Displaying Distributions - Quantitative Variables

Lecture 13
Sections 4.4.1 - 4.4.3

Robb T. Koether

Hampden-Sydney College

Wed, Feb 8, 2012

Outline

Outline

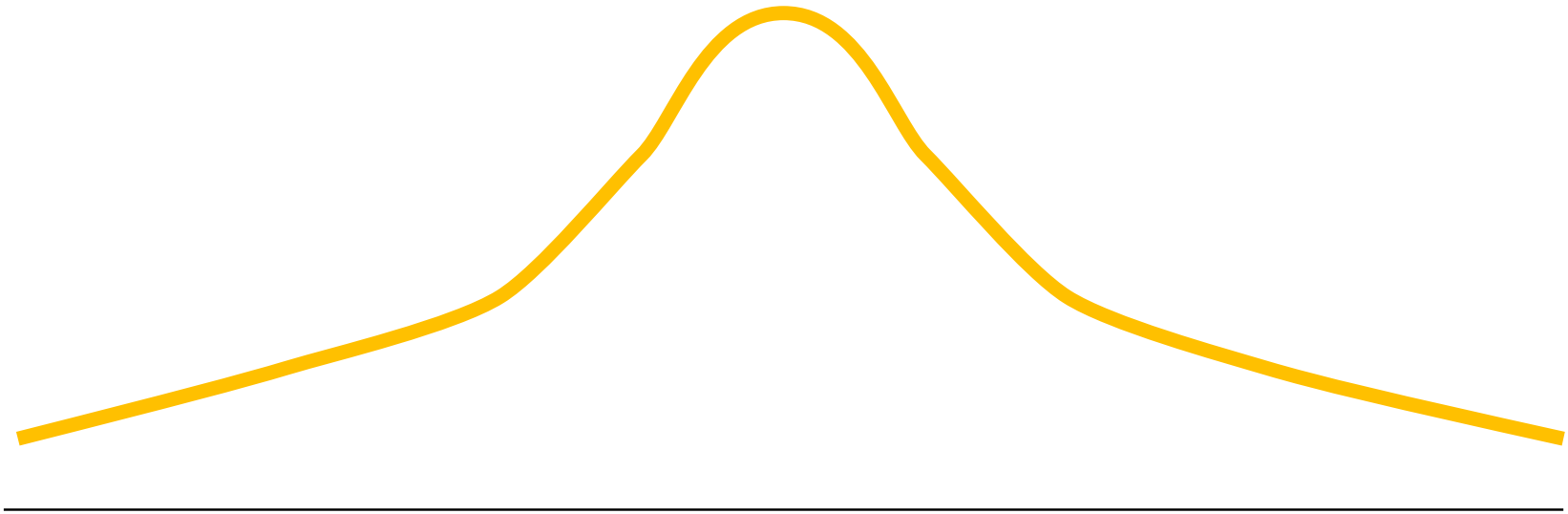
Introduction

- Now we wish to create displays of quantitative data.

Introduction

- Now we wish to create displays of quantitative data.
- With quantitative data, what is it that we wish to convey through our display?

