RESUMEN – Tidy Data, Capítulo 2: Defining tidy data

Todos los conjuntos de datos son iguales, en cambio los conjuntos de datos desordenados están desordenados a su manera, es decir, todos son diferentes.

Los datos ordenados vinculan sus datos (diseño físico) con su semántica (significado), la mayoría de los conjuntos de datos estadísticos son tablas rectangulares formadas por filas y columnas. Las columnas casi siempre están etiquetadas y las filas a veces están etiquetadas.

Un conjunto de datos es una colección de valores, generalmente números (si son cuantitativos) o cadenas (si son cualitativos). Los valores se pueden organizar de dos formas. Todo valor pertenece a una variable y a una observación. Una variable contiene todos los valores que miden el mismo atributo subyacente (como altura, temperatura, duración) en todas las unidades. Una observación contiene todos los valores medidos en la misma unidad (como una persona, un día o una carrera) en todos los atributos.

Los datos ordenados son una forma estándar de asignar el significado de un conjunto de datos a su estructura. Un conjunto de datos es desordenado u ordenado según como se combinen las filas, columnas y tablas con las observaciones, las variables y los tipos.

Los datos serás ordenados si: cada variable forma una columna, cada observación forma una fila y cada tipo de unidad de observación hace una tabla. Los datos ordenados facilitan que un analista o una computadora extraigan las variables necesarias porque proporcionan una forma estándar de estructurar un conjunto de datos. Las variables fijas deben ir primero, seguidas de las variables medidas, cada una ordenada de modo que las variables relacionadas sean contiguas. Luego, las filas se pueden ordenar por la primera variable, rompiendo los lazos con las variables.