



Guía 3 - Análisis de datos científicos en R

Ejercicio 3.1:

Usando el dataset *Birthdays* (que viene en el paquete *mosaicData*, el cual tienen que instalar, si no lo tienen en su sistema), responder las siguientes preguntas:

1. ¿Cuál es el estado que más nacimientos tiene en total? ¿Y el que menos?
2. Calcular la media del total nacimientos y guardarla en una variable llamada **medtot**. ¿Cuántos estados tienen una media mayor que **medtot**? ¿Cuáles son esos estados?
3. ¿Hay algún día del mes o de la semana que sea mucho mayor (más del 10%) que **medtot**?
4. ¿Qué mes del año tiene más nacimientos en promedio?
5. Considerando todo el país ¿Hay algún día (de cualquier año) sin nacimientos?
6. Generar un *data.frame* agrupando por estados y por año, y calcular el total de nacimientos para cada grupo. Graficar con **ggplot2** usando **geom_line**. ¿Cuál está en primer lugar a través de los años? ¿Cuales estados se reparten la última posición? (para esta última pregunta van a necesitar cambiar la ordenada a escala logarítmica).
7. **POSGRADO** – Usando los comandos **slice**, **which.max** y **which.min**, obtener los resultados de la pregunta anterior (es decir, sin graficar).
8. **POSGRADO** – Explorar el funcionamiento del comando **lag()** y, con él, ver en que año hubo la mayor diferencia de nacimientos. Graficar estas diferencias en función de la variable **year**.