

The history of statistics has its roots in biology

Sir Francis Galton (1822-1911)



Inventor of fingerprints,
study of heredity of quantitative traits

Regression & correlation

Karl Pearson (1857-1936)



Polymath

Studied genetics

Correlation coefficient
 χ^2 test
Standard deviation

Sir Ronald Fisher (1890-1962)



*The Genetical Theory of
Natural Selection*
Founder of population genetics

Analysis of variance
likelihood
P-value
randomized experiments
multiple regression
etc., etc., etc.

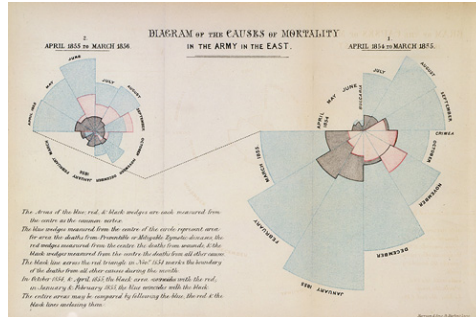
Florence Nightingale

(1820-1910)



Most famous as a nurse, especially tending for wounded soldiers

Made many contributions to statistical graphics



Chapter 2: Displaying Data

Types of data

- Categorical Variables (also known as Class variables, or Nominal variables)
- Numerical Variables (or Quantitative Variables)
 - Numerical variables are either *continuous* or *discrete*.

Categorical variables

- Sex
- Genotype
- Drug treatment (e.g. aspirin vs. ibuprofen)
- Province
- Survival (i.e., live or die)

Numerical variables

- Height
- Weight
- Tail length
- Dose (e.g., in micrograms/gram)
- Longevity (i.e., number of years)

Discrete vs. Continuous

Can be counted

- Number of limbs
- Number of offspring
- Number of petals

Can be measured

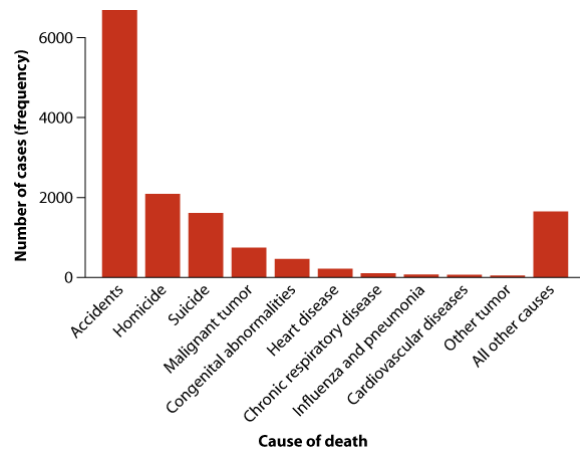
- Arm length
- Height
- Weight
- Age

Frequency table showing the ten most common causes of death in Americans between 15 and 19 years of age in 1999. The total number of deaths is $n = 13,778$.

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



Tarsus lengths of 34 Pacific Wrens (mm)

19.2	18.3	17.2	18.6
17.9	18.4	17.6	17.7
19.4	18.1	16.7	17.5
18.0	17.9	18.7	18.9
18.6	18.3	18.9	18.1
17.5	18.1	18.1	18.8
17.5	18.3	18.4	18.2
18.2	17.7	18.1	
19.0	17.6	17.5	

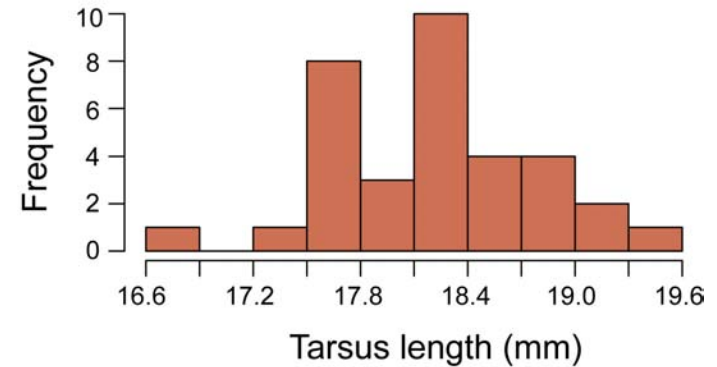
Frequency table:

16.7 18.1
 17.2 18.2
 17.5 18.2
 17.5 18.3
 17.5 18.3
 17.5 18.3
 17.6 18.4
 17.6 18.4
 17.7 18.6
 17.7 18.6
 17.9 18.7
 17.9 18.8
 18.0 18.9
 18.1 18.9
 18.1 19
 18.1 19.2
 18.1 19.4

