Statistical Methods

Lecture 2 – Describing Data-Sets

Luke Dickens

Autumn 2020

Python: List Comprehensions

Frequency tables and graphs

Grouped data and histograms

Other representations

List Comprehensions

List comprehensions are a simple, compact way to construct, and convert between, lists. Let's see some on this week's notebook.

Python: List Comprehensions

Frequency tables and graphs

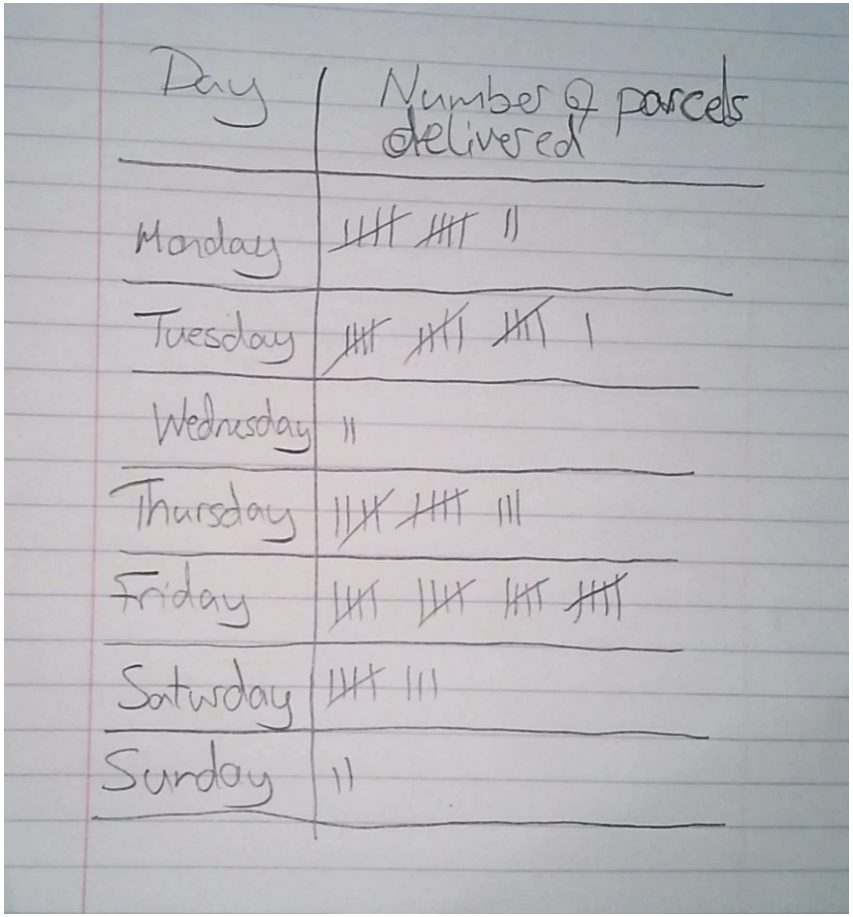
Grouped data and histograms

Other representations

Frequency counts

One of the most common form of statistical procedures is to count the frequency of occurrences.

- data can be numeric, but doesn't have to be
- can display as a collection of observations
- OR as a table of frequency counts
- OR as a plot
- OR ...



All forms can represent the same data. We sometimes want to transform data into representations that better show the patterns.

Some terms

- **variable** a characteristic that changes or varies over time and/or for different individuals/objects, e.g. body temperature
- **experimental unit** the individual or object on which a variable is measured.
- **measurement (or data value)** results when a variable is measured on an experimental unit
- **population** the set of all measurements of interest
- **sample** a subset of measurements selected from the population

Definitions as found in [MBB12]

