Construction Change Order analysis

CPSC 533C Analysis Project

Presented by
Chiu, Chao-Ying
Department of Civil Engineering
University of British Columbia

Problems of Using Construction Data

- Hybrid of physical and abstract data
  - Difficult to link, model, and organize
  - Costly to link, model, and organize
- Try to model, but lack of medium to interpret and see the benefit
- Vicious cycle between model and interpret data

But, Construction Data Really Valuable for Enhancing Project Performance!

Nature of Construction Data(1)

- Original Format
  - Filled out preprinted forms, workbooks, and logs
  - Plain text documents like contracts, memorandum, e-mail, meeting minutes
  - Pictorial documents like drawings, pictures, and videos

Nature of Construction Data(2)

- Digitized Format
  - Format for Abstract Data
    - Database or spreadsheet of workbooks and logs
  - Format for Physical Data
    - Collections of digital video and picture
    - Collection of electronically stored text documents (contracts, memorandum, e-mail, meeting minutes)
    - Collection of electronic product drawings (2D Cad, 3D Cad)

Nature of Construction Data(3)

- Link to Physical Data in Abstract Data
  - Categorize Physical Data by types
    - Memorandums from **Sub-trade** vs **Client**
    - Drawings of **Interior Components** vs **Structure Components**
  - Categorize Physical Data by contents
    - Two pictures show the site conditions are **Good** vs **Bad**
    - Problems encountered described in daily site report includes **Bad Weather, Equipment Down, etc.**

Lots of physical data can be transformed into categorical abstract data!

Nature of Construction Data(4)

- Construction data can be transformed to totally abstract data for the purpose of analysis
- In the abstract data form, they are **multidimensional!**

Current Practice of Using Construction Data

- Use Excel or Access to store abstract data
- Visualization is only focused on physical data and important abstract data like cost , resource, and schedule
- Seldom are full spectrum of construction data analyzed
- Excel charts used for rare presentation situation
- Too challenging when trying to analyze more data using Excel--ad hoc manner, time consuming. Graphing rather than analyzing

Key Criteria of Visualization Tools Selection

- Alleviate the burden of retrieving and visualizing abstract data
- Shorten the time of comprehending the meaning of data represented by visual encodings
- Link between visualization of abstract data and physical data
- Comprehensive, Robust and Easy to Use. Not necessarily the best.

Why Are the Criteria?

- Multidimensional data means you need to understand them from “multi-perspectives”
- Images conveying information from “multi-perspectives” just too many
- The iterative process of retrieving data->visualizing data->observing visualization->retrieving data is sheer tedious
- Visualization of physical data help validate users’ semantic perception of it’s abstract form

Key Visualization Techniques Targeted

- Query Data Faster
  - Query by slider
  - Query by Data Dimensions
  - Query by Brush
- Generate Visualization Faster
  - Automation of Graph Generation
- More effective Visualization
  - Built-in Visual Encoding Formalism based on Psychology Ground
- Link Between Abstract and Physical Data
  - Coordinated Multiple Views (or linked data views)

Tools Selected

- Query Data Faster
  - Query by slider **(Tableau)**
  - Query by Data Dimensions **(Tableau and Advizor)**
  - Query by Brush **(Advizor)**
- Generate Visualization Faster
  - Automation of Graph Generation **(Tableau)**
- More effective Visualization
  - Built-in Visual Encoding Formalism based on Psychology Ground **(Tableau and Advizor)**
- Link Between Abstract and Physical Data
  - Coordinated Multiple View**(Tableau and Advizor; Not implement exactly yet, but have extension potential)**

Scenario Formalisms (why)

- Information technology experts use handy tools in a technical thinking; but can not read the images of domain contexts
- Domain experts knows what the images mean; but do not use the tools in the underlying technology thinking
- Scenario Formalisms bridge the gap for domain experts so that they can systematically and mechanically use tools thereby focusing on “reading information”

Scenario Formalisms (what)

- Find quantity distributions along any Dimension type dimension
- Compare quantity distributions along the same Dimension type dimension (Correlations Finding)
- Compare trend and occurrence of time dependent data
- Association between Data of Different Dimensions

Definitions of Formalisms (1)

Number	Model	Year	Color	Sales	Profit	record count
N1	Chevy	Y1990	red	5	$10,000	1
N2	Chevy	Y1990	white	87	$15,000	1
N3	Chevy	Y1990	blue	62	$13,000	1
N4	Chevy	Y1991	red	54	$11,000	1
N5	Chevy	Y1991	white	95	$9,000	1
N6	Chevy	Y1991	blue	49	$8,700	1
N7	Chevy	Y1992	red	31	$7,600	1
N8	Chevy	Y1992	white	54	$9,450	1
N9	Chevy	Y1992	blue	71	$20,000	1
N10	Ford	Y1990	red	64	$12,000	1
N11	Ford	Y1990	white	62	$9,000	1
N12	Ford	Y1990	blue	63	$8,700	1
N13	Ford	Y1991	red	52	$7,600	1
N14	Ford	Y1991	white	9	$15,000	1
N15	Ford	Y1991	blue	55	$13,000	1
N16	Ford	Y1992	red	27	$11,000	1
N17	Ford	Y1992	white	62	$12,000	1
N18	Ford	Y1992	blue	39	$9,000	1

Dimension Type dimensions

Measure Type dimensions

Definitions of Formalisms (2)

Model	Total of Sales	Average of Sales	Total of Profits	Average of Profits	Total of Record Count
Chevy	508	56	$103,750	$11,528	9
Ford	433	48	$97,300	$10,811	9

Year	Total of Sales	Average of Sales	Total of Profits	Average of Profits	Total of Record Count
Y1990	343	57	$67,700	$11,283	6
Y1991	314	52	$64,300	$10,717	6
Y1992	284	47	$69,050	$11,508	6

"Color" to "Profit" : 1 to many

Color	Total of Sales	Average of Sales	Total of Profits	Average of Profits	Total of Record Count
red	233	39	$59,200	$9,867	6
white	369	62	$69,450	$11,575	6
blue	339	57	$72,400	$12,067	6

Statistics Values of Measure type dimension "Profit" for Dimension type dimension "car color"

Called "Quantity Measurement"!!

"Number" to "sales" : 1 to 1

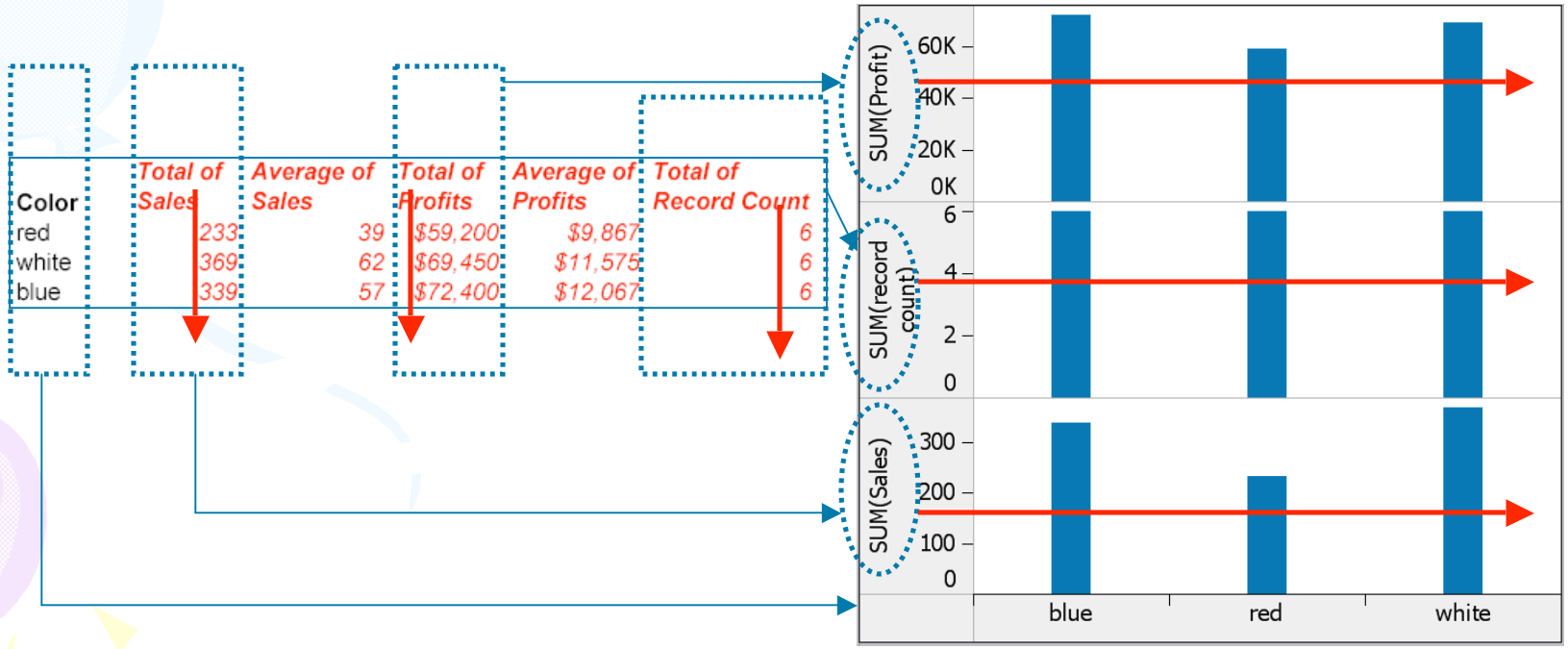
Number	Sales	Profit	record count
N1	5	$10,000	1
N2	87	$15,000	1
N3	62	$13,000	1
N4	54	$11,000	1
N5	95	$9,000	1
N6	49	$8,700	1
N7	31	$7,600	1
N8	54	$9,450	1
N9	71	$20,000	1
N10	64	$12,000	1
N11	62	$9,000	1
N12	63	$8,700	1
N13	52	$7,600	1
N14	9	$15,000	1
N15	55	$13,000	1
N16	27	$11,000	1
N17	62	$12,000	1
N18	39	$9,000	1

