ly Connection

The Data Science Newsletter

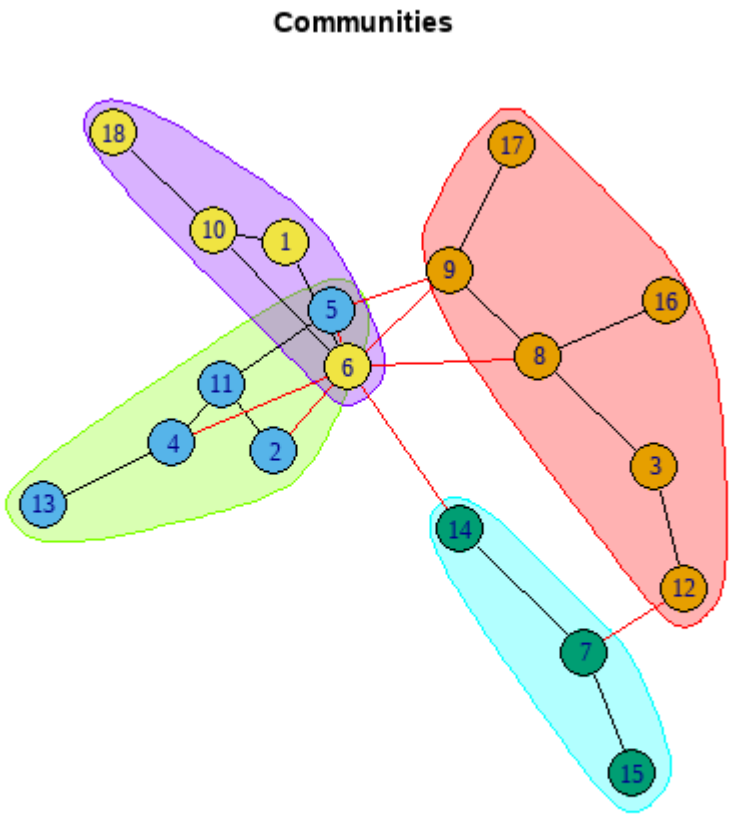
te on all things data science at MSI

PUBLISHED
January 31, 2024

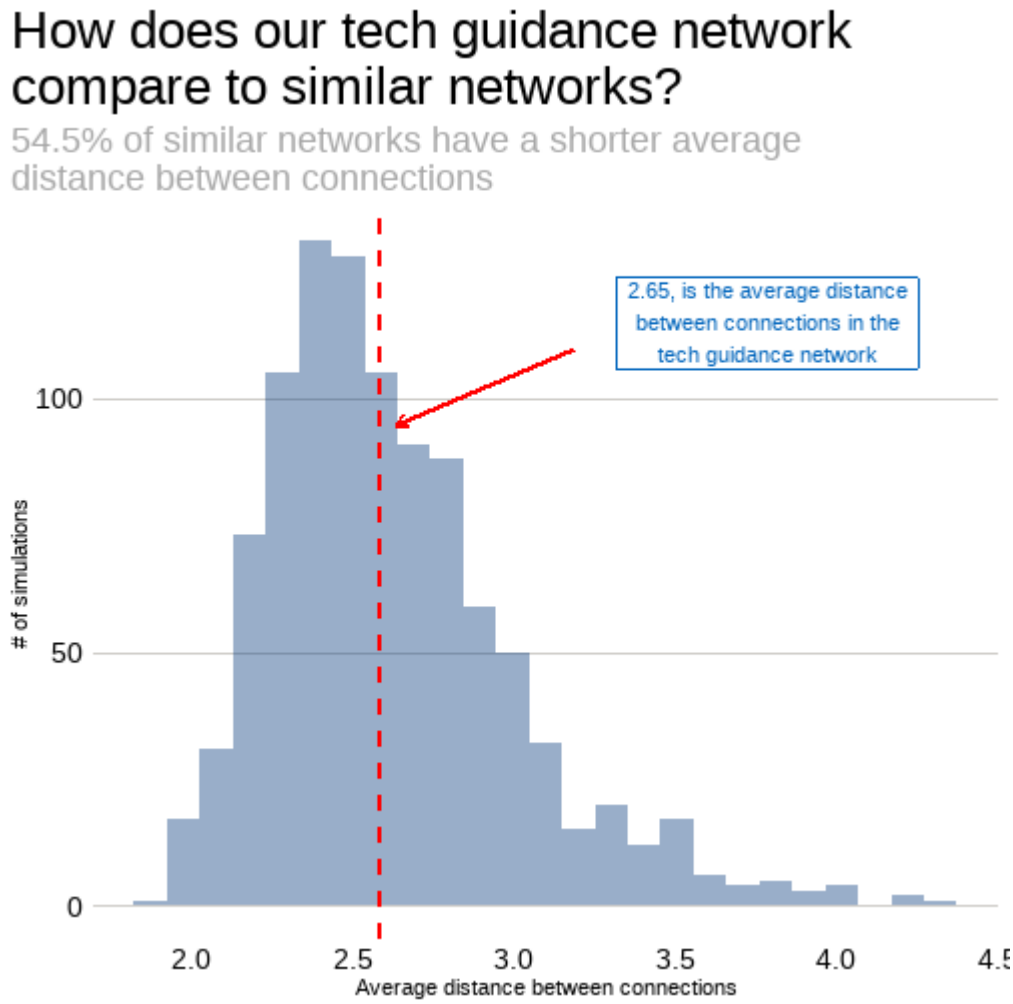
Happenings

Social Networking Analysis Brown Bag

On September 26, 2023 the Data Science Team hosted a brown bag on Social Network Analysis that was led by Brian Calhoon. Using data from our survey of DST members, Brian cleaned the data, transformed it into a graph object, and demonstrated how to calculate standard network statistics. In addition, he simulated similar networks in order to compare our actual network performance with what we should expect. Finally, he demonstrated how to use clustering algorithms to separate the network into communities.



