ACM Summer School @ IIT Gandhinagar

Machine Learning for the Planet

9 June 2025

Chocolates for answering my questions

OR

Asking good questions

Agenda

01 Energy Access & Emissions

02 Energy Markets

03 Power plants

04 Problem Statement

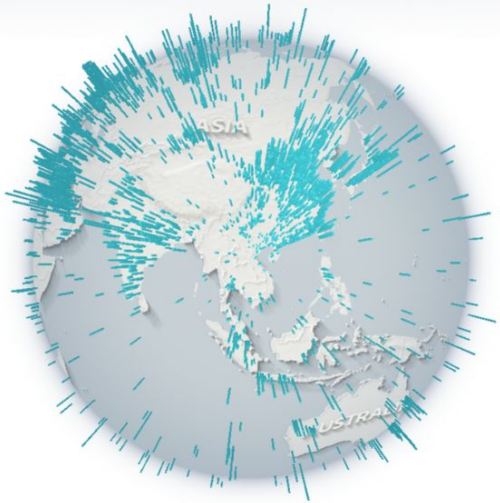
05 Climate Trace outcomes & impact

06 Datasets and modeling

CLIMATE TRACE

CLIMATE TRACE

INDEPENDENT GREENHOUSE
GAS EMISSIONS TRACKING



Self-reported emissions data and fossil fuel power plant generation varies by country in **detail, recency and spatiotemporal resolution**

AI Impact Challenge & Google.org Fellowship (WattTime)

AI Impact Challenge

- **What:** Global open call for organizations to submit ideas for how to use AI to address global problems
- **Awards:** $25M total funding from Google.org spread across 20 organizations
- **Goals**: Inspire and empower organizations to use AI to address global problems; show that AI is for everyone; demonstrate Google's commitment to and leadership in AI for good space

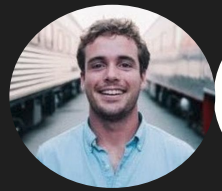
Google.org Fellowship with WattTime

- The Fellowship is Google.org's flagship employee engagement offering, enabling Googlers to complete up to six months of full-time pro bono work to accelerate the social impact of Google.org's top partners
- 11 total Fellowship programs, 6 as part of the AI Impact Challenge, 7 Googlers assigned to WattTime

