# 

# COMP 5070

# Statistical Programming for Data Science

## Assessment 1 SP5 2022

* You must submit your assessment through LearnOnline. Every submission should include two files – Jupyter Notebook file with all cells executed (that is, with all outputs, graphs, discussions) and PDF printout for the same file.
* Please think about your reader. Take care about readability of your code and your discussions. Make it easy to see that your job has been done perfect.
* No compressed files (e.g. .zip, .rar, .tar, .gz, .7z) are allowed for submission. These files will be ignored during marking.
* You do NOT need to include the data files provided to you, as it can be safely assumed I have them too.
* The assessment is out of 100 marks. To obtain the maximum available marks you should aim to:

1. Code all requested components (60%)
2. Aim for optimised code in terms of computational overhead (5%). It is not always possible to avoid loops, however you should aim to avoid loops where possible.
3. Use a clear coding style (5%). Code clarity is an important part of your submission. Thus, you should choose meaningful variable names and adopt the use of comments - you don't need to comment every single line, as this will affect readability - however you should aim to comment at least each section of code.
4. Have the code run successfully (5%).
5. Output the information in a presentable manner as decided by yourself and present the requested statistical analyses/discussions (25%).

• *Plagiarism is a specific form of academic misconduct.* Although the University encourages discussing work with others and the Social Forum will support this, ultimately this assessment is to represent your individual work. If plagiarism is found, all parties will be penalised.

You should retain copies of all assignment computer files used during development of the solution to the assessment. These files must remain unchanged after submission, for the purpose of checking if required.

Late submission will be penalized by 10-point deduction for each day or part of it after the due date.

# Global Warming: Fact or Fiction?

***Adelaide temperature data analysis***

In this assessment you will analyse data about a temperature in Adelaide. The data was collected by Australian Bureau of Meteorology – the best source of the climate related information in Australia. Two presented data set contain information about daily maximal and minimal temperatures for Adelaide weather station 023034 at Adelaide Airport. You can find these and other locations data at <http://www.bom.gov.au/climate/data/index.shtml?bookmark=200>

You get a chance to say your own word on the question of a climate change based on the hard-core facts from a rigorous data analysis. You must complete the required steps and provide answers on the following questions:

1. Write an introduction about what type of analysis you plan to execute. Provide a brief data description: what are your data about; how many variables and observations there; what are highest and lowest temperatures were observed and when?
2. Make a function to calculate and (nicely) present descriptive statistics (mean, standard deviation, skewness, kurtosis, median, IQR) for a given month and year. Run it for a maximal temperature for a month of your choice in 1961 and 2021. Compare results and discuss. Present appropriate data visualisation to support your discussion.
3. Combine two data sets together and calculate a daily range between maximal and minimal temperature. Get an average range for each year, discuss its distribution and change over the time. Provide appropriate data visualisations to support your discussion.
4. Count number of days per year with maximal temperature above 35 degrees. Build a graph. Discuss your observations.
5. Select and justify a threshold for a minimal temperature and repeat step 4 for the number of days below the chosen threshold.
6. Combine results from steps 4 and 5; discuss your findings and implications for the question of climate change.
7. Write a conclusion outlining your analysis and results.

You do not need to write too much as everyone hates reading long and/or boring reports. At the same time, if you need more space for your discussion then you can take it. There are no limits for word count or number of pages.

As a general rule, every question requires to present result in three ways – as a graph, as numbers and as text describing both. There might be exceptions when you cannot have all three, but these situations are not so common.