Identifying communities in our network



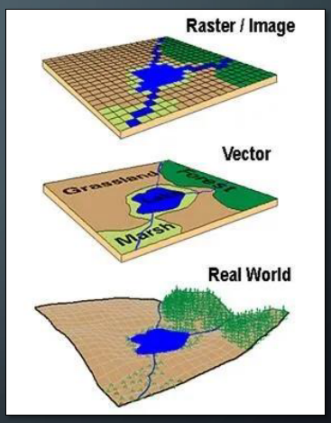
Leveraging simulations to understand network statistics

Mapping Brown Bag

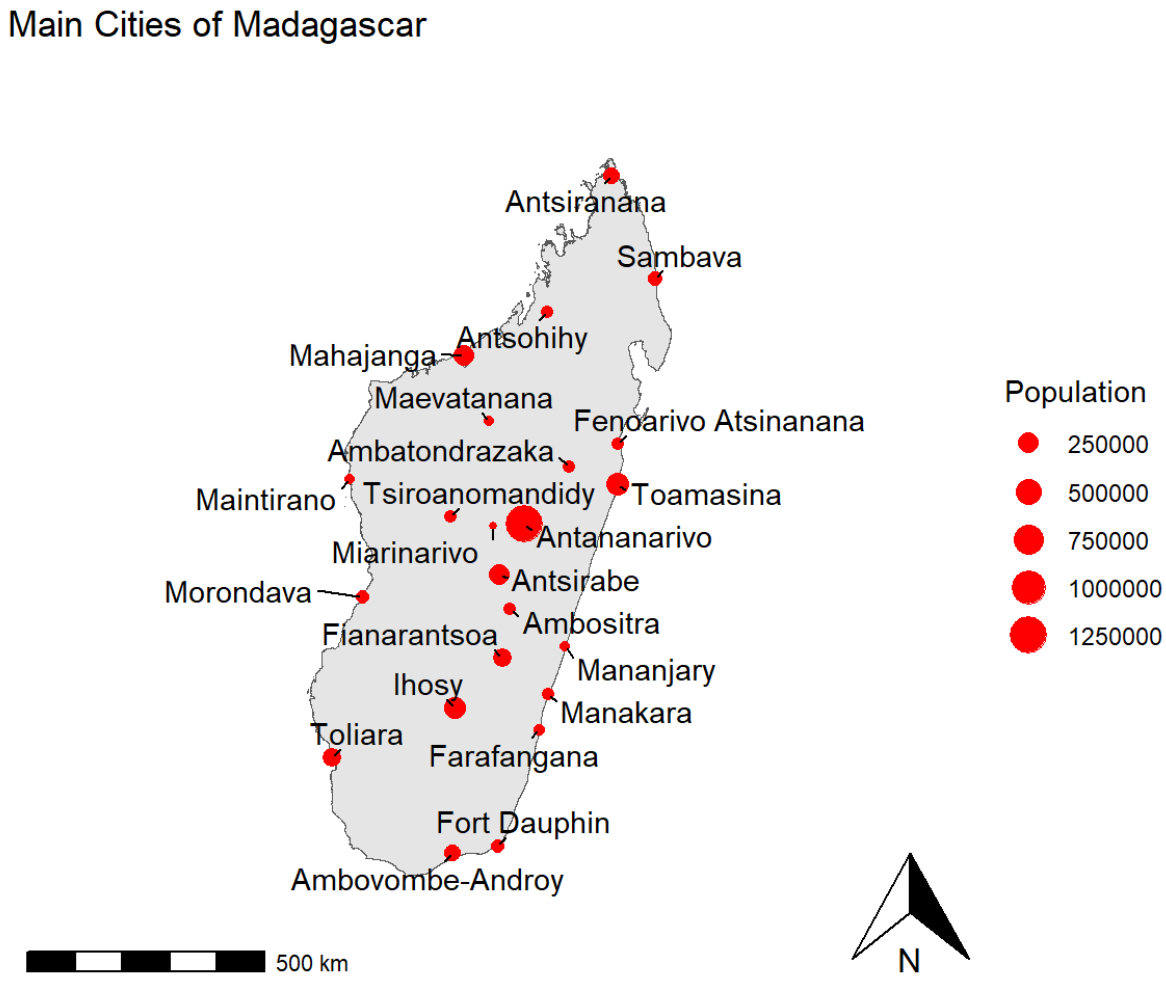
On November 16, 2024 the data science team hosted a brown bag to demonstrate basic mapping skills in QGIS and R. MSI's internal GIS Specialist, Jacob Milley led the session where he introduced GIS concepts, types of spatial data, and common data sources before demonstrating how to make a map in QGIS. Then, Brian Calhoon followed up by duplicating the same demonstration in R using the ggplot2 package.