Introduction

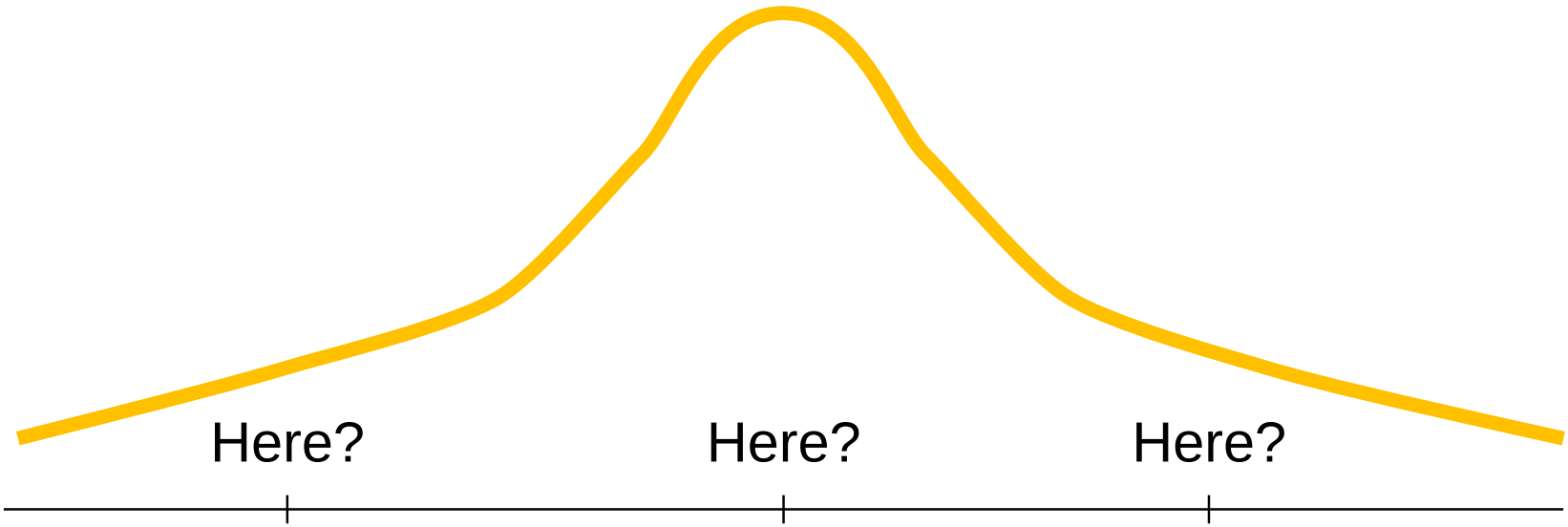


Are the data clustered around the middle?

Introduction

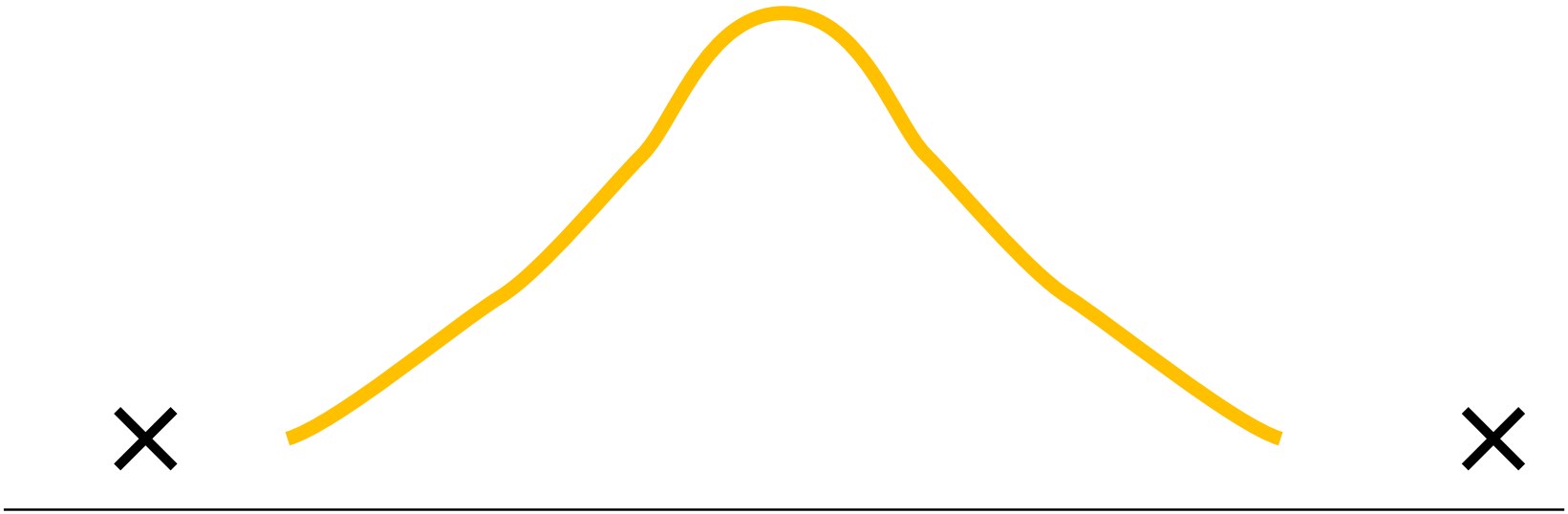
Or are they spread out?

