Beyond the Numbers: Unlocking Insights Through Visualization  
Abstract:

Effective data analysis goes beyond numbers—it requires seeing the underlying patterns that may be obscured in raw data and descriptive statistics. This session will explore the analytical power of visualization, focusing on how to encode data into visuals that bring clarity to complex relationships. We will critique ineffective visualizations, diagnose their flaws, and walk through the steps needed to refine them for more accurate analysis. Even in the age of AI, human visual interpretation remains a crucial tool for identifying anomalies, contextualizing trends, and ensuring that patterns are correctly understood. Attendees will gain practical strategies for transforming data into visuals that enhance, rather than distort, meaningful insights.

# Introduction:

2 minutes

**Reveal: Excel Spreadsheet**

“This... is a dataset.”  
*(Pause)*

“What do you see?

*(Pause)*  
Probably nothing. Just a grid of numbers.  
Overwhelming. Flat. Incompressible.  
And yet, as analysts and data scientists - this is where our work often begins.”  
“This talk is about translating numbers into a form the human eye understands.  
The emergence of patterns and the appearance of structure. Lets apply some formatting to reveal what is hidden here.

**Reveal RGB version**

“Wait a second, oh man, that didn’t help.”

*(Pretend to Struggle)*

**Reveal: Zoom out (a picture becomes visible)**  
its about the moment when the bigger invisible picture becomes obvious.”

**Part 1 (continued): Some Data Is Meant to Be Seen**

**3 minutes.**

**Reveal: Perlin Dataset Grid Display**  
Here’s another dataset.

At a glance, not so different from the one we started with - just another 50-by-50 grid. Every value is here.”

“But this time, our goal isn’t just to recognize a pattern.

We’re aiming higher, we want to extract an **actionable insight**.”

*(Pause)*

“So let me ask you: based on this, what decision can you make?”

*(Pause again - let the discomfort land)*

“What if I told you, these are elevation readings. Is that helpful?”

*(Beat)*

“Even with every value and knowledge of what it represents, we still come up empty .”  
You can’t see slope.  
You can’t see shape.  
You certainly can’t plan a route.”

“So let’s take a step.  
We add a little shading - just grayscale, to represent height.”

**Reveal: Applied Grayscale Gradient**

*(Pause)*

“Now it’s suggestive.  
You start to see hints - maybe hills and valleys.  
But it’s still not usable. You’d still be guessing.”

“But watch what happens when we encode it properly:  
Color. Contours. Elevation lines.”

**Reveal: Topographic Colored Version**

“Same data. Same grid.  
But now it’s no longer abstract - it’s navigable.”

“You can evaluate terrain.  
Spot bottlenecks.  
Pick a safe path from one side to the other.”

“You went from passive recognition to confident navigation.  
That’s the shift visualization enables.”

“Just for fun, I let our robot overlords take a crack at finding the best path.

**Reveal: Machine Generated Path**

“And it wasn’t great.”

“Right here it wants us to jump off a cliff.”

**Reveal: Human Path**

“Meanwhile, most of you could have traced a better route in seconds.  
Because your eyes did what the algorithm couldn’t:  
You *saw* the shape of the problem.”

*(Now wrap the insight with impact)*  
“This is an important role of visualization - not to decorate analysis, but to *drive* it.”

“The numbers were there all along - but they didn’t help us decide.  
Only when we gave them the right **structure** - the right encoding - did the insight appear.”

*(Land on your first major point)*

“That’s the first takeaway today:  
Some data isn’t just easier to see - it **has to** be seen.  
Without the right visual structure, the meaning is invisible.”

“Let me show you one more. A classic - but still wildly relevant.”

**Reveal: Anscombe Raw Data**

This is Anscombe’s Quartet.

This time we have four datasets, and you guessed it, they are x,y coordinates again.  
A quick analysis of these datasets reveals something fun.  
Identical summary statistics.  
Same means. Same Standard Deviation. Regression lines with identical slope and intercepts”

“If all you had were the numbers, you’d assume they were interchangeable.”  
“From a decision-making standpoint, you’d treat them the same.”

**Reveal: Scatter Plots**

“But this… is what they actually look like.”  
*(Pause. Let it land. Gesture across the plots as you speak.)*

“Same metrics. Four completely different situations.”

“Plot one? Clean linear relationship. Great - you might double down on marketing spend.”

“Plot two? Curved - maybe a saturation point, or a logistic growth pattern. You’d need a different model - and a different strategy.”

“Plot three? Almost pure noise, distorted by one massive outlier. That’s not insight - it’s a data quality issue.”

“Plot four? The regression line is trying to fit a vertical stack of identical x-values. That’s not a relationship - it’s a system failure.”

