = \text{which}(v_{\text{e}} = \min(v_{\text{e}}))$ # i.e. i : $v_{\text{e}}[i] = \min(v_{\text{e}})$
while $v_{\text{ex}}[\text{ind}] < \alpha_m$ **do**
    $v_{\text{e}} = v_{\text{e}}[-\text{ind}]$
    $v_{\text{ex}} = v_{\text{ex}}[-\text{ind}]$
    $v_{\text{d}} = v_{\text{d}}[-\text{ind}]$
    $\text{ind} = \text{which}(v_{\text{e}} = \min(v_{\text{e}}))$
return $v_{\text{d}}[\text{ind}]$

the round option, we got better ranges, with higher number of cases and a good fit of the exponential curve.

- Line 4: we define 3 empty vectors v_{ex}, v_{e} and v_{d} that will be used to collect the fit parameters, the error of the fit and the corresponding initial day of the range, respectively.

- Line 5-6: to select which days to consider to search for an appropriate 15-day range, we start with the first day for which the round number of the 7-day lagged moving average r is above 0, and try from that day, for 16 consecutive days.

- Line 7-14: in this for loop, we fit a linear model to the logarithm of r for different 15-day ranges and we collect the fit parameter α, the error of the fit ϵ (residuals) and the corresponding initial day for each considered range. The output of the fit gives the coefficients β and α (fit: $\log(\beta) + \alpha x$). If we consider r instead of its logarithm, we can write the function: $\beta e^{\alpha x}$. Note that we use a syntax similar to the programming language R, using the function lm to indicate the fitting procedure and the function $c()$ to indicate the concatenation of two vectors (in this case we expand the existing vectors with additional elements).

- Line 15-20: depending on the maximum value of v_{ex} (vector of coefficients α), we assign different values to α_{m}, which will be used in the while loop in line 22-26. If the maximum value results below 0.2, we return NA. We chose these values to have a threshold on the exponential growth, to capture a sufficiently vigorous epidemics near the start of the COVID-19 pandemic.[1]

- Line 21: we select the index which corresponds to the minimum error.

- Line 22-26: we look for the index corresponding to the lowest error and a coefficient α above α_{m}. While the condition on the coefficient α is not satisfied, we keep searching for the desired index by removing the already considered elements in the vectors v_{e}, v_{ex} and v_{d}. The selected "day 0" will be given by $v_{\mathrm{d}}[\mathrm{ind}]$.

4 Chosen CSAs

Given the "day 0" criterion, the final choice of the CSAs presents 29 CSAs. Using this criterion we excluded 3 CSAs which did not present a sufficiently high coefficient α to satisfy our criterion: Corpus Christi-Kingsville-Alice, TX, McAllen-Edinburg, TX and Montgomery-Selma-Alexander City, AL.

The extracted dates are presented and shown in Table 1 (number of days after the first day of data collection - 21 Jan 2020: 41, 49, 56, 50, 40, 43, NA, 41, 50, 50, 49, 48, 47, 46, 52, 43, 44, NA, 48, 44, NA, 45, 42, 48, 47, 46, 46, 50, 48, 47, 49, 40)

[1] Note that for the calculation of the growth rate and therefore of R_0 we used the output of our SCLAIV+D model simulation. Therefore, the extracted growth rate did not have necessarily to respect the threshold for the "day 0" choice, as we observe for the Tucson-Nogales, AZ results, with extracted growth rate $r = 0.12$

List number	CSA	Chosen day 0
1	Atlanta-Athens-Clarke County-Sandy Springs, GA-AL	1 March
2	Boston-Worcester-Providence, MA-RI-NH-CT	9 March
3	Brownsville-Harlingen-Raymondville, TX	16 March
4	Cape Coral-Fort Myers-Naples, FL	10 March
5	Chicago-Naperville, IL-IN-WI	29 February
6	Columbia-Orangeburg-Newberry, SC	3 March
7	Corpus Christi-Kingsville-Alice, TX	-
8	Denver-Aurora, CO	1 March
9	Detroit-Warren-Ann Arbor, MI	10 March
10	Greenville-Spartanburg-Anderson, SC	10 March
11	Hartford-East Hartford, CT	9 March
12	Houston-The Woodlands, TX	8 March
13	Indianapolis-Carmel-Muncie, IN	7 March
14	Jackson-Vicksburg-Brookhaven, MS	6 March
15	Lafayette-Opelousas-Morgan City, LA	12 March
16	Las Vegas-Henderson, NV	3 March
17	Los Angeles-Long Beach, CA	4 March
18	McAllen-Edinburg, TX	-
19	Miami-Port St. Lucie-Fort Lauderdale, FL	8 March
20	Minneapolis-St. Paul, MN-WI	4 March
21	Montgomery-Selma-Alexander City, AL	-
22	New Orleans-Metairie-Hammond, LA-MS	5 March
23	New York-Newark, NY-NJ-CT-PA	2 March
24	North Port-Sarasota, FL	8 March
25	Orlando-Lakeland-Deltona, FL	7 March
26	Philadelphia-Reading-Camden, PA-NJ-DE-MD	6 March
27	Phoenix-Mesa, AZ	6 March
28	San Antonio-New Braunfels-Pearsall, TX	10 March
29	Shreveport-Bossier City-Minden, LA	8 March
30	St. Louis-St. Charles-Farmington, MO-IL	7 March
31	Tucson-Nogales, AZ	9 March
32	Washington-Baltimore-Arlington, DC-MD-VA-WV-PA	29 February

Table 1: Chosen “day 0” (starting date of the exponential phase of community transmission) for each CSA.

