**A list of public data sources:**

**Public Datasets**

* Here is another collection <https://github.com/awesomedata/awesome-public-datasets>
* <https://github.com/apiad/datasets-list>
* <https://github.com/datasets/openml-datasets/tree/master/data> the same data as this list <https://www.openml.org/search?type=data>
* Machine Learning Data sets  <https://archive.ics.uci.edu/ml/datasets.html>
* <https://www.data.gov/>
* <https://www.kaggle.com/datasets>
* <https://sites.google.com/a/drwren.com/wmd/details>
* <https://data.cityofnewyork.us/data>
* <http://snap.stanford.edu/data/index.html>
* Amazon AWS Public Dataset Program <https://aws.amazon.com/opendata/public-datasets/>

**Event Data**

* GeoLife GPS Trajectories <http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/>
* Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors (MIT Project 2004) <http://courses.media.mit.edu/2004fall/mas622j/04.projects/home/>
* <https://sites.google.com/a/drwren.com/wmd/home>

**Frequent Itemset Mining Dataset Repository**

* <http://fimi.ua.ac.be/data/>

**Airline On-time Performance**

* <http://www.eecs.wsu.edu/~yyao/StreamingGraphs.html>
* <http://openflights.org/data.html>

**Collection and Streaming of Graph Datasets**

* <http://www.eecs.wsu.edu/~yyao/StreamingGraphs.html>

**Data Streams**

* [http://www.quora.com/Where-can-I-find-public-or-free-real-time-or-streaming-data-sources](https://www.quora.com/Where-can-I-find-public-or-free-real-time-or-streaming-data-sources)
* 3 hourly weather forecast and observational data – UK locations [http://data.gov.uk/dataset/metoffice\_uklocs3hr\_fc](https://data.gov.uk/dataset/metoffice_uklocs3hr_fc)
* There is also an IRC chan with the live log of wikipedia edits on the #en.wikipedia channel of the irc.wikimedia.org server. <http://blog.programmableweb.com/2010/12/26/64-new-apis-real-time-facebook-twitter-streaming-and-google-shopping/>

**New York Taxi Datasets**

* <https://data.ny.gov/>
* TLC Trip Record Data <http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml>
* Another Taxi Data set (including GPS) published by chris whong <http://chriswhong.com/open-data/foil_nyc_taxi/>
* Another Description of this dataset <http://publish.illinois.edu/dbwork/open-data/>

**OpenStreet Map**

* [http://wiki.openstreetmap.org/wiki/Planet.osm](https://wiki.openstreetmap.org/wiki/Planet.osm)

**Dublin Bus GPS sample data from Dublin City Council (Insight Project)**

* <https://data.gov.ie/dataset/dublin-bus-gps-sample-data-from-dublin-city-council-insight-project> Further data set at [https://data.gov.ie](https://data.gov.ie/)

**Wikipedia Dump**

* Wikpedia Dump <https://dumps.wikimedia.org/>