

Analysing Demographic Representation in Wikidata: A Visual Analytics Approach

Aysha Tolefat¹, Michelle Nwachukwu¹ and Martim Brandão¹

¹King's College London

Abstract

Wikidata is one of the largest open knowledge graphs, which contains data on real people's demographic data at scale. This data is also known to be unevenly covered across gender and geography, a problem with important downstream effects. Identifying and analysing such gaps is currently challenging in practice, however, especially for non-expert users and contributors. This paper presents a visual analytics system for exploring and investigating demographic representation gaps, bias and incompleteness in Wikidata biographies. The system gathers country-level demographic data across five demographic dimensions: gender, age, language, ethnic group, and occupation, with gender additionally analysed over time. These are visualised in an interactive dashboard that combines a global country overview with per-country views. The dashboard enables non-expert users to examine demographic patterns, identify representation gaps within Wikidata, and plan interventions towards more equitable representation of content. We used a controlled between-subjects study with forty (n=40) participants to evaluate the dashboard in representation-gap investigation tasks, showing fast task completion times, higher task accuracy and confidence than a tabular data baseline, and a System Usability Scale score of 84.25. Thus we demonstrate that incomplete knowledge graph records can be transformed into interpretable representation profiles that enable non-expert users to identify representation gaps and plan interventions towards more equitable content. Finally, we report on the current state of demographic representation in Wikidata, as of 2026, namely in terms of age distribution and the countries, age ranges and occupation sectors with highest (over 80%) gender gap.

Keywords

Wikidata, knowledge graphs, representation gaps, demographic bias, visual analytics, biographical data

1. Introduction

Wikidata is one of the largest open knowledge graphs, providing collaboratively curated structured data about millions of real-world entities [1]. It contains extensive biographical records described through properties such as gender (P21), date of birth (P569), occupation (P106), ethnic group (P172), and native and spoken languages (P103, P1412). As a socio-technical system shaped by the priorities and participation patterns of its editorial community, Wikidata cannot be treated as a neutral representation of the world [2]. Prior work has shown that biographical coverage is systematically uneven across gender and geography [3, 4], that data quality varies substantially by language community and region [5], and that such imbalances propagate into the downstream systems built on Wikidata as a structured source [2, 4]. These biases are not merely descriptive. They have direct consequences for any knowledge-based application that depends on Wikidata's biographical data.

Despite this, the demographic structure of Wikidata's biographical records remains difficult to examine in practice. SPARQL queries against the Wikidata Query Service are subject to strict timeout and rate-limiting constraints [6], and raw query outputs return identifier-heavy, incomplete records that are unsuitable for direct comparison or pattern recognition. Existing interfaces either target expert or SPARQL-literate editorial users [7], operate on Wikipedia's article-level content rather than structured knowledge graph properties [8, 9], or focus on

Wikidata'26: Wikidata Workshop at ISWC 2026

✉ aysha.tolefat@kcl.ac.uk (A. Tolefat); michelle.nwachukwu@kcl.ac.uk (M. Nwachukwu); martim.brandao@kcl.ac.uk (M. Brandão)

ORCID 0009-0002-6771-6014 (A. Tolefat); 0009-0007-7546-4888 (M. Nwachukwu); 0000-0002-2003-0675 (M. Brandão)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

domain-specific tasks such as scientometrics [10]. More importantly, they do not support direct analysis of demographic representation within Wikidata’s biographical records. No existing system supports country-level, multi-dimensional inspection of Wikidata’s demographic data in a form accessible to non-expert users. Such an interface can support social scientists, Wikidata editors, and non-specialist users in investigating incompleteness, representation gaps and bias in Wikidata’s biographical content.

This paper makes four contributions. First, we present a preprocessing pipeline that extracts Wikidata biographies into structured country-level datasets across five demographic dimensions, with gender additionally analysed over time. Second, we present an interactive visual analytics dashboard that enables country-level demographic exploration through linked visual views, requiring no SPARQL expertise. Third, we report a between-subjects user study ($n = 40$) evaluating task completion time, accuracy, confidence, and usability. Fourth, we analyse the processed datasets to identify representation gaps, namely in terms of age distribution and the countries, age ranges and occupation sectors with highest gender gap.

2. Related Work

2.1. Demographic Bias in Wikidata Biographical Data

Bias in collaborative knowledge systems arises not only from deliberate exclusion but from structural factors that determine whose knowledge is recorded, how completely, and under what conditions [11]. In the context of Wikidata, it is important to distinguish between representation gap and representation bias. A gap or imbalance refers to an uneven distribution of entities or demographic properties within the knowledge graph, whereas bias implies that this distribution systematically departs from an appropriate reference point, such as real-world populations or notability. Consequently, unequal demographic distributions do not necessarily indicate that Wikidata inaccurately reflects reality. Rather, they highlight patterns of representation that require careful interpretation alongside existing knowledge of editorial practices and data completeness.