Values of Measure type dimension "Sales" for Dimension type dimension "sales number"

Definitions of Formalisms (3)

Quantity Distribution of Data along Dimension Type Dimension:

The distribution of Quantity Measurements along a Dimension type dimension

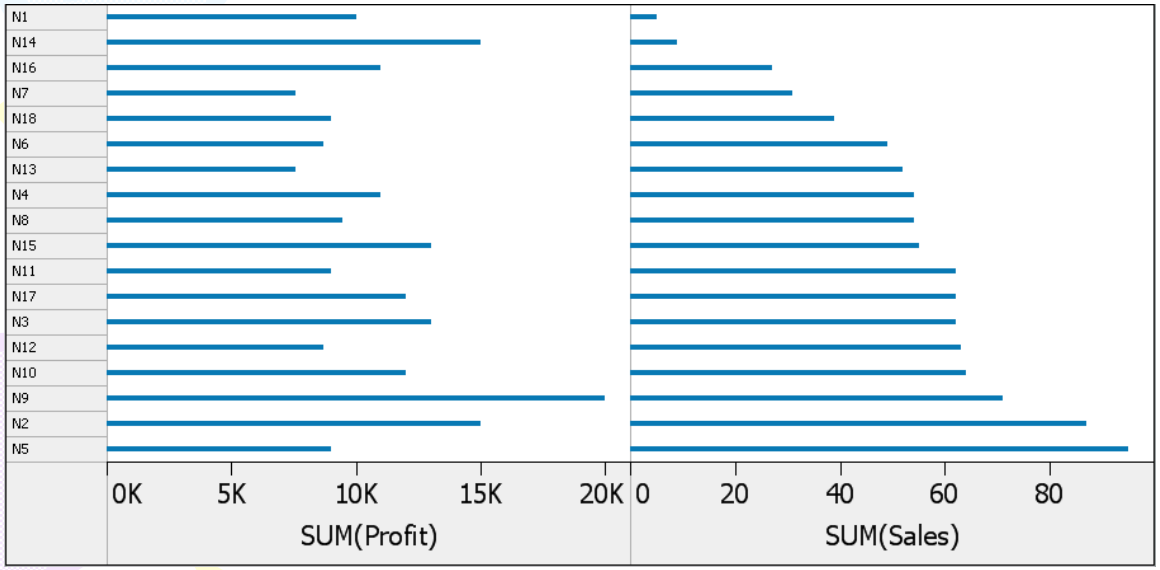


Find quantity distributions along any Dimension type dimension

- Find in the whole dataset
- Find in the subset of data

Compare quantity distributions along the same Dimension type dimension(1)

- Same data.
- Different quantity measurements



Number	Sales	Profit	record count
N1	5	$10,000	1
N2	87	$15,000	1
N3	62	$13,000	1
N4	54	$11,000	1
N5	95	$9,000	1
N6	49	$8,700	1
N7	31	$7,600	1
N8	54	$9,450	1
N9	71	$20,000	1
N10	64	$12,000	1
N11	62	$9,000	1
N12	63	$8,700	1
N13	52	$7,600	1
N14	9	$15,000	1
N15	55	$13,000	1
N16	27	$11,000	1
N17	62	$12,000	1
N18	39	$9,000	1

Different quantity measures

Compare quantity distributions along the same Dimension type dimension(2)

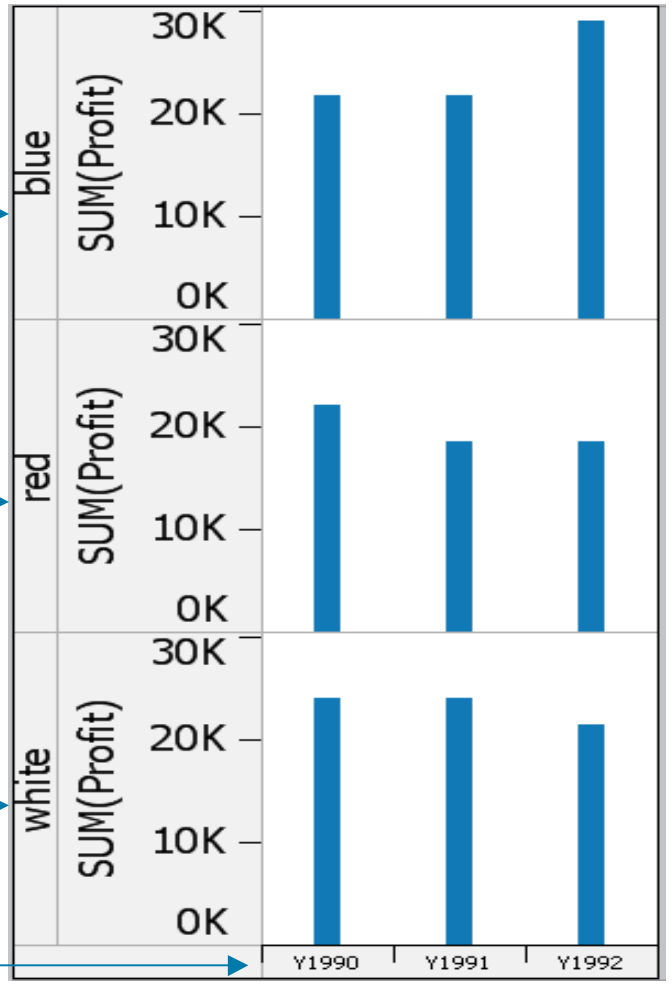
- Different subsets of data filtered from same dimension
- Same quantity measurement

Filter this dimension

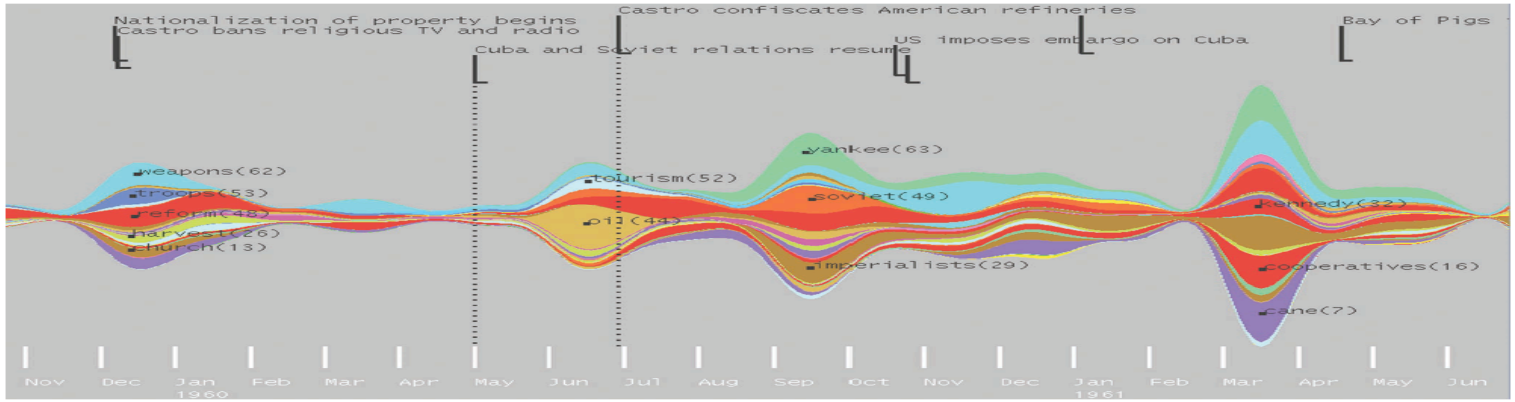
Color	Total of Sales	Average of Sales	Total of Profits	Average of Profits	Total of Record Count
red	233	39	$59,200	$9,867	6
white	369	62	$69,450	$11,575	6
blue	339	57	$72,400	$12,067	6

Year	Total of Sales	Average of Sales	Total of Profits	Average of Profits	Total of Record Count
Y1990	343	57	$67,700	$11,283	6
Y1991	314	52	$64,300	$10,717	6
Y1992	284	47	$69,050	$11,508	6

