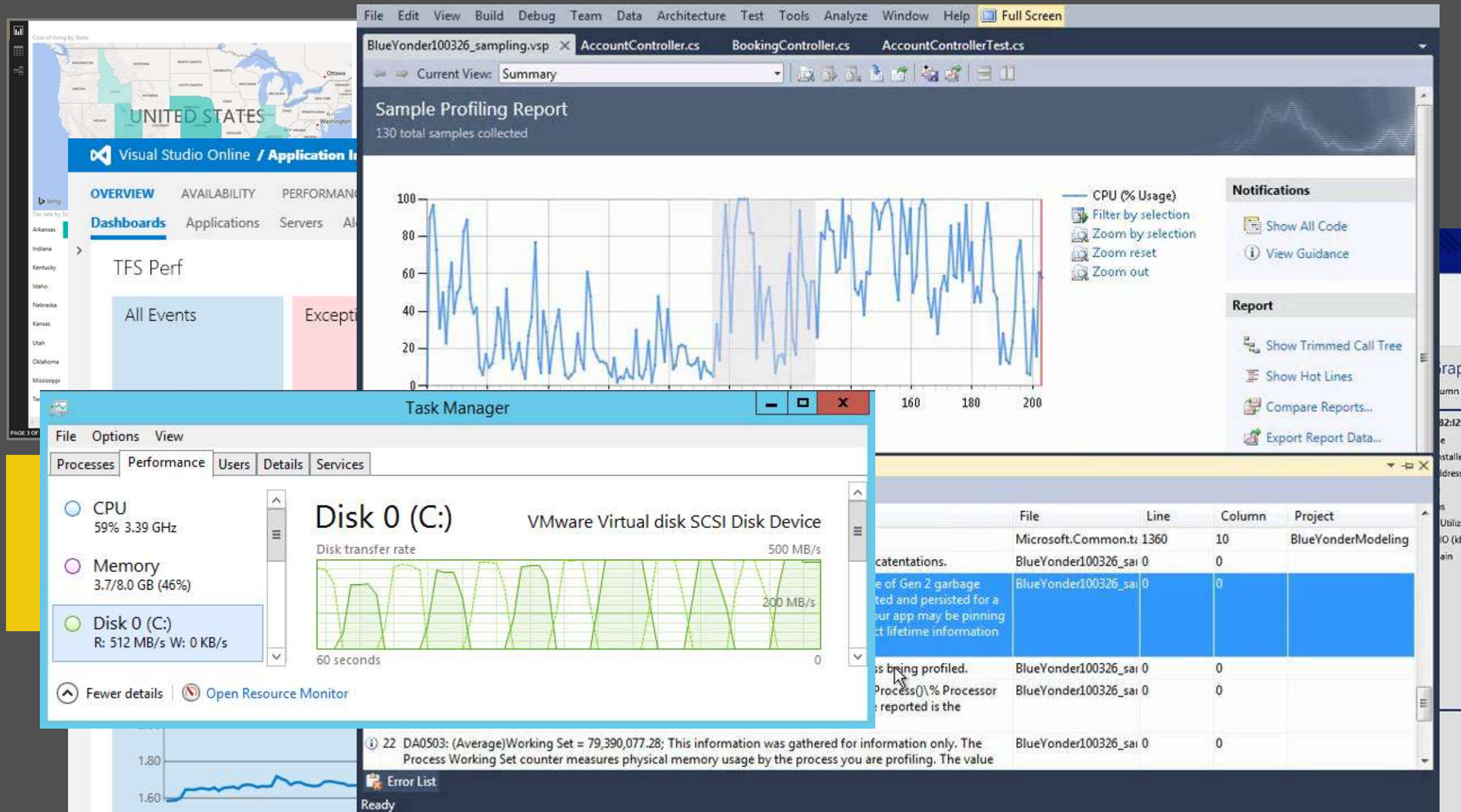
Making Sense of Data

Danyel Fisher
Senior Researcher at Microsoft Research

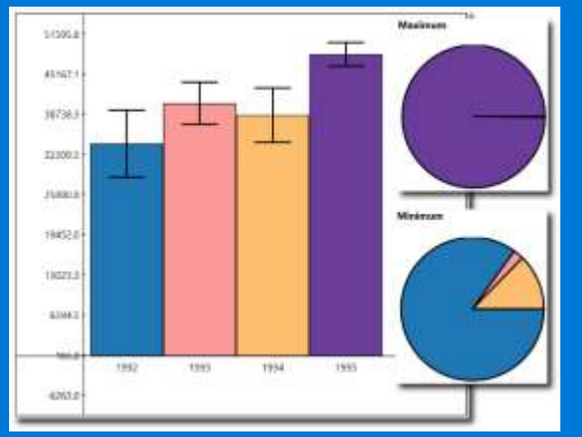
@danyelf

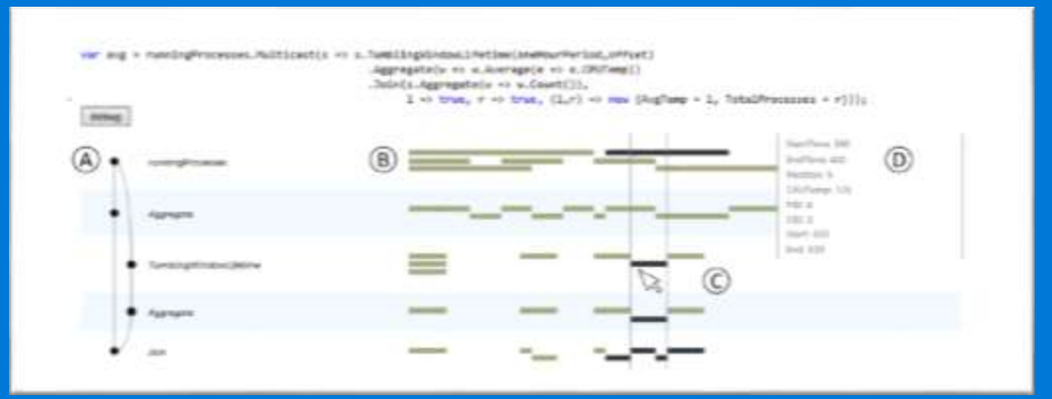
Microsoft


File Edit View Build Debug Team Data Architecture Test Tools Analyze Window Help Full Screen
BlueYonder100326_sampling.vsp
AccountController.cs
BookingController.cs
AccountControllerTest.cs
Current View: Summary
Sample Profiling Report
130 total samples collected
CPU (% Usage)
Filter by selection
Zoom by selection
Zoom reset
Zoom out
Notifications
Show All Code
View Guidance
Report
Show Trimmed Call Tree
Show Hot Lines
Compare Reports...
Export Report Data...
UNITED STATES
Visual Studio Online / Application I
OVERVIEW
AVAILABILITY
Dashboards
Applications
Servers
TFS Perf
All Events
Task Manager
File Options View
Processes
Performance
Users
Details
Services
CPU
59% 3.39 GHz
Memory
3.7/8.0 GB (46%)
Disk 0 (C:)
R: 512 MB/s W: 0 KB/s
Disk 0 (C:)
VMware Virtual disk SCSI Disk Device
Disk transfer rate
500 MB/s
200 MB/s
60 seconds
0
Fewer details
Open Resource Monitor
File
Line
Column
Project
Microsoft.Common.ta
1360
10
BlueYonderModeling
BlueYonder100326_sa
0
0
Error List
Ready

Who I Am

- I do research on visualization.
- I'm also often a consultant on visualizations for internal and external projects.





Other Talks I'm Not Giving Today

"Why Exploring Big Data is Hard"

"What Designers and Visualizers Can Learn From Each Other"

"Querying Temporal Event Sequences"

Where I Sit

- I use data visualization (largely) functionally
- Most stuff on the screen is meant to inform users
  - If the *audience* can't interpret it—it's not a visualization.
- Much of my work is about business intelligence

People have questions

I have data.
Can you visualize it for me?

That depends. What do you want to see?

I want to see my annual sales!

Maybe it’ll look like the Matrix!

Or the Minority Report!

Or something spinning in 3D!

And everything will become suddenly clear!

What's interesting about your annual sales?

Well, some of our products are selling very differently year after year

What varies about these products?

We change both the color and the sales pitch every year, and for each store

... and so on

Who is this for?

