


 **Tax Management Associates, Inc.**

 12,805

 Share

 Following 182



**GIS Solutions Challenge**  
📁 [Data Science](#), [Energy, Environment & Resources](#), [Infrastructure](#)  
Develop an open source solution to solve common GIS problems. [Read Overview...](#)

Phase 2

GeoTrue on Github

PREVIEW MY ENTRY

Phase 1

GeoTrue SQL Queries

PREVIEW MY ENTRY

GeoTrue, Open Source Map Routing

PREVIEW MY ENTRY

Legal agreement

Leave challenge

 **Robin Rowe's team**

 STAGE

**Phase 2 Submission Deadline**

 **\$45,000**

- [Overview](#)
- [Guidelines](#)
- [Timeline](#)
- [Updates](#) <sup>11</sup>
- [Forum](#) <sup>7</sup>
- [Community](#) <sup>182</sup>
- [FAQ](#)

The GIS Solutions Challenge asks innovators to develop scalable, efficient, and effective open source tools that generate useful business insights from geospatial data, which can solve the three specific GIS problems listed below.

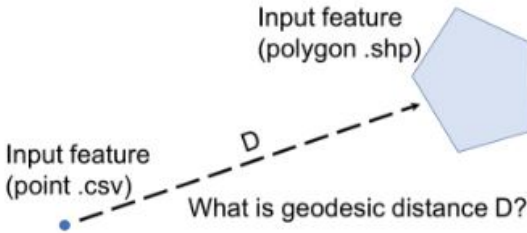
Challenge Goal

Competitors will need to develop an solution to answer one or more of the following questions using an open source analytics platform:

1) What is the geodesic distance between two features?

- Between a point and the closest edge of a polygon
- Between a point and another point
- Between the closest edge of a polygon to the closest edge of another polygon

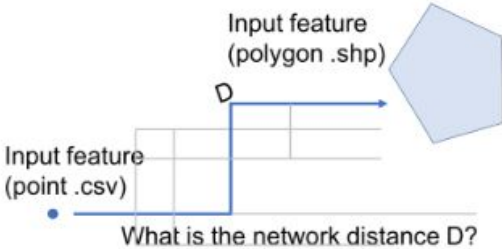
E.g., A particular street corner in Detroit is known to be a crime hotspot. How far is this hotspot from the area the police actively patrols? This distance would be measured as a straight line from point to the edge of a polygon.



2) What is the network distance between two features?

- Between a point and the closest edge of a polygon
- Between a point and another point
- Between the closest edge of a polygon to the closest edge of another polygon

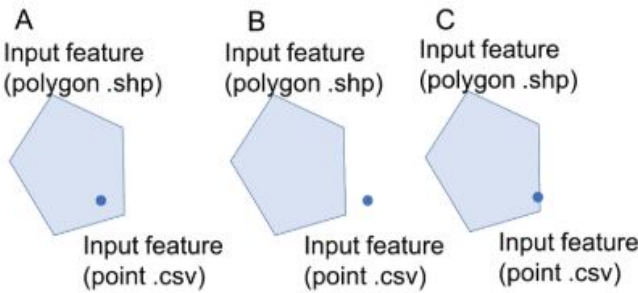
E.g., What is the actual distance police must travel from the edge of their patrol to reach a crime hotspot? This distance would take into account the specific route the police must travel to reach the hotspot.



3) Is a point inside or outside a polygon?

- Is the point completely within the polygon (not including features on the boundary) -- "completely contain within"
- Is the point within the polygon (including features on the boundary) -- "contain within"
- Is the point only touching the boundary (so neither in or out) -- Clementini

E.g., Is the crime hotspot within a police patrol area?



Timeline

The timeline for the challenge can be viewed [here](#).

Phase 1

Provide a proof-of-concept for your GIS Solution. This needs to include a diagram explaining the solution, a sample workflow of your solution, and descriptive information, including why the business insight generated from your tool will be useful to your target audience.

