

Manifiesto: Manejo y Exploración de datos

Carlos Iván Espinosa

3 de agosto de 2017

Prefacio

Durante los últimos años en múltiples áreas del conocimiento se habla mucho sobre datos, supongo que la persona que está leyendo este documento habrá oído mencionar el término “Big Data”. Como se imaginarán una de las ventajas de vivir en un mundo globalizado y con acceso a tecnologías es el poder conectar con miles de personas que hacen investigación o que sencillamente recogen datos para tomar decisiones, esto a su vez implica el acceso a una cantidad enorme de datos y por tanto de posibilidades de generar nueva información. Todas estas posibilidades ha hecho que los investigadores pongan en importancia mucho más que antes a los *datos*.

Los datos representan una parte fundamental en el desarrollo de la ciencia, por muchos años los investigadores se acostumbraron a mantener los datos con una estructura que solo ellos entendían, es posible que años atrás eso era lo más importante, sin embargo, hoy en día estos paradigmas están cambiando, cada vez es más importante que los datos sean guardados con formatos estandarizados y que puedan ser leídos por cualquier investigador. El presente curso trata sobre algunos de los puntos básicos para organizar y manejar datos. ¿Qué son los datos? ¿Cómo debo organizar mis datos? ¿Por qué y cómo debo limpiar mis datos?. Estas son algunas de las preguntas que intentaremos responder durante este curso.

Metodología del curso

El presente curso se realizará de una forma teórica-práctica en la cual se muestre la importancia del manejo de los datos y los procedimientos específicos para desarrollar algunas acciones específicas del manejo y exploración de datos. Durante el proceso del curso se propondrá varios ejercicios para procesar y analizar datos. En una primera fase del curso les pediré levantar (o inventarse) unos datos muy sencillos para que los pueda trabajar.

Con los datos levantados usaremos Microsoft Excel para organizarlos y explorarlos. Los ejercicios tienen por objetivo mostrarle como debe organizar los datos y como poder generar un procedimiento para revisar y graficar los datos. Al final del curso podrá ver como los datos se convierten en información. Les animo a intentar realizar la encuesta



Figure 1: Colectando datos de vegetación en el Bosque Seco. Mi ayudante preferido mi hijo

efectivamente así los resultados que obtengan le serán interesantes.

Recuerde repasar todos los ejercicios con el fin de afianzar los conocimientos.

Objetivo

El desarrollo de este curso permitirá a los estudiantes generar conocimientos y principalmente aptitudes de manejo y exploración de datos. Con el curso espero puedan adquirir conocimientos prácticos sobre el uso de herramientas en Excel que faciliten el procesamiento de sus datos.

¿Quién es su profesor?

Mi nombre es Carlos Iván Espinosa soy biólogo de profesión y tengo un Doctorado en Ciencias. Durante los últimos años mi labor como investigador me ha permitido tener contacto con diferentes personas que usan datos; estudiantes, profesionales e investigadores. Esta experiencia de más de 10 años me ha permitido entender las principales limitantes del manejo de datos.



Figure 2: En alguna de mis salidas de campo en Galapagos-Ecuador

Contenidos

1. Estructura de los datos y metadata
 - 1.1 Los datos y la información
 - 1.2 Los datos brutos
 - 1.3 La estructura de los datos
 - 1.4 La metadata un paso fundamental
 - 1.5 Unos consejos prácticos
2. Manipulación de datos y generación de variables
 - 2.1 Generación de variables
 - 2.2 Herramientas de organización y manejo de datos
 - 2.3 Reformar los datos
 - 2.4 Limpieza de variables
3. Representación gráfica: Los principios de una buena gráfica
 - 3.1 La visualización de datos
 - 3.2 Anatomía de una gráfica
 - 3.3 Gráficos de una variable
 - 3.4 Gráficos de dos variables