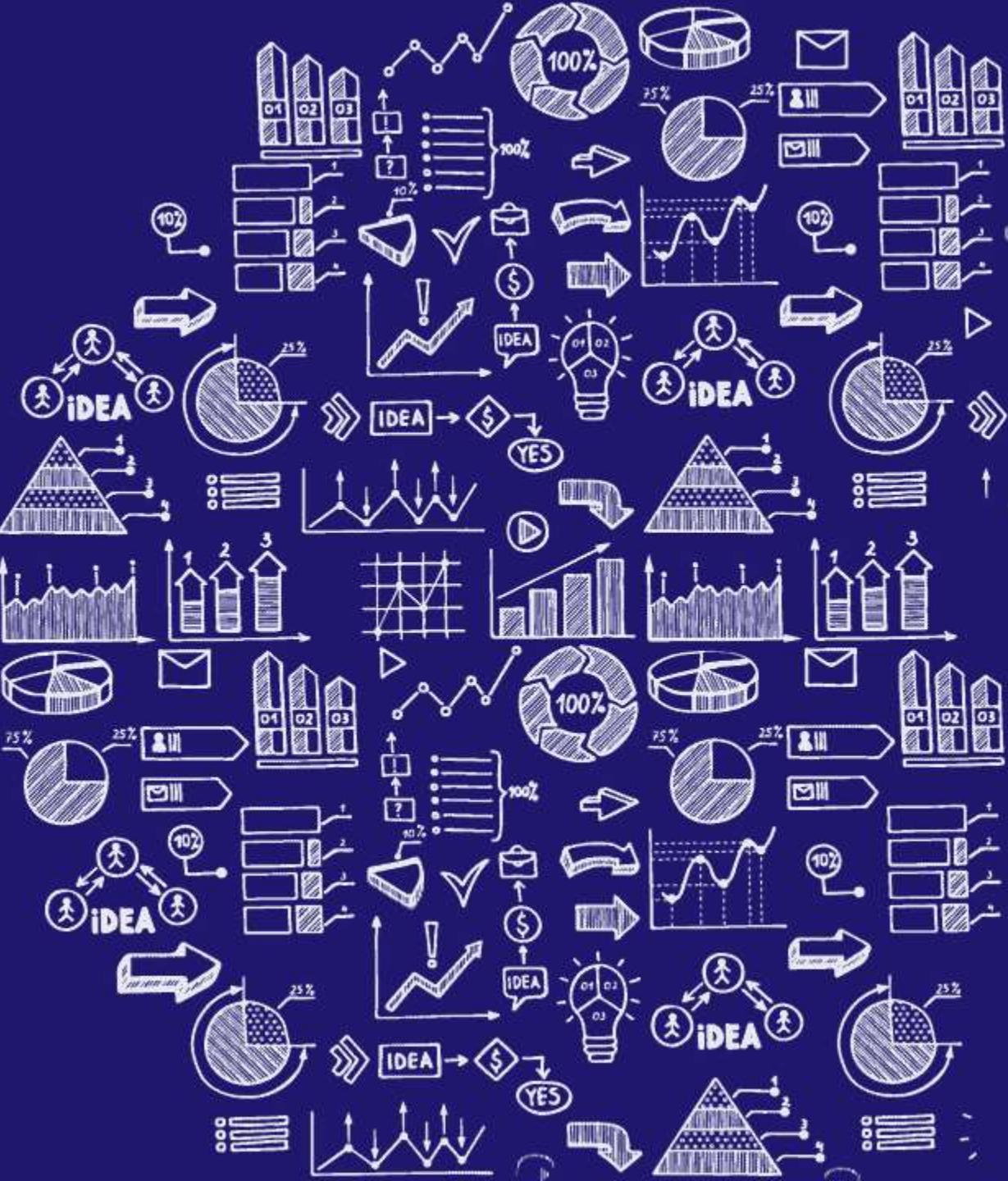


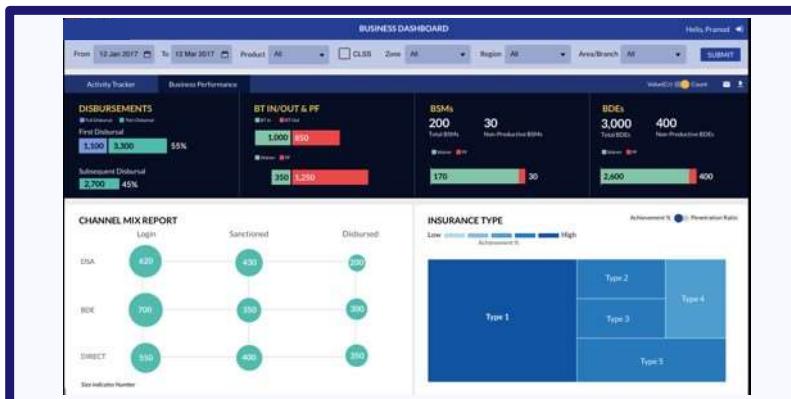
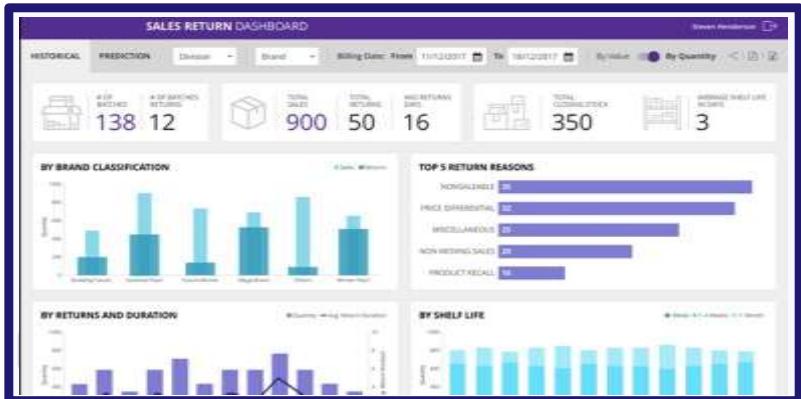
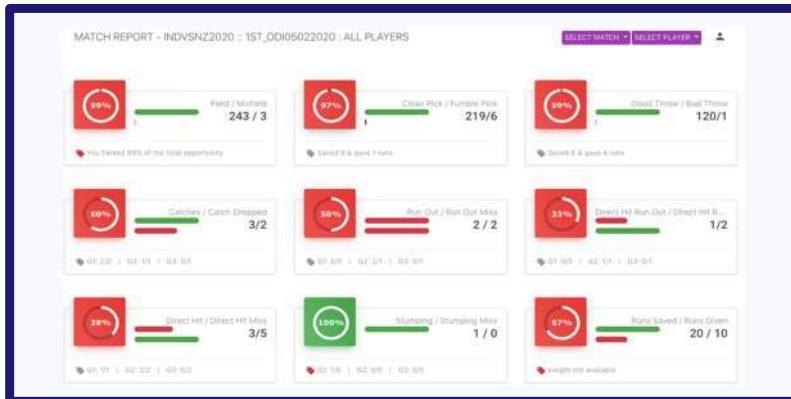
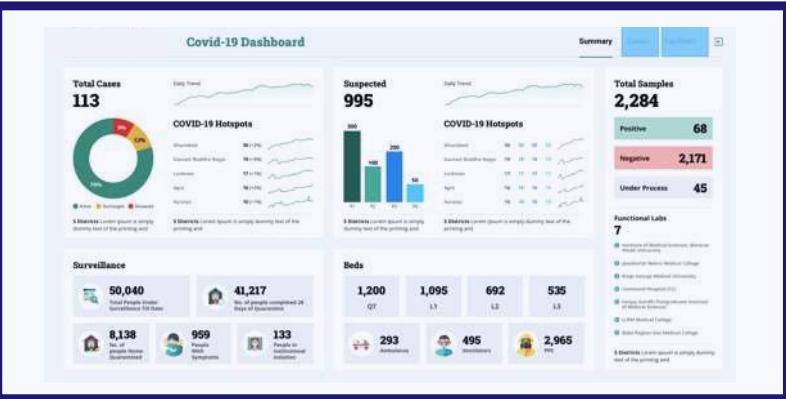
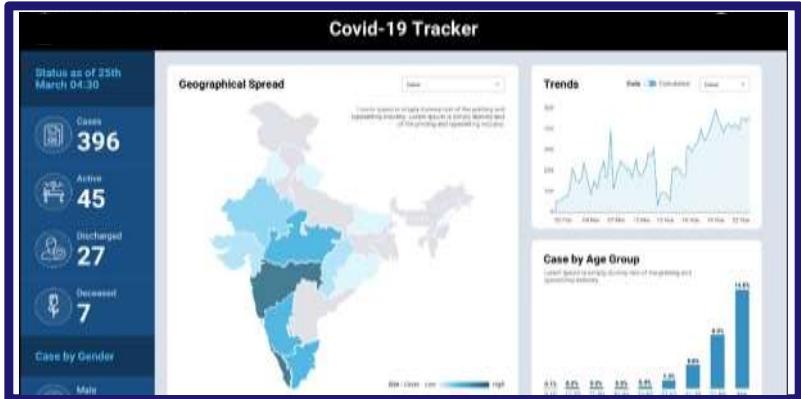
Introduction to  
**Data Storytelling**

**Rasagy Shamma**

Principal Information Designer,  
Gramener



# At Gramener, we narrate business insights as stories



“Can’t I just show the data?”  
“Why do I need charts?”

# How many numbers are **above** 100?

23	32	71	72	58	87	11	77	70	16
17	21	56	44	68	51	84	20	60	40
37	8	107	14	12	41	69	14	18	71
62	55	59	64	33	55	71	58	103	92
101	56	45	34	43	15	73	78	6	93
39	53	22	26	26	94	60	82	99	74
11	12	36	67	70	71	97	59	73	99
75	74	69	69	51	48	2	66	92	98
15	10	41	58	104	94	92	84	74	82
12	52	10	57	33	77	88	81	81	91
15	56	25	30	21	7	66	66	78	87
29	23	5	34	11	96	74	99	99	88
37	10	43	15	50	71	65	60	101	98
46	34	19	102	57	70	95	84	63	91
3	34	39	37	60	81	65	63	9	71
48	46	25	50	22	64	91	76	71	79

# How many numbers are **below** 10?

23	32	71	72	58	87	11	77	70	16
17	21	56	44	68	51	84	20	60	40
37	8	107	14	12	41	69	14	18	71
62	55	59	64	33	55	71	58	103	92
101	56	45	34	43	15	73	78	6	93
39	53	22	26	26	94	60	82	99	74
11	12	36	67	70	71	97	59	73	99
75	74	69	69	51	48	2	66	92	98
15	10	41	58	104	94	92	84	74	82
12	52	10	57	33	77	88	81	81	91
15	56	25	30	21	7	66	66	78	87
29	23	5	34	11	96	74	99	99	88
37	10	43	15	50	71	65	60	101	98
46	34	19	102	57	70	95	84	63	91
3	34	39	37	60	81	65	63	9	71
48	46	25	50	22	64	91	76	71	79

# Which quadrant has the **highest total**?

23	32	71	72	58	87	11	77	70	16
17	21	56	44	68	51	84	20	60	40
37	8	107	14	12	41	69	14	18	71
62	55	59	64	33	55	71	58	103	92
101	56	45	34	43	15	73	78	6	93
39	53	22	26	26	94	60	82	99	74
11	12	36	67	70	71	97	59	73	99
75	74	69	69	51	48	2	66	92	98
15	10	41	58	104	94	92	84	74	82
12	52	10	57	33	77	88	81	81	91
15	56	25	30	21	7	66	66	78	87
29	23	5	34	11	96	74	99	99	88
37	10	43	15	50	71	65	60	101	98
46	34	19	102	57	70	95	84	63	91
3	34	39	37	60	81	65	63	9	71
48	46	25	50	22	64	91	76	71	79

Answer them again,  
***with design changes..***

# How many numbers are **above** 100?

23	32	71	72	58	87	11	77	70	16
17	21	56	44	68	51	84	20	60	40
37	8	107	14	12	41	69	14	18	71
62	55	59	64	33	55	71	58	103	92
101	56	45	34	43	15	73	78	6	93
39	53	22	26	26	94	60	82	99	74
11	12	36	67	70	71	97	59	73	99
75	74	69	69	51	48	2	66	92	98
15	10	41	58	104	94	92	84	74	82
12	52	10	57	33	77	88	81	81	91
15	56	25	30	21	7	66	66	78	87
29	23	5	34	11	96	74	99	99	88
37	10	43	15	50	71	65	60	101	98
46	34	19	102	57	70	95	84	63	91
3	34	39	37	60	81	65	63	9	71

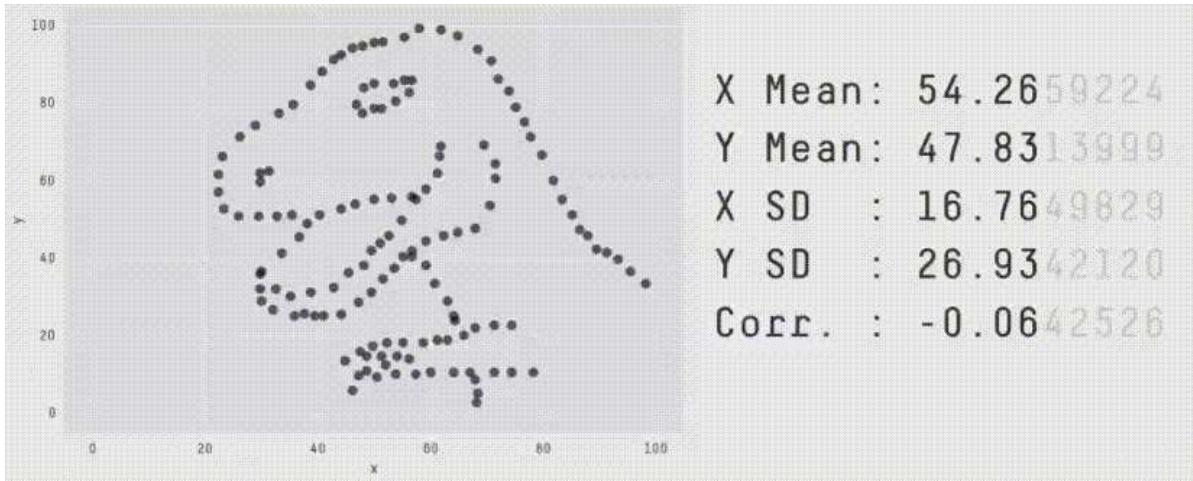
# How many numbers are **below** 10?

