

数据博览会 2009：航空公司准点数据 - ASA 统计计算 数据仓库 --- Data Expo 2009: Airline on time data - ASA Statistical Computing Dataverse

数据博览会 2009：航空公司准点数据 - ASA 统计计算数据仓库 --- Data Expo 2009: Airline
on time data - ASA Statistical Computing Dataverse [Skip to main content](#)

Toggle navigation



HARVARD
Dataverse

- [Add Data](#)
 - [New Dataverse](#)
 - [New Dataset](#)
- [Search](#)
 - Search
- [About](#)
- [User Guide](#)
- [Support](#)
- [Sign Up](#)
- [Log In](#)



[

ASA Statistical Computing Dataverse

=====

](/dataverse/amassc)(American Statistical Association)

[ASA Sections](#)

[Harvard Dataverse](#)

[ASA Statistical Computing Dataverse](#)

Data Expo 2009: Airline on time data

数据博览会 2009：航空公司准点数据

=====

Incomplete metadataVersion 1.0

元数据版本 1.0 不完整

2008, "Data Expo 2009: Airline on time data", <https://doi.org/10.7910/DVN/HG7NV7>,
Harvard Dataverse, V1

2008, "Data Expo 2009: Airline on time data", <https://doi.org/10.7910/DVN/HG7NV7>,
哈佛数据仓库, V1

Cite Dataset

- [EndNote XML](#)
- [RIS](#)
- [BibTeX](#)

Learn about [Data Citation Standards](#).

了解数据引用标准。

Access Dataset

- Download Options
- [Download ZIP \(1.5 GB\)](#)
- Explore Options
- [Binder](#)

[Contact Owner Share](#)

Dataset Metrics 数据集指标

1,022,374 Downloads 1,022,374 次下载

| Description 描述 |

Bi-Annual Data Exposition: Every other year, at the Joint Statistical Meetings, the Graphics Section and the Computing Section join in sponsoring a special Poster Session called The Data Exposition , but more commonly known as The Data Expo. All of the papers presented in this Poster Session are reports of analyses of a common data set provided for the occasion. In addition, all papers presented in the session are encouraged to report the use of graphical methods employed during the development of their analysis and to use graphics to convey their findings.

Airline on-time performance: Have you ever been stuck in an airport because your flight was delayed or cancelled and wondered if you could have predicted it if you'd had more data? This is your chance to find out.

The data: The data consists of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008. This is a large dataset: there are nearly 120 million records in total, and takes up 1.6 gigabytes of space compressed and 12 gigabytes when uncompressed. The data comes originally from RITA where it is described in detail. You can download the data there, or from the bziped csv files listed below. These files have derivable variables removed, are packaged in yearly chunks and have been more heavily compressed than the originals.

Variable descriptions:

Name	Description
1 Year	1987-2008
2 Month	1-12
3 DayofMonth	1-31
4 DayOfWeek	1 (Monday) - 7 (Sunday)
5 DepTime	actual departure time (local, hhmm)
6 CRSDepTime	scheduled departure time (local, hhmm)
7 ArrTime	actual arrival time (local, hhmm)
8 CRSArrTime	scheduled arrival time (local, hhmm)
9 UniqueCarrier	unique carrier code
10 FlightNum	flight number
11 TailNum	plane tail number
12 ActualElapsedTime	in minutes
13 CRSElapsedTime	in minutes
14 AirTime	in minutes
15 ArrDelay	arrival delay, in minutes
16 DepDelay	departure delay, in minutes

- 17 Origin origin IATA airport code
- 18 Dest destination IATA airport code
- 19 Distance in miles
- 20 TaxiIn taxi in time, in minutes
- 21 TaxiOut taxi out time in minutes
- 22 Cancelled was the flight cancelled?
- 23 CancellationCode reason for cancellation (A = carrier, B = weather, C = NAS, D = security)
- 24 Diverted 1 = yes, 0 = no
- 25 CarrierDelay in minutes
- 26 WeatherDelay in minutes
- 27 NASDelay in minutes
- 28 SecurityDelay in minutes
- 29 LateAircraftDelay in minutes

