Week 1

Data Types
- Categorical (eg. Nominal or Ordinal
  - Eye Colour, Nationality
- Continuous
  - Height, Weight, Income, Energy used

Principles of Summarisation/Compression
- "Over-smoothing" vs "Under-smoothing"
- Information loss vs Ease of understanding/interpretation
- Centre, spread, shape (eg. Symmetry and multi-modality)
  - Numerical as well as Graphical Approaches

Numerical Description
- Consider data $x_1, x_2, \dots, x_n$
- Measure of Centre (Location)
  - Mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
  - Median $\tilde{x}\ is\ the\ solution\ to$: $\sum_{i=1}^{n} I_{(x_i \leq \tilde{x})} \geq \frac{n}{2}\ and\ \sum_{i=1}^{n} I_{(x_i \geq \tilde{x})} \geq \frac{n}{2}$
    - Easier to find the Middle point of the data in ascending order
- Measure of Spread (Variation)
  - Standard Deviation $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
  - Range $R = x_{\{n\}} - x_{\{1\}}$
  - Inter-Quartile Range $IQR = Q_3 - Q_1$
    - $Q_1 = median\{x_i : x_i < \tilde{x}\}\ and\ Q_3 = median\{x_i : x_i > \tilde{x}\}$
- Measure of Shape (Symmetry)
  - Coefficient of Skewness $\gamma_3 = \frac{1}{ns^3}\sum_{i=1}^{n}(x_i - \tilde{x})^3$
  - Quartile Ratio $\delta = \frac{(Q_3 - \tilde{x}) + (Q_1 - \tilde{x})}{IQR}$
  - Best way to examine the "shape" of data is …

Graphical Description
- Categorical
  - Tables, bar charts and pie charts
- Continuous
  - Boxplots and histograms and stem and leaf plots

Probability Models

- Observe repeated outcomes of some phenomenon
- Select probability structure to describe distribution of outcome
- Fit particular distribution by estimating “parameters” (eg. mean, rate)
- Assess model choice using “diagnostics”

Week 2

Probability Basics

- The relative frequency with which a particular event occurs in a collection of identical trials
- Permutations – n objects can be ordered n! ways
- Combinations – A subset r objects can be selected from a collection of n things in $\frac{n!}{(r!(n-r)!)}$ = ${}_nC_r$ ways

Useful Probability Relationships

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A^c|B) = 1 - P(A|B)$$

$$P(AB) = P(A|B)P(B)$$

If $P(AB) = P(A) \times P(B)$ then A and B are independent $(A \perp B)$ therefore, $P(A|B) = P(A)$

$$P(B) = \sum_{j=1}^{k} P(B|A_j) \times P(A_j)$$

Baye's Rule

$$P(A_j|B) = \frac{P(BA_j)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_j)P(A_j)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}$$