Visual Analytics – Welcome!

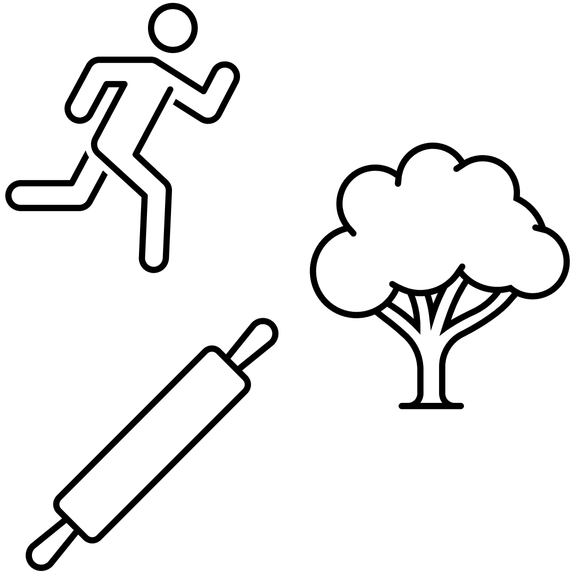
Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Jordan Crouser (https://jcrouser.github.io/)

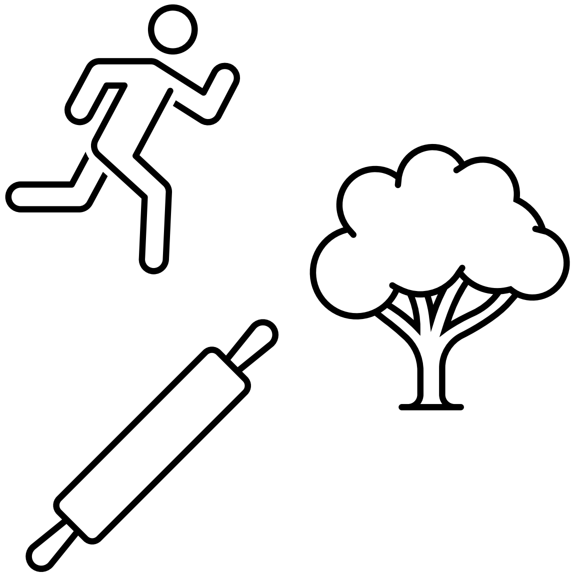
Plan for Today

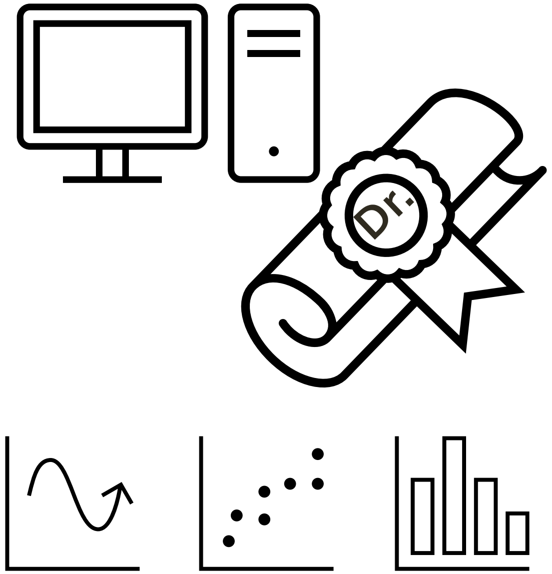
- Intros
- Big problem
- Quick history lesson
- Visual analytics: definition
- Discussion: how does it help?
- Structure of this course

Who Am I?

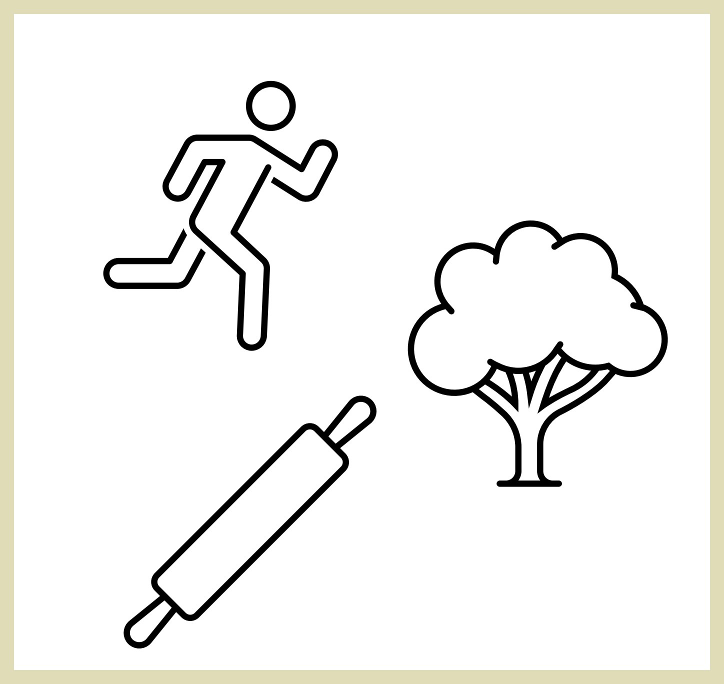


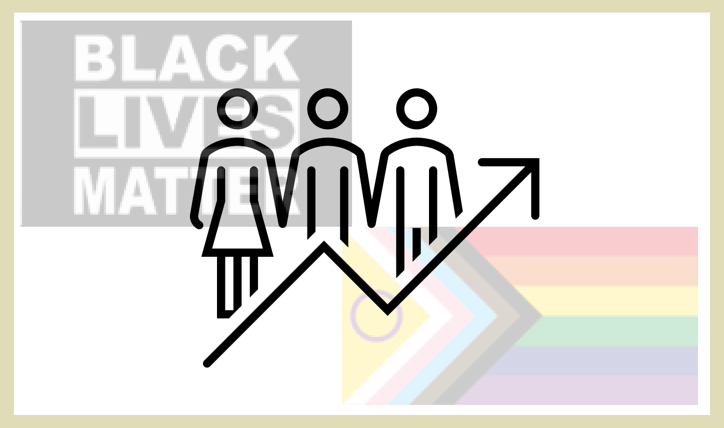
Who Am I?

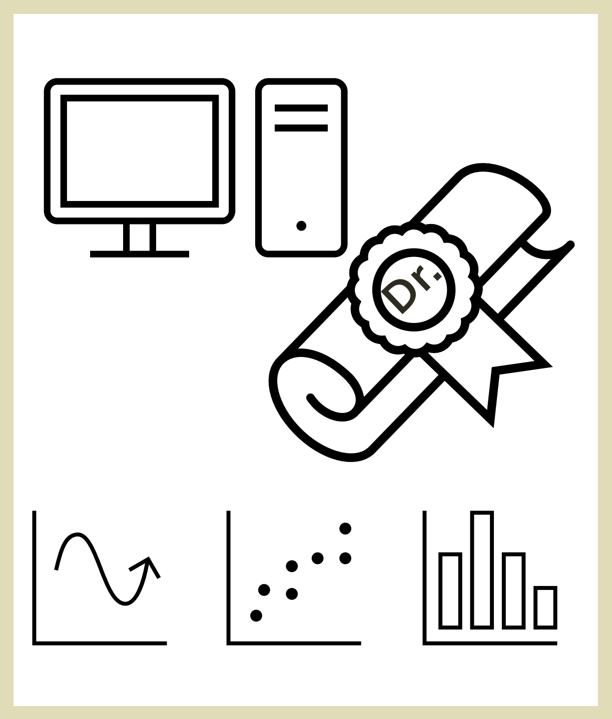




Who Am I?







Who Are You?

- Form groups of 3
- Introduce yourselves (name, pronouns)
- Share:
  - A highlight of your summer break
- Find 1 thing that your entire group has in common (favorite color? hometown? left-handed? Be creative!)
- After about 5 minutes we will go around, introduce ourselves, and share what each group has in common

Who Are You?

- Form **new groups** of 3 (move around!)
- Introduce yourselves (name, pronouns)
- Share:
  - Would you rather live with Cheeto dust on your fingers forever no matter what OR live with slightly damp clothes?
- After about 5 minutes we will go around, introduce ourselves, and share our would you rather answers

Who Are You?

- Form **new new groups** of 3 (move around!)
- Introduce yourselves (name, pronouns)
- Share:
  - Would you rather be able to touch a book and instantly read it but have your current retention OR read at your current speed but remember everything?
- After about 5 minutes we will go around, introduce ourselves, and share our would you rather answers

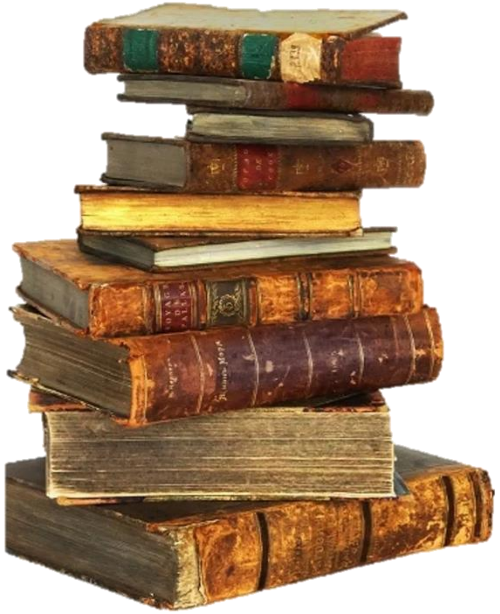
- Name tags!

