CORE Skills

Delivering Data Science
In Resources & Energy

Data & Communication Sandbox

DAY 7

15-Day Data Science Springboard

Dr Ying Yap & Dr Jeremy Mitchell, Data Mettle

2 February 2022

Program partners

_dm

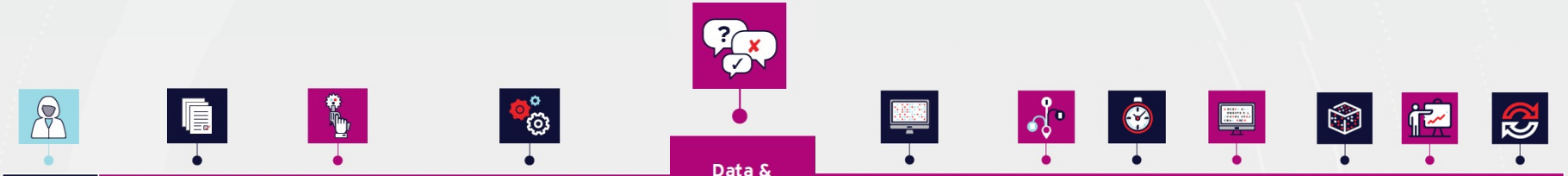
The University of Western Australia

Curtin University

A CORE partnership

Program Timeline

DAY 7: Data & Communication Sandbox



Leader Engagement	15 Day Professional Program														
2-hour Leading Data Scientists Leader Support	Preparatory		Introduction to Data Projects	Data Analysis			Data & Communication Sandbox	Data Fusion and Machine Learning		Data Fusion Sandbox	Special Data Types - Time-series & Networks	Special Data Types - Knowledge Discovery from Natural Language Processing & Text Mining	Special Data	Capstone Project Development & Presentation	Capstone Propeller
2-hours	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Day 14	Day 15
January 2022	Wed 24 Nov 2021	Thurs 25 Nov 2021	Wed 1 Dec 2021	Wed 8 Dec 2021	Wed 15 Dec 2021	Wed 19 Jan 2022	Wed 2 Feb 2022	Wed 9 Feb 2022	Wed 16 Feb 2022	Wed 23 Feb 2022	Wed 2 Mar 2022	Wed 9 Mar 2022	Wed 16 Mar 2022	Wed 23 Mar 2022	Wed 22 June 2022
Enabling your people's data science upskilling & project delivery in 15 day program	Introduction to the Program Tools: The Why?	Introduction to the Program Tools: Where does Data Science come from?	Zero to Data Science in a day	Getting to know the Program Tools: Data munging and exploratory data analysis	Simple predictions: Regression and statistical model building	Multivariate analysis and model building	Effective data storytelling - communicating results to non-technical audiences	Pros and cons of commonly used statistical and machine learning techniques I	Pros and cons of commonly used statistical and machine learning techniques II	Consolidate approaches covered and test on datasets	The 4th dimension and predictions	Finding needles in wordstacks	Spatial analytics and predictions	Capstone Project pitches to leadership	Project Review

Aims & Learning Outcomes

Day 7

Aims

- Employ a variety of effective data communication strategies.
- Appraise data driven narratives.
- Refine and appropriately frame questions.

Learning Outcomes

- You are invited to consider alternative avenues for communicating project results
- You are aware of interpretation biases in your audiences, and how design principles can be used to mitigate these.
- You can critique data presentations (both good and bad) in a variety of media.
- You can develop effective data-driven communication in different contexts.
- You examine 'data ethics' and appropriate handling of (personal) data.

Schedule

DAY 7

AWST	AEST	Agenda	Facilitator
07:30	**09:30**	**Q&A, Issues & Announcements**	Jeremy
07:45	09:45	**Clustering & Regression in Higher Dimensions (DAY 6 continued, from model NB. Partial Least Squares as homework)**	Jeremy
08:45	10:45	**Data Visualisation & Communication**	Ying
09:15	11:15	*Morning Tea*	
09:30	11:30	**Data Visualisation & Communication continued**	Ying
11:00	13:00	*Lunch*	
11:45	13:45	**Data Stories**	Ying
12:30	14:30	**Building Storyboards**	Ying
13:15	15:15	*Afternoon Tea*	
13:30	15:30	**Project Time**	All
14:00	16:00	**Review Storyboards**	Ying
14:30	16:30	**Closeout –Takeaways & Mid-Program Review**	Ying/Tamryn
14:55	16:55	**Menti** Feedback	Tamryn
15:00	17:00	**Close**	

Stories from Data: Understanding Context

(10-15 mins)

Who are you talking to?

- The more specifically you can define your audience, the better.
- If you're trying to motivate a few specific people, it's more likely to convince them to do something.
- If you have an established relationship, get to the point quicker.
- If not, you might want to spend more time on explanation to build trust.

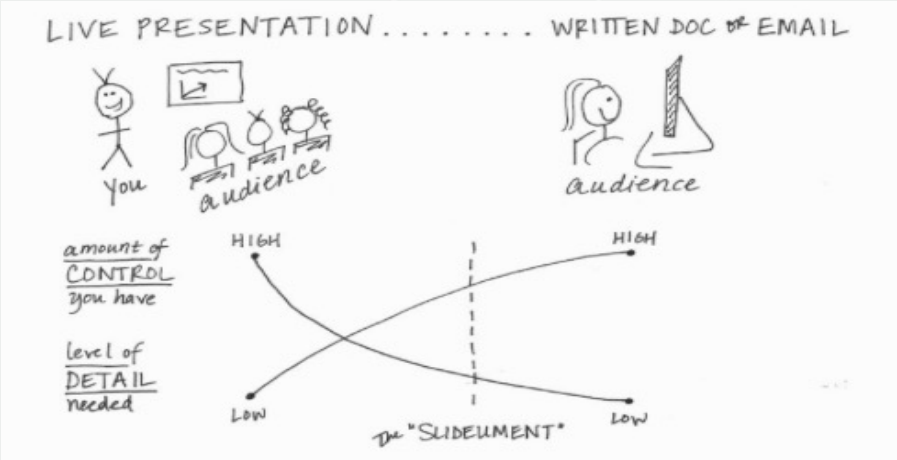
- **Who is your audience?**
- **What's their motivator?**
- **Who are you to them?**
- **What is the environment?**

What do you want out of this?

- **What do you want the audience *to know*?**
- **What do you want them *to do*?** (*action words!)*
- **How does authority/seniority play into this?**
  Can you recommend something directly, or just encourage discussion?
  You're typically the data expert.
- *We'll go through defining some of this in a succinct way shortly.*

A Continuum of Communication Styles

- How much detail do you need to include?
- How can you convey tone?
- How well can you control the reception of your message?
- How do you expect your audience to respond?
- Do you need to be able to have finer-grained control, and adapt to your audience in real time?



Avoid 'slideuments' (slides/document) – these forms are not usefully interchangeable!

How does this link to Data Culture and literacy?

- The environment you're in and your audience's level of data literacy constrains how you present your message.
- How data is linked to decisions will be different depending on the environment – so what you want people *to do* may be different (*data-centric, data-informed, …*).
- Your own data literacy is on show (*"argue with data"*).
- You don't need to take the audience through the entire **exploration** you have undertaken to reach your conclusion, but it's good to **explain** the key points.

Some Context Questions (we'll come back to this later)

What do you need to include for this audience?

- What data is available to support your point?
- What background information is relevant or essential?
- Who is the audience or decision maker? What do we know about them?
- What biases do our audience have that might make them supportive of or resistant?
- Is our audience familiar with this data or is it new?
- What are the risks? What factors could weaken our case, and do we need to proactively address them?
- What would a successful outcome look like?

