

INTRO TO TABLEAU



UNIVERSITY OF
CALGARY

WHILE WE'RE GETTING STARTED

Make sure **Tableau Desktop 2019.4**
is **installed and working**.

Download the dataset we'll be working
with from the course website.



<https://tinyurl.com/DATA605-W2020>

TODAY

Jan 9 — Visualization Tools.

Download and install Tableau before Class

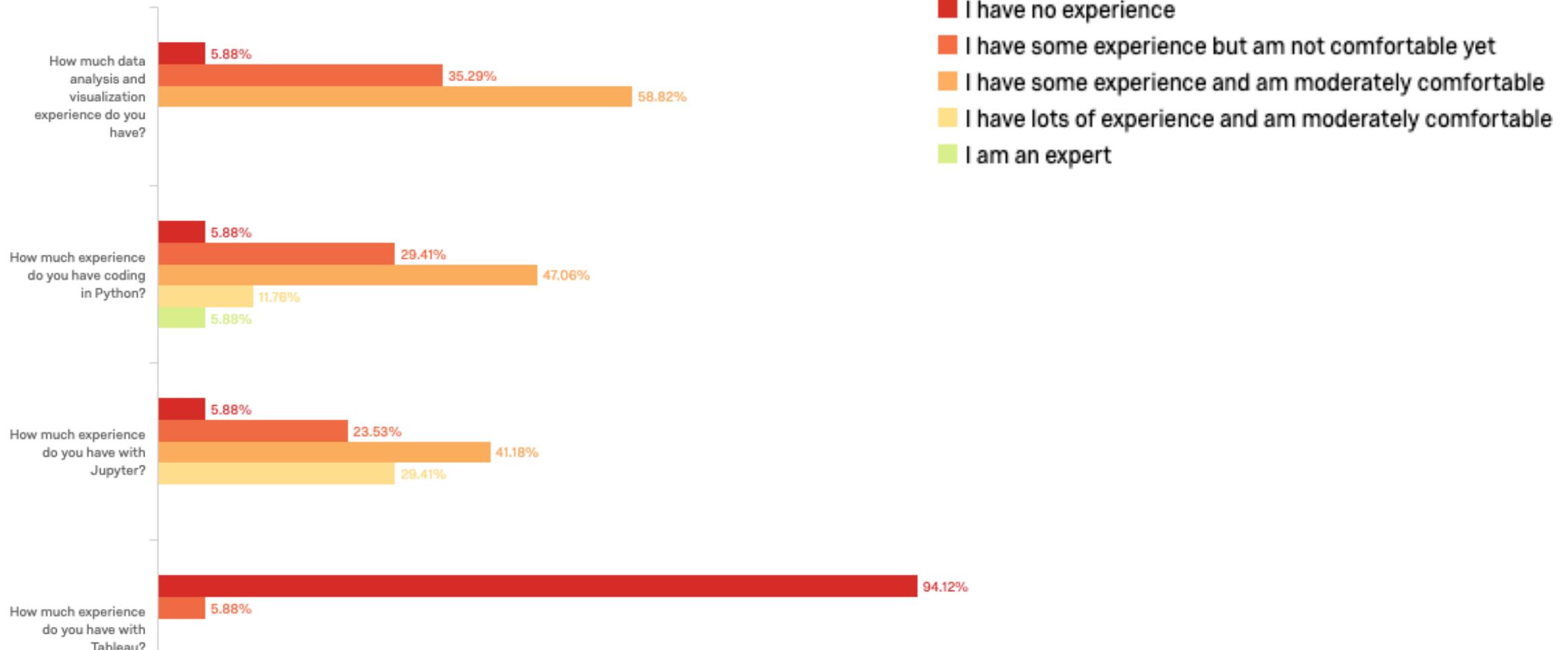
5:00pm - Tableau Tutorial

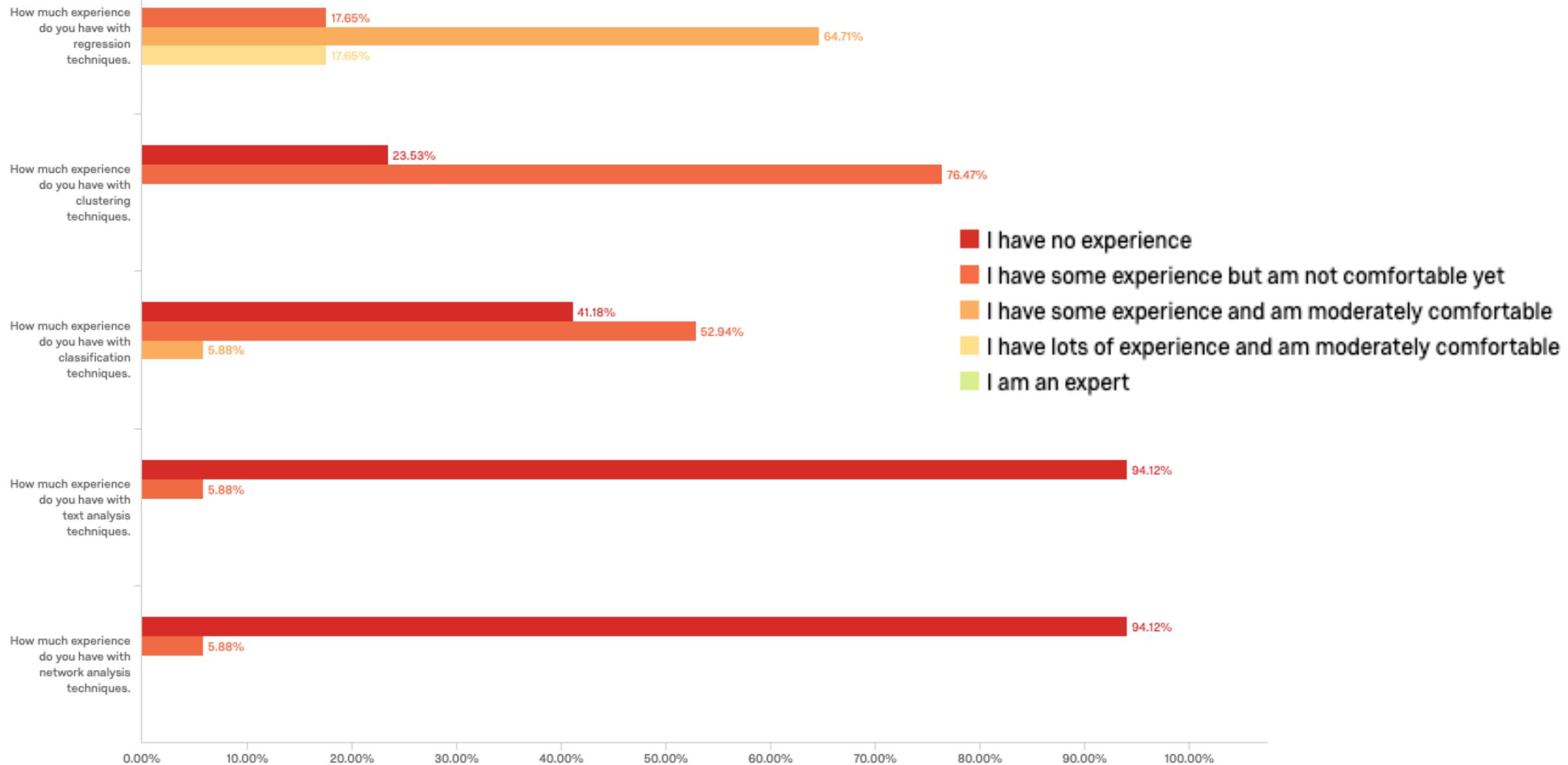
5:40pm - Intro to Visualization Tools

6:15pm - Break

6:30pm -  Datathon 1 (Movies) + team formation

SURVEY RESULTS





WHAT IS DATA VISUALIZATION?