Natalie Vais
Product Manager

Daniel Tyrrell
SWE

Wanda Czerwinski
SWE



Chris Dowd
PgM

Alok Talekar
SWE

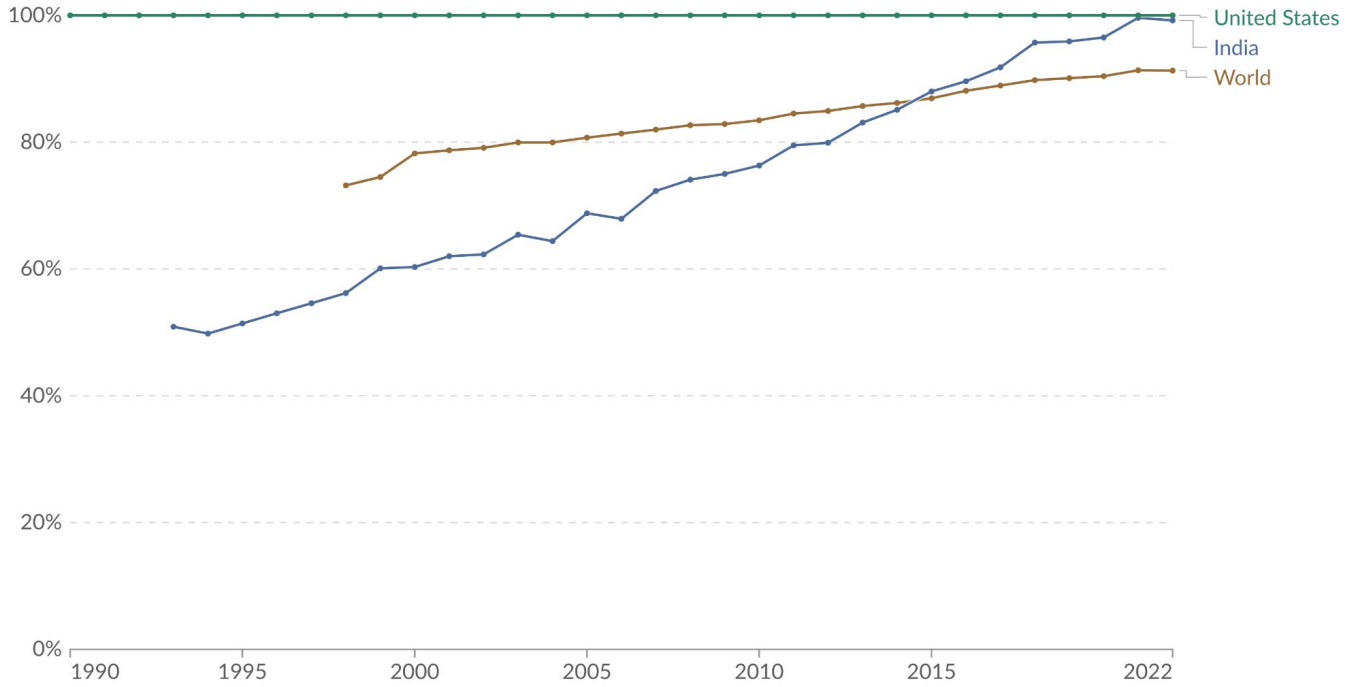
Kathi Kitner
UX Researcher

Brian Goldman
SWE

Share of the population with access to electricity

Our World in Data

Having access to electricity is defined in international statistics as having an electricity source that can provide very basic lighting, and charge a phone or power a radio for 4 hours per day.