两年一度的数据展示：每隔一年，在联合统计会议上，图形分会和计算分会共同赞助一个名为“数据展示”的特别海报会议，但更常被称为“数据博览会”。所有在本海报会议上发表的论文都是对为此次活动提供的通用数据集的分析报告。此外，鼓励所有在本届会议上发表的论文报告在分析开发过程中使用的图形方法，并使用图形来传达他们的发现。航空公司准点率：您是否曾经因为航班延误或取消而滞留在机场，并想知道如果您有更多数据是否可以预测？这是您找到答案的机会。数据：数据包括 1987 年 10 月至 2008 年 4 月美国所有商业航班的航班到达和出发详细信息。这是一个大型数据集：总共有近 1.2 亿条记录，压缩后占用 1.6 吉字节空间，解压缩后占用 12 吉字节空间。数据最初来自 RITA，在那里对其进行了详细描述。您可以从那里下载数据，或从下面列出的 bziped csv 文件中下载。这些文件已删除可推导变量，按年分块打包，并且比原始文件压缩得更多。变量描述：名称描述

- 1 年 1987-2008
- 2 月 1-12
- 3 月份中的日期 1-31
- 4 星期几 1（星期一） - 7（星期日）
- 5 出发时间 实际出发时间（当地时间，hhmm）
- 6 计划出发时间 计划出发时间（当地时间，hhmm）

- 7 到达时间 实际到达时间 (当地时间, hhmm)
- 8 计划到达时间 计划到达时间 (当地时间, hhmm)
- 9 唯一承运人 唯一承运人代码
- 10 航班号 航班号
- 11 尾号 飞机尾号
- 12 实际飞行时间 以分钟为单位
- 13 计划飞行时间 以分钟为单位
- 14 空中飞行时间 以分钟为单位
- 15 到达延误 到达延误, 以分钟为单位
- 16 出发延误 出发延误, 以分钟为单位
- 17 起点 起点 IATA 机场代码
- 18 目的地 目的地 IATA 机场代码
- 19 距离 以英里为单位
- 20 出租车到达时间, 以分钟为单位
- 21 出租车出场时间, 以分钟为单位
- 22 取消 航班是否取消?
- 23 CancellationCode 取消原因 (A = 航空公司, B = 天气, C = 国家航空系统, D = 安全)
- 24 备降 1 = 是, 0 = 否
- 25 航空公司延误 (分钟)
- 26 天气延误 (分钟)
- 27 国家航空系统延误 (分钟)
- 28 安全延误 (分钟)
- 29 飞机延误 (分钟)

Collapse Description [-]

|

| License/Data Use Agreement

许可/数据使用协议 |



[CC0 1.0](#)

|

ui-buttonui-按钮

- [Files文件](#)
- [Metadata元数据](#)
- [Terms术语](#)
- [Versions版本](#)

Search

Filter by筛选

File Type: All

- [All](#)
- [Unknown \(22\)](#)
- [Text \(4\)](#)

Access: All

- [All](#)
- [Public \(26\)](#)

File Tag: All

- [All](#)
- [2. Data \(22\)](#)
- [1. Documentation \(4\)](#)

Sort

- [Name \(A-Z\)](#)
- [Name \(Z-A\)](#)
- [Newest](#)
- [Oldest](#)
- [Size](#)

- [Type](#)

|

|

1 to 10 of 26 Files

1 到 26 个文件中的 1 到 10 个

|

[Download](#)



|

[1987.csv.bz2](#)

Unknown - 12.1 MB

未知 - 12.1 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

653,478 Downloads653,478 次下载

MD5: e5c...34e

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation
 - [EndNote XML](#)
 - [RIS](#)
 - [BibTeX](#)

|
|
|

[1988.csv.bz2](#)

Unknown - 47.2 MB

未知 - 47.2 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

14,611 Downloads14,611 次下载

MD5: 0be...a9c

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation
 - [EndNote XML](#)
 - [RIS](#)
 - [BibTeX](#)

|
|
|

[1989.csv.bz2](#)

Unknown - 46.9 MB

未知 - 46.9 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

13,881 Downloads13,881 次下载

MD5: 614...7fc

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation
 - [EndNote XML](#)
 - [RIS](#)
 - [BibTeX](#)

|

|

|

[1990.csv.bz2](#)

Unknown - 49.6 MB

未知 - 49.6 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

13,516 Downloads13,516 次下载

MD5: cbd...392

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)

- Download Metadata
- Data File Citation

- [EndNote XML](#)
- [RIS](#)
- [BibTeX](#)

|

|

|

[1991.csv.bz2](#)

Unknown - 47.6 MB

未知 - 47.6 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

13,049 Downloads13,049 次下载

MD5: 854...cff

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation

- [EndNote XML](#)
- [RIS](#)
- [BibTeX](#)

|

|

|

[1992.csv.bz2](#)

