



Intro to Data Science in Python Workshop

@unnati_xyz | @hasgeek

A person is holding a lit sparkler, with bright sparks emanating from the tip. The person's face is partially visible in the background, and they are wearing a dark, textured garment. The overall scene is dark, suggesting a nighttime setting.

Welcome



Facilitators



Amit

@amitkaps



Nischal
@nischalhp



Raghotham

@raghothams



Bargava

@bargava



Shrayas

@shrayasr

Three camera lenses are arranged horizontally on a dark, reflective surface. The lens in the center is the largest and most prominent, with its front element clearly visible. To its left and right are two smaller lenses, also showing their front elements. The background is a dark, gradient grey, and the overall lighting is soft, highlighting the metallic and glass textures of the lenses.

**See the world
through a data
lens**

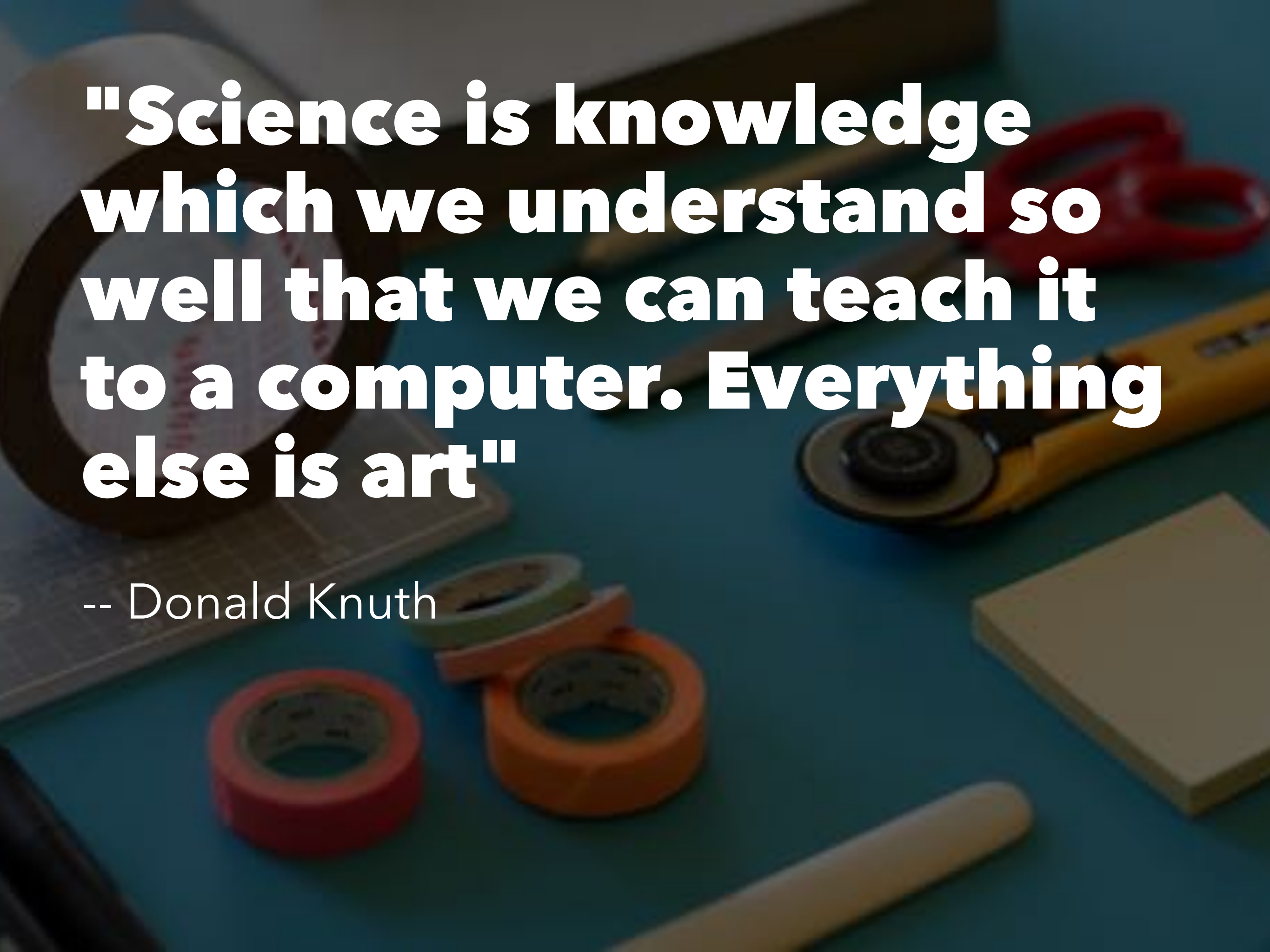


**"Data is just a clue to the
end truth"**

-- Josh Smith



Data Driven Decisions



**"Science is knowledge
which we understand so
