Bike-Sharing System

A bike-sharing system is a system providing public bikes for citizens' commutes. In a bike-sharing system, there are many bike stations distributed in a city. Users are required to swipe an RFID card when check out (borrow) and check in (return) a bike from/to a bike station. A record, consisting of the timestamp, station ID, bike ID and user's information is generated for each card swipe. These records constitute the bike dataset of each bike-sharing system.

Dataset Statistics

We have summarized four datasets, bike dataset and meteorology dataset of New York City from 1st Apr. to 30th Sep. 2014, bike dataset and meteorology dataset of Washington D.C. from 1st Apr. to 30th Sep. 2014. The statistics of these four datasets are illustrated in Table 1.

Table 1. Statistics of Datasets

Data Source			New York City	Washington D.C.
Time Span			1st Apr. – 30th Sep. 2014	1st Apr. – 30th Sep. 2014
Bike Data	# Stations		344	351
	# Bikes		6800+	3000+
	# Records		5,359,995	1,886,144
Meteorology Data	Weather / #hours	Snowy	2	0
		Rainy	231	149
		Foggy	303	150
		Sunny	3856	4093
	Temperature/ °C		[0, 33]	[-2, 36]
	Wind Speed/ mph		[0, 18]	[0, 29]

The datasets have been used in paper [1] to predict the check-out/in (the number of bikes that are borrowed/ returned from/ to a bike station) in a bike-sharing system. Please cite the following two papers when using the datasets.

[1] Yexin Li, Yu Zheng, Huichu Zhang and Lei Chen. Traffic prediction in a bike-sharing system. In *Proceedings of* the 23rd ACM SIGSPATIAL GIS, 2015.

[2] Yu Zheng, Licia Vapra, Ouri Wolfson, and Hai Yang. Urban Computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology, vol. 5, no. 3, pp. 38:1-38:55, 2014.

Dataset Details

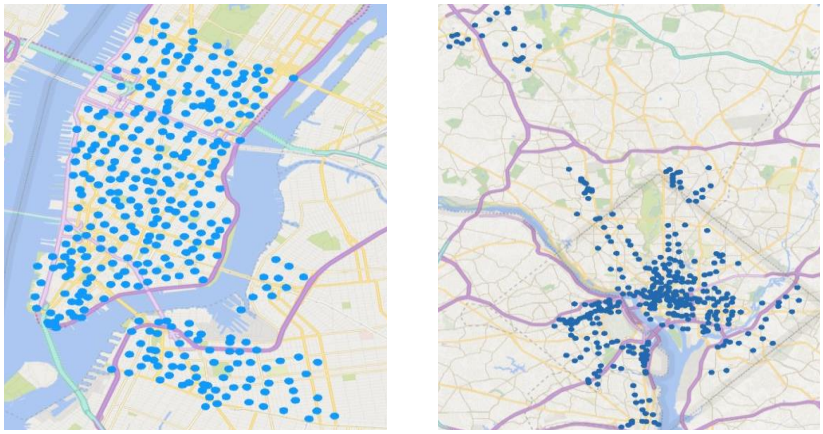
Regarding the size of the uploaded file, we only give four small subsets as example, to illustrate the format and details of the four datasets. However, the datasets are open online and their links are attached in corresponding paragraphs respectively. Users who need the whole datasets can click the links to download by themselves.

NYC Folder

- There is a **'csv'** document named "Bike-NYC" restoring the bike data from 1st, Aug. to 15th, Aug., 2014 in New York City, which contains the following information: origin station (station ID, station name, station latitude and longitude), destination station (station

ID, station name, station latitude and longitude), start time (when a bike is checked out), stop time (when a bike is checked in), trip duration in seconds (from when a bike is checked out to when it is checked in), bike ID (the bike which is borrowed), user type (subscriber, which means an annual member, and customer, which means a 24-hour or 7-day access holder), user's birth year and user gender. The bike-sharing system in New York City is called "Citi Bike". Its station map is shown in Figure 1. A). All the bike data in New York City from Jul. 2013 to current time can be downloaded at http://www.citibikenyc.com/system-data.

