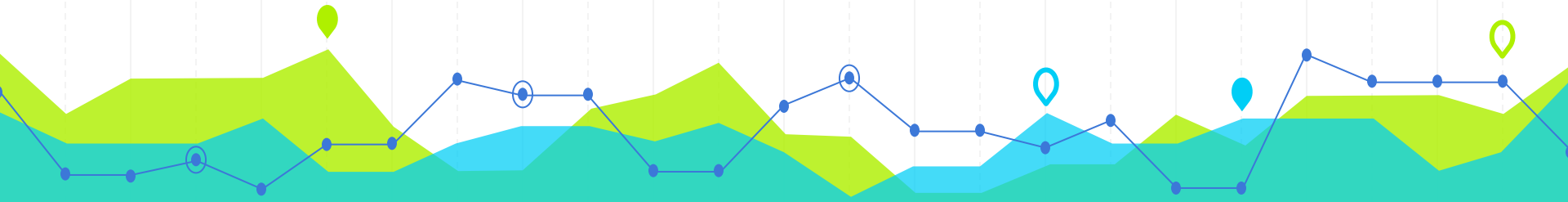
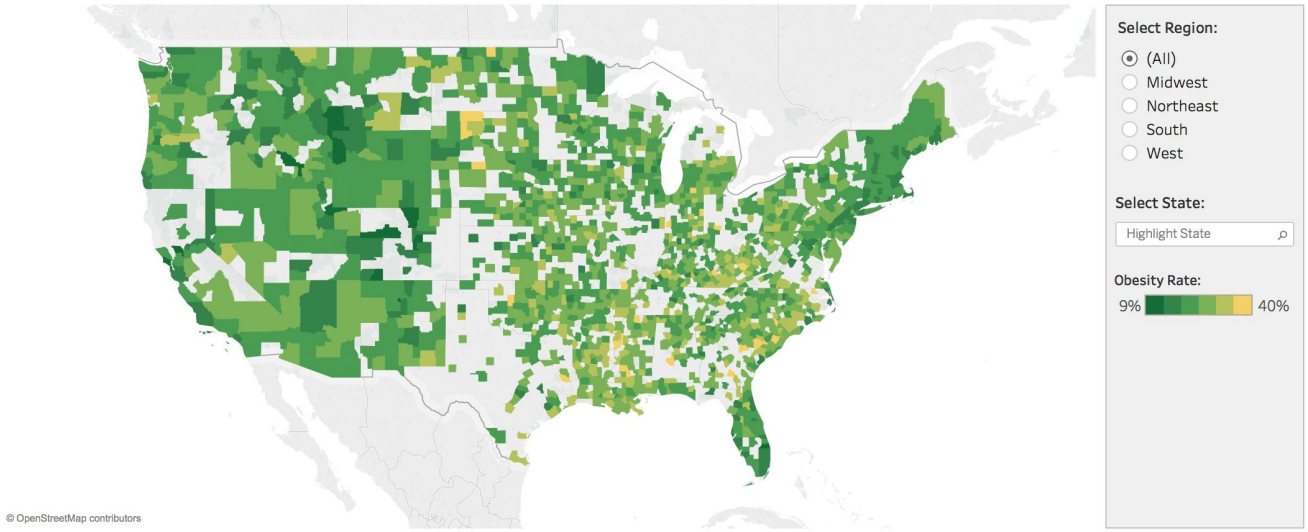


Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases

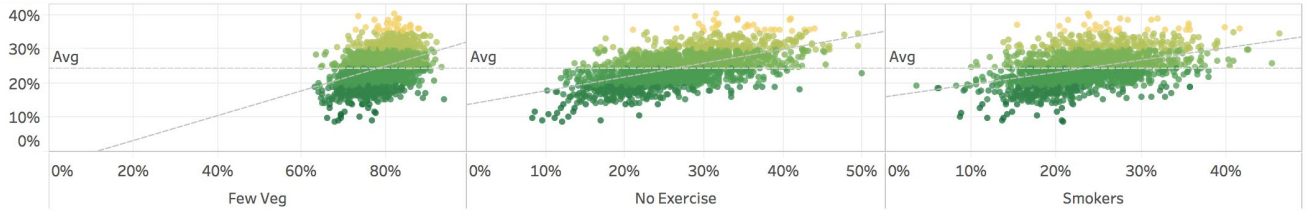
Chi-Hsien (Eric) Yen, Oct. 23

Is Your County Obese?