Prior work has shown that representation gaps are common across several demographic dimensions. Klein et al. [13] introduced the Wikidata Human Gender Indicators dataset for tracking gender gaps across time, geography, occupation, and language, showing that gender imbalance in Wikidata is structured and measurable. However, the analysis depends on the completeness of Wikidata’s gender property (P21), which is itself unevenly populated. Das et al. [4] demonstrate that biographical coverage in Wikidata is concentrated in the Global North, while Zhang and Terveen [14] show that estimates of gender disparity depend on category construction and data completeness. These findings motivate the normalisation and missing-value handling adopted in this work. Amaral et al. [5] show that data quality varies across language communities, meaning biographies from under-resourced communities are more likely to contain incomplete demographic information. Analytical systems should therefore account for, rather than suppress, missing values.

Broadly defined as the extent to which all required information is present in a dataset [38], data completeness is a well-established area of study [35, 36, 37] and remains important for the continuous maintenance and development of knowledge graphs. Methods for measuring these gaps span from external validation against gold standards [39] to relative, peer-based metrics [40]. Prior work shows that Wikidata completeness varies across classes, attributes, and users’ analytical needs [15, 16].

2.2. Visual Analytics Interfaces for Wikimedia and Knowledge Graphs

Various interfaces have been proposed for the visualisation of Knowledge Graphs. De Sabbata et al. [9] and the Wikipedia Diversity Observatory [8] use visual interfaces to expose representation

gaps in Wikipedia. However, both operate on article-level metadata rather than Wikidata’s structured properties and do not support multi-dimensional demographic analysis across countries. Vispedia [33] enables non-expert users to explore Wikipedia data visually through search-based integration. However, it operates on DBpedia rather than Wikidata and provides no support for analysing demographic representation across country-level biography collections. WikipediaViz [34] helps readers assess Wikipedia article quality using lightweight visual indicators, but focuses on edit-history metadata rather than demographic representation in Wikidata biographies. Within the Wikidata ecosystem, Graux et al. [7] demonstrate the feasibility of dashboard-based interfaces for monitoring edit activity, while Elhossary and Förstner [17] show how Wikidata can support exploration of scientific literature. Neither system addresses country-level demographic representation across biographies.

Our dashboard design follows established visual analytics principles that emphasise linked interactive views, task-appropriate visual encodings, and well-structured datasets to support analytical reasoning [18, 19, 20, 21].

3. Methodology

3.1. System Overview and Data Pipeline

Our system adopts a three-stage architecture consisting of data acquisition from Wikidata, offline preprocessing into structured datasets, and interactive presentation through a web-based dashboard. The system extracts biographical data for 195 sovereign countries and territories using a country-by-country acquisition strategy designed around the practical constraints of the Wikidata Query Service (WDQS). Rather than issuing a single global SPARQL query, separate queries are executed for each country to reduce timeouts. Human entities are identified through the Wikidata class Human (Q5) and linked to countries via the place of birth property (P19), resolving places to their corresponding countries through the country property (P17) where necessary. Where date of birth information is available, biographies are restricted to records from 1900 onwards to improve consistency across demographic analyses while reducing the influence of sparse historical records. Biographies without recorded birth dates are retained to avoid unnecessary exclusion of otherwise valid records.

The system also extracts four demographic dimensions—gender (P21), date of birth (P569), ethnic group (P172), and native and spoken languages (P103 and P1412)—as well as Occupation (P106).

Each demographic dimension is then normalised. Gender identifiers are mapped into four categories: Male, Female, Non-binary or other, and Unknown / not stated. Unlike conventional preprocessing pipelines that discard incomplete records, unknown values are retained explicitly because they represent an informative manifestation of incompleteness within Wikidata. Gender distributions are additionally aggregated by decade of birth rather than individual year to reduce sparsity and improve the interpretability of temporal trends, particularly for countries with relatively few biographies. Age was calculated from date of birth (P569) relative to the year of extraction. For this analysis, records with a recorded date of death (P570) were excluded. Remaining records were grouped according to United Nations demographic reporting conventions into four age cohorts: 0–14, 15–24, 25–59, and 60+ years. Native language and spoken language are preserved as separate datasets because they represent distinct aspects of biographical information and combining them would obscure meaningful differences in language representation. Ethnic-group values are retained using resolved labels, while absent properties are treated as missing rather than inferred from an assumed national or cultural default. Regarding occupation, Wikidata contains thousands of highly specific labels with substantial variation in granularity. Occupation labels were normalised and mapped to the International Standard Classification of Occupations 2008 (ISCO-08) major and, where possible, sub-major groups using a conservative rule-based procedure. Explicit label-pattern rules were applied to the normalised