Introduction



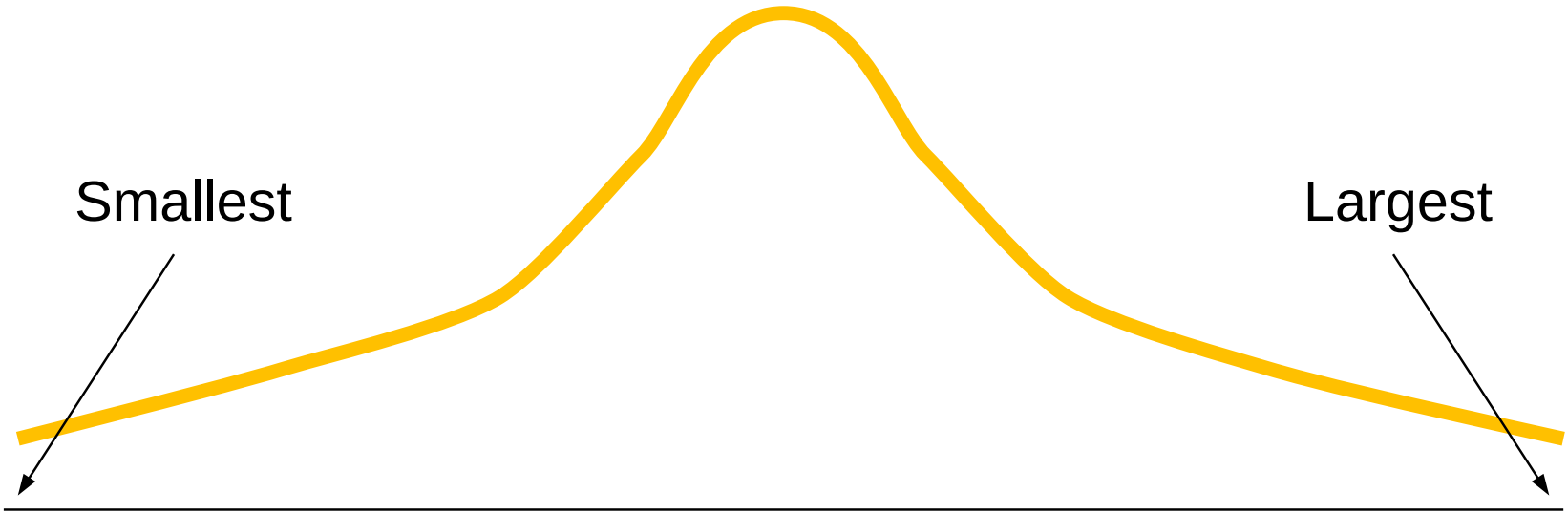
Where is the middle?

Introduction



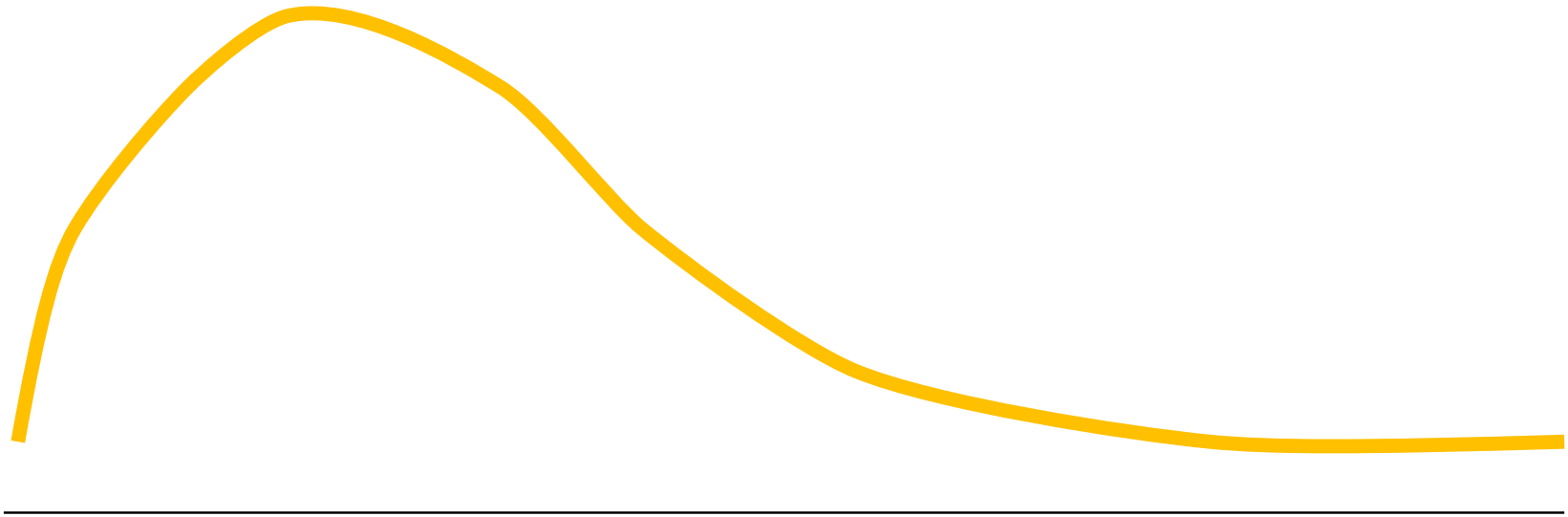
Are they any outliers?