Select your county to see how it compares with other counties in the country



Obesity's correlation with diet, exercise, and smoking (% of pop.) - All



Timeline

- 2002: Polaris paper published (Chris Stolte, Diane Tang, and Pat Hanrahan)
- 2003: Tableau Founded (Chris Stolte, Pat Hanrahan, and Christian Chabot)
- 2013: IPO launched ($250M)
- 2016: Full year revenue is $826.9M
- Now: A worldwide company with 80+ offices and 3000+ employees

Interesting to see a research project turned into a global company in 15 years!

Motivation

- Allow analysts to **rapidly specify and change visualizations** to explore large multidimensional databases
  - Easier to find trends, patterns, outliers, etc. in visualizations
  - Hypothesis testing and experimenting
  - Communicate insights and knowledge

Main Contributions

- **A unified way to specify visualizations**
  - Table Algebra
  - Type of Graphics
  - Visual Properties
- **An intuitive interface with useful features**

Visualization Specification - Table Algebra

- Specify the *x, y, z* axes of the table using the fields
  - **X: columns**
  - **Y: rows**
  - **Z: layers**
- Actual graph depends on ***types*** of fields

Visualization Specification - Table Algebra

- Types of data:
  - **Ratio (quantitative)**
  - **Interval**
  - **Ordinal**
  - **Nominal**

Visualization Specification - Table Algebra

- Types of data:
  - **Ratio (quantitative)** **<-> measure**
  - Interval -> **quantitative**
  - **Ordinal** **<-> dimension**
  - Nominal -> **ordinal**

Q: Is there any issue?

Visualization Specification - Table Algebra

- Operands: fields
  - **Quarter (O), Product (O), Profit (Q), Sales (Q)**
- Operators:
  - Concatenation (+)
  - Cross (x)
  - Nest (/)

Visualization Specification - Table Algebra

O + O = Quarter + Product = {Qtr1, Qtr2, Qtr3, Qtr4, Coffee, Espresso, Herbal Tea, Tea}:

Qtr1	Qtr2	Qtr3	Qtr4	Coffee	Espresso	Herbal Tea	Tea

O x O = Quarter x Product = {(Qtr1,Coffee), (Qtr1,Espresso), (Qtr1,Herbal Tea), (Qtr1, Tea), (Qtr2, Coffee) ... (Qtr4, Tea)}:

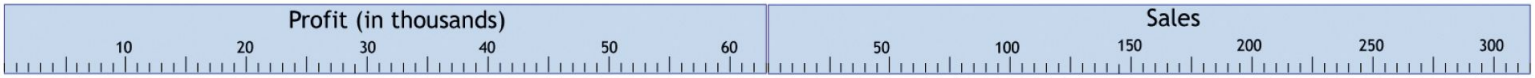
Qtr1				Qtr2				Qtr3				Qtr4			
Coffee	Espresso	Herbal Tea	Tea	Coffee	Espresso	Herbal Tea	Tea	Coffee	Espresso	Herbal Tea	Tea	Coffee	Espresso	Herbal Tea	Tea

O/O = Quarter / Month = {(Qtr1,Jan), (Qtr1,Feb), (Qtr1,Mar), (Qtr2, Apr), (Qtr2, May) ... (Qtr4, Dec)}:

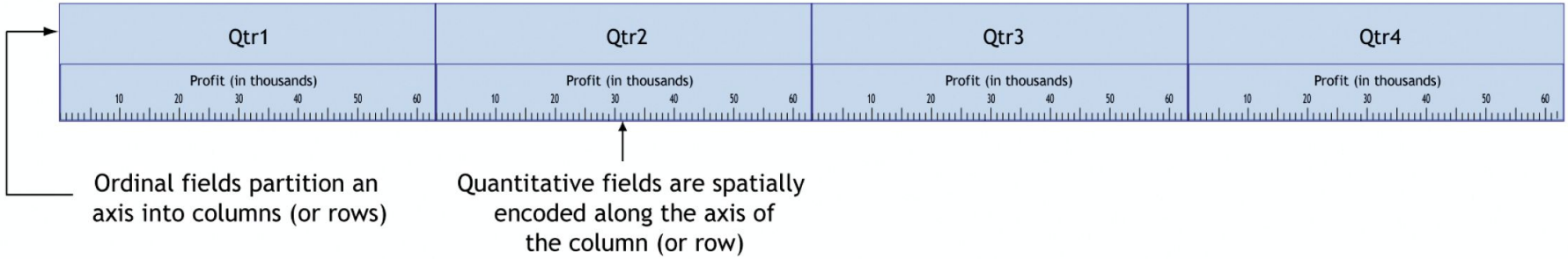
Qtr1			Qtr2			Qtr3			Qtr4		
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

