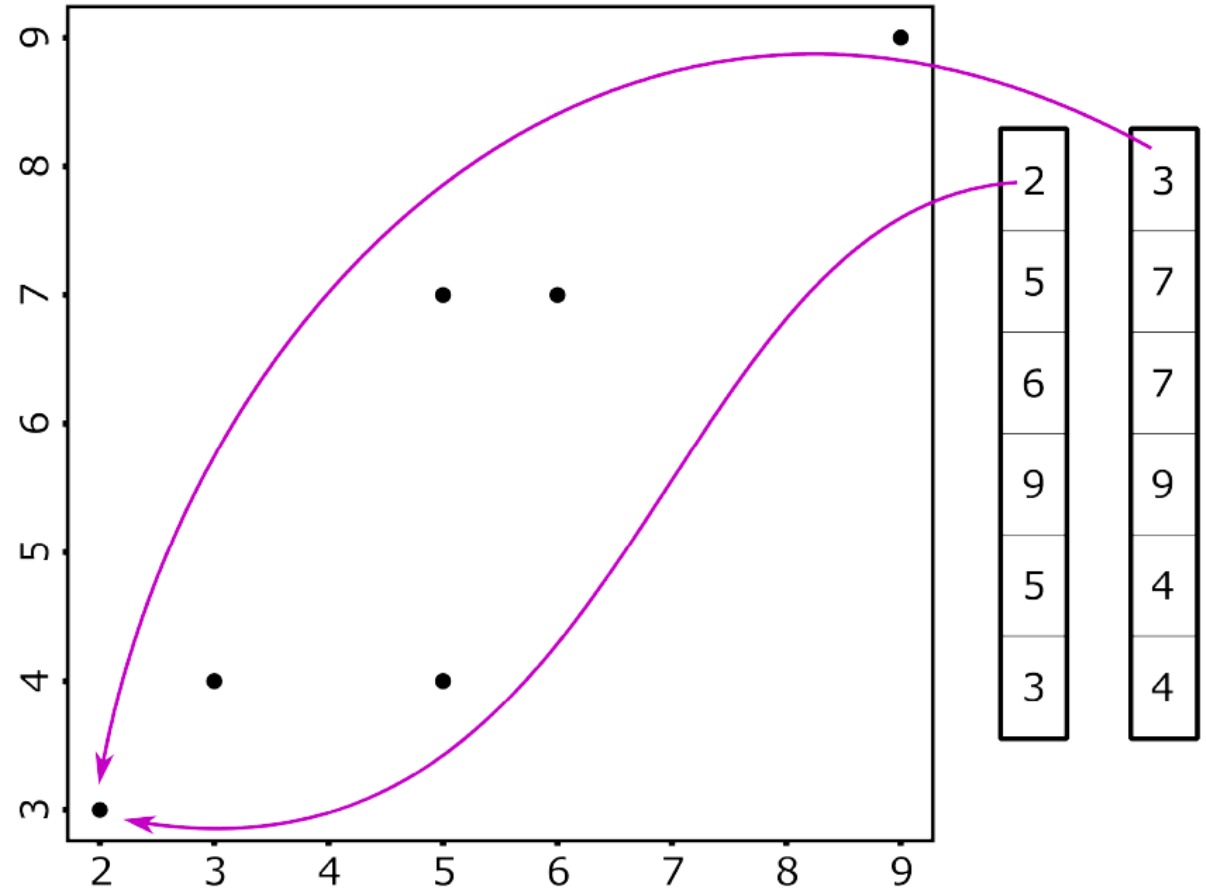# Scatterplots

IBEHS 4C03:  Statistical Methods
