Eruptions by Proxy
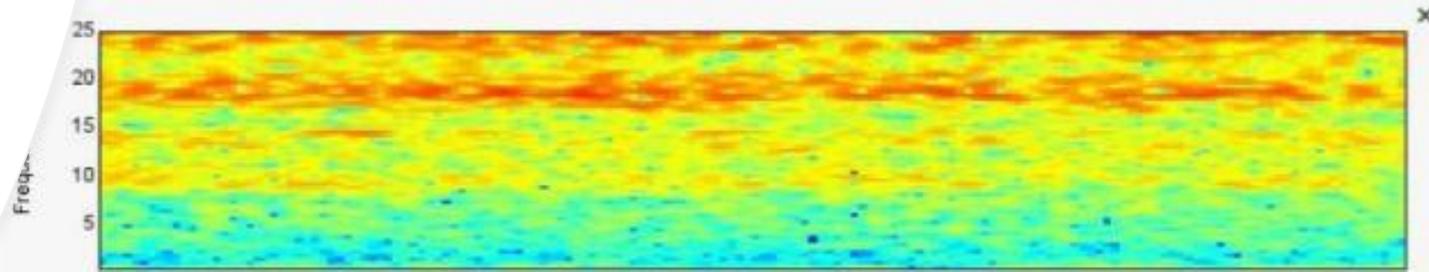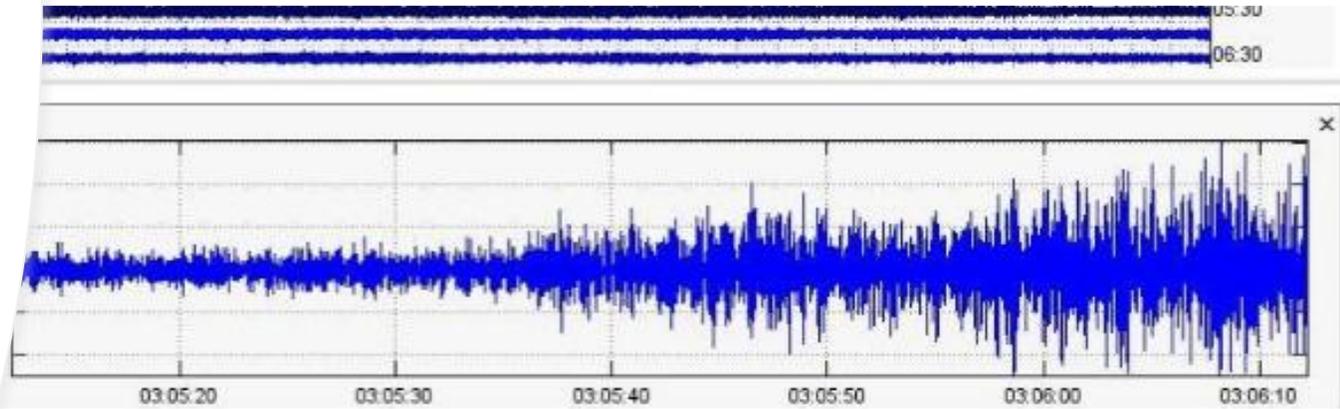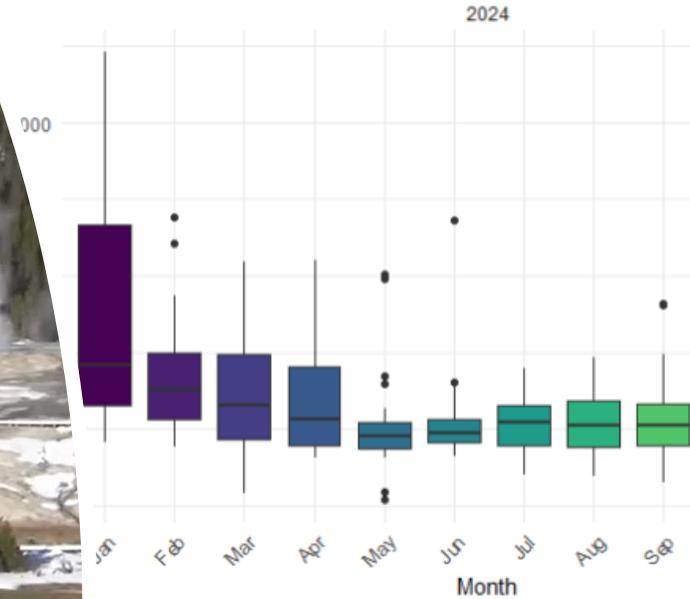
Remote Geyser Gazing and Data Interpretation

By Korben Cooper

Who am I?

- Geyser Gazing Regularly Since 2019

- Petroleum Engineering and Data Science

- Old Faithful Webcam Volunteer

- Research Impact of Climate on Subsurface Fluid Reservoirs

Presentation Outline

1. Remote Geyser Gazing
   - Old Faithful Webcam
   - Seismographs
   - Geysertimes
   - Other Loggers

2. Data and Geysers
   - Data Science Workflow
   - Issues with Geyser Data
   - Case Studies (Beehive and Strokkur)
   - AI and Geyser Data

```python
# --- Fetch Beehive eruption data ---
url = f"https://www.geysertimes.org/api/v5/entries/{one_year_ago_epoch}/{now_epoch}/{beehive_id}?iso=1"
response = requests.get(url)
data = response.json()


all_intervals, all_interval_times, all_interval_labels = [], [], []
if data["status"] == "success":
    eruptions = sorted(data["entries"], key=lambda x: x["time"])
    times = [datetime.fromisoformat(e["time"].replace("Z", "+00:00")) for e in eruptions]
    for i in range(1, len(times)):
        interval_hours = (times[i] - times[i - 1]).total_seconds() / 3600
        if interval_hours > 1:
            all_intervals.append(interval_hours)
            all_interval_times.append(times[i])
            all_interval_labels.append(times[i].strftime("%Y-%m"))
else:
    print("Failed to retrieve interval data:", data.get("message", "Unknown error"))


# --- Fetch durations ---
durations_by_time = []
if data["status"] == "success":
    for entry in data["entries"]:
        dur_sec = entry.get("durationSec")
        time_str = entry.get("time")
        res = entry.get("durationRes")
        if dur_sec is not None and time_str is not None and res != 1:
            try:
                dur_sec = int(dur_sec)
                if dur_sec > 0:
                    timestamp = datetime.fromisoformat(time_str.replace("Z", "+00:00"))
                    durations_by_time.append((timestamp, dur_sec / 60))
            except ValueError:
                continue
```

Webcam Resources

- **Old Faithful Webcam**
  - NPS Page
  https://www.nps.gov/yell/learn/photosmultimedia/webcams.htm
  - Geysertimes Chatpage
  https://chat.geysertimes.org/#
  - 12hr Webcam Loop
  https://www.youtube.com/watch?v=f7NDLBzILto

