The purpose of this dataset is to provide a detailed picture of the characteristics of Syrian towns in the years preceding the 2011 Syrian uprising and ensuing civil war. It incorporates the 2004 national census, the last before the uprising, and a newly collected set of data on ethnic identity. The level of analysis is the town (the Syrian Census Bureau's fourth administrative level[1]). The dataset was constructed jointly by Kheder Khaddour and Kevin Mazur, with funding from Princeton University's Mamdouha S. Bobst Center for Peace and Justice.

One important goal in assembling this dataset was to contextualize and add nuance to claims made about the role of ethnicity in Syrian political and social life. The ethnic composition of the Syrian population and government has been identified by scholars as an important factor in patterns of governance for decades; this is even more the case since the onset of the 2011 uprising. Yet ethnicity alone is not sufficient to explain patterns of governance, challenge, and conflict, nor is it a free-floating variable, disconnected from the other economic and social characteristics of an individual or local community. Understanding the ways in which ethnicity interacts with other forces to shape governance and challenge, therefore, requires a detailed picture of ethnic demography, and one that can be paired with other local-level data. By joining town-level data on ethnicity with official census figures, the Syria Town Database provides some of the data needed for this task.

We are making this dataset public in order to allow other researchers to benefit from the data, and to invite them to critique and improve the data. This is particularly the case for the ethnicity information; while we are presenting our best efforts to capture an accurate picture of Syria's ethnic demography in the 2000s, we fully realize that the database may have errors, and that there are many researchers who are in a position to provide more accurate information. For this reason, we view the project as an ongoing effort, and hope to continually update our coding where possible. Please contact us at kevin.a.mazur@gmail.com or kkhaddour@carnegie-mec.org with any questions, criticism, or updates you might have regarding our work.

Finally, we want to underscore that these data reflect the demographic composition of Syria at a specific point in time. The census was taken over half a decade before the uprising began, and displacement that has occurred during the uprising and ensuing civil war has permanently and, in some cases, dramatically altered the demographic composition of the towns in this database. The information presented here reflects Syria's pre-2011 settlement patterns; though it can be of assistance to researchers as a baseline for assessing the extent of change since the beginning of the uprising, understanding the modalities of change and subsequent demographic patterns requires additional data and detailed analysis. We intend to collect such data, on changes in demography over the course of the Syrian civil war and at its conclusion, and invite inquiries regarding this initiative.

The remainder of this document explains the contents of the database and coding procedures.

[1] Syrian administrative categorizations are as follows: the first level is the governorate (*muḥāfaẓa*, n=14), the second level is the district (*manṭiqa*, n=61), the third level is the sub-district (*nāḥiya*, n=270), and the fourth level is the town (*qarya*, n=5251). Official administrative categories and maps are taken from the UN Office for the Coordination of Humanitarian Affairs (OCHA). 2013. "OCHA's Common Operational Datasets for Syria." https://cod.humanitarianresponse.info/dataset/syrian-arab-republic-admin-level-1-boundaries-admin-level-2-boundaries-admin-level-3-0 (January 12, 2014).

2004 CENSUS DATA

The 2004 census data include basic demographic information (e.g. gender, age structure, education level, fertility), as well as data on employment, housing, and public goods provision. They are aggregated to the fourth-administrative level (the town, n=5204), and were taken from the Central Bureau of Statistics' website,[2] which has since been taken offline. Because these data did not come with a detailed codebook, we have left the variable names as we found them on the CBS website; an English translation of these names can be found in the third administrative level data from the United Nations Office for the Coordination of Humanitarian Affairs' (OCHA) website, which is included in this data package.

The census data were provided by CBS without geolocation information, and many of the place names do not match the standardized names used by international agencies (variations include non-standard spellings, typographical errors, and formatting issues). To make these data compatible with the ethnicity data and other sources, additional columns with spelling corrections to administrative levels were added, then merged to the official administrative classifications at the fourth level (taken from OCHA's official listing, included in this data package). Spelling corrections were done mostly by process of elimination within sub-districts (all governorates, districts, and sub-districts in the CBS data were easily identifiable, even if they required reformatting to correspond to the OCHA list). Merging was done on the basis of the corrected spellings, using the R statistical software. Not all OCHA entries corresponded to a CBS entry, and vice versa; the dropped units from each, with a justification, are listed at the end of this document.