About this course

What are some non-traditional forms of data?

About this course

What are some non-traditional forms of data?



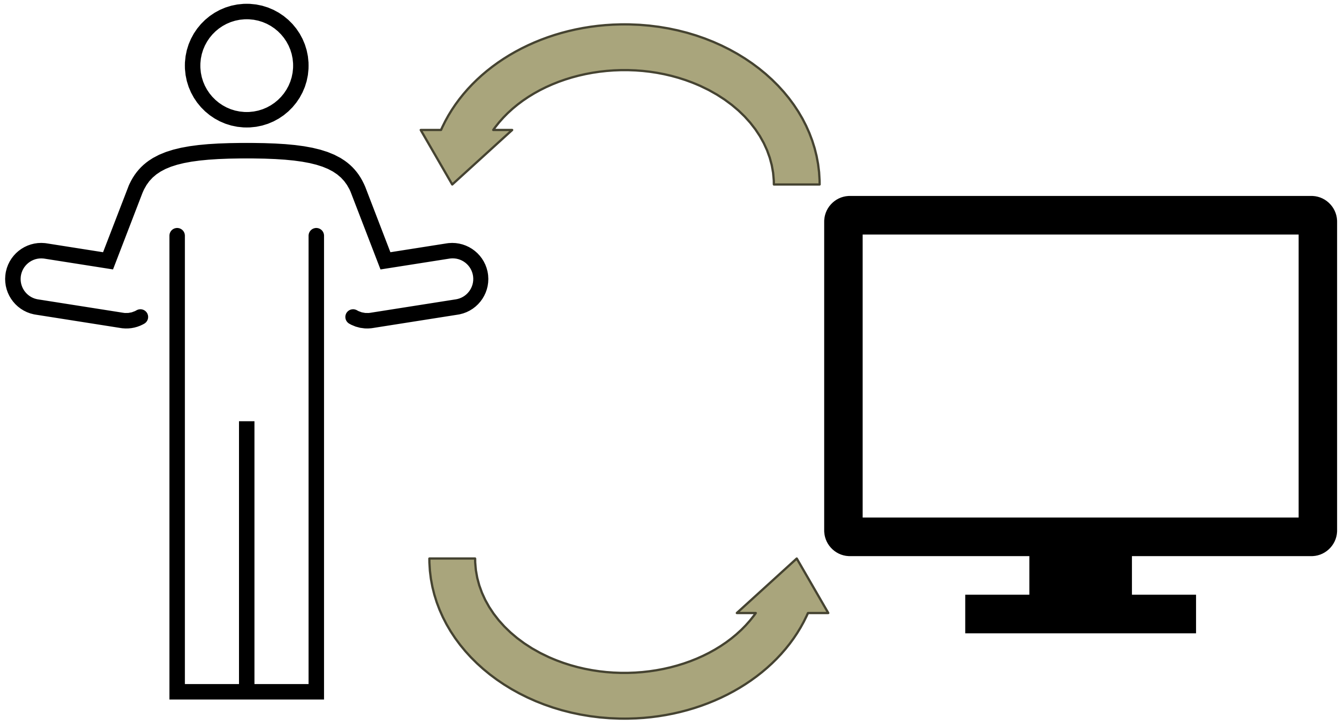
About this course

The problem:

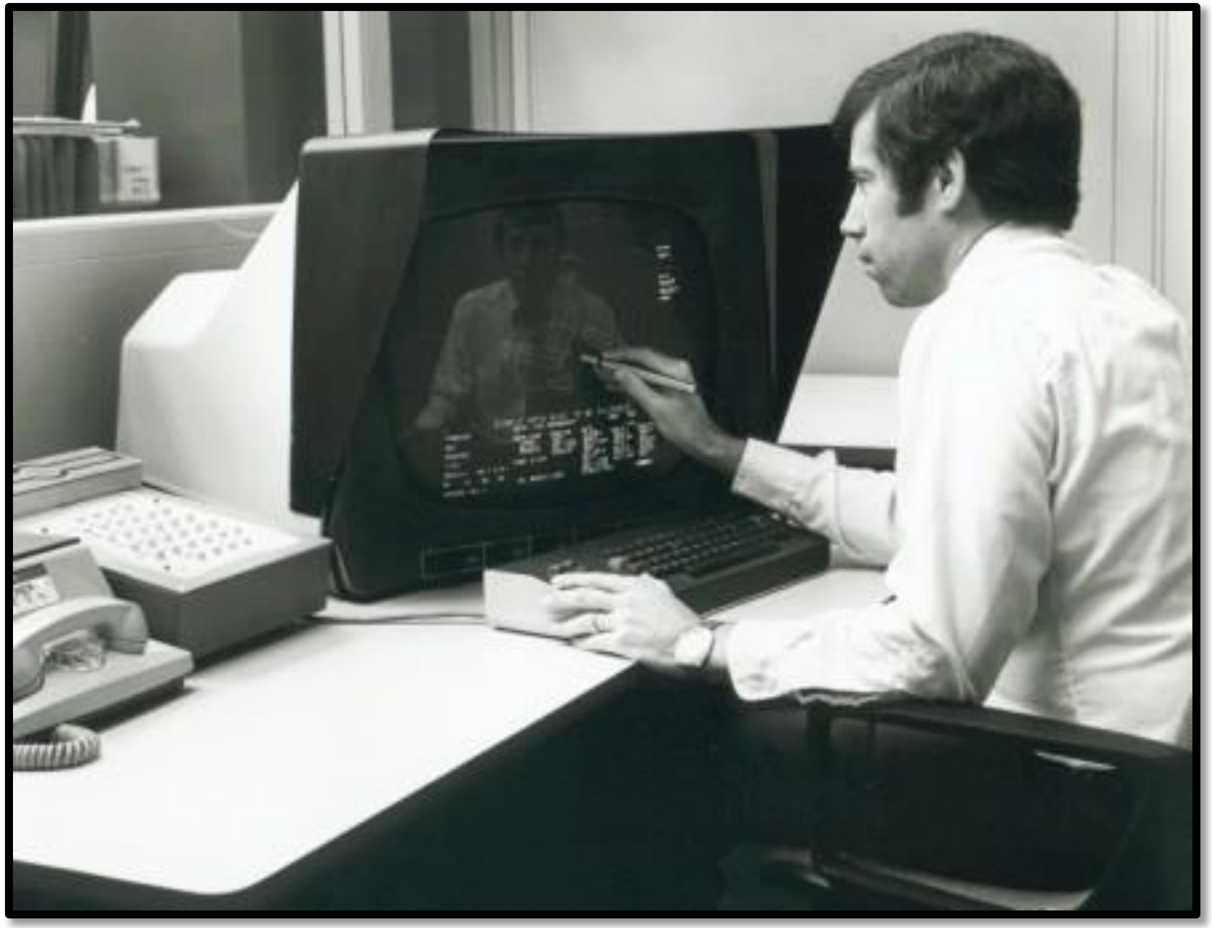
- We’re collecting and generating data **faster** than traditional methods can keep up
- Not only is the data unmanageably big, it’s usually **noisy and ambiguous** too ➔ requires interpretation
- **Major problem**: humans don’t scale

About this course

What are the strengths of each wrt data?

