

Student	Theresa Darzins	56
---------	-----------------	----

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	8
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	7
Commenting, white space and code presentation	5	5
TOTAL		32
Comment Q1: Files are not in the working directory for uploading. Good handling of city identifier. Q2: Good solution, but you don't need to do all of the extra stuff with the matrices at this point. Q3: Swansea and Wuhan plots are fine, but you haven't created the <i>displayGraph()</i> function to draw them. Q4: A workable graph, but again you didn't use a function to display it. Excellent commenting and use of white space.		

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	1
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	9
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	6
Commenting and use of white space	5	3
TOTAL		24
Comments Q1: The read statements require the data files to be in a folder other than that of the script file, otherwise the code will not run. Q2: You should not be reading data from a file at this point, but should instead be working with the data you've already imported. The charts are very good, but you've lost a lot of marks for not processing the data in R. Try to give your variables more descriptive names than DF1, DF2, DF3. Q3: Again, these should have been calculated from the main dataset not from a summary. Some of the calculations are fine although max, min and median produce unexpected results or errors. Q4: The statistical analysis is fine but the plots are not needed. Commenting and white space are generally good but more comments in places and more descriptive variable names would be useful.		