23	32	71	72	58	87	11	77	70	16
17	21	56	44	68	51	84	20	60	40
37	8	107	14	12	41	69	14	18	71
62	55	59	64	33	55	71	58	103	92
101	56	45	34	43	15	73	78	6	93
39	53	22	26	26	94	60	82	99	74
11	12	36	67	70	71	97	59	73	99
75	74	69	69	51	48	2	66	92	98
15	10	41	58	104	94	92	84	74	82
12	52	10	57	33	77	88	81	81	91
15	56	25	30	21	7	66	66	78	87
29	23	5	34	11	96	74	99	99	88
37	10	43	15	50	71	65	60	101	98
46	34	19	102	57	70	95	84	63	91
3	34	39	37	60	81	65	63	9	71

# Which quadrant has the **highest total**?

23	32	71	58	87	11	77	16
		72				70	
17	21	56	44	68	51	84	20
37	8	107	14		41	69	14
			12				18
62		59	33	55	71	58	103
55		6					92
101		45	34	43	15	73	78
39	53	22	26	26	94	82	99
					60		74
11		36	67		71	97	59
12			70				73
75		69	69	51	48	2	66
							92
							98

# Visually representing data helps us to see patterns in the data quickly



“The greatest value of a picture is when it forces us to notice what we never expected to see.”

— John Tukey

# Stories have a huge impact on humans



## Storytelling has a 30X Return on Investment

Rob Walker and Joshua Glenn [auctioned](#) common items like mugs, golf balls, toys, etc. The item descriptions were **stories** purpose-written by 200+ contributing writers.

Items that were bought for \$250 sold for over \$8,000 – a return of over 3,000% for storytelling!

## Stories are memorable and viral

People remember stories. They'll act on them. People share stories. That enables collective action.

We analyze data to improve people's decision making. For this to be effective, data stories are needed more than ever before.

# **Visual data storytelling is a critical skill for data scientists, analysts & managers**

## **But analysts present their work, not their message**

Data scientists present their analysis – what they did, and what they found. That's not what the audience needs.

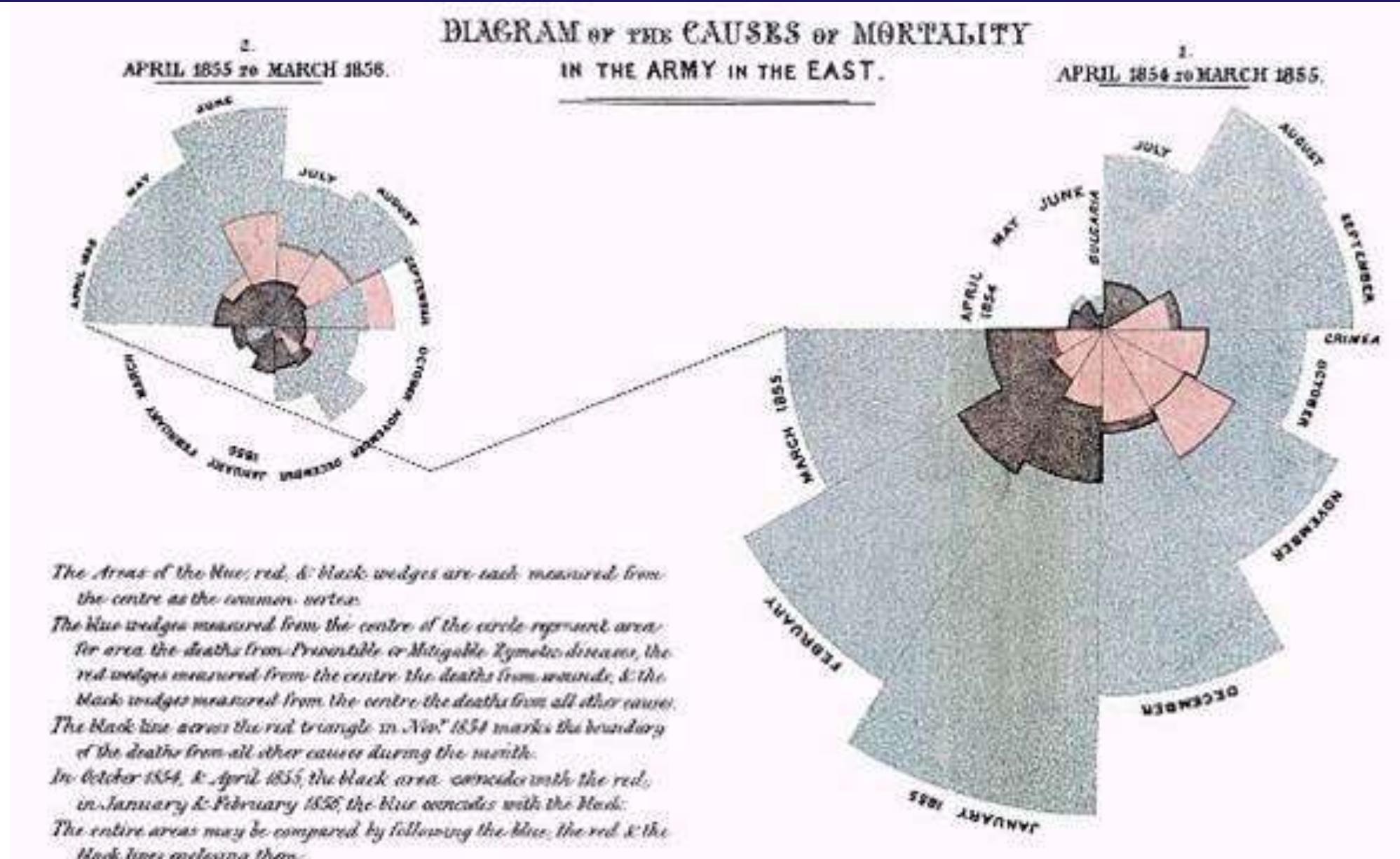
Audiences need a message that tells them what to do, and why. Told in an engaging way. As a story.

## **Share your data & analysis as data stories**

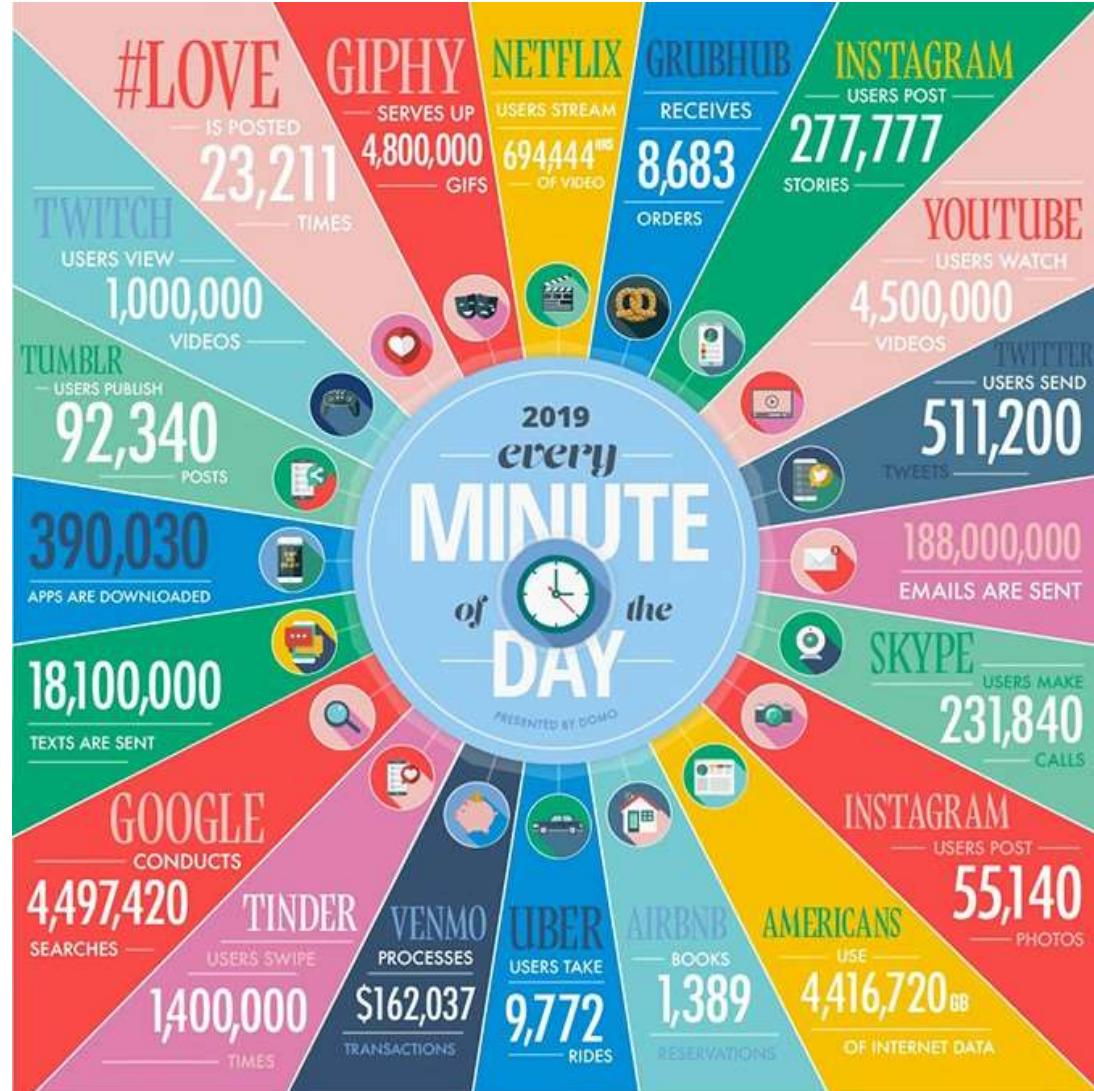
Whenever you share inferences from data – whether it's as a presentation, or an email or document with your analysis, or as a dashboard – craft it as a story.

This session will give you a glimpse of some of the data stories we've created at Gramener, and how you can make these yourself.

# We've been telling stories with data for a long time...



...but the overload of data in today's age makes this critical

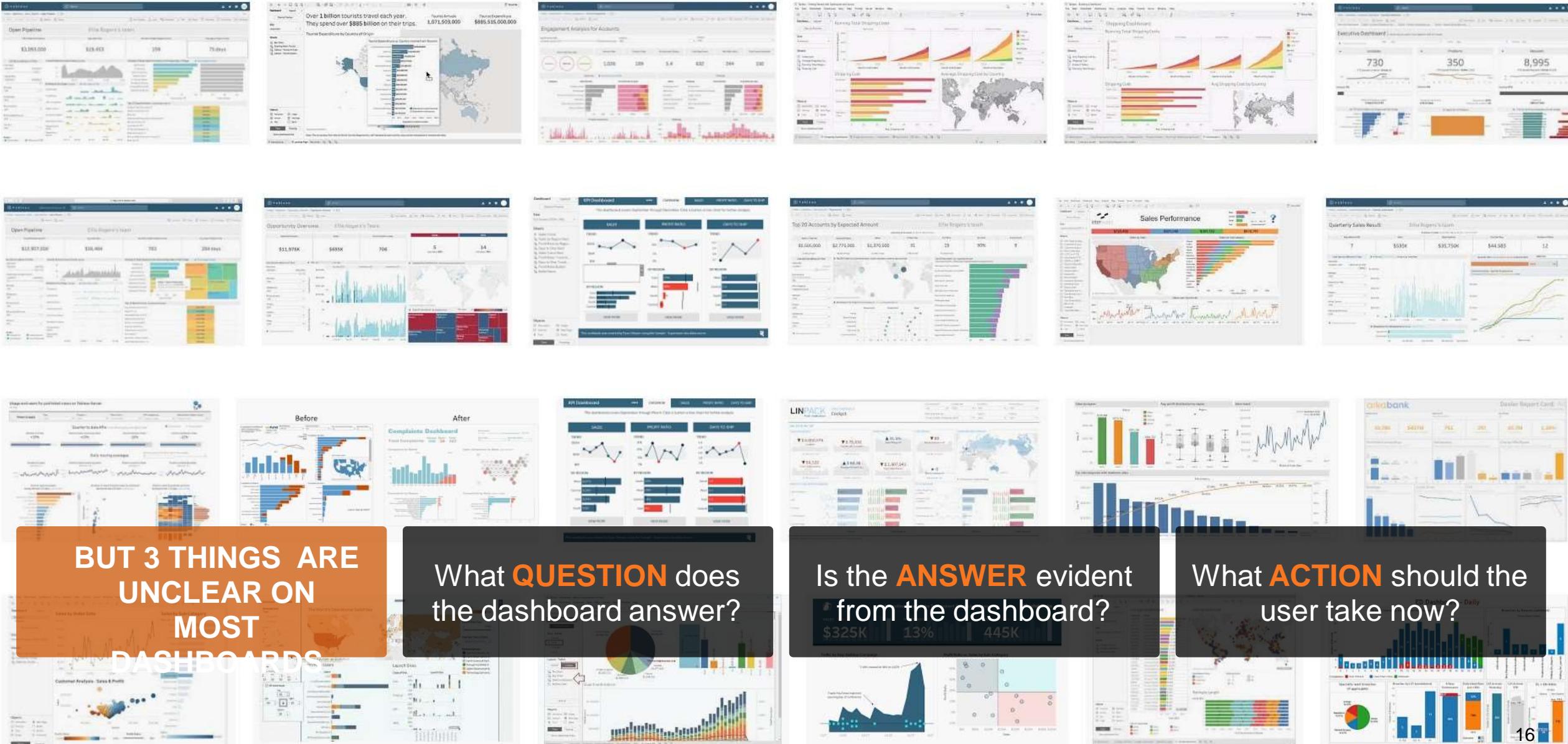


“Every second of every day, our senses bring in **way too much data** than we can possibly process in our brains.”

