STANFORD CS448B

Visualization Software

Jeffrey Heer

ASSISTANT: **Jason Chuang**

21 January 2009

http://cs448b.stanford.edu

Last Time: Visualization (Re-)Design

Administrivia

Assignment 1 Scores Returned

- Questions? Contact Jason and I

Visualization Programming Tutorial

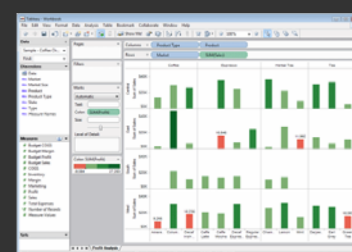
- Led by Jason Chuang
- Week of Jan 26, most likely Fri Jan 30^{th}

Assignment 2: Exploratory Data Analysis

Use visualization software (Tableau) to form & answer questions

First steps:

- Step 1: Pick a domain
- Step 2: Pose questions
- Step 3: Find Data
- Iterate



Create visualizations

- Interact with data
- Refine your questions

Make wiki notebook

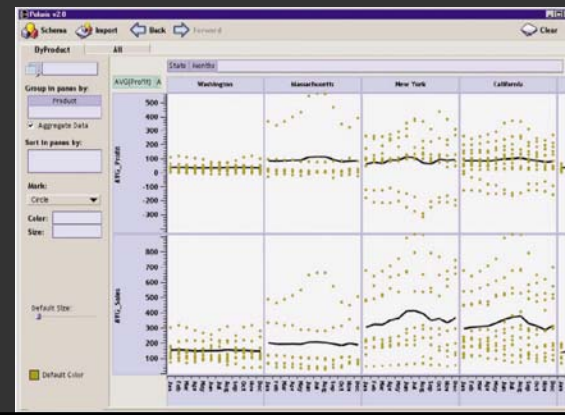
- Keep record of your analysis
- Prepare a final graphic and caption

Due by *end of day* on **Monday, January 26**

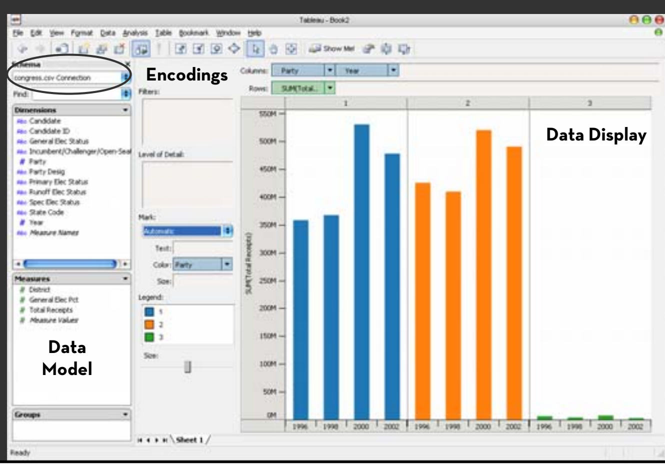
Tableau / Polaris

Polaris

Research at Stanford by Stolte, Tang, and Hanrahan.



Tableau



Polaris/Tableau Approach

Insight: can simultaneously specify both database queries and visualization

Choose data, then visualization, not vice versa

Use smart defaults for visual encodings

More recently: automate visualization design

Specifying Table Configurations

Operands are the database fields

- Each operand interpreted as a set {...}
- Quantitative and Ordinal fields treated differently

Three operators:

- **concatenation** (+)
- **cross product** (x)
- **nest** (/)

Table Algebra: Operands

Ordinal fields: interpret domain as a set that partitions table into rows and columns.

Quarter = {(Qtr1),(Qtr2),(Qtr3),(Qtr4)} →

Qtr1	Qtr2	Qtr3	Qtr4
95892	101760	105282	98225

Quantitative fields: treat domain as single element set and encode spatially as axes:

Profit = {(Profit[-410,650])} →

Concatenation (+) Operator

Ordered union of set interpretations

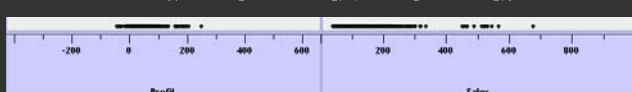
Quarter + Product Type

= {(Qtr1),(Qtr2),(Qtr3),(Qtr4)} + {(Coffee), (Espresso)}

= {(Qtr1),(Qtr2),(Qtr3),(Qtr4),(Coffee),(Espresso)}

Qtr1	Qtr2	Qtr3	Qtr4	Coffee	Espresso
48	59	57	53	151	21

Profit + Sales = {(Profit[-310,620]),(Sales[0,1000])}



Cross (x) Operator

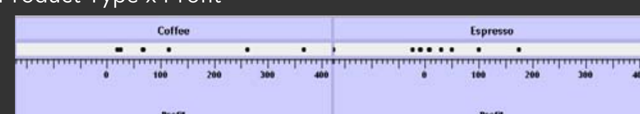
Cross-product of set interpretations

Quarter x Product Type

= {(Qtr1,Coffee), (Qtr1, Tea), (Qtr2, Coffee), (Qtr2, Tea), (Qtr3, Coffee), (Qtr3, Tea), (Qtr4, Coffee), (Qtr4,Tea)}