- **Biscuit Basin Webcam**
  - USGS Link
  https://www.usgs.gov/media/webcams/biscuit-basin-webcam
  - Joel's Webcam Loop
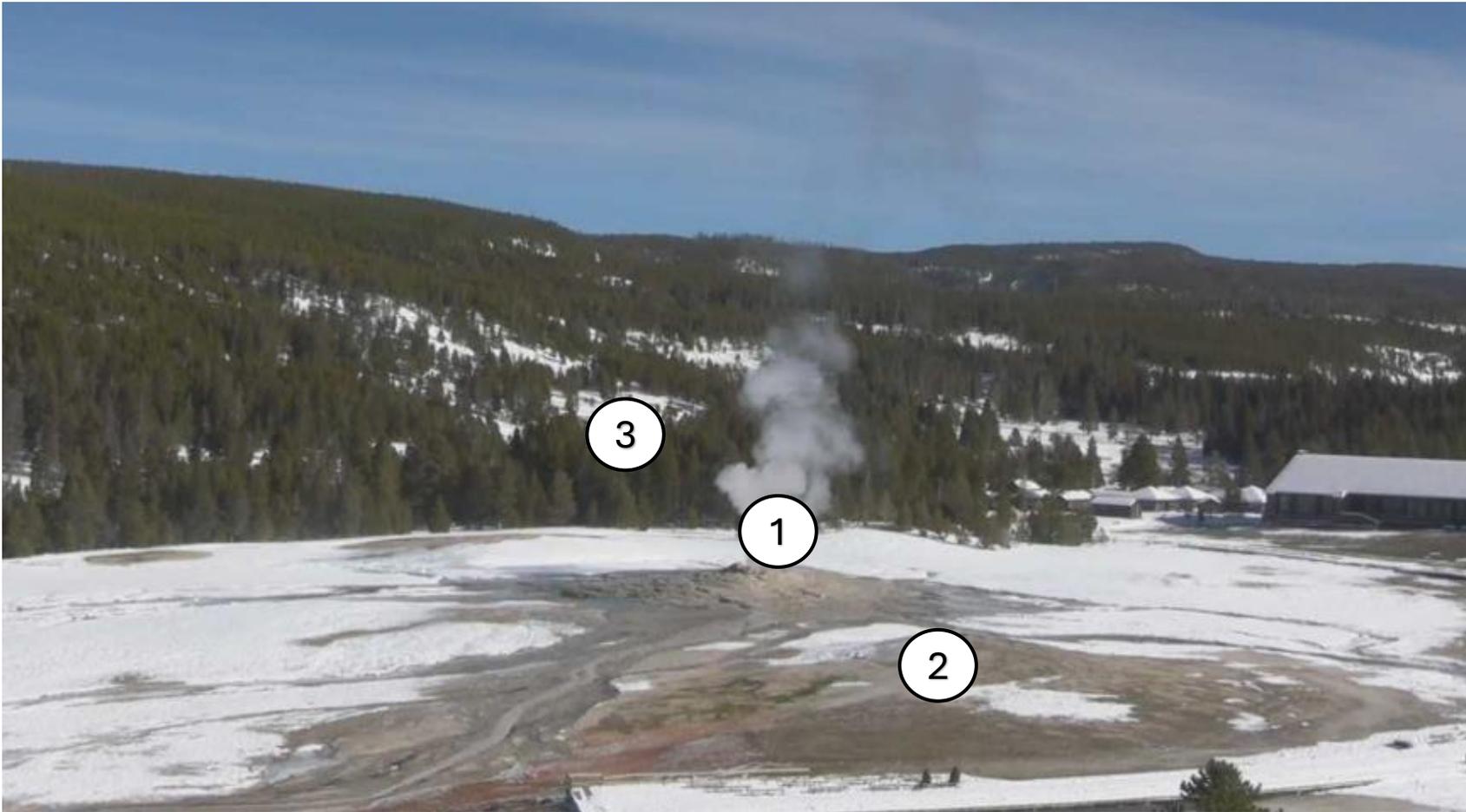 https://www.cs.drexel.edu/~jrh436/blackdiamond/

Old Faithful Webcam

- Located on top of the Old Faithful Inn

- Required to be on Old Faithful during Old Faithful's Window

- Moves elsewhere at Cam Operator discretion

Scope of Use: Eruption Times. Observing Behavior on Select Features.

Old Faithful: Feature Identification

1. Old Faithful
2. Split Cone
3. Dilapidated

Old Faithful: Observational Advice

Feature	Visible?	Start Time?	End Time?	Additional Details	Comments
Old Faithful	Yes	Exact	+/-5sec	Short vs. Long	
Split Cone	Yes	Ie	Not Possible		
Dilapidated	Behind Trees	Ns	+/-1min		

Geyser Hill: Feature Identification

1. Beehive
2. Depression
3. Lion
4. Aurum
5. Plate
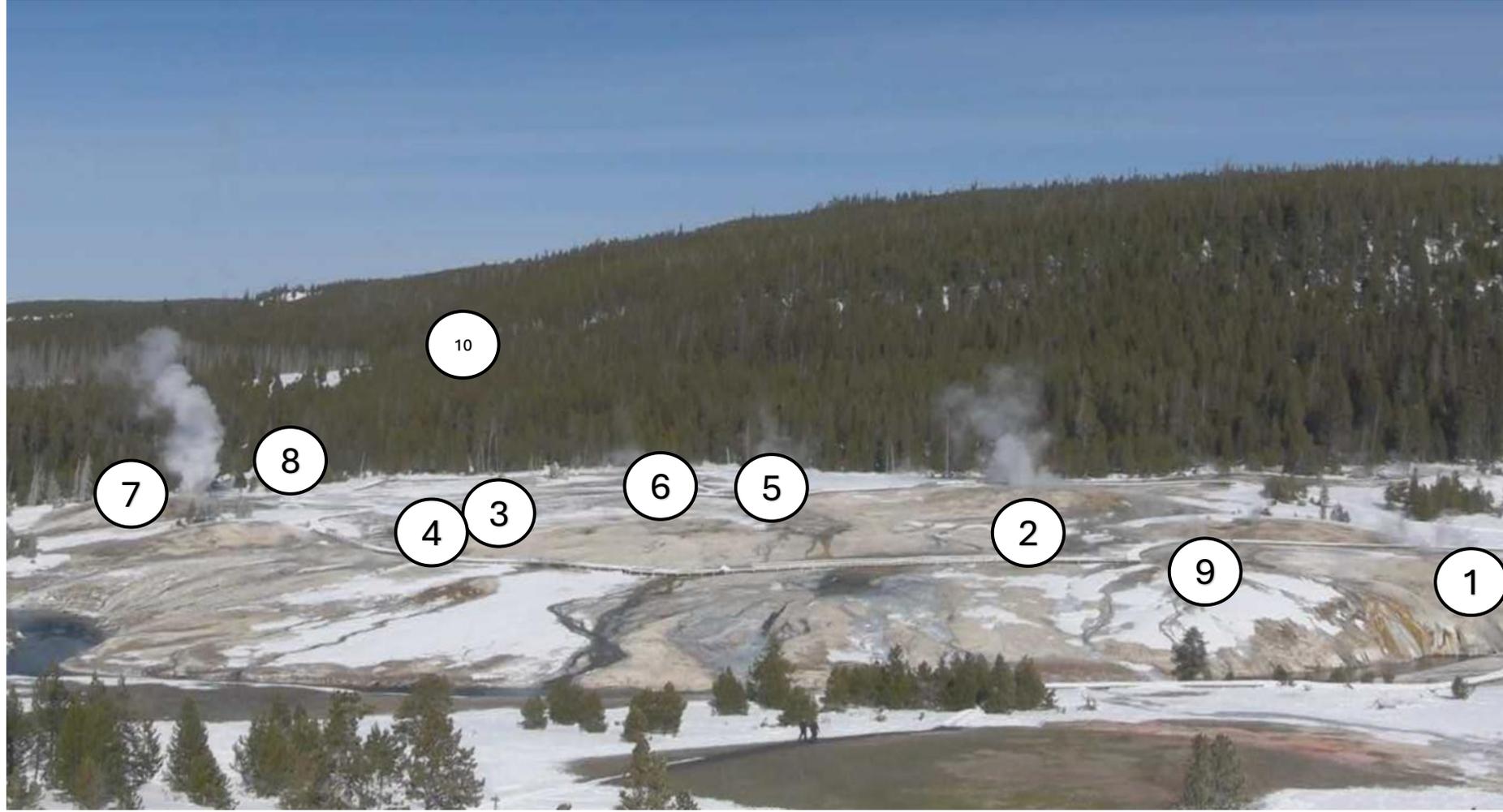6. Slot
7. Giantess
8. Dome
9. Little Squirt
10. Anemone
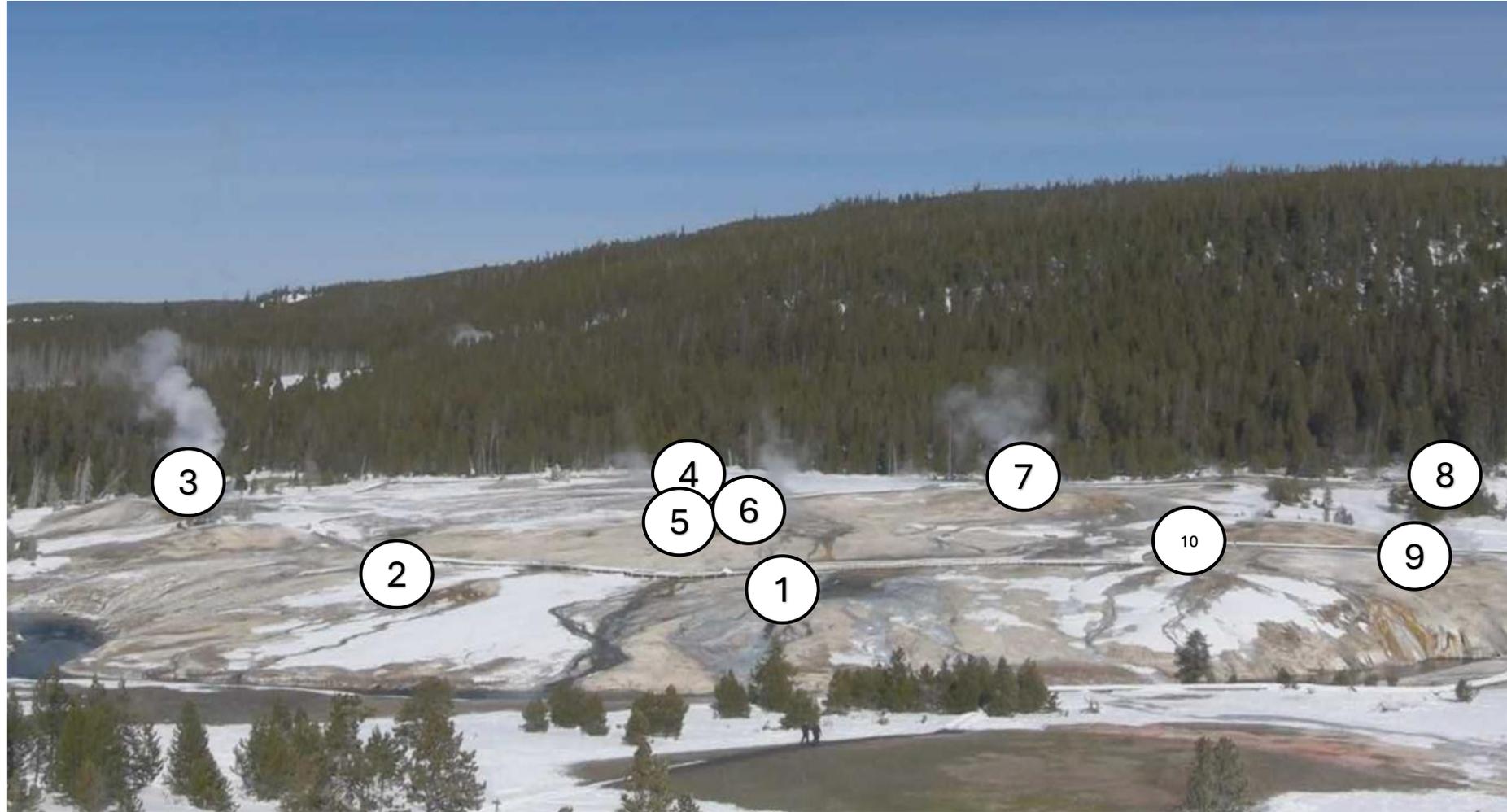
Geyser Hill: Observational Advice Pt 1