ETHNICITY DATA

The ethnicity variables in the Syria Town Dataset describe the ethnic composition of town populations.

In our understanding, ethnic identities are those relating to social actors' claims of shared ancestry and the cultural practices they associate with those claims.[3] This means that we consider both religious/sectarian and national differences—the primary axes of descent-based identification in Syria—as facets of ethnic identification. Ethnicity, on this view, is the overarching category of which sect and nationality are sub-types. This does not rule out either form of identification having distinctive properties, but makes this an empirical question, rather than a difference of type that can be assumed ex ante.[4]

The characteristics of ethnic identification presented in this database do not constitute a theory of ethnicity in Syria. Such a theory might offer an explanation of when and how ethnicity plays a role in social and political life, why a certain categorical distinction became cognitively available to Syrians, or how an ethnic boundary shifted to include new members in the group. Instead, the ethnicity data provide the building blocks for thinking about these questions by identifying the distribution of towns' populations vis-à-vis ethnic boundaries and relating them to other social factors. Pairing these ethnicity variables with census figures, for example, can shed light on spatial variation in the

[2] Central Bureau of Statistics, Syria. 2004. "2004 National Census." http://www.cbssyr.org/indicator/hp-f.htm (June 20, 2013).
[3] For further explication of this understanding of ethnicity, see Wimmer, Andreas. 2008. "The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory." *American Journal of Sociology* 113(4): 970–1022.
[4] See Brubaker, Rogers. 2015. "Religious Dimensions of Political Conflict and Violence." *Sociological Theory* 33(1), especially p. 12.

provision of public services among a single ethnic group, or help compare the magnitude of variation within an ethnic group to that across sub-national regions. Statistics of this sort can provide a baseline from which to launch analyses of dynamic historical processes and local-level sequences of contentious events, rather than produce definitive statements about static, homogeneous groups.

Data collection procedure

The ethnicity data were collected on the basis of expert interviews, conducted between October 2013 and March 2014; respondents were asked about ethnic composition before 2011 (i.e. before any displacement occurred during conflict). Seventy-five interviews were conducted in Gaziantep and Antakya, Turkey, twenty in Amman, Jordan, twenty-five in Beirut, Lebanon, twenty in Istanbul, Turkey, and twenty via Skype.[5] Interview subjects were selected for their knowledge of a given area and located through networks of contacts from the relevant region. Interviews ascertained the ethnic groups living in a given town and their approximate share of the population, from a predetermined list of ethnic identities and percentage categories. Possible percentage values were: (1) >81%, (2) 50-80%, (3) 33-50%, (4) 11-32%, (5) 6-10%. Coders were instructed to record only groups with populations greater than five percent of a town's total population. Possible values for ethnic identity are described in the next section.

For several regions with which we were less familiar, we employed interviewers with local knowledge and contacts to locate relevant interview subjects and collect data. For one governorate, al-Hassaka, we were only able to collect information on the major towns and cities; rural areas are sparsely populated and subject to frequent land disputes stemming from historical state policy resettling Arabs in Kurdish areas. A local research organization, *al-Tajammuʿ al-watani li-l-shabab al-ʿArabi*, produced a town-level map[6] in 2012, using the methods very similar to those employed in our data gathering effort. For villages and small towns in al-Hassaka, we used this map as a basis for verification, rather than new coding. We acknowledge that any bias in this organization's data would be to overstate Arab presence in the area. However, we choose to present its data because we were unable to locate any similarly detailed alternate source, and its coding was consistent with ours for the towns on which we were able to gather data. We invite further revision in this area of the data, in particular.

Identity categories

Ethnic identities in Syria fall along both religious and national lines. National identities represented in the database include Arabs, Kurds, Syriacs, Armenians, Circassians, and Turkmens, and the religious identities represented include Sunnis, Alawis, Christians, Druze, Ismailis, and Shias. In addition, within the Sunni Arab identity, we distinguish between populations having tribal or family social structure.[7] We also created a category of 'mixed cosmopolitan', for recently built industrial

[5] This research was approved by the Princeton University Institutional Review Board (IRB). Data collection was carried out in accordance with Princeton IRB Protocol Number 6282.

