# Introduction

# Data Source

## Origins

The data sources used in this project are publicly provided by the Bureau of Infrastructure, Transport, and Regional Economics (BITRE); a division of the Department of Infrastructure, Transport, Regional Development and Communications.

BITRE publishes various transportation statistics, including a wide variety of Aviation statistics. This report focuses on Domestic aviation activity and International airline activity. Time series data is provided, with monthly information for routes provided as far back as 1984.

One caveat with this dataset is that reporting of data by Qantas Airways changed in 2003 – for example, a flight reported as Adelaide to London in January 2002 (no direct services between these two cities), would be reported in January 2003 as either Adelaide to Singapore or Melbourne/Sydney to London. This makes it difficult to directly compare data before 2003 to data after 2003. To deal with this in my visualizations, most of my visualizations will only use data from 2003 to 2019 – if required, data before 2003 will not be compared to data after 2003.

## Processing

There was no element of data collection or web scraping in this project.

BITRE provided the data in multiple Excel spreadsheets – for example, the international data was split into 1985-1998, 1989-1993, 1994-1998, 1999-2003, 2004-2008, and 2009-2020.

To combine these spreadsheets together, the Python package **pandas** was used. Multiple spreadsheets are loaded at once using the **read\_excel** function, and then the data frames are appended together to create one large data frame.

Since the data is inherently multi-dimensional (Origin, Destination, Time); I also used the package **xarray** to easily work with the dimensions. Xarray provides a powerful n-dimensional data structure with dimensional labelling. These structures are much easier to work with for analysis and visualization purposes.

# Results

## Analysis

## Visualizations