HOBO Pendant Data Cleanup Walkthrough

Kara Dobson

July 13, 2020

This doc will describe the steps taken to clean the L0 HOBO data in the HOBO_pendantdata_cleanup_L0.R script. The cleanup script is located in the warmXtrophic Github repository.

Before the cleanup begins, the existing data is cleared and the graphic devices are closed. The functions are then read in from the HOBO_functions.R script in the Github repository, and the working directory is set to the shared warmXtrophic Google drive. Finally, the needed packages are loaded in.

```r
# Clear all existing data
rm(list=ls())

# Close graphics devices
graphics.off()

# Source functions
source("~/warmXtrophic/kara/HOBO_functions.R")

# Set working directory to Google Drive
# **** Update with the path to your Google drive on your computer
setwd("/Volumes/GoogleDrive/Shared drives/SpaCE_Lab_warmXtrophic/data/")

# Load packages
for (package in c("tidyverse", "weathermetrics")) {
  if (!require(package, character.only=T, quietly=T)) {
    install.packages("package")
    library(package, character.only=T)
  }
}
```

The csv files for each sensor, each year are then read in from the shared Google drive.

```r
### ***KBS*** ###
# Read in KBS HOBO data from all years
pend4P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_4P_09012017.csv", skip=1, header =T)
pend5P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_5P_09012017.csv", skip=1, header =T)
pend6P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_6P_09012017.csv", skip=1, header =T)
pend7P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_7P_09012017.csv", skip=1, header =T)
pend8P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_8P_09012017.csv", skip=1, header =T)
pend9P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_9P_09012017.csv", skip=1, header =T)
pend10P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_10P_09012017.csv", skip=1, header =T)
pend11P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_11P_09012017.csv", skip=1, header =T)
pend12P_17k<-read.csv("L0/KBS/sensor_data/2017/09_01_2017/KBS_12P_09012017.csv", skip=1, header =T)
```

```r
pend4P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_4P_09202018.csv", skip=1, heade
pend5P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_5P_09202018.csv", skip=1, heade
pend6P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_6P_09202018.csv", skip=1, heade
pend7P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_7P_09202018.csv", skip=1, heade
pend8P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_8P_09202018.csv", skip=1, heade
pend10P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_10P_09202018.csv", skip=1, head
pend11P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_11P_09202018.csv", skip=1, head
pend12P_18k<-read.csv("L0/KBS/sensor_data/2018/09_20_2018 (pendants)/KBS_12P_09202018.csv", skip=1, head

pend4P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBS_4P_09232019.csv", skip=1, header =T)[ ,1:4]
pend5P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBS_5P_09232019.csv", skip=1, header =T)[ ,1:4]
pend6P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBS_6P_09232019.csv", skip=1, header =T)[ ,1:4]
pend7P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBs_7P_09232019.csv", skip=1, header =T)[ ,1:4]
pend8P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBS_8P_09232019.csv", skip=1, header =T)[ ,1:4]
pend10P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBS_10P_09232019.csv", skip=1, header =T)[ ,1
pend11P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBS_11P_09232019.csv", skip=1, header =T)[ ,1
pend12P_19k<-read.csv("L0/KBS/sensor_data/2019/09_23_2019/KBS_12P_09232019.csv", skip=1, header =T)[ ,1

pend4P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_4P_04052020.csv", skip=1, header=T)[ ,1:4]
pend5P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_5P_04052020.csv", skip=1, header=T)[ ,1:4]
pend6P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_6P_04052020.csv", skip=1, header=T)[ ,1:4]
pend7P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_7P_04052020.csv", skip=1, header=T)[ ,1:4]
pend8P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_8P_04052020.csv", skip=1, header=T)[ ,1:4]
pend10P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_10P_04052020.csv", skip=1, header=T)[ ,1:4
pend11P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_11P_04052020.csv", skip=1, header=T)[ ,1:4
pend12P_20k<-read.csv("L0/KBS/sensor_data/2020/04_05_2020/KBS_12P_04052020.csv", skip=1, header=T)[ ,1:4
```