[6] Downloaded from: Kulna Shurakaʾ. 2013. "Dirasa hama hawla al-tawazzuʿ al-sukkani fi mohafazat al-Hassaka al-suriya [Important study about population distribution in al-Hassaka governorate of Syria]." *All4Syria*. http://all4syria.info/Archive/93123 (June 3, 2014).

[7] This distinction was developed inductively, based upon our preliminary open-ended interviews. Our interlocutors differentiated between communities organized along 'tribal (*ʿashāʾirī*)' and 'family (*ʿāʾilī*)' lines, describing the former as maintaining a wider net of extended family relations and identifying with the tribe's name, and the latter as denying or downplaying such an identification in favor a smaller family unit or a specific town identity. We recognize that there is a blurry line between these two categories. A more sustained investigation of tribal identification might conceptualize the

towns and factory housing complexes that have no clear ethnic majority, and where no religious or national identity is prominent in public space and public life. These identity categories, taken together, partition the ethnic space.[8] The possible values a population within a town can take are as follow, with the abbreviation code used in the "ethn_detail" field of the database:

Code	Group	Code	Group
al	Arab Alawi	rm	Armenian
ch	Arab Christian	sf	Arab Sunni families
ci	Circassian	sh	Arab Shia
dr	Arab Druze	st	Arab Sunni tribal
is	Arab Ismaili	sy	Syriac
ku	Kurdish	tk	Turkmen
mx	Mixed cosmopolitan		

Validation

Before discussing the limitations and weaknesses of our database, it is important to emphasize both the paucity of existing research of this kind and the difficulty of assembling this sort of information. The Syrian government does not report any information about ethnicity in official statistics and denies the importance of ethnicity to social life. Extant publicly available sources describe general regions, rather than specific towns, frustrating efforts at systematic comparison of ethnic versus other factors in routine social and political life, as well as in moments of contention. We thus believe that the Syria Town Database constitutes an improvement over currently available data sources on Syrian ethnicity.

Nonetheless, we recognize that our data collection strategy imposes certain limitations on the precision and accuracy of our data. Our survey relied on experts—some covering villages in more than one sub-district—and we asked the respondents to estimate the proportions of town residents of each ethnic identity.

We have taken several steps to address potential biases arising from this aspect of our data collection strategy ensure the validity of our data. First, we created a simpler coding scheme that reflects only the ethnic identity of a numerical majority of a town's residents and collapses the thirteen ethnic categories enumerated above into five. These data are in the "emaj5cat" column.

---

difference as a spectrum—a great number of Syria's contemporary rural populations were nomadic or semi-nomadic until the last several generations, and thus retain social structures and identifications characteristic of such populations. For the purposes of town-level classification at a national scale, however, we believe that this tribe/family distinction is defensible because there remain perceptible differences between town and village residents organized on each basis. Whereas Hanna Batatu observes that tribal identification diminished in the Darʿa countryside to the point that clan-based residential clustering largely ended, intramarriage norms broke down, and, for example, Al Miqdad clan failed to appoint a new chief, our interview subjects from Dayr al-Zur countryside had largely preserved clan-based residential and marriage patterns and continued to identify in tribal terms (e.g., as a member of al-Bulayl sub-tribe of al-ʿAgidat tribe) rather than with their town of origin. See Batatu, Hanna. 1999. *Syria's Peasantry, the Descendants of Its Lesser Rural Notables, and Their Politics*. Princeton, N.J.: Princeton University Press, p. 25-6, and compare with al-Mashhour, Faisal Dahmoush. 2017. *Abnaʾ al-ʿashaʿir fi Dayr al-Zur min al-istiqrar ila al-thawra, dinamikiyyat al-siraʿ wa-ʿawamil al-silm al-ahli [Sons of the tribes in Dayr al-Zur from stability to revolution, dynamics of struggle and factors of civil peace]*. Dayr al-Zur: Justice for Life Organiation. http://jfl.ngo/?p=4602 (September 11, 2018), pp. 4-5, 60-61.

[8] We acknowledge that these categories elide some identities into other categories (ex. Murshdis are included in the Alawi category) and exclude some small groups entirely (ex. Afghan and Somali refugees); these coding choices reflect our goal of developing a meso-level picture of ethnicity, at the town level, rather than national, regional, or neighborhood level.