well that we can teach it
to a computer. Everything
else is art"**

-- Donald Knuth



**Data Science is
an Art**

A blurred background image showing a person's hand holding a pen and writing on a document. The text 'Hypothesis Driven Approach' is overlaid in large, bold, white letters.

Hypothesis Driven Approach

Frame

"An approximate answer to the right problem is worth a good deal"

The background of the image is a dense, close-up shot of wood chips or shavings. The chips are in various shades of brown, tan, and grey, with some showing the natural grain of the wood. They are piled together, creating a textured, organic pattern.

Acquire

**"80% perspiration, 10%
great idea, 10% great
output"**

An axe with a wooden handle and metal head, resting on a tree stump in a forest setting. The background is blurred, showing trees and foliage.

Refine

"All data is messy."



Explore

**"I don't know, what I
don't know."**

Model

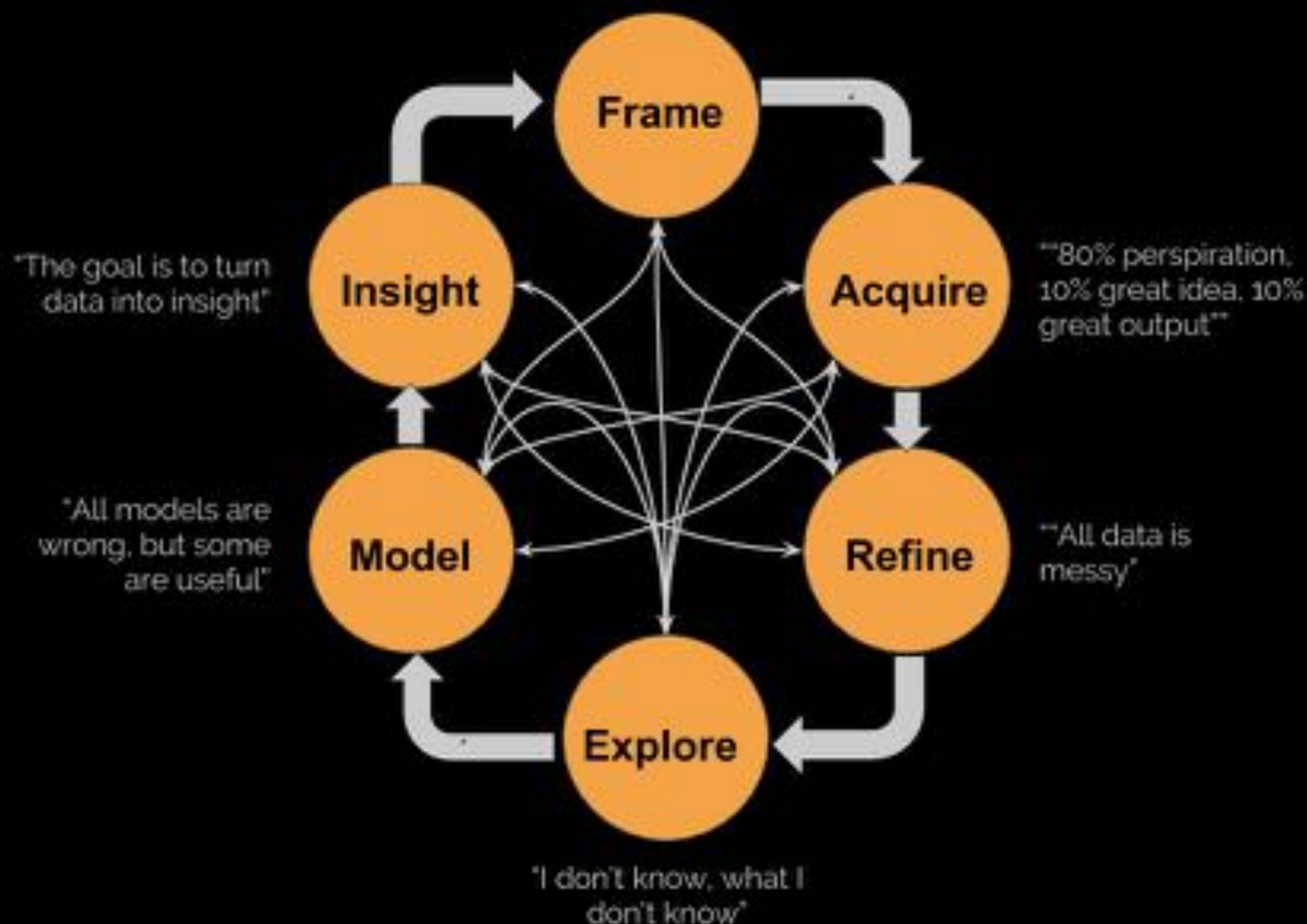
**"All models are wrong,
but some are useful"**

