

Paul Rosen

paul.rosen@utah.edu
@paulrosenphd
<https://cspaul.com>



Visualization for Data Science

DS-4630 / CS-5630 / CS-6630

FILTERING, AGGREGATION, & STATS

Reducing Items and Attributes

➔ Filter

➔ Items

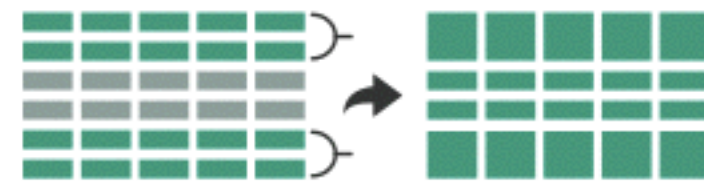


➔ Attributes

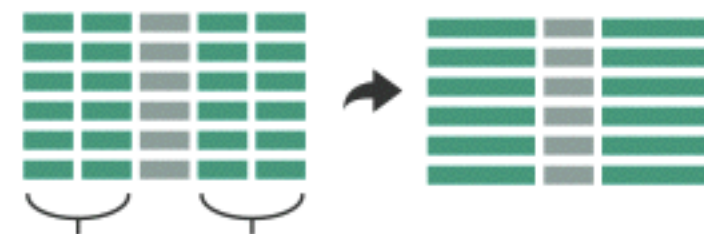


➔ Aggregate

➔ Items



➔ Attributes



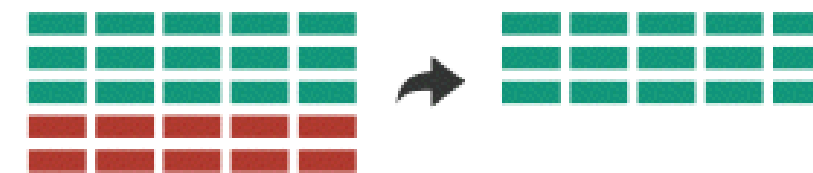
why reduce?

- Too many data items and/or too many attributes to focus on what is important in the data

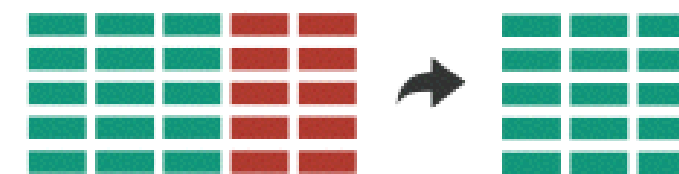
filter

- elements are eliminated to support dynamic queries
 - coupling between encoding and interaction so that user can immediately see the results of an action

→ Items



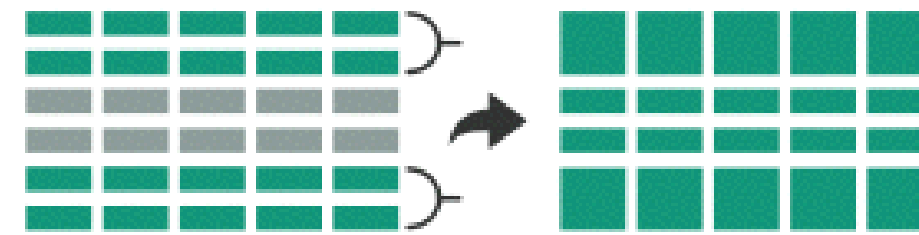
→ Attributes



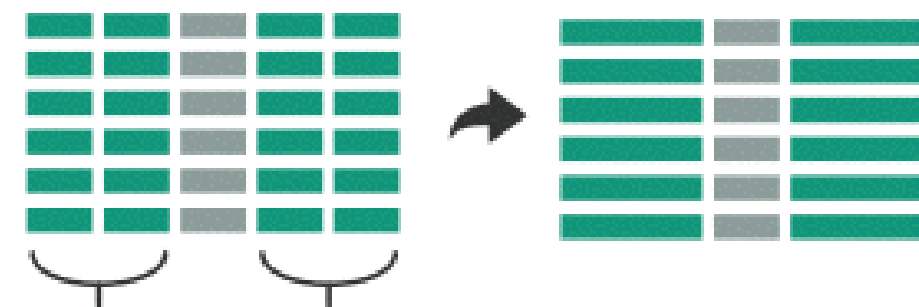
aggregate

- a group of elements is represented by a new derived element that stands in for the entire group

→ Items



→ Attributes



Numerous ways to reduce...

- statistics, topology, machine learning, etc.