– Peter Diamandis

Data Storytelling helps make sense of this data

# With the growth of self-service BI, 85% of companies have *lost track* of how many dashboards they generated

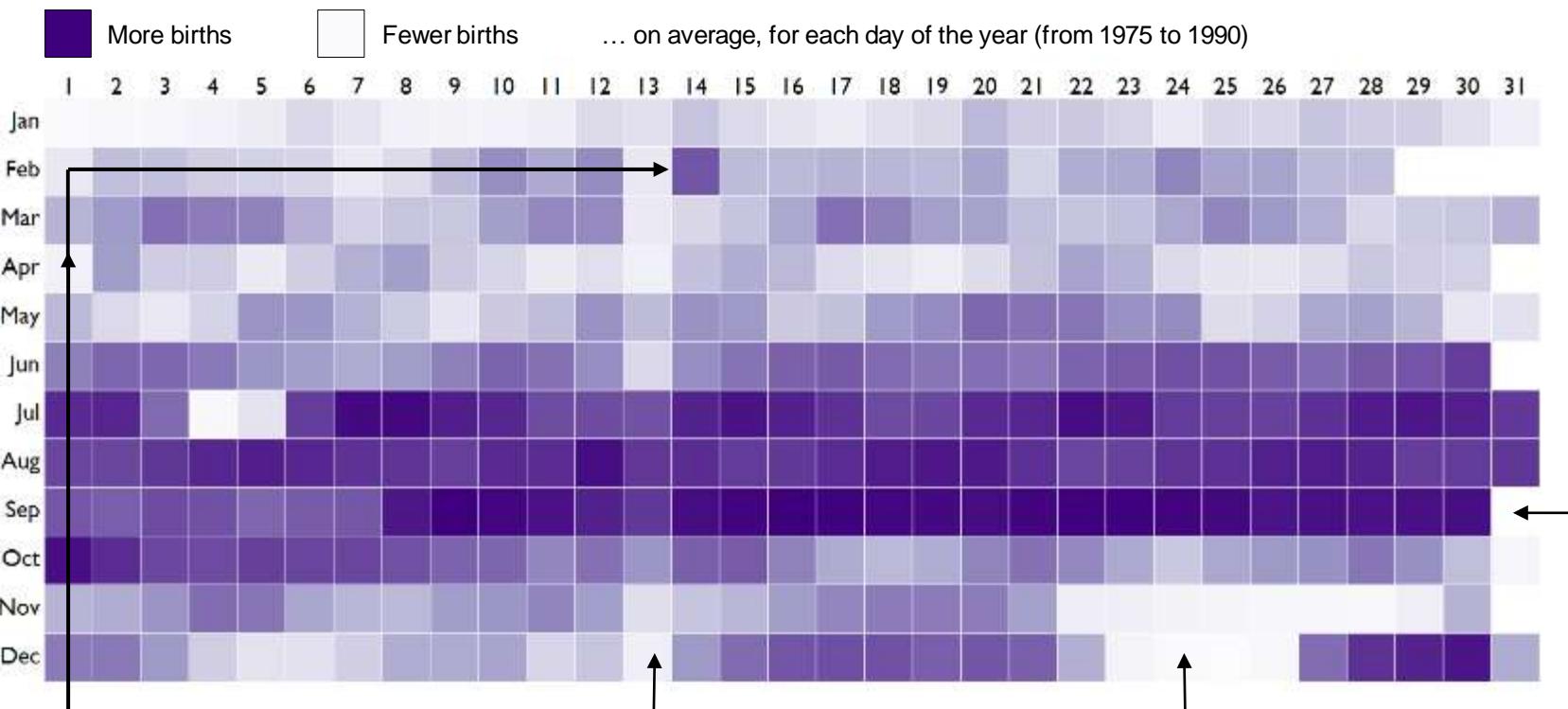


# Let's look at 15 years of US Birth Data

This is a dataset (1975 – 1990) that has been around for several years and has been studied extensively. Yet, a visualization can reveal patterns that are neither obvious nor well known.

For example,

- Are birthdays uniformly distributed?
- Do doctors or parents exercise the C-section option to move dates?
- Is there any day of the month that has unusually high or low births?
- Are there any months with relatively high or low births?



Some special days like April Fool's day are avoided, but Valentine's Day is quite popular

Most people prefer not to have children on the 13<sup>th</sup> of any month, given that it's an unlucky day

Relatively few births during the Christmas and Thanksgiving holidays, as well as New Year and Independence Day.

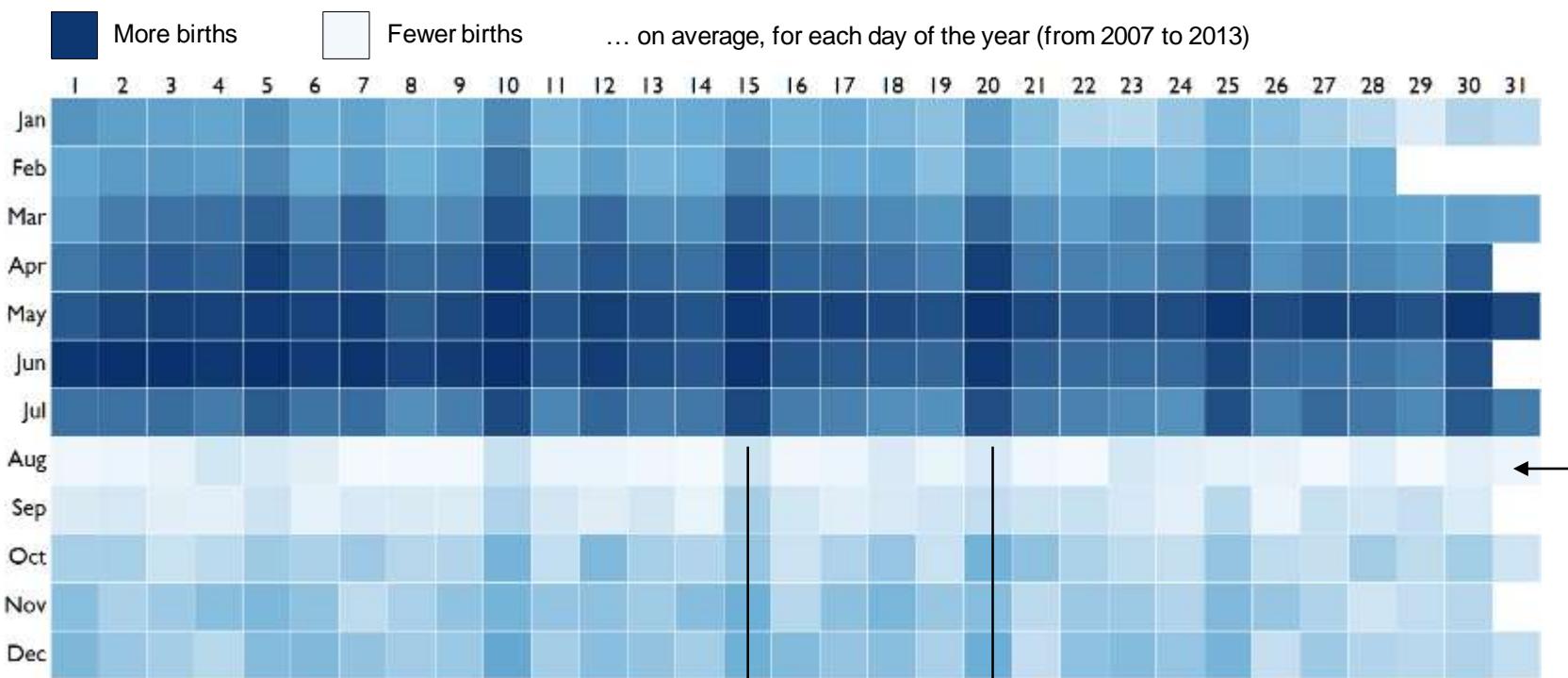
Very high births in September. But this is fairly well known. Most conceptions happen during the winter holiday season

# The pattern in India is quite different

This is a birth date dataset that's obtained from school admission data for over 10 million children. When we compare this with births in the US, we see none of the same patterns.

For example,

- Is there an aversion to the 13<sup>th</sup> or is there a local cultural nuance?
- Are holidays avoided for births?
- Which months have a higher propensity for births, and why?
- Are there any patterns not found in the US data?



Such round numbered patterns are typical indication of fraud. Here, birthdates are *brought forward to aid early school admission*

We see a large number of children born on the 5<sup>th</sup>, 10<sup>th</sup>, 15<sup>th</sup>, 20<sup>th</sup> and 25<sup>th</sup> of each month – that is, round numbered dates

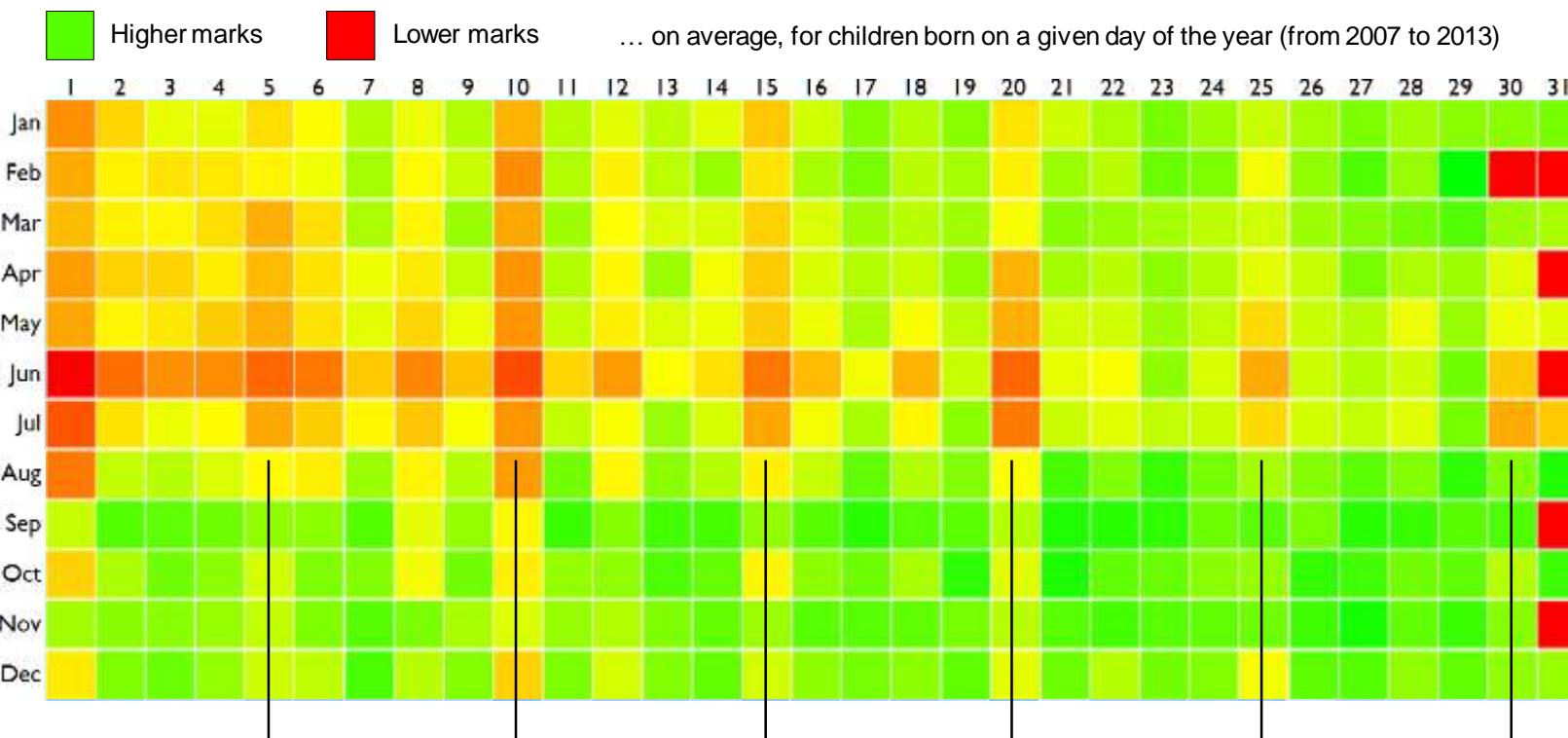
Very few children are born in the month of August, and thereafter. Most births are concentrated in the first half of the year

# This adversely impacts children's marks

It's a well-established fact that older children tend to do better at school in most activities. Since many children have had their birth dates brought forward, these younger children suffer.

The average marks of children "born" on the 1<sup>st</sup>, 5<sup>th</sup>, 10<sup>th</sup>, 15<sup>th</sup> etc.. of the month tend to score lower marks.

- Are holidays avoided for births?
- Which months have a higher propensity for births, and why?
- Are there any patterns not found in the US data?

