FLOWMINDER.ORG

Haiti Mobility Data Platform

Description of indicators

Last updated: November 2022

Description of indicators

Last updated: November 2022

List of content

Introduction

The Haiti Mobility Platform (haiti.mobility-dashboard.org) is a privacy-secure web-platform providing mobility indicators based on pseudonymised Call Detail Records (CDR) from Digicel Haiti to approved third parties. These indicators represent the distributions and movements of the Haiti population, and their dynamical changes. The platform allows users to visualise, interact and download the data for further analysis. To understand how we produce indicators from pseudonymised CDR data, please click here. You will be redirected to FlowGeek, our online knowledge centre on CDR data analytics.

This document is aimed at a technical / scientific audience and presents the indicators available on the Haiti Mobility Platform as of end of September 2022. In this document, you will learn more about available indicators, what they mean and measure and how they were calculated and any intermediate calculations required. Additionally, considerations for some calculations are described if applicable.

General remarks

- The *baseline period* referred to herein is defined as August 2020 - September 2021
- Some indicators have similar names. XXX_counts => comes from aggregates directly, XXX_subscribers => indicators derived from XXX_counts that is scaled / corrected for sample of subscribers, XXX_people => population scaled indicator

Residents

Arrived

Number of arrived *people* derived from scaling the arrived *subscribers*

We multiply arrived subscribers to scale up to the population level with (popstatic / median(residents_subscribers(baseline))). However the subs we see are not necessarily representative of the actual population so we use adjustment factors X (derived at the bottom of this doc) to adjust further [rounded to nearest 10 people]

Calculated as: arrived _people(month) = X * (popstatic / median(residents_subscribers(baseline))) * arrived_subscribers(month)

Departed

- See arrived, replacing arrived_subscribers with departed_subscribers [rounded to nearest 10 people]

Delta_arrived

- Difference in arrived and departed people
- Calculated as: arrived_people - departed_people [rounded to nearest 10 people]

Residents

Number of resident *people* derived from the number of resident *subscribers*

Resident subscribers are computed as a baseline level of subscribers plus some change in subscribers (baseline subs + change in subs). Goal is to scale this to (baseline population level + change in population)

We take the difference in observed subscribers from the baseline subscriber period (residents_subscribers(month) - median(residents_subscribers(baseline))) and scale this from subscriber level to population level (popstatic / median(residents_subscribers(baseline)). This difference is not representative of the whole population, so we compute scaling factors X (see end of doc) for the population scaled difference and then add this difference to popstatic to get an estimate of the resident people in the area. [rounded to nearest 10 people]

Calculated as: residents_people(month) = popstatic + X * (popstatic / median(residents_subscribers(baseline))) * (residents_subscribers(month) - median(residents_subscribers(baseline)))

Residents_perKm2

Number of resident *people* derived from the number of resident *subscribers* divided by the area of the spatial unit, per spatial unit **[rounded to nearest 10 people]**

Calculated as: resident_people(adm3) / adm3_km2(adm3)

Abnormality

For each month abnormality is defined as the number of standard deviations away from a reference baseline period - using modified z scores (link) which measures how many standard deviations the number of residents is away from the median number of residents in the baseline period.

Calculated as:

If MAD does equal 0
(resident(month)-median(residents(baseline))/(1.253314*MeanAD(baseline)).
If MAD does not equal 0
(resident(month)-median(residents(baseline))/(1.486*MAD(baseline)).

Where MAD and MeanAD are the median absolute deviation / mean absolute deviation of the number of residents in the baseline period (see https://en.wikipedia.org/wiki/Average_absolute_deviation),

[rounded to nearest 2 d.p.]