Introduction



What are the extremes?

Introduction



Are the values skewed to the left or right?

Introduction

- To obtain such information, we need a graph that plots the value of the variable on one axis and its frequency on the other axis.

Outline

Example

- The following data represent rainfall amounts at the Richmond airport in September for the years 1981 - 2010 (30 years).

9.52	0.08	6.14	8.68	2.93	2.03
3.60	14.71	4.01	0.85	6.89	11.07
4.42	3.41	2.85	2.56	1.92	5.15
1.58	4.44	0.77	4.76	1.15	3.02
1.73	2.60	2.56	10.01	2.46	6.49

- How might we display these data graphically?

Frequency Plots

Definition (Frequency plot)

A frequency plot is a display of quantitative data in which each data point is represented by an **X** over that value on a horizontal scale.

Drawing Frequency Plots

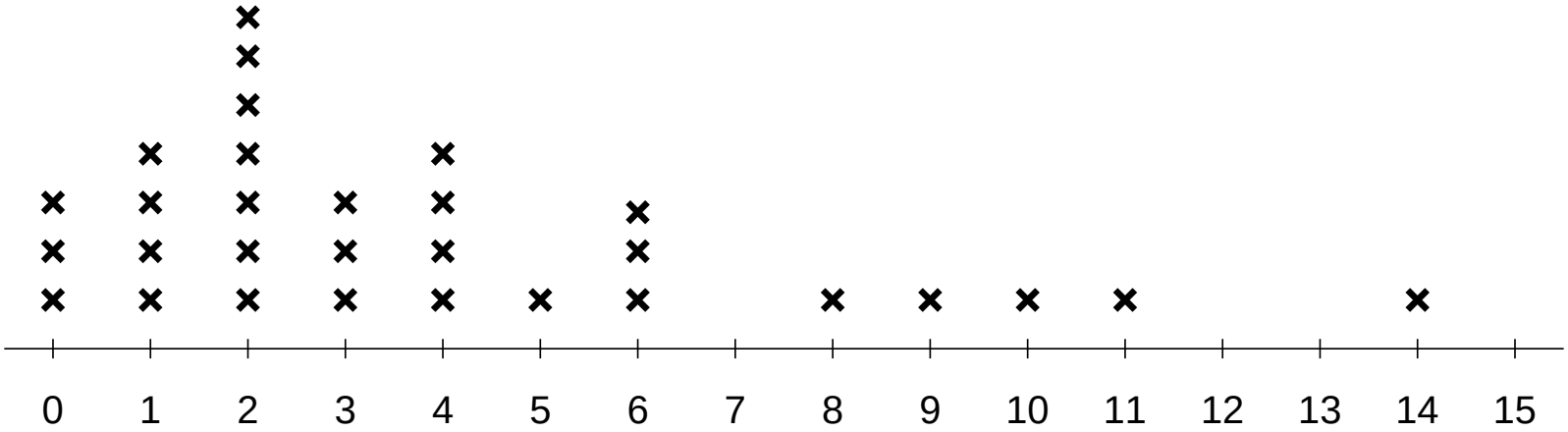
- Draw a horizontal line.
- Choose a resolution, e.g., 0.1.
- Keeping in mind the minimum and maximum values, mark reference points on the scale, as on a ruler.
  - Mark at *regular* intervals.
- For each data value, draw an **X** over that value on the scale.

Example

- Make a frequency plot of the rainfall data.

9.52	0.08	6.14	8.68	2.93	2.03
3.60	14.71	4.01	0.85	6.89	11.07
4.42	3.41	2.85	2.56	1.92	5.15
1.58	4.44	0.77	4.76	1.15	3.02
1.73	2.60	2.56	10.01	2.46	6.49

Frequency Plots



- What information is conveyed by this frequency plot?