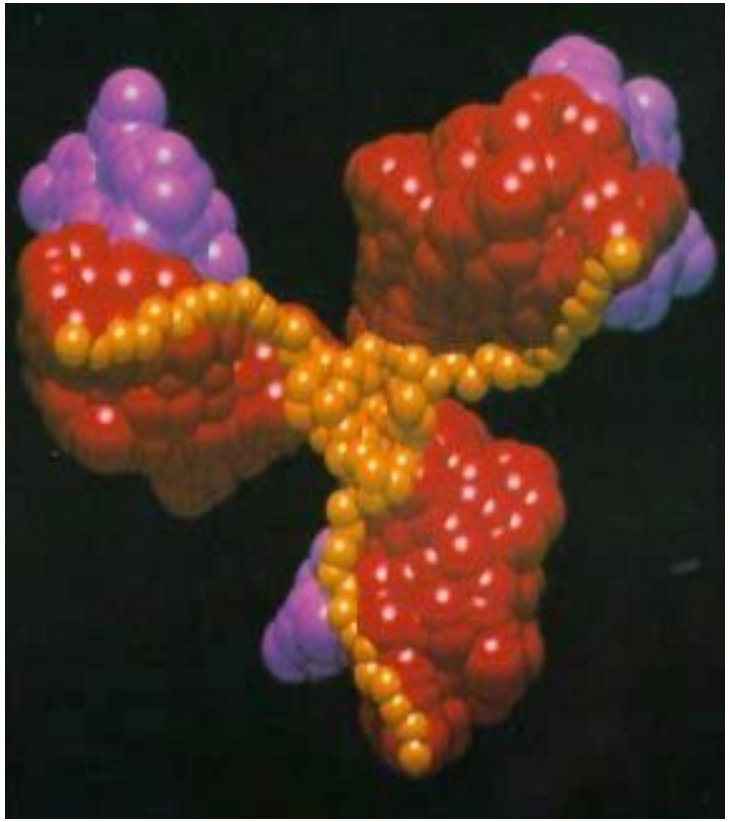


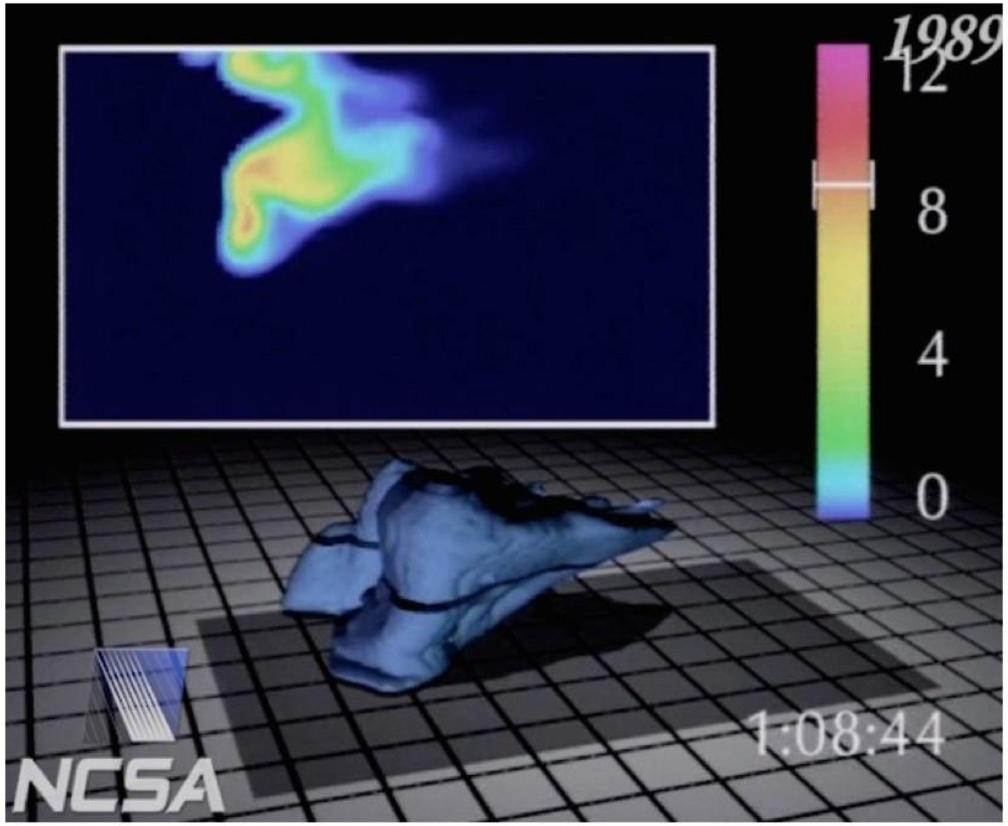
(Incomplete) History of Visual Analytics: 1970s



- CAD/CAM, building cars, planes, chips
- Starting to think about: 3D, animation, edu, medicine

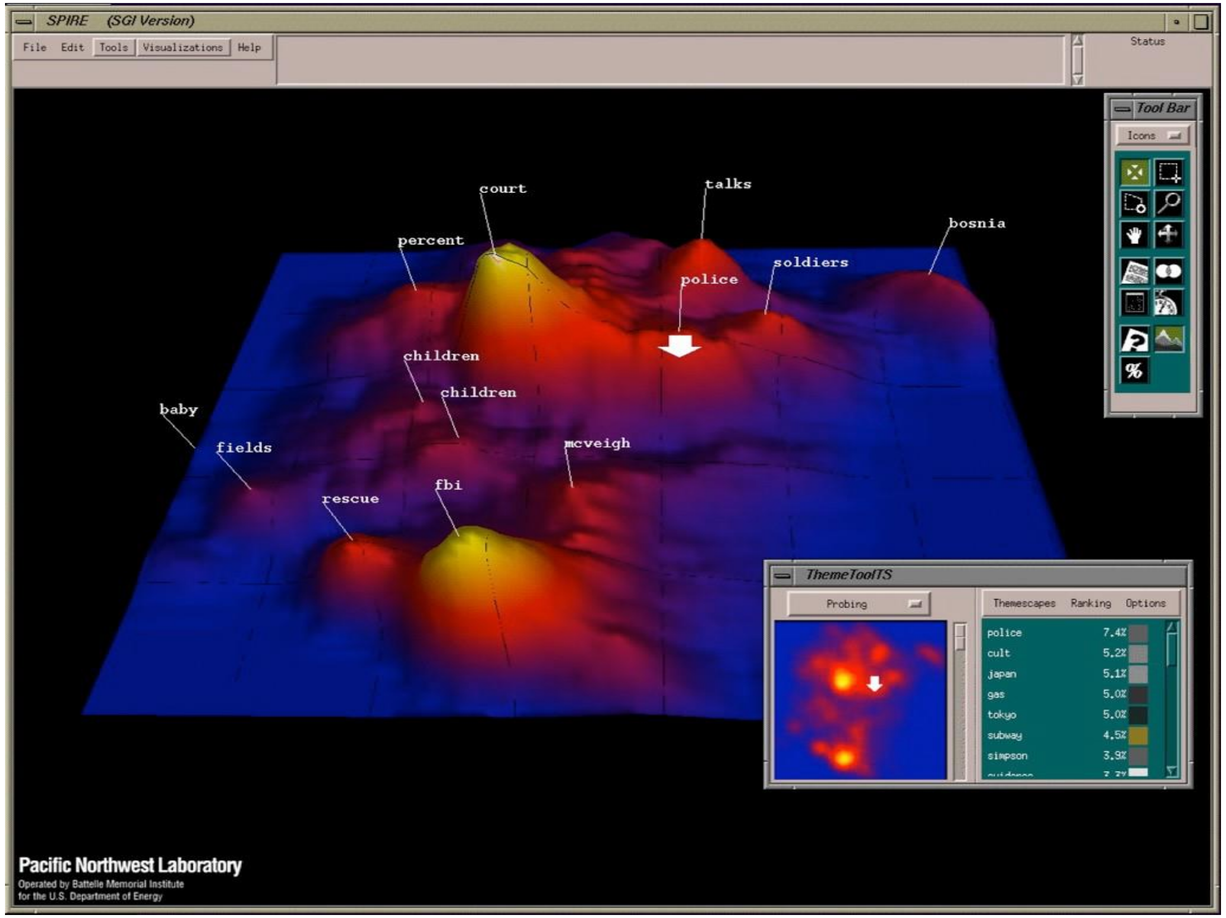
(Incomplete) History of Visual Analytics: 1980s





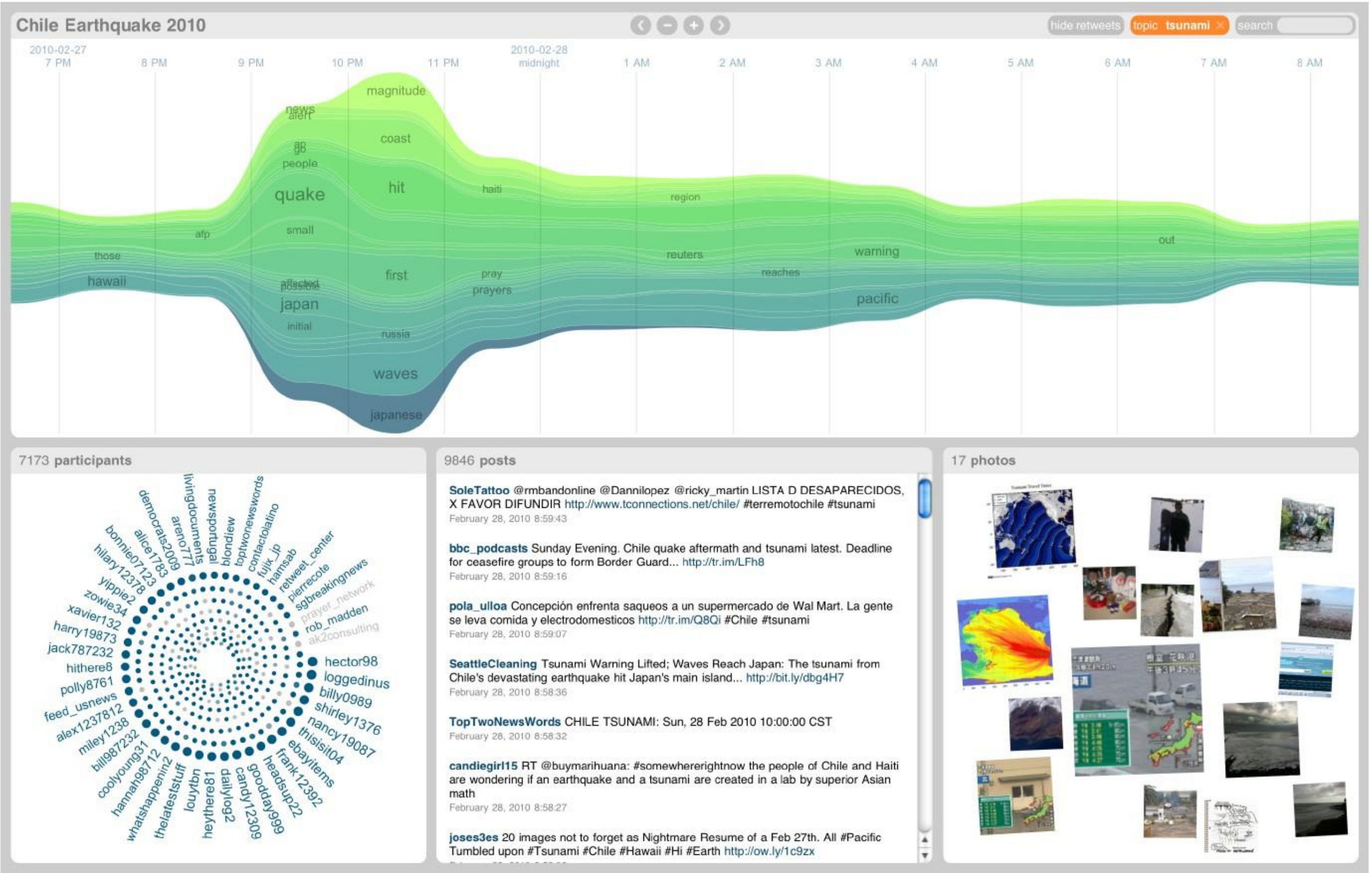
- Scientific visualization, physical phenomena
- Starting to think about: photorealism, entertainment

