



City Research Online

## City, University of London Institutional Repository

---

**Citation:** Hutchinson, M., Slingsby, A., Jianu, R. & Madhyastha, P. (2023). Towards Visualisation Specifications from Multilingual Natural Language Queries using Large Language Models. In: EuroVis 2023 - Posters. (pp. 77-79). Eindhoven, The Netherlands: Eurographics. ISBN 9783038682202 doi: 10.2312/evp.20231072

This is the published version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/31527/>

**Link to published version:** <https://doi.org/10.2312/evp.20231072>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

---



# Towards Visualisation Specifications from Multilingual Natural Language Queries using Large Language Models

Maeve Hutchinson\*, Aidan Slingsby, Radu Jianu, Pranava Madhyastha\*

Department of Computer Science, City, University of London

\* Corresponding Authors

Recent research into pre-trained Large Language Models (LLMs) suggests that generating data visualisations from natural language is an increasingly feasible technique, but current systems predominantly focus on English language. We demonstrate that a LLM approach can generate visualisations from multilingual queries, a first step towards the development of accessible multilingual interfaces for data visualisation.

## Experimental Setup

We use the **dataset** by Srinivasan, et al.<sup>1</sup> of 814 natural language queries, each relating to a specific Vega-Lite visualisation.

We use a **prompt-based learning** approach. This involves engineering a natural language prompt input that contains a task description and a set of valid in-context examples. We pair each query with another example from the dataset and its completed specification.

Examples are categorised as:

- (a) *in-domain*: semantically similar to the user query, relating to the same dataset
- (b) *out-of-domain*: not associated with the same dataset as the user query

For the **multilingual** step, we machine translate each query into Spanish, and pair with an in-domain example in English.

Generate a vega-lite specification for the task.

Task: Using data from <https://raw.githubusercontent.com/nlvcorp/nlvcorp/main/datasets/cars.csv>

Does higher cylinder count mean lower mileage?

Specification: 

```
{ "$schema": "https://vega.github.io/schema/vega-lite/v3.json", "data": ... }
```

Completed in-domain example

How does mileage relate to cylinder count?

Task: Using data from <https://raw.githubusercontent.com/nlvcorp/nlvcorp/main/datasets/cars.csv>

Specification: 

```
{ "$schema": "https://vega.github.io/schema/vega-lite/v3.json",
```