Pitches to Different Audiences

Geophysical Inversion

@ Domain Expert

Content developed between CORE Skills, Dr Morgan Williams & Dr Cericia Martinez | CSIRO

Inversion Algorithm

- Model objective function:

$$\phi(\rho) = \int_V (w(z)(\rho - \rho_0))^2 dv + L_x \int_V \left(\frac{\partial w(z)(\rho - \rho_0)}{\partial x} \right)^2 dv$$

$$+ L_y \int_V \left(\frac{\partial w(z)(\rho - \rho_0)}{\partial y} \right)^2 dv + L_z \int_V \left(\frac{\partial w(z)(\rho - \rho_0)}{\partial z} \right)^2 dv$$

- L_x, L_y, L_z: length scales

- ρ_0: reference model;

- w(z), depth weighting function of the form: $w(z) = (z - z_0)^{-\beta/2}$

- Misfit measure: $\phi_d = \left\| W_d (\vec{d} - \vec{d}_{obs}) \right\|^2$

Inversion Algorithm

- Minimize total objective function

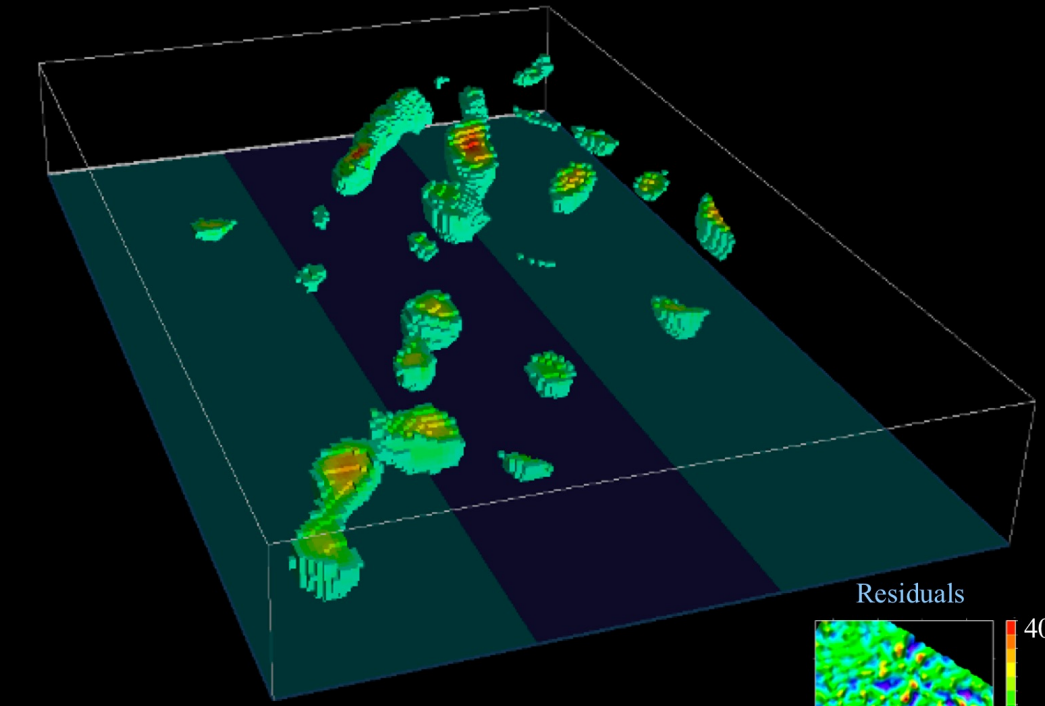
$$\min \phi = \phi_d + \mu \phi_m$$

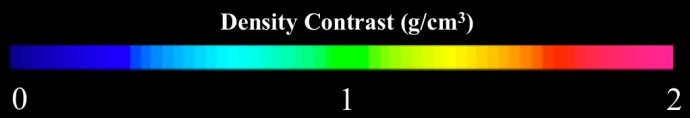
$$s.t.\ \ \rho_{\min} < \rho < \rho_{\max}$$

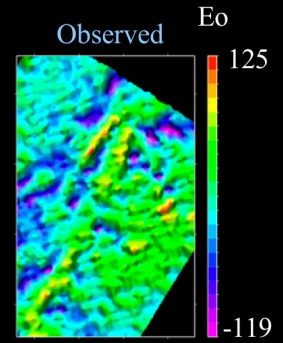
- Bound Constraints using Logarithmic barrier

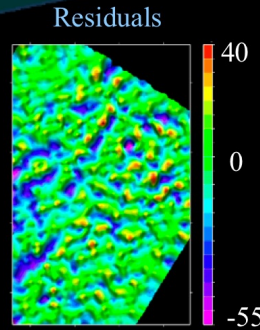
$$\phi(\lambda) = \phi_d + \mu \phi_m - 2\lambda \sum_{j=1}^{M} \left[\ln\left(\frac{\rho_j - \rho_j^{\min}}{\rho_j^{\max} - \rho_j^{\min}} \right) + \ln\left(\frac{\rho_j^{\max} - \rho_j}{\rho_j^{\max} - \rho_j^{\min}} \right) \right]$$

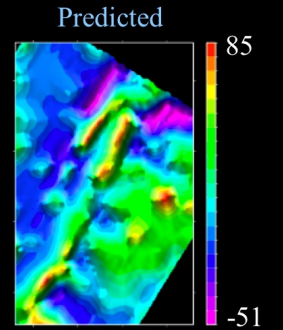
Recovered Model: T_{zz}











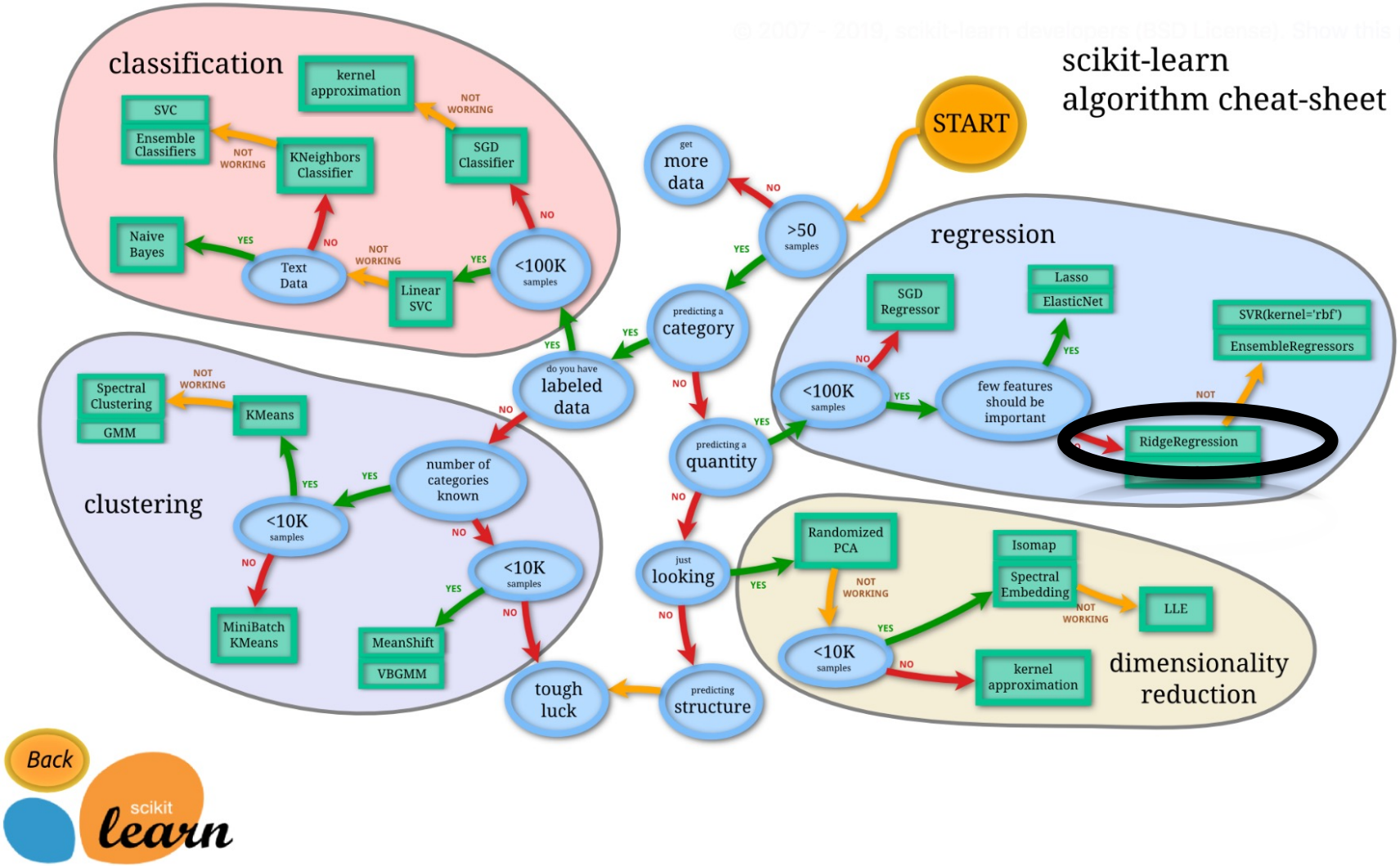
Pitches to Different Audiences

Geophysical Inversion

@ Data Science Springboard

Content developed between CORE Skills, Dr Morgan Williams & Dr Cericia Martinez | CSIRO

scikit-learn algorithm cheat-sheet



A *form* of ridge regression used to model subsurface properties

The regression variables represent:

- X: Physics kernels relating density to gravitational potential
- y: Measurements of gravitational field
- β: Estimates of subsurface density

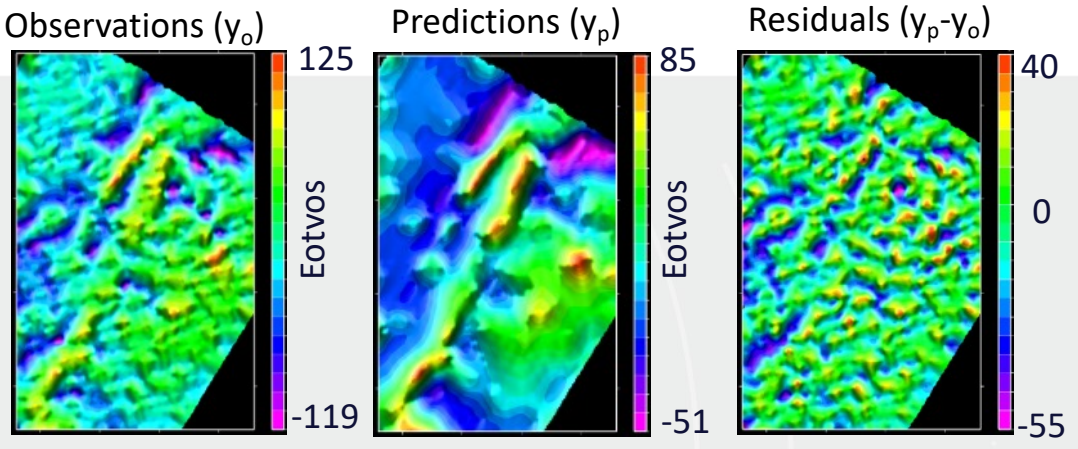
Fit Ridge regression model.

