Assignment 1

Sem 2, 2020

Marking	Student Id	Student Name
Karam	u3196825	Tuan Anh (Vincent) Nguyen

Assignment Part	Sub Mark	Out of
Part A	14	**15**
Part B	9	**15**
Part C	14	**20**
Part D	5	**5**
Total Mark	42	**55**

Part A

Items	Comments	Mark	Out of
Correctness Does the code of this part work correctly?		**10**	**10**
1) Have the data files loaded correctly?	☒		
2) Are the data concatenated correctly and produced one file with all the data records?	☒		
3) Has the students provided the code to check for problems in the previous two steps?	☒		
Programming Style and Code Layout (1 mark each)		**4**	**5**
• Has the student used good names for naming the variables and used blank line to separate the code blocks?	☐ Meaningful names which reflect the purpose of each object is beneficial to get the code readable and can be recycled.		
• Has the student used levels of indentation, so the code is easy to follow?	☒		
• Has the student used comments to make the code easy to understand?	☒		
• Has the submitted code included a prologue and a description at the beginning of the script?	☒		
• Has the student submitted the output of console window after running the code of this part and identified which output is for which component?	☒		

Task A total	14	15

Part B

Items	Comments	Mark	Out of
Correctness Does the code of this part work correctly?		7	12
1) Remove the variables, which have no data at all (i.e. all the records in these variables are NAs)	☒		
2) Drop the variables, which have few data (i.e. NAs values are more than 90% of number of records in these variables).	☐ Incorrect output. No columns should be removed		
3) Change the column names to have no spaces between the words and replace these spaces with underscore the '_' character.	☒		
4) Change the type of the column called "Date" from character to Date data type.	☐ Missing		
5) Add two new columns for the month and year of the data in each file, you may extract the contents of this column from the "Date" column. Please note that the data are collected for 19 months across 3 years (2018, 2019 and 2020).	☒		
6) Change the type of the "Month" and "Year" columns from Character to Ordinal with levels as the number of months in a year (i.e. 12) and number of years (3).	☒		
7) For all the numeric columns, replace the remaining NAs with the median of the values in the column, if exist.	☐ It is not efficient to do this step by defining the variables manually, think of if we have 1000s of these cols.		
Programming Style and Code Layout (1 mark each)		2	3
• Has the student used good names for naming the variables and used blank line to separate the code blocks?	☐ Meaningful names which reflect the purpose of each object is beneficial to get the code readable and can be recycled.		
• Has the student used comments to make the code easy to understand?	☒		
• Has the student submitted the output of console window after running the code of this part and	☒		

identified which output is for which component?			
Task B total		**9**	**15**

Part C

Items	Comments	Mark	Out of
Correctness Does the code of this part work correctly?		**12**	**17**
1) Printing the summary (i.e. minimum, median, mean, maximum)	☒		
2) Extracting the average of minimum temperature per month and year	☐ You should group the output by month and year		
3) Extracting the average of speed of maximum wind gust per direction of maximum wind gust	☒		
4) Which months were dry, if any, (i.e. no rainfall at all)? And in which year?	☒ Code to print which one?		
5) What about the humidity, which month in the ACT has the highest humidity level in the 2019 year?	☒		
6) For 2019, extract the minimum, maximum and average temperature, wind speed and humidity per month and per quarter in 2019 only.	☒		
7) Plot the histograms/bar-charts for each variable of the previous question.	☐ Missing		
Programming Style and Code Layout (1 mark each)		**2**	**3**
• Has the student used good names for naming the variables and used blank line to separate the code blocks?	☐		
• Has the student used comments to make the code easy to understand?	☒		
• Has the student submitted the output of console window after running the code of this part and identified which output is for which component?	☒		
Task C total		**14**	**20**

Part D			
Items	**Comments**	**Mark**	**Out of**
		5	**5**
1) Has the student generated reasonable (at least two) questions to be answered by this data?	☒		
Task 4 total		**5**	**5**