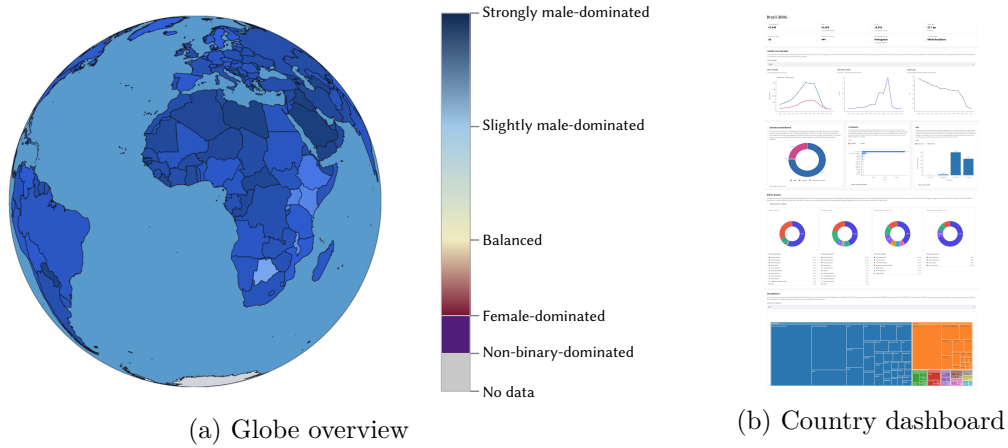


Figure 1: System interface. (a) Globe overview showing per-country gender balance as a diverging colour scheme. (b) Country dashboard for Brazil across all five demographic dimensions. Individual panels shown at readable scale in Figures 2-4.

occupation names to assign the most appropriate ISCO category. Each mapping retained its method and explanatory notes, while ambiguous labels and values not representing occupations were left unmapped. For example, labels such as engineer were mapped to Professionals, while farmer was mapped to Skilled Agricultural, Forestry and Fishery Workers.

The preprocessing pipeline ultimately produces six analytical datasets corresponding to the dashboard views: overall gender distribution, temporal gender trends, age groups, language representation, ethnic groups, and occupations. Each dataset stores normalised demographic categories together with aggregated country-level counts, enabling all dashboard interactions to operate entirely on precomputed data rather than issuing live queries during user exploration. Beyond improving responsiveness, this design ensures that all users analyse an identical snapshot of Wikidata, allowing demographic comparisons to be performed consistently without being affected by subsequent changes to the live knowledge graph.

3.2. Dashboard Design

We implemented the dashboard in Streamlit, concretely as two interface layers: a globe-based overview for country selection and global representation-gap visualization, and a dedicated country dashboard for in-depth demographic exploration (Figure 1). The deployed dashboard is publicly available at <https://wiki-dash.streamlit.app/>, with the source code and data-processing scripts available at <https://github.com/AyshaTolefat/Wiki-dash>. The globe encodes gender balance using a diverging colour scheme, with countries shaded toward one pole for male-dominant representation, the opposite pole for female-dominant representation, and a neutral midpoint for approximate balance. Countries for which no data is available are shown in grey. This encoding follows established guidance on diverging colour maps for data with a meaningful central reference [22], and allows users to identify gender gap patterns globally before selecting a country for deeper analysis.

The country dashboard follows a summary-first structure grounded in established dashboard design principles [23, 24]. A headline indicator panel presents total biographies, gender counts, the gender gap in percentage points, the top spoken language, and the most frequent ethnic group label before the user engages with the detailed visual sections. The dashboard then presents the five demographic dimensions through linked views, all operating within a shared country context derived from the preprocessed datasets (Figures 2-4).

Section-level interactivity supports details-on-demand without requiring users to navigate away from the country context [24]. Controls include a decade selector in the temporal gender section, a native/spoken toggle in the language section, count/percentage switching in the age



Figure 2: Dashboard panels for Brazil. Top: temporal gender trends showing male vs. female biography counts, non-binary counts, and absolute gender gap per decade of birth. Bottom left to right: overall gender breakdown; language representation for spoken languages; age group distribution across four bands.