**“THE USE OF COMPUTER-
GENERATED, INTERACTIVE,
VISUAL REPRESENTATIONS
OF DATA TO AMPLIFY
COGNITION.”**

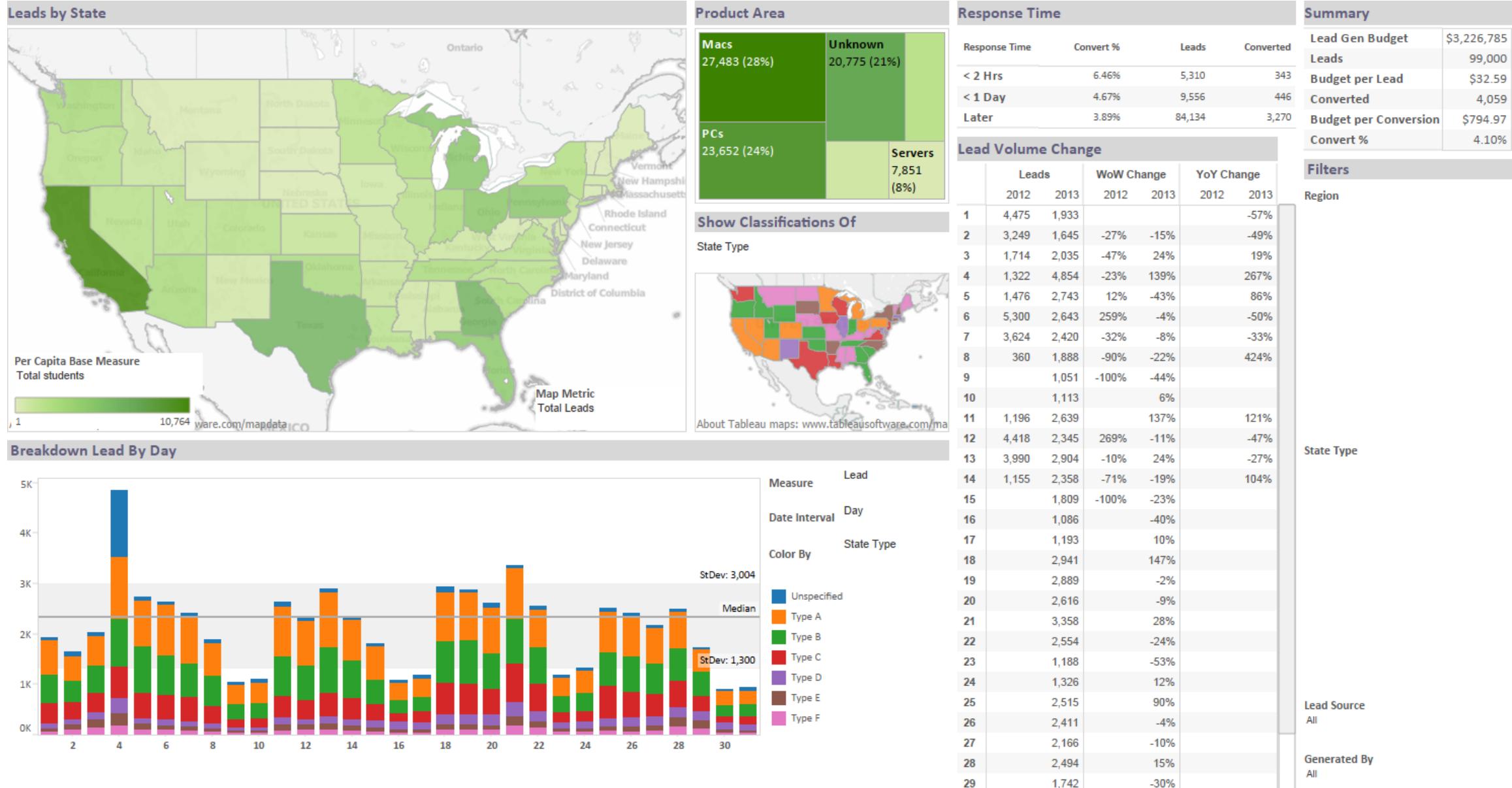


CARD | MACKINLAY | SHNEIDERMAN

**INFORMATION VISUALIZATION:
USING VISION TO THINK (1999)**

ENCODING
DATA → **MARKS**

**TO EFFICIENTLY LEVERAGE
HUMAN PERCEPTION**





S. S. STEVENS

NOMINAL, ORDINAL, AND QUANTITATIVE

N - NOMINAL (or “CATEGORICAL”) – LABELS

- Fruits: Apples, oranges, Ⓜ

O - ORDINAL - HAS A MEANINGFUL ORDERING

- Quality of meat: Grade A, AA, AAA
- Can be ordered and counted, but not measured

Q - QUANTITATIVE

- INTERVAL
 - No clear zero (or arbitrary)
 - E.g. dates, latitude, longitude
 - Usually compare differences (“intervals”)
- RATIO
 - Meaningful origin (often zero)
 - E.g. Physical measurement (Length, Mass, Ⓜ)
 - Counts and amounts

NOMINAL, ORDINAL, AND QUANTITATIVE

N - Nominal (labels)

- Operations: $=$, \neq

O - Ordered

- Operations: $=$, \neq , $<$, $>$

Q - Interval (Location of zero arbitrary)

- Operations: $=$, \neq , $<$, $>$, $-$
- Can measure distances or spans

Q - Ratio (zero fixed)

- Operations: $=$, \neq , $<$, $>$, $-$, \div
- Can measure ratios or proportions

DIMENSIONS AND MEASURES

DIMENSIONS

Discrete variables describing data

Categories, Names, Dates, etc. (independent vars)

MEASURES

Data values that can be aggregated

Numbers to be analyzed (dependent vars)

- Aggregate as sum, count, average, std. deviation, etc.

Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Blood Pressure Study (4 treatments, 6 months)

EXAMPLE: CENSUS DATA

Year: 1901- 2011 (every 5 years)

Age: 0 - 90+

Sex: Male, Female

Marital Status: Single, Married, Divorced, □

People: # of people in group

NOMINAL, ORDINAL, OR QUANTITATIVE?

Year

Q-Interval (O)

Age

Q-Ratio (O)

Sex

N

Marital Status

N

People

Q-Ratio

DIMENSION OR MEASURE?

Year

Age

Sex

Marital Status

People

Dimension

Depends!

Dimension

Dimension

Measure

WHY IS THIS IMPORTANT?

Nominal, Ordinal, and Quantitative data are best encoded in different ways.

Dimensions and Measures are important concepts in many analysis tools.

INTRO TO TABLEAU

POLARIS



Database Schema:

The user drags fields from the database schema to shelves to define the visual specification.

Layer Shelf:

The fields placed here determine how records are partitioned into layers.

Grouping and Sorting Shelves:

The fields placed here determine how records are grouped and sorted within the table panes.

Mark Pulldown:

Relations in each pane are mapped to marks of the selected type.

Retinal Property Shelves:

The fields placed here determine how data is encoded in the retinal properties of the marks.

Legends:

Legends enable the user to see and modify the mappings from data to retinal properties.