Parameters:

X : *{ndarray, sparse matrix} of shape (n_samples, n_features)*
Training data

y : *ndarray of shape (n_samples,) or (n_samples, n_targets)*
Target values

A *form* of ridge regression used to model subsurface properties



The regression variables represent:

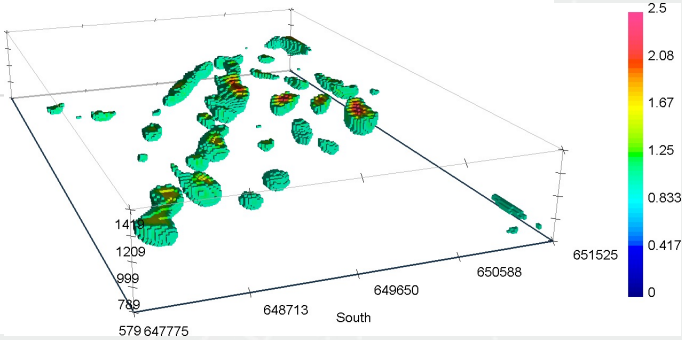
- X: Physics kernels relating density to gravitational potential
- y: Measurements of gravitational field
- β: Estimates of subsurface density

Fit Ridge regression model.

Parameters:

X : *{ndarray, sparse matrix} of shape (n_samples, n_features)*
Training data

y : *ndarray of shape (n_samples,) or (n_samples, n_targets)*
Target values



Pitches to Different Audiences

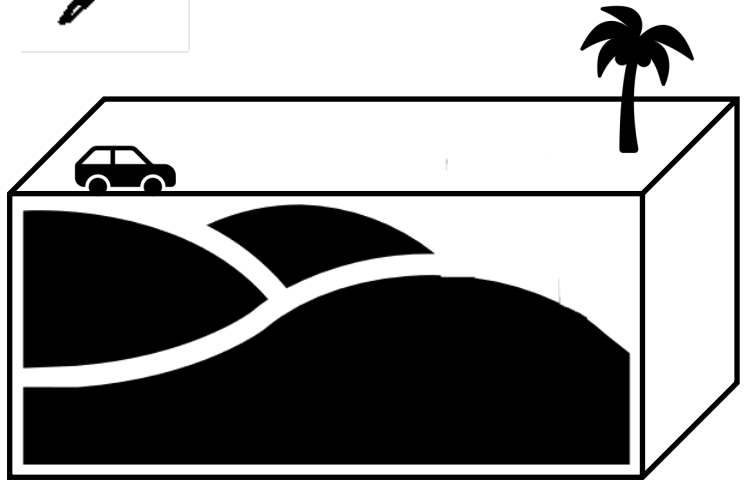
Geophysical Inversion

@ General Audience

Content developed between CORE Skills, Dr Morgan Williams & Dr Cericia Martinez | CSIRO

