Fundamental Concepts in Data Insight:

Data Mindset & Skills

Fundamentals for a General Audience

Data Mindset & Skills

- Why do you need a Data Mindset?
    - Discuss: What are the components of a Data Mindset?
    - Review: What are the components of a Data Mindset?
    - What mindset does a data practicioner need?
    - What mindsets does an organization need?
- Reasoning
    - What is Reasoning?
    - What is(n't) Evidence-Based reasoning?
        - Superstition
        - Speculation
        - Evidence-based reasoning
    - What techniques and technologies support evidence-based reasoning?
    - What are cognitive biases?
        - Anchoring and Adjustment Effect
        - Clustering illusion
        - Framing Effect
        - Availability heuristic
        - Base rate fallacy
        - Selective perception
        - Group think
    - How do you formulate a hypothesis?
    - How does data distinguish between hypotheses?
    - Demo: Amnes Room

- Storytelling & Visualization
    - What is Data Storytelling?
        - Project Valcri
    - What is a story?
    - Why are data projects presented as stories?
    - What is Visual Science?
    - How do you create good visuals?
    - DEMO: Ted Talk
- Group Project: Write a Data Story (30 min)

Why do you need a Data Mindset?

Intelligence work does not take place in a vacuum. The process of collecting and analysing information is influenced by a range of variables or "human issues". These issues impact both the process of intelligence and its outputs. They include but are not limited to ***cognitive factors such as sense-making, evidential reasoning and bias mitigation*** *(as described above).*

They also include non-cognitive variables including: the quality of one's data, the nature and quality of inter-and intra-team collaboration, legal and ethical constraints, operational factors, personal and interpersonal factors, situational and dispositional factors, socio-cultural factors, technology and socio-technological dynamics and so on.

Discuss: What are the components of a Data Mindset?

Review: What are the components of a Data Mindset?

- leadership
  - initivative
  - strategic
  - coordinative
  - prioritizing
- communication
  - storytelling
  - clarity and focus
  - priortization, audience-empathy
  - presentation
  - visualization
- reasoning thinking
  - depthwise
  - lateral
  - evidential
  - systematic...

What mindset does a data practicioner need?

- scepticism
    - disagreeable
- problem-oriented
    - proactive / self-motivated
    - "tasks" unlikely to be clear
        - derived by practicioner from problem
- democratic
    - cooperative with data
- accountability
    - data storytelling
        - communicates clearly to non-tech audiences
    - clarity of thought and justification

What mindsets does an organization need?

- data literacy
    - read: comprehend
    - work with: per-role specialism
    - analyse: insight
    - argue with: evidence-based thinking
    - .. data
- evidence-based reasoning
    - scepticism
        - towards tradition
        - AND! data
    - clarity of thought
    - systematic and analytical thinking
- problem-oriented
    - defining meaningful success criteria
    - working towards it
    - prioritising the problem (over people, ...)
- data democracy
    - jobs defined by value-add not data ownership
    - self-serve data systems
- accountability
    - explanations and justifications given with decisions
        - even if, "hunch"

Reasoning

What is Reasoning?

- a priori
- a posteriori
- instinct
    - action which appears immediately appropriate
- intuition
    - sense of plausibility based on prior experience
    - 'immediate' attempt to explain without expermimentation
- reasoning
    - dialectical (internal or external)
    - can be formalized and shared
    - systematic
- a priori reasoning
    - from concepts (may be learned a posteriori)
- a posteriori reasoning
    - from data/observation (will require concepts not in the data to interpret)

What is(n't) Evidence-Based reasoning?

- superstition
- vs. speculation
- vs. evidence-based reasoning

Superstition

- dogmatic vs dialetical
- ideas judged *only* by extrinsic properties
    - how much they are liked,
    - fit with other ideas, etc.
- no causal mechanism offered
- unrelated concepts connected by association
    - vs., plausible mechanism
- eg., china, 5g, etc. TODO

Speculation

- imaginative simulation
    - counterfactuals
- appropriate for unmeasurable domains
- imaginative process can be exposed and open to reivisoon
- possible causal mechanisms offered
- concepts related by intrinsic properties
    - with detailed explanation
- EG>, TODO

Evidence-based reasoning

- probability
- multiple hypotheses
- confidence intervals
- dialectical
- causal, mechanistic
- appropriate for measurable domains

What techniques and technologies support evidence-based reasoning?

