



MS5130 Assignment 2

“Data Analysis Project and Learning Diary”

For MS5130 Applied Analytics in Business and Society, 2024–2025 edition

Contact: Dr. Blair Wang, blair.wang@universityofgalway.ie

1. Introduction.....	1
2. Your Task.....	2
<i>Deliverable 1. Enhancements to your work for MS5130 Assignment 1.....</i>	<i>2</i>
<i>Deliverable 2. Learning Diary.....</i>	<i>3</i>
<i>Deliverable 3. MS5130 Assignment 2 Cover Page.....</i>	<i>5</i>
3. Grade Descriptors.....	6
4. Submitting Your Work & Due Date.....	8
5. Academic Integrity and Referencing.....	8

1. Introduction

This assignment is a formative assessment task for MS5130. The purpose of formative assessment is to provide an opportunity for students to undertake a low-stakes task that provides hands-on practical experience and lay foundations for later assessment tasks.

This assignment is worth **20% of your total marks** for MS5130 and is to be completed **individually**.

2. Your Task

This task involves multiple deliverables:

Deliverable 1. Enhancements to your work for MS5130 Assignment 1

- 14 of total 20 marks
- Format: ZIP (.zip) file containing your work (including Quarto QMD file, generated HTML file, and any other relevant assets such as image resources).

You have previously completed a basic analysis of one dataset, for MS5130 Assignment 1.

For this deliverable, you are required to make some enhancements to the work that you have previously completed for MS5130 Assignment 1.

To attain a passing grade (mark range 40%—49%) for this assignment, you are required to implement ALL of the following “**basic enhancements**” (BEs):

1. **(BE1) Executing the R code inside Quarto:** Ensure that all your R programming code is now saved in Quarto document format, following the same format as shown in various lecture demonstrations of Quarto. In other words, your R code should not be stored as “.r” or “.R” script files, but as R code inside code fences inside a “.qmd” Quarto document file. The outcome of the R code executing successfully should be visible inside the generated HTML document. Any references to sources that you have drawn on (e.g., journal articles and conference papers) are to be referenced within this Quarto document, both as in-text citations and in a reference list at the end of the generated Quarto document.¹
2. **(BE2) Use multiple datasets:** You were previous required to analyse only one dataset. For this assignment, MS5130 Assignment 2, you are now required to augment your previous analysis with TWO (2) additional datasets that are relevant to the topic of your analysis.



WARNING — Please ensure that your datasets are publicly available and permitted for general use. Exercise particular care to avoid using any dataset that is confidential or that you have obtained from proprietary sources (e.g., from your workplace).

3. **(BE3) Combine datasets together:** The THREE (3) datasets that you now have must be combined together in some way, i.e., you should NOT be performing three separate unrelated analyses. You may consider using ‘dplyr’ or other such R packages or techniques to combine datasets together.
4. **(BE4) Synergy of descriptive and inferential analysis:** Your analysis is required to demonstrate the usage of both descriptive (e.g., measures of central tendency and dispersion) and inferential analysis (e.g., regression models like GLM / GAM).
5. **(BE5) Explanatory text:** Write some text (sentences and paragraphs, not just dot points) in your Quarto document, explaining the work that you have completed.

¹ See <https://quarto.org/docs/authoring/citations.html>

Some students may be ambitious for higher grades beyond a passing grade. In order to be considered for a grade higher than a pass (i.e., mark of 50% or higher), you are required to complete some or all² of the following “**superior enhancements**” (SEs):

