

Adding Carbon to the Equation in Online Flight Search Datasets

Datasets available at: <https://doi.org/10.25338/B8FS5B> and
<https://doi.org/10.25338/B81S5M>

(This dataset supports report **Adding Carbon to the Equation in Online Flight Search**)

This U.S. Department of Transportation-funded dataset is preserved by the University of California in the digital repository Dryad (<https://datadryad.org/>), and is available at <https://doi.org/10.25338/B8FS5B> and <https://doi.org/10.25338/B81S5M>.

The related final report **Adding Carbon to the Equation in Online Flight Search**, is available from the National Transportation Library's Digital Repository at <https://rosap.ntl.bts.gov/view/dot/54762>.

Metadata from the Dryad Repository record for Popular flight costs and emissions data:

Citation:

Amenta, Annamaria; Sanguinetti, Angela (2019), Popular flight costs and emissions data, Dryad, Dataset, <https://doi.org/10.25338/B8FS5B>

Abstract:

This dataset gives information on 67 popular destinations for UC Davis employee business travel. For each destination, we give a selection of possible flights, collected in Summer and Fall 2019, each annotated with emissions as estimated by Atmosfair.de, and also the probability that each flight would be chosen, based on a model of employee flight choice preference that we compute from survey data.

Publication Date:

December 12, 2019

Methods:

The destinations were chosen based on data provided by the University's Accounting and Financial Services, and it cannot be sufficiently anonymized. We do include a file listing top destinations and the number of times which each one was flown to (`topDestinations.csv`). There is also a zip file, `scoredCombinedFlightOptions.zip`, containing 67 csv files, one for each of the popular destinations (some were excluded for various reasons). And, there is a README file, describing the files for the destinations.

Funding:

U.S. Department of Transportation, Award: USDOT Grant 69A3551747114

Dataset description:

This dataset contains 1 .zip file collection below.

doi_10.25338_B8FS5B_v1.zip:

The .zip file collection contains 1 .csv file, 1 .txt file, and .zip file collection, which are further described below.

- topDestination.csv
 - The .csv, Comma Separated Value, file is a simple format that is designed for a database table and supported by many applications. The .csv file is often used for moving tabular data between two different computer programs, due to its open format. The most common software used to open .csv files are Microsoft Excel and RecordEditor, (for more information on .csv files and software, please visit <https://www.file-extensions.org/csv-file-extension>).
- README.txt
 - .txt: The .txt file type is a common text file, which can be opened with a basic text editor. The most common software used to open .txt files are Microsoft Windows Notepad, Sublime Text, Atom, and TextEdit (for more information on .txt files and software, please visit <https://www.file-extensions.org/txt-file-extension>).
- ScoredCombinedFlightOptions.zip
 - This .zip file collection contains 68.csv files that can be found within the ScoredCombinedFlightOptions Folder, which can be found twice in .zip file. The files are titled “._Scorded_XXX.csv,” with the XXX representing three letters.

Metadata from the Dryad Repository record for Employee airline travel preferences survey data, UC Davis GreenFLY project:Citation:

Amenta, Annamaria; Sanguinetti, Angela (2019), Employee airline travel preferences survey data, UC Davis GreenFLY project, Dryad, Dataset, <https://doi.org/10.25338/B81S5M>

Abstract:

This survey data is from a study exploring the potential to promote lower-emissions air travel by providing consumers with information about the carbon emissions of possible flight choices in the context of online flight search and booking. We surveyed over 450 faculty, researchers, and staff at the University of California, Davis, and asked them to choose among hypothetical flight options for a domestic and an international university-related business trip. Emissions estimates for different flight alternatives were displayed as prominently as price; this simple intervention has been promoted in several demonstration projects, including GreenFLY, a demo we created at UC Davis.

