# INST0065 Data Visualization and GIS

## Assessment 2, 2022-23

### IMPORTANT NOTICE

This assessment forms part of your degree assessment. You may re-use work that has already been demonstrated as part of class practical exercises, but all additional work must be done entirely on your own from start to finish.

- You must not collaborate or work with other students at any stage
- You must not send or show other students your work
- You must not ask other students for help, or ask to see their work. As well as being against regulations, this is unfair to the other student concerned, since it may lead to them being accused of plagiarism
- You must not seek help from friends, relatives, online discussion groups (such as StackOverflow or similar), or anyone other than the lecturer and/or lab demonstrators for INST0065

If you are having difficulty in attempting any part of this assessment, you should discuss this in the first instance with the INST0065 lecturer.

### Submission deadline

The assignment should be submitted by **3pm (UK time) May 2nd, 2023**

### Submission link

Work should be submitted via: https://moodle.ucl.ac.uk/mod/assign/view.php?id=2798486

## Assignment

The website openplaques.org "... catalogues, curates, and promotes commemorative plaques and historical markers (often blue and round) installed on buildings and landmarks throughout the world."

The website allows visitors to explore a database detailing over 50,000 plaques around the world; it also provides a downloadable version of the base data, using an unrestrictive license.

You should create an RMarkdown file (see below) that uses this data to produce at least three data visualizations, using R. Data visualizations include various sorts of chart. You may also display information in the form of a table should you wish to do so.

Your RMarkdown file should incorporate commentary in which you cover these areas:

- Introduction
    - Identify and describe the original data sources
- Methods
    - Describe the purpose of modifications and changes that you make to the data (that is, you should say *why* you are merging tables or modifying values, or filtering for a given subset etc., but you do not need to explain what each R function is doing. The R code can contain comments if helpful to explain the code)
    - Explain the purpose of visualizations that you are going to produce
- Results
    - Show your visualizations
    - Discuss any features or patterns in the data that you think are notable
- Conclusion
    - Note findings, and comment on other visualizations that you think may be useful to produce using this data if you had more time and space etc

The commentary might be on characteristics of the data itself (what certain fields mean, whether the data has an effective structure etc), on characteristics of the plaques (how may plaques are there detailing different types of people, etc), or on some other aspect of the data that you wish to describe. The intention is to allow students to explore the data in different ways.

Any external references that you include should be detailed in a bibliography, as with any other piece of academic work.

## RMarkdown file requirements

The assessment requires you to submit one RMarkdown file. A template file has been provided, that uses the *read_csv()* function provided by *tidyverse*[1] to load the data. Note that the template uses *setwd()* to set the working directory; an example is shown, but you will need to adjust this for your work.

The same data file as used in the example will be used for grading, and thus your work must be compatible with this data rather than using an externally modified version.

The template file is included on Moodle, and is included as an appendix here.

The RMarkdown file also contains an initial blbliography with references set up in the document header. Bibliographies can be incorporated as regular marked-up text if you prefer.

After submission, the RMarkdown file will be run, and the results assessed. You should **ensure that the RMarkdown file knits properly and generates html output** before submitting.

## Word limit

This assignment has an upper limit of 2000 words.

- The word limit does not include bibliographies, or footnotes etc
- The word limit does not include R code contained within code blocks in your RMarkdown file, or the output of the code
- It is recognised that RMarkdown does not easily permit a word count to be made; it is suggested that text excluding code chunks etc is copied to a text editor or similar, and that a word count is done there

[1] The RMarkdown template loads the tidyverse library. It is assumed that the installation of R that you are using already has the library installed. If it does not, you should install it using the R console command: `install.packages("tidyverse")`

## Grade descriptors

### A (70% +)

A good range of visualizations which go beyond those produced in class practical sessions, are well presented, and form a coherent narrative in exploring aspects of the dataset.

Good description of the dataset and processes followed; clear evidence of critical thought in the selection and description of visualizations. RMarkdown output is well structured, with full references to source data, and to any additional materials cited.

### B (60-69%)

Visualizations are well presented, and go beyond the example ones produced in class practical sessions.

Good description of the data and justification of the visualizations produced, and evidence of critical thought. Acceptable level of literacy and bibliographical citation where required.

### C (50-59%)

Adequate production of visualizations, but might not go substantially beyond those covered in class practical sessions.

The descriptive content explains ideas to an acceptable standard, but may evidence some of the following: lack of depth, poor expression, little evidence of self-directed thought or curiosity, incomplete or poorly constructed bibliography, low standard of literacy.

### D (49% or below)

Unacceptable. May indicate significant ignorance or misunderstanding of the data, fail to produce suitable output, or rely heavily on other people's work. Where plagiarism is detected, appropriate action will be taken in line with UCL regulations

## Appendix: template RMarkdown file

````
---
title: "INST0065 Assessment 2"
author: "do not enter your name"
date: "14/03/2023"
output: html_document

# bibliography style; see: https://github.com/citation-style-
language/styles

csl: 'https://raw.githubusercontent.com/citation-style-
language/styles/master/apa-6th-edition.csl'

# for other examples of references
# see
https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html

references:
- id: openplaques2023
  title: openplaques.org database
  author:
  - literal: openplaques.org
  URL: 'https://openplaques.org/data/'
  type: webpage
  issued:
    year: 2023

- id: duke-williams2023
  title: INST0065 course notes
  author:
    - family: Duke-Williams
      given: Oliver
  URL: 'https://moodle.ucl.ac.uk/course/view.php?id=27657'
  issued:
    year: 2023

- id: r-error-messages2023
  title: r-error-messages
  author:
  - literal: programmingr.com
  URL: https://www.programmingr.com/r-error-messages/error-in-contrib-
urlrepos-source-r-markdown/#
  type: webpage
  issued:
    year: 2023


---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
```

# INST0065 Assessment 2

This is a template RMarkdown file, for use with an assessment for INST0065
Data Visualization and GIS (@duke-williams2023)
````

````
## Data

The data for this markdown file are taken from
[https://openplaques.org/data/](openplaques.org) (@openplaques2023)

see also (@r-error-messages2023)

```{r load-data, include=FALSE}
# include=FALSE means that code will run, but results will not be shown

# This section loads all the data; you do not need to use all tables

# NB This section uses read_csv(), which is included in tidyverse
# Students will need to alter directory path
setwd("N:/work/teaching/inst0065-dataviz-gis/assessments/2022-
23/assessment2/")
plaques <- read_csv("open-plaques-gb-2021-05-14.csv")
```

For initial testing, we can look at a sample of data, to confirm that it
has loaded properly

```{r test-data, include=TRUE}
# set to include=TRUE for testing, and then revert to include=FALSE
head(plaques)
```

<!-- student code to follow beyond this point -->


<!-- add references here -->
<!-- see
https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html -
->

# Bibliography
````