(Incomplete) History of Visual Analytics: 1990s



- Information visualization, storytelling
- Starting to think about: online spaces, interaction

(Incomplete) History of Visual Analytics: 2000s



- Coordination across multiple views, interaction
- Starting to think about: sensemaking, provenance

(Incomplete) History of Visual Analytics

- Early 2000s: US is reacting to 9/11
- 2003: Dept. of Homeland Security (DHS) est.
- DHS objectives:
  - Prevent terrorist attacks within the US
  - Reduce US vulnerability to terrorism
  - Minimize damage / aid recovery from attacks that do occur
- 2005: DHS charters the **National Visualization and Analytics Center** (NVAC) at PNNL

(Incomplete) History of Visual Analytics

NVAC mission:

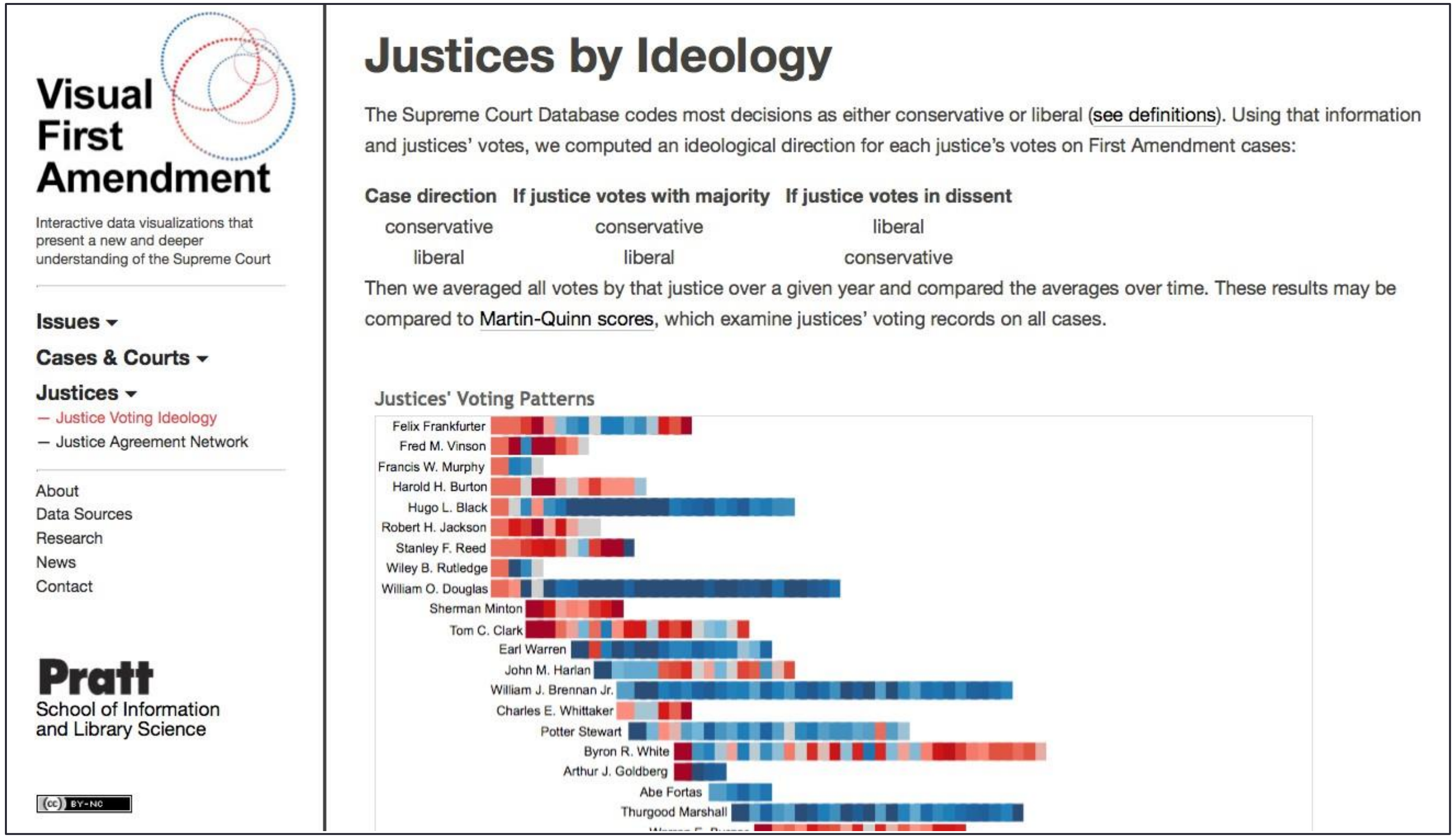
Develop advanced information technologies to support the Homeland Security mission with data that is massive, complex, incomplete, and uncertain in scenarios requiring human judgment.

Challenges:

- How do we support analytical reasoning under complex, changing circumstances?
- How do we make use of domain expertise, when domain experts are not computer scientists?

New idea: (visualization) ∩ (analytics)

(Incomplete) History of Visual Analytics: 2010s



- Human-machine collaboration, machine learning
- Starting to think about: evaluation, new media, **DH**

Visualization (def.)

Creating visual **representations** of data to reinforce human **cognition**

Analytics (def.)

Discovery and **communication** of meaningful patterns in data

Visual Analytics (def.)

"The science of **analytical reasoning** facilitated by **interactive visual interfaces**"[1]

[1]Thomas, James J., and Kristin A. Cook, eds. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.