5 R_0 calculation

Table 1 in the main paper reports the values for R_0 for each CSA considered in the analysis, given the generation time $\mathcal{T}$, equal to 5.40, as reported in (Rai et al., 2020), and the different epidemic growth rates r, as shown in Figure 2.

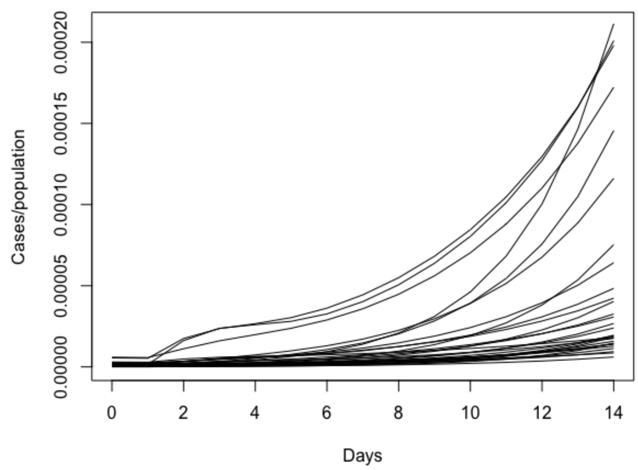


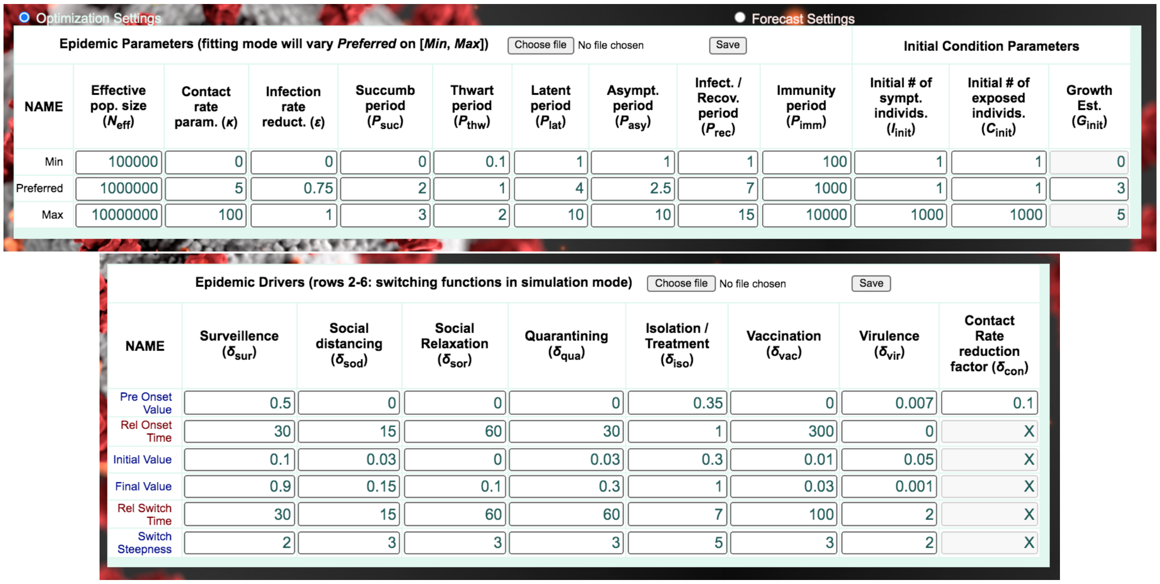
Figure 2: Exponential phase for each CSA. Population data downloaded from `data.census.gov`.

6 NMB-DASA web app: settings

In Figure 3 we provide the screenshots from the NMB-DASA web app to show the chosen parameters for the SCLAIV+D model and its fitting procedure. In particular, we chose:

- Infection rate reduction (ϵ): 0.75 (`https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html#table-1`, under "Infectiousness of asymptomatic individuals relative to symptomatic", Scenario 5)
- Latent period (P_{lat}): 4 (Getz et al., 2021)
- Asympt. period (P_{asy}): 2.5 (A person with COVID-19 may be contagious 48 to 72 hours (2 to 3 days) before starting to experience symptoms. (`https://www.health.harvard.edu/diseases-and-conditions/covid-19-basics`))
- Infectious period (P_{rec}): 7 (Getz et al., 2021)

We provide a separate supplementary online file SOF3 with the results of the Web App fits.



Optimization Settings | Forecast Settings

Epidemic Parameters (fitting mode will vary *Preferred* on [*Min*, *Max*]) | Choose file | No file chosen | Save | Initial Condition Parameters

NAME	Effective pop. size (N_{eff})	Contact rate param. (κ)	Infection rate reduct. (ε)	Succumb period (P_{suc})	Thwart period (P_{thw})	Latent period (P_{lat})	Asympt. period (P_{asy})	Infect. / Recov. period (P_{rec})	Immunity period (P_{imm})	Initial # of sympt. individs. (I_{init})	Initial # of exposed individs. (C_{init})	Growth Est. (G_{init})
Min	100000	0	0	0	0.1	1	1	1	100	1	1	0
Preferred	1000000	5	0.75	2	1	4	2.5	7	1000	1	1	3
Max	10000000	100	1	3	2	10	10	15	10000	1000	1000	5

Epidemic Drivers (rows 2-6: switching functions in simulation mode) | Choose file | No file chosen | Save