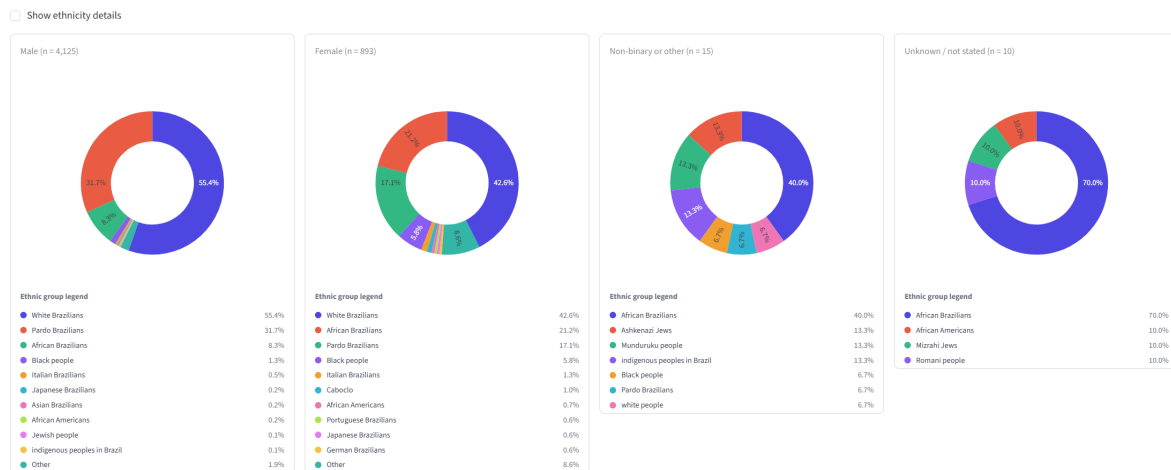


Figure 3: Ethnic group composition for Brazil by gender category. Each donut chart shows the distribution of ethnic group labels within one gender group.

section, and a gender filter in the occupation section. All interactions operate on the preloaded country slice, ensuring that the analytical context remains consistent across the full session.

3.3. User Study Design

We designed a user study in order to evaluate the extent to which the dashboard leads to efficient, accurate and confident use, compared to raw tabular inspection as a reference. Tabular data is the type of data Wikidata editors and non-expert users would currently (i.e. with existing means) be expected to analyse in order to solve the tasks they were given. We note that this is a “weak” baseline corresponding to off-the-shelf practice which was not designed specifically for the tasks, and therefore its purpose is to provide a point of reference to judge dashboard results, rather than to prove optimal design of our system.

A between-subjects design was chosen to eliminate learning effects that would arise if participants completed the same analytical tasks under both conditions. Forty participants were recruited through purposive sampling, targeting undergraduate students across STEM and humanities disciplines without Wikidata expertise, with snowball sampling used to reach the target number, and randomly assigned to one of two independent conditions: twenty interacted

accuracy, and confidence were all non-normally distributed in at least one condition, and the Wilcoxon Rank-Sum test was therefore applied as the statistical significance test. Statistical significance of task-level accuracy was analysed using Chi-Square tests, with Fisher’s Exact test substituted where expected cell counts fell below five.

4. Results

4.1. Qualitative Findings

Open-ended questionnaire responses showed satisfaction with dashboard, but also some frustrations. Dashboard participants frequently described the interface as “clear,” “intuitive,” and “easy to use,” with several noting that visual summaries allowed them to identify patterns and trends without manually scanning data. As one participant put it, “everything was very smooth and easy to use.” The most frequently raised limitation concerned the occupation treemap. One participant noted that it was “overwhelming since it displays lots of information in many different boxes,” and while it effectively communicates hierarchical structure, it imposed a greater interpretive burden than other sections and is the clearest target for future refinement. The final recurring concern involved the flow between the country search control and the globe. One participant described having to manually click the country on the map after already searching for it as “unnecessarily frustrating,” pointing to a specific workflow inefficiency resolvable through search-to-selection automation.

4.2. Quantitative Results

Figure 5a shows task accuracy by condition across all six tasks. The dashboard condition achieved an overall accuracy of 90.8%, compared to 62.5% in the tabular condition (statistically significant difference with $\chi^2 = 26.93$, $p < .001$). Significant differences were observed for tasks involving comparative reasoning, while other tasks were not significant (Figure 5a).

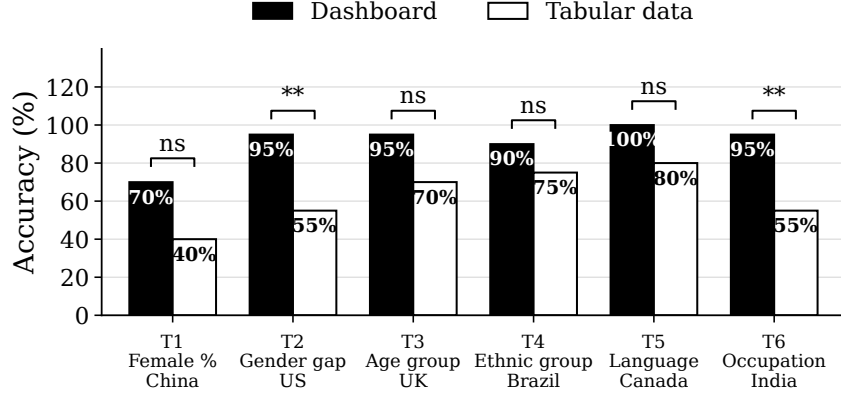
Figure 5b shows mean task completion time per condition across all six tasks. Dashboard participants completed tasks efficiently at an average of 48.94s, compared to 306.17s by participants using the tabular interface ($U = 0.00$, $p < .001$).

