Defining and Simulating Uncertainty

It has been hypothesized by many in the international conservation aid community that aid tends to flow towards areas with dense vegetation (i.e., forests) for protection, while less aid flows to areas with little vegetation for rehabilitation. Your team has been asked to provide evidence for or against this argument, in a country of your choice[1]. To date, our ability to answer this question has been driven by two key factors: first, we have not had information on international conservation aid, and second, when we do have this information, there is considerable uncertainty in if the "full envelope" of all aid is being measured. While simulation methods exist which can help overcome this challenge, they are rarely employed by development practitioners.

1 Project Deliverables

You will need to turn in three deliverables as a part of this project:
(1) A 2-page report summarizing your findings, including the following elements:

- Summary of Findings (1 paragraph summarizing everything you did and the key take-away)

- Figure(s) detailing your findings

- Data and Methods, with enough information for another practitioner to reproduce your approach (2-3 paragraphs). Make sure you simulate uncertainty using at least two distributions.

- Results, with a written description of any tables or figures you produce (1-2 paragraphs)

- Table(s) detailing your findings

- A discussion and conclusion, covering limitations of your approach, take-aways, and next steps; this is where you should discuss how changing your distributional assumptions changed (or not) your results. (1-2 paragraphs)

- A bibliography with any literature you cite.

(2) A CSV dataset containing the results from your simulation. (3) A brief one-paragraph description of the code you used to produce 1 and 2.
(4) You do not need to turn in your python code, as it will already be on the shared Jupyter hub!

   TIP: *Make sure to type in https:// , and that you are on campus! Off-campus access to SciClone is not allowed due to security reasons, unless you use a VPN. If you are attempting to access SciClone off campus, make sure you contact William and Mary IT*

---

[1]Previous research on this topic has been conducted in Honduras, so the donor suggests an alternative location for investigation.

to set up a VPN - it is recommended you do this at least one week before you will need it.

1.1   Getting the Data

Satellite (and aid) data is very difficult to download, process, and use. Luckily, there are a number of online tools available to assist non-expert users. For this excersize, we will use one such platform - www.geoquery.org - which allows you to download spatial data from regions of your choice.

Once on the site, you will select a country of interest, the geographic scale you would like to conduct your analysis at, and then the datasets you would like to acquire. Here you will want to pick at least NDVI (a measure of vegetation, where higher values indicate more vegetation) and General Environmental Protection aid from the World Bank. It is suggested you get NDVI data from 2014, but you can argue for other years. Note that some countries may have received no environmental aid, or limited data is available, so you may have to choose a different country after you get your dataset [2].

1.2   Simulating Uncertainty

Included with this lab is a Jupyter notebook which provides a simple example of simulating uncertainty in a dataset similar to the one you will produce. One of the biggest challenges in simulation is choosing appropriate distributions to model uncertainty with - for example, one might assume that estimates of international aid may be incorrect only in one direction (only omitting aid), or both directions (both omitting aid and having duplicate aid in records). Further, likelihoods of uncertainty might vary (it's likely aid is underestimated by 10%, but very unlikely it's underestimated by 40,000%). The distributions you choose will define your assumptions surrounding uncertainty.

2   Stretch Goals

These goals are optional, and worth a very small amount (up to 5% total of your assignment grade for all goals in total) of extra credit. Completing any one stretch goal gives you the opportunity to receive all 5 points of extra credit.

(1) Acquire and process a satellite dataset on your own, rather than using the online tool. This is very difficult, and only recommended for students with experience with GIS or Remote Sensing.

(2) Add an additional covariate to enhance your targeting analysis, and both (a) defend it theoretically, and (b) provide a justification for how you include uncertainty into the new variable in your simulation approach.

(3)Conduct your analysis at multiple geographic scales and contrast your results.

---

[2]Note that the data request you put in is processed on an HPC, and can take up to 24 hours to complete. Get started on this assignment early!