in Biomedical Engineering
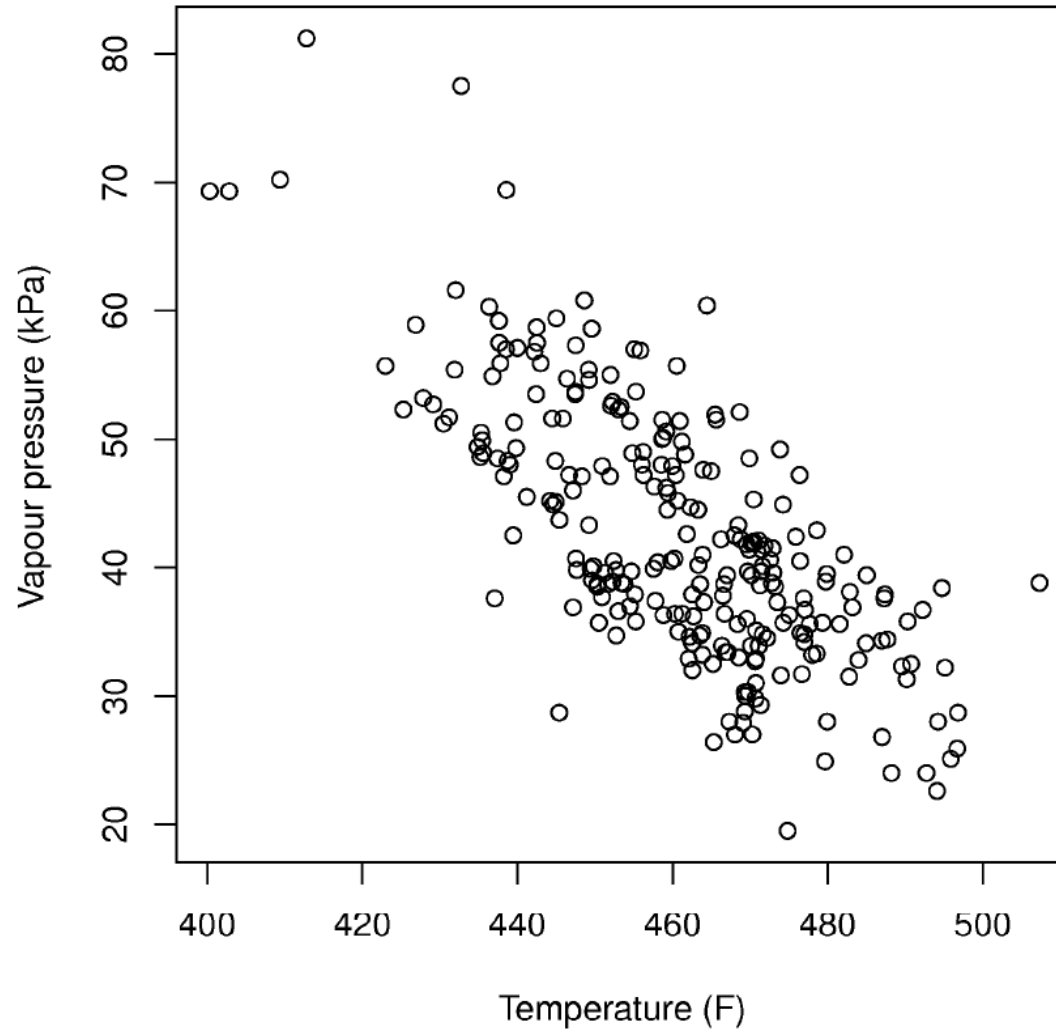
McMaster
University

