



MAILMAN SCHOOL
of PUBLIC HEALTH

EPIDEMIOLOGY



Driscoll Project 2015 Time Estimate: Moderate/Vigorous Activity (from GPS & ACC data) & the Built Environment

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
s13230@cumc.columbia.edu
&
Rachel L. Miller, MD
Professor of Medicine (In Pediatrics) and
Environmental Health Sciences
Columbia University
rlm14@cumc.columbia.edu

Time Estimate Introduction	3
Target Temporal Resolution for Data	3
Study Area	3
Abbreviations & Acronyms	3
Tasks	3
Preprocessing & Data Carpentry	3
Land Use and Built Variables	4
Road Type Variables	5
Air Pollution Variables	6
Time Estimate Totals	6

TIME ESTIMATE INTRODUCTION

This time estimate represents the total number of hours required to construct each Type or Category of GIS metric(s) and task(s) for the Driscoll Project 2015.

Target Temporal Resolution for Data

The target time to get data for this project is **2015**.

Study Area

New York City (the extent of the study area).

Abbreviations & Acronyms

BEH - Built Environment and Health Research Group.

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

ACC - Accelerometer data.

MVA - Moderate/Vigorous Activity as classified by accelerometer (ACC) data.

TASKS

Preprocessing & Data Carpentry

Write script(s) that opens each ACC (accelerometer) file and reads the epoch-by-epoch data into a data frame. For each 1 day ACC file there will probably only be at best about an hour of **vigorous** or **moderate** activity. The time stamps from these ACC epochs then need to be used to identify GPS points that occurred during the activity periods. These activity-associated waypoints then need to be characterized in the task categories following Preprocessing and Data Carpentry.

Data Source(s):

1. 2 ACC (accelerometer) file(s) for each study subject (n=280, 140 subjects).
 - A. 1 for continuous 24-hour monitoring.
 - B. 1 for final 24 hours of 5-day monitoring.
2. 2 GPS (global positioning system) files(s) for each study subject. (n=280, 140 subjects).
 - A. 1 for continuous 24-hour monitoring.
 - B. 1 for final 24 hours of 5-day monitoring.

There are **140** children for whom there is a **24 hour** monitoring period with both Global Position System (**GPS**) data and accelerometer (**ACC**) data and there is **5 day** GPS monitoring period where the last day also includes ACC monitoring. So there should be **280** ACC files.

Each ACC file has a header with;

1. Subject ID
2. Gender
3. Age

4. Summary Statistics
5. Time stamped epoch by epoch activity data with counts
6. Energy expenditure
7. Flag telling us if the activity epoch was;
 - A. Sedentary
 - B. Light
 - C. Moderate
 - D. Vigorous

Task(s):

Preprocessing & Data Carpentry	Low Estimate	High Estimate
1. Prepare and clean ACC data from initial format (csv) into workable project-ready format.	6	10
2. Prepare and clean GPS data from initial format (gpx) into workable project-ready format.	6	10
3. Join ACC and GPS data based on time stamp and create master input file.	12	20
Total:	24	40

Land Use and Built Variables

Determine what MVA takes place inside/outside buildings, in parks/park facilities, or built features road/sidewalk via intersect:

Data Source(s):

1. NYC Building Footprints [<https://data.cityofnewyork.us/Housing-Development/Building-Footprints/tb92-6tj8>]
2. Open Space, Parks, Sports facilities:
 - A. Open Space
 - B. Park
 - C. Facilities - 2009/2010 - Data Mine (polygons) - from BEH Archive as NYC Data Mine is now defunct.
 - i. Baseball
 - ii. Basketball
 - iii. Beaches
 - iv. Golf Course
 - v. Handball
 - vi. Multipurpose
 - vii. Parks
 - viii. Playgrounds
 - ix. Pools

- x. Soccer/Football
 - xi. Tennis
 - xii. Tracks
3. Built Features
- A. Sidewalk [<https://data.cityofnewyork.us/City-Government/Sidewalk-Features/vfx9-tbb6>]
 - B. Roadbed [<https://data.cityofnewyork.us/City-Government/Roadbed/xgwd-7vhg>]

Task(s):

Land Use and Built Variables	Low Estimate	High Estimate
1. GPS Point intersect with Building Footprints (indoor vs. outdoor)	8	12
2. GPS Point with Open Space, Parks and Park Facilities (eg. park, playground, ball fields, track, basketball court, etc.)	16	20
3. GPS point intersect with Sidewalks and Roadbeds	12	16
Total:	36	48

Road Type Variables

Determine what MVA takes place near (distance) different road types:

Data Source(s):

1. Road Type
2. Truck Route
3. Highway (might be same source as using road type.).

Task(s):

Road Type Variables	Low Estimate	High Estimate
1. Distance of GPS point to nearest Roadway \leq 2 lanes	3	6
2. Distance of GPS point to nearest Roadway $>$ 2 lanes	3	6
3. Distance of GPS point to nearest Truck Route	4	8
4. Distance of GPS point to nearest Highway	4	8
Total:	14	28

Air Pollution Variables

Categorize MVA exposure to Air Pollution (near, assign near value from raster grid):

Data Source(s):

1. New York City Community Air Survey (NYCCAS) [<http://www.nyc.gov/html/doh/html/environmental/community-air-survey.shtml>] as supplied via email from Dr. Andrew Rundle. The data that BEH-GIS received did not include some of the following variables (**marked in red**).

Task(s):

Air Pollution Variables	Low Estimate	High Estimate
1. Black carbon (this study has personal black carbon measurements for the study subjects so this would be a good variable to use to compare with our personal measurements)	4	8
2. PM2.5 (Particulate Matter)	4	8
3. Nitric Oxide (data not currently available or in BEH data archives)	0	0
4. Nitrogen Dioxide (data not currently available or in BEH data archives)	0	0
5. Ozone (data not currently available or in BEH data archives)	0	0
6. Sulfur Dioxide (data not currently available or in BEH data archives)	0	0
Total:	8	16

TIME ESTIMATE TOTALS

TIME ESTIMATE TOTALS

	Low Estimate	High Estimate
Preprocessing & Data Carpentry	24	40
Land Use and Built Variables	36	48
Road Type Variables	14	28
Air Pollution Variables	8	16
Total in Hours:	82	132
Total in Days:	10.25	16.5
Total in Weeks:	2.05	3.3