TYPES OF GIS DATA

- Vector: *A representation of the world using points, lines, and polygons*
  - Points
  - Lines
  - Polygons
- Raster: *A representation of the world divided into a grid of cells.*
  - Aerial photographs
  - Satellite images



Understanding GIS data



A map of Madagascar

For recordings, presentations, and any associated background material on any DST event, please visit our Egnyte repository.

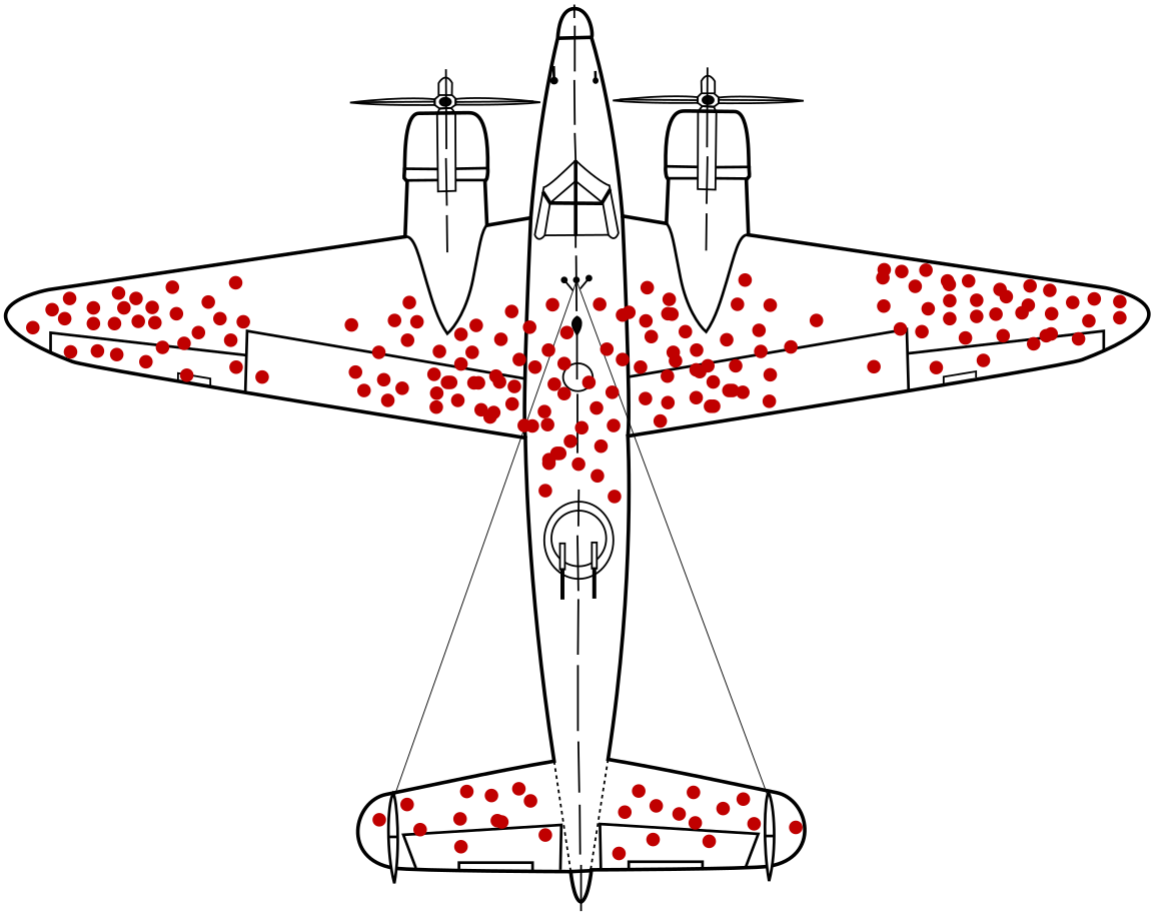
Stats corner

Today's puzzle

Situation:

You are an analyst for the United States Air Force during World War II. USAF has resources to add additional armor to its aircraft, and has asked for your advice as to where on the airplanes to add the extra protection. USAF has aggregated damage reports for all its fighters returning from active combat, and produced the following:

▶ Code



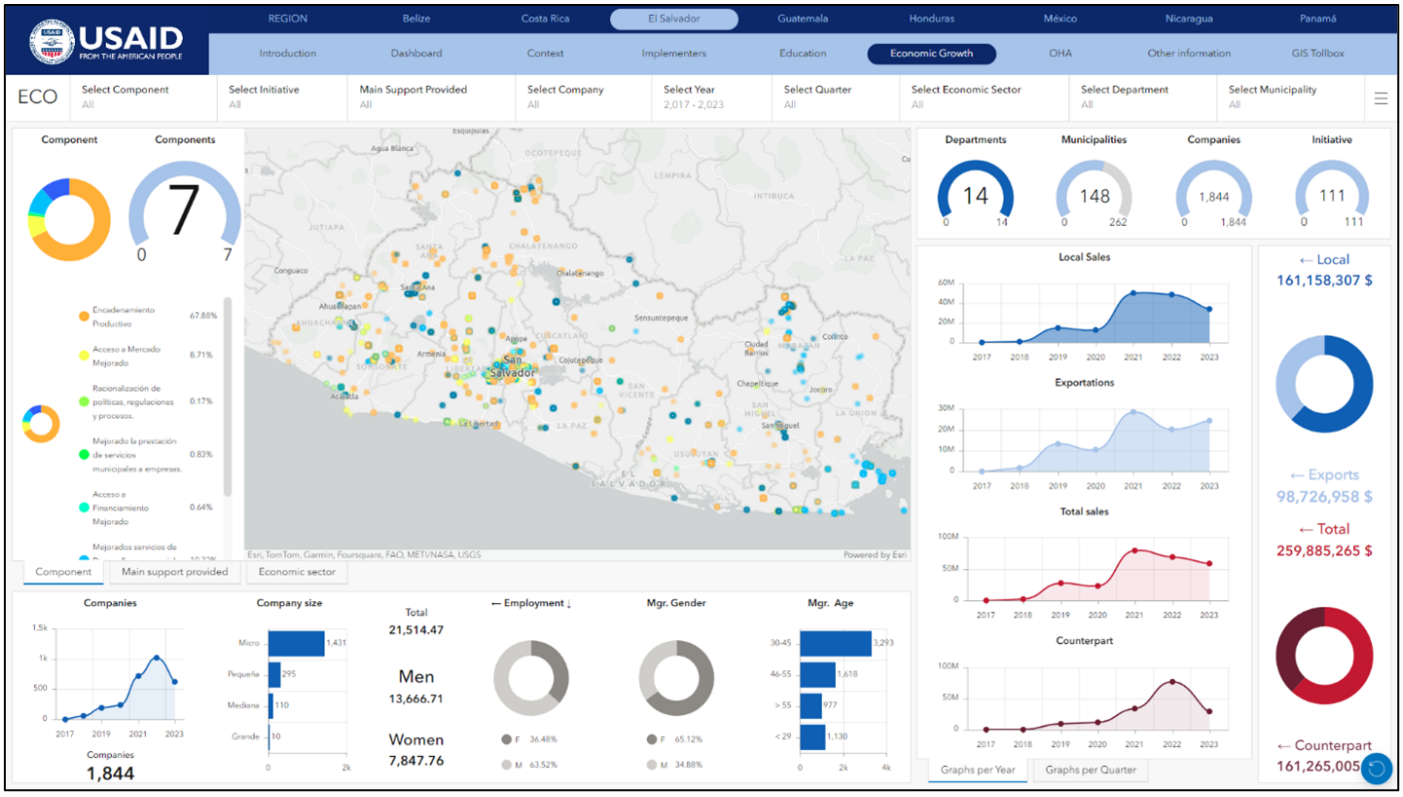
Your mission:

Based only on the information from the above figure, advise USAF on where it should add additional armor to its fighters.