BRIGHTER WORLD

# Scatter Plots

- 2D plot of two variables/vectors
- Each marker is the intersection of the values from the data vectors
- Asks the viewer to **draw a causal relationship** between the two variables
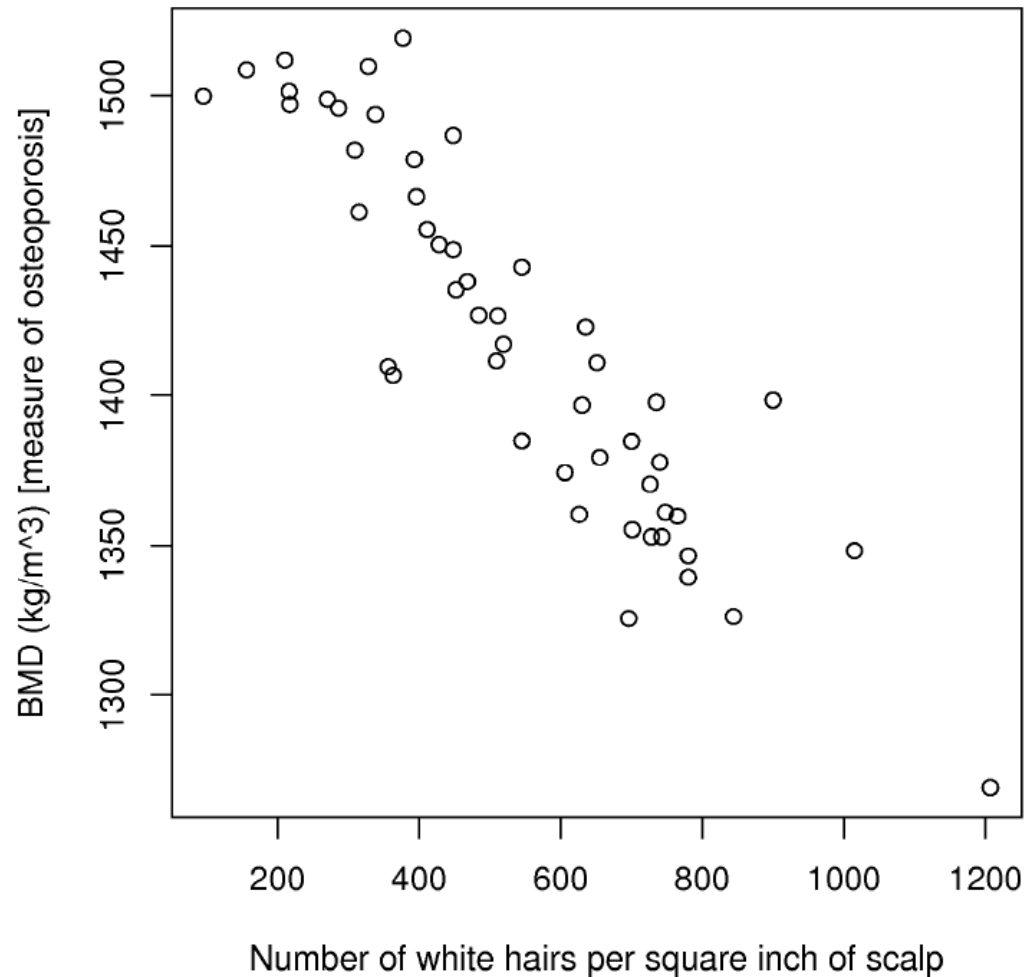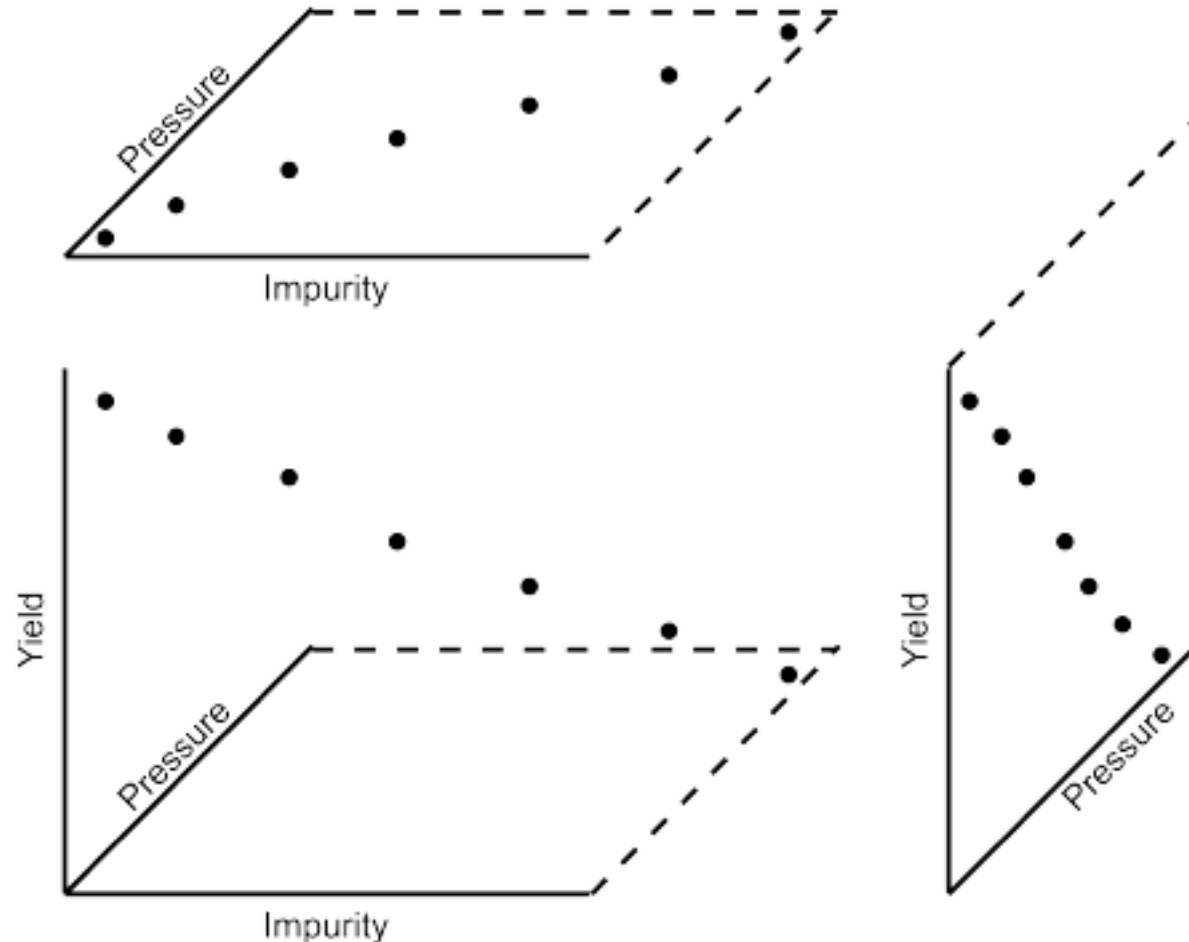
# Scatter Plots - Causality



