

THE BUILT ENVIRONMENT AND HEALTH PROJECT



Columbia Population Research Center
Advancing Research in Population, Health, and Society



COLUMBIA
UNIVERSITY

MAILMAN SCHOOL
of PUBLIC HEALTH

EPIDEMIOLOGY



Driscoll Project: Moderate/Vigorous Activity (from GPS & ACC data) & the Built Environment Data Dictionary

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:

Stephanie Lovinsky-Desir, MD
Assistant Professor of Pediatrics at CUMC
Columbia University
sl3230@cumc.columbia.edu
Rachel L. Miller, MD
Professor of Medicine (In Pediatrics) and
Environmental Health Sciences
Columbia University
rlm14@cumc.columbia.edu

Introduction	6
Study Area	6
Abbreviations & Acronyms	6
Summary Count Information	6
GPS/ACL Points MERGE MASTER DataSET	7
gps_vars_merge_acl.csv	7
Contributing master files:	7
gps.csv	7
acl_epoch.csv	7
sid	7
GPS point Variables	7
gpsuid	7
altitude	8
e_w	8
dage	8
date_takedown	8
datetime_rnd	8
distance	8
dsta	8
gpscnt	8
gps_day_index	8
gps_locaaldatetime_tmpr	8
gps_localgps_datetime	8
gps_sourcefile	8
hdop	8
heading	8
height	8
latitude	8
longitude	9
localdate	9
localdatetime	9
localtime	9
ms	9
n_s	9
nsat_used_view_	9
pdop	9
utcgps_datetime	9
utctime	9
rcr	9
satinfo_sid-ele-azi-snr_	9
speed	9
valid	9

vdop	9
utcdate	9
Actical File Variables	10
acluid	10
acl_datetime_tmp	10
acl_locaaldatetime	10
acl_sourcefile	10
activity_counts	10
activity_intensity	10
date	10
datetime	10
datetime_rnd	10
day	10
device_placed	11
dups_acl	11
elapsed_seconds	11
energy	11
epoch	11
event_marker	11
index_col	11
sampling_day	11
sampling_time	11
steps	11
time	11
Census Block 2010 Intersect	11
censusblocks2010_int	11
Intersect (_int) and Near (_nr_dist) Variable Suffixes	11
*** stands in for any possible variable prefix.	11
***_int	11
***_nr_dist	11
All near feature calculations had a search radius limit of 250 meters.	12
Built Features	12
^^^ stands in for the following possible variable suffixes (_int, _nr_dist, _nr_angle).	12
bldgftpt_^^^	12
sidewalk_^^^	12
roadbed_^^^	12
Parks Features	12
^^^ stands in for the following possible variable suffixes (_int, _nr_dist, _nr_angle).	12
baseball_^^^	12
basketball_^^^	12
beaches_^^^	12
golfcourses_^^^	12
handball_^^^	12
multipurpose_^^^	12
parks_^^^	13

playgrounds_^^^	13
pools_^^^	13
soccerfootball_^^^	13
sprayshower _ ^^^	13
tennis_^^^	13
tracks_^^^	13
waterfountain_^^^	13
Road Type Variables	13
centerline2013_^^^	13
centerline2013_number_tra	13
centerline2013_number_par	13
centerline2013_number_tot	13
centerline2013_width	13
Greater than 2 travel lanes (NUMBER_TRAVEL_LANES)	13
centerline2013lanesg2_^^^	13
centerline2013lanesg2_number_tra	13
centerline2013lanesg2_number_par	14
centerline2013lanesg2_number_tot	14
centerline2013lanesg2_width	14
Less than or equal to 2 travel lanes (NUMBER_TRAVEL_LANES)	14
centerline2013lanesl2_^^^	14
centerline2013lanesl2_number_tra	14
centerline2013lanesl2_number_par	14
centerline2013lanesl2_number_tot	14
centerline2013lanesl2_width	14
Roadway Type	14
centerline2014_^^^	14
centerline2014_rw_type	14
Highway Centerline	15
centerline2014highway_^^^	15
centerline2014highway_rw_type	15
Truck Routes	15
truckroutesall_^^^	15
truckrouteslcl_^^^	15
truckroutesthr_^^^	15
Air Pollution	15
airpollu_nr_latfid100m	15
airpollu_nr_dist	15
Round 1 NYCCAS Data Delivery Variables	15
airpollu_nr_pm25_pred	15
airpollu_nr_bc_pred	15
Round 2 NYCCAS Data Delivery Variables	15
airpollu_nr_no2w1	16

