

## PROGRAMA DEL CURSO

### Análisis Exploratorio y Visualización de Datos con R

---

#### **1. INFORMACIÓN GENERAL DEL CURSO**

**Nombre del curso:** Análisis Exploratorio y Visualización de Datos con R

**Modalidad:** Virtual

**Duración:** 8 videos × 40 minutos (aproximadamente 5 horas)

**Nivel:** Educación continua / Posgrado

**Requisitos:** Conocimientos básicos de R y estadística descriptiva. R 4.0 o superior instalado.

---

#### **2. DESCRIPCIÓN DEL CURSO**

Curso orientado al análisis exploratorio y la visualización de datos mediante el uso de R, con un enfoque aplicado y basado en scripts. Se utiliza el ecosistema tidyverse para explorar, resumir y visualizar datos reales, priorizando la interpretación de datos por encima de la formalización estadística. Los materiales (scripts y datos) están disponibles en GitHub, y el contenido se desarrolla a través de videos tutoriales alojados en YouTube.

---

#### **3. OBJETIVO GENERAL**

Desarrollar habilidades para explorar, visualizar e interpretar datos mediante R, utilizando estrategias gráficas y descriptivas adecuadas según el tipo de variable y el objetivo analítico.

---

#### **4. OBJETIVOS ESPECÍFICOS**

- Identificar el tipo de variables presentes en un conjunto de datos
  - Aplicar técnicas de exploración y visualización adecuadas para variables categóricas y cuantitativas
  - Interpretar patrones, asociaciones y diferencias entre variables
  - Utilizar R como herramienta para el análisis exploratorio aplicado
-

## 5. CONTENIDOS DEL CURSO

Capítulo	Tema	Contenidos
1	Introducción al análisis exploratorio de datos	Rol del AED, tipos de variables, escalas de medición, tipos de análisis según tipo de variable, R como herramienta de análisis
2	Repasso operativo de R (CSV, SPSS, Excel, Stata, Jamovi), funciones básicas de dplyr (filter, arrange, select, mutate), exploración de dataframes	Instalación y carga de paquetes, importación de datos
3	Exploración y Recodificación conforcats, tablas de frecuencias con visualización dejanitor, formatos conknitr ygt, gráficas de barras, análisis variables categóricas de múltiples variables categóricas	y Recodificación con forcats, tablas de frecuencias con dejanitor, formatos conknitr ygt, gráficas de barras, análisis de múltiples variables categóricas
4	Análisis univariante de variables cuantitativas	Estadísticas descriptivas (media, mediana, DE, cuartiles, asimetría, curtosis), histogramas y densidades, boxplots, QQ-plots, evaluación de normalidad, estandarización
5	Análisis bivariado entre variables categóricas	Tablas de contingencia (2x2 y RxC), gráficas de barras apiladas, análisis estratificado, medidas de asociación (Phi, Cramer, Gamma, Tau-b)
6	Análisis bivariado de correlación	Scatter plots con distribuciones marginales, matrices de correlación, coeficientes de correlación (Pearson, Spearman), mapas de calor, reordenamiento jerárquico
7	Análisis de diferencia	Comparación de dos o múltiples grupos, boxplots y violin plots, análisis estratificado, tablas comparativas
8	Análisis exploratorio automatizado	Tablas automáticas congtsummary, modelsummary y summarytools, tablas estratificadas, personalización, generación de reportes

---

## 6. METODOLOGÍA DE ENSEÑANZA-APRENDIZAJE

El curso se desarrolla mediante videos tutoriales de aproximadamente 40 minutos cada uno, alojados en YouTube. Cada video presenta un script de R completamente funcional y comentado, ejecutado paso a paso, mostrando resultados inmediatos y su interpretación.

Los materiales del curso (scripts, dataset, datos auxiliares) están disponibles en un repositorio GitHub. Los estudiantes pueden clonar o descargar el repositorio para trabajar localmente con el código, experimentar con modificaciones, y adaptarlo a sus propios datos.

La metodología enfatiza el aprendizaje práctico mediante la ejecución de código y la interpretación de resultados.

---

## 7. RECURSOS Y HERRAMIENTAS

### **Software requerido:**

- R (versión 4.0 o superior)
- RStudio Desktop (gratuito)

### **Paquetes principales:**

*Manipulación de datos:* readr, dplyr, tidyr,forcats, here

*Visualización:* ggplot2, ggpublisher, ggExtra, GGally

*Importación de datos:* haven, readxl, jmvReadWrite

*Tablas y resúmenes:* knitr, gt, janitor, gtsummary, modelsummary, summarytools

*Estadística:* moments, rstatix, DescTools

### **Materiales:**

- Scripts R (8 archivos, uno por capítulo)
  - Dataset: data\_covid.csv (153 observaciones, 45+ variables clínicas)
  - Diccionario de variables
  - Repositorio GitHub con todos los recursos
  - Videos en YouTube (8 videos × 40 minutos)
- 

## 8. BIBLIOGRAFÍA BÁSICA Y COMPLEMENTARIA

### **Textos principales:**

- Wickham, H., & Grolemund, G. (2023). *R for Data Science* (2nd ed.). O'Reilly Media.
- Wilke, C. O. (2019). *Fundamentals of Data Visualization*. O'Reilly Media.

- Chang, W. (2018). *R Graphics Cookbook* (2nd ed.). O'Reilly Media.
- Schwabish, J. (2021). *Better Data Visualizations*. Columbia University Press.

**Documentación oficial de paquetes:**

- ggplot2: <https://ggplot2.tidyverse.org/>
- dplyr: <https://dplyr.tidyverse.org/>
- tidyr: <https://tidyr.tidyverse.org/>
- forcats: <https://forcats.tidyverse.org/>
- gtsummary: <http://www.danielsjoberg.com/gtsummary/>
- modelsummary: <https://vincentarelbundock.github.io/modelsummary/>
- summarytools: <https://dcomtois.github.io/summarytools/>
- rstatix: <https://rpkgs.datanovia.com/rstatix/>

**Recursos en línea:**

- R Graph Gallery: <https://r-graph-gallery.com/>
  - RStudio Community: <https://community.rstudio.com/>
- 

**Versión:** 1.0

**Última actualización:** Diciembre 2025

**Autor:** Prof. Andre Chocó-Cedillos

**Contacto:** panteisme@yahoo.com

**Institución:** Universidad de San Carlos de Guatemala