

Character encodings

- ▶ **Encoding:** how digital binary signals are translated into human-readable characters.
 - e.g. 0100100 is displayed as 'd'
- ▶ This also includes characters such as á, ç, ü, etc.
- ▶ **Problem:** many different translation tables, sometimes hard to know which one is used
- ▶ R works with the default encoding scheme in your system:

```
> Sys.getlocale(category = "LC_CTYPE")  
[1] "en_US.UTF-8"
```
- ▶ For English Mac and Linux systems, generally UTF-8. For Windows systems, Windows-1252.
- ▶ UTF-8 (part of Unicode standard) is most popular scheme and used on many websites.