

# Data2Vis: Automatic Generation of Data Visualizations Using Sequence to Sequence Recurrent Neural Networks

Victor Dibia, Cagatay Demiralp

**Abstract—** Rapidly creating effective visualizations using expressive grammars is challenging for users who have limited time and limited skills in statistics and data visualization. Even high-level, dedicated visualization tools often require users to manually select among data attributes, decide which transformations to apply, and specify mappings between visual encoding variables and raw or transformed attributes. In this paper we introduce Data2Vis, an end-to-end trainable neural translation model for automatically generating visualizations from given datasets. We formulate visualization generation as a language translation problem where data specifications are mapped to visualization specifications in a declarative language (Vega-Lite). To this end, we train a multilayered attention-based encoder-decoder network with long short-term memory (LSTM) units on a corpus of visualization specifications. Qualitative results show that our model learns the vocabulary and syntax for a valid visualization, appropriate transformations (count, bins, mean) and how to use common data selection patterns. We introduce two metrics for evaluating the task of automated visualization generation (language syntax validity, visualization grammar syntax validity) and demonstrate the efficacy of bidirectional models with attention mechanisms for this task. Data2Vis generates visualizations that are comparable to manually-created visualizations in a fraction of the time, with potential to learn more complex visualization strategies at scale.

**For the published version of record document, go to:**

**<http://dx.doi.org/10.1109/MCG.2019.2924636>**