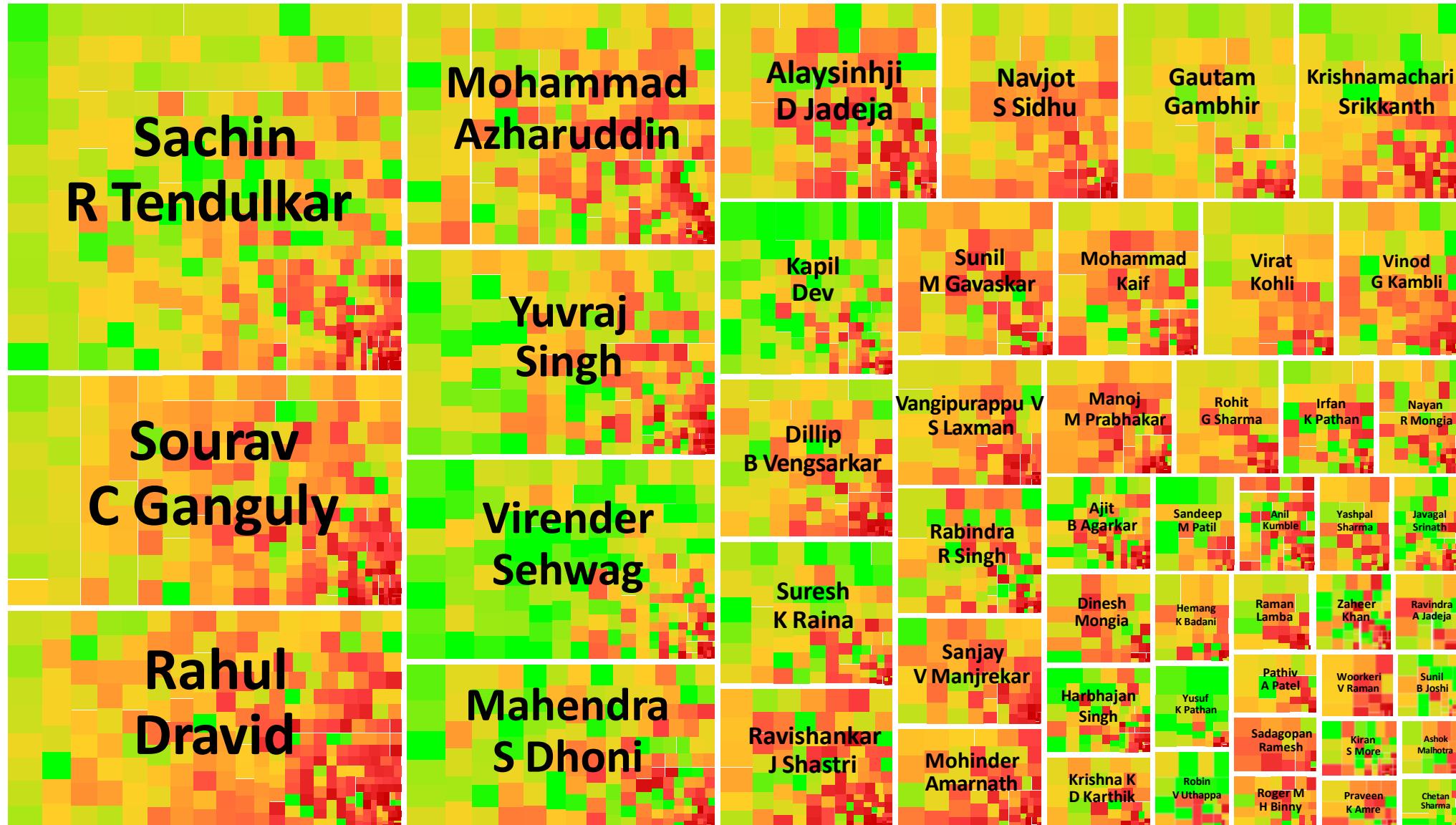


Children "born" on round numbered days score lower marks on average,  
due to a higher proportion of younger children

# Who is the \*best\* Batsman?

Sachin R Tendulkar	Mohammad Azharuddin	Alaysinhji D Jadeja	Navjot S Sidhu	Gautam Gambhir	Krishnamachari Srikkanth
	Yuvraj Singh	Kapil Dev	Sunil M Gavaskar	Mohammad Kaif	Virat Kohli
Sourav C Ganguly	Virender Sehwag	Dillip B Vengsarkar	Vangipurappu V S Laxman	Manoj M Prabhakar	Rohit G Sharma
Rahul Dravid	Mahendra S Dhoni	Suresh K Raina	Rabindra R Singh	Ajit B Agarkar	Sandeep M Patil
		Ravishankar J Shastri	Sanjay V Manjrekar	Dinesh Mongia	Hemang K Patel
			Harbhajan Singh	Yuvraj S Pathan	Pathiv A Patel
			Mohinder Amarnath	Krishna K D Karthik	Wasim V Patel
				Ravi Y Umrasa	Sadagopan Ramesh
					Rajat S Misra
					Arjun Mathur
					Roger M Hony
					Prajeet Kama
					Chetan Chauhan

# Seeing the best & the worst of the best



# World Bank used data stories to clarify impact of technology on innovation

The World Bank approached us to help communicating data stories from their economic development indicator data. Specifically: which countries have similar levels of innovation? **Does technology drive innovation?**

Gramener collated the diverse datasets and clustered the countries based on similarity of economic indicators.

We then ran a series of visual analytics that showed the impact of one on another – annotated with narrative explanations.

We discovered that innovation is enabled by access to latest technology and reliance on professional management. But it does not align with appetite for entrepreneurship in high income countries.

This interactive is featured on the World Bank website.

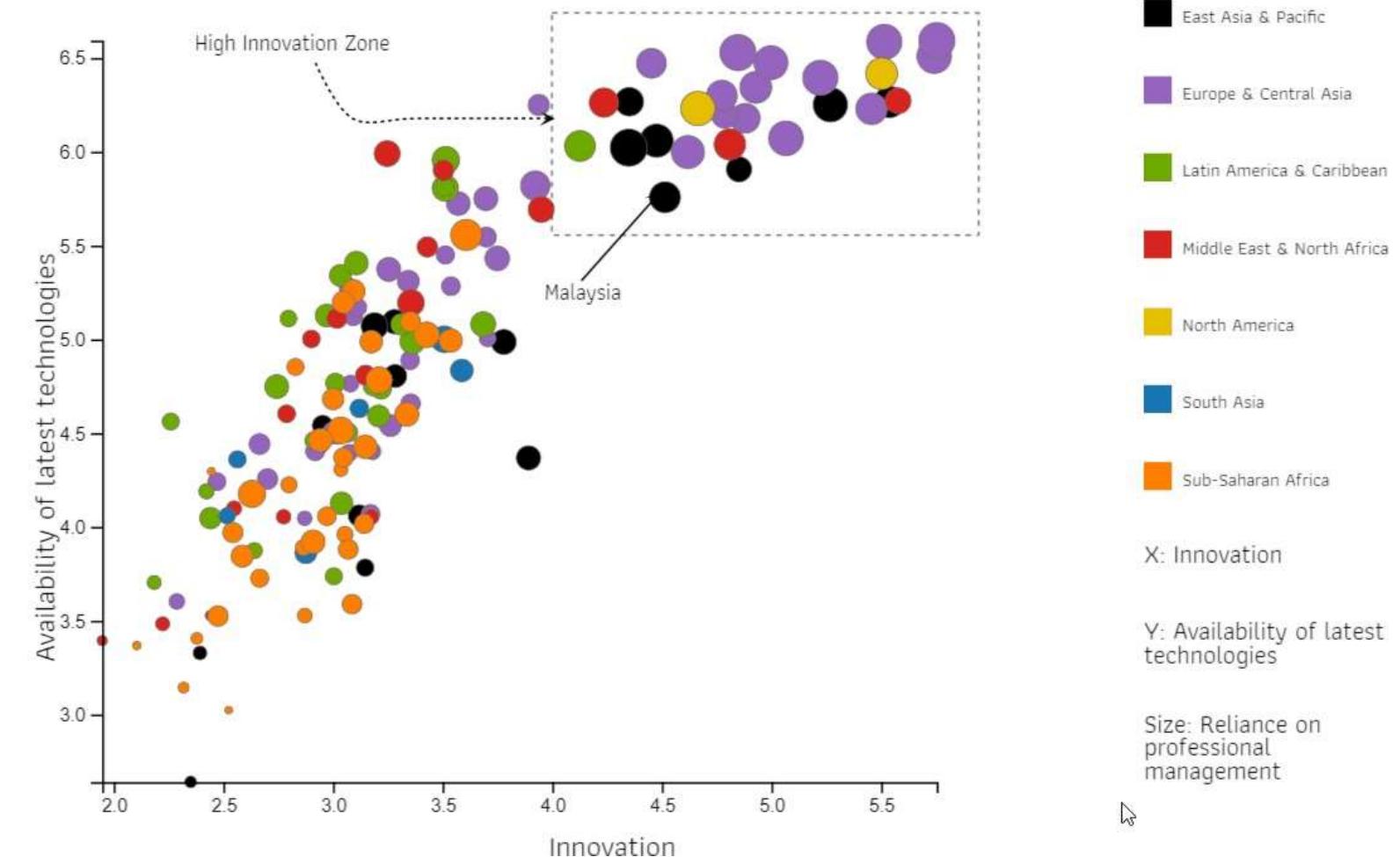
**1.3 m**

viewers read this interactive data story (as of Mar 2020)

**75%**

more people concluded the when shown the data story

Innovation vs Availability of latest technologies



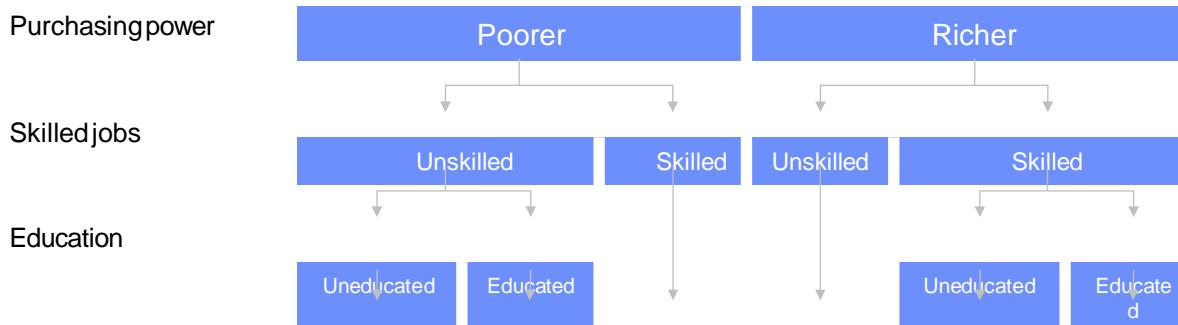
# Data stories can help communicate the data science process

Previously, the client was treating contiguous regions as a homogenous entity, from a channel content perspective.

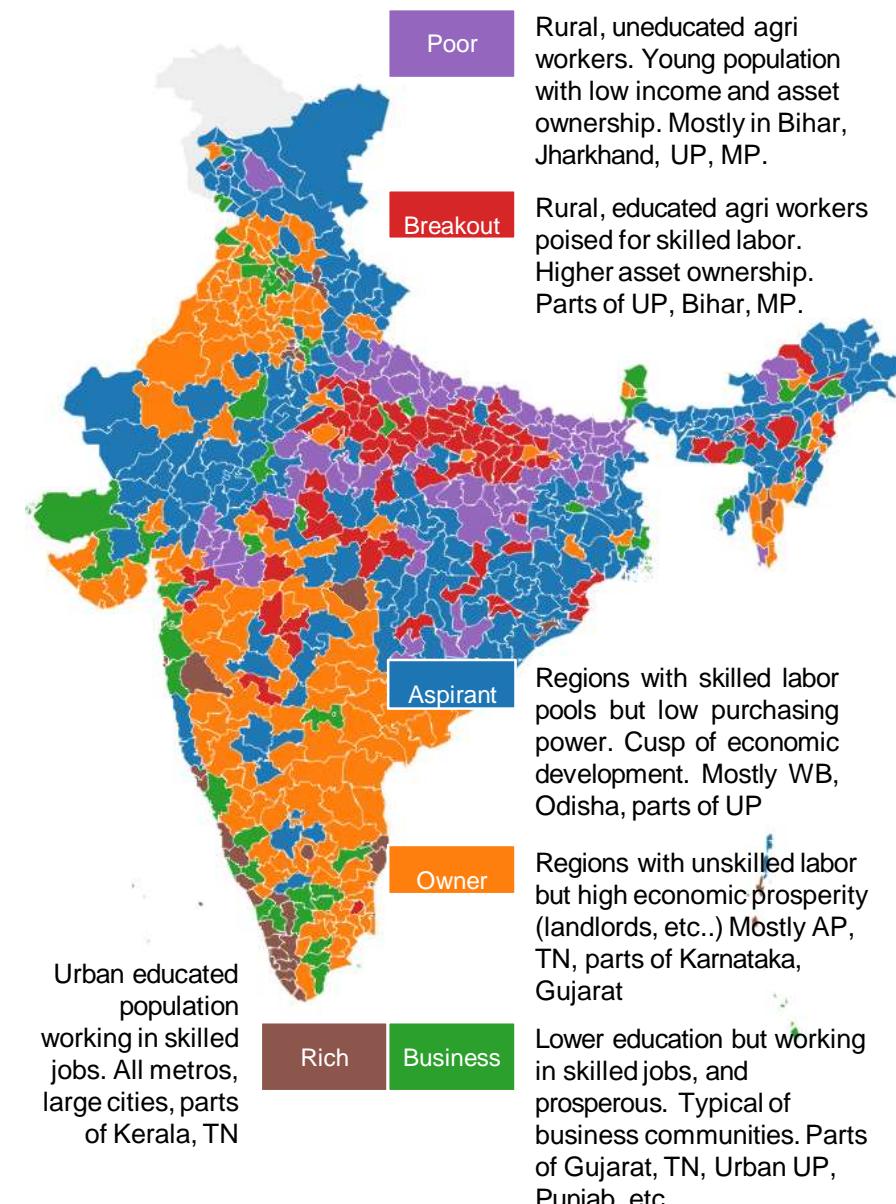
To deliver targeted content, we divided India into 6 clusters based on their demographic behavior. Specifically, three composite indices were created based on the economic development lifecycle:

- **Education** (literacy, higher education) that leads to...
- **Skilled jobs** (in mfg. or services) that leads to...
- **Purchasing power** (higher income, asset ownership)

Districts were divided (at the average cut-off) by:



The 6 clusters are

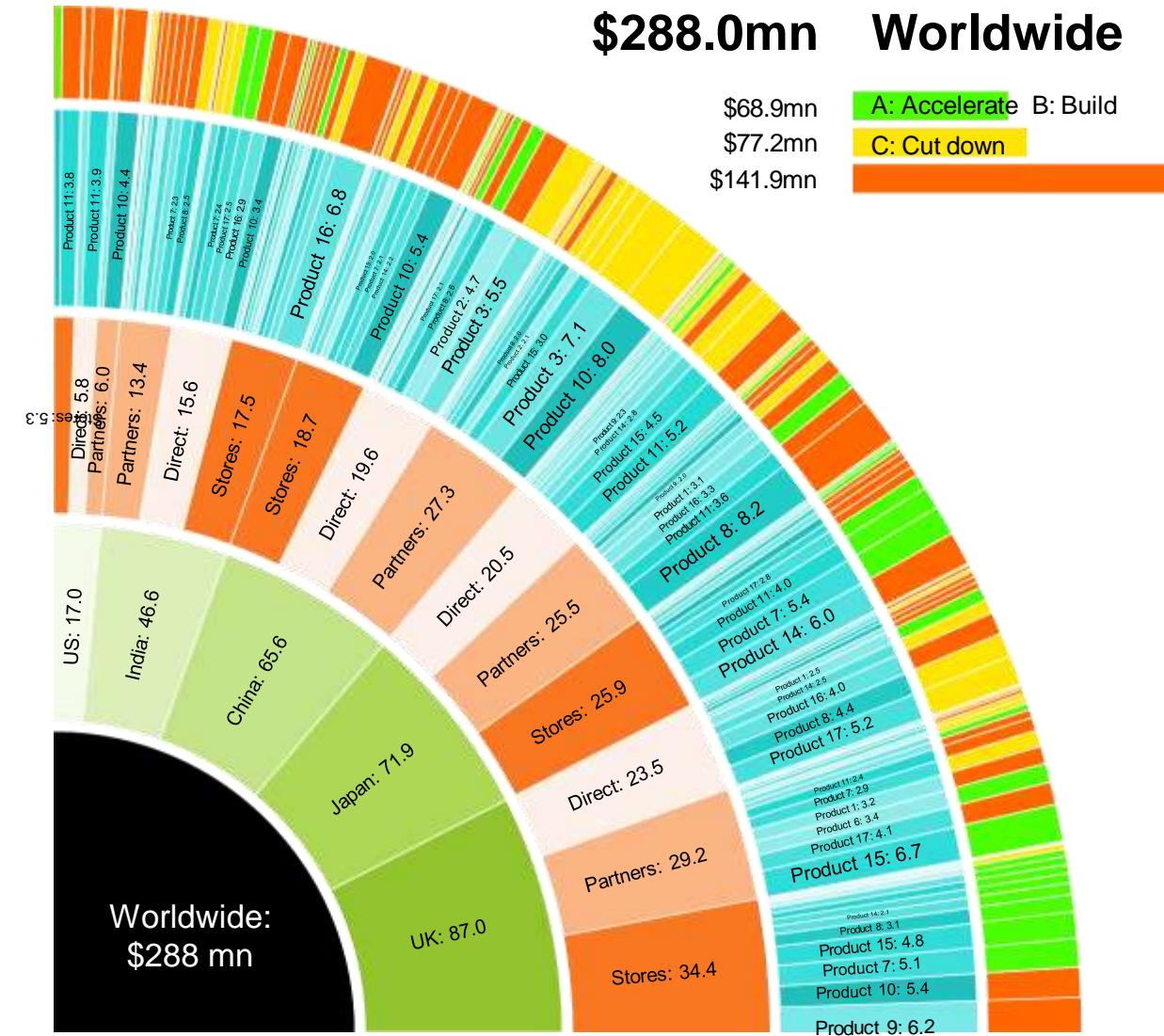


## Data stories can lead to higher engagement & answer new questions

A large FMCG organization wanted to create a visualization to review sales performance across geographies, channels and products for their board meetings.

Gramener built an interactive slide deck that allowed users to drill-down within powerpoint.

Dynamic presentations led to a complete revamp of the entire structure of board presentations.

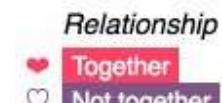


# Data stories can show you what you can't see

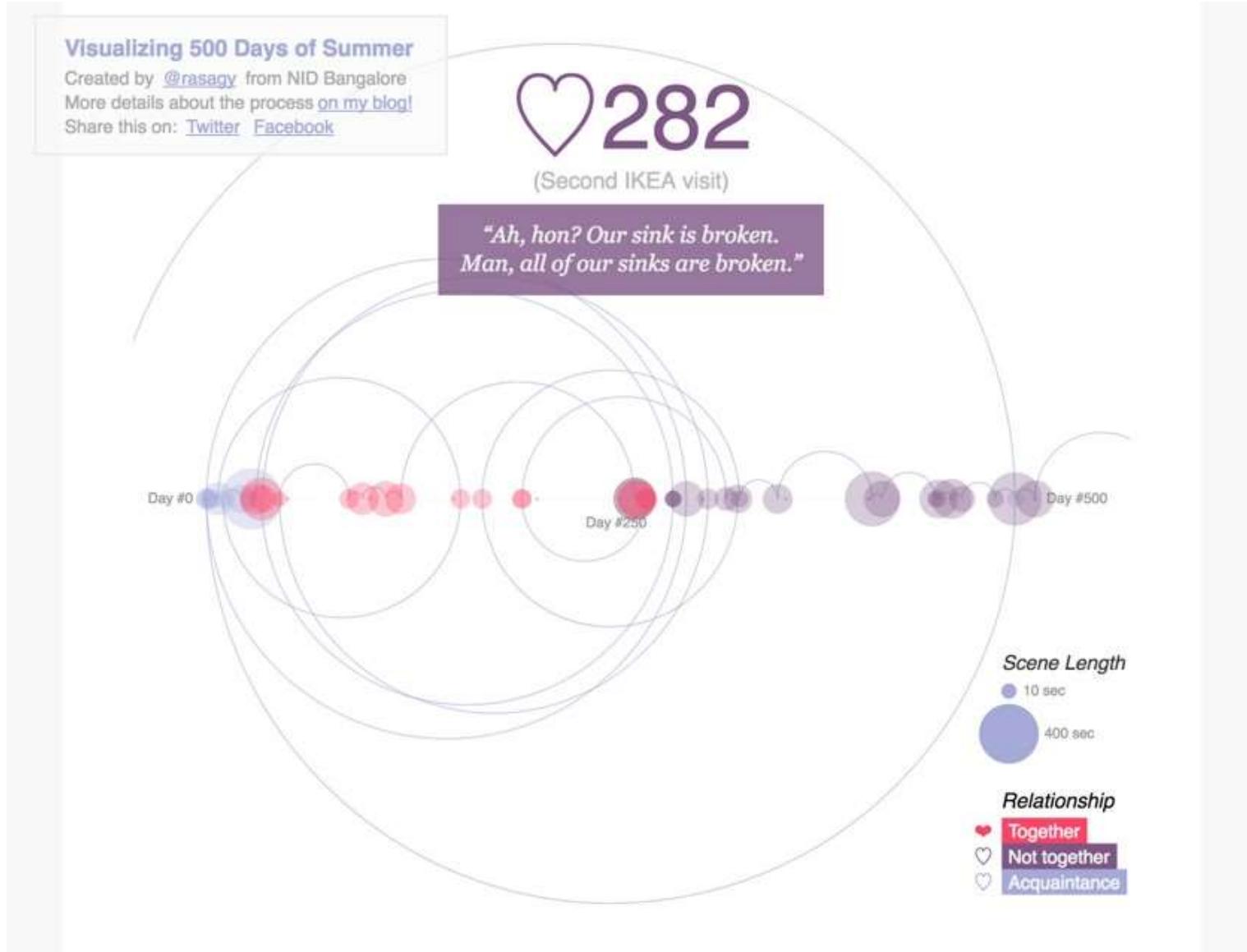


The movie starts with the Day #488, with Tom & Summer at the park...

Day #0      Day #250      Day #500



# Data stories can show you what you can't see

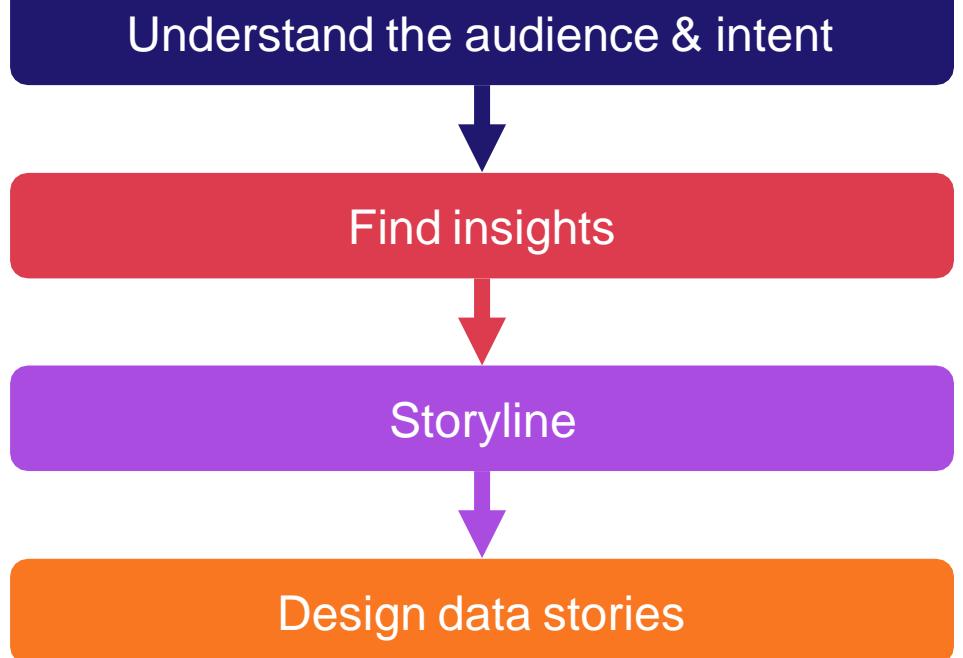


# Data stories can spark questions that you may have never asked



**“So how can I make data stories?”**

You have data. You  
have analysis.  
**Now, narrate your story.**



# We use these steps to go from data to a data story:

Who is your audience? They determine the story

What is their problem? That defines your analysis

Find the right analysis to solve the problem Filter for big, useful,  
surprising insights

Start with the takeaway. Summarize your entire story Add  
supporting analyses as a tree

Pick a format based on how your audience will consume the story

Pick a visual design based on the takeaway

Annotate to explain & engage. Use four types of narratives

# Who is your audience? They determine the story

**Different people want different things from the same data.**

Given sales data:

- The Board: “Predict next quarter’s sales”
- Product head: “Which product grew the most?”
- Sales head: “Did we meet our target?”

They are *not interested* in each others’ questions.

**DO IT: Who is the audience for your analysis?**

**Role:** \_\_\_\_\_

Be specific. “Head of sales”, not “executive”

**Example name:** \_\_\_\_\_

Name a real person. “Jim Fry”, not “any sales head”.

# What is their problem? That defines your analysis

For each person, answer the following questions:

1. What's their situation?
2. What problems do they face?
3. What action can they take?
4. What is the impact of this action?

## DO IT: Write it in this structure

"[Person, Role] is in [situation], and faces this [problem]. By taking [action], she can drive [impact]."

## Example

John, the Marketing head,  
must create a region-wise budget,  
and doesn't know the region-wise ROI.  
By prioritizing the region, she can  
maximize ROI.

person, role  
situation  
problem  
action  
impact

# Here are three examples in real life

	Purchasing Commodities	Cargo Delay	Customer Churn
Person, Role	Adam, the purchasing head of a leading European brewery	Cris, the operations head of a leading US airline	Ravi, the marketing manager of an Asian telecom company
Situation	had plants that purchased commodities from several vendors. Discounts were low. Number of weekly orders were high.	had an SLA to deliver cargo from the flight to the warehouse in under 1.5 hours – 15% lower than their current best performance.	Found that the cost of replacing customers was <i>thrice</i> the cost of retention.
Problem	But he didn't know which plants and commodities were a problem. Every plant denied it.	But she didn't know what were the biggest drivers of this delay – people, assets, or type of cargo.	But he didn't know which customers to make offers to in order to retain them.
Action	By consolidating vendors and reducing order frequency,	By adding resources only to the largest levers of delay,	By predicting which customer was likely to churn,
Impact	they could increase their discounts and reduce logistics cost.	she could reduce turnaround time with the lowest spend.	they could tailor a retention offer and reduce re-acquisition cost.

# Filter for big, useful, surprising insights

IS THE INSIGHT

**BIG**

IS THE INSIGHT

**USEFUL**

IS THE INSIGHT

**SURPRISING**

We want a result that **substantially** changes the outcome.

Can they take an **action** that improves their objective?  
What should they do next?

Is it **non-obvious**?  
Does it overturn an existing belief, or bring consensus?

## DO IT: Rate each analysis against B.U.S.

### Example

There are twice as many restaurants in NYC than any other city

Sales increased in every region except our largest branch, which dipped by 0.1%

Increase in rainfall increases the sale of umbrellas, and is the biggest driver of our sales

B	U	S
✓	✓	✗
✗	✓	✓
✓	✗	✗

Filter the analyses using this checklist

# Here are the analyses & filters for the problems we saw earlier

Purchasing Commodities S	B U	Cargo Delay S	B U	Customer Churn S	B U
The most common commodity was ordered 10 times a week across 2.4 vendors		Fragile cargo is a big factor in the delay, with a 20% impact		Number of inbound calls does not impact churn.	
The number of orders is correlated with the number of vendors. Reducing one will reduce the other		Fridays are when cargo is delayed the most		Customers who haven't made any calls in the last 15 days are the most likely to churn	
Plant P126 was the plant with the most violations, especially on largest commodity		Trained staff and forklifts impact delay the most		Customers making infrequent calls, recharging small amounts infrequently, are most at risk	

# Here are the analyses & filters for the problems we saw earlier

Purchasing Commodities	B U S	Cargo Delay	B U S	Customer Churn	B U
The most common commodity was ordered 10 times a week across 2.4 vendors		Fragile cargo is a big factor in the delay, with a 20% impact	B S	Number of inbound calls does not impact churn.	
The number of orders is correlated with the number of vendors. Reducing one will reduce the other	U	Fridays are when cargo is delayed the most		Customers who haven't made any calls in the last 15 days are the most likely to churn	B
Plant P126 was the plant with the most violations, especially on largest commodity	B U	Trained staff and forklifts impact delay the most	B U S	Customers making infrequent calls, recharging small amounts infrequently, are most at risk	B U

# Start with the takeaway. Summarize your entire story

**Close your eyes.** Think of a childhood tale. Summarize the moral of the story in one line

We easily remember these stories *and* their summary as a moral several years later.

**Close your eyes.** Think of a business presentation from last week. Can you easily summarize the message in one line?

Stories are *designed* around a moral. A single takeaway. An “elevator pitch”

## DO IT: Write your takeaway as one sentence

What's the *one thing* you want the audience to remember from your story?

What's the one message that the audience should take away?

