AirOnTime87to12: Airline On-Time Performance Data

09/07/2017 • 4 minutes to read • Contributors 

In this article

Description

Airline on-time performance data from 1987 to 2012.

Format

AirOnTime87to12 is an .xdf file with 148617414 observations on the following 29 variables. AirOnTime7Pct is a 7 percent subsample with a subset of 9 variables: ArrDelay, ArrDel15, CRSDepTime, DayOfWeek, DepDelay, Dest,Origin, UniqueCarrier, Year.

`Year`

year of flight (stored as integer).

`Month`

month of the flight(stored as integer).

`DayofMonth`

day of the month (1 to 31) (stored as integer).

`DayOfWeek`

day of the week (stored as factor).

`FlightDate`

flight date (stored as Date).

`UniqueCarrier`

unique carrier code (stored as factor). When the same code has been used by multiple carriers, a numeric suffix is used for earlier users, for example, PA, PA(1), PA(2).

`TailNum`

plane's tail number (stored as factor).

`FlightNum`

flight number (stored as factor).

`OriginAirportID`

originating airport, airport ID (stored as factor). An identification number assigned by US DOT to identify a unique airport. Use this field for airport analysis across a range of years because an airport can change its airport code and airport codes can be reused.

`Origin`

originating airport (stored as factor).

`OriginState`

originating airport, state code (stored as factor).

`DestAirportID`

destination airport, airport ID (stored as factor). An identification number assigned by US DOT to identify a unique airport. Use this field for airport analysis across a range of years because an airport can change its airport code and airport codes can be reused.

`Dest`

destination airport (stored as factor).

`DestState`

destination airport, state code (stored as factor).

`CRSDepTime`

scheduled local departure time (stored as decimal float, e.g., 12:45 is stored as 12.75).

`DepTime`

actual local departure time (stored as decimal float, e.g., 12:45 is stored as 12.75).

`DepDelay`

difference in minutes between scheduled and actual departure time (stored as integer). Early departures show negative numbers.

`DepDelayMinutes`

difference in minutes between scheduled and actual departure time (stored as integer). Early departures set to 0.

`DepDel15`

departure delay indicator, 15 minutes or more (stored as logical).

`DepDelayGroups`

departure delay intervals, every 15 minutes from < -15 to > 180 (stored as a factor).

`TaxiOut`

taxi time from departure from the gate to wheels off, in minutes (stored as integer).

`WheelsOff`

wheels off time in local time (stored as decimal float, e.g., 12:45 is stored as 12.75).

`WheelsOn`

wheels on time in local time (stored as decimal float, e.g., 12:45 is stored as 12.75).

`TaxiIn`

taxi in time in minutes (stored as integer).

`CRSArrTime`

scheduled arrival in local time (stored as decimal float, e.g., 12:45 is stored as 12.75).

`ArrTime`

actual arrival time in local time (stored as decimal float, e.g., 12:45 is stored as 12.75).

`ArrDelay`

difference in minutes between scheduled and actual arrival time (stored as integer). Early arrivals show negative numbers.

`ArrDelayMinutes`

difference in minutes between scheduled and actual arrival time (stored as integer). Early arrivals set to 0.

`ArrDel15`

arrival delay indicator, 15 minutes or more (stored as logical).

`ArrDelayGroups`

arrival delay intervals, every 15 minutes from < -15 to > 180 (stored as a factor).

`Cancelled`

cancelled flight indicator (stored as logical).

`CancellationCode`

cancellation code, if applicable (stored as factor).

`Diverted`

diverted flight indicator(stored as logical).

`CRSElapsedTime`

scheduled elapsed time of flight, in minutes (stored as integer).

`ActualElapsedTime`

actual elapsed time of flight, in minutes (stored as integer).

`AirTime`

flight time, in minutes (stored as integer).

`Flights`

number of flights (stored as integer).

`Distance`

distance between airports in miles (stored as integer).

`DistanceGroup`

distance intervals, every 250 miles, for flight segment (stored as a factor).

`CarrierDelay`

delay, in minutes, attributable to the carrier (stored as integer).

`WeatherDelay`

delay, in minutes, attributable to weather factors (stored as integer).

`NASDelay`

delay, in minutes, attributable to the National Aviation System (stored as integer).

`SecurityDelay`

delay, in minutes, attributable to security factors (stored as integer).

`LateAircraftDelay`

delay, in minutes, attributable to late-arriving aircraft (stored as integer).

`MonthsSince198710`

number of months since October 1987 (store as integer).

`DaysSince19871001`

number of days since 1 October 1987 (store as integer).

Details

These data set contain on-time performance data from 1987 to 2012. The data full data set is stored in 1024 *blocks* in an `zlib` compressed .xdf file.

Source

Research and Innovative Technology Administration (RITA), Bureau of Transportation Statistics.
`http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236`

References

American Statistical Association Statistical Computing Group, Data Expo '09.
`http://stat-computing.org/dataexpo/2009/the-data.html`

U.S. Department of Transportation, Bureau of Transportation Statistics, Research and Innovative Technology Administration. Airline On-Time Statistics.
`http://www.bts.gov/xml/ontimesummarystatistics/src/index.xml`

See Also

AirlineDemoSmall

Examples

Copy

```
 ## Not run:

#####################################################################
# Compute mean arrival delay by year using the full data set
#####################################################################
sumOut <- rxSummary(ArrDelayMinutes~Year = "AirOnTime87to12.xdf")
sumOut

#####################################################################
# To create the 7
# a subset of variables
#####################################################################
airVarsToKeep = c("Year", "DayOfWeek",  "UniqueCarrier","Origin", "Dest",
   "CRSDepTime", "DepDelay", "TaxiOut", "TaxiIn", "ArrDelay", "ArrDel15",
   "CRSElapsedTime", "Distance")

# Specify the locations for the data
bigAirData <- "C:/Microsoft/Data/AirOnTime87to12/AirOnTime87to12.xdf"
airData7Pct <- "C:/Microsoft/Data/AirOnTime7Pct.xdf"

set.seed(12)
rxDataStep(inData = bigAirData, outFile = airData7Pct,
   rowSelection = as.logical(rbinom(.rxNumRows, 1, .07)),
   varsToKeep = airVarsToKeep, blocksPerRead = 40,
   overwrite = TRUE)
rxGetInfo(airData7Pct)
rxGetVarInfo(airData7Pct)
## End(Not run)
```