NAME	Surveillence (δ_{sur})	Social distancing (δ_{sod})	Social Relaxation (δ_{sor})	Quarantining (δ_{qua})	Isolation / Treatment (δ_{iso})	Vaccination (δ_{vac})	Virulence (δ_{vir})	Contact Rate reduction factor (δ_{con})
Pre Onset Value	0.5	0	0	0	0.35	0	0.007	0.1
Rel Onset Time	30	15	60	30	1	300	0	X
Initial Value	0.1	0.03	0	0.03	0.3	0.01	0.05	X
Final Value	0.9	0.15	0.1	0.3	1	0.03	0.001	X
Rel Switch Time	30	15	60	60	7	100	2	X
Switch Steepness	2	3	3	3	5	3	2	X

Figure 3: Screenshot from the NMB-DASA web app showing the chosen parameters (Getz et al., 2021)

7 Clustering

In Algorithm 2 we present the calculation of the area under the curve, using the trapezoidal integration rule. The area under the curve is evaluated by dividing the total area into little trapezoids and summing their areas.

This procedure will return the list area20 (line 2-4) which will contain 20 vectors of 29 entries, showing the cumulative value of the area under the curve for each CSA. We start with a vector of 29 entries, all equal to 0 (line 5). Given the list ld of the $c_{\text{flatten}}(t)$ index for each CSA (line 8-9), we evaluate the area under the curve for each day starting for the last one corresponding to a $c_{\text{flatten}}(t)$ index equal to 0 (line 10-11). We evaluate the area under the curve for every single day (considering two consecutive days for constructing the trapezoid, line 12-13). In the area calculation (line 14), diff(x[id]) is always equal to 1 since we consider consecutive days, while rollmean(y[id],2) calculates the mean among the two considered values of the data (y), for the given id (a rolling 2-day window). The area under the curve (AUC) is calculated and summed with the previous values (line 15), returning a cumulative evaluation for each curve (line 16) which presents the value of the area for each of the 20-day range, for each CSA.

We use these values (presented in Table 2) to cluster the different CSAs. In particular, we considered the area values for day 5, 10, 15 and 20, and used the elbow method to extract the number of clusters (as shown in Figure 4). With this analysis, we extracted two clusters. The 20-day range considered for each CSA, by group, is also shown in Figure 4.

Additional clustering analysis

In this subsection we discuss some additional clustering analysis, considering area features and R_0, to provide further information. In Figure 5 (left) we report the results for a choice of 3 clusters, considering the 4 area features. In Figure 5 (right) we report the 20-day range of interest, coloured

Algorithm 2: c_{flatten} area calculation implementation

```
Input ld (list of c_flatten(t) for each CSA)
area20 = list()
for i in 1:20 do
    area20[[i]] = vector()
area = rep(0, 29)
for i in 1:20 do
    for j in 1:29 do
        y = ld[[j]]
        x = 1:length(y)
        start = min(which(y > 0))
        start = start - 1
        k = i-1
        id = c(start+k, start+i)
        AUC = sum(diff(x[id])*rollmean(y[id],2))
        area[j] = area[j] + AUC
    area20[[i]] = area
return area20
```