Residents_diffwithref

Per month the difference between the *estimated residents* and the *median residents in the baseline period*

Calculated as: residents(month) - median(residents(baseline))

Residents_pctchangewref

Per month the pct change between the estimated residents and the median residents in the baseline period

Calculated as: 100 * (residents(month) - median(residents(baseline)) / median(residents(baseline)))

Intermediary aggregates/indicators - used to compute release indicators

resident_count

'Residents' aggregate that counts the modal locations of subscribers each month

1. Using a 7 days rolling window, assign daily home location for each subscriber as the most frequent location of the last call of the day. For breaking ties, select the most recent home location

2. Using daily home location for a given month, assign monthly home location as a location with at least 50% daily home location in the current month or 30% of daily locations in a current month and 50 or more % of daily location in the previous month; otherwise assign as unlocatable

Considerations:
Removed any location with less than 200 median residents

Popstatic

See Region Scaling Factors X (end of doc) for a description of the population layer used

---

Adm3_km2

(pcod, area) pairs, computed by projecting WGS84 (lat lon coords) to EPSG:32618 (metre coords), taking area in km

---

Median_residents_count_baseline

Median of the residents counts between Aug 2020 and Sep 2021

Calculated as: median(resident_counts(baseline)) for each adm3 unit

---

Stayed_month

The number of users who are active in residents_counts vary from month to month. This intermediary variable computes for a given month, the number of people were active in a given month AND the month prior - used for calibration to adjust the sample we get in resident_counts to estimate resident_subscribers

Calculated as: stayed_month = sum(relocations(location, location)) for each location

---

Stayed_baseline

Median users that stay at a location over a given month

Calculated as: median(stayed_month(baseline))

Sample_size_adjustment_factor

Over different months we have a different sample of users - as in some months users are undetectable. Therefore we see fluctuations in residents counts that are attributed to users not being active in that given month (whereas we want to see fluctuations only based on the movement between adm3 units.

Calculated as: sample_adj_factor(month) = median(stayed_baseline) / stayed(month)

Considerations:

Where stayed_month << stayed baseline, the sample adjustment factor is very large. We remove any months where the adjustment for sample increases the counts by over 100% - as it is unclear whether changes in these months can be attributed to actual mobility or the fluctuation of detectable users over time. For instance in HT0542-05 we see months where under 10 people are detectable who were residents there in the previous month. This causes the post adjustment arrived / departed subscribers to be over 120 times larger for a given month.

Arrived_count

Number of inflows to an admin unit

Calculated as: arrived_count(location, month) = sum(relocations(x->location, month)) for each x in adm3_units where x not equal to location

---

Departed_count

Number of outflows from an admin unit

Calculated as: arrived_count(loc, month) = sum(relocations(loc->x, month)) for each x in adm3_units

---

Arrived_subscribers

Number of arrivals after adjustment for the difference in sample between each month

Calculated as: arrived_subscribers(adm3, month) = sample_size_adjustment_factor(adm3, month) * arrived_count(adm3, month)

Departed_subscribers

Calculated as: see above replacing with *departed_count*

Delta_arrived_subscribers

Net inflows to each adm3 unit per month

Calculated as: delta_arrived_subscribers(adm3, month) = arrived_subs(adm3, month) - departed_subs(adm3, month)

Delta_arrived_subscribers_abnormality

Number of std deviations away from median delta_in during baseline period. Used to identify large anomalous flows related to poor data instead of mobility. One example is when the number of stayed people are very small compared to the baseline (where stayed <<< stayed_baseline) and there are a large number of appeared people in an area, meaning we see upwards of 20000 people relocating to an area when in previous months it was ~500.

Calculated as: Modified z-scores used here

If MAD does equal 0
(delta_in_subs(month)-median(delta_in_subs(baseline))/(1.253314*MeanAD(baseline)).
If MAD does not equal 0
(delta_in_subs-median(delta_in_subs)/(1.486*MAD(baseline)).

Where MAD and MeanAD are the median absolute deviation / mean absolute deviation of the delta_arrived in the baseline period (see https://en.wikipedia.org/wiki/Average_absolute_deviation),

> Considerations:
> Removed any delta_in > 6 std away from baseline

Cumulative_sum(delta_arrived_subscribers)

Cumulative sum of delta_arrived subs. If no data for a month we do as follows
cumsum([1, nan, 2, 3, nan, 4] = [1, nan, 3, 6, nan, 10]

Calculated as: cumulative_sum(delta_arrived_subscribers)

Resident_subscribers

Estimated number of subscribers at an adm3 unit for a given month

Calculated as: resident_subscribers = median(resident_count(baseline)) + cumulated_sum(delta_arrived)