Data source: Data compiled from multiple sources by World Bank

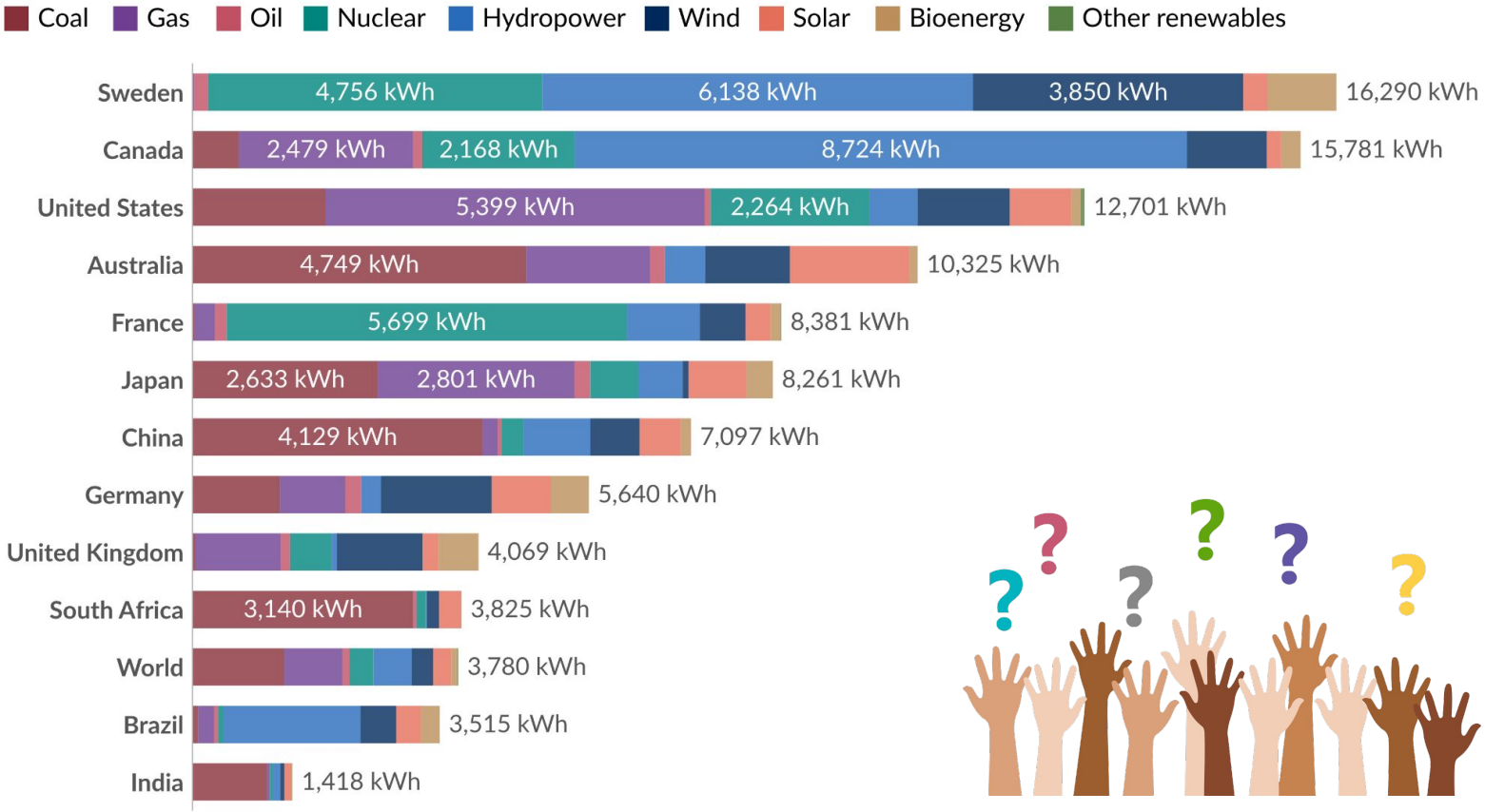
OurWorldinData.org/energy | CC BY

Note: Electricity access defined as having a source that can provide very basic lighting, and charge a phone or power a radio for 4 hours per day.

Per capita electricity generation by source, 2024

Our World in Data

Measured in kilowatt-hours[1]. "Other renewables" include geothermal, tidal, and wave generation.

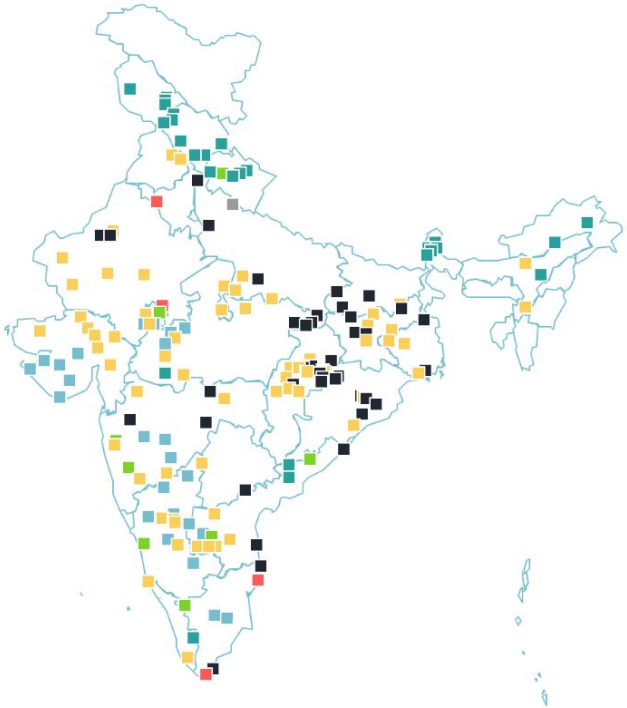


India Power Map - Pipeline

Operational | Pipeline | Retired

All India

- Select All
- Coal
- Oil & Gas
- Nuclear
- Hydro
- Wind
- Solar
- Bio Power
- Small-Hydro
- Pumped Storage



Capacity (MW)