Tarsus Group	Frequency
16.6-16.9	1
16.9-17.2	0
17.2-17.5	1
17.5-17.8	8
17.8-18.1	3
18.1-18.4	10
18.4-18.7	4
18.7-19.0	4
19.0-19.3	2
19.3-19.6	1

Histogram

Tarsus lengths of Pacific Wrens
 (n = 34)

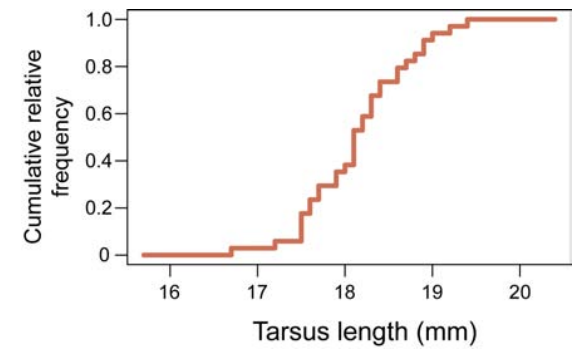


Cumulative Frequency Distribution

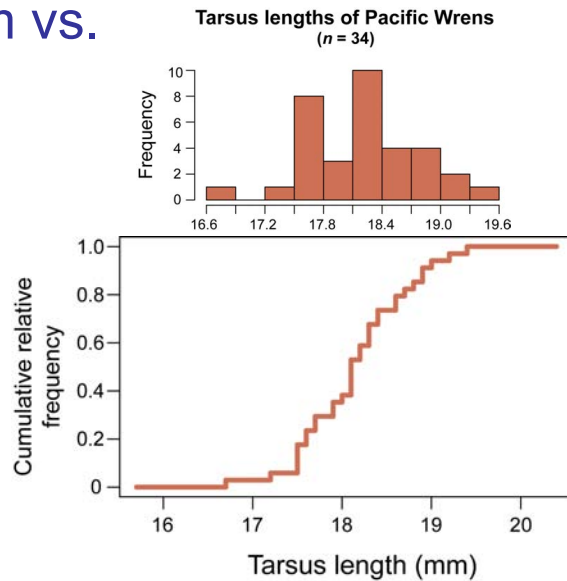
- The cumulative relative frequency of a value is the proportion of individuals equal to or less than that value.
- Example: In the Pacific Wrens, 13 out of 34 individuals have tarsus length less than or equal to 18.0.
 - Hence the cumulative frequency of 18.0 is $13/34 = 0.38$

Making a Cumulative Frequency Distribution

16.7 18.1
 17.2 18.2
 17.5 18.2
 17.5 18.3
 17.5 18.3
 17.5 18.3
 17.6 18.4
 17.6 18.4
 17.7 18.6
 17.7 18.6
 17.9 18.7
 17.9 18.8
 18.0 18.9
 18.1 18.9
 18.1 19
 18.1 19.2
 18.1 19.4



Histogram vs. CFD



Associations between two categorical variables

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.

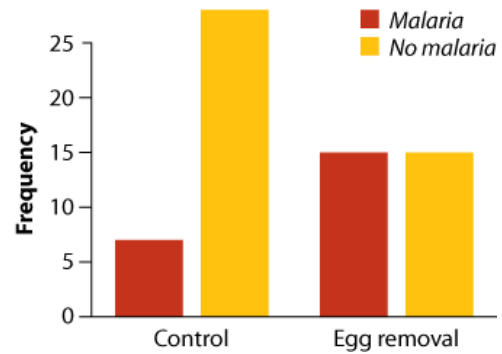


	control group	egg removal group	row total
malaria	7	15	22
no malaria	28	15	43
column total	35	30	65

