



**Curso:** Machine Learning en Agricultura.

**Tutor:** Hugo Andrés Dorado Betancourt

**Taller No 3:** Análisis exploratorio de datos con R

### 1. Realice los siguientes pasos para explorar la base de datos fgl.

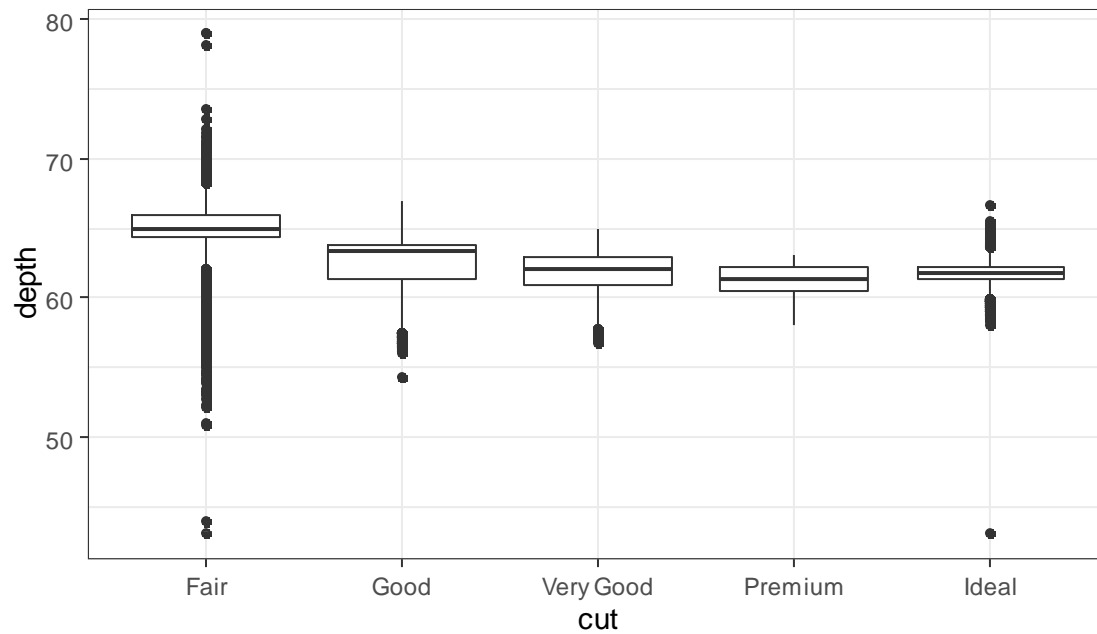
- Cargue la librería MASS
- Para la matriz fgl (que viene dentro del paquete MASS), Determine el número de columnas, el número de filas y nombre las variables utilizando las funciones vistas en clase en R
- **fgl** es un conjunto de datos que contiene muestras de vidrio forense en términos de su índice de refracción (RI), una designación de tipo (type) y porcentajes en peso de ocho óxidos elementales diferentes. Genere un gráfico de la concentración de óxido de magnesio (Mg) versus el índice de refracción (RI), con estas características:
  - Eje x: "Mg concentration"
  - Eje y: "RI concentration índice de refracción"
  - Agregue el argumento las=1 y compare el resultado de usar y no usar dicho argumento.
  - Agregue un titulo que le parezca apropiado.