Feature	Visible?	Start Time?	End Time?	Additional Details	Comments
Beehive	Yes	Exact	+/-15sec		Exact end difficult
Depression	Yes	Exact/ns	+/-10sec		End is when pool drops
Lion	Yes	Exact	+/-10sec	Initial, Normal, Minor	
Aurum	Yes	Exact	+/-10sec		
Plate	Yes	Exact	+/-30sec		Continuous Jetting
Slot	Yes	Exact	+/-30sec		First visible splash
Giantess	Yes	Exact	+/-1min		
Dome	Yes	Exact/ns	+/-1min	Initials/Near Start of Series based on strength	
Little Squirt	Yes	Ns/ie	+/-5min		May restart
Anemone	Yes	Exact	+/-5sec	Big vs. Little	

Geyser Hill: Feature Identification Pt 2

1. Silver Spring
2. Plume
3. UNNG-GHG-18
4. Marmot Cave
5. Pump
6. Doublet Pool
7. Little Cub
8. UNNG-GHG-5
9. Improbable
10. Solitary

Geyser Hill: Observational Advice Pt 2

Feature	Visible?	Start Time?	End Time?	Additional Details	Comments
Silver Spring	Yes	Ie	N/A		Look for bursting
Plume	Yes	Exact	Unknown	Please come back!!	
UNNG-GHG 18	Yes	Exact	+/-15s		
Marmot's Cave	Yes	Ns			
Pump	Yes	N/A	N/A		No Entry
Doublet Pool	Yes	N/A	N/A		Distracting steam
Little Cub	Yes	Exact/ns	+/-15s		
UNNG-GHG 5	Behind Trees	Ie	N/A		Look for large steam, water
Improbable	Yes	Exact	+/-15s	Only enter majors	
Solitary	Behind Trees	N/A	N/A		Rarely Enterable

Beehive: Feature Identification

1. Beehive
2. Beehive's Indicator
3. South Bubbler
4. West Bubbler
5. Close to Cone
6. Blooper
7. Frying Pan
8. SW Steam Vents
9. NW Bubbler
10. Far NW Bubbler
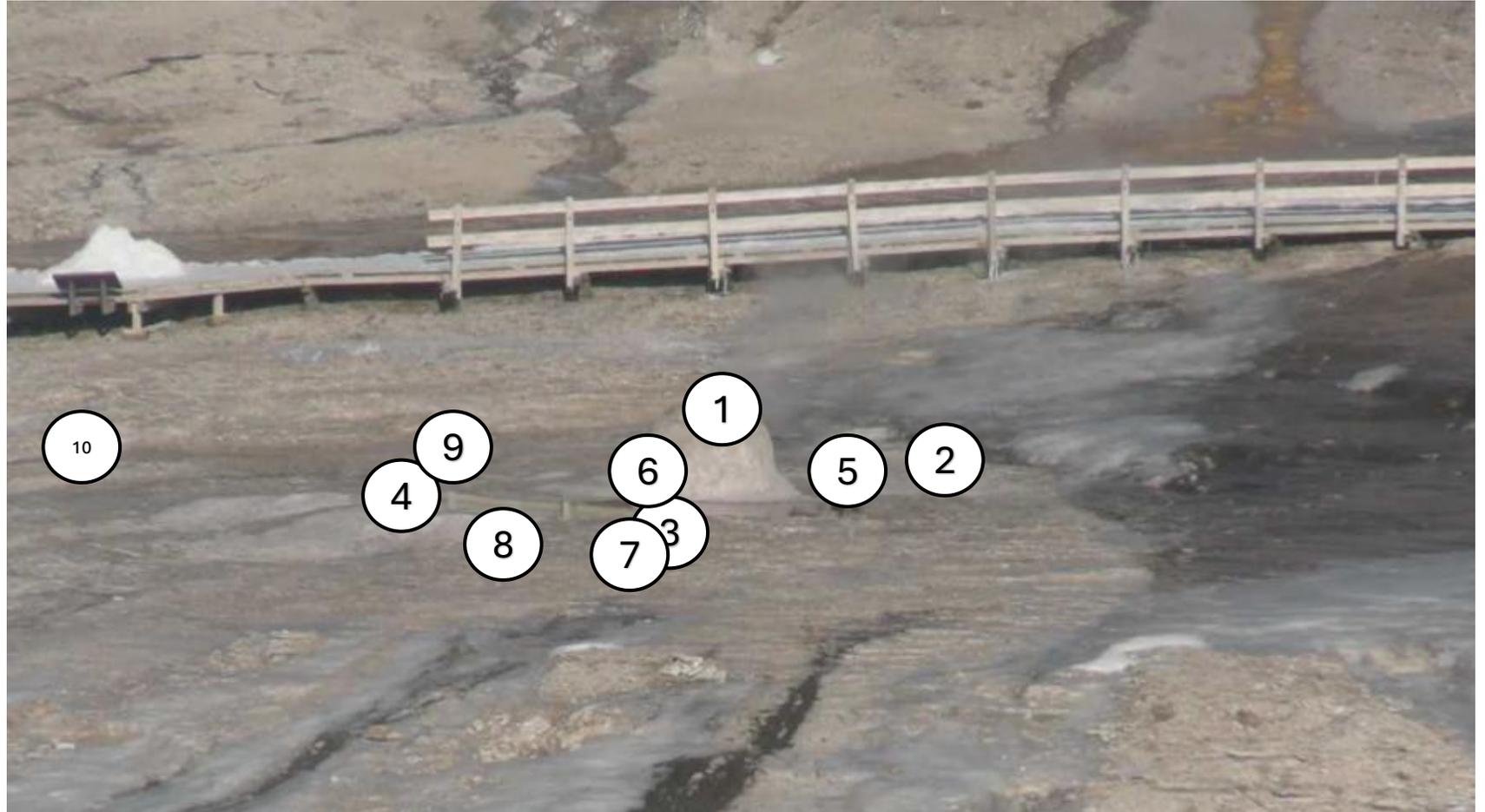
Beehive: Observational Advice

