



# Viszards XXXI

## Networks of US flight delays 1987-2008

Vladimir Batagelj, Joseph Biberstine, Katy Borner,  
Ulrik Brandes, Seok-Hee Hong, Hui Liu, Jeffrey C. Johnson,  
Lothar Krempel, Juergen Lerner, Andrej Mrvar,  
Juergen Pfeffer, Quan Nguyen

Sunbelt XXXI, St. Pete Beach, FL, USA,  
8-13 February, 2011



# Outline

Viszards XXXI

V. Batagelj  
etal.

The data

Some results

- 1 The data
- 2 Some results





# Networks of US flight delays 1987-2008

Viszards XXXI

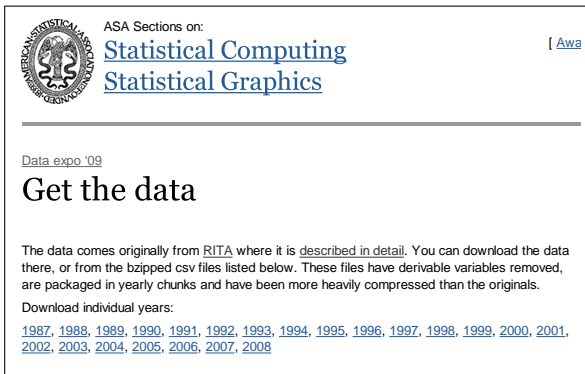
V. Batagelj  
etal.

The data

Some results

This year Vizards' data are networks of US flight delays 1987-2008. We "borrowed" the data from Data Expo 2009

<http://stat-computing.org/dataexpo/2009/>



ASA Sections on: [Statistical Computing](#) [Statistical Graphics](#) [ [Awa](#) ]

---

[Data expo '09](#)

## Get the data

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.

Download individual years:

[1987](#), [1988](#), [1989](#), [1990](#), [1991](#), [1992](#), [1993](#), [1994](#), [1995](#), [1996](#), [1997](#), [1998](#), [1999](#), [2000](#), [2001](#), [2002](#), [2003](#), [2004](#), [2005](#), [2006](#), [2007](#), [2008](#)

The visualizations proposed for Data Expo 2009 didn't deal with the relational aspects of the data.





# Data description

Viszards XXXI

V. Batagelj  
etal.

The data

Some results

## 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	<a href="#">unique carrier code</a>
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 <a href="#">IATA airport code</a>
18 Dest	destination <a href="#">IATA airport code</a>
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

1987	1311826
1988	5202096
1989	5041200
1990	5270893
1991	5076925
1992	5092157
1993	5070501
1994	5180048
1995	5327435
1996	5351983
1997	5411843
1998	5384721
1999	5527884
2000	5683047
2001	5967780
2002	5271359
2003	6488540
2004	7129270
2005	7140596
2006	7141922
2007	7453215
2008	7009728

total: 123534969



# The flight was canceled

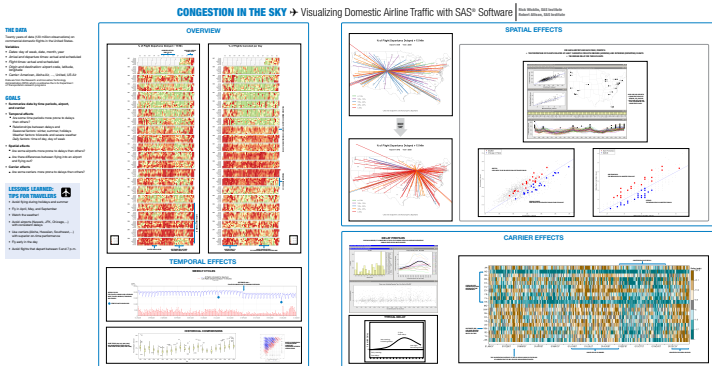
Viszards XXXI

V. Batagelj  
etal.

The data

Some results

The first place at the Data Expo 2009 was awarded to the poster **Congestion in the sky: Visualising domestic airline traffic with SAS** by Rick Wicklin and Robert Allison, SAS Institute.





# The flight was canceled

Viszards XXXI

V. Batagelj  
etal.

The data

Some results



On Sunday, 6 Februray 2011 the flight DL 5379: PITTSBURGH PIT 9:36 - ATLANTA ATL 11:30 was canceled. **Go, Steelers, go!!!**





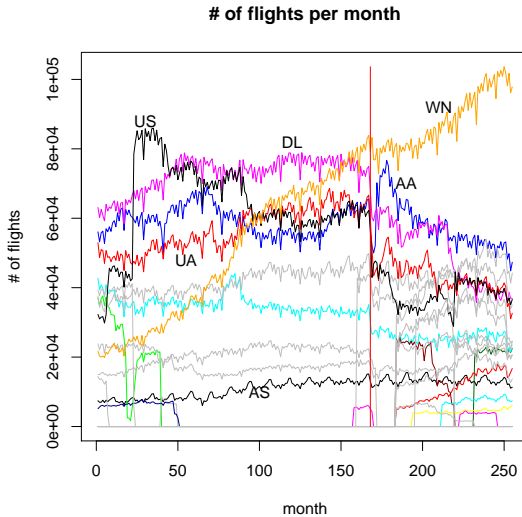
# Number of flights per month by companies

