

Student	Emily Stockdale	77
---------	-----------------	----

Part 1

	Available	Marks
Import data and generate data frame	5	5
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	10
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	11
Commenting, white space and code presentation	5	5
TOTAL		39
Comment Q1: Good Q2: Good. You could also have added the cumulative hours overall. Q3: This works well for producing the plot, and does indeed use a function, but the function doesn't take in the required parameters. Q4: Good graph. This works well for producing the plot, and does indeed use a function, but the function doesn't take in the required parameters. Excellent use of comments and 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	5
Plot the number of cases and number of deaths by month for both countries on the same appropriately-labelled graph.	10	9
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	14
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	5
Commenting and use of white space	5	5
TOTAL		38
Comments Q1: Good. Good handling of the required library. Q2: Good processing and graph. The comments explaining the problems are a good idea. Q3: For some reason dplyr didn't load automatically. Chaining is fine but needs the maggritr library, which also isn't part of the vanilla build. The code itself is good, but having to add in the extra libraries caused problems Q4: You need to test for normality and remove outliers if necessary, and define your hypotheses. In this case the null hypothesis would be "H0: there is no difference between the two samples" while the alternative hypothesis would be "H1: there is a difference between the two samples". Your decision to reject the null hypothesis is correct. Excellent commenting and use of white space.		