

Preprocessing

Denisa Mensatorisová a Adam Štuller

V tejto časti načítavame a prespracujeme vybrané stĺpce. Niektoré stĺpce sú zložené z viacerých vnorených stĺpcov, z ktorých nás zaujíma iba konkrétna hodnota merania. Ďalej chýbajúce hodnoty sú uvedené napr. ako 999 alebo 9999, takéto merania sme identifikovali a nahradili na jednotnú hodnotu - NA.

Stĺpce týkajúce sa zrážok - LP a LP24 majú inú štruktúru preto sú spracovávané samostatne počas prieskumnej analýzy a testovaní hypotéz.

Loading of documents (years 1973 - 2020)

```
selected_cols <- c('STATION', 'DATE', 'TMP', 'WND_ANGLE', 'WND_SPEED', 'VIS', 'SLP', 'CIG', 'DEW', 'SNOW', 'LP', 'LP24')

read.csv(file= "../data/all_data.csv") %>%
  dplyr::filter(STATION == 11903099999 | STATION == 11927599999 ) %>%
  dplyr::mutate(
    year = ymd_hms DATE) %>%
    lubridate::year() %>%
    map_chr(~ as.character(.x))
  ) %>%
  dplyr::filter(year >= 1973 & year <2021) %>%
  separate(TMP, c('TMP', NA), sep=',') %>%
  separate(WND, c('WND_ANGLE', NA, NA, 'WND_SPEED', NA)) %>%
  separate(VIS, c('VIS', NA, NA, NA)) %>%
  separate(SLP, c('SLP', NA)) %>%
  separate(CIG, c('CIG', NA, NA, NA)) %>%
  separate(DEW, c('DEW', NA), sep=',') %>%
  separate(AJ1, c('SNOW_DEPTH', NA, NA, NA, NA, NA, NA)) %>%
  separate(AY1, c('PWO', NA, NA, NA)) %>%
  separate(IA1, c('GSO', NA)) %>%
  dplyr::mutate(
    TMP = map_dbl(.$TMP, process_col, 10),
    TMP = na_if(TMP, 999.9),
    WND_ANGLE = map_dbl(.$WND_ANGLE, process_col),
    WND_SPEED = map_dbl(.$WND_SPEED, process_col, 10),
    VIS = map_dbl(.$VIS, process_col),
    SLP = map_dbl(.$SLP, process_col, 10),
    CIG = map_dbl(.$CIG, process_col),
    DEW = map_dbl(.$DEW, process_col, 10),
    SNOW_DEPTH = map_dbl(.$SNOW_DEPTH, process_col),
    PWO = dplyr::na_if(PWO, ""),
    GSO = dplyr::na_if(GSO, "")
  ) %>%
  dplyr::select(all_of(selected_cols)) %>%
  write.csv(file="../data/all.csv")
```