## CHECK IT: Verify these yourself

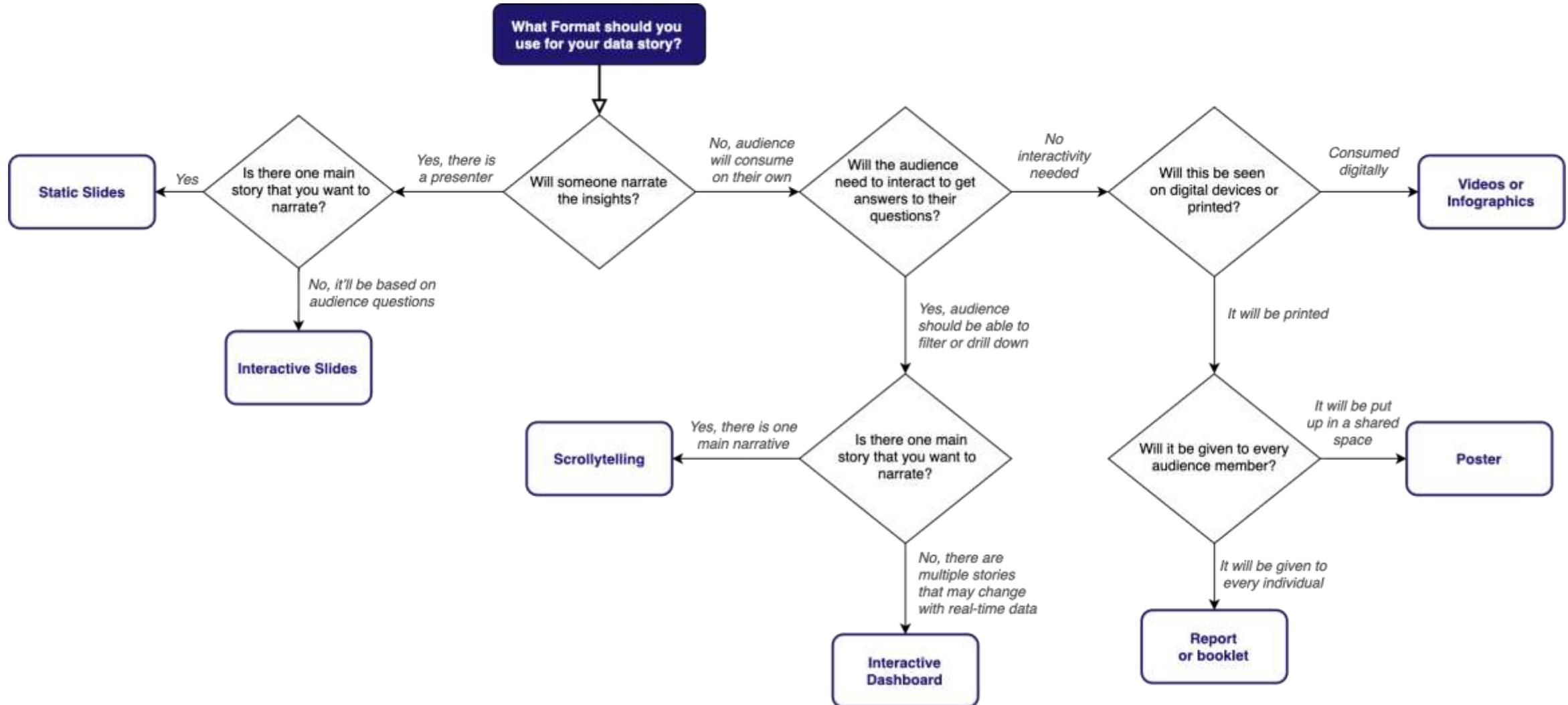
- Is it a single, complete, sentence?
- Does it deliver what you want the audience to remember?
- Will your audience care a lot about this?

It's a one-sentence summary of the most important message for the audience.

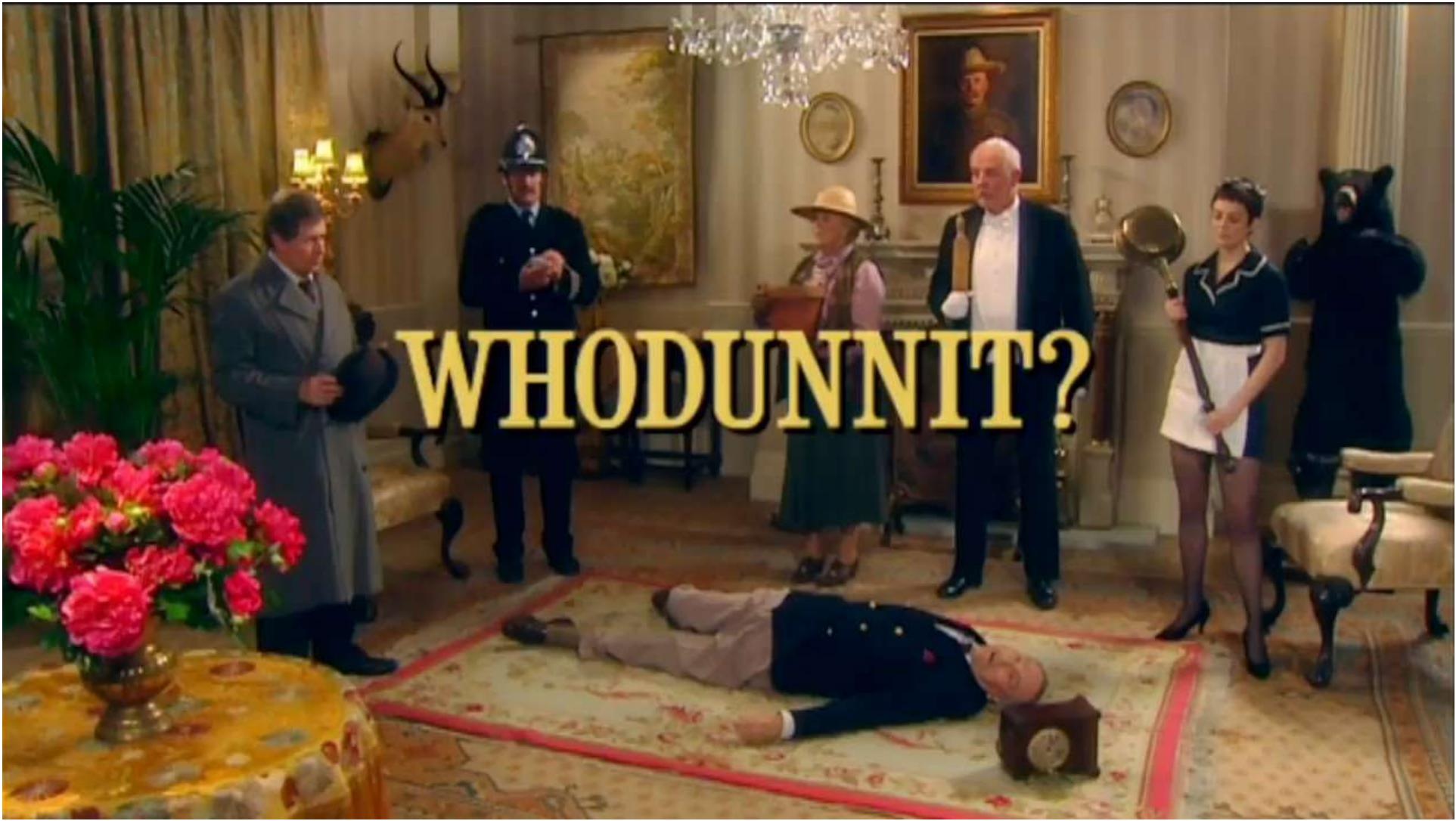
# Here is the storyline for the analyses we saw earlier

	Purchasing Commodities	Cargo Delay	Customer Churn
Takeaway	<p>Focus on reducing the number of vendors products ICG (in P126), FRS (in P121) and SWB (in P074) for a potential 40% reduction in logistics &amp; vendor cost.</p> <p>ICG spend is among the highest, at €6.9m. P126 typically orders 40 times a week, often from 15-20 vendors.</p> <p>FRS spend is €3.2m. P121 orders from 3 vendors 8-14 times a week.</p>	<p>To reduce the TAT to 1.5 hours at Airport XYZ, increase the number of forklifts from 1 to 2, and the number of trained staff from 4 to 6</p> <p>The number of forklifts is the biggest driver of TAT. Each forklift typically reduces TAT by 15-30%.</p> <p>Total staff count does not impact TAT. Increasing trained staff has a more tangible impact of ~5-10% per person.</p>	<p>If a customer has not called in the last 5-14 days, and they have made only 1 recharge under \$20 last quarter, make them an offer to retain them.</p> <p>The biggest driver of retention is when the customer made the outgoing call. The 5-14 days bucket has the highest variation.</p> <p>Customers who make at most 1 recharge under \$20 are 280% more likely to churn than others.</p>
Supporting points			

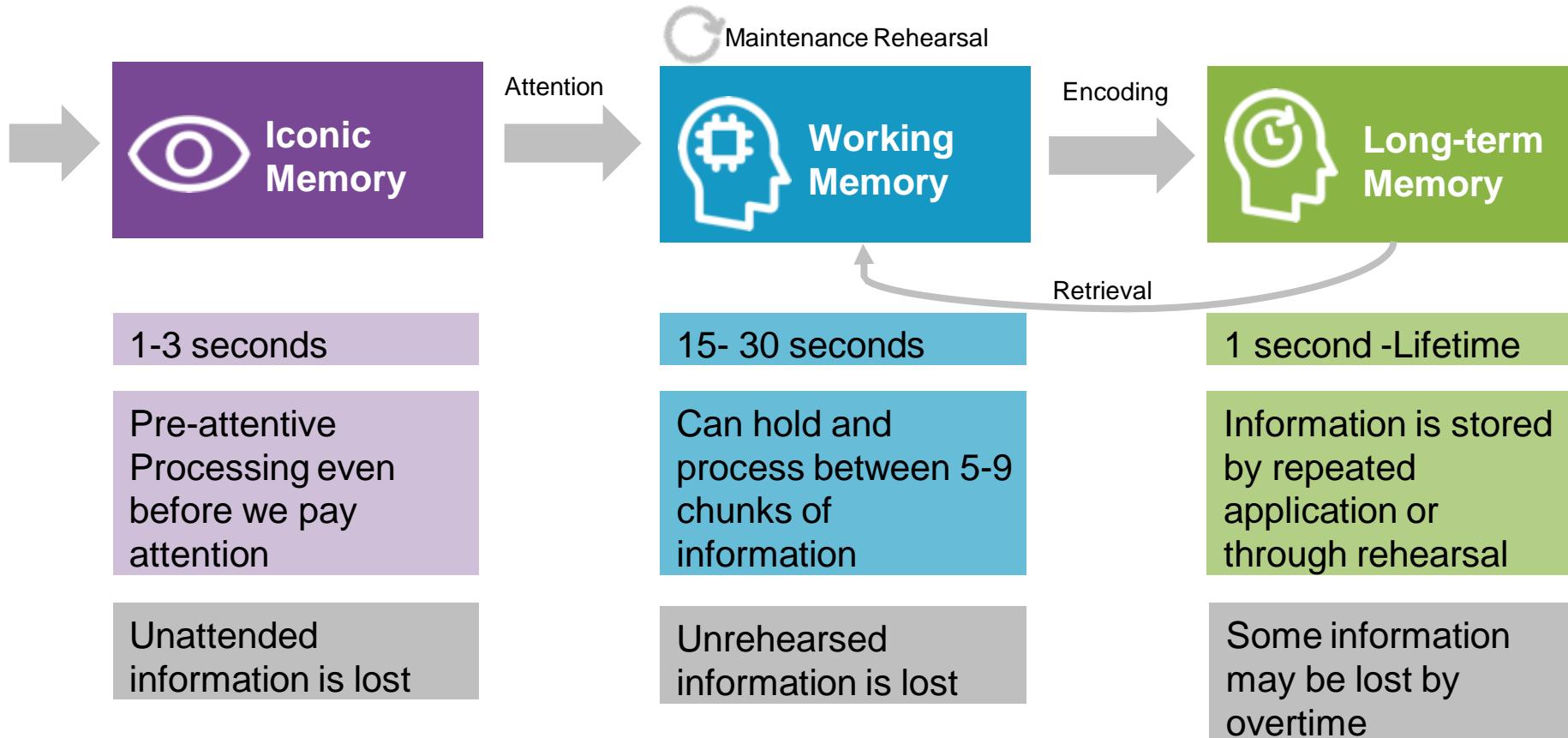
# Pick a format based on how your audience will consume the story



Here's a fun quiz



# Human memory is continuously capturing & forgetting information

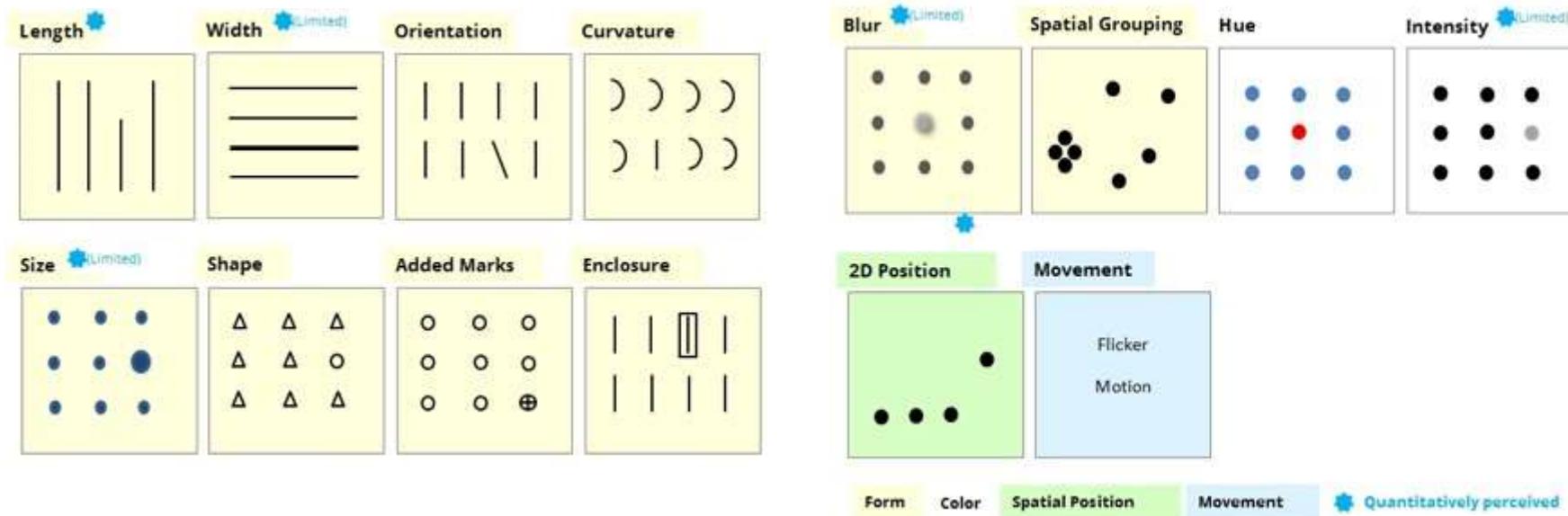


Visual perception as the **ability to interpret** the surrounding environment by **processing information** that is contained in visible light.