Big idea

Humans and machines have **complimentary strengths**

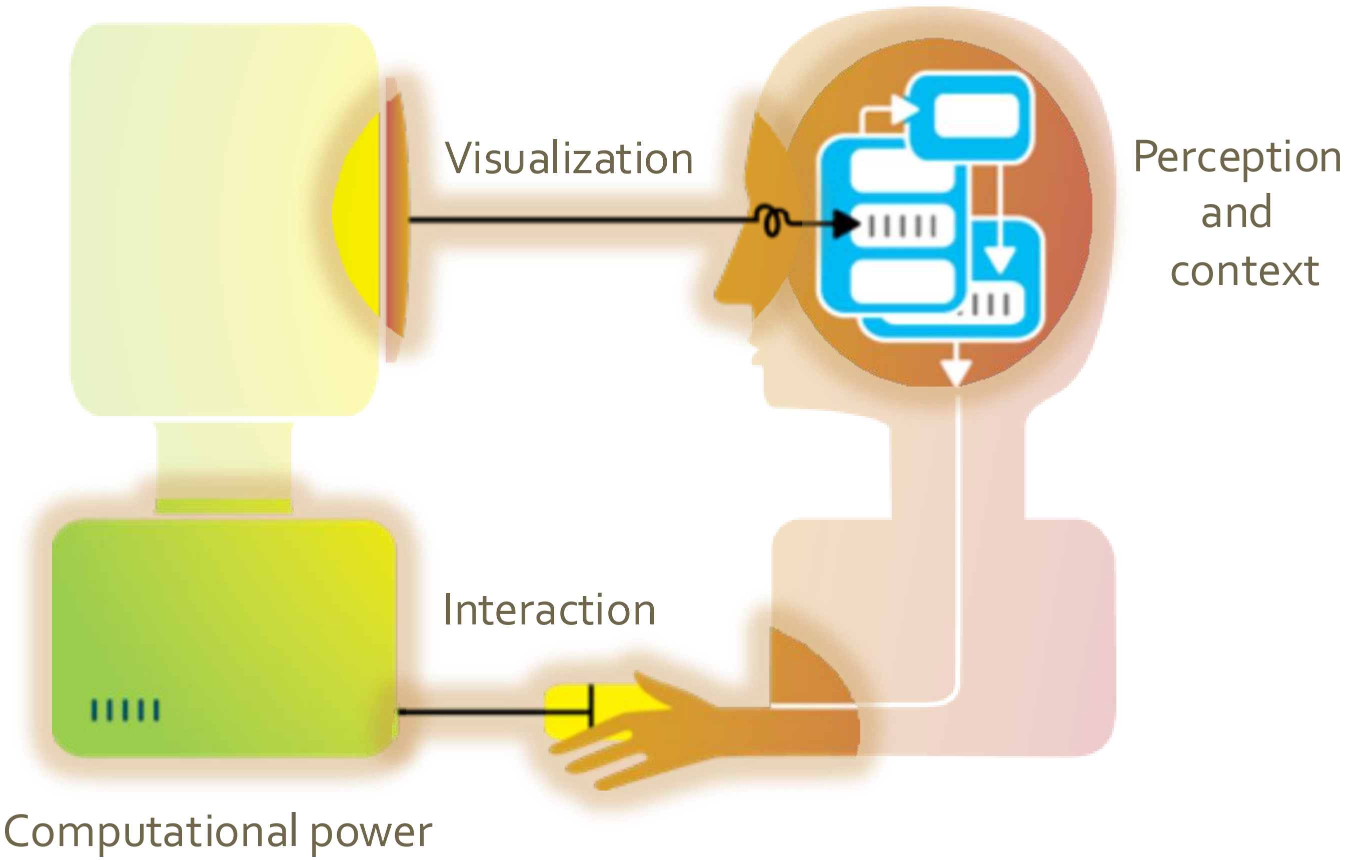


Image credit: Ali Ansari

VA: Human + Machine Collaboration

- Make use of what both humans and machine strengths using **interactive visual interfaces** to facilitate the "conversation"

- Key considerations:
  - Traditional **analytics** (stats, machine learning) can help make massive data tractable
  - Use what we know about **cognition and perception** to help humans put data in context

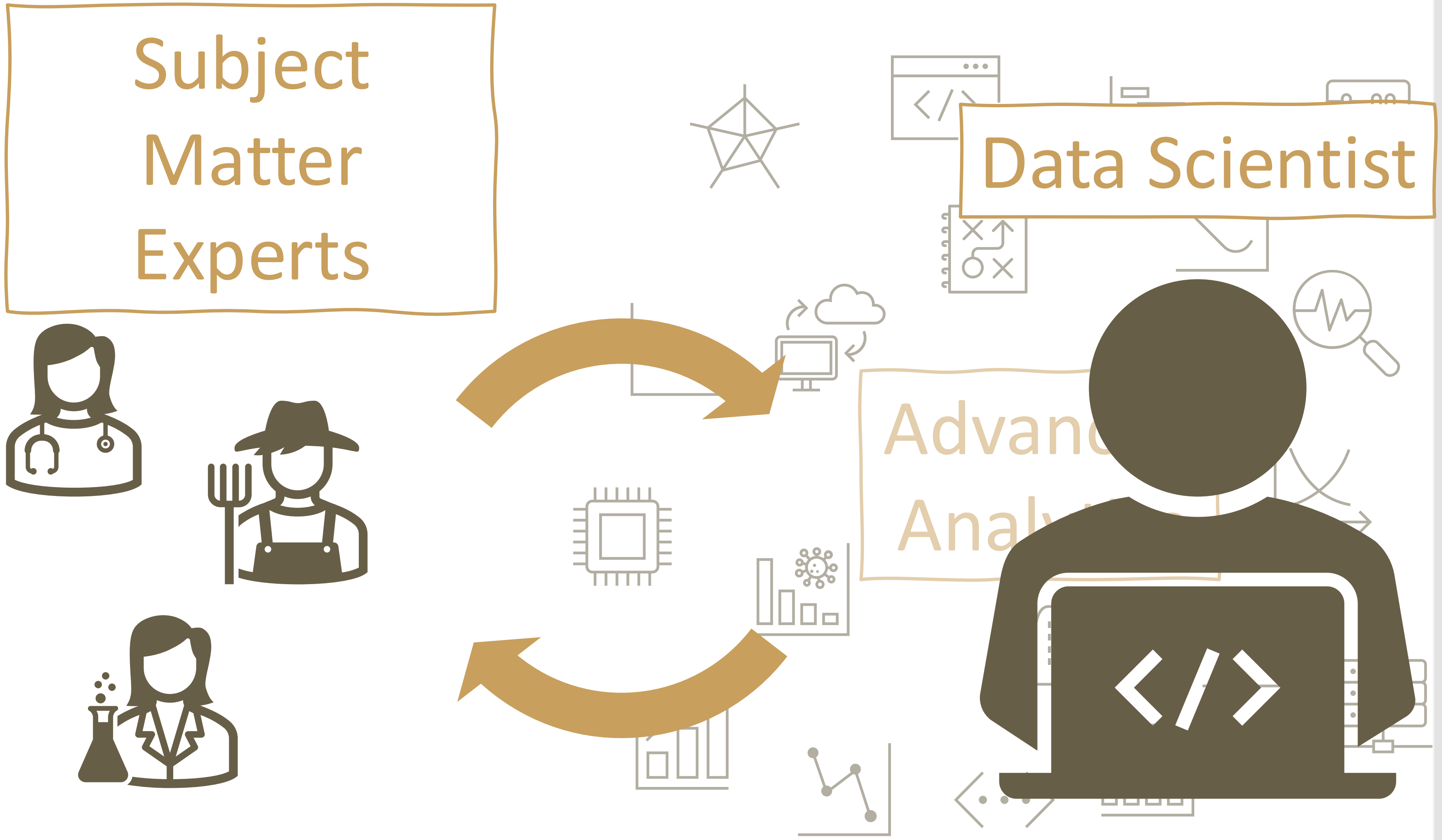
How does this help?

Idea 1:
do things more efficiently

How does this help?

Defining and Analysis: A Study of Client-Facing Data Scientists. Mosca, Robinson, Clarke, Redelmeier, Coates, Cashman, Chang.

How does this help?



Defining and Analysis: A Study of Client-Facing Data Scientists. Mosca, Robinson, Clarke, Redelmeier, Coates, Cashman, Chang.

[v]alt

How does this help?

Defining and Analysis: A Study of Client-Facing Data Scientists. Mosca, Robinson, Clarke, Redelmeier, Coates, Cashman, Chang.

How does this help?

Working backwards

Probing

Recommending

NLP **XAI**

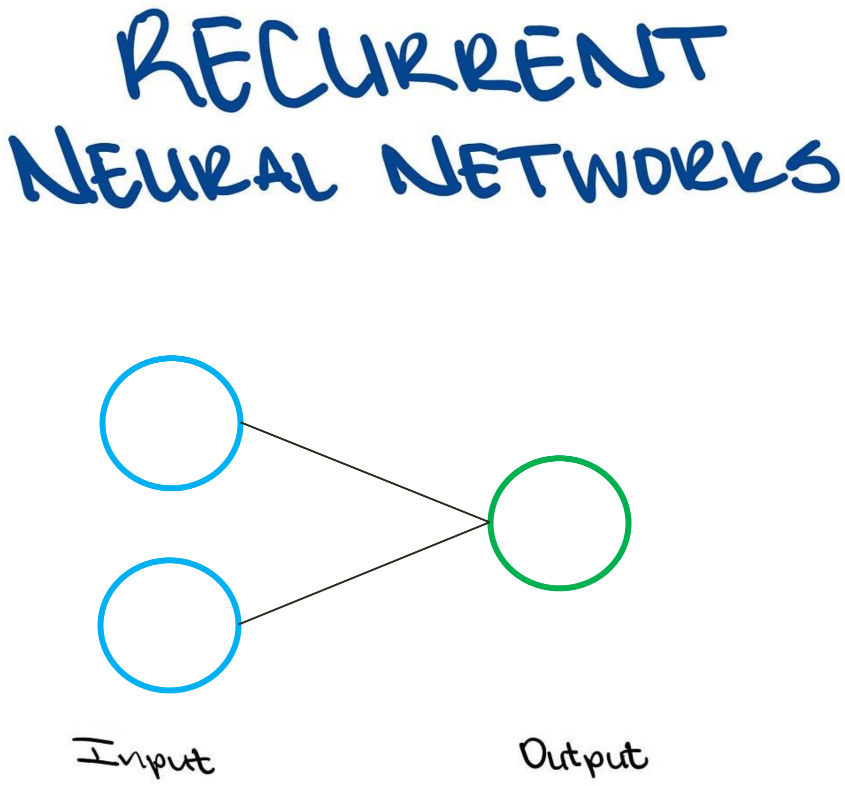
AutoML

Defining and Analysis: A Study of Client-Facing Data Scientists. Mosca, Robinson, Clarke, Redelmeier, Coates, Cashman, Chang.

