Character sets

- A **character set** is a collection of characters
  - E.g. the ASCII character set is 128 characters, mostly from the modern Latin alphabet
  - E.g. the Unicode character set is currently a little over 100,000 characters
- A **coded character set** assigns a unique number to each distinct character
  - E.g. in Unicode (and ASCII) 'A' is 65 and 'a' is 97 (decimal)
- A **character encoding** refers to the way the numbers are converted to bytes for storage and transmission
  - E.g. ASCII uses 7 bits for every character
  - E.g. UTF-32 uses 4 bytes for every character
  - E.g. UTF-8 uses 1 byte for ASCII characters and 2, 3 or 4 for others

The character encoding of your web page

- Browsers need to know which character encoding was used to create your web page
- When creating a web page,
  - Find out what character encoding your text editor is using
  - If it’s not using something sensible, e.g. UTF-8, change it (or use a better editor!)
  - Specify the character encoding in a **meta** element in the `<head>` of your HTML, e.g.
    ```html
    <meta charset="utf-8" />
    ```
- What happens if your editor uses one encoding but you specify a different one?
  - Some characters may display as other characters
  - Some characters may display as “

A better solution?

- Many web site designers make a complete mess of this
- The Apache web server can be configured so that, when it serves a text file, it converts it to, e.g., UTF-8 — irrespective of its original character encoding
- And it specifies the new character encoding in the **Content-Type** HTTP header
- Browsers treat the HTTP header as more authoritative than the `<meta>` element
Two major types of images

- **Bitmapped** images
  - Consist of pixels (coloured dots) in a grid
  - Their quality depends on their **resolution**: the number of pixels per inch
  - When software decreases the size of such images, it throws pixels away
  - When software increases the size of such images, it inserts new pixels and must guess their colour
- **Vector** graphics images, e.g. Scalable Vector Graphics (SVG)
  - Defined by mathematical equations which describe lines, rectangles, circles, etc.
  - Scalable without the kinds of loss of quality that we get when we resize bitmapped images

Two kinds of bitmapped images

- **Direct colour**
  - The image file specifies the colour of each pixel by giving an RGB code
  - E.g. JPEG
  - Good for images with continuous changing shades and soft transitions, e.g. photos
- **Indexed colour**
  - The image file contains a palette of colours
  - It then specifies the colour of each pixel by giving the colour's position on the palette
  - E.g. GIF — uses a palette of up to 256 colours
  - Good for images with large areas of solid, flat colour, e.g. logos, icons, charts, cartoons
- Some image formats have the advantages of both, e.g. PNG, WebP

HTML5 allows SVG

- You can include SVG images in separate files or directly in your HTML, e.g.

  ```xml
  <svg>
    <circle cx="50" cy="50" r="30"
      stroke="blue" fill="red"/>
  </svg>
  ```

- Inkscape, Adobe Illustrator and Apache Batik are drawing tools that can export images as SVG