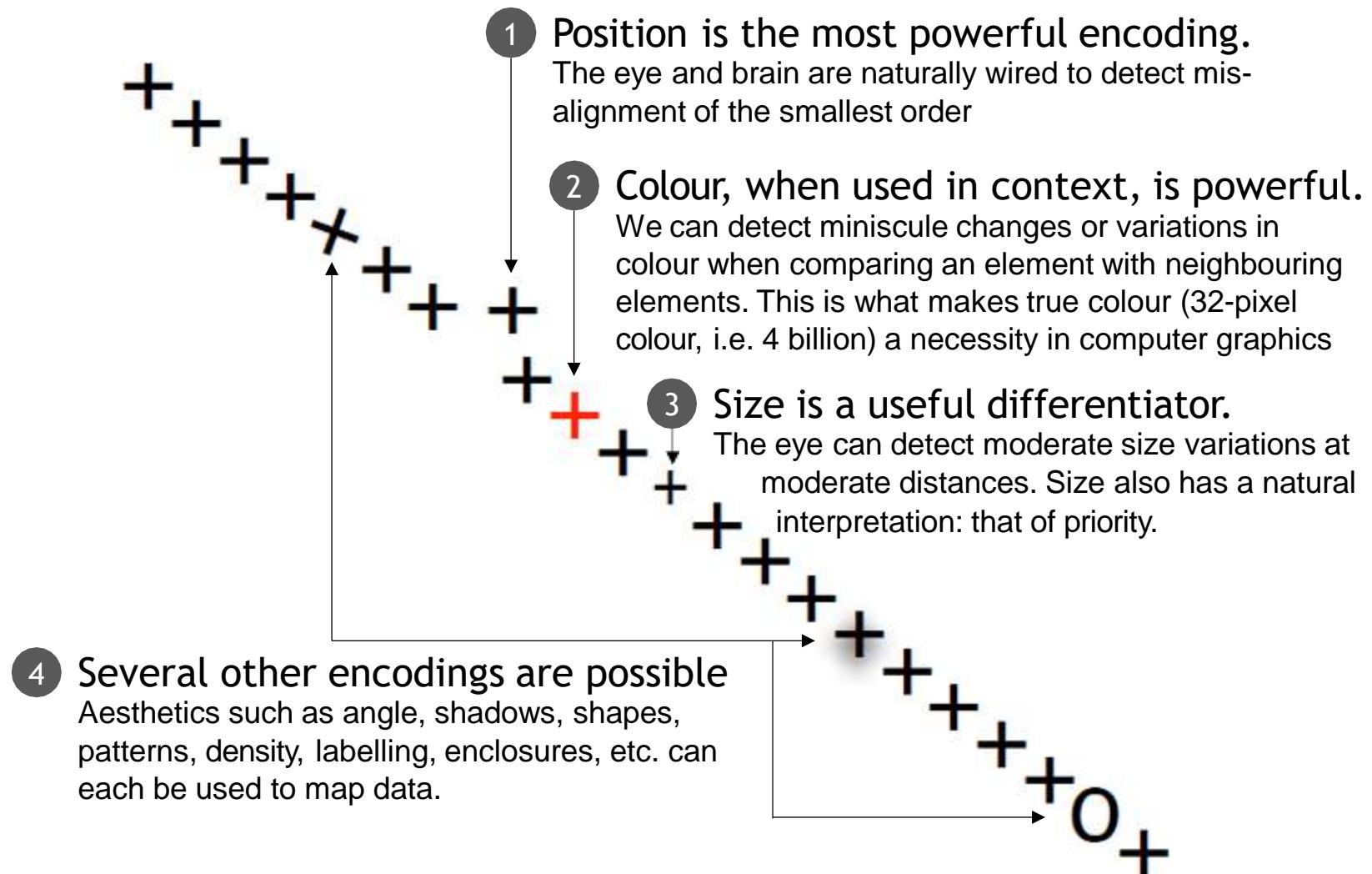
# Some visual attributes are noticed before we actively pay attention to them

4 categories of pre-attentive visual attributes.

**Form** | **Colour** | **Spatial Position** | **Movement**



# ...and these attributes vary in their effectiveness

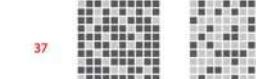
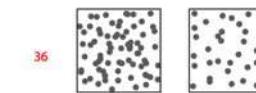
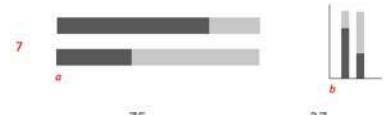
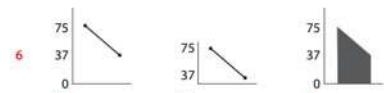
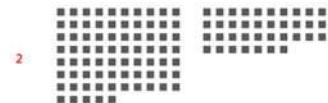


# Let's start small: visualize two numbers (2 & 8) from today's date

Sketch it out or watch others on Invision Freehand

# There are many ways to visualize just two numbers

75, 37 *multiple ways to communicate two quantities*



# Properties and Best Uses of Visual Encodings

Example	Encoding	Ordered	Useful Values	Quantitative	Ordinal	Categorical	Relational
	Position, Placement	Yes	Infinite				
1, 2, 3: A, B, C	Text Labels	Optional	Infinite				
	Length	Yes	Many				
	Size, Area	Yes	Many				
	Angle	Yes	Medium/Few				
	Pattern Density	Yes	Few				
	Weight, Boldness	Yes	Few				
	Saturation, Brightness	Yes	Few				

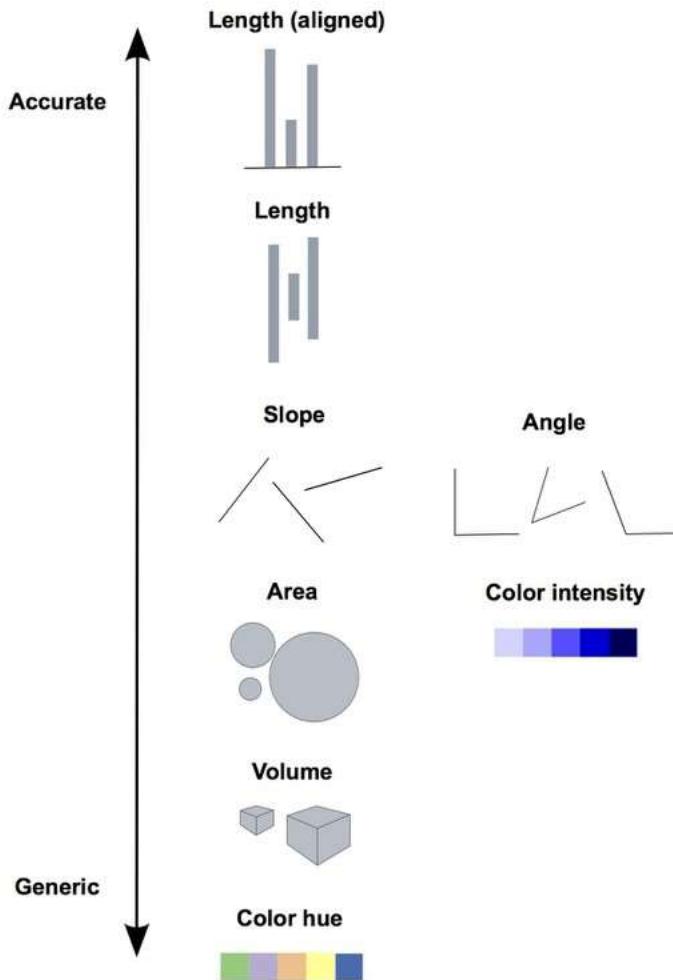
Noah Illinsky ComplexDiagrams.com/properties 2012-06

# Properties and Best Uses of Visual Encodings

Example	Encoding	Ordered	Useful Values	Quantitative	Ordinal	Categorical	Relational
	Color	No	Few (<20)				
	Shape, Icon	No	Medium				
	Pattern Texture	No	Medium				
	Enclosure, Connection	No	Infinite				
	Line pattern	No	Few				
	Line Endings	No	Few				
	Line Weight	Yes	Few				

Noah Illinsky ComplexDiagrams.com/properties 2012-06

# Some encoding methods are better\* than others



# Examples - Visual encoding

## Usain Bolt vs. 116 years of Olympic sprinters

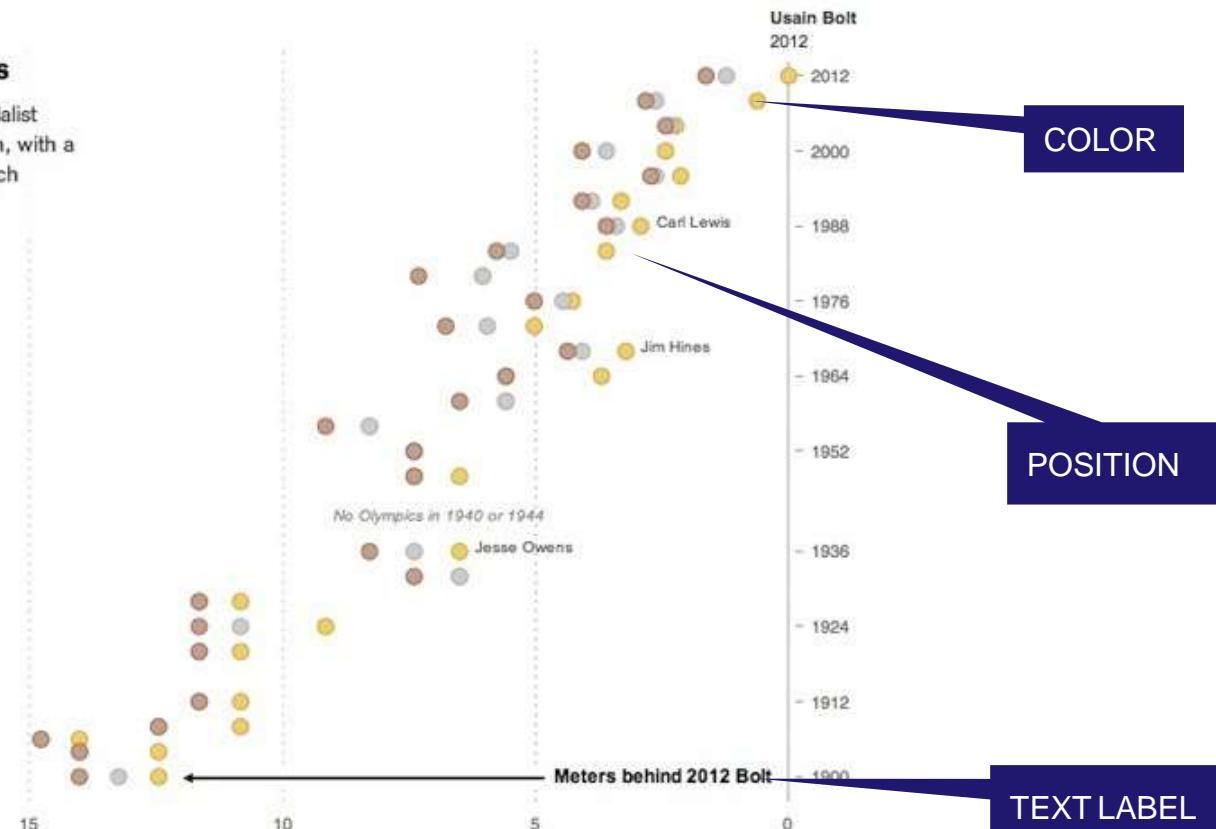
Based on the athletes' average speeds, if every Olympic medalist raced each other, Usain Bolt (the London version) would win, with a wide distribution of Olympians behind him. Below, where each sprinter would be when Bolt finishes his race.

### MEDALS BY COUNTRY

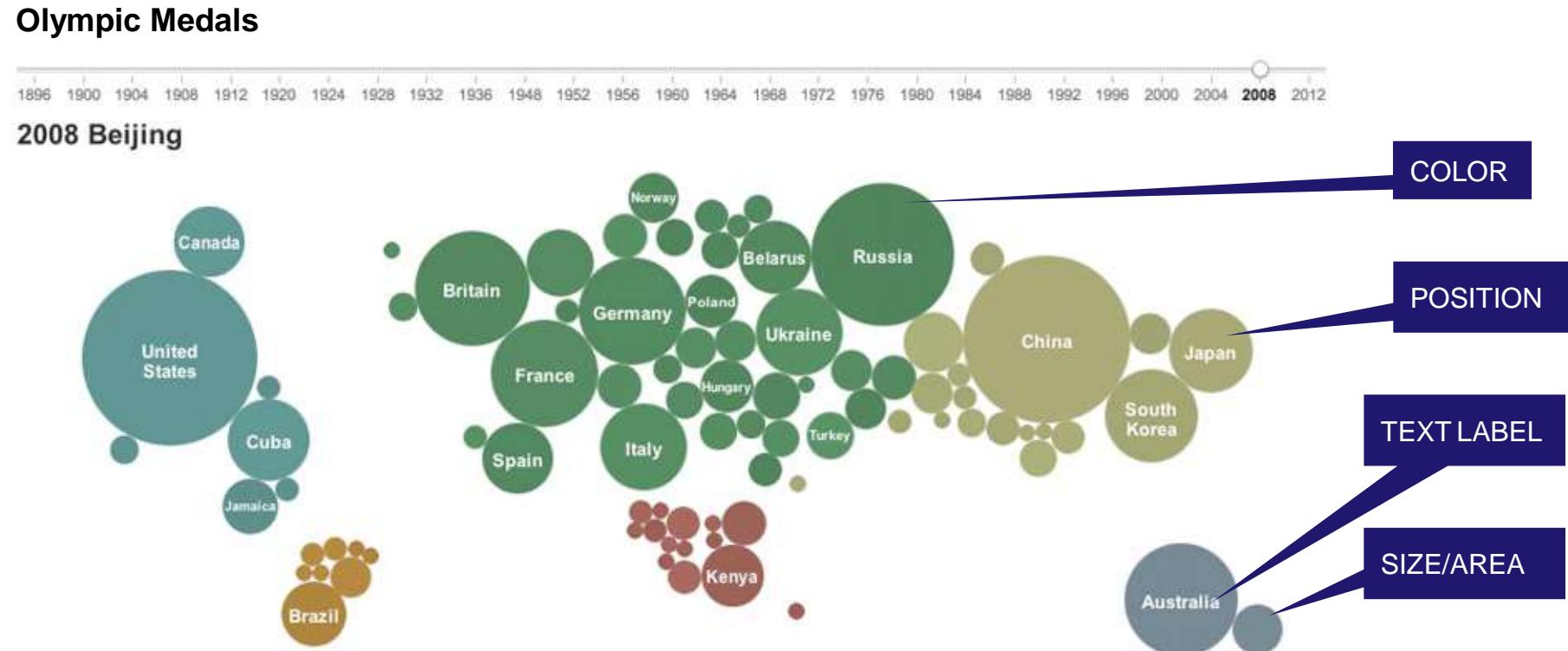
United States	40	Barbados	1
Britain	8	Bulgaria	1
Jamaica	7	Hungary	1
Canada	5	Netherlands	1
Trinidad and Tobago	4	New Zealand	1
Australia	3	Panama	1
Portugal	1	South Africa	1
Germany	3	United Team of Germany	1
Cuba	2		
Namibia	2		
Soviet Union	2		