How does this help?

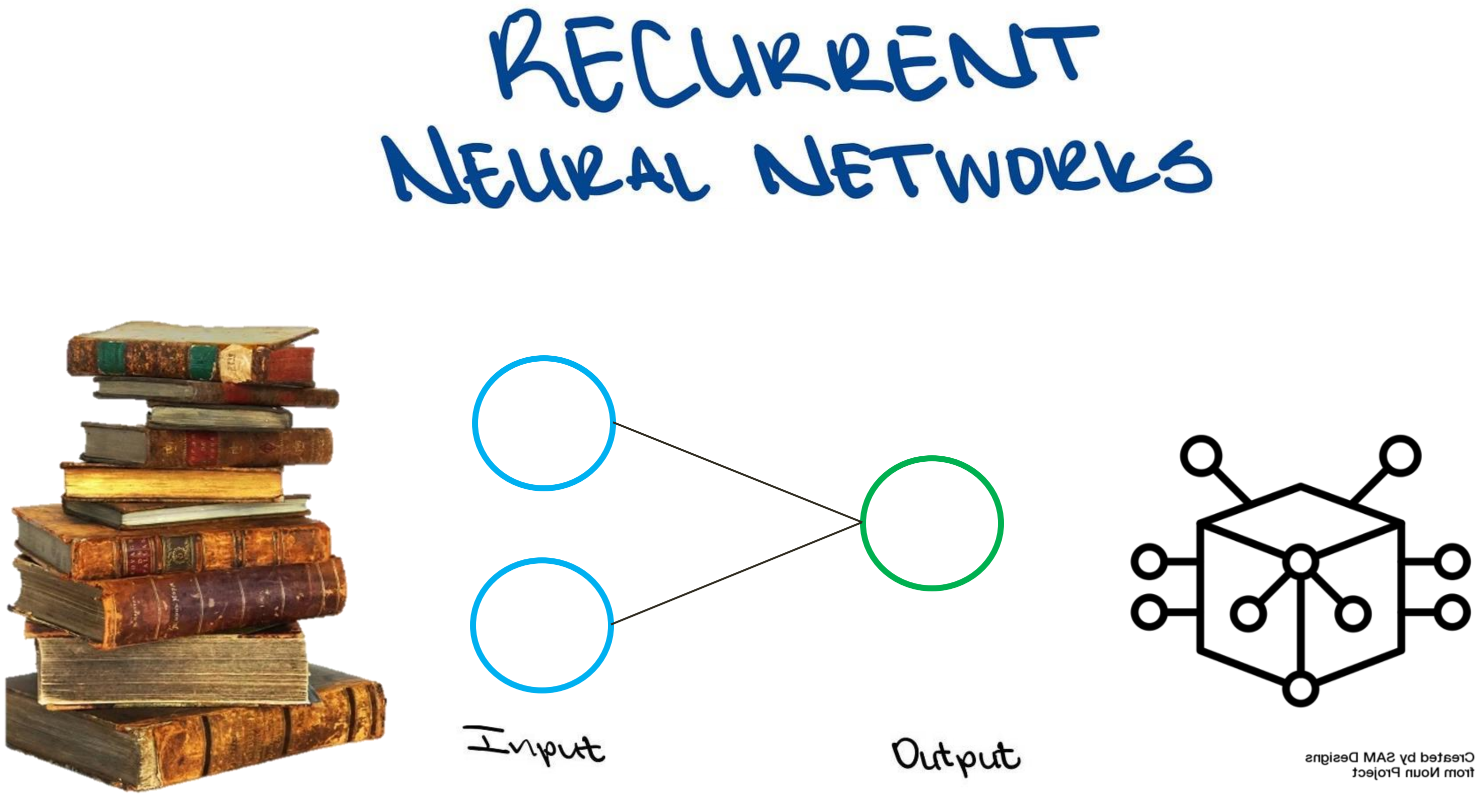
Idea 2:
See the bigger picture

How does this help?



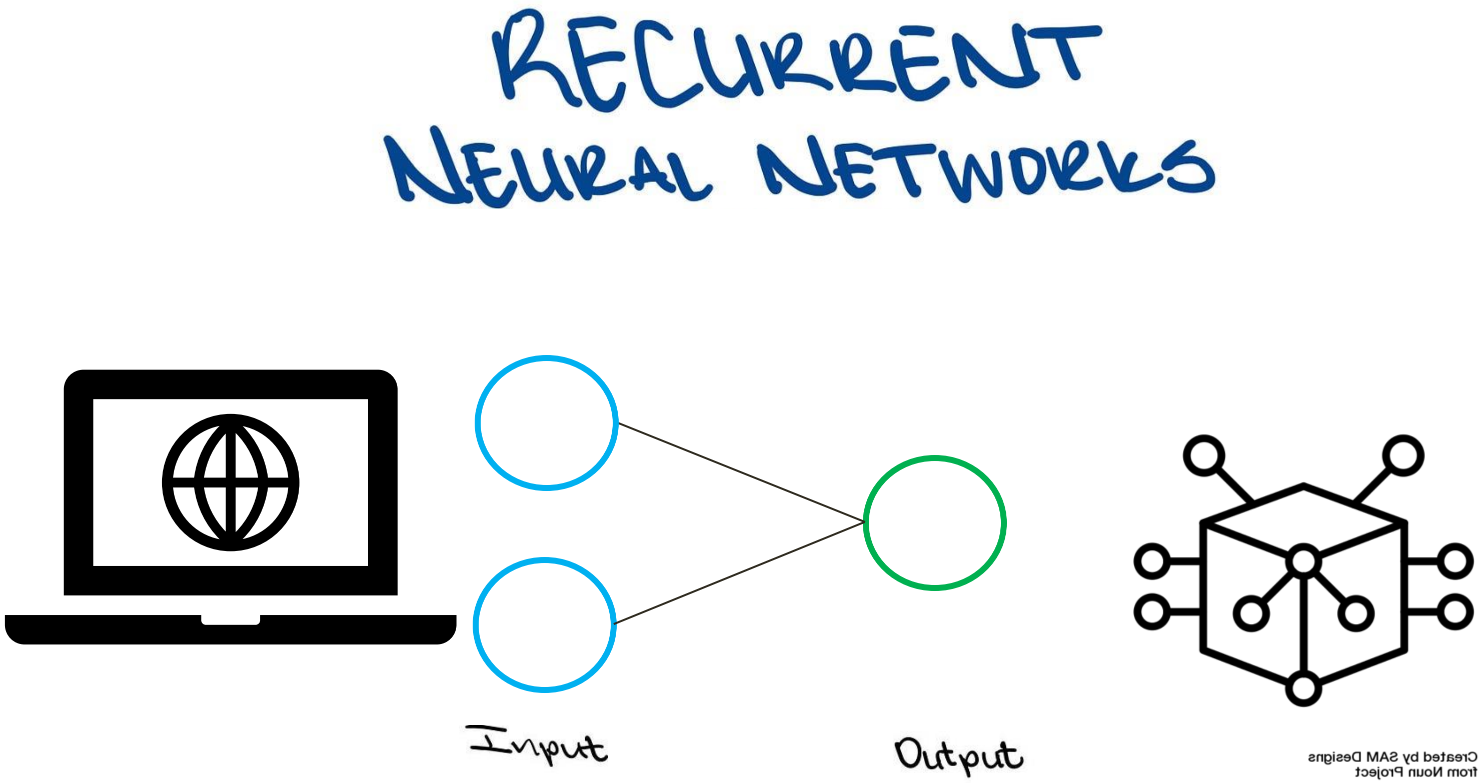
https://towardsdatascience.com/introducing-recurrent-neural-networks-f359653d7020

How does this help?