1. **(SE1) Depict your data streams using Mermaid:** Now that you have multiple files being combined together, draw a diagram depicting your data streams merging together. For a correct implementation, this diagram should be drawn using Mermaid code embedded inside your Quarto document.
2. **(SE2) Use of a private GitHub repository:** To correctly implement this enhancement, all your work should be stored and tracked using a Git repository hosted on GitHub. Better implementations of this enhancement are required to involve: (a) various commits to your work over a *meaningful period of time* (i.e., not just one commit just before the assignment is due); AND (b) changes made in commits that are *logically bundled* (i.e., representing one holistic improvement to the work, not just random or arbitrary modifications); AND (c) a brief introductory note in the README.md file. **Important note:** *To ensure that the repository is set to “private” for academic integrity reasons, you must contact the lecturer Dr. Blair Wang at least 1 week before the due date of the assignment to confirm that you wish to pursue this SE. Blair will then set up a private GitHub repository for you and add your username.*
3. **(SE3) Use of geographical data analysis using Leaflet:** Construct, using R code, a geographical analysis of your data using the Leaflet package for R. Your analysis should draw meaningful insights from this geographical data analysis.
4. **(SE4) Use of interactive charts/graphs/plots:** Construct, using R code, some charts, graphs, plots or other such data visualisations that are interactive (i.e., they are constructed as interactive HTML widgets, not static image files). You may consider using the Plotly package for R or similar such packages. Your analysis should draw meaningful insights from these data visualisations.
5. **(SE5) Explanation video:** Produce a short video (duration: 3-5 minutes) in which you briefly explain what you have produced in your work for this assignment. The format of this video should be similar to that which was required in MS5130 Assignment 1 (your face addressing the camera, your voice narrating your work, and a screen recording of you performing a ‘walkthrough’ of your work.)
6. **(SE6) Other superior enhancement:** If you have an idea for an alternative superior enhancement, it could be considered as such, provided that you have discussed your idea with Dr. Blair Wang and have received an email from him, confirming that your idea qualifies as “SE6” for your assignment.

To be clear, these SEs are all **optional** and you do **not** need to complete these SEs in order to receive a passing grade for this assignment; they are only listed here for students who are ambitious to receive a grade higher than a pass (i.e., mark of 50 or higher).

Deliverable 2. Learning Diary

- 6 of total 20 marks
- Format: ZIP (.zip) file containing your work (including Quarto QMD file, generated HTML file, and any other relevant assets such as image resources).

A very common question in job interviews, promotion meetings and performance reviews is something like the following: “Tell me about a time you faced a technical challenge. How did you overcome that challenge?”

² The grade awarded will be commensurate with the overall extent to which the superior enhancements (SEs) are successfully implemented, encompassing a consideration of both quantity and quality.

To support your ability to answer this kind of question in future, this deliverable requires you to keep a **learning diary** in which you record your experiences of overcoming technical challenges in MS5130, especially those related to the R programming language, open-source R packages, and the RStudio software environment. The focus of the learning diary is that you demonstrate your ability to exercise **resourcefulness**, a concept discussed throughout the semester in the MS5130 lectures.

Superior submissions would ensure that the **MS5130 learning outcomes** are in some way addressed or discussed by your overall learning diary:

1. Explain the different data types and objects used in statistical programming.
2. Develop and test models of business and societal data through statistical programming.
3. Install and exploit open source statistical programming packages.
4. Combine and reshape multiple datasets.
5. Create relevant visualisations using advanced statistical programming plotting packages.
6. Use data to identify, analyse, and solve business, policy and societal problems.

Further, superior submissions would ensure that the **MS5130 learning activities** are in some way addressed or discussed by your overall learning diary:

1. Your efforts to engage with the MS5130 Labs, either by direct participation during a live session or by engaging with the video recordings of those sessions (e.g., trying the exercises on your own computer);
2. The work that you have undertaken for MS5130 Assignment 1 (20% formative task);
3. The work that you have undertaken for MS5130 Assignment 3 (60% summative task).

Please note that **the learning diary is expected to be written in Markdown, using the RStudio Quarto publishing system, exported to a HTML file**. To ensure that you do not overlook this formatting requirement when you are glossing through this document: the Markdown, Quarto and HTML logos are copied below as a visual cue to draw your attention to this matter:



Please remember to use Markdown, in Quarto, and export to a HTML file!

Exemptions to this requirement can be requested by sending an email to Dr. Blair Wang.

Exemptions are typically granted only due to exceptional personal or technical circumstances such as chronic lack of access to a laptop on which one can install a version of RStudio that has Quarto installed.

