LUND UNIVERSITY

MNXB01 Introduction to programming and computing for scientists - Project Group D

Marie-Anne Bécard, Jakob Huhn, Florian Jörg, Jenny Tsin

November 15, 2019

Contents

1 Introduction

The Swedish climate is influenced by the Gulf Stream and the weather can vary quite a bit in the long stretched country. Milder winters can occur in the south, while the temperature can drop below -30 °C up in the north. In addition, February is usually the coldest month in Sweden. In historical times, people believed that the moon phase would affect the weather. When certain lunar maria were visible, good weather was predicted, while other visible maria suggested bad weather. Many questions arise due to the fact that Sweden's weather seems to be very diverse, .

The first research question studied is how the average temperature differs in the northern part, the central part, and the southern part of the country throughout a given day, over a period of roughly 50 years. It is expected that the frequency of the average temperature will be higher at a lower degree in the North and at a higher degree in the South, since the Northern part is at a higher latitude. In addition, climate change may have resulted in higher average temperatures over the years.

The second question covers what the hottest and the coldest temperature is per year in both Karlstad and Lund and compare those. It is expected that the coldest temperature is lower in Karlstad and the hottest temperature is higher in Lund.

The final question is whether there is a relationship in Lund between the average temperature per day in a moon cycle, compared to the moon phase in the year 2014 and previous years. According to the old folklore, the average temperature per day per moon cycle will follow a similar pattern as the moon phase.

2 Methods

In order to clean up the provided data and to structure it into the desired data, a bash script named cleanup.sh was used. This calculated the average temperature per day (one at a time) per year.

The master.sh program asks to the user the temperature of which city he wants to study for the first and second question. This program will run the coldestday.sh and hottestday.sh for that city and give the graph for the question 2.

For the first question, the function tempOnDay made a histogram of the average temperature in a given city on a given date and plotted the number of entries having the same value. This was done for the Northern part Luleå, the central part Karlstad and the Southern part Lund.

For the second question, the bash scripts coldestday.sh and hottestday.sh extracted the coldest and hottest temperature per year to new text files, respectively. This data was then combined and plotted in one graph, which displayed the coldest and highest temperature per year, in Karlstad and Lund. All the cpp files will automatically be compiled after every change thanks to the rootlogon.C file.

CHOICE OF PROGRAMMING LANGUAGE

To realize the project bash is used as well as C++/ROOT. A bash script is first used to analyse the data as bash is a very helpful tool to process data and restructure files. During the programming of the bash script new challenges arose, where solutions using bash code were attempted as well. This is why a lot of the code is done in bash. ROOT was mainly used to produce the graphs using the preprocessed files obtained by bash scripts.

3 PREPROCESSING THE DATA FILES

In the beginning, the given data files, containing temperatures on given days in different cities of Sweden were analyzed. Beside the date and the temperature, additional information, was included in the files as well, for example about the quality of the data. Therefore the first step was to structure and clean up the given data.

CLEANUP.SH

The cleanup of the data files was managed by a bash script called `cleanup.sh`. This files reads in the data file and generates an output file containing only the date and the corresponding temperature. The main part of this script are a `grep` command, which only greps lines in the original file, that contain at the beginning a date in the format `YYYY-MM-DD;`, as temperatures were measured at different times on a day the average of each day was taken.
The script also allows the option to take only high quality data into account. The quality is indicated by a "G" or "Y" in the original data files. Therefore, a grep command was used, to grep only high quality ("G") data. The generated output file was then used for further specific modification to answer the research questions.

4 QUESTION 1

For the first question the output file of `cleanup.sh` was directly used and loaded into ROOT scripts to plot a graph. It is useful to clean the original data file and calculate the average temperature on one day in one city.

TEMPONDAY

The implemented function to make a histogram for a given day was named `tempOnDay` inside the class `tempTrender` in the file tempTrender.cpp. The function takes two variables, the month and day chosen for the plot. The values can be chosen in the project.cpp file. However, this was later included in the master.sh, in which the user can choose the exact day the program should analyze.

In the code itself an `if`-statement handles the storage of only the temperatures of this specific month day combination in an array. In the tempTrender.cpp file, a variable was created for the cities, which was used to adapt the name of the output file when it would be saved. The city's name is found by the substring method in C++.

The cleaned file was read and the temperatures for the given month and day were stored in an array. Afterwards, this array was plotted as a histogram (in blue) with ROOT, which displayed the average temperature on the given day and the frequency in the past years. A statistics box is set to return the mean value of the temperature and the Root Mean Square.