Median(resident_subscribers(baseline))

Median number of subs in the baseline period

Calculated as: median(resident_subscribers(baseline))

Scaling_adjustment_factors (or 'X')

Adjustment factors to control for the

Calculated as: derived at the bottom of this doc

Relocations

Considerations: Remove any (loc A, loc X, month), (loc X, loc A, month) pair when (Loc A, month) is removed from residents

Relocations

Number of people who moved their home from region A to region B, between time period t and time period t+n (with n = 1, unless missing data)

Scales relocations seen by subscribers to the population scale using (popstatic_B / median(residents_subscribers_B(baseline))), with

Calculated as: relocations(A, B, month) = X_B * (popstatic_B / median(residents_subscribers_B(baseline))) * relocations_subscribers_AtoB(month)

Considerations: Remove any (loc A, loc B, month) tuple with less than 50 relocations

Impact on sum of relocations:
keeps: 2565100 (58.0%), tosses: 1868295 (42.0%)

clarification: we redact anything less than 50 after pop scaling

---

Abnormality

Compares the difference in the number of relocations for the current month with a reference period, to its variations during the reference period

Calculated as:

If MAD does equal 0
(resident(month)-median(relocations(A, B, baseline))/(1.253314*MeanAD(A, B, baseline)).
If MAD does not equal 0
(relocations(A, B, month)-median(relocations(A, B, baseline))/(1.486*MAD(A, B, baseline)).

Where MAD(A, B, baseline), and MeanAD(A, B, baseline) are the median/mean absolute difference of the relocation counts months in the baseline from loc A to loc B

Considerations
If there are not at least 6 data points in baseline, remove all (loc A, loc B, month) pairs as we cannot reliably say how abnormal a given relocation flow is (arguably, even 12 points is not enough)

Relocations_diffwithref

Difference in number of relocations between this month and a reference period

Calculated as:

relocations_diffwithref(a, b, month) = relocations_diffwithref(a, b, month) - median(relocations_diffwithref(a, b, [baseline_months]))

Considerations
If there are not at least 6 data points in baseline, remove all (loc A, loc B, month) pairs as we cannot reliably say how abnormal a given relocation flow is

---

Relocations_pctchangewithref

% change in number of relocations between this month and a reference period

Calculated as: relocations_diffwithref(a, b, month) = (relocations_diffwithref(a, b, month) - median(relocations_diffwithref(a, b, [baseline_months]))) / median(relocations_diffwithref(a, b, [baseline_months])

Considerations

If there are not at least 6 data points in baseline, remove all (loc A, loc B, month) pairs as we cannot reliably say how abnormal a given relocation flow is

---

Intermediary aggregates/indicators

Popstatic

See Region Scaling Factors X (end of doc) for a description of the population layer used

---

Median resident subscribers baseline

Median number of subs in the baseline period

Calculated as: median(resident_subscribers(baseline))

Stayed_month

The number of users who are active in residents_counts vary from month to month. This intermediary variable computes for a given month, the number of people were active in a given month AND the month prior - used for calibration to adjust the sample we get in resident_counts to estimate resident_subscribers

Calculated as: stayed_month = sum(relocations(location, location)) for each location

---

stayed_baseline

Median users that stay at a location over a given month

Calculated as: median(stayed_month(baseline))

Pop scaling factors / adjustment factors / 'X'

Calculated as: See bottom of document

Relocations_count_AtoB

Calculated as: Raw number of relocations from location A to location B

Relocations_subscribers_AtoB

Calculated as: relocations_subscribers_AtoB(A, B, month) = (stayed_baseline(locB) / stayed_month(month, locB)) * relocations_count_AtoB(A, B, month)

Daily presence

Presence

An estimated number of people presented in the region at time period t (presence include residents 'visiting' their own region) [rounded to the nearest 10 people]. Time period t is a day in this case.

Presence in people presence_people(day,adm3) for each administrative area (adm3, communal sections, in this case) per a day is computed as a change in static population in a given administrative area popstatic(adm3) by taking difference in observed subscribers' presence from the baseline period (presence_subscribers(day,adm3) - median(presence_subscribers(baseline, adm3) and scaling this to the population level by multiplying the difference by X(adm3) * (popstatic(adm3) / median(presence_subscribers(baseline, adm3)), where X(adm3) is a regional scaling factor calculated for each adm3 as described below in Regional scaling factors X.

