Introduction to the "Exploring Histograms" Online Essay

Aran Lunzer and Amelia McNamara

VPRI Research Note RN-2017-003

Introduction to the 'Exploring Histograms' online essay,
`https://tinlizzie.org/histograms/`
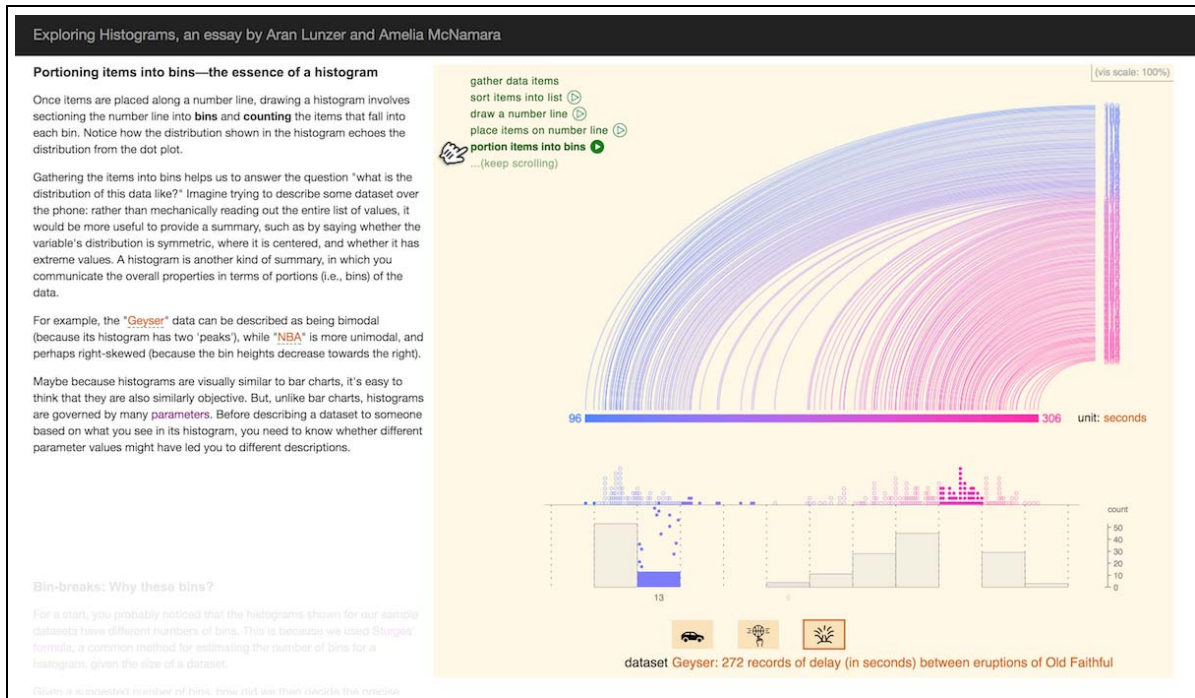
Aran Lunzer and Amelia McNamara

July 2017

Fig. 1: A snapshot of the *Exploring Histograms* essay. In this stage, the item values in a sample dataset have been gathered along a number line and are in the process of being portioned into equal-width bins. The curved paths, coloured smoothly from blue to pink along the values' overall range, connect each value's position on the number line to its place within the sorted (and densely packed) list at right. The clear appearance of a mostly-blue path cluster and a mostly-pink cluster helps to highlight the bimodal distribution in this dataset.

---

1 HOW WE CAME TO WRITE THIS ESSAY

Over the years, our work on histograms has taken a variety of forms. After initial explorations in R, we developed the LivelyR demonstrator to bring some of the interactivity of the Lively Web programming environment [2] to R charting. LivelyR showcases several forms of live parameter adjustment, including those involved in generating histograms: a user can adjust bin width and bin offset, and can set up a 'sweep' of parameter values to see a 'histogram cloud' of overlaid possibilities. We presented LivelyR in several venues [3, 4], and received especially positive feedback on the histogram features. Many people were struck by the way a single dataset can appear as a variety of histogram shapes, just depending on the parameter settings.

Our online essay was motivated by a desire to expose a broader audience to these insights. Inspired by "active essays" [7] and "explorable explanations" [1, 6], we set out to make an essay that would explain how to make a histogram from the ground up, and allow the reader to play along the way.

- *Aran Lunzer is a Senior Researcher at Viewpoints Research Institute, Los Angeles, USA. E-mail: aran@acm.org.*
- *Amelia McNamara is Visiting Assistant Professor of Statistical and Data Sciences at Smith College, Northampton, MA, USA. Twitter: @AmeliaMN.*

2 WHAT THE ESSAY SHOWS

The essay walks a reader through the process of building a histogram, from an unordered clump of values to the finished chart. Figure 1 shows the 'portioning items into bins' stage of the essay. After values have been sorted and dropped onto the number line (forming a dot plot), they then fall through to bins and are counted. At this stage, the bins have been chosen by us, but subsequent stages of the essay allow the reader to adjust the bin offset, bin width, and bin openness. Near the end of the essay we expose some pseudocode to show readers that what we've been doing isn't magical, and to allow them to explore the parameter space more fully.

3 WHAT WE HOPE PEOPLE WILL TAKE AWAY

The basic goal of the essay is to get people to think more critically about histograms. Many people find histograms hard to interpret [5], so we are keen to clarify what a histogram can (and can't) tell you. A further goal is to alert data analysts to the importance of trying out alternative parameter values when creating a histogram.

Our long-term hope is that creators of chart-generation tools will start to incorporate adjustment facilities like the ones we show—not just for histograms, but for all parameter-sensitive visualizations—to encourage experimentation by analysts and readers alike. We would like it to become common knowledge that summary visualizations, as useful as they are, should always be open to probing and enquiry.

REFERENCES

[1] N. Case. Explorable explanations. `http://explorabl.es/`, May 2016.

[2] D. Ingalls, R. Krahn *et al.* The Lively Web. `https://lively-web.org/`, 2013.

[3] A. Lunzer and A. McNamara. It ain't necessarily so: Checking charts for robustness. In *IEEE Vis 2014*, 2014.

[4] A. Lunzer, A. McNamara, and R. Krahn. LivelyR: Making R charts livelier. In *useR! Conference*, 2014.

[5] J. M. Shaughnessy. *Second handbook of research on mathematics teaching and learning*, chapter Research on statistics learning and reasoning. National Council of Teachers of Mathematics, 2007.

[6] B. Victor. Explorable explanations. `http://worrydream.com/ExplorableExplanations/`, March 2011.

[7] T. Yamamiya, A. Warth, and T. Kaehler. Active essays on the web. Technical report, Viewpoints Research Institute, 2009.