

Exam Details

Date:	Saturday 21 st of November 2020
Start Time:	9 a.m.
End Time:	11 a.m.
Duration:	2 hrs.
Grade:	15% of module

Note: If students have any questions, they can contact the lecturer during the exam period via david.leonard@tudublin.ie.

Programming Language

R or Python

(The language chosen should be the **same** as that used to complete the assignment.)

Permitted Libraries

R - any of the packages in **Tidyverse** and / or those that come baked in to R e.g. base.

Python – numpy, pandas, matplotlib, seaborn, re, and datetime

Exam Brief

Students will be provided with a reasonably sized dataset (e.g. 10,000 rows and 30 columns is an indicator of the ballpark size i.e. it is neither a toy dataset and nor a massive one) i.e. it will be available on **Brightspace** along with any information relevant to understand the purpose of the variables. There may also be 1 other data file provided for the purpose of merging with the first dataset.

After manually downloading the datasets, the tasks to be completed for the exam are as follows:

Task 1

Import the datasets and perform any operations necessary to clean them up in preparation for analysis.

Task 2

Perform an exploratory analysis of the entire dataset, the goal is to produce a variety of relevant summary statistics and a **single** interesting visualisation that captures the relationship between at least 3 of the **original** variables.

Task 3

Focus on a subset of the data and create 3 derived variables for the purposes of answering a single question of interest.

Submission details

File format for code: text or script files

Upload a **zip file** of the following to the link available on **Brightspace**:

- 1.) 1 file containing all code written during exam
- 2.) 1 much shorter file containing only the code to perform the 3 specific tasks.

Exams will **not** be corrected unless both files are submitted.

N.B. Students are advised to back their work up regularly.

Grading criteria

As the exam is taking place online there is an open ended character to it i.e. there is no correct answer. However, submissions will be graded in accordance with the following criteria:

Task 1 - 30%

Quality of Outcome	7.5%
Correctness	7.5%
Complexity of operations	15%

Task 2 - 35%

Quality of Outcome	15%
Correctness	5%
Complexity of operations	15%

Task 3 - 35%

Quality of Outcome	15%
Correctness	5%
Complexity of operations	15%