- There is an **'xls'** document named "Meteorology-NYC" restoring the meteorology data from 1st, Aug. to 15th, Aug., 2014 in New York City. The meteorology data are recorded by a station located near to Central Park, whose ID is KNYC and geographical location is (40.78333, -73.96667). Almost for every hour, there is at least one record of the weather, temperature and wind speed. In the document, each row corresponds to a timestamp when the record is taken. In the last column, which corresponds to weather, 'no value' means 'sunny'. We illustrate the proportion of each kind of weather in New York City for the half one year (from 1st, Apr. to 30th, Sep. 2014) by a sector diagram in Figure 2. A). As the proportion of snowy hours is very small, it does not appear in the diagram. Note that, if an hour is rainy & snowy, we consider it as snowy; if an hour is rainy & foggy, we consider it as rainy; if an hour is foggy & sunny, we consider it as foggy; etc. Meteorology data for more than 10 recent years in New York City can be downloaded at http://mesowest.utah.edu/cgi-bin/droman/download_ndb.cgi?stn=KNYC&year1=2015&day1=29&month1=10&hour1=&timetype=GMT&unit=0.



A) New York City B) Washington D.C.

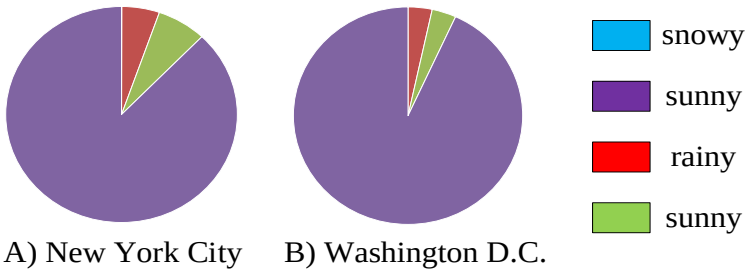
Fig 1. Station Map

D.C. Folder

- There is a '**txt**' document named "stations" restoring the geographical location of each bike station (station latitude, longitude, station ID and station name) in the bike-sharing system of Washington D.C. In the document, each row corresponds to one station and the columns separated by blank correspond to the station's latitude, longitude, ID and name. Station ID

is given by us for simplicity. Station latitude and longitude are collected from the official website. The bike-sharing system in Washington D.C. is called "Capital Bikeshare". Its station map is shown in Figure 1. B). All the current bike stations (until Jul. 2015) are included in the document. However, considering system development in the future, if necessary, the newest station information can be collected at https://www.capitalbikeshare.com/data/stations/bikeStations.xml.

- There is a **'csv'** document named "Bike-D.C." restoring the bike data from 1st, Aug. to 15th, Aug., 2014 in Washington D.C. The data contain the following information: origin station (station name), destination station (station name), start time, stop time, trip duration, bike ID and user type (casual and registered). All the bike data in Washington D.C. from Sep. 2010 to current time can be downloaded at https://www.capitalbikeshare.com/trip-history-data.
- There is an **'xls'** document named "Meteorology-D.C." restoring the meteorology data from 1st, Aug. to 15th, Aug., 2014 in Washington D.C. The meteorology data are recorded by a station located at (38.81667, -76.85), whose ID is KADW. There is another station located at (38.98060, -76.92230) with an ID KCGS, which seems more close to the Capital Bikeshare system. However, its records about the weather are not reliable. Thus we use the records of KADW as the meteorology data of Washington D.C. Its format is the same with that of New York City illustrated previously. We show the proportion of each kind of weather in Washington D.C. for the half one year (from 1st, Apr. to 30th, Sep. 2014) by a sector diagram in Figure 2. B). As the proportion of snowy hours is very small, it does not appear in the diagram. Meteorology data for more than 10 recent years in Washington D.C. can be downloaded at http://mesowest.utah.edu/cgi-bin/droman/download_ndb.cgi?stn=KADW&year1=2015&day1=29&month1=10&hour1=&timetype=GMT&unit=0.



A) New York City B) Washington D.C.

Fig 2. Weather Distribution