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
Atlanta	$2.76 \cdot 10^{-7}$	$9.83 \cdot 10^{-5}$	$5.03 \cdot 10^{-4}$	$1.46 \cdot 10^{-3}$	$3.38 \cdot 10^{-3}$	$6.97 \cdot 10^{-3}$	$1.42 \cdot 10^{-2}$	$3.43 \cdot 10^{-2}$	$9.37 \cdot 10^{-2}$	0.21	0.38	0.61	0.9	1.24	1.64	2.09	2.59	3.14	3.75	4.4
Boston	$5.34 \cdot 10^{-2}$	0.25	0.64	1.17	1.79	2.47	3.19	3.94	4.7	5.48	6.28	7.08	7.9	8.72	9.55	10.39	11.24	12.09	12.96	13.83
Brownsville	$3.22 \cdot 10^{-3}$	$1.89 \cdot 10^{-2}$	$6.51 \cdot 10^{-2}$	0.18	0.4	0.7	1.08	1.53	2.04	2.61	3.24	3.91	4.63	5.39	6.18	7.01	7.87	8.75	9.65	10.56
Cape Coral	$2.7 \cdot 10^{-2}$	0.13	0.43	0.91	1.46	2.06	2.72	3.42	4.16	4.93	5.72	6.54	7.38	8.23	9.11	9.99	10.89	11.81	12.73	13.66
Chicago	$2.59 \cdot 10^{-2}$	0.15	0.46	0.91	1.44	2.02	2.65	3.33	4.04	4.79	5.57	6.37	7.2	8.04	8.91	9.79	10.69	11.6	12.53	13.46
Columbia	$1.79 \cdot 10^{-2}$	$8.98 \cdot 10^{-2}$	0.3	0.69	1.17	1.69	2.25	2.84	3.46	4.1	4.77	5.45	6.16	6.89	7.64	8.4	9.18	9.97	10.78	11.6
Denver	$5.95 \cdot 10^{-35}$	$2.05 \cdot 10^{-6}$	$2.26 \cdot 10^{-5}$	$1.15 \cdot 10^{-4}$	$3.92 \cdot 10^{-4}$	$1.05 \cdot 10^{-3}$	$2.39 \cdot 10^{-3}$	$4.83 \cdot 10^{-3}$	$8.9 \cdot 10^{-3}$	$1.52 \cdot 10^{-2}$	$3.29 \cdot 10^{-2}$	$8.11 \cdot 10^{-2}$	0.19	0.4	0.71	1.11	1.58	2.12	2.73	3.38
Detroit	$8.88 \cdot 10^{-4}$	$6.67 \cdot 10^{-3}$	$2.03 \cdot 10^{-2}$	$4.15 \cdot 10^{-2}$	$7.3 \cdot 10^{-2}$	0.12	0.19	0.28	0.42	0.62	0.9	1.32	1.85	2.48	3.19	3.96	4.78	5.64	6.52	7.42
Greenville	$5.08 \cdot 10^{-3}$	$2.14 \cdot 10^{-2}$	$5.11 \cdot 10^{-2}$	$9.46 \cdot 10^{-2}$	0.15	0.22	0.33	0.48	0.68	0.96	1.32	1.76	2.3	2.92	3.58	4.28	5	5.75	6.52	7.31
Hartford	$1.14 \cdot 10^{-3}$	$4.53 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$2.04 \cdot 10^{-2}$	$4.03 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	0.12	0.18	0.27	0.41	0.62	0.89	1.25	1.71	2.27	2.93	3.64	4.4	5.2	6.03
Houston	$1.46 \cdot 10^{-2}$	$7.47 \cdot 10^{-2}$	0.2	0.43	0.8	1.3	1.88	2.52	3.21	3.93	4.69	5.47	6.28	7.1	7.94	8.8	9.67	10.54	11.43	12.33
Indianapolis	$2.45 \cdot 10^{-2}$	0.16	0.45	0.88	1.41	2.03	2.72	3.46	4.24	5.06	5.91	6.79	7.68	8.59	9.52	10.45	11.4	12.35	13.31	14.27
Jackson	$3.54 \cdot 10^{-3}$	$4.24 \cdot 10^{-2}$	0.16	0.37	0.7	1.12	1.6	2.14	2.72	3.35	4.02	4.72	5.45	6.2	6.98	7.78	8.59	9.42	10.27	11.12
Lafayette	$7.34 \cdot 10^{-3}$	$9.57 \cdot 10^{-2}$	0.48	1.18	2.01	2.9	3.82	4.75	5.7	6.66	7.62	8.59	9.55	10.53	11.5	12.47	13.45	14.43	15.41	16.39
Las Vegas	$1.16 \cdot 10^{-3}$	$6.05 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$	$3.55 \cdot 10^{-2}$	$7.99 \cdot 10^{-2}$	0.17	0.31	0.53	0.88	1.37	1.97	2.62	3.33	4.07	4.84	5.65	6.48	7.34	8.21	9.1
Los Angeles	$6.51 \cdot 10^{-4}$	$3.64 \cdot 10^{-3}$	$9.99 \cdot 10^{-3}$	$2.49 \cdot 10^{-2}$	$5.89 \cdot 10^{-2}$	0.13	0.3	0.57	0.95	1.41	1.94	2.54	3.2	3.9	4.65	5.44	6.26	7.11	7.99	8.88
Miami	$1.96 \cdot 10^{-8}$	$2.58 \cdot 10^{-2}$	0.13	0.36	0.71	1.18	1.73	2.34	3.02	3.74	4.51	5.31	6.14	6.99	7.86	8.75	9.65	10.57	11.49	12.43
Minneapolis	$1.31 \cdot 10^{-3}$	$8.6 \cdot 10^{-3}$	$2.49 \cdot 10^{-2}$	$4.96 \cdot 10^{-2}$	$8.22 \cdot 10^{-2}$	0.12	0.18	0.25	0.35	0.48	0.66	0.93	1.31	1.79	2.36	3.01	3.73	4.49	5.29	6.12
New Orleans	$2.14 \cdot 10^{-3}$	$8.49 \cdot 10^{-3}$	$1.9 \cdot 10^{-2}$	$3.34 \cdot 10^{-2}$	$5.17 \cdot 10^{-2}$	$7.39 \cdot 10^{-2}$	$9.98 \cdot 10^{-2}$	0.13	0.16	0.2	0.24	0.3	0.38	0.51	0.7	0.98	1.39	1.91	2.52	3.22
New York	$2.32 \cdot 10^{-3}$	$1.32 \cdot 10^{-2}$	$3.62 \cdot 10^{-2}$	$7.17 \cdot 10^{-2}$	0.12	0.18	0.26	0.36	0.48	0.64	0.85	1.12	1.5	1.99	2.58	3.26	4.02	4.83	5.69	6.59
North Port	$9.72 \cdot 10^{-4}$	$3.85 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$2.25 \cdot 10^{-2}$	$4.49 \cdot 10^{-2}$	$8.36 \cdot 10^{-2}$	0.14	0.23	0.35	0.55	0.84	1.22	1.69	2.27	2.93	3.66	4.44	5.25	6.09	6.96
Orlando	$1.36 \cdot 10^{-3}$	$5.38 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	$3.27 \cdot 10^{-2}$	$6.62 \cdot 10^{-2}$	0.12	0.21	0.34	0.52	0.74	1.05	1.47	1.97	2.58	3.28	4.05	4.87	5.72	6.61	7.51
Philadelphia	$8.46 \cdot 10^{-4}$	$3.38 \cdot 10^{-3}$	$1.21 \cdot 10^{-2}$	$3.15 \cdot 10^{-2}$	$6.11 \cdot 10^{-2}$	0.1	0.15	0.22	0.32	0.45	0.62	0.87	1.21	1.68	2.27	2.94	3.68	4.48	5.31	6.18
Phoenix	$2.86 \cdot 10^{-3}$	$1.82 \cdot 10^{-2}$	$5.96 \cdot 10^{-2}$	0.14	0.27	0.47	0.76	1.14	1.62	2.22	2.91	3.67	4.49	5.36	6.25	7.18	8.12	9.08	10.04	11.02
San Antonio	$4.08 \cdot 10^{-3}$	$1.72 \cdot 10^{-2}$	$4.24 \cdot 10^{-2}$	$8.22 \cdot 10^{-2}$	0.14	0.21	0.32	0.5	0.73	1.01	1.35	1.74	2.19	2.68	3.21	3.8	4.42	5.09	5.8	6.55
Shreveport	$3.07 \cdot 10^{-3}$	$1.19 \cdot 10^{-2}$	$2.58 \cdot 10^{-2}$	$4.85 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	0.18	0.33	0.55	0.85	1.26	1.77	2.4	3.1	3.88	4.7	5.55	6.43	7.34	8.26	9.19
St. Louis	$1.18 \cdot 10^{-3}$	$4.66 \cdot 10^{-3}$	$1.19 \cdot 10^{-2}$	$2.42 \cdot 10^{-2}$	$4.12 \cdot 10^{-2}$	$6.95 \cdot 10^{-2}$	0.12	0.19	0.3	0.44	0.63	0.87	1.17	1.55	2.05	2.66	3.34	4.08	4.86	5.68
Tucson	$7.76 \cdot 10^{-4}$	$3.09 \cdot 10^{-3}$	$6.93 \cdot 10^{-3}$	$1.23 \cdot 10^{-2}$	$2.14 \cdot 10^{-2}$	$3.6 \cdot 10^{-2}$	$5.83 \cdot 10^{-2}$	$9.01 \cdot 10^{-2}$	0.14	0.22	0.37	0.59	0.92	1.38	1.96	2.63	3.37	4.15	4.96	5.81
Washington	$1.99 \cdot 10^{-3}$	$7.46 \cdot 10^{-3}$	$1.58 \cdot 10^{-2}$	$2.68 \cdot 10^{-2}$	$4.23 \cdot 10^{-2}$	$6.41 \cdot 10^{-2}$	$9.93 \cdot 10^{-2}$	0.15	0.24	0.36	0.53	0.78	1.12	1.56	2.12	2.77	3.5	4.28	5.1	5.95

