

Asiste al:

Taller de Introducción a 'R' y Análisis de datos

Impartido por Paola y Nancy Manzo Romero

Todos los martes y jueves de 11:00 a 13:00 hrs.

MODALIDAD VIRTUAL · 16 SESIONES

Inicio: 16 de agosto 2022

Término: 6 de octubre 2022



Material requerido: laptop y programas
(previamente instalados) **R y RStudio**

*Si tu equipo es de la SISEMH, por favor indicarlo en la
liga de registro para solicitar la instalación de los
programas necesarios.*

Regístrate en la siguiente liga:

bit.ly/3QhMmGj

Fecha límite de registro: viernes 12 de agosto 2022

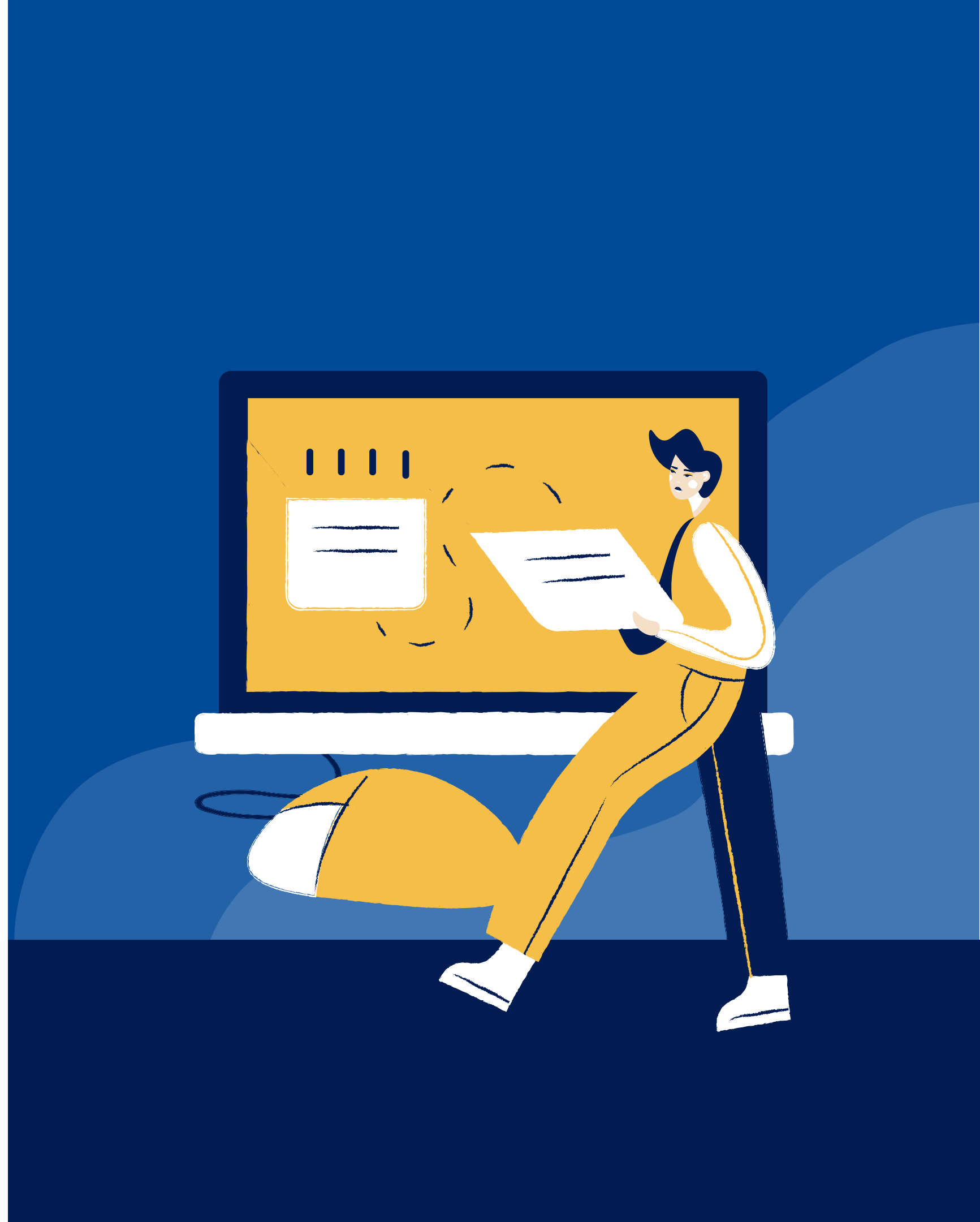
Taller introducción a R y análisis de datos

