



# **CLASE 1**

## **INTRODUCCIÓN AL CURSO ANÁLISIS AVANZADO DE DATOS II**

Gabriel Sotomayor  
Marzo 2024

# PRESENTACIÓN

El curso profundiza las principales técnicas multivariantes, que permiten analizar en forma conjunta la interacción de múltiples factores, lo cual da la posibilidad de estudiar de forma más compleja los problemas sociales.

Se enfatizan la comprensión de procedimientos e interpretación de resultados, así como el uso de herramientas computacionales, principalmente en R y RStudio. Se trabajará en la aplicación de los métodos sin necesidad de estudiar en profundidad los fundamentos matemáticos de todos ellos.

Al final del curso el alumno deberá ser capaz de realizar análisis descriptivo de los datos tanto primarios como secundarios y poder seleccionar la técnica estadística adecuada a utilizar cuando se tienen datos multivariantes.

# RESULTADOS DE APRENDIZAJE

**Resultado general:** Entregar herramientas al alumno para que pueda abordar el análisis de datos desde la estadística multivariante descriptiva e inferencial, elaborando y probando hipótesis y desarrollando modelos explicativos de determinados fenómenos de estudio.

# RESULTADOS ESPECÍFICOS

1. Manejar bases de datos complejas.
2. Plantear y estructurar adecuadamente un problema de análisis multivariante.
3. Identificar el tipo de técnica adecuada, o combinación de ellas, a aplicar en cada situación concreta.
4. Comprender las herramientas utilizada en cada una de las técnicas estudiadas.
5. Aprender a utilizar los programas computacionales necesarios para aplicar los distintos modelos estudiados, particularmente R.
6. Interpretar los resultados del análisis.
7. Redactar correctamente un informe con los resultados del análisis.

# CONTENIDOS

<b>1. Gestión de datos con R</b>	<ul style="list-style-type: none"><li>• Introducción a R y Rstudio</li><li>• Gestión de datos con tidyverse.</li><li>• Cálculo de estadísticos descriptivos.</li><li>• Visualización de datos con ggplot.</li></ul>
<b>2. Introducción a los modelos multivariados</b>	<ul style="list-style-type: none"><li>• Uso de modelos en ciencias sociales</li><li>• Modelos exploratorios y modelos confirmatorios</li><li>• Repaso de conceptos estadísticos: covarianza, correlación e inferencia.</li><li>• Supuestos de análisis.</li></ul>
<b>3. Formas funcionales de los Modelos de Regresión</b>	<ul style="list-style-type: none"><li>• Introducción</li><li>• Modelo Logarítmico bivariado</li><li>• Modelo Logarítmico Multivariado</li><li>• Interpretación de los coeficientes</li><li>• Ejemplos de aplicación en R.</li></ul>

# CONTENIDOS

<b>4. Análisis Factorial Exploratorio</b>	<ul style="list-style-type: none"><li>• Análisis factorial exploratorio y su aplicación en la investigación sociológica. El concepto de varianza y covarianza.</li><li>• Diferencias entre el análisis de componente principal y de factor común. Supuestos de la técnica.</li><li>• La extracción de factores comunes. Métodos de extracción de factores, criterios de selección del número de factores y rotación.</li><li>• La matriz factorial y su interpretación. Evaluación del modelo factorial y cálculo de puntuaciones factoriales.</li></ul>
<b>5. Análisis Factorial Confirmatorio</b>	<ul style="list-style-type: none"><li>• Análisis factorial confirmatorio y su aplicación en la investigación sociológica. El análisis confirmatorio frente al análisis exploratorio. Supuestos de la técnica.</li><li>• Especificación e identificación del modelo. Estimación de parámetros, evaluación del ajuste y reespecificación.</li><li>• Fiabilidad y validez de los modelos de medida. Validez convergente y validez discriminante. Evaluación de la capacidad confirmatoria del modelo.</li><li>• Ejemplo de análisis factorial confirmatorio.</li></ul>

