



Statistical Graphics:

Creating Histograms

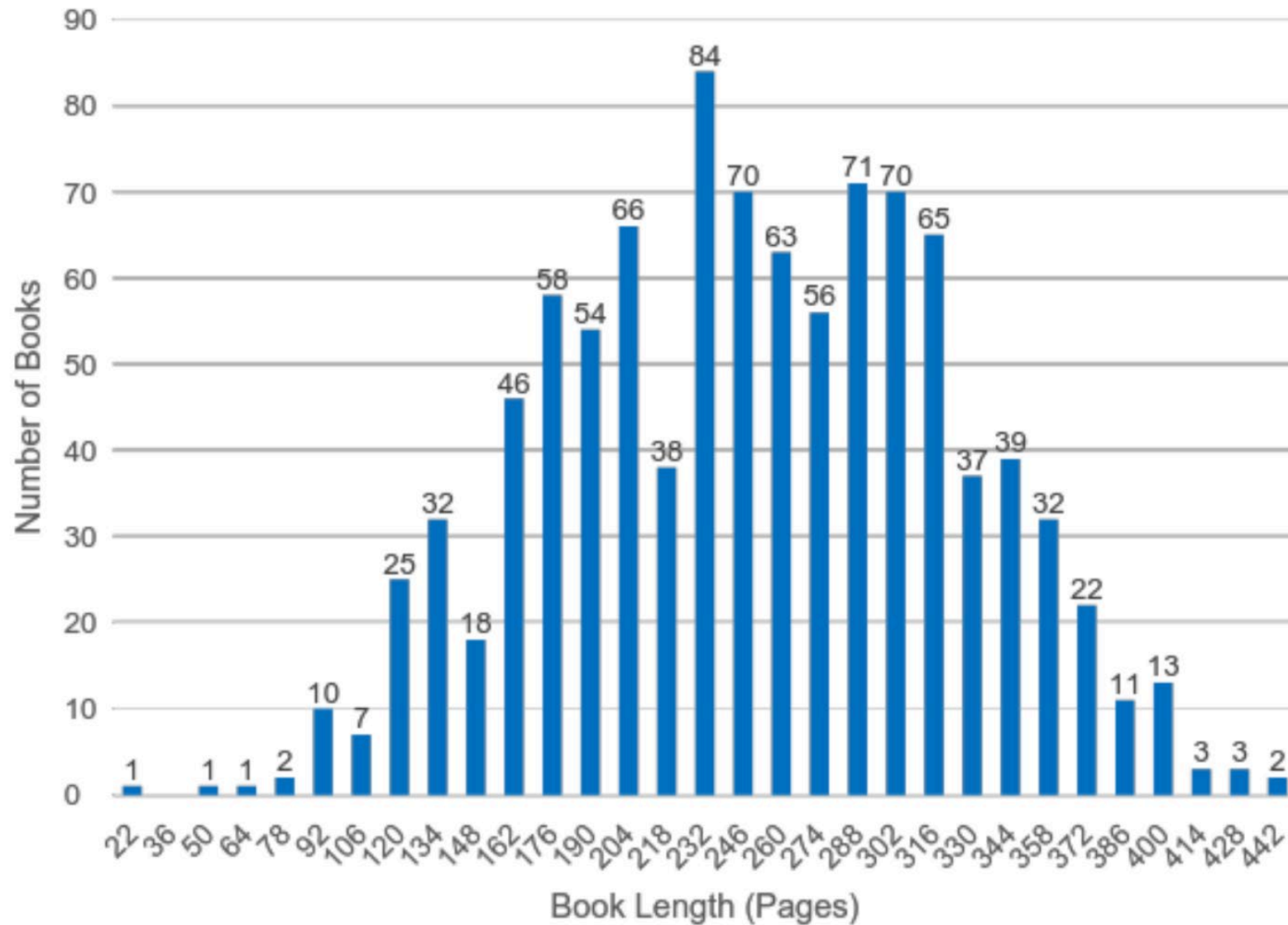
Objective



Objective

Define design principles for Histograms and the impact of parameter choices on the visualization.