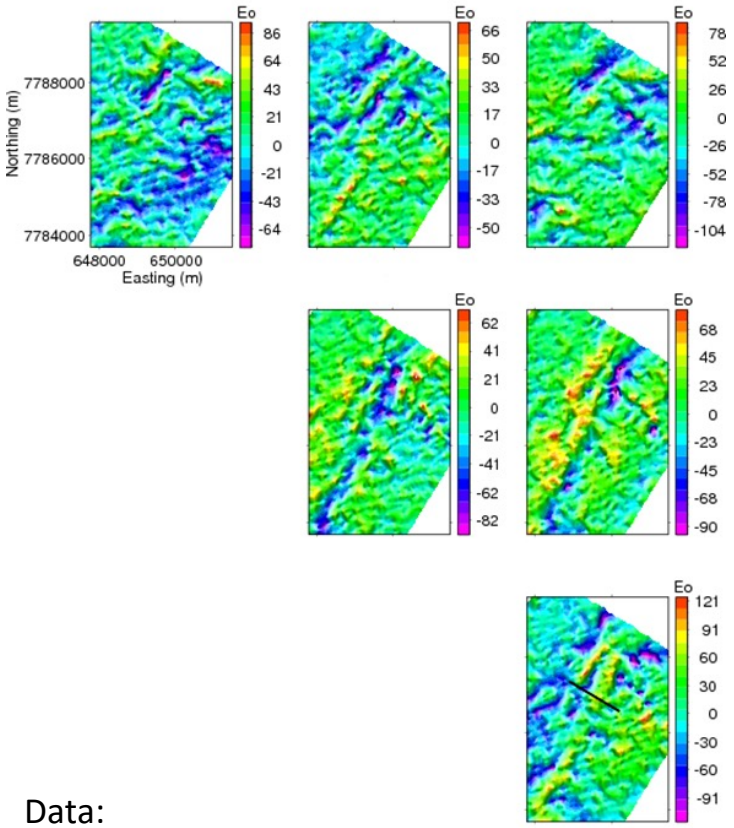
From data to subsurface models

Example from iron ore exploration



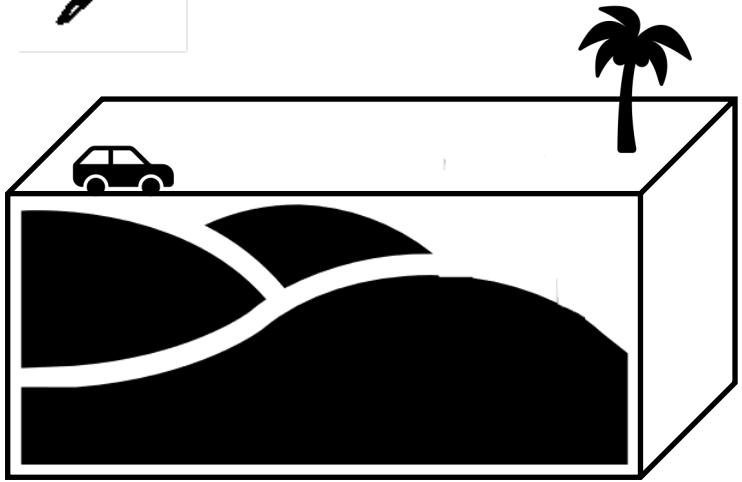
Subsurface geology

From data to subsurface models



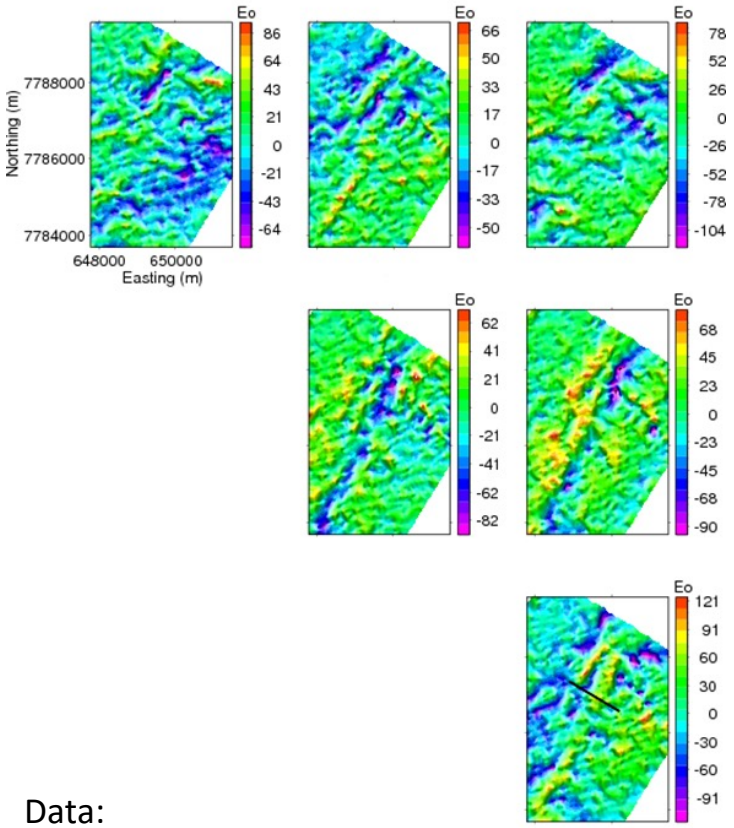
Data:
Airborne gravity gradiometry

Example from iron ore exploration



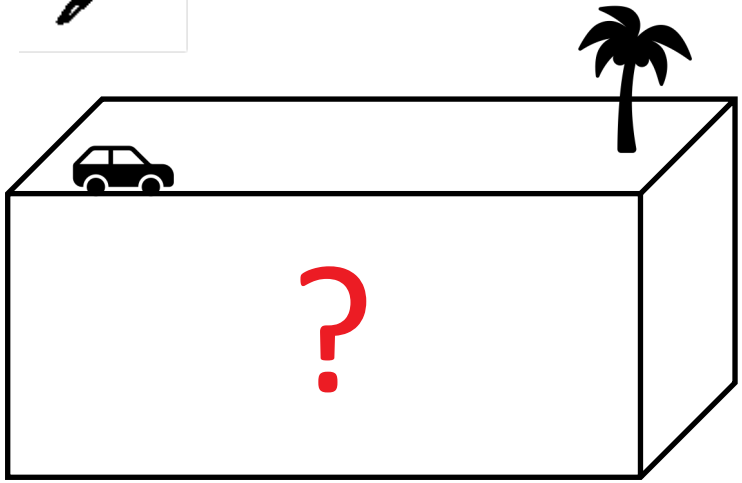
Subsurface geology

From data to subsurface models



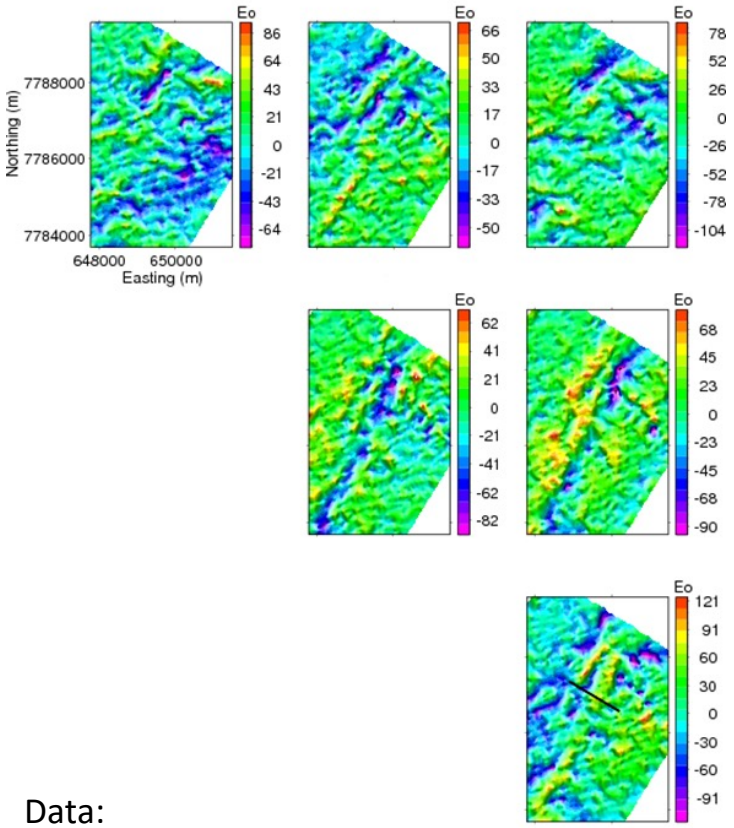
Data:
Airborne gravity gradiometry

Example from iron ore exploration



Subsurface geology

From data to subsurface models

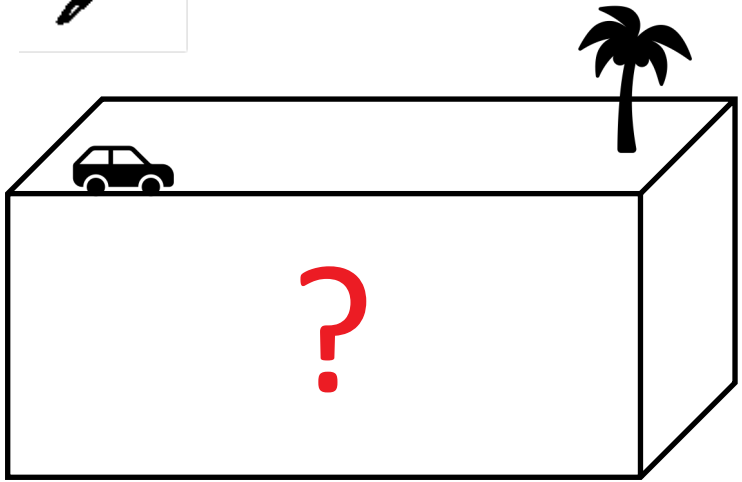


Data:
Airborne gravity gradiometry

Geophysical Inversion

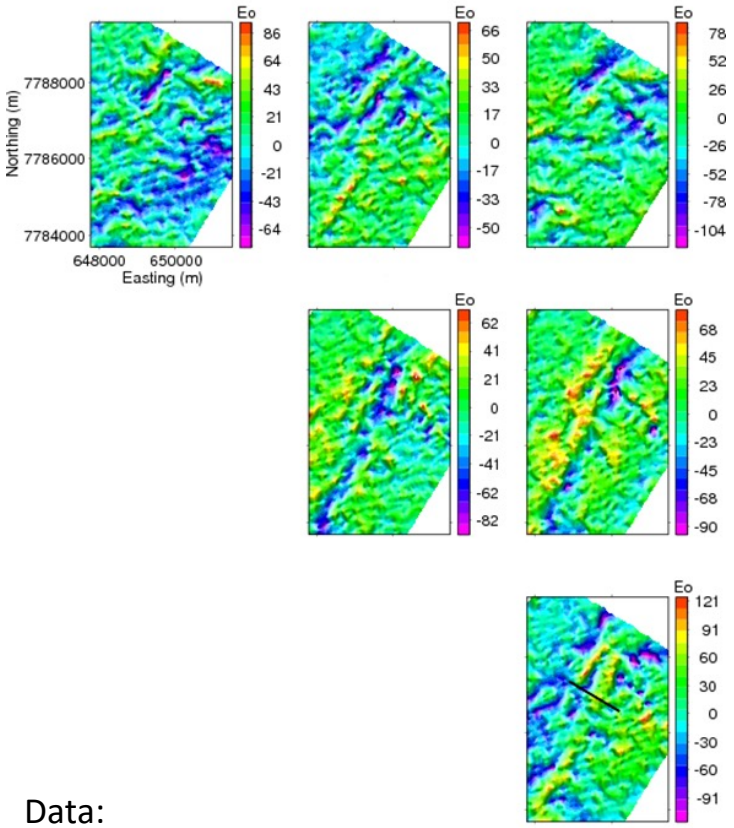
(calculate model)

Example from iron ore exploration



Subsurface geology

From data to subsurface models

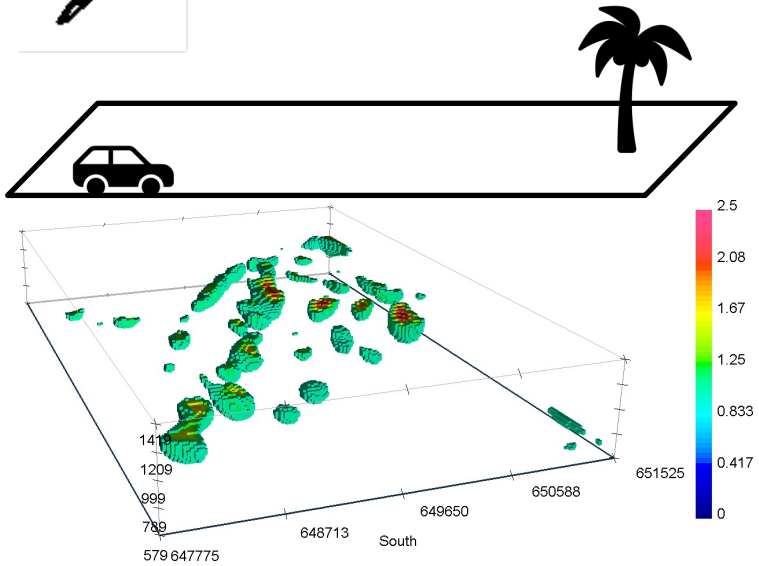


Data:
Airborne gravity gradiometry

Geophysical Inversion

(calculate model)

Example from iron ore exploration



Subsurface model:
3D density contrast

Martinez et al. (2013)

Telling Good Stories

(20 mins)

You’ve Got Three Minutes

Keep it Short

- Boil down your idea or message to a simple, direct form.
- Our 'so what’ becomes a paragraph which can be presented in just a few minutes (e.g. "elevator pitch).
- Its about them. Your pitch will be audience-specific!

