

The data set (and description) can be downloaded here:

<http://archive.ics.uci.edu/ml/machine-learning-databases/glass/glass.data>

**Description:**

1. Title: Glass Identification Database

2. Sources:

- (a) Creator: B. German
  - Central Research Establishment
  - Home Office Forensic Science Service
  - Aldermaston, Reading, Berkshire RG7 4PN
- (b) Donor: Vina Spiehler, Ph.D., DABFT
  - Diagnostic Products Corporation
  - (213) 776-0180 (ext 3014)
- (c) Date: September, 1987

3. Past Usage:

- Rule Induction in Forensic Science
- Ian W. Evett and Ernest J. Spiehler
- Central Research Establishment
- Home Office Forensic Science Service
- Aldermaston, Reading, Berkshire RG7 4PN
- Unknown technical note number (sorry, not listed here)
- General Results: nearest neighbor held its own with respect to the rule-based system

4. Relevant Information:

Vina conducted a comparison test of her rule-based system, BEAGLE, the nearest-neighbor algorithm, and discriminant analysis. BEAGLE is a product available through VRS Consulting, Inc.; 4676 Admiralty Way, Suite 206; Marina Del Ray, CA 90292 (213) 827-7890 and FAX: -3189. In determining whether the glass was a type of "float" glass or not, the following results were obtained (# incorrect answers):

Type of Sample	Beagle	NN	DA
windows that were float processed (87)	10	12	21
windows that were not: (76)	19	16	22

The study of classification of types of glass was motivated by criminological investigation. At the scene of the crime, the glass left can be used as evidence...if it is correctly identified!

5. Number of Instances: 214

6. Number of Attributes: 10 (including an Id#) plus the class attribute  
-- all attributes are continuously valued

7. Attribute Information:

- 1. Id number: 1 to 214
- 2. RI: refractive index
- 3. Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)
- 4. Mg: Magnesium
- 5. Al: Aluminum
- 6. Si: Silicon
- 7. K: Potassium
- 8. Ca: Calcium
- 9. Ba: Barium
- 10. Fe: Iron

11. Type of glass: (class attribute)
- 1 building\_windows\_float\_processed
  - 2 building\_windows\_non\_float\_processed
  - 3 vehicle\_windows\_float\_processed
  - 4 vehicle\_windows\_non\_float\_processed (none in this database)
  - 5 containers
  - 6 tableware
  - 7 headlamps

8. Missing Attribute Values: None

Summary Statistics:

Attribute:	Min	Max	Mean	SD	Correlation with class
2. RI:	1.5112	1.5339	1.5184	0.0030	-0.1642
3. Na:	10.73	17.38	13.4079	0.8166	0.5030
4. Mg:	0	4.49	2.6845	1.4424	-0.7447
5. Al:	0.29	3.5	1.4449	0.4993	0.5988
6. Si:	69.81	75.41	72.6509	0.7745	0.1515
7. K:	0	6.21	0.4971	0.6522	-0.0100
8. Ca:	5.43	16.19	8.9570	1.4232	0.0007
9. Ba:	0	3.15	0.1750	0.4972	0.5751
10. Fe:	0	0.51	0.0570	0.0974	-0.1879

9. Class Distribution: (out of 214 total instances)

- 163 window glass (building windows and vehicle windows)
  - 87 float processed
    - 70 building windows
    - 17 vehicle windows
  - 76 non-float processed
    - 76 building windows
    - 0 vehicle windows
- 51 Non-window glass
  - 13 containers
  - 9 tableware
  - 29 headlamps

Citation Request:

Please refer to the repository <http://archive.ics.uci.edu/ml> (see citation policy).

See also Frank, A. & Asuncion, A. (2010). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>].

Irvine, CA: University of California, School of Information and Computer Science.

Descriptive statistics:

Dataset= glass : n= 146 , d= 9

Class1: n= 70

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	0.0000	0.0006	0.0002	-0.0005	-0.0010	-0.0004	0.0011	0.0000	0.0000
[2,]	0.0006	0.2493	0.0479	-0.0816	-0.2295	-0.0762	0.0962	0.0062	-0.0150
[3,]	0.0002	0.0479	0.0610	-0.0267	-0.0707	-0.0271	0.0199	0.0001	-0.0043
[4,]	-0.0005	-0.0816	-0.0267	0.0746	0.0972	0.0440	-0.1163	0.0019	0.0023
[5,]	-0.0010	-0.2295	-0.0707	0.0972	0.3243	0.0999	-0.2234	-0.0098	0.0098
[6,]	-0.0004	-0.0762	-0.0271	0.0440	0.0999	0.0462	-0.0894	-0.0030	0.0028
[7,]	0.0011	0.0962	0.0199	-0.1163	-0.2234	-0.0894	0.3304	-0.0016	0.0016
[8,]	0.0000	0.0062	0.0001	0.0019	-0.0098	-0.0030	-0.0016	0.0070	-0.0004
[9,]	0.0000	-0.0150	-0.0043	0.0023	0.0098	0.0028	0.0016	-0.0004	0.0079

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.0000	0.5635	0.4004	-0.7268	-0.8072	-0.8213	0.8460	0.0518	-0.1672
[2,]	0.5635	1.0000	0.3880	-0.5981	-0.8072	-0.7101	0.3352	0.1481	-0.3370
[3,]	0.4004	0.3880	1.0000	-0.3950	-0.5024	-0.5097	0.1404	0.0070	-0.1959
[4,]	-0.7268	-0.5981	-0.3950	1.0000	0.6248	0.7498	-0.7406	0.0828	0.0937
[5,]	-0.8072	-0.8072	-0.5024	0.6248	1.0000	0.8163	-0.6825	-0.2058	0.1937
[6,]	-0.8213	-0.7101	-0.5097	0.7498	0.8163	1.0000	-0.7241	-0.1644	0.1482
[7,]	0.8460	0.3352	0.1404	-0.7406	-0.6825	-0.7241	1.0000	-0.0334	0.0315
[8,]	0.0518	0.1481	0.0070	0.0828	-0.2058	-0.1644	-0.0334	1.0000	-0.0600
[9,]	-0.1672	-0.3370	-0.1959	0.0937	0.1937	0.1482	0.0315	-0.0600	1.0000

Median: 1.5179 13.0924 3.5182 1.2406 72.8072 0.5326 8.625 0.0069 0.0537

Mean: 1.5187 13.2423 3.5524 1.1639 72.6191 0.4474 8.7973 0.0127 0.057

MCD-estimated:

MDC-0.975-Mean: 1.5185 13.3125 3.596 1.149 72.636 0.4555 8.6903 0 0

MDC-0.750-Mean: 1.5186 13.2841 3.5815 1.1715 72.6378 0.4583 8.7046 0 0

MDC-0.500-Mean: 1.5186 13.2841 3.5815 1.1715 72.6378 0.4583 8.7046 0 0

Class2: n= 76

Covariance matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	0.0000	-0.0010	-0.0036	-0.0005	-0.0018	-0.0005	0.0068	0.0005	0.0001
[2,]	-0.0010	0.4411	0.2792	0.0050	-0.0343	-0.0055	-0.5766	-0.0994	-0.0122
[3,]	-0.0036	0.2792	1.4778	0.1075	0.2421	0.1461	-2.1240	-0.1285	-0.0165
[4,]	-0.0005	0.0050	0.1075	0.1013	-0.0196	0.0435	-0.2718	0.0297	-0.0004
[5,]	-0.0018	-0.0343	0.2421	-0.0196	0.5250	0.0274	-0.6237	-0.1164	-0.0186
[6,]	-0.0005	-0.0055	0.1461	0.0435	0.0274	0.0457	-0.2638	0.0029	0.0007
[7,]	0.0068	-0.5766	-2.1240	-0.2718	-0.6237	-0.2638	3.6927	0.1769	0.0355
[8,]	0.0005	-0.0994	-0.1285	0.0297	-0.1164	0.0029	0.1769	0.1313	0.0090
[9,]	0.0001	-0.0122	-0.0165	-0.0004	-0.0186	0.0007	0.0355	0.0090	0.0113

