Goals:
Create a dot plot of the length of collatz sequences, for all numbers from 2 to 10000.
Create a dot plot of the maximum value of all collatz sequences starting with numbers 2 to 10000.
Create a histogram of the lengths of collatz sequences, for sequences starting with numbers 2 to 10000
Create a different, interesting plot using collatz sequences, which will consist of a dot plot of the length vs the maximum value of the collatz sequence, for all starting values 2 to 10000

Design:
Iterate over all numbers between 2 and 10000, as the graphs must be created based on the collatz sequence for all numbers between 2 and 10000.

> Create a collatz sequence for the iterated number.
>
> Count the number of lines in the collatz sequence, which is the length of the sequence, then record that and the current number to a file, which will create one point in the length of the collatz sequences graph.
> We also record the sequence length to a separate file, to be sorted and used to create the histogram data later.
>
> Next, we sort the collatz sequence for the iterated number
> We then take the last number in the sorted sequence, which should be the largest value in the sequence, and record that and the iterated number to a file, which should create one point in the collatz sequence max value graph.
>
> We also record the length and max value for the iterated number to another separate file, which should create one point on the length vs max value graph.

At this point we stop iterating

We then start work on arranging the histogram data properly.
To begin with, we sort the stored sequence lengths, and record the last value in the sorted data, which should be the maximum length.
We then count the number of times each length appears, and save these values in order from least to greatest.
Finally, we record the sequence lengths that repeat, and the sequence lengths that are unique and combine these values before sorting them, giving us a list of all sequence lengths that appear in order from least to greatest.

We then iterate over all numbers between 0 and the maximum length found, so we can give a height to all of the bars between 0 and that maximum length.

To begin with, we check to see if the next length that appears is equal to the current iterated value.
If it is not, we simply record a 0 for the height of that bar for that sequence length, as it did not appear.

Otherwise, we record the next number of occurrences as the height of the bar.
This should correspond to the next length, as both are in order.
We then remove the first number of occurrences and sequence length, as this will result in the next sequence length to be tested on the next loop.

We stop iterating here.

Finally, using all of the recorded data, we plot all four graphs, before removing any files during the running of the script

Pseudocode:
Compile collatz.c with the makerfile
For every number between 2 and 10000

Calculate the collatz sequence for the number, and save it into File temp

comparison_data will be used to graph the personal interesting plot

Creating length data and sorting it into dot plot format

Append the number and a whitespace to File length_data
Calculate the number of lines that are in File sequence, and by extension the collatz sequence
Append the number of lines to File length_data, to File raw_length_data, and to File comparison_data
Remove the last byte of File max_value_data, which should contain a newline

Creating value data and sorting it into dot plot format

Append the number and a whitespace to File max_value_data
Sort File sequence numerically, and save that to file ordered_sequence
Take the last line of File ordered_sequence, and append it to File max_value_data and to File comparison_data

Sorting length data into histogram format

Sort File raw_length_data numerically
Find the last line in File raw_length_data, which should be the maximum value found

Find all the unique items in File raw_length_data, and all the items in File raw_length_data that repeat, then append those two lists in File lengths_that_appear
Sort File lengths_that_appear numerically
Count the number of times each length appears in File raw_length_data, and save that in File number_of_occurrences

For every number between 2 and the maximum value found
    Get the first line of File lengths_that_appear
    If the value of the first line of File lengths_that_appear and the number are not equal,
        Append 0 to File histogram_length_data
    Otherwise,
        Save the number of times each length appears from File number_of_occurrences in File histogram_length_data
        Then get every line except the first from File number_of_occurrences, and save that to File number_of_occurrences
        Then get every line except the first from File lengths_that_appear, and save that to File lengths_that_appears

Activate gnuplot
    Set format of the plot to be pdf
    Set the name of the output to be figure 5
    Set the title of the plot to be Collatz Sequence Lengths VS Maximum Value
    Turn on visible x and y axis
    Set the x axis label to be length
    Set the y axis label to be value
    Plot File comparison_data as a dot plot

    Set the name of the output to be figure 2
    Set the title of the plot to be Collatz Sequence Lengths
    Set the x axis label to be n
    Set the y axis label to be length
    Plot File length_data as a dot plot

    Set the name of the output to be figure 3
    Set the title of the plot to be Maximum Collatz Sequence Value
    Set the x axis label to be n
    Set the y axis label to be value
    Set range to be from 0 to 100000
    Plot File max_value_data as a dot plot

    Set the name of the output to be figure 4
    Set the title of the plot to be Collatz Sequence Length Histogram
    Set the x axis label to be length
    Set the y axis label to be frequency

Set distance between bars to be 0
Set range to be from 0 to 200
Set domain to be from 0 to 225
Plot File histogram_length_data as a histogram

Remove all files created during runtime