Feature	Visible?	Start Time?	End Time?	Additional Details	Comments
Beehive	Yes	Exact	+/-15sec	Indy vs. No Indy	
BHI	Yes	Exact	+/-10sec		
BSB	Yes	Exact	+/-10sec		MUST have visible water
BWB	Yes	Exact/ns	+/-5min		Any visible water
CCI	Sometimes	Ie	N/A		Rarely enterable
Blooper	No	N/A	N/A		
Frying Pan	Yes	Note in GT			Easily confused with south
SW Steam Vents	Yes	Note in GT		Between BSB and BWB	
NW Bubbler	No	N/A	N/A		
Far NW Bubbler	No	N/A	N/A		

Down Basin: Feature Identification

1. Grand Group
2. Sawmill Group
3. Castle
4. Daisy
5. Artemisia
6. Grotto
7. Giant
8. Fan and Mortar
9. Riverside
10. Churn
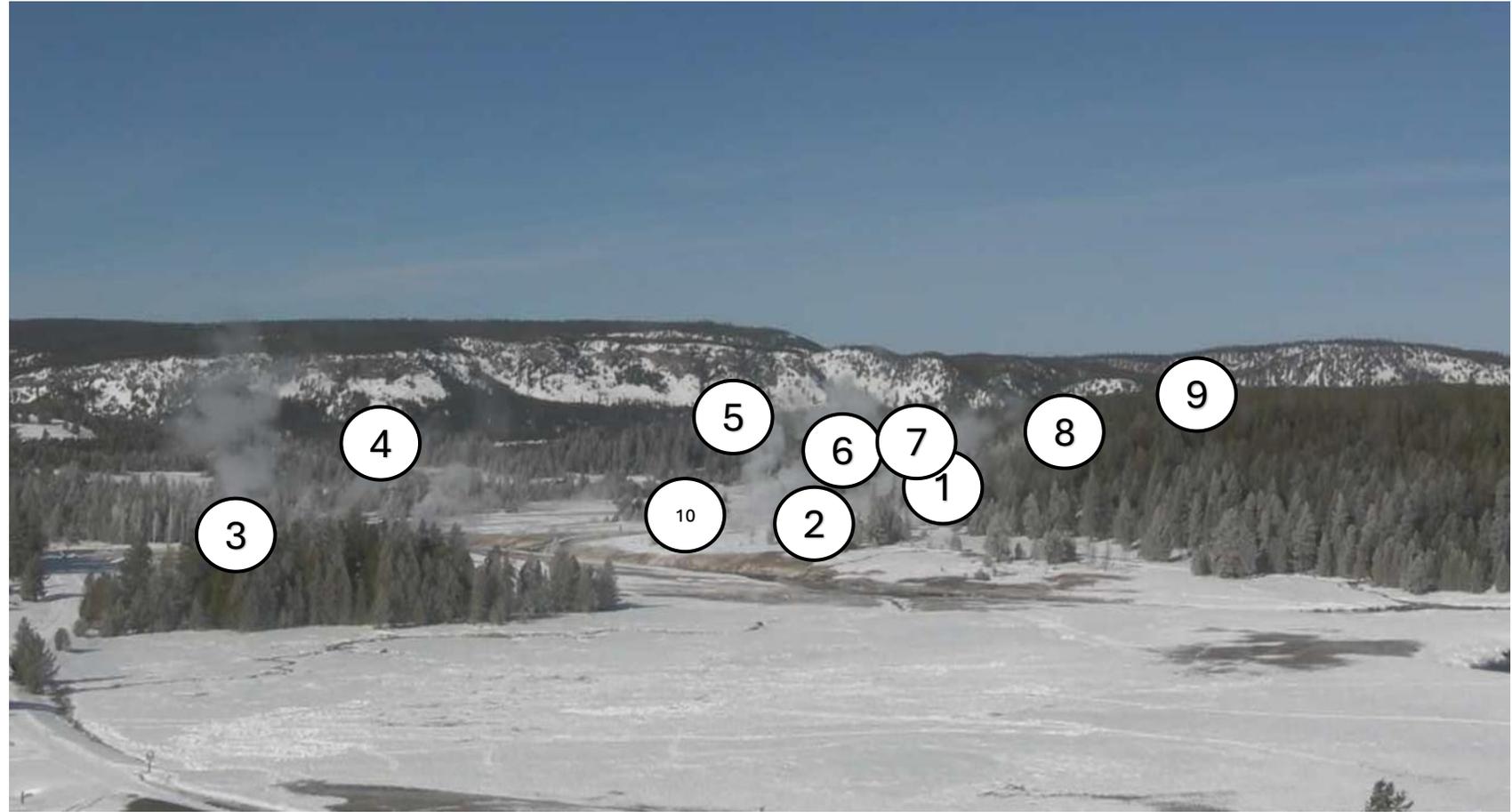
Down Basin: Observational Advice

Feature	Visible?	Start Time?	End Time?	Additional Details	Comments
Grand Group	Yes	See Details	See Details	See Details	See Details
Sawmill Group	Yes	See Details	See Details	See Details	See Details
Castle	Behind Trees	Exact	+/-1min	Able to tell maj/min from steam phase	Watch out for large preplay
Daisy	Behind Trees	Exact	+/-1min		
Artemisia	Behind Hill	Ie	Not Possible		Must be large, continuous steam plume
Grotto	Behind Trees	Ns	+/-5min	Possible to see Rocket, Grotto Ftn, and Spa	
Giant	Behind Trees	Exact	Unknown	Hot Periods will be impossible to differentiate	
F&M	Behind Hill	ns/	Not Possible	Can see pause/restart	Nuclear Steam Cloud
Riverside	Behind Hill	Ns/ie	Not Possible		
Churn	Yes	Exact	+/-30sec		Steam does not mean eruption

Grand and Sawmill: Feature Identification

1. Grand
2. West Triplet
3. Percolator
4. Bulger
5. Old Tardy
6. Penta
7. Spasmodic
8. Sawmill
9. Tardy
10. Oblong

Grand and Sawmill: Observational Advice

Feature	Visible?	Start Time?	End Time?	Additional Details	Comments
Grand	Behind Trees	Exact	+/- 10s	Burst Count, Sometimes C/Q	Cannot see pool
West Triplet	Behind Trees	ie	+/- 2min		
Percolator	Yes	ie	Difficult		
Bulger	Yes	Exact	+/- 10s	Major/Minor	
Old Tardy	Yes	ns	+/- 30s		
Penta	Yes	Exact/ns	+/- 1min	ID of Major/Minor Type	Cone slightly hidden
Spasmodic	Behind Trees	ie	+/- 10min	On/Off	
Sawmill	Yes	Exact	+/- 30s		
Tardy	Yes	Exact	+/- 30s	ID of Tardy mode possible	
Oblong	Behind Hill	ns/ie	Difficult	Eruption Strength	Hard to see flood

Black Diamond Webcam

- Eruption Indications:
  - Pool Level Drop
  - Wash/Rock Movement
  - Front Pools Drop

- Enter Approximate (A wc) time, with comments, based on best guess of when

YS-BBSN 13:00:03 2026-01-21

Webcam Considerations

- Everything comes back to visibility and eruption type!
  - Refer to TSB or resources of preference to crosscheck eruption style, height, duration, location to help with feature identification
- Start Time
  - If you can see pool/cone, exact.
  - Can almost see pool/cone, ns
  - Cannot see pool/cone, ie
- Always use webcam tag
- Enter Duration based on ability to see start/end
  - Able to see start and end? Use Approximate Duration
  - Not able to see start or end? Do not enter duration
- End Time
  - Consider eruption type: Fast Quit? +/- 15s or better. Not visible or slow quit? Less accurate

  Webcam has inherent limitations.

