Basic Linux and Coding for AA (BLAC) Exercise 2 (week 1)

Author:
Timo Halbesma, 6126561

Supervisor:
Dr. T. Coenen

Listing 1: TLRH's solution for the BLAC homework 2

```bash
#!/bin/bash
# blac_hw2.sh

# Shell script for Basic Linux and Coding for AA homework 2 (week 1).
# Usage:
# TLR Halbesma, 6126561, september 7, 2014. Version 1.0; implemented

set -o errexit

# Exercise 1. I have downloaded the following dataset:
# From 1950 January 1 untill 2000 January 1, All elements, All stations
# This dataset is stored in KNMI_20000101.txt
dataFile="./KNMI_20000101.txt"
if [ -f "$dataFile" ]; then
    totalLines=$(wc -l $dataFile | awk '{print $1}')
    echo "$dataFile has $totalLines linenumbers in total." # 344874
else
    echo "The datafile is not present"
    exit 1
fi

# Exercise 2. The header is predeceded by a #; no other lines have a #.
headerLines=$(grep -c "#" "$dataFile")
echo "$dataFile has $headerLines linenumbers in the header." # 85
dataLines=$(($totalLines - $headerLines))
echo "$dataFile has $dataLines linenumbers as data entries." # 344789

# Exericse 4.
hoogeveenEntries=$(grep -c "^[[:space:]]*279" "$dataFile")
echo "$dataFile has $hoogeveenEntries Hoogeveen entries in the dataset." # 4018
# NB Hoogeveen station id is 279. Do note that some entries are empty!

# Exercise 5.
# Hupsel temperature on August 4, 1980 not in dataset. Use Schiphol instead?
stationID=$(grep "SCHIPHOL" KNMI_20000101.txt | cut -c 3-5) # 240
schipholAugust41980=$(grep "^[[:space:]]*$stationID,19800804" "$dataFile")

# TN is the minimum temperature; TX is the maximum temperature. See KNMI header.
# TN is the 12th column; TX is the 14th column. Note $0 is the entire string.
minTemp=$(echo $schipholAugust41980 | awk 'BEGIN { FS = "," } ; { print $13}' )
# Temperatures in dataset are per 0.1 degrees centigrade. Bash arithmetic does
# not support floating points.
# https://stackoverflow.com/questions/24093798/how-to-divide-variable-by-10
minTemp=$(echo "scale=1; $minTemp/10" | bc)
echo "The minimum temperature at Schiphol on August 4, 1980 was $minTemp deg. C"
maxTemp=$(echo $schipholAugust41980 | awk 'BEGIN { FS = "," } ; { print $15}' )
maxTemp=$(echo "scale=1; $maxTemp/10" | bc)
echo "The maximum temperature at Schiphol on August 4, 1980 was $maxTemp deg. C"

# Exercise 6
# From header: "RH = Daily precipitation amount (in 0.1 mm)"
# Take entire dataset minus header as input to sort on column 22 (RH).
precipitationMax=$(grep -v "^#" "$dataFile" | sort -k 22 | tail -n 1)
stationID=$(echo $precipitationMax | awk 'BEGIN { FS = "," } ; { print $1}')
date=$(echo $precipitationMax | awk 'BEGIN { FS = "," } ; { print $2}')
level=$(echo $precipitationMax | awk 'BEGIN { FS = "," } ; { print $23}')
level=$(echo "scale=1; $level/10" | bc)
# 23-06-1975, stationID 344 = Rotterdam, 101,4 mm.
# http://members.home.nl/tianwa/noni/journaal/extremenpagina.html consistent.

echo -n "The most precipitation fell on $date at StationID $stationID. "
echo "The precipitation level was $level mm."

# Take only Arcen (stationID=391) as input to sort on column 22 (RH).
arcenMax=$(grep "^[[:space:]]*391" "$dataFile" | sort -k 22 | tail -n 1)
stationID=$(echo $arcenMax| awk 'BEGIN { FS = "," } ; { print $1}')
date=$(echo $arcenMax | awk 'BEGIN { FS = "," } ; { print $2}') # 19930925
level=$(echo $arcenMax | awk 'BEGIN { FS = "," } ; { print $23}')
level=$(echo "scale=1; $level/10" | bc) # 58.6 mm

echo -n "The most precipitation in Arcen($stationID) fell on $date. "
echo "The precipitation level was $level mm."

exit 0 # Exit with success. Strictly not necessary though.
```