Exercise Sheet 5:
Working with DWD Climate Data in R

Overview

In this exercise sheet you will work with daily climate data from the German Weather Service (DWD). You will learn how to download the data, load it into R, inspect the structure of the dataset, interpret each variable, and perform simple exploratory analyses.

All tasks are to be done in a new `.Rmd` file.

Exercise 1: Create an R Markdown File and Access DWD Data

Instructions

Generate a new **R Markdown** file. At the top of your `.Rmd`, load the required packages:

```
library(rdwd)
library(dplyr)
library(readr)
```

You may obtain the DWD climate data in one of the following two ways:

Option A: Automatic Download via the `rdwd` Package (Recommended)

1. Install `rdwd` (if not already installed):

```
install.packages("rdwd")
```

2. Use the DWD meta index to search for a station. For example, search for station 1420:

```
meta <- selectDWD(id = 1420, res = "daily", var = "kl", period = "recent")
meta
```

3. Download the file:

```
file <- dataDWD(meta, read = FALSE)
```

4. Read the dataset:

```
df <- readDWD(file)
```

This yields a data frame containing daily climate variables from the DWD station.

Option B: Manual Download from DWD Open Data

1. Go to `https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/`
2. Navigate to `recent/` for the newest values or `historical/` for older data.
3. Select a station folder, e.g. `tageswerte_KL_1420_akt/`
4. Download the ZIP file, e.g. `tageswerte_KL_1420_akt.zip`
5. Unzip the file manually.
6. Load the data in your `.Rmd` file:

```
df <- read_csv("produkt_klima_tag_20240501_20240531_1420.csv")
```

(Your file name may differ.)

Inspecting the Structure

Now inspect the structure of the dataset:

```
str(df)
```

Your output should look similar to:

```
$ STATIONS_ID: int 1420 1420 1420 ...
$ MESS_DATUM : Date, format: "2024-05-26" ...
$ QN_3       : int 10 10 10 ...
$ FX         : num 10.5 12.4 9.7 ...
$ FM         : num 2.9 2.9 3.6 ...
$ QN_4       : int 9 9 9 ...
$ RSK        : num 0 3.1 1.1 ...
$ RSKF       : int 6 6 6 ...
$ SDK        : num 10.6 4.1 9.3 ...
$ SHK_TAG    : int 0 0 0 ...
$ NM         : num 5.6 7 4.8 ...
$ VPM        : num 13.7 14.5 10.7 ...
$ PM         : num 1005 1005 1007 ...
$ TMK        : num 17.9 17.1 15.6 ...
$ UPM        : num 70 76 62 ...
$ TXK        : num 24.3 22.9 20 ...
$ TNK        : num 10.9 12.9 10 ...
$ TGK        : num 8.6 11 7.6 ...
$ eor        : Factor w/ 1 level "eor": 1 1 1 ...
```

Exercise 2: Interpret Each Variable

In your `.Rmd` file, create a table or bullet list explaining the meaning of each field:

- **STATIONS_ID**: Station identifier
- **MESS_DATUM**: Date of measurement
- **QN_3**: Quality level for precipitation-related variables
- **FX**: Maximum wind gust (m/s)
- **FM**: Mean wind speed (m/s)
- **QN_4**: Quality code for temperature, humidity, radiation
- **RSK**: Daily precipitation amount (mm)
- **RSKF**: Precipitation form (0=no precipitation, 6=rain, etc.)
- **SDK**: Sunshine duration (hours)
- **SHK_TAG**: Snow depth (cm)
- **NM**: Cloud cover
- **VPM**: Water vapor pressure (hPa)
- **PM**: Air pressure (hPa)
- **TMK**: Mean daily air temperature (°C)
- **UPM**: Relative humidity (%)
- **TXK**: Daily maximum temperature (°C)
- **TNK**: Daily minimum temperature (°C)
- **TGK**: Minimum ground-level temperature (°C)
- **eor**: End-of-record marker (can be removed)

Exercise 3: Basic Cleaning

Add the following steps to your `.Rmd`:

```
df <- df %>% select(-eor)        # remove end-of-record flag
df <- df %>% drop_na()           # remove missing entries
df$MESS_DATUM <- as.Date(df$MESS_DATUM)
```

Exercise 4: Exploratory Plots

(a) Plot the Temperature Time Series

(b) Plot the Precipitation Time Series

(c) Plot Sunshine Hours Distribution

(d) Do your Own Extended Exploratory Analysis

Exercise 5: Summary Statistics

Compute summary values

Discuss:

- What is the typical temperature?
- How often does rain occur?
- How variable is sunshine?
- How strong are maximum wind gusts?

Exercise 6: Optional Challenges

(a) Identify the Hottest Day

(b) Identify the Maximum Wind Day

(c) Investigate the Relationship Between Sunshine and Temperature