User query and incomplete solution

Show me a scatterplot of profitability by sales for each region

Muéstrame un diagrama de rentabilidad por ventas para cada

Plot a histogram of movie running time

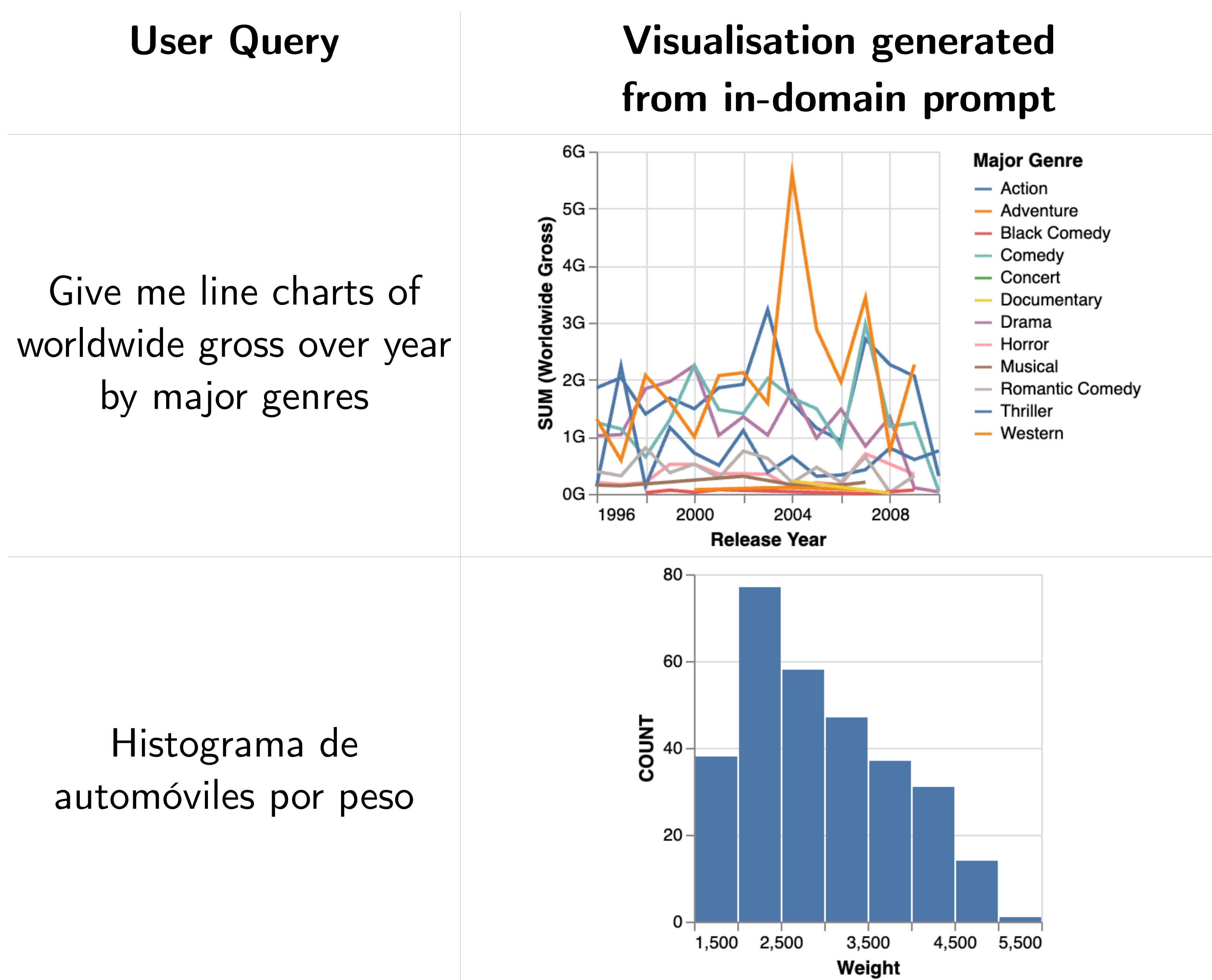
Trace un histograma del tiempo de ejecución de la película.

We implement a pre-trained **Large Language Model** that is able to generate text continuing from the input prompt. We use the openly available **BLOOM** model, which was built over data in multiple languages and large GitHub-based code repositories.

## Initial Results

We use a series of unit tests to **evaluate** the generated specifications: specifications that are not valid python dictionaries are categorised as *syntax errors*; dictionaries that do not compile to vega-lite are *validation errors*; and if the compiled chart does not match the target chart it is a *logic error*. The *Pass@1* metric represents the proportion of solutions that successfully match the target visualisation from the corpus.

| Prompt Type           | Pass@1 | Syntax Error | Validation Error | Logic Error |
|-----------------------|--------|--------------|------------------|-------------|
| English Out-of-domain | 0.12%  | 4.4%         | 2.7%             | 92.8%       |
| English In-domain     | 81.1%  | 3.4%         | 0%               | 15.5%       |
| Spanish In-domain     | 78.1%  | 4.3%         | 0.12%            | 17.4%       |



These results provide a proof of concept that a pre-trained LLM can produce a Vega-Lite specification from a user query in multiple languages using prompt engineering. Notably, the results for Spanish queries are very similar to English despite having no in-context examples in Spanish.

<sup>1</sup> SRINIVASAN, ARJUN, NYAPATHY, NIKHILA, LEE, BONGSHIN, et al. "Collecting and Characterizing Natural Language Utterances for Specifying Data Visualizations". Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21. New York, NY, USA: Association for Computing Machinery, May 2021, 1–10. ISBN: 978-1-4503-8096-6. DOI: 10.1145/3411764.3445400