



Guía 2 - Análisis de datos científicos en R

Ejercicio 2.1:

1. Cargen a la sesión (usando el comando **data**) el *data.frame* **msleep**. Vean que estructura tiene (cuántas observaciones –filas- y variables –columnas-, los nombres de las variables, el tipo de dato de cada columna. (Hay un único comando que permite ver todo eso.)
2. ¿Qué pasa si hacen solo **ggplot(msleep)**? ¿Es lógico que pase eso?
3. Hagan una figura tipo scatterplot (o sea, con puntos) de *sleep_total* vs. *bodywt*, esta última en el eje x con escala logarítmica.
4. ¿Qué pasa si intentan mapear *sleep_rem* (variable continua) a color, tamaño o forma? (**color=**, **size=** y **shape=**). ¿Qué pasa si intentan mapear *conservation* (variable categórica o discreta) a color, tamaño o forma? ¿Qué se puede concluir de estos dos puntos?.
5. **POSGRADO** – Podemos usar formas condicionales (**>**, **<**, etc) dentro de **aes()**. Investigar como representar la gráfica anterior con colores representando *vore* y con formas diferentes para las observaciones con *brainwt* **>** 0.1.

Ejercicio 2.2:

1. Hacer una figura utilizando una capa para barras (**geom_bar**) del *data.frame* *esoph*, para la variable *ncases*.
2. Hacer subplots con **facet_grid** dividiendo con la variable *alcgp* y con **facet_wrap** dividiendo tanto con *alcgp* como con *tobgp*. (Posiblemente tengan que agrandar el cuadrante de gráficos en Rstudio para ver mejor el resultado.) ¿En qué difieren ambas funciones?
3. Graficar con **geom_boxplot** *ncases* (o sea, asignar a y) comparando con *tobgp* (asignar a x). ¿Cuándo conviene usar este geom?
4. Graficar con **geom_smooth** las variables *ncontrol* en función de *ncases* ¿Cuándo conviene usar este geom?
5. **POSGRADO** – Encontrar, usando todas las variantes de ayuda posibles, para qué sirven los argumentos **jitter** y **stroke**.