Agosto 2022

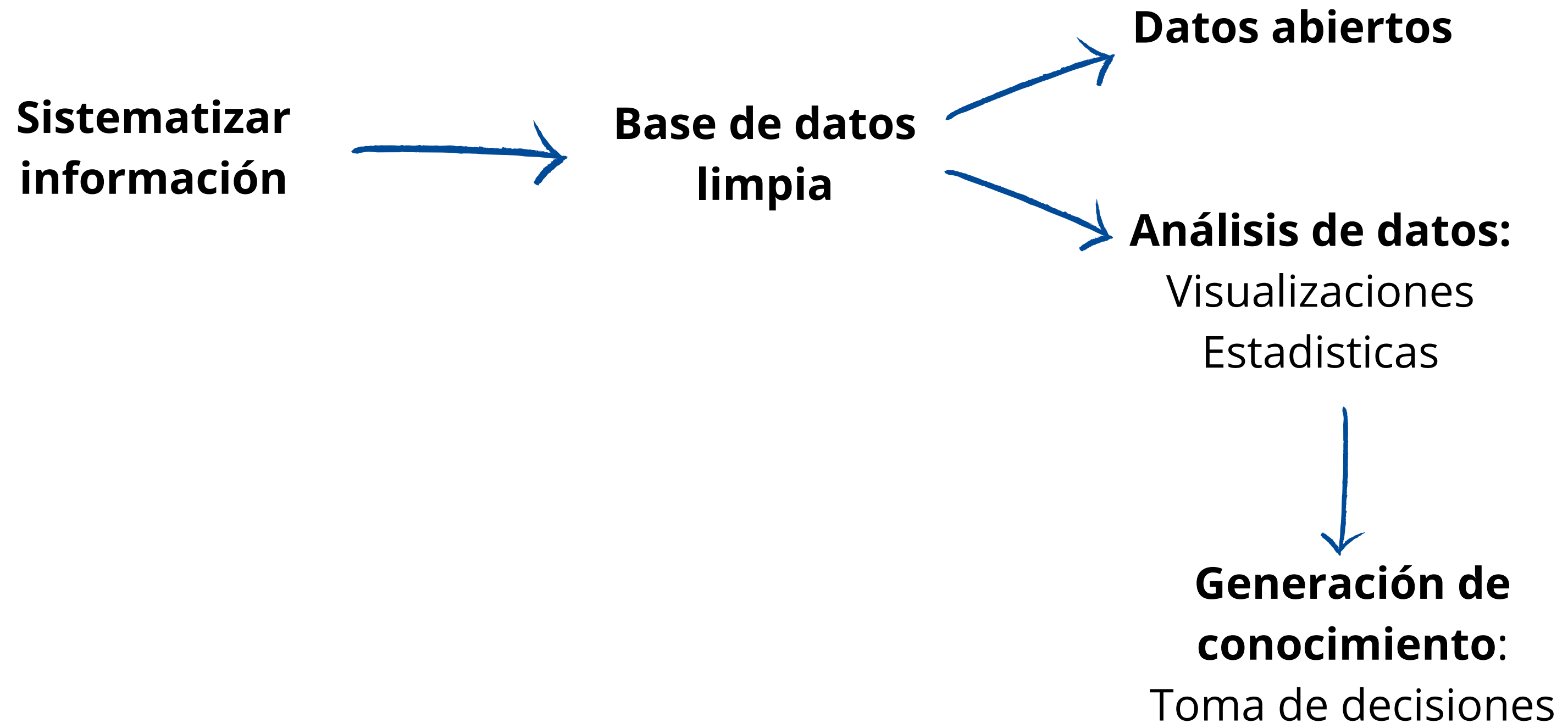
Sesion 1: Introducción general al taller

Hoy

1. Bienvenida
2. Introducción a la sistematización de datos
3. Valor en los datos
4. ¿Qué es R?
5. Instalación de R y R studio



Ciclo del dato



Breve introducción a la sistematización

La sistematización de información se refiere al **ordenamiento** y **clasificación** de todo tipo de datos e información, bajo determinados criterios, categorías, relaciones, etc



Sistematización de datos

La sistematización es la captura de un **registro administrativo**, un caso u observación que se almacena, se integra y se suele actualizar.

Por lo general se captura en **hojas de excel** y se almacenan en la nube.

El formato tradicional de una sistematización es un formato de bases de datos donde se integra por **columnas y filas**, siendo la primera fila el **encabezado**.

Mejoras en la sistematización de datos

Homologar tus datos

- Mayúsculas
- Sí acentos
- Formatos (dar fecha, texto, hora, moneda, etc.)
- Valores nulos/ vacíos / blancos: NA
- Encabezado en la primera fila
- Respuestas o información cerrada/categorica

Principales preguntas que deben responder mis datos:

¿Sobre quién?

¿Para quién?

¿Sirviendo a qué intereses?

¿Con qué valores?

Y sobre todo recordar que así como la sociedad evoluciona y no es neutral, también aplica en los datos.



Valor en los datos

Más allá de almacenar la información, la importancia de los datos **para la toma de decisiones**, en la generación de conocimiento.

Además,

Seguimiento y control de los procesos.

Extracción de información

Responder a necesidades de la ciudadanía.