# CONTENIDOS

<b>6. Análisis de Sendero</b>	<ul style="list-style-type: none"><li>• Análisis de sendero y su aplicación en la investigación sociológica. Supuestos de la técnica.</li><li>• Especificación del modelo.</li><li>• Evaluación del modelo de medición y de la capacidad confirmatoria del modelo.</li><li>• Ejemplo de análisis de sendero.</li></ul>
<b>7. Modelos de Ecuaciones estructurales</b>	<ul style="list-style-type: none"><li>• Análisis de ecuaciones estructurales y su aplicación en la investigación sociológica.</li><li>• La obtención de un modelo de ecuaciones estructurales: la estructura del modelo. Supuestos de la técnica.</li><li>• Estimación del modelo empírico.</li><li>• Evaluación del modelo.</li><li>• Introducción de modificaciones y/o conclusión del análisis. Evaluación de la capacidad confirmatoria del modelo.</li><li>• Ejemplos sobre modelamiento de ecuaciones estructurales.</li></ul>

# BIBLIOGRAFÍA

1	ANALISIS ESTADISTICO MULTIVARIANTE - UN ENFOQUE TEORICO Y PRACTICO	DE LA GARZA GARCIA, JORGE	2013
2	Análisis de datos multivariantes.	Peña, Daniel	2002
3	Análisis Multivariado Aplicado	URIEL Y ALDAS	2005, 1ª. Edición
4	Análisis Multivariante para las Ciencias Sociales	LEVI J.P. y VARELA J.;	2001
5	Modelos de Ecuaciones Estructurales Cuadernos de estadística	Batista,J.M. y Coenders G.,	2012,
6	Introducción al Análisis de Regresión Lineal	Montgomery, Peck y Vining	2006,
7	El análisis factorial como técnica de investigación en Psicología	Ferrando, P. J., & Anguiano-Carrasco, C.	2010
8	Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud.	. Batista-Foguet, J.M., Coenders, G., & Alonso, J.	2004
9	El Path Analysis: conceptos básicos y ejemplos de aplicación.	Pérez, E., Medrano, L. A.,& Rosas, J. S	2013
10	Modelos de ecuaciones estructurales	Ruiz, M.A., Pardo, A., & San Martín, R.	2010
11	RStudio para Estadística Descriptiva en Ciencias Sociales. Manual de apoyo docente para la asignatura Estadística Descriptiva	Boccardo, G. y Ruiz, F.	2019
12	R para Ciencia de Datos <a href="https://es.r4ds.hadley.nz/index.html">https://es.r4ds.hadley.nz/index.html</a>	Wickham, H	2019



# EVALUACIONES

Evaluación	Fechas	Porcentaje
Evaluación de R	25 de marzo 15 de abril 27 de mayo 17 de junio	30%
Prueba 1	29 de abril	35%
Trabajo de investigación	1 de julio	35%

# AYUDANTÍAS

El curso tiene dos ayudantes:

Patrio Alarcón [patricio.alarcon@mail.udp.cl](mailto:patricio.alarcon@mail.udp.cl)

Fernanda Hurtado [fernanda.hurtado@mail.udp.cl](mailto:fernanda.hurtado@mail.udp.cl)

Están disponibles para responder las dudas que puedan tener a lo largo del curso, tanto estadísticas como de uso de software.

Habrà sesiones de ayudantía cada 2 semanas aproximadamente, centradas en la aplicación de las técnicas que revisaremos en R.

También les acompañarán en la realización de tareas y trabajos de investigación.

# PÁGINA DEL CURSO

<https://aadii2024.netlify.app/>



Organización ▾ Clases Prácticos Recursos

## Análisis avanzado de datos II

Introducción al procesamiento, análisis y visualización  
para **Ciencias Sociales**

Profesor Gabriel Sotomayor • I° Semestre 2024  
Sociología • Universidad Diego Portales



# DELEGADO DE CURSO

Las comunicaciones del curso con el equipo docente para temas colectivos deberán gestionarse de manera centralizada mediante un delegado, especialmente considerando que hay alumnos de distintas generaciones.

Esto es particularmente relevante para solicitudes respecto a evaluaciones.

