Organizing Quantitative Data

To organize quantitative data, we start by organizing the data into classes (categories, or bins).

Guidelines for grouping:

1. The number of classes should be small enough to provide an effective summary, but large enough to display relevant characteristics of the data.
    - This can be a tricky balance! Usually 5-20 classes will do.
2. Each observation must belong to one - and only one - class.
3. Whenever possible, all classes should have the same width (same number of possible values).

Single-Value Grouping

If each class contains only one value, we call them single-value classes. This method of grouping is called single-value grouping.

This method is most appropriate when we have discrete data with a small number of distinct values.

Example: 50 randomly selected households were asked how many TVs they have. Values ranged from 0-6.

1	1	1	2	6	3	3	4	2	4
3	2	1	5	2	1	3	6	2	2
3	1	1	4	3	2	2	2	2	3
0	3	1	2	1	2	3	1	1	3
3	2	1	2	1	1	3	1	5	1

It makes the most sense to make each distinct number of TVs its own group:

Number of TVs	Frequency
0	1
1	16
2	14
3	12
4	3
5	2
6	2

Limit Grouping

In limit grouping, we use class limits. In this setting, each class consists of a **range of values**.

Each range has some smallest value - the lower limit - and some largest value - the upper limit.

Cheese consumption. The following data shows a year's worth of cheese consumption (lbs) for 35 randomly selected Americans who eat cheese.

45	28	32	37	41	39	33
32	31	35	27	46	25	41
35	31	44	23	38	27	32
43	32	25	36	26	30	35
36	36	35	21	43	35	28

We wish to use a limit grouping with a first class of 20-24.

The largest value is 46 and the smallest is 21.

Class	Data in Class
20-24	23, 21
25-29	28, 27, 25, 27, 25, 26, 28
30-34	32, 33, 32, 31, 31, 32, 32, 30
35-39	37, 39, 35, 35, 38, 36, 35, 36, 36, 35, 35
40-44	41, 41, 44, 43, 43
45-49	45, 46

Cutpoint Grouping

A third way to group quantitative data is cutpoint grouping. This uses class cutpoints. This is very similar to limit grouping, but can be used with continuous data.

Oxygen distribution in marine sediments ($mmol/m^2$).

1.8	2.0	1.8	2.3	3.8	3.4	2.7	1.1
3.3	1.2	3.6	1.9	7.6	2.0	1.5	2.0
1.1	0.7	1.0	1.8	1.8	6.7		

Use cutpoint grouping with first class 0 to < 1.

Class	Data
0 to < 1	0.7
1 to < 2	1.8, 1.8, 1.1, 1.2, 1.9, 1.5, 1.1, 1.0, 1.8, 1.8
2 to < 3	2.0, 2.3, 2.7, 2.0, 2.0
3 to < 4	3.8, 3.4, 3.3, 3.6
4 to < 5	
5 to < 6	
6 to < 7	6.7
7 to < 8	7.6

Choosing a Grouping Method

Grouping Method	Usage
Single-value Grouping	Discrete data with a small number of distinct values.
Limit Grouping	Discrete data with too many distinct values to reasonably use single-value grouping.
Cutpoint Grouping	Continuous data.

We can treat these classes as categories and use the same approaches as in organizing qualitative data.

However, often we prefer to use a method that's designed for numeric data.

Histograms

Def: A histogram displays the classes of the quantitative data on a horizontal axis and the frequencies on a vertical axis.

The frequency (relative frequency, percent) of each class is represented by a vertical bar whose height is equal to the frequency of that class.