Table 2: Values of the area for each CSA, in the considered 20-day range. The variable X_i represents the area under the curve after i days, starting from the last day with the index equal to 0. We coloured the table cells according to the values: the darker cells contain higher values.

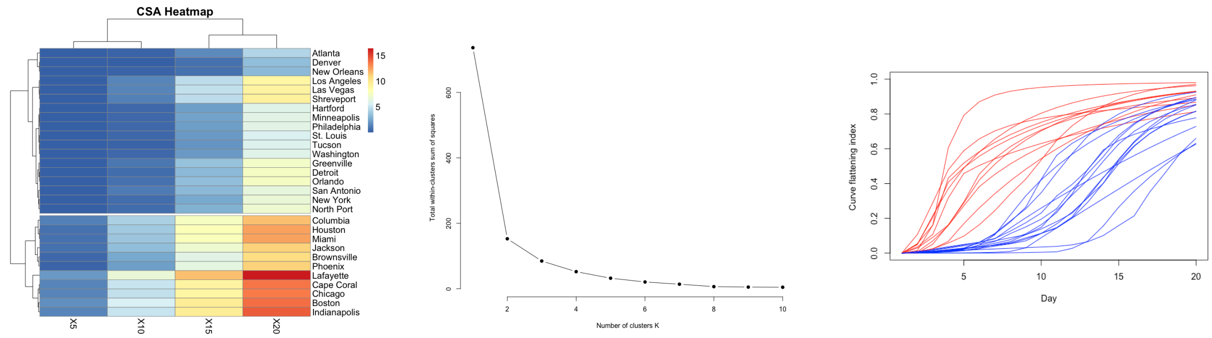


Figure 4: Heatmap and clusters of CSAs using 4 curve features (left). Elbow method results (center). The 20-day range considered for each CSA, by group (right).

by clusters, and observe that the strong social distancing response group was subdivided in strong (orange) and very strong (red) response. In Figure 6 we report the silhouette analysis: the average silhouette width suggests that the choice of two clusters, which is reported in the main paper, was more appropriate, as shown also by the elbow method. In addition, in Figure 7 we report the heatmap and clusters of CSAs considering all the 20 curve features. We observe that this choice yields the same extracted clusters as the ones reported in the main paper, which were extracted considering only 4 area features.

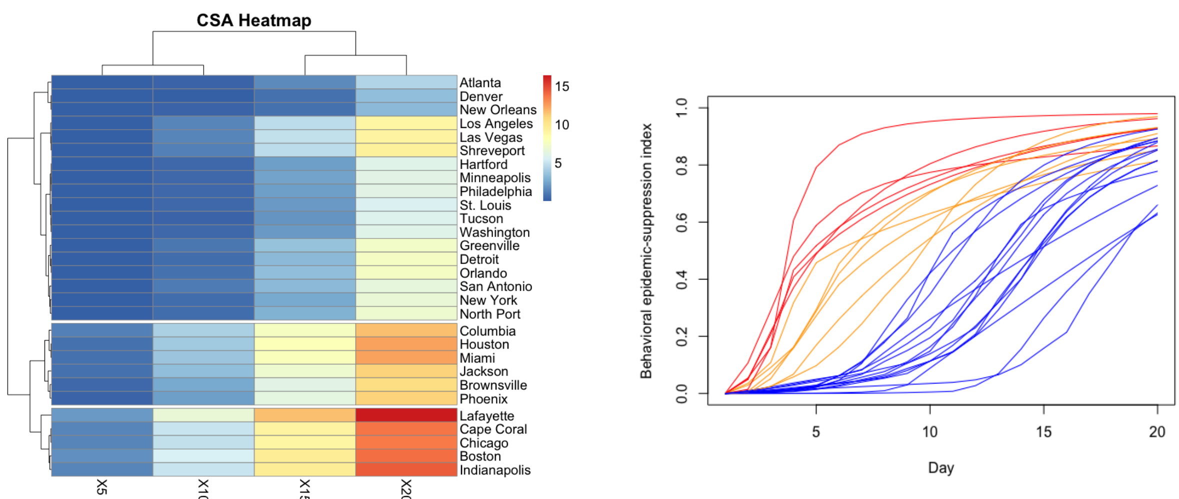


Figure 5: Heatmap and clusters of CSAs using 4 curve features (left), defining 3 clusters. 20-day range considered for each CSA, by group (right): weak (blue), strong (orange) and very strong (red) response.