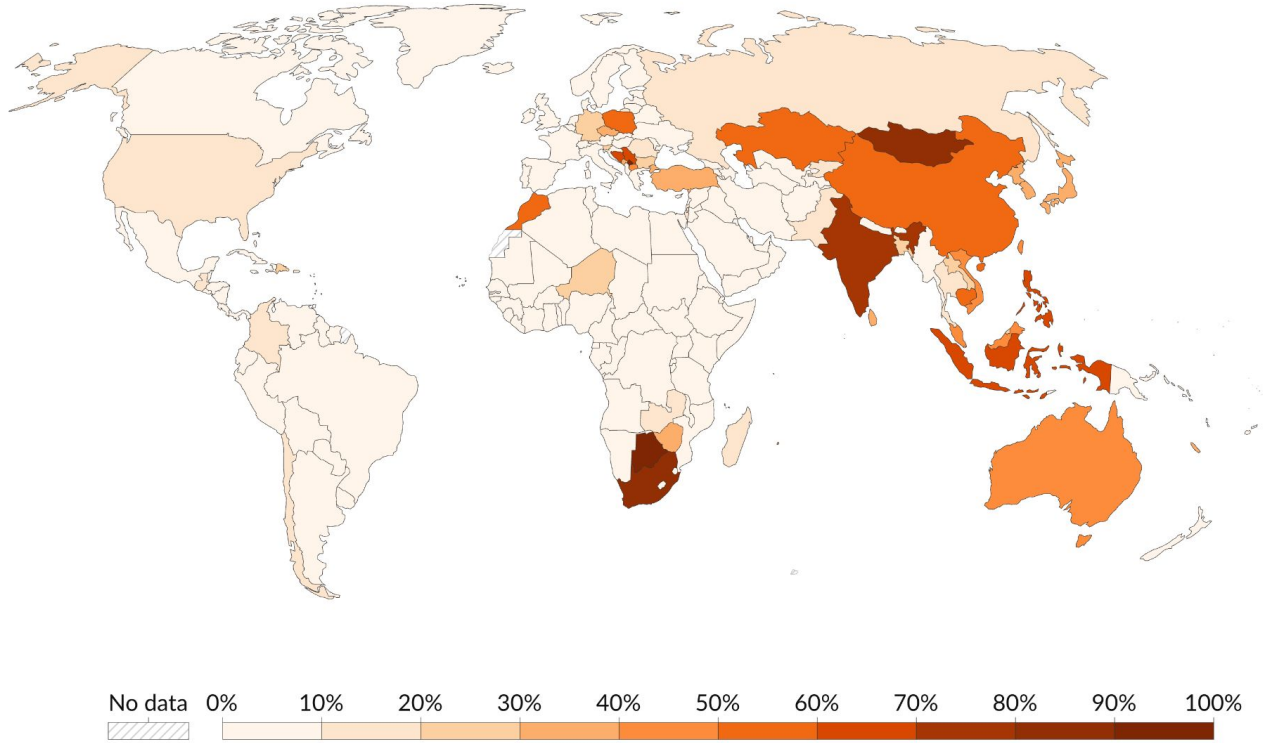
192,650.58

- Operational
- Pipeline
- Retired
- ISGS / RoofTop / Ground Mounted

Share of electricity production from coal, 2024

Our World in Data

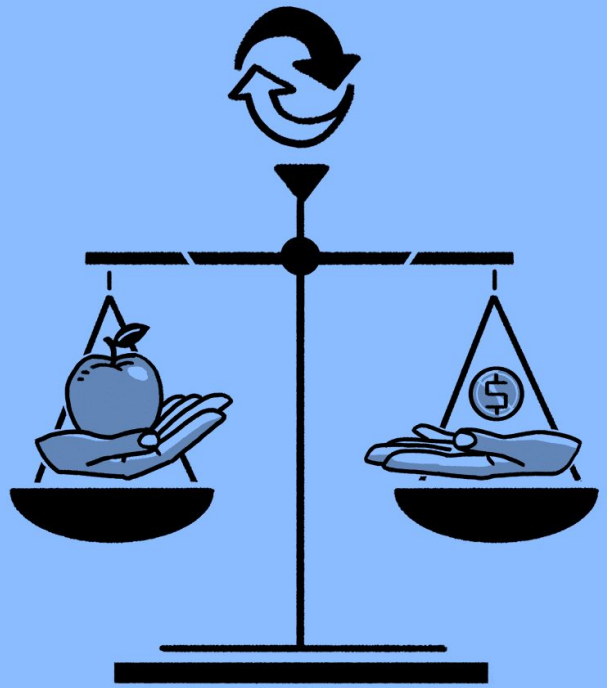
Measured as a percentage of total electricity produced in the country or region.