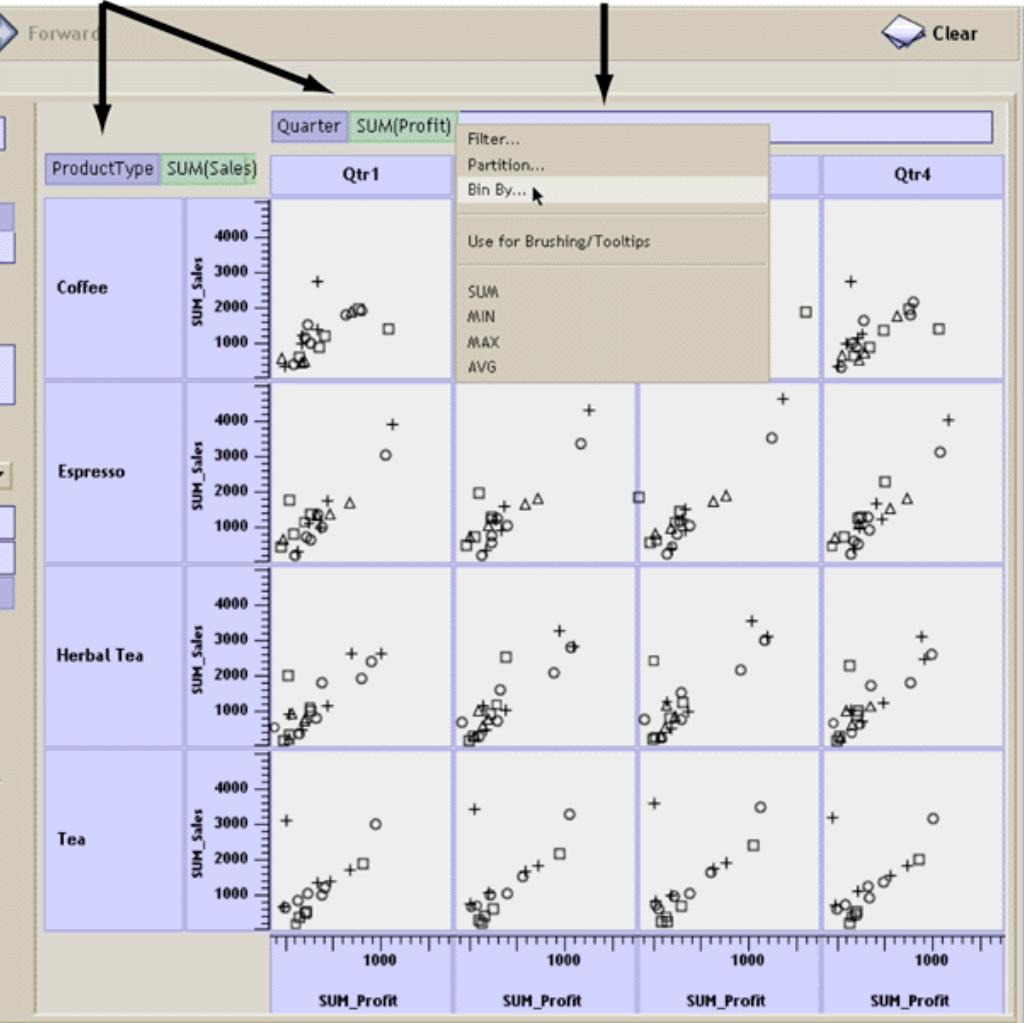
Layer Tabs:

Each layer has its own tab; different transformations and mappings can be specified for each layer.

The screenshot shows the Polaris schema editor interface. On the left, a vertical list of fields includes Quarter, Months, Market, State, MrktSize, ProductType, Product, Decaf, Profit, Margin, Sales, COGS, TotalExpenses, Marketing, Payroll, Misc, Inventory, Opening, Additions, Ending, MarginRate, ProfitRatio, BudgetProfit, BudgetMargin, BudgetSales, BudgetCOGS, BudgetPayroll, BudgetAdditions, and layer. Arrows point from specific sections of the text descriptions to the corresponding fields in this list. The right side of the interface shows a 'Layer Tabs' panel for the 'CoffeeChain' layer, containing tabs for 'Schema', 'Import', 'Back', 'Forward', and 'Clear'. Below this are sections for 'Group in panes by:' (set to 'State'), 'Sort in panes by:', 'Mark:' (set to 'Glyph'), 'Color:', 'Size:', 'Shape:' (set to 'Market'), and a 'Default Size:' dropdown. At the bottom is a 'Default Color' swatch.

Axis Shelves:

The fields placed here determine the structure of the table and the types of graphs in each table pane.



Context Menu:

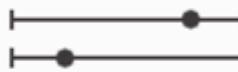
The context menu provides access to the data transformation and interaction capabilities of Polaris such as sorting, filtering, and aggregation.

COMPOSING VISUALIZATIONS BY MAPPING DATA TO SPACE/MARKS

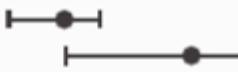
property	marks	ordinal/nominal mapping	quantitative mapping
shape	glyph	○ □ + △ S U	
size	rectangle, circle, glyph, text	• • • •	• • • • • • • • •
orientation	rectangle, line, text	— — / \ —	— — — / / / / / / / /
color	rectangle, circle, line, glyph, y-bar, x-bar, text, gantt bar	orange blue green purple yellow pink cyan brown black grey ...	min max

④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



⑤ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



Shape

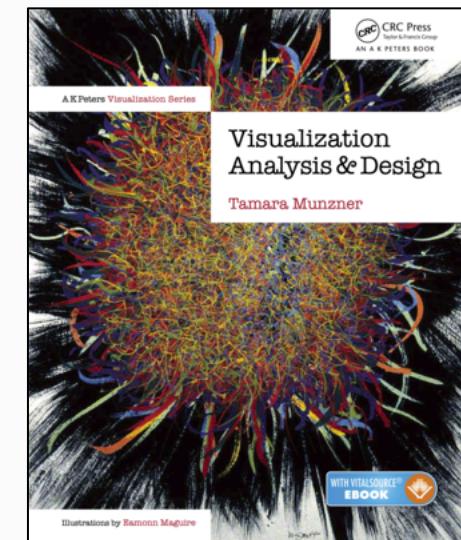


↑ Most

Effectiveness

▼ Least

Same



TAMARA MUNZNER

POLARIS/TABLEAU APPROACH

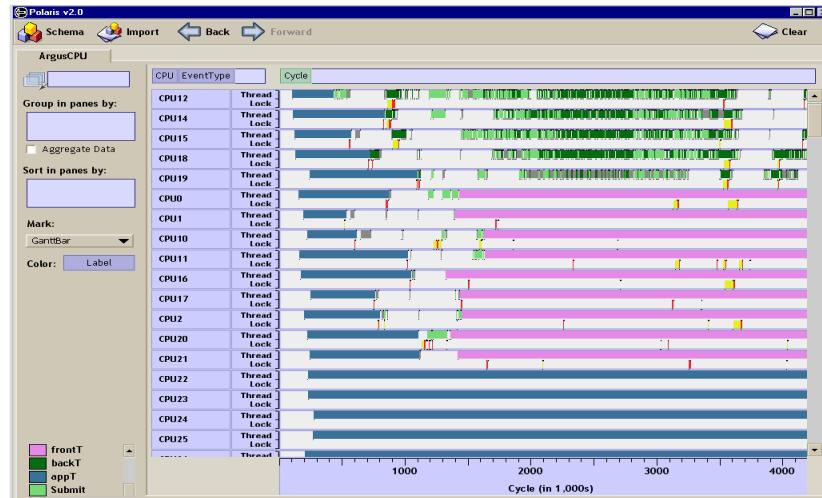
INSIGHT: Can simultaneously specify both database queries and visualization.