*(Pause slightly)*

“Same summary stats. But four very different courses of action - hiding in plain sight.”

“And without visualizing the data, you'd never know.”

**Part 2: Converting data to visual information**

“Okay—so you’re convinced: visualization reveals what statistics often miss.  
But now comes the practical question:

How do we actually build effective visuals?”

(Pause)

“Because here’s the hard truth: A bad chart doesn’t just fail to help - it actively misleads. The wrong encoding can bury the insight just as much as raw numbers.”

“To avoid that, we need a visual vocabulary -m a consistent way to map meaning onto visuals.

That’s what the Grammar of Graphics gives us:  
A blueprint for turning data into perception.

“Let’s talk about the core concept here: *encoding*.

Encoding just means translating information from one form to another.

In our world, that means converting data - numbers, categories, timestamps - into visual attributes:

* Position
* Color
* Size
* Shape
* Orientation”

(Pause)

“These aren’t just stylistic choices.  
They’re the channels through which meaning flows.”

**Building Blocks of a Visual**

*(Optional: flip to a visual diagram of chart anatomy—axes, scales, marks, encodings.)*

“Every good chart has structure. Let’s break it down.”

**1. The Canvas (Scaffolding)**

“Every chart starts with a canvas - a space to plot.”

“We use axes to define that space:  
How far? In what direction? What does one unit mean?”

“But the canvas isn’t neutral. The moment you define a scale, you’re shaping how someone interprets the data.”

“Here’s an example of how **visual structure shapes interpretation**

[Show both plots side-by-side]

“These two charts use the *exact same data* - a linear trend with some noise.”

“The chart on the **left** uses a *tight Y-axis range* - from 0 to 25. That makes every little fluctuation feel significant.  
This is useful volatility is important to understand.”

(Pause)

“Now look at the **right**. Same trend, but the Y-axis now goes up to 100.”

“A flatter, less dramatic line is produced”

“But that doesn’t make it wrong - it makes it **contextual**. If your goal is to reach 100, this view shows how far you still have to go.  
It de-emphasizes noise and highlights *distance from target*.”

“So which one is better?”

“It depends on your question.  
The axis you choose tells the audience *what matters*.”

“Even before we mark a single point, we’re already making decisions that guide perception.”

“The chart isn’t just about drawing what’s already there.  
We have to decide *how* to shape the data before it hits the page.”

“That’s where **data transformations** come in.”

“A chart is never just ‘showing the data.’  
It’s showing *a version* of the data - filtered, shaped, and framed by your decisions.”

**2. Data Transformations: Framing the View**

“Before we ever plot a single point, we’ve already made choices that shape the story.

Do we show raw values? Percent change? A log scale? Do we normalize across groups?

These are transformations - not of the picture, but of the data itself.”

(Show examples: raw vs log scale, or grouped averages vs raw scatter.)

“Think of it this way: If the canvas is the space for the chart, then the transformation is the lens through which we view it.

A poor transformation distorts the message. A thoughtful one brings clarity.”

“Even something as simple as bin edges in a histogram is a transformation - and it can completely change the impression you get from the same dataset.”

“Same for normalization: if you’re comparing product sales across categories of different sizes, normalizing is what turns a misleading chart into a meaningful one.”

**3: Marks and Encodings - How We Turn Data Into Meaning**

**“Alright - we’ve built the canvas. We’ve shaped the data. Now comes the moment we actually *draw* the chart.”**

“This part is about **marks** and **encodings**.  
They’re the core of every visual.”

**Part 1: Marks - The Building Blocks of Every Chart**

“Marks are the basic geometric elements we use to represent data visually.”

“You’ve seen them before—even if you didn’t think of them as a system.”

* **Points** for scatterplots, dot plots, geographic coordinates
* **Lines** for trends, time series, or continuous flow
* **Bars** for comparing quantities across categories
* **Areas** for showing accumulation or volume
* **Tiles, ribbons, rectangles** — for heatmaps, confidence intervals, density zones

“Every chart is made of marks. They’re what we actually see on the screen or page.”

“But marks alone are just shapes. What gives them meaning - what turns them into insight - is how we encode data into them.”

**Encodings - Mapping Data to Perception**

**## I need to add examples in my jupyter notebook that correspond with each topic in the encoding section. It could be one visual that demonstrates several topics.**

“Visual encodings are how we map information onto marks—using visual properties that our eyes and brains can instantly recognize.”

Let’s walk through the most powerful encodings we have:

**Position - Our Most Accurate Channel**

“Position is the gold standard.  
When you place a mark along an axis - horizontal, vertical, or both - you give it numerical meaning.”

* A dot’s location in a scatterplot
* The length of a bar
* A line’s rise over time

“Use position whenever you want precision. It’s the encoding we interpret fastest and most reliably.”

