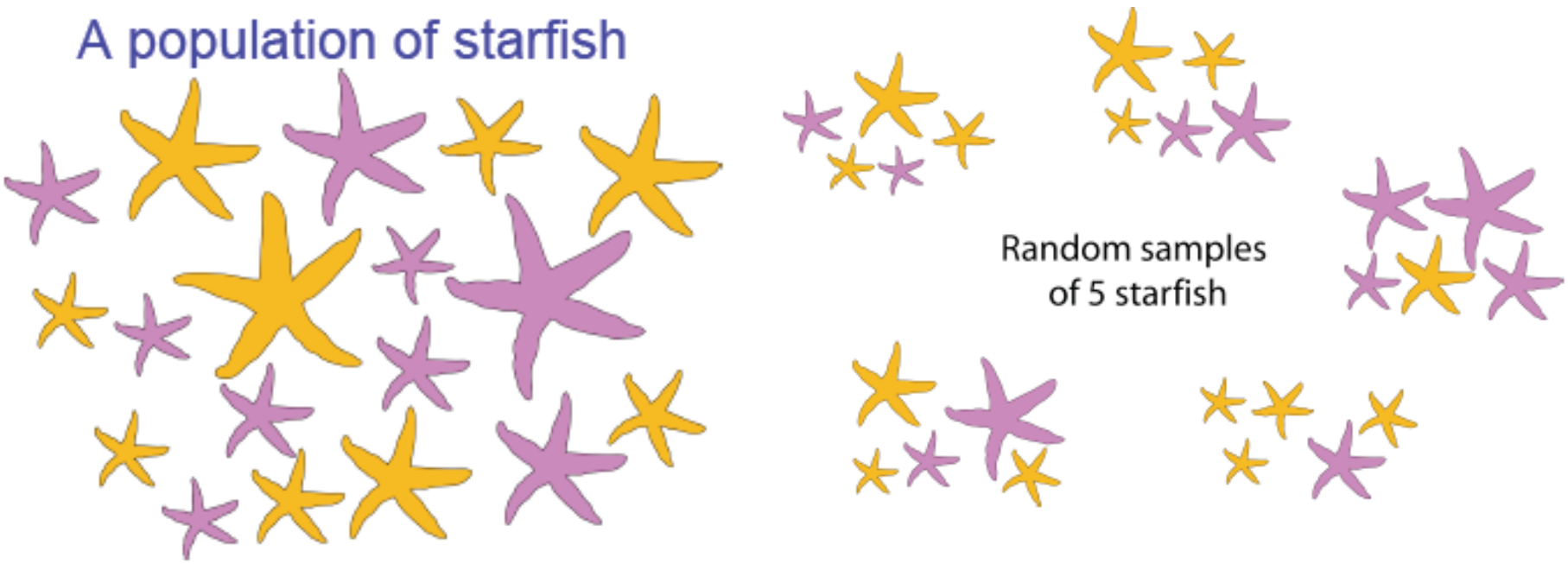
Graphing Data

Sampling Populations



Graphing Data

Random Variables:

- characteristics measured on individuals drawn from the population under study
- Value is not constant; it is subject to **variation**
- Examples include measurements such as height, weight, age, ethnicity, education
- **Categorical (Nominal, ordinal)** or **Numerical (Discrete, continuous)**

Types of data:

Categorical Variable

- AKA Class variables or Nominal variables
- They do not have magnitude on a numerical scale
- **Nominal**
  - Lack inherent order
- **Ordinal**
  - Inherent order
- Ex: blood type, genotype, sex, state, survival (live or die), drug treatment (aspirin vs ibuprofen)

Numerical Variables

- AKA Numerical variables
- Random Variable is a Quantitative variable
- **Continuous**
  - Ability to take any value ex.. Human weight, **age**
- **Discrete**
  - Spaces between possible values ex. Number of offspring, **age**

Graphing Data

Frequency Distributions and Probability Distributions

A ***frequency distribution*** is the number of times each value of a variable occurs in a sample

A ***probability distribution*** is the distribution of the variable in the entire population ie. the probability that the **variable** takes on particular values

Graphing Data

1. **Display frequency distribution**
   - **Bar graphs**
   - **Histograms**

2. **Display associations (or differences) between two variables**
   - **Grouped bar plot**
   - **Mosaic**
   - **Box plot**
   - **Scatter plot**
   - **Strip charts**

Graphing Categorical Variables

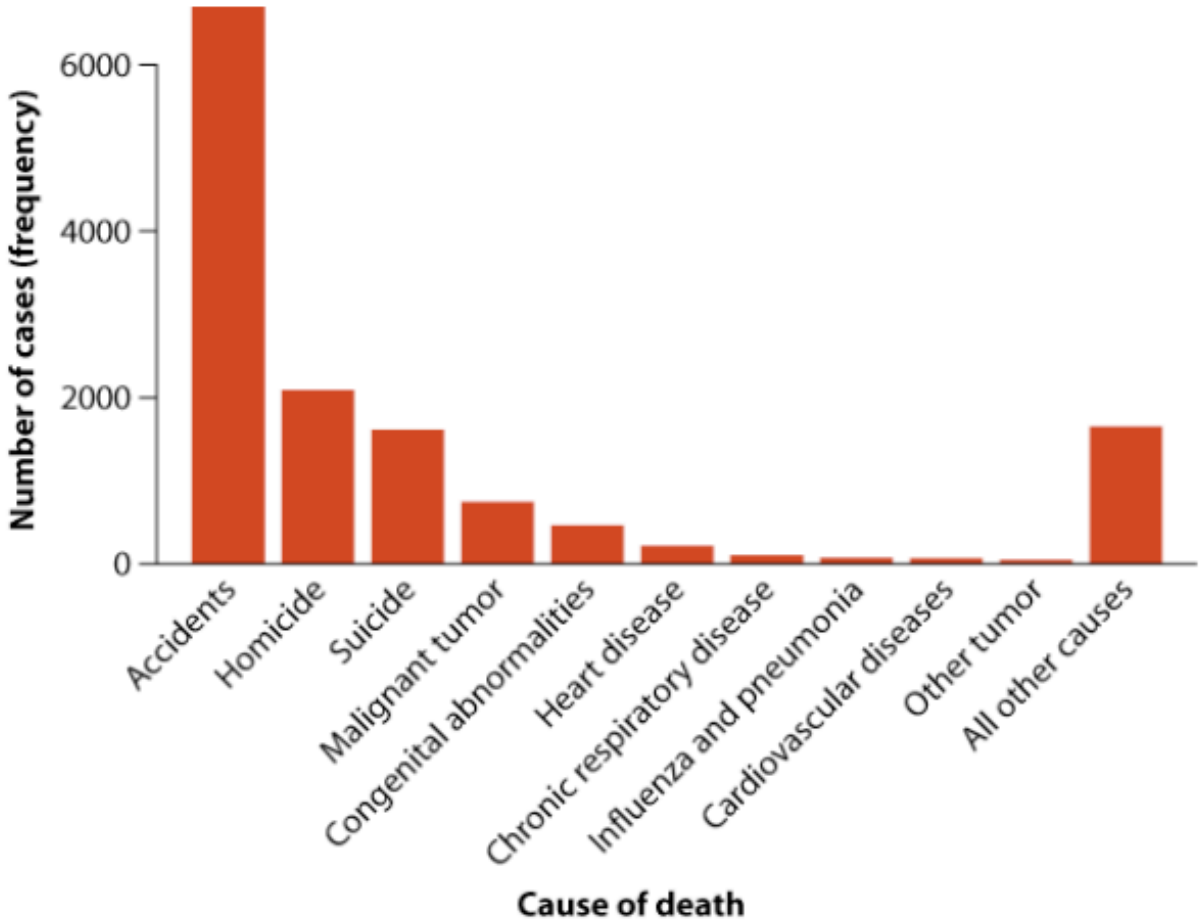
Graphing Categorical Variables:

- Frequency tables:

Cause of death	Frequency
Accidents	6,688
Homicide	2,093
Suicide	1,615
Malignant tumor	745
Heart disease	463
Congenital abnormalities	222
Chronic respiratory disease	107
Influenza and pneumonia	73
Cerebrovascular diseases	67
Other tumor	52
All other causes	1,653

Graphing Categorical Variables:

- Bar graph:



Graphing Numerical Variables

Graphing Data

Graphing Numerical Variables:

Heights of (theoretical) biostats students (cm):

165	168	163	173	170	163	170	155
152	190	170	168	142	160	154	165
156	177	173	165	165	175	155	166
168	165	180	165				

Graphing Numerical Data:

Frequency Table of heights:

Height Group (cm)	**Frequency**
141 - 150	1
151 - 160	6
161 - 170	15
171 - 180	5
181 - 190	1

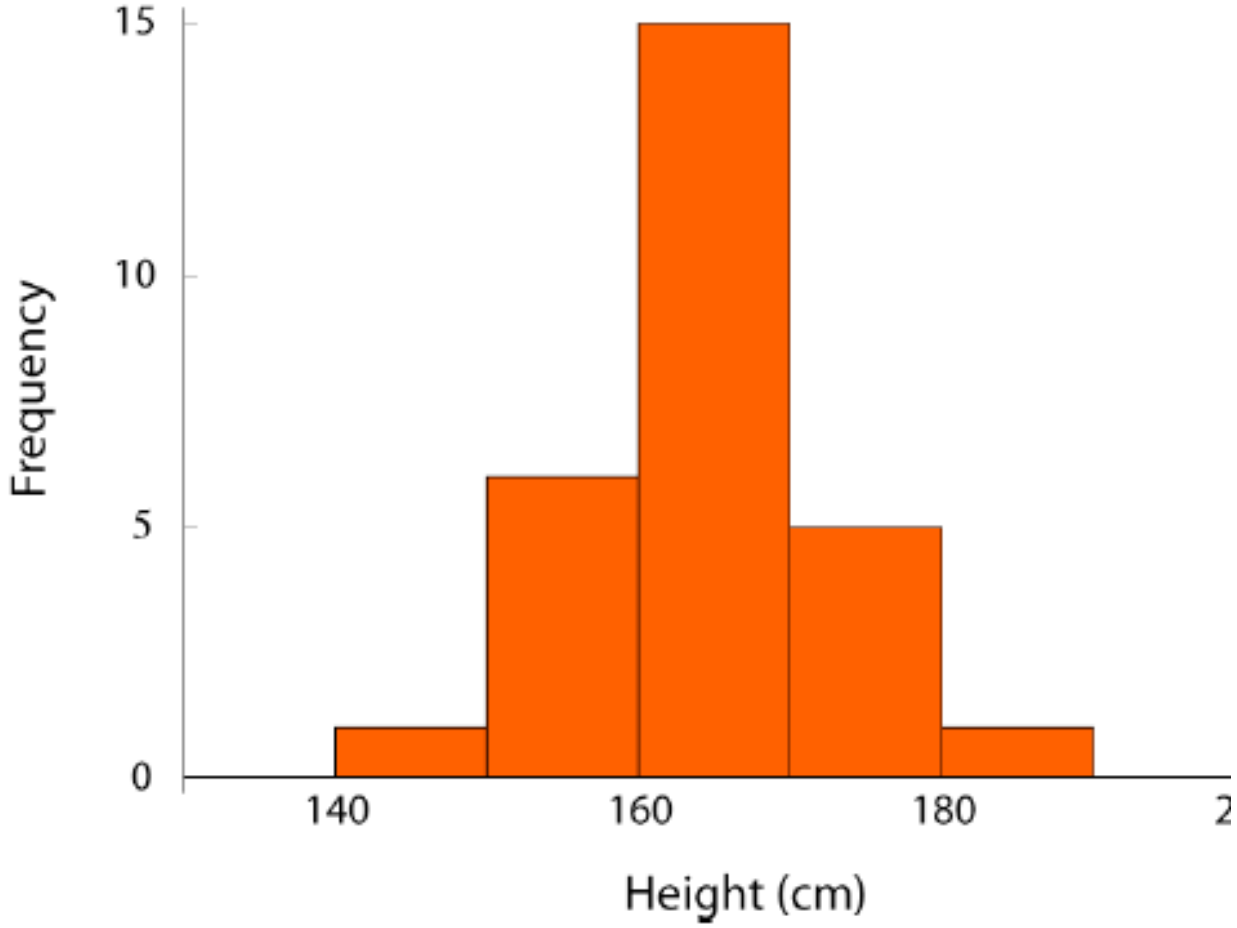
Graphing Data

Graphing Numerical Data:

Histogram: A bar graph of a frequency distribution (or relative frequency distribution) in which the heights of the bars are proportional to the class frequencies and the widths are proportional to the variable groupings

