The Lunch Problem
Restaurants Have

A Location
A type (French, ...)
Price ($ - $$$$$$)
Star rating (★ - ★★★★★★)
Different rating (0-10)
Open Mondays
Takes reservations online
(and more...)
How much on a glyph?

A thought exercise…

We are putting symbols on a map
Only restaurants
(don’t even worry about streets)
Glyph Design Problem

A type (French, …)  
Price ($ - $$$$$)  
Star rating (★-5)  
Different rating (0-10)  
Open Mondays  
Online reservations  
(and more…)

Color
Shape
Size
Texture
Angle
Thickness
Elongation
Curvature
Area
Mini-Experiment

Come up with 3 different glyph encodings

You do not need to consider location
You can’t use position

Pack as many variables as possible
Glyph Design Problem

A type (French, ...)
Price ($ - $$$$$$
Star rating (★-5)
Different rating (0-10)
Open Mondays
Online reservations
(and more...)

Color
Shape
Size
Texture
Angle
Thickness
Elongation
Curvature
Area
Exam Question...

What should you have asked first?
Task!

Where are there many restaurants with ...
What properties does this group have ...
What’s the closest restaurant with ...

Is there a restaurant with ...
Are there outliers?
(and many more)
Tasks to encodings

If the task doesn’t require location
  use position for something else

Does the task require all variables at the same time?

Search? Grouping? Averaging?
How much on a glyph?
How much on a glyph?

And, why not more?
What to think about?

What form is the data?
What visual variables do we have?

What tasks do we need to do?
How will those marks interact?
Encodings
How do we map data to marks?
Data Abstractions

What kind of data
Levels of Measurement

Categorical / Nominal
Ordinal
Interval
Ratio

Continuous vs. Discrete
Finite vs. Infinite
Other properties of types
set size
sequential / diverging
continuous / discrete
Conversions between types

Down-conversions easy

Up-conversions harder
imposed orderings on categorical
Visual Variables
# Data Types and Encodings

<table>
<thead>
<tr>
<th>Categorical</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinal</td>
<td>Shape</td>
</tr>
<tr>
<td>Quantitative</td>
<td>Size</td>
</tr>
<tr>
<td>Interval</td>
<td>Texture</td>
</tr>
<tr>
<td>Ratio</td>
<td>Angle</td>
</tr>
<tr>
<td>Small set/Large set</td>
<td>Thickness</td>
</tr>
<tr>
<td>Continuous/Discrete</td>
<td>Elongation</td>
</tr>
<tr>
<td>Diverging / Sequential</td>
<td>Curvature</td>
</tr>
<tr>
<td></td>
<td>Area</td>
</tr>
</tbody>
</table>
Bertin, 1967

O = Ordinal, Q = Quantitative
≠ = Differences = = Similarities

Variables of the Image

XY 2 Dimensions of the Plane

Z

Size

Value

Differential Variables

Texture

Color

Orientation

Shape

Point

Line

Area (Zone)
CHANNEL TYPES

*identity (what or where)*  *magnitude (how much)*

- **Position**
  - Horizontal
  - Vertical
  - Both

- **Color**

- **Shape**

- **Tilt**

- **Size**
  - Length
  - Area
  - Volume
A Gallery of Lines
<table>
<thead>
<tr>
<th>Points</th>
<th>Lines</th>
<th>Areas</th>
<th>Best to show</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shape</strong></td>
<td>possible, but too weird to show</td>
<td>cartogram</td>
<td>qualitative differences</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td></td>
<td>cartogram</td>
<td>quantitative differences</td>
</tr>
<tr>
<td><strong>Color Hue</strong></td>
<td></td>
<td></td>
<td>qualitative differences</td>
</tr>
<tr>
<td><strong>Color Value</strong></td>
<td></td>
<td></td>
<td>quantitative differences</td>
</tr>
<tr>
<td><strong>Color Intensity</strong></td>
<td></td>
<td></td>
<td>qualitative differences</td>
</tr>
<tr>
<td><strong>Texture</strong></td>
<td></td>
<td></td>
<td>qualitative &amp; quantitative differences</td>
</tr>
</tbody>
</table>
Types suggest encodings

Continuous1 x Continuous2

Continuous1 -> X Position
Continuous2 -> Y Position

Sample over range
Mappings (values to positions)
Continuous1 x Continuous2

Continuous1 -> X Position
Continuous2 -> Y Position

Sample over range
Mappings (values to positions)

Interval vs. Ratio?
Continuous1 x Continuous2

Continuous1 -> Y Position
Continuous2 -> Y Position

Specific Samples (Nx(R,R))

Mappings (values to positions)
Continuous1 x Continuous2

Continuous1 -> Y Position
Continuous2 -> Y Position

Specific Samples (Nx(R,R))

Interval -> Color
Derive!
Categorical X Ratio

Categorical -> X position
(mapping? Imposed order)

Ratio -> Y position
(range mapping)
Categorical X Ratio

Categorical -> X position
(mapping? Imposed order)

Ratio -> Y position
(range mapping)

Categorical -> Color
(redundant)
(mapping?)
Categorical X Ratio

Ratio/Category $\rightarrow$ X position
(sort by rank(Ratio))

Ratio $\rightarrow$ Y position
(range mapping)

Categorical $\rightarrow$ Color
(redundant)
(mapping?)
Where is this going?

Principles for choosing encodings mappings
Match designs to data

Not necessarily designs to task
Categorical X Ratio

Categorical -> X position
(mapping? Order by Ratio)

Ratio -> Color
(color encoding)
Categorical X Ratio

Categorical -> X position
(mapping? **Order by Ratio**)  

Ratio -> Color
(color encoding)

Ratio -> Size / Area
(mapping / ramp)
Categorical X
(Categorical, Ratio, Ratio)

Multi-class scatterplot
What data is this good for?

Even more data distinctions

Part/Whole Regions
What do you want to do with these?

Absolute judgment
Relative judgment
Identify / find / match against key
Form groups / regions

Count / quantify
Average / estimate statistics
Pick encodings that...

Match your data type
Are good for the low-level task
Work well with other things
Mackinlay, 1986
Cleveland & McGill, 1984

Figure 3. Graphs from position–angle experiment.

Figure 4. Graphs from position–length experiment.
Heer & Bostock, 2010

Figure 1: Stimuli for judgment tasks T1, T2 & T3. Subjects estimated percent differences between elements.

Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Right: Treemap (T9).
What do you want to do with these?

Absolute judgment
Relative judgment
Identify / find / locate
match against key
Form groups / regions

Count / quantify
Average
What **task(s)** is each of these good for?

Mackinlay, 1986
Position is good for most things

Absolute judgment
Relative judgment
Identify / find / locate
match against key
Form groups / regions

Count / quantify
Average
Position: Relative vs. Absolute

TechCrunch Top 8 Name References
(May 2006 - May 2008)
The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here’s a look at how movies have fared at the box office, after adjusting for inflation.
What do you want to do with these?

Absolute judgment
Relative judgment
Identify / find / locate
match against key
Form groups / regions

Count / quantify
Average
Most things OK for identification
(given sufficient contrast)

Healy web page
Popout examples

- line (blob) orientation
- length, width
- closure
- size
- curvature
- density, contrast
- number, estimation
- colour (hue)
- intensity, binocular lustre
- intersection
- terminators
- 3D depth cues, stereoscopic
Human Perception

A little bit on how we see
E. B. Goldstein “Sensation and Perception” (Adapted from Lindsay & Norman, 1977)
Brain pixels vary enormously in size over the visual field. This reflects differing amounts of neural processing power devoted to different regions of visual space.

At the edge of the visual field we can only barely see something the size of a fist at arm's length.

We can resolve about 100 points on the head of a pin held at arm's length in the very center of the visual field called the fovea.

Over half of our visual processing power is concentrated in a slightly larger area called the parafovea.
Some implications...

We see very little at a time

We build up an image over time using memory

Fast parallel sensing
Slow serial search
Some computer systems thinking...

Lots of sensors (retinal cells)  Lots of computing (neurons...)

- lens
- vitreous humor
- retina
- iris
- pupil
- central fovea
- optical nerve
- cornea
- aqueous humor
- ciliary muscles
- sclera
And a thin cable connected them!

Theory:
Clever coding gets lots of information in a little space

Practice:
Some things can be very efficient
Others have work-arounds
Other bottlenecks

Throughout the stages…
  Competition for Identification
  Competition for selection
Very limited short term memory

Bad news: resource limited
Good news: some parallel mechanisms
Your eyes do some things really well
Your brain needs help with some simple things
Attention

We control where to direct those resources
Pre-Attention

A misnomer!

Involuntary?

Pre-cognitive?

Efficient?

Parallel?
Pre-Attention

A misnomer!

Involuntary?

Pre-cognitive?

Efficient?

Parallel?
Other Perception Tricks
Some computer systems thinking...

Lots of sensors (retinal cells)

Lots of computing (neurons...)
What can you do efficiently?

Locate something that pops out

Sense ensemble properties
  estimate numerosity
  estimate statistics
Encodings for estimation?

Estimating values

Estimating relative differences

Estimating ensemble properties (!)

a diversion – to my work
Ensemble Encoding
Which Color Point is Higher on Average?

Gleicher, M., Correll, M., Nothelfer, C. and Franconeri, C. “Perception of Average Value in Multiclass Scatterplots.” InfoVis 2013
How did you do that?
**Key Results**

- Larger differences give better performance.
- More points do not hurt performance.
- Stronger cues (color) outperform weaker ones.
- Redundant cues do not help performance.
- Conflicting cues do not hurt performance.
- Distractor class does not hurt performance.

Gleicher, M., Correll, M., Nothelfer, C. and Franconeri, C. “Perception of Average Value in Multiclass Scatterplots.” InfoVis 2013
Factors to consider

Accuracy (estimate value)
Discriminability (tell apart)
Separability (see different things)

And others...
Grouping
Multi-Variate Glyph
(from the Paris Apartment Problem)

Position = Location

Price (ordinal) -> Size
Type (categorical) -> Hue
Rating (ordinal) -> Lightness/Saturation
Open Monday (categorical/binary) -> Shape
How much can you encode?

Restaurant
location

price ($, $$, $$$, $$$$)
type (French, Italian, …)
rating (★, ★★, ★★★, ★★★★)
open Mondays
Cue Interactions
Cue Interactions
SEPARABLE vs INTEGRAL

separable: can judge each channel individually

integral: two channels are viewed holistically

Ware 2004
SEPARABLE vs INTEGRAL

FIGURE 3.38. An example of the use of an ellipse as a map symbol in which the horizontal and vertical axes represent different (but presumably related) variables.
SEPARABLE vs INTEGRAL

- separable
  - color | location
  - color | motion

- integral
  - color | shape
  - size | orientation
  - x-size | y-size
  - red-green | yellow-blue

Ware 2004
Factors to consider

Accuracy
Discriminability
Separability

Popout, Search, Average, ...

Salience, Semantics, ...
Taxonomy-Based Glyph Design—with a Case Study on Visualizing Workflows of Biological Experiments

Eamonn Maguire, Philippe Rocca-Serra, Susanna-Assunta Sansone, Jim Davies, and Min Chen

Fig. 1. a) Workflow as rendered currently using toolkits such as GraphViz. b) We propose to replace the textual labels with glyphs, while allowing interactive access to detailed descriptions. This makes it easy to gain an overview, search components and compare workflows. The screenshot shows a prototype developed within ISACreator, a system for capturing biological experiment metadata.

Abstract—Glyph-based visualization can offer elegant and concise presentation of multivariate information while enhancing speed and ease in visual search experienced by users. As with icon designs, glyphs are usually created based on the designers' experience and intuition, often in a spontaneous manner. Such a process does not scale well with the requirements of applications where a large number of concepts are to be encoded using glyphs. To alleviate such limitations, we propose a new systematic process for glyph design by exploring the parallel between the hierarchy of concept categorization and the ordering of discriminative capacity of visual channels. We examine the feasibility of this approach in an application where there is a pressing need for an efficient and effective means to visualize workflows of biological experiments. By processing thousands of workflow records in a public archive of biological
<table>
<thead>
<tr>
<th>Inputs and Outputs</th>
<th>Input-output</th>
<th>Input-output</th>
<th>Input-output</th>
<th>Input-output</th>
<th>Input-output</th>
<th>Input-output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biological</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Device</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Vitro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Vivo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Silico</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material perturbation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material separation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material amplification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material combination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molecule.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cellular Port</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Material included perturbation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviourally included perturbation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physically included perturbation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>Example 1</td>
<td>Example 2</td>
<td>Example 3</td>
<td>Example 4</td>
<td>Example 5</td>
<td>Example 6</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Molecule.</td>
<td>![Image]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cellular Part</td>
<td>![Image]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell</td>
<td>![Image]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tissue</td>
<td>![Image]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organ</td>
<td>![Image]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>![Image]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>![Image]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some Examples

Name that encoding!
Critique that encoding!
The chance a child raised in the bottom fifth rose to the top fifth

- 35%
- 20%
- 15%
- 10%
- 4%

The top fifth is equal to family income of more than $70,000 for the child by age 30, or more than $100,000 by age 45.

In areas like Atlanta, upward mobility appears to be substantially lower than in any other rich country.

- Boston: 9.8%
- New York: 9.7%
- Charlotte: 4.3%
- Atlanta: 4.0%

Salt Lake City: 11.5%
L.A.: 9.6%
S.F.: 11.2%
Chicago: 6.1%
Houston: 8.4%
The Quantified Selfie

You've probably often wondered: Who tilts their heads more when taking selfies, the women of Bangkok or São Paulo? To find out, researcher Lev Manovich and his colleagues used Mechanical Turk to crowdsourced the analysis of 3,200 selfies posted on Instagram during one week in December 2013. (Results: São Paulans are the head-tiltingest selfie takers.) Now the team has gone back to the data, exploring the tags we attach to our smartphone photos. Here's what's trending. —RACHEL SOMERSTEIN

Who uses the widest variety of selfie tags (by number of unique tags):

- Bangkok: 1,000
- Moscow: 500
- São Paulo: 300
- New York: 0
- Berlin: 1,000

Instagrammers talk a lot. But Berlin and New York City have the biggest vocabularies — those cities use more unique tags than anywhere else.

Top 10 tags in each city (by percentage of selfies that use them):

- Bangkok: girl, me, thailand, love, like, morning, photooftheday, instalike, instalove, instalife
- Moscow: girl, me, Moscow, love, like, goodnight, instalike, instalove, instalife, instalife
- São Paulo: love, me, like, instalike, instalove, instalife, instalife, instalife, instalife, instalife
- New York: nyc, selfie, me, like, instalike, instalove, instalife, instalife, instalife, instalife
- Berlin: love, me, Berlin, like, instalike, instalove, instalife, instalife, instalife, instalife

Big surprise, "love" makes the top 10 in every city but NYC, the only one where "work" makes the top 10. And self-snappers seem to feel the need to say where they are: less so for Brazilians, though, who really like their "likes."