Workshop: Preliminary Data Discovery

Goal of the Workshop

Review some key concepts

- Enter-exit-update pattern
- Layout functions

Goal of the Workshop

Review some key concepts

- Enter-exit-update pattern
- Layout functions

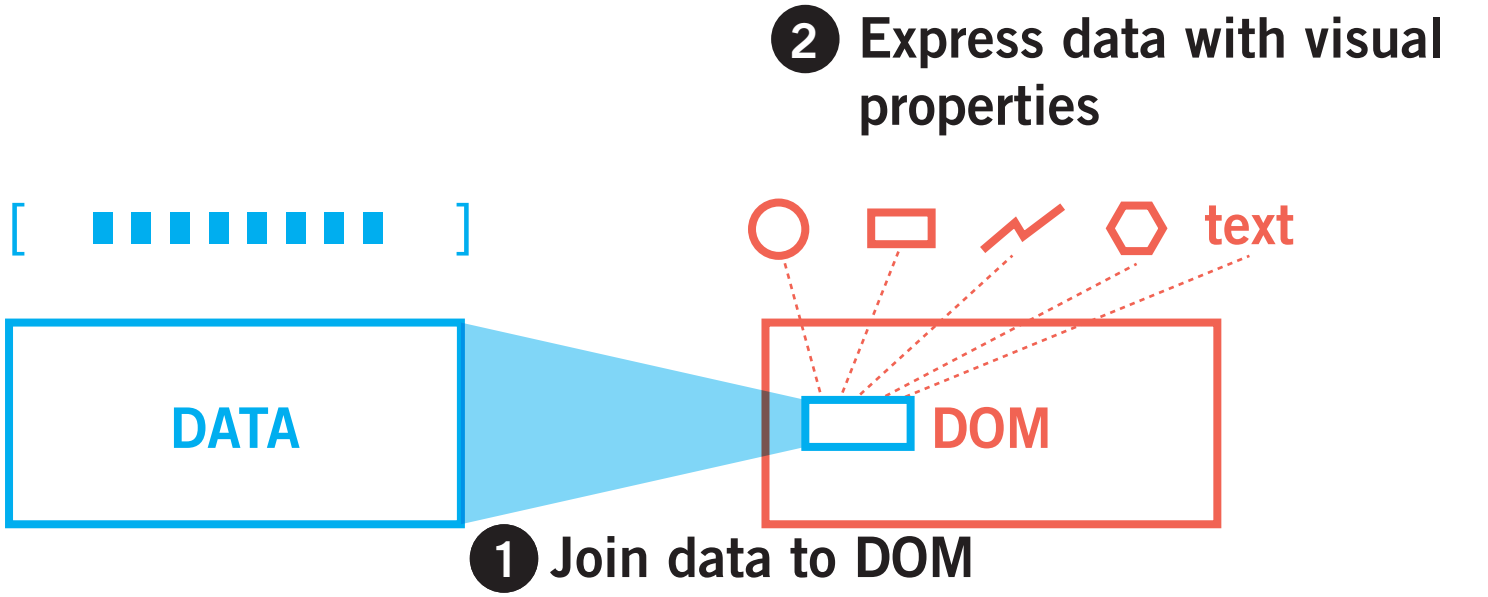
Perform some preliminary investigations into our data set, pose questions to it, and (hopefully) develop some broad insights and discover more questions.

Describing the Dataset

What is the language with which we describe a particular dataset?

- Sample
- Sample size

- Min and max
- Mean (“average”)
- Median


2 Express data with visual properties
text
DATA
DOM
1 Join data to DOM

Describing the Dataset, Visually

Visual strategies with which we can describe aspects of the data ("encoding")

Let's attempt it with a fictitious dataset.

Histogram

A graphical representation of the numerical **distribution** of a **continuous** variable.

Values are **binned** *i.e.* the range of values is divided into a series of intervals (which are consecutive and non-overlapping).

Let's see this in action.