**Color - Great, But Tricky**

“Color can show category or magnitude - but it’s easy to overdo.”

* **Discrete color** is for categories - like product types or regions
* **Continuous color** is for scales - like temperature or percent change

“Use color to draw attention or distinguish groups.  
But remember: color perception isn’t perfect, and it shouldn’t carry the whole message.”

**Shape - Helpful for Categories**

“Shape is good for categorical differences - especially when color is already doing something else.”

* But there’s a limit - most people can only distinguish a few shapes reliably

“I find it fairly weak when it comes to analysis.”

“I use it most often for adding readability to data stories rather than analysis.”

## example of a pictograph

**Size - Use With Care**

“Size can suggest quantity - but our perception of size, especially area, is nonlinear.”

* Works for emphasis or when combined with position
* Don’t expect your audience to compare circle areas accurately

## add an example where it is tough to distinguish

**Facets – The Power of Small Multiples**

“Sometimes, one chart isn’t enough.”

“Faceting breaks a dataset into small, repeated panels based on a category - like one chart per product or one row per region.”

* It reduces clutter
* It supports comparison
* It often tells a clearer story than one overloaded plot

“Facets let you see patterns emerge across slices. And they encourage better questions.”

“Pair plots are one of my go-to visuals for exploring data.”

# add a pair plot visual. Penguins or Iris?

**Animation & Interaction – Optional, Powerful Tools**

“When you’re working in a digital medium, you can also use **motion** and **interactivity**.”

* **Animation** helps show change over time or smooth transitions --.” Better for story telling, not great for analytics.
* **Interactive charts** let users explore - filter, hover, zoom - use for empowering users exploration.

**Wrap the Insight**

“Here’s the takeaway: **Marks show the data. Encodings give it meaning.**  
The better you match your encoding to the nature of your data, the clearer the insight becomes.”

(Pause)

“Good encodings aren’t just visual. They’re analytical.  
They shape *how people think* about the data.”

**Visual Design Is Analytical Design**

“This is what I mean when I say:  
**Visual design is part of analysis.** It’s not decoration at the end.”

“The way you choose to encode information *guides the questions people ask.*  
A good visual doesn’t just answer questions—it suggests better ones.”

**Transition to Part 3**

“So now you have the grammar.  
You understand what the building blocks are—and how to use them intentionally.”

“Let’s put that into practice by building some visuals together.”

**Part 3: Putting It Into Practice**

“So far, we’ve seen why visualization is essential.  
We’ve seen how numbers alone can mislead—and how visual structure can reveal truth.”

“Now let’s put this into action.”

“Let’s say you’re running an A/B test.  
Group A is your control. Group B is the variant.”

“You collect some results. Sales. Conversions. Bounce rate. Whatever it is.”

**Exercise 1: The Dangerous Lift**

*(Show summary statistics for both groups—mean, std, maybe t-test or p-value.)*

“Here’s what the numbers tell you:  
Group B has a higher mean. The difference is statistically significant.”

*(Pause)*

“So… do we roll it out to everyone?”

*(Gesture to the notebook - prompt a bit of doubt.)*

“Let’s slow down.”

“Here’s what the distributions look like.”

*(Reveal overlaid histograms or KDE plots - Group A vs Group B.)*

“Now we’re not just looking at the average - we’re seeing the shape.”

“Group A has a narrow distribution. Lower mean, but predictable.  
Group B is flatter. Higher mean, but much more spread.”

*(Add a reference line at 0—break-even point.)*

“Here’s the kicker: even though Group B has a higher average,  
**more of its values fall below zero.**”

“Translation?  
More risk. More volatility. Higher chance of losing money—even though the mean looks better.”

**Narrative Bridge**

“Statistically, you were ready to ship.  
But visually? You just saved yourself from a costly mistake.”

“This is why I keep saying:  
**Don’t just ask ‘is it better?’—ask ‘how is it better?’**  
Visualization gives you that answer.”

**Exercise 2: Reverse the Trap**

“Let’s flip it. What if Group B had a *lower* mean,  
but almost *none* of its values were below break-even?”

*(Show another KDE with narrow, high-confidence bump near the top.)*

“Now it looks like a more conservative choice—lower upside, but far less downside.”

“If you’re launching a new product, or pricing a risky offer,  
this might actually be the better option.”

**Wrap the Section**

“These are the kinds of decisions that visuals enable.”

“When you encode distributions visually, you stop making blind decisions based on averages,  
and start making smart decisions based on shape.”

“That’s the power of encoding:  
We’re not visualizing to decorate—we’re visualizing to evaluate.”

## Exercise 2:

Wolves: