

DEN

Data Exchange Network

The “AIRNET2000” Data and AIRNET Program

Zachary Neal
Michigan State University
East Lansing, Michigan

1. Overview

The AIRNET2000 dataset was derived from the U.S. Bureau of Transportation Statistics’ Origin and Destination Survey, using the AIRNET program. It includes four longitudinal networks that describe passenger air traffic among 108 U.S. metropolitan areas annually, from 2000 – 2011 (12 waves). The route network describes all passenger movements between cities, while the origin-destination network describes passenger movements between an initial origin and final destination cities, omitting intermediate connections. The business and leisure networks are subsets of the origin-destination network, reflecting the origin-destination movements of passengers likely traveling for business or leisure, respectively. A distance matrix, recording cities’ great circle distances in kilometers, is also provided to facilitate spatial visualization and analysis.

The AIRNET2000 dataset represents a sample of the network data that can be generated using the AIRNET program, written for Stata. The program can be installed by typing ‘ssc install airnet’ in the Stata command line, and allows users to specify options to customize the resulting networks. Following installation, a detailed helpfile documenting use of the AIRNET program is available by typing ‘help airnet.’ Using the AIRNET program to produce network data provides access to:

- Wider longitudinal timeframes: From 1993 to present
- Narrower intervals: Quarterly, rather than annual
- Alternate nodes: US airports or US metropolitan areas
- Alternate formats: Edgelist or matrices

- Detailed decomposition of passengers’ routes by use of connections and layovers
- Alternate parameters for identifying likely business & leisure passengers

2. Data Collection

The Origin and Destination Survey is conducted quarterly by the U.S. Bureau of Transportation Statistics, and provides details on a 10% random sample of all passenger air traffic in the United States. The raw data is provided in three nested data tables, each providing a different set of variables, and is available at: www.transtats.bts.gov/Tables.asp?DB_ID=125. The AIRNET2000 dataset was produced by executing the command “airnet, alpha(.05) minfare(20) maxfare(5000) metro(new) matrix distance” in Stata for each quarterly data release from 2000Q1 through 2011Q4, then aggregating the resulting networks annually. On a 2.5 GHz processor with 4 GB memory, this requires approximately 11 hours.

The basic networks reflect travel between all US airports, and are available in this form using the AIRNET program. The AIRNET program’s ‘metro’ option aggregates the 139 airports defined by the Federal Aviation Administration as “Primary Hubs” into US metropolitan areas. In the AIRNET2000 data, primary hub airports have been aggregated into the largest census-defined urban geographies – consolidated statistical areas (CSAs) and Metropolitan Statistical Areas (MSAs) using 2009 census definitions – to reflect that airports have large catchment areas.

3. Data Details

Response Rate	10% random sample, from 100% response
Non-Respondent Bias	n/a
Theoretical Grouping	Passenger air travel within the United States
Publications Using These Data	<ul style="list-style-type: none"> • Neal Z. In press. Evolution of the business air travel network in the US from 1993 – 2011: A descriptive analysis using AIRNET. <i>Research in Transportation Business and Management</i>. • Neal Z. In press. AIRNET: a program for generating intercity networks. <i>Urban Studies</i>. • Neal, Z. 2013. “Types of Hub Cities and Their Effects on Urban Creative Economies.” In <i>Hub Cities and the Knowledge Economy: Seaports, Airports, Brainports</i>, Ashgate. • Neal, Z. 2013. <i>The Connected City: How Networks are Shaping the Modern Metropolis</i>. New York: Routledge. • Liu, X, Neal, Z, Derudder, B. 2012. City networks in the United States: A comparison of four models. <i>Environment and Planning A</i>, 44, 255 – 256. • Neal, Z. 2012. “Creative Employment and Jet Set Cities: Disentangling Causal Effects.” <i>Urban Studies</i> 49. • Neal, Z, Kassens, E. 2011. “The Business Passenger Niche: Comparing Legacy Carriers and Southwest During a National Recession.” <i>Journal of Air Traffic Management</i> 17: 231 – 232. • Neal, Z. 2011. “The Causal Relationship between Employment and Business Networks in US Cities.” <i>Journal of Urban Affairs</i> 33: 167 – 184. • Neal, Z. 2011. “From Central Places to Network Bases: A Transition in the US Urban Hierarchy, 1900 – 2000.” <i>City and Community</i> 10: 49 – 74. • Neal, Z. 2010. “Refining the Air Traffic Approach to City Networks.” <i>Urban Studies</i> 47: 2195 – 2215.
Data Context	Derived from U.S. Bureau of Transportation Statistics’ Origin and Destination Survey
Respondents	Passenger air carriers operating in the US
Nodes	In AIRNET2000 Dataset: 108 US metropolitan areas Using AIRNET Program: US airports or US metropolitan areas
Edges	Airline passengers counts
Longitudinal	In AIRNET2000 Dataset: Yes, annual from 2000 – 2011 Using AIRNET Program: Yes, quarterly from 1993 – present
Temporality	n/a
Analytical or Pedagogical Utility	<p>Analytical –</p> <ul style="list-style-type: none"> • Changes in cities’ accessibility & air service adequacy • Route network: Diffusion of disease • Origin-Destination, Business, and Leisure Networks: Diffusion of information, wealth, etc. <p>Pedagogical –</p> <ul style="list-style-type: none"> • Demonstrating how valued graphs can be dichotomized • Demonstrating how directed graphs can be symmetrized • Illustrating spatial structure of graphs • Comparing flow networks carrying different resources
Known Issues	None