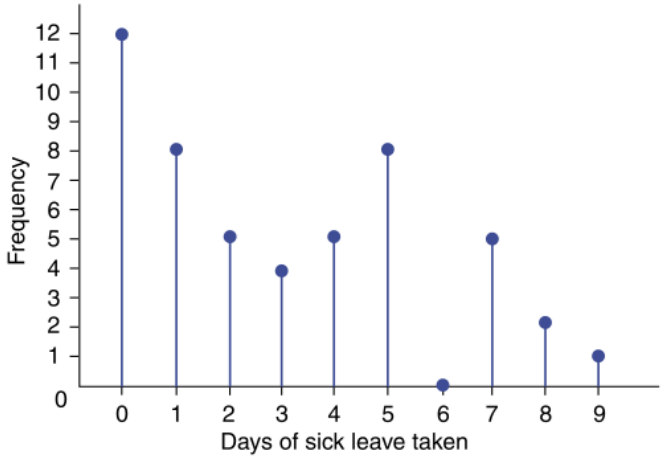
Different representations

Representations from [Ros17, Chp. 2] of the sick day data.

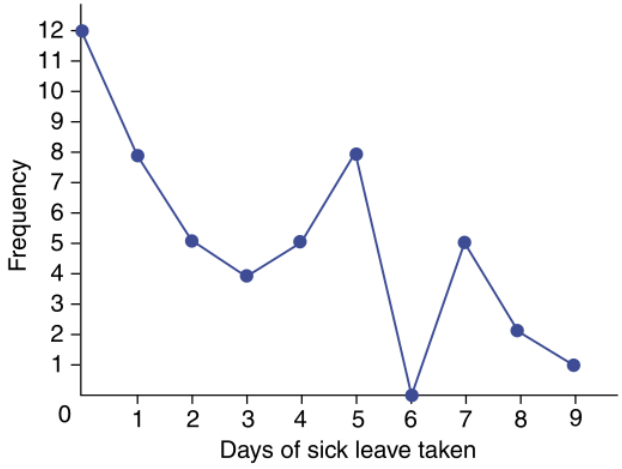
Table 2.1 A Frequency Table of Sick Leave Data

Value	Frequency	Value	Frequency
0	12	5	8
1	8	6	0
2	5	7	5
3	4	8	2
4	5	9	1

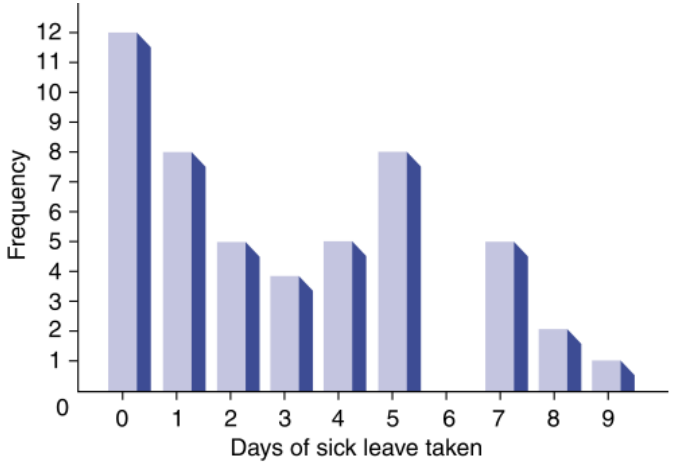
Frequency table



Line graph



Frequency polygon



Bar graph

Things to look for

Key features of a plot might include:

