This document provides additional information about the content of the following four datasets:

- MIG2012_DATA_dwelling.dta
- MIG2012_DATA_household.dta
- MIG2012_DATA_individual.dta
- MIG2012_DATA_movement.dta

These four datasets contain the same information. However, to aid the user we provide them at different units: a dwelling can consist out of multiple households, a household often consists out of multiple individuals, and an individual often makes multiple migration movements.

Any comments or errors, please send an email: petervanderwindt@nyu.edu

Generic codes across the datasets:

- - "-9" is don't know
- - "-8" is not applicable
- - "-7" is refused to respond

VARIABLES LEVELS

The first letters of the variable names indicate at what level the variables operate:

- "vill_" (all datasets) are variables that contain information at the village level.
- "dwelling_" (all datasets) are variables that contain information at the dwelling level.
- "hh_" (all datasets) are variables that contain information about the household.
- "p**X**hh**X**_" (dwelling dataset only): are variables that contain information at the individual level. For example, "p1hh1" indicates first individual in the first household, "p2hh3" indicates second individual in third household, etc.
- "p**X**_" (household dataset only): are variables that contain information at the individual level. For example, "p1_" indicates that household's first individual, etc.
- "v**X**_" (dwelling, household and individual datasets). For example, "v1_" contains information about the first village the individual has lived, "v2_" about the second village, etc.
- "d**X**_" (dwelling, household and individual datasets). For example, "d1_" contains information about the first displacement the individual made (between village 1 and village 2), "d2_" about the second displacement (from village 2 to village 3), etc.

IDENTIFYING VARIABLES

Every observations is uniquely identified in the datasets:

IDV: Unique village code (all datasets). This ranges from 1 to 25.[1]

IDD: Unique dwelling code (all datasets). IDD is IDV + number.

- The dwelling dataset is structured as follows: each row represents one dwelling. Variables starting with "p1hh1" refer to the head of the household ("p1") in the "main household" ("hh1"), "p2hh1" to the (first) wife in the main household, and "p3hh1" (if applicable) to the implied second wife, etc. Variables starting with "p1hh2" refer to the head ("p1") of the second or hosted household ("hh2"), "p2hh3" refer to the (first) wife ("p2") in the third household, and "p3hh4" refer to the second wife ("p3") in the fourth household, etc.

[1] Village 2 is missing, because it was dropped due to safety reasons.

IDH: Unique household code (household, individual and movement datasets). Each dwelling may have more than one household, but will always have only one "main household". IDH is IDD + number.

- Within column **HH**, HH1 is always the "main household", the subsequent household (HH2, HH3, etc.) are "hosted households".
- The household dataset is structured as follows: each row represents one household. "P1" is the head of the household. P2 is the spouse/wife. P3, P4 and P5 (if applicable) imply second, third and fourth wife, respectively.

IDI: Unique individual code (individual and movement datasets). IDI is IDH + number.

- Within the column **INDIV**: P1 is the head of the household. P2 is the spouse/wife. P3, P4 and P5 (if applicable) imply second, third and fourth wife, respectively.
- The individual dataset is structured as follows: each row represents one individual, therefore each survey question now only has one variable, instead of one for every individual (INDIV) in the household. This means that for all the individual-level variables, the 'prefix' "P1", "P2", "P3", "P4" and "P5" have been dropped. (E.g. p1_v1_name and p2_v1_name (in the household dataset) are replaced by v1_name in the individual and movement dataset.) In the cases of variables containing information about first language, gender and age, these prefixes have been replaced by "p_", to highlight that these variables are at the individual level.

IDM: Unique movement code (movement dataset only). IDM is IDI + number.

- Within the column **MOVE**: 1 is the first movement (i.e. from the place of birth to the second village), 2 refers to the second movement (i.e. from the second village to the third village), 3 refers to the third movement, etc.
- The movement dataset is structured as follows: each row represents one movement. Compared to the individual dataset, this means that for the variables containing information on movements, the variable names have been adjusted such that variables starting with "v1", "v2", "v3", etc. have dropped these prefixes and now begin with a mere "v", because their function has been taken over by the newly introduced column MOVE. Similarly, variables starting with "d1", "d2", "d3", etc. now start with "move_".

