MNXB11

Final Project

Authors
Samuel Anand
Sebastian Jakubiak
Amanda Sjölin

November 12, 2024

1 Goals

The goals of this project were to analyze big datasets of temperatures, using Bash, C++, and ROOT, and to understand and work collaborative using GitHub. Three things were decided to be analyzed; the warmest and coldest days of each year, the mean temperature of winter for every year, and the temperatures of 3rd of April, 26th of June, 29th of September, and Christmas eve every year.

2 Approaches used

The choice of what to analyze was made in the first meeting. At first the group worked together to create the repository, and start working on the code, but then tasks were divided among the members. The code was written in a group of three people working using GitHub as the common place for storing changes in the code. The group work consisted of individual work with changes submitted to the GitHub repository, mostly done in-person. The GitHub approach was chosen so that there was one person with the ability to push to the main branch, while other people had to create pull requests to do so. To that end, each person created their own branch and worked in it, while creating pull requests with changes.

The data was obtained using a program written in C++, with Vince's CSV Parser library, and ROOT. The data was pre-processed using a bash script. The filter/cleaner bash script created a file with removed lines up until the first line with data, and created files with filtered data appropriate for each goal.

The C++ code consisted of three source files used in three header files, for different types of data analysis, as well as the main code which compiled into the final executable.

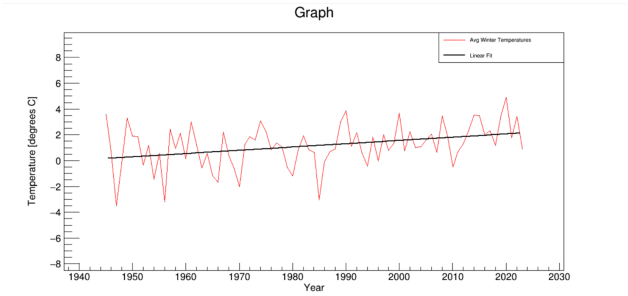
The source files had a similar structure; a function to extract the year from the date string, a parsing function to read the csv-files, a plotting function, and a main function. The parsing function was initialized with the filename, the years vector, and the temperature vectors. It went through the provided dataset, filling the vectors with appropriate data, using Vince's CSV Parser library, and the year extraction function. The plotting function, initialized using two vectors: years, and temperatures. This function simply plotted the provided vectors using ROOT. The main function for the source code which was initialized using the filepath for the dataset. This function would take the filepath, extract the filename, and add the appropriate prefix so that the result would correspond to the filtered dataset for the type of analysis done in that code. Then, it would initialize the years and temperatures vectors, fill them using the parsing function, and plot them using the plotting function.

In the main file; the three header files were included. The structure of this code was the following: two helper functions to determine whether the input filepath, and mode are correct. The main function did not have any initialization arguments. The body of the function consisted of running the aforementioned bash script, asking the user for the filepath input, checking if it is valid and terminating the program if not. The same was done for the mode with which the user chose which type of data analysis they wanted to conduct. Finally, depending on the chosen mode, the program would run the corresponding function from the appropriate include file described in the previous paragraph.

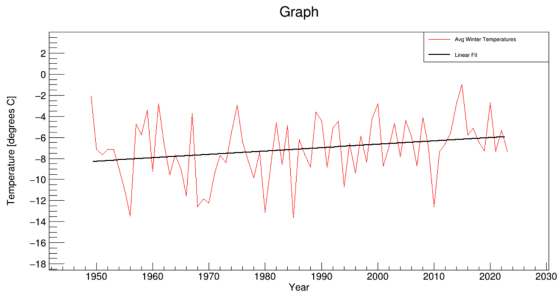
3 Mean temperatures for winter

This analysis considered the average winter temperature over each year and plotted those values against years in hopes to see the progression of climate change. To define winter, the months November through February were considered. It is to be noted that the temperature values were all taken at 6 or 7 am, since these times had values throughout all our datasets for all the locations. The temperature was calculated for 4 different locations, Visby, Falun, Lund and Luleå. Lund and Falun were selected as they had the longest measurement. Otherwise these locations were chosen for their geographical locations. Visby is on Gotland, an island and therefore would have been more affected by ocean currents. Luleå is in the north and thus it is expected to observe harsher winter weather there. Falun is quite central in Sweden and Lund is in the far south. The obtained figures can be found in Figure 1.

