# Information Visualisation

## Exploration & Communication Assignment (45% of grade)

The goal of this assignment is to design and create **three distinct visualisations** that illuminate some aspects of a complex quantitative dataset. A number of datasets have been provided – these have been selected because they contain a variety of categorical and quantitative attributes that lend themselves to different visualisation approaches. You must use one of the provided datasets.

Take a look at the different datasets provided and choose one to perform your analysis on. Develop a couple of initial questions you wish to address (e.g. What is the relationship between height and weight for Olympic athletes? How does it vary across sports?) and start to create visualisations to answer them. Your questions will likely evolve and be refined as you understand the dataset better. You may also wish to create initial exploratory visualisations (e.g. histograms, scatter plot matrices) to understand your data better and help frame your questions.

You may perform initial analysis in another tool (e.g. Tableau) but the submitted visualisations **must** be presented in Vega-Lite. You may perform data manipulation outside of the Vega-Lite environment, e.g. using Python, but you must make clear what transformations / filtering you have performed and make the processed data set available.

Your submission should contain distinct visualisations that address three different questions related to your dataset. By distinct I mean that they could not be simply generated using small multiples or using the repeat operators. Three bar charts that show the average height, average weight, and average age of Olympic athletes would not count as distinct visualisations. You must include a discussion of your visualisations – why you chose particular encodings, what alternatives you considered, and what the pros and cons are of the alternatives. You should give also provide some indication of the insight gained from the visualisations – what do they demonstrate and what have you learnt from them? See the sample solution for an example of what is expected in these sections.

The final visualisations must be presented in a html page (or pages) using Vega-Lite (or Vega if you prefer). A sample template is provided – if you're not familiar with HTML you can just open the files in a text editor and replace my text with yours. An online editor may help with basic HTML formatting.

I expect that you will generate many more visualisations than the three you submit. Make it clear that you have considered alternatives in your discussion of design issues. It's easy to create radically different visualisations with minor changes in Vega-Lite so I recommend taking screen shots throughout your process so you can remember all of the alternatives you explored.

The following will be considered when grading the assignment:

- Design considerations. What encodings have you used and are they appropriate for the task? Are they effective and expressive? What alternative designs have you considered? What are the advantages and disadvantages of the design you have submitted over the alternatives?
- Small touches (titles, axis labels, legend placement etc)
- Appropriate data filtering, aggregation, transformation
- Insight – what have you learned from your visualisation? Does the visualisation support your observations?

**Submission**

The key submission details for the assignment are as follows:

**Submission date:** Monday 23rd March 2020 before 23:55

**Submission format:** Submissions should be a zip file containing three html files, following the provided template.

**Late submissions:** Late submissions will be docked 5% per day or part thereof. They will only be accepted until Sunday 5th April 2020 at 23:55.

**Datasets:**

These datasets have been selected because they contain a variety of categorical and quantitative attributes that lend themselves to different visualisation approaches. Note that the Kickstarter data is extremely large and consequently may benefit from pre-processing to filter / aggregate data. You should store the dataset you are working with on Github or similar.

NCT Failure Rates 2016 - https://data.gov.ie/dataset/2016-make-model-year-failures-at-nct

Animal Age Dataset - http://genomics.senescence.info/download.html#anage

Arabica coffee ratings -https://github.com/jldbc/coffee-quality-database/tree/master/data

Kickstarter project details - https://www.kaggle.com/kemical/kickstarter-projects#ks-projects-201801.csv