

THE BUILT ENVIRONMENT AND HEALTH PROJECT



MAILMAN SCHOOL
of PUBLIC HEALTH

EPIDEMIOLOGY



NYC DOHMH PATS Data Dictionary for GPS Points

Prepared By:

Daniel M. Sheehan

Geographer, GIS Analyst

Built Environment and Health Research Group

Department of Epidemiology

Columbia University

722 West 168th Street, R735

New York, New York 10032

www.beh.columbia.edu

dms2203@cumc.columbia.edu

Prepared For:

Dr. Andrew Rundle

Associate Professor of Epidemiology

Built Environment and Health Research Group

Department of Epidemiology

Columbia University

722 West 168th Street, R730

New York, New York 10032

www.beh.columbia.edu

agr3@columbia.edu

Introduction	6
Target Temporal Resolution for Data	6
Study Area	6
Abbreviations & Acronyms	6
Data Structure	6
Points	6
GPS Points Data	7
doh_gps_points.csv	7
GPS point Variables	7
participantid	7
gpsuid	7
firsttravdate	7
gpstripid	7
gpstravdayid	7
gpstravdaytripid	7
stageid	7
localid	7
date_local	7
time_local	7
longitude	7
latitude	7
speed	8
pa	8
ped	8
dualactivity	8
dualaxis	8
thirdaxis	8
incline	8
point_x	8
point_y	8
did	8
pt_type	8
objectid	8
Intersect (_int) and Near (_nr_dist, _nr_angle) Variable Suffixes	9
*** stands in for any possible variable prefix.	9
***_int	9
***_nr_dist	9
***_nr_angle	9
All near feature calculations had a search radius limit of 250 meters.	9
Built Features	9
^^^ stands in for the following possible variable suffixes (_int, _nr_dist, _nr_angle).	9
bldgftpt_^^^	9

sidewalk_^^^	9
roadbed_^^^	9
Parks Features	10
^^^ stands in for the following possible variable suffixes (_int, _nr_dist, _nr_angle).	10
baseball_^^^	10
basketball_^^^	10
beaches_^^^	10
golfcourses_^^^	10
handball_^^^	10
multipurpose_^^^	10
parks_^^^	10
playgrounds_^^^	10
pools_^^^	10
soccerfootball_^^^	10
sprayshower_^^^	10
tennis_^^^	10
tracks_^^^	10
waterfountain_^^^	10
Road Type Variables	11
centerline2013_^^^	11
centerline2013_number_tra	11
centerline2013_number_par	11
centerline2013_number_tot	11
centerline2013_width	11
Greater than 2 travel lanes (NUMBER_TRAVEL_LANES)	11
centerline2013lanesg2_^^^	11
centerline2013lanesg2_number_tra	11
centerline2013lanesg2_number_par	11
centerline2013lanesg2_number_tot	11
centerline2013lanesg2_width	11
Less than or equal to 2 travel lanes (NUMBER_TRAVEL_LANES)	11
centerline2013lanesl2_^^^	11
centerline2013lanesl2_number_tra	11
centerline2013lanesl2_number_par	11
centerline2013lanesl2_number_tot	11
centerline2013lanesl2_width	12
Roadway Type	12
centerline2014_^^^	12
centerline2014_rw_type	12
Highway Centerline	12
centerline2014highway_^^^	12
centerline2014highway_rw_type	12
Truck Routes	12
truckroutesall_^^^	12

truckroutescl_^^^	12
truckroutesthr_^^^	12
Air Pollution	13
airpollu_nr_latfid100m	13
airpollu_nr_dist	13
airpollu_nr_angle	13
Round 1 NYCCAS Data Delivery Variables	13
airpollu_nr_pm25_pred	13
airpollu_nr_bc_pred	13
Round 2 NYCCAS Data Delivery Variables - no detailed information, just data with variable names	13
airpollu_nr_no2w1	13
airpollu_nr_no2s1	13
airpollu_nr_no2w2	13
airpollu_nr_no2s2	13
airpollu_nr_no2w3	14
airpollu_nr_no2s3	14
airpollu_nr_no2w4	14
airpollu_nr_no2s4	14
airpollu_nr_no2w5	14
airpollu_nr_no2s5	14
airpollu_nr_no2annavg1	14
airpollu_nr_no2annavg2	14
airpollu_nr_no2annavg3	14
airpollu_nr_no2annavg4	14
airpollu_nr_no2annavg5	14
airpollu_nr_pmannavg1	14
airpollu_nr_pmannavg2	14
airpollu_nr_pmannavg3	14
airpollu_nr_pmannavg4	14
airpollu_nr_pmannavg5	14
airpollu_nr_pmw1	14
airpollu_nr_pms1	14
airpollu_nr_pmw2	15
airpollu_nr_pms2	15
airpollu_nr_pmw3	15
airpollu_nr_pms3	15
airpollu_nr_pmw4	15
airpollu_nr_pms4	15
airpollu_nr_pmw5	15
airpollu_nr_pms5	15
airpollu_nr_anavg2010no	15
airpollu_nr_summer2010o3	15
airpollu_nr_so2w1	15
airpollu_nr_so2w2	15
airpollu_nr_so2w3	15
airpollu_nr_so2w4	15

INTRODUCTION

The Built Environment and Health Project [beh.columbia.edu] (BEH) received data from the New York City Department of Health and Mental Hygiene (DOHMH) tracking 803 New York City study participants from 2011.

