**FLIGHT DATA ANALYSIS**

The 2009 ASA Statistical Computing and Graphics Data Expo consisted of flight arrival and departure details for all commercial flights on major carriers within the USA, from October 1987 to April 2008. This is a large dataset; there are nearly 120 million records in total, and takes up 1.6 gigabytes of space compressed and 12 gigabytes when uncompressed. The complete dataset along with supplementary information and variable descriptions can be downloaded from the Harvard Dataverse at <https://doi.org/10.7910/DVN/HG7NV7>

Choosing any subset of (at least two) consecutive years and any of the supplementary information provided by the Harvard Dataverse to answer the following questions:

*1. When is the best time of day, day of the week, and time of year to fly to minimise delays?*

*2. Do older planes suffer more delays?*

*3. How does the number of people flying between different locations change over time?*

*4. Can you detect cascading failures as delays in one airport create delays in others?*

*5. Use the available variables to construct a model that predicts delays*.

All questions have been answered using BOTH R and Python for all tasks.