

SCIENTIFIC DATA



OPEN

DATA DESCRIPTOR

ClimActor, harmonized transnational data on climate network participation by city and regional governments

Angel Hsu¹✉, ZhiYi Yeo¹, Ross Rauber², James Sun³, Yunsoo Kim¹, Sowmya Raghavan¹, Nicholas Chin¹, Vasu Namdeo¹ & Amy Weinfurter¹

Cities and regions have become increasingly engaged in global climate change governance. They are pledging their own climate mitigation targets and participating in membership networks that typically are transnational in nature and engage thousands of subnational governments. Researching these growing trends in participation has been difficult due to the disparate and inconsistent nature of this self-reported data. To facilitate future analyses of these actors, we introduce ClimActor, the largest harmonized global dataset of more than 10,000 city and regional governments participating in networks like the Global Covenant of Mayors for Climate and Energy, C40 Cities for Climate Leadership, ICLEI Local Leaders for Sustainability, among others. We include key contextual information on each actor's population, geographic location, and administrative jurisdiction to facilitate disambiguation of potential overlaps in actions or emissions. We also provide a series of cleaning functions based on phonetic and fuzzy string matching algorithms within an open-source R package to make it easy for anyone to immediately use the ClimActor dataset with other relevant data.

Background & Summary

Cities and regions- defined as subnational administrative units generally broader in population and area than cities and the first administrative level below national governments¹ - are increasingly engaged in global climate change governance. They hold the majority of the world's population and are at high risk for experiencing the impacts of climate change². These subnational actors are pledging their own largely voluntary climate actions, which include setting emissions reduction targets, increasing renewable energy consumption, and financing for energy efficiency and other technology upgrades³. While some of these actions support implementation of national climate policies⁴, many subnational actors participate in transnational climate initiatives that engage actors working across borders towards shared goals^{5,6}, and often in collaboration with national governments¹. The EU Covenant of Mayors for Climate and Energy, for example, includes more than 9,000 primarily small cities in Europe (population less than 50,000) to commit to emission reduction targets more ambitious than the European Union's own pledges⁷. Participating in these networks creates opportunities for actors to enter into dialogues with other cities and to pool their global influence as critical sustainability actors^{8,9}. They can also provide benefits spanning agenda setting, information sharing, and capacity building¹⁰. Cities may join several initiatives to participate in multiple dialogues with slightly different participants, focus areas, or reputational benefits, particularly if the cost of entry or membership is low.

While the number of cities and regions participating in these transnational climate networks has risen dramatically over the last few decades^{11,12}, analyzing their collective impact has been challenging, largely due to lack of available data on the impact of subnational actors' climate actions and incomplete and heterogeneous data, when they are available^{13,14}. A recent survey of subnational climate actors that participate in reporting platforms found that just 58 percent had sufficient information to quantify their potential mitigation impact¹⁵. Further complicating the aggregation and estimation of these actors' contributions to global climate mitigation is the difficulty in

¹Yale-NUS College, 16 College Ave W, Singapore, 138609, Singapore. ²University of Chicago, Department of Computer Science, 5730 S. Ellis Ave, Chicago, IL, 60637, USA. ³Yale College, Department of Statistics and Data Science, 24 Hillhouse Ave, New Haven, CT, 06511, USA. ✉e-mail: angel.hsu@yale-nus.edu.sg