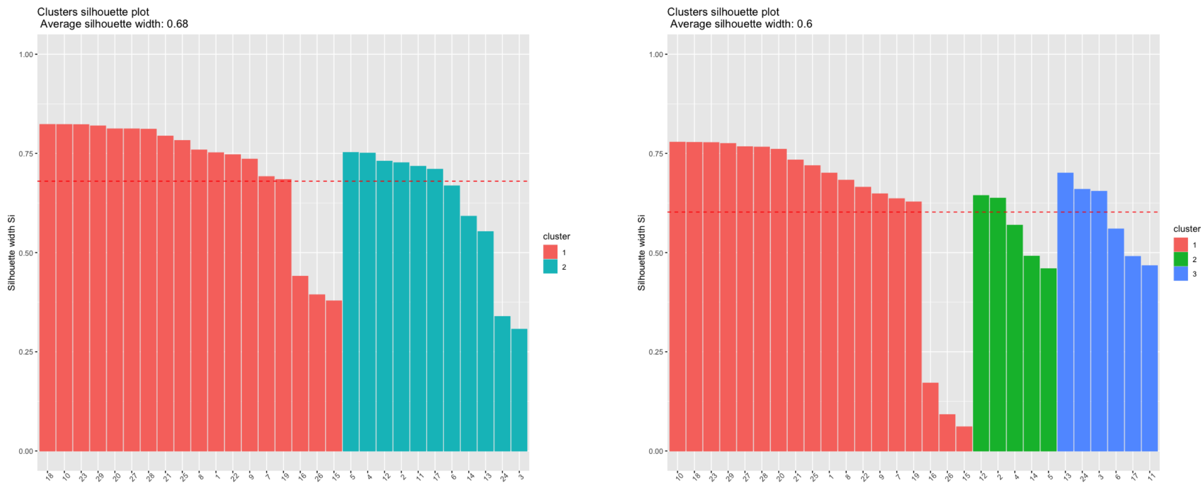


Figure 6: Silhouette information for the 2 clustering choices (2 and 3 clusters, considering the 4 area features). We observe a higher average silhouette width for the choice of 2 clusters (0.68 vs 0.60).

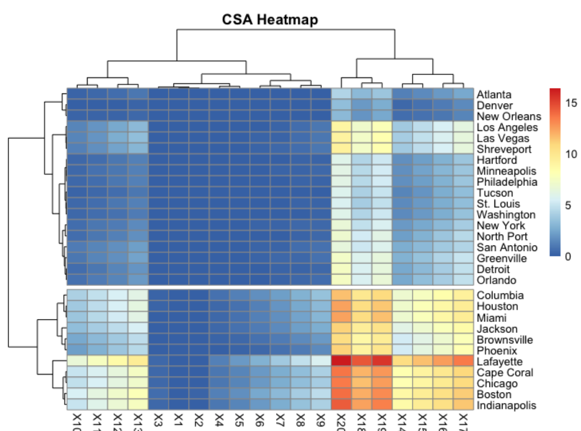


Figure 7: Heatmap and clusters of CSAs using all 20 curve features.

Additional clustering analysis that includes R_0

In Figure 8 (left) we report the results of the clustering analysis considering the 4 area features and R_0. To perform the clustering analysis, we normalised values by using the z-score transformation of the variables. By looking at the scatter plot of total 20-day area and R_0 in Figure 8 (right), we observe a clear separation of the 2 extracted clusters, with the area values being more influential to the clustering than R_0. The resulting clusters are the same as the previous analysis (considering only 4 area features), except for two CSAs (Brownsville-Harlingen-Raymondville, TX and Phoenix-Mesa, AZ). As shown in Figure 9, these two CSAs present an intermediate behaviour and they are now clustered with the *weak* response group. In Figure 10 we report the silhouette analysis for 2 and 3 clusters, observing the same average silhouette width for the two choices. We then report the 3 cluster analysis in Figure 11, observing that the new cluster is formed by the 2 CSAs New York-Newark, NY-NJ-CT-PA and Phoenix-Mesa, AZ, which present a particularly high R_0. The other two clusters are the remaining weak vs strong social distancing response CSAs.