Seismograph Resources

Yellowstone Volcano Observatory:

https://www.usgs.gov/observatories/yvo

Use map to navigate to preferred seismometer, click (black triangles)

https://www.isthisthingon.org/Yellowstone/daythumbs.php
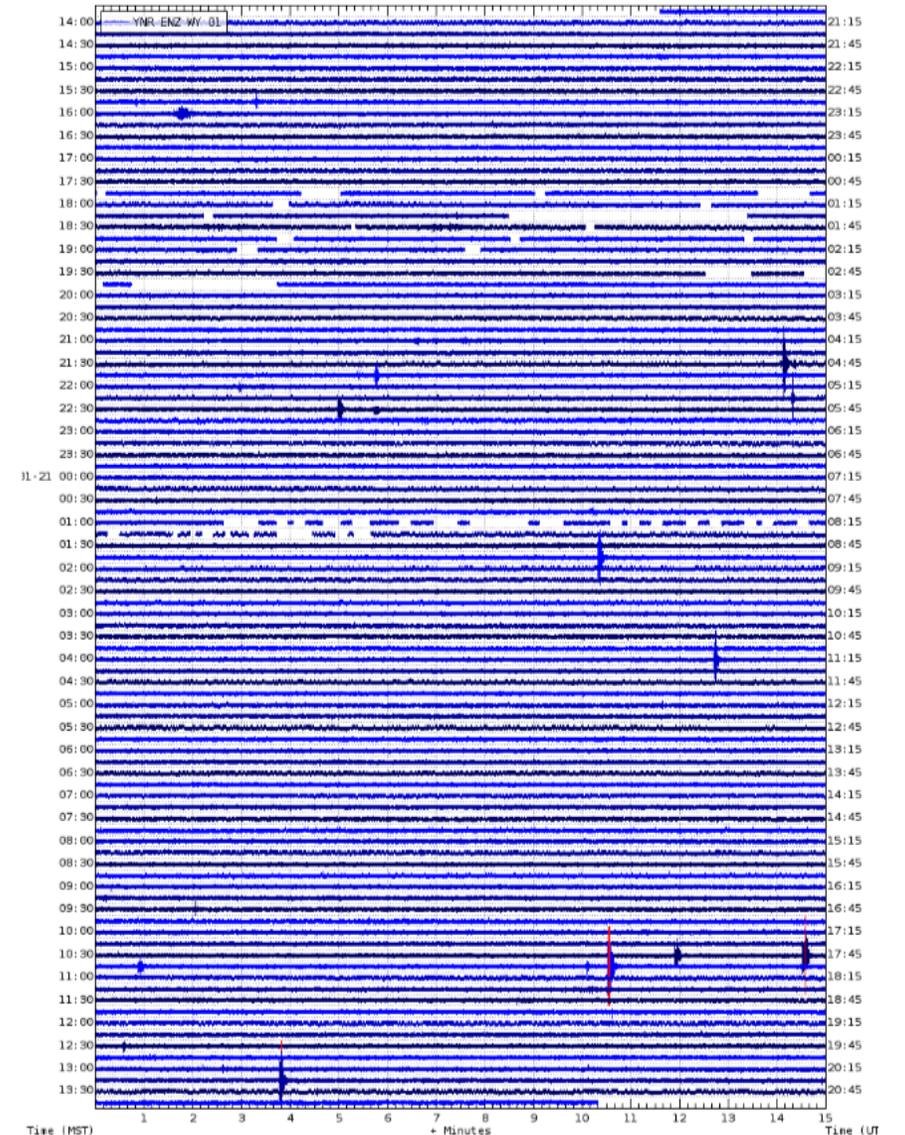
Displays some important seismographs

▲ Seismometer Monitoring Station
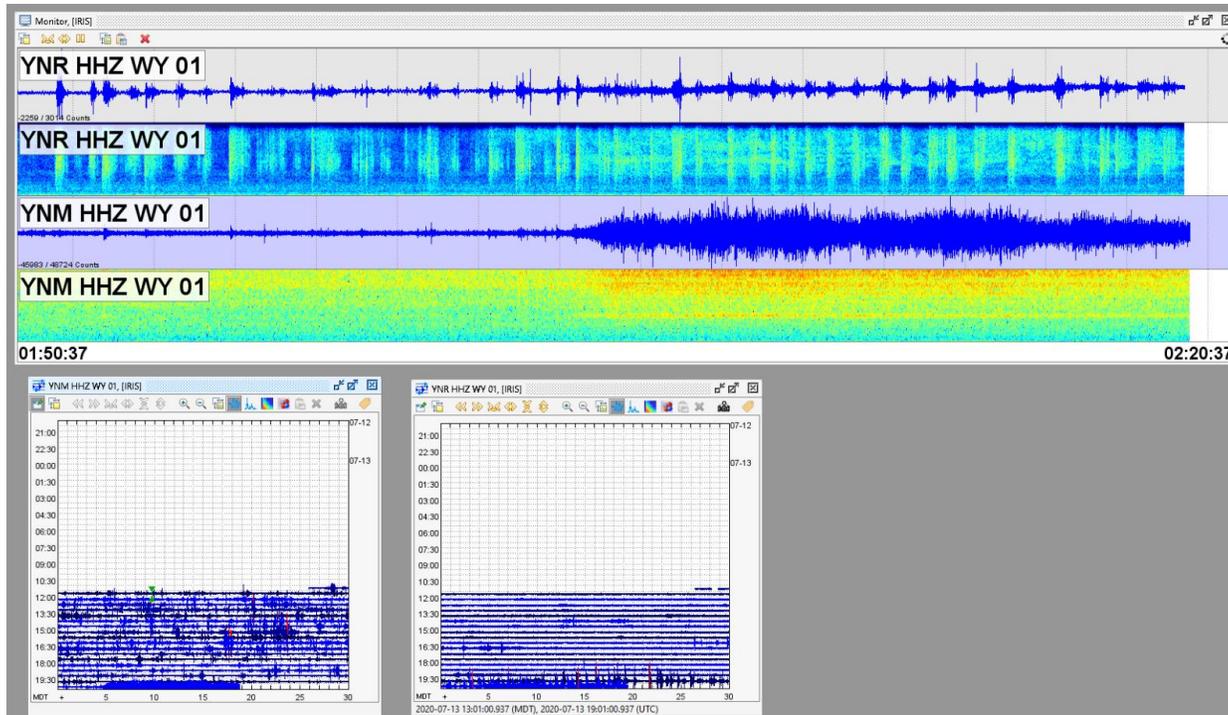
Station: YNR
Operator: UU

Last 24 Hours

Steamboat Geyser on Seismograph

- YNM or YNR
- Consider downloading Swarm software to look at spectrogram

Eruption Indications:
- Sustained increase in YNM and YNR
- Frequency Signature like this example
- Combine with Tantalus Creek Gauge (water output)
- Later: Confirm In Basin or via. Steamboat Temperature Gauge

Geysertimes Tricks

&n Trick to View More Eruptions

https://geysertimes.org/geyser.php?id=Beehive&n=1000

Adding &n=### onto end of link pulls ### eruptions instead of generic 100

Retrieve

https://geysertimes.org/retrieve.php

Detailed Entry Search Allows Exporting Data to Browser or CSV
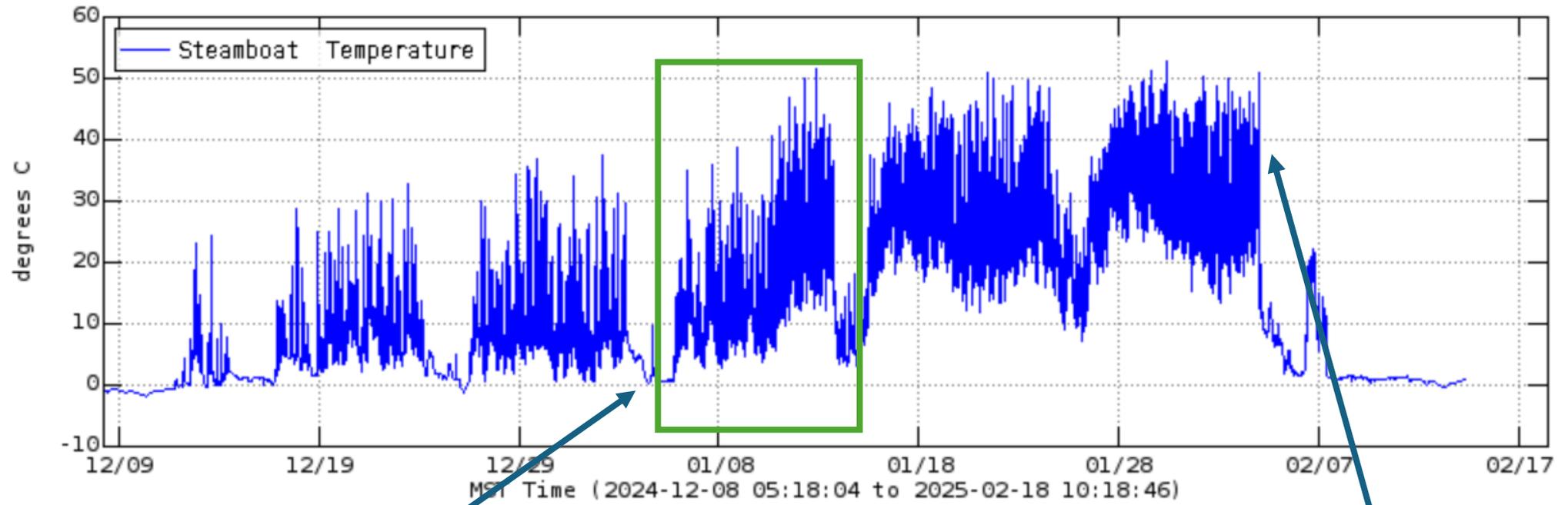
Analysis Page