- Machine Learning (Low Quality, Automated)
  - Deep Learning (Artifical Intelligence)
- Analytics (Medium Quality, Manual)
  - Fact-Finding
  - Diagnostics
- Statistical Modelling (High Quality, Automated/Manual Mix)
  - Experiment & Data Science

What are cognitive biases?

Confirmation bias: The tendency to search for or interpret information in a way that confirms one's preconceptions or hypotheses.

Anchoring and Adjustment Effect:

- The tendency to rely too heavily or 'anchor' on largely unrelated information when making decisions.

Clustering illusion:

- The tendency to see patterns where actually no patterns exist.

Framing Effect:

- The tendency to draw different conclusions from the same information, depending on how that information is presented.

Availability heuristic:

- The tendency to make judgments about the probability of events occurring by how easily these events are brought to mind.

Base rate fallacy:

- The tendency to base judgments on specifics, ignoring general statistical information.

Selective perception:

How do you formulate a hypothesis?

- a hypothesis is a claim which can be tested
- define the problem as a hypothesis
  - factual description (that may or may not be true)
  - neutral about solution
  - no normative / evaluative language

How does data distinguish between hypotheses?

- data cannot produce hypotheses
  - has no raw intepretation
- concepts are needed to intepret data
  - measurement of objects does not contain descriptiuons of those objects...
  - many (infinitely many) are consistent with data we observe (EG., TODO, Amnes ROOM)

Demo: Amnes Room

- multiple hypotheses about geometric structure
- how do we decide?

Storytelling & Visualization

What is Data Storytelling?

- taking a data project and telling a story about it
  - will likely involve a audio-visual component simplified for non-technical audience
  - will present a call-to-action

