



MAILMAN SCHOOL
of PUBLIC HEALTH

EPIDEMIOLOGY



Driscoll Project 2015 Time Estimate: GPS & Accelerometer

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

Time Estimate Introduction	3
Target Temporal Resolution for Data	3
Study Area	3
Abbreviations & Acronyms	3
Tasks Overview	3
Preprocessing & Data Munging	3
Land Use and Built Variables	4
Road Type Variables	5
Air Pollution Variables	6
Task Totals	7

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 Group.

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

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

TASKS OVERVIEW

Preprocessing & Data Munging

Write a script that opens each ACC (accelerometer) file and reads the epoch-by-epoch data into a data frame of spreadsheet file. 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 as follows.

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 are **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 epoch was a
 - A. sedentary
 - B. light
 - C. moderate
 - D. or vigorous activity epoch.

Task(s):

Preprocessing & Data Munging	Low Estimate	High Estimate
1. Prepare and clean ACC data from initial format (csv) into workable project-ready format.	8	16
2. Prepare and clean GPS data from initial format (gpx) into workable project-ready format.	8	16
3. Join ACC and GPS data based on time stamp	12	24
Total:	28	56

Land Use and Built Variables

The first step in achieving my overall goal would be to determine MVA that takes place:

Data Source(s):

1. NYC Building Footprints [<https://data.cityofnewyork.us/Housing-Development/Building-Footprints/tb92-6tj8>]
2. Open Space, Green Space, Sports facilities:
 - A. Open Space
 - B. Park
 - C. Sports Facilities
 - i. Basketball Court

- ii. Handball Court, etc. (find sources)
3. Built Features
- Sidewalk [<https://data.cityofnewyork.us/City-Government/Sidewalk-Features/vfx9-tbb6>]
 - Roadbed [<https://data.cityofnewyork.us/City-Government/Roadbed/xgwd-7vhg>]

Task(s):

Land Use and Built Variables	Low Estimate	High Estimate
1. indoor vs. outdoor	8	12
2. in outdoor green space (eg. park, playground, ball fields, track, basketball court, etc.)	16	20
3. GPS point intersect with Sidewalks or Roadbeds	8	12
Total:	32	44

Road Type Variables

The next set of variables, which actually get to the primary objectives of the study would be to look at:

Data Source(s):

- Road Type
- Truck Route
- 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	4	8
2. Distance of GPS point to nearest Roadway $>$ 2 lanes	4	8
3. Distance of GPS point to nearest Truck Route	4	8
4. Distance of GPS point to nearest Highway	4	8
Total:	16	32

Air Pollution Variables

I don't have experience with using the NYCCAS data so I am a bit unfamiliar with how the data is stored. However, if possible I think it would also be interesting to categorize during MVA, exposure to:

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.

Task(s):

Air Pollution Variables	Low Estimate	High Estimate
1. Black (elemental) 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

TASK TOTALS

TIME ESTIMATE TOTALS

	Low Estimate	High Estimate
Preprocessing & Data Munging	28	56
Land Use and Built Variables	32	44
Road Type Variables	16	32
Air Pollution Variables	8	16
Total in Hours:	84	148
Total in Days:	10.5	18.5
Total in Weeks:	2.1	3.7