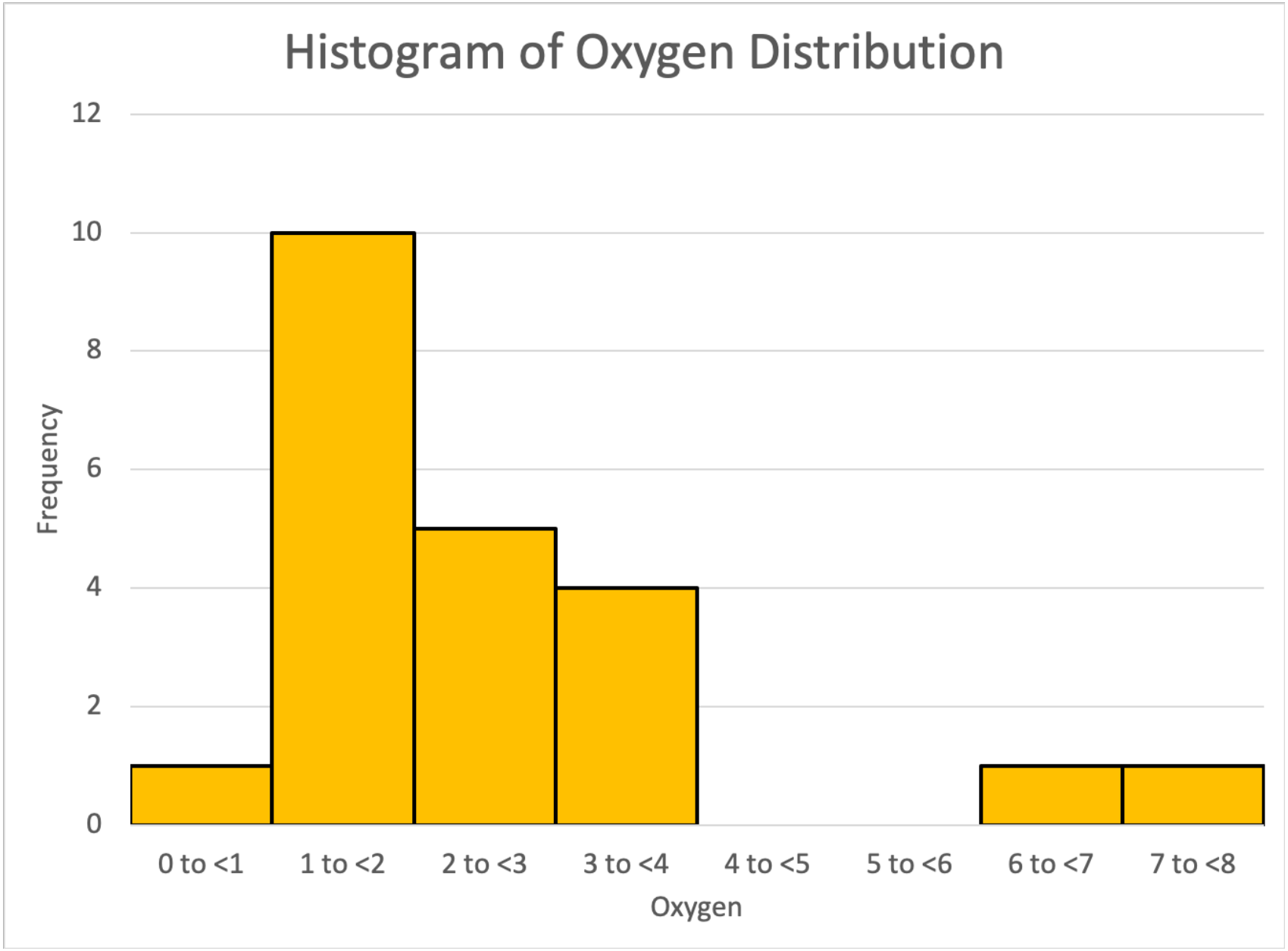
Typically the bars are positioned so that they touch each other.