- Prizes: Up to 10 prizes of \$1,000 each
- Feedback will be provided to winning entries
- Non-elimination: competitors may enter Phase 2 regardless of whether they entered or won Phase 1
- Duration: 6 weeks



Criteria	Description	Score
Technical feasibility	<ul style="list-style-type: none"><li>How feasible is the solution based on modern open source GIS capabilities?</li><li>Does the solution’s diagram flow in a reasonable manner?</li><li>Is the solution backed by any GIS science?</li></ul>	50
Usefulness	<ul style="list-style-type: none"><li>How useful is the business insight to the target audience?</li><li>How well does the solution address the described problem?</li><li>Does the solution generate an output that can be interpreted by someone without data science expertise (i.e., visual representation of output)?</li></ul>	40
Innovativeness	<ul style="list-style-type: none"><li>How innovative is the business insight?</li><li>How original is the development compared to existing tools and solutions?</li></ul>	10

## Phase 2

Develop a fully functional GIS Solution. This involves submission of open source code for the solution and finalised documentation, including metadata on each aspect of your solution. Solutions will be evaluated based on speed, scalability, overall architecture, and the business insight they produce. Please see complete judging criteria below.

- Duration: 6 weeks
- Prizes: A total prize pool of \$35,000 is available for Phase 2. TMA plans to award 3 or more solutions with the minimum prize being no lower than \$2,500.

### Phase 2 Minimum Requirements

All solutions must adequately meet the Phase 2 Minimum Requirements in order to be eligible for a prize. Solutions which meet these Phase 2 Requirements will be ranked against the Judging Criteria below.

#### Solution Languages

Java, R, or Python are the acceptable languages for the challenge. These languages have been selected for their broad open source community in relation to data science, as well as their compatibility with the KNIME Analytics platform. It is possible to use another language so long as that language can be run in a Java, R, or Python environment (such as Scala can be compiled into a JAR and run with Java).

#### Submission Format

Submissions should be contained in a git repository hosted on <https://github.com>. The repository can be public, or marked as private and shared with *tma1-dev*. At the root of the repository should be a README file that contains instructions for configuring and running the solution (see Documentation below for README requirements).

The solution must be able to run in a linux terminal (headlessly). The headless part of the solution will be used to create evaluations for quantitative judging metrics.

Although not a requirement, it is strongly recommended that your solution integrates with the KNIME platform. Integration with the KNIME platform makes GIS tools more accessible to local governments and other end users. Integration with the KNIME platform can be done as a new KNIME node, as a workflow containing the solution, or simply as a set of instructions that explains how to integrate the source code of the solution into a given set of KNIME nodes.

At a minimum, the solution must accept CSV and shapefiles as input and generate CSV files as output. Additional marks will be given to solutions that accept other types of data input and generate more user friendly output (e.g., map plots like a png or shapefile) as detailed in the judging criteria.

All components of the solution must be freely available for commercial use or licensed as *LICENSE\_LIST\_GOES\_HERE*.

#### Documentation

Submissions must be well documented for ease of use and ease of understanding. A submission with poor documentation will not be eligible for a prize.

Innovators must have a README file at the root of their git repository that contains instructions for setting up the solution. The README must contain the following:

- Clearly explain how to install all necessary dependencies
- If applicable, document how to configure a KNIME workflow to use the solution and how KNIME integration testing should be performed



Testing Environments

All solutions will be tested headlessly in a linux terminal on Google Cloud Compute Engine using an **n1-standard-4** instances in the us-east1-c region. The instances will run Ubuntu 18.04 and KNIME Analytics Platform (Desktop) 3.6. For the terminal based installation, any relevant dependencies will be installed based on the solution’s README in order to run the solution in a headless fashion. We reserve the right to reject dependencies that require insecure configurations to run (such as adding an unknown apt repository).

Performance

Solutions will be evaluated against a baseline and against all other submissions for speed. Each solution will be evaluated for accuracy, and scalability. For the Detroit crime dataset, comparing geodesic distances of patrol area centroids to crimes for 10,000 data points of crime takes 145 seconds on average according to our baseline. Performing a spatial join of 162,449 points from the Africa conflict dataset to a shapefile of the African Continent took an average of 6.4 seconds

Additional Requirements

- Solution runs as expected after following instructions.
- All solutions must analyze geo-spatial data to provide business insights in a format that is useful to a non-data scientist.
- All solutions must scale to other data sets of similar size and complexity
- Solution does not already exist in open source analytics platforms, such as KNIME, R, or similar
- If developed in an existing open source analytics platform, the solution must also adhere to guidelines for tools or nodes on that specific platform. For example, a node developed in KNIME must adhere to [KNIME node submission guidelines](#).
- If selected as a winner, the solution must be made open source

Phase 2 Judging Criteria

Criteria	Description	Score
<b>Quantitative evaluation</b> will be performed using the criteria provided below. Your solution will be ranked out of all available solutions and your position in the ranking will determine your score. In order to be eligible for a prize, your solution must meet or exceed the baseline for speed, accuracy, and scalability as detailed in the Performance section above.		
Speed	<ul style="list-style-type: none"><li>• Speed will be ranked based on the speed of processing 10,000 records.</li><li>• In the event there are multiple submissions that are competitive at 10,000 records, solutions will be ranked based on the speed of processing 100,000 records.</li></ul>	20
Accuracy	<ul style="list-style-type: none"><li>• Measured against our baseline methods for acquiring this data, how well does this solution meet that baseline.</li></ul>	10
Scalability	<ul style="list-style-type: none"><li>• Scalability will be ranked based on the number of records the solution is able to process without crashing. The solution will be tested with 1,000 records, 100,000 records and up to 500,000 records.</li><li>• Can the solution be run with datasets other than just the sample?</li></ul>	10
<b>Qualitative evaluation</b> will be performed by the judging panel. Solutions will be scored based on their ability to meet or exceed the judging criteria		
Usefulness	<ul style="list-style-type: none"><li>• How useful is the business insight to the target audience?</li><li>• How well does the solution address the described problem?</li><li>• Does the solution generate an output that can be interpreted by someone without data science expertise (i.e., visual representation of output)?</li><li>• Does the solution solve for multiple types of problems (e.g., allow for calculations of point-point distance and point-polygon distance)? See “Challenge Goal” for all options</li></ul>	30
Innovativeness	<ul style="list-style-type: none"><li>• How innovative is the business insight?</li><li>• How original is the development?</li></ul>	15
Ease of Understanding	<ul style="list-style-type: none"><li>• How easy was the solution to understand and evaluate on a technical level, based on documentation provided?</li><li>• Does the solution run as expected headlessly after following instructions?</li><li>• Did the solution make use of an integration with KNIME to ease understanding and evaluation? Was the integration easy to use?</li><li>• If applicable, does the solution run as expected when integrated into KNIME?</li></ul>	10

KNIME Integration	<ul style="list-style-type: none"><li>Does the solution work in KNIME Analytics Platform?</li><li>Is the solution a KNIME workflow?</li><li>Is the solution a custom KNIME node?</li></ul>	5
-------------------	--	---

## Data

We have included to sets of sample data for use when testing your submission. These sets are for sampling and testing purposes, and other datasets are welcome and encouraged. If possible, we strongly recommend providing any additionally tested data sets with your solution.

The Africa Excel file dataset contains 150,000+ incidents of conflict that have a geolocation available as a longitude and latitude. We have also provided a ESRI Shape file format of the African continent which was sourced from <http://www.maplibrary.org/library/stacks/Africa/index.htm>. These sets of data have been provided to explore point-in-polygon, resource leveling (optimal distance to center of most conflicts), and other solutions.

- <https://assets.tma1.io/herox/Africa.zip>
- [https://assets.tma1.io/herox/Africa\\_Crime.xlsx](https://assets.tma1.io/herox/Africa_Crime.xlsx)
- <https://nbviewer.jupyter.org/github/tma1/herox/blob/master/notebook/Africa.ipynb>

The Detroit CSV file contains 130,000+ crimes committed in Detroit available with a geolocation available as longitude and latitude. These can be juxtaposed with the Detroit patrol areas shape files that have also been provided. Some of the solutions that can be explored in this dataset are point-in-polygon as well as various distances. These distances can be point to point or point to closest polygon edge. They can be computed using euclidean, geodesic, or network distance algorithms. (In this case, the network is Detroit streets, a dataset that has not been provided). All of the Detroit data was gathered from <https://data.detroitmi.gov/>.

- [https://assets.tma1.io/herox/DPD\\_Crime\\_Incidents.csv.zip](https://assets.tma1.io/herox/DPD_Crime_Incidents.csv.zip)
- [https://assets.tma1.io/herox/DPD\\_Scout\\_Car\\_Areas.zip](https://assets.tma1.io/herox/DPD_Scout_Car_Areas.zip)
- <https://nbviewer.jupyter.org/github/tma1/herox/blob/master/notebook/detroit.ipynb>

## Rules

### Participation Eligibility:

The challenge is open to all adult individuals, private teams, public teams, and collegiate teams. Teams may originate from any country. Submissions must be made in English. All challenge-related communication will be in English.

No specific qualifications or expertise in the field of GIS is required. Prize organizers *encourage outside individuals and non-expert teams* to compete and propose new solutions.

To be eligible to compete, you must comply with all the terms of the challenge as defined in the [Challenge-Specific Agreement](#).

### Registration and Submissions:

Submissions must be made online (only), via upload to the [HeroX.com](#) website, on or before the deadlines outlined in the [Timeline](#). Please see the submission form for any document upload format requirements. No late submissions will be accepted.

### Intellectual Property Rights:

If an innovator is awarded a prize, the Sponsor will require all content and assets submitted as part of a Finalist’s Submission to be released under open source licenses that permit free distribution, derivative works, and use in commercial and non-commercial settings. Please see the [Challenge-Specific Agreement](#) for complete details.

All Innovators are welcome and encouraged to depend on or make use of other components, libraries, content, assets, and code. All such materials must be available under any Open Source Initiative (OSI) or Creative Commons license compatible with the OSI or Creative Commons license under which the Submission will be released. “Compatible” means that each Innovator’s entire Submission must be usable without violating the license terms of those components licensed under the CC BY 4.0 license, Apache License 2.0, or respective OSI license for the components. Source code licensed under the LGPL, BSD, MIT, or Apache licenses currently meets this criterion; other open source licenses may also meet it. If Innovators make modifications to existing open source projects, they are strongly encouraged to submit patches upstream and work to have them accepted. Patches that are not accepted upstream may be submitted as part of the code developed by the Innovator, under the same Apache License 2.0. Content and assets must be licensed under terms that permit commercial usage. The Creative Commons CC BY and CC-BY-SA licenses currently meet this criterion. Innovators cannot submit entries that include or rely on software or content that is either closed-source, proprietary, illegally sourced, or depends on per-seat licensing.

### Selection of Winners:

Based on the winning criteria, prizes will be awarded per the Judging Criteria section above. In the case of a

**Additional Information**

- Void wherever restricted or prohibited by law.
- No purchase or payment of any kind is necessary to enter or win the competition.
- All ineligible applicants will be automatically removed from the competition with no recourse or reimbursement.
- All applications will go through a process of due diligence; any application found to be misrepresentative, plagiarized, or sharing an idea that is not their own will be automatically disqualified.
- By participating in the challenge, each competitor agrees to submit only their original idea. Any indication of "copying" amongst competitors is grounds for disqualification.

**Was this page helpful?**

- ☐ Yes      ☐ No