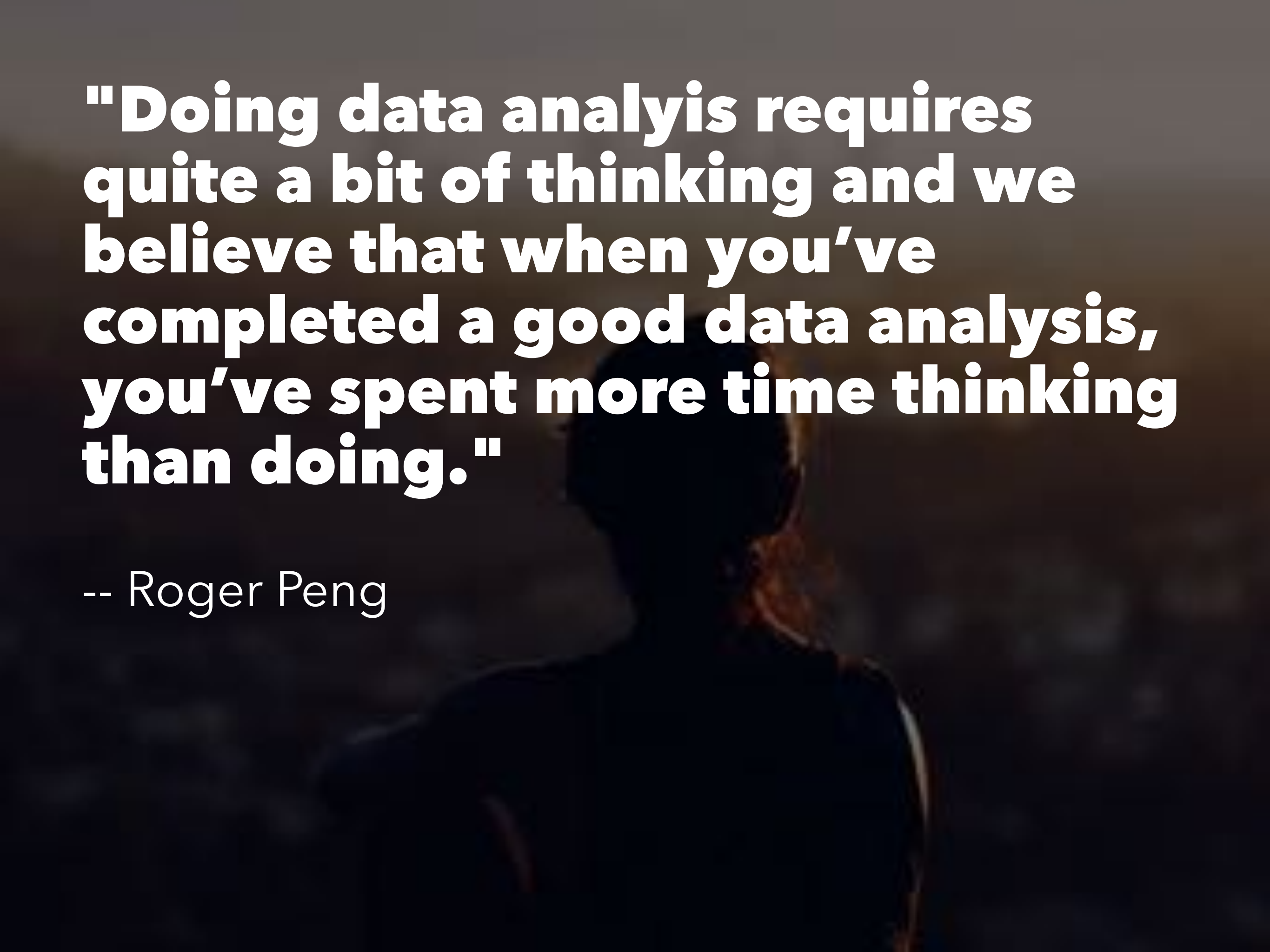
A rustic log cabin with a dark roof and a chimney, nestled in a dense forest of tall, thin trees. The scene is dimly lit, suggesting dusk or dawn, with a soft glow emanating from the cabin's windows. The text is overlaid in white, bold font.