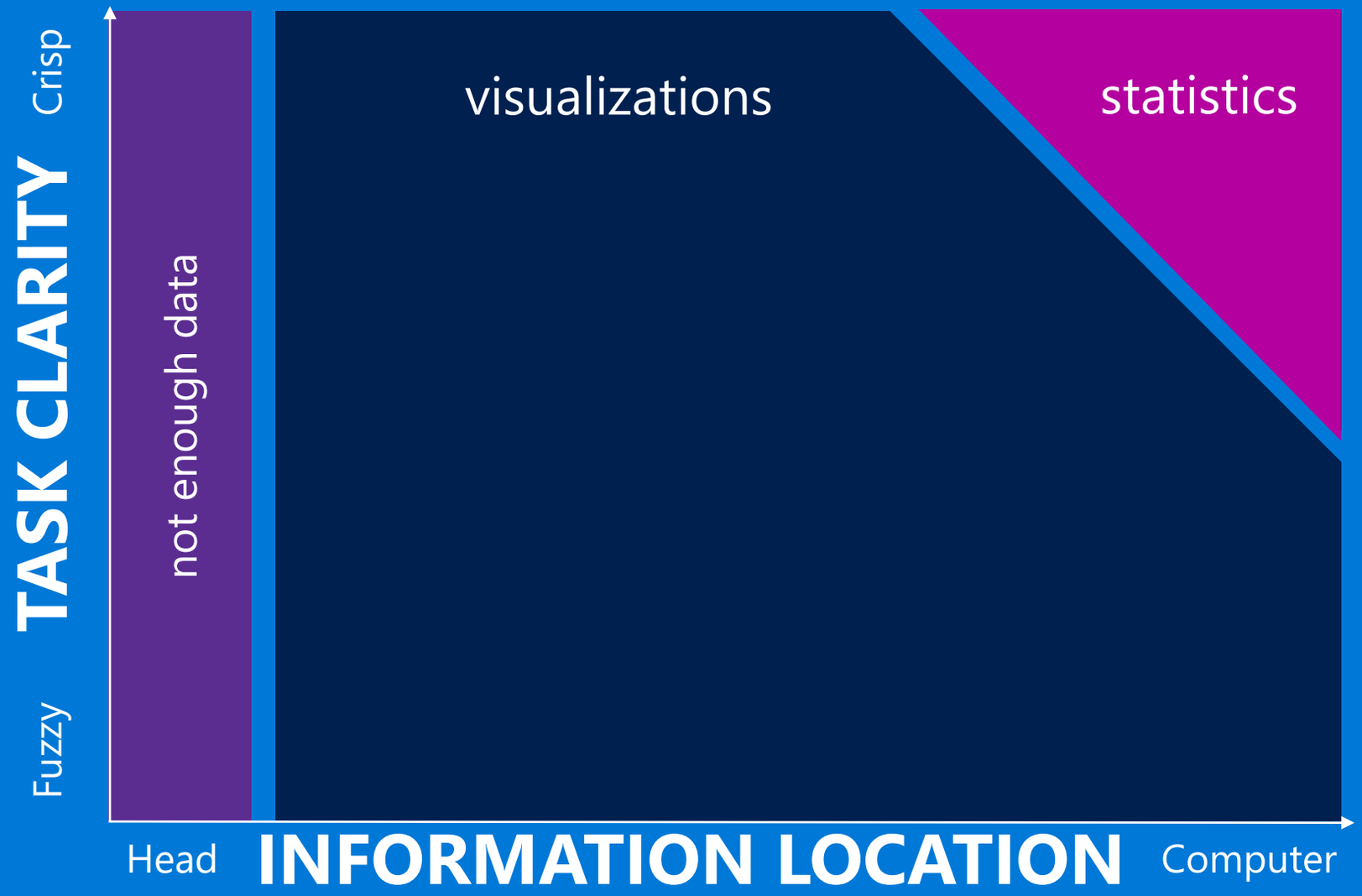
These People

- What visualization would make most sense here?
- How do I take on this dataset?
- I need to learn something from my dataset
- How do I describe what I do today?
- I need to communicate something about this data
- I want to monitor my data

Not These

- How do I get PowerBI to build my dashboard?
- How do I use this visualization in Excel?

Should I use visualization?



Outline of this talk

- Data Counseling
- The Nonprofit Thermometer
- Operation Tree
- A Task-Oriented View of Visualization
- Multiple Views
- Conclusion

Data Psychotherapy

What is “Data Counseling”?

Translating general **questions** into **concrete analysis** tasks that can be performed over the data

- interviews
- exploratory data analysis
- prototyping

Carrying Out Interviews

Identify Stakeholders

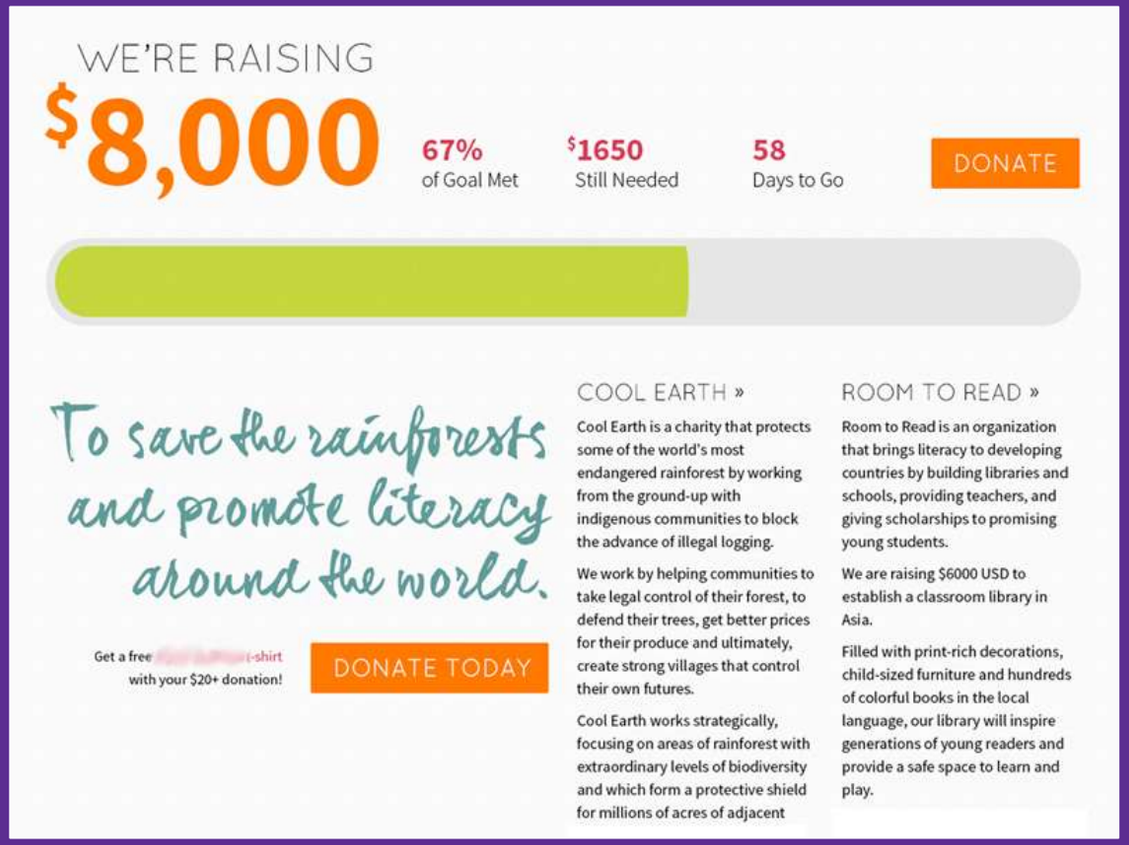
- Who *owns the data*
  - **Where it comes from**
  - **How it got generated**
  - **What its quirks are**
- Who *uses* it?
- Who makes decisions with it?

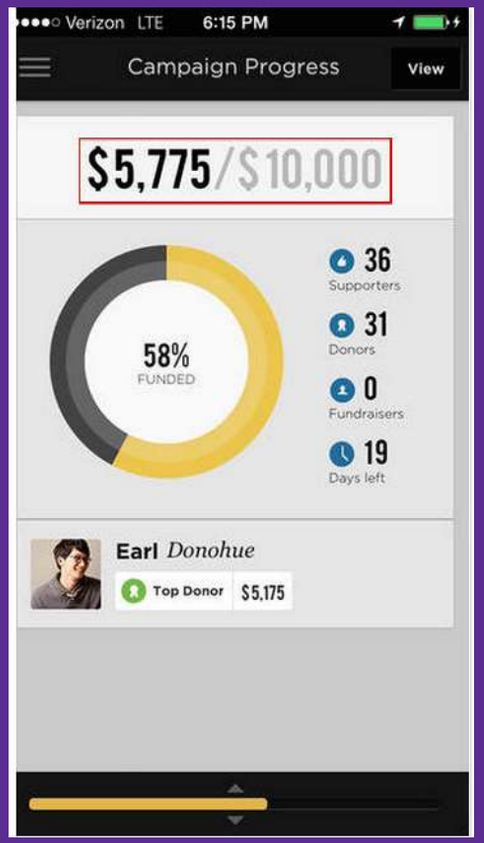
Forming Goals

- What would you like to know that you don't today?
- What story would you be able to tell?