Unknown - 47.7 MB

未知 - 47.7 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

15,130 Downloads15,130 次下载

MD5: 2d7...c98

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation

- [EndNote XML](#)
- [RIS](#)
- [BibTeX](#)

|

|

|

[1993.csv.bz2](#)

Unknown - 47.8 MB

未知 - 47.8 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

13,008 Downloads13,008 次下载

MD5: deb...545

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation
 - [EndNote XML](#)
 - [RIS](#)
 - [BibTeX](#)

|
|
|

[1994.csv.bz2](#)

Unknown - 48.8 MB

未知 - 48.8 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

13,106 Downloads13,106 次下载

MD5: d2c...27c

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation
 - [EndNote XML](#)

- [RIS](#)
- [BibTeX](#)

|

|

|

[1995.csv.bz2](#)

Unknown - 71.4 MB

未知 - 71.4 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

13,082 Downloads13,082 次下载

MD5: 949...b2f

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation

- [EndNote XML](#)
- [RIS](#)
- [BibTeX](#)

|

|

|

[1996.csv.bz2](#)

Unknown - 72.4 MB

未知 - 72.4 MB

Published 2008年10月6日发布时间: 2008 年 10 月 6 日

13,359 Downloads13,359 次下载

MD5: 7af...174

2. Data2. 数据

|

Access File

- File Access
- Public
- Download Options
- [Original File Format](#)
- Download Metadata
- Data File Citation
 - [EndNote XML](#)
 - [RIS](#)
 - [BibTeX](#)

|

[FP123NE](#)Files Per Page 每页文件 Rows Per Page102550

Export Metadata

- [OAI_ORE](#)
- [DataCite数据引用](#)
- [OpenAIRE开放科研平台](#)
- [Schema.org JSON-LD](#)
- [DDI](#)
- [Dublin Core都柏林核心元数据元素集](#)
- [DDI HTML CodebookDDI HTML 代码簿](#)
- [JSON](#)

Citation Metadata 引用元数据

| Persistent Identifier 持久标识符 | doi:10.7910/DVN/HG7NV7 |

| Previous Dataset Persistent ID

先前数据集持久 ID | hdl:1902.1/12012 |

| Publication Date 发布时间 | 2008-10-07 |

| Title 标题 | Data Expo 2009: Airline on time data

数据博览会 2009: 航空公司准点数据 |

| Description 描述 | Bi-Annual Data Exposition: Every other year, at the Joint Statistical Meetings, the Graphics Section and the Computing Section join in sponsoring a special Poster Session called The Data Exposition , but more commonly known as The Data Expo. All of the papers presented in this Poster Session are reports of analyses of a common data set provided for the occasion. In addition, all papers presented in the session are encouraged to report the use of graphical methods employed during the development of their analysis and to use graphics to convey their findings. Airline on-time performance: Have you ever been stuck in an airport because your flight was delayed or cancelled and wondered if you could have predicted it if you'd had more data? This is your chance to find out. The data: The data consists of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008. This is a large dataset: there are nearly 120 million records in total, and takes up 1.6 gigabytes of space compressed and 12 gigabytes when uncompressed. The data comes originally from RITA where it is described in detail. You can download the data there, or from the bziped csv files listed below. These files have derivable variables removed, are packaged in yearly chunks and have been more heavily compressed than the originals. Variable descriptions: Name Description 1 Year 1987-2008 2 Month 1-12 3 DayOfMonth 1-31 4 DayOfWeek 1 (Monday) - 7 (Sunday) 5 DepTime actual departure time (local, hhmm) 6 CRSDepTime scheduled departure time (local, hhmm) 7 ArrTime actual arrival time (local, hhmm) 8 CRSArrTime scheduled arrival time (local, hhmm) 9 UniqueCarrier unique carrier code 10 FlightNum flight number 11 TailNum plane tail number 12 ActualElapsedTime in minutes 13 CRSElapsedTime in minutes 14 AirTime in minutes 15 ArrDelay arrival delay, in minutes 16 DepDelay departure delay, in minutes 17 Origin origin IATA airport code 18 Dest destination IATA airport code 19 Distance in miles 20 TaxiIn taxi in time, in minutes 21 TaxiOut taxi out time in minutes 22 Cancelled was the flight cancelled? 23 CancellationCode reason for cancellation (A = carrier, B = weather, C = NAS, D = security) 24 Diverted 1 = yes, 0 = no 25 CarrierDelay in minutes 26 WeatherDelay in minutes 27 NASDelay in minutes 28 SecurityDelay in minutes 29 LateAircraftDelay in minutes