Visualization Specification - Table Algebra

Q + Q = Profit + Sales = {Profit, Sales}:



O x Q = Quarter x Profit = {(Qtr1,Profit), (Qtr2, Profit), (Qtr3, Profit), (Qtr4, Profit)}:



Visualization Specification - Graphics

- Depends on types of fields
  - Ordinal - Ordinal
  - Ordinal - Quantitative
  - Quantitative - Quantitative

Follow-up Research Paper: automatic presentation*

Mackinlay, J. D., Hanrahan, P., & Stolte, C. (2007). ***Show me: Automatic presentation for visual analysis***. Visualization and Computer Graphics, IEEE Transactions on, 13(6), 1137-114

Visualization Specification - Graphics

- Automatic rules

Table 1: Automatic marks rules

Pane Type		Mark Type	View Type
Field	Field		
C	C	Text	Cross-tab
Qd	C	Bar	Bar view
Qd	Cdate	Line	Line view
Qd	Qd	Shape	Scatter plot
Qi	C	Gantt	Gantt view
Qi	Qd	Line	Line view
Qi	Qi	Shape	Scatter plot

Mackinlay, J. D., Hanrahan, P., & Stolte, C. (2007). ***Show me: Automatic presentation for visual analysis***. Visualization and Computer Graphics, IEEE Transactions on, 13(6), 1137-114

Visualization Specification - Graphics

- Automatic rules

Text Tables: at least 1 field, rank 1

Text tables have the lowest rank because their primary utility is to look up specific values. The higher ranked commands present views that encode data graphically, which support other tasks such as comparison. Although text tables have a low rank, their condition is easily met. The text table command can handle a large number of fields and will always be available as a default for **Show Me Alternatives.** Heat maps are a related command that is not ranked.

Aligned Bars: at least 1 Q, rank 2

Bars are effective for comparing values because the human visual system is good at comparing bar lengths, particularly when they are aligned. Aligned Bars are a common default when the input includes a quantitative field unless the input also includes a date field or two quantitative fields. However, aligned bars can involve a lot of scrolling when multiple categorical fields are shown. The next command handles this case.

Stacked Bars: at least 2 C, at least 1 Q, rank 3 with at least 3 C

Stacked bars require less scrolling than aligned bars when there are multiple categorical fields in the view. There are two additional bar commands that are not ranked.

Discrete Lines: at least 1 Cdate, at least 1 Q, rank 4

A line view is a better default than a bar view when the input includes a date field because it is more effective for showing trends. This command treats the date field discretely. There is an unranked line command that treats the date

Mackinlay, J. D., Hanrahan, P., & Stolte, C. (2007). ***Show me: Automatic presentation for visual analysis***. Visualization and Computer Graphics, IEEE Transactions on, 13(6), 1137-114

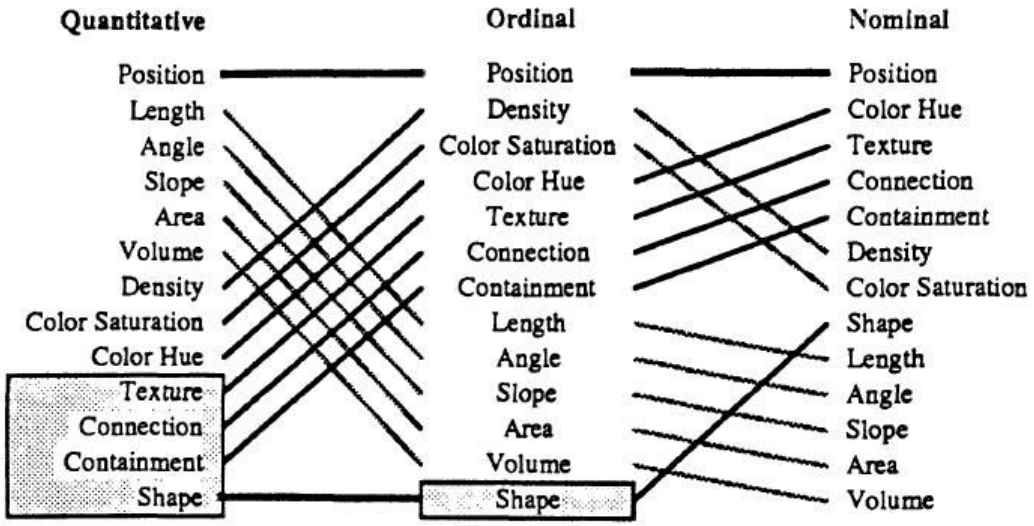
Visualization Specification - Visual Properties