Data source: Ember (2025); Energy Institute - Statistical Review of World Energy (2024)

OurWorldinData.org/energy | CC BY

What exactly happens when you switch on



Law of Supply and Demand

['lo əv sə-'plī ən(d) di-'mand]

A theory that explains the interaction between the sellers of a resource and the buyers for that resource.

Investopedia

Forecast Demand

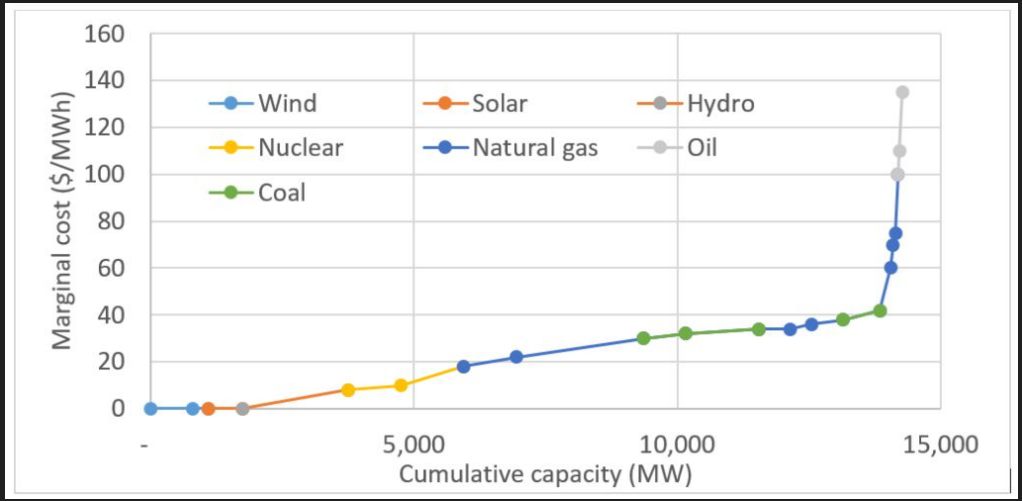
Forecast Supply

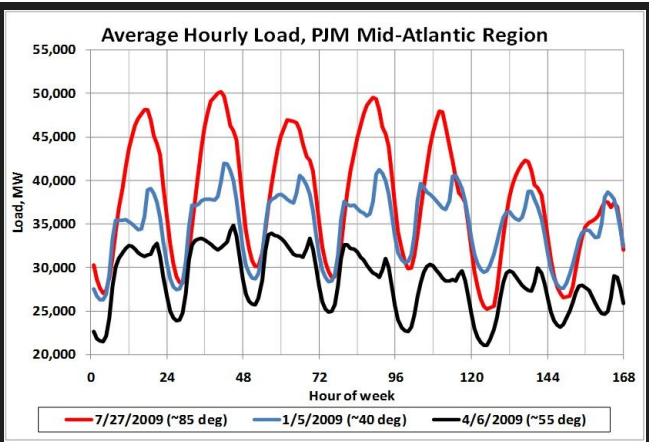
It’s a “Market”, stupid!

Price is governed by the highest bid power plant that needs to be run

Designed to give signals

- When each plant should run
- When new plants should be built, and what kind
- Real-time or day-ahead market with auctions to min cost.
- So economic theory gives a lot of information on how plants *should* operate - but not always true in practice