5 Question 2

coldestday.sh and hottestday.sh

To answer the second question, the output file of the cleanup was handled with a second bash script. This bash script coldestday.sh and hottestday.sh, respectively, used an `awk` statement in order to find the lowest and hottest temperature per year by looping through the data in a `while`-loop. The syntax of both scripts is the same, with the only difference of finding once the minimum (for the coldest day per year) and once the maximum (for the hottest day per year). The generated files, coldestday_highQuality_Lund.txt and coldestday_highQuality_Lund.txt were further used to plot the graph using ROOT.

hotCold

For this question, the hotCold function was implemented in the tempTrender.cpp. The user did not have to specify an exact date, since the input of the function was the text file, containing the year and the coldest/hottest temperature.

ROOT's TGraph were used to draw histograms with values read directly from files that have previously been created with the coldestday.sh and hottestday.sh bash programs, such as the coldestday_allEntries_Lund.txt and hottestday_allEntries_Lund.txt.

multi histogram plot on root

A multi graph was chosen to display the coldest (in blue) and the hottest (in red) temperatures of the past years in a given city. To achieve this, two graphs with different colors were created with TGraph and combined with the TMultiGraph.h library, in order to show the gap between the extremes temperatures.

6 Question 3

Key features of the moon were named after weather states, such as the "Sea of Cold" and the "Sea of Clouds". This is because it was believed that the moon phase correlates with the

weather. For the last and unique task we tested this hypothesis. Although we had temperature data for 9 cities over 50 years, it is actually a small package of data to test interesting hypothesis like climate change accurately. Therefore, we decided to test an historical theory to disprove it. In this task we investigated to relationship between the moon phase and a relative temperature and finally figured out that there was no relationship to be seen.

MOONTEMP.SH

The purpose of the bash script for question 3 was to make a file, that contains the difference between the temperature (in percent relative to the highest temperature of one moon cycle/28 days) and the percent of the moon (0% equals new moon, 100% equals full moon). The script loops over 28 day cycles and assigns the relative temperature in percent per cycle using the following formula:

$$T_{rel} = \frac{T - T_{min}}{T_{max} - T_{min}} \tag{6.1}$$

where T is the temperature on a given day, T_{min} is the lowest and T_{max} the highest temperature of this cycle. In the end, the difference between the percent of the relative temperature, T_{rel}, and the moon cycle are calculated.

MOONTEMP

The function `moonTemp` was implemented. It used the difference between the relative temperature and the moon phase and plotted it as a histogram.

7 MASTER.SH

In order to make sure the program would be compatible for all cities and qualities, a master bash script was implemented, which gave the user a menu to guide one through all of the options. To do so, the user was asked for a city and quality, which were then passed to the respective scripts. The scripts were run automatically, which ensured less input error from the user. Since passing multiple bash variables to c++ files was less convenient, it was decided to create the files with the correct variables every time the master.sh script was run and to overwrite any existing versions that existed. The program read the available date files from the data_files folder and stored the outputs from ROOT in the pictures folder. The respective commands in ROOT were automatically executed using EOF.

8 RESULTS

8.1 QUESTION 1

The histograms 8.1 to 8.3 show the average temperature on November 11th. It can be seen that the mean temperature in Lund was around 5.4°C, and around -1.7°C in Luleå. The mean

temperature in Karlstad is between the previous values, nearly 3°C, which is consistent with its location in the middle of Sweden between Lund and Luleå. The temperature range on November 11th is [0;11] °C in Lund and [-16;6] °C in Luleå. It can be concluded that the weather is rougher in the north of Sweden than in the south and differs according to the years. The temperatures of Karlstad and Lund are milder compared to Lule å , although temperatures in Karlstad became as cold as -10°C.