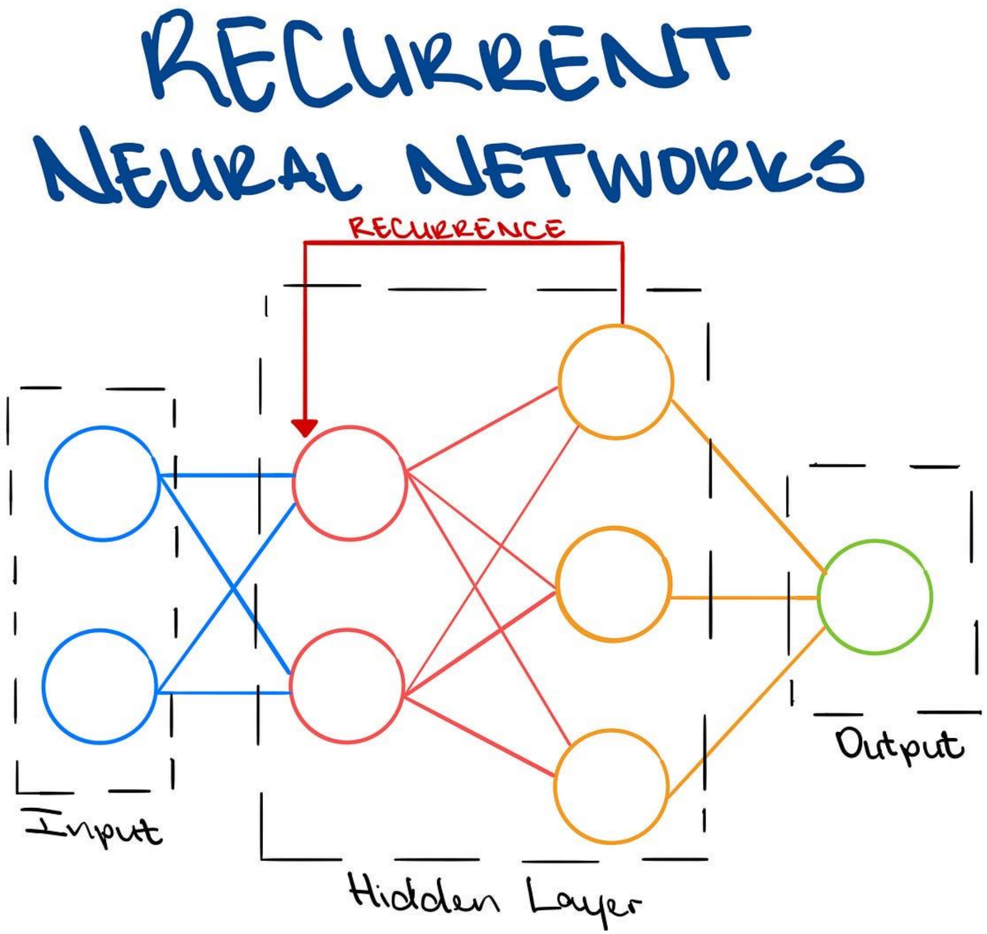
https://towardsdatascience.com/introducing-recurrent-neural-networks-f359653d7020

How does this help?



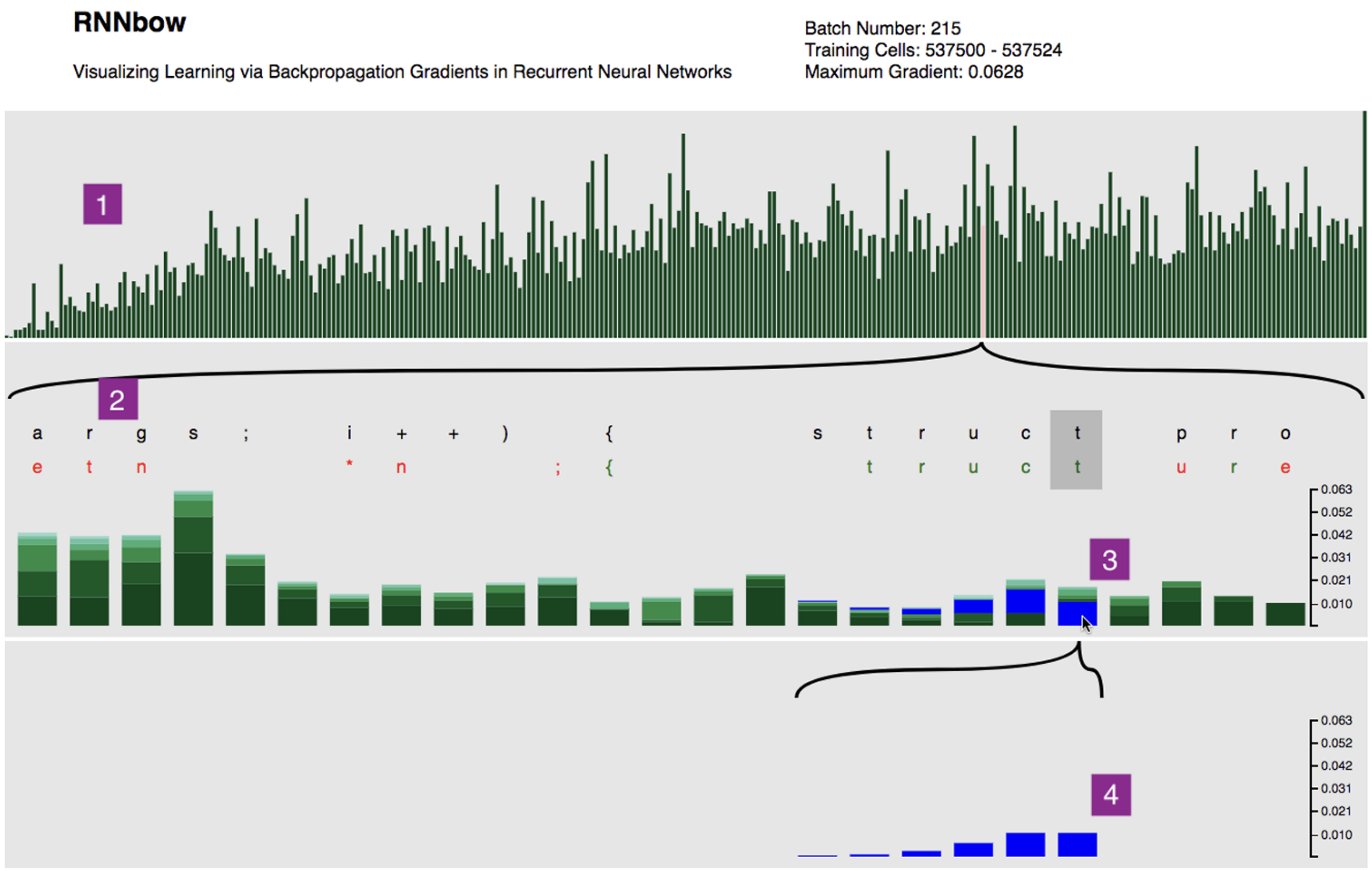
https://towardsdatascience.com/introducing-recurrent-neural-networks-f359653d7020

How does this help?



https://towardsdatascience.com/introducing-recurrent-neural-networks-f359653d7020

How does this help?

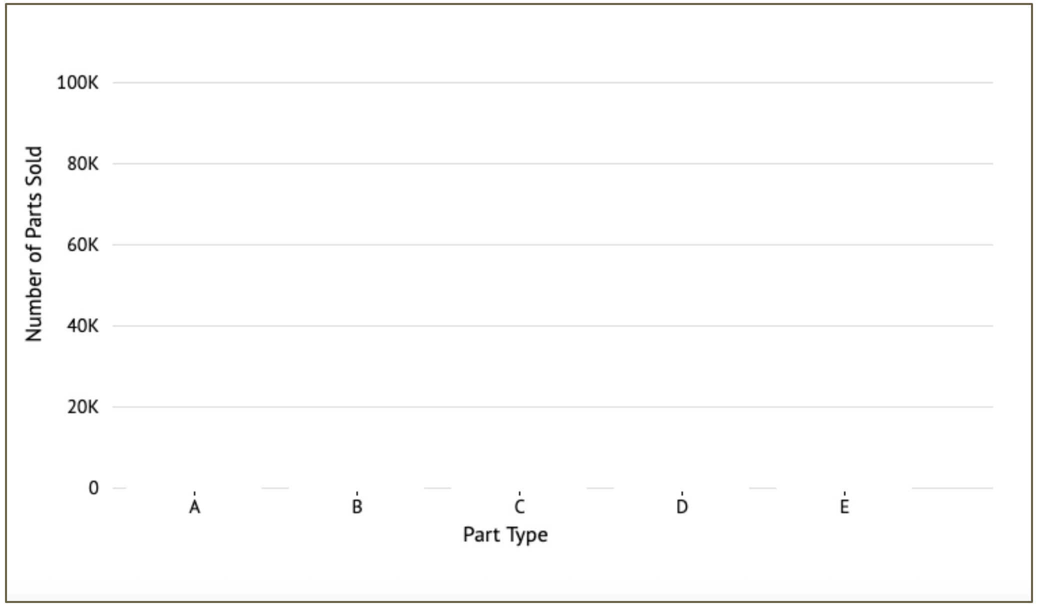


RNNbow: Visualizing Learning via Backpropagation Gradients in Recurrent Neural Networks. Cashman, Patterson, Mosca, Watts, Robinson, Chang.

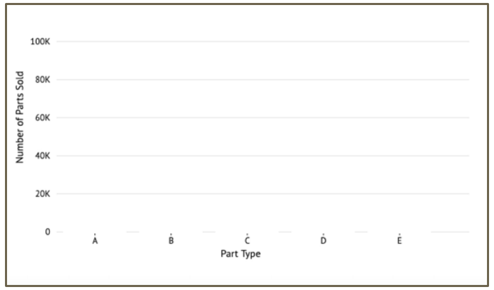
How does this help?