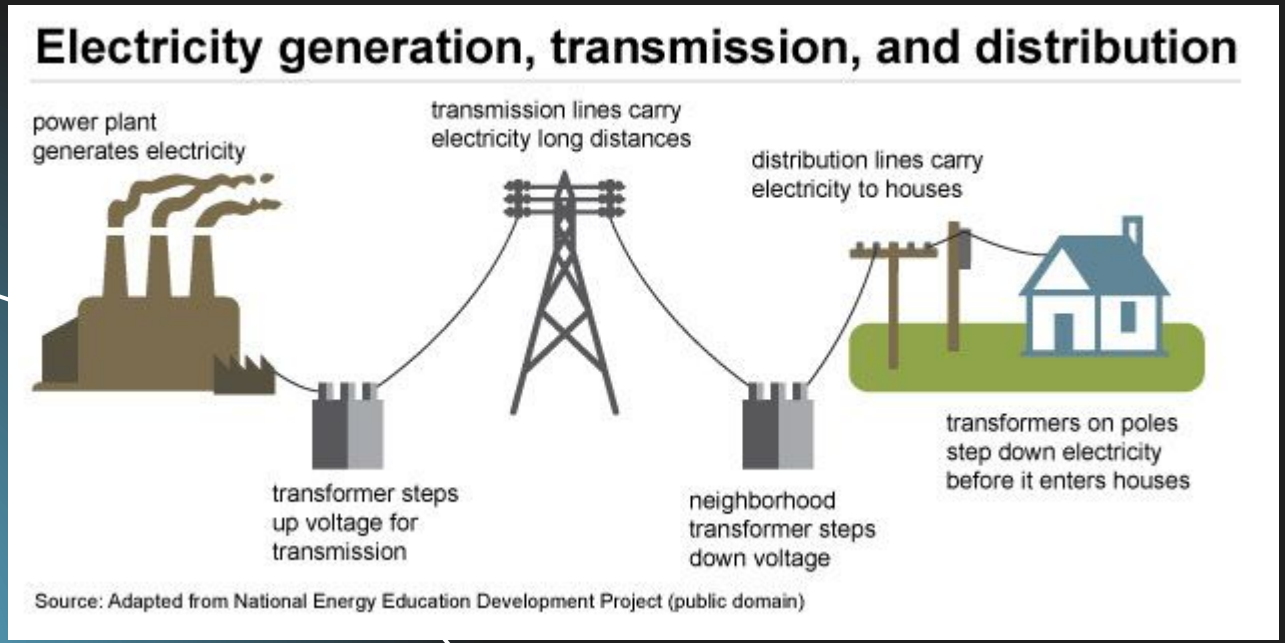


Energy Sector

GenCo

TransCo

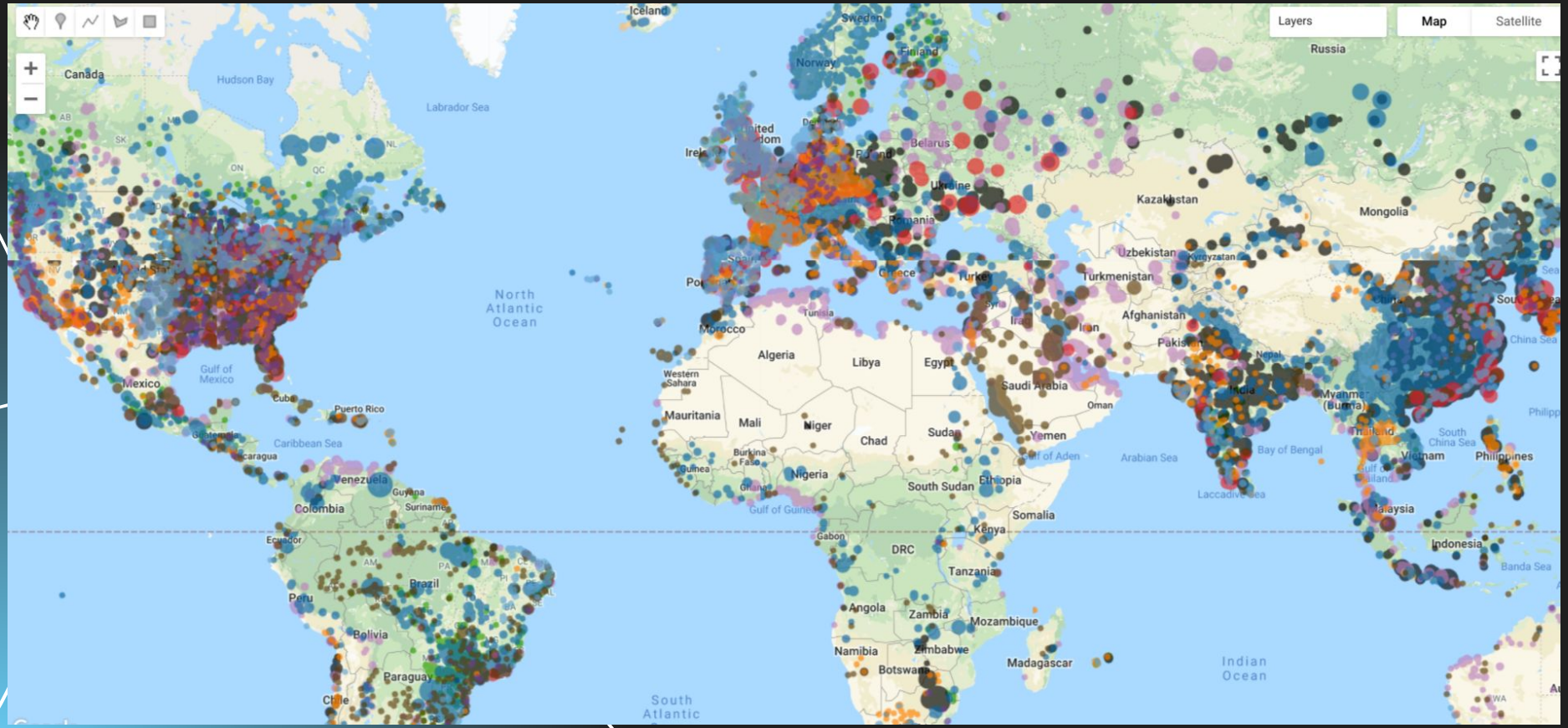
DisCom



Energy Sector Emissions

GenCo

Global power plants



https://code.earthengine.google.com/9efbd726e4a8ba9b8b56ba94f1267678

Anatomy of a thermal Power Plant

Multiple units per Power plant.

