Profesor: M.Sc. Fabio Fernández Jiménez

Visualización e Interpretación de Datos

Tarea Número 3

Notas:

- Las tareas tienen fecha de entrega una semana después a la clase, y deben ser entregadas **antes del inicio de la clase siguiente**. **Cada día de atraso** en la entrega **tendrá un castigo de 10 puntos.**
- Las tareas son estrictamente de carácter individual
- Debe entregar el resultado como un archivo auto-reproducible (HTML o PDF) con la solución a todas las preguntas. Debe mostrar la respuesta en prosa a lo que se consulte (si aplica), así como el código utilizado y los resultados de ese código (tablas, gráficos, etc.)
- En nombre del archivo debe tener el siguiente formato: *Tarea1_nombre_apellido.pdf(*o *.html)*. Por ejemplo, si el nombre del estudiante es Luis Pérez: ***Tarea1_luis_perez.pdf***. Para la tarea número 2 sería: ***Tarea2_luis_perez.pdf***, y así sucesivamente.
- El puntaje de cada pregunta se indica en su encabezado.

Para todas las siguientes preguntas apóyese en la herramienta visual para dar solución a las mismas. Sea elaborado con respecto a títulos, etiquetas de ejes, así como colores y leyendas en los gráficos. Todas las soluciones deben estar basadas estrictamente a funciones dentro de **lattice, rgl** o **maps**, según se indique.

I. El archivo "housing.csv" contiene información de viviendas en vecindarios de Boston. Las variables en el archivo son:
   - CRIM: tasa de delincuencia per cápita
   - ZN: proporción de zonas residenciales con lotes mayores a 25,000 pies cuadrados
   - INDUS: proporción de acres para negocios distintos al de menudeo
   - NOX: concentración de óxidos nítricos (partes por 10 millones)
   - RM: número promedio de cuartos por unidad
   - AGE: proporción de unidades ocupadas por dueños construidas antes de 1940
   - DIS: distancias ponderadas a 5 centros de empleo en Boston

(506) 2268.8823 - (506) 8708.9091 info@promidat.com facebook.com/promidat www.promidat.com

- RAD: índice de accesibilidad a autopistas
- TAX: tasa de impuesto a la propiedad por cada $10,000
- PTRATIO: tasa de estudiantes por maestro

Use el paquete *lattice* para dar solución a las siguientes consultas.

1. ¿Qué puede determinar de la relación entre tasa de delincuencia (CRIM), la tasa de impuesto a la propiedad (TAX), y la distancia a los centros de empleo? Utilice un *xyplot* para mostrar dicha relación, para esto especifique la tasa de delincuencia en función de la distancia, segmentada por el impuesto TAX. El impuesto es una variable continua, conviértalo a discreta creando cuatro grupos, use esta variable discreta en la segmentación.

   Al convertir el impuesto en discreta, podrán crearse algunos intervalos con *overlap*. Use el parámetro "*groups*" de *xyplot* para mostrar con colores el intervalo real al cual pertenece la observación. (20 pts.)
2. Adicione una "dimensión extra" en el gráfico anterior por medio de grupos usando el índice de accesibilidad a autopistas. ¿Aporta nueva información estos grupos? (5 pts.)
3. Muestre la distribución (histogramas) de edad (AGE), segmentando por la tasa de impuestos. Explique los resultados (10 pts.)

II. Prepare un set de datos (al menos 6 variables continuas) para utilizar las herramientas de plotly e interprete la información haciendo uso de las funcionalidades disponibles.

1. Explique en qué consisten los datos. (5 pts.)
2. Utilizando *plot_ly* muestre la relación entre 3 variables continuas. Explique lo que busca analizar e interprete los resultados (10 pts.). Sea detallado en la parametrización de colores y etiquetas.
3. Sobre el total del set de datos inicial aplique algún método de *clustering* para definir grupos. Asigne los grupos a los individuos y utilice algunos de los tipos de gráficos para explicar los resultados (no obtenga más de 3 grupos) (20 pts.)

No puede utilizar ningún archivo de datos usado en PROMiDAT o que venga por default con los paquetes de R.

(506) 2268.8823 - (506) 8708.9091 info@promidat.com facebook.com/promidat www.promidat.com

III. Los datos en WDS2014v3.csv contienen información de 214 países de todos los contienes relacionada con diversos indicadores:

- P2014: población total en el 2014
- TMI: Tasa de mortalidad infantil (por cada 1000 nacimientos)
- TFR: Tasa de fertilidad
- EVT: Esperanza de Vida al nacer (en años)
- EVH: Esperanza de Vida al nacer en hombres (en años)
- EVM: Esperanza de Vida al nacer en mujeres (en años)
- REG: Región
- ARE: Área geográfica (continente)
- INC: Ingreso per cápita (en US$)
- OOS: Porcentaje de adolescentes fuera del sistema educativo
- TNT: Tasa de natalidad
- TSB: Tasa de sobornos (% de empresas que han recibido al menos una solicitud)
- ABS: Área forestal (%)
- GDP: GDP per cápita (US$$)
- IPC: Inflación
- ISR: Número de usuarios con acceso a Internet (%)
- MCS: Tasa de uso de celulares
- TDS: Tasa de desempleo

Use el paquete *maps* para dar solución a las siguientes consultas.

1. ¿En qué regiones se concentra la mayor cantidad de países con mayor área forestal? Construya un mapa con *maps* y muestre únicamente 4 -grupos de- colores. Asegúrese que todos los países de América del Norte sean graficados. (15 pts.)

(506) 2268.8823 - (506) 8708.9091 info@promidat.com facebook.com/promidat www.promidat.com

2. A partir del código siguiente, grafique el continente americano y señale (escoja el método que desee) las capitales de los países de América del Norte, indique la expectativa de vida para cada país como etiquetas sobrepuestas en el mapa. Puede enfocar el mapa en esa región para una mejor visualización (15 pts.)

```
#Instalar y cargar el mapa
install.packages("RgoogleMaps", dependencies = TRUE)
library(RgoogleMaps)


#Datos de capitales
datos <- read.csv(file="world_cities.csv", head=TRUE, sep=",", dec = ".")
capital <- datos[datos$city=="Ottawa",]

lat <- c(capital$lat -20,capital$lat+20) #Rango en y
lon <- c(capital$lng-20,capital$lng+20) #Rango en x
center = c(capital$lat, capital$lng)  #Centro del gráfico
zoom <- 5  #zoom: 1 = Todo el globo,

#Mapa
terrmap <-  GetMap(center=center, zoom=zoom, maptype= "roadmap" , destfile =
"CA.png") #graficar mapa
PlotOnStaticMap(terrmap)
text(x=1, y= capital$lat, labels = "EVT:=82", cex = 0.8)
```


(506) 2268.8823 - (506) 8708.9091 info@promidat.com facebook.com/promidat www.promidat.com