

Academic Year	23-24
Semester	2
Module Number	CMM020
Module Title	Data Visualisation and Analysis
Module Co-ordinator	Ines Arana
Assessment Method	MOCK Practical Examination
Date of Exam	20 March 2024
Duration of Exam	3 hours examination plus 15 minutes submission. The quiz has a time limit of 30 minutes, which is included in the time specified above. The quiz must be completed during the first 30 minutes. Once completed, you cannot go back to it.
Open book or closed book?	Selected materials
On campus or Online?	On campus
Number of pages (including cover)	4
Calculator permitted?	No

Submission

You should submit two files to the practical examination dropbox as follows:

- An **R markdown (Rmd) file** called CMM020PE.Rmd – see detailed description below. You should save this file regularly.
- An **html file**, which has been produced by knitting the Rmd file to html.

You should also complete the time-limited **online quiz** which is available on the assessment section of the module area on CampusMoodle. The quiz is only available during the first 30 minutes.

.../Rmd file

Rmd file

This file should contain all your work for tasks 1, 2 and 3 in order. You are required to present the following for each task:

- A heading: the keyword “Task” followed by the task number, at the section level (i.e. use one hash and a blank space before the heading text). For example, for task 3 use the following in a single line, ensuring that the preceding and the following lines are blank.

Task 3

- R block(s): use one or more R blocks for the code for the task. Ensure that either they do not contain any code to view or print large amounts of data or that the relevant R blocks contain the right settings so that output is not produced. No R blocks are required for task 4, which is to be completed online.
- Text: use text outside the R blocks to justify your decisions and to interpret, discuss and evaluate the results obtained or given to you.

Time

You have THREE HOURS to complete the tasks described below, including saving your Rmd file and knitting it onto an HTML file.

Additionally, you will have 30 minutes to submit your work to the practical examination dropbox on the Assessment section of the CMM510 area on CampusMoodle. **You must not carry out any additional work during this time other than submitting.**

Working directory / folder

Create a folder called **CMM020** on the **H:** drive of your exam account.

R and RStudio

All the coding must be completed using R.

Launch R studio.

Create a new R markdown (Rmd) file. Delete unnecessary content from the file (from line 12 onwards).

Save this file on the CMM020 folder that you have just created as **CMM020PE.Rmd**.

The answers to all the tasks should be included in file **CMM020PE.Rmd** as follows:

- Where a task involves discussion or interpretation of results this should be included as text in your Markdown file.

- Where a task involves R code and output, this should be included as a code chunk in your Markdown file. All code must be written in R.

.../Datasets and downloads

Datasets and downloads

You will use three different datasets as follows:

- **arrivals.csv**: a file containing details of ship arrivals into a harbour in 2022, described by the following columns:
 - **Flag**: the country code for the vessel's flag.
 - **FlagCountry**: the name of the country the vessel is registered with.
 - **Name**: the vessel's name.
 - **Draft**: the vessels draft (an indicator of depth) in metres.
 - **Beam**: the vessels beam (width) in metres.
 - **Length**: the vessels length in metres.
 - **Type**: the type of vessel.
 - **Origin**: the name of the port of origin.
 - **Destination**: the name of the destination port.
 - **EstimatedDayArrival**: a number indicating the day of the year when the vessel was expected (between 1 and 365).
 - **ShippingCompany**: the name of the Shipping company in charge of the vessel.

Note that vessels may arrive several times.

- **vessel.csv**: details of individual vessels (one row per ship) with the following columns:
 - **Flag**: the country code for the vessel's flag.
 - **FlagCountry**: the name of the country the vessel is registered with.
 - **Name**: the vessel's name.
 - **Draft**: the vessels draft (an indicator of depth) in metres.
 - **Beam**: the vessels beam (width) in metres.
 - **Length**: the vessels length in metres.
 - **Type**: the type of vessel.
- **numbersPerDay.csv**: a file containing the number of arrivals for each of the 365 days in 2022. It is described by the following columns:
 - **Day**: the day of the year (a number between 1 and 365).
 - **ArrivalNumber**: the number of vessels arriving on the day.
 - **DayName**: a day of the week (Mo, Tu, We, Th, Fr, Sa, Su).

Download the 3 files above from the CMM020 assessment area on CampusMoodle. Put these 3 files in the CMM020 folder that you created earlier.

Load these 3 files onto RStudio using R code.

Libraries

Load the following libraries:

- tidyverse
- treemapify
- igraph
- ggraph
- ggwordcloud

.../Tasks

Tasks

Undertake the following FOUR tasks using R code for tasks 1 to 3. Task 4 requires no coding and should be undertaken first.

- **Task 1** – Analyse the following attributes for both the **vessel** and the **arrivals** datasets:
 - Draft.
 - Flag.Use at most 4 plots in total to show the distribution of the above variables. Choose colours "tomato", "skyblue1" , "coral3" "forestgreen" if you can. Discuss the results.
How do distributions compare between the 2 datasets?
- **Task 2** – Produce a plot for each of the scenarios below and comment on the plots, suggesting improvements where appropriate.
 - A line plot of number of arrivals per day using one line per day of the week. Use the **numbersPerDay** dataset.
 - A scatterplot of Beam vs Draft according to Type. Use the **vessel** dataset.
 - A word cloud of shipping company. Use the **arrivals** dataset.
 - A dendrogram of country name, vessel name and length for vessels where the country flag is Turkey. You can get the dataset of vessels from Turkey with the following code (which leaves the resulting dataset in variable vessels2)

```
vessel2 <- vessel |> filter(FlagCountry == "Turkey")
```

- **Task 3** – check whether a linear regression model is appropriate for predicting Beam (response) from Length (predictor) using the **vessel** dataset. Obtain the linear model and perform checks to further confirm or reject its suitability. Obtain a 2nd model for Beam (response) from Length and Draft (predictors). Comment on the new model.
- **Task 4** – Complete the online quiz which is available in the Assessment section of the module area on CampusMoodle. The time limit for this quiz appears in the cover page.

End of Practical Examination