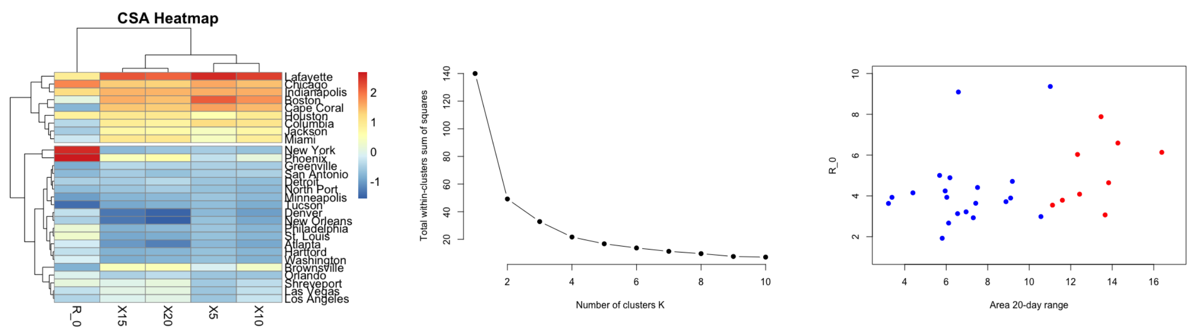


Figure 8: Heatmap and clusters of CSAs using 4 curve features and R_0 (left), considering the z-score transformation of the variables. Number of clusters chosen by performing the elbow method (center). Scatter plot of total 20-day area and R_0 (right), with CSA values coloured depending on the response: weak (blue) vs strong (red).

In Figure 12 we report the clustering analysis considering only two values: the final 20-day area value and R_0 (considering the z-score transformation of the variables). Also in this case, in Figure 12 (right) we observe that the 3 clusters represent the CSAs depending on weak vs strong response and by high values of R_0. In Figure 13 we observe a higher average silhouette width reported for 3 clusters.

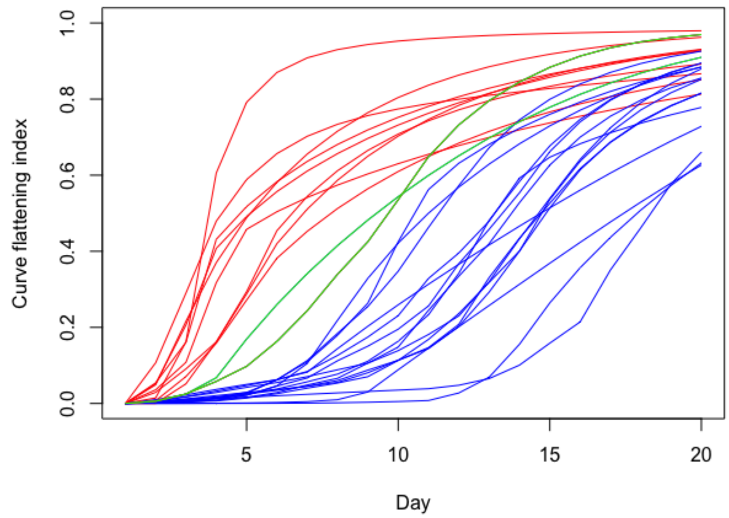


Figure 9: A 20-day range considered for each CSA, by group: weak (blue and green) and strong (red) response. We indicate the indices of Brownsville-Harlingen-Raymondville, TX and Phoenix-Mesa, AZ in green to highlight their behaviour: it is an intermediate response and now classified with the weak response group.

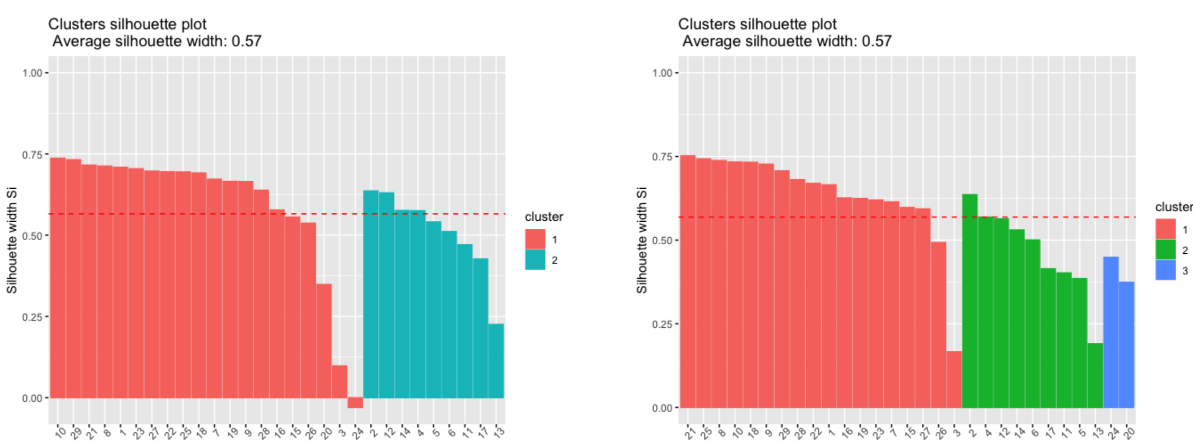


Figure 10: Silhouette information for 2 and 3 clusters, with the same average silhouette width for the two choices (0.57).

