Curso Ciencia de datos en R
Escuela de Doctorado, Universidad de Alcalá, Mayo-Junio 2021
Prof. Sara Villén Pérez

1) DATA MANIPULATION

0. EXAMPLE DATASETS

data(), data(database), ?database, data(package = .packages(all.available = TRUE))

1. DATA TYPE AND STRUCTURE

c(), factor(), ordered(), matrix(), array(), data.frame(), tibble(), list()
raster(), stack(), brick(), SpatialPoints(), SpatialPointsDataFrame()
is.numeric(), is.double(), is.integer(), is.logical()
is.vector(), is.factor(), is.matrix(), is.array(), is.data.frame(), is.list(), is_tibble()
as.numeric(), as.double(), as.integer(), factor(), ordered(), as.matrix(), as.array(), as.data.frame(), as_tibble(), as.list()

2. DATA EXPLORATION

View() , head(), tail()
nrow(), ncol(), dim(), length(), nchar()
names(), colnames(), rownames()
class(), str()
summary(), table()
unique(), duplicated(), which()
attributes()
levels()
hist(), plot()
match()

3. DATA SUBSETTING

database[lines, columns]
datalist[[position]]
dabase$column
==, !=, >, <, >=, <=, |, &
subset()
dplyr::slice()
dplyr::select()
dplyr::filter()

4. DATA EDITING

colnames(), rownames()
levels(), recode(), cut()
/, *, +, -, ^, dplyr::mutate()
is.na(), complete.cases(), na.omit(), mean(…, na.rm=T)
attributes()

5. DATA REORGANIZATION

rev(), t()
sort() # vector
reorder() # factor levels
order(), dplyr::arrange() # dataframe
sample()
cbind(), rbind()
merge()

6. TIDY | RESHAPE DATA

gather(), spread(), separate(), unite()

7. AGGREGATE AND ANALYZE DATA

rowMeans(), colMeans(), rowSums(), colSums()
summarise(), group_by(), ungroup()
apply(), lapply() , sapply(), vapply(), tapply(), rapply(), mapply()

8. CONCATENATE OPERATIONS

Parenthesis, Intermediate objects, Overwriting, Pipes %>%

9. SPECIFIC DATA MANIPULATION

9.1. FACTORS {forcats}

fct_count(), fct_relevel(), fct_infreq(), fct_reorder()

9.2. CHARACTER STRINGS {stringr}

str_detect(), str_which(), str_locate(), str_count()
writeLines("")
\n, \t , \\ , \"

9.3. DATES AND TIMES {lubridate}

as_datetime(), as_date(), hms::as.hms()
OlsonNames(), with_tz(), force_tz()
now(), today()
ymd_hms(), ymd_hm(), ymd_h(), ymd(), hms::hms()
ymd(), mdy(), dmy(), date_decimal()
date(), year(), month(), day(), hour(), minute(), second()
wday(), week(), quarter(), semester(), am(), pm(), dst(), leap_year()
round_date(), floor_date(), ceiling_date()
seconds(), minutes(), hours(), days(), months(), years()
dseconds(), dminutes(), dhours(), ddays(), dmonths(), dyears()
interval()

10. DATA IMPORTATION AND EXPORTATION

read_delim(), read_excel(), read_table(), readRDS(), load()
write_delim(), saveRDS(), save(), save.image()