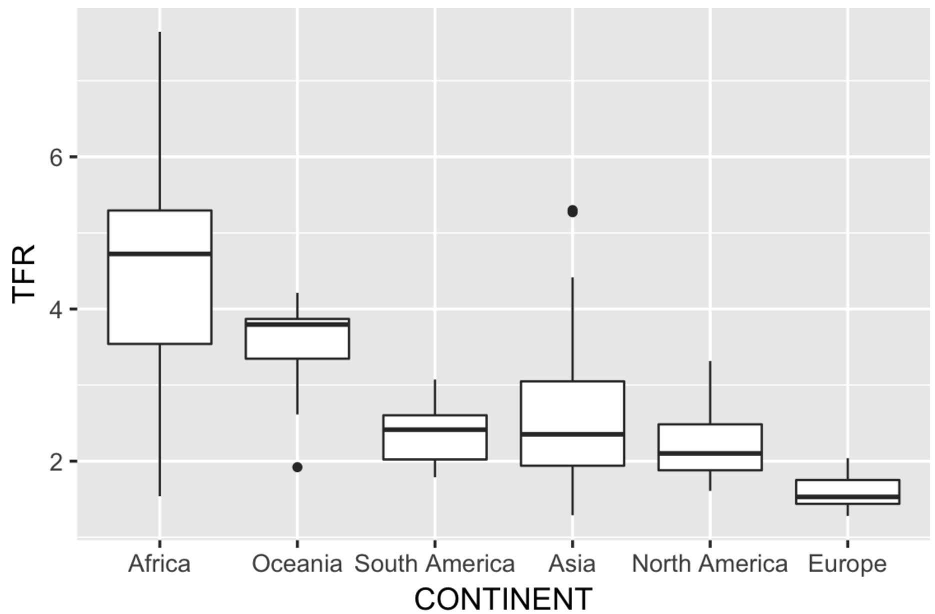
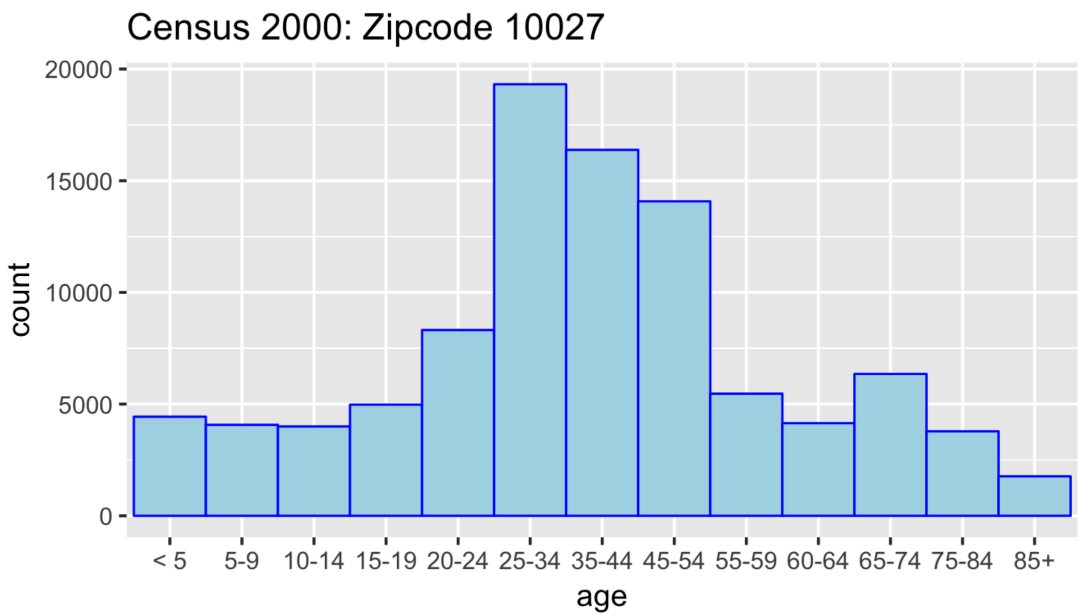
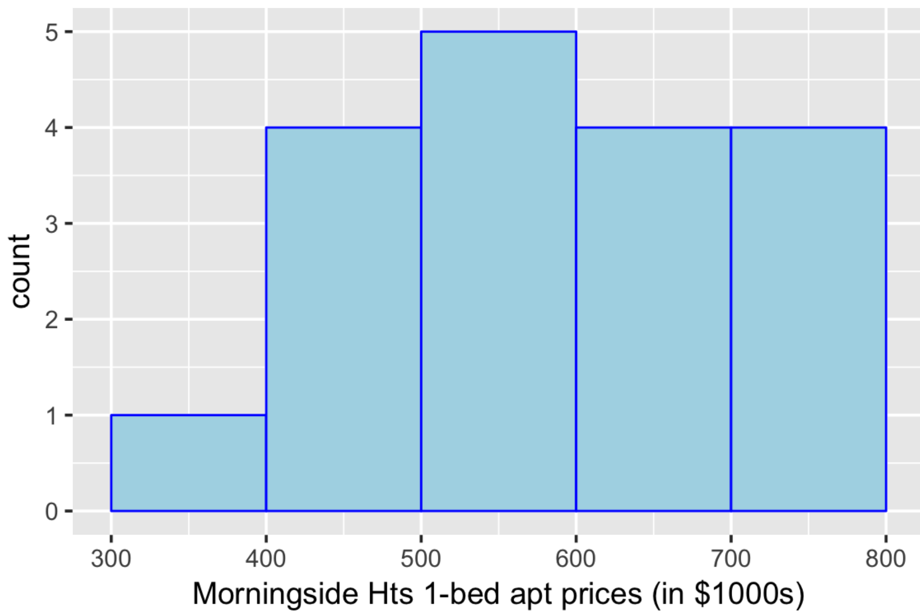
Multiple box plots