- Encode additional fields into visual properties, such as shape, size, orientation, color, etc.

property	marks	ordinal/nominal mapping	quantitative mapping
shape	glyph	○ □ + △ S U	
size	rectangle, circle, glyph, text		
orientation	rectangle, line, text		
color	rectangle, circle, line, glyph, y-bar, x-bar, text, gantt bar		min max

Visualization Specification - Visual Properties

Related Works: visual variable accuracy*



Mackinlay, J.D.: Automating the Design of Graphical Presentations of Relational Information, Computer Science Department, Stanford University, 1986

Interface & Features

- Deriving Additional Fields
  - Aggregation
  - Counting
  - Discrete Partitioning
  - Grouping
  - Threshold aggregation
- Sorting and Filtering
- Brushing and Tooltips
- Undo and Redo

Generating Database Queries

- Step1: Select Data

```
SELECT * WHERE {filters}
```

Generating Database Queries

- Step2: Partitioning data into panes

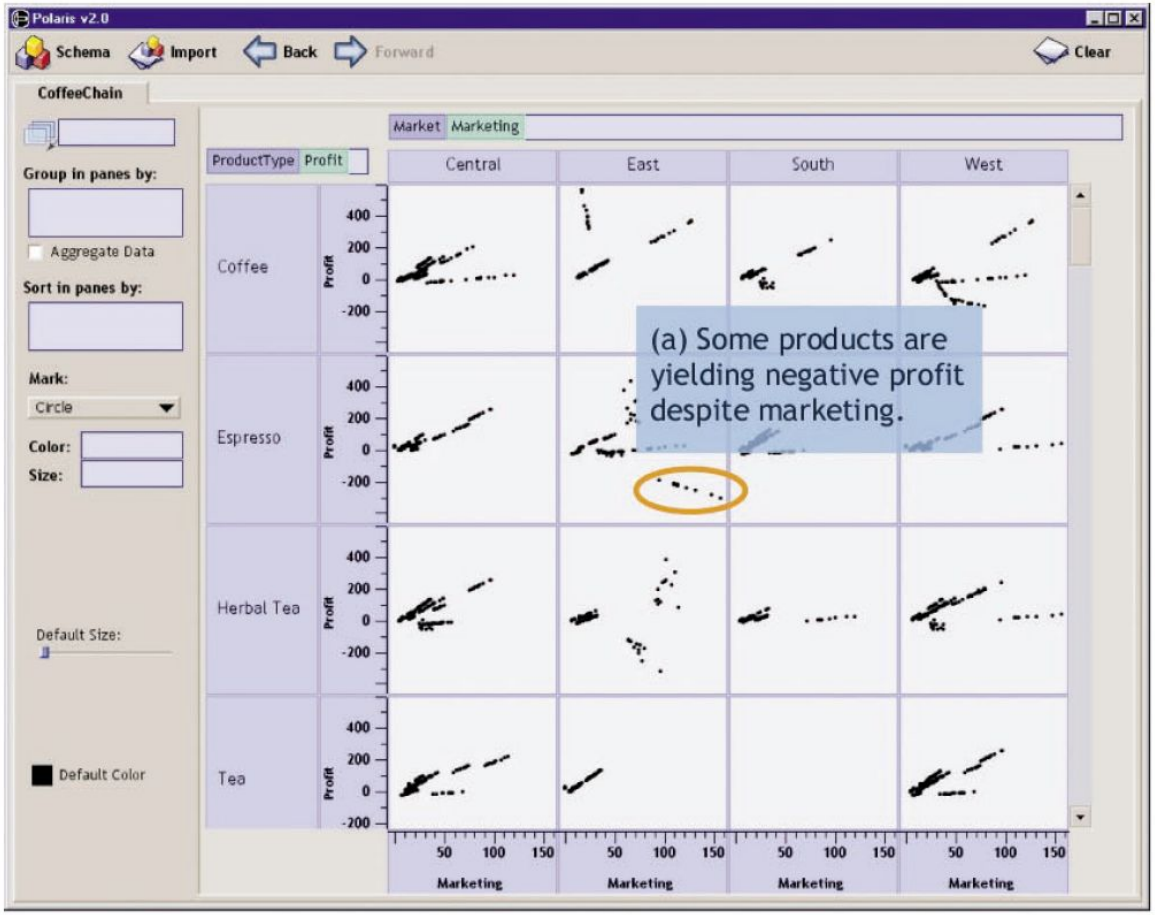
```
For i, j, k:
    SELECT * WHERE {Row(i) and Column(j) and Layer(k)}
```