Outline

Shapes of Distributions

Definition (Symmetric)

The distribution is symmetric if the left side is a mirror image of the right side.

Definition (Unimodal)

The distribution is unimodal if it has a single peak, showing the most common values.

Definition (Bimodal)

The distribution is bimodal if it has two peaks.

Shapes of Distributions

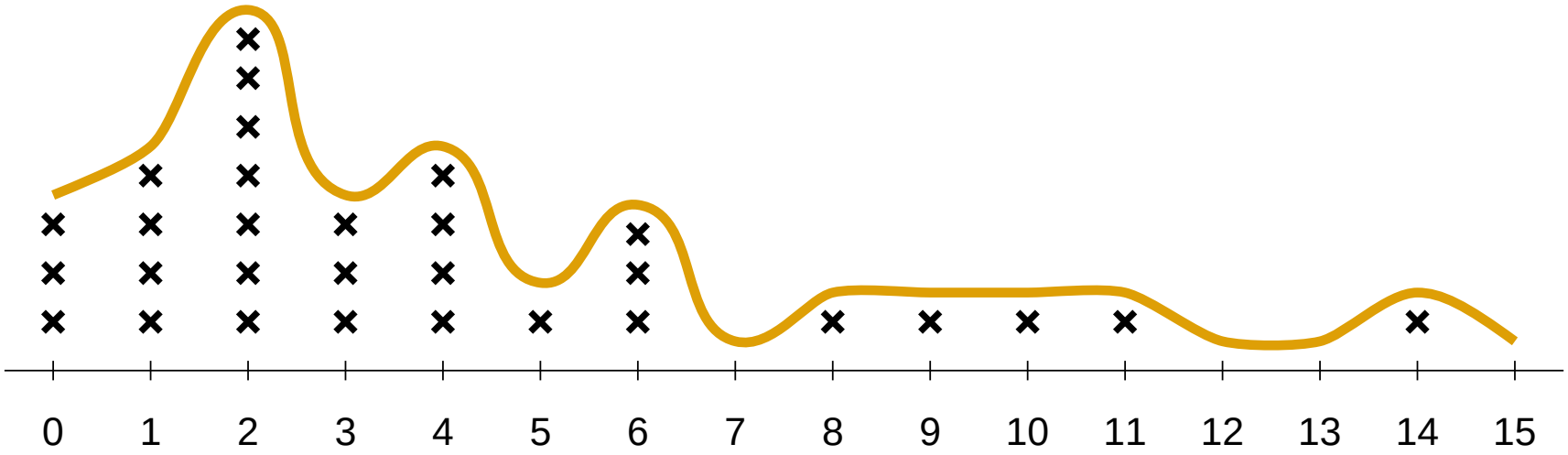
Definition (Uniform)

The distribution is uniform if all values have equal frequency.

Definition (Skewed)

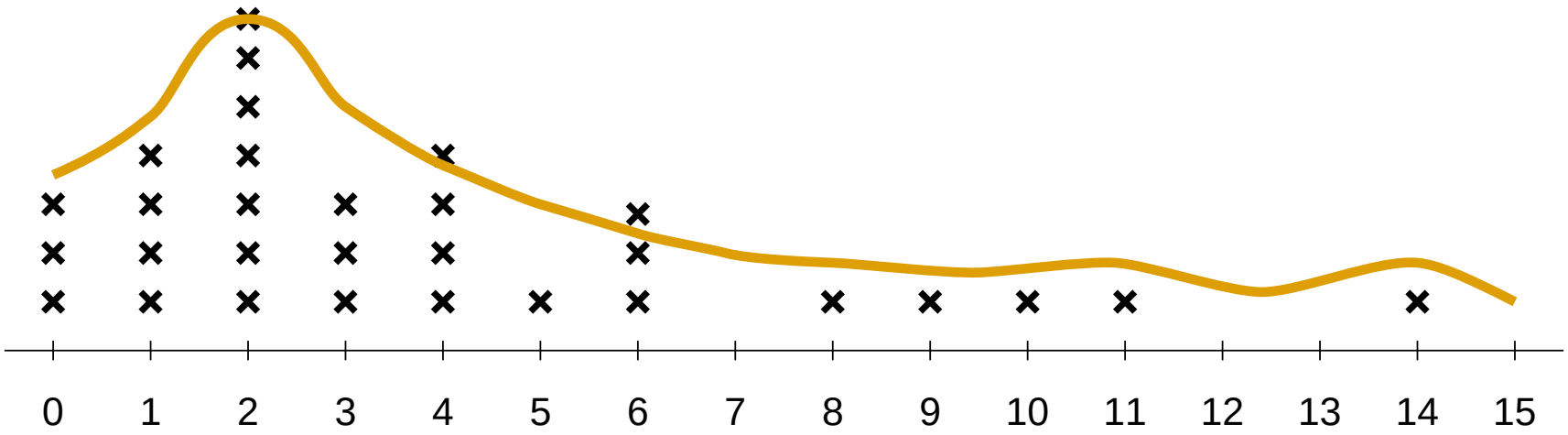
The distribution is skewed if it is stretched out more on one side than the other.