# OBJETIVOS DE LA SESIÓN

Reflexionar sobre el uso de software estadístico en ciencias sociales.

Introducir el uso de R y Rstudio para ciencias sociales.

# ¿DONDE SE SITÚA LA ESTADÍSTICA Y EL USO DE SOFTWARE EN EL PROCESO DE INVESTIGACIÓN?



El uso de software y estadística es una pequeña parte del proceso de diseño de una investigación, por lo que es importante no confundir ambos procesos ni anteponer el uso de determinadas herramientas al proceso general de investigación, debe primar el problema a investigar

# USO DE SOFTWARE EN CIENCIAS SOCIALES

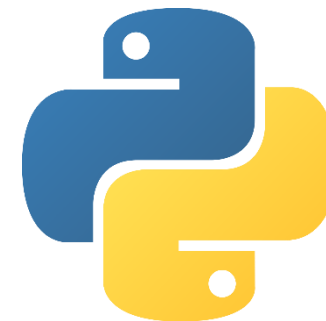
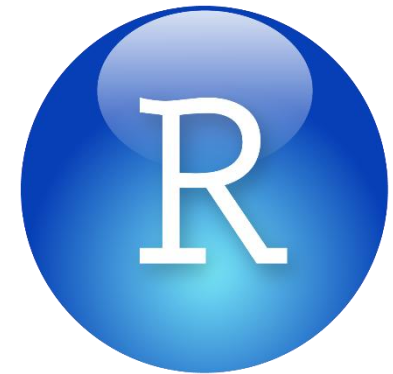
Progresivamente se ha generalizado el uso de software estadístico en ciencias sociales, abriendo grandes posibilidades de realizar análisis más complejos y facilitando su uso.

Riesgo de falta de formación estadística y reflexividad (calidad del análisis)

Dos niveles de manejo del software:

1. Ejecución básica y correcta interpretación de las salidas estadísticas (necesario mínimo para utilizar la herramienta e interpretar otras investigaciones).
2. Manejo avanzado de software y procedimientos estadísticos. Nivel deseable para un **uso reflexivo** de las herramientas y lectura crítica de otras investigaciones.

# ¿QUÉ SOFTWARE EXISTEN Y CUÁL UTILIZAR?





# ¿QUÉ SOFTWARE EXISTEN Y CUÁL UTILIZAR?

Dimensión / Lenguaje	R	Python	SPSS	Excel	Stata
Alcance	General, orientación multidisciplinar	General, orientación multidisciplinar	Limitado, orientado a Ciencias Sociales	Limitado, orientado a administración	Limitado, orientado a Economía
Licencia	Libre (freeware)	Libre (freeware)	Pagada (versión de prueba limitada)	Pagada (versión de prueba limitada)	Pagada (versión de prueba limitada)
Aprendizaje	Sintaxis, poco intuitivo	Sintaxis, poco intuitivo	Botones y sintaxis, intuitivo	Botones y sintaxis, intuitivo	Botones y sintaxis, intuitivo
Visualización	Avanzada	Intermedia	Básica	Intermedia	Intermedia
Análisis de texto	Intermedio, poca eficiencia	Avanzado, amplia eficiencia	No	No	No
Minería Datos	Intermedio, poca eficiencia	Avanzado, amplia eficiencia	No	No	No
Sistema operativo	Windows, Mac OS, Linux	Windows, Mac OS, Linux	Windows, Mac OS	Windows, Mac OS	Windows, Mac OS

(Boccardo y Ruiz, 2018)