Choose data, then visualization, **not vice-versa**

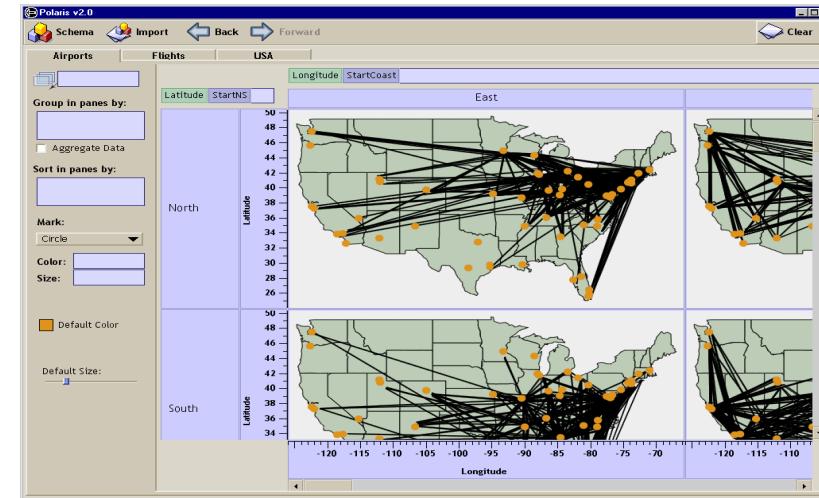
Interactively specify visualizations.

Use **smart defaults** for visual encodings.

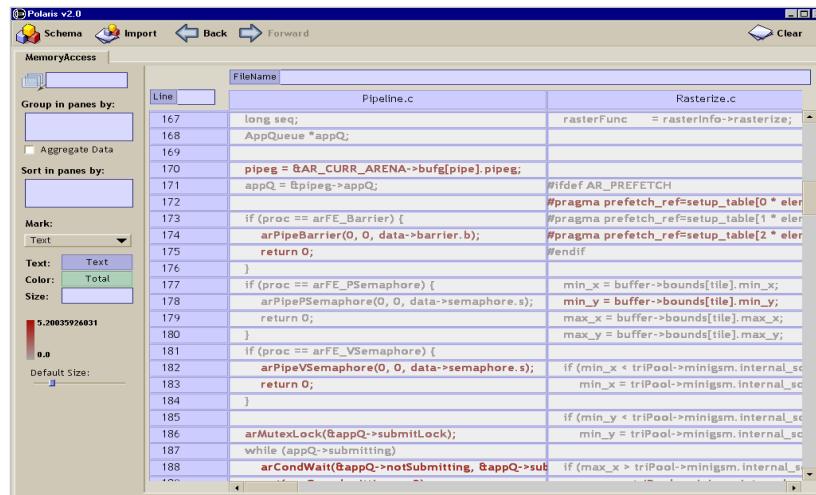
POLARIS



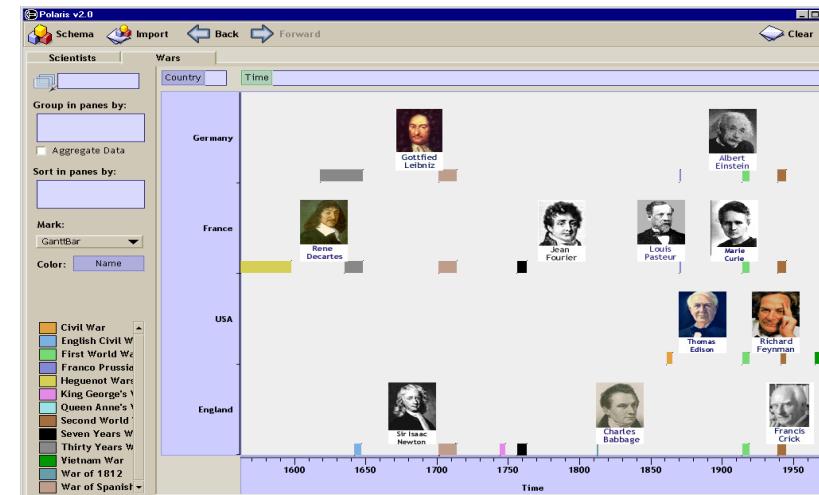
Gantt charts of events for a parallel graphics application on a 32-processor SGI machine.



Flights between major airports in the USA



Source code colored by cache misses for a parallel graphics application.



Major wars and the births of well known scientists as a timeline.

DIMENSIONS AND MEASURES

DIMENSIONS

Discrete variables describing data

Categories, Names, Dates, etc. (independent vars)

MEASURES

Data values that can be aggregated

Numbers to be analyzed (dependent vars)

- Aggregate as sum, count, average, std. deviation, etc.

