CSE 5544: Introduction to Data Visualization

Raghu Machiraju
machiraju.1@osu.edu
Design study methodology
Industrial Engineer

Industrial engineering is a branch of engineering which deals with the optimization of complex processes or systems. Industrial engineers work to eliminate waste of time, money, materials, man-hours, machine time, energy and other resources that do not generate value.
Nested model for vis design

- Domain situation
- Data/task abstraction
- Visual encoding/interaction idiom
- Algorithm
Nested model for vis design

Domain situation
What are people doing?
What are their goals?

Data/task abstraction

Visual encoding/interaction idiom

Algorithm
Nested model for vis design

- Domain situation
- Data/task abstraction
  - What are data/tasks to accomplish these goals?
- Visual encoding/interaction idiom
- Algorithm
Nested model for vis design

Domain situation

Data/task abstraction

Visual encoding/interaction idiom

Algorithm

How do I show/interact with the data?
Nested model for vis design

How do I make this all work?
Nested model for vis design

- **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
Nested model for vis design

- **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
Nested model for vis design

- **Threat**: Wrong problem
- **Validate**: Observe and interview target users

- **Threat**: Wrong task/data abstraction
  - **Threat**: Ineffective encoding/interaction idiom
    - **Validate**: Justify encoding/interaction design
    - **Validate**: Analyze computational complexity
    - **Implement system**
    - **Validate**: Measure system time/memory

- **Validate**: Qualitative/quantitative result image analysis
  - *Test on any users, informal usability study*
  - **Validate**: Lab study, measure human time/errors for task

- **Validate**: Test on target users, collect anecdotal evidence of utility
  - **Validate**: Field study, document human usage of deployed system

- **Validate**: Observe adoption rates
Workflow for designing a tool
Making the right tool

Questions
Data
Tasks

Vis researcher

http://halalfocus.net/wp-content/uploads/2014/05/question-mark-nothing.jpg
Making the right tool

Questions
Data
Tasks

Design study methodology

References:
- van Wijk:1999
Design study methodology

PRECONDITION
- Personal validation

CORE
- Inward-facing validation

ANALYSIS
- Outward-facing validation

Questions to ask during the precondition stage include:
- PF-1: What is the background of the visualization problem? Do we need to establish the visualization problem thread?
- PF-2: Where can we find the data? How much data is available? Does the data exist, is it enough, and can we access it?
- PF-3: Will any collaborators help us write the paper?

The goal of this stage is to identify the most promising collaborations.

In the core stage, knowledge of the current state-of-the-art is crucial for framing contributions relying on knowledge of previous work. Comparing and contrasting findings allows fast development of stable tool releases.

In the analysis stage, knowledge about visualization toolkits and algorithms assists in knowing how to properly evaluate the tool in the field; in the design stage, it informs data and task abstraction; in the problem analysis stage, it informs all later stages.

A crucial precondition for conducting an effective design study is a synergistic collaboration with domain experts. This visualization knowledge will inform all later stages.

Readers and reviewers of a paper. Munzner's nested model elaborates on how to choose appropriate methods at each stage.

The precondition stages of design study methodology are necessary: not all potential collaborations are a good match. Preparing the good from the bad and implying that careful selection is a lengthy process of separation is the goal of this stage.
Design study definition

Design study papers explore the choices made when applying infovis techniques in an application area, for example relating the visual encodings and interaction techniques to the requirements of the target task. Although a limited amount of application domain background information can be useful to provide a framing context in which to discuss the specifics of the target task, the primary focus of the case study must be the infovis content. Describing new techniques and algorithms developed to solve the target problem will strengthen a design study paper, but the requirements for novelty are less stringent than in a Technique paper.

[InfoVis03 CFP, infovis.org/infovis2003/CFP]

Munzner
Design study definition

Design study papers **explore the choices made when applying infovis techniques in an application area**, for example relating the visual encodings and interaction techniques to the requirements of the target task. Although a limited amount of application domain background information can be useful to provide a framing context in which to discuss the specifics of the target task, the primary focus of the case study must be the infovis content. Describing new techniques and algorithms developed to solve the target problem will strengthen a design study paper, but the requirements for novelty are less stringent than in a Technique paper.

[InfoVis03 CFP, infovis.org/infovis2003/CFP]  
Munzner
Design study definition

Design study papers explore the choices made when applying infovis techniques in an application area, for example relating the visual encodings and interaction techniques to the requirements of the target task. Although a limited amount of application domain background information can be useful to provide a framing context in which to discuss the specifics of the target task, the primary focus of the case study must be the infovis content. Describing new techniques and algorithms developed to solve the target problem will strengthen a design study paper, but the requirements for novelty are less stringent than in a Technique paper.

[InfoVis03 CFP, infovis.org/infovis2003/CFP] Munzner
Very Useful

http://cscheid.net/courses/spr15/cs444/lectures/week3b.html
d3’s Pipeline Model

[J. Heer, Prefuse]
d3’s Pipeline Model

[d3.csv('http://www.example.com/data.csv', /* callback */)]
[d3.tsv('http://www.example.com/data.tsv', /* callback */)]
[d3.json('http://www.example.com/data.json', /* callback */)]

Source Data → Data Tables → Visual Abstraction → Views
d3’s Pipeline Model

```javascript
d3.csv("f1.csv", function(error, data) {
    // Convert strings to numbers
    data.forEach(function(d) {
        d.year = +d.year;
        d.points = +d.points;
    });
    // Rest of vis code here
});
```

convert strings to numbers

Data > Data Tables > Visual Abstraction > Views

Source Data > Data Transformations > Visual Mappings > View Transformations

[J. Heer, Prefuse]
d3’s Pipeline Model

```javascript
d3.csv("f1.csv", function(error, data) {
  // Convert strings to numbers
  data.forEach(function(d) {
    d.year = +d.year;
    d.points = +d.points;
  });

  // Rest of vis code here
});
```

Data are arrays
d3’s Pipeline Model

[J. Heer, Prefuse]
d3’s Pipeline Model

```javascript
var circle = svg.selectAll("circle")
   .data(data);

circle.exit().remove();

circle.enter().append("circle")
   .attr("r", 5);

circle
   .attr("cx", function(d) { return d.x; })
   .attr("cy", function(d) { return d.y; });
```
Dre’s Pipeline Model

```javascript
var circle = svg.selectAll("circle")
  .data(data);

circle.exit().remove();

circle.enter().append("circle")
  .attr("r", 5);

circle
  .attr("cx", function(d) { return d.x; })
  .attr("cy", function(d) { return d.y; });
```

[M. Bostock, d3]

Data  Nodes

Enter  Update  Exit

Data elements

A B C D E F G H

Nodes

E F G H I J K L
d3’s Pipeline Model

```javascript
var circle = svg.selectAll("circle")
    .data(data);

circle.exit().remove();
circle.enter().append("circle")
    .attr("r", 5);
circle
    .attr("cx", function(d) { return d.x; })
    .attr("cy", function(d) { return d.y; });
```

[M. Bostock, d3]
elements, it is passed to functional operators as the first argument (by values, such as numbers, strings or objects. Once data is bound to agnostic processing [13]:

d3.selectAll("p").selectAll("b")

example, descendants of each selected element; we call this nested structures. The return the added nodes, thus allowing the convenient creation of

pie chart in Figure 7 is a single statement.) The

These operators wrap the W3C DOM API, setting attributes (styles), properties (html), HTML (text) content. Operator values are specified either as constants or means such as a third-party library or developer tool.

Whereas the top-level select methods query the entire document, D3 supports method chaining for brevity when applying multiple

can be applied to selected elements.

Any number of (f) Add reference lines. (g) Add reference labels. (h) Assign colors and other aesthetics with CSS.

Fig. 5. Specification of the area chart shown in Figure 4. (a) Define scale functions for position encoding. (b) Add an SVG container to the document

g.append("svg:path")

data

var g = svg.selectAll("g")

.attr("d", d3.svg.line().x(x).y(y));

var x = function(d, i) { return i * 25; }, y = function(d, i) { return 160 - d * 80; };

svg.append("svg:path")

data

var circle = svg.selectAll("circle")

data(data);

circle.exit().remove();

circle.enter().append("circle")

.attr("r", 5);

circle

.attr("cx", function(d) { return d.x; })
.attr("cy", function(d) { return d.y; });

[“enter” for data without nodes]
Elements, it is passed to functional operators as the first argument (by mat agnostic processing [13]).

### Data

Functional operators may depend on the numeric index of the current element.

Grouping plays an important role in the data join (see returning the added nodes, thus allowing the convenient creation of operators: the operator return value is the selection. (For example, the operators satisfy most needs, the

```javascript
var circle = svg.selectAll("circle")
  .data(data);

circle.exit().remove();

circle.enter().append("circle")
  .attr("r", 5);

circle
  .attr("cx", function(d) { return d.x; })
  .attr("cy", function(d) { return d.y; });
```

Whereas the top-level select methods query the entire document, D3 supports method chaining for brevity when applying multiple can be applied to selected elements.

Any number of methods for obtaining selections. These methods ac-

Selecting a selection's descendants of each selected element; we call this a selection's

These operators wrap the W3C DOM API, setting attributes (html), properties (html), and

*Fig. 5. Specification of the area chart shown in Figure 4. (a) Define scale functions for position encoding. (b) Add an SVG container to the document body and bind data. (c) Add a path element for the area. (d) Add a path element to emphasize the top line. (e) Add containers for reference values.*

Whereas the top-level select methods query the entire document, D3 supports method chaining for brevity when applying multiple can be applied to selected elements.

Any number of methods for obtaining selections. These methods ac-

Selecting a selection's descendants of each selected element; we call this a selection's

These operators wrap the W3C DOM API, setting attributes (html), properties (html), and

*Fig. 5. Specification of the area chart shown in Figure 4. (a) Define scale functions for position encoding. (b) Add an SVG container to the document body and bind data. (c) Add a path element for the area. (d) Add a path element to emphasize the top line. (e) Add containers for reference values.*
d3’s Pipeline Model

```javascript
var y = d3.scale.linear()
    .range([height, 0])
    .domain([minPoints, maxPoints]);
```
d3’s Pipeline Model

linear
log
ordinal
time
color

[J. Heer, Prefuse]
d3’s Pipeline Model

uses svg/html for visual elements

[J. Heer, Prefuse]
d3’s Pipeline Model

functions for axes, layouts, paths

```javascript
var xAxis = d3.svg.axis()
  .scale(mainX)
  .orient("bottom");
chart.append("g")
  .attr("class", "x axis")
  .attr("transform", "translate(0, height + ")")
  .call(xAxis);
```
d3’s Pipeline Model

functions for axes, layouts, paths

```javascript
var xAxis = d3.svg.axis()
    .scale(mainX)
    .orient("bottom");
chart.append("g")
    .attr("class", "x axis")
    .attr("transform", "translate(0, height + ")")
    .call(xAxis);
```

xAxis is a function!
d3’s Pipeline Model
functions for axes, layouts, paths

```javascript
var layout = d3.layout.force()
  .linkDistance(50)
  .charge(-150)
  .size([width, height])
  .nodes(graph.nodes)
  .links(graph.links)
  .start();

// This gets updated every time the layout iterates
layout.on("tick", function() {
  node.attr("cx", function(d) {return d.x;})
    .attr("cy", function(d) {return d.y;});
  link.attr("x1", function(d) {return d.source.x;})
    .attr("y1", function(d) {return d.source.y;})
    .attr("x2", function(d) {return d.target.x;})
    .attr("y2", function(d) {return d.target.y;});
});
```
d3’s Pipeline Model

functions for axes, layouts, paths

<path d="M 100 100 L 300 100 L 200 300 z" fill="red" stroke="blue" stroke-width="3" />

http://www.w3.org/TR/SVG/paths.html#PathElement
d3’s Pipeline Model

functions for axes, layouts, paths

```javascript
var lineFunction = d3.svg.line()
  .x(function(d) { return d.x; })
  .y(function(d) { return d.y; })
  .interpolate("linear");

var lineGraph = svgContainer.append("path")
  .attr("d", lineFunction(lineData))
  .attr("stroke", "blue")
  .attr("stroke-width", 2)
  .attr("fill", "none");
```

https://www.dashingd3js.com/svg-paths-and-d3js
d3’s Pipeline Model

rendering done by browser engine
hardware-accelerated transitions
vis is part of the DOM
many items can cause issues
can view all vis elements

chart.selectAll(".group")
  .data(data)
  .transition().delay(100)
  .attr("transform", function(d) {
    return "translate("+mainX(d[mainKey])+",0)";
  }).selectAll("rect")
  .attr("width", subX.rangeBand())
  .attr("x", function(d) {return subX(d[subKey]);})
  .style("fill", function(d) {return colors(d[subKey]);});
d3’s Pipeline Model

```
chart.selectAll(".group")
  .data(data)
  .transition().delay(100)
  .attr("transform", function(d) {
    return "translate(\"mainX(d\[mainKey\])+\",0\")";
  }).selectAll("rect")
  .attr("width", subX.rangeBand())
  .attr("x", function(d) {return subX(d\[subKey\]);})
  .style("fill", function(d) {return colors(d\[subKey\]);});
```

everything from here is animated
d3’s Pipeline Model

rendering done by browser engine

hardware-accelerated transitions

vis is part of the DOM

many items can cause issues

can view all vis elements

[J. Heer, Prefuse]
Design Example
Time-series data: Case A naive extruded curves: detailed comparisons impossible

[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]
Case A: Better Cluster-Calendar Solution

- derived data: cluster hierarchy
- juxtapose multiple views: calendar, superimposed 2D curves

[Cluster and Calendar based Visualization of Time Series Data. van Wijk and van Selow, Proc. InfoVis 99.]
Design study methodology
Design study methodology

What tools/techniques are available?

Read vis papers
Read vis books
Talk to vis practitioners
This course!
Design study methodology

Are these good collaborators?
Do they have interesting problems?
Do they need novel solutions?
Is there data?
Can I work with these people?
When can you do a design study?

![Diagram showing the suitability of design study methodology. The axes represent task clarity and information location, with regions indicating when a design study might be suitable or unsuitable.](image)

**DESIGN STUDY METHODOLOGY SUITABLE**

- **NOT ENOUGH DATA**
- **ALGORITHM AUTOMATION POSSIBLE**

**When can you do a design study?**

A design study might be suitable when:

- **Task Clarity** is crisp and **Information Location** is computer-based.

A design study might not be suitable when:

- **Task Clarity** is fuzzy and **Information Location** is head-based.

---

**Note:** The diagram and text are based on the information provided in the original image. The captions and labels are added for clarity and context. The diagram includes a color scheme and labels to indicate the suitability of design studies based on the axes of task clarity and information location.
Design study methodology

Who’s who?

Do people have time for a new project?
“Front-line analyst” is the domain expert
Are there false “front-line analysts”?
Do you need a “translator”?
Design study methodology

Requirements analysis
Critical reflection on requirements!
Abstraction is important for transferability
Need some domain-expert knowledge
Design study methodology

example: Cluster-Calendar, van Wijk and van Selow

Overall goal: are there temporal patterns in power consumption?
Design study methodology

example: Cluster-Calendar, van Wijk and van Selow

Data: ~50K pairs of (value, time)

Tasks

Find standard day patterns
Find out how patterns are distributed over year, week, season
Find outliers from standard daily patterns
Want overview first, details on demand
Design study methodology

Limitations of previous work:

- Predictive mathematical models: details lost
- Scale-space approaches (wavelet, fourier, fractal): hard to interpret, known scales lost
- 3D mountain diagram (x: hours, y: value, z: days)
Design study methodology

Design study methodology

- Pretty, not so useful
- Daily, weekly patterns are hard to see
Design study methodology

Data abstraction, visual encoding, interaction

What data transformations are needed?
What visual designs to use?
How to tie this together with interaction?
Don’t code!
Design study methodology

example: Cluster-Calendar, van Wijk and van Selow

Data transform: hierarchical clustering
Design study methodology

example: Cluster-Calendar, van Wijk and van Selow

Data transform: hierarchical clustering

start with M day patterns
- compute pair-wise differences, merge most similar
- now we have M-1 patterns
- repeat until we have 1 root cluster

result: binary hierarchy of clusters
Design study methodology

Example: Cluster-Calendar, van Wijk and van Selow

Data transform: hierarchical clustering

Issues:
- Distance metric to use?
- How to display the cluster?
Design study methodology

example: Cluster-Calendar, van Wijk and van Selow

dendrogram
Design study methodology

example: Cluster-Calendar, van Wijk and van Selow

Shows hierarchical structure but not time distribution!

PRECONDITION
personal validation

CORE
inward-facing validation

ANALYSIS
outward-facing validation

learn winnow cast discover design implement deploy reflect write
Several conclusions can be drawn from this image. We see that:

- Many people take a day off after a holiday (cluster 718). School vacations are visible in Spring (May 3rd to May 19th, December 25th and 26th.
- In the weekend and at holidays only very few people are present (cluster 718); On Fridays in the summer even fewer people are present.
- On Fridays and in the summer fewer people are present than on other weekdays.
- Employees present is slightly higher in the afternoons than in the mornings, with an exception on Fridays.
- Office hours are followed strictly. Most people arrive around 9:00 am and leave around 5:00 pm. Furthermore, in the morning the number of employees is lower in the summer.

Overview

<table>
<thead>
<tr>
<th>Cluster viewer</th>
<th>5/12/1997</th>
<th>31/12/1997</th>
<th>Cluster 710</th>
<th>Cluster 718</th>
<th>Cluster 719</th>
<th>Cluster 721</th>
<th>Cluster 722</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Detail

van Wijk:1999
Design study methodology

example: Cluster-Calendar, van Wijk and van Selow

clusters: data transformation to aggregate data
calendar: familiar visual representation for time
linking: interactive exploration of the data
task analysis guided choices: 3D extrusion and dendrogram don’t work
Design study methodology

Yay coding!

Need to test design hypotheses
Rapid prototyping (will probably throw away a lot of code)
Breaking bugs vs annoying bugs
Fast usability testing
Design study methodology

Hand-off to the users

Domain experts need to play with software
What works, what doesn’t?
How to evaluate?
May need to redesign/reimplement a lot
Design study methodology

Critique?
Design study methodology

Refine, reject, propose guidelines

Compare to existing design guidelines
Confirm which ones worked
Reject which ones didn’t work
Come up with new guidelines
Design study methodology

Forces clear articulation of problem, tasks, solution
Who else does my study help? - transferability!
Think carefully about what readers will care about
This takes time to do well!

Yay words!
Making the right tool

Questions
Data
Tasks

Design study methodology

van Wijk: 1999
Where are design studies?

- **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
Where are design studies?

- **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
Pitfalls
#1: Don’t skip steps!
Pitfalls

insufficient knowledge of literature
Pitfalls

collaboration with the wrong people
no real data available
insufficient time available from collaborators
no need for visualization: automate
no need for research: engineering project
Pitfalls

- is this interesting to me?
- existing tools are good enough
- not an important/recurring task
- no rapport with collaborators
Pitfalls

- not identifying front-line analyst and gatekeeper
- assuming same role distribution across projects
- mistaking tool-builders for real end users
Pitfalls

ignoring practices that currently work well
expecting just talking or fly on the wall to work
domain experts design the visualizations
too much/too little domain knowledge
too little abstraction

design consideration space too small

mistaking technique-driven and problem-driven work
non-rapid prototyping
usability: too little/too much
Pitfalls

insufficient deploy time

non-real task/data/user

*liking* a tool is not validation!