Qtr1		Qtr2		Qtr3		Qtr4	
Coffee	Espresso	Coffee	Espresso	Coffee	Espresso	Coffee	Espresso
131	19	160	20	178	12	134	33

Product Type x Profit =



Nest (/) Operator

Cross-product filtered by existing records

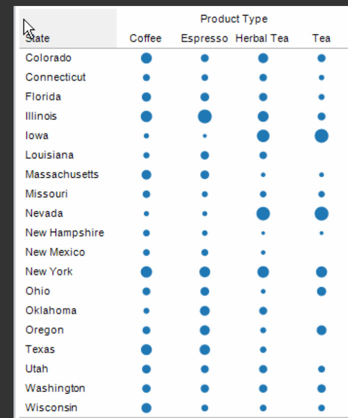
Quarter x Month

- creates twelve entries for each quarter. i.e., (Qtr1, December)

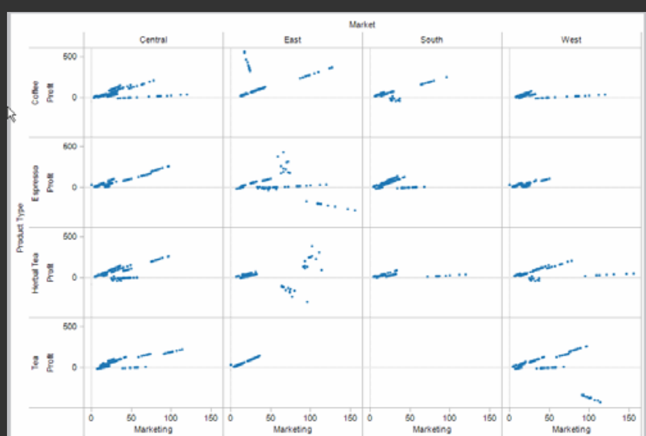
Quarter / Month

- creates three entries per quarter based on tuples in database (not semantics)

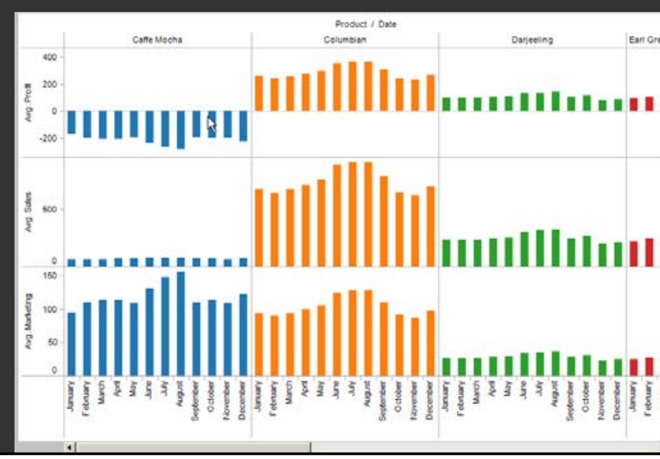
Ordinal - Ordinal



Quantitative - Quantitative



Ordinal - Quantitative



Querying the Database

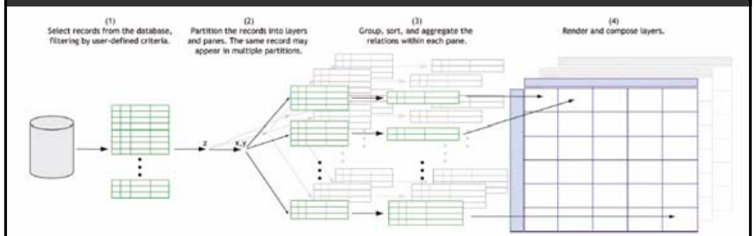


Tableau Demo

The dataset:

Federal Elections Commission Receipts

Every Congressional Candidate from 1996 to 2002

4 Election Cycles

9216 Candidacies

Data Set Schema

Year (Qi)

Candidate Code (N)

Candidate Name (N)

Incumbent / Challenger / Open-Seat (N)

Party Code (N) [1=Dem,2=Rep,3=Other]

Party Name (N)

Total Receipts (Qr)

State (N)

District (N)

This is a subset of the larger data set available from the FEC

Hypotheses?

