

## Data Loading and wrangling

```
library(rvest)

## Loading required package: xml2

library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.1.0      v dplyr  1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()      masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()         masks stats::lag()
## x purrr::pluck()       masks rvest::pluck()

library(readxl)
library(janitor)

##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test

library(stringr)
library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union

library(glue)

##
## Attaching package: 'glue'

## The following object is masked from 'package:dplyr':
##
##   collapse

Data scraping wikipedia

link <- read_html("https://en.wikipedia.org/wiki/List_of_dams_and_reservoirs_in_Maharashtra")
daminfo <- html_node(link, 'table') %>%
  html_table(fill = TRUE)
```

loading 2015 data

```
dam2015 <- read_excel("Data/2015.xlsx") %>%
  clean_names() %>%
  pivot_longer(cols = c(-reservoir_name, -district, -basin),
    names_to = "month") %>%
  mutate(day = 1) %>%
  separate(month, into = c("month", "year"), sep = "_") %>%
  mutate(month = case_when(
    month == "jan" ~ 1,
    month == "feb" ~ 2,
    month == "mar" ~ 3,
    month == "apr" ~ 4,
    month == "may" ~ 5,
    month == "jun" ~ 6,
    month == "jul" ~ 7,
    month == "aug" ~ 8,
    month == "sep" ~ 9,
    month == "oct" ~ 10,
    month == "nov" ~ 11,
    month == "dec" ~ 12),
    date = make_date(year, month, day)) %>%
  select(reservoir_name, district, basin, date, value, month, day, year)

daminfo <- daminfo %>%
  clean_names() %>%
  mutate(name_of_dam = case_when(
    name_of_dam == "Koyana" ~ "Koyana/Shivaji Sagar",
    name_of_dam == "Hatnur" ~ "Upper TapiHatnur Reservoir",
    name_of_dam == "Isapur" ~ "Isapur Reservoir",
    name_of_dam == "Upper Vaitarana" ~ "Upper Vaitarana Reservoir",
    name_of_dam == "Khadakwasla" ~ "Khadakwasla Reservoir",
    name_of_dam == "Yeldari" ~ "Yeldari Reservoir",
    name_of_dam == "Mula" ~ "Mula Reservoir",
    name_of_dam == "Jayakwadi" ~ "JayakwadiNath Sagar",
    name_of_dam == "Girna" ~ "Girna Reservoir",
    name_of_dam == "Mulshi" ~ "Mulshi Dam",
    name_of_dam == "Kanher" ~ "Kanher Dam",
    name_of_dam == "Ujani" ~ "BhimaUjjani Reservoir",
    TRUE ~ name_of_dam
  ) )

dam2015fin <- left_join(dam2015, daminfo, by = c("reservoir_name" = "name_of_dam"))
```