Introduction to Data - Multiple Choice Quiz (answers)"

## Exercise Answers

What does Fr[ea]nc[eh] match?

* this matches France, French, Frence, and Franch. It would find words where there were characters either side of these so Francer, foobarFrench, or Franch911.

What does Fr[ea]nc[eh]$ match?

* this matches France, French, Frence, and Franch at the end of a line. It would find words where there were characters before these so foobarFrench.

What would match the strings French and France only that appear at the beginning of a line?

* ^France|^French This would also find words where there were characters after French such as Frenchness.

How do you match the whole words colour and color (case insensitive)?

* In real life, you *should* only come across the case insensitive variations colour, color, Colour, Color, COLOUR, and COLOR (rather than, say, coLour. So one option would be \b[Cc]olou?r\b|\bCOLOU?R\b. This can, however, get quickly quite complex. An option we've not discussed is to take advantage of the / delimiters and add an ignore case flag: so /colou?r/i will match all case insensitive variants of colour and color.

How would you find the whole-word headrest or head rest but not head rest (that is, with two spaces between head and rest?

* \bhead ?rest\b. Note that although \bhead\s?rest\b does work, it would also match zero or one tabs or newline characters between head and rest. In most real world cases it should, however, be fine.

How would you find a 4-letter word that ends a string and is preceded by at least one zero?

* 0+[a-z]{4}\b

How do you match any 4-digit string anywhere?

* \d{4}. Note this will match 4 digit strings only but will find them within longer strings of numbers.

How would you match the date format dd-MM-yyyy?

* \b\d{2}-\d{2}-\d{4}\b In most real world situations, you are likely to want word bounding here (but it may depend on your data).

How would you match the date format dd-MM-yyyy or dd-MM-yy at the end of a line only?

* \d{2}-\d{2}-\d{2,4}$

How would you match publication formats such as British Library : London, 2015 and Manchester University Press: Manchester, 1999?

* .\* : .\*, \d{4} You will find that this matches any text you put before British or Manchester. In this case, this regular expression does a good job on the first look up and may be need to be refined on a second depending on your real world application.

## Multiple Choice Quiz Answers

* Q1. C
* Q2. A
* Q3. B
* Q4. B
* Q5. C
* Q6. B
* Q7. C
* Q8. A
* Q9. C
* Q10. A
* Q11. B
* Q12. A