Calculating a continuous histogram

- Given: $X = \{x_0, \dots, x_n\}$
- Select: k bins
- $\text{bin}_i = k * (x_i - \min X) / (\max X - \min X)$

Calculating a continuous histogram

- $X=\{1,2.5,3,4\}$
- $k = 3$



Calculating a continuous histogram

- $X = \{1, 2.5, 3, 4\}$
- $k = 3$



Calculating a continuous histogram

- $X = \{1, 2.5, 3, 4\}$
- $k = 3$
- $\text{bin}_i = \text{floor}(k * (x_i - \min X) / (\max X - \min X))$



Calculating a continuous histogram

- $X = \{1, 2.5, 3, 4\}$
- $k = 3$
- $\text{bin}_i = \text{floor}(3 * (x_i - 1) / (4 - 1))$



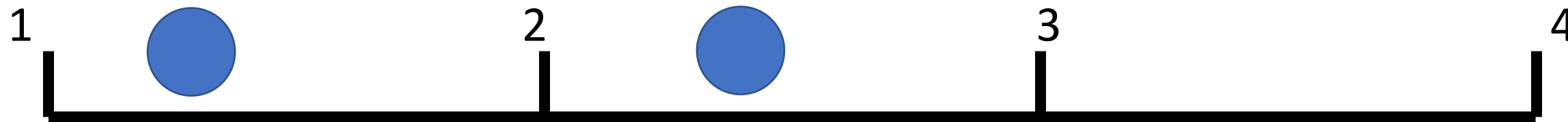
Calculating a continuous histogram

- $X=\{1,2.5,3,4\}$
- $k = 3$
- $1 \rightarrow \text{floor}(3 * (1 - 1) / (4 - 1)) = \text{Bin } 0$



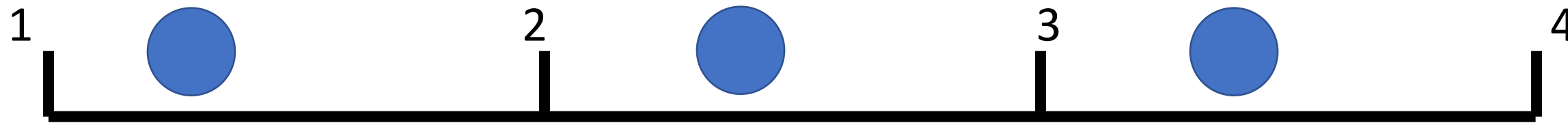
Calculating a continuous histogram

- $X=\{1,2.5,3,4\}$
- $k = 3$
- $2.5 \rightarrow \text{floor}(3 * (2.5 - 1) / (4 - 1)) = \text{Bin } 1$



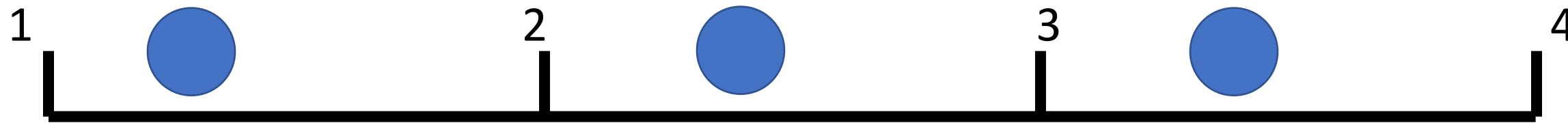
Calculating a continuous histogram

- $X=\{1,2.5,3,4\}$
- $k = 3$
- $3 \rightarrow \text{floor}(3 * (3 - 1) / (4 - 1)) = \text{Bin } 2$



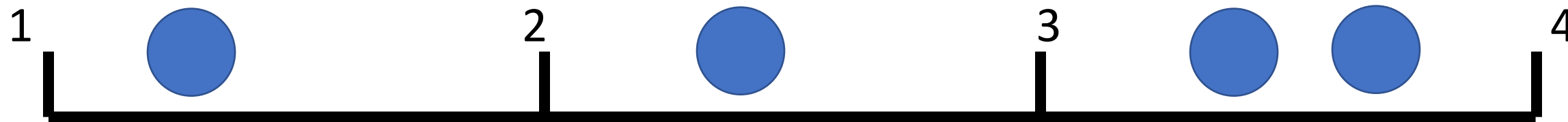
Calculating a continuous histogram

- $X = \{1, 2.5, 3, 4\}$
- $k = 3$
- $4 \rightarrow \text{floor}(3 * (4 - 1) / (4 - 1)) = \text{Bin } 3?$



Calculating a continuous histogram

- $X=\{1,2.5,3,4\}$
- $k = 3$
- $4 \rightarrow \text{floor}(3 * (4 - 1) / (4 - 1)) = \text{Bin } 2$