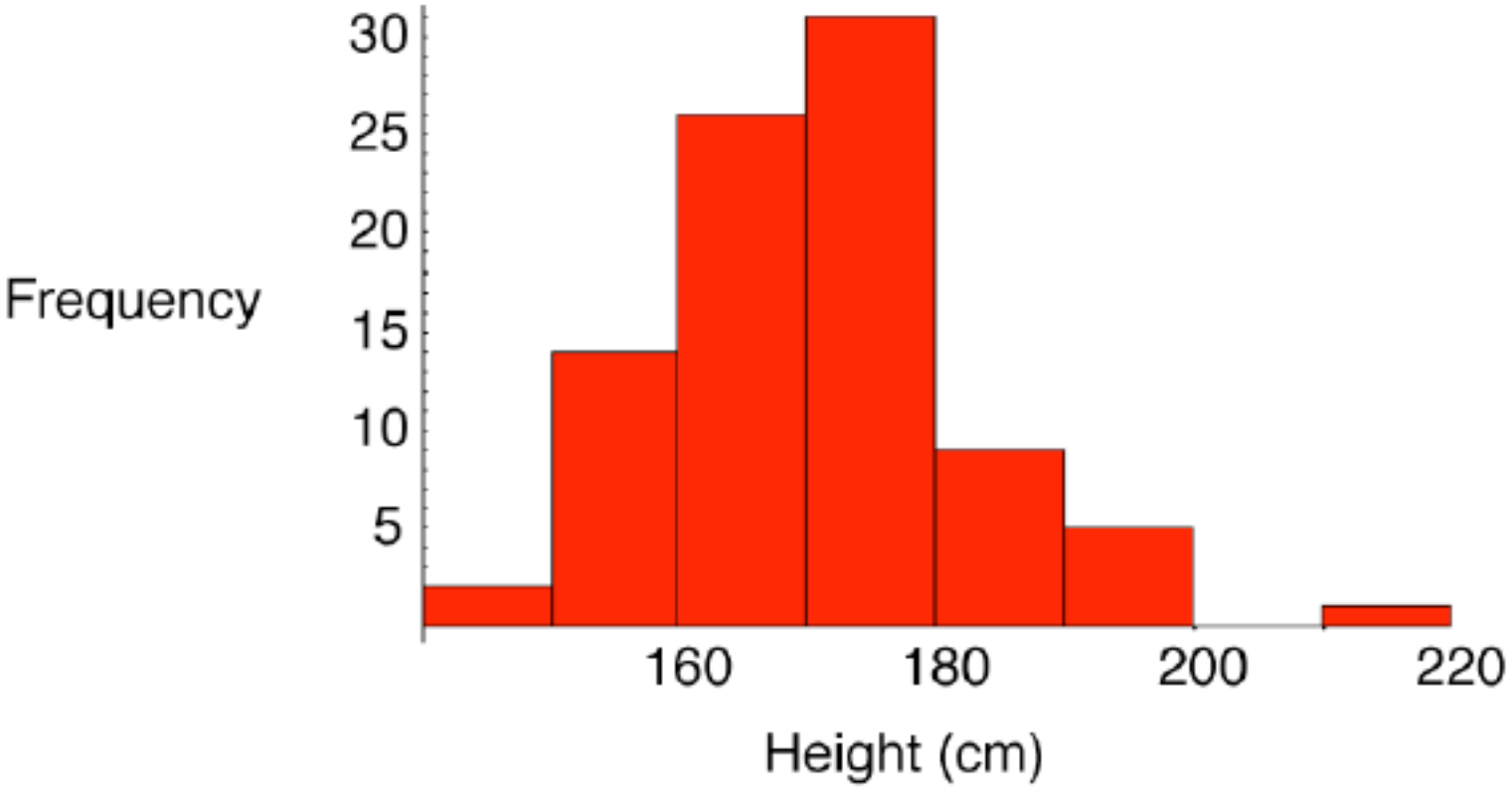
Graphing Data

Histogram of Height data:



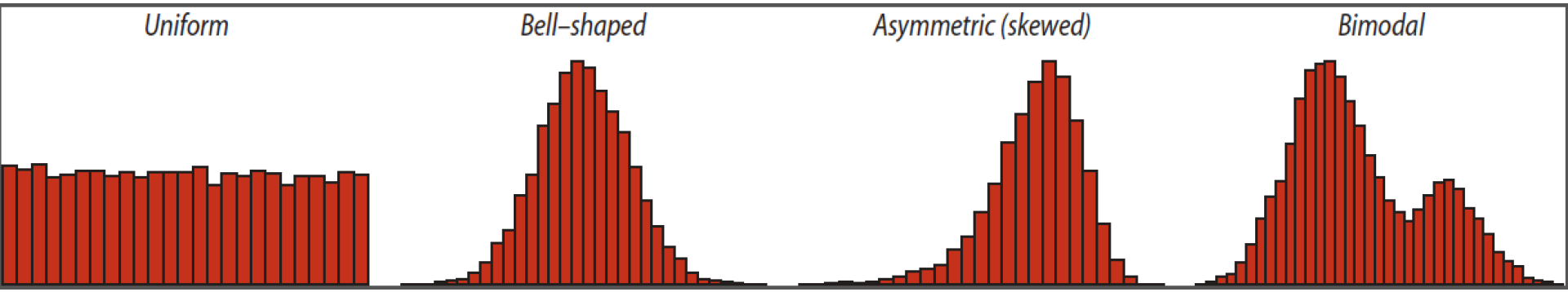
Histogram of Height data:

Now with more data!