Insight

**"The goal is to turn data
into insight"**

"An approximate answer to the right problem is worth a good deal"



A person is shown in silhouette, looking towards a bright light source, possibly a screen or window, which creates a strong backlighting effect. The person's head is tilted slightly upwards and to the right. The background is dark and out of focus.

"Doing data analysis requires quite a bit of thinking and we believe that when you've completed a good data analysis, you've spent more time thinking than doing."

-- Roger Peng

A collection of vintage tools, including axes, pliers, a saw, and a hammer, are arranged on a dark wooden surface. The tools are scattered across the frame, with some in the foreground and others in the background. The lighting is dim, creating a moody atmosphere. The text "Python Data Stack" is overlaid in the center in a large, white, sans-serif font.

Python Data Stack



Case Studies

A man with a beard, wearing a white shirt, is peeling onions in a market stall. The stall is filled with large piles of red onions. The background is slightly blurred, showing the structure of the stall and other market items.

Peeling the Onion



Hard Disk Failure



Shining Diamonds



Red Wine Selection

The Home of Data Science

COMPETITIONS • DATASETS • CUSTOMER SOLUTIONS • JOBS BOARD


Kaggle Competition

A close-up photograph of two hands, one light-skinned and one dark-skinned, gently holding a small, dark, textured object (possibly a seed or a small animal) over a light-colored, textured surface. The background is blurred, showing hints of red and white. The overall tone is warm and focused on the interaction between the hands and the object.

Learning Approach



Do the Exercises

A photograph of two men in an office environment. The man on the left, wearing a light-colored shirt, is pointing his right index finger towards a computer monitor. The man on the right, also in a light-colored shirt, is looking at the monitor and has his hand on a computer mouse. The desk in front of them holds a yellow coffee cup, some papers, and a keyboard. The background features a wall covered with numerous colorful sticky notes and a large whiteboard with faint diagrams. The text "Pair up & Learn" is overlaid in a large, bold, white font across the center of the image.

Pair up & Learn

A blurred image of a laptop computer. The screen displays a green rectangular area, possibly a video or a large sticky note. There are other yellow sticky notes visible on the screen. The text "Call for Help" is overlaid in the center in a large, white, bold font.

Call for Help



Schedule

Session 0 (0830 - 0930)

- Installation

Session 1 (0930 - 1115)

- Overview of Data Science
- Data Science Process
- How to use Jupyter Notebook
- Intro to Data Structures in Python

Session 2 (1135 - 1300)

- Case 1: Peeling the Onion
- Exploring Onion Price and Quantity

Session 3 (1400 - 1530)

- Fun visualization exercise
- Communicating the Onion Insights
- Acquiring the Onion data (Web Scraping)

Session 4 (1550 - 1730)

- Case 2: Hard Disk Failure
- Explore & build intuition around Failure Rates

Optional Advanced Session (1730 - 1830)

- Working with SQL to Acquire and Refine Hard Disk Data
- Office Hours

Food and Hydration

0830 - 0930: Breakfast

1115 - 1135: Tea Break

1300 - 1400: Lunch

1530 - 1550: Tea Break



Session 5 (0930 - 1115)

- Reflections
- Intro to Machine Learning
- Case 3: Shining Diamonds
- Linear Regression

Session 6 (1135 - 1300)

- Case 4: Red Wine
- Logistics Regression for Classification



Session 7 (1400 - 1530)