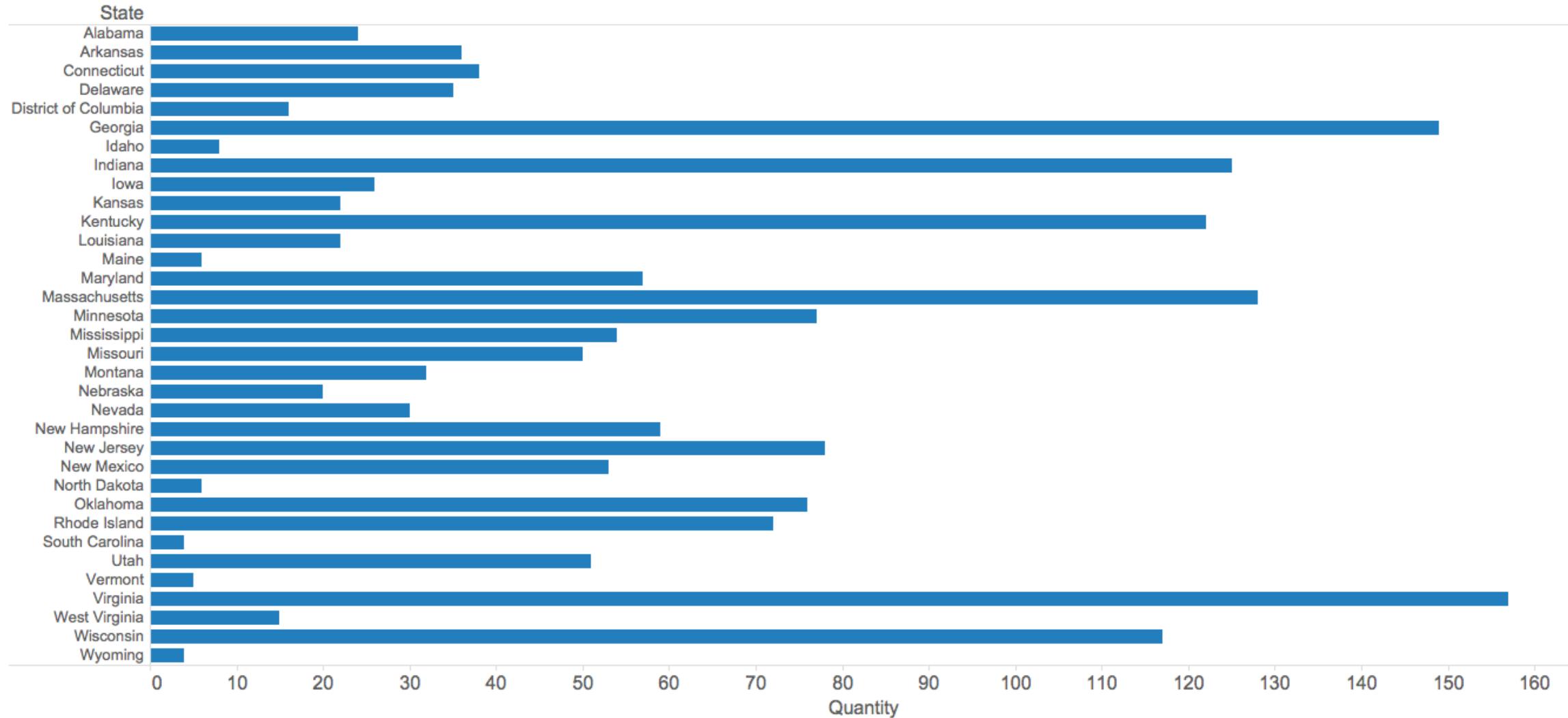
Month	Treatment	Pressure
March	Control	165
March	Placebo	163
March	300 mg	166
March	450 mg	168
April	Control	162
April	Placebo	159
April	300 mg	161
April	450 mg	163
May	Control	164

Blood Pressure Study (4 treatments, 6 months)

DIMENSION - DIMENSION

State	Ship Mode			Second Class	Standard Class
	First Class	Same Day			
Alabama				4	4
Arizona	10	8		2	
Arkansas		2			7
California		27			
Colorado		2			
Connecticut	3				9
Delaware	4			1	6
District of Columbia				3	
Florida		1		6	54
Georgia	3	13		3	21
Idaho					3
Illinois		11		29	
Indiana	4	1		1	29
Iowa				3	3

