## Predicting Building Heights—AWS Hackathon for Good

This repo contains all the information you will need to start working on the Urban Institute’s challenge to create a novel dataset on building heights.

## Overview

We want you to create a dataset of building heights in Washington, DC, using input building footprint data, LIDAR data, satellite imagery, and any other auxiliary data sources you find. Our goal is to generalize the methodologies you develop to create a national dataset of building footprints and heights, so please try to keep auxiliary data sources as generalizable as possible. This work is important for research organizations like us and city governments to achieve the following goals:

• understand what kind of housing currently exists in cities

• create data driven affordable housing plans

• plan effectively for the future growth of cities

The `docs` directory contains an executive summary, a press release, and a FAQ

list, which may be helpful for further background information on this project.

## Data

We recommend starting with the following datasets to generate building heights for DC. These datasets can either be accessed in the `data` folder or at the specified links:

1) Building footprint data: This is a listing of each of the 100,000+ buildings in DC and their polygon building footprints. These are the data we want you to append with height information. The data can be found in geojson format at `data\training-data\DC\_Buildings\_Footprint\_4326\_training.geojson`. Some additional notes about these data follow:

- Data include structures originally captured from 2005 aerial imagery, with updates based on 2010 aerial imagery.

- In many instances, row houses and other contiguous building rooftops could not be broken apart using the process methodology and appear as a single structure.

- In accordance with the geojson specification, the file uses the WGS 1984 projection (CRS: 4362).

- This is a modified version of the [original](https://opendata.dc.gov/datasets/274f7c2b5f7c4ae19f165d9951057a00) dataset, which was in ESRI multipatch format. We have simply stripped a few columns and converted it to the friendlier geojson format using ArcGIS. The original dataset actually has the building heights for each of the buildings because DC was one of the very few cities that contracted a private firm to compute building heights. However, we would like you to predict the building heights of our modified data so we can make the methodology publicly available. If you want, you can use the original dataset to calculate the accuracy of your methodologies.

2) LIDAR point cloud data: Light detection and ranging (LIDAR) is a remote sensing method that emits thousands of light pulses a second to measure distances between the ground and the height of structures. These data essentially contain a series of dense points along with heights. These data are available as tiled `.LAS` files on S3. Access instructions and more information about these files can be found [here](https://docs.opendata.aws/dc-lidar-2018/readme.html). A couple of notes on these data:

- These files are separated into numbered tiled “grids.” You can download a shapefile of the grids (and their associated numbers) by going to the link above.

- For those of you who haven’t worked with LAS files before, there are code examples for how to download and read in LAS files as sf dataframes using R or as geopandas dataframes using Python in the `scripts` directory.

- These files use the NAD83(2011) projection (CRS 6487).