airpollu_nr_no2s1	16
airpollu_nr_no2w2	16
airpollu_nr_no2s2	16
airpollu_nr_no2w3	16
airpollu_nr_no2s3	16
airpollu_nr_no2w4	16
airpollu_nr_no2s4	16
airpollu_nr_no2w5	16
airpollu_nr_no2s5	16
airpollu_nr_no2annavg1	16
airpollu_nr_no2annavg2	16
airpollu_nr_no2annavg3	16
airpollu_nr_no2annavg4	16
airpollu_nr_no2annavg5	17
airpollu_nr_pmnavg1	17
airpollu_nr_pmnavg2	17
airpollu_nr_pmnavg3	17
airpollu_nr_pmnavg4	17
airpollu_nr_pmnavg5	17
airpollu_nr_pmw1	17
airpollu_nr_pms1	17
airpollu_nr_pmw2	17
airpollu_nr_pms2	17
airpollu_nr_pmw3	17
airpollu_nr_pms3	17
airpollu_nr_pmw4	17
airpollu_nr_pms4	17
airpollu_nr_pmw5	17
airpollu_nr_pms5	17
airpollu_nr_anavg2010no	17
airpollu_nr_summer2010o3	18
airpollu_nr_so2w1	18
airpollu_nr_so2w2	18
airpollu_nr_so2w3	18
airpollu_nr_so2w4	18
airpollu_nr_so2w5	18
Projection Information	19

INTRODUCTION

The Built Environment and Health Project [beh.columbia.edu] (BEH) received data from the Driscoll Project Group (DPG) tracking New York City study participants.

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

DPG - Driscoll Project Group

GPS - Global Positioning System data, coordinate and geographic information.

ACL - Accelerometer data, actical file.

Summary Count Information

The master data set includes **366,076** records

Count of unique **acl** epochs is: **1,261,599**

Count of unique **gps** points is: **392,578**

Count of unique study subjects with **acl** epochs is: **162**

Count of unique study subjects with **gps** records is: **96**

Count of unique study subjects days with **acl** epochs is: **1,061**

Count of unique study subjects days with **gps** records is: **339**

Count of merged GPS points/corresponding ACL data: **366,076**

GPS/ACL POINTS MERGE MASTER DATASET

The joined master data set includes 366,076 records

Filename:

gps_vars_merge_acl.csv

Filepath:

/driscoll/tasks/201602_merge_gps_accel/data/output/gps_vars_merge_acl.csv

Contributing master files:

The input GPS data included 392,578 records

Filename:

gps.csv

CSV Filepath:

driscoll/tasks/201601_read_and_match_gps_accel/data/input/gps/gps.csv

GIS Filepath (including necessary accompanying shapefile file(s) and file(s) extensions):

driscoll/tasks/201601_read_and_match_gps_accel/data/output/gps/
gps_z18n.shp

The input ACL data included 1,261,599 records

Filename:

acl_epoch.csv

Filepath:

driscoll/tasks/201601_read_and_match_gps_accel/data/output/acl/
acl_epoch.csv

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

sid

Subject id number.

GPS point Variables

Note: Not all fields were in every input GPS file and many of these columns were not defined before being passed to BEH-GIS.

gpsuid

Unique GPS point ID assigned by BEH-GIS.

GPS points dataset (n = 392,578)

```
df['gpsuid'] = 10000000000 + (df['sid'].astype(float) * 1000000) +  
(df['set'].astype(float) * 1000000) + df['gps_day_index']
```

altitude

GPS Source file column, altitude.

e_w

GPS Source file column, input as E/W.

dage

GPS Source file column, dage.

date_takedown

GPS Source file column, date takedown.

datetime_rnd

Rounded datetime to the minute. Rounded down to the minute recorded. For example 11:37:49 (49 seconds) yields 11:37.

```
dfG[ 'datetime_rnd' ] = dfG[ 'gps_localdatetime_tmp' ].map(lambda x:  
    x.strftime( '%Y-%m-%d %H:%M' ) )
```

distance

GPS Source file column, distance.

dsta

GPS Source file column, dsta.

gpscnt

GPS Count flag, should be 1.

gps_day_index

GPS Source file column, day_index.

gps_localdatetime_tmp

Temporary datetime column for rounding purposes.

```
dfG[ 'gps_localdatetime_tmp' ] = pd.to_datetime(dfG[ 'localdatetime' ])
```

gps_localgps_datetime

Local datatime.

gps_sourcefile

Filename of input GPS data contributing to this row (record).

hdop

GPS Source file column, unknown.

heading

GPS Source file column, heading.

height

GPS Source file column, height.

latitude

GPS Source file column, input latitude (dd WGS84) of point.

longitude

GPS Source file column, input Longitude (dd WGS84) of point.

localdate

GPS Source file column, input as LOCAL DATE.