- Zero counts
- Symmetry
- Outliers
- Any other patterns

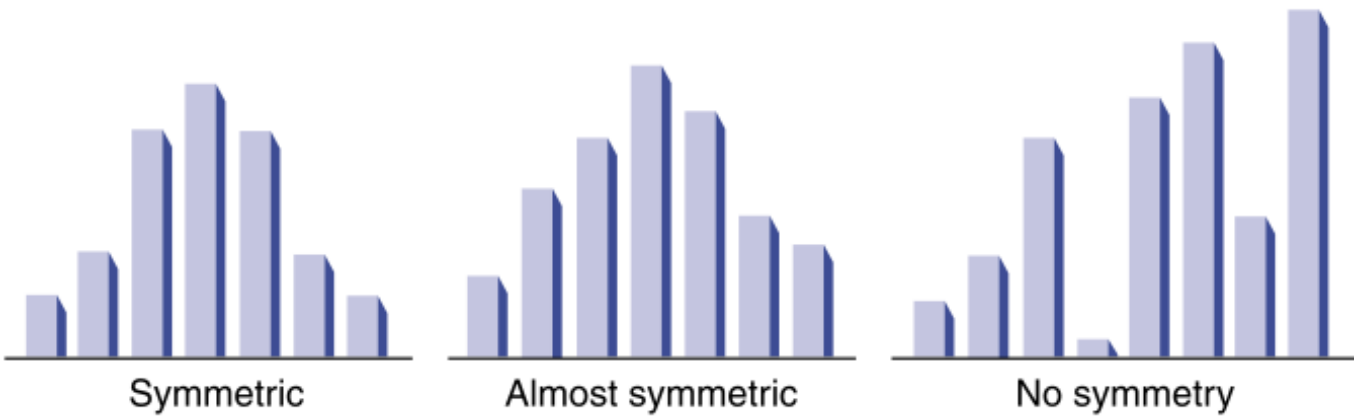


Image from [Ros17] showing types of symmetry.