Correlation matrix:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.0000	-0.3899	-0.7698	-0.3978	-0.6598	-0.5756	0.9327	0.3863	0.2022
[2,]	-0.3899	1.0000	0.3457	0.0234	-0.0712	-0.0388	-0.4518	-0.4129	-0.1726
[3,]	-0.7698	0.3457	1.0000	0.2777	0.2748	0.5624	-0.9092	-0.2918	-0.1277
[4,]	-0.3978	0.0234	0.2777	1.0000	-0.0848	0.6400	-0.4443	0.2575	-0.0122
[5,]	-0.6598	-0.0712	0.2748	-0.0848	1.0000	0.1767	-0.4480	-0.4434	-0.2407
[6,]	-0.5756	-0.0388	0.5624	0.6400	0.1767	1.0000	-0.6424	0.0376	0.0324
[7,]	0.9327	-0.4518	-0.9092	-0.4443	-0.4480	-0.6424	1.0000	0.2540	0.1734
[8,]	0.3863	-0.4129	-0.2918	0.2575	-0.4434	0.0376	0.2540	1.0000	0.2343
[9,]	0.2022	-0.1726	-0.1277	-0.0122	-0.2407	0.0324	0.1734	0.2343	1.0000

Median: 1.5173 13.151 3.4867 1.4489 72.7502 0.5785 8.3344 0.0108 0.0641

Mean: 1.5186 13.1117 3.0021 1.4082 72.598 0.5211 9.0737 0.0503 0.0797

MCD-estimated:

MDC-0.975-Mean: 1.5181 13.113 3.1188 1.4153 72.723 0.5293 8.8705 0 0

MDC-0.750-Mean: 1.5181 13.113 3.1188 1.4153 72.723 0.5293 8.8705 0 0

MDC-0.500-

Mean: 1.517 13.1778 3.6248 1.4789 72.73 0.5961 8.155 0 0.0591

Measures:

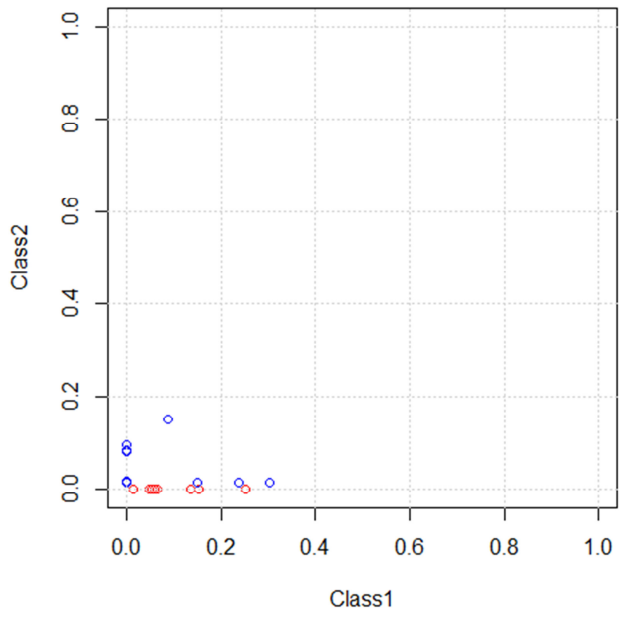
Mah.Dist: 1.3286

Mah.Dist-MCD-0.975: 1.2972

Mah.Dist-MCD-0.750: 1.2629

Mah.Dist-MCD-0.500: 1.2972

DD-Plot (zonoid): glass



DD-Plot (random Tukey): glass