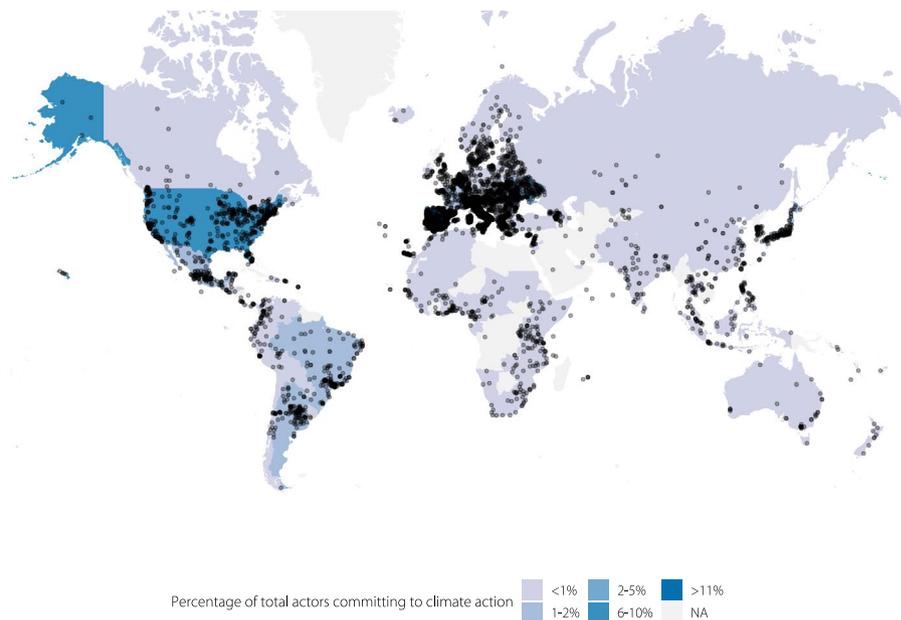


Fig. 1 Global distribution of subnational actors committing to climate action platforms included in the ClimActor dataset and R package.

determining overlaps between actors: city governments embedded within states often commit their own climate actions without coordinating with higher levels of jurisdictions, which could result in double counting of emissions reductions or other activities if not appropriately accounted for^{14,16,17}. Due to these challenges, most studies evaluating subnational climate actions are focused on a small number of cities with available data^{18,19} or localized to a specific region²⁰.

To obtain a full global picture of which subnational governments are taking climate actions as reported through primarily transnational climate initiatives, such as the Global Covenant of Mayors for Climate and Energy, C40 Cities for Climate Leadership, and ICLEI Local Leaders for Sustainability, this paper shares what we believe to be the most comprehensive and harmonized database of subnational climate action network participation and contextual information to date. It is intended to underpin research efforts to build on and extend analyses on the scope and impacts of subnational climate action. To create it, we collated data from multiple sources (see Online-only Table 1) and developed an R package (“ClimActor”) that provides a series of functions to aid in cleaning, matching and mapping these disparate datasets and enables users to match this information with other datasets (i.e., on emissions, commitments, etc.) in a standardized way. The ClimActor subnational climate action dataset includes more than 10,000 subnational actors’ location, population, area and participation in climate action networks (Figs. 1–3).

The ClimActor dataset helps navigate several key challenges and considerations around the analysis of subnational climate action. Data platforms often represent subnational actors in disparate ways: New York City, for example, can appear as New York, NY, New York, or the city of New York on multiple platforms but refer to the same city of more than 8 million people. Collecting information on which actors have signed up to various pledging platforms and then merging these data with critical information on population or emissions is then challenging because actor names must be cleaned and standardized to ensure actors’ data are correctly matched. Actor names can be misspelled or subnational governments’ could have similar names (i.e., London could refer to London, United Kingdom or London, Ontario or even London, Kentucky). The process of disambiguating these names and matching datasets can be incredibly laborious and time-consuming²¹. To address these obstacles, the ClimActor dataset compiles data on subnational climate network participation and relevant contextual data, enabling researchers to explore how climate action participation and trends vary within and across different geographic and economic contexts (Fig. 3)^{15,22}.

The dataset also addresses the challenges of disentangling overlaps and hierarchies in subnational climate action data. Identifying overlaps between different levels of jurisdiction and avoiding instances of “double-counting” is a vital part of many climate action analyses¹⁴. Identifying the links between local, regional, and national governments also aids efforts to explore the nuances of subnational climate action within specific national governance contexts⁶. The ClimActor database lists the networks an actor participates in as well as its higher-levels of political administration, making it possible to understand overlaps in climate commitments. This information can be used to avoid double-counting commitments reported across several different platforms and to facilitate explorations of subnational engagement with transnational climate initiatives. It can also help determine what commitments are not subsumed within a nested jurisdiction but instead in addition to (i.e., “additional”) to what a higher level of government has pledged, if any (i.e., if New York City and New York State both make climate pledges, understanding their nested relationship would allow for an analyst to make an evaluation of whether actions at the lower level are truly in addition to what a higher level of administration has pledged)¹.