Histograms



Histogram Binning



Number of bins (k) can be user-specified or chosen from a suggested bin width (h) such that:

$$k = \left\lceil \frac{\max x - \min x}{h} \right\rceil$$

Histogram Binning



Common choices for k include the square-root choice where:

$$k = \sqrt{N}$$

Histogram Binning

| Sturge's formula

$$k = \lceil \log N + 1 \rceil$$

| Scott's choice

$$h = \frac{3.5\sigma}{N^{\frac{1}{3}}}$$

| Freedman-Diaconis rule

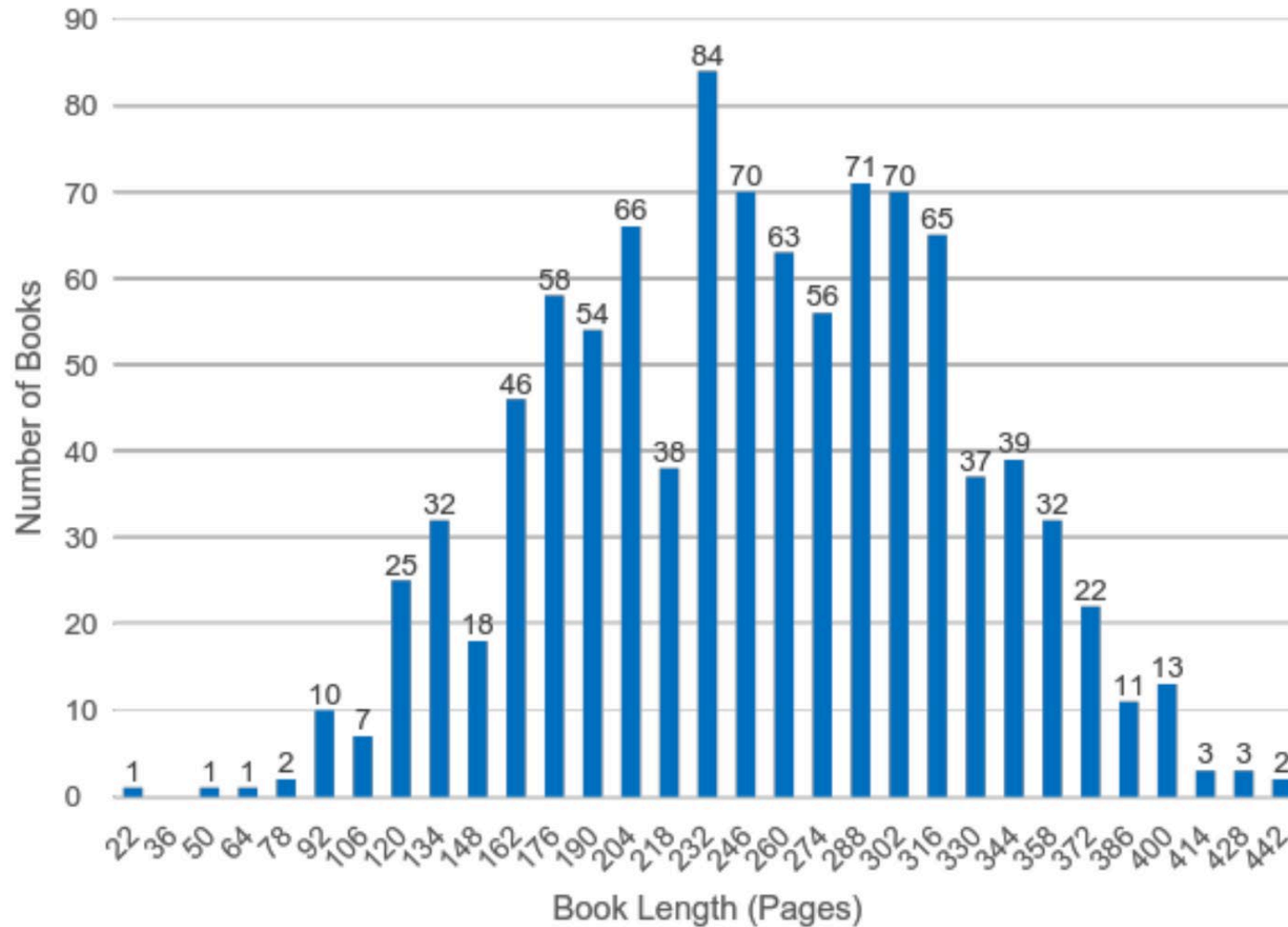
$$h = 2IQR(x)N^{-\frac{1}{3}}$$

Histogram Example

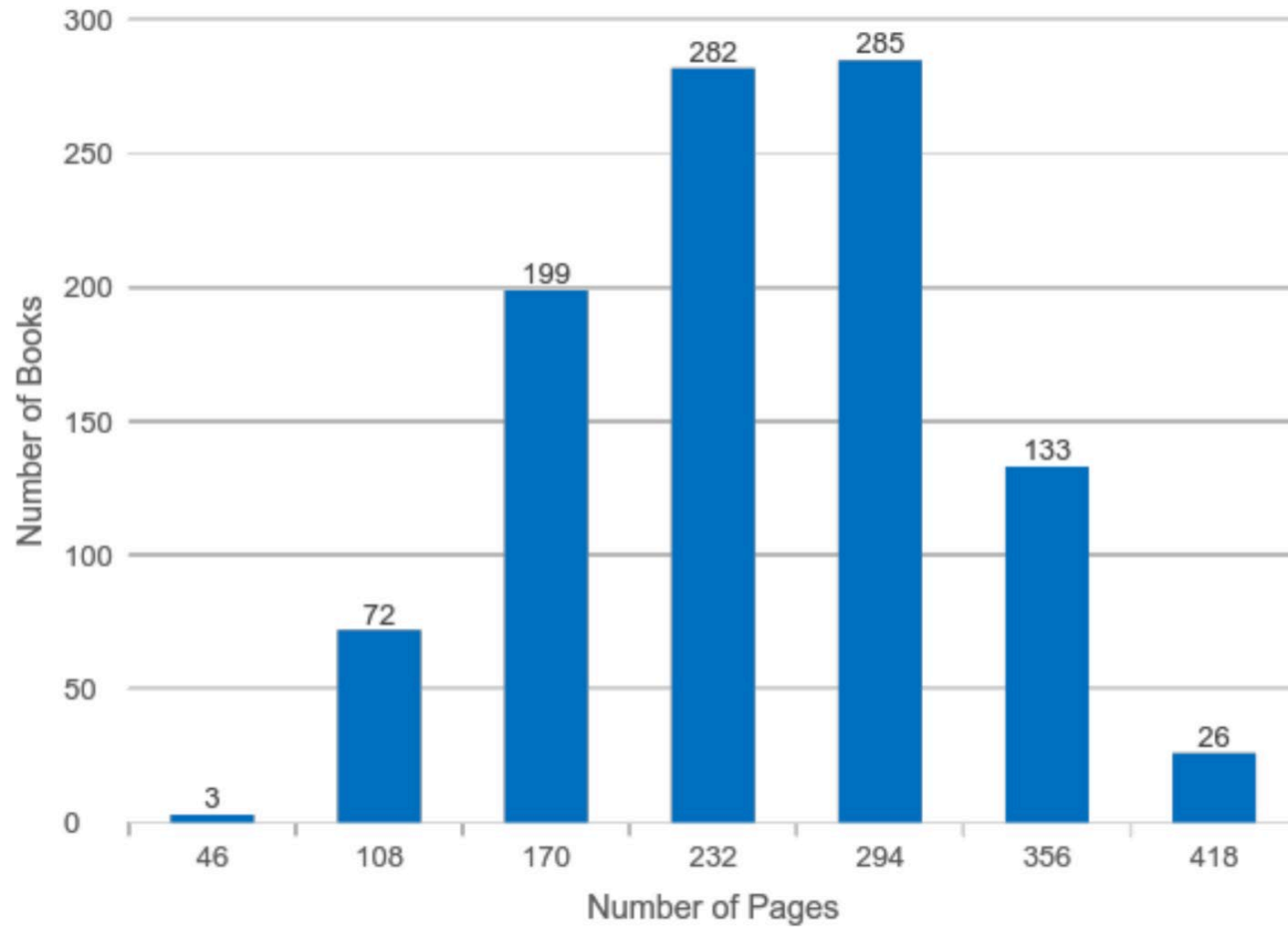


- Plot a histogram of 1000 book lengths.
- Use all four common choices for k or h .
- All x-axis labels indicate the center of the histogram bin.

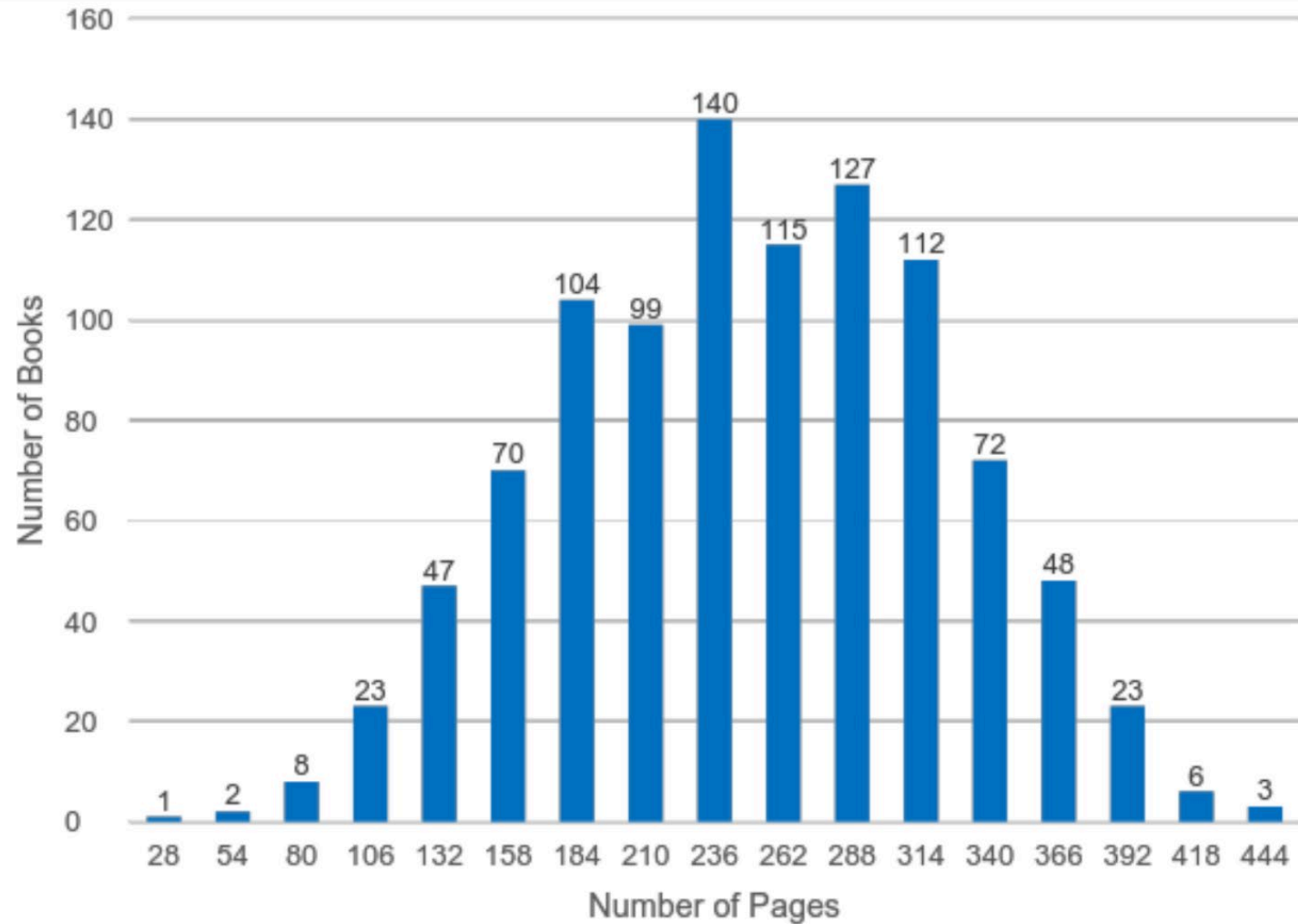
Square-Root Choice



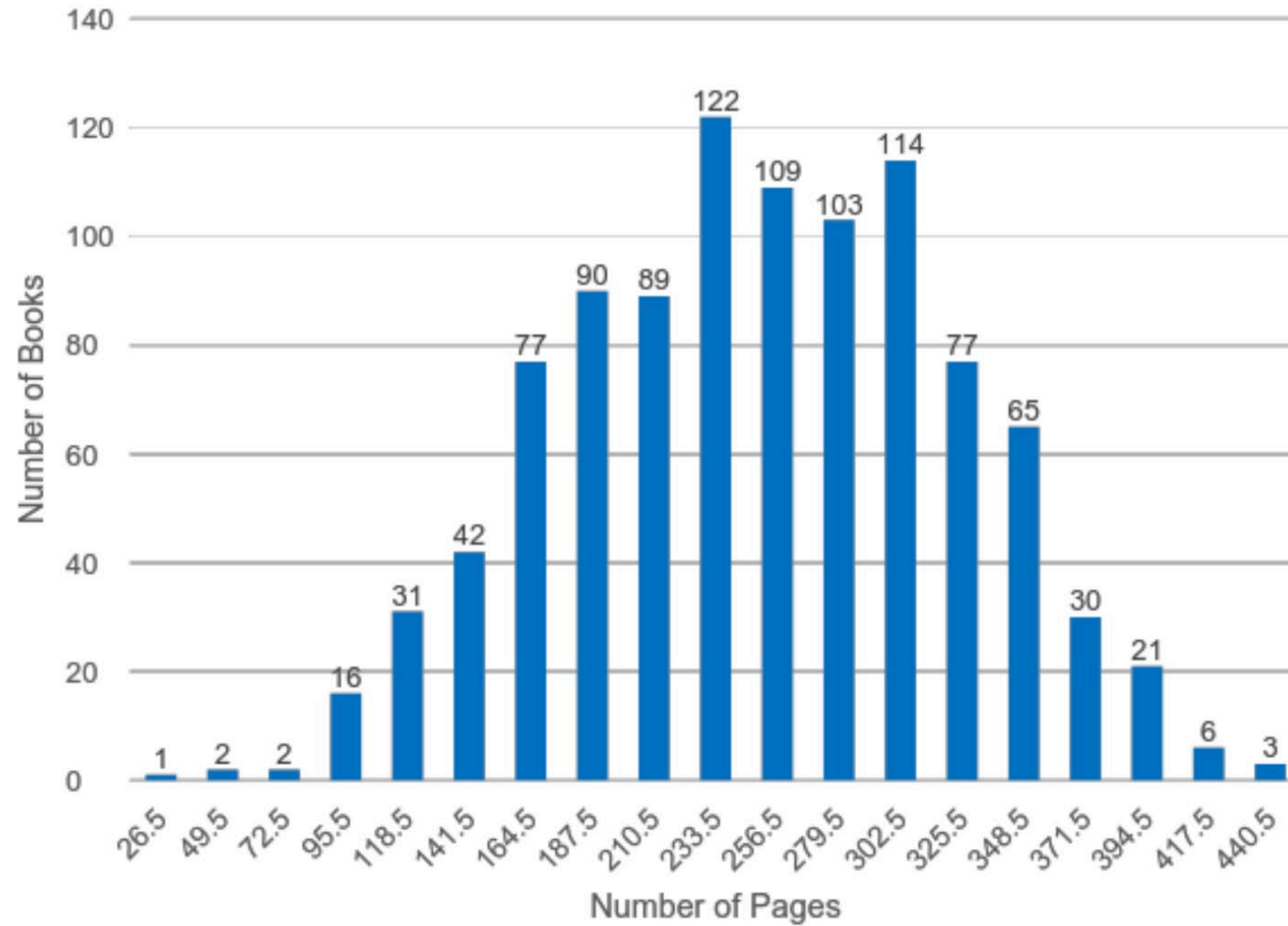
Sturge's Formula



Scott's Choice



Freedman-Diaconis Rule



Histograms