VARIABLES CONTENT

Information comes from two sources:

1. Data collected by the leader of the enumerator team about the village. Instrument is "MIG2012_village_survey_FR.pdf". This information starts with "vill_".
2. A survey was administered only to the head of the household (and if not present, his/her spouse): "MIG2012_dwelling_survey_FR.pdf". Information was asked about the head of the household, and his/her spouses, and the hosted families, if any.

We will go over some of the key variables now:

VILLAGE-LEVEL

- **vill_name**: Village name. This variable has been redacted to keep the village location anonymous.
- **vill_q2_nk1_name**: Name of the first "nyumba kumi". The nyumba kumi is a subchief in the village and responsible for ten households within a village. This variable has been redacted to keep the village location anonymous.
- **vill_q0_structure**: How does the village fit into the wider administrative structure? This variable contains descriptions of which smaller villages are part of the village, and/or whether it is part of a larger regional structure. This variable has been redacted to keep the village location anonymous.

- **vill_q0_how_arrived**: How did people arrive to the village? This variable contains descriptions of why many or most of the migrants arrived into the village. For example, because of safety concerns or because of family reasons. Answers have been documented in French.
- The surveyor was tasked to get a list of all the recent projects (that took place in the six months preceding the survey in the village) from the village chief, and to document information for each of these projects, which is contained in the following variables:
    - **Vill_q10_proj1_what**: What type of project took place?
    - **Vill_q10_proj1_lat:** What is the latitude of the project location? In DD. This variable has been redacted to keep the village location anonymous.
    - **Vill_q10_proj1_long:** What is the longitude of the project location? In DD. This variable has been redacted to keep the village location anonymous.
    - **Vill_q10_proj1_alt:** What is the altitude of the project location? In meters.
    - **Vill_q10_proj1_sal:** Was this project due to salongo? Salongo is a scheme/practice under which local chiefs on Saturdays request villagers to perform tasks for the community or work on projects (e.g. help build a school).
    - **Vill_q10_proj1_dollar**: Were people paid for the project?
    - **Vill_q10_proj1_ngo:** Was the project implemented by an NGO?
    - **Vill_q10_proj1_year:** In which year did this project take place?
    - **Vill_q10_proj1_month:** In which month did this project take place?

DWELLING-LEVEL

- **dwelling_longitude:** What is the longitude of the dwelling? In DD. This variable has been redacted to keep the village location anonymous.
- **dwelling_latitude:** What is the latitude of the dwelling? In DD. This variable has been redacted to keep the village location anonymous.

HOUSEHOLD-LEVEL

- **hh1_nom_chefmenage:** Name head of the household. This variable has been redacted.
- **hh_size_hh:** Household size. The household size includes the number of displaced people living in the household. For HH2 and beyond, this variable will therefore not return a value.
- **hh_maison_av** and **hh_maison_ap:** The number of houses before and after the arrival of displaced in the household. These two will be equal if (i) the main house is not a host family (ii) the main house is a host family, but has not constructed a separate house(s) for the hosted family.
- **hh_fem:** Number of wives the household head has. This was not asked to migrant households (i.e. HH2 and beyond).
- **hh_poul:** Number of chickens, turkeys and ducks in the household. This was not asked to migrant households (i.e. HH2 and beyond).
- **hh_ch_vo**: "If you have the means, would you like to buy a field in this village?" These two variables give information on any restrictions that the displaced households may face in buying fields in their destination village (the village chief may impose restrictions, for example).
- **hh_ch_pe**: "If you have the means, is it possible for you to buy a field in this village?" These two variables give information on any restrictions that the displaced households may face in buying fields in their destination village (the village chief may impose restrictions, for example).

Identity ranking exercise:

- The surveyor introduced this exercise to the household head using one of these two primes chosen randomly: In the university prime, the surveyor introduces the exercise as follows: "We are students. We work at the University of Columbia. We are not here for projects and we have nothing to distribute. This is only for purposes of scientific

research." In the NGO prime, the surveyor introduces the exercise as follows: "We are students. We work at the University of Columbia. You have no direct benefits. But we will share the results of this research with major organizations in Bukavu, for example the IRC, UNHCR and UNOCHA." Each household head was supposed to rank five identities from the most important to the least important.