Number of GPS subjects: **803**

Total number of GPS points collected: **9,645,374**

GPS points dataset (n = 2,783,574)

Blocked points dataset (n = 6,861,800)

Target Temporal Resolution for Data

The target time to get data for this project is 2011.

Study Area

New York City and surrounding area as well as extent of GPS data points. The data for this particular dictionary is mostly restricted to New York City.

Abbreviations & Acronyms

BEH - Built Environment and Health Group

BEH-GIS - Built Environment and Health Geographic Information Systems Team

DOHMH - New York City Department of Health and Mental Hygiene

DATA STRUCTURE

Points

Every GPS point (xyuid) is a row in the GPS points data.

GPS POINTS DATA

The following fields are fields that are in the master GPS Points dataset.

Filename:

doh_gps_points.csv

GPS point Variables

Many of the following variable definitions are copied directly from the GeoStats data dictionary excel files.

participantid

Six-digit participant id number.

gpsuid

Unique GPS point ID.

GPS points dataset (n = 2783574)

```
df[ 'gpsuid' ] = df[ 'uid' ] + 10000001
```

Blocked points dataset (n = 6861800)

```
df[ 'gpsuid' ] = df[ 'uid' ] + 20000001
```

firsttravdate

First travel date for specific deployment period.

gpstripid

Trip number within file.

gpstravdayid

Travel day number.

gpstravdaytripid

Trip number within the travel day.

stageid

Stage ID from trip stages table.

localid

Point index within original file.

date_local

Local date.

time_local

Local time.

longitude

Longitude (dd WGS84) of point.

latitude

Latitude (dd WGS84) of point.

speed

Point speed in MPH.

pa

Aggregate physical activity level assigned to the minute in which the gps point falls.

ped

Aggregate step count assigned to the minute in which the gps point falls.

dualactivity

See below.

dualaxis

Aggregate acceleration data on horizontal (x) axis and vertical (y) axis assigned to the minute in which the gps point falls

BLK POINTS AND GPS POINTS DON'T SHARE THESE FIELDS IN THEIR DATASET - DO THEY MAYBE MEAN THE SAME THING?

thirdaxis

Aggregate acceleration data on horizontal (x) axis,vertical (y) axis and horizontal front-back (z) axis assigned to the minute in which the gps point falls.

incline

0 = device is off, 1 = standing, 2 = lying horizontally, 3 = sitting upright.

point_x

X Coordinate (NAD_1983_UTM_Zone_18N) of point.

point_y

Y Coordinate (NAD_1983_UTM_Zone_18N) of point.

did

GIS variable id, not for analysis.

pt_type

Point type.

1 is from GPS point dataset.

2 is from Blocked point dataset.

objectid

GIS variable id, not for analysis.

Intersect (_int) and Near (_nr_dist, _nr_angle) Variable Suffixes

*** stands in for any possible variable prefix.

***_int

GPS point intersects with the feature.

1 - intersects feature.

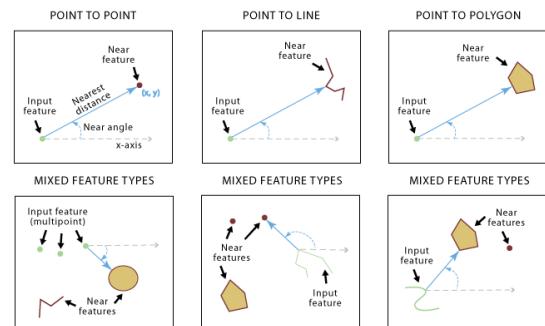
0 - does not intersect feature.

***_nr_dist

Distance to any feature (in meters). Distance of 0 meters is inside or on a feature.

Near function: Determines the distance from each feature in the input features to the nearest feature in the near features, within the search radius. Near is based on the units of the coordinate system of the input features.

Information from Esri: (<http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/00080000001q000000>)



***_nr_angle

Angle away from near boundary of Rio Das Pedras (in degrees - determine where 0 deg is).

Specifies whether the near angle values in decimal degrees will be calculated and written to a new field, NEAR_ANGLE. A near angle measures from the x-axis (horizontal axis) to the direction of the line connecting an input feature to its nearest feature at their closest locations, and it is within the range of 0 to 180 or 0 to -180 decimal degrees - 0 to the east, 90 to the north, 180 (-180°) to the west, and -90 to the south.

NO_ANGLE —Specifies that the near angle values will not be written. This is the default.

ANGLE —Specifies that the near angle values will be written to the NEAR_ANGLE field.

All near feature calculations had a search radius limit of 250 meters.

If no feature is found within the search radius the values of these fields will be -1.

Built Features

^^^ stands in for the following possible variable suffixes (_int, _nr_dist, _nr_angle).

These features were downloaded from NYC Open Data.

bldgftpt_^^^

NYC Building footprint polygon feature.

sidewalk_^^^

NYC Sidewalk polygon feature.

roadbed_^^^

NYC Roadbed polygon feature.

Parks Features

^^^ stands in for the following possible variable suffixes (`_int`, `_nr_dist`, `_nr_angle`).

These Parks features are based on the 2009/2010 NYC DPR Datamine downloads.

baseball_ ^^^

NYCDPR Baseball.

basketball_ ^^^

NYCDPR Basketball.

beaches_ ^^^

NYCDPR Beaches.

golfcourses_ ^^^

NYCDPR Golf Courses.

handball_ ^^^

NYCDPR Handball Courts.

multipurpose_ ^^^

NYCDPR Multipurpose use.

parks_ ^^^

NYCDPR Parks features.

playgrounds_ ^^^

NYCDPR Playgrounds.

pools_ ^^^

NYCDPR Pools.

soccerfootball_ ^^^

NYCDPR Soccer and/or Football fields.

sprayshower _ ^^^

NYCDPR Sprayshower. No Intersection variable as this feature is point.

tennis_ ^^^

NYCDPR Tennis Courts.

tracks_ ^^^

NYCDPR Tracks.

waterfountain_ ^^^

NYCDPR Waterfountains. No Intersection variable as this feature is point.

Road Type Variables

centerline2013_^^^

Nearest Centerline 2013. Feature was downloaded from NYC Open Data in 2013, but represents data derived from 2006 Photogrammetry data.

centerline2013_number_tra

Nearest Centerline 2013 number of travel lanes.

centerline2013_number_par

Nearest Centerline 2013 number of parking lanes.

centerline2013_number_tot

Nearest Centerline 2013 number of total lanes.

centerline2013_width

Nearest Centerline 2013 width (feet).

Greater than 2 travel lanes (NUMBER_TRAVEL_LANES)

centerline2013lanesgt2_^^^

Nearest Centerline 2013 greater than 2 travel lanes. Feature was downloaded from NYC Open Data in 2013, but represents data derived from 2006 Photogrammetry data.

centerline2013lanesgt2_number_tra

Nearest Centerline 2013 greater than 2 travel lanes number of travel lanes.

centerline2013lanesgt2_number_par

Nearest Centerline 2013 greater than 2 travel lanes number of parking lanes.

centerline2013lanesgt2_number_tot

Nearest Centerline 2013 greater than 2 travel lanes number of total lanes.

centerline2013lanesgt2_width

Nearest Centerline 2013 greater than 2 travel lanes width (feet).

Less than or equal to 2 travel lanes (NUMBER_TRAVEL_LANES)

centerline2013lanesle2_^^^

Nearest Centerline 2013 less than or equal to 2 travel lanes. Feature was downloaded from NYC Open Data in 2013, but represents data derived from 2006 Photogrammetry data.

centerline2013lanesle2_number_tra

Nearest Centerline 2013 less than or equal to 2 travel lanes number of travel lanes.

centerline2013lanesle2_number_par

Nearest Centerline 2013 less than or equal to 2 travel lanes number of parking lanes.

centerline2013lanesle2_number_tot

Nearest Centerline 2013 less than or equal to 2 travel lanes number of total lanes.

centerline2013lanesl2_width

Nearest Centerline 2013 less than or equal to 2 travel lanes width (feet).

Roadway Type

centerline2014_^^^

Centerline 2014, NYC Centerline File with Roadway Type designation.

centerline2014_rw_type

Roadway Type:

- 1 Street
- 2 Highway
- 3 Bridge
- 4 Tunnel
- 5 Boardwalk
- 6 Path/Trail
- 7 Step Street
- 8 Driveway
- 9 Ramp
- 10 Alley
- 11 Unknown
- 12 Non-Physical Street Segment
- 13 U-Turn
- 14 Ferry Route

Highway Centerline

centerline2014highway_^^^

Centerline 2014 Highway (Roadway Type). RW_Type = 2.

centerline2014highway_rw_type

Roadway Type:

- 2 Highway

Truck Routes

Data from NYC 2011 April.

truckroutesall_^^^

Truck Routes all.

truckrouteslcl_^^^

Local Truck Routes.

truckroutesthr_^^^

Through Truck Routes.