- Fun Demo : Music Visualization
- Case 5: Entering a Kaggle Competition

Session 8 (1550 - 1730)

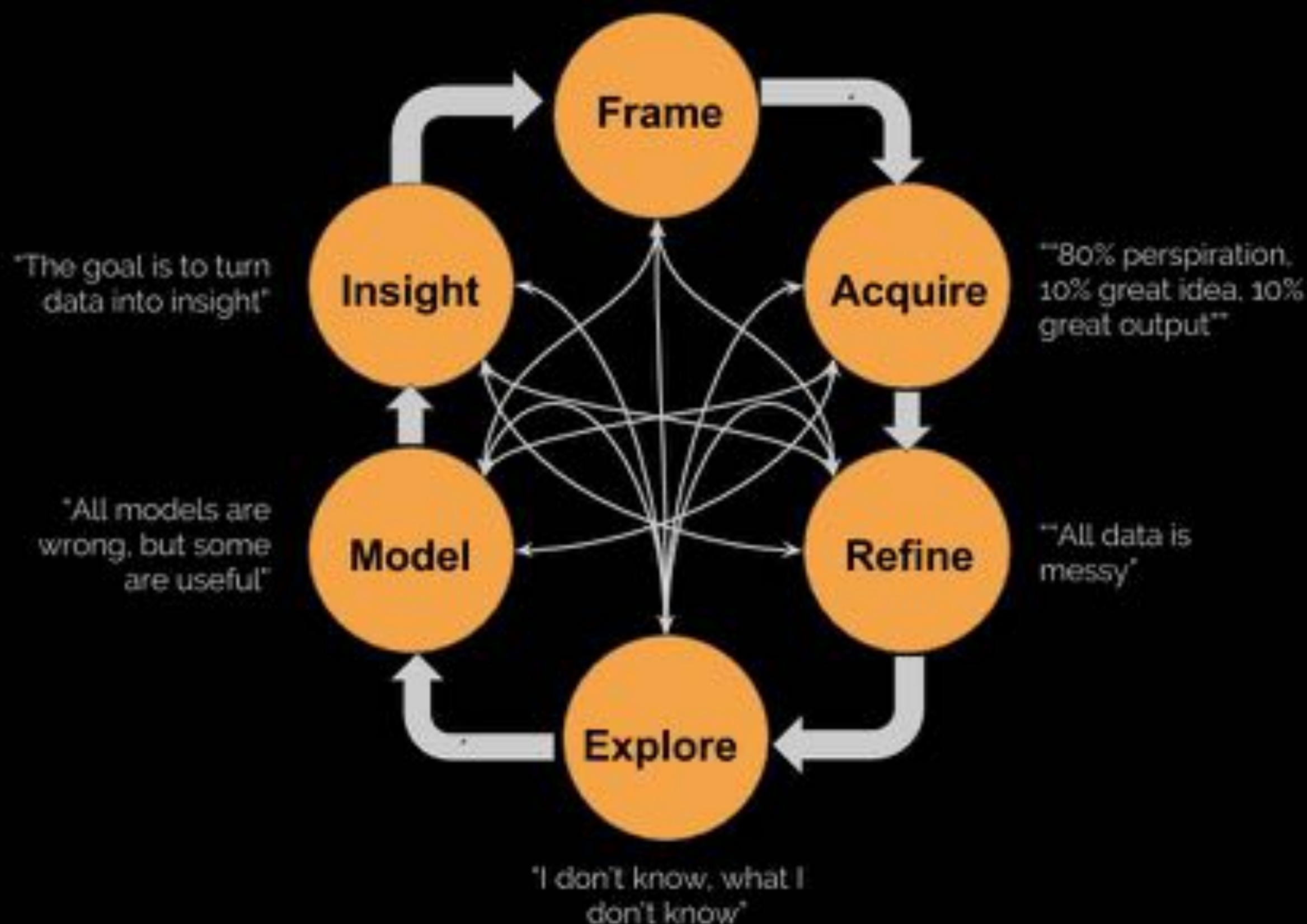
- Kaggle Competition (contd.)
- Recap and Way Forward
- Feedback



**Enjoy the
workshop**

Recap

"An approximate answer to the right problem is worth a good deal"



Frame

- **Toy Problems**
- **Simple Problems**
- Complex Problems
- Business Problems
- Research Problems



Acquire

- **Scraping** (structured, unstructured)
- **Files** (csv, xls, json, xml, pdf, ...)
- **Database** (sqlite, ...)
- APIs
- Streaming

Refine

- Data Cleaning (inconsistent, missing, ...)
- Data Refining (derive, parse, merge, filter, convert, ...)
- **Data Transformations** (group by, pivot, aggregate, sample, summarise, ...)

