Analyst Exercise
June 2019

In this exercise, you will be given a dataset with outputs from an agricultural runoff model, and you will be asked to explore, clean, and perform some basic analysis of the data, then briefly summarize your findings. The goal is to get some insight into how you work through a problem, and how you document and present your findings.

Please note that this is mock dataset and we do not expect you to come up with conclusions about how management actions affect nutrient and sediment runoff on each field. This is purely a data munging and visualization exercise.

The data:

1. *DATA.csv:* This dataset includes (fake) outputs from one of our models, which simulates the edge-of-field sediment and nutrient runoff on agricultural fields. See table 1 for additional details on the model outputs included in the dataset. Note that each row in the dataset represents a "scenario", which can either be the current conditions of the field, or an alternative scenario in which some conservation action has been implemented (e.g., edge-of-field buffer planted).

Your tasks:

1. Perform any necessary data cleaning and quality control to get the data ready for analysis.
2. Create two new columns, one for total phosphorus (OrganicP + SolubleP) and another for total nitrogen (OrganicN + NO3).
3. Calculate the "uplift" for total nitrogen, total phosphorus, and sediment by finding the difference between the current condition runoff values and each alternative for each field (i.e., what is the change in runoff under each suite of management actions for every field?).
    a. Note: pay attention to units – both "per acre" and "per field" runoff values can be important and insightful, depending on the situation.
4. Show your findings (assume your target audience is other TFT analysts).

How to deliver your work:

Option 1: RMarkdown or Jupyter notebook (or similar)
If you choose this method, you may include all of your work in a single notebook file, but please include a "findings" section at the end that includes a brief summary along with any graphs communicating your findings.

Option 2: PDF + supporting files
If you choose this method, please include only the brief summary (plus any graphs communicating your findings) in the PDF. All other work should be included in separate files that contain the R and/or Python code used.

Table 1 – description of column headers in DATA.csv

Column Header	Description
fieldID	Unique ID for each agricultural fields with the area of interest
irrigation	Type of irrigation simulated in that scenario
buffer	Type of edge-of-field buffer simulated in that scenario
fence	Type of livestock exclusion fence simulated in that scenario
area	The field area (acres)
scenario_type	-*Current condition*: baseline/current field conditions -*Alternative condition*: some modification from current conditions (e.g., implementation of one or more best management practices
OrganicN	The modeled annual organic nitrogen runoff (lbs/acre)
NO3	The modeled annual nitrate runoff (lbs/acre)
OrganicP	The modeled annual organic phosphorus runoff (lbs/acre)
SolubleP	The modeled annual soluble phosphorus runoff (lbs/acre)
Sediment	The modeled annual sediment runoff (lbs/acre)