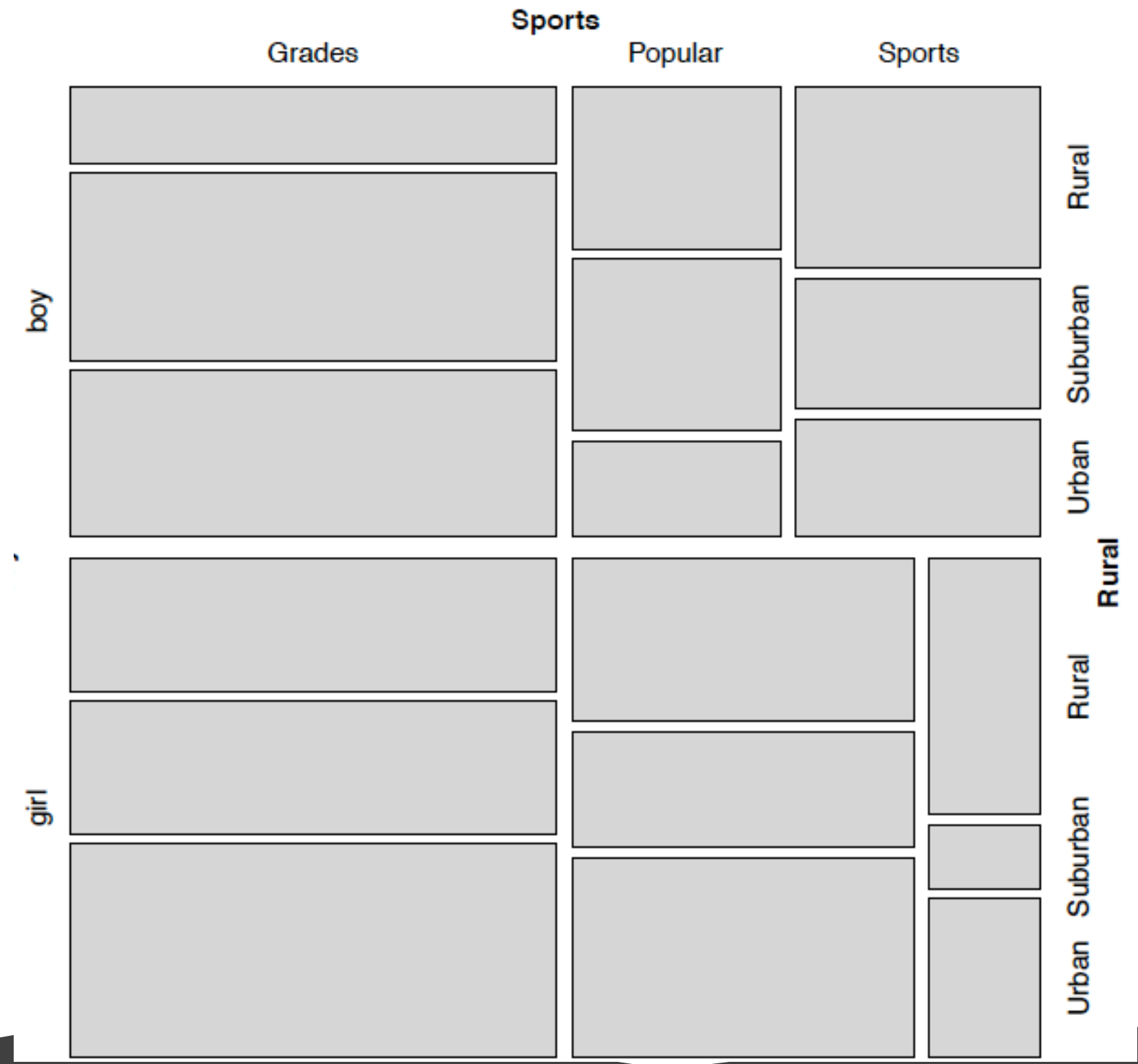
Calculating a continuous histogram

- $X = \{1, 2.5, 3, 4\}$
- $k = 3$



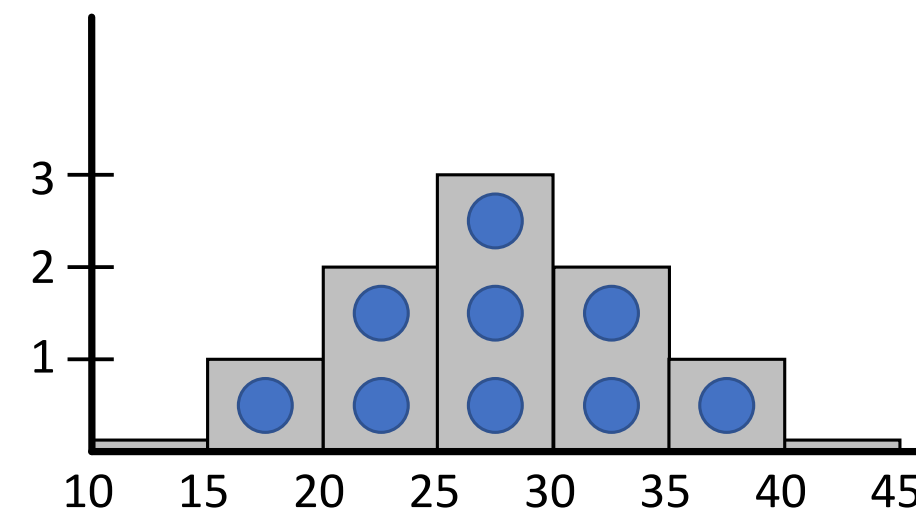
2D Histograms

Mosaic Plots



Histogram Challenges: Selecting Resolution

● 16
● 27
● 29
● 31
● 26
● 22
● 32
● 36
● 24



Mean (Average) = 27
Standard Deviation = 6

Boxplot