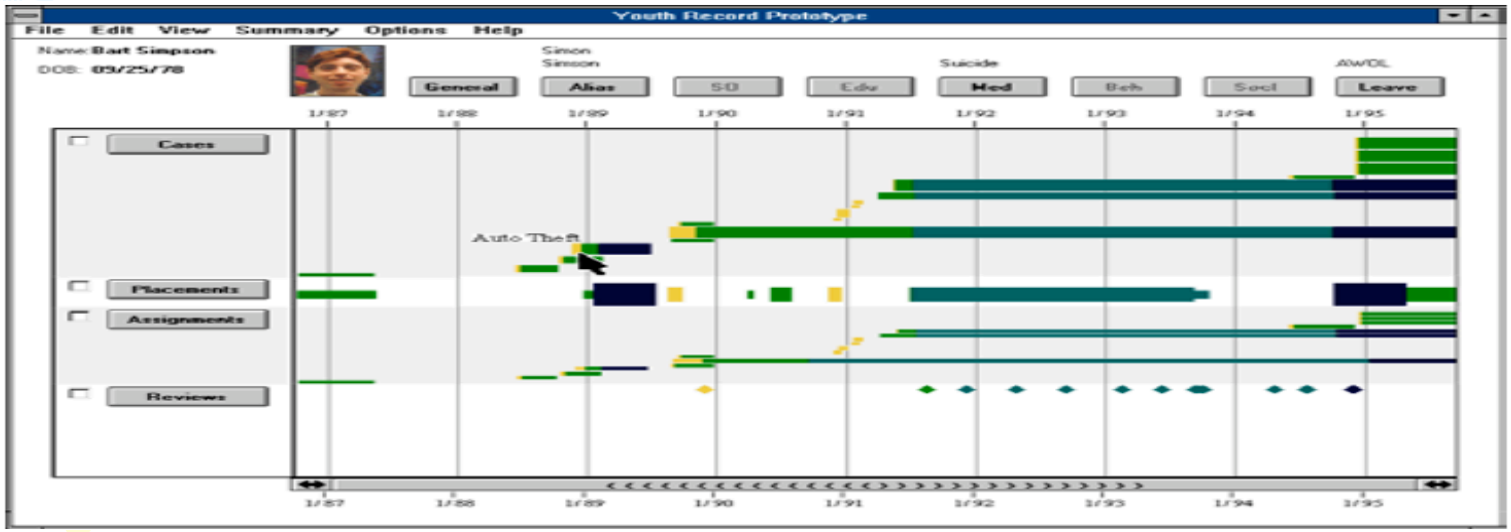
Along this dimension



Compare Trend and Occurrence of Time Dependent Data



Trend



Occurrence

Association Between Data of Different Dimensions

- Similar to finding quantity distribution (first type of scenario formalism)
- But quantity is not our concern now.
- Find “which data has something to do with what data”.

Dataset of Change Order Domain

Change Order Registry							
Data Field Name	*Change Order Number*	*Issued Date*	*Projected Cost*	*Date Approved*	*Approved Cost*	*Reason of Change*	*Reference Number*
Data Abstration Type	Raw	Raw	Raw	Raw	Raw	Raw	Raw
Data Field Type	Dimension	Dimension (Time)	Measure	Dimension (Time)	Measure	Dimension	Dimension

Affected Sub-Trade				
Data Field Name	*Change Order Number*	*Trade*	*Trade Revision Number*	*Trade Change Order Amount*
Data Abstration Type	Raw	Raw	Raw	Raw
Data Field Type	Dimension	Dimension	Dimension	Measure

Initiated Document		
Data Field Name	*Change Order Number*	*Initiated Document*
Data Abstration Type	Raw	Raw
Data Field Type	Dimension	Dimension

Affected Physical Location			
Data Field Name	*Change Order Number*	*Location*	*Sub-location*
Data Abstration Type	Raw	Raw	Raw
Data Field Type	Dimension	Dimension	Dimension

Affected Physical Component				
Data Field Name	*Change Order Number*	*Major Group Elements*	*Group Element*	*Individual Element*
Data Abstration Type	Raw	Raw	Raw	Raw
Data Field Type	Dimension	Dimension	Dimension	Dimension

Scenario Formalism+ Visualization Tools(1.1)

- Scenario Formalism:

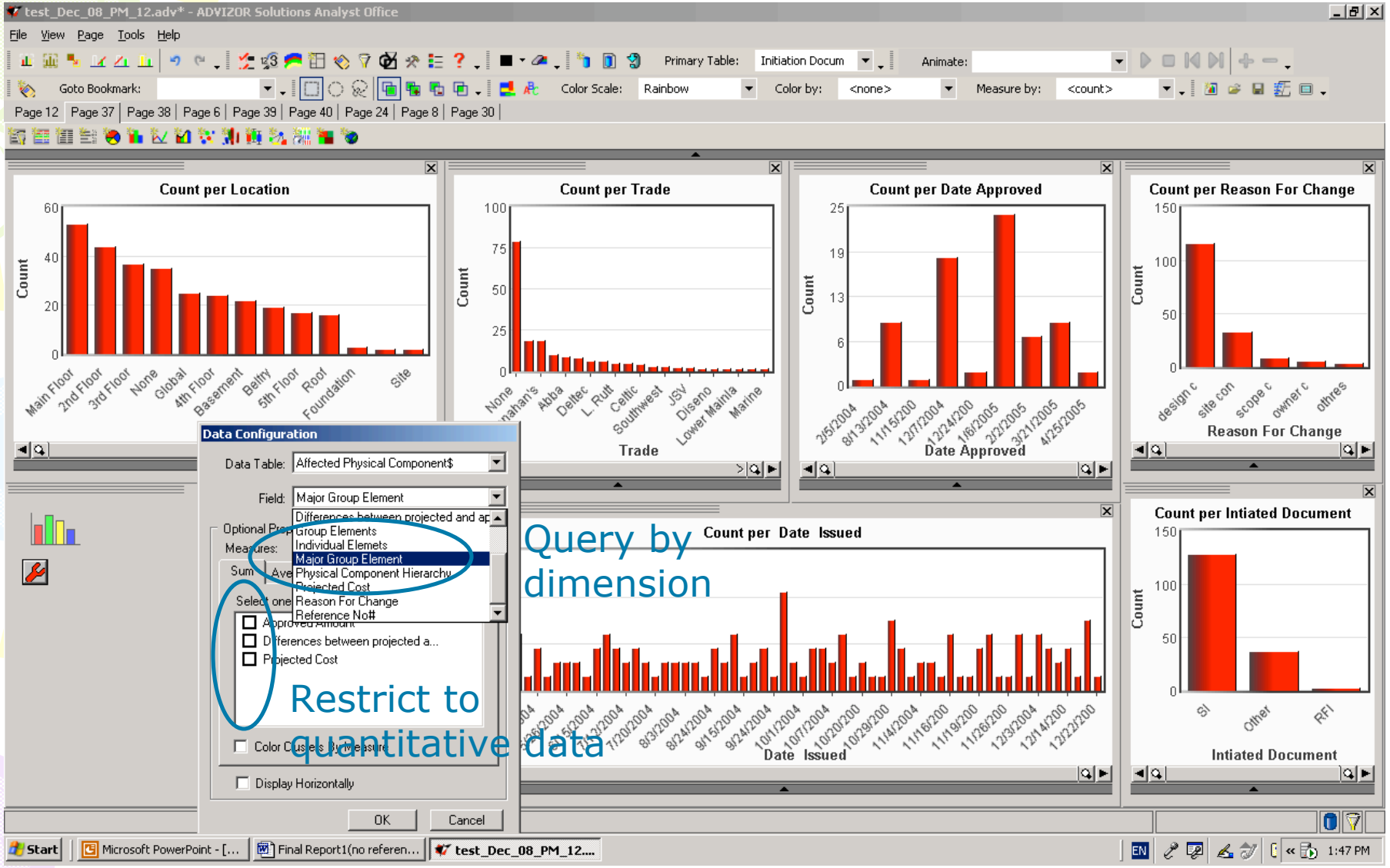
  Find quantity distribution

- Tool:

  Adivzor (better in gaining overview)

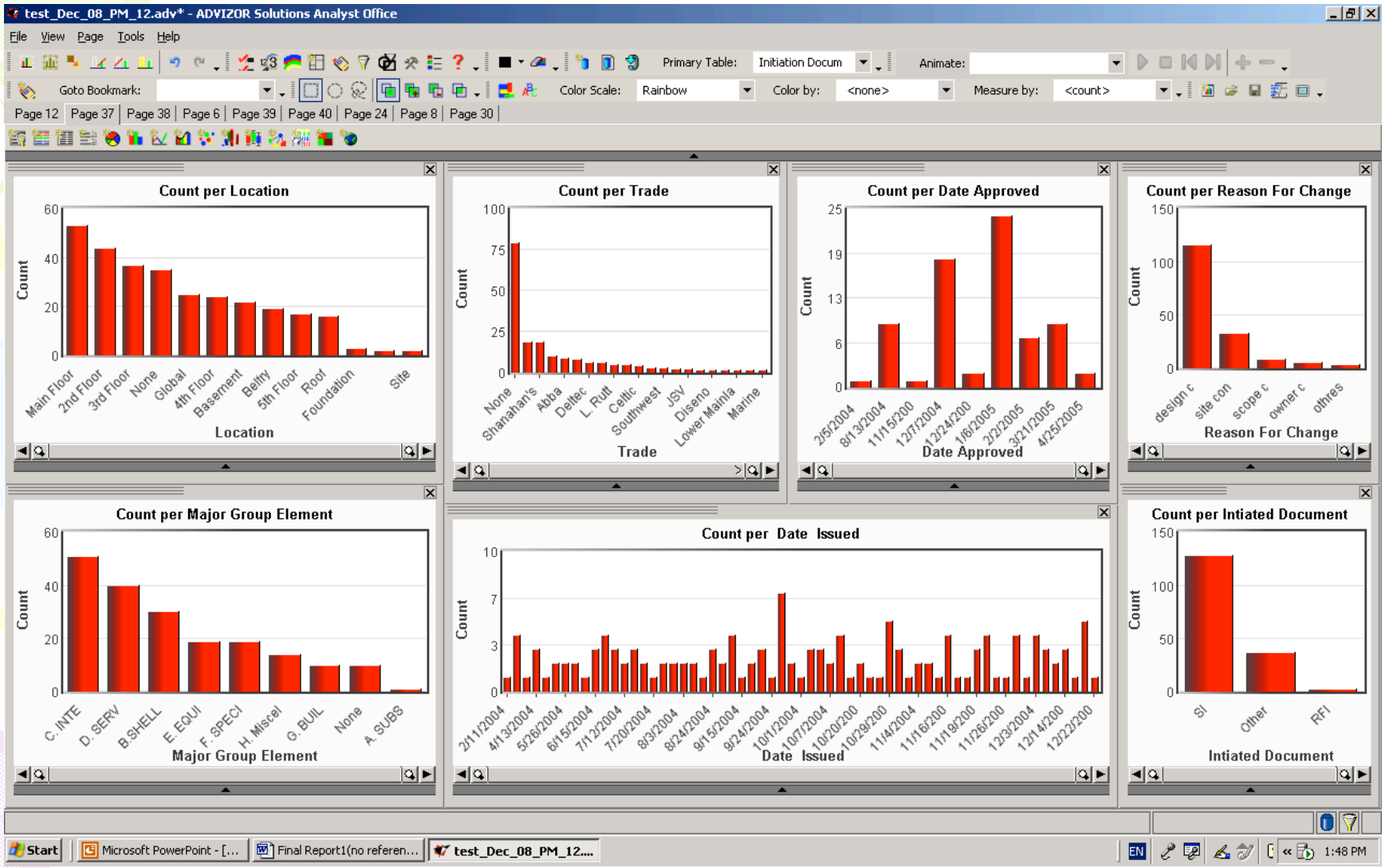
- Key Techniques:

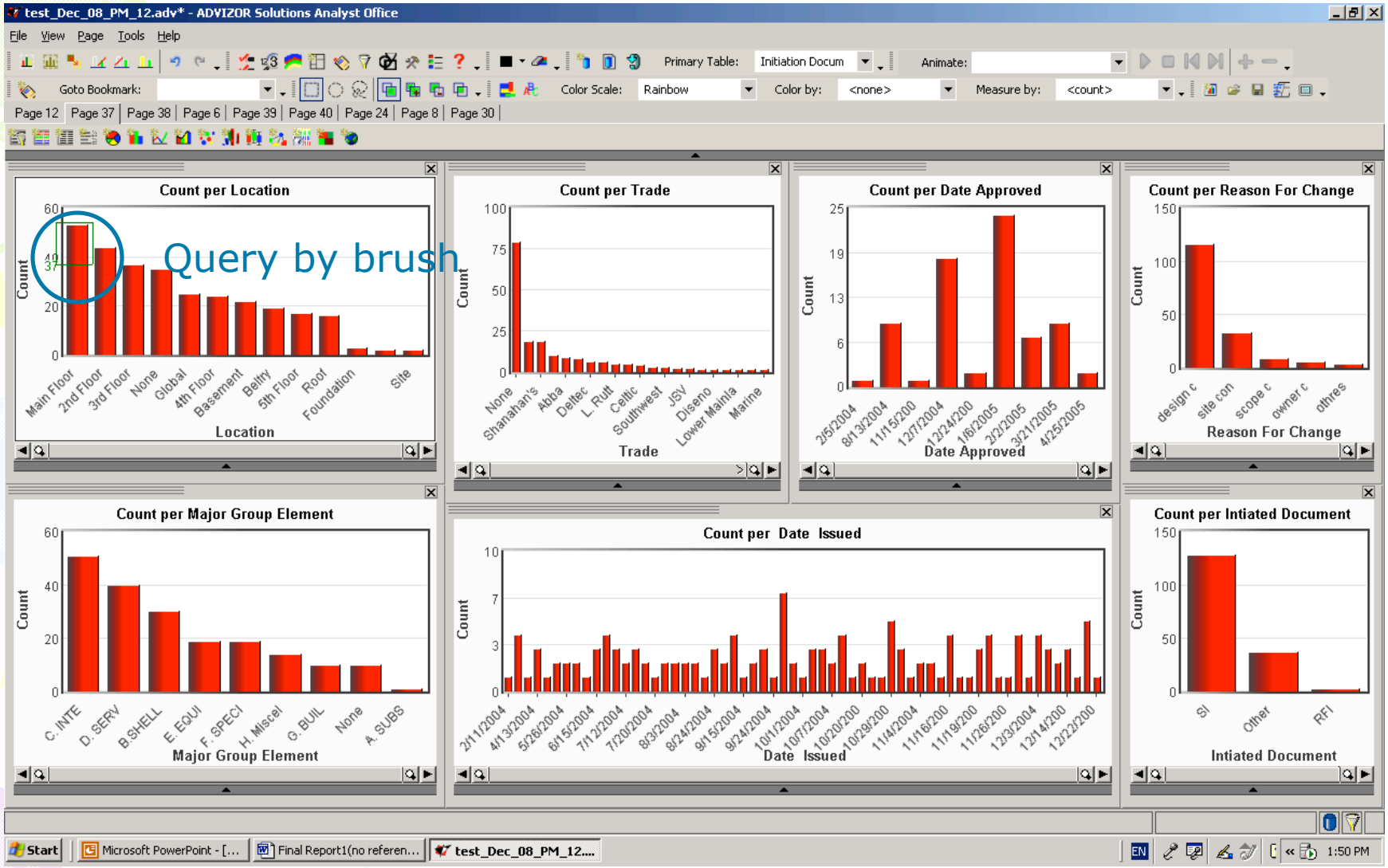
  Query by dimensions, query by brush, linked data views, visual encoding formalism



Create 7 bar charts to see quantity distributions of 7 dimensions

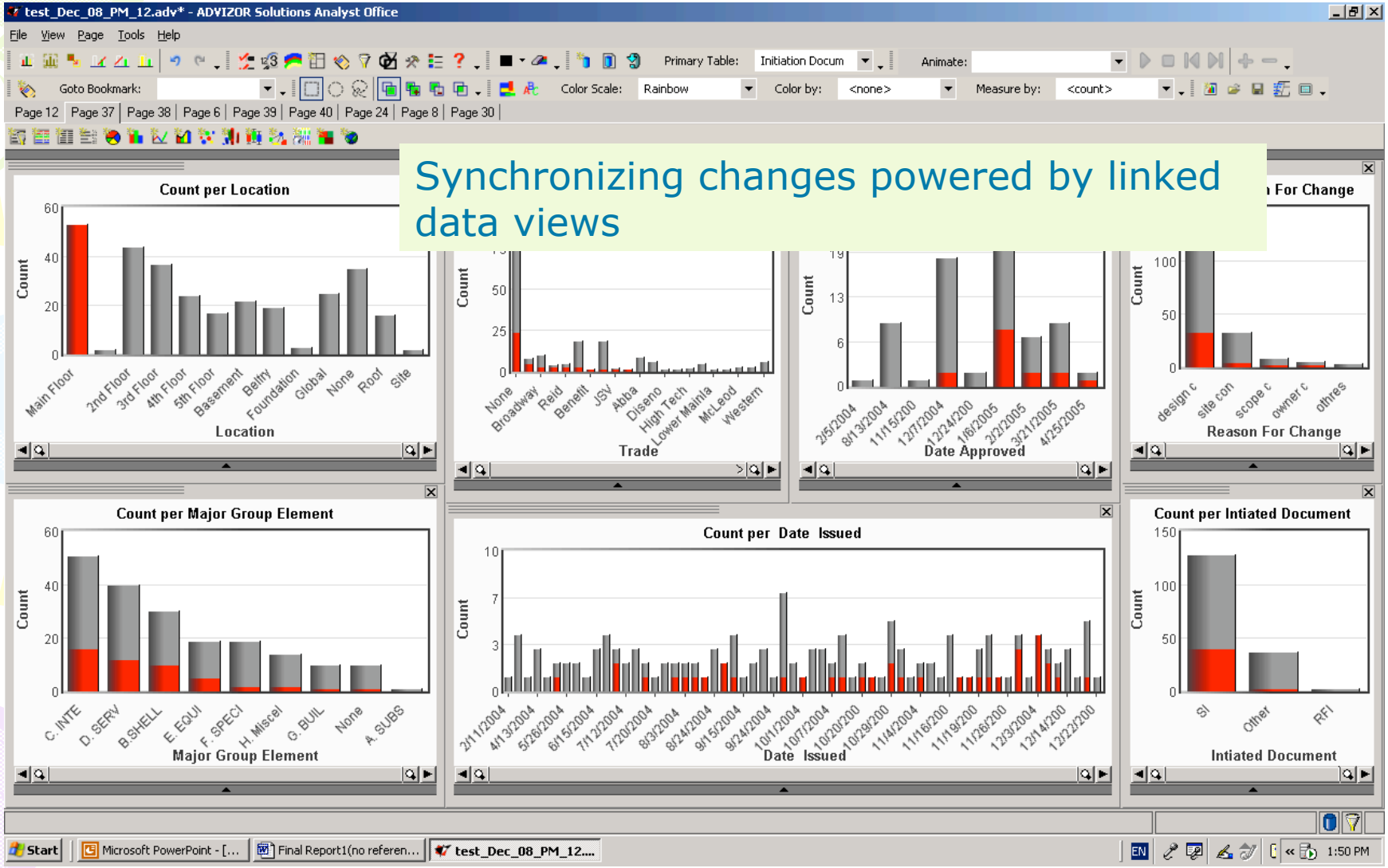
Quantity Distributions of the whole dataset along with different Dimension type dimensions





Filter to the subset of data whose location is “main floor”

Quantity Distributions of the subset of dataset along with different Dimension type dimensions



Scenario Formalism+ Visualization Tools(1.2)

- Information Extracted:
  - 9~12 information pertinent to change order obtained
  - Example Information: Interior construction and service construction encounter more change orders than other components of the building
- Time Spent on Retrieving and Visualizing : Less Than 5 minutes (excluding one time overhead of setting environment)