Shapes of Distributions



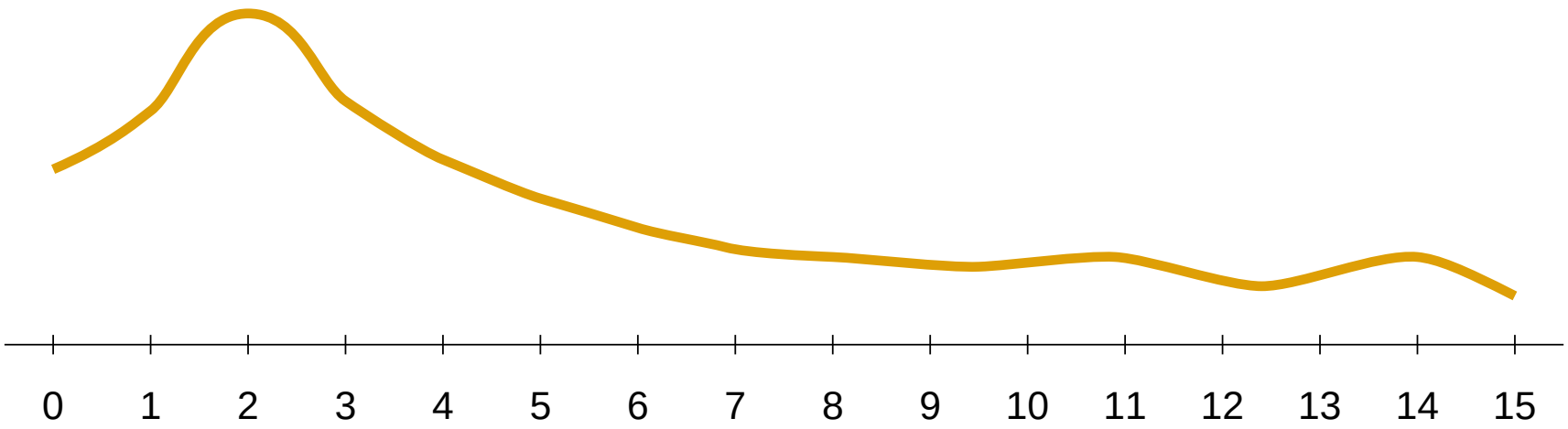
What properties does this distribution have?

Shapes of Distributions



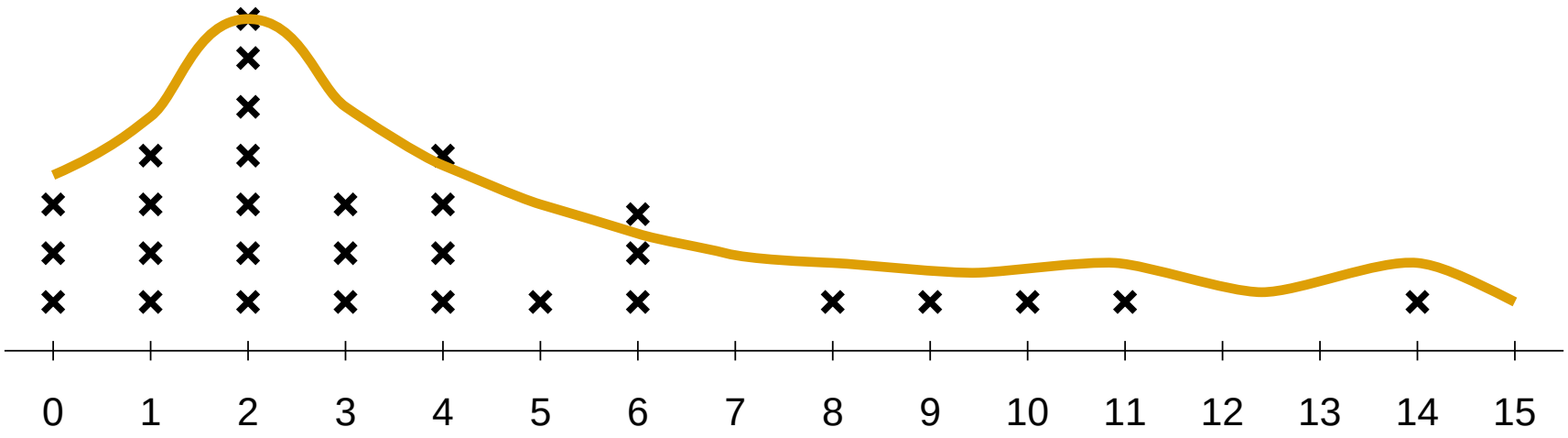
Maybe smooth it out a bit.