Dashboard participants reported high confidence in their answers across all tasks ($M = 4.63$, $SD = 0.46$), higher than tabular participants ($M = 3.12$, $SD = 0.88$), $U = 379.50$, $p < .001$. This higher confidence aligns with the faster completion times and higher accuracy observed across all tasks.

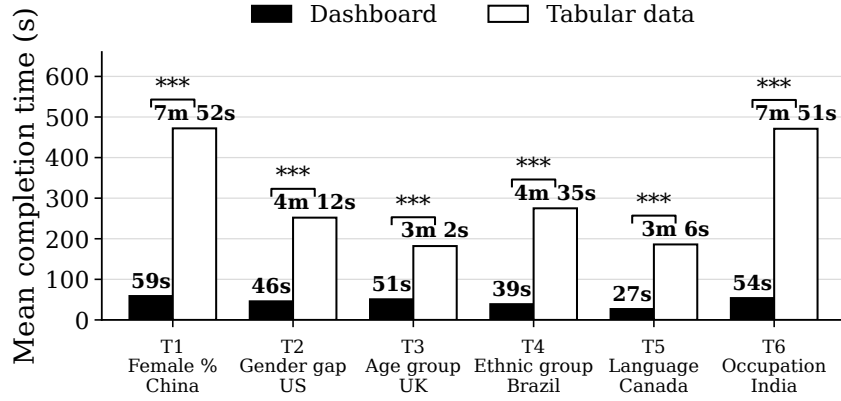
Finally, the dashboard condition achieved a mean System Usability Scale score of 84.25, corresponding to an “excellent” usability range [32], compared to 57.38 in the tabular condition (“acceptable” range).

4.3. Representation Gaps in Wikidata

Beyond the user evaluation, we analysed the processed Wikidata datasets to identify global representation gaps (Figure 6). Several countries exhibit substantial gender imbalance relative to the expected 50% parity, with Saudi Arabia (89.6% male biographies), Iraq (88.9%), and Paraguay (88.1%) showing the highest male representation among countries with at least 1,000 biographies recorded as male or female. The age distribution is concentrated in the 25–59 group (62.7% on average), differing from the global population distribution. This pattern is partly consistent with Wikidata’s focus on notable persons, since notability and available biographical documentation tend to accumulate with age, making children less likely to appear in the knowledge graph. Female representation also varies across occupational sectors, ranging from 6.3% in Military & Security to 32.2% in Arts & Entertainment, remaining below both women’s 40.1% share of global employment in 2025 [41] and gender parity. Finally, female



(a) Task accuracy.



(b) Task completion time.

Figure 5: Comparison of dashboard and tabular baseline performance across the six evaluation tasks. (a) Task accuracy. Brackets indicate pairwise significance from Chi-Square (χ^2) or Fisher’s Exact (FE) tests (overall: $\chi^2 = 26.93$, $p < .001$). (b) Mean task completion time. All pairwise differences were significant ($p < .001$, Wilcoxon Rank-Sum, $U = 0$). ** $p < .01$; *** $p < .001$; ns = not significant.

representation has generally increased across birth cohorts, rising from 14.7% among biographies born in the 1900s to approximately 31% among those born in the 2000s.

Together, these findings highlight systematic representation differences across countries, occupations, age groups, and time—highlighting the usefulness of our tool and assisting Wikidata editors in planning future interventions.

4.4. Discussion

The results provide consistent evidence that our visualization tool facilitates the identification of representation gaps on Wikidata. However, several limitations should be acknowledged.

First, the demographic patterns presented by the dashboard reflect the current state of Wikidata rather than real-world population demographics, and imbalances surfaced by our system are not necessarily indicative of representation bias—as such an assessment will require a comparison to the real-world distribution.

