

# DAPS&CO

## Data Access & Regulation, Module 3

**Instructor:** Robert Fahey <[robfahey@aoni.waseda.jp](mailto:robfahey@aoni.waseda.jp)>

### Assignment 1      8 Feb 2021 (Mon)

#### **Part One:**

In a short email (just a couple of paragraphs), describe the data from your existing project.

(This will usually be the data from a project you previously undertook for this module, or, if you prefer, it could be some other data you have collected and worked with for another project, e.g. your undergraduate thesis.)

When we say “describe” data, this means doing two things:

- 1) Explain what the data is. Where did it come from? What’s in it? Why did you collect it?
- 2) Explain the technical aspects of the data. For example:
  - Is it one table of data, or many?
  - Is it stored in an R file (.RData) or some other kind of file?
  - How many observations (i.e. rows of data) do you have?
  - What variables (i.e. columns of data) are included?
  - If it’s time series data, where does it begin and end?

Knowing these basic facts about a dataset is essential to starting to figure out what the best way to store, access and process that dataset is. Please write a description of your project’s data file(s), including all of these details (if relevant) and submit it to me by email.

## **Part Two:**

Next, I want you to repeat the same process - but this time, I first want you to find some new data that's interesting to you.

I strongly recommend going to **kaggle.com** and signing up for that site (it's free and very quick). It'll give you access to a huge number of datasets, which you can search through and download, and even use in your own research. (You don't have to use Kaggle if there's another dataset you're really interested in - this is just a suggestion.)

Find a dataset you think is interesting for some reason, and repeat the process from Part One - write a short description of the data, including both a few lines explaining what it is and why you chose it, and a technical description of how much data there is and what format it's stored in.

## **Submission:**

**Email both of these data descriptions to me ([robfahey@aoni.waseda.jp](mailto:robfahey@aoni.waseda.jp)) before the next class (Tuesday, Feb 9th, 9am CET).**

## **Late Submission:**

*Late submissions without a reason (illness etc.) will be accepted (with a grade penalty of 5% per weekday) up until the day of the module's final class (Tuesday, Feb 23rd). Any submission after that date will be rejected and your grade for this assignment will be zero.*