Shapes of Distributions



What properties does this distribution have?

Shapes of Distributions



Is 14 an outlier?

Outline

Stem-and-Leaf Displays

Definition (Stem-and-leaf display)

A stem-and-leaf display is a display of quantitative data in which the numerical representation of each data value is split into a stem and a leaf. The stem is the part to the left of the division point. The leaf is the first digit to the right of the division point.

Stem-and-Leaf Displays

- For example, the value 1.23 could be split as
  - 1|23 – stem = 1, leaf = 2, or
  - 12|3 – stem = 12, leaf = 3.
- Indeed, it could even be split as
  - |123 – stem = 0, leaf = 1, or
  - 123| – stem = 123, leaf = 0.

Stem-and-Leaf Displays

- The stem consists of the digits to the left of the division point.
- The leaf consists of the first digit to the right of the division point.
- A note should be added indicating how to interpret the numbers.
  - Note: 12|3 means 1.23.

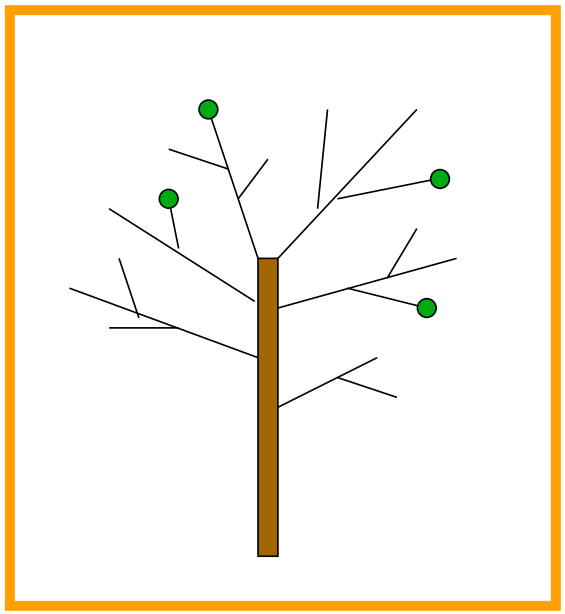
Example

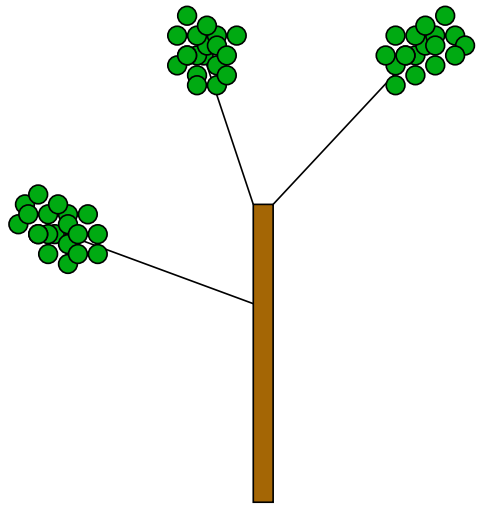
- Draw a stem-and-leaf display of the rainfall data.