Mosaic plot

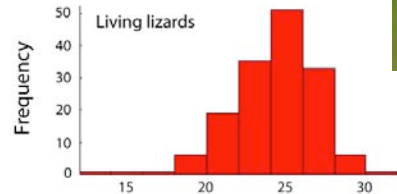


Grouped Bar Graph

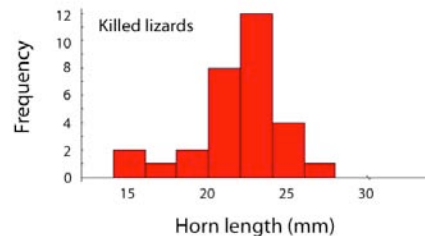


Associations between categorical and numerical variables

Multiple histograms



Shrike

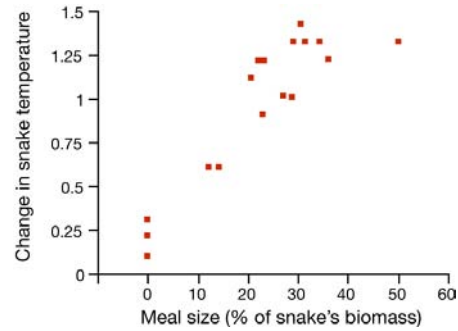


Associations between two numerical variables

Scatter plot

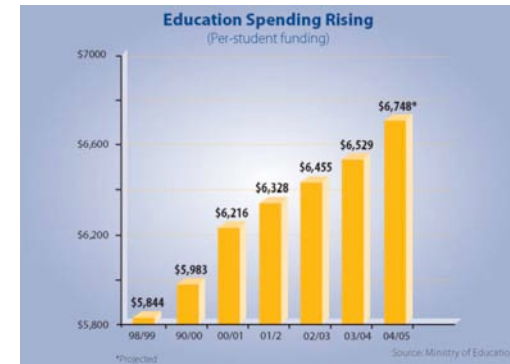


Tropical Rattlesnake (*Venomous*)

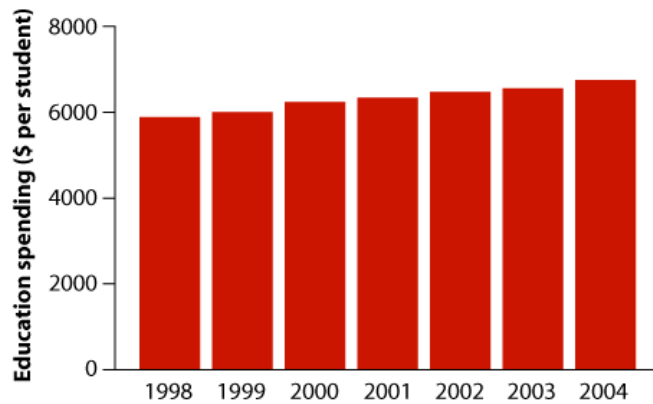


Tattersall et al. (2004) *Journal of Experimental Biology* 207:579-585

Don't mislead with graphics



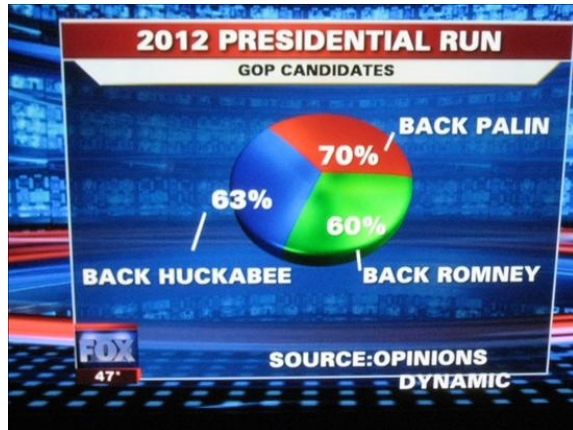
Better representation of truth



More examples of misleading graphics

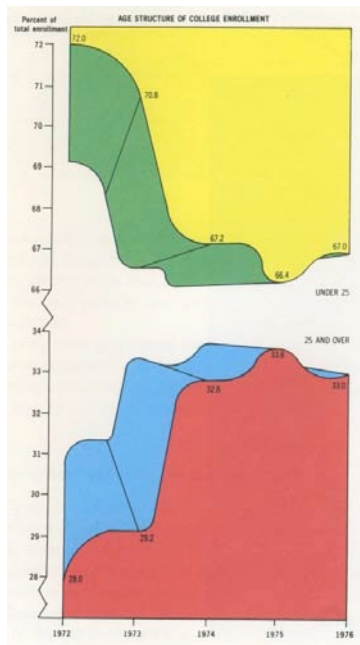
http://www.youtube.com/watch?v=-rbyhj8uTT8&feature=player_embedded

More examples of misleading graphics



Modified from Washington Post, October 25, 1978

Los Angeles Times, August 5, 1979



“Possibly the worst graphic ever to find its way into print”

-Edward Tufte, 1983, *The Visual Display of Quantitative Information*

From the magazine *American Education*, in the 1970' s

Summary: Graphical methods for frequency distributions

Type of Data	Method
Categorical data	Bar graph
Numerical data	Histogram Cumulative frequency distribution

Summary: Associations between variables

	Explanatory variable	
Response variable	Categorical	Numerical
Categorical	Contingency table Grouped bar graph Mosaic plot	Multiple histograms Cumulative frequency distributions
Numerical	Multiple histograms Cumulative frequency distributions	Scatter plot