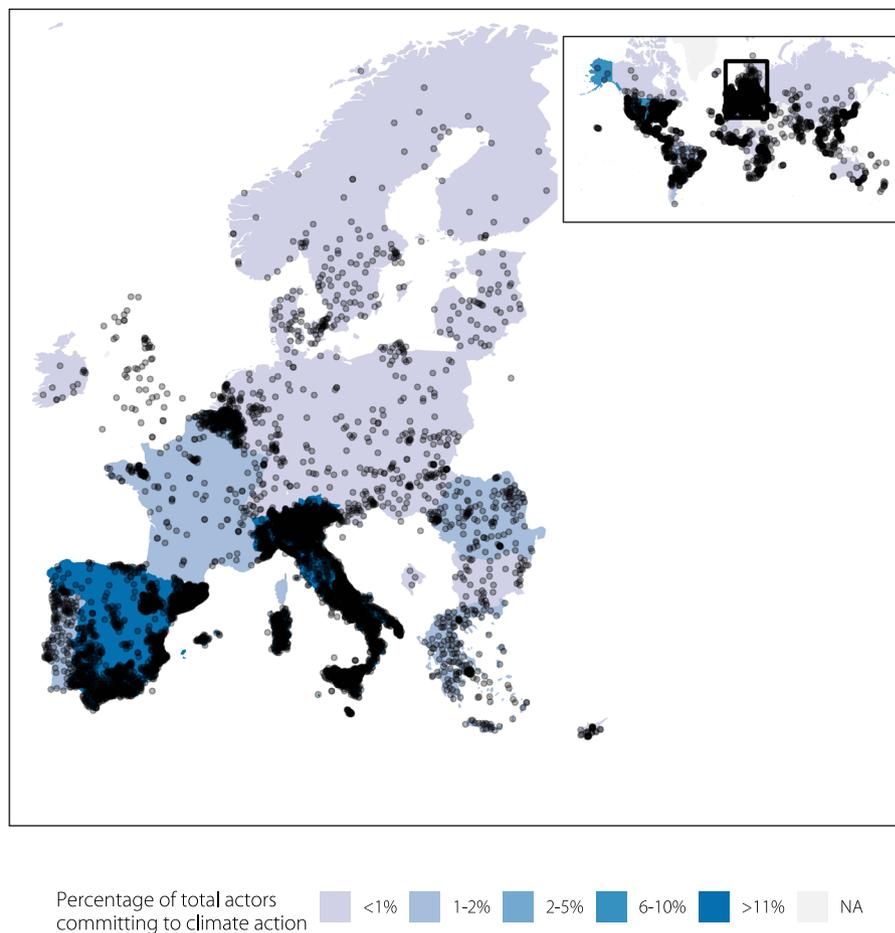


Fig. 2 Most subnational actors committing to climate action in the ClimActor dataset are concentrated in Europe, with Italy representing the highest number of actors ($n = 4,201$).

Methods

Participation data. Membership data for subnational actors participating in climate action networks (Online-only Table 1) such as the carbonn Climate Center and EU Covenant of Mayors for Climate and Energy initiatives and UNFCCC's Global Climate Action Portal were collected from the respective websites using Python v3.6.5²³ and the BeautifulSoup v4.6²⁴ and json²⁵ packages. Generally, the workflow to collect actor data from these websites is as follows: a complete list of actors within the database is first extracted from the site and if other relevant information is available for an actor (e.g., location, population, etc.), that information is collected and matched to an actor's name. Data are then cleaned using the pandas package in python²⁶ before being stored as csv files. Online-only Table 1 lists the complete climate action networks that were collated for the ClimActor dataset.

Contextual data. If a climate action platform or database does not include contextual information on a particular actor, such as population, location, area, etc., this data is collected from Wikipedia's English language pages. Although a crowdsourced website and information repository, Wikipedia has been shown to be an accurate source of political data and information²⁷. Wikipedia is accessed through the MediaWiki API using Python, and the BeautifulSoup, json, and requests²⁸ packages. Each actor in the database is minimally associated with a name, entity type (either City, Region, Company, or Country), and ISO country code, each of which is used to resolve which Wikipedia page corresponds to an actor in question. This procedure begins by passing an actor's name to the MediaWiki API's search function, which returns a list of potential matches ordered by relevance. The top 20 pages are each downloaded and prepared for further filtering by mapping each page to a likely entity type and country.

