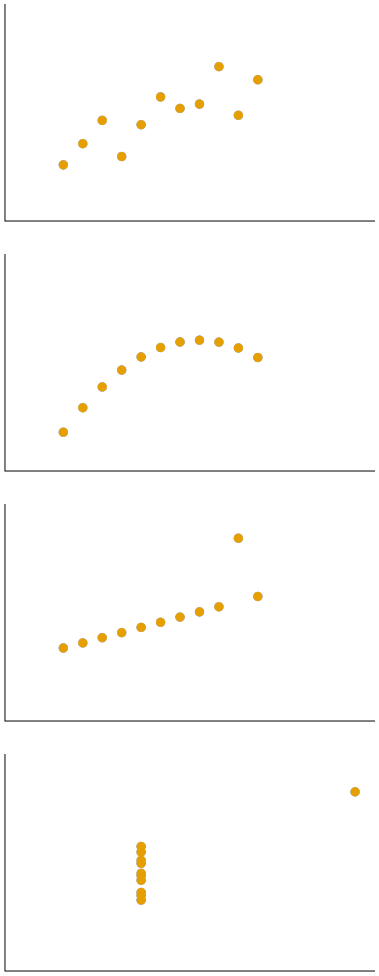
Exploring Data for Patterns

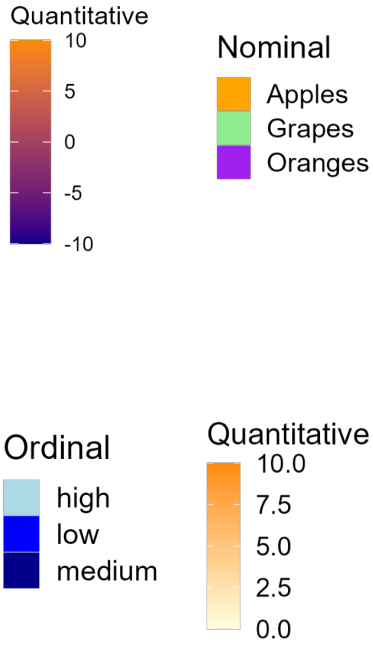
Patterns are the essence of data exploration and our eyes' ability to pick them out is integral to data understanding. Much of the data we work with, however, do not have a natural form and we need to make decisions about how they are to be represented. Try different ways to visualize the datasets so meaningful patterns may be found.

Consider the following concepts when creating your visualizations

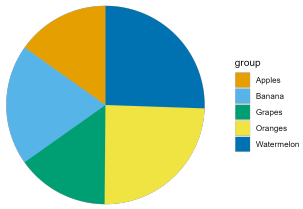
Patterns are the essence of data exploration. What kinds of representation will produce the most meaningful insights?

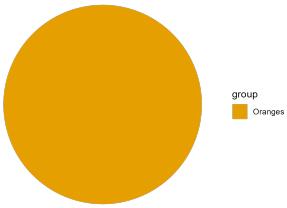


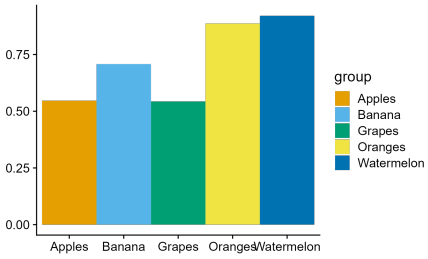
Color is a powerful encoding that presents several challenges. Have you chosen a color scale that is optimal for that data type?



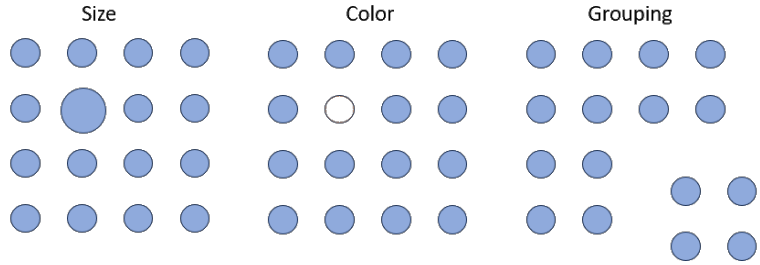
Encodings Some visual estimations are easier to make than others. How might you use encodings that are less accurate but otherwise better at conveying overall trends?







Salience and Relevance Pop-out effects enable quick recognition. Are the most noticeable elements of your visualizations also the most relevant?



1. Working in small groups, choose to work with dataset 1 or 2
2. Create different visualizations to highlight underlying patterns
3. Summarize the group's approach
4. Elect/volunteer a spokesperson to present the solution

Genetic profiles of cancer

These datasets contains 10 cancer samples. Table 1 describes the mutational status for a set of genes (A-E) and whether a mutation is absent (0) or present (1). Table 2 summarizes the

expression levels of those genes, ranging from no expression (0) to high expression (3).

Table 2: Mutational status for a set of genes

	1	2	3	4	5	6	7	8	9	10
A	0	0	0	0	0	1	0	0	0	0
B	0	0	0	0	1	1	1	0	1	1
C	0	0	1	0	0	0	1	1	1	1
D	1	1	0	0	1	1	0	0	0	0
E	0	1	1	0	1	0	0	0	1	0

Table 3: Expression levels for a set of genes

	1	2	3	4	5	6	7	8	9	10
A	2	1	1	2	2	0	2	1	1	2
B	1	1	2	1	0	0	0	2	0	0
C	1	1	3	1	2	2	3	0	3	0
D	0	0	2	1	3	3	2	1	1	1
E	1	3	3	1	3	1	2	1	3	2