Calculated as
presence_people(day,adm3) = popstatic(adm3) + X(adm3) * (popstatic(adm3) / median(presence_subscribers(baseline,adm3)) * (presence_subscribers(day,adm3) - median(presence_subscribers(baseline,adm3))

Considerations:

Presence represents both residents and visitors to a given area per day.
Several filters are applied to presence per admin3:

- Communal sections (admin3) with presence median of less than 300 subscribers during the study period is discarded from consideration
- Communal sections with presence less than 100 subscribers per day are discarded for this day
- Communal sections with presence less than 100 people (after scaling) per day are discarded for this day
- Communal sections with high abnormality scores (>6) per day was not discarded, however they were analysed

- Though an attempt was made to estimate a number of people based on the number of subscribers, more work on methodology on how to estimate people is needed.

Presence_perKm2

An estimated number of people presented per square kilometre in the region in a time period (day) [rounded to the nearest 10 people].

Calculated as: presence_people(day, adm3) / adm3_km2(adm3)

Presence_diffwithref

Difference in the number of people's presence between this day and the median of people's presence in a baseline period (a reference period).

Difference in the number of people's presence in comparison to the baseline period is calculated by scaling difference in subscribers' presence to population level for each adm3 per day.

Calculated as: presence_diffwithref(day,adm3) = X(adm3) * (popstatic(adm3) / median(presence_subscribers(baseline,adm3))) * presence_subscribers_diffwithref(day,adm3)

Presence_pctchangewithref

Percent change in a number of people's presence between this day and the median of people's presence in a baseline period (a reference period).

Percent change in a number of people's presence is calculated by scaling percent change in subscribers' presence to population level for each adm3 per day.

Calculated as: presence_pctchangewithref(day,adm3) = X(adm3) * presence_subscribers_pctchangewithref(day,adm3)

[rounded to nearest 2 d.p.]

Abnormality

Abnormality compares the difference in subscribers' presence for the current day with a baseline (reference) period to its variations during the baseline period. Daily abnormality is defined as the number of standard deviations away from a baseline period using modified z scores (link) which are calculated from the mean absolute deviation (MeanAD) or median absolute deviation (MAD). Absolute values above 3 indicate subscribers' presence is abnormal. This may be due to technical

issues or to a real event. Very large absolute values (above 6) are more likely to indicate technical issues.

Calculated as:

If MAD does equal 0:
abnormality(day,adm3)=(presence_subscribers(day,adm3) - median(presence_subscribers(baseline,adm3)))/ (1.253314*MeanAD(baseline,adm3))
If MAD does not equal 0:
abnormality(day,adm3)=(presence_subscribers(day,adm3) - median(presence_subscribers(baseline,adm3)))/ /(1.486*MAD(baseline,adm3)).

[rounded to nearest 2 d.p.]

Intermediary aggregates/indicators

Popstatic(adm3)
Static population per adm3 calculated using boundaries for adm3 and WorldPop population estimates, more details can be found below in Regional scaling factors X.

X(adm3)
A regional scaling factor calculated for each adm3 as described below Regional scaling factors X

Presence_subscribers(day,adm3)
Equals to presence_counts(day,adm3), computed as a number of active subscribers presented in a given area (adm3) in a given time period (day)

Median(presence_subscribers(baseline,adm3))
Median of the daily subscribers' presence in the baseline period between August 2020 and September 2022

Adm3_km2(adm3)
(pcod, area) pairs, computed by projecting WGS84 (lat lon coords) to EPSG:32618 (metre coords), taking area in km.

Presence_subscribers_diffwithref(day,adm3))
Difference in presence_subscribers(day,adm3) between a given day and the median subscribers' presence in the baseline period for each adm3.

Calculated as: presence_subscribers_diffwithref(day,adm3)

= presence_subscribers(day,adm3) - median(presence_subscribers(baseline,adm3))

Presence_subscribers_pctchangewithref(day,adm3))
Percent change in presence_subscribers(day,adm3) between a given day and the median subscribers' presence in the baseline period for each adm3.

