

Task 2

Tema: `ggplot2`

Profesor: Eduard F. Martínez González

Agosto 2025

Introducción

Este taller ha sido diseñado con el propósito de consolidar los conocimientos en torno a la manipulación de datos y la visualización de información, haciendo uso de la librería `ggplot2` y de los contenidos desarrollados en la semana 4 del curso.

La base de datos sintética `data-task-ggplot.csv` contiene información sobre clientes, productos y transacciones realizadas en varias ciudades de Colombia, entre los años 2022 y 2024. Incluye variables como el segmento de clientes, la categoría de productos, el precio base, los descuentos aplicados, la cantidad vendida y los ingresos generados. La base de datos también incorpora valores faltantes (NA) en algunas observaciones, con el fin de ejercitar procesos de limpieza y filtrado.

Repositorio

Para resolver este taller, el estudiante deberá **descargar y descomprimir** el repositorio `task-dplyr.zip`, disponible en el siguiente enlace: [aquí](#). El repositorio contiene la siguiente estructura:

- Un archivo `proj.Rproj`, el cual debe ser utilizado para **iniciar RStudio**. Esto garantizará que el directorio de trabajo se establezca correctamente en la carpeta principal del repositorio.
- Una carpeta `code`, que incluye el **script base** sobre el cual deberá desarrollar la solución del taller. En caso de requerir librerías adicionales, estas deberán ser instaladas y llamadas al inicio del script.
- Una carpeta `input`, que contiene la **base de datos necesaria** para resolver los ejercicios propuestos.

Se enfatiza que el correcto uso del archivo `proj.Rproj` es indispensable para evitar errores relacionados con rutas o directorios durante el desarrollo del taller.

Instrucciones

- No seguir correctamente **todas** las instrucciones del taller implicará una penalización del **20 % sobre la nota total**.
- El taller puede desarrollarse en grupos de hasta dos personas. Únicamente uno de los integrantes debe cargar la entrega en la plataforma Intu.
- Este documento presenta dos opciones de trabajo —*Taller 1* y *Taller 2*—. **Cada grupo debe escoger únicamente uno de los dos talleres para su desarrollo.**
- La fecha límite de entrega es el **martes 26 de agosto a las 14:00 para el Grupo 3 y viernes 29 de agosto a las 10:00 a.m. para el Grupo 7**. A partir de ese momento la plataforma Intu se cerrará y no permitirá subir más archivos.
- La plataforma Intu recibirá exclusivamente dos archivos:
 - Un archivo en formato **.R** (script), con el desarrollo de los puntos del taller, documentado de manera clara y ordenada.
 - Un archivo en formato **PDF** (no Word ni ningún otro formato), que contenga las respuestas interpretativas solicitadas.
- El script en **R** debe cumplir con las siguientes condiciones:
 - Incluir al inicio: los nombres de los integrantes, la versión de R utilizada y la carga de todas las librerías necesarias.
 - Estar limpio y organizado: elimine funciones o líneas de código innecesarias.
 - Documentar claramente el desarrollo de cada punto utilizando comentarios: **## Punto 1**, **## Punto 2**, etc.

```
## Nombre(s) de Autor(es)
## R version 4.5.0

## limpiar entorno
rm(list=ls())

## cargar librerías
require(dplyr)
require(skimr)
require(janitor)
require(rio)

## Punto 1
...
## Punto 2
...
```

Opción Taller 1: Exploración y Visualización de Datos de Ventas

1. (20 pts) Cargue la base de datos y realice un resumen general con `skimr::skim`. Reporte:
 - Número de observaciones y variables.
 - Variables con valores faltantes.
2. (20 pts) Genere un gráfico de barras con el ingreso total por `categoría` de producto. Interprete qué categorías concentran más ventas.
3. (20 pts) Construya un histograma de la variable `ingreso`. Interprete la forma de la distribución.
4. (20 pts) Realice un boxplot de ingresos por `segmento`. Explique qué diferencias existen entre segmentos.
5. (20 pts) Construya un diagrama de dispersión entre `edad` del cliente e `ingreso`, usando color para diferenciar el `segmento`. ¿Se observan patrones de consumo según segmento?

Total: 100 pts

Opción Taller 2: Análisis de Ventas y Storytelling

1. (20 pts) Cargue la base de datos y filtre únicamente las ventas con ingresos no nulos.
2. (20 pts) Construya un gráfico que muestre la evolución mensual de las ventas totales entre 2022 y 2024. Interprete tendencias generales.
3. (10 pts) Exporte al menos un gráfico en formato PNG utilizando `ggsave()`.
4. (50 pts) Diseñe una mini-narrativa:
 - Seleccione una pregunta de negocio (ejemplo: *¿qué segmentos y categorías impulsaron las ventas en 2024?*).
 - Construya al menos 4 gráficos distintos (barras, boxplot, dispersión, facetas, etc.) que respondan a la pregunta.
 - Personalice cada gráfico con títulos, subtítulos, etiquetas y escalas adecuadas.

Total: 100 pts