Tesis de MSc

Aguilera-González, J.F.

13 de abril de 2017

### Directorio

Para determinar la ubicación del directorio de trabajo y su contenido es necesario:

1. Emplear la función getwd()
2. Visualizar los archivos con dir()

### Base de datos

La Base de Datos se almacenó como archivo .xlsx (Microsoft Excel 2013) con el nombre "BD.xlsx".

* El archivo se compone de **tres hojas** de cálculo:
  + "BDT" *Datos Crudos*
  + "BDA" *Datos Hidrométricos e Hidrológicos*
  + "OUT" *Datos Atípicos*
* Los encabezados de las **columnas** corresponden a las **variables**
* Cada **observación** se almacenó en una **fila**
* La primer columna corresponde a la identificación **ID** de cada **observación**

Para visualizar la base de datos se requiere lo siguiente:

1. Con el **paquete** readxl importamos la Base de Datos al entorno de trabajo.
2. Creamos el **objeto** dbT <- que almacena la Base de Datos "BD.xlsx", especifcando la Hoja con sheet = "BDT"

### Estructura, dimensión y clase

Para determinar las características del objeto dbT <- empleamos:

* Estructura: str(dbT)
* Dimensión: dim(dbT)
* Clase: class(dbT)

### Recodificación

Variables almacenadas como caracteres chr las recodificamos a factor creando un nuevo objeto df <- con la función data.table

## Análisis Exploratorio de Datos (**EDA**)