But, increasingly time is at a premium and we may want to restrict it even more...

The Big Idea

From Nancy Duarte's *Resonate:*

- It must articulate your unique point of view.
- It must convey what's at stake.
- It must be a complete sentence.

- The shorter the better.
- Where relevant, make it personal.

A Big Idea?

Lunar Mission

A Big Idea

“The US should lead in space achievement, because it holds the key to *our* future on Earth.”

Facts a Story Maketh Not

- While a story often contains facts, they don't make it.
- A dot-point run through of a story illustrating just the key facts won't engage an audience – even if they’re the important bit.
- How you weave a narrative around them and connect to the audience is key.

Facts Won’t Win Friends

- If you're trying to convince someone, sometimes bombarding them with facts can be counter-productive.
- Especially where you have opposing views to start with (e.g climate change sceptics).
- Good stories are as much about inviting a conversation or questions from the audience as they are about conveying a message.

Structuring Narratives

A Dramatic Structure

- Beginning, Middle and End (à la Aristotle)
- Setup, Conflict and Resolution
- But what does this look like for us?

Structuring Narratives

A Dramatic Structure

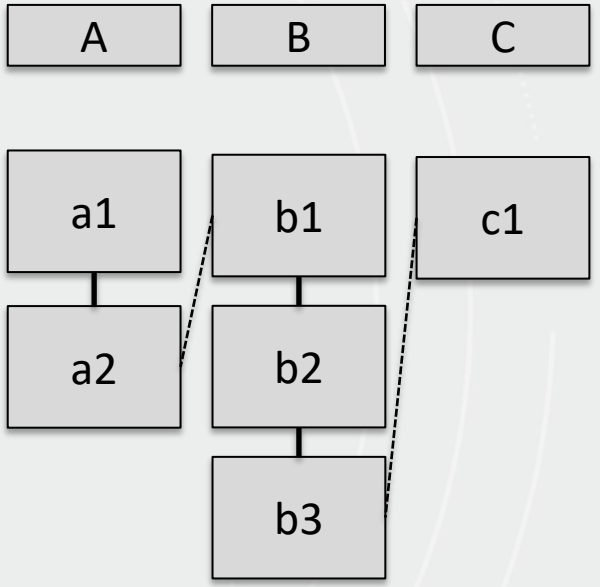
Setup	*The setting*: when, where and who?
Conflict	*The imbalance*: why is it necessary, what has changed?
	The balance: what do you want to see happen?
Resolution	*The solution*: how will you bring about the changes?
	The call to action.

Make sure you frame these in terms of your audience!

Storyboards

One Way to Outline Structure

- Starting with the elements of a broad narrative, how do you arrange the components?
- Typically components on loose paper/post-it notes.
- These can be as detailed as you want
- Today we'll be using them to structure narratives.
- **Where do you want your 'big idea'?** First? Last?



Refining Storyboards

- Check structure through headlines (horizontal logic)
- Making sure everything in a view is supportive (vertical logic)
- Reverse storyboard from a final product (start with headlines)
- Edit ruthlessly, keep it simple and be authentic
- Get a second opinion!

Stories from Data: Choosing Good Visualisations

(30 mins)

Why Apply Design to Visualisations?

Communication

- Make discrete arguments with visualisations
- Build visualisations to tell stories

Cognitive load

- Effort required to absorb the information

Data to Ink Ratio

- Declutter your plot!
- Complicated visuals are not conducive to telling stories, your audience won't know where to look.
- Start simple and highlight important features.
- If it’s visually complicated, it may *seem* difficult to understand - your audience are less likely to try!

Plot Types

What's the nature of the data?

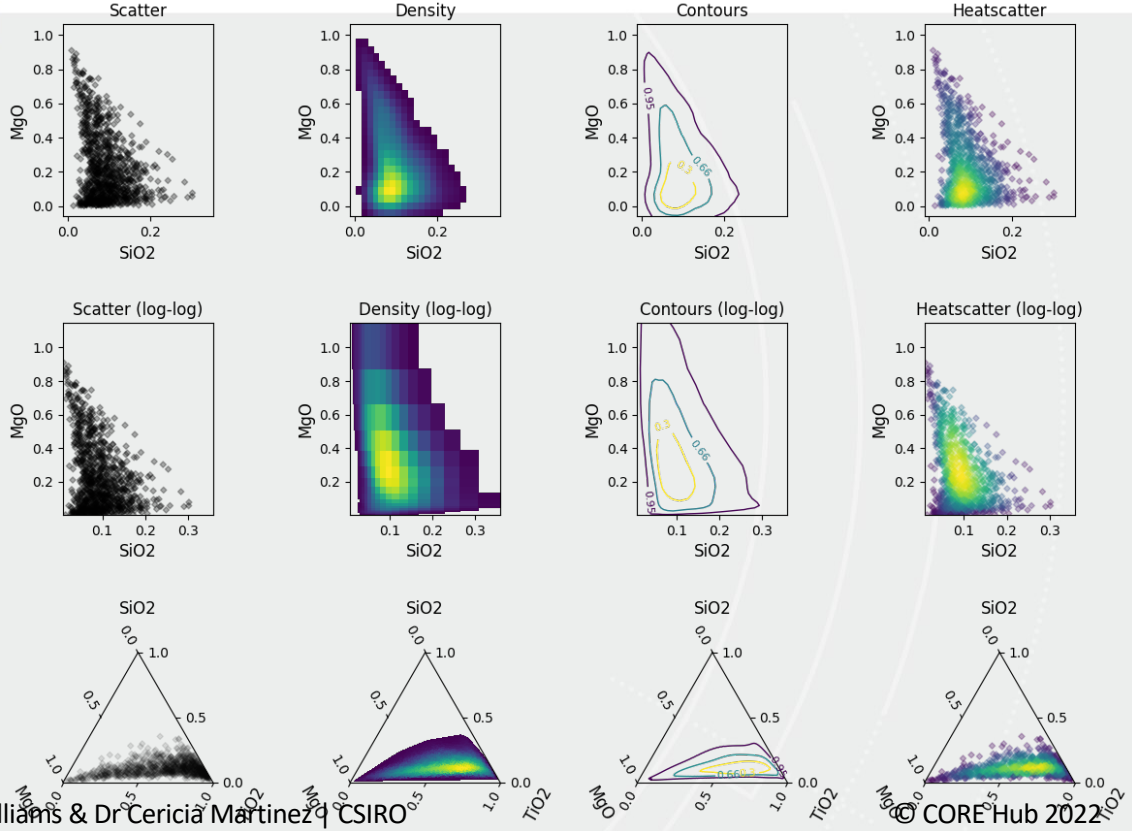
- Are you trying to visualise a summary, or the distribution?
- Is your data a continuous series (i.e. a line)?
- How will these things be *perceived*?
- What are humans good at judging? What are we not good at?

Avoid:

- 3D (depth perception)
- Multi-axes (confusing unless well labelled)
- Pie and area- or volume-based charts (aren't perceived well – maybe a stacked bar?)

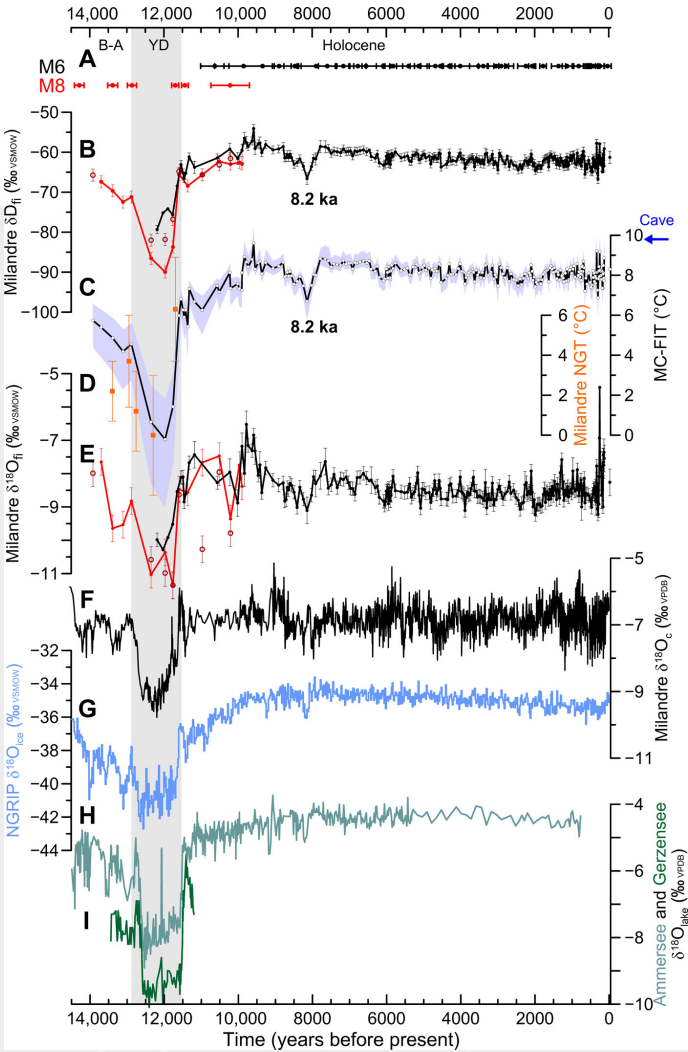
Plot Types

- There are multiple ways to represent the same data
- Avoid over-plotting!