Project Valcri

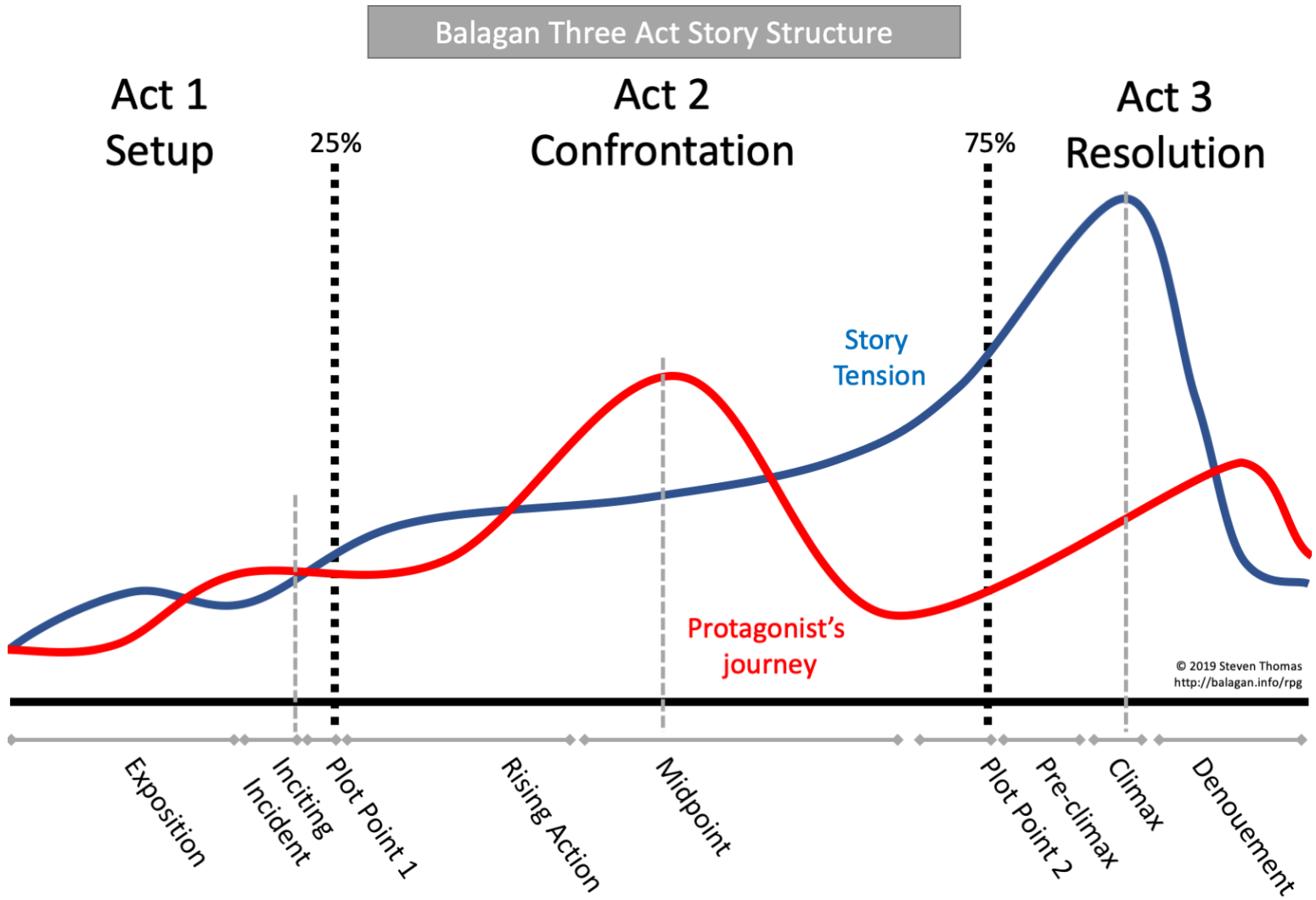
> *It has been said that making sense of a situation involves fitting its elements into some kind of structure that links them together (Klein, Phillips, Rall, and Peluso, 2007),* ***such as a story, a map, a script or a plan****. That when people make sense of stimuli they do so by placing them into some kind of framework which allows them to,* ***"Comprehend, understand, explain, attribute, extrapolate and predict"*** *(Starbuck and Milliken, 1988).*
>
> *Most accounts of sensemaking make similar reference to the role of structuring information, whether this be 'in the head' or 'in the world' (e.g. on paper or computer screens) (see also, Russell, Stefik, Pirolli and Card 1993; Pirolli and Card, 2005).*