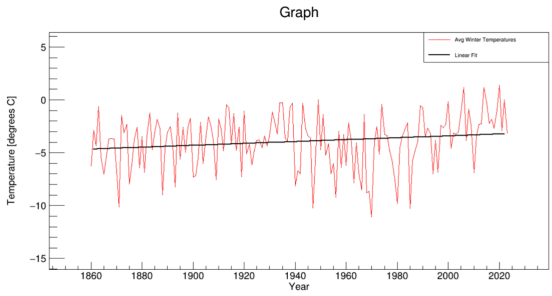
It is observed that the least variation in temperature occurs in Visby, which would make sense as Visby is a coastal city, with milder climate. However, both Luleå and Visby are coastal cities and their average increase in temperature over time were higher than the more inland cities of Lund and Falun. The average yearly increase in temperatures was the highest in Luleå and the lowest in Falun. This suggests the temperatures vary from year to year the least but increase the most in coastal cities. Furthermore this could suggest that cities in the north increase on average more than cities in the south, however this would require the studying of other cities in the north to prove this claim. It should also be noted that as the data varied quite a bit, the errors were rather significant for the linear fit, and therefore it cannot be fully trusted. It is presented more as a representation to show a plausible trend in increasing temperature.



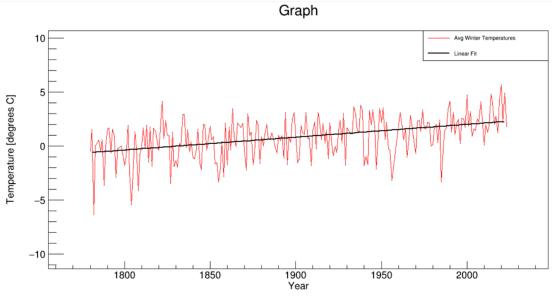
(a) Mean winter temperatures for Visby



(b) Mean winter temperatures for Luleå



(c) Mean winter temperatures for Falun

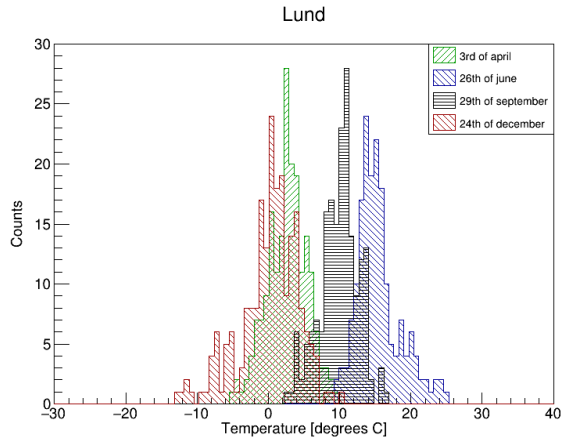


(d) Mean winter temperatures for Lund

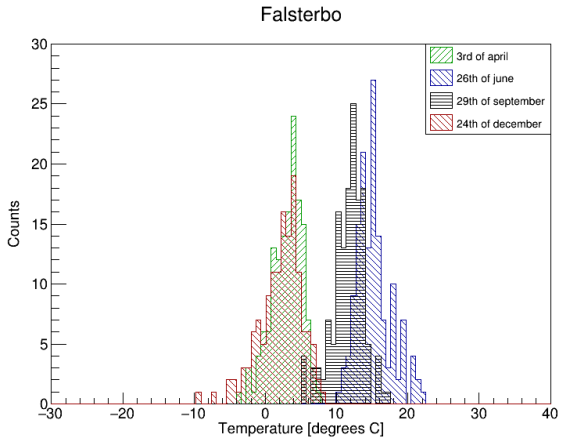
Figure 1: Mean winter temperatures for Visby, Luleå, Falun and Lund against years

4 Temperature for four specific dates

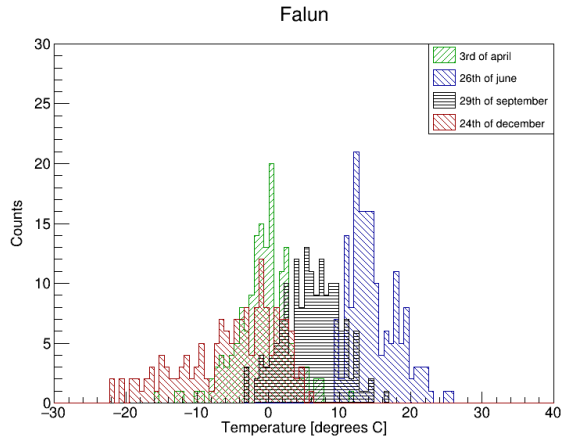
Figure 2 a), b), c), and d) shows the temperatures for 4 different dates at Lund, Falsterbo, Falun, and Luleå, respectively. The dates are the 3rd of April, the 26th of June, the 29th of September, and the 24th of December. One count in the histogram represent one instance where that temperature was recorded. From these one can deduct that the north of Sweden has more variation in seasons compared to the south cities. In Falsterbo and Lund the histograms of the overlap, while in Falun and Luleå they are more spread out.