Second, although missing and unknown demographic values were deliberately preserved as analytical signals, incompleteness is not uniformly distributed across countries or demographic dimensions. Some regions are substantially better documented than others, and certain properties, such as ethnic group or occupation, are less consistently populated than gender or date of birth. Ethnic-group and native-language properties may be selectively recorded when a value differs from an assumed national or cultural norm; consequently, an unrecorded value was treated as missing rather than inferred as the presumed default, and the observed distributions should be

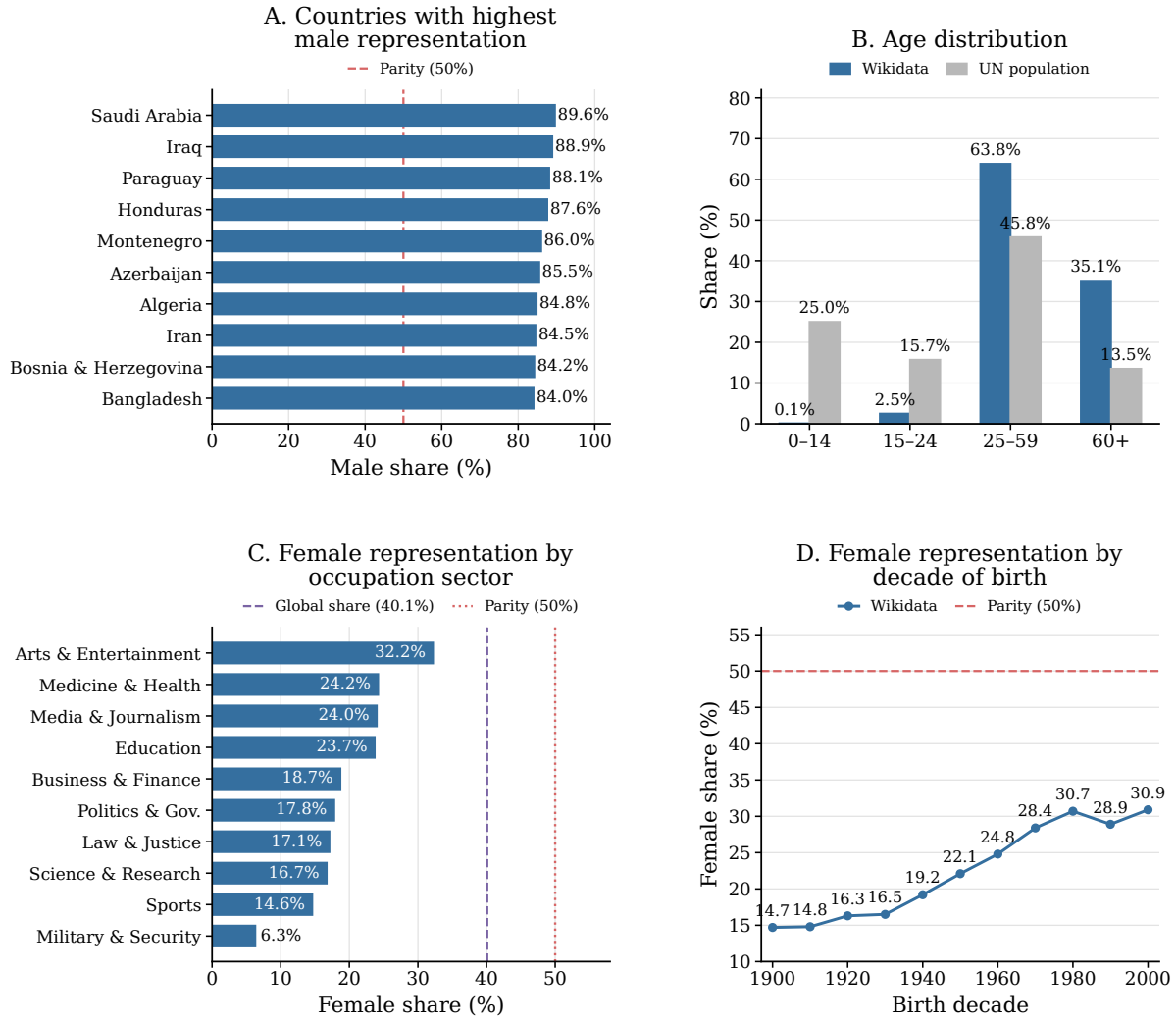


Figure 6: Representation gaps identified from the processed Wikidata biography datasets. (A) Countries with the highest male representation among countries containing at least 1,000 male-labelled or female-labelled biographies. (B) Mean country-level age-group distribution. (C) Female representation across the ten largest occupation sectors. (D) Female representation by decade of birth.

interpreted as reflecting both Wikidata representation and editorial recording practices rather than real-world population composition.

Third, the participant pool consisted primarily of students aged 18–27, supporting claims about non-expert accessibility but limiting generalisation to expert users or more diverse populations. Task completion times were self-reported, although the magnitude of the observed differences makes this unlikely to affect the overall conclusions. Finally, the evaluation compared the dashboard against a tabular baseline only, meaning the findings demonstrate its effectiveness relative to a “weak” baseline of raw tabular inspection rather than alternative visual analytics systems. Such comparison should be the subject of future work. The preprocessing-based architecture also requires periodic regeneration of the datasets as Wikidata evolves, meaning newly added or edited records are not reflected until the next preprocessing cycle.