Explore

- **Simple Vis**
- Multi Dimensional Vis
- Geographic Vis
- Large Data Vis (Bin - Summarise - Smooth)
- Interactive Vis

Model - Supervised Learning

- *Continuous*: Regression - **Linear**, Polynomial, Tree Based Methods - CART, Random Forest, Gradient Boosting Machines
- *Classification* - **Logistics Regression**, Tree, KNN, SVM, Naive-Bayes. Bayesian Network

Model - UnSupervised Learning

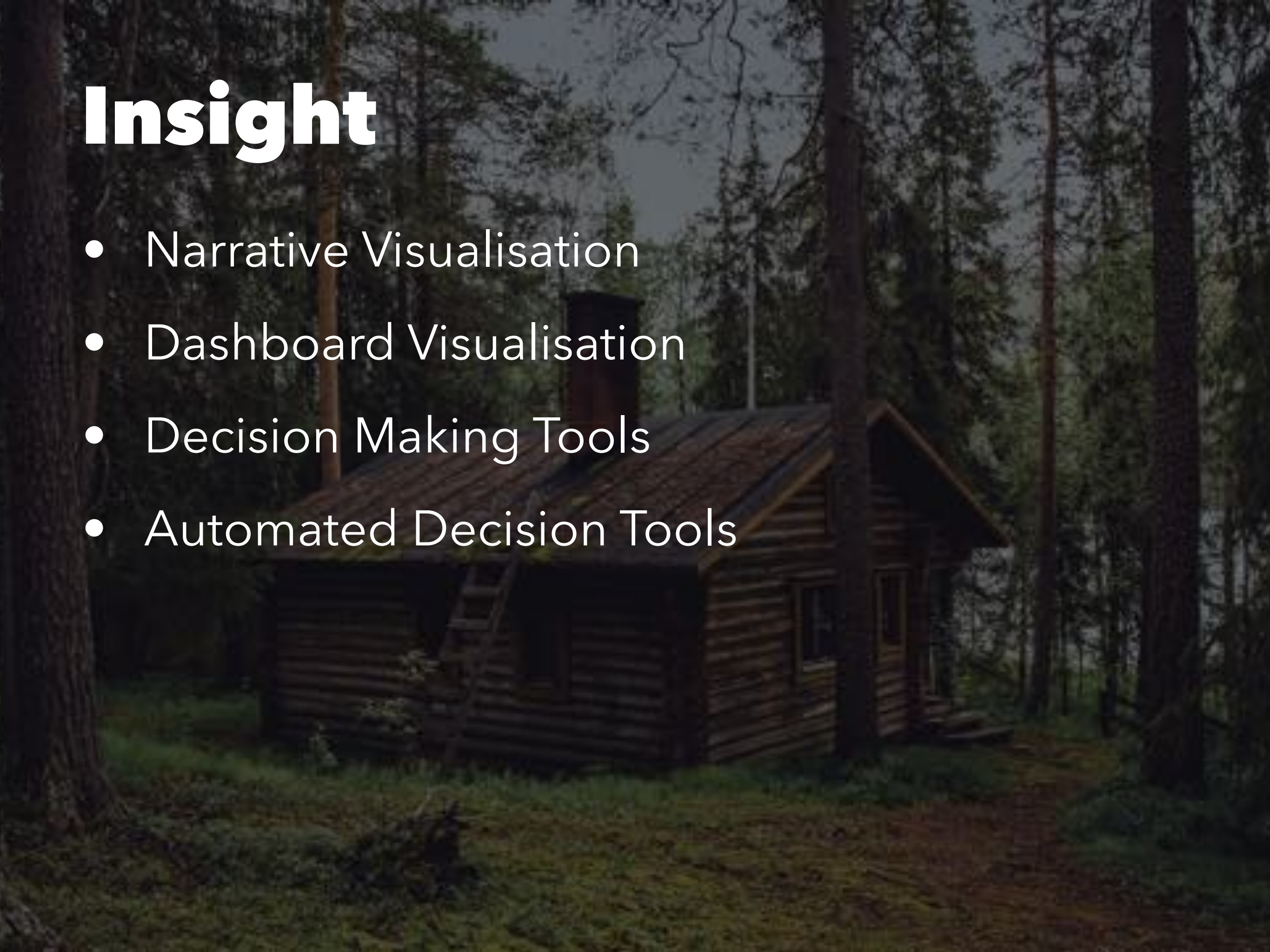
- *Continuous*: Clustering & Dimensionality Reduction like PCA, SVD, MDS, K-means
- *Categorical*: Association Analysis