Case Study: The Nonprofit Thermometer

Case Study: The Nonprofit Thermometer

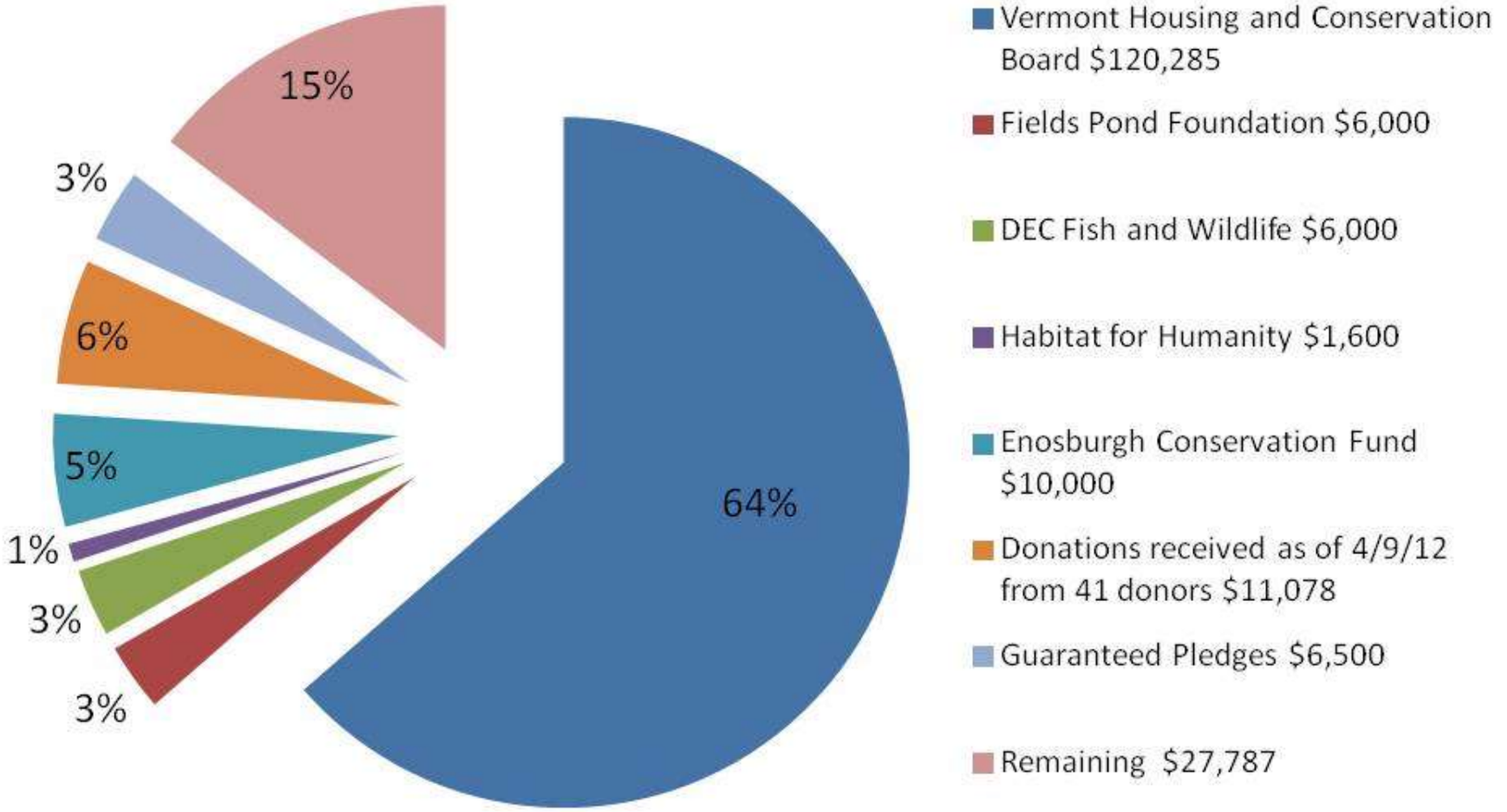




http://showcase.ixdetic.com/

Enosburg Falls River Access Fundraising

Help us raise $189,250 to conserve 9.28 acres and over 1,500' of shoreland on the Missisquoi River



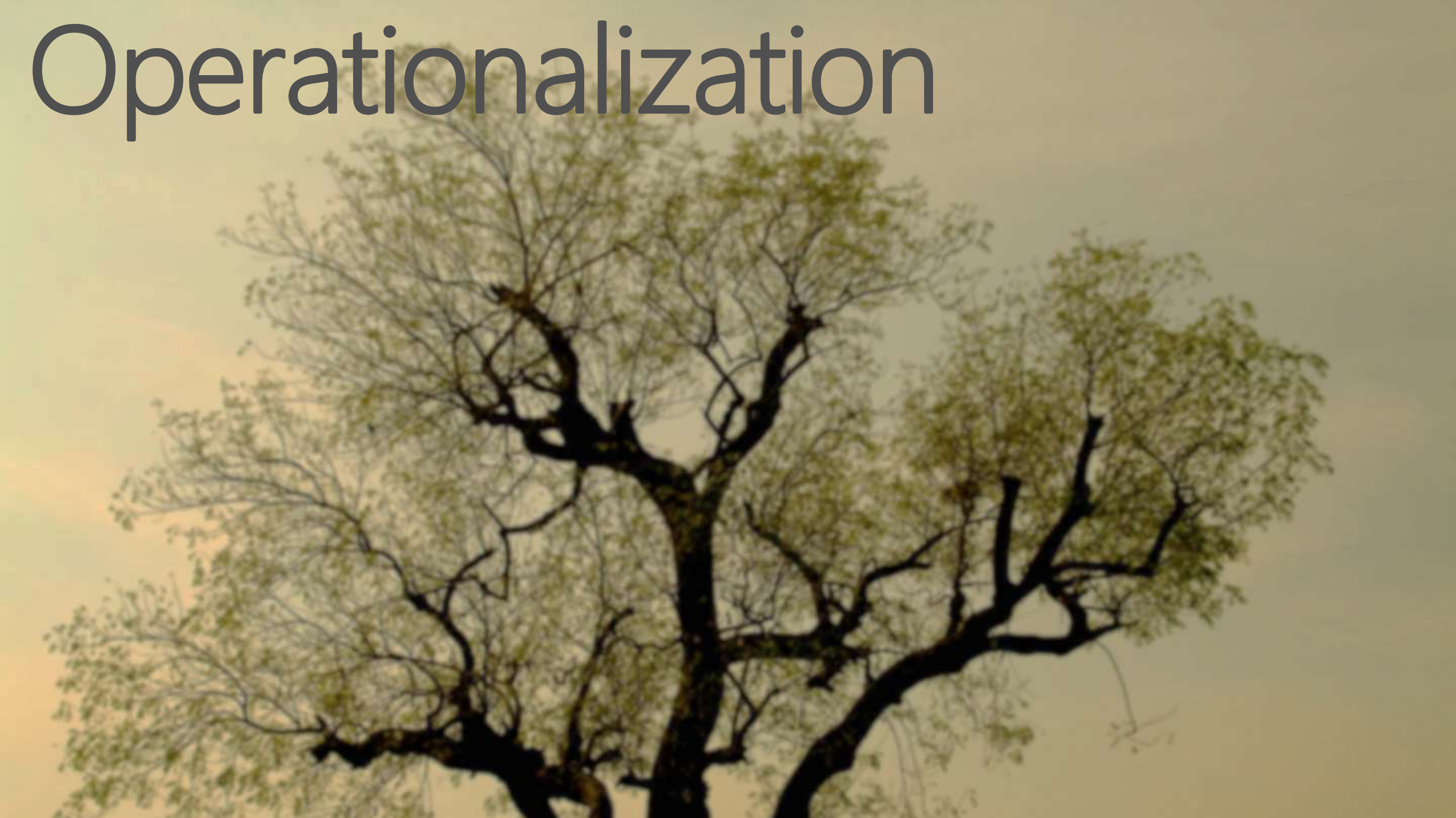
What's wrong with the current tool?

- *I just convinced a billionaire to give half-a-million dollars. That gets us 30% through the campaign, but I don't have two more billionaires sitting next to him.*

- *A segment of our donors are sitting this one out!*

How do I understand and communicate how our capital campaign is doing?

Operationalization



Operationalization

Systematically reducing a high-level, ambiguous, data-free question into specific tasks over the data, which can be addressed with definite answers and visualizations

Operationalization Tree

construct for guiding the operationalization

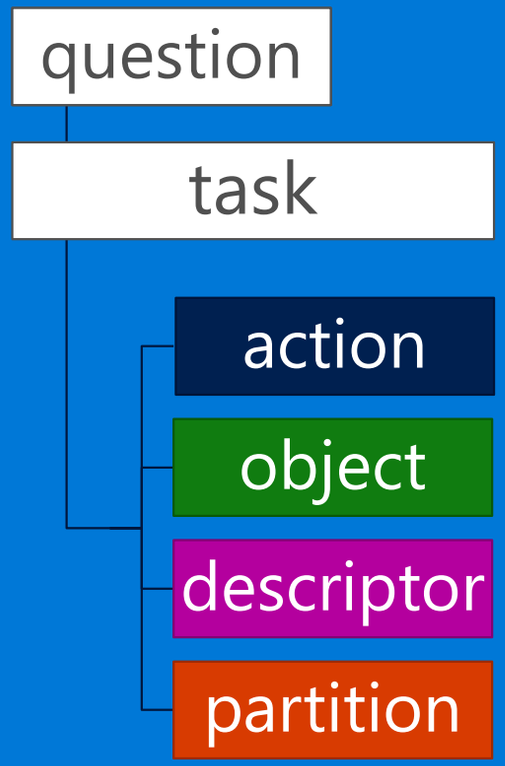
helps in identifying what is ambiguous

provides structure for designing a visualization solution

data constrains the problem

data adds a layer of specificity that is actionable

The Operationalization Tree



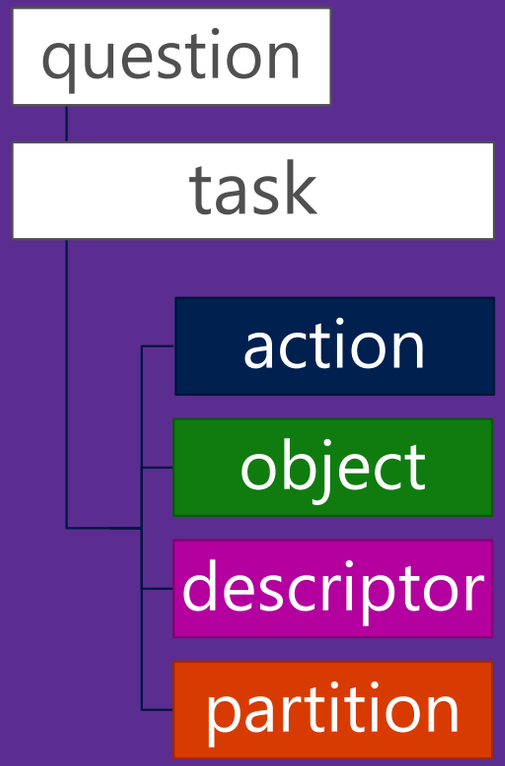
the thing you want to do

the items you want to take action on

the value you will measure for the objects

the way the data divides

Nonprofit



how is our capital campaign doing?

understand & communicate the current state of the capital campaign

describe

the capital campaign

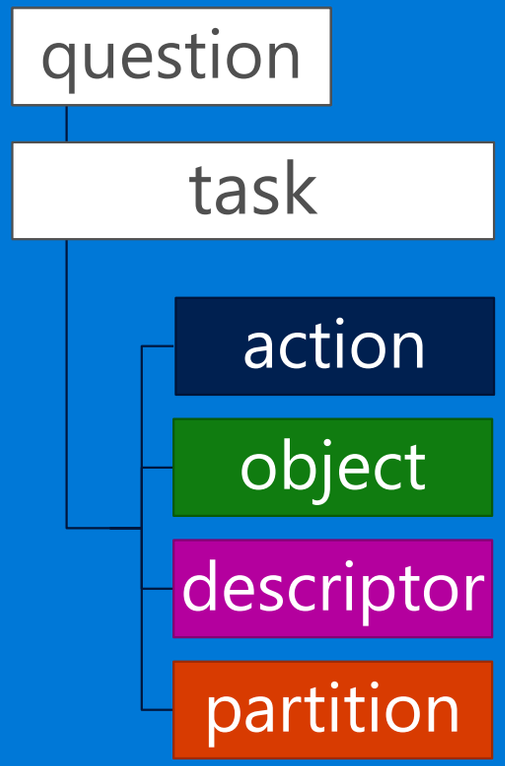
how well its doing

-

Tree Descent

- Look for ambiguous objects, then partitions and descriptors
  - Do we understand what this is?
  - Do we know how it corresponds to the data?
- Refine for a solution:
  - A datapoint
  - A chart
  - An equation

The Operationalization Tree



the thing you want to do
will help choose a specific visualization or interaction

the items you want to take action on
will correspond to rows of the data

the value you will measure for the objects
will correspond to measures

the way the data divides
will correspond to dimensions and grouping keys

Refine the Tree

Interviews

- What would it mean to have a “lot” of this?
- How can you tell groups apart?
- How does this vary?
- How does this look in the data?

Refinement

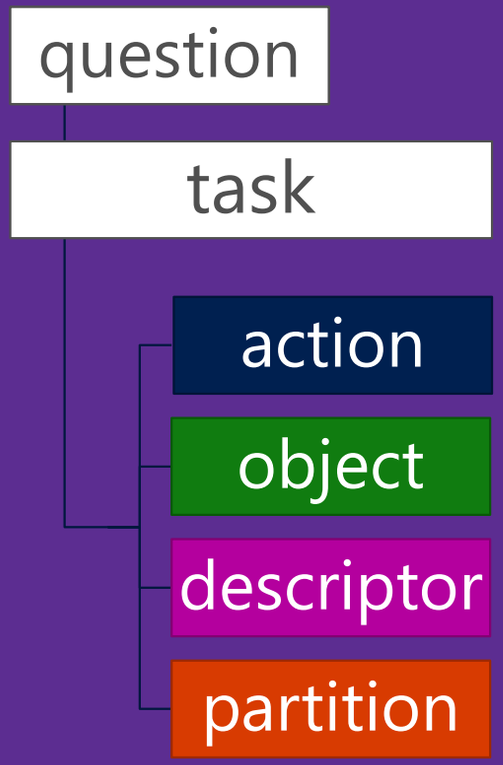
- Descriptor:
  "how well it is doing"

- *Interview: What would it mean for the campaign to be doing well? What would you need to understand?*
- *A: For each of three types of donors, I need to know how many we've reached out to, how many have responded, and what they're giving.*

Three Classes of Donors

- We have different groups of donors, each of which needs to be treated differently.
  - Type I donors give lots of money.
  - Type II donors give less money, or lots of money irregularly.
  - Type III donors give some money, irregularly.

Nonprofit



"How do I understand & communicate how many donors we've reached out to, and their giving rates, for our capital campaign – for each class of donors?

understand and communicate

donor, in the scope of this campaign

outreach, giving rate

donor group

How could you tell whether someone is a group 2 or a group 3 donor?

Donor Group

- LYBUNT: Last Year, But Unfortunately Not This Year
- SYBUNT: Some Year, But Unfortunately Not This Year

When did they last give?
How much did they give?
What is their capacity to give?

A Tangent, on Tangents

Domain experts *love* to tell you about additional complexity

- Will that additional complexity help inform this picture?
- What would we learn from this extra fact?

… how donors move between categories
… how long-term donor cultivation works
… the difference between annual fundraising and capital campaigns
…

Experts communicating with experts

What decisions would knowing this fact change?

What does the viewer need to know?

Donor Group

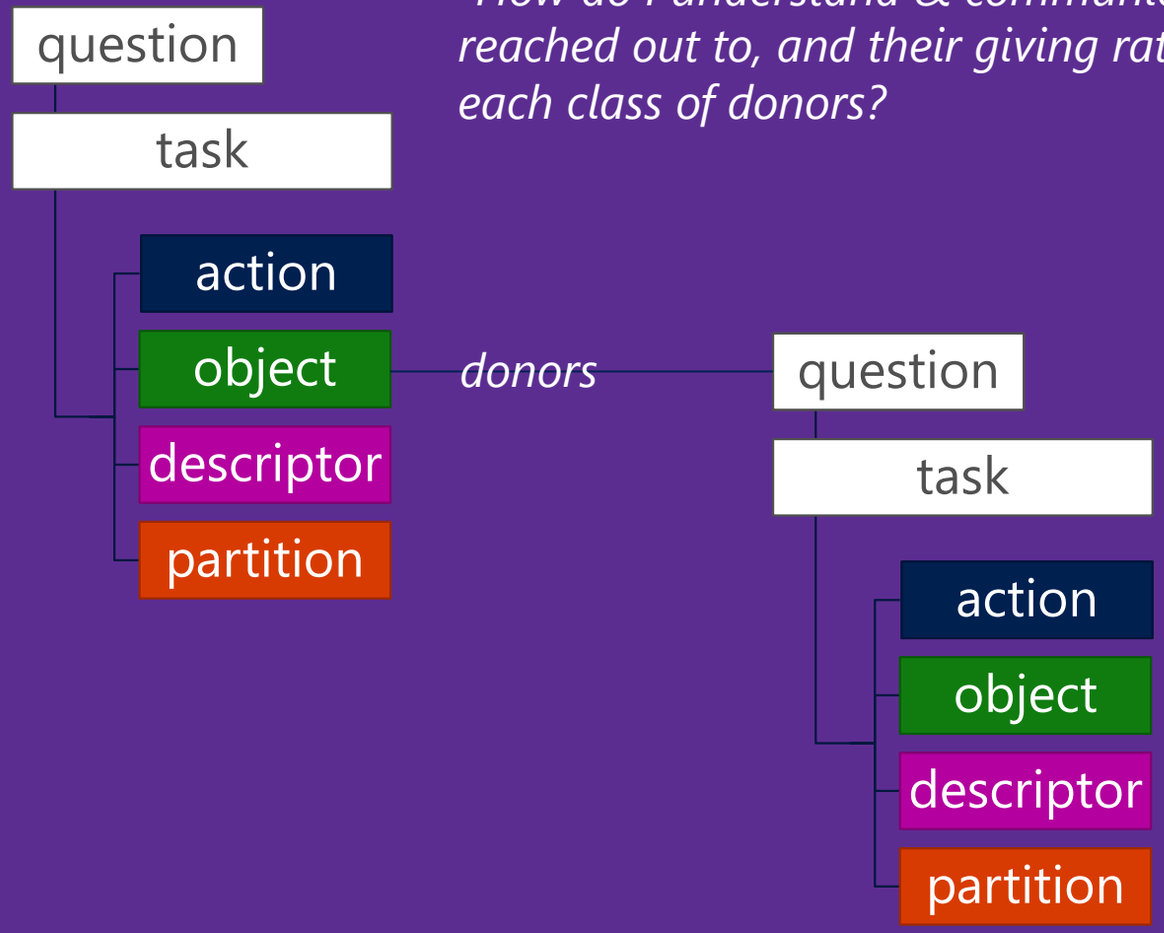
- When did they last give?
- How much did they give?
- What is their ca

- LYB tunately Not This Year
- SYBU , But Unfortunately Not This Year

Donor Group is a column In the Database

Reaching the leaves

"How do I understand & communicate how many donors we've reached out to, and their giving rates, for our capital campaign – for each class of donors?



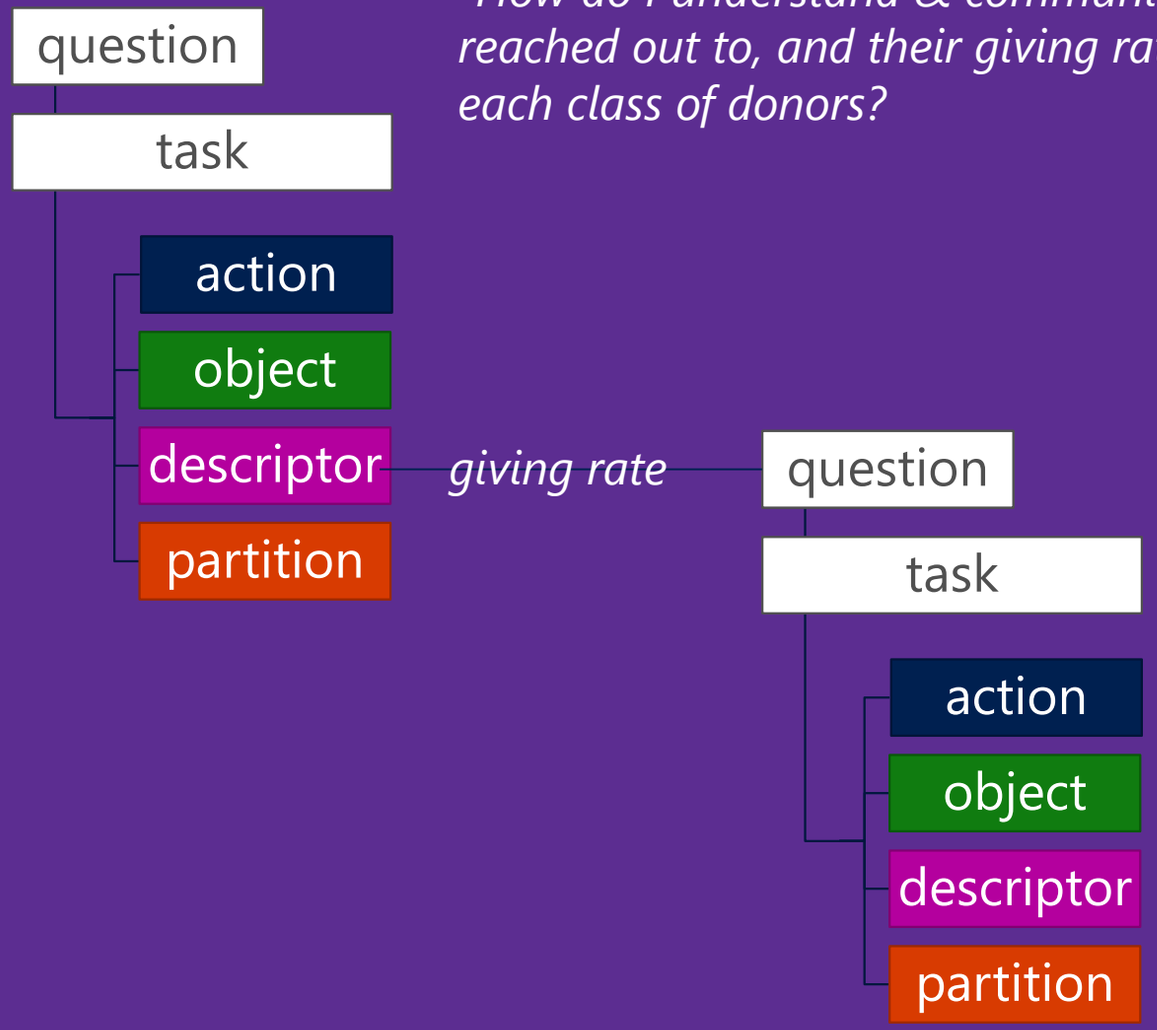
what is the group & giving history for a donor?

Identify group & giving history for a donor

Ok, looks pretty good!

Reaching the leaves

"How do I understand & communicate how many donors we've reached out to, and their giving rates, for our capital campaign – for each class of donors?

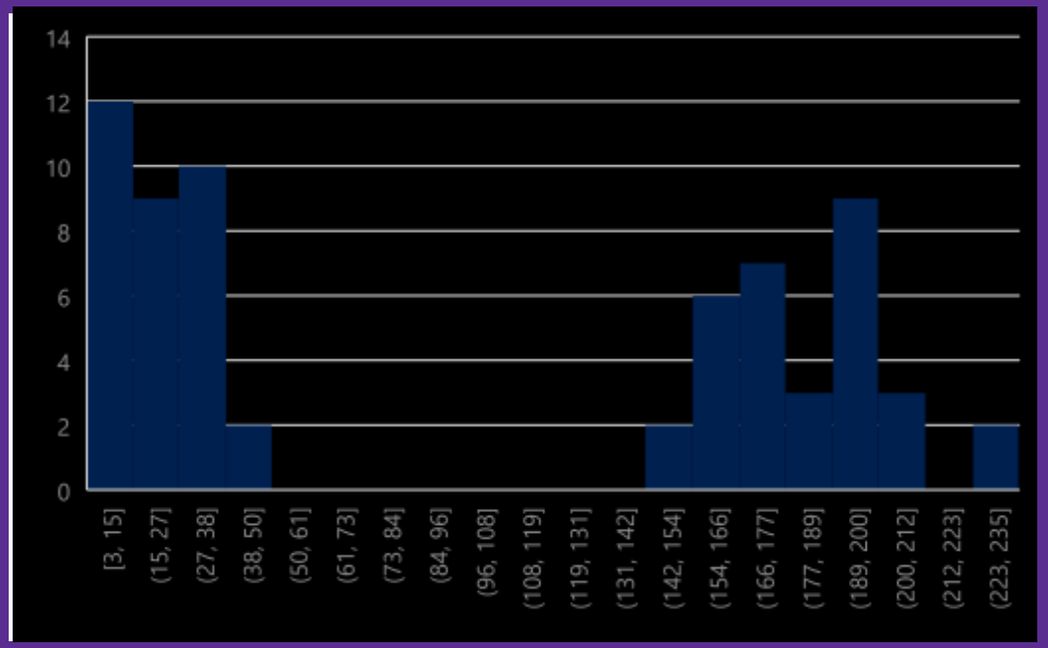


What is the giving rate for this donor group?

Describe the giving rate for a donor group?

Using EDA to understand leaves

- All the people in Tier II who gave during the phone campaign, by giving amount



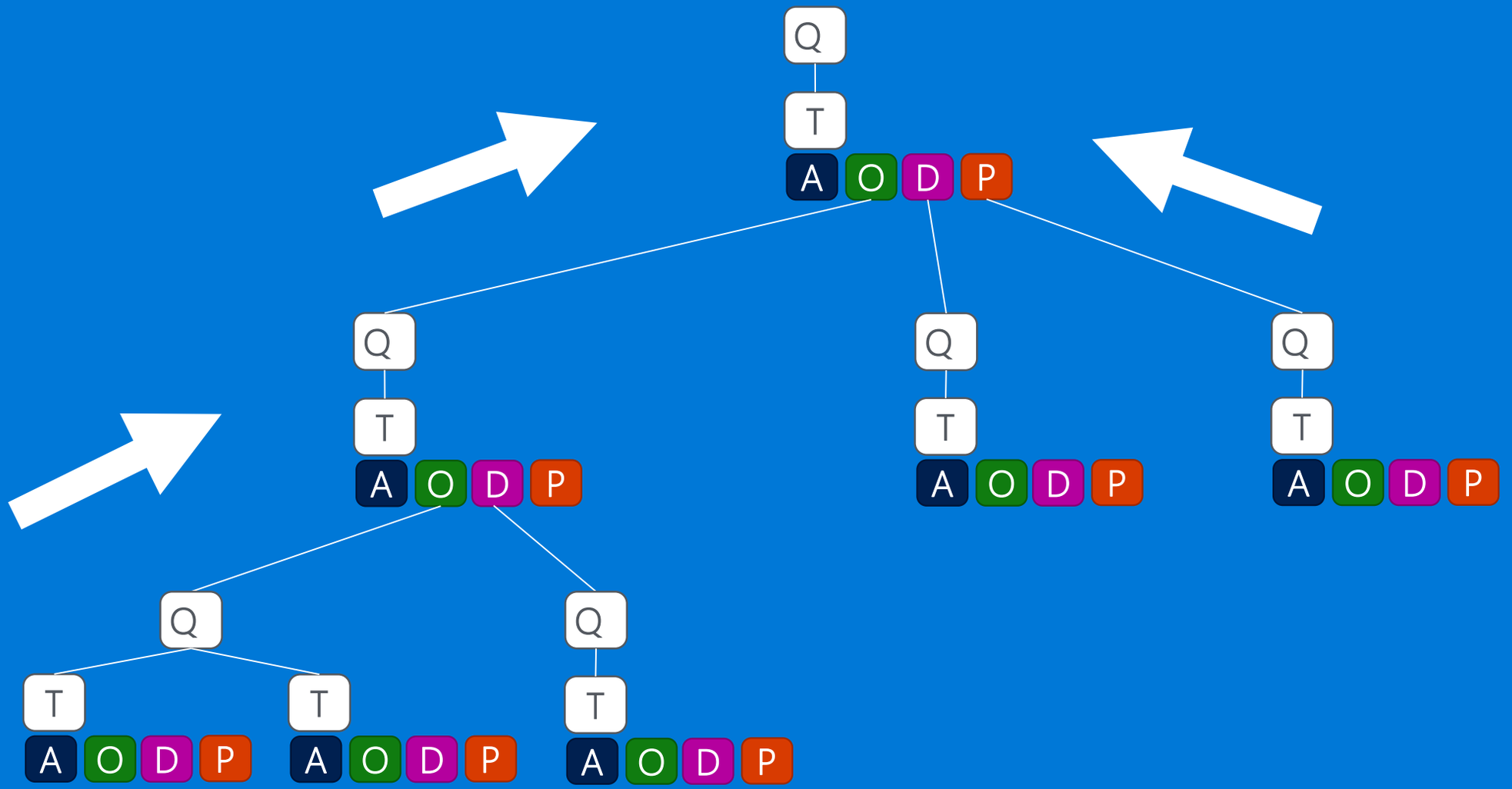
How many donors we've reached out to

Different methods of outreach depending on
- donor group (coffee vs galas vs newsletters)
- individual history

Escalating cost, but increasing returns

No need to reach out to someone who already gave

Database has CRM information with all contacts


Q
T
A O D P
Q
T
A O D P
Q
T
A O D P
Q
T
A O D P
Q
T
A O D P
Q
T
A O D P
Q
T
A O D P

Climbing Back Up

If the result is …

… a value, we can bubble it up

… a series of different values:

- multiple series
- Selector or pager

- If the result is a chart or a graph
  - Find breakpoints
  - Send up a selection
  - Create a compound visualization

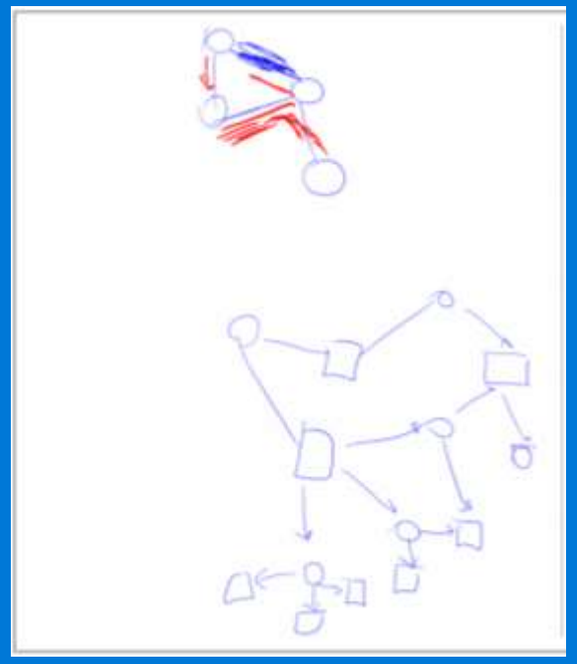
Prototypes & Design Probes

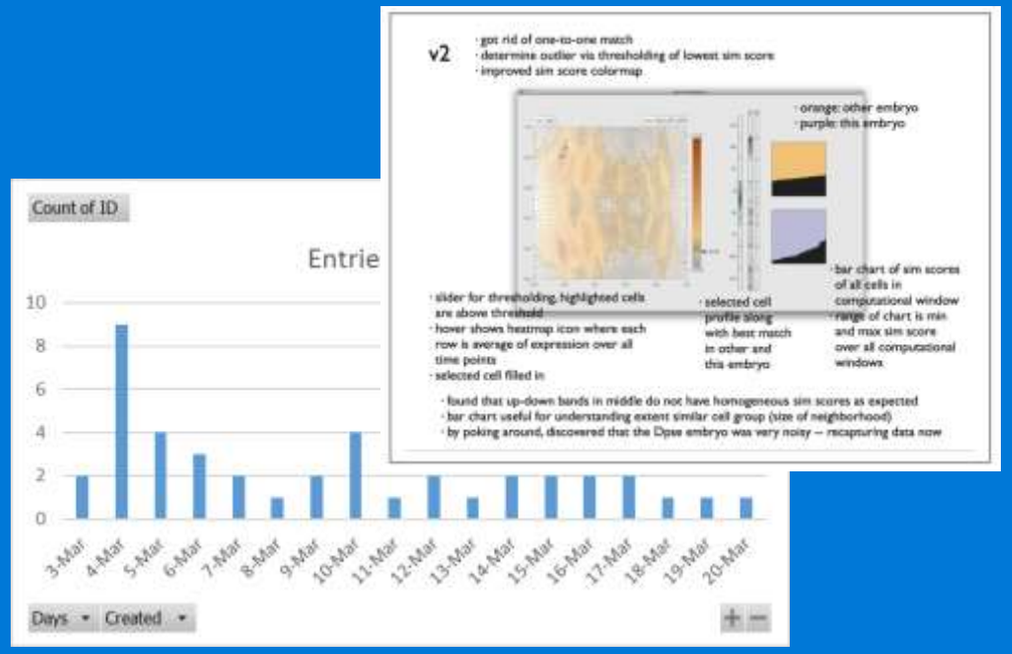
Low fidelity — High fidelity

paper mock-ups | digital mock-up | existing chart tools | plotting & scripting | bespoke viz

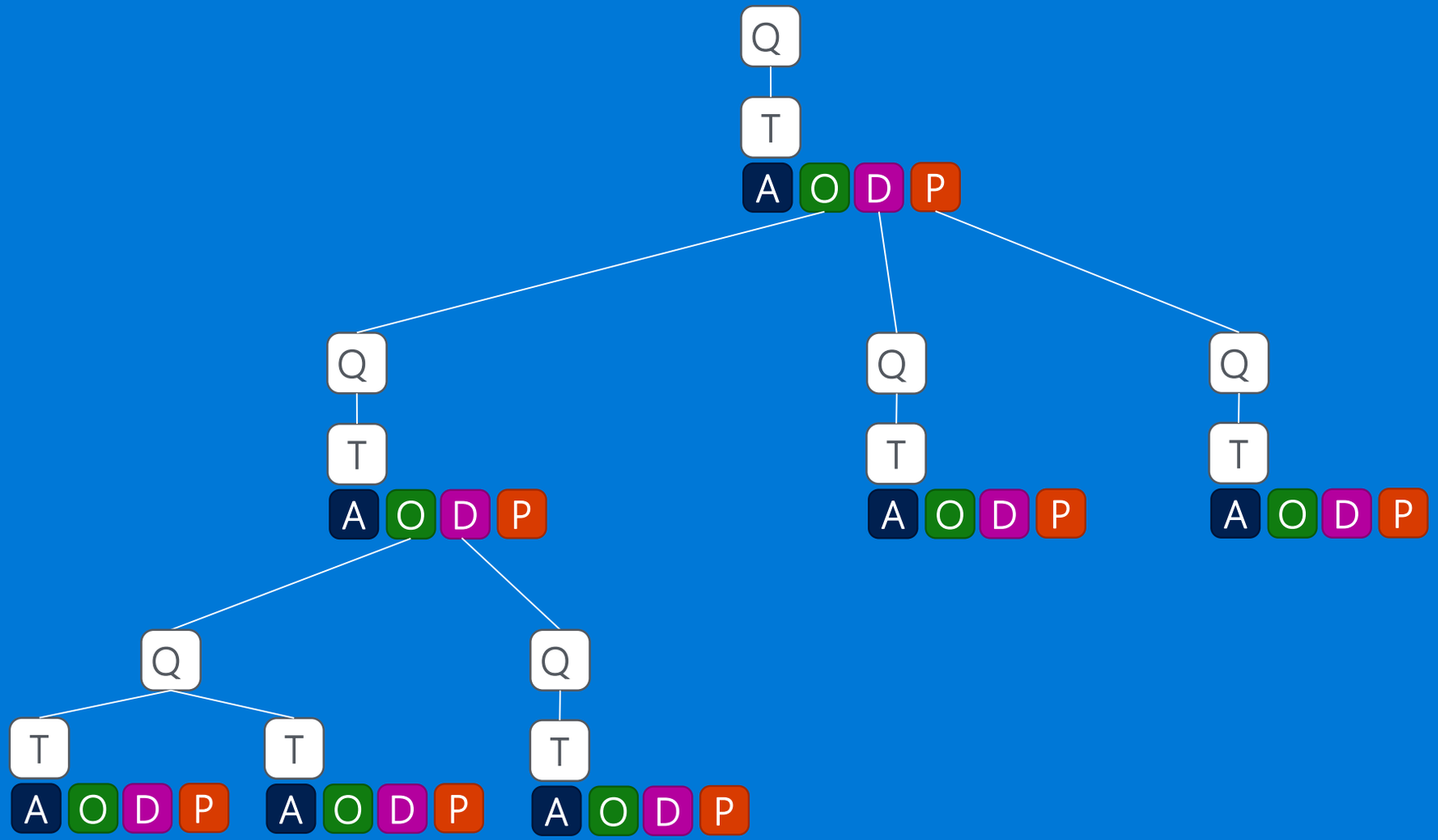
low time & energy commitment
fast iteration

high time & energy
slow iteration





SCORE<1> HI-SCORE SCORE<2>
0120 0000
3
CREDIT 00



Abstract questions

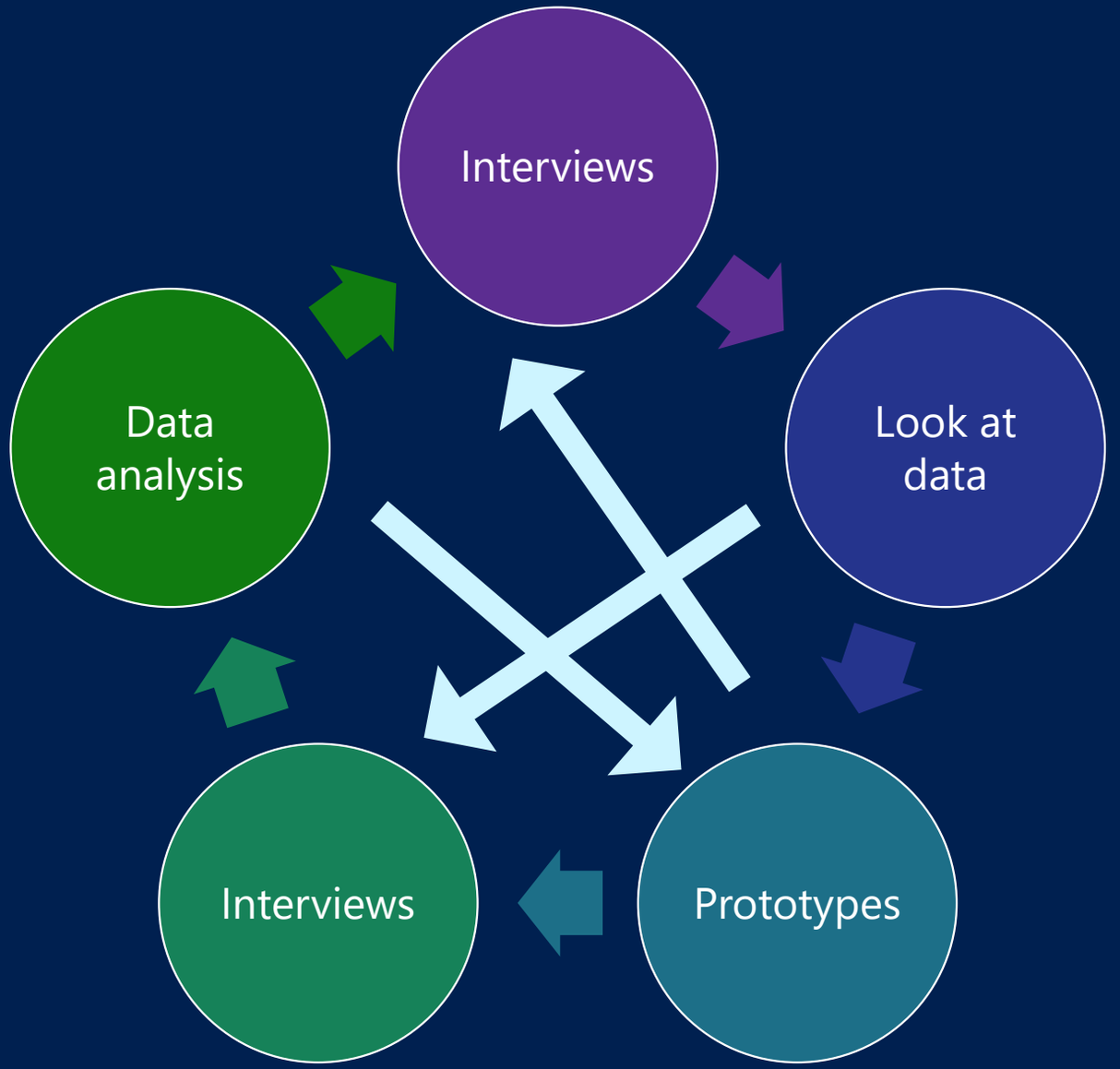
Interactive, compound visualizations

High-level prototypes

EDA, simple visualizations

Detailed questions

It’s an interactive cyclic process



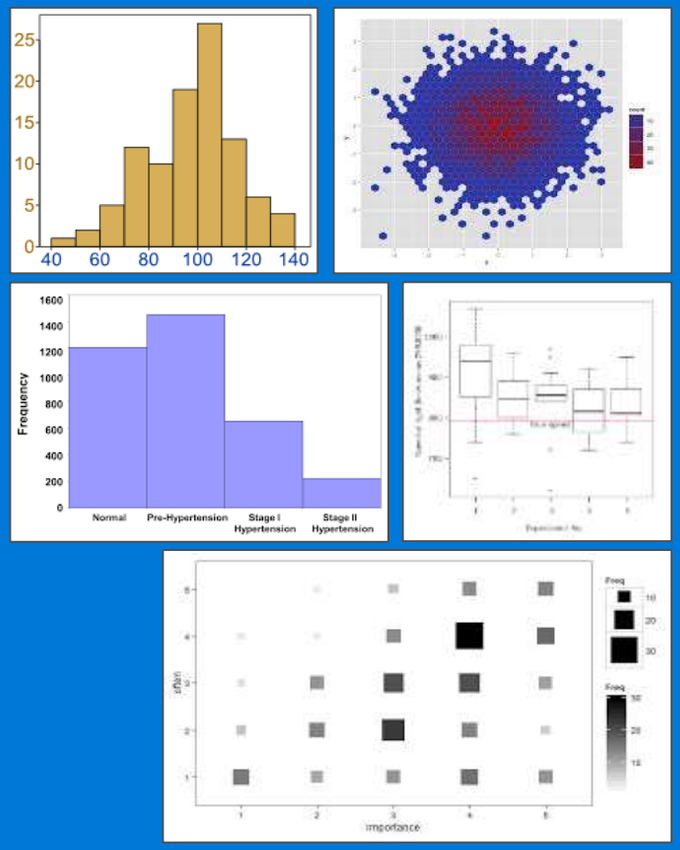
The Tree, and the Process

- Interviews
  - Find tasks and questions
  - figure out what data is available
  - Sanity check operationalization
- EDA
- Prototypes
  - “If we looked at THIS question, it would look like THAT”

- Going down…
  - Seek out ambiguous terms
  - Correct them, make them more specific
- Coming back up…
  - Fill in gaps with the results Propagate upwards to remove ambiguity

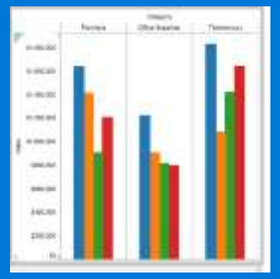
It’s a Task, Not a Visualization

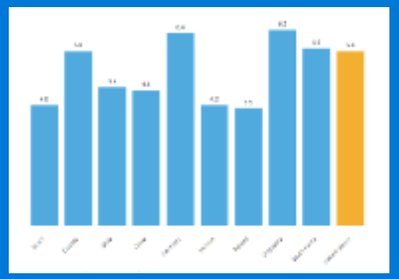
Linking Visuals to Tasks

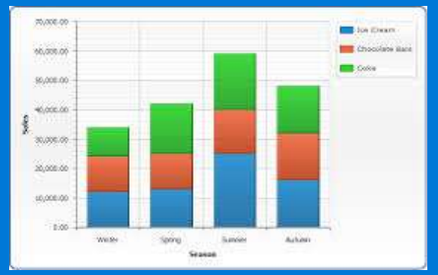


- How much of a thing is there?

Linking Visuals to Tasks

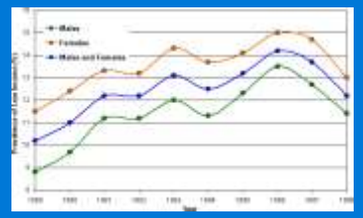


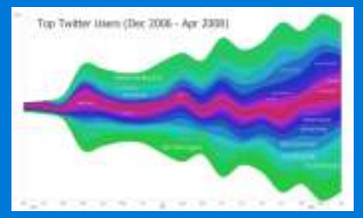




- How much of a thing is there?
- How are groups different from each other?

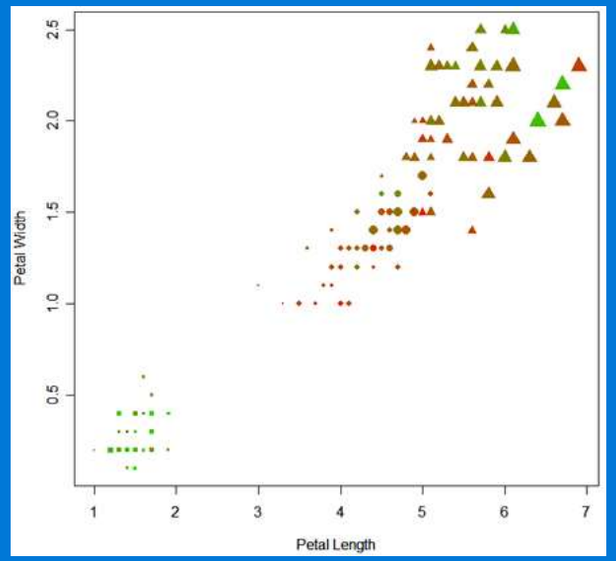
Linking Visuals to Tasks





- How much of a thing is there?
- How are groups different from each other?
- How does something change over time?

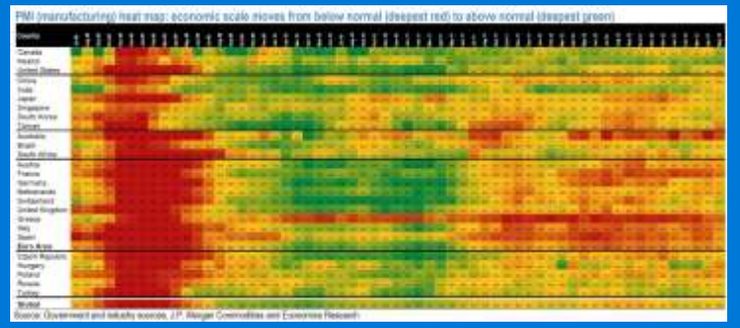
Linking Visuals to Tasks



- How much of a thing is there?
- How are groups different from each other?
- How does something change over time?
- Do things fall into groups, or follow a rule?

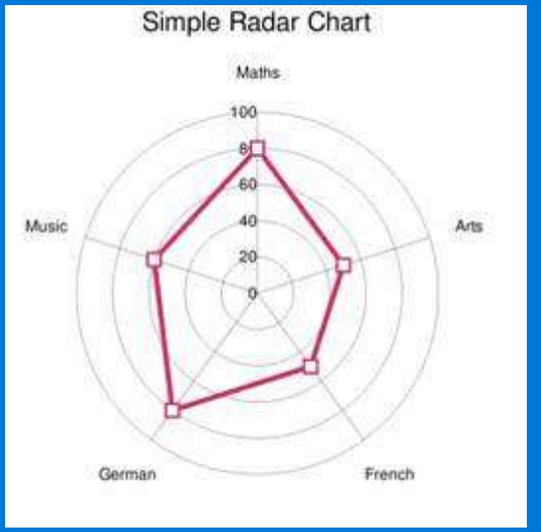
Linking Data to Tasks

- How much of a thing is there?
- How are groups different from each other?
- How does something change over time?
- Do things fall into groups, or follow a rule?
- How does a thing vary under conditions?



Cross-Cutting Concerns

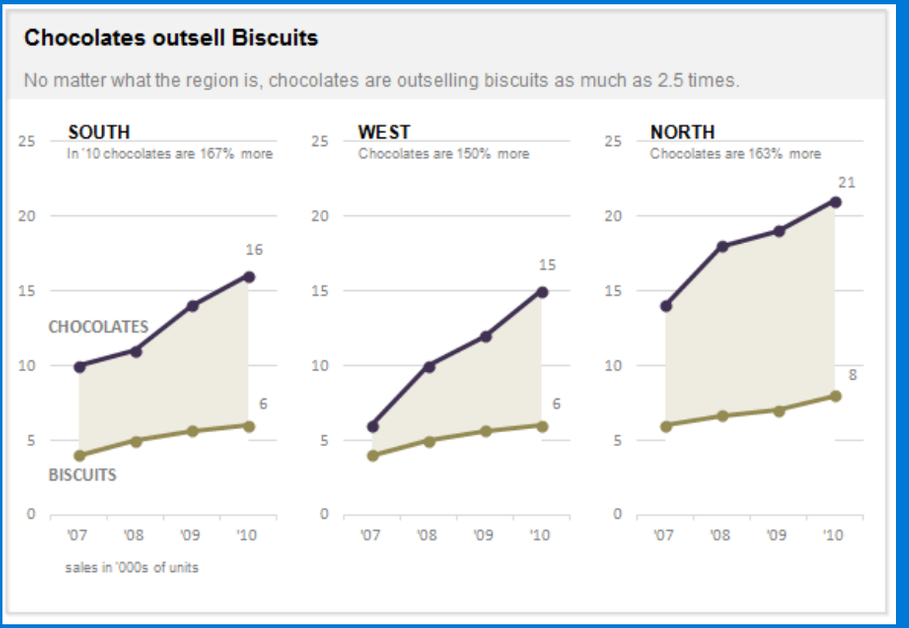
- Main encoding channels: **Space, Color, Size**
- Height & position are easy to compare;
  color is harder
- Lines connect things continuously
- Things that are closer together are easier to compare

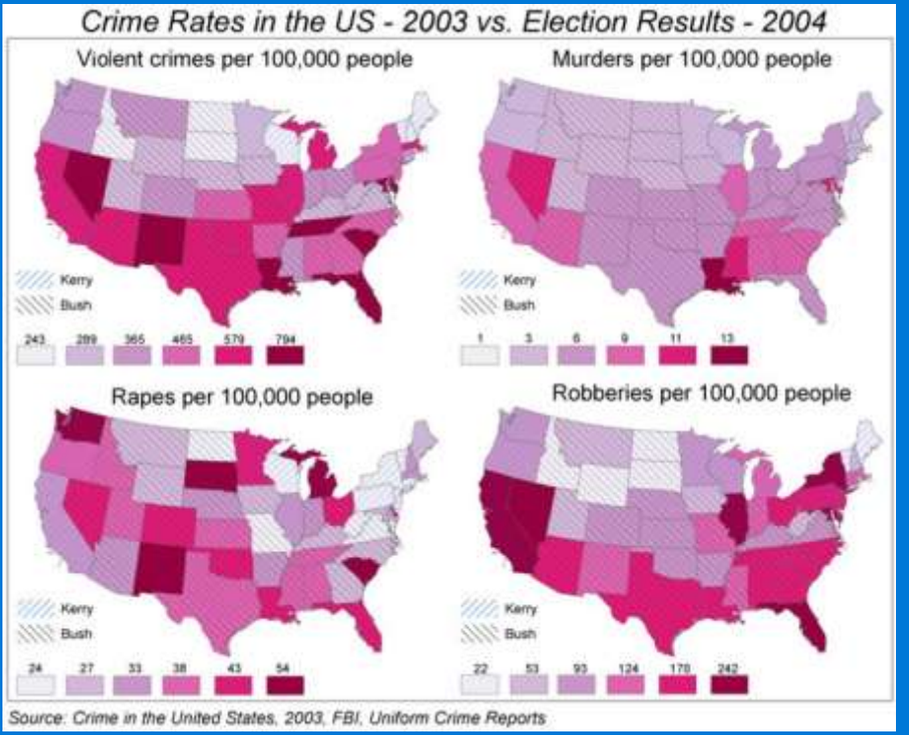


Multiple Views

Small Multiples

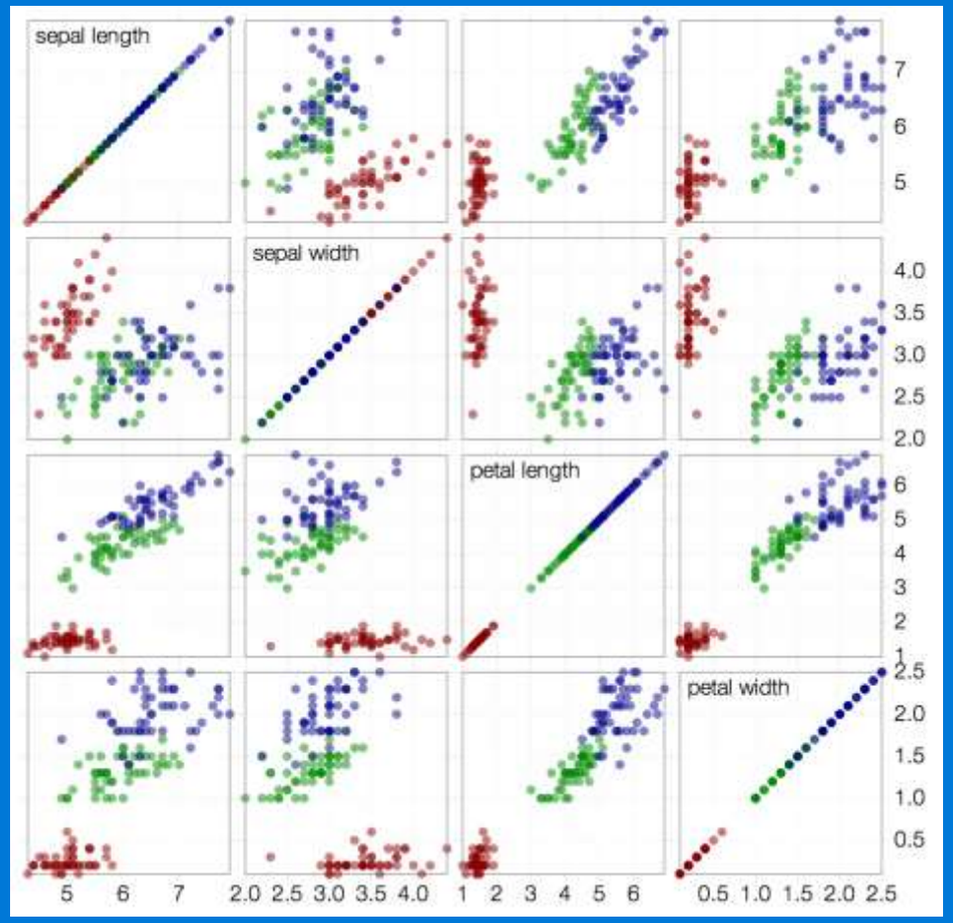
- Compare condition: “For each condition, compare behavior”





Multiple Views

- Brushing & linking
- Animation
- Overlays
- SPLOM: for multiple dimensions in all combinations



Conclusion

- The operationalization tree
- Visualizations fulfill *tasks*

- Visualization creators:
  - Can help systematize the questions you ask
- Visualization learners:
  - Can help direct your process

I'd love your feedback and questions

@fisherdanyel
danyelf@microsoft.com
http://research.microsoft.com/~dany

Microsoft

© 2016 Microsoft Corporation. All rights reserved.