5. Conclusion and Future Work

This paper presented a system for making demographic representation in Wikidata accessible to non-expert users through interactive visual analytics. We developed a reproducible preprocessing pipeline that extracts and structures Wikidata biographical data across five demographic

dimensions into coherent country-level datasets. Built on these datasets, an interactive dashboard enables country-level demographic exploration and representation-gap identification through linked visual views without requiring SPARQL expertise. A between-subjects user study ($n = 40$) demonstrated that the dashboard has “excellent” usability and leads to efficient and accurate investigation of representation gaps compared to a tabular data baseline: accuracy was high at 90.8% (vs 62.5% in tabular), confidence was 4.63/5 (vs 3.12). These results confirm that our system can make the demographic structure of a large collaborative knowledge graph substantially legible and interpretable for non-expert users. Finally, we analysed the demographic data to identify global representation gaps in the current state of Wikidata, namely in terms of age distribution and the countries, age ranges and occupation sectors with highest (80%+) gender gap.

Future work should support direct multi-country comparison, refine the occupation treemap, evaluate larger and more diverse samples, incorporate automated interaction logging, and assess completeness against classes with known real-world populations, such as national sports teams or award recipients.

Acknowledgments

This work was partially supported by the UK Research and Innovation Centre for Doctoral Training in Safe and Trusted Artificial Intelligence [grant number EP/S023356/1].

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Improve writing style, Paraphrase and reword. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

- [1] Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. *Communications of the ACM* 57(10), 78–85 (2014)
- [2] Piscopo, A., Simperl, E.: What we talk about when we talk about Wikidata quality: a literature survey. In: *Proceedings of the 15th International Symposium on Open Collaboration (OpenSym '19)*. ACM, New York (2019)
- [3] Wagner, C., Graells-Garrido, E., Garcia, D., Menczer, F.: Women through the glass ceiling: gender asymmetries in Wikipedia. *EPJ Data Science* 5, Article 5 (2016)
- [4] Das, P., Karnam, S.K., Soni, A.B., Mukherjee, A.: Social biases in knowledge representations of Wikidata separates Global North from Global South. In: *Proceedings of the 17th ACM Web Science Conference 2025*, pp. 12–21. ACM, New York (2025)
- [5] Amaral, G., Piscopo, A., Kaffee, L.A., Rodrigues, O., Simperl, E.: Assessing the quality of sources in Wikidata across languages: a hybrid approach. *Journal of Data and Information Quality* 13(4), Article 23 (2021)
- [6] Wikimedia Foundation: Wikidata Query Service — User Manual (2024), https://www.mediawiki.org/wiki/Wikidata_Query_Service/User_Manual
- [7] Graux, D., Orlandi, F., Lynch, B., Mahon, I., Mullen, O., Mahon, A., et al.: A real-time visual dashboard for Wikidata edits. In: *Proceedings of VOILA! 2020. CEUR Workshop Proceedings*, vol. 2778, pp. 41–46 (2020)
- [8] Miquel-Ribé, M., Laniado, D.: The Wikipedia Diversity Observatory: helping communities to bridge content gaps through interactive interfaces. *Journal of Internet Services and Applications* 12, Article 10 (2021)