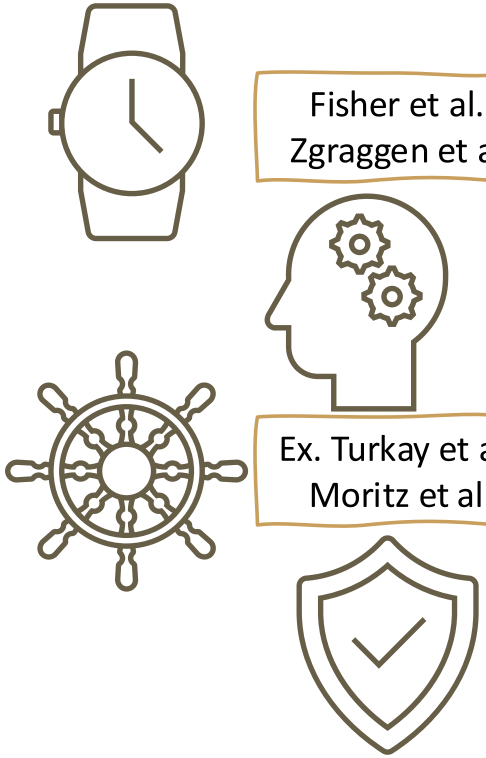
Idea 3:
Avoid getting stuck

How does this help?



How does this help?





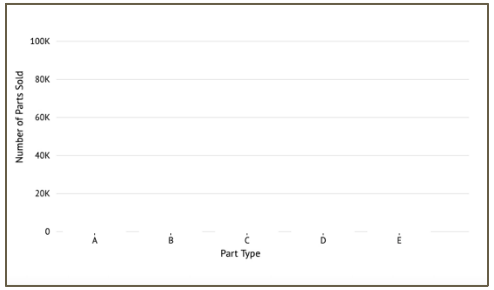
Fisher et al. 2012, Zgraggen et al. 2017

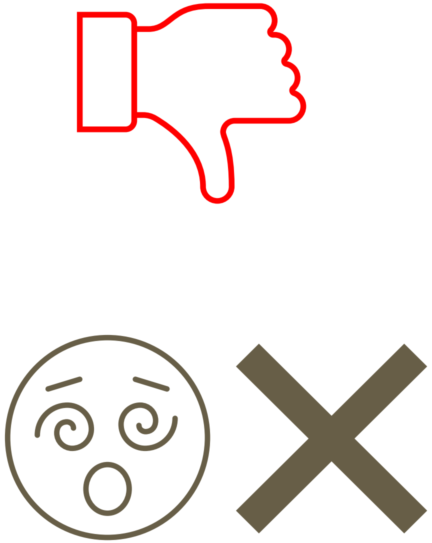
Angelini et al. 2018, Schulz et al. 2016

Ex. Turkay et al. 2017, Moritz et al. 2017

Badam 2017

How does this help?





Moritz et al. 2017,
Turkay et al. 2017

Micallef et al. 2019

How does this help?

Pitfalls

Cognitive Biases

Uncertainty

Illusion

Control

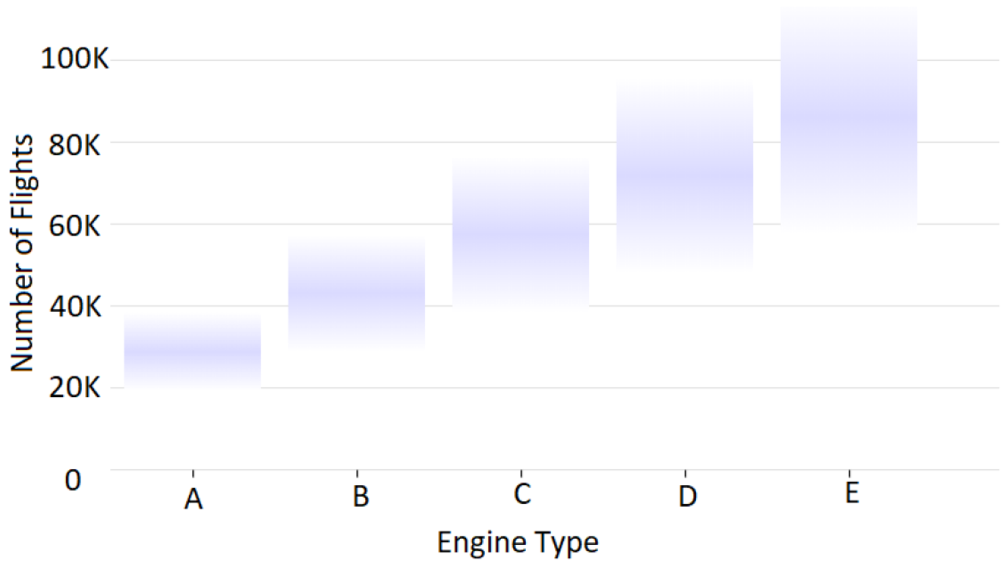
Anchoring

Impact Of Cognitive Biases On Progressive Visualization.
Procopio, Mosca, Scheidegger, Wu, Chang.

Columbia University

How does this help?

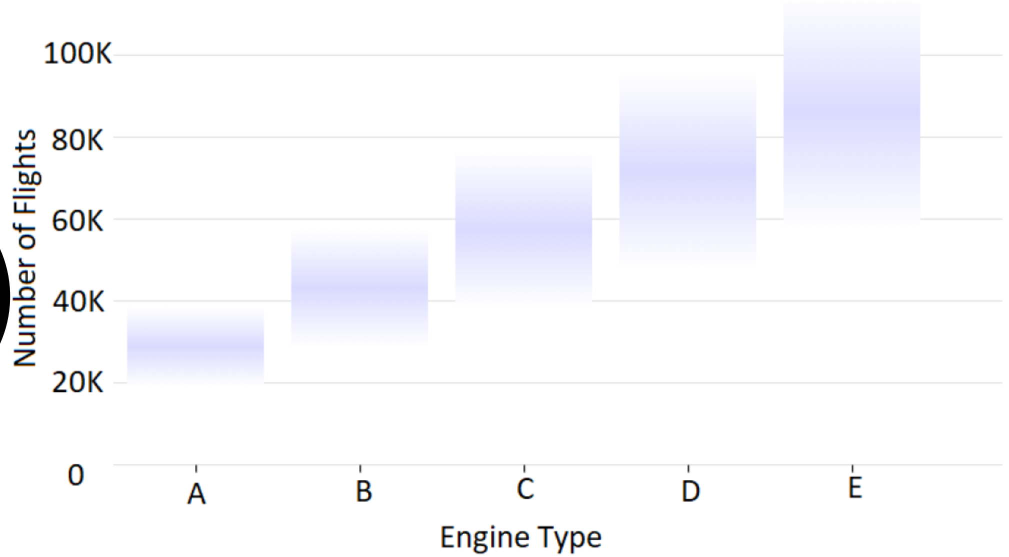
Reading something in incomplete results that is not there



Columbia University

How does this help?

Reading something in incomplete results that is not there



Impact Of Cognitive Biases On Progressive Visualization.
Procopio, Mosca, Scheidegger, Wu, Chang.

Columbia University

Structure of this course

- Disclaimer:

  this class is an experiment in **constructionism**

  (the idea that people learn most effectively
  when they're building personally-meaningful things)

- Our job as instructors:

What we'll cover

- Unit 1: Foundations
  - Next Class: **Mental and Visualization Models**
  - Data Wrangling and Modeling 101
  - Data Visualization 101
  - Interaction
  - Coordinating Multiple Views
- Unit 2: The Real Fun Starts
  - Quirky data (time-varying data, geospatial data, networks)
  - Scalability
  - Color and Visual Channels
  - High-dimensional data
- Unit 3: Ongoing Research
  - Provenance
  - Uncertainty
  - Prediction and Recommendation

What we WON'T cover

- CSC foundations:
  - Editors (vim, emacs, etc.)
  - Version control (git)
  - Debugging and profiling
  - Python basics

- SDS foundations:
  - Data collection
  - Databases and storage
  - Regression, etc.
  - R basics

Important note: VA systems are rarely built by a single person! You don't need to have mastered all these skill, just try to make some friends who have complementary experience !

General information

- Course website:
  - https://amosca01.github.io/SDS-CS235/

What you'll get

By the end of this course, you will:

- Understand what Visual Analytics is
- Know the foundational methods and tools available
- Be familiar with some ongoing research in VA
- Be able to perform analyses and report your findings
- Have (marketable!) experience developing useful visual analytics applications for real data

What I expect from you

- You enjoy grappling with difficult problems, and you’re excited about “figuring stuff out”
- You are (or are willing to work to become) proficient in programming and debugging
  - We’ll do crash courses in a few key topics, but this is NOT a course intended to teach you general programming techniques
  - You’re welcome to work in whatever language(s) you prefer, but I’m more helpful in the ones I know
- You’re comfortable asking questions

What you can expect from me

- I’m flexible w.r.t. the topics we cover:
  - This course is a collaboration
  - If there’s something you want to learn that’s not on the agenda, speak up!
  - If I’m doing something that doesn’t work for you (Font too small on presentations? Speaking too quickly? Using a marker or color you can’t see?), please let me know!

I’m here to help you succeed, and I believe you all have the ability to succeed

Course project

- Goals:
  - Learn how to break big, unwieldy questions down into clear, manageable problems
  - Figure out if/how the techniques we cover in class apply to
  - your specific problems
  - Build VA systems to address them
- Several (graded) milestones along the way
- Demos and discussion on the final day of class

gain real experience | solve real problems
build real relationships

Questions?