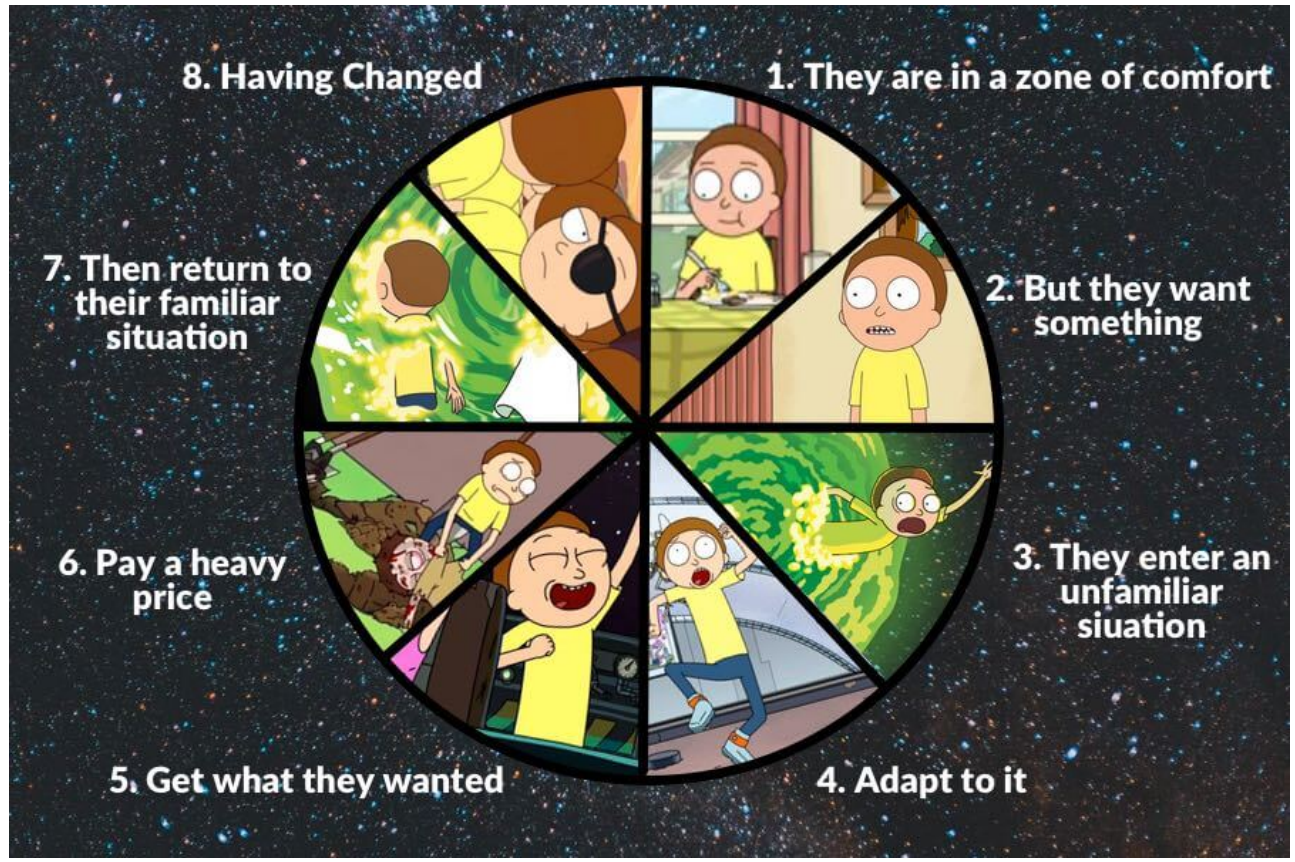
A narrative on the other hand, is a spoken or written account of events as they unfolded over time. It has the form of a story or chronicle.

From a psychological perspective, narrative seems to be an important form through which people make sense of criminal evidence (c.f. Pennington and Hasties' Story Model (1992) and Wagenaar and Crombags' Anchored Narrative Theory (Wagenaar, van Koppen & Crombag, 1993).

What is a story?

- stories communicate:
    - metaphor/framing
        - provides prioritiues
    - moral/ethical concerns
        - motivates caring
- stories have:
    - plot
        - events on a timeline
    - characters
        - with virtues/vices
    - emotional structure: tension/release
        - plot arcs: climax/denoument
        - character arcs: challenge/triumph
    - framing
        - core metahpors
        - authorial perspective