两年一度的数据展示: 每隔一年, 在联合统计会议上, 图形部分和计算部分共同赞助一个名为“数据展示”的特别海报会议, 但更常被称为“数据博览会”。所有在本海报会议上发表的论文都是对为此目的提供的通用数据集的分析报告。此外, 鼓励所有在本届会议上发表的论文报告在分析开发过程中使用的图形方法, 并使用图形来传达他们的发现。航空公司准点率: 您是否曾经因为航班延误或取消而被困在机场, 并想知道如果您有更多数据是否可以预测它? 这是您发现的

机会。数据：数据包括 1987 年 10 月至 2008 年 4 月美国所有商业航班的航班到达和出发详细信息。这是一个大型数据集：总共有近 1.2 亿条记录，压缩后占用 1.6 吉字节空间，解压缩后占用 12 吉字节空间。数据最初来自 RITA，在那里对其进行了详细描述。您可以从那里下载数据，或从下面列出的 bziped csv 文件中下载。这些文件已删除可推导变量，按年分块打包，并且比原始文件压缩得更厉害。变量描述：名称 描述 1 年 1987-2008 2 月 1-12 3 月份中的日期 1-31 4 星期几 1（星期一） - 7（星期日） 5 实际出发时间 实际出发时间（当地时间，hhmm） 6 计划出发时间 计划出发时间（当地时间，hhmm） 7 实际到达时间 实际到达时间（当地时间，hhmm） 8 计划到达时间 计划到达时间（当地时间，hhmm） 9 唯一承运人 唯一承运人代码 10 航班号 航班号 11 尾号 飞机尾号 12 实际飞行时间 以分钟计 13 计划飞行时间 以分钟计 14 空中飞行时间 以分钟计 15 到达延误 到达延误，以分钟计 16 出发延误 出发延误，以分钟计 17 起点 起点 IATA 机场代码 18 目的地 目的地 IATA 机场代码 19 距离 以英里计 20 滑行进场时间 滑行进场时间，以分钟计 21 滑行出港时间 滑行出港时间，以分钟计 22 取消 航班是否取消？ 23 CancellationCode 取消原因 (A = 航空公司, B = 天气, C = 国家航空系统, D = 安全) 24 备降 1 = 是, 0 = 否 25 航空公司延误 (分钟) 26 天气延误 (分钟) 27 国家航空系统延误 (分钟) 28 安全延误 (分钟) 29 飞机延误 (分钟) | | Deposit Date 存款日期 | 2008-10-06 |

Dataset Terms 数据集术语

License/Data Use Agreement

许可/数据使用协议

Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. Please use the data citation shown on the dataset page.

我们的社区规范以及良好的科学实践要求通过引用给予适当的署名。请使用数据集页面上显示的数据引用。



[CC0 1.0](#)

Direct

Dataset Version数据集版本	Summary摘要	Contributors贡献者	Published on发布于
No records found.未找到记录。			

Edit File[](#)

This file has already been deleted (or replaced) in the current version. It may not be edited.

Close

Restrict Access[](#)

Restricting limits access to published files. People who want to use the restricted files can request access by default. If you disable request access, you must add information about access to the Terms of Access field.

Learn about restricting files and dataset access in the [User Guide](#).

Request Access

Enable access request

You must enable request access or add terms of access to restrict file access.

Terms of Access for Restricted Files

[Save Changes](#) Cancel

Edit Embargo[](#)

The selected file or files have already been published. Contact an administrator to change the embargo date or reason of the file or files.

Cancel

Delete Files[](#)

The file will be deleted after you click on the Delete button.

Files will not be removed from previously published versions of the dataset.

Delete Cancel

Continue Cancel

Select File(s)[](#)

Please select one or more files.

Close

Share Dataset[](#)

Share this dataset on your favorite social media networks.

[](# "FaceBook") [](# "Twitter") [](# "LinkedIn")

Close

[Continue](#) Cancel

Dataset Citations[](#)

Citations for this dataset are retrieved from Crossref via DataCite using Make Data Count standards. For more information about dataset metrics, please refer to the [User Guide](#).

Sorry, no citations were found.

Close

Restricted Files, or Files only Accessible Via Globus Selected[](#)

The selected file(s) may not be downloaded because you have not been granted access or the files can only be transferred via Globus.

Close

Ineligible Files Selected[](#)

The selected file(s) may not be transferred because you have not been granted access or the files are not Globus accessible.

Close

Download Options[](#)

The files selected are too large to download as a ZIP.

You can select individual files that are below the 15.0 GB download limit from the files table, or use the [Data Access API](#) for programmatic access to the files.