# ¿QUÉ ES R?

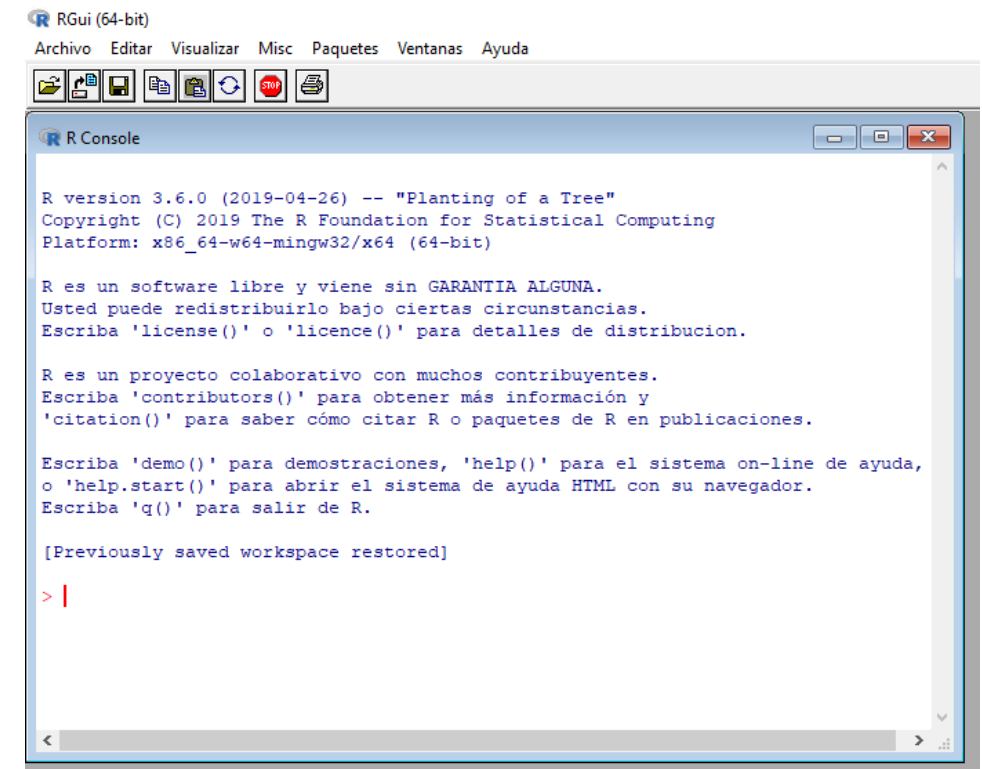
**R** es un software y un **lenguaje de programación**, ambos de carácter libre, enfocado principalmente en el análisis y visualización de datos. Es desarrollado por *The R Foundation for Statistical Computing*.

Funciona con una **lógica de objetos**, sobre los que podemos trabajar con **operadores y funciones**.