Python: List Comprehensions

Frequency tables and graphs

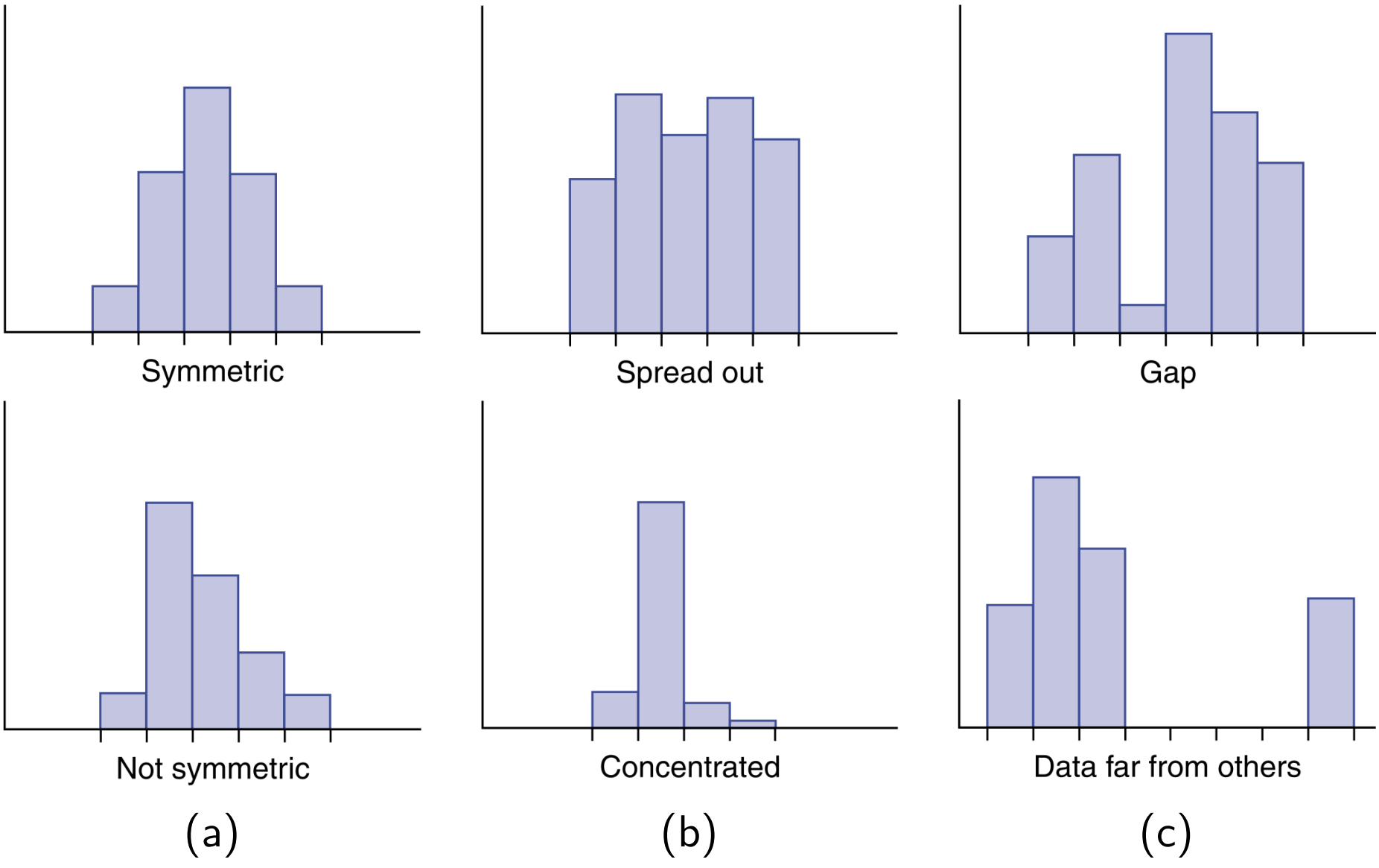
Grouped data and histograms

Other representations

Histograms

- Data grouped by class intervals or bins
- Fewer bins: clearer patterns but hides information
- More bins: retains information but patterns may be lost
- Bins typically same size
- End points must belong to just one group, e.g. use **left-end inclusion convention**
- Relative frequencies also possible

Characteristics detected by histograms



Plots from [Ros17]

Python: List Comprehensions

Frequency tables and graphs

Grouped data and histograms

Other representations

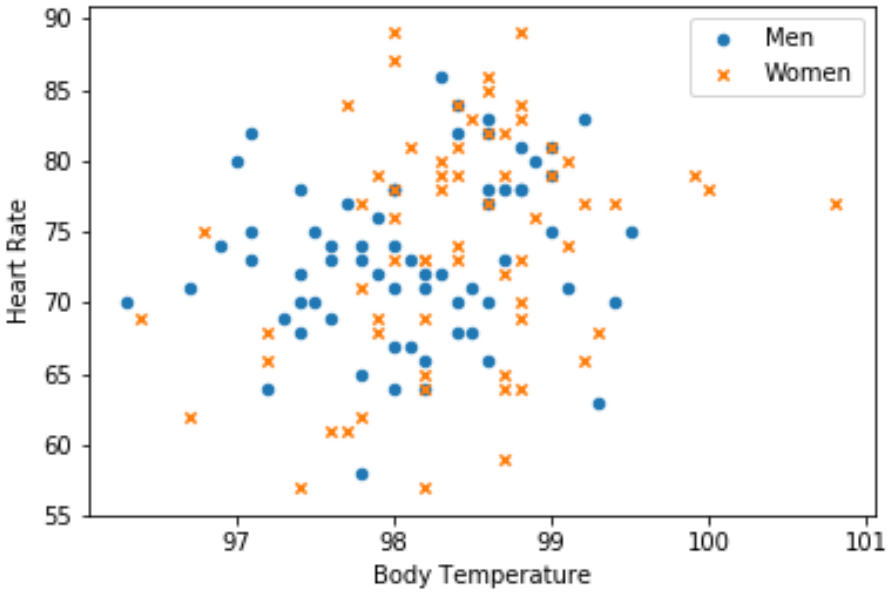
Stem and Leaf Plots

- Simplified view by grouping
- More information than histograms
- Order leaf entries for clarity
- Dual plots for comparison

Girls		Boys
5	14	
7, 5, 5, 5, 4	15	3, 8, 9
8, 4, 2, 1, 0	16	2, 5, 7, 7, 7, 8, 8, 9
9, 8, 7, 6, 6, 4, 2, 1, 1, 0, 0	17	0, 2, 3, 6, 6, 7, 7
	18	0, 1, 4, 5

Stem and leaf plot of children's heights by gender sourced from here.

Sets of paired-data



Plot of body temperature vs heart-rate by gender.

Data from [Sho96]

- When values are paired, two measurements for single observational unit
- Look for relationships between measurements
- Can identify outliers more readily

References I

[MBB12] William Mendenhall, Robert Beaver, and Barbara Beaver, *Introduction to Probability and Statistics*, 14th ed., Duxbury Press, 2012.

[Ros17] Sheldon M. Ross, *Introductory Statistics*, 4 ed., Academic Press, 2017.

[Sho96] Allen Shoemaker, *What's Normal? – Temperature, Gender and Heart Rate*, J. Stat. Educ. **4** (1996), no. 2.