Calculated as: presence_subscribers_pctchangewithref(day,adm3) = 100*(presence_subscribers(day,adm3) - median(presence_subscribers(baseline,adm3)))/ median(presence_subscribers(baseline,adm3))

Daily movements

Travellers
An estimated number of people who visited region A then region B, in time period t . The period is 1 day in this case.

Number of people travelled from A to B per day travellers_people_AtoB(day) is calculated from a number of subscribers travelled from A to B per day travellers_subscribers_AtoB(day) by scaling this by destination using adjustment factor X_B * (popstatic_B / median(presence_subscribers_B(baseline))), where X_B is a regional scaling factor calculated for each adm3 as described below in Regional scaling factors X, popstatic_B is static population in destination and median(presence_subscribers_B(baseline)) is median counts of subscribers' presence in destination during the baseline period,

Calculated as travellers_people_AtoB(day) = X_B * (popstatic_B / median(presence_subscribers_B(baseline))) * travellers_subscribers_AtoB(day)

[rounded to the nearest 10 people]

Considerations:

Travellers are calculated as consecutive pairs to avoid double counting. However, some long distance travels are missed in this case if a subscriber makes some calls during the travel.

Several filters are applied to travellers per admin3:

- Communal sections (admin3) with presence median of less than 300 subscribers during the study period is discarded from consideration
- Pairs with less than 50 subscribers per day are discarded for this day
- Pairs with less than 50 people (after scaling) per day are discarded for this day

Pairs with high abnormality scores (>6) per day was not discarded, however they were analysed

Though an attempt was made to estimate a number of people based on the number of subscribers, more work on methodology on how to estimate people is needed.

Travellers_diffwithref

Difference in number of travellers (people) between this day and a median of travellers (people) in a baseline (reference) period, for each pair.

Difference in the number of travellers (people) in comparison to the baseline period is calculated by scaling difference in travellers(subscribers) in comparison to the baseline period to population level for each pair per day.

Calculated as:
travellers_AtoB_diffwithref(day) = X_B * (popstatic_B / median(presence_subscribers_B(baseline))) * travellers_AtoB_subscribers_diffwithref(day)

Travellers_pctchangewithref

Percent change in the number of travellers (people) between this day and the median of in a baseline period (a reference period), for each pair.

Percent change in a number of travellers (people) is calculated by scaling the percent change in a number of travellers (subscribers) to population level for each pair per day.

Calculated as: travellers_AtoB_pctchangewithref(day) = X_B * travellers_AtoB_subscribers_pctchangewithref(day)

[rounded to nearest 2 d.p.]

Abnormality

Abnormality compares the difference in the number of travellers for the current day with a baseline (reference) period to its variations during the baseline (reference) period, for each pair. Daily abnormality is defined as the number of standard deviations away from a baseline period using modified z scores (link) which are calculated from the mean absolute deviation (MeanAD) or median absolute deviation (MAD). Absolute values above 3 indicate subscribers' presence is abnormal. This may be due to technical issues or to a real event. Very large absolute values (above 6) are more likely to indicate technical issues.

Calculated as:

If MAD does equal 0:

abnormality(day,AtoB)=(travellers_subscribers_AtoB(day) - median(travellers_AtoB_subscribers(baseline)))/ (1.253314*MeanAD(baseline,AtoB))
If MAD does not equal 0:
abnormality(day,AtoB)=(travellers_subscribers_AtoB(day) - median(travellers_AtoB_subscribers(baseline)))/ /(1.486*MAD(baseline,adm3)).

[rounded to nearest 2 d.p.]

Trips_in

An estimated number of trips (people) to a given area from any other location during a given day.

