

August 1-3, Foundation University, Dumaguete



PISTA NG MAPA

open data & software • free maps • community • talks • workshops



KAART



mapbox



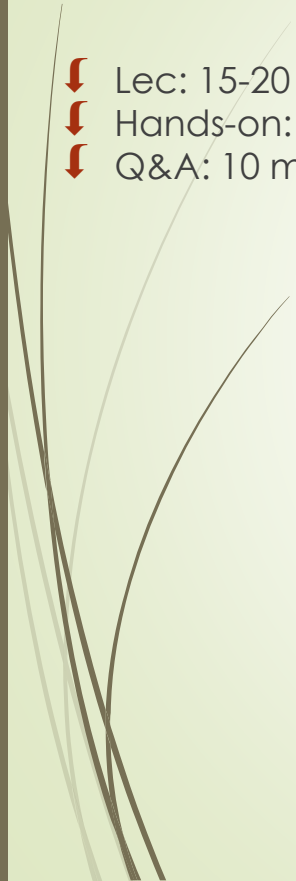
Mapillary UΔV4GEO



Geo* Data Science w/ Python

Ian Panganiban

Plan

- ↓ Lec: 15-20 mins
 - ↓ Hands-on: 50-60 mins
 - ↓ Q&A: 10 mins
- 

Hello, I'm Ian

Prior Engagements with



Currently working



*Head of Cloud Infrastructure

Twitter: @lkpanganiban
Email: lkp@noypimaps.com

hello from
the humans of



iraya

machine learning • geoscience



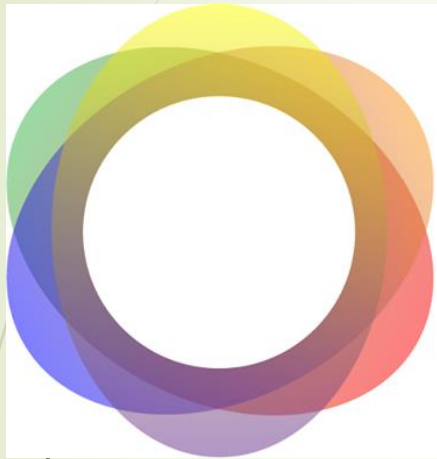
www.irayaenergies.com



info@irayaenergies.com



[/irayaenergies](https://www.linkedin.com/company/irayaenergies)



iraya

machine learning • geoscience

We're Hiring!!!!!!

Outline

- ↓ Geo* Data Science (GDS)?
- ↓ GDS Workflow
- ↓ Python in GDS
- ↓ Toolsets/Libraries
- ↓ Wrap Up
- ↓ Q & A



Energy



Agriculture



Transportation



Railroad



Construction



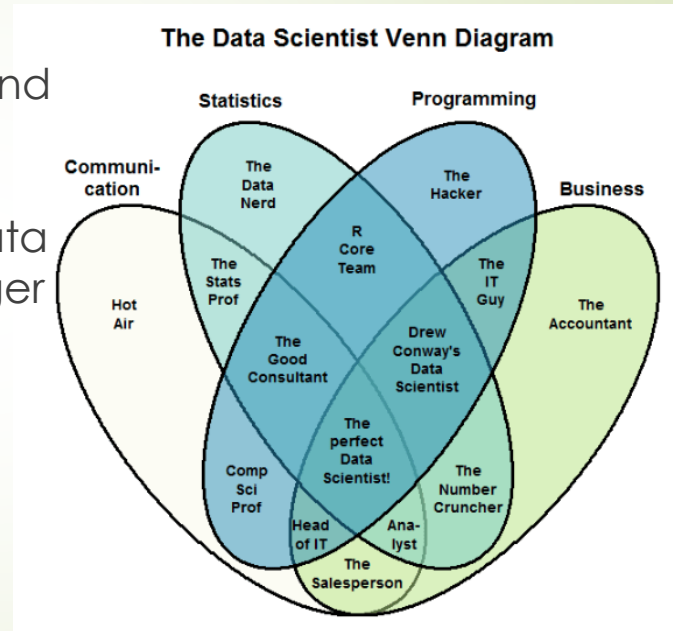
Industrial

What is Geo* Data Science?

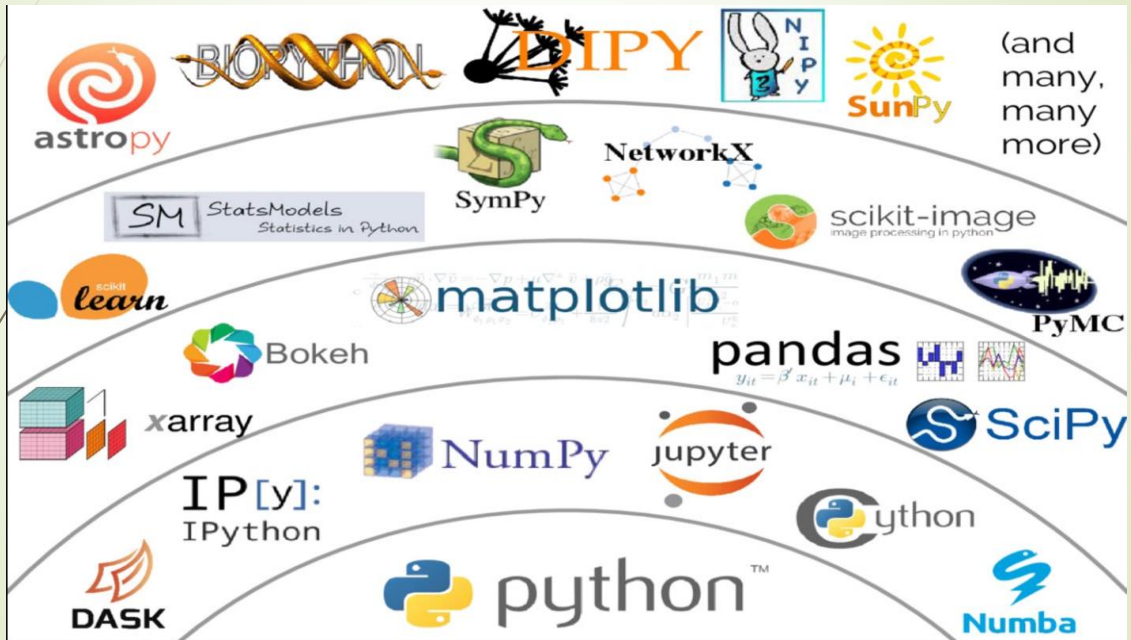
- Geospatial
- Geoscience
- Geodata
- Geology
- Geography