What might we learn from this data?

- ??

Hypotheses?

What might we learn from this data?

Correlation between receipts and winners?

Do receipts increase over time?

Which states spend the most?

Which party spends the most?

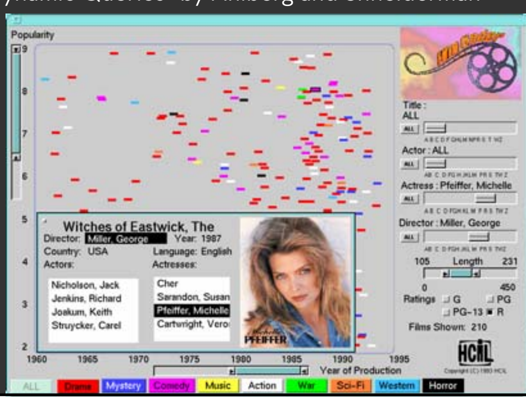
Margin of victory vs. amount spent?

Amount spent between competitors?

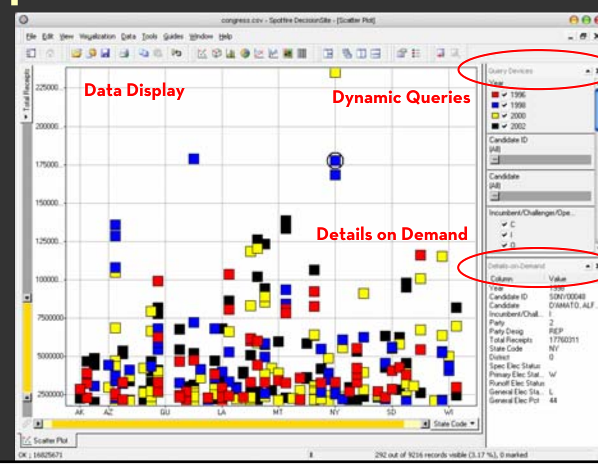
Spotfire

Spotfire

Research at UMD, College Park: “Starfield Displays” and “Dynamic Queries” by Ahlberg and Shneiderman



Spotfire



Parallel Coordinates

Parallel Coordinates [Inselberg]

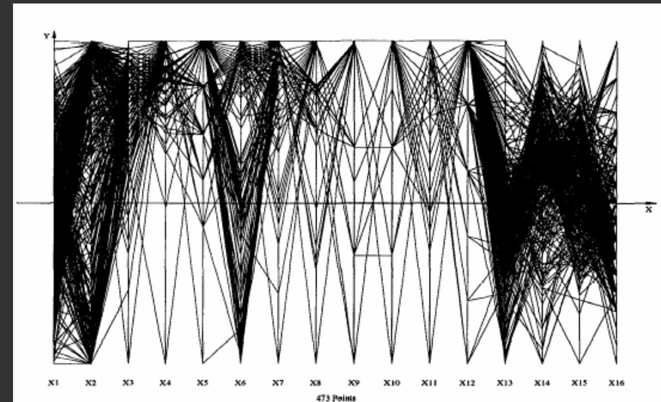


Figure 1: The full dataset consisting of 473 batches

The Multidimensional Detective

The Dataset:

- Production data for 473 batches of a VLSI chip
- 16 process parameters:

  X1: The yield: % of produced chips that are useful
  X2: The quality of the produced chips (speed)
  X3 ... X12: 10 types of defects (zero defects shown at top)
  X13 ... X16: 4 physical parameters

The Objective:
Raise the yield (X1) and maintain high quality (X2)

A. Inselberg, Multidimensional Detective, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997

Parallel Coordinates

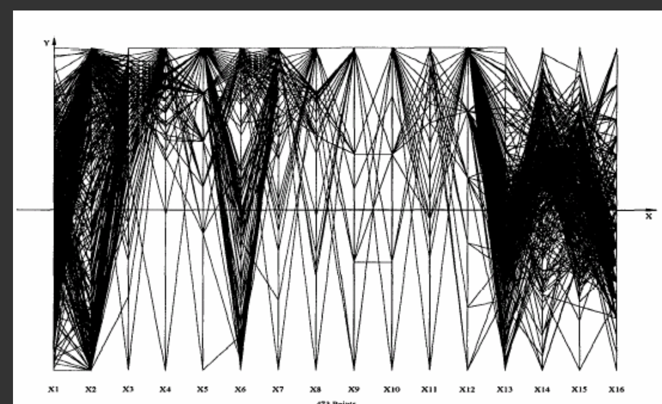


Figure 1: The full dataset consisting of 473 batches

Inselberg's Principles

1. Do not let the picture scare you
2. Understand your objectives
   - Use them to obtain visual cues
3. Carefully scrutinize the picture
4. Test your assumptions, especially the "I am really sure of's"
5. You can't be unlucky all the time!