Publication Date:

December 7, 2019

Methods:

The flight choice experiment involved an online survey in which UC Davis employees were asked to make a series of binary discrete choices between roundtrip flight alternatives, that varied in terms of cost, carbon emissions, layovers (0 or 2: one layover each way), and airport

(SMF or SFO), for two hypothetical UC Davis-related business trips, one to Washington, DC and the other to London. We based these hypothetical scenarios (trip destinations and attribute levels of flight alternatives) on data about actual UC Davis employee air travel.

For the layover flight alternatives, we created eight possible cost-carbon combinations, using each cost level and each carbon level twice, and not repeating any pairing. There are many ways to do this, and we chose one which tended to pair high cost with low carbon, to create trade-offs. Our eight layover flights to DC appear in Table 2. The same cost-carbon pairings were used for layover flight alternatives from SFO.

We organized the flight alternatives into sets of two for the choice experiment questions. Criteria for pairing flight alternatives were as follows:

1. Every flight alternative should appear roughly the same number of times in the survey,
2. The distribution of kinds of flights in the questions (eg. layover out of SFO) should match the distribution in the entire set,
3. Avoid questions in which the two flights have the same cost, or the same carbon, and
4. Focus on pairs that might have competitive utility (e.g. an alternative that is lower cost, lower carbon, nonstop and out of SMF is likely to be selected in most cases, so it is not useful for understanding potential trade-offs).

From this, we created seven "buckets" of questions for Washington, and seven for London, and asked each participant a randomly-chosen question from each bucket. We made an error in the online questionnaire-design software, which caused a random bucket to be skipped for London. Nonetheless, each flight option appears frequently (between 40 and 120 times) in the questions we asked.

The original data output from Qualtrics was processed into a format suitable for processing with the mlogit package in R.

Usage Notes:

The R file we used for analysis is included. It is called ModelsFromSurveyData.R

The data file is called Organized_Survey_Data.csv

There are two rows in the file for each question presented to a user, one for each of the two alternative flights the question asked the user to compare.

The columns in the file are:

rowIndex - The user's ID from the Qualtrics data. In the original data each user's entire questionnaire was one long row.

columnIndex - The question column from the Qualtrics data. Each question was a column in the original data.

QID - The question ID in the Qualtrics file for each question. These should be in 1:1 correspondence with the columnIndex values.

Code - The descriptive name of the specific flight corresponding to this row.

L.N - Layover or non-stop
Carbon - CO2 emissions for the flight, in pounds
Cost - cost in dollars
Airport - SMF or SFO
Choice - the flight the user chose has Choice=1, the other has Choice=0
preferred_airport - SFO or SMF. About 10% of participants live in the Bay Area and would prefer to fly out of SFO.

Other columns included other demographic data, which hopefully are self-explanatory.

Funding:

U.S. Department of Transportation, Award: USDOT Grant 69A3551747114

Dataset description:

This dataset contains 1 .zip file collection below.

doi_10.25338_B81S5M_v2.zip:

The .zip file collection contains 1 .csv file and 1 .R file, described below.

- Organized_Survey_Data.csv
 - The .csv, Comma Separated Value, file is a simple format that is designed for a database table and supported by many applications. The .csv file is often used for moving tabular data between two different computer programs, due to its open format. The most common software used to open .csv files are Microsoft Excel and RecordEditor, (for more information on .csv files and software, please visit <https://www.file-extensions.org/csv-file-extension>).
- modelsFromSurveyData.R
 - The .R file is a programming script written by the R programming language. .R files can be opened with any text editor.

National Transportation Library (NTL) Curation Note:

As this dataset is preserved in a repository outside U.S. DOT control, as allowed by the U.S. DOT's Public Access Plan (<https://doi.org/10.21949/1503647>) Section 7.4.2 Data, the NTL staff has performed **NO** additional curation actions on this dataset.

NTL staff last accessed this dataset at <https://doi.org/10.25338/B8FS5B> and <https://doi.org/10.25338/B81S5M> on 2021-03-10.

If, in the future, you have trouble accessing this dataset at the host repository, please email NTLDataCurator@dot.gov describing your problem. NTL staff will do its best to assist you at that time.