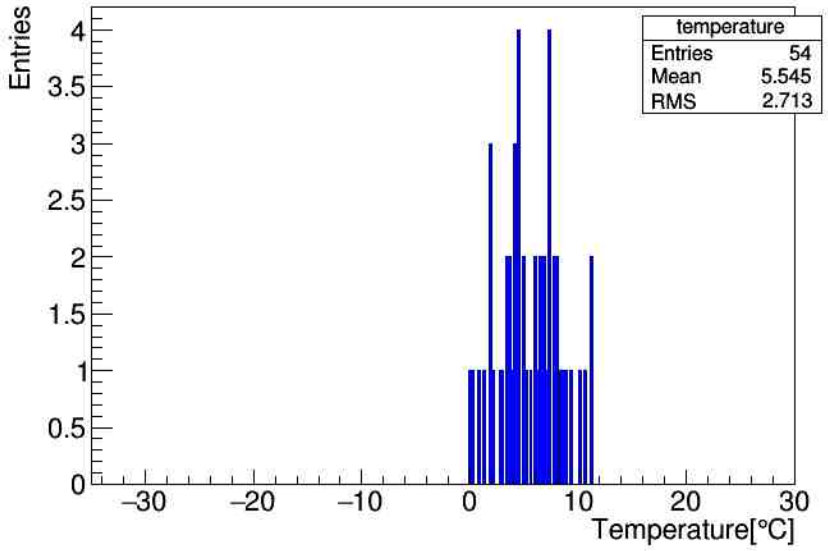

Figure 8.1: Lund with high quality on 11 November

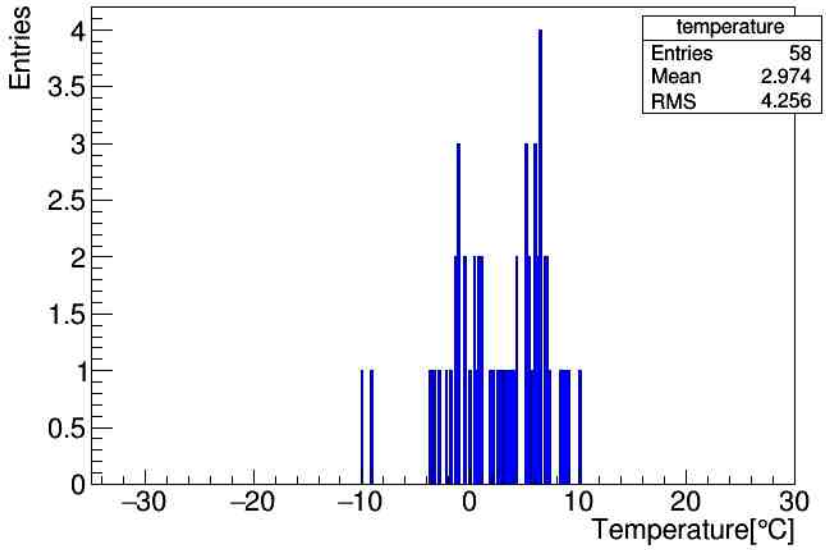

Figure 8.2: Karlstad with high quality on 11 November

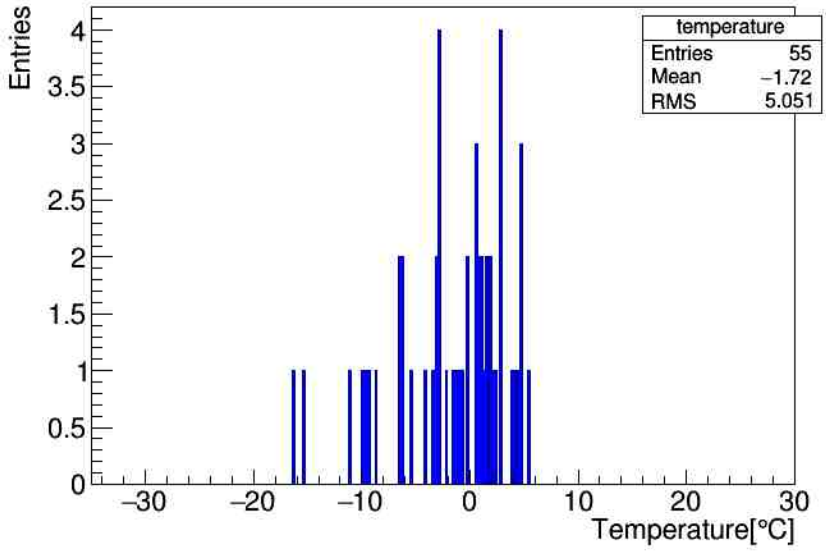

Figure 8.3: Luleå with high quality on 11 November

8.2 QUESTION 2

Figures 8.4 and 8.5 show the coldest and hottest days of each year in Lund and Karlstad, respectively. In Lund, the more southern location has hottest days between 25°C and 30°C,

whereas further north Karlstad has hottest days of only around 20°C. In contrast Karlstad's winters are much more colder, reaching as cold as -30°C, compared to the coldest winter in Lund with a minimum temperature reaching near -20°C during the past 50 years.

