

Student	Steven Moon	70
---------	-------------	----

Part 1

	Available	Marks
Import data and generate data frame	5	4
Create a new data frame <i>cumulativeData</i> that shows the cumulative rainfall and cumulative number of hours of sunshine for each city. Display this data frame.	10	8
Include a function <i>displayGraph()</i> that will take the city name weather parameter and will produce a line graph of that weather parameter over the year for that city. Display this using humidity in Wuhan and rainfall in Swansea.	15	14
Include a function <i>compareCities()</i> that will take the start month and end month and two weather parameters, and will plot the data for those parameters for both cities on the same graph. Mean temperature and humidity, if displayed, should be shown as line graphs, while all other weather parameters should be shown as bar plots. Appropriate labelling and colours should be used. Display this using mean temperature and hours of sunshine.	15	2
Commenting, white space and code presentation	5	3
TOTAL		31
Comment Q1: If you make any changes to the source datafiles you should include them in your uploaded submission. Q2: Good. You could also have added the cumulative hours overall. Q3: An excellent solution. The only thing missing is the city labels on the graphs. Q4: You've got the right skeleton for the function. Your use of white space is good but you need more comments, in particular including your name, student ID and the question numbers.		

Part 2

	Available	Marks
Load the data into a data frame <i>covidData</i> , ensuring that the headers are imported. Your script should assume that the data file is stored in the same folder as the script file.	5	5
Plot the number of cases and number of deaths by month for both countries on the same appropriately-labelled graph.	10	5
Create a data frame <i>CovidSummary</i> that shows descriptive statistics for the total number of new cases in China and the UK for April 2020, October 2020, April 2021 and October 2021. This should contain the following measures: mean, standard deviation, median, max, min, and range.	20	18
Perform a t-test to identify whether there is a significant difference in the numbers of new cases in the UK between October 2020 and October 2021. You should state your experimental hypotheses, the result of the test, and whether you accept or reject the null hypothesis in your comments.	10	7
Commenting and use of white space	5	4
TOTAL		39
Comments Q1: Good approach – load everything then process it in R. Your use of white space is good but you need more comments, in particular including your name, student ID and the question numbers. Q2: Calculation of summary data is good, but plotting the data produces this error: <pre>> #Plot chart of cases > barplot(summary,height = summary\$cases,beside = TRUE) Error in width/2 : non-numeric argument to binary operator In addition: warning message: In mean.default(width) : argument is not numeric or logical: returning NA</pre> Q3: Excellent use of a function to do the calculations. Good structure of summary dataframe. Q4 The process is correct, but the interpretation of the results is incorrect. 2.23-16 is much smaller than 0.05, so the null hypothesis can be rejected Your use of white space is good but you need to including your name, student ID and the question numbers.		