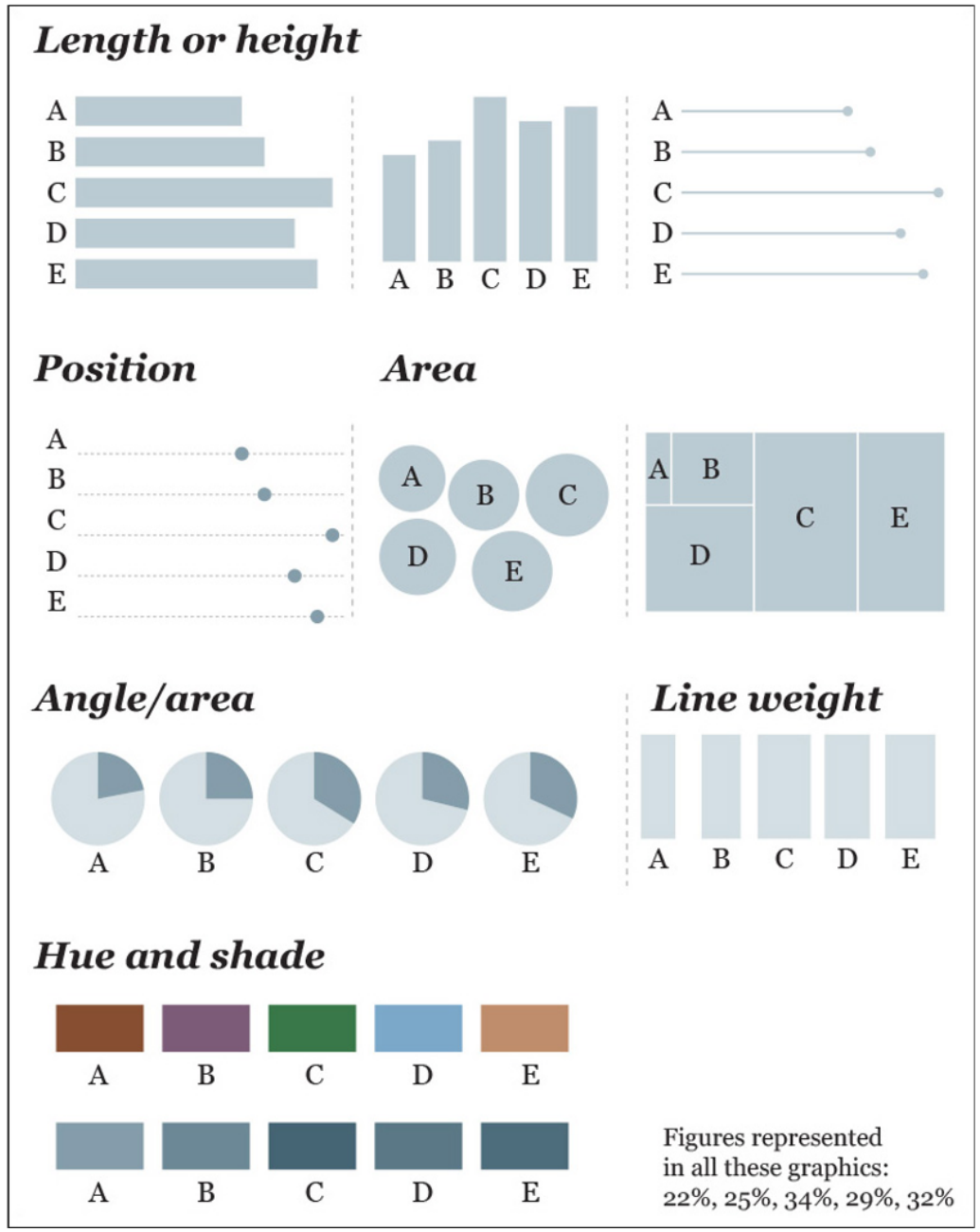

Balagan Three Act Story Structure
Act 1
Setup
25%
Act 2
Confrontation
75%
Act 3
Resolution
Story
Tension
Protagonist's
journey
© 2019 Steven Thomas
http://balagan.info/rpg
Exposition
Inciting
Incident
Plot Point 1
Rising Action
Midpoint
Plot Point 2
Pre-climax
Climax
Denouement