Model - Advanced / Specialized

- Network / Graph Analytics
- Optimization
- Reinforcement Learning
- Online Learning
- Deep Learning
- Applications: Time Series, Text, Image, Speech

Insight

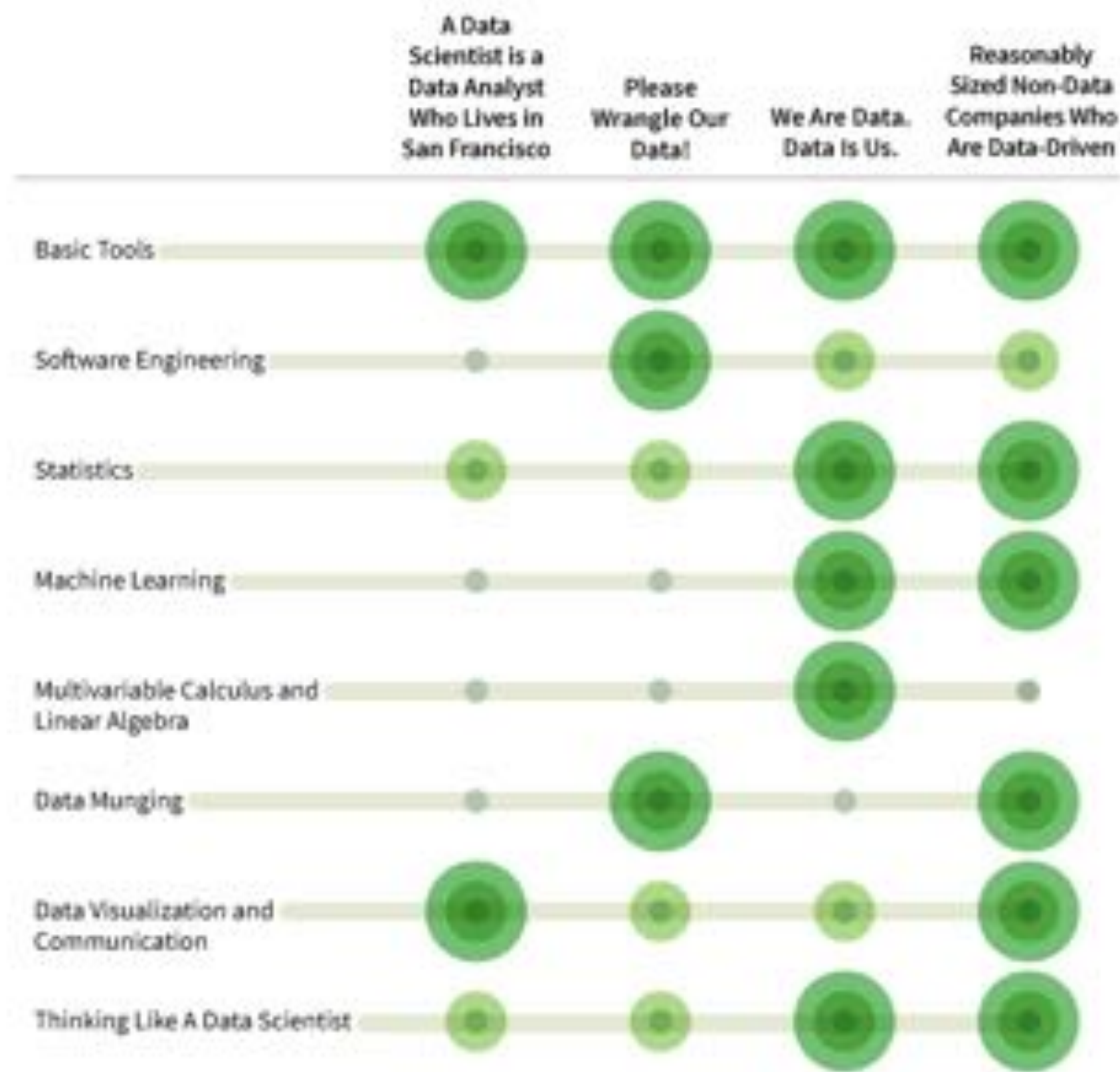
- Narrative Visualisation
- Dashboard Visualisation
- Decision Making Tools
- Automated Decision Tools