Viszards XXXI

V. Batagelj  
etal.

The data

Some results





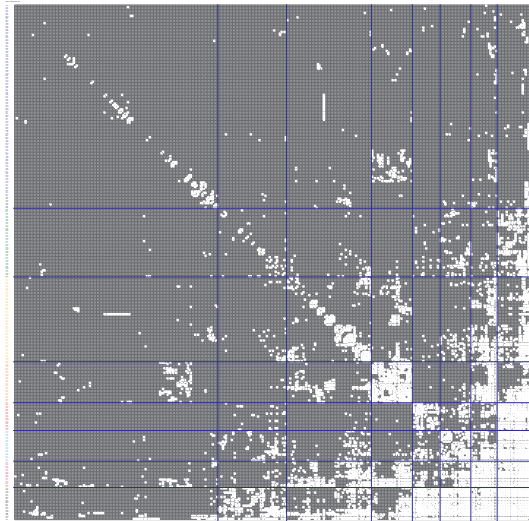
# Clustering of airports

Viszards XXXI

V. Batagelj  
etal.

The data

Some results







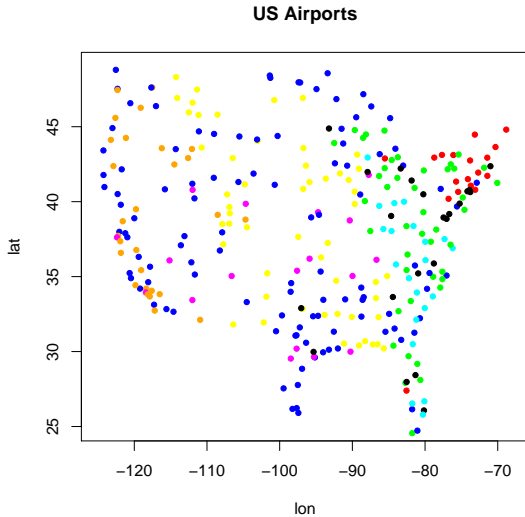
# Geographical view on clusters

Viszards XXXI

V. Batagelj  
etal.

The data

Some results





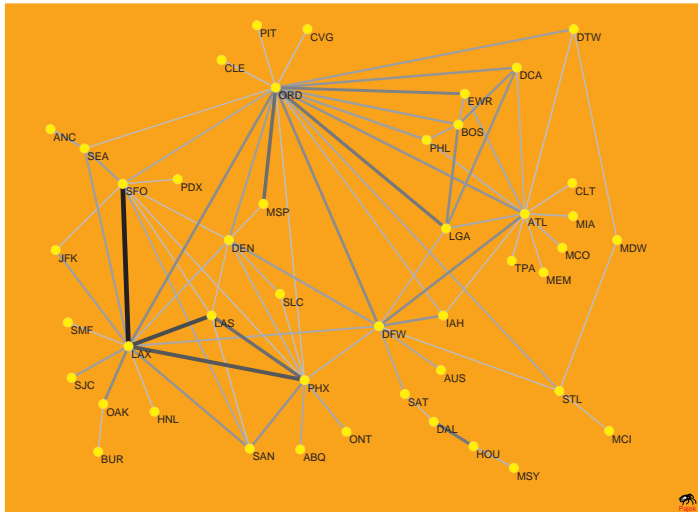
# Line cut for number of flights at 200000

Viszards XXXI

V. Batagelj et al.

The data

Some results







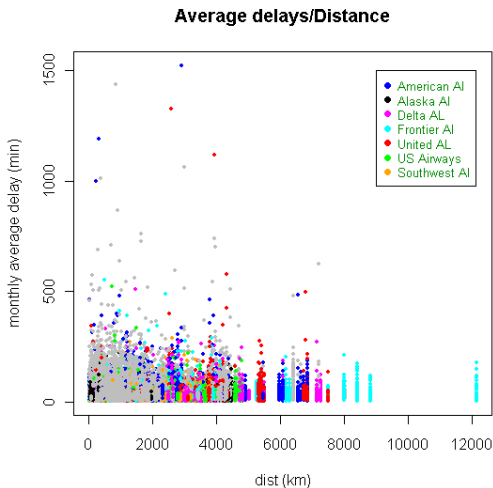
# Monthly average arrival delay / Distance

Viszards XXXI

V. Batagelj  
etal.

The data

Some results





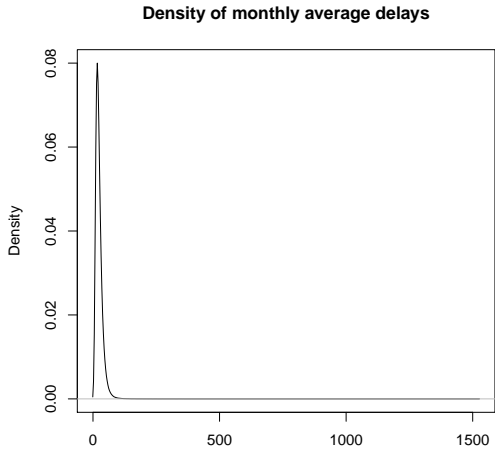
# Density of the monthly average arrival delay

Viszards XXXI

V. Batagelj  
etal.

The data

Some results



N = 1181339 Bandwidth = 0.6054

average delay = 23.88 min







# Monthly average delays over 1h in 2008

Viszards XXXI

V. Batagelj  
etal.

The data

Some results

Click on the picture to get next (January – December 2008) !



V. Batagelj etal.

Viszards XXXI