

1. Una muestra de 46 clientes se les consultó por el servicio eléctrico que recibían y lo clasificó de Excelente (6), Bueno (18), Regular (15) y el resto de Malo. Construya un gráfico que le parezca apropiado para representar esta información.
Además, si 20 de los encuestados eran mujeres, de las cuales la mitad se inclinó por Bueno, cinco por Regular y una por Malo. Construya el gráfico pertinente para representar esta información.
2. En años pasados, el Departamento de Personal de una empresa eléctrica creció como sigue, en 2018: 46 administrativos, de los cuales 26 eran mujeres; en 2019: aumentó en 10 mujeres y 3 hombres, en 2020: 32 hombres de un total de 69 administrativos. Haga todos los gráficos que pueda (completos) para representar esta información.
3. Considere el conjunto de datos iris. Estos datos consisten de mediciones de largo y ancho de los pétalos y sépalos de 150 flores de iris (50 flores de cada una de tres especies diferentes: Setosa, Versicolor y Virginica). Efectúe un análisis exploratorio para comparar esas tres especies, es decir, para cada especie, calcule: a) mínimo, b) máximo, c) media, d) mediana, e) moda, f) cuartiles y g) varianza.
¿Existen funciones o comandos en R que hagan las cuentas pedidas?
4. Considere el data frame “trees” disponible en R.
 - a) ¿Qué contiene? Haga un análisis exploratorio para cada columna de datos.
 - b) ¿Cuántos árboles:
 - i) tienen diámetro (d) entre 10 y 15?
 - ii) tienen altura (h) entre 60 y 70?
 - iii) tienen volumen (v) entre 20 y 60?
 - iv) cumplen dos o tres de las restricciones anteriores?, ¿cuáles?
 - c) Calcule el área basal de cada árbol, la cual está dada por $g = \frac{\pi}{4}d^2$.
 - d) Haga un gráfico de dispersión (plot).
5. Con la información del archivo “datos_titanic.csv”, que está en la plataforma adecca, genere los gráficos pedidos a continuación y solo cuando sea posible coloque título, rótulos en los ejes con las respectivas unidades de medida, leyendas apropiadas, colores distintivos y diferentes símbolos.
 - a) En un mismo gráfico represente la información del sexo versus grupos etarios y viceversa. ¿Cuántas mujeres viajaban en el Titanic? y ¿cuántos hombres?
 - b) Grafique la cantidad de personas respecto del lugar embarcado.
 - c) Haga un gráfico apropiado para representar la información de la tarifa (**Fare**) cobrada.
 - d) En un mismo gráfico represente el número de sobrevivientes según el sexo y lugar embarcado.
6. Instale el paquete “ggplot2” y luego use la librería “ggplot2” donde tendrá acceso a la base de datos “diamonds”. Para más información de estos datos, ejecute **?diamonds** en la consola de R.

Para cada una de las calidades de corte (**cut**) genere **listas** distintas. A cada lista dele el mismo nombre que la calidad (Fair, Good, Very Good, Premium, Ideal). En cada una de las listas, genere las siguientes tres columnas, con los respectivos nombres: a) **precio**, b) **características**: color, claridad y peso, y c) **coordenadas**: x, y, z.

7. Desde la plataforma adecca, descargue el archivo de texto “Medidas.txt” y guárdelo en su equipo. Luego, léalo desde R usando la función `read.table(file=“Medidas.txt”, header=TRUE, sep=“ ”)`.

Con estos datos, genere dos **listas**, una llamada “Mujer” y la otra “Hombre”. En cada lista, genere tantas columnas, de acuerdo a grupos de edad: a) Edad de 10 a 19, b) Edad de 20 a 29 y así sucesivamente hasta completar los datos de la respectiva lista. En estas columnas, ubique un bloque para el peso, otro para la altura y en otro bloque, llamado “diámetro”, deje los datos de la muñeca y del bíceps.

8. Considere las precipitaciones que están en el archivo “Precip.Estacion.8122002.txt”, alojado en la plataforma adecca, que contiene dos columnas, en la primera está Año-mes y en la segunda la precipitación respectiva.

- a) En un mismo gráfico y con distintos colores presente las precipitaciones mensuales para los diferentes años. En el eje horizontal coloque los meses (enero, febrero, . . . , diciembre) y en el eje vertical las precipitaciones. Ponga un título apropiado al gráfico y rótulos en los ejes.
- b) Forme un vector con los promedios anuales, solo de aquellos años que tienen registrados los 12 meses, es decir, no considere los años incompletos. Si hay datos con -9999, considérellos como “NA”.
- c) Ajuste una función de densidad de probabilidad a los datos del vector construido en el ítem 5. b). Para ello, siga los pasos desarrollados en la sección 5.5.6. del texto “El arte de programar en R. Un lenguaje para la estadística”.