Each line represents a tuple (e.g., VLSI batch)
Filtered below for high values of X1 and X2

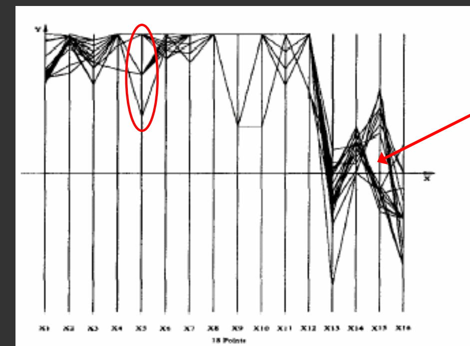
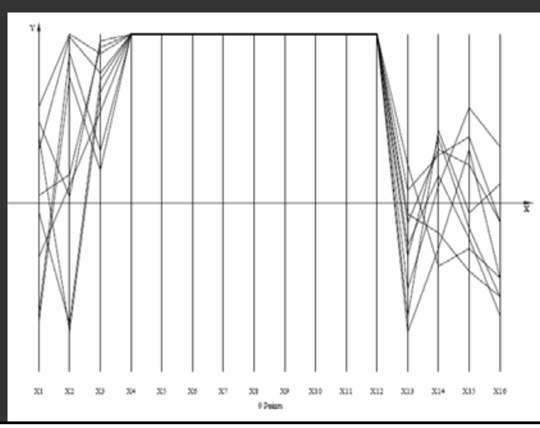


Figure 2: The batches high in Yield, $X1$, and Quality, $X2$.

Look for batches with *nearly* zero defects (9/10)
Most of these have low yields → defects OK.



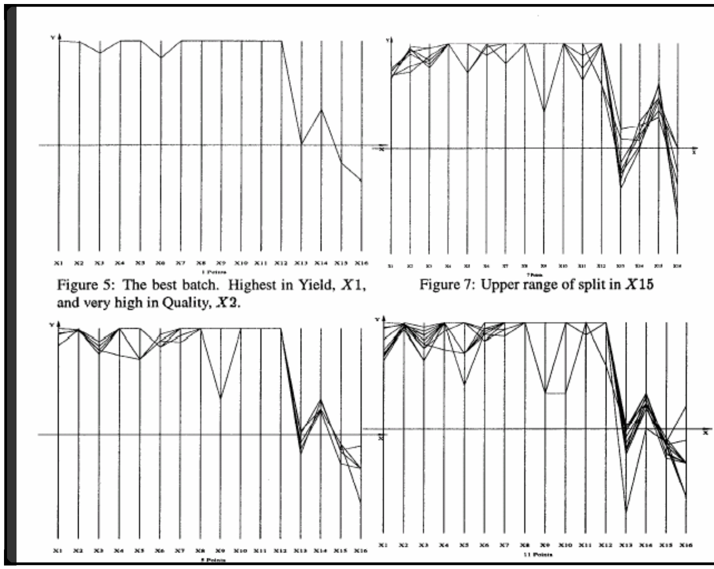


Figure 5: The best batch. Highest in Yield, $X1$, and very high in Quality, $X2$.

Figure 7: Upper range of split in $X15$

Notice that X6 behaves differently.

Allow 2 defects, including X6 → best batches

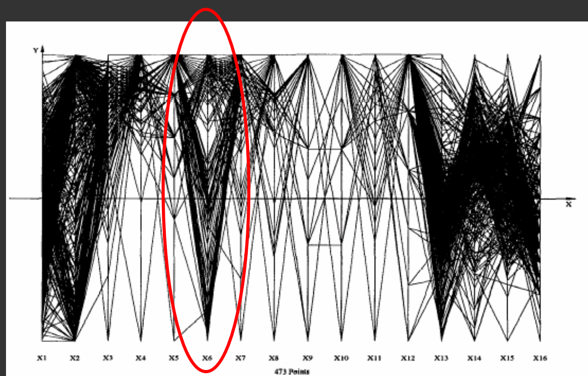


Figure 1: The full dataset consisting of 473 batches

Parallel Coordinates

Free implementation: Parvis by Ledermen

- http://home.subnet.at/flo/mv/parvis/