To ensure the correct Wikipedia page is selected for a subnational actor, the Wikipedia categories to which a page belongs are fed to a function that counts the occurrences of certain words that likely correspond to a specific entity type (e.g., words like municipality, town, capital, ward, and village are more commonly found in pages that should be in the City category). If the word "disambiguation" occurs in a page's categories, the page is taken to be a Disambiguation page²⁹. For instance, an actor in question that could have a name that refers to something else (e.g., Golf, Illinois could be mistaken for the sport of golf if no other contextual information is included). Otherwise, the entity type with the highest count of matching words is then taken to be the page's entity type. If no words match, the page is assigned no entity type.



Fig. 3 Population of subnational actors (cities and regions) recording climate action commitments as captured in the ClimActor dataset, grouped by region.

To infer a page's country, the page's "info-box"—the light grey table on the right of many Wikipedia pages—is searched for links to Wikipedia pages corresponding to the countries on the site's list of countries assigned an ISO code (ISO, 2020). The first of these links found in the infobox is taken to be the one corresponding to the country the page should be assigned to. If no links are found within the infobox, the page's body text is then searched according to the same procedure.

The list of 20 search results is then filtered such that only pages matching the actor in entity type and ISO code remain. The first of these pages—deemed by the Wikipedia search function to be most relevant—is then selected as the best candidate and is mined for contextual information according to the procedure described in the following paragraph. If no pages remain after filtering the top 20 for entity type and ISO code, the results are checked for any disambiguation pages. If one is found, then each of the links in the disambiguation page are processed and filtered as if they were the results of the search above, the first matching page being selected if multiple matches are found. If still no matches are found, then the process is repeated starting from the call of the search function including an actor's ISO code in the search query. If still no matches are found, the actor is not resolved to a page.

If a page is found during this process, contextual information is extracted from the page's info-box, a table containing rows usually several sections indicated by a single column spanning the entire table, each of which contains several related rows with two columns, the left one describing the data found in the right column. For example, under the section "population," a row may be found containing two columns with values "Metro" and some number respectively. This information is used to find relevant population and figures, matched with corresponding years where possible. In addition to this data, latitude and longitude are collected in decimal degrees where possible. To account for potential overlaps between city and region actors, the latter which often encompass the former, we also collect information on city actors' higher jurisdictions. For example, Manhattan (borough) is located in New York City (city) within New York (state). This information can be used to avoid double counting of climate actions, although there are some manual disambiguations that are still needed. Wallonia Region in Belgium, for instance, is a higher administrative unit than other regional actors, such as Antwerp Province and Brussels Capital Region, and encompasses five other provinces within Belgium (Liège, Hainaut, Luxembourg, Walloon Brabant, and Namur). Avoiding overlaps between Wallonia Region and other city or municipality actors within Belgium would need to account for overlaps within these provinces located in Wallonia as well. In some cases, regional councils commit to climate action and encompass several regions and cities, such as the Assembly of Regions of Cote d'Ivoire (ARDCI).

Because we only scraped English-language Wikipedia pages, we acknowledge that some contextual information is missed through Wikipedia pages in other languages (see Technical Validation:Data Limitations). Other data sources for country-level contextual data (i.e., population, land area) were obtained from the sources listed in Online-only Table 2.

Data Records

The ClimActor package includes three main datasets: subnational climate action participation and accompanying contextual information for each actor (stored as "contextuals"); a key dictionary ("key_dict") of subnational actor standardized and unstandardized names for purposes of string matching (see 'Usage Notes' below); and a country dictionary ("country_dict") for cleaning country names associated with each actor that is required for the cleaning functions in the ClimActor package. The datasets are available on figshare³⁰, and are also stored as Rdata objects and automatically loaded when the ClimActor package is loaded within a user's R environment. They are also available as CSV files and can be downloaded from the ClimActor package's GitHub page (www.github.com/datadrivenenvirolab/ClimActor). An overview of the included information within the datasets is provided in Online-only Table 2.