Mix of coal, oil/gas units is common.

Broad Goal

Use Machine learning and satellite imagery to predict emissions and energy generation from coal power plants

Capacity Factor

Capacity Factor (CF) =
(Actual Energy Output) / (Max Possible Energy Output)

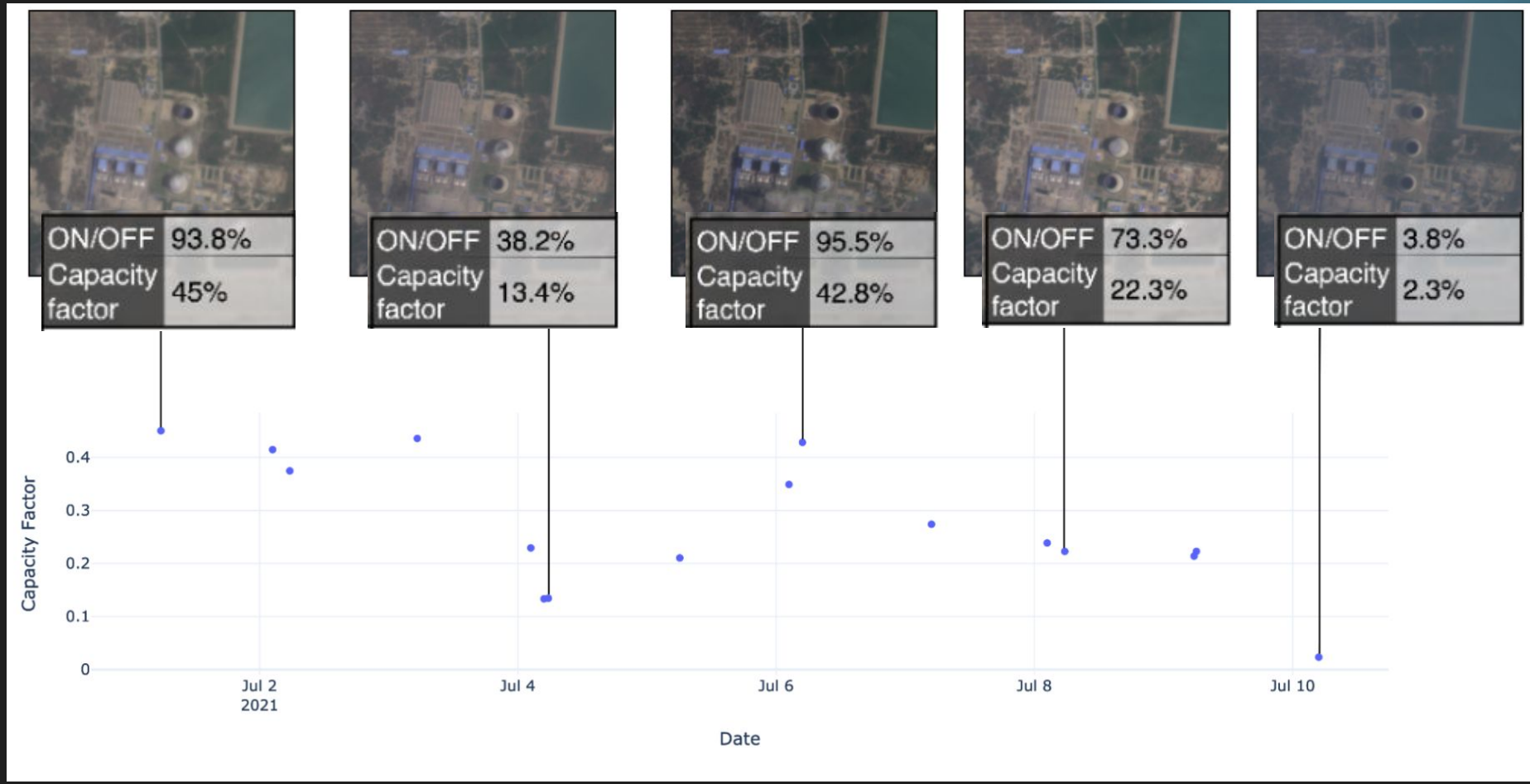
Coal plants - Base Load - high ramp up/ ramp down cost.