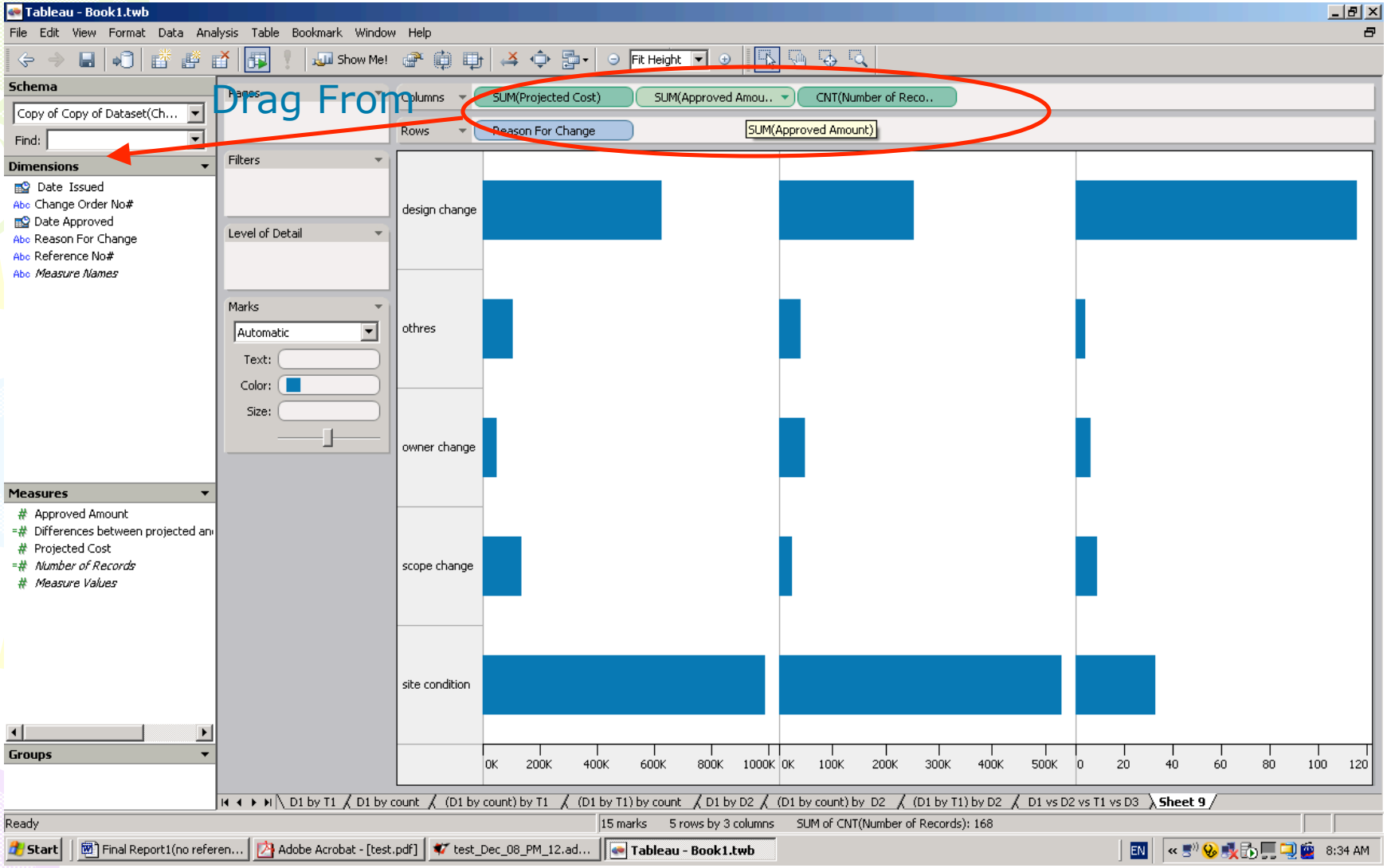
Scenario Formalism+ Visualization Tools(2.1)

- Scenario Formalism:

  Compare quantity distribution
- Tool:

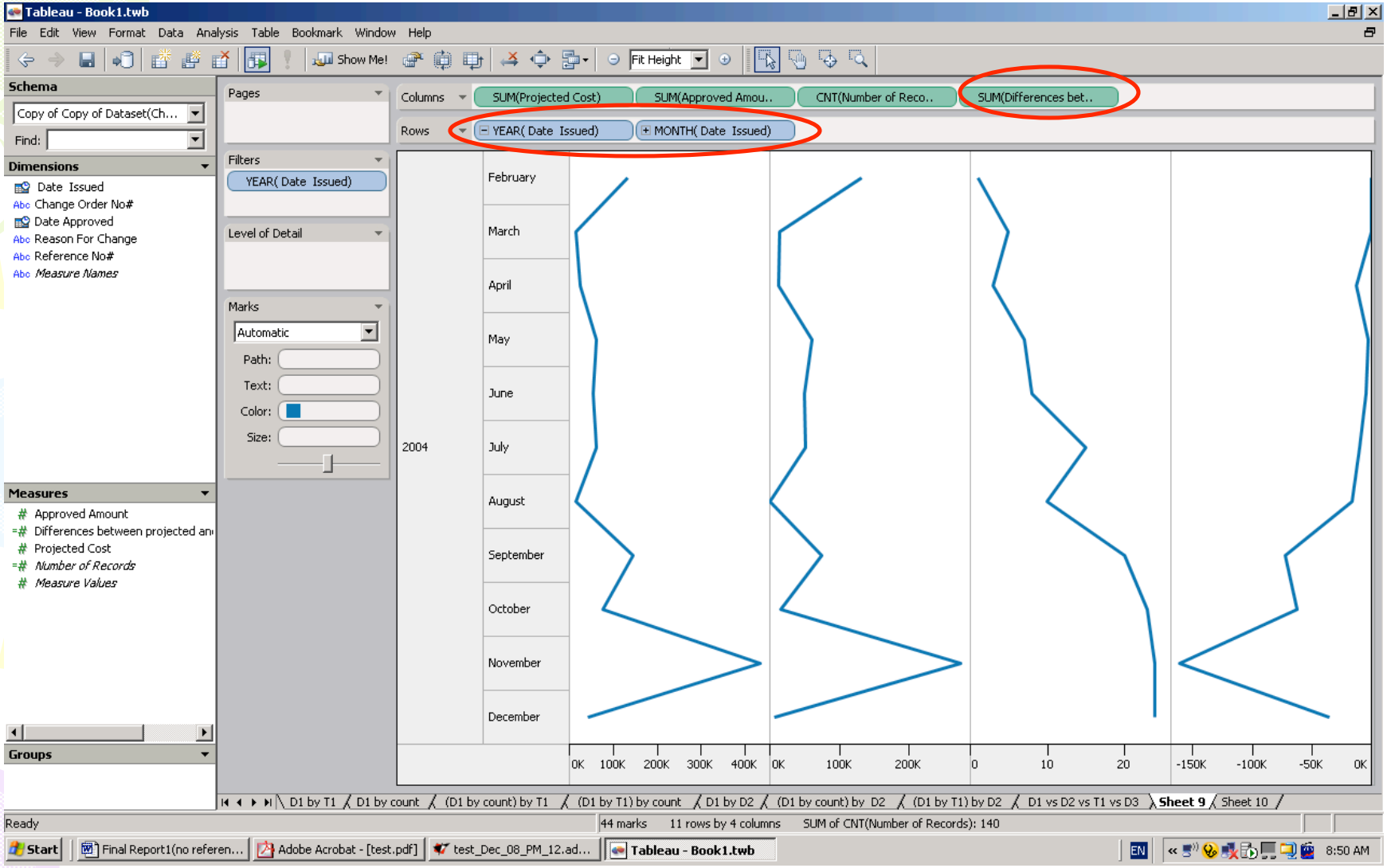
  Tableau (better in trend comparisons)
- Key Techniques:

  Query by dimensions, Visualization Automation, Visual Encoding Formalism

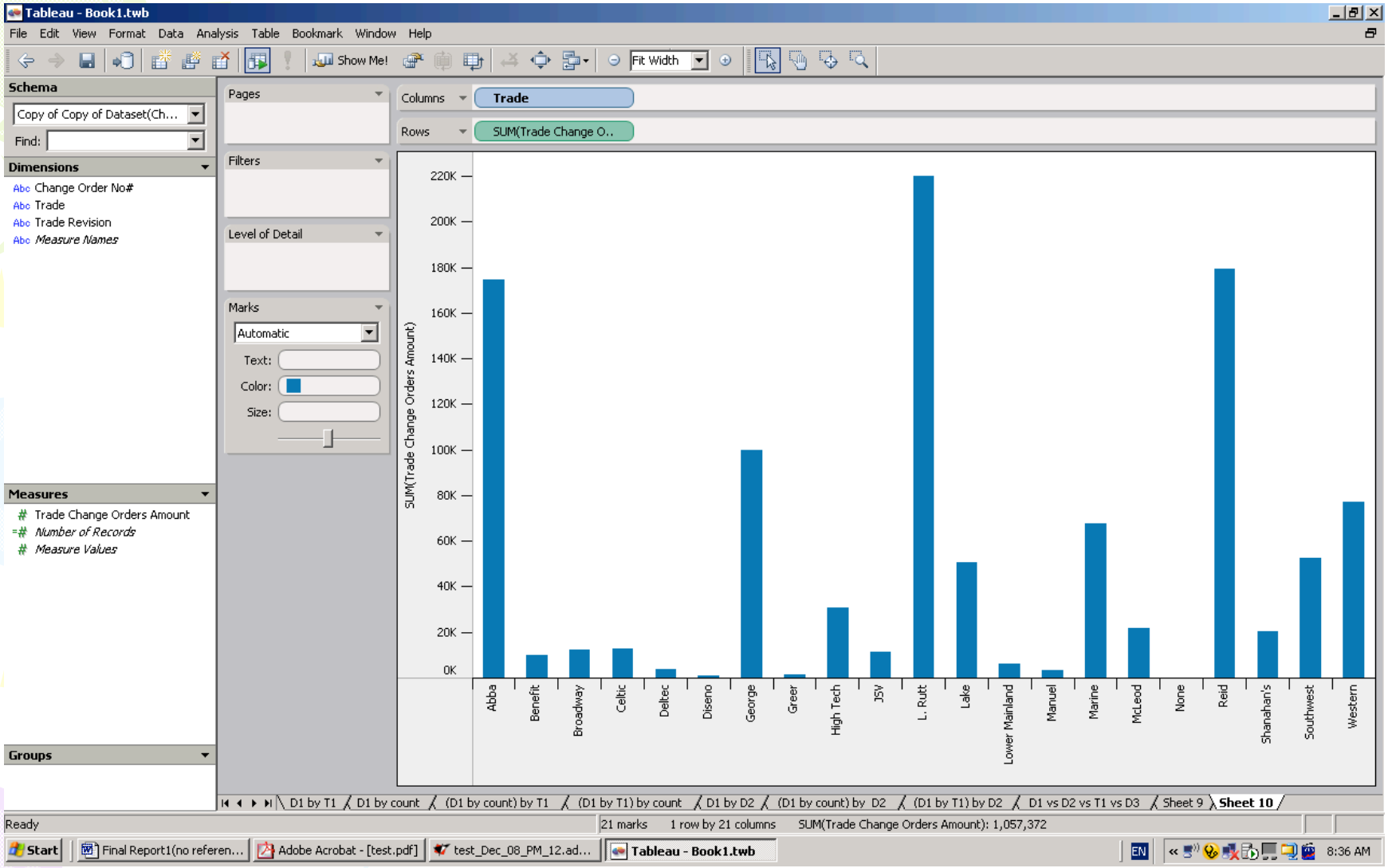


Drag Measure type dimension of “projected cost”, “approved cost”, and “record count” to Column Shelf, instruct the system to total these three dimensions, and drag the Dimension type dimension “reason of change” to Row Shelf

First type of quantity distribution comparison

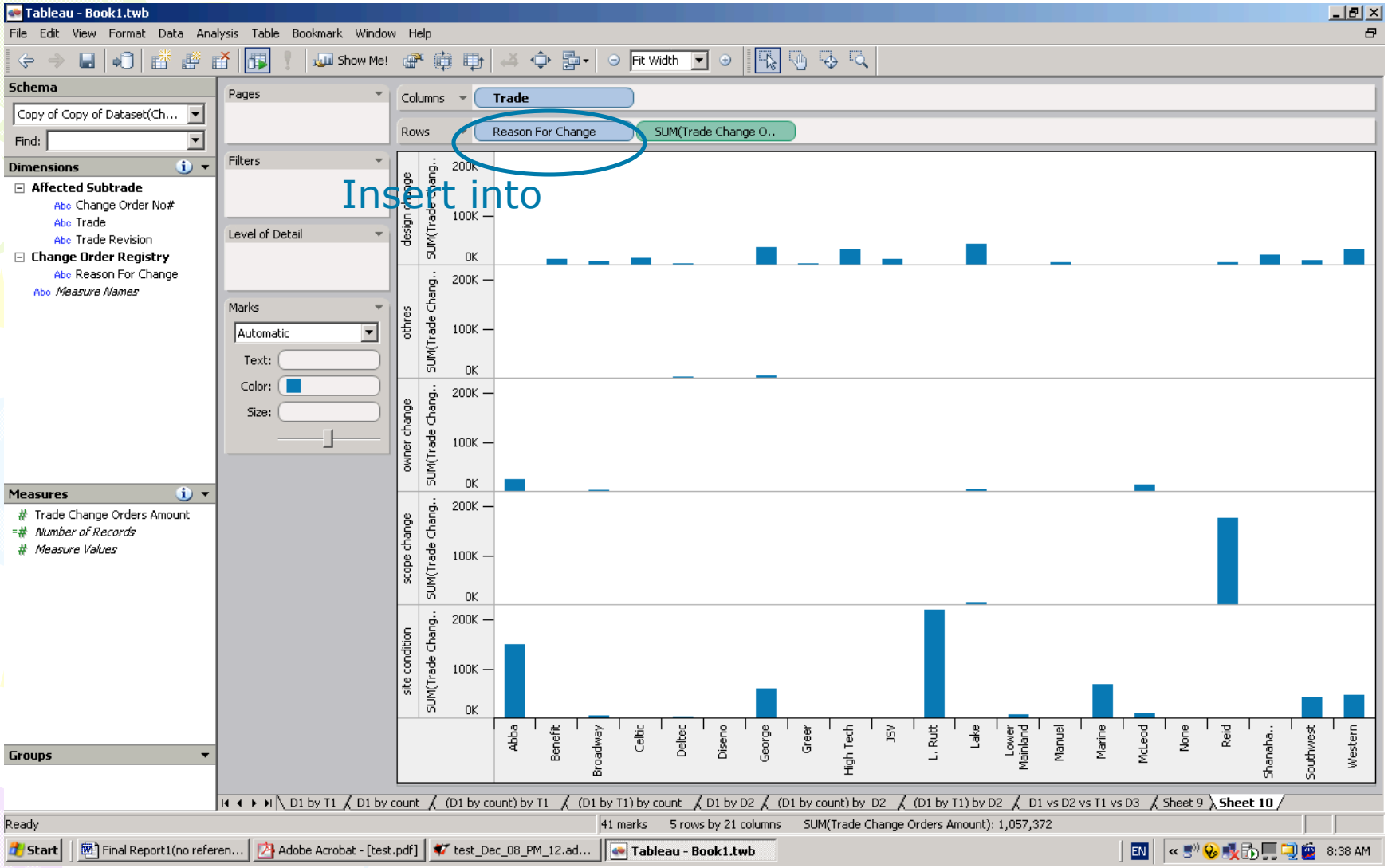


Replace the Dimension type dimension "reason of change" on the Row Shelf by "issued date". Also add another Measure type dimension "difference between projected and approved cost" **(same data, along same Dimension dimension, different quantity measurement)**



Drag Dimension type dimension “sub-trade” and Measure type dimension “trade change order cost” to Column Shelf and Row Shelf respectively. Also we instruct the system to do total on “trade change order cost”

Second type of quantity distribution comparison



Insert a different Dimension type dimension "reason of change" in front of the Measure type dimension "trade change order cost" on the Row Shelf. **(Align different subsets of data, along same Dimension dimension, same quantity measurement)**

Scenario Formalism+ Visualization Tools(2.2)

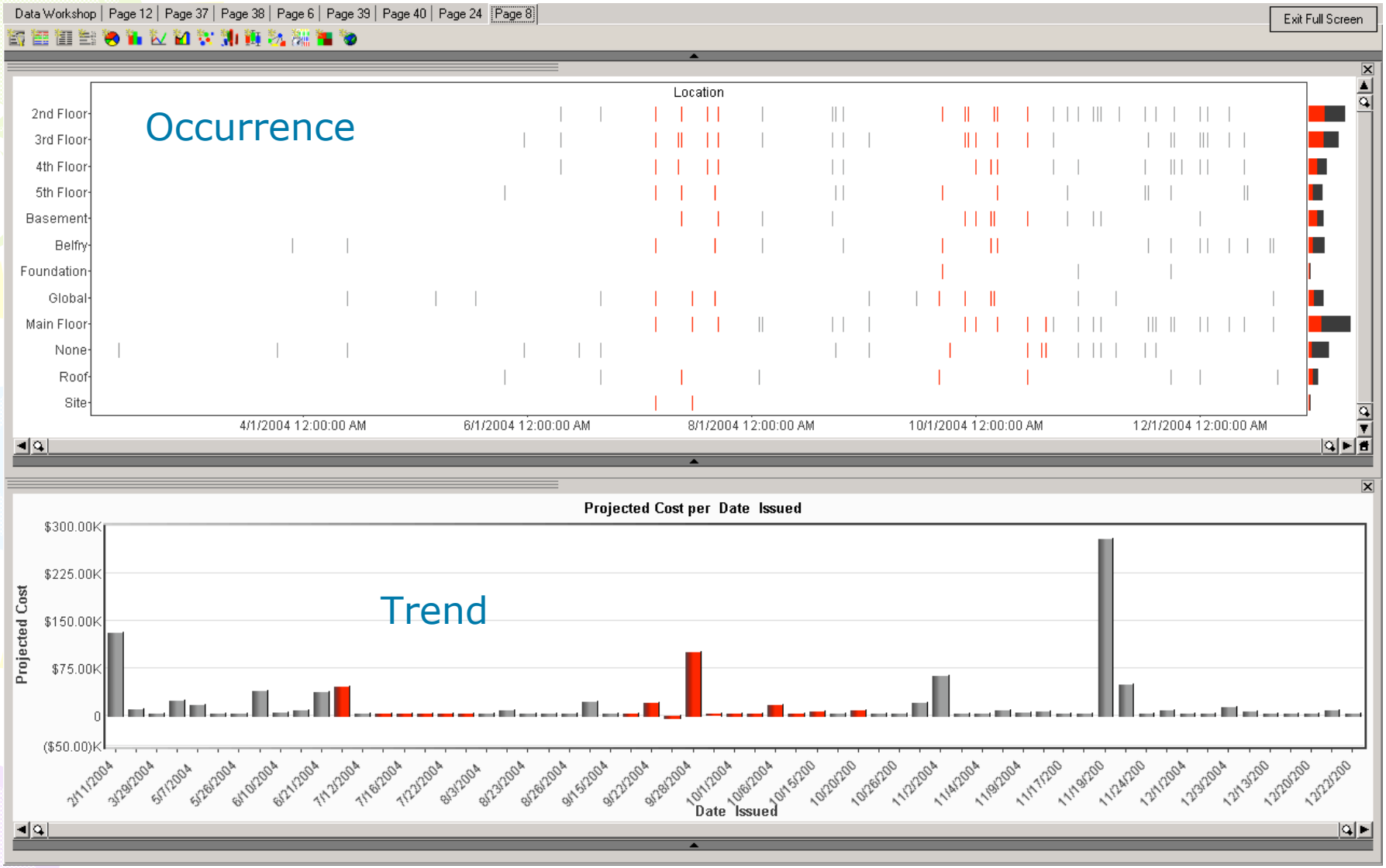
- Information Extracted:
  - 3 information pertinent to change order obtained
  - Example Info.: Client has no problem of approving request of extra cost at the beginning. However, along the increases of change orders, the amount of disagreement starts to rise!
- Time Spent on Retrieving and Visualizing: Less Than 2 minutes (excluding one time overhead of setting environment)

