

Fundamentos de la programación estadística y Data Mining en R

Unidad 1. Programación Estadística en R: Introducción a tidyverse

Dr. Germán Rosati (CONICET - UNSAM - UNTREF)

06 marzo, 2019

Las cinco instrucciones básicas de dplyr

- Elegir observaciones en función a sus valores: `filter()`.
- Reordenar las filas `arrange()`
- Seleccionar variables de acuerdo a su nombre: `select()`
- Crear variables nuevas en función de variables ya existentes: `mutate()`
- Colapsar valores a un solo resumen: `summarise()`
- Todas pueden ser usadas en conjunto con `group_by()` que cambia el “scope” de cada función: en lugar de operar sobre todo el dataset, lo hace al interior de cada grupo definido.

Las cinco instrucciones básicas de dplyr

- Estas 6 funciones proveen los verbos para un “lenguaje” de manipulación de datos.
 - Todas funcionan de forma parecida
- 1 El primer argumento es un dataframe
 - 2 Los siguientes argumentos describen qué hacer con el data frame, usando los nombres de variables sin comillas
 - 3 El resultado es un nuevo dataframe

Filtrando filas con `filter()`

- Permite subconjuntar observaciones basados en sus valores.
 - * El primer argumento es el dataframe
 - * Los siguientes, las expresiones filtran el dataset