Thomas Burke

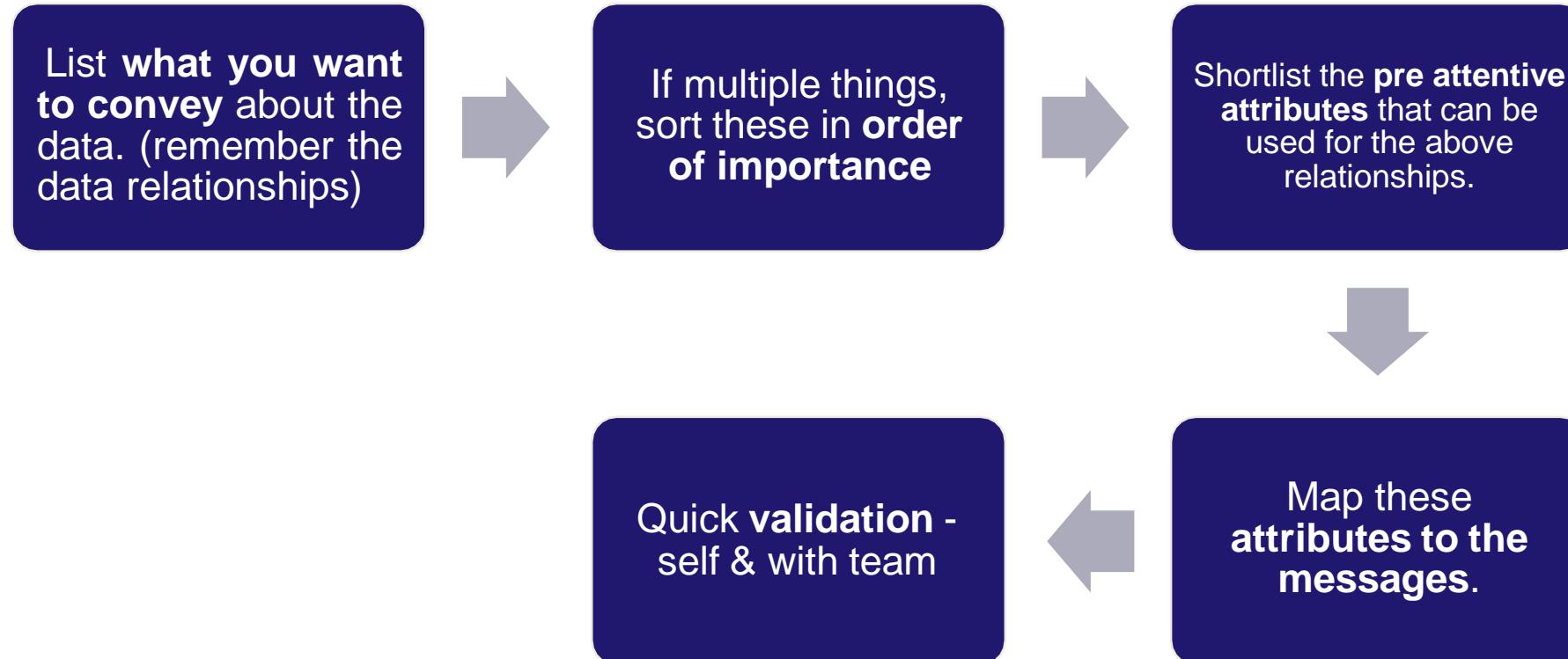
25 20



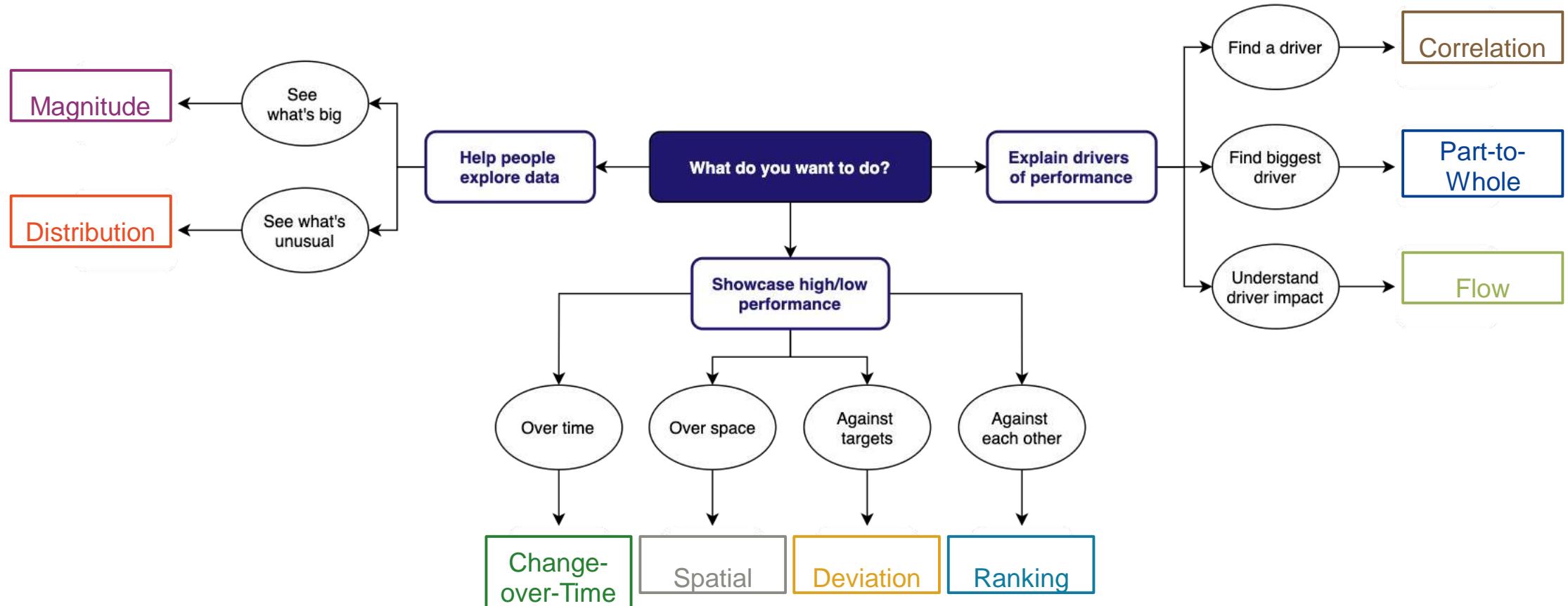
# Examples - Visual encoding



# Guidelines on Visual Encodings



# Pick a visual design based on the takeaway



# How the data should be interpreted decides the type of chart to be used

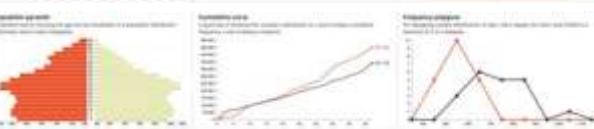
Help people explore data

## Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences).

## Distribution

Show values in a dataset and how often they occur. The shape (or skew) of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.



Showcase high/low performance

## Change-over-Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries.

## Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average.

## Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

## Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value.

Explain drivers of performance



<https://gramener.github.io/visual-vocabulary-vega/>

# Several other chart frameworks

Datavizcatalogue.com



Datavizproject.com



ft.com/vocabulary



# Annotate to explain & engage. Use four types of narratives

**Summarize the visual in its title**

Don't describe the chart.

Don't write the user's question.

Write the answer itself. Like a headline.

→ Teachers add marks to stop some students from failing

This chart shows Class 10 students' English marks in Tamil Nadu, India, in 2011. The X-axis has the mark a student has scored. The Y-axis has the # of students who scored that mark.

This is a bell curve. But the spike at 35 (the mark at which students pass) is unusual. Teachers must be adding marks to some of the students who are likely to fail by a small margin.

**Explain & interpret the visual**

How should the user *read* it?

What do you say when you talk through it?

Explain what the *visual* is. Then the *axes*.

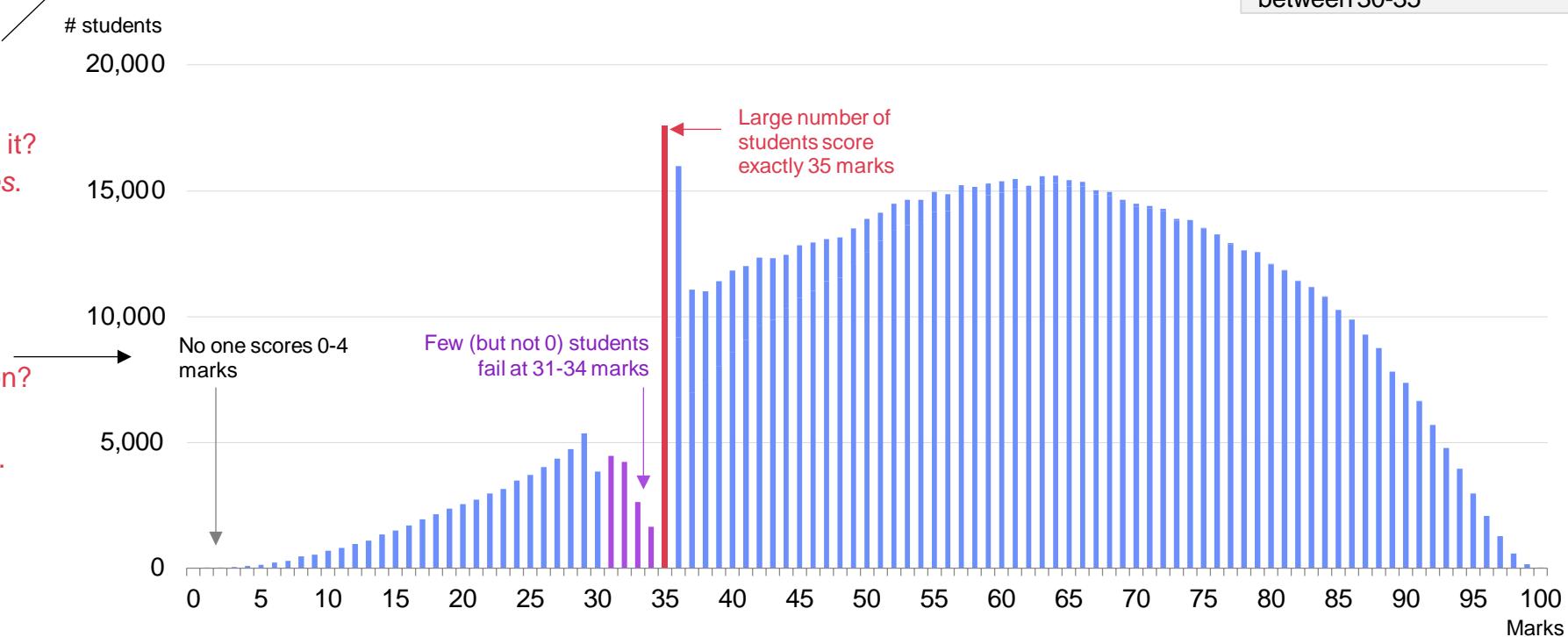
Then its *contents*. Then the *inference*.

**Annotate essential elements**

What should the user *focus* their eyes on?

Point it out, or highlight it with colors

Interpret what they're seeing – in words.



**Recommend an action**

How should I *act* on this?

You need to *change* the audience.

(Otherwise, you made no difference.)

→

Only some students get this benefit.  
Identify a fair policy that will be applied consistently.

**What's unusual**

Large number of students score 35 marks.

Few (but not 0) students score between 30-35

# In summary, here are the 9 steps to go from data to a data story

Who is your audience? They determine the story

What is their problem? That defines your analysis

Find the right analysis to solve the problem Filter for big, useful,  
surprising insights

Start with the takeaway. Summarize your entire story Add  
supporting analyses as a tree

Pick a format based on how your audience will consume the story

Pick a visual design based on the takeaway

Annotate to explain & engage. Use four types of narratives

**"Where can I learn more?"**

# Recommended Resources

## Data Storytelling at Gramener

1. Solutions on Gramener site:  
<https://gramener.com/solutions/>
2. Gramener's Blog  
<https://blog.gramener.com/>
3. Gramener's upcoming webinars:  
<https://linkedin.com/company/gramener/>

## Tools to learn

- **Paper & Pen** (Collaborative)
- **Excel**
- **Tableau & PowerBI**
- **JS** (D3, Vega, Plot.ly)
- **Python** (Bokeh/Matplotlib)
- **R** (ggplot)
- Raw graphs
- Illustrator / Sketch / Figma

## Books to read

- Resonate - Nancy Duarte
- Storytelling with Data - Cole Nussbaumer Knaflic
- Truthful Art - Alberto Cairo
- Design of everyday things - Don Norman
- Back of the napkin - Dan Roam

You can find me (Rasagy) on Twitter/LinkedIn/Instagram

“Most of us need to listen to the music  
to understand how **beautiful it is**.  
But often that’s how we present statistics:  
we just show the notes,  
**we don’t play the music.**”

— *Hans Rosling*

# THANK YOU