Scenario Formalism+ Visualization Tools(3.1)

- Scenario Formalism:

  Compare trend and occurrence of time dependent data
- Tool: Advizor
- Key Techniques:

  Query by dimensions, query by brush, linked data views, visual encoding formalism



Juxtapose Timetable chart with bar chart. Brush the clustering visual encodings in the TimeTable, corresponding data in the bar chart are highlighted too.

Scenario Formalism+ Visualization Tools(3.2)

- Information Extracted:
  - 2 information pertinent to change order obtained
  - Example Info.: There are two periods of time when the change orders involve almost all locations of the building. During those two time period, one sharply increase of projected cost is observed!!
- Time Spent on Retrieving and Visualizing: Less Than 2 minutes (excluding one time overhead of setting environment)

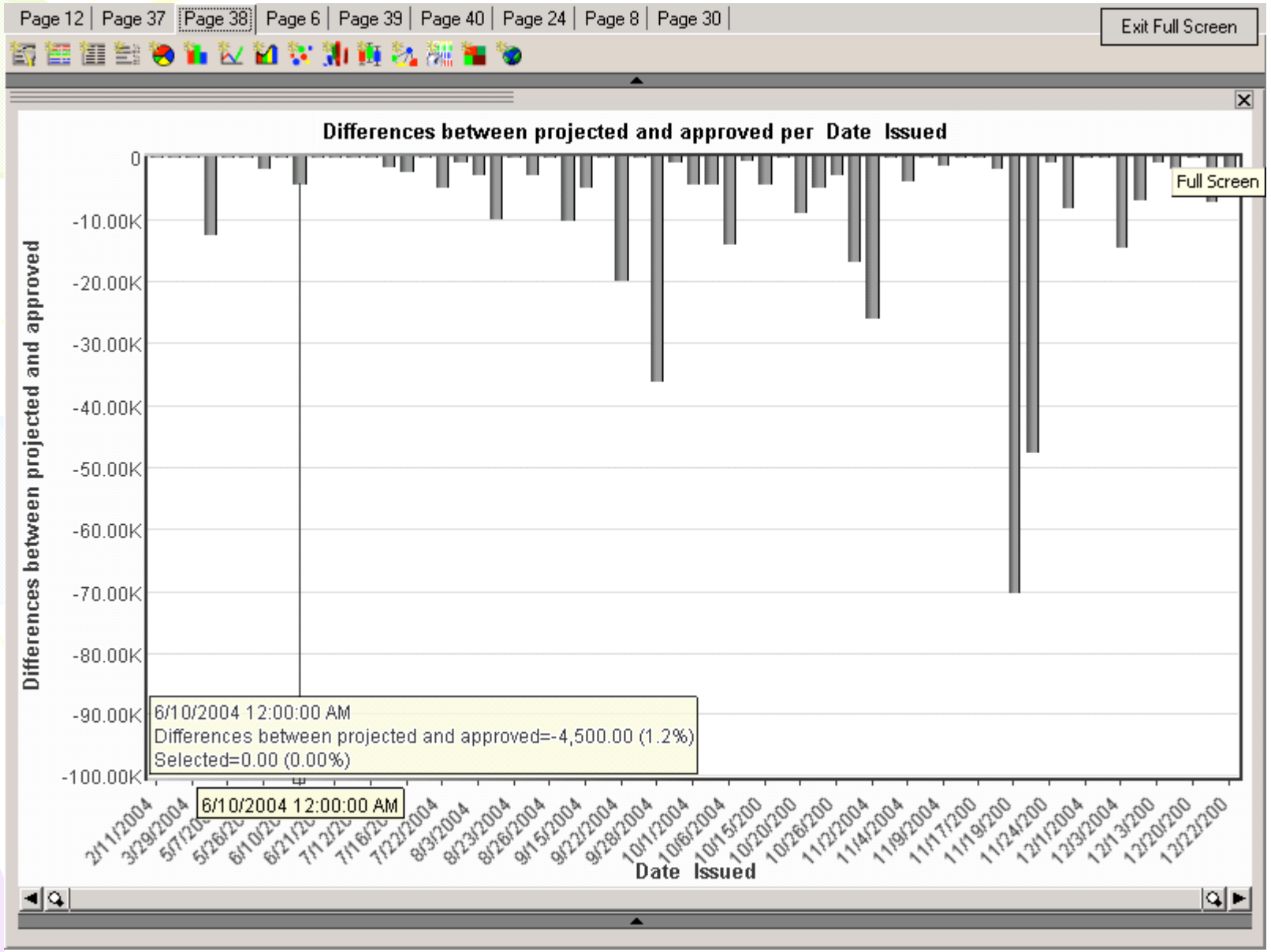
Scenario Formalism+ Visualization Tools(4.1)

- Scenario Formalism:

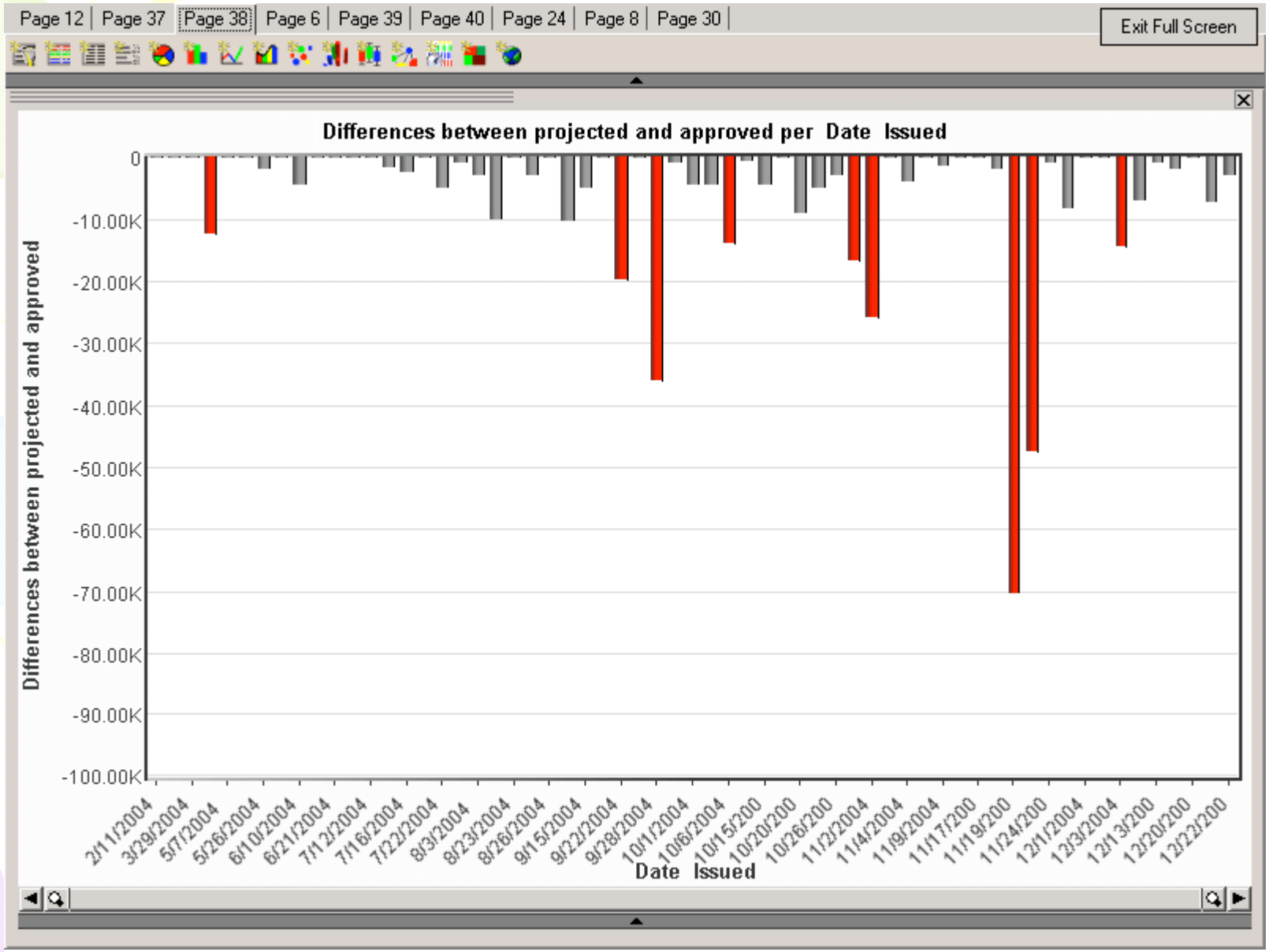
  Association between data of different dimensions

