Visualization: Abstractions & Idioms

Tamara Munzner
Department of Computer Science
University of British Columbia

University of Coimbra Guest Lecture
9 Mar 2022, virtual / Coimbra, Portugal

http://www.cs.ubc.ca/~tmn/talks.html#coimbra22
Visualization defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- suitable when human in the loop needs details
- interplay between human judgement and automatic computation

Anscombe’s Quartet

<table>
<thead>
<tr>
<th>Identical statistics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x mean</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>x variance</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>y mean</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>y variance</td>
<td>3.75</td>
<td></td>
</tr>
<tr>
<td>x/y correlation</td>
<td>0.816</td>
<td></td>
</tr>
</tbody>
</table>

Datasaurus Dozen

Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. CHI 2017. Matejka & Fitzmaurice
Why focus on tasks and effectiveness?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- effectiveness requires match between data/task and representation
  - set of representations is huge
  - many are ineffective mismatch for specific data/task combo
  - increases chance of finding good solutions if you understand full space of possibilities
Metaphor: Design space

+ good
○ okay
- poor
Metaphor: Design space

+ good
○ okay
- poor

some technique...
Metaphor: Design space

know
Metaphor: Design space

know

consider
Metaphor: Design space

know
consider
propose
select
Metaphor: Design space

Think broad!
Why focus on tasks and effectiveness?

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

• effectiveness requires match between data/task and representation
  – set of representations is huge
  – many are ineffective mismatch for specific data/task combo
  – increases chance of finding good solutions if you understand full space of possibilities

• what counts as effective?
  – novel: enable entirely new kinds of analysis
  – faster: speed up existing workflows

• how to validate effectiveness
  – many methods, must pick appropriate one for your context
Nested model: Four levels of visualization concerns

Nested model: Four levels of visualization concerns

• *domain* situation
  – *who* are the target users?
Nested model: Four levels of visualization concerns

• *domain* situation
  – **who** are the target users?

• *abstraction*
  – translate from specifics of domain to vocabulary of vis
Nested model: Four levels of visualization concerns

- **domain** situation
  - who are the target users?

- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - what is shown? **data abstraction**


Nested model: Four levels of visualization concerns

• **domain** situation
  – **who** are the target users?

• **abstraction**
  – translate from specifics of domain to vocabulary of vis
  – **what** is shown? **data abstraction**
    • often don’t just draw what you’re given: transform to new form

[A Multi-Level Typology of Abstract Visualization Tasks

[A Nested Model of Visualization Design and Validation.
Nested model: Four levels of visualization concerns

• **domain** situation
  – **who** are the target users?

• **abstraction**
  – translate from specifics of domain to vocabulary of vis
  – **what** is shown? data abstraction
    • often don’t just draw what you’re given: transform to new form
  – **why** is the user looking at it? task abstraction


Nested model: Four levels of visualization concerns

- **domain situation**
  - **who** are the target users?

- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don’t just draw what you’re given: transform to new form
  - **why** is the user looking at it? **task abstraction**

- **idiom**
  - **how** is it shown?
Nested model: Four levels of visualization concerns

• **domain** situation
  – **who** are the target users?

• **abstraction**
  – translate from specifics of domain to vocabulary of vis
  – **what** is shown? **data abstraction**
    • often don’t just draw what you’re given: transform to new form
  – **why** is the user looking at it? **task abstraction**

• **idiom**
  – **how** is it shown?
    • **visual encoding idiom**: how to draw

[A Multi-Level Typology of Abstract Visualization Tasks

[A Nested Model of Visualization Design and Validation.
Nested model: Four levels of visualization concerns

- **domain situation**
  - **who** are the target users?

- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don’t just draw what you’re given: transform to new form
  - **why** is the user looking at it? **task abstraction**

- **idiom**
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate
Nested model: Four levels of visualization concerns

- **domain situation**
  - **who** are the target users?

- **abstraction**
  - translate from specifics of domain to vocabulary of vis
  - **what** is shown? **data abstraction**
    - often don’t just draw what you’re given: transform to new form
  - **why** is the user looking at it? **task abstraction**

- **idiom**
  - **how** is it shown?
    - **visual encoding idiom**: how to draw
    - **interaction idiom**: how to manipulate

- **algorithm**
  - efficient computation

---


Why is validation difficult?

• different ways to get it wrong at each level

Why is validation difficult?

- different ways to get it wrong at each level

Domain situation
You misunderstood their needs

Why is validation difficult?

• different ways to get it wrong at each level

[Domain situation]
You misunderstood their needs

[Data/task abstraction]
You're showing them the wrong thing

Why is validation difficult?

• different ways to get it wrong at each level

- **Domain situation**
  You misunderstood their needs

- **Data/task abstraction**
  You’re showing them the wrong thing

- **Visual encoding/interaction idiom**
  The way you show it doesn’t work

Why is validation difficult?

- different ways to get it wrong at each level

- Domain situation
  - You misunderstood their needs

- Data/task abstraction
  - You're showing them the wrong thing

- Visual encoding/interaction idiom
  - The way you show it doesn’t work

- Algorithm
  - Your code is too slow

Validation solution: use methods from appropriate fields at each level

Validation solution: use methods from appropriate fields at each level

computer science

Algorithm
Measure system time/memory
Analyze computational complexity

Validation solution: use methods from appropriate fields at each level

- Algorithm
  - Measure system time/memory
  - Analyze computational complexity

Validation solution: use methods from appropriate fields at each level

Visual encoding/interaction idiom
Justify design with respect to alternatives

Algorithm
Measure system time/memory
Analyze computational complexity

Analyse results qualitatively
Measure human time with lab experiment (lab study)

technique-driven work

design
computer science
cognitive psychology

Validation solution: use methods from appropriate fields at each level

- **Domain situation**
  - Observe target users using existing tools

- **Data/task abstraction**
  - **Visual encoding/interaction idiom**
    - Justify design with respect to alternatives
  - **Algorithm**
    - Measure system time/memory
    - Analyze computational complexity
    - Analyze results qualitatively
    - Measure human time with lab experiment (*lab study*)
    - Observe target users after deployment (*field study*)

- **Anthropology/ethnography**
  - Measure adoption

Validation solution: use methods from appropriate fields at each level

- **Domain situation**: Observe target users using existing tools
- **Data/task abstraction**
  - **Visual encoding/interaction idiom**: Justify design with respect to alternatives
  - **Algorithm**: Measure system time/memory, Analyze computational complexity
  - Analyze results qualitatively
  - Measure human time with lab experiment (*lab study*)
  - Observe target users after deployment (*field study*)
  - Measure adoption

---

Validation solution: use methods from appropriate fields at each level

- avoid mismatches between level and validation

<table>
<thead>
<tr>
<th>Domain situation</th>
<th>Data/task abstraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observe target users using existing tools</td>
<td>Visual encoding/interaction idiom</td>
</tr>
<tr>
<td>Justify design with respect to alternatives</td>
<td>Algorithm</td>
</tr>
<tr>
<td>Measure system time/memory</td>
<td>Measure human time with lab experiment (lab study)</td>
</tr>
<tr>
<td>Analyze computational complexity</td>
<td>Analyze results qualitatively</td>
</tr>
<tr>
<td></td>
<td>Observe target users after deployment (field study)</td>
</tr>
<tr>
<td></td>
<td>Measure adoption</td>
</tr>
</tbody>
</table>

Visualization: Angles of attack

problem-driven work
Visualization: Angles of attack

problem-driven work  \(\rightarrow\)  technique-driven work
Visualization: Angles of attack

problem-driven work

evaluation

technique-driven work
Visualization: Angles of attack

problem-driven work

---

technique-driven work

---

theoretical foundations

---

evaluation
Three case studies: Abstractions & idioms

• e-commerce

• facilities management

• biology
Three case studies: Abstractions & idioms

- e-commerce
- facilities management
- biology
Segmentifier

Interactive Refinement of Clickstream Data


Segmentifier: Interactive Refinement of Clickstream Data.
E-commerce: mobile apps for large companies
What are the **Data and Task Abstractions** for **Clickstream Data Analysis**?
What is *Clickstream Data*?
Data: *Actions*
Data: Action Attributes
Data: Action Types

E-commerce
- addToCart
- removeFromCart
- search
- purchase

Site Functionality
- appStart
- appDisplayError
- offlineModeUsed

Pageviews
- pageview
Action Hierarchy
Action Hierarchy
Data: Sequences
Data: Sequences
Data: Client Sequences

*Client Sequences*: all actions performed by a single user
Session Sequences: all actions performed by a single user within a defined amount of time ($\Delta$) from each other. $\Delta$ is usually 30 min.
Data: Sequence Attributes

<table>
<thead>
<tr>
<th>Start time</th>
<th>End time</th>
<th>Duration</th>
<th>Action Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>: 5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>: 1</td>
</tr>
</tbody>
</table>
Data: **Segments**

**Segment**: any set of sequences
Data: Segment Attributes

Counts of sequences:
Absolute, Relative

Sequence Distributions:
Start Time, Duration, Action Counts

Action Distributions:
Action Transitions:
action before, action after
Real-world Clickstream Data
Real-world Clickstream Data

Scale is huge
Real-world Clickstream Data

Scale is huge

Variability is high
Real-world Clickstream Data

Scale is huge

Variability is high

Most work **fails** when applied to real-world data
What are *Clickstream Data Analysis Tasks*?
Tasks: Segment Behavior

Behavior: set of attribute constraints

Segment

Viewed 4 pages
Purchased
Between 9 - 10 am

Start time
Tasks: Segment Behavior

**Behavior:** set of attribute constraints

- **Expected**
  *Users add to cart before purchasing*

- **Unexpected**
  *No purchases on a certain month*

- **Favorable**
  *Purchased*

- **Unfavorable**
  *Bounced*
Tasks: Task Abstraction

Identify: Find some set of sequences that constitutes interesting behavior.
Tasks: Task Abstraction

Identify: Find some set of sequences that constitutes interesting behavior.

Drilldown: Distinguish more specific behaviors to further partition a segment previously defined by looser constraints.
**Tasks: Task Abstraction**

Identify: Find some set of sequences that constitutes interesting behavior

Drilldown: Distinguish more specific behaviors to further partition a segment previously defined by looser constraints

Frequency: Determine how many sequences are in the segment defined by behavior
**Tasks: Task Abstraction**

<table>
<thead>
<tr>
<th>Segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Segment Image]</td>
</tr>
</tbody>
</table>

**Identify:** Find some set of sequences that constitutes interesting *behavior*.

**Drilldown:** Distinguish more specific *behaviors* to further partition a segment previously defined by looser constraints.

**Frequency:** Determine how many sequences are in the segment defined by *behavior*.

**Ordering** within sequence: Match if one action subsequence occurs before (or after) another action subsequence in a sequence.
High-Level Segmentifier Analysis Model
High-Level Segmentifier Analysis Model

- Abstraction above task/data level to provide design rationale
- Take a giant, noisy dataset and refine it into small, clean segments for
  - actionable insights
  - downstream analysis
- Bridge the gap between real-world data and other techniques
High-Level Segmentifier Analysis Model

- Gives Insight into underlying data of segment
  - Action Attributes
  - Sequence Attributes
  - Segment Attributes
- Leads to:
  - Insights
  - New ways on how to refine
  - Whether segment should be abandoned
  - Whether segment should be exported
High-Level Segmentifier Analysis Model

- Apply operation to create new segments
- Type of Refinements
  - Filter
  - Partition
  - Transform
High-Level Segmentifier Analysis Model

- Record all refinement steps automatically
- Keep track of questions asked and hypotheses tested
- Ability to create and view multiple segments from the same segment
High-Level Segmentifier Analysis Model

- Export refined segments for further downstream analysis, to more specific tools:
  - Pattern mining
  - Clustering
High-Level Segmentifier Analysis Model

- Discover actionable insight by viewing segment
By viewing the segment, analyst abandons if:
- No actionable insights
- No further ways to refine
- Not suitable for export
Why Visual Analytics?

- Automation would be nice...
  - Put data in, actionable results appear
- … but it is not realistic
  - Many possible questions, data-driven interplay between finding answers and generating new questions
- Human-in-the-loop visual data analysis
  - Integrate computing power of machine with intuition of domain experts
Solution
The Segmentifier Interface
Segmentifier: Interactively Refining Clickstream Data into Actionable Segments

https://www.youtube.com/watch?v=TobYDFeISOg
Segmentifier Contributions

➤ Thorough **characterization of task and data abstraction** for clickstream data analysis
Segmentifier Contributions

➢ Thorough **characterization of task and data abstraction** for clickstream data analysis

➢ **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis
Segmentifier Contributions

➢ Thorough **characterization of task and data abstraction** for clickstream data analysis

➢ **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis

➢ Preliminary **evidence of utility**
Three case studies of problem-driven work

• e-commerce

• facilities management

• biology
Ocupado

Visualizing Location-Based Counts Over Time Across Buildings


Ocupado: Visualizing Location-Based Counts Over Time Across Buildings.

Video

https://www.youtube.com/watch?v=KcwjVK8eUdw
Location-Based Counts

- Regular intervals (e.g., every 5 minutes)
- Spatial hierarchy (Zone → Floor → Building → Campus)
- No trajectories or device identifiers are recorded
- Intrinsic privacy advantages
Data

Automated HVAC control
Data
Data

Decision making
WiFi connections as a proxy for occupancy
WiFi connections as a proxy for occupancy
Interviews with potential stakeholders
Focus Domains

- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control
Focus Domains

- Space planning
- Building management
- Custodial services
- Classroom management
- Data quality control

Semi-structured discussions and live demos
Tasks

Confirm assumptions or previous observations.
Do students occupy room x in evenings or on weekends?
Tasks

- **Confirm** assumptions or previous observations.

- **Monitor** the current/recent utilization rate.

  Which rooms are empty/busy?
Tasks

- Confirm assumptions or previous observations.
- Monitor the current/recent utilization rate.
- Communicate space usage and justify decisions.

Space usage improved after renovation.
Tasks

- **Confirm** assumptions or previous observations.
- **Monitor** the current/recent utilization rate.
- **Communicate** space usage and justify decisions.
- **Validate** the data (quality control).
  Check minimum size of a room that can be captured.
Spatial and Temporal Data Granularities
Visualization Prototypes

Sandbox

Data sketches,
static data export

Time
Visualization Prototypes

- original plan: different interface for each stakeholder
- realization: task & data abstractions match multiple stakeholders
  - if slice by space & time granularity

Sandbox

Data sketches, static data export
Spatial and Temporal Data Granularities

Regions of interest

Zone

Floor

Building
Spatial and Temporal Data Granularities

Regions of interest

- Zone
- Floor
- Building

Periods of interest

- Mondays
- Weekdays
- last 12 hours
- Summer term
- Fr 8-10am
- Weekends
Visualization Prototypes

**Sandbox**
- Data sketches, static data export

**Campus Explorer**
- Live-data stream, cross-building analysis

**Building Recent**

**Building Long-term**

**Region Compare**

Time
Reusable Visualization Components
# Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparkline</td>
<td></td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
</tr>
</tbody>
</table>

**How?**
- Encode
- Manipulate
- Facet
- Reduce
- Arrange
- Map
- Change
- Select
- Navigate
- Express
- Separate
- Order
- Align
- Use
- Juxtapose
- Partition
- Superimpose

**Why?**

**What?**
# Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
</tr>
<tr>
<td></td>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (non-contiguous)</td>
</tr>
</tbody>
</table>

**How?**
- Encode
- Manipulate
- Facet
- Reduce
- Arrange
- Map
- Change
- Select
- Navigate
- Express
- Separate
- Order
- Align
- Use
- Juxtapose
- Partition
- Superimpose
- Filter
- Aggregate
- Embed

**Why?**
- What?
## Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
</tr>
<tr>
<td></td>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (non-contiguous)</td>
</tr>
<tr>
<td></td>
<td>Confidence band</td>
<td>Aggregation</td>
<td>Typical utilization profiles</td>
</tr>
</tbody>
</table>

**How?**
- Encode
- Manipulate
- Facet
- Reduce
- Arrange
- Map
- Change
- Select
- Navigate
- Express
- Separate
- Order
- Align
- Use
- Juxtapose
- Partition
- Superimpose
- Filter
- Aggregate
- Embed

**Why?**
- How?
- What?
# Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
</tr>
<tr>
<td></td>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (non-contiguous)</td>
</tr>
<tr>
<td></td>
<td>Confidence band</td>
<td>Aggregation</td>
<td>Typical utilization profiles</td>
</tr>
<tr>
<td></td>
<td>line chart</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Superimposed line chart</td>
<td>Superposition</td>
<td>Within-session patterns, outliers</td>
</tr>
</tbody>
</table>
## Reusable Visualization Components

<table>
<thead>
<tr>
<th>Temporal</th>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sparkline</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (contiguous)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Box-plot-bars</td>
<td>Juxtaposition</td>
<td>Repeating patterns, trends, outliers (non-contiguous)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Confidence band line chart</td>
<td>Aggregation</td>
<td>Typical utilization profiles</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Superimposed line chart</td>
<td>Superposition</td>
<td>Within-session patterns, outliers</td>
<td></td>
</tr>
</tbody>
</table>

### Why? How? What?

- **Why?**
  - **Encode**
  - **Manipulate**
  - **Facet**
  - **Reduce**
  - **Arrange**
  - **Map**
  - **Change**
  - **Select**
  - **Navigate**
  - **Express**
  - **Separate**
  - **Order**
  - **Align**
  - **Use**
  - **Juxtapose**
  - **Partition**
  - **Superimpose**
  - **Filter**
  - **Aggregate**
  - **Embed**

- **How?**
  - **Color**
  - **Motion**
  - **Position, Size, Angle, Curvature, ...**
  - **Hue**
  - **Saturation**
  - **Luminance**
  - **Transparency**
  - **Region, Shape, ...**
  - **Direction, Rate, Frequency, ...**

- **What?**
Reusable Visualization Components

<table>
<thead>
<tr>
<th>Layout</th>
<th>Visual Encoding</th>
<th>Facet</th>
<th>Comparisons</th>
</tr>
</thead>
</table>
| Temporal
| Sparkline       | Juxtaposition    | Repeating patterns, trends, outliers (contiguous) |
| Box-plot-bars  | Juxtaposition    | Repeating patterns, trends, outliers (non-contiguous) |
| Confidence band line chart | Aggregation      | Typical utilization profiles                      |
| Superimposed line chart | Superposition    | Within-session patterns, outliers                 |
| Spatial
| Floor plan with symbols | Superposition    | Within local spatial neighborhood                 |
| Spatial heatmap | Containment (nested) | Across distributed regions                         |
Ocupado Interfaces
Ocupado Contributions

- Analysis and abstraction of data and tasks for studying space utilization
- Ocupado, a set of visual decision support tools
- Generalizable design choices for visualizing non-trajectory spatiotemporal data relating to large-scale indoor environments
MizBee

A Multiscale Synteny Browser

joint work with:
Miriah Meyer, Hanspeter Pfister

http://www.cs.utah.edu/~miriah/mizbee

MizBee: A Multiscale Synteny Browser.
https://www.youtube.com/watch?v=86p7brwuz2g
What: Data abstraction

• data: multiscale lists
  – features: hundreds of thousands
    • ordered attribute: position in chromosome sequence coordinates
    • categorical attributes: orientation, chromosome of matching feature
    • quantitative attributes: length, similarity score
  – syntenic blocks: thousands
    • contiguous sets of features on same chromosome
    • combine thresholded features if
      – destination chromosome and orientation match
      – close together
  – chromosomes: dozens
  – genomes: two
Why: Tasks in domain language

• analyze conservation (similarity) relationships between genomic features
  – high-level biology questions
    • evolution
      – how long ago did two species share common ancestor?
    • function
      – which segment of the genome is responsible for specific function in the cell?
    • ...
  – low-level data-centric questions
    • algorithm refinement
      – are paired features within a block contiguous?
      – which chromosomes share conserved blocks?
      – are similarity scores alike within block?
      – ...


## Why: Tasks abstraction

- **relationship types:** proximity, size, orientation, similarity
- **data scales:** genome, chromosome, block, feature
- **topics:** algorithm in/out, block reliability, high-level science

<table>
<thead>
<tr>
<th>Question</th>
<th>Genome</th>
<th>Chromosome</th>
<th>Block</th>
<th>Feature</th>
<th>Proximity/Location</th>
<th>Size</th>
<th>Orientation</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which chromosomes share conserved blocks?</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For one chromosome, how many other chromosomes does it share blocks with?</td>
<td>× ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the density of coverage and where are the gaps on: chromosomes? blocks?</td>
<td>× × ×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Where are the blocks: on chromosomes? around a specific location on a chromosome?</td>
<td>× ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What are the sizes and locations of other genomic features near a block?</td>
<td>× × ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How large are the blocks?</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do neighboring blocks go to the same: chromosomes? relative location on a chromosome?</td>
<td>× ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are the orientations matched or inverted for: block pairs? feature pairs?</td>
<td>× ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do the orientations match for pairs of: neighboring blocks? features within a block?</td>
<td>× ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are similarity scores alike: with respect to neighboring blocks? within a block?</td>
<td>× ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are the paired features within a block contiguous?</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How large is a feature relative to other genes within a block?</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What are the sizes, locations, and names of features within a block?</td>
<td>× × ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What are the differences between individual nucleotides of feature pairs?</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How: Idiom design choices

- encode match relationships between chromosome segments with both
  - color
    
    ![chromosome segments with color](chromosome_color.png)
  - connection marks
    
    ![chromosome segments with connection marks](chromosome_marks.png)

- **Identity Channels: Categorical Attributes**
  - Spatial region
  - Color hue
  - Motion
  - Shape

- **Marks As Links**
  - Containment
  - Connection
How: Arrange space

- design space of arrangements

<table>
<thead>
<tr>
<th>contiguous</th>
<th>discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>segregated</td>
</tr>
<tr>
<td>circular</td>
<td>interleaved</td>
</tr>
</tbody>
</table>

Mauve [Darling04]

Cinteny

MizBee

Apollo [Lewis02]
How: Idiom design choices

• juxtapose linked views
  – multiform overview-detail
    • three views: genome, chromosome, block
    • different visual encoding in each
How: Idiom design choices

• axis orientation
  – radial: genome
  – rectilinear: chromosome, block
    • aligned position more accurate than angle

Axial Orientation
  → Rectilinear → Parallel → Radial

Magnitude Channels: Ordered Attributes
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
How: Idiom design choices

• filter
How: Idiom design choices

• outer ring: summarize relationships with color
  – select one chromosome from set of source chromosomes

• inner ring:
  – destination chromosomes around copy of selected source chromosome
  – show relationship details with connection marks as well as color
MizBee contributions

• first synteny browser with side-by-side linked views
  – across the range of scales
  – encoding all four conservation relationship types
    • proximity, size, orientation, similarity

• open source
  http://www.cs.utah.edu/~miriah/mizbee
Visualization: Abstractions & idioms

• levels of design
  – identify abstractions
    • crucial & difficult, iterative process
  – select appropriate idioms
    • or create new ones if necessary

• three examples
  – different domains
  – different abstractions
  – different idioms
More information

• theoretical foundations: book (+ tutorial/course lecture slides)
  http://www.cs.ubc.ca/~tmm/vadbook

Visualization Analysis and Design.
  Munzner.
  AK Peters Visualization Series.
  CRC Press, 2014.

• papers, videos, software, talks, courses
  http://www.cs.ubc.ca/group/infovis
  http://www.cs.ubc.ca/~tmm

• this talk
  http://www.cs.ubc.ca/~tmm/talks.html#coimbra22