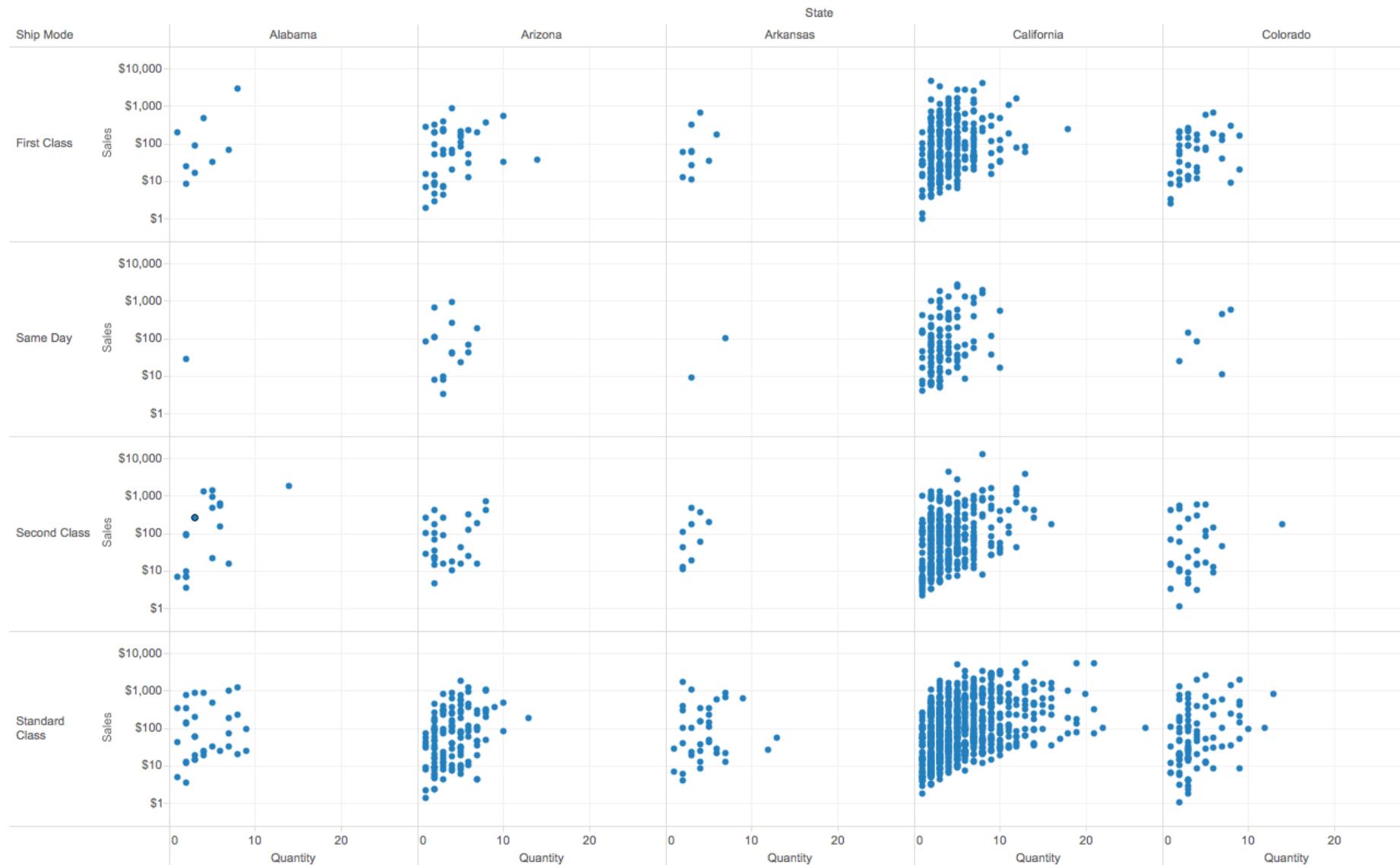
DIMENSION - MEASURE



MEASURE - MEASURE



NESTING



**TABLEAU
DEMO**

DATATHON

Make sure Tableau is installed and working.

Download the dataset from the class website.

movielens

Non-commercial, personalized movie recommendations.

From the GroupLens
researchers at the
University of Minnesota

<http://grouplens.org/datasets/movielens/>

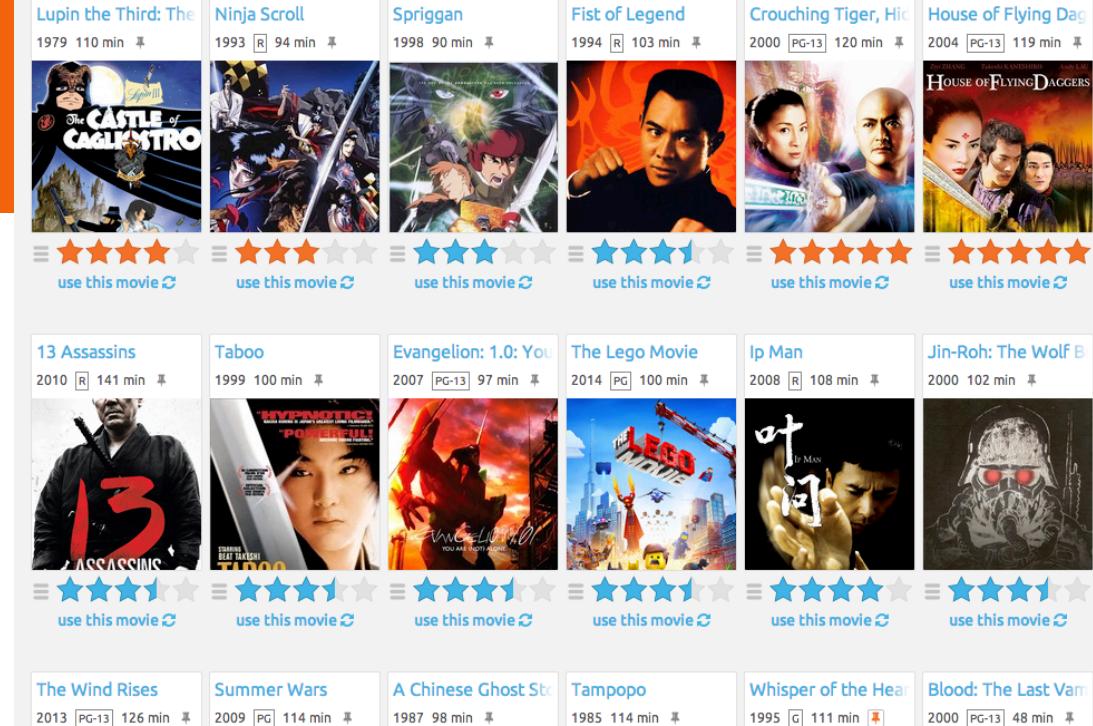
MovieLens Latest Datasets

These datasets will change over time, and are not appropriate for reporting research results. We will keep the download links stable for automated downloads. We will not archive or make available previously released versions.

Small: 100,000 ratings and 6,100 tag applications applied to 10,000 movies by 700 users. Last updated 1/2016.