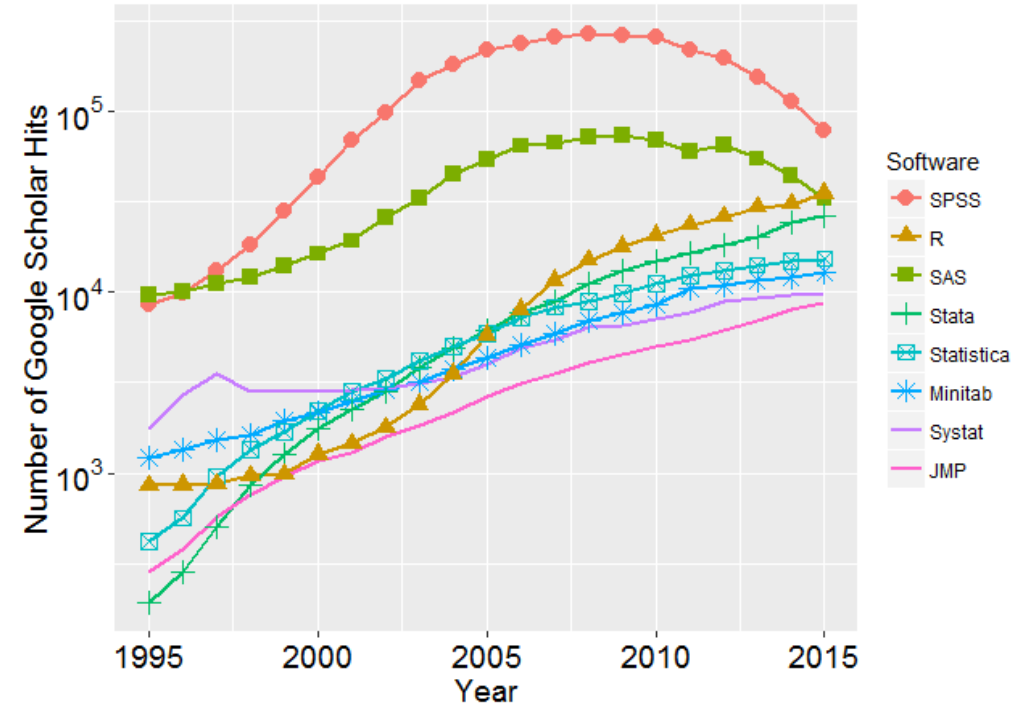
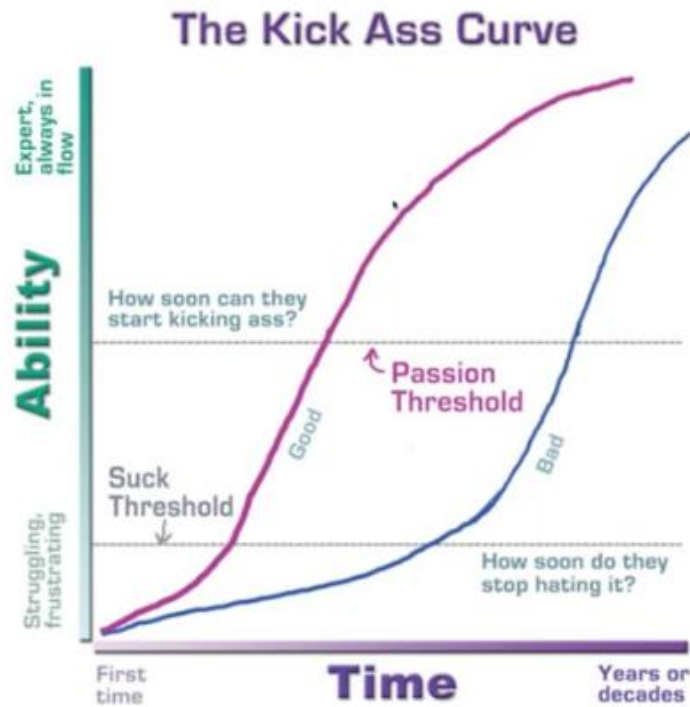
El hecho de que sea un lenguaje de programación nos permite **programar funciones**, por lo que es posible pasar de ser usuarios a programadores.

Instalamos una versión básica y podemos agregar paquetes con nuevas funciones.

Su interfaz es poco amigable, la solución: R Studio.



# ¿PORQUÉ APRENDER Y USAR R Y RSTUDIO?



## ¿PORQUÉ USAR R?

1. Es libre y gratis.
2. Por lo anterior ofrece grandes posibilidades de personalización de los análisis y amplitud de herramientas (en actualización).
3. Amplia y creciente comunidad de usuarios desarrollando funciones y resolviendo problemas en internet (la mayoría en inglés).
4. Su sintaxis es simple e intuitiva.

## ¿PORQUÉ TRABAJAR CON SINTAXIS?

- 1. Replicabilidad:** Permite a otros entender como construimos nuestros resultados y replicarlos, cada vez más importante en la ciencia.
- 2. Eficiencia:** En condiciones “reales” de trabajo resulta más eficiente que trabajar con botones (aunque ahora pueda parecer contraintuitivo)
- 3. Control:** Permite tener claridad de todas las etapas de nuestros análisis, cuestión que en otros software son realizadas por defecto.