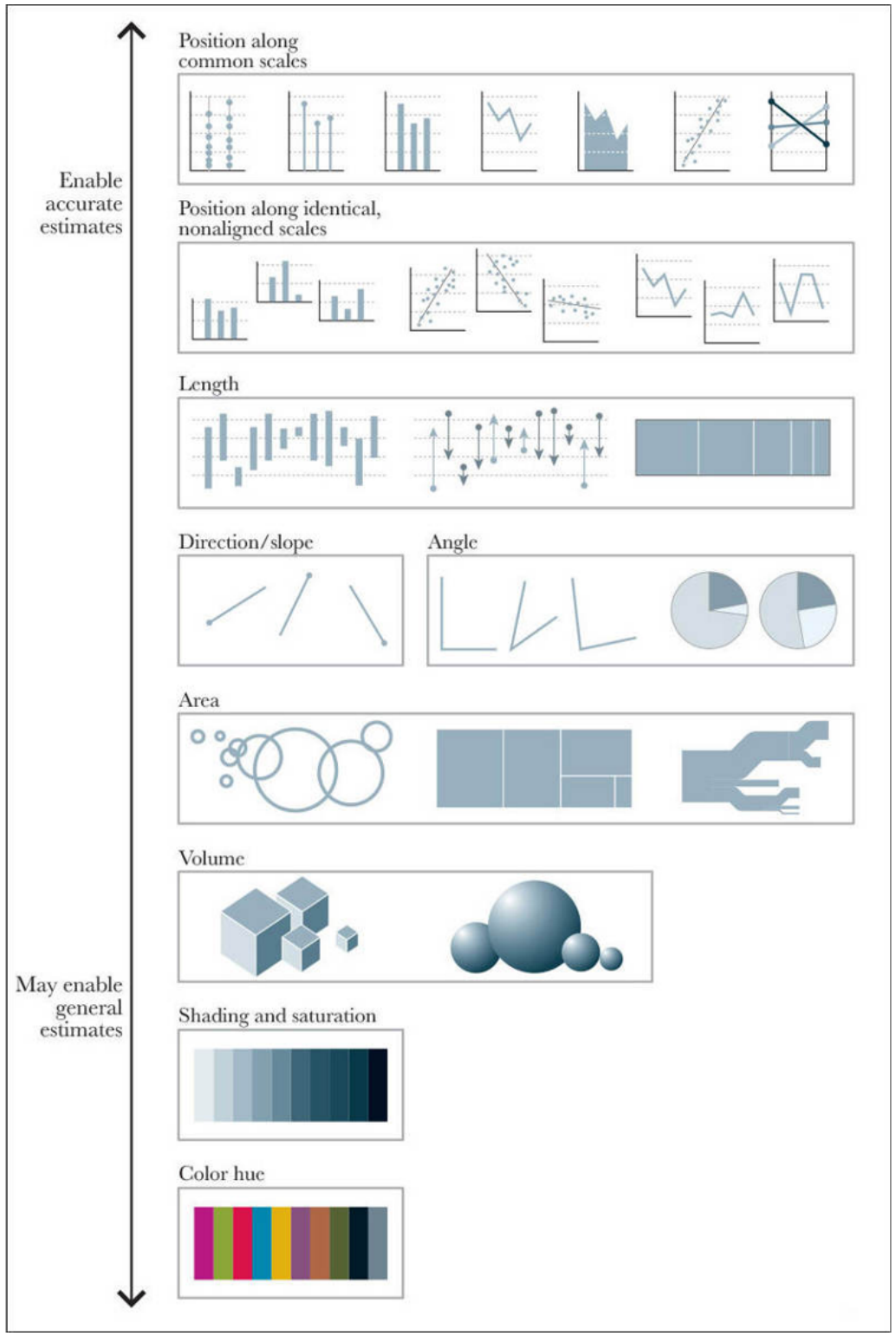

8. Having Changed
1. They are in a zone of comfort
7. Then return to their familiar situation
2. But they want something
6. Pay a heavy price
3. They enter an unfamiliar siuation
5. Get what they wanted
4. Adapt to it

Why are data projects presented as stories?

- a story is the simplest *universal* communication paradigm
- if something can be presented as a story, it is more likely to be universally understood
- stories are mostly about building trust and encouraging action
- stories (in my view) are NOT explanations!

What is Visual Science?


Length or height
A
B
C
D
E
A B C D E
A
B
C
D
E
Position
A
B
C
D
E
Area
A
B
C
D
E
A
B
C
D
E
Angle/area
A B C D E
Line weight
A B C D E
Hue and shade
A B C D E
A B C D E
Figures represented
in all these graphics:
22%, 25%, 34%, 29%, 32%


Position along common scales
Enable accurate estimates
Position along identical, nonaligned scales
Length
Direction/slope
Angle
Area
Volume
May enable general estimates
Shading and saturation
Color hue