Each of the files are then added to a list so that functions can be applied to every dataframe, and lapply is used to apply the needed functions to the list. The add_id_col() function is applied to the dataframes before they are added into a list.

```r
# Apply functions
add_id_col()
list_k <- list(pend4P_17k=pend4P_17k,pend5P_17k=pend5P_17k,pend6P_17k=pend6P_17k,pend7P_17k=pend7P_17k,
               pend4P_18k=pend4P_18k,pend5P_18k=pend5P_18k,pend6P_18k=pend6P_18k,pend7P_18k=pend7P_18k,
               pend4P_19k=pend4P_19k,pend5P_19k=pend5P_19k,pend6P_19k=pend6P_19k,pend7P_19k=pend7P_19k,
               pend4P_20k=pend4P_20k,pend5P_20k=pend5P_20k,pend6P_20k=pend6P_20k,pend7P_20k=pend7P_20k,
list_k <- lapply(list_k, change_pend_names)
list_k <- lapply(list_k, change_POSIX)
```

The dataframes in the list are then combined by year, resulting in a dataframe for each year that contains all of the pendants' data. These dataframes then have a column created to show what site they're from; "KBS"

```r
# Combine KBS pendant files
pend17k<-rbind(list_k$pend4P_17k,list_k$pend5P_17k,list_k$pend6P_17k,list_k$pend7P_17k,list_k$pend8P_17
pend18k<-rbind(list_k$pend4P_18k,list_k$pend5P_18k,list_k$pend6P_18k,list_k$pend7P_18k,list_k$pend8P_18
pend19k<-rbind(list_k$pend4P_19k,list_k$pend5P_19k,list_k$pend6P_19k,list_k$pend7P_19k,list_k$pend8P_19
pend20k<-rbind(list_k$pend4P_20k,list_k$pend5P_20k,list_k$pend6P_20k,list_k$pend7P_20k,list_k$pend8P_20
pend17k$Site<-"KBS"
pend18k$Site<-"KBS"
pend19k$Site<-"KBS"
pend20k$Site<-"KBS"
```

Finally, an .RData file is made that contains the yearly pendant dataframes. This .RData file is added to the L1 folder in the shared Google drive and is used in the HOBO_pendantdata_merge_L1.R script in the

Github repository, which combines the yearly pendant data into one file per site (KBS or UMBS)

```
#Create RData save file:
save(pend17k, pend18k, pend19k, pend20k, file="L1/HOBO_data/HOBO_pendant_data/KBS/KBS_HOBOpendant_L1.RD
```

The steps listed above are the same for the UMBS data, with the exception of the following code that manually changes some column names for the 10P UMBS sensor.

```
#Manually change 10p column names (they don't match the names of the others)
names(pend10P_17u)[names(pend10P_17u)=="Temp...F..LGR.S.N..10747441..SEN.S.N..10747441..LBL..B6_air_war
names(pend10P_17u)[names(pend10P_17u)=="Intensity..lum.ft...LGR.S.N..10747441..SEN.S.N..10747441..LBL..
names(pend10P_18u)[names(pend10P_18u)=="Temp...F..LGR.S.N..10747441..SEN.S.N..10747441..LBL..B6_air_war
names(pend10P_18u)[names(pend10P_18u)=="Intensity..lum.ft...LGR.S.N..10747441..SEN.S.N..10747441..LBL..
names(pend10P_19u)[names(pend10P_19u)=="Temp...F..LGR.S.N..10747441..SEN.S.N..10747441..LBL..B6_air_war
names(pend10P_19u)[names(pend10P_19u)=="Intensity..lum.ft...LGR.S.N..10747441..SEN.S.N..10747441..LBL..
names(pend10P_20u)[names(pend10P_20u)=="Temp...C..LGR.S.N..10747441..SEN.S.N..10747441..LBL..B6_air_war
names(pend10P_20u)[names(pend10P_20u)=="Intensity..Lux..LGR.S.N..10747441..SEN.S.N..10747441..LBL..B6_l
```