Exploring the observations and forecast data for a location

For the "location" assigned to your dyad, use the forecast data from ECMWF with a forecast horizon of 36 hours contained in `selected_location36_2009-2022.rds`. The file contains several forecast parameters for which acronyms, description and units are given in `ECMWF_selected_parameters.pdf`.

Perform the following tasks and document your code and results in a commonly used format and additionally as pdf and upload them to the `task` subfolder on OLAT.
In the following task descriptions `ibk` is used as an example location.

1. Load the package `zoo` using `library(zoo)` at the beginning of your R code. The data are in a "zoo-object", a special form of a data frame that allows to handle time series data more easily.
   To make sure that you are not running into problems with different time zones between the data set and the timezone set on your computer, tell R that the data are in UTC with `Sys.setenv('TZ'='GMT')`.

Use the `readRDS()` function to read your data into a variable called `ibk` (or the name of your location). Make sure you set the directory to the correct location of the data file. `getwd()` in R tells you in which directory you are and `setwd()` lets you change it (`wd` stands for "working directory").
A little extra info about zoo-objects: You get just the time/date of data points with `index()` and the data without the time/data with `coredata()`.
To see the column names (= acronyms of all forecast variables), use `colnames(ibk)` or simply with `names(ibk)`.
An easy way to see what the data look like is with `head(ibk)`, which prints the first few times of the data and `tail(ibk)` which prints the last lines.

2. Convert the forecasted temperatures `t2m` and `t850` from Kelvin to Celsius. One way to select rows and columns in a data frame in R is with square brackets: `[]`. For example, to select all rows (= all times here) of `ibk` for the variables `t2m` and `t850`, use `ibk[,c("t2m","t850")]`. By not specifying anything before the comma, we select all rows. The `c()` concatenates. Since we are using variable names, i.e. characters, they need to be in quotes. With this way of selecting variables, the unit conversion can be done in one single line of code for both variables.

3. Use the `View(ibk)` (or similarly the data view in RStudio) to get a tabular overview of the data.

4. Use `summary(ibk)`. What are the extreme values for `cape` and `ssshf6h` (with units!) and why is the third quantile, i.e. the 75percentile, of `sshf6h` still negative and only the maximum value positive? Hint: read the definition of that variable and consider the time of day.
   Note that you can address individual variables with `$`, e.g. `ibk$cape` to get just CAPE. If you do not want to have the summary of all variables printed, you can therefore just use `summary(ibk$cape)`, or `summary(ibk[,c("cape","sshf6h")])` to get a summary for both using the bracket-method.

5. Use `pairs()` to explore relationship among the variables. What happens when you use the option `cex = 0.1` in the `pairs` function? Select 2 variable pairs for which the relationship is linear and 2 for which it is clearly nonlinear (but not just a seemingly arbitrary cloud of points) and give an explanation for the shape of these 4 relationships.

6. Use `plot()` to plot a time series of the forecasted temperature `t2m` at Innsbruck. In which year where the coldest and warmest temperature, respectively forecast? Use `identify()` to obtain the values of these 2 extreme temperatures directly from the graphic. Then use the `max()` and `min()` functions to compute them.

7. Make a scatterplot for the forecasted 2m-temperature `t2m` and 850 hPa temperature `t850`. Try to explain its shape and peculiarities.

8. Plot a histogram for each variable using the function `hist()`. Are there any that have hard limiting value(s)?

9. Do `t2m` and `t850` follow a Gaussian distribution? Use a quantile-quantile (QQ) plot to show that graphically. `qqnorm(ibk$t2m)` with the diagonal line plotted by `qline(ibk$t2m)` will do that. Also plot the probability density functions of both variables (using the x-values and y-values of `density(ibk$t2m)`) and on top a Gaussian function of the same mean and standard deviation. Compare them to the interpretation about differences in dispersion and in length/heaviness of tails that you can draw from the QQ-plot.

10. Filter ("subset") data from the summer months (JJA) and winter months (DJF) and do a `summary()` for both of them. Are there any for which the median values are approximately the same in both seasons?