N.B. Please make sure to carefully read this document before beginning work on the project. The project is worth 70% of the final grade for the first part of the module i.e. 35% of the entire module. If you have any further questions, jusk ask.

Reports should not be difficult to write, as they simply represent an honest account of the work done. Ideally, a report has a voice, is interesting to read and tells the story of the work. For this project 10 pages would be an indicator of the expected length. The deadline for submission of the word processed document and accompanying files is Week 9, Friday 16th November - 23:55.

Notes:

1.) Plots should only be included in the report when they serve the purpose to illustrate a key point of the analysis i.e. not fill up the pages.

2.) Large sections of R code should not be included in the report, a simple referencing mechanism at the end of the paragraph is fine e.g. (thisFile.R lines 75-90).

3.) Plots not included in the report can be submitted in a separate folder or appendix along with the code used to generate them e.g. as part of this section of the analysis the plots in folder x were generated etc..

4.) Plain English is fine for the style of writing as your audience is the lecturer (**n.b.** please use the spell-check as marks cannot be awarded unless the text is clearly understood).

5.) The submitted code will be carefully examined against the following criteria:

- Complexity of operations being performed.
- Whether it is easy to understand e.g. variable names are informative (**not** temp161), indentation is used and comments are placed before different sections (line-by-line commenting obscures the structure of the code and is not necessary i.e. the variable names and indentation should suffice).
- Code sourced from elsewhere is appropriately cited (bear in mind your R skills are being examined).
- The code works correctly and is **not** reliant on hardcoded elements.

6.) Students are encouraged to present their reports in a professional manner, however bear in mind that the majority of the marks for written sections will be based on the structure of the paragraphs e.g. points made (what was done), flow of reasoning (motivation for steps), depth of analysis and conclusions reached. It is therefore better to focus attention on these areas.

7.) Anyone wishing to use the project as an opportunity to further explore a different presentation technology may write their report using R Markdown. Students availing of this option will have an extra 7 days to complete the project. This option should only be chosen by those who are confident that a presentation of an acceptable standard can be produced i.e. comparable to an ordinary word-processed document.

Deliverable:

Zip File containing:

Word Processed Document called “report” followed by an underscore “_” and the student number.

R files: Plots.R, BikeInfo.R, DublinBus.R, Footfall.R and Reflection.R

Any other files necessary to run the code.

Note: Statements made in the report will only be taken into account when there is runnable code available for confirmation.

Grade breakdown

The breakdown of the grade for all the tasks except 5.2, is half the marks for the R code and half the marks for the text in the report. These marks are based on the criteria outlined in the notes above.

Task 1

(Warm-up) Quarterly Economic Indicators

20% (Submission File - Plots.R) 1 Page

1.1 (**10**/**20**) Recreate the plot from file Q1a.pdf using the data from the first 3 columns in the file "*dublin employment trends.txt".* (It is **not** necessary to save the plot to the file system). Discuss any interesting observations in the plot.

1.2 (**10**/**20**) Recreate the plot from file Q1b.pdf using the data from the first 3 columns in the file "*dublin property trends.txt".* (It is **not** necessary to save the plot to the file system). Discuss any interesting observations in the plot.

Task 2

(Real-time) bike Info

20% (Submission File - BikeInfo.R) 2 Pages

Sign up to JCDecaux. Using the method outlined on Webcourses (I will put that up shortly), import some real-time bicycle information from a city of your choice (e.g. Dublin has information on 100 bike stands).

2.1 (**10**/**20**) Summarise the data available in the different variables e.g. suppose that you are having a 5 minute conversation with a manager and they are interested in obtaining a short but reasonably complete overview of the current situation with bikes.

2.2 (**10**/**20**) Think of a particular use-case where you or someone else (possibly a time-pressured individual) wishes to obtain relevant information from this dataset. Use qplot to create a clear and simple graph (3 variables) that facilitates an at-a-glance interpretation of the data for them.

Task 3

(Scale) Dublin Bus

20% (Submission File - DublinBus.R) 3 Pages

Use R to import the Dublin Bus GTFS dataset from **DublinBus.zip**. Info on the GTFS format may be found here.

3.1 (**10/20**) Shape the data and conduct an initial exploratory analysis of the entire dataset, focussing on all the routes.

3.2 (**10/20**) Pick a bus route (forward and reverse directions) and drill down into the details of it extracting statistics to answer a series of questions e.g. suppose that you were moving to a house along the route and relying on this bus to get to and from work on time everyday, what questions would need to be answered to make you an expert on the route?. Create a plot for the most interesting piece of this analysis.

Task 4

(Retail) Footfall

20% (Submission File - Footfall.R) 2 Pages

Use R to download and import the Footfall dataset from the year 2013.

4.1 (**10/20**) Pick two cameras (e.g. one on the Northside and the other on the Southside) and analyse the similarities and / or differences between the footfall patterns over the course of the 52 weeks in the year.

4.2 (**10/20**) Move outside the dataset and scrape some data from the web. Use this data to create a comparative plot against the footfall data from one of the cameras.

Task 5

Reflection

20% (Submission File - Reflection.R)

Key parts of any data analysis workflow include: 1.) Scheduling sessions to work on parts of the project. 2.) Keeping track of outcomes. 3.) Setting aside time at the end of the project to reflect on the results achieved. As such, the student is advised to maintain a personal log or diary of their activities.

5.1 Analytic (**10/20**) (1 Page)

The previous tasks have covered data generated by the activities of the citizens of Dublin in the course of their daily pursuits. This part of the project looks at the data generated during the completion of the project and applies data wrangling to the points outlined above in a bid to retrospectively analyse performance.

RStudio maintains a log of console activities in a file called history_database in a folder named RStudio-Desktop (Windows) or rstudio-desktop (Ubuntu). The format of the file is the time in mili-seconds followed by the command entered. Before beginning the project, the student is advised to confirm the existence of this file on their system and back it up at appropriate intervals to avoid data loss.

Task: Import a copy of this data (i.e. **not** the original file) into R and conduct a basic analysis of general patterns / evident trends.

5.2 Personal Reflection (**10/20**) (1 Page)

A reflection looks back on the experience of completing a project with a view to getting a sense of perspective and absorbing relevant lessons (positive or negative) to take forward to future projects. Topics to include might be 1.) challenges overcome (work-arounds), 2.) efficient work practices 3.) areas of frustration 4.) time management. A useful way to structure a reflection is by using the four I's e.g. **Identify** an element during the project (past) - results, quality, emotions, productivity. **Investigate** - determine the source of this (present). **Incorporate** - based on the investigation, is there a lesson to absorb (future)? **Improve** - the goal of reflection.

urls

https://developer.jcdecaux.com/#/opendata/vls?page=getstarted

https://developers.google.com/transit/gtfs/reference/

https://data.smartdublin.ie/dataset/pedestrian-footfall-index-in-dublin-city-centre

https://data.smartdublin.ie/dataset/8204be0a-6348-459e-96e9-65bb75600ec3/resource/384fe47a-2f25-4f52-8fc5-8e61899951e9/download/pedestrianfootfall2013.ods