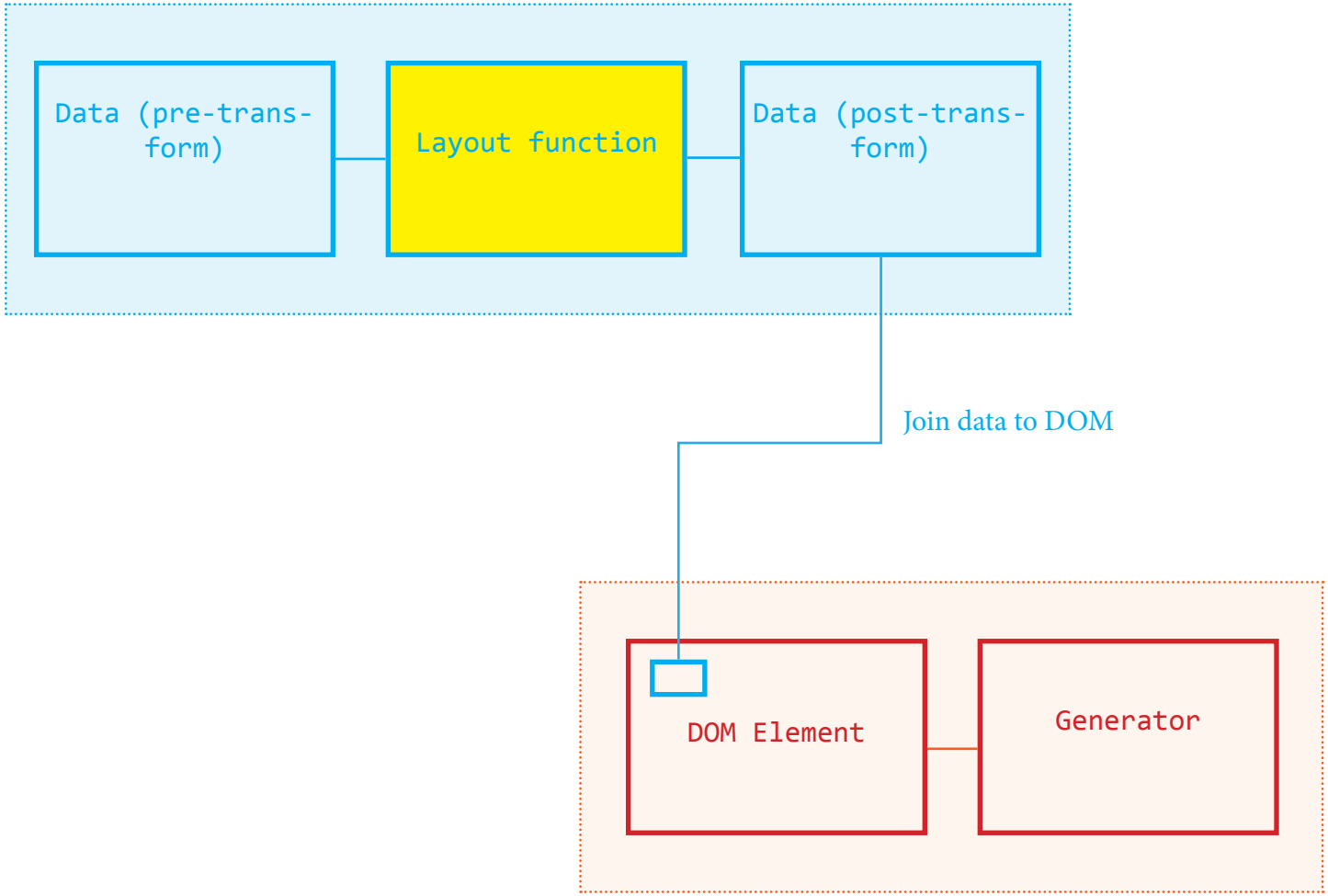
Histogram

Histograms provide a more nuanced representation of numerical distribution.

Consider issues of bin width.

Histogram

d3's implementation of the histogram is a **layout**.


Data (pre-trans-form)
Layout function
Data (post-trans-form)
Join data to DOM
DOM Element
Generator

Histogram

What values can we analyze with histogram?

Continuous vs. Categorical Values

See parallel coordinates vs. parallel set visualizations