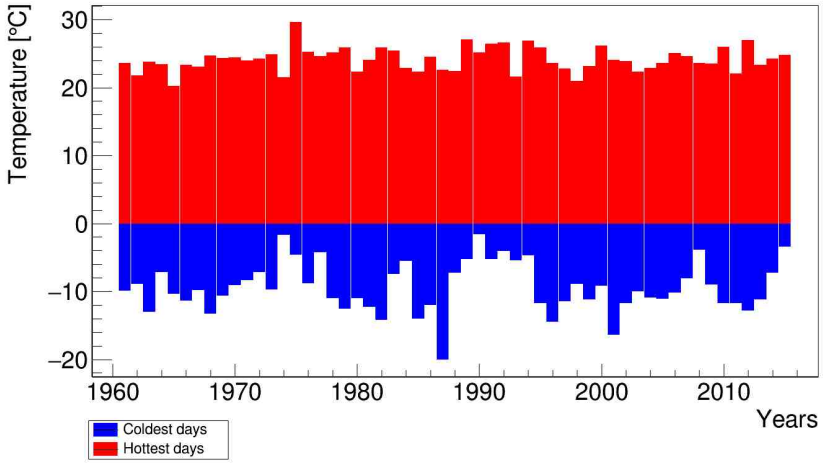

Figure 8.4: Hottest and coldest temperatures in Lund

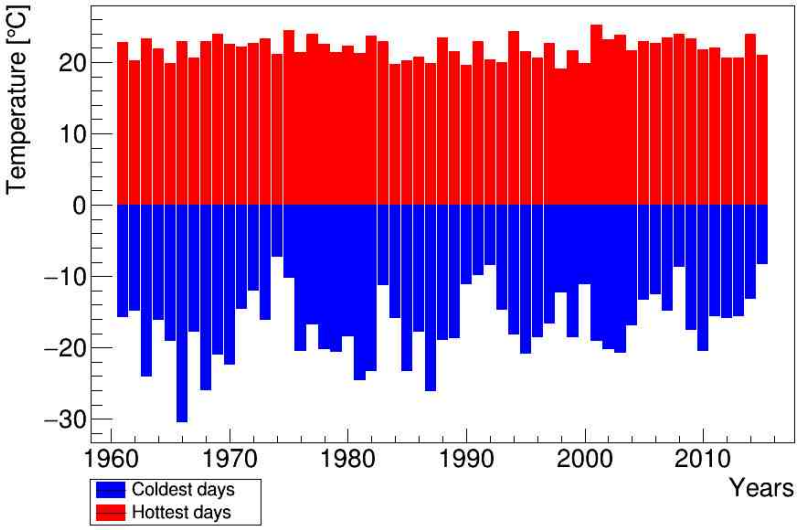

Figure 8.5: Hottest and coldest temperatures in Karlstad

8.3 Question 3

This histogram is the result of comparing the relative temperature described in the methods for moontemp to the moon phase. A scatter plot is not used as it becomes cluttered with the

amount of data points given (up to 12,000). Instead the difference is plotted in a histogram. A linear decline can clearly be seen. If a linear relationship between the moon phase and the temperature is seen, the histogram would have a normal distribution. A linear decline in the histogram is seen when there is a random correlation between the to sets. This is due to the linear decrease of area a random data point can fall into with increasing error.

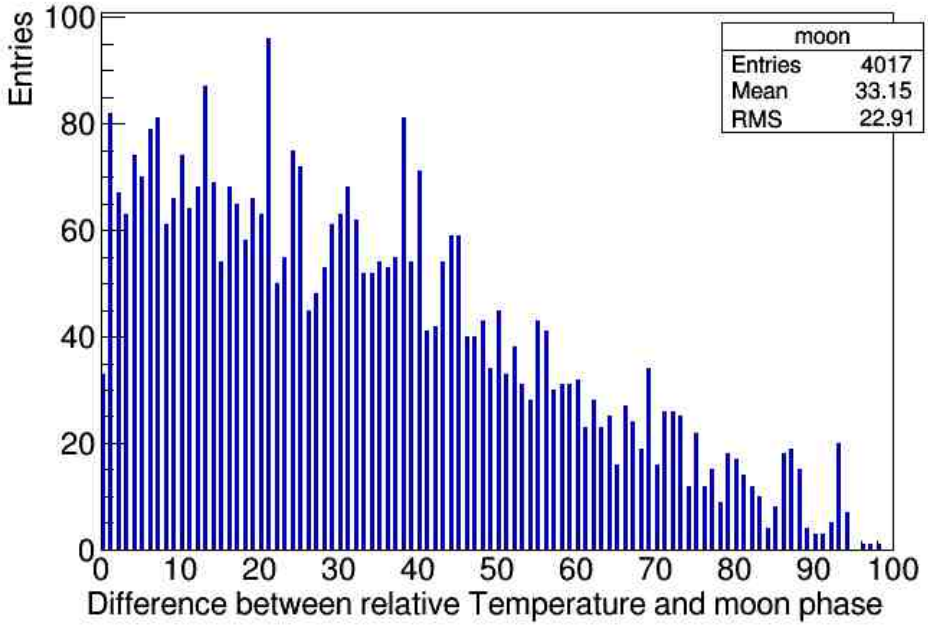

Figure 8.6: Histogram obtained by comparing the ralative temperature per moon cycle to the moon phase. Temperature data is from Lund.