Generating Database Queries

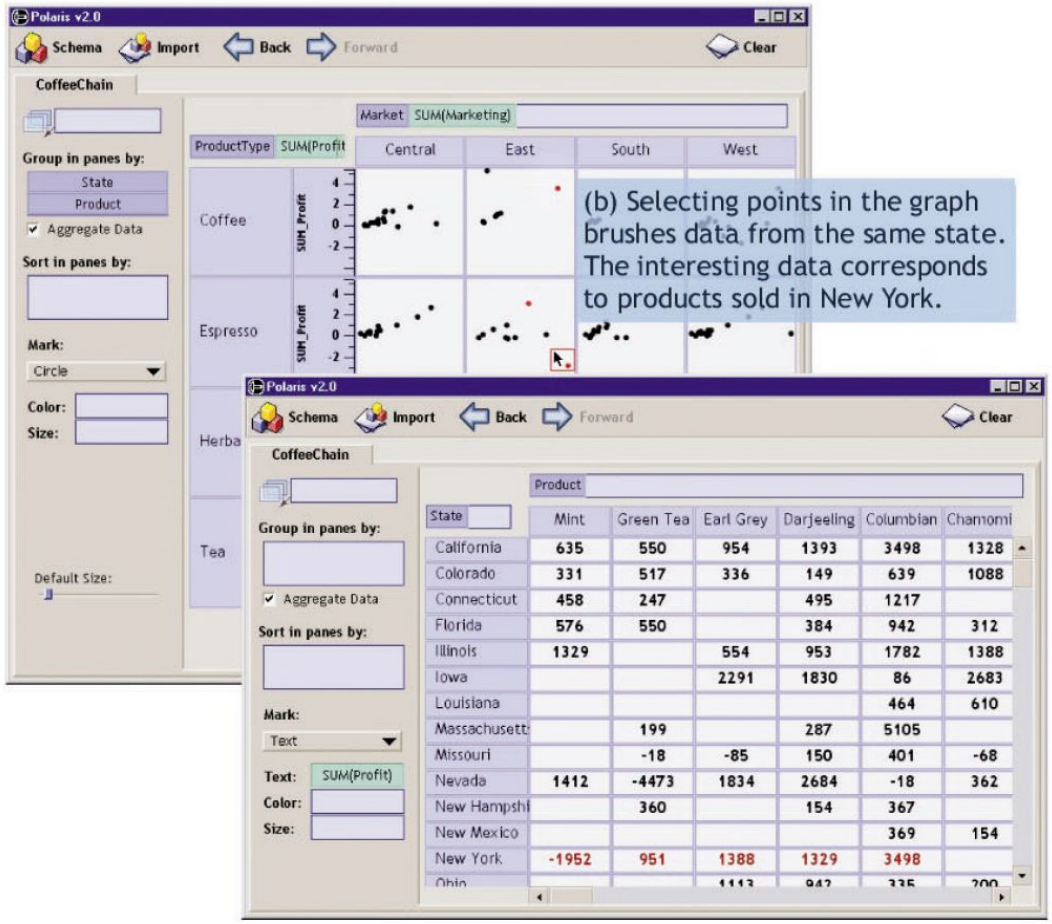
- Step3: Data transformation

```
SELECT {avg(), sum(), count(), etc.}
GROUP BY {groups}
HAVING {filters}
ORDER BY {drawing-order}
```