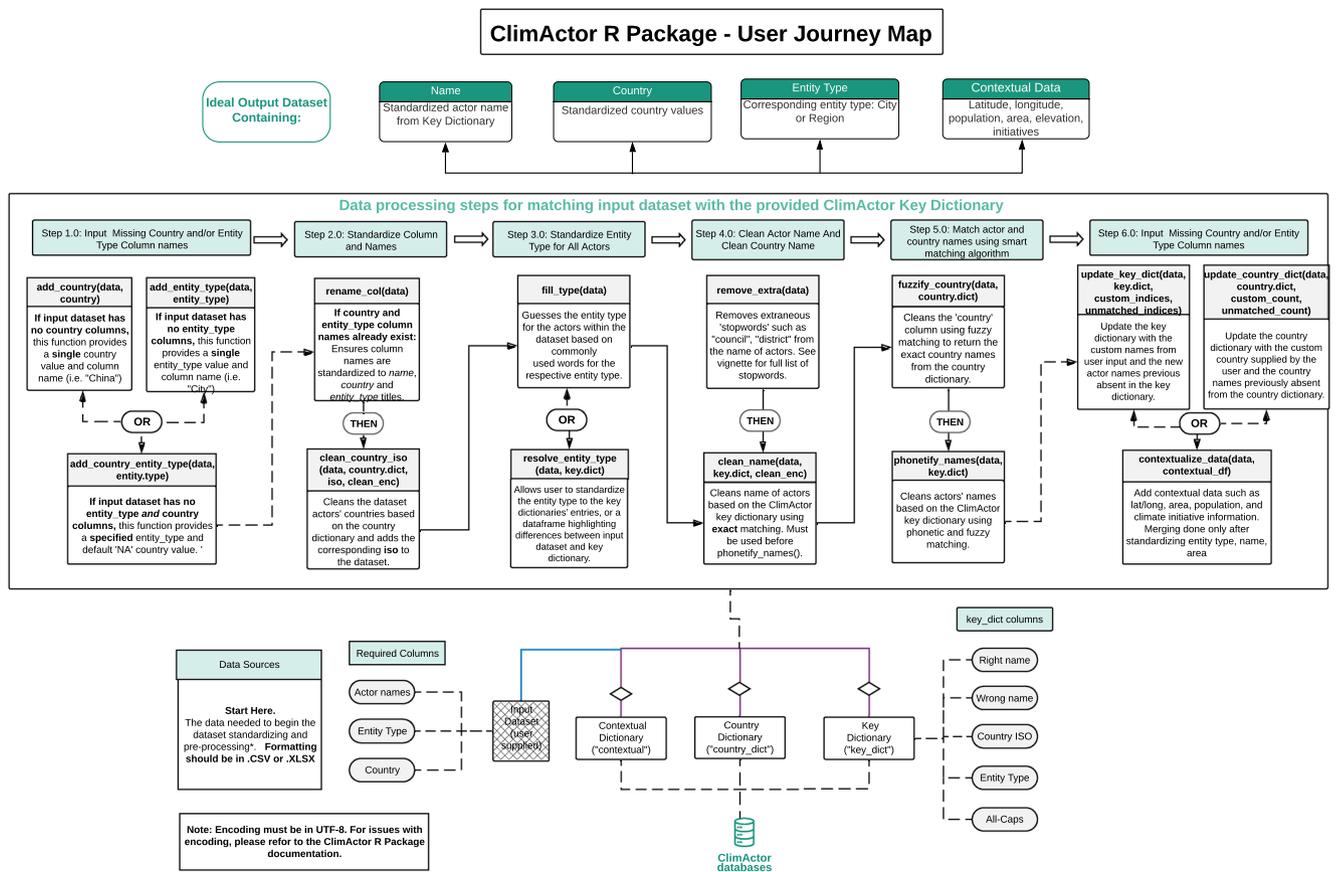


Fig. 4 Map of the ClimActor R package that describes the process of matching and combining a user's dataset with the ClimActor database.

Technical Validation

We performed validation checks primarily on the self-reported population and geographic location data from subnational actors via climate action reporting platforms. Platforms such as CDP and the Global Covenant of Mayors for Climate and Energy include population data, although sometimes the year associated with the data are not included and require validation. Because actors may simultaneously report to multiple climate action platforms at once and report conflicting data, we compared and validated the data to ensure the database reflects the most recent and accurate data available. Below we describe the validation checks applied to self-reported population and geographic location data.

Population data. In instances where subnational actors report to multiple climate action platforms and different population data are collected from these various sources, we compared the self-reported population numbers with each other. If the values varied within five percent of each other, we selected the first entry of that actor within the database. If the values varied more than five percent of each other, we prioritized the following data sources' population data: 1) CDP (formerly known as the Carbon Disclosure Project), as they require cities and regions to report on an annual basis; 2) Global Covenant of Mayors for Climate and Energy or EU Covenant of Mayors for Climate and Energy; 3) Carbon n Climate Registry. For actors in which no population data was reported to these three sources, population data was obtained from Wikipedia. For actors where population data was available but where the year corresponding to the population data was missing, the year corresponding to the population data was estimated by comparing the population data with the actor's Wikipedia page. For example, if an actor's population data matched the number on its Wikipedia page, we imputed the corresponding population year with the year listed on an actor's Wikipedia page.