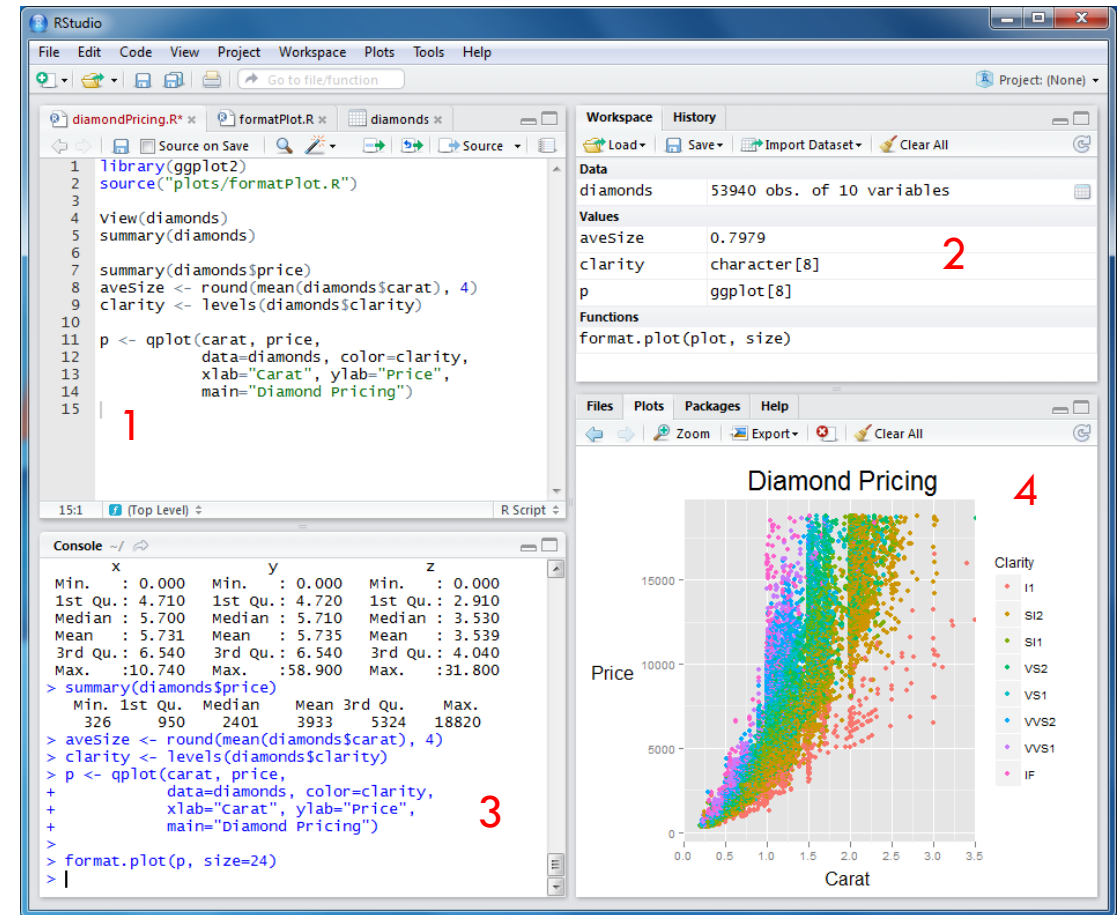
(Elousa, 2009 en Boccardo y Ruiz, 2018: 8-9)

# R STUDIO

Es una interfaz gráfica para R que nos permite tener mejor visualización y control de los procesos que realizamos.

Tiene 4 partes principales:

1. Editor de Sintaxis: formato de block de notas donde podemos construir nuestra sintaxis.
2. Ambiente de trabajo (Environment): muestra los objetos que hemos creado (memoria temporal – RAM)
3. Consola de R: Aquí es donde podemos ver los resultados de nuestros análisis, las operaciones ejecutadas y mensajes de error.
4. Panel de visualización: Aquí aparecen los gráficos que ejecutemos, además de tener la pestaña help y nuestro directorio de trabajo.