Allows basic chart creation

YVO Valve

Allows analysis of Seismometers, Deformation, and Norris Temperatures

Click "Norris Temperatures," Select Feature, Enter Start/End Time in form YYYYMMDDHHMM, Submit!

Clicking on chart twice, then collecting submit again creates new chart with your newly selected start and end times.

Temperature Gauges

Most Temperature Loggers Not Accessible to Public

Steamboat is primary logger within GOSA Scope

Can track cycles, and eruption times

Water Gauges

Accessible through YVO Valve.

Steamboat Direct Link:

https://waterdata.usgs.gov/monitoring-location/USGS-06036940/#dataTypeId=continuous-00065-0&period=P7D&showFieldMeasurements=true

- Change time frame, download data.

- For Steamboat: Look for spike

- **CAUTION**: Precipitation events may look like eruptions

at Norris Junction, YNP - USGS-06036940

Subscribe to Wate

- using custom time span -
December 25, 2025 - January 3, 2026
Gage height, feet

Dec 31, 2025 11:10:00 PM MST

Eruption

Dec 26   Dec 28   Dec 30   Jan 01   Jan 03
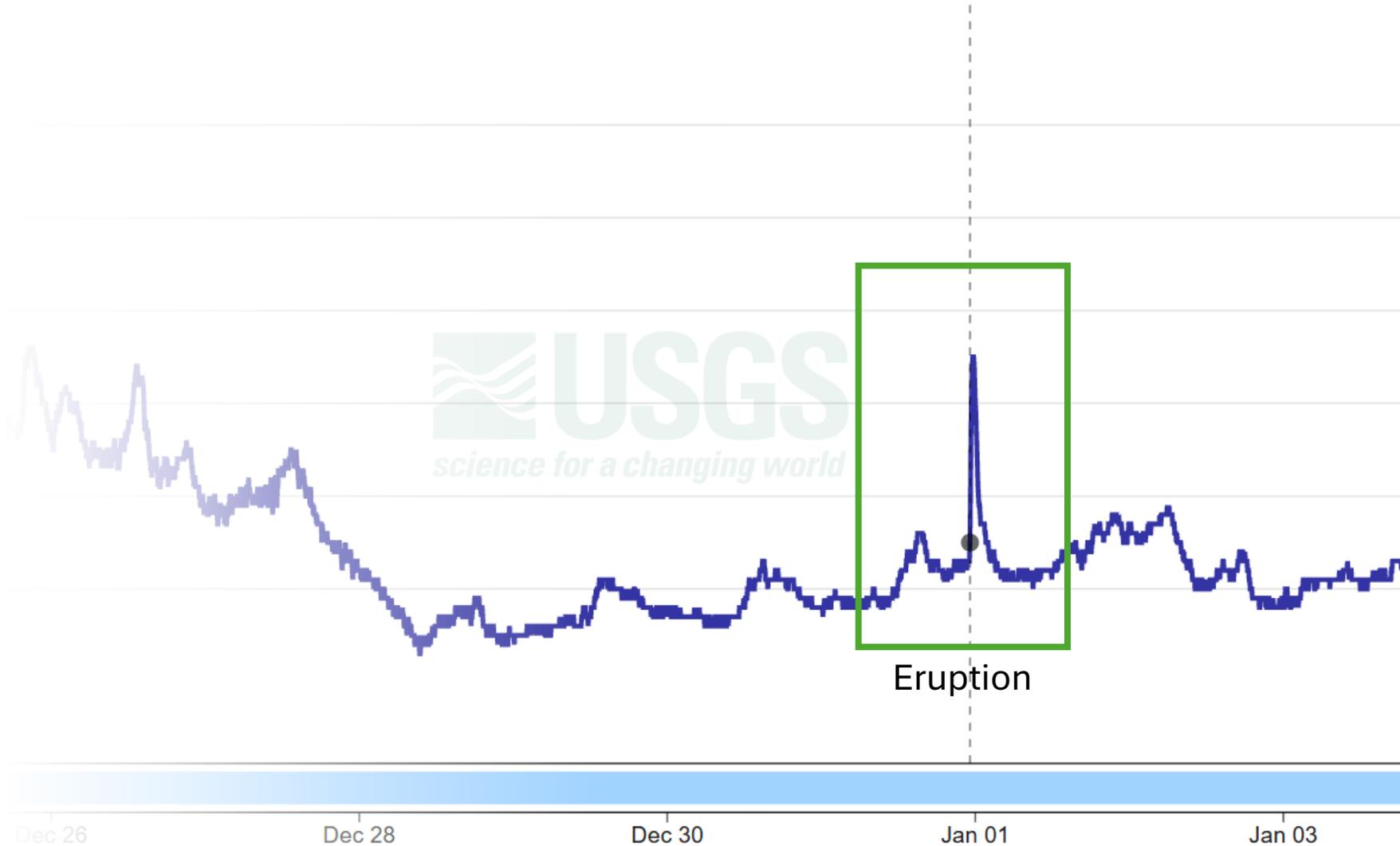
Introduction to Data Science

| IS			NOT		
	Asking a question	Clear, Simple, Defined Scope			
	Collecting Data	Variables of Interest, Possible Confounds		Running Code, Trusting Output	Must check each step
	Cleaning	Remove "Bad" Datapoints Manual or Filtering		Making Fancy Plots	Plots, Numbers serve purpose
	Analyzing	Calculation and Manipulation		Letting the computer find the answer	Interpret numbers in context
	Interpreting	Apply Analysis to Question			

How do we think about data?
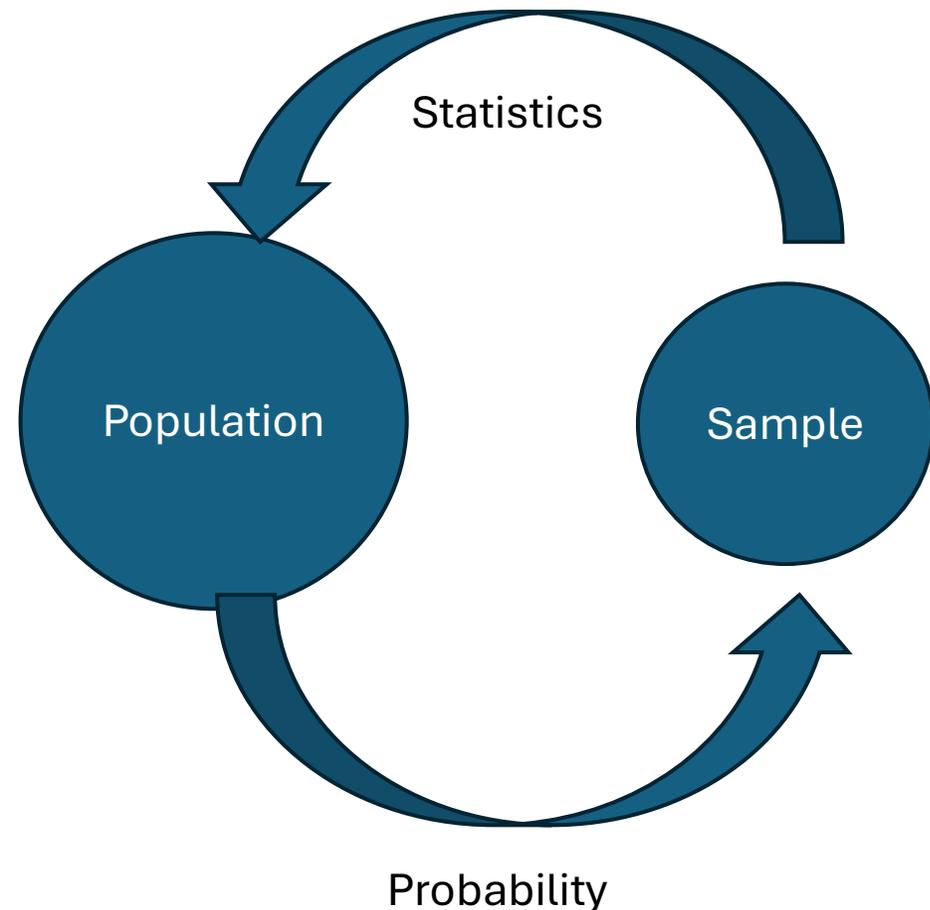
Population vs. Sample
- **Population:** The complete set of all possible geyser eruptions or behaviors.
- **Sample:** The subset of eruptions we actually observe and record.
- **Key idea:** Samples estimate populations, but never capture them fully.
- **Implication:** Conclusions always include uncertainty and possible bias.