$10 gift card to the first three responders with correct answers. Please send all responses to Dan Killian.

Member Spotlight

This month's GIS Spotlight is Joaquin Martin Lopez, our esteemed GIS Specialist on the EL Salvador MELS project who joined MSI in April 2023. As those of you who joined January's GIS COP meeting heard about, Joaquin has been working on a GIS "One Stop Shop" for the mission showcasing both contextual and project specific information for El Salvador as well as the surrounding Central American countries. The GIS "One Stop Shop" consists of a variety of interactive dashboards and story maps using ArcGIS online which allows users at USAID to dynamically view data submitted by implementing partners on themes such as education and economic growth. In addition, the Joaquin and the project will be conducting a mission-wide GIS assessment to gauge the current and future GIS capacity. The hope is that El Salvador MEL's usage of GIS will enable the mission and IPs to make informed decisions as to where to focus development work going forward. If you have any questions or are interested in learning a bit more about his GIS work for El Salvador MELS, feel free to contact him directly via email or Teams: Joaquin Martin-Lopez.



Recap: Quarterly Meeting

Our latest DST quarterly meeting was held on December 20th, 2023 with approximately 20 colleagues across both home and field offices.

We kicked off with a brief overview of our recent social network analysis and mapping brown bags. The team also highlighted contributions to a pivotal study by the Center for Global Development, focusing on early-grade reading assessment (EGRA) programs.

A significant portion of the quarterly meeting was dedicated to interactive breakout sessions. These sessions provided a platform for deep-diving into a variety of sampling scenarios. We explored how to effectively address and mitigate potential validity threats in these scenarios. Key discussions included:

- The influence of human choice on random sampling, emphasizing the need for enumerators to recognize and account for these characteristics.
- In Third-Party Monitoring (TPM), the importance of being cognizant of biases, especially those related to the locations selected for random sampling.
- Challenges and considerations in applying weights to ensure representativeness within the zone of influence.
- Addressing the issue of survivorship bias, which can skew data by overlooking non-representative elements.

These discussions set the stage for our upcoming brown bag in February 2024, where we will delve into sampling best practices and techniques, informed by our recent scenario planning exercises.

Recap: GIS COP January Meeting

Our in-house GIS Specialist Jacob Milley hosted the second GIS COP on January 23 with around 20 colleagues from varying projects/countries in attendance. The meeting included a recap of the first COP meeting, a presentation by Joaquin Martin Lopez on his GIS work for the El Salvador MEL project using ArcGIS Online, and finally a brief discussion on commonly used GIS data sources. The goal of this discussion was for the GIS COP to establish an authoritative a list of freely accessible GIS data sources to help other GIS users at MSI get their hands on more available data!

Upcoming events/opportunities

- Brown bag on sampling in mid-February
- Brown bag on Power BI Part 3 in March
- Brown bag on data collection instrument design and scripting in May

About the Data Science Team

Overview

The Data Science Team (DST) is a collection of colleagues at MSI that are interested in building and practicing data analytics skills, from the basics of descriptive analytics to advanced machine learning techniques. The core members of the Data Science Team are Dan Killian, Brian Calhoon, Melanie

Murphy, Tim Reilly, Tim Shifflett, Gunjan Maheshwari, Jack Payne, and Jacob Milley. Each member brings a particular expertise to ensure that the team's initiatives are in line with MSI's commitment to incorporate advanced data analytics into our work. The core team supports MSI staff by:

- Supporting staff conducting quantitative and qualitative analysis with guidance on best practice and available data sources.
- Helping staff learn new data science techniques and applying the techniques to existing activities.
- Providing trainings on the application of analyses across multiple analytical packages, as well as on core statistical concepts that underlie the analyses.
- Holding structured trainings on-demand, at least once per year and 4-12 brown bag sessions per year.
- Holding weekly office hours and on-demand consultation.
- Providing a knowledge management function for past and current data science work at MSI while ensuring everyone knows about these resources.

For support or to ask questions, reach out to us on our Teams Channel.