### 2. Análisis exploratorio de la base de datos de plátano.

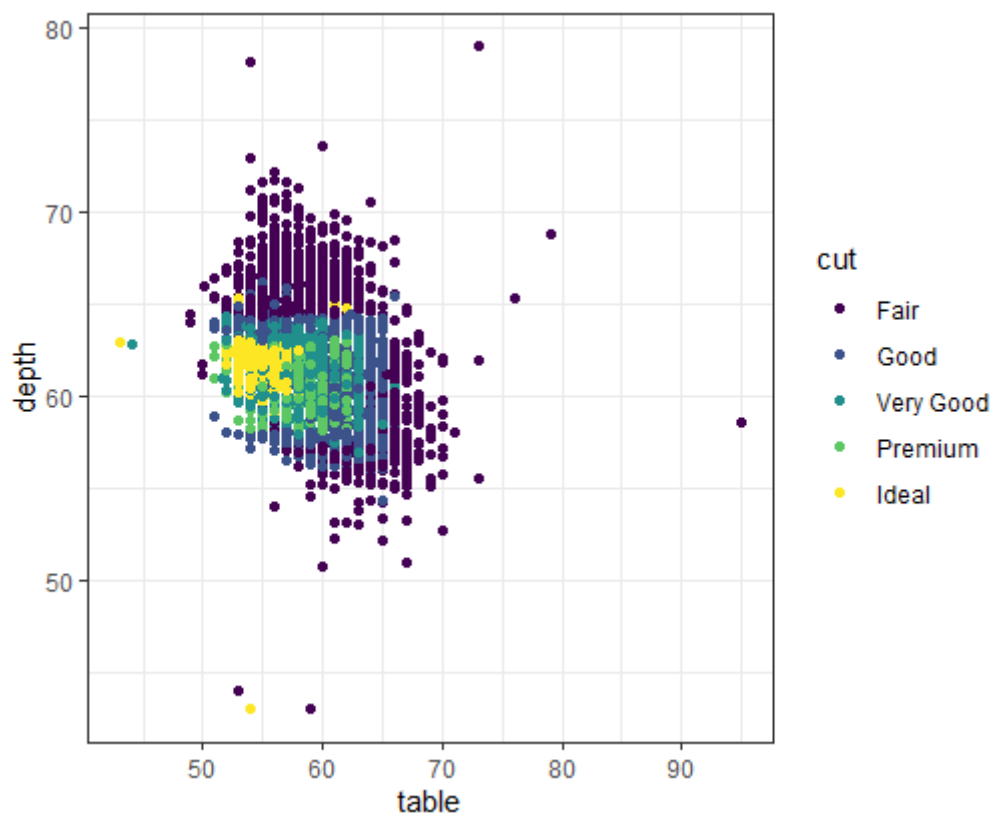
- Leer el archivo eventos\_de\_platano.csv, especificar row.names.
- Ajustar el formato de la fecha, utilizando la función as.Date.
- Calcular el rendimiento (PN\_ANIO/AREA\_UM) y agregarlo directamente usando \$.
- Realizar análisis exploratorio de datos para cada una de las variables de la base de datos. (Medidas de tendencia central y dispersión; complementar con gráficos; hacer matriz de correlación para variables cuantitativas)
- Hacer un gráfico de puntos Edad vs Rendimiento, utilizando la función plot; indique la media de rendimiento dentro del gráfico, utilizando abline.
- Realizar un gráfico de ggplot de barras combinando las variables PATRON\_USADO y dibujo de siembra como un color (revisar ejemplo de geom\_bar).
- Cruzar distancia de fecha vs Rendimiento (geom\_smooth) pero segmentar por dibujo de siembra y PATRON\_USADO.

### 3. Genere los siguientes gráficos en ggplot, utilizando el conjunto de datos diamonds del paquete ggplot.

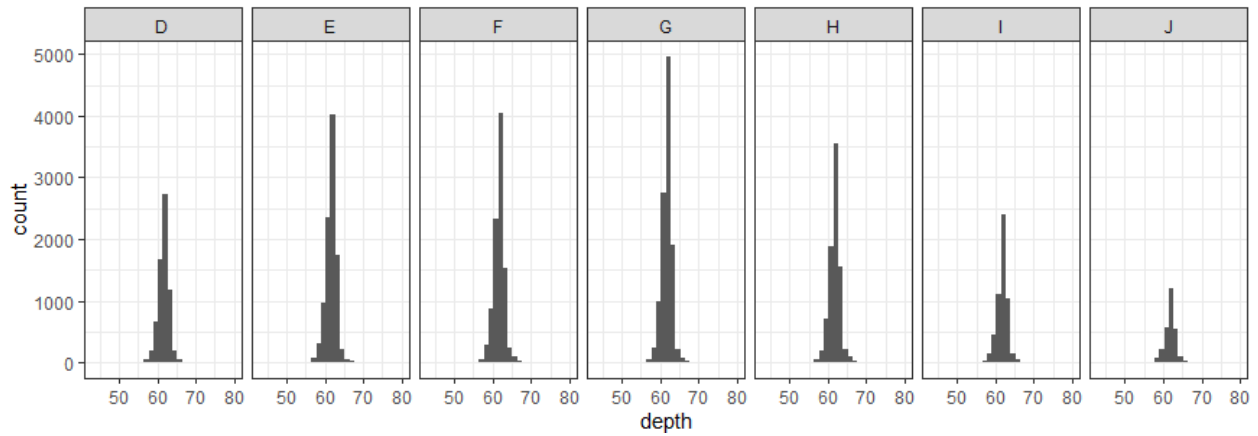
#### Gráfico 1



**Grafico 2**



**Grafico 3**



**4. Utilizando la base de datos de plátano la cual está en `eventos_de_platano.csv`, y la variable de rendimiento que ha creado, realice un gráfico de violín y uno de densidades múltiples utilizando los ejemplos posteados en <https://www.r-graph-gallery.com/135-stacked-density-graph.html> y <https://www.r-graph-gallery.com/95-violin-plot-with-ggplot2.html>, la variable categórica con la que debe cruzar es `PATRON_USADO`.**