4. Data Files and Formats

The network data are provided as valued and directed adjacency matrices in four Excel workbooks, each containing 13 worksheets (tabs). The first 12 worksheets in each workbook contain a single year's network, while the final worksheet contains a distance matrix reflecting the great circle distance (in kilometers) between the metropolitan areas included in the network. The distance matrix can be used to obtain a geographically organized visualization of the networks and as a control variable in analyses.

AIRNET2000R.xls contains longitudinal route networks, annually from 2000 – 2011. Passenger movement in the route network is defined as a single take-off and landing. Each cell R_{ij} indicates the number of passengers who took off in city i and landed in city j .

Using the AIRNET program, and specifying the 'legtype' option, will decompose the route network passenger counts into four categories indicating the leg's position in the passenger's trip from initial origin to final destination: first, last, middle, or only. This decomposition is generally not useful for constructing separate networks, but is useful for measuring cities' roles as hubs (e.g. identifying the number of a city's total passengers that are connecting vs. terminal).

AIRNET2000O.xls contains origin-destination networks, annually from 2000 – 2011. Passenger movement in the route network is defined as from the initial origin city, to the final destination city, omitting any intermediate layovers or connections. Each cell O_{ij} indicates the number of passengers who started their trip in city i and ended it in city j .

AIRNET2000B.xls contains origin-destination networks for passengers likely traveling for business, annually from 2000 – 2011. Passengers likely traveling for business were identified using two criteria: (1) traveling alone and (2) paid a fare that was statistically significantly ($\alpha = 0.05$) above the average fare for travel from the same origin and to the same destination in the same quarter. Each cell B_{ij} indicates the number of passengers meeting both criteria who started their trip in city i and ended it in city j .

AIRNET2000L.xls contains origin-destination networks for passengers likely traveling for leisure, annually from 2000 – 2011. Passengers likely traveling for leisure were identified using two criteria: (1) traveling with one or more companions and (2) paid a fare that was statistically significantly ($\alpha = 0.05$) below the average fare from the same origin and to the same destination in the same quarter. Each cell L_{ij} indicates the number of passengers meeting both criteria who started their trip in city i and ended it in city j .

In the AIRNET2000 dataset business and leisure networks, statistical significance is assessed at $\alpha = 0.05$, and fares below \$20 (representing frequent flyer redemption fees) or above \$5000 (mostly representing private charter flights) were excluded from fare mean and standard deviation computations. Different values for these parameters can be specified using the AIRNET program's 'alpha,' 'minfare,' and 'maxfare' options.

All business and leisure networks generated by the AIRNET program, including those in the AIRNET2000 dataset, are restricted to passengers taking single-destination round-trips. More complex itineraries (e.g. a multi-city circuit) are relatively rare, and the complexity of their fare pricing makes the proxy identification of business and leisure passengers using these criteria impractical.