Air Pollution

airpollu_nr_latfid100m

Near LATFID100m (air pollution grid point id) to GPS point.

airpollu_nr_dist

Near air pollution grid point to GPS point distance (meters).

airpollu_nr_angle

Near air pollution grid point to GPS point angle.

Round 1 NYCCAS Data Delivery Variables

airpollu_nr_pm25_pred

Near air pollution grid point pm25_pred - Average PM 2.5 value from NYCCAS data Round 1. PM2.5 units are ug/m3.

airpollu_nr_bc_pred

Near air pollution grid point bc_pred - Average Black Carbon value from NYCCAS data Round 1. BC is absorbance.

Round 2 NYCCAS Data Delivery Variables - no detailed information, just data with variable names

Year 1 -5 (Dec. 2008- Dec. 2013) annual average for NO2 (ppb) and PM2.5 (ug/m3)

Year 1-5 winter average for NO2, PM2.5 and SO2 (ppb)

Year 1-5 summer average for NO2 and PM2.5

Year 1 and 2 (2009-2010) annual average for NO (ppb)

Year 1 and 2 summer average for ozone (ppb)

Bureau of Environmental Surveillance and policy

NYC DOHMH

10/19/2015

Sarah Johnson

646-632-6543

airpollu_nr_no2w1

Near air pollution grid point no2w1.

airpollu_nr_no2s1

Near air pollution grid point no2s1.

airpollu_nr_no2w2

Near air pollution grid point no2w2.

airpollu_nr_no2s2

Near air pollution grid point no2s2.

airpollu_nr_no2w3

Near air pollution grid point no2w3.

airpollu_nr_no2s3

Near air pollution grid point no2s3.

airpollu_nr_no2w4

Near air pollution grid point no2w4.

airpollu_nr_no2s4

Near air pollution grid point no2s4.

airpollu_nr_no2w5

Near air pollution grid point no2w5.

airpollu_nr_no2s5

Near air pollution grid point no2s5.

airpollu_nr_no2annavg1

Near air pollution grid point no2annavg1.

airpollu_nr_no2annavg2

Near air pollution grid point no2annavg2.

airpollu_nr_no2annavg3

Near air pollution grid point no2annavg3.

airpollu_nr_no2annavg4

Near air pollution grid point no2annavg4.

airpollu_nr_no2annavg5

Near air pollution grid point no2annavg5.

airpollu_nr_pmnavg1

Near air pollution grid point pmnavg1.

airpollu_nr_pmnavg2

Near air pollution grid point pmnavg2.

airpollu_nr_pmnavg3

Near air pollution grid point pmnavg3.

airpollu_nr_pmnavg4

Near air pollution grid point pmnavg4.

airpollu_nr_pmnavg5

Near air pollution grid point pmnavg5.

airpollu_nr_pmw1

Near air pollution grid point pmw1.

airpollu_nr_pms1

Near air pollution grid point pms1.

airpollu_nr_pmw2

Near air pollution grid point pmw2.

airpollu_nr_pms2

Near air pollution grid point pms2.

airpollu_nr_pmw3

Near air pollution grid point pmw3.

airpollu_nr_pms3

Near air pollution grid point pms3.

airpollu_nr_pmw4

Near air pollution grid point pmw4.

airpollu_nr_pms4

Near air pollution grid point pms4.

airpollu_nr_pmw5

Near air pollution grid point pmw5.

airpollu_nr_pms5

Near air pollution grid point pms5.

airpollu_nr_anavg2010no

Near air pollution grid point anavg2010no.

airpollu_nr_summer2010o3

Near air pollution grid point summer2010o3.

airpollu_nr_so2w1

Near air pollution grid point so2w1.

airpollu_nr_so2w2

Near air pollution grid point so2w2.

airpollu_nr_so2w3

Near air pollution grid point so2w3

airpollu_nr_so2w4

Near air pollution grid point so2w4.

airpollu_nr_so2w5

Near air pollution grid point so2w5.

PROJECTION INFORMATION

NAD_1983_UTM_Zone_18N

WKID: 26918 Authority: EPSG

Projection: Transverse_Mercator

False_Easting: 500000.0

False_Northing: 0.0

Central_Meridian: -75.0

Scale_Factor: 0.9996

Latitude_Of-Origin: 0.0

Linear Unit: Meter (1.0)

Geographic Coordinate System: GCS_North_American_1983

Angular Unit: Degree (0.0174532925199433)

Prime Meridian: Greenwich (0.0)

Datum: D_North_American_1983

Spheroid: GRS_1980

Semimajor Axis: 6378137.0

Seminor Axis: 6356752.314140356

Inverse Flattening: 298.257222101