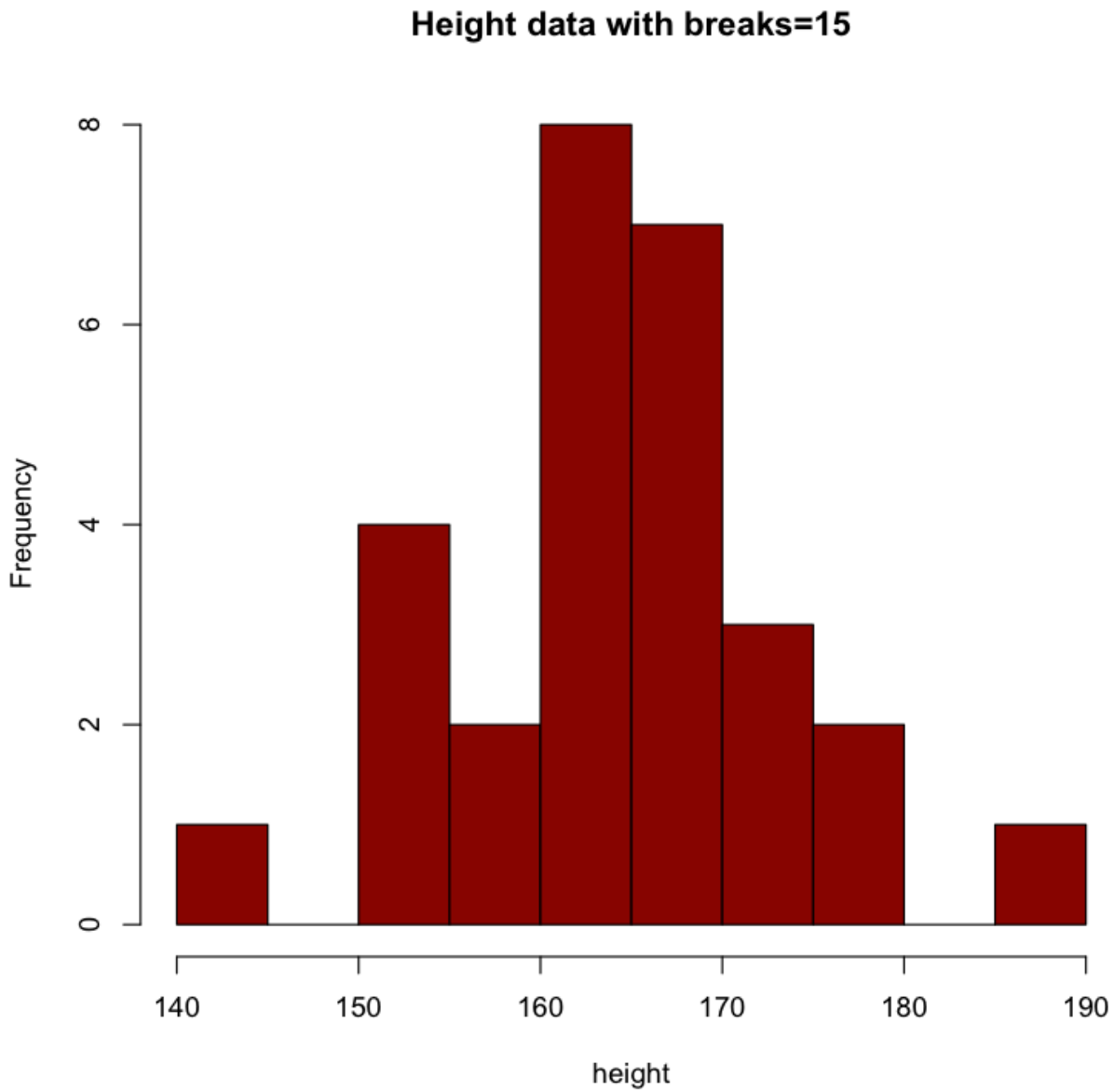


Graphing Data

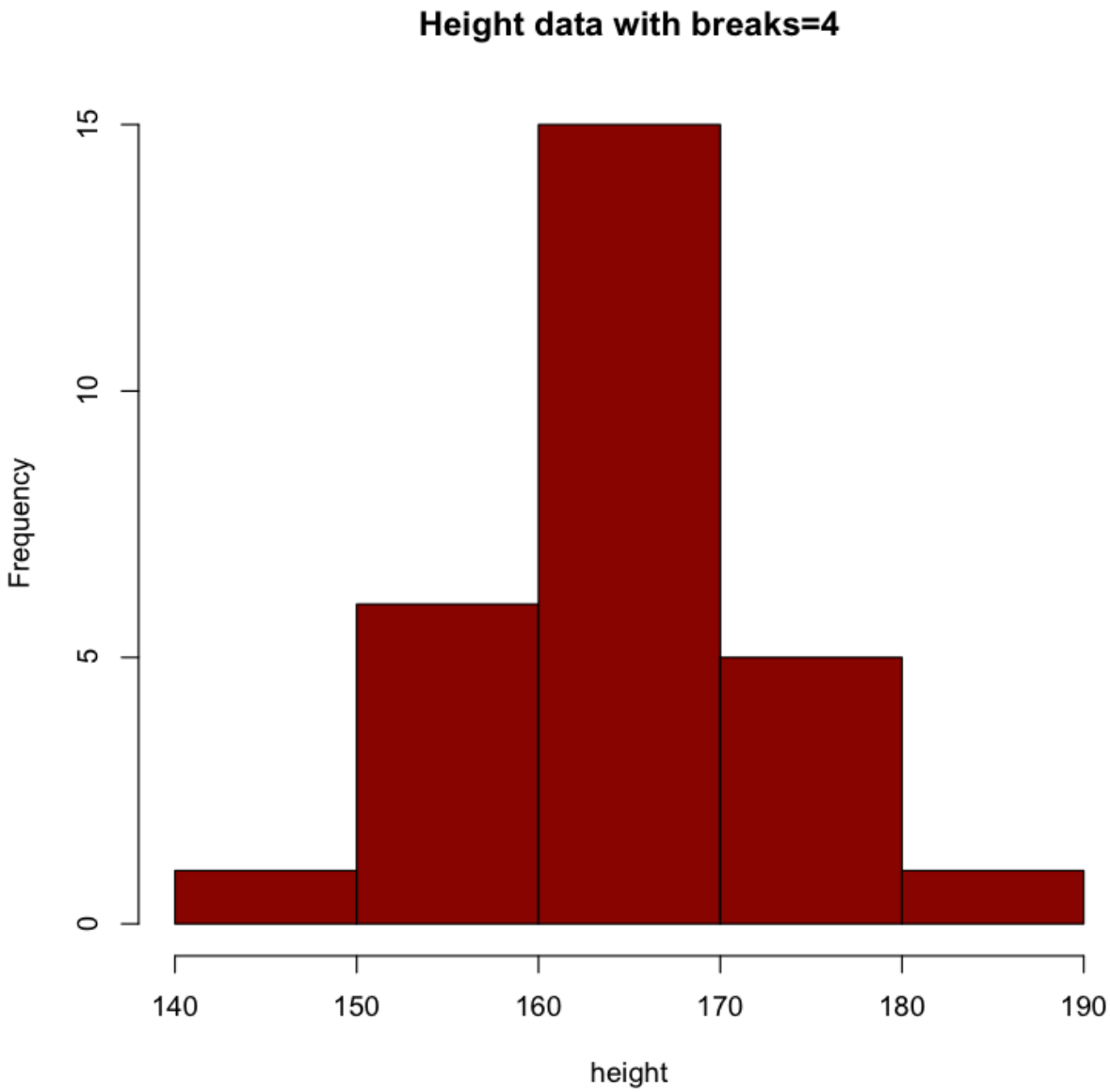
Shapes of frequency distributions:



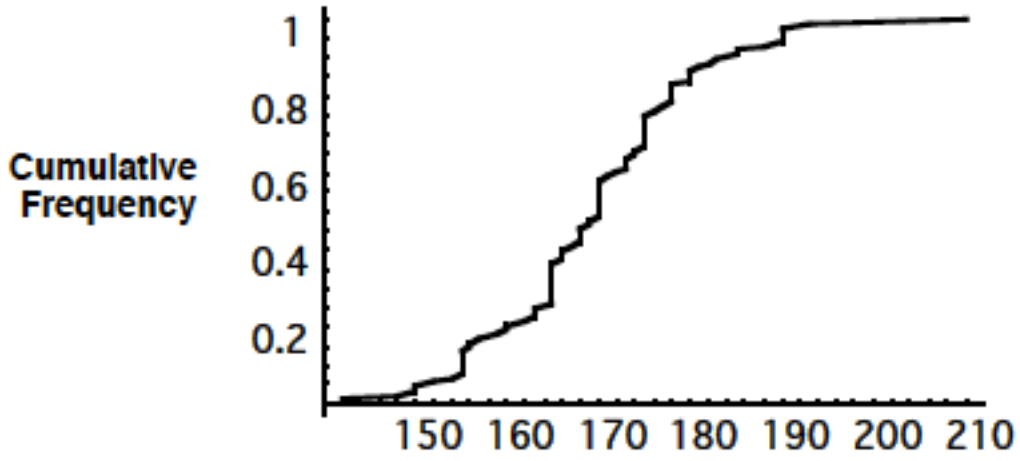
Histogram of Height data:

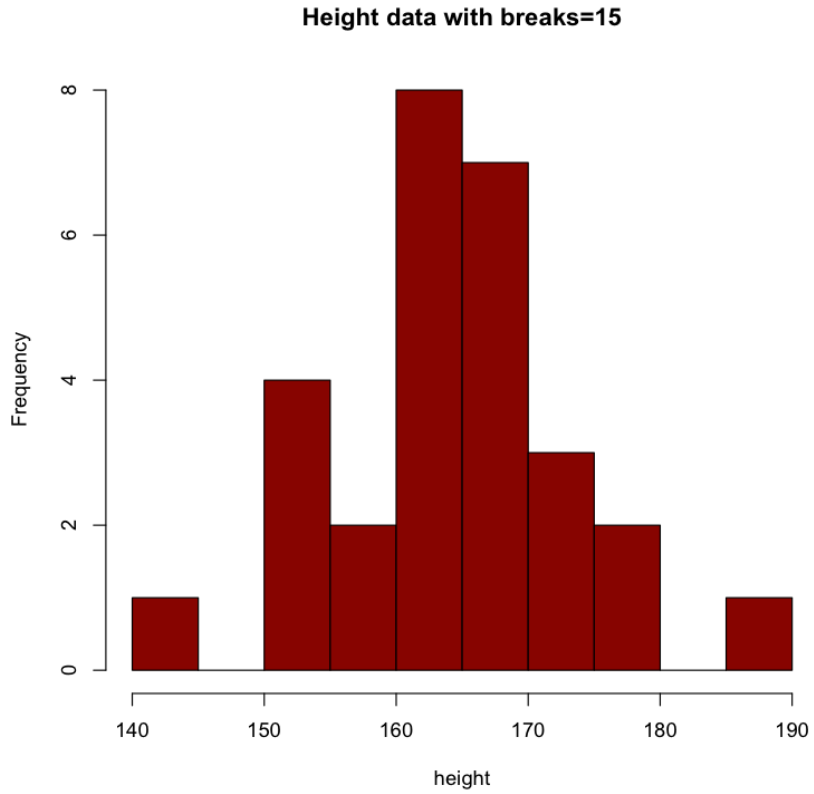


Histogram of Height data:

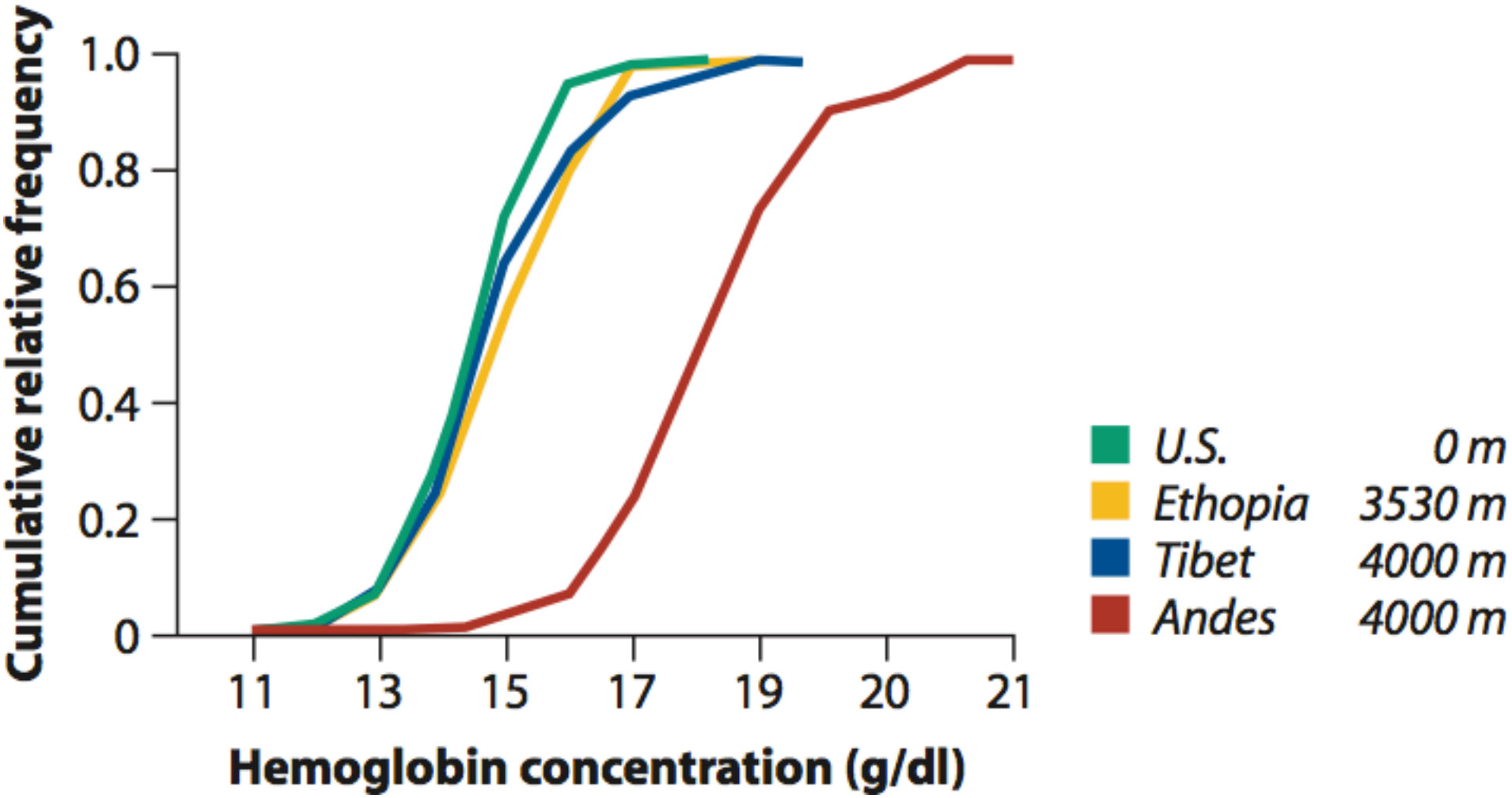


Graphing Data

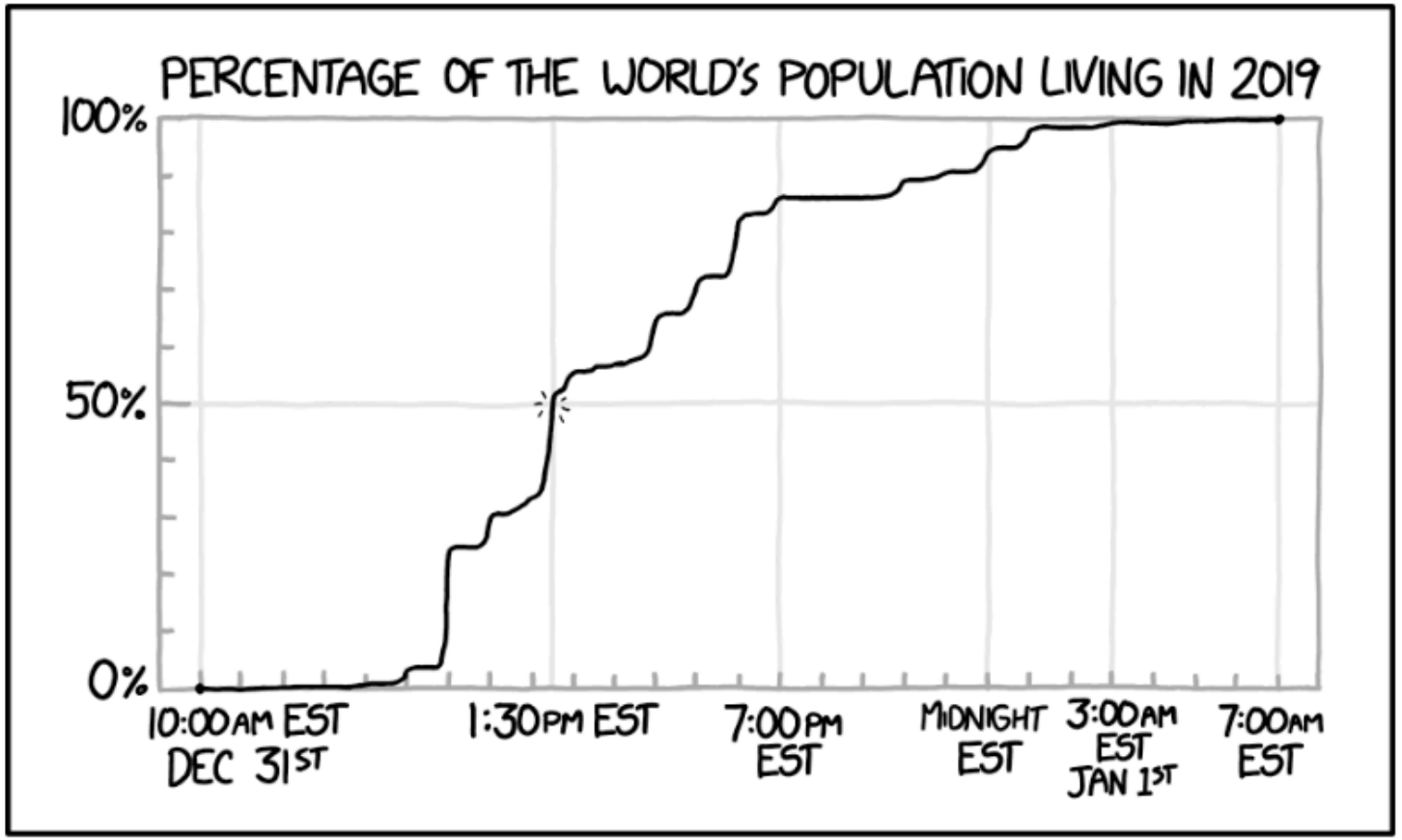




Multiple Cumulative Frequency Distribution:



https://xkcd.com/2092/



CONSENSUS NEW YEAR: AS OF 1:30PM EASTERN TIME (6:30PM UTC) A MAJORITY OF THE WORLD'S POPULATION WILL BE LIVING IN 2019.

Associations between two categorical variables

Associations between two categorical variables:

- Contingency table
- Grouped Bar Graph
- Mosaic Plot

Graphing Data

Associations between two categorical variables:

Contingency Table:

Association between reproductive effort and avian malaria

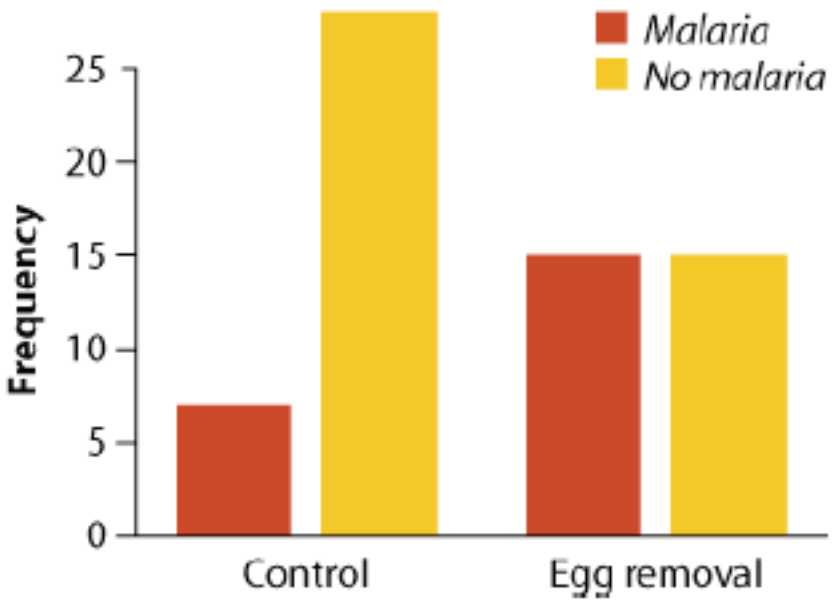
Table 2.3A. Contingency table showing incidence of malaria in female great tits subjected to experimental egg removal.

	control group	egg removal group	row total
malaria	7	15	22
no malaria	28	15	43
column total	35	30	65

Graphing Data

Associations between two categorical variables:

Grouped Bar Graph:



Graphing Data

Associations between two categorical variables:

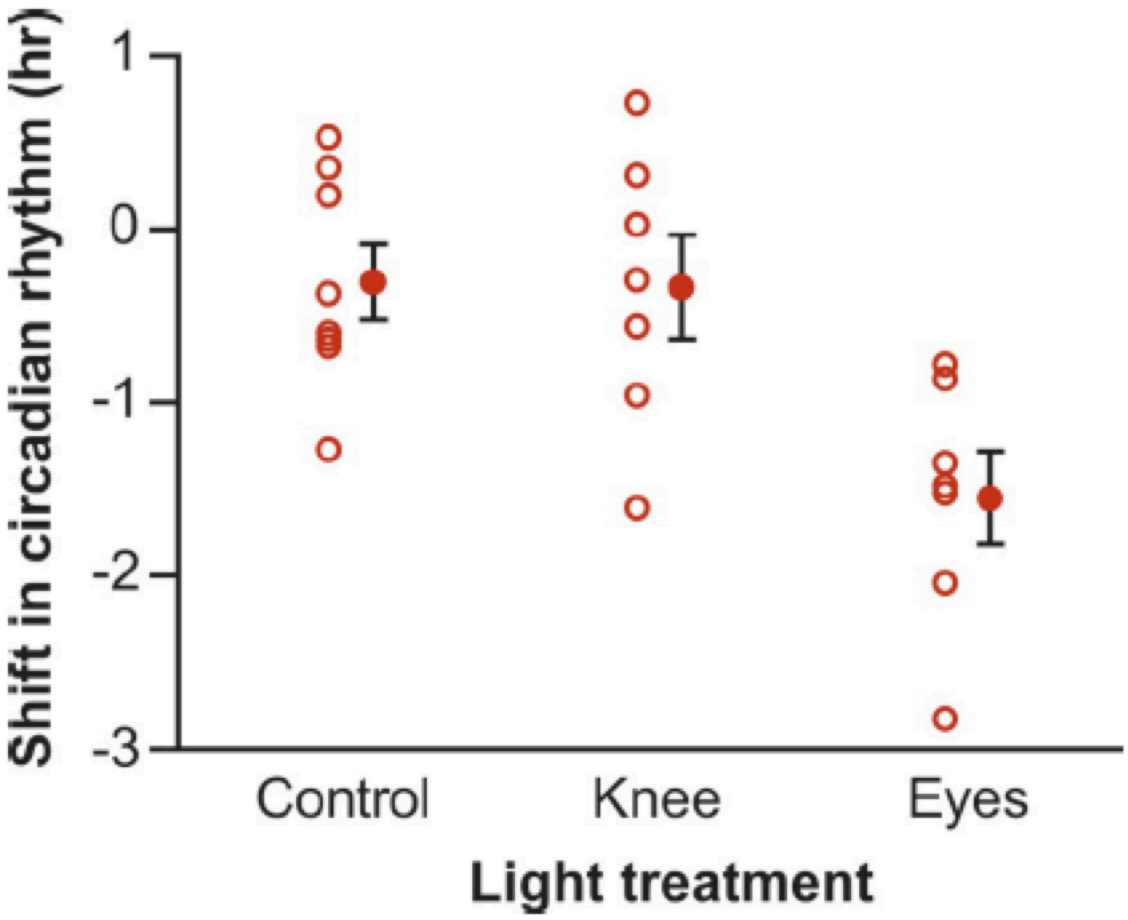
Mosaic Plot:

Associations between two categorical **and** numerical variables

Graphing Data

Associations between categorical **and** numerical variables:

Strip Charts: each observation is represented as a dot

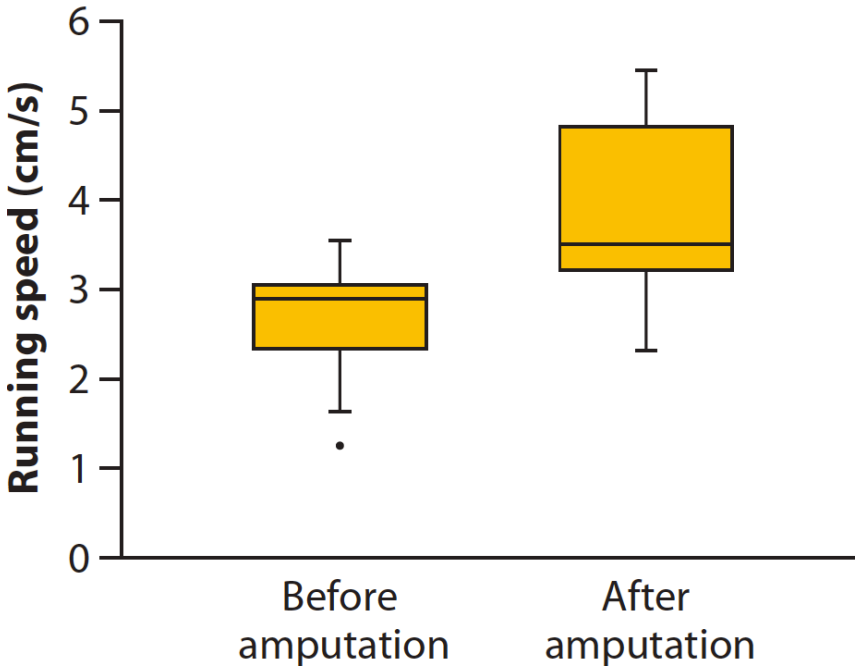


Graphing Data

Associations between categorical **and** numerical variables:

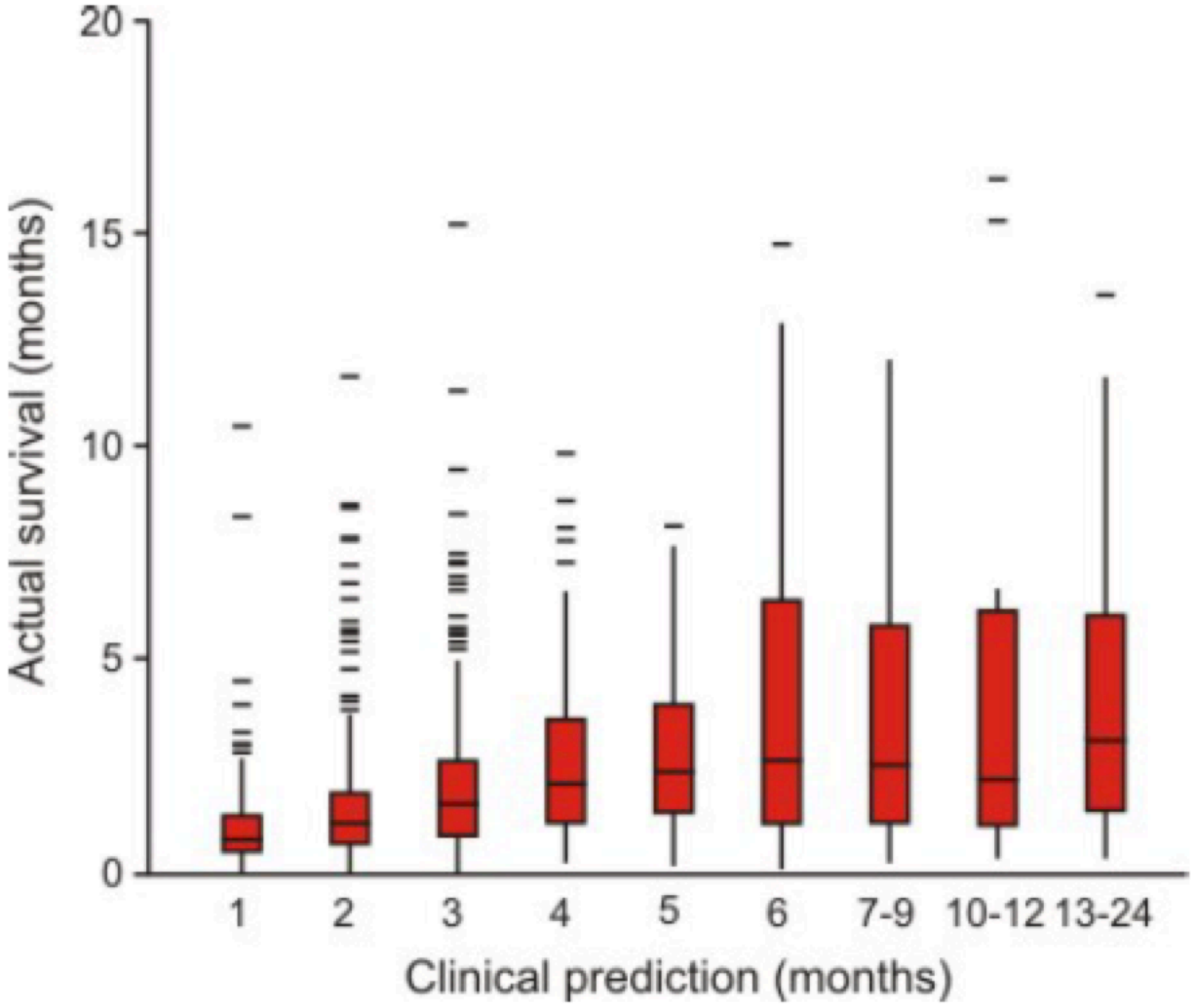
Strip Charts: each observation is represented as a dot