Your learning diary consists of a series of diary entries (i.e., it is a one-to-many relationship between learning diary and diary entry). At a minimum, each diary entry is expected to feature the following:

- **Date:** The date of the events described in the diary entry;
- **Problem Statement:** A brief description of the technical challenge that you faced;
- **Solution:** A brief description of how you overcame that challenge.

There is no minimum word count nor maximum word limit enforced for this deliverable. However, you are advised to exercise your best judgement when writing your learning diary. In the unlikely event that any student submits a learning diary in excess of 10,000 words, the MS5130 teaching team reserves the right to only assess the first 10,000 words.

There is also no “magic number” for how many diary entries you should write, either in total or for each of the topics outlined above. The most important thing is that you remain sufficiently self-aware so that you notice the moments when you naturally and authentically encounter technical challenges and apply efforts to overcome or solve those challenges *resourcefully*.

Please also note that **authenticity** is an important characteristic of a successful learning diary. This is not an assignment that can be “crammed” in the final days of the semester. Instead, it is expected that the assignment is gradually developed throughout the semester at a stable cadence. For example, it is expected that diary entries are spread out throughout the many weeks of the semester, not all concentrated in the final week of the semester. ***The MS5130 teaching team reserves the right to undertake an investigation where there is a reasonable suspicion that a learning diary has been plagiarised, is not authentic, or constructed in such a way that academic integrity has been compromised.***

Deliverable 3. MS5130 Assignment 2 Cover Page

For this deliverable, you are required to sign the MS5130 Assignment 2 cover page in the same style is what was required for MS5130 Assignment 1. ***Failure to include a cover page will result in a penalty of 1 mark from the total marks that you receive for this assignment.***

3. Grade Descriptors

The following grade descriptors for this assessment task are adapted from the grade descriptors of the Senate of the National University of Ireland (NUI), of which the University of Galway is a constituent university.

Grade	Marks range	Characteristics
1st Class Honours	90%—100% of total marks available	Supreme performance, engaging profoundly, systematically and comprehensively with the <i>assessment task</i> , brilliantly demonstrating: <ul style="list-style-type: none"> a superlative mastery of the subject matter, richly supported by evidence and <i>demonstrated resourcefulness</i>, reflecting deep and broad knowledge and understanding as well as extensive <i>independent research</i> an outstanding ability to organise, analyse and express ideas and arguments in an original, sophisticated and discriminating manner an optimal capacity for critical analysis the display of rare penetrative insight, originality and creativity
	80%—89% of total marks available	Exceptional performance, engaging deeply and systematically with the <i>assessment task</i> , with consistently impressive demonstration of <ul style="list-style-type: none"> a comprehensive mastery of the subject matter; amply supported by evidence and <i>demonstrated resourcefulness</i>, reflecting deep and broad knowledge and critical insight as well as extensive <i>independent research</i> an exceptional ability to organise, analyse and present arguments fluently and lucidly with a high level of critical analysis a highly-developed capacity for original, creative and logical thinking;
	70%—79% of total marks available	Excellent performance, engaging closely and systematically with the <i>assessment task</i> , with consistently strong evidence of <ul style="list-style-type: none"> a comprehensive mastery of the subject matter, ably supported by evidence and <i>demonstrated resourcefulness</i> excellent ability to organise, analyse and express arguments fluently and lucidly with a high level of critical analysis a highly-developed capacity for original, creative and logical thinking
2nd Class Honours (Grade 1)	60%—69% of total marks available	Very Good performance, engaging substantially with the <i>assessment task</i> , demonstrating strong grasp of the subject matter, well supported by evidence and <i>demonstrated resourcefulness</i> <ul style="list-style-type: none"> well-developed capacity to analyse issues, organise material, present arguments clearly and cogently some original insights and capacity for creative and logical thinking
2nd Class Honours (Grade 2)	50%—59% of total marks available	Good performance - intellectually competent answer (i.e. factually sound) with evidence of a reasonable familiarity with the relevant literature and techniques <ul style="list-style-type: none"> acceptable grasp of the subject material ideas stated rather than developed and insufficiently supported by evidence and <i>demonstrated resourcefulness</i> <i>communication</i> of sufficient quality to convey meaning but some lack of fluency and command of suitable vocabulary omission of parts of the subject in question or the appearance of several minor errors average critical awareness and analytical qualities limited evidence of capacity for original and logical thinking
Pass	45%—49% of total marks available	Satisfactory performance - intellectually adequate answer with evidence of some familiarity with the relevant literature and techniques <ul style="list-style-type: none"> basic grasp of subject matter, but somewhat lacking in focus and structure main points covered in answer, but lacking detail some effort to engage, but only a basic understanding of the topic portrayed some development of argument only some critical awareness displayed no evidence or <i>demonstrated resourcefulness</i> appearance of several minor errors or one major error lacking evidence of capacity for original and logical thinking