Second, we used work carried out by several other researchers to check our coding. We employed a map produced and shared privately by a researcher working for an international organization as the primary basis for comparison. Work by Khalifa[9] and an online forum listing Druze towns[10] were used for further validation. These maps, taken together, contained data for 1466 of the 5204 towns in the database (covering 72 percent of the total population). To produce an alternate coding for robustness testing, all codings were viewed alongside one another and the majority identity of a town was changed from that in the author's database if contradicted by one of the sources; if one contradicted and another agreed with the original coding, the original coding was retained. This led to a recoding of 146 towns (in which 2 percent of the total Syrian population resided). The towns recoded were primarily small villages; the largest town to be recoded had 30,519 residents, ten recoded towns had between 10,969 and 20,301 residents, all others had fewer than 10,000, and seventy had fewer than 1000 residents. These data are presented in the "eth5recode" column.

Finally, to facilitate further scrutiny and refinement of the database, we have also uploaded the raw database, separating out each ethnic group and the percentage of the population it represents, according to our survey, for each town in Syria. These data can be edited by other researchers, based upon their own inquiries, and used to improve on our work. We encourage researchers undertaking this task to be in communication with us about their work and its results.

[9] Khalifa, Mustafa. 2013. *The Impossible Partition of Syria*. Arab Reform Initiative. http://www.arab-reform.net/en/file/562/download?token=DImggNO9 (November 26, 2017).
[10] http://raia.ba7r.org/t1793-topic, accessed 22 January 2018.

DESCRIPTION OF ENCLOSED FILES

towndb_dataverse_12oct18.csv – This is the main file containing the Syria Town Database. The first twelve columns provide the administrative codes and Arabic and English names of locales. Columns 13-179 contain the 2004 census data.[11] Ethnicity data are in the last three columns of the file (180-182), comprised of the following: (1) a single line summary of the detailed ethnicity coding ("ethn_detail"), (2) a five-value variable indicating a town's majority ethnic group ("emaj5cat"), (3) the alternate ethnicity coding, based upon sources described in the *Verification* section above ("eth5recode"). Because the governorate of Quinetra is largely under Israeli occupation, towns in the governate are not included from this dataset.

Syria-gazetteer-2013-05-19.xls – This is a listing of all administrative units in Syria, downloaded from UN OCHA.

ocha_04census_l3.csv – This is the third level census data, downloaded from UN OCHA.

2004census_l4_spellchecked.csv – This is the fourth administrative level census, as downloaded from the CBS website. Four columns are added to standardize spelling of place names with the UN OCHA Gazetteer. "lX_unchanged" lists the administrative names as they appear in the file downloaded from CBS (with 'X' indicating the relevant administrative level), and "lX_okmerge" columns are those in which changed the spelling and formatting from CBS were modified manually to match up with UN OCHA spellings.

ethnicitydb_raw_8oct18.csv – This file expands the ethnicity coding detail in the "ethn_detail" column of the main file, giving each group and percentage a column to enable modification of our coding.

[11] Additional information on variable names, in English, is available in the third administrative level file from OCHA.

CENSUS MERGING DETAILS

PCODES in administrative data with no corresponding entry in the census (plus justification for dropping):

C5244 Dreamland is unbuilt area, just a plan.
C5245 Sarb Kawkab is not found on Google and near other areas of sparse inhabitation in al-Hassaka.
C5246 Looks to be in Dayr al-Zur, town on river but nothing comes up in Google.
C5247 Darʿa al-Balad (old core of Darʿa city), is not broken out separately in census data, so assumed to be included in Darʿa city figures.
C5248 Hamam Turkman is a historical mountain part of Tel Abyad, but is not inhabited per Wikipedia.
C5249 Jabila – nothing turned up on Google search.
C5250 Huseyiniya – nothing turned up on Google search.
C5251 Talil - in Taldo sub-district, but nothing comes up on Google.

Census towns with no PCODE (plus justification for dropping):

Hamdaniya – Google search turns up only one mention in official Syrian government site, no other information.
Ratla – Google maps locates next to Raqqa city, but no PCODE.
Masaada – nothing found on Google maps or Google search.
Aaliya – nothing found on Google maps or Google search.
Zahra – nothing found on Google maps or Google search.