Caso práctico

Asunto: Limpieza de base de datos de Interrupción legal del embarazo

Contexto:

4 años de datos (2016-2020)

Más de 70,000.
campos abiertos

Consideraciones:

La interpretación debe de considerar la totalidad de datos. Por ejemplo, si en la mayoría ponen “sí” y “no” el 1 será interpretado como sí y 0 como no.

En caso de controversia, podemos acudir con la Secretaría de Salud o el diccionario de datos.

El objetivo es conservar la mayor cantidad de datos vivos.

<https://datos.cdmx.gob.mx/dataset/interrupcion-legal-del-embarazo>

R y R studio

¿Por qué R studio?

Es un software libre

Fácil uso

Integración con sheets y otros programas

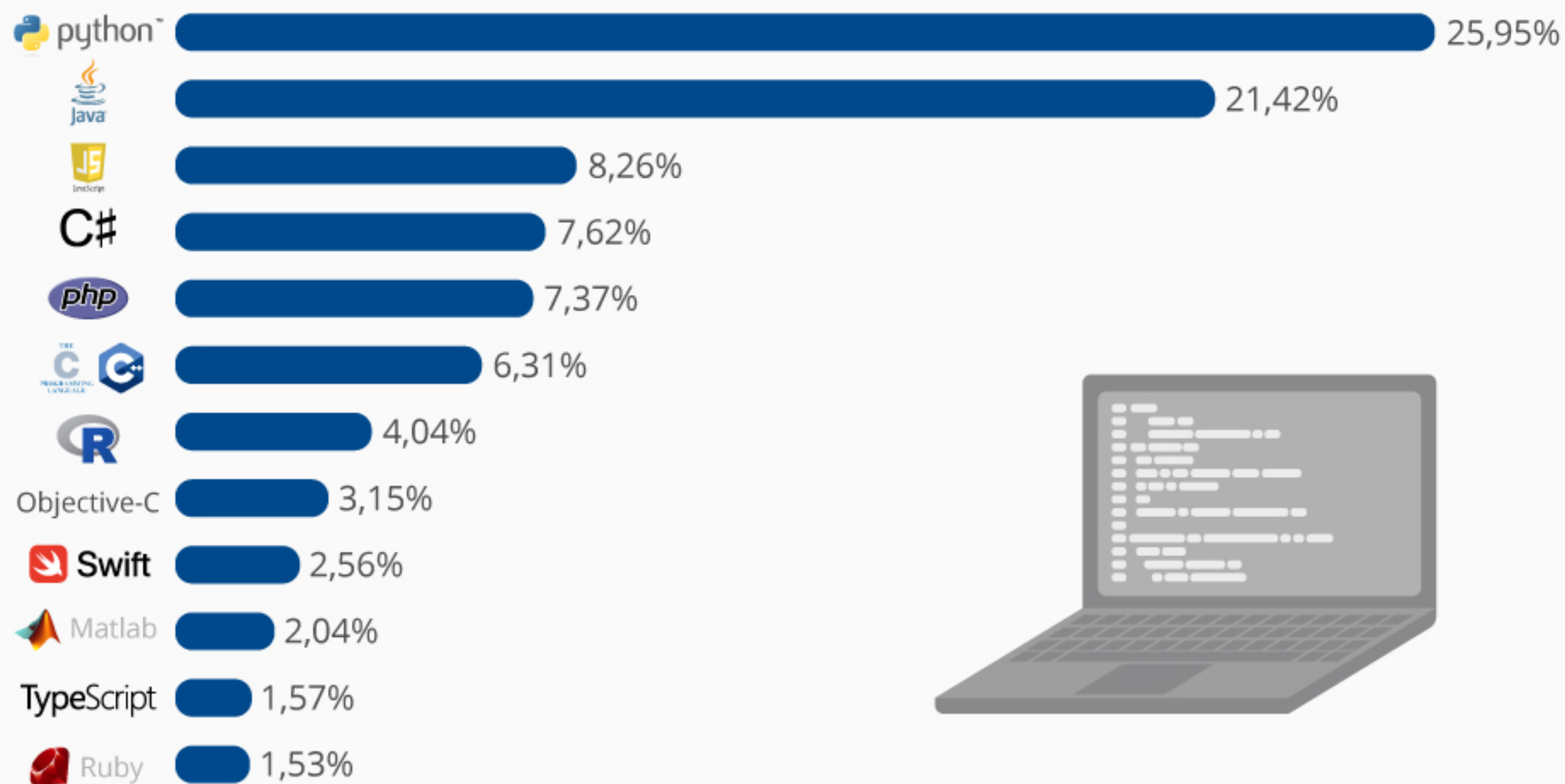
Manipulación y limpieza de datos

Generación de reporte visualizaciones y mucho más.



Los lenguajes de programación más usados

Porcentaje de uso de los lenguajes de programación más populares del mundo*



Instalación de R y Rstudio

- Deben descargar los dos archivos:
 - i. Instalar primero R (se dejan por default las opciones)
 - ii. Instalar R studio

Se muestra la instalación en la pantalla.