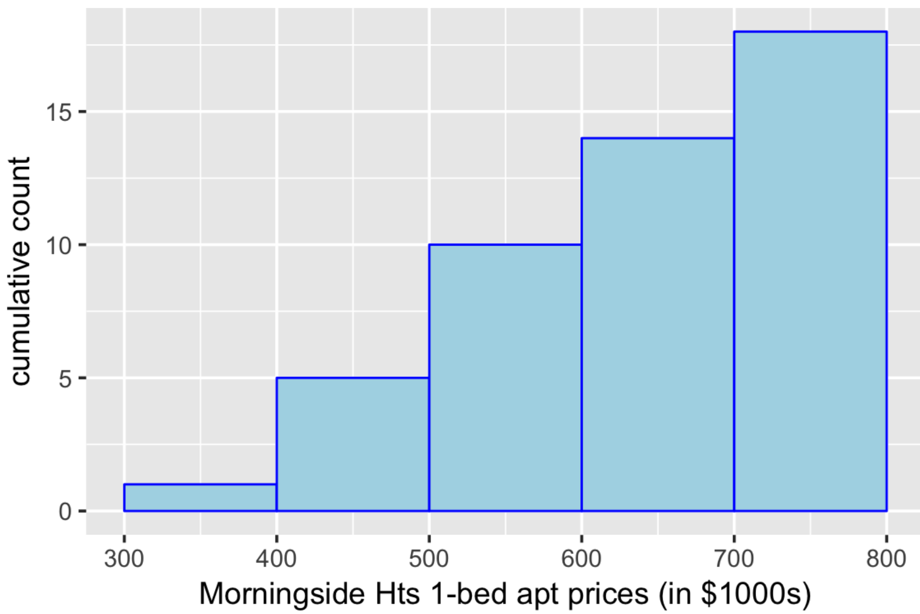
Histogram: what's wrong?