Geographic location data. In instances where a subnational actors' geographic location data (longitude, latitude) were provided, we validated by plotting all actors' coordinates on a map (Fig. 1). A global polygon of countries' boundaries were used to identify potentially erroneous coordinate information. Outliers were then visually inspected and the geographic coordinate information was removed and replaced with Wikipedia's geographic coordinates for that actor if they fell outside the bounding box.

Data limitations. Our dataset is limited by several constraints. First, data are self-reported by subnational actors that participate in transnational climate initiatives. While there are most certainly subnational governments outside of these initiatives that are taking climate actions, we are unable to systematically capture and assess

Phonetic Algorithm Name	Phonetic Representation for “City of New York”	Phonetic Representation “New York”	Similarity Score
Caverphone	STYFNW	NWYK11	0.38
Caverphone Modified	STFNWK1111	NWK1111111	0.68
Cologne	8236374	6374	0.65
Lein	C142	N350	0
Metaphone	STYFNRYK	NYRK	0.60
New York State Identification and Intelligence System (Nysiis)	CATYAF	NAYARC	0.57
Nysiis Modified	CATAFN	NARC	0.49
Oxford Name Compression Algorithm (Onca)	C315	N620	0
Onca Modified	C315	N620	0
Onca Refined	C306	N809	0.28
Onca Modified Refined	C306	N809	0.28
Phonex	C315	N200	0
Roger Root	182	2470	0.15
Soundex	C315	N620	0
Soundex Refined	C306028093	N8093	0.51
Statistics Canada Name Coding	CTFN	NWRK	0.13

Table 1. Different Phonetic Representations of alternative names for New York City, NY (“City of New York” and “New York”) and their similarity scores.

them due to the unwieldy nature and infeasibility of collecting every instance outside of common networks and reporting platforms we used in Online-only Table 1¹⁴. Second, our data collection efforts were primarily limited to English-language sources, although there are certainly non-English language websites and data sources we can research for future iterations of the database. The use of translation websites and application programmer interfaces (APIs) and social media text mining, combined with natural language processing (NLP) techniques, could also aid in further expanding the ClimActor database to include more non-English actors and those not currently reporting to the largest and most common networks available.

Usage Notes

We developed the ClimActor R package to assist researchers in utilizing our subnational climate action database alongside other datasets. For example, researchers may have data on a subset of subnational actors’ carbon emissions or GDP data that they wish to analyze alongside our dataset. Ensuring actors’ names are consistent (i.e., New York could refer to either New York City or New York state), “clean” (i.e., one dataset could refer to Dallas, TX as simply “Dallas” or “City of Dallas” while referring to the same actor), and disambiguated (i.e., the city of London could refer to London, Ontario in Canada, London, Kentucky in the United States or London, United Kingdom) can often present challenges in analysis. Data from different languages also may include character encodings, accent marks, or various spellings (e.g., Venice and Venezia), which the ClimActor package assists in disambiguating and standardizing. The ClimActor package streamlines the data cleaning process by:

- providing a “key dictionary” of consistent subnational actor names and contextual information (e.g., population, geographical coordinates, etc. whenever these data were available; see above Data Records);
- providing a “country dictionary” of consistent country names (i.e. South Korea could also be referred to as “Korea” or “Republic of Korea”; see above Data Records);
- a set of basic data preparation functions to ensure that the dataset in question has the columns and entries needed to work with the subsequent standardization functions;
- a set of data standardization functions to ensure standardized versions of accessory information (entity type, country, ISO code) that allows further functions to standardize actor names;
- a set of string matching functions (using a mixture of exact, phonetic, and fuzzy string matching) to standardize actors’ names based on our collected database of subnational actors;
- a set of post-cleaning functions to update and expand the key dictionary (with novel actors in the user’s database) and to include contextual information (i.e., data records listed in Online-only Table 2) from our database to the user’s dataset.