Movies like Castle in the Sky

but more ninja 



AN INTRO TO TABLEAU

IMPORTING DATA

BUILDING BASIC CHARTS

CALCULATED FIELDS

FILTERS

DATATHON ONE

Before class on Tuesday, add slides to the Google Slides deck for this Assignment (Details on Class Website).

Each team should prepare 5 slides, each with:

- A clearly stated **question** about the data.
- A clear **visualization** that answers the question.

Be ready to present your 5 questions and findings (in <10 minutes) as a team at the start of class on Tuesday.

DATATHONS AND MARKING

- You should be able to do the majority (but often not all) of the work in class/tutorial.
- You should demonstrate that you're thinking critically and using the tool well.
- You don't always need to get a decisive answer.
Clearly identifying why you can't answer a question with this data or documenting an analysis roadblock is okay too.
- You should present your results clearly and understandably.

AS A TEAM:

5/5 - Complete, satisfying, and clear presentation that explains all question/results and (if necessary) your analysis.

4/5 - Complete presentation but unclear or with obvious mistakes.

3/5 - Sloppy or incomplete presentation.

0/5 - No presentation.

INDIVIDUALLY:

-2 If you're missing during presentation or analysis sessions (unless instructor-approved).

**TABLEAU
(BACKUP SLIDES)**

Specifying VISUALIZATIONS

OPERANDS ARE THE DATABASE FIELDS

- EACH OPERAND INTERPRETED AS A SET {}
- DIMENSIONS AND MEASURES TREATED DIFFERENTLY

THREE OPERATORS:

CONCATENATION (+), CROSS PRODUCT (x), NEST (/)

Table Algebra

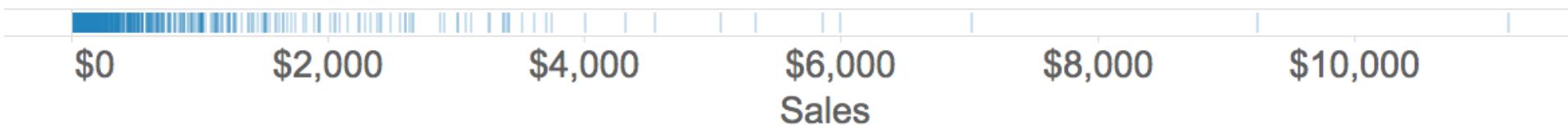
DIMENSIONS: INTERPRET DOMAIN AS A SET THAT PARTITIONS TABLE INTO ROWS AND COLUMNS.

STATE= {(ARIZONA),(ALABAMA),(CALIFORNIA),(COLORADO)} →

State						
Alabama	Arizona	Arkansas	California	Colorado	Connecticut	
61	224	60	2,001	182	82	

MEASURES: TREAT DOMAIN AS SINGLE ELEMENT SET AND ENCODE SPATIALLY AS AXES:

SALES= {SALES[0...410,000]} →



Concatenation (+) Operator

ORDERED UNION OF SET INTERPRETATIONS

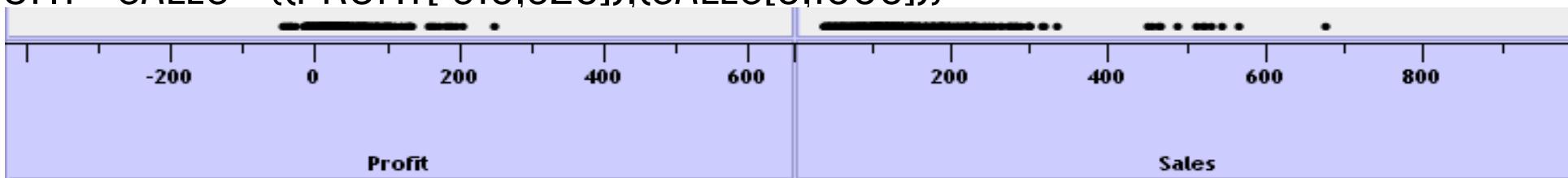
QUARTER + PRODUCT TYPE

$$= \{(QTR1), (QTR2), (QTR3), (QTR4)\} + \{(COFFEE), (ESPRESSO)\}$$

$$= \{(QTR1) (QTR2) (QTR3) (QTR4) (COFFEE) (ESPRESSO)\}$$

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

PROFIT + SALES = $\{(\text{PROFIT}[-310, 620]), (\text{SALES}[0, 1000])\}$



Cross (x) Operator

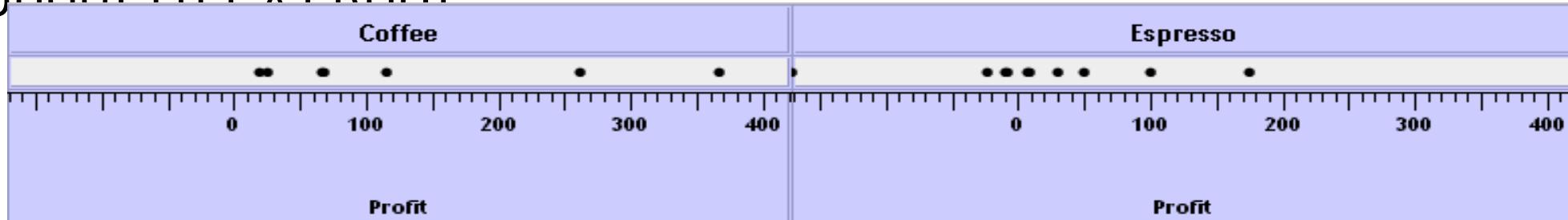
CROSS-PRODUCT OF SET INTERPRETATIONS

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)

Querying the Database