Grade	Marks range	Characteristics
	40%—44% of total marks available	<p>Acceptable performance - intellectually adequate answer with limited familiarity with the relevant literature and techniques</p> <ul style="list-style-type: none"> • basic grasp of subject matter but limited focus on <i>the assessment task</i> asked • unclear presentation of argument, random layout, with some omissions or inaccuracies in answer • argument insufficiently developed • no evidence or <i>demonstrated resourcefulness</i> • appearance of one major error and minor errors • inclusion of unsubstantiated statements and/or irrelevant material • descriptive rather than argumentative or analytical answer presented • an attempt to solve moderately difficult problems related to the subject material and an attempt to examine the material in a critical and analytical manner only partially successful • an incomplete or rushed answer e.g. <i>oversimplifying critical elements</i>
Fail	35%—39% of total marks available	<p>Unacceptable performance, with either</p> <ul style="list-style-type: none"> • insufficient understanding of the <i>assessment task</i> displayed • failure to address the <i>assessment task</i> resulting in a largely irrelevant answer • a display of some knowledge of material relative to the <i>assessment task</i> posed, but with very serious omissions / errors and/or major inaccuracies included in answer • or answer left somewhat incomplete for lack of time • limited understanding of question displayed • a random layout / underdeveloped structure - not planned sufficiently • poor analytical skills, with an absence of argument random and undisciplined development - limited structure lack of clarity, poor spelling • material of marginal relevance predominating
	<35% of total marks available	<p>Wholly unacceptable performance, with</p> <ul style="list-style-type: none"> • deficient understanding of the <i>assessment task</i> displayed • complete failure to address the <i>assessment task</i> resulting in an irrelevant answer • inadequate knowledge displayed relative to the <i>assessment task</i> posed • or answer left incomplete for lack of time • very poor analytical skills, with an absence of argument • random and undisciplined development - poorly structured answer confused expression, poor spelling • irrelevant material predominating

4. Submitting Your Work & Due Date

The due date of this assignment is **the 24th of March 2025 (i.e., Monday of Week 11)**.

Your work for the assignment must be submitted through the MS5130 Canvas system.

Late submission of your work for this assignment is not advisable. A **late penalty of 10% of the total marks available** will apply for each business day that a deliverable is submitted late. (Exceptions will be made in line with the University's QA209 Extenuating Circumstances Policy and QA208 Alternative Assessment Policy.)

5. Academic Integrity and Referencing

You are required to operate in alignment with the University's QA220 Academic Integrity Policy. Whilst all aspects of that policy apply, please particularly note that:

1. The policy prohibits *collusion* (p. 1), which is when part of your work for this assignment was produced by other students.
2. The policy prohibits the use of generative artificial intelligence *and claiming it as your own work* (p. 2). Therefore, the use of generative artificial intelligence is permitted if and only if it is acknowledged appropriately (including the name of the generative artificial intelligence product and the details of the prompts used).

The policy also requires that you must use appropriate referencing when building on the work of others, **including any work produced by generative artificial intelligence tools such as ChatGPT and Microsoft Copilot**.

Please familiarise with the University's resource on appropriate referencing, provided by the University Library and available at the following web address:

<https://libguides.library.universityofgalway.ie/friendly.php?s=research-skills/citingandreferencing>