Statistics vs. Probability
Statistics: Using data to describe patterns and estimate unknown parameters.
Probability: Using assumed models to predict outcomes before observing data.
Statistics asks: "What does this data tell us?"
Probability asks: "What should we expect to see?"

How do we think about data? Pt. 2

Observational vs. experimental: Geyser data is observed, not controlled or repeatable.

Data is not truth: Data records observations, methods, and limitations—not reality itself.

Uncertainty is expected: Variability is natural, not a flaw in the data.

Asking the right questions

- It is important to ask good questions
  - Clear variables, time scale, scope
  - Specific, Testable, Addresses an element
  - Big Projects: Asking many questions to create a complete picture

Bad Question:
  "Is this random" -> Not Testable
  "How does Beehive like weather" -> Need to be broken down

Good Question:
  "Is Grand more likely to start with Turban than without" -> Testable
  "How does Beehive respond to rain" -> Single Element

Variables and Workflow

Ask Question	"How does Beehive respond to rain" -> Simple, Clear
Determine Hypothesis	"Beehive intervals lengthen with rain" -> Testable
Decide Variables	Beehive intervals, precipitation, other environmental variables
Account for Confounding Variables	Other environmental variables are within the same "bucket"
Clean Data	Most Challenging Step, remove bad datapoints, unit conversions, formatting

Variables and Workflow

Analysis	What numbers do I need? What charts? Ie. Beehive intervals w/ rain versus w/o rain
Reproducibility	If I used data from July, test again with August, test again with July of last year
Stupid Check!	Does this make sense intuitively? Does it fit with your observations?
Back to Hypothesis	Confirm or reject your hypothesis. Ie. Beehive intervals are longer during rain
Answer Question	How does hypothesis go back to question? Ie. Beehive has a negative response to rain

Methods in Data Science

Descriptive Statistics: Summarize behavior using means, medians, ranges, ect.

Visualizations: Reveal patterns, trends, errors through plots and charts

Time-Series: Explore how variables change over time

Comparative Analysis: Compare Groups (like seasons)

Correlation: Measure Relationships between variables

Outlier Analysis: Identify unusual events

Modeling: Explore patterns predictively

Data Science Resources

Programming and Analysis Tools

- R: Excellent for statistical analysis and visualization

- Python: Flexible language for analysis, automation, modeling

- Excel/Google Sheets: Accessible tools for logging and basic analysis

Preference:

- Use Excel/Google Sheets to store data

- Use R or Python to perform analysis
  - Python is better at data manipulation
  - R may be easier to learn

Issues in Geyser Data

1 | **Correlation vs. Causation**

Correlation: Shows association
Causation: Physical Response

Bad: Dome causes short Beehive intervals
Good: Beehive has short intervals around the time Dome erupts

Guidance:
1) Observational data rarely allows causation
2) Physical processes may be suggested as causation
   Ie. Rain caused Beehive delays by an influx of cold water

2 | **Open vs. Closed Datasets**

Open: Incomplete, Missing Data Points
Closed: Curated or Complete

Bad: Mixing Open and Closed Datasets
Good: Curate data to remove missing data

Guidance:
1) Ensure datasets are as close to closed as possible
2) Be transparent: Explain the size of your dataset and the methods you used to curate it

Issues in Geyser Data

3 **Observational Bias**

Inherent impact of observation, beliefs, or expectations on a dataset

Examples:
- Geysers are observed more during day, and summer
- Well-known geysers are observed more frequently

Guidance:
- A lack of an observation does not mean a lack of event
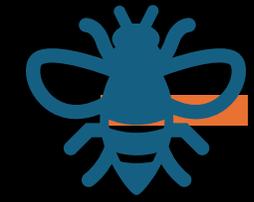- Random Sampling can help

4 **Errors in Data Curation**

Error introduced when deciding what data to include in analysis.

Examples:
- Filtering out intervals >8hrs misses 8.5hr intervals
- Not filtering out certain intervals introduces double intervals

Guidance:
- Hand curation may be necessary with geyser data
- Only work with datasets you know

Case Study:
Beehive
The Impact of
Indicator Lead time
on Beehive Duration

Case Study: Starting the Process

Ask the Question:

Do Indicator Lead Times Impact Beehive Durations?

Determine Hypothesis:

Longer Indicator Lead Times Result in Shorter Beehive Durations

Decide Variables:

Beehive Eruption Time, Preceding Indicator Time

Understand the Confounding Variables:

Mid Cycles, No Indicators, Environment

Case Study: Working with a Dataset

Clean Data:

- Pull Indicator and Beehive eruption times

- Find nearest Indicator time before Beehive's eruption time

- If not within 40 minutes, require manual check

- Record Indicator Lead Time (Beehive Time – Indicator Time) for any Beehive eruption with a duration

- Remove Inadequate Durations (~5m is not helpful, 5m14s is)

- Pair Data (x=Indicator Lead Time, y=Beehive duration)

- Print pairs and check manually to ensure proper calculation

Case Study: Working with your Dataset

Analyze Data:

- Bin Indicator Lead Times
  - X(1) represents all Beehive's w/ duration where X (indicator lead time) is 1 minute

- Treat Each Bin as it's own sample
  - Calculate mean, median, average, and standard deviation for X(n) where n is a bin between 0 (no Indicator) and 40 (maximum allowed lead time)
  - Create a "box a whiskers" plot to show distribution for each bin

- Compare Bins
  - Chart Duration= d[X(n)] where d is the distribution of each X(n) bin