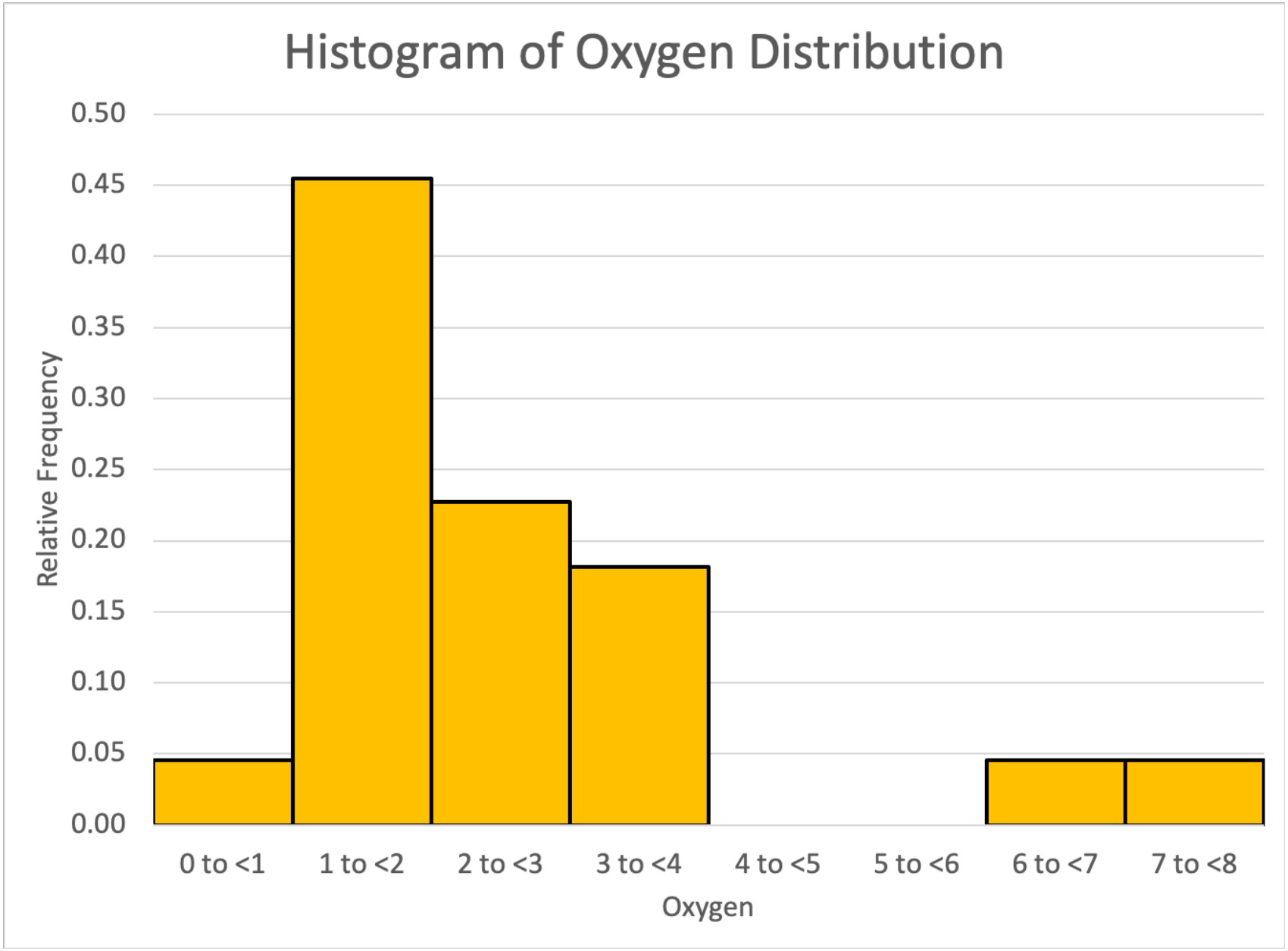
Constructing a Histogram

1. Determine the classes.
2. Obtain a (relative) frequency distribution.
3. For each class, construct a vertical bar whose height equals the frequency of that class.
4. Label everything.

For the marine oxygen data,

Class	Data
0 to < 1	1
1 to < 2	10
2 to < 3	5
3 to < 4	4
4 to < 5	
5 to < 6	
6 to < 7	1
7 to < 8	1


Histogram of Oxygen Distribution
12
10
8
6
4
2
0
Frequency
0 to <1
1 to <2
2 to <3
3 to <4
4 to <5
5 to <6
6 to <7
7 to <8
Oxygen


Histogram of Oxygen Distribution
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Relative Frequency
0 to <1
1 to <2
2 to <3
3 to <4
4 to <5
5 to <6
6 to <7
7 to <8
Oxygen

Dot plots

Def: A dot plot is a graph in which each observation is plotted as a dot at an appropriate place above a horizontal line. Observations with equal values are stacked vertically.

Constructing a dot plot:

1. Draw a horizontal axis that displays the possible values (or classes) of the quantitative data.
2. Record each observation by placing a dot over the appropriate value on the horizontal axis.
3. Label the horizontal axis with the name of the variable.