- Tool: Advizor
- Key Techniques:

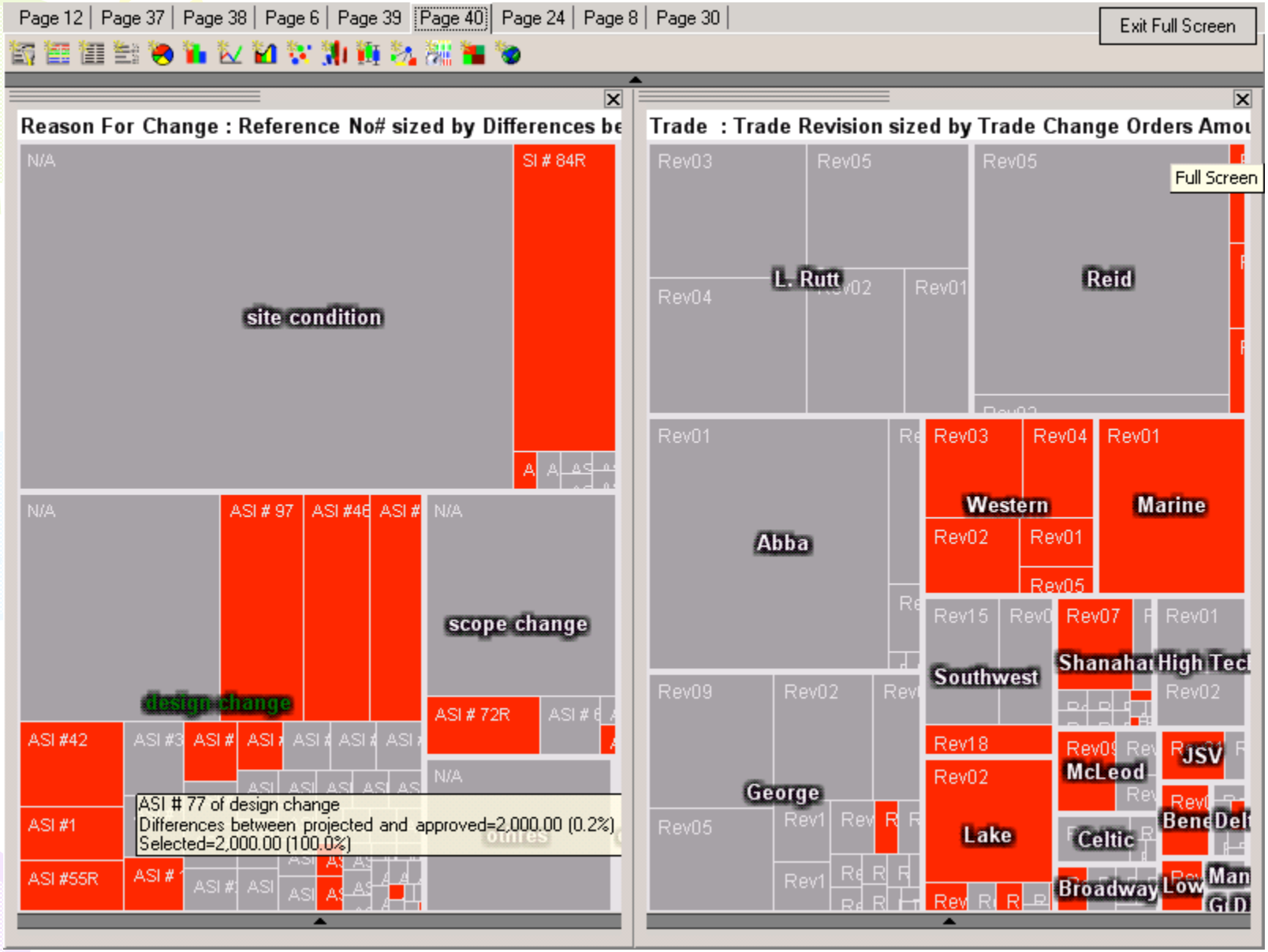
  Query by dimensions, query by brush, linked data views, visual encoding formalism, TreeMap



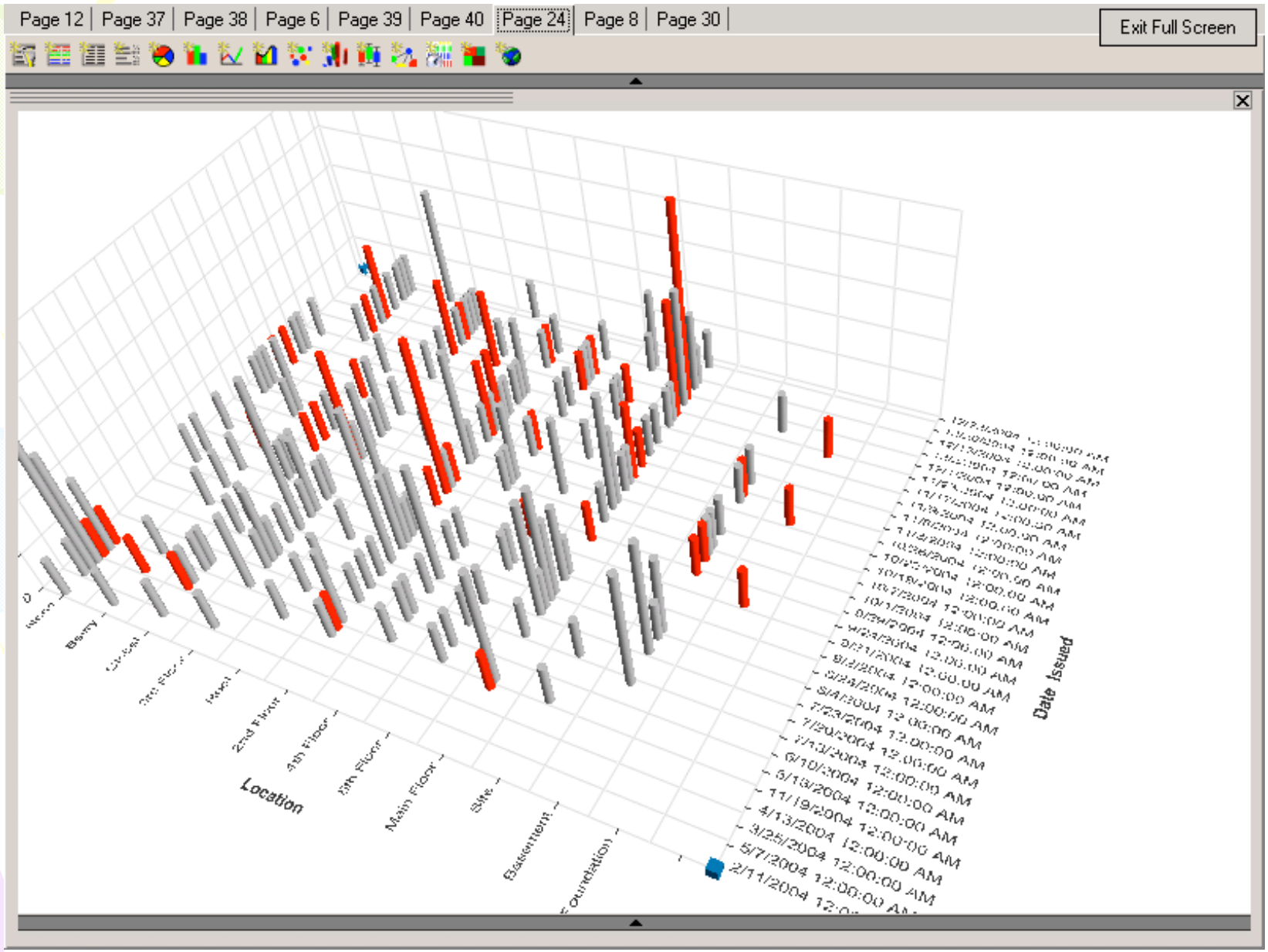
Using bar chart to see the quantity distribution along “issued date” in terms of “differences between projected cost and approved cost”



Brush to select data that have over $10,000 of differences of projected and approved cost



Switch to already generated Heatmap chart that visualize data of “reason of change” and “initiated documents”; data of of “affected sub-trade” and “sub-trade revision number”. All related data values are highlighted.



Switch to the already generated Multiscape chart that visualizes related data of combinations of two dimensions (time and space)

Scenario Formalism+ Visualization Tools(4.2)

- Information Extracted:
  - 5 information pertinent to change order obtained
  - Example Info.: Of those change orders that the client strongly disagree with, we find the coincidental time and space in which those change orders occurs
- Time Spent on Retrieving and Visualizing: Less Than 3 minutes (excluding one time overhead of setting environment)

Conclusion

- Difficult to say which still images better (infovis technology generated vs Excel or even hand made)
- Big improvement made mostly by technology of interaction
- We specifically value “linked data view” most:
  - It help link data and give overview of data
  - It give flexibility of graphing (no need to consolidate nD into a 2D or 3D)
  - It bridge the infovis with scientific visualization, which is major focus of engineering field.
- Scenario formalism further enhance improvement