Number of trips (people) to adm3 per day trips_in_people(day,adm3) is calculated from a number of trips (subscribers) to adm3 per day trips_in_subscribers(day,adm3) by scaling this using adjustment factor X(adm3) * (popstatic(adm3) / median(presence_subscribers(baseline,adm3))), where X(adm3) is a regional scaling factor calculated for each adm3 as described below in Regional scaling factors X, popstatic(adm3) is static population and median(presence_subscribers(baseline,adm3)) is median counts of subscribers' presence in adm3 during the baseline period,

Calculated as:
trips_in_people(day,adm3) = X(adm3) * (popstatic(adm3) / median(presence_subscribers(baseline,adm3))) * trips_in_subscribers(day,adm3)

[rounded to the nearest 10 people]

Considerations:

Trips_in includes trips made by some subscribers several times to the communal section of interest per day.

Several filters are applied to Trips_in per admin3:

- Communal sections (admin3) with presence median of less than 300 subscribers during the study period is discarded from consideration for Trips_in calculations
- Communal sections with Trips_in less than 50 subscribers per day are discarded for this day
- Communal sections with Trips_in less than 50 people (after scaling) per day are discarded for this day

Though an attempt was made to estimate a number of people based on the number of subscribers, more work on methodology on how to estimate people is needed.

Trips_out

An estimated number of trips (people) from a given area to any other location during a given day.

Number of trips (people) from adm3 per day trips_out_people(day,adm3) is calculated from a number of trips (subscribers) from adm3 per day trips_out_subscribers(day,adm3) by scaling this using adjustment factor X(adm3) * (popstatic(adm3) / median(presence_subscribers(baseline,adm3))), where X(adm3) is a regional scaling factor calculated for each adm3 as described below in Regional scaling factors X, popstatic(adm3) is static population and median(presence_subscribers(baseline,adm3)) is median counts of subscribers' presence in adm3 during the baseline period,

Calculated as:
trips_out_people(day,adm3) = X(adm3) * (popstatic(adm3) / median(presence_subscribers(baseline,adm3))) * trips_out_subscribers(day,adm3)

[rounded to the nearest 10 people]

Considerations:

Trips_out includes trips made by some subscribers several times to the communal section of interest per day.

Several filters are applied to Trips_out per admin3:

- Communal sections (admin3) with presence median of less than 300 subscribers during the study period is discarded from consideration for Trips_out calculations
- Communal sections with Trips_out less than 50 subscribers per day are discarded for this day
- Communal sections with Trips_out less than 50 people (after scaling) per day are discarded for this day

Though an attempt was made to estimate a number of people based on the number of subscribers, more work on methodology on how to estimate people is needed.

Intermediary aggregates/indicators

Travellers_subscribers_AtoB(day)

Equals to non-diagonal elements of the origin-destination matrix, consecutive pairs, which is the count of unique subscribers present in A then B per day.

X_B

A regional scaling factor calculated for destination B as described below in Regional scaling factors X.

Popstatic_B

Static population in destination B calculated using boundaries for destination B (adm3) and WorldPop population estimates, more details can be found below in Regional scaling factors X.

Median(presence_subscribers_B(baseline))

Median of the daily subscribers' presence in the destination B during the baseline period between August 2020 and September 2022.

Travellers_AtoB_subscribers_diffwithref(day)

Difference in the number of subscribers who travelled from A to B in a given day and the median number of subscribers who travelled from A to B in the baseline period.

Calculated as: travellers_AtoB_subscribers_diffwithref(day)
= travellers_AtoB_subscribers(day) - median(travellers_AtoB_subscribers(baseline))

---

travellers_AtoB_subscribers_pctchangewithref(day)

Percent change in the number of subscribers who travelled from A to B in a given day and the median number of subscribers who travelled from A to B in the baseline period.

Calculated as: travellers_AtoB_subscribers_pctchangewithref(day)
= 100*(travellers_AtoB_subscribers(day) - median(travellers_AtoB_subscribers(baseline)))/ median(travellers_AtoB_subscribers(baseline))

X(adm3)

A regional scaling factor calculated for each adm3 as described below Regional scaling factors X.

Popstatic(adm3)

Static population per adm3 calculated using boundaries for adm3 and WorldPop population estimates, more details can be found below in Regional scaling factors X.

Median(presence_subscribers(baseline,adm3))

Median of the daily subscribers' presence in the baseline period between August 2020 and September 2022.