Plot Types

- Sometimes climate scientists are bad at this
- As a reference figure… this is useful for highlighting a coincident event in multiple time-series
- But this isn't suited to direct communication!
- Consider how long it would take you to explain..



Content developed between CORE Skills, Dr Morgan Williams & Dr Cericia Martinez | CSIRO

Affolter et al. (2019)

Gestalt Principles

"Gestalt" follows from the German for form or shape. In psychology and design, it's used to refer to patterns or configuration.

The "Gestalt Principles" provide a framework for explaining why we see things the way we do, and in particular how we perceive the order of things. The principles include:

- **Proximity** (things closer are grouped)
- **Similarity** (things which appear more similar are grouped)
- **Enclosure** (things which share common enclosure are grouped)
- **Closure** (we prefer complete shapes, and will automatically fill gaps)
- **Continuity** (we create continuity in what we see, even if it does not explicitly exist)
- **Connection** (we group things which are connected)

Gestalt Principles

Good Figure

Objects groupped together tend to be perceived as a single figure. Tendency to simplify.

Proximity

Objects tend to be grouped together if they are close to each other.

Similarity

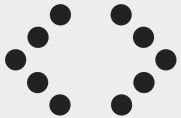
Objects tend to be grouped together if they are similar.

Continuation

When there is an intersection between two or more objects, people tend to perceive each object as a single uninterrupted object.

Closure

Visual connection or continuity between sets of elements which do not actually touch each other in a composition.



Symmetry

The object tend to be perceived as symmetrical shapes that form around their center.

Figure modified after UXcheat.com

Where are these messages being handled in the brain?

Attention and Memory

Iconic Memory

- The things you notice before you notice things.
- Pre-attentive = free-wins *or* quick losses.

Short Term Memory

- Low-capacity, but longer term. You can keep a few things in mind.
- *Three to five items (these should be the key points to your narrative)*
- Chunk the information!

Long Term Memory

- Visual and verbal memory.
- Not typically in use with regard to visualisation, but can be used to integrate external/bigger picture elements (e.g. via images, phrases).

Pre-Attentive Visual Properties

Colour

- Hue
- Intensity
- Contrast

Spatial position

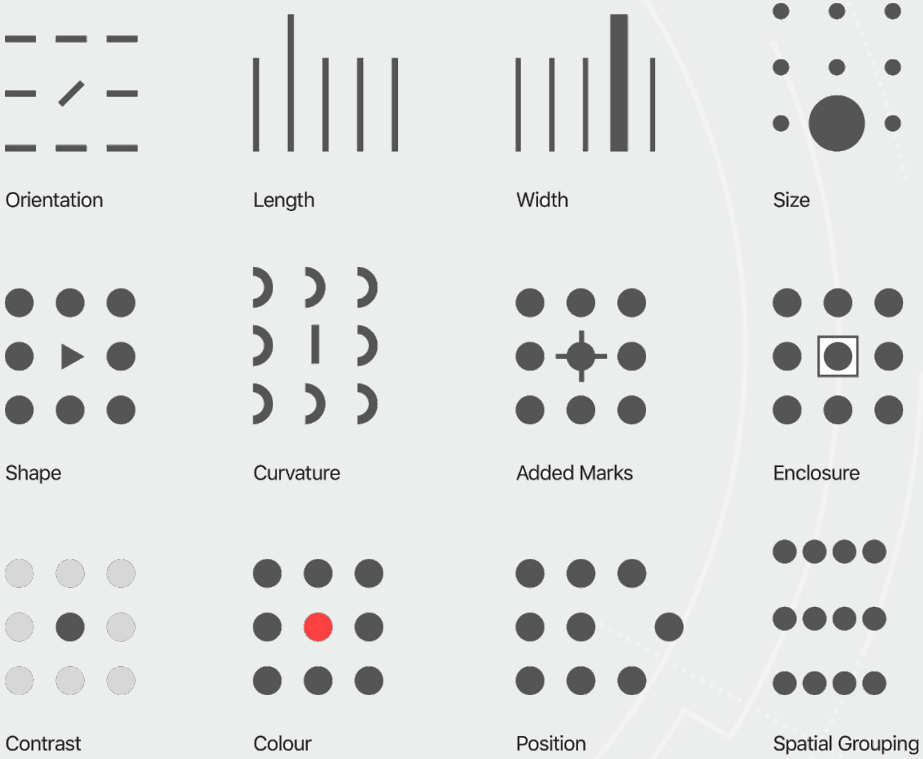
- Grouping or spatial associativity
- Connection

Motion

Form

- Orientation and co-linearity
- Shape
- Line length
- Size and width
- Curvature
- Added marks
- Enclosure

Pre-Attentive Visual Properties



Content developed between CORE Skills, Dr Morgan Williams & Dr Cericia Martinez | CSIRO

Figure modified after Healey & Enns, 2012

Using Pre-Attentive Visual Properties

A Few Things to Consider

- Sparingly – Focus the audience!
- You *can* use most of these for text. Consider that text will be observed differently (you read it!)
- Be aware that some colours, colour combinations and symbols might have cultural implications

Colour and Colourmaps

- Colour can be useful to highlight or divide
- But (mis)use of colour result in accessibility issues
  (e.g. consider the various types of colour-blindness)

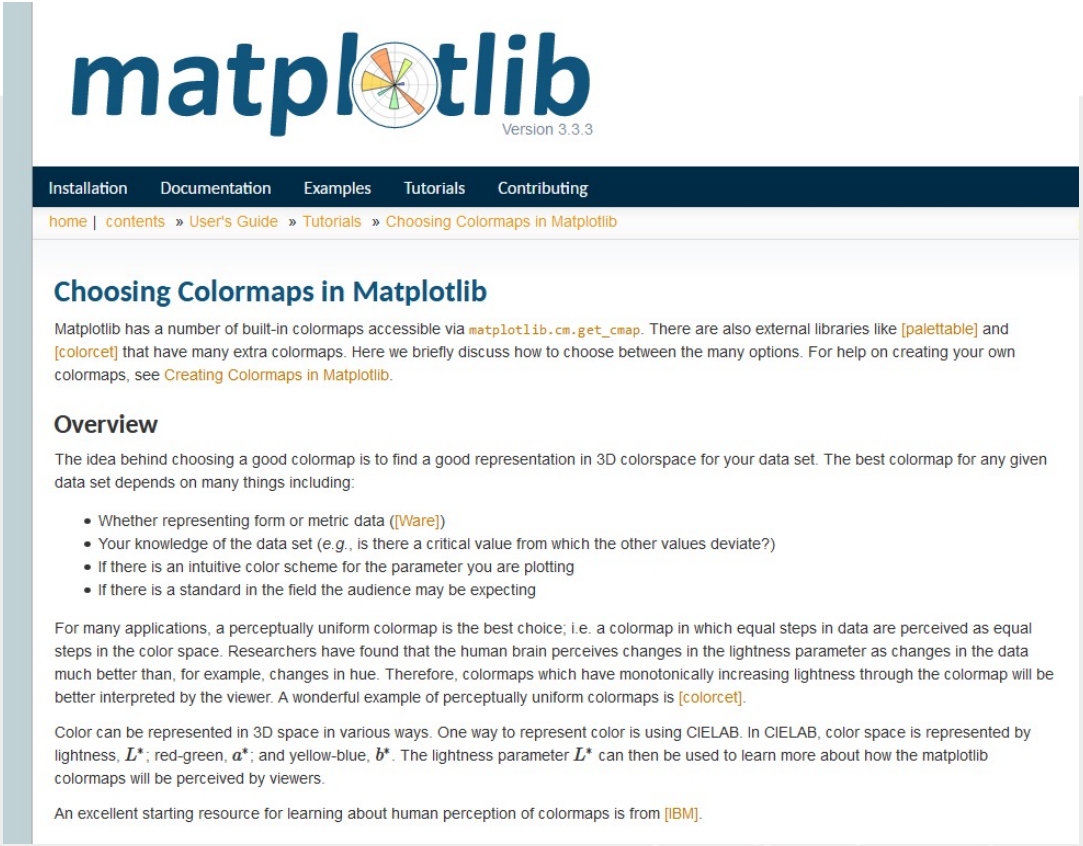
There are also significant implications for readability and perception, especially where you're mapping colours to non-categorical variables (i.e. using colormaps):

- Sequential, divergent or cyclic (e.g. is data centred about zero? Does it oscillate?)
- Linearity and perceptual uniformity

Using non-uniform colormaps can **distort** how your data appears, and result in **biases**!