Often times causal relationships between the variables do exist

# Scatter Plots - Causality



Number of white hairs per square inch of scalp
(y-axis: BMD (kg/m^3) [measure of osteoporosis])

- Other times not so much…
- "Correlation doesn't imply causality"
- Scatter plots aren't a 'tool' to test the existence causality
- Use your engineering judgement

# Scatter plots with three variables



- 3 scatter plot combinations can be drawn
- Which are correlation and which are causal?
- We'll come back to this later in the course (Design of Experiments)

# Will people understand my scatter plot?

- People often worry that scatter plots won't be correctly interpreted by their audience

  - Plant control rooms seldom have scatter plots

  - Tufte showed that scatter plots which are common in Japanese newspapers can be interpreted by 12-year-olds

- Producers of charts must assume their audience is capable of interpreting them

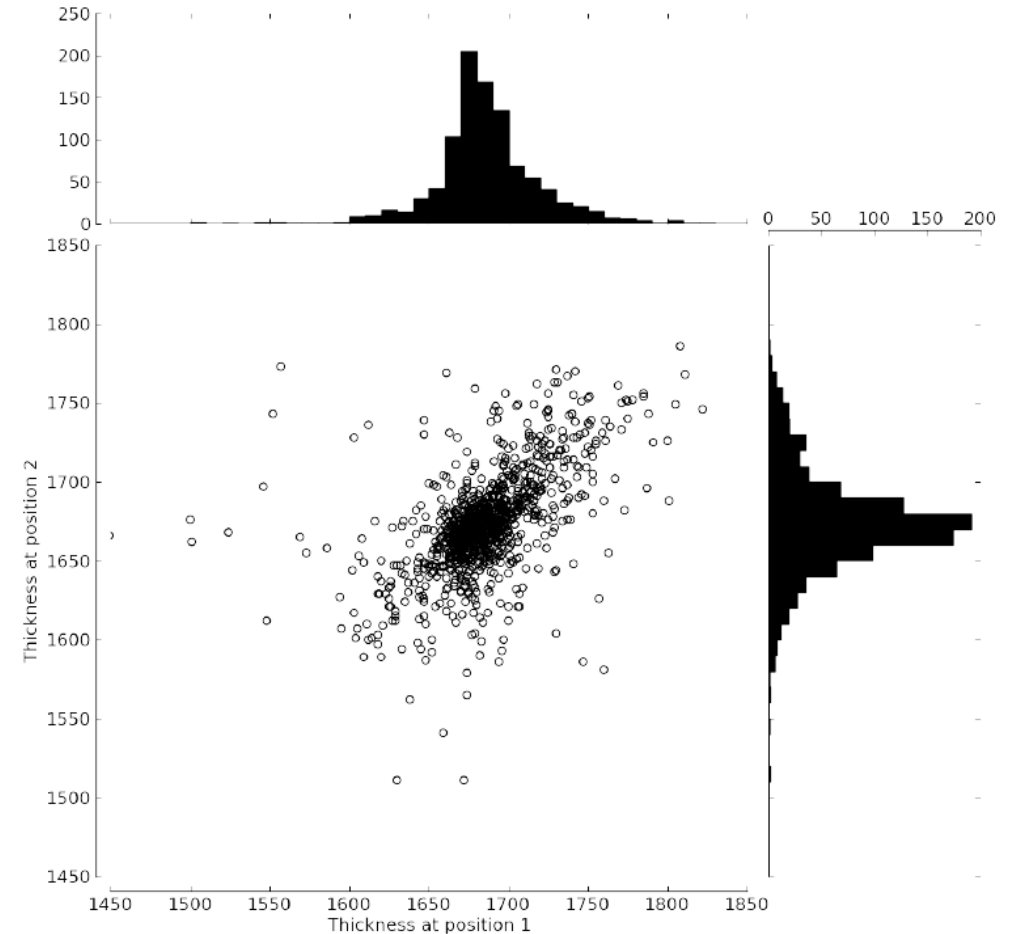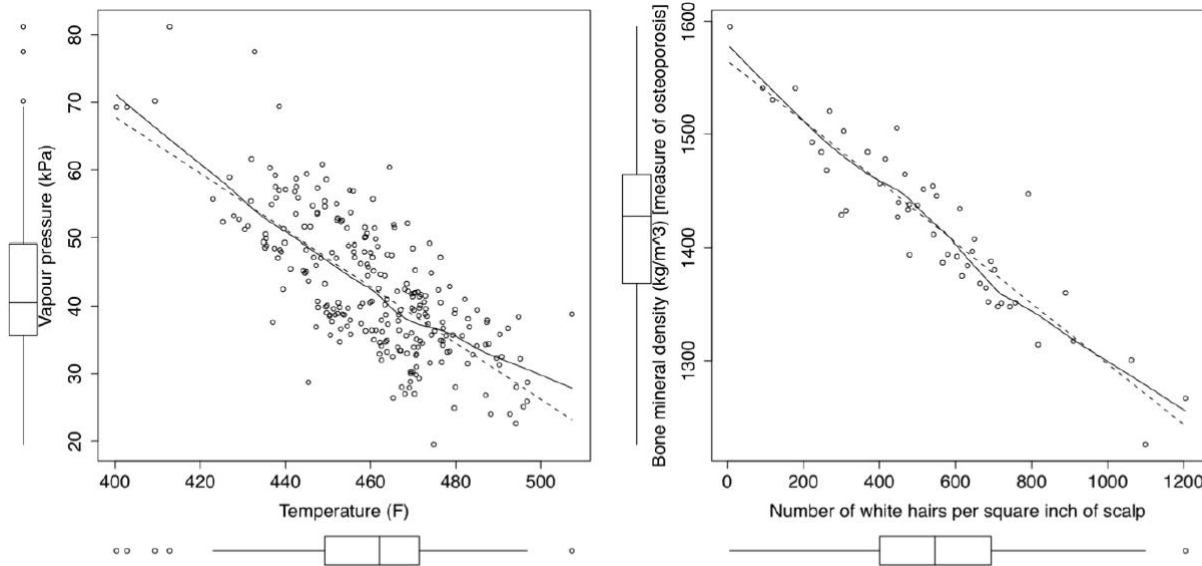- Assume that if you can understand the plot, so will your audience!

# Good scatter plot example

- 5 variables shown in the figure

- X-axis: income per person

- Y-axis: children per woman

- Marker area: population

- Colour: life expectancy

- Animation: evolution of the data over time

- Watch the video here

# Including distributions in scatter plots

- Distribution information can be included in scatter plots by adding box plots or histograms to the side

- Use when scatter plot suffers from overplotting

- Density based plots

# Pairplots