- [9] De Sabbata, S., Çöltekin, A., Eccles, K., Hale, S., Straumann, R.: Collaborative visualizations for Wikipedia critique and activism. *Proceedings of the International AAAI Conference on Web and Social Media* 9(5), 11–16 (2015)
- [10] Nielsen, F.Å., Mietchen, D., Willighagen, E.: Scholia, scientometrics and Wikidata. In: *The Semantic Web: ESWC 2017 Satellite Events. Lecture Notes in Computer Science*, vol. 10577, pp. 237–259. Springer, Cham (2017)
- [11] Redi, M., Gerlach, M., Johnson, I., Morgan, J., Zia, L.: A taxonomy of knowledge gaps for Wikimedia projects (second draft). *arXiv:2008.12314* (2020)
- [12] Graells-Garrido, E., Lalmas, M., Menczer, F.: First women, second sex: gender bias in Wikipedia. In: *Proceedings of the 26th ACM Conference on Hypertext & Social Media (HT '15)*, pp. 165–174. ACM, New York (2015)
- [13] Klein, M., Gupta, H., Meehan, C., Moss, H., Burns, K.: Monitoring the gender gap with Wikidata Human Gender Indicators. In: *Proceedings of the 12th International Symposium on Open Collaboration (OpenSym '16)*, Article 43. ACM, New York (2016)
- [14] Zhang, C.C., Terveen, L.: Quantifying the gap: a case study of Wikidata gender disparities. In: *Proceedings of the 17th International Symposium on Open Collaboration (OpenSym 2021)* (2021)
- [15] Wisesa, A., Darari, F., Krisnadhi, A.A., Nutt, W., Razniewski, S.: Wikidata completeness profiling using ProWD. In: *Proceedings of the 10th International Conference on Knowledge Capture (K-CAP '19)*, pp. 123–130. ACM, New York (2019). doi:10.1145/3360901.3364425
- [16] Abián, D., Meroño-Peñuela, A., Simperl, E.: An analysis of content gaps versus user needs in the Wikidata knowledge graph. In: *The Semantic Web – ISWC 2022*, pp. 354–374. Springer, Cham (2022)
- [17] Elhossary, M., Förstner, K.U.: InteractOA: showcasing the representation of knowledge from scientific literature in Wikidata. *Semantic Web* 15, 2381–2393 (2024)
- [18] Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: *Information Visualization. Lecture Notes in Computer Science*, vol. 4950, pp. 154–175. Springer, Berlin (2008)
- [19] Franconeri, S.L., Padilla, L.M., Shah, P., Zacks, J.M., Hullman, J.: The science of visual data communication: what works. *Psychological Science in the Public Interest* 22(3), 110–161 (2021)
- [20] Munzner, T.: *Visualization Analysis and Design*. CRC Press, Boca Raton, FL (2014)
- [21] Wickham, H.: Tidy data. *Journal of Statistical Software* 59(10), 1–23 (2014)
- [22] Moreland, K.: Diverging color maps for scientific visualization. In: *Advances in Visual Computing: 5th International Symposium, ISVC 2009, Proceedings, Part II. Lecture Notes in Computer Science*, vol. 5876, pp. 92–103. Springer, Berlin, Heidelberg (2009). doi:10.1007/978-3-642-10520-3_9
- [23] Few, S.: *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, Sebastopol, CA (2006)
- [24] Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings of the 1996 IEEE Symposium on Visual Languages*, pp. 336–343 (1996)
- [25] Shneiderman, B.: Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics* 11(1), 92–99 (1992)
- [26] Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. *Semantic Web* 9(1), 77–129 (2017)
- [27] Aigner, W., Miksch, S., Schumann, H., Tominski, C.: *Visualization of Time-Oriented Data*. Springer, London (2011)
- [28] United Nations Statistics Division: *National Reporting of Age and Sex-Specific Data*. United Nations, New York (2004)
- [29] International Labour Organization: *International Standard Classification of Occupations: ISCO-08*. ILO, Geneva (2012)

- [30] Brooke, J.: SUS: a quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (eds.) *Usability Evaluation in Industry*, pp. 189–194. Taylor & Francis, London (1996)
- [31] Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13(3), 319–340 (1989)
- [32] Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies* 4(3), 114–123 (2009)
- [33] Chan, B., Wu, L., Talbot, J., Cammarano, M., Hanrahan, P.: Vispedia: Interactive visual exploration of Wikipedia data via search-based integration. *IEEE Transactions on Visualization and Computer Graphics* 14(6), 1213–1220 (2008)
- [34] Chevalier, F., Huot, S., Fekete, J.D.: WikipediaViz: Conveying article quality for casual Wikipedia readers. In: *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, pp. 49–56 (2010)
- [35] Prasojo, R.E., Darari, F., Razniewski, S., Nutt, W.: Managing and consuming completeness information for wikidata using COOL-WD. In: *Proceedings of the 7th International Workshop on Consuming Linked Data, co-located with 15th International Semantic Web Conference (ISWC 2015)* (2016)
- [36] Ahmeti, A., Razniewski, S., Polleres, A.: Assessing the completeness of entities in knowledge bases. In: *European Semantic Web Conference*, pp. 7–11 (2017)
- [37] Darari, F., Razniewski, S., Prasojo, R.E., Nutt, W.: Enabling fine-grained RDF data completeness assessment. In: *International Conference on Web Engineering*, pp. 170–187 (2016)
- [38] Lehmann, J., Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Auer, S.: Quality assessment for Linked Data: A Survey. *Semantic Web* (1570-0844) 7(1) (2016)
- [39] Färber, M., Bartscherer, F., Menne, C., Rettinger, A.: Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic Web* 9(1), pp. 77–129 (2017)
- [40] Balaraman, V., Razniewski, S., Nutt, W.: Recoin: relative completeness in Wikidata. In: *Companion Proceedings of the The Web Conference 2018*, pp. 1787–1792 (2018)
- [41] International Labour Organization: The World of Work and the 2030 Agenda: A Ten-Year Review. *ILOSTAT* (2026), <https://ilostat.ilo.org/blog/the-world-of-work-and-the-2030-agenda-a-ten-year-review/>