Presence_subscribers(day,adm3)

Equals to presence_counts(day,adm3), computed as a number of active subscribers presented in a given area (adm3) in a given time period (day)

Trips_in_subscribers(day,adm3)

Sum of all subscribers who arrived in a given region from all other regions excluding a given region during a day. The same subscriber can arrive in this region several times from different regions.

Trips_out_subscribers(day,adm3)

Sum of all subscribers who departed from a given region to all other regions excluding a given region during a day. The same subscriber can depart from this region several times to different regions.

Regional scaling factors X

Here we scale the variation of an indicator to the *phone-using population* using X * (Static population / Resident subscribers during baseline period) e.g. for a given admin 3 unit,

Resident_people in current month = Static population + X * (Static population / Resident subscribers during baseline period) * (Resident subscribers in current month - Resident subscribers during baseline period)

$$\widehat{P}_{im} = P_{i0} + sf_i * (R_{im} - R_{io})$$

$$sf_i = \frac{P_{i0}}{R_{i0}} * x_i$$

$$x_i = \{p_j \qquad if\ c_{i0} \geq \frac{1}{25},\ i.e. \frac{P_{i0}}{R_{i0}} \leq 25$$

$$\{c_{i0} * median(\frac{p_j}{c_i}) \quad if\ c_{i0} < \frac{1}{25},\ i.e.\ \frac{P_{i0}}{R_{i0}} > 25$$

$$c_{i0} = \frac{R_{i0}}{P_{i0}}$$

$\hat{P}_{im}$ … Estimated nr of resident phone users in communal section i in month m
p_j … Mobile phone user rate in department j (static over time)
P_{i0} … Total population in communal section i in baseline period
sf_i … Scaling factor for communal section i (static over time)
R_{im} … Subscriber (Digicell) residents in communal section i in month m
R_{i0} … Subscriber (Digicell) residents in communal section i in baseline period
x_i … Scaling adjustment factor for communal section i (static over time)
c_{i0} … Subscriber coverage rate for communal section i in baseline period

If X is 1, then we are assuming that the variation of the indicator calculated from the CDR subset is representative of the variation in the general population.

If X is the phone-using penetration rate, then we are assuming that the variation of the indicator calculated from the CDR subset is representative of the variation in the phone-using population.

If X is the Digicel penetration rate, then we are assuming that the variation of the indicator calculated from the CDR subset is representative of the variation in the Digicel subscriber base..

The following are used to arrive at a definition of X for each admin 3 unit:

- HRSL 2020 population layer where pixel values are adjusted uniformly so that the total population equals the 2021 IHSI national population total
- Department (Admin 1 units + Port-au-Prince metro area) phone user penetration rate calculated from Digicel's June 2022 market share report with the denominator calculated from HRSL population layer i.e.
  PEN_DEPARTMENT = (DIGICEL_DEPARTMENT_SUBSCRIBERS + NATCOM_DEPARTMENT_SUBSCRIBERS) / HRSL_DEPARTMENT_POPULATION
- Median resident subscribers of each admin 3 unit during the baseline period (Jan-Nov 2021….)

First the coverage rate is calculated for each admin 3 unit, i:

c_i= R_i / P_i

Where R is the number of resident subscribers estimated from the CDR of sufficiently active subscribers, and P is the population count from the HRSL estimate.

Second the phone user penetration rate for each admin 1 unit j (department + PaP metro) is calculated:

p_j = (D_j + N_j) / P_j

Where D_j is the number of Digicel subscribers, N_j is the number of NATCOM subscribers, and P_j is the population count from the HRSL estimate.

- For all admin 3 units where c>=1 / 25, X_i is assigned as the Department penetration rate p_j.

- For all admin 3 units where c<1/25, X_i is assigned as c_i * median(p_j / c)

where median(p_j / c) is the admin 1 median of (p_j / c_i) of all admin 3 units i within admin 1 unit j. This means that for those admin 3 units with very high scaling factors, we replace X * P / R in the indicator equations with the admin 1 median value of X * P / R, thus suppressing the scale up of the indicator variations since they have higher uncertainty

Contact us

For queries or information about the Haiti Mobility Data Platform, the methods presented in this document or on mobile data analytics in general, please contact us at **info@flowminder.org**