Methodology for US Social Distancing Scoreboard

Pre-analysis decisions:

On how to assign devices to a county:

Devices were assigned to counties based on where a specific device was recorded for the longest time on a specific day. This allowed us to:

- Be independent from the Home and Work logic
- Account for people who moved to other places (e.g., to their partners, parents, friends, summer cottages, etc.) to live during the quarantine

On how to define PRE- and POST-COVID19 periods:

PRE-COVID19 period was defined as 4 weeks before the 8th of March, 2020 (incl). POST-COVID19 period is defined as everything after the 8th of March, 2020.

On how to define baselines:

For Metrics 1 and 2:

For each POST weekday, a baseline is the average of the metric on 4 corresponding PRE weekdays. For example, for the 30th of March, Monday, the baseline is the average of daily metrics on 4 PRE Mondays: 10, 17, 24th of February and 2nd of March.

For Metric 3:

Baseline is the national average of the metric taken over 4 weeks that immediately precede Covid-19 outbreak on March 8th, i.e, Feb 10 2020 - Mar 8 2020 incl.

On how to increase the reliability of the metrics:

Two filters were applied:

- Only those devices that had at least 2 dwells per day OR one dwell > 3h in duration were taken into the analysis
- We excluded counties with population < 1000 people OR where < 100 devices were observed on 70% of the days of the PRE-COVID19 period
- No counties with fewer than 5 venues or fewer than 100 visits during the baseline period were used in visitation

Metric 1: Difference (in %) in distance traveled

Definition: Percent reduction in total distance traveled per device, averaged across all devices located in the county.

- Reflects the difference in distance travelled per identifier, averaged on the county level
    - The metric is calculated each day as a percentage difference vs the avg distance traveled in the same county on the same weekday during the 4 weeks pre-Corona
    - The metric was chosen for its strong correlation with confirmed cases and policy measures in combination with its simplicity: no HOME assumed, relatively independent of supply fluctuations, more signals from people in motion, baseline is county and weekday specific
- In an initial phase, the U.S county with the strongest decrease was taken as an anchor. This proved to be problematic as travel distances were still decreasing.
- To recalibrate cut-offs for the grades, the highest decrease in traveled distance in Italy was taken as the baseline for what can be expected under a total shut-down. At the time of recalibration, no U.S state achieved the new cut-off, but no state had chosen to go into a full quarantine either. We defined steps for each grade so that F has a larger range compared to the other grades as a cautionary measure.
    - **A**: > 70% decrease
    - **B**: 55-70% decrease
    - **C**: 40-55% decrease
    - **D**: 25-40% decrease
    - **F**: <25% decrease or increase

Metric 2: Difference (in %) in visitation of non-essential POIs

Definition: Difference in visitation of non-essential POIs on a specific POST-COVID19 day compared to a corresponding PRE-COVID19 baseline.

- **Non-essential POIs** combine (1) venues from the VDP output and (2) additional BYOPOI based on categories from OpenSteetMaps (added to get more data → increase the accuracy of the metric) and they belong to the following groups :
    - ***Retail and Services***: Restaurant (multiple kinds), Department Store, Clothing (multiple kinds), Footwear, Discount Stores, Jewelry, Computers + Consumer Electronics, Gifts, Seasonal, Books, Office Supplies, Hair, Cosmetics + Beauty Supplies, Gyms + Fitness Facilities, Communications, New/Used Car Dealers, Hotels, Used Products, "Crafts, Toys, and Hobbies", Travel, "Spa, Massage, + Esthetics", Sports + Recreation, Weight Loss, Furnishings, Home + Housewares, Home Improvement +Building Supplies, "Printing, Copying + Publishing", Theatres, Music, Amusement, Furnishing Rentals, Shared Offices + Coworking, Car Wash, Cannabis Retail, Flowers
    - ***OpenSteetMaps POIs:***

        - amenity: bar, pub, cafe, restaurant, theatre, nightclub, cinema, casino
        - leisure: bowling_alley, fitness_centre, cafe, restaurant, theatre, nightclub
        - shop: department_store, mall, clothes, shoes, doityourself, furniture, sports
- Cut-offs for the grades were chosen based on counties which already implemented a shut down of non-essential business lines. The residual visitations consist mostly of take-away restaurants and some percentage of false visitation assignment:
    - **A**: >70% decrease
    - **B**: 65-70% decrease
    - **C**: 60-65% decrease
    - **D**: 55-60% decrease
    - **F**: <55% decrease or increase

Metric 3: Rate of Human Encounters compared to National Average pre Corona

Definition: Estimate of close encounters between two devices per square km, expressed as a fraction of pre-Corona baseline.

M3 = ((#unique encounters per day / km^2) / baseline) – 1

- **Encounter**
    - Defined as two clusters observed in spatial distance, Δd, of 50 m or less and in temporal distance, Δt, of 60 min or less. The metric was inspired by the one used by Pepe et al.*("To create such a network, we assert proximity between any two users of the same province who were seen within a circle of radius R = 50m over a 1 hour period."* Pepe et al.).
- **Cut-offs for the grades**
    - Defined based on the average x̃ from the distribution of the metric in PRE-COVID19 period and fixed percentages of this value; the baseline value represents "business as usual" and based on how much the number of encounters decreases (in comparison to this baseline), we assign grades:
        - **A**: > 94% Decrease Compared to National Baseline
        - **B**: 94% - 82% Decrease in Encounters Density Compared to National Baseline
        - **C**: 82% - 74% Decrease in Encounters Density Compared to National Baseline
        - **D**: 74% - 40% Decrease in Encounters Density Compared to National Baseline
        - **F**: < 40% Decrease in Encounters Density Compared to National Baseline

Example: given that median value from the pre-Corona x̃= 100, the range for B grade is computed as (100-94 ; 100-82) = (6 ; 18).

- **Normalization by area**
  - Applied to the total count of encounters in the given area. Area is expressed in Km^2 (=1,000,000 m^2).
- **Locations restrictions**
  - As we normalize by area (i.e., #encounters / area), we only consider land and exclude water areas. This is done based on area_land_meters column in bigquery-public-data.geo_census_tracts.census_tracts_* tables

Metric 1 + 2 + 3

- The total grade based on Metric 1, Metric 2, and Metric 3 was calculated as the average between the three numeric grades. This method was used because the two metrics have different cut-offs and because some counties have only one metric. The final scores were assigned in the following way:

Final Grades and Cut-offs		
Grade	**Lower cut-off**	**Higher cut-off**
A	5	NA
A -	4.5 (inclusive)	5
B +	4	4.5
B	4	NA
B -	3.5 (inclusive)	4
C +	3	3.5
C	3	NA

C -	2.5 (inclusive)	3
D +	2	2.5
D	2	NA
D -	1.5 (inclusive)	2
F	1	1.5