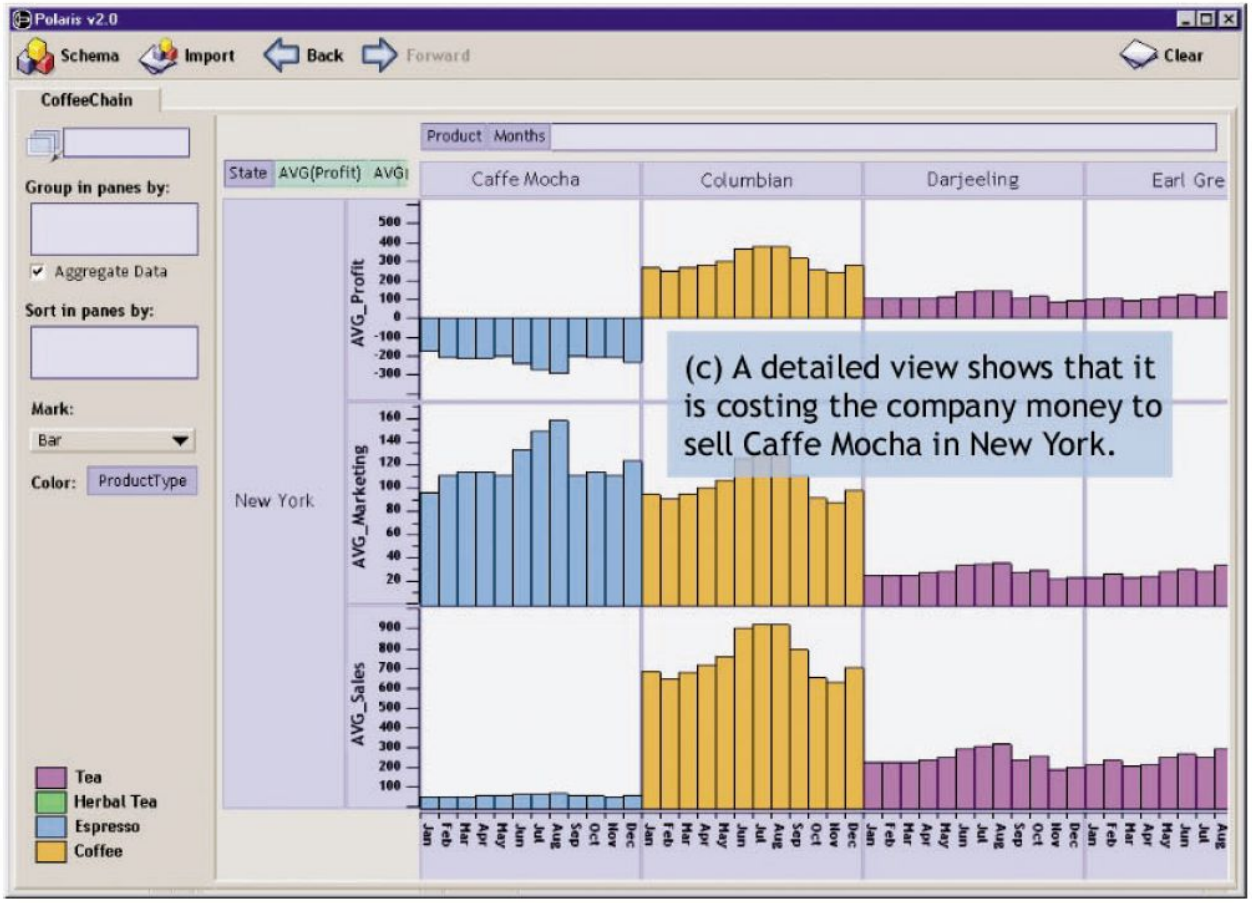
Evaluation - Financial Analysis of a Coffee Chain



Evaluation - Financial Analysis of a Coffee Chain



Evaluation - Financial Analysis of a Coffee Chain



Limitations?

- No actual performance evaluation or usability tests
- Limited to 2D visualizations
- Does not suggest graph specification (beyond two fields)
- Other ideas?