Select File(s)[](#)

Please select a file or files to be downloaded.

Close

Restricted Files, or Files only Accessible Via Globus Selected[](#)

The restricted file(s) selected may not be downloaded because you have not been granted access.

Click Continue to download the files you have access to download.

Continue Cancel

Ineligible Files Selected[](#)

Some file(s) cannot be transferred. (They are restricted, embargoed, or not Globus accessible.)

Click Continue to transfer the eligible files.

Continue Cancel

Delete Dataset[](#)

Are you sure you want to delete this dataset and all of its files? You cannot undelete this dataset.

Continue Cancel

Delete Draft Version[](#)

Are you sure you want to delete this draft version? Files will be reverted to the most recently published version. You cannot undelete this draft.

Continue Cancel

Unpublished Dataset Private URL[](#)

Private URL can only be used with unpublished versions of datasets.

Cancel

Unpublished Dataset Private URL

Are you sure you want to disable the Private URL? If you have shared the Private URL with others they will no longer be able to use it to access your unpublished dataset.

Yes, Disable Private URL Cancel

Delete Files[](#)

The file(s) will be deleted after you click on the Delete button.

Files will not be removed from previously published versions of the dataset.

Delete Cancel

Compute[](#)

This dataset contains restricted files you may not compute on because you have not been granted access.

Close

Deaccession Dataset[](#)

Are you sure you want to deaccession? This is permanent and the selected version(s) will no longer be viewable by the public.

No

Deaccession Dataset[](#)

Are you sure you want to deaccession this dataset? This is permanent an it will no longer be viewable by the public.

No

Version Differences Details[](#)

Please select two versions to view the differences.

Close

Version Differences Details[](#)

Version:

Last Updated:

Version:

Last Updated:

Done

Select File(s)[](#)

Please select a file or files for access request.

Close

Select File(s)[](#)

Embargoed files cannot be accessed. Please select an unembargoed file or files for your access request.

Close

Edit Tags[](#)

Select existing file tags or create new tags to describe your files. Each file can have more than one tag.

Save ChangesCancel

Request Access[](#)

You need to [Sign Up](#) or [Log In](#) to request access.

Close

Dataset Terms[](#)

Please confirm and/or complete the information needed below in order to request access to files in this dataset.

This dataset is made available under the following terms. Please confirm and/or complete the information needed below in order to continue.

License/Data Use Agreement

Our [Community Norms](#) as well as good scientific practices expect that proper credit is given via citation. Please use the data citation shown on the dataset page.



[CC0 1.0](#)

Preview Guestbook[](#)

Upon downloading files the guestbook asks for the following information.

Guestbook Name

Collected Data

Account Information

Close

Package File Download[](#)

Use the Download URL in a Wget command or a download manager to download this package file. Download via web browser is not recommended. [User Guide - Downloading a Dataverse Package via URL](#)

Download URL

`https://dataverse.harvard.edu/api/access/datafile/`

Close

Compute Batch[](#)

Clear Batchui-button

Dataset	Persistent Identifier	Change Compute Batch
----------------	------------------------------	-----------------------------

[Compute Batch](#) Cancel

Submit for Review[](#)

You will not be able to make changes to this dataset while it is in review.

This draft version has incomplete metadata that needs to be edited before it can be published.

Cancel

Publish Dataset[](#)

Are you sure you want to republish this dataset?

By default datasets are published with the CC0- "Public Domain Dedication" waiver. Learn more about the CC0 waiver [here](#).

To publish with custom Terms of Use, click the Cancel button and go to the Terms tab for this dataset.

Select if this is a minor or major version update.

|

Minor Release (1.1) |

Major Release (2.0) |

Cancel

Publish Dataset[](#)

This dataset cannot be published until [ASA Statistical Computing Dataverse](#) is published by its administrator.

Close

Publish Dataset[](#)

This dataset cannot be published until [ASA Statistical Computing Dataverse](#) and [Harvard Dataverse](#) are published.

Close


Return to Author[](#)

Return this dataset to contributor for modification. The reason for return entered below will be sent by email to the author.

2000 characters remaining.

Continue Cancel

Copyright © 2024, The President & Fellows of Harvard College | [Privacy Policy](#)

Powered by  <http://dataverse.org/> v. 6.2 build
"v6.2+10451+10463+10383+10547+10555-iqss"

[DataCite-Logo](#)

[Feedback](#)



Contact Harvard Dataverse Support[](#)

To

Harvard Dataverse Support

From

Subject

Message

Please fill this out to prove you are not a robot.

+ =

[Send Message](#) [Cancel](#)