The string matching functions are based on phonetic matching algorithms and string distance measures compiled in the phonics R package³¹ and stringdist R package³² respectively. The process reduces a character string (i.e., a subnational government actor’s name) to a symbolic representation that approximates its English pronunciation before comparing string distance between these measures. This reduced form is then used for matching and avoids complications when there may be common words that increase the length of an actor’s name (i.e. Prefeitura de Pedreira and Prefeitura de Pedra Bela) or special characters such as accent marks that may complicate fuzzy string matching³³. Five different phonetic matching algorithms are used in the package

(Metaphone, Nysiis modified, Onca modified refined, Phonex, and Roger Root). These algorithms were selected after evaluating all 16 (nine core and seven modified) that were included in the phonics R package and selecting the five that consistently yielded the most accurate match results in our validation. To validate the phonetic algorithms, the 16 different phonetic representations were first created for all entries in the ClimActor key dictionary. Representations were then visually inspected for duplication, before string similarity measures (based on the average of the Full Damerau-Levenshtein distance, q-gram distance, cosine distance q-gram profiles, Jaccard distance between q-gram profiles, and the Jaro-Winker distance) were used to calculate string similarities between alternative names for entries in the key dictionary (See Table 1 for an example). For detailed descriptions of these algorithms and their respective trade-offs, see Howard (2019)³¹.

A more detailed user guide outlining the workflow of using the ClimActor package is visible in Fig. 4 and also available at: <https://github.com/datadrivenenvirolab/ClimActor>. Because the landscape of subnational climate action changes almost daily, the online repository will be updated periodically.

Code availability

Data and code for the ClimActor R package functions is available on GitHub: <https://github.com/datadrivenenvirolab/ClimActor>.

Received: 17 June 2020; Accepted: 11 September 2020;

Published online: 06 November 2020

References

- Kuramochi, T. *et al.* Beyond national climate action: the impact of region, city, and business commitments on global greenhouse gas emissions. *Clim. Policy* **20**, 275–291 (2020).
- Revi, A. *et al.* Urban areas. In *Climate Change 2014 Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Field, C.B. *et al.*) Ch. 8 (Intergovernmental Panel on Climate Change, 2015).
- Palermo *et al.* Assessment of climate change mitigation policies in 315 cities in the Covenant of Mayors initiative. *Sustain. Cities Soc.* **60**, 102258 (2020).
- Reckien, D. *et al.* How are cities planning to respond to climate change? Assessment of local climate plans from cities in the EU-28. *J. Clean. Prod.* **191**, 207–219 (2018).
- Andonova, L. B., Betsill, M. M., & Bulkeley, H. Transnational climate governance. *Global environmental politics*, **9**, 52–73 (2009).
- Andonova, L. B., Hale, T. N. & Roger, C. B. National policy and transnational governance of climate change: Substitutes or complements? *Int. Stud. Q.* **61**, 253–268 (2017).
- Kona, A., Bertoldi, P., Monforti-Ferrario, F., Rivas, S. & Dallemand, J. F. Covenant of mayors signatories leading the way towards 1.5 degree global warming pathway. *Sustain. Cities Soc.* **41**, 568–575 (2018).
- Toly, N. J. Transnational municipal networks in climate politics: From global governance to global politics. *Globalizations* **5**, 341–356 (2008).
- Erickson, P., Lazarus, M., Chandler, C. & Schultz, S. Technologies, policies and measures for GHG abatement at the urban scale. *Greenh. Gas Meas. Manag.* **3**, 37–54 (2013).
- Bulkeley, H. *et al.* Governing climate change transnationally: assessing the evidence from a database of sixty initiatives. *Environ. Plan. C Gov. Policy* **30**, 591–612 (2012).
- Castán Broto, V. & Bulkeley, H. A survey of urban climate change experiments in 100 cities. *Glob. Environ. Chang.* **23**, 92–102 (2013).
- Gordon, D. J. & Johnson, C. A. The orchestration of global urban climate governance: conducting power in the post-Paris climate regime. *Env. Polit.* **26**, 694–714 (2017).
- Intergovernmental Panel on Climate Change. Human Settlements, Infrastructure, and Spatial Planning. In *Climate Change 2014 Mitigation of Climate Change* (2015).
- Hsu, A. *et al.* A research roadmap for quantifying non-state and subnational climate mitigation action. *Nat. Clim. Chang.* **9**, 11–17 (2019).
- NewClimate Institute, Data-Driven Lab, PBL, German Development Institute/Deutsches Institut für Entwicklungspolitik (DIE), Blavatnik School of Government, U. of O. *Global climate action from cities, regions and businesses: Impact of individual actors and cooperative initiatives on global and national emissions.* http://datadrivenlab.org/wp-content/uploads/2019/11/Report-Global-Climate-Action-from-Cities-Regions-and-Businesses_2019.pdf. (2019).
- Roelfsema, M. Assessment of US City Reduction Commitments, from a Country Perspective. PBL Netherlands Environmental Assessment Agency. *PBL Policy Brief* (2017).
- Hsu, A., Weinfurter, A. J. & Xu, K. Aligning subnational climate actions for the new post-Paris climate regime. *Clim. Change* **142**, 419–432 (2017).
- Kennedy, C., Demoullin, S. & Mohareb, E. Cities reducing their greenhouse gas emissions. *Energy Policy* **49**, 774–777 (2012).
- Khan, F. & Sovacool, B. K. Testing the efficacy of voluntary urban greenhouse gas emissions inventories. *Clim. Change* **139**, 141–154 (2016).
- Croci, E., Lucchitta, B., Janssens-Maenhout, G., Martelli, S. & Molteni, T. Urban CO₂ mitigation strategies under the Covenant of Mayors: An assessment of 124 European cities. *J. Clean. Prod.* **169**, 161–177 (2017).
- Rahm, E. & Do, H. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.* **23**, 3–13 (2000).
- Hsu, A., Cheng, Y., Weinfurter, A., Xu, K. & Yick, C. Track climate pledges of cities and companies. *Nature* **532**, 303–306 (2016).
- Python Software Foundation. Python Language Reference, version 2.7. *Python Software Foundation*, <https://www.python.org/> (2013).
- Richardson, L. Beautiful soup documentation. *Dosegljivo*: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (2007).
- Sphinx. JSON encoder and decoder — Python documentation. *JSON encoder and decoder* (2016).
- McKinney, W. pandas: a Foundational Python Library for Data Analysis and Statistics. *Python High Perform. Sci. Comput.* **14**, 1–9 (2011).
- Brown, A. R. Wikipedia as a Data Source for Political Scientists: Accuracy and Completeness of Coverage. *PS - Polit. Sci. Polit.* (2011).
- Chandra, R. V., & Varanasi, B. S. *Python requests essentials.* (Packt Publishing Ltd, 2015).
- Cucerzan, S. Large-scale named entity disambiguation based on Wikipedia data. In *EMNLP-CoNLL 2007 - Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* 708–716 (2007).
- Hsu, A. *et al.* ClimActor, a harmonized transnational dataset of climate network participation by city and regional governments. *figshare* <https://doi.org/10.6084/m9.figshare.c.5022878> (2020).