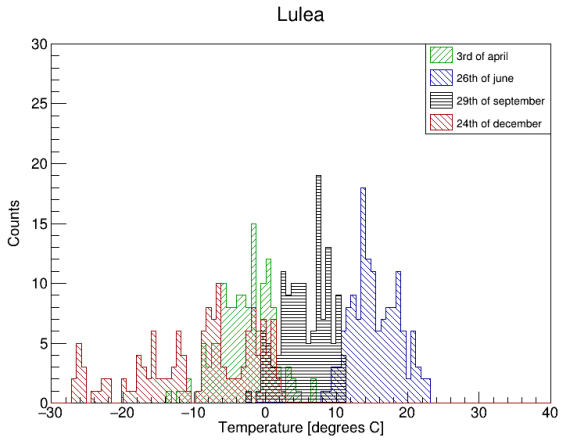
(a) The four temperatues in lund



(b) The four temperatues in Falsterbo.



(c) The four temperatues in Falun



(d) The four temperatues in Luleå

Figure 2: Temperatures for the four dates for Lund, Falsterbo, Falun, and Luleå

5 Warmest and coldest days

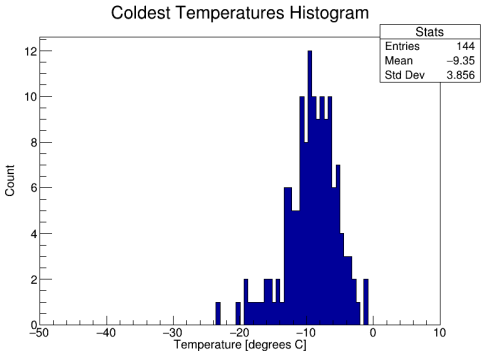
For this analysis, the pre-processing was done in a way that resulted in a dataset consisting of only the coldest and warmest temperatures for each year. That is, there were two entries per year, one with the coldest temperature and one with the warmest. As such, the temperature measurements used could have been taken at any time of the day. The results were then plotted as histograms with each count corresponding to one

recorded temperature on the x-axis. In the graphs, the mean and standard deviation, as well as the number of datapoints, are given. The graphs and the aforementioned statistics were then compared for four different locations: Lund, Falsterbo, Falun, and Luleå. The graphs with the statstics boxes can be seen in Figure 3, below. The statistics for each location are also given in Table 1, also below.

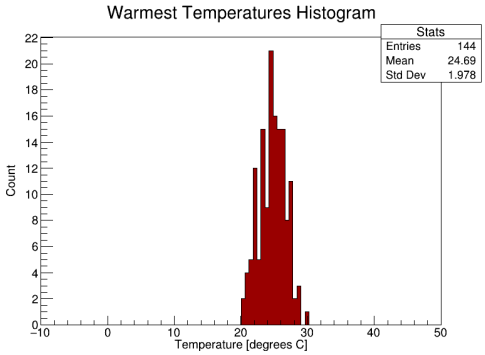
As can be seen, there are definite differences in mean coldest temperatures between locations. However, the mean warmest temperatures do not differ by much in comparison. Moreover, the standard deviations for the coldest temperatures across all locations are greater than the ones for warmest temperatures. This suggests that there are greater variations in coldest temperatures than warmest ones. Additionally, we can see that the location with the highest mean coldest temperature is Falsterbo, which also has the lowest mean warmest temperature. Since Falsterbo is a seaside location, we can surmise that the seaside climate is more mild than that of inland locations, such as Falun. What is more, both Lund and Falsterbo, located in the south, have smaller differences between the warmest and coldest temperatures compared to Falsterbo and Luleå, which are located further north. This suggests that the climate in the southern parts of Sweden is milder than that of the northern parts.

Table 1: Table containing the summary of the information from the plots obtained in this analysis. The σ in this case was chosen to represent standard deviation.

