



Técnicas de procesamiento y estructuración de la información

[DATA REPÚBLICA]

Margarita Vaca | *Cepei*

Marzo de 2019



movistar



Data



Information



Presentation



Knowledge



Contexto

Estructurado

Datos perfectamente definidos y sujetos a un formato específico que no permite ambigüedad
Nombre, fecha, edad...



Tipo de formato

Tipo de archivo en que esta siendo presentado el dato:
JPG, PDF, Excel

No estructurado

Datos variados y no sujetos a un formato específico
Fotos, vídeos, Publicaciones, correos electrónicos

Datos

Unidad mínima de recolección

Cualitativo

Asociados a categorías, cualidades, contextos, preferencias.
¿Por qué?

Tipo de categoría

Clases de variables utilizadas para la medición del dato

Cuantitativos

Expresados por medio de valores numéricos que pueden ser medidos, contados o ponderados
¿Qué?, ¿Dónde?, ¿Cuándo?

Contexto



Definición de ruta



Definición de temáticas

- Educación básica por departamento - Colombia



Periodo de análisis

- Anual: 2015-2017



Tipos de datos según formatos

- Datos estructurados - Excel



Definición de variables

- Año, departamento, deserción, aprobación, reprobación, repitencia, nivel educativo



Tipos de variables

Cuantitativas - Cualitativa



Fuentes de datos

Datos Abiertos

Estadísticas en educación básica por departamento



Procesos para el análisis de variables

Series de tiempo – No se requiere codificación de datos cualitativos

Recolección de Datos

Técnicas de recolección de datos

- **Entrevistas:** estructurada, semiestructurada y no estructurada
- **Cuestionarios y encuestas:** resultados numéricos y de perspectiva u opinión
- **Observaciones:** detalle de fenómenos o comportamientos --> Patrones
- **Grupos focales:** comparación de puntos de vista
- **Registros, documentos y literatura:** encontrar, analizar y seleccionar
- **Etnografía:** Comprender las dinámicas sociales de un grupo determinado
- **Técnica Delphi:** grupo de expertos, como un todo, tratan un problema complejo.

● Cuantitativa

● Cualitativa

● Mixta



movistar



Estructuración

Limpieza de datos



Organizar información por división de caracteres



Eliminar datos duplicados



Revisión de ortografía y tipo de letra



Quitar espacios y caracteres no imprimibles



Verificar separador de miles y porcentajes (punto y coma)



Corregir números y signos de número



Combinar y dividir celdas



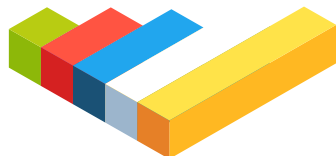
Conciliar datos de tablas (buscarv)



Completar vacíos

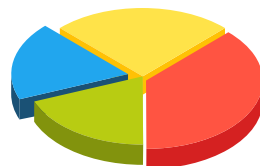


Estructuración



Clasificación:

Variables transversales –
variables complementarias



Organización de valores:

Es fundamental establecer una
secuencia de valores lógica que
no interfiera en cálculos futuros



Etiquetas:

Asignación de nombres
cortos que sean de fácil
interpretación y asociación

Base maestra:

recopilación de información a través de una
base de datos que se encuentre vinculada entre
hojas o fichas de trabajo del archivo

Procesamiento

Codificación

Crear una identidad cuantificable a variables cualitativas

- Encuestas de satisfacción
- Características de los encuestados

Conformar categorías

- Preguntas abiertas que deben ser reorganizadas en categorías que permitan su interpretación.

Detectar valores anómalos o faltantes

- La asignación de un código a valores atípicos o faltantes cuando estos no pueden ser eliminados de la base de datos.

