

Incomplete data

The data sets listed below are either widely used in general in the missing data community or used for illustration of different methods handling missing values in the tutorials from the [Tutorials](#) and [R packages](#) sections. This presentation scheme is inspired by the [UCI Machine Learning Repository](#).

Click on a table entry to obtain further information about the data set.

Name	Data Types	Attribute Types	# Instances	# Attributes	% Missing entries	Complete data available	Year
Airquality	Multivariate, Time Series	Real	154	6	7	No	1973
chorizonDL	Multivariate	Integer, Real	606	110	15	Yes	1998
Health Nutrition And Population Statistics	Multivariate, Time Series	Integer, Real	15,022	397	54	No	2017
NHANES	Multivariate	Categorical, Integer, Real	10,000	75	37	No	2012
oceanbuoys	Multivariate, Time Series	Real	736	8	3	No	1997
Ozone	Multivariate	Categorical, Integer, Real	366	13	6	No	1976

Los Angeles Ozone Pollution Data, 1976. This data set contains daily measurements of ozone concentration and meteorological quantities. It can be found in R in the [mlbench](#) package and is loaded by calling `data(Ozone)`.

[More information on the dataset.](#)

Tutorials illustrating methods on this data:

- Julie Josse's [course](#) on missing values imputation using PC methods.
- Julie Josse's and Nick Tierney's tutorial on handling missing values. Download the data set from this tutorial: [ozoneNA.csv](#)
- Nick Tierney's [naniar vignette](#) for missing data visualization.

pedestrian	Multivariate, Time series	Categorical, Integer	37,700	9	2	No	2016
----------------------------	---------------------------	----------------------	--------	---	---	----	------