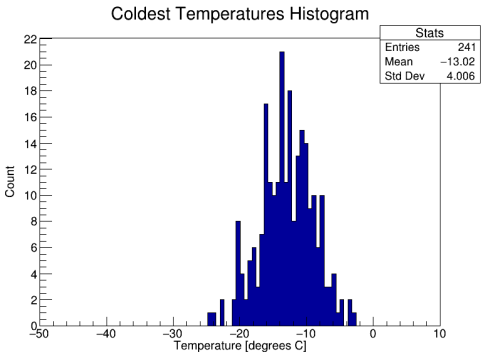
	Mean coldest temp.	Mean warmest temp.	Coldest temperature σ	Warmest temperature σ
Falsterbo	-9.35 °C	24.69 °C	3.856 °C	1.978 °C
Lund	-13.02 °C	27.66 °C	4.006 °C	2.494 °C
Luleå	-30.23 °C	26.58 °C	3.765 °C	2.059 °C
Falun	-25.71 °C	28.09 °C	5.160 °C	2.015 °C



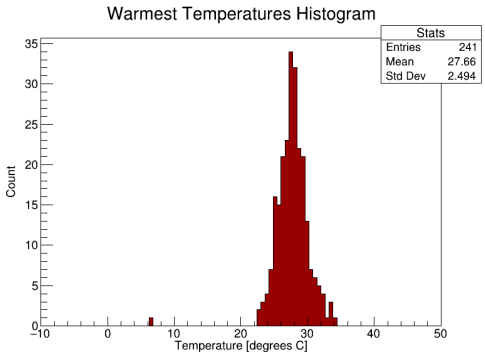
(a) Coldest temperatures histogram for Falsterbo



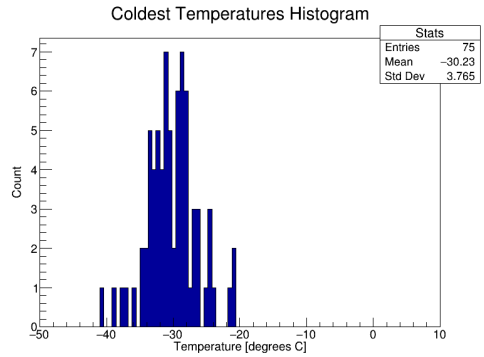
(b) Warmest temperatures histogram for Falsterbo



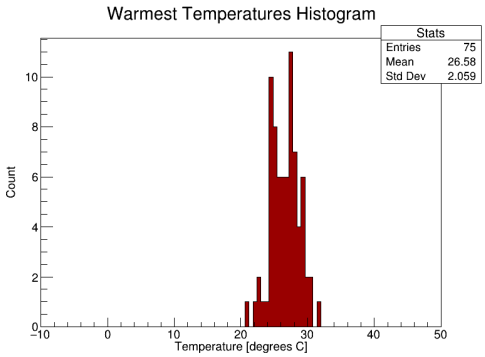
(c) Coldest temperatures histogram for Lund



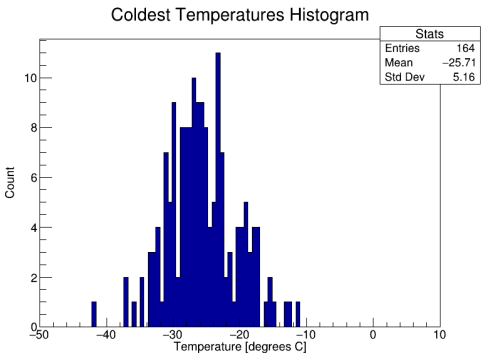
(d) Warmest temperatures histogram for Lund



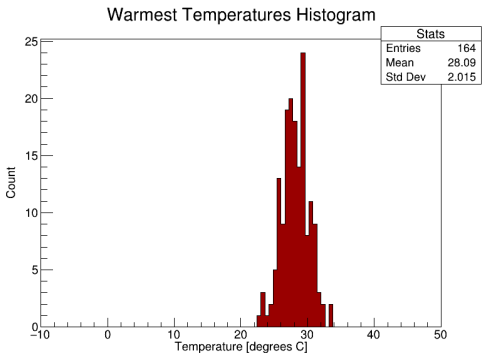
(e) Coldest temperatures histogram for Luleå



(f) Warmest temperatures histogram for Luleå



(g) Coldest temperatures histogram for Falun



(h) Warmest temperatures histogram for Falun

Figure 3: The coldest and warmest temperatures histograms for the four locations: Falsterbo, Lund, Luleå, and Falun.