What is the role of a GDS?

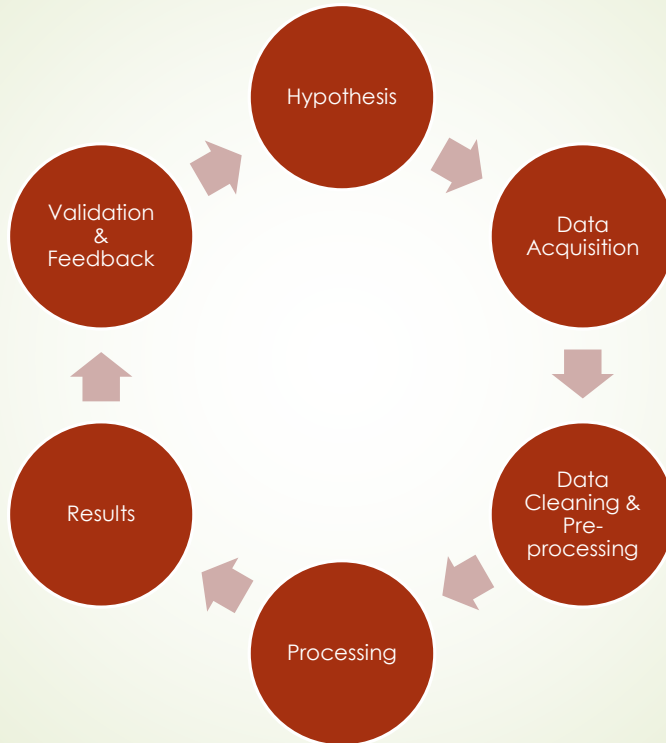
- ↓ Sometimes creating models and visualization
- ↓ Sometimes running algos
- ↓ Most of the time you clean data aka data janitor & data munger



Why Python in GDS?



GDS Workflow



GDS Workflow - Hypothesis

- ↓ What are the problems that you are trying to solve?
- ↓ Who are your stakeholders/users?
- ↓ What is the end product?
- ↓ What is the value of my solution?

GDS Workflow - Data Acquisition

- ↓ What type of data should I acquire?
- ↓ What type of data do I have?
- ↓ Where is my data source? (Database, API, File, Reports, Drones, Survey, Public, Private)
- ↓ What are the limitations of my data? (spatial resolution, time, sensors, licensing)

GDS Workflow - Data Cleaning/Conditioning

- ↓ How should I handle null values? (0, N/A, None, Null)
- ↓ Do I need to apply corrections? (sensor, environmental)
- ↓ What other pre-processing steps? (rectification, georeference, coordinate consistency, file naming)
- ↓ This takes around 70% of your time.

GDS Workflow - Processing

- ↓ What type of algorithm should I use?
- ↓ Do I need to scale my solution?
- ↓ Can this be automated?
- ↓ Do I have enough data?

GDS Workflow - Results

- ↓ How should I visualize my results? (Maps, Charts, Dashboard)
- ↓ What are the metrics should I use? (F1-score, precision, std dev, performance benchmarks/runtime)
- ↓ How should I package my results? (Web app, model, algorithm, workflow)
- ↓ How should I communicate my results?

GDS Workflow - Validation and Feedback

- ↓ How should I test my results in the real world?
- ↓ How do I get user/stakeholder feedback?
- ↓ How do I monitor my solution? (performance, accuracy)
- ↓ Is my solution being used?



Hands-on

<http://52.187.204.123>



Google Cloud Platform





Q & A

References

↓ Awesome Lists

↓ Awesome GIS

↓ Awesome Data Science

↓ Awesome Python

↓ Awesome Geospatial

↓ Repository:

https://github.com/lkpanganiban/pistangmapa_geopython

Final Note

“The most important part of being a data scientist in general is your skill in identifying and solving problems”

“A great calligrapher doesn't blame the brush”

"Regardless of what we discover, we understand and truly believe that everyone did the best job they could, given what they knew at the time, their skills and abilities, the resources available, and the situation at hand."
- Norman Kerth

August 1-3, Foundation University, Dumaguete



PISTA NG MAPA

open data & software • free maps • community • talks • workshops



KAART



mapbox



Mapillary UΔV4GEO