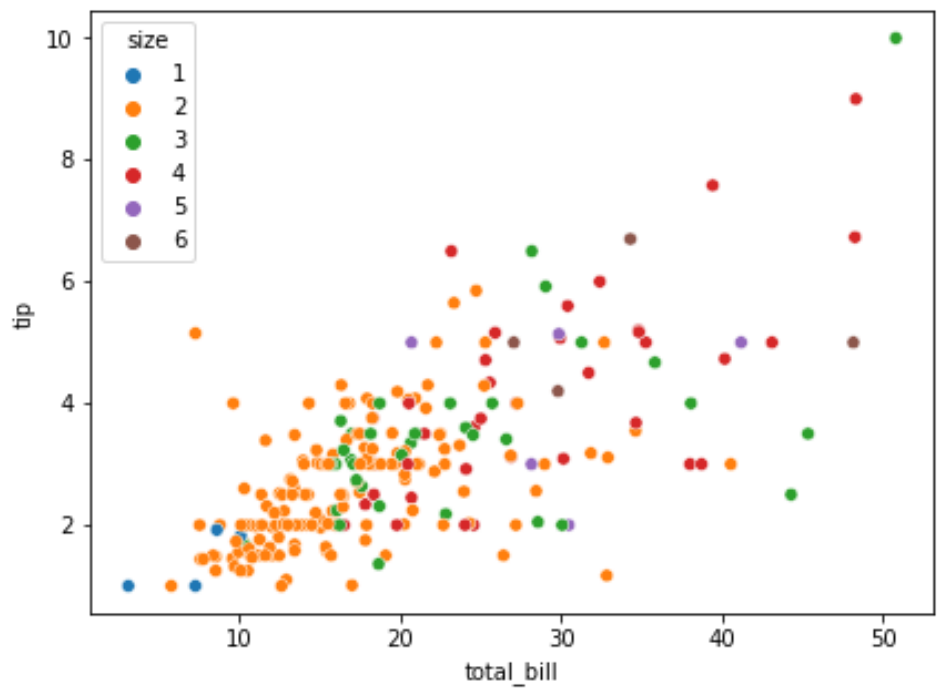
How do you create good visuals?

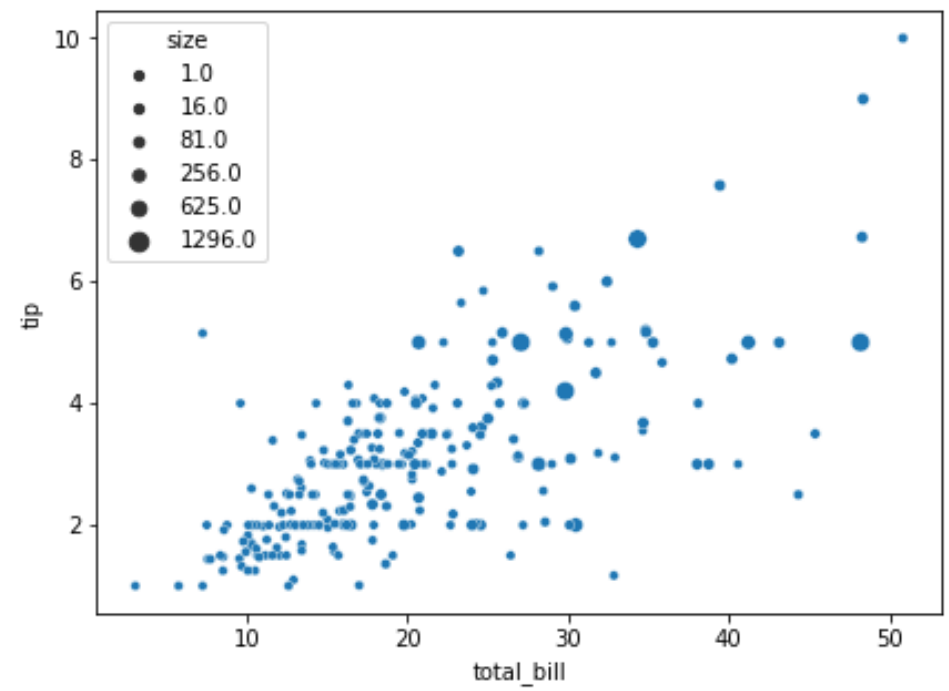
```
import seaborn as sns

d = sns.load_dataset('tips').astype({'size': 'category'})
```

```
d.sample(3)
```