31. Howard, J. P. Phonetic algorithms in R. *Journal of Open Source Software*, **3**, 480 (2018).
32. van der Loo, M. P. J. The stringdist package for approximate string matching. *R J.* **6**, 111–122 (2014).
33. Cohen, W. W., Ravikumar, P. & Fienberg, S. E. A comparison of string distance metrics for name-matching tasks. *Proc. IJCAI-2003 Work. Inf. Integr. Web* (2003).

Acknowledgements

This work was supported by a 2018 National University of Singapore Early Career Award (Grant number NUS_ECRA_FY18_P15) to A.Hsu and a Yale-NUS College Summer Research Programme Award to M. Raghavan and Y. Kim. We'd also like to thank Claire Kruppenacher, Xiyao Fu, Yin Xi Tan of Yale-NUS College for data collection and validation support; Nihit Goyal of Yale-NUS College for road-testing the ClimActor R package; and Luke Hellum of Yale College for contributing to an earlier version of this work. We also thank the transnational climate initiatives that are included in this dataset.

Author contributions

Each author's contribution to the work should be described briefly, on a separate line, in the Author Contributions section. A.H. conceived of the dataset, compiled data, contributed code, and wrote the article. Z.Y. led the development of the ClimActor package, collected data, produced figures, conducted data validation, assisted with writing of the article. R.R. developed the Wikipedia python scraper and contributed to writing the article. J.S. assisted in developing an earlier version of the ClimActor package. Y.K. assisted in data collection, validation, and contributed to writing the article. S.R. assisted in developing the ClimActor package and the article's figures. N.C. assisted in data collection and validation. V.N. assisted in the Wikipedia python scraper and data validation. A.W. assisted with data collection and contributed to writing the article.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020