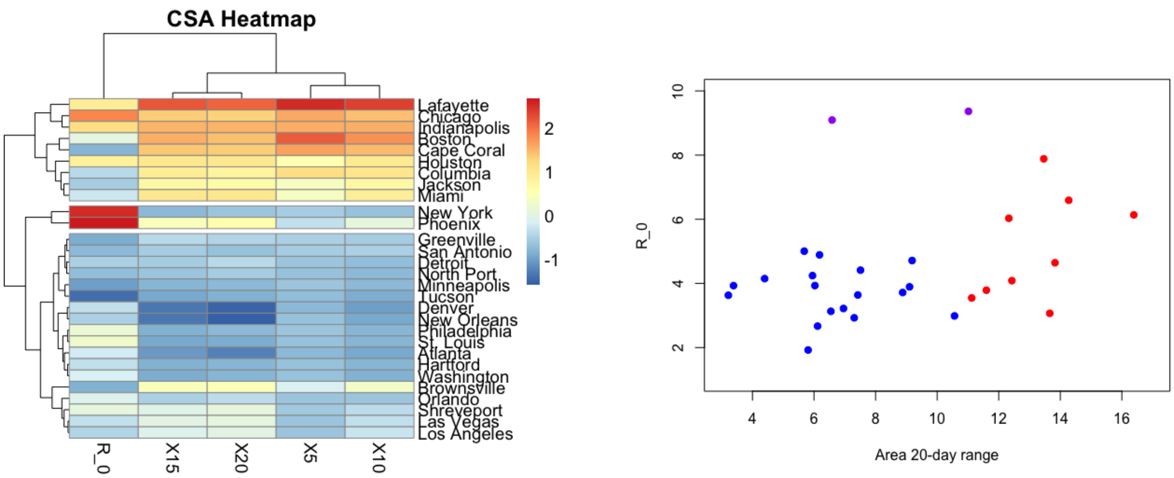


Figure 11: Heatmap showing the 3 clusters (left). Scatter plot of total 20-day area and R_0 (right), with CSA values coloured depending on the response or R_0: weak (blue) vs strong (red) response, and particularly high R_0 value (purple).

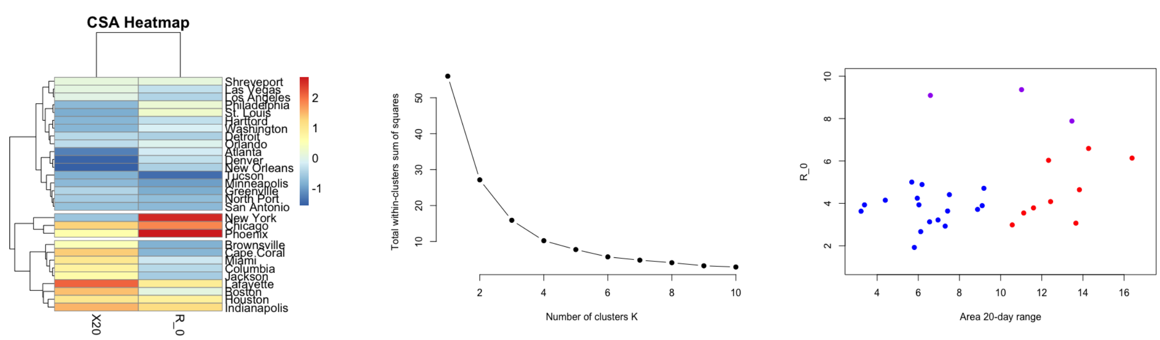


Figure 12: Heatmap and clusters of CSAs using the total 20-day area and R_0 (left), considering the z-score transformation of the variables. Number of clusters chosen by performing the elbow method (center). We observe a similar separation (right) as in Figure 8, with an additional separate cluster for the 3 CSAs reporting a high value of R_0 (Chicago-Naperville, IL-IN-WI, New York-Newark, NY-NJ-CT-PA and Phoenix-Mesa, AZ)

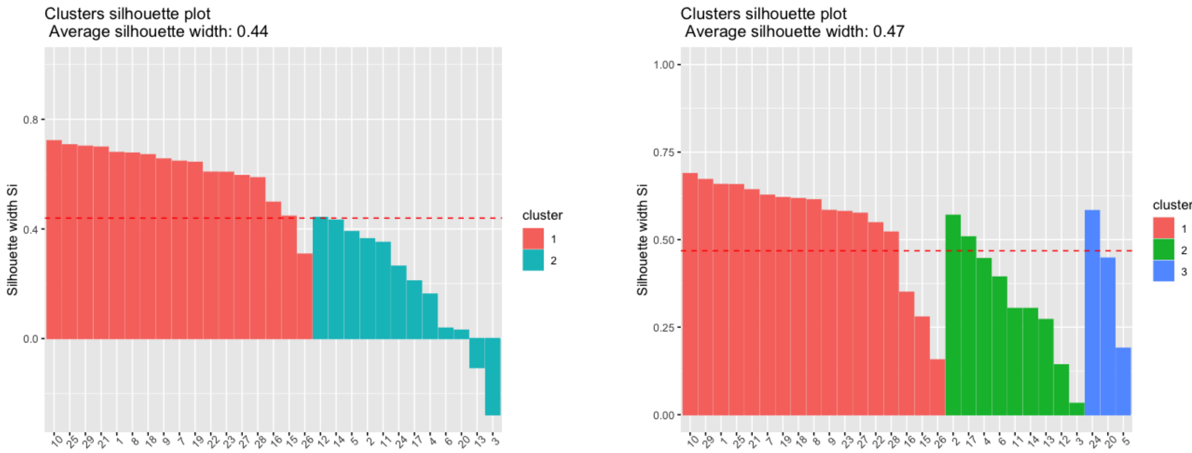


Figure 13: Silhouette information for the clustering of total 20-day area and R_0. We observe a higher average silhouette width for 3 clusters (0.47 vs 0.44).

References

Getz, W.M., Salter, R., Luisa Vissat, L., Horvitz, N., (2021). A versatile web app for identifying the drivers of COVID-19 epidemics. *Journal of Translational Medicine 19*, 1–20. `https://doi.org/10.1186/s12967-021-02736-2`.

Rai, B., Shukla, A., Dwivedi, L.K., (2020). Estimates of serial interval for COVID-19: A systematic review and meta-analysis. *Clinical epidemiology and global health*. `https://doi.org/10.1016/j.cegh.2020.08.007`.