Measuring actual emissions is limited by temporal frequency of satellite imagery

- Lack of diurnal variation due to sun synchronous orbit
- Public satellite imagery at weekly cadence. Private daily.

Specific Problem Statement

Use Machine learning and satellite imagery to predict capacity factor of coal power plants by detecting their on/off state



Talwandi Sabo Power Project, India
Activity estimated using PlantetScope Imagery with CNN model

Plants using NDT produce visible vapor plumes during cooling.

FGD removes sulfur dioxide from gases released at coal-fired power plants, wherein flue gases become saturated with water, producing visible plumes.

A power plant may have one, both, or neither of these technologies.

Models can Evaluate the Impact of Policies/Events on Emissions

Climate TRACE models emissions estimates track with local events

Datasets

1. **Global power plant name and location**
2. **EPA hourly emissions data from US power plants (Q1 2024)**
3. **Satellite imagery for Q1 2024 at the power plant locations**

Questions 1 - Feel the data

1. **Which countries have more than 50 coal power plants?**
2. **Which US coal power plant has the cleanest and most polluting power generation in terms of CO2 emissions per megawatt ?**
3. **What is the average operating time for US coal power plants? Does it vary by state?**
4. **Is there a correlation between the commissioning year of a power plant and its CO2 emission rate?**
5. **What are the top 5 states in the US with the highest total CO2 emissions from coal power plants?**

Questions 2 - Feel the data

1. **How do the SO2 and NOx emissions compare between the cleanest and most polluting CO2-emitting plants?**

2. **Can you visualize the locations of all coal power plants in India on a map?**

3. **What is the distribution of power plant capacities (in MW) for different primary fuel types in India?**

4. **Which is the largest power plant in India ?**

5. **Which state in India has the most power generation capacity ?**

Data Prep

For simplicity, let's work with only 2 power plants to begin with. We can expand once setup is complete

```
us_training_plants = {
   'Barry': {'lat': 31.060, 'lon': -88.020},
   'Gorgas': {'lat': 33.644, 'lon': -87.196}
}

indian_target_plants = {
   'Mundra': {'lat': 22.823, 'lon': 69.553},
   'Sabarmati (Ahmedabad)': {'lat': 23.070, 'lon': 72.594}
}
```

Modeling tasks

1. **Model 1: Predicting Hourly Emissions (Regression)**

   Let's build a simple linear regression model.
   Features (X): Our satellite-derived feature.
   Target (y): The hourly CO2 emissions from the EPA data.

2. **Predicting Power Plant On/Off State (Classification)**

   This is a simpler but very useful task. Can we tell if a power plant is running just by looking at it from space?

   Features (X): Satellite feature.

   Target (y): A binary on/off status (1 for on, 0 for off). We'll create this from the EPA's 'Operating Time' column.

Thank you!