Duration vs Lead Time - Last 365 days
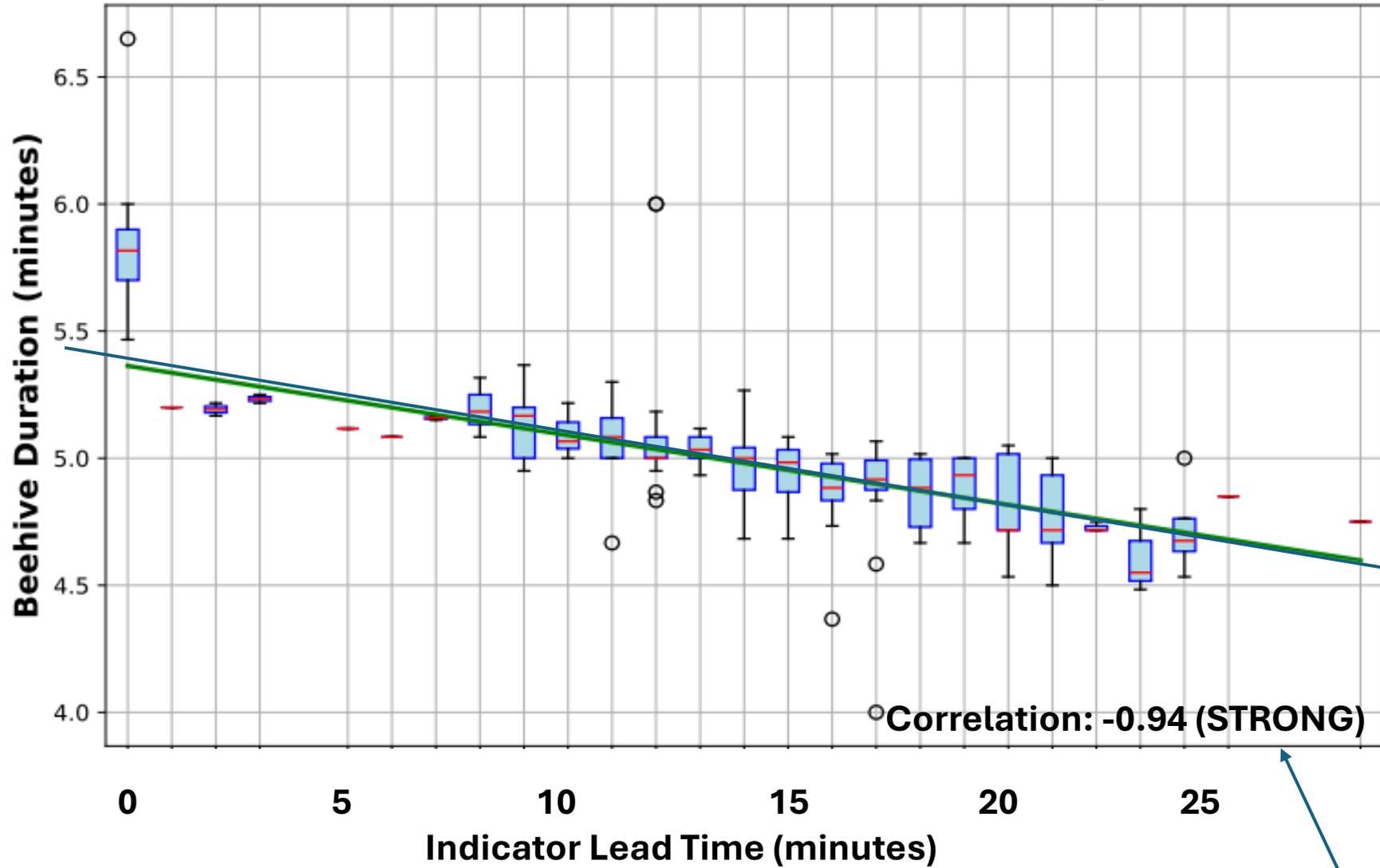
Clear Negative Trend

Correlation: -0.94 (STRONG)

STRONG Correlation

Case Study: Interpret Results

Reproducibility:

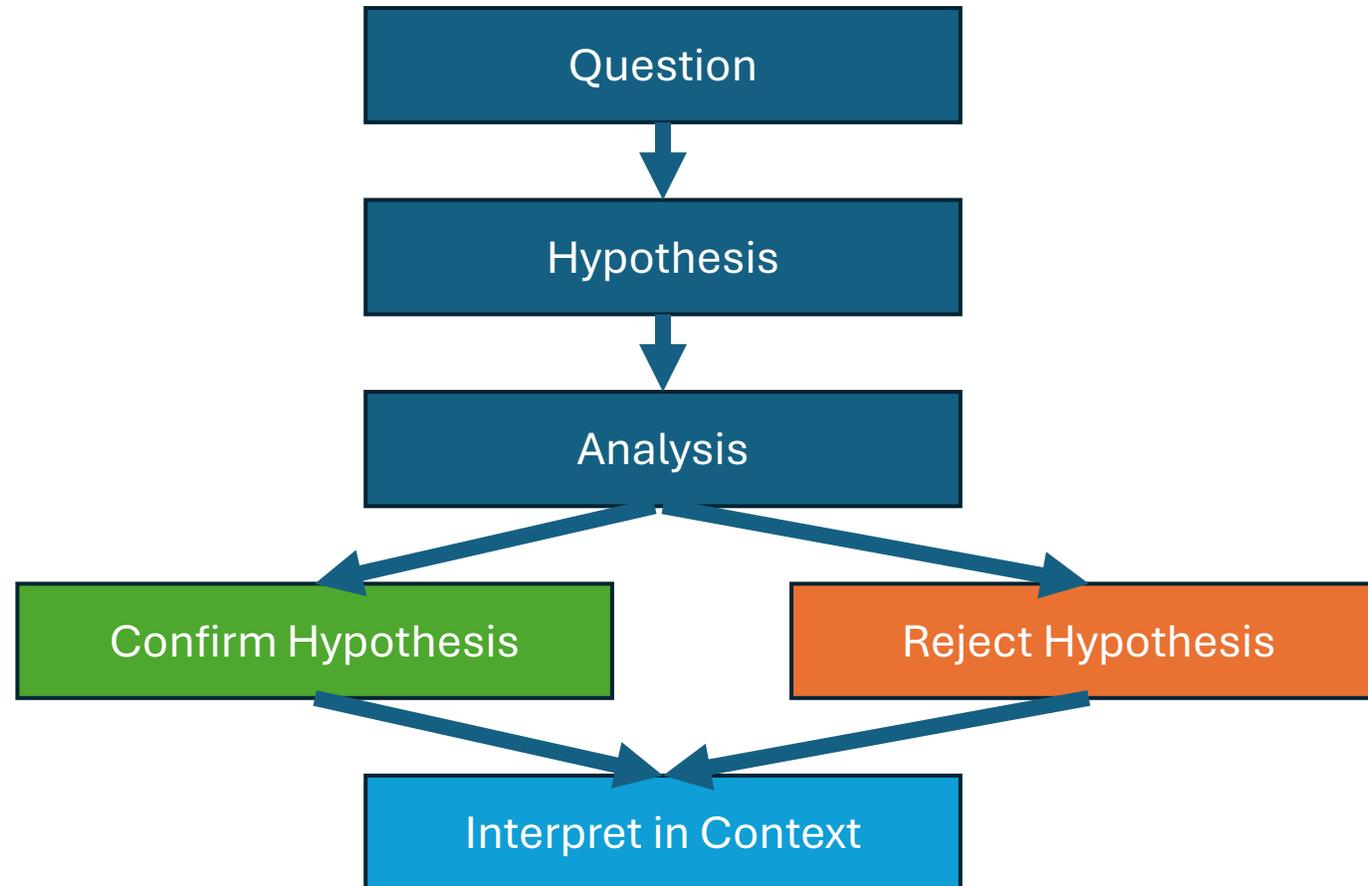   Tested across multiple time frames, all showed same results

Stupid Check:

    Has a physical explanation, makes sense, can see it in the data

Back to Hypothesis and Question:

    CONFIRMED! Longer Lead Times = Shorter Duration

AI in Geyser Data

AI is a tool, not a scientist: It does not understand geysers or physical processes

Best Use: Code Acceleration. Writing Scripts when the goal is already defined.

Human Judgement is Required: User must choose variables, filters, assumptions, and analysis

Do not feed it raw data: AI cannot evaluate data quality, and often makes computational errors

Not a substitute for analysis: AI speeds up workflows, but does not replace reasoning

Data Spotlight: CDF Optimization of Sit Time

- CDF calculates probability of some event (eruption) having happened by some time
- Effectively, function for % of eruptions occurring by X hour.
- Allows for optimization of sit time

Grand Geyser Optimal Sit Table:

Coverage (%)	CDF Lower	CDF Upper	Interval Start (hrs)	Interval End (hrs)
25	0.375	0.625	7.942778	8.596944
50	0.250	0.750	7.637222	8.903333
60	0.200	0.800	7.360278	9.022222
70	0.150	0.850	7.122778	9.164722
75	0.125	0.875	7.044722	9.164722
80	0.100	0.900	6.873611	9.552222
85	0.075	0.925	6.806111	9.622778
90	0.050	0.950	6.328333	9.972222
95	0.025	0.975	6.109722	10.691944

Remote Geyser Gazing Review

Know what the webcam can and cannot see

Timing is approximate, not exact

Use multiple data sources when possible

Observation bias built in

Document context and observation, not just events

Data Science Review

? Good Questions Matter: Use Statistics/Probability to Confirm or Reject a Hypothesis

👁 Observational Data has Limits

⚙ Sanity Checks are Essential: Geyser Gazers will often know the answer before the analysis

💬 Visualiation is Analysis and Communication

📈 Context is Part of Data: Must understand your data

Questions?

Contact Info:

Korben Cooper

korben_cooper@mines.edu