9. Considere la base de datos “airquality”, que ya está en R, la cual contiene medidas de la calidad del aire de New York, desde el 1 de mayo de 1973 (martes) al 30 de septiembre de 1973. Para saber más detalles de esta base puede usar el comando `?airquality`. Todo lo que se solicita a continuación debe ser automatizado.

- a) En un mismo plot, pero en ventanas distintas, presente la información mensual de todas las características que contiene el data frame (Ozone, Solar.R, Wind y Temp). En cada ventana debe presentar un mes. Ponga título y rótulos apropiados. Presente una sola leyenda para todos los gráficos. En el eje horizontal deben estar los días (1, 2, . . . , 31).
- b) Obtenga un promedio mensual de: Ozone, Solar.R, Wind y Temp, e imprima estos promedios en la consola.
- c) ¿En qué día y mes fue más alto el Ozone?, Solar.R?, Wind? y Temp? ¿En qué día y mes fueron más bajas dichas mediciones?
- d) Para el mes en que el Ozone fue más alto, haga un gráfico que presente en forma semanal todas las características que contiene el data frame (Ozone, Solar.R, Wind y Temp). En el eje horizontal presente los días de la semana (Lunes, Martes, . . .).

10. Considere el data frame “CO2” disponible en R. Para más información puede revisar el link <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/zCO2.html>
- a) Haga un análisis exploratorio para comparar la absorción (**uptake**) de los dos tipos (**Type**) de planta (Quebec y Mississippi) a las distintas concentraciones (95,175,...). Para ello, compare mínimos, máximos, medias, cuartiles y desviaciones estándar. Intente comparar las modas. Le puede ser útil hacer una tabla para cada uno de los estadísticos a comparar.
 - b) Compare la absorción, como en a), de los dos tipos de planta para cada tratamiento (**Treatment**).
 - c) Compare la absorción, como en a), desglosado ahora por identificador (**Plant**).
 - d) Haga gráficos de las absorciones para los siguientes casos:
 - por cada concentración.
 - por tratamiento.
 - por identificador.
 - e) Para cada tipo de planta, use el coeficiente de variación para:
 - i) comparar la absorción entre los tratamientos. ¿qué tratamiento es mejor?
 - ii) comparar la absorción entre los identificadores. ¿qué identificador resultó mejor?
 - f) Para cada tipo de planta, grafique las absorciones en las siguientes situaciones:
 - por cada concentración.
 - por tratamiento.
 - por identificador.
 - g) Para cada tratamiento, grafique las absorciones para los siguientes casos:
 - por cada concentración.
 - por tipo de planta.
 - por identificador.
11. Considere el paquete “datos” que contiene los 336 776 vuelos que partieron de la ciudad de Nueva York durante el 2013. Para ello ejecute **install.packages(“datos”)**, se abrirá una ventana y elija **Chile (Santiago)** [https], luego pulse **OK**. Para acceder a la base ejecute **library(datos)**. Usaremos los datos de “vuelos”, si desea saber más sobre el contenido de esta base, ejecute **?vuelos**. Con **View(vuelos)** verá todos los datos.
- a) Haga un gráfico mensual del número de vuelos, ubique los meses como abscisas y un título apropiado. ¿En qué mes hubo más vuelos y en cuál hubo menos?
 - b) Obtenga los 10 vuelos más retrasados. ¿Cómo quiere manejar los empates?
 - c) Obtenga los 10 vuelos que viajaron más lejos. ¿Cuál viajó más cerca?
 - d) Haga un diagrama de dispersión del número de vuelos versus la demora promedio.
 - e) Agrupe los vuelos por destino, ¿Cuántos destinos hay? Para cada destino, calcule:
 - distancia promedio,
 - demora promedio,

- el número de vuelos.
- f)* Haga un gráfico diario del número de vuelos del mes que hubo más vuelos. En el mismo gráfico, presente diariamente el número de vuelos del mes en que hubo menos vuelos. De un título y nombres apropiados a los ejes.
- g)* Obtenga los 10 vuelos con menor tiempo de vuelo.
- h)* Grafique el tiempo de vuelo versus la distancia recorrida y calcule e interprete el coeficiente de correlación entre ellos, para esto emplee la función **cor**, con el argumento **use="complete.obs"**.
- i)* Agrupe los vuelos por aerolínea, ¿cuántas aerolíneas hay? Para cada aerolínea, calcule:
- distancia promedio,
 - demora promedio,
 - el número de vuelos.