**Workflow for data analysis**

**Use Master Script to source all the other scripts in correct order!**

* Read in data and create list
  + Separate scripts for separate datasets
* Set all loggers to same start time (same interval)
  + Interpolate data to 1 min intervals and
  + Reduce again to 10 min values
* Use calculated offset (reference thermometer) to correct data
  + Use offset from test in the lab and subtract the offset from temperature data
  + Offset for logger 33 is missing -> check that
* Decompose the time series (plot seasonality, trend and noise)
* Tidy up data regarding spikes
  + Set a threshold for a rise in temperature that are regarded as spikes and therefore set to NA
    - For water: threshold of 2.5°C/10min, remove 2h of data
    - For Air: threshold of 5°C/10min, remove 30mins of data
* Plots
  + plot the tidy data in pairs (water, settlement, vegetation)
  + plot overviews for all water logger/ vegetation logger/ sealed area logger
* Split data into day and night datasets (creates separate lists)
  + Two hours per day for dawn are removed
  + Works for every dataset
* Plot the day and night datasets and save to file
  + Add the sunrise and sunset as vertical lines to the plots
  + Add description and type of location to plots

**To Do:**

* Create (working heatmap) -> check out other possibilities
* Get daily means for the data (and add to plot?)
* Check where NAs have gone
* Try and puzzle matching time frames together
* Check time differences between sealed/veg data (different Logger starttime)
  + Use time series functions to prove that besides seasonality the temperature is linear to time (Prove that linearity can be assumed and therefore interpolation is correct) 🡪 R squared value?
  + Do spikes go missing when interpolating?

**Statistics:**

* Significance tests for difference between sealed areas and vegetation