Box plot: graph that uses lines and a rectangular box to display the median, quartiles, range and outliers of the data (we'll see more of this is chapter 3)



Graphing Data

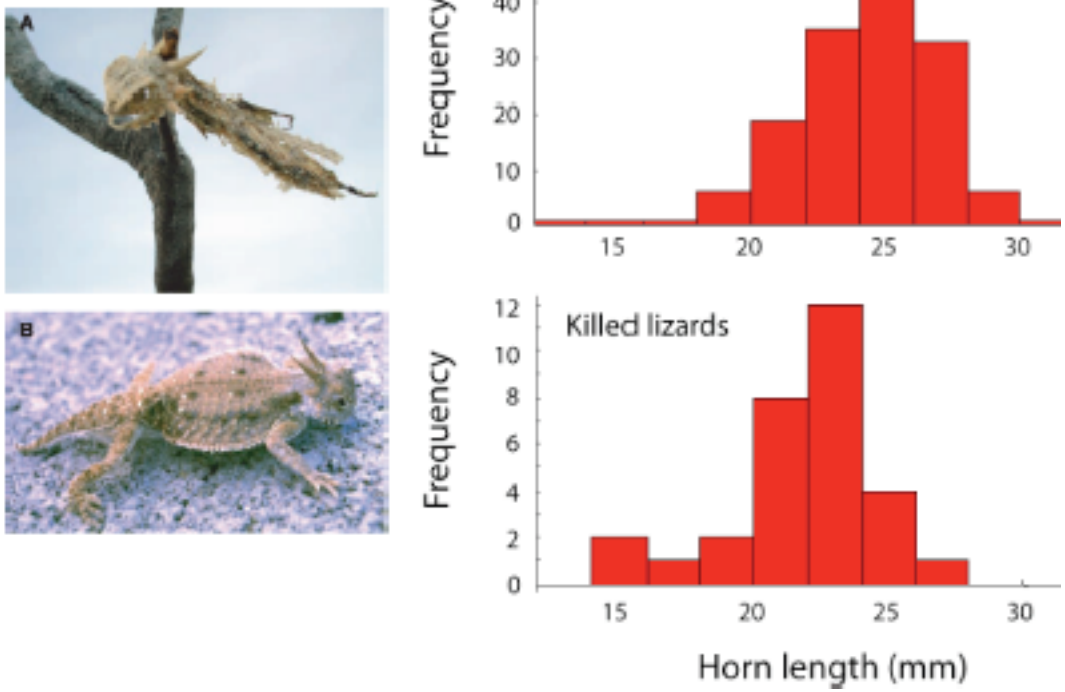
Associations between categorical **and** numerical variables:



Graphing Data

Associations between categorical **and** numerical variables:

Multiple histograms:



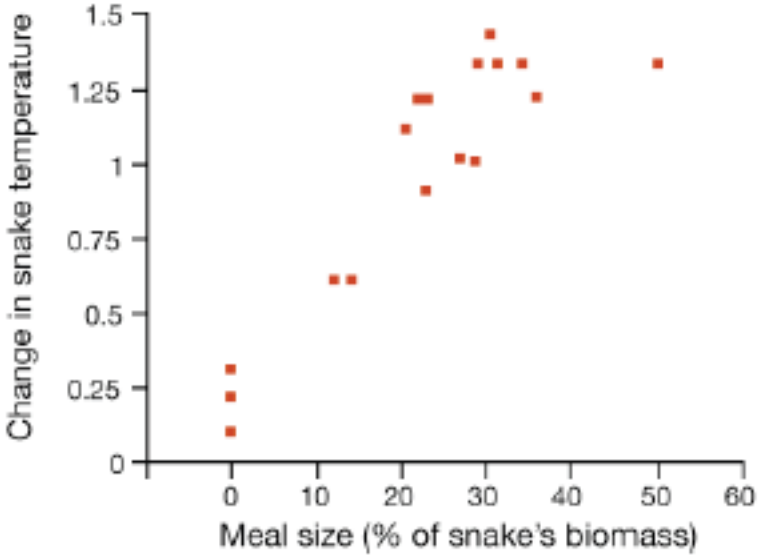
Young, K.V., E.D. Brodie Jr. and E.D.Brodie III. 2004. How the horned lizard got its horns. Science 304:65.

Associations between two numerical variables

Graphing Data

Associations between two numerical variables:

Scatter Plots:



Tattersall et al. (2004) *Journal of Experimental Biology*. 207: 579-585

Question 6: The following data are the occurrences in 2002 of the different taxa in the list of endangered and threatened species under the U.S. Endangered Species Act. The taxa are listed in no particular order:

Taxon	Number of species
Birds	92
Clams	70
Reptiles	36
Fish	115
Crustaceans	21
Mammals	74
Snails	32
Plants	745
Amphibians	22
Insects	44
Arachnids	12

- How would you re-write this table to make it more informative?
- Choosing the most appropriate graphical method, display the number of species in each taxon. What kind of graph did you choose and why?
- What should the baseline for the number of species be in your graph and why?

Graphing Data

Summary: Graphical Methods for frequency distributions

Types of Data	**Method**
Categorical	Bar Graph
Numerical	• Histogram • Cumulative freq distribution

Graphing Data

Summary: Associations between variables

	Explanatory Variable: Categorical	Explanatory Variable: Numerical
Response Variable		
Categorical	• Contingency table • Grouped Bar Graph • Mosaic Plot	
Numerical	• Multiple histograms • Cumulative freq distr.	• Scatter Plot • Line Graph