- **hh_univ_ong**: whether the household head received the "university" or the "NGO" prime. These identities are (i) member of ethnic group xx, (ii) member of village xx, (iii) migrant or displaced, (iv) poor, or (v) Congolese. Individuals were randomly assigned to one of these two treatments. The respective variables are:
    - **hh1_groupe_ethnique**: From 1 to 5, how do you rank your identity as member of your ethnic group?
    - **hh1_membre_village**: From 1 to 5, how do you rank your identity as member of your village?
    - **hh1_deplace**: From 1 to 5, how do you rank your identity as a migrant or displaced?
    - **hh1_pauvre**: From 1 to 5, how do you rank your identity as being poor?
    - **hh1_congolais**: From 1 to 5, how do you rank your identity as a Congolese?

More information on Projects. The surveyor was tasked to get a list of all the recent projects (last six months) in a village from the village chief. For each of these recent projects, information was also collected from the household head (and his wife with the prefix P2, P3 and so on) on the following:

- **hh_proj1_what**: What type of project took place?
- **hh_proj1_sais**: Do you (household head) know about this project?
- **hh_proj1_qui**: Who from the household participated in the project?
- **hh_proj1_co**: Did you contribute in labor?
- **hh_proj1_cdollar**: Did you contribute money?
- **hh_proj1_cmat**: Did you contribute in material?
- **hh_proj1_sal**: Is this a *salongo* project? Salongo is a scheme/practice under which local chiefs on Saturdays request villagers to perform tasks for the community or work on projects (e.g. help build a school, clean the road in front of your dwelling, etc.).
- **hh_proj1_ong**: Is this a project implemented by an NGO?
- **hh_proj1_recudollar**: Did you receive money for your participation?
- **hh_proj1_month** and **hh_p1_year**: When did the project take place?
- **hh_proj1_c_reason**: Why did the household contribute?

INDIVIDUAL-LEVEL

Note that the questions were answered by the household head, but pertain to individuals within the household.

Migration histories of each person surveyed were recorded (household head and his wife/wives). In case the household head was not available, the wife was asked. The surveyor drew circles, each circle corresponding to a place where the individual stayed for at least one month. For each circle (location), information was collected on: the name of the village and chiefdom, the reason for moving, number of people that moved together, the number of people that arrived in the destination village, the year of movement. Information was also collected on fields and houses owned in each village. Each individual would thus start with his village of birth and speak about each village he moved to in his life. For example, the following data was collected for the household head:

- **p1_v1_name**: Name of the birth village of the household head. This variable has been redacted to keep the village location anonymous.
- **p1_v2_name**: The village he moved to from his birth village. This variable has been redacted to keep the village location anonymous.
- **p1_v3_name**, **p1_v4_name**, etc.: Subsequent villages he moved to. This variable has been redacted to keep the village location anonymous.

- **p1_ve_name**: The name of a village where the individual has not lived in, but where he does own fields. This variable has been redacted to keep the village location anonymous.
- For each of these villages, data was collected on:
    - **p1_v1_fields**: Number of fields owned in village 1
    - **p1_v1_field1**, **p1_v1_field2,** etc.: The year he obtained each field in village 1
    - **p1_v1_houses**: Number of houses owned in village 1
    - **p1_v1_house1**, **p1_v1_house2**, etc.: The year he obtained each house in village 1
    - **p1_d1_reason**: The reason for the movement to the next village
    - **p1_d1_nbrleave**: The number of people that left in total to the next village
    - **p1_d1_nbrarrive**: The number of people that arrived in total in the destination village
    - **p1_d1_year**: Year of the first move

PHASING IN OF QUESTIONS

A number of questions were added after the first village. For the relevant variables, this empty data is indicated (at least for a subset of all applicable variables) by "**NE**" (Not Exist).

- **q0**, **q2**-**q9**, **q14**-**q21**, **q39**-**q42**: from village 5 onwards
- **maison_av**, **maison_ap**, **five_yrs**: from village 5 onwards
- **fem**: from village 4 onwards
- **poul**, **Swahili**, **ch_vo**, **ch_pe**: from village 7 onwards
- **time**, **univ_ong** + variables related to the identity game: from village 6 onwards