Frequency histogram



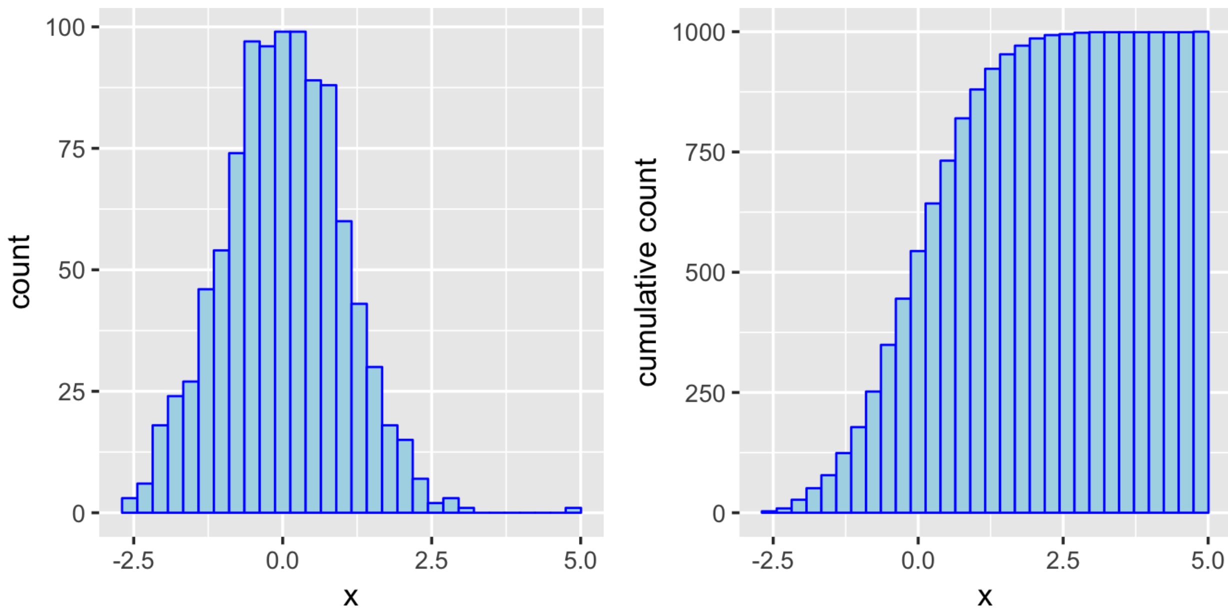
Cumulative frequency histogram



Cumulative frequency histogram

Class	Freq	CumulativeFreq
300-400	1	1
400-500	4	5
500-600	5	10
600-700	4	14
700-800	4	18

Cumulative frequency histogram



EXERCISE

(based on #17, p. 26)
Construction industry data:

bidders	contracts
2	7
3	20
4	26
5	16
6	11
7	9
8	6
9	8
10	3

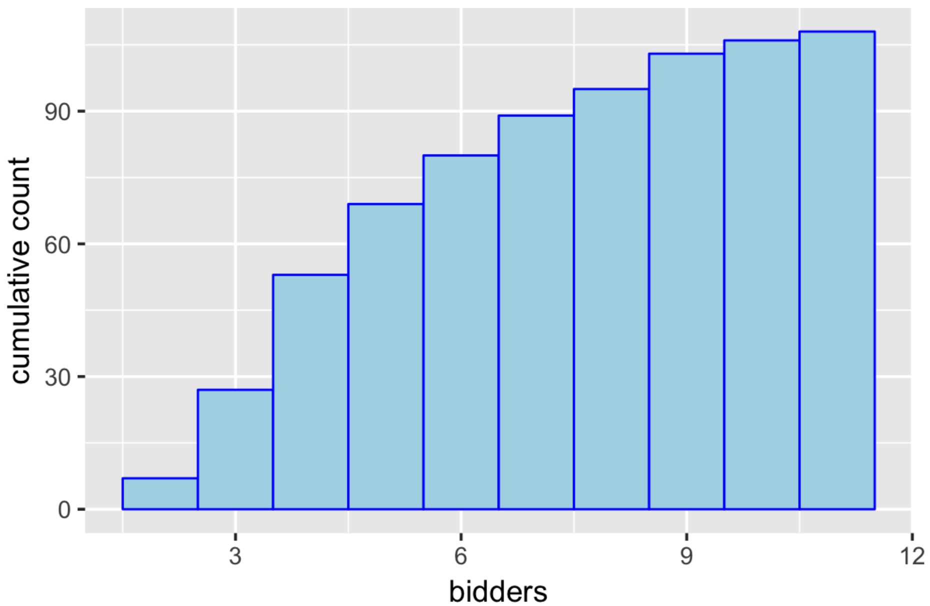
a) What proportion of the contracts involved at most five bidders?

