CSPro Airports Lab

In this lab you will learn:

- How to use the Excel to CSPro tool.
- How to write a simple batch processing application.
- How to use the Tabulate Frequencies, Export Data, and Sort Data tools.
- How to use write batch export applications.

There are two datasets that will be used in this lab:

- List of worldwide airports with their IATA codes.
  - The original data[1], in JSON, was concerted to Excel and slightly modified.
- List of countries and territories and their ISO codes.
  - The original data[2], in CSV, was concerted to Excel.

These datasets, along with dictionaries to describe them, can be found on Josh's website.

Activity 1: Create CSPro data files from the Excel data

Using the provided Excel files and the CSPro data dictionaries, create data files that can be used by CSPro to read the airport and country data.

Activity 2: Write a batch application to merge and process the data from the airport and country data files

Batch applications are CSPro applications that only contain logic and do not require interaction by a user. They are generally used to parse, recode, or edit data. Batch applications accept an input data file, perform work on that file, and then write out the modified cases to a new file. The contents of the input data file are not modified.

In this lab, we will:

- Modify the airports data dictionary to add fields for the country name, continent name, and a numeric representation of the airport size.
- Using the countries data dictionary as a lookup, query the country name based on the airport's country code. If the country code is not in the countries file, then write out an error message and don't output that airport. (See the `skip case` statement.)
- Recode the continent code to add the continent name to the output data file.
- Recode the airport size so that it can be used as a numeric using the following scheme:
  - Small: 1
  - Medium: 2
  - Large: 3

[1] https://github.com/jbrooksuk/JSON-Airports/blob/master/airports.json
[2] https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv

Activity 3: Look at the data with the Tabulate Frequencies and Export Data tools

You may want to:

- Use the Tabulate Frequencies tool to find out how many airports exist in your country. You can use the Universe option to filter your query.
- Use the Export Data tool to output, to CSV format (a format readable by Excel), all of the information about the airports in your country.

Activity 4: Learn about Production Runner

The CSPro Production Runner is a tool for executing many CSPro applications, or DOS batch files, in a systematic way. It was designed for Kenya for the processing of the 2009 census tables. There were many things that needed to occur before the tables could be produced, including processing and editing the data, and there were many gigabytes worth of data to process, so this tool helped automate execution of the programs so that the tables could be generated overnight without human intervention.

The Production Runner can be accessed from Tools menu within CSPro. In this lab, we will:

1. Write a batch application that looks up an airport code and writes to an external file the latitude and longitude of the airport. (Use the `write` or `filewrite` statements.)
2. Write a batch application to add to the airports data file the distance from the above airport to each airport in the file. We will use `gps` function's distance functionality, which returns the great circle distance (in meters) between two sets of latitude and longitude.
3. Use the Sort Data tool to sort the above file by distance.
4. Use the Export Data tool to export the sorted data to a format that can be read by Excel.
5. Use the Production Runner to automate the above four steps.
6. Add to the Production Runner replacement parameters so that we can run these steps on multiple airports. This requires adding the %%parameter%% replacement text to your PFF files.
7. Add a DOS batch file to delete the working files created during the processing of the above steps.

(Unfortunately, there is a bug with the Sort Data tool in that it overwrites a PFF if it has the default generated name based on the SSF file, so make sure that the SSF and PFF files have different names.)

Activity 5: Create a data entry program to record flight journeys

We want to keep track of journeys flown around the world. Each journey will consist of two or more airport codes, indicating each segment of the journey. In addition, we will keep track of the date of the journey. No more than one journey can begin in a given day. Add these checks to the data entry program:

- Prefill the date with today's date.
- Make sure that the journey date is valid and not in the future.
- Ensure that all entered airport codes are valid.
- Do not allow the operator to enter the same airport code twice in a row.
- Ensure that at least two airport codes are entered (to properly complete a journey).
- When the operator tries to end the journey, check if the beginning and ending airport codes are the same. If not, query the operator as to whether this was intentional.

Activity 6: Obtain the code basis for exporting journey data

Some of the CSPro tools, including the Tabulate Frequencies and Export Data tools, are simply interfaces that write out CSPro batch applications that execute the desired task. The Export Data tool has a feature on the Options menu, *Copy Logic to Clipboard*, which allows you to see what code CSPro would execute to run the export options that you have selected. Obtain the code for the following export:

- Exporting all IDs and records in the journey dictionary.
- Exporting multiple occurrences to separate records.
- Exporting to CSV format.

Activity 7: Create a batch export application

Create a new batch application, using your journey dictionary as the base file. You will run this program on data that you have entered using the data entry program. Use the export code that you copied above as the basis for your batch application. We want to export the following information:

- The journey date
- For each segment:
    - Segment number
    - Starting airport code, airport name, and country name
    - Ending airport code, airport name, and country name
    - The distance between the airports in both kilometers and miles

Only the journey date and airport codes exist in the journey data file. The airport names, country names, and geographic coordinates can be obtained from the airport lookup file.

You may want to use a working storage dictionary to help with this export. A working storage dictionary is a dictionary attached to your application that has no external file associated with it. You can put temporary variables in this dictionary in the same way that you can add numeric and string variables to your application in the PROC GLOBAL section of your logic. Working storage dictionary items can be exported along with items from your main dictionary and lookup file.

Hint 1: You may have to access your airport lookup file twice for each segment, once for the starting airport and once for the ending airport. However, because the end of one segment is the start of the next segment, you can optimize your code to reduce the number of `loadcase` statements necessary.

Hint 2: One kilometer is 0.621371 miles.