Colour and Colourmaps

- See the `matplotlib` docs for an overview of choosing appropriate colourmaps and a reference of those available



matplotlib
Version 3.3.3

Installation Documentation Examples Tutorials Contributing

home | contents » User's Guide » Tutorials » Choosing Colormaps in Matplotlib

Choosing Colormaps in Matplotlib

Matplotlib has a number of built-in colormaps accessible via `matplotlib.cm.get_cmap`. There are also external libraries like [palettable] and [colorcet] that have many extra colormaps. Here we briefly discuss how to choose between the many options. For help on creating your own colormaps, see Creating Colormaps in Matplotlib.

Overview

The idea behind choosing a good colormap is to find a good representation in 3D colorspace for your data set. The best colormap for any given data set depends on many things including:

- Whether representing form or metric data ([Ware])
- Your knowledge of the data set (*e.g.*, is there a critical value from which the other values deviate?)
- If there is an intuitive color scheme for the parameter you are plotting
- If there is a standard in the field the audience may be expecting

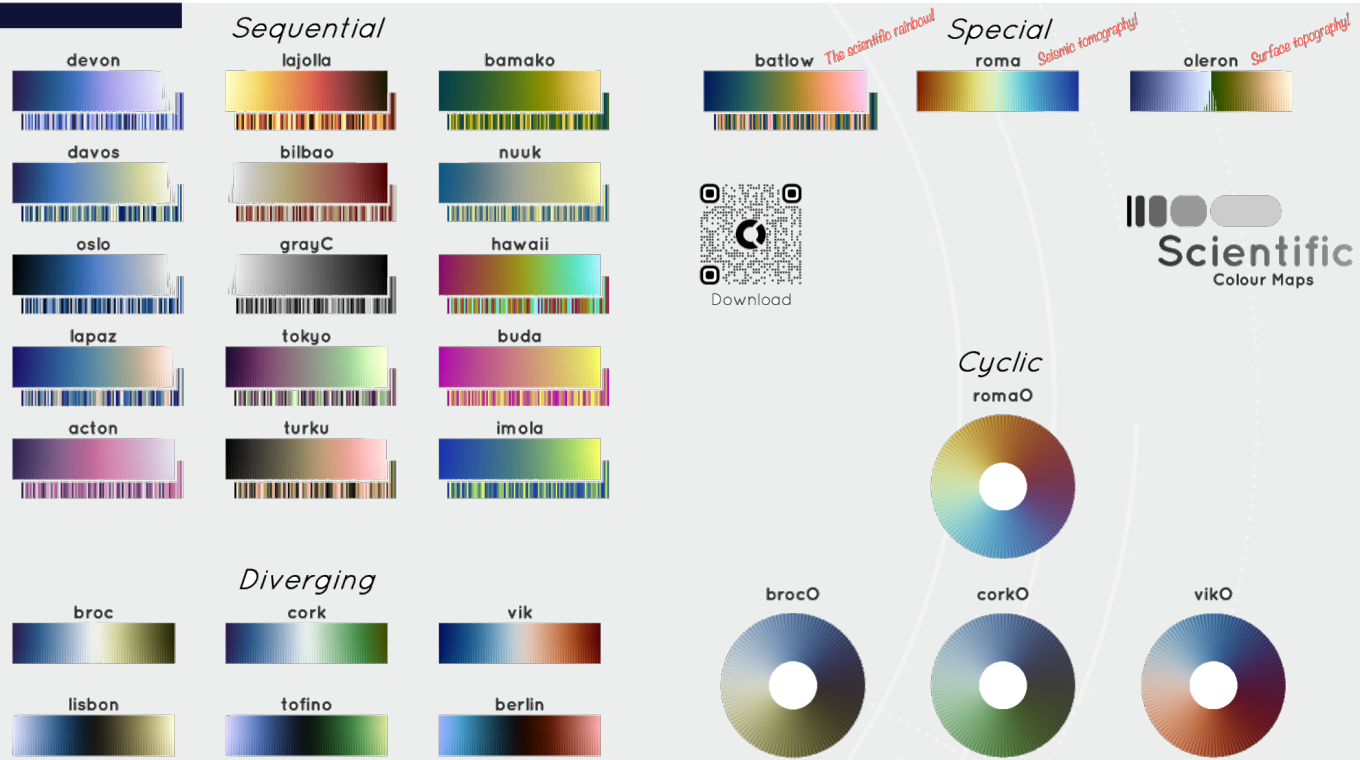
For many applications, a perceptually uniform colormap is the best choice; i.e. a colormap in which equal steps in data are perceived as equal steps in the color space. Researchers have found that the human brain perceives changes in the lightness parameter as changes in the data much better than, for example, changes in hue. Therefore, colormaps which have monotonically increasing lightness through the colormap will be better interpreted by the viewer. A wonderful example of perceptually uniform colormaps is [colorcet].

Color can be represented in 3D space in various ways. One way to represent color is using CIELAB. In CIELAB, color space is represented by lightness, L^*; red-green, a^*; and yellow-blue, b^*. The lightness parameter L^* can then be used to learn more about how the matplotlib colormaps will be perceived by viewers.

An excellent starting resource for learning about human perception of colormaps is from [IBM].

Colour and Colourmaps

- There are other options beyond `matplotlib`, if you're interested – like these from Fabio Crameri



Aesthetics

- Time spent making visualisations aesthetically pleasing can be time well spent.
- Design elements should all be intentional and used sparingly.
- White space is important.
- Alignment and symmetry are also important.
- If you aren't particularly comfortable with visual design, find something you find appealing, deconstruct the design, and perhaps and attempt to replicate it.

Interactivity and Affordance

Interactive Plots are Different – You're Giving Over Some Control

- How do you guide your audience?

Affordance refers to the properties or features of something which hints towards its use.

- For most tools you'll work with, some of these affordances will be built-in.
- It's worth considering accessibility too!
- *We'll go through some examples of this in the next section.*

Summary: Good Data Visualisations

Make it Pretty but Keep it Simple

- Acknowledge the Nature of Your Data
- Free wins: Use pre-attentive signals to emphasise something (or de-emphasise others).
- Well-designed visualisations will be more engaging.
- If you’re going for interactivity, affordances and visual ‘hints’ are important.

Schedule

DAY 7

AWST	AEST	Agenda	Facilitator
07:30	**09:30**	**Q&A, Issues & Announcements**	Jeremy
07:45	09:45	**Clustering & Regression in Higher Dimensions (DAY 6 continued, from model NB. Partial Least Squares as homework)**	Jeremy
08:45	10:45	**Data Visualisation & Communication**	Ying
09:15	11:15	*Morning Tea*	
09:30	11:30	**Data Visualisation & Communication continued**	Ying
11:00	13:00	*Lunch*	
11:45	13:45	**Data Stories**	Ying
12:30	14:30	**Building Storyboards**	Ying
13:15	15:15	*Afternoon Tea*	
13:30	15:30	**Project Time**	All
14:00	16:00	**Review Storyboards**	Ying
14:30	16:30	**Closeout –Takeaways & Mid-Program Review**	Ying/Tamryn
14:55	16:55	**Menti** Feedback	Tamryn
15:00	17:00	**Close**	