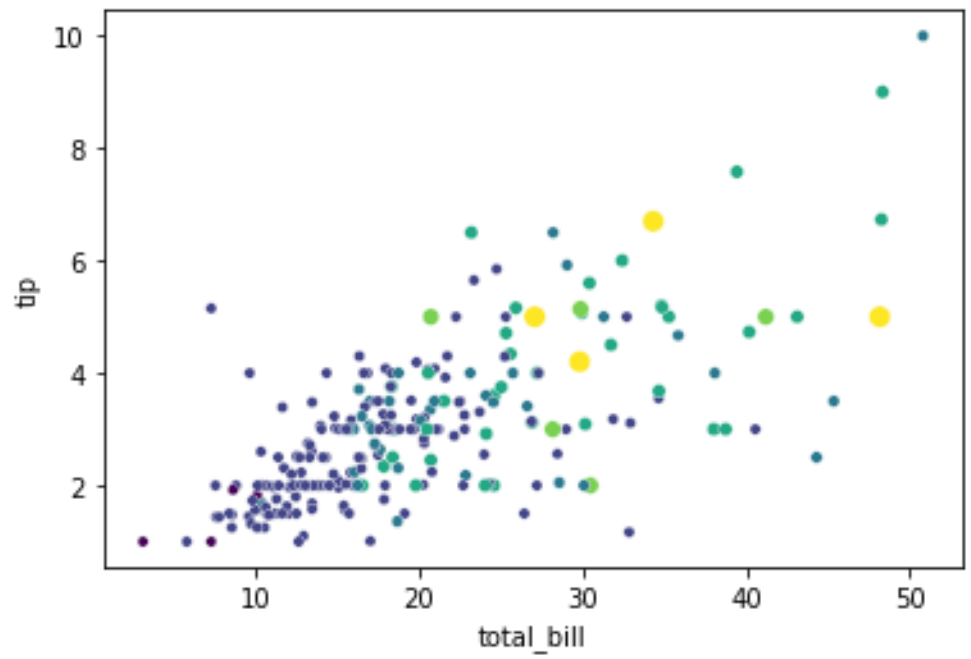
	total_bill	tip	sex	smoker	day	time	size
191	19.81	4.19	Female	Yes	Thur	Lunch	2
172	7.25	5.15	Male	Yes	Sun	Dinner	2
226	10.09	2.00	Female	Yes	Fri	Lunch	2

```
fig, ax =plt.subplots(1, 2, figsize=(15,5))

sns.scatterplot(x=d['total_bill'], y=d['tip'], hue=d['size'], ax=ax
[0])
sns.scatterplot(x=d['total_bill'], y=d['tip'], size=d['size'].astyp
e(float)**4, ax=ax[1]);
```





```
sns.scatterplot(
    x=d['total_bill'], y=d['tip'], hue=d['size'].astype(float), siz
e=d['size'].astype(float)**4,
    palette="viridis"
).legend_.remove()
```



DEMO: Ted Talk

Group Project: Write a Data Story (30 min)

- Group Project: Produce a powerpoint to argue for a new strategy

Eg.

Two roosters fought for supremacy in the farmyard.

Finally one was defeated and he went and hid himself in a corner of the hen-house.

The victor flew up to the roof of the barn and begin to crow, “I’ve won, I’ve won!"

An eagle swooped down and carried him away.

The rooster that had been defeated suddenly found himself the master of the farmyard.

Background ”Implied” Message

- Who is the audience?
- What is your message for them?
- Where does the story take place?
- What team, company, project environment?

Possibly Explicit Components

- What is the inciting incident?
- Why did the project begin? (What failed, changed, was uncertain?)
- Who are the characters?
- Who is the protagonist? (Who will be the key character we follow?)
- Who is the antagonist? (Were you trying to prove the competition wrong?)
- What are the complicating factors?
- What is the climax?
- What is the memorable event of the project?
- How was the conflict/challenge settled?
- What lessons have been learnt?

Tips

- Review the 3-act structure and Dan Harmon's story-circle
- The story circle approach works well for TV shows which require protagonists to be in the same situation at the beginning of each episode
- Stories that attempt to provoke action (almost always) use outrage
    - outrage = the world *should be* like X, but *it isnt* and the *reason* is character fault
    - eg., we were naive to think X, X caused a lot of problem, therefore we should do Y
    - eg., the PM's pride in the UK has lead to blindness about risks from X, ...

TED Links

Watch the following ted talks

- https://www.youtube.com/watch?v=eIho2S0ZahI
- https://www.youtube.com/watch?v=6Af6b_wyiwI