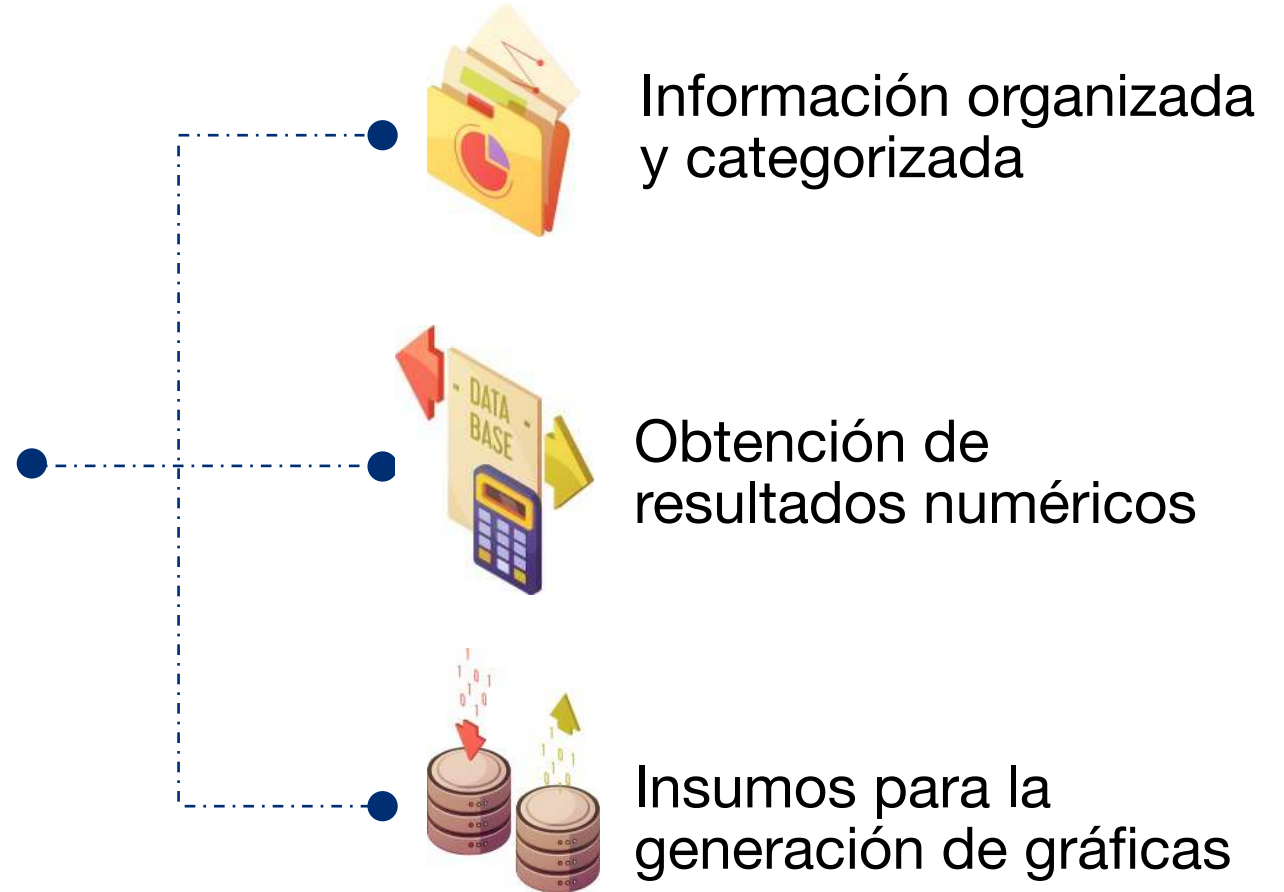
Definir perfil de las variables cuantitativas

- La asignación de un código a valores atípicos o faltantes cuando estos no pueden ser eliminados de la base de datos.

Procesamiento

Tabulación:

determina qué resultados de variables se presentarán y qué relaciones entre esas variables son necesarias para el análisis.



Procesamiento

Indicadores

Es una relación de dos o más variables que describen características, comportamientos o fenómenos de la realidad que cumple con tres características: simplificación, medición y comunicación.

Objetivos

-  Generar información útil para mejorar el proceso de toma de decisiones, diseño, implementación o evaluación
-  Monitorear el cumplimiento de acuerdos y compromisos
-  Cuantificar cambios de una situación problemática
-  Efectuar seguimiento a los planes, programas y proyectos.

Procesamiento

Tipos de indicadores



Procesamiento



Formulación del problema

- Identificación del objeto de medición.
- Definición de aspecto específico.



Definición de las variables

- Unidad de medida
- Variables principales y secundarias.
- Fuente de producción de variables



Selección de indicadores y calidad de los datos

- Características del proyecto, datos, relación con el problema de análisis y utilidad para el usuario.



Diseño del indicador

- Identificación de contexto
- Determinación de usos y actores
- Identificación de fuentes de información y procedimientos
- Definición de responsabilidades
- Documento del indicador

Procesamiento



Formulación del problema

- Identificación del objeto de medición

- Definición de aspecto específico



Definición de las variables

- Unidad de medida
- Variables principales y secundarias.
- Fuente de producción de variables



Selección de indicadores y calidad de los datos

Características del proyecto, datos, relación con el problema de análisis y utilidad para el usuario.



- Identificación de contexto
- Determinación de usos y actores
- Identificación de fuentes de información y procedimientos
- Definición de responsabilidades
- Documento del indicador

The logo for Data Living LAB is positioned on the left side of the slide. It features the words "Data", "Living", and "LAB" stacked vertically in a white, sans-serif font. To the left of the text is a graphic element consisting of several vertical lines of varying heights, some ending in small circles, resembling a stylized signal or data visualization. The background of the left half of the slide is a dark blue isometric illustration of a city skyline with various buildings and a wind turbine.

**Data
Living
LAB**

¡MUCHAS GRACIAS!

DATA REPÚBLICA



movistar