b) What proportion of the contracts involved between five and ten bidders, inclusive?

c) Draw a cumulative frequency histogram.

bidders	contracts
11	2

Cumulative frequency histogram



Sample and population means

population mean: μ = sum of N population values / N

sample mean: $\bar{x} = \frac{x_1+x_2+\ldots+x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$

population median: $\tilde{\mu}$

sample median: $\tilde{x}$

Measures of variability

deviations from the mean

Data: 3, 8, 11, 14
Mean: 9

$value$	$deviation$	$deviation^2$
3	-6	36
8	-1	1
11	2	4
14	5	25

Sum of squared deviations

S_{xx}: 36 + 1 + 4 + 25 = 66

Population variance

σ^2 = 66/4 = 16.5

$$\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 / N$$

Sample variance

Sum of squared deviations:

S_{xx}: 36 + 1 + 4 + 25 = 66

Sample variance:

s^2 = 66 / **3** = 22

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Why n-1?

Short answer: using **n** would result in an underestimation, since the values in the sample are closer to the sample mean than to the true population mean (which we don't know)

Standard deviation

Square root of variance

- Population s.d. = $\sqrt{\sigma^2}$
- Sample s.d. = $\sqrt{s^2}$
- *same units as original values*

EXERCISE (p. 47, #62)

Consider the following information on ultimate tensile strength (lb/in^2) for a sample of n = 4 hard zirconium copper wire specimens:

$\bar{x}$ = 76,831
s = 180
smallest x_i = 76,683
largest x_i = 77,048

Set up equations to determine the values of the two middle sample observations. *Do not solve.*

Chapter 2

Probability

In 1654, writer Antoine Gombaud “Chevalier de Méré” wanted to know if the following bets are profitable:

- getting at least one six on 4 dice rolls
- getting at least one double-six on 24 dice rolls

Vocabulary (2.1)

- **experiment** – process whose outcome is subject to uncertainty
  (ex. rolling a die)

- **sample space** – set of all possible outcomes of an experiment
  S = {1, 2, 3, 4, 5, 6}

- **event** – *collection* of outcomes contained in the sample space

Experiment with an infinite sample space

- ex. flip a coin until you get tails
- **sample space**

  S = {T, HT, HHT, HHHT, …}
- **event**

  you get tails in fewer than 8 flips

  A = {T, HT, HHT, HHHT, HHHHT, HHHHHT, HHHHHHT}

Set theory 1

- **complement** of an event – all outcomes in the sample space that are *not* in the event

  S = {1, 2, 3, 4, 5, 6}

  A = {1, 3}

  A' (“not A”) = {2, 4, 5, 6}

Set theory 2

- **union** of two events: all outcomes in *either* event or in *both* $A \cup B$ (“A *or* B”)
  $A = \{1, 3\}$
  $B = \{3, 5\}$
  $A \cup B = \{1, 3, 5\}$

Set theory 3

- **intersection** of two events: all outcomes in *both* events
  $A \cap B$ ("A *and* B")
  $A = \{1, 3\}$
  $B = \{3, 5\}$
  $A \cap B = \{3\}$

Set theory 4

- **null event**: no outcomes ∅ or {}
  C = {1, 2}
  D = {3, 4}
  $C \cap D = \varnothing$
- **mutually exclusive** – events that cannot occur at the same time
  if $A \cap B = \varnothing$, then A and B are *mutually exclusive* or *disjoint*

Set theory: more than two events 1

- $A \cup B \cup C$: all outcomes in at least one of A, B, & C

$A = \{1, 2, 3\}$
$B = \{5\}$
$C = \{1, 5, 10\}$

$A \cup B \cup C = \{1, 2, 3, 5, 10\}$

- $A \cap B \cap C$: all outcomes in A, B, *and* C

  $A \cap B \cap C = \{\}$