

Tarea y Solución

Pregunta 1

Indica qué función y parámetros usarías para leer ficheros separados con “|”

Solución

```
read_delim(file, delim = "|")
```

Pregunta 2

Además de `file`, `skip` y `comment` que hemos visto en el curso, ¿qué otros argumentos tienen `read_csv` y `read_tsv` en común? Indica para qué sirve cada uno de ellos.

Solución

De la documentación oficial:

- `col_names` and `col_types` are used to specify the column names and how to parse the columns
- `locale` is important for determining things like the encoding and whether “.” or “,” is used as a decimal mark.
- `na` and `quoted_na` control which strings are treated as missing values when parsing vectors
- `trim_ws` trims whitespace before and after cells before parsing
- `n_max` sets how many rows to read
- `guess_max` sets how many rows to use when guessing the column type
- `progress` determines whether a progress bar is shown.

Pregunta 3

Indica los argumentos más importantes de `read_fwf()`

Solución

El más importante de los argumentos de `read_fwf` para ficheros “fixed-width formats”, es `col_positions` que nos indica en qué posiciones empieza cada una de las columnas de datos.

Pregunta 4

A veces un csv contiene necesariamente comas en los campos que son strings. Para evitar problemas en la carga, suelen ir rodeadas de comillas dobles " o de comillas simples '. La convención de `read_csv()` es que asume que cualquier caracter vendrá rodeado por comillas dobles " y si lo queremos cambiar tenemos que usar la función `read_delim()`.

Indica qué argumentos tendríamos que especificar para poder leer el texto del siguiente data frame correctamente

```
"x,y\n1,'a,b'"
```

Solución

```
data <- "x,y\n1,'a,b'"  
read_delim(data, ",", quote = "'")
```

Pregunta 5

Indica qué está mal en la siguiente línea de lectura de CSV:

```
read_csv("x,y\n1,2,3\n4,5,6")
```

Solución

Solo hay dos columnas en la cabecera pero tres en los datos. Se eliminará la última columna de datos para cada fila de información.

Pregunta 6

Indica qué está mal en la siguiente línea de lectura de CSV:

```
read_csv("x,y,z\n1,2\n3,4,5,6")
```

Solución

Las filas de datos tienen menos o más columnas de información de la cabecera. En la lectura, la primera fila tendrá un NA mientras que la última eliminará la última columna de información.

Pregunta 7

Indica qué está mal en la siguiente línea de lectura de CSV:

```
read_csv("x,y\n\"1")
```

Solución

El carácter escapante está mal indicado.

Pregunta 8

Indica qué está mal en la siguiente línea de lectura de CSV:

```
read_csv("x,y\n1,2\na,b")
```

Solución

Las columnas de datos no son homogéneas: contienen números y caracteres a la vez.

Pregunta 9

Indica qué está mal en la siguiente línea de lectura de CSV:

```
read_csv("x;y\n1;2")
```

Solución

Si el separador es un punto y coma, debemos usar `read_csv2` en lugar del estándar.