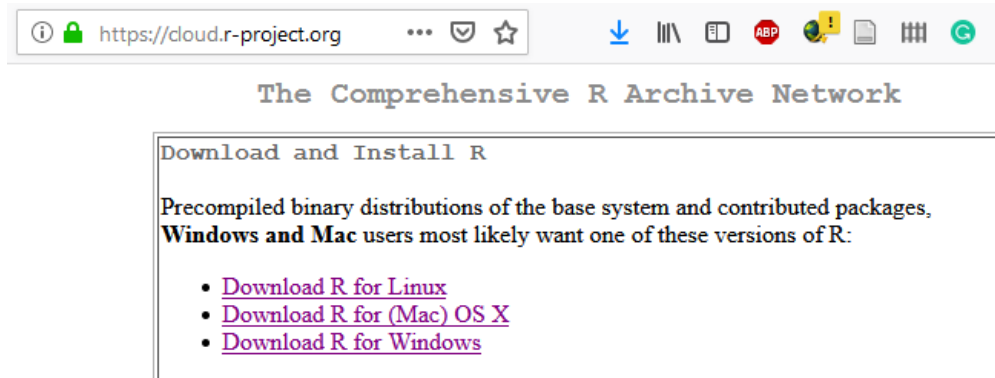


# INSTALAR R Y RSTUDIO

## Instalar R

Podemos encontrar R para Windows, Linux y Mac en <https://cloud.r-project.org/>

Debemos descargar el paquete base, luego ejecutar el instalador, e instalarlo según las opciones predeterminadas

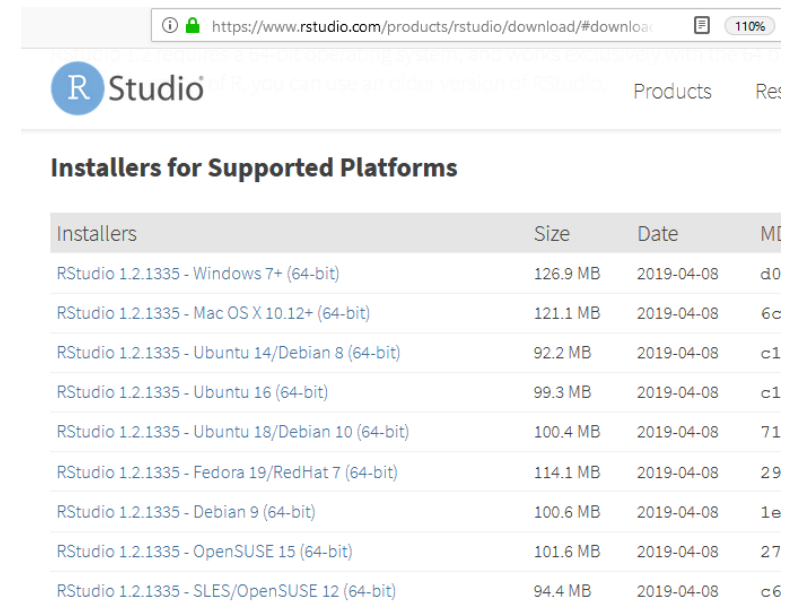


Podemos encontrar R Studio en su sitio web, y bajar la versión gratis, que contiene todas las funciones

<https://www.rstudio.com/products/rstudio/download/>

Debemos fijarnos en elegir el sistema operativo correcto.

Igualmente lo ejecutamos e instalamos según las opciones predeterminadas

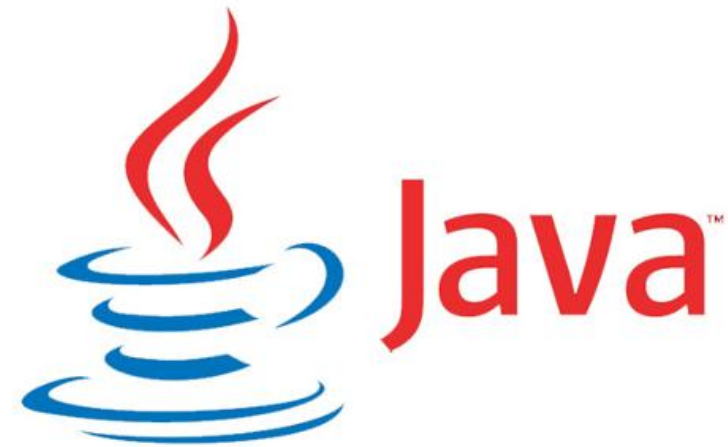


# EL PROBLEMA DE JAVA X64

Para algunos análisis R se basa en Java, el problema es que nuestros computadores suelen instalar de manera automática la versión de 32 bits, porque los navegadores de internet funcionan (o funcionaban) en esta arquitectura, pero nuestro computadores y R funcionan a 64 bits.