A survey of "How long does it take you to eat breakfast?" has these results[1]:

Minutes	Frequency
0	6
1	2
2	3
3	5
4	2
5	5
6	0
7	0
8	2
9	3
10	7
11	4
12	1

[1] Example from mathisfun.com

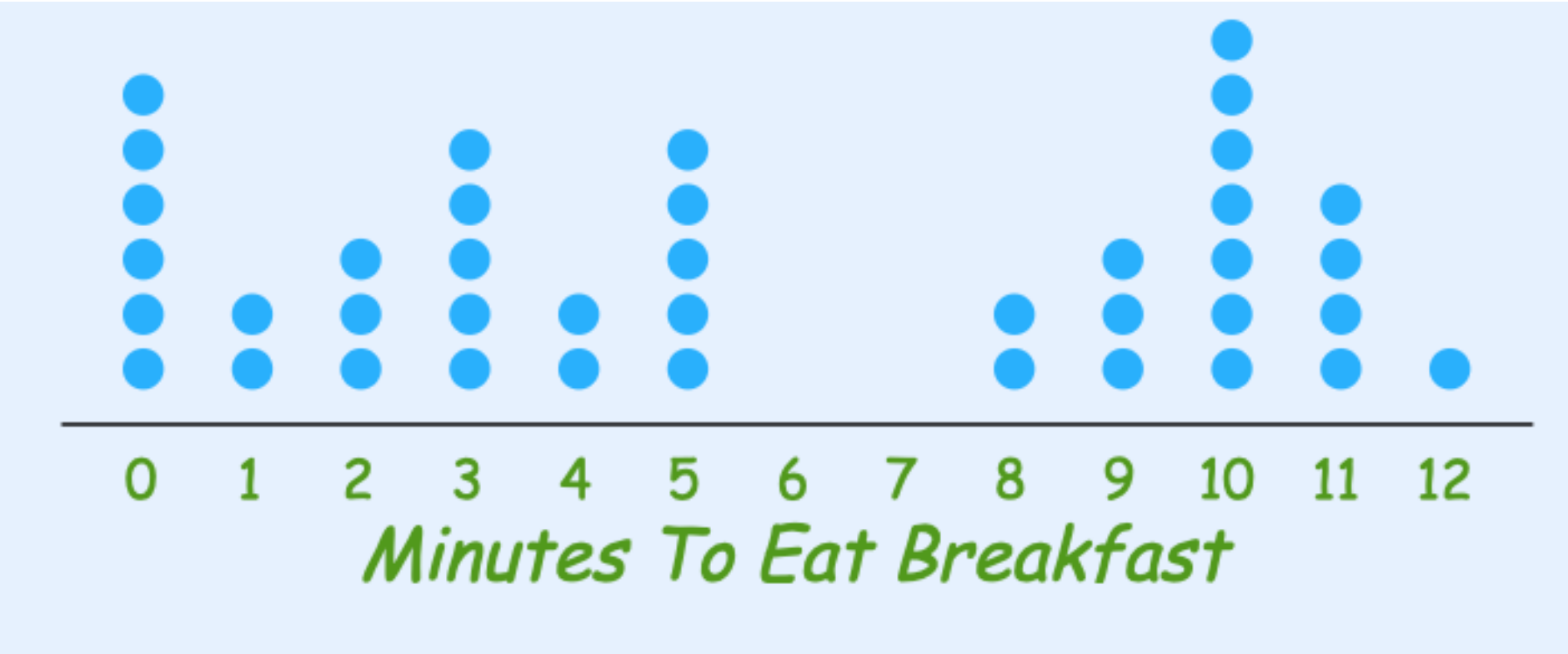


Image from mathisfun.com

Stem and Leaf Diagrams

Def: In a stem and leaf diagram, each observation is separated into two parts - a stem and a leaf.

Typically, the leaf is the ones digit and the stem is everything else.

Constructing a stem and leaf diagram:

1. Think of each observation as a stem and a leaf, the rightmost digit.
2. Write the stems from smallest to largest in a vertical column. Draw a vertical line down the right side of this column.
3. Write each leaf to the right of the vertical line on the line with the appropriate stem.
4. Arrange the leaves in each row from smallest to largest.

For the cheese consumption data,

45	28	32	37	41	39	33
32	31	35	27	46	25	41
35	31	44	23	38	27	32
43	32	25	36	26	30	35
36	36	35	21	43	35	28

the stem and leaf for the first observation, 45, will look like 4|5.

The stem and leaf plot is

2	135567788
3	01122223555556667 89
4	1133456

Based on this plot, we can identify $2|1 = 21$, $2|7 = 27$, $4|3 = 43$, etc.