localdatetime

GPS Source file column, localdatetime.

```
df['localdatetime'] = pd.to_datetime(df['localdate'] + ' ' +  
df['localtime'])
```

localtime

GPS Source file column, , input as LOCAL TIME

ms

GPS Source file column, ms.

n_s

GPS Source file column, nsat_used_view_ , input as N/S.

nsat_used_view_

GPS Source file column, nsat_used_view_.

pdop

GPS Source file column, pdop.

utcgps_datetime

UTC time for gps record.

```
df['utcdatetime'] = pd.to_datetime(df['utcdate'] + ' ' + df['utctime'])
```

utctime

GPS Source file column, utctime.

rcr

GPS Source file column, rcr.

satinfo_sid-ele-azi-snr_

GPS Source file column, satinfo_sid-ele-azi-snr_.

speed

GPS Source file column, speed.

valid

GPS Source file column, valid.

vdop

GPS Source file column, vdop.

utcdate

GPS Source file column, utcdate.

Actical File Variables

Note: Not all fields were in every input ACL file and many of these columns were not defined before being passed to BEH-GIS.

acluid

Unique Actical record ID assigned by BEH-GIS.

ACL points dataset (n = 1,261,599)

```
df[ 'acluid' ] = 90000000000 + (df[ 'sid' ].astype(float) * 1000000) +
(df[ 'sampling_time' ].astype(float) * 100000) +
(df[ 'sampling_day' ].astype(float) * 10000) + df[ 'index_col' ]
```

acl_datetime_tmp

```
dfa[ 'acl_datetime_tmp' ] = pd.to_datetime(dfa[ 'datetime' ])
```

acl_locaaldatetime

ACL Source file column, localdatetime.

```
dfa = dfa.rename(columns=lambda x: x.replace('locaaldatetime',
'acl_locaaldatetime'))
```

acl_sourcefile

Filename of input ACL data contributing to this row (record).

activity_counts

ACL Source file column, activity counts.

activity_intensity

ACL Source file column, activity intensity.

(1=sedentary)
(2=light)
(3=moderate)
(4=vigorous)

date

ACL Source file column, date.

datetime

ACL Source file column, datetime.

datetime_rnd

Datetime round, duplicate column for ACL file.

```
dfa[ 'datetime_rnd' ] = dfa[ 'acl_datetime_tmp' ].map(lambda x: x.strftime( '%Y-
%m-%d %H:%M' ))
```

day

ACL Source file column, day.

device_placed

Location where device was placed, mined from filename/file header.

dups_acl

Duplicate check column for ACL data, should be 0.

elapsed_seconds

ACL Source file column, Elapsed seconds.

energy

ACL Source file column, energy.

epoch

ACL Source file column, epoch.

event_marker

ACL Source file column, event marker.

index_col

ACL Source file column, index_col.

sampling_day

ACL Source file column, sampling day.

sampling_time

ACL Source file column, sampling time.

steps

ACL Source file column, steps.

time

ACL Source file column, time.

Census Block 2010 Intersect

censusblocks2010_int

Census block id intersected by GPS points in New York City 2010. Department of City Planning.

Intersect (_int) and Near (_nr_dist) Variable Suffixes

*** stands in for any possible variable prefix.

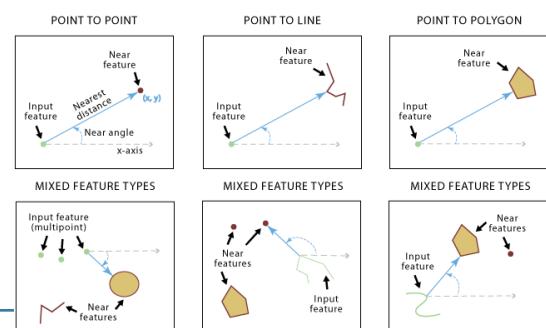
*****_int**

GPS point intersects with the feature.

1 - intersects feature.

null/NaN - does not intersect feature.

*****_nr_dist**



Distance to any feature (in meters). Distance of 0 meters is inside or on a feature. -1 means no feature was found within 250 meter search radius.

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>)

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)

centerline2013lanesl2_^^^

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.

centerline2013lanesl2_number_tra

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

centerline2013lanesl2_number_par

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

centerline2013lanesl2_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).

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

Important Note: Many have no detailed information, just data with variable names as delivered.

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_pmannavg1

Near air pollution grid point pmannavg1.

airpollu_nr_pmannavg2

Near air pollution grid point pmannavg2.

airpollu_nr_pmannavg3

Near air pollution grid point pmannavg3.

airpollu_nr_pmannavg4

Near air pollution grid point pmannavg4.

airpollu_nr_pmannavg5

Near air pollution grid point pmannavg5.

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