9.52	0.08	6.14	8.68	2.93	2.03
3.60	14.71	4.01	0.85	6.89	11.07
4.42	3.41	2.85	2.56	1.92	5.15
1.58	4.44	0.77	4.76	1.15	3.02
1.73	2.60	2.56	10.01	2.46	6.49

Splitting the Numbers

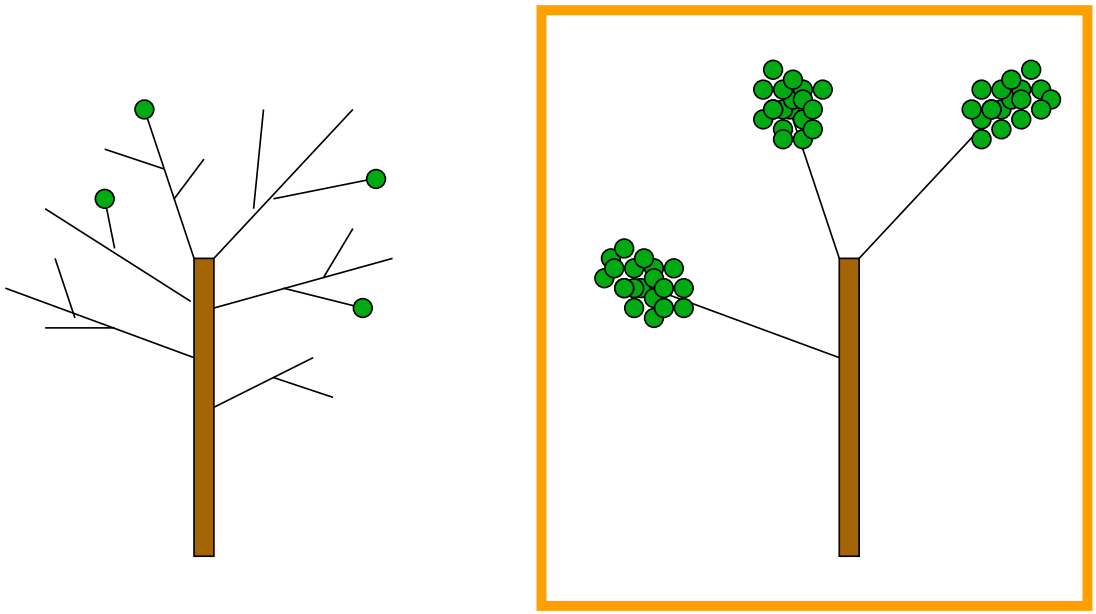
- We choose where to split the numbers in order to avoid
  - Too many stems, each with too few leaves.
  - Too few stems, each with too many leaves.





Splitting the Numbers

- We choose where to split the numbers in order to avoid
  - Too many stems, each with too few leaves.
  - Too few stems, each with too many leaves



Example

- Draw a stem-and-leaf display of the rainfall data, in centimeters.

24.18	0.20	15.60	22.04	7.44	5.16
9.14	37.36	10.19	2.16	17.50	28.12
11.23	8.66	7.24	6.50	4.88	13.08
4.01	11.28	1.96	12.09	2.92	7.67
4.39	6.60	6.50	25.43	6.25	16.48

Example

- We may split the values after the 10's digit:

Stem	Leaf
0	0 7 5 9 2 8 7 6 4 4 1 2 7 4 6 6 6
1	5 0 7 1 3 1 2 6
2	4 2 8 5
3	7

- Note: 1|2 means 12.

Example

- Or we may split the values at the decimal point:

Stem	Leaf
0	2
1	9
2	1 9
3	
4	8 0 3
5	1
6	5 6 5 2
7	4 2 6
⋮	⋮
37	3

- Note: 12|3 means 12.3.

Example

- Which is better?
- Is either one particularly good?

Outline

Splitting the Stems

- We can obtain a good compromise (in this example) by splitting the stems.
- Each stems appears twice.
- The first time for leaves 0 - 4.
- The second time for leaves 5 - 9.

Splitting the Stems

- Display with split stems:

Stem	Leaf
0	0 2 4 4 1 2 4
0	7 5 9 8 7 6 7 6 6 6
1	0 1 3 1 2
1	5 7 6
2	4 2
2	8 5
3	
3	7

- Note: 1|2 means 12.

Outline

Example

- Draw back-to-back stem-and-leaf displays of the September rainfall vs. the April rainfall 1981 - 2010 (shown below).

2.63	4.25	3.82	3.88	3.04	2.48
2.14	2.60	4.48	2.25	1.05	2.72
4.86	4.05	14.59	5.33	3.61	5.72
1.33	1.26	2.47	0.80	4.12	1.87
1.39	1.47	2.06	3.41	0.74	2.82

Outline

Assignment

Homework

- Read Section 4.4.1 - 4.4.3, pages 238 - 241, 242 - 248.
- Let's Do It! 4.10, 4.12.
- Page 241, exercise 18.
- Page 248, exercises 22 - 27, 29.