Para resolverlo debemos descargar e instalar la versión offline de 64 bits de Java

<https://www.java.com/es/download/manual.jsp>



# PROGRAMACIÓN POR OBJETOS Y FUNCIONES

## Tipos de estructuras de datos en R

Vector (vector): columna o fila de datos de un mismo tipo (una variable individual)

Listas (list): Nos permiten agrupar objetos de distinto tipo.

Matrices (matrix): arreglo de dos dimensiones de datos de un mismo tipo (conjunto de variables)

Data.frame (base de datos): Matriz de datos en el que las columnas tienen asignado nombres, y que permite usar todo tipo de datos.

## Tipos de variables en R

Numérico (numeric; números con decimales)

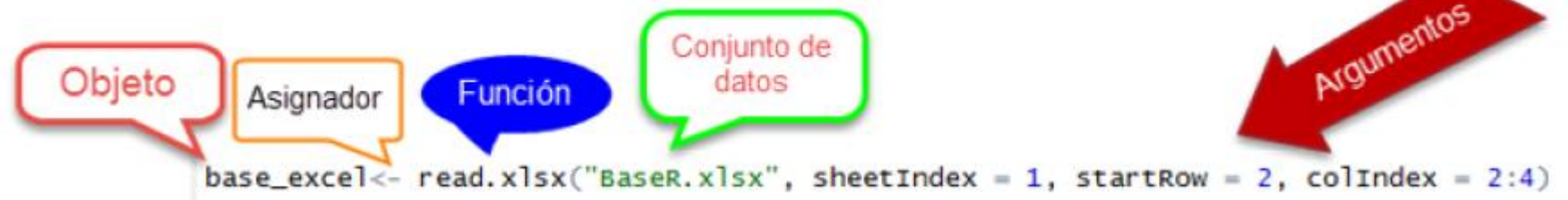
Números enteros (integer)

Lógicos: Verdadero o Falso

Variables cadena o texto (string / character)

Factores (factor): para variables cualitativas (nominales)

Estructura básica de la sintaxis de R





# ¿COMO CONTINUAR?

Dominar R es un camino de mediano plazo y es importante desarrollar habilidad de “hacking”, es decir:

1. estar dispuesto a buscar respuestas de manera autónoma
2. saber donde buscar la información
3. utilizar nuevos datos y paquetes por su cuenta

Donde buscar:

- **GOOGLE!** (preferentemente en inglés) y youtube
- Pedir ayuda a otros usuarios
- Pestaña “help”
- <https://www.rdocumentation.org/>
- <https://stackoverflow.com/>
- Chat GPT\*

Además hay múltiples recursos web para aprender R de manera profunda y guiada.

Cuestiones importantes que no revisamos: visualización de datos con ggplot2, el tidyverse (paquete dplyr), y profundizar en programación en R (loops y funciones), programación literada (Rmarkdown).

# BIBLIOGRAFÍA Y OTROS RECURSOS PARA PROFUNDIZAR

Boccardo, Giorgio y Ruiz, Felipe. *Uso de RStudio para Estadística Univariada en Ciencias Sociales. Manual de apoyo docente para la asignatura Estadística Descriptiva*. Santiago: Departamento de Sociología, Facultad de Ciencias Sociales, Universidad de Chile. [Disponible aquí](#).

Grolemund, Garrett y Wickham Hadley. *R para Ciencia de Datos*. <https://es.r4ds.hadley.nz/>

Paradis, E. (2003). *R para Principiantes*. Francia: Institut des Sciences de l'Évolution. Disponible en: [https://cran.r-project.org/doc/contrib/rdebuts\\_es.pdf](https://cran.r-project.org/doc/contrib/rdebuts_es.pdf)

## Sitios web útiles

<http://www.sthda.com>

## Cursos online (inglés)

Especialización en Ciencia de Datos Universidad John Hopkins  
<https://www.coursera.org/specializations/jhu-data-science>

Data Camp <https://www.datacamp.com/courses/tech:r>