PyData Stack

- **Acquire / Refine:** Pandas, BeautifulSoup, Selenium, Requests, SQL Alchemy, Numpy, Blaze
- **Explore:** Matplotlib, Seaborn, Bokeh, Plotly, Vega, Folium
- **Model:** Scikit-Learn, StatsModels, SciPy, Gensim, Keras, Tensor Flow, PySpark
- **Insight:** Django, Flask





Very important



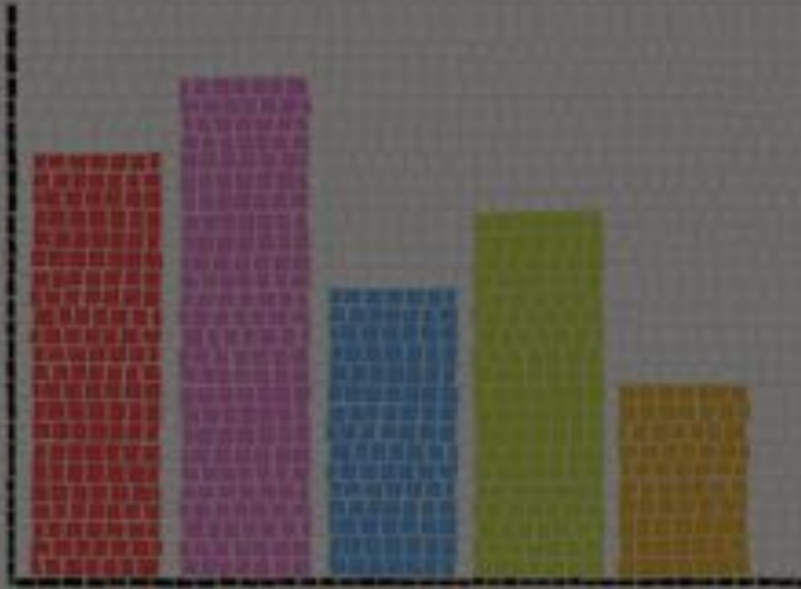
Somewhat important



Not that important

The Art of Data Science

A Guide for Anyone Who Works with Data



Roger D. Peng & Elizabeth Matsui

Data Wrangling with Pandas, NumPy, and IPython

Python for Data Analysis



O'REILLY®

Wes McKinney

O'REILLY®

2nd Edition

Think Stats

EXPLORATORY
DATA ANALYSIS



Allen B. Downey

Books

The Art of Data Science

A Guide for Anyone Who Works with Data



Roger D. Peng & Elizabeth Matsui

Data Wrangling with Pandas, NumPy, and IPython

Python for Data Analysis



O'REILLY*

Wes McKinney

O'REILLY*

2nd Edition

Think Stats

EXPLORATORY
DATA ANALYSIS



Allen B. Downey

Online Course

- Harvard Data Science Course - [CS 109 Course](#)
- Structured in similar way to the approach we shared
- Many more on Coursera & Udacity...

Upcoming Workshops

- Advanced Data Science (Machine Learning, Statistics)
- Data Science at Scale (Spark)
- Visualisation (Multi-Dimensional, Geographic, Large Data)
- Deep Learning (Text, Speech, Image, Video)

A dark, atmospheric background featuring a microphone on a stand in the lower-left foreground. Several bright stage lights are visible in the upper half of the image, creating a bokeh effect with soft, out-of-focus circles of light. The overall tone is professional and focused on the theme of public speaking or presentation.

Speak to Us!

Custom Workshops

Data Science Consulting



Thank you

unnati.xyz/workshop-feedback



has
geek