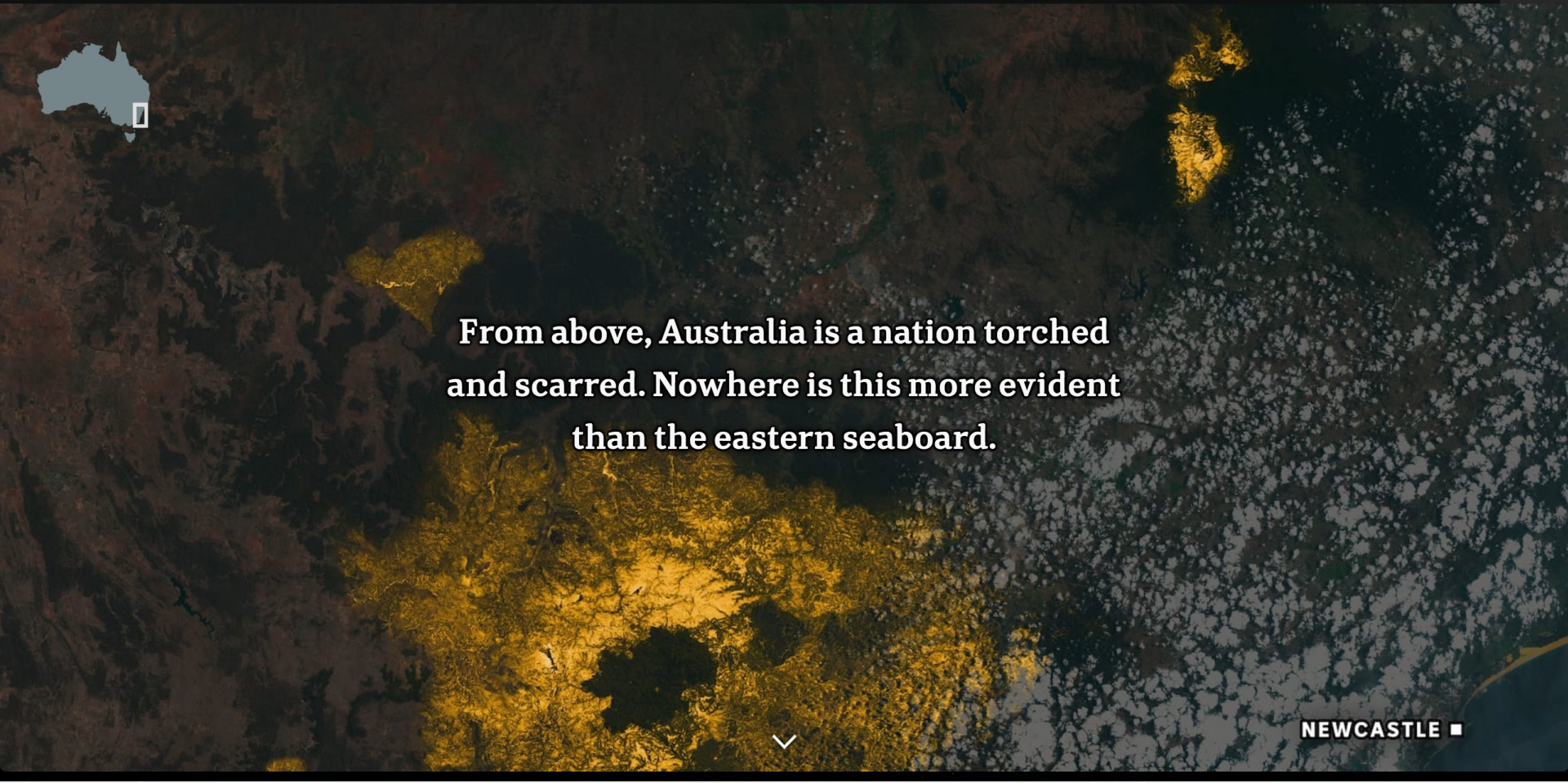
Journalistic Data Stories

Journalistic Data Stories

- We will look at some data-driven articles with the aim of understanding how design elements discussed this morning tell a story
- Read / Pair / Share Agenda:
  - Individually (~20 min): Read assigned story and summarise
  - Pair (~10 min): Share summaries in small groups
  - Group (~3 min per group): Nominate a group representative to share collective summary with all of us in a 3-min summary

Journalistic Data Stories

https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6


From above, Australia is a nation torched and scarred. Nowhere is this more evident than the eastern seaboard.
NEWCASTLE

Deep reading of a professional story

Prepare the following to share with others

- Three bullet points (for setup, conflict and resolution)
- ‘Big Idea’ sentence for this story to summarize it for everyone
- One example of good use of technical elements
  - structure, design, visualisation or narrative
- One example where you thought the authors could improve.

Framework to investigate data design elements

	Questions to investigate how the story has been put together
Context	❓ How have the authors identified their audience?
	❓ What knowledge do they assume the readers know, what new concepts are they going to introduce?
Structure	❓ How is the story structured?
	❓ Is the takeaway up front? Or does it get revealed gradually?
	❓ Have the authors used a setup/conflict/resolution structure the narrative?
	❓ If not, if so, how are they using tension or conflict to maintain your interest?
Narrative	❓ How have the authors made the data personal or relevant for their audience?
	❓ What techniques have they used to make the narrative memorable?

Framework to investigate data design elements

Data Visualisation Interrogation
❓ How do the team walk readers through the story in this data viz?
❓ How do they provide structure to the data (e.g. consider Gestalt principles)?
❓ How do they focus the audience's attention (e.g. consider pre-attentive attributes)?
❓ If interactive, how do they use *design affordances* for the audience to interact with the data?

Gestalt principles: Proximity, Similarity, Enclosure, Closure, Continuity, Connection. ↩️

Pre-attentive attributes: Orientation, Colour, Size, Curvature or Added Marks, Intensity, Location, Motion. ↩️

Deep reading of a professional story

Prepare the following to share with others

- Three bullet points (for Setup, Conflict and Resolution)
- ‘Big Idea’ sentence for this story to *summarise* it for everyone
- One example of good use of technical elements
  - structure, design, visualisation or narrative
- One example where you thought the authors could improve

Link to activity overview, framework, and assigned data stories:

https://github.com/core-skills/07-data-storytelling/blob/master/program/02_datastories.md

Communication Design & Storyboarding

Putting Together a Data-Driven Storyboard

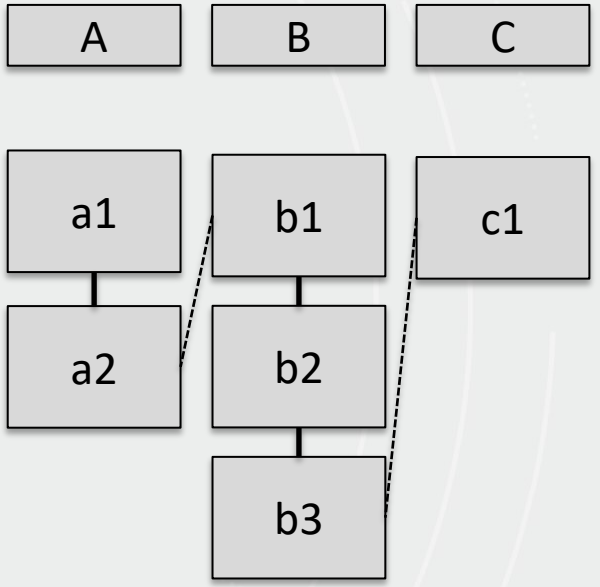
- **What's the narrative behind your Capstone Project?**
- What can you extract from your dataset to frame the problem?
- Who's the audience?
- How do you want your audience to respond?
- What's your call to action?

These might be useful for your capstone overviews in a few weeks time!

Storyboards

One Way to Outline Structure

- Starting with the elements of a broad narrative, how do you arrange the components?
- Typically components on loose paper/post-it notes.
- These can be as detailed as you want.
- Today we'll be using them to structure narratives.
- **Where do you want your 'big idea'?** First? Last?



Putting Together a Data-Driven Storyboard

You have 90 minutes:

- Put together 1 slide to present (90 seconds)
- Focus on the introduction/set up of your Capstone Project: What is the Problem you are trying to solve.
- Once you've finished, spend the rest of the time working on project

Putting Together a Data-Driven Storyboard

- Aim for **90 second** presentation
- **Go back to some of the notes** from this morning about context, audiences, narrative structure and visualisations
- Consider where ethical concerns might come in
  (who are you talking to? who are you talking about?)
- Build it up, and slim it down – keep it simple and to the point.

Schedule

AWST	AEST	Agenda	Facilitator
07:30	**09:30**	Q&A, Issues & Announcements	
07:45	09:45	**Data Visualisation & Communication**	Morgan
09:15	11:15	*Morning Tea*	
09:30	11:30	**Data Stories**	Cici
10:15	12:15	**Building Storyboards**	Morgan
11:00	**13:00**	***Lunch***	
11:45	13:45	**Building Storyboards Cont.**	
13:15	15:15	*Afternoon Tea*	
13:30	15:30	**Storyboards and Review**	Cici
14:45	16:45	**Closeout**	

Storyboarding & Project Session

Storyboard Presentations

Storyboard sharing

Share Capstone Pitches/Storyboards

- 90 seconds to present a pitch
- What's your "big-idea" for your Capstone Project?
- Who do you need to convince, and what do you want/need from them?
- An image or visualisation framing the problem or solution
- Note any thoughts around how you can structure and frame the pitch to engage the audience, and any visual aspects which might be useful for this or the next draft

Review: First 7 Days

Review & Reflection

Honest critique of the first 7 days

- Onboarding from your prior knowledge comfortable?
- Did the pre-requisite prepare you for the first module of the program?
- Have we pitched the content at the right level? Enough detail? Have we assumed too much?
- Pace – what too fast what too slow what just right?
- Do you see how things relate to your organization function?
- Do you see areas in your own work you can apply these steps?
- Have you had enough examples and opportunities to practice (which)?
- Are we meeting your expectations?

Project Time

CORE Skills

A CORE partnership

COREHUB.COM.AU/SKILLS